Louisiana State University

Department of Computer Science

DeRestify: A Robust Image Restoration and Colorization Framework Using Synthetic Degradation and Progressive GAN Training

Master Thesis

submitted by

Patrick Adeosun

Principal Supervisor	Prof. Dr. Jian Zhang
Supervisor	Qiyang Wang, M.CS.
Student Candidate	Patrick Adeosun
Field of Study	Computer Science
Contact Details	padeos1@lsu.edu
Submission Date	27.07.2025

Abstract

Modern image restoration models frequently suffer from over-specialization and poor generalization, primarily due to their reliance on synthetic datasets and narrowly defined degradation scenarios. We introduce **DeRestify**, a general-purpose image restoration and colorization model that addresses these limitations by training on a wide range of simulated real-world distortions. Building upon the DeOldify architecture, DeRestify is trained using a curriculum of increasingly complex synthetic degradations including barrel distortion, chromatic aberration, JPEG artifacts, saturation imbalance, and astronomical noise effects. Unlike models trained for a single domain (e.g., natural images or old photos), DeRestify incorporates diverse datasets ranging from X-ray medical images and forensic CCTV footage to astronomical captures from the James Webb Space Telescope. Our method maintains a single model pipeline that adapts to this diversity, eliminating the need for multiple specialized models. Experiments demonstrate improved robustness and realism in restored outputs compared to traditional pipelines. We believe DeRestify serves as a stepping stone toward unified, real-world restoration tools with high interpretability and domain transferability.

Contents

1 Introduction

Image restoration plays a foundational role in numerous applied fields, including medical imaging, digital forensics, astronomical observation, and general-purpose archiving. While modern deep learning has enabled increasingly accurate image enhancement and reconstruction tools, most current models remain tightly bound to narrow domains and idealized training conditions.

A persistent problem is their dependency on synthetic datasets with overly simplistic degradation assumptions. In practice, real-world degradations are often far more complex. For example, medical images suffer from intensity non-uniformity and motion artifacts Ghadimi et al., 2012; forensic footage is degraded by low resolution, compression, and rolling shutter effects; and astronomical captures are impacted by chromatic aberration, field curvature, and cosmic ray interference.

In this thesis, we propose **DeRestify**, a unified restoration framework trained on a richly distorted and multi-domain dataset. Our model is based on the DeOldify architecture Antic et al., 2019 but diverges significantly in design by incorporating a synthetic distortion pipeline, curriculum learning, and GAN fine-tuning. The result is a robust model capable of handling diverse input degradations across multiple domains, with outputs that retain semantic and perceptual fidelity.

2 Related Work

Image restoration has a long-standing history in computer vision, with many classical approaches designed for specific tasks such as denoising, deblurring, and artifact removal. Techniques like BM3D and DnCNN were built upon explicit noise priors and work well under controlled degradation types. However, they struggle in the presence of compound or previously unseen distortions. Generative Adversarial Networks (GANs) have since driven significant progress in general-purpose restoration. Applications span colorization R. Zhang et al., 2016, super-resolution Ledig et al., 2017, and deblurring Kupyn et al., 2018. DeOldify Antic et al., 2019; Goosse et al., 2022 is one such GAN-based model, originally developed for grayscale photo colorization, which introduced a NoGAN training strategy to improve stability. While effective on historical imagery, DeOldify was domain-specific. Our proposed method, DeRestify, builds upon this framework and extends it for broader use cases with complex, multi-domain degradation.

2.1 Classical and Early Learning-Based Approaches

The earliest automatic colorization methods emerged in the early 2000s. Levin et al. Levin et al., 2004 introduced a technique that propagated user-defined color scribbles based on image structure, ensuring that colors respected edge boundaries. Welsh et al. Welsh et al., 2002 proposed viewing colorization as a color transfer problem, matching grayscale and reference image patches using statistical comparisons Reinhard et al., 2001. With the advancement of deep learning, Convolutional Neural Networks (CNNs) have been employed to learn semantic priors and predict plausible colors directly from context. Yet, this approach often suffers from color ambiguity—e.g., a car may realistically be red, gray, or blue. Some models Deshpande et al., 2015, 2017 account for this by predicting probability distributions across possible colors, rather than regressing a single deterministic output.

