Enhancing U-Net for Breast Tumor Segmentation with Automatic Data Augmentation

Dat T. Tran, Thanh N. Nguyen, Minh N. Nguyen and Phuc H. Tran

March 16, 2024

Enhancing U-Net for Breast Tumor Segmentation with Automatic Data Augmentation

Dat T. Tran[1], Thanh N. Nguyen[2], Minh N. Nguyen[3], and Phuc H. Tran[4]

FPT University, Ho Chi Minh City, Vietnam

Abstract. Segmentation plays a crucial role in computer-aided diagnosis (CAD) for the early detection and diagnosis of diseases. This involves identifying and locating the regions of interest (ROI). Despite recent successes in computer vision technology, challenges persist due to the limited accessibility of medical data, resulting in models that may not generalize well to unseen data. This paper proposes an approach utilizing automatic data augmentation techniques for enhancing U-Net performance in breast tumor segmentation. We employ data augmentation in both the training and prediction phases. During training, augmentation increases the diversity and quantity of training data. In the prediction phase, Test Time Augmentation (TTA) is used to aggregate segmentation outputs from several versions of the original input, generated by the augmentation policy. Experimental results on the BrEaST dataset of 256 ultrasound breast tumor images and corresponding masks demonstrate the effectiveness of our proposed method. It improves U-Net's Dice Similarity Coefficient (DSC) from 0.7078 to 0.7541, concurrently mitigating the risk of overfitting and ensuring robust generalizability to unseen data. This research contributes to improving segmentation outcomes on limited datasets and promises to enhance the robustness of CAD systems in medical image analysis.

Keywords: Automatic Data Augmentation · U-Net · Breast Tumor Segmentation · Medical Image Analysis · Test Time Augmentation

1 Introduction

Medical imaging plays an important role in the diagnosis and treatment of various diseases, including cancer. Among different imaging technologies, ultrasound imaging offers several advantages such as real-time imaging capabilities, non-invasiveness, and cost-effectiveness [1]. It is particularly valuable for detecting and characterizing tumors, providing insights for treatment planning and monitoring. One of the critical tasks in ultrasound image analysis is tumor segmentation, which involves identifying and locating the tumor region and its boundary.

In recent years, deep learning segmentation has shown promising performance, and it has the potential for real-time and nearly real-time processing. The well-known neural network architecture in medical image segmentation is

U-Net [2], which was initially designed to address limited training data issues in medical image analysis and has achieved significant performance. Since U-Net's development, there have been several attempts to improve its performance by modifying and extending this architecture. Attention U-Net [3] employs attention gates that learn to focus on relevant regions while ignoring irrelevant ones. UNet++ [4] is based on nested and dense skip connections that gradually increase the semantic level of encoder feature maps to be closer to the corresponding decoder feature maps. Dense-UNet [5] utilizes dense blocks to deepen the network for extracting more semantic features, hence improving segmentation accuracy.

Besides the above efforts, research on data augmentation techniques for medical image analysis remains limited, specifically for breast ultrasound image analysis. Even though basic methods such as image manipulation, erasure, and mixing have been explored to some extent [6, 7], there is a distinct lack of exploration into more advanced techniques. This gap in research not only hinders progress in improving the accuracy and robustness of medical image analysis models but also highlights the potential for innovative approaches to address the unique challenges posed by medical imaging data.

In this paper, we aim to fill the identified gap in research by focusing on Automatic Data Augmentation (Auto Augment) methods, particularly exploring Faster AutoAugment [8], within the context of breast tumor segmentation tasks. By leveraging the capabilities of Auto Augment techniques, which automatically adjust data augmentation parameters to generate synthetic images closely resembling real ones, we endeavor to enhance the efficacy of the U-Net model on breast ultrasound tumor segmentation. To summarize, our contributions can be outlined as follows:

1. We introduce the application of Auto Augment techniques for medical datasets.
2. We utilize the policy searched by Faster AutoAugment in Test Time Augmentation.
3. We demonstrate that our proposed method enhances the performance of the U-Net model.

