Remote Sensing Image Semantic Segmentation Using Feedforward CNNs

Raghad Lotfy (*Media Engineering Technology*) and **Prof. Mohammed Salem** (*Digital Media Engineering*)

The German University in Cairo

amina.sakr@student.guc.edu.eg

Remote sensing image semantic segmentation is the practice of labelling every single pixel in an image that is taken from a satellite or drone according to some designated classes. This is very useful in many aspects. For instance, scientists can easily analyze large areas of land without constantly having to monitor it since doing so is very time consuming and expensive when it comes to resources. Moreover, semantic segmentation of aerial imagery is used by civil engineers for land use analysis, where semantically segmented aerial images of a certain land help them to determine whether or not to construct buildings in such areas.

Literature Review

In an effort to give more focus to small objects in remote sensing images, Kampfmeyyer et. al [6] focus on combining two approaches: patch-based classification and pixel-to-pixel classification, in order to achieve a good accuracy for small objects while still achieving high accuracy for normal objects. The paper starts by testing each approach on its own and records the accuracies of both the patch-based (also known as PB) classification and pixel-to-pixel classification (also known as a Fully Convolutional Network or FCN) on its own.

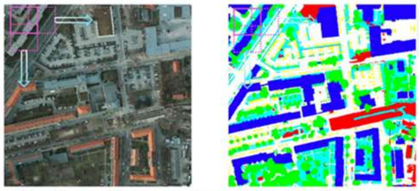
A patch-based pixel classification approach basically involves training a convolutional neural network on small image patches which are taken from large training images. On the other hand, a pixel-to-pixel approach, or a fully convolutional network allows end-to-end learning of pixel labelling. The metrics that are taken into account in order to determine which approach has the highest overall accuracy. These metrics are IoU(Intersection over Union) and overall accuracy for 6 classes, which are: impervious surfaces, buildings, low vegetation, trees and cars. The paper also provides uncertainty maps for each model, in order to determine which model can perform semantic segmentation with the highest accuracy since they are a good measure for pixel-wise uncertainty. The dataset used in this paper is the ISPRS Vaihingen 2D semantic labeling contest dataset [3]. It consists of 33 images of different sizes and each image's pixel count ranges from 3 million to 10 million pixels. The dataset also contains a Digital Surface Model for each of these 33 images and ground truth for 16 of the 33 images. The results of this paper show that the overall accuracy of the patch-based approach is 83.74 percent while the overall accuracy of the Fully Convolutional Network is 86.65 percent (the FCN was trained using cross-entropy loss function). Another FCN model was trained using median frequency balancing (also known as FCN-MFB) in order to take lost classes into consideration, its overall accuracy is 86.48 percent, which is very close to the normal FCN model. For the combined approach, there were 4 results that show different approaches models within this combined approach. Firstly, PB and FCN were combined, and their overall accuracy was 86.84 percent. Then, PB and FCN-MFB were combined to give an overall accuracy of 86.74 percent. Next, FCN and FCN-MFB were combined to give an overall accuracy of 86.98 percent. Finally, all these models were combined to give the highest overall accuracy which is 87.03 percent.

Methodology

1. Dataset

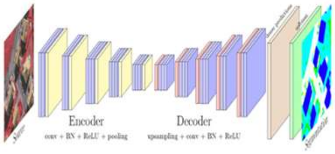
The dataset we're using consists of 38 aerial images of the city of Potsdam.

2. Dataset pre-processing



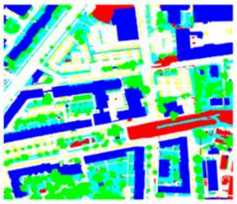
Since the models cannot accommodate the images due to their large dimensions, we will pre-process the images using patch extraction and data augmentation techniques like cropping and rotating.

3. Training the models



Using the suitable learning rate and the optimal number of epochs, we train the models to obtain an output that resembles the ground truth that corresponds to the input image.

4. Predicting the models' performance using the validation set

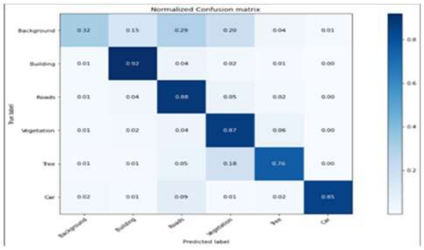


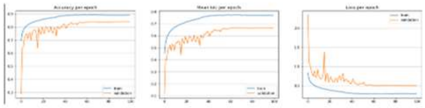
We test the performance of the models by letting them predict the segmented output of a different part of the dataset, which is the validation, or testing part.

5. Comparing the models using evaluation metrics

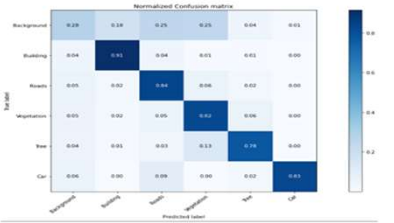
We use evaluation metrics like the confusion matrix and per class IoU of each model in order to determine which model best performs the operation of semantic segmentation on aerial images.

Results





These results were obtained after training the UNet model. The first figure shows the confusion matrix and the second figure shows the accuracy, mean IoU, and loss per epoch.



This confusion matrix is obtained after testing the UNet model by using the validation dataset.

References

[1] S. Kalderon, "A simple guide to semantic segmentation," Jul 2021.
[2] m. m, "Cnn for deep learning: Convolutional neural networks," Jul 2021.
[3] F. Rottensteiner, G. Sohn, M. Gerke, and J. D. Wegner, "Isprs semantic labeling contest,"ISPRS: Leopoldshohe, Germany, vol. 1, p. 4, 2014.
[4] J. Jeong, T. S. Yoon, and J. B. Park, "Towards a meaningful 3d map using a 3d lidar and a camera," Aug 2018.
[5] M. Sebai, "Mrsebai/aerial-tile-segmentation: Large satellite image semantic segmentation into 6 classes using tensorflow 2.0 and isprs benchmark dataset.," Mar 2021.
[6] M. Kampffmeyer, A.-B. Salberg, and R. Jenssen, "Semantic segmentation of small objects and modeling of uncertainty in urban remote sensing images using deep convolutional neural networks," in Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp. 1–9, 2016.
[7] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. 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, 2016.
[8] V. Mnih, Machine learning for aerial image labeling. University of Toronto (Canada), 2013.
[9] M. Kampffmeyer, R. Jenssen, A.-B. Salberg, et al., "Dense dilated convolutions merging network for semantic mapping of remote sensing images," in 2019 Joint Urban Remote Sensing Event (JURSE) , pp. 1–4, IEEE, 2019.

Prepared for Thesis Poster Display Conference

11th -12th June 2022

Faculty of Management Technology
THESIS Poster Display Conference 2022