Adversarial Open Domain Adaption for Sketch-to-Photo Synthesis

Xiaoyu Xiang[1]*, Ding Liu[2], Xiao Yang[2], Yiheng Zhu[2], Xiaohui Shen[2], Jan P. Allebach[1]
[1]Purdue University, [2]ByteDance Inc.
{xiang43,allebach}@purdue.edu, {liuding,yangxiao.0,yiheng.zhu,shenxiaohui}@bytedance.com

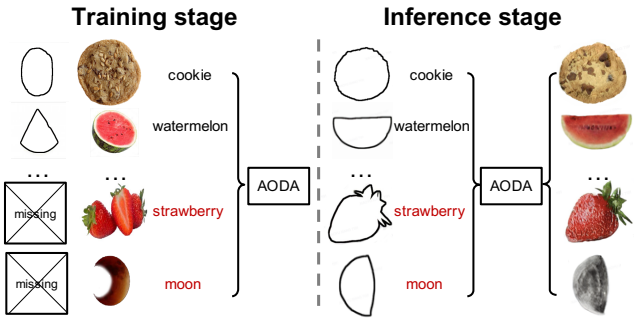

Figure 1: Illustration of open-domain sketch-to-photo synthesis problem. During the training stage of multi-class sketch-to-photo generation, sketches of some categories are missing. In the inference stage, our algorithm synthesizes photos from the input sketches for not only known classes, but also the classes that were missing during the training.

1. Introduction

Sketch-to-photo translation aims to automatically translate a sketch in the source domain S to the target photo-realistic domain P. However, the sketch-to-photo translation task suffers from the open-domain adaptation problem, where the majority of data is unlabeled and unpaired [5, 21, 17, 9, 34, 2, 18], and the freehand sketch covers only a small portion of the photo categories [26, 31, 28, 23, 7] due to the fact that they require a large number of human annotations. Therefore, some works [12, 19, 3, 20] use edges extracted from the target photos as substitution. Still, edges and freehand sketches are very different: freehand sketches are human abstractions of an object, usually with more deformations. Due to this domain gap, models trained on the edge inputs easily fail to generalize to freehand sketches.

Well-labeled freehand sketches and photos can help the translation model better understand the geometry correspondence. [33, 22, 11, 16, 14, 23] aim to learn from unpaired sketches and photos collected separately. Even so, the existing sketch datasets cannot cover all types of photos in the open domain [24]: the largest sketch dataset *Quick, Draw!* [9] has 345 categories, while the full ImageNet [4] has as many as 21,841 class labels. Therefore, most categories even lack corresponding freehand sketches to train a sketch-to-image translation model.

To resolve this challenging task, we propose an Adversarial Open Domain Adaption (AODA) framework that for the first time learns to synthesize the absent freehand sketches and makes the unsupervised open-domain adaption possible, as illustrated in Figure 1. We propose to jointly learn a sketch-to-photo translation network and a photo-to-sketch translation network for mapping the open-domain photos into the sketches with the GAN priors. With the bridge of the photo-to-sketch generation, we can generalize the learned correspondence between in-domain freehand sketches and photos to open-domain categories. Still, there is an unignorable domain gap between synthesized sketches and real ones, which prevents the generator from generalizing the learned correspondence to real sketches and synthesizing realistic photos for open-domain classes. To further mitigate its influence on the generator and leverage the output quality of open-domain translation, we introduce a simple yet effective random-mixed sampling strategy that considers a certain proportion of fake sketches as real ones blindly for all categories. With the proposed framework and training strategy, our model is able to synthesize a photo-realistic output even for sketches of unseen classes. We compare the proposed AODA to existing unpaired sketch-to-image generation approaches. Both qualitative and quantitative results show that our proposed method achieves significantly superior performance on both seen and unseen data. Our source code is available at https://github.com/Mukosame/AODA.

2. Adversarial Open Domain Adaption

Framework. Our AODA framework is trained with the unpaired sketch and photo data. During the training process, G_s extracts the sketch $G_s(p)$ from the given photo p. Then, the synthesized sketch $G_s(p)$ and the real sketch s are sent to G_p along with their labels η_p and η_s, and turned into the reconstructed photo $G_p(G_s(p), \eta_p)$ and the synthesized photo $G_p(s, \eta_s)$, respectively. Note that we only send the sketch with its true label to ensure that G_p learns the cor-

*This work was done as a part of internship at ByteDance.

