# Benchmarking YOLOv11 against RF-DETR for Defect Detection in Automotive Heat Shields

Saber Yu*,†, Mengchen Liu*, Timothy Reinhart*,†, Seshasai Srinivasan*, Zhen Gao*, Majan Alavi*

*W Booth School of Engineering Practice and Technology, McMaster University, Hamilton, ON, Canada
†Ascension Automation Solutions Ltd, Cambridge, ON, Canada

Emails: {yux8, liu2583, reinhat, ssriniv, gaozhen, alavis2}@mcmaster.ca

***Abstract*—Industrial defect detection often faces challenges from limited defective samples and low data variations. Through evaluating various augmentation techniques on YOLOv11-Nano using a custom heat shield dataset, we developed an optimized augmentation pipeline that improved the model's average recall from 76.43% (without augmentation) to 97.86% (+21.43%). Comparative benchmarking of YOLOv11 against RF-DETR, RT-DETR, and YOLOv12, demonstrated that RF-DETR-Base (100% recall — 98.26% precision — 40 FPS) and YOLOv11-Nano (97.86% recall — 98.04% precision — 103 FPS) were the best performing architectures, with the former excelling in detection accuracy and the latter in inference speed. Our findings align with recent COCO benchmarks, where RF-DETR emerges as the latest state-of-the-art architecture, and confirm its capability for industrial applications.**

***Index Terms*—Computer Vision, Object Detection, Defect Detection, Data Augmentation, YOLO, RF-DETR.**

## I. Introduction

Specialized industrial datasets often contain limited defective samples. This data scarcity poses significant challenges for model training and generalization, highlighting the value and necessity of data augmentation [25]. Traditional augmentation techniques include geometric transformations like translation, scaling, and rotation, and photometric transformation such as hue, brightness, and saturation adjustments. More advanced methods include structural rearrangement such as mix-up [23], cut-mix [21], and mosaic [1]. Beyond manual augmentations, researchers have invented automatic algorithms like AutoAugment [4], RandAugment [5], and TrivialAugment [11], all of which have proven effective in image classification. The latest research in this field explores generating completely synthetic images using Variational Autoencoders (VAE) and Generative Adversarial Networks (GAN) [17], [20] and creating new data based on domain knowledge [6], [9]. In practice, the performance difference between a model with and without proper augmentation can exceed the gap between different architectures. As a result, optimizing the augmentation pipeline is crucial for both application development and fair model comparison.

While image augmentation addresses the data scarcity challenge, another critical consideration lies within the selection of pre-trained models. The field of computer vision has undergone very rapid evolutions in the last decades, creating a large pool of options for model selection (Figure 1). While the latest DETR-based [3] transformer architectures showcased promising results on large standard datasets like COCO [10], their performance on small industrial datasets remains unclear. Our work aims to address this gap by benchmarking YOLOv11 against RF-DETR [14], RT-DETR [24], and YOLOv12 [19] on an industrial heat shield dataset.

## II. Methods

### A. Data Acquisition

Automotive heat shields, as shown in Figure 2, are designed to isolate heat, noise, and vibration from high-temperature components in the vehicle such as engines and turbochargers. For this study, we acquired a total of 21 heat shields from the industry, including 3 defect-free and 18 defective specimens (of which 11 had splits, 6 had crushed edges, and 1 had both types of defects). Seven images per specimen were taken to increase the size of the dataset. The camera location and field of view were fixed during image acquisition to simulate in-fixture manufacturing inspection, but slight positional variations existed due to manual loadings of samples. Lighting and camera parameters (exposure, gain, gamma, contrast, sharpness) were deliberately varied to increase data diversity. A monochrome camera was used to increase light sensitivity and highlight defect details. The final dataset contained 147 images of 2592 x 2048 pixels with 84 split instances, 49 crushed edge instances, and 21 null annotations for defect-free parts. To prevent data leakage, we manually and carefully split these images into 105 for training, 21 for validation, and 21 for testing, ensuring that data from the same physical sample have not been assigned to different splits.

