Using layer-Wise Training for Road Segmentation in Autonomous Cars

Yashavardhan SG
Dept of ISE (student)
Alva's insitute of Engineeing And Technology
Mijar,Moodbhiri-574225
yashavardhansg@gmail.com

Pradeep Nayak
Assistant Professor
Alva's insitute of Engineeing And Technology
Mijar,Moodbhiri-574225
pradeep@aiet.org.in

Diya HB
Dept of ISE (student)
Alva's insitute of Engineeing And Technology
Mijar,Moodbhiri-574225
diyahb20@gmail.com

Sushma KN
dept of ISE (student)
Alva's insitute of Engineeing And Technology
Mijar,Mudbidri, Karnataka-574225
sushma.kn998@gmail.com

Sudeep K
dept of ISE (student)
Alva's insitute of Engineeing And Technology
Mijar,Moodbhiri-574225
sudeepkumar07@gmail.com

Abstract—**Path finding in autonomous vehicles is a new use for computer vision.Autonomous driving depends on two key subfields of computer vision: semantic segmentation and semantic scene interpretation. A number of sizable sample datasets and deep learning techniques are used to create an appropriate model for path finding semantic segmentation. Given the significance of this work, reliable and accurate models must be trained to function well across a range of lighting and weather scenarios as well as in the presence of noisy input data. The study assesses layer-wise training, a novel learning approach for semantic segmentation, on an efficient neural network (ENet), a lightweight architecture.On two RGB photo data sets covering road (CamVid) and off- road (Freiburg Forest) pathways, the performance of the proposed learning method is compared with the classical learning approaches in terms of mIoU, . Transfer learning is only partially required when using this approach. Additionally, it enhances network performance when input is loud.**

Keywords- **Autonomouscars,layer-wisetrains, Computer vision,convolution neural networks.**

I. INTRODUCTION

Pixel-level classification is the foundation for image semantic segmentation.Unlike instance segmentation, which is frequently used, this type of segmentation does not distinguish between different objects of the same class. Since segmentation may extract important information from a photograph pixel by pixel, it is frequently utilized when the shape of an item is unknown or varies depending on the scene. Rather than creating a final label for the given image, this method labels every pixel at the end of the process. Numerous computer vision-related professions, such as those in wearable augmented reality, home automation, self-driving automobiles, etc., Convolutional Neural Networks (CNNs) can find, identify, and classify each pixel in a picture. Thanks to big labeled datasets and powerful processors, Deep Convolutional Neural Networks (DCNNs) have lately outperformed many popular computer vision techniques. These factors have led to the application of several CNN designs, including Alexnet, ResNet, VGGnet, and GoogleNet, for segmentation techniques in a variety of study fields.An extensively trained basmodel is used to boost the degree of precision. Subsequently, the pre-trained model is periodically modified and Transfer Learning (TL) is applied to improve the current model using the destination dataset. Because of the changing and varied conditions, including weather, light, and color, perception outside is more difficult. even in controlled outdoor settings like city streets.lots of obstacles, such as puddles, to aid in finding strange objects. Robust and comprehensive scene comprehension data are necessary for the segmentation process in autonomous driving. Autonomous driving has made extensive use of image segmentation in both on- and off-road scenarios. For supervised learning in road and off- road path segmentation algorithms, a large number of labeled datasets are available. RGB camera-based labeled datasets make up the majority of currently accessible datasets for semantic segmentation. Low sample datasets based on other technologies, such as near-infrared sensors, RGB-D, and LiDAR, are also available. In this work, we compress the so-called DCNN and speed up the training process without significantly compromising accuracy by using a well-designed and efficient DCNN named ENet with layer-wise training.

The following is a summary of the primary contributions made by the suggested training approach:

- Transfer learning is limited to training on the target dataset; it is not required for R-wise training.

As layer-wise training has been used to extract features more precisely, adding noise to the input photos can yield consistent results.

- There's a chance that the final training will have a much less number of training epochs.
- A layer-wise training approach can shrink the model without impairing IoU by eliminating some encoder layers.

Use the DCNN architecture on two well- characterized datasets (CamVid and Freiburg Forest) that contain both off-road and urban areas. Before we get into the ENet structure and our layer-wise training methodology, we first cover some of the existing techniques and datasets are used.

