Face Aging on Realistic Photo in Cross-Dataset Implementation

Nabilah Hamzah, Fadhlil Hafizheli Kamaru Zaman
Faculty of Electrical Engineering, Universiti Teknologi Mara (UiTM), Shah Alam

nabilahhamzah91@gmail.com

Abstract. Face aging or age progression is a prediction on how a person looks at the future. Face aging image-to-image translation is a process of translating an image of young people to their older version or vice versa. The need for a paired training dataset to train the generative adversarial networks (GANs) is a major problem with face aging image-to-image translation. Nowadays, there is a method where an unpaired training dataset can be used to do an image-to-image translation. CycleGANs is a GANs extended method where there is no need for paired training dataset to train the CycleGANs. From the result, it shows that CycleGANs can do face aging image-to-image translation without using the paired training dataset.

1. Introduction

Image translation is a method of mapping between an input and output image using a training dataset of aligned image pairs [1]. Image translation or image-to-image translation is a very challenging task because the translating image must be realistic, relevant and high in quality. Much of the current literature on image translation pays particular attention to the deep-learning algorithm. Deep-learning is an unsupervised machine learning model that inspired by neurons cell that can learn complex problem and able to solve a multi-layered computational problem [2]. In deep-learning there are different types of architecture or framework i.e. Generative Adversarial Networks (GANs), Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Autoencoders, Resnet, etc. Each type of architecture is used to counter a different type of problem.

In image translation, the most popular approach is by using GANs. GANs is generative modeling where this model automatically discovering and learning the pattern in input data and can generate a new output that plausibly could have been drawn from the original dataset. Preliminary work on GANs was undertaken by Goodfellow [3], where the researcher proposes to train two different models: generative (G) and discriminative (D) model concurrently. The advantages of GANs is that it can train a large dataset, fast to simulate and the trained image produces a visually compelling sample image [4]. However, the main drawbacks of GANs are it unstable to train, the generative (G) model
often produces a nonsensical output, experienced with mode collapse and have no inference capability [3]–[5]. In recent years, there has been an increasing amount of literature on GANs where many of the research venture GANs with other techniques to mitigate GANs disadvantages. GANs have been used in many image processing tasks such as image coloring and sketching [6], medical image synthesis [7], [8] and image manipulation [9] and especially in image translation [1], [3]–[5], [10]–[14].

GANs allow for training on a large dataset but this makes the GANs model very unstable and experience mode collapse. Auto-encoder based GANs (AE-GANs) aims to address this problem by using auto-encoder to boost the model to better represent all the data that training and by this, it also can discouraging the mode collapse problem [4]. In another example, plug and play generative networks (PPGNs) [15] combine auto-encoder, GANs, and classification loss to optimize the network by using a pre-trained classifier. There is more model use a separate encoder to stabilize GANs training such as mode-regularized GANs (MRGANs) [16] and adversarial generator encoder (AGE) [17]. In [18], Radford and Metz propose a method to use a generative (G) and discriminative (D) network model as a feature extractor so that the GANs model will produce more sensical output.

As mention before, image translation is a mapping between an input image and an output image by using a training dataset of aligned image pairs. However, the paired image dataset may not be available for some tasks. To overcome that problem, Jun-Yan [1] suggests a cycle consistent loss where this method is extended GANs but the difference is it using an unpaired training dataset. Figure 1 shows an example of the paired and unpaired training dataset. Based on figure 1 below, unpaired training dataset (right) is a dataset consist of two different domains X and Y. Where these two domains do not have any information as paired training dataset (left) where x_i matches with y_i.

In this research, an unpaired training dataset is used where this research aims to translate the image of young people to an older version of them or vice versa this process is being called as face aging process. CycleGANs is trained to learn the different features between young and older people. From the training session, CycleGANs learn a specific feature and apply it back to the testing image. There are many researches that venture GANs face aging as image-to-image translation [19]–[23]. Face aging approaches can be divided into two main categories: model-based and prototype-based [20]. Model-based approach is using parametric and non-parametric learning where prototype-based approach is using the different age classification method. For this research, a prototype-based approach is using where the dataset is being divided into two groups: young and old of front face image.
In this paper, we investigate the age progression or image translation of face aging by using two different datasets for training and testing. In section 2 it discussed about the research methodology, section 3 discussed the result and discussion and section 4 is the conclusion.

