Differential Morphological Profile Neural Network for Object Detection in Overhead Imagery

Grant J. Scott, J. Alex Hurt, Alex Yang, Muhammad Aminul Islam, Derek T. Anderson and Curt H. Davis
Department of Electrical Engineering & Computer Science
University of Missouri, Columbia, MO, USA

***Abstract*—Deep convolutional neural networks (DCNN) have been the dominant methodology in the field of computer vision over the last decade, using various architectural organizations of successive convolutional layers to extract and assemble low level image features into visual component detectors. One of the trade-offs that have been made as the community has migrated to deep neural models is the loss of explainability and understanding of which salient visual components are being recognized by a model for a particular task. However, there exists a significant heritage in the remote sensing community that has developed advanced algorithms to analyze the signal and structural characteristics of anthropogenic features. One such approach is the use of morphological image processing techniques to extract objects from imagery and aid in the structural analysis of shapes. In particular, the differential morphological profile (DMP) has had great success extracting object shapes, while naturally grouping the extracted shapes into scale ranges. In this research, we present a novel architecture that integrates an explicit (definable and explainable) scaled object extraction into the network architecture, allowing shallower convolutional layers and lower complexity neural models. The architecture is evaluated on a challenging remote sensing dataset of object classes, providing insights to this approach and illuminating future directions of integrating morphology into neural architectures for enhanced explainability.**

***Index Terms*—Convolutional neural network, differential morphological profile, object detection, overhead imagery**

I. Introduction

Image processing for high-resolution remote sensing imagery (HR-RSI) has a long history that includes a wide variety of application domains, such as environmental and urban monitoring, land cover classification, and object detection just to name a few. Object detection is particularly challenging in remote sensing due to the variability of the numerous sensor platforms, environmental effects, naturally occurring object class variances, as well as diverse collection geometries (elevation angle, capture altitude, etc.). The sheer scale, variety, and complexity of HR-RSI presents challenges for even the most basic image processing, such as edge detection and object segmentation; which compounds further the difficulty of computer vision tasks such as land cover classification and object detection. Typical HR-RSI scenes may cover hundreds of square kilometers, with billions of pixels and a significant variety of collection characteristics; such as *ground sample distance* (GSD) (i.e., ground resolution) based on sensor geometry respective to the imaged Earth region (e.g., elevation angle from ground to sensor, azimuth orientation, and the compounding effects of Earth topology). This is further compounded by the variety of platforms (i.e., airborne or spaceborne) and sensors (e.g., different organizations and platforms, such as WorldView and Planet satellites).

Deep convolutional neural networks (DCNN) have been the dominant architecture in recent literature for land cover classification and object detection applications in remote sensing imagery [1]. There exists a variety of DCNN architectures that are constructed with organizations of successive convolutional layers to extract and assemble low level image features into visual component detectors, such as found in the ubiquitous VGG networks [2]. Furthermore, the research and commercial community has provided us with a variety of DCNN design enhancements.

Building upon the basic convolutional network layers, architectures such as GoogLeNet [3] and InceptionNetv3 [4] leverage *inception* modules that are composed of increasingly larger convolutions and max-pooling operations that are concatenated within a layer module. Alternative techniques include *residual* connections, such as found in the ResNet architectures [5], which also have been combined with inception modules for architectures such as the *InceptionResNet* varieties [6]. Architectures such as DenseNet [7] and Xception [8] build upon these concepts even further. DenseNet with complete residual connectivity, and Xception using depth separable convolutions to maintain independence of color channels.

Interesting architectural modifications to phases of DCNN beyond the convolutional feature extraction have been explored in the development of capsule networks as dynamic routing [9]. In [10], a Superpixel Capsule network is developed to map convolutional feature extraction into superpixel segmentation and applied to HR-RSI imagery data. Many of these architectures have been evaluated on a variety of benchmark overhead imagery datasets in research such as [11]–[14].

Object detection and localization in satellite imagery has been explored with a variety of neural architectures and techniques. The *single shot detectors* (SSD) [15] rely on a multi-box detectors, which test each location for each class to perform detection and localization. The most commonly used object detection and localization algorithm in recent years, however, is the *you only look once* (YOLO) [16] algorithm. YOLO is optimized for real time performance and needs only a single pass on an image to predict both the detection score and bounding boxes. Recently, YOLOv3 [17] introduces multi-class detection as well as object detection at three scales: small, medium, and large. Additional variants of the

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

YOLO architecture that are designed specifically for overhead imagery include YOLT [18] and SIMRDWN [19]. Similar to object detection, several techniques have been developed to perform semantic segmentation in overhead imagery. One leading technique is the U-Net algorithm [20], a convolutional-deconvolutional network that was originally developed for biomedical imaging, but has seen success in several other fields. Other techniques for semantic segmentation include the *mask regional convolutional neural network* (Mask R-CNN) [21]. This algorithm relies on finding *regions of interest* (ROI) and performing object localization before finally computing a semantic segmentation.