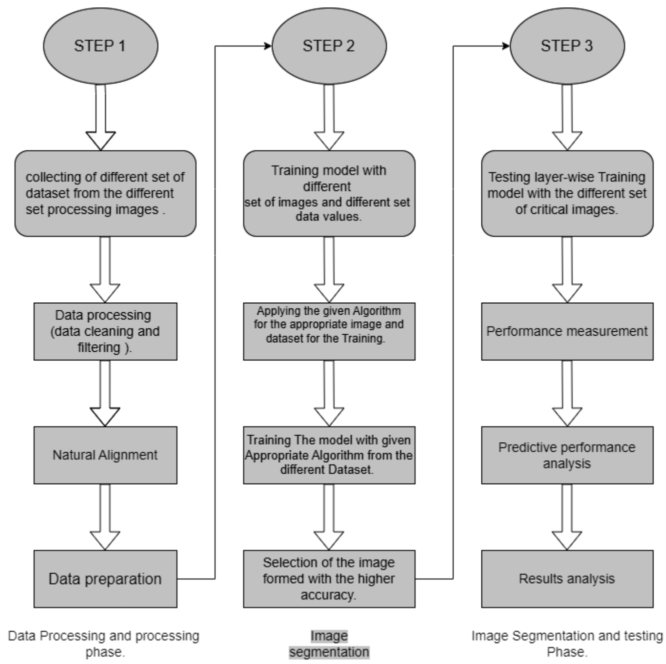
II. LITERATURE REVIEW

In order to validate a machine learning model's accuracy and dependability in a variety of agricultural industries, it is imperative to collect data from multiple geographies. Every area has distinct soil compositions, climate quirks, and environmental circumstances, all of which have an impact on crop growth. It is possible to guarantee that the model is reliable, flexible, and able to provide appropriate recommendations across a range of farming circumstances by gathering data from various places.[1] A significant area of computer vision, image semantic segmentation has many real-world applications, including autonomous driving, virtual or augmented reality, medical picture analysis, and more. Owing to the transformer and multilayer perceptron (MLP)'s exceptional performance in computer vision—which is comparable to that of convolutional neural networks (CNNs)—a significant number of image semantic segmentation works have been conducted recently with the goal of creating various deep learning architectures. The goal of this paper is to present a thorough overview of deep learning techniques for general image semantic segmentation. First, a list of frequently used datasets for picture segmentation is provided.[2] One of the main areas of research for remote sensing analysis is "road detection," which is considered to be difficult because of the data's complexity, which can vary greatly within and across classes and result in errors and gaps in the road's extraction. Furthermore, most supervised learning methods suffer from insufficient training data or the high cost of manual annotation. Thus, the purpose of this research is to present a novel road detection model.[3]In recent years, deep neural networks have been widely applied to semantic scene interpretation. Safe autonomous vehicle navigation requires strong and efficient segmentation in outdoor environments. Instead of employing a single RGB modality, the goal of this work is to determine the optimal utilization of several imaging modalities for road scene segmentation. We investigate early and later fusion patterns for semantic segmentation based on deep learning, and we suggest a novel multi-level feature fusion network. The network may include more contextual information and reach faster convergence given a pair of aligned multimodal images.[4] Semantic scene understanding is crucial for robust and safe autonomous navigation, particularly so in off-road environments. Recent deep learning advances for 3D semantic segmentation rely heavily on large sets of training data, however existing autonomy datasets either represent urban environments or lack multimodal off-road data. We fill this gap with RELLIS-3D, a multimodal dataset collected in an off-road environment, which contains annotations for 13,556 LiDAR scans and 6,235 images. The data was collected on the Rellis Campus of Texas A\&M University and presents challenges to existing algorithms related to class imbalance and environmental topography. Additionally, we evaluate the current state-of-the-art deep learning semantic segmentation models on this dataset.[5]The structure and operation of a dataset created to help autonomous cars identify off-road terrain from a single monocular image are described in this work. More than 12,000 off-road terrain photos and the associated sensor data from a wheel rotation speed sensor, an inertial measurement unit (IMU), and a global positioning system (GPS) are included in this dataset.

III PROPSED WORK

For off-road path segmentation and supervised learning, a multitude of labeled datasets containing road data from diverse types of sensors are accessible. Path detection usually makes use of sensors such as light detection and ranging (LiDAR), RGB cameras, RGB-D, and near- infrared sensors. Nonetheless, the majority of labeled datasets now in use for path semantic segmentation come from RGB and LiDAR cameras.In the following sections, we will review some CNN architectures that have been proposed for semantic segmentation tasks in autonomous driving. Autonomous driving on different types of highways relies heavily on computer vision to recognize routes and avoid both moving and stationary things. Semantic segmentation using large sample datasets has yielded dependable models for road segmentation tasks.

