Share Comment Star

Autonomous Driving Segmentation Report

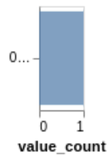
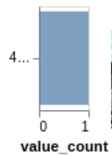
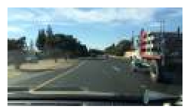
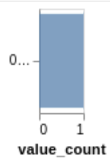
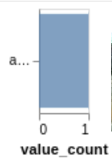
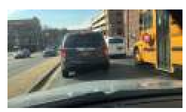
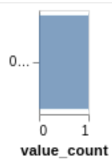
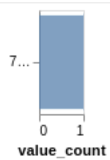
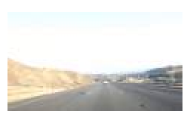
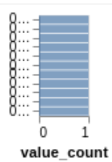
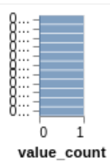
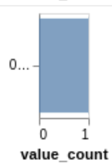
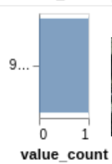
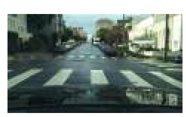
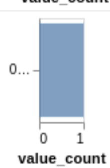
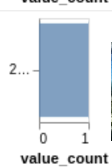
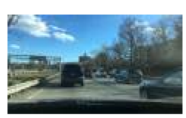
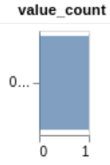
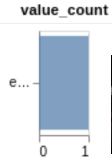
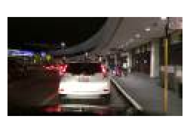
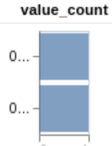
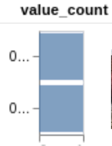
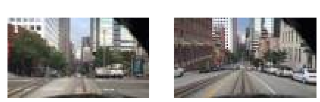
Amr Farghaly

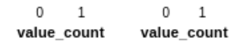
Created on October 2 | Last edited on October 10

1. Dataset Exploration and Analysis

We explore the dataset by analyzing key attributes such as P1, P2, Images, and Classes Represented. Grouping the data by these attributes helps us identify imbalances, similarities, and potential data leakage, ensuring a balanced dataset split for effective training.

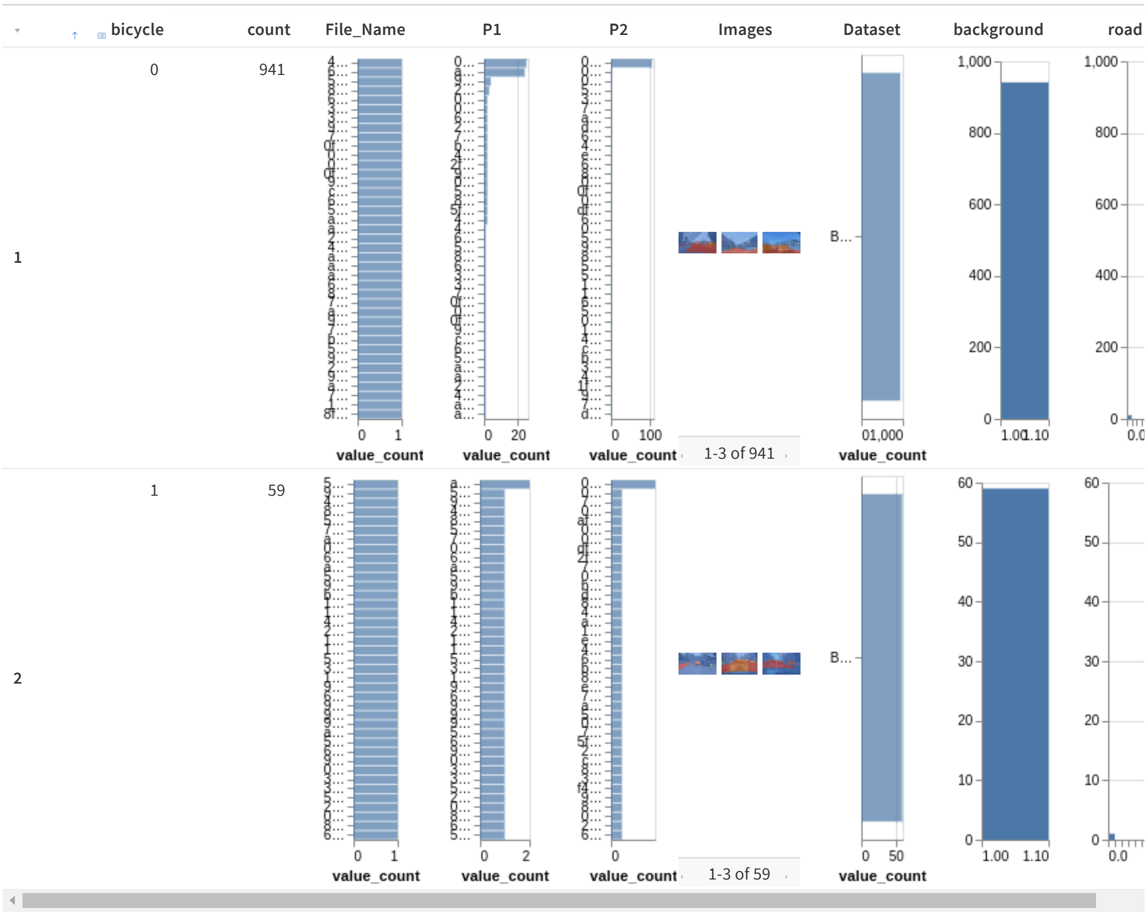
```
project("amribrahim-amer-2024", "Semantic-Segmentation-Model-for-Autonomous-Vehicle").artifact("dataset").membershipForAlias("v0").artifactVersion.file("Samples_data.table.json"
)
```

	P1	count	File_Name	P2	Images
1	01b29c03	1			
2	01ab0f6d	1			
3	016e76c0	1			
4	00e9be89	25			 1-5 of 25
5	00e69ee0	1			
6	00d79c0a	1			
7	00aad4a0	1			
8	0027eed2	2			



Images with the same P1 attribute tend to be similar or appear to be consecutive frames from the same video. This should be taken into account when splitting the dataset into training, validation, and testing sets to avoid data leakage.

```
project("amribrahim-amer-2024", "Semantic-Segmentation-Model-for-Autonomous-Vehicle").artifact("dataset").membershipForAlias("v0").artifactVersion.file("Samples_data.table.json"
)
```

