

Learning-Based Shadow Detection in Aerial Imagery Using Automatic Training Supervision from 3D Point Clouds

Deniz Kavzak Ufuktepe
dkyb5@mail.missouri.edu

Jaired Collins
jrcqw5@mail.missouri.edu

Ekincan Ufuktepe
euh46@missouri.edu

Joshua Fraser
jbf8cf@missouri.edu

Timothy Krock
tkrock@mail.missouri.edu

Kannappan Palaniappan
pal@missouri.edu

University of Missouri - Columbia
Electrical Engineering and Computer Science Dept., Columbia, MO, USA

Abstract

Shadows, motion parallax, and occlusions pose significant challenges to vision tasks in wide area motion imagery (WAMI) including object identification and tracking. Although there are many successful shadow detection approaches that work well in indoor scenes, close range outdoor scenes, and spaceborne satellite images, the methods tend to fail in intermediate altitude aerial WAMI. We propose an automatic shadow mask estimation approach using self-supervised learning without manual labeling to provide a large amount of training data for deep learning-based aerial shadow extraction. Analytical ground-truth shadow masks are generated using 3D point clouds combined with known solar angles. FSDNet, a deep network for shadow detection, is evaluated on aerial imagery. Preliminary results indicate that training using automated shadow mask self-supervision improves performance, and opens the door for developing new deep architectures for shadow detection and enhancement in WAMI.

1. Introduction

Deep shadows often degrade the performance of vision-based tracking systems in aerial imagery [33, 37]. As vehicles or pedestrians enter a shadow region where the sun is occluded by surrounding tall structures like buildings or trees, objects become difficult to detect and track. Figure 1 shows examples of complex and dark shadows in Albuquerque, New Mexico (ABQ) and Columbia, Missouri (COU) aerial WAMI datasets. Shadow segmentation can be used to improve ob-

ject tracking by adding resilience to abrupt appearance changes and occlusions [12]. Shadow mask regions can be selectively enhanced, in-filled, or combined with temporal imagery to reduce artifacts and improve visual quality. The critical step of shadow detection is a challenging problem for which training machine learning algorithms with supervision at city scale requires labeled pixel accurate shadow masks for thousands of objects over different time scales. However, generating such a large volume of manually ground-truthed shadows is impractical, expensive, and error prone. In this paper we explore automated methods for shadow generation when prior 3D scene models, like multiview point clouds or LiDAR data, are available.



Figure 1: Close look of Albuquerque, New Mexico at 7:10 PM UTC (left) and Columbia, Missouri at 8:50 PM UTC (right) shadow examples. The bottom row shows analytical automatically labeled shadows in red using Algorithm 1 for their respective images in the top row.

A number of successful shadow detection and shadow removal algorithms have been developed for close-up [23], indoors [20], surveillance [11], and satellite images [30, 54]. Although these methods can produce state-of-the-art results for their target domains, they fail for shadow detection in WAMI image sequences due to complexities caused by motion parallax, camera movement, occlusions, illumination changes, light filtering, partly transparent objects like vegetation, the scale of the image, and the small size of the objects of interest. An example failure case using the recent Fast Shadow Detection Network (FSDNet) deep learning method [23], is shown in Figure 2 tested on a subimage from the ABQ WAMI dataset. The FSDNet model was originally trained and performs well on a variety of natural images, but fails in complex WAMI data.

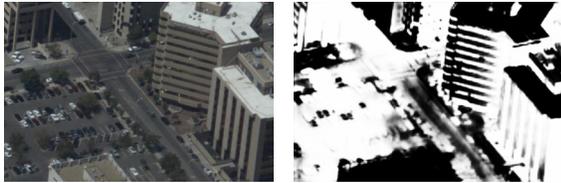


Figure 2: Example of weak inference using FSDNet trained on non-WAMI shadow data fails to detect shadows properly applied to a sample subimage from ABQ.

A few studies focus on shadow detection in aerial imagery and WAMI data, especially traditional methods [8, 40], while others use shadows for different analytics in aerial imagery without detection as the main focus [3, 4, 27, 38, 42]. Some [2, 43, 50] specifically consider the shadows to improve detection for vehicle tracking. A recent work provides a shadow mask dataset of aerial imagery [29] that consists of very small low resolution crops taken from aerial imagery with nadir views of satellite images. Yet there is no available annotated dataset for shadow detection or shadow removal in WAMI sequence data at city scale.

Our goal is to investigate the difficulties of shadow detection specifically in WAMI data compared to existing non-WAMI datasets and methods. For example, extracting a background model that is needed for some of the traditional methods is impossible in WAMI data. The natural complexity, scale, parallax, and moving camera in WAMI make traditional methods too complex for real-time applications such as tracking. Due to the lack of WAMI shadow data and the inherent difference between WAMI data and existing shadow datasets, existing deep convolutional neural networks perform poorly for shadow detection on WAMI data. To address these

problems and to leverage the power and speed of deep learning methods, we make the following contributions: (i) developing an automated shadow mask generation pipeline using a 3D point cloud, accurate camera orientations, and foundational geospatial data to produce high volume data for learning-based shadow detection methods, and (ii) performing a case study to train a deep learning shadow detection network using the analytic shadow masks to evaluate the quality of the shadow segmentation. The preliminary results of the shadow mask generation are promising and show that training with the analytic shadow masks improves deep neural model performance that is originally trained on non-WAMI data.

