**Framework to create cloud-free remote sensing data using passenger aircraft as the platform**

Chisheng Wang1, Shuying Wang1, Hongxing Cui1,2,3\*, Monja B. Šebela4, Ce Zhang5, Xiaowei Gu6, Xu Fang1, Zhongwen Hu1, Qiandi Tang1 and Yongquan Wang1

*1Ministry of Natural Resources (MNR) Key Laboratory for Geo-Environmental Monitoring of Great Bay Area & Guangdong Key Laboratory of Urban Informatics & Shenzhen Key Laboratory of Spatial Smart Sensing and Services, the School of Architecture and Urban Planning, Shenzhen 518000, China; 2South Marine Science and Engineering Guangdong Laboratory, Guangzhou, China; 3Department of Ocean Science, Hong Kong University of Science and Technology, Hong Kong, SAR, China; 4Sinergise Laboratory for Geographical Information Systems Ltd., Cvetkova ulica 29, 1000 Ljubljana, Slovenia; 5Lancaster Environment Centre, Lancaster University, Lancaster, United Kingdom LA1 4YQ; 6Department of Computer Science, Aberystwyth University, Aberystwyth, United Kingdom, SY23 3DB*

***Abstract***

Cloud removal in optical remote sensing imagery is essential for many Earth observation applications. To recover the cloud obscured information, some preconditions must be satisfied. For example, the cloud must be semi-transparent or relationships between contaminated and cloud-free pixels must be assumed. Deep-learning-based cloud removal methods are widely used to predict cloud-free imagery; however, the uncertainties in repairing fully contaminated cloud imagery are significant. Due to the inherent imaging geometry features in satellite remote sensing, it is impossible to observe the ground under the clouds directly; therefore, cloud removal algorithms are not completely accurate owing to the loss of ground truth. Recently, the use of passenger aircraft as a platform for remote sensing has been proposed by some researchers and institutes, including Airbus and the Japan Aerospace Exploration Agency. Passenger aircraft have the advantages of short visitation frequency and low cost. Additionally, because passenger aircraft fly at lower altitudes compared to satellites, they can observe the ground under the clouds at an oblique viewing angle. In this study, we examine the possibility of creating cloud-free remote sensing data by stacking multi-angle images captured by passenger aircraft. To accomplish this, a processing framework is proposed, which includes four main steps: 1) multi-angle image acquisition from passenger aircraft, 2) cloud detection based on state-of-the-art deep learning semantic segmentation models, 3) cloud removal by image stacking, and 4) image quality enhancement via haze removal. This method is intended to remove cloud contamination without the requirements of reference images and pre-determination of cloud types. The proposed method was tested in multiple case studies, wherein the resultant cloud- and haze-free orthophotos were visualized and quantitatively analyzed in various land cover type scenes. The results of the case studies demonstrated that the proposed method could generate high quality, cloud-free orthophotos. Therefore, we conclude that this framework has great potential for creating cloud-free remote sensing images when the cloud removal of satellite imagery is difficult or inaccurate.

***Keywords:*** *Passenger aircraft, Multiple viewing angles, Cloud removal, Haze removal, Deep learning, Photogrammetry*

# 1. Introduction

With the rapid development of remote sensing technology in recent decades, remote sensing satellite images have been widely applied in various Earth observation activities, such as climate change assessment, land use and land cover identification, crop mapping, and change detection [1-5]. Optical satellite images from space-based remote sensing platforms have played a critical role in Earth observation activities; however, cloud coverage is problematic in the retrieval of surface or atmospheric parameters [6,7], feature extraction [8], and dynamic detection [9] from optical images due to their spectral band distribution from visible to near-wavelengths [10]. With the dramatic increase in satellite data obtained from remote sensing observations, problematic cloud-contamination in optical remote sensing images has become more apparent. Approximately 67% of the Moderate Resolution Imaging Spectroradiometer images are affected by clouds [11]. In remote sensing imagery, thick cloud coverage creates opaque pixels that obscure the ground surface; therefore, precise identification and removal of cloud coverage are essential for using remote sensing data.

Many cloud removal methods for optical remote sensing data have been presented recently. Traditional methods can be divided into three major categories: multitemporal [12-14], multispectral [15-17], and spatial-based approaches [18-20]. Lin et al. [12] proposed a cloud removal method that uses multitemporal satellite images and information cloning, wherein cloudy areas are cloned from corresponding cloud-free (CF) areas based on a global optimization process and the Poisson equation. A non-negative matrix factorization and error correction method has been used to remove clouds using multitemporal remote sensing data from different sensors [14]. Additionally, Xu et al. [13] introduced a cloud removal method using sparse representation and multitemporal dictionary learning techniques. Multitemporal approaches, which use temporal images with different acquisition dates to retrieve images without corrupted pixels to yield a CF image, are the most prominent techniques. However, CF reference imagery is required for this method, which complicates scene reconstruction due to rapidly changing surface conditions. Multispectral approaches are suitable for the removal of semi-transparent clouds and haze. A thin cloud removal approach based on multidirectional dual-tree complex wavelet transform and transfer least square support vector regression was proposed in a previous study [15]. Xu et al. [16] developed a thin cloud removal method using signal transmission principles and spectral mixture analysis. These multispectral methods are applied for cloud removal and do not require additional imagery; However, they are limited to semi-transparent clouds that transmit the spectral properties of the surface to some extent. Spatial-based techniques use the hypothetical relationship between contaminated and CF pixels based on spatial and geometric information. A cloud removal method based on similar pixel replacement driven by a spatio–temporal Markov random field model was previously introduced [18]. Meng et al. [20] applied a sparse-dictionary-learning-based adaptive patch inpainting method based on high-spatial-resolution remote sensing imagery. However, these methods only generate a visually plausible image without clouds and are more complicated due to the increased data requirement. Many additional deep-learning-based methods for cloud removal have been presented [21-24]. Recently, CF data has been produced using deep-learning-based methods regardless of cloud type; however, some uncertainties occur when using one-scene imagery that is fully contaminated by clouds. Therefore, inherent limitations exist in traditional and deep-learning-based methods to generate accurate CF imagery from cloudy imagery. These limitations are difficult to overcome using cloud removal algorithms only. Cloud coverage in remote sensing data is associated with the satellite platform used to obtain the imagery. Generally, remote sensing data with greater cloud coverage percentages are obtained from satellite platforms located at significantly higher altitudes. Meanwhile, the satellite’s field of view (FOV) is relatively fixed, making it difficult to avoid cloud contamination during the imaging process.

Recently, using passenger aircraft as the remote sensing platform has been proposed due to the large coverage area, short visitation frequency, and low cost [25-27]. Passenger aircraft also overcome cloud interference due to their flight altitude and multi-viewing angles. Figure 1 shows that in the image acquisition process, cloud interference increases with satellite platform altitude [28]. However, the passenger aircraft platform has the advantage of multi-view angle observations and can therefore obtain CF surface information, whereas the satellite platform observes fully contaminated surface information. Successful earth observation applications using an aircraft platform have been developed [29,30]. In a previous study [29], a Civil Aircraft for the Regular Investigation of the Atmosphere Based on an Instrument Container Project was used to monitor the atmosphere and was able to provide less costly, real-time meteorological information similar to traditional remote sensing platforms. Passenger aircraft observations have also obtained meteorological data, which has been used in the investigation of the relationship between the COVID-19 pandemic and weather forecast [30]. Recently, programs using passenger aircraft as remote sensing platforms have been instituted by different countries globally [31-33]. Ray20 uses the Airbus aircraft to build autonomous remote sensing systems over Europe and North America [31]. The Norwegian Research Centre conducted a project that used passenger aircraft equipped with high-resolution imaging systems for environmental monitoring. Additionally, a passenger aircraft used as an observation platform can increase safety and emergency response in the Arctic [32]. These results demonstrate that using passenger aircraft as a remote sensing platform has significant potential for earth observation activities in the future.

Figure 1. Map showing cloud impact differences of the satellite and aircraft platforms. A1 is the cloud coverage area using a satellite platform and A2 is the CF area produced using the passenger aircraft platform. H1 is the altitude of the passenger aircraft platform. H2 and H3 are the cloud heights of the different clouds above the earth’s surface. H4 is the altitude of the satellite platform. F1 and F2 are the fields of view of the satellite and passenger aircraft, respectively.

In this study, to overcome the limitations mentioned above, a novel framework was developed to automatically generate truly CF orthophotos from a set of time series imagery taken with a smartphone camera onboard a passenger aircraft. The proposed method can remove cloud contamination without distinguishing between cloud types or requiring a reference image. The proposed framework included four processing steps. First, a series of images were captured using a passenger aircraft as a platform. Second, state-of-the-art semantic segmentation architectures were modified to serve as cloud detection models, which can identify clouds in the imagery with only red, green, and blue channels in different scenes. Third, large-scale CF orthophotos were generated by integrating photogrammetry methods and a set of time-series images with the corresponding cloud masks under the various scenes. Finally, a high-quality haze-free (HF) orthophoto for each scene was obtained. The structure of this study is as follows: Section 2 describes the proposed method in detail, Section 3 introduces the dataset, Section 4 evaluates the metrics, Section 5 presents an analysis and discussion of the results, and Section 6 presents the conclusions.

