Statistical Learning and Inference
Final Report

Lei Liu (120033910045, loklz@sjtu.edu.cn),
Menghang Yu (120033910050, ymh18516525237@163.com),
Xianglong Lyu (120033910107, kyle_lyu@sjtu.edu.cn)

Department of Computer Science,
Shanghai Jiao Tong University, Shanghai, China

1 Introduction

This is an era of big data and artificial intelligence. On one hand, the ubiquitous cameras, sensors and other data collectors provide access to extremely large datasets [1], enabling the machines to learn models even more accurate than human beings. On the other hand, the tremendous data collection is valuable only when analyzed by deep and complex AI models. Among all the AI technologies, Deep Neural Network (DNN) is one of the most popular and promising [2, 3, 4]. DNNs have been utilized in a variety of domains ranging from image classification, image generation, natural language processing, automatic translation, protein structure prediction, speech recognition, reinforcement learning among others.

Nevertheless, DNNs comprise a large quantity of parameters. To train them, three computation phases, i.e. forward propagation, backpropagation and kernel update, are repeated millions of times, making it both time and energy consuming. Under certain conditions it is indispensable to train a DNN model with scarce datasets; in those case deep neural networks seem to fall short, over-fitting on the training set and producing poor generalization on the test set.

In a few-shot image generation task, we pour an abundant amount of labelled images, which are tagged as seen categories, into the training phase to form the model. Subsequently, we input test datasets comprising unseen categories, then the learned model is expected to output diverse and lifelike images of unseen categories. Having surveyed a lot in the field of few-shot image generation, we find that there are several effective implementations of few-shot image generation. A domain adaptive few shot generation framework (DAWSON) [5] is a plug-and-play framework that supports a broad family of meta-learning algorithms and various GANs with architectural-variants. Data augmentation generative adversarial networks (DAGAN) [6] is in fact a method of one shot image generation, which injects random noise into the generator to produce a slightly different image from the same category. Fusing-and-filling GAN for few-shot image generation (F2GAN) [7] fuses the high-level features of conditional images and fills in the details of generated image with relevant low-level features of conditional images.

In this report, we train a network which combines two techniques together, i.e. DAGAN [6] and F2GAN [7], and then we employ an evaluation method to acquire the better images which are produced by these two methods. The high-level idea is comparing the generated images by an estimate technique and outputing the best ones, which is portrayed in Figure 1. We tested our network on 2 datasets: Omniglot and VGGFace.

2 Related Work

Transfer Learning and Dataset Shift: the one shot learning problem is a particular case of dataset shift. The term dataset shift [8] generalises the concept of covariate shift [9, 10, 11] to multiple cases of changes between domains. For one shot learning there is an extreme shift in the class distributions - the two distributions share no support: the old classes have zero probability and the new ones move from zero to non-zero probability. However there is an assumption that the class conditional distributions share some commonality and so information can be transferred from the source domain to the one-shot target domain.

Data Augmentation: Data Augmentation [12] is routinely used in classification problems. Often it is non-trivial to encode known invariances in a model. It can be easier to encode those invariances in the data instead by generating additional data items through transformations from existing data items.

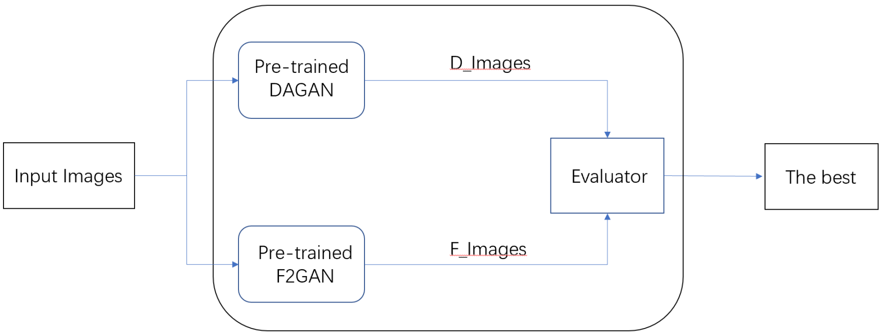


Fig. 1. The high-level idea of our method.

For example the labels of handwritten characters should be invariant to small shifts in location, small rotations or shears, changes in intensity, changes in stroke thickness, changes in size etc. Almost all cases of data augmentation are from a priori known invariance. Prior to this paper we know of few works that attempt to learn data augmentation strategies. One paper that is worthy of note is the work of [13], where the authors learn augmentation strategies on a class by class basis. This approach does not transfer to the one-shot setting where completely new classes are considered.

