Proceedings of CONF-CDS 2025 Symposium: Application of Machine Learning in Engineering
DOI: 10.54254/2755-2721/2025.AST26110

Data Augmentation Techniques and Their Impact on the Performance of Instance Segmentation Models

Yuehang Yang

School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, Singapore
YUEHANG001@e.ntu.edu.sg

Abstract. In fields such as medical diagnosis and intelligent driving, there is a growing demand for accurate instance segmentation. This paper mainly reviews data augmentation techniques and their impact on the performance of instance segmentation models. The purpose of this study is to evaluate the effectiveness of different augmentation strategies in alleviating data insufficiency and improving model generalization. This paper uses a document analysis method to summarize the results of image transformation-based (geometry, color, filters, random erasure, image blending) and depth network-based (GAN, neural style transfer) enhancement methods. The results show that geometric and color transformations enhance the model's adaptability to lighting variation, random erasing and image mixing enhance its robustness to occlusion and complex backgrounds, and synthetic samples generated by GAN and neural style transfer further optimize segmentation accuracy. The study found that a reasonable combination of multiple augmentation strategies can significantly improve instance segmentation performance, providing practical references for related applications and algorithm improvements.

Keywords: Data Augmentation, Image Segmentation, Neural Networks

1. Introduction

Image segmentation is the technology and process of dividing an image into several specific regions with unique properties and identifying objects of interest. It is typically divided into semantic segmentation and instance segmentation. Semantic segmentation assigns each pixel in the input image to a specific category, thereby identifying different object regions; instance segmentation can even distinguish between different instances of the same category in an image. Traditional segmentation methods mainly rely on manually designed features or rules for region division. These methods have limited performance when dealing with complex shapes, dynamic backgrounds, and noise, and often require manually designed features for different situations [1]. With the development of deep learning, models such as convolutional neural networks can automatically learn image features without manual design, thereby demonstrating significant advantages in image segmentation tasks. However, deep learning segmentation models have extremely high requirements for large-scale labelled data. In fields such as medicine, images involve personal privacy, and large amounts of labelled data are costly.

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2025.AST26110

Data augmentation techniques can alleviate the problem of insufficient labelled data. Data augmentation is a technique that expands the training set by generating new samples that have been perturbed or synthesized based on existing samples. Therefore, data augmentation can effectively alleviate the risk of overfitting caused by insufficient training data or category imbalance without increasing collection costs, thereby improving the model's generalization and prediction accuracy.

This paper focuses on the performance improvement of instance segmentation models using different data augmentation techniques and their applicable situations. The study uses literature review and comparative analysis methods to systematically summarize and evaluate the performance of various data augmentation strategies in multiple instance segmentation tasks. This study helps provide efficient data augmentation solutions for training instance segmentation models in practical applications, and also provides theoretical and practical references for subsequent researchers to select or design augmentation strategies.

2. Introduction to data augmentation techniques

Data augmentation methods can be divided into two main categories: image transformation-based methods and deep network-based methods.

2.1. Image transformation

Image transformation data augmentation techniques generate new samples by performing geometric or optical operations on the original image, including geometric transformations, color space transformations, kernel filters, random erasing, and image mixing. Common geometric transformations include rotation, translation, flipping, scaling, cropping, and shearing; color space transformations include operations such as changing brightness, contrast, saturation, and hue; kernel filters use convolution kernel filters (such as smoothing, sharpening, edge detection, etc.) to change image details, making the model more robust to texture and noise; random erasing randomly erases a region of an image and fills it with random values, increasing the model's robustness to local occlusion; image mixing generates new training samples by mixing two images at a certain ratio, enabling the model to better capture cross-sample information [2].

2.2. Deep networks

Data augmentation methods based on deep networks utilize generative models or learning models to create new samples. Generative adversarial networs (GAN) is a typical method, which generates high-quality synthetic images through adversarial training. GAN consists of a generator and a discriminator. The generator continuously learns to create new images with extremely high realism, while the discriminator learns to distinguish between real and fake images. Extensive research has shown that synthetic images generated by GANs can significantly improve the robustness of models to data distributions. Neural style transfer is a technique that applies the artistic style of one image to another, altering the texture of the image while preserving its semantic content, thereby providing more sample variations for the training set.

