

Article

Not peer-reviewed version

Data Augmentation Through Random Style Replacement

[Qikai Yang](#)^{*}, Cheng Ji, Huaiying Luo, [Panfeng Li](#), Zhicheng Ding

Posted Date: 31 March 2025

doi: 10.20944/preprints202503.2367.v1

Keywords: data augmentation; style transfer; style augmentation



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

Data Augmentation Through Random Style Replacement

Qikai Yang ^{1,*}, Cheng Ji ¹, Huaiying Luo ², Panfeng Li ³ and Zhicheng Ding ⁴

¹ University of Illinois Urbana-Champaign, Urbana, USA

² Cornell University, Ithaca, USA

³ University of Michigan, Ann Arbor, USA

⁴ Columbia University, New York, USA

* Correspondence: qikaiy2@illinois.edu

Abstract: In this paper, we introduce a novel data augmentation technique that combines the advantages of style augmentation and random erasing by selectively replacing image subregions with style-transferred patches. Our approach first applies a random style transfer to training images, then randomly substitutes selected areas of these images with patches derived from the style-transferred versions. This method is able to seamlessly accommodate a wide range of existing style transfer algorithms and can be readily integrated into diverse data augmentation pipelines. By incorporating our strategy, the training process becomes more robust and less prone to overfitting. Comparative experiments demonstrate that, relative to previous style augmentation methods, our technique achieves superior performance and faster convergence.

Keywords: data augmentation; style transfer; style augmentation

1. Introduction

Recent advancements in deep learning have driven significant progress in a wide range of computer vision tasks, including image classification, object detection, and semantic segmentation [1–10]. Despite these advances, many of these tasks continue to face a fundamental bottleneck: a lack of sufficient labeled data [11–14]. Annotating large-scale datasets is both time-consuming and costly, which can limit the applicability of deep neural networks in specialized or rapidly evolving domains. To mitigate this issue, data augmentation techniques are heavily utilized, artificially expanding and diversifying the training set so that models generalize more effectively.

Data augmentation has garnered significant attention in supervised learning research across a wide range of domains—including computer vision [15–17], natural language processing [18,19], graph learning [20–25], large language model [26], and the Internet of Things [27,28]—due to its ability to increase both the volume and diversity of training data, thereby enhancing model generalization and mitigating overfitting. Broadly, data augmentation strategies can be grouped into two categories: generative methods, which utilize models like Variational Autoencoders (VAEs), Generative Adversarial Networks (GANs), Large Language Models (LLMs), or diffusion-based frameworks to synthesize new data [29–36]; and traditional methods, which rely on transformations such as random cropping, flipping, rotations, color jittering, and histogram equalization to modify existing samples. While both approaches aim to expose the model to a wider variety of conditions, thus reducing overfitting, traditional augmentation strategies may not fully capture the complexity of real-world variability. Consequently, several studies have explored more refined methods, including style augmentation and random erasing [37–41]. Style augmentation employs style transfer to alter the visual attributes of training images while preserving their semantic content, thereby increasing robustness to differences in texture, color, and contrast. Random erasing, on the other hand, randomly occludes or replaces subregions of an image, making models more resilient to missing or corrupted information. In this

paper, we revisit these traditional approaches—particularly focusing on their potential to advance the efficacy of data augmentation in supervised learning.

In this paper, we introduce a novel data augmentation method that merges style augmentation with random erasing. Our approach involves applying a random style transfer to an image, followed by replacing specific subregions with patches from the style-transferred version. This technique enhances robustness against style variations and occlusion-like effects. It integrates smoothly with existing style transfer frameworks and fits easily into standard data augmentation pipelines, as illustrated in Figure 1.