With the rare exceptions, such as [9] and [10], one of the trade-offs that have been made as the community has migrated to deep neural models is the loss of explainability and understanding of which salient visual components are being recognized by the models for particular object classes. For instance, in research such as [13] *state-of-the-art* performance on benchmark datasets is achieved using DCNN architectures such as ResNet50 [5], DenseNet [7], and Xception [8], just to list a few. In contrast to opaque techniques, such as these deep neural architectures, there exist a significant heritage in the remote sensing community that has developed advanced algorithms to analyze the signal and structural characteristics of anthropogenic features, such as objects in overhead imagery.

In particular, morphological image processing techniques have been shown to extract objects from imagery and aid in the structural analysis of shapes in a meaningful, human-interpretable way. Specifically, the *differential morphological profile* (DMP), first defined by Pesaresi and Benediktsson [22], has had great success extracting object shapes, while naturally grouping the extracted shapes into scale ranges. Various research, such as [23]–[27], has been enabled by the DMP; which extracts image regions that are either lighter or darker than their contextual setting using *morphological opening* or *closing* operations, respectively. An in-depth summary of the use of the DMP for remote sensing objects, as well as using information fusion constructs, to achieve multi-scale object extractions is presented in [28].

In this paper, we present a novel neural architecture that integrates non-linear image morphology based on the DMP as an initial stage for a convolutional neural architecture that is, by contemporary standards, only moderately deep. Specifically, we present a preliminary *Differential Morphological Profile Neural Net*, **DMPNet**, and we evaluate its performance on a large, challenging HR-RSI benchmark dataset. We explore an initial architecture combining a neural DMP phase for light and dark component segmentation, feeding into convolutional feature extraction, then classical fully connected layers, and then a softmax classifier.

A basic overview of morphology and the DMP is provided in Sect. II. Sect. III details the implementation of the DMP as a pre-convolutional phase of a neural architecture. Experimental data, design, and results are discussed in Sect. IV. Finally, Sect. V provides summary remarks and our future research directions.

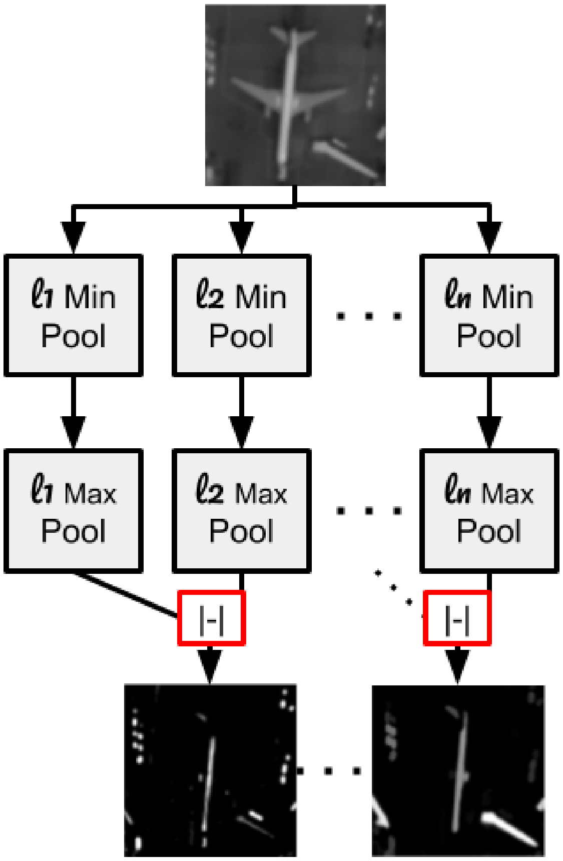


Fig. 1. Differential opening profile component tensor flows. Concurrent flows of an input image pass through an erosion (min pool) then dilation (max pool) sequence, effectively $\gamma^{SE=l}(I)$. The piecewise differential is computed with the absolute difference operation, noted as $|-|$.

II. Differential Morphological Profile Object Extraction

Image morphology is traditionally defined with two base operations on an image I, dilation (δ) and erosion (ϵ),

$$\delta^{SE}(I) = \delta(I) \wedge \texttt{SE}, \tag{1}$$

$$\epsilon^{SE}(I) = \epsilon(I) \vee \texttt{SE}, \tag{2}$$

respectively, where SE is the structuring element, with $\wedge$ and $\vee$ as setwise maximum and minumum, over each pixel neighborhood in I. From the two base operations, we can construct higher level operations such as opening,

$$\gamma^{SE}(I) = \delta^{SE}(\epsilon^{SE}(I)), \tag{3}$$

and closing

$$\varphi^{SE}(I) = \epsilon^{SE}(\delta^{SE}(I)). \tag{4}$$

In terms of a signal, the closing operation with a SE of size n fills in signal valleys (i.e., dark image holes), and conversely the opening operation removes signal peaks. In the 2-D image space, the valleys and peaks represent dark objects in light background context and light objects in dark background context, respectively.