2. Research Methodology
In this section GANs formulation (section 2.1) is reviewed first as CycleGANs is the extended method from GANs. Next, the CycleGANs formulation (section 2.2), network architecture (section 2.3) and training details (section 2.4) is explained.

2.1. Generative Adversarial Networks (GANs)
GANs learning process is to train generative (G) and discriminative (D) model simultaneously. The target of generative (G) model is to generate an input from the distribution of \( P_{Y} \) over data \( X \). The generative (G) model will start with sampling the input variables \( y \) from a Gaussian distribution \( P_{Y}(y) \), then maps the input variables \( y \) to data space \( G(y; \theta_{g}) \) through a differentiable network. The discriminative (D) model will act as a classifier \( D(x; \theta_{d}) \) that aims to identify whether the image is from training data or the image produce by the generative (G) model. The generative (G) and discriminative model minimax objective can be formulated as follows [3]:

\[
\min_{G} \max_{D} V(D,G) = E_{x \sim p_{data(x)}}[\log D(x)] + E_{z \sim p_{y}(y)}[\log(1 - D(G(z)))]
\]  

(1)

2.2. Cycle-Consistent Adversarial Networks (CycleGANs)
CycleGANs main goal is to learn a mapping function between two different domain \( X \) and \( Y \) where the training samples is given as \( \{x_{i}\}_{i=1}^{N} \in X \) and \( \{y_{j}\}_{j=1}^{M} \in Y \). CycleGANs model consist of two mappings function \( G : X \rightarrow Y \) and \( F : Y \rightarrow X \). In cycleGANs model network there will be two generative (G) model: \( G_{X} \) and \( G_{Y} \) and discriminative (D) model: \( D_{X} \) and \( D_{Y} \). \( G_{X} \) goal is to generate an image from input domain \( \{x\} \) and the generated image is \( \{G(x)\} \) and \( G_{Y} \) goal to generate an image from input domain \( \{y\} \). \( D_{X} \) and \( D_{Y} \) model to distinguish between images from domain \( \{x\} \) and the generated images \( \{G(x)\} \). Same goes to \( D_{Y} \) model to discriminate between images from domain \( \{y\} \) and the generated images \( \{G(x)\} \). In cycleGANs there are two losses, first is adversarial loss where this loss is for matching the distribution of the generated images to the data distribution in the target domain. Another loss is cycle consistency loss to prevent the learned mappings of generative (G) and discriminative (D) model from contradicting each other.

2.2.1. Adversarial Loss
Adversarial loss is applied at both mapping functions, \( G : X \rightarrow Y \) and \( F : Y \rightarrow X \) [1], [3] adversarial loss is applied at both mapping functions, \( G : X \rightarrow Y \) and \( F : Y \rightarrow X \)

\[
L_{GAN}(G,F,X,Y) = E_{y \sim p_{data(y)}}[\log D_{Y}(y)] + E_{x \sim p_{data(x)}}[\log(1 - D_{Y}(G(x)))]
\]  

(2)

Where \( G \) is the generative model that tries to generate images \( G(x) \) that look similar to the images at the domain \( Y \), while \( D_{Y} \) tries to distinguish between the generated samples produce by \( G \) and the real sample from domain \( Y \).