# 2. Methodology

Data processing was performed in three steps: 1) cloud detection using deep learning algorithms, 2) cloud removal from the orthophotos using photogrammetry methods, and 3) haze removal from the orthophotos using image-enhanced methods (Fig. 2). In the first step, three semantic segmentation models are presented, including U-Net, feature pyramid network (FPN), and pyramid scene parsing network (PSPNet), which were trained and optimized to assign the appropriate cloud mask to the corresponding images. In the second step, the assigned masks were integrated into the structure’s procedure by motion and orthophoto mosaic generation to generate a CF orthophoto. In the third step, the haze in the CF orthophoto was removed using the developed haze removal method that combines a dark channel prior (DCP) [34] and histogram statistics to generate a high-quality HF orthophoto. The data acquisition process using passenger aircraft as the remote sensing platform is discussed in detail in Section 2.1.

## 2.1. Passenger Aircraft Data Acquisition

This section briefly describes data acquisition using passenger aircraft as the platform. In common passenger aircraft, window seats are available from which passengers can capture high-quality images of the ground using handheld cameras (e.g., smartphone camera). To meet photogrammetric processing requirements, two essential rules must be implemented when capturing an image. First, adjacent images with a certain degree of coverage (50%) must be taken to ensure that the images can be aligned successfully. Second, a low incident is required to ensure that more ground details are captured in one picture, as well as to improve dense point matching. Side-look imaging geometry (Fig. 3) can be obtained when pictures are taken by passengers from both sides of the plane; however, the varying scales of the images are a defect in the oblique images [26,35]. To obtain high-quality images, the incidence angle should be controlled within a small range. The average FOV of a smartphone camera is approximately 60°. The minimum incidence angle can be adjusted to approximately half the FOV plus the inclination of the seat (approximately 3° measured from a Boeing 787-8 schematic) during a smooth flight. Generally, the altitude of commercial aircraft ranges between 9 and 11 km, the stripe width is approximately 17 km based on the slant imaging geometry, and the ground resolution ranges between 3.3 and 6.6 m if there are 3000 image pixels along the width.

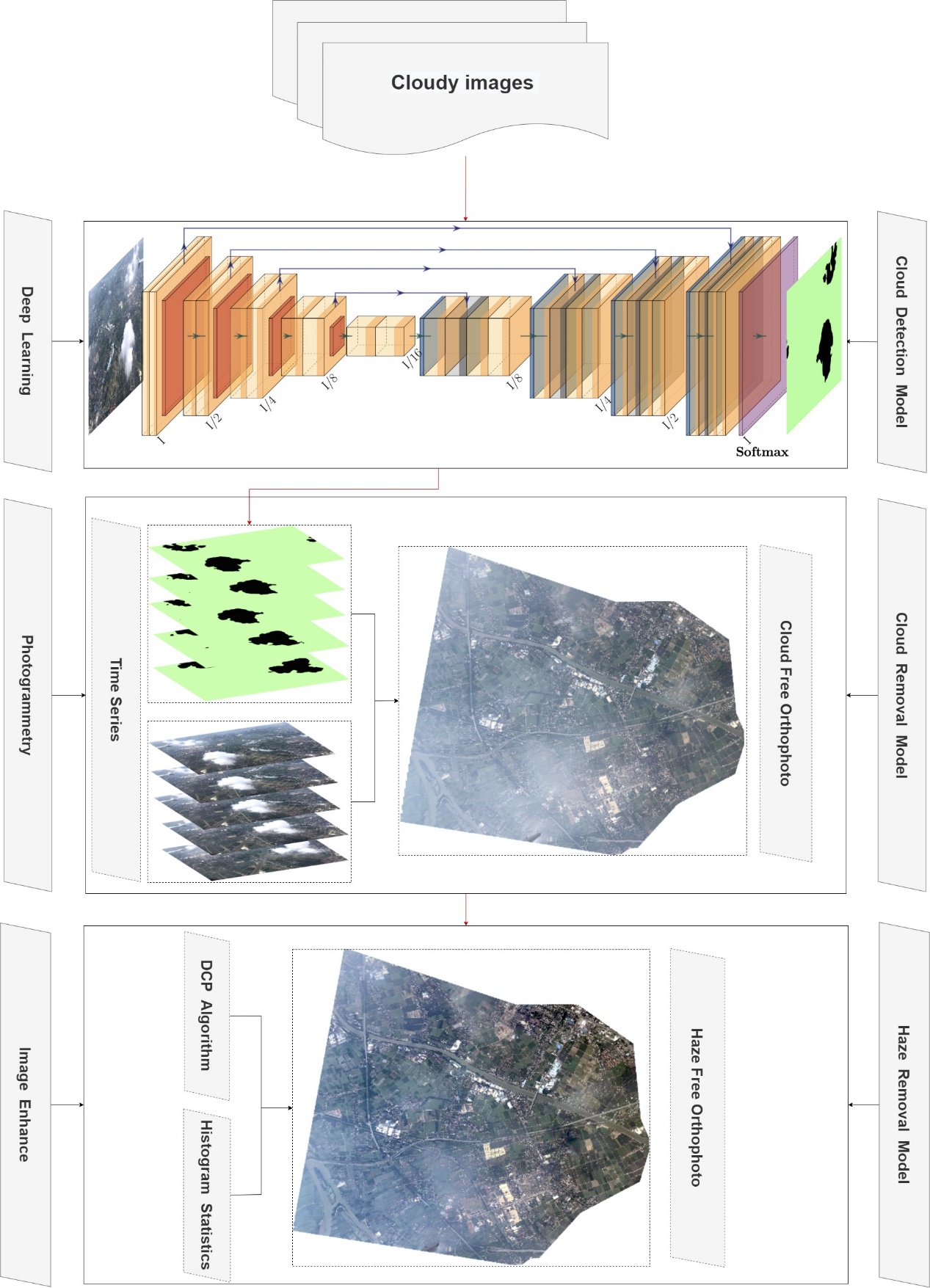


Figure 2. Flowchart of the orthophoto generation process consisting of cloud detection,

cloud removal, and haze removal

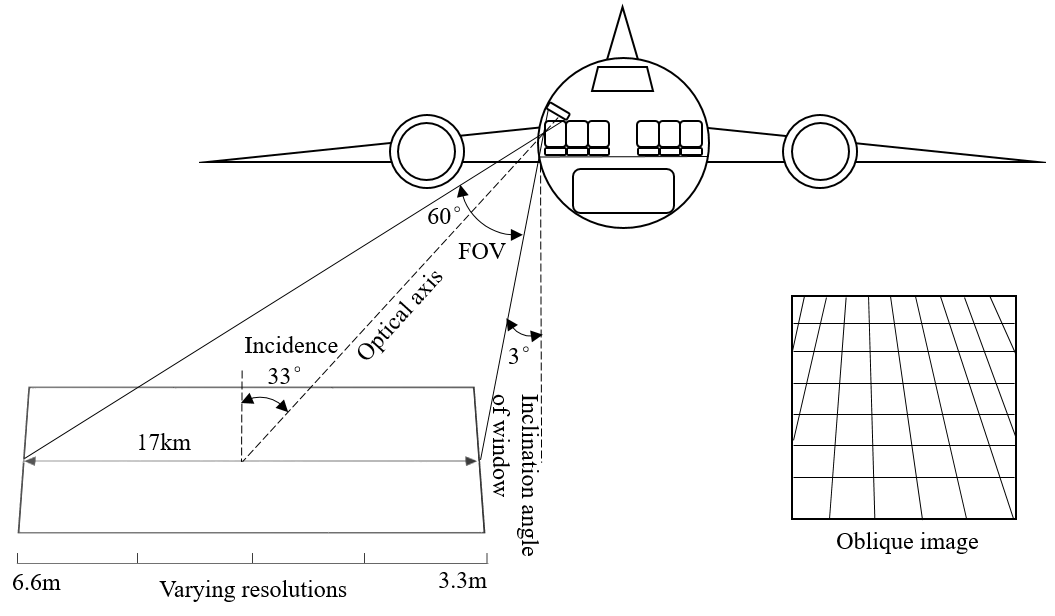


Figure 3. Imaging geometry using a passenger aircraft as the platform

## 2.2. Cloud Detection Model Architecture

Many deep learning architectures based on semantic segmentation methods have recently been introduced, and some of which are used in satellite imagery [36,37]. Three representative methods having an encoder-decoder structure, U-Net, FPN, and PSPNet were trained and tested for suitability for use in the proposed approach. The adopted models were pre-trained on the PASCAL VOC-2012 semantic segmentation dataset [38]. To compare the performance of the different methods objectively, three stages of the cloud detection process were studied. First, the binary cross-entropy function was used to calculate the loss between the predicted and true cloud masks using the different methods. Second, the stochastic gradient descent was selected using momentum [39] as the optimizer in each method's training stage. Third, each method produced a model adapted from the ResNet-101 network [40].

