

Article

# A Transfer Learning for Line-Based Portrait Sketch

Hyungbum Kim <sup>1</sup>, Junyoung Oh <sup>1</sup> and Heekyung Yang <sup>2,\*</sup> <sup>1</sup> Department of Computer Science, Sangmyung University, Seoul 03016, Korea<sup>2</sup> Division of SW Convergence, Sangmyung University, Seoul 03016, Korea

\* Correspondence: yanghk@smu.ac.kr; Tel.: +82-2-2287-6156

**Abstract:** This paper presents a transfer learning-based framework that produces line-based portrait sketch images from portraits. The proposed framework produces sketch images using a GAN architecture, which is trained through a pseudo-sketch image dataset. The pseudo-sketch image dataset is constructed from a single artist-created portrait sketch using a style transfer model with a series of postprocessing schemes. The proposed framework successfully produces portrait sketch images for portraits of various poses, expressions and illuminations. The excellence of the proposed model is proved by comparing the produced results with those from the existing works.

**Keywords:** sketch; transfer learning; portrait; GAN; AdaIN

**MSC:** 68T07



**Citation:** Kim, H.; Oh, J.; Yang, H. A Transfer Learning for Line-Based Portrait Sketch. *Mathematics* **2022**, *10*, 3869. <https://doi.org/10.3390/math10203869>

Academic Editors: Andrea Prati, Luis Javier García Villalba and Vincent A. Cicirello

Received: 9 September 2022

Accepted: 11 October 2022

Published: 18 October 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

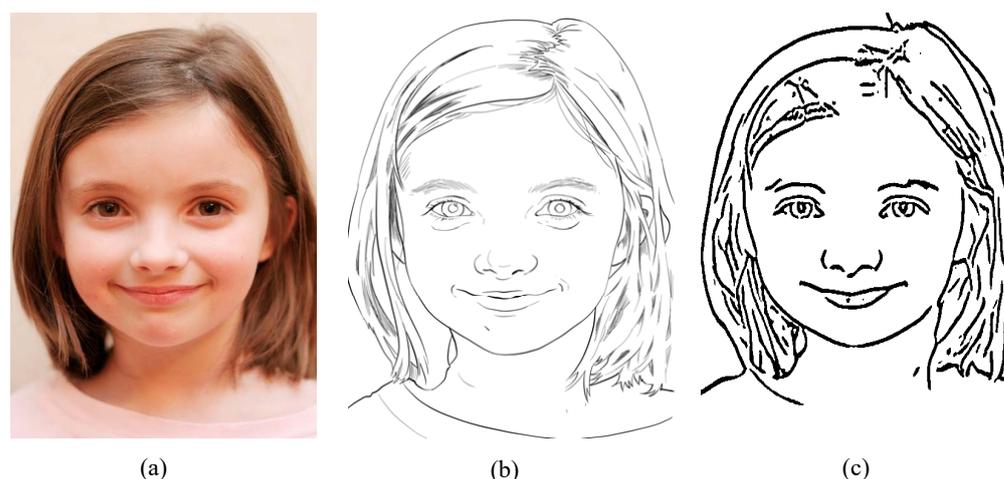
In computer graphics and computer vision research, many techniques that generate sketches from images have been presented. In the early days, the sketch generation schemes have been developed in the domain of image processing as edge-detection schemes. Widely used schemes including Canny edge [1], XDoG [2], and FDoG [3] belong to this category. After deep learning has been introduced, example-based methods that require the pairs of target images and their corresponding sketches are widely studied. Among those schemes, pix2pix [4], CycleGAN [5], DualGAN [6] or MUNIT [7] produces sketches from their target images. Recently, several dedicated schemes based on CNN or RNN have been presented for extracting sketches from images [8–13]. These techniques, however, have a limitation that the quality of the results is deeply influenced by the quality of training dataset. Therefore, deep-learning-based sketch extraction models require training sets of high-quality image–sketch pairs for extracting visually convincing quality sketches from images.

In computer graphics and computer vision society, many schemes that produce face sketches expressing faces with salient lines and smooth tone have been presented [14–19]. However, the schemes that produces line-based sketches from various face images are rarely found.

Artistic line-based face sketches have the following characteristics: (i) the salient features in a face such as eyes, lips, nasal wings or hair are clearly preserved; (ii) the shadows or highlights on a face that may produce more salient lines than eyes or eyelids in a face are diminished; (iii) the subtle deformation of eyes or lips that comes from expressions are clearly preserved; and (iv) many artifacts which frequently appear in many automatic sketch production models are not observed.

The aim of this paper is to produce a deep-learning-based framework that produce line-based sketch from portrait images with diverse poses, expressions and illuminations. The most straightforward strategy is to build a dataset of high-quality sketches with corresponding face images. This strategy has a limitation that such a dataset is very expensive and time consuming.

In the proposed framework, an automatic sketch generation scheme that extracts salient lines from face images and reduces artifacts is presented. In this framework, a pseudo-sketch dataset is constructed from a single-artist-created, high-quality sketch image. Transfer learning is executed by re-training an existing deep-learning-based sketch production model with this newly constructed sketch dataset. This framework produces line-based facial sketch images of visually convincing quality. Figure 1 compares our results with the line-based sketch image created by a professional artist.



**Figure 1.** Comparison of the produced results with artistic line-based sketch: (a) input hedcut portrait image, (b) a line-based sketch artwork created by a professional artist, (c) the produced result.

The main contribution of this paper is to present an efficient transfer learning framework that produces visually convincing sketch portrait images from a single-artist-created portrait. Instead of constructing a dataset of artistic sketch images, which requires a lot of time and effort, the proposed transfer-learning-based technique presents an efficient framework from a single-artist-created sketch image. The proposed framework consists of two modules: a module that constructs the pseudo-sketch dataset and a module that produces sketch images from a portrait image. The first module employs a style transfer scheme, semantic segmentation scheme, and a threshold-based post-processing scheme. The whole process and the processes of the two modules are presented in Section 3. The techniques compared to the proposed framework include CycleGAN [5], MUNIT [7], RCCL [20], FDoG [3], and Photo-sketching [11].

## 2. Related Work

Studies on sketches can be classified into two categories: sketch-based retrieval and sketch generation. The sketch generation schemes are further classified into sub-categories according to the target images where sketch generation framework is applied.

### 2.1. Sketch-Based Retrieval

Sketch-Based Image Retrieval (SBIR) is a technique that searches images using sketching that describes a key content of the images. The sketch used in SBIR is drawn as free-hand style instead of formal and artistic style. Many category-level SBIR studies [21–26] achieve very impressive performance. Among them, classical SBIR studies [21–24] execute conventional training and inferencing processes for their purpose, and some zero-shot SBIR studies [25,26] separate the training and inferencing process to reduce annotation costs.

