A Computer Vision-Based Framework for Industrial Corrosion Detection

Bayan Abdullah Aldahlawi
Student IDs: g202317310
King Fahd University of Petroleum and Minerals
Dhahran, Saudi Arabia

Supervised by: Dr. Muzammil Behzad
muzammil.behzad@kfupm.edu.sa
King Fahd University of Petroleum and Minerals
Dhahran, Saudi Arabia

***Abstract*—This study addresses the challenge of detecting corrosion effectiveness on industrial surfaces, a critical task for ensuring structural safety and reducing maintenance costs. The research focuses on binary image classification using computer vision and advanced deep learning techniques. It evaluates the performance of Convolutional Neural Network (CNN) based ResNet models [5] in three variants: ResNet18, ResNet50, and ResNet101, with integrated architectural enhancements. The models are trained using transfer learning on the labeled Phase5 Capstone Project dataset [15]. The proposed approach implements a modified focal loss function [7], which help to improve classification accuracy and address class imbalance. Additionally, an attention block [17] is incorporated into the network architecture to support the model focus on important regions within the image, enhancing feature extraction. The enhanced model improves performance, robustness, and generalization compared to the baseline models. The results show that the ResNet50 model benefits the most from the architectural enhancements, achieving an increased accuracy of 95.60% with reduced overfitting. In contrast, while ResNet18 and ResNet101 exhibit improved regularization and reduced overfitting, their classification performance decreased with accuracies to 90.11% and 91.21%, respectively.**

***Index Terms*—Corrosion Detection, Image Classification, Deep Learning, Computer Vision, Transfer Learning, ResNet**

I. Introduction

A. Background and Significance

Corrosion is a common and critical problem faced by many sectors, such as construction and industries, where metallic structures and instruments are widely used [1]. It is caused by environmental conditions such as oxidation, moisture, temperature, and many other factors that significantly reduce the strength of steel structures [9]. As a result, there is the risk of catastrophic failures, such as bridge and building collapses. In addition, corrosion requires costly maintenance and repairs [1]. Corrosion inspection uses traditional methods, including visual inspection, radiography, ultrasonic testing, and acoustic emission testing, often fail to detect corrosion accurately at an early stage, especially in locations that are difficult to reach or hidden areas [8] [11]. These constraints create the need for more effective and reliable inspection techniques. Nowadays, deep learning based advanced image recognition methods have demonstrated effectiveness in early stage corrosion detection, providing faster, safer, and more accurate assessments compared to conventional methods [8].

In this paper, the proposed enhancement to the standard CNN architecture ResNet [5], using three variants: 18, 50, and 101, integrates a modified focal loss function [7] with Multi-Head Self-Attention (MHSA) [17] for effective training on images in corrosion classification across various types of industrial surfaces. Their performances was compared to enhance corrosion classification accuracy and improve generalization to unseen structural contexts.

B. Challenges in Current Techniques

Many researchers struggle with the limited size of corrosion datasets [9], and the lack of publicly available datasets increases the risk of overfitting. Furthermore, class imbalance remains a significant challenge, negatively affecting model performance and highlighting the need for large and diverse datasets [1]. There are also concerns about external factors that affect image data, such as weather conditions and lighting variations, which can influence model predictions [4]. These issues have been addressed in some studies by collecting large and diverse image datasets [8] or by applying data augmentation techniques [11] to increase dataset size and reduce overfitting.

Some research has performed binary classification without considering the damage size or severity of localized corrosion, which presents challenges in complex backgrounds [8]. In contrast, other work, such as [13], has investigated to estimate corrosion depth, providing a measure of severity.

Most existing studies are developed with a narrow scope, focusing on specific industrial components such as pipelines [1] or metal plates [14]. This limits the ability of the models to detect corrosion on more diverse

and complex surfaces, reducing their applicability in broader industrial contexts [4], [11].

C. Problem Statement

The primary challenge is the limited size of the dataset, which contains only 1,819 labeled images, making the model prone to overfitting. Convolutional neural networks [5] for image classification require large and diverse datasets to learn different visual patterns. To address this issue, data augmentation [3] techniques were applied to increase the dataset size by creating copies of modified versions of the original images. This approach enables the model to learn a wider range of features, thereby improving its robustness.

Another critical limitation is the class imbalance within the dataset. Using binary cross-entropy loss with ResNet models tends to focus more on the majority class, leading to biased learning and poor results on the minority class. To tackle this problem, a modified focal loss was used. It adjusts the model's learning by assigning greater weight to less frequent examples, thus emphasizing hard-to-classify samples. In addition, an attention block was incorporated into the model architecture to focus on important regions within the image, improving feature extraction and reducing background noise. These modifications help mitigate overfitting and enhance the model's ability to learn from imbalanced data.

Overall, these improvements address the key limitations of existing methods. Previous research [9] was extended by focusing on generalized corrosion detection across various structures using 2D images, excluding real-time monitoring and video analysis. By expanding the training data, improving focus on minority classes, and enhancing attention to critical regions, the proposed approach achieves more accurate across diverse industrial surfaces.

