Data-Driven Modelling of Perceptual Properties of 3D Shapes

A DISSERTATION PRESENTED
BY
KAPIL DEV
TO
THE SCHOOL OF COMPUTING AND COMMUNICATIONS

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN THE SUBJECT OF
COMPUTER SCIENCE

LANCASTER UNIVERSITY
LANCASTER, LANCASHIRE
JUNE 2018

Data-Driven Modelling of Perceptual Properties of 3D Shapes

Abstract

The recent surge in 3D content generation has led to the evolution of difficult to search, organise and re-use massive online 3D visual content libraries. We explore crowdsourcing and machine learning techniques to help alleviate these difficulties by focusing on the visual perceptual properties of 3D shapes. We study "style similarity" and "aesthetics" as two fundamental perceptual properties of 3D shapes and build data-driven models. We rely on crowdsourcing platforms to collect large number of human judgements on style matching and aesthetics of 3D shapes. The judgement data collected directly from humans is used to learn metrics of style matching and aesthetics.

Our style similarity measure can be used to compute style distance between a pair of input 3D shapes. In contrast to previous work, we incorporate colour and texture in addition to geometric features to build a colour and texture aware style similarity metric. We also experiment with learning objective and personalised style metrics 3D shapes. The application prototypes we build demonstrate the use of style based search and scene composition. Further, our style distance metric is built iteratively to consume lesser amount of human style judgement data compared to previous methods.

We study the problem of building a data-driven model of 3D shape aesthetics in two steps. We first focus on designing a study to crowdsource human aesthetics judgement data. We then formulate a deep learning based strategy to learn a measure of 3D shape aesthetics from collected data. The results of the study in first step helped us choose an appropriate shape representation i.e. voxels as an input to deep neural networks for learning a measure of visual aesthetics. In the same crowdsourcing study, we experiment with the use of polygonal, volumetric, and point based shape representations to create shape stimuli to collect and compare human shape aesthetics judgements. On analysis of the collected data we found that that humans can reliably distinguish more aesthetic shape in a pair even from coarser shape representations such as voxels. This observation implies that detailed shape representations are not needed to compare aesthetics in pairs.

The aesthetic value of a 3D shape has traditionally been explored in terms of specific visual features (or handcrafted features) such as curvature and symmetry. For example, more symmetric and curved shapes are considered aesthetic compared to less curved and symmetric shapes. We call such properties as pre-existing notion (or rules) of aesthetics. In order to develop a measure of perceptual aesthetics of 3D shapes which is independent of any pre-existing notion or shape features, we train deep neural networks directly on human aesthetics judgement data. We demonstrate the usefulness of the learned measure by designing applications to rank a collection of shapes based on their aesthetics scores and interactively build scenes using shapes with high aesthetics scores.

The overarching goal of this thesis is to demonstrate the use of machine learning and crowdsourcing approaches to build data-driven models of visual perceptual properties of 3D shapes for applications in search, organisation, scene composition, and visualisation of 3D shape data present in ever increasing online 3D shape content libraries. We believe that our exploration of perceptual properties of 3D shapes will motivate further research by looking into other important perceptual properties related to our vision system and will also fuel development of techniques to automatically enhance such properties of a given 3D shape.

Family and Friends

Acknowledgments

This thesis is the result of direct and indirect work and support of many, especially the below mentioned and who remain unnamed. First and foremost, I am deeply indebted to my advisor, Prof. Manfred Lau, not just for supporting my ideas but also for fruitful discussions that contributed to this thesis. He has remained a source of constant motivation and encouragement for me due to his high standards of research excellence. Second, I would like to thank Dr. Nicolas Villar, Microsoft Research Cambridge and acknowledge Microsoft PhD Scholarship that supported the research presented in this thesis. Third, special thanks goes to Dr. Kwang In Kim for his inputs to one of our projects, which got me started in machine learning and to Luther Power for sharing his desktop computer for running deep neural networks. Fourth, I greatly appreciate Prof. Hans Gellersen and Dr. Yukun Lai for giving constructive comments to uplift the quality of this thesis. Finally, I would like to thank my mother, my wife, and my daughter, for their precious emotional support and encouragement, especially my patient wife who has spent almost four years in small university flats by helping me stretch my student stipend very far.

Declaration

I, Kapil Dev, declare that the work presented in this thesis titled, "Data-Driven modelling of Perceptual Properties of 3D Shapes" is my own. I confirm that where I have consulted the published work of other researchers, this is clearly indicated in the thesis.

Kapil Dev
June 2018

Publications

- *Kapil Dev*, Nicolas Villar and Manfred Lau. 2017. Polygons, Points, or Voxels? Stimuli Selection for Crowdsourcing Aesthetics Preferences of 3D Shape Pairs. *In Proceedings of the Symposium on Computational Aesthetics*.
- *Kapil Dev*, Manfred Lau, and Ligang Liu. 2016. A Perceptual Aesthetics Measure for 3D Shapes. *arXiv preprint arXiv:1608.04953*.
- *Kapil Dev*, Nicolas Villar, and Manfred Lau. 2016. StandUp: understanding body-part and gestural preferences for first-person 3D modeling. *In Proceedings of the Joint Symposium on Computational Aesthetics and Sketch Based Interfaces and Modeling and Non-Photorealistic Animation and Rendering, Eurographics Association*.
- *Kapil Dev* and Manfred Lau. 2015. Democratizing Digital Content Creation Using Mobile Devices with Inbuilt Sensors. *IEEE Computer Graphics and Applications, 35, 1, 84-94*.
- *Kapil Dev* and Manfred Lau. 2016. Improving Style Similarity Metrics of 3D Shapes. *In Proceedings of the 42nd Graphics Interface Conference, Canadian Human-Computer Communications Society*.
- Manfred Lau, *Kapil Dev*, Julie Dorsey, and Holly Rushmeier. 2016. Learning a Human-Perceived Softness Measure of Virtual 3D Objects. *In Proceedings of the ACM Symposium on Applied Perception*.
- Manfred Lau, *Kapil Dev*, Weiqi Shi, Julie Dorsey, Holly Rushmeier. 2016. Tactile Mesh Saliency. *ACM Transactions on Graphics, 35, 4, 52:1-52:11*.

Contents

Listing of figures

It is entirely possible that behind the perception of our senses, worlds are hidden of which we are unaware.

Albert Einstein

1 Introduction

In this thesis we propose data-driven methods to analyse, learn, and build applications with style and aesthetics as two perceptual attributes of 3D shapes, which are present in abundance in modern day online 3D shape repositories. In this chapter, we motivate our research (Section 1.1), describe the problems we solve (Section 1.2), propose the solution methodology (Section 1.3), present our main contributions (Section 1.4) and outline the whole thesis (Section 1.5).

1.1 Background and Motivation

We are witnessing an explosion in the growth of 3D shape data freely available in online repositories. This explosion is fuelled by easy availability of cheap geometry acquisition devices and 3D modelling tools in the hands of ordinary users. These shape repositories are also emerging as popular platforms for 3D content sharing not just by novices and professionals, but also by product manufacturers to showcase their brand. Furthermore, many augmented reality [93, 114] based apps allow an ordinary consumer to virtually try products before effective buying can take place.

The exponential growths in the amount of free 3D data has created many fundamental challenges and opportunities in the analysis, search, organisation, synthesis, and reuse of shapes present in online repositories. Many of these challenges are arising due to the way these on-

line repositories are created in the first place. Specifically, online shape repositories do not possess any organisational structure and use inaccurate and ambiguous tags for object shapes. For example, a 'vase' is wrongly placed with 'coffee mugs' and a 'table' may be tagged as an 'umbrella'. Consequently, there is an emerging need to develop tools and techniques that can help effectively search, organise and reuse these large shape repositories.

Although, there exists a large body of work in 3D shape analysis and retrieval, little attention has been paid to understanding and learning perceptual properties of 3D shapes for meeting the challenges mentioned above. These perceptual properties can be related to different human senses such as vision and touch. For example, perceptual aesthetics and softness of 3D objects are two properties related to visual and touch senses. The recent popularity of crowdsourcing platforms in computer graphics make it easier to conduct studies on perceptual attributes of 3D shapes. The large amounts of data collected from such studies can be used for content analysis and content creation by employing machine learning methods.

Shape perceptual property analysis and prediction is an emerging topic [131] in computer geometry analysis and processing. Several applications in computer graphics related to search, organisation, visualisation, reuse, and editing can directly benefit from data-driven modelling of perceptual properties. Consider someone looking for an aesthetic chair model on 3D Warehouse first uses 'chair' keyword to search the repository, and then manually goes through 37, 997 search results (at the time of writing this thesis) to look for the most aesthetic model. Further, if the same person is able to find a model that is the most beautiful, how can he use that model to search a table or other furniture that matches in style with it. Although, the existing body of research allows one to use shape analysis to infer consistent tags, it is relatively very hard to convey stylistic and aesthetics traits using only text. To this end, we investigate the problems related to "matching styles" and "predicting beauty" of shapes in large 3D shape repositories. The terms 'style' and 'aesthetics' are defined as "a distinctive quality, form, or type of something" [Merriam-Webster, 2018] and "the qualities in a person or thing that gives pleasure to the senses" [Merriam-Webster, 2018], respectively.

Our thesis is that developing tools that utilise data-driven models of visual perceptual properties of man-made 3D shapes allows for efficient search, reuse, and visualisation of large amounts of geometric data. We demonstrate this by developing data-driven models of style similarity and aesthetics as two visual perceptual properties of 3D shapes. Our data-driven models consume large amounts of human visual perception data collected from workers registered at popular Amazon Mechanical Turk (AMT) crowdsourcing platform.

The key idea of our approach is that by taking advantage of the large amounts of crowdsourced human visual perceptual judgements along with 3D shape data of man-made objects, we can learn to predict aesthetic value and match styles of 3D shapes, without relying on hard-coded rules or any pre-existing notion of these visual properties [1]. This suggests that

[1] By pre-existing notion, we mean that quantifiable properties such as curvature and symmetry are believed

Figure 1.1.1: Example style matching task. Which coffee table do you think matches more in style with the sofa on the left? Please note the variation in shape, colour, and texture.

a knowledge-based system can be build that uses such hard-coded rules to measure aesthetics, for example more curved shapes are evaluated as more aesthetic. However, in our evaluation, we learn directly from human judgement data rather than considering curvature, symmetry and any other pre-existing notion of aesthetics. Specifically, we collect human judgement data as relative comparison tasks, for example, for aesthetics study we show shapes in pairs and ask which shape is more aesthetic, thus we avoid biasing the data collection to a specific attribute such as curvature or symmetry.

The input to our data-driven methods is a collection of semantically-related man-made shapes, mainly taken from ShapeNet [17] large-scale online repository, which provides multitude of semantic categories and organises them under the WordNet [84] taxonomy [2]

1.2 Problem Statements

As mentioned in the section above, new tools and techniques are needed to explore, organise, and reuse the "big geometric data" [131], we see it as an opportunity to formulate and solve the following problems.

Problem#1 How to learn objective and personalised 3D shape style matching metrics that take into account colour and texture along with shape features? The existing approaches to style similarity metric learning work only on shape geometry features without considering colour and texture attributes. We argue that incorporating colour and texture in learning can make style similarity metrics more useful and throw light on the role of colour, texture, and

to contribute to aesthetic appeal of 3D shapes.

[2] To remove any confusion, by man-made, we mean 3D shapes representing real life objects constructed and built by humans. Specifically, objects of varying sizes and topology, such as mugs, chairs, tables, air planes, buildings, and vases are examples of the shape categories used in our study.

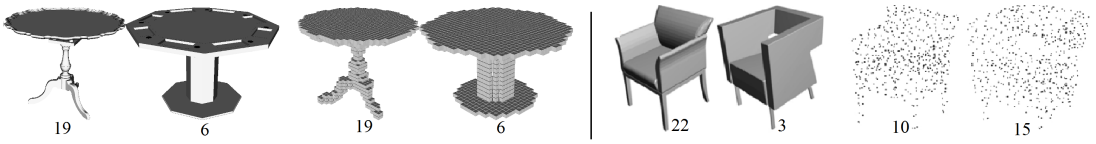


Figure 1.2.1: Examples of shape representations and their aesthetics preference scores (numbers below each shape). First, we show a pair of tables rendered in polygonal representation, followed by the same pair rendered in voxel representations. The number below each shape shows, the participants, out of 25 total, who prefer it to be more aesthetic than the other. For example, for the first pair of tables, 19 prefer first and 6 prefer second for aesthetics. Similarly, next two pairs are for two chairs in polygonal and point representations. Please note that for chairs, we get different proportions of user preferences for polygonal and points representations.

shape properties in style matching. Further, we call the metrics learned on crowdsourcing judgements as objective metrics. We let users explore and adapt these objective metrics to create personalised style metrics.

Problem#2 How does human perception of shape aesthetics vary between different 3D shape representations and how can this help design data-driven model of 3D shape aesthetics? The recent data-driven approaches in computer graphics employ a variety of 3D shape representations as input, namely voxels, images, or point clouds. We use these shape representation to study perception of aesthetics of 3D shapes in pairs to help us decide the appropriate shape representation for data-driven modelling of 3D shape aesthetics (Problem #3). Further, this study on perception of shape aesthetics helps evaluate if humans can discriminate more aesthetic shapes from coarse shape representations. The study is designed to collect and compare human aesthetics judgements on stimuli created using shape representations with varying levels of shape details. The used shape representations include polygons, voxels, point-clouds, and wire-meshes.

Problem#3 How to build a data-driven model of visual aesthetics of 3D shapes from crowdsourced human aesthetics judgements, and without relying on pre-existing shape aesthetics rules? We use deep ranking approach to build a data-driven model of shape aesthetics. The previous study helped us choose voxels as the shape representation to input to deep ranking technique. Compared to the previous approaches, which focus their evaluation of aesthetics on specific properties such as curvature or symmetry, our approach learns directly from human perception judgement data collected by showing shapes in pairs and asking which shape they think is more aesthetic.

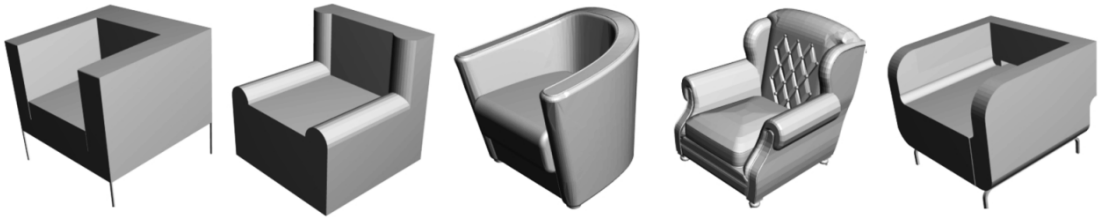


Figure 1.2.2: Example shapes used for the study of aesthetics. These show variation in curvature, structure, and perceived ergonomics.

1.3 Methodology

One basic approach to solve the above problems is to design a knowledge-based system [35]. The hard-coded knowledge is embedded into such systems and then a computer is able to reason using logical inference rules. However, such systems have their drawbacks in terms of generalisation and thus do not lead to any major success.

Another approach involves building a machine learning model that can acquire knowledge by extracting patterns from raw data. However the usefulness of such models depend largely upon the representation of the data used to input to the algorithms. For example, in case of image data, if a right set of features are not extracted, the learning mechanism can fail to learn any useful mappings. In many cases, we do not even know beforehand what right set of features are suitable for learning, so researchers end up using an over-complete set of features. We use this approach in our style similarity problem by learning on an over complete set of features, which works well in our problem.

Finally, to overcome the input representational problem, a new set of machine learning algorithms have recently been used and are called representational learning [11]. In these techniques, we not only learn the mapping from representations to output, but also the representations themselves. Recently, the advances in deep learning help us achieve both goals. We use deep learning to predict and learn shape aesthetics.

We can summarise the overall method as follows. Since we are interested in building computational models of perceptual attributes, we aim to learn from human perception data. We don't base our evaluation on manually constructed rules using any pre-existing notion of aesthetics, such as more curved shape are perceive more aesthetic. Thus, to solve above problems, we follow data-driven approach to allow aggregation of key information from a collection of shapes to support analysis and reasoning relating to various attributes. We demonstrate that a data-driven model is able to reason about shape characteristics without relying on hard-coded set of rules. Since a large amount of perceptual preference data is needed for such problems, we use Amazon Mechanical Turk for conducting perceptual studies. We post perceptual study tasks, called Human Intelligence Tasks (HITs) in crowdsourcing terminology, analyse the collected responses on the predefined set of rules and use the data for learning the shape charac-

teristics.

Shape Style Matching We aim to learn a metric for matching styles of 3D shapes based on their shape, colour and texture by learning directly from human style preference data. We propose to extend and evaluate the effectiveness of existing feature-based style metric learning approaches by using a shape descriptor for metric learning that combines geometric, colour and textural information. We demonstrate using empirical results that this approach works very well for developing texture and colour aware shape style similarity metrics. To this end, our method involves the following steps: first, crowdsourcing a large amount of style preference data as it helps define styles by example rather than using any manually defined rules; second, using a pairwise metric learning method to compute a style distance function based on the crowdsourced style preference data. The pairwise metric learning method is inspired from the existing techniques in learning distance metrics. Our crowdsourcing approach involves showing three shapes to participants, consisting of a source shape (A) and two target shapes (B and C), and asking "Is shape A more similar in style to shape B or C?" This method of presenting stimuli is termed as 'relative comparison triplets' and helps participants actively provide their responses. The metric learning algorithm takes as input two parameters: first, the features computed on 3D meshes and their texture and colour attributes; second, the triplet responses received from crowdsourcing platform. To learn a personalised style metric, we let users interact with the application user interface to provide their style preferences, which can be used to personalise the objective style metric.

Stimuli Selection for Aesthetic Preferences We compare the use of polygonal, wire-mesh, voxel, and point cloud as four different shape representations (or stimuli) to crowdsource shape aesthetic judgements. We pair a set of shapes (A and B) for each stimuli type (e.g. polygonal with polygonal, and voxel with voxel) to ask humans whether they perceive A or B to be more aesthetic. We then compare the received preferences using Fisher's test.

Predicting Perceptual Aesthetics In this problem, our main motive is to learn to predict perceptual aesthetics scores of 3D shapes and use the learned scores to demonstrate a variety of graphics applications. Unlike the previous work, which explored shape aesthetics using predefined features, such as curvature and symmetry, our method employs deep neural networks. We build a deep neural network based data-driven model of shape aesthetics, which takes two inputs. The first input is human aesthetics judgements and the second input is 3D shape's voxel grid. In our method, we treat visual aesthetics as a visual perceptual concept, and thus crowdsource a large number of shape aesthetics judgements. The relevant features or shape descriptors to learn the aesthetic function are discovered automatically. In addition to the above, we investigate the link between the learnt shape aesthetic scores and their computable

statistical features. We do this by first computing several shape descriptors such as Gaussian discrete curvature [83], shape diameter function [109], and D2 shape descriptor [115] etc., and then training a perception to see the prediction accuracy.

We formulate a method to learn a measure of visual aesthetics of 3D shapes by first focusing specifically on geometric aesthetics. Although colour, texture, and any other kind of information could also be included along with shape information, we don't choose to do so. Instead, we first investigate into learning and predicting using "form" of an object alone, which in its own a challenging problem. Our deep neural network based method is easy to extend to use any other properties of object 3D shapes. We are inspired to use deep learning for this problem as it is the state of the art for modelling and learning perceptual concepts, and has the ability to automatically discover features relevant to learning a function of shape aesthetics directly from input 3D shape. Further, our formulation of these two problems is fundamentally different. We learn a function of 3D shape aesthetics that outputs a real value as a measure of aesthetics of input object 3D shape, while for style similarity we learn a distance function that gives a real value as a measure of style distance between two given input object 3D shape descriptors.

Finally, it must be observed that our methodology is general and can be applied to any class of objects. The primary requirement is the need to have human perception data for new object category under consideration. With this data, data-driven models can be built in a way similar to we do and describe in the thesis for other categories of objects.

1.4 Contributions

This thesis makes the following main contributions to advance the state of computer graphics research.

We conduct large scale shape style and beauty perception studies to collect human preference data, which we use to learn computational models of human perception of such characteristics, in order to demonstrate novel applications for content creation and reuse, and to suggest important implications for conducting perception studies and design of machine learning mechanisms.

The above central idea is carried out in three parts as follows:

- We use metric learning for building both objective and subjective models of style matching of 3D models based on their shape, colour and texture. We demonstrate this by building user interfaces to perform style based search, scene composition, and personalised style matching.
- We conduct a crowdsourcing study by showing shapes in pairs and asking which shape is more aesthetic. The shape stimuli shown to participants is created using four different

shape representations prevalent in data-driven analysis of 3D shapes. Our results show that humans' aesthetics judgements on coarser a shape representation such as voxels are comparable to their judgements on polygonal shape representations.

- We exploit deep learning to learn to predict aesthetics scores of 3D shapes and demonstrate aesthetics based applications by building user interfaces to rank shapes in large data-sets and creating visualisations.

1.5 Structure

This thesis is organised as follows. In Chapter 2, we discuss the related work. Chapter 3 details style similarity metrics for 3D shapes. We report the results of our experiments with different shape representations for collecting aesthetics preferences in Chapter 4. In Chapter 5 we propose a data-driven system to learn and predict shape aesthetics and applications built around it. Chapter 6 concludes the thesis.

What you see when you see a thing depends on what the thing you see is. But what you see the thing as depends upon what you know about what you are seeing.

Fodor and Pylshyn

2 Related Work

In this thesis, we visualise our work as having four major components: perception of 3D shape attributes, crowdsourcing perceptual preference data, learning and prediction techniques for 3D shape attributes, and penalisation of perceptual attribute learning. Therefore, we give an overview of the related work partitioned along these areas.

2.1 3D Shape Attribute Perception

In this section we discuss the studies and experiments that look into the problems related to study and analysis of perception of 'style' and 'aesthetics'.

2.1.1 Shape Style

Although the main focus of our work is on learning a measure of perceptual style similarity, for completeness, we also discuss related work in analysis of style. While style analysis allows us to probe defining characteristics of an object, style matching involves more than one object, to say how they relate in terms of their visual styles. In domains other than computer graphics, a significant effort has been made to analyse the styles of a variety of media such as images, videos, and audios [5, 117]. However, in computer graphics there is very limited work on perceptual matching and learning a measure of style similarity of visual content (learning approaches to style matching are discussed in next section).

The focus of style analysis techniques has remained on identification of a set of features or characteristics unique to that style [50]. The work of [1] considers the problem of pottery style. They use statistical analysis and a questionnaire to explore the actual process by which people determine that one object is similar to, or different from another. On using statistical technique of Exploratory Data Analysis (E.D.A.), they identify objective stylistic grouping. Their objective analysis is based on measures of physical characteristics, including width, height, weight, thickness etc. Instead of using object-specific features [50], Dorothy K. Washburn [125] presents a discussion on the concept of style. He emphasises on the basic properties of form or shape, including lines, colour, texture, symmetry, and orientation. His arguments are based on perceptual data analysis using mathematical concepts derived from Euclid and other researchers.

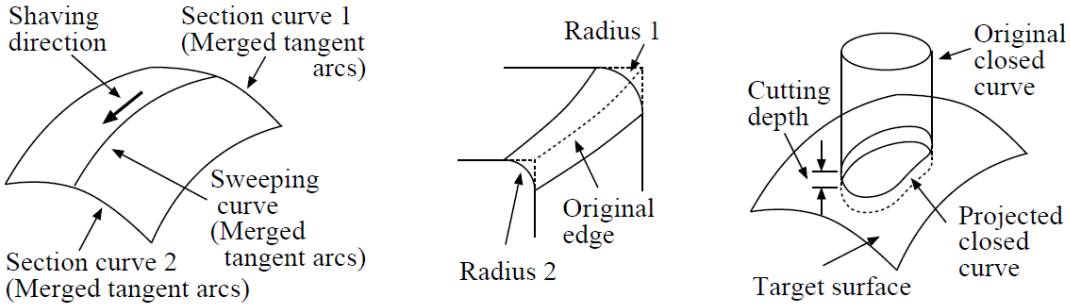
In [16], author examines design studies to establish important concepts related to evolution of architectural styles. He observes that repeated forms can be used to identify an architectural style. This observation is based on looking at design sessions of artists who repeat forms in 'plans' and 'elevations' of architectural drawings. T. Chiu-Shui Chan [15] argue that cultural circumstances and social aspects can be used to identify a style. Thus, a set common features appearing in an object are used as a fundamental unit of style measurement. Further, the proposed algorithmic measure to identification of style uses the concept of features related to physical characteristics such as colour, texture, shape, materials, and patterns.

Martin Stacey [111] suggests that in order to have 'style similarity' between objects, it is important to have shared features. He differentiates between different ways of defining 'style', by considering 'perceptual style', 'creative style', 'preferential style', 'analytical style', and 'generative style'. Further, he presents a review of the literature on how expert designers and novices apply perceptual similarity between objects to create artefacts. Human perception system uses two different ways to access style similarity in [81]. First, based on object features, which allows fast judgements of similarity. This way does not take structure of the representation into account. The second way uses structure into account and provides slower but more thoughtful judgements. Chensheng Wang et al. [121] explore shape style using form features and also investigate into the influence of form features on a style by the topology structure and geometric variations. They found out that the design style can be varied distinctly by introducing variations in topological structure, while modification to local features can enhance the characteristic appearance of a style.

2.1.2 Shape Aesthetics

In this subsection, we report on literature connected to shape aesthetics features such as curvature and symmetry. We also discuss the use of aesthetics for design applications.

Curvature It has been known for a long time that more curved a shape is more beautiful



(a) Face shaving feature (b) Fillet feature (c) Pocket cut feature

Figure 2.1.1: Specific aesthetic features used in [39] as a template to translate an industrial design into a CAD system.

or attractive it is perceived. For example, while buying a product, people have strong preferences for curvilinear designs [73]. The results of psychological experiments done in [6] suggest a strong link between perceptual aesthetics and curvature of geometric shapes. The authors use eight different object classes in their study, including visual art, landscapes, faces and different design classes to establish the link between visual aesthetics and curvature. They ask participants to describe their aesthetic impressions about objects using a set of words. In a study that uses functional magnetic resonance imaging, authors [119] ask participants to judge whether architectural spaces are beautiful. When presented with stimuli, 200 photographs of architectural spaces, each participant was instructed to respond by 'beautiful' or 'not beautiful'. The authors found that participants were more likely to judge them as beautiful if they were curvilinear than rectilinear. Another study asks art gallery visitors to observe an image set containing shape variations of a sculpture [40]. Participants were asked to note their 'most preferred' and 'least preferred' shapes on a ballot. This experiment found that the visitors prefer shapes with gentle curves as opposed to those with sharp points. For aesthetic design applications, a formalisation of aesthetic curves and surfaces is presented in [86]. Authors use specific shape criteria such as 'fairness metric', 'bending energy', and 'minimum variation surface' as a way to describe the relation between curvature and aesthetic surfaces. Two mathematical equations for aesthetic curves are also provided for implementations purposes.

Symmetry The term 'symmetry' in day to day language signifies harmony and beauty in proportion and balance. Authors in [12] associate symmetry [85] as a feature to shape aesthetic. Using quantifiable properties of surface geometry, namely, entropy, complexity, deviation from normality, noise, and symmetry; image aesthetics models are extended towards the evolution of 3-dimensional structures. Locher and Nodine [69], demonstrate through psychological experiments that the symmetrical compositions are linked

to arousal and aesthetic judgements. Using flash experiments it is demonstrated that symmetry is detected by subjects during the first glance. Further, the authors found that the axis of symmetry is used as a perceptual landmark for visual exploration of visual stimuli. Considering polygons as an art form, the work in [38] first asks observers to rate the attractiveness of octagonal polygons that varied in contour length but had approximate constant area, and then ask for judgements about polygons with different numbers of concavities but with constant contour length. On analysing the responses, they show that shapes with partial symmetry are judged more attractive and also the shapes with more total contour length.

Aesthetic Applications There is much work in developing CAD tools that aid in designing products that could satisfy consumer emotional needs. With an aim to build CAD tools for aesthetic product modelling, a study to find a link between a user's emotional reactions and a product's basic geometric elements has been performed in [45]. The study involved an analysis of the design activities carried out by stylists and surfacer in automotive and household supplies fields. As an outcome of the study are two languages for aesthetic product expression. Sequin [108] introduces the idea of "optimising a surface by maximising some beauty functional". His work focuses on abstract sculptural forms for artistic purposes and he mathematically defines 'beauty functionals' that have properties that lead to more beautiful shapes. Their work has been applied to the automotive and household supplies fields. Similarly, a fuzzy shape specification system to support design for aesthetics is described in [99]. It uses pre-defined aesthetic descriptors for designing shapes by allowing designers to specify and work with rough models in a more intuitive fashion. Again, from a product design perspective, elements of design including colour, light, line and shape, texture, and space and movement are considered useful for understanding aesthetics by designers in [37]. The features shown in Figure 2.1.1 are used as a template to translate an industrial design into a CAD system. For example, Figure 2.1.1 (a) shows the face shaving feature that is used for defining fundamental exterior of a product with free surfaces [39].

For building exteriors and urban design, high-level features that determine the aesthetic quality of buildings and their surroundings have been studied in [87]. The high level building aesthetic variables that are studied in this work include: enclosure, naturalness, style, complexity and order. Following the similar pattern of research, that is focusing on specific features, researchers have studied aesthetics of objects such as furniture and jewellery. For example, for office chair design, researchers have considered user satisfaction criteria such as luxuriousness, balance, and attractiveness in [95]. They build a fuzzy rule-based model based on specific variables that are related to these user satisfaction criteria. For jewellery design, the aesthetics of jewellery shapes have been con-

sidered [124]. They allow the user to adjust specific shape features such as golden ratio, mirror symmetry, and rotational symmetry to design more aesthetic shapes.

2.1.3 Shape Feature Perception

Humans perceive an object as a two dimensional image in their retina. The perception of shape can happen from real 3D objects or from 2D images of objects. In computer graphics research, several different representations are possible for an object depending upon the rendering style used. For example, a shape can be rendered as a line drawing or as a polygonal surface. Researchers have investigated how shape perception is related to rendering style used for an object. The usage of line drawings to convey a 3D shape is presented in [26]. The idea of 'suggestive contours' is proposed, which are lines drawn on clearly visible parts of the surface of a 3D shape. Such line drawings convey reliable shape visual information. The important work of Todd et al. [118] investigates the use of different sources of information (e.g. shading, texture, contours) that humans can use to visually perceive 3D shapes. The authors found that the perceptual representation of 3D shape in human brain involves a relatively abstract or less detailed data structure. This data structure is based primarily on qualitative properties of the shape that can be reliably determined from visual information. Using a set of perception studies, Ferwerda et al. [36] find that rendering methods (such as global illumination) and viewpoint have a significant effect on the ability to discriminate shape differences. Their work also suggests that changes in viewpoint increase the human ability to discriminate shapes.

McDonnell et al. [82] conduct a series of psycho-physical experiments to study the effect of different rendering styles on the perception of virtual humans. They found that render style does not change the interpretation of content in a positive or negative manner. This finding is important in the sense that for an animation movie, it is the content not the rendering style that contributes to success or failure. In a similar work, Zell et al. [135] study how shape and material stylisation affect the perception of characters and facial expressions. They show that realism alone is a bad predictor of attractiveness, and shape is the important factor for realism and material is important for appeal. Further to above, in domains related to human perception, researches have explored aesthetics in many different ways [110, 128], suggesting a transcendental nature of aesthetics.

Our work is different from the previous works along two important lines. First, we focus on 3D shape aesthetic judgements on crowdsourcing platforms. Second, our stimuli selection is motivated by use of different shape representations used in data-driven shape analysis and processing. Specifically, we investigate the question of whether rendering 3D shapes in different shape representations (such as polygon meshes, point clouds, voxels) affect the perceptual aesthetics judgements.

2.2 Crowdsourcing in Graphics

The use of crowdsourcing platforms such as Amazon Mechanical Turk (AMT) is constantly on the rise in computer graphics and human-computer interaction research. Crowdsourcing is defined as "the process of obtaining needed services, ideas, or content by soliciting contributions from a large group of people, and especially from an online community, rather than from traditional employees or supplier [Merriam-Webster 2005]". Crowdsourcing platforms offer many benefits for surveying, user studies, data collection, including: low costs per participant, online recruitment, procedural analysis of the data, and variety of participant backgrounds to choose from [8].

In this section we discuss the ways researchers have utilised crowdsourcing in graphics related projects, how they control the quality of collected data, and a brief comparison to our data collection approach.

2.2.1 Crowdsourcing Visual Perceptual Data

Adrian Second et al. [107] crowdsource a user study to measure 3D object viewpoint goodness, in which participants are shown two different viewpoint images of an 'object' and are asked: "Which of these two views do you prefer?". The results of the large user study are then used to optimise the parameters of a computational model for viewpoint goodness. The learned model is used to predict people's preferred views for different varieties of objects. Authors in [20] propose the concept of 'schelling points', which are 'salient' feature points on 3D surface having several fundamental applications in computer graphics. In their study, they ask users to select points on the surface of an object mesh that they think will also be selected by other users. The collected data from humans is analysed using local and global shape properties such as symmetry and curvature. The same properties are further used to predict where 'schelling' points on the surface of an input mesh will be.

Gingold et al [46] introduce human micro-tasks to solve perceptual problems in graphics, for example in the problem of augmenting an image with high-level semantic information such as symmetry can be aided by human input (Figure 2.2.1). They emphasise on the idea that humans are good at visual tasks such as tagging an image, while computers are good at numerical computations. Thus, in their approach they define an algorithm that uses 'human processors' for small visual tasks along with digital processors.

Nghiem et al. [88] demonstrate a system to exploit workers at crowdsourcing platforms to build semantic links between a product's textual description and it's corresponding 3D visualisation. On a web-page based interface, participants read textual product description and locate product features on the rendered 3D model of the same product. For example, in the textual description of a digital camera, participants may read 'shutter button to capture photo'

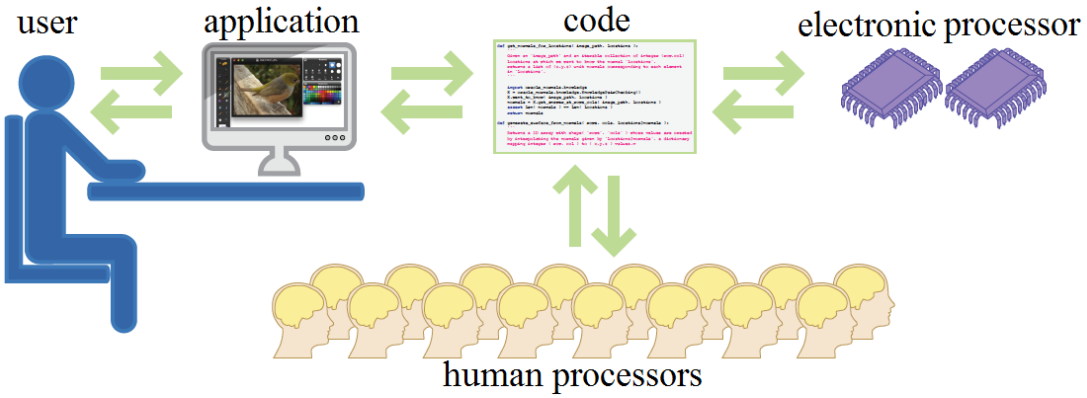


Figure 2.2.1: The process envisioned in [46] to involve human computation (HC) to solve problems in computer graphics and vision. The figure depicts a human who is using an interactive application to solve a perceptual problem such as "create depth layers". The application code invokes a Human Computation (HC) algorithm to utilise human processors (HP). Specifically, The HC algorithm takes advantage of crowd of human workers to solve perceptual tasks and give results back to the main application.