2.2.2. Cycle Consistency Loss
Cycle consistency loss aim to bring back the generated image into it original image. For each generated image \( \{G(x)\} \) and \( \{F(y)\} \).

\[
L_{cycle}(G,F) = E_{x \sim p_{data(x)}}[\|G(F(x)) - x\|_1] + E_{y \sim p_{data(y)}}[\|G(F(y)) - y\|_1]
\]  

(3)
2.3. Network Architecture

For the network architecture for this research is adapt from Johnson et al. [24] who have shown an impressive results for neural style transfer and super resolution. This network contains two stride-2 convolutions network and several blocks of residual network at the generative (G) model [25] and two \( \frac{1}{2} \) - stride convolutions. An instance normalization is apply to the generative (G) and the discriminative (D) model [26]. For the discriminative (D) networks, a 70 x 70 patch-GANs to classify overlapping image patches are real or fake [11], [25], [26].

2.4. Training Details

In this section it discussed about the training procedure and dataset where it focusses on how training is being held and how the dataset is prepared. In section 2.4.1 it discussed about the training procedure and on section 2.4.2 it discussed about the training dataset.

2.4.1 Training Procedure

The training session is using a python language with TensorFlow 2.0 as a deep learning platform. Even though there are many others deep learning platform but TensorFlow is much easier to implement and TensorFlow have their own community that can help with the research. Plus, a GPU GeForce RTX 2070 Super is used in the training session.

2.4.2 Training Dataset

Dataset used in this research is from UTKFace and LFW dataset, where UTKFace dataset consists of images from long age span range from 0 to 116 years old. The UTKFace images is used as training image and the LFW is used as the testing image. The UTKFace dataset consists of 20,000 face images, in the wild. The training dataset is divided into two domains \( X \) and \( Y \) where domain \( X \) consists of images from age range of 18 to 25 years old male and domain \( Y \) consists of images from age range of 58 to 70 years old male. The range for the youngest subject is chose between 18 to 25 years old because the facial feature for that range of age is most likely the same.

3. Result and Discussion

In this section it discussed the result obtain from the training and testing sessions. Figure 2 shows the result for image-to-image translation for face aging. From the figure it can be observed that the algorithm is able to translate and image of young people to older version. From the images it can be observed that there is no alignment needed in all the images. So, the algorithm can successfully translate all the images.
From figure 3 below, it can be observed that the image is being align as there are a black patch on the images. So, from the images, it can be observed that all images that being align cannot be translate successfully. Another observation is image with low resolution also cannot be translate by the algorithm.

| Input | Output |
|-------|--------|

**Figure 2: Testing Images**

From figure 3 below, it can be observed that the image is being align as there are a black patch on the images. So, from the images, it can be observed that all images that being align cannot be translate successfully. Another observation is image with low resolution also cannot be translate by the algorithm.
Figure 3: Example of Images not Successfully Translated

4. Conclusion
From the result it shows that CycleGANs is able to translate an image of young people to their older version by learning the feature from image of older people face. But, for image that tilt and not centre and need an alignment reduces the CycleGANs performance where the algorithm not successfully to translate the images. In the future work, need a good alignment method to increase the CycleGANs performance.

Acknowledgments
We would like to extend our acknowledgment to Universiti Teknologi Mara (UiTM) and to those who have directly or indirectly contribute to our research. This research is funded by the Trans-Disciplinary Research Grant Scheme (600-IRMI/TRGS 5/3 (001/2019)-1), Ministry of Higher Education Malaysia.

References
[1] A. E. A. Zhu Jun-Yan, Taesung Park, Isola Phillip, “Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks Jun-Yan,” Proc. IEEE Int. Conf. Comput. Vis., pp. 183–202, 2010.
[2] Z. Kadim, M. A. Zulkifley, and N. Hamzah, “Deep-learning based single object tracker for night surveillance,” Int. J. Electr. Comput. Eng., vol. 10, no. 4, pp. 3576–3587, 2020.
[3] I. J. Goodfellow et al., “Generative Adversarial Networks,” pp. 1–9, 2014.
[4] M. Rosca, B. Lakshminarayanan, D. Warde-Farley, and S. Mohamed, “Variational Approaches for Auto-Encoding Generative Adversarial Networks,” 2017.
[5] A. Heljakka, A. Solin, and J. Kannala, “Pioneer Networks: Progressively Growing Generative Autoencoder,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 11361 LNCS, pp. 22–38, 2019.
[6] Sangkloy, J. Lu, C. Fang, F. Yu, and J. Hays, “Scribbler: Controlling deep image synthesis with sketch and color,” Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017, vol.
[7] C. Han, L. Rundo, R. Araki, and Y. Furukawa, “Infinite Brain Tumor Images: Can GAN-based Data Augmentation Improve Tumor Detection on MR Images?,” pp. 1–4.

