# Generating Road Maps Automatically from Aerial Images

Yun Ye
ye.721@osu.edu

## ABSTRACT

Digital maps have been playing an important role in our everyday life. They are in our car navigation systems, smart phones and smart watches *etc*. And we have many online map providers like Google, Bing and Here.

However, traditional ways of surveying and mapping of roads by human beings working on the ground usually takes a lot of time and human labor work. Thus we are proposing here a totally automatic solution by utilizing aerial images with geographical information.

Road maps are automatically generated by: First processing aerial images using a modified deep neural network based on GoogleNet to identify the most likely road part. Then the road parts are assigned with their corresponding geographical coordinates. Finally the image data and geographical data are fusion together in Geotiff format, which could be used for map updating, car navigation, urban planning.

This work differs from previous ones in that besides its implementation of identifying roads from aerial images using a deep learning neural network, it also correlate image pixel information with geographical coordinations. The whole map generation is in a totally automatic way.

## CCS CONCEPTS

• **Information systems** → **Geographic information systems**; • **Computing methodologies** → **Neural networks**; **Image processing**;

## KEYWORDS

Digital maps, deep neural network, image processing, aerial image classification, geotiff images

**ACM Reference format:**
Yun Ye. 2017. Generating Road Maps Automatically from Aerial Images. In *Proceedings of 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Los Angeles, California, USA, November 2017 (ACM SIGSPATIAL 2017),* 4 pages.
https://doi.org/10.1145/1234567.1234567


*ACM SIGSPATIAL 2017, November 2017, Los Angeles, California, USA*

ACM ISBN 123-4567-24-567/08/06. . . $15.00
https://doi.org/10.1145/1234567.1234567

## 1 INTRODUCTION

Road maps are always playing a crucial role in a wide area of human activities and services such as urban planning, disaster management, land use analysis and traffic analysis *etc*., besides its most direct use for navigation.

Previously, extracting road information from aerial images is done by human experts, so the processing work is very costly and time consuming, especially for very large areas. Now people are trying to use more advanced new methods to accomplish this work in a partly, semi or fully automatic way, with the help of computer technologies [19, 20].

We are proposing here an automatic road map generation solution by making use of aerial images with geographical information. We have experimented with data shared by Mnih [18], who firstly deals with aerial images in a deep neural network way [19, 20].

We use a modified network based on GoogleNet[27] and do the binary labeling of different scales of patches for the image processing step. Besides acquiring the roads information of the input images, we also associate the image information with geographical information using the same projection and coordinate system as the images from the Massachusetts road data set[18], which has made this work distinct from previous work. The final output results are Geotiff files containing both image information of roads and geographical information of their corresponding coordinates, which could be used for map updating, navigation, and imported into other softwares like ArcGIS *(*etc).

The rest sections of the paper is organized as follows: First we give a brief survey of the advancing of road extracting methods, especially those related to deep neural network. Second, we talk about how we form up this road identifying problem. Following that, we will start "Experiment and Results" section, with subsections talking about the data sets, data pre-processing, deep neural network framework to be used and how we manipulate the image pixels step by step. The last subsection discusses how we output the final data. In the last "Results and Discussions" section, we briefly summary our work and contributions. Also, we list out several possible directions for continuous future endeavors.

## 2 RELATED WORK

Extracting road networks from aerial imagery has long been a challenging research problem, despite over 40 years' research efforts [1, 7, 8, 13, 16, 17, 19–21, 29]. It has draw decades' of attentions from researches in computer vision, remote sensing, urban planning and traffic analysis *etc* communities as well.

However, it is still not perfectly solved problem to get a satisfying road map Barzohar *et al* introduce a geometric-stochastic model with dynamic programming to find the MAP estimation in each partitioned window [2]. Stoica *et al* uses a Gibbs point process to extract road from remote sensed images [26]. However, such approaches need humans to tune the parameters. Wegner *et al* proposes a higher-order CRF model for road network extraction based on $P^N$-Potts model, the road labeling of which is based on a prior represented by cliques that exhibit linear network structures [29]. Chai *et al* use image consistency and shape priors to sample junction points [3].