D. Scope of Study

Corrosion occurs in a wide range of structures, including bridges, towers, railways, industrial buildings, support frames, and pipelines, affecting various industries such as construction, marine, oil and gas, and vehicle manufacturing [9] . The application of 2D image based deep learning models for early corrosion detection in these structures is crucial, as it reduces repair costs maintenance, extends the lifespan of assets, improves operational safety, and minimizes environmental risks [1].

II. Literature Review

A. Overview of Existing Techniques

With the evolution of artificial intelligence, deep learning has become important across multiple tasks, outperforming traditional techniques [8]. It demonstrates strong performance in detecting and classifying objects that are difficult to access compared to manual inspection or conventional methods, because the neural networks have ability to learn complex visual patterns and features [11]. One of the most widely used deep learning architectures in image analysis is the Convolutional Neural Network (CNN) [5], which processes spatial data using multiple convolutional layers. CNNs have demonstrated remarkable performance in image classification, object detection, and segmentation tasks.

Recent studies have explored enhancements to standard CNN architecture by integrating recent and more efficient models to improve performance. For example, EfficientNetB0 [9] is a powerful model used for image classification and feature extraction. It predicts the class label of an image to differentiate between corrosion and non-corrosion surfaces with high accuracy while maintaining computational efficiency. Alternatively, YOLOv8 [9] is an object detection model capable of identifying and localizing specific objects within an image, such as corrosion areas, using bounding boxes. This model contributes to enhanced accuracy and precision in corrosion detection, reducing false positives compared to YOLOv5.

B. Related Work

The study [9] presents a deep learning approach using Convolutional Neural Networks (CNN) by applying three models: The first model is a CNN designed and trained from scratch, while the other two models, YOLOv8 for object detection and EfficientNetB0 for image classification, were pretrained and fine-tuned to detect general corrosion in industrial images across various surfaces. Compared to other studies that concentrate on specific areas and industrial parts such as pipelines [2], [6], and bolts [16]. The models were trained on a combined dataset consisting of an open-source augmented dataset (Phase5-Capstone-Project) [15] and additional industrial images contributed by the authors of a related study [1], resulting in a total of 1,000 images. However, the models were tested primarily on oil and gas images, which may limit their generalization to other industries. They introduced an anchor free architecture on YOLOv8 and a decoupled detection head, demonstrating improved precision and adaptability in detecting industrial corrosion compared to earlier YOLO versions. For EfficientNetB0, they replaced the original classification layer with a new output layer tailored specifically for the binary corrosion detection task and employed a compound scaling method and lightweight architecture, enabling efficient feature extraction and achieving perfect classification performance in binary corrosion detection. The study aims to enhance the performance of

corrosion detection evaluated using accuracy, precision, recall, and F1-score, automate the inspection process, and reduce human error. By comparing these three models, the results showed that both the custom CNN and EfficientNetB0 achieved 100% accuracy, 100% precision, 100% recall, and 100% F1-score, while YOLOv8 achieved 95% accuracy, 100% precision, 90% recall, and a 94.74% F1-score.

The authors [8] introduced a deep learning approach for the detection of pitting corrosion in gas pipelines, focusing on identifying deep seated depressions or holes in surface images, rather than detecting various grades of rust as done in other studies [1] [18] [14] . The main challenge addressed in this work is the difficulty of detecting small corrosion defects that resemble ordinary surface irregularities. To solve this, they used localization and binary classification task to differentiate between corrosion and benign surface alterations. The dataset was collected by the authors themselves through 576,000 images extraction from site visits and pipeline inspection videos. To handle the class imbalance, they applied class weighting and label smoothing. After facing overfitting during early experiments, they incorporated Global Average Pooling (GAP) instead of fully connected layers to reduce the number of parameters and minimize overfitting, compared to ZFNet. The hyperparameters of the CNN were then optimized using Bayesian optimization techniques. Then, the custom CNN was compared against several standard CNN architectures including AlexNet, ZFNet, VGGNet, Inception, ResNet, and Xception and achieved superior performance, with an accuracy of 98.44%, an F1 score of 97.30%, a ROC AUC of 0.99, and a PR AUC of 0.99.

