Residual Conditional Expert Block for Adaptive Weld Defect Recognition in Radiographic NDT

Janga Bharat Reddy
Ingenious Research Solutions Pvt. Ltd
Noida, Uttar Pradesh, India
jbharatreddy333@gmail.com

Shivani Verma
Amity Institute of Space Science and Technology
Noida, Uttar Pradesh, India
sverma2@amity.edu

MS Prasad
Amity Institute of Space Science and Technology
Noida, Uttar Pradesh, India
msprasad@amity.edu

***Abstract*— In the non-destructive testing (NDT) industry, radiographic images are widely used to check the safety and reliability of equipment. Accurately identifying defects becomes challenging when there is limited labelled data, a wide variation in defect sizes, and the presence of noisy, low-contrast images. To tackle this problem, this work introduces a pre-trained ResNet-50 model integrated with a conditional expert block, which routes features through multiple convolutional experts. The proposed model achieves an accuracy of 95.5%.**

Keywords*— *weld defect classification, radiographic NDT, conditional computation, mixture-of-experts, adaptive feature learning, deep learning, ResNet-50

I. INTRODUCTION

Welded component quality is frequently assessed using Radiographic Non-Destructive Testing. This technique is often used when the welded components need to meet strict safety standards in applications such as pipelines, aircraft and power systems [2]. Any type of defect on one of these welded components has the potential to cause failure due to progressive degradation from exposure over time [2]. Usually today, experts conduct inspections manually, which often leads to a slow process and inconsistency [3]. As such, there is a need for automated solutions to facilitate quicker inspections with consistently reliable results [4].

One of the main challenges in this task is the variation in how defects appear [5]. Even defects from the same category can look different in different images [6]. For example, a crack may appear as a thin line in one image but look irregular in another [7]. Similarly, porosity can appear as a few small spots or as a group of clustered regions [8]. These differences make it difficult for models to learn patterns that work well in all cases [9].

Most deep learning models process all images in the same way using a fixed structure [10], [11]. While this works well in controlled conditions, it is less effective when the data is more varied [12]. In real-world situations, some images are simple while others are more complex, so treating them the same is not always ideal [13]. To address this, researchers have started using models where different parts of the network focus on different types of features [14]. Expert-based models follow this idea by using multiple branches and selecting the most useful ones during processing [15].

Another approach that has gained attention is conditional computation [16]. In this method, the model adjusts how it processes each input instead of following the same steps every time [17]. This can improve both performance and efficiency [18]. However, this approach has not been widely used for weld defect classification [19]. In this work, we combine residual learning with conditional expert selection to better handle variations in radiographic images [20].

II. RELATED WORK

Earlier methods for weld defect detection mainly used handcrafted features and basic image processing techniques such as edge detection and thresholding [3], [4]. While these methods were useful at the beginning, they were sensitive to noise and changes in image quality [5]. This limited their performance in real-world applications [6]. With the development of DL, especially CNNs, performance improved in weld defect detection and classification tasks [7], [8]. Models such as VGG, DenseNet, and ResNet became popular because they can automatically learn useful features from data [11], [12]. Among these, ResNet-based models perform well because they allow deeper networks to be trained more effectively [11].

To further improve results, attention mechanisms have been added to CNN models. Methods such as Squeeze-and-Excitation Networks and CBAM help the model focus on important regions in the image while ignoring unnecessary background information [13], [14]. Even with these improvements, challenges like low contrast, noise, and variation within the same defect type remain [9], [10]. Also, most of these models use a fixed structure, which limits their ability to adapt to different types of data [7].

More recent approaches include Vision Transformers and Swin Transformers for image classification tasks [18], [19]. In addition, object detection models such as YOLO, Faster R-CNN, and Mask R-CNN have been used in industrial inspection [15], [17]. Conditional computation methods like CondConv and Switch Transformers show that selecting different computation paths can improve both efficiency and performance [21], [22]. However, the use of these methods for weld defect classification in radiographic images is still limited [20].

