FACE RECOGNITION WITH DISENTANGLED FACIAL REPRESENTATION LEARNING AND DATA AUGMENTATION

Chia-Hao Tang[1], *Gee-Sern Jison Hsu*[1], *Moi Hoon Yap*[2]

[1]National Taiwan University of Science and Technology, [2]Manchester Metropolitan University

ABSTRACT

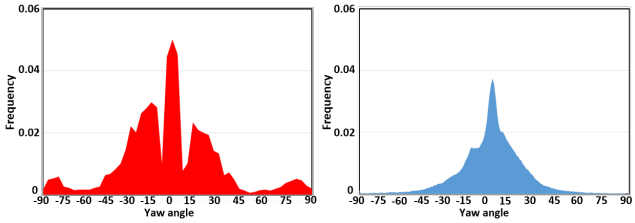
We address two issues for tackling face recognition across pose, one is disentangled representation learning and the other is training data augmentation. To have better training properties, we propose the Representation-Learning Wasserstein-GAN (RL-WGAN) with three component networks for learning the disentangled facial representation. As the learning based on imbalanced data often leads to biased estimation, we proposed a data augmentation scheme that exploits the 3D Morphable Model (3DMM) for generating faces of desired poses. The RL-WGAN and the data augmentation are verified in the experiments with benchmark databases, and compared with contemporary approaches for performance evaluation.

1. INTRODUCTION

Two issues are addressed in this paper, one is disentangled representation learning and the other is data augmentation for tackling imbalanced training data. Both issues are addressed for tackling face recognition across pose. We revise and restructure a state-of-the-art framework for better learning of disentangled facial representation, and propose an effective approach for handling imbalanced data. Our proposed framework and approach are verified by running experiments on benchmark databases.

In general, there are three categories of approaches to tackle cross-pose recognition. One category aims to rotate a non-frontal face back to the frontal view for better extracting facial features [1, 2, 3]. Another category aims to learn the pose-invariant features directly from non-frontal faces [4, 5, 6]. The third category aims to search for disentangled facial representation so that the identity-preserving features can be disentangled from pose, illumination and other parameters for better handling cross-pose recognition [7]. Our framework belongs to the third category.

The DR-GAN (Disentangled Representation learning-GAN) [7] learns a generative and discriminative facial representation which disentangles the face identity from pose. The DR-GAN is built on the common two-player GAN setup, but its generator explores an encoder-decoder structure, leading to the desired disentanglement. The encoder's input is a face image of any pose, the decoder's output is a synthetic face at a target pose, and the learned facial representation connects the encoder and decoder. The discriminator follows the same discriminator design in the Categorical Generative Adversarial (CGA) network [8] which is trained to not only distinguish synthetic (fake) images from real images, but also predict the identity and pose of a face.



(a) IJB-A (b) CASIA WebFace

Fig. 1. Data distribution across the angle in yaw, (a) IJB-A, (b) CASIA-WebFace.

Although the DR-GAN demonstrates a good performance for cross-pose recognition, our experiments show the following issues when training the DR-GAN: 1) The training is hard to converge; 2) In the few converged cases, the mode collapse is often observed; 3) For the cases without mode collapse, the convergence of different tasks appears different one and another, i.e., the training for one task may take longer than for another task. In addition, we also address the tackling of imbalanced training data. It is commonly known that learning based on imbalanced data would result in biased estimation. As shown in the latest work by Masi et al. [9], several common databases all present imbalanced pose distribution. Two examples, the IARPA Janus Benchmark A (IJB-A) [10] and the CASIA WebFace [11], are illustrated in Fig. 1. Most of the faces are within 40° in yaw, and this unbalanced pose distribution can be one of the major causes for the poor performance when recognizing profile faces.

In the following, we first present the Representation-Learning Wasserstein-GAN (RL-WGAN) proposed for tackling the aforementioned issues and the data augmentation scheme for handling the imbalanced data in Sec.2. The RL-WGAN and the data augmentation scheme are verified in the experiments reported in Sec.3, followed by a conclusion

Thanks to the Ministry of Science and Technology (MOST) for funding.

