Preliminary Study on Shadow Detection in Drone-Acquired Images with U-NET

Siti-Aisyah Zali, Shahbe M-Desa(✉), Zarina Che-Embi, and Wan-Noorshahida Mohd-Isa

Faculty of Computing and Informatics, Multimedia University, Cyberjaya, Selangor, Malaysia
shahbe@mmu.edu.my

Abstract. This study shows a preliminary investigation of shadow detection in drone-acquired images using a deep learning method with minimal labelled shadow images. The aim is to discuss how the selected U-Net architecture performs in a small-sized dataset consisting of various types of shadow brightness and objects. Two types of data augmentation methods, which are shadow variant and geometric transformation are implemented, aiming to improve the segmentation accuracy. Several experimental procedures are performed to observe the model performance. The study shows that adding images for training increases the accuracy of shadow detection in drone images from 0.95 to 0.96, and geometric transformation data augmentation method increases the accuracy from 0.961 to 0.963, while the shadow variant method increases the flexibility of detection.

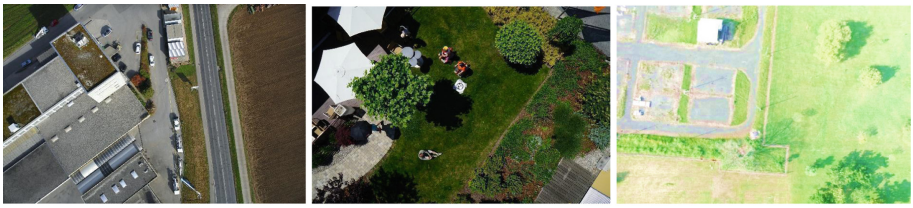
Keywords: Shadow detection · Deep learning · Aerial images · Data augmentation

1 Introduction

Shadow detection is an important image pre-processing step as the occurrence of shadow brings errors in numerous applications. For example, in an attempt to detect drones, removing shadows of the drone is necessary as it causes mirror images of the drone that bring confusion to the detector [1]. The shadows also become a constraint of the tested methods in road traffic analysis [2] and weed classification [3]. On the other hand, shadow properties are useful for some tasks such as building height estimation [4]. Thus, the need for shadow detection is significant as drone usage has been popularised, and many industries have used drones to capture their surroundings, creating different kinds of drone images such as for industrial areas, indoor and outdoor applications. Sample outdoor drone-acquired images are shown in Fig. 1.

Shadow detection step on drone images has been a challenging task due to a lack of labelled shadow data as data collection and annotation are costly and time-consuming [5]. Moreover, drone images contain complicated shadows compared to non-aerial images as they are captured at various heights and times which results in different sizes and brightness of shadows. The uneven characteristics of different types of shadow also make it harder to accurately detect shadow regions. There are two major groups of

S.-C. Haw and K. Sonai Muthu (Eds.): CITIC 2022, 10, pp. 357–368, 2022.
https://doi.org/10.2991/978-94-6463-094-7_28



(a) (b) (c)

Fig. 1. Sample Drone-Acquired Images from (**a**) SenseFly Drone Dataset [7], (**b**) Aerial Semantic Segmentation Dataset by Kaggle [8], and (**c**) Mendeley Thermal and Visible Aerial Imagery [9]

shadows: Cast shadows and self-cast shadows [6]. Cast shadow is produced by an object onto another object, while self-cast shadow is the shadow that is cast by the object on itself. Shadows also can be classified into two subcategories, umbra, and penumbra [6]. Penumbra is the region of shadow where only a part of the ray of light is concealed by an object, and umbra is when the light is fully halted by the object.

Various characteristics of shadows have led to different shadow detection solutions using physical information, such as the location of the scene and the angles of the drone. The methods that rely so much on this knowledge and involve trial-and-error processes, show the inefficiency of the methods to be applied and are also prone to device errors and reading errors. The unsupervised and supervised machine learning approaches are alternative solutions to the shadow physical information dependencies. Yet many have stated the importance of implementing lightweight networks to increase efficiency. Another problem that comes with the implementation of machine learning is the lack of aerial shadow datasets.