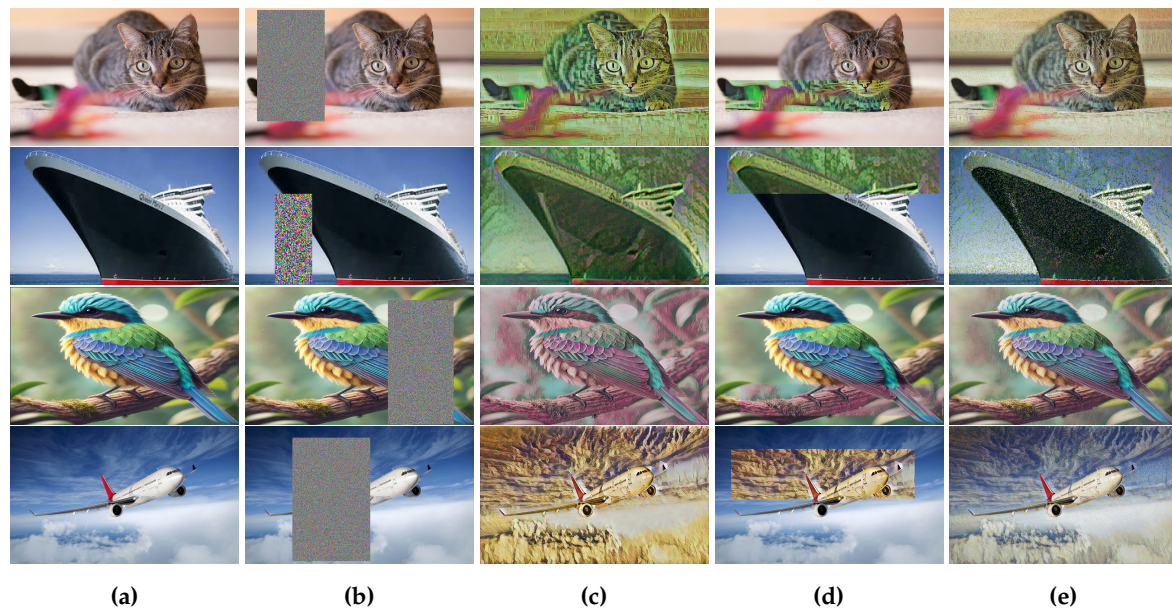


Figure 1. Examples of random style transfer: we generate a style-transferred image and use it to patch original image in different ways (a) Input Image. (b) Random-Region-Erased Image. (c) Style Transfer. (d) Random Style Replacement. (e) Random Style Replacement.

The main contributions of our work are as below:

1. We propose a technique that merges style augmentation and random erasing, offering benefits from both texture variation and structured occlusion.
2. We demonstrate through experiments that our approach reduces the risk of overfitting while achieving faster convergence compared to established style augmentation methods. Ease of Integration: Our strategy is parameter-free and can be readily adapted to a broad spectrum of computer vision tasks, making it a highly practical solution for data augmentation.

By leveraging this new augmentation method, we observe notable gains in model performance across different tasks, highlighting its potential to address the persistent challenge of limited labeled data in computer vision research.

2. Dataset

We tested our random style replacement method on the STL-10 dataset, which includes 5,000 training images and 8,000 test images, each with a resolution of 96×96 pixels across 10 classes [11]. We chose STL-10 due to its complex backgrounds and high resolution, which pose a substantial challenge for image classification, making it a robust benchmark. Additionally, the limited size of the training set highlights the effectiveness of our data augmentation technique in enhancing training data.

3. Methods

This section introduces our random style replacement method in details. We described the overall process of random style replacement and explain how we perform image patch.

3.1. Random Style Replacement

During training, random style replacement is applied with a certain probability p : for each image I in a mini-batch, there's a probability p that it undergoes style replacement and a probability $1 - p$ that it remains unchanged.

If selected, the image will be transformed into a new version with a partial style change. This random style replacement process consists of two steps: generating a complete-style-transferred image and merging it with the original image by certain patching methods. The procedure is shown in Algorithm 1.

Style transfer refers to a class of image processing algorithms that alter the visual style of an image while preserving its semantic content. For style transfer to be part of a data augmentation technique, it needs to be a both fast and random algorithm capable of applying a broad range of styles. Therefore, we adopt the approach of Jackson et al., which efficiently generates a completely style-transferred image by incorporating randomness on the fly without requiring heavy computations [37].

The generated style-transferred image will then be used to patch the original image, creating an augmented image. There are multiple patching methods, and we adopt the two most common ones: patching by a random subregion and patching randomly selecting individual pixels. To avoid bias in data augmentation, we employed random style transferring to ensure diverse and uniform modifications across all image types, enhancing model generalization.

