International Journal of Engineering Research and Science & Technology

A LIGHTWEIGHT AND MULTI - BRANCH MODULE IN FACIAL SEMANTIC SEGMENTATION FEATURE EXTRACTION

M. V. NARAYANA [1], SEKURI SAI PRANAV [2], GALIMUDI MAHESHWARI [3], YATHARLA AASHRITH REDDY [4], SHIVARATHRI UDAY KUMAR [5]

[1]Professor, Dept. of CSE, Guru Nanak Institutions Technical Campus**,** Ibrahimpatnam, Telangana, India

[2, 3, 4, 5] B.Tech, Dept. of CSE, Guru Nanak Institutions Technical Campus**,** Ibrahimpatnam, Telangana, India

ABSTRACT— Facial semantic segmentation is a fundamental task in computer vision that focuses on accurately identifying and separating different facial regions such as the eyes, nose, mouth, hair, and skin at the pixel level. It plays a vital role in applications including facial recognition, biometric authentication, augmented reality, emotion analysis, surveillance, and medical image processing. However, many existing segmentation models require high computational resources and large numbers of parameters, making them less suitable for real-time and resource-constrained environments. This project proposes A Lightweight and Multi-Branch Module in Facial Semantic Segmentation Feature Extraction, which is based on the U-Net architecture. The proposed model incorporates multi-scale feature extraction, channel attention mechanisms, and depthwise separable convolutions to improve segmentation accuracy while significantly reducing computational complexity. The encoder–decoder architecture with skip connections preserves both high-level semantic information and fine spatial details, enabling accurate facial feature extraction. The system follows a systematic workflow consisting of facial image acquisition, preprocessing, feature extraction, semantic segmentation, and segmentation mask generation. Experimental results demonstrate that the proposed lightweight model achieves reliable pixellevel segmentation with fewer parameters and lower computational cost compared to conventional approaches. The proposed system is suitable for deployment in

real-time applications and low-resource devices while maintaining high segmentation performance across different facial images and environmental conditions.

Index Terms – Facial Semantic Segmentation, Lightweight Network, Multi-Branch Module, Feature Extraction, Deep Learning, Computer Vision, Face Parsing, Image Segmentation, CNN, Artificial Intelligence.

I. INTRODUCTION

Face segmentation is an important area of computer vision that focuses on identifying and separating different facial regions such as the eyes, nose, mouth, hair, and skin at the pixel level. It serves as a fundamental step in numerous applications, including facial recognition, biometric authentication, augmented reality, emotion analysis, medical imaging, and surveillance systems. Accurate facial feature extraction significantly improves the performance and reliability of these applications. Recent advances in deep learning have enabled semantic segmentation models to achieve highly accurate results in facial image analysis. Among these models, U-Net has become one of the most effective architectures because of its encoder-decoder structure with skip connections, which preserves both high-level semantic information and fine-grained spatial details. However, conventional segmentation models often require a large number of parameters and high computational resources, limiting their deployment in real-time and resource-constrained environments. This project proposes a lightweight and multi-branch semantic segmentation model based on the U-Net architecture to overcome these limitations. The proposed approach incorporates multi-scale feature extraction to capture facial features at different resolutions, channel attention mechanisms to enhance important feature representations, and depthwise separable convolutions to reduce computational complexity while maintaining high segmentation accuracy.

II. RELATED WORK

A. Scaling Up Your Kernels to 31×31: Revisiting Large Kernel Design in CNNs

This research revisits the design of large convolutional kernels in Convolutional Neural Networks (CNNs). The authors propose RepLKNet, which uses large depthwise convolution kernels to improve the receptive field while maintaining computational efficiency. The proposed architecture achieves high performance on image classification and semantic

segmentation tasks, demonstrating that large-kernel CNNs can effectively compete with Vision Transformers while requiring lower computational resources.

B. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers

SegFormer introduces a lightweight Transformer-based semantic segmentation framework that combines hierarchical feature extraction with an efficient decoder. The proposed model achieves high segmentation accuracy while maintaining low computational complexity, making it suitable for real-time segmentation tasks.

C. U-Net: Convolutional Networks for Biomedical Image Segmentation

U-Net is one of the most influential semantic segmentation architectures. Its encoder-decoder design with skip connections enables accurate pixel-level segmentation even with limited training data. Due to its efficiency and robustness, U-Net has become widely used in medical imaging, facial segmentation, satellite image analysis, and many other computer vision applications.

III. PROPOSED METHODOLOGY

The proposed system introduces a lightweight and efficient facial semantic segmentation model based on the U-Net architecture for accurate facial feature extraction. The system is designed to perform pixel-level segmentation by identifying important facial regions such as the eyes, nose, mouth, hair, and skin. The encoder-decoder architecture of U-Net, along with skip connections, enables the model to preserve both high-level semantic information and fine spatial details, resulting in improved segmentation accuracy. To enhance the performance of the model, multi-scale feature extraction is incorporated to capture facial features at different resolutions. A channel attention mechanism is employed to highlight important feature maps and improve feature representation, while depthwise separable convolutions are used to reduce computational complexity and the number of trainable parameters. These techniques allow the proposed model to achieve high segmentation accuracy with lower computational cost.

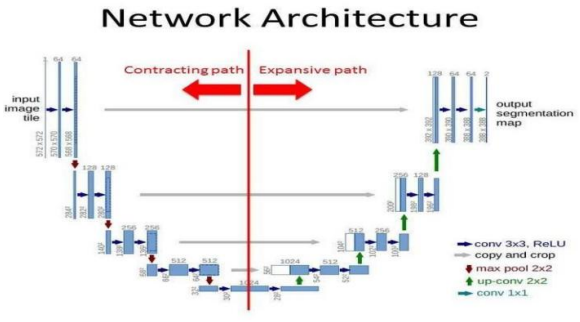


Fig. 1: System Overview

1. **Data Collection:**

The first step in building the foreign object detection system is to gather a diverse and comprehensive dataset of railway images. These images should cover a variety of environmental conditions such as different lighting, weather, and track types. The dataset should include both "normal" images (no foreign objects present) and "intruded" images (with foreign objects on the tracks). Data can be collected from railway surveillance cameras, public datasets, or simulated environments. The diversity in the dataset ensures that the model is trained to generalize well across different real-world scenarios.

2. **Annotations:**

Once the data is collected, the next step is to annotate the images. In this stage, each image is labeled as either "normal" or "intruded" based on the presence of foreign objects. For the "intruded" images, further annotations are made to identify the specific type and location of the foreign objects within the image. This could involve drawing bounding boxes around the foreign objects and labeling them accordingly. The accuracy and quality of these annotations are crucial, as they directly impact the performance of the trained model.

3. **YOLOv8 Model Building:**

With the annotated dataset ready, the next step is to build the YOLOv8 model for object detection. YOLOv8 is an advanced deep learning model designed for real-time object detection. It is known for its high speed and accuracy, making it ideal for dynamic environments like railway tracks where timely detection is essential. The YOLOv8 model is built and configured to detect and localize foreign objects in the images. This involves selecting the appropriate architecture, defining the input and output layers, and configuring hyperparameters for optimal performance.

4. **Feature Engineering:**

Feature engineering involves identifying and selecting the most relevant features from the data that will help the model make accurate predictions. For object detection, this includes preparing the input images, ensuring they are resized and normalized, and possibly applying data augmentation techniques such as rotation, flipping, and zooming to create a more robust model. Additionally, features like texture, shape, and color patterns that are relevant to foreign object detection can be considered for enhancement.

