International Journal of Engineering Applied Sciences and Technology, 2021
Vol. 6, Issue 2, ISSN No. 2455-2143, Pages 251-257
Published Online June 2021 in IJEAST (http://www.ijeast.com)

ADVERSARIAL OPEN DOMAIN ADAPTION FRAMEWORK (AODA): SKETCH-TO-PHOTO SYNTHESIS

Amey Thakur
Department of Computer Engineering,
University of Mumbai,
Mumbai, MH, India

Mega Satish
Department of Computer Engineering,
University of Mumbai,
Mumbai, MH, India

***Abstract*— This paper aims to demonstrate the efficiency of the Adversarial Open Domain Adaption framework for sketch-to-photo synthesis. The unsupervised open domain adaption for generating realistic photos from a hand-drawn sketch is challenging as there is no such sketch of that class for training data. The absence of learning supervision and the huge domain gap between both the freehand drawing and picture domains make it hard. We present an approach that learns both sketch-to-photo and photo-to-sketch generation to synthesise the missing freehand drawings from pictures. Due to the domain gap between synthetic sketches and genuine ones, the generator trained on false drawings may produce unsatisfactory results when dealing with drawings of lacking classes. To address this problem, we offer a simple but effective open-domain sampling and optimization method that "tricks" the generator into considering false drawings as genuine. Our approach generalises the learnt sketch-to-photo and photo-to-sketch mappings from in-domain input to open-domain categories. On the Scribble and SketchyCOCO datasets, we compared our technique to the most current competing methods. For many types of open-domain drawings, our model outperforms impressive results in synthesising accurate colour, substance, and retaining the structural layout.**

Keywords— **Adversarial Open Domain Adaption (AODA), Sketch-to-Photo Synthesis, Photo-to-Sketch Synthesis, Sketches, In-domain, Open-domain Generator, Discriminator, Generative Adversarial Networks (GAN).**

I. INTRODUCTION

Generative Adversarial Networks [1] have been a recent breakthrough in the field of machine learning. When it comes to the generation of images from various inputs, GANs have been widely used for the desired outcome. Lately, new research regarding GANs allowed the generation of cartoon images from real high-quality pictures. The generator outputs cartoonized images and tries to fool the discriminator whereas the latter one tries to distinguish the real images from fake ones. The results of White Box Cartoonization using an extended GAN framework [2][3] were similar to the real photo yet it had characteristics of a cartoon image. The synthesis of cartoon images from real ones is a real challenge but a GAN framework was able to do it with real precision in less time.

Sketches are drawn by anyone and they can be anything. Sketches can describe ideas of any product or artwork. It might be used to depict people, landscapes, or create art. The popularity of artwork created by computers has risen in recent years. Artists, as well as Social Media Users, can interact with visual media and communicate their goals via a freehand sketch. Given the limitation of a sketched object, the purpose of sketch-based image synthesis is to construct some image, photorealistic or non-photorealistic. This enables non-artists to transform simple black-and-white sketches into more abstract, intricate paintings. Also, with the widespread use of touch screens, new scenarios for sketch-based applications are emerging, such as sketch-based photo editing [4], sketch-based image retrieval for 2D [5] and 3D shapes [6], and 3D modelling from sketches.

The aim of the sketch to image translation is to convert a sketch of source domain *S* to destination photo domain *P*. Generative Adversarial Networks are used for the sketch-to-photo synthesis [7] and use paired data for learning purposes. But there are some limitations to the open domain adaption [8]: the source domain and the target domain both have images that are not labelled or paired [9]. Moreover, images in the target domain are unrelated to images in the source domain, and vice versa. As a result, the data is essentially unlabeled and unpaired. Also, the freehand sketches represent a minority of the categories of photos as they need to be annotated. So some systems [10] change the sketches with extracted edges from the target photo. However edges and freehand still remain two distinct sketches as freehand sketches are more deformed. Because of this domain difference, Models trained on edge inputs frequently fail to generalise to freehand sketches. A decent sketch-based picture generator should adjust the object structure depending on the input composition as well as fill the right textures within the lines.