2 Related works

2.1 Data Augmentation

Data augmentation is a crucial technique for enhancing model accuracy by artificially creating more data samples from existing datasets. This not only improves model performance but also reduces the costs associated with collecting additional data. There are two main categories of data augmentation methods [9]: basic methods and advanced methods, as illustrated in Fig. 1. The most common methods are basic augmentations, such as image manipulation (flipping, rotating, scaling,...), erasure and mixing. These methods offer simplicity and ease of implementation and have proven effective in medical image analysis tasks [6,

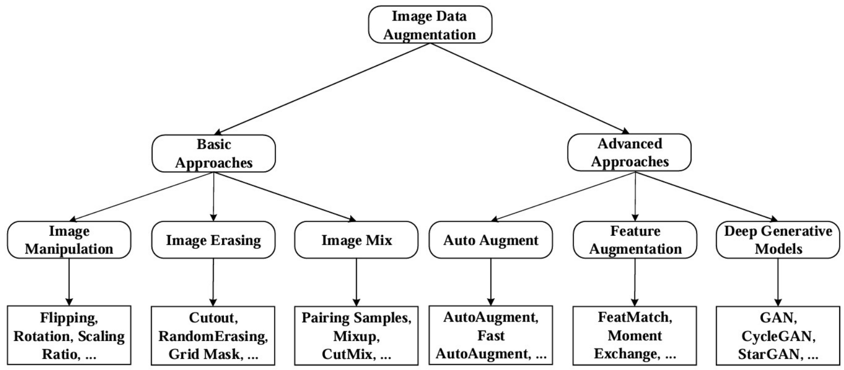

Fig. 1. A taxonomy of image data augmentation methods [9].

7]. However, due to their simplicity, they may not capture the complex structures present in real-world data and could potentially introduce unrealistic data samples.

2.2 Automatic Data Augmentation

One type of advanced method is Automatic Data Augmentation (Auto Augment) [8, 10–12]. It seeks to automatically design a data augmentation strategy or policy. Auto Augment defines a search space that includes all possible transformations and employs a search algorithm to select the most effective transformations from this space. While Auto Augment offers the ability to find unique augmentation policies customized for datasets with distinct characteristics, it requires additional training time and computational resources [13].

2.3 Test Time Augmentation

Test Time Augmentation (TTA) is an effective and robust technique for enhancing model performance across various computer vision applications. By predicting on and aggregating the results from multiple versions of the original input generated through augmented transformations, TTA significantly improves model performance. Several approaches exist for aggregating predictions, ranging from simple averaging to learning-based methods [14, 15]. Research has shown that TTA tends to reduce the expected error compared to the original model's average error [15]. However, it's important to note that TTA increases the computational cost and processing time during inference [16].

3 Method

In this section, we introduce a comprehensive method (Fig. 2) that incorporates augmentation policy with the U-Net architecture, and the application of Test

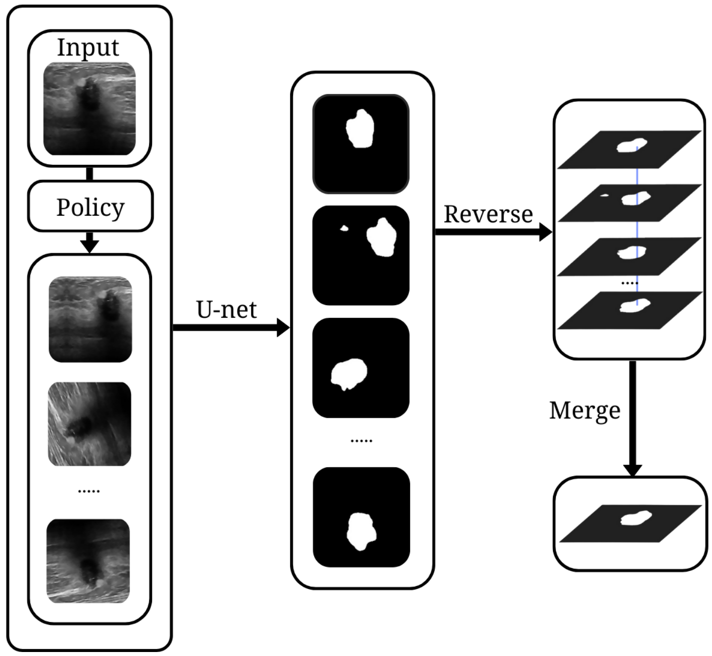

Fig. 2. Prediction process using policy U-Net and Test Time Augmentation.