In this preliminary study, we discuss a deep learning method, specifically using U-Net architecture [17], for solving shadow detection in drone images. We aim to investigate the following:

i. The implementation of U-Net architecture for small-sized datasets with a small number of parameters.
ii. The effectiveness of data augmentation methods for improving shadow detection performance.
iii. The effect of using different sources of datasets to train shadow models.
iv. The suitable split ratios of test and train sets to observe the performance of the shadow models.

2 Related Work

Shadow detection methods can be divided into three categories which are property-based methods, model-based methods, and machine learning methods [5]. Property-based methods include several subcategories of methods such as thresholding, invariant color model and object segmentation method. Model-based methods use the information obtained from the environment of the scene captured and the physical properties of the drone. Machine learning techniques which include unsupervised learning and supervised

learning are preferred mainly because the techniques require less knowledge of shadow properties compared to older techniques. Unsupervised learning is often implemented by clustering techniques such as Colour Quantization, K-means and Gaussian Mixture Model [10]. Supervised learning uses training samples to build shadow and non-shadow models using Support Vector Machine [11, 12] and deep learning [5, 13–16].

Considering the constraint in the unsupervised and supervised learning caused by irregular shapes of shadows and noises, more improved deep learning approaches were proposed. The incorporation of progressive feature fusion in deep learning architecture by [5] has been shown to improve shadow model performance. At the same time, a new aerial shadow imagery dataset (AISD) was created [5] and an approach based on edge aware spatial pyramid fusion network for detecting salient shadows was later proposed [13]. Another work in [14] proposed GSCA-UNet, a model composed of a U-shaped encoder and decoder, and GSCA module for flexibility. U-Net [17] which performs well with a small training dataset has been selected to perform shadow detection in satellite images [14, 15]. [14] mentioned several problems encountered in the implementation of GSCA-UNet such as high cost for computation and number of parameters. Therefore, U-Net network architecture with a small number of parameters is used which also requires less cost of computation.

An insufficient amount of data always leads to model overfitting in supervised learning approaches and it greatly impacts the shadow detection on complex aerial images. As a way to reduce annotation errors as well as reduce the time consumed for annotation, a data augmentation technique that produces multiple images by changing the brightness level of the shadow region was used [16]. Another augmentation method using geometrical transformation involving flipping and rotation was also implemented [14]. The augmentation outcomes are realistic, where the same images can be generated when a drone changes its direction when capturing images.

3 Methodology

The research methodology consists of five levels, in which the first level starts with dataset preparation. Second, the images are pre-processed by cropping into patches of size 512 x 512. Third, is the building of the U-Net architecture. Next, the level continues with the training of shadow models using several setups which differ in types of datasets, splitting ratios, and data augmentation methods. Lastly, the results using different training models are quantitatively and visually evaluated. Figure 2 represents the workflow of the methodology.

3.1 Dataset Preparation

The drone-acquired images are collected from various available datasets: (i) SenseFly Drone Dataset [7], (ii) Aerial Semantic Segmentation Dataset by Kaggle [8], and (iii) Mendeley Thermal and Visible Aerial Imagery [9]. The images with significant shadows are selected to ease annotation work as shadow masks are not included in these datasets. The shadow masks were manually annotated using PhotoScapeX software. The summary of the dataset description is shown in Table 1.

