Face Images to Sketch Image Generation Using CycleGan

Abil Varkichan Jose, Aravindh Gopalsamy
axv5324@mavs.uta.edu, axg8538@mavs.uta.edu

Abstract

This project focuses on using CycleGAN to generate artistic sketches from face images. The motivation behind this work stems from the growing interest in leveraging generative adversarial networks (GANs) for image-to-image translation tasks. The primary objective is to explore the potential of CycleGAN in transforming realistic face images into expressive and artistic sketches, thereby enhancing the creative possibilities in digital art and design. The project involves the utilization of a dataset consisting of paired face images and corresponding sketches, enabling the training of the CycleGAN model to learn the mapping between the two domains. By incorporating cycle-consistency and adversarial loss functions, the model aims to capture the essential characteristics of facial features and translate them into visually appealing sketches while preserving the identity and structural details of the original images. Through comparative analysis with existing approaches, the project seeks to evaluate the effectiveness of the proposed method in generating high-quality artistic sketches from face images. The analysis will focus on assessing the visual fidelity, artistic expression, and identity preservation aspects of the generated sketches, thereby providing insights into the strengths and potential areas for improvement of the proposed approach. The findings of this project are expected to contribute to the advancement of image-to-image translation techniques, particularly in the domain of artistic content generation. Additionally, the project aims to shed light on the creative and artistic applications of GAN-based models in transforming visual content, with implications for digital art, design, and creative expression.

Introduction

In the realm of computer vision and image processing, the task of image-to-image translation has garnered substantial attention due to its potential applications in various domains, including art, entertainment, and biometric recognition. One particularly intriguing application within this domain is the generation of artistic sketches from real face images, a task that presents unique challenges related to preserving facial features, capturing expressive details, and ensuring the fidelity of the transformed images.

This paper presents an innovative approach to address the aforementioned challenge through the utilization of CycleGAN, a type of generative adversarial network (GAN) known for its proficiency in unsupervised image translation tasks. The proposed CycleGAN model is tailored to facilitate the seamless transformation of facial photographs into compelling and expressive sketches, thereby bridging the gap between photographic realism and artistic interpretation.

The significance of this endeavor lies in its potential to empower digital artists, photographers, and creators with a tool for effortlessly translating real-world facial images into captivating sketches, thereby expanding the creative possibilities in digital artistry. Moreover, the implications extend to domains such as law enforcement and forensic investigations, where sketch-based facial recognition systems could benefit from the generation of high-quality artistic sketches from photographic sources.

By leveraging the inherent capabilities of CycleGAN and custom loss functions such as cycle consistency loss and identity loss, the proposed approach aims to not only achieve faithful image translation but also imbue the generated sketches with artistic flair and expressive depth. The introduction of this paper sets the stage for a comprehensive exploration of the technical underpinnings, experimental evaluations, and insightful analyses that underpin the development and validation of the CycleGAN model for artistic sketch generation from face images.

Dataset Description

The dataset used in this study comprises the CUHK Face Sketch Database (CUFS) and the CUSK Dataset. The CUFS dataset is tailored for research on face sketch synthesis and recognition, encompassing 188 faces from the Chinese University of Hong Kong (CUHK) student database, 123 faces from the AR database, and 295 faces from the XM2VTS database, resulting in a total of 606 faces. Each face in the dataset is accompanied by a sketch drawn by an artist based on a photo taken in a frontal pose, under normal lighting conditions, and with a neutral expression.This augmentation aims to improve facial features and dataset diversity. The CUSK Dataset, containing 5000+ pairs of data, is incorporated to enhance the model's accuracy and performance. This dataset will be utilized for training the model, contributing to improved accuracy and performance. Model aims to capture a wide range of facial characteristics and artistic styles, ultimately producing a high-quality and expressive sketch-to-photo translation.

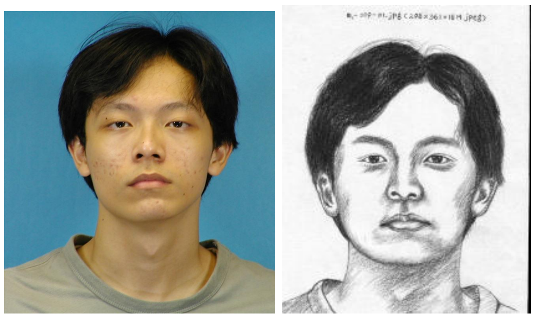


Fig 1. A sample face photo and the corresponding artistic sketch drawn from our dataset

The face photographs having the neutral expression and proper lighting and uniform background help in generating sketches of high quality. The pivotal factor in choosing this dataset was the uniformity of the images so as it can be utilized to limit preprocessing efforts to standardize the input dataset and thus aid the proposed Gan Model to generate artistic images.

