

Content-Aware Diffusion Stylization for Data Augmentation in Low-Resource Visual Recognition

Tymon Turon

Department of Electrical Engineering and Computer Science, University of Missouri,
Columbia, MO, USA.
turontymon@missouri.edu

Arin Chaudry

School of Electrical Engineering and Computer Science, Oregon State University, Corvallis,
OR, USA.
achaudry@oregonstate.edu

Abstract

Content-aware diffusion stylization has emerged as a promising mechanism for augmenting visual recognition datasets when labeled examples are scarce, domain coverage is uneven, or collection costs are prohibitive. This paper examines the architecture, governance, and systems implications of using diffusion-based stylization as a data augmentation strategy rather than as a purely aesthetic image transformation. The central argument is that content-aware stylization must be understood as a socio-technical pipeline in which semantic preservation, style diversification, compute budgeting, evaluation validity, and policy accountability are jointly optimized. The paper reviews conceptual foundations spanning data augmentation, style transfer, conditional generative modeling, and denoising diffusion. It then develops a system-level account of conditioning pathways, latent representations, adaptation modules, and deployment topologies suitable for low-resource recognition. Particular attention is paid to structural trade-offs between realism and diversity, semantic fidelity and stylistic novelty, and augmentation yield and computational sustainability. The discussion extends to robustness and fairness, arguing that stylized augmentation can improve resilience to nuisance variation while also amplifying spurious correlations or narrowing the effective visual distribution if governance is weak. Evaluation frameworks must therefore combine recognition accuracy with subgroup diagnostics, calibration, corruption robustness, and provenance documentation. The paper concludes by outlining institutional practices for responsible adoption, including data cards, model cards, energy accounting, audit trails, and domain-specific review. Overall, content-aware diffusion stylization is best governed as an infrastructural capability whose value depends less on generative novelty alone than on disciplined integration into recognition pipelines and accountable data ecosystems.

Keywords

data augmentation, diffusion models, content-aware stylization, low-resource visual recognition, robustness, fairness, machine learning systems, governance.

1. Introduction

Low-resource visual recognition remains a persistent challenge in domains where labeled data are expensive, sensitive, unevenly distributed, or difficult to collect at scale. Agricultural monitoring, rare medical imaging, industrial defect inspection, wildlife conservation, and disaster response all exhibit visual recognition problems in which the long tail of conditions

matters more than the average case. In such settings, conventional supervised learning tends to overfit to the limited visual distribution represented in the training set, producing models that perform well on familiar viewpoints and poorly on rare but consequential variations [1]. Data augmentation has therefore become a foundational technique for expanding the effective training distribution without collecting new labels, but classical augmentation methods often struggle to produce meaningful semantic and stylistic diversity beyond geometric transforms, color jitter, occlusion, and noise injection [2]. More advanced learned augmentation policies can improve recognition performance, yet they remain constrained by the transformations available in their search spaces and by the labels and domains from which those policies are learned [3].

Diffusion models have changed the generative landscape by offering high-fidelity image synthesis with stable training dynamics and strong controllability relative to earlier generative adversarial approaches [5], [6]. Latent diffusion models further reduce computational cost by operating in compressed representations while preserving semantic detail and visual realism [7]. These developments suggest that diffusion-based augmentation could move beyond simple perturbation and toward content-aware synthesis, where the semantic identity of an object or scene is preserved while style, texture, lighting, background, and appearance factors are systematically varied. The promise is substantial for low-resource recognition because the bottleneck is often not the number of images alone but the narrowness of visual conditions represented in the available data. A classifier trained only on daylight, clear-weather, single-camera, or single-institution images may fail when deployed under different sensors, geographies, or operational conditions.

However, treating diffusion stylization as a straightforward preprocessing step obscures the system-level complexities. Generative augmentation introduces new dependencies on model weights, sampling schedules, conditioning mechanisms, safety filters, storage systems, and evaluation protocols. It also changes the statistical assumptions of the training pipeline. Synthetic or stylized images are not neutral replacements for real examples; they carry the biases of the generative model, the style references, and the selection criteria used to decide which images to retain. If content preservation is weak, the augmentation may alter class-defining features. If style diversity is too narrow, it may produce only superficial variation. If generation is too permissive, it may introduce artifacts that classifiers learn to exploit. These trade-offs are architectural, operational, and institutional rather than merely algorithmic. This paper argues that content-aware diffusion stylization should be designed as a governed augmentation infrastructure, with explicit controls for semantic fidelity, style coverage, compute intensity, and downstream fairness.

