
CADA-GAN: Context-Aware GAN with Data Augmentation

Sofie Daniëls^{* 1} Jiugeng Sun^{* 1} Jiaqing Xie^{* 1}

Abstract

Current child face generators are restricted by the limited size of the available datasets. In addition, feature selection can prove to be a significant challenge, especially due to the large amount of features that need to be trained for. To manage these problems, we proposed CADA-GAN, a Context-Aware GAN that allows optimal feature extraction, with added robustness from additional Data Augmentation. CADA-GAN is adapted from the popular StyleGAN2-Ada model, with attention on augmentation and segmentation of the parent images. The model has the lowest *Mean Squared Error Loss* (MSEloss) on latent feature representations and the generated child image is robust compared with the one that generated from baseline models.

1. Introduction

Generative adversarial networks (GANs) have become increasingly popular in recent years, in particular due to their ability to solve complex image-based problems, such as image generation (Bao et al., 2017), image-to-image translation (Liu et al., 2017a), and face recognition (Tran et al., 2017). More specifically, GANs can be used to generate realistic faces in high-resolution (Wang et al., 2022; Kammoun et al., 2022; Wu et al., 2017). One application of that is kinship face generation.

Prospective parents are typically highly interested in what their children will look like, and historians dream of automatically figuring out family trees from a limited number of portraits. Kinship face generation can greatly help with this. In this paper, we specifically focus on the generation of faces of children using the images of their biological parents as input.

Some earlier papers rely on a one-to-one kinship relation (Dahan & Keller, 2018), but as better datasets are being published, more recent papers have allowed the use of multiple relatives as input (Zhang et al., 2020; Ghatas & Hemayed, 2020; Gao et al., 2021; Cui et al., 2021; Lin et al., 2021a).

A major limitation of current works, however, is the limited dataset size. Obtaining father-mother-child triplets is difficult, and only few such datasets are available (TSK; Robinson et al., 2016). In addition, some datasets are highly

pre-processed and do not allow generalization of existing algorithms to less homogeneous datasets.

To combat this, we propose two additional pre-processing steps. We first augment the images with various existing data augmentation techniques, in order to improve robustness of our model. In addition, we segment the images of the parents to allow the adversarial network to focus on the main features of the face.

The main contributions of this work are as follows:

1. We propose a model that improves robustness of child face generation by applying simple augmentations to the parent images.
2. Using segmentation, our model additionally allows targeted highlighting of facial features for optimal child feature prediction.
3. We employ transfer learning in order to facilitate downstream latent vector extraction of both parents and image reconstruction of the child.

2. Related Work

Child Face Generation. Relatively few papers have been written on the topic of child face generation. Still, promising results were achieved using CDFS-GAN, GANKin, DNA-net, ChildGAN, and StyleDNA (Zhang et al., 2020; Ghatas & Hemayed, 2020; Gao et al., 2021; Cui et al., 2021; Lin et al., 2021a).

CDFS-GAN is one of the first child face generators that uses both parents as input (Zhang et al., 2020). By extracting main facial features, such as nose, mouth, and eyes separately, CDFS-GAN models feature-dependent genetic inheritance.

In contrast, GANKin directly extracts the features of the parents using FaceNET (Schroff et al., 2015) and uses a four-layer fully connected network to determine the child feature vector (Ghatas & Hemayed, 2020). The child face is generated from this feature vector using PGGAN (Karras et al., 2018).

DNA-net maps features to genes and back to incorporate genetic knowledge in the child face prediction (Gao et al., 2021). The parent images are mapped to the feature space by training a conditional adversarial auto-encoder (CAAE) as proposed in (Zhang et al., 2017). Features are selected at random, in order to model the real world more accurately. The final child face is generated using the decoder from the

^{*}Equal contribution ¹Computer Science, ETH Zurich.