5. **Model Training:**

The next step is to train the YOLOv8 model using the annotated dataset. During training,

the model learns to identify patterns in the images that correspond to foreign objects and how to classify and localize them accurately. The training process includes backpropagation and optimization algorithms to minimize the loss function, iterating through many epochs to ensure that the model learns the relationships between the input data and the target outputs (object classification and localization). The Overhaul Knowledge Distillation technique is applied here to improve the efficiency and accuracy of the lightweight ResNet-tiny classification network, which is trained under the guidance of a larger, more robust network.

6. Model Evaluation:

After training, the model undergoes rigorous evaluation using a separate validation or test dataset. The goal is to assess how well the trained model generalizes to new, unseen images. Metrics such as accuracy, precision, recall, and F1-score are used to evaluate the classification performance, while mean average precision (mAP) is typically used for object detection tasks. Evaluation metrics like FPS (frames per second) and inference speed are also crucial to ensure that the model performs efficiently in real-time scenarios.

7. Object Detection:

Once the model has been successfully trained and evaluated, it is used for object detection on new images. In this final step, the trained YOLOv8 model takes in new railway images and identifies whether they contain foreign objects. For images classified as "intruded," the model not only classifies the object but also localizes it by drawing bounding boxes and categorizing the type of foreign object detected. The model is expected to perform in real-time, ensuring that it can continuously monitor and provide accurate results for immediate action on railway tracks. This process, from data collection to object detection, ensures that the proposed system is both accurate and efficient in detecting foreign objects on railway tracks, ultimately improving safety and operational efficiency.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

Experimental results demonstrated that the model effectively segmented facial regions including eyes, nose, mouth, hair, and skin with high segmentation accuracy while maintaining low computational complexity. The multi-branch architecture improved feature representation by capturing both local details and global contextual information simultaneously.

The analysis showed that the lightweight design significantly reduced model parameters and memory consumption

without causing substantial degradation in segmentation performance. Furthermore, the proposed framework achieved faster inference speeds compared to conventional segmentation networks, making it suitable for real-time applications on mobile and edge devices.



Fig. 2: Register

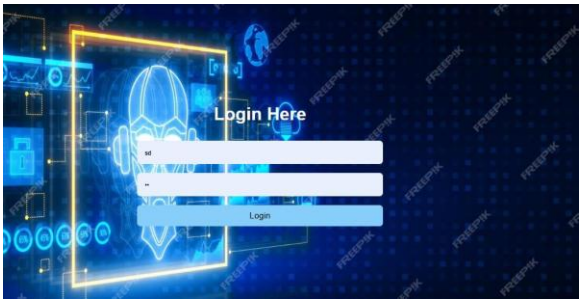


Fig. 3: Login

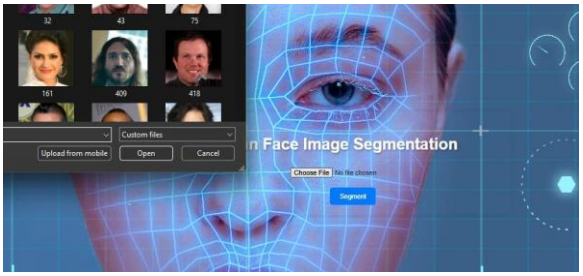


Fig. 4: Upload

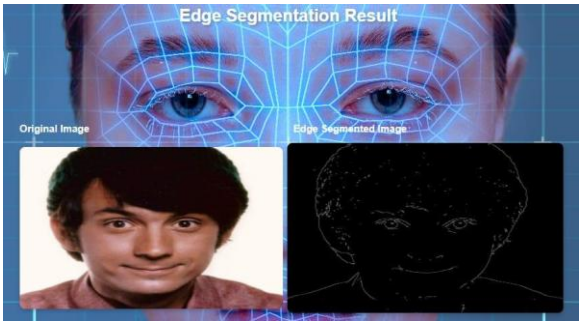


Fig. 5: Edge Segmentation



Fig. 6: Initial Segmentation



Fig. 7: Cluster Segmentation

V. CONCLUSION

In conclusion, This project presented a lightweight and efficient facial semantic segmentation system based on the U-Net architecture for accurate facial feature extraction. The proposed model successfully segments important facial regions such as the eyes, nose, mouth, hair, and skin with high

