Accelerating Corrosion Surface-Area Measurements with Computer Vision and Deep Learning: An Ensemble Approach

HAI D. NGUYEN, SHENGYI WANG, REBEKAH WILSON, BRIAN EICK and NATALIE BECERRA-STASIEWICZ

ABSTRACT

The corrosion of infrastructure and facilities poses a significant challenge for the United States Department of Defense (DoD) in terms of cost and military readiness. To tackle this challenge, our research introduces a data-driven corrosion segmentation method that combines three deep learning-based models and an ensemble learning approach for the automatic identification and segmentation of corroded regions within high-resolution images. The method involves several stages, such as data annotation, preprocessing, augmentation, model implementation, and performance evaluation. The deep learning models used include Feature Pyramid Network (FPN) with Residual Network (ResNet)-34 encoder, UNet with ResNet-34 encoder, and UNet++ with Visual Geometry Group (VGG)-19 encoder. Ensemble learning, a technique that integrates these deep learning models, was employed to improve prediction accuracy and overall performance. The proposed method is evaluated using both the Dice score and the Intersection over Union (IoU) score metrics. Experimental results demonstrate that the ensemble learning approach outperforms individual models, achieving a Dice score of 90.1% and an IoU score of 83.9%. The approach shows promise to automatically detect and measure corrosion, which can reduce inspection costs and identify major issues to aid in prevention of structural failure. The tool developed in this study will be expanded to provide similar capabilities for large-scale civil infrastructure.

INTRODUCTION

Mitigating corrosion currently accounts for 40% of the national maintenance budget, totaling over $20 billion a year for the military and equally impacting civil works infrastructure. Corrosion poses a significant challenge for the U.S. Army Corps of Engineers, leading to substantial economic and environmental consequences.

Hai D. Nguyen, The U.S. Army Corps of Engineers, Champaign, IL 61822
Shengyi Wang, University of Illinois Urbana-Champaign, Urbana, IL 61801
Rebekah Wilson, The U.S. Army Corps of Engineers, Champaign, IL 61822
Brian Eick, The U.S. Army Corps of Engineers, 2902 Newmark Dr, Champaign, IL 61822
Natalie Becerra-Stasiewicz, The U.S. Army Corps of Engineers, Champaign, IL 61822

It contributes to air and water pollution, soil and vegetation contamination, and generates waste, among other issues. Additionally, corroded infrastructure is prone to structural failure, resulting in catastrophic disasters and loss of life, as demonstrated by incidents such as the collapse of the Morandi bridge in Genoa, Italy and the sinking of the Erika ship in Brittany, France. Moreover, corrosion significantly increases maintenance expenses and reduces the lifespan of infrastructure. Therefore, early detection of corroded areas is crucial for maintaining structural integrity, preventing disasters, and minimizing long-term maintenance costs.

In recent years, there has been a significant increase in publications related to automated corrosion detection using machine learning and image-based methods. Several studies, including Spencer et al. [1], Flah et al. [2], Jahanshahi et al. [3], and Bai et al. [4], have investigated various techniques for automated inspection and structural damage detection. These techniques include Convolutional Neural Network (CNN) classifiers, Otsu image processing, image registration, morphological image processing, edge detection, texture and color analysis, wavelet transform, pattern recognition, image classification, object detection, and semantic segmentation. By using these methods, researchers have been able to identify different types of damage, such as cracks, spalling, and corrosion. While many researchers have used popular CNN networks like AlexNet, GoogLeNet, Residual Network (ResNet), and VGGNet for corrosion classification [5-7], there has been limited research on recent advancements such as Feature Pyramid Network (FPN) [8], U-Net [9], and ensemble learning [10]. These methods have shown promising results and could be beneficial for improving the accuracy and efficiency of automated corrosion detection.

This paper proposes a deep learning-based corrosion segmentation method that combines three deep learning models and ensemble learning to automatically identify and segment corroded regions in high-resolution images. The study focuses on developing a method to replace or supplement manual corrosion measurement according to ASTM D1654. The goal is to use a computer vision-based approach for automatic surface area measurement of corrosion on test specimens. This would drastically reduce analysis time and can then be further developed to enable automated rapid assessment for large-scale infrastructure and buildings. The following sections will discuss laboratory-tested specimens, image annotation, preprocessing, deep neural networks architectures including FPN, U-Net, ResNet, and VGGNet, ensemble training, and corrosion prediction results from three deep learning models and ensemble learning.

