International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-11 Issue-6, August 2022

CNN based System for Road, Vehicle Detection and Segmentation from Aerial Images

A. Jyothi, P. Chandra Sekhar

Abstract*: *Image segmentation is crucial for computer vision. Visual segmentation simplifies image analysis. Detecting and dividing highways is a major difficulty in aerial traffic monitoring, autonomous driving, and border surveillance. This is tough. Image segmentation traditional algorithms are unsuccessful, according to the literature. Segmentation of the semantic field divides an image into semantically relevant components and assigns each to a class. Deep convolutional neural networks accurately segregate semantics. In deep learning, a convolutional neural network utilizes an input image to rate the relevance of various things. This work used convolutional neural networks to recognize and segment roads in aerial photos. SegNet and DeepLabv3+ use Vgg16 and ResNet-18 pre-trained models for road recognition and segmentation from aerial photos. SegNet based on Vgg16 produced high accuracy, whereas DeepLabv3+ based on Resnet-18 was efficient in terms of accuracy and training time. The suggested system designs use MATLAB. Learning and operational phases are included in the algorithm for image segmentation. In MATLAB, convolutional neural network analyses tagged aerial photographs. This work trains a convolutional neural network to accurately extract road features from aerial images. This system identifies roads and automobiles quickly for traffic monitoring.

Keywords*: *ANN, CNN, Segmentation, Road Detection, Vehicle Detection, ResNet, SegNet.

I. INTRODUCTION

Extraction of accurate information from aerial photos may be used for catastrophe monitoring (earthquakes, floods, and fires), border surveillance, autonomous driving, and precision agricultural crop monitoring. Airborne ground surface monitoring requires road identification and segmentation. Image processing techniques are inefficient and need human participation. Image segmentation uses deep learning for computer vision which often uses convolutional neural networks. Road scene applications must display look and form. This work met this need. Deep convolutional neural networks must be used for semantic segmentation in traffic monitoring and self-driving automobiles. To extract proper information from a picture, it may be divided into segments (pixel sets) depending on particular qualities. This method clarifies and simplifies the picture. Human-engineered feature extraction methods affect the quality and reliability of classic picture segmentation models. CNNs are used to analyze visual images in deep learning. Convolutional neural networks, for example, take an image as input, add weights and biases, and then discriminate. ConvNet needs less preprocessing than other classification methods. With adequate training, ConvNets can design their own filters and attributes. As the model's layers increase, it can fit increasingly complicated functions. Deep convolutional neural networks identified and segmented highways in aerial photos. [1] Implemented fully convolutional network by adding deconvolution layers to VGG16. 5 convolution layers were used and the batch size is 2. Accuracy is good when compared to other traditional algorithms. [2] Used tradition sliding box algorithm - CNN with 4 convolution and pooling layers along with 1 fully connected, softmax layer. Accuracy is good but slightly less when compared to SegNet. [3] Designed SegNet for image segmentation and tested using Camvid dataset with 12 as batch size and 33 epochs. Accuracy is good. [4] Implemented DeepLabV3+, an advanced image semantic segmentation model with ResNet-50 as an encoder. Results showed that the method significantly improves the accuracy of water extraction. After reviewing the various deep convolutional neural network models, VGG16 and ResNet-18 pre-trained models showed high test accuracy and less error rate. Hence, in this work, SegNet using VGG16 and DeepLabv3+ using ResNet-18 deep convolutional neural network models are designed for implementing road detection and segmentation from aerial images.

II. METHODOLOGY

A. Image Segmentation

Image segmentation is a computer vision approach that splits digital images. Segmentation reduces a picture's complexity and/or representation to make it more intelligible. Image segmentation identifies objects and picture borders (such as lines, curves, and so on). Picture segmentation involves labelling individual pixels such that only photos with matching labels are joined. Create a collection of image segments or extract the image's outlines using image segmentation. All pixels in the same region have the same color, intensity, or texture. Neighborhoods are quite distinct. Interpolation techniques like Marching cubes, segmented contours may be used to create 3D reconstructions from a stack of photographs.

Manuscript received on 29 July 2022 | Revised Manuscript received on 04 August 2022 | Manuscript Accepted on 15 August 2022 | Manuscript published on 30 August 2022.