Algorithm 1: Random Style Replacement Procedure

Input : Input image I ;
Augmentation probability p ;
Patch mode $pMode$;
Output: Augmented image I^* .
Initialization: $p_1 \leftarrow \text{Rand}(0, 1)$.

```

1 if  $p_1 \geq p$  then
2    $I^* \leftarrow I$ ;
3   return  $I^*$ .
4 else
5    $I' \leftarrow \text{randomStyleTransfer}(I)$ ;
6    $I^* \leftarrow \text{randomPatch}(I, I', pMode)$ ;
7   return  $I^*$ .
8 end

```

3.2. Random Patch

Random patch is to patch a image based on another image. Here, we provided a detailed explanation of random patch by subregion. This method copies a randomly selected region from the style-transferred image onto the original image. Specifically, it randomly selects a rectangle region I_e within the image and overwrite all its pixels with those from the style-transferred image.

Firstly we will determine the shape of the patching area I_e . Assume the training image has dimensions $W \times H$ and an area $S = W \times H$. We randomly initialize the area of the patched rectangle region to S_e , where $\frac{S_e}{S}$ falls within the range defined by the minimum s_l and maximum s_h . Similarly, the aspect ratio of the rectangle region, denoted as r_e , is randomly chosen between r_l and r_h . Given those, the dimensions of I_e are computed as $H_e = \sqrt{S_e \times r_e}$ and $W_e = \sqrt{\frac{S_e}{r_e}}$.

Next, we randomly select a point $\mathcal{P} = (x_e, y_e)$ within I to serve as the lower-left corner of I_e . If the selected region I_e are completely inside I (i.e. $x_e + W_e \leq W$ and $y_e + H_e \leq H$), we define it as the selected rectangular region. Otherwise, we repeat the selection process until a valid I_e is found. The whole procedure for selecting the rectangular region and applying the patch to original image is illustrated in Algorithm 2.

Algorithm 2: Random Patch by Subregion

Input : Input image I ;
Utility image I' ;
Patched area ratio range s_l and s_h ;
Patched aspect ratio range r_l and r_h .

Output: Patched image I^* .

```

1  $S_e \leftarrow \text{Rand}(s_l, s_h) \times S$ ;
2  $r_e \leftarrow \text{Rand}(r_l, r_h)$ ;
3  $H_e \leftarrow \sqrt{S_e \times r_e}$ ,  $W_e \leftarrow \sqrt{\frac{S_e}{r_e}}$ ;
4 while True do
5    $x_e \leftarrow \text{Rand}(0, W)$ ,  $y_e \leftarrow \text{Rand}(0, H)$ ;
6   if  $x_e + W_e \leq W$  and  $y_e + H_e \leq H$  then
7      $I_e \leftarrow (x_e, y_e, x_e + W_e, y_e + H_e)$ ;
8      $I(I_e) \leftarrow I'(I_e)$ ;
9      $I^* \leftarrow I$ ;
10    return  $I^*$ .
11  end
12 end

```

4. Experiment

4.1. Experiment Settings

As mentioned in previous sections, we evaluated our random style replacement method for image classification using the well-known STL-10 dataset [11]. To ensure the effectiveness and fairness of our evaluation, we set our experiment conditions mostly the same as [37,38]. The benchmark networks we selected are ResNet18, ResNet50, ResNet101 and ResNet152 without pre-trained parameters [42].

In all experiments, instead of introducing more advanced optimizers or training procedures such as [43], we selected the Adam optimizer (momentum $\beta_1 = 0.5$, $\beta_2 = 0.999$, initial learning rate of 0.001) to align with the settings of other data augmentation methods [37,38,44]. For our setting of the style augmentation parameter, we selected the style interpolation parameter α as 0.5 and augmentation ratio as 1:1 [37]. All experiments are trained on RTX 4090 with 100 epochs.

1. Original dataset without any advanced data augmentation techniques.
2. Dataset with naive data augmentation by simply copying and stacking the original dataset to match the same augmentation ratio as other groups, along with some routine augmentation operations.
3. Dataset with random style replacement by subregion.
4. Dataset with random style replacement at the pixel level (with an independent probability $p = 0.5$).