2. Related Work

2.1. Deep Learning Methods for Shadow Masks

Deep neural architectures have emerged as a powerful learning model and the detection of natural shadows is no exception. Different visual tasks such as robotic navigation [7], vehicle tracking in WAMI [13, 15, 17, 32, 34], and multiview 3D reconstruction [5] can use shadow detection to improve performance. Deep learning-based methods can be accelerated for real-time low memory footprint applications like robotics and visual tracking with a tradeoff in performance [21]. Fast Shadow Detection Network (FSDNet) by Hu et al. [23] is a state-of-the-art method but does not perform well on aerial imagery. It is unclear if deep neural networks perform poorly due to the lack of sufficient city-scale labeled shadow mask training data, training protocol or limitations of the architecture.

Semi-supervised approaches can be used with unlabeled data but usually need a much larger amount of data [14]. Recent papers contribute a new dataset in addition to their trained neural network [23, 29], which are also shown to be more accurate than a popular state-of-the-art algorithm, bidirectional feature pyramid network with recurrent attention residual modules for shadow detection (BDRAR) [53].

FSDNet [23] was shown to be more accurate and faster than other state-of-the-art shadow segmentation methods, and so was chosen to be used in our case study. FSDNet uses MobileNetV2 trained on ImageNet as a backbone [36]. After the backbone, a direction-aware spatial module (DSC) [22, 24] is combined with a detail enhancement module (DEM). The DEM and DSC find and concatenate low, mid, and high level features, followed by a series of convolution layers operating on the found features to predict shadow masks. For training, the only data augmentation used was a horizontal flip.

The aforementioned deep learning-based shadow detection models are trained with close-up and satellite images causing the pre-trained weights to fail in detecting WAMI shadows accurately. These proposed networks are also designed according to the characteristics of non-WAMI data and fail to adequately address issues specific to shadows in urban WAMI data.

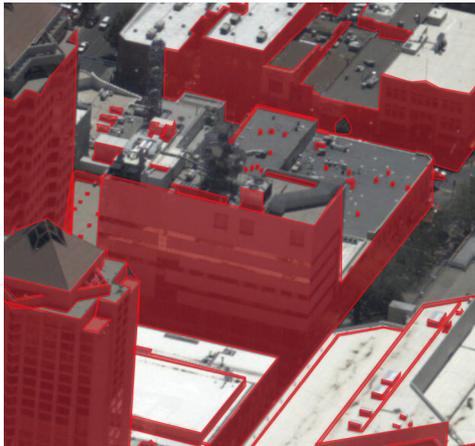


Figure 3: Subjective manually labeled area of downtown ABQ, New Mexico at 7:10 PM UTC. Structures are complicated and cast many umbra and penumbra shadows. This crop took about 2 hours to label using the Computer Vision Annotation Tool (CVAT) [39] and represents just 1.3% of one image at 6600×4400 pixels.

2.2. Other Shadow Detection Methods

Non-deep learning methods often use photometric models, background subtraction or change detection. For background subtraction, pixels can be classified as background, motion, or shadow [20, 26, 46]. However, this method does not work when the camera pose changes every frame unless egomotion is accounted for. Statistical distributions including Gaussian mixture models (GMM) can be used as well [26]. Multispectral approaches such as the near-infrared band in Landsat imagery and channel ratios can be used to predict shadow locations of clouds [1, 54], but not buildings due to resolution. A guided filter, tricolor attenuation model, intensity information, and infilling can produce good masks [10]. Decision trees can also be used to achieve impressive results, but the dataset used was not made publicly available [8]. Additional methods for shadow detection include histogram thresholding, invariant color models, object segmentation, geometrical methods, and physics-based methods [1, 49].

Due to the unique characteristics of WAMI data, these methods either fail or perform very poorly in aerial

Table 1: Popular shadow detection datasets.

Dataset	# Images	Sizes
UCF [52]	245	Various
UIUC [18]	108	Various
SBU [44]	4727	285×177 to 800×600
ISTD [45]	1870	640×480
AISD [29]	514	256×256 to 1688×1688
CUHK [23]	10,500	115×107 to 4800×3872
Our WAMI	137,180	Tiles 660×440

urban scenes. It is difficult to develop a traditional method that is both fast and accurate for WAMI data due to its complexity. These other methods may also have high time complexity or energy use on large imagery which is often critical for realtime onboard tracking. Therefore, learning-based approaches should be evaluated for shadow detection in WAMI to have embedded real-time, or near real-time performance.

2.3. Datasets for General Shadow Detection

Training datasets with a variety of images and corresponding labeled shadow masks are rare, and have been a bottleneck in developing deep neural models for shadow detection. A few, exemplar datasets like the University of Central Florida (UCF) [52] and University of Illinois Urbana-Champaign (UIUC) [18] although extensively used have only a few hundred labeled images. As interest grew in this field, more publications released datasets with labeled shadows for larger number of images: Stony Brook University (SBU) [44], Image Shadow Triplets (ISTD) [45], Aerial Imagery dataset for Shadow Detection (AISD) [29], and Chinese University of Hong Kong (CUHK) [23], summarized in Table 1. UCF and UIUC have a relatively small number of images [18, 52], for which SBU created a much larger collection with 4727 images [44]. The images in UCF, UIUC, and SBU usually focus on a single object plus shadow and not a scene. AISD was created for shadows on a larger scale, like a city, but the images are all orthorectified satellite images, offering no oblique views [29]. CUHK is a very recent dataset that features many close-up scenes, often complex but are general in its theme and not guaranteed to be urban [23].