Generative Adversarial Networks: Generative Adversarial Network (GAN) [14] is a powerful generative model based on adversarial learning. In the early stage, unconditional GANs [15] generated images with random vectors by learning the distribution of training images. Then, GANs conditioned on a single image [6, 16] were proposed to transform the conditional image to a target image. Recently, a few conditional GANs attempted to accomplish more challenging tasks conditioned on more than one image, such as few shot image translation [17] and few-shot image generation [18, 19].

Few-Shot Learning and Meta-Learning: there have been a number of approaches to few-shot learning, from [20] where they use a hierarchical Boltzmann machine, through to modern deep learning architectures for one-shot conditional generation in [21], hierarchical variational autoencoders in [22] and most recently a GAN based one-shot generative model [23]. One early but effective approach to oneshot learning involved the use of Siamese networks [24]. Others have used nonparametric Bayesian approaches [20], and conditional variational autoencoders [21]. With few examples a nearest neighbour classifier or kernel classifier is an obvious choice. Hence meta-learning distance metrics or weighting kernels has clear potential. Skip-residual pairwise networks have also proven particularly effective [23]. Various forms of memory augmented networks have also been used to collate the critical sparse information in an incremental way. None of these approaches consider an augmentation model as the basis for meta-learning.

Few-shot Image Generation: Few-shot generation is a challenging problem which can generate new images with a few conditional images. Early few-shot image generation works are limited to certain application scenario. For example, Bayesian learning and reasoning were applied in [21, 25] to learn simple concepts like pen stroke and combine the concepts hierarchically to generater new images. More recently, FIGR [19] was proposed to combine adversarial learning with optimization-based few-shot learning method Reptile to generate new images. Similar to FIGR [19], DAWSON [5] applied meta-learning MAML algorithms [26] to GAN-based generative models to achieve domain adaptation between seen categories and unseen categories. Metric-based few-shot learning method Matching Network [27] was combined with Variational Auto-Encoder [28] in GMN [18] to generate new images without finetuning in the test phase. MatchingGAN [29] attempted to use learned metric to generate images based on a single or a few conditional images.

3 Our Method

Other than proposing a totally brand-new method to implement few-shot image generation, we decide to put together two existing networks, which have demonstrated their effectiveness in prior work, to build our network. In order to pick the best produced images, we utilize an evaluation to compare the images, and then the model outputs the splendid ones, which are the final results we expect to get.

3.1 Image Generation

In accordance with the prior experimental consequences, we choose two few-shot image generation techniques, DAGAN [6] and F2GAN [7], to generate new images of unseen categories. DAGAN [6]. F2GAN [7].

3.2 Evaluation

Having obtained the images generated by those two models, we add an evaluation phase to estimate the generated images. After the estimation, we output the most similarity images between the two network and generated a new result. What we use to calculate the similarity of the generated image is the Color histogram similarity algorithm. Firstly, collect the histogram data of the input image and the generated image, and then normalize the histogram of each image.

Next, the image similarity value is obtained through the calculation of histogram data with the Bhattacharyya coefficient algorithm(the function is shown below), and the value is between [0,1],which means that 0 for very different, 1 for very similar.

$$\rho(p, p') = \sum_{i=1}^{N} \sqrt{p(i)p'(i)}$$

For each image that we generated, calculate the Bhattacharyya coefficient of the 2 generated image based on the same image, and choose the bigger one as output.

4 Experimental Setting

4.1 Datasets and Baselines

To ensure the correctness and precision of our method, we determine to select Omniglot as a low-level dataset and VGGFace as a high-level dataset. After training the model based on the 2 datasets with DAGAN and F2GAN, we've got the model. Then using the model to generated images based on the model, and it's been shown below.