4.2. Classification Evaluation

We evaluated our proposed data augmentation technique on the STL-10 dataset, with only 5,000 training images and 8,000 test color images. After the augmentation process, as mentioned in previous sections, the size of the augmented training set will double to 10,000 and the corresponding augmentation treatment will be randomly applied accordingly. We applied the same training settings as prior work, whose effectiveness in strategy and hyperparameter selection, including learning rate, has already been verified.

Our method achieved 81.6% classification accuracy in just 100 training epochs, as shown in Figure 2. This result is both faster and more accurate than the 80.8% accuracy after 100,000 epochs reported by Jackson et al. [37], highlighting our approach's efficiency and scalability.

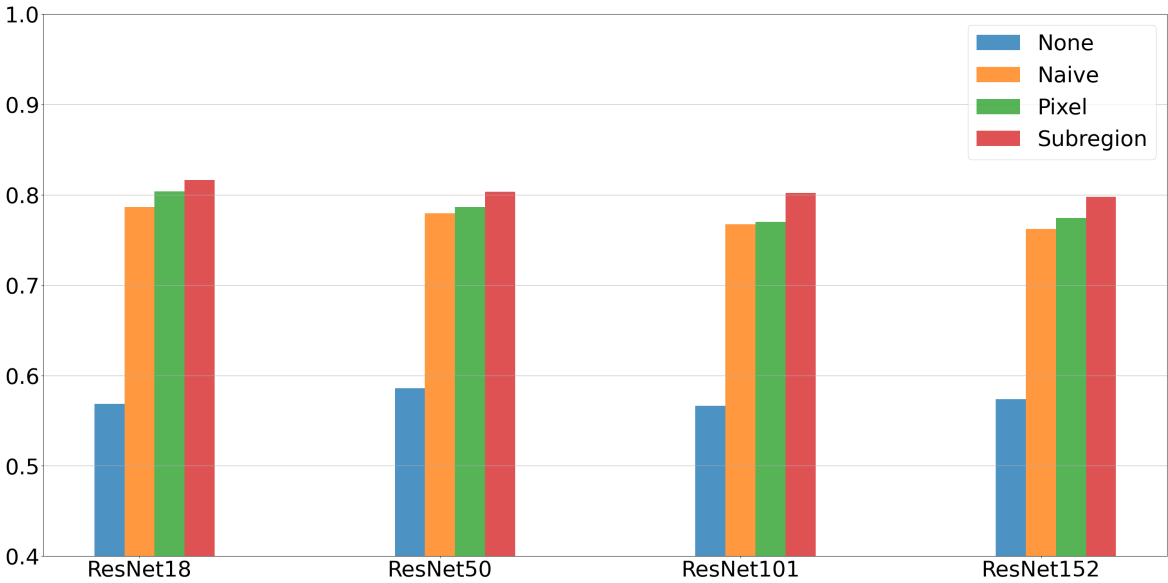


Figure 2. Classification Accuracy of ResNets on STL-10 test set. "None" represents original dataset. "Naive" represents dataset with naive data augmentation by simply stacking the original dataset. "Pixel" represents dataset with random style replacement at the pixel level. "Subregion" represents dataset with random style replacement by subregion.

We also tested our data augmentation technique across various network architectures including ResNet18, ResNet50, ResNet101, and ResNet152, where it consistently outperformed others, demonstrating its robustness and versatility for a wide range of computer vision tasks.

Furthermore, our findings support those of Zhong et al. [38], who found that erasing entire subregions is more effective than pixel-level erasing. Similarly, our data shows that random style replacement within subregions is a superior augmentation strategy, enhancing the training data’s representational richness and contributing to faster model convergence and improved performance. This strategy maintains structural integrity and introduces variations that reflect the natural diversity of real-world datasets.

To confirm the effectiveness of our data augmentation technique, we analyzed the test loss of each method on the STL-10 test set, as shown in Fig. 3. In contrast to the naive dataset, whose test loss stops converging after the 20th epoch, all augmented datasets show improved convergence speed and reduced loss variability. Notably, the style augmentation strategy that randomly replaces subregions achieves the fastest convergence and the most stable training process. Despite varying effectiveness in stabilizing training loss across ResNets, our method’s performance remains consistently stable.