III. PROPOSED METHODOLOGY

A. Model Architecture

The proposed system works on weld images with four classes: crack, porosity, lack of penetration, and no defect. Since the images are collected under different conditions, a preprocessing step is applied to make the dataset more consistent [10]. All images are resized to a fixed size and normalized using standard values. As the images are grayscale, intensity normalization is also applied to improve contrast and reduce variations caused by different imaging conditions [5]. Basic data augmentation techniques are used during training, including horizontal flipping, rotations at different angles, and slight changes in intensity [5], [13].

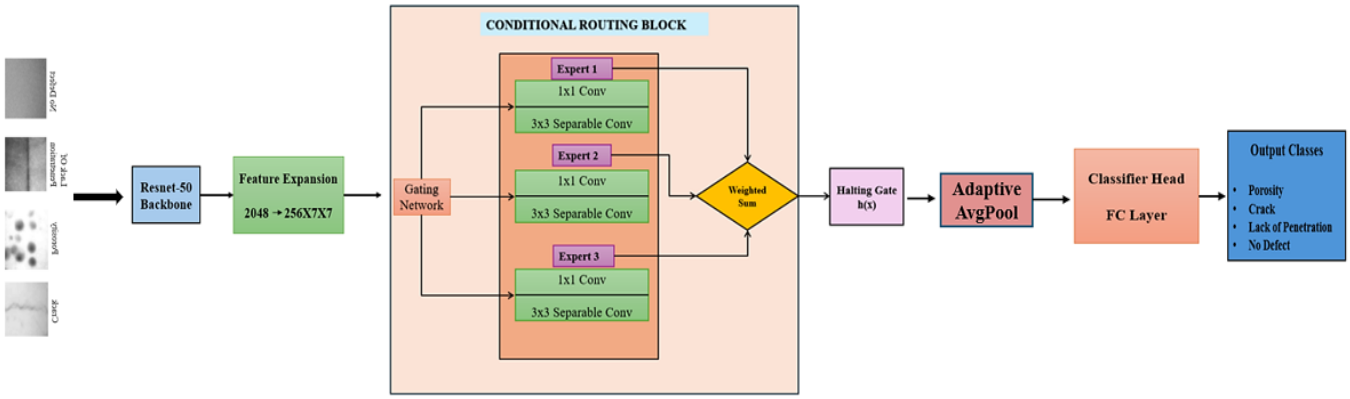

Fig. 1. Overall system architecture for radiographic weld defect classification

An overview of the architecture is shown in Fig. 1. After preprocessing, the images are passed through a ResNet-50 backbone, which is used to extract features [11]. Because of its residual structure, the network can learn both simple features such as edges and textures, as well as more complex defect-related patterns.

The extracted features are then passed through a feature expansion step, where they are reshaped into a spatial form. This helps the model preserve important spatial information and focus on local defect details [9]. As a result, the model becomes better at distinguishing between defect types that appear visually similar [12].

Finally, these features are passed into the Conditional Routing Block, which is the main component of the proposed system. Instead of using a fixed processing path, this block employs multiple expert branches along with a gating network to determine how the features should be combined. The gating mechanism assigns adaptive weights to each branch based on the input, allowing the model to focus on the most relevant features [20], [21]. This adaptive approach improves classification performance, especially when defect appearances vary significantly [22].

B. Spatial Parameters

The expert branches and the gating mechanism work on feature maps of size 7×7. Keeping this resolution unchanged throughout the block helps retain the spatial structure of the features, allowing the model to preserve important local details while processing the input. This resolution retains sufficient local structural detail while maintaining computational efficiency, particularly for modelling fine-grained weld defect geometries. A 3×3 convolutional kernel within each expert captures neighbourhood-level texture variations. The gating network uses global average pooling to obtain a global descriptor for expert selection. The halting gate performs scalar-controlled interpolation uniformly across all spatial locations.