Project Description

1. Description

The project aims to implement a CycleGAN model for the transformation of facial photographs into expressive and artistic sketches, and vice versa. This endeavor is rooted in the broader context of generative adversarial networks (GANs) and their application in image-to-image translation tasks. The primary objective is to explore the potential of CycleGAN in enabling accurate and reversible image translation between the photo and sketch domains, with a specific emphasis on preserving facial features and capturing expressive details.

Generative Adversarial Networks (GANs) have emerged as a powerful framework for generating realistic synthetic data, particularly in the domain - computer vision. GANs consist of two neural networks, the generator and the discriminator, which are trained simultaneously in a competitive manner. The generator aims to produce synthetic data that is indistinguishable from real data, while the discriminator seeks to differentiate between real and generated data. This adversarial training process drives the generator to continually improve its ability to produce high-quality synthetic samples.In the context of image-to-image translation, GANs have been leveraged to learn the mapping between two different visual domains, such as photographs and sketches. This involves training the GAN to transform images from one domain to another in a manner that preserves essential visual characteristics and semantic information. However, traditional GANs are designed for paired data, where corresponding images in the two domains are explicitly aligned and paired. This constraint limits their applicability to scenarios where paired data is not readily available.

CycleGAN, introduced by Jun-Yan Zhu et al., addresses the limitation of paired data by enabling unpaired image-to-image translation. It achieves this by introducing a cycle-consistency loss, which enforces the property that an image translated from domain A to domain B and then back to domain A should closely resemble the original image from domain A. This property allows CycleGAN to learn the mapping between two domains using unpaired data, making it well-suited for a wide range of image translation tasks. The implementation of CycleGAN for transforming facial photographs into artistic sketches involves training the model on the CUHK Face Sketch Database (CUFS) and the CUSK Dataset. The CUHK Face Sketch Database contains 188 faces from the Chinese University of Hong Kong (CUHK) student database, while the CUSK Dataset comprises 123 faces from the AR database, 123 faces from the XM2VTS database, and 295 faces from the CUSK database, totaling 606 faces. This diverse dataset provides a rich source of facial images and sketches for training the CycleGAN model.

The training process involves optimizing the CycleGAN model using adversarial loss, cycle-consistency loss, and identity loss. Adversarial loss encourages the generator to produce realistic sketches, while cycle-consistency loss ensures that the translation process is reversible and consistent. Additionally, identity loss promotes the preservation of essential facial features during the translation process. By leveraging the capabilities of CycleGAN and the rich dataset comprising facial photographs and sketches, the project aims to achieve accurate and visually appealing transformations between the photo and sketch domains. The potential applications of this work extend to digital art, creative expression, and the development of tools for artists and designers to explore new possibilities in image manipulation and artistic creation.

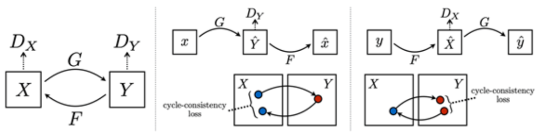


Fig.2 Training of a cycleGAN

In summary, the project represents an exploration of cutting-edge techniques in generative adversarial networks, specifically focusing on the innovative capabilities of CycleGAN for unpaired image-to-image translation. The utilization of facial photographs and sketches as the target domains underscores the potential for creative and artistic applications, with implications for digital art, design, and visual expression.

2. Project References

The primary reference 'Identity-Aware CycleGAN for Face Photo-Sketch Synthesis and Recognition' introduces an innovative approach called Identity-Aware Cycle Generative Adversarial Network (IACycleGAN) for the synthesis and recognition of face photos and sketches. The primary objective is to generate realistic facial photos from sketches and vice versa, while also preserving the identity of the individuals in the synthesized images. This is achieved by integrating a recognition model into the image generation process to improve the quality and identity preservation of the synthesized images.

The implementation of the proposed method involves several key components and techniques. Firstly, a feedback training method is designed to enhance the deep recognition model by integrating feedback from the synthesis network. Various loss functions, including cycle-consistency loss, identity perception loss, and pixel-wise consistency loss, are used to train the synthesis network and ensure the preservation of identity information. The training process includes training the synthesis networks from scratch using instance normalization and the Adam optimizer. Data augmentation techniques such as horizontal image flipping are employed, and the learning rate is decayed linearly over epochs. Additionally, the real images are fine-tuned to be closer to the corresponding fake samples from the same person and far away from images of different people, involving retraining with a triplet-loss to improve recognition performance. The recognition networks are trained using momentum, weight decay, and step learning policies, with hard negative mining and score fusion techniques used to improve recognition rates.

