PAPER • OPEN ACCESS

Facial Expression Transfer using Generative Adversarial Network : A Review

View the article online for updates and enhancements.

Facial Expression Transfer using Generative Adversarial Network : A Review

Noor Adibah Najihah Mat Noor[2], Norhaida Mohd Suaib[1]

[1]School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia, Skudai, Johor, MALAYSIA

[2]UTM Big Data Centre/School of Computing, Faculty of Engineering , Universiti Teknologi Malaysia, Skudai, Johor, MALAYSIA

E-mail: nooradibahnajihah@gmail.com

Abstract. There is high demand of realistic facial expression in current computer graphics and multimedia research. Realistic and accurate facial expression can guarantee the animated character to deliver the expression correctly. However, generating facial expression requires hard work, effort and time since high realism of facial expression need to be in details. There are some available methods in current research area such as face warping to the target, re-use the existing images and also models for generating facial image with certain attribute. Based on literature reviews, current trend for facial expression is using the deep learning method such as generative model like Generative Adversarial Network (GANs). Some of GANs that recently available are Conditional Generative Adversarial Network (cGANs), Double Encoder Conditional GAN (DECGAN), Conditional Difference Adversarial AutoEncoder (CDAAE), Geometry-Guided Generative Adversarial Network (G2GAN), and Geometry-Contrastive Generative Adversarial Network (GC-GAN). These methods actually helped in creating more realistic images, reaching out the realistic facial expression and good identity preservation. This paper aims to review available GANs, find out related features to these methods and also performance of these methods that are useful in facial expression transfer process

1. Introduction

High demand ing realistic facial expression for animated character boost up expression animation as hot topic in present computer graphics research area. Current approach for high realism facial expression is complicated as it counted on accurate facial expression in animated character to ensure animated character deliver the right expression well. However, generating facial expression from scratch is hard since create facial expression involve different muscles on facial and need many efforts. Variety of methods is applicable in generating facial expression of animated character. Transferring facial expression is one of the options in saving the painstaking work other than reduce hard work and time in ensuring facial expression of animated character well-generated. Facial expression is crucial in processing animated character to secure the high realism of animation. Since then, initiatives are suggested to make the facial expression easy, less time consumption and low cost of computational. Transferring facial expression remarks simpler way to generate facial expression on animated character. Following are methods that applied in facial expression transfer. First is input face warping directly to the target [1][2] and second is re-using existing images [3] while other researcher synthesize images with predefined attribute using generative model [4]. Recent years, new trend in

handling facial expression is using Generative Adversarial Network (GANs) which eventually works well in generating high quality sample of face images. This method shows impressive result and options when using GANs. It is either conditional or additional approach for better result and performance. Figure 1 shows the types of GANs that are recently available.
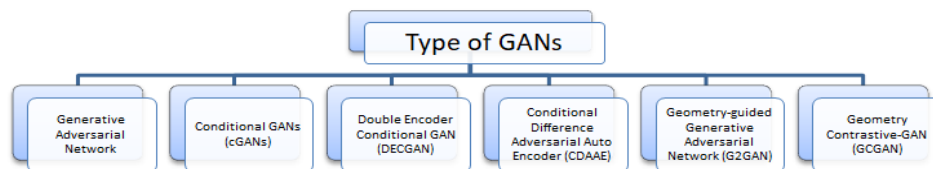

Figure 1. Types of GANs

This paper is organized as follows; Section 2 will describe on Generative Adversarial Network (GAN) and type of available GANs. Section 3 will be discussing on GANs function and summarize the related information of GAN usage and function. Section 4 will conclude the review on Generative Adversarial Networks.

2. Generative Adversarial Network (GAN)

Generative Adversarial Network Model (GAN) was developed recently as in natural environment, it managed to generate high quality sample for face images [5]. Face image synthesis using GANs has been applied well by [6][4][7]. The good part of GANs is the optimization is directly and producing high realism data. Other than that, conditional extended GANs can be applied to manage characteristic of image produced like new images by manipulate potential vectors and decode face image of low dimensional representation in latent space. The following are applicable GANs for facial expression.

