

Comparison and Analysis of Common Models for Special Effect Makeup Transfer

Xiaoyan Wang 36920221153121¹ Wenfang Xiang 36920221153126¹

Zhiqun Yang 36920221153137¹ Xiaorui Huang 36920221153085¹ Ximing Zhang 36920221153148¹

¹ Artificial Intelligence Research Institute

Abstract

Makeup transfer is one of the applications of image style transfer, which refers to transfer the reference makeup to the face without makeup, and maintaining the original appearance of the plain face and the makeup style of the reference face. Compared with the traditional makeup technology, makeup transfer has a high degree of freedom, which allows users to get makeup from the real model pictures independently and arbitrarily, which greatly enriches the diversity of makeup. Transferring makeup from the misaligned reference image is challenging. There are many algorithms related to makeup transfer to solve problems in different directions, such as PSGAN, SCGAN, CPM, etc. We compare the above three methods, and repeat the relevant algorithms, analyze their advantages and disadvantages and possible directions for improvement, which is helpful to quickly understand the new solutions in the direction of makeup transfer in recent years and the problems still to be solved. Since CPM proposed a unified template, which could align 3D head posture, facial shape and facial expression of source image and target image, and the generated effect was the best among the three methods, we attempt to add quantization to the reasoning process of CPM to speed up its reasoning process. In view of the excessive and redundant use of computing resources and model parameters caused by U-net in the original model, we also try to replace the U-net module with attention U-net, and suppress irrelevant information in images through the attention mechanism to highlight local important features.

1. Introduction

Virtual make-up testing technology has always been one of the most important technologies in the beauty market. At present, the popular mainstream technology in this field is traditional material makeup. This technology refers to that professional designers design makeup materials according to the specified format, and then apply makeup materials to the corresponding position of five facial features by using face key point detection. However, this is not convenient for the operation of imitation makeup, and the transfer of makeup can solve such problems well. By providing the reference image of imitation makeup, the makeup in the reference image can be transferred to the source image. Compared with traditional makeup technology, the transfer of

makeup has a high degree of freedom, which enables users to get makeup from real models independently and arbitrarily, instead of being limited to the makeup designed by designers, greatly enriching the diversity of makeup. Makeup transfer technology can not only transfer the makeup information of five facial features, but also transfer the overall information of skin color, light and shadow. Plus, the makeup information comes directly from real reference images, so it's a great way to keep the makeup realistic.

However, Makeup in real life is rich in diversity, including not only color changes but also patterns, such as stickers, blush, jewelry, etc. However, existing works ignore the latter components, limit the transfer of makeup to color manipulation, and only pay attention to the light makeup style, which is not good for the transfer effect of special effects makeup. The transfer of special effects makeup is of great research significance. For example, in the process of actor audition, the effect of audition can be viewed through the transfer of special effects makeup, which greatly saves the cost and time of makeup and has very important commercial value.

In this paper, we choose PSGAN(Jiang et al. 2020) and SCGAN(Deng et al. 2021) to illustrate the existing cutting-edge methods of ordinary makeup transfer, but these two methods cannot transfer special effects makeup. We reproduced the relevant algorithm and inputted our own test pictures to conduct the transfer test of special effect makeup, as shown in Fig.1. It can be seen that SCGAN's transfer effect of special effect makeup is not very good. Obviously, the transfer of effects such as eye makeup and texture is not very good, so to speak, only the base makeup has been transferred. We introduce and analyze their advantages and disadvantages in detail, and give the corresponding improvement direction. Secondly, because CPM(Nguyen, Tran, and Hoai 2021) proposed a unified template, this template can align the 3D head posture, facial shape and facial expression of the source image and the target image, which can carry out the transfer of special effect makeup. We compare the differences between CPM(Nguyen, Tran, and Hoai 2021) and SCGAN(Deng et al. 2021) to solve spatial dislocation, and analyze the existing shortcomings of CPM(Nguyen, Tran, and Hoai 2021). Finally, as CPM's reasoning process takes a long time, we tried to add quantization to speed up its reasoning process.

