Performance Evaluation of Road Marking Detection Models in Indian Conditions

Khushi Agrawal[1], Jafri Syed Mujtaba[2] and Samarth Bankar[3]
Ahmedabad University, Ahmedabad, Gujarat 380009, India
[1]khushi.a2@ahduni.edu.in, [2]jafri.h@ahduni.edu.in, [3]samarth.b2@ahduni.edu.in

***Abstract*—Road marker detection is essential for infrastructure monitoring, traffic control, and autonomous driving. Large-scale road marker identification can now be effectively done thanks to the growing availability of UAV (Unmanned Aerial Vehicle) footage. The purpose of this work is to use the AU-Drone dataset to assess the performance of many Indian road marking identification technologies, mostly segmentation-based techniques. Examining current publicly accessible models, determining their efficacy, and applying transfer learning to modify them for Indian road conditions are all part of the project. Using ground-truth data, an assessment system based on Python will be created to benchmark the performance of the model. The finished framework will be open source, offering a strong instrument for further study and real-world uses in autonomous navigation and urban planning.**

***Index Terms*—Road Marking Detection, UAV Imagery, Semantic Segmentation, Deep Learning, Transfer Learning, Indian Road Infrastructure, Autonomous Navigation, Model Evaluation Framework**

I. Introduction

Road markers are crucial visual aids that improve road safety and help autonomous cars and human drivers navigate their lanes. Ground-level cameras and LiDAR are frequently used in traditional road marker detection techniques, however, their scalability and affordability are limited. A viable substitute is aerial photography, which offers high-resolution aerial views of highways that can be processed by deep learning-based segmentation models.

The evaluation of current segmentation algorithms for the identification of road markers in the Indian setting is the main objective of this study. Indian roads offer particular difficulties due to their irregular signage, different lane lines, and environmental elements, including fading paint and occlusion. The search for publicly accessible models, the examination of their designs, and the comparison of their results using the AU-Drone dataset and IDD-20K Part I will be the first steps in the study process. Transfer learning will be used if necessary to modify these models for Indian road conditions. A Python-based system will be created that incorporates important performance indicators, including precision, recall, IoU (Intersection over Union), and F1 score, to guarantee objective evaluation. To confirm the prediction of the model, a manually labeled ground truth dataset will also be created. An open-source framework that makes it easier to detect road markings will be the end product, which will help with applications like traffic control, smart city planning, and autonomous car navigation.

II. Related Works

The extraction of road markings is a pivotal task in high-definition mapping for autonomous driving and transportation management. Given the challenges posed by variable illumination, occlusions, and the small size of many lane marking features, recent research has explored various deep learning approaches to improve segmentation accuracy.

Several studies have pursued model architectures that explicitly address the spatial and structural complexity of road markings. For example, Chen et al. [1] developed an Attentive Capsule Feature Pyramid Network (ACapsFPN) that integrates capsule networks with a feature pyramid structure and attention mechanisms. Unlike traditional convolutional neural networks (CNNs) that use scalar activations, capsule networks preserve orientation and pose information, thereby enhancing the detection of irregular patterns in road scenes.

In another line of research, Azimi et al. [2] introduced an architecture that combines a Fully Convolutional Neural Network (FCNN) with Discrete Wavelet Transform (DWT) for lane marking segmentation. The incorporation of DWT allows the model to retain high-frequency details that are critical for segmenting thin and faded markings, while a cost-sensitive loss function mitigates the impact of severe class imbalance inherent in aerial imagery.

More recent work by Zhang et al. [3] has focused on benchmarking a wide range of semantic segmentation models—both CNN-based and transformer-based—using transfer learning. Their comparative analysis, which includes models such as U-Net, DeepLabV3+, and SegFormer, demonstrates that transformer architectures excel in capturing long-range dependencies and contextual information, yielding superior performance metrics such as mean Intersection over Union (mIoU) and F1-score. However, the increased computational overhead of transformer models poses practical challenges for real-time applications.

