2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT) | 979-8-3503-3509-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCCNT56998.2023.10308219

A Conditional Generative Adversarial Network Model for Sketch-to-Image Translation

Ashish Kumar
Apex Institute of Technology (CSE)
Chandigarh University
India
ashish29300@gmail.com

Pulkit Dwivedi
Apex Institute of Technology (CSE)
Chandigarh University
India
pdwivedi1990@gmail.com

***Abstract*—The sketch-to-image translation is a challenging task that involves generating realistic photographs from hand-drawn sketches. While Generative Adversarial Networks (GANs) have achieved remarkable success in generating real images, sketch-to-image translation remains difficult due to the limited information in the input sketches. Generative modeling problems can be addressed using GANs, which are a category of artificial intelligence algorithms. The objective of a generative model is to discover the probability distribution by analyzing training samples. This estimated probability distribution is then utilized by GANs to produce additional instances. In this paper, we propose a conditional GAN-based model for sketch-to-image translation tasks. By incorporating a loss function in the training process and learning mappings from input sketches to output images, these networks can handle various problems that previously required separate loss formulations. This approach allows for a versatile problem-solving technique using the same fundamental principles. Our suggested approach involves adversarial training of a generator and a discriminator, both of which are integrated into our model. The generator network takes the sketch and a textual description as input and generates a corresponding image. The discriminator network within our model is trained to distinguish between generated and authentic images. Consequently, the generator network is trained to produce images that can deceive the discriminator network into perceiving them as genuine.**

***Index Terms*—Computer Vision, Generative Adversarial Networks (GANs), Image Synthesis, Conditional GANs, Deep Learning, Automation**

I. Introduction

Drawing a sketch is among the simplest methods to immediately picture a scene or an object. In contrast to photography, sketching doesn't need any capture equipment. Unfortunately, sketches are frequently poor and unrefined, making it difficult to create realistic images from amateur sketches. Sketches can be considered edge-bounded images with a minimum amount of pixel information that can be transformed into photorealistic images containing substantial characteristics and pixel material. The edge information of such sketches may include crucial structural details that help provide a high-quality visual representation. In general, an image can be thought of as a collection of constrained pixels that combine style and content. However, regardless of the drawing mode, sketches lack a universal style template, making it challenging to convert them into photos.

Sketch-to-image translation continues to be an occupying research topic in computer vision and graphics for many years that has acheived considerable attention. It consists of creating realistic images from given sketches, and it has numerous real-world uses in areas like computer-aided design, virtual reality, video games, design proposals, animations, and image editing software. People who lack creative talent or specialized knowledge in image synthesis can produce realistic images by using sketch-to-image translation.

The gap between the high-level semantic information in the sketches and the lower-level visual characteristics in the target pictures makes the task of translating sketches into images particularly difficult. Furthermore, sketches frequently miss key visual clues seen in the target images and are left unfinished. Researchers have proposed a number of methods based on deep learning techniques to address these issues, including conditional models, autoencoders, and GAN [1] models. GANs [1] have been frequently employed in image generation tasks, such as Sketch to Image Generation, because of developments in deep learning. Yet, because of the complexities of the image generation process, producing high-quality and diversified images from simple sketches is still a difficult task.

The phrase converting the given input image into a target image is used to describe issues in image processing tasks, computer vision, and graphics. Rendering techniques for a scene can encompass various representations, like RGB [2] image and semantic label map, among others. If provided with sufficient training data, the process of converting one possible representation of a scene into another is commonly referred to as ”image-to-image translation.” In particular, the quick growth of GANs [1] has substantially aided the research in this field, improving both the quality and efficiency of this work. Deep learning models and similar network frameworks are the main subjects for research. A GAN [1] considers the training process as a competition among the generator and discriminator. The discriminator's objective is to discover whether the target picture is real or false, and the generator's goal is to produce feasible visuals that the discriminator will mistake for real.

The primary objective of sketch-to-image translation is related to discovering mapping starting from sketch to colored images domain. It is a unique kind of I2I problem, but it frequently encounters information inconsistency among the source domain and target domain as well as a wide domain gap. After the release of the GANs [1], image-to-image translation has gained popularity. Many studies have been conducted on supervised and unsupervised GAN-based approaches. For this study, we use Conditional GANs (cGANs) [3]. The ability to conditionally create an output picture based on input picture makes cGANs [3] excellent for I2I conversion problems. Conditional GANs are used in different areas which include translating from maps to aerial photographs and vice versa, mapping from cityscapes to photographs, mapping from daylight photos to night, and mapping from input sketches to colored images (one that we are going to build).

