Residual-Connected Convolutional Neural Network for Road Segmentation

Eric Wolf, Pengfei Ji, Haohui Deng and Guanju Li

group: Coding until midnight, Department of Computer Science, ETH Zurich, Switzerland

July 3, 2017

Abstract

With the advent of artificial intelligence age, road segmentation, using machine learning techniques, has been developing at an astonishing pace. In this paper, we investigate the approaches to build a system that can directly detect and extract roads from aerial images. We start with a basic patch-based convolutional neural network. In order to improve the performance, we propose 1) a powerful pipeline of preprocessing upon the input images; 2) an adaptive-stride approach to generate training samples; 3) an residual-connected structure for better model fitting. Finally, we obtain an accuracy of 0.92712 for the competition.

I. Introduction

Road segmentation is a research topic of both academic significance and practical application nowadays. Currently, road segmentation in real life is still mostly performed by manual labor, which is not only costly but also inefficient. An automatic road detection system would definitely bring many economic and social benefits to the society. As the development of deep learning, many approaches have been proposed to solve semantic segmentation problems from neural nets perspective by patchwise training [1, 2, 3, 4]. Also, pixelwise Fully Convolutional Network shows good result [5].

For patchwise methods, padding is used to make the input compatible with the network architecture. But for deep networks, the original image (before padded) is represented as a very small map, which makes the learning process much more difficult. In this paper, we present an Residual-Connected Convolutional Neural Network, a patchwise segmentation neural net, which has significant performance improvement in comparison to the basic model by tackling the small map problem. We will discuss this explicitly in section II.

For our task, positive samples (road) are much less than negative samples (other objects). At the same time, border (patches that contain both road and non-road pixels) is much more difficult to capture than pure part. To deal with these, we propose a comprehensive preprocessing pipeline and adaptive stride. The details will be discussed in section II.

Section III will give the experimental result we obtain and we will conclude our work in section IV.

II. Methods

i. Preprocessing

- The images are randomly rotated and flipped. Since the convolution is done by only horizontally and vertically shifting a window, we propose to rotate the patches at different time epoch to strengthen the generalization ability of our model.
- Brightness and contrast are changed at random.

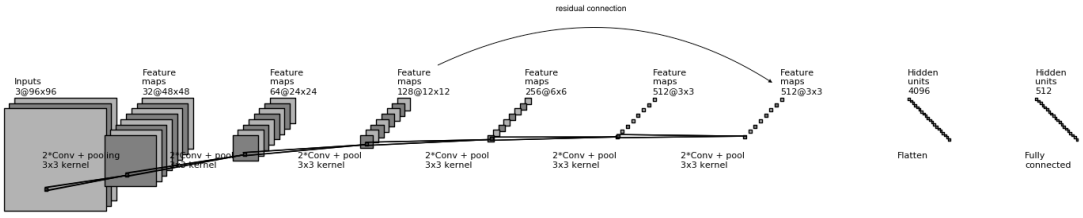

Figure 1: *Architecture of Residual-connected Convolutional Neural Network*

- Small channel shifts are introduced.

ii. Adaptive Stride

Our model is a patchwise segmentation nerual net. Patches are cropped out from each training image. Since border parts are the difficulty for a segmentation model to learn, we care more about them. We use different stride size when cropping patches. When cropping border part the stride size will be smaller. In the contrary, the stride size will be larger if the window is in a pure area. Patches are weighted by the inverse of the presence of the classes in the whole training dataset. This is designed to cope with the class imbalance problem.

iii. Residual Connection

The patches are padded with the reflection of their corresponding images like what has been done in [6]. After doing this, the information about the center of a padded patch should be considered particularly. This is done by bypassing the last two convolution layers in a kind of residual connection. We extract the center part of the feature maps at the convolution layer before the last two convolution layers and concatenate them with the last feature maps, and finally feed them into the fully connected layers. See 1.

III. Experiments

i. Batch Normalization

For a faster convergence and a better performance, we add a batch normalization layer after each convolution layer. This enables a faster convergence of the model and thus in turn improve the experiment results.

ii. Post-processing

Here, we use a scenario-based intuitive post-processing technique for performance improvement. One technique is the single-pixel correction. If one of the pixel, surrounded by other eight pixels, is of the opposite color of the surrounding eight pixels, then the color of the pixel will be modified to correspond to the color of the surrounding pixels. The other technique is the aligning-strip concatenation. If two aligning strips are only separated by a couples of pixels, then the colors of the pixels will be modified to correspond to the color of the two aligning strips. Both of methods are intuitive and based on common knowledge.

iii. Results

We report the accuracy of the basic CNN model, along with each CNN model with our improvements in table 1. This supports our claims that these improvements significantly improve the performance

Here, six models will be discussed respectively to illustrate the significance of our modification on the architecture and preprocessing pipeline. Improvements will be added gradually to present a clear view of the extent of improvement in terms of performance of each technique. The first model is the basic convolutional neural network that starts with a convolutional layer(convolution + relu) and a max pooling layer, followed by a convolutional layer

Table 1: *Average accuracy on baseline and improved models*

Model	Public Score
CNN (2 cov+pooling)	0.76388
CNN (5 cov+cov+pooling)	0.80238
CNN (5 cov+cov+pooling) + preprocessing	0.85687
CNN (5 cov+cov+pooling) + preprocessing + adaptive stride	0.87335
CNN (5 cov+cov+pooling) + preprocessing + adaptive stride + batch normalization	0.90729
CNN (5 cov+cov+pooling) + preprocessing + adaptive stride + batch normalization + residual connection	**0.92712**

and a pooling layer. The second model is a deeper convolutional neural network based on the first model. The third model is an improvement based on the second convolutional neural network, with 5 consecutive convolutional layers and max pooling layer and the input data is preprocessed. The fourth model is the third model, added with adaptive stride.The fifth model is the fourth model with additional batch normalization. The final model is fifth model, added with a residual connection.