Recently, they focus on fine-grained SBIR (FG-SBIR) studies that searches the instance of input images instead of their domain [27–32]. FG-SBIR studies employ attention mechanism and triplet loss functions for fine grained details. Their direction aims to lessen retrieval time and lessen the number of strokes in the sketch. Bhunia et al. [31] proposed a real-time framework that executes sketch drawing and image retrieval simultaneously.

## 2.2. Sketch Generation

### 2.2.1. Image-to-Image Translation

Many image-to-image translation schemes [4–7,33,34] transform the images in one domain to another domain and vice versa. Therefore, the translation between the photo domain and sketch domain can produce sketches from input images. From the seminal work of [4], several cycle-consistency-based GANs such as CycleGAN [5] or DualGAN [6] employ two pairs of generators and discriminators for the cross-domain image transfer without paired image sets. MUNIT [7] improved the visual quality of the result images by decomposing feature space into content space and domain-wise style space.

The results of these image-to-image translation frameworks are stored in raster format, which include tonal depictions in their results. These works can produce sketches from input images in a straightforward way, but have the limitation that the strokes of the sketch, which can be expressed as vector format, are not produced.

### 2.2.2. Photo-to-Sketch Generation

Photo-to-sketch generation studies that produce sketches from input images can be categorized according to their result formats: raster sketches or vector sketches.

Classical edge detection schemes such as Canny edge [1] extract high-frequency regions of images without understanding the content of the image. Other contour extraction schemes [35] can extract only the outer contours. Recently, some deep learning models such as CNN or GAN have been employed for photo-to-sketch generation [11,13].

Among them, Li et al. [11] presented a deep-learning-based model that understands the contents of the images by collecting contour drawing images. This model successfully produced both internal and external contours from the input images. However, these schemes have a limitation that the quality of the resulting sketch heavily depends on the input image domain.

Photo-to-vector sketch studies [10,12] employ the encoder and decoder structure of Sketch RNN [9]. They employ LSTM-based encoders and decoders since they express the sketch as a set of strokes. Pix2seq [8] produces various vector sketches from input images by employing a CNN encoder.

These models require paired datasets of corresponding photo and vector sketches for their training. This dataset is very rare in the public domain. Most of these works employ QMUL-Shoe, Chair dataset [36]. Therefore, the result images are restricted to the images in the QMUL-Shoe, Chair dataset. Since the quality of the images in [36] is equivalent to free-hand sketches, the quality of the resulting sketches is far from that of artistic sketches. Song et al. [10] improved the quality of the resulting sketch by reducing the domain difference through shortcut cycle consistency.

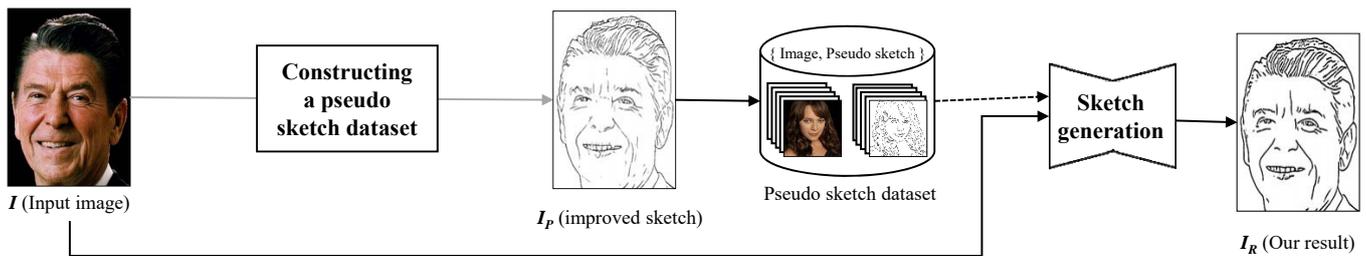
## 3. The Proposed Method

The key idea in this paper is to construct a pseudo-sketch dataset from a single-artist-created sketch image and train a sketch generation model with the constructed pseudo-sketch dataset. The first stage of the proposed model constructs a pseudo-sketch dataset using an existing style transfer scheme and a series of post-processing steps. The second stage produces sketch images from the pseudo-sketch dataset using a GAN-based approach. This process is illustrated in Figure 2.

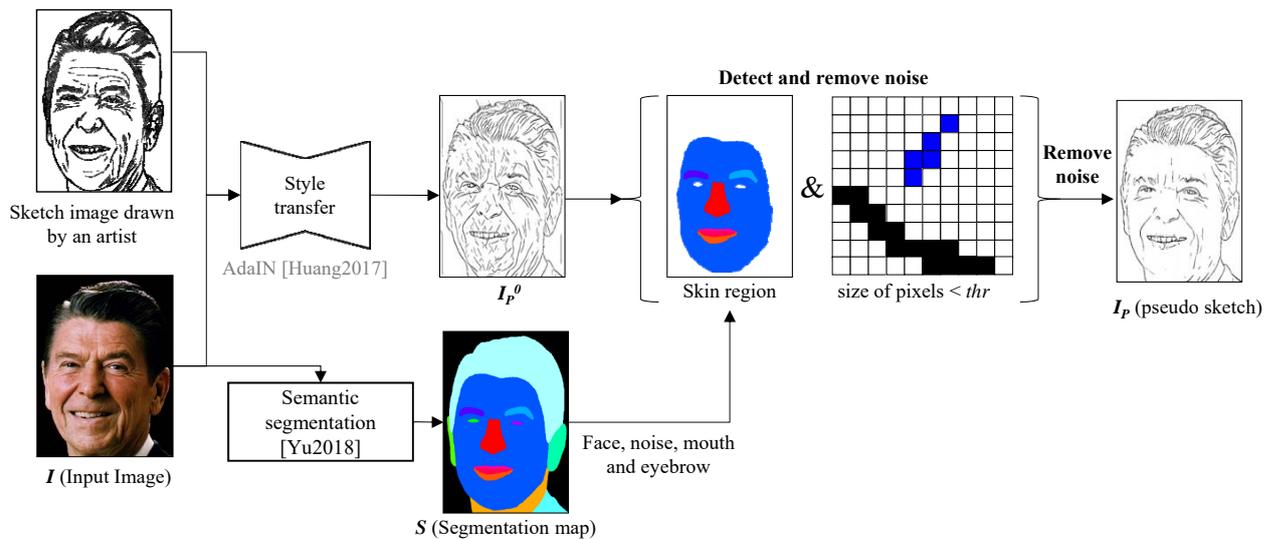
### 3.1. Constructing a Pseudo-Sketch Dataset

The first step of the proposed approach is to build a set of pseudo-sketch images from input hedcut portrait images using an existing style transfer model. The AdaIN-based style transfer model [37], which is recognized as one of the most effective style transfer models, is used. The AdaIN model successfully extracts important lines for a sketch ( $I_p^0$ ), but it also produces a lot of unwanted artifacts in the primitive sketch images. To diminish these artifacts, a semantic segmentation model [38] was used to estimate a segmentation map that distinguishes important regions in a face image. This map is applied to the primitive

sketches to diminish the artifacts. The short lines in a face that do not correspond to salient features in the segmentation map are recognized as artifacts and removed. To remove the artifacts, an adaptive thresholding technique is employed to compute average within the block size of the central pixel, which is experimentally assigned as 55. The average is set as a threshold. This operation is performed on all pixels to remove the artifacts. This post-processing produces pseudo-sketches ( $I_p$ ), which compose a pseudo-sketch dataset, which is a paired set of the input hedcut portrait image and its pseudo-sketch image. This process is illustrated in Figure 3.