Although the lack of large shadow detection datasets with ground-truth has been addressed by several groups, they are often application specific [18, 23, 29, 44, 45, 52]. This work focuses on aerial WAMI sequence data with oblique views, and training using these existing datasets was insufficient to detect shadows in WAMI. Manual ground-truthing requires too much effort due to complicated structures as shown in Figure 3. As an alternative

to complex and time-consuming methods for generating WAMI shadow mask training data, there is a need to efficiently create labeled shadow segmentations in real aerial imagery that can be used to train and validate learning-based shadow detection algorithms.

Synthetic Scene Generation for Labeled Shadow Maps Synthetic data generation has advantages over other traditional datasets. Since it offers a fully controlled environment, 3D scene objects and lighting can be set explicitly, allowing generation of photorealistic rendering with soft and hard shadows relatively easily. Several recent methods exist for synthetic data generation [25, 28, 51], where most of the data generated feature shadows placed in small scenes and not city-scale. Recent approaches [9, 31] produce synthetic city-scale simulations for testing computer vision algorithms.

In this paper we take a hybrid approach combining real urban 3D point clouds with analytic shadows for automatic self-supervised (noisy) labels to provide more realistic training data that can be scaled geospatially.

3. Geometric Shadow Masks for WAMI

3.1. Automatic Shadow Mask Generation

Our analytical geometric shadow mask algorithm utilizes a dense 3D point cloud reconstructed from *city-scale* (WAMI) sequences, sun position, and precise camera poses and camera intrinsics to automatically estimate shadows to provide self-supervised training data. The dense 3D point clouds are estimated using Yao et al. [48], and the bundle adjusted camera poses for each view use the BA4S method by Aliakbarpour et al. [6].

The sun location is calculated from metadata acquired during the data collect containing the precise time and GPS location of the scene.¹ Solar position is described by two angles: azimuth A as an angle measured clockwise with respect to North, and elevation α as the vertical angle between the sun and the horizon as shown in Figure 4 and given in Eq. (1) and (2):

$$\alpha = \sin^{-1} [\sin \delta \sin \phi + \cos \delta \cos \phi \cos(\text{HRA})] \quad (1)$$

$$A = \cos^{-1} \left[\frac{\sin \delta \cos \phi - \cos \delta \sin \phi \cos(\text{HRA})}{\cos \alpha} \right] \quad (2)$$

where δ is the latitude, ϕ the declination angle, and HRA is the hour angle converted from local solar time [35]. Hour angle is based on the Earth’s 15° rotation per hour away from solar noon and is negative in the morning, positive in the afternoon, and zero at solar noon.

¹<https://github.com/CIVA-Lab/solar-position-calculator>

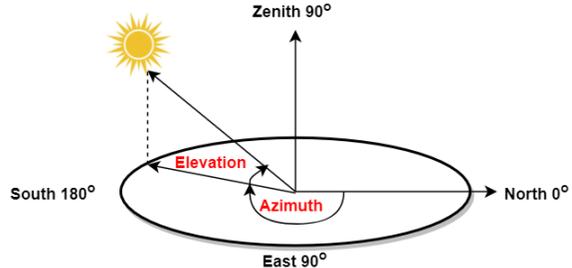


Figure 4: Azimuth and elevation angles for measuring solar position with respect to local north.

Our *Projected Geometric Shadow*² algorithm (Algorithm 1) is based on the classical two-pass shadow map technique from computer graphics [47]. Depth maps encode the distance to the nearest visible fragment for a view into a two-dimensional array initialized to infinity. The first pass determines per-fragment solar occlusion by rendering the point cloud orthographically to a depth buffer (*sun_map*) as shown in Figure 5a whose coordinate system L is aligned to the solar angle. The second pass renders a depth map for the point cloud from a perspective precisely matched to an original view by the complete camera matrix (current view) in Eq. (3):

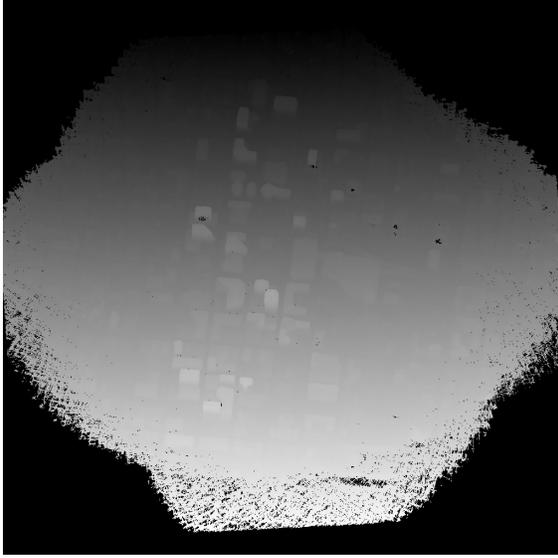
$$C = K [R|t] \quad (3)$$

where K contains the camera intrinsics including focal length, principal point, and optional skew, R is the camera orientation, and t the translation of the camera in R by the convention of [19]. This second pass first tests if a fragment is visible in the current view C using a depth map aligned to the original image. If the fragment is not occluded in that view, its position is found in L and tested against the *sun_map* to determine whether or not the visible fragment is in shadow. This final depth test utilizes a scene dependent *bias* parameter to reduce noise resulting from self-shadowing and the limited precision of the depth buffers in scene units (default 0.005). The point cloud is rendered as voxels to depth and mask fragments using the recursive method described in [16].