In this work, the concept of the DMP is simplified, namely replacing the morphological reconstruction operations using geodesic SE (see [28] for details) to instead use simplified morphological operations with flat-square SE. As discussed, $\gamma^{SE}(I)$ removes signal peaks, or image regions lighter than their surrounding context, yet smaller than SE. Therefore, given an increasing scale of SE, such as square edges of 3, 5, 7, 9, we can expect that increasingly larger *light objects* are removed. From the strictly increasing set, L, of SE sizes, we can construct an opening profile:

$$\Pi\gamma(I) = \{\Pi\gamma_l : \gamma^{SE=l}(I), \forall l \in L\}. \tag{5}$$

To generate a set of scaled light object extractions, we compute a piecewise derivative of the opening profile (differential opening profile) as

$$\Delta\gamma(I) = \{\Delta\gamma_l : \Delta\gamma_l = |\Pi\gamma^{SE=l}(I) - \Pi\gamma^{SE=l-1}|, \forall l \in L'\}, \tag{6}$$

where $L' = L \backslash min(L)$. In an equivalent manner we can define the closing profile,

$$\Pi\varphi(I) = \{\Pi\varphi_l : \varphi^{SE=l}(I), \forall l \in L\}, \tag{7}$$

and differential closing profile,

$$\Delta\varphi(I) = \{\Delta\varphi_l : \Delta\varphi_l = |\Pi\varphi^{SE=l}(I) - \Pi\varphi^{SE=l-1}|, \forall l \in L'\}. \tag{8}$$

Figure 1 shows the internal tensor operations that implement the differential opening profile, $\Delta\gamma$. The opening operation is computed concurrently for each SE, $l \in L$. Once the set of morphological openings are computed, the piecewise differentials are concurrently computed with an *absolute difference* (i.e., $| - |$) tensor operation within the network. As can be seen in Fig. 1, light objects include an airplane and various ground equipment. The smaller objects are bright responses from the smaller SE in the left most scaled extraction. The large fuselage, which is brighter than the wings, and the jet-way have strong responses, shown in the right-most final image chip.

III. DMP Neural Network

A key motivation of the DMPNet is to simplify the normally extensive network complexity (number of learnable parameters) that results from layers and layers of convolutional feature extraction in traditional DCNN. As discussed in Sect. II, the DMP is a series of non-linear operations that produce scaled object extractions of either light (opening) or dark (closing) objects. In this sense, the two DMP, $\Delta\gamma(I)$ and $\Delta\varphi(I)$ can be defined in network layers, opening-stack ($\Delta\gamma$) and closing-stack ($\Delta\varphi$); specifically, as a set of user specified scaled extractions that are parameter-free in regards to network training. In this way, $\Delta\gamma$ and $\Delta\varphi$ become architectural components of neural network design that perform explicit scaled object extraction tasks within a larger network.

Figure 2 is a diagram of a simple DMPNet architecture. To facilitate a simple architecture, the color image is first transformed into a grayscale image using a 1×1 convolution. This grayscale signal is then concurrently passed into the $\Delta\gamma$ and $\Delta\varphi$ components. In this preliminary examination of the DMPNet architecture, we have used SE $\in \{3, 5, 7, 9\}$ for both the opening and closing profiles. Figure 3 shows the opening profile and closing profile generated by the neural components for a particular image chip. The figure shows a sample from the *Cross-walk* class along with the learned grayscale (1×1 convolved) image, followed by the resulting piecewise differential images for both opening and closing profiles. It can be seen that the stripping pattern from the crosswalk is extracted as both small light objects in dark context, as well as the space between the paint extracted as dark objects in light context. for the lower differentials ($5 - 3$ and $7 - 5$). However, the larger SE differential, $9 - 7$, loses discernible features of the crosswalk.

