LAND USE CLASSIFICATION USING DEEP MULTITASK NETWORKS

J. R. Bergado[1],* C. Persello[1], A. Stein[1]

[1] Department of Observation Science, ITC, University of Twente, The Netherlands - (j.r.bergado, c.persello, a.stein)@utwente.nl

Commission III, WG 1

KEY WORDS: Land Use Classification, VHR Imagery, Multitask Learning, Convolutional Networks

ABSTRACT:

Updated information on urban land use allows city planners and decision makers to conduct large scale monitoring of urban areas for sustainable urban growth. Remote sensing data and classification methods offer an efficient and reliable way to update such land use maps. Features extracted from land cover maps are helpful on performing a land use classification task. Such prior information can be embedded in the design of a deep learning based land use classifier by applying a multitask learning setup—simultaneously solving a land use and a land cover classification task. In this study, we explore a fully convolutional multitask network to classify urban land use from very high resolution (VHR) imagery. We experimented with three different setups of the fully convolutional network and compared it against a baseline random forest classifier. The first setup is a standard network only predicting the land use class of each pixel in the image. The second setup is a multitask network that concatenates the land use and land cover class labels in the same output layer of the network while the other setup accept as an input the land cover predictions, predicted by a subpart of the network, concatenated to the original input image patches. The two deep multitask networks outperforms the other two classifiers by at least 30% in average F1-score.

1. INTRODUCTION

Urban land use maps provide essential information on the utilization of urban spaces. Updated information on urban land use allows city planners and decision makers to conduct large scale monitoring of urban areas for sustainable urban growth. Remote sensing data and methods offer an efficient and reliable way to update such land use maps. Using images regularly acquired by spaceborne and airborne sensors provide a much higher degree of objectivity and automation than traditional in-situ mapping methods. In this manner, extensive and updated information on urban land use can be made available on a regular basis.

Transforming large scale remote sensing data into functional land use maps requires advanced classification methods. Knowledge-driven rule sets from object-based classification techniques have been employed for such purpose (Voltersen et al., 2014). However, those require tedious crafting of features extracted from the input data. More recently, deep learning techniques applied on remote sensing data further automated this feature crafting step by learning empirical data representations that are optimized for the classification task (Bergado et al., 2016; Mboga et al., 2017; Huang et al., 2018; Persello et al., 2019; Zhang et al., 2018). Particularly, Huang et al. (2018) used a patch-based convolutional network combined with a post-classification trimming step to classify land use from high spatial resolution multispectral imagery; while Zhang et al. (2018) used an object based CNN to classify land use from very high resolution imagery.

A subset of handcrafted features employed in knowledge-driven land use classification can also be extracted from land cover maps (Voltersen et al., 2014). Such prior information can be embedded in the design of a deep learning based land use classifier by applying a multitask learning setup—simultaneously solving a land use and a land cover classification task. Fully convolutional network variants have also been recently found to be more effective than their patch-based counterparts (Volpi and Tuia, 2017). In this study, we explore a fully convolutional multitask network to classify urban land use from very high resolution (VHR) imagery. To the best of our knowledge, this is the first study to explore performing a land use and a land cover classification simultaneously, in an end-to-end manner.

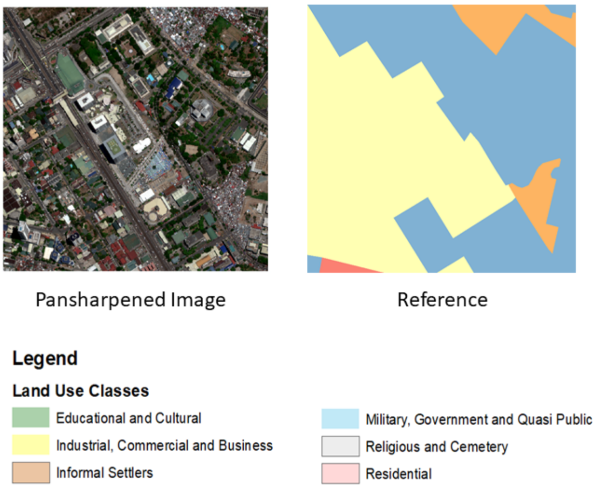

Figure 1. Sample image and corresponding reference for test Tile 1.