***** Correspondence Author

A. Jyothi*, Faculty, Department of Electronics & Communication Engineering, Government Institute of Electronics, Hyderabad (Telangana), India. Email: jyothisagar.a@gmail.com

Prof. P. Chandra Sekhar, Professor, Department of Electronics & Communication Engineering, University College of Engineering, Osmania University, Hyderabad (Telangana), India. Email: sekharp@osmania.ac.in

B. Image segmentation techniques.

Two sorts of segmentation techniques exist. Semantic segmentation divides pixels into meaningful clusters. They're semantically interpretable andmatch real world categories. Instance Segmentation identifies every object in an image. This approach doesn't categorize every pixel like semantic segmentation

C. Semantic Segmentation using Deep Learning

Classifying every pixel in an image, results in an image that is segmented according to its class. As an example, semantic segmentation may be used for autonomous driving and cancer cell segmentation. Semantic segmentation using deep learning is seen in Fig. 1. Encoder network is followed by decoder network in generic semantic segmentation architecture. The encoder is a pre-trained classification network, such as VGG/ResNet. Decoding involves converting the encoder's lower-resolution discriminative characteristics into higher-resolution pixel space. As well as the ability to discriminate pixels, semantic segmentation also requires a way to map the encoder's learned discriminative features into pixel space. As part of the decoding procedure, several techniques use a variety of mechanisms.

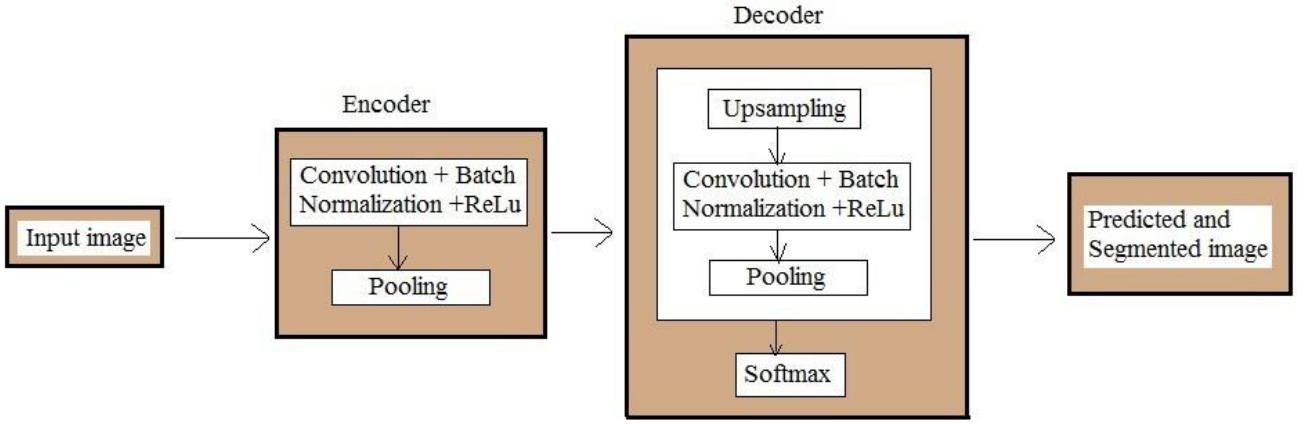


Fig.1: Block diagram of Semantic Segmentation using Deep Learning

D. SegNet

SegNet is an efficient pixel-wise semantic segmentation framework. To completely comprehend the scene, road scene apps must present the appearance. Road and sidewalk are two examples of the spatial relationship between different classes of objects. Most pixels in typical road scenes fall into big classes like road and building and hence the network needs to create seamless segmentation.

