

Semi Few-Shot Attribute Translation

Ricard Durall^{1,2,3}

Franz-Josef Pfrendt¹

Janis Keuper^{1,4}

¹Fraunhofer ITWM, Germany

²IWR, University of Heidelberg, Germany

³Fraunhofer Center Machine Learning, Germany

⁴Institute for Machine Learning and Analytics, Offenburg University, Germany

Abstract

Recent studies have shown remarkable success in image-to-image translation for attribute transfer applications. However, most of existing approaches are based on deep learning and require an abundant amount of labeled data to produce good results, therefore limiting their applicability. In the same vein, recent advances in meta-learning have led to successful implementations with limited available data, allowing so-called *few-shot learning*.

In this paper, we address this limitation of supervised methods, by proposing a novel approach based on GANs. These are trained in a meta-training manner, which allows them to perform image-to-image translations using just a few labeled samples from a new target class. This work empirically demonstrates the potential of training a GAN for few shot image-to-image translation on hair color attribute synthesis tasks, opening the door to further research on generative transfer learning.

1 Introduction

Deep learning models have achieved state-of-the-art performance in large variety of tasks, from image synthesis [41, 17, 3], text style transfer [34], video generation [2, 38] to image-to-image translation [30, 15, 44, 5, 27]. The latter task, image-to-image translation, is a computer vision problem that aims at translating images from one domain to another, including colorization [42], super-resolution [23], style transfer [16, 43, 44, 14], inpainting [30, 39, 24, 15, 40] and attribute transfer [25, 18, 8, 5]. The strong performance, however, heavily relies on training a network with abundant labeled instances with diverse visual variations. The hu-

man annotation cost as well as the scarcity of data in some classes significantly limit the applicability of current vision systems to learn new visual concepts efficiently. In contrast, the human visual systems can recognize new instances with extremely few labeled examples. It is thus of great interest to learn to generalize to new cases with a limited amount of labeled examples for each novel case.

The problem of learning to generalize to unseen classes during training, known as few-shot classification, has attracted considerable attention [37, 31, 10, 35, 36]. One promising direction to few-shot classification is the meta-learning paradigm where transferable knowledge is extracted and propagated from a collection of tasks to prevent overfitting and improve generalization. Recent advances in meta-learning algorithm includes metric-based methods [20, 37, 35, 36], model-based methods [33, 28] and optimization-based methods [31, 10, 1, 29, 11]. These models have allowed learning tasks to perform well on novel data sampled from the same distribution as the training data. These meta-learning algorithms have seen direct applications in supervised and reinforcement learning. Additionally, due to their general applicability, recent works based on meta-learning have successfully been utilized for image generation [21, 32, 4, 7].

The objective of attribute transfer is to synthesize realistic appearing images for a pre-defined target domain. For instance, given an image with a particular attribute "blond hair" (original domain), change it to "black hair" (target domain). We refer to a domain as a set of images sharing the same attributes. Such attributes are meaningful semantic feature inherent in an image such as "smiling" or "face with eyeglasses". After the introduction of generative adversarial networks (GANs) [12], transfer domain algorithms have experienced significant improvements achieving state-of-the-

art results in style transfer [16, 43, 44, 14] and in attribute transfer [18, 8, 5]. However, GANs require several orders of magnitude more data points than humans in order to generate comprehensible images successfully from a given class of images [7]. This impairs the ability of GANs to generate novelty. Additionally, in many cases, if the data is abundant enough to successfully train a GAN, there is little purpose to generating more of this data.

In this work we focus on the challenging scenario, where we define the problem of semi few-shot image-to-image translation. In particular, we propose a novel approach capable of performing attribute transfer on a target domain with a very limited amount of labeled data. In order to achieve this goal, we use two independent but equal networks and we train them as proposed in [29]. Most of the existing few-shot algorithm are applied to classification tasks where one class remains unseen. In our approach, we apply the same principle but with attributes and we name it semi few-shot because there are not different classes. Nevertheless, having only one kind of images but with different attributes does not make the problem trivial since still the network needs to learn the ability to perform image-to-image transformation for untrained target domain with a few examples. We apply our model to the CelebA [26] dataset of faces and control several hair color attributes.

Overall, our contributions are summarized as follows

- We propose a novel generative adversarial network based on meta-learning, trained for end-to-end image-to-image translation.
- We demonstrate, how we can successfully learn to transfer hair attributes by using a generative few-shot approach. These first results are opening the door to further research.
- We provide qualitative results based on CelebA dataset, showing the effectiveness of our proposal.