**Figure 2.** The overview of the proposed framework. From a series of input images, a pseudo-sketch dataset, which is further employed for training a sketch generation module, is constructed. Finally, the proposed framework produces a line-based sketch that preserves important lines and reduces unwanted artifacts.



**Figure 3.** The process please moved Figure 3 body after its first citation in the main text according to our layout rule, the same with all figures. please confirm. of constructing a pseudo-sketch dataset. An AdaIN-based style transfer model [37] to an input image  $I$  is applied in order to extract a primitive sketch image  $I_p^0$ , which contains important lines as well as a lot of artifacts. To preserve the important lines and to diminish the artifacts, a semantic segmentation is executed and the resulting segmentation map is employed to build a pseudo sketch,  $I_p$ . Improved sketch dataset is constructed from this set of  $I_p$ 's. This dataset is applied to re-train a sketch generation model, which produces the resulting sketch image,  $I_g$ .

### 3.2. Sketch Generation

Line-based sketch images using a GAN structure [39] is generated. The proposed model is pre-trained by Li et al.'s strategy [11] that produces line-based sketch images from various ranges of input images. Since the domain is restricted to human portraits, a transfer learning model that re-trains the pre-trained model is executed with the produced pseudo-sketch dataset, which was constructed in the previous stage.

The loss function in the proposed model includes adversarial loss and L1 loss. The adversarial loss is defined as in Equation (1):

$$L_{adv}(\mathbf{I}_p, \mathbf{I}_g, \mathbf{I}) = \mathbb{E}[\log D(\mathbf{I}_g, \mathbf{I})] + \mathbb{E}[\log(1 - D(\mathbf{I}_p, \mathbf{I}))]. \tag{1}$$

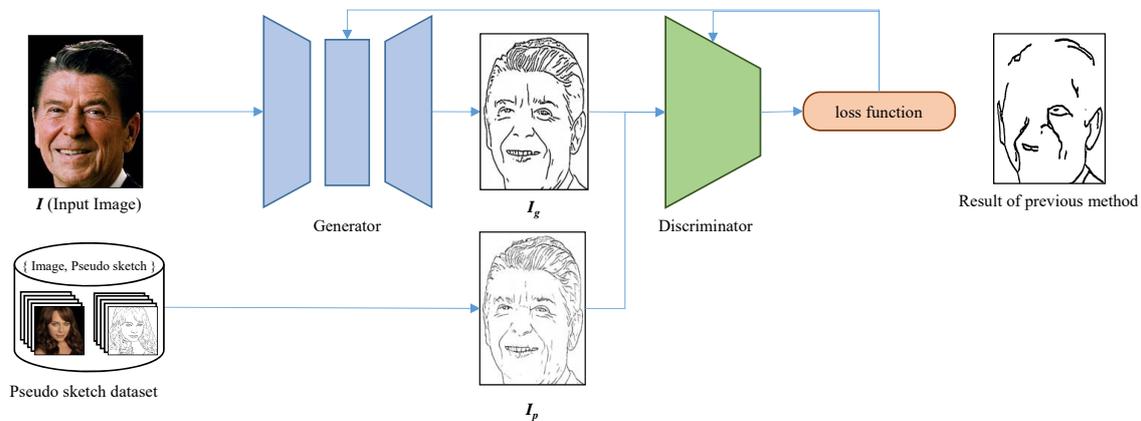
Another loss in the proposed model is the L1 loss that compares the difference of the pseudo-sketch image  $\mathbf{I}_p$  and the generated image  $\mathbf{I}_g$ . The L1 loss is defined as in Equation (2):

$$L_{L1}(\mathbf{I}_p, \mathbf{I}_g) = \|\mathbf{I}_p - \mathbf{I}_g\|_1. \tag{2}$$

Finally, the proposed loss function is defined as Equation (3):

$$L_{total} = \lambda_1 L_{adv} + \lambda_2 L_{L1}, \tag{3}$$

where  $\lambda_1$  and  $\lambda_2$  are set as 0.5. These values are determined experimentally. The process of sketch generation is illustrated in Figure 4.



**Figure 4.** The process of generating line-based sketch images. This initial generator produces a sketch image  $I_g^0$ , which shows the quality of the existing work [11]. The initial generator is further trained in the discriminator with the pseudo-sketch image  $I_p$  from the improved sketch dataset. This generator after training produces a sketch image  $I_g$ . Note the difference in quality between the produced result  $I_g$  and the existing result  $I_g^0$  [11].

## 4. Implementation and Results

### 4.1. Implementation

The initial generation model is trained with pseudo sketch dataset. Adam is employed as an optimization method with 0.0002 as the learning rate. The initial value of this learning rate is aligned with that of Li et al.’ work [11]. The hyperparameters including learning rate and its decay rate further evolve experimentally. The proposed model is trained for 100 epochs, and then the learning rate is decreased by a factor of 10 for a further 100 epochs. The training took approximately one day.

The proposed portrait sketch generation model is implemented on a personal computer with a single Intel Pentium I7 CPU and double nVidia GTX 3090Ti GPUs. The operating system is Linux Ubuntu. The proposed model is implemented using Python with Pytorch library. Fifty-five artists-created sketch images are employed.