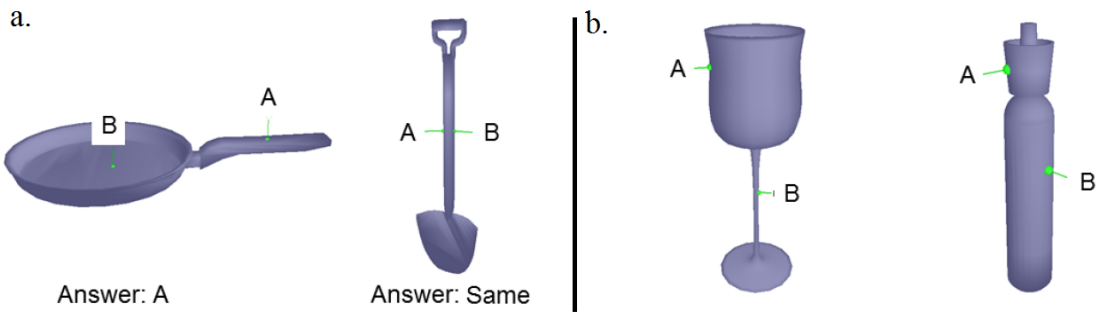


Figure 2.2.2: Crowdsourcing interface used in [61] to learn tactile mesh saliency. In (a) authors show, as part of instructions before attempting the task (or 'HIT' in Mechanical Turk terms), two examples of images with correct answers. The participants are asked to 'imagine the virtual shape as if it were a real-world object, and to choose which point is more salient (i.e. grasp to pick up, press, or touch for statue) compared to the other or that they have the same saliency'. In (b), authors show two examples of real questions.

in the product description and locate the same on the rendered 3D model of the camera. This linkage is useful in enhancing online 3D product browsing experience for customers.

Lau et al. [61] use Amazon Mechanical Turk crowdsourcing platform to collect mesh saliency data to measure tactile mesh saliency. They ask humans to compare between pairs of vertices of a mesh and decide which vertex is more salient (Figure 2.2.2). In another work [60], they learn a model of perceived softness of virtual 3D objects. Similar to previous work, they collect crowdsourced data where humans rank their perception of the softness of vertex pairs on virtual 3D models.

Authors in [112] use Amazon Mechnical Turk to collect ratings of geometric human bodies with respect to 30 body attributes, such as curvy, fit, heavyset, round apple etc. Using the

collected data they learn a linear function relating these ratings to 3D human shape parameters. Specifically, they learn a mapping between a linguistic body space and a geometric body space. An important finding of their work is that humans share an understanding of the 3D meaning of shape attributes used in their work.

The work presented in [41] builds a data-driven model of style similarity for 2D clip art with crowdsourced data. In order to collect large amount of data on human style preferences, authors use crowdsourcing platform to show each participant questions having three pieces of clip art A, B, and C, and ask: "Is A more similar to B or to C?" (Figure 2.2.3 a) The collected data is used in a linear 'metric learning' method to develop a style distance measure between two given pieces of clip-art. The distance metric is build by computing an over-complete set of features encoding 'colour', 'texture', 'strokes', and 'shading'.

Jun-Yan Zhu et al. [139], use popular crowdsourcing platform Amazon Mechanical Turk to collect pairwise comparisons (e.g., "Is expression A more attractive than B?") to score the attractiveness of facial expressions to train a model to automatically predict facial attractiveness of different expressions of a person. They note that collecting data for such studies in pairwise comparisons is a common approach since it is much harder for people to provide an absolute score. In their approach, a novel active learning scheme is also described to help both customise the learned model to the user's data and select the user's top expressions across a range of seriousness levels.

Liu et al. [67] learn a measure of style compatibility for furniture models using a combination of crowdsourcing and machine learning. In each task they pair one furniture item (say chair) with six other different pieces of furniture from another category (say tables), and ask participants to select two pairs that are stylistically similar. In this way, using one task they are able to gather more style similarity data. Lun et al. [74] use similar setup as in [41], except the participants are given two additional options: "can not tell-both B and C", "can not tell-neither B nor C".

Sean Bell and Kavita Bala [9], learn an embedding for visual search in interior design. Specifically, they learn a distance metric between an object in-situ (i.e., a chair in a drawing room image) and independent product image of that object (i.e., product image with white background). They use a unique crowdsourced pipeline to collect a large number of pairings between scene images and the individual product images. For data collection, participants are asked to draw bounding boxes around a product appearing in a scene with other objects. With this data, they design a deep convolutional neural network to learn an a distance function.

Koyama et al. [58], present a method to incorporate crowdsourced human computations for a traditional design optimisation problem dealing with parameter tweaking in graphics design. An example of such problems is when a designer has to spend a lot of time in color enhancement of photographs because parameters such as 'brightness' and 'contrast' need a careful tweaking to obtain pleasing results. In this method, the participants are asked to perform a

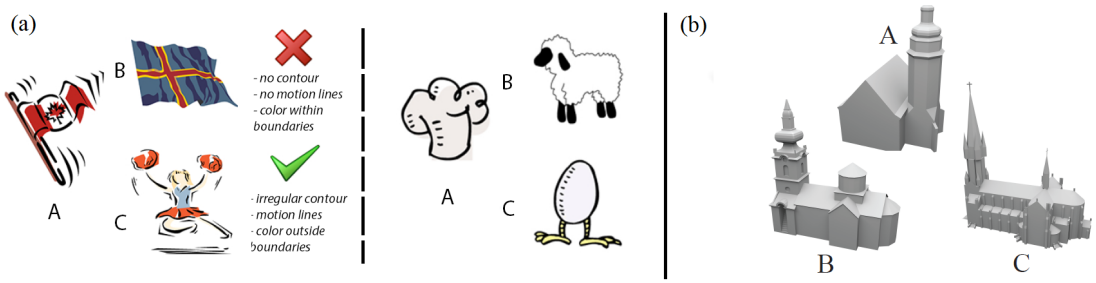


Figure 2.2.3: Examples of tasks presented to participants for collecting style similarity data for clip-art (a) and 3D shapes (b) in [41] and [74], respectively. In (a), left image shows an example in which style of source clip-art 'A' is matched with target clip-art 'C', and right triplet shows a real style matching task. In (b), there buildings 'A', 'B', and 'C' are shown where participants are asked to click on either 'B' or 'C' based on which they think matches more in style with 'A'.

sequence of single-slider manipulation micro-tasks to adjust the parameters. Data collected in this manner from crow is used in a novel technique extending Bayesian optimisation to allow many manipulation tasks using a single slider.

2.2.2 Issues and Quality Control

Amazon runs a survey website called Mechanical Turk (MTurk), on which a large number of workers agree to fill out surveys in exchange for some monetary benefits. Typically, the results of a survey are available within a few hours to a few days. The process of posting and retrieving results is easy and efficient. A typical survey takes a couple of minutes to complete, so hourly rates per person are low. In short, for a majority of work available on Mechanical Turk, crowdsourcing is a bargain for researchers but not for workers.

Issues

Although crowdsourcing allows large amounts of data collection at low prices and in relatively lesser time, it presents challenges for recruiting crowds and collecting quality data. We now briefly discuss some issues concerning crowdsourcing studies used to conduct surveys.

Privacy and Confidentiality Two important downsides of crowdsourcing surveying are privacy of workers' personal data and confidentiality of requester' tasks. To the best of our knowledge, enforcement of such responsibilities is not technically supported by on line platforms. However, the anonymity of workers helps respect privacy of their data. For example, a worker on Amazon Mechanical Turk is identified by a code like "A3IZSXSSGW8oFN", which we don't share outside of our research. A careful study design is still the best way to reduce privacy and confidentiality risks. The details of method or technology should be kept hidden as much as possible.

Worker Communication When the number of crowd workers in a study is large, say a few thousands, communicating with all of them in a meaningful way becomes very difficult. This issue gets more complicated due to anonymous nature of the workers. This results in workers making assumptions if some aspects of the survey are not clearly defined.

Right Crowd Selection The selection of right crowd workers is essential for collection of good quality data. It is expected that some crowdsourcing workers would prefer certain kind of task more over others. For example, performing visual tasks using a mouse-click is considered easier over text entry task. Thus, it is desirable to have such workers for visual tasks. Further, for visual tasks related surveys, lack of worker background on deficiencies such as colour blindness exposes limitations of crowdsourcing. The research reported by Julie S. Downs et al. [31], suggests that some respondents may be participating in Amazon Mechanical Turk (AMT) studies for quick cash rather than inherent interest, and may not be inclined to answer conscientiously

Ethical Considerations The basic idea of crowdsourcing is based on reciprocity or mutual benefit. The workers registered on a crowdsourcing platform volunteer to perform tasks in exchange of money. This process may result in two issues: unfair compensation and not everybody gets paid. Our compensation is designed based on what other researchers pay for similar studies. Some researchers choose to pay bonus to those workers whose work is highly satisfactory, however we do not do so as our tasks are relatively easy, for example for aesthetics, workers click on one shape out of a pair to choose the one they think is more beautiful. We reject and do not pay only those workers who do not correctly answer less than fifty percent of control questions (see below). Further, we block those workers from attempting our tasks whose work get rejected at least three times.

Quality Control

Julie S. Downs et al. [31] develop a method to disqualify participants who participate but don't take the study tasks seriously. They design a set of qualification tests and find that young men seem to be most likely to fail the qualification task. However, participants with professional, student, or non-workers backgrounds seem to be more likely to take the task seriously than financial workers, hourly workers, and other workers. Jeffrey Heer and Michael Bostock, [49], replicate a set of previous experiments on Mechanical Turk to conclude that it is a useful tool to conduct graphical perception studies. However they raise some concerns related to ecological validity, subject motivation and expertise, display configuration and viewing environment are specific to visual perception. They found that crowdsourcing provides up to an order of magnitude cost reduction. Garces et al. [41], use short training sessions and control questions to improve the reliability of the crowdsourced data. Specifically, they use a set of

questions with obvious answers to test the participants. The participants could only access the real test if they answer all the questions correctly on the qualification test. As a second measure of quality control, a set of questions, called control questions with obvious answers, are embedded in the set of actual test questions. Only those workers who provide correct answers to a specific percentage of control questions are paid for their word and their data used for further consideration.

We believe that collecting data as relative comparisons is more efficient compared to using a rating scale. In literature, researchers have investigated different presentation methods for collecting perceptual data on images and shapes. For example, a binary like/dislike rating and a numerical 10-point scale [3] have been tested with crowdsourced voters to understand image aesthetics. Authors in [80] explore different experimental setups to collect data to assess image quality to suggest that forced-choice pairwise comparison method allows collection of most accurate data. In our experiments we also use forced-choice relative comparison method, however in slightly different formats. Similar to [41, 67, 74, 105], we collect relative similarity comparison responses of the form "is object A more stylistically similar to object B or to C", using popular crowdsourcing platform Amazon Mechanical Turk as it offers more flexibility, especially when a quick response to many thousands of queries is required. In shape aesthetics work (Chapters 4 and 5), we collect aesthetics preference data by showing shape pairs belonging to the same high level category (e.g. chairs) and asking which is more aesthetic. Although, subtle, the "key feature" of our crowdsourcing studies is the use of multi-view images (gifs) showing multi-viewpoints of 3D shapes.

We use several techniques to ensure that participants provide quality data. First, we design qualifications test for the participants before they can complete the original tasks. The qualification tests serve as training session for participants before they can start the real test. Second, we embed control questions in the main test and discard the collected data if number of correct answers to control questions falls below the pre-set thresholds. Third, if some participants repeatedly give incorrect answers (to control questions), they are detected and are blocked from doing any further work. Fourth, for aesthetics perception study, we make participants spend at least 2.5 second on each shape pair to select the more aesthetic shape. Finally, we use Mechanical Turk's built-in qualifications to define the audience of workers to work on survey questions. These qualifications are a set of requirements that a potential worker has to satisfy in order be eligible to complete surveys on Mechanical Turk. We define our questions only for those workers who have an overall HIT (Human Intelligence Task) acceptance rate of 95% or more. In the instructions before every survey, we ask workers to use only desktop or laptop devices to complete the surveys. Although, this feature wasn't available at the time we collected data for our projects, however recently, Mechanical Turk offers the possibility where a researcher can specify different parameters related to a worker, such as location (e.g. US, Asia, Africa etc), age, gender, or type of smart phone etc. By doing so, researchers can

recruit workers that meet a certain profile.

2.3 Learning and Predicting Perceptual Attributes

In computer graphics research community, the use of machine learning techniques is rapidly growing. These techniques have the potential to help save both human and computational times in creating, editing, and organising both 2D and 3D content. We focus on developing data-driven models of shape visual perceptual properties using machine learning. These methods allow learning "measures of visual perceptual properties" to allow search and scene composition among other uses. Data-driven visual computing and machine learning has the potential to overcome the current bottlenecks prevalent in the use of computer graphics techniques. For example, computer generated film production requires a lot of manual work (e.g. modelling and texturing) by artists, which can be aided by building data-driven models using "big geometric data".

Here we discuss the background and related work in metric learning and deep learning. These are two specific machine learning techniques which we use in learning measures of style similarity and aesthetics.

2.3.1 Style Metric Learning

Pairwise metric learning methods have long been used [129] to compute distance functions between data points in fields such as computer vision, data mining, computer graphics, and pattern recognition. As stated in [10], "the goal of metric learning is to adapt some pairwise real-valued metric function, say the Mahalanobis distance to the problem of interest using the information brought by training examples", we follow the same paradigm in our style metric learning problem.

Recently, there is a significant interest in research in style-based retrieval of both 2D and 3D content, such as clip-art, info graphics, and 3D furniture, for applications in design and composition. In this section, we discuss style similarity based related work by separating into 2D and 3D content style matching.

2D Content Style

There have been continuous attempts to learn or classify styles of different kinds of media. The most recent ones include the work of [53], which looks into the problem of furniture style classification, both using hand-crafted features and using deep learning (learning-based features). They explore features for style classification into categories such as American Style, Gothic Style, Rococo style etc. Their results show the superiority of learning-based features and also the comprehensiveness of handcrafted features. The techniques proposed in [41, 56, 89], mea-

sure style similarity metric between clip art and fonts using crowdsourcing are also directly related to our work. These techniques use a combination of machine leaning and crowdsourcing to construct functions that compute style distance between two pieces of clip art or fonts.

Also related to our work is the area of design style classification from object images. Aruna Lorensuhewa [70] use an image based questionnaire to collect data about the properties of objects for use with a machine learning based classification. Specifically, they use Bayesian Network to classify furniture designs into different styles, such as Jacobian, William and Marry, and Quessn Anne etc. The classification utilises the data collected from humans for a set of furniture features, including 'appearance', 'proportions', 'chair arms', 'back material', and 'leg type' etc. The work in [7] presents a method to learn the typographical style and produce characters in same style. For this, a deep neural network is build on large amount of training data. In their setup they explore 60 neural network architectures and observe that deeper networks do not lead to significantly higher accuracy.

Authors in [42] extend the work of [41], by reasoning about the illustration attributes people consider more important when they respond to style similarity matching tasks. Their crowdsourcing setup involves directly asking people what they look at when comparing two pieces by style. A deep network to learn image style, aesthetic quality, and image quality is presented in [72]. The networks take as input multiple patches taken from an image.

Andreas Veit et al. [120] propose a novel Siamese Convolutional Neural Network (CNN) architecture to learn cross-category fit or compatibility for fashion items. The network is trained on pairs of items from clothing category that are either compatible or incompatible. The trained network is able to learn interesting semantic information about clothing styles and lets a user create outfits of clothes using items from different categories, that go well together. The work presented in [56] describes a method to learn to predict style of images using deep convolution neural networks. Authors use annotation information to define several different types of image style, such as of visual style, including photographic techniques (Macro, HDR), composition styles (Minimal, Geometric), moods (Serene, Melancholy), genres (Vintage, Romantic, Horror), and types of scenes (Hazy, Sunny). Kai Xu and colleagues [130] present a shape co-analysis method to allow style transfer between 3D models. This is achieved by analysing shapes at part level and treating the anisotropic part scales as a shape style. A data driven method to study style and abstraction in human face sketches is presented in [13]. This method uses properties of strokes and geometry to define style and build models of abstraction and styles of different artists. A similar deep neural network that creates artistic imagery is introduced in [43].

3D Content Style

Ruizhen et al. [52] develop a method to characterise the styles of furniture models by localising geometric elements or regions over shapes. This is done by extracting style defining elements and co-locating them over the set of shapes. Very closely related to our work are the methods presented by Liu et al. [67] and Lun et al. [74] to study style compatibility and style similarity respectively. Both methods compute a weight matrix using distance metric learning. The main inputs to metric learning algorithm are human style judgements and shape descriptors capturing local and global shape properties. The style compatibility work [67] uses co-segmentation to construct part-aware feature vectors for style metric learning using crowdsourced data.

In a very recent work, Isaak Lim et al. [64] employ deep neural networks to tackle 3D shape style similarity problem. They are motivated to use deep learning as all previous approaches used hand-crafted geometric descriptors to learn a measure of similarity. Use of deep learning offers several advantages including: learning style metric on the shape collection directly and avoiding to search and match element level similarity. Authors in [33] analyse the decorative style of 3D Heritage Collections. Their analysis is based on shape saliency. Main contributions of the article include an ontology for documenting 3D representations of heritage artefacts decorated with ornament. Authors in [75] solves the problem of shape style transfer without changing target shape functionality. They view it as a constrained optimisation problem that tries to minimise the style distance between elements and maximises the functionality characteristics.

We highlight the key differences of our approach here. First, we compute geometric features directly on the 3D meshes i.e. without segmentation, and compute colour and texture properties on the associated material files. Second, our triplet construction method is different from the previous work. We use an iterative approach to construct triplets rather than doing it randomly [67] or with subjective bias [74]. Since, due to large data size, constructing triplets at random results in a large number of possible queries, many of which are hard to answer and lead to inconsistencies in learning. Although, [67] uses co-segmentation to construct part-aware feature vectors, there is no correspondence between shape parts, e.g. chair legs getting matched with table legs. In contrast, our method produces fixed length feature vectors and we use an over complete set of features to capture shapes numerically. Lun et al. [74] perform a very detailed analysis of the shapes to define style similarity based on similarity in individual parts. It is important to note that in these methods, the accuracy of results depend on the quality of segmentation. Finally, our work is different from these methods as we use a combination of geometry and material (colour and texture) features to study style similarity of 3D shapes.

2.3.2 Deep Learning

The term deep learning refers to the class of machine learning algorithms that employ a chain of non-linear information processing layers for feature learning, data encoding and transformation. Li Deng [27] provides an informative tutorial on architectures, algorithms, and applications on deep learning categorising them into three classes: generative, discriminative, and hybrid. Recently, deep learning has been used to build data-driven models of perceptual properties and solving 3D modelling problems. Examples of these include, computing human body correspondences [126], 3D shape recognition [113], and tactile mesh saliency [61]. Our formulation to learn and predict 3D the shape aesthetics also uses the concept of deep learning, however our problem definition is different.

We observe that much of the work in deep ranking is motivated by the idea of learning directly from the raw data rather than using any hard-coded rules or hand crafted features. In the paragraphs below, we first talk about the input to deep learning techniques followed by a discussion of recent deep learning based techniques for 3D graphics.

Input to Deep Learning

It is very common to use different shape representations as input to deep neural networks, such as multi-view images, depth images, volumetric or voxels, polygonal mesh, point cloud, and primitive based models etc. While the first three can be considered as regular grids the next three are irregular geometric forms. The choice of 3D representation should allow easy formulation of input-output for the neural network.

Charles R. Qi et al [101] use symmetric max pool function to build a novel deep learning neural network that can reason directly from point cloud data, either sampled from a shape or pre-segmented from a scene point cloud. The point sample are represented as a set of three coordinates (x, y, z), taken from 3D mesh surfaces. They demonstrate the applicability of such network in the areas ranging from object classification, part segmentation, to scene semantic parsing. Mehmet Ersin Yumer and Levent Burak Kara [133] present a data-driven, learning-based surface creation method for unstructured point sets. The method first embeds the given point cloud to 2D space. This is followed by training of the learner. Finally, creation of the tessellation and generation of the surface in three dimensions is done. Authors in [123] present an octree based convolutional neural network for 3D shape analysis tasks such as classification, shape retrieval, and shape segmentation. The novel octree data structure allows efficient storage of octant information and CNN features into the graphics memory and execute the entire O-CNN training and evaluation on the GPU. Rohit Girdhar et al. [47] use two components: an autoencoder and a novel convolutional network architecture to build predictable and generative vector representations of 3D shapes. The novel architecture helps learn an embedding space to allow generation of new 3D shapes and make predictions from 2D images.

Alexey Dosovitskiy et al. [30] implement a generative neural network to generate object images from three inputs: object style, viewpoint and color. The network is essentially an upsampling CNN, trained on images rendered from 3D shapes.

Learning Aesthetics

Although, our focus is on 3D shape aesthetics, we begin by looking at works that assess image aesthetics in both computer vision and computer graphics domains. Many of the works treat the challenge of automatically inferring aesthetic and other picture qualities as a machine learning problem. Datta et al. [24] use a online peer-rated photo sharing site as the data source to build a classifier for automatic assessment of image aesthetics. They use their intuition to extract 56 image visual features which are useful to discriminate between aesthetically pleasing and displeasing images. The selected image features include, measure of colorfulness, rule of thirds, saturation and hue, and familiarity measure etc.

An approach to optimize photo composition using rules taught in photography community is presented in [66]. In addition to optimising composition, this method computes a score of image beauty using features such as rule of thirds, diagonal dominance, visual balance, and size region. The optimisation procedure works by devising a compound operator of crop-and-retarget that selects a subset of the image objects and then the re-targeting operator allows adjustment of their relative locations. The work of Christoph Redies et al. [103] focuses on aesthetic quality assessment of paintings by devising a novel set of features. The proposed features, calculated using computational method called Pyramid of Histograms of Orientation Gradients (PHOG), provide values that are linked to aesthetic perception as suggested by different psychologists. Eisenthal et al. [34] train a predictor for the attractiveness of face images, where facial images are represented as raw gray scale pixels. The human ratings are collected for training the predictor along with vectorized representation of facial images. A data-driven method to enhance facial shape beauty is presented in [62]. Human rates are first asked to provide their ratings on facial beauty, which are then used to train a facial attractiveness engine. This process works by identifying a set of points on face images, called feature points, and optimising their position to improve facial beauty. Said et al. [104] build a regression model to predict facial beauty from facial images. It works by relating attractiveness of input faces to a high dimensional face space. As symmetry of a shape is central to aesthetic perception. Liao et al. [63] use unsupervised learning with geometric priors to improve the perceived beauty of 3D face models, without deviating much from the original faces. Their method enhances a 3D face model by performing symmetrization and adjusting various facial proportions based on the golden ratio [1].

Researchers in [79] design a deep neural network that learns image aesthetics directly from

[1] For two quantities to be in 'golden ratio', their ratio must be same as the ratio of their sum to the larger of the two quantities.

input images i.e. without performing any transformation which destroys the image composition. This is done by the use of pooling layers to directly handle input images with original sizes and aspect ratios. Jiajing Zhang et al. [136] first define a set of metrics to evaluate important aspects in design principles such as balance, contrast, and harmony of logos. The ratings on design principles are collected from 60 volunteer participants, which are then used to build a regression model to predict aesthetics. Yubin Ding et al. [28] review the recent computer vision techniques used in assessment of image aesthetic quality. Chew et al. [21] directly measure the aesthetic perception of virtual 3D shapes with electroencephalogram (EEG) signals. The work of [71] develop a novel double-column deep convolutional neural network to handle both global and local image views for learning a model of image aesthetics. By doing such architecture they claim to capture both global and local characteristics of images, as one column takes a global view of the image and the other column takes a local view of the image. Authors in [137] describe a CNN based framework to model how humans perceive aesthetically pleasing regions in an image. They first associate textual attributes obtained from user specified tags to pixel level in image regions or patches. The importance of this process is that it suggests where humans look at as appealing regions with respect to each textual attribute. Then such patches are used in a convolutional neural network (CNN) to model how humans perceive the visual attributes.

A deep metric learning approach has been used to compare between the facial images in [51]. Deep metric learning has advantages to this challenging problem of comparing faces as it allows modelling a complex and non-linear transformations of face images to useful features. Following the basic principle of deep neural networks, this method passes a face pair through multiple layers of nonlinear transformations. At the last layers, the reduced representations are used to compute the Euclidean distance between two faces. In an attempt to show how to learn directly from image data (i.e. without manually defined features), a convolutional neural network architecture is demonstrated in [134]. This method, trained using gradient descent backpropagation, operates on and compares image patches. Jiang Wang et al. [122] describe a deep ranking method to learn a similarity measure directly from the images and without using any hand-crafted features. They employ a triplet-based convolutional neural network architecture to realise a ranking loss function. The research presented in [19] solves the problem of person re-identification using deep convolutional neural network (CNN) directly from raw image pixels. The proposed network is able to learn the relation between input image pairs and their similarity scores through a joint representation. Authors in [132] propose a deep ranking method to model relationship between video segments considered 'highlight' and 'non-highlight'. The model can be used to assign highlight scores to segments in a long video sequence for the purpose of video summarising. The work in [138] improves the existing learning-to-rank approaches using a novel joint learning-to-rank technique. This technique allows for effectively modelling the intrinsic interaction relationships between the

feature-level and ranking-level components of a ranking model. With the goal of relating physical compliance to perceived compliance of 3D objects, authors [100] collect perceptual elasticity data from a user study using 3D printed metamatrials. The collected stimuli data is then used to learn and predict perceived compliance.

In our approach to study the aesthetics problem, we also apply the concept of deep ranking to attain representational learning paradigm. A representation learning paradigm doesn't require manually selected and computed shape descriptors to represent a shape for learning. Our novelty lies in learning 3D shape aesthetics from human preference data rather than using any predefined aesthetics rules and without using any predefined shape descriptors.

2.4 Shape Retrieval and Personalisation

Directly related to our work is the area of content-based shape retrieval from 3D shape databases. Over the past few years, interest in content-based retrieval methods has resulted in large number of geometry analysis techniques [14, 68, 116]. These techniques allow computing features on a range of aspects of 3D meshes to accurately describe them numerically.

However, these methods perform simple 3D similarity search based on the feature-vectors or on any other representation of three-dimensional meshes. Our work is different from these works as we use human perception studies to understand similarity in 3D shapes. Our focus is on content reuse for content creation; however our technique could be used to perform content-based retrieval as well. Moreover, we focus on colour and texture properties, which traditionally have been ignored in this domain of research.

Personalised information retrieval [96] involves learning a user specific model of perceived relevance to present reordered search results. While there is a considerable work in personalised image search [57, 106], all prior works in 3D content-based retrieval focus on learning a monolithic model using many users' preference data from crowdsourcing platforms. To our knowledge, no effort has been made to learn a user-specific perceptual model of style similarity for 3D content.

2.5 Summary

In this chapter we have introduced the related work by means of grouping in four categories, resulting in four sections, namely, "3D Shape Attribute Perception", "Crowdsourcing in Graphics", "Learning and Predicting Perceptual Attributes", and "Shape Retrieval and Personalisation". In the very first section, we introduce the techniques that explore 'style' and 'aesthetics' in different domains. We introduce works that look into identification of style and the visual characteristics of a style. The published research shows that the style of objects such as 'pottery' and 'architecture' is unique to these categories. We then introduce works in under-

standing "similarity in styles", which show that having shared visual features is important for similarity in style. We also survey the relevant works in the broad area of shape aesthetics. The review of shape aesthetics literature throws light on techniques that pay attention to analysis of curvature and symmetry, and on the development of specific applications. We also review works that study perception of 3D objects from different representations, such as line drawings and under different lighting conditions.

The second section, "Crowdsourcing in Graphics", not only provides a review of the crowdsourcing techniques for collecting human perceptual judgements but also mentions the design and data quality related issues in surveys. The main idea of the reviewed work is to collect large amount of perceptual data and use the same for building data-driven models using machine learning. The reviewed work in "quality issues subsection" suggests several considerations for designing data collection studies for crowdsourcing. These consideration are related to participant payment, communication, and recruitment. In the next section, "Learning and Predicting Perceptual Attributes", we provide an introduction and up to date listing of strongly related metric and deep learning oriented data-driven modelling methods in computer graphics. These methods utilise some 3D shape representation, such as voxels, as input. Finally, we review works that implement 'personalised' applications for shape retrieval in computer graphics. This is motivated from our desire to implement a 'personalised' style similarity metric.

Our work is different from the previous approaches on several key points. We explore shape style similarity not only with shape geometry but also include colour and textural attributes. We let participants try style metrics and adjust the results to produce subjective style judgement data, which we use to learn subjective style similarity metrics. Our setup to build a data-driven model of 3D shape aesthetics and study on perception of 3D shape aesthetics are novel.

Fashion fades; only style remains the same.

Coco Chanel

3

Matching Styles of 3D Shapes

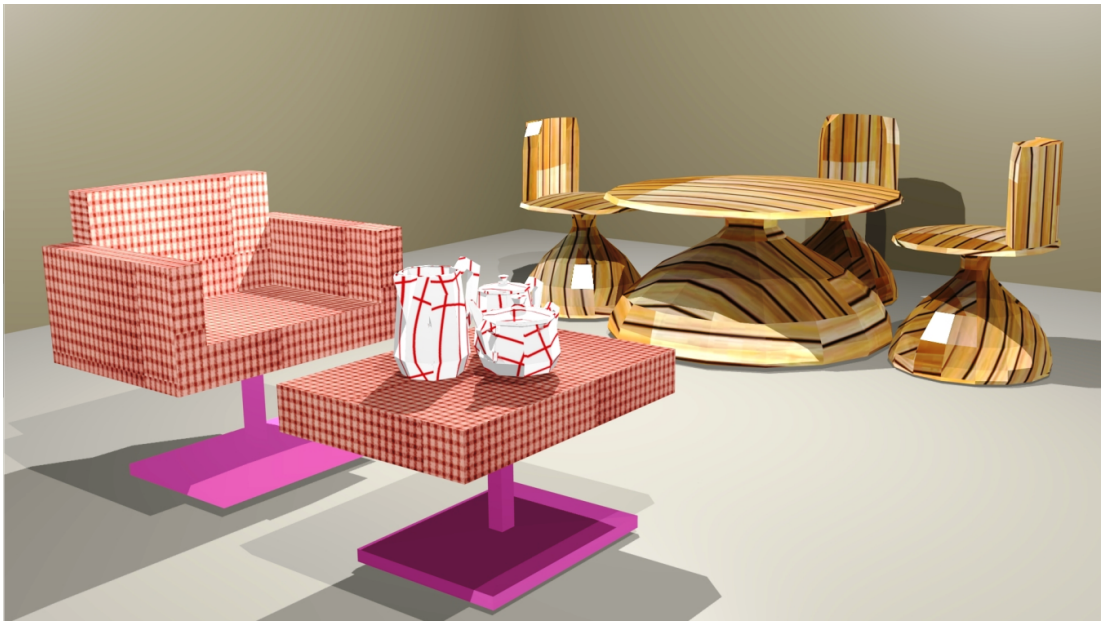


Figure 3.0.1: Example scene depicting two groups of 3D shapes having similar styles.

3.1 Introduction

The word 'style' is inseparable from our lives as it relates very deeply to literally everything around us. Whether it is our dressing style, hair style, living room style, style of our car, style of the music we listen to, writing style, the style of paintings, or the food style, the list is endless; emphasising the omnipresence of style. Further, it is not just the style of individual objects that matters, also matters is when we group a set of objects together(say in our living room Figure 3.0.1). We try to mix in such a way that there is a harmony in overall style, or to be precise we tend to match styles of shapes. The style matching can be considered with objects belonging to: same category (chair to chair), different categories (chair to table or table lamp), and different media (chair to sound). The earliest known source on the study of similarity is the popular "law of similarity" from gestalt psychology, which lays the foundation of understanding perceptual similarity. According to this law, things that appear similar are automatically linked and grouped by our brain based on spatial relations, shape, colour, size or texture. A strategy called "perceptual matching" is used in [76] to investigate style matching of 20 wheel hubs and 6 car types to provide guidelines on choosing wheel hubs for a given car style. The important work of Ming-huang Lin et al. [65], record brain activity when participants match styles of shapes. Their study involves showing tables and chairs in a sequence and recording their brain activity and also asking them to provide 'match' or 'mismatch' judgements. The important result of their study is that a stronger variation in style elicits stronger N400 (record-able brain response to different types of stimuli) effects within the same semantic category.

To build a useful model of style similarity, the colour and texture attributes of 3D models need to be taken into account as these play an important role in overall look and feel of a style in addition to the shape characteristics. To our knowledge, no previous work has endeavoured to learn the model of style matching of objects based on their form, colour and texture. We are inspired by the previous works in defining style similarity metrics for different kinds of media [41, 67, 74, 105]. Our focus is on learning a holistic metric of style similarity of 3D shapes based on their geometry, colour and textures. We make the following contributions to advance the state of the art of this area.

- We build a colour and texture aware shape style matching metric. The real-life observation about style matching leads us to hypothesise that colour and texture features contribute more to style matching than shape features. The learned weights for geometry, colour and material features allow us to shed light on this hypothesis. Our metric learning algorithm uses colour and texture features in addition to geometric features.

- We introduce novel high level grouping as a way to define cross-category metrics. Although, an attempt [67] to learn a weak cross-category (e.g. between table and lamp) style similarity metric has already been made, we apply a more holistic approach by

learning a cross-cluster style similarity metric. We derive the concept of clusters by observing that the day-to-day use words like 'furniture' can be used to represent different kinds of movable items such as chairs, tables, lamps, and beds etc. We define clusters of items using words such as 'furniture', 'cutlery', and 'tableware', for instance, items such as chairs and tables can be placed in 'furniture' cluster. Thus, in addition to learning a metric between individual object classes (e.g. chairs and tables), we learn metrics between the high level clusters (e.g 'furniture' and 'cutlery'), allowing us to substantially reduce the number of learned metrics given N object categories, assuming these can be clustered into M clusters, where $M << N$.

- We demonstrate learning a metric adapted to individual style preferences. Since style perception can be a subjective process, style matching judgements received from individuals can be used to adapt a style metric according to their preferences. The generic metric is based on the crowdsourced style matching preferences, while a user-guided metric is based only on one user's style preferences. If a user is not satisfied with the search results from a crowdsourced metric, our interface allows the user to provide information (e.g. re-rank the results) and create new training data for learning a user-guided style similarity metric.

