Remote Sensing of Environment 247 (2020) 111945

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Recurrent Shadow Attention Model (RSAM) for shadow removal in high-resolution urban land-cover mapping

Yindan Zhang[a,b,c], Gang Chen[c,*], Jelena Vukomanovic[b,d], Kunwar K. Singh[b,e], Yong Liu[a], Samuel Holden[c], Ross K. Meentemeyer[b,f]

[a] College of Earth and Environmental Sciences, Lanzhou University, Lanzhou 730000, China
[b] Center for Geospatial Analytics, North Carolina State University, Raleigh, NC 27695, USA
[c] Laboratory for Remote Sensing and Environmental Change (LRSEC), Department of Geography and Earth Sciences, University of North Carolina at Charlotte, NC 28223, USA
[d] Department of Parks, Recreation and Tourism Management, North Carolina State University, Raleigh, NC 27695, USA
[e] Global Research Institute, AidData, The College of William and Mary, Williamsburg, VA 23185, USA
[f] Department of Forestry and Environmental Resources, North Carolina State University, Raleigh, NC 27695, USA

ARTICLE INFO

Keywords:
Shadow removal
Urban land-cover mapping
High resolution
Recurrent Shadow Attention Model (RSAM)
Urban development patterns
Deep learning

ABSTRACT

Shadows are prevalent in urban environments, introducing high uncertainties to fine-scale urban land-cover mapping. In this study, we developed a Recurrent Shadow Attention Model (RSAM), capitalizing on state-of-the-art deep learning architectures, to retrieve fine-scale land-cover classes within cast and self shadows along the urban-rural gradient. The RSAM differs from the other existing shadow removal models by progressively refining the shadow detection result with two attention-based interacting modules – Shadow Detection Module (SDM) and Shadow Classification Module (SCM). To facilitate model training and validation, we also created a Shadow Semantic Annotation Database (SSAD) using the 1 m resolution (National Agriculture Imagery Program) NAIP aerial imagery. The SSAD comprises 103 image patches (500 × 500 pixels each) containing various types of shadows and six major land-cover classes – building, tree, grass/shrub, road, water, and farmland. Our results show an overall accuracy of 90.6% and Kappa of 0.82 for RSAM to extract the six land-cover classes within shadows. The model performance was stable along the urban-rural gradient, although it was slightly better in rural areas than in urban centers or suburban neighborhoods. Findings suggest that RSAM is a robust solution to eliminate the effects in high-resolution mapping both from cast and self shadows that have not received equal attention in previous studies.

1. Introduction

The presence of shadows is frequent in high-resolution remote sensing imagery of urban landscapes due to tall objects, both natural (e.g., trees) and human-made (e.g., buildings) (Fig. 1). Depending on the source, shadows can be categorized into cast shadows and self shadows (Arévalo et al., 2008). Cast shadows are caused by tall objects in the vicinity blocking the light source, while self shadows arise from the object surface not being directly illuminated by the light source (Su et al., 2016). Shadows can reduce the urban heat island effect providing outdoor thermal comfort (Lin et al., 2010) and can serve as a clue for building identification (Shackelford and Davis, 2003). However, they are generally considered as a nuisance obscuring details of fine-scale geographic objects (Dare, 2005), and hence pose a significant challenge for accurate land-cover mapping. More specifically, shadows reduce spectral radiance in the shaded landscape, which makes detection of spectral and spatial features within shadows using a high-resolution imagery a problematic task (Wang et al., 2017). Typically, it is also common for shadows to be misclassified as water or water-related land-cover types (e.g., wetland, marsh, etc.) due to high similarities in spectral signatures and texture (Kang et al., 2017).

To deal with the undesirable shadow effects on high-resolution imagery of urban landscapes, researchers for land cover mapping have two strategies, such as treating shaded areas as one single 'shadow class' and correcting the distorted radiance. The implementation of the former approach is straightforward. However, shadows are not a real land-cover type, and therefore, mapped products are highly erroneous, making the product useless for many urban studies. The second strategy capitalizes on the advancement of inpainting, a process of re-constructing missing or damaged areas of digital photographs and

* Corresponding author.
E-mail address: gang.chen@uncc.edu (G. Chen).

https://doi.org/10.1016/j.rse.2020.111945
Received 5 March 2020; Received in revised form 15 May 2020; Accepted 8 June 2020

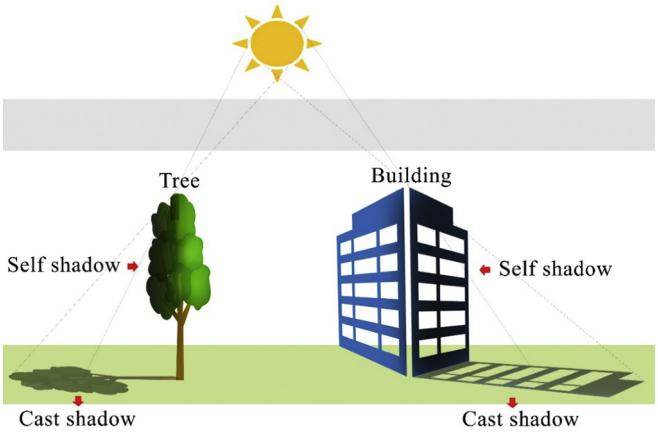

Fig. 1. This is a simple example of shadows caused by a built (e.g., a building) and natural (e.g., a tree). When light is blocked by the object projected on the ground is called cast shadows while projected on the object surface itself is called self shadows.

videos (Buyssens et al., 2015). This is applied in two steps: shadow detection and shadow removal, often stated as de-shadowing or shadow correction/compensation process, to eliminate shadows in remote-sensing-based land-cover mapping. Precisely, shadow detection locates shadow pixels. Wide range of models have been developed to consider the sensor-sun-object geometry for accurately simulating shadow locations (e.g., Arévalo et al., 2008; Li et al., 2004). While promising, those models are often tied to specific scene conditions and require a priori knowledge of the viewing geometry that is not always available (Zhang et al., 2014). To overcome these limitations, property-based models have been developed to identify the spectral contrast between the sunlit land cover and their shaded counterparts using thresholding (e.g., Chen et al., 2007; Milas et al., 2017; Shedlovska and Hnatushenko, 2019), shadow index (e.g., Mostafa and Abdelhafiz, 2017; Zhang et al., 2014), segmentation (e.g., Azevedo et al., 2019; Mo et al., 2018), and classification (e.g., Kang et al., 2019). Those models are typically calibrated on a scene-by-scene basis to achieve optimal results. Once the shadow-contaminated pixels are located, their spectral radiance is enhanced to simulate the corresponding sunlit condition through shadow removal procedure. The basic concept behind the majority of shadow removal models is to correct the spectral difference between the sunlit and the shaded pixels representing the same or similar surface materials, such as using histogram or region matching (e.g., Sarabandi et al., 2004; Shedlovska and Hnatushenko, 2019), illumination correction (Luo et al., 2019; Zhang et al., 2015a), linear correlation correction (Chang and Tsay, 2010; Chen et al., 2007), and gamma correction (Jain and Khunteta, 2017; Massalabi et al., 2004). Relying on a single image to correct shadow effects in high-resolution imagery has been a standard practice, while data integration from multiple sources and/or dates has drawn increasing attention (Zhang et al., 2014).

Recent studies have reported 85–95% accuracies for correcting shadow effects in high-resolution remote sensing imagery (e.g., Jain and Khunteta, 2017; Luo et al., 2019; Qiao et al., 2017; Silva et al., 2018). While promising, those models require carefully defining thresholds, parameters, or image features for optimized performance on individual scenes over specific types of urban neighborhoods. Compared to traditional methods, deep-learning-based methods for shadow processing have proven effective in natural images, such as using Generative Adversarial Network (Hu et al., 2018) and attention-based Convolutional Neural Networks (Ding et al., 2019; Zhu et al., 2018). However, these methods failed to process high-resolution remote sensing imagery that has a wide spectral range, varying types of shadows, and sophisticated features (Wang et al., 2017). In addition, previous studies has emphasized cast shadow corrections (e.g., Adeline et al., 2018; Arévalo et al., 2008; Zhang et al., 2014), while self shadows have received less attention. Compared to cast shadows, self shadows exhibit distinct spectral variation and semantic features, e.g., highly fragmented shaded tree canopies (Chen et al., 2011) that entail unique consideration. Another challenge lies in the two-step sequential process for shadow elimination is the lack of a mechanism to correct errors from shadow detection once they propagate to the succeeding shadow removal process.

Based on those considerations, the goal of this study is to develop an accurate model to eliminate shadow effects in high-resolution urban land-cover mapping. Our model aims to (i) retrieve fine-scale land-cover classes within cast and self shadows, and (ii) generate reliable results across various types of urban neighborhoods. To do so, we used state-of-the-art deep learning (DL; LeCun et al., 2015) architectures for high-performance semantic segmentation of urban land cover within cast and self shadows. Further inspired by the bidirectional recurrent neural networks (Schuster and Paliwal, 1997) and the attention mechanism in DL (Vaswani et al., 2017), we designed the model structure to progressively refine the shadow detection result preventing the propagation of significant errors to the retrieval of land-cover classes in the shadow.

2. Study area

The Research Triangle metropolitan area of North Carolina (4030 km^2, Fig. 2) is located in the southern Piedmont physiographic region and characterized by the rolling landscapes and mosaics of oak-pine-hickory forests adjoining to the Atlantic Coastal Plain region. The Raleigh-Durham-Chapel Hill CSA (Combined Statistical Area) has a population of over 2.2 million (Census, 2020). With high-tech enterprises advancing the “new economy,” the three major cities Raleigh, Durham, and Chapel Hill and the surrounding towns all expanded tremendously. The rapid growth and sprawling patterns typify the land uses of various neighborhood types (e.g., residential vs. commercial) of urban areas representing different development densities (e.g., high-density urban centers vs. low-density rural regions). The diverse land-cover classes and spatially heterogeneous patterns in the region were ideal for us to incorporate a wide variety of representative shadow types in model development.

3. Data

3.1. NAIP (National Agriculture Imagery Program) imagery

The NAIP (National Agriculture Imagery Program) images covering the study area were downloaded from the USGS Earth Explorer data portal (USGS, 2019). Original NAIP images were taken during the leaf-on seasons from 2016 to 2018 at the 1.0 m spatial resolution with four spectral bands (blue - 400-580 nm; green - 500-650 nm; red - 590-675 nm; and near-infrared - 675-850 nm). The images were orthorectified with data quality inspected before being delivered by the vendor.