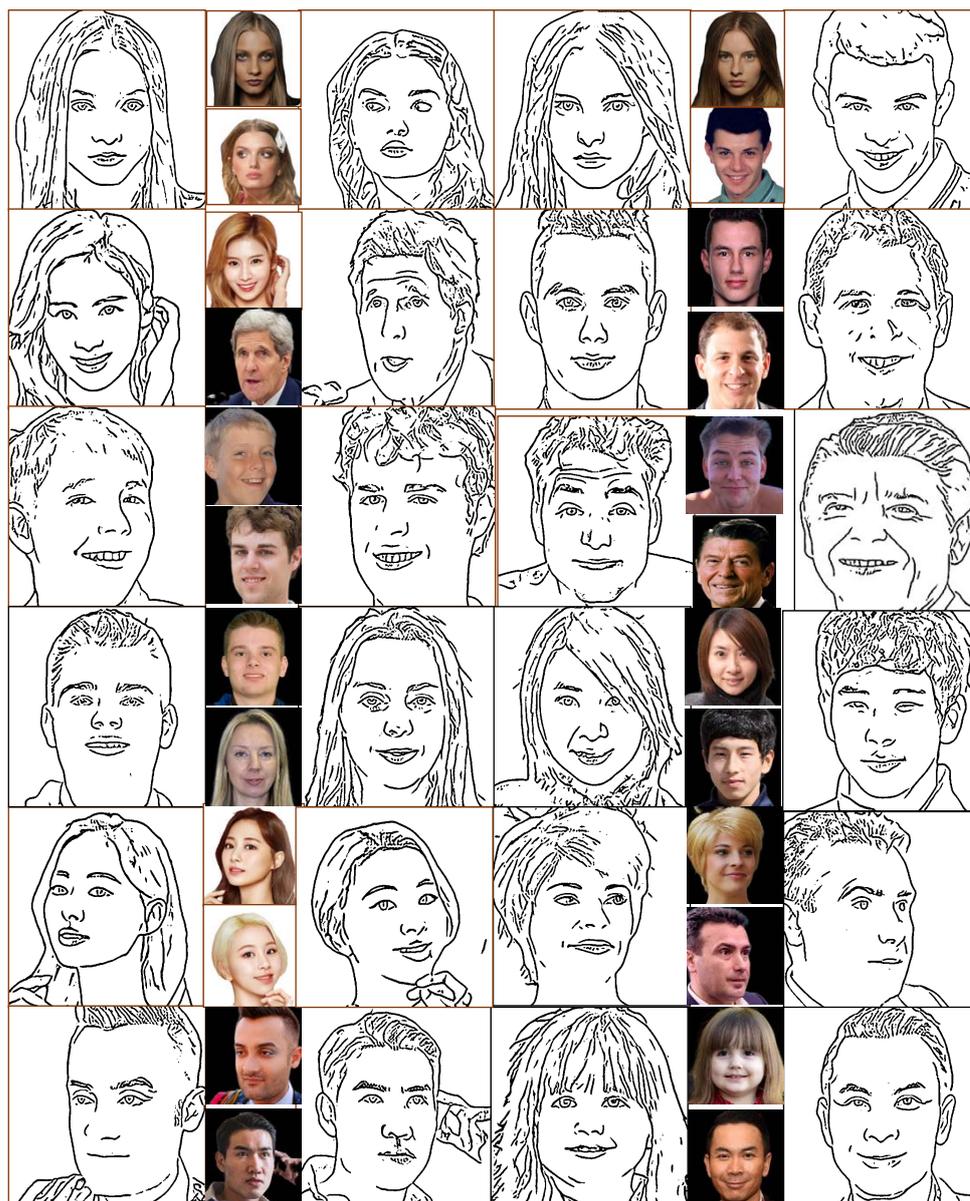
### 4.2. Training

A professional artist was hired to draw nine artist-created sketches. One of them is selected to train the AdaIN model of the pseudo-sketch construction module in Figure 3. This module produces thirty pseudo-sketches. This pseudo-sketch dataset is employed to train the sketch generation module in Figure 4. After training, the nine portraits with GT artist-created sketches are tested for the validation of the proposed framework. Finally,

portrait images without corresponding sketch images are collected, and their sketch images are generated using the proposed sketch generation framework.

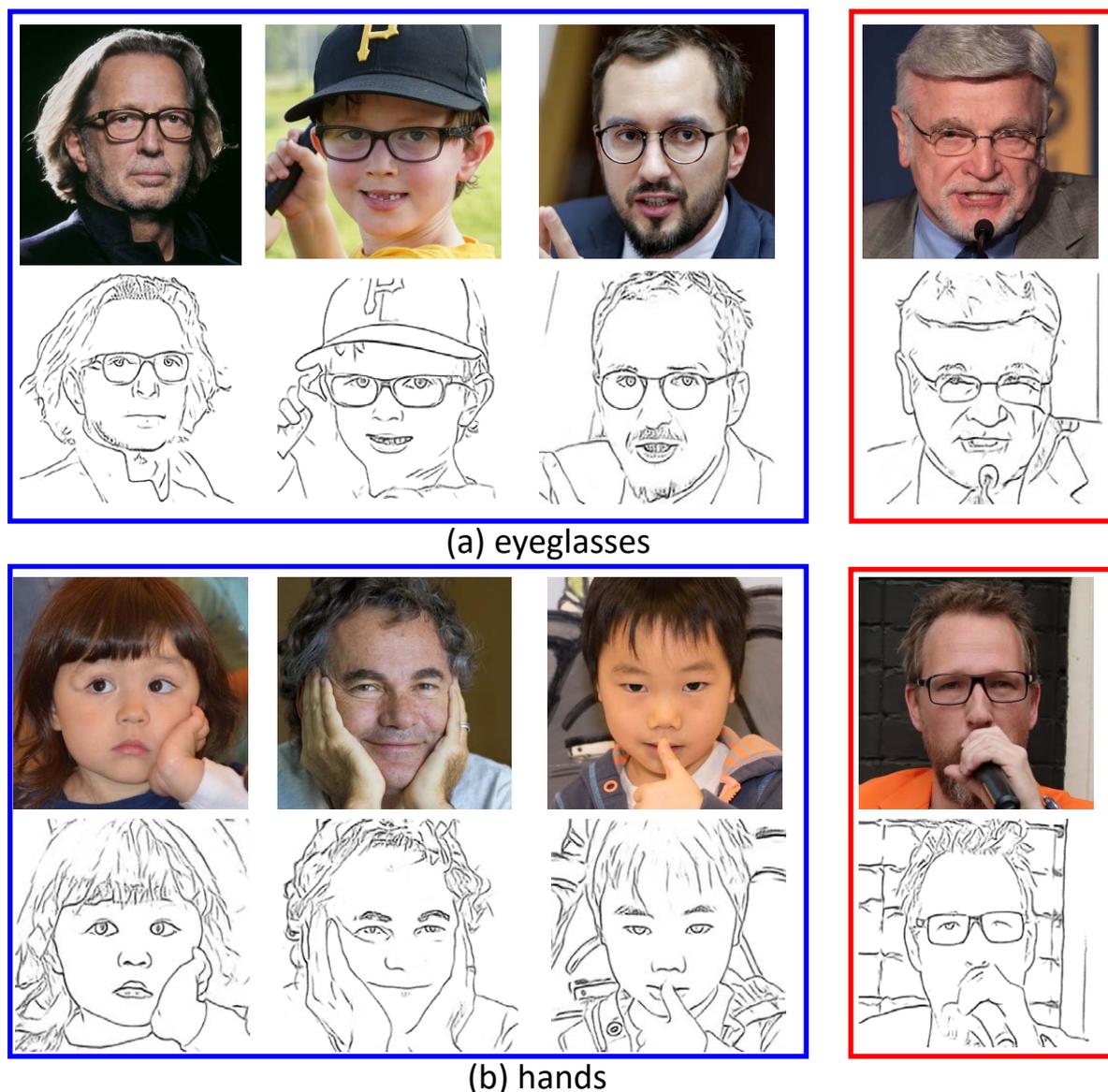
#### 4.3. Results

Various hedcut portrait images of various people with diverse attributes, including gender, race and age, are applied to the proposed model. The input portrait images with their resulting sketch images are presented in Figure 5. In Figure 5, salient features as well as detailed information in the portrait images are conveyed in a line-based style on the result images.



