Image segmentation and region classification on Multi-Spectral Images using Deep Learning

Philip Haddad

Eötvös Loránd University, Budapest, Hungary

1 Introduction

Semantic segmentation algorithms assign a label to every pixel in an image. In remote sensing, semantic segmentation is often referred to as image classification, and semantic segmentation of non-RGB imagery has numerous applications, such as land-cover classification, vegetation classification, urban planning, and natural disasters such as floods, earthquakes, hurricanes..etc. And after the breakthrough in Deep Convolutional Neural Networks (DCNNs) which had a significant performance in the different computer vision topics such as Semantic segmentation, however, it's all about the RGB images and not about the non-RGB or Multi-Spectral Images (MSI) -which are the images that have more than three channels.- and with the lack of the annotated datasets that DCNNs need, so in this paper, we are going to work on methods that will enable us to use the power of DCNNs with MSI.

Keywords: Multi-Spectral Images· Synthetic imagery · Aerial Images · Image Segmentation.

2 Related Works

2.1 Synthetic Image Generation using DIRSIG

Because there are no publicly-available ImageNet sized datasets for non-RGB sensor modalities makes it difficult to directly train a deep neural network for semantic segmentation, because it will be prone to overfitting, for that int the literature [1] they used DIRSIG software to build a large synthetic labeled dataset for the semantic segmentation of aerial scenes and call it RIT-18, we will talk about it in the following section.

2.2 Fully-Convolutional Deep Networks for Semantic Segmentation

We can divide this section into many branches; the first one is the DCNNs whose architecture is based on the U-Net [2] architecture and the second one is based on SegNet [3].

2.3 U-Net based

In the literature [1]; they adapted two fully-convolutional deep neural networks for the semantic segmentation of MSI: SharpMask and RefineNet, both of which were applied to the RIT-18 dataset.

SharpMask was developed by Facebook to be a lightweight and faster version of DeepMask network to augment feedforward nets for object segmentation with a novel top-down refinement approach for RGB Images [4], The network is broken into the convolution, bridge, and segmentation sub-networks and it use layers of the ResNet model. It only uses the first four macro-layers. A macro-layer contains the convolution, batch normalisation, and ReLU activation layers right up to where the feature map is down-sampled by a factor of 2x. This corresponds to the first 40 ResNet convolution layers.[1]

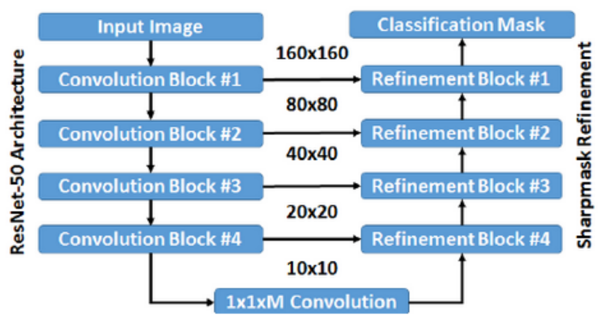

Fig. 1: SharpMask Architecture[1]

RefineNet a generic multi-path refinement network that explicitly exploits all the information available along the down-sampling process to enable high-resolution prediction using long-range residual connections [5]. In [1] they used the same basic structure as the Sharpmask model with a few minor changes.

2.4 Seg-Net based

Copy Initialization Network in the literature [6] is transferred from SegNet [2] whose basis is VGG16. CoinNet is composed of encoder-decoder pairs. An encoder includes several convolutional, batch normalization, Rectified Linear Unit (ReLU) layers as well as a max-pooling operation. The corresponding decoder has similar components, where the max-pooling is replaced by an upsampling. The indices of max-pooling locations were stored and passed to the decoder. where a total of 13 convolutional layers are observed. Different from the original SegNet, the last convolution layers of the decoders contain 16 filters, and correspondingly they will generate 16 label maps for the test images.