3.2. Shadow Semantic Annotation Database (SSAD)

We created a Shadow Semantic Annotation Database (SSAD) from the collected NAIP images for the training and validation of the model proposed in this study. The SSAD has a total of 103 image patches containing various types of shadows along the urban-rural gradient of the Research Triangle region while balancing the distribution of the six major land-cover classes (e.g., building, tree, grass/shrub, road, water, and farmland). The size of each image patch is 500 × 500 pixels. Two categories of annotation in the SSAD are: (i) shadow annotation (i.e., shadow and non-shadow) for straightforward shadow detection; and (ii) land-cover annotation including the six land-cover classes within

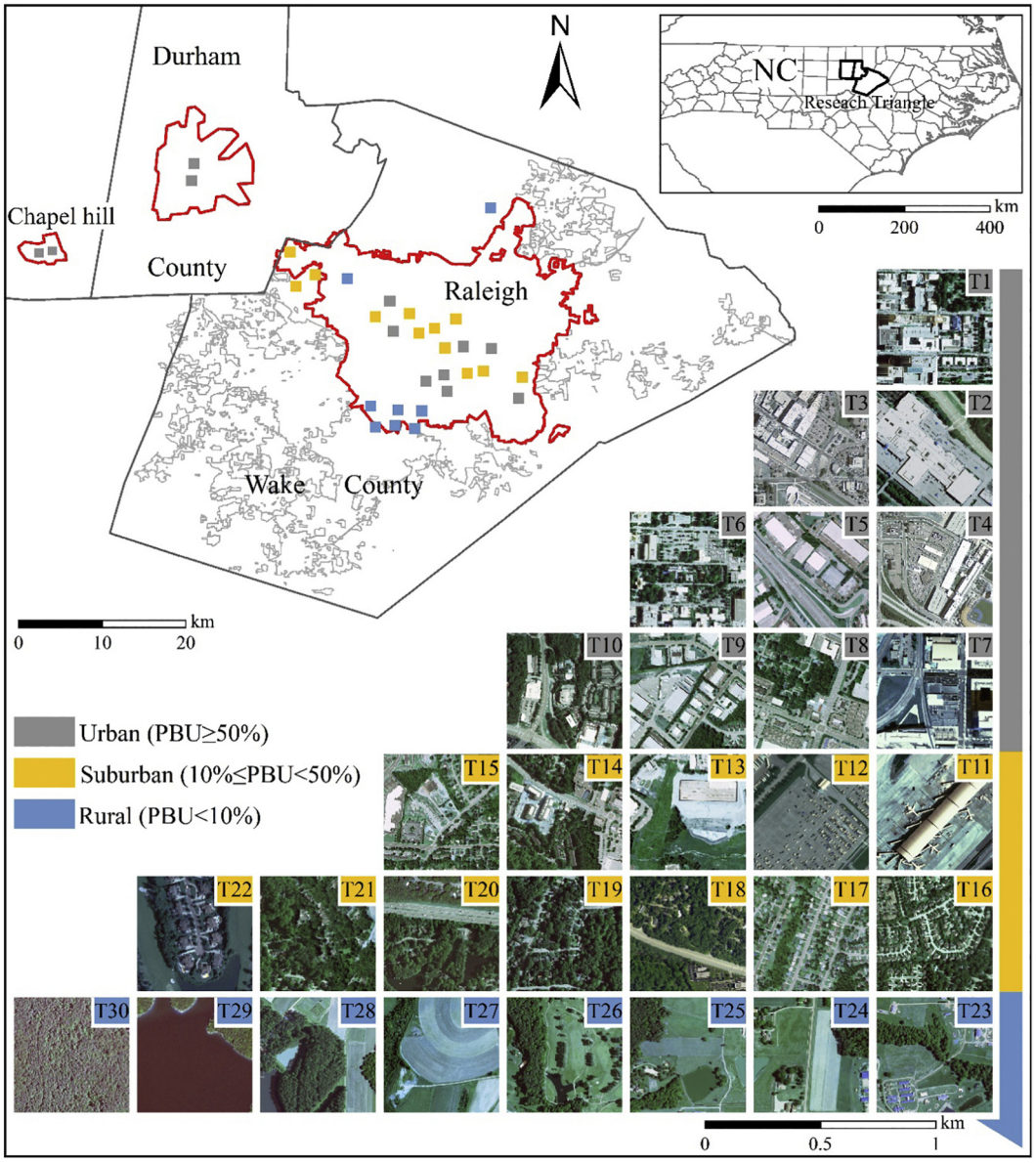

Fig. 2. Research Triangle is a rapidly urbanizing region with a spatially heterogeneous land cover that ranges from high-density city centers to low-density rural areas (PBU: Percent Built-Up). The thirty sample neighborhoods are shown in insets T1-T30.

shadows. Sample image patches and the corresponding semantic annotations are shown in Fig. 3.

We utilized transfer learning (Torrey and Shavlik, 2009) and manual correction to annotate 103 image patches. For transfer learning, we employed SegNet (Badrinarayanan et al., 2015) to produce pre-annotated sample patches, using two available semantic databases: (i) SBU Shadow Dataset (Tomas et al., 2016) for shadow annotation, and (ii) ISPRS 2D Semantic Labeling Dataset (ISPRS, 2019a, 2019b) for land cover semantic annotation. With this process, we were able to quickly obtain initial sample patches. However, because the two databases were not specifically developed for the same research purpose as ours, noises and errors were prevalent in the obtained annotations. To improve the quality, at each training process, we randomly selected five annotated sample patches, carefully conducted manual correction, and put them back for further learning. Through an incremental process, we gradually corrected all the sample patches to ensure the high annotation quality in the SSAD. Here, the Oxford's renowned VGG-16 network architecture was initially fine-turned to the annotation network, mainly due to its proven success in image object recognition and the publicly available network structure and weights for ease-of-use (Zhang et al., 2015b).

4. Methods

4.1. Recurrent Shadow Attention Model (RSAM)

4.1.1. Overview

Our proposed Recurrent Shadow Attention Model (RSAM) retrieves fine-scale urban land-cover classes within shadows (Fig. 4) and is comprised of two interacting modules: Shadow Detection Module (SDM) and Shadow Classification Module (SCM) (Fig. 4a). The SDM was intended to focus primarily on cast shadows, capitalizing on a dilated Convolutional Neural Network (CNN), and a dual attention network (i.e., position and channel attention), while the SCM was designed to give more attention to self shadows. Using two sequential convolutional Long Short-Term Memory (LSTM) units, SDM and SCM reciprocally influenced further learning, which agglomerated over epochs to refine the shadow and the shaded land-cover feature attention in a

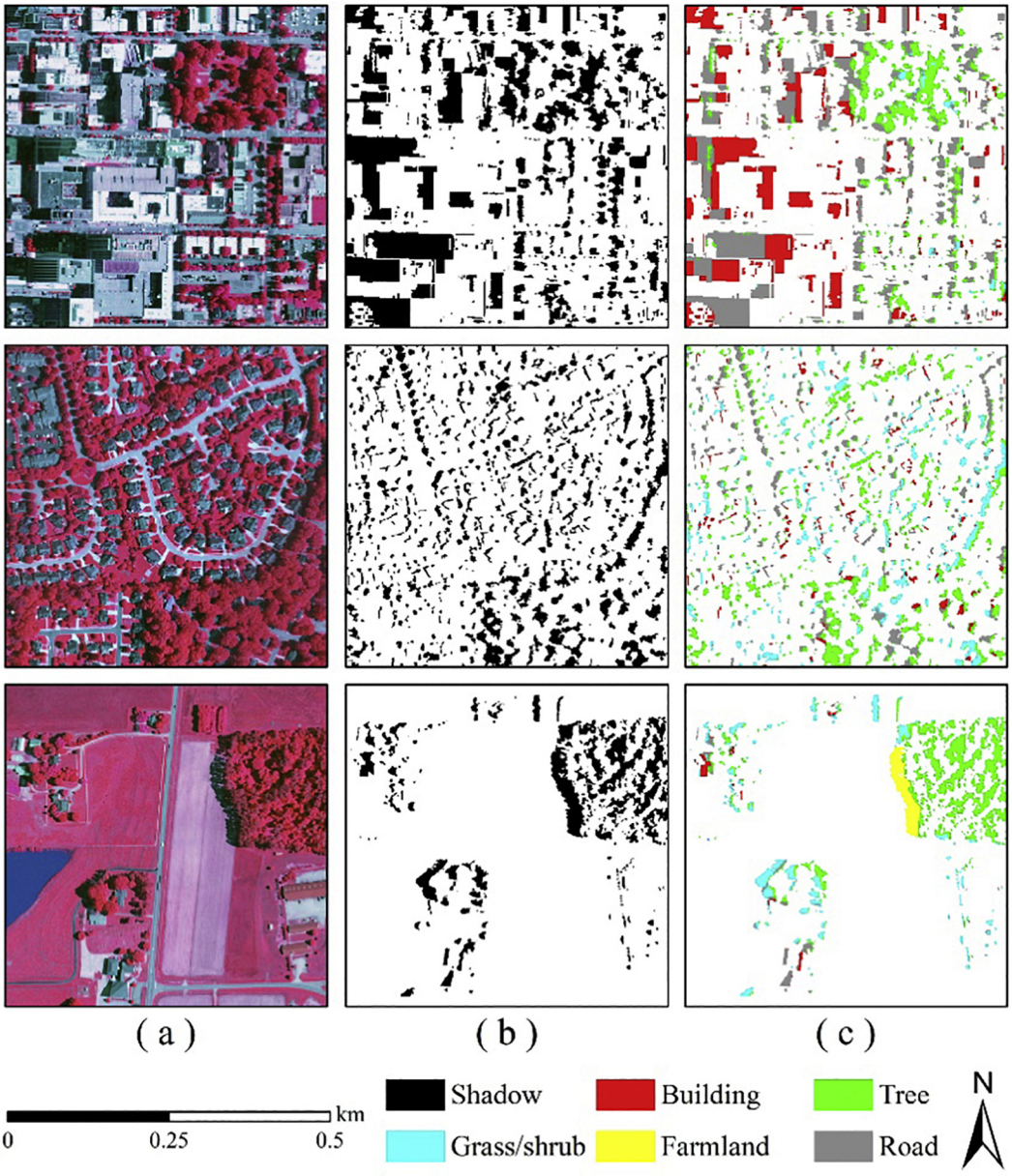

Fig. 3. Three sample patches along an urban-rural gradient from the SSAD: (a) NAIP Infrared-Red-Green image composites, (b) SSAD Category (i) results, and (c) SSAD Category (ii) results. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Recurrent Neural Network (RNN) architecture (Fig. 4b). After the refinement, both the final shadow attention and the shaded land-cover feature attention were bidirectionally fed into SegNet to retrieve detailed land-cover classes within shadows (Fig. 4c).

4.1.2. Shadow Detection Module (SDM)

To obtain the best local and global shadow feature representations for pixel-level prediction, SDM used a dilated SegNet (Badrinarayanan et al., 2015) architecture due to its robust performance in semantic segmentation (Audebert et al., 2017; Hamida et al., 2017; Jiang et al., 2020; Panboonyuen et al., 2017; Zhang et al., 2019), which was applied to detect shadow versus non-shadow areas in the NAIP imagery. The novelty of the proposed SDM lies in the integration of the position (spatial) and the channel (band) attention via the Convolutional Block Attention Module (CBAM) (Woo et al., 2018). The choice of the module was mostly due to its effectiveness in enhancing shadow contextual dependencies in the scene segmentation task (Fu et al., 2019), which ensured the key advantages of SDM: (i) channel attention was used to understand 'what' is meaningful in the high-resolution imagery by exploiting the inter-channel relationship of shadow features; and (ii) position attention allowed the channel attention to focus on local semantic information through prioritizing the shadow areas, i.e., 'where' is an informative part.