This paper [4] investigated the effect of U-shaped encoder–decoder neural networks with three different backbones: ResNet34, DenseNet121, and EfficientNetB7 within a UNet architecture for corrosion detection in steel structures. A new dataset was created, consisting of 300 manually annotated images of steel structures, primarily due to the limited availability of labelled datasets in civil infrastructure applications. To address the limited dataset size, the Python Albumentations library [3] was used for data augmentation, increasing the dataset to 4000 images with corresponding masks. Additional images collected using UAS and from the internet were used for testing. On EfficientNetB7, they implemented a compound scaling method on the baseline network, which optimizes performance by uniformly scaling the network's depth, width, and resolution. In the case of ResNet34, they used residual blocks with skip connections to help mitigate the vanishing gradient problem, for more stable training in deeper networks. For DenseNet121, they applied dense connectivity, where each layer is connected to all previous layers, improving gradient flow, reducing redundant feature learning, and enabling more efficient training. The authors compared pre-trained with transfer learning. The results showed that models using transfer learning performed significantly better than pre-trained models. The study highlighted those deep networks, such as EfficientNetB7, suffer from performance degradation when trained from scratch due to insufficient data, but transfer learning effectively mitigates this issue. Pre-trained DenseNet121 and EfficientNetB7 yielded average pixel-level accuracies of 96.78% and 98.5%, respectively.

The authors [11] implemented a fine-tuned YOLOv5 object detection model to detect and classify corrosion in two metal types: copper and iron. The model also considered the amount of corroded area for each metal. Nonetheless, the scope of the study was limited to these two types. They collected and labeled 411 images, which constitutes a relatively small dataset, potentially leading to poor generalization and a higher risk of overfitting. As a result, the model achieved an accuracy of 89.56%.

Table I provides a summary and comparison of the hyperparameter settings used for model training across corrosion detection studies.

C. Limitations in Existing Approaches

The paper [9] studied the use of recent deep learning models, including EfficientNetB0 and YOLOv8, for training in general corrosion. However, the limitation was that the test was performed only on oil and gas images, which constrains applicability across industries and unseen structures, and increases the risk of overfitting. In this study, the models will be tested on images from different sectors to evaluate generalization and true performance.

D. Contributions

Considering that image data has a more complex structure, it requires significant computation using deep network architectures. In this paper, the models are explored for corrosion detection: the standard CNN-based ResNet model [5], which learns residual functions instead of unreferenced mappings, helping to reduce both training and validation errors. This improves performance by enhancing accuracy and enabling better generalization, especially in very deep networks.

Deep learning methods typically require large-scale image datasets; however, existing corrosion datasets are often limited in size. To address this issue, data augmentation techniques [10] are applied to generate synthetic images, which help reduce overfitting and improve class balance, ultimately enhancing model performance for corrosion classification.

TABLE I
SUMMARY OF MODEL CONFIGURATIONS, HYPERPARAMETERS, AND DATASET SPLITS USED ACROSS CORROSION DETECTION STUDIES

Paper	Models	Image Size	Batch Size	Epochs	Optimizer	Loss Function	LR
[9] (Farooqui et al.)	CNN (custom)	NA	22	NA	Adam	Binary Cross-Entropy	0.001
	YOLOv8	NA	44	NA	AdamW	NA	0.001
	EfficientNetB0	NA	22	30	Adam	Categorical Cross-Entropy	0.001
[8] (Malashin et al.)	Custom CNN	224×224	64	NA	RMSprop	Class weighting	1×10^{-5}
[11] (Mundada et al.)	YOLOv5s	416×416	150	7500	NA	GIoU, objectness, and classification losses	NA
[4] (Das et al.)	UNet	227×227	16	25	Adam	Binary Cross-Entropy	0.001
	ResNet34	227×227	16	25	Adam	Binary Cross-Entropy	0.001
	DenseNet121	227×227	16	25	Adam	Binary Cross-Entropy	0.001
	EfficientNetB7	227×227	16	25	Adam	Binary Cross-Entropy	0.001

III. PROPOSED METHODOLOGY

A. Existing Model and Challenges

The Deep Residual Learning for Image Recognition (ResNet) model is a deep convolutional neural network (CNN) that incorporates shortcut connections. Introduced by He et al. (2015) [5], it is derived from the VGG architecture [12]. The model was developed to improve optimization by using skip connections, which help prevent the degradation of training accuracy as network depth increases and mitigate the problem of vanishing or exploding gradients in very deep networks.

However, experiments on extremely deep models applied to small datasets showed signs of overfitting. Another issue observed is that projection layers, which are used to adjust dimensional in residual blocks, can increase model size, computational cost, and time complexity.

In the original study, the researchers investigated ResNet variants with 18, 34, 50, 101, and 152 layers. This work compares the performance of ResNet18, ResNet50, and ResNet101 for corrosion detection.

B. Proposed Enhancements

Focal loss: class imbalance on the dataset, is the major issue for obstacle to evaulate the model. Base model use cross entropy loss tend to focus more on the majority class, which causes overfitting and poor results on the minority class. To mitigate this issue, modify focal loss was used instead [7]. It adjusts the amount the model learns from each sample by giving more weight to less frequent examples, resulting in more emphasis on hard-to-classify examples. Furthermore, applied adjustment on the loss function to improve the performance.

Attention block: On the base model, all spatial regions learn patterns, and extract features equally in each layer, which may effect on noise background of image and mislead the result. An attention block [17] was added to the model architecture to help the network to focus on important regions in the image by assign higher weight , improving feature learning and reducing irrelevant noise.