Authorized licensed use limited to: CHONNAM NATIONAL UNIVERSITY. Downloaded on July 26,2021 at 06:21:35 UTC from IEEE Xplore. Restrictions apply.

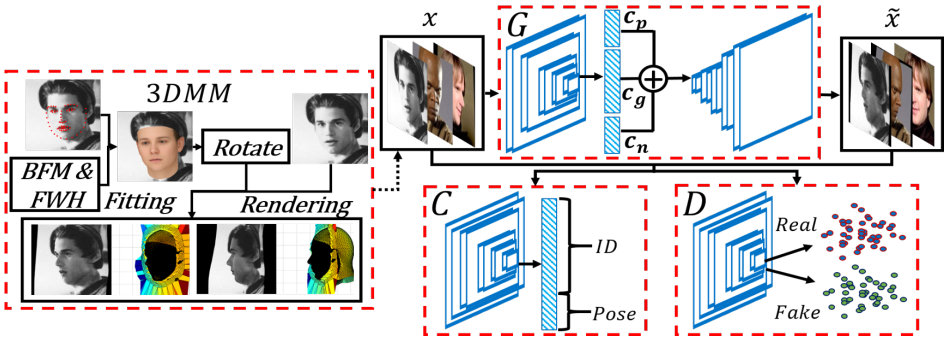


Fig. 2. The proposed framework composed of the RL-WGAN for disentangled representation learning and the 3DMM-based face profiling for data augmentation

given in Sec.4.

2. PROPOSED APPROACH

The proposed framework is illustrated in Fig. 2, which includes the Representation-Learning Wasserstein-GAN (RL-WGAN) and the training data augmentation module. The core part of the training data augmentation module is a 2D-to-3D face mapping approach. The RL-WGAN is revised from the DR-GAN with the following restructuring and amendments: 1) a separate identity/pose classifier from the discriminator which, in the DR-GAN, shares the same convolution layers with the discriminator; 2) a discriminator built upon the Wasserstein loss for better training properties; and 3) the batch norm in the DR-GAN replaced by the group norm [12] for better feature extraction across the convolution layers. The details are presented below.

2.1. Representation-Learning Wasserstein GAN

The configuration of the proposed Representation-Learning Wasserstein GAN (RL-WGAN) is composed of a generator G, a discriminator D and a classifier C, and all are built on the same structure of a base network. We choose the modified CASIA Net [11], same as that used in the DR-GAN, as the base network, denoted by N_0. The modified CASIA Net N_0 is developed on a relatively simple architecture but offers a comparable performance to the DeepFace [13] and DeepID2 [14] for face recognition. It consists of 5 convolution blocks, including 1 double-convolution block and 4 triple-convolution blocks, followed by an average pooling (AvePool) layer for feature code extraction. The extracted feature code c_d is processed differently in G and D. In D, the AvePool layer is connected to a fully connected (FC) layer for handling classification. Note that the G and D in the DR-GAN all use the batch normalization to stabilize training. The generator G is composed of an encoder G_e and a decoder G_d, i.e., $G = [G_e, G_d]$. We follow the design for making G in the DR-GAN. Given a face image x, the encoder's output code $c_d = G_e(x) \in R^{N_c}$ from the AvePool layer is concatenated with a pose code $c_p \in R^{N_p}$ and a noise $z \in R^{N_n}$ to form $[c_d, c_p, z]$, which is used as the input to G_d. G_d is a deconvolution neural network that transforms $[c_d, c_p, z]$ to a decoded face image, i.e., $\hat{x} = G_d([c_d, c_p, z])$. G aims to make the classifier C classify $\hat{x}$ as the same identity as x but in the desired pose c_p, and to fool the discriminator D into determining $\hat{x}$ to be real. Therefore, in the RL-WGAN, the loss L_g considered when training G is evaluated via the losses considered in the discriminator D and the classifier C.