The below Figure shows the proposed model presented in Fig 1.



1) *Road Semantic Segmentation Methods*

For DCNNs to function, an appropriate and sufficient amount of data is needed. There are datasets with fairly substantial data and accurate labeling for path detection in urban environments. For off-road path detection, there are, however, far less datasets than in urban areas, and the tagging of existing datasets has been less precise. A precision prediction in supervised learning necessitates a large amount of well labeled data. Because there are few datasets available for off- road sites, many studies use TL.

SegNet, a cutting-edge and useful deep fully CNN architecture, is introduced in. SegNet features a 13-layer convolutional encoder network that issimilar to the VGG16 network, followed by a matching decoder network that concludes with a pixel-wise classification layer. We have tested this model using the SUN RGB-D and CamVid datasets. Dual-Path Dense-Block Networks (DPDBNets) are encoder- decoder designs that incorporate the orthogonal ideas. Only the encoder's features are reused in the dense block. The Freiburg Forest and CamVid datasets were used to assess the suggested architecture.

Introduced in, SegNet is a state-of-the-art and practical deep fully CNN architecture. SegNet includes a VGG16-like 13-layer convolutional encoder network, a corresponding decoder network, and a pixel-wise classification layer at the end. We have used the SUN RGB-D and CamVid datasets to test our model. The encoder- decoder architectures known as Dual-Path Dense- Block Networks (DPDBNets) use the orthogonal concepts. In the dense block, just the encoder's features are utilized again. The recommended architecture was evaluated using the Freiburg Forest and CamVid datasets.

A novel efficient deep neural network design named ENet is introduced in for tasks demanding low-latency operations. This method has an accuracy comparable to or better than existing models such as SegNet, with 79 times fewer parameters and up to 18 times faster processing speed. Moreover, 75 times fewer FLOPs are needed. This model was evaluated using the CamVid, Cityscapes, and SUN datasets, and its accuracy and processing speed trade-offs were compared to those of other state-of-the-art models.

The investigations additionally employ robust segmentation through mixed techniques approaches. Learning from fused representations is one of these. For example, the research proposed a unique semantic segmentation architecture and the Convoluted Mixture of Deep Experts (CMoDE) fusion techniques. CMoDE enables a multi-stream Deep Neural Network (DNN) to learn features from complementary modalities and spectra.

The model analyzes and evaluates expert network class-specific features depending on scene conditions in order to learn fused representations. This model is evaluated using three publicly available datasets: Freiburg Forest, Cityscapes, and Synthia. Using a multi-task technique, robust segmentation can also be accomplished by sharing a common latent space.

The aforementioned dataset contains a variety of dataset types that are used to measure both on- and off-road paths with various set values provided to determine whether an object is traveling in the intended direction.This is trained using an alternative algorithm.

2) *System Requirements And Specification*

The entire description of the behavior of the system that has to be constructed is contained in a Software Requirements Specification (SRS). SRS is a document that outlines all the functions that the suggested program should have without going into detail about how it will execute those functions. It is a two-way policy that at all times the company and the client are aware of what is expected of them. The SRS document itself is accurate and offers the features and functionalities of the system that it ought to. SRS's primary goal is to improve communication amongst the various parties engaged in the software development process. It contributes to the design specification.

4) *Properties of SRS*

- True: When all software replies to input data classes are included in the SRS, the SRS is considered full.
- Complete: A set of standards is considered clear if and only if each requirement is understood in one way.
- Unambiguous**:** A SRS can only be verified if the specified requirement can be verified.
- Verifiable requirements are those that can be independently verified using an affordable approach to determine whether the finished software satisfies the need.
- An SRS is considered changeable if its style and structure allow for easy modifications that maintain consistency and completeness.

3) *Methodology And Architecture*

We conducted tests on two distinct datasets for the task of road semantic segmentation. Initially, we utilize the Freiburg Forest dataset, which is a real-world off-road autonomous vehicle dataset. The second dataset, named Camvid, is a real-world road scene interpretation dataset for urban semantic segmentation tasks. We provide a brief description of each in the section that follows.

