future internet

MDPI

Article

Post-Processing for Shadow Detection in Drone-Acquired Images Using U-NET

Siti-Aisyah Zali, Shahbe Mat-Desa *, Zarina Che-Embi and Wan-Noorshahida Mohd-Isa

Faculty of Computing and Informatics, Multimedia University, Jalan Multimedia, Cyberjaya 63100, Malaysia; sitiaisyahzali25@gmail.com (S.-A.Z.); zarina.embi@mmu.edu.my (Z.C.-E.); wan.noorshahida.isa@mmu.edu.my (W.-N.M.-I.)

* Correspondence: shahbe@mmu.edu.my; Tel.: +60-3-83125458

Abstract: Shadows in drone images commonly appear in various shapes, sizes, and brightness levels, as the images capture a wide view of scenery under many conditions, such as varied flying height and weather. This property of drone images leads to a major problem when it comes to detecting shadow and causes the presence of noise in the predicted shadow mask. The purpose of this study is to improve shadow detection results by implementing post-processing methods related to automatic thresholding and binary mask refinement. The aim is to discuss how the selected automatic thresholding and two methods of binary mask refinement perform to increase the efficiency and accuracy of shadow detection. The selected automatic thresholding method is Otsu's thresholding, and methods for binary mask refinement are morphological operation and dense CRF. The study shows that the proposed methods achieve an acceptable accuracy of 96.43%.

Keywords: shadow detection; deep learning; U-Net; aerial images; automatic thresholding; binary mask refinement; post-processing

check for updates

Citation: Zali, S.-A.; Mat-Desa, S.; Che-Embi, Z.; Mohd-Isa, W.-N. Post-Processing for Shadow Detection in Drone-Acquired Images Using U-NET. *Future Internet* **2022**, *14*, 231. https://doi.org/10.3390/fi14080231

Academic Editors: Remus Brad and Arpad Gellert

Received: 10 June 2022
Accepted: 14 July 2022
Published: 28 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

As technology has grown in this modern age, small, unmanned aircraft, more widely known as drones, have achieved tremendous growth from being just a mere toy to contributing to solving many problems in various industries. The study of emerging drone technologies and image analysis has been an interesting topic, as it offers opportunities for improvement in numerous disciplines. Detecting shadows is one of the problems that is widely discussed in the field, as shadows are frequently visible in images captured by drones, and hence become obstacles.

Currently, a lot of studies on object detection and tracking in drone images listed the presence of shadows as the reason for errors in their findings [1]. Shadows consistently cause detection inaccuracies in many image vision applications due to developed algorithms failing to tackle shadow properties in images. Thus, this problem remains unsolved, and more efforts are necessary to improve shadow detection, considering high-resolution drone images pose complex shadows. As many researchers treated shadows as obstacles, others benefited from the shadows. The presence of shadow can be utilized in some applications, such as object height estimation through drone images [2]. Unfortunately, this possibility is limited due to the limitations of existing shadow detection methods.

The early implementations of shadow detection in drone images mainly rely on shadow image properties and the physical properties of surroundings. These methods require more effort from researchers to study the suitability of different properties and are prone to errors resulting from technical malfunctions. Solutions using machine learning have been introduced to many computer vision problems, including shadow detection as an alternative to conventional methods. The early solution was to use unsupervised learning, and it was enhanced to supervised learning to increase accuracy. However, there

were still disadvantages in implementing supervised learning, such as the fact that some classifiers might not be flexible enough to categorize a large number of objects from various scenes, which leads to the implementation of deep learning [3]. Several deep learning implementations have been proposed throughout the years, yet many have stated the importance of implementing lightweight networks to increase efficiency. For the study of shadow detection in drone images, another problem arises, which is the lack of drone image datasets with the shadow ground truth. This limits the research on the machine learning methods, as dataset preparation is laborious and time-consuming.

Drone images, which are high resolution and tend to contain shadows of different brightness and sizes, bring up more challenges in developing shadow detection algorithms. One of the issues highlighted in the deep learning approaches was the difficulty of relying on a trial-and-error approach to find the optimum threshold value [4]. An unreliable threshold value always results in noises in predicted shadow masks, consequently leading to errors when removing the shadowed regions.

In the present paper, we implement a deep learning method to solve shadow detection in a dataset of drone images. We use the established algorithm in machine learning that requires minimal knowledge of shadow or physical properties. Furthermore, given the complicated shadows which are common in drone images, post-processing strategies for handling the inefficiency of finding optimal threshold values and refining the shadow mask are also proposed. In addition, drone images from various sources are collected and labeled due to the lack of shadow annotated datasets to be used in this study. The aim is to widen the study of deep learning methods specifically for shadow detection in aerial images. The contributions of our work are summarized as follows:

- Implementation of U-Net architecture with a small number of parameters for shadow detection using a small-sized aerial images dataset.
- Implementation of an automatic thresholding method to increase efficiency in obtaining final binary mask prediction.
- Implementation of morphological operations and dense Conditional Random Field (CRF) methods for binary mask refinement.
- Creating an annotated shadow image dataset using existing aerial images from several online sources.

This paper is organized as follows: Section 2 discusses the related studies; Section 3 describes the methodology; while Section 4 discusses the results. Section 5 outlines conclusions and future works.

2. Related Studies

2.1. Shadows in Drone Images

By definition, a shadow is a dark-colored figure of an object that is produced when the object intercepts a ray of light. There are two major groups of shadows, which are cast shadows and self-cast shadows [5]. Cast shadows can be defined as shadows that are produced by an object onto another object. Meanwhile, self-cast shadows are shadows cast by the object onto itself. Defining the presence of shadow regions in images is a complex problem, as shadows can be further categorized into two other subcategories, umbra and penumbra [6]. Penumbra is the upper region of the shadow produced by a certain object, where only a part of the ray of light is concealed by the object. On the other hand, the shadow is categorized as umbra or the inner shadow region, when the light is fully halted by the object. This shows the characteristics of shadows in an image are uneven or defined by different types of pixel values, which makes it harder to detect the shape of shadows accurately.

Shadow regions often cause drawbacks in many image processing tasks. This problem gives researchers not choice but to include shadow removal as a preprocessing step before going into the main part of the processing such as object detection. The presence of shadow in aerial images is always mentioned to cause errors, as it mirrors the properties of objects. For example, in an attempt to detect drones, an improvement is needed specifically in

removing shadows, as it causes mirror images of the drone that bring confusion to the detector [1]. The same problem also occurred in road traffic detection, such that duplicate images detected from the shadow have become the limitation of the tested method [7]. Another example is weed detection in agriculture fields; the shadows of weeds have caused a high number of errors [8]. Other than removing shadows to enhance the visual representation of images, shadow detection also helps to define shadow properties, which benefits some tasks such as building height estimation [2]. The need for shadow detection has become more significant as drone usage has been popularized. Many industries have used drones to facilitate their operations in different kinds of surroundings, including indoor, industrial, cities, forestry, and farming areas. Drone images commonly contain complicated shadows compared to ground-based images, as they are captured across various heights and times, which result in shadows with different sizes and levels of brightness. Thus, a method used for shadow detection in ground-based images may not be suitable to detect shadows in aerial images. This further proves that it is crucial to explore a method that is flexible enough for all types of images, which is also a big challenge in this problem.