- We introduce an iterative approach to construct and crowdsource triplet queries and consequently learn a distance metric with them. Rather than creating all the triplets randomly [67] or with subjective bias [74], we use the learned metric in steps to generate the most informative triplet queries (except in the first step).

In this chapter, we will demonstrate the above four contributions with various classes of 3D shapes (e.g. furniture, tableware, and cutlery) and build tools to show the applications of style-based similarity search and 3D scene composition. Our results will help to improve the development of style similarity metrics of 3D shapes.

3.2 Approach

In this section, we introduce our approach to realise the four contributions mentioned above. Figure 3.2.1 shows an overview of our approach. We collect 3D models from online sources (step 1). We then compute various shape descriptors or features (including colour/texture features) for each 3D model (step 2). We generate queries containing triplets of 3D models and place them on Amazon Mechanical Turk to collect crowdsourced data regarding style preferences of the 3D shapes (step 3). The features and collected data are then used to compute a style similarity measure with an iterative approach (step 4). The style metric can be

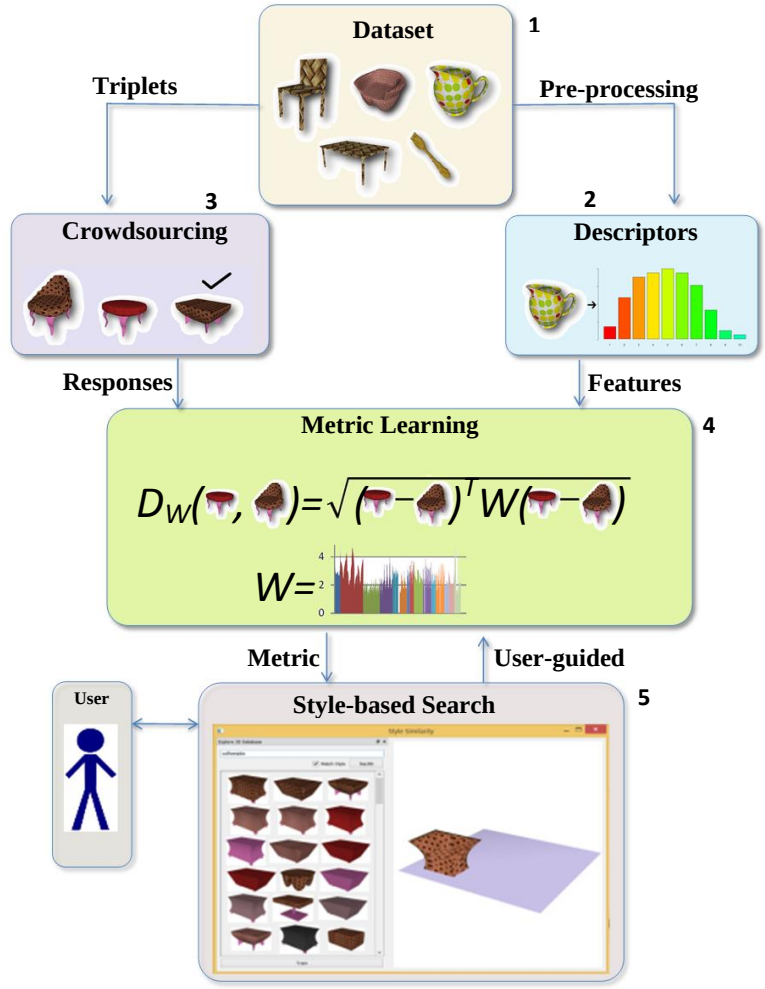


Figure 3.2.1: Overview of our approach. Shapes in the dataset (1) are first preprocessed to compute descriptors (2). Triplets are constructed for posting on crowdsourcing platform (3). Triplet responses and shape features are used to learn the style metric (4), which is then used in style-based search application and user guided metric learning (5).

used in various applications, including style-based search of 3D models (step 5). An individual user can re-rank the models in our interface according to their style preferences, and this information can then be used to compute a user-guided style metric.

3.2.1 Datasets

We collected 3D models from the following sources: 3D Warehouse, Threeding.com, Thingiverse, Lun et al.[67, 74], and ShapeNet [17]. 3D Warehouse is the most popular online open library for sharing 3D models created by 3D modelling tool SketchUp. It offers a keyword based search interface to find relevant models. Threefing.com provides an online marketplace where 3D models can be bough or sold or exchanged freely. The high quality textured models in this dataset are separated into thirteen categories. Similarly, Thingiverse is an online platform for sharing user-created digital design files. The models are high quality designs suitable for use

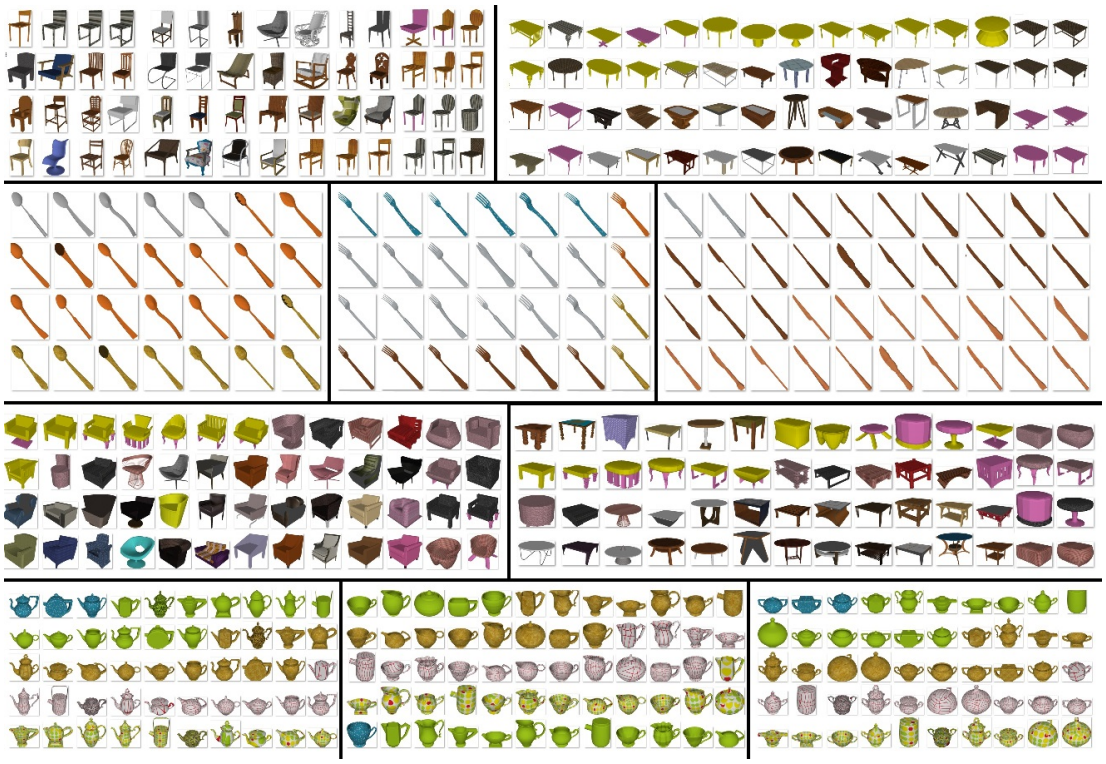


Figure 3.2.2: Example shapes demonstrating different shape classes. (Left to right and top to bottom), chairs, tables, spoons, forks, knives, sofas, coffee tables, teapots, sugar bowls, and creamers.

with 3D printers, laser cutters, milling machines and many other technologies to physically create the files shared by the users on Thingiverse. Finally, mainly for academic research purposes, ShapeNet provides an online 3D content library of richly-annotated, large-scale dataset of 3D shapes. This dataset is organised into two sets, namely ShapeNetCore and ShapeNetSem, based on the quality of mesh data. The object types (Figure 3.2.2) can be categorised into 'dining' room furniture (chairs, tables), 'cutlery' (knives, spoons, forks), 'living' room furniture (sofas, coffee tables), and 'tableware' (teapots, sugar bowls, creamers).

In order to investigate the interplay between shape, colour, and texture, we collected 3D models in two ways. First, we specifically build a set of shapes such that we can easily construct queries (Figure 3.2.4) where the user can specify whether his/her style preferences of 3D shapes is more dependent on geometry or colour/texture. For example, in the 'living' room furniture example in Figure 3.2.4, 'A' matches with 'B' more in its geometry aspects while 'A' matches with 'C' more in its colour/texture. Understanding whether users prefer the geometry or colour/texture aspects more is important for our analysis of style similarity. Hence we map a selected set of textures onto a set of 3D shapes (Figure 3.2.3). For texture mapping, we use 3D modelling tool called Blender that allows interactive mapping of a selected texture image to a 3D shapes. Our choice of texture images is inspired by commonly-used patterns in

Figure 3.2.3: Example texture images used to study style similarity.

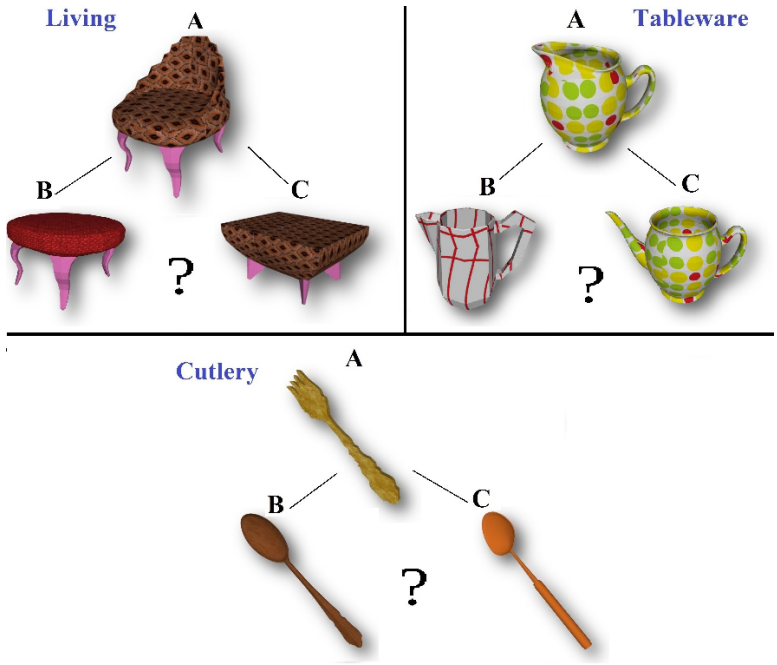


Figure 3.2.4: A subset of our dataset is purpose built to allow us carefully experiment with shape, colour, and texture perception. Specifically, we have models to allow us to test whether users prefer to match the style of 3D shapes based on geometry, colour/texture, or both. We show some examples in various categories. For the 'living' category, B is more similar to A than C in geometry but is less similar in colour/texture. For the 'tableware' category, C is more similar to A than B in both geometry and colour/texture. For the 'cutlery' category, both B and C are different from A in their geometry and colour/texture.

real life. For example, dining room furniture mainly uses wooden shades and textures while living room furniture uses fabric shades and textures. We have 17 Models $\times 5$ textures for each type of tableware (i.e. 17×5 teapots, 17×5 creamers, 17×5 sugar bowls), 21×7 for each type of cutlery (i.e. 21×7 spoons, 21×7 forks, 21×7 knives), 18×7 for each type of living room furniture (i.e. 18×7 sofas and 18×7 coffee tables), and 21×7 for each type of dining room furniture (i.e. 21×7 chairs and 21×7 tables). Second, we downloaded a general set of 3D models from the ShapeNet [17] online repository. For the 'living' and 'dining' categories, we have the following types and numbers of models: chairs (37), tables (35), sofas (35), and coffee tables

(27). These models have much variety in both their geometry and colour/texture.

3.2.2 Geometric Features

We compute shape descriptors on the dataset of 3D models to obtain a 2728-dimensional feature vector for each 3D model, which includes shape, colour and texture features. Before computing features, all models are oriented in the same direction and scaled to have similar proportions within each object type. We use an over-complete set of features and let the learning decide the relative importance of each feature. We aim to capture both global and local shape properties. The features are not new on their own. Please refer to previous work [18, 74, 78, 90] for details of them. We compute histograms of the following (with the number of histogram bins in brackets): shape distribution (128), curvature (Gaussian, mean, max, min: 128 each), shape diameter (128), light field descriptor (470), voxel gradient (192), voxel gradient direction (128), silhouette centroid distances (192), silhouette Fourier descriptor (57), silhouette Zernike moments (108), silhouette D2 descriptor (192), silhouette gradient (192), silhouette gradient direction (96), and shape histogram (192). For these geometric features, there are a total of 2587 dimensions in the feature vector.

Before computing features, all models are oriented in the same direction and scaled to have similar proportions within each object type. Since the input meshes have different resolutions, computing some features (e.g. shape diameter) directly on the 3D mesh results in incomparable or incompatible feature vectors for the learning stage. To rectify this, we use uniformly sampled (with 10,000 samples) surface versions of 3D models to compute the first three features. Specifically, we compute the shape distribution or D2 descriptor as histograms of 128 bins histogram; Gaussian, mean, min and max curvatures histograms with 128 bins each; and shape diameter descriptor with 128 bins. The remaining features above are computed using the volumetric representation of the 3D model by a voxel grid of size 300^3. We rasterize the models into binary voxel grids, where a voxel has value 1 if it is on the boundary of the model, and a voxel has a value 0 if it lies elsewhere. To compute voxel gradient from voxel representation, we use 3^3 sobel filter along x, y, and z axis. The voxel gradient direction histograms of 64 bins each along x-z and x-y are computed from the voxel gradient. The silhouettes along x, y, and z directions are obtained by projecting the voxel space along the three axes respectively. Silhouettes centroid histogram for each projection (64 bins) is constructed from Euclidian distances between center and boundary points. Additionally, on silhouettes, we compute Fourier, Zernike moments, D2 (between each point on boundary), and silhouette gradient (2D sobel filter) and silhouette gradient direction histograms. Lastly, we compute the shape shell histogram descriptor, which is similar to the shape distribution descriptor. The histogram bins for each shell (3) in this case represent the distance of each point to the barycenter of the 3D mesh. Finally, we combine all the features above to give a feature vector of 2587 dimensions.

3.2.3 COLOUR AND TEXTURE FEATURES

We compute the following features (inspired by [41]) to capture colour and texture properties: average HSV of the top five dominant colours (3), hue histogram (32), saturation histogram (32), value histogram (32), and local binary patterns (42). These features are computed directly from the texture images, not from the 3D shape domain. Our feature vector has a total of 141 dimensions of these colour/texture features. To maintain uniformity, all texture images are resized to 512×512. A 3D shape may have more than one texture. For example, a 3D model of a chair may consist of a wooden texture for its seating area and a steel texture for its legs. We handle multiple textures by computing the same features above for each texture and combining their histograms.

To compute the above mentioned features for each model, we use the .obj and .mtl files along with texture images available in folders associated with each object. Specifically, the .obj file gives the information about total number of materials used and which material is assigned to which face. The .mtl file specifies colour and texture properties for each material, and the image folder contains texture images used. We begin by computing the fraction (F) of total surface area covered by each material based on the number of faces using it. For this we use the .obj files. Next, for each material specified in the .obj file, we construct two lists using the information in .mtl files. The first list represents colour information present in diffuse parameters and second list represents texture images used. Given these two lists, we use the following steps to compute colour features:

- For each material: (a) If no texture exists, use diffuse parameters to get 32 bins each for the hue, saturation, and value histograms. (b) If texture exists, use texture image to get 32 bins each for the hue, saturation, and value histograms.
- Finally, combine the above histograms for all materials to get 32 bins for each hue, saturation, and value histograms.

To compute texture features we use Local Binary Pattern (LBP) on the associated texture images. We use three different resolutions of rotation invariant LBP to construct a 42 dimensions vector. We also include the average HSV of the top five dominant materials based on surface area covered (F). These features give us a colour feature vector of length 141 dimensions.

3.2.4 CROWDSOURCING STYLE SIMILARITY DATA

Since it is difficult for humans to provide absolute similarity values (for example, to provide a real number to say how stylistically similar a chair model is to a table model), we ask humans to provide relative values. We differentiate between crowdsourced data collection with many

Figure 3.2.5: Example showing four example Human Intelligence Tasks (HITs) for four different shape categories. In each task, users selected two pairs of models out of the six that are more similar in style compared to the others. They were instructed to base their selection on: number of parts and their arrangement, colour, texture, dimensions of parts and the overall shape, and curviness of parts and the overall shape.

users and user-guided data collection. We collect data by gathering the preferences of a large number of humans by posting tasks on Amazon Mechanical Turk. This idea is similar to previous work [41, 67, 74] and we describe our process here for completeness. The key is to collect data in the form of triplets where we have three objects (A, B, C) and A is more similar in style to B than C. To collect such triplets, we create queries where a human is presented with a 3D model of one object type X and six models of object type Y. Figure 3.2.5 shows some example queries. The task is to identify which two of the six of type Y are more similar in style to the model of type X. For each task, we get eight triplets. If we let the two preferred type Y be Y_1 and Y_2 and the rest be Y_3 to Y_6, the eight triplets are of the form $(X, Y_1, Y_3 - Y_6)$ and $(X, Y_2, Y_3 - Y_6)$.

We post these tasks as HITs (Human Intelligence Tasks) on Amazon Mechanical Turk. Each HIT contains 25 tasks and we paid \$0.15 for each HIT. We can choose the 3D models in these tasks manually or with an iterative approach (Section 4.2). We generate tasks with various pairs of object types, as indicated in Figure 3.2.6. Each human 'Turker' is initially given written instructions and an example task with the responses (two pairs) already chosen by us. We ask Turkers to specifically pay attention to the overall shape, shape of parts, colour, and texture before providing their preferences. For the crowdsourced data collection, we had 220 users and collected 48,000 triplets.

3.2.5 Used-Guided Data Collection

We also investigate if we can learn user-guided metrics and hence we collect data from individuals. We do not use Mechanical Turk here as it can be difficult to require a specific Turker

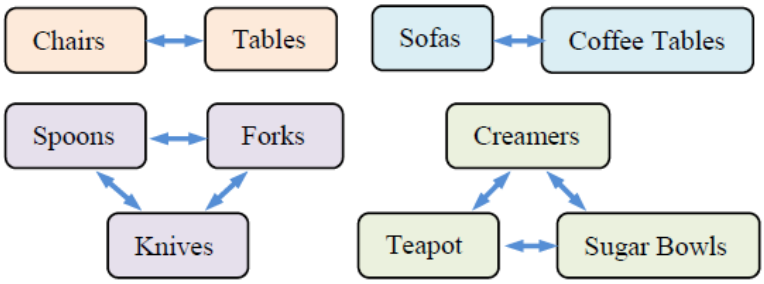


Figure 3.2.6: Bidirectional arrows show pairings of 3D model types for which crowdsourcing queries were generated in this work.

to work on many HITs to collect the needed data (as many Turkers would do just one HIT). Hence we have users who directly use our tools in our lab to collect personalised data. We have two ways to collect such data and we use both of them and combine the data. First, we built a tool to allow users to specify their own preferences by interactively re-ranking search results. The idea is that if a user is not satisfied with the results from the crowdsourced metric, he/she can re-rank the results to generate training data which can then be used to learn a user-guided metric. The tool (Figure 3.2.7: top) allows a user to visualise all 3D models of an object type on the left scrollable panel, where the models can be ranked according to their style similarity to a selected 3D model in the current environment on the right. A user is initially asked to perform a search with the tool using the crowdsourced metric. The user can then re-arrange the ranked results based on his/her preferences of how well they match in style with a model selected in the current environment. The user is asked to specifically place the ten closest match at the top since we use them to generate triplets data. The user interface consists of dragging and dropping the images of the 3D models interactively with the mouse to re-order them. If there is a long list of 3D models in the scrollable panel, the user can also move the mouse cursor over a 3D model and press a key on the keyboard to move it to the top of the ranking. After re-arranging the models, the user clicks a button to generate new triplets according to the ranking. The triplets are of the form (A, B, C) where A is the selected model in the environment, B is one of the top ten ranked models, and C is one of the other models (not ranked as top ten). Such triplets indicate that for the selected model A, the model B is more similar in style to it than C. This process generates $10(n-10)$ triplets where n is the number of models we have for the object type being ranked. Second, we provide another tool (Figure 3.2.7: bottom) for the user to generate triplets data. For this tool, the user can choose two object types X and Y, and the system randomly chooses one model of type X and six models of type Y. The user chooses two of the six and the system generates eight corresponding triplets (as described in the HIT tasks above). For this user-guided data collection process, we had eight users who collected data for various object types. Each user generated just over 30,000 triplets and took about 45 minutes.

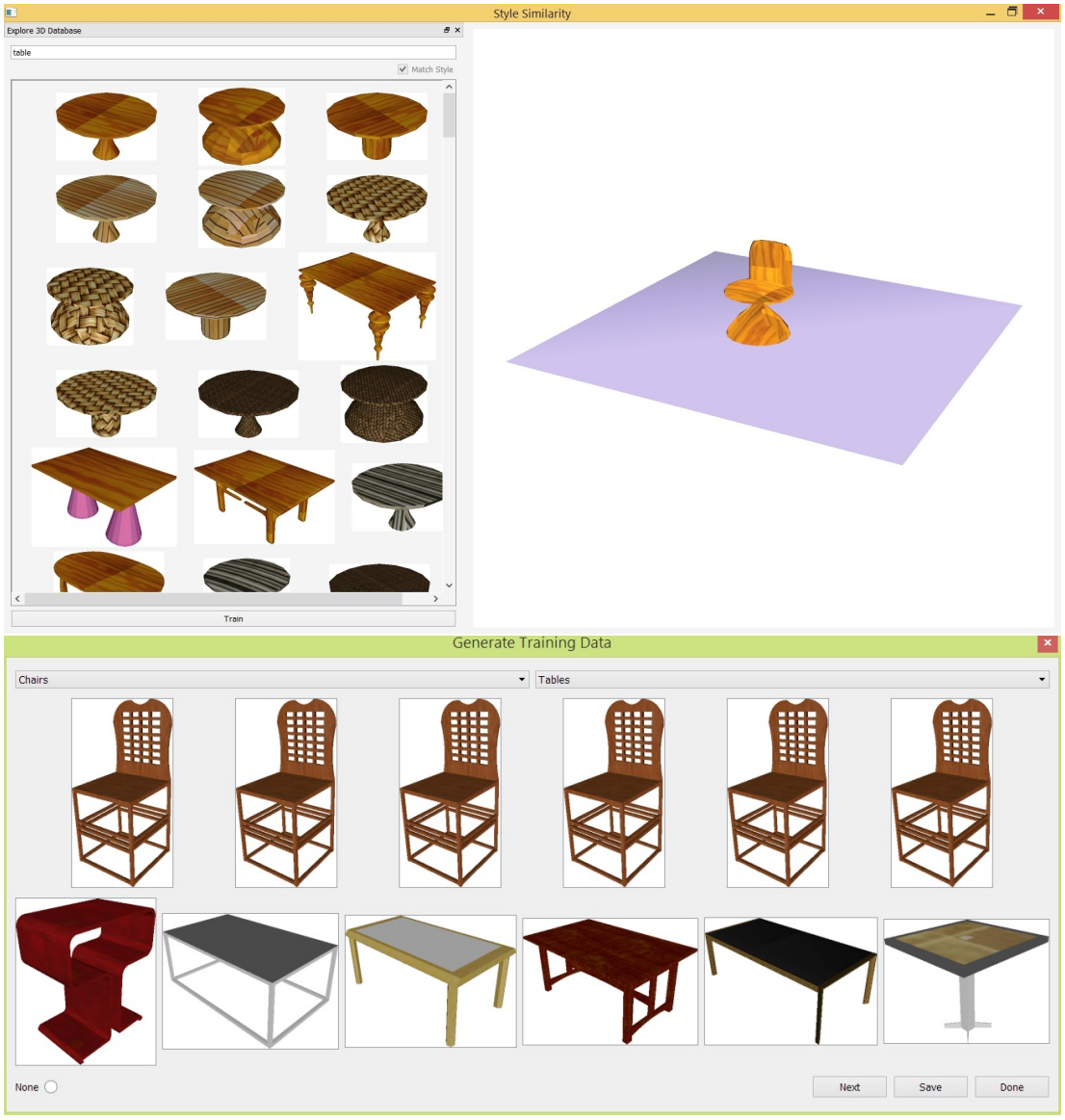


Figure 3.2.7: Our application user-interface and user-guided data collection. (Top) Style-based search and scene composition tool. On the left, we have the model view panel showing the available shapes in the form of a list, while on the right, a panel shows a 3D environment. The list of models can be ranked based on the style similarity compared to the selected model on the right. We allow the user to interactively drag and drop these models to re-rank them to specify their own style preferences, and then the metric can be re-trained. (Bottom) Screen-shot of our tool to allow a user to generate personal style matching triplets similar to the a format used on crowdsourcing platform.

3.2.6 Learning Style Similarity Metric

We wish to develop a method to learn a style distance function between a given pair of 3D shapes. Metric learning [59] approach allows one to learn such functions and hence develop a measure of style similarity. It is a form of supervised machine learning within the domain of artificial intelligence. The general model of metric learning can be considered to have two parts:

a loss function and a regulariser. A typical metric learning method thus can be formulated by choosing a regulariser and a loss function. We use a metric learning approach to compute a distance between two 3D models based on their style similarity. Let $\mathbf{x}_A$ and $\mathbf{y}_B$ be the feature vectors for two 3D models **A** and **B**, and we wish to compute the distance $d_{\mathbf{W}}$, parameterized by **W**, between them as follows:

$$d_{\mathbf{W}}(\mathbf{x}_A, \mathbf{y}_B) = \sqrt{(\mathbf{x}_A - \mathbf{y}_B)^T \mathbf{W} (\mathbf{x}_A - \mathbf{y}_B)} \tag{3.1}$$

The existing methods in metric learning are closely related to regression and classification, with a goal to learn a similarity function from examples. Such function can be used to measures how similar or related two objects are. Broadly speaking, the metric learning techniques fall under three categories: regression, classification, and ranking. We focus on "ranking-based" style similarity learning, to allow search by style or to rank large number of shapes in a dataset based on their style distance from a selected shape. Our method takes advantage of human style judgements collected in the form of style triplets of 3D shapes **(A, B, C)**, where a participant can choose **A** to be more similar in style to **B** than **C**. We define **q=1** if a participant selects B as more similar to A, and **q=0** if a participant selects **C** as more similar to **A**. The probability [41] that a participant's response is **q=1** for a triplet **(A, B, C)** can be defined as:

$$P^A_{BC}(q = 1) = \sigma(d_{\mathbf{W}}(\mathbf{x}_A, \mathbf{x}_C) - d_{\mathbf{W}}(\mathbf{x}_A, \mathbf{x}_B)) \tag{3.2}$$

$$\sigma(x) = \frac{1}{1 + \exp(-x)} \tag{3.3}$$

An objective function can be defined with this probability function and then minimised using a suitable optimisation algorithm [59]. Our framework is inspired by previous methods in metric learning [41, 105]. The learning formulation and solution to solve for **W** is the same as in previous approaches [41, 105], and hence we do not repeat the details here but refer the reader to the previous works.

We take an iterative approach to learn a metric and to gradually build a better **W** metric. The algorithm presented in (Algorithm 1) captures the main idea of this iterative approach. We begin with an identity weight matrix (Step 2) to construct an initial set of random triplets (Step 9). This initial random set of triplets is used to adjust the weights to produce a new weight matrix. The new weight matrix is used to select Y_i in each triplets (X_i, Y_i, Z_j) to produce a set of triplets, which are used to produce the new weight matrix. This process is repeated to a fixed number of times or when learning performance declines (as discussed below). We post HITs on Amazon Mechanical Turk to collect data and learn the weight matrix in each iteration of (Algorithm 1). We can either stop the iterative process after a fixed number of iterations or until the accuracy starts to decrease. We compute the prediction accuracy of a metric learned

with a set of triplets by performing five-fold cross validation on them. Specifically, we randomly divide the triplets data set into five subsets. We use one of the five subsets as the test set and the rest four subsets are used for training the metric. We report the prediction accuracy as the average across all five trials.

Specifically, we take the steps in (Algorithm 1) for each pair of object types X and Y, with $\mathbf{x}_i$ and $\mathbf{y}_i$ denoting the feature vectors of X_i and Y_i respectively. The *random_pick_from*(X) function returns a random shape from object type X, and the function *learn_matrix_using_triplets*$(\mathbf{S})$ returns the weight matrix using the set of triplets $\mathbf{S}$ and the corresponding feature vectors. The main idea of iterative approach is 'information gain' by selectively constructing the triplet queries. The information gain happens by constructing triplets with object shapes with clear majority answer. We avoid generating large number of triplets that don't help much in learning style metric. We argue that the final metric learned using either iterative approach (ours) or completely random approach would result in similar prediction accuracy. However, random generation of triplets incurs more cost in terms of human style judgements crowdsourcing.

Algorithm 1 Iterative learning algorithm

procedure ITERATIVE–LEARN
$\mathbf{W}_0$ = identity matrix or random matrix
for p=1:N
    $\mathbf{S} = \varnothing$
    for q=1:M
        $X_i = \textit{random\_pick\_from}(X)$
        $Y_i = \textit{argmin}_Y \, d_{\mathbf{W}_{p-1}}(\mathbf{x}_i, \mathbf{y})$
        $Y_j = \textit{random\_pick\_from}(Y \setminus Y_i)$
        $\mathbf{S} = \mathbf{S} \cup \{(X_i, Y_i, Y_j)\}$
    $\mathbf{W}_p = \textit{learn\_matrix\_using\_triplets}(\mathbf{S})$
end procedure

The reliability of Turkers was an issue when collecting crowdsourced data. For each HIT, we have 5 tasks out of 25 as control questions to check the quality of the responses. We only accept a HIT if 80% or more of the control questions match with our responses. This is similar to the idea of control questions in previous work [41], and these control questions have clear answers that are meant to check if Turkers are realistically attempting the questions or just randomly selecting answers. In each iteration, we keep re-posting the rejected HITs (which can be done by new Turkers) until we get the desired number of HITs.

We noticed that the HIT rejection rate tends to be high in the initial iterations (as high as 60% in some cases). This is because the initial iterations produce essentially 'random' triplets (i.e. Y_i and Y_j being random due to the initial $\mathbf{W}$). Hence it was difficult for Turkers to provide good responses without paying proper attention and many of them gave responses that seem random. As we progress towards more iterations, the learned $\mathbf{W}$ matrix becomes more

effective and the triplets become less 'random'.

3.3 Results

We present the results towards each of our four contributions. We use the applications of style similarity based 3D model search and 3D scene composition to demonstrate our work.

3.3.1 Colour and Texture

We learn the weight matrices for various object categories with the iterative approach. We experimented with both diagonal and full matrices and empirically observe no significant differences. Hence we choose to learn diagonal matrices and plot the log of the diagonal values (Figure 3.3.1) which correspond to the relative importance of the feature values. The plots show that colour-related features consistently dominate over the geometry features (range of colour-features is 10^5 compared to 10^3 of geometry features). These results are observed for our data collection which includes 3D shapes we manually textured and general 3D shapes downloaded from online datasets. Visually looking at the weights' plots gives the impression that the distributions are same and weights learned on one dataset would work with another dataset. However this is not true. There are both large and subtle differences between actual weight values for different datasets. Similarly, the values differ for the corresponding feature values for shapes in different datasets. These differences for both weights and features for one dataset result in very poor prediction (<55%) accuracy when used with a dataset not used for learning the weights.

Figure 3.3.2 shows the results of style similarity based search with our style metric. The top five search results for each query 3D model show that while both geometry and colour/texture are important, colour/texture is considered first when attempting to match style before geom-

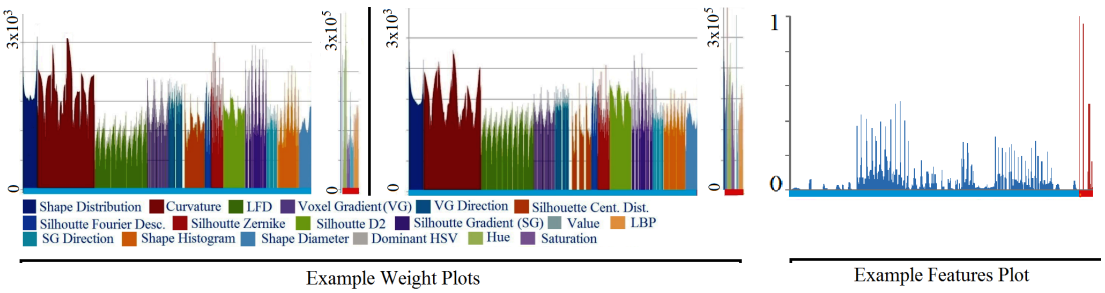


Figure 3.3.1: Example weights (first two) and features (last) plots. The first two example plots show the learned weights for 'dining' and 'cutlery' categories (y-axis show the actual range of values). The weights correspond to features in the feature vector (last plot). There are 13 geometric features (blue bar on bottom of each plot) and 5 colour/texture features (red bar).

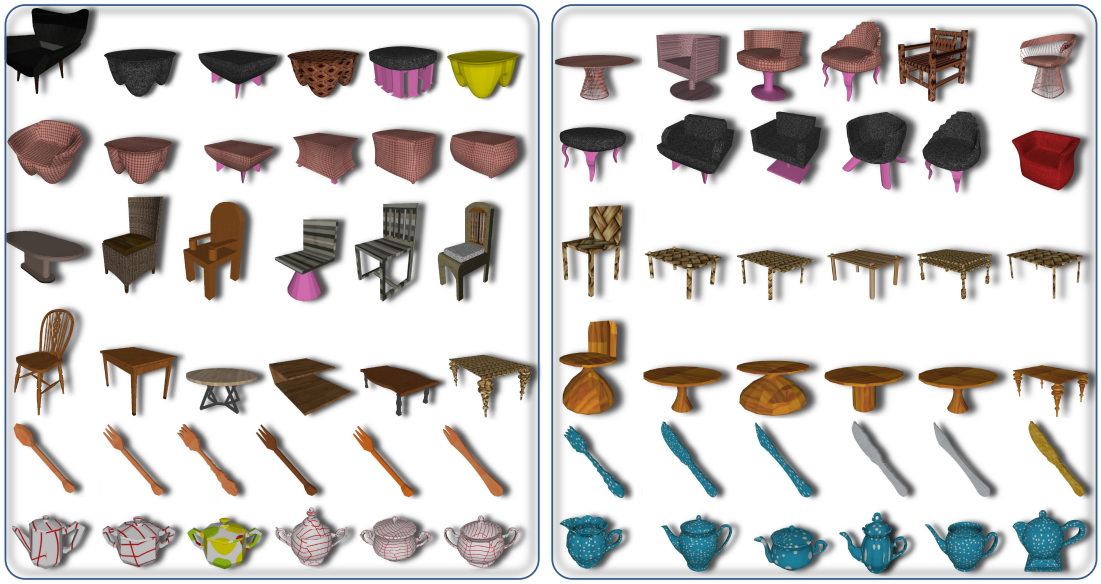


Figure 3.3.2: Style similarity based sample search results with our crowdsourced metric. There are two columns of results. First model in each column is the query model which is followed by top five models that best match in style with the query.

etry is considered. This is true across the different types of shapes that are shown, which again include 3D shapes we manually textured and general 3D shapes downloaded online. In the second row (right column) of Figure 3.3.2, the red sofa happens to match well with the query model as their curvatures are similar. In Figure 3.0.1, we use our style similarity metric with our search tool to compose 3D scenes. As the 3D models that are preferred by the crowdsourced metric are placed at the top of the search results, it is easier to find models that match in style with a selected shape. Hence we have empirical evidence to support our hypothesis that colour/texture features dominate over geometry features, in the plots of weights and in the style based search results.