The encoder of SDM performed convolution using a filter bank to generate a series of shadow feature maps, comprising five convolution blocks. Feature maps at the shallower layers encoded the fine details that helped to preserve the shadow boundaries, while feature maps at the deep layers carried global semantics that helped to recognize the shadow and non-shadow regions. In pre-training, SDM started initially with the weights of the inceptions-v4 net (Szegedy et al., 2017) pre-trained with ImageNet, and all the weights were fine-tuned using the SSAD Category (i) training data (Section 3.2). Compared with the other pre-trained networks, e.g., ResNet-50 (Targ et al., 2016) and VGGNets (Wang et al., 2015), Inception-v4 balanced efficiency and performance well (Szegedy et al., 2017). With a very deep convolutional network, it was especially suitable for high-resolution shadow context recognition

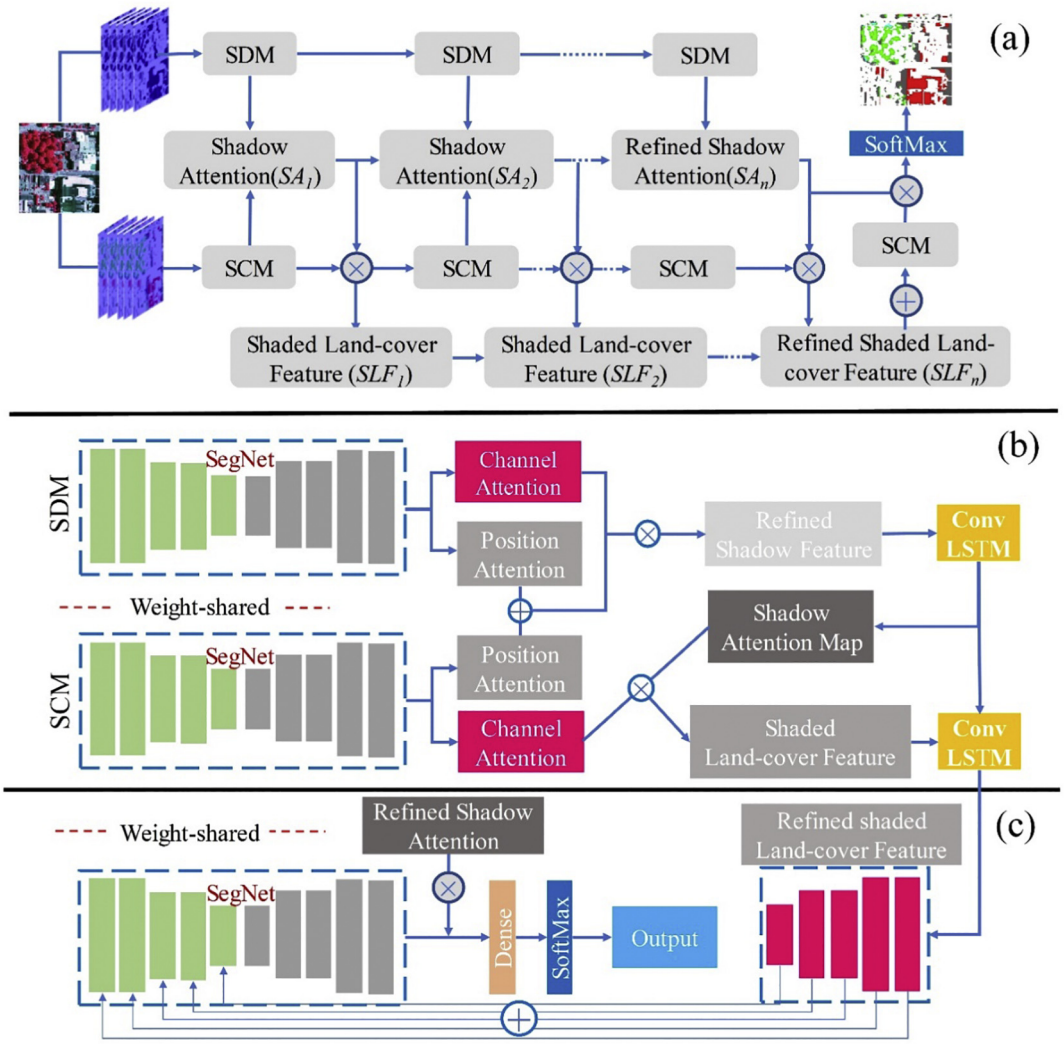

Fig. 4. (a) RSAM workflow; (b) the detailed structure for generating recurrent shadow attention maps with two interacting modules – the Shadow Detection Module (SDM) and the Shadow Classification Module (SCM); (c) the structure for retrieving fine-scale shaded land-cover classes using shadow attention map, shaded land-cover attention map, and SegNet.

due to the proven success of fitting databases with a similar size as SSAD (Liu and Deng, 2015). At the end of the encoder, we generated the final shadow feature maps based on the features aggregated at multiple layers. The decoder of SDM unsampled and reconstructed the shadow feature maps by five symmetrical deconvolution blocks with respect encoder, and utilized the memorized max-pooling operation of the corresponding encoder feature maps. CBAM was then integrated into our SDM architectures to compute the channel and the position attention.

4.1.3. Shadow Classification Module (SCM)

Self shadows are highly variable, containing objects of different sizes, shapes, and spatial patterns. While the SDM may work well for cast shadows that are often connected, it is challenging to characterize self shadows, which tend to be fragmented (Fu et al., 2019). Hence, the land-cover contextual information within self shadows is often different from those in cast shadows. To address this issue, SCM adapted the same neural network structure as SDM. Consequently, the parameters and pre-training remained the same. However, SCM was fine-tuned using the SSAD Category (ii) training data (Section 3.2) to produce position and channel attention maps. This is because the Category (ii) data included detailed land-cover classes within shadows. Our preliminary evaluations found that such fragmented land-cover information helped detect and remove self shadows at similar fragmentation levels. Compared to SDM, SCM has two unique operations: (a) assigning the position attention to update the shadow attention of SDM, i.e., an element-wise sum operation on the above-resulting position attention of SDM and SCM, and (b) refining shaded land-cover features within the shadow attention map obtained in the first steps, i.e., a matrix multiplication between the two attention maps. The purpose was to refine the shadow attention map, and suppress the non-shaded land-cover contextual information to be introduced into the final feature maps.

We used the two sequential convolutional LSTM units (Graves and Schmidhuber, 2005) to progressively refine the shadow attention map for developing the RSAM (Fig. 4a). At each epoch, we employed a convolutional LSTM unit to adaptively generate a recurrent shadow attention map relying on the shadow position attention from SDM and SCM, which consisted of two stages: (i) producing shadow attention map by a 1×1 convolutional layer with stride 1, and (ii) progressing as memory in the RASM. The integrated position attention map was then fed into SDM and SCM in the next epoch to improve the estimation of shadow cover and land-cover classes within shadows, respectively. The process completed at the end of the epochs. The shadow attention map was a matrix ranging from 0 to 1, rather than a binary mask. The larger the value, the more attention should be given to this region.

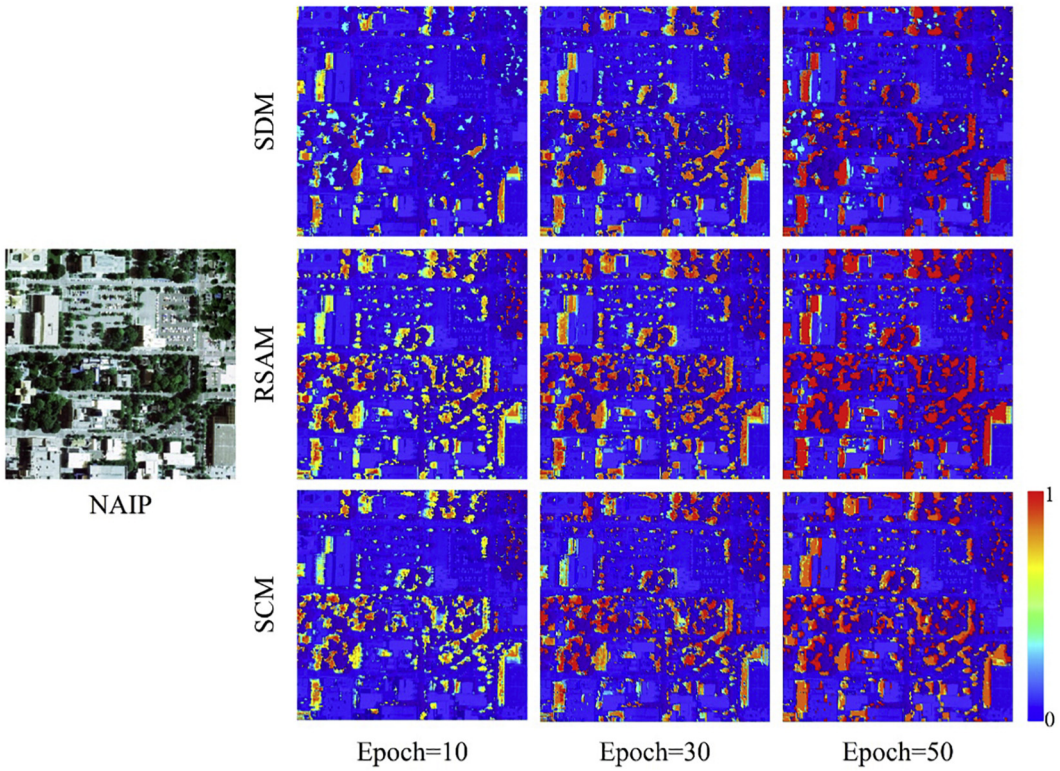

Fig. 5. Visualization of sample shadow attention maps generated by one NAIP image at epochs of 10, 30, and 50 using three proposed modules: top - Shadow Detection Module (SDM), middle - Recurrent Shadow Attention Model (RSAM), bottom - Shadow Classification Module (SCM).

Here, the attention mechanism was employed to direct the model's emphasis on shadow areas. Such a device is one of the most influential trends in deep learning attempting to mimic human brain actions by selectively concentrating on relevant objects (Bahdanau et al., 2014; Sutskever et al., 2014), which are shadows in our study. As shown in Fig. 5, the red areas with higher attention values (i.e., closer to 1) are more likely to be shadow, whereas the blue areas with lower attention values are more likely to be non-shadow. The shadow attention maps gradually improved through recurrent iterations.