These modifications help combat overfitting and improve the model's ability to learn from imbalanced data, ultimately resulting in higher accuracy and better generalization. The scaled dot-product attention is computed as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

where Q, K, and V are the query, key, and value matrices, and d_k is the dimension of the key vectors.

In the Multi-Head Attention setup, multiple attention heads are computed independently and then concatenated:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \ldots, \text{head}_h)W^O$$
$$\text{where} \quad \text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \quad (2)$$

C. Algorithm and Implementation

In this paper, fine-tuning was applied to a pretrained ResNet model by freezing the pretrained layers, including the residual blocks and the average pooling layer, to prevent backpropagation through them reducing computational cost and making training faster and more memory efficient. A custom classification head was added, incorporating a Multi-Head Attention block to help the network focus on significant regions within the image. The ReLU activation function was used after the fully connected layers to enable the model to learn nonlinear patterns. A dropout rate of 0.5 was applied to randomly deactivate half of the neurons during training, helping to reduce overfitting. For the output layer, a sigmoid activation function was used for binary classification, instead of the softmax function that used for multi-class tasks. Additionally, the binary cross-entropy loss was replaced with a modified focal loss to better handle class imbalance and improve model optimization. Table II illustrates the enhanced ResNet architecture with the proposed modifications.

D. Loss Function and Optimization

The original focal loss, introduced by Lin et al. (2017) [7], is designed to address class imbalance by focusing on hard-to-classify examples and down-weighting the

TABLE II
SUMMARY OF CUSTOM RESNET ARCHITECTURE WITH FOCAL LOSS AND ATTENTION

Stage	Output Size	Operation / Component	Source	Notes
Input	128×128×3	Image Input	–	Resized and normalized using ImageNet statistics
conv1	∼64×64×64	7×7 Conv, stride 2	Pretrained	From ResNet backbone
conv2_x to conv5_x	∼64×64 to 4×4×512	Residual Blocks (18 / 50 / 101)	Pretrained	All backbone layers frozen
Attention Block	4×4×512	Multi-Head Self-Attention (MHSA)	Custom	Applied in **ResNet18, ResNet50, and ResNet101**
Global Pooling	1×1×512	Adaptive Average Pooling	Pretrained	Converts feature map to feature vector
FC Layer 1	512	Dense → ReLU → Dropout	Custom	First stage of classification head
FC Layer 2	256	Dense → ReLU → Dropout	Custom	Second stage of classification head
Output Layer	1	Dense → Sigmoid	Custom	Binary output for corrosion classification
Loss Function	–	Modified Focal Loss	Custom	Replaces BCE to handle class imbalance and overfitting

contribution of easy examples using the factor $(1-p_t)^\gamma$. It is defined as follows:

$$\mathcal{L}_{\text{focal}} = -\alpha_t (1-p_t)^\gamma \log(p_t) \tag{3}$$

- α_t is the class-balancing factor
- γ is the focusing parameter
- p_t is the predicted probability for the correct class:

$$p_t = \begin{cases} p & \text{if } y = 1 \\ 1-p & \text{if } y = 0 \end{cases}$$

The proposed modified focal loss introduces an extra multiplicative penalty term $\log\left(1+\frac{1}{p_t}\right)$ to handle the issue of overconfident incorrect predictions. Consequently, it improves recall on minority classes and enhances the model's robustness in imbalanced classification settings. The modified focal loss is formulated as:

$$\mathcal{L}_{\text{modified}} = -\alpha_t (1-p_t)^\gamma \log(p_t) \cdot \log\left(1+\frac{1}{p_t}\right) \tag{4}$$

Here:

- α_t is the class-balancing factor
- γ is the focusing parameter
- p_t is the predicted probability for the true class:

$$p_t = \begin{cases} p & \text{if } y = 1 \\ 1-p & \text{if } y = 0 \end{cases}$$

- The term $\log\left(1+\frac{1}{p_t}\right)$ increases the penalty when the model is confidently wrong.

For optimization, the Adam optimizer is utilized to minimize the loss function during training, with a learning rate of 1×10^{-4} to avoid overshooting the minima. The model is trained for 100 epochs to ensure stable convergence given the small dataset size.

IV. EXPERIMENTAL DESIGN AND EVALUATION

A. Datasets and Preprocessing

Dataset: In this experiment, the corrosion classification model was trained on the Phase5 Capstone Project dataset [15], which consisting of diverse samples of corrosion across multiple infrastructure types, including steel corrosion, ship corrosion, propeller damage, car corrosion, oil and gas pipeline corrosion, concrete rebar degradation, water tank corrosion, and stainless-steel corrosion. It is a binary labelled dataset with two classes: "CORROSION" and "NOCORROSION". Figure 1 provides an overview of sample images from the dataset, illustrating both corrosion and non-corrosion examples.

