Image Style Transfer

Swapping Autoencoder

NeurIPS

2020

Swapping Autoencoder for Deep Image Manipulation

Taesung Park[1,2] Jun-Yan Zhu[2] Oliver Wang[2] Jingwan Lu[2]
Eli Shechtman[2] Alexei A. Efros[1,2] Richard Zhang[2]

[1]UC Berkeley [2]Adobe Research

Abstract

Deep generative models have become increasingly effective at producing realistic images from randomly sampled seeds, but using such models for *controllable manipulation of existing images* remains challenging. We propose the Swapping Autoencoder, a deep model designed specifically for image manipulation, rather than random sampling. The key idea is to encode an image into two independent components and enforce that any swapped combination maps to a realistic image. In particular, we encourage the components to represent structure and texture, by enforcing one component to encode co-occurrent patch statistics across different parts of the image. As our method is trained with an encoder, finding the latent codes for a new input image becomes trivial, rather than cumbersome. As a result, our method enables us to manipulate real input images in various ways, including texture swapping, local and global editing, and latent code vector arithmetic. Experiments on multiple datasets show that our model produces better results and is substantially more efficient compared to recent generative models.

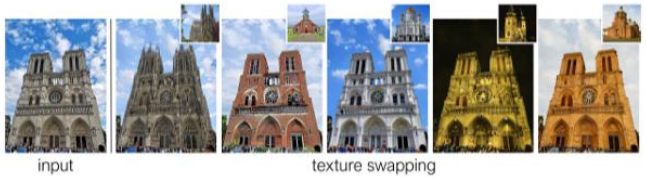


Figure 1: Our Swapping Autoencoder learns to disentangle texture from structure for image editing tasks. One such task is texture swapping, shown here. Please see our project webpage for a demo video of our editing method.

1 Introduction

Traditional photo-editing tools, such as Photoshop, operate solely within the confines of the input image, i.e. they can only "recycle" the pixels that are already there. The promise of using machine learning for image manipulation has been to incorporate the *generic visual knowledge* drawn

Park, T., Zhu, J. Y., Wang, O., Lu, J., Shechtman, E., Efros, A., & Zhang, R. (2020). Swapping autoencoder for deep image manipulation. *Advances in Neural Information Processing Systems*, *33*, 7198-7211.

Image Style Transfer

Swapping Autoencoder

Background & Goal

Previous research limitation

- Challenges in image editing : GAN-based models lack precise control over the latent vector, making direct image m odification difficult.
- Challenges in disentanglement: Traditional style transfer changes only colors and low-level textures, struggling to pr eserve structure.

Goal

- To aim for efficient disentanglement of structure and texture for image editing.

Park, T., Zhu, J. Y., Wang, O., Lu, J., Shechtman, E., Efros, A., & Zhang, R. (2020). Swapping autoencoder for deep image manipulation. *Advances in Neural Information Processing Systems*, *33*, 7198-7211.