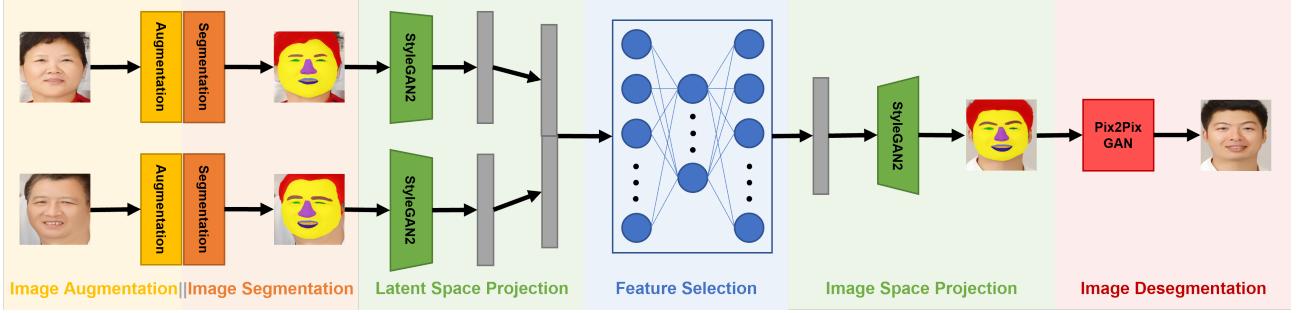


Figure 1. Pipeline of context-aware GAN with data augmentation.

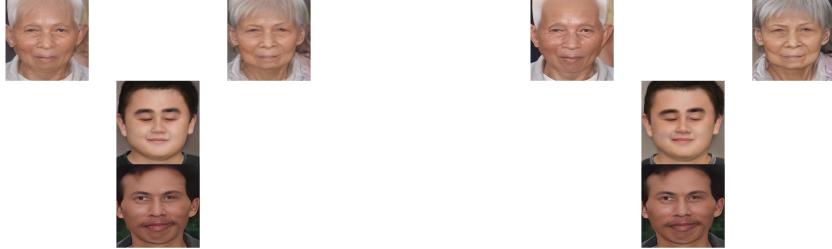


Figure 2. Generated Child Images (Baseline)

Figure 3. Generated Child Images (Mixup)



Figure 4. Generated Child Images (Segmentation)

CAAE.

Similar to DNA-net, ChildGAN (Cui et al., 2021) also relies on genetic knowledge to improve child face generation, but combines it with a semantic learning framework. First, the parent images are projected into the latent space using Image2StyleGAN (Abdal et al., 2019). After obtaining and finetuning the child vector with a macro- and microfusion step, the final child image is generated using StyleGAN (Karras et al., 2019).

(Lin et al., 2021a) instead use StyleGAN2 for latent space embedding. They compare an improved DNA-net based model, StyleDNA, with both a k-nearest neighbors and an eigenvector projection approach.

Most papers perform feature selection in the latent space to account for the imbalance between the limited dataset

size and the number of learnable parameters. In this paper, we will therefore use StyleGAN2-Ada (Karras et al., 2020) for image-to-latent and latent-to-image space projection.

Image Augmentation. Data augmentation encompasses a range of techniques that increase the size of the dataset and improve robustness of training (Shorten & Khoshgoftaar, 2019). Augmentation can be performed by deep neural networks or more fundamental image manipulation methods like CutOut (DeVries & Taylor, 2017) and MixUp (Hendrycks et al., 2020). One exciting strategy for augmentation is generative modelling, such as in (Radford et al., 2016), (Karras et al., 2018), and (Mirza & Osindero, 2014). GANs are a method for extracting new information from a dataset, according to (Bowles et al., 2018). In Deep Learning research, the idea of meta-learning mainly relates to the

idea of neural network optimization via neural networks (Shorten & Khoshgoftaar, 2019), which has been utilised in the Data Augmentation field as well (Perez & Wang, 2017; Lemley et al., 2017; Cubuk et al., 2019). Though its creative uses are arguably where neural style transfer is best recognized, it is also a fantastic tool for data augmentation (Johnson et al., 2016; Gatys et al., 2015). The techniques we are utilising are MixUp (Zhang et al., 2018) and AugMix (Hendrycks et al., 2020).

Image Segmentation. Segmentation of facial features has many applications in the field of computer vision, such as landmark detection, head pose estimation, recognition of facial expressions, and face recognition (Khan et al., 2020).