2.1. Conditional Generative Adversarial Network(cGAN)

Once the parameter of potential vector is controlled, the information in image is in well-organized and called as conditional Generative Adversarial Network (cGAN). It is justified as generating images that are controlled by labels and attributes [8]. This cGANs also being applied face manipulation [9], face aging [10] and many other facial expression transfer by considering the output result [11][12].

2.2. Double Encoder Conditional GAN (DECGAN)

GANs can directly optimized to generate most realistic and reasonable data. There are conditional GANs to control features of generated images. Encoders of GANs used for face images representation and produce new images by decoding using potential vector manipulation. DECGAN is proposed by [13] to help produce certain target expression. Two encoder share features by adopting associative learning. Associative learning associates unpaired data with feature similarity. Data can correlate from target domain to original domain by reconstruct on consistency set of intrinsic properties even different domain and image data. [14] geared up on association approach. One target domain extract and another for vector input mapping. Generator takes output and another reconstructs generated image. Desired facial expressions on face images produced yet maintain the identity information [13].

2.3. Conditional Difference Adversarial Autoencoder (CDAAE)

Other types of GANs proposed by [12] that produced conditioned faces on emotion states were CDAAE. Many proposed approach on expression human emotion in continuous way before, but failed to satisfy detailed characteristic facial expression description and [15][16] include continuous information like geometry into cGANs. CDAAE framework included source expression, target

expression and process in handing expression. The input passes through encoder for raw pixel mapping in latent space distribution. Then, target face is generated through decode by linking chain of labels latent and target talent. CDAAE proposed long-range feedforward connection from encode to decode which reused part low-level facial attribute and learn high-level facial expression changes. This allows identity information preservation and helps network to rest and focus only on modeling the facial expression changes

2.4. Geometry- Guided Generative Adversarial Network (G2GAN)
There are impressive result based on other generative model [17][18][19] but some lack in details and low in resolution. Encoded latent feature space with certain aligned semantic properties lead flexible image but less control on synthesized images. [20] proposed Geometry-Guided Generative Adversarial Network (G2GAN) as facial geometry information to guide the facial expression transfer. Image of facial is photorealistic with continuously-adjusting expression as variety of facial geometry input used as controller for target expression. The input directly linked to the facial landmarks. Facial expression removal and synthesis performed concurrently by dual generator. However, neutral face used for facial expression transfer produced from removal network in this method may affect the performance mainly if the subject emotion in variation [21].

2.5. Geometry-Contrastive Adversarial Network (GC-GAN)
GC-GAN is being introduced by [21] in order to transfer emotion on different shape and subjects while preserving the identity. Contrastive learning is used to manage the different alignment of subject and emotion as it changes manifold of geometry into facial expression manifold of embedded semantic. For effectively emotion control, the latent space of GANs is injected by embedded geometry. Hence, the consistency of facial image expression remains [21].

3. Discussions
Based on review, deep learning approach has become one of the chosen methods to create realistic facial expression. Deep generative models in deep leaning are low-impact since they had difficulties in computational and leveraging the generative context benefit. Hence, new generative model estimation proposed to overcome the difficulties [5]. This new model eventually leads towards adversarial net framework, which is a new framework on generative model estimation using adversarial process. The model that can represent the probability distribution over kinds of data and able to directly optimized to ensure most reasonable and realistic data produced. There are a variety of types that can be found in recent years of research are in adversarial network area for expression image synthesis. Generating facial expression with discrete and limited emotion states mainly focusing on previous work related to GAN-based facial expression transfer. However, expression of human emotions are available in continuous was, thus discrete states are not adequate for detailed characteristic of facial expression description. Hence, many researcher proposing variation of methods in generating facial expression transfer using variety of condition using GAN order to incorporate the continuous information and improving the result of facial expression transfer. GAN-based is actually an active recent study since it gains more attention on creating realistic faces. For image generation, this method is one of the state-of-the-art approaches. [5]. Based on previous work, the limitation of original GAN is that this kind approach is actually less controllable for larger pictures and higher number of pixels. GAN framework also lacking in latent representation of input image finding where this mechanism is important in reconstruct and real images modification [6]. Improvement is required to allow GAN control image and solve problem related. To generate controlled images by labels or attribute, conditional GAN (cGAN) was introduced by [8] and applied by [4][11][12]. Determine specific relationship between images is cGAN ability as it contains external information. Modification cGAN introduced conditional version as image synthesis process guidance [8]. DECGAN is proposed [13] inspired by structure and function of GAN and cGAN where framework includes modification structure and condition using