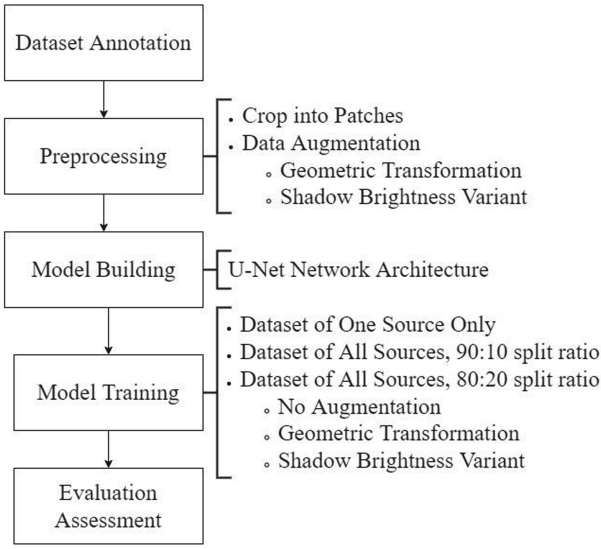


Fig. 2. Workflow of methodology

3.2 Crop into Patches

All the images and annotated masks are cropped into 512 x 512 patches in order to fit into the network architecture. The cropping is done using the Patchify function where the dimension and step are specified. The final number of patches obtained from each dataset and used for training and testing are shown in Table 1.

3.3 U-Net Implementation

U-Net with a low number of parameters is implemented in order to see the performance of a lightweight network in solving shadow detection. This network architecture will perform a semantic segmentation task, which is performing classification at pixel level that generates a binary mask consisting of shadow area only. The contraction path of U-Net is the same as the common architecture of a convolutional network. It contains four convolution layers. Each layer consists of two 2D convolutions with a 3 x 3 size kernel, followed by a rectified linear unit (ReLU), stride two down-sampling, and max-pooling operation. The expansive path consists of four deconvolution layers. Each deconvolution layer consists of a 2D transposed convolution operation for up-sampling, followed by concatenation with a corresponding cropped feature map from the contraction path and two convolutions, ending with a rectified linear unit (ReLU). The output of the expansive path is then applied to another 1 x 1 convolution to generate the final output. For the training process, Adam algorithm and binary cross entropy as loss function is used for optimization.

Table 1. Description of dataset

Dataset	Properties of shadowed images
Dataset 1: Semantic Drone Dataset	– 1540 shadowed patches – Various Brightness – Contain objects such as people, houses, trees and cars – Captured by drone at a low altitude
Dataset 2: SenseFly dataset	– 700 shadowed patches – Low Brightness – Contain objects such as vehicles, buildings and trees – Captured by drone at a high altitude
Dataset 3: Mendeley Thermal and Visible Aerial Imagery	– 20 shadowed patches – High Brightness – Contain objects such as trees and buildings – Captured by drone at a high altitude

3.4 Data Augmentation

As a way to reduce annotation errors as well as reduce time consumed for annotation, two data augmentation techniques are applied: geometric transformation and brightness variant. The images are cropped into patches before augmentation. For the geometric transformation technique, the images are rotated 90° clockwise and 90° counter clockwise. To generate different brightness of the shadow region, a gamma correction formula is applied using four gamma values, $\gamma_1 = 0.6$, $\gamma_2 = 0.8$, $\gamma_3 = 1.2$, and $\gamma_4 = 1.4$. Given $\gamma =$ gamma values, $I =$ pixel values, the formula of gamma correction is as Eq. (1).

$$I' = 255 \times \left(\frac{I}{255}\right)^{\gamma} \tag{1}$$

The procedure of generating brightness variant images on shadow regions is described as follows:

1. Shadow region is extracted into an image using bitwise AND operation of mask and image.
2. Gamma correction is performed by implementing the formula with four gamma values (0.6, 0.8, 1.2, 1.4) on the shadow ROI image.
3. For each gamma value (γ), a copy of the original image is created. Shadow ROI from step 2 replaces the original shadow region in the copy of the original image.
4. The new corrected images are used in the training.

3.5 Experimental Procedure

The comparison of different models is done to observe the following factors:

Table 2. Details of Experiment

Model	Description	Total Patches
1	• Images are cropped into patches, Split ratio 80:20, • Without data augmentation	1816
2	• Images are cropped into patches, Split ratio 80:20, • Data augmentation (Shadow brightness variant)	5448
3	• Images are cropped into patches, Split ratio 80:20, • Data augmentation (Geometric Transformation)	5448

Initial Study 1: Models of single and multiple datasets are used to observe the effect of using a dataset with a different variety of drone images and total images used.