4.1.4. Implementation of RSAM

At the training stage, we randomly selected 73 image patch samples out of 103 (approximately 70%) from the SSAD and their corresponding NAIP images (Infrared Red Green, IRRG). We used a sliding window strategy (Lampert et al., 2008) to extract the patches of 250 × 250 pixels and a 32-pixel stride to recognize shadow at varying scales and locations in each patch. We conducted 50 epochs with the bath size 10 (i.e., 500 iterations). In our evaluation, the optimal number of epochs was set to 50, as this number allowed all the SSAD sample patches to be used for training and the model performance achieved a stable level (Fig. 6). At the testing stage, we used a 32-pixel stride window for the selected NAIP image. The value of 32 was chosen to reduce border effects and predict results efficiently while being permitted by the GPU cache. RSAM was implemented in the Caffe framework (Jia et al., 2014), using the eminent Stochastic Gradient Descent (Sra et al., 2012) to optimize parameters with a base learning rate of 0.01 adaptively, a momentum of 0.9, a weight decay of 0.0005 and a batch size of 10 for high-resolution mapping in urban environments (Audebert et al., 2018). For SegNet-based architectures, the weights of the encoder were initialized with those of inception-v4 trained on ImageNet, while the decoder weights were randomly initialized (He et al., 2015). The main computation in both the convolution and deconvolution stages was filtering, which was implemented as the standard dot product of two vectors. Our networks used a SoftMax layer to compute the multinomial logistic losses (Audebert et al., 2018), which were averaged over all the patches. In this study, a total of six shaded land-cover classes across the 30 tested urban neighborhoods were generated, i.e., building, tree, grass/shrub, road, water, and farmland.

4.1.5. Accuracy assessment

We used the remaining 30 urban neighborhood patches (set aside from the training data) from the SSAD for model validation. Those patches (a total area of 7.50 km^2) cover diverse shadow types along the urban-rural gradient. Overall accuracy (OA) and Kappa statistic (Kappa) were calculated and reported along with F1-score as an additional accuracy metric, as this is suitable for comparing different methods running on different portions of the dataset (Goutte and Gaussier, 2005).

4.2. Effects of urban development patterns on model performance

In the study, we selected widely used landscape metrics (McGarigal, 2014) to evaluate RSAM's performance for urban neighborhoods of various development patterns, including landscape-level metrics (e.g., CONTANG: Contagion Index; SHDI: Shannon's Diversity Index) and class-level metrics (e.g., ED: edge density; PLAND: percentage of land; COHESION: patch cohesion index) (Lechner et al., 2009; Smith et al., 2003). The 8-neighbor rule was chosen for patch delineation, treating both cardinal and diagonal pixels/cells as adjacent neighbors. This rule has been found to generate appropriate patches in previous urban studies (Godwin et al., 2015). These metrics capture urban landscape patterns from various perspectives – geometry, dispersion/interspersion, diversity, and connectivity.

We summarized and compared all the selected landscape metrics for the 30 validation neighborhood patches (T1-T30, Fig. 2). Those patches are divided according to the percent built-up (PBU) of each neighborhood (Angel et al., 2012), representing three primary urban development intensities along the urban-rural gradient: urban center (hereafter

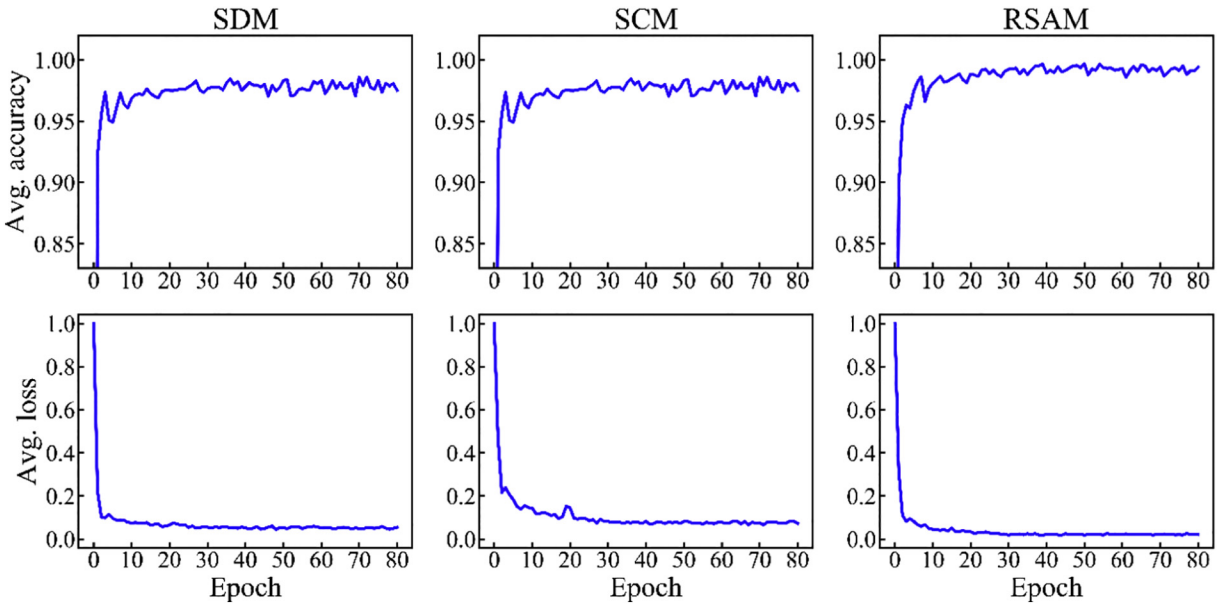

Fig. 6. The blue line corresponds to the average values of accuracy (top) or loss (bottom) with the change of epoch (0–80) for SDM, SCM, and RSAM. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

urban), suburban, and rural. Spearman's correlation coefficients were calculated to evaluate the relationship between model accuracy and the selected landscape metrics for the three types of neighborhoods, respectively. Spearman's correlation was chosen, instead of the commonly used Pearson's correlation, to describe the potentially nonlinear, monotonic relationship between variables (Hollander and Wolfe, 1973).

5. Results and discussion

5.1. Overall performance of RSAM

To facilitate the performance evaluation of the proposed model RSAM, we applied state-of-the-art deep learning architecture ResNet-50 (Targ et al., 2016) as a baseline (i.e., direct classification of land-cover classes within shadows) for result comparison. There are two reasons. First, recent studies (e.g., Luo et al., 2019; Mostafa and Abdelhafiz, 2017; Su et al., 2016) focused extensively on recovering the spectral reflectance of the ground objects to their non-shadow conditions. No other shadow algorithms were known to directly generate land cover maps, as we did with RSAM. Second, ResNet-50 has been successfully applied to urban semantic segmentation (Zhong et al., 2018; Zhu et al., 2017) and shadow processing of natural imagery (Zhu et al., 2018), demonstrating a good balance between model complexity and accuracy. We initialized RSAM and ResNet-50, respectively, using parameters from inception-v4 pre-trained with ImageNet. Compared with RSAM, ResNet-50's configuration remained the same to ensure a fair comparison (Section 4.1.4). Although the structure of RSAM was more complicated due to using the bidirectional attention mechanism, we relied on parallel processing to run the model. As a result, ResNet-50 and RSAM had similar runtime. Using the same training and validation samples for the two models, RSAM was found to outperform ResNet-50, with OA of 90.6% versus 76.2%, and Kappa of 0.82 versus 0.54. Similarly, RSAM revealed higher accuracies (F1-scores) than ResNet-50 in mapping each of the six land-cover classes within shadows – building: 72.8% versus 33.1%, tree: 90.2% versus 63.9%, grass/shrub: 79.7% versus 41.4%, road: 84.8% versus 54.3%, water: 83.9% versus 32.2%, farmland: 91.9% versus 35.8% (see Fig. 7). Our visual interpretation of the results indicates a relatively robust model performance of ResNet-50 across urban neighborhood types (Fig. 8); however, problems

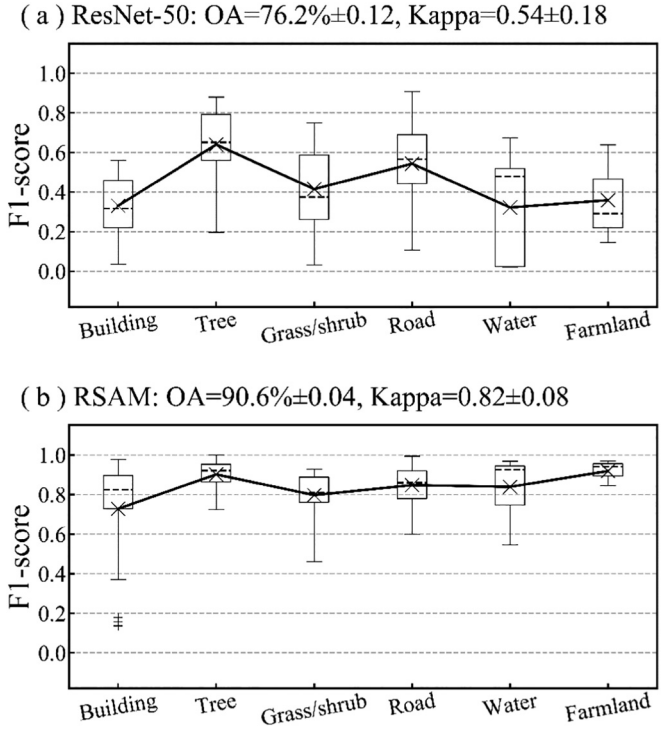

Fig. 7. Comparison of the model performance across 30 tested neighborhood patches using (a) ResNet-50 versus (b) RSAM. Boxplots showed the maximum, minimum, median, standard deviation, and average values of the F1-score.

remained primarily for extracting roads and water (Fig. 8). This may be due to the fact that high-resolution remote sensing imagery typically has low feature resolution and broad data range, causing the shadow areas to have highly fragmented and detailed land-cover features, i.e., uncertain semantic and fuzzy boundary (Wang et al., 2017). Compared with ResNet-50, RSAM improved the average performance for extracting shaded roads by 30.5% in F1-score, while the standard deviation decreased by 0.44 across the tested neighborhood patches. Similarly, for extracting shaded water, the model performance improved by 51.7% in F1-score, with a decrease of 0.17 in standard deviation

