Road detection and segmentation by aerial images using CNN based system

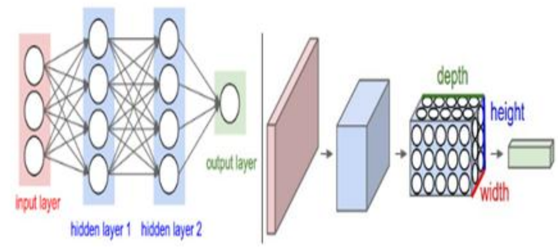
Kurapati.Priyanka[1], Dr. Patnala S.R Chandra Murthy[2]
[1]M.tech, Dept. of CSE, Acharya Nagarjuna University, India
[2]Asst.Profeesor, Dept. of CSE, Acharya Nagarjuna University, India

***Abstract*- In This paper, proposes a system architecture based on deep convolutional neural network (CNN) for road detection and segmentation from aerial images.images are acquired by an unmanned aerial vehicle implemented by the authors. The algorithm for image segmentation has two phases.one is learning phase and another one is operating phase.the input images are decomposed and preprocessed in matlab and partitioned in dimension of 33x33 pixels using a sliding box algorithm these are considered as input into deep CNN.and CNN was design by MatConvNet and some structures. those are four convolutional layers, four pooling layers, one ReLu layer, one full connected layer, and a Softmax layer. The CNN was implemented using programming in MATLAB on GPU and the results are promising.**

Index terms- Machine learning, Classifier, random forest

I. INTRODUCTION

The Extraction of reliable information from aerial images is difficult problem, but it has numerous important utilizations: The disaster monitoring crop monitoring in precision agriculture, border surveillance, traffic monitoring, and so on. different image processing techniques were considered. Texture analysis techniques are used to detect and segment regions of interest and, particularly roads, from aerial images in and but the choice of the representative features depends on the specific context of the application that uses it. The authors in consider also a supervised learning approach to detect road textures using a neural network. UAV (multi-copter type) is proposed for efficient road detection and tracking. Different road features and information as the Stroke Width Transform, colors, and width, are combined to highlight possible road candidates [6]. In order to increase the accuracy and robustness of road detection in a deep Convolutional Neural Network (CNN) was successfully used. For efficient training, in this project the authors proposed the parallel image processing in GPU.a road structure refined CNN (RSRCNN) approach for automatic road extraction in aerial images was proposed.



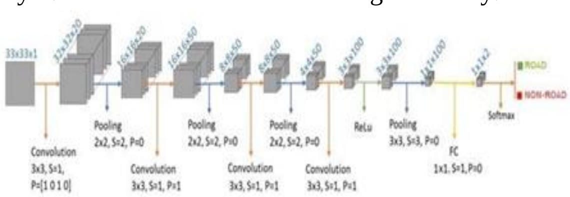
A simple ConvNet is a sequence of layers, and every layer of a ConvNet transforms one volume of activations to another through a differentiable function. We use three main types of layers to build ConvNet architectures: Convolutional Layer, Pooling Layer, and Fully-Connected Layer (exactly as seen in regular Neural Networks). We will stack these layers to form a full ConvNet architecture.
ConvNets transform the original image layer by layer from the original pixel values to the final class scores. Note that some layers contain parameters and other don't. In particular, the CONV/FC layers perform transformations that are a function of not only the activations in the input volume, but also of the parameters (the weights and biases of the neurons). On the other hand, the- RELU/POOL layers will implement a fixed function. The parameters in the CONV/FC layers will be trained with gradient descent so that the class scores that the ConvNet computes are consistent with the labels in the training set for each image. The Convlayer is the core building block of a Convolutional Network that does most of the computational heavy lifting. Each convolutional neuron processes data only for its receptive field. Although fully connected feed forward neural networks can be used to learn features

as well as classify data, it is not practical to apply this architecture to images. A very high number of neurons would be necessary, even in a shallow (opposite of deep) architecture, due to the very large input sizes associated with images, where each pixel is a relevant variable. For instance, a fully connected layer for a (small) image of size 100 x 100 has 10000 weights for each neuron in the second layer. CNNs are often used in image recognition systems. In 2012 an error rate of 0.23 percent on the MNIST database was reported. Another paper on using CNN for image classification reported that the learning process was "surprisingly fast"; in the same paper, the best published results as of 2011 were achieved in the MNIST database and the NORB database.