Image Style Transfer

Swapping Autoencoder

Network Architecture

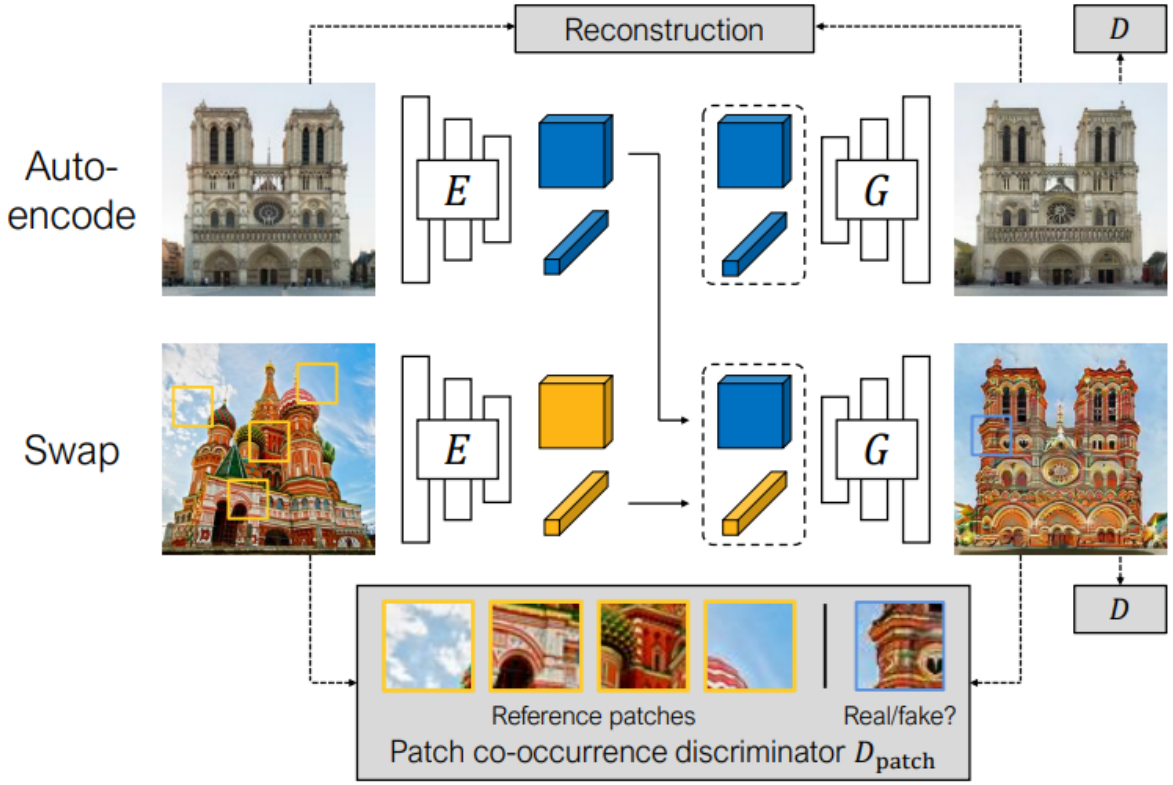


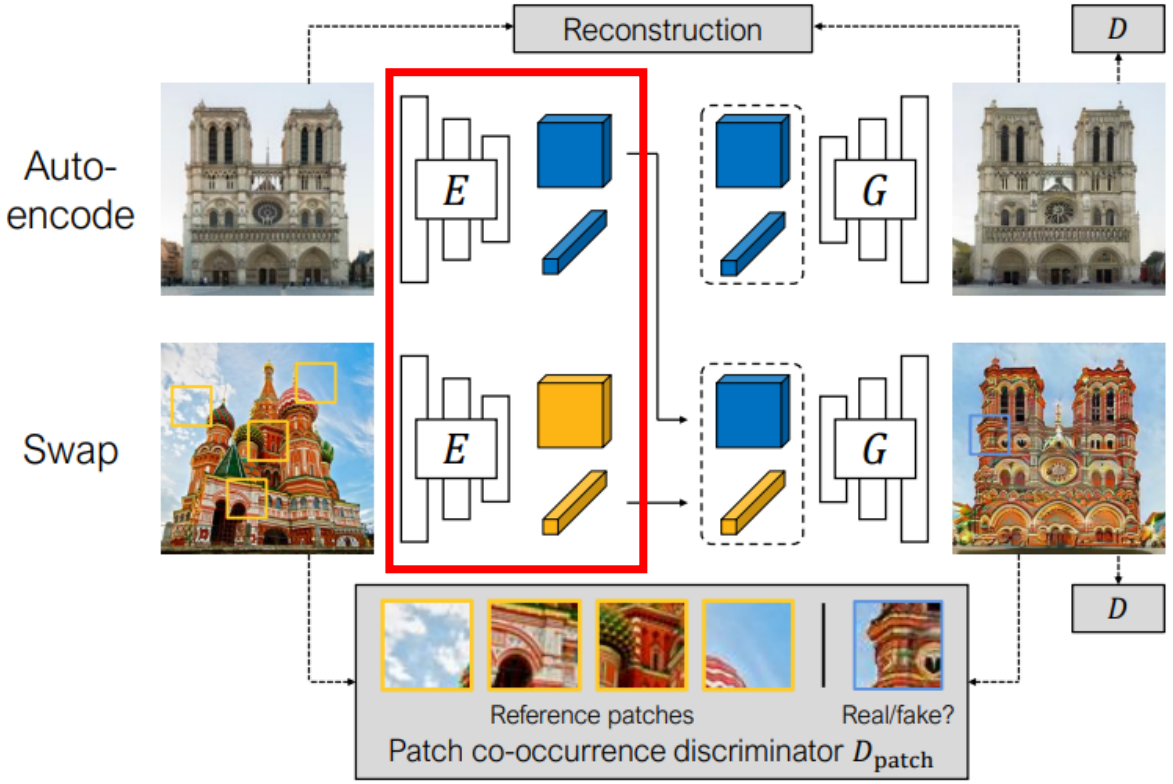
Fig 1. Overview of network architecture

Park, T., Zhu, J. Y., Wang, O., Lu, J., Shechtman, E., Efros, A., & Zhang, R. (2020). Swapping autoencoder for deep image manipulation. *Advances in Neural Information Processing Systems*, *33*, 7198-7211.