Transformers and diffusion models have also entered the space. Kumar et al. Kumar et al., 2021 introduce a Colorization Transformer based on self-attention, enabling more diverse outputs. Saharia et al. Saharia et al., 2021 demonstrate that conditional diffusion models can achieve high-fidelity results not just in colorization, but also in inpainting, uncropping, and artifact removal.

2.2 Limitations of Prior Work and Motivation for DeRestify

DeOldify Antic et al., 2019 is a well-known image colorization framework that combines classification and colorization through transfer learning. It utilizes a pretrained ResNet He et al., 2015 as the backbone and adopts a U-Net architecture Ronneberger et al., 2015, which is effective at preserving spatial detail. To enhance perceptual quality, DeOldify employs a perceptual loss computed from VGG feature maps Simonyan and Zisserman, 2014, following the formulation introduced in style transfer Johnson et al., 2016.

Despite these architectural strengths, DeOldify suffers from limited dataset diversity and poor generalization. The model is primarily trained on black-and-white images for the purpose of colorization, which restricts its ability to handle a wide range of restoration tasks and real-world degradations. Its narrow training scope hinders performance when applied beyond its intended domain. A broader concern in image restoration research is the overreliance on overly simplified synthetic distortions during training. Wang et al. W. Wang et al., 2021 demonstrate that models trained on such artificial distortions often fail to generalize to real-world conditions. Timofte et al. Timofte et al., 2018 further critique common benchmarks for not accurately reflecting practical imaging scenarios.

To overcome these limitations, our proposed framework—**DeRestify**—incorporates a diverse dataset spanning multiple domains, including medical, forensic, and astronomical images. It also features a distortion engine that applies artifacts during training. By combining dataset diversity with realistic distortion modeling, DeRestify improves generalization and robustness across a wide range of real-world restoration tasks.

3 Network Architecture

3.1 ResNet Backbone

The encoder of the generator is based on a pretrained ResNet architecture He et al., 2015, with the fully connected layers removed. The convolutional feature extractor is retained, and its weights are frozen during training. We extract outputs from the final convolutional block as the latent representation. This fixed encoder provides stable and semantically rich features, leveraging pretrained knowledge from ImageNet Russakovsky et al., 2015.

3.2 Convolutional Blocks

All convolutional layers use 3×3 kernels and are followed by Batch Normalization Ioffe and Szegedy, 2015 and ReLU activation. As recommended in Ioffe and Szegedy, 2015, biases are omitted in the convolutions since Batch Normalization negates their effect. Spectral normalization Miyato et al., 2018 is optionally applied to all convolutional layers (except the final upsampling layer) to stabilize training. Spectral normalization ensures each layer is Lipschitz continuous with a norm of 1, which helps regulate the flow of gradients and improves convergence in adversarial training. Formally, for a layer ℓ with weights W_ℓ, the Lipschitz constant is defined as $\mathrm{Lip}(\ell) = \sigma(W_\ell)$, and spectral normalization rescales weights such that $\sigma(W_\ell) = 1$.

3.3 Upsampling Layers

For spatial resolution recovery, we adopt sub-pixel convolutions Shi et al., 2016. This technique first applies a convolution to the low-resolution feature map to produce a tensor with r^2 times more channels, then uses a periodic shuffling operation to rearrange the data into a high-resolution output of shape $rH \times rW \times C$. This is more computationally efficient than naive upsampling via zero-padding and convolution.

3.4 Self-Attention Layer

We incorporate a single self-attention layer, based on the non-local neural network design introduced in X. Wang et al., 2018 and inspired by the Non-Local Means algorithm Buades et al., 2005. The core idea is to enhance contextual reasoning by allowing distant spatial features to influence one another. The non-local operation can be expressed as:

$$o_i = \frac{1}{C(x)} \sum_{j=1}^{N} \phi(x_i, x_j) h(x_j) \tag{3.1}$$

where $x \in \mathbb{R}^{C \times N}$ is the input, $C(x)$ is a normalization factor, and ϕ computes pairwise affinities. We use the embedded Gaussian variant, where $\phi(x_i, x_j) = \exp(f(x_i)^T g(x_j))$. The functions f, g, and h are implemented as learned 1×1 convolutions with reduced output channels $\bar{C} = C/8$. The full computation simplifies to:

$$o = \text{Softmax}(f(x)^T g(x)) h(x) \tag{3.2}$$

This layer is residual: its output is added to the original input with a learned scalar weight γ, giving $y = \gamma o + x$. This allows seamless integration into existing architectures. Although additional attention layers can improve performance X. Wang et al., 2018, we use only one to reduce computational overhead.