3.3.2 Clustering of Object Types

Table 3.3.1 shows the accuracy results for different pairings of object types and clusters. We do not take the iterative approach here to ensure that the randomness does not affect the results. We instead created HITs manually to cover the range of 3D models in each object type. We took 5 HITs with acceptable responses (after control questions) for each pair of object types, and generated a total of 6040 triplets. For the clustering into groups, the idea is that chairs/tables can be a 'furniture' cluster and forks/spoons can be a 'cutlery' cluster. Since the models in our dataset have already been labelled (e.g. as chairs, forks), we manually cluster them into higher-level categories (e.g. chairs/tables is 'furniture'). We combine the collected triplets from the separate types to create the triplets data for the clusters.

	Chairs	Tables	Forks	Spoons
Chairs	-	66.73	34.52	50.44
Tables	73.29	-	41.67	47.52
Forks	64.25	51.19	-	80.19
Spoons	38.87	61.17	61.38	-

	Chairs and Tables	Forks and Spoons
Chairs	66.73	42.24
Tables	73.29	42.99
Forks	61.94	80.19
Spoons	50.16	61.38

	Chairs and Tables	Forks and Spoons
Chairs and Tables	73.15	42.65
Forks and Spoons	51.83	72.21

	Chairs, Tables, Forks and Spoons
Chairs, Tables, Forks, and Spoons	56.84

Table 3.3.1: Cross-validation percentages for different pairings of object types and clusters. We learn metrics for $X \rightarrow Y$, where X (and Y) is the type or cluster in each row (and column).

Observing the results from Table 3.3.1, we see that the percentages for some object types (e.g. chairs and tables) are comparable to the results with the iterative approach (shown below). We intentionally compared across different object types here (e.g. forks $\rightarrow$ tables) and hence some pairings give low percentages as it may be difficult to compare between some object types. This does not affect what we aim to show: the trade off between learning metrics for specific object types versus clusters of object types.

We observe that the percentages of the clustered pairings are somewhat averaged from the percentages of the separated pairings. We hypothesised that the clustered metrics would be less accurate, as they may be mixing object types that are quite different. However, our empirical results show no clear consensus of whether the metrics from specific object pairings or clustered pairings is better.

	Chairs and Tables	Forks and Spoons
Chairs and Tables	72.30	40.68
Forks and Spoons	52.12	71.10

	Chairs, Tables, Forks and Spoons
Chairs, Tables, Forks, and Spoons	56.36

Table 3.3.2: We randomly take half of the original triplets (compared to Table 3.3.1) in each of these five cases and re-calculate the cross-validation percentages.

Since we combine the triplets data to learn a style metric during this 'clustering' process, we also tested whether the number of triplets would have been a variable that affects the percentages (i.e. more triplets data may lead to a higher percentage). Table 3.3.2 shows the results

Figure 3.3.3: Example search results for crowdsourced and user-guided metrics for three source shapes shown on the left in each row: chair, sofa, and spoon. We show top five crowdsourced metric results immediately after each source shape, followed by top five shapes using user-guided metric, all in the same row.

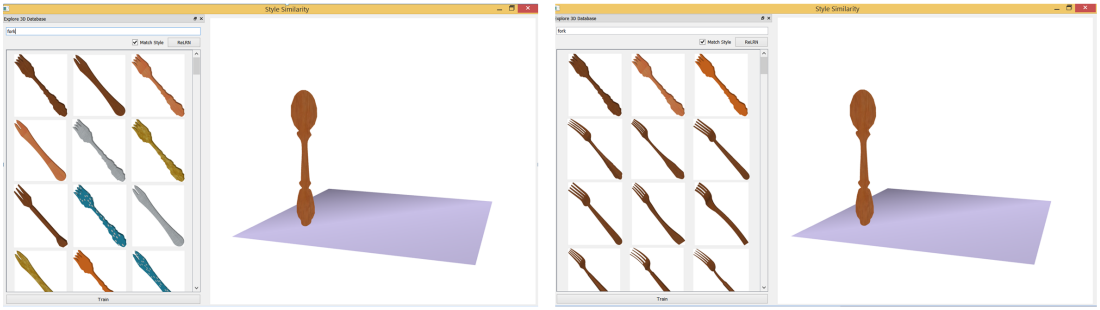


Figure 3.3.4: An example re-ranking (right) done by a participant of original ranking (left). Specifically, left image shows original ranked results using the generic metric and the right image shows rearranged results by a participant. Please note that the user gave more weightage to the geometry.

where we randomly take half of the triplets in each case and re-calculate the percentage. These results show that the number of triplets does not affect the percentage.

3.3.3 User-Guided Style Similarity Metric

Figure 3.3.3 shows that the user-guided results are interestingly somewhat different from the crowdsourced results. For example, the colour and shape of the legs of the furniture pieces are different between the two results. For the 'cutlery' example, the crowdsourced results mainly match with the colour/texture features, while the user-guided results include a spoon that has very similar geometry (i.e. round-shaped handle) but different colour/texture. The individual user in this case was attentive to the geometry of the cutlery in addition to their colour/texture. These results demonstrate that we can learn a style metric for individual users that is different from the crowdsourced metric, providing evidence for our hypothesis that the user-guided concept can be beneficial in some cases.

The participants in our study rearranged (or re-ranked) the position of shapes within the ranking produced as a result of style-based search (Figure 3.3.4). We found that three proper-

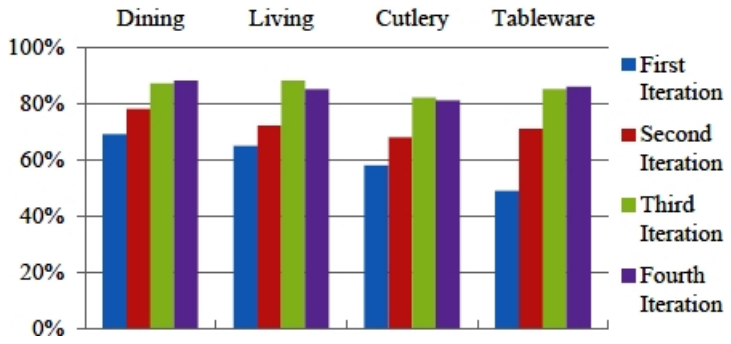


Figure 3.3.5: Cross-validation percentages for the iterative learning for 'dining' (chairs→tables), 'living' (sofas→coffee tables), 'cutlery' (spoons→forks), and 'tableware' (teapots→sugar bowls). First bar (blue) in each category shows the accuracy of the metric learned on randomly generated triplets.

ties of objects led to re-ranking of shapes, namely geometry, colour and texture. We also found that one third of the participants preferred colour over geometry while re-ranking the shapes. For a set of common reference shapes, for which all participants performed style-based search, and re-ranked the search results, we report Spearman's Rank correlation. Specifically, we use Spearman's Rank correlation on re-ranked results to report how much ranking results differ between participants. We compute Spearman's Rank correlation between first fifteen re-ranked shapes by pairs of participants. The average value of this correlation is high (0.88), signifying consistency between re-rankings done by participants. This result signifies that the idea of personalised data collection by allowing re-ranking of the results if useful. Further, since the overall colour and shape preferences among different people are still mostly consistent, the differences in the user-guided metrics are subtle.

3.3.4 Iterative Learning Approach

For each category of shapes, we started with posting randomly generated triplets. We collected the same number of triplets for each iteration, and we have between 1600 and 2000 of triplets for each of our four object categories (in each iteration). The accuracy achieved with the metric learned on such triplets in each iteration are shown in Figure 3.3.5. As the first iteration represents the non-iterative method since it is the same as randomly generating triplets as done in previous work, we can compare between the percentages for the non-iterative method and the iterative method (our last iteration). The percentages increased from 69% to 87% for 'dining', from 65% to 85% for 'living', from 57% to 82% for 'cutlery', and from 49% to 86% for 'tableware'. These results support our hypothesis that the iterative process can generate useful HITs, and can avoid having to randomly generate triplets or to manually group the 3D models in advance [74]. We found that stopping after a fixed number of three iterations worked well, and this was consistent across the four object categories.

3.4 Discussion

Learning a pairwise distance metric between 2D or 3D data pairs has recently become popular in computer graphics due to applications in shape retrieval, scene composition, and visualisation. In this chapter, we build a style similarity metric of 3D shapes that overcomes the limitations of existing style metrics. The methodology presented in this chapter is general and can be applied to problems having similar formulation. For example, in addition to colour and texture characteristics, we may build more advanced style metrics that take into account the construction material of the shapes, since, in general, style matched furniture sets are made with similar material. Next, we present a discussion of our technique by focusing on limitations, human aligned metric learning, and crowdsourcing style data, followed by drawing the main conclusion.

3.4.1 Limitations

Our style similarity metric has two minor limitations that arise due to use of specific dataset of shapes and learning using a simple linear distance metric. Fist, we build part of our dataset by manually mapping a set of texture images to a set of shapes in order to better understand the relative importance of colour, texture and shape characteristics, please see Figure 3.2.4. Specifically, the use of subjectively built subset of dataset used in this study makes our contribution slightly biased to our dataset. Further, manually mapping textures to shapes is time consuming and also limits the ways the texture can be mapped to geometry.

Second, the metric learning method used in our study is very naive. Specifically, the modelled distance function is simply the Euclidean distance metric. Hence the learned style metrics are limited in their expressive power, and more complex non-linear functions can allow the metrics to better represent human preferences. To this end, Isaak Lim et al. [64] use deep learning to model shape style similarity. However they don't consider other useful visual stimuli possessed by the shapes in online repositories, like colour, texture, and material, to build a practical style similarity metric.

3.4.2 Human Aligned Style Metric

Our method is well aligned with the human perception of style considering colour, texture and shape attributes of the object models present in current day shape repositories. We demonstrate that a single shape descriptor capturing shape, colour and texture features can be used to compute style distance using metric learning. Additionally, the set of shape features (Section 3.2.2) used in our study robustly allow for learning the style distance providing prediction accuracy comparable to previous approaches [67]. This is significant as previous methods rely on more advanced part-based segmentation methods to compute shape features before per-

forming metric learning. The overall approach used in our paper can be easily extended to include other prominent model attributes such as construction material to further make the style similarity metric more aligned to various perceptual characteristic. As desired by previous work, our results highlight the usefulness of building cross-category (Section 3.3.2) shape style metrics e.g. between furniture and cutlery. We notice that cross category metrics tend to be less accurate owing to structural and categorical differences between the shapes being compared.

3.4.3 Crowdsourcing Style Data

Our system relies on large amount of crowdsourced data for parameter learning. One consideration for collecting such data is the how we present the queries to humans. We choose to ask humans about their preference by showing triplets of shapes (Figure 3.2.3). Although, this approach allowed us to collect useful and reliable data for our problem, other approaches for collecting style similarity judgements may also be considered. For example, rather than using relative comparison triplets, we may employ an absolute value of style similarity between pair of shapes and then see how consistent humans are in providing data in such format. Another consideration is the selection of stimuli for style comparisons. Specifically, our data collection process employs single-view object images to shown in triple queries, we argue that usage of multi-view images may allow participants to perceive more shape details and thus aid in collecting more accurate style similarity judgement data.

3.4.4 Conclusion

In this chapter, we presented a colour and texture aware style similarity model for 3D shapes. Our method is based on crowdsourcing and learning to build a model of perceptual style similarity of 3D shapes. Our work is different from previous works in four ways. We use colour and texture properties to match styles, introduce a way to build inter-class style similarity without requiring large amounts of perceptual data, learn and demonstrate personalised style metrics, and use iterative approach to build a style metric. Our method operates in both image and object space since texture images and geometry are both used to generate a unique shape descriptor for each shape in the dataset.

It's not what you look at that matters, it's what you see.

Henry David Thoreau

4

Stimuli for Aesthetics Judgements and Beyond

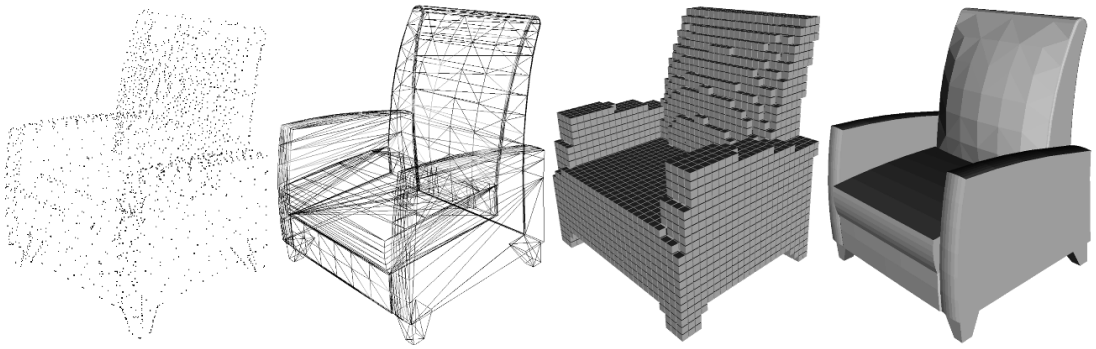


Figure 4.0.1: Example shape demonstrating stimuli created from four shape representations, namely points, wires, voxels, polygons, used in our study to collect and compare perceptual aesthetics judgements of 3D shapes.

In previous chapter we described our computational model for measuring perceptual style similarity of 3D shapes. The model was build on perceptual preference data collected from human participants. In this chapter, we study human perception of 3D shape aesthetics. To this end, we show participants 3D shapes in pairs to choose the one they think is more beautiful. The experiments are repeated using stimuli created from four different shape representations, namely, polygonal, voxel, point-clouds, and wire-mesh representations of 3D shapes (Figure 4.0.1). Since these representations offer varying degrees of shape details, based on this, one of our main goal is to investigate the affect that shape details have on perception of 3D shape

aesthetics.

Although there are other possible shape representations, the motivation to use polygonal, voxel, point-cloud, and wire-mesh shape representations is derived from their use in recent data-driven research in 3D graphics [47, 101, 133]. Using these representations in mind, we first wish to study human perception of shape aesthetics, which has implication for crowdsourcing human aesthetics judgement data collection and also in the further use of such data for modelling and analysis problems. We use renderings produced from these representations as shape stimuli in our experiments. The renderings are created in a consistent manner by fixing the rendering parameters (e.g. lighting, camera, position).

4.1 Introduction

Historically, the perception of 3D shape has remained an active research area in many different disciplines, including psychology, neuroscience, computer science, physics and mathematics, and we argue that it will remain to do so until the secrets behind working of human brain are completely unearthed. An important outcome of the research in these areas suggests that human brain uses relatively abstract data structure for the perceptual representation of 3D shapes [118]. In this work, considering aesthetics or beauty as a perceptual concept, we present our first step towards computational exploration of this concept using machine learning and without any pre-defined rules traditionally used to evaluate aesthetics of shapes. Specifically, we focus on collecting human aesthetics judgements for use in data-driven techniques to learn and predict shape aesthetics.

Recently, data-driven techniques have been successfully used to explore aesthetics of images [62, 66] and 3D shapes [12, 40, 91, 108]. These techniques allow learning from human preference data rather than on subjective hard-coded rules [1]. The input to these is a range of shape representations, such as voxels, points samples, depth images, multi-view images, and polygonal meshes, to name a few. It is relatively unclear as how to decide which representation to use and how many input samples, pixels, points or voxels are enough for learning shape aesthetics for instance.

The primary input to many data-driven methods [131] in computer graphics is the visual perception data crowdsourced from large population of participants. A typical crowdsourcing study involves showing shape stimuli as images on which participants are required to provide their judgements about a specific perceptual attribute, such as style or aesthetics [29, 41, 74]. Among many other parameters in a study, selection of stimuli is the key ingredient, and in our view has not received the due research attentions in more recent crowdsourcing based computer graphics research endeavours. As [64] rightly observes, a single view may hide important visual content, we investigate if any differences can be seen in perceptual judgement

[1] Please see Chapter 1

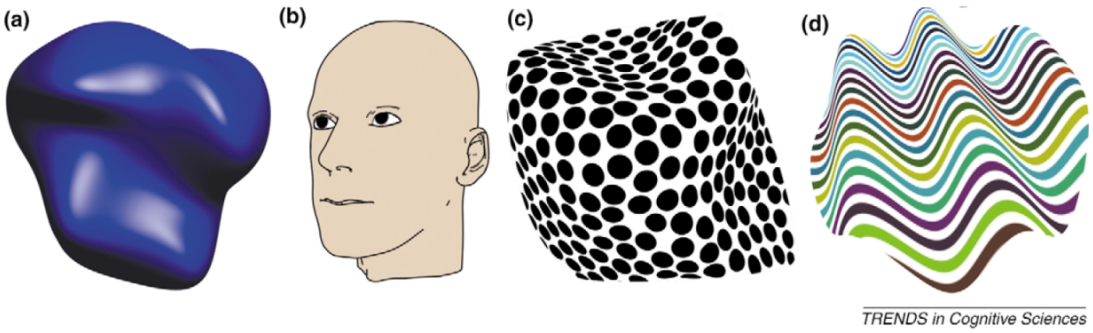


Figure 4.1.1: Example showing possible ways to depict a 3D shape using: (a) shading, (b) line drawing, (c) pattern of dots, and (d) pattern of parallel surface contours [118]

when a multi-view stimuli is used for the aesthetics problem. Further, inspired by the use of different input shape representations to (as mentioned in previous paragraph) learning mechanisms, we experiment with collecting shape aesthetics judgements with such representations and compare the received responses.

How to collect large number of perceptual aesthetics preferences of 3D shapes is a challenging task. We believe, there are two important considerations: how to and what to present as the stimuli, and how do humans provide their judgements to the presented stimuli. The stimuli could be single image and participants respond by providing an absolute aesthetics score. The stimuli could be paired shapes and humans say which is more aesthetic by clicking on it. The ease with which humans can give their judgements is critical to the effective data collection process. One method is to show users a single shape and ask them to give an absolute aesthetics score. However, an absolute scale may not be consistent across individuals. One person might give a score of 0.95 to indicate an aesthetic shape while another might say that a score of 0.7 is already very aesthetic. Instead of an absolute scale, we choose a relative scale of scores. This is motivated by recent work in collecting crowdsourced data [61, 67, 74, 89] where triplets or pairs of media types (including fonts, 3D shapes, and vertices) are shown to users. We collect aesthetics data by showing participants two shapes (or a shape pair) and asking them to choose one that they perceive to be more aesthetic.

In this work, in addition to comparing single-view and multi-view images, the important problem we explore is whether the shape representation affects the human aesthetic preferences of 3D shapes, where 3D shapes are presented in pairs (Figure 4.0.1). We believe that this problem is interesting from two important perspectives. First, the quality of human aesthetics judgement data is important for the study of shape aesthetics [2, 29, 40, 91]. Second, to aid in the design of learning based shape analysis methods that use various shape representations (e.g. voxels [127] and point clouds [101]) as input. An understanding of the effects of different representations would be useful for robust data collection for shape aesthetics and for

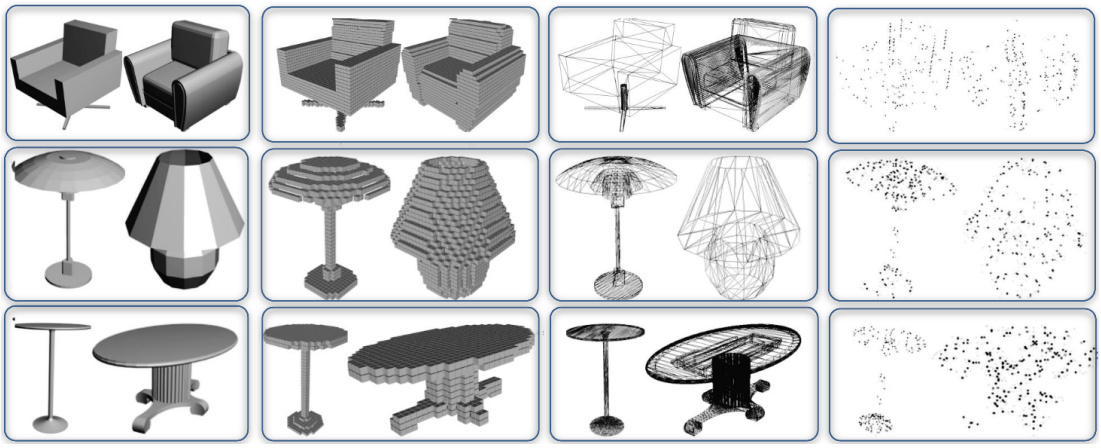


Figure 4.1.2: Example of a 3D shape pair of chairs in four shape modelling representations: polygon mesh, wireframe mesh, point cloud (250 points), and voxels (resolution of 32x32x32).

general shape analysis problems.

Specifically, the following points summarise the key idea and the approach used in this work.

- We first compare between single-view and multi-view representations of 3D shapes and investigate whether there is a difference between them towards the human aesthetics preferences of shape pairs.
- We then compare between the polygon mesh and wireframe mesh representations, between the polygon mesh and point cloud (for various numbers of points) representations, and between the polygon mesh and voxel (for various voxel resolutions) representations.
- We take a shape pair (shapes A and B) from each style and ask humans whether they perceive A or B to be more aesthetic. For example, to compare between polygon meshes and voxels, we take the polygon mesh renderings of A and B and collect the preferences for a number of participants.
- We use Fisher's exact test to test the null hypothesis that there is no difference in the proportions of preferences (of A and B) between the polygon and voxel representations. We perform this test for many shape pairs to compare between the two representations.

Finally, comparing between various types of stimuli for collecting the aesthetics preferences of 3D shape pairs is the main contribution of this work. Contrary to what is believed in general, our analysis of collected data suggests that abstracted shape representations, such as voxels, can be reliably used to collect aesthetics preferences. For example, aesthetics preferences with voxel resolution of 32^3 are comparable to that with polygon meshes. We find that between

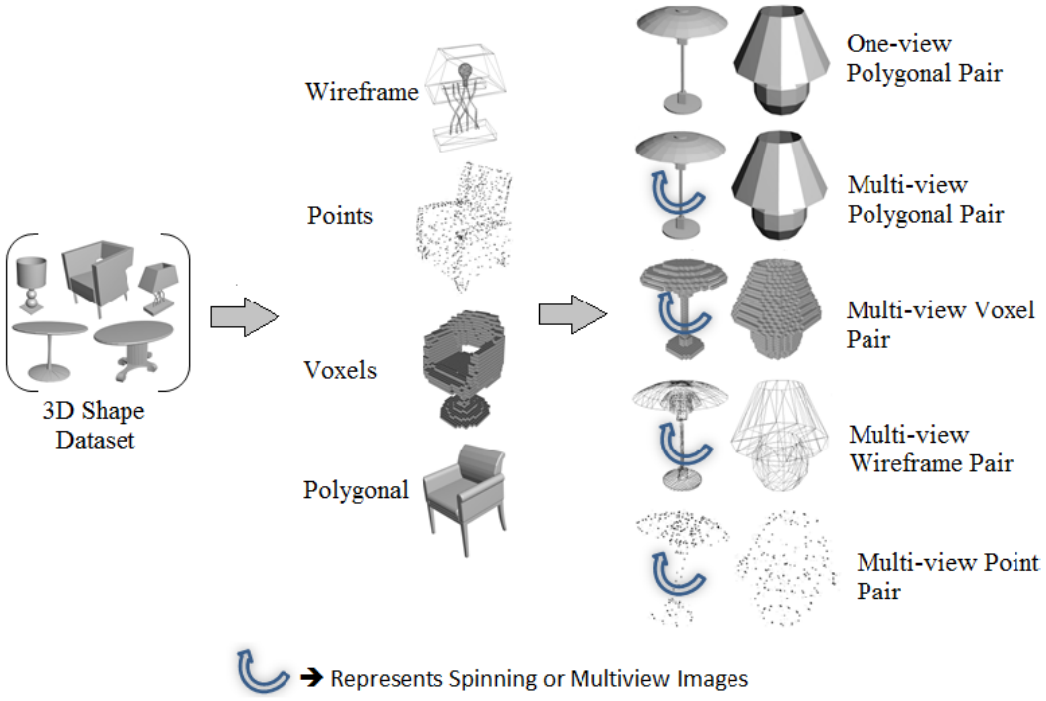


Figure 4.1.3: Steps used in generation of shape pairings. (a) Dataset of shapes having 12 categories, namely: chairs, tables, table lamps, air planes, abstracts, ashcans, bags, birdhouses, buildings, dishes, teapots, and vases. (b) Generation of images for four rendering styles: (top to bottom) wire-frame, points, voxels, and polygonal. (c) Pairing of images for crowdsourcing: (top to bottom) single view polygonal, multi-view polygonal, multi-view voxels, multi-view wire frames, and multi-view points.

single-view and multi-view representations, there is no significant differences in the human aesthetics preferences, suggesting that having just a single-view is enough. These observations lead us to say that humans tend to look at the global structure for comparing aesthetics of shapes in pairs and need not observe the details, and very detailed (continuous or high resolution) representations of shapes are not needed.

4.2 Approach

Our experiments are aimed at exploring whether the stimuli created using different rendering methods affect the human aesthetics preferences of shape pairs and how they compare between different methods. In this section, we describe the 3D shapes we used, the 3D modelling representations, the crowdsourced data collection, and the method used to compare whether different modelling representations give significantly different user aesthetic responses (Figure 4.1.3).

4.2.1 3D Shapes

Our dataset of 3D shapes comes from ShapeNet [17] online shape repository. We collect shapes belonging to twelve categories: Abstract (30), Air plane (20), Ashcan (18), Bag (15), Basket (15), Candelabra (28), Bench (15), Birdhouse(18), Building (15), Chair (30), Dish (15), Teapot (10), Lamp (30), Vase (20), Table (30), the number in brackets show randomly selected shapes for each category from ShapeNet online 3D repository. Since the downloaded meshes come with texture and colour information, we manually remove such information before using the shapes in our study. The shapes are already oriented and scaled. We generate pairs of shapes, where each pair comes from the same category. It makes more intuitive sense to compare a chair against another chair, rather than a chair against a lamp. For each category, we generate 60-30 shape pairs randomly.

4.2.2 Stimuli Creation

We convert each shape into these 3D modelling representations (stimuli) before using them in our data collection process:

- Single-view polygon mesh (or just 'single-view'). We create a single-view image that shows a representative forward-facing viewpoint of the polygon mesh.
- Multi-view polygon mesh (or just 'multi-view'). We rotate the mesh along the up-axis and have a slightly slanted up-direction to better show the 3D shape. We choose to take three seconds for each complete rotation followed by half a second of pause at the same representative viewpoint for single-view. These are then repeated continuously as a gif image and rendered with the same shading parameters as the single-view case.
- Wireframe mesh. It is created with original wireframe and with remeshed wireframe (appearing more uniform). We then apply a quadric-based edge collapse method to reduce the number of polygons to a desired number while maintaining the shape. For mesh simplification, we use MeshLab [22] built in filters.
- Point clouds. We use geodesic surface sampling [98] to get the desired number of points. We tested various cases in our experiments and eventually used meshes with 250, 500, 1000, 2000 points.
- Voxels. We tested various cases and took voxel resolutions of 16^3, 32^3, and 64^3. The voxels are rendered as small cubes.

The viewpoints are chosen by us and the shapes are consistently rendered in the same way by choosing similar lighting and shading parameters.

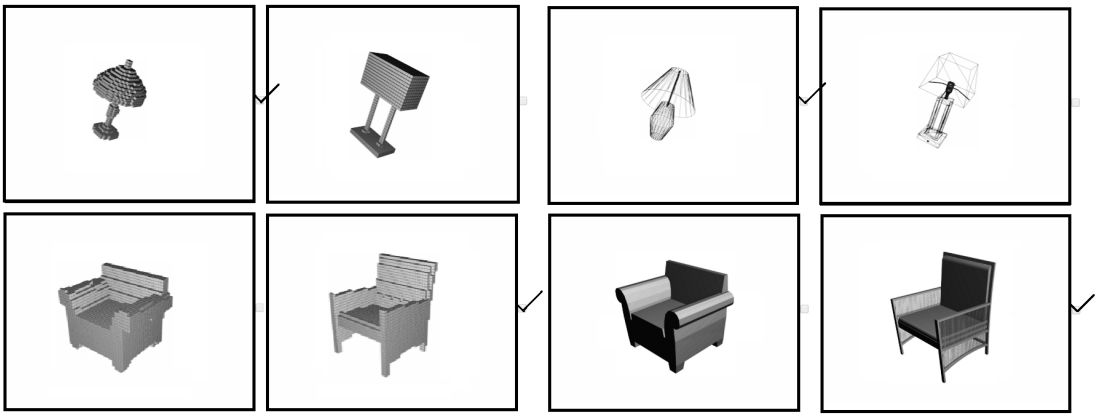


Figure 4.2.1: The interface on Amazon Mechanical Turk allows users to click anywhere on the image or the small box to the right to indicate the one they perceive to be more aesthetic. The left pair is for voxel resolution of 32^3 and the right pair is for polygon meshes.

4.2.3 Human Intelligence Tasks

Our study to collect crowdsourced aesthetics data is designed by taking inspirations from similar studies in evaluation of image aesthetics (Please see the 2 Related Work Section). We use double stimulus method to collect human preference data for our study [80]. All participants are shown pairs of static or animating images. In case of animations (gifs of 3secs in length) the first frame is shown for 1 secs. We ask the participants to select one shape from each pair that visually appears more aesthetic to them. They are required to select one shape from each pair.

Each participant is paid \$0.05 to \$0.10 on completing a HIT comprising 60 pairs of shapes. Each HIT was done by 25 participants. Since our setup requires data collection for a large number of 3D shapes, we use Amazon Mechanical Turk (AMT) crowdsourcing platform to collect aesthetic preferences. However, data collected from crowdsourcing platforms for such studies can be very noisy and unreliable. In order to collect quality responses, we use three strategies. First, we create a qualification test with 10 tasks. Only those participants who answered all the 10 tasks correctly were allowed to take the main HITs. The qualification test was designed with pairs showing an ugly shape with an aesthetics shape chosen by us manually. We manually created some distorted or ugly shapes (Figure 4.2.2) and paired them with normal shapes. Second, in the instructions before working on the HIT, we clearly state that un-honest workers Ids will be detected and blocked from future work offered by us. Third, to ensure that participants actually spend time to have multi-view aspect of the shape, when a participant has provided the response to 'xth' pair by clicking on the check box, the next pair's (i.e. 'x+1th') check box becomes clickable only after 4secs when pair 'x' was clicked. We filter out 'bad' participants by allowing only those with HIT acceptance ratios of 95% or more. We

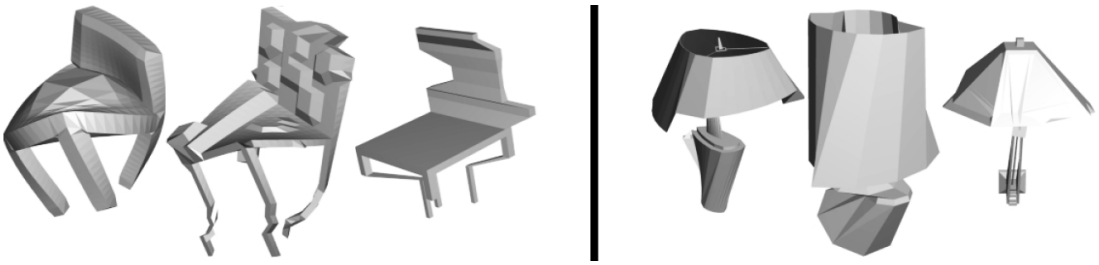


Figure 4.2.2: Manually created distorted shapes paired with normal shapes in qualification tests. If a participant does not answer pairs with ugly chairs correctly then he can't qualify to work on our tasks.

also collect age, gender, and geographic location as participants demographic information.

4.2.4 Demography

Since workers on a typical crowdsourcing platform come from all over the globe, collecting data about their backgrounds could help us better understand the nature of aesthetic responses collected. For example, people living in different regions (Africa, Asia, or America) may have different preferences for aesthetics owing to the cultural differences. Thus, in addition to collecting the aesthetic preferences, we ask user to provide the following information: gender, age, and the region where they come from.

In total, we had 1165 unique participants working on 2850 HITs, with 37.42%(436) males and 62.58% (729) females. The percentages of participants belonging to Africa, North America, S. America, Asia, Europe, Australia are 0.65%(8), 74.08%(863), 1.37%(16), 16.57%(193), 6.87%(80), 0.00%, and percentages of participants belonging to age groups 0-20, 21-30, 31-40, 41-50, 51-60, 61-100 are 3.35%(39), 37.34%(435), 30.04%(350), 17.00%(198), 7.55%(88), 4.64%(54) respectively. Based on these figures, we can say that the collected data represents diverse regional backgrounds and different age groups.

4.2.5 Comparing Responses

Our main aim is to make observations about perceptibility of shape aesthetics using stimuli created using several different representations. This hasn't been studied before especially in the context of shape aesthetics judgements on crowdsourcing platforms for input to data-driven methods. Specifically, using the collected data, we wish to study the effect of the modelling representation on participant choice, by comparing the results of two modelling representations each time. First, we compare between single-view and multi-view (of polygon meshes). We then use polygon mesh as a basis to compare between: polygon mesh and wireframe mesh (original and re-meshed wireframe), polygon mesh and point cloud (125, 250, and 500 points), and polygon mesh and voxels (16^3, 32^3, 64^3 resolutions). In layman terms,

if most people choose shape A for one representation of the pair (A,B), while most people choose shape B for another representation, then these two modelling representations lead to a significantly different user choice. We describe the modelling representations that we compare against, and the method to compare the data collected for two modelling representations to decide whether they lead to significantly different user responses.

We compare between two modelling representations by using Fisher's exact test [4]. This test tells us whether any differences that we observe in the proportions of (A,B) choices between two modelling representations is significant. As the number of responses for a particular choice can be small or even be zero, we choose Fisher's exact test which can handle these cases. The null hypothesis for each of the comparisons above is that the two modelling representations are equally likely to have the same proportions of choices. As an example of this test, to compare between polygon mesh and voxels (at a specific resolution), we take each shape pair (A,B) and observe the choices of 25 participants. For polygon mesh, we may have 18 participants choosing A and 7 choosing B. For voxels, we may have 9 choosing A and 16 choosing B. Intuitively (18,7) and (9,16) are quite different. Fisher's exact test gives a p-value of 0.022. Since the p-value is less than 0.05, this provides evidence to reject the null hypothesis at the 5 significance level or that the two modelling representations lead to significantly different proportions of responses (note that this was just an example to illustrate the process). We perform Fisher's exact test with the data for each shape pair. Then for all shape pairs for each shape category, we note the percentage of pairs where the null hypothesis is rejected or where the two modelling representations lead to different responses. The results in the next section show these percentages.