International Journal of Engineering Applied Sciences and Technology, 2021
Vol. 6, Issue 2, ISSN No. 2455-2143, Pages 251-257
Published Online June 2021 in IJEAST (http://www.ijeast.com)

Paired and labelled data could help with the synthesis of the sketch to photo. Recent adversarial networks have learned from unpaired sketches and pictures that were gathered separately. But this still doesn't cover all types of pictures in the open domain as many datasets don't have enough freehand drawings for the training of sketch-to-photo synthesis.

To address this difficult problem, we present an Adversarial Open Domain Adaption (AODA) [11]. This framework will, firstly, train to reconstruct missing freehand drawings and enable unsupervised open-domain adaptation. We are proposing a concurrent framework: a generating sketch-to-photo translation network and a photo-to-sketch translation network for translating open-domain pictures into drawings. We may extend the learnt correlation between in-domain hand-drawn designs and photographs to open-domain classes using the photo-to-sketch synthesis link. Yet, there is a significant domain gap between created drawings and genuine sketches, which prohibits the generalisation of the learnt correlation to real sketches as well as the synthesis of realistic pictures for open-domain classes. To reduce this gap which affects the generator and exploits the results, we propose a random-mixed sampling algorithm that takes a number of false drawings as real ones arbitrarily for all categories. With this learning strategy, our model is capable of generating a realistic photo for unlabeled sketches. The suggested AODA [11,12] is compared to existing unpaired sketch-to-image generating methods. Both qualitative and quantitative findings demonstrate that our suggested technique outperforms the competition on both seen and unseen data.

II. RELATED WORK

A. Pix2Pix –

Conditional Generative Adversarial Network (cGAN) [13] are a branch of GANs and are widely used for image-to-image translation. They have two components: a generator and a discriminator. CGANs learn the mapping from the source image to the output image, with the help of a loss function. We can apply the same basic method to others examples that would need different loss formulae. The framework has been tested on a variety of image-to-image translation [14] tasks, including translating maps to satellite pictures, black-and-white photographs to class label graphs, and product drawings to product photographs.

B. Sketch to image generation with limited data –

The generation of the sketch to photo synthesis has gone through a few phases. The first phase is to turn edges into images like pix2pix i.e. Image-to-image translation with a conditional GAN. The second phase is to turn the freehand sketches into images compared to edge to image freehand sketches that have more deformation and the connections between input and output can be loose. Some freehand sketches can be really ambiguous. Some works take class labels as inputs to specify the output still it takes a lot of human labour to draw the freehand sketches that is why we always lack nicely labelled and pure sketches data. Here we introduce a surface open-domain freehand sketch to image synthesis. We want to learn a multi-class generation with one generation and even generate output from unseen sketches in the training phase.

C. Challenges –

We might encounter several obstacles in order to tackle this problem. The first is Freehand sketches. Most of them don't look like their target photo. Also, the sketches and target photos are unpaired. The sketch data is also limited for some classes; there are only a few or no sketches. Besides, we aim to generate multiple classes of photos with only one model.

In the training phase, sketches of some categories are missing and in the inference stage, input sketches are not only from known classes but also from the classes that were missing during the training.

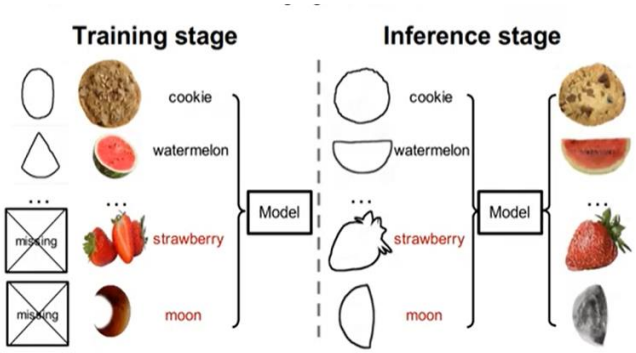


Figure 1: Training Stage and Inference Stage

D. Previous Solutions –

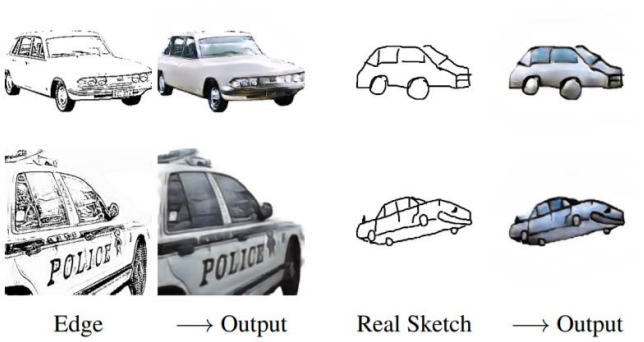
It allows a network to learn to synthesise pictures from both in-domain and open-domain classes. The difficulties in the preceding techniques can be solved in two ways. The first is to train a model using extracted edges maps, while the second is to supplement open domain classes with synthesised drawings generated by a pre-trained picture to sketch extractor.