To sum up, our work mainly includes the following three parts:

- Three makeup transfer algorithms, PSGAN, SCGAN and CPM, are reproduced and compared, and their advantages and disadvantages, as well as the future improvement directions are summarized and analyzed.
- Attempts were made to add quantization to CPM to accelerate its reasoning process, which was accelerated by 2 seconds.
- We also attempt to replace the U-net module with attention U-net, and suppress irrelevant information in images through the attention mechanism to highlight local important features.



Figure 1: A demonstration of the transfer of special effects makeup using SCGAN.

2. Related Work

In recent years, there have been a large number of studies on facial makeup transfer based on deep learning. We first reviewed the traditional methods of makeup transfer, and then briefly introduced the work similar to PSGAN(Jiang et al. 2020), SCGAN(Deng et al. 2021) and CPM(Nguyen, Tran, and Hoai 2021).

Makeup migration can be seen as a problem of image domain transformation(Taigman, Polyak, and Wolf 2016), that is, learning a generation function that maps a sample image from the source domain to the target domain. There has also been work proposed on pixel to pixel frame mapping(Isola et al. 2017), but this approach requires pairs of pictures during training. CycleGAN(Zhu et al. 2017) incorporates cyclic consistency losses into the generation network to produce pictures that fulfill the specified domain distribution without requiring strict pairs of pictures to train. Since the generative adversarial networks(GANs)(Goodfellow et al. 2020) was proposed, the global discriminator has been used in a lot of makeup transfer work. Although the global discriminator can discriminate between pictures from many domains in the domain transformation issue, it can only capture the global structure, by assuming the independence of pixels split by the patch diameter and modeling the picture as a Markov random field to account for this, the local (patch) discriminator can learn. The local discriminator was first used in neural network training process by Li *et al.*(Li and Wand

2016). The local discriminator helps to prompt the generator to capture local details with L1 loss by using patchGAN structure(Isola et al. 2017). You can also use a combination of the global discriminator to capture the global structure and the local discriminator to complete the details so that the generated path and context information are consistent.

While CycleGAN(Zhu et al. 2017) can relax its requirements for pairing images, it can't use specific reference makeup images for migration, which greatly limits the flexibility of makeup migration. PairedCycleGAN(Chang et al. 2018) solves this problem by making it possible to transfer makeup based on a specific reference image. BeautyGlow(Chen et al. 2019) divided the latent features into two parts: non-makeup features and makeup features using Glow framework(Kingma and Dhariwal 2018). BeautyGAN(Li et al. 2018) combines the global domain-level loss and the local instance-level loss, using two GAN(Goodfellow et al. 2020) networks to solve the makeup transfer problem. One of the biggest problems with the above methods is that they can not solve the spatial misalignment problem, which is very important in the makeup transfer, because the makeup transfer requires the original image and the reference image, and the images must be aligned before the makeup transfer can be carried out. PSGAN(Jiang et al. 2020) solves the problem of image alignment by introducing an attention mechanism, but this method requires pixel-level computation, which is expensive for local migration. SCGAN(Deng et al. 2021) can take into account the calculation cost and alignment effect, and has spatial invariance for the face in the image.

For the parts of lip and eye makeup transfer, CAGAN(Kips et al. 2020) uses a color discriminator to improve the transfer. Various previous approaches have focused on the transfer of makeup colors, while real-life makeup styles can also cover pattern-based components such as stickers, face painting, and embellishments. LADN(Gu et al. 2019) solves part of the problem of pattern-based makeup transfer, but it does not handle the details of complex patterns well. When a pair of original images and reference images have different head positions, the transfer effect may appear artifacts. In addition, the quality of the image after style transfer will be lower, the edge of the image will become blurred, and the original features of the image will be degraded. However, CPM(Nguyen, Tran, and Hoai 2021) can overcome the above shortcomings, showing good robustness to the head posture in the pictures, and good transfer effect for complex makeup patterns. The quality of the final output pictures can also be guaranteed.