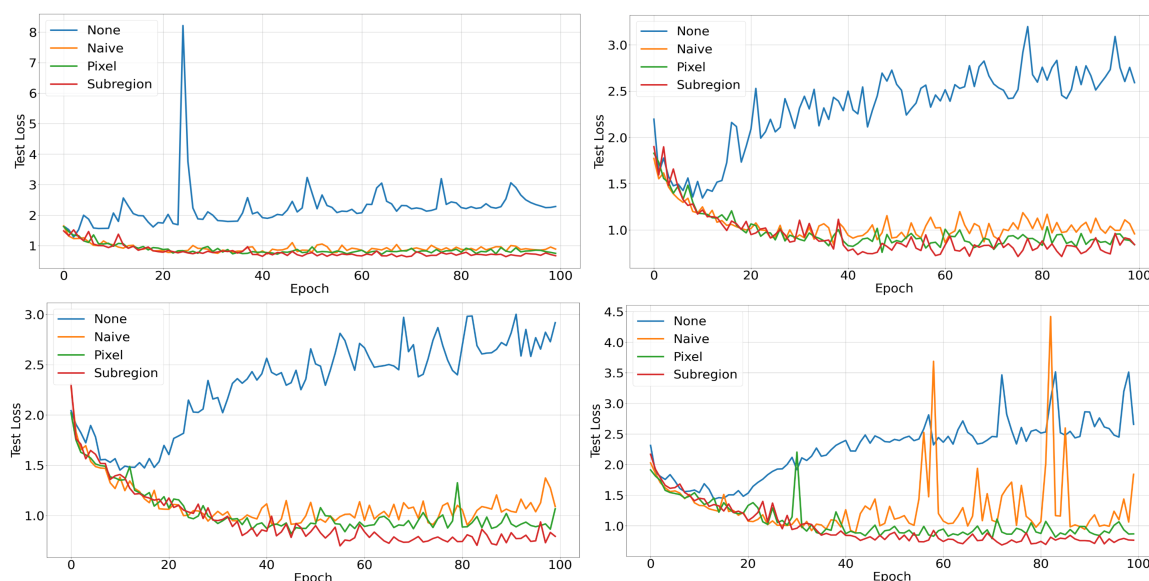


Figure 3. Loss of ResNets on STL-10 test set. "None" represents original dataset. "Naive" represents dataset with naive data augmentation by simply stacking the original dataset. "Pixel" represents dataset with random style replacement at the pixel level. "Subregion" represents dataset with random style replacement by subregion.

5. Conclusions

In conclusion, our random style replacement strategy offers a practical and scalable data augmentation solution for the STL-10 dataset and beyond. By innovatively combining [37] and [38], our proposed data augmentation framework demonstrates superior performance and achieves faster convergence. By randomly replacing subregions rather than individual pixels, we preserve critical structural information while introducing meaningful variability, resulting in faster training convergence and higher accuracy. Our experiments with multiple ResNet architectures consistently verify the robustness of this method, showcasing its versatility for diverse computer vision applications.

6. Future Work

Our random style replacement method has shown promising results, yet further validation is needed to confirm its wider applicability. It is crucial to test this technique across various datasets and tasks to establish its generalizability and identify any limitations. Additionally, the convergence speed observations require confirmation through further experiments involving diverse datasets and network architectures. Moreover, integrating Large Language Models (LLMs) guided approaches [45–52] could enhance the method. These approaches would use LLMs to guide style replacement, potentially selecting optimal subregions for style transfer based on foreground and background information, thus enabling more meaningful and effective transformations.

References

1. Li, K.; et al. Deep reinforcement learning-based obstacle avoidance for robot movement in warehouse environments. In Proceedings of the ICCASIT. IEEE, 2024.
2. Li, K.; Liu, L.; et al. Research on reinforcement learning based warehouse robot navigation algorithm in complex warehouse layout. In Proceedings of the ICAICA. IEEE, 2024, pp. 296–301.
3. Zhou, Y.; et al. Evaluating modern approaches in 3d scene reconstruction: Nerf vs gaussian-based methods. In Proceedings of the DOCS, 2024, pp. 926–931.
4. Li, X.; Chang, J.; Li, T.; Fan, W.; Ma, Y.; Ni, H. A vehicle classification method based on machine learning. In Proceedings of the PRML, 2024, pp. 210–216.
5. Jin, Y.; et al. Scam Detection for Ethereum Smart Contracts: Leveraging Graph Representation Learning for Secure Blockchain. *arXiv* 2024.

