Fine-Tuning Pre-Trained Faster R-CNN to Reduce False Positives in Guardrail Damage Detection from Dashcam Images

S. T. Jones[1], A. H. Le[1], R. Raj[1], D. Sacharny[1], Z. Al-Halah[1], T. C. Henderson[1], and S. Kundur[2]

[1]University of Utah, Salt Lake City, UT
[2]Bentley Systems

***Abstract*— Highway guardrails are critical safety features that require timely maintenance to remain effective. However, current inspection methods are manual, reactive, costly, and prone to human error, often failing to detect damage from/due to unreported incidents. We develop a deep learning fine-tuning pipeline to reduce false positives for automated guardrail damage detection using dashcam images implemented via supervised fine-tuning (SFT) of a pre-trained Faster R-CNN model using focal loss to address class imbalance and improve sensitivity to subtle damage features. We further evaluate the impact of data augmentation, custom anchor box design, two-stage detection, and post-processing techniques, such as DBSCAN clustering and Soft Non-Maximum Suppression (Soft-NMS). Our results demonstrate that SFT with focal loss reduces false positives (from 2.84 to 0.25 per image) while maintaining precision and recall. This work presents a scalable and effective approach for guardrail damage detection, with potential applications in other infrastructure monitoring systems for improved roadway safety.**

I. Introduction and Background

Highway guardrails play a critical role in roadway safety by preventing vehicles from veering off the road or colliding with hazardous obstacles [1]. Over time, their effectiveness can degrade due to vehicle impacts, environmental corrosion, and natural aging processes [2]. Timely maintenance of damaged guardrail sections is essential to prevent serious traffic accidents. However, current inspection practices rely heavily on accident reports and manual field inspections. These limitations highlight the need for reliable automated and scalable methods for damage assessment. Dashboard camera (dashcam) footage with GPS presents a promising solution for large-scale, cost-effective guardrail monitoring [3]. Unlike scheduled patrols, dashcam data can be collected passively during regular driving, allowing broad coverage without additional effort.

The application of computer vision techniques for infrastructure monitoring has been widely explored in recent years [4]–[8]. Convolutional neural networks (CNNs) and object detection models, such as YOLO and Faster R-CNN, have been adapted to detect structural damage in various transportation assets, including bridges and pavement [1], [2], [6], [9], [10]. While these approaches show promise, their application to guardrail damage detection, particularly from real-world dashcam imagery, remains relatively limited. The complexity of guardrail scenes, variations in lighting and camera angles, and the subtlety of visual damage all contribute to frequent false positive detections.

Fig. 1. Example image from Blyncsy, Inc. dataset with guardrail damage ground truth box shown in red.

To address this, Lin et al. introduced focal loss, a loss function designed to reduce the influence of easy negatives and focus training on hard, misclassified examples [11], [12]. Focal loss has been successfully used to improve precision in object detection tasks characterized by significant class imbalance. Applying this concept to guardrail damage detection offers a promising direction for reducing false positives while preserving high recall.

II. Method

The training and validation datasets for guardrail damage detection consisted of dashcam images collected from vehicles provided by Blyncsy, Inc. Each image was manually annotated with bounding boxes identifying guardrail damage based on NCHRP guidelines [13]. The dataset includes 30,383 training images and 8,682 validation images, labeled with two classes: guardrail damage (1) and background (0). All images in the dataset contained visible guardrail damage and, therefore, included at least one ground truth bounding box. As a result, the dataset does not include true negative examples. An example image with its corresponding ground truth bounding box is shown in Figure 1.

Blyncsy, Inc. provided two pre-trained models for: detection of (1) guardrails and (2) damage. Both models are based on the Faster R-CNN architecture, a two-stage object detection framework known for its accuracy and flexibility. The architecture includes a convolutional backbone network

to extract feature maps from input images, followed by a Region Proposal Network (RPN) that generates candidate object regions using anchor boxes [14]. These proposals are then passed through ROI pooling to produce fixed-size feature maps, which are processed by fully connected layers for classification and bounding box refinement [14].

Augmentations were chosen to make the model invariant to damage position, improve its robustness under different lighting conditions, and enable it to generalize well on low-quality images. Of the 30,383 training images, augmentations were applied to 50% of the data.