DOI: 10.54254/2755-2721/2025.AST26110

3. The impact of data augmentation techniques on the performance of image segmentation models

3.1. Evaluation metrics

Common metrics used to evaluate the performance of image segmentation models include intersection over union (IoU), average precision (AP), Dice similarity coefficient (DSC), and average accuracy (AA).

3.1.1. Intersection over Union (IoU)

Intersection over Union (IoU) is a classic metric used in computer vision and pattern recognition to measure the degree of overlap between two sets or two regions. The mathematical definition of IoU is that, given a predicted region A and a true region B, IoU is equal to the intersection area of the two regions divided by the union area, which is

$$IoU = \frac{\text{Segmented Image} \cap \text{Ground Truth Image}}{\text{Segmented Image} \cup \text{Ground Truth Image}} \tag{1}$$

The IoU value is in [0,1]. When two regions completely overlap, the value is 1; when there is no overlap, the value is 0. IoU is insensitive to scale, position, and rotation, focusing only on the spatial overlap ratio. Therefore, it can objectively reflect the positioning accuracy without being affected by absolute size.

3.1.2. Average Precision (AP)

Average Precision (AP) is an important evaluation metric in image segmentation, particularly in instance segmentation. It assesses model performance by calculating precision and recall at different confidence thresholds and computing the area under the precision-recall curve (AUC). The AP value ranges from 0 to 1, with higher values indicating better model performance.

When calculating AP, first collect the model's predictions for the test images, including the predicted instance masks and confidence scores, as well as the true ground truth instance masks. Set an IoU threshold to determine whether the predicted instances match the ground truth instances. Sort all predictions from highest to lowest confidence. For each predicted instance, calculate its mask IoU with the ground truth instance. If the IoU exceeds the threshold and the ground truth instance has not yet been matched, it is recorded as a true positive (TP); otherwise, it is a false positive (FP). While iterating through the sorted list of predictions, accumulate TP and FP, calculate the accuracy and recall rate for each point, and plot the PR curve. AP is the area under the PR curve, which is usually calculated by interpolation.

3.1.3. Dice Similarity Coefficient (DSC)

The Dice similarity coefficient (DSC) is another method for measuring the similarity between predicted segmentation results and actual conditions. It takes into account both false positives and false negatives. The formula for calculating this coefficient is

DOI: 10.54254/2755-2721/2025.AST26110

$$DSC = \frac{2 \times (\text{Segmented Image} \cap \text{Ground Truth Image})}{\text{Area of Predicted Segmentation} + \text{Area of Ground Truth}} \quad (2)$$

In the context of image segmentation, DSC is used to measure the degree of overlap between the predicted segmentation and the actual segmentation, taking into account both true positives and false positives. DSC values range from 0 to 1, with higher values indicating a better degree of overlap between the predicted segmentation and the actual segmentation. A DSC value of 1 means that the predicted segmentation completely matches the actual segmentation.

3.1.4. Average Accuracy (AA)

The average accuracy rate (AA) is the average of the pixel classification accuracy rates for all categories, which is the average value of the sum of all pixel categories and the total number of pixels, reflecting the overall segmentation accuracy of the model. The calculation formula is

$$\left(\frac{1}{N}\right) \times \sum_{i=1}^{N} \Sigma \left(\frac{\text{True Positives}_i + \text{True Negatives}_i}{\text{Total Pixels}_i}\right) \quad (3)$$

Among these, the category True Positivesi represents the number of pixels correctly predicted as i, and True Negativesi represents the number of pixels correctly predicted as i. Among these, the category True Positivesi represents the number of pixels correctly predicted as i, and True Negativesi represents the number of pixels correctly predicted as i. In the evaluation framework for semantic segmentation, calculating the average accuracy helps prevent the 'majority class'—which holds an absolute advantage in terms of pixels or samples—from exerting an excessive influence on model training and evaluation results in datasets with extremely imbalanced category distributions. This prevents the phenomenon where minority classes are overlooked or misclassified yet remain difficult to reflect in the metrics.