2. DATA AND METHODS

In this study, we utilized a deep fully convolutional multitask network to perform urban land use classification from VHR imagery. The dataset comprises of a Worldview-03 satellite image of Quezon City, Philippines acquired on 17th April 2016

* Corresponding author

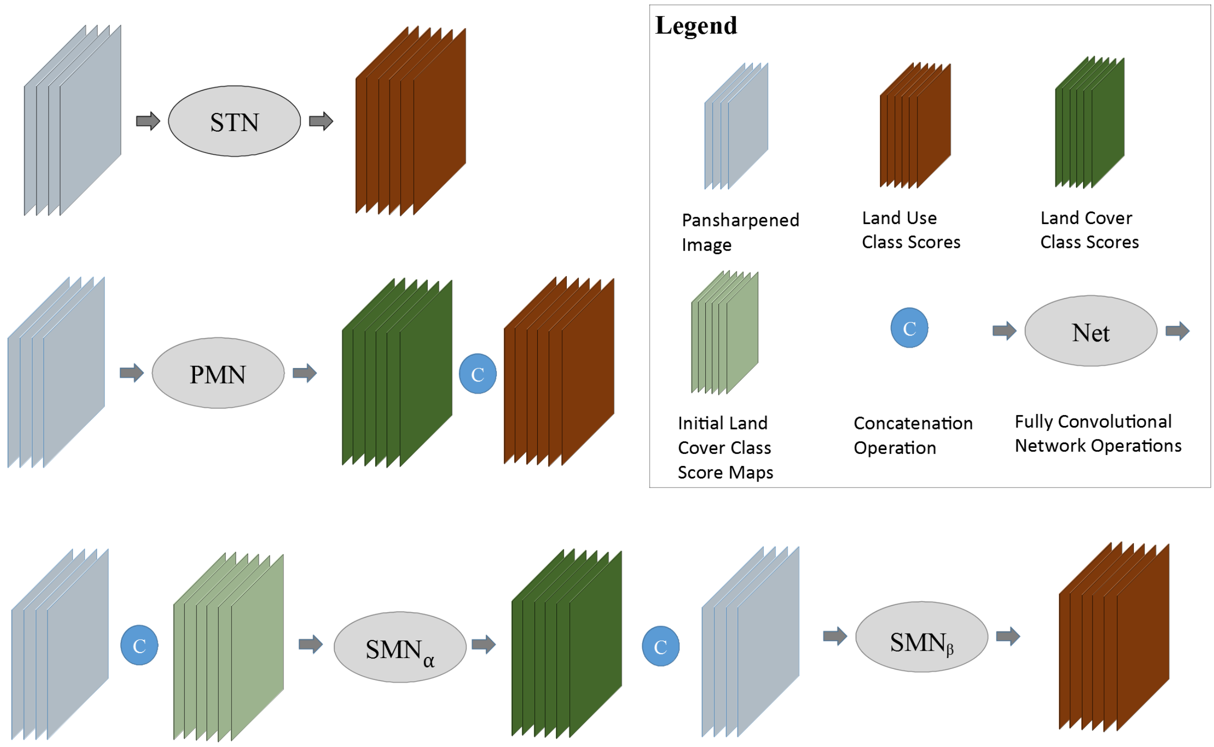

Figure 2. Illustration of the three different fully convolutional networks. STN (single task network) is a standard network only predicting the land use class of each pixel in the image. PMN (parallel multitask network) is a multitask network that concatenates the land use and land cover class labels in the same output layer of the network while SMN (sequential multitask network) accept as an input the land cover predictions, predicted by a subpart of the network, concatenated to the original input image patches.

and manually prepared reference images extracted by updating a Land Use Map of Metro Manila, the capital region where Quezon City is. Fully labeled reference images of land use classes were obtained from this step. Sparsely labeled reference images for the land cover classes to be used by the multitask networks were manually prepared via photointerpretation. The satellite image has a panchromatic band of 0.3 m resolution and four multispectral bands (near-infrared, red, green, and blue) of 1.2 m resolution.

The satellite image was pan-sharpened using the Gram-Schmidt pansharpening technique (Laben and Brower, 2000) and was subdivided into smaller non-overlapping image tiles of size 3200 × 3200 pixels. Twelve tiles were chosen, taking into account the presence of land use classes of interest, and grouped into training, validation, and testing set—six for training, three for validation, and three for testing. Reference images corresponding to the 12 input image tiles were prepared with 6 land use classes: i) educational and cultural, ii) residential, iii) religious and cemetery, iv) informal settlements, v) commercial and industrial, vi) government and military. The image tiles were systematically sampled into smaller non-overlapping 128 × 128 image patches that are then fed as input to the network. The training set was further augmented by two flips and three 90°rotation transformations. Figure 1 shows a sample image and reference tile from the test set.