**Figure 5.** The result of the proposed framework.

The proposed model is also applied for portrait images with obstacles such as eyeglasses or hands. Figure 6 presents the result images. As illustrated in Figure 6, most of the results are convincing. However, some portraits show somewhat unsatisfactory results. For an analysis, some portraits with unclear obstacles tend to produce unsatisfactory results. The man in the rightmost portrait wears frameless eyeglasses in Figure 6a, and the fingers of the man in the rightmost portrait are not clear in Figure 6b.



**Figure 6.** Results from portraits with obstacles. The images in the blue box show convincing results, while the images in the red box show somewhat unsatisfactory results.

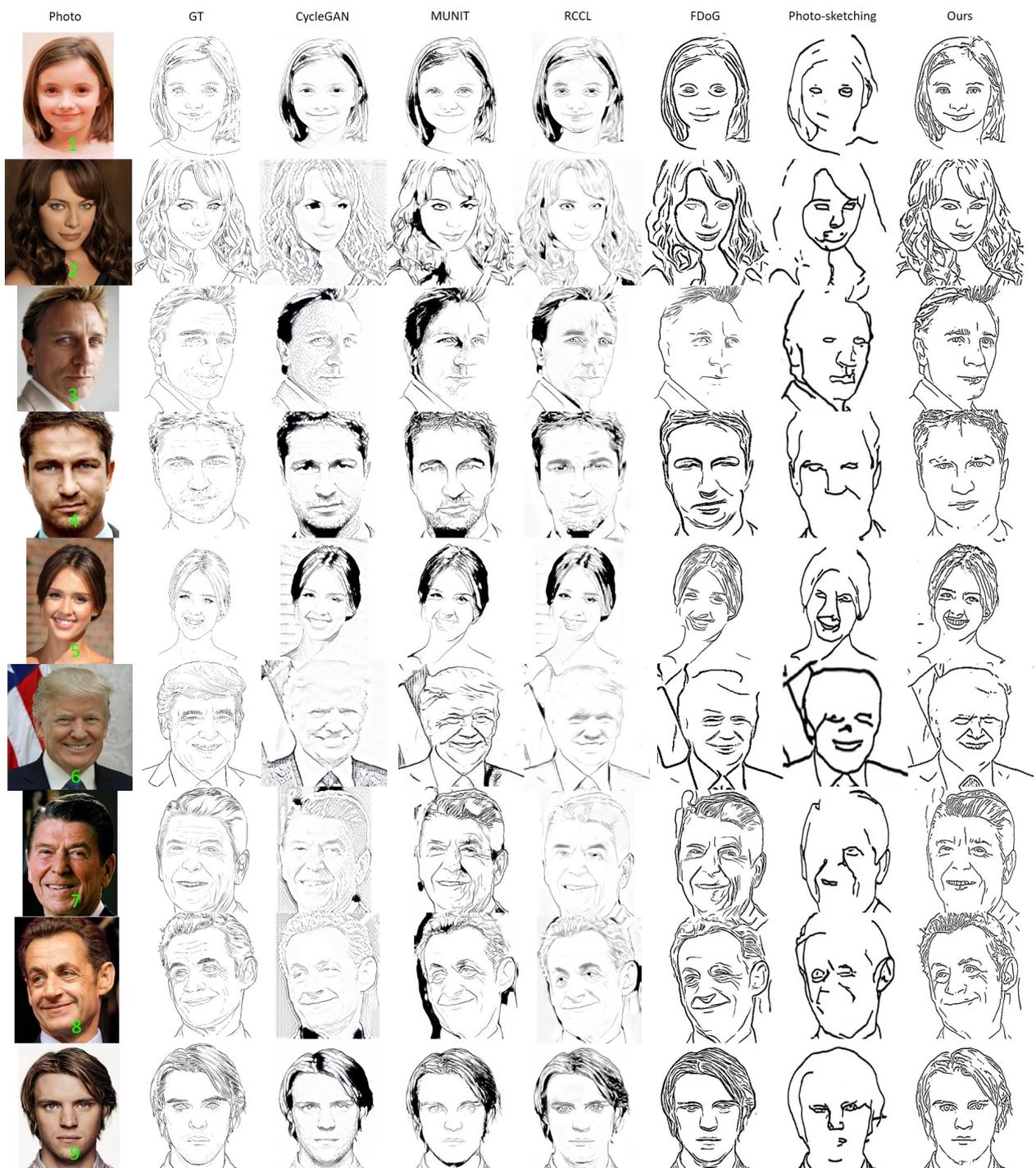
## 5. Evaluation and Analysis

### 5.1. Comparison

The produced results are compared with several existing schemes including CycleGAN [5], MUNIT [7], RCCL [20], FDoG [3], and Photo-sketching [11]. The produced results are further compared with GT images, which are created by professional artists. The visual comparison of the produced results is suggested in Figure 7.

### 5.2. Evaluation

The evaluation of the generated sketch images is hardly achieved through the widely used Frechet Inception Distance (FID). Therefore, instead of using FID, the precision, recall, and F1 scores are estimated on the sketch images generated from the produced model and the compared existing models. Professional artists are hired to create hand-drawn sketches of the portrait images in the leftmost column in Figure 7 as GT images for this estimation. For a comparison, the hired artists are required to draw the lines on the exact border of the images. As little artistic deformation as possible is allowed.



**Figure 7.** Comparison with existing studies: GT is created by professional artists, CycleGAN [5], MUNIT [7], RCCL [20], FDoG [3], and Photo-sketching [11] are compared.

The  $TP$  (true positive),  $FP$  (false positive), and  $FN$  (false negative) rates are estimated from the generated sketch image  $I_s$  and GT image  $I_G$ .  $TP$  is defined as the number of pixels which belong to sketch lines in both the generated sketch image and the GT image;  $FP$  is the number of pixels which belong to sketch lines in the generated image, but not in the GT image; and  $FN$  is the number of pixels which do not belong to sketch lines in the generated image, but belong to sketch lines in the GT image. The values of the pixels in the line-based sketch images are either black (0) for the sketch lines or white (1) for the empty regions. Therefore, the  $TP$ ,  $FP$ , and  $FN$  are defined as follows:

$$TP = \sum_p (I_s(p) = 0 \text{ and } I_G(p) = 0)$$

$$FP = \sum_p (I_s(p) = 0 \text{ and } I_G(p) = 1)$$

$$FN = \sum_p (I_s(p) = 1 \text{ and } I_G(p) = 0)$$

Therefore, the precision of a generated sketch image is estimated as follows:

$$precision = \frac{TP}{TP + FP}$$

The recall is estimated as follows:

$$recall = \frac{TP}{TP + FN}$$

The F1 score is estimated as the harmonic average of precision and recall.

