Recent Progress of Face Image Synthesis

Zhihe Lu[2,3,4], Zhihang Li[1,2,4], Jie Cao[1,2,4], Ran He[1,2,3,4], and Zhenan Sun[1,2,4]

[1]National Laboratory of Pattern Recognition, CASIA
[2]Center for Research on Intelligent Perception and Computing, CASIA
[3]Center for Excellence in Brain Science and Intelligence Technology, CAS
[4]University of Chinese Academy of Sciences, Beijing, 100049, China
{luzhihe2016, lizhihang2016, caojie2016}@ia.ac.cn, {rhe, znsun}@nlpr.ia.ac.cn

Abstract

Face synthesis has been a fascinating yet challenging problem in computer vision and machine learning. Its main research effort is to design algorithms to generate photo-realistic face images via given semantic domain. It has been a crucial prepossessing step of main-stream face recognition approaches and an excellent test of AI ability to use complicated probability distributions. In this paper, we provide a comprehensive review of typical face synthesis works that involve traditional methods as well as advanced deep learning approaches. Particularly, Generative Adversarial Net (GAN) is highlighted to generate photo-realistic and identity preserving results. Furthermore, the public available databases and evaluation metrics are introduced in details. We end the review with discussing unsolved difficulties and promising directions for future research.

1. Introduction

As one of the most successful applications of image analysis and understanding, face synthesis has recently received significant attention, especially during the last decade. The major purpose of face synthesis is to synthesize desired images, *e.g*. photo-realistic, some artistic style, or clearer faces via given input in some semantic domains. In this paper, traditional methods as well as advanced deep learning approaches for face synthesis will be extensively reviewed and discussed.

Non-neural-network approaches are summarized as traditional methods which are divided into three broad categories, *i.e*. subspace representation, geometry modeling, and statistical models. Early researchers mainly focused on the cure of the curse of dimensionality, *e.g*. Huang *et al*. [22] utilized manifold estimation, and [41, 64] proposed tensor-based subspace learning method. Inspired by massive successful works in computer vision from geometry knowledge, Halder *et al*. [17] provided a face synthesis system based on geometry modeling and Patel and Smith [37] used the information of the obscure part as auxiliary condition to obtain 3D morphable model. In order to build or expand the database for face recognition, a method by Ramalingam and Viet [43] focused on using inexpensive 3D sensors to synthesize face. Simultaneously, a bunch of works have been proposed with statistical learning theory.

Deep learning has achieved great breakthroughs in image classification [27], image generation [69, 45], biometrics [51, 59, 39, 73], *etc*. Particularly, convolution neural network does remarkably well in tasks related to image [51, 59, 39, 45]. A tremendous amount of works based on deep learning have been developed in face synthesis. Peng *et al*. [39] concentrated on a synthesis CNN to generate non-frontal view from a single frontal face and Richardson *et al*. [45] extracted the face geometry from its image directly by a CNN based approach. As an extension of traditional AAM models, Duong *et al*. [12] proposed a Deep Appearance Models (DAMs) approach, which uses Deep Boltzmann Machines (DBM) to robustly capture the variations of facial shapes and appearances.

Goodfellow *et al*. [15] proposed the conception of Generative Adversarial network (GAN) in 2014, and later several successful revamped versions [2, 66] improved the stability of training process. Researchers fused the information and prior knowledge in face synthesis field and produced fake images that almost have no distinction with real ones if observed by human. Meanwhile, a lot of fascinating adapted GAN models have been brought up to deal with plenty of problems in face synthesis. Till now, many state-of-the-art results produced by deep models are based on the GAN framework.

2. Traditional Methods

Many traditional methods of face synthesis have been proposed, which including subspace representation, geometry modeling and statistical models. In this part, relevant researches as well as the differences of various methods will be discussed.

2.1. Subspace Representation

Face image lies in an embedded non-linear manifold within the high-dimensional space [22], which brings great challenges to face synthesis. Subspace learning based methods tackle this problem by mapping high-dimensional data to a low-dimensional space while retaining as much information as possible [22].