2.2. Shadow Detection Methods in Drone Images

There has been a lot of research using various techniques to detect shadows as the importance of shadow detection in drone images has become more significant. Based on various research studies [9–11], shadow detection methods in drone images can be divided into three categories, which are property-based methods, model-based methods, and machine learning methods.

2.2.1. Property-Based

Property-based methods mainly focus on detecting shadows based on shadow properties in images. Considering that shadow properties in images can be explored in different ways, property-based methods are then further classified into thresholding, invariant color model, and object segmentation. Thresholding is a simple image processing method that creates binary images by setting a pixel value threshold. The global thresholding and local thresholding [12], bimodal histograms, and Otsu's threshold [13] are examples of approaches that are generally used due to the simplicity of the implementation. Unfortunately, there are a lot of weaknesses of these methods. Some thresholding methods apply various threshold values to different segments of the image [12]. Moreover, implementation of the thresholding methods highly depends on prior knowledge of shadow properties. One of the shadow properties used is the significant changes in the red band of RGB components in the Optimized Shadow Index of the shadow region [3].

For the invariant color model methods, color information is exploited for shadow detection by using invariant properties of some color spaces, such as CIELAB [14] and C1C2C3 [15]. The drawback of these methods is that they depend mainly on the knowledge of the presence of shadow in different channels of color model. Proposing an object-oriented method, [16] performed a Gaussian Mixture Model to refine soft shadow and added the result to another result from image segmentation. While it managed to recover incomplete shapes and holes in the detected shadow, the accuracy obtained was not higher than 90%, while a lot of shadow analysis needed to be performed, such as clustering analysis. Another object-oriented segmentation approach was performed by [17], who used shadow features in the image segmentation step, followed by extraction of shadow regions using the statistical features of the images. They also used spatial information to avoid false shadow detection, but it still faced difficulty detecting small shadows.

2.2.2. Model-Based

Model-based is a method that includes physical and environmental information of shadows, such as light source direction, object geometry, and sensor localization. For example, a study by [18] detected shadows on UAV images of forested areas using in situ

spectral reflectance measurement. A digital surface model (DSM) was used to represent the elevation of the terrain. In [19], geometrical shadow detection was proposed, using solar position to map out shadow and non-shadow areas in VHR images. It was mentioned that sun position and DSM were used to reconstruct the geometric relationship between the ground and sun when generating binary shadow masks. One of the reasons why the model-based methods are not favorable is due to the presence of errors from the geometrical measurements that affect the accuracy of shadow detection [19].

2.2.3. Machine Learning

Shadow detection using machine learning, a technique where a model is built and fit into data to accomplish a task, is a method that has been preferred recently. This is mainly because machine learning techniques require less knowledge of shadow properties and offer more flexibility in detection compared to older techniques. Machine learning techniques can be divided into three more categories, which are unsupervised learning, supervised learning, and deep learning.

Unsupervised learning is popular due to its simplicity of implementation, and it does not require sample data to train the model, which saves a lot of time on data preparation. Shadow detection using unsupervised learning is often implemented by clustering technique, where the pixels of the images are grouped by their similarities, and a cluster that possibly represents shadow is selected. An algorithm that can be seen to be popular for performing shadow detection in aerial images is K-Means. Ref. [20] used improved K-Means, which is K-Means-AP, to solve the drawbacks of traditional K-Means, where the random number of clusters produces a random result. Another implementation was by [21], which compared the clustering results of two algorithms, K-means and Gaussian Mixture Model (GMM), after preprocessing the image using morphological operations. While using unsupervised learning guaranteed some benefits, the accuracy of the detection still needed improvement, as it was prone to noise [21].

Compared to unsupervised learning, supervised learning requires training samples to build classifiers, which usually produce more accurate detection in classification. The famous model used in shadow detection based on supervised learning is the Support Vector Machine (SVM). SVM is used to generate an initial map containing the classification of shadow class and non-shadow class. Various enhancements were also included in SVM, such as the implementation of LSSVM to reduce cross validation error [22] and adding Extended Random Walker to reduce the noisiness result of SVM [23].

As the machine learning field has been widely discussed to solve problems in visual information processing, implementation of deep learning in shadow detection has been a growing discussion, considering there are still many weaknesses shown in traditional unsupervised and supervised learning. Much research has discussed the implementation of deep learning in shadow detection, specifically in trying different neural network architectures and adding enhanced modules to achieve state-of-the-art results. When using drone images, deep learning has always been a consideration, as they usually produce very high-resolution images with complicated shadow areas. In deep learning, shadow detection is considered a semantic segmentation task, where each pixel is assigned as either shadow or non-shadow. Few works related to shadow detection in aerial images implementing deep learning have been studied. One of them proposed DSSDNet, a network that chose a deeply supervised CNN as its network architecture with the implementation of DSPF to improve model performance [10]. Alongside proposing a new model with enhancement, this work also aimed to solve shadow detection in aerial image problems related to insufficient training datasets by creating a publicly available dataset called AISD. The images were captured using drones flying at high altitudes; thus, many small shadowed regions cast by objects such as vehicles are not clearly noticeable and are not labeled. This contribution allows many researchers to propose and test state-of-the-art methods using this dataset, which widens the possibilities for different aspects of enhancement within this topic. One of the research projects which used the AISD dataset was by [4], who

proposed the Edge-aware Spatial Pyramid Fusion network (ESPFNet) for enhancing the detection of salient shadows. One of the main components in this network is the parallel spatial pyramid structure, which extracts multiscale features from the input image. The approach to aggregate the feature representation was similar to the pooling operation used in PSPNet [24]. Instead of aiming to detect the most noticeable shadows only, [9] proposed a shadow detection network based on a multiscale spatial attention mechanism (MSASDNet) to solve false detection in weak illumination regions and small shadow regions.

Another architecture that is widely used in various types of segmentation tasks is U-Net. It is an architecture that performs well with a small training dataset and was proven to improve image localization [25]. For shadow detection in aerial images, [26] considered non-local spatial contextual information, and proposed GSCA-UNet, a model composed of a U-shaped encoder and decoder, and a GSCA module for flexibility. Much other research in shadow detection also chose to build network architecture inspired by U-Net, such as [27–29], which indicates that U-Net architecture is one of the popular architectures in this domain. For this reason, given a small number of images, U-Net architecture is applied in this project to produce acceptable shadow detection results.

In the study, some limitations are identified to be further discussed, which are an insufficient training dataset, along with the need to spend too much time on dataset preparation due to complex shadow regions [10], unexpected noises in the resulting shadow masks [26], and inefficiency of trial-and-error post-processing steps [4]. Instead of using existing annotated datasets, such as AISD, which exclude the labeling of small shadowed regions, a new dataset is prepared for this study to focus on low altitude aerial images. Through this process, this study contributes a new annotated dataset with more of a variety of images and allows for more method comparisons in the future.

2.3. Post-Processing

Binary thresholding is a crucial part of binary segmentation using deep learning, as the pixels of the mask produced at the end of the process contain a number in the range of 0 to 1, which indicates the probability of the pixel belonging to the shadow class. Therefore, it is important to select a suitable value to decide on the resulting mask, as a wrong value will lead to enormous errors. Inspired by the implementation of the thresholding method in [3], which is able to threshold fine shadow, this project will analyze the result of Otsu's thresholding in computing the binary mask, aiming to propose an automatic thresholding method in deep learning implementation for shadow detection. In addition, Otsu's thresholding method has also been commonly implemented in medical segmentation works [30,31] and it has been proven that this method is sufficient to produce an accurate binary mask in the deep learning implementation.