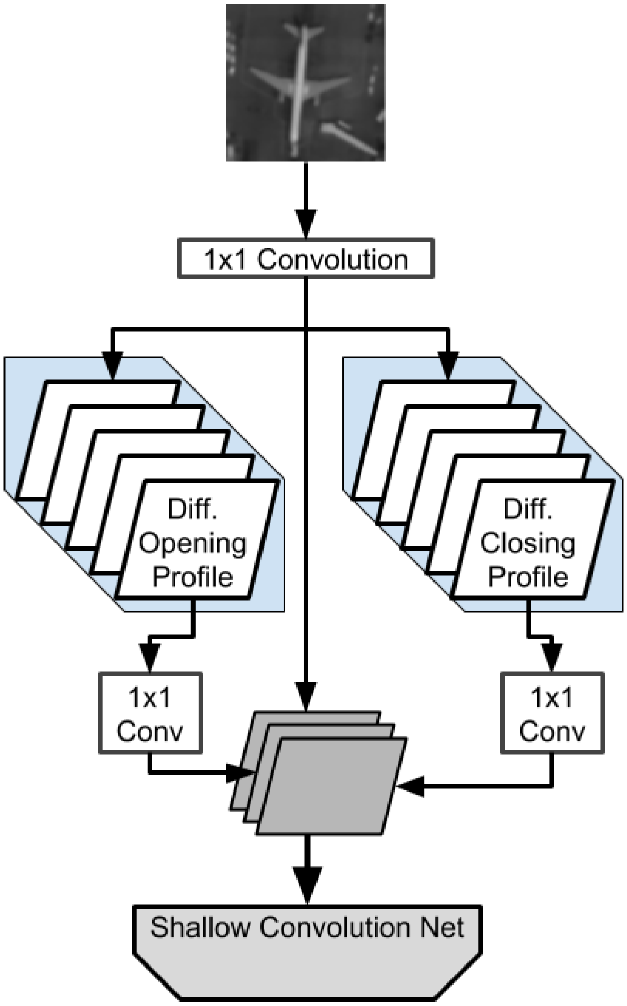
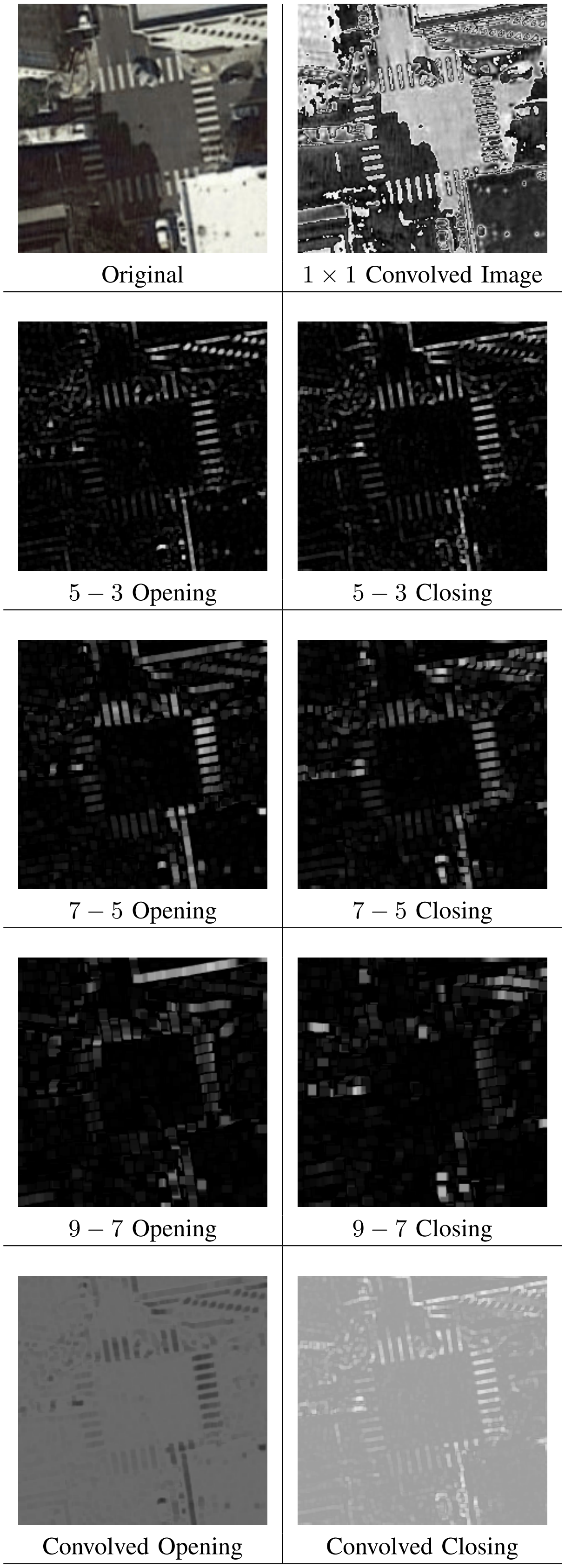


Fig. 2. Differential Morphological Profile Network: The input color image is flattened to grayscale with a 1×1 convolution, then passed simultaneously into $\Delta\gamma$ and $\Delta\varphi$. Each profile is compressed via 1×1 convolutions, then stacked with the grayscale image and fed as a 3-band input into a shallow convolutional neural network.

The output from both $\Delta\gamma$ and $\Delta\varphi$ are independently passed into learnable 1×1 convolutions to flatten into a 2-D tensor; and then they are both stacked with the original grayscale 2-D



Original

1×1 Convolved Image

$5 - 3$ Opening

$5 - 3$ Closing

$7 - 5$ Opening

$7 - 5$ Closing

$9 - 7$ Opening

$9 - 7$ Closing

Convolved Opening

Convolved Closing

Fig. 3. The flattened input image is concurrently processed for opening and closing responses with SE $\in \{3, 5, 7, 9\}$, which are then used to compute $\Delta\gamma$ and $\Delta\varphi$.

tensor into a 3-band, 2-D tensor and passed into a traditional shallow convolutional network. In Fig. 3, the final pair of images in the figure show the resulting (i.e., learned) fusion of the DMP into a multiscale light and dark component extraction. It is important to note here that when comparing the resulting fused DMP images, open for light objects and close for dark objects, the characteristic salient visual cues of a cross walk are present in a mostly gray background. We observe that the unimportant areas appear to be gray and the profiles are convolved into peaks and valleys when rendered as an image signal.

In this initial work, we have used *VGG16* convolutional layers–initialized with *ImageNet* weights–as the shallow convolutional phase. The network then uses a two fully connected hidden layers (1024 and 1024), followed by a traditional *softmax* classifier layer. A key aspect of training the DMPNet architecture is that the convolutions are learned, however, the SEs are actually network design hyperparameters instead of learnable weights. This implies that we are able to extract structural components without relying on learning the extraction with convolutions alone.