2 Related Work

While machine learning systems might have surpassed humans at many tasks [9], they generally need far more data to reach the same level of performance. Nonetheless, it is not completely fair to compare humans to algorithms directly, since humans enter a task with a large amount of prior knowledge. In other words, humans do not learn from scratch, but they fine-tune and recombine sets of pre-existing skills. Meta-learning has emerged recently as an approach for learning from small amounts of data as human do. This machine learning field has already been employed in many different domains such as classification, reinforcement learning and even image generation.

2.1 Meta-learning

Meta-learning, also known as few-shot learning, intends to design models that can learn new skills or adapt to new environments rapidly with a few training examples. Specifically, we assume to have access to a problem, which is split into a set of tasks. Each of these tasks can be for example: a set of images (belonging to the same class) for a classification problem, or a set of state, action and reward for a reinforcement algorithm, or even an attribute for an image-to-image transformation. Then, from this problem, we sample a training set and a test set of tasks, and we feed the training set into our algorithm, so that, eventually it will produce good performance on the test set. Since each task corresponds to a learning problem, performing well on a task corresponds to learning quickly.

A variety of different approaches to meta-learning have been proposed, each with its own flavors. There are three common approaches: metric-based that learns an efficient distance metric [20, 37, 35, 36], model-based that uses network with external or internal memory [33, 28] and optimization-based that optimizes the model parameters explicitly for fast learning [31, 10, 1, 29, 11].

Optimization-based. Optimization-based models, also known as initialization-based methods, tackle the meta-learning problem by "learning to fine-tune". The idea behind this approach is to learn a good model initialization (i.e. the weights) so that in the case of a classification task, given an unseen class, the model can classify correctly using only a limited number of labeled examples and a small number of gradient update steps [31, 10, 1, 29, 11]. These initialization based methods are capable of achieving fast and effective adaption with a limited number of training examples for new classes, however they still have difficulty in handling domain shifts between base and novel classes.

Few-Shot Image Generation. Most of meta-learning applications focus on classification tasks defined as the ability to learn a classifier to recognize unseen classes during training with limited labeled examples. However, it is possible to extend such a definition of few-shot learning to other domains such as image generation. To best of our knowledge, [21] pioneered the successful combination of few-shot techniques with image generation. In this first approach, the model is fed with images and their strokes, and trained on the Omniglot dataset [21]. This yields a system that can generate novel binary samples. Trying to make a more general approach, [32] presents a sequential generative model which is only trained on pure images (no stroke information is required). Even though, this second approach improves previous results, it suffers from lengthy sequential inference as a consequence. [4] tackles this issue by suggesting matching networks (memory-assisted networks). This im-

plementation generates binary images on Omniglot dataset using few-shot learning and with fast inference periods. Finally, [7] proposes a new approach integrating GANs in the structure, surpassing in this way, the scalability limitation found in the other models. Furthermore, this work presents more extensive experiments including additional datasets (MNIST [22] and FIGR-8 [7]).

2.2 Generative Adversarial Network

Generative adversarial network [12] is capable of learning deep generative models. It can be described as a min-max game between the generator G , which learns how to generate samples which resemble real data, and a discriminator D , which learns to discriminate between real and fake data. Throughout this process, G indirectly learns how to model the input image distribution p_{data} by taking samples z from a fixed distribution p_z (e.g. Gaussian) and forcing the generated samples $G(z)$ to match the natural images x . The objective loss function is defined as

$$\min_G \max_D \mathcal{L}(D, G) = \mathbb{E}_{\mathbf{x} \sim p_{data}} [\log(D(\mathbf{x}))] + \mathbb{E}_{\mathbf{z} \sim p_z} [\log(1 - D(G(\mathbf{z})))] \quad (1)$$

Domain Transfer. GAN-based approaches for image-to-image translation have been actively studied. One of the first proposals [16] capable of learning consistent image domain transforms, employed a pair of images that could be used to create models that convert input from an original domain to different target domain (e.g. segmentation labels to the original image). But this system requires that both images and target images must exist as pairs in the training dataset in order to learn the transformation between domains. Several works [43, 5] try to address this drawback. Their suggestion is to use the virtual result in the target domain. Therefore, if the virtual result is inverted again, the inverted result must match with the original image. In these works, the framework can control the image translation into different target domains. Recently, numerous works have focused on transferring visual attributes such as color [42], texture [16, 43, 44, 14], facial features [25, 18, 8, 5] and more. However, although most of these approaches synthesize new images that belong to the target domain, they lack in generalizing attributes since they are designed to transfer a specific type of visual attribute.