Another important post-processing step in binary segmentation is mask refinement, which is a step to improve the result of segmentation by removing noise and reshaping the region of interest. Noise is a common problem faced in detecting shadows in drone images due to the complexity of shadows, even after implementing deep learning. For example, unexpected noise often occurs in the implementation of GSCA U-Net architecture that causes errors [26]. Mask refinement also reduces detection errors caused by the coarse shadow boundaries. In this project, two types of refinements will be analyzed to see how they are able to increase the accuracy of the detection compared to ground truth.

The first technique to be analyzed is dense CRF. This technique has been implemented in many semantic segmentation works to detect clear boundaries of regions. Even though U-Net concatenation features are implemented to improve fine-grained segmentation of high-resolution images, the max-pooling features used still have difficulties detecting boundaries clearly, as they are intended to improve high-level features extraction [27]. Therefore, CRF is implemented in this work to improve shadow boundary detection. In several shadow detection works, CRF has also been implemented as a post-processing step to smoothen the predicted result [32,33]. Dense CRF is selected by [27] to refine shadow boundaries in satellite images, resulting in a sharper boundary between shadow and cloud

regions. To solve problems such as noisy results and coarse boundaries, as well as to observe the effectiveness of this method in drone images, this project will implement dense CRF instead of normal CRF, as it has a long-range interaction between nodes and has no shrinking bias.

The second technique is morphological operation. This technique is inspired by the previous implementation of shadow detection [14], where the shadow mask produced is applied with morphological erosion, followed by dilation in order to reduce noise and fix the boundary. This method is also used to recover the penumbra region, where both dilation and erosion are also implemented [13]. Hence, to solve the refinement and noise problems, this project will apply and compare the performance of the morphological and dense CRF methods.

3. Materials and Methods

The proposed methodology consists of six steps, as shown in Figure 1, which start with dataset preparation. The second step is the implementation of preprocessing methods, which include cropping images into patches and data augmentation. Next is the building of U-Net network architecture and model training using the annotated dataset. The fifth step is the implementation of selected post-processing methods, which are Otsu's thresholding using OpenCV function, and two methods of binary mask refinement, morphological operations also using OpenCV functions, and dense CRF. Lastly, all the results from the implementation of post-processing on model prediction are evaluated using suitable evaluation metrics, and the predicted mask is visually compared. Figure 1 represents the flow chart of the proposed methodology.

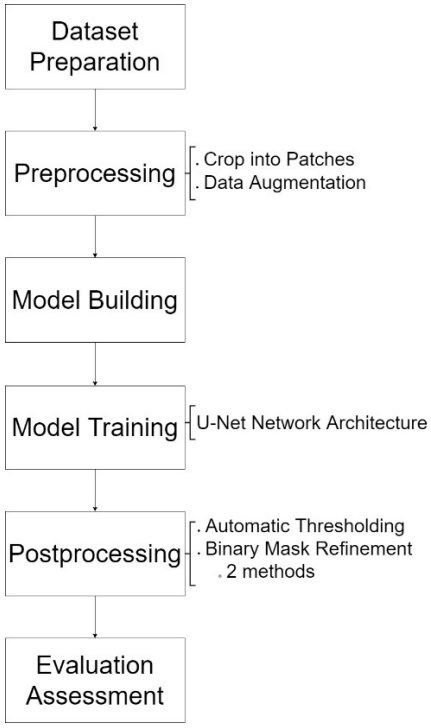


Figure 1. Workflow of methodology. The main steps include dataset preparation, image preprocessing, building of the shadow model, model training, post-processing, and evaluation of results.

The next sections will describe each step in the proposed methodology of this work in more detail.

3.1. Dataset Preparation

3.1.1. Dataset Sources

In this project, aerial images are collected and annotated, as there is a lack of public datasets to evaluate shadow detection algorithms in aerial images [10]. A new set of low-altitude aerial images is collected to be used instead of the AISD dataset, because the publicly available dataset excludes the labeling of small-shadowed regions. The images for the dataset in this project are collected from various publicly available drone images. The collected images are described below:

SenseFly Drone Dataset

The sample image is given in Figure 2. Ten images are selected from SenseFly Drone Dataset [34], specifically from the industrial estate area, where the majority of objects that can be seen are buildings, cars, and some trees.



Figure 2. Sample image of SenseFly Drone Dataset.

Aerial Semantic Segmentation Dataset by Kaggle

Another source is the Aerial Semantic Segmentation Dataset by Kaggle [35]. The sample image is given in Figure 3. The images from this dataset were taken from lower heights compared to the previous dataset, where the shadows of objects are bigger and more visible. The images contain shadows from various objects, mainly people, buildings, and trees. Twenty images are selected from this dataset.

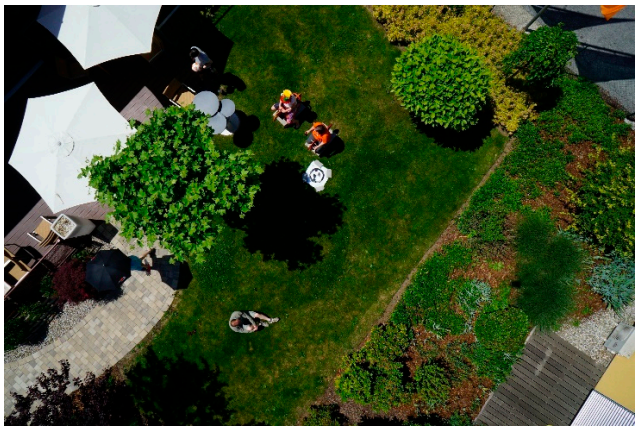


Figure 3. Sample image of Aerial Semantic Segmentation Dataset by Kaggle.

Mendeley Thermal and Visible Aerial Imagery

The final source of the dataset is from Mendeley Thermal and Visible Aerial Imagery [36]. The sample image is given in Figure 4. Twenty images are selected, in which the shadowy parts are mainly cast by trees with high brightness of shadows.



Figure 4. Sample image of Mendeley Thermal and Visible Aerial Imagery.

In image selection, the images with clear shadow regions are selected to ease annotation work, as shadow masks are not included in all datasets. For preliminary investigation, we also trained a shadow model using the non-drone dataset ISTD [29] and used the model for shadow detection in aerial images. This dataset provides shadow images and their corresponding binary shadow masks.

3.1.2. Dataset Annotation

The dataset consists of images taken from public sources, as stated in Section 3.1. All images are filtered by selecting only the images containing shadow areas. Then, the corresponding shadow mask (ground truth image) is annotated using PhotoScapeX software. The shadow mask is annotated using a magic wand, pen and eraser, and masking tool. A binary shadow mask for each image is then generated one by one.

3.2. Preprocessing

In order to reduce annotation errors and time consumed for annotation, the geometric transformation is applied as data augmentation, which includes rotation by 90 degrees clockwise and 90 degrees counterclockwise. The images are cropped into 512 × 512-sized patches before augmentation. The summary of the dataset description is shown in Table 1.

Table 1. Description of dataset.