III. Dataset Details

The effectiveness of segmentation models largely depends on the diversity and richness of training data. To ensure robust model evaluation in Indian road conditions, we utilize two publicly available datasets:

A. AU-Drone Dataset

The AU-Drone dataset is an annotated collection of aerial images captured using Unmanned Aerial Vehicles (UAVs), providing high-resolution imagery with lane markings, traffic signs, pedestrian crossings, and road boundaries. It is specifically designed for road scene analysis and marker detection.

- **Total Images:** 14,021 images
- **Resolution:** 1920×1080 pixels
- **Annotation Format:** Pixel-wise semantic masks
- **Classes:** Lane markings, crosswalks, stop lines, arrows, zebra crossings
- **Geographical Region:** Indian urban and suburban areas
- **Use Cases:** Road infrastructure mapping, autonomous driving applications

B. IDD 20K Part I Dataset

The Indian Driving Dataset (IDD) Part I is a large-scale dataset with over 20,000 annotated images of urban and rural Indian roads. This dataset captures the unique challenges of Indian traffic, including diverse weather conditions, crowded lanes, and unstructured road markings.

- **Total Images:** 20,000 images
- **Resolution:** Variable (ranges from 1280×720 to 1920×1080 pixels)
- **Annotation Format:** Pixel-wise segmentation masks with 20 semantic classes
- **Classes:** Lane markings, vehicles, pedestrians, traffic signs, road boundaries, and others
- **Geographical Region:** Urban and rural regions of India
- **Use Cases:** Semantic segmentation, autonomous driving, and lane detection models

C. Comparison of Datasets

A side-by-side comparison of key features between the AU-Drone and IDD 20K datasets is presented in Table I.

TABLE I
Comparison of AU-Drone and IDD 20K Datasets

Dataset Attribute	**AU-Drone Dataset**	**IDD 20K Part I Dataset**
Total Images	14,021	20,000
Resolution	1920×1080 pixels	Variable (1280×720 to 1920×1080 pixels)
Number of Classes	6	20
Annotation Type	Pixel-wise semantic segmentation	Pixel-wise semantic segmentation
Geographical Region	Indian Urban/Suburban	Indian Urban/Rural
Application	UAV-based Road Monitoring	Autonomous Driving and Lane Detection
Use Cases	Infrastructure Mapping, Smart City Analysis	Object Detection, Lane Boundary Analysis

D. Preprocessing Strategy

For consistency during model training and evaluation, the following preprocessing steps are applied uniformly across both datasets:

- **Resizing:** Standardizing all images to 512×512 dimensions.
- **Normalization:** Scaling pixel intensity values to the range $[0, 1]$.
- **Data Augmentation:** Applying random horizontal flips, rotations, and brightness adjustments to enhance training diversity.

E. Ground Truth Annotation and Quality Control

The annotation masks in both datasets are manually labeled and validated to ensure precision in ground truth generation. Each semantic class is assigned a unique class ID, which is used for model benchmarking and evaluation.

IV. Methodology

A. Dataset Preparation and Preprocessing

The datasets used for model evaluation include:

- **AU-Drone Dataset:** A high-resolution aerial dataset containing road marking images captured by UAVs. The dataset includes annotated masks for lane boundaries, road signs, and pedestrian markings.
- **IDD 20K Part I Dataset:** Indian Driving Dataset comprising over 20,000 annotated urban and rural road images with 20 semantic classes, including lane markings, vehicles, traffic signs, and pedestrians.

Preprocessing Steps:

1) **Image Resizing:** Resize images and corresponding masks to a consistent size, typically (512×512) for model input.
2) **Normalization:** Pixel values normalized to $[0, 1]$ to enhance model convergence.
3) **Data Augmentation:** Random horizontal flipping, rotation, and brightness adjustments to introduce variance in training samples.