6. Yang, J.; Liu, J.; Yao, Z.; Ma, C. Measuring digitalization capabilities using machine learning. *Research in International Business and Finance* **2024**, *70*, 102380. <https://doi.org/https://doi.org/10.1016/j.ribaf.2024.102380>.
7. Dan, H.C.; Lu, B.; Li, M. Evaluation of asphalt pavement texture using multiview stereo reconstruction based on deep learning. *Construction and Building Materials* **2024**, *412*, 134837.
8. Dan, H.C.; Huang, Z.; Lu, B.; Li, M. Image-driven prediction system: Automatic extraction of aggregate gradation of pavement core samples integrating deep learning and interactive image processing framework. *Construction and Building Materials* **2024**, *453*, 139056.
9. Li, P.; Yang, Q.; Geng, X.; Zhou, W.; Ding, Z.; Nian, Y. Exploring diverse methods in visual question answering. In Proceedings of the 2024 5th International Conference on Electronic Communication and Artificial Intelligence (ICECAI). IEEE, 2024, pp. 681–685.
10. He, L.; Ka, D.H.; Ehtesham-Ul-Haque, M.; Billah, S.M.; Tehranchi, F. Cognitive models for abacus gesture learning. In Proceedings of the Proceedings of the Annual Meeting of the Cognitive Science Society, 2023, Vol. 46.
11. Coates, A.; Ng, A.; Lee, H. An analysis of single-layer networks in unsupervised feature learning. In Proceedings of the PMLR, 2011, pp. 215–223.
12. Hastie, T.; et al. Unsupervised learning. *The elements of statistical learning: Data mining, inference, and prediction* **2009**, pp. 485–585.
13. Zheng, L.; et al. Heterogeneous contrastive learning for foundation models and beyond. In Proceedings of the ACM SIGKDD, 2024, pp. 6666–6676.
14. Li, Z.; et al. Can Graph Neural Networks Learn Language with Extremely Weak Text Supervision? *arXiv preprint arXiv:2412.08174* **2024**.
15. Ding, K.; et al. Data augmentation for deep graph learning: A survey. *ACM SIGKDD Explorations Newsletter* **2022**, *24*, 61–77.
16. Kumar, T.; et al. Image data augmentation approaches: A comprehensive survey and future directions. *IEEE Access* **2024**.
17. Huang, S.; et al. AR Overlay: Training Image Pose Estimation on Curved Surface in a Synthetic Way. In Proceedings of the CS & IT Conference Proceedings, 2024.
18. Bayer, M.; et al. Data augmentation in natural language processing: a novel text generation approach for long and short text classifiers. *IJMLC* **2022**, *14*, 135–150.
19. Chen, X.; Li, K.; Song, T.; Guo, J. Few-shot name entity recognition on stackoverflow. In Proceedings of the 2024 9th International Conference on Intelligent Computing and Signal Processing (ICSP). IEEE, 2024, pp. 961–965.
20. He, J.; et al. GIVE: Structured Reasoning with Knowledge Graph Inspired Veracity Extrapolation. *arXiv preprint arXiv:2410.08475* **2024**.
21. He, J.; Kanatsoulis, C.I.; Ribeiro, A. T-GAE: Transferable Graph Autoencoder for Network Alignment. *arXiv* **2023**.
22. Yang, Q.; et al. A comparative study on enhancing prediction in social network advertisement through data augmentation. In Proceedings of the MLISE, 2024.
23. Jin, Y.; et al. Representation and extraction of diesel engine maintenance knowledge graph with bidirectional relations based on BERT and the Bi-LSTM-CRF model. In Proceedings of the ICEBE. IEEE, 2021, pp. 126–133.
24. Sankar, A.; et al. Self-supervised role learning for graph neural networks. *Knowledge and Information Systems* **2022**.
25. Narang, K.; et al. Ranking User-Generated Content via Multi-Relational Graph Convolution. In Proceedings of the SIGIR, 2021, pp. 470–480.
26. Li, S.; Xu, H.; Chen, H. Focused ReAct: Improving ReAct through Reiterate and Early Stop. In Proceedings of the WinLP Workshop, 2024.
27. Wang, T.; Chen, Y.; et al. Data augmentation for human activity recognition via condition space interpolation within a generative model. In Proceedings of the ICCCN. IEEE, 2024, pp. 1–9.
28. Wang, T.; et al. Fine-grained Control of Generative Data Augmentation in IoT Sensing. *Advances in Neural Information Processing Systems* **2025**.
29. Lu, Y.; et al. Machine learning for synthetic data generation: a review. *arXiv preprint arXiv:2302.04062* **2023**.
30. Kingma, D.P.; Welling, M. Auto-Encoding Variational Bayes. *arXiv preprint arXiv:1312.6114* **2013**.
31. Goodfellow, I.J.; et al. Generative Adversarial Networks. *arXiv* **2014**.