The paper contributes a system-oriented synthesis of the topic. It first reviews the conceptual foundations of data augmentation, style transfer, and diffusion conditioning. It then develops an architectural account of content-aware stylization pipelines and their integration into recognition workflows. Subsequent sections examine robustness and fairness, deployment and sustainability, governance and policy, and cross-domain case illustrations. The final sections discuss limitations and forward-looking directions. The central claim is that the value of diffusion stylization for low-resource visual recognition depends not on generative realism alone but on whether the augmentation pipeline preserves label semantics, diversifies nuisance factors, supports reproducible evaluation, and remains accountable to affected communities and institutions.

2. Conceptual Foundations of Content-Aware Diffusion Stylization

Content-aware diffusion stylization sits at the intersection of three research traditions: data augmentation, neural style transfer, and conditional generative modeling. Neural style transfer demonstrated that convolutional representations can separate content structure from artistic style, enabling images to be recomposed with new textures and color palettes while retaining recognizable scene layout [9]. Subsequent work improved flexibility and speed through adaptive normalization and arbitrary style transfer, allowing a single model to apply diverse styles without retraining for each style pair [10]. Conditional generative adversarial networks extended this line by incorporating semantic maps, labels, and other structural conditions into synthesis, making it possible to control object layout and appearance at higher resolution [11]. Style-based generative architectures further showed that disentangled latent spaces can support controllable variation in identity, pose, texture, and fine detail [12]. These developments established the conceptual vocabulary for thinking about content and style as partially separable factors, even though the separation is never perfect in practice.

Diffusion models recast generative learning as an iterative denoising process, which provides a different balance of fidelity, diversity, and controllability. Denoising diffusion probabilistic models learn to reverse a gradual noise process, producing samples through many small refinement steps [5]. Their improved training stability and sample quality helped them surpass earlier generative paradigms on several image synthesis benchmarks [6]. Latent diffusion models then moved the process into a compressed latent space, making high-resolution synthesis more practical for real-world deployment [7]. For augmentation, this is significant because the generation process can be conditioned on content cues, textual descriptions, segmentation masks, depth maps, or reference style statistics. Content-aware stylization specifically seeks to preserve the label-relevant structure of an image while changing surface appearance and contextual statistics. The goal is not to replace the original image but to produce plausible counterfactual views that improve recognition under distribution shift.

A recent direction has examined scale-separated conditioning for style-encoder-free diffusion stylization [8]. In system terms, the relevance of such approaches is that they reduce dependence on a dedicated style encoder and instead condition the diffusion process at multiple scales, allowing coarse structure and fine texture to be influenced differently. This matters for low-resource recognition because the same object may appear under varied lighting, weather, sensor characteristics, or institutional imaging protocols. A stylization pipeline that can alter fine texture without disturbing object boundaries is more useful for augmentation than one that changes both indiscriminately. At the same time, scale-separated conditioning introduces complexity in scheduling, model compatibility, and evaluation. If coarse conditioning is too strong, the class label may be preserved while the scene becomes implausible for the target domain. If fine conditioning is too weak, stylistic variation may be insufficient to improve robustness. The conceptual foundation therefore points toward a multi-objective view in which semantic preservation, style novelty, visual plausibility, and computational cost are jointly managed.

3. System Architecture and Conditioning Pathways

A content-aware diffusion stylization system for data augmentation can be understood as a layered architecture comprising data ingestion, content representation, style conditioning, diffusion generation, filtering, labeling, storage, and training integration. The ingestion layer determines which source images are eligible for augmentation and which metadata are available. In low-resource settings, metadata may include class labels, segmentation masks, bounding boxes, acquisition device, location, time, and domain identifiers. The content

representation layer extracts or preserves semantic structure, often through latent encodings, edge maps, depth cues, or segmentation conditions. The style conditioning layer determines how appearance factors are sampled or transferred. The diffusion generation layer produces candidate images. The filtering layer assesses whether generated images are usable for training. The storage layer manages provenance and versioning. The training integration layer mixes real and stylized examples according to a controlled schedule. Each layer introduces design choices that affect downstream recognition.