While modeling a face directly with a linear subspace model is difficult, Nguyen *et al*. [35] presented an approach that can extract different layers automatically to support the process of face synthesis. Later, Huang *et al*. [22] proposed a method to synthesize 3D face images from a single image with manifold estimation. One sub-problem, sketch synthesis, is to convert a photo into a sketch. For instance, Chang *et al*. [7] divided the photos and sketches into patches and computed the sparse representation. In order to decrease the computational complexity, Wang *et al*. [53] employed a novel method of face sketch synthesis based on random sampling and locality constraint.

With generating multi-pose images or 3D face images from a single image being a significant research direction, Qiao *et al*. [41] presented a tensor-based subspace learning method (TSL). Moreover, Gender and Ethnicity Specific GEMs (GE-GEMs) is considered in [19]. Heo and Savvides [18] further proposed a 3D Generic Elastic Model (3D-GEM) to model faces utilizing Combination of ASMs and AAMs (CASAAMs) to represent a sparse 2-D shape. Confronted with the problem of data sparsity, Xu and Zha [56] deployed transfer learning approach to generate auxiliary data from original sparse data. It was reported in [67] that a triangulation-based partition criterion ensures the strict alignment of corresponding triangular patches and sparse representation is utilized to produce high quality frontal face images. Xu and Savvides [24] proposed the Fukunaga Koontz Transform (FKT) approach to model dual subspace for synthesizing facial ethnic appearances.

Based on the Lambertian reflection model and linear observation model, Zhang *et al*. [64] presented a tensor analysis method to synthesize an artificial high-resolution (HR) visual light (VIS) face image from a low-resolution (LR) near-infrared (NIR) input image. Besides, Dou *et al*. [11] formulated the monocular face shape reconstruction problem as a Two-Fold Coupled Structure Learning (2FCSL) process.

2.2. Geometry Modeling

Face image contains rich and complicated structural information resulting from the variations in expressions, poses, textures, illumination, *etc*., which is the major stumbling block of face synthesis. Considering facial key points, texture, shape and other graphics characteristics from geometric perspective, geometry based methods can model face directly.

To parameterize the facial images with multi-modeling is proposed by [74] and they aim to synthesize accurate facial expression images. A Face Synthesis system (FASY) is introduced in [17]. Similarly, Bhattacharjee *et al*. [4] proposed face construction method to generate expected face image with textual description via stored facial components from distinct databases. Furthermore, Patel and Smith [37] utilized shape-from-shading as an auxiliary approach to gain a better fitting 3D morphable model. To create databases for face recognition, Ramalingam and Viet [43] proposed a method using inexpensive 3D sensors to synthesize face image. Occlusion problem is studied in [68]. In addition, Ferrari *et al*. [13] extended a 3D based frontalization for synthesizing unconstrained face images. To reconstruct individual 3D shapes from multiple single images of one person, Piotraschke and Blanz [40] proposed a quality measure that judges a reconstruction without information about the true shape. A novel data augmentation method is presented in [3] that can solve the unbalanced dataset problem. It is reported in [23] that multiple synthesized facial images can improve the results of recognizing the unfamiliar faces compared with only a single image.

2.3. Statistical models

Statistical learning based models have widely applied in computer vision, machine learning and biometrics, which can discover statistical properties from big data.

Wang *et al*. [55] proposed a 3D spherical harmonic basis morphable model (SHBMM) that combines spherical harmonics with the morphable model framework. To synthesize face sketch, Wang and Tang [54] used multi-scale Markov Random Fields to address the shortcomings of the linear processes. The embedded hidden Markov model(E-HMM) [70], Bayesian inference [63], multi-representation approach [38] are proposed to improve effectiveness. An architecture that consists of probabilistic face diffuse model and generic face specular map is presented by [49]. Sagonas *et al*. [47] employed a novel method for joint frontal view reconstruction. Liu *et al*. [30] investigated thoroughly cascaded regression based 3D face reconstruction approaches. Then Liu *et al*. [31] further proposed an approach to simultaneously solve face alignment and 3D face reconstruction by iteratively and alternately applying two sets of cascaded regressors. Introducing guided image filters to capture detailed features is proposed in [9]. Moreover, Ren *et al*. [44]

achieved face image synthesis by upscaling and refining the estimated image several times. A two-step method for cross-modality face synthesis is proposed in [50]. Recently, Bordes *et al*. [5] presented a method for training a generative model via an iterative denoising procedure as the transition operator of Markov chain.