U-Net, which is based on a fully convolutional neural network [41], is built on an encoder-decoder architecture that is comprised of a contracting path to capture context and a symmetric expanding path to enable accurate location. The two main features of the U-Net architecture are the U-shaped network structure and the skip connection. The U-shaped network structure consists of downsampling on the left and symmetric upsampling on the right. To obtain the feature map of the image, the encoder performs feature extraction, which consists of convolution and downsampling, and the fusion method used by the skip connection concatenates on the feature map channels. Finally, the feature map is restored to the original resolution of the corresponding cloud mask based on the downsampling and convolution of the decoder.

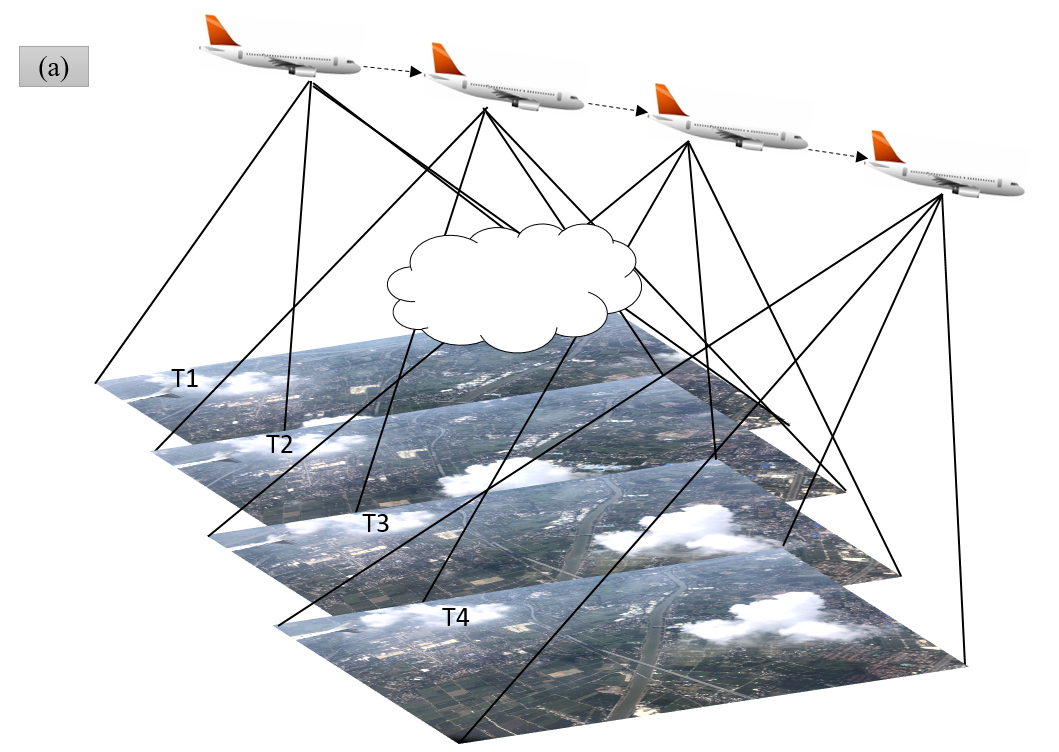
The fully convolutional neural network has been further improved and is now known as the FPN [42], which extracts features at different scales to form a pyramidal hierarchy and efficiently uses the semantic information of the different scales in the FPN. The structure of the FPN can be summarized as feature extraction, upsampling, feature fusion, and multi-scale feature output. The input and output images of the FPN are feature maps of different scales. The FPN architecture is divided into bottom-up and top-down pathways. The bottom-up pathway is a feature encoder process using the ResNet-101 network that is pre-trained on the PASCAL VOC-2012 semantic segmentation dataset. The top-down pathway with upsampling and lateral connections builds high-level semantic feature maps at different scales using the corresponding cloud masks.

The PSPNet architecture also contains an encoder and a decoder. Specifically, the encoder contains a pyramid pooling module and a convolutional neural network [43] backbone with dilated convolutions instead of the fully convolutional neural network. The dilated convolution layers can capture a more receptive field and the pyramid-pooling module is used for capturing the global context from an input image, which helps the PSPNet network to classify the pixels from the global information present in the image. After the encoder features of the image are extracted, the decoder takes the features and converts them into feature representations using upsampling and concatenation layers. Finally, the representation is fed into a convolutional layer to obtain the corresponding per-pixel cloud mask.

## 2.3. Cloud Removal Model

Considering the characteristics of the obtained images, such as the oblique acquisition angle and uneven illumination of the images caused by the different acquisition angles, a traditional photogrammetric method is not suitable for generating the orthophoto. However, the varied viewing angle geometry at which the cloudy and CF images of the same area are obtained is important because CF orthophoto generation is possible by combining these multi-angled photos (e.g., Fig. 4). To overcome obstacles such as the oblique acquisition angle and uneven illumination of the images, a processing procedure for CF orthophoto generation is proposed in this study, which consists of the following four steps:

1. Camera position initialization: Approximate camera positions are necessary for rapid conjugate point detection and georeferencing of the sparse point cloud. Additionally, the cloud masks detected in the VAPRS images are not calculated during conjugate point detection. Because aircraft GPS information is obtained by commercial companies and publicized for flight tracking purposes, the flight GPS positions can be obtained from the flight tracking information. The GPS position of each image can be obtained using piecewise linear interpolation of the GPS position of the flight.
2. Interior orientation parameters initialized from the exchangeable image file header: In addition to the camera positioning information, internal positioning parameters such as focal length and sensor size are required to facilitate conjugate point searching. The interior camera orientation parameters are further refined in Step 3.
3. Structure from motion (SfM) processing: SfM is a 3D reconstruction algorithm based on the initialized GPS information and interior orientation parameters [44]. SfM processing can be performed by many aerial photogrammetry software programs (e.g., Agisoft Metashape, Pix4d, Menci APS, and MicMac). On comparing the performances of these software in processing a large number of images, the results showed that Metashape provides acceptable accuracy and satisfactory computational performance with graphics processing unit acceleration.
4. Orthophoto mosaic generation: A dense point cloud can be generated by a dense image stereo when the SfM stage is completed. Meanwhile, a digital surface model in the form of a raster image can be obtained based on the regular interpolation of the point cloud. The input photos are orthorectified to orthophotos using the individual camera positions and digital surface model. To generate mosaic CF orthophotos, several strategies can be used, including averaging the values of all pixels from the individual photos, taking pixels from the photo observations closest to the normal direction, and using a frequency domain approach.



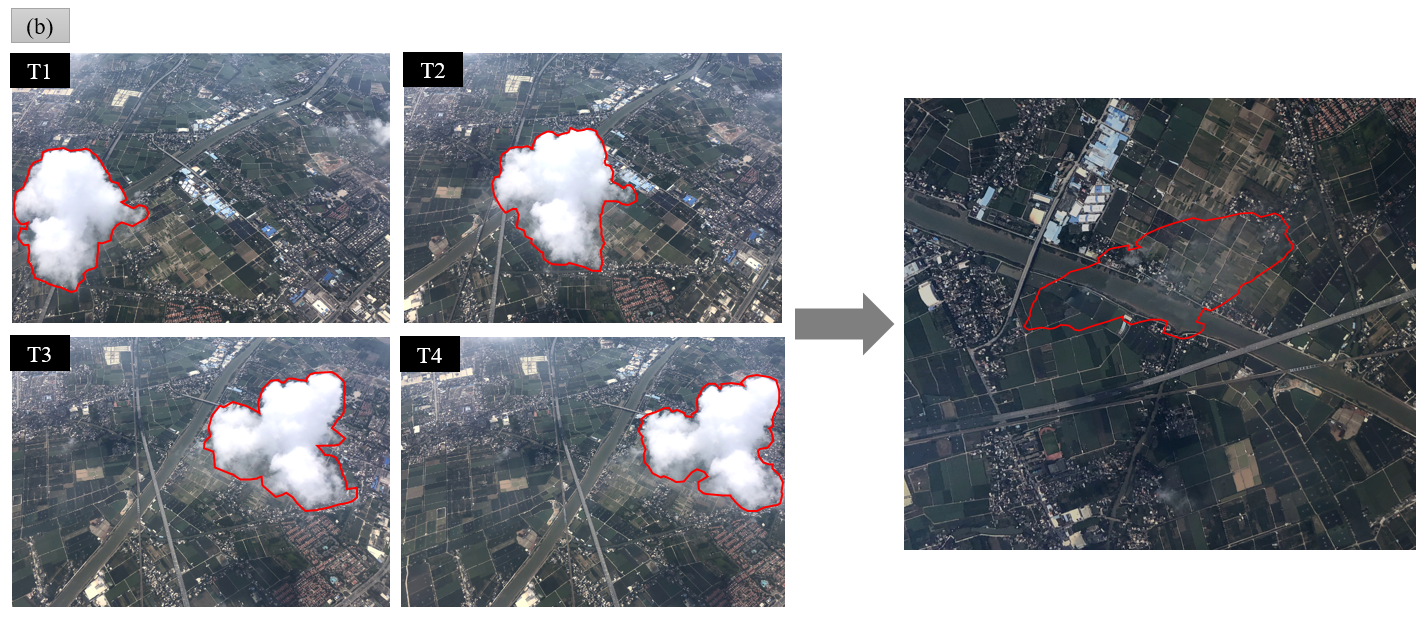


Figure 4. Cloud removal performed using images produced by multiple viewing angles (a) diagram illustrating how time-series images are obtained and (b) an example of cloud removal using images captured at multiple viewing angles.

