Copyright Warning & Restrictions

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material.

Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be "used for any purpose other than private study, scholarship, or research." If a, user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use" that user may be liable for copyright infringement,

This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve violation of copyright law.

Please Note: The author retains the copyright while the New Jersey Institute of Technology reserves the right to distribute this thesis or dissertation

Printing note: If you do not wish to print this page, then select "Pages from: first page # to: last page #" on the print dialog screen

ABSTRACT

LAND COVER IMAGE SEGMENTATION BASED ON INDIVIDUAL CLASS BINARY SEGMENTATION

by
Sathyanarayanan Somasunder

Remote sensing techniques have been developed over the past decades to acquire data without being in contact of the target object or data source. Their application on land-cover image segmentation has attracted significant attention in recent years. With the help of satellites, scientists and researchers can collect and store high resolution image data that can be further processed, segmented, and classified. However, these research results have not yet been synthesized to provide coherent guidance on the effect of variant land-cover segmentation processes. In this paper, we present a novel model that augments segmentation using smaller networks to segment individual classes. The combined network trains on the same data but with the masks, combined and trained using categorical cross entropy. Experimental results show that the proposed method produces the highest mean IoU (Intersection of Union) as compared against several existing state-of-the-art models on the DeepGlobe dataset.

LAND COVER IMAGE SEGMENTATION BASED ON INDIVIDUAL CLASS BINARY SEGMENTATION

by
Sathyanarayanan Somasunder

A Thesis
Submitted to the Faculty of
New Jersey Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Computer Science

Department of Computer Science

May 2021

APPROVAL PAGE

LAND COVER IMAGE SEGMENTATION BASED ON INDIVIDUAL CLASS BINARY SEGMENTATION

Sathyanarayanan Somasunder

Frank Y. Shih, Thesis Advisor Professor of Computer Science, NJIT	Date
Usman Roshan, Committee Member Associate Professor of Computer Science, NJIT	Date
Vincent Oria, Committee Member Professor of Computer Science, NJIT	Date
	Date
	Date

BIOGRAPHICAL SKETCH

Author: Sathyanarayanan Somasunder

Degree: Master of Computer Science

Date: May 2021

Undergraduate and Graduate Education:

- Bachelor of Technology,
  SRM Institute of Science and Technology, 2018

Major: Computer Science

Presentations and Publications:

S.Sathyanarayanan, Sumant Agnihotri, Sasidhar Anbazhagan, G.Kalaimagal, "AIMBOT," *International Journal of Pure and Applied Mathematics*, vol.120, pp 1023-1035, 2018.

Either we're going to create simulations that are indistinguishable from reality or civilization will cease to exist, those are the two options.

Elon Musk

ACKNOWLEDGMENT

I would like to thank my advisor, Professor Frank Shih, for his support and expert opinion in completion of my thesis and research. He guided me when I needed direction and gave excellent advice on the subject of research. I would like to thank my committee members Usman Roshan and Vincent Oria, who have helped me in completing my thesis, not only this semester but also parted knowledge on the subject in previous years, without which I would not be able to finish this. I would also thank the Department of Computer Science in giving me the opportunity to finish my thesis.

TABLE OF CONTENTS

LIST OF TABLES

Table **Page**

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Remote sensing techniques have been developed over the past decades to acquire data without being in contact of the target object or data source. Their application on land-cover image segmentation has attracted significant attention in recent years. Image segmentation is the process of partitioning an image into multiple segments based on region and class of the segments. This procedure is intended to better understand and process the image for further analysis. It lays the foundation for future tasks and helps in eliminating outliers in the image data. Without segmentation, a huge amount of image data needs to be manually cropped or cleaned, which is a very time-consuming process.

Of all the many tasks involved in image processing and cleaning of visual data, image segmentation has always been one of the most arduous tasks. Labeling a dataset of thousands of images region by region is a herculean task, limiting the number of datasets and growth in the field. It has always been a precursor to multiple higher-level tasks, such as object detection and recognition. Image segmentation lays the groundwork of bounding boxes, that help other algorithms perform better and remove noise from the images. However, shoddy segmentation results could cause outliers and problems in the future classification processes, and therefore it is one of the most important [23][3].

Image segmentation intends to separate coherent regions from one another. Thresholding is used to extract foreground from background based on pixel intensity or color. This technique can be applied to text and document images and other images with similar clear-cut difference between the two [11].Researchers also use texture and add other factors for a better split across the regions. A more complex

version can range from based on color to using more complex domains to sort out regions. Many filters like the Gabor filter can be used to group texture of a region of pixels[20][15]. The clustering itself can be done with k means or histogram methods. The K means clustering algorithm clusters all the color regions within groups of k. Hence, the regions near to a center chosen by the algorithm are represented by the same color [26][1][12].

But a better way to represent and group data is semantic image segmentation or grouping regions based on their class. This makes a lot more sense in real world especially, where images are filled with noise and a simple addition in the background or outlier can cause large variations in the output classes and trained data. Semantic image segmentation classifies and partitions different classes of an image, with an output mask of different values denoting different regions. This method has been extensively employed with the advent of convolutional neural networks. A general layout adopted from the VGG 16-layer net, and further deconvolution and unpooling has led to pixel-wise masks that segment regions effectively. This method has largely been successful in many industries and various types of images[17].

