Attention Modules Improve Efficient Anomaly Localization for Industrial Inspection in the Wild

André Luiz Vieira e Silva[1,*] Felipe Battisti[1] Veronica Teichrieb[1] Francisco Simões[1,2]
[1] Voxar Labs, Centro de Informática, Universidade Federal de Pernambuco, Brazil
[2] Visual Computing Lab, DC, Universidade Federal Rural de Pernambuco, Brazil
[*] albvs@cin.ufpe.br

Abstract

The automation of visual defect inspection has become indispensable across various industries like manufacturing, healthcare, surveillance, and power delivery due to its potential to minimize inspection costs and enhance security. However, the scarcity of annotated data presents a significant challenge, especially in scenarios where defects are rare. Recent advances in anomaly detection methods, propelled by new datasets, have paved the way for research in image- and pixel-level anomaly detection. In this paper, we focus on pixel-level anomaly detection for defect segmentation in the wild, particularly in the context of power line asset inspection. We introduce InsPLAD-seg, a novel dataset for pixel-level anomaly detection of power line components. Moreover, we propose the integration of attention modules into efficient anomaly detection methods to improve their performance in identifying defects at a high inference rate, ready for real-world inspections. Our experiments demonstrate that attention-enhanced models not only enhance image-level anomaly detection but also improve pixel-level defect segmentation. The results highlight the potential of the InsPLAD-seg dataset to stimulate further research in power line inspection and the efficacy of attention modules for industrial inspection in the wild.

1. Introduction

Recently, the automation of visual defect inspection has emerged as a critical solution in various industries such as manufacturing [13, 16], healthcare [10, 15], surveillance [1, 5], and power delivery [8, 9]. This popularity is due to the potential to reduce inspection costs and security risks, particularly in sectors where defects have severe financial and social consequences [14, 18]. However, industries often face the challenge of scarce annotated data to train deep learning models, especially when defects are rare.

Advances in anomaly detection methods, boosted by many new datasets [3, 4, 7, 19], have gained traction due to their ability to operate with limited labeled anomalous data. These new methods have opened new avenues for research on image- and pixel-level anomaly detection. However, industrial inspection in the wild poses several challenges, such as the scarcity of publicly available datasets, particularly those annotated at pixel level.

On another front, many new methods for anomaly detection methods for industrial inspection have been proposed in recent years [2, 9, 12, 14, 16], leveraging the power of deep CNNs and Vision Transformers. These methods have shown promising results in detecting defects in controlled industrial settings. However, many of them require a large amount of computational power and memory, making them unsuitable for deployment on edge devices or resource-constrained environments, which are important in in-the-wild inspections. Furthermore, the lack of interpretability in these models poses a significant challenge in understanding the decision-making process, particularly in safety-critical applications.

Our hypothesis is that attention modules can boost anomaly detection methods in the wild, not just at the image level as shown in [8], but also at the pixel level, increasing the model's anomaly localization capabilities. In this work, we propose a new dataset for defect segmentation of power line components based on the publicly available InsPLAD dataset, the InsPLAD-seg. It is an anomaly detection dataset that consists of three classes with defects annotated at the pixel level. We also experiment with attention modules as a way to improve real-time state-of-the-art anomaly detection methods [2, 12] by helping them focus on the relevant image areas since they were idealized for controlled industrial environments. In summary, the main contributions of our work are as follows.

- We propose InsPLAD-seg[1], which is, to the best of our knowledge, the first real-world dataset for industrial in-

[1] https://drive.google.com/file/d/1l1XhVdn5z2f84jIQkNQ4FAiX2wbu0Xmw

spection in the wild with pixel-level annotations.

- We show that combining attention modules with real-time state-of-the-art anomaly detection methods improves anomaly detection (image-level) and localization (pixel-level) in the wild.

2. Methodology

2.1. Anomaly Detection Methods

Here, we detail two state-of-the-art anomaly detection methods, SimpleNet and EfficientAD, and how to integrate them with the attention modules explored in this work.

2.1.1 SimpleNet

SimpleNet [12] is a state-of-the-art method for detecting and localizing anomalies in industrial inspection images. During training, local features are extracted from normal samples using a pre-trained Feature Extractor, and the Feature Adaptor aligns these extracted features with the target domain characteristics. Anomalous samples are synthesized by introducing Gaussian noise to the adapted features, and the Discriminator distinguishes normal and anomalous samples using these adapted and synthesized features. During inference, the Anomalous Feature Generator is removed for real-time detection, and ResNet-18 is used as the backbone network to optimize performance. Attention modules are strategically inserted into the ResNet-18 architecture following [8] to ensure the attention mechanisms guide the network's essential feature focus.

