

De-mark GAN: Removing Dense Watermark With Generative Adversarial Network

Jinlin Wu, Hailin Shi*, Shu Zhang, Zhen Lei, Yang Yang, Stan Z. Li

Center for Biometrics and Security Research & National Laboratory of Pattern Recognition

Institute of Automation, Chinese Academy of Sciences

University of Chinese Academy of Sciences

jinlin.wu@cbsr.ia.ac.cn, {hailin.shi, shu.zhang, zlei, yang.yang, szli}@nlpr.ia.ac.cn

Abstract

This paper mainly considers the MeshFace verification problem with dense watermarks. A dense watermark often covers the crucial parts of face photo, thus degenerating the performance in the existing face verification system. The key to solving it is to preserve the ID information while removing the dense watermark. In this paper, we propose an improved GAN model, named De-mark GAN, for MeshFace verification. It consists of one generator and one global-internal discriminator. The generator is an encoder-decoder architecture with a pixel reconstruction loss and a feature loss. It maps a MeshFace photo to a representation vector, and then decodes the vector to a RGB ID photo. The succedent global-internal discriminator integrates a global discriminator and an internal discriminator with a global loss and internal loss, respectively. It can ensure the generated image quality and preserve the ID information of recovered ID photos. Experimental results show that the verification benefits well from the recovered ID photos with high quality and our proposed De-mark GAN can achieve a competitive result in both image quality and verification.

1. Introduction

In recent years, benefitting from the deep learning[2], e.g., ConvNet[9], face recognition has made a great progress[5]. On the LFW benchmark[7], ConvNet continues to create new records, and even a high verification accuracy beyond human[12]. The face verification between ID photos and daily life photos (FVBID)[20] is a specific task in face recognition. FVBID is widely used in clearance at airport, opening bank account from remote, etc.

However, in order to protect the user's privacy, ID photos are usually corrupted by random watermarks. Corrupted

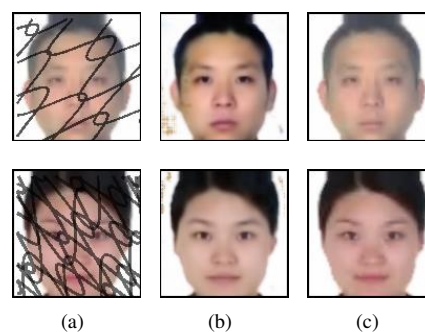


Figure 1. Samples in the MeshFace dataset. From left to right: (a) ID photo corrupted by watermark, (b) ID photo recovered by De-mark GAN, (c) original ID photo.

ID photos are called MeshFace [19]. Watermarks cover some parts of the face and produce a large disturbance in the pixel space, shown in the top row of Fig. 1. When the lines of the watermark become dense (e.g., thicker, darker), many crucial parts of face are covered, shown in the bottom row of Fig. 1. Due to this, the verification performance of the ConvNet is severely reduced by the dense Watermarks. MeshFace verification is becoming a challenging problem in FVBID. The key to solving MeshFace verification is to reduce the loss of the ID information, while recovering high quality ID photos.

Some efforts have been made to solve MeshFace verification. Zhang *et al.* [1] treat this problem as a blind face inpainting problem, proposing a multi-task SRCNN [2, 4] to handle the MeshFace verification. This method preserves ID information of MeshFace by detecting the corrupted regions and recovering the corresponding part. However, the verification performance of ID photos recovered by SRCNN is sensitive to pixel perturbation. To avoid this, Zhang *et al.* [19] propose the DeMeshNet by using a feature loss to train the Deep FCN (Fully Convolutional Net-

*Corresponding author.