Image Style Transfer

Swapping Autoencoder

Network Architecture



Content-style disentanglement

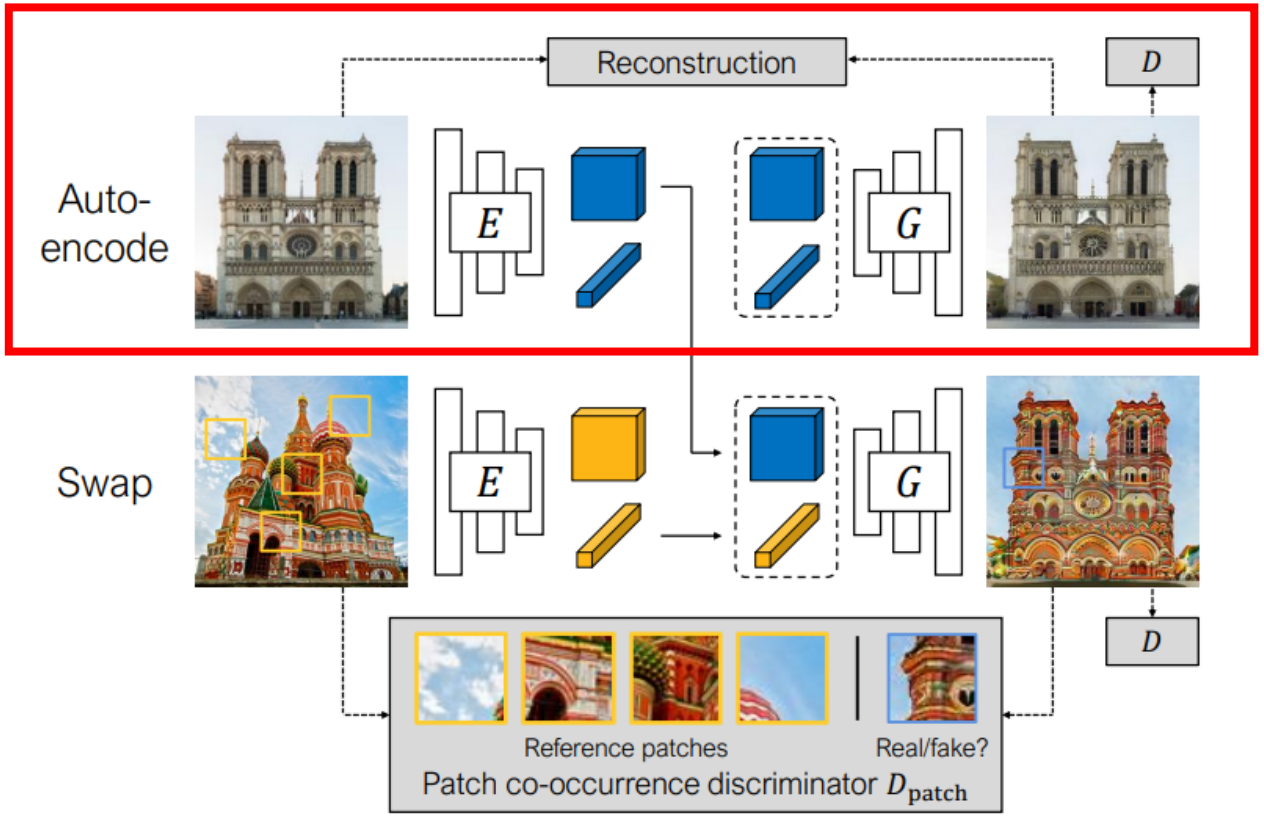
- Dual latent code representation
- Structure code : encodes spatial information, deter mining the image's shape and outline.
- Texture code : represents overall style, color, and patterns.