II. METHODOLOGY

2.1 CNN ARCHITECTURE:

However, some extensions of CNNs into the video domain have been explored. One approach is to treat space and time as equivalent dimensions of the input and perform convolutions in both time and space. The proposed architecture of CNN contains four convolutional layers followed by four layers of spatial reduction (pooling layers), one fully connected layer, one ReLu activation layer and a Softmax layer (Fig. 2).It is a custom created neural network structure, which was trained in a supervised mode. The input images were selected from the UAS implemented by the authors in the research project MUROS [13].The result of a CNN-specific operation (convolution or pooling) is defined as a map of features. Each of the results obtained will have a third dimension, the depth (the number of neurons) of each layer. CNNs have been used in drug discovery.



Predicting the interaction between molecules and biological proteins can identify potential treatments. The system trains directly on 3-dimensional representations of chemical interactions. similar to how image recognition networks learn to compose smaller, spatially proximate features into larger, complex structures, Atom Net discovers chemical features, such as aromaticity, sp3 carbons and hydrogen bonding. Subsequently, Atom Net was used to predict novel candidate biomolecules for multiple disease targets, most notably treatments for the Ebola virus and multiple sclerosis.

III.RESEARCH DATA

This paper develops a new method based on intelligent feature selection for aerial image segmentation. The features investigated derived from the inter-spectral co-occurrence matrices between the color channels RGB and HSV and from LBP approach. The proposed algorithm consists of two phases: the learning and the segmentation, respectively. In the learning phase we proposed a selection algorithm of features for each region of interest based on trace of confusion matrices.

In the segmentation phase we proposed a classification algorithm based on a voting scheme for each selected feature. The experimental results were done on aerial images taken by a fixed-wing type UAV. A new co-occurrence matrix was computed from images represented as three-dimensional tensors, taking into consideration the occurrences between color components of the image (H, S and V). The proposed method has been implemented, and experiments with 1003 general road images demonstrate that it is effective at detecting road-regions in challenging conditions. A set of 100 aerial images from UAV was tested for establishing the rate of correct classification.

IV. PROPOSED METHODOLOGY

The system for the road detection and segmentation from aerial images is presented in Fig. 1 and contains two main modules: UAV module (fixed wing type) and GROUND module. The images taken from UAV' camera, are transmitted via digital data link to GROUND module. In order to detect and segment the roads, successive images are taken with constant rate on the programmed trajectory. The images are saved in the Image Buffer in order to be next processed. The image is firstly decomposed in color components and, for shadow attenuation, only H (Hue) component is considered. Two important operations of primary image processing are included: noise rejection and contrast enhancement. ,the input data for the proposed convolution neural network are monochrome images.

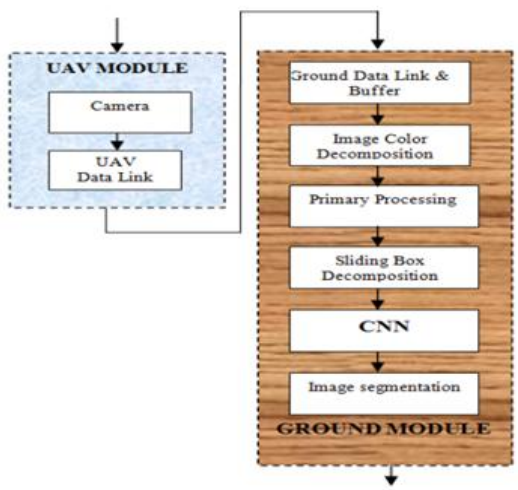
4.1 FIRST NORMAL FORM:

A relation is said to be in first normal form if the input is not considered as a network layer. The first layer is a convolutional layer of 3 × 3 pixels, with the stride 1 and the padding equal to segment. This means that the results are boxes with dimension of 32 × 32 pixels.