The dataset "FREIBURG FOREST" includes six classes related to forest scenes: sky, road, tree, grass, vegetation, and obstacle. Different from highly structured urban sceneries (buildings, for example) are unstructured off-road environments, such as trails. For the training and test sets, the dataset contains 230 and 136 samples.

	Data Type		
Name	RGB	LiDAR	Type
Freiburg Forest [11]	✓		Off-road track
Yamaha-CMU [12]	✓		Off-road track
RELLIS-3D [13]	✓	✓	Off-road track
Off-Road Terrain [14]	✓		Off-road track
RUGD [15]	✓		Off-road track
KITTI [16]	✓	✓	Road track
CamVid [17]	✓		Road track
CityScapes [18]	✓		Road track
Mapillary [19]	✓		Road track

Table-1 Showing Different set dataset that available

For the road as well as the off-road segment.

4) *Model Testing with Essential Features*

Model Testing: Similar to crop recommendation, this involves evaluating the trained model's accuracy on unseen datasets of road images. Here, the unseen data represents real-world driving scenarios the car might encounter.

- *Model Evaluation*:You analyze the model's performance in segmenting roads in the unseen test set. Metrics like:
- *Intersection over Union (IoU):* Measures the overlap between the predicted and actual road segmentation.
- *Pixel-wise accuracy:* Calculates the percentage of pixels correctly classified as road or non-road.
- *Mean Absolute Error (MAE):* Measures the average difference between predicted and actual road boundaries.

*Analyzing Outcomes:*This stage involves assessing the model's effectiveness in real-world road segmentation tasks.Identify potential biases, such as the model performing poorly under certain lighting conditions.Refine the model based on the evaluation results. This could involve adjusting hyper parameters, collecting more diverse training data, or potentially fine-tuning additional layers in the network.Taken at a frequency of 20 Hz and a resolution of 1024 × 768 pixels in order to measure the lighting-related variability in the data. All of the utilized samples are completely labeled RGB pictures.

DCNNs require appropriate and substantial amounts of data. For path detection in cities, there are datasets with a comparatively large amount of data and precise labeling. However, compared to the datasets in metropolitan settings, the offroad path detection datasets are substantially smaller and have less accurate labeling. In supervised learning, a precision prediction requires a high volume of data with precise labeling. Therefore, many studies employ TL because there are limited datasets available for off-road locations. According to, the ENet DNN architecture is a new, quick, and effective DNN architecture.provides a brief summary of this network. The block's several internal parts each have their own dimensions for input and output. As can be observed, three successive convolution layers are present in the prescribed blocks. As such, we are working with a deep structure where the influence of the mistake on the last layers (decoder) is greater than that on the initial layers (encoder), regardless of how much we try to prevent gradient vanishing by applying activation functions, such as ReLU and its derivatives. The encoder section is the first layer that the input goes through and is in charge of feature extraction.Consequently, it is crucial to extract the relevant and accurate features and move them to the following layer. The output of the network in the decoder layers is less dependable the weaker the characteristics extracted in the network's early layers. In order for the whole convolution portion to decide on semantic segmentation and properly label each pixel, these layers decode the retrieved features. Because the first layers serve as the input pathway to the subsequent layers, improper training of these levels results in a decline in the network's overall performance.

B.Our Requirements lie in Terms

According to, the ENet DNN architecture is a new, quick, and effective DNN architecture. a brief analysis of this network. various internal block components, together with the size of each input and output. As can be observed, three successive convolution layers are present in the prescribed blocks. As such, we are working with a deep structure where the influence of the mistake on the last layers (decoder) is greater than that on the initial layers (encoder), regardless of how much we try to prevent gradient vanishing by applying activation functions, such as ReLU and its derivatives. The encoder section is the first layer that the input goes through and is in charge of feature extraction.

Data preprocessing involves resizing photos while preserving aspect ratio to a standard size appropriate for network input. A similar scale, such as [0, 1] or [-1, 1], should be applied to the pixel values to improve convergence during training.

- *Training Strategy:* Divide the datasets into sets for training and validation (e.g., training and validation with an 80-20 split). For semantic segmentation tasks, use a loss function such as cross-entropy loss that is appropriate. Use transfer learning by starting the network with pre-trained weights (such as those found in ImageNet) and optimizing the model using the particular set of data.
- *Assessment:* Utilize evaluation metrics like as IoU, accuracy, precision, recall, and F1-score to appraise the trained model's performance on the test set. Examine the model's predictions graphically to see where it performs well and poorly.
- *Considering:* Use class-weighted loss functions or oversampling strategies to address any class imbalance that may exist in the datasets. To enhance the performance of the model, adjust the hyperparameters in light of validation findings. Try out various network configurations or adjustments to enhance the precision of segmentation, particularly for unstructured settings such as the Freiburg Forest dataset.