## 2.4. Haze Removal Model

Haze is a common characteristic in remote optical sensing images. Commercial aircraft typically fly between 9 and 11 km above the surface; therefore, objects observed at this height will be obstructed by different cloud types such as cumulus, altocumulus, and stratus, and haze generally occurs with certain cloud types. Haze is treated as a cloud to generate a larger number of cloud masks; otherwise, a portion of the resultant orthophoto will be missing. In this study, to obtain a complete orthophoto, a haze mask was not used as a cloud mask in the cloud detection process. Instead, a haze removal method was developed that combined the DCP algorithm and histogram statistics based on the characteristics of the orthophoto. The primary reason for employing this method is to use the DCP algorithm to remove haze from the orthophoto and then use the histogram statistics to restore the color of the HF orthophoto.

The DCP algorithm [34] indicated that at least one color channel has very low intensity or approaches 0 for some pixels in most of the non-sky patches of the image. This algorithm can be divided into three steps. First, the dark channel is calculated and the atmospheric light intensity is estimated. Subsequently, a refined transmission function with soft matting is obtained. Finally, the atmospheric light intensity is calculated. The HF image can be reconstructed based on the unknown values calculated. To restore the color of the orthophoto, the histogram method was used to enhance the HF orthophoto quality.

# 3. Experimental Data

The experimental data were divided into two parts. The first part was used for training and testing based on deep learning methods. To obtain suitable cloud detection models in specific scenes, a new dataset was obtained from passenger aircraft platforms that included 2 sets of 20 images for training and 12 images to be used as test images. The cloud mask delineation procedure was performed using 32 images. First, the cloudy image was stretched into a proper visual contrast using Adobe Photoshop. Subsequently, the magic wand tool was used to mark the cloudy locations in the image. Finally, a manually labeled reference mask was generated by assigning the cloud and CF pixel values of 0 and 255, respectively [45]. To expedite this process, the images were input into a cloud detection model network, and multiple non-overlapping patches and the corresponding cloud mask were extracted from each of the images. Overall, 3788 and 2552 training and testing patches were generated, respectively.

The second part of the dataset was used to predict cloud masks and then the corresponding orthophoto was generated. To evaluate the quality of orthophotos generated under different environmental conditions, the prediction dataset from three representative scenes was selected based on cloud type, haze density, and vegetation coverage, and images for two of the scenes were captured during flight CZ6591 on June 23, 2019. During this flight, a set of 17 time-series images of Shenzhen City (herein referred to as scene1) and a set of 12 images of Huizhou City (herein referred to as scene2) were taken at a 5 s time interval between adjacent images. During flight CZ3192 on August 5, 2018, a set of 10 time-series images were taken of Guangzhou City (herein referred to as scene3). Evaluating the quality of the CF orthophotos based on the various cloud types present at the time they were taken facilitated an understanding of the effect of cloud characteristics on CF orthophotos. Haze is frequently found in remote sensing images; therefore, haze density was considered during this evaluation, which contributed to obtaining comprehensive cloud removal knowledge in a specific scene. The spectral feature is one of the most important features of cloud detection [46]. The normalized green–red difference index (NGRDI = [Green DN – Red DN]/[Green DN + Red DN]), with values ranging between -1 and 1, was selected as the spectral feature [47] in this study because the image channels only had red, green, and blue bands. Based on the mean NGRDI threshold of the image, the representative scenes were classified into three categories: low-level vegetation cover (LVC), medium-level vegetation cover (MVC), and high-level vegetation cover (HVC) with thresholds of (-1, -0.1), (-0.1, 0.2), and (0.2, 1.0), respectively. Scene1, scene2, and scene3 had average NGRDI values of -0.23, 0.36, and 0.07 and were therefore classified as LVC, HVC, and MVC, respectively.

# 4. Evaluation Metrics

To evaluate the performance of the cloud detection results and image quality after haze removal objectively, different quantitative indicators in the two sections were selected.

Because the cloud masks obtained via U-Net, FPN, and PSPNet used different datasets, the predicted masks were compared against the corresponding ground truth masks and classified as "cloud" (positive) or "clear" (negative). The datasets were evaluated quantitatively based on the metrics of accuracy, recall, precision, F-score, and Jaccard index. Accuracy is defined as the percentage of accurately predicted masks in the total sample, which can be used as an indicator for evaluating the accuracy of different models. However, accuracy is not an objective indicator to evaluate the performance of models when the data types are unbalanced. Precision is defined as the number of classified clouds that were literally clouds. Recall is defined as the number of cloud pixels that were classified. The F-score provides insight into the optimum balance between recall and precision, and the Jaccard index is a measure of the similarity between the truth masks and predicted masks [48-50]. These five metrics are calculated as follows:

where *tp, tn, fp*, and *fn* are the numbers of true positive, true negative, false positive, and false negative pixels in each test image, respectively, and M denotes the total number of images in each test dataset.

In this study, two full-reference metrics [51] and no-reference image quality assessment (IQA) models were used to fully evaluate the HF orthophotos. The full-reference metrics used were the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM). The PSNR is calculated by the error of the corresponding pixels, and a large PSNR value indicates small distortion [52,53]. The SSIM is used to measure image similarity based on brightness, contrast, and structure [54], and a large SSIM value indicates less image distortion. The no-reference image evaluation method was the blind/referenceless image spatial quality evaluator (BRISQUE) in which the mean subtracted contrast normalized coefficients, and neighborhood coefficients are fitted with the generalized and asymmetric generalized Gaussian distribution models, and then the image quality is evaluated using these model parameters [55]. The two full-reference IQA metrics are defined as follows:

where and represent the HF and dehazed images, respectively.

where and are the averages of and , respectively; is the average variance of is the variance of . is the variance of . is the covariance with and and are constants used to avoid system instability caused by a denominator of 0.

# 5. Results and Discussion

## 5.1. Cloud Detection Results

The cloud detection quantitative accuracy evaluation results are presented in Table 1, and the qualitative evaluation of the cloud masks generated based on various cloud detection methods is presented in Fig. 5.

Among the three ground vegetation cover types, the evaluation indices of the three cloud detection models were highest with MVC. According to the results in Fig. 5a, the strong contrast between the clouds and the surrounding objects shows that in certain areas, thick regularly shaped clouds completely occluded ground objects, making it easy for the model to mask the clouds. Additionally, the haze densities in the sky were low. These environmental factors provided good conditions for the cloud detection model architectures to learn local and global features from the MVC image. Table 1 shows that all the cloud detection models performed well in each evaluation index, with PSPNet achieving the highest accuracy, recall, F-score, and Jaccard values and U-Net achieving the highest precision value (96.71%). For the same evaluation index, the difference between the results of the methods was less than 0.3%, which means all the methods had high cloud detection accuracy in MVC.

Table 1 Evaluation results for the cloud detection methods of percent accuracy of predicted data for the three scenes. Bold values denote the highest accuracy for each scene among the different cloud detection methods.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Dataset (no. scenes) | Method | Overall acc. | Recall | Precision | F-score | Jaccard |
| MVC (10 images) | U-Net | 98.87 | 97.48 | **96.71** | 97.09 | 94.44 |
| FPN | 98.84 | 97.57 | 96.49 | 97.02 | 94.32 |
| PSPNet | **98.89** | **97.74** | 96.58 | **97.15** | **94.56** |
| LVC (17 images) | U-Net | 97.80 | 94.08 | **95.25** | 94.65 | 90.14 |
| FPN | **97.81** | **94.19** | 95.17 | **94.67** | **90.17** |
| PSPNet | 97.67 | 93.81 | 94.89 | 94.34 | 89.61 |
| HVC (12 images) | U-Net | 94.66 | 91.71 | 92.29 | 92.00 | 85.51 |
| FPN | 94.97 | **91.96** | 92.92 | 92.43 | 86.24 |
| PSPNet | **95.17** | 91.51 | **93.87** | **92.62** | **86.56** |

Fig. 5b shows that the LVC scene is primarily covered by urban areas, has minimal vegetation, and the radiation intensities of some of the buildings and clouds are similar; therefore, the contrast between them is poor. Additionally, the cloud boundaries are blurred, and the thin cloud cover is fragmented. These environmental factors create difficulties for cloud detection models to differentiate between clouds and ground objects in the extraction of global and local features. The haze density in the sky, however, is the lowest among the three scenes, which contributes positively to training the model architecture. In the LVC scene, FPN achieved the highest values for each evaluation index, except precision, which was only 95.17%. The highest precision value of 92.25% was achieved by U-Net. Notably, the difference in the calculation results of the various methods was less than 0.5% when comparing the same evaluation index, indicating that the three cloud detection models showed good consistency with regard to the various accuracy evaluation results.