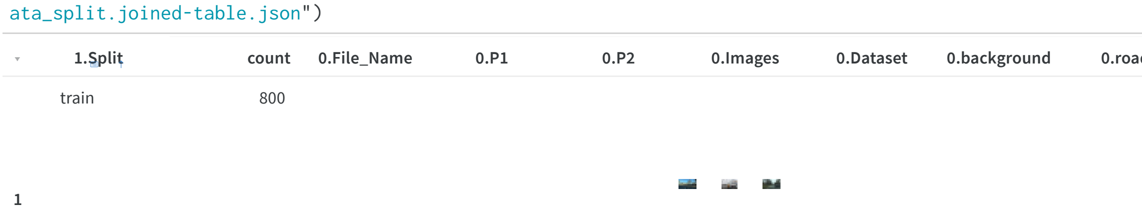


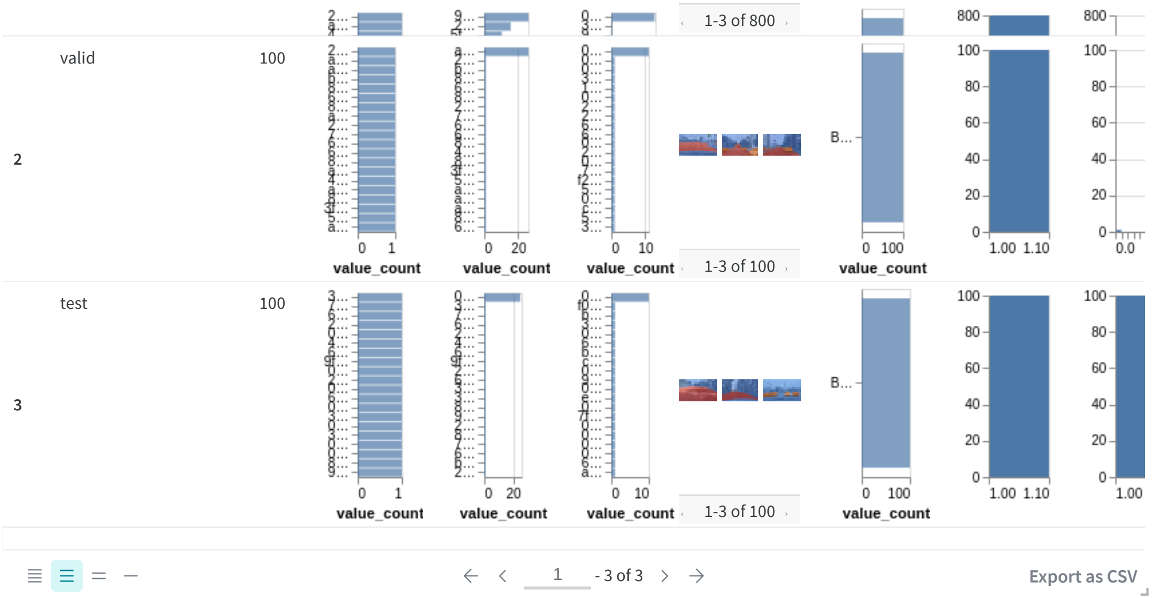
The dataset contains only 59 images with bicycles, and they are often small or noisy in appearance. This may make it challenging for the model to learn and accurately detect this class.

2. Data Preparation

We prepare the dataset for training by carefully splitting it into training, validation, and test sets. Our focus is on ensuring a balanced distribution of classes and avoiding any potential data leakage by grouping images based on their video identifier (P1). This step is critical to ensure the model generalizes well and performs effectively on unseen data.

```
project("amribrahim-amer-2024", "Semantic-Segmentation-Model-for-Autonomous-Vehicle").artifact("dataset_split").membershipForAlias("latest").artifactVersion.file("sample_table_data_split.joined-table.json")
```





After applying stratification, we can observe from the table and histogram that the rare 'bicycle' class is now more evenly distributed across the training, validation, and test splits. This ensures that the model will have enough representation of this class in each split, addressing the initial imbalance issue.

Additionally, we ensured that images with the same P1 attribute, which represent frames from the same video, are grouped together in the same split. This prevents data leakage and helps the model generalize better, as it won't be exposed to similar images from the same video during both training and testing phases.

3. Model Training

We trained a baseline model for semantic segmentation using the fastai framework. The model is a U-Net architecture with a ResNet-18 backbone, pretrained on ImageNet to take advantage of transfer learning, allowing for faster convergence and better performance. The training was performed on an 80%-10%-10% split of the dataset for training, validation, and testing, respectively. We ensured that the dataset was stratified to handle class imbalance (particularly the rare "bicycle" class) and avoided data leakage by grouping images based on video frames.