Primarily in the medical image segmentation, with U-Net and further into V-net, medical image segmentation has never had such higher accuracy with traditional methods of segmentation and classification [10]. Medical imaging and segmentation have always gone hand in hand. Grouping and organization for images has always been a requirement in many fields. But it is more pronounced in the medical field, where such details equate presence of a condition and abnormality. It has been used extensively in segmentation of images of brain, cornea, skin lesions and individual organs. It has also been predominantly used in the segmentation and analysis of cancer and disease detection [29][4][28].

In real life, image segmentation is used to extract more information out of images by segmenting them into classes, especially in the field of automated cars and

security, where image segmentation can identify the different objects and living things in an image [27][18][2]. Another important use of image segmentation is in the field of land cover segmentation. Here, we segment terrain, to classify them into different types of land and water bodies, using images from satellites. This is prominently used in many applications like planning and construction, where gauging a terrain is of tantamount importance. Even in google maps, the terrain is mapped according to its type, to indicate the importance of terrain segmentation.

Land cover image segmentation has also been one of the more difficult types of segmentation, due to the difference in quality of images, noises in the different images and also due to the high number of different features even within the same class of images, although the segmented class is not required to be of high accuracy or precision, this leads to a more challenging case of segmentation [14][9][22].

CHAPTER 2

LITERATURE SURVEY

2.1 Convolutional Neural Networks

Artificial neural networks have come a long way since its inception in the 90s. With advancements in both computer hardware and techniques, we are able to process more data, extract more features and create more complex networks. In this regard convolutional neural networks have been a milestone in the world of computer vision, unlike traditional neural networks, they are able to correlate spatial differences in images and extract features that make a lot more sense than traditional networks.

Basically, as the name implies, convolutional neural networks use the operation of convolution to extract features, an age-old technique used for generations of image processing. Originally, convolution shows how one signal is modified when passed through another. In image processing and computer vision, convolution is based on filters. The output of convolution be how an image is changed when it is passed through a filter. Convolutional Neural Networks use the same principle, to extract features from images.

There have been many a neural network that have built on this principle. But the main usage of convolutional neural networks, lie in their worth in computer vision. With large image sizes, artificial neural networks also need to be sufficiently large to accommodate the data, but with convolutional neural nets, this problem is reduced exponentially. Since with each layer the data size becomes smaller and smaller. So, more features are extracted with lesser weights to be stored. These models have extensively been used in object detection and classification models such as VGG16 and ResNet families[24][10]. This method also serves as the backbone of image segmentation.

Fully convoluted network (FCN) was the first milestone in semantic segmentation. It produced class wise pixel by pixel classification or segmentation masks using convolutional neural networks. Traditionally, convolutional neural networks produce a single output, based on class or detection of objects. But in FCN we extrapolate this data to produce masks that are based on the class or detected object and represent the regions of the image[16].

2.2 Convolution-Deconvolution Based Networks

With the advent of FCNs, new possibilities were opened to the world of semantic segmentation. The features extracted could not only be output for image classification, but now it can also be used as input for image segmentation tasks. With this, many a network were built to profit from the features extracted by doing so.

As the name implies, these networks have two parts to them – the backbone and classifier. The backbone is a general convolutional neural network, which is usually a VGG16 or a ResNet model, that act to extract features using the conventional convolution, downsampling and maxpooling layers, that extract spatial features for identifying and classifying objects in the image.

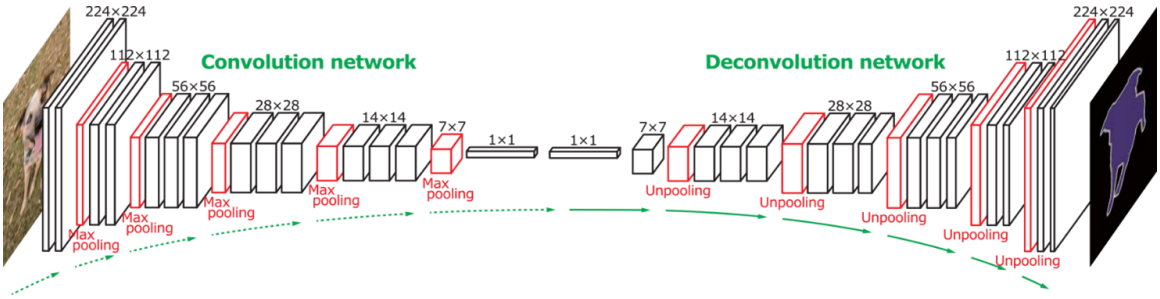

Figure 2.1 Convolution-Deconvolution architecture.
Source: [17]

But what is different from traditional convolutional neural networks, is that this is then fed forward to a decoder – deconvolution network. The deconvolution network does the exact opposite of the backbone. It upsamples the data and deconvolutes to

produce more pixels. Finally, with the help of a softmax layer, the mask that contains the segmentation is then produced.

U-Net, introduced by Ronneberger et al., originally invented to be used on brain image segmentation, has been very robust in its usage and has produced exceptional results in all types of image segmentation. It has two parts – down-sampling and up-sampling, at each down-sample the feature map is convoluted by 2 x 2, which halves the feature channels followed by a ReLu. They are up-convoluted along the up-sampling phase by 2 x 2 [6]. Figure 2 details the same.