4.3 Results

We show and analyse the results to give insights into whether the modelling representation of 3D shapes affects the human aesthetic preferences of shape pairs. Please see the Table 4.2.1 for a quick summary of results.

4.3.1 Single-View vs. Multi-View

We compare between the single-view and multi-view representations by using the method described in the previous section i.e using Fisher's test. The percentages of shape pairs where we observe significant differences (according to Fisher's exact test at the 5% significance level) in the proportions of (A,B) aesthetic choices between single-view and multi-view are shown in the second column of the Table 4.2.1. The overall percentage for all shape pairs is 3.31%. These are the percentages of shape pairs where the null hypothesis is rejected. To consistently compare between the various types of modelling representations, we decide on a percentage

Category	Static	Points				Voxels		
		250	500	1000	2000	16	32	64
Abstract	2.22	17.78	11.11	13.33	8.89	20.0	8.89	11.11
Air plane	0.00	6.67	6.67	0.00	0.00	0.00	3.33	0.00
Ashcan	2.22	17.78	11.11	13.33	6.67	17.78	2.22	0.00
Bag	3.33	6.67	3.33	6.67	6.67	20.0	0.00	3.33
Basket	6.67	16.67	10.00	26.67	26.67	6.67	3.33	6.67
Bench	4.44	22.22	11.11	31.11	26.67	35.56	22.22	11.11
Birdhouse	9.30	11.63	11.63	11.63	11.63	6.98	6.98	0.00
Buildings	0.00	24.44	8.89	6.67	4.44	15.56	4.44	0.00
Candelabra	6.82	6.82	4.55	2.27	13.64	29.55	20.45	6.82
Chair	1.70	10.00	6.70	8.33	5.00	13.30	5.0	5.0
Dish	2.22	2.22	8.89	0.00	0.00	8.89	0.00	2.22
Teapot	0.00	10.34	0.00	3.45	0.00	6.90	0.00	3.45
Lamp	6.70	6.70	6.70	5.00	5.00	11.7	10.0	4.65
Vase	2.33	20.93	11.63	16.28	9.30	13.95	16.28	11.11
Table	1.70	18.30	8.30	3.33	5.00	11.70	1.7	0.00
Average	3.31	13.28	8.04	9.87	8.64	14.57	6.99	4.36

Table 4.2.1: Numbers representing percentages of shape pairs where null hypothesis is rejected by Fisher's test between polygonal and three other representations: static (column 2), points (columns 3-6 for 250, 500, 1000, and 2000 points respectively), and voxels (columns 7-9 for 16, 32, and 64 voxel resolutions respectively).

of shape pairs where the null hypothesis is rejected that is still acceptable. This is a parameter that we choose. We choose that 10% or less is 'acceptable' meaning that two modelling representations have similar proportions of (A,B) responses.

Figure 4.3.1 shows examples of shape pairs where the single-view and multi-view cases give either the same or quite different (A,B) responses. In (c), we see that the left chair when rotated has a bottom part that appears to be more hollow, while the right chair's bottom part looks more normal and appealing. This explains the difference in the (A,B) choices shown in (b). In (f), we see that although the left lamp looks simple from one viewpoint, it has a nice curved geometric shape that looks appealing when rotated. This explains the difference in the (A,B) choices shown in (e). In (i), we see that the left table has a nice shape when rotated although it looks a bit squashed from a single viewpoint. The right table looks nice from one view, while the rotation would expose more of the rectangular table top (which users seem to prefer less). This explains the difference in the (A,B) choices shown in (h). Based on the results, we conclude surprisingly that the single view polygon mesh and multi-view polygon mesh have similar proportions of aesthetic responses. The implication is that having a single-view is enough even though the shapes are in 3D, at least for the shape categories we tested. However, there were a few cases where the single-view and multi-view lead to significantly different proportions of (A,B) responses. These tend to be cases where some additional in-

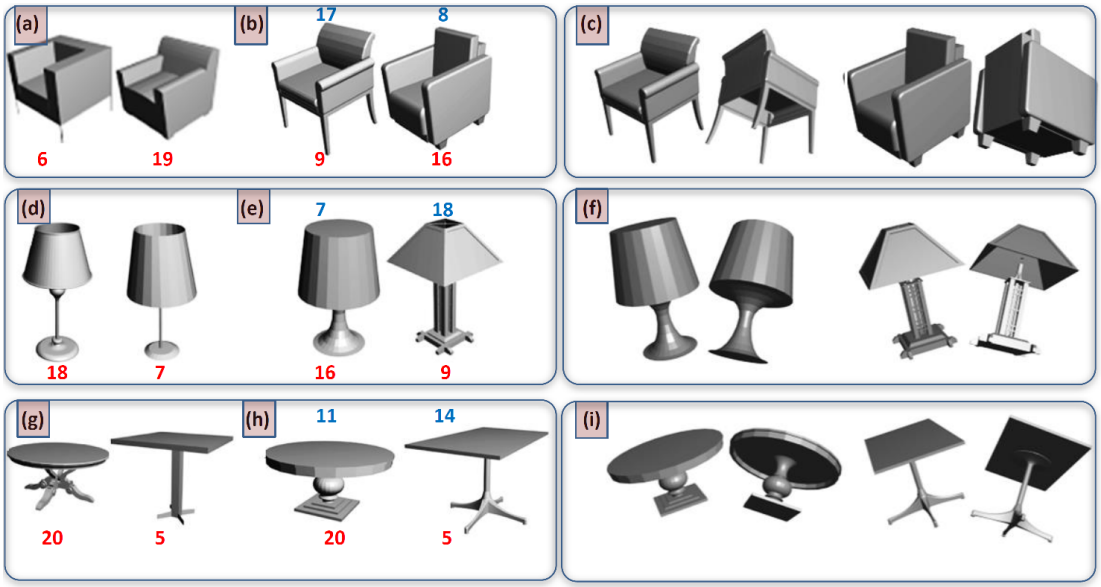


Figure 4.3.1: Single-View vs. Multi-View: Examples of shape pairs with user (A,B) responses. (a) A shape pair of chairs with the same (A,B) responses (numbers below shapes) for both single-view and multi-view. (b) A shape pair of chairs with quite different responses between single-view (numbers above shapes) and multi-view (numbers below shapes). (c) Two views for each of the shapes in (b). The second row shows the same type of examples as in first row but for lamps, and the third row is for tables.

formation about a 3D shape is provided by other viewpoints. Therefore, we suggest that the shape category and the shapes themselves are important when considering to select between the single-view and multi-view representations.

4.3.2 Polygon Mesh vs. Point Clouds

We compare between the polygon mesh and point cloud representations. The percentages of shape pairs where we observe significant differences (according to Fisher's exact test at the 5% significance level) in the proportions of (A,B) aesthetic choices between polygon mesh and point clouds (for 250, 500, 1000, and 2000 points respectively) are shown in Table 4.2.1 in third, fourth, fifth, and sixth columns respectively. We can see that the average values of Fisher's test score gradually decrease with increase in number of points. Figure 4.3.2 shows examples of shape pairs where the polygon mesh and point cloud cases give either the same or quite different (A,B) responses. With 250 points per mesh, we can see that it is a sparse representation of the overall shape. For (b, d, f),we attempt to explain the difference in the (A,B) choices between the two cases. In (b), the left chair's point cloud does not show as much curvature compared to its polygon mesh, while the right chair's point cloud seemingly show a better structure of a chair compared to the left chair. In (d), the right lamp's intricate pattern in the polygon mesh is much less visualisable in the point cloud. The structure of the lamp base and shade are also less clear and more planar in the right lamp's point cloud. In (f),

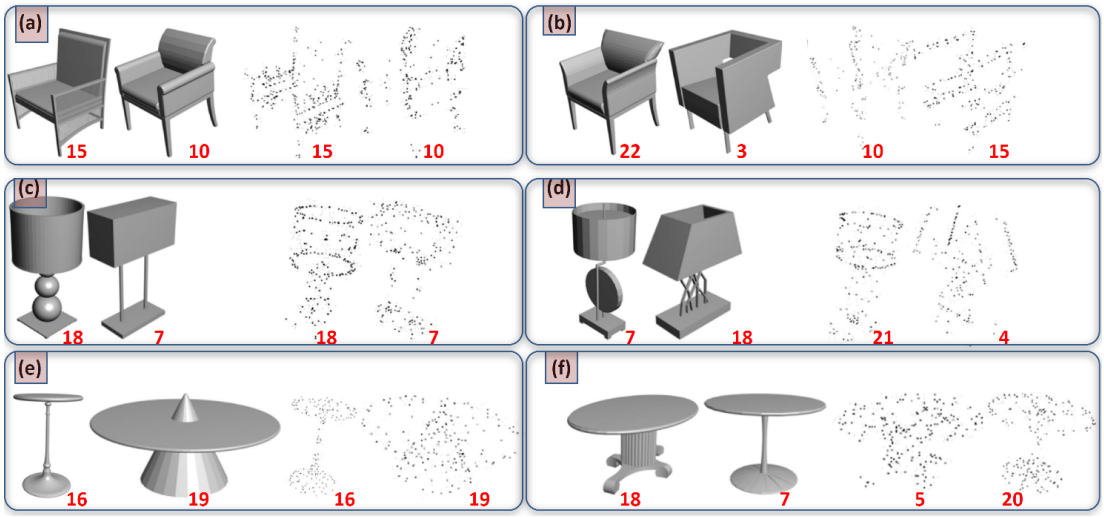


Figure 4.3.2: Polygon Mesh vs. Point Clouds (250 points): Examples of shape pairs with user (A,B) responses (numbers below shapes). (a) A shape pair of chairs with the same (A,B) responses for both polygon mesh and point clouds. (b) A shape pair of chairs with quite different responses between polygon mesh and point clouds. The second row shows the same type of examples as in first row but for lamps, and the third row is for tables.

the left table's polygon mesh shows a more elegant table, while the right table's point cloud shows a more normal-looking table. Based on the results, we do not find a consistent number of points across the shape categories that are just comparable to polygon meshes. However, it is clear that a relatively small number of points is enough. For instance, a good number of points to use for chairs is 250, for lamps is 125, and for tables is 500. This is surprising as these numbers of points are very small compared to the thousands of vertices that some shapes have. In some cases, the point representation can miss some shape details or even parts of the shape (e.g. the lamp pole). This implies that participants typically do not need to observe the details of a shape to make the aesthetics choices.

4.3.3 Polygon Mesh vs. Voxels

We compare between the polygon mesh and voxel representations. The percentages of shape pairs where we observe significant differences (according to Fisher's exact test at the 5% significance level) in the proportions of (A,B) aesthetic choices between polygon mesh and voxels (for resolutions of 16^3, 32^3, and 64^3 respectively) are shown in the last three columns of the Table 4.2.1. Figure 4.3.3 shows examples of shape pairs where the polygon mesh and voxel cases give either the same or quite different (A,B) responses. For (b, d, f), we attempt to explain the difference in the (A,B) choices between the two cases. In (b), the right chair as a polygon mesh looks fancy, while its voxel representation makes it look more planar. In (d), the left lamp as a polygon mesh look fancy, while its voxel representation smooths some details

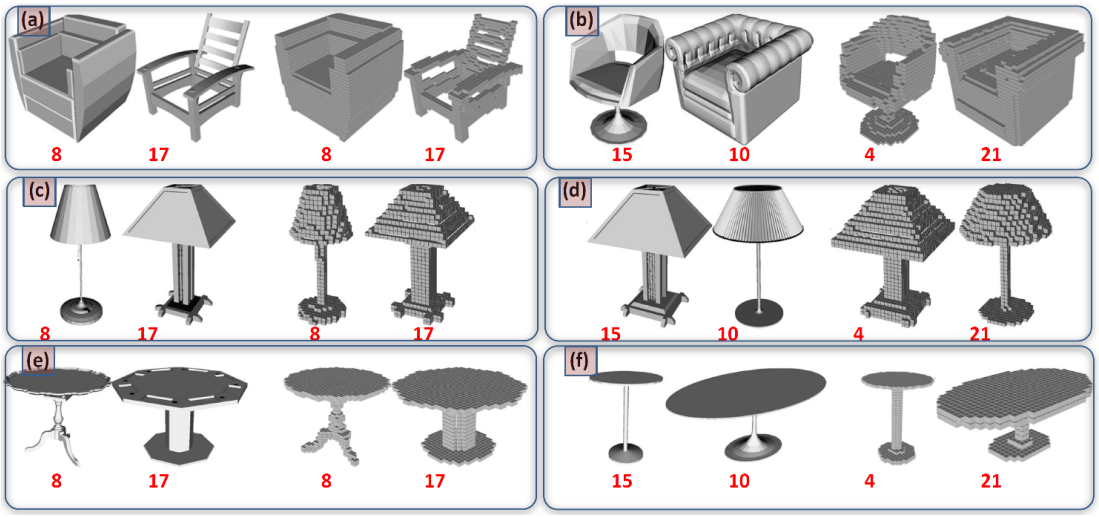


Figure 4.3.3: Polygon Mesh vs. Voxels (resolution of 32^3): Examples of shape pairs with user (A,B) responses (numbers below shapes). (a) A shape pair of chairs with the same (A,B) responses for both polygon mesh and voxels. (b) A shape pair of chairs with quite different responses between polygon mesh and voxels. The second row shows the same type of examples as in first row but for lamps, and the third row is for tables.

in the lamp pole and creates some jagged artefacts in the lamp shade. In (f), the right table as a polygon mesh looks nicer, while the left table's voxel representation looks more normal due to its circular shapes (as opposed to more oval for the right table). Based on the results, we conclude that the polygon mesh and a voxel resolution of 32^3 have similar proportions of aesthetic responses, while a resolution of 16^3 clearly leads to higher percentages of 'different' responses. We therefore suggest using a voxel resolution of 32^3 for collecting aesthetics data of shape pairs. This result is surprising as a resolution of 32^3 is quite small and in some cases can miss many details of the shape. Similar to the previous subsection for point clouds, this also provides evidence that participants typically do not need to observe the details of a shape to make their choices.

4.3.4 Polygon Mesh vs. Wireframe Mesh

We compare between the polygon mesh and wireframe mesh representations for three categories only: chairs, lamps and tables. The percentages of shape pairs where we observe significant differences (according to Fisher's exact test at the 5% significance level) in the proportions of (A,B) aesthetic choices between polygon mesh and 'original' wireframe are 8.3% for chairs, 5.0% for lamps, and 10.0% for tables. The overall percentage for all shape pairs is 7.8%. The percentages between polygon mesh and 're-meshed' wireframe are 5.0% of chairs, 1.7% for lamps, and 5.0% for tables. The overall percentage for all shape pairs is 3.9%. Figure 4.3.4 shows examples of shape pairs where the polygon mesh and wireframe mesh cases give either the same or quite different (A,B) responses. For (b, d, f), we attempt to explain the dif-

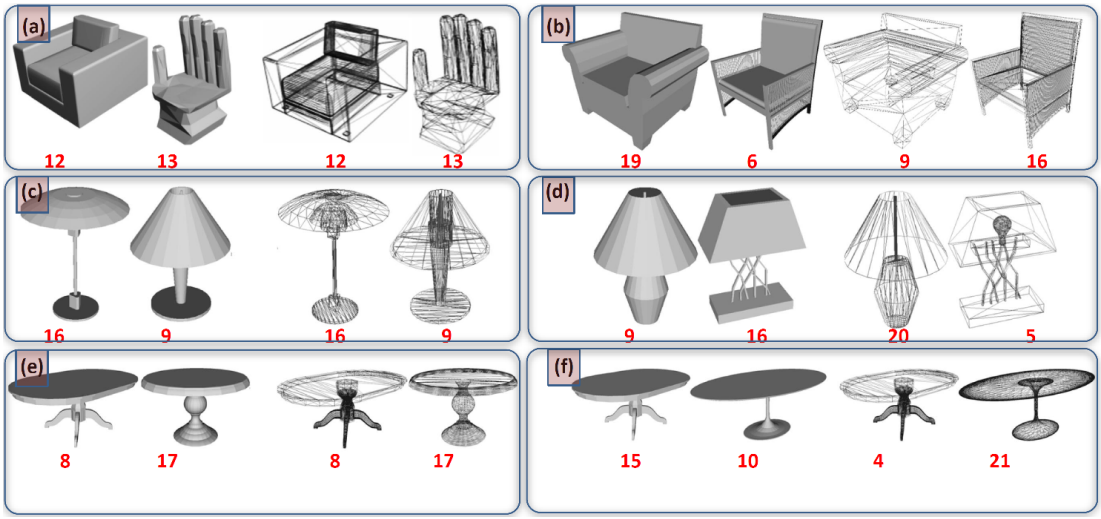


Figure 4.3.4: Polygon Mesh vs. Wireframe Mesh (original): Examples of shape pairs with user (A,B) responses (numbers below shapes). (a) A shape pair of chairs with the same (A,B) responses for both polygon mesh and wireframe mesh. (b) A shape pair of chairs with quite different responses between polygon mesh and wireframe mesh. The second row shows the same type of examples as in first row but for lamps, and the third row is for tables.

ference in the (A,B) choices between the two cases. In (b), the left chair's wireframe is more coarse as some large planar parts are composed of a small number of large triangles. In (d), the right lamp's wireframe shows the light bulb and more of the intricate pattern to make the overall shape look more strange. In (f), the left table's wireframe is not uniform in the sizes of the triangles, as the table top has larger triangles and the table leg has smaller triangles. Based on the results, we conclude that the polygon mesh can be the 'same' as the original wireframe as they have similar proportions of aesthetic responses. If we create a re-meshed wireframe that has a more uniform triangle size throughout the shape (such that for example large planar surfaces are represented by more and smaller triangles), the re-meshed wireframe is better in that it leads to a lower number of different aesthetics (A,B) cases compared to the original wireframe. The implication is that showing the wireframe mesh either way is just as good as the polygon version for collecting aesthetics data for shape pairs. In addition, from our experiences, the transparency of a wireframe mesh is sometimes helpful to show more details of the shape as we may not need to rotate the shape to show the 3D aspects or any occluded parts. However, the transparency may also reveal some inner parts that are not necessary or makes the overall shape more strange to visualise (e.g. seeing the details of a light bulb and wiring inside the lamp shade that one normally would not observe).

4.4 Discussion

With a goal to build a data-driven model of shape aesthetics, we first conduct a crowdsourcing study having implications in quality of human aesthetics judgement data and design of machine-learning methods. We argue that several design parameters are important for such studies needed to collect data for building data-driven models. In our preliminary analysis of the problem of learning a measure of shape aesthetics, we realise that such parameters are related not only to "human perception and discrimination of shape aesthetics" but also for "helping choose appropriate shape representation as input" to data-driven technique. We clarify these two points further. First from the perspective of "human perception and discrimination of shape aesthetics", we demonstrate that human judgement data collected using single-view object pairs is similar to data collected using multi-view object pairs. However there are subtle differences (Section 4.3.1) due to multi-view stimuli presenting more shape information due to having several shape viewpoints. The results suggest that crowdsourcing studies can be designed to show single-view shape pairs to gather human aesthetics judgement data. Our original impression was that, since a multi-view object image allows perception of shape details from several viewpoints, usage of single-view and multi-view images would result in substantial difference in collected judgements. However, this is not the case even for non-symmetric shapes used in our study. Second, for choosing appropriate shape representation as input to data-driven techniques, our comparison of collected aesthetics judgements using different shape representations shows that shape details don't matter much when humans compare aesthetics of shapes in pairs. Specifically, the voxelized shape representations are as good as the polygonal shape representations for collecting shape aesthetics judgement data. This is an important result as much of the deep learning based data driven methods learn on non-polygonal shape representations, such as voxels or point clouds. In our view, the general belief is that using one of such representations, say voxels, as input to data-driven algorithm is not an ideal case as these results in loss in shape details. For example, compared to polygonal representation, voxelized representations have non-smooth surfaces.

4.4.1 Shape Aesthetics Perception

The focus of this work is crowdsourcing study design and perceptual aesthetics data collection for data-driven modelling of visual aesthetics of 3D shapes. We argue that ideas from our study can be generalised to study perception of aesthetics from representations Figure 4.1.1 not used in our study. For example, non-photorealistic rendering techniques [23] in computer graphics have studied shape depiction using different styles, such as line drawings. It is unclear that how perception of aesthetics is affected by usage of such abstracted shape representations. Although the participants are provided clear instructions to click on a shape they think is more aesthetic, however we believe that their responses may have been influenced by

participants unconsciously thinking of other forms of aesthetics, such as functional aesthetics. Specifically, consider a chair for example, it may appear beautiful structurally however looking at it may give the impression that it may not be comfortable to sit on, or it may have low functional aesthetics. Thus, investigating the question "do participants unconsciously think about other shape properties such as perceived functional aesthetics or perceived ergonomics or interestingness?" may help throw more light on the gamut between structural and functional aesthetics. Our study collects aesthetics judgement data by showing shape pairs belonging to shape categories, for examples, chairs are paired with chairs and buildings are paired with buildings. We make this pairing choice by assuming that it is easier for humans to compare shapes from same structural and semantic category. It is relatively unclear how humans would respond if they are asked to compare shapes coming from different object categories, for example a chair is compared with a table. Finally, in this chapter we show that crowdsourcing allows collection of large amount of human perception data from participants coming from diverse geographic and cultural backgrounds, hence, replicating the already existing lab based low participant perception studies or conducting new ones [44] on crowdsourcing platforms allows for better generalisations of results and findings.

4.4.2 Stimuli Presentations

We would like to highlight the considerations in the presentation of stimuli that have implications in the design of algorithms trying to use the collected data for modelling the perceptual property under consideration. The most important is the collection of data as relative comparison judgements for pairs of shapes rather than as an absolute aesthetics score for each shape. The way judgement data is collected has implications in the formulation of data-driven algorithm. For example, an absolute aesthetics score collected for each shape allows formulation of this problem as a supervised learning problem with availability of direct labels. Hence relatively less complex deep neural network formulation can be used to model aesthetics in this case. However in case of data collected as paired responses, formulation of deep learning neural networks is more complex as there are no direct labels for each shape. Since it is easier for humans to provide their aesthetics judgements as relative comparisons, we adopt pairwise data collection approach. We suggest that other forms of data collection approaches such as using a continuous rating bar, allowing data collection as 'bad', 'poor', 'fair', 'good', and 'excellent' could take ideas from our data collection approach. Further, while collecting data in our study, participants are forced to choose one shape as more aesthetic for each pair. Another possibility for data collection process that has implication in the data-driven modelling, is to let participants to provide options such as 'same' or 'none', if they think for the given pair both objects have similar aesthetics or none is aesthetic, respectively.

4.4.3 Conclusion

Visual aesthetics is one of the fundamental properties of man made 3D shapes. To our knowledge, this is the first work attempting to evaluate how the shape representations affect the human judgements of 3D shape aesthetics. Although, results of this study are restricted to the shape representations used in our experiments, but can help provide useful insights on perception of aesthetics and building data driven models. The major contribution of this work is in suggesting that human aesthetics judgements for shape pairs are not affected by the usage of less detailed shape representations. This result has implications for the selections of shape representations as input to deep learning based data-driven models.

Beauty depends on size as well as symmetry. No very small animal can be beautiful, for looking at it takes so small a portion of time that the impression of it will be confused. Nor can any very large one, for a whole view of it cannot be had at once, and so there will be no unity and completeness.

Aristotle

5

Measuring Perceptual Aesthetics of 3D Shapes

In this chapter, we learn to predict visual aesthetic value or beauty of 3D shapes. To do so, we employ a more natural way to learn the aesthetics using deep neural networks rather than learning on pre-computed shape features, though the crowdsourcing setup is similar to previous two problems. The observations made in previous chapter on aesthetics perception from different 3D shape representations form an important input to the method presented in this chapter. We learn a model of 3D shape visual aesthetics that aligns with human aesthetic judgements well as it is built on large amount of visual aesthetic judgement data collected on Amazon Mechanical Turk crowdsourcing platform. In this chapter, we describe our approach, demonstrate results and applications, and discuss the potential future work.

5.1 Introduction

We are hardwired to seek beauty or aesthetics in everything around us to experience pleasure and satisfaction [32]. The idea of aesthetics can be applied to a variety of completely unrelated things, such as poems, clothes, landscapes, chairs, flowers, dance performances, music, and food etc. The traditional attempts to explore aesthetics focused on mathematical criteria such as minimum description length and self-similarity [97]. Although, previous work has tried to define aesthetics in different ways [6, 55], in our work we consider beauty as a widely accepted and simplest notion of aesthetics and thus we interchangeably use the terms 'aesthetic value', 'beauty', and 'visually pleasing' to have the same meaning. In this chapter, we focus on learning

Figure 5.1.1: Shape design (first row) and image (second row) aesthetics examples. Curvilinear design (top left) results in stronger pleasure rating than rectiliner design (top right) [25]. Next row, first two images with high aesthetic value, followed by two images with low aesthetic appeal [71]

to predict aesthetic value of 3D shapes, where we consider aesthetic shapes as those that are visually attractive or pleasing. In addition, we present an analysis of computable shape features in relation to shape aesthetics.

Although, perception of aesthetics can be subjective [92], irrespective of our backgrounds and experiences, our brains find certain shapes as universally pleasing [37, 40, 45] (consider two scenes in top row in Figure 5.1.1). The visual attractiveness of the form of a product plays an important role in the decision towards purchasing that product [25]. We argue that computational understanding of shape aesthetics can help us to more effectively use large datasets of 3D shapes, for search, scene composition, and visualisation for example. Furthermore, learning to predict shape aesthetics can contribute to other fields such as psychology, neurology, and philosophy.

As in our style similarity work (Chapter 3), we consider aesthetics as a perceptual concept. Consequently, to build a data-driven system, we collect a large amount of shape aesthetics preferences data from humans and learn from that data. Our data collection study is carried out on Amazon Mechanical Turk, where each participant records their responses for a set of questions showing paired shapes, by selecting one shape as more aesthetic or pleasing. We believe, it is relatively difficult for humans to give an absolute aesthetics score to a single shape [41, 61, 67, 74], so we use paired comparison task [92]. Since we use several shape categories to experiment with, an important question to ask is, "how to choose which shapes to pair for aesthetic judgement data collection?" To this end, we can either pair shapes that belong to same high level category [1] or sub-category (e.g. dining chair with dining chair) or allow pair-

[1] The input to our data-driven methods is a collection of semantically-related man-made shapes, mainly taken

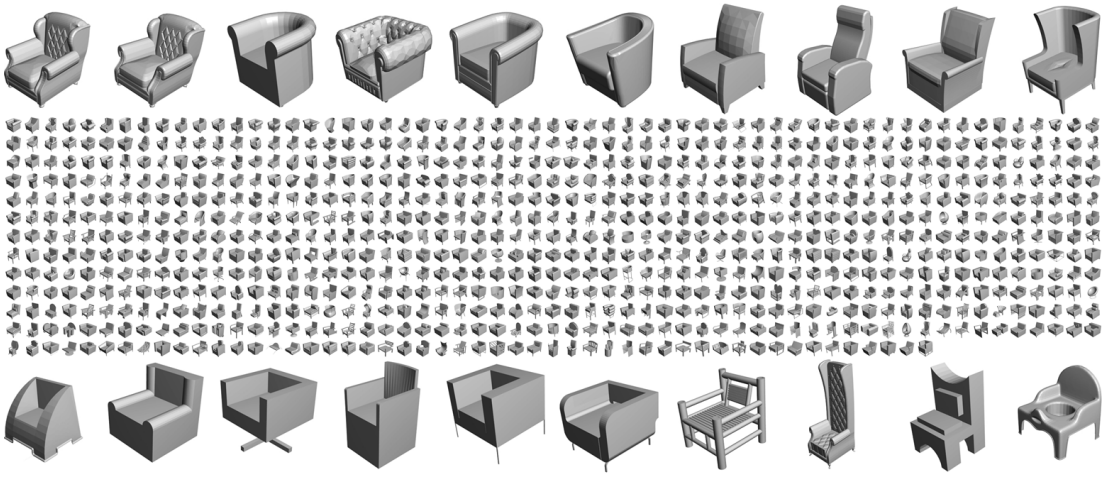


Figure 5.1.2: Example showing a large number of 3D shapes (i.e. chairs) ranked from high to low (left to right and top to bottom) aesthetic scores. Top and bottom rows are shown in large size to see the difference clearly. Given this kind of aesthetics ranking, a user can easily find what she is looking for.

ings between shapes belonging to two completely different categories (e.g. chair with bed). However, in this work we focus on pairing shapes belonging to same category and leave the exploration of other way of pairing as future work, since it is interesting to see from perceptual and consistency points of view how humans compare aesthetics of two shapes that come from two different categories.

There exist several interesting works that look into the problem of learning and analysing aesthetics of 2D content, such as images [66, 103] (Figure 5.1.1) and human faces [34, 62, 104]. In case of 3D shapes, almost every work [12, 45, 86, 119] focuses on manually defined features to evaluate aesthetics, prominent ones include curvature, symmetry, and mathematical properties such as bending energy and minimum variation surface. In contrast, our approach to model 3D shape aesthetics isn't based on the use any hard-coded rules or any manually-crafted features. We learn directly from raw shape data using human aesthetic judgements, allowing us build a model of shape aesthetics that aligns with human aesthetics preferences.

We demonstrated the ability of humans to perceive shape aesthetics from different shape representations in previous chapter. We take inspiration from there and aim to learn directly from the volumetric shape representation or voxelized volume. To this end, we exploit deep neural networks, which have been shown to learn complex and non-linear functions, to compute the shape aesthetic measure. Specifically, we take a deep convolutional 3D shape ranking approach to compute our aesthetics measure since our collected data is ranking-based (e.g. one shape is more aesthetic than another). The deep architecture allows us to autonomously

from ShapeNet [17] large-scale online repository, which provides multitude of semantic categories and organises them under the WordNet [84] taxonomy.

Class of 3D Shapes	Number	$\lvert\mathcal{I}_{train}\rvert$	$\lvert\mathcal{I}_{validation}\rvert$
Club Chairs	778	7600	400
Pedestal Tables	40	2578	297
Mugs	75	743	82
Lamps	88	2250	250
Dining Chairs	277	4790	310
Abstract Shapes	40	600	60

Table 5.2.1: We separate the total number of collected samples in each class into a training data set $\mathcal{I}_{train}$ and a validation data set $\mathcal{I}_{validation}$. $|\mathcal{I}|$ is the number of samples in $\mathcal{I}$. Please note that one sample of data here is one human response to one pair of shapes.

learn features from raw voxel data. The raw voxel data is the first layer to a deep neural network that computes an aesthetics score for each mesh. To learn the weights for the network with our ranking-based data, we use a deep convolutional ranking formulation and backpropagation that uses two copies of the deep neural network. After training the network, we can use one copy of the learned network to compute an aesthetics score for a new 3D mesh of the corresponding object type.

This chapter is organised as follows. In the next section, we describe the design of crowdsourcing study to collect shape aesthetic judgements, analysis of collected data to shed light on participant consistencies, deep learning formulation, and testing strategy. The results sections addresses the qualitative patterns, quantity of training data, comparison of network architectures, failure cases, link between shape features and their aesthetic scores, aesthetic duality, and applications. Finally, we present the conclusion.

5.2 Approach

To model 3D shape aesthetics, our overall framework involves first to collect data on the human perception of visual shape aesthetics and then learn from the data to get an aesthetics measure.

5.2.1 Crowdsourcing Shape Aesthetics Judgement Data

Our crowdsourcing setup is similar to one used in previous chapter, where we show shapes in pairs and ask participants to click on more aesthetic shape. We collect a 3D shape dataset (Table 5.2.1) from ShapeNet [17], where models are already grouped into human understandable categories, and also already rotated and scaled correspondingly with the other models in the same category.

Quality Control and Participant Backgrounds. In each HIT (Human Intelligence Task is a set of questions) Figure 5.2.1, we collect data for 30 shape pairs and pay $0.10. In an attempt

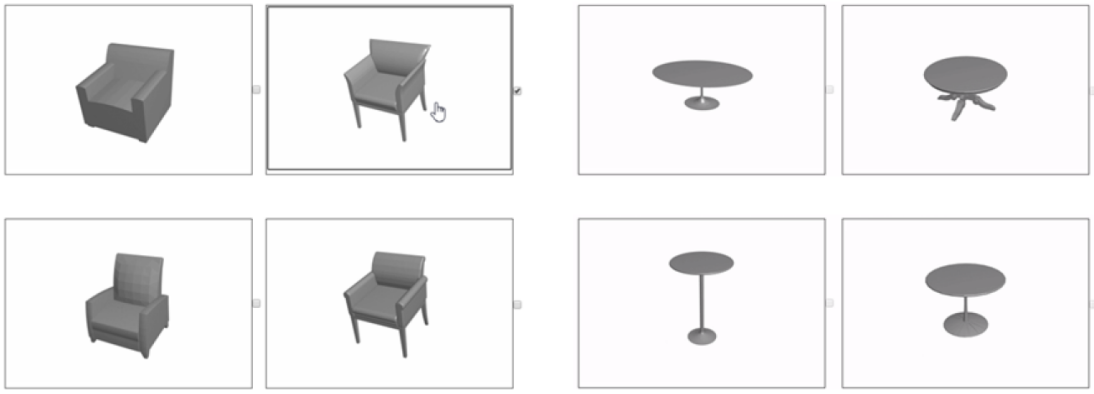


Figure 5.2.1: Example shape pairs for use in Human Intelligence Task on Amazon Mechanical Turk crowdsourcing platform.

to collect as good data as possible, we use various methods during the data collection. First, we provide clear instructions to tell the potential participants that dishonest workers can be blocked from doing future work. We classify a worker as dishonest if he repeatedly answers control questions incorrectly. Second, we force participants to spend enough time to view the shape and then provide their response. To do this, after users click on a response, we have a 4 second delay before they can click on the next response. Third, we include control questions as in previous work [41] where one shape from the pair is intentionally made to be ugly (see Figure 5.2.2 for some examples). In each HIT, we include five control questions and the user must correctly answer all of them for us to accept the tasks in the HIT. At the start of each HIT, we also collect some demographics data from the participants. This data includes their gender, age group, and region. We had 403 male and 360 female Turkers (and 12 who did not provide their gender). The HIT acceptance rates based on gender are 87.1% for males and 82.8% for females. We had the following age groups: (0-20, 21-30, 31-40, 41-50, 51-60, 60-100) and the percentages of Turkers in each group respectively are: (1.6%, 36.0%, 37.1%, 14.3%, 9.6%, 1.3%). We had the following regions: (Africa, Asia, Australia, Europe, North America, South America). The HIT acceptance rates based on region are: (N/A due to no Turkers, 85.1%, 100%, 87.9%, 85.0%, 77.3%).