In the DR-GAN, the input to the multi-task discriminator D is either a real or generated face image, and the output is composed of three multi-task parts denoted as $[d_d, d_p, d_r]$. $d_d \in R^{N_d}$ is the output of the softmax function for classifying N_d subjects in the training set, $d_p \in R^{N_p}$ is the output of the softmax function for classifying N_p poses, and d_r is a 1D softmax output for fake/real image classification. Experiments show that the training of G and D with the above settings is unstable and hard to converge.

We explore the Wasserstein Generative Adversarial Network (WGAN) proposed by Arjovsky et al. [15] to handle the training stability. The discriminator in the WGAN considers a cost function based on the Wasserstein-1 distance between the data distribution p_d and the model distribution p_g, denoted as $W(p_d, p_g)$, converting the core problem to the cost of transporting the mass of p_g to that of p_d. The following min-max objective is considered in the WGAN:

$$\min_{\theta_G} \max_{D \in \mathbf{D}_L} \mathbf{E}_{x \sim p_d}[D(x)] - \mathbf{E}_{\hat{x} \sim p_g}[D(\hat{x})] \quad (1)$$

where $\mathbf{D}_L$ is the set of 1-Lipschitz functions. When the discriminator D is being optimized, the parameter update involved in $\min_{\theta_G}$ leads to the minimization of $W(p_d, p_g)$, which yields a *critic* function whose gradient behaves better than does the gradient considered in the typical GAN. To meet more requirements, the same team improves the WGAN with a gradient penalty (GP) added in, resulting in the WGAN-GP [15], where a constraint is imposed on the gradient norm of the discriminator's output and the following objective L_d is

Authorized licensed use limited to: CHONNAM NATIONAL UNIVERSITY. Downloaded on July 26,2021 at 06:21:35 UTC from IEEE Xplore. Restrictions apply.

constituted:

$$\min_{\theta_G} \max_{D \in \mathbf{D}_L} \quad \mathbf{E}_{x \sim p_d}[D(x)] - \mathbf{E}_{\hat{x} \sim p_g}[D(\hat{x})] + \lambda \mathbf{E}_{\tilde{x} \sim p_{\tilde{x}}}\left[(||\nabla_{\tilde{x}} D(\tilde{x}) - 1||_2)^2\right] \quad (2)$$

The last term is the penalty on the gradient norm computed at random samples $\tilde{x} \sim p_{\tilde{x}}$. $p_{\tilde{x}}$ is implicitly defined as the distribution of the uniform samples along the straight lines between the pairs of the data sampled from p_d and p_g. To make D, we connect the output of the average pooling layer in N_0 to a scalar output. When training D with a given image x_s, the images are entered in real-fake pairs $(x_s, \hat{x}_s)$, and the interpolated data $\tilde{x}_s = \eta x_s + (1-\eta)\hat{x}_s$ can be determined by choosing $\eta \sim U[0,1]$. Given x_s, $\hat{x}_s$ and $\tilde{x}_s$, the loss $L_d(x_s, \hat{x}_s)$ as shown in (2) can be computed.

In the RL-WGAN, the classifier has two parts $C = [C_d, C_p]$, where $C_d(x_s) \in R^{D_d}$ is for identity classification and $C_p(x_s) \in R^{D_p}$ for pose identification. To make C, we connect the output of the average pooling layer in N_0 to a $(D_d + D_p)$-dimensional fully connected layer with softmax outputs. The loss L_c is the sum of the cross entropy losses from the two parts,

$$\begin{aligned} L_c(x_s) &= L_{c,d}(x_s) + L_{c,p}(x_s) \\ &= \mathbf{E}[\log C_d(x_s)] + \mathbf{E}[\log C_p(x_s)] \end{aligned} \quad (3)$$

Given the objective considered in D and the loss considered in C, the loss considered when training G, L_g, can be written as follows:

$$\begin{aligned} L_g(\hat{x}_s) = &\; L_d(\hat{x}_s) + L_{c,d}(\hat{x}_s) + L_{c,p}(\hat{x}_s) \\ = &\; -\mathbf{E}[D(\hat{x}_s)] + \mathbf{E}[\log C_d(\hat{x}_s)] + \mathbf{E}[\log C_p(\hat{x}_s)] \end{aligned} \quad (4)$$