3 Method

In the following section, we describe our approach which addresses image-to-image translation, where the input image contains the original attributes and the output image the

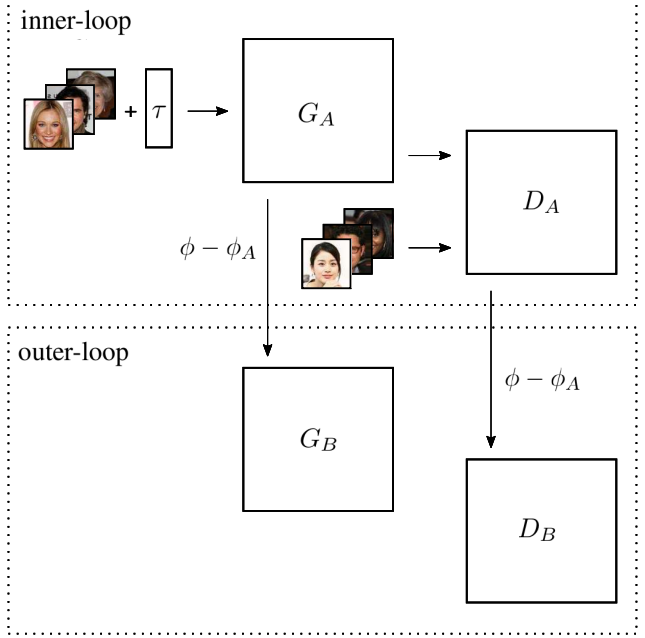


Figure 1: The figure shows the few-shot attribute transfer structure, which consists of two distinguishable parts: inner-loop and outer-loop. At the same time, each of this blocks is a GAN system formed for a generator G and a discriminator D . The inner-loop optimizes G_A using samples A and the task τ judged by D_A . After k updates, we update the outer-loop by setting its gradient to $\phi - \phi_A$ and performing one step of optimizer.

target attributes. We explain the training of the model in two levels of abstraction that train independently, following an adversarial fashion which allows generating realistic samples containing the target attributes.

3.1 Meta-learning Model Architecture

We define the few-shot attribute transfer problem $P(\mathcal{T})$ based on the meta-learning reptile algorithm introduced in [7]. In this scenario, we assume a set of tasks $\{\mathcal{T}_i\}_{i=1}^M$, where each individual task τ is an attribute transfer problem with its corresponding loss L_τ . The intuition behind L_τ is that it accounts for the ability of generating realistic samples. We approach our problem by defining a meta-learning model that optimizes towards convergence, by modifying the neural network parameters ϕ within a limited k updates. Therefore, we can define the minimization problem as

$$\min_{\phi} \mathbb{E}_{\tau} [L_{\tau}(U_{\tau}^k(\phi))], \quad (2)$$

where $U_{\tau}^k(\phi)$ is the operator (usually stochastic gradient

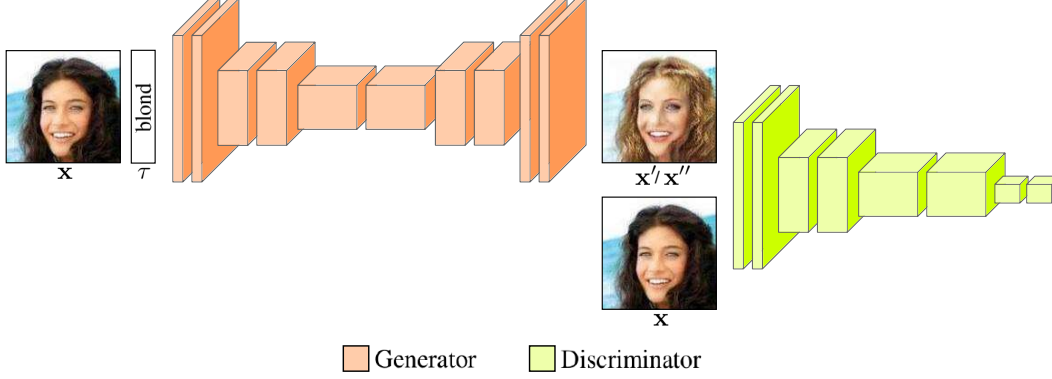


Figure 2: Overview of the attribute transfer network structure while training. Our model contains a generator G and a discriminator D , which are trained independently. For example, given an input image and a target domain label (blond), G learns to generate images within the target domain. At the same time, D learns to distinguish between real and generated images, and to classify them to their corresponding domain.

descent) that updates ϕ parameters k times using data sampled from τ .

