Shadow Removal via Diffusion Model

Neelesh Verma (neverma), Bhavesh Kumar Vasnani (bvasnani) @cs.stonybrook.edu

Abstract

Diffusion Denoising Probabilistic Models (DDPM) have recently showcased the benefits over traditional Generative Adversarial Networks (GANs) for several image generation scenarios.[12][7] This makes us believe that there's a scope for incorporating DDPM into shadow removal tasks. Shadow removal is however different from any image generation task as the model needs to preserve the hidden features while denoising to only produce the most contextual result. This conditionality has been the key motivation to use Diffusion models for shadow removal as recent works have shown promising results of methods to guide image generative process from a reference image.[2] The code for the entire project is available at https://github.com/neeleshverma/Shadow_Removal

Introduction

Image generation techniques have recently been heavily invested in since the coming Generative Networks. GANs followed by Diffusion models all allow us to generate samples with accuracy from any distribution. This has broadened the scope of sampling and several important papers like Pix2Pix[6] and DALL-E[9] gained popularity being the initial ones to portray the merits of image generation in several scenarios.

Shadow removal is a rather comparatively new problem to solve using Generative models and there has been some work to showcase that it's possible up to some extent. Generative models were originally developed to generate random samples until conditionality and the use of priors were accommodated. Shadow removal is different where we do need to generate the shadow region again but also ensure that the patterns, textures, and features hidden previously also come out properly in the generated image. The same output needs to match the illumination, and chromaticity of the rest of the shadow-free region.

Hence, DDPM models could possibly be the right direction allowing us to probabilistically weigh the diffusion through an inference that redirects the image generation closer to the inference image.[3] We try to depend on the same by passing the shadow-free images as the reference images so the model tries to generate samples closer to the shadow-free images when passed an image with shadow.

Related Works

Choi et al. [2] proposed Iterative Latent Variable Refinement (ILVR). This conditional DDPM learns by feeding the desired information in the training procedure and has shown remarkable success in various tasks. In this, the forward diffusion process will remain the same as in unconditional DDPM. They simply add Gaussian noise to the data. Assuming $x_0 \sim q(x_0)$, the forward diffusion process will generate random variables x_1, x_2,..., x_t.

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I) \tag{1}$$

where β_1, β_2, ..., β_t is a fixed variance schedule (chosen ahead of model training) and I is the identity matrix. Unconditional DDPM learns $p_\theta(x_{t-1}|x_t)$ in the reverse diffusion process parameterized by θ. However, we already have the original information of the shadowed region, let's represent it by sf. Therefore, the reverse diffusion process will be modeled by $p_\theta(x_{t-1}|x_t, sf)$. The ILVR algorithm suggests that this distribution can be approximated as -

$$p_\theta(x_{t-1}|x_t, sf) \sim p_\theta(x_{t-1}|x_t, \phi_N(x_{t-1}) = \phi_N(sf_t)) \tag{2}$$

where ϕ_N is a low-pass filtering operation, a sequence of downsampling and upsampling by N factor.

RePaint[7] proposed by Andreas et al. on the other hand was designed for inpainting tasks to generate high-quality images in the masked region. RePaint also does no change to the forward diffusion process like unconditional DDPM and only alters the reverse diffusion process. In each step of the diffusion process, RePaint samples the known region from the input passed and masks out the known region only. The unknown or masked region is generated from the DDPM pipeline with the inverted mask applied on

top. They further blend the two generated samples and use the resulting image for the next iteration. RePaint thus allows maintaining the features of the image in the unmasked region which is crucial for the Shadow Removal task.

Jiaming et al.[10] recently proposed Denoising Diffusion Implicit Models claiming it to accelerate slow sampling of Denoising Diffusion models via a non-Markovian diffusion process. Their inference process modifies the existing dependency of the DDPM on $q(x_t|x_0)$ by $q(x_{1:t}|x_0)$. The modification proposed to accelerate inference through the use of a non-Markovian process has led to 10x and 50x faster generation with very low error.

Objectives

We have divided our entire project into multiple objectives for clearer understanding:

- Validate results of RePaint on Places2 and CelebA-HQ
- Validate results of ILVR for faces
- Train Guided Diffusion model on augmented ISTD Dataset
- Extend ILVR, RePaint, DDIM for shadow removal tasks
- Generate best results from the combination of pipelines

Implementation Details

ILVR Method

We initially started with training the Guided Diffusion model on the ISTD Dataset. In order for the training to be successful, we accommodated Data Augmentation techniques to increase the dataset size. After training for about 110K Epochs, we used this trained model for further evaluation.