Dataset	Number of Patches	Objects	Properties of Shadows
Semantic Drone Dataset	1540	Contain objects such as people, houses, trees, and cars, captured at low altitude.	Varying Brightness
SenseFly dataset (industrial state)	700	Contain objects such as vehicles, buildings, and trees, mostly captured at high altitude.	Low Brightness
Mendeley Thermal and Visible Aerial Imagery	20	Contain objects such as trees and buildings, captured at high altitude.	High Brightness

3.3. Network Architecture

In performing shadow detection using deep learning methods, shadow detection is considered a semantic segmentation task, specifically binary segmentation [10]. Binary segmentation is assigning pixels in an image to two types of labels; in this case, shadow and non-shadow. The result of the segmentation is binary masks, which typically can be visualized as white (shadow) and black (non-shadow) labels. To perform semantic segmentation tasks in this project, U-Net architecture is applied, which is a U-Shaped architecture that consists of an encoder and a decoder, and concatenation operation is performed in the decoder. This work also implemented U-Net with a low number of parameters to see the performance of a lightweight network in solving shadow detection, which is important for achieving real-time detection [26].

U-Net network architecture will perform a semantic segmentation task, which is performing classification at pixel level that generates a binary mask consisting of the shadow area only. It consists of an encoder and decoder which create a U-shaped operation, where the encoder is in the beginning of the architecture and is followed by the decoder. The encoder, which is also called the contraction path of U-Net architecture, is the same as the common architecture of a convolutional network. It contains four convolution layers, where each layer consists of two 2D convolutions with a 3×3 size kernel, followed by a rectified linear unit (ReLU), stride two downsamplings, and max pooling operation. The decoder, also called the expansive path, follows the contraction path, which consists of four deconvolution layers. Each deconvolution layer consists of a 2D transposed convolution operation for upsampling, followed by concatenation with a corresponding cropped feature map from the contraction path and two convolutions, which are then followed by a rectified linear unit (ReLU). The output of the expansive path is then applied to another 1×1 convolution to generate the final output. For the training process, Adam's algorithm and binary cross entropy as the loss function were used for optimization, while the train and validation split ratio was set to be 80:20. Before the training, the training images were cropped into small patches 512×512 in size to fit into the network architecture. Data augmentation is applied by rotating each patch 90 degrees clockwise and counterclockwise. The total number of training patches used during training is 5448. The summary of the training parameters is shown in Table 2.

Table 2. Summary of training parameter for U-Net architecture.

Training Parameters	Value/Type
Initial learning rate	0.001
Number of epochs	15
Filter size	3×3
Pooling size	2×2
Batch size	16
Optimizer	Adam
Loss function	Binary cross entropy
Dropout	8

In order to evaluate the performance of U-Net with the post-processing methods, another commonly used deep learning method is taken into consideration. In this study, we also implement the pyramid scene parsing network (PSPNet) [24] and retrain the network on the training set, and the results are compared. Due to the limitation of CPU, only 1816 training images were used for the comparison of U-Net and PSPNet, where a model is trained for each network architecture. The total images are from the dataset without implementing augmentation.

3.4. Comparison of U-Net Shadow Model Using Drone Dataset and Non-Drone Dataset

A comparison of results is made between two models that are trained using different datasets, which are the non-drone dataset and drone dataset. This comparison is performed

to observe the complexity of detecting shadows in drone images compared to non-drone images. Both models are trained by U-Net architecture using the same parameters and split ratio of 80:20. The datasets used in each model are described in Table 3.

Table 3. Comparison between non-drone and drone dataset.

Model	Dataset	Description
1	ISTD dataset	The shadow detection dataset that is available online consists of non-drone images with simple shadows. The images are resized into 512 × 512 patches, and all 1330 images from the train set are used for training.
2	Self-annotated drone dataset	The main dataset used in this project consists of drone images selected from multiple sources mentioned in Table 1. A total of 1330 patches with the size of 512 × 512 are selected for training.

3.5. Post-Processing

3.5.1. Otsu's Thresholding

Otsu's method of thresholding is used to automatically classify the image pixels into two classes. The method calculates the optimal threshold value based on the between-class variances between the foreground and background. Otsu's thresholding iterates over multiple threshold values to find the maximum between-class variance as the optimum threshold value and applies it to the shadow mask. The equation of between-class variance [37] is derived using weight W and mean μ; that is:

$$Between\ Class\ Variance,\ \sigma_B^2 = W_{background} W_{foreground} \left(\mu_{background} - \mu_{foreground}\right)^2 \quad (1)$$

3.5.2. Morphological Operation

Opening and closing operations are applied to remove noise pixels in the resulting binary masks. The opening process is the implementation of erosion followed by dilation, with the purpose of opening gaps in the mask. The closing process is to close the gap while maintaining the foreground shape via the implementation of a dilation, followed by erosion. The operations are applied on image pixel I with structuring element s, and defined as:

$$Opening,\ I \circ s = (I\ \ominus s) \oplus s \quad (2)$$

$$Closing,\ I \cdot s = (I \oplus s) \ominus s \quad (3)$$

where I is dilation that grows the foreground pixels based on s. Meanwhile, $\ominus$ is erosion that shrinks the foreground pixels based on s. In this project, two shapes of structuring elements s are applied for analysis, which are a rectangle and a cross, where the visualization of the shapes is shown in Figures 5 and 6. For both operations and all shapes of structuring elements, the same size of 3 × 3 structuring element is used to obtain a fine result.

Figure 5 shows the visualization of the rectangle shape structuring element used for morphological operation, while Figure 6 shows the cross shape of the structuring element. Both structuring elements are in the size of 3 × 3, which is the size used in the experiment.

1	1	1
1	1	1
1	1	1

Figure 5. Visualization of 3 × 3 rectangle shape structuring element.

0	1	0
1	1	1
0	1	0

Figure 6. Visualization of 3 × 3 cross shape structuring element.

3.5.3. Fully Connected Conditional Random Field

In this work, the implementation of dense CRF post-processing is similar to binary segmentation implementation in [38,39]. CRF is pixel-wise labeling, forming a graph called Markov Random Field based on pixel labels as random variables, modeled as nodes, and global observation from image for the edges. The pixel-wise labeling can be modeled as CRF characterized by Gibbs distribution [27]. Gibbs' energy function in dense CRF is defined as:

$$E(x) = \sum_i \psi_u(x_i) + \sum_i \sum_{i<j} \psi_p(x_i, x_j) \tag{4}$$

where x denotes the assignment of labels for pixels. u represents unary potential, where $\psi_u(x_i)$ is taken from outputs from U-Net prediction. ψ_p represents pairwise potential, which $\psi_p(x_i)$ is described as:

$$\psi_p(x_i, x_j) = \mu(x_i, x_j) \underbrace{\sum_{m=1}^{K} w^{(m)} k^{(m)}(f_i, f_j)}_{k(f_i, f_j)} \tag{5}$$

where $\mu(x_i, x_j)$ represents the compatibility of label and feature vectors f_i and f_j. Each $k^{(m)}$ for $m = 1, \ldots, K$ is a Gaussian Kernel that is applied on feature vectors, and $w^{(m)}$ are the linear combination weights.