EXPERIMENTAL PROCEDURE

In this study, hundreds of painted/coated test panels were scribed and exposed to extreme weathering conditions to accelerate corrosion at the predefined scribe marks. The purpose of these scribe marks is to assess the performance of coatings in accordance with ASTM D1654. Standardized flaws were created on these test panels by making line and X-shaped scribes on the specimens. Accelerated weathering conditions were then simulated in the laboratory to rapidly induce corrosion on the test panels. The corrosion acceleration process followed ISO 12944 guidelines for corrosion protection of steel structures by protective paint systems. This involved subjecting the panels to alternating exposures of ultraviolet (UV) radiation and water condensation for four hours each over three days, followed by three days in a salt fog chamber and one day in a negative 20-degree Celsius freezer. This one-week cycle was repeated 12 times. Afterward, the test panels were photographed using a high-resolution digital microscope. These images were subsequently annotated and utilized for training, testing, and validation of the deep learning models and ensemble learning techniques.

PROPOSED DEEP LEARNING-BASED CORROSION SEGMENTATION METHOD

This study presents a data-driven approach that utilizes three deep learning-based models and an ensemble learning technique to automatically detect and segment corroded areas in high-resolution images, where "segment" refers to the partitioning of an image into multiple regions based on specific criteria. The proposed method involves several stages (as illustrated in Figure 1), including: (1) Annotation, preprocessing, and augmentation of image data; (2) Application of three deep learning-based models/learners for corroded region segmentation, namely Feature Pyramid Network (FPN) [8] with Residual Network (ResNet)-34 encoder [11], UNet [9] with ResNet-34 encoder, and UNet++ [12] with Visual Geometry Group (VGG)-19 encoder [13]; (3) Assessment of the performance of the three models/learners; and (4) Ensemble learning involving the three models/learners and subsequent evaluation.

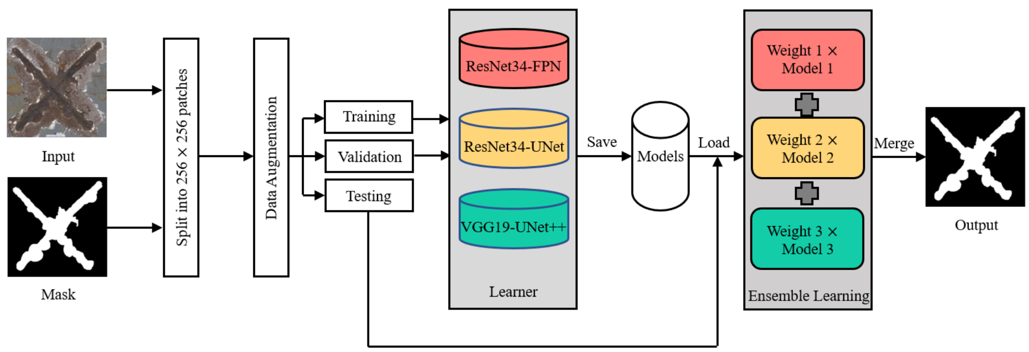

Figure 1. Proposed deep learning-based method.

Data Annotation, Preprocessing, and Augmentation

This study utilized high-resolution images of varying sizes, ranging from 2,880 × 2,160 to 11,020 × 9,295 pixels. These images were manually annotated with the Labkit extension in ImageJ, a robust image processing software [14]. Figure 2 exhibits some examples of the annotated images and their corresponding corrosion masks. The dataset comprised a total of 262 images, which were divided into training, validation, and testing subsets in an 8:1:1 ratio. The training, validation, and testing subsets contained 209, 26, and 27 images, respectively. To accommodate the original images within GPU memory, both the images and their corresponding masks were divided into non-overlapping 256 × 256 patches, rather than downsampling [15], which allowed for the preservation of maximum detail. Splitting resulted in 424,194 patches for training and validation. During testing, the test images were predicted using 256 × 256 patches and reconstructed by merging them to the original resolution.