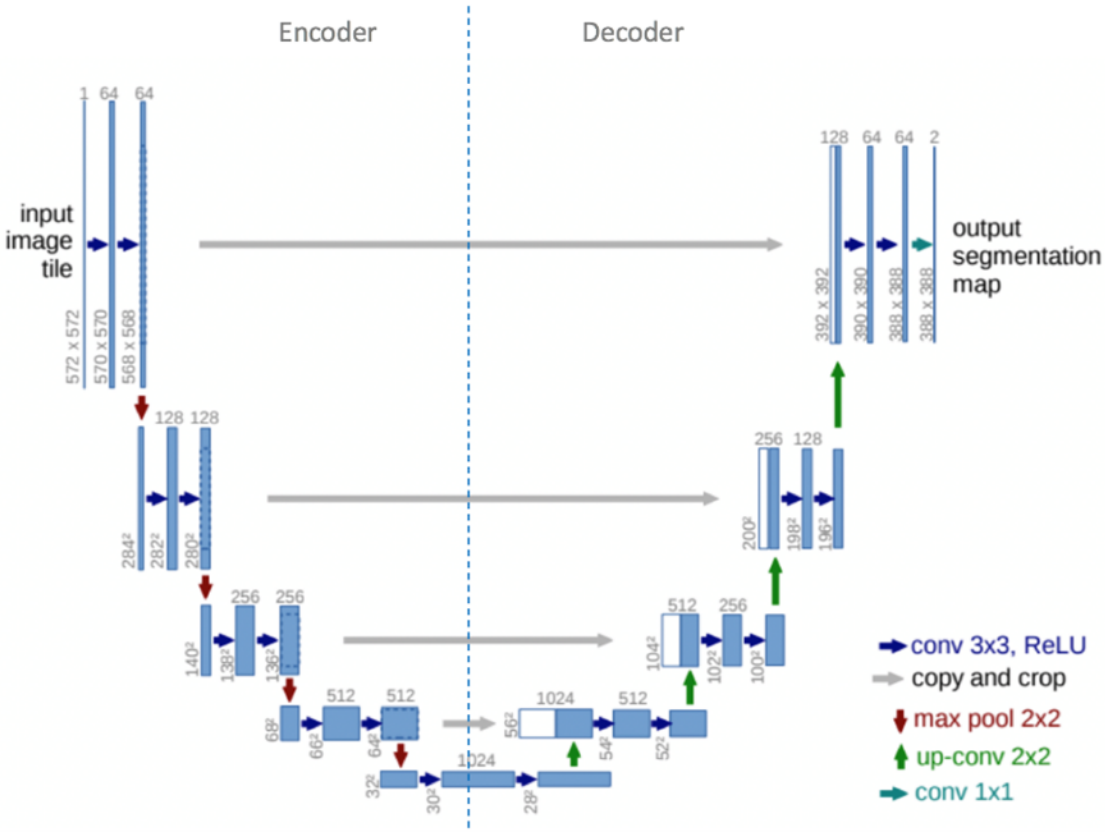

Figure 2.2 U-Net architecture.
Source: [6]

Another leading image segmentation technique is SegNet, by Badrinarayanan et al. In this approach, the convolution is carried out on the original image unlike other approaches where the images are resized for lower processing time since convolutional

networks require the full range of features from the original image. It consists of two parts of convolution / downscaling and deconvolution / upscaling. During the downscaling, the images are convoluted using a 2 x 2 max pooling and when upscaling, SoftMax is used to classify each pixel [19]. Figure 3 details the architecture.

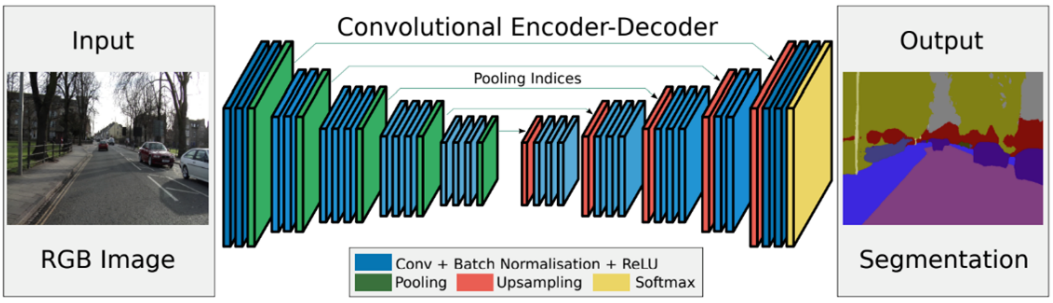

Figure 2.3 SegNet architecture.
Source: [19]

MaskLab by Chen et al, builds on R-CNN of 5x5 and 64 filters, based model and gathers regional logits, to separate different segments and classes of an image [4]. DeepLab by chen et al, uses DCNN based model, uses up sampling to extract dense features, implemented in VGG16 [5]. In it, instead of deconvolution, deepLab uses Atrous convolution of features to extract dense features, but without an increase in parameters. Figure 4 illustrates the atrous convolution used in DeepLab.