Fig. 1. Sample images from the dataset showing corroded and non-corroded industrial surfaces

The original dataset contains 3,624 images; however, some were found corrupted and excluded. The final dataset consists of 1,819 images (990 CORROSION and 829 NOCORROSION). Figure 2 shows the number of images per class. Following that, the dataset was split into training (70%), validation (20%), and testing (10%) sets, and stored in three separate folders. Table III summarizes the number of images in each split.

TABLE III
DATASET DISTRIBUTION ACROSS TRAIN, VALIDATION, AND TEST SPLITS

Split	CORROSION	NOCORROSION	Total
Train	693	580	1273
Validation	198	166	364
Test	99	83	182
Total Images	**990**	**829**	**1819**

Due to the small dataset size, data augmentation [3] was applied to increase dataset diversity and improve model generalization by using the five techniques, including random rotation, width and height shift, zoom, and horizontal flip. Figure 3 shows an example of augmented image.

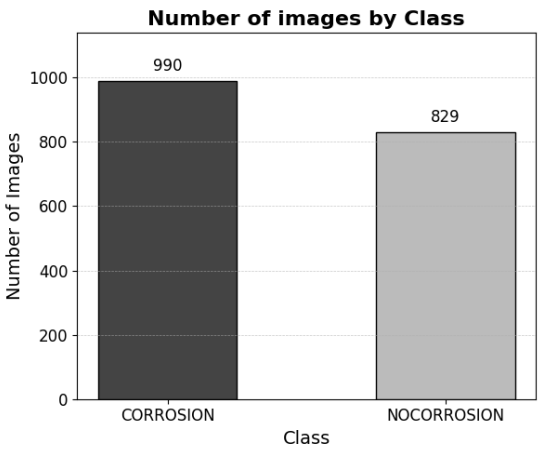

Fig. 2. Class distribution of the dataset showing the number of images labeled as *CORROSION* and *NOCORROSION*

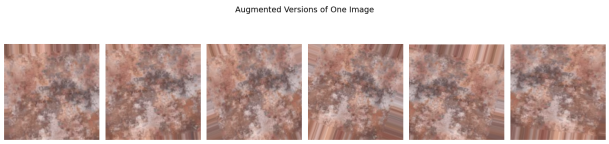

Fig. 3. Sample augmented image

Preprocessing: After reviewing the dataset, it was observed that the images varied in size. All images were resized to a fixed resolution of 128×128 to ensure the model process them correctly, training the batch efficient, and avoid errors. The batch size was set to 128 images per iteration to optimize memory usage and speed up training. Images were normalized using the preprocess_input function in Keras, which converts images from RGB to BGR and subtracts the ImageNet mean pixel values from each channel. In PyTorch, normalization was performed using the mean and standard deviation of each color channel (RGB) from the ImageNet dataset. This normalization helps scale the data to have unit variance and centers it, enabling the model to learn more effectively and improving training stability.

B. Performance Metrics

The goal of the ResNet model is to evaluate the accuracy of different architectures: ResNet101, ResNet50, and ResNet18 and analyze their behavior by comparing training and validation loss curves to identify the version with the highest accuracy and reduced overfitting. This supports the model's ability to generalize and classify corrosion efficiently on unseen data. The accuracy formula is given by:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

C. Experiment Setup

The experiments were conducted using three ResNet architectures [5]: ResNet18, ResNet50, and ResNet101, all trained on the Phase5 Capstone Project dataset [15]. A random seed was set to 1 before training to ensure reproducibility and avoid variation in results due to random initialization. Each model was fine-tuned from a pretrained base by freezing the early and intermediate convolutional layers. The architecture was modified by adding custom layers, including a Multi-Head Self-Attention block and fully connected layers associated with ReLU activation. A modified focal loss function was employed to improve performance, particularly on imbalanced data. For regularization, Dropout with a rate of 0.5 was applied. For optimization, the Adam optimizer was used with a learning rate of 1×10^{-4}, and a sigmoid activation function was applied for binary classification. Each model was trained for 100 epochs to ensure stable convergence.

ResNet50 and ResNet101 were implemented using the Keras framework, while ResNet18 was built using PyTorch due to the lack of direct support for the 18 layer in Keras. All models were trained using consistent dataset splits, preprocessing steps, learning rate, batch size, and optimization strategy to ensure a fair comparison. Figure 4 illustrates the proposed methodology.