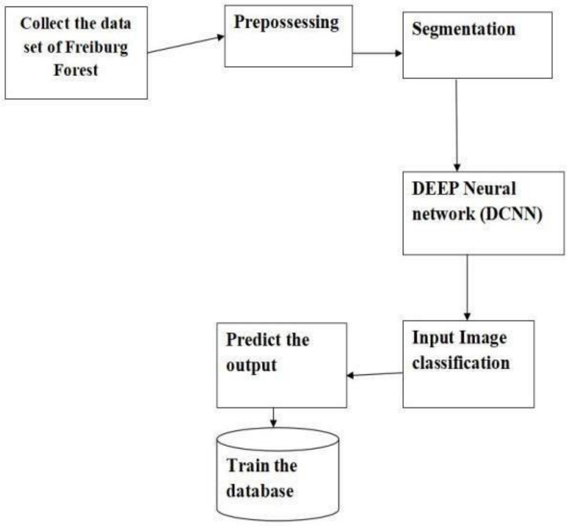


Fig-2 Proposed Flow chart for the Image Segmentation

C.Alogorithm

1. *Distinct Instruction (Layers 1 through L_init)*:
- Within the interval [1, L_init + 1), for each layer i:Take the complete ENet design and extract the sub-network Net_i that contains layers 1 through i.
- Train Net_i for N_iters_init iterations on X_train and Y_train using the suitable optimizer (e.g., Adam).
- Update the weights of layers 1 through i over the whole ENet using the training weights from Net_i.

2. *Comprehensive Network Education*:

- Train the whole ENet (updated layers 1 to L_init) on X_train and Y_train for N_iters_full iterations using the optimizer of choice.
- Projection: For each new image (x_new),
- Feed x_new to the trained ENet model.
- The expected segmentation mask for the new image to be acquired.

3.*Prediction*:

- For every fresh picture x_new:
- Feed x_new into the ENet model that has been trained.
- Obtain the segmentation mask that is projected for x_new.

IV.RESULTS AND DISCUSSIONS

Layer-wise training for road segmentation in autonomous cars shows promise. Research suggests it achieves good accuracy (IoU, pixel accuracy) while requiring less training data compared to training from scratch. This efficiency makes it attractive for real-time applications.

Future work might explore different pre-trained models or data augmentation techniques to potentially improve performance further. Additionally, researchers will need to consider factors like computational efficiency and safety when deploying such models in real-world autonomous vehicles.

.

A. Datasets of different Images and different Areas:

Different set of collection of dataset that are provided for the model for the segmentation process.

Method	Building	Tree	Sky	Car	Sign	Road	Pedestrian	Fence	Pole	Sidewalk	Bicycle	mIoU
Dilation8 [41]	82.6	76.2	89	84	46.9	92.2	56.3	35.8	23.4	75.3	55.5	65.3
BiSeNet [42]	83	75.8	92	83.7	46.5	94.6	58.8	53.6	31.9	81.4	54	68.7
VideoGCRF [43]	86.1	78.3	91.2	92.2	63.7	96.4	67.3	63	34.4	87.8	66.4	75.2
DeepLabv3Plus+SDCNetAug (SS) [44]	90.9	82.9	92.8	94.2	69.9	97.7	76.2	74.7	51	91.1	78	81.7
DeepLabv3Plus+SDCNetAug (MS) [44]	91.2	83.4	93.1	93.9	71.5	97.7	79.2	76.8	54.7	91.3	79.7	82.9
SERNet-Former (ours)	**93**	**88.8**	**95.1**	91.9	**73.9**	97.7	76.4	**83.4**	**57.3**	90.3	**83.1**	**84.62**

Table-2 Camvid Dataset for the segmentation of the Images.

Fig-3 Represtation of the Freiburg Forest Image Segmentation.

B. Selection of the Image based on the Area and the type Road.

Extraction of Image for the Segmentation and Layer-wise Training for Road Segmentation with the real world images And LIDAR,cameras Images that taken with real-time sensors data for the environmental parameters (like fog,rain,sunny).the model might be adjusted slightly to account for these conditions, essentially adapting its road segmentation behavior.