The precision, recall, and F1 score values are shown in Tables 1, 2, and 3, respectively.

From Table 1, the proposed model shows the best precision values on eight images among nine total comparison images in Figure 7, the best recall values on five images, and the best F1 score values on all nine comparison images.

**Table 1.** The precision estimated on the sketches in Figure 7. The red figures denote maximum precision in each row.

No.	CycleGAN	MUNIT	RCCL	FDoG	Photo Sketching	Ours
1	0.200	0.220	0.195	0.253	0.255	0.305
2	0.280	0.280	0.280	0.309	0.280	0.312
3	0.135	0.150	0.135	0.160	0.165	0.170
4	0.139	0.175	0.145	0.194	0.194	0.185
5	0.130	0.115	0.132	0.170	0.128	0.189
6	0.075	0.065	0.075	0.095	0.065	0.105
7	0.255	0.255	0.260	0.295	0.260	0.330
8	0.265	0.260	0.245	0.312	0.270	0.380
9	0.295	0.365	0.310	0.465	0.385	0.545

**Table 2.** The recalls estimated on the sketches in Figure 7. The red figures denote maximum recall in each row.

No.	CycleGAN	MUNIT	RCCL	FDoG	Photo Sketching	Ours
1	0.532	0.518	0.521	0.513	0.514	0.505
2	0.524	0.561	0.523	0.490	0.541	0.510
3	0.510	0.512	0.505	0.490	0.509	0.501
4	0.502	0.518	0.512	0.470	0.506	0.499
5	0.511	0.508	0.498	0.503	0.507	0.528
6	0.493	0.499	0.494	0.502	0.508	0.514
7	0.496	0.515	0.487	0.513	0.510	0.530
8	0.483	0.500	0.464	0.502	0.504	0.524
9	0.509	0.506	0.502	0.505	0.508	0.527

**Table 3.** The F1 scores estimated on the sketches in Figure 7. The red figures denote maximum F1 score in each row.

No.	CycleGAN	MUNIT	RCCL	FDoG	Photo Sketching	Ours
1	0.291	0.309	0.284	0.339	0.313	0.380
2	0.365	0.374	0.365	0.379	0.369	0.387
3	0.213	0.232	0.213	0.241	0.249	0.254
4	0.219	0.262	0.226	0.275	0.280	0.270
5	0.207	0.188	0.209	0.254	0.204	0.278
6	0.130	0.115	0.130	0.160	0.115	0.174
7	0.337	0.341	0.339	0.373	0.334	0.407
8	0.342	0.342	0.321	0.385	0.352	0.441
9	0.374	0.424	0.383	0.484	0.438	0.536

### 5.3. User Study

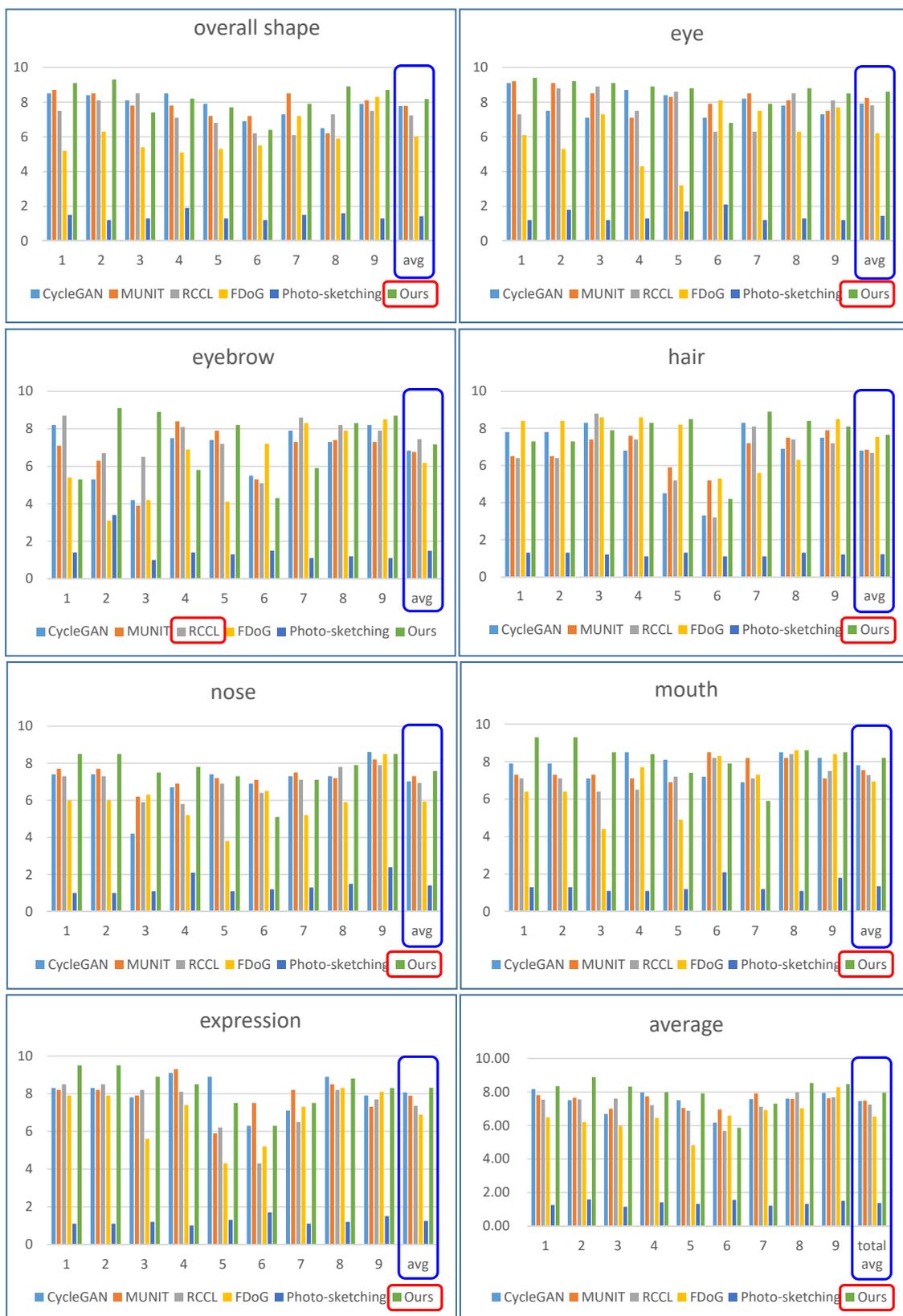
A user study is executed to evaluate the produced results and compare them with those from several existing studies. Thirty human participants were hired to evaluate the generated sketch images and compare them with the existing works. Twenty of the participants are in their twenties and ten are in their thirties. Thirteen of them are male, and seventeen are female. None of them have a fine art background. The participants were hired under only one criterion: no background in fine arts. Since fine art experts have their own standards for artistic sketches, this criterion is required for an unbiased evaluation of the results.