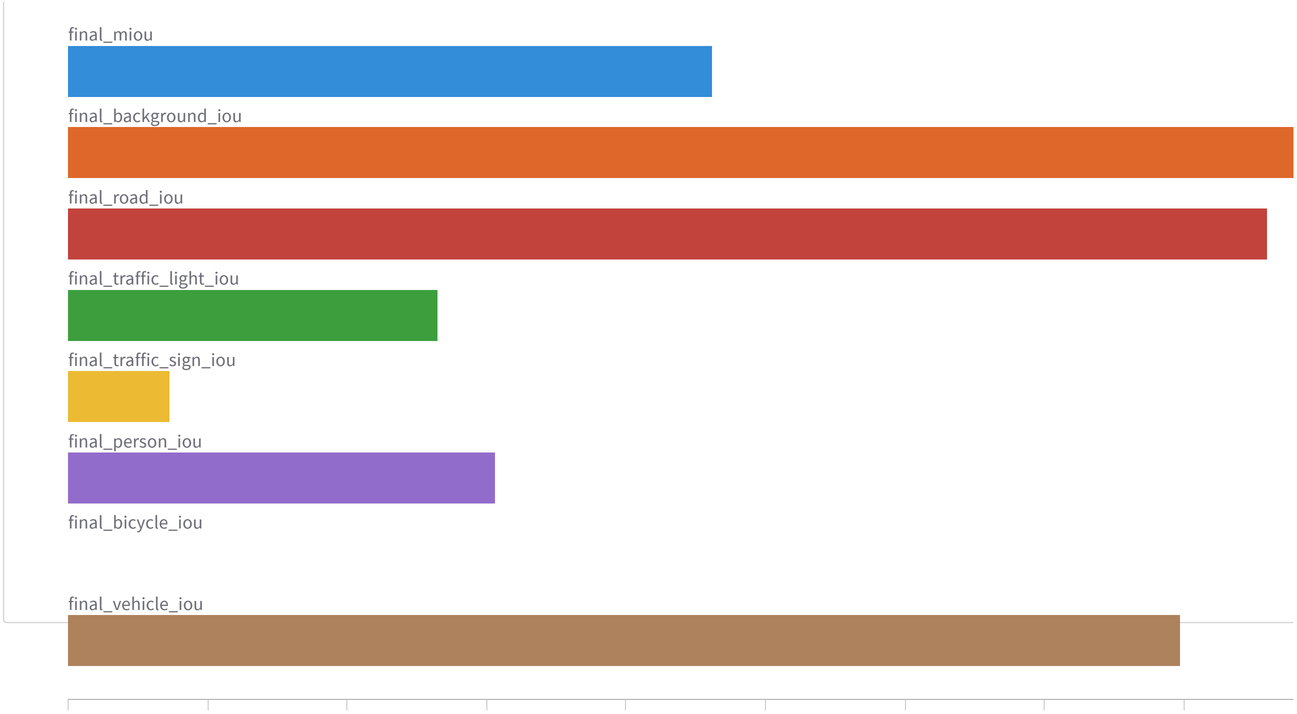
Key hyperparameters used during training included:

- Image size: 180x320 pixels
- Batch size: 8
- Number of epochs: 50
- Intial learning rate: 2e-3
- Data augmentation: Enabled to improve model generalization

To monitor performance, we logged metrics such as Mean Intersection over Union (MIOU) and class-specific IoUs (for 'background', 'road', 'traffic light', 'person', 'vehicle', etc.) using Weights & Biases (W&B). The best model was saved based on MIOU performance, and we also performed error analysis by logging model predictions and ground truth samples in W&B. After training, we reloaded the best model checkpoint for final validation and summary reporting.

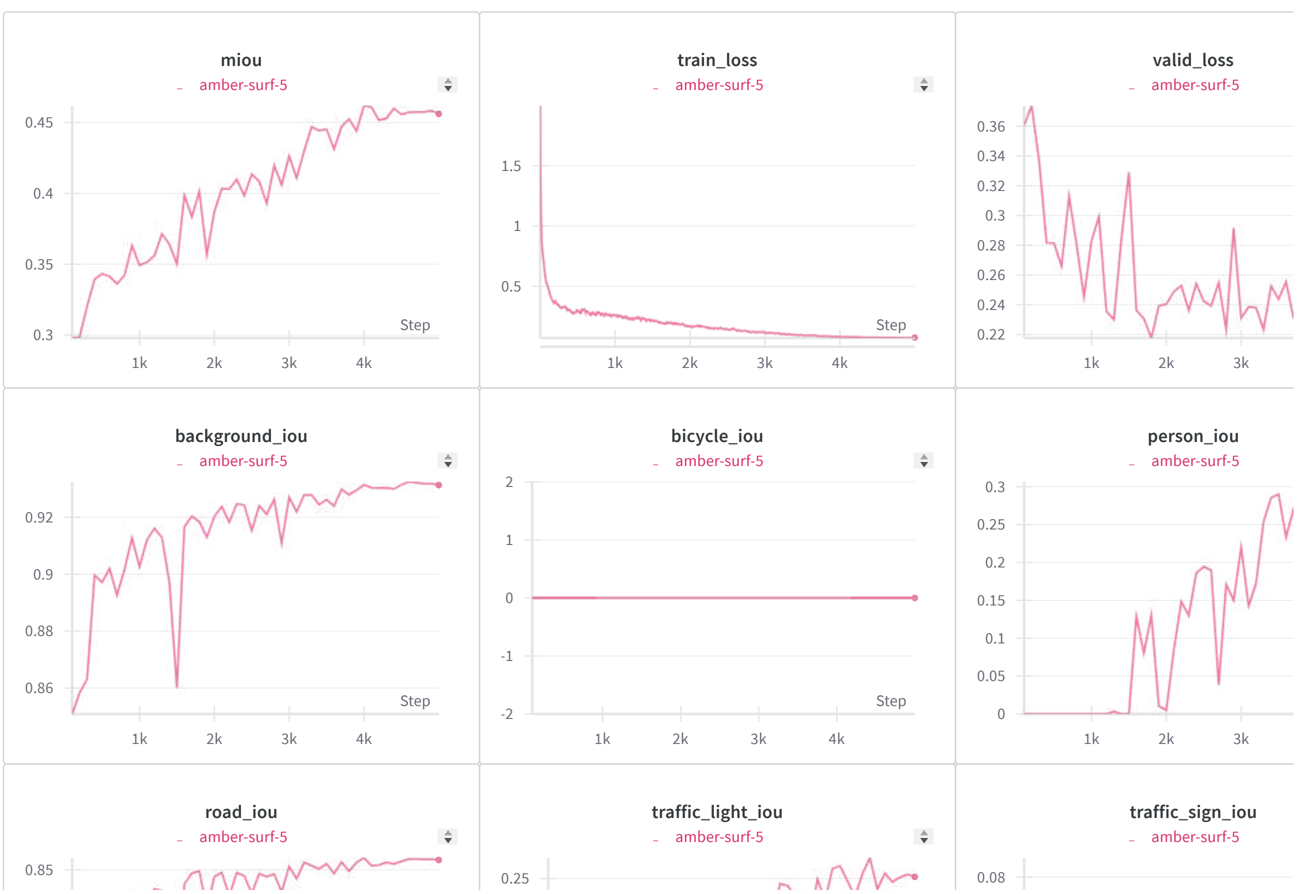
The bar chart below highlights the key metrics. It shows that common classes like 'background,' 'road,' and 'vehicle' are easier for the model to learn, while rarer classes or those with fewer pixels, such as the 'bicycle' class, are more challenging for the model to capture accurately.