Conditional random fields (CRFs) are commonly used as the basis for human face segmentation, either implemented as is (Warrell & Prince, 2009; Khan et al., 2017) or in the form of a convolutional network (Liu et al., 2015; Zhou et al., 2017). Convolutional networks can also be combined with recurrent networks, such as in (Liu et al., 2017b) and (Zhou, 2017). Many other algorithms can be used for facial feature extraction, like random forests (Khan et al., 2015) and support vector machines (SVMs) (Khan et al., 2018). Still, most of the available segmentation models require extensive and error-prone preprocessing steps, in the form of either facial landmark detection, image cropping, or image alignment. In this paper, we therefore rely on the pretrained model of (Lin et al., 2021b), who propose a scale, rotation, and transformation equivariant model with competitive accuracy.

3. Proposed method: CADA-GAN

Our proposed model consists of three main steps: 1. data augmentation, 2. segmentation, and 3. face generation. The full pipeline can be found in Figure 1.

Image Augmentation. In the first step, we augment the images of the parents using either MixUp or AugMix. MixUp (Zhang et al., 2018) generates a weighted combination of image pairs from the training data, while AugMix (Hendrycks et al., 2020) mixes augmented images through linear interpolation. All images are augmented with a probability P .

For MixUp, we follow the setup from (Zhang et al., 2018):

$$\begin{cases} \text{img}'_f = \alpha * \text{img}_f + (1 - \alpha) * \text{img}_m \\ \text{img}'_m = \beta * \text{img}_m + (1 - \beta) * \text{img}_f \end{cases} \quad (1)$$

with img_f and img_m denoting the original images of the father and mother respectively. α and β are drawn independently and at random from the interval $[0,1]$. Images of the children are not altered in any way. The goal of MixUp is to approximate the feature space by combining as many instances of the input feature vector as possible, thus strengthening the generality of the generation. For an

example, see Figure 5.



Figure 5. Comparison of original images of a father and mother (left) with images augmented with MixUp (right).

AugMix chooses an appropriate image manipulation method at random and applies it to the pictures of the parents. The available techniques are shearing, translation, small-degree rotation, and horizontal flipping.

$$\text{Img} = \mathcal{F}(\text{Img}) \quad (2)$$

where $\mathcal{F} \in \{\text{shear}, \text{translate}, \text{rotate}, \text{flip}\}$.

Translation and shearing can be applied with respect to either the x-axis or the y-axis. We omit the more drastic modifications from the original literature (Hendrycks et al., 2020), as we intend to retain most of the facial features. Like in MixUp, we only apply transformations to the parent images.

Image Segmentation. In the second preprocessing step, we apply segmentation to facilitate the downstream extraction of the main facial features of the parents. For this, we use the pretrained model from (Lin et al., 2021b), which segments images of faces into 11 classes: background, hair, nose, left eye, right eye, left eyebrow, right eyebrow, upper lip, lower lip, inner mouth, and skin. (Figure ?? in Appendix.)

Child Face Generation After preprocessing, we convert the segmented faces of the parents into latent vectors \mathbf{g}_{m_i} and \mathbf{g}_{f_i} using (Karras et al., 2020), where $\mathbf{g}()$ performs augmentation and segmentation, and embeds vectors to latent space. (m_i, f_i) is a pair of parent images sampled from dataset $M \odot F$. We first finetune a pretrained model trained on the FFHQ dataset with images from the TSKinFace Dataset. Considering the slow convergence of tuning pretrained parameters on a new model, we apply transfer learning with a training session of 15 ticks. The overall training loss for generating a child image in the latent space is:

$$\text{rloss} = \sum_{\substack{(m_i, f_i, c_i) \\ \in \{M, F, C\}}} \mathcal{L}(G(\text{aggr}(\mathbf{g}_{m_i}, \mathbf{g}_{f_i})), \mathbf{c}_i), \quad (3)$$

, where G is the generator from the GAN and \mathcal{L} is the mean squared error (MSE) loss. Similar to GANkin (Ghatas & Hemayed, 2020), we then use a MLP with two fully connected layers (see Table 1) as our aggregation function aggr to predict the latent features of the child. The MSE loss is calculated between the original and predicted latent vectors of the child. We train with a batch size of 16 and

use Adam as the optimization algorithm with a learning rate of 0.00001.