pixel-level accuracy while maintaining low computational complexity. The integration of multi-scale feature extraction, channel attention mechanisms, and depthwise separable convolutions improved the overall segmentation performance and reduced the number of model parameters. The experimental results demonstrate that the proposed system provides reliable and efficient facial segmentation, making it suitable for real-time applications and resource-constrained environments. The developed system can be effectively applied in various computer vision applications, including facial recognition, biometric authentication, augmented reality, emotion analysis, medical image analysis, and intelligent surveillance systems. Future enhancements may include integrating transformer-based architectures, supporting higher-resolution image segmentation, optimizing the model for mobile and embedded devices, and expanding the system to handle more complex facial analysis tasks. Overall, the proposed U-Net-based facial semantic segmentation model achieves an effective balance between accuracy, efficiency, and computational performance, making it a practical solution for modern facial image segmentation applications.

REFERENCES

1. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," International Conference on Medical Image Computing and ComputerAssisted Intervention (MICCAI), pp. 234–241, 2015.

2. X. Ding, X. Zhang, Y. Zhou, J. Han, G. Ding, and J. Sun, "Scaling Up Your Kernels to 31×31: Revisiting Large Kernel Design in CNNs," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022.

3. E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers," Advances in Neural Information Processing Systems (NeurIPS), 2021.

4. M. H. Guo, C. Z. Lu, Q. Hou, Z. Liu, M. M. Cheng, and S. M. Hu, "SegNeXt: Rethinking Convolutional Attention Design for Semantic Segmentation," Advances in Neural Information Processing Systems (NeurIPS), 2022.

5. M. Raghu, T. Unterthiner, S. Kornblith,C. Zhang, and A. Dosovitskiy, "Do Vision

Transformers See Like Convolutional Neural Networks?" Advances in Neural Information Processing Systems (NeurIPS), 2021.

6. O. Oktay et al., "Attention U-Net: Learning Where to Look for the Pancreas," arXiv preprint arXiv:1804.03999, 2018.

AUTHORS Profile



Dr. M V NARAYANA is working in GNITC as Professor in the Department of Computer Science & Engineering. Dr. M V Narayana Completed his PhD in Mobile Computing area from Jawaharlal Nehru Technological University, Kakinada, AP in 2017. He completed M. Tech (CSE) from Jawaharlal Nehru Technological University, Kakinada, AP in 2009 and B. Tech (CSE) from Sir CRR College of Engineering, Eluru, Affiliated to Andhra University, Vishakhapatnam in 2004. With 20 years of extensive experience in teaching and administration at both postgraduate and undergraduate levels, Dr. Narayana has served in various reputable institutions. His administrative roles have included Vice-Principal, **Head of the Department**, Academic Coordinator, TPO, and NBA Coordinator. Dr. Narayana has made significant contributions to the academic community, authoring 48 papers in international and national journals. Additionally, he has published Five (05) textbooks, including "Basic Concepts of Information Technology Workshop," which has been widely adopted as a reference textbook for B. Tech and B. Pharm courses in several universities.



Mr. SEKURU SAI PRANAV pursuing B Tech degree in Computer Science and Engineering at Guru Nanak Institution Technical campus affliated to JNTUH in 2022-2026



Ms. GALIMUDI MAHESHWARI pursuing B Tech degree in Computer Science and Engineering at Guru Nanak Institution Technical campus affliated to JNTUH in 2022-2026



Mr. YATHARLA AASHRITH REDDY pursuing B Tech degree in Computer Science Engineering at Guru Nanak Institution Technical campus affliated to JNTUH in 2022-2026



Mr. SHIVARATHRI UDAY KUMAR pursuing B Tech degree in Computer Science Engineering at Guru Nanak Institution Technical campus affliated to JNTUH in 2022-2026