1. Edge Maps –

An edge map is a graphic that shows where the image's edges are. With an edge detection algorithm, the edges of the image are filtered and a map is created. The edge map indicates the sharpness of the edges as well as other metrics. The picture may already have been quantized with the edge() method to produce a binary image. An efficient hashing approach

International Journal of Engineering Applied Sciences and Technology, 2021
Vol. 6, Issue 2, ISSN No. 2455-2143, Pages 251-257
Published Online June 2021 in IJEAST (http://www.ijeast.com)

associated with the internal sequencing of the edges is used to create edge maps. The anticipated time for an access operation is *O(1).*

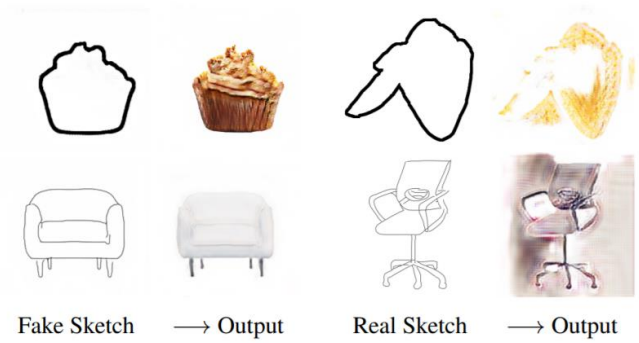


(a) Edge -> Output (b) Real Sketch -> Output

Figure 2

XDoG-extracted edges were used to train a model. The model can't fix the deformed forms of freehand doodles because it's only trained with edges. The result from drawings isn't as realistic as the pictures created by edge mapping, which are relatively realistic.

2. Synthesized Sketches –



(a) Fake Sketch -> Output (b) Real Sketch -> Output

Figure 3

Pre-extracted drawings were used to train a model. On genuine sketches, the model was unable to generalise. From the synthesised drawings, the model may create ever more realistic approaches.

III. FRAMEWORK

A. Network Structure –

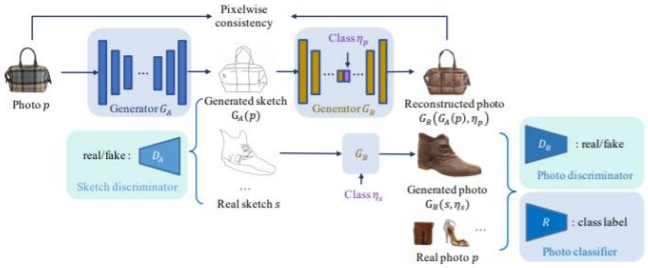


Figure 4: Network Structure

Our framework is made up of the following components, as illustrated in Figure. Two generators and two discriminators. The first generator is a Photo-to-Sketch generator i.e. G_A and the second generator is a multi-class Sketch-to-photo generator i.e. G_B which take inputs sketch and class label η_s. The first discriminator is D_A which is a sketch discriminator and the second discriminator is D_B which is a Photo discriminator. Unpaired drawing and picture data are used to train the AODA framework.

Generator G_B derives the drawing $G_A(p\ )$ from the provided picture p throughout the training phase. By passing the synthesised drawing $G_A(p\ )$ and the original sketch s to G_B, the photo that has been rebuilt$G_B(G_A(p), \eta_p)$ and the synthesised photo $G_B(s, \eta_s)$ are formed accompanied by the labels p and s. To verify that G_B acquires the proper form of rectification from sketch to picture domain for each class, we only transmit the drawing with its actual label. A pixel-wise consistency loss imposes a similarity requirement on the rebuilt photo. We don't impose a consistency restriction on the sketch domain since we want the synthesised sketches to be as varied as possible. The generated photo is sent to discriminator D_B as inputs to check if it's real or fake and the classifier R confirms that it shares the target class's perceptual properties.