De-Segmentation. To reverse the predicted segmentation of the child face into a real image at the end of the pipeline, we rely on a pix2pix GAN as proposed in (Isola et al., 2017). To account for the color variations in the StyleGAN2 output, we add a random value \mathbf{x} to the hue of the original image and modify the saturation by \mathbf{y} . We restrict the range of \mathbf{x} and \mathbf{y} to [-5, 5].

$$\text{Img}'_h = (\text{Img}_h + \mathbf{x}) \bmod 180 \quad (4)$$

$$\text{Img}'_s = \min(255, \max(0, \text{Img}_s + \mathbf{y})) \quad (5)$$

Datasets. We train and test the entire model with the TSKinFace Dataset (TSK). TSKinFace is grouped into three family compositions, Father-Mother-Daughter (FM-D), Father-Mother-Son (FM-S), and Father-Mother-Son-Daughter (FM-SD), which allows straightforward extraction of the parent and corresponding child images. For computational reasons, we will only work with the FMS dataset.

The problem of poor reconstruction results from StyleGAN on low-resolution data inspires us to first transform our images to a high resolution image (512x512). We take advantage of SPARNET (Chen et al., 2020) to perform such transformations.

The pix2pix GAN is trained on images from the Flickr-Faces-HQ (FFHQ) datasets (Lin et al., 2021b; Karras et al., 2019). All images are resized to a resolution of 256x256 and segmented using the model from (Lin et al., 2021b). Due to the vast size of the FFHQ-dataset, we only use 10 000 images to train the pix2pix GAN and are able to segment 9 872 correctly.

Code and Execution. All code was executed in the High Performance Computing Clusters of ETH Zürich on an NVIDIA GeForce GTX 2080 Ti GPU, where the full pipeline can be trained and tested in 18 hours on average. We also tested our code on an NVIDIA A100 GPU, where training of the baseline took only 15 hours.

4. Evaluation

Transfer learning 15 ticks are performed according to our transfer learning model. At the very beginning, generated parents and child images are unlike to the original individual input since the pretrained model has learned nothing from the dataset. After 15 ticks, the models are retrained well so

Layer	Input Size	Output Size
Fully Connected	(None, 2x16x512)	(None, 512)
ReLU		
Dropout (p=0.25)		
Fully Connected	(None, 512)	(None, 1x16x512)

Table 1. Layers of MLP used for feature extraction.

we are convinced to take it as the new pretrained model to be further fine-tuned quickly ??.

Experiments For evaluation, we compare four different scenarios. Our baseline consists of the original GAN implementation with a four-layer network used for child feature prediction. In the first and second experiment, we apply image augmentation to the parent images in order to overcome the limited dataset size. We use either MixUp or AugMix for this. Both experiments did not show improvement when compared to the baseline, see Table 2. In the third experiment, we test the effects of segmentation and parse the parent faces with 11 labels. After feature prediction and image generation, the resulting child segmentation is converted back into a final realistic image with the pix2pix GAN. The results were significantly better than the baseline, as can be seen in Table 2. The final experiment on the entire pipeline showed an averaged out result between segmentation and augmentation. Still, we believe the added robustness is worth the loss in accuracy.

Evaluation Metric Many papers measure the cosine similarity between the predicted latent vector of the child and the latent vectors of the parents, and compare it with the cosine similarity between the real child and parental latent vectors (Ghatas & Hemayed, 2020; Cui et al., 2021; Gao et al., 2021). The cosine similarity equation is given by:

$$\cos_{\text{sim}} = 1 - \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\| \|\mathbf{v}\|} \quad (6)$$