3. Face Synthesis based on Deep Learning

3.1. Convolutional Neural Networks

Recent years have witnessed the significant breakthroughs of CNN in face recognition. Similarly, CNN has been widely applied to face synthesis. Taigman *et al*. [51] proposed a DeepFace system to obtain the human-level performance in face recognition and Yi *et al*. [59] presented a 11-layers convolutional neural network to learning face representation. A gradient ascent approach is deployed by Zhmoginov and Sandler [69] to generate real-time face images. In order to recover high-quality facial pose, shape, expression, reflectance and illumination, Kim *et al*. [25] focused on an InverseFaceNet to obtain high-quality estimation.

Without requiring extensive pose coverage in training data, Peng *et al*. [39] exerted a synthesis network to generate non-frontal view from a single frontal face. Richardson *et al*. [46] deployed an end-to-end convolutional neural network, which utilizes two parts to achieve transformation from coarse to fine. Capturing long-term dependencies along a sequence of transformations of shifts and rotations by recurrent structure, Yang *et al*. [58] designed an end-to-end recurrent convolutional encoder-decoder network. Analogously inspired by encoder-decoder framework, Cole *et al*. [8] encoded the input face images into 1024-D feature vectors by FaceNet, and then used decoder network to generate the final images. Richardson *et al*. [45] put forward a CNN based approach which extracts the face geometry directly from original image. Zhang *et al*. [76] focused on inpainting corrupted face images through their DeMeshNet. They also proposed multi-task ConvNet [75] with skip connection to further improve synthetic image quality.

3.2. Generative Adversarial Nets

Goodfellow *et al*. [15] proposed a brand new deep framework called generative adversarial network (GAN). GAN can be understood as a two-player non-cooperative game model. Generator and discriminator, the main components of GAN, are fighting against each other, which can be formulated as the equation below:

$$\min_G \max_D V(D,G) = \boldsymbol{E}_{x\sim p_{data}(x)}[\log D(x)] + \boldsymbol{E}_{x\sim p_z(z)}[\log(1-D(G(z)))]$$

where G is the generator and D is the discriminator, p_x is the data distribution and p_z is a known noise distribution. Satisfactory results have shown that training the generator and discriminator alternatively through an iterative process with some stochastic gradient descent methods is experimentally effective.

Differing from traditional CNN, GAN is a generative model that can learn to fit the target data distribution. What's more, neither Markov chains nor inference is needed during learning, only back propagation is used to obtain gradients. However, original GAN suffers from some computational problems, *e.g*. the inferior performance caused by training generator without updating discriminator. Collapsed generator maps too many z to the same value of x and loses the capacity to fit the target data distribution. To address the aforementioned model collapse, Zhao *et al*. [66] proposed energy based GAN (EBGAN) and viewed generator and discriminator as energy functions. Further, Arjovsky *et al*. [2] presented Wasserstein GAN (WGAN) based on Earth Mover distance. They proved that WGAN gets rid of the collapse problem to some extent.

Current GAN models can handle the most headache cases, in which the PIE (pose, illumination and expression) of the input are under totally unconstrained situation, *e.g*. Radford *et al*. [42] designed a variation of GAN called DCGAN. Huang *et al*. [21] focused on the local patches that has some semantic meaning and proposed TPGAN. Li *et al*. [29] put their interest in the situation that parts of the face images are completely missing and came up with novel two adversarial losses as well as a semantic parsing loss to complete the faces. [14, 34] applied an extension of GAN to a conditional setting. VariGANs model is proposed by Zhao *et al*. [65] to solve the problem that generating multiview images from only single view point. Tran *et al*. [52] put forward DR-GAN which fuses the pose information and takes one or multiple face images with yaw angles as input to achieve pose invariant feature learning. Similarly, Antipov *et al*. [1] concentrated on improving face synthesis in cross-age scenarios. Considering scene structure and context, Yang *et al*. [57] presented LR-GAN that learns generated image background and foreground separately and recursively to produce a complete natural or face image.

Inspired by the capacity of fusing information of GAN, the application of face synthesis is also broadened. Kim *et al*. [26] built a pair of dual models that are capable of two-way image domain transfer. Differing from Kim's model that can alter images semantically, Liu and Tuzel [32] proposed a model to produce images in two different domains simultaneously. Zhou *et al*. [71] extracted feature representation with abstract semantic meanings and then "crossbred" to finish domain transfer. Yin *et al*. [60] presented a new method to learn to generate and modify the facial image coherently. A novel algorithm is put forward by Donahue *et al*. [10] to jointly learn latent codes for both identities and observations.