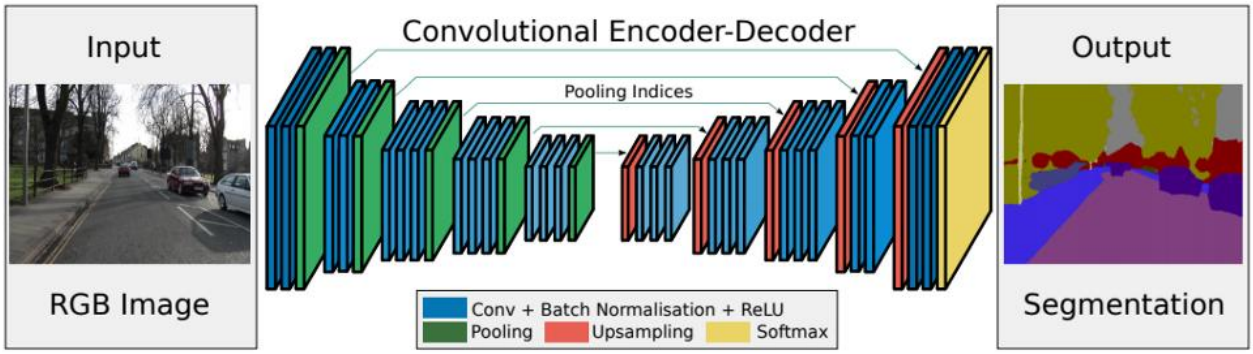


Fig.2: SegNet Encoder-Decoder Architecture

As in Fig.2, SegNet has an encoder, decoder, and pixel-wise classification layer. The encoder network contains 13 convolutional layers that match the VGG16 network's initial 13 layers. The encoder's deepest output is a high-resolution feature map, not connected layers. SegNet contains fewer parameters than other modern models. Each encoding layer has one decoder layer. Multi-class soft-max classifiers employ decoder output to calculate pixel class probabilities. The Conv layer allows 224 224 RGB images. The picture is processed using convolutional filters with a narrow receptive field, 3x3 (the smallest size capable of capturing left/right, up/down, and centre). In one design, 11 convolution filters achieve linear input channel modification (followed by non-linearity).

After convolution, the spatial resolution of layers is maintained using 1-pixel padding. Five max-pooling levels are then added for spatial pooling after several conventional layers have been completed. In 2x2 frames, Stride 2 maximizes pooling. Three Fully-Connected (FC) layers follow the convolutional layers: the first two have 4096 channels each, while the third has 1000 channels and performs a 1000-way classification of ILSVRC data (one for each class). The final touch is a soft-max fabric. Every network has linked layers. Every buried layer is non-linear (ReLU).

Retrieval Number: 100.1/ijeat.F37640811622
DOI: 10.35940/ijeat.F3764.0811622
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

Except for one network, none employ Local Response Normalization, which doesn't enhance ILSVRC performance but increases memory and calculation time.

E. ResNet

A residual neural network (ResNet) is a type of ANN that builds on constructs from pyramidal cells in the cerebral cortex. Residual neural networks utilize skip connections, or shortcuts to jump over a few layers. Typical ResNet models are implemented with double or triple-layer skips which contain nonlinearities (ReLU) and batch normalization in between. An additional weight matrix called HighwayNets is used to learn the skip weights.

III. EXPERIMENTAL RESULTS

A. Road detection and segmentation from aerial images using SegNet(VGG16):

Each picture rotated 90, 180, and 270 degrees clockwise to facilitate efficient learning from aerial photographs taken from freeways, woodlands, and deserts at various elevations. The Matlab - Image Labeler programme used to conduct the pixel labelling. There are 140 photos in this collection. Matlab programming used to develop the SegNet network architecture based on the VGG16 convolutional neural network. 80% of the photos in the dataset are used to train the SegNet network. The remainders of pictures are for testing purposes. Stochastic gradient descent with momentum used to train the network with options mentioned in Table-I.

Table-I: Training options

Option Name	Values
Momentum	0.9
InitialLearnRate	0.01
LearnRateScheduleSettings	['1x1 Struct']
L2Regularization	1.00E-04
GradientThresholdMethod	'l2norm'
GradientThreshold	Inf
MaxEpochs	100
MiniBatchSize	16
Verbose	1
VerboseFrequency	50
ValidationFrequency	50
ValidationPatience	5
Shuffle	once
CheckpointPath	" "
ExecutionEnvironment	'auto'
WorkerLoad	[]
OutputFcn	[]
Plots	'Training-progress'
SequenceLength	'longest'
SequencePaddingValue	0

It is seen in Table-II that around 80% of photos from a dataset used for training, while the remaining 20% used for testing purposes. Fig.3 shows the pixels breakup for road, background from the given dataset. SegNet layers plot shown in Fig.4 and Fig.5 shows training progress plot.

