Corrosion Detection at Transmission Accessories Using Combination of Object Detection, Image Classification and Background Removal

Edy Sucipto & Nugraha Priya Utama

School of Electrical Engineering and Informatics, Institut Teknologi Bandung, Bandung, Indonesia
Email: 23522307@std.stei.itb.ac.id, utama@staff.stei.itb.ac.id

Abstract. Inspection of electrical transmission accessories is an important aspect of maintaining the reliability of electricity supply. Lately there has been a trend to inspect it autonomously using Deep Learning. And one of state of the art in object detection models is YOLO. Using it in object detection and image classification task using new dataset and exclusive form Indonesia that researcher never has been done, that is Clevis, Dead end, Shackle, Tension clamp, Hole and Bolt. Data sets are also provided in several types to target the highest score. Also use another model like YOLOX and RT-DETR for object detection, and VGG-19, DenseNet-201, etc. to image classification. The result is that using smaller number of class and larger object can improve metric score, and it get by model YOLOv9e, it can reach 0,972 in mAP@0,5. And the removal of background will lead to poor metric score less than 0,20 poin in mAP@0,5. Combination between object detection and image classification seems good at training and testing part, but when combined the result decreases, it is about 0,82~0,84 precision. This is likely due to IoU limitations when extracting it from object detection which causes lower quality dataset that deliver to image classification process.

Keywords: *background removal, corrosion, defect anomaly, high voltage equipment, transmission accessories.*

1 Introduction

In this part we will talk about the background of research, related work and the hypothesis we used.

1.1 Background

Align with "Review dan Revisi Kepdir No. 0520 K/DIR/2014 tahun 2022 (Divisi RJT PT. Perusahaan Listrik Negara (Persero) Regional JAMALI, 2022)" that every high voltage transmission in PT PLN (Persero) must be implemented Climb up Inspection every 5 years. There are 36 types of equipment and accessories that must be checked. In line with 2021 annual report of PT PLN (Persero) [1] number of high voltage transmission (70 kV, 150 kV, 250 kV, 500 kV) in Indonesia reaches 68.206 kilo meter sirkit (kms) long. And if average it with 250 meters for

distances each tower, approximately number of towers is about 136.412 units. Therefore, the cost to maintain it is so high. In the interval range of inspection 5 years apart, there is a risk that an unknown anomaly will occur and become a danger. In the worst case it can cause breakdown tower transmission because structure conditions are unable to withstand the load, because some structure is missing, or in bad condition like corrosion.

To overcome that issue this research proposes to do inspection of accessories tower using Deep Learning, especially to detect corrosion anomaly. Because corrosion can occur in every part of the transmission tower, especially in important parts that support equipment loads and tensile loads such as clevis, dead end and clevis. Corrosion can also occur quickly in highly polluted areas such as beaches and industrial areas. So, it's expected that with this research, corrosion anomalies can be detected early so that they do not endanger the transmission network. In addition to reducing time consumption, it can eliminate different aspects among experts about what it calls corrosion, and another human error aspect.

1.2 Related Works

Detection of anomalies in transmission has been carried out several times. Like previously research that has been done by T, Mao et al. [2] using combination of DAG-SVM and HOG to inspect condition of conductor, so it can get accuracy about 84,3%. There is another research from J. Hao et al. [3] that uses SSD and HSV color space filters, so it finds another object (bird nest) in structure of tower. Another research from Z. A. Siddiqui et al. [4] that capable to detect object and condition of accessories like conductor, isolator, dumper, spacer, lighting arrester, sag adjuster, balisor, and type of tower, even it can get average precision at 86.34% by using YOLO v3 to detect object and segmentation for classification of its condition. Research also has been conducted by Y. Chen et al. [5] which examines the condition of bolt in high voltage transmission. By using Double iterative learning + sample mining + feature pyramid + deformable convolution + faster RCNN, it can obtain mAP score 0.785 with using about 7614 datasets and up to 5000 iterations. Another research from T. Su [6] that examines condition of conductor, damper, sag adjuster and bird nest. This research using Context Enhancement-SSD and obtain variation mAP score ranging from 24.96% for broken strands class to 98.05% for misplacement class with average mAP score 66.65%.