II. Related Work

The problem of sketch-to-image translation has drawn a lot of attention recently, and various strategies have been put forth to handle this difficult process. The edge-based image synthesis method, which creates images from edges using a parametric texture model, is one of the oldest techniques. Although various different techniques have been suggested to enhance this approach's performance, it has only had a limited amount of success in producing realistic images. By developing the network on a large dataset consisting of sketch-image pairs, deep neural networks, such as CNNs [4] [18] [19], are used to create images from sketches. By first learning a latent representation of the sketches and then decoding the representation to produce images, variational autoencoders (VAEs) have also been used to create images from sketches.

A. Image to Image Translation

The ”image-to-image translation” seeks for discovering relation between the input and output photo with the help of a training data set of matched picture pairs. Image-to-image translation is process of transforming an input photo into an output photo while preserving certain properties of the input image. This can be applied on various tasks such as style transfer, colorization, super-resolution, and more.

The field of image-to-image translation relates to Hertzmann et al.'s [5] Image Analogies, where a non-parametric texturing model was used on the single training pair of input-output image. The framework is divided into two phases. An image pair is presented as ”training data”, one of which appears to be a ”filtered” version of the other in designing phase. In the application phase, the learned filter is used upon a new target image to generate an ”analogous” filtered output.

Resales et al. [6] proposed a probabilistic inference and learning system using a Bayesian technique that is computationally efficient for predicting the most probable output image and understanding the rendering style.

The ability to produce natural, genuine images has been demonstrated by generative adversarial networks (GANs). GANs [1] employ a discriminator to separate realistic photos from unreal photos. This forces the generator to create clearer pictures. Isola et al.'s [7] ”pix2pix” work illustrates a simple method for translating one image to another using conditional GANs. Using paired input-output samples, it creates realistic visuals using conditional adversarial networks. In addition to learning the input image's mapping relation to the output image, the proposed scheme also feeds back this relation using a loss function. As a consequence, it is possible to tackle problems that previously required the employment of entirely different loss formulas by using the same general technique.

Zhu et al. [8] suggested a network architecture known as Cycle GANs that can transform input photos into target photos to overcome the issue of the lack of paired training data in some situations. These suggested solutions, however, are unable to regulate how the facial features change while translating from one image to another.

B. Sketch to Image Translation

A sketch-based image synthesis technique was introduced by Sangkloy et al. [9] that makes use of a Siamese GAN architecture to produce realistic images. This technique uses a perceptual loss function that takes into account how similar the input photo and generated photo appear to the eye. The network can learn across the sketch and image domains thanks to the adaptation, and it can even learn a shared feature space between the two. They get 37.1 accuracy for K = 1 recall using this architecture.

The GAN model, developed by Chen et al. [10], proposes a revolutionary approach that synthesizes realistic images from various categories. It demonstrates an entirely automatic data augmentation method for sketches and demonstrates how the supplemented data is beneficial to the work at hand. Another interactive GAN-based technique for sketch-to-image translation was introduced by Ghosh et al. [11] to make it simple for beginners to create images of basic things. The network contains a feedback loop where the user can alter the drawing in accordance with the network's suggestions, and the network can more accurately synthesize any images the user may be thinking about. They offer a gating-based strategy for class conditioning, which enables the generation of separate classes without feature mixing from a single generator network in order to employ a single model for a wide range of object classes.

Lately, sketch-to-image translation has used attention methods to improve the output images' quality. In order to preserve local details and textures, attention techniques enable the network to concentrate on particular areas of the image when producing the output. Because normalization layers have a tendency to ”wash away” semantic information, it is not ideal to feed the semantic layout directly into the deep network. Park et al. [12] introduced a straightforward but efficient layer for creating realistic images from the input semantic layout. The spatially adaptive normalization (SPADE) [12] technique adapts the input characteristics based on the semantic labels of the sketches by using an attention-based normalization layer. Deep convolutional networks have been employed in several recent efforts to produce realistic pictures.