Conditioning pathways are the core architectural mechanism. Text conditioning offers flexibility but can be ambiguous and difficult to control for fine-grained visual recognition. Image-based conditioning can preserve object identity but may leak unwanted content from the reference. Segmentation and depth conditioning can maintain spatial layout but require additional annotations. Latent conditioning can be efficient but may lose fine detail. Diffusion implicit models and improved denoising schedules can reduce sampling cost while maintaining quality, which is important when augmentation must be repeated across many classes and domains [13], [14]. Subject-driven adaptation methods such as DreamBooth demonstrate how a generative model can be personalized to specific visual concepts, but they also raise concerns about overfitting to a small reference set and about the computational overhead of fine-tuning [15]. Parameter-efficient adaptation methods such as low-rank adaptation provide a more scalable way to specialize a diffusion model to a domain or style family without updating all weights [16]. In practice, a robust system may combine a general diffusion backbone with lightweight domain adapters and a retrieval mechanism that selects style references appropriate to the target recognition task.

The architecture must also specify where augmentation occurs in the recognition lifecycle. Pre-training augmentation uses stylized images to shape general representations before fine-tuning on real data. Fine-tuning augmentation mixes stylized and real examples during task adaptation. Test-time augmentation applies stylization to input images and aggregates predictions, which can improve robustness but increases inference cost and may introduce instability if the stylization distribution differs from training. Each placement creates different trade-offs. Pre-training augmentation can improve transfer but may waste computation if the synthetic distribution is poorly aligned. Fine-tuning augmentation can directly target minority conditions but risks overfitting to generative artifacts. Test-time augmentation can help with nuisance variation but complicates latency and explainability. A well-designed system therefore treats stylization as a configurable stage rather than a universal replacement for real data.

4. Data Augmentation Pipelines for Low-Resource Visual Recognition

In low-resource recognition, the augmentation pipeline must address both scarcity and imbalance. Scarcity means that some classes have too few examples to learn robust representations. Imbalance means that some visual conditions, geographic regions, devices, or demographic groups are underrepresented relative to others. Content-aware stylization can address both by generating additional views that preserve class semantics while varying nuisance factors. However, the pipeline must be carefully orchestrated. The source dataset should be audited for label quality, domain coverage, and known biases. Style references should be selected to cover meaningful variation rather than arbitrary aesthetic diversity. Generated images should be filtered for semantic consistency, realism, and label preservation. The final training mixture should be tuned to avoid overwhelming real examples with synthetic ones. These steps resemble established augmentation practices, but generative

stylization expands the space of possible transformations and therefore increases the importance of validation [2], [3].

Classical augmentation methods such as geometric transforms, color changes, and cutout remain valuable because they are cheap, interpretable, and unlikely to alter labels. Learned policies such as AutoAugment can discover combinations that improve accuracy on benchmark datasets [4]. Mixup and related interpolation methods encourage smoother decision boundaries and can improve calibration [17]. Content-aware stylization can complement these methods by targeting higher-level appearance variation that classical transforms cannot produce. For example, in medical imaging, stylization might simulate different scanner vendors or staining protocols while preserving anatomical structure. In remote sensing, it might vary atmospheric conditions, seasonal texture, and sensor characteristics while preserving land-cover class. In industrial inspection, it might render defects under different lighting and surface finishes while preserving defect type. The pipeline must therefore define which factors are nuisance variables that should vary and which factors are label-defining and must remain stable.

A critical design question is how to label stylized images. If content preservation is strong, the original label can be inherited with high confidence. If stylization changes context in ways that affect class identity, inherited labels become unreliable. Weakly supervised or self-supervised filtering can help estimate semantic consistency, but it introduces another model whose errors may propagate. Human review may be necessary for high-stakes domains, but it does not scale to large augmentation budgets. A practical compromise is tiered augmentation, in which low-risk stylizations inherit labels automatically, medium-risk stylizations require model-based verification, and high-risk stylizations require human audit or are excluded. This tiering reflects a broader governance principle: the more a generative transformation can alter label-relevant content, the stronger the validation burden. The pipeline should also preserve provenance, recording which source image, style reference, model version, and filtering decision produced each augmented sample.

5. Robustness, Fairness, and Evaluation

Robustness is often the primary motivation for content-aware stylization. Recognition models trained on narrow data distributions can fail under common corruptions such as blur, noise, weather, and compression [18]. Stylized augmentation can simulate some of these variations and thereby improve resilience. However, robustness gains are not automatic. If the stylization distribution is too narrow, the model may overfit to the specific styles generated during training. If it is too broad, the model may learn invariances that erase useful discriminative cues. If generated images contain artifacts, the model may learn to recognize the generator rather than the underlying object. Robustness evaluation must therefore test against real distribution shifts, not only against the same stylization pipeline used for training. Otherwise, reported improvements may reflect a closed-loop effect in which training and testing share generative biases.