Time Augmentation to achieve accurate results. Section 3.1 presents the method for selecting the policy. In section 3.2, we introduce the U-Net architecture used and its components. Finally, in the next section 3.3, we explain how to perform Test Time Augmentation effectively for result prediction.

3.1 Policy Search

Affine Transformation	**Color Enhancing**	**Other**
Shear x	Solarize	Cutout
Shear y	Posterize	Sample pairing
Translate x	Contrast	
Translate y	Color	
Rotate	Brightness	
Flip	Sharpness	
	Auto contrast	
	Equalize	

Table 1. The operations use in the policy.

Policy
The policy [8, 10–12] consists of distinct sub-policies, each containing consecutive operations selected from a set of 16 operations (as shown in table 1). When an image is passed to the policy, a sub-policy is chosen randomly to produce a new image by executing consecutive augmentation operations. Each operation is characterized by two parameters, p and μ, which indicate the probability of applying the operation and the magnitude of the operation, respectively (illustrated in Fig. 3).

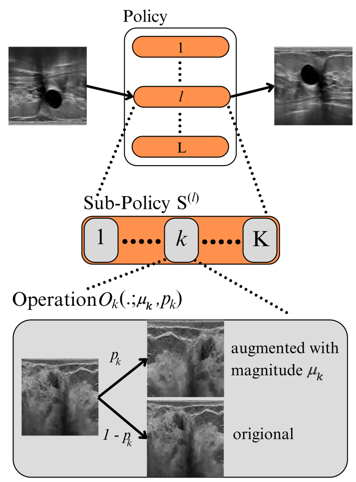

Fig. 3. The process of augmenting an image. [8]

The output of an operation is defined as:

$$X' = bO(X; \mu) + (1 - b)X \tag{1}$$

where $b \in \{0, 1\}$ is sampled formed Relaxed Bernoulli distribution with a low temperature of λ and b = 1 with a probability of p [8].

Back-propagation-based policy searching
The objective of the search process is to find the optimal policy that is capable of generating synthetic images that closely resemble real images. Wasserstein Generative Adversarial Network [17] (WGAN) with gradient penalty [18] is employed to achieve this goal. In contrast to traditional Generative Adversarial Networks [19] (GANs), where the generator typically takes the form of a Convolutional Neural Network [20] (CNN), in this scenario, the policy itself functions as the generator.

To enable back-propagation for policy search, it's crucial that each policy parameter, p and μ, be differentiable. To achieve this, a Relaxed Bernoulli distribution is used instead of a Bernoulli distribution. This modification ensures that the operation $O(.;\mu,p)$ remains differentiable with respect to its probability parameter p [8].

In the case of the magnitude parameter μ, some operations have discretized magnitude, which hinders back-propagation. To address this problem, each element (i, j)th of an augmented image by operation O is approximated as:

$$\tilde{O}(X;\mu)_{i,j} = \mathrm{StopGrad}\left(O(X;\mu)_{i,j} - \mu\right) + \mu \tag{2}$$

where StopGrad is a stop gradient operation that treats its operand as a constant. Additionally, only the backward pass uses the approximated operator, the forward pass uses the exact operator [8].

During searching, each operation is approximated by the weighted sum of the outputs of all operations as:

$$\sum_{n=1}^{16} [\sigma_\eta(w_k)]_n O_k^{(n)}(X;\mu_k^{(n)},p_k^{(n)}) \tag{3}$$

where σ is the Softmax function with a positive temperature parameter and w_k is a learnable parameter. This approach enables efficient training through backpropagation-based optimization, facilitating faster policy search [8].