We fine-tuned the baseline pre-trained Faster R-CNN model for guardrail damage detection using supervised learning with focal loss as the classification objective to improve the model's ability to detect subtle guardrail damage while minimizing false positives.

Focal loss modifies standard cross-entropy loss to down-weight well-classified examples and focus training on harder, misclassified examples [11]. Formally, focal loss is defined as:

$$FL(p_t) = -\alpha_t(1-p_t)^{\gamma} log(p_t) \quad (1)$$

where p_t is the model's estimated probability for the ground-truth class, $\gamma \geq 0$ is the focusing parameter, and α_t is a weighting factor that balances the importance of positive and negative examples [11]. When $\gamma = 0$, focal loss reduces to the standard cross-entropy. As γ increases, the loss assigned to well-classified examples decreases, allowing the model to concentrate on hard negative and ambiguous foreground instances.

Hyperparameter tuning used Bayesian optimization across 20 trials of 10 epochs each. Tested parameters included batch sizes (4, 8, 16), learning rates (1 x 10^{-5} to 110^{-2}), momentum (0.85–0.99), and focal loss parameters α (0.1–1.0) and γ (0.5–5.0), with weight decay fixed at 1×10^{-4}. The final SFT model with focal loss was trained twice—on original images and on a mix of original and augmented images—using the best-performing hyperparameters (Table I).

To refine the raw detection from our models, we employed multiple post-processing approaches to reduce false positives and overlapping boxes. The approaches consist of 1) DBSCAN clustering and box fusion, 2) Soft Non-Maximum Suppression (Soft-NMS), and 3) a combination of both techniques. To evaluate detection performance, we used mean average precision (mAP) and false positive rate.

TABLE I

TUNED HYPERPARAMETERS FOR SFT WITH FOCAL LOSS

Hyperparameter	Value
Learning Rate	0.002604
Momentum	0.9898
α	0.103
γ	2.527
Batch Size	16
Weight Decay	0.0001
Number of Epochs	60

We created an automated QA process using the initially provided robust model to score the ground truth bounding boxes based on the model's current focus. Theoretical bounding boxes were generated from the saliency maps from the images and compared to the ground truth using IoU overlap and center distance.

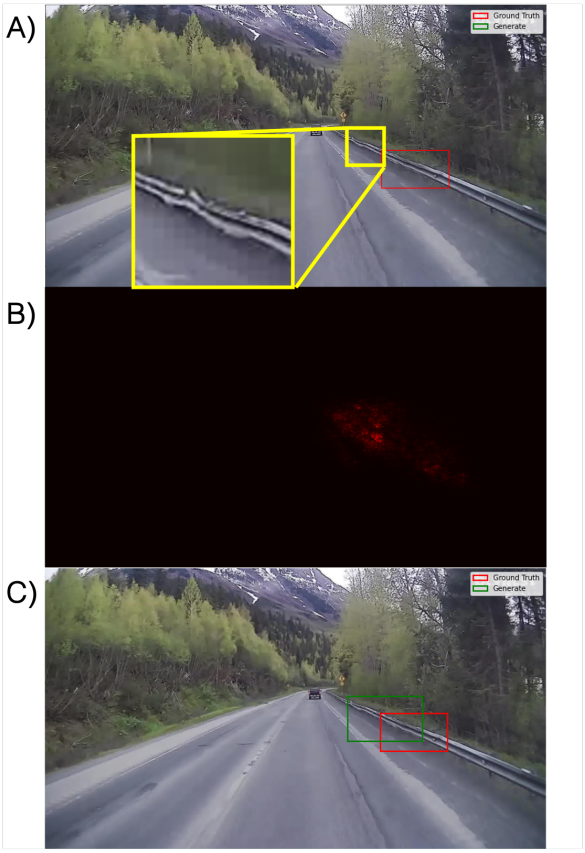

Fig. 2. Generated bounding box from saliency pipeline. A) The original image with the ground truth bounding box marked in red and missed damage marked in green; B) Saliency map from the model of the original image input; C) Generated bounding box from the saliency map with damage now included.

. We developed an automated QA method using a pre-trained model to assess ground truth bounding box accuracy in training and validation sets. The baseline model detected guardrail damage but revealed mislabeled or missing annotations. Saliency maps generated theoretical bounding boxes compared with originals using IoU overlap and center distance. A QA score (0–1) was calculated with alpha weight of 0.7 emphasizing IoU overlap. The scoring pipeline is illustrated in Figure 2.

