International Journal of Science and Research Archive
eISSN: 2582-8185
Cross Ref DOI: 10.30574/ijsra
Journal homepage: https://ijsra.net/

(Research Article)

Forensic sketch-to-photo transformation with improved Generative Adversarial Network (GAN)

Bhargavi Jangam, Ashwin Bharatha *, Tagore Lavishetty and M D Rehan

Department of CSE (Artificial Intelligence and Machine Learning) of ACE Engineering College, India.

International Journal of Science and Research Archive, 2025, 14(01), 1216-1220

Publication history: Received on 11 December 2024; revised on 18 January 2025; accepted on 21 January 2025

Article DOI: https://doi.org/10.30574/ijsra.2025.14.1.0204

Abstract

Forensic sketch-to-photo transformation is a critical technique in criminal investigations in which forensic artists detail their drawings based on eyewitness descriptions. However, sketch-to-photo traditional methods are not an exception to the usual differences between sketch detail and photograph style. This system aims at improving the accuracy and realism of sketch-to photo transformation through an improved Generative Adversarial Network. The system aims to leverage advanced GAN architectures to bridge the gap between sketches and photos by learning intricate mappings and stylistic nuances. The GAN model comprises two neural networks: a generator and a discriminator. The generator synthesizes photo-realistic images from forensic sketches while the discriminator evaluates the authenticity of the generated images, iteratively refining the generator's output.

Keywords: Forensic Sketch Recognition; Image-to-Image Translation; Generative Adversarial Networks; GAN-based Image Generation; Face Reconstruction; Face Synthesis; Neural Networks for Image Generation

1. Introduction

This project explores the use of Generative Adversarial Networks (GANs) to transform forensic sketches into realistic facial images, addressing the limitations of manual sketch interpretation in law enforcement and forensic science. By training a GAN on augmented sketches and corresponding real-life photos, the model generates high-fidelity facial images, improving accuracy in criminal investigations. A web-based interface allows users to upload sketches and view generated images, enhancing forensic investigations and reducing human bias, showcasing AI's potential to revolutionize forensic science.

2. Literature survey

Goodfellow et al. introduced GANs, which have become a robust framework for generating images with realism through adversarial learning. This foundation work highlighted how GANs can generate good-quality images from random noise in the process of translating forensic sketches into realistic facial images. GANs can learn complicated patterns and generate accurate images and this feature can be very effective for addressing the challenge of transforming basic sketches into detailed facial representations.

Li et al. refined GAN architectures to make them more realistic facial images generated from sketches while paying emphasis on the reproduction of facial features in detail. The approach is thus very relevant for forensic applications, where lack of detail in sketches often hampers or prevents the identification of suspects. The work, thus, demonstrates potential avenues in deep learning that can improve the sketch-to-photo translation task toward more reliable forensic application purposes.

* Corresponding author: Ashwin Bharatha

Choi et al. introduced the model known as StarGAN that could perform cross-domain image-to-image translation. Such a technique provides the versatility to translate a sketch into the relevant realistic image while adapting to diverse styles in sketching forensic work. Thus, whatever be the quality and the aesthetic presentation of forensic sketches, their reliability is assured with respect to facial representations through this adaptation process of StarGAN.

Chen et al. focused on improving the fidelity of generated images from sketches, using advanced GAN techniques to ensure high-quality outputs. Their research emphasizes the importance of generating realistic and recognizable faces from sketches which is critical in forensic investigations. By enhancing the detail and accuracy of generated faces, this approach helps bridge the gap between vague sketches and precise real-life images.

This reference of research depicts the significant advancement made in the application of GANs in forensic sketch-to-image translation. These studies depict how GANs can be used to overcome the challenges related to sketch quality, data availability, and face recognition, with transformative potential for forensic investigations.

2.1. Architecture

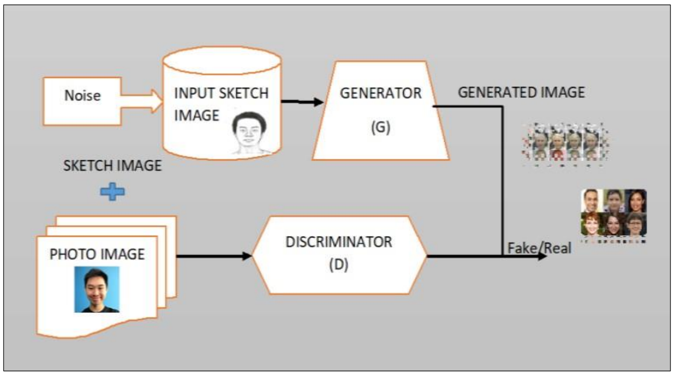

Figure 1 GAN Architecture

The proposed architecture utilizes a Generative Adversarial Network (GAN) to generate photorealistic images from forensic sketches. The system has two main components: the Generator (G) and the Discriminator (D). The process begins with the input of two elements: a forensic sketch image and a noise vector. The sketch provides the structural details, such as the outline of facial features, while the noise vector adds randomness, allowing the generator to produce diverse and realistic images.