In this research we will use a combination of object detection, image classification and background removal. Like research from Zhendong Yin, Dasen Li and Zhilu Wu [7] that say background removal will increase the result. And experiments from Alexey Dzyuba [10] have resulted that using a smaller number

of classes will improve result. It is already known that image classification has simpler tasks, so they tend to have better results than object detection that have complex architecture and parameters. In line with all related work above, this research can implement and can complement previous research.

2 Dataset

In this research, will use self-processed data, that specifically obtained from high voltage transmission in around of Indonesia, covering various region like urban area that contain a lot of pollution, rural area, mountainous area, forest, beach coast that contain a lot of salinization, and industrial area that contain iron dust, and another chemical dust. Data was obtained from previous climb up inspections using various devices, such as smart phones, digital cameras, DSLRs, and drones. The data used is only clearly visible, not backlight, and not blurry. Annotation is done manually using the Roboflow system, and the occlusion value is a minimum of 50%. Due to the limited amount and imbalance data, augmentation was carried out, namely by oversampling only for classes with a small number. Fill black color to the bounding boxes in classes where there are already enough, so that class will not include when image augmented. Augmentation is done by rotating 90^{o} clockwise or counterclockwise, and flip horizontally and vertically so that 1 original dataset will produce 3 images new dataset. It can be seen in figure 1 regarding the class distribution in dataset A1.

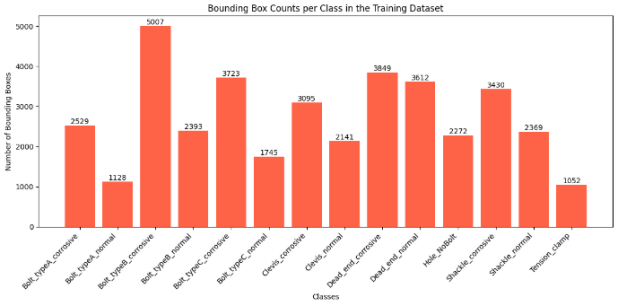

Figure 1 Distribution of class in dataset A1.

In this research, we will detect corrosion anomaly in various transmission accessories as shown in table 1. Definition of corrosion in this research is where the majority part (>60%) of accessories is contaminated corrosion, in this research not enforced level of corrosion, so only normal or corrosion class. The definition of Hole_NoBolt class is if there is an accessory that has a hole, whether due to it's a feature or because it is missing bolt which makes an anomaly, so special for this class need further examination. This occurs because an anomaly missing bolt and a hole that features in accessories (like adjusters) have the same shape, thus a model object detection cannot distinguish it. There is also Bolt class

that divides into 3 types, that is bolt_typeA that means bolt that has pin, whose function is to maintain that bolt not easily detached, then bolt_typeB that is regular bolt, and bolt_typeC which means long bolt whose function serves as a foothold in tower transmission. In this research, we contain nested class relationships that are clevis, shackle, dead end compression, and tension clamp class that have bolt_typeA and bolt_typeB class inside it.

Table 1 Types of main class and sub class

Main Class	Sub Class	Sub Class Anomaly	
		Normal	Corrosion
Clevis		✓	✓
Shackle		✓	✓
Dead_end		✓	✓
Tension_clamp		×	×
Bolt	Bolt_typeA	✓	✓
	Bolt_typeB	✓	✓
	Bolt_typeC	✓	✓
Hole-NoBolt		×	×

The original dataset in this research is raw data from the result of annotations and resize step using Roboflow. Dataset A consists of 14 classes, dataset B is like dataset A buat have removal background first in data train and data validation but not in data test, which is expected to be able to improve metric score result [7] [8] example in figure 3. Dataset C is followed up from dataset A or B which has better results and changed to only 6 classes to obtain better results [9]. Dataset D is dataset for image classification which contains only 4 classes, as shown in figure 4. The dataset split ratio is 70:20:10 for data train, data validation and data test. Details amount of dataset can be seen in table 2, and example image of each class can be seen in figure 2.