Table-II: Breakup of image pixels for road and background

Class name	Pixel Count	Image Pixel Count
'road'	1.9381e+06	5.6197e+06
'Background'	3.5894e+06	5.6197e+06

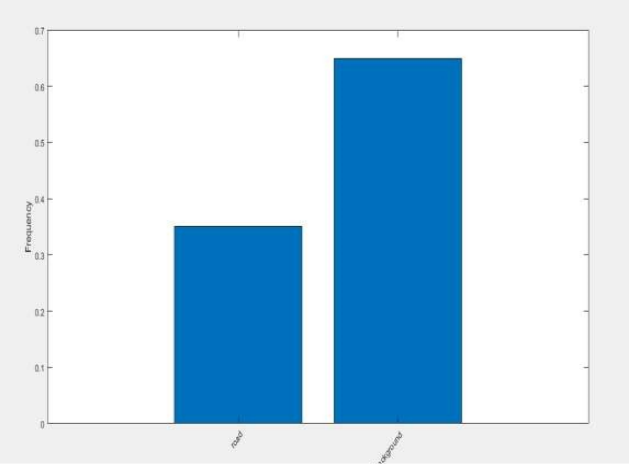


Fig.3: Graph showing the breakup of image pixels for road and background

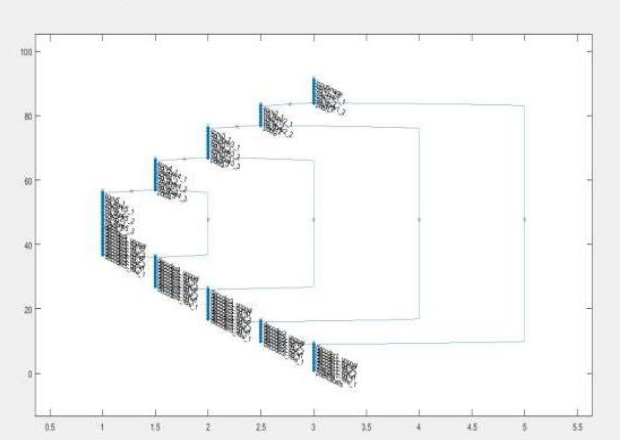


Fig.4: Plot of Seg Net layers using VGG16

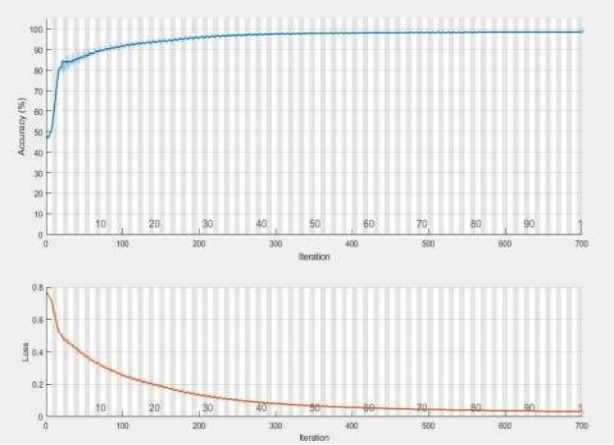


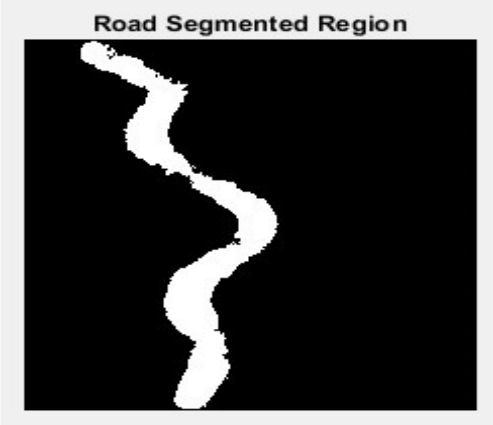
Fig.5: Plot of training progress

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

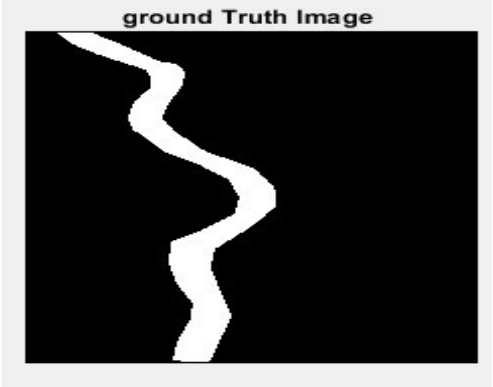
Using photos from a test data set, we put our newly-trained network to the test. The results, shown in the Fig.6, are for road recognition and segmentation from aerial photographs.

a)

b)



c)



d)

Fig.6: Output images for road detection and segmentation from aerial images.