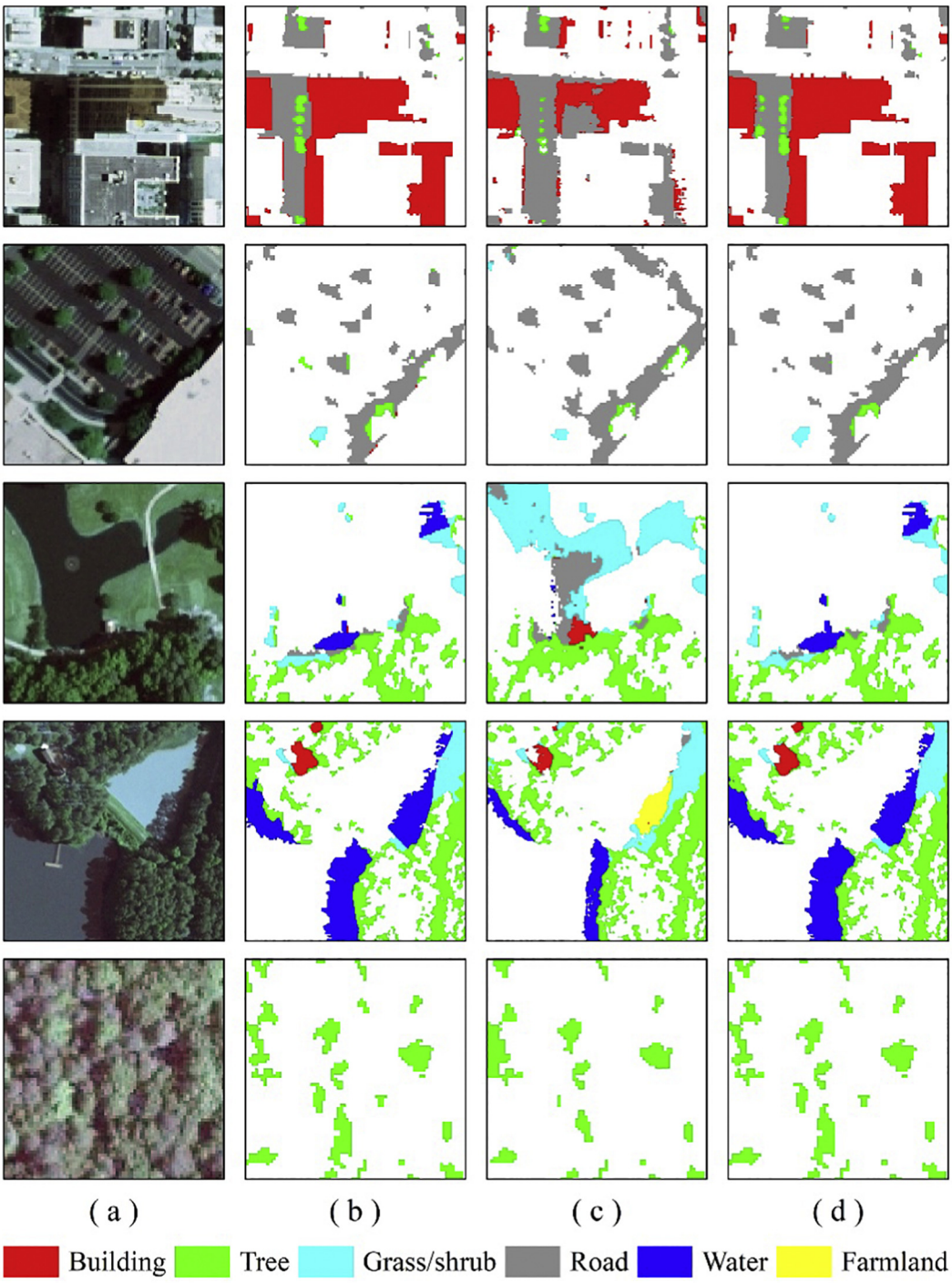

Fig. 8. Five typical examples for model performance in building, road, water, and tree: (a) NAIP, (b) Ground truth; (c) ResNet-50, and (d) RSAM, respectively.

(Fig. 7). While ResNet-50 (with the residual learning technique) has improved feature refinement by learning the residual of input features, it introduced non-shadow areas into the results because it lacks the bidirectional shadow attention mechanism developed for RSAM. Capitalizing on this mechanism, RSAM was able to suppress non-shaded land-cover context information (including those from low-albedo objects) to be introduced into the final result and exhibit smaller variation in the results along the urban-rural gradient, e.g., lower standard deviation values for OA, Kappa, and F1-score (RSAM: 0.04, 0.08, and 0.12; ResNet-50: 0.12, 0.18, and 0.19) across the tested neighborhood patches.

While recent studies focused extensively on using two sequential steps – shadow detection and shadow removal to address the shadow concern in high-resolution imagery (Luo et al., 2019; Mostafa and Abdelhafiz, 2017; Su et al., 2016), errors propagating through the steps raised new concerns about model robustness and its generalizability capacity across land-cover types (Mostafa, 2017). Our study, for the first time, capitalizes on the deep learning bidirectional attention mechanism to iteratively refine shadow detection for estimating land-cover classes within shadows. The strategy circumvents the dependence on two separate steps. We also note that most of the previous studies attempted to recover the spectral reflectance of the ground objects to their non-shadow conditions. Our model moved one step forward to extract land cover types within shadows. This product can be directly

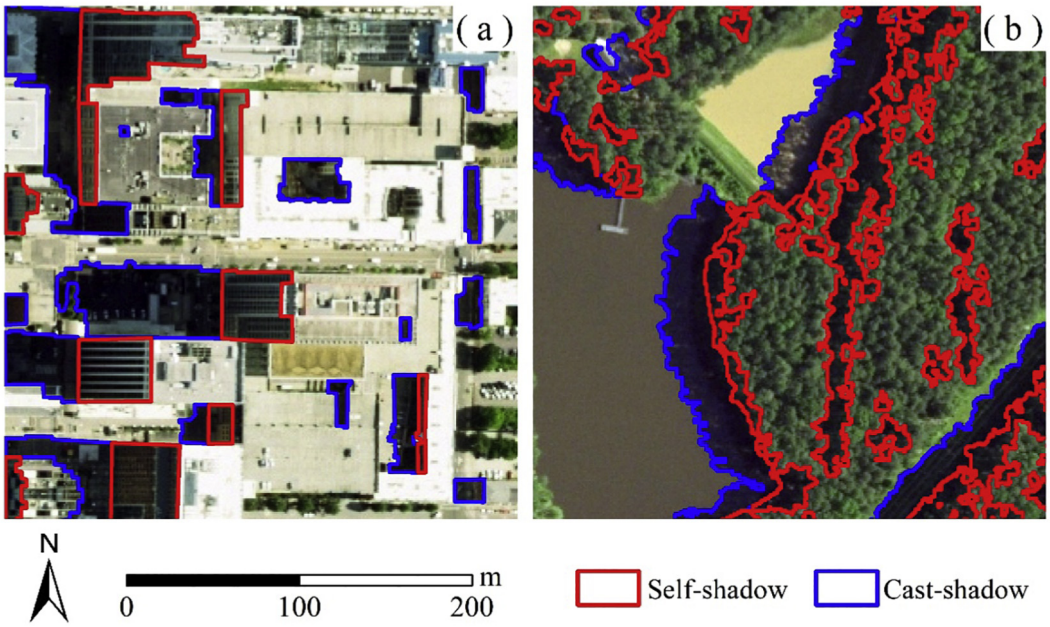

Fig. 9. Sample cast and self shadows for buildings (a) and trees (b). The red polygons include self shadows, while the blue polygons indicate cast shadows. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

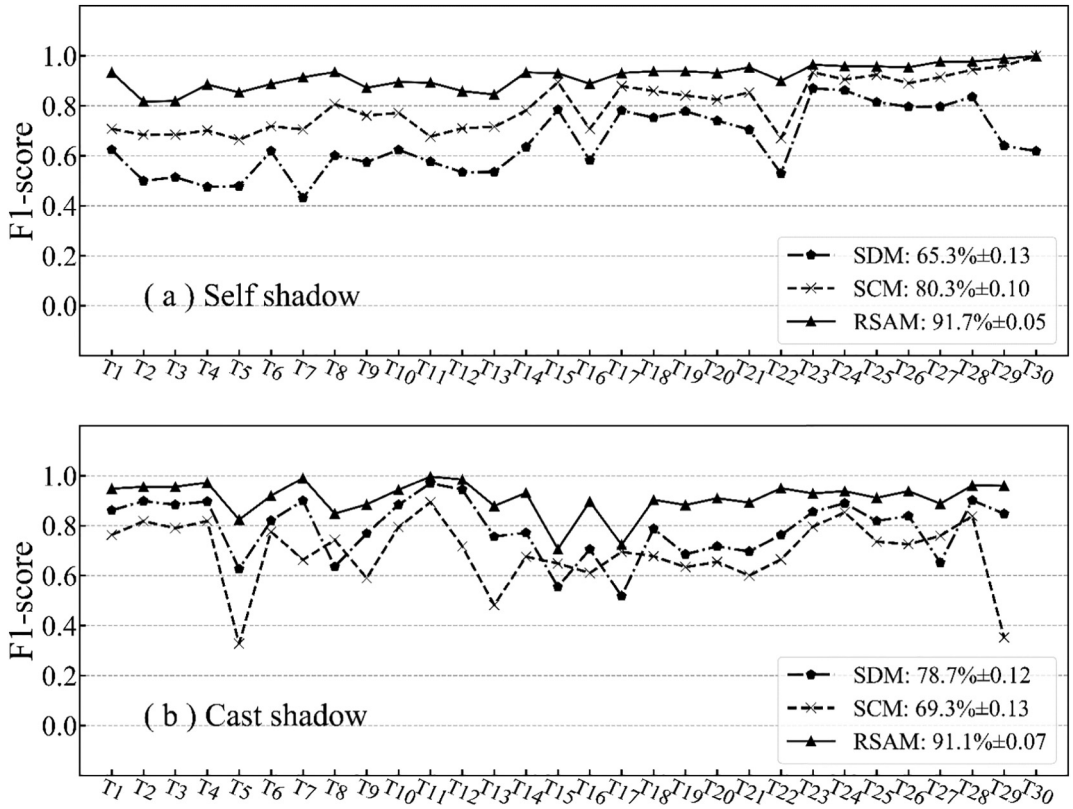

Fig. 10. Comparison of overall performance in F1-score for (a) self shadow versus (b) cast shadow detection across 30 tested urban neighborhood patches (T1-T30) using three scenarios: SDM, SCM, and RSAM.

used by researchers and practitioners in urban studies. In addition, incorporating ancillary data (e.g., LiDAR – light detection and ranging) has proven effective in shadow detection and classification (Milas et al., 2017; Sharma and Singhai, 2019); however, it demands computational resources and expertise, and sometimes impossible to obtain such auxiliary data for specific regions. Our model only requires high-resolution imagery as input, which is increasingly available in urbanized areas. Due to certain illumination conditions (e.g., a low solar elevation angle) or land cover types (e.g., the shadow cast onto a water surface), heavy shadows add additional challenges to ascertaining accurate land cover even with visual interpretation. While the spectral and spatial information of shadows is used in our modeling, RSAM

