Road Markings Detection and Measurement in Aerial Imagery

Kadiwala Adnan
School of Engineering & Applied Science
Computer Science & Engineering
Ahmedabad, India
adnan.k@ahduni.edu.in

Shah Vandit
School of Engineering & Applied Science
Computer Science & Engineering
Ahmedabad, India
vandit.s@ahduni.edu.in

Hingu Dhruv
School of Engineering & Applied Science
Computer Science & Engineering
Ahmedabad, India
dhruv.h@ahduni.edu.in

Patel Het
School of Engineering & Applied Science
Computer Science & Engineering
Ahmedabad, India
het.p5@ahduni.edu.in

***Abstract*—Road markings detection and measurement in aerial imagery is a crucial task for enhancing road safety and supporting autonomous driving technologies. In this work, we explore the use of computer vision techniques, specifically leveraging the U-Net architecture, to segment and classify road markings from AU drone dataset images. We present a comprehensive methodology that includes dataset preparation, model selection, training, and evaluation. Our approach involves converting segmentation masks into one-hot encoded formats for training, utilizing categorical loss for multiclass segmentation, and evaluating model performance using Intersection over Union (IoU). The trained model achieves an impressive accuracy of 98.51% and an IoU coefficient of 0.93, demonstrating its efficacy in accurately localizing road markings. Post-segmentation, we conduct contour analysis and dimension calculation to determine the type, size, and location of road markings. Our findings highlight the potential of computer vision and deep learning in advancing road safety and transportation systems.**

***Index Terms*—Computer Vision, U-Net, Aerial Imagery, Road Markings Detection, Semantic Segmentation, Multiclass Segmentation, Intersection over Union (IoU), One Hot Encoding, Region of Interest (ROI) Detection.**

I. Introduction

Detecting road markings using segmentation approaches involves employing computer vision techniques to identify and delineate road markings from images or videos captured by AU drone dataset. Once these markings are identified, converting the detected pixels into real-world measurements like centimeters involves calibrating the camera parameters like height from sea-level and latitude and longitude which will give us distance between the camera and the road surface. This calibration allows for mapping the detected pixels to actual physical dimensions on the road, facilitating precise localization of road markings.

In simple terms, imagine you're driving on the road, and you notice those painted lines that separate lanes or indicate where you should stop at an intersection. The problem is, for both human drivers and upcoming self-driving cars, it's crucial to not just see these markings but to understand them accurately. By using smart technology that can detect and measure these markings pixel by pixel, we're essentially teaching our cars to "see" the road just like we do. This isn't just about making sure cars stay in their lanes; it's about making driving safer and smoother for everyone, whether we're navigating through busy city streets or letting our cars take the wheel on long road trips.

II. Methodology

A. Dataset

We used AU's Drone Dataset for our project, which was provided to us by Ph.D. scholar Yagnik Bhavsar at Ahmedabad University.
The data contains around 1500 high-quality images along with the ground truth segmented masks of road markings for each image. The dataset included various marking schemas like zebra crossings, Direction marks, lane separators, and lane markings. The images and masks exhibit robustness, capturing occlusions like trees and vehicles within the scenes. Here is an illustrative example from the dataset.

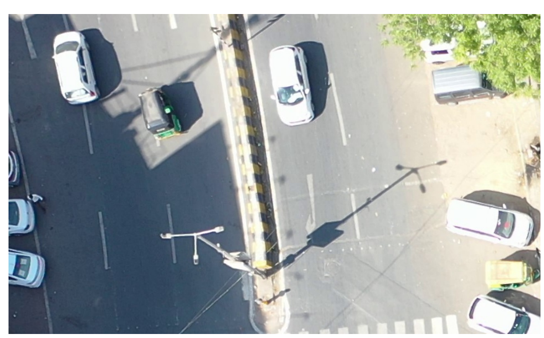

Fig. 1. Original image

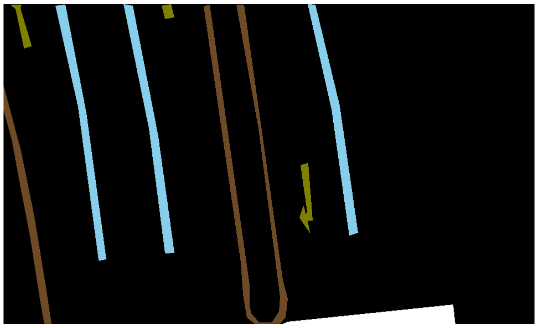

Fig. 2. Segmentation Masks

For our use case, we split the data into training and testing sets, with 1250 images allocated for training and 250 images for testing purposes. The data quality was satisfactory, requiring minimal cleaning or preprocessing.