2.1.2 EfficientAD

EfficientAD [2] integrates multiple components: first, a patch description network (PDN) extracts features efficiently using a shallow convolutional architecture, bolstered by distillation from a deep pre-trained classification network to ensure accurate anomaly localization in each patch independently. Using a lightweight student-teacher (S-T) approach, EfficientAD employs the distilled PDN as the teacher and reuses only the PDN architecture as the student, improving anomaly detection via a novel training loss while maintaining computational efficiency at test time. Furthermore, it learns logical anomaly detection through an autoencoder, learning constraints from training images to detect deviations. By combining results from the S-T model and autoencoder, EfficientAD achieves comprehensive anomaly detection with minimal computational burden. Figure 1 shows the proposed combination of attention modules with EfficientAD to expand its application to anomaly detection and localization in the wild.

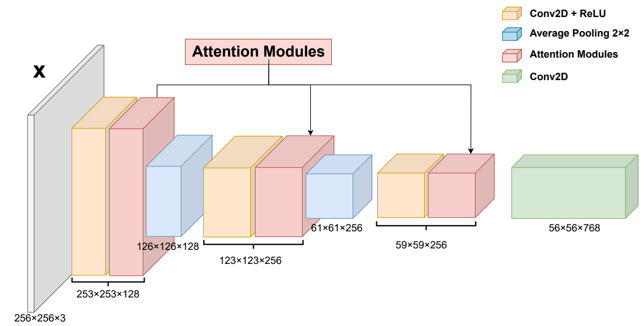


Figure 1. Attention-enhanced EfficientAD. Network architecture of the proposed version of EfficientAD's student network.

2.2. Attention Modules

SENet [11] and CBAM [17] are two attention modules that can be integrated in CNNs to improve their performance in image-related tasks. SENet uses attention mechanisms to direct the network's focus towards important areas of an image while reducing the emphasis on less significant features. CBAM incorporates two key submodules: the channel attention module and the spatial attention module, which generate channel attention maps and spatial attention maps, respectively. SENet and CBAM improve the network's representational power and performance across diverse tasks by extracting and leveraging crucial information.

2.3. InsPLAD-seg

The proposed InsPLAD-seg dataset is an adaptation of the previous InsPLAD-fault dataset [7], focusing specifically on semantic segmentation. It contains real-world inspection imagery of operating overhead power lines, including images of normal and anomalous components of different classes. It is a preliminary version that includes segmentation masks for three components: Glass Insulators, Vari-grips, and Lightning Rod Suspensions.

For the annotation process, we utilized the Roboflow annotation tool [6]. Two annotators, with domain knowledge of inspection of power line components, meticulously annotated segmentation masks for the selected object defects. Table 1 summarizes the current properties of InsPLAD-seg, while Figure 2 shows some samples of the dataset, including normal and anomalous components with their respective anomaly masks.

3. Experiments

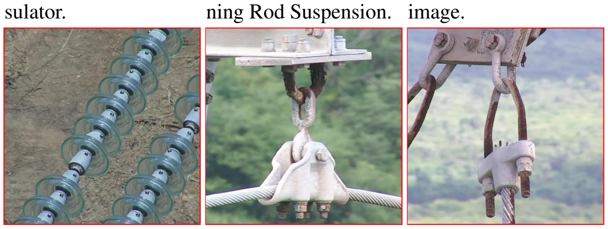
We use an NVIDIA RTX 2080 Ti GPU in a Linux OS to conduct our experiments. We apply all standard parameters from the SimpleNet[2] and EfficientAD[3] code repositories.

[2]https://github.com/DonaldRR/SimpleNet
[3]https://github.com/nelson1425/EfficientAD

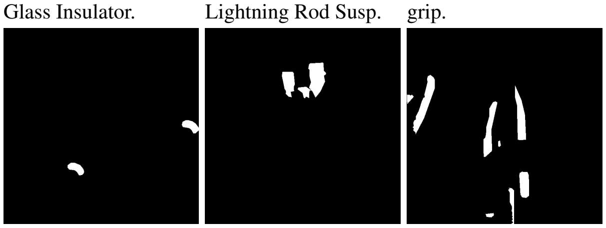
Asset category	Anomaly detection		
	Train	Test	
	Flawless	Flawless	Anomalous
Glass Insulator	1104	591	88
Lightning Rod Suspension	462	117	46
Vari-grip	477	114	40

Table 1. Description of the InsPLAD-seg anomaly detection and localization dataset. Glass Insulator anomalies are missing caps, while the rest are corrosion-related. The amount of mask annotations matches the number of anomalous images in each category.