3.6. Evaluation

In order to observe the performance using different image conditions, three images are chosen from each of the selected datasets. The different image conditions include varying illumination characteristics, the variety of objects, and the amount of shadow in the images. The image in the testing set is then cropped into 444 patches and evaluated using qualitative and quantitative analysis.

The predicted shadow mask obtained with the implementation of the proposed post-processing methods will be compared with the corresponding ground truth image. Different evaluation metrics are used to quantitatively examine the performance of the applied methods. To monitor the overall performance, overall accuracy (OA) is calculated based on Equation (6). F1-score is calculated as the weighted average of precision and recall, as given in Equation (9). Precision and recall are calculated using Equation (7) and Equation (8), respectively. Intersection Over Union (IoU), as in Equation (10), which is also called the Jaccard index, is used to measure the percentage of overlap between the ground truth and output predicted mask.

$$Overall\ Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Negative + False\ Positive} \quad (6)$$

$$Precision = \frac{True\ Positive}{(True\ Positive + False\ Positive)} \quad (7)$$

$$Recall = \frac{True\ Positive}{(True\ Positive + False\ Negative)} \quad (8)$$

$$F1\ Score = 2 \times \frac{(Recall \times Precision)}{(Recall + Precision)} \quad (9)$$

$$IoU = \frac{True\ Positive}{(True\ Positive + False\ Positive + False\ Negative)} \quad (10)$$

4. Results

4.1. Result of Shadow Models Using Drone Dataset and Non-Drone Dataset

Quantitative analysis is performed using a test set consisting of drone images prepared during this project. The result of evaluation metrics in Table 4 shows that the model trained with the drone dataset performs much better than the non-drone ISTD model. The overall accuracies of shadow detection in aerial images using shadow models based on the ISTD and drone images are 0.7962 and 0.9551, respectively.

Table 4. Results of using non-drone and drone dataset models.

Evaluation Metrics	Model Trained with ISTD Dataset	Model Trained with Drone Dataset
Overall Accuracy	0.7962	0.9551
F1 Score	0.2916	0.3855
IoU/Jaccard	0.2351	0.3268

4.2. Result Comparison between U-Net and PSPNet

Quantitative analysis is performed using a test set on the models trained with U-Net and PSPNet using the same dataset. The result of evaluation metrics in Table 5 shows that the U-Net model performs better than the PSPNet model, where the U-Net model has a higher number for accuracy, F1 score, and IoU. The overall accuracies of shadow detection in aerial images using U-Net and PSPNet are 0.9624 and 0.8614, respectively.

Table 5. Results of U-Net and PSPNet models.

Evaluation Metrics	U-Net	PSPNet
Overall Accuracy	0.9624	0.8614
F1 Score	0.9620	0.8654
IoU/Jaccard	0.9390	0.8516

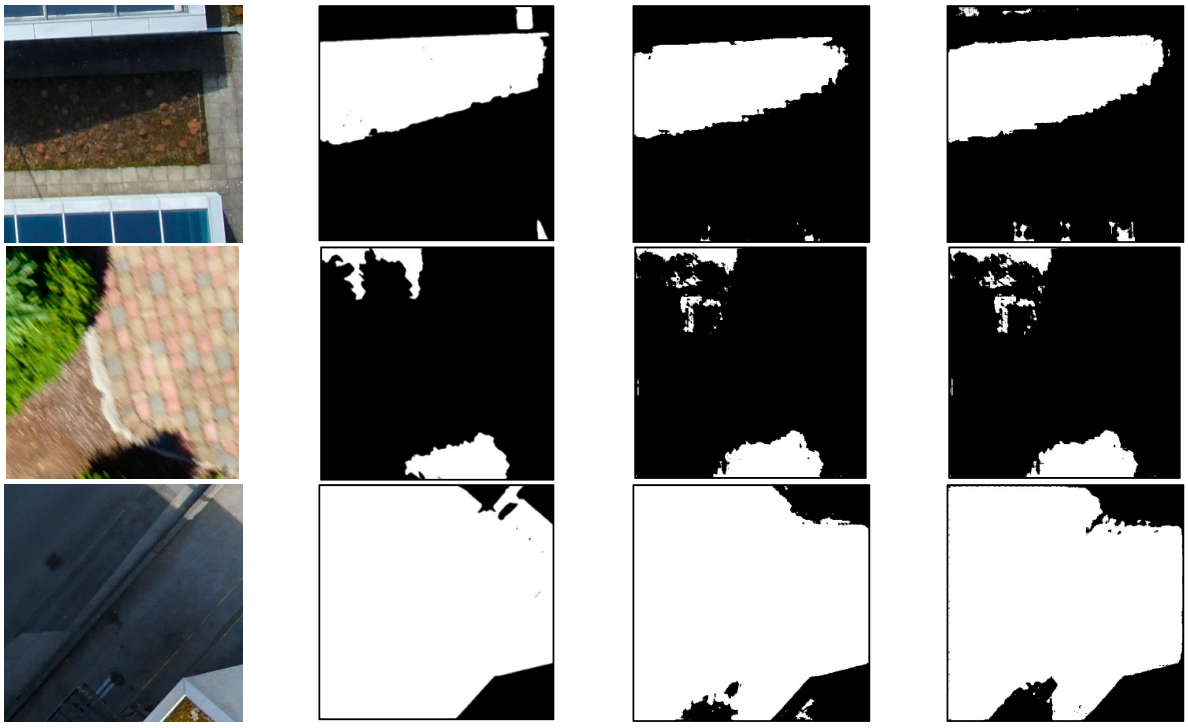
4.3. Result of Automatic Thresholding

Table 6 provides the quantitative analysis of manual thresholding and Otsu's thresholding. It demonstrates that Otsu's thresholding manages to obtain results comparable to manual thresholding (0.5). Even though there is a slight decrease in the overall accuracy, F1 score, IoU, and recall, this is acceptable, as this automatic thresholding manages to overcome the tedious identification of the trial-and-error thresholding value.

Table 6. Results of manual thresholding and Otsu's thresholding.

Evaluation Metrics	Manual Thresholding = 0.5	Otsu's Thresholding
Overall Accuracy	0.9638	0.9412
F1 Score	0.3728	0.4061
IoU/Jaccard	0.3212	0.3490
Recall	0.3885	0.4839
Precision	0.4194	0.3866

Three test images are selected to better visualize the impact of this method in thresholding binary masks. The visual comparison is illustrated in Figure 7.



(**a**) (**b**) (**c**) (**d**)

Figure 7. Visual comparison of thresholding method for Test Image 1 (row 1), Test Image 2 (row 2), and Test Image 3 (row 3). (**a**) input image, (**b**) ground truth, (**c**) manual thresholding, and (**d**) Otsu's thresholding. The outputs using Otsu's thresholding are comparable to the results using manual thresholding.

4.4. Result of Binary Mask Refinement

This section shows the results of binary mask refinement methods, which are morphological operations and dense CRF. Based on the overall accuracy of test images, as shown in Table 7, except for closing operation with cross kernel shape, all the selected methods cause slight increases in accuracy. Dense CRF can be seen to have the highest accuracy, which equals 0.9643.

Table 7. Quantitative analysis of binary mask refinement methods.