Since a deep learning based method is used in this paper, we will focus on the endeavors and progress of this research line from existing publications. Researches on deep neural network have enjoyed a boom since the term "deep learning" was first used in 2006 by Professor Geoffrey Hinton [9]. Since then, deep learning has shown significant improvements in a number of areas including image classification [5], image super resolution [6], image denoising and inpainting [30], recognition [22], classification of hyperspectral data [4], tracking [28] *etc*. We refer readers to [12] for an authoritative and detailed literature review and comparison. And deep neural networks has become the state of art for various visual learning tasks since the pioneering work by Krizhevsky *et al* [11] in 2012. Krizhevsky and Simonyan *et al* has demonstrated that Convolutional Neural Networks (CNNs) are especially effective in image recognition related problems [11, 25]. Actually, CNNs has out performed all other methods for a lot of image problems currently [23].

Mnih and Hinton [19, 20] have proposed learning based methods to be working robustly for appearance changes, where neural network is implemented to label patches according to what categories they belong to. They are the first to use deep neural network to deal with aerial images. Based on their pioneering work, Saito *et al* have successfully applied CNNs to the semantic segmentation of aerial images [24].

For most recent work on aerial images, Mattyus *et al* learn to enhance road maps by combining ground truth labels acquired from Open Street Map (OSM) with aerial images in [16]. And Marcu *et al* proposes to use dual-stream network to combine local and global information complimentarily [14]. Marmanis *et al* try using an ensemble of CNNs to semantic aerial images classification [15].

## 3 PROBLEM DESCRIPTION

We have experimented with the Massachusetts road data sets shared by [18]. And we have followed the similar problem settings in [19] and [24] to perform a semantic segmentation to categorize each pixel in an image with the most likely label it belongs to. The difference is that we will only want to get the information of roads, so we will only need to do a binary classification for each input image, thus *1* standing for road pixel and *0* otherwise. And we will implement a different network structure modified based on GoogleNet [27]. So let $A$ stands for the aerial image and $M$ denotes the corresponding result of road map image. We set $M(i,j)$ to be 1 if the corresponding location $(i,j)$ in the aerial image A is considered to be road, and set it to 0 if otherwise. While theoretically it could predict one pixel at one time in practice we would predict a small patch of labels. Thus, we model a distribution as $p(n(M(i,j),\omega_m)|n(S(i,j),\omega_s))$, where $n(I(i,j),\omega)$ denotes a $\omega * \omega$ patch of image $I$ centered at location $(i,j)$. Thus, We are to learn from data to make predictions for $\omega_m * \omega_m$ road map, dealing with a corresponding $\omega_s * \omega_s$ aerial images centering at the same location.

**Figure 1: Four example images from Massachusetts roads data set**

## 4 MAPPING ROADS FROM AERIAL IMAGES

### 4.1 Data Set

We have done our experiments with Massachusetts roads data set provided by Mnih [18]. The example images are as shown in Figure 1.

The corresponding reference maps to the above mentioned example images in Figure 1 are shown as in Figure 2.

The training set patches contain both positive and negative ones, and we use different sizes of 32∗32, 48∗48 and 64∗64. Besides, we have considered images patches of different directions and we do a rotation step too.

### 4.2 Data Pre-Processing

In the data preprocessing step, several actions have been taken to enhance the input RGB aerial image patches. Normalization is done to all pixel values for all the data by subtracting 128 for each channel, which helps the network convergence. Also, we perform a data augmentation step by horizontally flipping the images which will be used for training.

**Figure 2: Corresponding target maps of Figure 1**

**Figure 3: Corresponding results of Figure 1**

## 4.3 Deep Learning Framework

We implement our learning framework based on GoogleNet [27]. However, we have done some modifications. We add multi-scale patches as input to the network. Also we have removed down-sampling pooling layers of GoogleNet. Besides, we have removed LRN layers too.

## 4.4 Image Processing

We implement the model using the Caffe library [10]. During training, we apply data augmentation by horizontally flipping the images. We show some results in Figure 3 corresponding to example images in Figure 1.