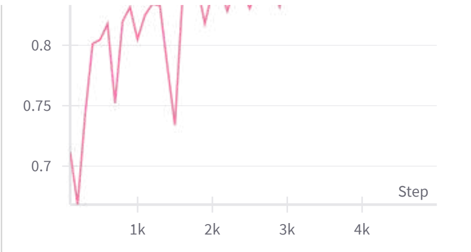
Final Metrics IOU Bar charts

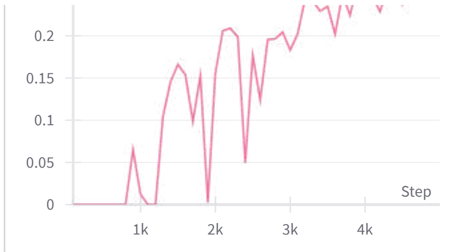


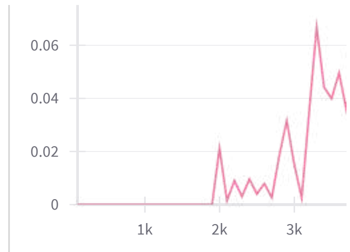
Throughout the training process, several key metrics were logged, including mean Intersection over Union (mIoU), individual IoU for each class, as well as training and validation losses. A consistent trend was observed where both training and validation losses decreased as training progressed, signifying that the model was learning effectively. Additionally, the mIoU increased, reflecting improved performance in segmenting most classes.

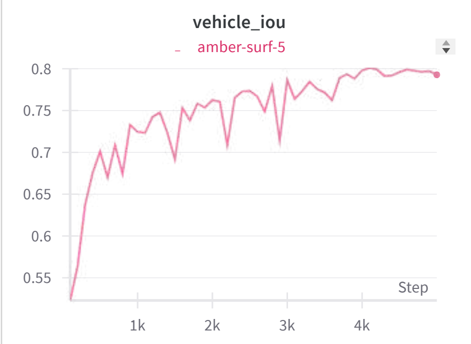
However, despite the overall improvement, the bicycle class presented challenges, with its IoU remaining at zero throughout training, indicating that the model struggled to learn this class. In contrast, classes with more frequent or larger representations in the dataset, such as background and road, showed notable improvements in IoU as the model learned more effectively from those examples. This discrepancy highlights the difficulty the model faces when learning from rare or small-object classes.





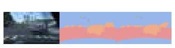
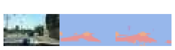
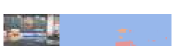






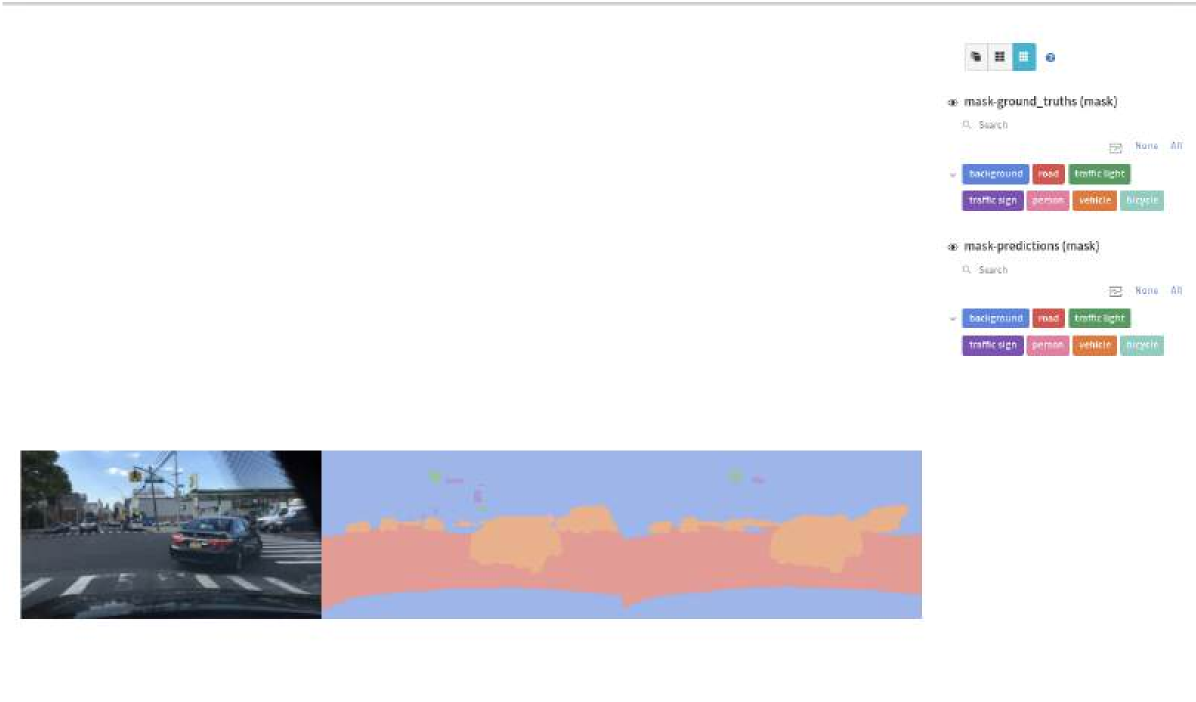
Below is the prediction table, showing segmented images along with the corresponding IOU metrics for each class.

```
runs.summary["pred_table"]
```

	Image	background IoU	road IoU	traffic light IoU	traffic sign IoU	person IoU	vehicle IoU
11		0.9502	0.7677	-	0	-	0.4483
12		0.9421	0.9405	0.4336	0.09091	0.482	0.779
13		0.9494	0.8992	0	-	-	0.7105
14		0.9517	0.9167	0.6275	0	0.7071	0.7791
15		0.9501	0.9493	0.6743	0.2242	0	0.8611
16		0.9826	0.9743	-	0	0.1429	0.8273
17		0.9304	0.6911	0	0	0	0.5966
18		0.9285	0	-	0	-	0
19		0.9383	0.8259	0	0	0	0.8446
20		0.8309	0.9145	0	0	0	0.8098

11 - 20 of 100 Export as CSV Columns...