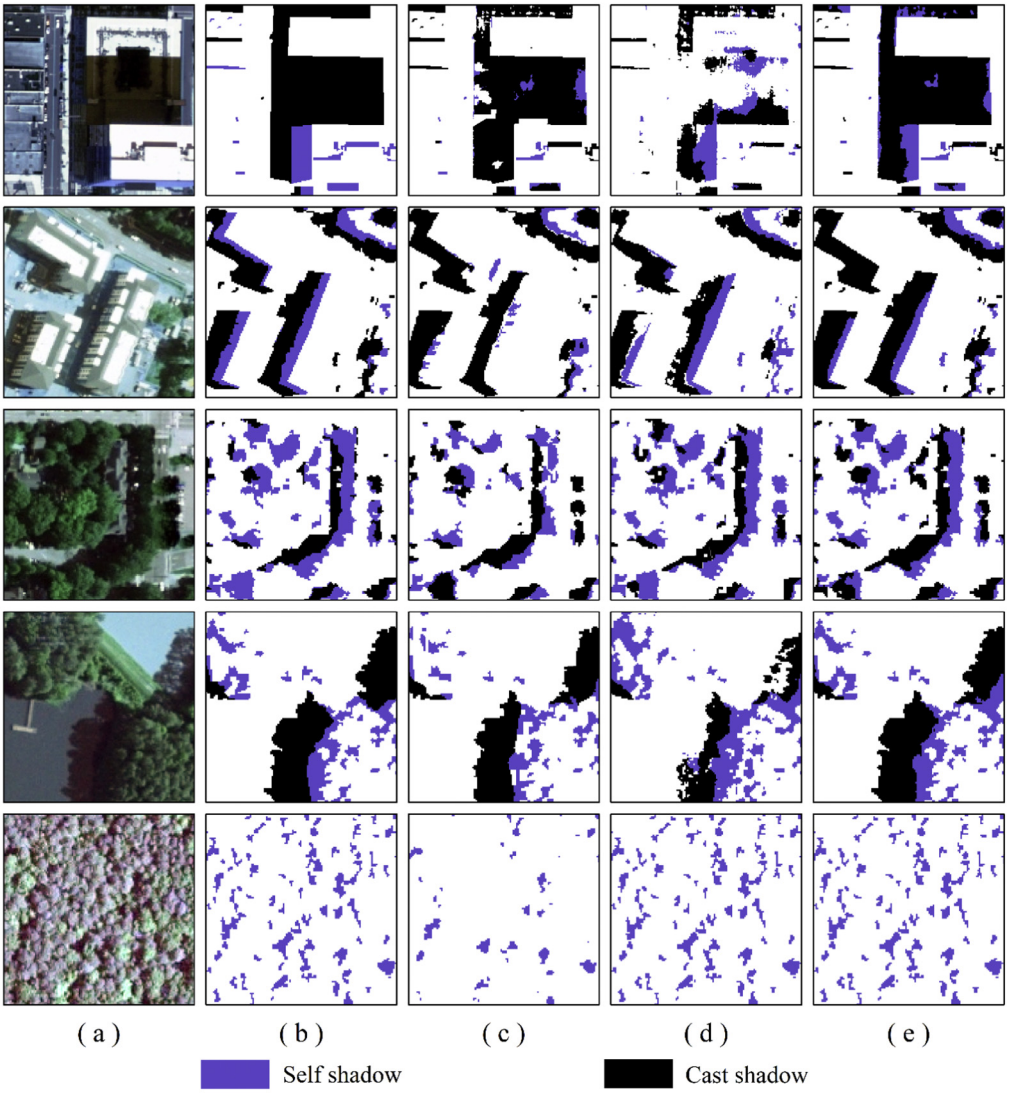

Fig. 11. Five sample patches, i.e., building, road, water, and tree in model performance for self shadow versus cast shadow detection between three scenarios: (a) NAIP, (b) Ground truth, (c) SDM, (d) SCM, and (e) RSAM, respectively.

further capitalizes on shadow's contextual variation (neighboring land covers) – one key feature of deep learning – to infer the landscape types underneath the shadow. Another unique contribution of this study is the development of a shadow semantic annotation database – SSAD, which is accurate, diverse, and extendable. Such knowledge can be easily used to help train a deep learning model for high-resolution shadow removal in other urban regions.

5.2. Comparison between cast and self shadows

Our study evaluated cast and self shadows over various types of urban neighborhoods. Compared to cast shadows, we found that self shadows typically demonstrated high spectral variation and high fragmentation. For instance, the self shadow of a building or a tree (or a tree cluster) is often a mixture of light and dark shadow patches that are smaller than the corresponding cast shadow (Fig. 9). Aside from the causes of shadow, secondary lighting from the surrounding illuminated objects may also be contributed to the difference (Dare, 2005).

We compared the performance of RSAM in detecting cast versus self shadows. Since the SDM and SCM modules of RSAM were designed to identify the two shadow types, respectively, they were also used as standalone models for the purposes of comparison. Here, we reported our findings from three shadow detection scenarios: SDM, SCM, and RSAM. For all the tested neighborhoods (Fig. 10), on average, RSAM gained superior performance for detecting self shadows with higher F1-scores (91.7%) than using the SDM (65.3%) or the SCM model (80.3%). Similarly, RSAM achieved higher performance (F1-score) for cast shadows in (91.1%) than SDM (78.8%) or SCM (69.3%). RSAM extracted and integrated the semantic of both shadows and land-cover classes into a position attention map through SDM and SCM, achieving higher sensitivity and more accurate detection performance for self and cast shadows. When comparing SDM and SCM, we found that SDM performed better than SCM, with more accurate results in cast shadows, especially those from buildings with accurate boundary detection (Fig. 11). However, SCM was more suitable for detecting self shadows with higher sensitivity to shaded trees (Fig. 11). While previous studies mainly investigated cast shadows (Liu et al., 2017; Su et al., 2016; Zhang et al., 2014) or treated cast and self shadows as single objects (Zhou et al., 2009), our study is one of the first to focus on cast and self shadows separately and report on model performance for both shadow types.

5.3. Effects of urban development patterns

We observed that shadow occupied 14.8% urban, 13.4% suburban, and 10.3% rural in the test database. The proposed RSAM demonstrated

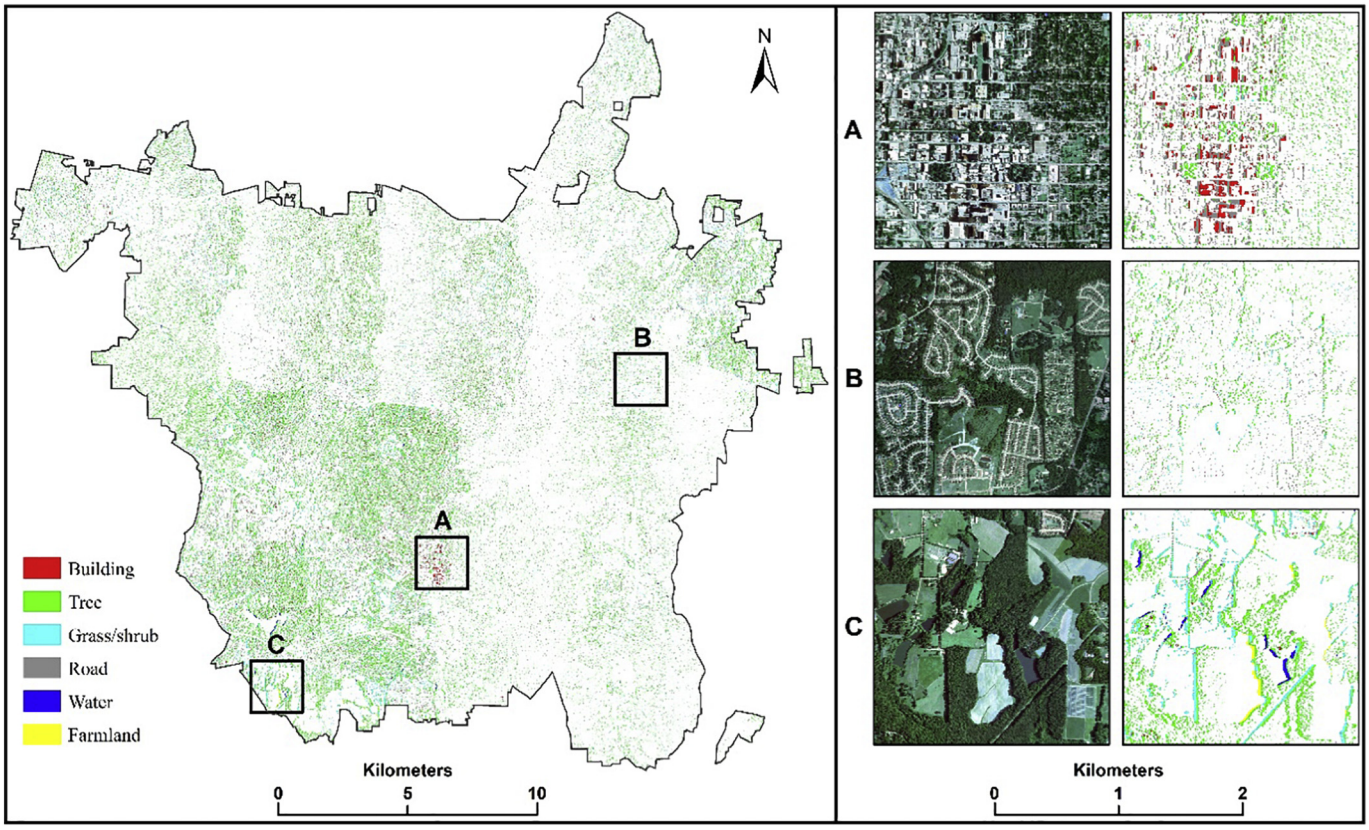

Fig. 12. Shaded land cover estimating result for the city of Raleigh using RSAM. The insets correspond to the neighborhood-level results for sample urban center, suburban, and rural regions.

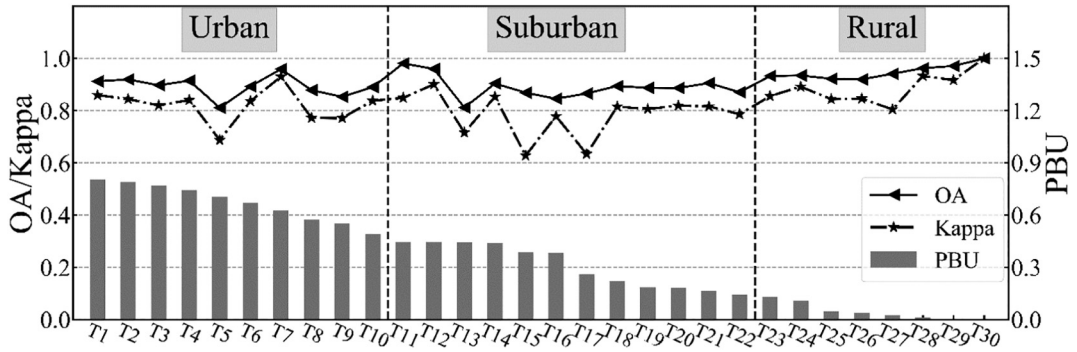

Fig. 13. The overall performance of RASM in OA (Overall Accuracy) and Kappa across various types of urban neighborhoods. PBU is a percent build-up representing various development intensities in the tested neighborhoods T1-T30.

relatively stable performance across the three types of urban development patterns (Fig. 13), with average OA of 89.3%, 89.0%, and 94.8%, and Kappa of 0.82, 0.78, and 0.89 for urban, suburban and rural neighborhoods, respectively. We also estimated the city of Raleigh and the model performed well (Fig. 12), especially in the rural regions due to simple shadow conditions (e.g., less tall buildings and larger forest patches) than the highly urbanized areas. The six land-cover types were detected within shadows with various levels of success along the urban-rural gradient. For instance, RSAM achieved stable performance for all the land-cover types over the urban core neighborhoods except farmlands that did not exist in the region (Fig. 14). Specifically, the model performance (average F1-scores) across urban, suburban and rural neighborhoods – building: 80.3%, 68.9%, and 67.9%; Tree: 84.7%, 90.2%, and 97.1%; Grass/shrub: 76.7%, 77.8%, and 88.6%; Road: 89.8%, 84.6%, and 74.8%, Water: 78.4%, 84.4%, and 85.7%. However, when moving to the neighborhoods with lower urban development intensities (e.g., suburban and rural), buildings within shadows were harder to detect than trees. It was possibly due to the discontinuous urban fabric (i.e., decentralized building distribution) in the suburban, causing the shaded buildings to be fragmented at a higher level than the shaded trees (Awuah et al., 2018).