Fairness introduces additional concerns. Generative models trained on biased data can reproduce or amplify stereotypes, underrepresent minority groups, or generate implausible combinations for rare subgroups [19]. In visual recognition, stylization may change skin tone, clothing, lighting, or context in ways that correlate with demographic attributes. If such changes are not controlled, the augmented dataset may improve average accuracy while worsening subgroup performance. Evaluation should therefore include disaggregated metrics across classes, domains, devices, geographies, and demographic groups where legally and

ethically appropriate. Calibration, false positive rates, false negative rates, and uncertainty estimates should be examined alongside top-line accuracy. The goal is not only to improve recognition under variation but to ensure that the benefits and harms of augmentation are distributed fairly.

Evaluation frameworks must also account for the provenance and quality of generated data. Data cards and model cards provide structured documentation of dataset composition, intended use, limitations, and ethical considerations [22], [23]. In the context of diffusion stylization, documentation should record the base generative model, adaptation modules, style references, filtering thresholds, human review procedures, and known failure modes. If the augmentation pipeline is updated, downstream models should be re-evaluated because changes in stylization can alter learned representations. This is particularly important in regulated domains such as healthcare, transportation, and public safety, where model updates may require audit trails and impact assessments. A robust evaluation protocol should combine quantitative benchmarks with qualitative case analysis, because some failure modes are visible only when inspecting generated images and their effects on specific decisions.

6. Deployment, Infrastructure, and Sustainability

Deploying content-aware diffusion stylization as part of a recognition system requires significant infrastructure. Diffusion sampling is computationally expensive relative to classical augmentation, especially when generating high-resolution images or running many denoising steps [13], [14]. Latent diffusion reduces cost but still requires accelerator hardware, memory, and storage for generated datasets [7]. In low-resource settings, the organizations that most need augmentation may have the least capacity to run large generative models. This creates an equity problem: if high-quality stylization depends on expensive infrastructure, then well-funded institutions may gain a further advantage in visual recognition performance. Deployment architectures should therefore consider shared services, model distillation, caching, and parameter-efficient adaptation to reduce barriers [16]. Edge deployment may be possible for inference but is rarely practical for large-scale generation.

Storage and data management are also major concerns. A stylization pipeline can produce many variants per source image, leading to dataset inflation. If every generated image is stored, the storage footprint may exceed that of the original dataset by orders of magnitude. If images are generated on the fly, reproducibility becomes harder because the same seed, model version, and conditioning must be preserved. A hybrid approach may store compact conditioning metadata and regenerate images as needed, but this requires deterministic pipelines and stable model weights. Version control is essential because a change in the diffusion model or style adapter can alter the entire augmented dataset. Organizations should treat generative augmentation as a data product with release notes, checksums, and rollback procedures.

Sustainability is inseparable from deployment. Training and running large generative models consume energy and produce carbon emissions, and these costs must be weighed against recognition benefits [20]. In some cases, a smaller augmentation budget with targeted stylization may yield most of the robustness improvement at a fraction of the energy cost. In other cases, the environmental cost may be justified by safety or access benefits, but the decision should be explicit rather than hidden in technical defaults. Energy accounting should include model training, fine-tuning, sampling, storage, and repeated evaluation. The system should support early stopping, batch generation, and reuse of generated datasets across related tasks. Sustainable design also means avoiding unnecessary style diversity that does not

improve downstream recognition. Augmentation is not valuable merely because it is abundant; it is valuable when it addresses a measured gap in the training distribution.

7. Governance, Policy, and Institutional Implications

Governance is necessary because content-aware stylization affects data provenance, representation, and accountability. An institution that uses synthetic or stylized images to train a recognition model should be able to explain where those images came from, what transformations were applied, and how they were validated. This is especially important when the model affects access to services, safety, or resource allocation. Data protection laws, sector-specific regulations, and institutional review boards may impose constraints on the use of generative models trained on personal or sensitive data. Even when generated images do not directly correspond to real individuals, they may encode statistical patterns derived from protected groups. Governance should therefore address consent, licensing, data minimization, and purpose limitation.