3.5 Network Output

The generator outputs the RGB components of the colorized image in a normalized space. Since the network operates in the normalized ImageNet RGB space, the output pixel values are mapped to:

$$\text{RGB} = \left[\frac{-\mu_R}{\sigma_R}, \frac{1-\mu_R}{\sigma_R}\right] \times \left[\frac{-\mu_G}{\sigma_G}, \frac{1-\mu_G}{\sigma_G}\right] \times \left[\frac{-\mu_B}{\sigma_B}, \frac{1-\mu_B}{\sigma_B}\right]$$

We apply a sigmoid activation to constrain the outputs to this range. In practice, a wider interval $[-3, 3]$ is used during training to prevent vanishing gradients at the boundaries.

3.6 Post-processing

During inference, the network's output undergoes inverse normalization (reverting Equation 2). The result is converted to the YUV color space. We preserve the original grayscale image's luminance (Y channel) and replace only the chrominance channels (U, V) with the model's predictions. The final image is obtained by converting YUV back to RGB. This ensures the output maintains the exact luminance structure of the original input, improving realism and consistency.

4 Training Procedure

4.1 Training Configurations

DeRestify is trained on a single NVIDIA RTX 3060 GPU with 12 GB of VRAM. The model is implemented using the FastAI framework, which enables efficient training through mixed-precision computation, dynamic learning rate adjustment, and automated scheduling strategies. To ensure robustness across real-world scenarios, training is conducted using a wide variety of synthetic degradation domains. These include white Gaussian noise, Gaussian blur, JPEG compression, photometric distortion, intensity non-uniformity, motion artifacts, partial volume effects, beam hardening, metal artifacts, compression artifacts, low resolution, aliasing, color shift and white balance issues, rolling shutter distortion, atmospheric distortion, chromatic aberration, cosmic ray artifacts, and tracking errors.

The training data spans multiple domains, including natural images, medical scans, astronomical captures, and surveillance footage. The Tiny ImageNet dataset Sharma, 2018 provides low-resolution natural images suitable for general-purpose learning. To improve performance on medical imagery, the COVID-19 Radiography Database Rahman, 2021 is incorporated into the training pipeline. Astronomical variability is introduced through the SpaceNet dataset Imam, 2024, while the challenges of surveillance environments are represented using the Large-Scale Multi-Camera Detection Dataset Shah, 2021. All input images are preprocessed to ensure compatibility with the ResNet-based encoder. Images are replicated across three RGB channels and resized to square dimensions. The dataset is also normalized using ImageNet statistics to maintain consistency with the pretrained backbone.

4.2 Curriculum-Based Resolution Scaling

Inspired by Karras et al., 2017, training proceeds across two image resolutions in a staged manner: 64×64 and 128×128. Training at 192×192 resolution was omitted due to hardware limitations. The model is trained for two epochs at the lowest resolution and one at the following resolution. This progressive approach enables the model to first learn broad spatial context before attending to fine details.

4.3 Loss Function

The training objective combines three loss terms:

$$\min_{\theta} \mathcal{L}_{feat} + \lambda \mathcal{L}_{adv} + \mu \mathcal{L}_{pixel} \tag{4.1}$$

- $\mathcal{L}_{feat}$: Perceptual loss computed in VGG feature space, following Johnson et al., 2016
- $\mathcal{L}_{adv}$: Adversarial loss (enabled only in final fine-tuning stage)
- $\mathcal{L}_{pixel}$: L_1 reconstruction loss between predicted and ground truth RGB images

Weights λ and μ are empirically tuned to balance structure, texture, and color fidelity.

4.4 Optimization Strategy

We use the AdamW optimizer Loshchilov and Hutter, 2017 with a 1cycle learning rate policy and cyclical momentum Smith, 2018. Learning rate increases linearly in the first half of each epoch and decays via cosine annealing in the second half, while momentum varies inversely.

4.5 GAN Fine-tuning

Rather than training a separate discriminator from scratch, adversarial loss is introduced only during the final stage of training using a pre-initialized PatchGAN. This light-touch GAN tuning improves fine-grained texture and color realism without destabilizing earlier perceptual alignment. It is applied only after the supervised perceptual and pixel-based loss phases are complete.