Method		Accuracy
Without refinement		0.9638
Morphological operation opening	Rectangle kernel shape	0.9640
	Cross kernel shape	0.9639
Morphological operation closing	Rectangle kernel shape	0.9640
	Cross kernel shape	0.9637
Dense CRF		0.9643

For qualitative evaluation, two test images are selected to display the results of the selected methods. This analysis observes how the selected method improves or worsens the detection of shadow regions with unclear boundaries. The results using Test Image 1 are shown in Figure 8. The second test image has a more complicated and unclear boundary, as shown in Figure 9. It can be observed that noises in the predicted mask without refinement are more visible compared to the results with refined methods.

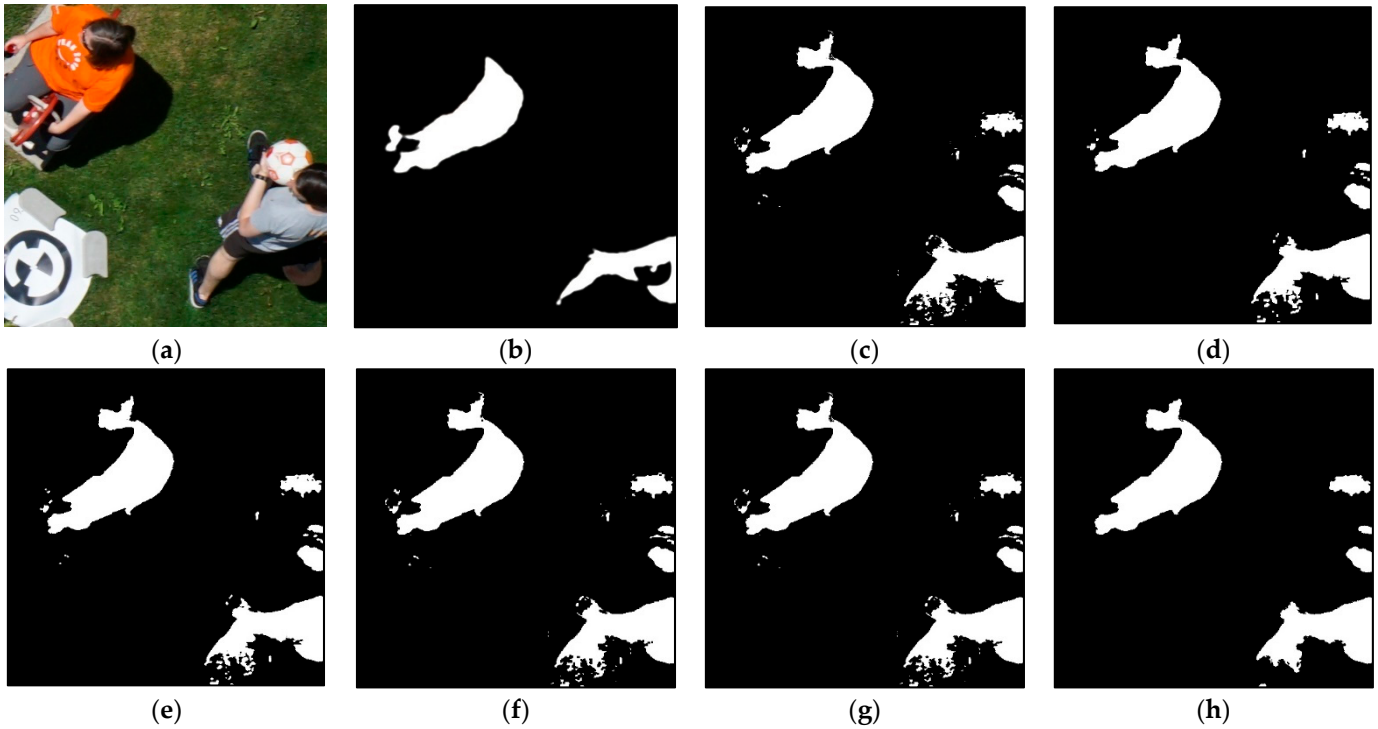


Figure 8. Visual comparison of refinement methods on Test Image 1: (**a**) input image, (**b**) ground truth, (**c**) without refinement, (**d**) opening using rectangle kernel shape (**e**), opening using cross kernel shape, (**f**) closing using rectangle kernel shape, (**g**) closing using cross kernel shape, and (**h**) dense CRF. Dense CRF performs considerably well, as most of the white noise is removed.

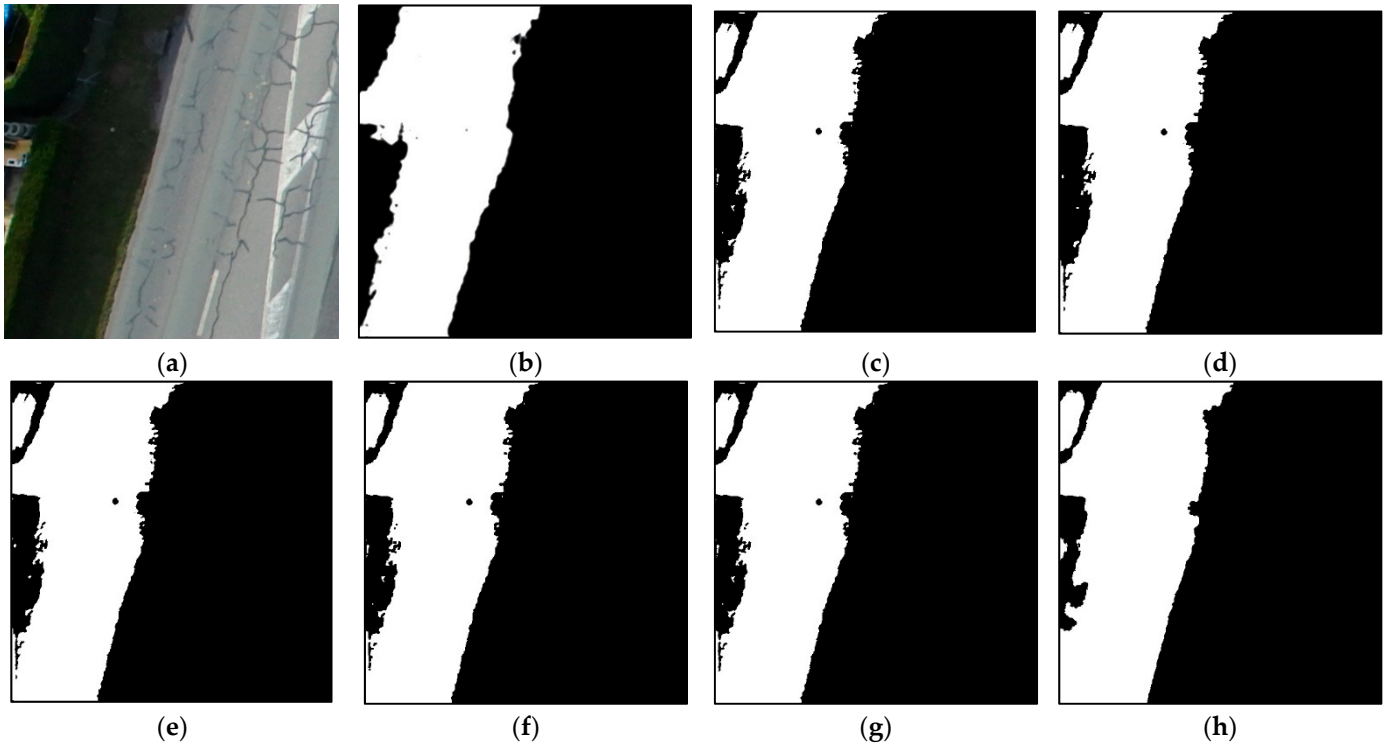


Figure 9. Visual comparison of refinement methods on Test Image 2: (**a**) input image, (**b**) ground truth, (**c**) without refinement, (**d**) opening using rectangle kernel shape (**e**), opening using cross kernel shape, (**f**) closing using rectangle kernel shape, (**g**) closing using cross kernel shape, and (**h**) dense CRF. Dense CRF performs considerably well, as most of the white noise is removed.