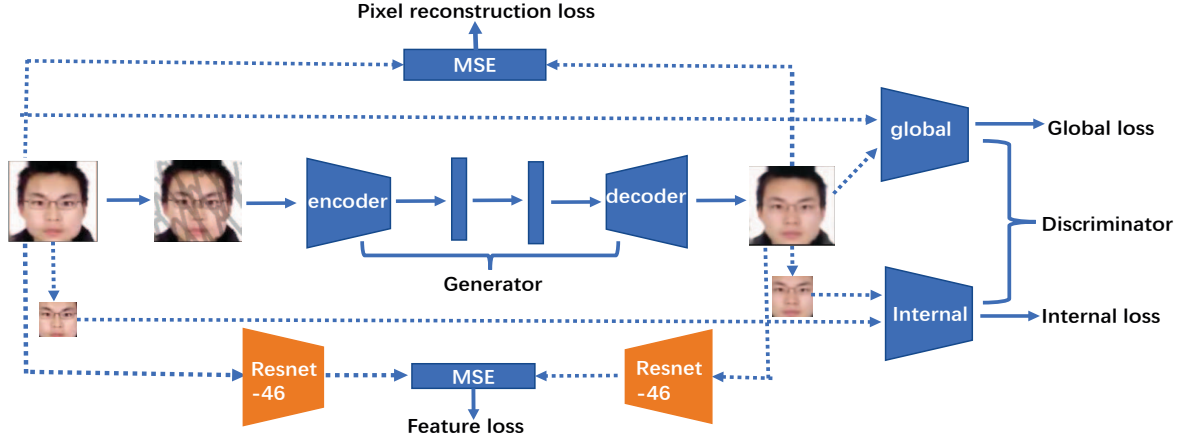


Figure 2. Architecture of De-mark GAN. It consists of one generator and one global-internal discriminator. A pre-trained Resnet-46 is used to compute feature loss. In this scheme, the whole face image includes the whole area of face image. The internal face image is a fixed area (70×70) in the center of face image.

works) model [14]. DeMeshNet recovers a gray ID photo from MeshFace with a high verification accuracy. ID photos recovered by the DeMeshNet are robust to perturbation in pixel space. For these methods, FCN loses some ID information of MeshFace and can't achieve a high verification accuracy. Especially for the dense watermark, the quality of recovered images becomes worse. In recent years, GAN [6] is proposed as a generative model which can generate samples similar to the training data, through an adversarial training between a generator (G) and a discriminator (D). Many existing works indicate that GAN has a powerful ability for image recovering and completing [13, 17].

In this paper, we regard MeshFace recovering as an image generating problem. A GAN model is proposed to solve it, named De-mark GAN. The structure of De-mark GAN is shown in Fig. 2. The main objectives of our De-mark GAN are improving the quality and verification performance of recovered ID photos. The generator of De-mark GAN is an auto-encoder structure, which maps MeshFace to a representation vector, and then decodes the vector to a RGB ID photo. The discriminator is a global-internal structure, which consists of a global and an internal discriminator. Different from the conventional GAN, De-mark GAN can control the content of the generated image owing to the auto-encoder structure generator. The global-internal structure discriminator of De-mark GAN is more effective than discriminator of conventional GAN in improving the generated images quality. De-mark GAN is trained with a combination of a pixel reconstruction loss, global and internal adversarial losses and a feature loss. Pixel reconstruction loss is the distance between recovered ID photos and ground truth ID photos in pixel level. Global and internal adversarial loss are the loss of GAN, ensuring recovered ID photos are realistic enough. We use a Resnet-46 to extract features

of recovered ID photos and ground truth photos. The feature loss is the distance between ID photos and ground truth photos in feature space. This measure improves the verification performance and avoids the pixel perturbation.

2. Related Work

2.1. Image recovering

Image recovering refers to the process of reconstructing a clear image from the corrupted image. Some contemporary works [3] are proposed to solve it. These works have shown that ConvNet has powerful abilities in recovering problem [3]. Zhang *et al.* [18] propose a multi-task SRCNN [4] to recover clear ID photos from MeshFace. They regard MeshFace recovering as a blind face inpainting problem since the position of corruptions face part is unknown during test. Zhang *et al.* [19] propose the De-MehNet model, which solves MeshFace recovering in feature space. DeMeshNet recovers clear ID photos with a high verification accuracy.

2.2. GAN

GAN is a generative model proposed by Goodfellow *et al.* [6]. With a minimax two-player game, the generator can generate samples similar to the training data. GAN is widely used in images synthesis, style transfer, image recovering and completion. More recent works focus on using images to synthesis images [11]. Tran *et al.* [16] propose a DR-GAN (Disentangled Representation Learning GAN) which can convert profile picture to front picture, improving the performance pose-invariant face recognition (PIFR). Li *et al.* [13] use global and internal adversarial loss to train GAN, generating the corrupted face part.

