FLOODED AREA DETECTION FROM UAV IMAGES BASED ON DENSELY CONNECTED RECURRENT NEURAL NETWORKS

Maryam Rahnemoonfar[*1], *Robin Murphy*[2], *Marina Vicens Miquel* [1], *Dugan Dobbs* [1], *Ashton Adams* [1]

1: School of Engineering and Computing Sciences, Texas A&M University-Copus Christi, TX
2: Department of Computer Science and Engineering, Texas A&M University, TX

ABSTRACT

The emergence of small unmanned aerial vehicles (UAV) along with inexpensive sensors presents the opportunity to collect thousands of images after each natural disaster with high flexibility and easy maneuverability for rapid response and recovery. Despite the ease of data collection, data analysis of the big datasets remains a significant barrier for scientists and analysts. Here we propose an integration of densely connected CNN and RNN networks, which is able to accurately segment out semantically meaningful object boundaries with end-to-end learning. The proposed network is applied on UAV aerial images of flooded areas in Houston, TX. We achieved 96% accuracy in detecting flooded areas on a large UAV dataset.

1. INTRODUCTION

For quick response and recovery in large scale, access to aerial images are critically important for the response team. The emergence of small unmanned aerial vehicles (UAV) along with inexpensive sensors presents the opportunity to collect thousands of images after each natural disaster with high flexibility and easy maneuverability for rapid response and recovery. Moreover, UAVs can access hard-to reach areas and perform data-gathering tasks that are otherwise unsafe or impossible for humans. Despite the ease of data collection, data analysis of the big datasets remains a significant barrier for scientists and analysts. While traditional analyses provide some insights into the data, the complexity, scale, and multi-disciplinary nature of the data necessitate advanced intelligent solutions. In this paper we present a new deep Convolutional Neural Networks (CNN) architecture for automatically detecting flooded areas in Houston from data collected by UAVs. In recent years CNN have been widely used in computer vision research including classification, human action recognition, object recognition, and semantic segmentation. Semantic segmentation is very important in comprehensive scene understanding, which aims at predicting a class label for every pixel in a given image.

* Corresponding author.

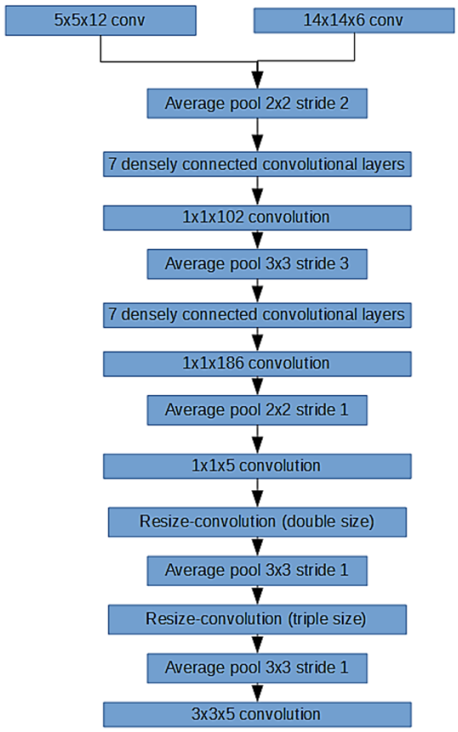

Fig. 1. Our convolutional neural network architecture.

In recent years, UAV has been used in variety of crucial and sensitive applications, including surveillance, reconnaissance, cartography, landslide monitoring, inspection, traffic monitoring, search and rescue, etc. Region-based methods have been applied traditionally for semantic segmentation [1, 2], where only human centric images have been used with limited objects. In [3] CNN has been explored for semantic segmentation of aerial images. In [4] hand-crafted features and CNN learning are combined for semantic labeling satellite imagery. Although CNNs have been shown to be able to automatically learn discriminative features, there are still a lot of ambiguities near object boundaries in the resulting labeling map. For example, sometimes when most areas

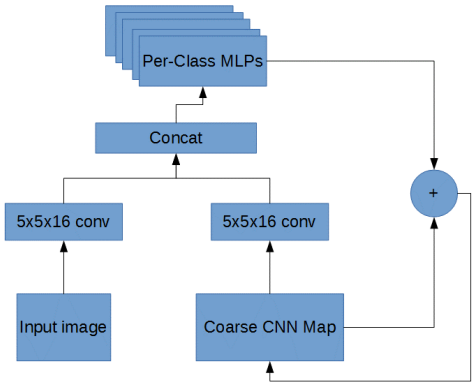

Fig. 2. Our recursive neural network architecture.