2.1 Standard approach

We used two methods as baseline approaches to be compared to our proposed methods. Firstly, a pixel-based random forest classifier trained to classify land use from the input pansharpened images. Secondly, a standard fully convolutional network (FCN) classifying land use from the same input pansharpened images; but instead of accepting a 1D input vector of pixel values as done in the random forest classifier, accepts a 3D array of values from the 128 × 128 image patches, and thus, takes spatial context into account. For notational purposes, we call this network STN (single task network).

We used a modified version of U-Net (Ronneberger et al., 2015) as the base network architecture of all our convolutional networks. The weights of the encoder is also initialized from a pretrained VGG16 (Simonyan and Zisserman, 2015). The modification, similarly done by Sherrah (2016), involves adding an additional set of kernels in the first convolutional layer, this additional kernel is initialized from randomly choosing one of the three original kernels from the pretrained VGG16. The number of channels of the output layer was also correspondingly changed to be equal to the number of our target land use and land cover classes.

2.2 Proposed approach

The proposed approach multitask LULC networks has two variants: first is a multitask network that concatenates the land use and land cover class labels in the same output layer of the network; second is a variant that accept as an input the land cover predictions, predicted by a subpart of the network, concatenated to the original input image patches. We call the first variant PMN (parallel multitask network) and the second one SMN (sequential multitask network). The three networks can be represented by the following functions:

$$y_u = \mathrm{STN}(x) \tag{1a}$$

$$[y_u, y_c] = \mathrm{PMN}(x) \tag{1b}$$

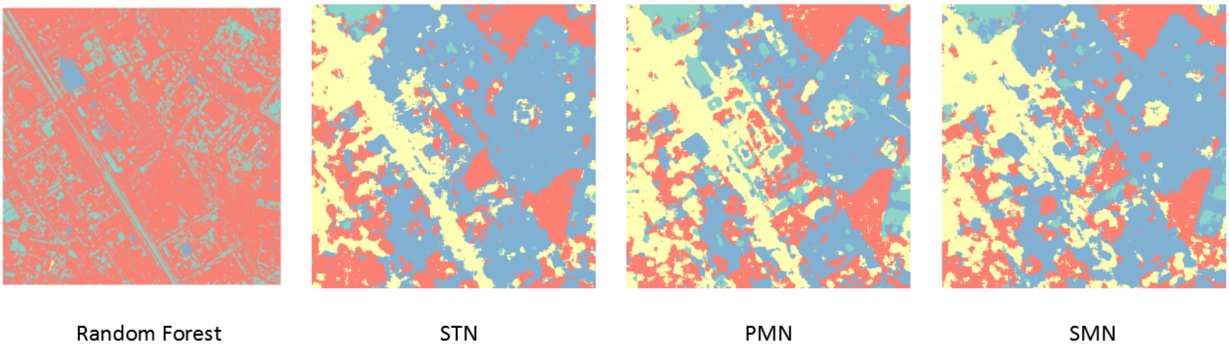

Figure 3. Predicted land use maps of the four classifiers on test Tile 1.

$$\begin{cases} y_c = \mathrm{SMN}_\alpha([x, y_c^0]) & \text{(1c)} \\ y_u = \mathrm{SMN}_\beta([x, y_c]) & \text{(1d)} \end{cases}$$

where x is the pansharpened input image patch, y_u is the output land use predictions, y_c is the output land cover predictions, SMN_α and SMN_β are two sub networks of SMN, y_c^0 is land cover class initialization (in the experiments we initialized all the values to zero), and [ ] is a channel-wise concatenation operation. Equations 1c and 1d are jointly optimized. Figure 2 shows a diagram of the three networks, highlighting the difference between the input and output layers of each network.