In the prediction example below, we compare the ground truth mask (middle) with the predicted mask (right) for the test image (left). We can see that the model's prediction is close to the ground truth.

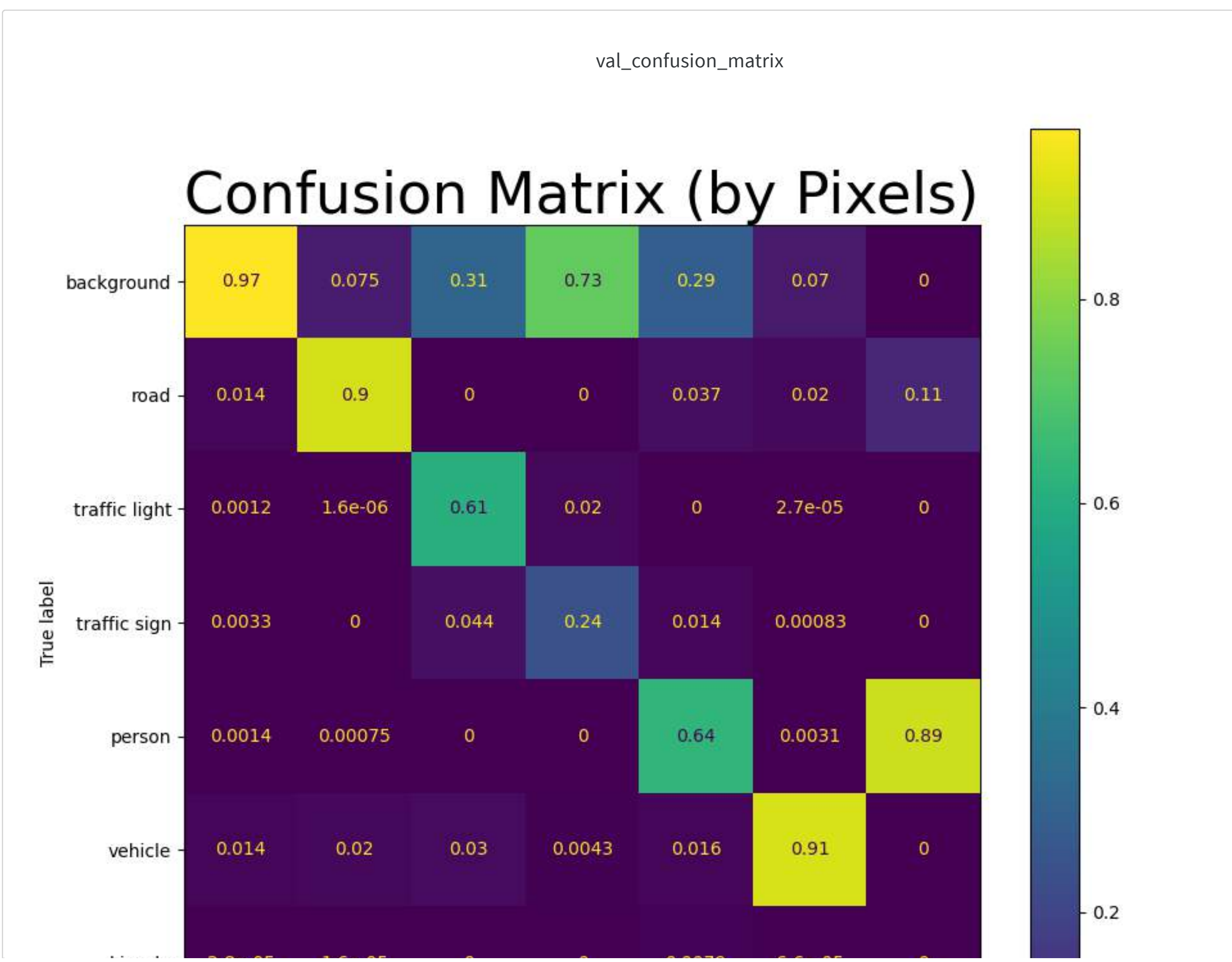


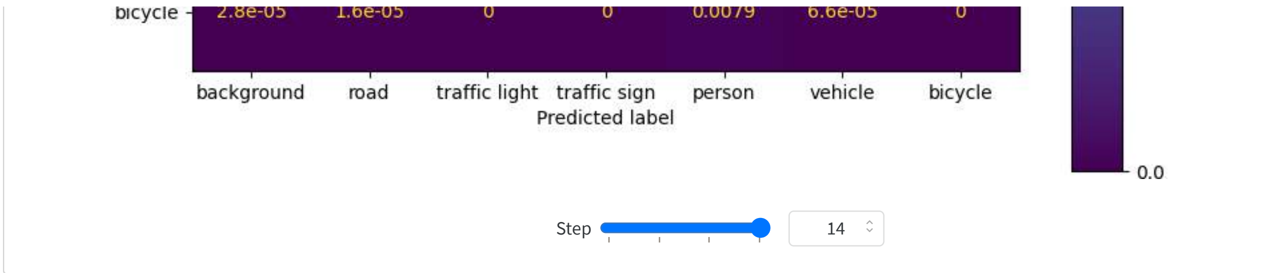
4. Model Evaluation and Analysis

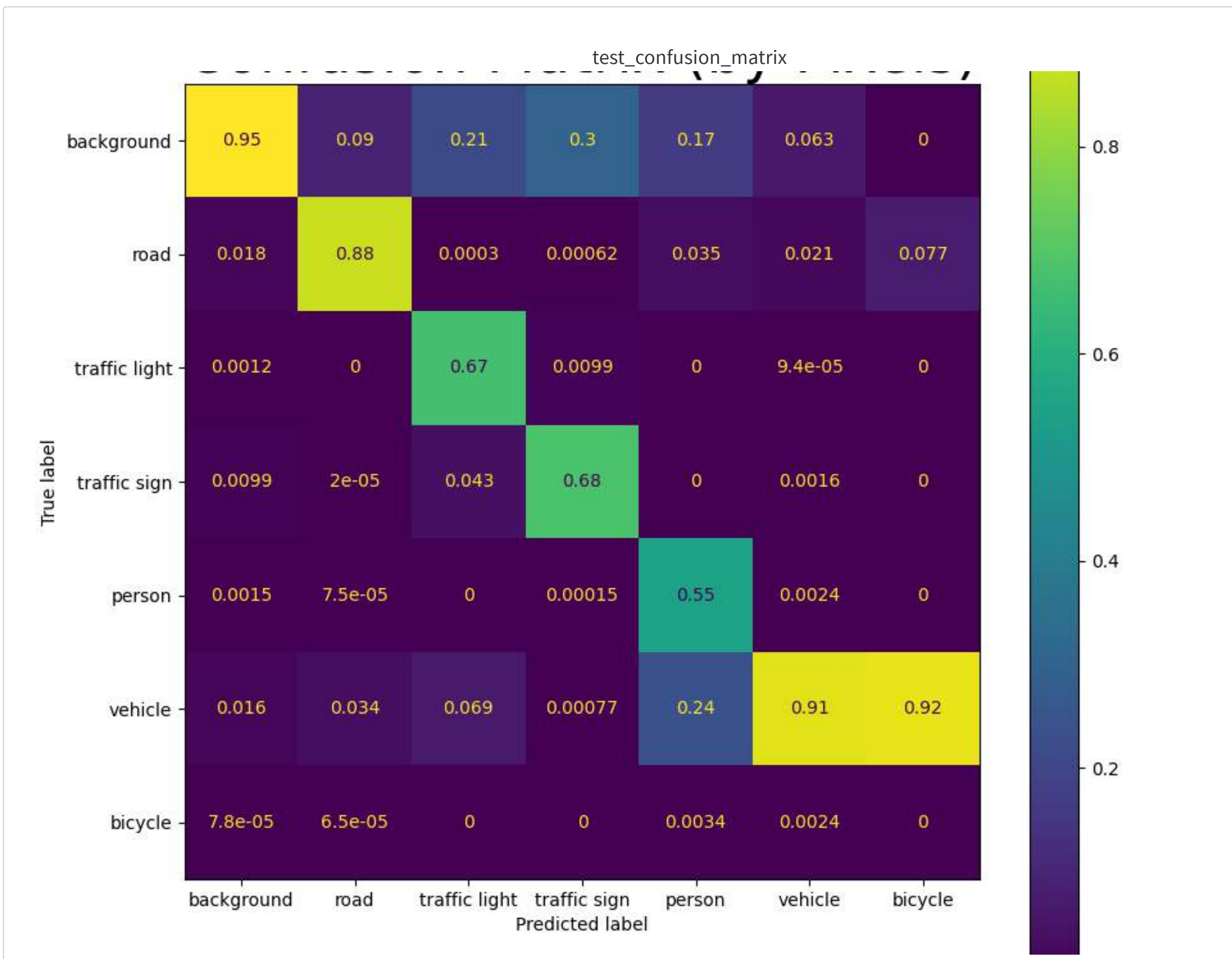
In this section, we present the evaluation results of our segmentation model, focusing on the confusion matrices - histograms - losses & metrics and tables obtained from both the validation and test datasets.

Confusion Matrix:

The confusion matrices for both the validation and test datasets are presented below.







Key Findings and Analysis

- **Background and Road:** The model excels in identifying the 'background' and 'road' classes, achieving impressive accuracy rates of 97% and 90% in the validation set, respectively. This strong performance is consistent in the test set, with 95% for background and 88% for road, indicating robust detection capabilities.
- **Traffic Light:** The accuracy for the 'traffic light' class stands at 61% in the validation set, improving slightly to 67% in the test set. This suggests that while the model is reasonably effective at recognizing traffic lights, there is room for improvement.