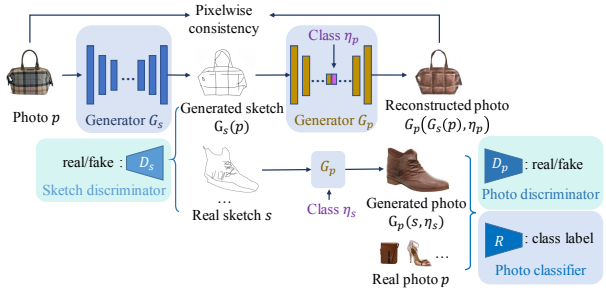

Figure 2: Our framework has two generators G_s : photo $\rightarrow$ sketch and G_p : sketch $\rightarrow$ photo conditioned on the input label, and two discriminators D_s and D_p for the sketch and photo domains, respectively. In addition, we use a photo classifier R to encourage G_p to generate indistinguishable photos from the real ones of the same class.

rect shape rectification from sketch to image domain for each class. The reconstructed photo is supposed to look similar to the original photo, which is imposed by a pixel-wise consistency loss. We do not add such a consistency constraint onto the sketch domain since we wish the synthesized sketches to be diverse. The generated photo is finally sent to the discriminator D_p to ensure it is photo-realistic, and the classifier R to ensure it has the same perceptual features as the target class. In summary, the generator loss includes four parts: the adversarial loss of photo-to-sketch generation $\mathcal{L}_{G_s}$, the adversarial loss of sketch-to-photo translation $\mathcal{L}_{G_p}$, the pixel-wise consistency of photo reconstruction $\mathcal{L}_{pix}$, and the classification loss for the synthesized photo $\mathcal{L}_{\eta}$:

$$\mathcal{L}_{GAN} = \lambda_s \mathcal{L}_{G_s}(G_s, D_s, p) + \lambda_p \mathcal{L}_{G_p}(G_p, D_p, s, \eta_s) + \lambda_{pix} \mathcal{L}_{pix}(G_s, G_p, p, \eta_p) + \lambda_{\eta} \mathcal{L}_{\eta}(R, G_p, s, \eta_s). \quad (1)$$

However, if we directly train the multi-class generator with the loss defined in Equation 1, the training objectives for open-domain classes $\mathcal{M}$ become the following form due to the missing sketches s:

$$\mathcal{L}^{\mathcal{M}}_{GAN} = \lambda_s \mathcal{L}_{G_s}(G_s, D_s, p) + \lambda_{pix} \mathcal{L}_{pix}(G_s, G_p, p, \eta_p), \quad (2)$$

where $\eta_p \in \mathcal{M}$. As a result, the sketch-to-photo generator G_p is solely supervised by the pixel-wise consistency. Since the commonly used $\mathcal{L}_1$ and $\mathcal{L}_2$ loss lead to the median and mean of pixels, respectively, this bias will make G_p generate blurry photos for the open-domain classes.

To solve this problem, we propose the random-mixed sampling strategy to minimize the domain gap between real and fake sketch inputs for the generator and improve its output quality with the open-domain classes.

Random-mixed strategy. This strategy aims to "fool" the generator into treating fake sketches as real ones. Algorithm 1 describes the detailed steps for the random-mixed sampling and modified optimization: *Pool* denotes the buffer that stores the minibatch of sketch-label pairs. Querying the pool returns either the current minibatch or a previously stored one (and inserts the current minibatch in the pool) with a certain likelihood. U denotes uniform sampling in the given range, and t denotes the threshold that is set according to the ratio of open-domain classes and in-domain classes to match the photo data distribution.

Algorithm 1: 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**
  $s_{fake} \leftarrow G_s(p)$;
  $s_c \leftarrow s$;
  $\eta_c \leftarrow \eta_s$;
  if $t < u \sim U(0, 1)$ **then**
    $s_c, \eta_c \leftarrow pool$.query$(s_{fake}, \eta_p)$;
  end
  $p_{rec} \leftarrow G_p(s_{fake}, \eta_p)$;
  $p_{fake} \leftarrow G_p(s, \eta_s)$
  Calculate $\mathcal{L}_{GAN}$ with $(p, s_c, p_{rec}, \eta_c)$ and update G_s and G_p;
  Calculate $\mathcal{L}_{D_s}(s, s_{fake})$ and $\mathcal{L}_{D_s}(p, p_{fake})$, update D_s and D_p;
  Calculate $\mathcal{L}_R(p, p_{fake}, \eta_p, \eta_s)$ and update the classifier.
end