Consistency Analysis

In data-driven shape analysis and processing, the term consistency is used to check robustness of the collected data and thus can be computed in different ways [67, 74]. In this work we perform two types of consistency checks. First, to see if crowdsourcing is a good option to collect large number of judgements, and then the level to which judgements agree to a majority. As demonstrated in [102], there can be differences between paid and unpaid participant responses. To verify this in our case, we collect and compare responses for each 25 shape pairs,

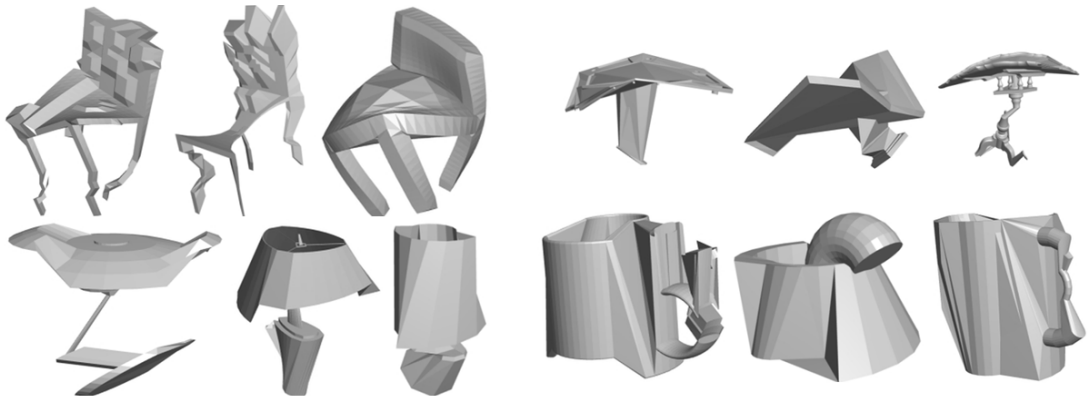


Figure 5.2.2: Example shapes intentionally distorted to look ugly for pairing with normal shapes to check participants responses in crowdsourcing study. First row, three ugly chairs and three ugly tables. Similarly, in second row we have three lamps and three mugs.

first from 15 unpaid participants recruited on Facebook and then from 15 paid participants recruited on Amazon Mechanical Turk. We perform Fishers Exact Test at 0.05 significance level on all shape pairs and found that it does not reject the null hypothesis of non-random association between the responses collected from two platforms i.e. on Facebook and Amazon Mechanical Turk. This result provides the necessary motivation to use crowdsourcing to collect large number of shape aesthetic judgements. In a second test, after splitting the responses for 25 shape pairs by 42 participants into male or female categories, and then performing Fisher's test as in previous case, we get the same results i.e. non-rejection of null hypothesis, allowing us to conclude that collected responses are consistent based on the populations.

5.2.2 Deep Ranking

This section describes how we learn an aesthetics measure from the collected data described in the previous section. We take a deep multi-layer neural network architecture to allow us to learn a potentially complex and non-linear function of shape aesthetics. The learned function maps input shape data to its aesthetics score. We use voxels to represent 3D shapes which has been shown to be an effective representation for deep learning [127]. Further, the study in previous chapter gives us confidence to choose voxel based shape representation as input to our deep ranking techniques. Specifically, we found in Chapter 4 that slight loss in shape detail with voxel representation does not affect human judgement of shape aesthetics in pairs. This implies that voxel representation carries enough visual information to help learn a measure of 3D shape aesthetics. We choose to experiment with voxels at different resolutions as input to the neural network.

As our collected data is based on rankings of pairs of 3D shapes, we use a learning technique commonly known as learning-to-rank [94] to compute an overall measure that 'best fits' with

the paired rankings in the data. As we collect pairs of data, we take a deep ranking formulation that is inspired by [51, 61, 134] and that fits well with our collected data and problem. The novelety of our method include: our data compares between pairs of 3D shapes instead of points on the shapes, and we convert the shapes into their voxel representations instead of working with images taken from different viewpoints of the shapes. If the voxel resolution is high, we use a convolutional neural network architecture as otherwise a fully-connected network would not be practical to train. We experimented with various voxel resolutions and neural network architectures to gain insight into what works well for this shape aesthetics problem. Furthermore, we have two copies of the neural network (which takes the concept of Siamese networks [9] instead of four copies in the backpropagation. We train a separate network for each category of shapes. Specifically, 3D shapes representing real life objects of varying sizes and topology, such as mugs, chairs, tables, air planes, buildings, and vases are examples of the shape categories used in our study of objects for which we collect aesthetics judgement data.

Since the perception of aesthetics of a 3D object shape, for example a chair, can be subjective, we leave this to our neural network to model. Ours is a data-driven model that takes human aesthetic judgements for shapes in pairs and adjusts the network weights to best fit the judgements.

We first describe the voxel data representation and the neural network architectures. We next describe the deep ranking formulation and the backpropagation in the neural network that works with the collected data pairs. After the training process, we have an aesthetics measure that gives a score for each 3D shape.

Voxel Data Representation

Our choice of voxel representation is motivated from two key observations: first this is a basic representation from which more complex features may be computed and second, as demonstrated in previous chapter, the human perception of shape aesthetics can be done at manageable voxel resolution. We voxelize each mesh and the voxels become the input to the first layer of the deep neural network.

We experimented with different neural network architectures (Figure 5.2.3). We can have a low resolution voxel representation, where the nodes between each successive layers are fully-connected. As we increase the voxel resolution, we need a convolutional architecture. We do not use any pooling layers as we wish to keep the details of the shapes in the voxel representation. The motivation for specifically experimenting with different resolutions is that when humans make decisions on shape aesthetics, we may observe the overall shape and this corresponds to a lower resolution and fully-connected layers for the whole shape (Figure 5.2.3 top diagram). We may also observe the details of the shape and this corresponds to a higher resolution and some convolutional layers to recognise more local features of the shape (Figure

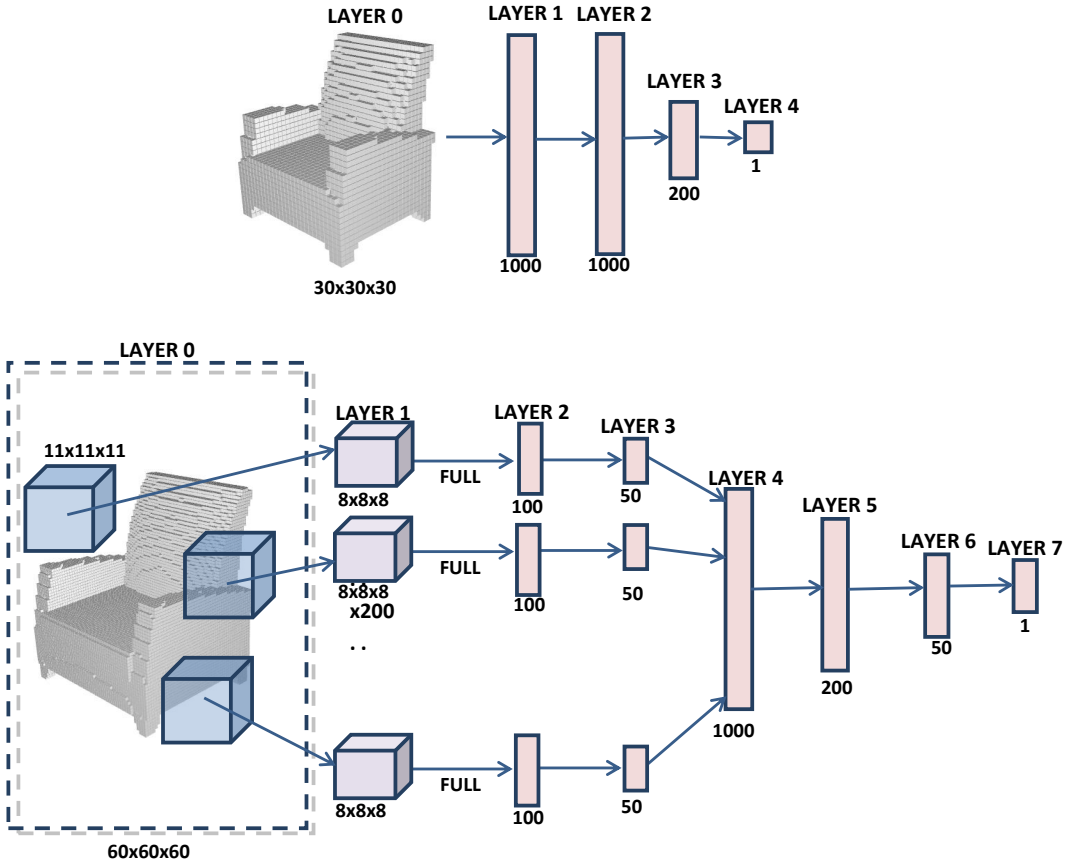


Figure 5.2.3: Example deep neural network architectures: fully connected (top) and convolution network (bottom). The input in layer 0 is the voxel representation of a 3D shape and the output in the last layer is the shape's aesthetics score. We experiment with convolution neural network with 128^3 voxel resolution.

5.2.3 bottom diagram). The convolution network takes as input voxel volume and transforms it to a score using a set of convolution and fully connected layers. For example, if we have a kernel size of 15^3 [3375 values] and a stride of 7, the total number of volume patches (or kernels) will be 512 for a input voxel size of 64^3. We choose to reduce 3375 values in each volume patch to 200 (can be a different value) values resulting in a [512x200] matrix as activations to first layer, which can be represented as [8x8x8x200] activation matrix.

We let $\mathbf{W}$ be the set of all weights consisting of $\mathbf{W}^{(l)}$ between each successive layers, where $\mathbf{W}^{(l)}$ is the matrix of weights for the connections between layers $l-1$ and l. We use the neural network in Figure 5.2.3b to provide some examples. In this case, $\mathbf{W}^{(6)}$ has 50x200 (between Layer 5 and Layer 6) values. Between layers 2 and 3, there are multiple (e.g. 200 in this case) sets of such weights. For the convolutional layer between layers 0 and 1, there are 1331x200 weight values. We let $\mathbf{b}$ be the set of all biases consisting of $\mathbf{b}^{(l)}$ for each layer except for layer 0, where $\mathbf{b}^{(l)}$ is the vector of biases for the connections to layer l. For example, $\mathbf{b}^{(6)}$ has 50x1 values. Between layers 2 and 3, there are multiple sets of such biases. For the convolutional

layer between layers 0 and 1, there are 1x200 bias values.

Deep Ranking Formulation and Backpropagation

Our algorithm takes the set $\mathcal{I}_{train}$ and learns a deep neural network that maps the voxels of a 3D shape $\mathbf{x}$ to the shape's aesthetics score $y = h_{\mathbf{W},\mathbf{b}}(\mathbf{x})$ (Figure 5.2.3). We follow the deep ranking formulation in [61], but there are many subtle and important differences including: the 3D shape and voxel representation, the formulation for the 3D convolutional architecture, and the two copies of the neural network for the A and B cases.

While supervised learning frameworks have the target values y in the training data, we do not directly have such target values. Our data is ranking-based and provides rankings of pairs of 3D shapes. This is the motivation for taking a learning-to-rank formulation and we learn $\mathbf{W}$ and $\mathbf{b}$ to minimise this ranking loss function:

$$\mathcal{L}(\mathbf{W},\mathbf{b}) = \frac{1}{2}\|\mathbf{W}\|_2^2 + \frac{C_p}{|\mathcal{I}_{train}|}\sum_{(\mathbf{x}_A,\mathbf{x}_B)\in\mathcal{I}_{train}} l(y_A - y_B) \tag{5.1}$$

where $\|\mathbf{W}\|_2^2$ is the L^2 regularizer (2-norm for matrix) to prevent over-fitting, C_p is a parameter, $|\mathcal{I}_{train}|$ is the number of elements in $\mathcal{I}_{train}$, $l(t) = \max(0, 1-t)^2$ is a suitable loss function for the inequality constraints, and $y_A = h_{\mathbf{W},\mathbf{b}}(\mathbf{x}_A)$.

The training set $\mathcal{I}_{train}$ contains inequality constraints. If $(\mathbf{x}_A, \mathbf{x}_B) \in \mathcal{I}_{train}$, our neural network should give a higher aesthetics score for shape A than for shape B (i.e. $h(\mathbf{x}_A)$ should be greater than $h(\mathbf{x}_B)$). The loss function $l(t)$ enforces prescribed inequalities in $\mathcal{I}_{train}$ with a standard margin of 1.

To minimise $\mathcal{L}(\mathbf{W},\mathbf{b})$, we perform an end-to-end neural network backpropagation with gradient descent. First, we have a forward propagation step that takes each pair $(\mathbf{x}_A, \mathbf{x}_B) \in \mathcal{I}_{train}$ and propagates $\mathbf{x}_A$ and $\mathbf{x}_B$ through the network with the current $(\mathbf{W},\mathbf{b})$ to get y_A and y_B respectively. Hence there are two copies of the network for each of the A and B cases. Note that in some cases there are multiple sets of weights and biases between layers and the forward propagation proceeds as usual between each set of corresponding nodes. In the convolutional layer, the same weights and biases in each 3D convolutional mask are forward propagated multiple times.

We then perform a backward propagation step for each of the two copies of the network and compute these delta (δ) values:

$$\delta^{(n_l)} = 1 - y^2 \qquad \text{for output layer} \tag{5.2}$$

$$\delta_i^{(l)} = \Big(\sum_{k=1}^{s_{l+1}} \delta_k^{(l+1)} w_{ki}^{(l+1)}\Big)\ \big(1 - (a_i^{(l)})^2\big) \qquad \text{for inner layers} \tag{5.3}$$

where the δ and y values are indexed as δ_{Ai} and y_A in the case for A. The index i in δ is the neuron in the corresponding layer and there is only one node in our output layers. n_l is the number of layers, s_{l+1} is the number of neurons in layer $l+1$, $w_{ki}^{(l+1)}$ is the weight for the connection between neuron i in layer l and neuron k in layer $(l+1)$, and $a_i^{(l)}$ is the output after the activation function for neuron i in layer l. We use the *tanh* activation function which leads to these δ formulas. Because of the learning-to-rank aspect, we define these δ to be different from the usual δ in the standard neural network back-propagation. Note that in some cases there are multiple sets of weights and biases between layers and the backward propagation proceeds as usual between each set of corresponding nodes. The backward propagation computes these δ values from the last layer up to layer 1.

We now compute the partial derivatives for the gradient descent. For $\frac{\partial \mathcal{L}}{\partial w_{ij}^{(l)}}$, we split this into a $\frac{\partial \mathcal{L}}{\partial \|\mathbf{W}\|_2} \frac{\partial \|\mathbf{W}\|_2}{\partial w_{ij}^{(l)}}$ term and $\frac{\partial \mathcal{L}}{\partial y} \frac{\partial y}{\partial w_{ij}^{(l)}}$ terms (a term for each y_A and each y_B computed from each $(\mathbf{x}_A, \mathbf{x}_B)$ pair). The $\frac{\partial \mathcal{L}}{\partial y} \frac{\partial y}{\partial w_{ij}^{(l)}}$ term is expanded for the A case for example to $\frac{\partial \mathcal{L}}{\partial y_A} \frac{\partial y_A}{\partial a_i} \frac{\partial a_i}{\partial z_i} \frac{\partial z_i}{\partial w_{ij}^{(l)}}$ where the last three partial derivatives are computed with the copy of the network for the A case. z_i is the value of a neuron before the activation function.

The entire partial derivative is:

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial w_{ij}^{(l)}} &= w_{ij}^{(l)} \\ &+ C \sum_{(A,B)} max(0, 1 - y_A + y_B)\, chk(y_A - y_B)\, \delta_{Ai}^{(l)} a_{Aj}^{(l-1)} \\ &- C \sum_{(A,B)} max(0, 1 - y_A + y_B)\, chk(y_A - y_B)\, \delta_{Bi}^{(l)} a_{Bj}^{(l-1)}\end{aligned} \tag{5.4}$$

where $C = \frac{2C_p}{|\mathcal{I}_{train}|}$. There is one term for each of the A and B cases. (A, B) represents $(\mathbf{x}_A, \mathbf{x}_B) \in \mathcal{I}_{train}$ and all terms in the summation can be computed with the corresponding $(\mathbf{x}_A, \mathbf{x}_B)$ pair. For the $chk(t)$ function, if $t \geq 1$ then $chk(t) = 0$, and if $t < 1$ then $chk(t) = -1$. For each (A, B) pair, we can check the value of $chk(y_A - y_B)$ before doing the backpropagation. If it is zero, we do not have to perform the backpropagation for that pair as the term in the summation is zero.

For the weights in a convolutional layer, the partial derivative is:

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial w_{ij}^{(l)}} &= w_{ij}^{(l)} \\ &+ C \sum_{(A,B)} \sum_{k} max(0, 1 - y_A + y_B)\, chk(y_A - y_B)\, a_{Ajk}^{(l-1)} \delta_{Aki}^{(l)} \\ &- C \sum_{(A,B)} \sum_{k} max(0, 1 - y_A + y_B)\, chk(y_A - y_B)\, a_{Bjk}^{(l-1)} \delta_{Bki}^{(l)}\end{aligned} \tag{5.5}$$

where k is the number of times each 3D convolutional mask is used. The last part $a_{Ajk}^{(l-1)} \delta_{Aki}^{(l)}$

is reversed from the previous formula, since in the implementation it is easier to consider the computation as a multiplication of the corresponding vectors in this order into a matrix. This computation typically takes the longest time in the whole training process. To speed up the overall process for the first layer, since our activations represent voxels and are only binary values, we can explicitly construct the matrix with this knowledge rather than performing the actual matrix multiplications.

The partial derivative for the biases is:

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial b_i^{(l)}} = & C \sum_{(A,B)} max(0, 1 - y_A + y_B)\, chk(y_A - y_B)\, \delta_{Ai}^{(l)} \\ & - C \sum_{(A,B)} max(0, 1 - y_A + y_B)\, chk(y_A - y_B)\, \delta_{Bi}^{(l)}\end{aligned} \tag{5.6}$$

For the biases in a convolutional layer, the partial derivative is:

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial b_i^{(l)}} = & C \sum_{(A,B)} \sum_k max(0, 1 - y_A + y_B)\, chk(y_A - y_B)\, \delta_{Aki}^{(l)} \\ & - C \sum_{(A,B)} \sum_k max(0, 1 - y_A + y_B)\, chk(y_A - y_B)\, \delta_{Bki}^{(l)}\end{aligned} \tag{5.7}$$

The gradient descent (Algorithm 2) starts by initialising $\mathbf{W}$ and $\mathbf{b}$ randomly. We then go through the training data for a fixed number of iterations (epoch), where each iteration involves taking a set of data pairs and performing the forward and backward propagation steps and computing the partial derivatives. Each iteration of gradient descent sums (in $\Delta W^{(l)} and \Delta b^{(l)}$) the partial derivatives from a set of data pairs and updates $\mathbf{W}$ and $\mathbf{b}$ with a learning rate a.

Learned Aesthetics Measure

After gradient descent learns $\mathbf{W}$ and $\mathbf{b}$, we can use them to compute an aesthetics score for a 3D shape. For a new shape of the corresponding category, we voxelize it into $\mathbf{x}$ and use one copy of the neural network and a forward propagation pass to get the score $h_{\mathbf{W},\mathbf{b}}(\mathbf{x})$. This score is an absolute value, but since the data and method are ranking-based, it has more meaning in a relative sense when the score of a shape is compared to that of another.

Validation Data Sets

We use a validation dataset to set the parameters of the neural network. For each category, we keep about 5 to 10% of the collected data as a separate validation set $\mathcal{I}_{validation}$ which has the

Algorithm 2 Deep Ranking Training Algorithm

procedure TRAIN DEEP NETWORK
  $W^{(l)} \leftarrow RandomWeights$, $b^{(l)} \leftarrow RandomBiases$
  for $i \leftarrow 1 : epoch$ **do**
    $\Delta W^{(l)} \leftarrow 0$, $\Delta b^{(l)} \leftarrow 0$
    for $eachpair(x_A, x_B) in I_{train}$ **do**
      Feed-forward (x_A, x_B) to get $Out_A and Out_B$ as network outputs
      if $((Out_A - Out_B) < 1)$ **then**
        Back-propagate to compute partial derivatives $\frac{\partial \mathcal{L}}{\partial w^{(l)}} and \frac{\partial \mathcal{L}}{\partial b^{(l)}}$
        $\Delta W^{(l)} \leftarrow \Delta W^{(l)} + \frac{\partial \mathcal{L}}{\partial w_{ij}^{(l)}}$
        $\Delta b^{(l)} \leftarrow \Delta b^{(l)} + \frac{\partial \mathcal{L}}{\partial b_i^{(l)}}$
      end if
    end for
    $W^{(l)} \leftarrow W^{(l)} + a\Delta W^{(l)}$
    $b^{(l)} \leftarrow b^{(l)} + a\Delta b^{(l)}$
  end for
end procedure

same format as the training data $\mathcal{I}_{train}$. We use only those pairs where there is a difference of three or more between the number of users who select 'A' as more aesthetic and the number of users who select 'B' as more aesthetic. For each pair of shapes in $\mathcal{I}_{validation}$, the prediction from the measure learned with $\mathcal{I}_{train}$ is correct if the collected data says shape A is more aesthetic than shape B and our score of shape A is greater than that of B. To select the parameter for a, for example, we can let a be $\{1e^{-1}, 1e^{-2}, 1e^{-3}, 1e^{-4}, 1e^{-5}\}$. The selected a is the one that minimises the validation error. There is typically a wide range of parameters that works well.

NEURAL NETWORK PARAMETERS

In each iteration of the gradient descent, we use all the data samples or sampled pairs in $\mathcal{I}_{train}$. We typically perform 10 iterations of all samples. The weights **W** and biases **b** are initialised by sampling from a normal distribution with mean 0 and standard deviation 0.1. The parameter C_p is set to 100 and the learning rate a is set to 0.0001. The learning process is done offline and it can take up to one hour of execution time in MATLAB to perform 10 iterations of gradient descent for 1000 data samples. After the weights and biases have been learned, computing the score for a shape is interactive as this only requires straightforward forward propagation.

5.3 RESULTS

We demonstrate our learned aesthetics measure by showing the rankings of a large number of various classes of 3D shapes based on their aesthetics scores. Our aesthetics measure is learned from crowdsourced data and contains the collective preferences of many people. A single score

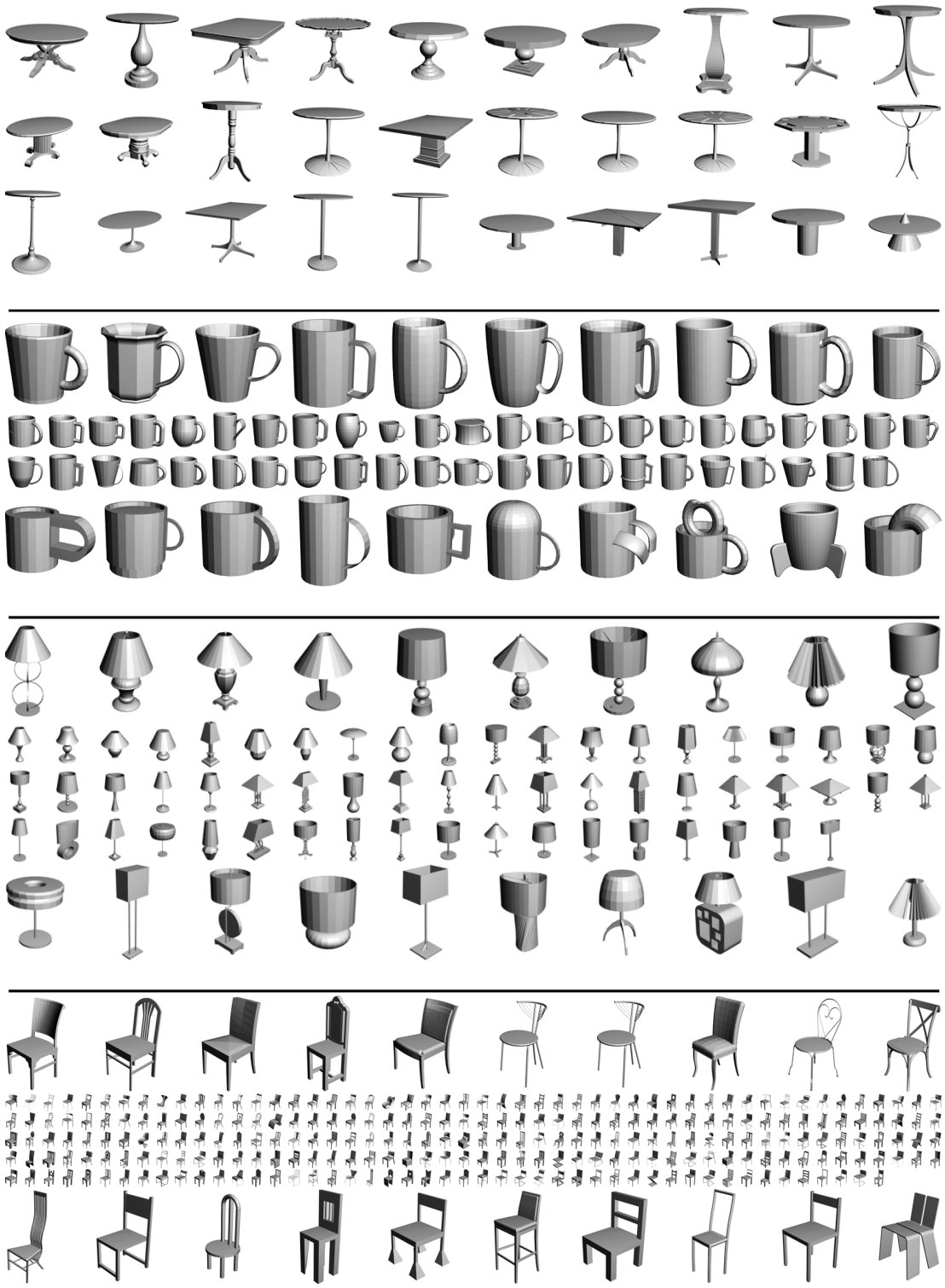


Figure 5.2.4: Shapes ranked (from top to bottom and left to right in each row) according to our aesthetics measure. There are 30 pedestal tables, 65 mugs, 78 lamps, 267 dining chairs, and 30 abstract shapes.

for one shape is not meaningful, while the scores for multiple shapes can be compared against each other to give information about their relative aesthetics.

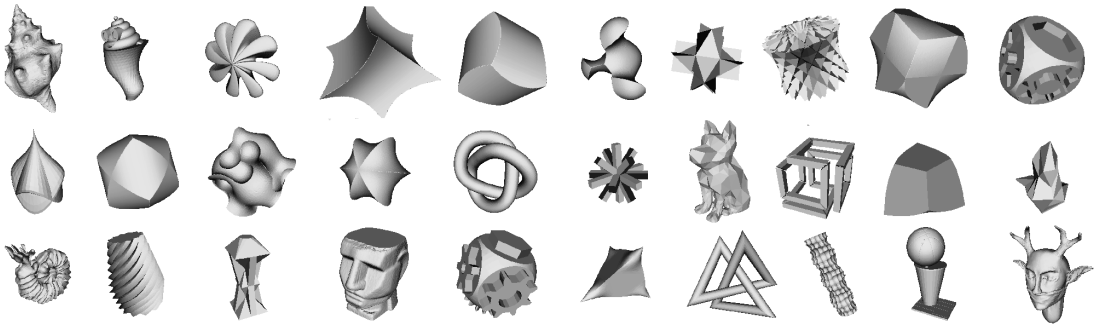


Figure 5.2.5: Abstract shapes (30) ranked (from top to bottom and left to right in each row) according to our aesthetics measure.

Figure 5.2.6: Test sets of shapes ranked (from high to low scores) by our aesthetics measures. There are 4 classes of 10 shapes each: pedestal tables, dining chairs, mugs, and lamps. The last 2 shapes in each row are intentionally created to be ugly shapes and have the lowest scores. The ugly shapes are also used as part of the control questions in the data collection process and are not included in the training data.

5.3.1 Qualitative Patterns in Results

Figure 5.1.2 shows the results for a large set of club chairs. We can observe some clear patterns in these results. The highest ranked chair models tend to have more curved surfaces and/or tall (but not too tall) backs. The lowest ranked models tend to have more planar surfaces and the lowest few in particular are somewhat ugly. All of these aspects are learned autonomously from the human aesthetics data and no geometric features that for example correspond to curved, round, or planar surfaces are specified.

There are a few examples of models in the club chair dataset that are the same and they are ranked beside each other. For the fourth chair in the first row of Figure 5.1.2, for example, there are a few other chairs that are similar but slightly different. While they are not ranked immediately beside each other, all of them are still ranked near the top. For all the classes of models in general, models that are very similar tend to be ranked near each other. A small change to a 3D shape tends to result in a small change to its ranking and this shows that our algorithm is robust.

Figure 5.2.4 shows the aesthetics rankings for four classes of shapes. We describe the patterns that we can observe in the images. For the pedestal tables, the top few models have fancy

and/or rounded table legs. The middle row has four similar rounded tables near each other. In the last row, there are a few taller tables followed by the most ugly tables at the end. For the mugs, the top models tend to be tall (but not too tall) and not wide. The handle shapes typically match with the corresponding body shapes of the mugs: they are not too thin and neither too big nor too small. Many mugs in the middle are similar, with subtle differences in their shapes and handles in contrast to the top ones. The bottom ones tend to be the opposite: taller, shorter, wider, and/or with a handle that is thick or thin and too large or more rectangular. The last five are relatively ugly and the upside down one is ranked low (there just happened to be an upside down mug in the downloaded dataset). For the lamps, the top ones tend to be rounded in some way and have some spherical or circular shapes. The bottom ones are wider, taller, planar, and/or not symmetric. The last lamp appears to be broken due to the separated parts of the model. For the dining chairs, the top models tend to have nice proportions, somewhat curved backs, and/or some nice patterns on the backs. The bottom models tend to have taller or simpler backs, and/or planar surfaces. In case of abstract shapes (Figure 5.2.5), although the differences are very subtle, we can see the presence of smooth surfaces among the top ranked shapes. Further, the sharpness of curves or edges contributes to less aesthetic shapes as exhibited by low ranking shapes.

The results for all classes typically show a distribution where there are some particularly aesthetic shapes, some particularly ugly shapes, and many shapes that are in between. If we take two shapes that are relatively far apart in their computed scores, they typically look quite different in their aesthetics. If we take two shapes that are relatively close in their scores, it can sometimes be difficult to say which one is more aesthetic. There can be cases where given the same two shapes to a group of people, half of them may prefer one shape while the other half may prefer the other shape. We show examples of these cases and analyse them in the evaluation section.

Test Data Sets

In addition to ranking many existing shapes for each class, we can take separate testing sets of 3D models and rank them. Figure 5.2.6 shows the rankings of ten such models for four classes and these results show similar patterns as in the rankings above. There are then two ugly shapes for each class that we intentionally made and these have the lowest aesthetics scores. The results here also provide a good way to test that these models can be considered ugly for our control questions in the data collection process.

Category	15^3	30^3	60^3	128^3	128^3*
Abstract	74.10	78.48	81.32	81.32	79.32
Club Chairs	79.38	79.38	76.29	77.32	76.29
Pedestal Tables	78.30	84.80	82.18	82.80	80.48
Mugs	80.08	79.78	81.29	81.29	80.78
Lamps	76.54	79.01	85.19	85.19	81.48
Dining Chairs	81.82	83.84	84.85	83.43	82.84

Table 5.3.1: Comparison of Network Architectures and Voxel Resolutions. The last column is for convolutional network with 128^3 voxel resolution, while rest are for fully-connected networks with resolution of 15^3, 30^3, and 60^3, in first, second, and third columns respectively. The percentages are the percent of samples that are correctly predicted.

5.3.2 Quantitative Evaluation

We perform various types of evaluation to gain a better understanding of our method. Throughout this section, we consider the validation datasets $\mathcal{I}_{validation}$ as 'correct' or ground truth data and use them to evaluate the accuracy of the learned measure. For each data sample in $\mathcal{I}_{validation}$, the learned measure is correct if the collected data says shape A is more aesthetic than shape B and our score of shape A is greater than that of B.

Comparison of Network Architectures

We show a comparison of different network architectures and voxel resolutions (Table 5.3.1). These architectures are non-linear functions that can already represent complex relations from raw voxel data to an aesthetics score, and hence we do not compare them with linear functions. We start with a 15^3 voxel resolution for a fully-connected architecture (Figure 5.2.3a). This resolution is relatively low but still gives a reasonable representation of the 3D shapes. For this case, there are 200 nodes in the first and second layers and 50 nodes in the third layer. The next resolutions are 30^3, 60^3, 128^3 with the same architecture. We also use a convolutional network (Figure 5.2.3b) for 128^3 resolution. For this case, the 3D convolution mask or filter in layer 0 is of size 18 (with a stride of 10) and layer 1 is a cube (of nodes) of size 12.

Quantity of Training Data

The effectiveness of the learned aesthetics measure depends on the quantity of data. We show the percentage accuracy on $\mathcal{I}_{validation}$ as the amount of training data increases for each class of shapes (Figure 5.3.1). In each case, we train with the number of data samples for 10 iterations of gradient descent and compute the accuracy with the full $\mathcal{I}_{validation}$ set. For each class, the voxel resolution (we take the best in each class from the previous subsection even if it is better by only a small amount) and amount of data (four units in the x-axis of the graph) are different. The voxel resolutions for club chairs, pedestal tables, mugs, lamps, and dining chairs are 15,

Figure 5.3.1: Plots of percent accuracy on $\mathcal{I}_{validation}$ versus the amount of data samples in $\mathcal{I}_{train}$ for five classes of shapes. We show this to highlight the relationship between amount of training data and prediction accuracy.

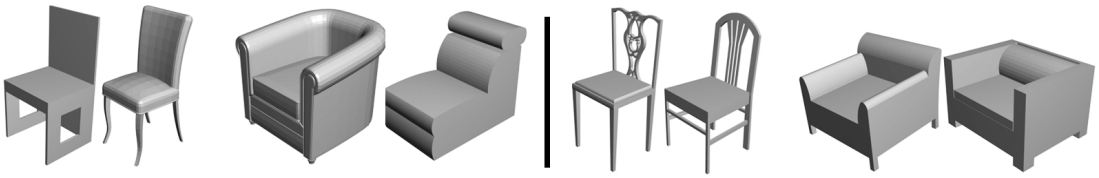


Figure 5.3.2: First two example pairs where all ten Turkers chose the same shape (right dining chair and left club chair) as being more aesthetic. Next two example pairs where five chose one shape and five chose the other.

15, 60, 45, and 15 respectively. Each unit of data is 1900, 640, 180, 550, and 1150 samples respectively. In the graph, the main idea is to show that the plots exhibit decreasing returns. As the amount of data keeps increasing, the percentage increases but this increase will slow down. Observing these plots provides one empirical way of knowing whether we have enough training data.