In general, methods based on GAN are basically on the behalf of state-of-the-art. Even human observers may be confused by synthetic face obtained from GAN model under some circumstances. However, ill-pose problem is yet an unresolved issue. It is worth noting that a framework combining 3DMM with GAN called FF-GAN was recently proposed by Yin *et al*. [61]. In their work, 3DMM is utilized to provide shape and appearance priors to converge fleetly and they add a novel masked symmetry loss to recover visual quality under occlusions.

3.3. Other Deep Neural Networks

Apart from CNN and GAN, lots of other deep neural network based methods have been developed. A novel face representation called the face identity-preserving (FIP) feature is proposed by Zhu *et al*. [72]. In order to better understand facial features and generate multi-view images, Zhu *et al*. [73] proposed multi-view perceptron (MVP). In addition, their architecture has the ability to interpolate and synthesize the viewpoints that are not appeared in the training set. To avoid the high dependence on training set for AAM models, Duong *et al*. [12] proposed a Deep Appearance Models (DAMs) approach, which uses Deep Boltzmann Machine (DBM) to robustly capture the variations of facial shapes and appearances.

4. Public Available Databases and Evaluation Metrics

4.1. Databases

To demonstrate the effectiveness of proposed methods, experimental results are analyzed on the data collected from Internet or existing public databases. Those databases have been established originally for face detection, recognition or alignment. Widely used databases are listed below:

Multi-PIE [16]: The CMU Multi-PIE face database contains more than 750,000 images of 337 people. The database contains more than 305 GB of high resolution face images in total. The pose, illumination, and expression are the main interesting factors of the database. In total, 15 view points and 19 illuminations and 7 expression conditions are recorded in controlled environment.

CASIA WebFace Database [59]: CASIA WebFace Database is a large-scale dataset containing 10,575 subjects and 494,414 images. The collection process started from the well-structured information in IMDB and then continued with the mean of web crawler.

CelebA [33]: CelebFaces Attributes Dataset is a large-scale face attributes database that includes 10,177 identities, 202,599 face images and each image has 5 landmark locations and 40 binary attributes. In addition, this database covers large pose variations and background clutters.

VGG Face Dataset [36]: The dataset consists of 2.6M images of 2,622 identities. The images are collected from Internet using Google and Bing API via a series of filtering strategies including manual operations. The dataset has quite large images outside publicly available industrial datasets as well as substantial diversity in PIE (pose, illumination, expression).

LFW [20, 28]: Labeled Faces in the Wild [20, 28] is an unconstraint face recognition database. It collects 13,000 face images from the web and each face has a label. In addition, there are 1680 people have two or more distinct photos in database.

FaceWarehouse [6]: It is a 3D facial expression database for visual computing. It includes 150 people aged from 7 to 80 that have different ethnic backgrounds and each person has various expressions.

4.2. Metrics

The quality of face synthesis is a semantic concept which can not be accurately quantified. Employing human annotators to judge the visual quality of synthetic samples can only work for situation with very limited data amount. Therefore, Kim *et al*. [25] presented 3 error estimation methods, the first is photometric error and it computes the RMSE of RGB pixel values by comparing the input image and a rendering of the reconstructed face model. The second is geometric error that computes the RMSE of corresponding error in millimeter between the ground-truth geometry and their 3D model. The third is overlap of face masks and it estimates the intersection over union of face masks between their model and the input image. Besides, the improvement of synthetic samples for downstream algorithm dealing with some visual tasks face reconstruction can reflect the quality in some sense. As an alternative to human annotators, Salimans *et al*. [48] proposed an automatic method to evaluate the performance of a GAN based model:

$$Inception\ score = \exp(\boldsymbol{E}_{\boldsymbol{x}} KL(p(y|\boldsymbol{x})||p(y)))$$

The evaluation above is derived under the consideration that an admirable model should produce images that have close analogies to real data as many as possible, where x denotes a generated sample and y is the label predicted by a some out-of-the-box classifier. To measure the error in 3D model, Zhao *et al*. [68] utilized the Euclidean distances to compute all the pairs of 3D points from ground truth model and their reconstructed model respectively.