C. ResNet-50 Backbone Feature Extraction

ResNet-50 serves as the backbone network for feature extraction. The residual architecture enables efficient training of deep networks by removing the vanishing gradient problem. Input weld radiographic images are resized to 227×227 pixels and normalized using ImageNet mean/standard deviation. They are passed through a frozen ResNet-50 backbone up to the global average pooled layer to obtain a 2048-D feature vector. A fully connected expansion layer maps this vector into a feature map of shape C×H×W = 256×7×7, C is channel H is height and W is width. While the original input is single-channel grayscale, expansion into 256 channels enables the model to encode a diverse set of high-dimensional patterns such as crack textures and porosity densities. This compact spatial representation serves as a bridge between global semantics and the Conditional Routing Block.

D. Conditional Expert Block

To better handle the wide variation in weld defect patterns, a Conditional Expert Block is introduced. This allows the model to adapt its processing by directing feature representations through multiple specialized convolutional expert branches, depending on the input Unlike conventional convolutional layers that apply a fixed transformation to all inputs, the proposed block consists of K=3 depthwise-separable convolutional experts and a lightweight gating network that determines the contribution of each expert. The architecture of the block is illustrated in Fig. 2.

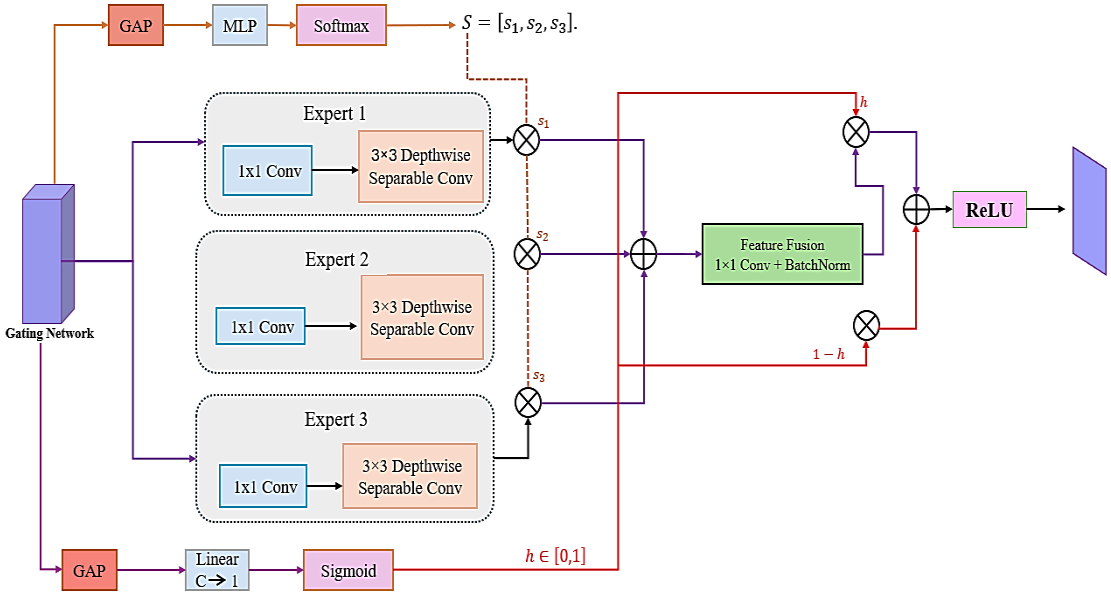

Fig. 2. Architecture of the Conditional Routing Block

Given an input feature map x ∈ R^(C×H×W), the routing process begins by applying global average pooling to extract a compact channel descriptor. This descriptor is passed through a squeeze-and-excitation style multilayer perceptron to generate routing weights g ∈ R^K using a SoftMax activation, ensuring that (i) takes input map x ∈ R^(C×H×W); (ii) computes GAP and uses an SE-style MLP to produce expert weights g ∈ R^K (SoftMax) and halting gate h ∈ [0,1]; (iii) passes x through K depthwise-separable convolutional experts; (iv) performs mixture-of-experts fusion via weighted sum; (v) applies 1×1 fusion convolution with batch normalization; and (vi) performs halting interpolation between residual and transformed path. This allows the network to adjust how features are processed depending on the input image, instead of treating all samples in the same way. Since weld defects can vary significantly in appearance, this flexibility helps the model capture subtle differences more effectively.