, as per the official SciPy documentation, with u and v two vectors or flattened matrices. If they are similar to each other element-wise, then the subtracted term in equation 6 is equal to one, meaning that the lefthand term \cos_{sim} is equal to zero. Recent papers consider it a success if the predicted latent vector is more similar to the parents than the original one. However, since some feature selection mechanisms automatically predict latent features similar to those of the parents, this metric does not account for the true genetic variety in the population. Therefore, we opt to evaluate the cosine similarity between the predicted and real child image. Since the main goal of child face generation isn't to generate as many different potential siblings as possible, we find that our metric is less subjective and less biased

Methods	MSE (lat)	MSE (images)
StyleGAN2	2.45	3372.53
StyleGAN2 + Mixup	2.46	3485.57
StyleGAN2 + AugMix	2.47	3481.18
StyleGAN2 + Segmentation	2.06	3343.21
CADAGAN (Ours)	2.17	5518.13

Table 2. Validation MSE loss for both latent vectors and original images

. During the training, we do consider the similarity between latent vectors. Therefore the MSE loss would become smaller.

towards the feature selection method. The results can be seen in Table 3.

Methods	cosine similarity
StyleGAN2	2
StyleGAN2 + Mixup	2
StyleGAN2 + AugMix	2
CADAGAN (Ours)	0.15

Table 3. Cosine Similarity between generated images and original images. The lower is better. The maximum value is 2.

5. Conclusion and Future work

We proposed CADA-GAN, a novel GAN model to deal with child face synthesis considering face segmentation and augmentation from original data. We tried different augmentation methods to expand our datasets, which brought invariance effects when applied to convolutional networks and strengthened the model. Applying segmentations lowered the general reconstruction loss in comparison to the baseline model. In addition, latent vector extractions for StyleGAN2 with a transfer learning model greatly speeds up training.

There could be two limitations according to our observations on CADA-GAN. We only tested two picture alteration techniques for the augmentation part. We might try some other methods based on equivariant or invariant operations on images, such as SmartAugment (Lemley et al., 2017), which can be integrated into the pipeline. In addition, instead of highlighting facial features with a color segmentation, blurring of the non-important features, such as background, hair, and skin, could potentially further improve feature extraction, while simultaneously reducing the color artifacts produced by the pix2pix GAN.

Although our model did not perform adequately, we still believe in the potential added robustness and awareness of our proposed architecture.

References

- The Closed Eyes in the Wild (CEW) dataset. URL http://parnec.nuua.edu.cn/_upload/tpl/02/db/731/template731/pages/xtan/TSkinFace.html.
- Abdal, R., Qin, Y., and Wonka, P. Image2stylegan: How to embed images into the stylegan latent space? *CoRR*, abs/1904.03189, 2019. URL <http://arxiv.org/abs/1904.03189>.
- Bao, J., Chen, D., Wen, F., Li, H., and Hua, G. Cvae-gan: fine-grained image generation through asymmetric training. In *Proceedings of the IEEE international conference on computer vision*, pp. 2745–2754, 2017.
- Bowles, C., Chen, L., Guerrero, R., Bentley, P., Gunn, R., Hammers, A., Dickie, D. A., Hernández, M. V., Wardlaw, J., and Rueckert, D. GAN Augmentation: Augmenting Training Data using Generative Adversarial Networks, October 2018. URL <http://arxiv.org/abs/1810.10863>. arXiv:1810.10863 [cs].
- Chen, C., Gong, D., Wang, H., Li, Z., and Wong, K.-Y. K. Learning spatial attention for face super-resolution. 2020.
- Cubuk, E. D., Zoph, B., Mane, D., Vasudevan, V., and Le, Q. V. AutoAugment: Learning Augmentation Policies from Data, April 2019. URL <http://arxiv.org/abs/1805.09501>. arXiv:1805.09501 [cs, stat].
- Cui, X., Zhou, W., Hu, Y., Wang, W., and Li, H. Heredity-aware child face image generation with latent space disentanglement. *arXiv preprint arXiv:2108.11080*, 2021.
- Dahan, E. and Keller, Y. Selfkin: Self adjusted deep model for kinship verification. *CoRR*, abs/1809.08493, 2018. URL <http://arxiv.org/abs/1809.08493>.
- DeVries, T. and Taylor, G. W. Improved regularization of convolutional neural networks with cutout, November 2017. URL <http://arxiv.org/abs/1708.04552>. arXiv:1708.04552 [cs].
- Gao, P., Robinson, J., Zhu, J., Xia, C., Shao, M., and Xia, S. Dna-net: Age and gender aware kin face synthesizer. In *2021 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6. IEEE, 2021.
- Gatys, L. A., Ecker, A. S., and Bethge, M. A Neural Algorithm of Artistic Style, September 2015. URL <http://arxiv.org/abs/1508.06576>. arXiv:1508.06576 [cs, q-bio].
- Ghatas, F. and Hemayed, E. Gankin: generating kin faces using disentangled gan. *SN Applied Sciences*, 2, 02 2020. doi: 10.1007/s42452-020-1949-3.
- Hendrycks, D., Mu, N., Cubuk, E. D., Zoph, B., Gilmer, J., and Lakshminarayanan, B. AugMix: A Simple Data Processing Method to Improve Robustness and Uncertainty, February 2020. URL <http://arxiv.org/abs/1912.02781>. arXiv:1912.02781 [cs, stat].
- Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. Image-to-image translation with conditional adversarial networks. *CVPR*, 2017.
- Johnson, J., Alahi, A., and Fei-Fei, L. Perceptual Losses for Real-Time Style Transfer and Super-Resolution, March 2016. URL <http://arxiv.org/abs/1603.08155>. arXiv:1603.08155 [cs].