Table 2 Details of datasets

Dataset	Image	Class	Object	Train	Val	Test
A1	6120	14	38.345	4334	1192	594
A2	6076	14	26.323	4255	1211	610
B	6076	14	26.323	4255	1211	610
C	6076	6	26.323	4253	1227	596
D	24.435	4	-	Different each class		

Dataset A is divided into dataset A1 and dataset A2, selecting the best dataset by eliminating small objects. Removing small objects lower than 7 pixels for dataset A1, and less than 15 pixels for dataset A2. This is done to achieve a target minimal of metric score 90% [10].

Figure 2 1st Row: Clevis normal and corrosion (left) and Shackle normal and corrosion (right), 2nd Row: Dead End normal and corrosion (left) and Tension Clamp (right), 3rd Row: Bolt type A (left), type B (middle) and type C (right), 4th Row: Hole Feature (left) and Hole because missing bolt (right)

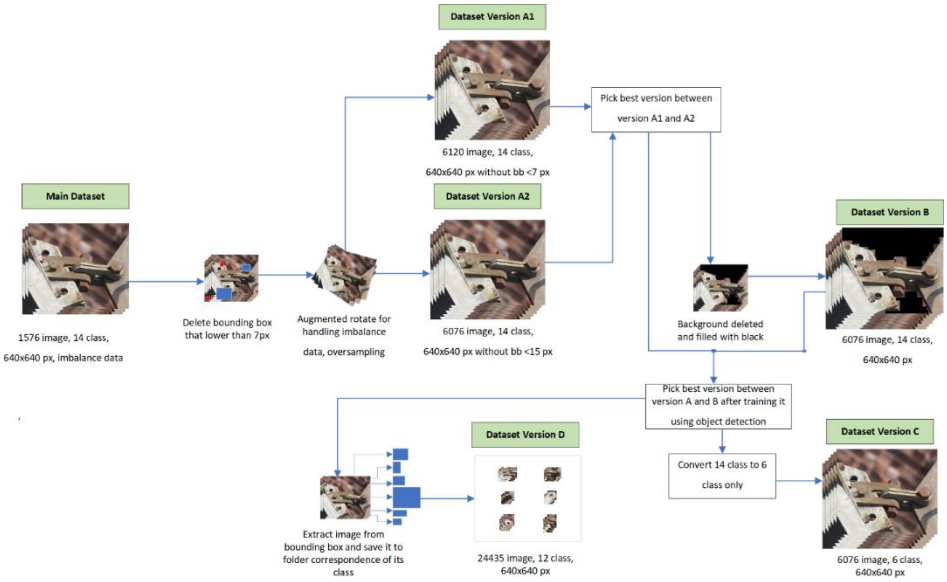

Figure 3 Flow process to create various datasets

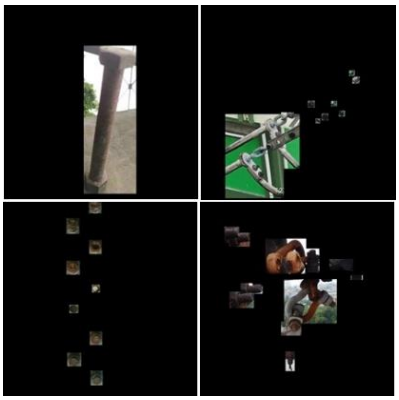

Figure 4 Random example of image from dataset B

3 Scenario of experiment

The objective of this research is to achieve the best metric score, so it will run several scenarios. The model that uses is YOLO variants from Ultralystic, because that is one of Sate of the Art baik (SotA) for object detection and image classification tasks. Especially on YOLOv8x and YOLOv9e, according to research by Saenprasert W et al. [10] that the model gets first and second place for AP values in detecting small objects. Meanwhile, YOLO v10 and v11 are

models aimed at a balance between speed and precision. RT-DETR is also used as a transformer-based object detection model, which obtains good results [11]. And using YOLOX, which has a different system from the latest YOLO variant as a comparison [12]

In this experiment will use 2 kind of scenario, scenario 1 is using only YOLO, RT-DETR or YOLOX directly for object detection with dataset A and dataset B. and scenario 2 will sing a combination of object detection from YOLO, RT-DETR or YOLOX using dataset C with image classification form YOLOv8x-cls, YOLOv11x-cls or another image classification model using dataset D. The Final metric score result of scenario 1 and scenario 2 will be compared to decide where the best is. The method to compare it is by picking the best result in scenario 1 with various models and datasets, and only using object detection method. For scenario 2, using the best model from various model in object detection step, then bounding box from that prediction will be extracted and then resize before performing image classification using best model from classification step.