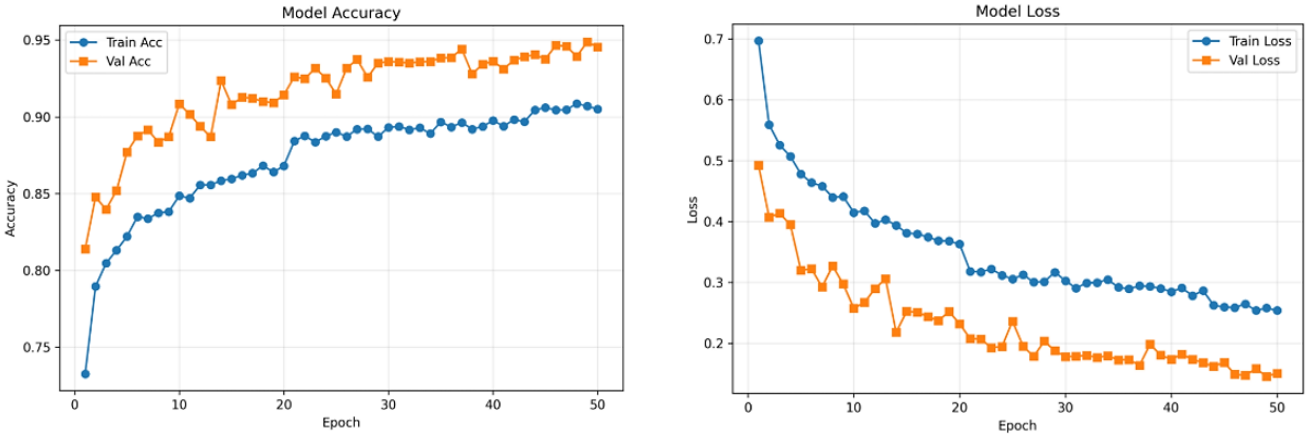

Fig. 3. Training accuracy vs Validation accuracy and training loss vs. Validation loss

E. *Gating Network*

The routing weights are produced by a lightweight gating network that operates on the globally pooled feature representation. Global Average Pooling first reduces the spatial feature map into a compact channel descriptor. This descriptor is passed through a two-layer multilayer perceptron with non-linear activation to generate expert routing logits, which are normalized using a SoftMax function to obtain the expert weight vector g(x). $g(x) = Softmax\left(W2\sigma(W1 \cdot GAP(x))\right)$ where W1, W2 are learned weight matrices, σ is the ReLU activation, and g(x) is the routing weight vector. The fusion equation is:

$$y = \sum_{k=1}^{K} g_k(x) \cdot f_k(x)$$

where $f_k(x)$represents the output of the k-th expert and $g_k(x)$ indicates how much importance is given to that expert for a given input.

F. *Depthwise Separable Convolutions*

Each expert uses a 1×1 bottleneck followed by a depth wise separable 3×3 convolution to reduce computation and parameters while preserving spatial expressivity.

G. *Global Pooling-Based Gating*

Global Average Pooling is used to summarize spatial information into a compact channel descriptor, enabling the gating network to capture global contextual information for expert selection. The routing weights are produced by a squeeze-and-excitation-style gating network: global average pooling over spatial dimensions, followed by a two-layer

MLP and SoftMax analogous to SE-Net channel attention but applied over experts rather than channels.

H. Halting Gate

Inspired by Adaptive Computation Time and dynamic routing in ResNet, we introduce a scalar halting gate per CRB. A sigmoidal MLP predicts h ∈ [0,1] that continuously interpolates between the identity residual path and the transformed expert output:

$$out = (1 - h) \cdot residual + h \cdot transformed$$

IV. Training Configuration

The training configuration is summarized in Table I.