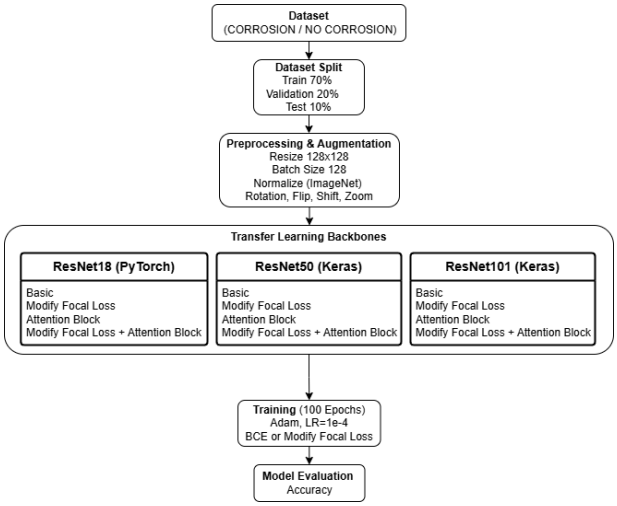

Fig. 4. Methodology

D. Results Comparative Analysis

The three variants of ResNet were analyzed by comparing training and validation loss curves under different architectural enhancements: the base model, the model with modified focal loss only, the model with an attention block only, and the model with both modified focal loss and an attention block. Figure 5 summarizes the results

across all configurations, illustrating the training and validation loss trends for each structural improvement.

ResNet18: In the basic model, after 20 epochs, the training and validation loss curves begin to move apart. The training loss continues to decrease slowly, while the validation loss fluctuates and increases, indicating memorization model. When modified focal loss is applied, the gap between training and validation loss is reduced, revealing better generalization. This improvement is due to the ability of the loss function to effectively handle class imbalance. In contrast, adding an attention block alone leads to poor generalization. After 20 epochs, the training loss decreases while the validation loss increases significantly, showing high variance. This is likely due to the shallow depth of the model, which causes it to over focus on the training data. By combining modified focal loss with the attention block, the model shows better generalization than the attention only version. Training becomes more stable with the lowest training loss, and validation loss initially improves, though it still fluctuates slightly, because of the increased model complexity.

ResNet50: Both the base model and with the attention block show signs of overfitting, with a gap between training and validation loss. However, the base model maintains relatively stable validation loss, while the attention block version shows an upward trend. Despite this, both perform better than ResNet18, resulting from the increased number of layers. Applying the modified focal loss leads to more stable performance and improved generalization, particularly for minority classes, effectively reducing overfitting. The combination of focal loss and attention block further improves performance, leading to lower loss and better model behavior.

ResNet101: This model demonstrates behavior similar to ResNet18 and ResNet50 but achieves significantly better performance, which is attributed to its deeper architecture. The additional layers enhance both generalization and learning capacity.

Accuracy is used to evaluate the classification performance and efficiency of the models. Table IV summarizes the accuracy results across ResNet architectures under different architectural enhancements. All models demonstrate relatively high performance, but variations are observed depending on the applied enhancements.

ResNet18: The model shows lower accuracy compared to deeper architectures due to its limited number of layers. The base model achieves 0.9286 accuracy. However, applying focal loss or attention block individually results in decreased performance with 0.9231 and 0.9176, respectively. Using both focal loss and attention results in the accuracy dropping further to 0.9011. The outcomes are because ResNet18 has limited capacity to model complex features, and the added enhancements introduce unnecessary complexity.

ResNet50: While applying focal loss does not change the base accuracy and remains at 0.9451, it contributes to reducing overfitting through the loss curve. Introducing the attention block alone improves accuracy to 0.9505 due to its ability to focus on important regions and filter out noisy backgrounds, leading to enhanced context-aware feature extraction. Combining both focal loss and attention block yields the highest accuracy of 0.9560. By emphasizing hard samples and enhancing feature focus.

ResNet101: The base model performs at 0.9451, similar to ResNet50. However, applying architectural enhancements either individually or together, leads to decreased performance: 0.9231 with focal loss, 0.9341 with attention block, and 0.9121 when both are applied. This decline is likely due to over regularization and increased model complexity, which can negatively impact performance when training on a relatively small dataset.

TABLE IV
ACCURACY COMPARISON OF RESNET ARCHITECTURES WITH DIFFERENT ENHANCEMENTS

Model	Architectural Enhancement	Accuracy
ResNet18	Basic Model	0.9286
	Focal loss	0.9231
	Attention Block	0.9176
	Focal loss + Attention Block	0.9011
ResNet50	Basic Model	0.9451
	Focal loss	0.9451
	Attention Block	0.9505
	Focal loss + Attention Block	0.9560
ResNet101	Basic Model	0.9451
	Focal loss	0.9231
	Attention Block	0.9341
	Focal loss + Attention Block	0.9121

E. Ablation Study

An ablation study was conducted across three ResNet variants: 18, 50, and 101, using four configurations: the baseline model, the model with custom focal loss only, the model with an attention block only, and the model with both enhancements. As shown in Table IV, the ResNet50 model showed the most balanced and effective integration of both enhancements. While focal loss alone stabilized generalization without affecting accuracy. Attention improved feature focus and led to a modest accuracy gain. Their combination produced the highest accuracy at 95.60% and the most stable loss curves, demonstrating that the two enhancements work together to improve generalization and accuracy. While ResNet18 lacks sufficient capacity to benefit from the enhancements, and ResNet101 may exhibit reduced performance resulting from an overly complex architecture.