Background noise affects accuracy, so a two-stage pipeline was employed: the baseline guardrail detection model first identified guardrail segments, then the damage detection model located damage. While two-stage detection successfully segmented guardrails and identified damage in many cases (see Figure 3A), it yielded lower precision and recall compared to the baseline (Table II). An example is shown in Figure 3.

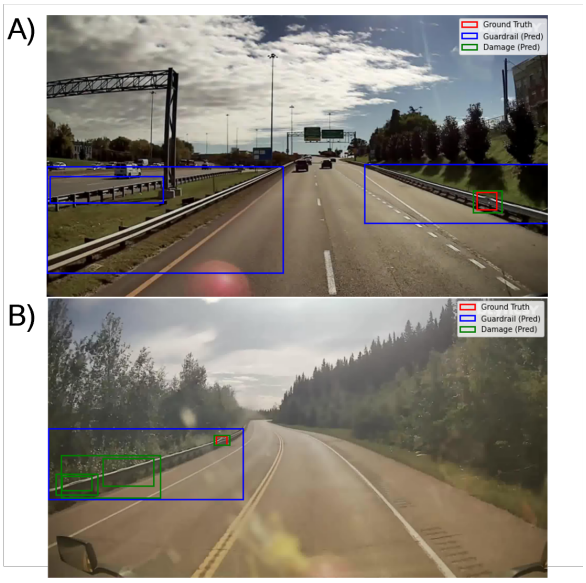

Fig. 3. Outputs from two-stage detection pipeline. A) An image showing detected guardrails and one predicting bounding box with good IoU overlap with the ground truth; B) An image using two-stage detection with many false positives.

III. Results

We evaluated the performance of various configurations of the Faster R-CNN model for guardrail damage detection using the validation dataset. Table II summarizes the results across four key metrics: mAP at IoU thresholds of 0.5 and 0.5:0.95, mean Average Recall (mAR), and average False Positive Rate (FPR). The models included the baseline pre-trained model, fine-tuned variants with and without augmentations, and post-processing and architectural modifications to reduce false positives.

The baseline model, a pre-trained Faster R-CNN with no task-specific tuning, achieved an mAP of 0.262 at IoU 0.5:0.95 and 0.643 at IoU 0.5. The mAR was 0.444, with a high false positive rate of 2.84. SFT significantly improved detection precision. The mAP increased to 0.277 (IoU 0.5:0.95) and 0.665 (IoU 0.5), while mAR lifted to 0.482. The false positive rate dropped sharply to 0.34, suggesting that fine-tuning enabled the model to better distinguish damage and background. Data augmentations were added to the fine-tuning process to improve generalization. The model maintained a similar mAP (0.276 at IoU 0.5:0.95 and 0.658 at IoU 0.5), but showed an additional lift in mAR, reaching 0.491. The false positive rate further dropped to 0.25, the lowest among all models.

Example images below a threshold of 0.1 are shown in Figure 4A and B. Images with scores above 0.5 are generally considered correct and not suggested for additional QA. An image with a near-perfect score is shown in Figure 4D. Two-stage detection resulted in increased false positives, from an average of 2.84 per age up to 2.95 per image.