The segmentation masks demonstrate these classes:
* **White masks**: zebra crossing.
* **Brown masks**: lane-edge markings.
* **Blue masks**: lane markings.
* **Orange masks**: in-lane markings.
* **Black masks**: everything else.

As we needed to perform multiclass segmentation, we converted the provided masks into one-hot encoded masks for each class to facilitate better and faster training.

B. Deep Learning Model Selection and Training

We chose the 'U-Net architecture' [1] for our project due to the availability of ground truth masks accompanying each image in the dataset. These masks provided essential segmentation annotations required for training our model and achieving our goal of multiclass segmentation accurately. Also, the U-Net architecture is well suited for multiclass segmentation, including tasks like road marking detection (segmentation), and medical imagining.

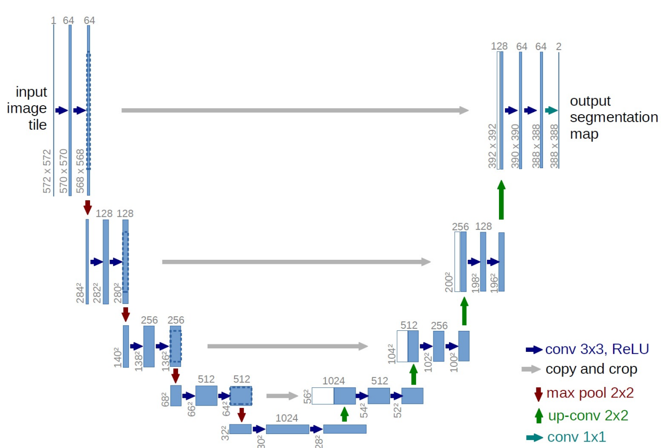

Fig. 3. U-Net Architecture

U-Net has an **encoder-decoder** type of architecture. During the **encoder** phase of the U-Net, the images go through a series of convolutional as well as pooling layers which leads to downsampling of the images due to which the spatial information is reduced while it extracts the feature from the image and leading in increase of essential feature information. This downsampling of the image helps in capturing the low-level features (such as edges) to the high-level features such as shapes and patterns, which eventually helps in finding different types of road marking and provides accurate results.

After the input image completes the encoder phase, it passes throug the **decoder** phase that involves upsampling the features that were learnt back to the spatial region of the input image. This process helps in retaining the spatial information and features in the input image and ensures that the segmented region accurately maps the original input image structure. This spatial information gained helps in the localization and segmentation of road marking within and helps detect lane marking, zebra crossing, in-lane marking and background.

U-Net not only provides segmented output semantically meaningful but also provides spatially accurate results. This is quite important in the particular task as the precise localization of the road marking is essential for its further application like autonomous car and traffic management.

Now as our problem is on **multiclass segmentation** using U-net, it required the one-hot encoding of the masks correspoding to its road marking label. During this one-hot encoding, it changes the categorical data (different classes of the road marking) into a binary vector where each road marking is given a binary value (0 or 1). This method eventually helps in providing better ground truth to the model and helps in identifying and differentiating between classes during training. During the training, we used **categorical loss** for the multiclass segmentation and provided the model with the ability to classify and segment different road markings.

The categorical loss (categorical cross entropy) calculates the dissimilarities between the predicted segmentation mask and the ground truth for each road marking class and eventually helps the model to detect and segment the different road markings with minimal loss. By employing both methods, the model becomes quite robust and accurate and can classify road markings like zebra crossings, lane-edge markings, lane markings, and in-lane markings and could also handle complex input where the model could find multiple road markings within a single image.

For evaluating the multiclass segmentation model performance, we utilized **Intersection over Union** (IoU). This metric measures the overlap between predicted masks and ground truth masks for each road marking class, indicating segmentation accuracy and spatial alignment. Implementing early stopping during training prevented overfitting and ensured an optimal model for road marking detection. Early stopping monitored IoU on the validation dataset; training halted when performance ceased to improve or declined, promoting generalization on new images. Using IoU as an evaluation metric quantified segmentation quality, assessing the model's ability to classify zebra crossings, lane-edge markings, and other road markings. Iteratively improving the model based on IoU scores and early stopping criteria enhanced accuracy, robustness, and suitability for real-world applications. This approach yielded optimal results for multiclass segmentation of aerial imagery in road marking detection.