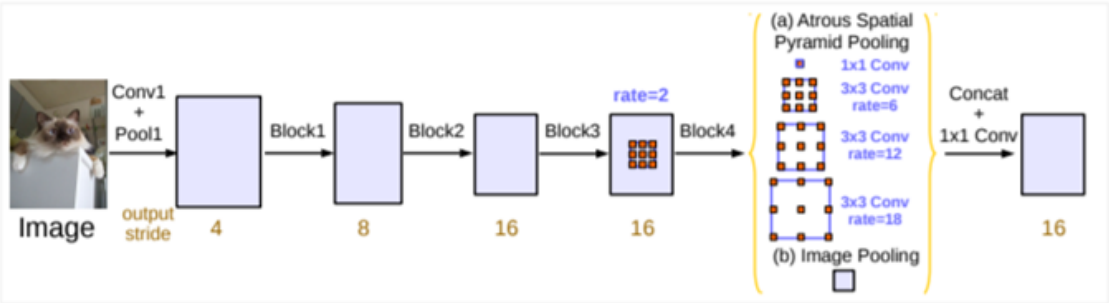

Figure 2.4 DeepLab architecture.
Source: [5]

CHAPTER 3

PROPOSED MODEL

In our model, we take a different approach to segmentation. We first get the segmentation mask of separate individual classes by individual decoder networks then combine and postprocess to produce the required output mask. The basic outline of the model is given in Figure 3.1.

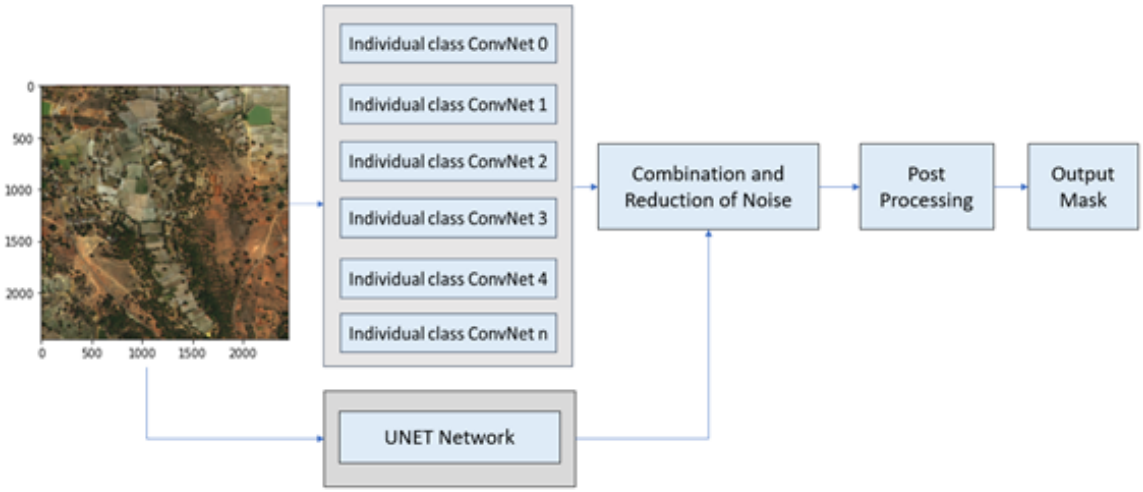

Figure 3.1 Proposed model.

We noticed that the precision of the convolutional net with only a single class is far above compared to metrics of models which segment all the classes together. We tried to use this property to our advantage, in creating a model that incorporates individual segmented masks into a whole, to produce a mask based on individual masks. Thus, we propose we have an array of individual convolutional nets for separate classes, each class feeds into the output from the combined class model, to augment and incorporate with higher results.

Ensemble have already been tried and tested methodology to improve metrics of different networks. There have been many models, that use the process of popular

voting, or other mechanisms to combine multiple models to create a model that has higher results. Some of the popular models in image segmentation can be seen in the paper by Noh et al [17], Choudhury et al [7] and Ali et al [2]. But in their models, they had used multiple networks to perfect the same mask. In our model we propose, we build the mask with the multiple separate neural networks. We then combine the multiple masks produced using a combination engine to perfect the mask, which gives us the resulting output segmented mask.

3.1 Detailed Model

In our model, as proposed in Figure 3.2, the input is split into images of size 512 x 512, then these input images are simultaneously passed through the multiple Individual networks and the Combined network, each of these networks output a mask, which is then processed for Blob detection. Additionally, regions that do not overlap are removed. Then by the order of the class with the largest region in the base mask, the regions are grown, using the previously detected blobs, to remove any regions which are not segmented. Finally, it is then post processed, to produce the output mask. The subsequent subsections detail the minutiae of the blocks in the model, to give a better picture.

3.1.1 Individual Networks

The Individual networks are a set of UNets, trained to produce masks that are single class, corresponding to 1 for the class that is being trained and 0 otherwise. The input to the networks is the same image input given to the entire model. Their output is the single class mask of 1s and 0s. Each of these UNets are based on the MobileNetV2 [21][25] architecture and their input and output images are of size 512 x 512.

They are trained by separating the specific class from the ground truth image, to produce the binary ground truth masks for each of the classes in the training dataset.