Park, T., Zhu, J. Y., Wang, O., Lu, J., Shechtman, E., Efros, A., & Zhang, R. (2020). Swapping autoencoder for deep image manipulation. *Advances in Neural Information Processing Systems*, *33*, 7198-7211.

Image Style Transfer

Swapping Autoencoder

Loss Functions



Autoencoder

- Reconstruction loss : ensures the generated image closely resembles the original

$$\mathcal{L}_{\text{rec}}(E,G) = \mathbb{E}_{\mathbf{x}\sim\mathbf{X}}[\|\mathbf{x} - G(E(\mathbf{x}))\|_1].$$

Eq 1. Reconstruction loss function

- GAN loss : the generated image to appear real, resulting in more natural outcomes.

$$\mathcal{L}_{\text{GAN,rec}}(E,G,D) = \mathbb{E}_{\mathbf{x}\sim\mathbf{X}}[-\log(D(G(E(\mathbf{x}))))].$$

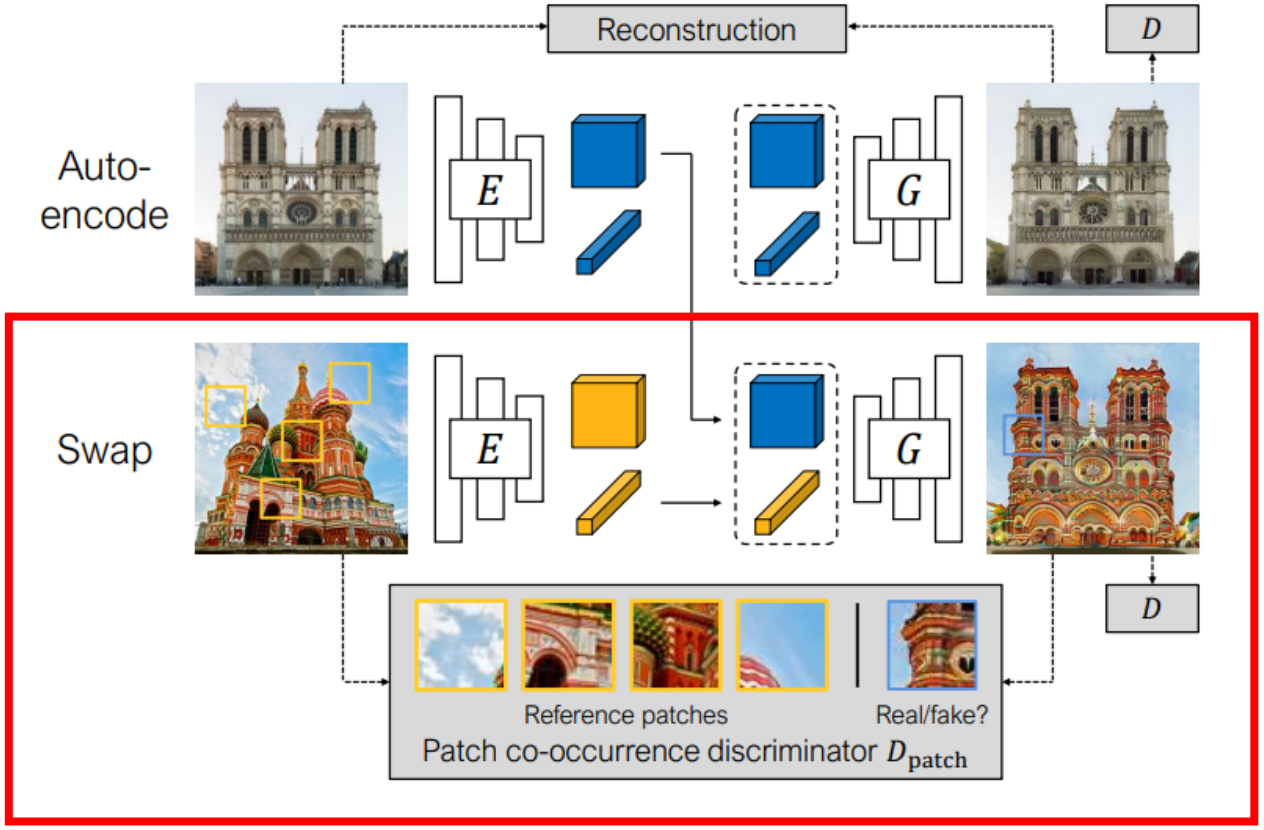
Eq 2. GAN loss function

Park, T., Zhu, J. Y., Wang, O., Lu, J., Shechtman, E., Efros, A., & Zhang, R. (2020). Swapping autoencoder for deep image manipulation. *Advances in Neural Information Processing Systems*, *33*, 7198-7211.

Image Style Transfer

Swapping Autoencoder

Loss Functions



Swapping

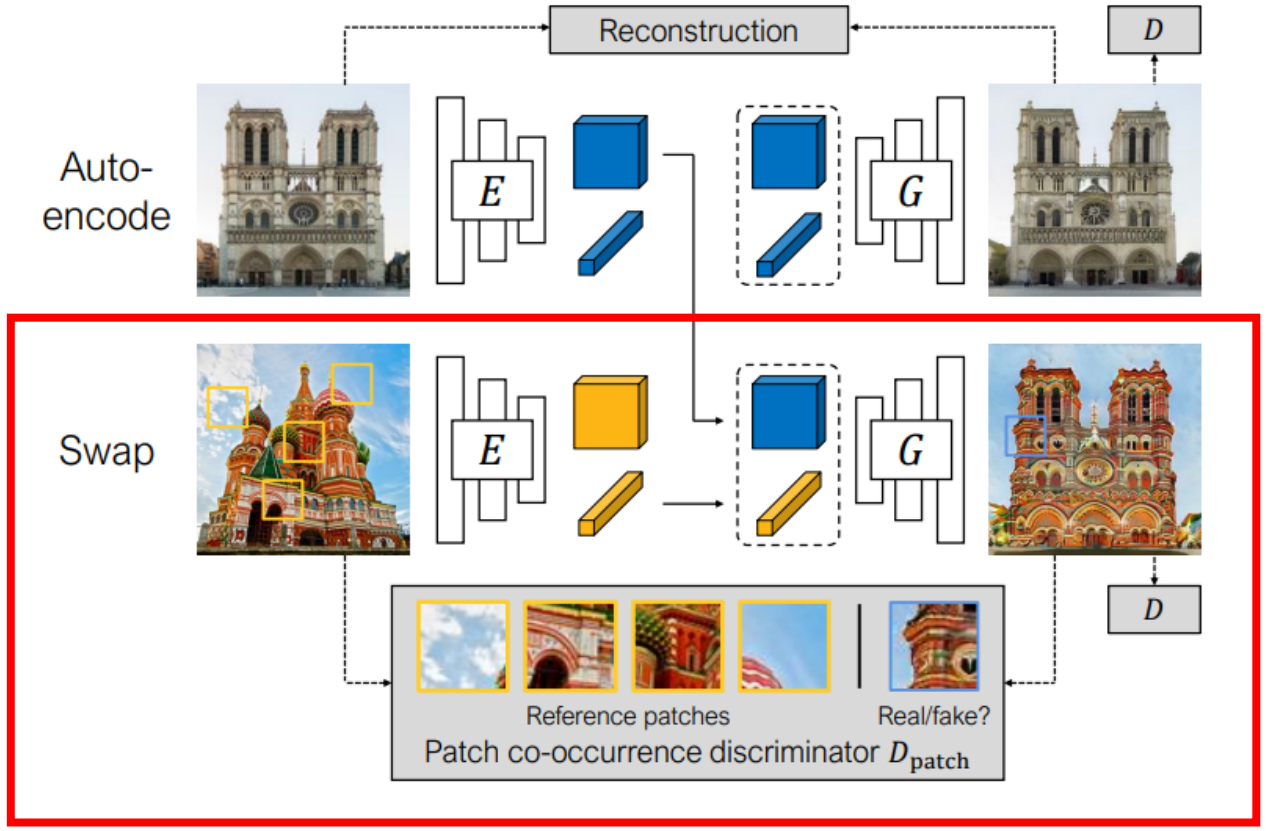
- Swap the structure code and texture code of two i mages.
- Trains to manipulate structure and texture indepen dently.
- Enables flexible style transformation while maintai ning realistic images.