We trained our model with careful consideration of all aspects discussed, for about 50 epochs. The model's performance was evaluated using test and validation datasets, resulting in an **accuracy of 98.51%** and an **IoU coefficient of 0.93.**

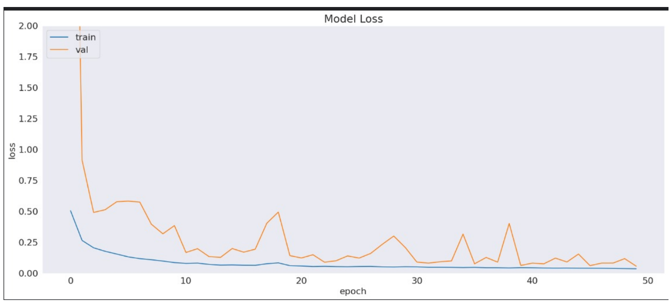

Fig. 4. Model Loss Curve vs. Epoch

Figure 4 shows the model's loss curve plotted against epochs for both test and validation data.

C. Prediction by the Model

The model provided us with one-hot encoded masks for each example, as illustrated below. The next step involves assigning Marking Labels to the class labels obtained from the U-Net output based on specific criteria.

The received class labels are initially in a one-hot encoded format, representing different categories such as "Lane marking," "Lane Edge," "Zebra Crossing," and "Traffic sign." Each class label corresponds to a specific RGB color according to the following mapping:

- "Lane marking": (0, 0, 255) (Blue)
- "Lane Edge": (0, 255, 0) (Green)
- "Zebra Crossing": (255, 255, 255) (White)
- "Traffic sign": (255, 0, 0) (Red)

This color mapping process is crucial for facilitating visual differentiation among various road marking categories within the segmentation output. It allows for easier interpretation and analysis of the segmentation results.

To ensure accurate alignment of the segmented masks with the original image dimensions, a resize function is applied to the color-encoded mask. This resizing step is essential for

Fig. 5. Model's Prediction

precise visualization and overlay of detected road markings onto the original scene.

Subsequently, detailed contour analysis and dimension calculation are performed using OpenCV functionalities. Each contour is examined to compute its bounding box, ensuring maximum overlap (IoU) with the contour. This will give a bounding box to detect our Region of Interest.

Fig. 6. Mark Detection

The bounding box extracted from each contour is utilized to determine its dimensions (width and height) and to identify the dominant colour within the contour region.

For colour analysis of each contour, the mean colour within the contour region is computed. A **colour-based classification** is then conducted using a predefined set of reference colours as mentioned earlier. This classification assists in identifying the type of road marking based on the closest match to the predefined colours, such as lane markings, zebra crossings, traffic signs, and lane edge markings.

Finally, the annotated visualization of the original image is generated. Bounding boxes are drawn around the detected contours, and text labels indicating the type of road marking are overlaid within these boxes. This annotated visualization offers a clear depiction of the detected road markings along with their respective types, dimensions, and colours.

III. Final Output

Fig. 7. Mark Detection and Classification

Along with the detection it gives the details on each marking as follows:

- Mark 1: ’Zebra Crossing’
  - Width = 116.00px
  - Height = 44.00px
- Mark 2: ’Zebra Crossing’
  - Width = 614.21px
  - Height = 117.52px
- Mark 3: ’Lane Edge’
  - Width = 17.12px
  - Height = 142.90px
- Mark 4: ’Lane marking’
  - Width = 14.00px
  - Height = 47.00px
- Mark 5: ’Lane Edge’
  - Width = 179.40px
  - Height = 37.05px

IV. Conclusion

In conclusion of our journey into computer vision and neural networks for detecting road markings and translating them into real-world measurements, we’ve uncovered some amazing discoveries. We’ve seen how technology can help us better understand roads, making them safer and traffic smoother. Along the way, we’ve learned how artificial intelligence, like segmentation, plays a big role in this. From making traffic flow better to setting the stage for self-driving cars, this adventure has shown us how powerful technology can be in making our world a better place. Looking forward, the lessons we’ve learned will keep shaping how we travel and innovate, making transportation even more efficient and safe.

V. References

[1] Ronneberger, O., Fischer, P., Brox, T. (2015, May 18). U-NET: Convolutional Networks for Biomedical Image Segmentation. arXiv.org. https://arxiv.org/abs/1505.04597

[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,” in IEEE Transactions on Geoscience and Remote Sensing, vol. 57, no. 5, pp. 2920-2938, May 2019, doi: 10.1109/TGRS.2018.2878510.

[3] Bhavsar, Y. M., Zaveri, M., Raval, M. S., Zaveri, S. B. (2023). Vision-based Investigation of Road Traffic and Violations at Urban Roundabout in India using UAV Video: A Case Study. Transportation Engineering (Oxford), 14, 100207. https://doi.org/10.1016/j.treng.2023.100207