The second test image shown in Figure 9 has a more complicated and unclear boundary. This analysis observes how the selected method improves or worsens the detection of shadow regions with unclear boundaries. Comparing the result of opening and closing operations, the opening method manages to erase white noise, while the closing method erases black noise. Both methods maintain the boundary of the shadow, as predicted. Dense CRF seems to erase most of the noises and perform better than the other two methods, but it smoothens the boundary and fills up the wrong non-shadow area into a shadow area.

5. Discussion

From the results, we observe that a thorough study is necessary to perform when detecting shadows in drone images. Our initial investigation demonstrated that the non-drone ISTD images could not be used to obtain shadow detection in drone images, as shown in Table 4. We expect similar unsatisfactory results as those in Table 4 when using a model created based on other non-drone datasets that were widely used in the literature (i.e., SBU and UCF). This is because the images of the non-drone dataset are insufficient to model the complex shadows in drone images.

Furthermore, the use of automatic thresholding in U-Net architecture is necessary considering the variant aerial images. Even though the comparison of the results between manual and automatic thresholding in Table 5 is not apparent, the automatic Otsu's thresholding is necessary to adaptively distinguish shadow and no-shadow pixels according to different images. Based on the visual observation in Figure 7, regardless of whether using manual or automatic thresholding, U-Net shadow detection has difficulty identifying the boundary of shadow regions correctly. The unsmooth shadow boundary detection is clearly visible in Figure 7c,d.

As for the refinement methods, dense CRF outperforms the two morphological operations, as the results show that dense CRF achieves the best performance, quantitatively and qualitatively (refer to Figures 8h and 9h). Despite its performance in refining the final

shadow masks, it erases some small shadow areas. Furthermore, it smoothens the boundary and fills up the non-shadow area into a shadow area. For the opening operations, although it incompletely eliminates the small noises, lots of small white noise in several areas is removed (refer to Figure 8d,e and Figure 9d,e). Both kernel shapes show the same effect on the predicted mask. As for the closing operation, it is unable to improve the predicted mask, as it is unable to remove most of the white noise (refer to Figure 8f,g and Figure 9f,g). It is consistent with the purpose of the closing operation implementation, which is to close gaps in white regions.

Overall, the poor performance of shadow detection in this study is contributed by three factors: (i) unclear shadow boundary due to unstructured objects, such as trees, (ii) dark color pixels, such as tarmac, misclassified as shadows (False Positive), and (iii) failed detection (False Negative) of small shadowed regions cast by objects that appear small, such as vehicles, humans, and bushes. These problems are non-trivial when dealing with high-spatial-resolution images captured at high altitudes. Thus, a careful investigation should be conducted when detecting shadows in drone images.

6. Conclusions

This study presents reliable post-processing methods to improve predicted binary shadow masks. According to the experimental results, overall, Otsu's automatic thresholding is comparable to time-consuming manual thresholding. Furthermore, the investigation shows that the refinement method based on dense CRF outperforms the morphological operators. In the analysis of binary mask refinement, we show the effectiveness of morphological operations and dense CRF, where the methods increase the accuracy of shadow detection. Further improvement can be made by building a network architecture with a higher number of parameters to improve recall and precision. Based on our results, further study on automatic thresholding methods can also be made to better improve efficiency and selection of optimum threshold values. Methods to refine binary masks without erasing small shadow areas are also to be investigated in future work.

Author Contributions: Conceptualization, S.-A.Z., S.M.-D., Z.C.-E. and W.-N.M.-I.; methodology, S.-A.Z.; validation, S.-A.Z., S.M.-D., Z.C.-E. and W.-N.M.-I.; formal analysis, S.-A.Z.; investigation, S.-A.Z.; data curation, S.-A.Z. and S.M.-D.; writing—original draft preparation, S.-A.Z.; writing—review and editing, S.-A.Z., S.M.-D., Z.C.-E. and W.-N.M.-I.; visualization, S.-A.Z.; supervision, S.M.-D. and Z.C.-E.; funding acquisition, S.M.-D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the IR Fund of Multimedia University, Malaysia (MMUI/210095).

Institutional Review Board Statement: Not Applicable.

Data Availability Statement: Restrictions apply to the availability of these data. Data are available from the authors with the permission of Multimedia University.

Acknowledgments: The authors would like to thank Noramiza Hashim and Mohamad Zamfirdaus bin Mohd Saberi for providing technical support for completion of the study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Trapal, D.D.C.; Leong, B.C.C.; Ng, H.W.; Zhong, J.T.G.; Srigrarom, S.; Chan, T.H. Improvement of Vision-based Drone Detection and Tracking by Removing Cluttered Background, Shadow and Water Reflection with Super Resolution. In Proceedings of the 2021 6th International Conference on Control and Robotics Engineering (ICCRE), Beijing, China, 16–18 April 2021; pp. 162–168. [CrossRef]
2. Sharma, D.; Singhai, J. An Object-Based Shadow Detection Method for Building Delineation in High-Resolution Satellite Images. *PFG J. Photogramm. Remote Sens. Geoinf. Sci.* **2019**, *87*, 103–118. [CrossRef]
3. Mostafa, Y.; Abdelhafiz, A. Shadow Identification in High Resolution Satellite Images in the Presence of Water Regions. *Photogramm. Eng. Remote Sens.* **2017**, *83*, 87–94. [CrossRef]
4. Luo, S.; Li, H.; Zhu, R.; Gong, Y.; Shen, H. ESPFNet: An Edge-Aware Spatial Pyramid Fusion Network for Salient Shadow Detection in Aerial Remote Sensing Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 4633–4646. [CrossRef]