In order to solve Equation 2, we employ a solution composed of two independent neural structures, but with the same topology, called outer-loop structure and an inner-loop structure. Given a task τ and the initial parameters ϕ , the inner-loop optimization trains on a set of samples, while updating a copy of the parameters ϕ_A . After k updates, we optimize the outer-loop, where we set the gradient of ϕ_B to be equal to $\phi - \phi_A$, take one step on the optimizer and compute the loss. In this way, the algorithm optimizes for generalization. We write it as

$$\min_{\phi} \mathbb{E}_{\tau} [L_{\tau}(U_{\tau,A}^k(\phi_A))]. \quad (3)$$

Notice that this dual structure gives enough flexibility to cope with different problems using the same framework. In this work, we make use of such a generalization feature by implementing a GAN system inside of each loop (see Figure 1). For a given image \mathbf{x} and a target domain label τ , our goal is to translate \mathbf{x} into an output image y , which now belongs to the target domain. In other words, we can transfer attributes by generating synthetic data in a meta-learning fashion.

3.2 Attribute Transfer Model Architecture

We construct our baseline attribute transfer generative network based on recent state-of-the-art image-to-image translation model [5] which is an adaptation from [43]. In particular, our approach employs an adversarial structure made of a generator and a discriminator. By combining these elements, our model can self-estimate the difference

between the generated samples (with attribute transfer) and the real samples, and then update itself to produce more realistic samples. The network architecture of our proposal is shown in Figure 2.

Generator. The goal of the generator G is to learn mappings among multiple attribute domains. To achieve this goal, we train G to translate \mathbf{x} into an output image \mathbf{x}' conditioned on the target domain label τ . To achieve this objective, the generator loss consists of $\mathcal{L}_{\text{disc}}$ that penalizes inappropriately generated images and $\mathcal{L}_{\text{cycle}}$ that guarantees that translated images \mathbf{x}' preserve the content of its input images \mathbf{x} while changing only the domain-related part of the inputs. We write the generator loss as

$$\mathcal{L}_{\text{gen}} = \mathcal{L}_{\text{disc}} + \lambda_{\text{cycle}} \mathcal{L}_{\text{cycle}}. \quad (4)$$

Notice that the generator performs an entire cyclic translation $\mathbf{x} \rightarrow \mathbf{x}' \rightarrow \mathbf{x}''$ for every sample, forcing τ to be crucial for moving among domains. First, to translate an original image \mathbf{x} into an image in the target domain \mathbf{x}' and then to reconstruct the original image from the translated image \mathbf{x}'' . This procedure can be written as

$$\mathcal{L}_{\text{cycle}} = \|\mathbf{x} - \mathbf{x}''\|_1 \quad (5)$$

where

$$\begin{aligned} \mathbf{x}' &= G(\mathbf{x}, \tau_{\text{target}}) \\ \mathbf{x}'' &= G(\mathbf{x}', \tau_{\text{original}}). \end{aligned} \quad (6)$$

Discriminator. The second element of the model is the discriminator D . It takes samples of true and generated data

and tries to classify them correctly. This classification procedure takes place after reconstruction and generation tasks, and it consists of two parts. One part that implements \mathcal{L}_{adv} , which employs Wasserstein distance to determine if an image looks realistic and penalize it otherwise, and a second part similarly to other approaches like [6], where we have an additional loss function $\mathcal{L}_{\text{class}}$ which accounts for domain classification. This term computes the binary cross entropy loss between $\tau_{\text{generated}}$ and τ_{target} penalizing incorrect image-domain transformations. Therefore, the discriminator loss can be defined as

$$\mathcal{L}_{\text{disc}} = \lambda_{\text{adv}} \mathcal{L}_{\text{adv}} + \lambda_{\text{class}} \mathcal{L}_{\text{class}}. \quad (7)$$

4 Experiments

In this section, we present experimental results evaluating the effectiveness of the proposed method. First, we give a detailed introduction of the experimental setup. Then, we discuss the results and its possible interpretation.

4.1 Experimental Settings

We train our model on the CelebFaces Attributes (CelebA) dataset [26]. It consists of 202,599 celebrity face images with variations in facial attributes. For the experiments, since we focus on hair attributes, we select a set of 32,502 images containing a balanced amount of hair attributes (blond, black, brown and gray hair). It is indispensable to have such an even distribution, otherwise the algorithm might fail at transferring marginal attributes. We randomly select 2,000 images for testing and use all remaining images as the training dataset. In training, we crop and resize the initially 178x218 pixel image to 128x128 pixels. All experiments presented in this paper have been conducted on a single NVIDIA GeForce GTX 1080 GPU.