V. EXTENDED CONTRIBUTIONS

This research aims to improve CNN architectures by integrating focal loss and attention blocks. The proposed

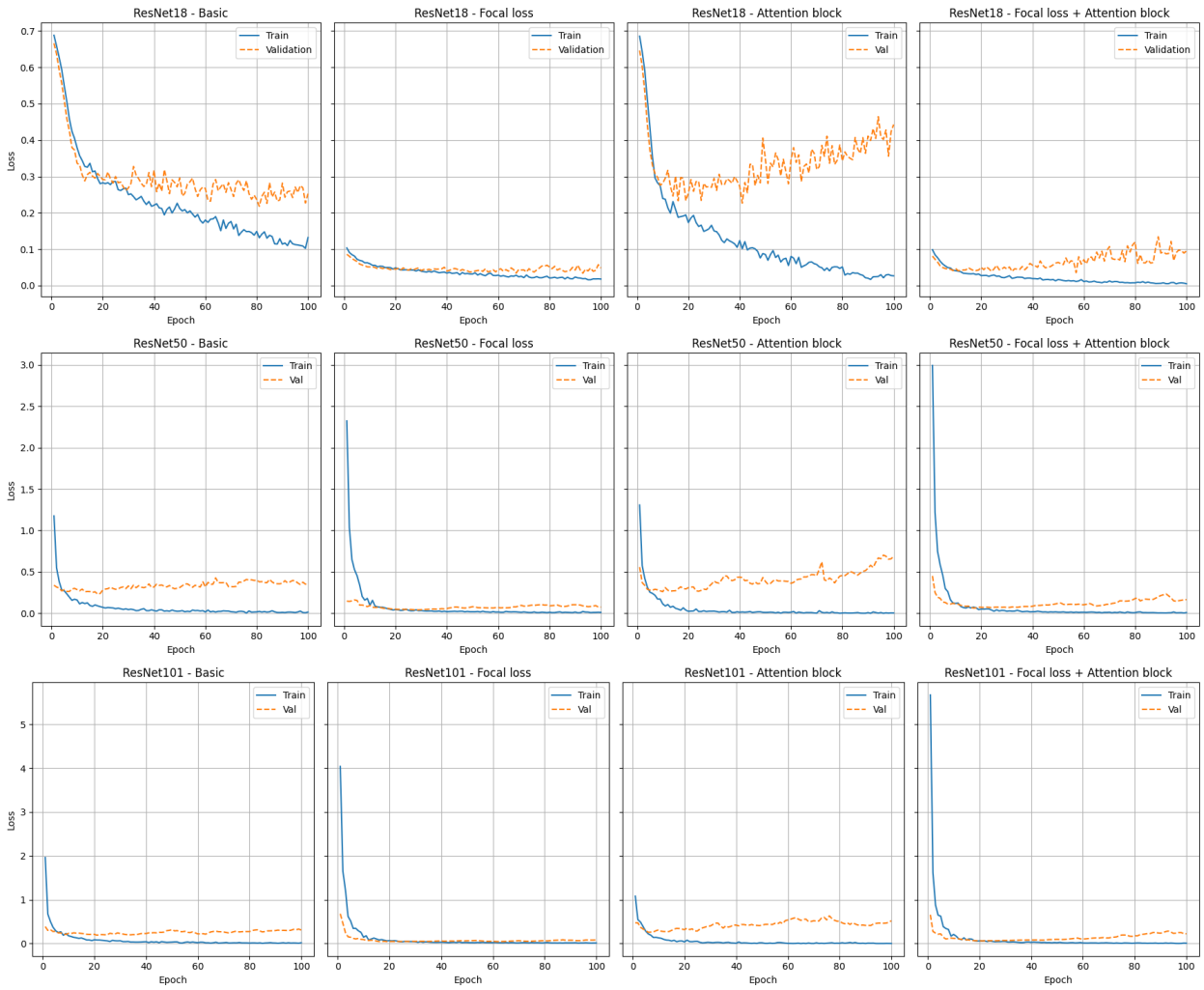

Fig. 5. Comparison of training and validation loss curves across ResNet architectures with different enhancements

approach enhances classification accuracy on imbalanced datasets and helps reduce overfitting. The study also highlights how model depth influences performance outcomes. Additionally, it provides a flexible model design that can be adapted to various industrial tasks, such as corrosion detection and defect localization.

VI. Conclusion and Future Work

The proposed paper investigates the original ResNet architecture [5] and explores performance improvements by modifying the design through the integration of custom focal loss and Multi-Head Self-Attention mechanisms. The results demonstrate that the adjusted focal loss helps address class imbalance and reduces overfitting by narrowing the gap between training and validation loss, resulting in better generalization of the model on unseen data. Incorporating attention alongside focal loss enhances the model by focusing on specific important regions and minimizing noise.