- **Traffic Sign:** The model struggles with the 'traffic sign' class, recording an accuracy of only 24% in the validation set, which increases to 68% in the test set. This inconsistency indicates potential issues with

generalization and highlights the need for better training data representation.

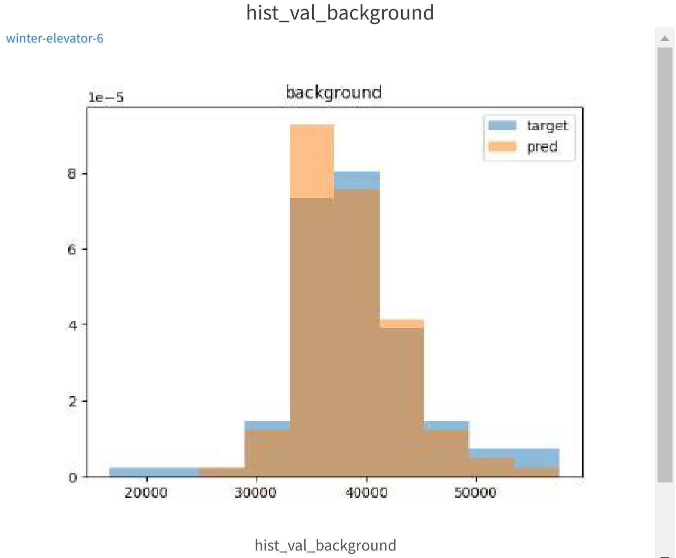
- **Person:** The detection of 'persons' shows a 64% accuracy in the validation set, which decreases to 55% in testing. This drop suggests variability in the model's ability to identify people, indicating the need for more diverse training samples.
- **Bicycle Class Difficulty:** The 'bicycle' class is a significant challenge, with the model recording 0% accuracy in both the validation and test sets. This stark finding underscores the necessity for enhanced data representation or specific modeling techniques to improve detection of bicycles or including more images for this class in the dataset.

Summary:

The confusion matrices reveal that the model performs well in detecting 'background' and 'road' but faces notable challenges with 'traffic signs', 'persons', and particularly 'bicycles'. By addressing these weaknesses through more data augmentation, and model refinement, we can enhance the segmentation model overall effectiveness.

