Image Processing in Autonomous Car using Deep Learning

Abstract

This work presents an investigation into the application of VGG16+U-Net and ResNet50+U-Net architectures for image processing tasks, specifically **segmentation** and **edge detection**, in the context of **driverless car** pictures. With the aim of enhancing the perception capabilities of autonomous vehicles, these **deep learning** models are trained and evaluated using the Cityscapes dataset.

The VGG16+U-Net architecture combines the **VGG16** model, pre-trained on ImageNet, as a feature extractor, with the U-Net architecture for pixel-level segmentation. This hybrid approach allows for efficient feature extraction and accurate pixel-wise classification, enabling precise segmentation of objects and regions of interest in driverless car images.

Similarly, the ResNet50+U-Net architecture utilizes the **ResNet50** model, pre-trained on ImageNet, for feature extraction and integrates it with the **U-Net** architecture. By leveraging the skip connections and residual blocks of ResNet50, this model captures both low-level and high-level features, leading to improved edge detection performance.

The **Cityscapes dataset**, with its high-resolution images and pixel-level segmentation annotations, serves as a valuable resource for training and evaluating these models. It provides diverse driving scenarios, including urban environments, with annotated ground truth labels for accurate evaluation of segmentation and edge detection performance.

Experimental evaluations are conducted to assess the effectiveness of the VGG16+U-Net and ResNet50+U-Net architectures on the Cityscapes dataset. Performance metrics such as precision, recall, mIoU and F1 score are measured, highlighting the models' ability to accurately segment objects and detect edges in driverless car images.

What is Deep Learning?

Deep learning is a type of machine learning that uses artificial neural networks to learn from data. Neural networks are inspired by the human brain, and they are able to learn complex patterns from data [6].

Analogy - Imagine that you're trying to learn how to recognize different types of animals. You could start by reading a book about animals, but that wouldn't be enough. You would also need to see pictures of animals, and you would need to practice identifying them. The more you practice, the better you would become at recognizing animals.

Deep learning models work in a similar way. They are trained on data, and the more data they are trained on, the better they become at performing their tasks. The data that is used to train deep learning models for animal recognition is typically images of animals. The models are trained to identify the different features of animals, such as their shape, size, and colour. Figure 1 shows the model takes in an image of an animal, and it uses the features that it has learned to identify the animal.

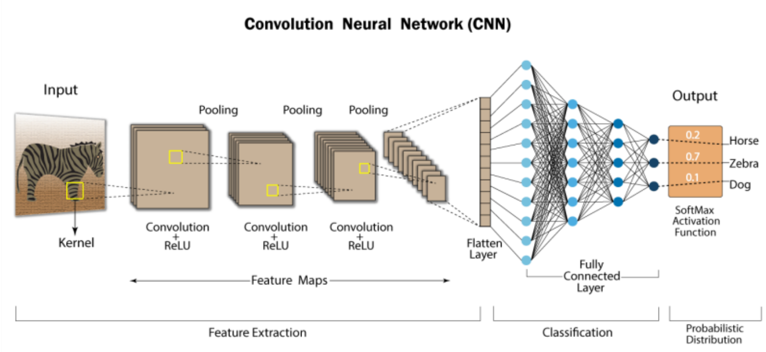

Figure 1, CNN – a class of artificial neural network

Deep learning is a powerful tool that is being used to solve a variety of problems. It is being used to develop self-driving cars, create virtual assistants, and improve medical diagnosis. As deep learning models become more sophisticated, they will be able to solve even more complex problems.

How good is it (state of the art)

Deep learning models can achieve state-of-the-art results on a variety of tasks. For example, deep learning models have been shown to be able to recognize objects in images with an accuracy that is better than humans can. They have also been shown to be able to translate languages with an accuracy that is close to the human level.

The reason why deep learning models are so good at these tasks is that they can learn from large amounts of data. The data that is used to train deep learning models are typically very complex, and it would be difficult for humans to learn from this data on their own. However, deep learning models can learn from this data by using a technique called artificial neural networks.

Artificial neural networks are inspired by the human brain. They are made up of a network of interconnected nodes, and each node can learn to recognize a particular pattern in the data. As the model is trained, the nodes in the network learn to recognize more and more patterns, and this allows the model to make better predictions.