Because limitations of this research are not considered about specification of hardware will be used, it will use the highest type of each version of model, because it will perform best in each version. Therefore, YOLO object detection version that use is YOLOv8x, YOLOv9e, YOLOv10x, YOLOv11x, YOLOX-X [12] and RT-DETR Extra-Large [11] Whereas for image classification we will use another State of the Art like ResNet-101, VGG-19, Inception-v3, DenseNet-201, YOLOv8x-cls and YOLOv11x-cls. In this research we will use 200 epoch with early stopping, conf is 0,001 and the majority of default parameter setting of each model.

4 Results and Discussion

The following are results of experiment from scenario 1 and scenario 2.

4.1 Scenario 1

This is the result of comparing metric score result of using dataset A1 and dataset A2.

From table 3 it can be shown that experiments using dataset A2 can achieve better metric score than using dataset A1, and from figure 5 (left) can be shown that difference is Precision 0.051, Recall 0.075, mAP@0.5 0.071, and F1 Score 0.064. Best model generated from YOLOv9e. Therefore, dataset A2 used for dataset A.

Table 3 Result of experiment using dataset A1 and A2

Model	Dataset A1				Dataset A2			
	Precision	Recall	mAP @0.5	F1 Score	Precision	Recall	mAP @0.5	F1 Score
Yolo v8x	0,882	0,787	0,853	0,832	0,878	0,827	0,892	0,852
Yolo v9e	0,874	0,817	0,870	0,845	0,925	0,892	0,941	0,908
Yolo v10x	0,849	0,789	0,843	0,818	0,887	0,834	0,891	0,860
Yolo v11x	0,877	0,806	0,869	0,840	0,908	0,882	0,928	0,895
YoloX	0,359	0,424	0,628	0,374	0,294	0,400	0,546	0,338
RT-DTR	0,855	0,778	0,822	0,815	0,858	0,843	0,868	0,850

And from table 4 it can be shown experiments using dataset A is way better than using dataset B. This makes using remove background for data training and validation then only left bounding box will not increase metric score, quite the opposite, even metric score when training and validation is good.

Table 4 Result of experiment scenario 1 using Dataset A and B

	Dataset A				Dataset B			
	Precision	Recall	mAP @0.5	F1 Score	Precision	Recall	mAP @0.5	F1 Score
YOLOv8x	0,878	0,827	0,892	0,852	0,236	0,153	0,125	0,186
YOLOv9e	0,925	0,892	0,941	0,908	0,231	0,130	0,102	0,166
YOLOv10x	0,887	0,834	0,891	0,860	0,241	0,143	0,130	0,179
YOLOv11x	0,908	0,882	0,928	0,895	0,159	0,145	0,098	0,152
YOLOX-X	0,294	0,400	0,546	0,338	0,569	0,725	0,661	0,637
RT-DETR-X	0,858	0,843	0,868	0,850	0,343	0,167	0,137	0,225

Figure 5 Difference in metrics using datasets A2 and A1 (left) and dataset A and B (right)

From figure 6 (left), there are no signs of overfitting in training and validation process. Similar patterns occurred in model YOLOv9e, YOLOv10x, and YOLOv11x. Whereas YOLOX-X obtained the best metric score result using

dataset B, even better than using dataset A. Background removal may remove important visual context that the model needs for accurate detection, especially to distinguish between positive objects and background objects, as shown in figure 6 (right).

Therefor experiments object detection in scenario 1 using dataset A1, A2, and B, the best result conducted by YOLOv9e using dataset A2 (or we called dataset A now) that has metric score Precision 0,925, Recall 0,892, mAP@0.5 0,941, and F1 Score 0,908, slightly different from the second best model, it is YOLOv11x with Precision 0.017, Recall 0.010, mAP@0.5 0.013, and F1 Score 0.013.