The Generator (G) processes the input sketch and noise to create a photorealistic image, enhancing details like textures, skin tones, and lighting. This image is then passed to the Discriminator (D), which evaluates whether the image is real or fake based on a reference dataset of real photos. The Discriminator's feedback informs the Generator, helping it refine its output over successive iterations. Through adversarial training, the system improves, ultimately producing high-quality images that resemble real-life photographs. This architecture is particularly beneficial for forensic investigations, aiding in the identification of suspects from sketches.

3. Proposed Methodology

The methodology in this project was based on pix2pix, a type of conditional Generative Adversarial Network (cGAN) designed specifically for image-to-image translation tasks. Pix2pix is specially effective for transforming an input image into an output image where both are paired and have the same spatial dimensions, for example, turning sketches into photos.

3.1. Pix2Pix Architecture

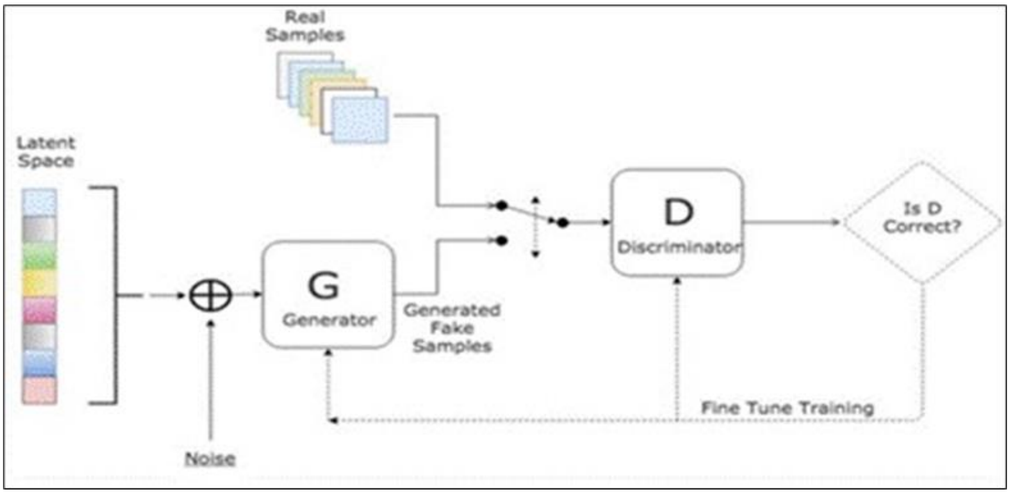

Figure 2 pix2pix Architecture

In the context of sketch-to-image translation in forensic applications, the pix2pix model uses a Generator (G) to map the input sketch directly to the corresponding photo-realistic image. The Generator is conditioned to generate high-quality images using the input sketch and random noise. The input sketch acts as the structural reference providing the facial features outline while noise introduces variability and makes the model able to produce diversified realistic images. The Discriminator (D) is designed to determine whether a generated image is real or fake. It is given a set of paired real photos and sketches. The Discriminator learns the distinction between the actual images and those generated. As such, the Discriminator informs the Generator as to how similar its output looks to the actual photos.

The Generator and the Discriminator, in fact work in opposition against each other but maintain an adversarial relationship through which the Generator is always fine-tuning itself on the negative feedback from the Discriminator so that it produces output images which seem to be much more realistic the longer the generation process is used. As the Generator improves its ability to generate realistic images, the Discriminator becomes more effective at identifying imperfections, so that the quality of the generated images improves with each iteration. This adversarial training makes pix2pix especially well-suited for applications like forensic sketch enhancement in criminal investigations, where high-fidelity facial images are critical for identification.

3.2. Training Process and Loss Function

The training of the pix2pix model takes the form of a competitive adversarial setup where there is competition between the Generator G and the Discriminator D. That is, primarily, to make the Generator capable of generating images from the sketches, but then the Discriminator learns how to distinguish the difference between a real and fake image. A random noise vector was used with a sketch to enable diversity in output. The initial images generated may not be real, but with more training, the Generator starts to generate output images that resemble real photos.

The Discriminator (D) is designed to determine the authenticity of images. It receives real images from the dataset and fake images produced by the Generator. Its function is to classify images as real or coming from the dataset or fake images generated by the Generator. The Discriminator gains more and more proficiency in discerning subtle distinctions between real images and generated ones. This process of mutual advancement for the Generator and Discriminator increases the quality of the generated image.