of a building are detected successfully, the segmentation map still lacks semantically meaningful shape. This is partially due to the reason of having fewer number of learnable features in CNN in order to learn long-range contextual features might lead to relatively lower spatial accuracy [5]. To mitigate the issue of coarser thematic map, CNN output has been incorporated into the original image as an initialization for further fine-grain classification techniques, such as CRF [6, 7], edge stopping function [8], etc. RNN also has been shown to achieve more accurate segmentation results [9] by using relevant data dependent features automatically. In this paper, we propose to achieve accurate labeling of aerial images with more semantically meaningful segmentation map by developing a dense CNN network incorporated into a Recurrent neural network architecture as initialization for more semantical segmentation of the original input image. Such dense connection helps to improve the information flow and gradients throughout the network, which results in efficient training of deeper networks while still reducing over-fitting problems. Our CNN and RNN architectures are shown in Fig.1 and Fig.2, respectively.

2. DENSE CONVOLUTIONAL NEURAL NETWORKS

Convolutional Neural Networks (CNN) consist of various convolutional and pooling layers that resemble human visual system.Generally, image data is fed into the network that constitutes an input layer and produces a vector of reasonably discriminative features associated to object classes. From input to output layers there are many hidden layers including convolution layers, pooling layers and fully connected layers.

Very deep CNNs has a problem of information wash out as it passes through more layers in the network. Bypassing the information from one layer to the next via individual connections can be a solution to this problem [10, 11]. Another method that can be useful is random dropping of layers while training with better information flow in the network [12]. Huang et al. [13] proposed an architecture to ensure maximum information flow between layers in the network by connecting all layers directly from one to another. In order to preserve the feed-forward nature, each layer obtains additional inputs from all preceding layers and passes on its own feature maps to all following layers. As a result, features are not combined through summation before they are passed into a layer, instead they provide separate inputs to all layers.

The main advantage of the dense connections is the efficiency for training. Moreover, the improved information flow and gradients throughout the network helps to prevent information loss. Each layer has direct access to the gradients from the loss function and the original input signal. This improves training especially for deeper network architectures. Moreover, the short connections have a regularizing effect on the network, which tends to avoid over-fitting in experimental settings with smaller datasets.

3. NETWORK ARCHITECTURE

Our network begins with a 5×5 and 14×14 convolution concatenated together for multi-scale feature extraction. Followed by average pooling with a 2×2 window and a stride of 2, the bulk of the network is then in two modules of 7 densely connected convolutional layers. These modules consist of a sequence of 3×3 convolutions in which every convolution has the results of all previous convolutions as its input. This serves to increase the efficiency of the parameters learned by the network [13]. The sequence of densely connected convolutions is followed by 1×1 convolution and average pooling. The first densely connected module uses 3×3 average pooling with stride 3. The second densely connected module uses 2×2 average pooling with stride 1. Following this, a 1x1 linear convolution is used to produce a low resolution segmentation map. Rather than using deconvolution as in [14], we instead use resize-convolution [15], which alleviate the problems of checkerboard artifacts better compared with deconvolution layers. This technique consists of a nearest neighbor resize followed by convolution. The first resize-convolution doubles the size of the image. It is followed by 3×3 average pooling with stride 1 and zero-padding to keep the image size invariant. The segmentation map is then tripled in size by resize-convolution, which is followed by another padded average pooling with a 3x3 window and stride 1. The final layer in the network is a 3×3 convolution. This produces a segmented map.

On top of this convolutional neural network, a recurrent neural network (RNN) is applied to refine the output. The structure of the RNN is as follows. Separate 5×5 convolutional filters are applied to both the coarse CNN result and the original input image. These filters are concatenated together, and then a multi-layer perceptron (MLP) is used for