Sangkloy et al. [13] suggested an architecture that is dependent on sketch boundaries and sparse color strokes to create real images. They showed off a technique for sketch-based image synthesis that enables users to annotate sketches with their favorite color choices for items. Because of the feed-forward network, users may instantly see their edits' results. Jo et al. [14] introduced a revolutionary image editing system that creates images according to the user's inputs. They trained the network using an extra style loss, allowing them to provide realistic results despite the removal of significant chunks of the image. With the use of simple user inputs, high-quality synthetic images can be produced utilizing this network design, SC-FEGAN. Devis et al. [15] suggested a transfer learning strategy to learn the sketch's features. For training, they made use of the 75,471 sketches in the Sketchy Database, which is divided into 125 categories. The cosine similarity function is used to determine how similar the drawing and picture databases are using a pre-trained VGG-19 network [20] [21]. It recognizes a set of related photos that correlate to the drawing based on the similarity function's output. Galea et al. [16]suggested a deep CNN in order to identify a subject in photo by referencing it to face images. This network was generated and trained by using transfer learning on a pre-trained model for face photo recognition. To add the existing training data, both photographs, and sketches are synthesized using a 3D morphable model. Osahor et al. [17] introduced an approach that harnesses the power of generative adversarial networks to generate multiple synthetic images with diverse features, such as gender and hair color, from a single sketch. They utilized several components within their framework, including an identity-preserving network, a hybrid discriminator that performs attribute classification, and a network that reduces perceptual dissimilarity. Kazemi et al. [22] presented a novel framework that represents a conditional variation of CycleGAN, where facial features are used as the condition. Their network is trained without a pre-defined set of aligned pair of face and sketch and is designed to incorporate specific facial characteristics into the synthetic photos. Although numerous methods have been put forth for translating sketches into photographs, the generated photographs still need to be of higher quality, particularly in terms of keeping regional details and textures.

III. Proposed Methodology

GANs [1] are generative models that establish a relationship between noise vector c and generated output image b, which is described by a mapping function G: c → b. While, Conditional GANs learn a mapping function G: {a, c} → b that maps the input image a and noise vector c to an output image b. During the training process, generator G produce output pictures that are indistinguishable from "real" pictures by an adversarially trained discriminator, D. The discriminator detects the generator's "fakes" as accurately as possible.

A. Objective

The objective of a conditional GAN can be expressed as:

$$\begin{aligned} L_{cGAN}(G,D) =& E_{a,b}[logD(a,b)]+ \\ & E_{a,c}[log(1-D(a,G(a,c))] \end{aligned} \quad (1)$$

where G tries to minimize this objective function while D aims to maximize it, i.e.

$$G^{*} = argmin_{G}max_{D}L_{cGAN}(G,D).$$

To assess the significance of conditioning the discriminator on input a, we compare it with an unconditional version in which the discriminator does not have access to a:

$$\begin{aligned} L_{GAN}(G,D) =& E_{b}[logD(b)]+ \\ & E_{a,c}[log(1-D(G(a,c))]. \end{aligned} \quad (2)$$

Previous studies have shown that combining the GAN objective with a traditional loss, such as L2 distance, can be advantageous. In this approach, the generator aims not only to fool the discriminator but also to produce outputs that are similar to the ground truth in an L2 sense. However, this does not affect the discriminator's task. We also explore this approach and utilize L1 distance instead of L2 because L1 promotes less blurring in the generated images:

$$L_{L1}(G) = E_{a,b,c}[||b-G(a,c)||_{1}]. \quad (3)$$

The final objective is:

$$G^{*} = argmin_{G}max_{D}L_{cGAN}(G,D) + \lambda L_{L1}(G). \quad (4)$$

Even without c, the network could still learn to map a to b, but it would produce deterministic outputs that are limited to a delta function distribution. To address this issue, previous conditional GAN models have included Gaussian noise c as an additional input to the generator, along with a. This allows the generator to produce a diverse set of outputs that are not solely determined by the input a and can match a variety of distributions.

B. Network Architecture

We adopted the generator and discriminator model architectures from Radford et al.'s [24] work. Figure 1 shows the complete conditional GAN network architecture. The convolution-BatchNorm-ReLu type of module is used by both the generator and discriminator. Key characteristics of the architecture are covered here, along with details of them.