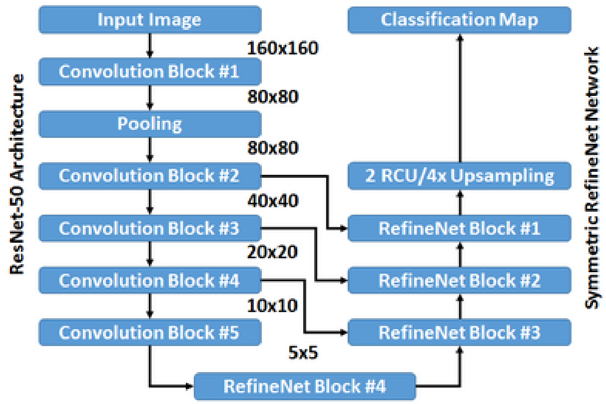

Fig. 2: RefineNet Architecture [1]

Encoder ↓	Decoder ↑
	softmax
Input (6-band image)	conv3-16-1
conv3-64-2	conv3-64-1
maxpool	upsample
conv3-128-2	conv3-128-2
maxpool	upsample
conv3-256-3	conv3-256-3
maxpool	upsample
conv3-512-3	conv3-512-3
maxpool	upsample
conv3-512-3	conv3-512-3
maxpool	upsample

Fig. 3: CoinNet Architecture[6]

3 Dataset Overview

RIT-18 is a high-resolution benchmark, designed to evaluate the semantic segmentation of MSI, collected by a UAS. multispectral data set for validation. Compared with other more popular multispectral/hyperspectral data sets, the main advantage of RIT-18 is that the training and testing sets are separated. Unfortunately the ground truth for the testing set is not published This is an excellent characteristic. Current hyperspectral data sets usually consist of a single image that is randomly sampled for training and testing sets.

The Dataset is a six-band (channels) multispectral data set covering visible and near-infrared regions. Because the ground truth for the testing set is not published, here, we use the original validating set for testing. The size of training data is 9394×5642, that of validation data is 8833×6918 and the test data is 12446x7654.

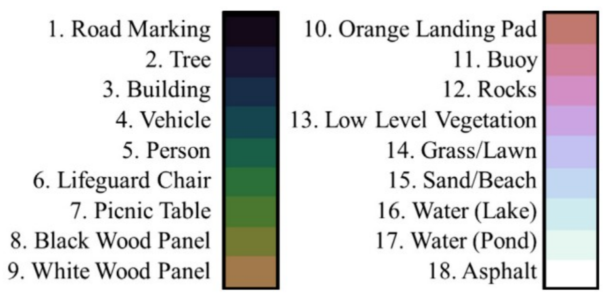

Fig. 4: Class labels for RIT-18

And because of the un-published ground truth of the test dataset, the original validation dataset is used as test set the 20% of the train dataset is used for validation.

The dataset is originally a high-resolution images, so to make it usebale, it must be converting two a smaller images (patches) and each patch has 256x256 pixels and each patch has 6 channels the RGB plus the infrared channels.

Train	Test (prev val)
(792, 6)	(918, 6)

Table 1: Patched dataset

s

4 Proposed Method

As we saw in the literature the main idea to start solving an image segmentation problem is to pick one of the famous architecture in this field and modify it, by changing the encoder or use one of the other machine learning technique.

The proposed model is to use LinkNet[7] which is a novel deep neural network architecture which allows it to learn without any significant increase in number of parameters, and it must must be faster because it has much less number of parameters than the other model and that's lead to work efficiently on the embedded devices or with drones.