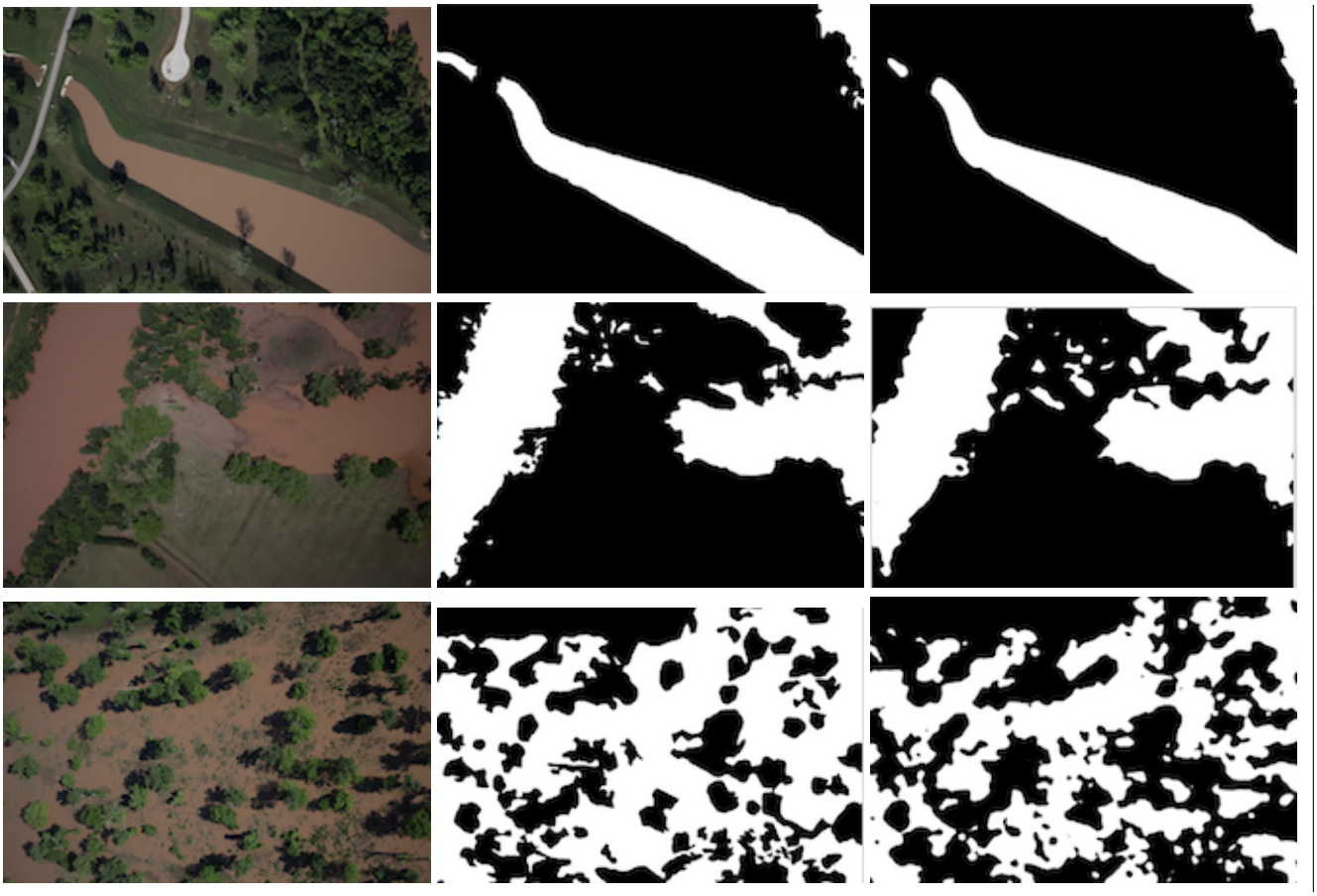

Fig. 3. First column: Original Images, second column: Labeled Images, third column: Our proposed method

each class. The result of these MLPs is added to the coarse CNN result. Due to the larger images used as inputs in this work, only 3 iterations of the RNN can be trained on a single GPU.

4. EXPERIMENTAL RESULTS

Images in this work are captured by optical cameras mounted on UAV. We study the Houston area, Texas, after the flood. The CNN and RNN were trained separately due to the memory constraints imposed by high resolution images. The CNN was trained using Adam optimization [16] with a learning rate of $1e$-5 and batch size 12. The RNN was trained using Adagrad [17] with a learning rate of 0.01 and a batch size of 8. Both parts of the network were trained for $35,000$ steps.