1) Generator with skips: An encoder-decoder network has been widely employed in the past to solve issues in this field. In an encoder-decoder network with a bottleneck layer, the input data is first transmitted through a succession of convolutional layers that gradually downsample the input. Once it reaches the bottleneck layer, the representation is typically at its lowest dimensionality. Then, the data is transmitted through a succession of deconvolutional layers that gradually upsample the representation until it reaches the original resolution of the input. Each layer in the network must be traversed by the information flow during both the forward and backward passes of training.

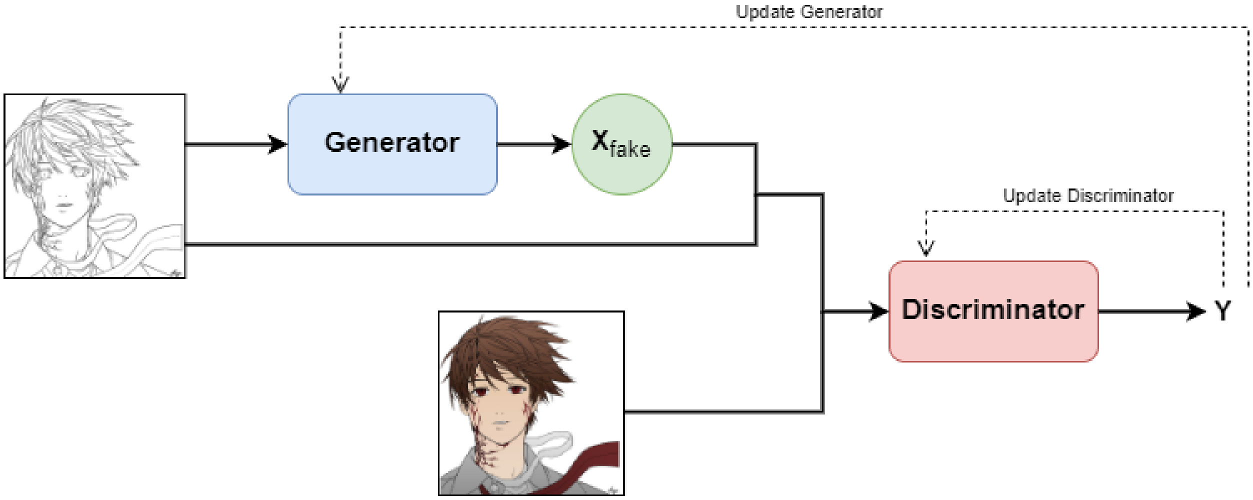

Fig. 1. Conditional GAN Network Architecture

Skip connections are a way to provide shortcuts for the information flow in a neural network, allowing it to bypass certain layers and facilitate the training process. In the generator architecture mentioned in the passage, skip connections are added to provide a way for the information to flow around the bottleneck layer. Specifically, for total n number of layers: for each layer i, a skip connection is added to its corresponding layer n-i, which concatenates all layer i channels with layer n-i channels. This is roughly following the shape of a ”U-Net” [25], a commonly used neural network for image segmentation.

2) Markovian discriminator: L2 and L1 losses are commonly used to measure the difference among input and generated images. L2 loss emphasizes the importance of large errors and tends to produce blurry images, while L1 loss promotes less blurring but is less sensitive to outliers.

The approach employed here involves modelling structures having high- frequency using the GAN discriminator while enforcing low-frequency accuracy through an L1 term (4). By concentrating on the structure of individual image patches, it becomes possible to model high frequencies effectively. In order to achieve this, structure is punished only at the patch level in devised discriminator. The PatchGAN discriminator evaluates the authenticity of each patch in an N x N image, enabling the generation of the final output by applying it to the entire image and averaging the responses through convolution.

Based on findings, N can be considerably reduced from the original image size, while still achieving exceptional outcomes. A smaller PatchGAN can be utilized for images of any dimension, which would involve fewer parameters and execute more quicker. Assuming that there is relation among pixels separated by a distance greater than a patch diameter, this discriminator can model the image as a Markov random field. This correlation has previously been investigated, and models of texture and style generally rely on it.

C. Optimization and Inference

For the purpose of network optimization, we employ the conventional technique outlined in Goodfellow et al. [1]: we applied gradient descent step alternatively between discriminator D and generator G. In accordance with the GAN paper's initial advice, we train the generator G to maximize log D(x, G(x, z)) rather than minimize the log (1-D(x, G(x, z)). Furthermore, during the optimization of the discriminator D, we half the objective to make discriminator learning rate match to generator G.