It is pointed out by Gulrajani et al. [16] that the WGAN-GP does not work with batch normalization (BN), which changes the discriminators processing from mapping a single input to a single output to mapping from a batch of inputs to a batch of outputs. Although the layer normalization (LN) is recommended for the WGAN-GP, we have found that the group normalization [12] performs better. The group normalization (GN) was proposed as a simple alternative to BN. GN divides the input channels into groups and computes the mean and variance within each group for normalization. However, different from BN, the GN computation is independent of the batch size, and the obtained parameters are stable over a wide range of batch sizes. Although the WGAN-GP processes single inputs, there are multiple channels for each single input, making the GN an appropriate choice for normalization. Note that the normalization discussed above is for the discriminator D with the WGAN-GP built in. The normalization for the G and C can still be BN or GN, and we report the performance comparison of both in Sec. 3.

2.2. 3DMM Face Profiling for Data Augmentation

Face profiling refers to making the profile view of a frontal face. The face profiling proposed by Zhu et al. [17] begins with a 3D Morphable Model (3DMM) [18] fitted to the 2D face by following the Multi-Features Fitting (MFF) [19]. MFF can be directly applied to faces labeled with landmarks. To obtain the facial landmarks on a given face, we employ two state-of-the-art algorithms, the Face Alignment Network (FAN) [20] and Multi-Dropout Network (MDN) [21], jointly. When the error between the two algorithms is below a threshold, the landmarks are located at the average of two detected locations. The advantages of this joint referencing localization include the removal of low-quality/undesired images and better localization of landmarks.

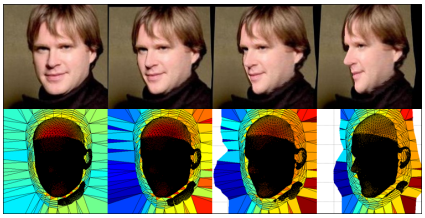


Fig. 3. Face profiling with the original face on the upper left. The top row shows rotated faces and background, bottom row is 3DMM fitted face model and meshed background.

Given a landmark labeled 2D face, the MFF fitting will be appropriately constrained by the landmarks and deliver a well fitted 3D model. To include expression variation to the 3DMM, Zhu et al. [22] combine the identity shape from the Basel Face Model (BFM) [23] and the expression shape from the Face Warehouse [24]. In addition, they propose the *landmark marching* for fitting the 3DMM to a face with pose variation, and the 3D meshing with depth estimated on the background. The 3D facial model and its 3D meshed background form a 3D object that can be rotated to a specific view. We use the code available via the link provided in [22] for the face profiling. Fig.3 shows a sample face with its 3D meshed face and background model rotated in yaw and pitch, and the rendered 2D images. We apply this approach to the faces in the CASIA-WebFace, which is a major part of our training set, to augment the database with the poses needed to make up the missing parts in the distribution in Fig. 1.

3. EXPERIMENTAL EVALUATION

The experiments are designed to highlight the advantages of the following items: 1) the discriminator D_W with Wasserstein distance as the measure for discriminating real/fake images; 2) the RL-WGAN with three component networks; 3) the GN for normalization; and 4) the data augmentation with face profiling. To attain these objectives, we have trained the following networks: 1) DR-WGAN, which is the two-player DR-GAN with D replaced by D_W; 2) RL-WGAN$_{BN}$, which is the RL-WGAN with BN on G and C and LN on D_W; 3) RL-WGAN$_{GN}$, which is the RL-WGAN with GN on all three networks, i.e., G, C and D_W; 4) RL-WGAN$^{ag}_{GN}$, which is the

Authorized licensed use limited to: CHONNAM NATIONAL UNIVERSITY. Downloaded on July 26,2021 at 06:21:35 UTC from IEEE Xplore. Restrictions apply.