Failure and Limitation Cases

In the data samples, there are some samples where the pairs of shapes have consistent responses. Given the same pair of shapes to different people, they will choose the same response (see examples in Figure 5.3.2). In these cases, our aesthetics measure works well. On the other hand, there are some pairs of shapes that can be very close in their aesthetics. Given the same pair of shapes to different people, half of them will choose one shape while half will choose the other (Figure 5.3.2). In these cases, our aesthetics measure fails since we will get an accuracy of 50% regardless of what we predict, and a random measure will also get this accuracy.

We show a numerical analysis of these cases (Figure 5.3.3). The idea is to have some samples of pairs of shapes where we collect the aesthetics preferences from multiple people. We then separate the samples into groups based on the multiple responses. For example, a sample pair

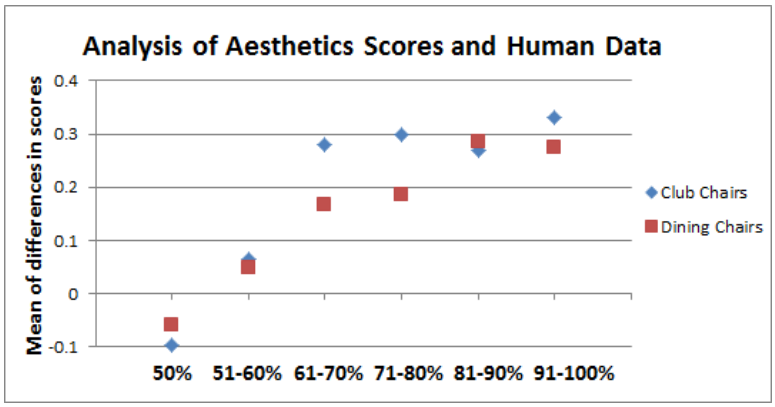


Figure 5.3.3: We post 5 HITs and have 10 Turkers provide responses to each HIT. For some HIT tasks, all ten gave the same response (A or B), and these are placed into the 91-100% group. There are some tasks where five chose A and five chose B, and these are placed into the 50% group. For each data sample, we use our learned measure to compute the difference in aesthetics scores. If A is the more common response, we take the score of shape A minus that of shape B. We plot the mean of these differences for each group.

(A,B) given to 10 people with responses (9,1) (i.e. 9 choose A and 1 choose B), (1,9), or (1,8) goes into the 81-90% group [2]. Note that some Turkers are rejected so we may not have 10 responses for each sample pair. We wish to compute the difference in the aesthetics scores from our learned measure for each sample pair of shapes. If the user chooses A, the difference is the score of A minus the score of B.

Figure 5.3.3 shows the results for club chairs and dining chairs. We observe an increasing trend in the mean of differences of scores and this trend matches with our intuition. For the 50% group, the two shapes in each pair tend to be similar in aesthetics and the difference in their scores tend to be smaller. If these types of pairs are in $\mathcal{I}_{validation}$, it may not be useful to consider them. For the 91-100% group, the two shapes in each pair tend to have a clear difference in aesthetics and the difference in their scores tend to be larger. These are also cases that the learned measure can predict well.

5.3.3 What makes a 3D shape aesthetic?

In this section, we use different qualitative and quantitative shape attributes to predict aesthetics scores with. We compute quantitative attributes such as: bounding box volume, area, intrinsic volume, mean curvature, Gaussian curvature, D2 shape distribution, 3D histogram of gradients, shape diameter function, light field descriptor, and show that some of these correlates positively with aesthetics scores. We also measure correlation with a qualitative attribute related to shape functionality or ergonomics which we call as 'functional aesthetics'. Although,

[2] Please note that before learning we order (A, B) based on which of A or B is preferred by the participant.

Category	GCUR	MCUR	HOG	D2	SDF	LFD	VOX
Abstract	80.00%	70.00%	70.00%	40.00%	50.00%	50.00%	81.32%
Club Chairs	63.64%	54.55%	54.55%	56.25%	59.09%	72.73%	79.38%
Tables	80.00%	80.00%	70.00%	60.00%	50.00%	70.00%	82.80%
Mugs	57.14%	42.86%	28.57%	71.43%	28.57%	78.57%	81.29%
Lamps	70.00%	70.00%	50.00%	60.00%	40.00%	70.00%	85.19%
Dining Chairs	75.00%	75.00%	55.00%	55.00%	50.00%	70.00%	84.85%

Table 5.3.2: Prediction accuracy with different shape features and voxels (VOX in last columns). **GCUR** is Gaussian Curvature, **MCUR** is Mean Curvature, **HOG** is 3D Histogram of Voxel Gradients, **D2** is D2 Shape Distribution, **SDF** is Shape Diameter Function, **LFD** is Light Field Descriptor Curvature, The percentages are the percent of samples that are correctly predicted.

the main idea is to study the link between individual feature and shape aesthetics, we notice that learning with a combination of these features is less effective than input voxel representation.

Simple Features

Are simple 3D shape features enough to decide if a shape will be perceive as more or less aesthetic? We computed three simple shape features: bounding box volume (BBV), surface area (SA), and intrinsic volume (INV) and plot (Figure 5.3.4) these against the sorted aesthetics scores. While bounding box volume gives the volume of the smallest box that encloses the given shape, intrinsic volume is related to the inner volume occupied by the shape and gives an estimate of how thick or thin a shape is perceived. We found that these feature alone do not have any correlation with shape aesthetics, thus more descriptive characteristics are needed.

Curvature

As discussed before, curvature alone has been established to contribute a lot to the beauty of a form. In order to understand this, we compute 256 bin histograms of Gaussian curvature, mean curvatures, and voxel gradient on all shapes and train two layer ranking neural net to see if these descriptors alone can predict aesthetics, if yes to what accuracy (Table. 5.3.2). We compute curvature on uniformly sampled 10,000 surface points and compute voxel gradients along x, y, and z direction on a voxelized volume of 128^3. The choice of these descriptors is motivated by the observation that pair (s) where structure is similar but more curved shape (even with minor difference in curvature) is selected by majority (80% or more). For example, for the first pair of shapes in Figure 5.3.6), the left shape is the majority vote (90%).

We found that both Gaussian and mean curvature descriptors provides good prediction 5.3.5. These results concord with research in psychology and other fields establishing curva-

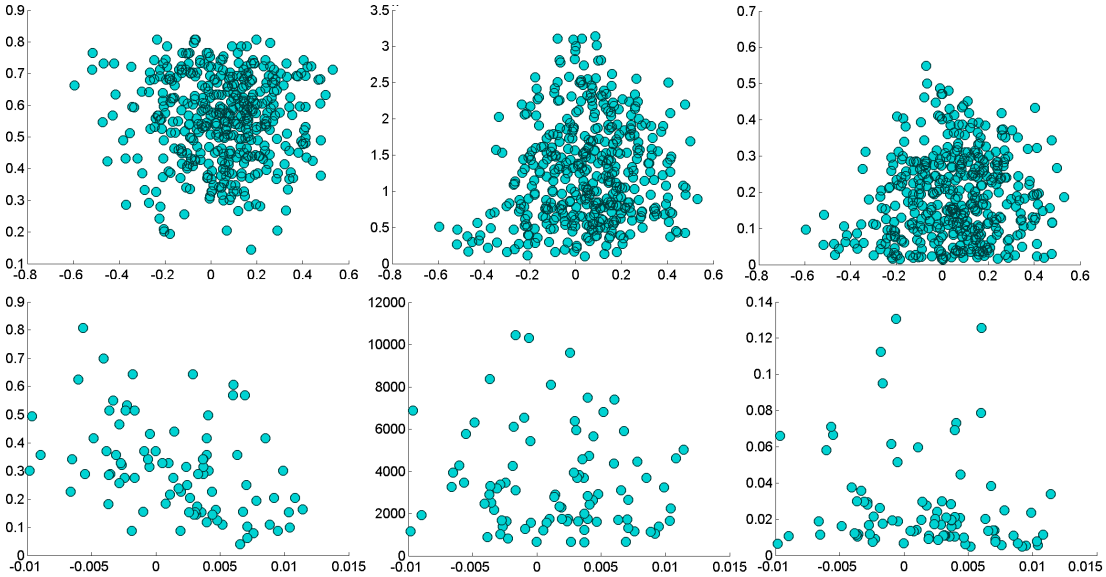


Figure 5.3.4: Aesthetics and simple 3D shape features. For all features and plots, we first sort the aesthetic scores and plot along x-axis. First row plots are for club chairs bounding-box volume, intrinsic volume, surface area respectively. The next row shows the plots for table lamps. Plots for the rest of the categories are in the supplementary material.

ture as an important feature in perceptual shape aesthetics.

Looking at the first two columns suggests slightly different prediction accuracy for different shape categories. For example abstract shapes category has a prediction accuracy of 80% while mugs have an accuracy of 57.14%. This may be due to the reason that abstract shapes exhibit more variation in curvature than the coffee mugs. Also, we note that both Gaussian and mean curvatures are equally good at predicting aesthetics. Another descriptor 3D HOG related to shape curvature doesn't predict the aesthetics very well.

Structure

We observed that participants on a majority prefer some structures as more aesthetic than the others. For example, the first shape in last pair in Figure 5.3.6 is the majority vote (90%). Motivated by these observations, we study the role several structural attributes such as 'shape diameter', 'D2 shape distribution', and 'light field descriptor' in prediction of aesthetics of a 3D shapes (Figure 5.3.5). The first two descriptors are computed on uniformly sampled 10,000 shape surface samples, while the third one is generated from 100 views of the shape.

As given in the Table. 5.3.2, the shape diameter function doesn't provide a good prediction, while the other two descriptors seem to have a better contribution in prediction, especially the light field descriptor.

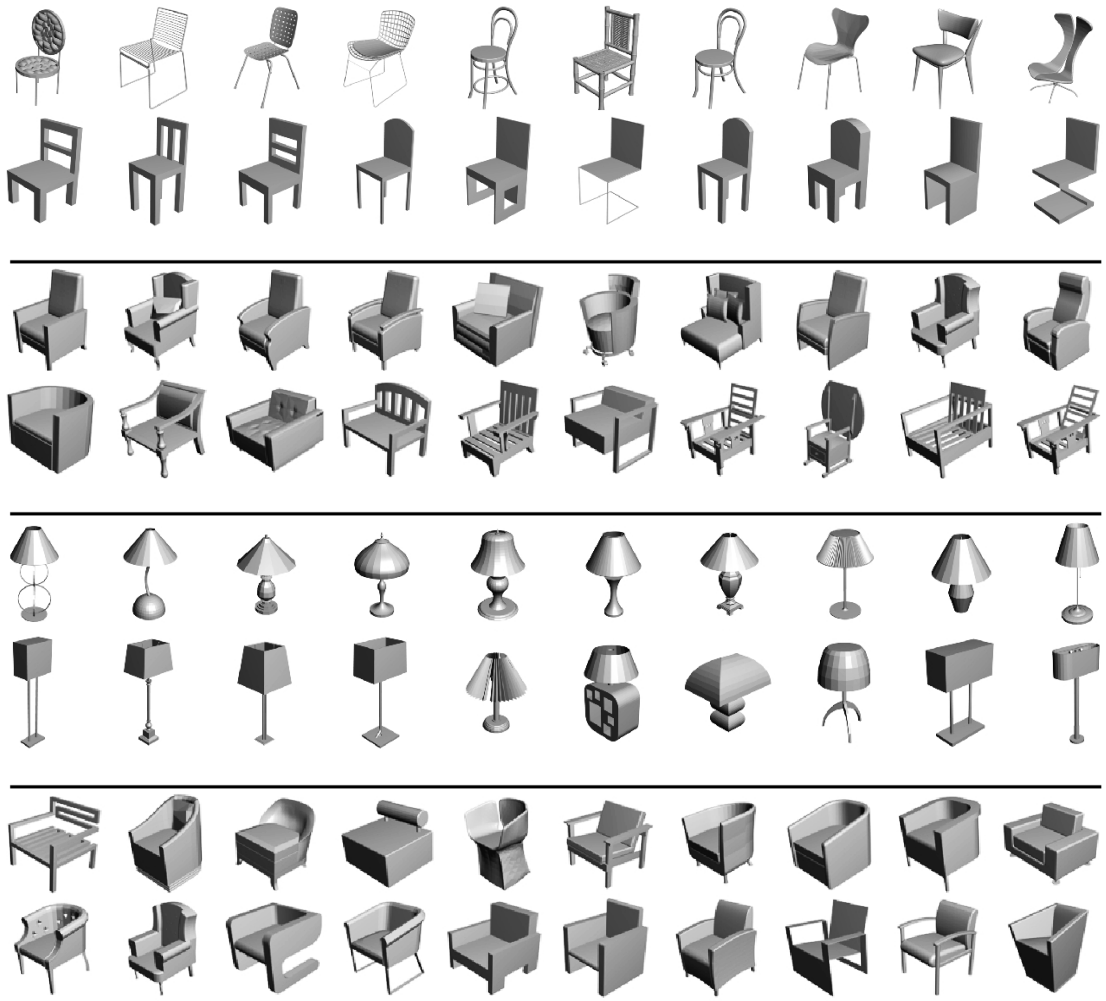


Figure 5.3.5: Ranking of shapes based on aesthetic scores learned on different shape descriptors. First two rows show top and bottom ten aesthetic dining chairs using Gaussian curvature descriptor respectively. Similarly, in the next two rows we have results for club chairs using d2 shape descriptor, followed by results for lamps using light field shape descriptor, and results for club chairs using shape diameter function.

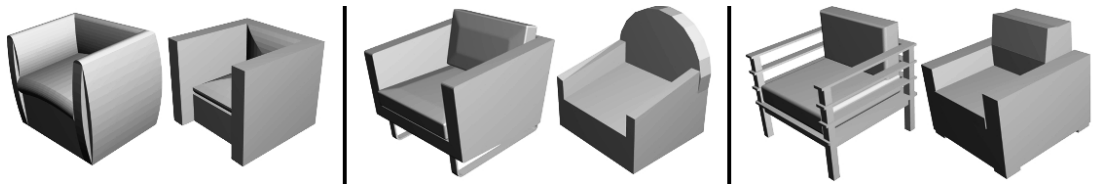


Figure 5.3.6: Some example pairs where left shape in each pair is selected as more aesthetic by more than 90% participants. First pair shows similar structures however the more curved one (left) is the majority vote. Second pair shows similarity in curvature, however more functionally aesthetic shape (left) is the majority choice. Third example shows, structural difference guiding users to select more aesthetic shape (left).

5.3.4 Aesthetics Duality

As Steve Jobs observed, "Design is not just what it looks like and feels like. Design is how it works". In our daily lives, we use objects that have both form and functionality. Our aesthetic

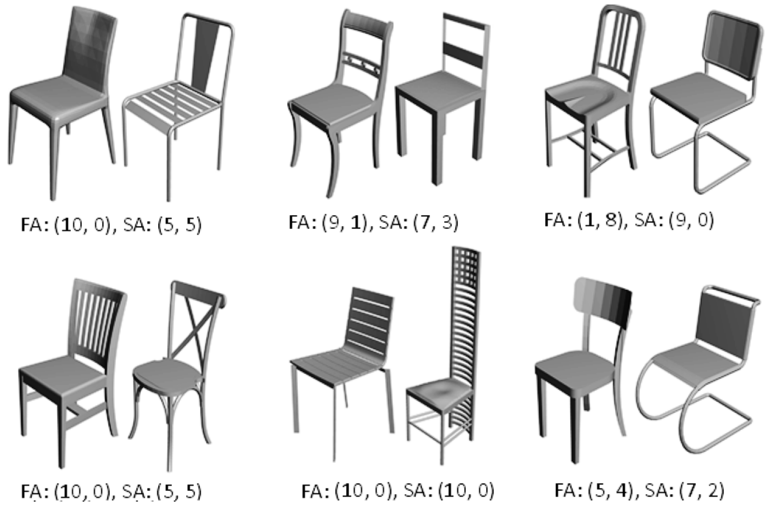


Figure 5.3.7: Example showing differences and similarities in majority vote for functional and shape aesthetics responses. In the first column, both the pairs have clear majority in functional aesthetics responses while this is not true for shape aesthetics responses. In the second column, for both the pairs, participants agree on both functional and shape aesthetics. In the last column, the first pairs have opposite majority vote, while the second pair has no clear majority on functional aesthetics however participants clearly agree on more aesthetic shape. *FA-Functional Aesthetics, *SA-Shape Aesthetics, (x,y) means x and y number of participants choose first shape and second shapes as more functionally aesthetic.

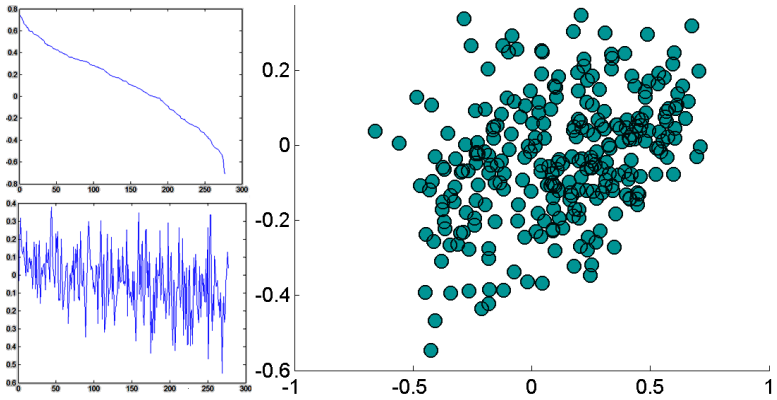


Figure 5.3.8: Correlating shape and functional aesthetics scores. Two line plots on the left: (top) sorted shape aesthetics scores plot and (bottom) functional aesthetics scores plot of the same shape order. Scatter plot on the right showing sorted shape aesthetic scores along x-axis and functional aesthetics scores along y-axis. Clearly, these two can not be correlated.

judgements about such objects may be influenced by either the form or the functionality. We refer to the degree to which a shape or object serves its function as 'functional aesthetics'. For example, a chair when viewed gives perceptual clues about how comfortable to sit on, or ergonomic, or functionally aesthetic it will be when used for sitting on, second pair of shapes in Figure 5.3.6. We conduct a study on MTurk to study the influence of perceived functional

Figure 5.3.9: Comparing functional and shape aesthetic predictions. First row shows top 5 and bottom 5 functionally aesthetic shapes respectively, as predicted by ranking network trained on functional aesthetic responses. Similarly, the next rows show the results of shape aesthetics predictions using the network trained on shape aesthetic responses.

aesthetics on shape aesthetic judgements. Specifically, we do the following:

First, we collect perceptual functional aesthetics response data on MTurk by showing pairs of chairs in a way similar to normal shape aesthetic data collection (Section 5.2). We ask participants to choose which shape they think is functionally more aesthetic, or comfortable, or ergonomic to use. We used 277 dining chairs for this purpose and collected data from 120 participants (4 rejections based on incorrect control question answers) with a total of 4050 responses.

Second, we compare the collected functional aesthetics data with shape aesthetics data in following ways: First, we counted the percentage of responses that users agree (or disagree) on a majority (70% or more) in a pair for both functional and shape aesthetics. We found that 64.8% (84/125) participants agree on a majority on functional aesthetic responses, while this figure is 35.2% for the shape aesthetic responses. This suggests that it is relatively easier for people to agree on a majority on functional aesthetic responses than on shape aesthetic responses. Additionally, we found that out of 84 pairs in functional aesthetic responses, 28 (33.33%) pairs have similar majority vote in shape aesthetic responses. This figure has a value of 63.63% (28/44) on shape aesthetic to functional aesthetic responses. This suggests that while responding to functional aesthetic questions participants focus more on function than form, however this is not true in the opposite direction. Second, we trained a ranking network on functional aesthetic responses to rank the shape collection and compare with the rankings received from shape aesthetics responses. We found 0.18 Spearman's rank correlation and predicted scores are plotted in Figure 5.3.8, also rankings of top 5 and bottom 5 models is shown in Figure 5.3.9. Clearly, in Figure 5.3.9 the shapes in the first row and the shapes in the third row have several differences. For example, the shapes in the first row looks more ergonomic while those in the third row depicts more variety and aesthetics.

5.3.5 Applications

The learned aesthetics measure can be used for various applications. We demonstrate the applications of aesthetics-based visualisation, search, and scene composition.

Aesthetics-based Visualisation

The idea is to visualise a large dataset of 3D shapes in one image based on aesthetics. First, the aesthetics scores can affect the size of each shape icon in the overall image. We take the aesthetics scores to scale the size of the shape icons. This helps to create a more aesthetics overall image (as the aesthetics shapes are larger) and makes it easier to view the most aesthetic shapes.

Second, the aesthetics measure can affect the 2D positions of the shape icons within the overall image. We can take a voxel resolution of 15 and use t-SNE [77] to map the raw voxel data to two dimensions. The input to t-SNE is 1D vectorised representation from 3D voxel grid and output is the position (x, y coordinates)in 2D space. Since this is based on the raw voxels, aesthetics is not considered in this case. If we increase the voxel resolution to 30, t-SNE typically does not work well and the shape icons become mostly laid out with uniform spacing in the overall image. In this case, we take the activation values in the neurons of an inner layer of the neural network (which can be considered as a dimension reduction based on aesthetics) and use t-SNE to map these values to two dimensions. This works well and in this case the aesthetics features learned in the inner layers can influence the 2D positions.

Figure 5.3.10 shows some examples of visualisations created this way. We can observe some patterns in these images. For dining chairs, a voxel resolution of 15 works well as input to t-SNE. There are many regions of similar shapes grouped together. Chairs with taller backs tend to be near the top, and there is a group of taller back and aesthetic chairs near the top middle of the overall image. For mugs, taking a voxel resolution of 15 and the 200 activation values of the first layer of the architecture in Figure 5.2.3a works well. There are aesthetic mugs around the middle column part of the image while the ugly mugs are very small. The shorter mugs tend to be near the top and the taller mugs tend to be near the bottom. The mugs on the right side of the image tend to have a body shape that is more vertical or cylindrical. For club chairs, taking a voxel resolution of 60 and the activation values of a later layer (e.g. fourth to sixth layer) of the architecture in Figure 5.2.3b works well. The more aesthetic club chairs tend to be on the right side of the image. Some chairs with tall backs and curved surfaces are near the bottom right while some chairs with curved arms are near the top right of the image.

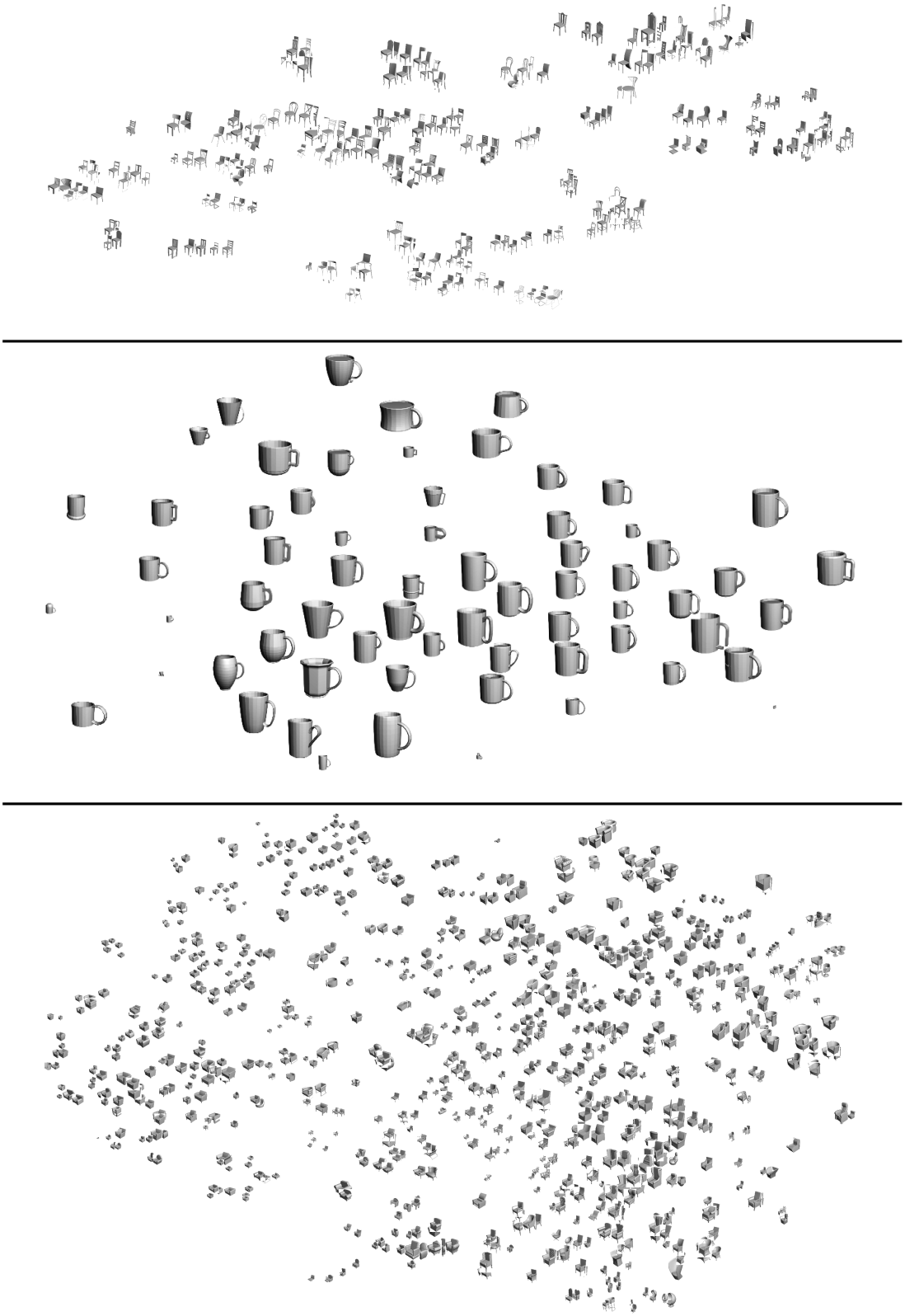


Figure 5.3.10: Aesthetics-based Visualisation, where size of each shape depends on its aesthetic score and its 2D position, showing regions of shapes similar in both geometry and aesthetics.

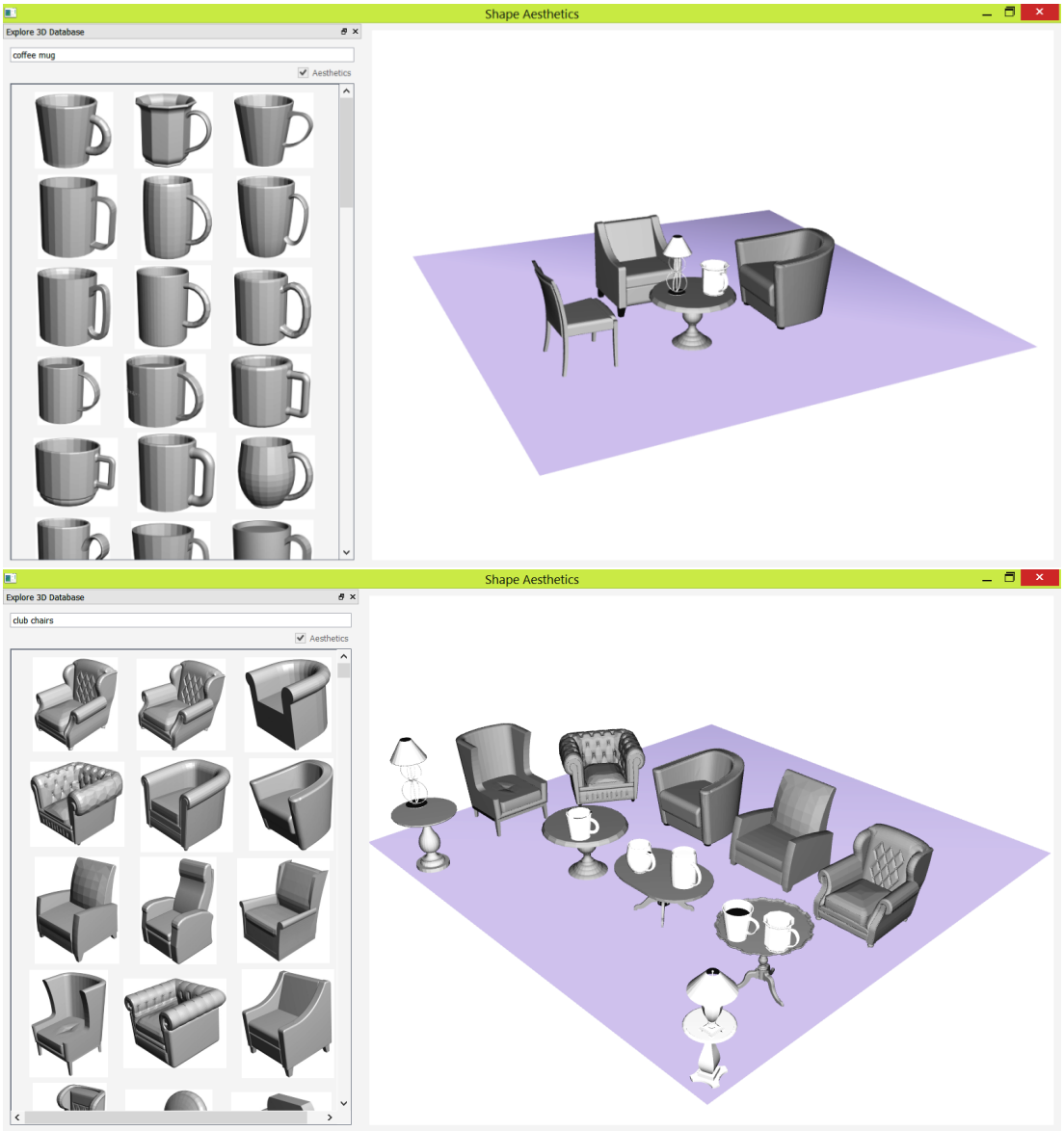


Figure 5.3.11: Aesthetics-based Search and Scene Composition. Our search tool displays each class of 3D shapes in the left panel and they can be ranked according to our aesthetics scores. We can use the tool to compose 3D scenes (two examples in image).

Aesthetics-based Search and Scene Composition

We built a search tool where we can rank each class of shapes according to the aesthetics scores, allowing us to to browse through and choose from the most aesthetic shapes that are at the top. Figure 5.3.11 shows some example screenshots of our search tool. We can use our search tool for 3D scene composition. The idea is that the tool makes it easier to compose more aesthetic shapes together if there are a large number of shapes in each class. Figure 5.3.11 shows some example scenes made with our tool.

5.4 Discussion

To our knowledge, this is the first work to build a data-driven model of 3D shape aesthetics by utilising the concept of perceptual aesthetics, crowdsourcing, and deep ranking. The way we crowdsource perception data by showing shapes in pairs has strong influence on our deep network design. We maintain two copies of neural network to represent two shapes in pairwise data collection approach. The learned measure of aesthetics is a real value between 1 and -1. This restriction is due to the selection of 'tanh' as the neural network activation function. Further, the aesthetic learned in our approach is relative, and can be used to rank large number of shapes from high to low aesthetics. We believe there are several interesting avenues to extend and improve the techniques presented in this chapter.

5.4.1 Image, shape, and scene aesthetics

Computer vision community has paid a lot of attention to advance the technical understanding of several different perceptual attributes of images, such as aesthetics [137], interestingness [48], and memorability [54]. Their focus has been on characterising the attributes of aesthetic images using different machine learning methods including recent deep learning based techniques. We argue that, since an image is a 2D representation of underlying 3D object (s), developing a measure of image or scene aesthetics directly from 3D shapes comprising it, is more natural way to deal with image aesthetics modelling problem. One approach to understand this is by learning a function that takes as input the set of 3D shapes that comprise a scene, a representation of their arrangement, and viewpoint direction to provide a measure of image aesthetics i.e. aesthetics of the image representing a snapshot of the scene created by 3D objects and their arrangement. Another approach would be to first create scene images with different sets of objects and then tying to relate the data-driven modelled aesthetics measure of scene images with the objects comprising the scenes.

5.4.2 Modelling Functioning of Aesthetic Regions of Human Brain

Our results show that it is possible to build a data-driven model of visual perceptual aesthetics of 3D shapes. This data driven model uses artificial neural networks to model the reasoning followed by a natural neural network i.e. human brain. A number of key questions can be raised from this observation which could be the potential fundamental problems to explore. The first question is what we can say about the reasoning followed by human brain to judge more aesthetic shapes based on the learned function of shape aesthetics. Further, what shape features and representations do human brain use to process aesthetics aspects of 3D shapes and how do these relate to representations used by the artificial neural network's hidden layer. Our work is the preliminary investigation into this direction and if extended has the potential

to through light on functioning of the parts of the human brain that help decide if an object is aesthetic or not.

5.4.3 Neural Network Design

We train separate deep networks for different object categories. This means that for each object category we learn a different function to approximate aesthetics. There are two possibilities in the way we train our neural networks. First, we can build a data-driven model of aesthetics that is object category oblivious. It means that we try to learn a single function to predict shape aesthetics. In our view, although it would be possible to learn such a function, however the prediction accuracy would always be less than a category specific model. Second, we can train a neural network by collecting data by paring shapes belonging to different object categories, such as a chair paired with a table. The learned network can thus be compared with a object category specific network. Our intuition is that owing to structural and semantic difference between different object categories, collecting data for shapes belonging to different categories is difficult and needs consistency analysis before using for leaning.

5.4.4 Conclusion

In this work, we conducted crowdsourcing studies to collect shape aesthetics preferences to build ranking based data-driven model of shape aesthetics. The key advantage of our setting is the use of human aesthetics judgements and deep ranking networks to build a computational model of shape aesthetics, rather than manually defining shape aesthetics and using any pre-computed shape features for learning. We demonstrated the usefulness of our technique by ranking a large number of shapes belonging to different categories on the basis of their perceived visual aesthetics, creating visualisations by emphasising more aesthetics shapes, and interactively building virtual 3D scenes using more aesthetics shapes. In our analysis, we also try to relate computable shape features, such as for curvature and shape structure, to aesthetics scores. We found that curvature features such as Gaussian curvatures or mean curvature alone say a lot about the perceived aesthetics of shapes. This result agrees with the results presented in the perceptual study [6]. We would like to emphasise that the learned measure is not a formal measurement but is based on human perception as the data is collected based on human perception. The learned measure is based on data from many people and there can be cases where one person's perception may not agree with it.

I hate that aesthetic game of the eye and the mind, played by these connoisseurs, these mandarins who "appreciate" beauty. What is beauty, anyway? There's no such thing. I never "appreciate," any more than I "like." I love or I hate.

Pablo Picasso

6 Conclusion

The key motivation for the work presented in this thesis is the need to find, analyse, and re-use shapes present in ever growing online 3D shape repositories, a form of precise "big geometric data" [131]. We focus on perceptual attribute learning, for search, reuse by scene composition, and visualisation. We view "crowdsourcing perceptual attributes for building data-driven models without predefined rules" as the central theme of this thesis. We next present the contribution made in this thesis by the way of discussing the lessons learned followed by the discussion on related aspects and how to improve and extend this work.