Park, T., Zhu, J. Y., Wang, O., Lu, J., Shechtman, E., Efros, A., & Zhang, R. (2020). Swapping autoencoder for deep image manipulation. *Advances in Neural Information Processing Systems*, *33*, 7198-7211.

Image Style Transfer

Swapping Autoencoder

Loss Functions



Swapping

- GAN loss : the generated image to appear real, re sulting in more natural outcomes.

$$\mathcal{L}_{\text{GAN,swap}}(E,G,D) = \mathbb{E}_{\mathbf{x}^1,\mathbf{x}^2\sim\mathbf{X},\mathbf{x}^1\neq\mathbf{x}^2}\left[-\log(D(G(\mathbf{z}_s^1,\mathbf{z}_t^2)))\right],$$

Eq 3. GAN loss function with swapped images

- Co-occurrence loss : trains the generated image to maintain the same texture.

$$\mathcal{L}_{\text{CooccurGAN}}(E,G,D_{\text{patch}}) =$$

$$\mathbb{E}_{\mathbf{x}^1,\mathbf{x}^2\sim\mathbf{X}}\left[-\log\left(D_{\text{patch}}\left(\texttt{crop}(G(\mathbf{z}_s^1,\mathbf{z}_t^2)),\texttt{crops}(\mathbf{x}^2)\right)\right)\right]$$

Eq 4. Co-occurrence loss function

Park, T., Zhu, J. Y., Wang, O., Lu, J., Shechtman, E., Efros, A., & Zhang, R. (2020). Swapping autoencoder for deep image manipulation. *Advances in Neural Information Processing Systems*, *33*, 7198-7211.