All the networks were trained to optimize a cross-entropy loss:

$$E_N = -\sum_{n=1}^{N} \mathbf{t}_n \bullet \log(\mathbf{y}_n) \tag{2}$$

where E is the loss function value evaluated over N samples, $\mathbf{t}_n$ is a binary vector encoding the the target class labels (with the index corrresponding to a class having a value of 1 and 0 otherwise), $\bullet$ denotes the dot product, and $\mathbf{y}_n$ is the class score maps of a sample n calculated using a *softmax* activation function. STN and PMN defines one cross-entropy loss function in their output layers while SMN decomposes the total loss function into two equally-weighted cross-entropy losses at the output layers of SMN_α and SMN_β.

The loss was optimized using Adam (Kingma and Ba, 2014) for 150 epochs utilizing a batch size of 64. The base learning rate used was 0.0001 which was reduced by a factor of 10 every 50 epochs.

Class	Frequency (%)
Educational and Cultural	22
Residential	39
Religious and Cemetery	2
Informal Settlements	3
Commercial and Industrial	19
Government and Military	15

Table 1. Land use class frequency averaged over the whole set of image tiles

Since there is an imbalance in the distribution of the land use classes present in our image tiles (see Table 1), we assessed the classification performance of the four classifiers using the average class F1-score. This metric will be more robust to the class frequency imbalance than the standard overall classification accuracy, the latter generally giving overly optimistic estimates of the classifier performance.

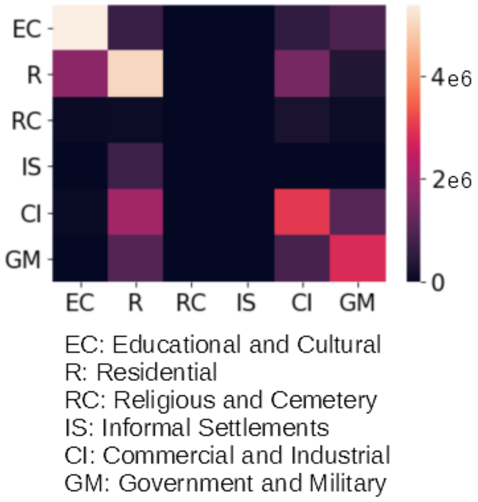

Figure 4. Confusion matrix of PMN on the three test tiles.

3. RESULTS

Classifier	Tile 1	Tile 2	Tile 3
random forest	1.07	14.27	12.36
STN	25.48	21.59	29.63
PMN	57.90	**52.89**	**57.44**
SMN	**59.87**	52.53	53.34

Table 2. Average land use class F1 scores of the classifiers on the three test tiles

The land use classification accuracy of the different classifiers assessed on the three test image tiles (1, 2, and 3) are shown in Table 2. PMN achieves the highest average class F1 score in two of the three test tiles, with SMN having better results for Tile 1. There is a considerable increase in the classification accuracy of the classifier by using a standard fully convolutional network over the baseline random forest classifier. This shows that the features learned in the hidden layers of the convolutional network is helpful for this land use classification task. There is also an observable increase in performance, at least about 30% in average F1-score, when using the two multitask network over the standard one. Such improvements are consistent with the intuition that features extracted from land cover can help the land use classification task.

Figure 3 shows the predicted maps of all the classifiers on Tile 1. All the three networks produced maps of better quality than the baseline random forest classifier. All of three confuses the

