Applied Artificial Intelligence

COMP534 ***Assignment – III***

Name	Student ID	Email
Santhosh Arunagiri	201586816	sgsaruna@liverpool.ac.uk
Roshan Prakash	201670489	sgrprak2@liverpool.ac.uk

PROJECT GOAL:

This project aims to solve a semantic segmentation task over 32 object classes in a driving environment using 3 deep learning models.

- PSPNET
- UNET
- SEGNET

DATASET:

The Cambridge Labelled Objects dataset is highly benchmarked as a rich and comprehensive dataset in object recognition, scene analysis for computer vision and semantic segmentation tasks. It has 32 object classes from 101 images classified in Red Green Blue values from 0 to 255, diverse content, and detailed annotations make it a valuable resource for training and evaluating computer vision algorithms.

MODEL DESIGN:

- **Pyramid Scene Parsing Network** architecture is a popular approach for pixel-level scene parsing which labels pixels in an image. PSPNet makes use of the pyramid pooling module to effectively gather contextual information, allowing the model to have a global awareness of the scene while retaining local details over 4 layers such as Feature Extraction, Pyramid Pooling Module, Convolutional Decoder and SoftMax or Sigmoid. This combination yields precise and thorough semantic segmentation outcomes.

- **U-Net** architecture performs semantic segmentation at the pixel level and extract key information from the images, U–net is widely used in medical imaging tasks. The U-Net architecture is intended to incorporate the benefits of both localization and contextual information from the expanding path. By retaining geographical features across the network, skip links enable exact segmentation over 6 layers such as 2 Convolutional layers, Bottleneck, Skip Connections and Expansion of path. This structure has been shown to be useful for a variety of medical picture segmentation tasks as well as general semantic segmentation tasks.

- **SEGNET** architecture is known for efficiency and effectiveness of the model from pixel level, it fetches detailed analysis and observation of the visual content from the images. The usage of pooling indices is a major element of SegNet, which aids in exact localisation while lowering the number of learnable parameters when compared to other designs which uses the similar 4 layers as PSP. SegNet can rebuild the segmented output with pixel-level accuracy by exploiting the pooling indices during the up-sampling phase.

HYPERPARAMETERS:

- Learning rate scheduler, activation functions and optimizers are some of the hyperparameters investigated. The find_best_hyperparameters function receives a model function, a training dataset, a criterion function, and an optional number of cross-validation folds. The function outputs the best accuracy and hyperparameters found during the search.

OVERFITTING / UNDERFITTING:

- In machine learning models, overfitting and underfitting are frequent problems. When a model performs exceptionally well on the training dataset but poorly on the validation or testing datasets, this is known as overfitting. Underfitting, on the other hand, happens when the model does not perform well on both the training and the validation or testing datasets. there is no overfitting problem. However, the difference between the training and testing accuracies is not large, which indicates that the model might be underfitting the data. To deal with underfitting, we need to try increasing the complexity of the model, for example, by adding more layers or increasing the number of neurons in the existing layers. Also, can explore different activation functions, optimisers, and learning rates. since this model is simple, increasing the number of neurons in the layers, adding more layers, or trying different activation functions would be good idea.

TECHNIQUES:

- There are three distinct model architectures defined: UNet, SEGNet, and PSPNet. These models are made up of encoder and decoder components, with the encoder extracting features from the input image and the decoder up sampling the features to form a segmentation map.

- Each model is made up of convolutional, pooling, and up-convolutional layers for up sampling. The DoubleConv class depicts a block with two convolutional layers that are followed by batch normalization and ReLU activation. The Up class is used to upsample the feature maps and concatenate them using encoder skip connections.

- The UNet model is symmetric, with skip links between the encoder and decoder. The SEGNet model employs a similar design, but adds a max-pooling and up-convolutional layer.

- Once the model is Pretrained on the dataset, we further fine tune the algorithm with a pretrained models on certain tasks or domains can boost their performance even further. The model can learn to provide more contextually relevant replies for those specific tasks by exposing it to task-specific data and fine-tuning its parameters. This strategy assists in aligning the model's behaviour with the target task, resulting in a more effective and accurate model. However, if the fine-tuning data is inadequate or unrepresentative of the goal task, it may result in overfitting or poor generalisation to previously unseen samples.

- Reinforcement learning (RL) can be used to fine-tune a language model by rewarding or punishing it based on the quality of its responses. RL can help enhance model performance by optimising for certain criteria such as relevance, coherence, or human input. The model can use RL to experiment with alternative response techniques and learn from the feedback it receives. RL training, on the other hand, can be computationally and time-consuming. Furthermore, RL might result in unstable or undesired behaviour if the reward function is not adequately designed or the training process is not carefully regulated.