The state-of-the-art results that have been achieved by deep learning models are very impressive. However, there are still some challenges that need to be addressed before deep learning models can be used in real-world applications. For example, deep learning models can be computationally expensive to train, and they can be sensitive to the quality of the data that is used to train them.

Despite these challenges, deep learning is a very promising technology. As deep learning models continue to improve, they will be able to solve even more complex problems. They may one day be able to develop self-driving cars, create virtual assistants that can understand our natural language, and even cure diseases.

Achievements in Computer Vision

Deep learning has revolutionized computer vision, leading to remarkable achievements in various domains. These achievements highlight the power of deep learning algorithms in analysing visual data and making intelligent decisions. Some of the key achievements are:

1. **Driverless Cars:** Driverless cars have emerged as one of the ground-breaking achievements in computer vision and autonomous systems. By leveraging deep learning techniques, driverless cars can perceive the environment, analyse sensor data, and make informed decisions for safe and efficient navigation. Deep learning plays a crucial role in perception systems, enabling accurate detection of objects, pedestrians, and obstacles [1]. It also contributes to decision-making algorithms, allowing driverless cars to interpret complex traffic situations and react appropriately.

2. **Healthcare:** Deep learning has made significant contributions to medical image analysis. It has enabled automated detection and diagnosis of diseases, such as cancer, through techniques like image segmentation, lesion detection, and classification [2]. Deep learning models have shown promising results in detecting abnormalities in medical scans and assisting healthcare professionals in making accurate diagnoses.

How Does Deep Learning Work in Driverless Cars?

In driverless cars, deep learning is used to perform a variety of tasks, such as:

- **Object detection:** This is the task of identifying objects in the environment, such as cars, pedestrians, and traffic signs [10].
- **Scene understanding:** This is the task of understanding the context of a scene, such as the layout of the road and the traffic rules that apply [9].
- **Path planning:** This is the task of planning a safe and efficient path for the car to follow [8].

- **Decision making:** This is the task of making decisions about what actions the car should take, such as when to brake, turn, or accelerate [7].

Deep learning is able to perform these tasks because it is able to learn from large amounts of data. The data that is used to train deep learning models for driverless cars typically include images, sensor data, and maps.
Analogy - Imagine that you're trying to learn how to drive a car. You could start by reading a book about driving, but that wouldn't be enough. You would also need to get behind the wheel and practice driving. The more you practice, the better you would become at driving.

Deep learning models work in a similar way. They are trained on data, and the more data they are trained on, the better they become at performing their tasks. The data that is used to train deep-learning models for driverless cars is typically collected from real-world driving scenarios. This data is used to train the models to identify objects, understand scenes, plan paths, and make decisions.

Once a deep learning model is trained, it can be used to control a driverless car. The model takes in data from the car's sensors, such as images from cameras and radar data, and it uses this data to make decisions about how to control the car.

Deep learning is a powerful tool that is being used to develop driverless cars. As deep learning models become more sophisticated, driverless cars will become safer and more efficient.

Figure 2 shows the autonomous car training model at the steering level. The data is collected from different positioned camera sensors and is given to the CNN model to predict the steering angle (right, left or middle). The model's predicted value is then validated against the actual value of the steering angle. Error adjustments are then made if any and weights are updated by using Back Propagation. This process continues for many epochs until the error is minimal or zero.

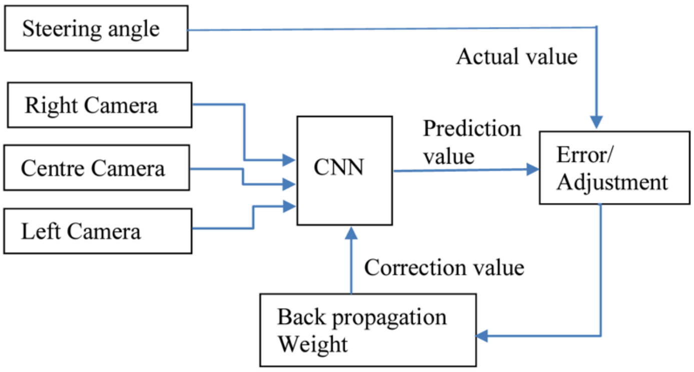

Figure 2, Autonomous car training model [5]

<u>Model Training and Evaluation</u>

The CNN-based network used here is U-Net, one of the most widely used networks in real-time semantic segmentation, in conjunction with ResNet-50 and VGG-16 to demonstrate better performance. Extensive experiments are carried out on the Cityscapes Semantic Segmentation dataset [3].