[8] Y. Z. Jun, K. Philipp, S. Eli, and A. E. Alexei, “Generative Visual Manipulation on the Natural Image Manifold,” ECCV 2016, vol. 9909, pp. 597–613, 2016.

[9] C. Han, H. Hayashi, L. Rundo, R. Araki, and W. Shimoda, “GAN-Based Synthetic Brain MR Image Generation,” 2018 IEEE 15th Int. Symp. Biomed. Imaging (ISBI 2018), no. Isbi, pp. 734–738, 2018.

[10] A. Nguyen, J. Clune, and J. Yosinski, “Plug & Play Generative Networks: Conditional Iterative Generation of Images in Latent Space,” no. 1, pp. 4467–4477.

[11] L. W. Che Tong, Li Yanran, Jacob Athul Paul, Bengio Yoshua, “Mode Regularized Generative Adversarial Networks,” ICLR 2017, pp. 1–13, 2017.

[12] D. Ulyanov, A. Vedaldi, and V. Lempitsky, “It Takes (Only) Two: Adversarial Generator-Encoder Networks,” Thirty-Second AAAI Conf. Artif. Intell., pp. 1250–1257, 2016.

[13] Z. Yi, H. Zhang, P. Tan, and M. Gong, “DualGAN: Unsupervised Dual Learning for Image-to-Image Translation,” Proc. IEEE Int. Conf. Comput. Vis., vol. 2017-Octob., pp. 2868–2876, 2017.

[14] K. He, J. Sun, Z. Xiangyu, and S. Ren, “Deep Residual Learning for Image Recognition,” IEEE Conf. Comput. Vis. Pattern Recognit., pp. 770–778, 2016.

[15] Z. Murez, S. Kolouri, D. Kriegman, R. Ramamoorthi, and K. Kim, “Image to Image Translation for Domain Adaptation,” Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 4500–4509, 2018.

[16] Y. Liu and Q. Li, “Attribute-aware Face Aging with Wavelet-based Generative Adversarial Networks,” IEEE, pp. 11877–11886, 2019.

[17] Y. Liu and Q. Li, “Attribute-aware Face Aging with Wavelet-based Generative Adversarial Networks,” IEEE, pp. 11877–11886, 2019.

[18] Johnson, A. Alahi, and L. Fei-Fei, “Perceptual Losses for Real-Time Style Transfer and Super-Resolution,” Eur. Conf. Comput. Vis. (ECCV 2016), vol. 9906, no. Lecture Notes in Computer Science, pp. 306–310, 2018.

[19] J. Song, J. Zhang, L. Gao, X. Liu, and H. T. Shen, “Dual Conditional GANs for Face Aging and Rejuvenation,” pp. 899–905, 2017.

[20] S. Liu and W. Wang, “Face Aging with Contextual Generative Adversarial Nets,” 2017.

[21] J. Song, J. Zhang, L. Gao, X. Liu, and H. T. Shen, “Dual Conditional GANs for Face Aging and Rejuvenation,” pp. 899–905, 2017.

[22] S. Liu and W. Wang, “Face Aging with Contextual Generative Adversarial Nets,” 2017.

[23] Y. Liu and Q. Li, “Attribute-aware Face Aging with Wavelet-based Generative Adversarial Networks,” IEEE, pp. 11877–11886, 2019.

[24] Johnson, A. Alahi, and L. Fei-Fei, “Perceptual Losses for Real-Time Style Transfer and Super-Resolution,” Eur. Conf. Comput. Vis. (ECCV 2016), vol. 9906, no. Lecture Notes in Computer Science, pp. 306–310, 2018.