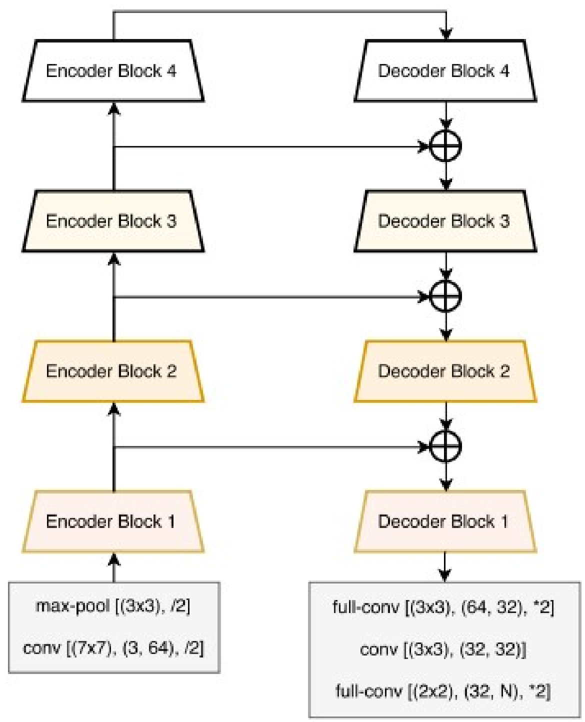

Fig. 5: LinkNet Architecture [7]

Due to the small size of the dataset there is a need to use transfer learning technique and because there is no other huge Multipsectral dataset like ImageNet, we can use RGB-Models and modify the first layer to make it usable with 6-channels input and for that I used ResNet as encoder instead of the original encoder, keeping the connections between the encoder and the decoder.

5 Experiments and results

During the training processes, I tried manually different learning rates (1-2, 1e-3, 1e-4) with Adam Optimizer and Stochastic gradient descent using batch size (16, 32, 64) and Regularization using weight-decay (1e-2, 1e-3, 1e-4).

And the best Average Accuracy is 27%, and here I choose AA instead of other mercies like F1-score) to compare results with the CoinNet in the literature [7] because they tested on the "original" validation data with 52% with considering some differences in the classes appearance during the patching process.

In general the method perform well when materials are conjoint to large areas. However, the results in some small-scale classes are unwarranted. For example, class "Person" and "Buoy" are seldom correctly classified. This drawback is can be explain that during the encoder process small materials can hardly be observed before upsampling. Since the spatial resolution of RIT-18 is 4.7 cm, a single "Person" may account for only dozens of pixels.

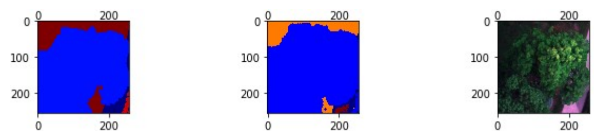

Fig. 6: Target - Pred - RGB-Channels [7]

6 Conclusion

Even with using transfer learning is not that feasible to rely mainly on it to get a robust End-to-End DCNN Model for the semantic segmentation of remote sensing MSI.

References

1. Ronald Kemker, Carl Salvaggio, Christopher Kanan 2017. Algorithms for Semantic Segmentation of Multispectral Remote Sensing Imagery using Deep Learning `https://arxiv.org/abs/1703.06452`
2. Olaf Ronneberger, Philipp Fischer, Thomas Brox 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. `https://arxiv.org/abs/1505.04597`
3. Vijay Badrinarayanan, Alex Kendall, Roberto Cipolla 2016. U-Net: SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation `https://arxiv.org/abs/1511.00561`
4. Pedro O. Pinheiro, Tsung-Yi Lin, Ronan Collobert, Piotr Dollàr 2016. Learning to Refine Object Segments `https://arxiv.org/abs/1603.08695`

5. G. Lin, A. Milan, C. Shen and I. Reid, "RefineNet: Multi-path Refinement Networks for High-Resolution Semantic Segmentation," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 5168-5177, https://doi.org/10.1109/CVPR.2017.549.
6. B. Pan, Z. Shi, X. Xu, T. Shi, N. Zhang and X. Zhu, "CoinNet: Copy Initialization Network for Multispectral Imagery Semantic Segmentation," in IEEE Geoscience and Remote Sensing Letters, vol. 16, no. 5, pp. 816-820, May 2019, https://doi.org/10.1109/LGRS.2018.2880756.
7. Abhishek Chaurasia, Eugenio Culurciello 2017, LinkNet: Exploiting Encoder Representations for Efficient Semantic Segmentation https://arxiv.org/abs/1707.03718.