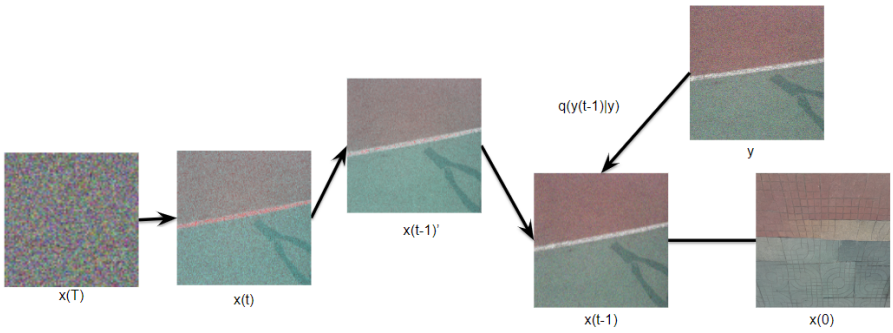

Figure 1: ILVR Pipeline for Shadow Removal Task

For ILVR, we only changed the reverse diffusion phase where we start from a completely noisy image and gradually reverse the noise effects to generate the sample. An inference image is presented to it at some timestamp; from there the losses are compared against this reference image so the end result generated is closer to this shadow-free reference image passes. This reference image can be passed at different stages of the inference process.

RePaint Method

We had to resize the images for RePaint and also generate inverted masks for the testing set to be used in the denoising step. The process is carried out separately for known and unknown regions. The unknown regions are generated by denoising noisy images and generating features within the shadow region which the known region is made noisy to match the noise levels for blending. This blended image was then used as the input for the next iteration.

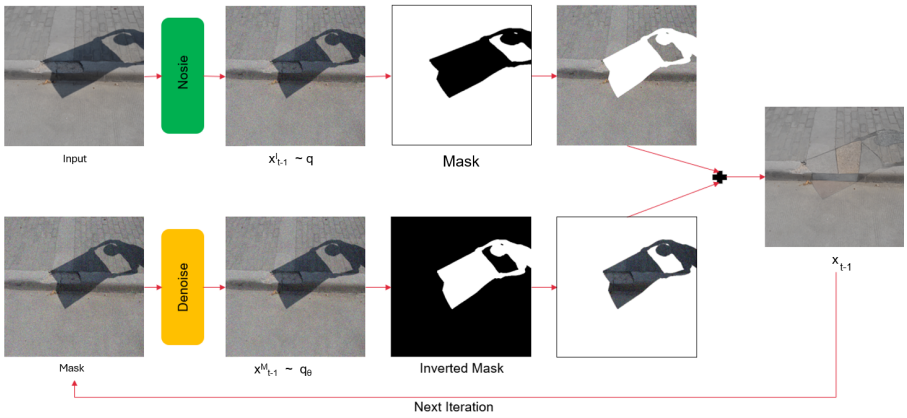

Figure 2: RePaint Pipeline for Shadow Removal Task

In order to produce patterns within the shadow region, we incorporated a Decay rate-based method to also pass in shadow information. With this method, we are not completely taking the generated image for the shadow region but also using some information from the original image. This decay rate is initially high and forces the blending to be closer to the shadow region features and gradually decreases over diffusion steps to produce a smooth reconstruction.

$$x_t = (\text{mask} * (\text{WeightedGT}) + (1 - \text{mask}) * ((1.0 - \text{decay}) * x_{t-1} + (\text{decay} * \text{WeightedGT})))$$

Experiments and Results

ILVR: We validated the results of the original ILVR paper on faces as reference images from the CelebA-HQ paper to test the claims. We then extended ILVR to Shadow Removal by first retraining the Guided Diffusion model on the ISTD Dataset and then using Shadow free image as a reference image during the inference process. ILVR didn't show promising results and couldn't complete the shadow removal task. We believe this is because the Dataset was not sufficient for ILVR to learn that it is supposed to remove the shadow or lack of proper mask and just shadow enough sufficient information for the iterative latent variable technique to generate appropriate samples. Some results are shown in figure 3.