3.2. Post-processing of Shadow Masks

After shadow masks are generated using the analytical approach given in Section 3.1, we apply post-processing steps to improve the quality of the masks. The quality of the generated shadow masks is directly affected by the quality of the 3D point cloud. Holes and noise in the 3D point cloud corresponds to the holes and

²<https://github.com/CIVA-Lab/depthshadowmask>



(a) Solar depth map ($2K \times 2K$) from L aligned to the solar vector.



(b) Point cloud rendered from view C matched to frame 775.



(c) Visible depth map for view C used to compute shadow mask.

Figure 5: Two pass projected geometric shadow algorithm. (a) Solar depth map resulting from Pass 1 of Algorithm 1 using a rotated orthographic projection of the a priori 3D point cloud. (b) ABQ 3D point cloud aligned with view 775. (c) Non-shadowed depth from Eq. 3 used in Pass 2 to compute the shadow mask.

Algorithm 1: Projected Geometric Shadows (PGS) – Depth map-based analytical shadow mask generation

Input: Array of 3D Points, Camera matrix $C = K[R|t]$, Depth bias, Orthographic projection matrix L aligned with solar vector

Output: Shadow Mask

```

/* Pass 1: Generate an orthographic
depth map (sun_map) of the entire
scene aligned with solar vector */
for each 3D point  $p$  in point cloud do
     $p_l \leftarrow L \cdot p$ ; // project  $p$  to solar view
    if  $p_l.z < sun\_map[p_l.xy]$  then
        |  $sun\_map[p_l.xy] \leftarrow p_l.z$ ;
    end
end
end

/* Pass 2: Determine solar occlusion of
all visible points */
for each 3D point  $p$  in point cloud do
     $p_v \leftarrow C \cdot p$  // project  $p$  to view
     $p_v.xy \leftarrow p_v.xy / p_v.z$ ; // perspective
    // if point is visible in C
    if  $p_v.z < depth[p_v.xy]$  then
        |  $depth[p_v.xy] \leftarrow p_v.z$  // update depth
        |  $p_l \leftarrow L \cdot p$  // project  $p$  into  $L$ 
        | if  $sun\_map[p_l.xy] < p_l.z - bias$  then
            | |  $mask[p_v.xy] \leftarrow shadow$ ;
        | end
    end
end
end

```

noise in the shadow masks. Since the 3D point cloud is not perfect, the predicted shadow masks also have imperfections. Some examples of problems include very small shadows of small objects, shadows of short trees and shrubs, holes in buildings that do not have sufficiently dense 3D points, and other errors caused by missing points in the point cloud.

For filtering and improvement of the generated masks, two steps of morphological operations on the generated masks followed by smoothing are performed. To remove the small, noisy shadows in the masks, we apply morphology operations on the predicted mask to remove the small objects. This operation calculates the size of each connected component in the mask and removes the components that have a smaller size than the threshold. This threshold is selected empirically as 5. Following the removal of small components, to fill the holes in the buildings caused by missing points in the

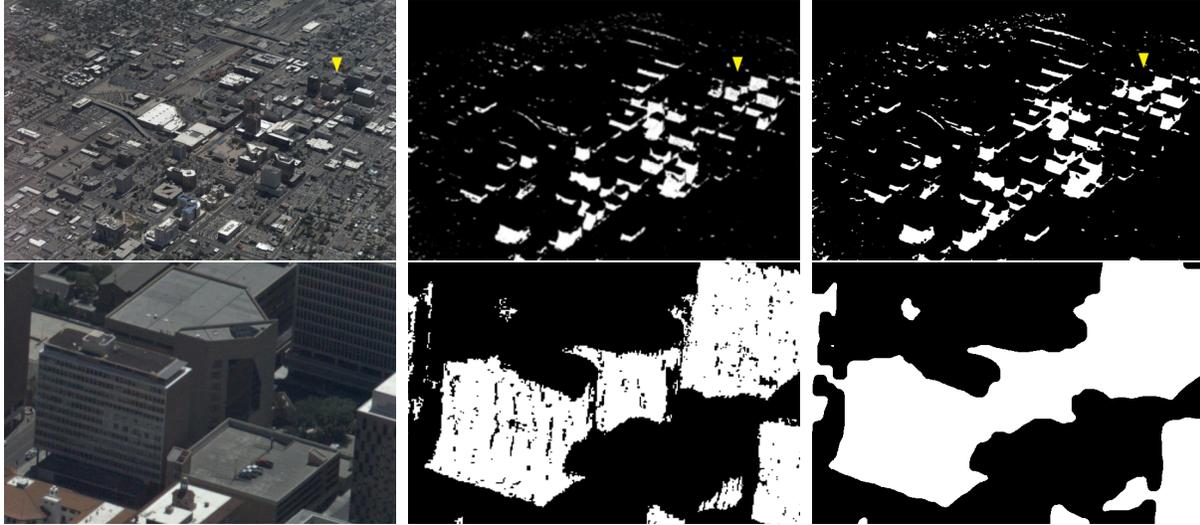


Figure 6: Automatic shadow mask estimation example. Left to right: An original image from the ABQ sequence, the geometry-based automatically generated (synthetic) shadow masks, and after post-processing. First row shows the full image resolution, and the second row shows a zoomed region of interest (yellow arrows in top row).

3D point cloud, a morphological binary closing is applied to the processed masks. The closing operator performs dilation followed by erosion to close the gaps in the masks of the connected components. The structure element used in the closing operator is a 20 by 20 kernel, which is also determined experimentally. Finally, processed masks are smoothed by using a median filter of size 11x11, chosen empirically.