### B. Applying Augmentation

Geometric, photometric, and image-structural transformations were applied (in this order) with parameters detailed in Table 1. Mosaic augmentation used a 2x2 grid following the original approach [1]. YOLO's default Albumentations [2] settings, consisting of a minor blur, togrey, and CLAHE were maintained throughout the study. We performed 10 runs for each experiment and fixed the training seeds used across

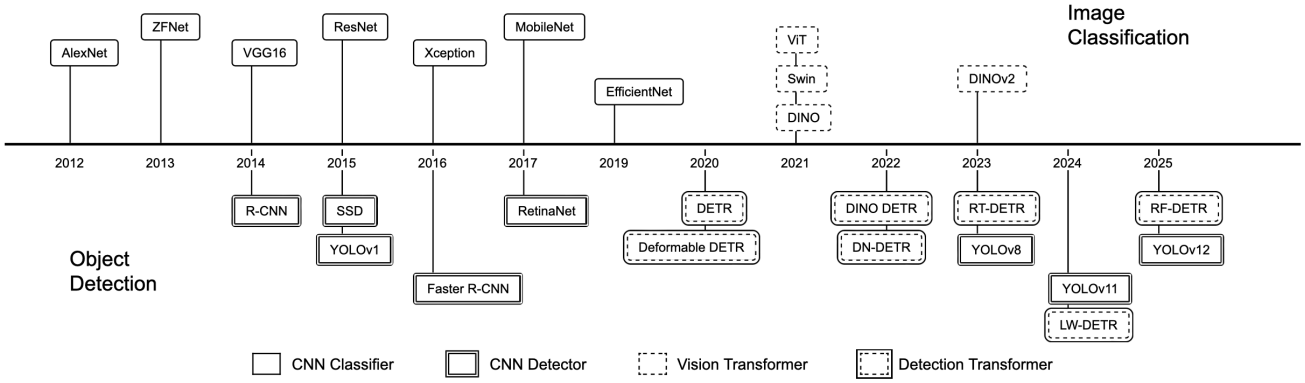


Fig. 1: Timeline of the recent advancements of vision models, highlighting key architectures such as AlexNet [8], ResNet [7], EfficientNet [18], Faster R-CNN [13], YOLO [12], DETR [3], DINO DETR [22], RT-DETR [24], and RF-DETR [14].

Fig. 2: An example of an annotated heat shield image showing a split defect (blue box) and a crushed edge (red box).

TABLE I: Data Augmentation Techniques and Parameters.

| Category | Code | Description | Strength | Probability |
|---|---|---|---|---|
| Geometric | G1 | Horizontal Flip | NA | 50% |
| | G2 | Scale | 10% | 100% |
| | G3 | Translation | 5% | 100% |
| | G4 | Shear (s/l) | 3 / 5 deg | 100% |
| | G5 | Rotate | 10 deg | 100% |
| | G6 | Perspective (s/l) | 2 / 5% | 100% |
| Photometric | P1 | Brightness | 20% | 100% |
| | P2 | Hue | 20% | 100% |
| | P3 | Saturation | 20% | 100% |
| Structural | S1 | Mosaic | NA | 100% |

trials. Recall (coverage) was prioritized in this application over precision to minimize the risk of undetected defects.

### C. Model Evaluation

For fair model comparison, we created a new training dataset, containing 525 augmented images (5x transformation for each of the original 105 images in this split), using the best policy achieved from our augmentation experiment. All models were trained on these images at their original resolution with model-level resizing. Experiments were conducted on an **NVDIA RTX 4070 Laptop** GPU and **Intel Core i7-14700HX** CPU.