One key operation of this strategy is to construct pseudo sketches for G_p by randomly mix the synthesized sketches with real ones in a batch-wise manner. In this step, the pseudo sketches are treated as the real ones by the generator. Thus, the open-domain classes' $\mathcal{L}^{\mathcal{M}}_{GAN}$ becomes:

$$\mathcal{L}^{\mathcal{M}}_{GAN} = \lambda_s \mathcal{L}_{G_s}(G_s, D_s, p) + \lambda_p \mathcal{L}_{G_p}(G_p, D_p, s_{fake}, \eta_p) + \lambda_{pix} \mathcal{L}_{pix}(G_s, G_p, p, \eta_p) + \lambda_{\eta} \mathcal{L}_{\eta}(R, G_p, s_{fake}, \eta_p), \quad (3)$$

where $\eta_p \in \mathcal{M}$. Another key of the strategy is on optimization: the sampling strategy is only for G_p. The classifier and discriminators are still updated with real/fake data to guarantee their discriminative powers.

The random mixing operation is blind to in-domain and open-domain classes. As a result, the training sketches include both real and pseudo sketches from all categories. By including pseudo sketches from both the in-domain and open-domain classes, it would further enforce the sketch-to-image generator to ignore the domain gap in the inputs and synthesize realistic photos from both real and fake sketches. Note that since G_s's parameters are consistently updated during training, the pseudo sketches also change for each batch. Moreover, the pseudo sketch-label pairs are acquired

from a history of generated sketches and their labels rather than the latest produced ones by G_s. We maintain a buffer that stores the 50 previously added minibatch of sketch-label pairs [27, 33].

Mixing real sketches with fake ones can be regarded as an online data augmentation technique for training G_p. Compared with augmentation using edges, G_s can learn the distortions from real freehand sketches by approaching the real data distribution [8, 15, 32], and enable G_p to learn shape rectification on the fly. Benefiting from the joint training mechanism, as the training progresses, the sketches generated by G_s change epoch by epoch. The loose consistency constraint on sketch generation further increases the diverseness of the sketch data in the open-domain. Compared with using pre-extracted sketches, the open-domain buffer maintains a broad spectrum of sketches: from the very coarse ones generated in early epochs to the more human-like sketches in later epochs as G_s converges.

3. Experiments

We train and evaluate the performance of sketch-to-photo synthesis methods on two datasets: Scribble [7], and SketchyCOCO [6]. Scribble contains ten object classes with photos and simple outline sketches. Six out of ten classes have similar round outlines, which imposes more stringent requirements on the network: whether it can generate the correct structure and texture conditioned on the input class label. In the open-domain setting, we only have the sketches of four classes for training: *pineapple*, *cookie*, *orange*, and *watermelon*, which means that 60% classes are open-domain. SketchyCOCO includes 14 object classes. The 14,081 photos for each object class are segmented from the natural images of COCO Stuff [2] under unconstrained conditions, thereby making it more difficult for existing methods to map the freehand sketches to the photo domain. The two open-domain classes are: *sheep* and *giraffe*.

We quantitatively evaluate the generation results with Fréchet Inception Distance (FID) [10], classification Accuracy (Acc) [1] and user Preference Study (Human): we show the participants a given sketch and the class label, and ask them to pick one photo with the best quality and realism from generated results. We randomly sample 31 groups of images. For each evaluation, we shuffle the options and show them to 25 users. We collect 775 answers in total.

Sketch-to-Photo Synthesis. To better illustrate the effectiveness of our proposed solution, here we adopt CycleGAN [33] as the baseline in building our network and include the original CycleGAN in the following comparison. To make it able to accept sketch class labels, we modified the sketch-to-photo translator to be a conditional generator. Besides, we also compare a recent work EdgeGAN [6] on each dataset. We mark the open-domain sketch with a ★ for better visualization.

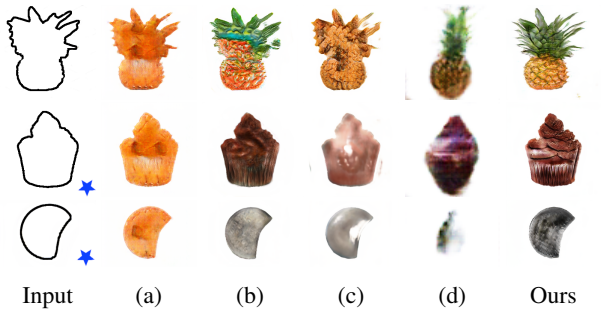

Figure 3: Results on Scribble dataset [7]. The following columns are outputs of (a) CycleGAN [33], (b) conditional CycleGAN, (c) classifier+(b), (d) EdgeGAN [6], and ours.