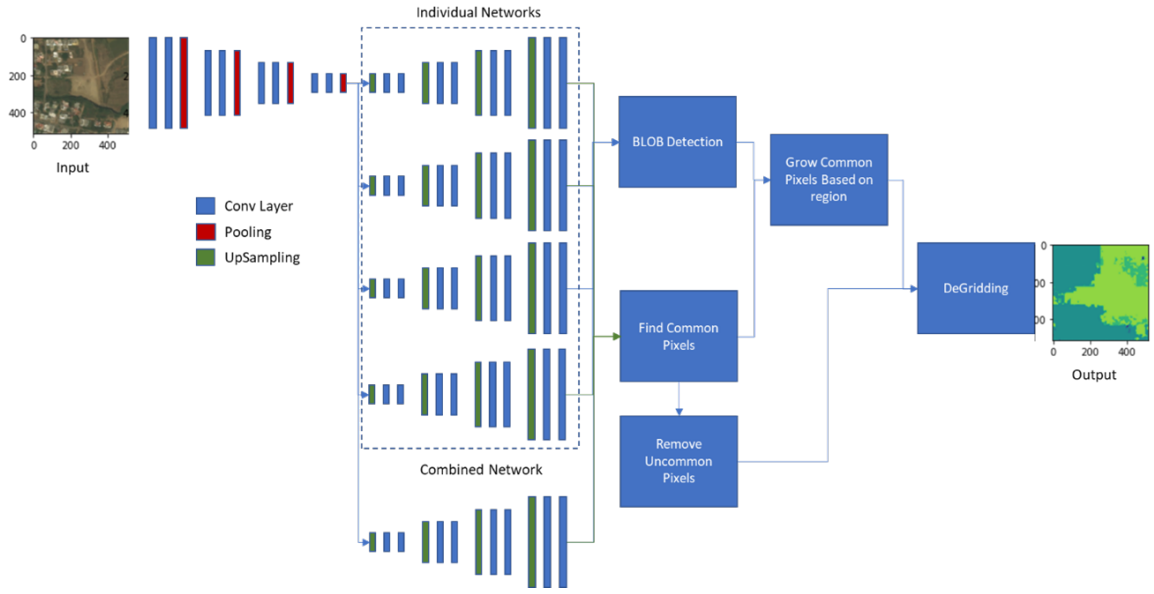

Figure 3.2 Detailed model.

By testing, we have concluded that binary cross entropy and binary masks yield higher Mean IoU than categorical, even considering the same accuracy. Additionally, we have removed the class 0, since if all the networks equal to 0 then it would indicate class 0. Empirically, for our problem, we have concluded that the classes 5 and 6 induce more errors and thus removed them from the array of networks.

3.1.2 Combined Networks

The Combined Network works as the base for our model, the image on which we do operations to produce better results. Similar to the Individual Networks, the Combined Network is also a UNet with a MobileNetV2 as the backbone. But instead of training on a binary mask, the network trains on the complete ground truth and thus, produces the complete prediction mask, containing all the classes of the dataset. The Combined mask, also produces an image of 512 x 512.

For our Combined and Individual Networks, we used UNet, because through implementation, this network produced better results compared to other similar

models, the results of these, along with other state-of-the-art model results will be illustrated in the upcoming sections.

3.1.3 Blob Detection

In image processing, Blobs are defined as a region of connected pixels, with the same color or intensity. In our case, each Blob is a group of pixels that are 8-connected, meaning the pixels in the region can be either directly or diagonally connected to each other. If a region is not connected to another through any of its pixels, through neither directly nor diagonally, then it is considered a separate Blob, from the existing one.

The Blobs are detected based on binary masks for the outputs from the Individual Networks. Each Blob is labeled, and a map of the Blobs and their labels are passed forward to the next block.

In the case of the Combined Network, a separate set of grayscale image, for each of the classes are prepared, to denote the corresponding classes. This is then similarly processed for Blob Detection, producing the set of label images. This set of images is then passed to the next block.

3.1.4 Remove Uncommon Pixels and Find Common Pixels

As the title of the block indicates, in this procedure, we remove pixels from the base image that are not overlapping with their corresponding class image masks. We first, create a grayscale mask for each of the classes in the base image, using which, we sequentially compare with the masks of the class images produced from the Individual Networks. If a pixel is not set to 1 in both the grayscale images, the pixel is set to 0.

With this we also have isolated the common pixels from the images, producing an image with only the common regions from both the masks. But this also produces

a lot of pixels with 0 in the base mask. We will address and rectify this issue in the subsequent block.

3.1.5 Growing Regions

In this block, we address the issue of empty spots in the base images based on the election process in the previous block. The input to this block, is the Blobs detected previously and the base image processed and without any uncommon pixels.

We first find the order of class pixels to grow by finding the number of pixels congruent with a class, then finding the class that has the highest number of pixels in the image. Thus, ordering the classes on this criterion, and the following process is done on each of the classes based on the same ordering.

For each pixel in the base image, we select the Blob it corresponds to in both the Individual mask and the Combined mask, therefore producing the possible pixel growth location for that pixel. We then choose pixels that have the value 0 in the base mask and were not originally classified as class 0. Then we fill the empty location with the value of the class if the pixel is located in one of the possible growth locations. We repeat this process for each of the different classes to produce the final augmented mask.

3.1.6 Post Processing

UNets are famously beset by the issue of gridding artifacts. To get rid of these artifacts that occur after we have acquired our output image, we deGrid, by taking mean of segmented pixels across an 8 x 8 block. Which is iterated across the entire image to produce solid regions of pixels.

This entire process produces a 512 x 512 output image, which we combine with the other cropped parts of the whole image to get the final output image, which correlates to the original input image.

CHAPTER 4

EVALUATION AND RESULTS

4.1 Dataset

Image segmentation datasets are difficult to create because of the sheer manpower required to label and classify each segment. Many datasets are updated over years and require a lot work hours. Each segmentation mask, although not be perfect should be able to classify the semantic components correctly. And each of these masks must be done by hand. There are various semi-automatic tools that take user input only for correcting wrongly segmented masks. A popular tool is the GUIDE tool in MATLAB.

The DeepGlobe Dataset [8] is a highly detailed satellite imagery dataset, which has semantic segmentation from building and road extraction and terrain classification. It contains 803 images of high-resolution satellite images, throughout the globe, each with a semantic segmentation mask that details the type of terrain. There are 6 types of terrain – urban land, agricultural, rangeland, forest lang, water, barren land and unknown. This dataset is a benchmark for remote sensing models and segmentation tasks.