4.2 SECOND NORMAL FORM:

Successive divisions with 2 and pooling operations without losses are permitted. The layer contains 20 filters considered as neurons. They are initialized with random numbers from a Gaussian distribution. The second layer in the CNN structure is a pooling layer which reduces the space dimension by half with a sliding box of 2 × 2 pixels. The step is also 2 pixels because there is no need to overlap two blocks.

Transitive Dependency: The second layer in this is convolutional layer does not change the size of the feature maps, but through it the deep is increased to 50 neurons. Its parameters are similar to those of the first convolutional layer, with the difference that the classical padding (P = 1) is made in order to preserve the box size.



4.3 DATA FLOW DIAGRAMS:

The ImageNet Large Scale Visual Recognition Challenge is a benchmark in object classification and detection, with millions of images and hundreds of object classes. In the ILSVRC 2014, a large-scale visual recognition challenge, almost every highly ranked team used CNN as their basic framework. The winner GoogLeNet (the foundation of DeepDream) increased the mean average precision of object detection to 0.439329, and reduced classification error to 0.06656, the best result to date. Its network applied more than 30 layers. That performance of convolutional neural networks on the ImageNet tests was close to that of humans. The best algorithms still struggle with objects that are small or thin, such as a small ant on a stem of a flower or a person holding a quill in their hand. They also have trouble with images that have been distorted with filters, an increasingly common phenomenon with modern digital cameras. By contrast, those kinds of images rarely trouble humans. Humans, however, tend to have trouble with other issues. some extensions of CNNs into the video domain have been explored. Another way is to fuse the features of two convolutional neural networks, one for the spatial and one for the temporal stream. LSTM units are typically incorporated after the CNN to account for inter-frame or inter-clip dependencies. Unsupervised learning schemes for training spatio-temporal features have been introduced, based on Convolutional Gated Restricted Boltzmann Machines and Independent Subspace Analysis.

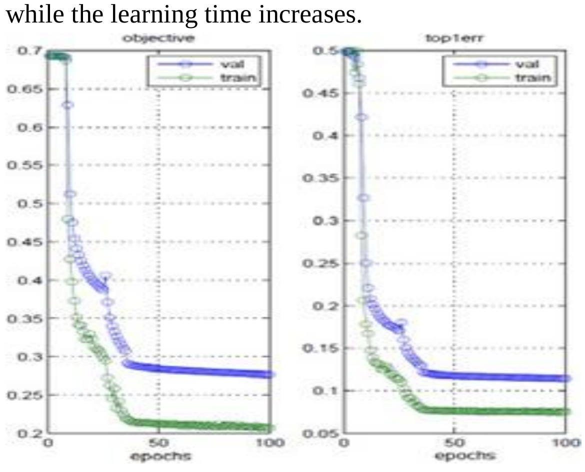
V. EXPERIMENTAL RESULTS

The images were captured along a path generated from parallel lines, with distances between lines of 75m, altitude of200 m, and speed of 70 km/h. Camera type was Sony Nex7, objective 50mm, 24.3 megapixels, and 10 fps. The images were taken by a fixed wing type UAV (MUROS – Fig.3.a) in a real flight. The camera was mounted in gyro-stabilized payload (Fig.3.b).

The images were generated with Agisoft Photoscan Professional Edition [14]. The primary processing filters and color component have very efficient effects on the image. Therefore, the primary processing of images was used both in the learning phase and in the operating phase. The CNN was also implemented in the Ground module, on a computing system for target image evaluation which has the following characteristics: Intel Core i7-4790 CPU, 4.00 GHz, 16.0 GB RAM, Windows 8.1, x 64, GPU Programming in MATLAB [15]. For CNN we used GPU implementation. which ensures high processing speed. The segmentation time was about of 10 s for an image.