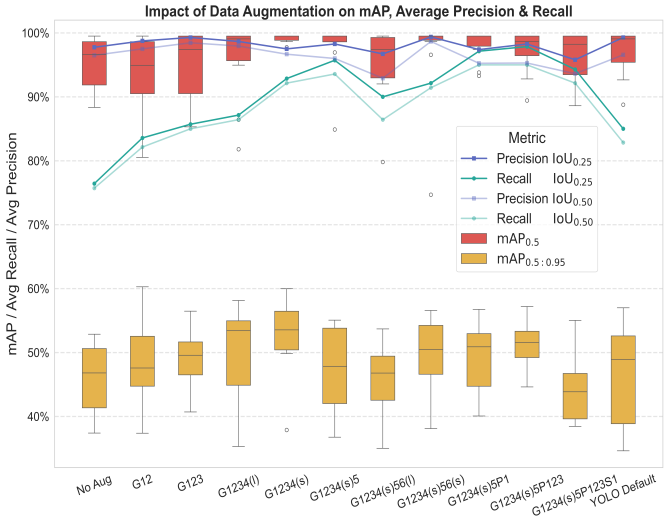


Fig. 3: The effects of data augmentation on YOLOv11-Nano. Eg. A **G1234(s)5P1** strategy contains G1, G2, G3, G4(s), G5, and P1 augmentations. YOLO default consists of horizontal flip (50% chance), scale (0.5), translate (0.1), brightness (0.4), hue (0.015), saturation (0.7), and mosaic (always applied).

## III. Results

### A. Data Augmentation

As shown in Figure 3, the baseline (no augmentation) achieved 76.43% average recall at 0.25 IoU and 0.25 confidence threshold. Affine augmentations such as flip, scale, and translation boosted this value to 85.71% (+9.28%). Adding a minor shear and rotation further improved recall to 95.71% (+19.29% from baseline), but more aggressive transformations (with higher shear intensities and perspective distortions) were counter-productive. Color space augmentations (saturation and hue) showed limited effects since the images were in monochrome, but brightness changes further optimized recall by 1.43%. Finally, mosaic was found to be detrimental, with YOLO's default augmentation settings, which included this technique, achieving only 85.00% average recall on YOLOv11-Nano.

TABLE II: The Average Inference Speed of Each Model in Milliseconds and Frame Per Second.

| Model | YOLOv11-N | YOLOv12-N | YOLOv11-S | YOLOv11-M | YOLOv11-L | YOLOv11-XL | RF-DETR-Base | RT-DETR-L |
|---|---|---|---|---|---|---|---|---|
| **Milliseconds** | 9.7 | 11.9 | 9.5 | 12.2 | 15.4 | 26.1 | 25.0 | 23.3 |
| **FPS** | 103 | 84 | 105 | 82 | 65 | 38 | 40 | 43 |

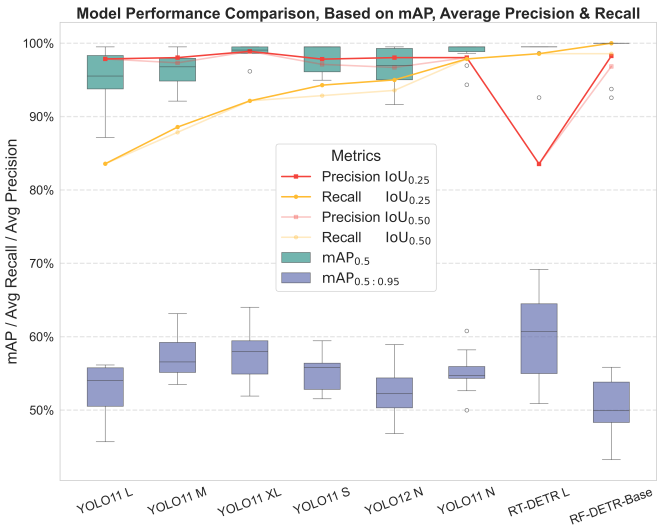


Fig. 4: A comparison of model performance across YOLOv11, RF-DETR, RT-DETR, and YOLOv12, highlighting average recall and precision, as well as $mAP_{0.5}$ and $mAP_{0.5:0.95}$.