target facial expression feature. DECGAN can synthesize new face expression and solve problem of specified target expression lacking and unpaired data training. Similar theme of face images generated with preferred facial expression while maintain identity information. CDAAE is another option on GAN-based image generation. Face with condition on emotion states or AU labels by CDAAE that combine and interpolate facial expression or action units within training set can generate novel expression even or unseen subject faithfully and preserved identity information. It is worthy especially for small database training. High-level semantic presence of large and non-linear face geometry variation are among of the challenges in GAN. G2GAN is being proposed as continuously-adjusting and identity-preserving facial expression synthesis using facial geometry [22] as controllable condition to guide. The facial geometry helps photo-realistic synthesis and target expression specification. G2GAN combined facial editing subnetworks and form cycle mapping for training without paired data and individual-specific shape model for facial geometry due to individuality consideration. Continuous emotion transfer across variety subject made GC-GAN proposed as it made up of facial geometry embedding network, image generator network and image discriminator network. GC-GAN introduces geometry information into cGAN as continuous condition to guide identity-preserving face with expression characteristic. Basically, generative adversarial network in facial image synthesis shows impressive result in high realism images of face. Some methods manage to reduce processing time to build expression on target facial images as certain methods speed up process by adding condition to framework to ensure less process work during facial expression transfer. The identity information preserved even after source/ input is move across process through whole framework of these methods to reach target expression. Table 1 shows GANs type and the information related such as database applied, function, advantage, and purposes applicable for each GANs for future research reference.

Table 1. Types of GAN

Type of GAN	Database Applied	Function	Advantage	Purpose
Generative Adversarial Network (GAN)	MNIST, Toronto Face Database (TFD), CIFAR	Generate a quality of high quality sample for face image using data distribution probability	State-of-the-art approach of image generation	Face synthesis, image synthesis
Conditional GANs (cGANs)	MNIST, MIR Flickr	Generating potential images using parameter, labels and attribute	Control generated image features with conditional extension	Face synthesis, image synthesis, facial expression transfer
Double Encoder Conditional GAN (DECGAN)	CK+, Real-world Affective Faces Database (RAF-DB)	Generate potential expression by decoding vector potential manipulation	Learn faces diversity and manage to create face image of same theme with desired facial expression, preserve identity information	Facial expression synthesis
Conditional Difference Adversarial Auto Encoder (CDAAE)	Denver Intensity Spontaneous Facial Action (DISFA), Radbound Faces Database (RaFD)	Generate data directly upon the condition information	Easy to manipulate generated facial expression and image, preserved identity information when generate facial expression	Facial expression synthesis
Geometry-guided Generative Adversarial Network (G2GAN)	CK+, MultiPIE	Generate facial with certain expression by employing the facial geometry	Continuously adjusting and identity-preserving facial expression synthesis	Facial expression synthesis, face transfer, expression interpolation, expression-invariant face recognition
Geometry Contrastive-GAN (GCGAN)	MultiPIE, CK+, BU-4DFE, CelebA	Generate facial expression with facial landmark which act as guiding landmark as additional feature map	Transfer continuous emotion across different subject, identity preservation	Facial expression synthesis, facial expression transfer

4. Conclusions

Many improvements had done by reviewing methods applied for facial expression transfer. The performance upgraded, consistency exists in transferring facial information from source to target and the increasing realism level on facial expression model. However, as much as attempt and experimental result demonstrated by researcher, there are still gaps and space for to be developed.

References

[1]   Yang, F., Bourdev, L., Shechtman, E., Wang, J. and Metaxas, D., 2012. Facial expression editing in video using a temporally-smooth factorization. *IEEE Conference on Computer Vision and Pattern Recognition.* pp 861-8.