The overall design idea is to feed into the convolutional feature extraction phase the grayscale image, a light object structural analysis, and a dark object structural analysis. It should be noted that this is a preliminary design of the DMPNet architecture and numerous improvements will be contemplated for the SE sets, fusion of scaled object extractions, and related transitional architecture that passes the profiles into the shallow convolutional network.

IV. Experimental Evaluation

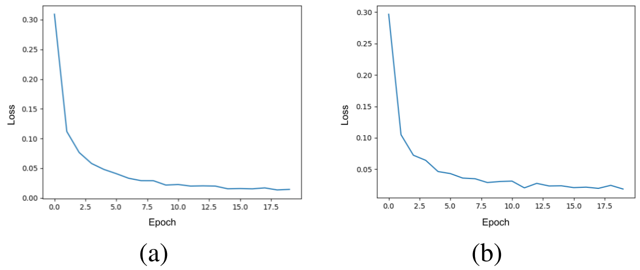
To evaluate this preliminary DMPNet, we leverage a large and challenging HR-RSI benchmark dataset. Following DCNN training insights for HR-RSI from [1], we explore cross-validation performance of the network and compare it to the base VGG-16 model.

A. Evaluation Dataset

In [29], a benchmark meta-dataset (MDS) was developed as an agglomeration of object classes from four previously existing land cover and object detection remote sensing datasets. The MDS was designed to have increased variability and resolutions within class for objects (intra-class variability). The dataset consists of 33 object classes, with the instance counts per class ranging from 700 for classes such as *Church* or *Palace* to 1655 for class *Overpass/Viaduct*. Table II list all classes and the corresponding class counts. As noted, the MDS is an agglomeration of object classes and is specifically designed for training object detection machine learning models. To this end, it is a natural dataset for evaluation of the DMPNet, which has been designed to generate scaled object extractions within a neural model.

B. Object Classification

To evaluate the suitability of the DMPNet for detection of objects within overhead imagery we conducted five-fold



(a) (b)

Fig. 4. Network Learning: Fold-A loss function during 20 epochs of training for (a) DMPNet, (b) VGG-16, using a 4× rotation augmentation.

TABLE I
CLASSIFICATION PERFORMANCE FOR 4× AUGMENTATION, 20 EPOCHS OF 5-FOLD CROSS VALIDATION

	F1	Weighted F1	Precision	Recall
DMPNet	0.95	0.96	0.96	0.96
VGG-16	0.96	0.97	0.97	0.96

cross-validation experiments using both the DMPNet and the standard VGG-16 network. In the experiments, the folds are identically generated for both the DMPNet and VGG-16, and all training hyperparameters are the same. VGG-16 is initialized with ImageNet weights for the convolutional phase, and new fully connected and softmax classification layers are learned for the MDS dataset. This DMPNet uses $\mathtt{SE} \in \{3, 5, 7, 9\}$ for its object extraction phase, followed by an identical setup of the VGG-16 as the shallow convolutional network.

During the network training, a $4\times$ data augmentation was applied by rotating the image chips through the cardinal orientations, i.e., successive $90°$ rotations. Based on lessons from [1], the directed augmentation as opposed to randomization of training samples is used to enhance the generalization of network learning. The augmented data was passed through for 20 epochs of training in the networks, which should be noted is equivalent to 80 epochs of training on the original dataset. The network has been implemented in the PyTorch deep learning framework and was trained with the Adam optimizer using a 1×10^{-4} initial learning rate. It can be observed in Fig. 4 that the training loss curve for both networks is quite similar. However, we can see that DMPNet appears to have a more smooth loss curve, which may be indicative of a more stable network architecture.

Table I provides the summary cross-validation performance of the DMPNet and the VGG-16. We see that the *recall* of both networks is 0.96, while in *precision*, *F1*, and *weighted F1*, the two networks are very comparable. Table II list the *F1* score for all classes, for both DMPNet and VGG-16. We see that, in most cases, the results are similar. A critical item of note to consider in this context is that the VGG-16 network is processing the color information; whereas the DMPNet is limited to a single band compression (i.e., grayscale) of the color information from the input 1×1 convolutional layer that is prior to the concurrent neural DMP profiles.

TABLE II
PER CLASS F1 SCORES FOR DMPNET AND VGG-16

Class		**Count**	**DMPNet**	**VGG-16**
1	Airplane	1600	0.984	0.988
2	Baseball Field	1600	0.969	0.977
3	Basketball Court	1500	0.965	0.978
4	Bridge	1550	0.966	0.966
5	Church	700	0.767	0.807
6	Coastal Mansion	800	0.988	0.997
7	Crosswalk	800	0.997	0.983
8	Ferry Terminal	800	0.936	0.976
9	Football Field	850	0.959	0.986
10	Freeway	1600	0.949	0.958
11	Golf Course	1600	0.983	0.985
12	Intersection	1600	0.955	0.967
13	Mobile Home Park	1600	0.977	0.987
14	Nursing Home	800	0.941	0.975
15	Oil Well	800	1.000	0.999
16	Overpass / Viaduct	1655	0.954	0.971
17	Palace	700	0.728	0.760
18	Parking Lot	1650	0.981	0.988
19	Parking Space	800	0.995	0.997
20	Railway Station	1550	0.961	0.962
21	Roundabout	700	0.954	0.960
22	Runway	1600	0.973	0.975
23	Runway Marker	800	0.993	0.998
24	Ship	700	0.923	0.943
25	Solar Panel	800	0.996	0.995
26	Stadium	700	0.932	0.931
27	Storage Tanks	1600	0.970	0.980
28	Swimming Pool	800	0.979	0.993
29	Tennis Court	1600	0.976	0.975
30	Thermal Power Station	700	0.921	0.942
31	Track Field	700	0.923	0.952
32	Transformer Station	800	0.980	0.993
33	Wastewater Treatment	800	0.975	0.991