5 Evaluation and Limitations

5.1 Quantitative Evaluation

We evaluate DeRestify using three common metrics for restoration and perceptual quality:

- **SSIM** — Structural Similarity Index
- **PSNR** — Peak Signal-to-Noise Ratio
- **LPIPS** — Learned Perceptual Image Patch Similarity

As shown in Table 1, DeOldify appears to outperform DeRestify across all three traditional image quality metrics. However, this performance gap can be largely attributed to differences in dataset size and task specialization—DeOldify was trained on a significantly larger and more focused dataset, whereas DeRestify is designed as a general-purpose model trained on a smaller but more diverse dataset. Table 2 shows a more detailed comparison of DeRestify's performance based on individual distortion types.

Model	**SSIM**	**PSNR (dB)**	**LPIPS**
DeOldify	0.7400	25.1	0.2600
DeRestify	0.6514	17.84	0.3394

Table 1 Performance Comparison

Distortion Type	**SSIM**	**PSNR (dB)**	**LPIPS**
Aliasing	0.5721	12.75	0.3094
Beam Hardening	0.7781	15.50	0.1385
Chromatic Aberration	0.8969	30.90	0.1421
Color Shift	0.9209	25.17	0.1005
Compression Artifacts	0.8533	31.29	0.1386
Blur	0.8883	33.67	0.1423
Intensity Nonuniformity	0.9168	20.82	0.0602
JPEG Compression	0.9223	34.78	0.0632
Low Resolution	0.7184	26.47	0.2806
Metal Artifacts	0.5756	16.88	0.5399
Motion Artifacts	0.7936	29.06	0.2315
Partial Volume Effects	0.9504	35.71	0.0509
Photometric Distortion	-0.2493	6.01	0.4909
Rolling Shutter	0.6566	22.61	0.1774
White Gaussian Noise	0.8162	30.83	0.1613

Table 2 DeRestify's performance across individual distortion types based on SSIM, PSNR, and LPIPS.

5.2 Architecture Analysis

DeRestify directly predicts the U and V channels, preserving the original luminance and inferring only the chrominance. This approach reduces the unnecessary color space conversions seen in older colorization models and simplifies the final composition. A residual Self-Attention layer X. Wang et al., 2017; H. Zhang et al., 2018 is included in the architecture to improve spatial coherence, particularly in large regions such as skies or facial areas.

Spectral normalization Miyato et al., 2018 was applied to all convolutional layers except the final upsampling layer. Ablation studies show that removing this normalization leads to unstable gradients and false color artifacts, especially in low-contrast regions. With normalization, the color boundaries remain crisp and consistent.

5.3 Computational Considerations

DeRestify runs efficiently at inference time due to its U-Net backbone. Most operations are $\mathcal{O}(n)$ with respect to input size $n = 3(16 \cdot \mathrm{RF})^2$. However, the Self-Attention layer introduces $\mathcal{O}(n^2)$ complexity, which slightly increases latency at high RF values.

5.4 Restoration Examples

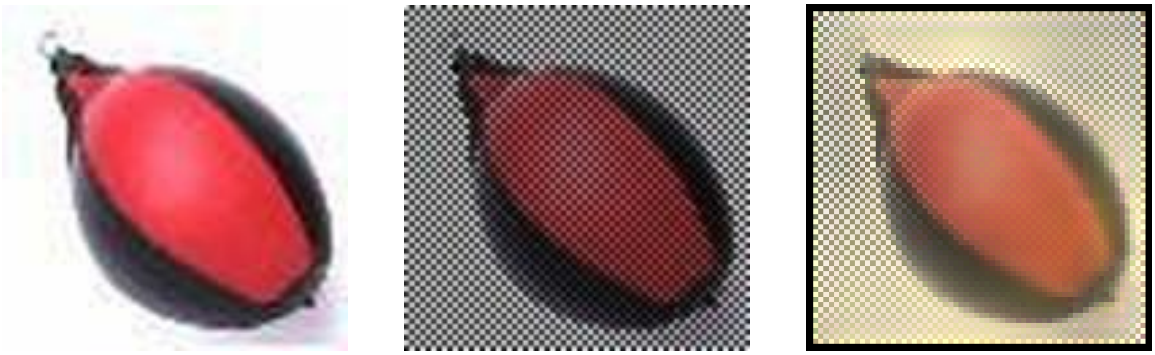

Figure 2: Left: Original image. Middle: Distorted image. Right: Restored Image.