6.1 Contributions

We build two data-driven models of perceptual attributes of 3D shapes. The first model learns a metric of style similarity between pairs of 3D shapes while the other allows computing a real value as a measure of visual aesthetics of a 3D shape. The two measures are essentially represented as two different real valued functions, one modelled using metric learning and the other using deep learning.

We demonstrate that it is possible to extend metric learning (Chapter 3) by using descriptors that include colour and texture information in addition to shape features. We build two metrics of style similarity. One using triplets collected from large number of participants on crowdsourcing platform Amazon Mechanical Turk. We call this 'objective' style similarity metric as it reflects the style judgements of a large crowd. The second metric is built by adapt-

ing the objective metric to one specific user's style preferences. We thus call it 'personalised' style metric. We found that some users, although a minority, desired to adapt the objective metric to their style preferences. The GUI-based application we designed to validate our technique allows a user to perform style based search, scene composition by interactively searching, matching, and viewing collections of 3D shapes, and personalised style matching. Our results show that in addition to shape characteristics, the colour and texture attributes play a vital role in style similarity of 3D shapes.

The approach to build a data-driven model of 3D shape aesthetics is carried in two steps. In the first step (Chapter 4), we run a crowdsourcing study to conclude two important results. First, we show that using either one viewpoint or multi-viewpoints images to collect aesthetics is equally good. Our original assumption was that there will be difference between human judgements collected on one-view images and on multi-view images, since multi-view images show more shape information compared to shape information shown by single-view image. Second, on comparison of collected aesthetics judgements using different shape representations we found that for humans shape details don't matter much when they compare and discriminate aesthetics of shapes in pairs. This means that the coarse shape representations such as voxels are as good as the polygonal shape representations for collecting shape aesthetics judgement data. Since polygonal shape representations can't be used directly in deep learning based data-driven methods. We use the results of our study to choose voxels as the input shape representation for data-driven model of shape aesthetics.

The second step to build a data-driven model involved designing and training a deep ranking neural network. The formulation of learning technique follows our paired data collection approach by maintaining two copies of the deep neural network. Our deep neural network produces a real value as a measure of aesthetic score of an input 3D shape. This computed aesthetic value is relative and can be used to rank shapes from high to low aesthetics, for instance. Our aesthetics measure is aligned with the large amount of data collected on human aesthetics preferences. This means that we don't base our formulation on a specific set of shape features such as curvature or symmetry. We let the deep neural network learn what features are important for aesthetics. We demonstrate the learned aesthetics measure by building user interface to rank shapes in large data-sets and creating visualisations.

6.2 Discussion

6.2.1 Quality of geometry

The quality and organisation of geometric data available in online shape repositories is still a concern. This lack in quality sometimes requires manual pre-processing of large number of shapes before they can be used for research. We found that many shapes are classified into in-

correct object categories. For example, there are many shapes in chairs category which are not semantically chairs. These need to be manually removed from the collection to not to bias the learning results and perceptual data collection process. Further, a large number of shapes have abnormalities in their geometry. For example, much geometry have disoriented faces, holes, very high number of polygons, misalignment in relation to others, and non-uniform scaling etc. The presence of these makes it impossible to rely on these repositories without either fixing or discarding these geometries. In our research, we spent a lot of time by first manually going through the shapes and then either fixing or discarding many. We argue that since a large number of researchers rely on such online shape repositories, the quality of data available there for research can thus be improved to save time wasted in manually going through the geometries.

6.2.2 Aesthetic shape modelling and auto-enhancing shape aesthetics

The key challenge still faced by computer graphics research community is to develop easy-to-use tools for casual and novice users to create 3D content. One way to overcome this challenge is by using data-driven models to synthesise new shapes automatically. For example, data-driven models can learn to generate shape models from a set of exemplars so that the synthesised shapes are novel. Our shape aesthetics measure can be extended by such generative models to automate generation of visually more aesthetics shapes. Further this approach could be embedded in an interactive shape modelling interface to suggest aesthetic value of shape being modelled and possible edits to enhance it. One way to do this is by building a data-driven shape editing method. This allows learning a model that characterises the plausible variation of the shapes from a collection of closely related more aesthetic shapes. The learned model can then be used to constrain the user's edit to maintain plausibility.

6.2.3 Shape representations for learning perceptual properties

The geometry of an object can be digitally represented in different ways. The polygon based representations are typically stored as a list of vertices, edges, and faces. When it comes to using these in machine learning, researchers employ either image-based or object based representations. Examples of images based representations are depth-images or single- or multi-viewpoint images rendered from different camera locations. Object based representations include voxelized volume or point clouds. Since, time and memory constraints are critical in deep learning algorithms, the chosen shape representation can have implications on these. We argue that a systematic evaluation of different shape representations as inputs to deep networks is needed to learn their relation with time and memory requirements.

There are many other fundamental perceptual attributes of 3D shapes related to human vision (or visual perceptual attributes) and touch (or tactile perceptual attributes) [60, 61]. In

addition to the two perceptual problems, namely style and aesthetics, which we explored in this thesis, we envision several other interesting perceptual problems to look into to allow better organisation and exploration of large 3D shape repositories. These problems include taste perception from 3D shapes, interestingness of 3D shape, and memorability of 3D shapes. We argue that learning to predict perceptual taste from 3D printed food has applications in 3D food printing, a buzzword in 3D printing industry. A system that can learn to 3D print personalised food could be advantageous to hospitality industry.

6.2.4 Shape perception

We argue that from perception perspective more work is need to explore how humans compare shapes to conclude if one is more or less aesthetic, say in a pair. It is relatively unclear if they look at the overall shapes quickly to give their responses or they perform a more local or part level comparison. We think that use of shape saliency and eye tracking while viewing shapes can be useful. Further, more work is needed to investigate the link between 3D shape aesthetics and other perceptual attributes [54] such as interestingness and memorability, in order for us to be able to answer questions such as "Are aesthetic 3D shapes also interesting?". Furthermore, when source shapes are selected for scene composition, more work is needed to investigate the relative weighting of style similarity between and aesthetic value of source and target shapes.

6.2.5 Crowdsourcing Perceptual Judgements

Although we took inspirations from already published works in computer graphics to design and conduct our crowdsourcing studies, there are several directions to improve the data collection process. First, we can inquire about the details of the display screen used to conduct the study. For example the type of display (mobile or stationary), screen resolution (high or low), dimensions (width and height), or colour range etc. This could allow us to build an understanding of the dependency between display device and perceptual aesthetic judgements. Since focus of our data collection process was visual tasks completion, we do not poll participants on these specific details.

We use several metrics provided by Amazon Mechanical Turk (AMT) to filter out reliable workers. These worker metrics, such as HIT approval rate, are automatically computed based on the work done by workers. For example if a worker attempts a total of 100 tasks and gets only 80 accepted by the requester (or employer who posted the work on AMT), then AMT would assign a HIT approval rate of 80 percent to this worker. These values are like confidence values of the workers measured over all the tasks in which a worker has participated. It is important to understand that this metric is not necessarily an indicator of reliability with regards to any specific task. Thus we suggest that a more rigorous, task-specific subject reliability method needs to be developed to crowdsource perceptual data. Further, future studies could

benefit by paying attention to parameters such as participants viewing conditions, performing visual tests, and collecting reports on time spent on performing tasks.

6.2.6 Style and Aesthetics for Scene composition

In addition to explosion of 3D shapes in online repositories, the number of 3D scenes is also rapidly growing in digital repositories. This growth provides new opportunities and challenges for data-driven scene analysis, editing, synthesis, and reuse. In our work we developed measures of style similarity and shape aesthetics independently. These measures allow for style and aesthetics based scene composition among other uses. However when 3D scenes are composed either interactively or automatically, which property between style and aesthetics contribute to choosing the new shape to add to the existing shapes in the scene needs to be evaluated. For example some users may prefer matching the overall style of the shapes in scene while other may prefer choosing more aesthetic shapes. Developing an understanding of relative weightings between style and aesthetics for building 3D scenes can be beneficial to automating the building of data-driven models for optimisation of these properties to help produce perceptually more stylish and aesthetic scenes.

References

[1] Israel Abramov, Ann Farkas, and Edward Ochsenschlager. A study in classification: style and visual perception. *Visual Anthropology*, 19(3-4):255–274, 2006.

[2] Olivia C Adkins and J Farley Norman. The visual aesthetics of snowflakes. *Perception*, 45(11):1304–1319, 2016.

[3] Abhishek Agrawal, Vittal Premachandran, and Ramakrishna Kakarala. Rating image aesthetics using a crowd sourcing approach. In *Pacific-Rim Symposium on Image and Video Technology*, pages 24–32. Springer, 2013.

[4] Alan Agresti. A survey of exact inference for contingency tables. *Statistical science*, pages 131–153, 1992.

[5] Jean-Julien Aucouturier, Francois Pachet, et al. Music similarity measures: What's the use? In *ISMIR*, pages 13–17, 2002.

[6] M Dorothee Augustin, Johan Wagemans, and Claus-Christian Carbon. All is beautiful? generality vs. specificity of word usage in visual aesthetics. *Acta Psychologica*, 139(1): 187–201, 2012.

[7] Shumeet Baluja. Learning typographic style: from discrimination to synthesis. *Machine Vision and Applications*, pages 1–18, 2017.

[8] Tara S Behrend, David J Sharek, Adam W Meade, and Eric N Wiebe. The viability of crowdsourcing for survey research. *Behavior research methods*, 43(3):800, 2011.

[9] Sean Bell and Kavita Bala. Learning visual similarity for product design with convolutional neural networks. *ACM Transactions on Graphics (TOG)*, 34(4):98, 2015.

[10] Aurélien Bellet, Amaury Habrard, and Marc Sebban. A survey on metric learning for feature vectors and structured data. *arXiv preprint arXiv:1306.6709*, 2013.

[11] Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1798–1828, 2013.

[12] Steven Bergen and Brian J Ross. Aesthetic 3D model evolution. In *EvoMUSART*, pages 11–22. Springer, 2012.

[13] Itamar Berger, Ariel Shamir, Moshe Mahler, Elizabeth Carter, and Jessica Hodgins. Style and abstraction in portrait sketching. *ACM Transactions on Graphics (TOG)*, 32 (4):55, 2013.

[14] Benjamin Bustos, Daniel A Keim, Dietmar Saupe, Tobias Schreck, and Dejan V Vranić. Feature-based similarity search in 3D object databases. *ACM Computing Surveys (CSUR)*, 37(4):345–387, 2005.

[15] Chiu-Shui Chan. Can style be measured? *Design studies*, 21(3):277–291, 2000.

[16] Chiu-Shui Chan. An examination of the forces that generate a style. *Design Studies*, 22 (4):319–346, 2001.

[17] Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, et al. Shapenet: An information-rich 3D model repository. *arXiv preprint arXiv:1512.03012*, 2015.

[18] Ding-Yun Chen, Xiao-Pei Tian, Yu-Te Shen, and Ming Ouhyoung. On visual similarity based 3D model retrieval. In *Computer graphics forum*, pages 223–232. Wiley Online Library, 2003.

[19] Shi-Zhe Chen, Chun-Chao Guo, and Jian-Huang Lai. Deep ranking for person re-identification via joint representation learning. *IEEE Transactions on Image Processing*, 25(5):2353–2367, 2016.

[20] Xiaobai Chen, Abulhair Saparov, Bill Pang, and Thomas Funkhouser. Schelling points on 3D surface meshes. *ACM Transactions on Graphics (TOG)*, 31(4):29, 2012.

[21] Lin Hou Chew, Jason Teo, and James Mountstephens. Aesthetic preference recognition of 3D shapes using eeg. *Cognitive neurodynamics*, 10(2):165–173, 2016.

[22] Paolo Cignoni, Marco Callieri, Massimiliano Corsini, Matteo Dellepiane, Fabio Ganovelli, and Guido Ranzuglia. Meshlab: an open-source mesh processing tool. In *Eurographics Italian Chapter Conference*, volume 2008, pages 129–136, 2008.

[23] Forrester Cole, Kevin Sanik, Doug DeCarlo, Adam Finkelstein, Thomas Funkhouser, Szymon Rusinkiewicz, and Manish Singh. How well do line drawings depict shape? *ACM Transactions on Graphics (ToG)*, 28(3):28, 2009.

[24] Ritendra Datta, Dhiraj Joshi, Jia Li, and James Z Wang. Studying aesthetics in photographic images using a computational approach. In *European Conference on Computer Vision*, pages 288–301. Springer, 2006.

[25] Sibel S Dazkir and Marilyn A Read. Furniture forms and their influence on our emotional responses toward interior environments. *Environment and Behavior*, 44(5):722–732, 2012.

[26] Doug DeCarlo, Adam Finkelstein, Szymon Rusinkiewicz, and Anthony Santella. Suggestive contours for conveying shape. *ACM Transactions on Graphics (TOG)*, 22(3): 848–855, 2003.

[27] Li Deng. A tutorial survey of architectures, algorithms, and applications for deep learning. *APSIPA Transactions on Signal and Information Processing*, 3, 2014.

[28] Yubin Deng, Chen Change Loy, and Xiaoou Tang. Image aesthetic assessment: An experimental survey. *IEEE Signal Processing Magazine*, 34(4):80–106, 2017.

[29] Kapil Dev, Manfred Lau, and Ligang Liu. A perceptual aesthetics measure for 3D shapes. *arXiv preprint arXiv:1608.04953*, 2016.

[30] Alexey Dosovitskiy, Jost Tobias Springenberg, Maxim Tatarchenko, and Thomas Brox. Learning to generate chairs, tables and cars with convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 39(4):692–705, 2017.

[31] Julie S Downs, Mandy B Holbrook, Steve Sheng, and Lorrie Faith Cranor. Are your participants gaming the system?: screening mechanical turk workers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2399–2402. ACM, 2010.

[32] Denis Dutton. *The art instinct: Beauty, pleasure, & human evolution*. Oxford University Press, USA, 2009.

[33] Karina Rodriguez Echavarria and Ran Song. Analyzing the decorative style of 3D heritage collections based on shape saliency. *Journal on Computing and Cultural Heritage (JOCCH)*, 9(4):20, 2016.

[34] Yael Eisenthal, Gideon Dror, and Eytan Ruppin. Facial attractiveness: Beauty and the machine. *Neural Computation*, 18(1):119–142, 2006.

[35] Steven Feiner, Blair Macintyre, and Dorée Seligmann. Knowledge-based augmented reality. *Communications of the ACM*, 36(7):53–62, 1993.

[36] James A Ferwerda, Stephen H Westin, Randall C Smith, and Richard Pawlicki. Effects of rendering on shape perception in automobile design. In *Proceedings of the 1st Symposium on Applied perception in graphics and visualization*, pages 107–114. ACM, 2004.

[37] Ann Marie Fiore. *Understanding aesthetics for the merchandising and design professional*. A&C Black, 2010.

[38] Jay Friedenberg and Marco Bertamini. Aesthetic preference for polygon shape. *Empirical Studies of the Arts*, 33(2):144–160, 2015.

[39] Kikuo Fujita, Takafumi Nakayama, and Shinsuke Akagi. Integrated product design methodology for aesthetics, functions and geometry with feature based modeling and constraint management. In *International Conference on Engineering Design, ICED*, volume 99, pages 24–26, 1999.

[40] Megan Gambino. Do our brains find certain shapes more attractive than others? http://www.smithsonianmag.com/science-nature/do-our-brains-find-certain-shapes-more-attractive/, 2013. [Online; accessed 20-August-2017].

[41] Elena Garces, Aseem Agarwala, Diego Gutierrez, and Aaron Hertzmann. A similarity measure for illustration style. *ACM Transactions on Graphics (TOG)*, 33(4):93, 2014.

[42] Elena Garces, Aseem Agarwala, Aaron Hertzmann, and Diego Gutierrez. Style-based exploration of illustration datasets. *Multimedia Tools and Applications*, 76(11):13067–13086, 2017.

[43] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. A neural algorithm of artistic style. *arXiv preprint arXiv:1508.06576*, 2015.

[44] Deepti Ghadiyaram and Alan C Bovik. Massive online crowdsourced study of subjective and objective picture quality. *IEEE Transactions on Image Processing*, 25(1):372–387, 2016.

[45] Franca Giannini and Marina Monti. Cad tools based on aesthetic properties. *Eurographics, Italian Chapter, July*, pages 11–12, 2002.

[46] Yotam Gingold, Ariel Shamir, and Daniel Cohen-Or. Micro perceptual human computation for visual tasks. *ACM Transactions on Graphics (TOG)*, 31(5):119, 2012.

[47] Rohit Girdhar, David F Fouhey, Mikel Rodriguez, and Abhinav Gupta. Learning a predictable and generative vector representation for objects. In *European Conference on Computer Vision*, pages 484–499. Springer, 2016.

[48] Michael Gygli, Helmut Grabner, Hayko Riemenschneider, Fabian Nater, and Luc Van Gool. The interestingness of images. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 1633–1640. IEEE, 2013.

[49] Jeffrey Heer and Michael Bostock. Crowdsourcing graphical perception: using mechanical turk to assess visualization design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 203–212. ACM, 2010.

[50] Ian Hodder. *Symbols in action: ethnoarchaeological studies of material culture*. Cambridge University Press, 1982.

[51] Junlin Hu, Jiwen Lu, and Yap-Peng Tan. Discriminative deep metric learning for face verification in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1875–1882, 2014.

[52] Ruizhen Hu, Wenchao Li, Oliver Van Kaick, Hui Huang, Melinos Averkiou, Daniel Cohen-Or, and Hao Zhang. Co-locating style-defining elements on 3D shapes. *ACM Transactions on Graphics (TOG)*, 36(3):33, 2017.

[53] Zhenhen Hu, Yonggang Wen, Luoqi Liu, Jianguo Jiang, Richang Hong, Meng Wang, and Shuicheng Yan. Visual classification of furniture styles. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 8(5):67, 2017.

[54] Phillip Isola, Jianxiong Xiao, Devi Parikh, Antonio Torralba, and Aude Oliva. What makes a photograph memorable? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(7):1469–1482, 2014.

[55] Dhiraj Joshi, Ritendra Datta, Elena Fedorovskaya, Quang-Tuan Luong, James Z Wang, Jia Li, and Jiebo Luo. Aesthetics and emotions in images. *IEEE Signal Processing Magazine*, 28(5):94–115, 2011.

[56] Sergey Karayev, Matthew Trentacoste, Helen Han, Aseem Agarwala, Trevor Darrell, Aaron Hertzmann, and Holger Winnemoeller. Recognizing image style. *arXiv preprint arXiv:1311.3715*, 2013.

[57] Adriana Kovashka and Kristen Grauman. Attribute adaptation for personalized image search. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3432–3439, 2013.

[58] Yuki Koyama, Issei Sato, Daisuke Sakamoto, and Takeo Igarashi. Sequential line search for efficient visual design optimization by crowds. *ACM Transactions on Graphics (TOG)*, 36(4):48, 2017.

[59] Brian Kulis et al. Metric learning: A survey. *Foundations and Trends® in Machine Learning*, 5(4):287–364, 2013.

[60] Manfred Lau, Kapil Dev, Julie Dorsey, and Holly Rushmeier. Learning a human-perceived softness measure of virtual 3D objects. In *Proceedings of the ACM Symposium on Applied Perception*, pages 65–68. ACM, 2016.

[61] Manfred Lau, Kapil Dev, Weiqi Shi, Julie Dorsey, and Holly Rushmeier. Tactile mesh saliency. *ACM Trans. Graph.*, 35(4):52:1–52:11, July 2016. ISSN 0730-0301.

[62] Tommer Leyvand, Daniel Cohen-Or, Gideon Dror, and Dani Lischinski. Data-driven enhancement of facial attractiveness. *ACM Transactions on Graphics (TOG)*, 27(3):38, 2008.

[63] Qiqi Liao, Xiaogang Jin, and Wenting Zeng. Enhancing the symmetry and proportion of 3D face geometry. *IEEE transactions on visualization and computer graphics*, 18(10): 1704–1716, 2012.

[64] Isaak Lim, Anne Gehre, and Leif Kobbelt. Identifying style of 3D shapes using deep metric learning. *Computer Graphics Forum*, 35(5):207, 2016.

[65] Ming-huang Lin, Ching-yi Wang, Shih-kuen Cheng, and Shih-hung Cheng. An event-related potential study of semantic style-match judgments of artistic furniture. *International Journal of Psychophysiology*, 82(2):188–195, 2011.

[66] Ligang Liu, Renjie Chen, Lior Wolf, and Daniel Cohen-Or. Optimizing photo composition. *Computer Graphics Forum*, 29(2):469–478, 2010.

[67] Tianqiang Liu, Aaron Hertzmann, Wilmot Li, and Thomas Funkhouser. Style compatibility for 3D furniture models. *ACM Transactions on Graphics (TOG)*, 34(4):85, 2015.

[68] Zhen-Bao Liu, Shu-Hui Bu, Kun Zhou, Shu-Ming Gao, Jun-Wei Han, and Jun Wu. A survey on partial retrieval of 3D shapes. *Journal of Computer Science and Technology*, 28 (5):836–851, 2013.

[69] P Locher and C Nodine. The perceptual value of symmetry. *Computers & mathematics with applications*, 17(4-6):475–484, 1989.

[70] Aruna Lorensuhewa, Shlomo Geva, and Binh Pham. Inferencing design styles using bayesian networks. *Ruhuna Journal of Science*, 1(1), 2012.

[71] Xin Lu, Zhe Lin, Hailin Jin, Jianchao Yang, and James Z Wang. Rating image aesthetics using deep learning. *IEEE Transactions on Multimedia*, 17(11):2021–2034, 2015.

[72] Xin Lu, Zhe Lin, Xiaohui Shen, Radomir Mech, and James Z Wang. Deep multi-patch aggregation network for image style, aesthetics, and quality estimation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 990–998, 2015.

[73] Yen-nien Lu and Chun-heng Ho. Impact of curvature of product shape on aesthetic judgments. In *KEER2014. Proceedings of the 5th Kanesi Engineering and Emotion Research; International Conference; Linköping; Sweden; June 11-13*, pages 743–754. Linköping University Electronic Press, 2014.

[74] Zhaoliang Lun, Evangelos Kalogerakis, and Alla Sheffer. Elements of style: learning perceptual shape style similarity. *ACM Transactions on Graphics (TOG)*, 34(4):84, 2015.

[75] Zhaoliang Lun, Evangelos Kalogerakis, Rui Wang, and Alla Sheffer. Functionality preserving shape style transfer. *ACM Transactions on Graphics (TOG)*, 35(6):209, 2016.

[76] Shi-Jian Luo, Ye-Tao Fu, and Yu-Xiao Zhou. Perceptual matching of shape design style between wheel hub and car type. *International Journal of Industrial Ergonomics*, 42(1): 90–102, 2012.

[77] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(Nov):2579–2605, 2008.

[78] Evgeni Magid, Octavian Soldea, and Ehud Rivlin. A comparison of gaussian and mean curvature estimation methods on triangular meshes of range image data. *Computer Vision and Image Understanding*, 107(3):139–159, 2007.

[79] Long Mai, Hailin Jin, and Feng Liu. Composition-preserving deep photo aesthetics assessment. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 497–506, 2016.

[80] Rafał K Mantiuk, Anna Tomaszewska, and Radosław Mantiuk. Comparison of four subjective methods for image quality assessment. *Computer Graphics Forum*, 31(8): 2478–2491, 2012.

[81] Arthur B Markman and Dedre Gentner. Nonintentional similarity processing. *The new unconscious*, pages 107–137, 2005.

[82] Rachel McDonnell, Martin Breidt, and Heinrich H Bülthoff. Render me real?: investigating the effect of render style on the perception of animated virtual humans. *ACM Transactions on Graphics (TOG)*, 31(4):91, 2012.

[83] Mark Meyer, Mathieu Desbrun, Peter Schröder, and Alan H Barr. Discrete differential-geometry operators for triangulated 2-manifolds. In *Visualization and mathematics III*, pages 35–57. Springer, 2003.

[84] George A Miller. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41, 1995.

[85] Niloy J Mitra, Leonidas J Guibas, and Mark Pauly. Symmetrization. *ACM Transactions on Graphics (TOG)*, 26(3):63, 2007.

[86] Kenjiro T Miura and RU Gobithaasan. Aesthetic curves and surfaces in computer aided geometric design. *IJAT*, 8(3):304–316, 2014.

[87] Jack L Nasar. Urban design aesthetics: The evaluative qualities of building exteriors. *Environment and behavior*, 26(3):377–401, 1994.

[88] Thi Phuong Nghiem, Axel Carlier, Geraldine Morin, and Vincent Charvillat. Enhancing online 3D products through crowdsourcing. In *Proceedings of the ACM multimedia 2012 workshop on Crowdsourcing for multimedia*, pages 47–52. ACM, 2012.

[89] Peter O'Donovan, Jānis Lībeks, Aseem Agarwala, and Aaron Hertzmann. Exploratory font selection using crowdsourced attributes. *ACM Transactions on Graphics (TOG)*, 33(4):92, 2014.

[90] Robert Osada, Thomas Funkhouser, Bernard Chazelle, and David Dobkin. Matching 3D models with shape distributions. In *Shape Modeling and Applications, SMI 2001 International Conference On.*, pages 154–166. IEEE, 2001.

[91] Alice J O'Toole, Theodore Price, Thomas Vetter, James C Bartlett, and Volker Blanz. 3D shape and 2d surface textures of human faces: the role of "averages" in attractiveness and age. *Image and Vision Computing*, 18(1):9–19, 1999.

[92] Stephen E Palmer, Karen B Schloss, and Jonathan Sammartino. Visual aesthetics and human preference. *Annual review of psychology*, 64:77–107, 2013.

[93] Eleonora Pantano, Alexandra Rese, and Daniel Baier. Enhancing the online decision-making process by using augmented reality: A two country comparison of youth markets. *Journal of Retailing and Consumer Services*, 38:81–95, 2017.

[94] Devi Parikh and Kristen Grauman. Relative attributes. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 503–510. IEEE, 2011.

[95] Jungchul Park and Sung H Han. A fuzzy rule-based approach to modeling affective user satisfaction towards office chair design. *International Journal of Industrial Ergonomics*, 34(1):31–47, 2004.

[96] Gabriella Pasi. Flexible information retrieval: some research trends. *Mathware & soft computing. 2002 Vol. 9 Núm. 1*, 2002.

[97] Heinz-Otto Peitgen and Peter H Richter. *The beauty of fractals: images of complex dynamical systems*. Springer Science & Business Media, 2013.

[98] Gabriel Peyré and Laurent D Cohen. Geodesic remeshing using front propagation. *International Journal of Computer Vision*, 69(1):145, 2006.

[99] Binh Pham and Jinglan Zhang. A fuzzy shape specification system to support design for aesthetics. *Studies in Fuzziness and Soft Computing*, 127:39–50, 2003.

[100] Michal Piovarči, David IW Levin, Jason Rebello, Desai Chen, Roman Ďurikovič, Hanspeter Pfister, Wojciech Matusik, and Piotr Didyk. An interaction-aware, perceptual model for non-linear elastic objects. *ACM Transactions on Graphics (TOG)*, 35(4): 55, 2016.

[101] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3D classification and segmentation. *arXiv preprint arXiv:1612.00593*, 2016.

[102] Judith Redi and Isabel Povoa. Crowdsourcing for rating image aesthetic appeal: Better a paid or a volunteer crowd? In *Proceedings of the 2014 International ACM Workshop on Crowdsourcing for Multimedia*, pages 25–30. ACM, 2014.

[103] Christoph Redies, Seyed Ali Amirshahi, Michael Koch, and Joachim Denzler. Phog-derived aesthetic measures applied to color photographs of artworks, natural scenes and objects. In *European Conference on Computer Vision*, pages 522–531. Springer, 2012.

[104] Christopher P Said and Alexander Todorov. A statistical model of facial attractiveness. *Psychological Science*, 22(9):1183–1190, 2011.

[105] Babak Saleh, Mira Dontcheva, Aaron Hertzmann, and Zhicheng Liu. Learning style similarity for searching infographics. In *Proceedings of the 41st graphics interface conference*, pages 59–64. Canadian Information Processing Society, 2015.

[106] Jitao Sang, Changsheng Xu, and Dongyuan Lu. Learn to personalized image search from the photo sharing websites. *IEEE Transactions on Multimedia*, 14(4):963–974, 2012.

[107] Adrian Secord, Jingwan Lu, Adam Finkelstein, Manish Singh, and Andrew Nealen. Perceptual models of viewpoint preference. *ACM Transactions on Graphics (TOG)*, 30(5): 109, 2011.

[108] Carlo H Séquin. Cad tools for aesthetic engineering. *Computer-Aided Design*, 37(7): 737–750, 2005.

[109] Lior Shapira, Ariel Shamir, and Daniel Cohen-Or. Consistent mesh partitioning and skeletonisation using the shape diameter function. *The Visual Computer*, 24(4):249, 2008.

[110] Richard C Smardon. Perception and aesthetics of the urban environment: Review of the role of vegetation. *Landscape and Urban Planning*, 15(1-2):85–106, 1988.

[111] Martin Stacey. Psychological challenges for the analysis of style. *Ai Edam*, 20(3):167–184, 2006.

[112] Stephan Streuber, M Alejandra Quiros-Ramirez, Matthew Q Hill, Carina A Hahn, Silvia Zuffi, Alice O'Toole, and Michael J Black. Body talk: Crowdshaping realistic 3D avatars with words. *ACM Transactions on Graphics (TOG)*, 35(4):54, 2016.

[113] Hang Su, Subhransu Maji, Evangelos Kalogerakis, and Erik Learned-Miller. Multi-view convolutional neural networks for 3D shape recognition. In *Proceedings of the IEEE international conference on computer vision*, pages 945–953, 2015.

[114] Jeff KT Tang, Wan-Man Lau, Kwun-Kit Chan, and Kwok-Ho To. Ar interior designer: Automatic furniture arrangement using spatial and functional relationships. In *Virtual Systems & Multimedia (VSMM), 2014 International Conference on*, pages 345–352. IEEE, 2014.

[115] Johan WH Tangelder and Remco C Veltkamp. A survey of content based 3D shape retrieval methods. In *Shape Modeling Applications, 2004. Proceedings*, pages 145–156. IEEE, 2004.

[116] Johan WH Tangelder and Remco C Veltkamp. A survey of content based 3D shape retrieval methods. *Multimedia tools and applications*, 39(3):441, 2008.

[117] Joshua B Tenenbaum and William T Freeman. Separating style and content. In *Advances in neural information processing systems*, pages 662–668, 1997.

[118] James T Todd. The visual perception of 3D shape. *Trends in cognitive sciences*, 8(3): 115–121, 2004.

[119] Oshin Vartanian, Gorka Navarrete, Anjan Chatterjee, Lars Brorson Fich, Helmut Leder, Cristián Modroño, Marcos Nadal, Nicolai Rostrup, and Martin Skov. Impact of contour on aesthetic judgments and approach-avoidance decisions in architecture. *Proceedings of the National Academy of Sciences*, 110(Supplement 2):10446–10453, 2013.

[120] Andreas Veit, Balazs Kovacs, Sean Bell, Julian McAuley, Kavita Bala, and Serge Belongie. Learning visual clothing style with heterogeneous dyadic co-occurrences. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4642–4650, 2015.

[121] Chensheng Wang, Joris SM Vergeest, Tjamme Wiegers, DJ van der Pant, and TMC van den Berg. Exploring the influence of feature geometry on design style. *Electrical and Computer Engineering Series: Computational Methods in Circuits and Systems Application*, pages 21–28, 2003.

[122] Jiang Wang, Yang Song, Thomas Leung, Chuck Rosenberg, Jingbin Wang, James Philbin, Bo Chen, and Ying Wu. Learning fine-grained image similarity with deep ranking. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1386–1393, 2014.

[123] Peng-Shuai Wang, Yang Liu, Yu-Xiao Guo, Chun-Yu Sun, and Xin Tong. O-cnn: octree-based convolutional neural networks for 3D shape analysis. *ACM Transactions on Graphics (TOG)*, 36(4):72, 2017.

[124] Somlak Wannarumon. An aesthetics driven approach to jewelry design. *Computer-Aided Design and Applications*, 7(4):489–503, 2010.

[125] Dorothy K Washburn. Style, perception, and geometry. In *Style, Society, and Person*, pages 101–122. Springer, 1995.

[126] Lingyu Wei, Qixing Huang, Duygu Ceylan, Etienne Vouga, and Hao Li. Dense human body correspondences using convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1544–1553, 2016.

[127] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3D shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1912–1920, 2015.

[128] Ioannis Xenakis and Argyris Arnellos. Aesthetic perception and its minimal content: a naturalistic perspective. *Frontiers in psychology*, 5, 2014.

[129] Eric P Xing, Michael I Jordan, Stuart J Russell, and Andrew Y Ng. Distance metric learning with application to clustering with side-information. In *Advances in neural information processing systems*, pages 521–528, 2003.

[130] Kai Xu, Honghua Li, Hao Zhang, Daniel Cohen-Or, Yueshan Xiong, and Zhi-Quan Cheng. Style-content separation by anisotropic part scales. *ACM Transactions on Graphics (TOG)*, 29(6):184, 2010.

[131] Kai Xu, Vladimir G Kim, Qixing Huang, and Evangelos Kalogerakis. Data-driven shape analysis and processing. In *Computer Graphics Forum*, pages 101–132. Wiley Online Library, 2017.

[132] Ting Yao, Tao Mei, and Yong Rui. Highlight detection with pairwise deep ranking for first-person video summarization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 982–990, 2016.

[133] Mehmet Ersin Yumer and Levent Burak Kara. Surface creation on unstructured point sets using neural networks. *Computer-Aided Design*, 44(7):644–656, 2012.

[134] Sergey Zagoruyko and Nikos Komodakis. Learning to compare image patches via convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4353–4361, 2015.

[135] Eduard Zell, Carlos Aliaga, Adrian Jarabo, Katja Zibrek, Diego Gutierrez, Rachel McDonnell, and Mario Botsch. To stylize or not to stylize?: the effect of shape and material stylization on the perception of computer-generated faces. *ACM Transactions on Graphics (TOG)*, 34(6):184, 2015.

[136] Jiajing Zhang, Jinhui Yu, Kang Zhang, Xianjun Sam Zheng, and Junsong Zhang. Computational aesthetic evaluation of logos. *ACM Transactions on Applied Perception (TAP)*, 14(3):20, 2017.

[137] Luming Zhang. Describing human aesthetic perception by deeply-learned attributes from flickr. *arXiv preprint arXiv:1605.07699*, 2016.

[138] Xueyi Zhao, Xi Li, and Zhongfei Zhang. Multimedia retrieval via deep learning to rank. *IEEE Signal Processing Letters*, 22(9):1487–1491, 2015.

[139] Jun-Yan Zhu, Aseem Agarwala, Alexei A Efros, Eli Shechtman, and Jue Wang. Mirror mirror: Crowdsourcing better portraits. *ACM Transactions on Graphics (TOG)*, 33(6): 234, 2014.