Figures 5, 6, and 7 highlight some key insights and characteristics that affect the performance of the DMPNet. In Figs. 5 and 6, two classes with higher F1 score for the DMPNet over VGG-16 are shown. In each, a pair of sample images is provided for review, followed by the result of the 1×1 convolution of the opening DMP layers, then the 1×1 convolution of the closing DMP layers. The performance of the DMPNet was higher for classes *Cross-walk* and *Oil Well* than for VGG-16. In the samples, we can see that these classes have small components that are key structural aspects of the objects. It is expected that the choice of SE used in the DMPNet is well aligned for these object classes. Specifically, the *Cross-walk* has the white painted lines forming pathway patterns that are extracted in the opening profile as a key salient visual element of the class. In the context of the *Oil Well*, the piping is fairly low contrast, however, the cast shadows of the piping form the salient visual components that are extracted from the closing profile.

In contrast, Fig. 7 shows image samples from two classes in which VGG-16 suitably outperforms DMPNet. These are *Church* and *Ferry Terminal*, both of which are larger structures. In the case of the *Church*, we see the vaulted roofing of the structure that is characteristically large, beyond the scope of the SE used in the DMPNet; instead the edge features are highlighted by the DMP versus more relevant structural

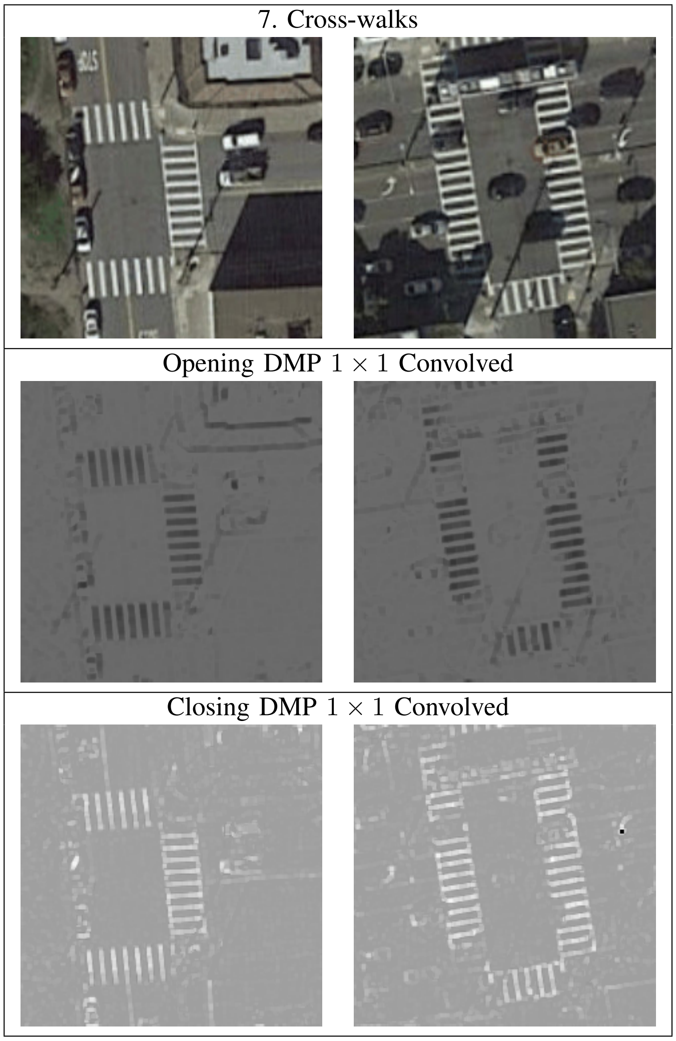


Fig. 5. Example images from the *Cross-walk* class where DMPNet outperformed VGG-16 in F1 score. We see small complex patterns and sub-structures dominate the visual content and were amplified by the set of SE.