Table-III: Dataset Metrics

GlobalAccuracy	MeanAccuracy	MeanIoU	WeightedIoU	MeanBFScore
0.95416	0.94155	0.89075	0.8978	0.64748

Table-IV: Class metrics

Class	Accuracy	IoU	MeanBFScore
road	0.89763	0.86621	0.56395
Background	0.98547	0.91529	0.731

B. Road detection and segmentation from aerial-images using DeepLabv3+(ResNet-18) :

Created a collection of aerial photos of a road taken from various heights, and used Matlab Image Labeler tool to identify the pixels as either the road or a backdrop. The ResNet-18 convolutional neural network architecture for DeepLabv3+ was built using Matlab programming and the corresponding toolboxes. To train the DeepLabv3+ network, 80 percent of the photos in the dataset were used. For validation and testing, 10% of the photos were used. Stochastic gradient descent with momentum used to train the network with options mentioned in Table-V.

Table-V: Training options

Option Name	Values
Momentum	0.9
InitialLearnRate	1.00E-03
LearnRateSchedule	'piecewise'
LearnRateDropFactor	0.3
LearnRateDropPeriod	10
L2Regularization	0.005
GradientThresholdMethod	'l2norm'
GradientThreshold	Inf
MaxEpochs	50
MiniBatchSize	8
Verbose	1
VerboseFrequency	2
ValidationData	pximdsVal
ValidationFrequency	50
ValidationPatience	4
Shuffle	every-epoch'
CheckpointPath	tempdir
ExecutionEnvironment	'auto'
WorkerLoad	[]
OutputFcn	[]
Plots	'training-progress'
SequenceLength	'longest'
SequencePaddingValue	0
SequencePaddingDirection	right'
DispatchInBackground	0
ResetInputNormalization	1

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Table-VI: Breakup of image pixels

Class name	PixelCount	ImagePixelCount
'road'	2.4307e+06	7.0246e+06
'Background'	4.4789e+06	7.0246e+06

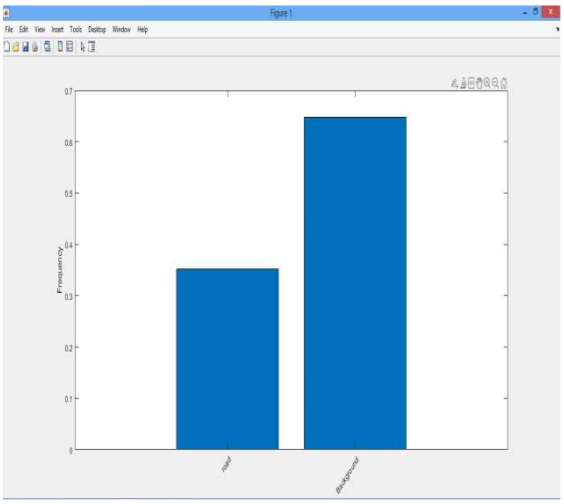


Fig.7: Graph showing the breakup of image pixels for road and background.