C. Performance Metrics of the proposed work.

Segmentation Accuracy: The structure and operation of a dataset created to help autonomous cars identify off-road terrain from a single monocular image are described in this work. More than 12,000 off-road terrain photos and the associated sensor data from a wheel rotation speed sensor, an inertial measurement unit (IMU), and a global positioning system (GPS) are included in this dataset.

Generalization:

- *Performance in Various Conditions:* Assess the model's accuracy in a range of road conditions, taking into account changes to:

1. Lighting (day, nightly, and shadowed).
2. Climate (sunny, cloudy, or snowy).
3. Types of roads: city streets, rural roads, and highways.

- *Mean Absolute Error (MAE):* This statistic calculates the average (in pixels) difference between the actual and forecasted road borders. A smaller MAE denotes more accurate segmentation.

Training Effectiveness:

- *Training Time:* Examine the difference in training time between building a model from scratch and layer-wise training. This illustrates how using prior knowledge can increase efficiency.
- *Information Needs:* Compare the amount of training data required for layer-wise training to that of a fully trained model in order to determine how much performance is achieved. This demonstrates possible increases in data efficiency.

D. *Data set of the proposed*

From the below shows the Different segmentation for the images that is to be segmented and the image processing for the segmentation of the given Data set images with testing and the training code.

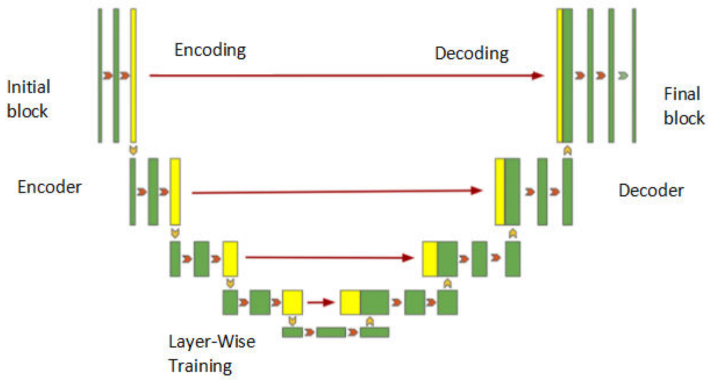


Fig-4 Segmenatation of the Image Using the Layer-Wise training by making use of the convaluted Horizontal and vertical Strping of the image.

This is the visual analysis of the image segmentation using the Layer-wise Training Algorithm that Segments the image and get different values that are identified in the image representsClear image given to the model. Also, the represents the Image being Segmented using Model.

E. Output Of The Proposed Mode

The selection of the datsets for the image segmentation of the cars moving in the plain road as well as the off-road in the first process is the selection phase/process here different set of images are taken from the dataset to process into the model for the segmentation of the image using the Algorithm.The system with the specific requirements are processed for the prediction of the encoding and decoding of the image to yield the overall increase in the accuracy,performance,precision,image processing.. The input and output of the proposed model is as shown in the Fig 5.

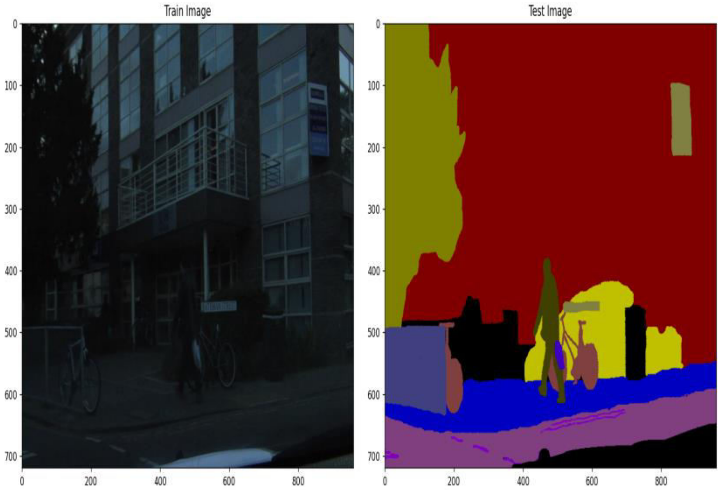


Fig-5 Input and output for the proposed model

CONCLUSION

In general, layer-wise training strengthens the learned model by improving the efficiency of feature learning in the first and middle layers of encoders. We introduced noise into the input to illustrate its robustness. When compared to traditional training, the enhanced feature extraction in the early layers produced a greater final accuracy.