3.2 U-Net Architecture

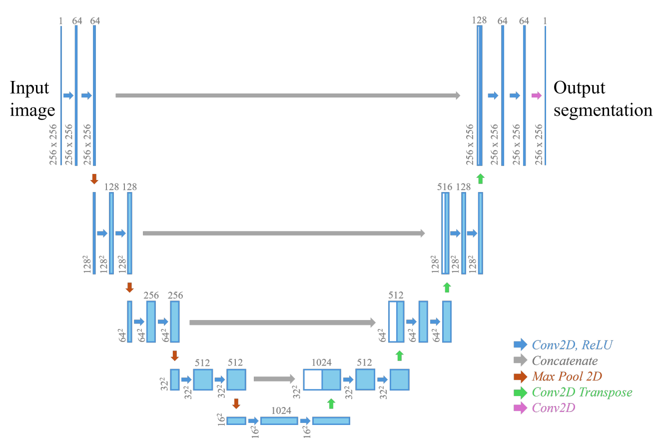

Fig. 4. U-net [2]

U-Net [2] in Fig. 4 was originally designed for biomedical image segmentation and has proven highly effective, especially for datasets with limited samples. It is a fully convolutional neural network comprised of two main components: the encoder and the decoder. The encoder is responsible for extracting relevant features from the input image, while the decoder utilizes these features to reconstruct a segmentation map with the same dimensions as the original image.

The encoder network consists of four repeated double-convolutional layers followed by a 2x2 max pooling layer for downsampling. The decoder mirrors the encoder's architecture, utilizing four repeated double-convolutional layers following each 2x2 transposed convolutional layer for upsampling. Each convolutional layer is followed by a rectified linear unit (ReLU) activation function. There are two primary types of connections between the encoder and decoder: skip connections and the bottleneck. Skip connections involve the concatenation of feature maps from corresponding levels of the encoder and decoder. The bottleneck, on the other hand, is a single double-convolutional layer.

3.3 Test Time Augumentation

Data augmentation was employed during the prediction stage to improve accuracy and minimize errors illustrated in Fig. 2. We leveraged the searched policy to generate ten augmented images from the input image. Both the original and augmented images were then processed to generate segmentation masks using trained U-Net. To ensure consistency, these masks were aligned by applying the reverse transformations initially used on the original image. Finally, the predicted masks were averaged to synthesize the final result [21].

4 Experiments

4.1 Data Preprocessing

We evaluated the performance of our method on the BrEaST breast tumor dataset [22], which consists of 256 ultrasound images. The dataset includes images from 154 benign, 98 malignant, and 4 normal cases, all obtained from different patients. Each case was manually annotated and labeled by radiologists working at medical centers in Poland between 2019 and 2022.

Ultrasound images are full of noise, especially speckle noise, which can significantly impact the accuracy of segmentation models [23, 24]. To mitigate this, a non-local mean filter [25] is employed for denoising Fig. 5.

Following denoising, the images were resized to a resolution of 256x256 pixels. We utilized a total of 252 tumor cases for our experiments, dividing the dataset into training, validation, and testing sets. The training set comprised 176 samples, the validation set contained 38 samples, and the testing set consisted of 38 samples. Notably, each set contained approximately 61.3% benign cases.

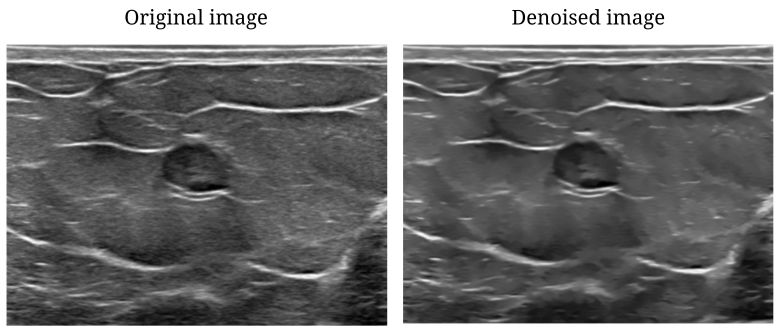

Fig. 5. Data preprocessing

4.2 Evaluation Metrics

To assess the effectiveness of our proposed approach, we employed a range of metrics commonly used for evaluating segmentation performance. These metrics include accuracy, sensitivity, specificity, Dice Similarity Coefficient (DSC), and Intersection-over-Union (IoU) [26].