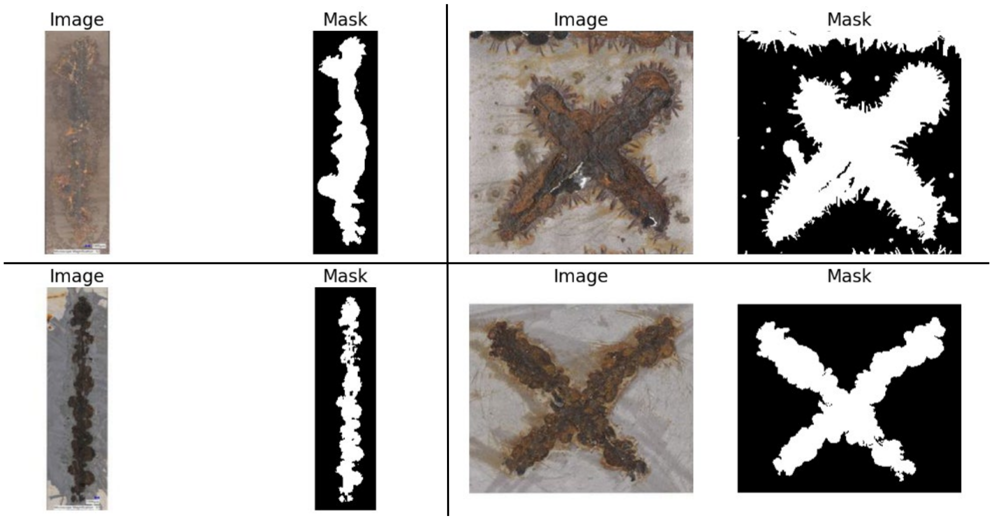

Figure 2. Examples of four individual annotated and preprocessed images with corresponding masks.

This study employed horizontal flipping as a data augmentation method to increase the diversity of the dataset [16]. By doing so, the dataset was enriched with a wider range of corrosion patterns, which could help the model learn more robust features for segmentation. Data augmentation enhanced the model's generalization capabilities by reducing the risk of overfitting to the training data. Furthermore, by introducing variations in the images, data augmentation strengthened the model's capability to handle diverse situations and conditions that may occur in real-world scenarios, thus enhancing its practical applicability.

Model Architecture

Semantic segmentation is typically performed using a two-stage network consisting of an encoder and a decoder. The encoder, made up of convolutional layers, extracts high-level features from the input image, while the decoder is an upsampling network that generates a probability map for each pixel and its corresponding class, recovers spatial resolution, and produces segmentation masks that match the input image size. This study employed three deep learning-based models for semantic segmentation: FPN with ResNet-34 encoder, UNet with ResNet-34 encoder, and UNet++ with VGG-19 encoder.

Encoder: The authors leveraged ResNet-34 and VGG-19 as the backbone networks to extract features from images. Both of these networks have been pre-trained on the ImageNet dataset [17], an extensive annotated collection of over one million images. VGG-19 utilizes multi-layered blocks of 3x3 convolutional filters, which allows it to extract complex features from images. On the other hand, ResNet-34, implied by its name, is comprised of 34 layers, each with shortcut connections that employ residual learning. This approach of learning the difference (or residual) between the input and output of a layer rather than the underlying function directly, enables ResNet to overcome the vanishing gradient problem [11] and train deep networks efficiently without any loss in performance.

Feature Pyramid Network (FPN): FPN is a segmentation model that uses a top-down pathway and lateral connections to construct a feature pyramid from a single-scale input image. This architecture enables FPN to generate feature maps with rich semantic information and high spatial resolution at multiple scales, leading to improved segmentation accuracy. FPN achieves this by using the top-down pathway to upsample feature maps from higher encoder levels and merging them with corresponding feature maps from lower encoder levels through lateral connections. By doing so, FPN constructs a feature pyramid where the feature maps at each level retain the semantic information from higher levels while maintaining the spatial resolution of the lower levels. In this study, FPN employed ResNet-34 as the encoder. ResNet-34 is a deep neural network architecture that has shown excellent performance in various computer vision tasks, including image classification, object detection, and segmentation. With ResNet-34 as the encoder, FPN was able to leverage the powerful representation capabilities of the encoder to construct a feature pyramid that captures rich semantic information at multiple scales, leading to improved segmentation performance.