B. Training –

Four elements make up the generator loss:

1. The adversarial loss of photo generation to sketch generation. I.e. $L_{G\ \ A}$
2. The adversarial loss of sketch translation to photo translation. I.e. $L_{G\ \ B}$
3. The pixel consistency of photo reconstruction. I.e. L_{pix}
4. The classification loss for synthesizing to photos. I.e. L_{η}

C. Generator Loss –

International Journal of Engineering Applied Sciences and Technology, 2021
Vol. 6, Issue 2, ISSN No. 2455-2143, Pages 251-257
Published Online June 2021 in IJEAST (http://www.ijeast.com)

$$L_{GAN} = \lambda_A L_{G_A}(G_A, D_A, p) + \lambda_B L_{G_B}(G_B, D_B, s, \eta_s) + \lambda_{pix} L_{pix}(G_A, G_B, p, \eta_p) + \lambda_\eta L_\eta(R, G_B, s, \eta_s)$$

Due to the missing sketches s, the training objectives for open-domain classes M take the following form if we simply train the multi-class generator with the aforementioned loss:

$$L^M_{GAN} = \lambda_A L_{G_A}(G_A, D_A, p) + \lambda_{pix} L_{pix}(G_A, G_B, p, \eta_p)$$

$$\text{Where } \eta_p \in M.$$

As a result, the open-domain classes get pixelated images from the drawing to photo generator G_B Since L_1 and L_2 loss leads to the median and mean of pixels, respectively.

D. Training strategy for limited data: zero-shot/open-domain –

For some categories, we have no sketch for training. For example,

Index	Category	Sketch Number
0	Pineapple	151
1	Strawberry	0
2	Basketball	147
3	Chicken	0
4	Cookie	146
5	Cupcake	0
6	Moon	0
7	Orange	146
8	Soccer	0
9	Watermelon	146

Table 1: Example

To address this issue, we suggest a random-mixed sampling method to reduce the impact of the domain gap between real and false sketch inputs on the generator. This approach mainly consists of generating fake drawings from pictures and mixing them randomly with the actual ones in the pre-built batch. Consequently, it will enhance the quality of the output with open-domain categories. All in-domain and open-domain classes are invisible to the random mixing procedure. As a direct consequence, the batch pool gets enhanced with both genuine and fake produced drawings and related tags from various epochs during the learning process.

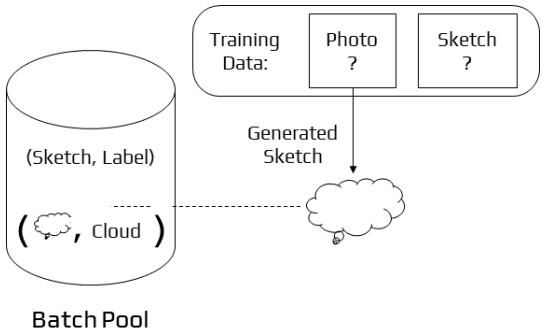


Figure 5: Proposed Solution: Mixed Sampling with Batchwise Substitution

E. Random Mixed Strategy –

The goal of this method is to deceive the generator into thinking bogus sketches are real.

Algorithm: Minibatch Random-Mixed Sampling and Updating

Input: In training set $\mathcal{D}$, each minibatch contains photo set p, freehand sketch set s, the class label of photo η_p, and the class label of sketch η_s;
for *number of training iterations* **do**
  $s_{fake} \leftarrow G_A(p)$;
  $s_c \leftarrow s$;
  $\eta_c \leftarrow \eta_s$;
  if $t < u \sim U(0,1)$ **then**
    $s_c, \eta_c \leftarrow pool.\text{query}(s_{fake}, \eta_p)$;
  end
  $p_{rec} \leftarrow G_B(s_{fake}, \eta_p)$;
  $p_{fake} \leftarrow G_B(s, \eta_s)$
  Calculate $\mathcal{L}_{GAN}$ with $(p, s_c, p_{rec}, \eta_c)$ and update G_A and G_B;
  Calculate $\mathcal{L}_{D_A}(s, s_{fake})$ and $\mathcal{L}_{D_A}(p, p_{fake})$, update D_A and D_B;
  Calculate $\mathcal{L}_R(p, p_{fake}, \eta_p, \eta_s)$ and update the classifier.
end

Algorithm

International Journal of Engineering Applied Sciences and Technology, 2021
Vol. 6, Issue 2, ISSN No. 2455-2143, Pages 251-257
Published Online June 2021 in IJEAST (http://www.ijeast.com)

IV. NETWORK ARCHITECTURE

We show how our framework's architecture works, including generators, discriminators, and a classifier. Note that our proposed method is not tied to a certain network design; we chose the CycleGAN as a baseline to demonstrate its effectiveness. As a result, we merely alter the G_B to a multi-class generator and leave the rest of the structures alone, as seen below.