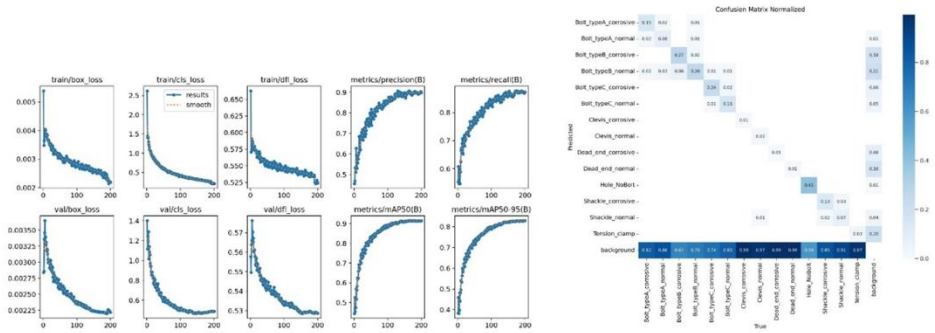

Figure 6 Visuals from training and validation process using dataset B and model YOLOv8x

4.2 Scenario 2

This is the result of experiments using dataset C object detection section in scenario 2. An overall pipeline for scenario 2 is shown in figure 7. In experiments scenario 2 object detection section using dataset C, the best result also achieved from model YOLOv9e with metric score Precision 0,947, Recall 0,947, mAP@0.5 0,972, F1 Score 0,947 and mAP@[.5:.95] 0,786 detail in table 5.

Table 5 Experiments result of scenario 2 using dataset C

	Precision	**Recall**	**mAP@0.5**	**F1-Score**	**mAP@[.5:.95]**
YOLOv8x	0,926	0,924	0,954	0,925	0,762
YOLOv9e	0,947	0,947	0,972	0,947	0,786
YOLOv10x	0,924	0,91	0,954	0,917	0,759
YOLOv11x	0,938	0,944	0,967	0,941	0,778
YOLOX-X	0,356	0,434	0,628	0,391	0,356
RT-DETR-X	0,931	0,913	0,949	0,922	0,705

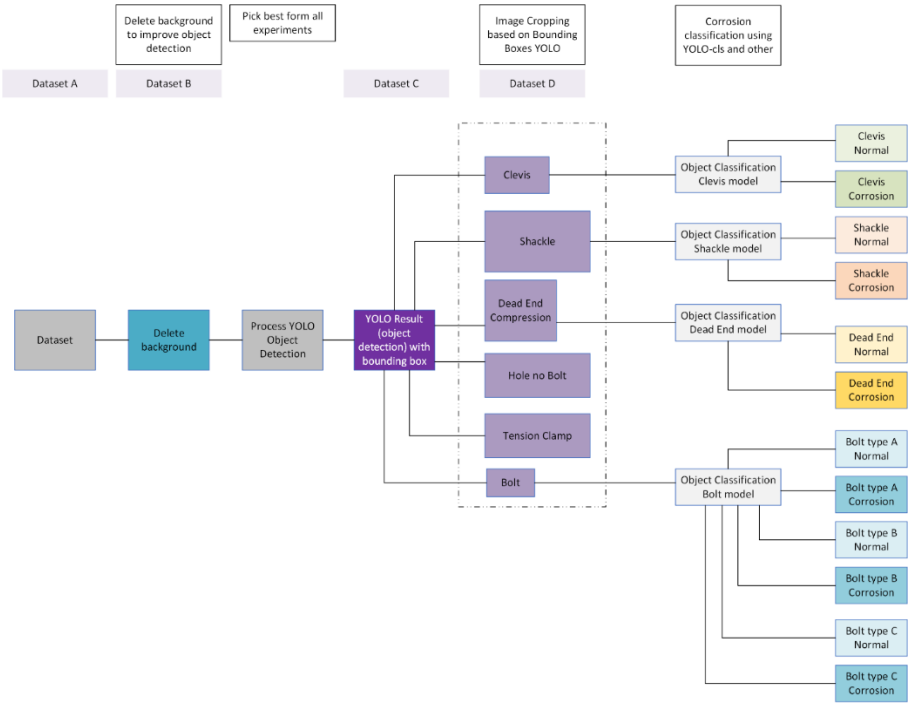

Figure 7 Pipeline of experiment scenario 2

In experiment scenario 2 image classification section using dataset D in table 6, the result can be debated who is the best, that is result from YOLOv8x or YOLOv11x. This occurs because only there are slightly different about those results. If we use T-Test and ANOVA to compare it, the results say both models perform similarly across all metrics like in table 7.