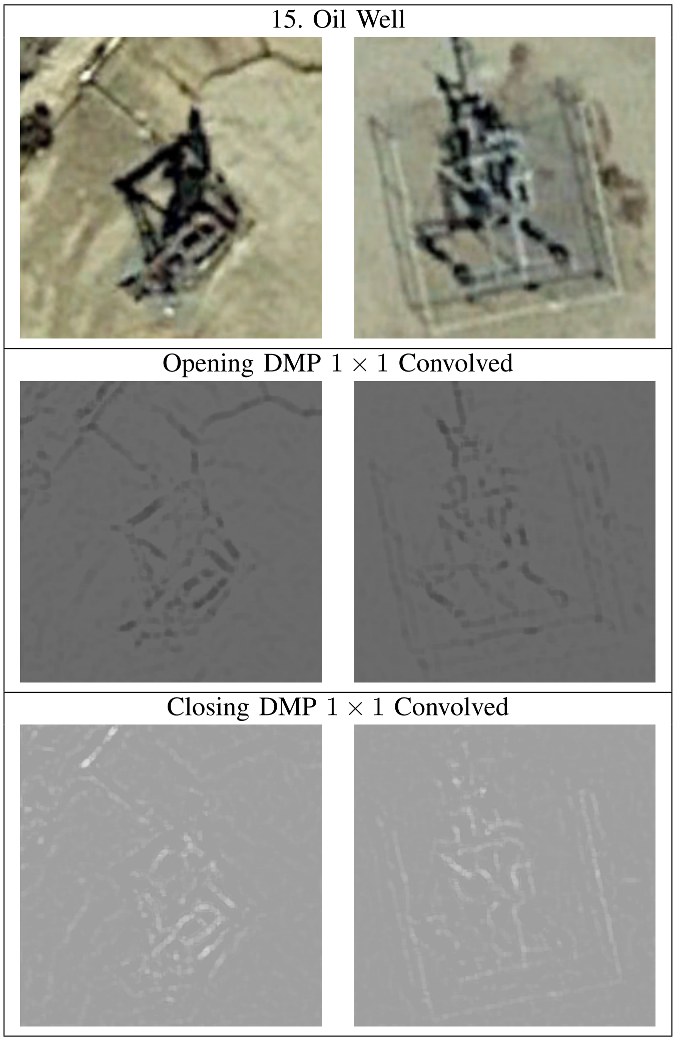


Fig. 6. Example images from *Oil Well* class where DMPNet outperformed VGG-16 in F1 score. We see small complex patterns and sub-structures dominate the visual content and were amplified by the set of SE.

aspects. Similar characteristics exist in the *Ferry Terminal* class, where the long planking is co-existing with boats and resulting DMP images lack discernible visual components.

V. Summary and Future Work

In this paper we have presented a novel neural architecture, the *Differential Morphological Profile Neural Net* – **DMPNet**. This architecture integrates non-linear image morphology as an initial stage of light and dark component extraction for input into convolutional neural components. The DMPNet was evaluated on a HR-RSI benchmark dataset and shown to have comparable performance to an existing convolutional network without the benefit of color information. We have shown that by exploring the DMP activations within the DMPNet, we can gain insights and understanding about the salient visual features that are contributing to the network's object detection. These insights make clear that the structural component object extraction achieved within the DMPNet neural structure can enhance object detection and classification in particular cases.

Future investigations will explore a variety of pathways forward for the DMPNet. These include broader experimentation of the set of SE used, as well as alternative techniques to transition the scaled light and dark component extraction into the convolutional layers. Specifically, we will explore alternatives to the 1×1 flattening of the profiles. Furthermore, techniques from the recently developed neural architecture search (NAS) can also be explored to find optimal differential steps. Additionally, other techniques that are being introduced into contemporary neural architectures, such as capsules, will be explored.

References

[1] G. J. Scott, M. R. England, W. A. Starms, R. A. Marcum, and C. H. Davis, "Training deep convolutional neural networks for land-cover classification of high-resolution imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 4, pp. 549–553, 2017.

[2] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

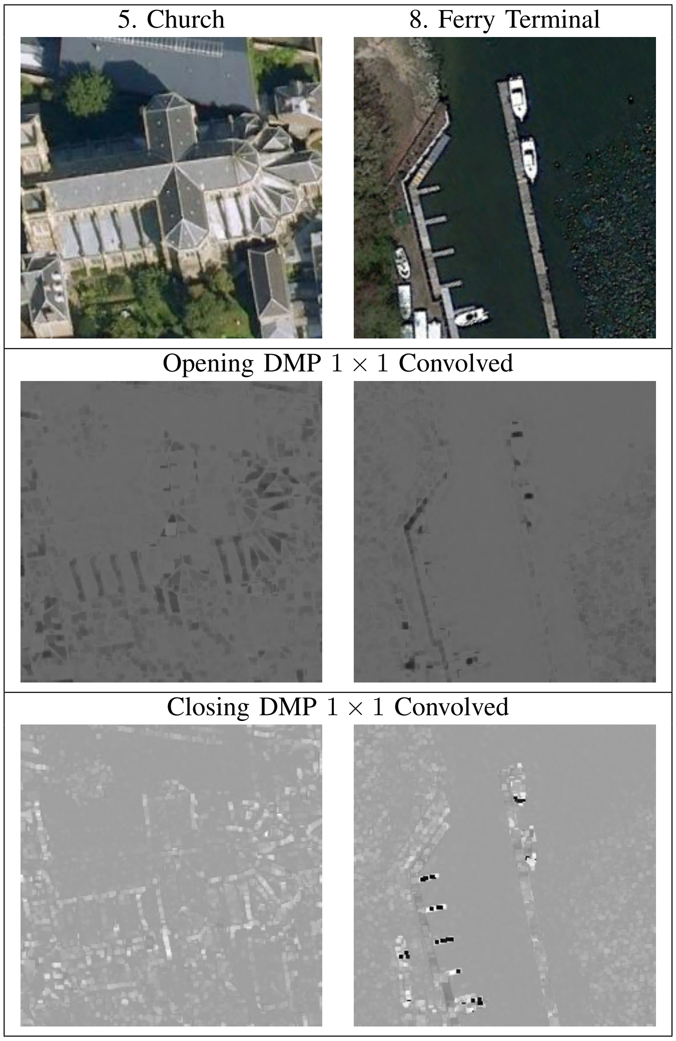


Fig. 7. Example images from *Church* and *Ferry Terminal* classes where VGG-16 outperforms DMPNet. We see large structures lead to only basic edges extracted by DMP with the investigated set of SE.