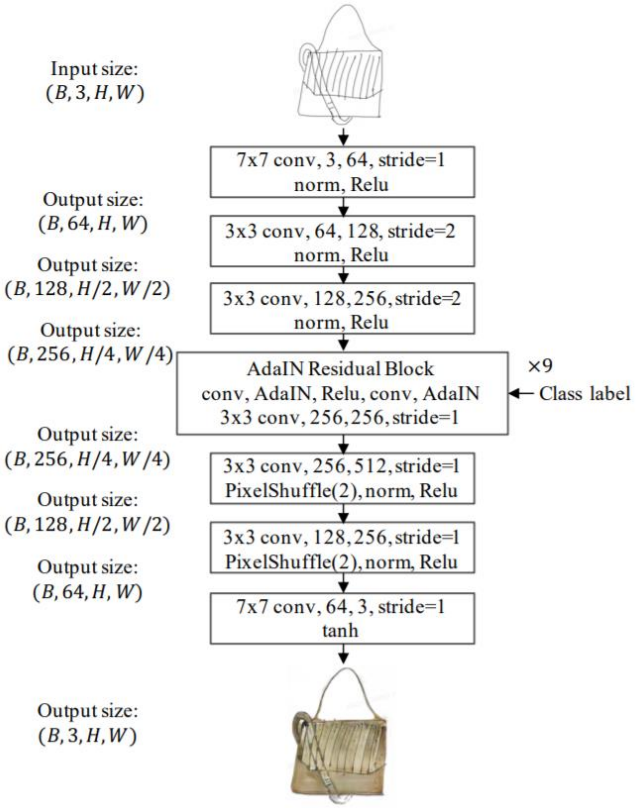


Figure 6: Multi-Class Sketch-to-Photo Generator Architecture

V. SKETCH-TO-PHOTO SYNTHESIS

To show the efficacy of our method, we use CycleGAN [14] as the basis for creating our network. We converted the sketch-to-photo translator into a conditional generator so that it could receive sketch class labels. We compared a recent EdgeGAN [15] study on each dataset with different other models. On all datasets, we can state that our model is chosen by more people than other approaches and delivers greater performance in terms of FID [16] rating and classification accuracy. We also put our produced robustness to the test against the input. Our approach works well for drawings that have been altered by deleting and introducing strokes.

VI. PHOTO-TO-SKETCH SYNTHESIS

For a given photo, our network can also generate a high-quality freehand sketch generator G_A. Beyond the edge map of a photo, our model can create numerous types of freehand drawings like human drawers, even for invisible things. The weights of the generator are continually updated as the training advances, which is characterised by joint training. As a byproduct, the drawings produced by G_A evolve epoch by epoch. The changing drawings broaden the sketch's diversity, which can help the sketch-to-photo generator generalise to a wider range of hand-drawn sketch sources.

VII. EXPERIMENTS

A. Implementation –

PyTorch was used for the implementation of the model and training was done with the Adam algorithm [17], an optimiser for deep learning. We trained using 1 NVIDIA V100 GPU. The loss function to train the generator contains four elements: the adversarial loss of photo-to-sketch generation $L_{G\ \ A}$, the adversarial loss of sketch-to-photo translation $L_{G\ \ B}$, the pixel-wise consistency of photo reconstruction L_{pix}, and the classification loss for synthesised photo L_{η}. For the discriminator's DA and DB, two-loss functions are used.

For all datasets, the batch size is set at 1 and the initial learning rate is set at 2e - 4.

The epoch numbers for the Scribble[18, QMUL-Sketch[19], and SketchyCOCO [15] are 200, 400, and 100, respectively. In the second half of the epochs, the learning rates are multiplied by 0.5.