Initial Study 2: Effect of using train and validation sets of different split ratios. 80:20 and 90:10 ratios are selected, where a higher size of validation set is not further tested due to insufficient data for training. The model with the best result is labelled as Model 1 and the selected split ratio is used for Model 2 and Model 3 training. Both models are trained using the same number of training patches.

Model 2 & 3: Effect of different implementation of data augmentation methods on the models with the same size of dataset, split ratio, and parameters.

The details of the trained models are given in Table 2.

The model performance is measured by observing the loss and accuracy of the training and validation sets. In order to quantitatively evaluate the shadow detection performance, different metrics are used: overall accuracy (OA), F1-score as the weighted average of precision and recall, and Intersection Over Union (IoU) to measure the percentage of overlap between the ground truth and output mask [14]. As it is commonly used for shadow detection works [16], Balanced Error Rate (BER) is selected to observe the average of errors in unbalanced classes. For visual comparison, several test images with simple and complex shadow boundaries are selected and the effect of shadow complexity present in the images are discussed.

4 Results and Discussion

4.1 Initial Study

The first initial study is done to compare a model trained with one dataset source containing only 1309 patches and a model trained with all dataset sources (Model 1). Both training processes show that the models are a good fit for the dataset as there is a relatively small gap between training and validation for loss and accuracy (refer to Fig. 3 and Fig. 4). However, on epoch 9th during training of Model 1 (Fig. 4), validation loss and accuracy drop drastically, which may be due to insufficient variety of validation sets to validate the training. Overall, the training models have a good performance as both models reach a stable state at a point during training. It is recorded that the addition of

Table 3. Quantitative evaluation

Evaluation metrics	Model 1	Model 2	Model 3
Overall accuracy	0.9617	0.9555	0.9638
F1 score	0.3807	0.3748	0.3728
IoU	0.3265	0.3186	0.3212
BER	0.0307	0.0325	0.0295

dataset sources and the number of images increases the overall accuracy from 0.9524 to 0.9624 and reduces the BER from 0.0369 to 0.0319.

For the second initial study, Fig. 5 shows that model trained with a split ratio of 90:10 has a more volatile validation loss and accuracy compared to the training performance of Model 1 (split ratio of 80:20) as shown in Fig. 4. This indicates that the validation set from 90:10 ratio splitting is inadequate to evaluate the training process. The validation loss for the 90:10 model is lower than the training loss, showing a case of overfitting. The overall accuracies of shadow detection using a model trained with 90:10 split ratio and a model trained with 80:20 split ratio (Model 1) are 0.9624 and 0.9617 respectively. Based on this result, the split ratio selected for Model 2 and Model 3 are 80:20.

4.2 Augmentation Methods

For models trained with augmented samples (Model 2 and Model 3), the training performance has seemingly the same observation and patterns (refer to Fig. 6 and Fig. 7). It also can be observed that both models are not fully converged, which means they require a greater number of epochs compared to the model without augmentation. The comparison of the two models in Table 3 shows that models with geometric transformation augmentation (Model 3) have better performance compared to the model with shadow brightness variants (Model 2) based on overall accuracy, IoU, and BER. However, Model 2 shows a higher F1 score value compared to Model 3, which indicates that it slightly reduces the false predictions of shadows.

4.3 Visual Evaluation

4.3.1 Non-complex Images

As observed in Fig. 8, Model 2 can be seen to have the best prediction compared to other models. Prediction of the second test image using Model 3 highlighted that the model is the best at predicting shadows of different brightness. Model 3 is observed to detect clear shadow boundaries based on its prediction on the first test image shown in Fig. 8.