To date, studies of shadow detection and/or removal have mainly focused on urban (Ma et al., 2015; Qiao et al., 2017; Su et al., 2016; Zhang et al., 2014) and suburban residential regions (Zhou et al., 2009). However, none of them have quantified the relationship between the spatial patterns of urban development and model performance. In this study, our statistical analysis discovered consistent, significant effects of urban forests (i.e., PLAND_C2 and COHESION_C2) and edge density (ED) of ground patches on RSAM's performance along the entire urban-rural gradient (Table 1). This was probably due to a large number of trees in all the neighborhoods. The intricate edges of tree patches and high variation in forest structure (caused by diverse

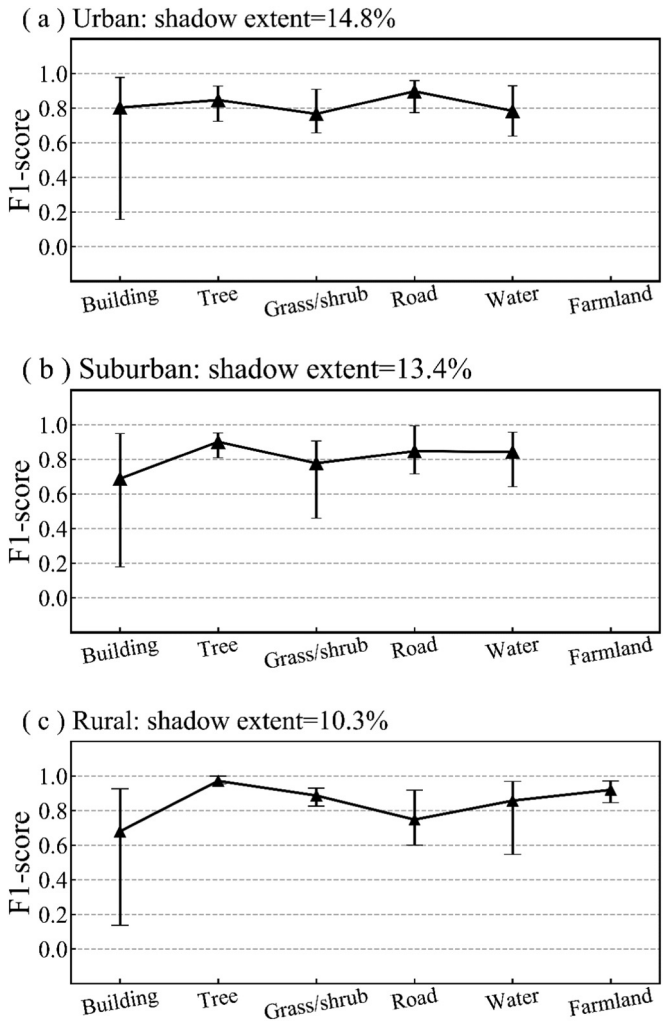

Fig. 14. Comparison of the RSAM's model performance in F1-score for all the land-cover types within shadows (i.e., building, tree, grass/shrub, water, and farmland) over the urban core neighborhoods such as (a) urban, (b) suburban, and (c) rural. Stock charts presented maximum, minimum, and average values.

Table 1
Spearman's correlation coefficients between RSAM's performance and the selected landscape metrics along the urban-rural gradient in urban, suburban, and rural neighborhoods, respectively.

Landscape metric	Rural	Suburban	Urban
PLAND_C1	−0.15 ns	−0.03 ns	0.61**
PLAND_C2	0.79**	0.57**	0.68**
PLAND_C3	0.07 ns	0.03 ns	0.47**
PLAND_C4	0.16 ns	0.60**	0.21 ns
PLAND_C5	0.38 ns	0.74**	−0.77 ns
PLAND_C6	−0.09 ns	n/a	n/a
COHESION_C1	−0.30 ns	0.04 ns	0.69**
COHESION_C2	0.84**	0.63**	0.80**
COHESION_C3	−0.07 ns	0.20 ns	0.52**
COHESION_C4	0.17 ns	0.55**	0.10 ns
COHESION_C5	0.31 ns	0.64**	−0.26 ns
COHESION_C6	−0.40 ns	n/a	n/a
ED	−0.72**	−0.43**	−0.36*
CONTAG	0.03 ns	0.02 ns	0.29 ns
SHDI	−0.60**	−0.12 ns	−0.32 ns

Level of significance: *$p < .05$; **$p < .01$; ns: $p > .05$.
Ci: shaded land-cover class i (1: building, 2: tree, 3: grass/shrub, 4: road, 5: water, 6: farmland); PLAND_Ci: percentage of landscape area for class i; COHESION_Ci: patch cohesion index for class i; edge density(ED); contagion index (CONTAG); and Shannon's diversity index (SHDI).

species types, growing stages, and forest management practices) may have introduced high uncertainties to retrieving land cover in shadows. When comparing neighborhoods of increasing urban development intensity from rural to urban centers, we found increasing effects of non-forest land-cover classes on model performance. For example, roads and water bodies revealed significant correlations with model accuracy in the suburban areas, while buildings and grass/shrub were more influential to the results in areas close to urban centers. We note that the specific relationship between landscape patterns and model performance may vary from one city to another. However, such quantitative analysis informs valid deep-learning-based model calibration by collecting training data for the land cover that is highly relevant to model performance.

6. Conclusion

In this study, we developed a novel deep-learning-based shadow removal model RSAM to eliminate shadow effects in high-resolution urban land-cover mapping. We also quantified the relationship between the spatial patterns of urban development and model performance. Based on the results of this study, the following conclusions can be drawn: (i) RSAM capitalizes on the deep learning attention mechanism to progressively refine the shadow detection results, which provides a viable and accurate solution to retrieve land-cover classes within shadows in an urban context (overall accuracy of 90.6% and Kappa of 0.82). (ii) Our model has the capacity to detect both self and cast shadows, which is significant because self shadows were often neglected in previous studies due to their complex and distinct contextual features in the urban environment. (iii) RSAM shows relatively stable performance along the evaluated urban-rural gradient with diverse spatial development patterns, demonstrating its generalizability potential to other urban regions. (iv) This study produced a database of shadow semantic annotation SSAD, the first of its kind for high-resolution shadow detection. It is accurate, diverse, and extendable, and hence can directly benefit future shadow removal studies as effective training and validation data. The SSAD currently has 103 patches, and we are keen to enhance the database for removing more complex shadows by sharing it with the broader community: https://pages.uncc.edu/gang-chen/research-products/.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was jointly supported by the National Science Foundation of China (Grant NO. 41271360), the Fundamental Research Funds for the Central Universities (LZUJBKY-2016-248), and the China Scholarship Council (CSC#201806180076). The authors also appreciate the generous support from the University of North Carolina at Charlotte and North Carolina State University.

References

Adeline, K., Briottet, X., Ceamanos, X., Dartigalongue, T., Gastellu-Etchegorry, J.-P., 2018. ICARE-VEG: a 3D physics-based atmospheric correction method for tree shadows in urban areas. ISPRS J. Photogramm. Remote Sens. 142, 311–327.
Angel, S., Parent, J., Civco, D.L., 2012. The fragmentation of urban landscapes: global evidence of a key attribute of the spatial structure of cities, 1990–2000. Environ. Urban. 24, 249–283.
Arévalo, V., González, J., Ambrosio, G., 2008. Shadow detection in colour high-resolution satellite images. Int. J. Remote Sens. 29, 1945–1963.
Audebert, N., Boulch, A., Randrianarivo, H., Le Saux, B., Ferecatu, M., Lefèvre, S., Marlet, R., 2017. Deep learning for urban remote sensing. In, 2017 IEEE Joint Urban Remote Sensing Event (JURSE), 1-4.