- Kammoun, A., Slama, R., Tabia, H., Ouni, T., and Abid, M. Generative adversarial networks for face generation: A survey. *ACM Computing Surveys (CSUR)*, 2022.
- Karras, T., Aila, T., Laine, S., and Lehtinen, J. Progressive Growing of GANs for Improved Quality, Stability, and Variation, February 2018. URL <http://arxiv.org/abs/1710.10196>. arXiv:1710.10196 [cs, stat].
- Karras, T., Laine, S., and Aila, T. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4401–4410, 2019.
- Karras, T., Aittala, M., Hellsten, J., Laine, S., Lehtinen, J., and Aila, T. Training generative adversarial networks with limited data. *CoRR*, abs/2006.06676, 2020. URL <https://arxiv.org/abs/2006.06676>.
- Khan, K., Mauro, M., and Leonardi, R. Multi-class semantic segmentation of faces. In *2015 IEEE International Conference on Image Processing (ICIP)*, pp. 827–831, 2015. doi: 10.1109/ICIP.2015.7350915.
- Khan, K., Ahmad, N., Ullah, K., and Din, I. Multi-class semantic segmentation of faces using crfs. *TURKISH JOURNAL OF ELECTRICAL ENGINEERING & COMPUTER SCIENCES*, 25:3164–3174, 01 2017. doi: 10.3906/elk-1607-332.
- Khan, K., Syed, I., Khan, M., Ehsan, M., Uddin, I., and Ahmad, N. Fpl-an end-to-end face parts labeling framework. In *2018 24th International Conference on Automation and Computing (ICAC)*, pp. 1–6, 09 2018. doi: 10.23919/IConAC.2018.8748976.
- Khan, K., Khan, R. U., Ahmad, K., Ali, F., and Kwak, K.-S. Face segmentation: A journey from classical to deep learning paradigm, approaches, trends, and directions. *IEEE Access*, 8:58683–58699, 2020. doi: 10.1109/ACCESS.2020.2982970.
- Lemley, J., Bazrafkan, S., and Corcoran, P. Smart Augmentation - Learning an Optimal Data Augmentation Strategy. *IEEE Access*, 5:5858–5869, 2017. ISSN 2169-3536. doi: 10.1109/ACCESS.2017.2696121. URL <http://arxiv.org/abs/1703.08383>. arXiv:1703.08383 [cs, stat].
- Lin, C.-H., Chen, H.-C., Cheng, L. C., Hsu, S.-C., Chen, J.-C., and Wang, C.-Y. Styledna: A high-fidelity age and gender aware kinship face synthesizer. In *2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021)*, pp. 1–8, 2021a. doi: 10.1109/FG52635.2021.9666998.
- Lin, Y., Shen, J., Wang, Y., and Pantic, M. Roi tanh-polar transformer network for face parsing in the wild. *Image and Vision Computing*, 112:104190, 2021b. ISSN 0262-8856. doi: <https://doi.org/10.1016/j.imavis.2021.104190>. URL <https://www.sciencedirect.com/science/article/pii/S0262885621000950>.
- Liu, M.-Y., Breuel, T., and Kautz, J. Unsupervised image-to-image translation networks. *Advances in neural information processing systems*, 30, 2017a.
- Liu, S., Yang, J., Huang, C., and Yang, M.-H. Multi-objective convolutional learning for face labeling. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3451–3459, 2015. doi: 10.1109/CVPR.2015.7298967.
- Liu, S., Shi, J., Liang, J., and Yang, M. Face parsing via recurrent propagation. *CoRR*, abs/1708.01936, 2017b. URL <http://arxiv.org/abs/1708.01936>.
- Mirza, M. and Osindero, S. Conditional Generative Adversarial Nets, November 2014. URL <http://arxiv.org/abs/1411.1784>. arXiv:1411.1784 [cs, stat].
- Perez, L. and Wang, J. The Effectiveness of Data Augmentation in Image Classification using Deep Learning, December 2017. URL <http://arxiv.org/abs/1712.04621>. arXiv:1712.04621 [cs].
- Radford, A., Metz, L., and Chintala, S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, January 2016. URL <http://arxiv.org/abs/1511.06434>. arXiv:1511.06434 [cs].
- Robinson, J. P., Shao, M., Wu, Y., and Fu, Y. Family in the wild (FIW): A large-scale kinship recognition database. *CoRR*, abs/1604.02182, 2016. URL <http://arxiv.org/abs/1604.02182>.
- Schroff, F., Kalenichenko, D., and Philbin, J. Facenet: A unified embedding for face recognition and clustering. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 815–823, 2015. doi: 10.1109/CVPR.2015.7298682.
- Shorten, C. and Khoshgoftaar, T. M. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1):60, July 2019. ISSN 2196-1115. doi: 10.1186/s40537-019-0197-0. URL <https://doi.org/10.1186/s40537-019-0197-0>.
- Tran, L., Yin, X., and Liu, X. Disentangled representation learning gan for pose-invariant face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1415–1424, 2017.