TABLE I. Training Hyperparameters

Parameter	Value
Optimizer	Adam
Learning Rate Scheduler	ReduceLROnPlateau (×0.5, patience = 5)
Batch Size	32
Max Epochs	50
Early Stopping Patience	10
Backbone	ResNet-50 (frozen)
Input Size	227 × 227

V. Experimental Results and Analysis

Fig. 3 shows the training and validation accuracy and loss curves across epochs. The model demonstrates stable learning behaviour, with training accuracy steadily increasing and gradually converging. The validation accuracy follows a similar trend, remaining close to the training curve, which indicates good generalization and no significant overfitting. At the same time, both training and validation losses decrease consistently, with the validation loss reaching a slightly lower value. The overall alignment between training and validation curves suggests that the model achieves a balanced trade-off between bias and variance and converges reliably during training.

TABLE II. Performance Comparison of the Proposed Model with Baseline CNN Architectures

Model	Accuracy	Precision	Recall	F1
ResNet50	90.8	0.89	0.88	0.88
VGG16	88.5	0.86	0.86	0.86
DenseNet121	91.7	0.90	0.89	0.89
ResNet +CBAM	92.9	0.92	0.91	0.91
Proposed Conditional Expert	**95.5**	**0.95**	**0.95**	**0.95**

Table II compares the performance of the proposed Conditional Expert model with widely used CNN architectures for weld defect classification. Conventional models such as VGG16 and ResNet50 achieve accuracies of 88.5% and 90.8%, while DenseNet121 slightly improves performance to 91.7%. The attention-based ResNet+CBAM model further enhances the results with an accuracy of 92.9%. In comparison, the proposed Conditional Expert model achieves the best performance with an accuracy of 95.5% and balanced precision, recall, and F1-score of 0.95, demonstrating its effectiveness in adaptively capturing diverse weld defect features in radiographic image.

VI. Performance Evaluation

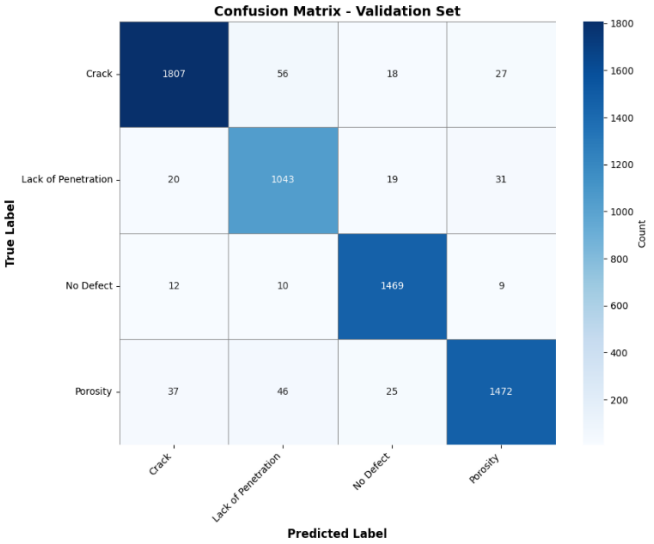

Fig. 4. Confusion matrix for four class weld defect classification

The confusion matrix demonstrates strong classification performance across all four weld defect categories, with high true-positive rates for Crack, Lack of Penetration, No Defect, and Porosity. The results show that most predictions fall along the diagonal, which means the model is correctly identifying many of the samples. Only a few misclassifications are observed, and these mainly occur between defect types that appear quite similar in radiographic images. For instance, confusion is sometimes seen between crack and porosity, as well as between lack of penetration and porosity, since these defects can share overlapping texture patterns.

VII. Practical Implementation: Weld Defect System GUI

A GUI was built using PyQt to make the model easier to use in real applications. The process is straightforward. First, the user loads the trained model weights. Then, a weld defect image is selected for analysis. After processing, the system shows the predicted defect class along with its confidence score.