Audebert, N., Le Saux, B., Lefèvre, S., 2018. Beyond RGB: very high resolution urban remote sensing with multimodal deep networks. ISPRS J. Photogramm. Remote Sens. 140, 20–32.
Awuah, K.T., Nölke, N., Freudenberg, M., Diwakara, B.N., Tewari, V.P., Kleinn, C., 2018. Spatial resolution and landscape structure along an urban-rural gradient: do they relate to remote sensing classification accuracy? – a case study in the megacity of Bengaluru, India. Remote Sens. Appl. Soc. Environ. 12, 89–98.
Azevedo, S., Silva, E., Colnago, M., Negri, R., Casaca, W., 2019. Shadow detection using object area-based and morphological filtering for very high-resolution satellite imagery of urban areas. J. Appl. Remote. Sens. 13, 036506.
Badrinarayanan, V., Handa, A., Cipolla, R., 2015. Segnet: a deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling. *arXiv preprint. arXiv:1505.07293*.
Bahdanau, D., Cho, K., Bengio, Y., 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint. arXiv:1409.0473*.
Buyssens, P., Daisy, M., Tschumperlé, D., Lézoray, O., 2015. Exemplar-based inpainting: technical review and new heuristics for better geometric reconstructions. IEEE Trans. Image Process. 24, 1809–1824.
Census, 2020. Raleigh-Durham-Chapel Hill, NC CSA - Profile data. https://censusreporter.org/profiles/33000US33450-raleigh-durham-chapel-hill-nc-csa Last accessed on May 13, 2020.
Chang, C.W., Tsay, J.-R., 2010. Shadow detection and information recovery in aerial images. In, 31st Asian Conference on Remote Sensing 2010, 392-397.
Chen, Y., Wen, D., Jing, L., Shi, P., 2007. Shadow information recovery in urban areas from very high resolution satellite imagery. Int. J. Remote Sens. 28, 3249–3254.
Chen, G., Hay, G.J., Castilla, G., St-Onge, B., Powers, R., 2011. A multiscale geographic object-based image analysis to estimate lidar-measured forest canopy height using Quickbird imagery. Int. J. Geogr. Inf. Sci. 25, 877–893.
Dare, P.M., 2005. Shadow analysis in high-resolution satellite imagery of urban areas. Photogramm. Eng. Remote. Sens. 71, 169–177.
Ding, B., Long, C., Zhang, L., Xiao, C., 2019. ARGAN: attentive recurrent generative adversarial network for shadow detection and removal. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 10213–10222.
Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z., Lu, H., 2019. Dual attention network for scene segmentation. In, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 3146-3154.
Godwin, C., Chen, G., Singh, K.K., 2015. The impact of urban residential development patterns on forest carbon density: an integration of LiDAR, aerial photography and field mensuration. Landsc. Urban Plan. 136, 97–109.
Goutte, C., Gaussier, E., 2005. A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In, 27th European Conference on Information Retrieval, 345–359.
Graves, A., Schmidhuber, J., 2005. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural Netw. 18, 602–610.
Hamida, A.B., Benoit, A., Lambert, P., Klein, L., Amar, C.B., Audebert, N., Lefèvre, S., 2017. Deep learning for semantic segmentation of remote sensing images with rich spectral content. In, 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2569–2572.
He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1026–1034.
Hollander, M., Wolfe, D.A., 1973. Nonparametric Statistical Methods, John Wiley & Sons. Inc. New York.
Hu, X., Zhu, L., Fu, C.-W., Qin, J., Heng, P.-A., 2018. Direction-aware spatial context features for shadow detection. In, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 7454-7462.
ISPRS, 2019a. 2D Semantic Labeling Contest. http://www2.isprs.org/commissions/comm3/wg4/2d-sem-label-potsdam.html Last accessed on December 9, 2019.
ISPRS, 2019b. 2D Semantic Labeling Contest. http://www2.isprs.org/commissions/comm3/wg4/2d-sem-label-vaihingen.html Last accessed on December 9, 2019.
Jain, V., Khunteta, A., 2017. Shadow removal for umbrageous information recovery in aerial images. In: In, 2017 IEEE International Conference on Computer, Communications and Electronics (Comptelix), pp. 536–540.
Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T., 2014. Caffe: convolutional architecture for fast feature embedding. In, Proceedings of the 22nd ACM international conference on Multimedia, 675-678.
Jiang, J., Lyu, C., Liu, S., He, Y., Hao, X., 2020. RWSNet: a semantic segmentation network based on SegNet combined with random walk for remote sensing. Int. J. Remote Sens. 41, 487–505.
Kang, X., Huang, Y., Li, S., Lin, H., Benediktsson, J.A., 2017. Extended random walker for shadow detection in very high resolution remote sensing images. IEEE Trans. Geosci. Remote Sens. 56, 867–876.
Kang, J., Ci, T., Dang, A., Wang, Y., 2019. An automatic method for water extraction from high spatial resolution GF-1 imagery based on a deep learning algorithm. In: In, 2019 International Conference on Computer Intelligent Systems and Network Remote Control (CISNRC 2019), pp. 555–562.
Lampert, C.H., Blaschko, M.B., Hofmann, T., 2008. Beyond sliding windows: object localization by efficient subwindow search. In, 2008 IEEE conference on computer vision and pattern recognition, 1-8.
Lechner, A.M., Stein, A., Jones, S.D., Ferwerda, J.G., 2009. Remote sensing of small and linear features: quantifying the effects of patch size and length, grid position and detectability on land cover mapping. Remote Sens. Environ. 113, 2194–2204.
LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.
Li, Y., Sasagawa, T., Gong, P., 2004. A system of the shadow detection and shadow removal for high resolution city aerial photo. In: In, XXth ISPRS Congress, pp. 12–23.
Lin, T.-P., Matzarakis, A., Hwang, R.-L., 2010. Shading effect on long-term outdoor thermal comfort. Build. Environ. 45, 213–221.
Liu, S., Deng, W., 2015. Very deep convolutional neural network based image classification using small training sample size. In, 2015 3rd IAPR Asian conference on pattern recognition (ACPR), 730-734.
Liu, X., Hou, Z., Shi, Z., Bo, Y., Cheng, J., 2017. A shadow identification method using vegetation indices derived from hyperspectral data. Int. J. Remote Sens. 38, 5357–5373.
Luo, S., Shen, H., Li, H., Chen, Y., 2019. Shadow removal based on separated illumination correction for urban aerial remote sensing images. Signal Process. 165, 197–208.
Ma, L., Jiang, B., Jiang, X., Tian, Y., 2015. Shadow removal in remote sensing images using features sample matting. In, 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 4412–4415.
Massalabi, A., He, D.-C., Benie, G.B., Beaudry, E., 2004. Detecting information under and from shadow in panchromatic Ikonos images of the city of Sherbrooke. In, 2004 IEEE International Geoscience and Remote Sensing Symposium, 2000–2003.
McGarigal, K., 2014. Landscape Pattern Metrics. Statistics Reference Online, Wiley StatsRef.
Milas, A.S., Arend, K., Mayer, C., Simonson, M.A., Mackey, S., 2017. Different colours of shadows: classification of UAV images. Int. J. Remote Sens. 38, 3084–3100.
Mo, N., Zhu, R., Yan, L., Zhao, Z., 2018. Deshadowing of urban airborne imagery based on object-oriented automatic shadow detection and regional matching compensation. IEEE J. Select. Topics Appl. Earth Observ. Remote Sens. 11, 585–605.
Mostafa, Y., 2017. A review on various shadow detection and compensation techniques in remote sensing images. Can. J. Remote. Sens. 43, 545–562.
Mostafa, Y., Abdelhafiz, A., 2017. Shadow identification in high resolution satellite images in the presence of water regions. Photogramm. Eng. Remote. Sens. 83, 87–94.
Panboonyuen, T., Jitkajornwanich, K., Lawawirojwong, S., Srestasathiern, P., Vateekul, P., 2017. Road segmentation of remotely-sensed images using deep convolutional neural networks with landscape metrics and conditional random fields. Remote Sens. 9, 680.
Qiao, X., Yuan, D., Li, H., 2017. Urban shadow detection and classification using hyperspectral image. J. Indian Soc. Remote Sens. 45, 945–952.
Sarabandi, P., Yamazaki, F., Matsuoka, M., Kiremidjian, A., 2004. Shadow detection and radiometric restoration in satellite high resolution images. In, 2004 IEEE International Geoscience and Remote Sensing Symposium, 3744–3747.
Schuster, M., Paliwal, K.K., 1997. Bidirectional recurrent neural networks. IEEE Trans. Signal Process. 45, 2673–2681.
Shackelford, A.K., Davis, C.H., 2003. A combined fuzzy pixel-based and object-based approach for classification of high-resolution multispectral data over urban areas. IEEE Trans. Geosci. Remote Sens. 41, 2354–2363.
Sharma, D., Singhai, J., 2019. An object-based shadow detection method for building delineation in high-resolution satellite images. PFG – J. Photogrammetry, Remote Sens. Geoinform. Sci. 87, 103–118.
Shedlovska, Y., Hnatushenko, V., 2019. Shadow removal algorithm for remote sensing imagery. In: In, 2019 IEEE 39th International Conference on Electronics and Nanotechnology (ELNANO), pp. 818–822.
Silva, G.F., Carneiro, G.B., Doth, R., Amaral, L.A., de Azevedo, D.F., 2018. Near real-time shadow detection and removal in aerial motion imagery application. ISPRS J. Photogramm. Remote Sens. 140, 104–121.
Smith, J.H., Stehman, S.V., Wickham, J.D., Yang, L., 2003. Effects of landscape characteristics on land-cover class accuracy. Remote Sens. Environ. 84, 342–349.
Sra, S., Nowozin, S., Wright, S.J., 2012. Optimization for Machine Learning. Mit Press.
Su, N., Zhang, Y., Tian, S., Yan, Y., Miao, X., 2016. Shadow detection and removal for occluded object information recovery in urban high-resolution panchromatic satellite images. IEEE J. Select. Topics Appl. Earth Observ. Remote Sens. 9, 2568–2582.
Sutskever, I., Vinyals, O., Le, Q., 2014. Sequence to sequence learning with neural networks. In: In, Advances in Neural Information Processing Systems, pp. 3104–3112.
Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A., 2017. Inception-v4, inception-resnet and the impact of residual connections on learning. In, Proceedings of the AAAI Conference on Artificial Intelligence, 4278-4284.
Targ, S., Almeida, D., Lyman, K., 2016. Resnet in resnet: generalizing residual architectures. arXiv preprint. *arXiv:1603.08029*.
Tomas, F.Y.V., Hou, Le, Chenping, Y., Minh, H., Dimitris, S., 2016. SBU shadow dataset. https://www3.cs.stonybrook.edu/~cvl/projects/shadow_noisy_label/index.html.
Torrey, L., Shavlik, J., 2009. Transfer learning. In: Soria, E., Martin, J., Magdalena, R., Martinez, M., Serrano, A. (Eds.), Handbook of Research on Machine Learning Applications. IGI Global, pp. 242–264.
USGS, 2019. Earth Explorer. https://earthexplorer.usgs.gov Last accessed on December 9, 2019.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. In, Advances in Neural Information Processing Systems. pp. 5998–6008.
Wang, L., Guo, S., Huang, W., Qiao, Y., 2015. Places205-vggnet models for scene recognition. *arXiv preprint. arXiv:1508.01667*.
Wang, Q., Yan, L., Yuan, Q., Ma, Z., 2017. An automatic shadow detection method for VHR remote sensing orthoimagery. Remote Sens. 9, 469.
Woo, S., Park, J., Lee, J.-Y., So Kweon, I., 2018. Cbam: convolutional block attention module. In, Proceedings of the European Conference on Computer Vision (ECCV), 3-19.
Zhang, H., Sun, K., Li, W., 2014. Object-oriented shadow detection and removal from urban high-resolution remote sensing images. IEEE Trans. Geosci. Remote Sens. 52, 6972–6982.
Zhang, L., Zhang, Q., Xiao, C., 2015a. Shadow remover: image shadow removal based on illumination recovering optimization. IEEE Trans. Image Process. 24, 4623–4636.
Zhang, X., Zou, J., He, K., Sun, J., 2015b. Accelerating very deep convolutional networks for classification and detection. IEEE Trans. Pattern Anal. Mach. Intell. 38,

1943–1955.
Zhang, C., Wei, S., Ji, S., Lu, M., 2019. Detecting large-scale urban land cover changes from very high resolution remote sensing images using CNN-based classification. ISPRS Int. J. Geo Inf. 8, 189.
Zhong, Y., Han, X., Zhang, L., 2018. Multi-class geospatial object detection based on a position-sensitive balancing framework for high spatial resolution remote sensing imagery. ISPRS J. Photogramm. Remote Sens. 138, 281–294.
Zhou, W., Huang, G., Troy, A., Cadenasso, M.L., 2009. Object-based land cover classification of shaded areas in high spatial resolution imagery of urban areas: a comparison study. Remote Sens. Environ. 113, 1769–1777.
Zhu, T., Li, Y., Ye, Q., Huo, H., Fang, T., 2017. Integrating saliency and ResNet for airport detection in large-size remote sensing images. In: In, 2017 2nd IEEE International Conference on Image, Vision and Computing (ICIVC), pp. 20–25.
Zhu, L., Deng, Z., Hu, X., Fu, C.-W., Xu, X., Qin, J., Heng, P.-A., 2018. Bidirectional feature pyramid network with recurrent attention residual modules for shadow detection. In, Proceedings of the European Conference on Computer Vision (ECCV), 121-136.