4.2 Model Parameters

4.2.1 Loss

Cross entropy is the most popular loss used in semantic segmentation. It is defined as given in the formula below. Where y_{truth} is the ground truth classification each pixel and y_{pred} is the predicted classification of the pixel, predicted by the model. Which is summed individually to produce the loss function.

$$H(p, q) = -\sum y_{truth} log(y_{pred}) \tag{4.1}$$

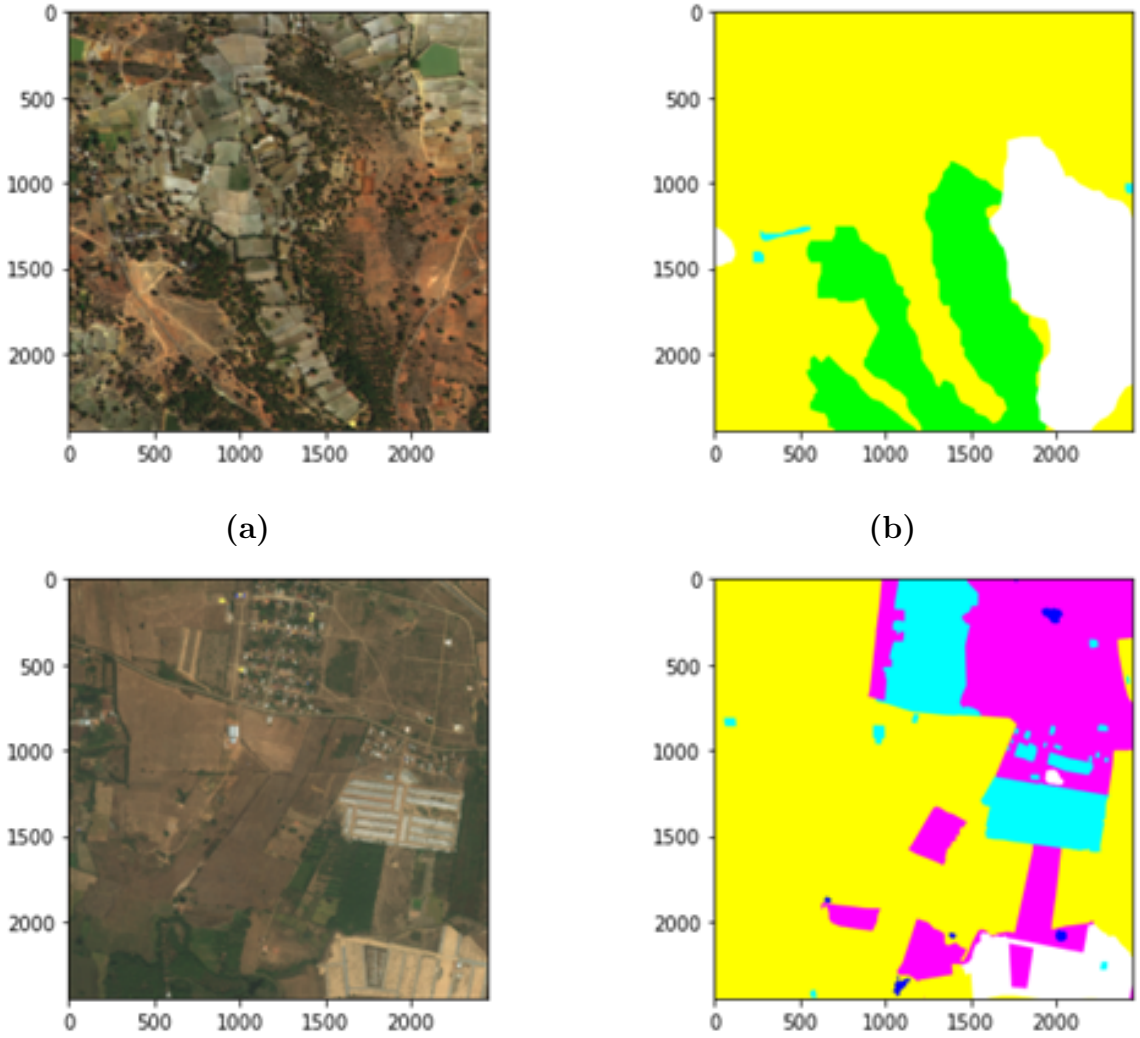

Figure 4.1 DeepGlobe dataset. (a) and (c) are the original images, (c) and (d) are the ground-truth segmentation masks.

4.2.2 Optimizer

We use the Adam Optimizer to train our model. Unlike traditional Stochastic Gradient optimizer, Adam works to change learning rate dynamically, throughout the training to get the best results. In addition to this, each weight in the network are seperately adapted, to a learning rate specific to the weight. Adam optimizer combines RMSProp and AdaGrad algorithms to give a better performance overall in larger datasets[13].

4.3 Metrics

Accuracy is a straightforward metric that is dictated as the number of correctly classified pixels to the total number of pixels averaged across all the pixels across all the images in the test dataset. Although it is a very easy metric, it is not the most reliable since if a class covers 90 percent of the image then, the accuracy will be 90 percent, which is not reliable.