UNet: The UNet is an advanced neural network architecture used for image segmentation tasks. It consists of two main paths: an encoding path and a decoding path. The encoding path is responsible for feature extraction from the input image and the decoding path recovers the feature representations to produce the segmentation map. One of the unique features of UNet is the incorporation of skip connections, which allow for the preservation of spatial information and the recovery of fine details during segmentation. These skip connections link the encoding and decoding paths at different levels, allowing the model to capture features at multiple scales. To enhance the feature extraction capability of the UNet, ResNet-34 was used as the encoder in this study to capture high-level features from images. Incorporating ResNet-34 as the encoder in UNet can improve the model's ability to extract meaningful features from the input image, which can lead to more accurate segmentation results.

UNet++: The UNet++ is an enhanced version of the UNet model for image segmentation. UNet++ improves upon the original architecture by incorporating nested and dense skip connections between the encoder and decoder paths. These connections facilitate better preservation of spatial and contextual information, enabling the network to adapt to varying levels of complexity. In this study, VGG-19 was employed as the encoder for UNet++, providing more diverse features for segmentation. This allows for greater flexibility in capturing the relevant image features for accurate segmentation. The nested and dense skip connections in UNet++ enable the model to capture both low-level and high-level features by concatenating feature maps of different sizes. This allows the network to better integrate contextual information and preserve spatial resolution throughout the segmentation process.

Model Evaluation

In this study, the Dice score [18] and Intersection over Union (IoU) score were selected to assess the efficacy of the proposed deep learning trainers and ensemble learning approach. The Dice score, also known as the F-1 score, is a measure of the overlap between the ground truth and predicted outcomes. It ranges from 0 to 1, with higher values indicating better agreement between the prediction and the ground truth. The IoU score is a metric that measures the similarity between the predicted segmentation mask and the ground truth segmentation mask. It is calculated by dividing the area of overlap between the two sets by the area of union between the two sets. IoU scores range from 0 to 1, where 0 means no overlap and 1 means perfect overlap. The formulae for calculating the Dice score and IoU score are as follows:

$$Dice\ Score = \frac{2 \times |X \cap Y|}{|X| + |Y|} \tag{1}$$

$$IoU\ Score = \frac{|X \cap Y|}{|X \cup Y|} = \frac{|X \cap Y|}{|X| + |Y| - |X \cap Y|} \tag{2}$$

where $|X|$ and $|Y|$ denote the pixel counts for the ground truth and predicted outcomes, respectively.

Ensemble Learning

Ensemble learning method was proposed in this study. Three learners were trained separately and then combined to improve overall performance. The weights w_1, w_2, and w_3 represent the relative importance of each learner in the ensemble. These weights are constrained to be greater than zero and their sum must equal 1. The ensemble prediction *F(x)* for a given input *x* is calculated as the weighted sum of the predictions from the three learners, as illustrated in Equation 3. For pixel-level classification, a binary output is obtained by applying a threshold function, *Mask(x)*, which assigns 0 to the ensemble prediction if it falls between 0 and the selected threshold, and 1 if it falls between the selected threshold and 1, as depicted in Equation 4.

$$F(x) = w_1 \times Model1(x) + w_2 \times Model2(x) + w_3 \times Model3(x)$$
$$(w_1, w_2, w_3 > 0,\ w_1 + w_2 + w_3 = 1) \tag{3}$$

$$Mask(x) = \begin{cases} 0, & 0 < F(x) \leq threshold \\ 1, & threshold < F(x) \leq 1 \end{cases} \tag{4}$$

RESULT AND DISCUSSION

The deep learning-based methods were implemented using PyTorch [19], a flexible and widely adopted Python deep learning framework that enables GPU acceleration. All three models were trained for 10 epochs separately, using the Stochastic Gradient Descent (SGD) optimizer with a momentum of 0.9 to overcome local minima [20]. The learning rate, determining the step size that the optimizer takes to reach the minimum of the loss function, was set to 0.0001. A weight decay parameter of 1e−4 was employed to penalize large weights and prevent overfitting. Dice loss served as the loss function for all three models. The weights for the ensemble learning method were selected using the grid search method (a hyperparameter optimization technique used to find the best combination of hyperparameter values for a machine learning mode) to optimize performance. The weights for ResNet34-FPN, ResNet34-UNet, and VGG19-UNet++ are 0.1, 0.5, and 0.4, respectively.