Out of this result, when viewed from number of parameters, YOLOv11x has less parameters, that is 28.4 million, and YOLOv8x have 57,4 million, where this can be considered. Therefore, in the final we will use YOLOv11x-cls model.

4.3 Comparison of scenario 1 and scenario 2

Table 8 shows the result of image classification using YOLOv11x-cls and dataset that resulting from extraction bounding box prediction by model YOLOv9e using dataset C. Because it will use object detection and image classification, only precision and recall metric than can be compared, as shown in figure 8. And if result of image classification combined with best result of object detection experiments using dataset C that is by YOLOv9e 6 class thus will make Precision and Recall like as shown in table 9.

Table 6 Result of experiment scenario 2 using dataset D

Class		YOLO v8x-cls	YOLO v11x-cls	Dense Net-201	Inception-v3	ResNet-101	VGG-19
Bolt	Accuracy	94,93	97,35	93,94	48,15	94,55	44,11
	Precision	93,67	96,98	93,94	66,48	94,61	31,52
	Recall	94,39	97,52	93,94	34,21	94,55	44,11
	F1-Score	93,97	97,23	93,88	34,97	94,5	34,99
Clevis	Accuracy	94,63	94,1	90,43	41,49	92,91	52,48
	Precision	94,64	94,09	90,43	41,36	92,94	75,39
	Recall	94,63	94,11	90,43	41,7	92,91	52,48
	F1-Score	94,63	94,09	90,43	41,03	92,9	37,61
Dead_end	Accuracy	94,26	83,54	90,77	48,62	92,42	51,24
	Precision	94,23	83,49	90,79	48,63	92,47	26,26
	Recall	94,3	83,57	90,77	49,74	92,42	51,24
	F1-Score	94,25	83,51	90,77	36,52	92,43	34,72
Shackle	Accuracy	97,38	97,05	93,46	51,31	94,44	60,13
	Precision	96,78	97,06	93,45	54,79	94,45	36,16
	Recall	97,81	96,81	93,46	54,54	94,44	60,13
	F1-Score	97,24	96,93	93,45	51,19	94,45	45,16

Table 7 Comparison results between YOLOv8x-cls and YOLOv11x-cls using ANOVA and T-Test

Metric	ANOVA F-Statistic	ANOVA P-Value	T-Test T-Statistic	T-Test P-Value
Accuracy	0,476664	0,515742	0,690409	0,515742
Precision	0,343484	0,579189	0,586075	0,579189
Recall	0,466645	0,520026	0,683114	0,520026
F1-Score	0,396214	0,552249	0,629455	0,552249

Table 8 Experiment result in scenario 2 image classification using YOLOv11x-cls and dataset result of extraction using YOLOv9e

Metric	Value per class			
	Bolt	Clevis	Dead_end	Shackle
Accuracy	81,97%	82,63%	78,49%	87,35%
Precision	74,34%	81,89%	78,39%	87,28%
Recall	82,03%	83,14%	78,45%	86,97%
F1-Score	76,95%	82,20%	78,41%	87,11%

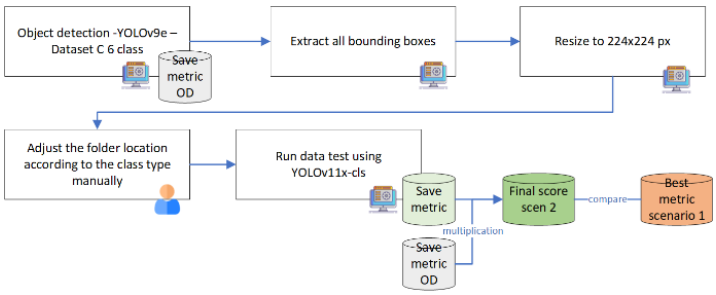

Figure 8 Pipeline to compare the results of scenario 1 and 2

Comparing scenario 1 result in table 4 and scenario 2 in table 9, it will be shown that scenario 1 result is better in all precision and recall. The difference in results is Precision 6,06% and Recall 16,78%.