3. Proposed Solution

3.1 PSGAN

In order to make the makeup adapt to the source picture and address the misalignment issue brought on by the attitude difference, PSGAN(Jiang et al. 2020) extracted the reference image into two makeup matrices, and then processed the deformation using the AMM module. The suggested DRNet(Hasan et al. 2021) applies the makeup matrix and, assigning pixel-by-pixel weight to the outcome

of makeup removal to implement re-makeup, to the original picture. By modifying the makeup matrix and, the design may achieve partial makeup transfer and overcomes the spatial dislocation problem. PSGAN primarily separates cosmetics-related features from face features (such as lip gloss and eye appeal) and depicts makeup-related features as two makeup matrices.

Existing problems and improvement directions:

- Since PSGAN is only sensitive to color and not to line and pattern makeup, this model can't transmit special effects makeup.

3.2 SCGAN

SCGAN divides the makeup transfer into 3 parts: The two extraction module: Part-specific Style Encoder and Face identity Encoder; One assignment module: Makeup fusion Decoder. Part-specific Style Encoder is used to extract style information; Face identity Encoder is used to extract face id information; Makeup fusion Decoder fuses style information with face id information and migrates it onto the source image. The Part-specific Style Encoder (PSEnc) proposed by SCGAN makes the transfer of makeup not limited to using only the front-facing photos. Even at different angles, even rotating the image so that there is large spatial misalignment between the source image and the reference image can be successfully migrated.

Existing problems and improvement directions:

- SCGAN can't do special effects transfer because the skin tone is transferred as a whole (the features are transferred separately) and special effects makeup is generally distributed on the skin.

3.3 CPM

A uniform template put out by CPM(Nguyen, Tran, and Hoai 2021) aligns the source and destination photos' 3D head positions, facial shapes, and expressions. The cosmetics transfer employed uv texture rather than the original picture and was built on the BeautyGAN(Li et al. 2018) platform. UV space can eliminates differences in 3D poses, shapes and expressions (also spatial misalignments).In CPM, it is proposed to combine color transformation and pattern, which can transfer both light makeup and special effect makeup. Two branches were designed for color and pattern transfer respectively, and the face was designed to be placed in UV space to eliminate the differences in the input image in face shape, head posture and expression. A new makeup dataset has also been introduced, containing extreme makeup styles.

Existing problems and improvement directions:

- The two branches of Color and Pattern in CPM tend to misidentify face shadows as colors, as shown in Fig.2
- The inference process of CPM is slow, probably because of the two branch operation.



Figure 2: The shadow of the folded face on the left cheek is identified as black.

- In view of the excessive and redundant use of computing resources and model parameters caused by U-net in the original model, we also attempt to replace the U-net module with attention U-net, and suppress irrelevant information in images through the attention mechanism to highlight local important features.

3.4 Add Quantization into CPM

We use quantitative techniques(Gholami et al. 2021; Polino, Pascanu, and Alistarh 2018; Li et al. 2021) to compress the model and improve the efficiency of the network. We round the 32-bit floating point values used by the model, leaving the lower precision 8-bit integer values, so that the network has approximately 4 times the compression rate. This is because the processor's integer computation instructions are more efficient than the corresponding floating-point computation instructions. At the same time, low bit values occupy less memory and storage space, which can reduce memory footprint and improve network efficiency. When quantization is added, the inference time is shortened by 2 seconds.

4. Experiments

In this section, we provide a comprehensive analysis of the experiment from several aspects. Specifically, (i) Section 4.1 describes the dataset used in the experiment and the related experimental settings, (ii) the experimental results are presented and analyzed in Section 4.2, and (iii) we compare the three models in Section 4.3 and select the CPM model for optimization.

4.1 Datasets and Experimental Settings

4.1.1 Datasets The dataset collected and used in the original PSGAN paper contains 772 facial images with a resolution of 256 × 256, including 403 makeup images and 269 non-makeup images.

The Makeup Transfer dataset, which contains 3834 photos, of which 1115 are makeup images and 2719 are not, was utilized in the SCGAN article. 100 non-makeup photographs and 250 make-up images were chosen for testing, and the rest were used for training after the original authors scaled the image to 256 × 256.