3. Method

3.1. Generator

As illustrated in Fig. 2, the generator G is an auto-encoder structure. The input of G is the MeshFace image. The encoder of G encodes the MeshFace image to a hidden representation vector, and the decoder of G generates a clear ID photo with the hidden representation vector. After the process of encoding and decoding, the generator preserves the ID information and removes watermark.

The details of the generator network is shown in the supplementary material. We adopt the residual blocks [9] to build G . The encoder consists of 6 DownResidual blocks (the structure of DownResidual block is also illustrated in supplementary material), encoding the 120×120 RGB MeshFace image to 512-d hidden representation vector. This vector is then mapped to a second vector through a FC (fully-connected) layer. The decoder consists of 5 UpResidual blocks, generating a 120×120 RGB non-watermark ID photo from the second hidden representation vector.

3.2. Discriminator

Mere using of pixel reconstruction loss will lead to a small error to the average face. Especially for the dense watermark which covers most part of the face, the generator generates a blurry and smooth average face, losing the ID information heavily. To preserve more ID information and improve the quality of recovered ID photos, we apply a global-internal structure in the discriminator. The global-internal discriminator contains two discriminators, the global discriminator discriminates the faithfulness of the entire image recovered by generator. (The details of the discriminator network is shown in the supplementary material.) The internal discriminator discriminates the realistic of the internal part of ID photos, including eyes, mouth and nose. The internal discriminator enforces the generator recovering more details in the internal part.

The structure of the global discriminator has 5 Residual blocks and 5 DownResidual blocks (The structures of Residual block and DownResidual block are illustrated supplementary material). A fully-connected layer is added after the residual blocks as a classifier, to discriminate the input is real or fake. The internal discriminator adopts the structure of 4 Residual blocks and 4 DownResidual blocks. It is also added by a fully-connected layer after the residual blocks as the classifier.

3.3. Feature Loss

ConvNet achieves a high accuracy in image verification. However, the ConvNet can be fooled by adding a tiny amount of noise to original images [15]. Therefore, we adopt a feature loss as an additional constraint in the training of the generator. The feature loss reduces the distance

between recovered ID photos and ground truth ID photos in the feature space, improving the verification performance on the recovered ID photos. We use a pre-trained Resnet-46 to extract features. The Resnet-46 model is pre-trained on MS-Celeb-1M dataset [8], achieving a 99.3% verification accuracy on the LFW [10] benchmark. The pre-trained Resnet-46 is denoted as $\phi(\cdot)$. The x is MeshFace. $G(x)$ is the ID photo recovered by the generator G . The y is the ground truth ID photo. The feature loss is defined as the Euclidean distance between the features $\phi(G(x))$ and $\phi(y)$:

$$L_f = \|\phi(G(x)) - \phi(y)\|_2 \quad (1)$$

3.4. Objective Functions

We adopt a pixel reconstruction loss L_{pixel} to encourage generator to generate smooth ID photos with basic face outline. The pixel reconstruction loss is the L2 distance between the recovered ID photo $G(x)$ and the ground truth ID photo y . The pixel reconstruction can be defined as:

$$L_{pixel} = \|G(x) - y\|_2 \quad (2)$$

To improve the realistic level and the quality of recovered ID photos, we employ the adversarial loss to encourage more realistic images to fool the discriminator. It can be defined as:

$$L_a = \min_G \max_D \mathbb{E}_{y \sim p_{ground_truth}(y)} [\log D(y)] + \mathbb{E}_{x \sim p_{MeshFace}(x)} [\log(1 - D(G(x)))] \quad (3)$$

In the above formula, the input x is either the global face image or the internal part of face. When x is global, L_a is denoted as L_{global} in Eq. (4). When x is internal, L_a is denoted as $L_{internal}$. Different from the conventional GAN, De-mark GAN generates images by decoding the representation vector, instead of randomly sampling from the Normal distribution. $p_{ground_truth}(y)$ represents the distribution of ground truth ID photos. We apply the feature loss L_f in Eq. (1) to make the recovered photos favorable to the following face verification. We adopt three weights λ_1 , λ_2 and λ_3 to balance the effects of different losses. The entire objective function is defined as follow:

$$Loss = L_{pixel} + \lambda_1 L_{global} + \lambda_2 L_{internal} + \lambda_3 L_f \quad (4)$$

4. Experiment

4.1. Datasets

In the experiment, we adopt five models to recover ID photos from MeshFace and compare their verification performance and quality of recovered ID photos. These five models are trained on the MeshFace dataset, which contains

Table 1. The De-mark GAN and the competitors in our experiments.