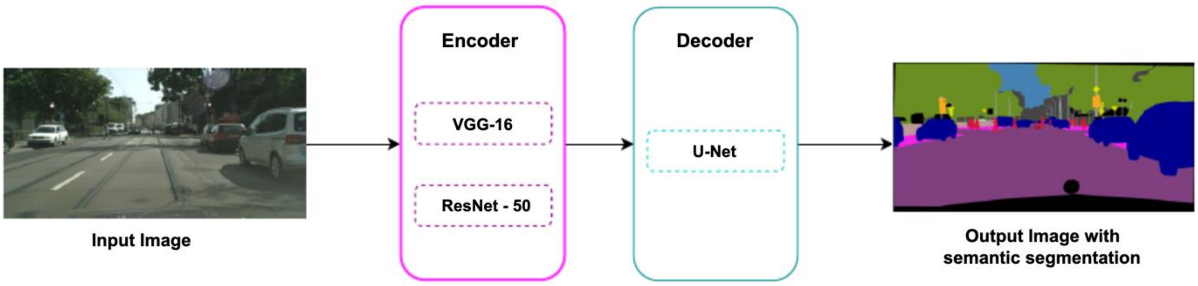

Figure 3, Block diagram of the model

Dataset: Cityscapes dataset is the most widely used large-scale autonomous driving dataset which includes images from urban street scenes. It comprises 5000 well-labelled photos and 20,000 crude labels. In order to train, validate, and test, I used 2975, 500, and 1525 photos respectively. There are 19 different classes in it.

Training: The model was trained on Google Colab with GPU enabled for 50 epochs. The learning rate was set to 0.001 and Adam optimizer and ReLU activation function were used.

Evaluation Metrics: The metrics used for evaluation are mIoU and F1-score.

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{Recall} = \frac{TP}{TP+FN} \qquad \text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \qquad MIoU = \frac{1}{k}\sum_{i=0}^{k} \frac{TP}{TP+FP+FN}$$

<u>Test Results</u>

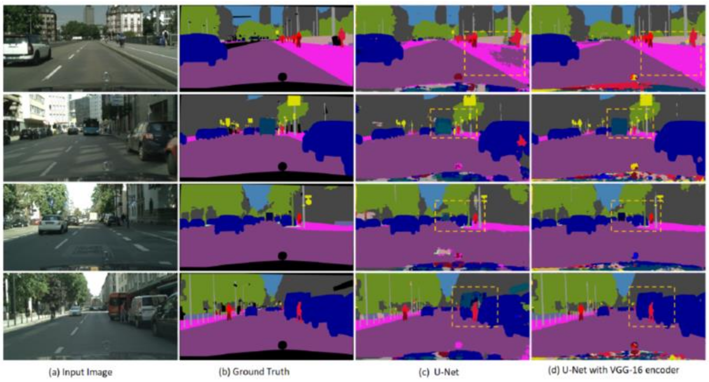

Figure 4, Semantic Segmentation with VGG16+U-Net architecture

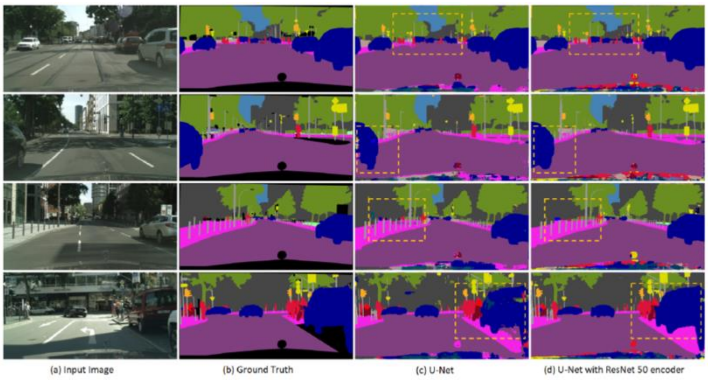

Figure 5, Semantic Segmentation with RestNet50+U-Net architecture

Results: The F1-score and mIoU are 0.86 and 79.4 respectively for U-Net+RestNet-50 and 0.87 and 80.5 respectively for U-Net+VGG-16. It is evident that the U-Net+VGG-16 encoder performed better than U-Net+RestNet-50. The dotted rectangles in the above figures show the potions where VGG16+U-Net and RestNet-50+U-Net performed better than the typical U-Net.