Table 9 Detail result of metric score model YOLOv9e using dataset C and combined object detection YOLOv9e 6 class and image classification using YOLOv11x-cls

Class	Object Detection YOLOv9e 6 class			Combine Object Detection YOLOv9e with Image Classification YOLOv11x-cls	
	Precision	Recall	mAP@0,5	Precision	Recall
all	94,7%	94,7%	0,972	82,48%	83,94%
Bolt	95,4%	85,7%	0,946	70,92%	70,30%
Clevis	90,1%	93,5%	0,956	73,78%	77,74%
Dead_end	96,5%	98,7%	0,989	75,65%	77,43%
Hole_NoBolt	96,6%	100,0%	0,995	96,60%	100,00%
Shackle	89,3%	92,4%	0,957	77,94%	80,36%
Tension_clamp	100,0%	97,8%	0,989	100,00%	97,80%

Scenario 2 in object detection part has better results because it removes small objects and uses smaller class numbers, similarly with image classification part, but when it combines instead decreases the result. It is estimated that this is due to the quality of dataset resulting from extraction by YOLOv9e 6 class is not quite good, because setting IoU in 0,5, thus making the dataset that pass to the image classification is not quite fitting, while image classification model use IoU 1,0 when training, validation and test process, direct from manual annotations results. Or it is due to researcher that must be sorting the image resulting from extraction in right sub class so its performance can be measure have some human error, because when manual annotations using high resolution image, and it only use extraction bounding box that can only have 15 pixels.

Table 10 Comparison metric score with previous research

Researcher	Equipment / accessories	Method	Metric score
Tianqi Mao et al, 2019	conductor	SAD+HOG+PCA	accuracy 80.3%
Jiang Hao et al, 2019	bird nest	SSD+HSV	precision 98.23% Recall 67,37%
Siddiqui and Park 2020	Realtime: conductor, dumper, spacer, LA, sag adjuster, and balisor	YOLOv3 edited	precision 86.34% Recall 91.44%
	isolator	YOLOv3 edited	precision 93,42% Recall 97,47%
Yuquan Chen et al, 2021	bolt	Faster-RCNN	mAP 0.785
Chunyang Liu et al, 2023	isolator and damper	YOLOv7-CSM	precision 98,7% Recall 96,8% mAP @0.5 0.989
Our proposed scenario 1	Clevis, Dead end, Shackle, Tension clamp, Hole and Bolt	YOLOv9e	mAP @0.5 94.1% Precision 92,5%
Our proposed scenario 2		YOLOv9e + YOLOv8x-cls	Precision 83.19% Recall 84,35%

In table 10 shown comparison between the results of this research and previous research. Even though all this research cannot be compared equally, because the type of data is not the same. But all of this can be used as a general perspective about detection in transmission equipment and accessories. Especially if it only uses small class numbers, the result tends to be higher.

5 Conclusions

In this section, conclusions, limitations and suggestions for further research will be presented.

5.1 Comparison of scenario 1 and scenario 2

Removing objects that are tiny size can improve the result of experiments. If we add it with removal background in object detection, it will increase speed of training process, but the result is not improved, quite the opposite. Reducing class numbers will improve the result of experiments in object detection, even this can achieve perfect score in some classes. But when it combines with image classification, the result decreases and lower than experiment that only use object detection.

The results of this research obtained an overall good score, where the results were more than 90%, especially in scenario 1 using the YOLOv9e model. Therefore, implementation of a YOLOv9e-based automated corrosion detection system can improve inspection efficiency and reduce the risk of transmission equipment and accessories failure, potentially saving significant maintenance costs, and avoid the risk of unknown anomalies within the 5-year routine climb-up inspection period.