Mean Intersection of the Union (mIoU) is another commonly used metric, that can measure the quality of the segmentation mask with respect to the ground truth. It is as its name is given mean across the classes of intersection of true mask and predicted mask divided by their union. If the value is above a threshold (usually 0.5), it is said to be true positive otherwise it is false negative. With this in mind the formula is given as below.

$$IoU = \frac{TruePositive}{TruePositive + FalsePositive + FalseNegative} \tag{4.2}$$

4.4 Results

The following results and experiments were conducted on a machine with a RTX 2070 mobile GPU, which has 2,304 CUDA cores, i7–9750H processor and 16GB of RAM. The images from the DeepGlobe dataset were separated into two sets of validation

sets and one training dataset, with the validation datasets having 20% of the original dataset. To accommodate the model in the GPU, the images of resolution 2048 x 2048 were cropped to 16 images of 512 x 512 and were then fed to the model.

4.4.1 Induvidual Networks

The Individual models are plagued with the problem of overfitting, hence, we had to introduce noise to the training dataset. We augment the training data by randomly – flipping, adding Gaussian noise and mixing images. We also augment the UNet model that we had used for binary classification, to a smaller and lightweight model based on MobileNetV2. With this we were able to get the testing results similar to the training results. The training results are shown in Table 4.1 and testing in Table 4.2.

Table 4.1 Augmented Segmentation Testing Result By Class

	Class 1	***Class 2***	***Class 3***	***Class 4***	***Class 5***	***Class 6***
Mean IoU	79.7	70.23	75.05	65.29	64.77	64.27
Accuracy	92.93	85.72	87.68	83.58	85.27	78.46

Table 4.2 Augmented Segmentation Training Result By Class

	Class 1	***Class 2***	***Class 3***	***Class 4***	***Class 5***	***Class 6***
Mean IoU	86.03	67.29	78.20	58.03	66.98	64.27
Accuracy	95.23	84.42	88.97	79.26	86.02	79.20

4.4.2 Final Results and Comparison

For performance comparisons, we use the existing state-of-the-art models: UNet, SegNet, and Deep Lab. The ensemble model consists of popular voting of 3 networks of SegNet, DeepLab, based on Xception and UNet, based on MobileNetV2, for reference purposes. The ensemble model images were considered from results of the three networks, when there was not a clear majority, the result of UNet was taken. The Deep Aggregation model from Kuo et al, uses the DeepLab network as base and adds extra networks to make the model [20]. The Stacked UNet model, from Gosh et al, stacks a series of UNets, to improve segmentation results [21].

Table 4.3., details the results of other models and papers on the DeepGlobe dataset with a crop of 512. With models both open sourced and implemented on our own. As we can see the results of our model is the highest when compared to similar models.

Table 4.3 Segmentation Result Of Similar Models

Model	***Mean IoU***
UNet	48.1
SegNet	47.5
DeepLab	39.1
FCN	45.8
Ensemble	50.2
DeepAggregation Net	52.7
SUNET	50.1
Ours	53.3

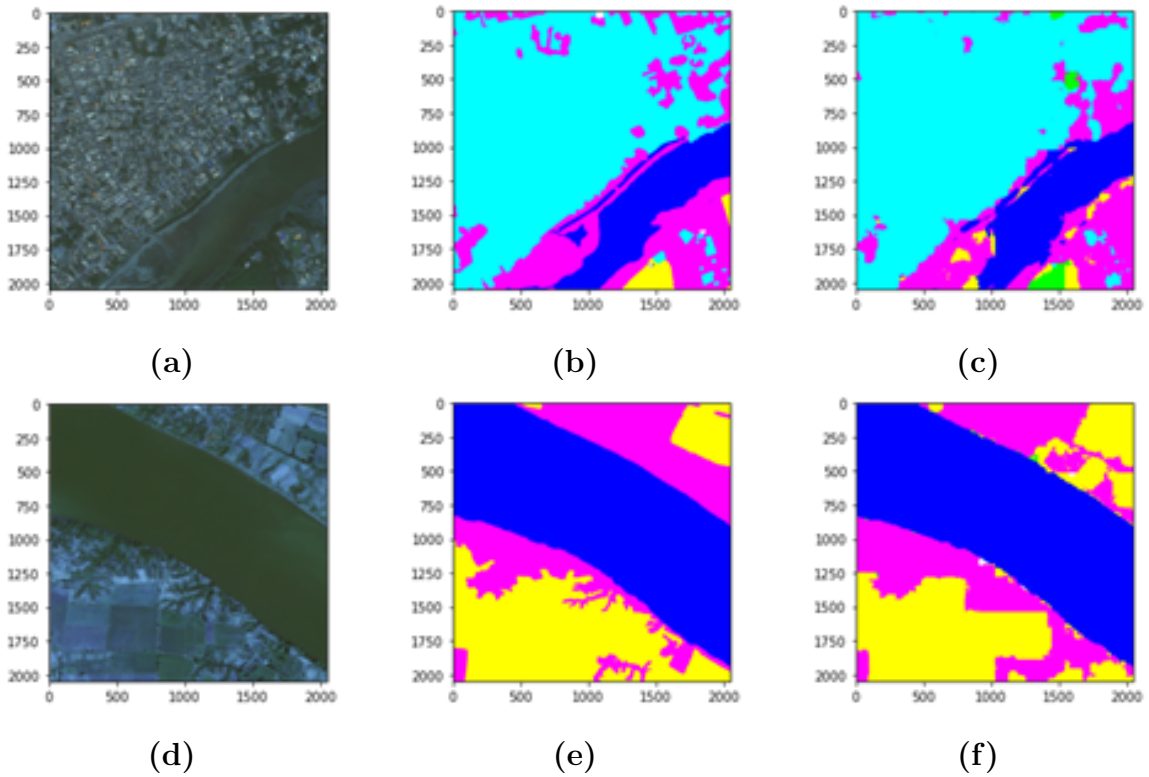

Figure 4.2 Segmentation results. (a) and (d) are the original images, (b) and (e) are the ground-truth images, and (c) and (f) are the segmented images.