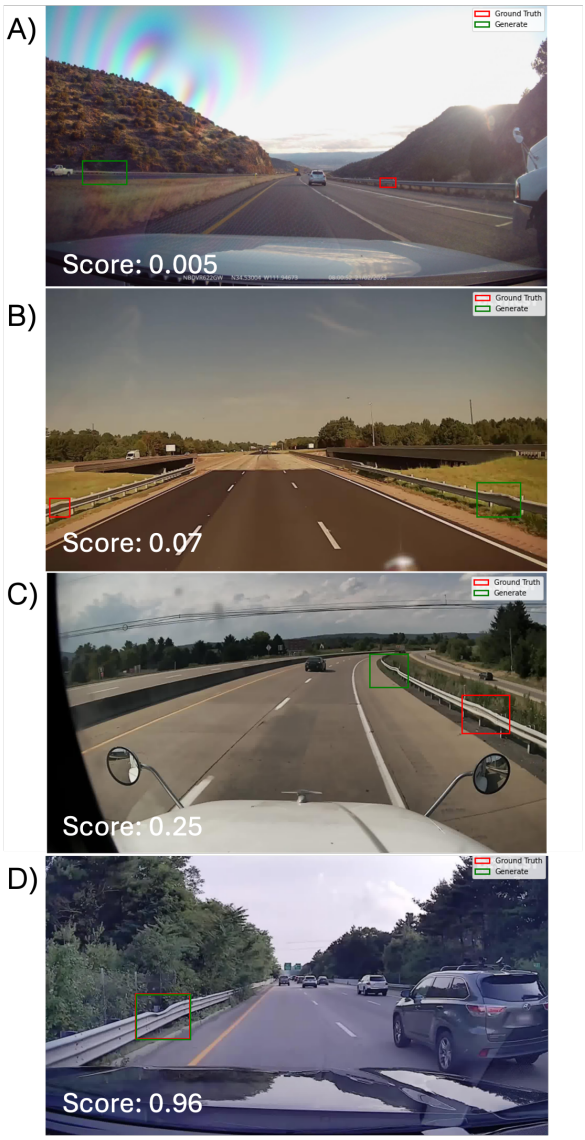

Fig. 4. Scoring based on generated bounding boxes from saliency maps and their associated images. A) Low-scoring image with large windshield distortion; B) Low-scoring image with additional instances of damage located across the highway; C) Slightly below threshold scoring image with additional instances of damage higher up on the guardrail and correctly labeled damage; D) High-scoring image with maximal overlapping ground truth and generated bounding boxes.

IV. Discussion and Conclusions

Blyncsy, Inc. presented us with the task of reducing false positives in guardrail detection. We utilize an SFT pipeline to improve model accuracy. SFT with focal loss was introduced in this study to address one of the central challenges in guardrail damage detection from dashcam images: the high rate of false positives resulting from subtle visual distinctions between damaged and undamaged guardrail sections [11], [12]. Our results demonstrate that fine-tuning the pre-trained Faster R-CNN model using focal loss led to a significant reduction in false positives, while improving both the precision and reliability of the detector. Compared to the baseline model, which exhibited a high false positive rate of 2.84,

TABLE II

COMPARISON OF VALIDATION METRICS FOR BASELINE AND SUPERVISED FINE-TUNED FASTER R-CNN MODEL FOR GUARDRAIL DAMAGE DETECTION

Validation Metric	Baseline	Fine-Tuned	Fine-Tuned with Augmentations	DBSCAN	Soft NMS	DBSCAN + Soft NMS	Two-Stage Detection	Custom Bounding Boxes
mAP @ IoU=0.5:0.95	0.262	0.277	0.276	0.216	0.237	0.196	0.185	0.236
mAP @ IoU=0.5	0.643	0.665	0.658	0.533	0.588	0.480	0.485	0.619
mAR @ IoU=0.5:0.95	0.444	0.482	0.491	0.389	0.362	0.332	0.352	0.439
False Positive Rate	2.84	0.34	0.25	1.83	0.64	0.49	2.95	0.42

the fine-tuned model with focal loss reduced this to just 0.34. This large improvement indicates that focal loss was effective in helping the model suppress noisy or spurious detections of background regions. Interestingly, while the mAP improvements were modest (from 0.262 to 0.277), the mAR increased more substantially, suggesting that focal loss improved the model's ability to detect true positives without overwhelming it with background noise.

Despite its advantages, focal loss is not without limitations. The choice of hyperparameters γ and α significantly affects performance and may require tuning specific to the dataset and object class imbalance [11], [12]. Additionally, focal loss assumes that hard examples are informative, which may not always hold true in noisy datasets. In cases where poor-quality annotations or occlusions contribute to model confusion, focal loss could amplify the influence of mislabeled or ambiguous data points. Overall, SFT with focal loss proved to be a highly effective strategy for improving model robustness and reducing false positive detections in guardrail damage detection.

Future development may use the current baseline model to evaluate the training and validation sets, identifying mislabeled or ambiguous images that may require additional quality assurance, as well as examples where the model consistently underperforms. We also recommend training a new model with custom anchor boxes, as the default sizes in Faster R-CNN are often too large for many damage instances in this dataset. Finally, incorporating focal loss can further improve performance by addressing class imbalance, reducing the influence of easy background examples and encouraging the model to focus on challenging, underrepresented damage cases.