Figure 3 shows the qualitative results on Scribble dataset: (a) CycleGAN, (b) conditional CycleGAN, (c) conditional CycleGAN with classification loss, (d) EdgeGAN and our method, where the bottom two rows are open-domain. The original CycleGAN exhibits mode collapse and synthesizes identical textures for all categories, probably due to the fact that the sketches in the Scribble dataset barely imply their class labels. This problem is alleviated in (b). Still, it fails to synthesize natural photos for some categories due to the gap between open-domain and in-domain data. Such a domain gap is even worse in (c), where the in-domain result is with realistic but wrong texture, and the open-domain results are texture-less. This might be because that classifier implicitly increases the domain gap while maximizing the class discrepancy. Thus, we do not include this model for comparison on the other two datasets. Compared with (d), our results are more consistent with the input sketch shape, demonstrating that our model is better at understanding the composition in sketches and learning more faithful shape rectification in sketch-to-photo domain mapping.

The qualitative results of SketchyCOCO are shown in Figure 4, where the top two rows are of in-domain categories, and the bottom two are open-domain. The photos generated by CycleGAN suffer from mode collapse. As shown in column (b), conditional CycleGAN cannot generate vivid textures for open-domain categories. Compared with EdgeGAN in (c), the poses in our generated photos are more faithful to the input sketches.

The quantitative results for the three datasets are summarized in Table 1. We can see that our model is preferred by more users than the other compared methods, and achieves the best results in terms of the FID score and classification accuracy on all datasets. These results confirm our observations of the qualitative outputs, as discussed above. Besides, we have an interesting observation: compared with the baseline CycleGAN and conditional CycleGAN, our random-mixed strategy improves not only the open-domain results, but also in-domain results. We find a possible explanation from [29]: the "fake-as-real" operation can effectively alleviate the gradient exploding issue during GAN

Dataset	Method	CycleGAN [33]			conditional CycleGAN			EdgeGAN [6]			Ours		
	Metric	full	in-domain	open-domain	full	in-domain	open-domain	full	in-domain	open-domain	full	in-domain	open-domain
Scribble	FID ↓	279.5	252.7	355.9	213.6	210.9	253.6	259.7	256.3	298.5	**209.5**	**204.6**	**252.8**
	Acc (%) ↑	16.0	30.0	6.7	68.0	70.0	66.7	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
	Human (%) ↑	5.60	1.00	8.67	19.20	17.00	20.67	25.20	17.00	30.67	**48.80**	**65.00**	**38.00**
SketchyCOCO	FID ↓	201.7	218.7	237.2	124.3	138.7	171.6	169.7	177.8	221	**114.8**	**128.4**	**139.2**
	Acc (%) ↑	8.4	10.8	1.9	57.0	58.7	52.4	75.8	68.8	98.3	**78.3**	**70.5**	**100.0**
	Human (%) ↑	0.36	0.00	0.67	5.09	5.60	4.67	22.55	32.00	14.67	**72.00**	**59.20**	**82.67**

Table 1: Quantitative evaluation and user study on Scribble and SkethyCOCO datasets. We show the full testset results, in-domain results, and open-domain results, respectively. Best results are shown in **bold**.

Figure 4: Results on SketchyCOCO dataset [6] for the compared methods: (a) CycleGAN [33], (b) conditional CycleGAN, (c) EdgeGAN [6], and ours.

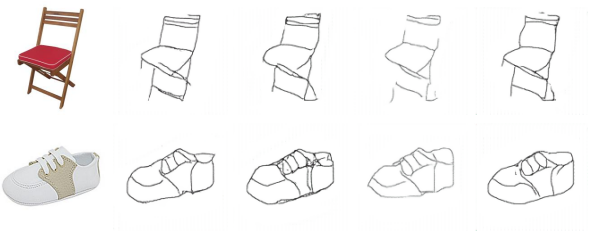

Figure 5: Photo-based sketch synthesis results at different epochs. Given a photo input shown in the first column, the synthesized sketches from our photo-to-sketch generator change at different epochs.

training and result in a more faithful generated distribution.