There are four datasets used in CPM: Stickers, CPM-Real, CPM-Synt-1, and CPM-Synt-2. Stickers has 577 RGBA images of different sizes; CPM-Real is an in-the-wild makeup dataset, and there are 3895 RGB images of different sizes in total; CPM-Synt-1 is a transferred pattern dataset used to train pattern-branch which contains 5555 RGB texture images and its size is 256×256 ; CPM-Synt-2 is also a transferred pattern dataset while it contains 1625 RGB triplets, which is consisted of source, reference and ground truth. And the image size is 256×768 .

On the basis of pre-training, we trained these models in our own environment using the Makeup Transfer dataset. There are 1115 photographs without makeup and 2719 images with makeup in the dataset, the resolution of which is 361×361 .

4.1.2 Experimental settings In this paper, we replicate three algorithms utilizing PyTorch(Paszke et al. 2019) on the MoPaaS cloud platform. We used an RTX-3090 graphics card, 10GB of RAM, and other resources.

We employ the Adam(Kingma and Ba 2014) optimizer in PSGAN. The experiment has trained for 50 epochs. The learning rate of the model is fixed at 0.0002, and the batch size is 1.

In SCGAN, we use VGG16(Simonyan and Zisserman 2014) to extract features. The optimizer used by the generator and discriminator is also Adam where $1=0.52=0.999$. The model’s learning rate is set to 0.0002 and its batch size is 1.

There are 100 epochs of CPM training, and the batch size is 1. Adam is also used as the optimizer. In particular, when the number of training epochs is an integral multiple of 25, the learning rate is adjusted to $1e-5$, and the learning rate remains unchanged $(1e-4)/4$ at other times.

4.2 Experimental results and analysis

Since there is no particularly effective method to evaluate the quality of generated images in academic circles(He et al. 2022), one of the most popular methods is to give some generated images to users of Amazon Mechanical Turk (AMT)(Paolacci, Chandler, and Ipeirotis 2010). However, this manual scoring method is not reproducible. Therefore, the results of the AMT scores were only used as an aid in this paper, and the qualitative analysis of the experimental results was mainly conducted.

We first conducted experiments with PSGAN(Liu et al. 2020), which refines the reference image into two matrices. The deformation process is performed by the AMM module to adapt the makeup to the source image waiting for makeup application, solving the misalignment problem caused by pose differences. The effect of PSGAN is shown in Fig.3. It can be seen that PSGAN has the most obvious migration of makeup skin tone and lip color both on and off, but it does not successfully migrate the details of makeup such as eye shadow and coloring. Secondly, we used SCGAN to conduct the experiment of special effect makeup transfer. SCGAN decomposes makeup transfer into two steps: extraction and distribution. From Fig.4, we can see that SCGAN has almost no transfer effect on special effect makeup, and there



(a) special effect makeup transfer with PSGAN



(b) automatic de-makeup with PSGAN

Figure 3: A display of makeup transfer using PSGAN.

are some problems such as the edge transition in local areas. Then we tested and duplicated CPM. Makeup transfer is divided into two categories by CPM: color branch and pattern branch. Fig.5 and Fig.6 depict the transfer effects of solely executing the color branch and the pattern branch, respectively, while Fig.7 depicts the transfer effects of both. These figures make it evident that CPM’s particular makeup transfer effect is effective, and that it performs better on the color branch than the pattern branch.

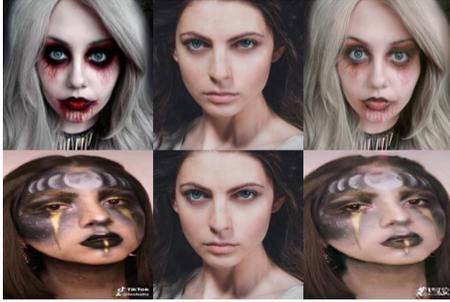
4.3 Comparison and optimization

We can observe from Sections 4.2 that PSGAN and SCGAN struggled with the process of transferring special effect makeup. After careful examination, we discovered that the PSGAN and SCGAN models are highly sensitive to characteristics like skin color but not to characteristics like patterns and textures. Neither model can successfully convey the special effect makeup because the majority of it concentrates on patterns.