Table 1. Performance on MPIE, Avg is the average rate for $0^\circ \sim 60^\circ$ and $(\cdot)$ is the average for $0^\circ \sim 90^\circ$

Method	0°	15°	30°	45°	60°	75°	90°	Avg
Zhu et al.[28]	95.7	92.8	83.7	72.9	60.1	-	-	79.3
Yim et al.[3]	99.5	95	88.5	79.9	61.9	-	-	83.3
DR-GAN[7]	97	94.01	90	86.2	83.2	-	-	89.2
DR-WGAN	99.1	97.0	96.2	93.2	88.2	80.4	70.1	94.3 (88.4)
RL-WGAN$_{BN}$	99.4	97.8	97.1	95.1	90.5	83.8	72.8	95.6 (90.3)
RL-WGAN$_{GN}$	99.4	98.2	97.5	96.7	91.7	86.1	75.1	96.4 (91.6)

RL-WGAN$_{GN}$ trained on the augmented data.

Similar to the experiments reported in [25, 7], we also study the performance of the proposed framework by assessing the capability of handling cross-pose recognition. We select the MPIE [26] and CASIA-WebFace for training, and the CFP (Celebrities in Frontal-Profile) database [27] and IJB-A for testing. MPIE is one of the most popular in-the-house databases with 337 subjects, 13 poses, 6 expressions and 20 illumination conditions. The CASIA-WebFace offers 494,414 faces of 10,575 subjects, with pose distributed as shown in Fig. 1. We use the MDN pose regressor [21] to identify the pose of each face in the database. *Following the proposed data augmentation approach, we have generated 361,782 additional faces with pose* $30^\circ \sim 90^\circ$ *in yaw.* The CFP offers 500 subjects with 10 frontal and 4 profile images per subject. The CFP protocol contains frontal-to-frontal (FF) and frontal-to-profile (FP) verification, each with 350 intra and 350 extra pairs. The IJB-A contains images and videos of 500 subjects captured in the wild. The protocol contains identification (search) and verification (compare).

In the experiments, all faces are aligned to the landmarks and cropped to 100×100. We employ CASIA Net [11] as basic net and the network weights are initialized to $N(0, 0.02)$. We use the Adam optimizer with a fixed learning rate 0.0001 and momentum 0.5. The batch size is set to be 64. All experiments are run on a GTX 1080 Ti GPU with CUDA 8.0 and cuDNN 6.0 on Pytorch.

To have better training, we adopt a two-phase training scheme by training the networks on the in-the-house MPIE first as it offers good quality images with manually labeled pose groups, and then on the in-the-wild CASIA WebFace. In Phase-1, we split the 337 subjects in MPIE into 188 for training and 149 for validation (this split is the same as that in [29]). In Phase-2, we retrain the validated networks by the 149 subjects from MPIE and the CASIA WebFace.

Table 1 shows the Phase-1 validation on MPIE and a comparison to other latest approaches. It can be seen that the RL-WGAN$_{GN}$ outperforms all, followed by RL-WGAN$_{BN}$, and then DR-WGAN, which performs better than DR-GAN and other selected methods. This shows the following observations: 1) GN appears better than BN for normalization; 2) the 3-player framework RL-WGAN works better than the 2-player DR-WGAN, and 3) the D_W (discriminator with Wasserstein distance) does not just smooth the training, but also improve the performance. Our experiments show that D_W makes the training way more stable than training the network with regular D. Table 2 shows the performance evaluation on CFP. The performance improvement made by D_W is not apparent as that shown in Table 1, revealing the difficulty between the in-the-house and in-the-wild face recognition. However, the aforementioned three observations can still be considered valid. Additionally, the RL-WGAN$^{ag}_{GN}$ performs better than RL-WGAN$_{GN}$, indicating that the proposed data augmentation helps to tackle data imbalance and further improve the performance. As Tables 1 and 2 show that both RL-WGAN$^{ag}_{GN}$ and RL-WGAN$_{GN}$ perform better than RL-WGAN$_{BN}$ and DR-WGAN, we only present RL-WGAN$^{ag}_{GN}$ and RL-WGAN$_{GN}$ in Table 3 for the performance on IJB-A. It shows performance drop slightly when the loss breaks out 1-Lipschitz constraint since the samples are more complexity in unconstrained condition. However, it shows that the data augmented RL-WGAN$^{ag}_{GN}$ outperforms RL-WGAN$_{GN}$, which outperforms other state-of-the-art approaches.