Photo-to-Sketch Synthesis. As a byproduct, our network can also provide a high-quality freehand sketch generator G_s for a given photo [25, 30, 13]. Characterized by the joint training, the weights of the photo-to-sketch generator are constantly updated as the training progresses. So the sketches generated by G_s change epoch by epoch. Figure 5 shows the extracted sketches at different epochs. The changing sketches increase the diverseness of the sketch, thus can further augment the data and help the generator to better generalize to various freehand sketch inputs.

Ablation Study. To illustrate the effect of the proposed open-domain training strategy, we simplify the dataset to

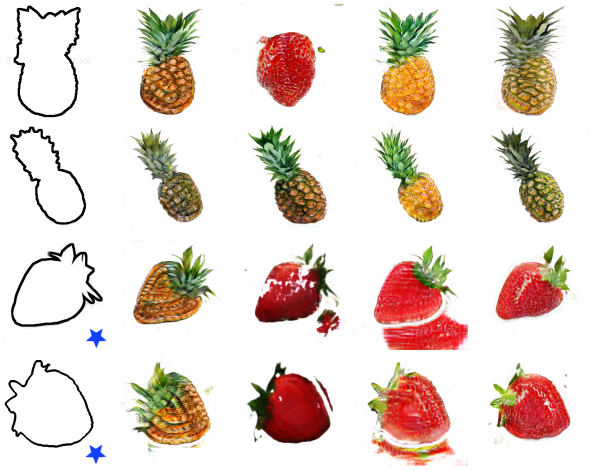

Figure 6: Ablation study of the proposed solution. (a): baseline without classifier or strategy; (b): our framework without strategy; (c) trained with pre-extracted open-domain and real in-domain sketches; (d): random-mixed sampling strategy.

two classes, including the in-domain class *pineapple* and the open-domain class *strawberry*. We compare four models: (a) baseline CycleGAN without classifier or strategy; (b) AODA framework without applying any strategy; (c) AODA trained with synthesized open-domain sketches and real in-domain sketches; (d) AODA trained with the random-mixed sampling strategy. From Figure 6, we can see that the base model in column (a) translates all inputs to the in-domain category; (b) generates texture-less images with correct colors for the open-domain class due to the pixel-wise consistency, as discussed in Equation 2. For in-domain sketches, it generates photo-realistic outputs with the shape and texture of any category, which indicates that the model associates the class label with real/fake sketches, and thus fails to generalize to open-domain data. For column (c), the model trained with fake open-domain sketches can barely generate realistic textures for *strawberries*. Besides, it fails to distinguish the object region from the background due to the weak generalization capability, as the extracted sketches actually impair the discriminative power of D_s. Column (d) shows that our open-domain sampling and training strategy can alleviate the above issues, and bring superior performance for the multi-class generation.

References

[1] Oron Ashual and Lior Wolf. Specifying object attributes and relations in interactive scene generation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4561–4569, 2019. 3

[2] Holger Caesar, Jasper Uijlings, and Vittorio Ferrari. Cocostuff: Thing and stuff classes in context. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1209–1218, 2018. 1, 3

[3] Shu-Yu Chen, Wanchao Su, Lin Gao, Shihong Xia, and Hongbo Fu. Deepfacedrawing: deep generation of face images from sketches. *ACM Transactions on Graphics (TOG)*, 39(4):72–1, 2020. 1

[4] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. IEEE, 2009. 1

[5] Mathias Eitz, James Hays, and Marc Alexa. How do humans sketch objects? *ACM Transactions on Graphics (TOG)*, 31(4):1–10, 2012. 1

[6] Chengying Gao, Qi Liu, Qi Xu, Limin Wang, Jianzhuang Liu, and Changqing Zou. Sketchycoco: Image generation from freehand scene sketches. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5174–5183, 2020. 3, 4

[7] Arnab Ghosh, 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 International Conference on Computer Vision*, pages 1171–1180, 2019. 1, 3

[8] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, pages 2672–2680, 2014. 3

[9] David Ha and Douglas Eck. A neural representation of sketch drawings. *arXiv preprint arXiv:1704.03477*, 2017. 1

[10] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Advances in Neural Information Processing Systems*, pages 6626–6637, 2017. 3

[11] Xun Huang, Ming-Yu Liu, Serge Belongie, and Jan Kautz. Multimodal unsupervised image-to-image translation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 172–189, 2018. 1