B. Model Selection and Transfer Learning

To ensure optimal performance, pre-trained segmentation models such as U-Net, DeepLabV3+, and SegFormer are evaluated and fine-tuned using transfer learning. The key advantage of transfer learning lies in leveraging pre-trained feature extractors trained on large-scale datasets like ImageNet.

Transfer Learning Formulation:

$$\hat{y} = f(W_{pre}, X) \tag{1}$$

where:

- W_{pre} denotes the pre-trained model weights.
- X represents the input image tensor.
- $\hat{y}$ is the predicted segmentation mask.

Fine-tuning is performed by freezing the initial layers of the model and retraining the final layers with a reduced learning rate:

$$W^* = \text{argmin}_W \frac{1}{N} \sum_{i=1}^{N} L(f(W, X_i), y_i) \tag{2}$$

where:

- L denotes the loss function (e.g., Cross-Entropy or Dice Loss).
- N is the total number of training samples.

C. Segmentation Model Architectures

1) U-Net: U-Net uses a symmetric encoder-decoder structure with skip connections to preserve spatial information.

$$f(x) = \sigma(W_d * g(W_e * x) + b) \tag{3}$$

2) DeepLabV3+: DeepLabV3+ leverages Atrous Spatial Pyramid Pooling (ASPP) to capture multi-scale context.

$$f(x) = \sigma \left(\sum_{r \in R} W_r * x \right) \tag{4}$$

D. Performance Evaluation Metrics

Model evaluation is conducted using the following key metrics:

1) Intersection over Union (IoU):

$$\text{IoU} = \frac{|P \cap G|}{|P \cup G|} \tag{5}$$

2) Mean IoU (mIoU):

$$\text{mIoU} = \frac{1}{N} \sum_{i=1}^{N} \text{IoU}_i \tag{6}$$

3) Dice Coefficient:

$$\text{Dice} = \frac{2|P \cap G|}{|P| + |G|} \tag{7}$$

4) F1-Score:

$$\text{F1} = 2 \cdot \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{8}$$

E. Ground Truth Generation for Benchmarking

Manually labeled ground-truth masks are generated for the AU-Drone dataset and IDD-20K using the CVAT (Computer Vision Annotation Tool). Each mask is encoded using class IDs from 0 to 19 corresponding to segmentation classes.

F. Python-Based Evaluation Framework

A Python-based evaluation framework is designed to:

- Preprocess and normalize input images.
- Apply trained segmentation models to generate predicted masks.
- Calculate IoU, mIoU, and F1-score for performance evaluation.
- Generate detailed performance reports for comparative analysis.

The framework is modular, supporting multiple model architectures and batch processing of test images.

V. Conclusion and Future Work

This study proposes a systematic evaluation of segmentation models applied to Indian road marking detection using aerial imagery from the AU-Drone and IDD 20K datasets. By leveraging pre-trained models and fine-tuning them with transfer learning, we aim to optimize segmentation accuracy while addressing class imbalances, occlusions, and environmental variations. Future work will involve extending the evaluation framework to include additional models and integrating real-time deployment capabilities for smart city applications.

Acknowledgments

References

[1] H. Guan, X. Lei, Y. Yu, H. Zhao, D. Peng, J. M. Junior, and J. Li, “Road marking extraction in uav imagery using attentive capsule feature pyramid network,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 107, p. 102677, 2022.

[2] S. M. Azimi, P. Fischer, M. Körner, and P. Reinartz, “Aerial lanenet: Lane-marking semantic segmentation in aerial imagery using wavelet-enhanced cost-sensitive symmetric fully convolutional neural networks,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 5, pp. 2920–2938, 2018.

[3] W. Liu, S. Qiao, K. Gao, H. He, M. A. Chapman, L. Xu, and J. Li, “Advancements in road lane mapping: Comparative fine-tuning analysis of deep learning-based semantic segmentation methods using aerial imagery,” *arXiv preprint arXiv:2410.05717*, 2024.