Table I presents the results obtained from three deep learning models and the ensemble learning method. It demonstrates that ResNet34-UNet achieves the highest Dice and IoU scores among the individual models, followed by VGG19-UNet++. ResNet34-FPN has the lowest scores among the single models. Ensemble learning, which combines the predictions of multiple models, yields the highest scores overall, indicating that it benefits from the diversity and complementarity of different learners.

Figure 3 displays examples of original images, corresponding ground truth masks, and predicted masks of the ensemble learning method. As shown, the proposed method can successfully detect most corrosion areas and produce accurate segmentation masks.

However, there are some limitations. First, the method tends to undersegment parts of the corrosion that have low contrast or irregular shapes, resulting in false negatives. Second, the method suffers from edge artifacts caused by the splitting and merging operations that divide the image into patches and then stitch them together. These artifacts can affect the smoothness and continuity of the segmentation boundaries, leading to false positives. More robust and adaptive techniques for splitting and merging need to be explored to overcome these challenges.

TABLE I. RESULTS OBTAINED FROM THREE DEEP LEARNING-BASED MODELS AND ENSEMBLE LEARNING

Method	**Dice Score**	**IoU Score**
ResNet34 - FPN	86.1%	78.2%
ResNet34 - UNet	89.5%	83.0%
VGG19 - UNet++	87.9%	80.8%
Ensemble Learning	**90.1%**	**83.9%**

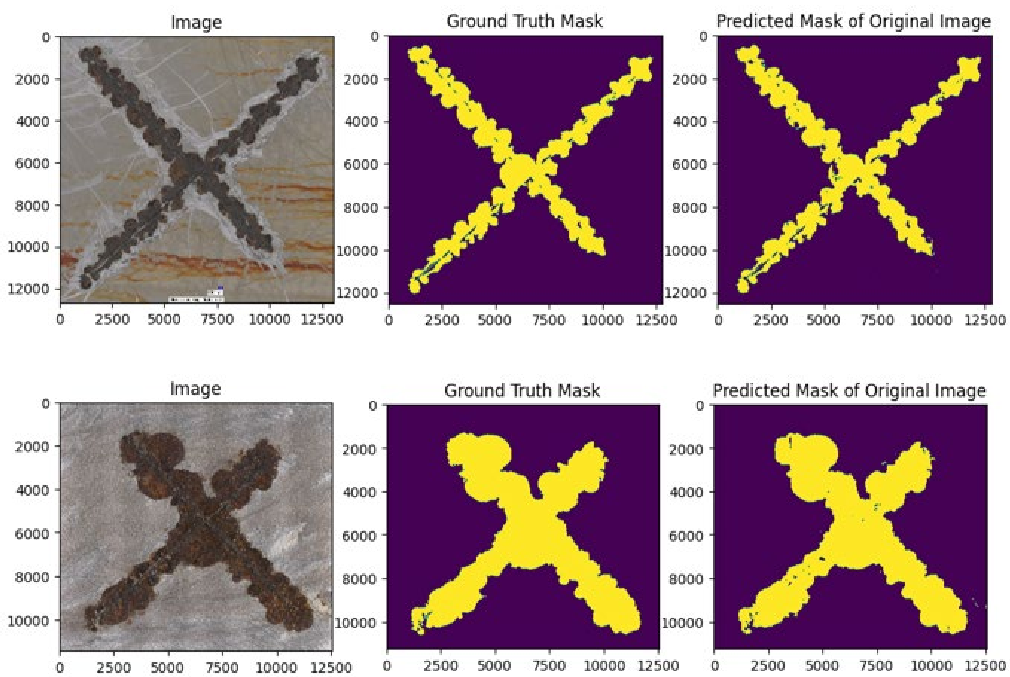

Figure 3. Examples of original images and their corresponding ground-truth and predicted masks.