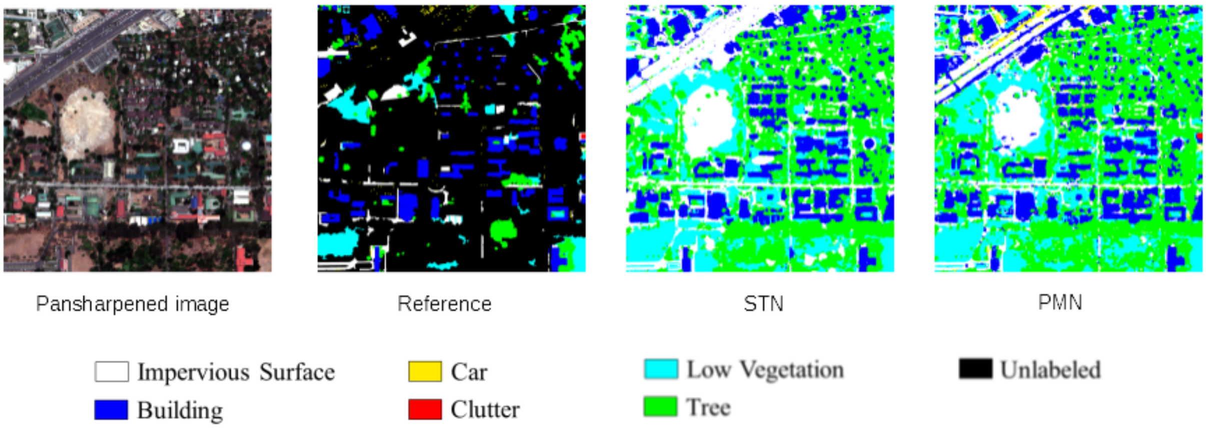

Figure 5. Comparison of predicted land cover maps of Tile 3 using the plain network STN and one of the multitask network PMN.

underrepresented (see Table 1) informal settlement classes as residential areas. This is due to the visual similarity of high density residential areas in the city to informal settlements. This is further affected by the limited number of training labels for this class. This can also be observed in the resulting confusion matrix (see Figure 4) of PMN on the three test tiles where it can clearly be seen the poor performance of the classifier on both the underrepresented two classes (religious and cemetery and informal settlements). On the other hand, the educational and cultural land use class appears to have the least misclassification compared to other classes.

Figure 5 shows a comparison of predicted land cover maps of Tile 3 using the plain network STN and one of the multitask network PMN. The plain network poorly classifies the car class which are confused with building pixels. Predictions of the underrepresented car class were greatly improve by using the multitask learning setup. There is also less overclassification of the impervious surface class after using the multitask network PMN. This shows that the learned features shared by both tasks help on improving the each other's predictions.

4. CONCLUSION

Classification of urban land use maps is essential to provide updated information on utilization of urban spaces. This study shows that performing land use classification simultaneously with classifying land cover improves the resulting classified land use maps. Comparing two multitask networks, we obtained an improvement of at least 30% in the average F1-score as compared to standard classification approaches. Such an approach can be embedded in the design of a deep learning based classifier. The multitask network also improves the predictions on the additionally embedded land cover prediction task.

ACKNOWLEDGEMENTS

We thank the Metro Manila Development Authority (MMDA) for providing us the Metro Manila Earthquake Impact Reduction Study (MMIERS) Land Use Map dataset.

References

Bergado, J. R., Persello, C., Gevaert, C., 2016. A deep learning approach to the classification of sub-decimetre resolution aerial images. *Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS) 2016*, 2016-November, 1516–1519.

Huang, B., Zhao, B., Song, Y., 2018. Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery. *Remote Sensing of Environment*, 214, 73–86.

Kingma, D. P., Ba, J., 2014. Adam: A method for stochastic optimization. cite arxiv:1412.6980Comment: Published as a conference paper at the 3rd International Conference for Learning Representations, San Diego, 2015.

Laben., C., Brower, B., 2000. Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening. U.S. Patent 6011875, 2000.

Mboga, N., Persello, C., Bergado, J., Stein, A., 2017. Detection of Informal Settlements from VHR Images Using Convolutional Neural Networks. *Remote Sensing*, 9(11), 1106.

Persello, C., Tolpekin, V., Bergado, J., [de By], R., 2019. Delineation of agricultural fields in smallholder farms from satellite images using fully convolutional networks and combinatorial grouping. *Remote Sensing of Environment*, 231, 111253.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. N. Navab, J. Hornegger, W. M. Wells, A. F. Frangi (eds), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Springer International Publishing, Cham, 234–241.

Sherrah, J., 2016. Fully Convolutional Networks for Dense Semantic Labelling of High-Resolution Aerial Imagery. *arXiv preprint arXiv:1606.02585*.

Simonyan, K., Zisserman, A., 2015. Very deep convolutional networks for large-scale image recognition. *International Conference on Learning Representations*.

Volpi, M., Tuia, D., 2017. Dense semantic labeling of subdecimeter resolution images with convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 55(2), 881–893.

Voltersen, M., Berger, C., Hese, S., Schmullius, C., 2014. Object-based land cover mapping and comprehensive feature calculation for an automated derivation of urban structure types at block level. *Remote Sensing of Environment*, 154, 192–201.

Zhang, C., Sargent, I., Pan, X., Li, H., Gardiner, A., Hare, J., Atkinson, P. M., 2018. An object-based convolutional neural network (OCNN) for urban land use classification. *Remote Sensing of Environment*, 216, 57 - 70.