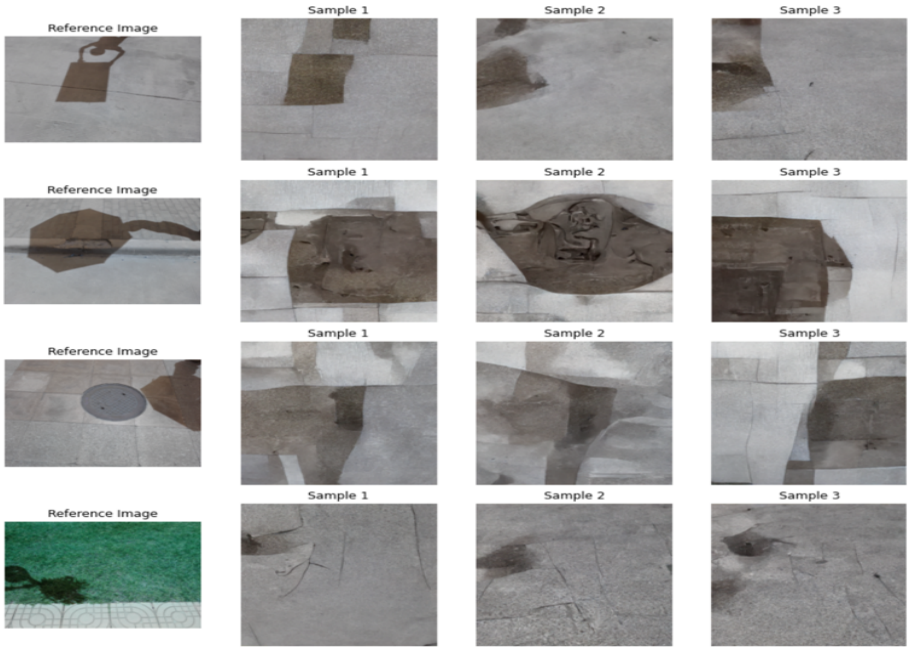

Figure 3: Shadow Removal Results using ILVR on ISTD

RePaint: RePaint was first validated by generating samples using pre-trained Places2 and CelebA-HQ models. Their results matched the claims made in the paper. To adapt RePaint-based inpainting for shadow removal application, we used the shadow image and also used shadow mask so the inpainting through DDPM only takes care inside the shadow region. This showed improved results and we could see the merits of this for inpainting tasks mapped to shadow removal. Some examples are shown in figure 4. However, the results inside the shadow region were randomly generated regions but not exact.