The post-processing steps applied on the generated shadow masks make them more suitable for training and evaluation purposes, emphasizing the shadows of taller structures, like buildings and trees, which are more important for a model to learn for shadow detection in WAMI data. Such shadows greatly affect image processing and computer vision tasks for WAMI such as detection, tracking, and 3D visualization [33,37]. An example of an image from ABQ and a crop, corresponding generated shadow masks, and post-processed shadow masks are given in Figure 6.

4. Case Study Using FSDNet

4.1. Datasets for Aerial Shadow Detection

Two large city-scale aerial image sequences, Albuquerque (ABQ) and Columbia (COU), provided by Transparent Sky [41] were used. Each WAMI frame is 6600x4400 pixels. The high-resolution images were captured using a manned aircraft flying a circular path around the downtown area of each city. Figure 7 shows ~72% of the image sequence that was used for training

(red dots), ~9% for validation (yellow dots), and ~19% for testing (green dots) the deep learning shadow detection network. Table 2 shows the total number of images and cropped tiles used for training, validation, and testing using aerial WAMI. To train FSDNet, overlapping image tiles were cropped from each image in the training, validation, and test image sets. In each image, 361 overlapping cropped tiles each of size 660x440 with the associated synthetically generated analytic shadow masks were used as labeled ground-truth for training. This provided a total of 137,180 image tiles.

Table 2: Training, validation, and testing datasets. Camera poses are full circular orbits around the downtown area of Albuquerque, NM (ABQ), and Columbia, MO (COU). Image sizes: 6600x4400. Total tiles: 137,180.

Dataset	#Images	#Train. Images	#Valid. Images	#Test Images
ABQ	215	155	20	40
COU	165	120	15	30
Tiles	137,180	99,275	12,635	25,270

4.2. Shadow Detection Using FSDNet

We use a recent deep learning network for shadow detection, FSDNet [23], that we retrained using transfer learning with aerial WAMI images and the self-supervised shadow masks described in the next section. First, a FSDNet model was trained using CUHK, which is a non-WAMI dataset, and was used in the original

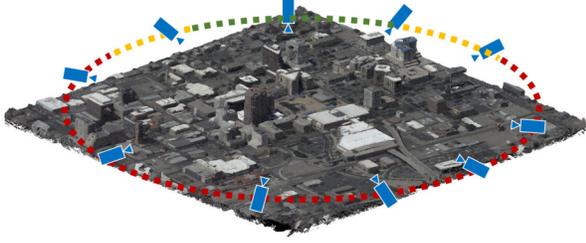


Figure 7: Visualization of the camera flight path for the ABQ WAMI city-scale dataset. Red dots indicate training views, yellow dots indicate validation views, and green dots indicate the testing views.

FSDNet paper [23]. This baseline model was trained for 50,000 iterations, starting with ImageNet weights for the MobileNetV2 [36] backbone, and we refer to it as FSDNet-nonWAMI. Using transfer learning with FSDNet-nonWAMI weights, we trained an additional 50,000 iterations (about 12 epochs), using the aforementioned WAMI data and shadow masks. The training is done on an Nvidia Quadro RTX 8000 GPU with a batch size of 64. The loss function is L1 (Mean Absolute Error) with stochastic gradient descent optimization.

Qualitative assessment of the preliminary FSDNet shadow detection results on aerial imagery is shown in Figure 8 for sample regions of interest comparing the performance of FSDNet-nonWAMI to FSDNet-WAMI. FSDNet-nonWAMI model contains a lot of false positive shadows in the predicted mask for WAMI images, due to the differences between non-WAMI and WAMI data, like scale, angle, and the complexity of the shadows. Significant improvement in accuracy, both area and shape of the shadow, using the self-supervised FSDNet-WAMI shadow detection network is clearly evident. However, shadow detection accuracy using FSDNet-WAMI still needs further improvement.

Quantitative accuracy of the detected shadows using FSDNet is compared to the analytic shadows using average Dice and average mean squared error (AMSE),

$$\text{Average Dice} = \frac{1}{N} \sum_{i=0}^{N-1} \frac{2|X_i \cap Y_i|}{|X_i| + |Y_i|} \quad (4)$$

$$\text{AMSE} = \frac{1}{N} \sum_{i=0}^{N-1} \left(\frac{1}{D} \sum_{j=0}^{D-1} (X_{ij} - Y_{ij})^2 \right) \quad (5)$$

where X_i represents the predicted shadow mask of the i^{th} test image and Y_i represents the analytic ground-truth shadow mask of the i^{th} test image. In Eq (4), the

intersection over union of all N test images are averaged. In Eq (5), i, j represents the j^{th} pixel of the corresponding i^{th} test image. AMSE is calculated as pixel-wise squared error between the predicted real valued and analytically generated masks, then averaged over the N test images. MSE is widely used to evaluate the error in model predictions across a range of applications, while the Dice similarity coefficient is commonly used in binary segmentation and classification evaluation.

The AMSE and average Dice quantitative evaluation of shadow detection accuracy using the testing data in Table 3 shows the significant improvement in performance of FSDNet with self-supervision. The test set uses images cropped from the ABQ and COU datasets (see Table 2). While the AMSE is low, it is dominated by the background non-shadow pixels and consequently, the average Dice values are also very low. The performance of FSDNet-WAMI can be further improved by increasing the number of training epochs, additional data augmentation, using WAMI images from different times of day with shadows of different length, and using improved photorealistic ray tracing and global illumination for hard and soft shadow mask training data.

Table 3: Test data results for average MSE error and average Dice similarity coefficient values on the inferences of the retrained FSDNet-WAMI model on crops of test images. Total of 137,180 tiles and tile size: 660×440 .