[2]   Yeh, R., Liu, Z., Goldman, D. B., and Agarwala, A. 2016. Semantic facial expression editing using autoencoded flow. *arXiv preprint arXiv:1611.09961.*

[3]   Umar Mohammed, Simon J.D.Prince, and Jan Kautz.2009. Visio-lization: generating novel facial images. ACM SIGGRAPH.

[4]   Ding, H., Sricharan, K., and Chellappa, R. 2018. Exprgan: Facial expression editing with controllable expression intensity. *Thirty-Second AAAI Conference on Artificial Intelligence.*

[5]   Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., and Bengio, Y. 2014. Generative adversarial nets. *Advances in neural information processing systems.* pp 2672-2680.

[6]   Perarnau, G., van de Weijer, J., Raducanu, B., and Álvarez, J. M. 2016. Invertible conditional gans for image editing. *arXiv preprint arXiv:1611.06355.*

[7]   Zhu, J. Y., Park, T., Isola, P., and Efros, A. A. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv preprint.*

[8]   Mirza, M., and Osindero, S. 2014. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784.*

[9]   Yin, X., Yu, X., Sohn, K., Liu, X., and Chandraker, M. 2017. Towards large-pose face frontalization in the wild. *Proceedings of the IEEE International Conference on Computer Vision.* pp 3990-9.

[10]   Zhang, Z., Song, Y., and Qi, H. 2017. Age progression/regression by conditional adversarial autoencoder. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* pp 5810-18.

[11]   Choi, Y., Choi, M., Kim, M., Ha, J. W., Kim, S., and Choo, J. 2018. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* pp 8789-97.

[12]   Zhou, Y., and Shi, B. E. 2017. Photorealistic facial expression synthesis by the conditional difference adversarial autoencoder. *Affective Computing and Intelligent Interaction (ACII), 2017 Seventh International Conference on .* pp 370-6.

[13]   Chen, M., Li, C., Li, K., Zhang, H., and He, X. 2018. Double Encoder Conditional GAN for Facial Expression Synthesis. *2018 37th Chinese Control Conference (CCC.)* pp 9286-91.

[14]   Haeusser, P., Frerix, T., Mordvintsev, A., and Cremers, D. 2017. Associative domain adaptation. *Proceedings of the IEEE International Conference on Computer Vision.* pp 2765-73.

[15]   Ma, L., Jia, X., Sun, Q., Schiele, B., Tuytelaars, T., and Van Gool, L. 2017. Pose guided person image generation. *Advances in Neural Information Processing Systems.* pp 406-16.

[16]   Balakrishnan, G., Zhao, A., Dalca, A. V., Durand, F., and Guttag, J. 2018. Synthesizing Images of Humans in Unseen Poses. *arXiv preprint arXiv:1804.07739.*

[17]   Huang, R., Zhang, S., Li, T., and He, R. 2017. Beyond face rotation: Global and local perception gan for photorealistic and identity preserving frontal view synthesis. *Proceedings of the IEEE International Conference on Computer Vision.* pp 2439-48.

[18]   Isola, P., Zhu, J. Y., Zhou, T., and Efros, A. A. 2017. Image-to-image translation with conditional adversarial networks. *Proceedings of the IEEE conference on computer vision and pattern recognition.* pp 1125-34.

[19]   Zhu, J. Y., Park, T., Isola, P., and Efros, A. A. 2017. Unpaired image-to-image translation using cycle-

consistent adversarial networks. *Proceedings of the IEEE international conference on computer vision.* pp 2223-32.

[20]　Song, L., Lu, Z., He, R., Sun, Z., and Tan, T. 2018. Geometry guided adversarial facial expression synthesis. *2018 ACM Multimedia Conference on Multimedia Conference.* pp. 627-35.

[21]　Qiao, F., Yao, N., Jiao, Z., Li, Z., Chen, H., and Wang, H. 2018. Geometry-Contrastive Generative Adversarial Network for Facial Expression Synthesis. *arXiv preprint arXiv:1802.01822.*

[22]　Matthews, I., and Baker, S. 2004. Active appearance models revisited. *International journal of computer vision*, *60*(2), pp 135-64.