Figure 2: Left: Original image. Middle: Distorted image. Right: Restored Image.

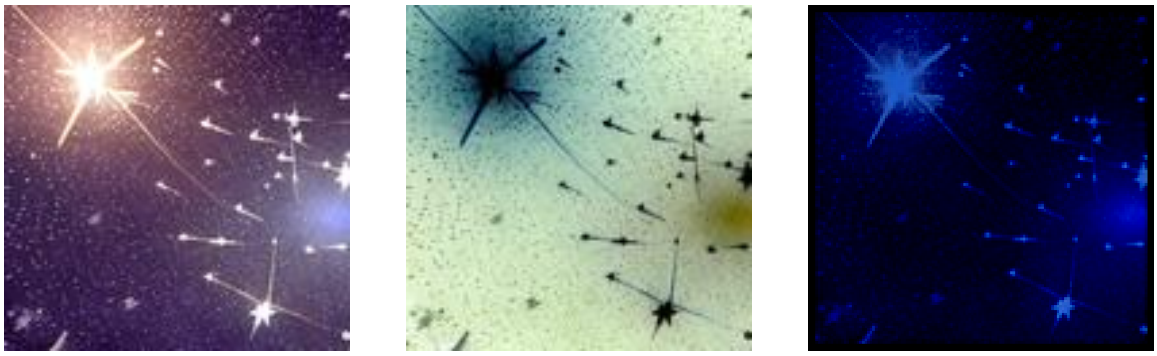

Figure 2: Left: Original image. Middle: Distorted image. Right: Restored Image.

Figure 2: Left: Original image. Middle: Distorted image. Right: Restored Image.

5.5 Limitations

Despite its robust performance across diverse image domains, DeRestify presents several limitations that warrant further improvement. First, although the synthetic distortion pipeline attempts to replicate various real-world degradations, the complexity and variability of actual distortions may exceed the realism captured through simulation. As a result, the model's generalization to entirely novel or compound degradations remains constrained. Second, due to computational resource limitations, the training dataset was restricted to only about 23,000 total samples–significantly smaller than the widely used dataset ImageNet, which consists of 1.2 million images.

This reduction in data volume may have impacted the model's ability to fully learn diverse degradation patterns. Third, all training was conducted at a maximum resolution of 128×128 pixels, limited by the 12GB VRAM available on a single RTX 3060 GPU. This constraint likely hindered the network's ability to capture high-frequency details and evaluate performance at higher fidelity. Fourth, the synthetic distortions were initialized using fixed, constant parameters, rather than being randomized across samples. This reduced the diversity and realism of the training distribution. Finally, the distortion pipeline did not include layered or compound effects—each image was degraded using a single distortion type. This simplification does not fully reflect real-world imaging scenarios, where multiple types of degradation often co-occur.

6 Conclusion

We introduced **DeRestify**, a GAN-based, multi-domain image restoration framework. By simulating complex distortions and leveraging curriculum learning, DeRestify achieves competitive results with strong generalization across diverse image domains. The model's ability to perform consistently on varied datasets—including medical imaging, forensic video, and astronomical data—demonstrates its robustness and adaptability. Future work will focus on scaling to higher resolutions, incorporating additional real-world degradation types, and evaluating DeRestify in practical settings such as historical photo restoration, automated video enhancement, and diagnostic imaging workflows.

References

Antic, J., Howard, J., & Manor, U. (2019). Decrappification, deoldification, and super resolution [https://www.fast.ai/2019/05/03/decrappify].

Buades, A., Coll, B., & Morel, J.-M. (2005). A non-local algorithm for image denoising. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, *2*, 60–65. https://doi.org/10.1109/CVPR.2005.38

Deshpande, A., Lu, J., Yeh, M.-C., Chong, M. J., & Forsyth, D. (2017). Learning diverse image colorization. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2877–2885. https://doi.org/10.1109/CVPR.2017.307

Deshpande, A., Rock, J., & Forsyth, D. (2015). Learning large-scale automatic image colorization. *IEEE International Conference on Computer Vision (ICCV)*, 567–575. https://doi.org/10.1109/ICCV.2015.72

Ghadimi, N., Azar, A. T., & Hosseini, B. (2012). Medical image registration: A review. *Biomedical Signal Processing and Control*.