5. Conclusion

In this paper, we survey the recently representative methods in face synthesis. Existing face synthesis methods are roughly divided into two categories: traditional methods and deep learning models. GAN has become a popular generative model recently and simultaneously obtains rapid progress, so generating more realistic, diverse and high-quality face images becomes possible. Although great progress in face synthesis has been made, still there is a space for improvement and we propose the following promising directions as a guide for future development.

- **High-resolution face synthesis**: Notwithstanding GAN and other models have greatly improved the quality of face, high-resolution face synthesis is still an open problem.
- **Cross-modal face synthesis**: NIR imaging has been widely employed as a way to avoid illumination changes in outdoor circumstances. How to combine the advantages of VIS and NIR to improve face synthesis is an interesting topic. Incorporating 3D model to rectify face image is also a promising direction, but 3D reconstruction itself is still an intricate issue.
- **Face synthesis via unpaired data**: It is expensive to collect large-scale paired data for face synthesis training. Hence how to use deep learning methods to learn from unpaired face data is an encouraging direction.
- **Lack of databases**: Because of the limitation of current face databases, more large-scale multi-pose and in-the-wild face databases are urged to build for the improvement of face synthesis.
- **Evaluation criterion**: Existing evaluation methods in face synthesis mainly contain two classes. One is evaluating indirectly by the improvement of down-stream algorithm that taking synthetic samples as input. Another method is to evaluate by human, which is a time-consuming and laborious work. How to design an objective, feasible, easily understanding and suitable calculating evaluation criterion remains challenging.

In the future, we would expect that more application of face synthesis will be developed and widely used in real-life. Furthermore, face synthesis will promote development of other tasks like face recognition, face detection, *etc*. Finally, the combination of several advanced techniques from multiple aspects is also helpful for face synthesis.

Acknowledgements

This work is funded by the National Natural Science Foundation of China (Grant No. 61622310, 61473289), the Youth Innovation Promotion Association CAS (Grant No. 2015190), the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDB02000000).

References

[1] G. Antipov, M. Baccouche, and J.-L. Dugelay. Face aging with conditional generative adversarial networks. In *International Conference on Image Processing*, 2017. 3

[2] M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein generative adversarial networks. In *International Conference on Machine Learning*, pages 214–223. 2017. 1, 3

[3] S. Banerjee, J. S. Bernhard, W. J. Scheirer, K. W. Bowyer, and P. J. Flynn. SREFI: Synthesis of realistic example face images. *arXiv preprint arXiv:1704.06693*, 2017. 2

[4] D. Bhattacharjee, S. Halder, M. Nasipuri, D. K. Basu, and M. Kundu. Construction of human faces from textual descriptions. *Soft Computing*, 15(3):429–447, 2011. 2

[5] F. Bordes, S. Honari, and P. Vincent. Learning to generate samples from noise through infusion training. In *International Conference on Learning Representations*, 2017. 3

[6] C. Cao, Y. Weng, S. Zhou, Y. Tong, and K. Zhou. Facewarehouse: A 3d facial expression database for visual computing. *IEEE Transactions on Visualization and Computer Graphics*, 20(3):413–425, 2014. 4

[7] L. Chang, M. Zhou, Y. Han, and X. Deng. Face sketch synthesis via sparse representation. In *International Conference on Pattern Recognition*, pages 2146–2149. IEEE, 2010. 2

[8] F. Cole, D. Belanger, D. Krishnan, A. Sarna, I. Mosseri, and W. T. Freeman. Synthesizing Normalized Faces from Facial Identity Features. In *International Conference on Computer Vision and Pattern Recognition*, 2017. 3

[9] Y. Dang, F. Li, Z. Li, and W. Zuo. Detail-enhanced cross-modality face synthesis via guided image filtering. In *CCF Chinese Conference on Computer Vision*, pages 200–209. Springer, 2015. 2

[10] C. Donahue, A. Balsubramani, J. McAuley, and Z. C. Lipton. Semantically decomposing the latent spaces of generative adversarial networks. *arXiv preprint arXiv:1705.07904*, 2017. 3

[11] P. Dou, Y. Wu, S. K. Shah, and I. A. Kakadiaris. Robust 3d face shape reconstruction from single images via two-fold coupled structure learning. In *British Conference on Machine Vision*, pages 1–13, 2014. 2