[3] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich, "Going deeper with convolutions," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.

[4] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna, "Rethinking the Inception Architecture for Computer Vision," *arXiv:1512.00567 [cs]*, Dec. 2015, arXiv: 1512.00567.

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," *arXiv preprint arXiv:1512.03385*, 2015.

[6] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[7] Gao Huang, Zhuang Liu, Kilian Q Weinberger, and Laurens van der Maaten, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, vol. 1, p. 3.

[8] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017, pp. 1800–1807.

[9] Sara Sabour, Nicholas Frosst, and Geoffrey E Hinton, "Dynamic routing between capsules," in *Advances in neural information processing systems*, 2017, pp. 3856–3866.

[10] Alex Yang, J.Alex Hurt, Charlie T.Veal, and Grant J.Scott, "Remote sensing object localization with deep heterogeneous superpixel features," in *2019 IEEE International Conference on Big Data (BIGDATA)*, Dec 2019.

[11] G. J. Scott, R. A. Marcum, C. H. Davis, and T. W. Nivin, "Fusion of deep convolutional neural networks for land cover classification of high-resolution imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 9, pp. 1638–1642, 2017.

[12] D.T. Anderson, G.J Scott, M. Islam, B. Murray, and R. Marcum, "Fuzzy choquet integration of deep convolutional neural networks for remote sensing," in *Computational Intelligence for Pattern Recognition*, Witold and Shyi-Ming, Eds., pp. pp–pp. Springer Berlin Heidelberg, 2018.

[13] G. J. Scott, K. C. Hagan, R. A. Marcum, J. A. Hurt, D. T. Anderson, and C. H. Davis, "Enhanced fusion of deep neural networks for classification of benchmark high-resolution image data sets," *IEEE Geoscience and Remote Sensing Letters*, pp. 1–5, 2018.

[14] J. A. Hurt, G. J. Scott, and C. H. Davis, "Comparison of deep learning model performance between meta-dataset training versus deep neural ensembles," in *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, July 2019, pp. 1326–1329.

[15] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg, "Ssd: Single shot multibox detector," in *European conference on computer vision*. Springer, 2016, pp. 21–37.

[16] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi, "You only look once: unified, real-time object detection (2015)," *arXiv preprint arXiv:1506.02640*, 2015.

[17] Joseph Redmon and Ali Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[18] Adam Van Etten, "You only look twice: Rapid multi-scale object detection in satellite imagery," *arXiv preprint arXiv:1805.09512*, 2018.

[19] Adam Van Etten, "Satellite imagery multiscale rapid detection with windowed networks," in *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2019, pp. 735–743.

[20] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[21] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross B Girshick, "Mask r-cnn. corr abs/1703.06870 (2017)," *arXiv preprint arXiv:1703.06870*, 2017.

[22] M. Pesaresi and J. A. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 2, pp. 309–320, Feb 2001.

[23] M. N. Klaric, G. J. Scott, and C. Shyu, "Multi-index multi-object content-based retrieval," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 10, pp. 4036–4049, Oct 2012.

[24] G. J. Scott and D. T. Anderson, "Importance-weighted multi-scale texture and shape descriptor for object recognition in satellite imagery," in *2012 IEEE International Geoscience and Remote Sensing Symposium*, July 2012, pp. 79–82.

[25] S. R. Price, D. T. Anderson, M. R. England, and G. J. Scott, "Soft segmentation weighted ieco descriptors for object recognition in satellite imagery," in *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2015, pp. 4939–4942.

[26] Aaron K Shackelford and Curt H Davis, "A combined fuzzy pixel-based and object-based approach for classification of high-resolution multispectral data over urban areas," *IEEE Transactions on GeoScience and Remote sensing*, vol. 41, no. 10, pp. 2354–2363, 2003.

[27] G. J. Scott, M. N. Klaric, C. H. Davis, and C. Shyu, "Entropy-balanced bitmap tree for shape-based object retrieval from large-scale satellite imagery databases," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 5, pp. 1603–1616, May 2011.

[28] G. J. Scott and D. T. Anderson, "Fusion of differential morphological profiles for multi-scale weighted feature pyramid generation in remotely sensed imagery," in *2011 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, Oct 2011, pp. 1–8.

[29] J. A. Hurt, G. J. Scott, D. T. Anderson, and C. H. Davis, "Benchmark meta-dataset of high-resolution remote sensing imagery for training robust deep learning models in machine-assisted visual analytics," in *2018 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, Oct 2018, pp. 1–9.