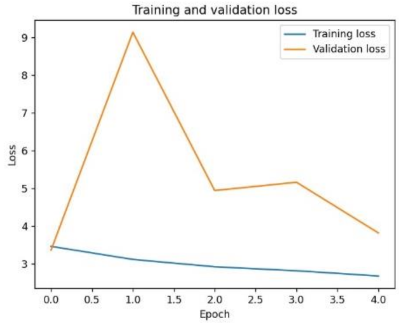
- Filtering and moderating input and output data during training can help improve the model's performance in terms of safety, ethics, and compliance with norms. The model's responses can be more trustworthy and aligned with society standards if biassed, damaging, or improper stuff is excluded or mitigated. Excessive filtering or moderation, on the other hand, can cause the model to be overly cautious and produce overly generic or unhelpful results, reducing its effectiveness.

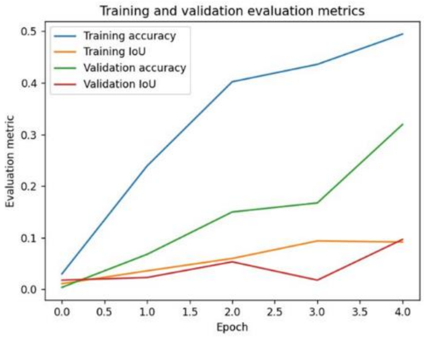
- Active learning is iteratively selecting and annotating data to train the model, with a focus on difficult or ambiguous examples. Human reviewers or experts can profit from their expertise and improve the model's performance on specific tasks or areas by actively involving them in the training process. Active learning assists the model in addressing its flaws and focusing on areas that require improvement.

RESULTS:

After performing Cross validation of the hyperparameters with an epoch of 5 we were able to get a lot of results, and here we display some of our better results for all models.

UNET:





TRAIN LOSS	TRAIN IOU	TRAIN ACCURACY
2.67773	0.09185	0.49470

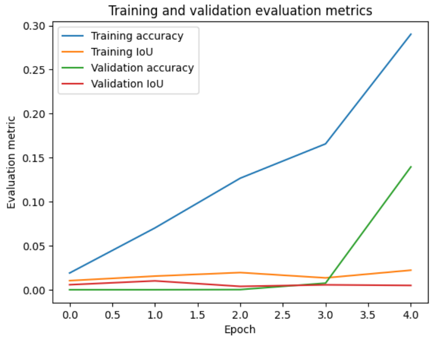
VALIDATION LOSS	VALIDATION IOU	VALIDATION ACCURACY
3.81991	0.09687	0.31949

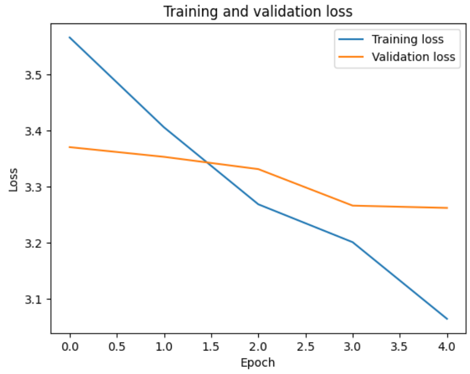
TEST LOSS	TEST IOU	TEST ACCURACY
4.29168	0.09997	0.22393

BEST HYPERPARAMETERS:

- Learning rate: 0.001
- loss function: Cross Entropy Loss
- optimizer: Adam

SEGNET:





TRAIN LOSS	TRAIN IOU	TRAIN ACCURACY
2.48583	0.13126	0.53669

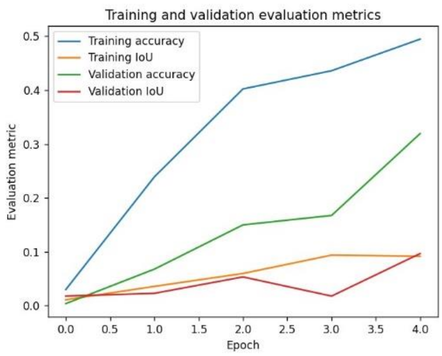
VALIDATION LOSS	VALIDATION IOU	VALIDATION ACCURACY
5.72603	0.05981	0.33409

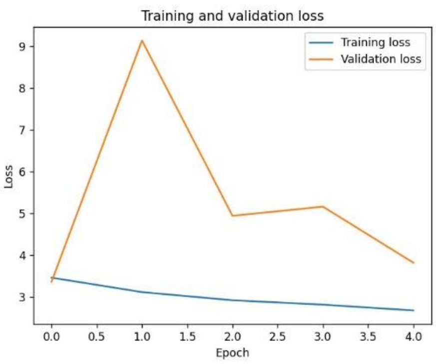
TEST LOSS	TEST IOU	TEST ACCURACY
6.16995	0.15318	0.29454