This work developed an SFT pipeline for guardrail damage detection using Faster R-CNN, with a particular focus on reducing false positives through the use of focal loss. Future work should consider fine-tuning on a curated dataset that includes either more obvious and visually distinct damage examples or a mix of true negative images containing intact guardrails without damage. This could help the model better differentiate between damaged and undamaged states and reduce confusion during training.

ACKNOWLEDGMENT

Thanks to Blyncsy, Inc. and Bentley Systems for providing models, training datasets, and technical guidance, and the Deep Learning Program at the University of Utah.

REFERENCES

[1] M. Soltani, T. B. Moghaddam, M. R. Karim, and N. H. R. S. and, "The safety performance of guardrail systems: review and analysis of crash tests data," *International Journal of Crashworthiness*, vol. 18, no. 5, pp. 530–543, 2013.

[2] X. Jin, M. Gao, D. Li, and T. Zhao, "Damage detection of road domain waveform guardrail structure based on machine learning multi-module fusion," *PLOS ONE*, vol. 19, no. 3, pp. 1–27, 03 2024. [Online]. Available: https://doi.org/10.1371/journal.pone.0299116

[3] C. Naude, T. Serre, M. Dubois-Lounis, J.-Y. Fournier, D. Lechner, M. Guilbot, and V. Ledoux, "Acquisition and analysis of road incidents based on vehicle dynamics," *Accident Analysis & Prevention*, vol. 130, pp. 117–124, 2019.

[4] S. Li and Y. Huang, "Damage Detection Algorithm Based on Faster-RCNN," in *2023 5th International Conference on Electronics and Communication, Network and Computer Technology (ECNCT)*. Guangzhou, China: IEEE, Aug. 2023, pp. 177–180. [Online]. Available: https://ieeexplore.ieee.org/document/10280967/

[5] W. Ding, X. Zhao, B. Zhu, Y. Du, G. Zhu, T. Yu, L. Li, and J. Wang, "An Ensemble of One-Stage and Two-Stage Detectors Approach for Road Damage Detection," in *2022 IEEE International Conference on Big Data (Big Data)*. Osaka, Japan: IEEE, Dec. 2022, pp. 6395–6400. [Online]. Available: https://ieeexplore.ieee.org/document/10021000/

[6] N. Youssouf, "Traffic sign classification using cnn and detection using faster-rcnn and yolov4," *Heliyon*, vol. 8, no. 12, 2022.

[7] W. Wang, B. Wu, S. Yang, and Z. Wang, "Road damage detection and classification with faster R-CNN," in *IEEE International Conference on Big Data (IEEE BigData 2018), Seattle, WA, USA, December 10-13, 2018*, N. Abe, H. Liu, C. Pu, X. Hu, N. K. Ahmed, M. Qiao, Y. Song, D. Kossmann, B. Liu, K. Lee, J. Tang, J. He, and J. S. Saltz, Eds. IEEE, 2018, pp. 5220–5223. [Online]. Available: https://doi.org/10.1109/BigData.2018.8622354

[8] T. Getahun, A. Karimoddini, and P. Mudalige, "A deep learning approach for lane detection," in *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*. IEEE, 2021, pp. 1527–1532.

[9] I. García-Aguilar, J. García-González, R. M. Luque-Baena, and E. López-Rubio, "Automated labeling of training data for improved object detection in traffic videos by fine-tuned deep convolutional neural networks," *Pattern Recognition Letters*, vol. 167, pp. 45–52, 2023.

[10] H. Kim, Y. Lee, B. Yim, E. Park, and H. Kim, "On-road object detection using deep neural network," in *2016 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia)*. IEEE, 2016, pp. 1–4.

[11] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.

[12] X. Li, W. Wang, L. Wu, S. Chen, X. Hu, J. Li, J. Tang, and J. Yang, "Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection," *Advances in neural information processing systems*, vol. 33, pp. 21 002–21 012, 2020.

[13] RoadSafe, LLC, *Field Manual for Guardrail Inspection and Damage Assessment*, 3rd ed., RoadSafe, LLC, n.d., accessed: YYYY-MM-DD. [Online]. Available: https://www.roadsafellc.com/Field_Manual3b.pdf

[14] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 6, pp. 1137–1149, 2016.