CONCLUSION

The proposed deep learning-based corrosion segmentation method offers a promising approach to automatically detect and segment corroded regions from high-resolution images. The ensemble learning method, which combines the predictions of FPN with ResNet-34 encoder, UNet with ResNet-34 encoder, and UNet++ with VGG-19 encoder, outperforms the individual models, achieving a Dice score of 90.1% and an IoU score of 83.9%. Despite some limitations, such as undersegmentation and edge artifacts, the proposed method shows great potential for real-world corrosion detection and segmentation applications. Future work may explore more robust and adaptive techniques for splitting and merging operations to overcome these challenges and further improve the segmentation accuracy.

REFERENCES

1. Spencer Jr, B. F., Hoskere, V., & Narazaki, Y. (2019). Advances in computer vision-based civil infrastructure inspection and monitoring. *Engineering*, *5*(2), 199-222.
2. Flah, M., Suleiman, A. R., & Nehdi, M. L. (2020). Classification and quantification of cracks in concrete structures using deep learning image-based techniques. *Cement and Concrete Composites*, *114*, 103781.
3. Jahanshahi, M. R., Kelly, J. S., Masri, S. F., & Sukhatme, G. S. (2009). A survey and evaluation of promising approaches for automatic image-based defect detection of bridge structures. *Structure and Infrastructure Engineering*, *5*(6), 455-486.
4. Bai, Y., Zha, B., Sezen, H., & Yilmaz, A. (2023). Engineering deep learning methods on automatic detection of damage in infrastructure due to extreme events. *SHM*, *22*(1), 338-352.
5. Holm, E., Transeth, A. A., Knudsen, O. Ø., & Stahl, A. (2020, January). Classification of corrosion and coating damages on bridge constructions from images using convolutional neural networks. In *Twelfth International Conference on Machine Vision* (Vol. 11433, pp. 549-556). SPIE.
6. Shi, J., Dang, J., Cui, M., Zuo, R., Shimizu, K., Tsunoda, A., & Suzuki, Y. (2021). Improvement of damage segmentation based on pixel-level data balance using vgg-unet. *Applied sciences*, *11*(2), 518.
7. Luo, C., Yu, L., Yan, J., Li, Z., Ren, P., Bai, X., Yang, E., & Liu, Y. (2021). Autonomous detection of damage to multiple steel surfaces from 360 panoramas using deep neural networks. *Computer-Aided Civil and Infrastructure Engineering*, *36*(12), 1585-1599.
8. Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2117-2125).
9. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the *MICCAI 2015 conference on Medical Image Computing and Computer-Assisted Intervention,* pp. 234-241).
10. Sagi, O., & Rokach, L. (2018). Ensemble learning: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(4), e1249.
11. He, K., Zhang, X., Ren, S., & Sun, J. (2016). “Deep residual learning for image recognition.” In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
12. Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., & Liang, J. (2019). “Unet++: Redesigning skip connections to exploit multiscale features in image segmentation.” *IEEE transactions on medical imaging*, *39*(6), 1856-1867.
13. Simonyan, K., & Zisserman, A. (2014). “Very deep convolutional networks for large-scale image recognition.” *arXiv preprint arXiv:1409.1556*.
14. Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). “NIH Image to ImageJ: 25 years of image analysis.” *Nature methods*, *9*(7), 671-675.
15. Nguyen, H., Wang, S., Wilson, R., Eick, B., & Becerra-Stasiewicz, N. (2023). Machine Learning and Computer Vision Approaches for Corrosion Detection. The *2023 DoD Corrosion Prevention Technology and Innovation Symposium*.
16. Shorten, C., & Khoshgoftaar, T. M. (2019). “A survey on image data augmentation for deep learning.” *Journal of big data*, 6(1), 1-48.
17. Deng, W. Dong, R. Socher, L. -J. Li, Kai Li and Li Fei-Fei, “ImageNet: A large-scale hierarchical image database.” *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248-255.
18. Zijdenbos, A. P., Dawant, B. M., Margolin, R. A., & Palmer, A. C. (1994). Morphometric analysis of white matter lesions in MR images: method and validation. *IEEE transactions on medical imaging*, *13*(4), 716-724.
19. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... & Chintala, S. (2019). “Pytorch: An imperative style, high-performance deep learning library.” *Advances in neural information processing systems*, *32*.
20. Qian, N. (1999). “On the momentum term in gradient descent learning algorithms.” *Neural networks*, 12(1), 145-151.