We finally settle on the CPM model for optimization because its effect is clearly superior to that of the first two models. The transfer of CPM on the color branch is outstanding, but the transfer of makeup information on the pattern branch is subpar, and the model’s ongoing calculations consume a lot of resources. We hypothesize that this phenomena may be brought on by the model’s use of Resnet-50’s(He et al. 2016) U-net for picture segmentation. Cascaded neural networks with many levels are used by U-Net. The cascaded architecture isolates the region of interest (ROI) and does in-depth forecasting for this particular ROI. However, U-net causes a wasteful and repeated usage of processing power and model parameters. For instance, the cascade’s models



(a) special effect makeup transfer with SCGAN



(b) automatic de-makeup with SCGAN

Figure 4: A display of makeup transfer using SCGAN.



Figure 5: CPM transfer effect of Color-Only

all repeatedly extract the same low-level characteristics.

To this end, we substitute Attention U-net(Oktay et al. 2018) for U-net. The core idea of Attention U-net is to propose an Attention gate module, replace hard attention with soft attendance, and integrate attention into U-net's jump connection and up sampling modules to realize spatial attention mechanism. The attention mechanism is used to suppress the irrelevant information in the image and highlight the local important features.

When we use Attention U-net, the CPM model is still trained with Adam optimizer for 100 epochs and the batch size is changed to 4. Due to certain conditions, we only compared the performance of CPM with U-net and CPM with



Figure 6: CPM transfer effect of Pattern-Only



Figure 7: A display of makeup transfer using CPM.

	Dice Loss	IoU Score	Accuracy
CPM with U-net	0.1992	0.6774	0.9647
CPM with Attention U-net	0.2004	0.6756	0.9646

Table 1: Comparison of best performance on the validation set between CPM with U-net and CPM with Attention U-net for the first 30 epochs.

Attention U-net in the first thirty epochs under the same conditions, and the results are shown in Table1. Although the highest IoU score of CPM with Attention U-net is lower than the one with U-net in the first thirty epochs, under the complete 100 epochs of training, the dice loss of the CPM with Attention U-net on the validation set is reduced to about 0.12, the IoU score reaches 0.6888, the accuracy reaches 0.9625, which means that the model performs well.

In order to speed up the execution of model, we also incorporated quantization(Gray and Neuhoff 1998) to CPM, reducing the running time by two seconds.

Despite the addition of two optimizations, the model still has a great deal of space for improvement. In the future, the model can be improved from the two angles listed below to enhance the pattern branch's transfer effect in CPM: 1) Increase the accuracy of image segmentation. 2) Consider including the diffusion model.

5. Conclusion

Makeup transfer, which has drawn a lot of attention, offers users a quick and effective alternative for visualizing and replacing facial makeup. Face makeup transfer based on deep learning is still a challenging emerging topic. In this paper, we replicate and compare three excellent makeup transfer algorithms and apply them to special effect makeup and de-makeup tasks. We chose the CPM model that performed the best out of the three, refined its network architecture, incorporated quantization, and significantly enhanced the effect of special effect makeup transfer while shortening the running time of the model by two seconds. The model we obtain currently still has the drawback of low generalization capabilities. Future efforts will see us continue to learn and improve.