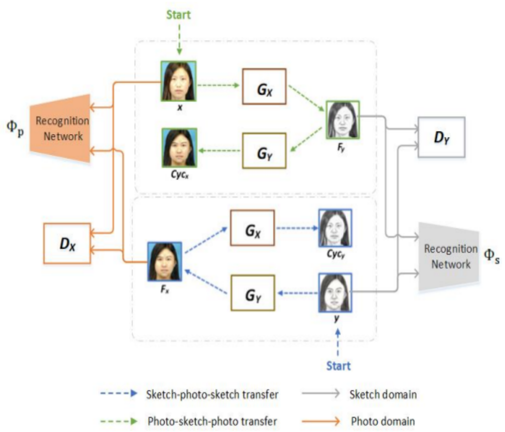


Fig.3 Framework of IACycleGAN

The second reference 'Semi-supervised Cycle-GAN for face photo-sketch translation in the wild' provides a broader perspective to implement CycleGAN in general. The paper discusses the significant advancements in face photo-sketch translation achieved through a semi-supervised learning framework based on Cycle-GAN. The authors introduce a pseudo-sketch feature (PSF) that enables the training of Gp2s using a small reference set of photo-sketch pairs, along with a large face photo dataset without ground-truth sketches. This innovative approach allows the networks to generalize well to face photos in real-world scenarios.

Furthermore, the paper presents a self-supervised approach to train the sketch-to-photo generator Gs2p without using real sketches through cycle-consistency. The authors address the challenge of invisible steganography affecting cycle-consistency loss by employing a simple noise-injection strategy. This strategy disrupts invisible steganography and significantly improves the training of Gs2p. The authors observe that, despite the noisy inputs during training, Gs2p can effectively handle clean sketches during testing, demonstrating the effectiveness of the noise-injection strategy.

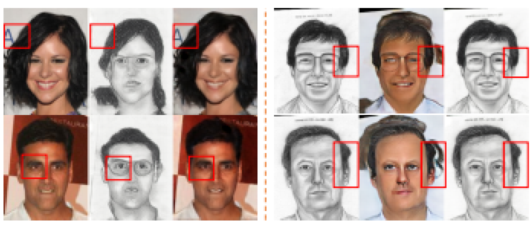


Fig.4 Illustration from source while training unpaired data

The main contributions of the paper are threefold. Firstly, the proposal of a semi-supervised learning framework, named Semi-Cycle-GAN, for face photo-sketch translation. Secondly, the introduction of the PSF, which allows training Gp2s using a small reference set of photo-sketch pairs and a large face photo dataset without ground-truth sketches, enables the networks to generalize well to face photos in real-world scenarios. Lastly, the introduction of a self-supervised approach to train the sketch-to-photo generator Gs2p without using real sketches through cycle-consistency, along with the innovative noise-injection strategy to address invisible steganography.

The authors extend their preliminary work in five aspects. They combine the previously proposed semi-supervised learning framework with cycle-consistency to conduct both photo-to-sketch and sketch-to-photo translations. Additionally, they address the challenge of invisible steganography affecting cycle-consistency loss and introduce a Gram matrix loss based on PSF, providing second-order style supervision. The paper also provides comprehensive comparisons with recently proposed methods and adopts recent perceptual-oriented metrics for performance evaluation, demonstrating the extended framework's superior performance.

Overall, the paper presents a comprehensive and innovative approach to face photo-sketch translation, addressing key challenges and significantly improving the performance of the semi-supervised learning framework.

3. Difference between project approach and references

The major difference for this project is itself the implementation strategy of the CycleGAN. The project has exclusively used upsampling and downsampling layers for this image translation tasks whereas the references uses Identity-Aware paired data translation for CycleGAN and the using unpaired data without sketches and utilizing cycle consistency loss for better training. Also our specific case is face to sketch translation. Due to insufficient data for the same , we have a limited face-sketch combination for training, extensive and efficient preprocessing steps were adopted for the project implementation. Various tryouts were made in choosing the resultant optimizers and hyperparameters for generating almost perfect face sketches. Choosing the number of layers and strides also followed the same format.