Most areas in the HVC scene (Fig. 5c) were covered by natural vegetation, except for the river basin, and the high-density vegetation contributed to a strong contrast between ground objects and clouds, which is conducive to accurate cloud detection. However, haze densities were high, which reduced the contrast between the clouds and other features, and there were a large number of fragmented clouds, which caused difficulties in determining the thin cloud boundaries. Additionally, several pixels containing both ground objects and extremely thin clouds were semi-transparent. These environmental factors are extremely unfavorable for thin cloud detection. PSPNet achieved the highest accuracy, recall, F-score, and Jaccard values, and FPN achieved the highest precision of 91.96%. Based on these results, the cloud detection accuracy was the lowest for the HVC case, and the differences in evaluation index values between the three methods were greater than 1%, which shows that the performance of the three methods was not stable compared to the performance of the LVC and MVC scenes.

According to the above results, we conclude that cloud detection performance is significantly influenced by specific environmental conditions (e.g., cloud characteristics and the haze and vegetation density).

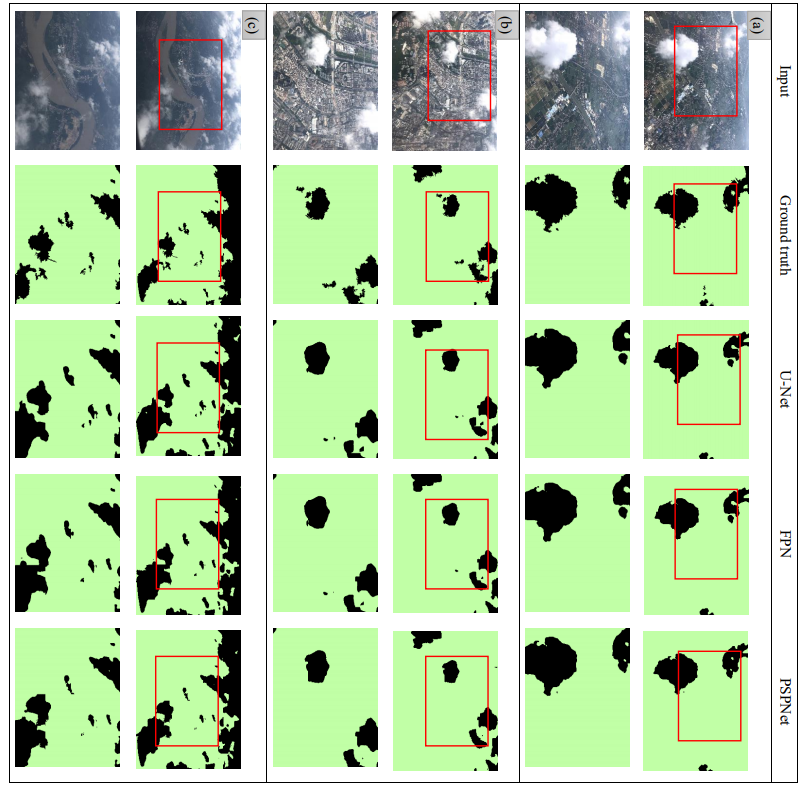


Figure 5. Examples of cloud and cloud mask detection results using the three cloud detection methods on the three different scenes (a) MVC, (b) LVC, and (c) HVC

## 5.2. CF Orthophoto Results

To obtain a better understanding of the spatial visualization of the HF orthophotos using the different cloud detection methods, orthophotos of the MVC, LVC, and HVC scenes were used for subjective visual evaluation (Fig. 6).

In the MVC scene, most of the cloud boundaries were distinct, and the contrast between ground objects and clouds was evident (Fig. 5a). These factors are favorable for the extraction of cloud features using deep-learning-based cloud detection methods; therefore, highly accurate cloud mask results were generated by the three cloud detection methods, and the evaluation results were similar (Table 1). The quantitative results were consistent with what can be observed, although a few thin clouds remained in the orthophotos (yellow circles in Fig. 6a) because they were difficult to separate from the surrounding objects. Moreover, some pixels containing both clouds and ground objects were semi-transparent. The generated orthophoto would contain fewer thin clouds if the semi-transparent pixels had been classified as non-clouds for the training and testing procedures of the cloud detection methods. If the semi-transparent pixels had been classified as clouds, the orthophoto would return no values in these areas. To obtain a comprehensive orthophoto, some semi-transparent pixels had to be classified as non-clouds in this study.

For the LVC scene, the high reflectivity of buildings reduced the contrast between the clouds and ground objects and blurred some thin cloud boundaries (Fig. 5b). The cloud detection accuracy was lower than the MVC scene, and the results of PSPNet were not as accurate as other cloud detection methods (Table 1). The differences between the methods are evident from the visual differences in the orthophoto subset (yellow circles in Fig 6b). A few blurry thin clouds found in the subset are associated with PSPNet, which classified many semi-transparent pixels as non-clouds (Fig. 5b). Therefore, the U-Net and FPN cloud detection results, which were similar, were better than the PSPNet results because they classified the same subset area as no-cloud. This indicates the high accuracy of the cloud detection results, which was consistent with the visual orthophotos without clouds in the LVC scene.

The HVC scene had the highest haze densities compared with other scenes, and the cloud fragmentation was the highest and had many semi-transparent pixels (Fig. 5c). These specific factors are detrimental to accurately capturing cloud features; hence, the performances of the models were the lowest among the scenes (Table 1). The high incidence of semi-transparent pixels increased the uncertainty when matching feature points between two adjacent images, and due to the high haze density, the visual performance of the HVC orthophoto was worse than those of the MVC and LVC orthophotos (Fig. 6). There were significant differences in the quantitative evaluation metrics produced by the three cloud detection methods and PSPNet performed better than U-Net and FPN. Additionally, the visual performance of the orthophoto subset (yellow circles in Fig. 6c) could not be easily distinguished by the cloud detection methods.

To clarify the advantages of cloud removal using passenger aircraft as platforms in cloudy weather, we created a visualized map based on a Landsat 8 scene and obtained two CF orthophotos during flight CZ6591, wherein the Landsat 8 scene overlaid the orthophotos (Fig. 7a). Clouds in the Landsat 8 scene completely covered the LVC and surrounding areas (Fig. 7b). Under these conditions, the contaminated area cannot be reconstructed using spatial-based and multispectral approaches. Spatial-based methods require a hypothetical relationship between cloudy and CF pixels [18,19], and semi-transparent cloud or haze conditions are required for using multispectral methods [15,17]. Landsat 8 images only covered half of the HVC areas. Such conditions limit the use of spatial-based and multispectral methods, as well as multitemporal approaches with a reference CF image [13]. This is because additional CF images are taken from the same satellite at the same position and recovered after at least 1 revisit period. It is difficult to meet the requirements necessary to perform quality dynamic monitoring of the Earth’s surface activities. Deep-learning-based approaches [22,23] used to generate CF images encounter uncertainties during the learning process if a satellite image is completely contaminated by clouds; therefore, only using cloud removal algorithms on fully contaminated satellite images obtained by optical sensors on remote sensing platforms cannot generate truly CF images.