Model	Avg. MSE ↓	Avg. Dice ↑
FSDNet-nonWAMI	0.29	0.192
FSDNet-WAMI	0.07	0.376

5. Conclusions

We developed the analytic Projected Geometric Shadows algorithm for self-supervised learning to detect shadows in aerial urban imagery. This avoids the need for manual labeling of thousands of shadow regions in city-scale WAMI data. The proposed algorithm requires an accurate 3D point cloud, precise time and GPS flight metadata to calculate the solar position, bundle adjusted camera poses, and intrinsics to automatically generate the shadow segmentation masks in aerial imagery that closely match the real image shadows. The analytic shadows can be noisy but are of sufficient accuracy to improve the performance of deep learning models. The shadow detection accuracy of FSDNet nearly doubled in terms of the Dice overlap ratio from 19.2% to 37.6%, when trained with self-supervised analytic aerial shadows from ABQ and COU WAMI datasets. The

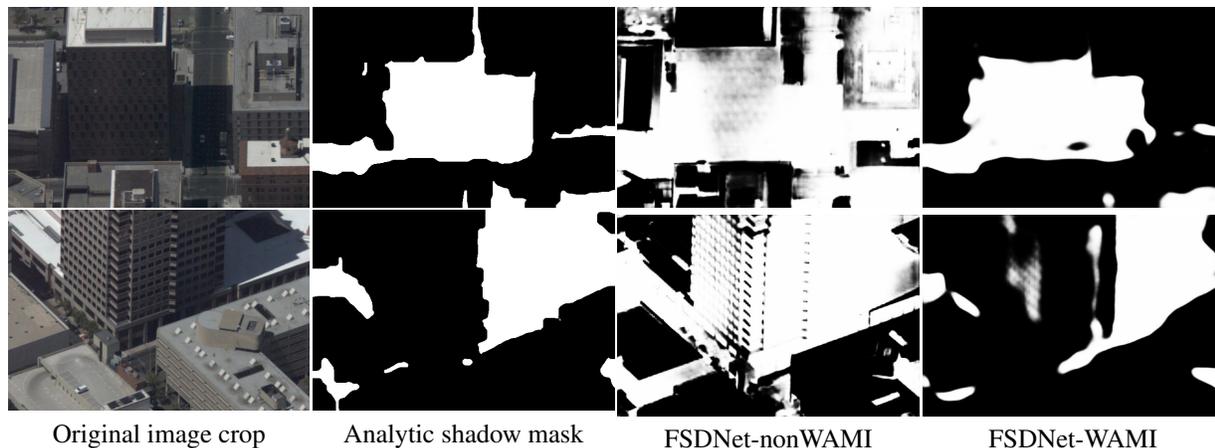


Figure 8: Performance of FSDNet for shadow detection with self-supervision. Crops from ABQ showing original image, analytic post-processed shadow mask, inference using FSDNet-nonWAMI model and FSDNet-WAMI model.

accuracy and generalization capability of FSDNet can be improved further by: modifying the training protocols, selecting alternate hyperparameters, utilizing data augmentation, supplementing with aerial imagery from additional cities, improving the accuracy of 3D point clouds, applying adversarial networks, using photorealistic synthetic scenes and shadows for greater variation in the training data, or improving the analytic shadow algorithm. The results show that current state-of-the-art shadow detection deep networks trained for different applications are not sufficiently accurate for shadow detection in aerial imagery, even when trained with WAMI data. This highlights the need to enhance current deep architectures (like the FSDNet backbone) or develop new deep neural models to detect shadows in aerial imagery due to its unique characteristics.

6. Acknowledgments

The dense 3D point clouds were generated by Shizeng Yao using the Multiscale Plane Sweep (MuPS) algorithm. Research partially supported by U.S. Army Research Laboratory cooperative agreement W911NF-1820285 and Army Research Office DURIP W911NF-1910181. The raw wide area motion imagery (WAMI) aerial datasets for ABQ and COU were collected by TransparentSky. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the U. S. Government or agency.

References

[1] K.R.M. Adeline, M. Chen, X. Briottet, S.K. Pang, and N. Paparoditis. Shadow detection in very high spatial reso-

lution aerial images: A comparative study. *ISPRS Journal of Photogrammetry and Remote Sensing*, 80:21–38, 2013.

- [2] Chad Aeschliman, Johnny Park, and Avinash C. Kak. Tracking vehicles through shadows and occlusions in wide-area aerial video. *IEEE Transactions on Aerospace and Electronic Systems*, 50(1), 2014.
- [3] Noor Al-Shakarji, Filiz Bunyak, Hadi Aliakbarpour, Guna Seetharaman, and Kannappan Palaniappan. Multi-cue vehicle detection for semantic video compression in georegistered aerial videos. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019.
- [4] Noor M. Al-Shakarji, Filiz Bunyak, Guna Seetharaman, and Kannappan Palaniappan. Robust multi-object tracking for wide area motion imagery. In *IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, 2018.
- [5] H. Aliakbarpour, Joao F. Ferreira, V. B. Surya Prasath, Kannappan Palaniappan, Guna Seetharaman, and Jorge Dias. A probabilistic framework for 3D reconstruction using heterogeneous sensors. *IEEE Sensors*, 17(9):2640–2641, May 2017.
- [6] Hadi Aliakbarpour, Kannappan Palaniappan, and Guna Seetharaman. Robust camera pose refinement and rapid SfM for multiview aerial imagery – without RANSAC. *IEEE Geoscience and Remote Sensing Letters*, 12(11), 2015.
- [7] Hadi Aliakbarpour, V. B. Surya Prasath, Kannappan Palaniappan, Guna Seetharaman, and Jorge Dias. Heterogeneous multi-view information fusion: Review of 3-D reconstruction methods and a new registration with uncertainty modeling. *IEEE Access*, 4(1):8264–8285, 2016.
- [8] Gilberto Alvarado-Robles, Roque A. Osornio-Rios, Francisco J. Solis-Munoz, and Luis Alberto Morales-Hernandez. An approach for shadow detection in aerial images based on multi-channel statistics. *IEEE Access*, 9, 2021.