4.2 Training Setting

Since our model is divided into two distinguishable parts, two independent Adam optimizer [19] with $\beta_1 = 0.5$, $\beta_2 = 0.999$. are used during training. We set the batch size to 16 and run the experiments for 200K iterations. We update the generator after every five discriminator updates as in [13, 5]. The learning rate used in the implementation is 0.0001 for the first 10 epochs and then linearly decreased to 0 over the next 10 epochs. The losses are weighted by the factors: λ_{class} and λ_{cycle} set to 10, and λ_{adv} to 1. The inner-loop k_A runs 1 iteration and the outer-loop k_B 10 iterations. The procedure described above follows a first-order gradient-based meta-learning algorithm and its described in Algorithm 1.

Algorithm 1 Training of the proposed architecture model. All conducted experiments in the paper used the default values: $n_{\text{iter}} = 200,000$, $\alpha_{\text{disc}} = \alpha_{\text{gen}} = 0.0001$, $m = 16$, $k_A = 1$, $k_B = 10$, $n_{\text{gen}} = 5$.

```

1: Require:  $n_{\text{iter}}$ , number of iterations.  $\alpha$ 's, learning rate.
    $m$ , batch size.  $n_{\text{gen}}$ ,  $k_A$ , number of iterations inner-loop.
    $k_B$ , number of iterations outer-loop. Number of skipped iterations
   of the generator per discriminator iteration.
2: Require:  $\phi_0$ , initial generator and discriminator parameters.
3: Initialize  $\phi_B = \phi_0$ 
4: for  $i < n_{\text{iter}}$  do
5:   Sample a batch of images  $\{\mathbf{x}^{(z)}\}_{j=0}^m$ 
6:   Sample a task  $\tau$ 
7:   Initialize  $\phi_A = \phi_B$ 
8:   for  $j < k_A$  do
9:     # Train discriminator  $D_A$ 
10:     $\phi_{A,\text{disc}} \leftarrow \phi_{A,\text{disc}} + \alpha_{\text{disc}} \nabla_{\phi_A} \{\mathcal{L}_{\text{disc}}(\mathbf{x}, G_A(\mathbf{x}))\}$ 
11:    # Train generator  $G_A$ 
12:    if  $\text{mod}(i, n_{\text{gen}}) = 0$  then
13:       $\phi_{A,\text{gen}} \leftarrow \phi_{A,\text{gen}} + \alpha_{\text{gen}} \nabla_{\phi_A} \{\mathcal{L}_{\text{gen}}(\mathbf{x})\}$ 
14:    end if
15:  end for
16:  Set gradients of  $D_B$  and  $G_B$  to  $\phi_0 - \phi_A$ 
17:  Perform step on  $D_B$  and  $G_B$  optimizers
18: end for
19: # Train few-shot
20: for  $j < k_B$  do
21:   # Train discriminator  $D_B$ 
22:    $\phi_{B,\text{disc}} \leftarrow \phi_{B,\text{disc}} + \alpha_{\text{disc}} \nabla_{\phi_B} \{\mathcal{L}_{\text{disc}}(\mathbf{x}, G_B(\mathbf{x}))\}$ 
23:   # Train generator  $G_B$ 
24:   if  $\text{mod}(j, n_{\text{gen}}) = 0$  then
25:      $\phi_{B,\text{gen}} \leftarrow \phi_{B,\text{gen}} + \alpha_{\text{gen}} \nabla_{\phi_B} \{\mathcal{L}_{\text{gen}}(\mathbf{x})\}$ 
26:   end if
27: end for

```

4.3 Empirical Validation

In this subsection, we present an empirical study of the results of our proposed method. We validate, that the attribute in the generated image, continuously changes with the coding vector τ (task). This phenomenon is known as attribute transfer or morphing. In particular, we focus on hair color attributes: blond, black, brown and gray hair.