5.2 Limitations and future work

There are limitations in this research, particularly the very small number of datasets, which can affect the final model results. There is also a class that cannot be directly determined whether it is an anomaly or not, it is the Hole_NoBolt class. And one of the causes of scenario 2 results is that there is still manual intervention in data selection, so it is very possible for bias or human error to occur. Although the dataset covers a wide range of environmental conditions in Indonesia, seasonal variations and extreme weather conditions may not have been fully represented, which could affect model performance in certain scenarios.

Therefor future works for detecting anomalies in transmission are still wide open, there are still many equipment or accessories that have never been discussed before, like arching horn, armour rod, joint sleeve etc. Or detect corrosion levels to differentiate levels of urgency. Or focus on class Hole that can differentiate if it is an anomaly or not is quite challenging, like from its location.

6 References

[1] PT (PLN) Persero, “Laporan Tahunan PLN tahun 2022,” pp. 185–185, 2022, [Online]. Available: https://web.pln.co.id/stakeholder/laporan-tahunan

[2] T. Mao *et al.*, “Defect recognition method based on HOG and SVM for drone inspection images of power transmission line,” *2019 International Conference on High Performance Big Data and Intelligent Systems, HPBD and IS 2019*, no. 61701404, pp. 254–257, 2019, doi: 10.1109/HPBDIS.2019.8735466.

[3] J. Hao, H. Wulin, C. Jing, L. Xinyu, M. Xiren, and Z. Shengbin, “Detection of Bird Nests on Power Line Patrol Using Single Shot Detector,” *Proceedings - 2019 Chinese Automation Congress, CAC 2019*, pp. 3409–3414, 2019, doi: 10.1109/CAC48633.2019.8997204.

[4] Z. A. Siddiqui and U. Park, “A Drone Based Transmission Line Components Inspection System with Deep Learning Technique,” *Energies (Basel)*, vol. 13, no. 13, 2020, doi: 10.3390/en13133348.

[5] Y. Chen, H. Wang, J. Shen, X. Zhang, and X. Gao, "Application of Data-Driven Iterative Learning Algorithm in Transmission Line Defect Detection," *Sci Program*, vol. 2021, 2021, doi: 10.1155/2021/9976209.

[6] T. Su, "Transmission line defect detection based on feature enhancement Content courtesy of Springer Nature , terms of use apply . Rights reserved . Content courtesy of Springer Nature , terms of use apply . Rights reserved .," 2023.

[7] K. Kc, Z. Yin, D. Li, and Z. Wu, "Impacts of background removal on convolutional neural networks for plant disease classification in-situ," *Agriculture (Switzerland)*, vol. 11, no. 9, Sep. 2021, doi: 10.3390/agriculture11090827.

[8] J. Liang, Y. Liu, and V. Vlassov, "The Impact of Background Removal on Performance of Neural Networks for Fashion Image Classification and Segmentation," in *Proceedings - 2023 Congress in Computer Science, Computer Engineering, and Applied Computing, CSCE 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 1960–1968. doi: 10.1109/CSCE60160.2023.00323.

[9] A. Dzyuba, "Can you improve YOLOv5 by reducing the number of classes?" Accessed: Nov. 20, 2024. [Online]. Available: https://medium.com/@poseydon.iq/can-you-improve-yolov5-by-reducing-the-number-of-classes-6105f49764a8

[10] W. Saenprasert, E. E. Tun, A. Hajian, W. Ruangsang, and S. Aramvith, "YOLO for Small Objects in Aerial Imagery: A Performance Evaluation," in *Proceedings - 21st International Joint Conference on Computer Science and Software Engineering, JCSSE 2024*, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 720–727. doi: 10.1109/JCSSE61278.2024.10613680.

[11] Ultralytics, "Baidu's RT-DETR: A Vision Transformer-Based Real-Time Object Detector." Accessed: Nov. 28, 2024. [Online]. Available: https://docs.ultralytics.com/models/rtdetr/

[12] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding YOLO Series in 2021," Jul. 2021, [Online]. Available: http://arxiv.org/abs/2107.08430