### B. Model Performance

All models demonstrated excellent levels of average recall ($> 90\%$) with the augmented data. Our results (Figure 4) showed that RF-DETR-Base achieved 100% recall across all 10 experiments at 0.25 IoU and 0.25 confidence thresholds, while maintaining excellent precision at 98.26%. Second to was RT-DETR-Large, with an average recall of 98.57% but a significantly lower precision (83.54%). This was then followed by YOLOv11-Nano (97.86% recall and 98.04% precision) and YOLOv12-Nano (95.00% recall and 98.04% precision). All larger variants within the YOLOv11 family showed inferior performance compared to the Nano model. Lastly, as shown in Table 2, YOLOv11-Nano and YOLOv11-Small demonstrated the quickest inference speeds (9.7 ms and 9.5 ms per inference, relatively) whereas RT-DETR, RF-DETR, and YOLOv11-XL were slower (23.3 ms, 25.0 ms, and 26.1 ms).

## IV. Discussion

### A. Data Augmentation

Despite a 2:1 class imbalance in the training data favoring the split class, its detection was a more difficult task compared to crushed edges, especially when no augmentations were applied (61.43% vs. 91.43% recall on YOLOv11-Nano, respectively). Geometric transformations at optimal intensities substantially improved split recall to 97.14% (+35.72% from baseline), while further photometric adjustments primarily benefited crushed edge detections. This unexpectedly large impact of geometric augmentations could be partially attributed to the positional variations introduced by the manual part loading process in data collection. During actual manufacturing, heat shields are stamped and inspected in fixtures where their orientations remain unchanged, and as a result, the real performance gap between baselines and augmented trials in production settings should be less significant. In addition, perspective distortions and mosaic were found to damage model performance. This is likely because they have excessively altered the locations where defects might appear and therefore biased the model towards impractical cases. These findings highlight the importance of domain-specific augmentation designs, as emphasized by [16], particularly for industrial applications with fixed inspection geometries.

### B. Model Performance

While benchmarking model performance, our results indicated that RF-DETR achieved the best average recall and precision at 0.25 IoU and 0.25 confidence thresholds. The model outperformed RT-DETR and the latest CNN architectures (YOLOv11 and YOLOv12), exceeding the best YOLO variant by 2.14% in recall (100% vs. 97.86%) and 0.22% in precision (98.26% vs. 98.04%). The only metric where RF-DETR underperformed was average $mAP_{0.5:0.95}$, however, since this value measures mean model performance at higher IoU thresholds across all confidence levels, its impact was considered less practically significant for our application.

Although RF-DETR excelled in detection accuracy, YOLO remained superior in inference speed, with smaller variants like YOLOv11-Nano processing approximately 2.5 times faster than RF-DETR (9.7 ms vs. 25.0 ms per inference). Larger YOLO models, however, demonstrated poorer performance on both accuracy and speeds, likely due to our limited dataset size. Future work could investigate retraining these larger models as additional data become available. To summarize, we recommend using RF-DETR for most applications while reserving YOLO only for cases where inference speed is critical. Our findings generally align with [14] and [15] and confirm the potential of RF-DETR for industrial defect detection.

## V. Conclusion

This study evaluated data augmentation techniques for training YOLOv11-Nano on an industrial heat shield dataset and benchmarked the model's performance against RF-DETR-Base, RT-DETR-Large, YOLOv12-Nano, and other YOLOv11 variants. The optimized augmentation pipeline was found to increase average recall on YOLOv11-Nano from 76.43% (no augmentation) to 97.86% (+21.43%). Trainings on this augmentation strategy revealed RF-DETR-Base as the best

architecture for detection accuracy while YOLOv11-Nano remained superior in inference response.

## ACKNOWLEDGMENT

The authors would like to thank Patrick Gilbert and Thomas Bachle at Dana Inc. for their assistance with data collection, and Salman Bawa and Richard Allen at McMaster University for coordinating this project with Ascension Automation Solutions Ltd.

## REFERENCES

[1] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. Yolov4: Optimal speed and accuracy of object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 163–178, 2020.

[2] Alexander Buslaev, Alex Parinov, Eugene Khvedchenya, Vladimir I Iglovikov, and Alexandr A Kalinin. Albumentations: fast and flexible image augmentations. *arXiv preprint arXiv:1809.06839*, 2018.