Edge Detection: The RGB image is first converted into a grayscale (1 colour channel) by using OpenCV. Then, added Gaussian noise as it is a way to super-pass noise and spurious gradients (kernel_size=5). Then applied, the Canny edge detection method to detect the road lane boundary [4].

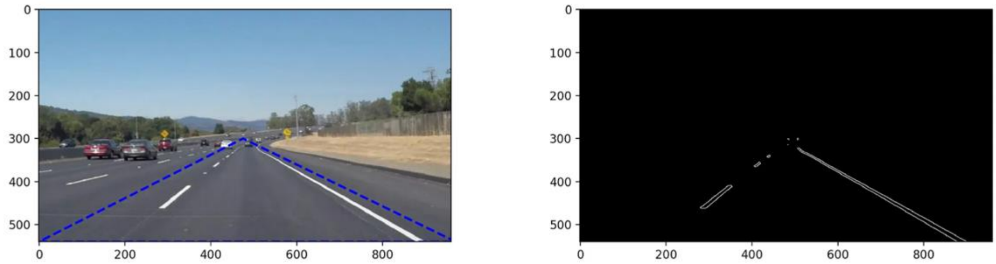

Figure 6, Original Image with selected region (left), Canny masked image (right)

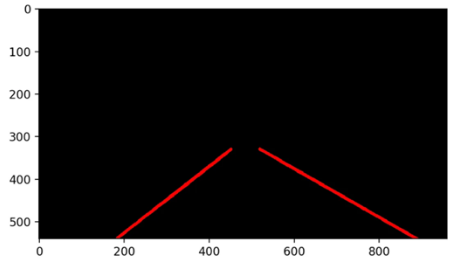

Figure 7, Generated coloured edge of the lane

Conclusion

The findings of this study contribute to the growing body of knowledge in deep learning-based image processing techniques for driverless cars. The superiority of the VGG16+U-Net architecture over the U-Net+ResNet50 architecture emphasizes the importance of carefully selecting and combining models to achieve optimal performance in specific applications. Moreover, the successful integration of the Canny method for lane edge detection highlights the effectiveness of combining traditional image processing techniques with deep learning approaches to enhance the capabilities of autonomous driving systems.

These results provide valuable insights for researchers and practitioners working on developing robust and efficient algorithms for image processing in autonomous vehicles. The knowledge gained from this study can drive further advancements in deep learning-based perception systems, paving the way for safer and more reliable driverless cars in the future.

References

1. Pan, X., Shi, J., Luo, P., Wang, X., & Tang, X. (2017). Spatial as deep: Spatial CNN for traffic scene understanding. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 4879-4888).
2. Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115-118.
3. Thayalan, Sugirtha & Muthukumarasamy, Sridevi. (2022). Semantic Segmentation using Modified U-Net for Autonomous Driving. 1-7. 10.1109/IEMTRONICS55184.2022.9795710.
4. GitHub Repository: AutonomousCar GitHub Repository
5. I Sonata et al 2021 J. Phys.: Conf. Ser. 1869 012071
6. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press
7. Barla, N. (2023, April 25). Self-Driving Cars with Convolutional Neural Networks (CNN). Neptune.ai Blog. Retrieved from https://neptune.ai/blog/self-driving-cars-with-convolutional-neural-networks-cnn
8. Ma, Hao, Wenhui Pei, and Qi Zhang. 2022. "Research on Path Planning Algorithm for Driverless Vehicles" *Mathematics* 10, no. 15: 2555. https://doi.org/10.3390/math10152555
9. K. Muhammad et al., "Vision-Based Semantic Segmentation in Scene Understanding for Autonomous Driving: Recent Achievements, Challenges, and Outlooks," in IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 12, pp. 22694-22715, Dec. 2022, doi: 10.1109/TITS.2022.3207665.
10. Jin Liu and Dongquan Zhang 2020 *J. Phys.: Conf. Ser.* 1575 012150, doi 10.1088/1742-6596/1575/1/012150