4.3.2 Complex Images

As observed in Fig. 9, Model 2 shows a good detection, but the presence of noise is more apparent compared to other models. Model 1 and Model 2 have many cases of

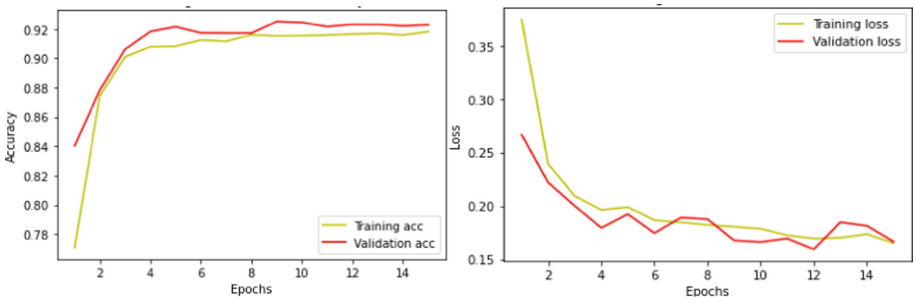


Fig. 3. Accuracy and loss of model trained with single dataset

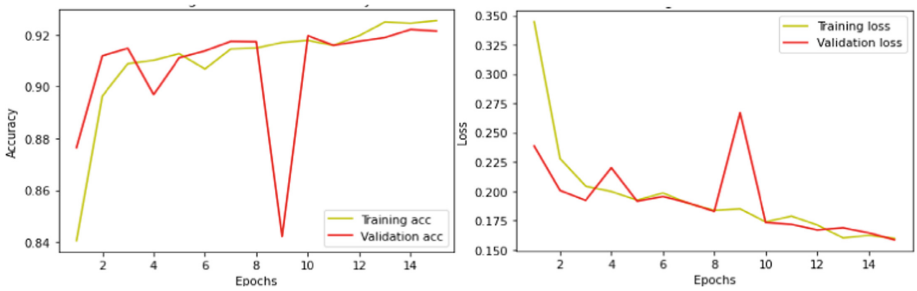


Fig. 4. Accuracy and loss of model trained with multiple datasets

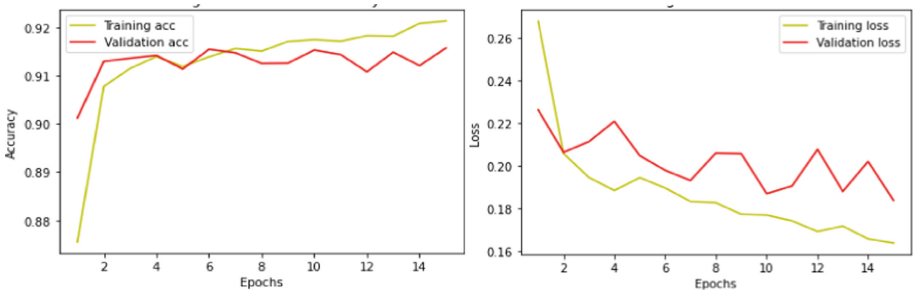


Fig. 5. Accuracy and loss of model trained with split ratio of 90:10

false positive detection, while Model 3 detected clear boundaries with a low case of false positive detection. It can be seen that Model 3 reduces the false positive errors better compared to other models. Similar to non-complex images, Model 2 and Model 3 have better predictions compared to Model 1, suggesting that both augmentation methods improve the model performance.

5 Discussion

Overall, the best model to detect shadow boundaries is Model 5, which is trained with augmented samples of geometric transformation. The improvement can prevent errors