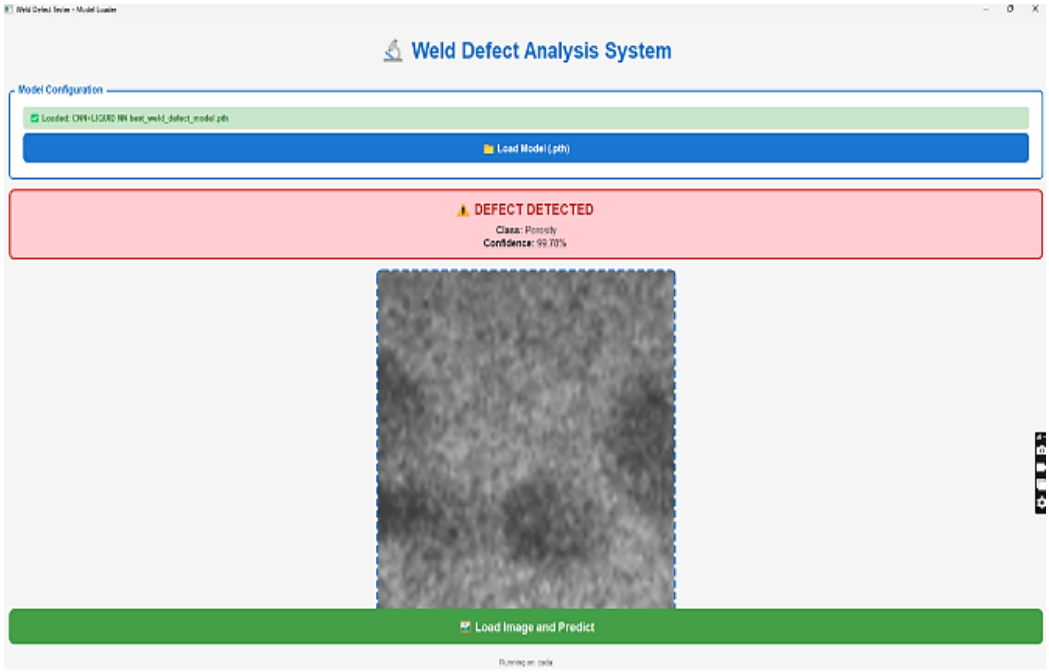

Fig. 5. Weld Defect Analysis System

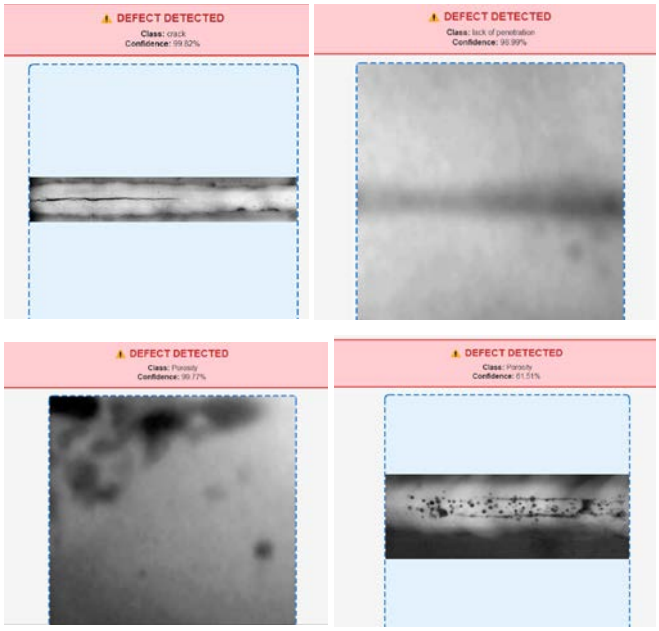

Fig. 6. GUI outputs of various weld defect types

VIII. Conclusion

This study presents a method that combines ResNet-50 with an expert-based system to detect weld defects. The model uses multiple branches, where each branch focuses on different types of patterns. This helps the model handle the variations commonly seen in radiographic images. The model achieves an accuracy of 95.5% on the RIAWELC dataset. This shows that it can correctly classify defects even when they look similar. Another important feature of this model is that it can adjust how it processes each image by selecting the most relevant features. This improves its performance and makes it useful for real-world applications. In future work, we plan to study the computational efficiency of the model and evaluate its performance in real-time conditions. We will also test the model on different datasets to check how well it works across various types of defects.