We have verified the results in ArcGIS as shown in Figure 4. From left to right in the first row, they are original input image and output images from deep learning processing step. In the second row, the first image is showing ground truth data of Massachusetts( in green color ) laid on the original image, while the second one the same ground truth data layered on the result image. The vector Shape data is highlighted in green. The Massachusetts Shape file is also provide by [18].

## 4.5 Generating Geotiff Outputs

The experimented data used in this paper are Geotiff images, which are images with embedded georeferencing information [18]. While we have acquired the binary images of roads from the previous steps, we perform an Otsu operation and then get the binary images which will be used for computing geographical information next. we associate the pixels with geographical coordinates from the source '.tiff' files by assigning pixel information with geographical information. The final results are in Geotif format too, which could be directly used as road maps.

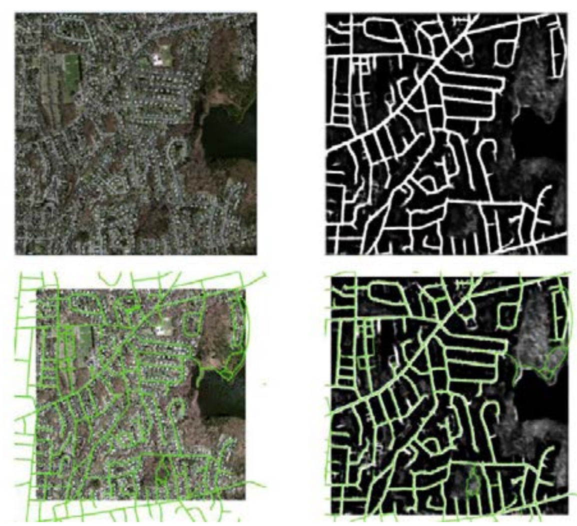

**Figure 4: Experiment results compared with shape files of Massachusetts in software Arcgis. from left to right, they are input and output in the first row, corresponding comparison with ground truth in second row.**

# 5 RESULTS AND DISCUSSIONS

In this paper, we have proposed a fully automatic solution to generate road maps from aerial images with embedded geographical information in Geotiff formats.

Other contributions include: Firstly, we take care of the images using a modified neural network based on GoogleNet to get a road network binary image. Then we filter out possibly less likely road segment parts. Following this, we compute out only the roads' coordinates. Fourthly, we output the road network results in Geotiff format too, the format of which contain both pixel information and corresponding geographical coordinates. Hopefully, researcher from other areas like urban planning, traffic analysis, cartography *etc* could incorporated the results into their future research and work, besides for digital maps updating and navigation systems.

For future work, the author has seen at least the following places where future efforts could be laid on. A first point is about the improved neural network structures as people are always devising and improving new ones that might perform better on certain tasks. A second possible work could be at the geographical information computing stage. For example based on the assumption that roads do not change abruptly, data analysis work could be done to help verify road and none road.

## REFERENCES


[1] Ruzena Bajcsy and Mohamad Tavakoli. 1976. Computer recognition of roads from satellite pictures. *Systems, Man and Cybernetics, IEEE Transactions on* 9 (1976), 623–637.

[2] Meir Barzohar and David B Coope. 1996. Automatic finding of main roads in aerial images by using geometric-stochastic models and estimation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 18, 7 (1996), 707–721.

[3] Dengfeng Chai, Wolfgang Forstner, and Florent Lafarge. 2013. Recovering Line-Networks in Images by Junction-Point Processes. In *Computer Vision and Pattern Recognition*. 1894–1901.

[4] Yushi Chen, Zhouhan Lin, Xing Zhao, Gang Wang, and Yanfeng Gu. 2014. Deep learning-based classification of hyperspectral data. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of* 7, 6 (2014), 2094–2107.

[5] Dan Ciresan, Ueli Meier, and Jürgen Schmidhuber. 2012. Multi-column deep neural networks for image classification. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 3642–3649.

[6] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. 2014. Learning a deep convolutional network for image super-resolution. In *Computer Vision–ECCV 2014*. Springer, 184–199.