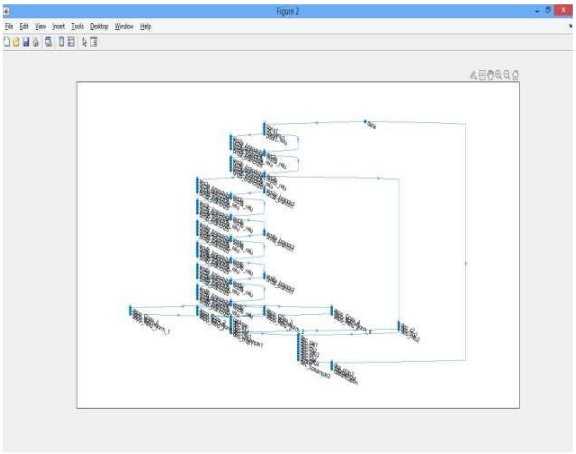


Fig.8: Plot of DeepLabv3+ layers using ResNet-18

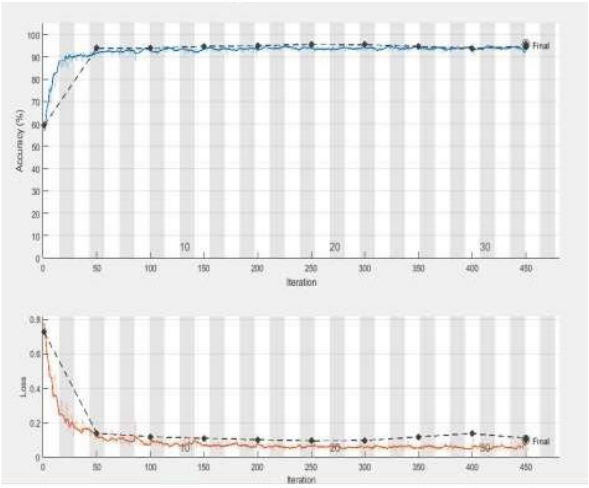
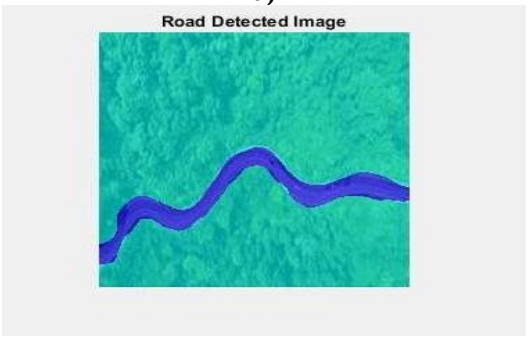


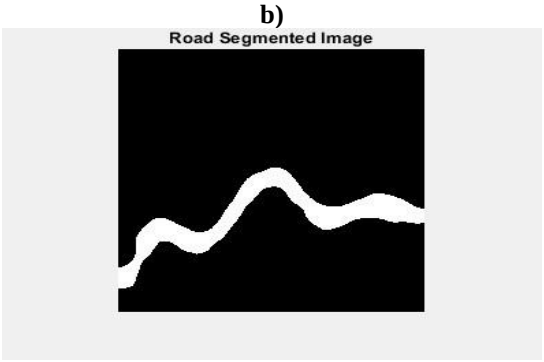
Fig.9: Plot of training progress with elapsed training time of 68 min. 35 sec.

Using photos from a test dataset, we put our newly-trained network to the test. Fig.10 shows the results for road recognition and segmentation from aerial photographs.

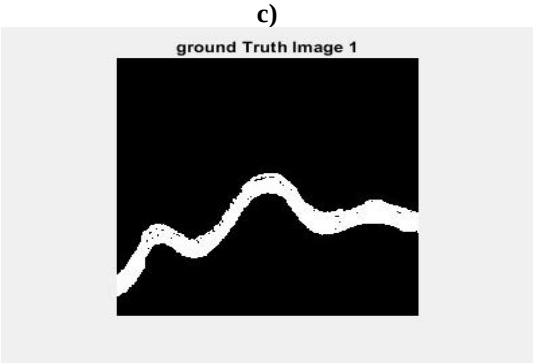
a)



b)



c)



d)

Fig.10: Result photos for road recognition and segmentation from aerial photographs.

Table-VII: Dataset Metrics

Class	Accuracy	IoU	MeanBFScore
road	0.95186	0.88421	0.72627
Background	0.9609	0.93783	0.83662

Table-VIII: Class Metrics

GlobalAccuracy	MeanAccuracy	MeanIoU	WeightedIoU	MeanBFScore
0.95784	0.95638	0.91102	0.91969	0.78144

C. Performance evaluation of SegNet(VGG16) and DeepLabv3+(ResNe t-18):

SegNet (VGG16) and DeepLabv3+ (ResNet-18) evaluated the photographs. DeepLabv3 ResNet-18 accuracy was better than SegNet's VGG16 despite its quicker learning curve and better test outcomes. DeepLabv3+ (ResNet-18) avoids over fitting using a validation criterion.

Table-IX: Comparison of the Metrics from the Trained Networks

Metric	SegNet(VGG16)	SegNet(VGG16)	DeepLabv3+(ResNet-18)
Image dataset size	140	140	140
Network training time	609 min 42 sec	1044 min 34 sec	68 min. 35 sec.
Number of epochs	50	100	50 (training ended at 33rd epoch as it met validation accuracy)
Minimum batch size	8	16	8
GlobalAccuracy	0.84001	0.95416	0.95784
MeanAccuracy	0.79327	0.94155	0.95638
MeanIoU	0.6781	0.89075	0.91102
WeightedIoU	0.712	0.8978	0.91969
MeanBFScore	0.40848	0.64748	0.78144

IV. APPLICATIONS

Traffic monitoring may benefit from road detection and segmentation. Below is the output of a dataset created for segmenting automobiles and roads. This application can be coupled with cameras installed at the traffic junctions, which can detect and identify vehicles.