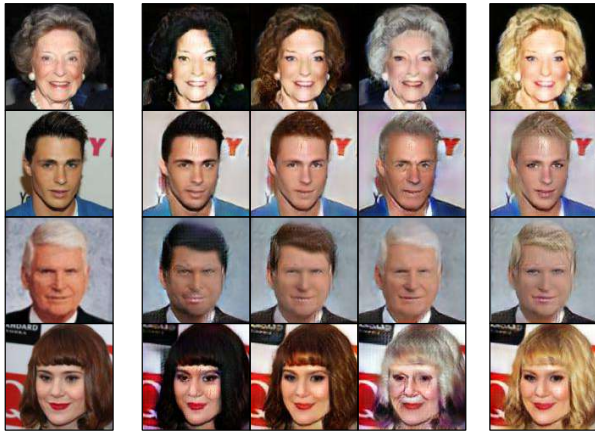
Figure 3 depicts four different experiments for hair attribute transfer. For example, Figure 3a shows the result of training the images using the tasks blond, black and gray, during the normal training and then apply few-shot learning for the task brown hair. By implementing this procedure, the algorithm can find really suitable initialization weights, so that, when we fine-tune for the new task (e.g. brown hair)



(a) Brown hair attribute transform.



(b) Gray hair attribute transform.



(c) Blond hair attribute transform.



(d) Black hair attribute transform.

Figure 3: Illustration of various few-shot hair attribute transfer testing results. Each sub-figure is an independent experiment, where the first column is the original image. The following three columns are the attribute learnt during the training and the last column is the target attribute which has been trained within a few-shot.

just with a few samples and a few iterations, the algorithm already converges towards a good local minima. Additionally, the algorithm keeps the ability of transfer the attributes for which is originally trained (e.g. blond, black and gray hair).

By judging the results, we can conduct a qualitative evaluation that suggests a good behaviour of the model. It clearly shows the ability of generating natural-looking faces after applying the image-to-image transformation. Furthermore, it is important to notice that the gray hair attribute incorporates more information than just the color of the hair. We can observe that results which this attribute also look older. The reason for such an event is the entanglement of some attributes. In this dataset, the attribute gray hair is almost always related with old people. Therefore, when the model is pushed to learn the attribute, it cannot decouple the attributes old and gray hair.

4.4 Ablation Study

We present further experiments that support the proposed few-shot algorithm and try to build a general intuition about how sensitive our model is. We conduct a large set of tests for several hair attributes, where we modify the amount of samples of the unseen target class (4,8,16 or 32 samples) and the number of gradient steps (10, 100 or 1000) building in this way, a grid search space. Figure 4 shows four examples and the results after applying the different configuration setup. As expected, the more samples are used for fine-tuning towards the target task, the better the results are. Therefore, experiments where 32 samples are used, offer in general the best performances. Surprisingly, the number of gradient updates shows a counter-intuitive results. The more we optimize towards the class, the worse results. This phenomenon can be explained by a