This model is based on a patch-wise method, which is originally a $16 * 16$ patch. Here, we padded the patch with the reflection of their corresponding images around it to obtain a $96 * 96$ patch eventually for further processing.

The experiments are conducted under the hyper-parameter parameters listed in 2. Certain hyperparameters, such as threshold , are fined-tuned based on precision, recall and F1 score .

The actual predication image of the residual-connected Convolutional Neural Network can be see at 2. Majority of pixels are well predicted and misclassification rate is low.

Table 2: *Major hyperparameter of residual connected CNN model*

Parameter	Value
Batch size	128
Learning rate	0.0001
Input patch size	96
Threshold	0.44
Epoch	30

IV. Discussion

In this aforesaid experiment, we examine the effectiveness of various improvement on the original model. A deeper model is firstly introduced and examined, which improves the performance significantly since a deeper model have a better capability in terms of modeling non-linear interactions. In addition to this deeper model, we introduce a preprocessing pipeline which shuffle the training data in terms of properties of images, such as orientation, contrast, brightness. This improves our result since it increase the variety of training samples thus in turn the model has a stronger generalization ability. After meticulously examining our predication images, an adaptive stride strategy is introduced to tackle with border pixel misclassification. This is intuitive for semantic segmentation task. Finally, a residual connection is added to retain more information about the center of the image.

The residual-connected is CNN is indeed a significant improvement over the original model, however, certain aspects are still left for discussions and further improvements. One major drawback is that the model fails to deal with the scenario of multi-object segmentation task. The adaptive stride, introduced in this paper, is a rather simplified and intuitive approach for this two-object (road and non road) task. But in the case of multi-object segmentation, the relationship between neighboring

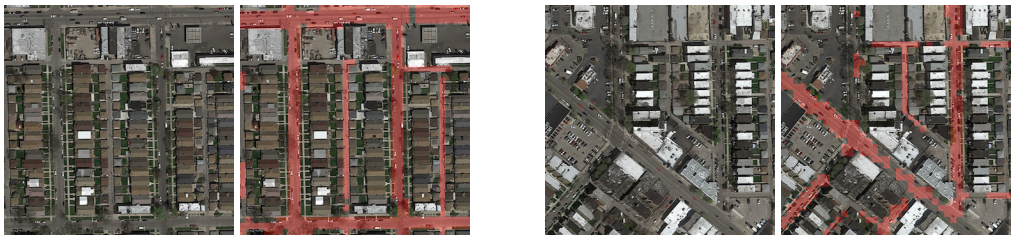

Figure 2: *Sample Predication Image*

pixels may be too sophisticated for such a simplistic strategy.

V. Summary

In this paper, we propose a residual-connected convolutional neural network for road segmentation from aerial images, which combines processing of training data with architecture improvement. Empirical study on given aerial images dataset demonstrates that our model can classify road in a image patch-wise at a high accuracy, which proves the significance of our techniques. A further investigation can be conducted to examine the generalization capability of the residual-connected convolutional neural network, along with the effectiveness of a even deeper model for better performance.

References

[1] D. C. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber. Deep neural networks segment neuronal membranes in electron microscopy images. In NIPS, pages 2852–2860, 2012.

[2] C. Farabet, C. Couprie, L. Najman, and Y. LeCun. Learning hierarchical features for scene labeling. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2013.

[3] Y. Ganin and V. Lempitsky. N4-fields: Neural network nearest neighbor fields for image transforms. In ACCV, 2014.

[4] P. H. Pinheiro and R. Collobert. Recurrent convolutional neural networks for scene labeling. In ICML, 2014.

[5] Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015.

[6] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2015.

ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Declaration of originality

The signed declaration of originality is a component of every semester paper, Bachelor's thesis, Master's thesis and any other degree paper undertaken during the course of studies, including the respective electronic versions.

Lecturers may also require a declaration of originality for other written papers compiled for their courses.

I hereby confirm that I am the sole author of the written work here enclosed and that I have compiled it in my own words. Parts excepted are corrections of form and content by the supervisor.

Title of work (in block letters):

~~Rest~~RESIDUAL-CONNECTED CONVOLUTIONAL NEURAL NETWORK FOR ROAD SEGMENTATION

Authored by (in block letters):

For papers written by groups the names of all authors are required.

Name(s):	First name(s):
WOLF	ERIC
JI	PENGFEI
LI	GUANJU
DENG	HAOHUI

With my signature I confirm that

- I have committed none of the forms of plagiarism described in the 'Citation etiquette' information sheet.
- I have documented all methods, data and processes truthfully.
- I have not manipulated any data.
- I have mentioned all persons who were significant facilitators of the work.

I am aware that the work may be screened electronically for plagiarism.

Place, date

Zurich 03/07

Signature(s)

Eric Wolf
Pengfei Ji
Guanju Li
邓浩辉

For papers written by groups the names of all authors are required. Their signatures collectively guarantee the entire content of the written paper.