Fig.5 Samples of sketches after preprocessing (random_jitter, standardization,crop)

The importance of upsampling and downsampling layers are such that In CycleGAN, upsampling and downsampling layers are fundamental components used in the generator and discriminator networks for image translation tasks. Upsampling layers, also known as transpose convolution or deconvolution layers, are used in the generator network to increase the spatial resolution of the feature maps. These layers help in transforming low-resolution feature maps into higher-resolution feature maps, allowing the generator to create more detailed and realistic images. Upsampling layers are typically followed by convolutional layers to further process the upsampled feature maps. On the other hand, downsampling layers, often implemented as max-pooling or convolutional layers with a stride greater than 1, are used in both the generator and discriminator networks to reduce the spatial dimensions of the feature maps. Downsampling layers help in capturing the essential features of the input images by reducing the spatial resolution, which is particularly useful for learning abstract representations and reducing computational complexity.

In the project implementation of the CycleGAN, these upsampling and downsampling layers play a crucial role in the generator's ability to transform images between domains and in the discriminator's capacity to distinguish between real and generated images. By strategically using these layers, CycleGAN can effectively learn the mapping between different domains and generate high-quality translated images.

The project uses Cycle-Consistency loss from one of the references in our implementation for CycleGAN. Along with Adversarial Loss, CycleGAN uses cycle-consistency loss to enable training without paired images and this additional loss help the model to minimize reconstruction loss $F(G(x)) \approx X$ and $G(F(Y)) \approx Y$

$$\mathcal{L}(G,F,D_X,D_Y)=\mathcal{L}_{GAN}(G,D_Y,X,Y)+\mathcal{L}_{GAN}(F,D_X,Y,X)+\lambda\mathcal{L}_{cyc}(G,F)$$

Fig.6 Combination of all the losses in CycleGAN

4. Difference between project performance and references

For the generation tasks, we have improved results for face-sketch generation by optimizing the number of epochs for training and providing such a hyperparameter tuning to optimally result in artistic sketches.

Since accuracy is not typically used as a performance metric in the same way it is used in classification tasks. Instead, the quality of the generated images is often assessed visually or using metrics such as the Inception Score. In that perspective, we are almost getting 80 percent matching scores for images using external tools.

Also the references don't exactly match the implementation done in the project, so it can be inferred as CyclicGAN and the other optimized method being referred to implement the project for tailoring the output in the maximized way.

Fig.7 Implementation of one of the references using a different method(Unpaired pix-2-pix)

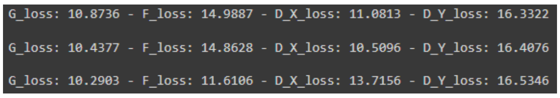
Analysis

1.Works done and what is working well

The CycleGAN implementation for face-sketch conversion demonstrates several key strengths that contribute to its effectiveness in generating high-quality sketch outputs from face photographs. The model excels at unpaired image-to-image translation, enabling it to learn the mapping between face photographs and sketches without explicitly paired data. This flexibility allows the model to generalize well to diverse facial features and artistic styles present in the input datasets.

Incorporating cycle-consistency loss ensures that the translated images maintain a consistent visual structure and content. This property is crucial for face-sketch conversion, as it helps preserve facial features, expressions, and identity across the translation process. The model effectively preserves essential facial features during the conversion process through the combination of adversarial loss and cycle-consistency loss, which collectively encourage the preservation of semantic information and visual characteristics specific to facial images.

Along with this, a critical measure is the loss graphs.



G_loss: 10.8736 - F_loss: 14.9887 - D_X_loss: 11.0813 - D_Y_loss: 16.3322

G_loss: 10.4377 - F_loss: 14.8628 - D_X_loss: 10.5096 - D_Y_loss: 16.4076

G_loss: 10.2903 - F_loss: 11.6106 - D_X_loss: 13.7156 - D_Y_loss: 16.5346

Fig.8 Loss value comparison as epochs proceed

Initially, the adversarial loss for the generators may be high as the generators start producing poor-quality translations, leading to easy discrimination by the discriminators. As training progresses, the adversarial loss for the generators should decrease, indicating that the generators are learning to produce more realistic and convincing translations. At the beginning of training, the adversarial loss for the discriminators may be high as they struggle to differentiate between real and generated images. As training advances, the adversarial loss for the discriminators should decrease, reflecting their improved

ability to distinguish between real and generated images.If identity loss is incorporated, its value should stabilize at a low level as training proceeds. The identity loss encourages the generators to preserve specific attributes of the input images. As the generators become proficient at preserving these attributes, the identity loss should remain low and relatively constant. The cycle-consistency loss should ideally decrease as training progresses. Initially, the generators may struggle to produce consistent and reversible translations, leading to a higher cycle-consistency loss. As the generators improve their translation capabilities, the cycle-consistency loss should decrease, indicating that the translated images are becoming more consistent and reversible.
Overall, the ideal behavior of these loss values as the epoch proceeds involves a decrease in adversarial and cycle-consistency losses, indicating improved translation quality and consistency. The identity loss should stabilize at a low level, reflecting the preservation of specific attributes during the translation process. These trends collectively signify the progressive improvement of the CycleGAN model in generating high-quality and consistent translations as training advances through the epochs.