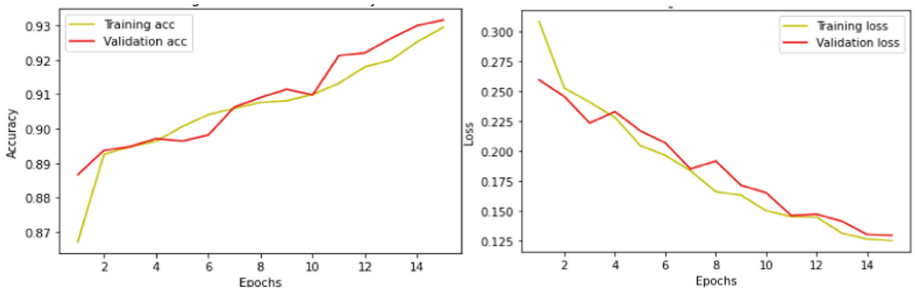


Fig. 6. Accuracy and loss of model trained with augmented samples of shadow brightness variant

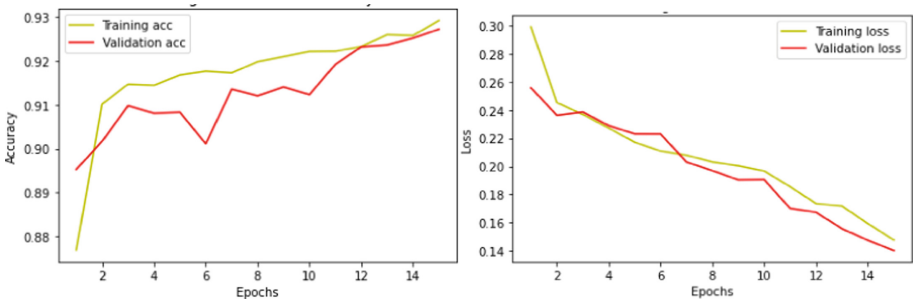


Fig. 7. Accuracy and loss of model trained with augmented samples of geometric transformation

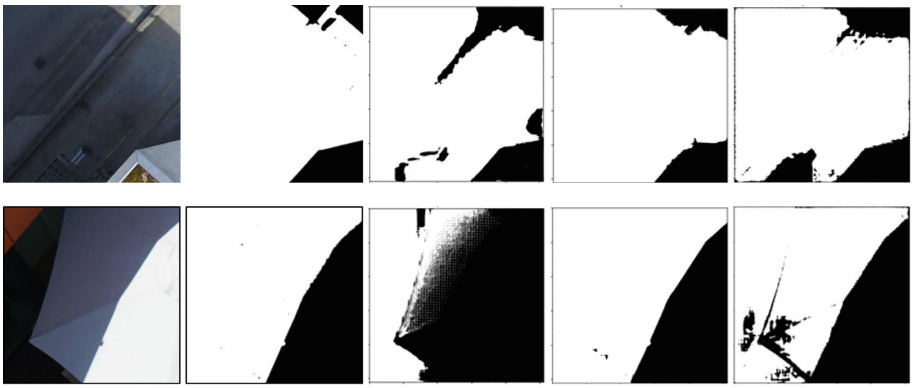


Fig. 8. Shadow detection results for a non-complex image: (**a**) Input image [7], (**b**) ground truth, (**c**) Model 3: Without augmented samples, (**d**) Model 4: Shadow brightness variant, (**e**) Model 5: Geometric transformation

in extracting shadow from an image that commonly coming from unclear boundaries. Meanwhile, Model 4 shows the flexibility to detect shadows in various brightness, shown in Fig. 8. This flexibility allows the model to suit in many different scenes under different weather and lighting. Observed from the overall results, predictions of all models have