Table 2. Performance on CFP

Method	Frontal-Frontal	Frontal-Profile
Sengupta et al.[27]	96.40±0.69	84.91±1.82
Sankarana et al.[30]	96.93±0.61	89.17±2.35
Chen et al.[31]	98.67±0.36	91.97±1.70
DR-GAN[7]	97.13±0.62	90.82±0.28
DR-WGAN	97.56±1.04	91.07±1.23
RL-WGAN$_{BN}$	98.29±1.07	91.39±1.65
RL-WGAN$_{GN}$	98.67±0.82	91.78±1.72
RL-WGAN$^{ag}_{GN}$	98.67±1.01	92.74±1.85

Table 3. Performance comparison on IJB-A

	Verification		Identitfication	
Method	@FAR=.01	@FAR=.001	@Rank-1	@Rank-5
Wang et al.[32]	72.9±3.5	51.0±6.1	82.2±2.3	93.1±1.4
PAM [5]	73.3±1.8	55.2±3.2	77.1±1.6	88.7±0.9
DR-GAN [7]	77.4±2.7	53.9±4.3	85.5±1.5	94.7±1.1
RL-WGAN$_{GN}$	79.1±2.14	55.7±3.3	87.8±1.5	94.5±1.1
RL-WGAN$^{ag}_{GN}$	81.6±2.14	57.7±2.7	89.3±0.9	96.5±1.2

4. CONCLUSION

We propose the 3-player Representation-Learning WGAN (RL-WGAN) for better learning of disentangled facial representation for cross-pose recognition, and a data augmentation scheme for tackling data imbalance. The proposed framework incorporates the Wasserstein discriminator and the group normalization for improving representation learning. Experiments show that the framework is competitive to state-of-the-art approaches.

Authorized licensed use limited to: CHONNAM NATIONAL UNIVERSITY. Downloaded on July 26,2021 at 06:21:35 UTC from IEEE Xplore. Restrictions apply.

5. REFERENCES

[1] Tal Hassner, Shai Harel, Eran Paz, and Roee Enbar, "Effective face frontalization in unconstrained images," *CVPR*, 2015.

[2] Christos Sagonas, Yannis Panagakis, Stefanos Zafeiriou, and Maja Pantic, "Robust statistical face frontalization," *ICCV*, 2015.

[3] Junho Yim, Heechul Jung, ByungIn Yoo, Changkyu Choi, Dusik Park, and Junmo Kim, "Rotating your face using multi-task deep neural network," *CVPR*, 2015.

[4] Jui-Shan Chan, Gee-Sern Jison Hsu, Hung-Cheng Shie, and Yan-Xiang Chen, "Face recognition by facial attribute assisted network," *ICIP*, 2017.

[5] Iacopo Masi, Stephen Rawls, Gérard Medioni, and Prem Natarajan, "Pose-aware face recognition in the wild," *CVPR*, 2016.

[6] Florian Schroff, Dmitry Kalenichenko, and James Philbin, "Facenet: A unified embedding for face recognition and clustering," *CVPR*, 2015.

[7] Luan Tran, Xi Yin, and Xiaoming Liu, "Disentangled representation learning gan for pose-invariant face recognition," *CVPR*, 2017.

[8] Jost Tobias Springenberg, "Unsupervised and semi-supervised learning with categorical generative adversarial networks," *arXiv*, 2015.

[9] Iacopo Masi, Feng-Ju Chang, Jongmoo Choi, Shai Harel, Jungyeon Kim, KangGeon Kim, Jatuporn Leksut, Stephen Rawls, Yue Wu, Tal Hassner, et al., "Learning pose-aware models for pose-invariant face recognition in the wild," *TPAMI*, 2019.

[10] Brendan F Klare, Ben Klein, Emma Taborsky, Austin Blanton, Jordan Cheney, Kristen Allen, Patrick Grother, Alan Mah, and Anil K Jain, "Pushing the frontiers of unconstrained face detection and recognition: Iarpa janus benchmark a," *CVPR*, 2015.