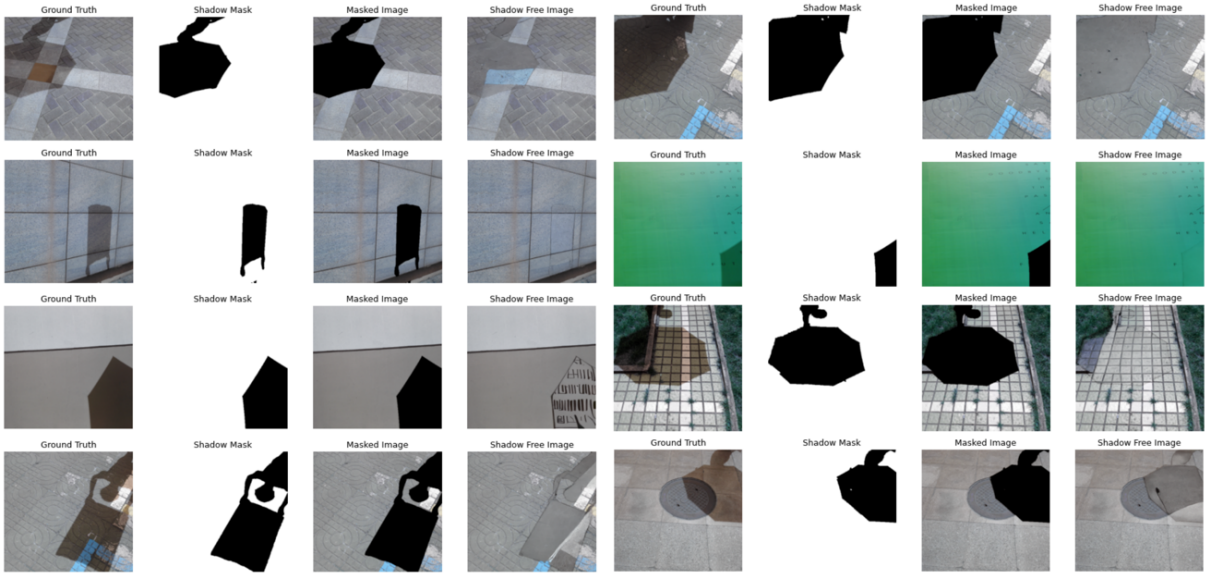

Figure 4: Shadow Removal Results using RePaint on ISTD

To also confirm that the results were not improper due to lack of sufficient training, we also experimented with removing shadow using the Places2 dataset as it is supposed to be the closest one trained for several scenes. Several results with Places2 were slightly off but the images that matched the trained images and features of Places2 showed that even patterns were regenerated correctly. This could be seen in figure 5. This means that with proper training of the model, this could be a potential direction for future work.

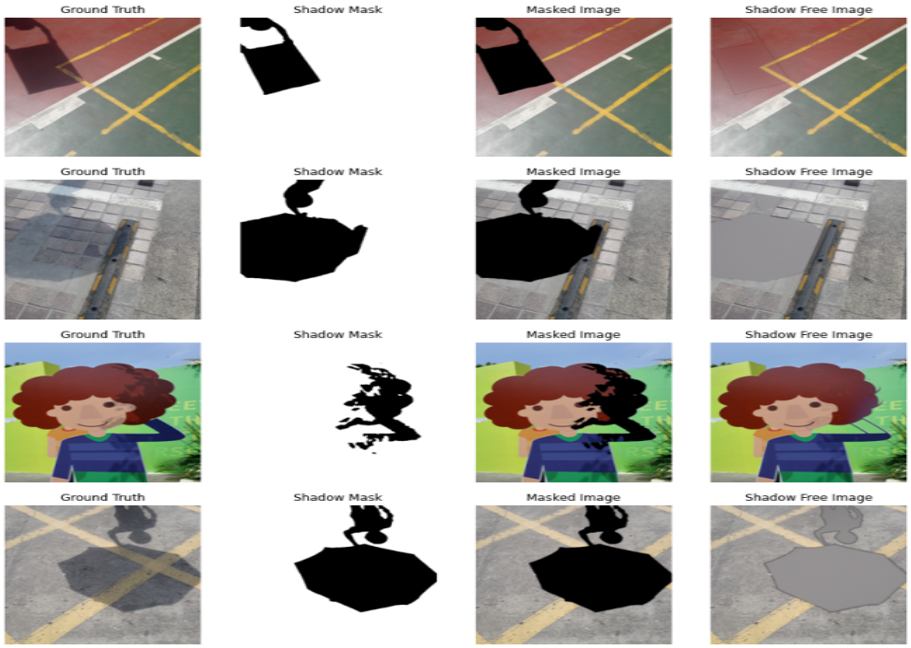

Figure 5: Shadow Removal Results using RePaint using Places2

RePaint doesn't reproduce patterns during the generation process as it misses shadow information. Therefore we pass shadow information through a decay rate logic. We use the decay rate to gradually influence the DDPM process for the unknown region and we are able to generate a pattern closer to the shadow region. We try the inference for varying decay rates to also see the ideal decay rate. Through this analysis, we can also see that the generation gets closer to the pattern information passed until a convergence point after which it remains constant and doesn't improve.
In figure 6, we can observe that the underlying pattern is generated as we lower the decay rate, however, it stops improving after a certain decay rate value.