Image Style Transfer

Swapping Autoencoder

Experiments

Fig 2. Results of the style transfer

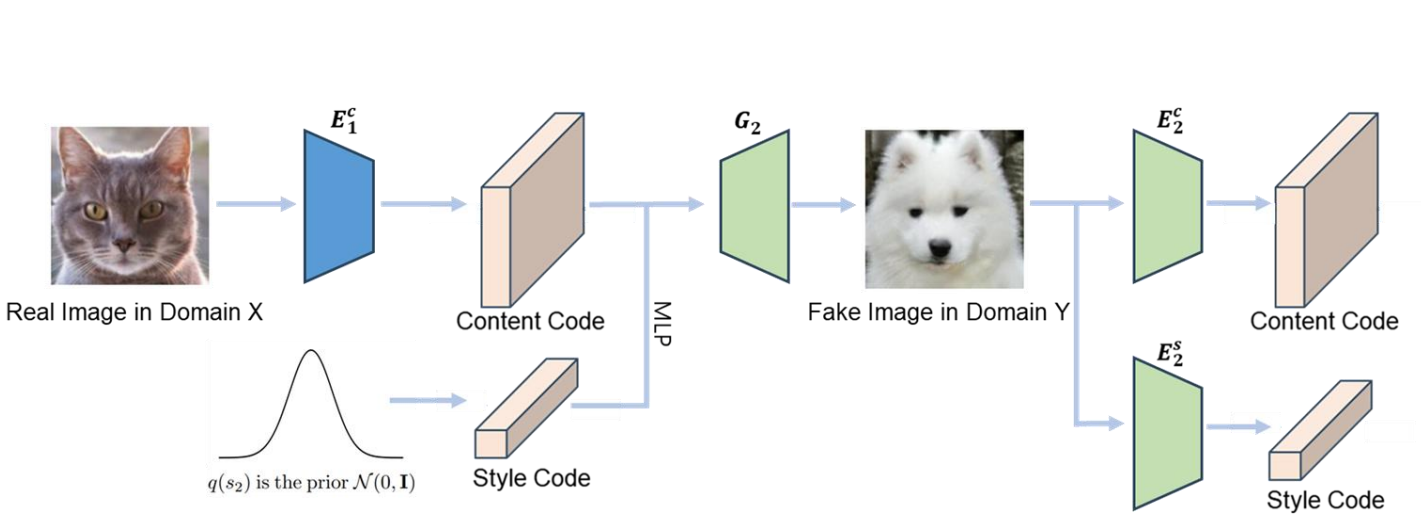
Park, T., Zhu, J. Y., Wang, O., Lu, J., Shechtman, E., Efros, A., & Zhang, R. (2020). Swapping autoencoder for deep image manipulation. *Advances in Neural Information Processing Systems*, *33*, 7198-7211.

Image Style Transfer

Swapping Autoencoder

Experiments

- MUNIT vs Swapping Autoencoder



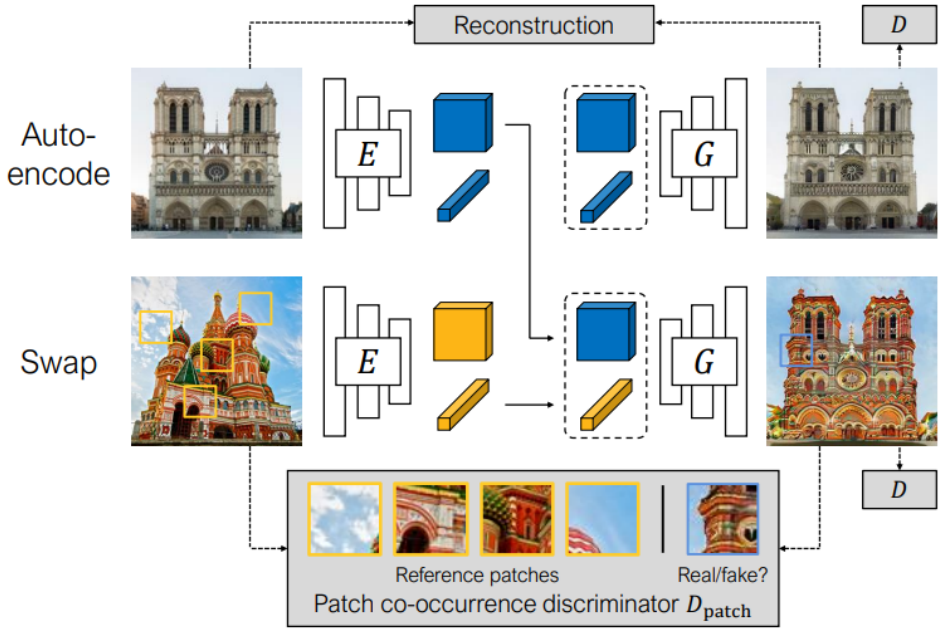


Fig 3. Compare MUNIT

Park, T., Zhu, J. Y., Wang, O., Lu, J., Shechtman, E., Efros, A., & Zhang, R. (2020). Swapping autoencoder for deep image manipulation. *Advances in Neural Information Processing Systems*, *33*, 7198-7211.

Image Style Transfer

Swapping Autoencoder

Implications & Limitations

Implications

- Enables image translation across diverse domains without data labeling.

Limitations

- The patch-based discriminator has limitations, leading to inconsistencies in style uniformity.