Model	Generator	Discriminator	Loss
FCN	FCN	None	pixel reconstruction loss
DeMeshNet	auto-encoder	None	pixel reconstruction loss, feature loss
Pixel GAN	auto-encoder	a global discriminator	pixel reconstruction loss, global adversarial loss
Pixel De-mark GAN	auto-encoder	a global discriminator and an internal discriminator	pixel reconstruction loss, global adversarial loss, internal adversarial loss
De-mark GAN	auto-encoder	a global discriminator and an internal discriminator	pixel reconstruction loss, global adversarial loss, internal adversarial loss, feature loss

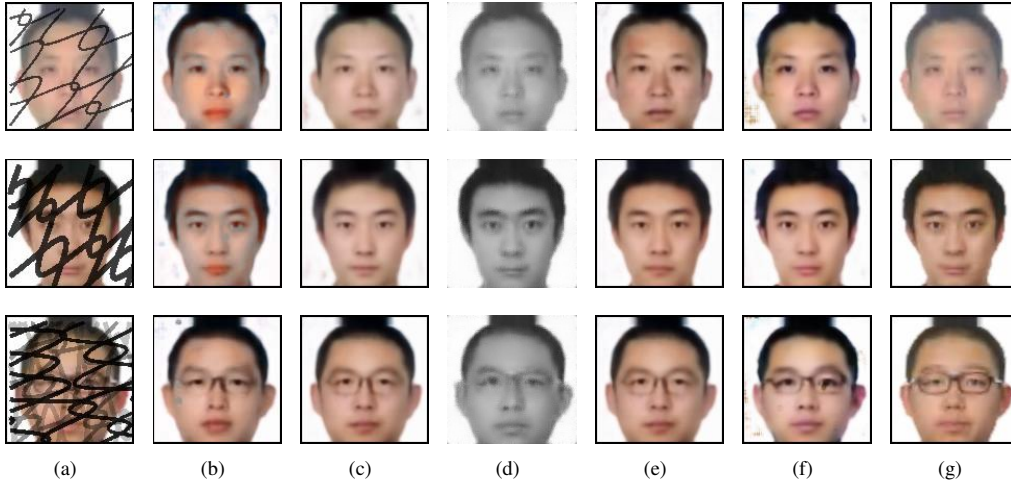


Figure 3. From left to right: (a) MeshFace(120×120), (b) FCN, (c) Pixel GAN, (d) DeMeshNet, (e) Pixel De-mark GAN, (f) De-mark GAN, (g) ground truth ID photos. From the top row to the bottom row, the watermark becomes dense. We show more examples in the supplementary material. Best viewed in color.

more than 80,000 subjects. Each subject has one ID photo and one daily photo. The ID photos are used for the process of watermark and de-watermark. The spot photos are used to perform the face verification.

We randomly sample 10,000 subjects for validation and 10,000 for test. The remaining 60,000 subjects are used as training set. We evaluate these models in term of recovery quality and verification accuracy. For evaluating the quality of recovered ID photos, we choose the value of PSNR (peak signal-to-noise ratio) and SSIM (structural similarity index) to which can directly measure the difference in pixel values and reflect the quality of the recovery ID photo. The high PSNR and SSIM values generally points to a high quality of the recover ID photo. For the verification, we adopt the face comparison protocol of DeMeshNet [19]. Face comparison is conducted with cosine similarity in feature space between all the ID-daily pair. In the experiments, the inputs of discriminators are the whole face image and the inter-

nal face image. The whole face image includes the whole area of face image. The internal face image is a fixed area (70×70) in the center of face image. This area is estimated by experience. It can cover almost all the facial features.

4.2. Implementation Details

We propose a deep FCN and four GANs in the experiments. The structures of each network are shown in the supplementary material. The losses of each model are shown in Table 1. For balancing the effects of different losses in Eq. 4, $\lambda_1 = 0.1$, $\lambda_2 = 1$ and $\lambda_3 = 0.1$ are determined on the validation set. In the training process, we set the learning rate to 0.00005 and use RMSprop as the optimization method. The input image size is 120×120 with RGB channels. The batch size is 64. The De-mark GAN training takes approximately 240k iterations to converge. All the experiments are implemented with the Pytorch framework on 3

Table 2. Testing accuracy of verification on the test set of MeshFace.