References

- Chang, H.; Lu, J.; Yu, F.; and Finkelstein, A. 2018. Pairedcyclegan: Asymmetric style transfer for applying and removing makeup. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 40–48.
- Chen, H.-J.; Hui, K.-M.; Wang, S.-Y.; Tsao, L.-W.; Shuai, H.-H.; and Cheng, W.-H. 2019. Beautyglow: On-demand makeup transfer framework with reversible generative network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10042–10050.
- Deng, H.; Han, C.; Cai, H.; Han, G.; and He, S. 2021. Spatially-invariant style-codes controlled makeup transfer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6549–6557.
- Gholami, A.; Kim, S.; Dong, Z.; Yao, Z.; Mahoney, M. W.; and Keutzer, K. 2021. A survey of quantization methods for efficient neural network inference. *arXiv preprint arXiv:2103.13630*.
- Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2020. Generative adversarial networks. *Communications of the ACM*, 63(11): 139–144.
- Gray, R. M.; and Neuhoff, D. L. 1998. Quantization. *IEEE transactions on information theory*, 44(6): 2325–2383.
- Gu, Q.; Wang, G.; Chiu, M. T.; Tai, Y.-W.; and Tang, C.-K. 2019. Ladrn: Local adversarial disentangling network for facial makeup and de-makeup. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 10481–10490.
- Hasan, M. K.; Alam, M. A.; Elahi, M. T. E.; Roy, S.; and Mart, R. 2021. DRNet: Segmentation and localization of optic disc and fovea from diabetic retinopathy image. *Artificial Intelligence in Medicine*, 111: 102001.
- He, F.; Bai, K.; Zong, Y.; Zhou, Y.; Jing, Y.; Wu, G.; and Wang, C. 2022. Makeup transfer: A review. *IET Computer Vision*.
- He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.
- Isola, P.; Zhu, J.-Y.; Zhou, T.; and Efros, A. A. 2017. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1125–1134.
- Jiang, W.; Liu, S.; Gao, C.; Cao, J.; He, R.; Feng, J.; and Yan, S. 2020. Psgan: Pose and expression robust spatial-aware gan for customizable makeup transfer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5194–5202.
- Kingma, D. P.; and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Kingma, D. P.; and Dhariwal, P. 2018. Glow: Generative flow with invertible 1x1 convolutions. *Advances in neural information processing systems*, 31.
- Kips, R.; Gori, P.; Perrot, M.; and Bloch, I. 2020. Ca-gan: Weakly supervised color aware gan for controllable makeup transfer. In *European Conference on Computer Vision*, 280–296. Springer.
- Li, C.; and Wand, M. 2016. Precomputed real-time texture synthesis with markovian generative adversarial networks. In *European conference on computer vision*, 702–716. Springer.
- Li, T.; Qian, R.; Dong, C.; Liu, S.; Yan, Q.; Zhu, W.; and Lin, L. 2018. Beautygan: Instance-level facial makeup transfer with deep generative adversarial network. In *Proceedings of the 26th ACM international conference on Multimedia*, 645–653.
- Li, Y.; Gong, R.; Tan, X.; Yang, Y.; Hu, P.; Zhang, Q.; Yu, F.; Wang, W.; and Gu, S. 2021. Brecq: Pushing the limit of post-training quantization by block reconstruction. *arXiv preprint arXiv:2102.05426*.
- Liu, Q.; Zhou, H.; Xu, Q.; Liu, X.; and Wang, Y. 2020. PS-GAN: A generative adversarial network for remote sensing image pan-sharpening. *IEEE Transactions on Geoscience and Remote Sensing*, 59(12): 10227–10242.
- Nguyen, T.; Tran, A. T.; and Hoai, M. 2021. Lipstick ain’t enough: beyond color matching for in-the-wild makeup transfer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 13305–13314.
- Oktay, O.; Schlemper, J.; Folgoc, L. L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N. Y.; Kainz, B.; et al. 2018. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*.
- Paolacci, G.; Chandler, J.; and Ipeirotis, P. G. 2010. Running experiments on amazon mechanical turk. *Judgment and Decision making*, 5(5): 411–419.
- Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32.
- Polino, A.; Pascanu, R.; and Alistarh, D. 2018. Model compression via distillation and quantization. *arXiv preprint arXiv:1802.05668*.
- Simonyan, K.; and Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Taigman, Y.; Polyak, A.; and Wolf, L. 2016. Unsupervised cross-domain image generation. *arXiv preprint arXiv:1611.02200*.
- Zhu, J.-Y.; Park, T.; Isola, P.; and Efros, A. A. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, 2223–2232.