During inference, we use the generator network as while training. However, our approach is different from the standard technique as we dropout during testing and utilize batch normalization [26] based on test batch statistics, rather than aggregate training batch statistics. Depending on the experiment, we employ batch sizes ranging from 1 to 10 in our evaluations.

IV. Experimental Results

For this study, we investigated a conditional GAN-based model for the sketch-to-image translation problem. To assess the versatility of conditional GANs, we performed experiments on various datasets suitable for tasks, including graphics-related tasks such as photo generation, as well as vision-related tasks such as translation between semantic labels and photos, trained on Cityscapes dataset [27], converting black and white images to color images, Day-to-night translation, trained on [28] and translation from sketches to images.

TABLE I
FCN-SCORES FOR VARIOUS LOSSES

Loss	Per-pixel acc.	Per-class acc.	Class IOU
L1	0.42	0.15	0.11
GAN	0.22	0.05	0.01
cGAN	0.57	0.22	0.16
L1+GAN	0.64	0.20	0.15
L1+cGAN	0.66	0.23	0.17
Ground Truth	0.80	0.26	0.21

TABLE II
FCN-SCORES FOR VARIOUS GENERATOR ARCHITECTURES

Loss	Per-pixel acc.	Per-class acc.	Class IOU
Encoder-decoder (L1 only)	0.34	0.11	0.09
Encoder-decoder (L1 and cGAN)	0.28	0.08	0.06
U-net (L1 only)	0.46	0.17	0.14
U-net (L1 and cGAN)	0.53	0.19	0.15

A. Dataset Details

The ”Anime Sketch Colorization Pair” [29] dataset available on Kaggle is applied for training and testing the cGAN model. This dataset is a collection of over 15,000 image pairs of anime sketches and their corresponding colorized versions. The dataset consists of 14,200 pairs of sketches and colored images for training purposes and 3,545 pairs of sketches and colored images for validation purposes. The dataset was created by scraping images from various anime forums and websites. Each image pair consists of a black-and-white sketch and its colorized version. This dataset can be used for a variety of purposes, including image colorization research, deep learning, and computer vision applications.

B. Evaluation Metrics Analysis

Assessing the generated images' quality is a complex and unresolved issue [30]. Conventional metrics do not consider the combined statistics of the output. To obtain a more comprehensive evaluation of the visual quality of our synthesized images, we utilize two strategies. First, we conduct perceptual studies on AMT, comparing ”real vs. fake” scenarios. For graphical problems such as colorization and photo creation, the ultimate goal is often to achieve believability for a human observer. We leverage this approach to evaluate our image colorization models. In addition to perceptual studies on AMT, we also evaluate the effectiveness of our synthesized cityscapes by testing whether commercial recognition systems can accurately identify all the objects in them, like ”inception score” [30].

C. Objective Function Analysis

To determine the crucial components of the objective function in (4), we conduct studies of ablation that isolate impact of L1 term and GAN term. Additionally, we evaluated the performance of a discriminator where condition is applied on input (cGAN,(1)) to an unconditional discriminator (GAN,(2)).

Results from L1 alone are reasonable but hazy. While the cGAN alone (with $\lambda = 0$ in (4)) produces results that are much sharper, it also introduces visual artifacts in some applications. These artifacts are diminished when both terms are added together (with $\lambda = 100$).

The FCN score method was employed on a dataset of cityscapes (Table I) to quantify certain observations: The objectives based on GAN gain higher scores, which shows that generated photos consist of more identifiable structured elements. Also, we investigate the results about discarding conditioning factor from the discriminator (referred to as GAN).

For this scenario, the focus of the loss function is solely on ensuring that the output is believable, without penalizing any discrepancies between the input and output. However, upon reviewing the results, it became apparent that this approach was not effective as the generator produced almost identical outputs regardless of the input photos. Thus, it is essential to evaluate the similarity between the input and output in such situations, and the cGAN outperforms the GAN in this aspect. Specifically, the L1 loss function is designed to penalize any dissimilarities between the ground truth outputs that correspond to the input and the generated outputs that may not match the input accurately. This approach encourages the generated output to respect the input and its associated features. As a result, the L1 loss, when used in combination with GAN, proves effective in creating precise renderings that are faithful to the input label maps. The L1+cGAN approach, which merges both techniques, produces similarly satisfactory results.