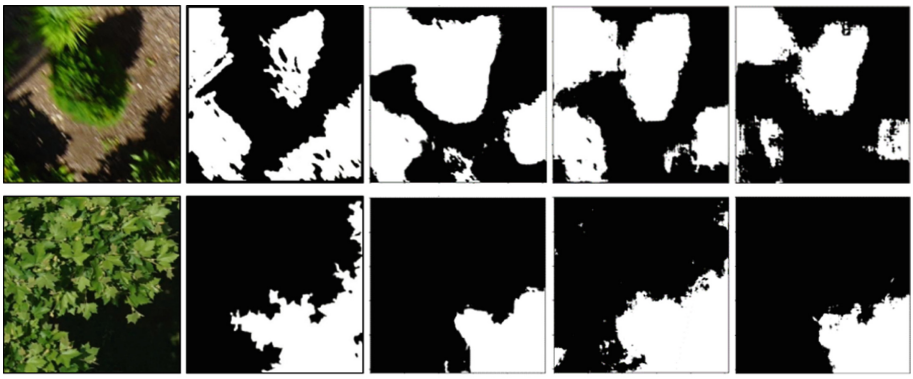


Fig. 9. Shadow detection results for a complex image: (**a**) Input image [8], (**b**) ground truth, (**c**) Model 3: Without augmented samples, (**d**) Model 4: Shadow brightness variant, (**e**) Model 5: Geometric transformation

high false negative and false positive regions. This reflects the quantitative evaluation here of the low values of IoU and F1 scores obtained in the overall results, which are less than 0.39. Low IoU value indicates that the models struggle to detect the true cases accurately compared to the ground truth. F1 score results are affected by the low recall and precision values. Another insight obtained from the results is increasing the size of the training dataset and using a variety of scenes in the images increases the accuracy of shadow detection, as discussed in Sect. 4.1.

Through these experiments, the results prove that the deep learning approach eases shadow detection works. Moreover, this paper also discussed the method to improve shadow detection in aerial images using data augmentation. While shadow detection work by [16] implemented a shadow brightness variant method in non-drone images, this paper was able to discuss the implementation of the method on drone images and the improvement that the method brings. The geometric transformation method was implemented separately to observe the difference in improvement it brings compared to the other method. Through analysis, this paper proves that both data augmentation methods improve the accuracy of detection and flexibility in detecting different brightness of shadow and complex shadow boundaries, as shown in Fig. 8 and Fig. 9. Hence, implementing the combination of both augmentation approaches, including post-processing such as binary mask refinement, is expected to provide better results.

6 Conclusion

In this work, to deal with limited labelled shadow data in aerial images, we implement several data augmentation methods. According to the results obtained from various tests, the following conclusions can be drawn. Using a higher number of train datasets and a variety of images is seen to improve the accuracy and flexibility of the training model. For the data augmentation method, the shadow brightness variant is proven to increase flexibility in detecting various brightness of shadow, while the geometric transformation

method is able to detect the boundary of shadow clearly. As noises are consistently present in the predictions due to the complexity of the shadows, our future work will contain but not limited to the refinement of the binary shadow mask. The refinement step is in progress.

Acknowledgments. This research is funded by the IR Fund of Multimedia University, Malaysia (MMUI/210095).

Authors' Contributions. Zali, S: Formal Analysis, Validation, Writing, Data Curation; M-Desa, S: Project Administration, Supervision, Data Curation, Review & Editing; Che-Embi, Z: Supervision, Review & Editing; Mohd-Isa, W: Review & Editing.