Policy implications extend to procurement and audit. Organizations that purchase recognition systems should ask vendors whether generative augmentation was used, how it was evaluated, and what safeguards prevent harmful distributional shifts. Auditors should have access to sufficient documentation to assess claims about robustness and fairness. Model cards and datasheets provide a starting point, but they must be tailored to generative augmentation, which introduces additional layers of provenance and uncertainty [22], [23]. Certification regimes may eventually require standardized reporting on synthetic data proportion, style reference sources, filtering methods, and subgroup performance. Such requirements should be proportionate to risk and should not become mere compliance exercises that fail to address substantive harms.

Institutional culture also matters. Researchers and engineers may treat generative augmentation as a technical optimization problem, but its consequences are organizational and social. Teams should include domain experts, ethicists, legal reviewers, and affected community representatives where appropriate. Decision-making should be documented, and disagreements should be surfaced rather than resolved by default settings. In low-resource contexts, governance should be participatory, recognizing that communities may have legitimate concerns about how their visual data are used and represented. The goal is not to slow innovation but to ensure that augmentation improves recognition in ways that are legitimate, equitable, and sustainable.

8. Case Illustrations and Cross-Domain Comparisons

Different domains reveal distinct trade-offs in content-aware diffusion stylization. In medical imaging, content preservation is paramount because anatomical structures define diagnosis. Stylization may simulate scanner variation, staining differences, or acquisition protocols, but it must not alter lesions or tissue boundaries. The cost of false augmentation is high because a mislabeled synthetic image can teach a model incorrect clinical associations. Governance must therefore be stringent, with expert review, conservative filtering, and extensive external validation. In contrast, remote sensing may tolerate greater stylistic variation because land-cover classes are often defined by broad spatial and spectral patterns. Stylization can simulate seasonal changes, atmospheric conditions, and sensor differences, but it must preserve geographic and temporal plausibility. The evaluation challenge is to test on real geographic and temporal holdouts rather than on synthetic variants alone.

Industrial inspection occupies a middle ground. Defect detection requires preserving subtle visual cues that distinguish acceptable from unacceptable products. Stylization can vary lighting, surface finish, and camera angle, but excessive texture generation may create false defects or obscure real ones. The deployment context is often high-speed and cost-sensitive, so the augmentation budget must be justified by reduced inspection errors or lower labeling costs. Wildlife conservation and ecological monitoring involve long-tailed species distributions and variable camera traps. Stylization can simulate weather, vegetation, and illumination changes, but it must not create biologically implausible appearances that mislead classifiers. In all these domains, the common lesson is that content-aware stylization is most useful when it targets known nuisance factors and is evaluated against real-world shifts.

Cross-domain comparison also highlights the importance of style reference selection. A style reference set drawn from a different domain may introduce artifacts or semantic shifts that are irrelevant or harmful. For example, artistic styles may be useful for general object recognition but inappropriate for medical diagnosis. Style references should be curated with domain expertise and documented as part of the data pipeline. The same diffusion backbone may be adapted differently across domains using parameter-efficient modules [16]. This modularity supports reuse while allowing domain-specific control. However, modularity also creates complexity: each adapter must be versioned, evaluated, and governed. The system-level lesson is that flexibility without accountability can produce brittle datasets and unclear responsibility.

9. Limitations and Future Directions

Content-aware diffusion stylization is not a substitute for representative data collection. It can interpolate and extrapolate within learned distributions, but it cannot reliably create genuinely novel causal factors that were absent from training [1]. If the base generative model has not seen a particular sensor, geography, or demographic group, stylization may produce implausible outputs or fail to cover the missing distribution. Low-resource settings may therefore benefit from combining stylization with active learning, targeted annotation, and community-based data collection. The most sustainable approach is often a hybrid: use generative augmentation to fill measured gaps while investing in real data where possible.

Evaluation remains a major limitation. There is no single metric that captures semantic fidelity, style diversity, realism, and downstream utility. Accuracy improvements on a benchmark may not generalize to deployment. Robustness tests may not cover the full range of real-world shifts. Fairness audits may be constrained by missing demographic metadata or legal restrictions. Future work should develop evaluation suites that are domain-specific, adversarially informed, and participatory. These suites should include stress tests for rare conditions, artifact detection, label preservation, and subgroup performance. They should also measure computational cost and environmental impact, because augmentation strategies that are effective but unsustainable may be inappropriate for widespread use [20].