[3] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In *European Conference on Computer Vision*, pages 213–229. Springer, 2020.

[4] Ekin D. Cubuk, Barret Zoph, Dandelion Mane, Vijay Vasudevan, and Quoc V. Le. Autoaugment: Learning augmentation strategies from data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 113–123, 2019.

[5] Ekin D. Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V. Le. Randaugment: Practical automated data augmentation with a reduced search space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 702–703, 2020.

[6] Yu Gong, Xiaoqiao Wang, and Chichun Zhou. Human-machine knowledge hybrid augmentation method for surface defect detection based few-data learning. *[Journal name not visible in excerpt]*, pages 1–24, 2023.

[7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.

[8] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, volume 25, pages 1097–1105, 2012.

[9] Adrian Shuai Li, Elisa Bertino, Rih-Teng Wu, and Ting-Yan Wu. Building manufacturing deep learning models with minimal and imbalanced training data using domain adaptation and data augmentation. In *2023 IEEE International Conference on Industrial Technology (ICIT)*, pages 1–8, Orlando, FL, USA, April 2023.

[10] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. *arXiv preprint arXiv:1405.0312*, 2014.

[11] Samuel G. Müller and Frank Hutter. Trivialaugment: Tuning-free yet state-of-the-art data augmentation. *arXiv preprint arXiv:2103.10158*, 2021.

[12] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 779–788, 2016.

[13] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6):1137–1149, 2017.

[14] Isaac Robinson, Peter Robicheaux, and Matvei Popov. Rf-detr. https://github.com/roboflow/rf-detr, 2025. SOTA Real-Time Object Detection Model.

[15] Ranjan Sapkota, Rahul Harsha Cheppally, Ajay Sharda, and Manoj Karkee. Rf-detr object detection vs yolov12 : A study of transformer-based and cnn-based architectures for single-class and multi-class greenfruit detection in complex orchard environments under label ambiguity. *arXiv preprint arXiv:2504.13099*, 2025. Submitted to Elsevier.

[16] Connor Shorten and Taghi M. Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(60), 2019.

[17] Chaitanya Singla, Rajat Bhardwaj, Nilesh Shelke, and Gurpreet Singh. Data augmentation: Synthetic image generation for medical images using vector quantized variational autoencoders. In *2025 3rd International Conference on Disruptive Technologies (ICDT)*, pages 1502–1507. IEEE, 2025.

[18] Mingxing Tan and Quoc V Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *PMLR*, pages 6105–6114, 2019.

[19] Yunjie Tian, Qixiang Ye, and David Doermann. Yolov12: Attention-centric real-time object detectors. 2025.

[20] Wei Xiong, Janghwan Lee, Shuhui Qu, and Wonhyouk Jang. Data augmentation for applying deep learning to display manufacturing defect detection. In *2020 International Display Workshops (IDW)*, pages 81–1. The Institute of Image Information and Television Engineers, 2020.

[21] Sangdoo Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, and Youngjoon Yoo. Cutmix: Regularization strategy to train strong classifiers with localizable features. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 6023–6032, 2019.

[22] Hao Zhang, Feng Li, Shilong Liu, Lei Zhang, Hang Su, Jun Zhu, Lionel M Ni, and Heung-Yeung Shum. Dino: Detr with improved denoising anchor boxes for end-to-end object detection. *arXiv preprint arXiv:2203.03605*, 2022.

[23] Hongyi Zhang, Moustapha Cisse, Yann N. Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. In *International Conference on Learning Representations (ICLR)*, 2018.

[24] Yian Zhao, Wenyu Lv, Shangliang Xu, Jinman Wei, Guanzhong Wang, Qingqing Dang, Yi Liu, and Jie Chen. Detrs beat yolos on real-time object detection. *arXiv preprint arXiv:2304.08069*, 2024.

[25] Xiaopin Zhong, Junwei Zhu, Weixiang Liu, Chongxin Hu, Yuanlong Deng, and Zongze Wu. An overview of image generation of industrial surface defects. *Sensors*, 23(19):8160, 2023. College of Mechatronics and Control Engineering, Shenzhen University.