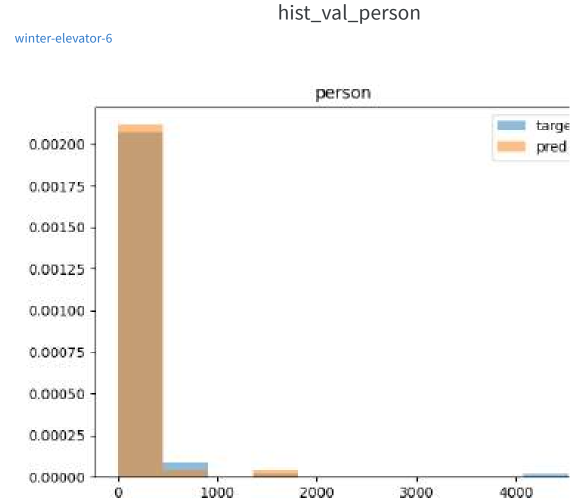
Histograms:

hist_val_background



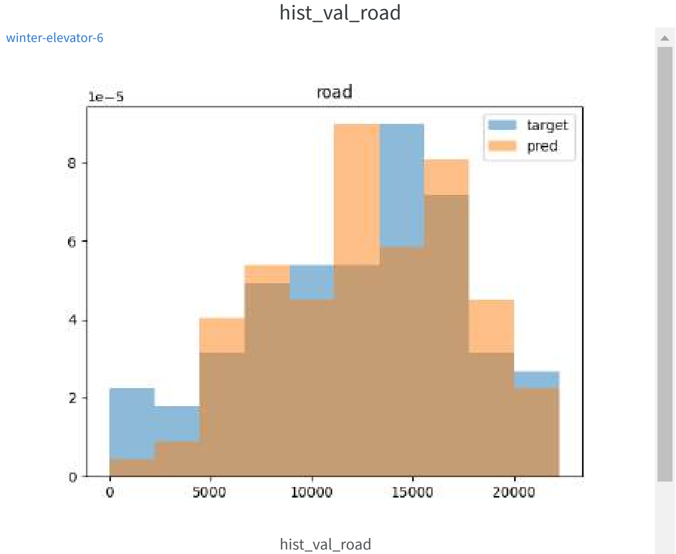
hist_val_background

hist_val_person



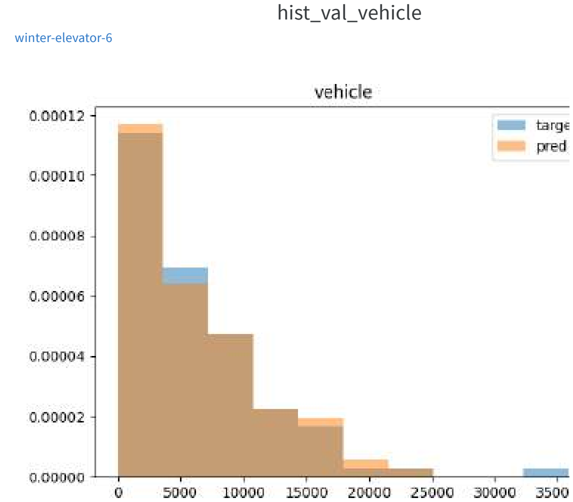
hist_val_person

hist_val_road



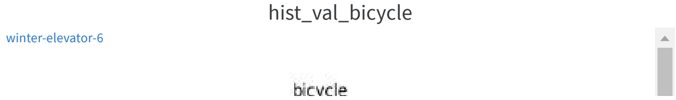
hist_val_road

hist_val_vehicle

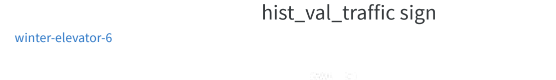


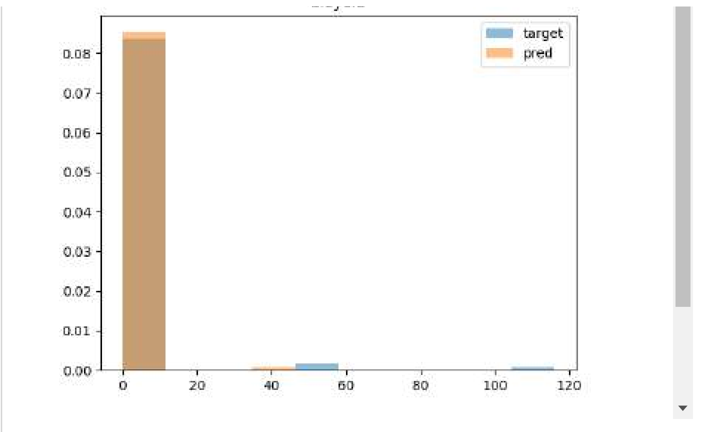
hist_val_vehicle

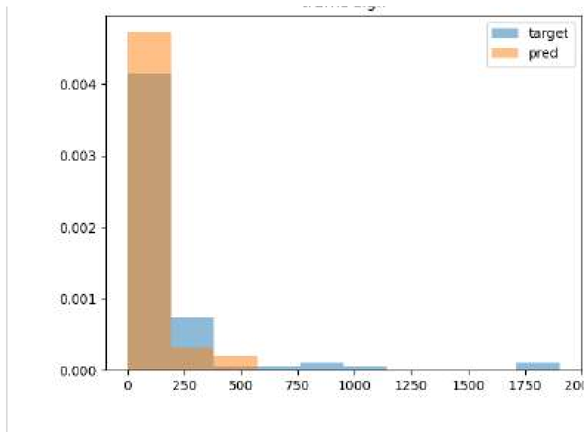
hist_val_bicycle

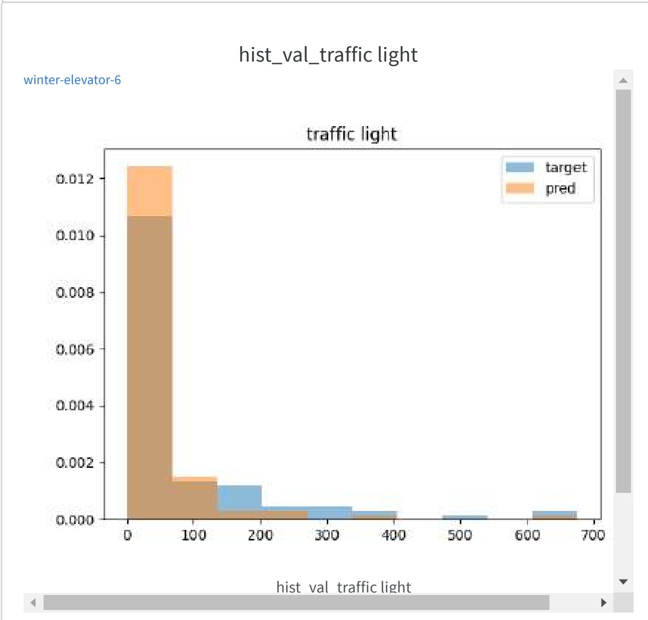


hist_val_traffic sign









Here are the key findings based on the histograms for each class:

- **Background Class:** The predicted values closely follow the target distribution with minor deviations, particularly around the 30,000–40,000 pixel range, where predictions slightly overestimate the pixel count compared to the target. Both target and predicted values peak at around 35,000 pixels, indicating that the model is generally accurate for this class.
- **Road Class:** Predicted values tend to slightly overestimate the target in the middle of the range (around 10,000 to 20,000 pixels), which aligns with the road being a predominant feature in images. Both target and prediction distributions are well aligned, especially towards the middle and higher pixel counts, indicating that the model performs well in detecting the road class.
- **Traffic Sign Class:** The model significantly overpredicts smaller pixel counts (below 250 pixels), where the largest discrepancies between target and prediction appear. The target distribution falls off quickly after this peak, showing that the model struggles with finer details like traffic signs, likely because they occupy a smaller portion of the image.
- **Person Class:** The histogram shows that both target and predicted pixel distributions are primarily concentrated in the 0 to 500 pixel range, indicating some success in identifying persons in the images. However, the predicted counts drop sharply beyond this range, suggesting that the model is less effective at capturing the presence of persons in more complex scenes or when there are multiple individuals.
- **Bicycle Class:** The predicted and target pixel counts are centered around zero, indicating that the model struggles to detect instances of the bicycle class effectively. This reinforces the notion that the model may require additional training or improvements to better identify bicycles, which are likely less prevalent in the dataset.

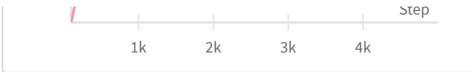
Summary:

The histogram analysis reinforces the findings from the confusion matrices, highlighting the model strengths in detecting background and road classes while exposing weaknesses in recognizing traffic signs, persons, and bicycles. The discrepancies between target and predicted values suggest specific areas for improvement, emphasizing the need for further refinement to enhance detection accuracy across all classes, particularly for those with lower pixel counts and representation in the dataset.

Losses and Metrics:

We trained the model for 50 epochs with a batch size of 8. The results indicate that the mean Intersection over Union (IoU) is achieving an overall accuracy of 45%. From the charts below, it is evident that both training and validation losses are decreasing consistently, reflecting effective model training.

Additionally, we logged the IoU metrics for each class. The results show that the background, road, and vehicle classes are performing well, with accuracies of approximately 97%, 90%, and 80%, respectively. However, the model struggles with detecting certain classes, such as traffic signs and traffic lights, and is unable to detect bicycles at all. These findings align with our earlier analyses of the confusion matrix and histograms.



Tables:

runs.summary["val_count_table"]

	y_true	y_pred
1	0.6747	0.6721
2	0.2128	0.2229
3	0.0012	0.000622
4	0.002578	0.001
5	0.002282	0.001336
6	0.1064	0.1021

1 - 6 of 7 Export as CSV Columns... Rese

runs.summary["test_count_table"]

	y_true
1	0.6984
2	0.1777
3	0.00125
4	0.007957
5	0.0021
6	0.1123
7	0.0003177

1 - 7 of 7 Export as CSV Co

Comparative Analysis of Validation and Test Sets:

1. **Background Class:**

- Validation: 67.5% (True) vs. 67.2% (Predicted)
- Test: 69.8% (True) vs. 70.7% (Predicted)
- The model shows consistent performance in predicting the background class across both sets, with slight overestimations in predictions, confirming strong alignment with previous findings from confusion matrices and histograms.

2. Road Class:

- Validation: 21.3% (True) vs. 22.3% (Predicted)
- Test: 17.8% (True) vs. 18.6% (Predicted)
- While the model performs well in detecting roads, it shows a notable drop in true pixel percentage from validation to test, suggesting potential overfitting or challenges in generalizing to new data.

3. Traffic Light Class:

- Validation: 0.1% (True) vs. 0.1% (Predicted)
- Test: 0.1% (True) vs. 0.1% (Predicted)
- The model's inability to effectively detect traffic lights remains consistent across both sets, mirroring previous findings that indicated difficulty in recognizing smaller or less prominent objects.

4. Traffic Sign Class:

- Validation: 0.3% (True) vs. 0.1% (Predicted)
- Test: 0.8% (True) vs. 0.1% (Predicted)
- The model continues to underestimate traffic signs in both sets, with a slight improvement in true pixel percentage in the test set, aligning with previous insights that highlighted struggles in detecting finer details.

5. Person Class:

- Validation: 0.2% (True) vs. 0.1% (Predicted)
- Test: 0.2% (True) vs. 0.1% (Predicted)
- The model consistently struggles to detect persons in both sets, corroborating findings from the confusion matrix and histogram analyses.

6. Vehicle Class:

- Validation: 10.6% (True) vs. 10.2% (Predicted)
- Test: 11.2% (True) vs. 10.4% (Predicted)
- Vehicle detection remains robust, with predictions closely aligned with true values across both sets, confirming previous positive insights.

7. Bicycle Class:

- Validation: 0.0% (True) vs. 0.0% (Predicted)
- Test: 0.0% (True) vs. 0.0% (Predicted)
- The model consistently fails to detect bicycles, mirroring earlier analyses that emphasized this significant gap in performance.

Summary:
The comparison between validation and test sets reveals strong performance for background and road detection, with some inconsistencies noted for the road class between sets. While vehicle detection remains reliable, the model continues to struggle with detecting smaller classes such as traffic lights, traffic signs, persons, and bicycles. These findings align with earlier evaluations from confusion matrices and histogram analyses, indicating areas where the model requires further improvement and refinement.

Created with ❤️ on Weights & Biases.

https://wandb.ai/amribrahim-amer-2024/Semantic-Segmentation-Model-for-Autonomous-Vehicle/reports/Autonomous-Driving-Segmentation-Report--Vmlldzo5NTcxMzUx?accessToken=in3t56dygyohbdug7yh3cqo33ds3ydkj71oih8nqr7vw98wmorr228ccg357h56y