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{4}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{5}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FN + FP} \tag{6}$$

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \tag{7}$$

$$\text{DSC} = \frac{2 \cdot \text{TP}}{2 \cdot \text{TP} + \text{FP} + \text{FN}} \tag{8}$$

Where:

- TP is true positive.
- TN is true negative.
- FP is false positive.
- FN is false negative.

4.3 Implementation Details

Our method was implemented using the PyTorch[1] and the Albumentations[2] libraries. For policy search, we utilized AutoAlbument, a module within Albumentations, which leverages Faster AutoAugment to find the optimal policy. The U-Net implementation, however, employed PyTorch. Albumentations was used to generate augmented images based on the searched policy and to revert the transformations for Test Time Augmentation. The entire training process was conducted on a server with the following hardware configuration:

1. CPU: 16 vCPUs with a clock speed of 2667v4.
2. RAM: 48 GB RAM, providing large processing capabilities and high performance for tasks that require a lot of memory.
3. GPU: Nvidia Tesla P40 GPU with 24GB of memory, used for parallel computation tasks and high-end graphics processing. It has 3840 CUDA cores, providing powerful parallel computation capability and high performance.
4. Storage: 500GB SSD NVME, offering fast data access speed and high storage performance.

4.4 Experimental Settings

Four separate experiments were conducted on the BrEaST dataset to evaluate the effectiveness of our proposed method:

1. U-Net without data augmentation: This experiment evaluated the performance of the U-Net model trained solely on the original dataset, without incorporating any data augmentation techniques.
2. U-Net with data augmentation: In this experiment, the U-Net model was trained on training data that had been augmented using the searched policy. Ten unique augmented images were generated for each original image.
3. U-Net with Test Time Augmentation: This experiment employed the trained U-Net model from experiment 1 to make predictions using the Test Time Augmentation approach described earlier (Section 3.3).
4. U-Net with both data augmentation and Test Time Augmentation: This experiment leveraged the trained U-Net model from experiment 2 to make predictions while incorporating Test Time Augmentation.

For all experiments, the U-Net model was trained using the ADAM optimizer with a learning rate of 1e-4 and a batch size of 8. During policy searching, the learning rate for the policy was set to 1e-4, and the critic learning rate was set to 9e-4. Both learning rates utilized the ADAM optimizer with initial decay rates of 0 and 0.999, respectively. The batch size for policy searching was set to 16.

[1] https://pytorch.org/
[2] https://albumentations.ai/

4.5 Main Result

Method	DSC	IoU	Acc.	Sen.	Spec.
U-Net	0.708	0.592	0.961	0.746	**0.986**
U-Net + policy	0.723	0.624	0.963	0.751	**0.986**
U-Net + TTA	0.694	0.582	0.958	0.759	0.967
U-Net + policy + TTA (ours)	**0.754**	**0.652**	**0.968**	**0.818**	0.977

Table 2. Segmentation performance of U-Net model with and without policy and Test Time Augmentation.

Table 2 presents the segmentation performance metrics for the four experiments conducted. Our proposed method, which incorporates both data augmentation and Test Time Augmentation, achieves a significant improvement of 0.046 compared to the baseline U-Net model. While enhancements are observed in other metrics as well, there is a slight decrease in specificity. Interestingly, using U-Net with Test Time Augmentation alone (experiment 3) leads to a reduction in performance compared to the baseline.