3.2. Effects of image transformation enhancement

3.2.1. Geometric transformation

Geometric transformations are an important method of data augmentation, which can increase the robustness of models to changes in target space position, posture, size, etc. In scenarios where there are insufficient samples or category imbalances, geometric augmentation such as rotation and flipping of minority categories can effectively expand their sample size, thereby improving model performance. Lopez de la Rosa et al. used the ResNet50 convolutional neural network as the base model for the semiconductor defect classification task. They increased data diversity by applying geometric transformations such as rotation, translation, and scaling to the training data. The experimental results showed that the dataset enhanced by geometric transformations significantly improved the model's performance, with the classification model's F1 score increasing by 3.74% compared to the original dataset [3]. Chen et al. found that in tongue segmentation tasks, the application of geometric transformations such as rotation, flipping, and cropping can improve segmentation accuracy by 5%–20% compared to non-augmentation strategies [4].

DOI: 10.54254/2755-2721/2025.AST26110

3.2.2. Color transformation

Color transformation improves the model's robustness to lighting and color changes by adjusting image attributes such as brightness, contrast, and hue. Contrast enhancement improves visual effects by adjusting the contrast between light and dark areas in an image to make details more prominent. Sharpness enhancement improves the clarity of image edges and reinforces texture features. Color shift alters the color distribution of an image, such as by adjusting hue or saturation, to simulate different color conditions. Nanni et al. used color space transformation technology to generate diverse training images to enhance data richness and enable the model to capture a wider range of image features. The experimental results show that combining color space transformation with other enhancement methods (such as geometric transformation) significantly improves classification accuracy on multiple benchmark datasets [5].

3.2.3. Random erasing

Random Erasing is a common data augmentation technique widely used in image processing and computer vision tasks, aimed at enhancing the robustness and generalization capabilities of models by randomly removing portions of an image. Ma et al. proposed a novel data augmentation method called ‘Erase, then Redraw,’ which extends the concept of traditional random erasing to improve the performance of free space detection in image segmentation tasks. Experimental results on the KITTI road dataset and Cityscapes dataset demonstrate that this method outperforms traditional data augmentation techniques such as RandomErasing, Cutout, GridMask, and CutMix, achieving the best performance across metrics including accuracy, precision, recall, F1 score, and mean intersection over union (mIoU) [6].

3.2.4. Image mixing

Image mixing is a data augmentation technique that aims to generate virtual samples by linearly combining two training images and their corresponding labels, thereby improving the generalization ability and robustness of deep learning models. Isaksson et al. conducted experiments on a dataset of 100 T2-weighted MRI images, using mixing weights drawn from a symmetric beta distribution (parameter $\alpha = 0.5$) in combination with a deep convolutional neural network for five-fold cross-validation. The results showed that compared with training without Mixup, Mixup training improved the average Dice score by 0.19 ($p=0.05$) and reduced the mean surface distance (MSD) by 0.78 ($p=0.05$), corresponding to performance improvements of 3.15% and -13.9%, respectively [7].

3.3. Impact based on deep networks

3.3.1. Generative Adversarial Network (GAN)

Generative adversarial network (GAN) can synthesize high-fidelity data through adversarial learning between generators and discriminators, thereby producing high-quality synthetic samples to expand the training set. The synthetic images generated by GAN are indistinguishable from real images and can effectively improve the robustness of the model. In segmentation tasks, training with GAN-synthesized images can improve boundary recognition and small object detection. Kugelman et al.'s research shows that the combined design of conditional generation, discriminator architecture reuse and semi-supervised pre-training can significantly improve the accuracy of OCT layer segmentation in small sample scenarios of 1-5 cases. However, it still needs to be optimised for challenging

DOI: 10.54254/2755-2721/2025.AST26110

categories (such as CSI boundaries), reflecting the constraints of pathological complexity on generation quality [8].