Another direction is greater integration between generative augmentation and recognition architecture. Instead of treating stylization as a preprocessing step, future systems might use uncertainty-aware conditioning to generate examples that specifically challenge the current model. This could resemble curriculum learning, where the augmentation pipeline targets classes or conditions with high error rates. Such systems would need careful safeguards to avoid feedback loops in which the model reinforces its own biases [19]. They would also require robust provenance and human oversight. Finally, policy frameworks must evolve to address synthetic data in high-stakes recognition. Standards for documentation, auditing, and

liability will shape how diffusion stylization is adopted. The technical community can contribute by building systems that make these governance requirements feasible rather than treating them as external burdens.

10. Conclusion

Content-aware diffusion stylization offers a powerful approach to data augmentation for low-resource visual recognition, but its value depends on disciplined system design and governance. Diffusion models provide high-fidelity synthesis and flexible conditioning, while style transfer and conditional generation offer conceptual tools for separating content from appearance. When integrated carefully, stylization can diversify nuisance factors, improve robustness, and reduce the label burden in domains where real data are scarce or unevenly distributed. However, the same flexibility that makes diffusion stylization useful also creates risks of semantic drift, artifact exploitation, bias amplification, and unsustainable compute use. These risks cannot be resolved by model architecture alone. They require evaluation frameworks that measure subgroup performance, robustness, calibration, and provenance; deployment architectures that manage cost, storage, and reproducibility; and governance mechanisms that ensure accountability to affected communities and institutions.

The most promising path is not to treat generative augmentation as a replacement for real data but as one component of a broader data strategy. Organizations should audit their datasets, identify specific gaps, and use stylization where it can address those gaps without compromising label integrity. They should document style references, filtering decisions, and model versions. They should evaluate downstream models on real distribution shifts and disaggregated metrics. They should account for energy and infrastructure costs. Above all, they should recognize that content-aware diffusion stylization is a socio-technical capability whose legitimacy depends on transparency, participation, and careful integration into recognition pipelines. If these conditions are met, diffusion stylization can meaningfully expand the reach of visual recognition in low-resource settings while maintaining the trust required for responsible deployment.

References

1. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.
2. Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1), 60. <https://doi.org/10.1186/s40537-019-0197-0>
3. Perez, L., & Wang, J. (2017). The effectiveness of data augmentation in image classification using deep learning. arXiv preprint arXiv:1712.04621.
4. Cubuk, E. D., Zoph, B., Mane, D., Vasudevan, V., & Le, Q. V. (2019). AutoAugment: Learning augmentation strategies from data. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 113-123.
5. Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33, 6840-6851.
6. Dhariwal, P., & Nichol, A. (2021). Diffusion models beat GANs on image synthesis. *Advances in Neural Information Processing Systems*, 34, 8780-8794.
7. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10684-10695.

8. Zhang, Jingtao, et al. "Scale-separated conditioning for style-encoder-free diffusion stylization." arXiv preprint arXiv:2608.19719 (2026).
9. Gatys, L. A., Ecker, A. S., & Bethge, M. (2016). Image style transfer using convolutional neural networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2414-2423.
10. Huang, X., & Belongie, S. (2017). Arbitrary style transfer in real-time with adaptive instance normalization. *Proceedings of the IEEE International Conference on Computer Vision*, 1501-1510.
11. Wang, T. C., Liu, M. Y., Zhu, J. Y., Tao, A., Kautz, J., & Catanzaro, B. (2018). High-resolution image synthesis and semantic manipulation with conditional GANs. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 8798-8807.
12. Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4401-4410.
13. Song, J., Meng, C., & Ermon, S. (2021). Denoising diffusion implicit models. *International Conference on Learning Representations*.
14. Nichol, A. Q., & Dhariwal, P. (2021). Improved denoising diffusion probabilistic models. *International Conference on Machine Learning*, 8162-8171.
15. Ruiz, N., Li, Y., Jampani, V., Pritch, Y., Rubinstein, M., & Aberman, K. (2023). DreamBooth: Fine tuning text-to-image diffusion models for subject-driven generation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 22500-22510.
16. Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., & Chen, W. (2022). LoRA: Low-rank adaptation of large language models. *International Conference on Learning Representations*.
17. Zhang, H., Cisse, M., Dauphin, Y. N., & Lopez-Paz, D. (2018). mixup: Beyond empirical risk minimization. *International Conference on Learning Representations*.
18. Hendrycks, D., & Dietterich, T. (2019). Benchmarking neural network robustness to common corruptions and perturbations. *International Conference on Learning Representations*.
19. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys*, 54(6), 1-35. <https://doi.org/10.1145/3457607>
20. Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 3645-3650.