[7] Martin A Fischler, Jay M Tenenbaum, and Hans Christoph Wolf. 1981. Detection of roads and linear structures in low-resolution aerial imagery using a multisource knowledge integration technique. *Computer graphics and image processing* 15, 3 (1981), 201–223.

[8] Christian Heipke, H Mayer, C Wiedemann, and O Jamet. 1997. Evaluation of automatic road extraction. *International Archives of Photogrammetry and Remote Sensing* 32, 3 SECT 4W2 (1997), 151–160.

[9] Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh. 2006. A fast learning algorithm for deep belief nets. *Neural computation* 18, 7 (2006), 1527–1554.

[10] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. 2014. Caffe: Convolutional Architecture for Fast Feature Embedding. (2014), 675–678.

[11] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. ImageNet classification with deep convolutional neural networks. In *International Conference on Neural Information Processing Systems*. 1097–1105.

[12] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *Nature* 521, 7553 (2015), 436–444.

[13] Mengmeng Li, Alfred Stein, Wietske Bijker, and Qingming Zhan. 2016. Region-based urban road extraction from VHR satellite images using Binary Partition Tree. *International Journal of Applied Earth Observation and Geoinformation* 44 (2016), 217–225.

[14] Alina Elena Marcu. 2016. A Local-Global Approach to Semantic Segmentation in Aerial Images. (2016).

[15] D Marmanis, J. D Wegner, S Galliani, K Schindler, M Datcu, and U Stilla. 2016. Semantic Segmentation of Aerial Images with AN Ensemble of Cnns. III-3 (2016), 473–480.

[16] Gellert Mattyus, Shenlong Wang, Sanja Fidler, and Raquel Urtasun. 2015. Enhancing Road Maps by Parsing Aerial Images Around the World. In *Proceedings of the IEEE International Conference on Computer Vision*. 1689–1697.

[17] Helmut Mayer, Stefan Hinz, Uwe Bacher, and Emmanuel Baltsavias. 2006. A test of automatic road extraction approaches. *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences* 36, 3 (2006), 209–214.

[18] Volodymyr Mnih. 2013. *Machine Learning for Aerial Image Labeling*. Ph.D. Dissertation. University of Toronto.

[19] Volodymyr Mnih and Geoffrey E. Hinton. 2010. Learning to Detect Roads in High-Resolution Aerial Images. In *Computer Vision - ECCV 2010 - European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings*. Springer, 210–223.

[20] Volodymyr Mnih and Geoffrey E Hinton. 2012. Learning to label aerial images from noisy data. In *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*. 567–574.

[21] Mehdi Mokhtarzade and MJ Valadan Zoej. 2007. Road detection from high-resolution satellite images using artificial neural networks. *International journal of applied earth observation and geoinformation* 9, 1 (2007), 32–40.

[22] Ali Razavian, Hossein Azizpour, Josephine Sullivan, and Stefan Carlsson. 2014. CNN features off-the-shelf: an astounding baseline for recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 806–813.

[23] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, and Michael Bernstein. 2015. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision* 115, 3 (2015), 211–252.

[24] Shunta Saito and Yoshimitsu Aoki. 2015. Building and road detection from large aerial imagery. In *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 94050K–94050K.

[25] Karen Simonyan and Andrew Zisserman. 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. *Computer Science* (2014).

[26] Radu Stoica, Xavier Descombes, and Josiane Zerubia. 2004. A Gibbs point process for road extraction from remotely sensed images. *International Journal of Computer Vision* 57, 2 (2004), 121–136.

[27] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. 2015. Going deeper with convolutions. In *Computer Vision and Pattern Recognition*. 1–9.

[28] Naiyan Wang and Dit-Yan Yeung. 2013. Learning a deep compact image representation for visual tracking. In *Advances in neural information processing systems*. 809–817.

[29] Jan Wegner, Javier Montoya-Zegarra, and Konrad Schindler. 2013. A higher-order CRF model for road network extraction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1698–1705.

[30] Junyuan Xie, Linli Xu, and Enhong Chen. 2012. Image denoising and inpainting with deep neural networks. In *Advances in Neural Information Processing Systems*. 341–349.