Method	TPR@FPR =1%	TPR@FPR =0.1%	TPR@FPR =0.01%	TPR@FPR =0.001%	PSNR	SSIM
corrupted	20.71%	1.42%	0.62%	0.022%	14.33	0.288
FCN	84.22%	59.75%	33.58%	14.56%	23.54	0.883
DeMeshNet	95.02%	85.62%	71.74%	49.75%	21.79	0.904
pixel GAN	75.12%	45.08%	20.76%	7.46%	23.58	0.351
pixel De-mark GAN	92.95%	79.34%	58.65%	33.78%	23.77	0.885
De-mark GAN	96.36%	87.86%	75.12%	49.45%	23.37	0.884

TitanX. The code is realised on the github*.

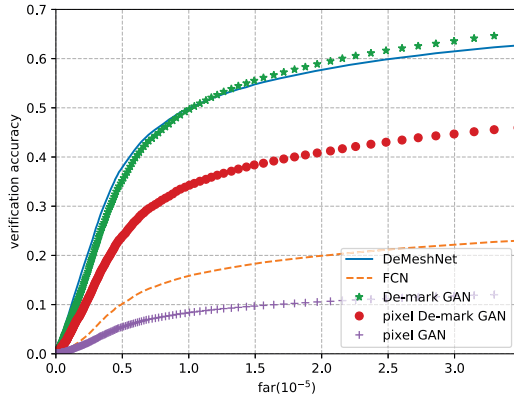


Figure 4. ROC curves on the test set of MeshFace

4.3. Qualitative Evaluation of Recovered Results

Some recovered results are shown in Fig. 3. From top to bottom, the watermark changes from sparse to dense. Column (b) contains the results of FCN, which are similar to each other. The reason is that minimizing the pixel reconstruction loss can't preserve the ID information of the MeshFace. Column (c) contains the results of pixel GAN. The recovered images are slightly better than those of FCN. This indicates that the discriminator improves the recovered photos with more ID-specific details. Column (d), (e), (f) are the results of DeMeshNet, pixel De-mark GAN, and De-mark GAN. The recovered images are much better than those of FCN and pixel GAN. For example, in the last row, the photo belongs to a man wearing a black glasses. From column (b) to column (f), the black glasses become more and more clear. It reveals that discriminator and the feature loss can enhance the details of recovered images. De-mark GAN can preserve colors and more details of MeshFace. More recovered results are shown in the supplementary material. From these recovered images, we can see that our De-mark GAN is an effective method for removing watermark, especially for the dense watermark.

*https://github.com/yichuan9527/demark_gan

4.4. Quantitative Evaluation of Recovered Results

In this section, we quantitatively evaluate the recovered ID photo in term of the values of PSNR, SSIM and verification accuracy. PSNR and SSIM values are shown in Table 2. The pixel De-mark GAN model gives the highest PSNR value and high SSIM value. The internal discriminator is effective in improving the image quality (pixel De-mark GAN outperforms pixel GAN owing to the internal discriminator). The feature loss causes more noise in recovering, while enforcing the generator pay more attention in details. Due to this, De-mark GAN is worse than pixel GAN and pixel De-mark GAN in recovering high quality images, but better in verification performance. The accuracy of verification is also shown in the Table 2. TPR@FPR represent the true positive rates, when false positive is 1%, 0.1%, 0.01% and 0.001%, respectively. The verification on the raw MeshFace is failed. The accuracy suffers a severe drop due to the face covered by dense watermarks. After processed by each of the five models, the verification performance on the recovered ID photos is much better than MeshFace. The ROC curves are shown in Fig. 4. FCN is our baseline model. The image quality of pixel GAN is better than FCN. However, the pixel GAN is worse than FCN in verification. This reveals that the global discriminator causes the degeneration of ID information, regardless of its improvements in PSNR. Compared with pixel GAN, pixel De-mark GAN benefits from the internal discriminator which reduces the pixel perturbations. Therefore, the verification accuracy increases greatly. On the basis of pixel De-mark GAN, De-mark GAN adopts the feature loss in training process. De-mark GAN performs better than pixel De-mark GAN in verification. As shown in Fig. 3, the results of De-mark GAN has rich details and clear edges. This improves the verification of De-mark GAN by a large gap than the others. Compared with the DeMeshNet, De-mark GAN gives the higher PSNR value and higher verification accuracy at the FAR points 1%, 0.1% and 0.01%.