(a) A normal Glass Insulator. (b) A normal Lightning Rod Suspension. (c) A normal Vari-grip image.



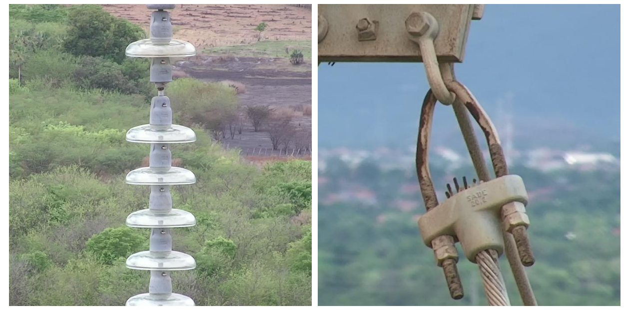
(d) An anomalous Glass Insulator. (e) An anomalous Lightning Rod Susp. (f) An anomalous Vari-grip.



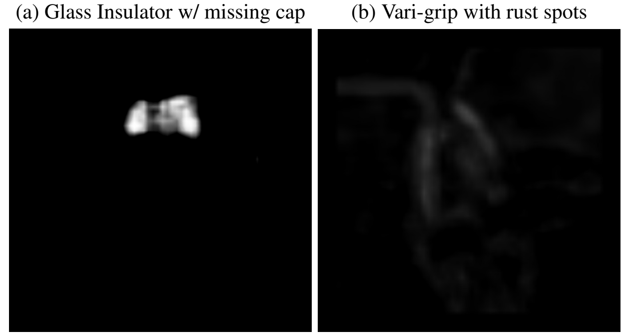
(g) Glass Insulator anomaly mask. (h) Lightning Rod Susp. anomaly mask. (i) Vari-grip anomaly mask.

Figure 2. Samples of normal and anomalous objects from InsPLAD-seg, with respective defect masks for the latter.

Quantitative Results Tables 2 and 3 present findings from the InsPLAD-seg dataset analysis. Across both image and pixel-level evaluations, attention-enhanced methods consistently outperform standard approaches in SimpleNet and EfficientAD models, with one exception for SimpleNet at the pixel level. At the image level, significant enhancements were seen, such as a 4.9% increase in the SimpleNet+SENet configuration for the Lightning Rod Suspension (rust) class. Similarly, at the pixel level, notable advancements were achieved, with SimpleNet+CBAM elevating AUPRO from 59.40% to an impressive 70.10% for the Lightning Rod Suspension (corrosion) class.



(a) Glass Insulator w/ missing cap (b) Vari-grip with rust spots



(c) Glass Insulator anomaly map (d) Vari-grip anomaly map

Figure 3. Test anomalous images and their resulting anomaly maps by EfficientAD (SENet) inferences.

Qualitative Results Figure 3 shows defective components and their respective anomaly maps inferred by EfficientAD (SENet). As in the mask annotations shown in Figure 2, darker pixels indicate normality, while lighter pixels indicate anomaly. The Vari-grip anomaly map is correctly inferred since lighter pixels appear on the handle near the upper piece junction. Interestingly, the lighter pixels in the Glass Insulator anomaly map cover the entire region where the missing cap should be, showing the potential of EfficientAD (SENet) for logical anomaly detection in the wild.

Performance analysis Table 4 shows the computational performance and necessary resources to run predictions in the analyzed and proposed methods. Clearly, the attention modules addition presents negligible impact on all evaluated aspects of computational performance.