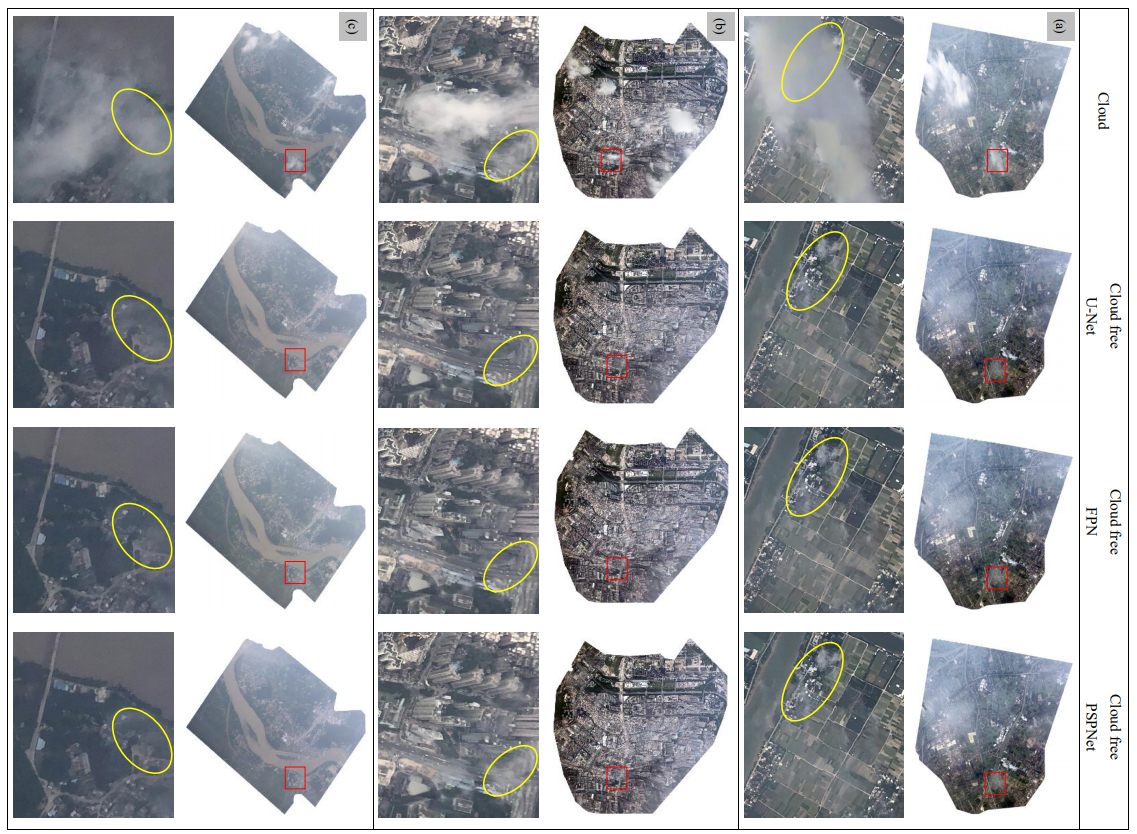


Figure 6. Qualitative comparison of CF orthophotos generated by different cloud detection methods for various scenes (a) MVC, (b) LVC, and (c) HVC. The cloudy and corresponding CF locations are indicated by yellow circles.

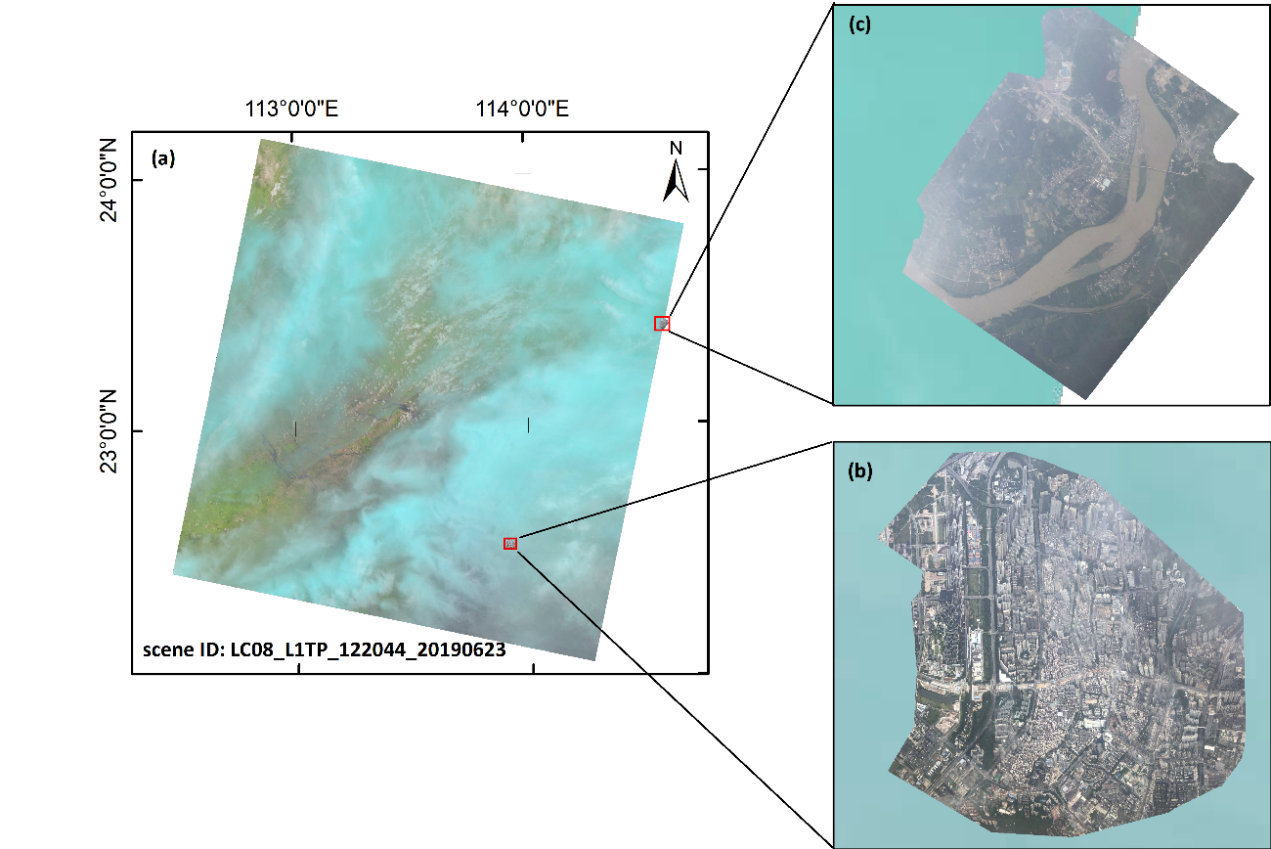


Figure 7. Two orthophotos from flight CZ6591 and the overlaid Landsat 8 scene taken on the same day. Zoom images of the (a) LVC and (b) HVC areas

## 5.3. HF Orthophoto Quality Assessments

In this section, to evaluate the image quality of HF orthophotos objectively and comprehensively, HF and CF orthophoto subsets (obtained via the haze and cloud removal methods, respectively) of the MVC, LVC, and HVC scenes, using the different cloud detection methods, were compared. The results are presented in Fig. 6, and the corresponding quantitative results comparing the three described IQA metrics are shown in Table 2.

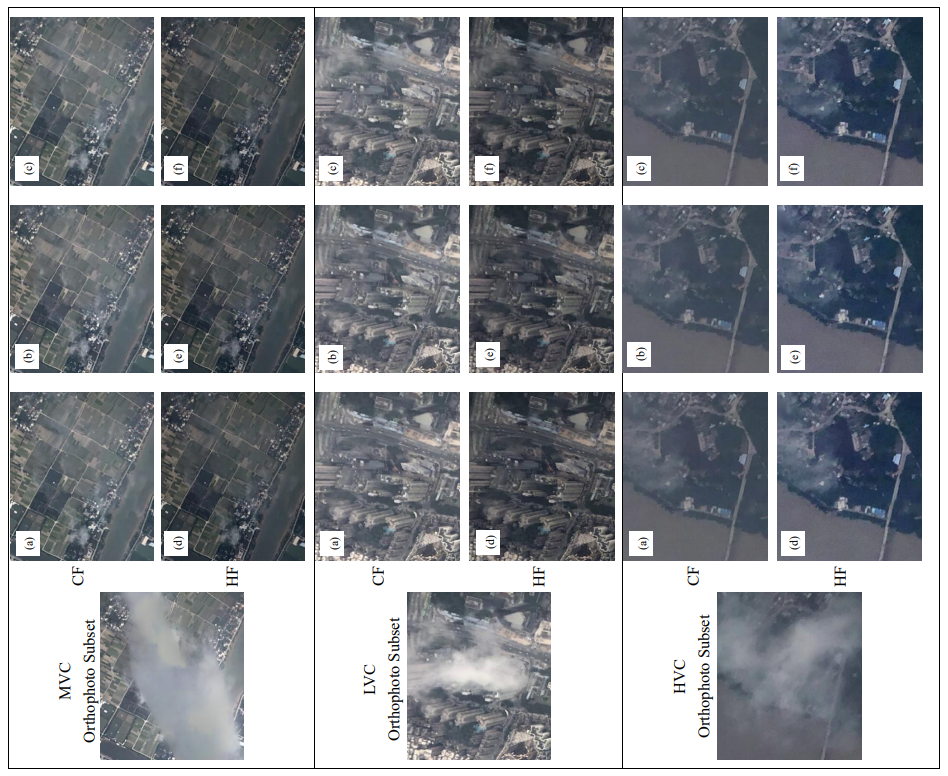


Figure 8. Qualitative comparison of the CF and HF orthophoto results obtained by the cloud and haze removal methods, respectively, of the MVC, LVC, and HVC scenes using the different cloud detection methods. The cloud removal results using (a) U-Net, (b) FPN, and (c) PSPNet and the haze removal results using (d) U-Net, (e) FPN, and (f) PSPNet.