5. Conclusion

In this paper, we propose an effective generative adversarial network, named De-mark GAN, for the MeshFace verification problem with dense watermarks. Under the

supervision of the global-internal discriminator, De-mark GAN can recover ID photos with a high quality. Combined with the generator, the recovered image of De-mark GAN achieves a high verification accuracy. In the experiments, we compare the recovered image quality and verification performance of each model in the test set of MeshFace. The pixel De-mark GAN achieves the best result in image quality. Our De-mark GAN achieves the best verification accuracy at the FAR points 1%, 0.1% and 0.01%, and competitive results at other points.

Acknowledgement

This work was supported by the National Key Research and Development Plan (Grant No.2016YFC0801003), the Chinese National Natural Science Foundation Projects #61572536, #61473291, #61572501, #61502491, Science and Technology Development Fund of Macau (No.112/2014/A3, 151/2017/A) and AuthenMetric R&D Funds.

References

- [1] V. Badrinarayanan, A. Handa, and R. Cipolla. Segnet: A deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling. *arXiv preprint arXiv:1505.07293*, 2015. 1
- [2] X. P. Burgos-Artizzu, P. Perona, and P. Dollár. Robust face landmark estimation under occlusion. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1513–1520, 2013. 1
- [3] A. Criminisi, P. Pérez, and K. Toyama. Region filling and object removal by exemplar-based image inpainting. *IEEE Transactions on image processing*, 13(9):1200–1212, 2004. 2
- [4] C. Dong, C. C. Loy, K. He, and X. Tang. Learning a deep convolutional network for image super-resolution. In *European Conference on Computer Vision*, pages 184–199. Springer, 2014. 1, 2
- [5] D. Eigen, C. Puhrsch, and R. Fergus. Depth map prediction from a single image using a multi-scale deep network. In *Advances in neural information processing systems*, pages 2366–2374, 2014. 1
- [6] 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. 2
- [7] I. J. Goodfellow, J. Shlens, and C. Szegedy. Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572*, 2014. 1
- [8] Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao. Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In *European Conference on Computer Vision*, pages 87–102. Springer, 2016. 3
- [9] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 1, 3
- [10] 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, Technical Report 07-49, University of Massachusetts, Amherst, 2007. 3
- [11] 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. *arXiv preprint arXiv:1704.04086*, 2017. 2
- [12] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the 22nd ACM international conference on Multimedia*, pages 675–678. ACM, 2014. 1
- [13] Y. Li, S. Liu, J. Yang, and M.-H. Yang. Generative face completion. *arXiv preprint arXiv:1704.05838*, 2017. 2
- [14] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3431–3440, 2015. 2
- [15] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus. Intriguing properties of neural networks. *arXiv preprint arXiv:1312.6199*, 2013. 3
- [16] L. Tran, X. Yin, and X. Liu. Disentangled representation learning gan for pose-invariant face recognition. In *CVPR*, volume 4, page 7, 2017. 2
- [17] R. Yeh, C. Chen, T. Y. Lim, M. Hasegawa-Johnson, and M. N. Do. Semantic image inpainting with perceptual and contextual losses. *arXiv preprint arXiv:1607.07539*, 2016. 2
- [18] S. Zhang, R. He, Z. Sun, and T. Tan. Multi-task convnet for blind face inpainting with application to face verification. In *Biometrics (ICB), 2016 International Conference on*, pages 1–8. IEEE, 2016. 2
- [19] S. Zhang, R. He, and T. Tan. Demeshnet: Blind face inpainting for deep meshface verification. *arXiv preprint arXiv:1611.05271*, 2016. 1, 2, 4
- [20] E. Zhou, Z. Cao, and Q. Yin. Naive-deep face recognition: Touching the limit of lfw benchmark or not? *arXiv preprint arXiv:1501.04690*, 2015. 1