5. Freitas, V.L.D.S.; Reis, B.M.D.F.; Tommaselli, A.M.G. Automatic shadow detection in aerial and terrestrial images. *Bol. Ciências Geodésicas* **2017**, *23*, 578–590. [CrossRef]
6. Min, S.; Lee, J.; Won, J.; Lee, J. Soft shadow art. In Proceedings of the Joint Symposium on Computational Aesthetics and Sketch Based Interfaces and Modeling and Non-Photorealistic Animation and Rendering, Los Angeles, CA, USA, 29–30 July 2017. [CrossRef]
7. Gheorghe, C.; Gheorghe, C.; Filip, N. Image Processing Technique Used in Road Traffic Analysis—Opportunities and Challenges. *Acta Tech. Napocensis Ser. Appl. Math. Mech. Eng.* **2021**, *64*, S1–S2. Available online: https://atna-mam.utcluj.ro/index.php/Acta/article/view/1532 (accessed on 13 October 2021).
8. Che'Ya, N.; Dunwoody, E.; Gupta, M. Assessment of Weed Classification Using Hyperspectral Reflectance and Optimal Multispectral UAV Imagery. *Agronomy* **2021**, *11*, 1435. [CrossRef]
9. Liu, D.; Zhang, J.; Wu, Y.; Zhang, Y. A Shadow Detection Algorithm Based on Multiscale Spatial Attention Mechanism for Aerial Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 6003905. [CrossRef]
10. Luo, S.; Li, H.; Shen, H. Deeply supervised convolutional neural network for shadow detection based on a novel aerial shadow imagery dataset. *ISPRS J. Photogramm. Remote Sens.* **2020**, *167*, 443–457. [CrossRef]
11. Movia, A.; Beinat, A.; Crosilla, F. Shadow detection and removal in RGB VHR images for land use unsupervised classification. *ISPRS J. Photogramm. Remote Sens.* **2016**, *119*, 485–495. [CrossRef]
12. Truptirajendraghewari, M.; Khot, A.S.R.; Pise, A.P.S. Successive Thresholding Scheme for Shadow Detection of Aerial Images. Available online: https://www.ripublication.com/irph/ijert_spl17/ijertv10n1spl_89.pdf (accessed on 13 October 2021).
13. Su, N.; Zhang, Y.; Tian, S.; Yan, Y.; Miao, X. Shadow Detection and Removal for Occluded Object Information Recovery in Urban High-Resolution Panchromatic Satellite Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 2568–2582. [CrossRef]
14. Silva, G.F.; Carneiro, G.B.; Doth, R.; Amaral, L.A.; de Azevedo, D.F. Near real-time shadow detection and removal in aerial motion imagery application. *ISPRS J. Photogramm. Remote Sens.* **2018**, *140*, 104–121. [CrossRef]
15. Mostafa, Y.; Nady, B. Study on shadow detection from high-resolution satellite images using color model. *Sohag Eng. J.* **2021**, *1*, 85–95. [CrossRef]
16. Mo, N.; Zhu, R.; Yan, L.; Zhao, Z. Deshadowing of Urban Airborne Imagery Based on Object-Oriented Automatic Shadow Detection and Regional Matching Compensation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 585–605. [CrossRef]
17. Zhang, H.; Sun, K.; Li, W. Object-Oriented Shadow Detection and Removal from Urban High-Resolution Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 6972–6982. [CrossRef]
18. Pons, X.; Padró, J.C. An Empirical Approach on Shadow Reduction of UAV Imagery in Forests. In Proceedings of the IGARSS 2019–2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 2463–2466. [CrossRef]
19. Wang, Q.; Yan, L.; Yuan, Q.; Ma, Z. An Automatic Shadow Detection Method for VHR Remote Sensing Orthoimagery. *Remote Sens.* **2017**, *9*, 469. [CrossRef]
20. Du, Y.; Li, J.; Wang, Y. Shadow Detection in High-Resolution Remote Sensing Image Based on Improved K-means. In Proceedings of the ICIMCS'16: International Conference on Internet Multimedia Computing and Service, Xi'an, China, 19–21 August 2016; pp. 281–286.
21. Deshpande, A.M.; Gaikwad, M.; Patki, S.; Rathi, A.; Roy, S. Shadow detection from aerial imagery with morphological preprocessing and pixel clustering methods. *ICTACT J. Image Video Process.* **2021**, *11*, 3. [CrossRef]
22. Vicente, T.F.Y.; Hoai, M.; Samaras, D. Leave-One-Out Kernel Optimization for Shadow Detection and Removal. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 682–695. [CrossRef]
23. Kang, X.; Huang, Y.; Li, S.; Lin, H.; Benediktsson, J.A. Extended Random Walker for Shadow Detection in Very High Resolution Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 867–876. [CrossRef]
24. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 6230–6239. [CrossRef]
25. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention 2015*; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2015; Volume 9351, pp. 234–241. [CrossRef]
26. Jin, Y.; Xu, W.; Hu, Z.; Jia, H.; Luo, X.; Shao, D. GSCA-UNet: Towards Automatic Shadow Detection in Urban Aerial Imagery with Global-Spatial-Context Attention Module. *Remote Sens.* **2020**, *12*, 2864. [CrossRef]
27. Jiao, L.; Huo, L.; Hu, C.; Tang, P. Refined UNet: UNet-Based Refinement Network for Cloud and Shadow Precise Segmentation. *Remote Sens.* **2020**, *12*, 2001. [CrossRef]
28. Le, H.; Vicente, T.F.Y.; Nguyen, V.; Hoai, M.; Samaras, D. A+D Net: Training a Shadow Detector with Adversarial Shadow Attenuation. *arXiv* **2017**, arXiv:1712.01361.
29. Wang, J.; Li, X.; Hui, L.; Yang, J. Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 1788–1797. [CrossRef]
30. Horwath, J.P.; Zakharov, D.N.; Mégret, R.; Stach, E.A. Understanding important features of deep learning models for segmentation of high-resolution transmission electron microscopy images. *npj Comput. Mater.* **2020**, *6*, 108. [CrossRef]

31. Xu, Y.; Gao, F.; Wu, T.; Bennett, K.M.; Charlton, J.R.; Sarkar, S. U-Net with optimal thresholding for small blob detection in medical images. In Proceedings of the 2019 IEEE 15th International Conference on Automation Science and Engineering (CASE), Vancouver, BC, Canada, 22–26 August 2019; pp. 1761–1767. [CrossRef]
32. Chen, Z.; Zhu, L.; Wan, L.; Wang, S.; Feng, W.; Heng, P.A. A Multi-Task Mean Teacher for Semi-Supervised Shadow Detection. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 5610–5619. [CrossRef]
33. Zhu, L.; Deng, Z.; Hu, X.; Fu, C.-W.; Xu, X.; Qin, J.; Heng, P.-A. Bidirectional Feature Pyramid Network with Recurrent Attention Residual Modules for Shadow Detection. In Proceedings of the Computer Vision—ECCV 2018, Munich, Germany, 8–14 September 2018; Volume 11210, pp. 122–137. [CrossRef]
34. Discover a Wide Range of Drone Datasets—SenseFly. Available online: https://www.sensefly.com/education/datasets/ (accessed on 17 November 2021).
35. ICG—DroneDataset. Available online: https://www.tugraz.at/index.php?id=22387 (accessed on 17 November 2021).
36. Garcia, L.; Diaz, J.; Correa, H.L.; Restrepo-Girón, A.D. Thermal and Visible Aerial Imagery. Mendeley Data, V2, 2020. Available online: https://data.mendeley.com/datasets/ffgxxzx298/2 (accessed on 20 June 2022).
37. Greensted, A. The Lab Book Pages Sitewide RSS. Available online: http://www.labbookpages.co.uk/software/imgProc/otsuthreshold.html (accessed on 8 June 2022).
38. Zheng, S.; Jayasumana, S.; Romera-Paredes, B.; Vineet, V.; Su, Z.; Du, D.; Huang, C.; Torr, P.H.S. Conditional random fields as recurrent neural networks. *arXiv* **2016**, arXiv:1502.03240v3.
39. Arnab, A.; Jayasumana, S.; Zheng, S.; Torr, P.H.S. Higher Order Conditional Random Fields in Deep Neural Networks. In Proceedings of the 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Volume 9906, pp. 524–540. [CrossRef]