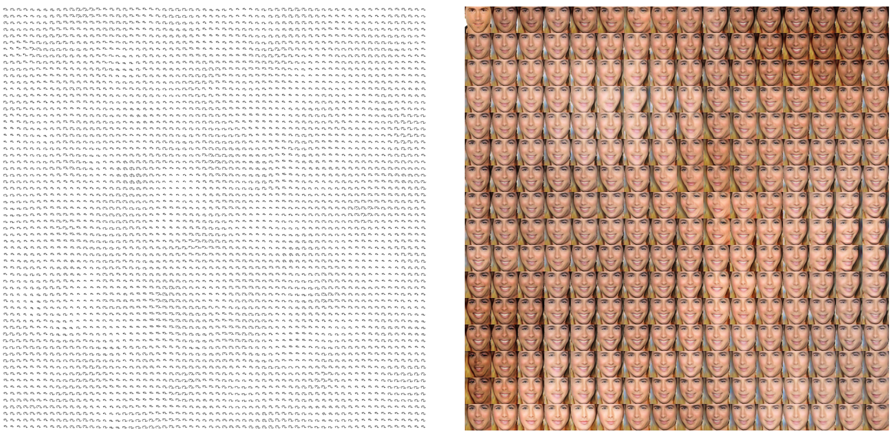


Fig. 2. The generation image of Omniglot and VGGFace.

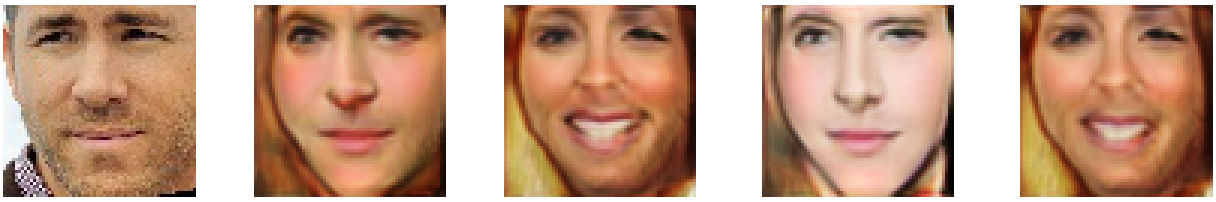
4.2 Evaluation Metrics

According to the outputs of the experiments, the ratio of the 2 mode count for the similarity of the original input.In a way,it means more effective.

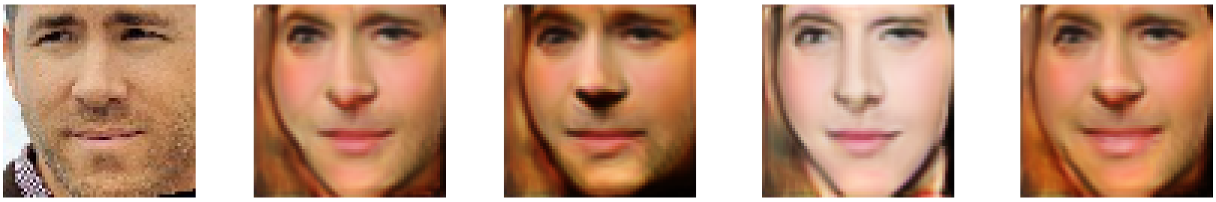
4.3 Experimental Results

After the experiments, we've got a set of outputs

Fig. 3. The generation image of Omniglot after the evalution.



(a) Input (b) Experiment(1) (c) Experiment(2) (d) Experiment(3) (e) Experiment(4)



(f) Input (g) DAGAN(1) (h) DAGAN(2) (i) DAGAN(3) (j) DAGAN(4)

(k) Input (l) F2GAN(1) (m) F2GAN(2) (n) F2GAN(3) (o) F2GAN(4)

Fig. 4. Compare of the Experiment Outputs And the Original Outputs of the 2 Model

According to the comparison of the experiment outputs and the 2 model's outputs,it could get the results more similar to the input and more features.

Table 1. Numbers in Each Model

Dataset	DAGAN	F2GAN
Omniglot	1531	2564
VGGFace	92	163

The table shows the numbers choose to output in one shot, and the times to choose the F2GAN is more often than DAGAN.

5 Conclusion

According to the outputs of the 2 model with the Omniglot and VGGFace datesets, and the ratio of each model outputs in the experiment's outputs.The F2GAN model accounts for a larger proportion. And it means that this model is more similar to the original input compare to the DAGAN model.