[12] C. N. Duong, K. Luu, K. G. Quach, and T. D. Bui. Deep appearance models: A deep boltzmann machine approach for face modeling. *arXiv preprint arXiv:1607.06871*, 2016. 1, 4

[13] C. Ferrari, G. Lisanti, S. Berretti, and A. Del Bimbo. Effective 3d based frontalization for unconstrained face recognition. In *International Conference on Pattern Recognition*. 2

[14] J. Gauthier. Conditional generative adversarial nets for convolutional face generation. *Class Project for Stanford CS231N: Convolutional Neural Networks for Visual Recognition, Winter semester*, 2014(5):2, 2014. 3

[15] 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. 1, 3

[16] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-pie. *Image and Vision Computing*, 28(5):807–813, 2010. 4

[17] S. Halder, D. Bhattacharjee, M. Nasipuri, D. K. Basu, and M. Kundu. Face synthesis (fasy) system for generation of a face image from human description. *arXiv preprint arXiv:1005.4034*, 2010. 1, 2

[18] J. Heo and M. Savvides. 3-d generic elastic models for fast and texture preserving 2-d novel pose synthesis. *IEEE Transactions on Information Forensics and Security*, 7(2):563–576, 2012. 2

[19] J. Heo and M. Savvides. Gender and ethnicity specific generic elastic models from a single 2d image for novel 2d pose face synthesis and recognition. *IEEE transactions on pattern analysis and machine intelligence*, 34(12):2341–2350, 2012. 2

[20] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, October 2007. 4

[21] R. Huang, S. Zhang, T. Li, and R. He. Beyond face rotation: Global and local perception gan for photorealistic and identity preserving frontal view synthesis. In *International Conference on Computer Vision*, 2017. 3

[22] X. Huang, J. Gao, S.-c. Cheung, and R. Yang. Manifold estimation in view-based feature space for face synthesis across poses. *Asian Conference on Computer Vision*, pages 37–47, 2010. 1, 2

[23] S. P. Jones, D. M. Dwyer, and M. B. Lewis. The utility of multiple synthesized views in the recognition of unfamiliar faces. *The Quarterly Journal of Experimental Psychology*, 70(5):906–918, 2017. 2

[24] F. Juefei-Xu and M. Savvides. Facial ethnic appearance synthesis. In *European Conference on Computer Vision*, pages 825–840. Springer, 2014. 2

[25] H. Kim, M. Zollhöfer, A. Tewari, J. Thies, C. Richardt, and C. Theobalt. Inversefacenet: Deep single-shot inverse face rendering from a single image. *arXiv preprint arXiv:1703.10956*, 2017. 3, 4

[26] T. Kim, M. Cha, H. Kim, J. Lee, and J. Kim. Learning to discover cross-domain relations with generative adversarial networks. In *International Conference on Machine Learning*, 2017. 3

[27] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012. 1

[28] G. B. H. E. Learned-Miller. Labeled faces in the wild: Updates and new reporting procedures. Technical Report UM-CS-2014-003, University of Massachusetts, Amherst, May 2014. 4

[29] Y. Li, S. Liu, J. Yang, and M.-H. Yang. Generative face completion. In *International Conference on Computer Vision and Pattern Recognition*, 2017. 3

[30] F. Liu, D. Zeng, J. Li, and Q. Zhao. 3d face reconstruction via cascaded regression in shape space. *arXiv preprint arXiv:1509.06161*, 2015. 2

[31] F. Liu, D. Zeng, Q. Zhao, and X. Liu. Joint face alignment and 3d face reconstruction. In *European Conference on Computer Vision*, pages 545–560. Springer, 2016. 2

[32] M.-Y. Liu and O. Tuzel. Coupled generative adversarial networks. In *Advances in Neural Information Processing Systems*, pages 469–477, 2016. 3

[33] Z. Liu, P. Luo, X. Wang, and X. Tang. Deep learning face attributes in the wild. In *International Conference on Computer Vision*, 2015. 4

[34] Y. Lu, Y.-W. Tai, and C.-K. Tang. Conditional cyclegan for attribute guided face image generation. *arXiv preprint arXiv:1705.09966*, 2017. 3