B. Dataset –

The model was trained using three sets of data: Scribble which contained 10 classes, SketchyCOCO which contained 14 classes of objects, and open domain setting. To satisfy the open-domain conditions, the drawings of some classes are eliminated during the training stage.

C. Results –

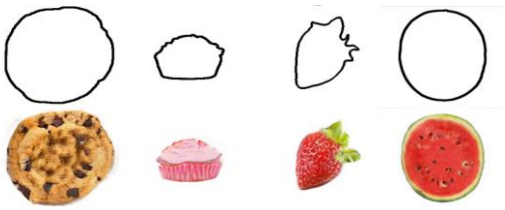


Figure 7: Results on Scribble Dataset

International Journal of Engineering Applied Sciences and Technology, 2021
Vol. 6, Issue 2, ISSN No. 2455-2143, Pages 251-257
Published Online June 2021 in IJEAST (http://www.ijeast.com)

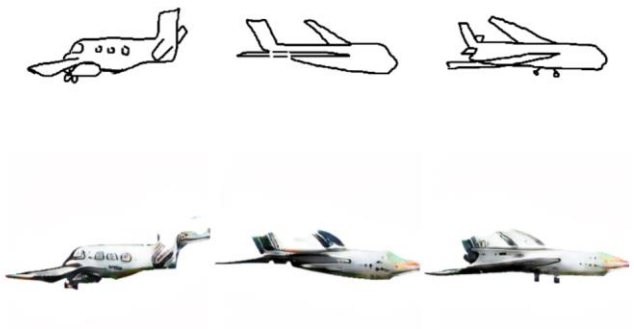


Figure 8: Results on SketchCOCO Dataset

Figure 9: Results on QMUL-Sketch Dataset

As we can see, the output photos are very similar to the open domain inputs. Based on the shape of the input image, the framework can synthesise the output as well as colouring the photo.

D. Evaluation Metrics –

Quantitative evaluation was done with three distinct metrics. The Frechet Inception Distance (FID) analyses the feature similarity of generated and actual image collections. A low FID score indicates that the produced pictures are less dissimilar to the genuine ones and hence have good integrity. Classification Accuracy of produced pictures. Having higher accuracy implies realistic images. And finally, User Preference Study in which we display the users a provided drawing and the class label, and ask them to choose one photo from the produced results that has the highest quality and authenticity.

E. Comparison with Other Methods –

To demonstrate the efficiency of our model, we have compared it with several other solutions like CycleGAN, conditional CycleGAN, Classifier and CycleGAN, EdgeGAN. We have used two datasets for this comparison: Scribble and SketchyCOCO. For some, we have given as input open domain sketches.

Scribble: The initial CycleGAN displays mode collapse and synthesises similar textures for all categories, most likely because the sketches in the Scribble dataset hardly suggest their class names. This issue is resolved with the conditional CycleGAN but it still fails to generate real pictures for some classes. For the classifier and conditional CycleGAN, the gap between the open domain and in-domain data is even bigger, because the classifier widens current differences. The EdgeGAN is closer to the input shape but still doesn't entirely translate the sketch. The consistency of the model is shown in our results where the mapping of the images is done with accuracy.

SketchyCOCO: CycleGAN's outputs struggle from mode collapse whereas conditional CycleGAN cannot create rich textures for open-domain categories. In comparison to EdgeGAN, the postures in our produced photographs are more accurate than the original drawings.

Our model is favoured by more participants than the other techniques evaluated, and that it produces the best results in terms of FID score and classification accuracy across all datasets. And we can see from our comparison that our random mixed sampling algorithm improves both the in the domain and open domain results.

VIII. EFFECTIVENESS OF AODA

A. Open-domain classes –

1. Baseline without classifier or strategy.
2. AODA framework without the strategy.
3. Trained with pre-extracted open-domain and real in-domain sketches.
4. Random Mixed sampling strategy

B. Observation –

1. Turn everything to the in-domain category.
2. Generates texture-less images.
3. Fails on open-domain category.
4. AODA strategy can alleviate the above issues and brings superior performance for a multi-class generation

IX. CONCLUSION

Synthesis of sketch-to-photo is a challenging task, especially for freehand sketches The open set domain, where the source domain has labelled categories and the target domain has undefined categories, raises an issue as the data is unpaired. To solve this problem, we are proposing an adversarial open domain adaption that can learn how to develop the missing

hand-drawn sketches. The framework enables the unsupervised open domain adaption by learning sketch-to-image translation and vice versa. Also, the random mixing sampling method reduces the domain gap's effect on the generator, thus generating accurate images. Additional testing and user assessments on a variety of samples show that our model can properly synthesize real images for many types of open-domain hand-drawn sketches. Further work in this model can lead to the generation of high-quality photos by improving the design and accuracy.