D. Generator Architecture Analysis

The U-Net architecture facilitates the rapid propagation of low-level information throughout the network. However, does this actually result in improved outcomes? In the case of generating cityscapes, Table II compares the performance of the U-Net architecture to that of an encoder-decoder architecture, which is essentially a U-Net without skip connections. Our findings show that the encoder-decoder approach fails to generate realistic visuals. On the other hand, when both the U-Net and the encoder-decoder model are trained using an L1 loss, then U-Net achieves better results, indicating that its advantages extend beyond conditional GANs.

V. Conclusion

Sketch-to-image translation using Generative Adversarial Networks (GANs) is a cutting-edge technique aimed at generating realistic images from hand-drawn sketches. GANs leverage the power of adversarial training, where a generator network learns to produce images that can deceive a discriminator network into perceiving them as authentic. This approach enables the generation of high-quality and visually coherent images, bridging the gap between sketches and photorealistic representations. The use of GANs in sketch-to-image translation holds great promise for various applications, including art, design, animation, and virtual reality.

In conclusion, our investigation into the use of conditional GANs for converting sketches into images has yielded highly promising results. Through extensive quantitative and qualitative evaluations, our proposed approach demonstrated excellent performance. By harnessing the power of conditional GANs, we successfully generated colorized images from black-and-white sketches, providing valuable insights into the inner workings of the network. The findings of this study suggest that conditional adversarial networks show promise in numerous image-to-image translation applications, particularly those that require precise and well-organized graphical outcomes. These networks can learn a loss function tailored to the specific process and available data, making them valuable in various situations.

REFERENCES

[1] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2020. Generative adversarial networks. Commun. ACM 63, 11 (November 2020), 139–144. https://doi.org/10.1145/3422622

[2] H. Altun, R. Sinekli, U. Tekbas, F. Karakaya and M. Peker, ”An efficient color detection in RGB space using hierarchical neural network structure,” 2011 International Symposium on Innovations in Intelligent Systems and Applications, Istanbul, Turkey, 2011, pp. 154-158, doi: 10.1109/INISTA.2011.5946088.

[3] M. G. Blanch, M. Mrak, A. F. Smeaton and N. E. O'Connor, ”End-to-End Conditional GAN-based Architectures for Image Colourisation,” 2019 IEEE 21st International Workshop on Multimedia Signal Processing (MMSP), Kuala Lumpur, Malaysia, 2019, pp. 1-6, doi: 10.1109/MMSP.2019.8901712.

[4] S. Albawi, T. A. Mohammed and S. Al-Zawi, ”Understanding of a convolutional neural network,” 2017 International Conference on Engineering and Technology (ICET), Antalya, Turkey, 2017, pp. 1-6, doi: 10.1109/ICEngTechnol.2017.8308186.

[5] Aaron Hertzmann, Charles E. Jacobs, Nuria Oliver, Brian Curless, and David H. Salesin. 2001. Image analogies. In Proceedings of the 28th annual conference on Computer graphics and interactive techniques (SIGGRAPH '01). Association for Computing Machinery, New York, NY, USA, 327–340. https://doi.org/10.1145/383259.383295

[6] Resales, Achan and Frey, ”Unsupervised image translation,” Proceedings Ninth IEEE International Conference on Computer Vision, Nice, France, 2003, pp. 472-478 vol.1, doi: 10.1109/ICCV.2003.1238384.

[7] P. Isola, J. -Y. Zhu, T. Zhou and A. A. Efros, ”Image-to-Image Translation with Conditional Adversarial Networks,” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 5967-5976, doi: 10.1109/CVPR.2017.632.

[8] J. -Y. Zhu, T. Park, P. Isola and A. A. Efros, ”Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks,” 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017, pp. 2242-2251, doi: 10.1109/ICCV.2017.244.

[9] Patsorn Sangkloy, Nathan Burnell, Cusuh Ham, and James Hays. 2016. The sketchy database: learning to retrieve badly drawn bunnies. ACM Trans. Graph. 35, 4, Article 119 (July 2016), 12 pages. https://doi.org/10.1145/2897824.2925954

[10] W. Chen and J. Hays, ”SketchyGAN: Towards Diverse and Realistic Sketch to Image Synthesis,” 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, pp. 9416-9425, doi: 10.1109/CVPR.2018.00981.

[11] A. Ghosh et al., ”Interactive Sketch & Fill: Multiclass Sketch-to-Image Translation,” 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), 2019, pp. 1171-1180, doi: 10.1109/ICCV.2019.00126.