4. Conclusions

We introduce the InsPLAD-seg dataset, the first real-world dataset with pixel-level anomaly detection annotations tailored for industrial inspection in the wild. Leveraging this dataset, we demonstrate the efficacy of integrating attention modules with state-of-the-art efficient anomaly detection methods, SimpleNet and EfficientAD. Our experiments revealed improvements in both image-level anomaly detection and pixel-level anomaly detection. In particular, for the Lightning Rod Suspension class, SimpleNet+SENet led to a

Category	SimpleNet (RN18)	SimpleNet (RN18-SENet)	SimpleNet (RN18-CBAM)	EfficientAD (RN18)	EfficientAD (RN18-SENet)	EfficientAD (RN18-CBAM)
Glass Insulator (missing cap)	78.50	81.17	78.40	79.13	81.98	79.18
Lightning Rod Suspension (corrosion)	79.70	84.60	82.70	88.33	88.25	89.48
Vari-Grip (corrosion)	79.50	82.40	78.60	85.61	89.65	85.15
Average AUROC	79.23	**82.72**	79.90	84.36	**86.63**	84.60

Table 2. Comparison of AUROC (image-level) results on the InsPLAD-seg dataset. Bold font indicates the best average results for SimpleNet variations and also for the EfficientAD variations, while underlined values show the second-best results. It should be noted that attention-enhanced methods always outperform their standard methods.

Category	SimpleNet (RN18)	SimpleNet (RN18-SENet)	SimpleNet (RN18-CBAM)	EfficientAD (RN18)	EfficientAD (RN18-SENet)	EfficientAD (RN18-CBAM)
Glass Insulator (missing cap)	58.20	59.40	53.50	49.48	58.50	50.20
Lightning Rod Suspension (corrosion)	59.46	54.44	70.10	43.74	42.01	44.65
Vari-Grip (corrosion)	62.80	52.20	60.00	49.45	46.45	47.91
Average AUROC	60.13	55.35	**61.20**	47.56	**48.99**	47.59

Table 3. Comparison of AUPRO (pixel-level) results on the InsPLAD-seg dataset. Bold font indicates the best average results for SimpleNet variations and also for the EfficientAD variations, while underlined values show the second-best results. Here, the attention-enhanced methods also outperform their standard versions.

Aspect	SimpleNet (RN18)	SimpleNet (RN18-SENet)	SimpleNet (RN18-CBAM)	EfficientAD (RN18)	EfficientAD (RN18-SENet)	EfficientAD (RN18-CBAM)
Model size (MB)	15.39	15.4	15.7	32.3	32.4	32.4
Trainable parameters ($\times 10^6$)	1.58	1.58	1.58	5.36	5.38	5.41
Inference time (ms)	87	87	87	3.4	4	5

Table 4. Performance analysis summary including the size of models, model trainable parameters, and average inference time. Clearly the impact of the attention modules addition is virtually irrelevant in terms of computational performance.

4.9% increase in image level, while SimpleNet+CBAM significantly elevated AUPRO from 59.40% to an impressive 70.10%. These findings highlight the potential of our approach for real-time applications in power line inspection. Due to the proposition of InsPLAD-seg, we expect that our research will open new avenues for future exploration in industrial inspection in the wild.

Limitations Despite the usage of attention modules, we noticed that the background still affects the model's inference to some extent. Figure 4 illustrates that problem, in which EfficientAD (SENet) mistakenly identifies a colorful part of a house in the background as an anomaly.

Future Works As future steps, we intend to:

- Increase InsPLAD-seg from three to five object types (categories) with mask defect annotations.
- Explore more efficient anomaly detection methods for attention-enhanced defect segmentation.
- Add a semantic segmentation step as a preprocessing step to remove the background in InsPLAD-seg images.

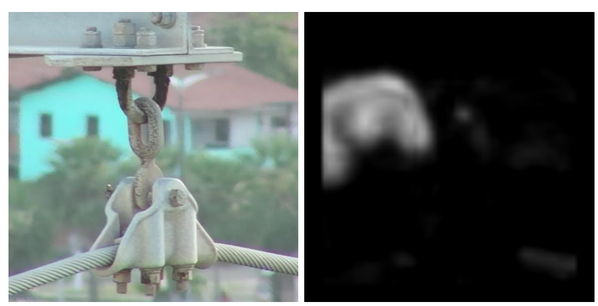


Figure 4. Incorrect anomaly map inference. It highlights part of the house in the background, probably due to its color.

- Experiment a different annotation strategy for Glass Insulator (missing cap): annotate the whole area where the missing cap should be.