[35] M. H. Nguyen, J.-F. Lalonde, A. A. Efros, and F. De la Torre. Image-based shaving. In *Computer Graphics Forum*, volume 27, pages 627–635. Wiley Online Library, 2008. 2

[36] O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. In *British Conference on Machine Vision*, 2015. 4

[37] A. Patel and W. A. Smith. Driving 3d morphable models using shading cues. *Pattern Recognition*, 45(5):1993–2004, 2012. 1, 2

[38] C. Peng, X. Gao, N. Wang, D. Tao, X. Li, and J. Li. Multiple representations-based face sketch–photo synthesis. *IEEE transactions on neural networks and learning systems*, 27(11):2201–2215, 2016. 2

[39] X. Peng, X. Yu, K. Sohn, D. Metaxas, and M. Chandraker. Reconstruction for feature disentanglement in pose-invariant face recognition. *arXiv preprint arXiv:1702.03041*, 2017. 1, 3

[40] M. Piotraschke and V. Blanz. Automated 3d face reconstruction from multiple images using quality measures. In *International Conference on Computer Vision and Pattern Recognition*, pages 3418–3427, 2016. 2

[41] X. Qiao, X.-H. Han, T. Igarashi, K. Nakao, and Y.-W. Chen. Tensor-based subspace learning and its applications in multi-pose face synthesis. *Neurocomputing*, 73(13):2727–2736, 2010. 1, 2

[42] A. Radford, L. Metz, and S. Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015. 3

[43] S. Ramalingam and N. T. Viet. 3d face synthesis with kinect. In *International Conference on Systems, Man, and Cybernetics*, pages 4433–4438. IEEE, 2013. 1, 2

[44] Y. Ren, Y. Romano, and M. Elad. Example-based image synthesis via randomized patch-matching. *arXiv preprint arXiv:1609.07370*, 2016. 2

[45] E. Richardson, M. Sela, and R. Kimmel. 3d face reconstruction by learning from synthetic data. In *International Conference on 3D Vision*, pages 460–469. IEEE, 2016. 1, 3

[46] E. Richardson, M. Sela, R. Or-El, and R. Kimmel. Learning detailed face reconstruction from a single image. In *International Conference on Computer Vision and Pattern Recognition*, 2017. 3

[47] C. Sagonas, Y. Panagakis, S. Zafeiriou, and M. Pantic. Robust statistical face frontalization. In *International Conference on Computer Vision*, pages 3871–3879, 2015. 2

[48] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen. Improved techniques for training gans. In *Conference on Neural Information Processing Systems*, 2016. Yan, Y. Ding, P. Li, Q. Wang, Y. Xu, and W. Zuo. Mind the Class Weight Bias: Weighted Maximum Mean Discrepancy for Unsupervised Domain Adaptation. In *International Conference on Computer Vision and Pattern Recognition*, 2017. 4

[49] H. Shim, I. Ha, T. Rhee, J. D. Kim, and C. Kim. A probabilistic approach to realistic face synthesis. In *International Conference on Image Processing*, pages 1817–1820. IEEE, 2010. 2

[50] C. Song, F. Li, Y. Dang, H. Gao, Z. Yan, and W. Zuo. Structured detail enhancement for cross-modality face synthesis. *Neurocomputing*, 212:107–120, 2016. 3

[51] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf. Deepface: Closing the gap to human-level performance in face verification. In *International Conference on Computer Vision and Pattern Recognition*, pages 1701–1708, 2014. 1, 3

[52] L. Tran, X. Yin, and X. Liu. Disentangled representation learning gan for pose-invariant face recognition. In *International Conference on Computer Vision and Pattern Recognition*, volume 3, page 7, 2017. 3

[53] N. Wang, X. Gao, and J. Li. Random sampling and locality constraint for face sketch. *arXiv preprint arXiv:1701.01911*, 2017. 2

[54] X. Wang and X. Tang. Face photo-sketch synthesis and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(11):1955–1967, 2009. 2

[55] Y. Wang, L. Zhang, Z. Liu, G. Hua, Z. Wen, Z. Zhang, and D. Samaras. Face relighting from a single image under arbitrary unknown lighting conditions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(11):1968–1984, 2009. 2

[56] H. Xu and H. Zha. Manifold based face synthesis from sparse samples. In *International Conference on Computer Vision*, pages 2208–2215, 2013. 2