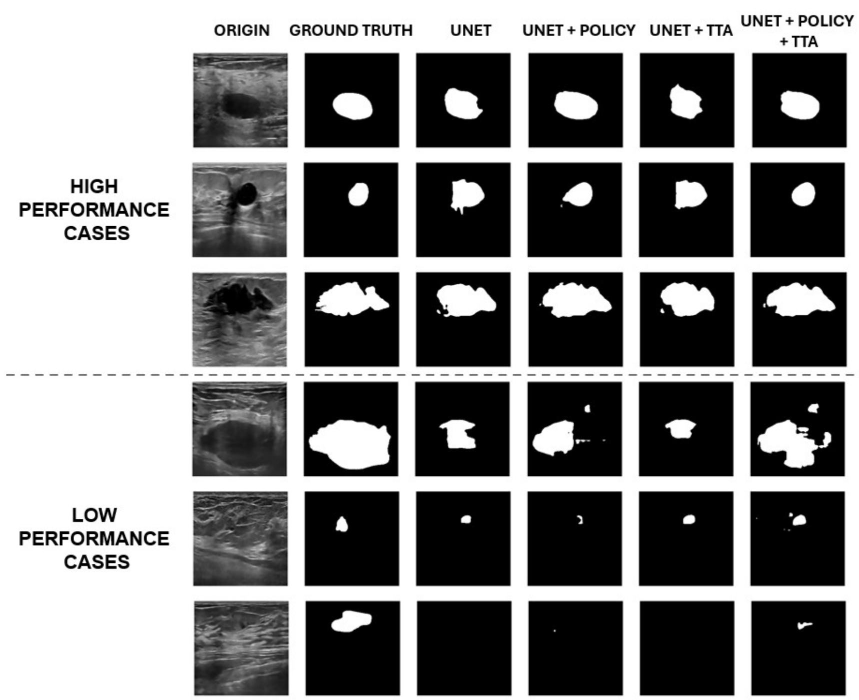

Fig. 6. High and Low performance cases.

High performance cases
In Fig. 6, The U-Net model without data augmentation or Test Time Augmentation performs well on images with bright tumors that have a clear contrast with the background. However, in the second example, this model fails to distinguish the tumor from adjacent black streaks. Additionally, the predicted mask generated by the U-Net with Test Time Augmentation tends to be slightly smaller compared to the baseline U-Net.

In Fig. 6, The U-Net with data augmentation (experiment 2) significantly improves the accuracy of the predicted mask compared to the baseline U-Net. Our proposed method (combining both data augmentation and Test Time Augmentation) further refines the segmentation, enabling a clear distinction between the tumor and the black streaks, which posed a challenge for the baseline U-Net.

Low performance cases
In cases where the brightness levels are low and both the tumor and background have similar intensities, particularly for small tumors, the U-Net model struggles with mask prediction, failing entirely in the last case and achieving low accuracy in two others. Despite these challenges, our proposed method demonstrates potential for improved accuracy in these difficult scenarios.

5 Conclusion

This paper presents a novel method that leverages Auto Augment techniques for medical image segmentation, specifically focusing on breast ultrasound images. Our approach aims to address the challenge of limited datasets, a common hurdle in the medical imaging domain. Overall, the method demonstrates promising results in enhancing the performance of the U-Net model when dealing with small datasets.

Our future goal is to refine the proposed method by developing a more robust Test Time Augmentation approach that utilizes the policy learned from the Auto Augment algorithms. This could potentially lead to further improvements in segmentation accuracy.

6 Acknowledgement

We would like to express our sincere gratitude to the individuals, organizations, and especially the teacher who significantly contributed to the completion of this data generation research. Their unwavering support and guidance were invaluable throughout the entire process. We thank you all for your significant contributions.

References

1. G. Farhat, "Diagnostic ultrasound imaging : Inside out," *Z Ultraschall in der*, vol. 25, pp. 407 – 407, 2004. [Online]. Available: https://api.semanticscholar.org/CorpusID:60603067