[12] T. Park, M. -Y. Liu, T. -C. Wang and J. -Y. Zhu, ”Semantic Image Synthesis With Spatially-Adaptive Normalization,” 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019, pp. 2332-2341, doi: 10.1109/CVPR.2019.00244.

[13] P. Sangkloy, J. Lu, C. Fang, F. Yu and J. Hays, ”Scribbler: Controlling Deep Image Synthesis with Sketch and Color,” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 6836-6845, doi: 10.1109/CVPR.2017.723.

[14] Y. Jo and J. Park, ”SC-FEGAN: Face Editing Generative Adversarial Network With User's Sketch and Color,” 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), 2019, pp. 1745-1753, doi: 10.1109/ICCV.2019.00183.

[15] N. Devis, N. J. Pattara, S. Shoni, S. Mathew and V. A. Kumar, ”Sketch Based Image Retrieval using Transfer Learning,” 2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2019, pp. 642-646, doi: 10.1109/ICECA.2019.8822021.

[16] C. Galea and R. A. Farrugia, ”Matching Software-Generated Sketches to Face Photographs With a Very Deep CNN, Morphed Faces, and Transfer Learning,” in IEEE Transactions on Information Forensics and Security, vol. 13, no. 6, pp. 1421-1431, June 2018, doi: 10.1109/TIFS.2017.2788002.

[17] U. Osahor, H. Kazemi, A. Dabouei and N. Nasrabadi, ”Quality Guided Sketch-to-Photo Image Synthesis,” 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, 2020, pp. 3575-3584, doi: 10.1109/CVPRW50498.2020.00418.

[18] T. P. Kancharlapalli and P. Dwivedi, ”A Novel Approach for Age and Gender Detection using Deep Convolution Neural Network,” 2023 10th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2023, pp. 873-878.

[19] P. Dwivedi and B. Sharan, ”Deep Inception Based Convolutional Neural Network Model for Facial Key-Points Detection,” 2022 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), Greater Noida, India, 2022, pp. 792-799, doi: 10.1109/ICCCIS56430.2022.10037639.

[20] P. Dwivedi and A. Upadhyaya, ”A Novel Deep Learning Model for Accurate Prediction of Image Captions in Fashion Industry,” 2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2022, pp. 207-212, doi: 10.1109/Confluence52989.2022.9734171.

[21] P. Dwivedi and B. Islam, ”An Item-based Collaborative Filtering Approach for Movie Recommendation System,” 2023 10th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2023, pp. 153-158.

[22] H. Kazemi, M. Iranmanesh, A. Dabouei, S. Soleymani and N. M. Nasrabadi, ”Facial Attributes Guided Deep Sketch-to-Photo Synthesis,” 2018 IEEE Winter Applications of Computer Vision Workshops (WACVW), Lake Tahoe, NV, USA, 2018, pp. 1-8, doi: 10.1109/WACVW.2018.00006.

[23] M. Mathieu, C. Couprie, and Y. LeCun, “Deep multi-scale video prediction beyond mean square error,” arXiv.org, 26-Feb-2016. https://arxiv.org/abs/1511.05440.

[24] A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative Adversarial Networks,” arXiv.org, 07-Jan-2016. https://arxiv.org/abs/1511.06434.

[25] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” arXiv.org, 18-May-2015. https://arxiv.org/abs/1505.04597.

[26] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” arXiv.org, 02-Mar-2015. https://arxiv.org/abs/1502.03167.

[27] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, “The cityscapes dataset for Semantic Urban Scene understanding,” arXiv.org, 07-Apr-2016. https://arxiv.org/abs/1604.01685.

[28] Pierre-Yves Laffont, Zhile Ren, Xiaofeng Tao, Chao Qian, and James Hays. 2014. Transient attributes for high-level understanding and editing of outdoor scenes. ACM Trans. Graph. 33, 4, Article 149 (July 2014), 11 pages. https://doi.org/10.1145/2601097.2601101

[29] T. Kim, “Anime sketch colorization pair,” Kaggle, 14-Dec-2018. https://www.kaggle.com/datasets/ktaebum/anime-sketch-colorization-pair.

[30] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, “Improved techniques for training gans,” arXiv.org, 10-Jun-2016. https://arxiv.org/abs/1606.03498.