Generator loss includes two major parts: adversarial loss and reconstruction loss or L1 loss. The adversarial loss is a binary cross-entropy loss that encourages the Generator to create images that are able to deceive the Discriminator into classifying them as real. The reconstruction loss ensures that the generated images closely match the real images by minimizing the pixel-wise differences between the generated output and the target photo. Both losses combined teach the Generator to generate realistic images while keeping the structural integrity.

Discriminator loss focuses on differentiating between real and fake images. It is fed with both real and generated images, then labeled as either real or fake. The Discriminator will be trained in order to minimize the binary cross-entropy loss by getting it right for image classification. The Discriminator loss is very essential in refining the output of the Generator since it gives the Generator feedback on how well the generated images compare with real images and pushes the Generator to improve.

4. Results and Analysis

It showed the tremendous possibilities of the pix2pix GAN model in bringing forensic sketches close to photo-realistic images through transformation. Key facial features could be mapped and transformed by the Generator throughout its training, indicating how the models would be interpreted and rendered: textures, lighting, and different skin tones of the faces can be achieved on the generated image. Through more adversarial trainings, improved images with even more realism than the previous version are shown from the generator's end. It is the iterative process between the Generator and Discriminator that enhanced the authenticity of the facial images generated. The final output was highly faithful, with identifiable facial structures and features that very closely matched the real photos. This means the model learned well the underlying patterns in the dataset, effectively turning simple sketches into detailed and realistic images.

Even though the model was promising with some results, some limitations are present. For example, complex hairstyles and accessories, which were not well-represented in the dataset, posed difficulties for the model. More sophisticated techniques could be employed to address these challenges. Nonetheless, the model's ability to generate realistic and detailed facial features makes it a valuable tool in forensic applications. This capability could significantly enhance the identification process in criminal investigation, where sketches are often the only visual reference available. It will provide clearer and more accurate images of suspects, enhancing the reliability of forensic investigations to ultimately help law enforcement solve cases and ensure public safety.

4.1. Evaluation Metrics

In the evaluation of your pix2pix GAN model for forensic sketch-to-image translation two primary metrics, SSIM (Structural Similarity Index) and L2-norm are used to assess the quality and performance of the generated images. SSIM is a perceptual metric that measures the similarity between two images based on three key factors: luminance, texture, and structural information. Unlike PSNR, which is more pixel-wise difference-based SSIM captures the structural integrity of images, thus being more in line with human visual perception. For your project, SSIM is particularly useful since it evaluates how well the pix2pix GAN model preserves the overall structure of facial features, textures and details from the input sketch to the generated image. A higher SSIM score means that the generated image closely resembles the real image in terms of these critical aspects which provides valuable insight into the model's ability to produce realistic and high-quality images.

On the other hand, L2-norm (Euclidean Distance) offers a more quantitative measure of the pixel-wise difference between the generated image and the real photo. This metric computes the Euclidean distance between the pixel values of the generated image and the corresponding real image directly evaluating how accurately the model is reproducing the exact facial features, lighting and texture. L2-norm is useful for evaluating the pixel-level accuracy of the generated images, though it fails to capture the more perceptual qualities that SSIM does. Despite this, L2-norm is also crucial since it measures the closeness of the generated image to the real photo in terms of pixel values with great detail. Combining both these metrics gives a balanced measure of the pix2pix GAN model as they give an evaluation on both perceptual quality and pixel level accuracy in the generated images, which is quite essential for applications in forensic investigation such as suspect identification.

5. Conclusion

This project represents an example of how GANs, particularly the pix2pix model may be applied to transform a forensic sketch into a photorealistic image. This represents an important advancement in forensic research being capable of using paired sketches and real images to learn high-resolution images that can closely resemble the suspect in a particular criminal case. The Generator produces diverse and realistic images by conditioning on the input sketch and random noise, while the Discriminator evaluates the authenticity of these images. Through the adversarial training process, both networks continually improve with the Generator refining its outputs over time.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, et al.,

[2] "Generative Adversarial Nets" (Goodfellow et al., 2014)

[3] Y. Li, X. Wang, and X. Tang, "Sketch to Photo: Image Generation from Sketch Using Deep Learning" (Li et al., 2017)

[4] J. Choi, M. Choi, H. Kim, et al., "StarGAN: Unified Generative Adversarial Networks for Multi-Domain

[5] Image-to-Image Translation" (Choi et al., 2018)

[6] Y. Chen, X. Liu, L. Zhou, et al., "High-Fidelity Sketch-to-Image Translation with a Dual-Discriminator Generative Adversarial Network" (Chen et al., 2021)

[7] I. Goodfellow, J. Mirza, J. Pouget-Abadie, "Generative Adversarial Networks for Face Image Synthesis"