X. REFERENCE

[1] Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. "Generative adversarial networks." *Communications of the ACM* 63, no. 11 (2020): 139-144.

[2] Wang, Xinrui, and Jinze Yu. "Learning to cartoonize using white-box cartoon representations." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8090-8099. 2020.

[3] Thakur, Amey, Hasan Rizvi, and Mega Satish. "White-Box Cartoonization Using An Extended GAN Framework." *arXiv preprint arXiv:2107.04551* (2021)

[4] Dekel, Tali, Chuang Gan, Dilip Krishnan, Ce Liu, and William T. Freeman. "Sparse, smart contours to represent and edit images." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3511-3520. 2018.

[5] Bhunia, Ayan Kumar, Yongxin Yang, Timothy M. Hospedales, Tao Xiang, and Yi-Zhe Song. "Sketch less for more: On-the-fly fine-grained sketch-based image retrieval." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9779-9788. 2020.

[6] Han, Xiaoguang, Chang Gao, and Yizhou Yu. "Deepsketch2face: A deep learning based sketching system for 3d face and caricature modeling." *ACM Transactions on graphics (TOG)* 36, no. 4 (2017): 1-12.

[7] Chen, Wengling, and James Hays. "Sketchygan: Towards diverse and realistic sketch to image synthesis." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 9416-9425. 2018.

[8] Fang, Zhen, Jie Lu, Feng Liu, Junyu Xuan, and Guangquan Zhang. "Open set domain adaptation: Theoretical bound and algorithm." *IEEE Transactions on Neural Networks and Learning Systems* (2020).

[9] Li, Da, Yongxin Yang, Yi-Zhe Song, and Timothy M. Hospedales. "Deeper, broader and artier domain generalization." In *Proceedings of the IEEE international conference on computer vision*, pp. 5542-5550. 2017.

[10] Isola, Phillip, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. "Image-to-image translation with conditional adversarial networks." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1125-1134. 2017.

[11] Xiang, Xiaoyu, Ding Liu, Xiao Yang, Yiheng Zhu, Xiaohui Shen, and Jan P. Allebach. "Adversarial Open Domain Adaption for Sketch-to-Photo Synthesis." *arXiv preprint arXiv:2104.05703* (2021).

[12] Shermin, Tasfia, Guojun Lu, Shyh Wei Teng, Manzur Murshed, and Ferdous Sohel. "Adversarial Network with Multiple Classifiers for Open Set Domain Adaptation." *arXiv preprint arXiv:2007.00384* (2020).

[13] Mirza, Mehdi, and Simon Osindero. "Conditional generative adversarial nets." *arXiv preprint arXiv:1411.1784* (2014).

[14] Liu, Ming-Yu, Thomas Breuel, and Jan Kautz. "Unsupervised image-to-image translation networks." *arXiv preprint arXiv:1703.00848* (2017).

[15] Gao, Chengying, Qi Liu, Qi Xu, Limin Wang, Jianzhuang Liu, and Changqing Zou. "SketchyCOCO: Image Generation from Freehand Scene Sketches." *arXiv preprint arXiv:2003.02683*(2020).

[16] Obukhov, Artem, and Mikhail Krasnyanskiy. "Quality Assessment Method for GAN Based on Modified Metrics Inception Score and Fréchet Inception Distance." In *Proceedings of the Computational Methods in Systems and Software*, pp. 102-114. Springer, Cham, 2020.

[17] Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." *arXiv preprint arXiv:1412.6980*(2014).

[18] Ghosh, Arnab, Richard Zhang, Puneet K. Dokania, Oliver Wang, Alexei A. Efros, Philip HS Torr, and Eli Shechtman. "Interactive sketch & fill: Multiclass sketch-to-image translation." In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1171-1180. 2019.

[19] Liu, Runtao, Qian Yu, and Stella Yu. "Unsupervised Sketch to Photo Synthesis." *arXiv preprint arXiv:1909.08313* (2019).