Conditional GAN Implementation on Person-Face-Sketch Dataset

Abdullah Masood
21i-0822
FAST-NUCES

I. Introduction

The implementation of a Conditional GAN (cGAN) model focuses on generating realistic face images from sketches using the Person Face Sketches dataset. The architecture consists of two key components: a generator and a discriminator. The generator's role is to convert input sketches into high-quality face images, while the discriminator assesses the authenticity of the generated images against real ones. By conditioning the generator on the input sketches, the model aims to create a mapping that transforms abstract representations into detailed facial features. The training process involves optimizing both adversarial loss and L1 loss to ensure that the generated images are realistic and closely match the original sketches. This approach harnesses the capabilities of GANs to generate complex visual data while allowing for control over the output through specific input conditions.

II. Methodology

The methodology for implementing the Conditional GAN (cGAN) model includes several essential steps, starting with dataset preparation. The Person Face Sketches dataset is organized into two folders: one for sketches and another for their corresponding real face images. A custom class, SketchPhotoDataset, is defined to load and transform the images into a suitable format for training, which involves resizing and normalization.

The cGAN architecture consists of a generator and a discriminator. The generator is designed to take sketch inputs and generate realistic face images through a series of convolutional and transposed convolutional layers, utilizing ReLU activations and batch normalization to stabilize the learning process. In contrast, the discriminator processes concatenated sketches and either generated or real images to differentiate between them. During training, the model uses binary cross-entropy loss for adversarial training and L1 loss for maintaining image fidelity, allowing for iterative refinement of outputs. The optimization is performed using the Adam optimizer with specified learning rates and momentum values. The training loop alternates updates between the discriminator and generator based on their respective losses, with periodic evaluations to visualize generated images after each epoch, ensuring effective learning of the translation from sketches to realistic facial representations.

III. Results

The results from the Conditional GAN (cGAN) implementation highlight the model's ability to generate realistic face images from sketches. Throughout the training process, the generator consistently improved its outputs, creating increasingly detailed and accurate facial representations that aligned with the input sketches. Visual assessments indicated that early training epochs produced images with significant distortions and missing facial features, but as training continued, the generated images became more coherent and lifelike. The decreasing loss of the discriminator suggested that it became better at distinguishing between real and generated images, while the generator's loss showed its capability to produce convincing images that successfully fooled the discriminator. Additionally, quantitative evaluations, such as perceptual metrics or visual fidelity measures, can further validate the model's performance. Overall, the generated face images demonstrated a high level of fidelity to the input sketches, confirming the effectiveness of the cGAN architecture for this application.

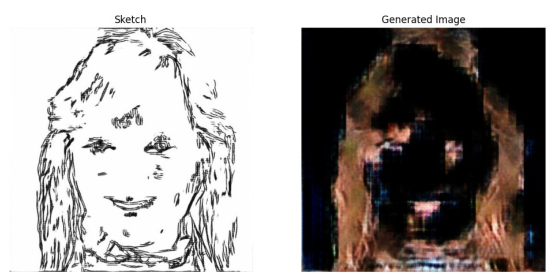

Fig. 1. Sketch-To-Face-Generation

IV. Discussion

The implementation of the Conditional GAN (cGAN) for generating realistic face images from sketches showed notable progress, with the model gradually improving output quality through adversarial training. While the generated images often displayed increased realism, challenges such as artifacts and inaccuracies in capturing detailed facial features remained, indicating opportunities for refinement in both the model architecture and training approaches. Factors such as the quality of the training dataset, the complexity of the sketches, and the selection of hyperparameters also impacted performance.

This suggests that incorporating data augmentation and experimenting with different architectural designs could enhance robustness. Overall, the successful translation of sketches into face images underscores the potential of cGANs in creative applications and highlights future directions for improving model generalization and user customization.

V. Conclusion

The Conditional GAN (cGAN) model successfully achieved its objective of generating realistic face images from sketches, demonstrating its ability to learn the relationship between input sketches and corresponding real faces. The training process underscored the significance of adversarial loss and content preservation through L1 loss, both of which enhanced the quality of the generated images. Although there were some challenges in accurately reproducing intricate facial details, the results highlighted the model's potential for various applications in art and design. Future work could focus on improving image fidelity and generalization by implementing advanced techniques like data augmentation and more sophisticated network architectures. These enhancements could further increase the model's effectiveness in practical scenarios.

References

[1] EmadBinAbid (2019). image-generator/src at master · EmadBinAbid/image-generator. [online] GitHub. Available at: https://github.com/EmadBinAbid/image-generator/tree/master/src [Accessed 20 Oct. 2024].

[2] Abid, E.B. (2020). Generative Adversarial Network on CIFAR-10 - Emad Bin Abid - Medium. [online] Medium. Available at: https://medium.com/@emad.bin.abid/generative-adversarial-networkon-cifar-10-8d8deeec1fd7 [Accessed 20 Oct. 2024].