5.1 Algorithm Performance.

The above layer segmentation helps to the background layers by that we can processes the target images. the below are the road detected image, targeted image that is result for the testful image. The training files are represented in MatConvNet [11] by.mat files, each file containing a number of lines. Each line icomposed by concatenating the lines from a box and also contains the box label. Thus, a training file is a structure with two fields: data and labels. Each line from a training file contains a vector of 1089 elements, in the data field, which represents a positive or a negative box (33 × 33 pixels), and also the corresponding label, in the label field. For training we used 1000 boxes with asphaltic road (examples in Fig. 4.a) and 1000 boxes with non-road (trees, grass, crops and ground examples in Fig. 4.b) from 50 images. We trained the proposed CNN for 50 epochs (experimentally chosen). For a greater number of epochs, it can be seen (Fig. 5) that both the objective function and the errors decrease very little, while the learning time increases.



VI.CONCLUSION

Road detection is a difficult task in aerial image segmentation due to different size and texture. One of the most important steps in training a CNN is the preprocessing stage. In the case of road segmentation, noise rejection and contrast enhancement techniques had been applied. The second important stage is the selection of the training data. The selected boxes have to cover all the road types from the over flown area (thin and thick roads, with ramifications or without ramifications).The proposed system for road detection and segmentation has the advantage of processing speed, simplicity and possible application to pipeline or river segmentation from aerial images.

REFERENCES

[1] D.Popescu and L.Ichim,"Aerial image segmentation by use of textural features," In Proc.20th International Conference on System Theory,Control and Computing(ICSTCC), Sinaia,Romania, pp. 721–726, October 2016.

[2] H. Kong, J.Y. Audient, and J. Ponce, "General road detection from a single image," in IEEE Trans. Image Process. vol. 19, pp. 2211–2220, 2010.

[3] D. Popescu and L. Ichim, "Image recognition in UAV application based on texture analysis," in ACIVS 2015. LNCS, Springer, Heidelberg, vol. 9386, pp. 693–704, 2015.

[4] V. Mnih and G.E. Hinton, "Learning to detect roads in high-resolution aerial images," In ECCV 2010. LNCS, Springer, Heidelberg, vol. 6316, pp. 210–223, 2010.

[5] H. Zhou, H. Kong, L. Wei, D. Creighton, and S. Nahavandi, "Efficient road detection and tracking for Unmanned Aerial Vehicle," IEEE Trans. on Intell. Transportation Systems, vol. 16(1), pp. 297–309, Feb. 2015.

[6] H. Zhou, H. Kong, L. Wei, D. Creighton, and S. Nahavandi, "On detecting road regions in a single UAV image," IEEE Trans. on Intell. Transportation Systems, vol. 18(7), pp. 1713–1722, July 2017.

[7] T. Ayoul, T. Buckley, and F. Crevier, "UAV navigation above roads using convolutional neural networks," http://cs231n.stanford.edu/reports/2017/pdfs/553.pdf, 2017.

[8] P. Kaiser, J.D. Wegner, A. Lucchi, M. Jaggi, T. Hofmann, and K. Schindler, "Learning aerial image segmentation from online maps," IEEE Transactions on Geoscience and Remote Sensing, vol. 55(11), pp. 6054–6068, Nov. 2017.

[9] Y. Wei, Z. Wang, and M. Xu, "Road structure refined cnn for road extraction in aerial image," IEEE Geoscience and Remote Sensing Letters, vol. 14(5), pp. 709–713, 2017.

[10] Convolutional neural networks for visual recognition. http://cs231n.github. io/ understanding-cnn/. Data Science and Advanced Analytics, pp. 435-441, 2014.

[11] A. Vedaldi, K. Lenc, and A. Gupta, "MatConvNet convolutional neural networks for MATLAB," https://arxiv.org/pdf/1412.4564.pdf, 2018.

[12] CLAHE, Contrast Limited Adaptive Histogram Equalization; Site: https://www. mathworks. com/help/images/ref/adapthisteq.html; Accesat 2017.

[13] MUROS,https://trimis.ec.europa.eu/project/multi sensory-roboticsystem-aerial-monitoring-critical-infrastructure-systems.

[14] Agisoft PhotoScan, www.agisoft.com.

[15] http://www.mathworks.com/products/matlab/.

[16] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," Inf. Process. Manage. vol. 45, pp. 427-437, July 2009.