- [9] Brendan Alvey, Derek Anderson, Andrew Buck, Matthew Deardorff, Grant Scott, and James Keller. Simulated photorealistic deep learning framework and workflows to accelerate computer vision and unmanned aerial vehicle research. In *ICCV Workshop on Analysis of Aerial Motion Imagery*, 2021.
- [10] Benish Amin, M. Mohsin Riaz, and Abdul Ghafoor. Automatic shadow detection and removal using image matting. *Signal Processing*, 170:107415, 2020.
- [11] C. Benedek and T. Sziranyi. Bayesian foreground and shadow detection in uncertain frame rate surveillance videos. *IEEE Transactions on Image Processing*, 17(4):608–621, 2008.
- [12] Erik Blasch, Guna Seetharaman, Steve Sudderth, Kannappan Palaniappan, Genshe Chen, Haibin Ling, and Arlsan Basharat. Summary of methods in wide-area motion imagery (wami). *Geospatial InfoFusion and Video Analytics IV; and Motion Imagery for ISR and Situational Awareness II*, 2014.
- [13] Sema Candemir, Kannappan Palaniappan, Filiz Bunyak, and Guna Seetharaman. Feature fusion using ranking for object tracking in aerial imagery. In *Proc. SPIE Conf. Geospatial InfoFusion II (Defense, Security and Sensing: Sensor Data and Information Exploitation)*, volume 8396, 2012.
- [14] Zhihao Chen, Lei Zhu, Liang Wan, Song Wang, Wei Feng, and Pheng-Ann Heng. A multi-task mean teacher for semi-supervised shadow detection. *IEEE Conference on Computer Vision and Pattern Recognition*, 2020.
- [15] Ilker Ersoy, Kannappan Palaniappan, Guna Seetharaman, and Raghuvveer Rao. Interactive tracking for persistent wide-area surveillance. In *Proc. SPIE Conf. Geospatial InfoFusion II (Defense, Security and Sensing: Sensor Data and Information Exploitation)*, volume 8396, 2012.
- [16] Ke Gao, Hadi Aliakbarpour, Joshua Fraser, Koundinya Nouduri, Filiz Bunyak, Ricky Massaro, Guna Seetharaman, and Kannappan Palaniappan. Local feature performance evaluation for structure-from-motion and multi-view stereo using simulated city-scale aerial imagery. *IEEE Sensors Journal*, 21(10), 2021.
- [17] A. Gning, W. T. L. Teacy, R. Pelapur, H. AliAkbarpour, K. Palaniappan, G. Seetharaman, and S. J. Julier. The effect of state dependent probability of detection in multi-target tracking applications. In *Proc. SPIE Conf. Geospatial InfoFusion and Video Analytics IV; and Motion Imagery for ISR and Situational Awareness II (Defense, Security and Sensing)*, volume 9089, 2014.
- [18] Ruiqi Guo, Qieyun Dai, and Derek Hoiem. Single-image shadow detection and removal using paired regions. *IEEE Conference on Computer Vision and Pattern Recognition*, 2011.
- [19] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2nd edition, 2003.
- [20] Thanarat Horprasert, David Harwood, and Larry Davis. A statistical approach for real-time robust background subtraction and shadow detection. *IEEE International Conference on Computer Vision*, 1999.
- [21] Sepideh Hosseinzadeh, Moein Shakeri, and Hong Zhang. Fast shadow detection from a single image using a patched convolutional neural network. *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2018.
- [22] Xiaowei Hu, Chi-Wing Fu, Lei Zhu, Jing Qin, and Pheng-Ann Heng. Direction-aware spatial context features for shadow detection and removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(11), 2020.
- [23] Xiaowei Hu, Tianyu Wang, Chi-Wing Fu, Yitong Jiang, Qiong Wang, and Pheng-Ann Heng. Revisiting shadow detection: A new benchmark dataset for complex world. *IEEE Transactions on Image Processing*, 2021.
- [24] Xiaowei Hu, Lei Zhu, Chi-Wing Fu, Jing Qin, and Pheng-Ann Heng. Direction-aware spatial context features for shadow detection. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018.
- [25] Naoto Inoue and Toshihiko Yamasaki. Learning from synthetic shadows for shadow detection and removal. *IEEE Transactions on Circuits and Systems for Video Technology*, 2020.
- [26] Pakorn KaewTraKulPong and Richard Bowden. An improved adaptive background mixture model for real-time tracking with shadow detection. *Video-Based Surveillance Systems*, 2002.
- [27] Rodney LaLonde, Dong Zhang, and Mubarak Shah. Clusternet: Detecting small objects in large scenes by exploiting spatio-temporal information. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018.
- [28] Zhihao Liu, Hui Yin, Zhenyao Wu, Yang Mi, and Song Wang. From shadow generation to shadow removal. *IEEE Conference on Computer Vision and Pattern Recognition*, 2021.
- [29] Shuang Luo, Huifang Li, and Huanfeng Shen. Deeply supervised convolutional neural network for shadow detection based on a novel aerial shadow imagery dataset. *ISPRS Journal of Photogrammetry and Remote Sensing*, 167, 2020.
- [30] Uttam K. Majumder, Erik Blasch, and David A. Garren. *Deep Learning for Radar and Communications Automatic Target Recognition*. Artech House, 2020.
- [31] Koundinya Nouduri, Ke Gao, Joshua Fraser, Shizeng Yao, Hadi AliAkbarpour, Filiz Bunyak, and Kannappan Palaniappan. Deep realistic novel view generation for city-scale aerial images. In *International Conference on Pattern Recognition (ICPR)*, pages 10561–10567, 2021.
- [32] Kannappan Palaniappan, Filiz Bunyak, Praveen Kumar, Ilker Ersoy, S. Jaeger, K. Ganguli, A. Haridas, J. Fraser, R. Rao, and G. Seetharaman. Efficient feature extraction and likelihood fusion for vehicle tracking in low frame rate airborne video. In *13th Int. Conf. Information Fusion*, 2010.
- [33] Kannappan Palaniappan, Mahdieh Poostchi, Hadi Aliakbarpour, Raphael Viguier, Joshua Fraser, Filiz Bunyak,