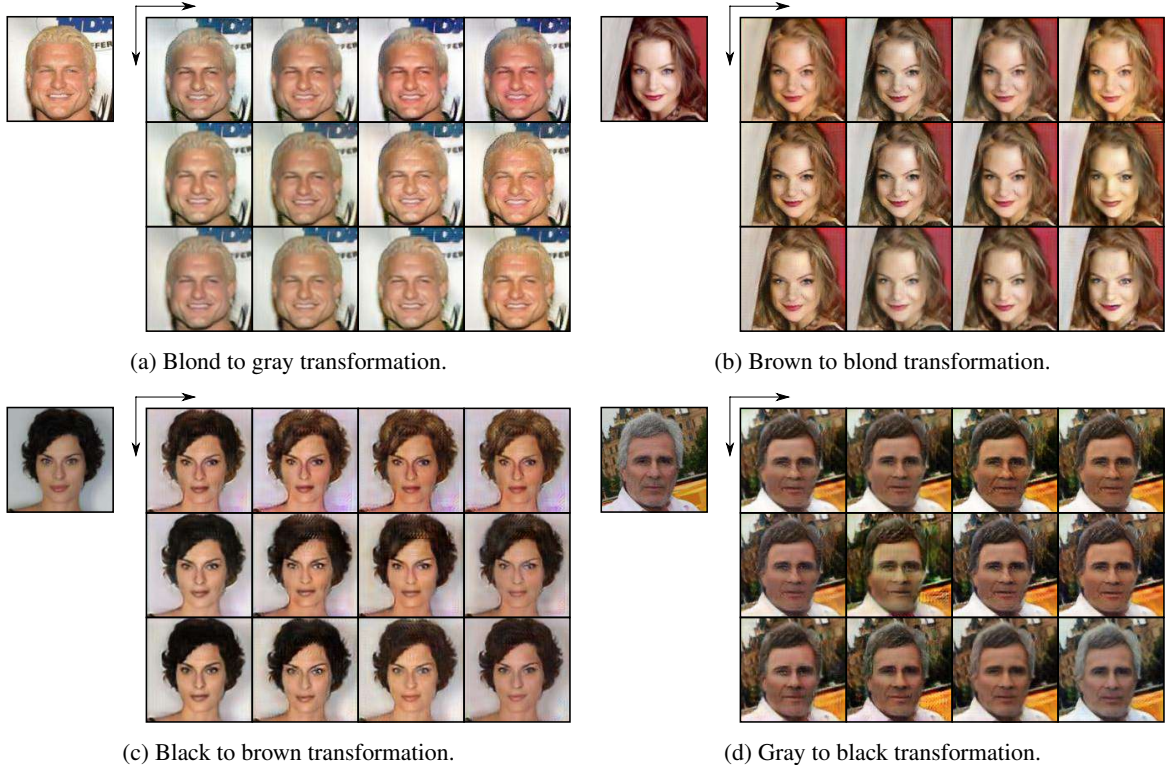


Figure 4: Illustration of various few-shot hair attribute transfer testing results. Each sub-figure is an independent experiment, where we have evaluated the grid search space made of number of samples used for fine-tuning (4,8,16 or 32 column-wise from left to right), and number of gradient updates (10, 100, or 1000 row-size from top to bottom). We can see that best results are always found when we used 32 samples and 10 gradient step updates (upper-right corner).

generalization intuition. Given a loss landscape, after 10 iterations we might find in a wide local minima, however, if we keep on optimizing it is likely that we end up in a much sharper minima where at testing time will yield much worse results. This is what we can observe, the more gradient updates, the less attribute transfer is present on the images.

5 Discussion and Conclusion

We have shown that meta-learning can be used to effectively train generative models for few-shot attribute transfer. Using these techniques on attributes, we can learn to generate images containing unseen attributes with just a few samples. This is done with no lengthy inference time, no external memory and no additional data. Results show that our approach is able to learn and generate attribute given complex image structures like faces. The low amount of data required to generate images, once the model is pre-trained, opens the door to several applications that were previously gated by the high amount of data required. We see many

interesting avenues of future work including combining different types of attribute transform such as hair color with smiling, eyeglasses, and other facial attributes.

References

- [1] A. Antoniou, H. Edwards, and A. Storkey. How to train your maml. *arXiv preprint arXiv:1810.09502*, 2018.
- [2] A. Bansal, S. Ma, D. Ramanan, and Y. Sheikh. Recycle-gan: Unsupervised video retargeting. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 119–135, 2018.
- [3] J. Bao, D. Chen, F. Wen, H. Li, and G. Hua. Cvae-gan: fine-grained image generation through asymmetric training. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2745–2754, 2017.
- [4] S. Bartunov and D. Vetrov. Few-shot generative modelling with generative matching networks. In *International Conference on Artificial Intelligence and Statistics*, pages 670–678, 2018.
- [5] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proceedings of the*