32. Wang, Y.; et al. TWIN-GPT: digital twins for clinical trials via large language model. *ACM Transactions on Multimedia Computing, Communications and Applications*.
33. Ho, J.; et al. Denoising diffusion probabilistic models. *Advances in neural information processing systems* **2020**, *33*, 6840–6851.
34. Ji, Y.; et al. Mitigating the risk of health inequity exacerbated by large language models. *arXiv preprint arXiv:2410.05180* **2024**.
35. Ji, Y.; Li, Z.; et al. RAG-RLRC-LaySum at BioLaySumm: Integrating Retrieval-Augmented Generation and Readability Control for Layman Summarization of Biomedical Texts. *arXiv preprint arXiv:2405.13179* **2024**.
36. Yang, Z.; et al. HADES: Hardware Accelerated Decoding for Efficient Speculation in Large Language Models. *arXiv:2412.19925* **2024**.
37. Jackson, P.T.; et al. Style augmentation: data augmentation via style randomization. In *Proceedings of the CVPR workshops, 2019*, Vol. 6, pp. 10–11.
38. Zhong, Z.; et al. Random erasing data augmentation. In *Proceedings of the AAAI, 2020*, Vol. 34, pp. 13001–13008.
39. Ding, Z.; Li, P.; Yang, Q.; Li, S.; Gong, Q. Regional style and color transfer. In *Proceedings of the CVIDL. IEEE, 2024*, pp. 593–597.
40. Chen, Y.; Kang, Y.; Song, Y.; Vachha, C.; Huang, S. AI-Driven Stylization of 3D Environments. *arXiv preprint arXiv:2411.06067* **2024**.
41. Xu, X.; et al. Style Transfer: From Stitching to Neural Networks. In *Proceedings of the ICBASE. IEEE, 2024*, pp. 526–530.
42. He, K.; et al. Deep residual learning for image recognition. In *Proceedings of the CVPR, 2016*, pp. 770–778.
43. Xu, Q.; et al. A stochastic gda method with backtracking for solving nonconvex (strongly) concave minimax problems. *arXiv* **2024**.
44. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* **2014**.
45. Ding, Z.; Li, P.; Yang, Q.; Li, S. Enhance image-to-image generation with llava-generated prompts. In *Proceedings of the ISPDS. IEEE, 2024*, pp. 77–81.
46. Deng, Q.; et al. Composerx: Multi-agent symbolic music composition with llms. *arXiv preprint arXiv:2404.18081* **2024**.
47. Xu, H.; Ye, J.; Li, Y.; Chen, H. Can speculative sampling accelerate react without compromising reasoning quality? In *Proceedings of the ICLR, 2024*.
48. Xu, H.; et al. Restful-llama: Connecting user queries to restful apis. In *Proceedings of the EMNLP: Industry Track, 2024*, pp. 1433–1443.
49. Xu, H. Towards seamless user query to rest api conversion. In *Proceedings of the CIKM, 2024*, pp. 5495–5498.
50. Jin, Y.; et al. Adaptive Fault Tolerance Mechanisms of Large Language Models in Cloud Computing Environments. <https://arxiv.org/abs/2503.12228> **2025**.
51. Yang, Z.; et al. Research on Large Language Model Cross-Cloud Privacy Protection and Collaborative Training based on Federated Learning. <https://arxiv.org/abs/2503.12226> **2025**.
52. Li, K.; Chen, X.; Song, T.; Zhang, H.; Zhang, W.; Shan, Q. GPTDrawer: Enhancing Visual Synthesis through ChatGPT. *arXiv preprint arXiv:2412.10429* **2024**.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.