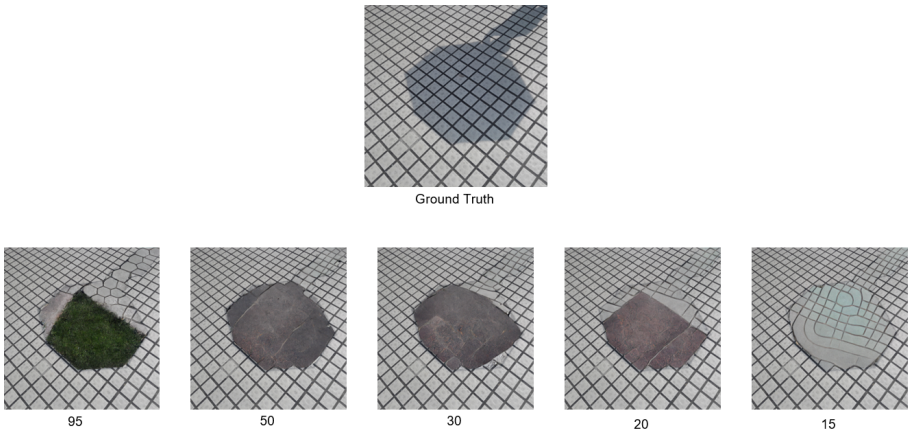

Figure 6: Shadow Removal Results with Decay Rate and RePaint

DDIM: DDIM's inference times were the fastest like its claims. Results from DDIM were also able to generate the patterns previously hidden in the image accurately but the results were off with colors. Results from DDIM had a major use of the color blue which we believe is because of the pre-trained church model used. Since the model has seen most images in the blue sky it tried to generate that but we believe seeing its pattern-generating merits, it could potentially prove beneficial for shadow removal.

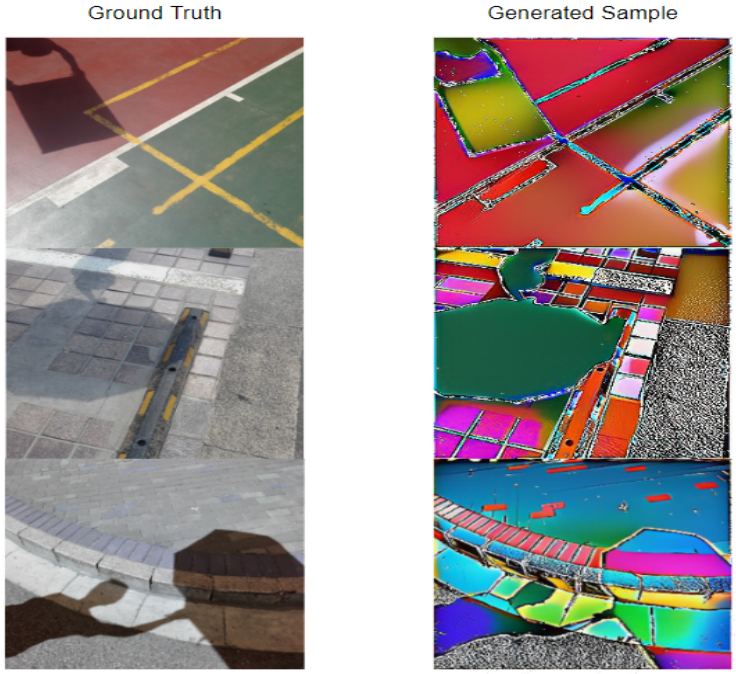

Figure 7: Shadow Removal Results DDIM

We can infer from the results that the DDIM7 is giving poor coloring on our images. Also, it is not able to generate the underlying patterns. We were using the model given by the paper which was trained

on **LSUN-Church** dataset. Since there is a big difference between our ISTD dataset and the church dataset, the poor performance of the model is on the expected line.

Evaluation: We used the most commonly used Learned Perceptual Image Patch Similarity (LPIPS) score which generates a similarity score between the activations of two patches for networks. It is widely acclaimed as it closely matches the human perception and thus a lower score means more similar patches.[13]

Method	LPIPS
MaskShadowGAN[5]	0.25
CycleGAN[14]	0.118
DeShadowNet[8]	0.080
DSC[4]	0.202
ST-CGAN[11]	0.067
Ours(RePaint Based)	0.3258
Ours(RePaint + decay)	0.3094

The LPIPS score for Ours(RePaint Based) method is generated for 52 images while the score for the Ours(RePaint + Decay) is generated for 16 images. Based on the LPIPS score, we understand that the generative capacity of our model is not as effective. We can see that the Ours(RePaint + Decay) based improvement on RePaint showed an improvement of 0.0164 on the LPIPS score showing that the improvement actually suits the shadow removal task. However, even for the decay-based method which generates patterns, the LPIPS is lesser than others as we believe that this is because of a combination of less dataset sample size leading to poor learning of the model and also that all these methods are inference-based methods. This could help us conclude that inference-based methods are not as effective for shadow removal and the future direction could be to change training of diffusion models through custom priors or different losses.

Conclusion

From the experimentation and results obtained, we can conclude that Diffusion Models have not yet reached the stage to be directly suited for the Shadow removal task. Controlling and guiding the inference process has proved to be difficult in order to generate expected shadow-free results. This could primarily be due to the small size of datasets available for shadow tasks. Diffusion models have shown significant results when trained with millions of images but with a small dataset, the results could be off from the expectation.