[11] Dong Yi, Zhen Lei, Shengcai Liao, and Stan Z Li, "Learning face representation from scratch," *arXiv*, 2014.

[12] Yuxin Wu and Kaiming He, "Group normalization," *arXiv*, 2018.

[13] Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf, "Deepface: Closing the gap to human-level performance in face verification," *CVPR*, 2014.

[14] Yi Sun, Yuheng Chen, Xiaogang Wang, and Xiaoou Tang, "Deep learning face representation by joint identification-verification," *NIPS*, 2014.

[15] Martin Arjovsky, Soumith Chintala, and Léon Bottou, "Wasserstein gan," *arXiv*, 2017.

[16] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville, "Improved training of wasserstein gans," *NIPS*, 2017.

[17] Xiangyu Zhu, Zhen Lei, Xiaoming Liu, Hailin Shi, and Stan Z Li, "Face alignment across large poses: A 3d solution," *CVPR*, 2016.

[18] Volker Blanz and Thomas Vetter, "Face recognition based on fitting a 3D morphable model," *TPAMI*, 2003.

[19] Sami Romdhani and Thomas Vetter, "Estimating 3d shape and texture using pixel intensity, edges, specular highlights, texture constraints and a prior," *CVPR*, 2005.

[20] Adrian Bulat and Georgios Tzimiropoulos, "How far are we from solving the 2d & 3d face alignment problem?(and a dataset of 230,000 3d facial landmarks)," *ICCV*, 2017.

[21] Gee-Sern Hsu and Cheng-Hua Hsieh, "Cross-pose landmark localization using multi-dropout framework," *IJCB*, 2017.

[22] Xiangyu Zhu, Zhen Lei, Junjie Yan, Dong Yi, and Stan Z Li, "High-fidelity pose and expression normalization for face recognition in the wild," *CVPR*, 2015.

[23] Pascal Paysan, Reinhard Knothe, Brian Amberg, Sami Romdhani, and Thomas Vetter, "A 3d face model for pose and illumination invariant face recognition," *AVSS*, 2009.

[24] Chen Cao, Yanlin Weng, Shun Zhou, Yiying Tong, and Kun Zhou, "Facewarehouse: A 3d facial expression database for visual computing," *TVCG*, 2014.

[25] Xi Peng, Xiang Yu, Kihyuk Sohn, Dimitris N Metaxas, and Manmohan Chandraker, "Reconstruction-based disentanglement for pose-invariant face recognition," *ICCV*, 2017.

[26] Ralph Gross, Iain Matthews, Jeffrey Cohn, Takeo Kanade, and Simon Baker, "Multi-pie," *IVC*, 2010.

[27] Soumyadip Sengupta, Jun-Cheng Chen, Carlos Castillo, Vishal M Patel, Rama Chellappa, and David W Jacobs, "Frontal to profile face verification in the wild," *WACV*, 2016.

[28] Zhenyao Zhu, Ping Luo, Xiaogang Wang, and Xiaoou Tang, "Multi-view perceptron: a deep model for learning face identity and view representations," *NIPS*, 2014.

[29] Annan Li, Shiguang Shan, and Wen Gao, "Coupled bias–variance tradeoff for cross-pose face recognition," *IEEE Transactions on Image Processing*, vol. 21, no. 1, pp. 305–315, 2011.

[30] Swami Sankaranarayanan, Azadeh Alavi, Carlos Castillo, and Rama Chellappa, "Triplet probabilistic embedding for face verification and clustering," *arXiv*, 2016.

[31] Jun-Cheng Chen, Jingxiao Zheng, Vishal M Patel, and Rama Chellappa, "Fisher vector encoded deep convolutional features for unconstrained face verification," *ICIP*, 2016.

[32] Dayong Wang, Charles Otto, and Anil K Jain, "Face search at scale.," *TPAMI*, 2017.

Authorized licensed use limited to: CHONNAM NATIONAL UNIVERSITY. Downloaded on July 26,2021 at 06:21:35 UTC from IEEE Xplore. Restrictions apply.