[57] J. Yang, A. Kannan, D. Batra, and D. Parikh. Lr-gan: Layered recursive generative adversarial networks for image generation. In *International Conference on Learning Representations*, 2017. 3

[58] J. Yang, S. E. Reed, M.-H. Yang, and H. Lee. Weakly-supervised disentangling with recurrent transformations for 3d view synthesis. In *Advances in Neural Information Processing Systems*, pages 1099–1107, 2015. 3

[59] D. Yi, Z. Lei, S. Liao, and S. Z. Li. Learning face representation from scratch. *arXiv preprint arXiv:1411.7923*, 2014. 1, 3, 4

[60] W. Yin, Y. Fu, L. Sigal, and X. Xue. Semi-latent gan: Learning to generate and modify facial images from attributes. In *International Conference on Computer Vision*, 2017. 3

[61] X. Yin, X. Yu, K. Sohn, X. Liu, and M. Chandraker. Towards large-pose face frontalization in the wild. In *International Conference on Computer Vision*, 2017. 4

[62] Y. Song and H. Zhang. A framework of face synthesis based on multilinear analysis. In *ACM SIGGRAPH Conference on Virtual-Reality Continuum and Its Applications in Industry-Volume 1*, pages 111–114. ACM, 2016.

[63] S. Zhang, X. Gao, N. Wang, J. Li, and M. Zhang. Face sketch synthesis via sparse representation-based greedy search. *IEEE transactions on image processing*, 24(8):2466–2477, 2015. 2

[64] Z. Zhang, Y. Wang, and Z. Zhang. Face synthesis from low-resolution near-infrared to high-resolution visual light spectrum based on tensor analysis. *Neurocomputing*, 140:146–154, 2014. 1, 2

[65] B. Zhao, X. Wu, Z.-Q. Cheng, H. Liu, and J. Feng. Multi-view image generation from a single-view. *arXiv preprint arXiv:1704.04886*, 2017. 3

[66] J. Zhao, M. Mathieu, and Y. LeCun. Energy-based generative adversarial network. In *International Conference on Learning Representations*, 2017. 1, 3

[67] L. Zhao, X. Gao, Y. Yuan, and D. Tao. Sparse frontal face image synthesis from an arbitrary profile image. *Neurocomputing*, 128:466–475, 2014. 2

[68] Y. Zhao, Q. Xu, X. Huang, and R. Yang. Mask-off: Synthesizing face images in the presence of head-mounted displays. *arXiv preprint arXiv:1610.08481*, 2016. 2, 4

[69] A. Zhmoginov and M. Sandler. Inverting face embeddings with convolutional neural networks. *arXiv preprint arXiv:1606.04189*, 2016. 1, 3

[70] J. Zhong, X. Gao, and C. Tian. Face sketch synthesis using e-hmm and selective ensemble. In *International Conference on Acoustics, Speech and Signal Processing*, volume 1, pages I–485. IEEE, 2007. 2

[71] S. Zhou, T. Xiao, Y. Yang, D. Feng, Q. He, and W. He. Genegan: Learning object transfiguration and attribute subspace from unpaired data. *arXiv preprint arXiv:1705.04932*, 2017. 3

[72] Z. Zhu, P. Luo, X. Wang, and X. Tang. Deep learning identity-preserving face space. In *International Conference on Computer Vision*, pages 113–120, 2013. 4

[73] Z. Zhu, P. Luo, X. Wang, and X. Tang. Multi-view perceptron: a deep model for learning face identity and view representations. In *Advances in Neural Information Processing Systems*, pages 217–225, 2014. 1, 4

[74] Y. Zhuang, Y. Wang, J. Xiao, and Y. Wang. Expressive 3d face synthesis by multi-space modeling. In *International Conference on Ubi-Media Computing*, pages 207–212. IEEE, 2008. 2

[75] S. Zhang, R. He, Z. Sun, and T. Tan. Multi-task convnet for blind face inpainting with application to face verification. In *International Conference on Biometrics*, pages 1–8. IEEE, 2016. 3

[76] S. Zhang, R. He, and T. Tan. DeMeshNet: Blind Face Inpainting for Deep MeshFace Verification. *arXiv preprint arXiv:1611.05271*, 2016. 3