For the quantitative analysis of the results, we calculated accuracy, precision, recall, and IoU (intersection over union), where IoU is a widely used metric for evaluating segmentation performance. The mean IoU is defined as,

$$IoU = \frac{1}{N} \sum_{i} \frac{n_{ii}}{t_i + \sum_j n_{ji} - n_{ii}}, \quad (1)$$

where n_{ji} is the number of pixels of class j predicted to belong to class i out of N classes, and t_i is the total number of pixels of class i. For the semantic segmentation of flooded area in our UAV dataset we got accuracy of 96% and MIoU of 92%. We have also compared our results with an structure similar to Maggiori, et.al. [18]. Accuracy and MIoU based on [18] are 92% and 84%, respectively.

Fig 3 shows some qualitative segmentation results. First column shows the original image from UAV. Second column shows the manually labeled image (ground-truth). Third column present the segmentation results obtained by our proposed network. By visual interpretation of the results, we can find out that our proposed method is able to detect all flooded areas, even some times better than manually labeled data.

5. CONCLUSION

In this paper we propose a novel deep convolutional and recurrent neural network for semantic segmentation of UAV imagery, where it is extremely difficult to semantically label small objects with insufficient object features and large ambiguities of target signatures. To address the issue of large

segmentation ambiguities along object boundaries, we propose an end-to-end integration network where a densely connected CNN is incorporated with RNN to produce more accurate and semantically meaningful segmentation maps. Experimental results demonstrate the efficiency and effectiveness of the proposed network for semantic segmentation both qualitatively and quantitatively.

6. ACKNOWLEDGMENT

This project is supported in part by Amazon Academic Research Award (AARA) and Texas Comprehensive Research Fund (TCRF).

7. REFERENCES

[1] X. Ren and J. Malik, "Learning a classification model for segmentation," in *Proceedings Ninth IEEE International Conference on Computer Vision*, Oct 2003, pp. 10–17 vol.1.

[2] J. M. Gonfaus, X. Boix, J. van de Weijer, A. D. Bagdanov, J. Serrat, and J. Gonzlez, "Harmony potentials for joint classification and segmentation," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June 2010, pp. 3280–3287.

[3] Volodymyr Mnih, "Machine learning for aerial image labeling," *University of Toronto*, 2013.

[4] S. Paisitkriangkrai, J. Sherrah, P. Janney, and A. van den Hengel, "Semantic labeling of aerial and satellite imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 7, pp. 2868–2881, July 2016.

[5] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 640–651, April 2017.

[6] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. H. S. Torr, "Conditional random fields as recurrent neural networks," in *2015 IEEE International Conference on Computer Vision (ICCV)*, Dec 2015, pp. 1529–1537.

[7] H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," in *2015 IEEE International Conference on Computer Vision (ICCV)*, Dec 2015, pp. 1520–1528.

[8] Liang-Chieh Chen, Jonathan T. Barron, George Papandreou, Kevin Murphy, and Alan L. Yuille, "Semantic image segmentation with task-specific edge detection using cnns and a discriminatively trained domain transform," *CoRR*, vol. abs/1511.03328, 2015.

[9] E. Maggiori, G. Charpiat, Y. Tarabalka, and P. Alliez, "Recurrent neural networks to correct satellite image classification maps," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 9, pp. 4962–4971, Sept 2017.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, pp. 770–778.

[11] Rupesh Kumar Srivastava, Klaus Greff, and Jürgen Schmidhuber, "Highway networks," *CoRR*, vol. abs/1505.00387, 2015.

[12] Gao Huang, Yu Sun, Zhuang Liu, Daniel Sedra, and Kilian Q. Weinberger, "Deep networks with stochastic depth," *CoRR*, vol. abs/1603.09382, 2016.

[13] Gao Huang, Zhuang Liu, and Kilian Q. Weinberger, "Densely connected convolutional networks," *CoRR*, vol. abs/1608.06993, 2016.

[14] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015, pp. 3431–3440.

[15] Augustus Odena, Vincent Dumoulin, and Chris Olah, "Deconvolution and checkerboard artifacts," *Distill*, 2016.

[16] Diederik P. Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *CoRR*, vol. abs/1412.6980, 2014.

[17] John Duchi, Elad Hazan, and Yoram Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *J. Mach. Learn. Res.*, vol. 12, pp. 2121–2159, July 2011.

[18] Emmanuel Maggiori, Yuliya Tarabalka, Guillaume Charpiat, and Pierre Alliez, "Convolutional neural networks for large-scale remote-sensing image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 2, pp. 645–657, 2017.