BEST HYPERPARAMETERS:

- Learning rate: 0.001
- Loss function: Cross Entropy Loss
- Optimizer: Adam

PSPNET:





TRAIN LOSS	TRAIN IOU	TRAIN ACCURACY
3.00270	0.02048	0.31792

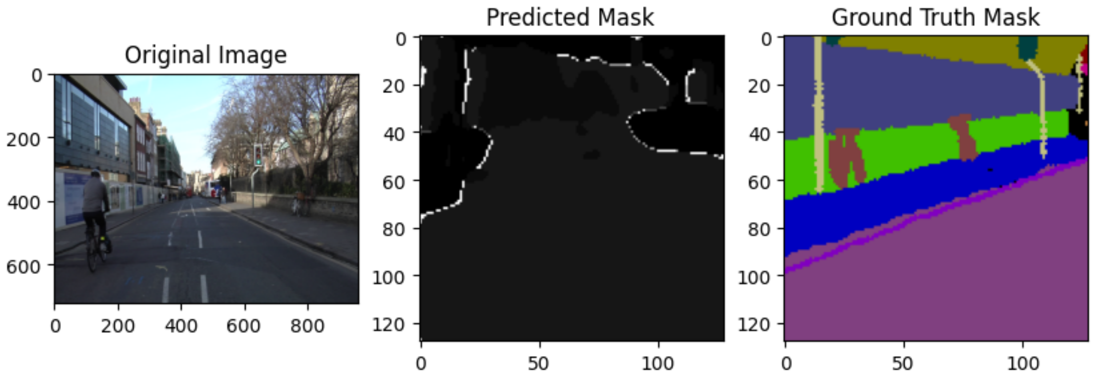
VALIDATION LOSS	VALIDATION IOU	VALIDATION ACCURACY
1.39267	0.00243	0.90745

TEST LOSS	TEST IOU	TEST ACCURACY
2.79847	00.04774	0.49393

BEST HYPERPARAMETERS:

- Learning rate: 0.001
- Loss function: Cross Entropy Loss
- Optimizer: Adam

And upon the completion of the running, we were able to visualize some results as



OBSERVATION:

Among these three models, PSPNet performed well in terms of validation IOU and validation accuracy, followed by SEGNet and then UNet. Though we received best, when it comes to the test set, PSPNet and SEGNet both outperformed UNet, which had the lowest test accuracy and IOU.

UNet, being one of the most widely used models, was easy to implement but the simple structure of it makes the model work less efficient when compared to other models.

SEGNet has also been applied in various applications and has shown good results in terms of accuracy and efficiency. In the observed results, SEGNet performed better than UNet but was outperformed by PSPNet in terms of validation accuracy and IOU.

The pyramid pooling module in the PSPNet model allows the model to effectively capture context at multiple scales, which improvs the model's ability to understand the context of the scene.
PSPNet achieved the highest validation accuracy and IOU in the provided results, indicating its superiority for the given dataset. But still it failed to give the expected output in the test set.

We believe increasing the number of epochs and variety of hyperpaparameters will result us a better result.

CONCLUSION:

Convolutional neural networks (CNNs), in particular, have revolutionised computer vision and perception in autonomous vehicles. Neural networks can reliably detect and classify objects, identify traffic signs, and comprehend complicated scenarios by training on enormous volumes of data, greatly enhancing the vehicle's capacity for autonomous navigation.

Vehicle perception and understanding of their surroundings are greatly aided by picture segmentation, the division of an image into relevant parts. Lane markers, road limits, and obstructions can be found and distinguished with the use of image segmentation. Autonomous vehicles may efficiently plan their trajectories, uphold lane discipline, and confidently negotiate challenging road conditions by extracting this crucial information from the visual data.

By these models we can clearly see that PSPNET performs better on this dataset in comparison with other two models. but still the best model is not providing the expected result.
We might need to explore other models. Or we need to improvise the PSPNet in terms of layers and parameters.

We had difficulty in terms of hardware as the models were exhausting the ram limits were not able to run the models in scheduled time. Performing crossvalidation made it more difficult.

This assignment helped us understand the working of the neural network models for image recognisation. Loading the image data, building a suitable model, crossvalidating with different hyperpaparameters. All were very helpful for us in learning process of the machine learning model building.