Fig.11: Target Image

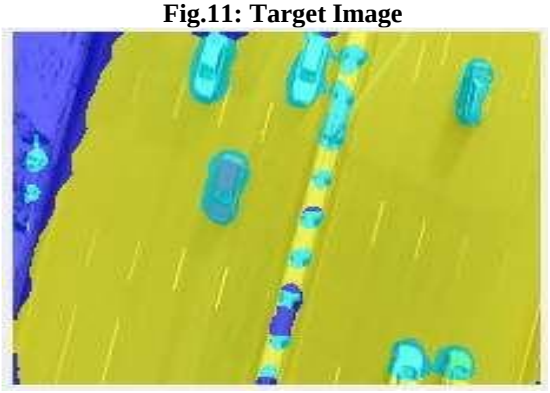


Fig.12: Road and vehicle detected image

V. CONCLUSION

Aerial ground surface monitoring has several applications, including border surveillance, traffic detection, and autonomous driving, where roadway identification and segmentation are critical challenges to overcome. As part of the project work, DeepLabv3+ and ResNet-18 deep convolutional neural networks were built to implement road recognition and segmentation from aerial photos. A deep convolutional neural network was successfully trained using photos taken from various heights. Aerial photos could be used to train the network to recognize and segment roads accurately. The most recent DeepLabv3+(ResNet-18) convolutional neural network achieved global accuracy of 96 percent with 33 epochs and 68 minutes of network training time, whereas SegNet(VGG16) convolutional neural network achieved global accuracy of 95 percent with 100 epochs and 1044 minutes of network training time. As a result, it is preferable to use the most recent DeepLabv3+ (ResNet-18) deep convolutional neural network for road recognition and segmentation from aerial photos because it has higher accuracy and requires less network training time than SegNet.

REFERENCES

1. Baiyan Li; Xiang Zhang ; Hui Song, "A New Method for Road Element Extraction Based on Fully Convolutional Network" 2019 IEEE International Conference on Computational Electromagnetics (ICCEM), published in IEEE on 29, July 2019.
2. Loretta Ichim and Dan Popescu, "Road Detection and Segmentation from Aerial Images using a CNN based System" in 2018 41st International Conference on Telecommunications and Signal Processing (TSP), published in IEEE, Aug, 2018. [CrossRef]
3. Vijay Badrinarayanan ; Alex Kendall ; Roberto Cipolla, SegNet: "A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation"–IEEE - Journal Article - Dec. 1 2017, Volume: 39, Issue: 12. [CrossRef]
4. Ziyao Li ; Rui Wang ; Wen Zhang ; Fengmin Hu ; Lingkui Meng -"Multiscale Features Supported DeepLabv3+Optimization Scheme for Accurate Water Semantic Segmentation" -25 October 2019, IEEE Access,Volume:7. [CrossRef]

AUTHORS PROFILE

A. Jyothi, received B.Tech degree from Jawaharlal Nehru Technological University Hyderabad, M.Tech degree from Osmania University in 2009 and 2020 respectively. Prior to teaching, has 6 years of industrial experience in implementation of Oracle ERP systems. Working as Associate Lecturer in Electronics and Communication Engineering, Government Institute of Electronics, Hyderabad from 2018. Research interests include Artificial intelligence, Machine learning, Image processing.

Prof. Chandra Sekhar Paidimarry, received BE degree from Nagpur University, M.Tech degree from JNTU Hyderabad and PhD from Osmania University (OU) in 1991, 1999 and 2009 respectively. He had been awarded with Post Doctoral Fellowship by Shizuoka University, Japan for one year. Prior to teaching, has eight years of industrial experience of design and development of Embedded Systems. Working as a Professor in the Department of Electronics and Communication Engineering, University College of Engineering, OU, Hyderabad from 2001. Served as the Head of Department and Chairman BOS in ECE Dept., Osmania University. Have more than 50 research publications, delivered more than 15 invited talks and guest lecturers in various conference and events. Research interests include Development of high performance Computational Electro-magnetic and efficient FPGA based signal processing algorithms and Design Automation.

Retrieval Number: 100.1/ijeat.F37640811622
DOI: 10.35940/ijeat.F3764.0811622
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)