Furthermore, the notion of Transfer Learning has been applied to the bulk of datasets with fewer data, such Freiburg Forest. The network must initially be trained on a larger dataset in order to use this method. Subsequently, the network ought to undergo another training cycle on the target dataset, occasionally with and occasionally without network layer alterations.

It takes longer to train and requires a larger dataset for this task. This shows that layer-wise training, which does not involve transfer learning, is used to train the network using only the target data set. Due to the limited training data without Transfer Learning, the suggested learning technique has not reduced the network's ability to detect threats, and as shown, layer- wise trained networks are resilient to noisy input.

REFERENCES

[1] "ENet: A deep neural network architecture for real-time semantic segmentation," by A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello arXiv:1606.02147, 2016.

[2] "ImageNet classification with deep convolutional neural networks," by A. Krizhevsky, I. Sutskever, and G. E. Hinton ACM Communications, vol. 60, no. 2, June 2012, pp. 84–90.

[3] "Drivable area detection using deep learning models for autonomous driving," by D. Qiao and F. Zulkernine, in Proceedings of the IEEE Int. Conf. Big Data (Big Data), December 2021, pp. 5233–5238.

[4]"Optimized convolutional neural network for road detection with structured contour and spatial information for intelligent vehicle system," by D. K. Dewangan and S. P. Sahu Journal of Pattern Recognition and Artificial Intelligence, volume 36, issue 6, May 2022, art. no. 2252002.

[5] A survey of image semantic segmentation using deep learning approaches was conducted by J. Cheng, H. Li, D. Li, S. Hua, and V. S. Sheng in Comput., Mater. Continua, vol. 74, no. 1, pp. 1941–1957, 2023.

[6] "Semantic segmentation with transfer learning for off-road autonomous driving," by S. Sharma, J. E. Ball, B. Tang, D. W. Carruth, M. Doude, and M. A. Islam, Sensors, vol. 19, no. 11, p. 2577, Jun. 2019.

[7] "Exploration of deep learning-based multimodal fusion for semantic road scene segmentation," by Y. Zhang, O. Morel, M. Blanchon, R. Seulin, M. Rastgoo, and D. Sidibé, is published in Proc. 14th Int. Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl., 2019, pp. 336–343.

[8] The paper "In2I: Unsupervised multiimage-to- image translation using generative adversarial networks" was presented by Perera, M. Abavisani, and V. M. Patel in the Proceedings of the 24th International Conference on Pattern Recognition (ICPR), August 2018, pp. 140–146.

[9] SegNet: A deep convolutional encoder–decoder architecture for image segmentation, V. Badrinarayanan, A. Kendall, and R. Cipolla, IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[10] Deep multispectral semantic scene understanding of forested landscapes via multimodal fusion, A. Valada, G. Oliveira, T. Brox, and W. Burgard, Proc. Int. Symp. Exp. Robot. (ISER), 2016, pp. 465–477.

[11]Real-time semantic mapping for autonomous off- road navigation, D. Maturana, P.-W. Chou, M. Uenoyama, and S. Scherer, Field and Service Robotics. 335–350 in Springer, 2018, pp.

[12]"An off-road terrain dataset including images labeled with measures of terrain roughness," by G. Gresenz, J. White, and D. C. Schmidt, in Proc. IEEE Int. Conf. Auto. Syst. (ICAS), Aug. 2021, pp. 1–5.

[13]A new performance metric and assessment standard for road detection algorithms, J. Fritsch, T. Kuhnl, and A. Geiger, in Proc. 16th Int. IEEE Conf. Intell. Transp. Syst. (ITSC), Oct. 2013, pp. 1693–1700.

[14]"Segmentation and recognition using structure from motion point clouds," Proc. ECCV, 2008, pp. 44–57, G. J. Brostow, J. Shotton, J. Fauqueur, and R. Cipolla.

[15] "The cityscapes dataset for semantic urban scene understanding," by M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, was published in the IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), June 2016, pp. 3213–3223.

[16] In the IEEE Int. Conf. Comput. Vis. (ICCV), October 2017, pp. 4990–4999, G. Neuhold, T. Ollmann, S. R. Bulo, and P. Kontschieder presented their work, "The mapillary vistas dataset for semantic understanding of street scenes."