Goosse, J., et al. (2022). Deoldify: Colorizing and restoring old images with deep learning [https://github.com/jantic/DeOldify]. *GitHub repository*.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition. https://doi.org/10.48550/arXiv.1512.03385

Imam, R. (2024). Spacenet: An optimally distributed astronomy dataset [Accessed: 2025-07-28].

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *International conference on machine learning*, 448–456. http://proceedings.mlr.press/v37/ioffe15.html

Johnson, J., Alahi, A., & Fei-Fei, L. (2016). Perceptual losses for real-time style transfer and super-resolution. https://doi.org/10.48550/arXiv.1603.08155

Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2017). Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*. https://doi.org/10.48550/arXiv.1710.10196

Kumar, M., Weissenborn, D., & Kalchbrenner, N. (2021). Colorization transformer. https://doi.org/10.48550/arXiv.2102.04432

Kupyn, O., Budzan, V., Mykhailych, M., Mishkin, D., & Matas, J. (2018). Deblurgan: Blind motion deblurring using conditional adversarial networks. *CVPR*.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Aitken, A., Tejani, A., Totz, J., Wang, Z., & Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. *CVPR*.

Levin, A., Lischinski, D., & Weiss, Y. (2004). Colorization using optimization. *ACM Transactions on Graphics*, *23*(3), 689–694. https://doi.org/10.1145/1015706.1015780

Loshchilov, I., & Hutter, F. (2017). Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*. https://doi.org/10.48550/arXiv.1711.05101

Miyato, T., Kataoka, T., Koyama, M., & Yoshida, Y. (2018). Spectral normalization for generative adversarial networks. *arXiv preprint arXiv:1802.05957*. https://doi.org/10.48550/arXiv.1802.05957

Rahman, T. (2021). Covid-19 radiography database [Accessed: 2025-07-28].

Reinhard, E., Ashikhmin, M., Gooch, B., & Shirley, P. (2001). Color transfer between images. *IEEE Computer Graphics and Applications*, *21*(5), 34–41. https://doi.org/10.1109/38.946629

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. https://doi.org/10.48550/arXiv.1505.04597

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al. (2015). Imagenet large scale visual recognition challenge. *International journal of computer vision*, *115*(3), 211–252. https://doi.org/10.1007/s11263-015-0816-y

Saharia, C., Chan, W., Chang, H., Lee, C. A., Ho, J., Salimans, T., Fleet, D. J., & Norouzi, M. (2021). Palette: Image-to-image diffusion models. https://doi.org/10.48550/arXiv.2111.05826

Shah, A. (2021). Large scale multi-camera detection dataset [Accessed: 2025-07-28].

Sharma, A. (2018). Tiny imagenet [Accessed: 2025-07-28].

Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., Rueckert, D., & Wang, Z. (2016). Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. *arXiv preprint arXiv:1609.05158*. https://doi.org/10.48550/arXiv.1609.05158

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. https://doi.org/10.48550/arXiv.1409.1556

Smith, L. N. (2018). A disciplined approach to neural network hyper-parameters: Part 1–learning rate, batch size, momentum, and weight decay. *arXiv preprint arXiv:1803.09820*. https://doi.org/10.48550/arXiv.1803.09820

Timofte, R., Agustsson, E., Van Gool, L., et al. (2018). Ntire 2018 challenge on single image super-resolution: Methods and results. *CVPR Workshops*.

Wang, W., Yang, J., Lin, X., Liu, J., Yang, J., Xu, Y., Qiao, Y., & Dai, Q. (2021). Blind real image denoising via deep hybrid denoiser prior. *CVPR*.

Wang, X., Girshick, R., Gupta, A., & He, K. (2017). Non-local neural networks. *arXiv preprint arXiv:1711.07971*. https://doi.org/10.48550/arXiv.1711.07971

Wang, X., Girshick, R., Gupta, A., & He, K. (2018). Non-local neural networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 7794–7803. https://doi.org/10.1109/CVPR.2018.00813

Welsh, T., Ashikhmin, M., & Mueller, K. (2002). Transferring color to greyscale images. *ACM Transactions on Graphics*, *21*(3), 277–280. https://doi.org/10.1145/566570.566576

Zhang, H., Goodfellow, I., Metaxas, D., & Odena, A. (2018). Self-attention generative adversarial networks. *arXiv preprint arXiv:1805.08318*. https://doi.org/10.48550/arXiv.1805.08318

Zhang, R., Isola, P., & Efros, A. A. (2016). Colorful image colorization. *ECCV*.