The visual examples of the three scenes show that the details of the HF orthophoto subset are more refined and display ground objects more clearly. No significant differences in the HF results could be observed between the different cloud detection methods. The results of the CF orthophoto subset are consistent with those of the HF results. By comparing the different scenes, better image quality enhancement occurred in the HVC scene compared with the other two scenes.

The visual results are consistent with the statistical results, where all the IQA metrics values are similar for the same scene. For the PSNR and SSIM metrics, the MVC orthophoto subset yielded the best results, which indicates that the CF and HF data are the closest and have the lowest haze densities. The PSNR and SSIM results were better in the LVC scene than the HVC scene, indicating that the haze densities in the LVC scene were lower than those in the HVC scene, which concurs with the visual results between these scenes (Fig. 8). Obvious differences in the PSNR and SSIM metrics existed between the PSPNet results and the U-Net and FPN results. This is attributed to fewer thin clouds in the orthophoto subset generated by PSPNet in the LVC scene. According to the no-reference IQA results, the MVC data yielded the best BRISQUE results that were similar to the corresponding HVC results, while the statistical results of the LVC were the worst. Because the BRISQUE was used to generate the HF orthophoto without reference to the CF orthophoto, the shapes of the ground objects were simple and regular with few color distortions in the MCV areas and, the BRISQUE had excellent performance. In the LVC scene, the buildings made feature extraction difficult, the generated orthophoto was slightly distorted, and a large number of shadows were present in the scene. Consequently, the BRISQUE result was worse for the LVC scene than those for the other scenes. Based on the statistical results, the image quality of the HF orthophoto strongly affects the quality of the generated CF orthophoto.

Table 2 Quantitative comparison of orthophoto IQA metrics for the LVC, MVC, and HVC scenes by different cloud detection methods based on the haze removal method. The bold values denote the highest IQA metrics.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Scenes | Methods | PSNR | SSIM | BRISQUE |
| MVC | U-Net | 25.569 | 0.951 | 30.012 |
| FPN | 25.388 | 0.939 | 30.023 |
| PSPNet | **25.621** | **0.962** | **30.437** |
| LVC | U-Net | 18.122 | 0.826 | 35.481 |
| FPN | **18.297** | **0.889** | **35.053** |
| PSPNet | 15.146 | 0.688 | 37.142 |
| HVC | U-Net | 15.195 | 0.859 | 31.782 |
| FPN | 15.241 | 0.865 | 31.182 |
| PSPNet | **15.460** | **0.890** | **31.039** |

# 6. Conclusion

Optical satellites are inevitably hindered by clouds when obtaining remote sensing images. Traditional cloud removal methods have specific preconditions, and deep-learning-based cloud removal methods generate predicted CF imagery with uncertainties caused by contaminated cloud imagery. The issue of cloud coverage in satellite imagery is associated with the altitude and FOV of the satellite platform. The extent of cloud contamination in satellite imagery is more evident as the satellite’s altitude increases. However, the optical sensor on the satellite platform captures the surface information at the same angle due to its fixed FOV, causing information located in the same place to be contaminated by clouds in time-series remote sensing data. Therefore, generating a truly CF image from a fully contaminated image is difficult using only cloud removal approaches.

Compared with the FOV and altitude limitations of the satellite platform, the passenger aircraft platform has the advantages of suitable altitude and multi-viewing angles to capture time-series images in cloudy conditions. In this study, an automatic framework was presented to generate a CF orthophoto using passenger aircraft as the remote sensing platform. The proposed method can combine images from multiple viewing angles and remove cloud contamination without distinguishing cloud types or the need for a reference image. This study presented the cloud removal results from three representative scenes using the proposed framework, which considered cloud type, haze density, and vegetation coverage. The clouds in the orthophotos of the different scenes were effectively removed. The image quality of the CF orthophotos was highly associated with the accuracy of detected cloud masks, and the performances of the cloud detection models were significantly influenced by cloud characteristics, haze density, and image contrast. For the same scene, there were no obvious differences in the detected cloud masks generated by the three cloud detection methods. Subsequently, the CF orthophoto was processed by the haze removal model to yield a high-quality HF orthophoto. This demonstrates that the framework can create high quality, truly CF orthophotos and can be highly beneficial when used in conjunction with cloud removal methods for satellite imagery.

# Acknowledgments

This work is supported by National Key Research and Development Project (2020YFC1512001), National Natural Science Foundation of China (41974006), Shenzhen Scientific Research and Development Funding Program (KQJSCX20180328093453763, JCYJ20180305125101282, JCYJ20170412142239369), and Department of Education of Guangdong Province (2018KTSCX196).

**References**

1. Kalisa, W.; Igbawua, T.; Henchiri, M.; Ali, S.; Zhang, S.; Bai, Y.; Zhang, J. Assessment of climate impact on vegetation dynamics over East Africa from 1982 to 2015. *Scientific Reports* **2019**, *9* 16865.

2. Fisher, A.; Flood, N.; Danaher, T. Comparing Landsat water index methods for automated water classification in eastern Australia. *Remote Sensing of Environment* **2016**, *175* 167-182.

3. Zhang, C.; Harrison, P.A.; Pan, X.; Li, H.; Sargent, I.; Atkinson, P.M. Scale Sequence Joint Deep Learning (SS-JDL) for land use and land cover classification. *Remote Sensing of Environment* **2020**, *237* 111593.

4. Asokan, A.; Anitha, J. Change detection techniques for remote sensing applications: a survey. *Earth Science Informatics* **2019**, *12* 143-160.

5. Wei, J.; Li, Z.; Guo, J.; Sun, L.; Huang, W.; Xue, W.; Fan, T.; Cribb, M. Satellite-Derived 1-km-Resolution PM1 Concentrations from 2014 to 2018 across China. *Environmental Science & Technology* **2019**, *53* 13265-13274.

6. Wei, J.; Huang, W.; Li, Z.; Xue, W.; Peng, Y.; Sun, L.; Cribb, M. Estimating 1-km-resolution PM2.5 concentrations across China using the space-time random forest approach. *Remote Sensing of Environment* **2019**, *231* 111221.

7. Bai, T.; Li, D.; Sun, K.; Chen, Y.; Wenzhuo, L. Cloud Detection for High-Resolution Satellite Imagery Using Machine Learning and Multi-Feature Fusion. *Remote Sensing* **2016**, *8* 715.

8. Mahajan, S.; Fataniya, B. Cloud detection methodologies: variants and development—a review. *Complex & Intelligent Systems* **2019**, 10.1007/s40747-019-00128-0.

9. Ju, J.; Roy, D.P. The availability of cloud-free Landsat ETM+ data over the conterminous United States and globally. *Remote Sensing of Environment* **2008**, *112* 1196-1211.

10. King, M.D.; Platnick, S.; Menzel, W.P.; Ackerman, S.A.; Hubanks, P.A. Spatial and Temporal Distribution of Clouds Observed by MODIS Onboard the Terra and Aqua Satellites. *IEEE Transactions on Geoscience and Remote Sensing* **2013**, *51* 3826-3852.

11. Lin, C.; Tsai, P.; Lai, K.; Chen, J. Cloud Removal From Multitemporal Satellite Images Using Information Cloning. *IEEE Transactions on Geoscience and Remote Sensing* **2013**, *51* 232-241.

12. Xu, M.; Jia, X.; Pickering, M.; Plaza, A.J. Cloud Removal Based on Sparse Representation via Multitemporal Dictionary Learning. *IEEE Transactions on Geoscience and Remote Sensing* **2016**, *54* 2998-3006.

13. Li, X.; Wang, L.; Cheng, Q.; Wu, P.; Gan, W.; Fang, L. Cloud removal in remote sensing images using nonnegative matrix factorization and error correction. *ISPRS Journal of Photogrammetry and Remote Sensing* **2019**, *148* 103-113.

14. Hu, G.; Li, X.; Liang, D. Thin cloud removal from remote sensing images using multidirectional dual tree complex wavelet transform and transfer least square support vector regression. **2015**, *9 %J Journal of Applied Remote Sensing* 095053.

15. Xu, M.; Pickering, M.; Plaza, A.J.; Jia, X. Thin Cloud Removal Based on Signal Transmission Principles and Spectral Mixture Analysis. *IEEE Transactions on Geoscience and Remote Sensing* **2016**, *54* 1659-1669.

16. Xu, M.; Jia, X.; Pickering, M.; Jia, S. Thin cloud removal from optical remote sensing images using the noise-adjusted principal components transform. *ISPRS Journal of Photogrammetry and Remote Sensing* **2019**, *149* 215-225.

17. Cheng, Q.; Shen, H.; Zhang, L.; Yuan, Q.; Zeng, C. Cloud removal for remotely sensed images by similar pixel replacement guided with a spatio-temporal MRF model. *ISPRS Journal of Photogrammetry and Remote Sensing* **2014**, *92* 54-68.