Also, we believe that shadow removal task along with diffusion models could be improved with the use of custom priors. The approach of injecting a custom prior to diffusion models is an area under research with not much concrete work published so far. Ideally, we would want to not pass a noisy image but instead pass a shadow image and the model understands shadow as noise and tries to reverse the effects of shadow. Cold Diffusion[1] recently showcased that instead of Gaussian noise, we could also use blur, snowification, and inpainting but these are yet performed for very basic operations. With more research in this field, this could be a viable option for Shadow Removal tasks.

Future Work

Since DDIM showcased merits in generating patterns during the inference phase, we wish to understand how that could be improved to fix the quality of the image received. We would also like to explore more on the use of custom prior and test the efficiency of Diffusion models. Interpreting shadow as noise could be the most natural and fast way for diffusion models could regenerate the hidden information to produce high-quality shadow-free images. Also, a significant effort has to be made to properly train the diffusion model, implying that we need to find avenues to extend the available datasets or modify training so the efficiency is not traded off because of the small dataset.

References

[1] Arpit Bansal et al. *Cold Diffusion: Inverting Arbitrary Image Transforms Without Noise*. 2022. DOI: 10.48550/ARXIV.2208.09392. URL: https://arxiv.org/abs/2208.09392.

[2] Jooyoung Choi et al. *ILVR: Conditioning Method for Denoising Diffusion Probabilistic Models*. 2021. DOI: 10.48550/ARXIV.2108.02938. URL: https://arxiv.org/abs/2108.02938.

[3] Jonathan Ho, Ajay Jain, and Pieter Abbeel. *Denoising Diffusion Probabilistic Models*. 2020. DOI: 10.48550/ARXIV.2006.11239. URL: https://arxiv.org/abs/2006.11239.

[4] Xiaowei Hu et al. "Direction-Aware Spatial Context Features for Shadow Detection and Removal". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 42.11 (Nov. 2020), pp. 2795–2808. DOI: 10.1109/tpami.2019.2919616. URL: https://doi.org/10.1109%2Ftpami.2019.2919616.

[5] Xiaowei Hu et al. "Mask-ShadowGAN: Learning to Remove Shadows From Unpaired Data". In: *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, Oct. 2019. DOI: 10.1109/iccv.2019.00256. URL: https://doi.org/10.1109%2Ficcv.2019.00256.

[6] Phillip Isola et al. *Image-to-Image Translation with Conditional Adversarial Networks*. 2016. DOI: 10.48550/ARXIV.1611.07004. URL: https://arxiv.org/abs/1611.07004.

[7] Andreas Lugmayr et al. *RePaint: Inpainting using Denoising Diffusion Probabilistic Models*. 2022. DOI: 10.48550/ARXIV.2201.09865. URL: https://arxiv.org/abs/2201.09865.

[8] Liangqiong Qu et al. "DeshadowNet: A Multi-context Embedding Deep Network for Shadow Removal". In: July 2017, pp. 2308–2316. DOI: 10.1109/CVPR.2017.248.

[9] Aditya Ramesh et al. *Zero-Shot Text-to-Image Generation*. 2021. DOI: 10.48550/ARXIV.2102.12092. URL: https://arxiv.org/abs/2102.12092.

[10] Jiaming Song, Chenlin Meng, and Stefano Ermon. *Denoising Diffusion Implicit Models*. 2020. DOI: 10.48550/ARXIV.2010.02502. URL: https://arxiv.org/abs/2010.02502.

[11] Jifeng Wang et al. *Stacked Conditional Generative Adversarial Networks for Jointly Learning Shadow Detection and Shadow Removal*. 2017. DOI: 10.48550/ARXIV.1712.02478. URL: https://arxiv.org/abs/1712.02478.

[12] Ling Yang et al. "Diffusion Models: A Comprehensive Survey of Methods and Applications". In: (2022). DOI: 10.48550/ARXIV.2209.00796. URL: https://arxiv.org/abs/2209.00796.

[13] Richard Zhang et al. *The Unreasonable Effectiveness of Deep Features as a Perceptual Metric*. 2018. DOI: 10.48550/ARXIV.1801.03924. URL: https://arxiv.org/abs/1801.03924.

[14] Jun-Yan Zhu et al. *Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks*. 2017. DOI: 10.48550/ARXIV.1703.10593. URL: https://arxiv.org/abs/1703.10593.