The results show that the ResNet50 model benefits the most from these architectural enhancements, achieving an accuracy of 95.60% with stable training and validation loss curves and no overfit across 100 epochs. In contrast, Although ResNet18 and ResNet101 exhibit improved regularization and reduced overfitting, their classification performance declines, with accuracy dropping to 90.11% and 91.21%, respectively. Therefore, ResNet50 provides a balanced architecture that effectively leverages the enhancements, resulting in the best overall performance.

In future work, these architectural enhancements can be extended to other models such as EfficientNet, or Vision Transformers, and integrating techniques such as advanced data augmentation, or lightweight attention modules could further boost performance and general-

ization across diverse datasets.

REFERENCES

[1] Blossom Treesa Bastian, Jaspreeth N, S. Kumar Ranjith, and C.V. Jiji. Visual inspection and characterization of external corrosion in pipelines using deep neural network. *NDT E International*, 107:102134, 2019.

[2] Venkatasainath Bondada, Dilip Kumar Pratihar, and Cheruvu Siva Kumar. Detection and quantitative assessment of corrosion on pipelines through image analysis. *Procedia Computer Science*, 133:804–811, 2018. International Conference on Robotics and Smart Manufacturing (RoSMa2018).

[3] Alexander Buslaev, Vladimir I. Iglovikov, Eugene Khvedchenya, Alex Parinov, Mikhail Druzhinin, and Alexandr A. Kalinin. Albumentations: Fast and flexible image augmentations. *Information*, 11(2), 2020.

[4] Amrita Das, Sattar Dorafshan, and Naima Kaabouch. Autonomous image-based corrosion detection in steel structures using deep learning. *Sensors*, 24(11), 2024.

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

[6] Iason Katsamenis, Eftychios Protopapadakis, Anastasios Doulamis, Nikolaos Doulamis, and Athanasios Voulodimos. Pixel-level corrosion detection on metal constructions by fusion of deep learning semantic and contour segmentation. In *Advances in Visual Computing: 15th International Symposium, ISVC 2020, San Diego, CA, USA, October 5–7, 2020, Proceedings, Part I*, page 160–169, Berlin, Heidelberg, 2020. Springer-Verlag.

[7] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2):318–327, 2020.

[8] Ivan Malashin, Vadim Tynchenko, Vladimir Nelyub, Aleksei Borodulin, Andrei Gantimurov, Nikolay V. Krysko, Nikita A. Shchipakov, Denis M. Kozlov, Andrey G. Kusyy, Dmitry Martysyuk, and Andrey Galinovsky. Deep learning approach for pitting corrosion detection in gas pipelines. *Sensors*, 24(11), 2024.

[9] Latifa Alsuliman Zainab Alsaif Fatimah Albaik Cadi Alshammari Razan Sharaf Sunday Olatunji Sara Waslallah Althubaiti Hina Gull Mehwash Farooqui, Atta Rahman. A deep learning approach to industrial corrosion detection. *Computers, Materials & Continua*, 81(2):2587–2605, 2024.

[10] Agnieszka Mikołajczyk and Michał Grochowski. Data augmentation for improving deep learning in image classification problem. In *2018 International Interdisciplinary PhD Workshop (IIPhDW)*, pages 117–122, 2018.

[11] Kapil Mundada, Mihir Kulkarni, Pranjali Sagar, Rohit Asegaonkar, and Nikhil Mulgir. Corrosion detection using deep learning and custom object detection. In *2022 2nd Asian Conference on Innovation in Technology (ASIANCON)*, pages 1–5, 2022.

[12] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014.

[13] Eun-Young Son, Dayeon Jeong, and Min-Jae Oh. Corrosion area detection and depth prediction using machine learning. *International Journal of Naval Architecture and Ocean Engineering*, 16:100617, 2024.

[14] Afzal Ahmed Soomro, Ainul Akmar Mokhtar, Jundika Candra Kurnia, Najeebullah Lashari, Umair Sarwar, Syed Muslim Jameel, Muddasser Inayat, and Temidayo Lekan Oladosu. A review on bayesian modeling approach to quantify failure risk assessment of oil and gas pipelines due to corrosion. *International Journal of Pressure Vessels and Piping*, 200:104841, 2022.

[15] P. Sun. Phase 5 capstone project dataset. https://github.com/pjsun2012/Phase5_Capstone-Project, 2023. Accessed: April 2025.

[16] Lei Tan, Xiaohan Chen, Dajun Yuan, and Tao Tang. Dsnet: A computer vision-based detection and corrosion segmentation network for corroded bolt detection in tunnel. *Structural Control and Health Monitoring*, 2024:1–16, 02 2024.

[17] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 6000–6010, Red Hook, NY, USA, 2017. Curran Associates Inc.

[18] Bingqin Wang, Yunquan Mu, Faming Shen, Renzheng Zhu, Yiran Li, Chao Liu, Xuequn Cheng, Dawei Zhang, and Xiaogang Li. Identification of corrosion factors in blast furnace gas pipe network with corrosion big data online monitoring technology. *Corrosion Science*, 230:111906, 2024.