References

1. D. D. C. Trapal, B. C. C. Leong, H. W. Ng, J. T. G. Zhong, S. Srigrarom, and T. H. Chan, "Improvement of Vision-based Drone Detection and Tracking by Removing Cluttered Background, Shadow and Water Reflection with Super Resolution," *2021 6th Int. Conf. Control Robot. Eng. ICCRE 2021*, pp. 162–168, 2021, https://doi.org/10.1109/ICCRE51898.2021.9435671.
2. C. Gheorghe, C. Gheorghe, And N. Filip, "Image Processing Technique Used in Road Traffic Analysis – Opportunities and Challenges," *ACTA Tech. NAPOCENSIS - Ser. Appl. Math. Mech. Eng.*, vol. 64, no. 1-S2, Feb. 2021, Accessed: Oct. 13, 2021. [Online]. Available: https://atna-mam.utcluj.ro/index.php/Acta/article/view/1532.
3. N. N. Che'ya, E. Dunwoody, and M. Gupta, "Assessment of weed classification using hyperspectral reflectance and optimal multispectral UAV imagery," *Agronomy*, vol. 11, no. 7, Jul. 2021, https://doi.org/10.3390/AGRONOMY11071435.
4. D. Sharma and J. Singhai, "An Object-Based Shadow Detection Method for Building Delineation in High-Resolution Satellite Images," *PFG – J. Photogramm. Remote Sens. Geoinf. Sci. 2019 873*, vol. 87, no. 3, pp. 103–118, Jun. 2019, https://doi.org/10.1007/S41064-019-00070-3.
5. S. Luo, H. Li, and H. Shen, "Deeply supervised convolutional neural network for shadow detection based on a novel aerial shadow imagery dataset," *ISPRS J. Photogramm. Remote Sens.*, vol. 167, pp. 443–457, Sep. 2020, https://doi.org/10.1016/J.ISPRSJPRS.2020.07.016.
6. V. L. de S. Freitas, B. M. da F. Reis, and A. M. G. Tommaselli, "Automatic Shadow Detection in Aerial and Terrestrial Images," *Bol. Ciências Geodésicas*, vol. 23, no. 4, pp. 578–590, 2017, https://doi.org/10.1590/S1982-21702017000400038.
7. "Discover a wide range of drone datasets - senseFly." https://www.sensefly.com/education/datasets/ (accessed Nov. 17, 2021).
8. "ICG - DroneDataset." https://www.tugraz.at/index.php?id=22387 (accessed Nov. 17, 2021).
9. L. Garcia, J. Diaz, H. Loaiza Correa, and A.-D. Restrepo-Girón, "Thermal and visible aerial imagery," vol. 2, 2020, https://doi.org/10.17632/FFGXXZX298.2.
10. A. M. Deshpande, M. Gaikwad, S. Patki, A. Rathi, and S. Roy, "Shadow Detection from Aerial Imagery with Morphological Preprocessing and Pixel Clustering Methods," *ICTACT J. IMAGE VIDEO Process.*, p. 3, 2021, https://doi.org/10.21917/ijivp.2021.0340.
11. T. F. Y. Vicente, M. Hoai, and D. Samaras, "Leave-One-Out Kernel Optimization for Shadow Detection and Removal," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 03, pp. 682–695, Mar. 2018, https://doi.org/10.1109/TPAMI.2017.2691703.

12. X. Kang, Y. Huang, S. Li, H. Lin, and J. A. Benediktsson, "Extended Random Walker for Shadow Detection in Very High Resolution Remote Sensing Images," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 867–876, Feb. 2018, https://doi.org/10.1109/TGRS.2017.2755773.
13. S. Luo, H. Li, R. Zhu, Y. Gong, and H. Shen, "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.*, vol. 14, pp. 4633–4646, 2021, https://doi.org/10.1109/JSTARS.2021.3066791.
14. Y. Jin, W. Xu, Z. Hu, H. Jia, X. Luo, and D. Shao, "GSCA-UNet: Towards Automatic Shadow Detection in Urban Aerial Imagery with Global-Spatial-Context Attention Module," *Remote Sens. 2020, Vol. 12, Page 2864*, vol. 12, no. 17, p. 2864, Sep. 2020, https://doi.org/10.3390/RS12172864.
15. L. Jiao, L. Huo, C. Hu, and P. Tang, "Refined UNet: UNet-Based Refinement Network for Cloud and Shadow Precise Segmentation," *Remote Sens. 2020, Vol. 12, Page 2001*, vol. 12, no. 12, p. 2001, Jun. 2020, https://doi.org/10.3390/RS12122001.
16. H. Le, T. F. Y. Vicente, V. Nguyen, M. Hoai, and D. Samaras, "A+D Net: Training a Shadow Detector with Adversarial Shadow Attenuation," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11206 LNCS, pp. 680–696, Dec. 2017, Accessed: Oct. 14, 2021. [Online]. Available: https://arxiv.org/abs/1712.01361v2.
17. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 9351, pp. 234–241, 2015,* https://doi.org/10.1007/978-3-319-24574-4_28.