2. O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” *ArXiv*, vol. abs/1505.04597, 2015. [Online]. Available: https://api.semanticscholar.org/CorpusID:3719281
3. O. Oktay, J. Schlemper, L. L. Folgoc, M. J. Lee, M. P. Heinrich, K. Misawa, K. Mori, S. G. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, “Attention u-net: Learning where to look for the pancreas,” *ArXiv*, vol. abs/1804.03999, 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:4861068
4. Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, “Unet++: A nested u-net architecture for medical image segmentation,” *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support : 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, held in conjunction with MICCAI 2018, Granada, Spain, S...*, vol. 11045, pp. 3–11, 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:50786304
5. S. Cai, Y. Tian, H. Lui, H. Zeng, Y. Wu, and G. Chen, “Dense-unet: a novel multi-photon in vivo cellular image segmentation model based on a convolutional neural network.” *Quantitative imaging in medicine and surgery*, vol. 10 6, pp. 1275–1285, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:219684633
6. E. Göçeri, “Medical image data augmentation: techniques, comparisons and interpretations,” *Artificial Intelligence Review*, pp. 1 – 45, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:257642844
7. Z. Hussain, F. Gimenez, D. Yi, and D. Rubin, “Differential data augmentation techniques for medical imaging classification tasks,” *AMIA ... Annual Symposium proceedings. AMIA Symposium*, vol. 2017, pp. 979–984, 2017. [Online]. Available: https://api.semanticscholar.org/CorpusID:21158732
8. R. Hataya, J. Zdenek, K. Yoshizoe, and H. Nakayama, “Faster autoaugment: Learning augmentation strategies using backpropagation,” *ArXiv*, vol. abs/1911.06987, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:208139159
9. S. Yang, W.-T. Xiao, M. Zhang, S. Guo, J. Zhao, and S. Furao, “Image data augmentation for deep learning: A survey,” *ArXiv*, vol. abs/2204.08610, 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:248240105
10. E. D. Cubuk, B. Zoph, D. Mané, V. Vasudevan, and Q. V. Le, “Autoaugment: Learning augmentation strategies from data,” *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 113–123, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:196208260
11. D. Ho, E. Liang, I. Stoica, P. Abbeel, and X. Chen, “Population based augmentation: Efficient learning of augmentation policy schedules,” *ArXiv*, vol. abs/1905.05393, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:153312991
12. S. Lim, I. Kim, T. Kim, C. Kim, and S. Kim, “Fast autoaugment,” 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:141477988
13. M. Negassi, D. Wagner, and A. Reiterer, “Smart(sampling)augment: Optimal and efficient data augmentation for semantic segmentation,” *ArXiv*, vol. abs/2111.00487, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:240354515
14. D. Shanmugam, D. W. Blalock, G. Balakrishnan, and J. V. Guttag, “Better aggregation in test-time augmentation,” *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 1194–1203, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:238634826
15. M. Kimura, “Understanding test-time augmentation,” *ArXiv*, vol. abs/2402.06892, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:244916966

16. J. Son and S. Kang, "Efficient improvement of classification accuracy via selective test-time augmentation," *Inf. Sci.*, vol. 642, p. 119148, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:258779527
17. M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein gan," *ArXiv*, vol. abs/1701.07875, 2017. [Online]. Available: https://api.semanticscholar.org/CorpusID:13943041
18. I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of wasserstein gans," 2017. [Online]. Available: https://api.semanticscholar.org/CorpusID:10894094
19. I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," 2014.
20. K. O'Shea and R. Nash, "An introduction to convolutional neural networks," *ArXiv*, vol. abs/1511.08458, 2015. [Online]. Available: https://api.semanticscholar.org/CorpusID:9398408
21. N. Moshkov, B. Máthé, A. Kertész-Farkas, R. Hollandi, and P. Horváth, "Test-time augmentation for deep learning-based cell segmentation on microscopy images," *Scientific Reports*, vol. 10, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:208583332
22. A. Pawlowska, A. Ćwierz-Pieńkowska, A. Domalik, D. Jaguś, P. Kasprzak, R. Matkowski, Łukasz Fura, A. Nowicki, and N. Żołek, "Curated benchmark dataset for ultrasound based breast lesion analysis," *Scientific Data*, vol. 11, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:267362150
23. O. V. Michailovich and A. Tannenbaum, "Despeckling of medical ultrasound images," *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, vol. 53, pp. 64–78, 2006. [Online]. Available: https://api.semanticscholar.org/CorpusID:13987272
24. X. Li, Y. Wang, Y. Zhao, and Y. Wei, "Fast speckle noise suppression algorithm in breast ultrasound image using three-dimensional deep learning," *Frontiers in Physiology*, vol. 13, 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:248152976
25. B. K. S. Kumar, "Image denoising based on non-local means filter and its method noise thresholding," *Signal, Image and Video Processing*, vol. 7, pp. 1211 – 1227, 2012. [Online]. Available: https://api.semanticscholar.org/CorpusID:2711386
26. D. Müller, I. Soto-Rey, and F. Kramer, "Towards a guideline for evaluation metrics in medical image segmentation," *BMC Research Notes*, vol. 15, 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:246824012