For the evaluation, two overall estimations and five local estimations were executed. For the overall estimation, participants were asked to evaluate the overall shape of the sketch and the preservation of the expression conveyed in the portrait in the sketch. The artist-created portraits are not presented for the participants. Participants were instructed to evaluate them on a 10-point scale. They were guided to give 1 point for the worst result and 10 for the best. The nine images in Figure 7 were employed for the user study. The answers from the participants are averaged and presented in the eight graphs in Figure 8, where the proposed model shows the best performance for six components. The proposed model shows the best overall average score.

### 5.4. Limitation

Even though the proposed model produces visually convincing line-based sketch images from input portraits, the following limitations are observed in processing several parts of a portrait. The most critical limitation is hair, which requires abstracted expression for the long and smooth flow of hair. The proposed model, however, produces many short and disconnected lines as well as long and salient lines. Another limitation is short line segments, which is observed in the face or hair. These artifacts are assumed to correspond to wrinkles, mustaches, or tiny features on faces. Since the proposed model does not consider such features, these features produce unwanted artifacts.

These limitations are found to come from the proposed strategy that a single model is applied to produce the strokes of various lengths lying on a sketch. A model that produces longer and more abstracted strokes should be different from a model that produces shorter and more detailed strokes. A model that generates sketch strokes should have an adaptive structure to the components of a portrait. The contingency plan for the limitations is to design a model with an adaptive structure for the strokes on a sketch.



**Figure 8.** The analysis of the user study. The average scores are presented in the blue box, and the model with the best score is presented in the red box. The proposed model records the six best scores for the individual scores and the best score for the overall average score.

## 6. Conclusions and Future Work

In this paper, a transfer learning approach for line-based portrait sketches is presented. A pseudo-sketch dataset from a single-artist-created sketch image is constructed by applying a style transfer model and several post-processing operations. This pseudo-sketch dataset is employed for training the proposed GAN-based architecture that produces line-based sketch images. Instead of using a lot of artist-created portrait sketch images that require a lot of time and effort, pseudo-sketch images play a very effective role in training a GAN-based sketch-generating architecture. The proposed model can successfully produce visually convincing sketch images from various portraits.

Since the proposed framework shows limitations in depicting long strokes and short strokes effectively, the next plan is to design a model that applies adaptive sketch drawing strategies to different parts of portrait images. After decomposing a portrait image into several components such as hair, eyes, eyebrows, and overall shape, a sketch-generation model with a shared core and an adaptive structure will be designed to produce an improved line-based sketch for the components of a portrait. Another direction is to apply tonal sketch representations that can improve the expressive domain of sketch drawings. The third direction is to develop an automatic sketch generation tool for a society of digital artists. Nowadays, webtoons, for example, are one of the hottest media in digital art, and sketching is a fundamental requirement for webtoon creators. The proposed model with a user-friendly interface can be a very effective tool for this community.

**Author Contributions:** Conceptualization, J.O.; Data curation, H.K.; Methodology, J.O.; Software, J.O.; Supervision, H.Y.; Validation, H.K.; Visualization, H.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by Sangmyung Univ. at 2020.

**Data Availability Statement:** Not available.

**Acknowledgments:** Not available.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

Acronym	Meaning
GAN	Generative Adversarial Network
xDoG	eXtended Difference of Gaussian
FDoG	Flow-based Difference of Gaussian
MUNIT	Multimodal UNsupervised Image-to-image Translation
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
SBIR	Sketch-Based Image Retrieval
FG-SBIR	Fine-Grained Sketch-Based Image Retrieval
LSTM	Long Short-Term Memory
AdaIN	Adaptive Instance Normalization
RCCL	Relaxed Cycle Consistency Loss
FID	Frechet Inception Distance
GT	Ground Truth

## References

1. Canny, J. A computational approach to edge detection. *IEEE Trans. Patt. Anal. Mach. Intel.* **1986**, *8*, 679–698. [[CrossRef](#)]
2. Winnemoller, H.; Kyprianidis, J.; Olsen, S. XDoG: An eXtended difference-of-Gaussians compendium including advanced image synthesis. *Comput. Graph.* **2012**, *36*, 740–753. [[CrossRef](#)]
3. Kang, H.; Lee, S.; Chui, C. Coherent line drawing. In Proceedings of the NPAR 2007, San Diego, CA, USA, 4–5 August 2007; pp. 43–50.

4. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A. Image-to-image translation with conditional adversarial networks. In Proceedings of the CVPR 2017, Honolulu, HI, USA, 21–26 July 2017; pp. 1125–1134.
5. Zhu, J.; Park, T.; Isola, P.; Efros, A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the ICCV 2017, Venice, Italy, 22–29 October 2017; pp. 2223–2232.
6. Yi, Z.; Zhang, H.; Tan, P.; Gong, M. Dualgan: Unsupervised dual learning for image-to-image translation. In Proceedings of the ICCV 2017, Venice, Italy, 22–29 October 2017; pp. 2849–2857.
7. Huang, X.; Liu, M.; Belongie, S.; Kautz, J. Multimodal unsupervised image-to-image translation. In Proceedings of the ECCV, Munich, Germany, 8–14 September 2018.
8. Chen, Y.; Tu, S.; Yi, Y.; Xu, L. Sketch-pix2seq: A model to generate sketches of multiple categories. *arXiv* **2017**, arXiv:1709.04121.
9. Ha, D.; Douglas, E. A neural representation of sketch drawings. *arXiv* **2017**, arXiv:1704.03477.
10. Song, J.; Pang, K.; Song, Y.; Xiang, T.; Hospedales, T. Learning to sketch with shortcut cycle consistency. In Proceedings of the CVPR 2018, Salt Lake City, UT, USA, 18–22 June 2018; pp. 801–810.
11. Li, M.; Lin, Z.; Mech, R.; Yumer, E.; Ramana, D. Photo-sketching: Inferring contour drawings from images. In Proceedings of the WACV 2019, Waikoloa Village, HI, USA, 7–11 January 2019; pp. 1403–1412.
12. Zhang, Y.; Su, G.; Qi, Y.; Yang, J. Unpaired image-to-sketch translation network for sketch synthesis. In Proceedings of the VCIP 2019, Sydney, Australia, 1–4 December 2019; pp. 1–4.
13. McGonigle, D.; Wang, T.; Yuan, J.; He, K.; Li, B. I2S2: Image-to-scene sketch translation using conditional input and adversarial networks. In Proceedings of the ICTAI 2020, Baltimore, MD, USA, 9–11 November 2020; pp. 773–778.
14. Wang, X.; Tang, X. Face photo-sketch synthesis and recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *31*, 1955–1967. [[CrossRef](#)] [[PubMed](#)]
15. Zhang, L.; Lin, L.; Wu, X.; Ding, S.; Zhang, L. End-to-end photosketch generation via fully convolutional representation learning. In Proceedings of the ICMR 2015, Shanghai, China, 23–26 June 2015; pp. 627–634.
16. Zhu, M.; Wang, N. A simple and fast method for face sketch synthesis. In Proceedings of the ICMCS 2016, Marrakech, Morocco, 29 September–1 October 2016; pp. 168–171.
17. Yu, J.; Xu, X.; Gao, F.; Shi, S.; Wang, M.; Tao, D.; Huang, Q. Toward realistic face photo-sketch synthesis via composition-aided GANs. In Proceedings of the Cybern 2020, Velké Karlovice, Czech Republic, 29 January–1 February 2020; pp. 1–13.
18. Qi, X.; Sun, M.; Wang, W.; Dong, X.; Li, Q.; Shan, C. Face sketch synthesis via semantic-driven generative adversarial network. In Proceedings of the IJCB 2021, Shenzhen, China, 4–7 August 2021; pp. 1–8.
19. Qi, X.; Sun, M.; Li, Q.; Shan, C. Biphasic Face Photo-Sketch Synthesis via Semantic-Driven Generative Adversarial Network with Graph Representation Learning. *arXiv* **2022**, arXiv:2201.01592.
20. Yi, R.; Liu, Y.J.; Lai, Y.K.; Rosin, P.L. Unpaired Portrait Drawing Generation via Asymmetric Cycle Mapping. In Proceedings of the CVPR 2020, Virtual, 14–19 June 2020; pp. 8217–8225.
21. Bui, T.; Ribeiro, L.; Ponti, M.; Collomosse, J. Deep manifold alignment for mid-grain sketch based image retrieval. In Proceedings of the ACCV 2018, Perth, Australia, 2–6 December 2018; pp. 314–329.
22. Cao, Y.; Wang, C.; Zhang, L.; Zhang, L. Edgel index for large-scale sketch-based image search. In Proceedings of the CVPR 2011, Springs, CO, USA, 20–25 June 2011; pp. 761–768.
23. Collomosse, J.; Bui, T.; Jin, H. Livesketch: Query perturbations for guided sketch-based visual search. In Proceedings of the CVPR 2019, Long Beach, CA, USA, 15–20 June 2019; pp. 2879–2887.
24. Collomosse, J.; Bui, T.; Wilber, M.; Fang, C.; Jin, H. Sketching with style: Visual search with sketches and aesthetic context. In Proceedings of the ICCV 2017, Venice, Italy, 22–29 October 2017; pp. 2660–2668.
25. Dutta, A.; Akata, Z. Semantically tied paired cycle consistency for zero-shot sketch-based image retrieval. In Proceedings of the CVPR 2019, Long Beach, CA, USA, 15–20 June 2019; pp. 5089–5098.
26. Dey, S.; Riba, P.; Dutta, A.; Lladós, J.; Song, Y. Doodle to search: Practical zero-shot sketchbased image retrieval. In Proceedings of the CVPR 2019, Long Beach, CA, USA, 15–20 June 2019; pp. 2179–2188.
27. Pang, K.; Song, Y.; Xiang, T.; Hospedales, T. Crossdomain generative learning for fine-grained sketch-based image retrieval. In Proceedings of the BMVC 2017, London, UK, 4–7 September 2017; pp. 1–12.
28. Song, J.; Qian, Y.; Song, Y.; Xiang, T.; Hospedales, T. Deep spatial-semantic attention for fine-grained sketch-based image retrieval. In Proceedings of the ICCV 2017, Venice, Italy, 22–29 October 2017; pp. 5551–5560.
29. Xu, P.; Yin, Q.; Huang, Y.; Song, Y.; Ma, Z.; Wang, L.; Xiang, T.; Kleijn, W.; Guo, J. Cross-modal subspace learning for fine-grained sketch-based image retrieval. *Neurocomputing* **2018**, *278*, 75–86. [[CrossRef](#)]
30. Yu, Q.; Liu, F.; Song, Y.; Xiang, T.; Hospedales, T.; Loy, C. Sketch me that shoe. In Proceedings of the CVPR 2016, Las Vegas, NV, USA, 27–30 June 2016; pp. 799–807.
31. Bhunia, A.; Chowdhury, P.; Sain, A.; Yang, Y. More photos are all you need: Semi-supervised learning for fine-grained sketch based image retrieval. In Proceedings of the CVPR 2021, Virtual, 21–23 June 2021; pp. 4247–4256.
32. Bhunia, A.; Yang, Y.; Hospedales, T.; Xiang, T.; Song, Y. Sketch less for more: On-the-fly fine-grained sketch-based image retrieval. In Proceedings of the CVPR 2020, Virtual, 13–19 June 2020; pp. 9779–9788.
33. Kim, J.; Kim, M.; Kang, H.; Lee, K. U-gat-it: Unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation. In Proceedings of the ICLR, Addis Ababa, Ethiopia, 26–30 April 2020.

34. Liu, M.; Breuel, T.; Kautz, J. Unsupervised image-to-image translation networks. In Proceedings of the NIPS, Long Beach, CA, USA, 4–9 December 2017.
35. Arbelaez, P.; Maire, M.; Fowlkes, C.; Malik, J. Contour detection and hierarchical image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 898–916. [[CrossRef](#)]
36. Yu, Q.; Song, Y.; Xiang, T.; Hospedales, T. SketchX!—Shoe/Chair Fine-Grained SBIR Dataset. Available online: <http://sketchx.eecs.qmul.ac.uk> (accessed on 1 March 2022).
37. Huang, X.; Belongie, S. Arbitrary style transfer in real-time adaptive instance normalization. In Proceedings of the ICCV 2017, Venice, Italy, 22–29 October 2017; pp. 1501–1510.
38. Yu, C.; Wang, J.; Peng, C.; Gao, C.; Yu, G.; Sang, N. BiSeNet Bilateral Segmentation Network for Real-time Semantic Segmentation. In Proceedings of the ECCV 2018, Munich, Germany, 8–14 September 2018; pp. 334–349.
39. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Bengio, Y. Generative Adversarial Networks. In Proceedings of the NIPS 2014, Montreal, QC, Canada, 8–13 December 2014; pp. 2672–2680.