- Arslan Basharat, Steve Sudderth, Erik Blasch, Raghuveer M. Rao, and Guna Seetharaman. Moving object detection for vehicle tracking in wide area motion imagery using 4D filtering. *International Conference on Pattern Recognition*, 2016.
- [34] Jing Peng, Kannappan Palaniappan, Sema Candemir, and Guna Seetharaman. Kullback-leibler divergence-based data fusion for target tracking. In *Int. Conf. Pattern Recognition*, pages 3480–3483, 2012.
- [35] Ibrahim Reda and Afshin Andreas. Solar position algorithm for solar radiation applications. *Solar Energy*, 76(5), 2004.
- [36] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018.
- [37] Andres Sanin, Conrad Sanderson, and Brian C. Lovell. Shadow detection: A survey and comparative evaluation of recent methods. *Pattern Recognition*, 45(4):1684–1695, 2012.
- [38] Varun Santhaseelan and Vijayan Asari. Tracking in wide area motion imagery using phase vector fields. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2013.
- [39] Boris Sekachev et al. OpenCV/CVAT: v1.1.0, <https://github.com/opencv/opencv4/cvat>, Aug. 2020.
- [40] Guilherme F. Silva, Grace B. Carneiro, Ricardo Doth, Leonardo A. Amaral, and Dario F. G. de Azevedo. Near real-time shadow detection and removal in aerial motion imagery application. *ISPRS Journal of photogrammetry and remote sensing*, 140, 2018.
- [41] Transparent Sky. <https://transparensky.net/>. [Online; Accessed on July 22, 2021].
- [42] Michael Teutsch and Michael Grinberg. Robust detection of moving vehicles in wide area motion imagery. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2016.
- [43] Elena Vella, Anee Azim, Han X. Gaetjens, Boris Repasky, and Timothy Payne. Improved detection for wami using background contextual information. In *Digital Image Computing: Techniques and Applications (DICTA)*, 2019.
- [44] Tomás F. Vicente, Le Hou, Chen-Ping Yu, Minh Hoai, and Dimitris Samaras. Large-scale training of shadow detectors with noisily-annotated shadow examples. *European Conference on Computer Vision*, 2016.
- [45] Jifeng Wang, Xiang Li, and Jian Yang. Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. *IEEE Conference on Computer Vision and Pattern Recognition*, 2018.
- [46] Rui Wang, Filiz Bunyak, Guna Seetharaman, and Kannappan Palaniappan. Static and moving object detection using flux tensor with split gaussian models. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2014.
- [47] Lance Williams. Casting curved shadows on curved surfaces. In *ACM Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques*, 1978.
- [48] Shizeng Yao, Hadi AliAkbarpour, Guna Seetharaman, and Kannappan Palaniappan. 3D patch-based multi-view stereo for high-resolution imagery”. *SPIE Geospatial Informatics, Motion Imagery, and Network Analytics VIII*”, 10645, 2018.
- [49] Q. Ye, H. Xie, and Q. Xu. Removing shadows from high-resolution urban aerial images based on color constancy. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXIX-B3, 2012.
- [50] Meng Yi, Dong Wang, Fan Yang, Jonathan Xu, Yiran Cai, Erik Blasch, Carolyn Sheaff, Genshe Chen, and Haibin Ling. Vehicle pose estimation in WAMI imagery via deep convolutional neural networks. In *IEEE National Aerospace and Electronics Conference (NAECON) and Ohio Innovation Summit (OIS)*, 2016.
- [51] Shuyang Zhang, Runze Liang, and Miao Wang. ShadowGAN: shadow synthesis for virtual objects with conditional adversarial networks. *Computational Visual Media*, 5(1), 2019.
- [52] Jiejie Zhu, Kegan G. Samuel, Syed Z. Masood, and Marshall F. Tappen. Learning to recognize shadows in monochromatic natural images. *IEEE Conference on Computer Vision and Pattern Recognition*, 2010.
- [53] Lei Zhu, Zijun Deng, Xiaowei Hu, Chi-Wing Fu, Xuemiao Xu, Jing Qin, and Pheng-Ann Heng. Bidirectional feature pyramid network with recurrent attention residual modules for shadow detection. In *European Conference on Computer Vision*, 2018.
- [54] Zhe Zhu and Curtis E. Woodcock. Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sensing of Environment*, 118, 2012.