References

[1] Samet Akçay, Amir Atapour-Abarghouei, and Toby P. Breckon. Skip-GANomaly: Skip Connected and Adversarially Trained Encoder-Decoder Anomaly Detection. In

2019 International Joint Conference on Neural Networks (IJCNN), pages 1–8, 2019. 1

[2] Kilian Batzner, Lars Heckler, and Rebecca König. Efficientad: Accurate visual anomaly detection at millisecond-level latencies. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 128–138, 2024. 1, 2

[3] Paul Bergmann, Kilian Batzner, Michael Fauser, David Sattlegger, and Carsten Steger. The MVTec Anomaly Detection Dataset: A Comprehensive Real-world Dataset for Unsupervised Anomaly Detection. *International Journal of Computer Vision*, 129(4):1038–1059, 2021. 1

[4] Paul Bergmann, Kilian Batzner, Michael Fauser, David Sattlegger, and Carsten Steger. Beyond dents and scratches: Logical constraints in unsupervised anomaly detection and localization. *Int. J. Comput. Vision*, 130(4):947–969, 2022. 1

[5] Thomas Defard, Aleksandr Setkov, Angelique Loesch, and Romaric Audigier. PaDiM: A Patch Distribution Modeling Framework for Anomaly Detection and Localization. In *Pattern Recognition. ICPR International Workshops and Challenges*, pages 475–489, Cham, 2021. Springer International Publishing. 1

[6] B. Dwyer, J. Nelson, and T. Hansen. Roboflow (Version 1.0) [Software], 2024. 2

[7] André Luiz Buarque Vieira e Silva, Heitor de Castro Felix, Franscisco Paulo Magalhães Simões, Veronica Teichrieb, Michel dos Santos, Hemir Santiago, Virginia Sgotti, and Henrique Lott Neto. Insplad: A dataset and benchmark for power line asset inspection in uav images. *International Journal of Remote Sensing*, 44(23):7294–7320, 2023. 1, 2

[8] André Luiz Vieira e Silva, Francisco Simões, Danny Kowerko, Tobias Schlosser, Felipe Battisti, and Veronica Teichrieb. Attention modules improve image-level anomaly detection for industrial inspection: A differnet case study. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 8246–8255, 2024. 1, 2

[9] Bangbang Ge, Chunping Hou, Yang Liu, Zhipeng Wang, and Ruiheng Wu. Anomaly Detection of Power Line Insulator from Aerial Imagery with Attribute Self-supervised Learning. *International Journal of Remote Sensing*, 42(23):8819–8839, 2021. 1

[10] Changhee Han, Leonardo Rundo, Kohei Murao, Tomoyuki Noguchi, Yuki Shimahara, Zoltán Ádám Milacski, Saori Koshino, Evis Sala, Hideki Nakayama, and Shin'ichi Satoh. MADGAN: Unsupervised Medical Anomaly Detection GAN Using Multiple Adjacent Brain MRI Slice Reconstruction. *BMC bioinformatics*, 22(2):1–20, 2021. 1

[11] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-Excitation Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 2

[12] Zhikang Liu, Yiming Zhou, Yuansheng Xu, and Zilei Wang. Simplenet: A simple network for image anomaly detection and localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 20402–20411, 2023. 1, 2

[13] Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter Gehler. Towards total recall in industrial anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14318–14328, 2022. 1

[14] Marco Rudolph, Bastian Wandt, and Bodo Rosenhahn. Same Same but DifferNet: Semi-Supervised Defect Detection With Normalizing Flows. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 1907–1916, 2021. 1

[15] Thomas Schlegl, Philipp Seeböck, Sebastian M. Waldstein, Georg Langs, and Ursula Schmidt-Erfurth. f-AnoGAN: Fast Unsupervised Anomaly Detection with Generative Adversarial Networks. *Medical Image Analysis*, 54:30–44, 2019. 1

[16] Tran Dinh Tien, Anh Tuan Nguyen, Nguyen Hoang Tran, Ta Duc Huy, Soan T.M. Duong, Chanh D. Tr. Nguyen, and Steven Q. H. Truong. Revisiting reverse distillation for anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 24511–24520, 2023. 1

[17] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. CBAM: Convolutional Block Attention Module. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018. 2

[18] Jiawei Yu, Ye Zheng, Xiang Wang, Wei Li, Yushuang Wu, Rui Zhao, and Liwei Wu. FastFlow: Unsupervised Anomaly Detection and Localization via 2D Normalizing Flows, 2021. 1

[19] Zilong Zhang, Zhibin Zhao, Xingwu Zhang, Chuang Sun, and Xuefeng Chen. Industrial anomaly detection with domain shift: A real-world dataset and masked multi-scale reconstruction. *Computers in Industry*, 151:103990, 2023. 1