3.3.2. Neural style transfer

Neural style transfer provides more visual variants of training samples by changing the overall style of an image without altering its semantic structure. ZhiJie Lin et al. explored the application potential of neural style transfer in image segmentation data augmentation by proposing an image style transfer algorithm based on semantic segmentation. The algorithm uses a semantic segmentation network to automatically extract semantic information from style images and content images, and uses this information to precisely guide style transfer, effectively avoiding semantic mismatch issues that may arise in semantically sensitive scenarios with traditional style transfer algorithms. This method is applied to the data augmentation stage of image segmentation tasks, generating images with diverse styles but consistent semantic information, significantly enhancing the diversity of the dataset [9].

4. Conclusion

This paper reviews commonly used data augmentation techniques in the field of image segmentation and their impact on the performance of instance segmentation models. Among them, geometric transformations can effectively enhance the model's adaptability to changes in object shape and perspective, usually resulting in significant performance improvements; color and filter transformations enhance the model's robustness to lighting and texture variations; and strategies such as random erasing and image mixing further improve the model's generalization ability by increasing occlusion and composite samples. In deep network-based generation methods, generative adversarial network can synthesize high-quality, diverse images, expanding the semantic and statistical diversity of the training set and thereby improving the segmentation model's ability to recognize complex scenes; neural style transfer provides the model with new texture variations without altering semantic content by altering the visual style of images, thereby introducing additional training sample diversity. Overall, existing research shows that reasonable data augmentation strategies generally help improve the accuracy and robustness of instance segmentation models.

In future research, with the development of generative models (such as diffusion models) and automatic optimization methods, future work can continue to focus on designing more intelligent and task-customized data augmentation schemes to further advance the development of instance segmentation technology.

References

[1] Archana, R., & Jeevaraj, P. E. (2024). Deep learning models for digital image processing: a review. Artificial Intelligence Review, 57(1), 11.

[2] Alomar, K., Aysel, H. I., & Cai, X. (2023). Data augmentation in classification and segmentation: A survey and new strategies. Journal of Imaging, 9(2), 46.

[3] de la Rosa, F. L., Gómez-Sirvent, J. L., Sánchez-Reolid, R., Morales, R., & Fernández-Caballero, A. (2022). Geometric transformation-based data augmentation on defect classification of segmented images of semiconductor materials using a ResNet50 convolutional neural network. Expert Systems with Applications, 206, 117731.

[4] Ranjan, R., Sankaranarayanan, S., Bansal, A., Bodla, N., Chen, J. C., Patel, V. M., ... & Chellappa, R. (2018). Deep learning for understanding faces: Machines may be just as good, or better, than humans. IEEE Signal Processing Magazine, 35(1), 66-83.

DOI: 10.54254/2755-2721/2025.AST26110

[5] Nanni, L., Paci, M., Brahnam, S., & Lumini, A. (2021). Comparison of different image data augmentation approaches. Journal of imaging, 7(12), 254.
[6] Ma F, Qi W, Zhao G, et al. Erase, then redraw: a novel data augmentation approach for free space detection using diffusion model [J]. arXiv preprint arXiv: 2409.20164, 2024.
[7] Isaksson, L. J., Summers, P., Raimondi, S., Gandini, S., Bhalerao, A., Marvaso, G., ... & Jereczek-Fossa, B. A. (2022). Mixup (sample pairing) can improve the performance of deep segmentation networks. Journal of Artificial Intelligence and Soft Computing Research, 12.
[8] Kugelman, J., Alonso-Caneiro, D., Read, S. A., Vincent, S. J., & Collins, M. J. (2024). Enhancing OCT patch-based segmentation with improved GAN data augmentation and semi-supervised learning. Neural Computing and Applications, 36(29), 18087-18105.
[9] Lin, Z., Wang, Z., Chen, H., Ma, X., Xie, C., Xing, W., ... & Song, W. (2021). Image style transfer algorithm based on semantic segmentation. IEEE Access, 9, 54518-54529.