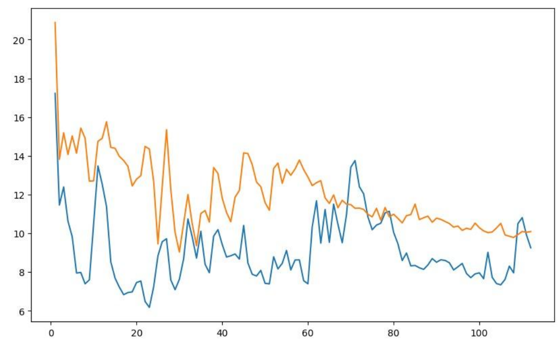


Fig.9 Loss value comparison graph as epochs proceed(orange-discriminator, blue-generator)

Furthermore, the model's ability to capture and transfer artistic styles from face photographs to sketches contributes to the generation of expressive and visually appealing outputs. This is particularly important for artistic sketch generation, where the preservation of style and expressive details is paramount. The CycleGAN implementation demonstrates the generation of diverse and realistic sketches from face photographs, capturing a wide range of artistic variations and styles. This diversity enhances the creative potential of the model and enables the production of varied and expressive sketch outputs.The utilization of datasets such as the CUHK Face Sketch Database (CUFS) and the CUSK Dataset provides the model with a rich and diverse set of facial images and sketches for training.

2.What could have been done better?

There is no doubt that cooperation with other data sources and different data types will significantly increase the robustness and versatility of the CyclicGAN application. By adding multiple facial features and images, the image can better capture facial features, faces and artwork. This collaborative effort will not only improve the quality of the training data but will also help reduce bias and improve the applicability of the model across sectors. Collaborations can include collaborations with institutions, researchers, or arts groups to collect and share data, creating an inclusive and comprehensive approach to design.

3. What is left for future work?

Future work may involve diversifying the data sets, fine-tuning the hyperparameters, and exploring new analytical metrics for better generalization. Conducting user research, optimizing real-time applications, and enabling design processes are essential. Evaluating conditional generation, enhancing opponent training, and ethical considerations of designs are essential for ongoing improvement. Application strategies, such as integrating creative tools, also warrant exploration to increase accessibility and usability, thus contributing to a growing image-to-image translation from face painting to artistic content generation.

In a broader perspective, the CyclicGAN can be tweaked for an artistic image generation translating full size photographs to sketches. The current application of faces can be expanded to tailored use cases.

Conclusion:

This model work thoroughly investigates the ability of CycleGAN to convert facial images into expressive images, and demonstrates its effectiveness in artistic content production. Using non-duplicate data and cycle-consistency loss, the model excels in preserving facial characteristics and form information. Datasets contribute to the creative use of the model. Despite industry standard results, future work includes increasing the availability of data sets, tuning hyperparameters, and searching for user-centered features This work paves the way for digital artistic applications and emphasize the importance of GAN in pushing the boundaries of image-to-image interpretation. Overall, the implementation of CycleGAN for face-sketch conversion represents a significant step forward in the realm of image translation and holds promise for applications in art, design, and digital entertainment. As the field of generative modeling continues to evolve, the insights gained from this work can contribute to the ongoing advancement of image synthesis techniques and their practical applications.

References

1. Identity-aware CycleGAN for face photo-sketch synthesis and recognition.
- Link - https://arxiv.org/pdf/2103.16019.pdf

2. Semi-supervised Cycle-GAN for face photo-sketch translation in the wild
- Link - https://arxiv.org/pdf/2307.10281.pdf

3. Landmark Assisted CycleGAN for Cartoon Face
- Link - https://arxiv.org/pdf/1907.01424.pdf

4. Dataset credits: CUHK Face Sketch Database (CUFS)
- Authors:Zhang, Wei and Wang, Xiaogan and Tang, Xiaoou

5. Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks." In Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017.

6. Ming-Yu Liu, Thomas Breuel, and Jan Kautz. "Unsupervised Image-to-Image Translation Networks." In Advances in Neural Information Processing Systems (NeurIPS), 2017.