Wang, X., Guo, H., Hu, S., Chang, M.-C., and Lyu, S. Gan-generated faces detection: A survey and new perspectives. *arXiv preprint arXiv:2202.07145*, 2022.

Warrell, J. and Prince, S. J. Labelfaces: Parsing facial features by multiclass labeling with an epitome prior. In *2009 16th IEEE International Conference on Image Processing (ICIP)*, pp. 2481–2484, 2009. doi: 10.1109/ICIP.2009.5413918.

Wu, X., Xu, K., and Hall, P. A survey of image synthesis and editing with generative adversarial networks. *Tsinghua Science and Technology*, 22(6):660–674, 2017.

Zhang, H., Cisse, M., Dauphin, Y. N., and Lopez-Paz, D. Mixup: Beyond empirical risk minimization. pp. 13, 2018.

Zhang, Y., Li, L., Liu, Z., Wu, B., Fan, Y., and Li, Z. Controllable descendant face synthesis. *CoRR*, abs/2002.11376, 2020. URL <https://arxiv.org/abs/2002.11376>.

Zhang, Z., Song, Y., and Qi, H. Age progression/regression by conditional adversarial autoencoder. *CoRR*, abs/1702.08423, 2017. URL <http://arxiv.org/abs/1702.08423>.

Zhou, L., Liu, Z., and He, X. Face parsing via a fully-convolutional continuous CRF neural network. *CoRR*, abs/1708.03736, 2017. URL <http://arxiv.org/abs/1708.03736>.

Zhou, Y. Top-down sampling convolution network for face segmentation. In *2017 3rd IEEE International Conference on Computer and Communications (ICCC)*, pp. 1893–1897, 2017. doi: 10.1109/CompComm.2017.8322867.