Figure 3 shows the segmentation results.

CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1 Conclusion

In this paper, we have proposed a novel method and a unique viewpoint on multiclass segmentation, where instead of looking at the problem together and parsing through a neural network, worked and implemented a network, consisting of separate single class networks working to augment the segmentation results and proven that this method produces higher metrics than when similar models. This can be considered a method to combine the advantages of binary segmentation techniques, and also subsequently improve segmentation results in future models.

5.2 Future Work

This new method provides new ways to build and improve image segmentation models and problems. We await to try and see the results of our network in different datasets and computer vision tasks.

BIBLIOGRAPHY

[1] Nadeem Akhtar, Nishi Agarwal, and Armita Burjwal. K-mean algorithm for image segmentation using neutrosophy. *2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2014.

[2] Redha Ali, Russell C. Hardie, Barath Narayanan Narayanan, and Supun De Silva. Deep learning ensemble methods for skin lesion analysis towards melanoma detection. *2019 IEEE National Aerospace and Electronics Conference (NAECON)*, 2019.

[3] A. Antonacopoulos. Segmentation and classification of document images. *IEE Colloquium on Document Image Processing and Multimedia Environments*, 1995.

[4] Y. Ben Fadhel, S. Ktata, and T. Kraiem. Cardiac scintigraphic images segmentation techniques. *2016 2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, 2016.

[5] Liang-Chieh Chen, Alexander Hermans, George Papandreou, Florian Schroff, Peng Wang, and Hartwig Adam. Masklab: Instance segmentation by refining object detection with semantic and direction features. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018.

[6] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4):834–848, 2018.

[7] Ahana Roy Choudhury, Biswas Parajuli, and Piyush Kumar. Quadroad: An ensemble of cnns for road segmentation. *Procedia Computer Science*, 176:138–147, 2020.

[8] Ilke Demir, Krzysztof Koperski, David Lindenbaum, Guan Pang, Jing Huang, Saikat Basu, Forest Hughes, Devis Tuia, and Ramesh Raskar. Deepglobe 2018: A challenge to parse the earth through satellite images. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2018.

[9] Arthita Ghosh, Max Ehrlich, Sohil Shah, Larry Davis, and Rama Chellappa. Stacked u-nets for ground material segmentation in remote sensing imagery. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2018.

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[11] Prashant Devidas Ingle and Parminder Kaur. Adaptive thresholding to robust image binarization for degraded document images. *2017 1st International Conference on Intelligent Systems and Information Management (ICISIM)*, 2017.

[12] Rehna Kalam and K. Manikandan. Enhancing k-means algorithm for image segmentation. *2011 International Conference on Process Automation, Control and Computing*, 2011.

[13] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *International Conference on Learning Representations*, 12 2014.

[14] Tzu-Sheng Kuo, Keng-Sen Tseng, Jia-Wei Yan, Yen-Cheng Liu, and Yu-Chiang Frank Wang. Deep aggregation net for land cover classification. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2018.

[15] Ying Liu and Xiaofang Zhou. Automatic texture segmentation for texture-based image retrieval. *10th International Multimedia Modelling Conference, 2004. Proceedings.*

[16] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[17] Hyeonwoo Noh, Seunghoon Hong, and Bohyung Han. Learning deconvolution network for semantic segmentation. *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015.

[18] Wenqi Ren, Jinshan Pan, Xiaochun Cao, and Ming-Hsuan Yang. Video deblurring via semantic segmentation and pixel-wise non-linear kernel. *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017.

[19] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science*, page 234–241, 2015.

[20] Mahdi Sabri and Paul Fieguth. A new gabor filter based kernel for texture classification with svm. *Lecture Notes in Computer Science*, page 314–322, 2004.

[21] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018.

[22] Lianlei Shan and Weiqiang Wang. Densenet-based land cover classification network with deep fusion. *IEEE Geoscience and Remote Sensing Letters*, page 1–5, 2021.

[23] Jianbo Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):888–905, 2000.

[24] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv 1409.1556*, 2014.

[25] Debjyoti Sinha and Mohamed El-Sharkawy. Thin mobilenet: An enhanced mobilenet architecture. *2019 IEEE 10th Annual Ubiquitous Computing, Electronics amp; Mobile Communication Conference (UEMCON)*, 2019.

[26] Prachi Surlakar, Sufola Araujo, and K. Meenakshi Sundaram. Comparative analysis of k-means and k-nearest neighbor image segmentation techniques. *2016 IEEE 6th International Conference on Advanced Computing (IACC)*, 2016.

[27] Yu-Ho Tseng and Shau-Shiun Jan. Combination of computer vision detection and segmentation for autonomous driving. *2018 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, 2018.

[28] Yading Yuan, Ming Chao, and Yeh-Chi Lo. Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance. *IEEE Transactions on Medical Imaging*, 36(9):1876–1886, 2017.

[29] Y. Zhang, M. Brady, and S. Smith. Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, 20(1):45–57, 2001.