References

[1] J. R. Moussa, D. Harmon, and S. Rane, “Industrial Radiography: Trends, Market Drivers, and Alternatives to Gamma-based Devices,” Health Physics, vol. 129, no. 3, pp. 174–183, Sep. 2025.

[2] S. Vishnuvardhan, A. Ramachandra Murthy, and A. Choudhary, “A review on pipeline failures, defects in pipelines and their assessment and fatigue life prediction methods,” Int. J. Pressure Vessels and Piping, vol. 201, p. 104853, 2023.

[3] X. Zhang, S. Wang, and Y. Zhang, “Automatic weld defect detection in radiographic images using deep learning,” NDT & E Int., vol. 110, p. 102202, 2020.

[4] Y. Wang, J. Chen, and Z. Li, “Automatic classification of weld defects in radiographic images using deep convolutional neural networks,” NDT & E Int., vol. 102, pp. 1–9, 2019.

[5] L. Zhang, X. Wang, and H. Liu, “Weld defect recognition in radiographic images using deep learning with data augmentation,” J. Manuf. Syst., vol. 54, pp. 295–305, 2020.

[6] S. Chaudhury, A. K. Sinha, and S. Banerjee, “Automated weld defect detection from radiographic images using convolutional neural networks,” IEEE Trans. Ind. Informat., vol. 17, no. 6, pp. 4127–4136, 2021.

[7] Y. Li, H. Zhang, and L. Wang, “Deep learning-based weld defect detection in X-ray images: A comprehensive review,” IEEE Access, vol. 10, pp. 45123–45145, 2022.

[8] H. Zhao, X. Liu, and J. Chen, “Automated weld defect detection in radiographic testing using deep convolutional neural networks,” NDT & E Int., vol. 124, p. 102567, 2021.

[9] M. Kumar, R. Singh, and A. Verma, “Radiographic weld defect classification using deep learning techniques,” J. Manuf. Syst., vol. 61, pp. 250–261, 2022.

[10] B. Totino, F. Spagnolo, and S. Perri, “RIAWELC: A novel dataset of radiographic images for automatic weld defects classification,” Int. J. Electr. Comput. Eng. Res., vol. 3, no. 1, pp. 13–17, 2023.

[11] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE CVPR, 2016, pp. 770–778.

[12] M. Tan and Q. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” in Proc. ICML, 2019, pp. 6105–6114.

[13] J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 42, no. 8, pp. 2011–2023, 2020.

[14] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, “CBAM: Convolutional block attention module,” in Proc. ECCV, 2018, pp. 3–19.

[15] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You Only Look Once: Unified, real-time object detection,” in Proc. IEEE CVPR, 2016, pp. 779–788.

[16] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1137–1149, 2017.

[17] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask R-CNN,” in Proc. IEEE ICCV, 2017, pp. 2961–2969.

[18] A. Dosovitskiy et al., “An image is worth 16×16 words: Transformers for image recognition at scale,” in Proc. ICLR, 2021.

[19] Z. Liu et al., “Swin Transformer: Hierarchical vision transformer using shifted windows,” in Proc. IEEE ICCV, 2021, pp. 10012–10022.

[20] N. Shazeer et al., “Outrageously large neural networks: The sparsely gated mixture-of-experts layer,” in Proc. ICLR, 2017.

[21] Y. Yang, X. Dai, M. Tan, J. Liu, and Q. Le, “CondConv: Conditionally parameterized convolutions for efficient inference,” in Proc. NeurIPS, 2019.

[22] W. Fedus, B. Zoph, and N. Shazeer, “Switch Transformers: Scaling to trillion parameter models with simple and efficient sparsity,” JMLR, vol. 23, no. 120, pp. 1–39, 2022.

[23] D. Eigen, M. Ranzato, and I. Sutskever, “Learning factored representations in a deep mixture of experts,” in Proc. ICLR, 2014.

[24] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016.

[25] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” in Proc. NeurIPS, 2012.

[26] C. Szegedy et al., “Going deeper with convolutions,” in Proc. IEEE CVPR, 2015.

[27] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in Proc. IEEE CVPR, 2017.