References

1. Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Fei-Fei Li: ImageNet: A large-scale hierarchical image database. In *CVPR*, 2009.
2. Henry A. Rowley, Shumeet Baluja, and Takeo Kanade: Neural Network-Based Face Detection. In *TPAMI*, 1998.
3. Zhaohui Zhang, Xinxin Zhou, Xiaobo Zhang, LizhiWang, and Pengwei Wang: A Model Based on Convolutional Neural Network for Online Transaction Fraud Detection. In *Secur. Commun. Networks*, 2018.
4. Carey E. Floyd, Joseph Y. Lo, A. Joon Yun, Daniel C. Sullivan, and Phyllis J. Kornguth: Prediction of breast cancer malignancy using an artificial neural network. In *Cancer*, 1994.
5. Weixin Liang, Zixuan Liu, and Can Liu: . DAWSON: A domain adaptive few shot generation framework. *arXiv:Learning*, 2020.
6. Antreas Antoniou, Amos Storkey, and Harrison Edwards: Data augmentation generative adversarial networks. *arXiv:ML*, 2018.
7. Yan Hong, Li Niu, Jianfu Zhang, Weijie Zhao, Chen Fu, Liqing Zhang: F2GAN: Fusing-and-Filling GAN for Few-shot Image Generation. *arXiv:CV*, 2020.
8. A.J. Storkey.: When training and test sets are different. In Candela Sugiyama Schwaighofer Lawrence (ed.), *Dataset Shift in Machine Learning*, chapter 1, pp. 3–28. MIT Press, 2009.
9. H. Shimodaira: Improving predictive inference under covariate shift by weighting the log-likelihood function. *Journal of Statistical Planning and Inference*, 90:227–244, 2000.
10. A.J. Storkey and M. Sugiyama: Mixture regression for covariate shift. In *Advances in Neural Information Processing Systems 19 (NIPS2006)*, 2007.
11. M. Sugiyama and K. -R. Müller: Input-dependent estimation of generalisation error under covariate shift. *Statistics and Decisions*, 23:249–279, 2005.
12. Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton: Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pp. 1097–1105, 2012.
13. Søren Hauberg, Oren Freifeld, Anders Boesen Lindbo Larsen, John Fisher, and Lars Hansen: Dreaming more data: Class-dependent distributions over diffeomorphisms for learned data augmentation. In *Artificial Intelligence and Statistics*, pp. 342–350, 2016.
14. Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron C. Courville, and Yoshua Bengio: Generative adversarial networks. In *NeurIPS*, 2014.
15. Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida: Spectral normalization for generative adversarial networks. *arXiv preprint arXiv:1802.05957*, 2018.
16. Takeru Miyato and Masanori Koyama: cGANs with projection discriminator. *arXiv preprint arXiv:1802.05637*, 2018.
17. Ming-Yu Liu, Xun Huang, Arun Mallya, Tero Karras, Timo Aila, Jaakko Lehtinen, and Jan Kautz: Few-shot unsupervised image-to-image translation. *arXiv preprint arXiv:1905.01723*, 2019.
18. Sergey Bartunov and Dmitry Vetrov: Few-shot generative modelling with generative matching networks. In *ICAIS*, 2018.
19. Louis Clouâtre and Marc Demers: FIGR: Few-shot image generation with reptile. *arXiv preprint arXiv:1901.02199*, 2019.
20. Ruslan Salakhutdinov, Joshua B Tenenbaum, and Antonio Torralba: One-shot learning with a hierarchical nonparametric bayesian model. In *ICML*, 2012.
21. Danilo Rezende, Ivo Danihelka, Karol Gregor, Daan Wierstra, et al: One-shot generalization in deep generative models. In *ICML*, 2016.
22. Harrison Edwards and Amos Storkey: Towards a neural statistician. In *ICLR*, 2017.
23. Akshay Mehrotra and Ambedkar Dukkipati: Generative adversarial residual pairwise networks for one shot learning. *arXiv preprint arXiv:1704.00028*, 2017.
24. Gregory Koch: Siamese neural networks for one-shot image recognition. *Doctoral dissertation, University of Toronto*, 2015.
25. Brenden M Lake, Ruslan Salakhutdinov, Jason Gross, and Joshua B Tenenbaum: One shot learning of simple visual concepts. *Cognitive Science* 33, 33, 2011.
26. Chelsea Finn, Pieter Abbeel, and Sergey Levine: Model-agnostic meta-learning for fast adaptation of deep networks. In *ICML*, 2017.

27. Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, et al: Matching networks for one shot learning. In *NeurIPS*, 2016.
28. Yunchen Pu, Gan Zhe, Ricardo Henao, Yuan Xin, and Lawrence Carin: Variational autoencoder for deep learning of images, labels and captions. In *NeurIPS*, 2016.
29. Yan Hong, Li Niu, Jianfu Zhang, and Liqing Zhang: MatchingGAN: Matching-based few-shot image generation. In *ICME*, 2020.