18. Lorenzi, L.; Melgani, F.; Mercier, G. Missing-Area Reconstruction in Multispectral Images Under a Compressive Sensing Perspective. *IEEE Transactions on Geoscience and Remote Sensing* **2013**, *51* 3998-4008.

19. Meng, F.; Yang, X.; Zhou, C.; Li, Z. A Sparse Dictionary Learning-Based Adaptive Patch Inpainting Method for Thick Clouds Removal from High-Spatial Resolution Remote Sensing Imagery. **2017**, *17* 2130.

20. Zhang, Q.; Yuan, Q.; Zeng, C.; Li, X.; Wei, Y. Missing Data Reconstruction in Remote Sensing Image With a Unified Spatial–Temporal–Spectral Deep Convolutional Neural Network. *IEEE Transactions on Geoscience and Remote Sensing* **2018**, *56* 4274-4288.

21. Grohnfeldt, C.; Schmitt, M.; Zhu, X. A Conditional Generative Adversarial Network to Fuse Sar And Multispectral Optical Data For Cloud Removal From Sentinel-2 Images. In Proceedings of IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, 22-27 July 2018; pp. 1726-1729.

22. Bermudez, J.D.; Happ, P.N.; Feitosa, R.Q.; Oliveira, D.A.B. Synthesis of Multispectral Optical Images From SAR/Optical Multitemporal Data Using Conditional Generative Adversarial Networks. *IEEE Geoscience and Remote Sensing Letters* **2019**, *16* 1220-1224.

23. Meraner, A.; Ebel, P.; Zhu, X.X.; Schmitt, M. Cloud removal in Sentinel-2 imagery using a deep residual neural network and SAR-optical data fusion. *ISPRS Journal of Photogrammetry and Remote Sensing* **2020**, *166* 333-346.

24. Wang, C.; Ke, J.; Xiu, W.; Ye, K.; Li, Q. Emergency Response Using Volunteered Passenger Aircraft Remote Sensing Data: A Case Study on Flood Damage Mapping. *Sensors (Basel)* **2019**, *19* 4163.

25. Wang, C.; Wang, Y.; Wang, L.; Hu, Z.; Zhang, S.; Wang, S.; Xiu, W.; Cui, H.; Wang, D.; Li, Q. Volunteered remote sensing data generation with air passengers as sensors. *International Journal of Digital Earth* **2020**, 10.1080/17538947.2020.18087211-23.

26. Mastelic, T.; Lorincz, J.; Ivandic, I.; Boban, M. Aerial Imagery Based on Commercial Flights as Remote Sensing Platform. **2020**, *20* 1658.

27. Dehnavi, S.; Maghsoudi, Y.; Zakšek, K.; Valadan Zoej, M.J.; Seckmeyer, G.; Skripachev, V. Cloud Detection Based on High Resolution Stereo Pairs of the Geostationary Meteosat Images. **2020**, *12* 371.

28. Fleming, R.J. The Use of Commercial Aircraft as Platforms for Environmental Measurements. *Bulletin of the American Meteorological Society* **1996**, *77* 2229-2242.

29. Chen, Y. COVID-19 Pandemic Imperils Weather Forecast. **2020**, *47* e2020GL088613.

30. Maturing computer vision and connectivity solutions within swarm networks for autonomous Earth Observation and GIS. [*https://acubed.airbus.com/projects/ray-20/*](https://acubed.airbus.com/projects/ray-20/)**2020**.

31. Norway: World’s first passenger aircraft for environmental monitoring. [*https://www.norceresearch.no/en/insight/verdens-forste-passasjerfly-for-miljoovervaking-er-norsk*](https://www.norceresearch.no/en/insight/verdens-forste-passasjerfly-for-miljoovervaking-er-norsk)**2019**.

32. AXA and ANA launch Joint Research on Remote Sensing Observation of Atmospheric Components using Passenger Aircraft and Satellites

-Contributing global warming measures from sky and space-. [*https://global.jaxa.jp/press/2020/09/20200928-1\_e.html*](https://global.jaxa.jp/press/2020/09/20200928-1_e.html)**2020**.

33. He, K.; Sun, J.; Tang, X. Single Image Haze Removal Using Dark Channel Prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2011**, *33* 2341-2353.

34. Verykokou, S.; Ioannidis, C. Oblique aerial images: a review focusing on georeferencing procedures. *International Journal of Remote Sensing* **2018**, *39* 3452-3496.

35. Jeppesen, J.H.; Jacobsen, R.H.; Inceoglu, F.; Toftegaard, T.S. A cloud detection algorithm for satellite imagery based on deep learning. *Remote Sensing of Environment* **2019**, *229* 247-259.

36. Liu, C.-C.; Zhang, Y.-C.; Chen, P.-Y.; Lai, C.-C.; Chen, Y.-H.; Cheng, J.-H.; Ko, M.-H. Clouds Classification from Sentinel-2 Imagery with Deep Residual Learning and Semantic Image Segmentation. *Remote Sensing* **2019**, *11*.

37. Everingham, M.; Eslami, S.M.A.; Van Gool, L.; Williams, C.K.I.; Winn, J.; Zisserman, A. The Pascal Visual Object Classes Challenge: A Retrospective. *International Journal of Computer Vision* **2015**, *111* 98-136.

38. Ruder, S. An overview of gradient descent optimization algorithms. **2016**.

39. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 27-30 June 2016; pp. 770-778.

40. Shelhamer, E.; Long, J.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2017**, *39* 640-651.

41. Lin, T.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 21-26 July 2017; pp. 936-944.

42. LeCun, Y.; Boser, B.; Denker, J.S.; Henderson, D.; Howard, R.E.; Hubbard, W.; Jackel, L.D. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation* **1989**, *1* 541-551.

43. Westoby, M.J.; Brasington, J.; Glasser, N.F.; Hambrey, M.J.; Reynolds, J.M. ‘Structure-from-Motion’ photogrammetry: A low-cost, effective tool for geoscience applications. *Geomorphology* **2012**, *179* 300-314.

44. Li, Z.; Shen, H.; Cheng, Q.; Liu, Y.; You, S.; He, Z. Deep learning based cloud detection for medium and high resolution remote sensing images of different sensors. *ISPRS Journal of Photogrammetry and Remote Sensing* **2019**, *150* 197-212.

45. Xiong, Q.; Wang, Y.; Liu, D.; Ye, S.; Du, Z.; Liu, W.; Huang, J.; Su, W.; Zhu, D.; Yao, X., et al. A Cloud Detection Approach Based on Hybrid Multispectral Features with Dynamic Thresholds for GF-1 Remote Sensing Images. **2020**, *12* 450.

46. Hunt, E.R.; Cavigelli, M.; Daughtry, C.S.T.; McMurtrey, J.E.; Walthall, C.L. Evaluation of Digital Photography from Model Aircraft for Remote Sensing of Crop Biomass and Nitrogen Status. *Precision Agriculture* **2005**, *6* 359-378.

47. Mohajerani, S.; Krammer, T.A.; Saeedi, P. A Cloud Detection Algorithm for Remote Sensing Images Using Fully Convolutional Neural Networks. In Proceedings of 2018 IEEE 20th International Workshop on Multimedia Signal Processing (MMSP), 29-31 Aug. 2018; pp. 1-5.

48. Mohajerani, S.; Saeedi, P. CPNet: A Context Preserver Convolutional Neural Network for Detecting Shadows in Single RGB Images. In Proceedings of 2018 IEEE 20th International Workshop on Multimedia Signal Processing (MMSP), 29-31 Aug. 2018; pp. 1-5.

49. Mohajerani, S.; Saeedi, P. Cloud-Net: An End-To-End Cloud Detection Algorithm for Landsat 8 Imagery. In Proceedings of IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, 28 July-2 Aug. 2019; pp. 1029-1032.

50. Avcibas, I.; Sankur, B.; Sayood, K. Statistical evaluation of quality measures. J Electron Imaging. *J. Electronic Imaging* **2002**, *11* 206-223.

51. Jobson, D.; Rahman, Z.-U.; Woodell, G.; Hines, G. A comparison of visual statistics for the image enhancement of FORESITE aerial images with those of major image classes. *Proceedings of SPIE - The International Society for Optical Engineering* **2006**, 10.1117/12.664591.

52. Shao, S.; Guo, Y.; Zhang, Z.; Yuan, H. Single Remote Sensing Multispectral Image Dehazing Based on a Learning Framework. *Mathematical Problems in Engineering* **2019**, *2019* 4131378.

53. Zhou, W.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing* **2004**, *13* 600-612.

54. Mittal, A.; Moorthy, A.K.; Bovik, A.C. No-Reference Image Quality Assessment in the Spatial Domain. *IEEE Transactions on Image Processing* **2012**, *21* 4695-4708.