- IEEE Conference on Computer Vision and Pattern Recognition*, pages 8789–8797, 2018.
- [6] L. Chongxuan, T. Xu, J. Zhu, and B. Zhang. Triple generative adversarial nets. In *Advances in neural information processing systems*, pages 4088–4098, 2017.
 - [7] L. Clouâtre and M. Demers. Figr: Few-shot image generation with reptile. *arXiv preprint arXiv:1901.02199*, 2019.
 - [8] A. Creswell, Y. Mohamied, B. Sengupta, and A. A. Bharath. Adversarial information factorization. *arXiv preprint arXiv:1711.05175*, 2017.
 - [9] S. Dodge and L. Karam. A study and comparison of human and deep learning recognition performance under visual distortions. In *2017 26th international conference on computer communication and networks (ICCCN)*, pages 1–7. IEEE, 2017.
 - [10] C. Finn, P. Abbeel, and S. Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 1126–1135. JMLR. org, 2017.
 - [11] C. Finn, K. Xu, and S. Levine. Probabilistic model-agnostic meta-learning. In *Advances in Neural Information Processing Systems*, pages 9516–9527, 2018.
 - [12] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.
 - [13] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville. Improved training of wasserstein gans. In *Advances in Neural Information Processing Systems*, pages 5767–5777, 2017.
 - [14] X. Huang, M.-Y. Liu, S. Belongie, and J. Kautz. Multimodal unsupervised image-to-image translation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 172–189, 2018.
 - [15] S. Iizuka, E. Simo-Serra, and H. Ishikawa. Globally and locally consistent image completion. *ACM Transactions on Graphics (ToG)*, 36(4):107, 2017.
 - [16] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. *arXiv preprint*, 2017.
 - [17] T. Karras, T. Aila, S. Laine, and J. Lehtinen. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*, 2017.
 - [18] T. Kim, M. Cha, H. Kim, J. K. Lee, and J. Kim. Learning to discover cross-domain relations with generative adversarial networks. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 1857–1865. JMLR. org, 2017.
 - [19] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
 - [20] G. Koch, R. Zemel, and R. Salakhutdinov. Siamese neural networks for one-shot image recognition. In *ICML deep learning workshop*, volume 2, 2015.
 - [21] B. Lake, R. Salakhutdinov, J. Gross, and J. Tenenbaum. One shot learning of simple visual concepts. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 33, 2011.
 - [22] Y. LeCun, C. Cortes, and C. Burges. Mnist handwritten digit database. *AT&T Labs [Online]*. Available: <http://yann.lecun.com/exdb/mnist>, 2:18, 2010.
 - [23] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. 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.
 - [24] Y. Li, S. Liu, J. Yang, and M.-H. Yang. Generative face completion. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3911–3919, 2017.
 - [25] J. Liao, Y. Yao, L. Yuan, G. Hua, and S. B. Kang. Visual attribute transfer through deep image analogy. *arXiv preprint arXiv:1705.01088*, 2017.
 - [26] Z. Liu, P. Luo, X. Wang, and X. Tang. Deep learning face attributes in the wild. In *Proceedings of the IEEE international conference on computer vision*, pages 3730–3738, 2015.
 - [27] S. Mo, M. Cho, and J. Shin. Instance-aware image-to-image translation. In *International Conference on Learning Representations*, 2019.
 - [28] T. Munkhdalai and H. Yu. Meta networks. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 2554–2563. JMLR. org, 2017.
 - [29] A. Nichol, J. Achiam, and J. Schulman. On first-order meta-learning algorithms. *arXiv preprint arXiv:1803.02999*, 2018.
 - [30] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros. Context encoders: Feature learning by inpainting. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2536–2544, 2016.
 - [31] S. Ravi and H. Larochelle. Optimization as a model for few-shot learning. 2016.
 - [32] D. J. Rezende, S. Mohamed, I. Danihelka, K. Gregor, and D. Wierstra. One-shot generalization in deep generative models. *arXiv preprint arXiv:1603.05106*, 2016.
 - [33] A. Santoro, S. Bartunov, M. Botvinick, D. Wierstra, and T. Lillicrap. Meta-learning with memory-augmented neural networks. In *International conference on machine learning*, pages 1842–1850, 2016.
 - [34] T. Shen, T. Lei, R. Barzilay, and T. Jaakkola. Style transfer from non-parallel text by cross-alignment. In *Advances in neural information processing systems*, pages 6830–6841, 2017.
 - [35] J. Snell, K. Swersky, and R. Zemel. Prototypical networks for few-shot learning. In *Advances in Neural Information Processing Systems*, pages 4077–4087, 2017.
 - [36] F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. Torr, and T. M. Hospedales. Learning to compare: Relation network for few-shot learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1199–1208, 2018.

- [37] O. Vinyals, C. Blundell, T. Lillicrap, D. Wierstra, et al. Matching networks for one shot learning. In *Advances in neural information processing systems*, pages 3630–3638, 2016.
- [38] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, G. Liu, A. Tao, J. Kautz, and B. Catanzaro. Video-to-video synthesis. *arXiv preprint arXiv:1808.06601*, 2018.
- [39] R. A. Yeh, C. Chen, T. Yian Lim, A. G. Schwing, M. Hasegawa-Johnson, and M. N. Do. Semantic image inpainting with deep generative models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5485–5493, 2017.
- [40] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang. Generative image inpainting with contextual attention. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5505–5514, 2018.
- [41] H. Zhang, T. Xu, H. Li, S. Zhang, X. Wang, X. Huang, and D. N. Metaxas. Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 5907–5915, 2017.
- [42] R. Zhang, P. Isola, and A. A. Efros. Colorful image colorization. In *European conference on computer vision*, pages 649–666. Springer, 2016.
- [43] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv preprint*, 2017.
- [44] J.-Y. Zhu, R. Zhang, D. Pathak, T. Darrell, A. A. Efros, O. Wang, and E. Shechtman. Toward multimodal image-to-image translation. In *Advances in Neural Information Processing Systems*, pages 465–476, 2017.