[12] Phillip Isola, 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*, pages 1125–1134, 2017. 1

[13] Moritz Kampelmuhler and Axel Pinz. Synthesizing human-like sketches from natural images using a conditional convolutional decoder. In *The IEEE Winter Conference on Applications of Computer Vision*, pages 3203–3211, 2020. 4

[14] Junho Kim, Minjae Kim, Hyeonwoo Kang, and Kwanghee Lee. U-gat-it: unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation. *arXiv preprint arXiv:1907.10830*, 2019. 1

[15] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4681–4690, 2017. 3

[16] Hsin-Ying Lee, Hung-Yu Tseng, Jia-Bin Huang, Maneesh Singh, and Ming-Hsuan Yang. Diverse image-to-image translation via disentangled representations. In *Proceedings of the European conference on computer vision (ECCV)*, pages 35–51, 2018. 1

[17] Da Li, 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*, pages 5542–5550, 2017. 1

[18] Mengtian Li, Zhe Lin, Radomir Mech, Ersin Yumer, and Deva Ramanan. Photo-sketching: Inferring contour drawings from images. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1403–1412. IEEE, 2019. 1

[19] Yuhang Li, Xuejin Chen, Feng Wu, and Zheng-Jun Zha. Linestofacephoto: Face photo generation from lines with conditional self-attention generative adversarial networks. In *Proceedings of the 27th ACM International Conference on Multimedia*, pages 2323–2331, 2019. 1

[20] Yuhang Li, Xuejin Chen, Binxin Yang, Zihan Chen, Zhihua Cheng, and Zheng-Jun Zha. Deepfacepencil: Creating face images from freehand sketches. In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 991–999, 2020. 1

[21] Yi Li, Timothy M. Hospedales, Yi-Zhe Song, and Shaogang Gong. Fine-grained sketch-based image retrieval by matching deformable part models. In *In British Machine Vision Conference (BMVC)*, 2014. 1

[22] Ming-Yu Liu, Thomas Breuel, and Jan Kautz. Unsupervised image-to-image translation networks. In *Advances in Neural Information Processing Systems*, pages 700–708, 2017. 1

[23] Runtao Liu, Qian Yu, and Stella Yu. Unsupervised sketch-to-photo synthesis. *arXiv preprint arXiv:1909.08313*, 2019. 1

[24] Pau Panareda Busto and Juergen Gall. Open set domain adaptation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 754–763, 2017. 1

[25] Kaiyue Pang, Da Li, Jifei Song, Yi-Zhe Song, Tao Xiang, and Timothy M Hospedales. Deep factorised inverse-sketching. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 36–52, 2018. 4

[26] Patsorn Sangkloy, Nathan Burnell, Cusuh Ham, and James Hays. The sketchy database: learning to retrieve badly drawn bunnies. *ACM Transactions on Graphics (TOG)*, 35(4):1–12, 2016. 1

[27] Ashish Shrivastava, Tomas Pfister, Oncel Tuzel, Joshua Susskind, Wenda Wang, and Russell Webb. Learning from simulated and unsupervised images through adversarial training. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2107–2116, 2017. 3

[28] Jifei Song, Qian Yu, Yi-Zhe Song, Tao Xiang, and Timothy M Hospedales. Deep spatial-semantic attention for fine-grained sketch-based image retrieval. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 5551–5560, 2017. 1

[29] Song Tao and Jia Wang. Alleviation of gradient exploding in gans: Fake can be real. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1191–1200, 2020. 3

[30] Ran Yi, Yong-Jin Liu, Yu-Kun Lai, and Paul L Rosin. Apdrawinggan: Generating artistic portrait drawings from face photos with hierarchical gans. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 10743–10752, 2019. 4

[31] Qian Yu, Feng Liu, Yi-Zhe SonG, Tao Xiang, Timothy Hospedales, and Chen Change Loy. Sketch me that shoe. In *Computer Vision and Pattern Recognition*, 2016. 1

[32] Ruofan Zhou and Sabine Susstrunk. Kernel modeling super-resolution on real low-resolution images. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2433–2443, 2019. 3

[33] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2223–2232, 2017. 1, 3, 4

[34] Changqing Zou, Qian Yu, Ruofei Du, Haoran Mo, Yi-Zhe Song, Tao Xiang, Chengying Gao, Baoquan Chen, and Hao Zhang. Sketchyscene: Richly-annotated scene sketches. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 421–436, 2018. 1