OCCLUSION-AWARE FACE INPAINTING VIA GENERATIVE ADVERSARIAL NETWORKS

Yu-An Chen[1*], *Wei-Che Chen*[1*], *Chia-Po Wei*[2*], *and Yu-Chiang Frank Wang*[1]

[1]Dept. Electrical Engineering, National Taiwan University, Taipei, Taiwan
[2]Research Center for Information Technology Innovation, Academia Sinica, Taipei, Taiwan

ABSTRACT

Face inpainting aims to restore the corrupted regions of face images due to extreme lighting variations, occlusion, or even disguise. This task becomes especially challenging, when the face images are taken in an unconstrained environment (i.e., with pose, illumination, and expression variations) and the type of corruption is not known in advance. In this paper, we propose a deep-learning based approach of occlusion-aware generative adversarial networks (GAN) for solving this problem. By utilizing GAN pre-trained on occlusion-free images, we are able to detect corrupted image regions automatically with the associated image pixels properly recovered. We produce promising performances on images from the benchmark dataset of LFW, and show that recognition of such face images would be benefited from our proposed approach.

Index Terms— Face inpainting, generative adversarial networks

1. INTRODUCTION

Recent deep-learning based face recognition approaches have reached or surpassed human-level performances on large-scale and real-world datasets like LFW [1, 2]. However, such face recognition systems are typically trained on face datasets in which most of the face images are occlusion-free. Their recognition performances might be degraded if the test image are corrupted due to occlusion/disguise such as sunglasses, scarves, and hands. An intuitive solution for addressing this issue is to restore the pixels of occluded regions in the input image, which is known as the task of face inpainting.

Existing approaches on face inpainting can be divided into two categories. The first category considers input images captured under constrained scenarios (e.g., at frontal pose or with a predetermined type of occlusion) [3, 4, 5]. In [3], Park *et al.* applied recursive PCA to reconstruct eye regions occluded by glasses/sunglasses. In [4], Zhou *et al.* employed Markov random field to enforce the local spatial continuity of the support of reconstruction errors. In [5], Deng *et al.* proposed a graph Laplace method to transform the inpainting problem into the classic graph-labeling task. An concern of the above methods is that their training data need to include image samples of the subject to be inpainted. However, such requirements might not be be practical in real-world applications.

The approaches in the second category focus on dealing with unconstrained input images [6, 7, 8]. For example, Xie *et al.* [6] combined sparse coding and a denoising auto-encoder to address the task of image denoising and blind inpainting. In [7], Pathak *et al.* introduced the context encoder which learned a representation capturing both appearance and semantics of visual structures. In [8], Yeh *et al.* proposed to inpaint occluded or corrupted images based on a pre-trained generative adversarial network. However, the above methods assume that the locations/sizes of image regions to be inpainted are known in advance. This assumption might not hold for practical inpainting tasks (or it would require manual segmentation of occluded regions from the input image).

In this paper, we address the face inpainting problem for unconstrained face images that exhibit pose, illumination, and expression variations plus occlusion. We present a generative adversarial network (GAN) based algorithm, *Occlusion-Aware GAN*. Our approach is able to determine the optimal parameter for GAN to restore the corrupted face regions, while such regions will be automatically identified during the learning process. The inpainting process of our proposed method iteratively observes/updates a binary matrix, with each entry indicating whether the associated pixel is occlusion free. In contrast, previous face inpainting approaches [6, 7, 8] all require the knowledge of locations of occluded regions. Later in our experiments, we will qualitatively and quantitatively verify our proposed method on the Labeled Faces in the Wild (LFW) database [9].

2. OUR PROPOSED METHOD

2.1. A Brief Review of Generative Adversarial Networks

Since our proposed learning model is based on the recent advance of generative adversarial networks (GAN) [10], it is necessary for us to briefly review GAN for the sake of completeness of this paper.

Generally, a GAN framework consists of a generative model $G(\mathbf{z}) \in \mathbb{R}^p \rightarrow \mathbb{R}^{m \times n \times 3}$ and a discriminative model $D(\mathbf{d}) \in \mathbb{R}^{m \times n \times 3} \rightarrow \mathbb{R}$. GAN optimizes a two-player mini-

*The authors contributed equally to this work.

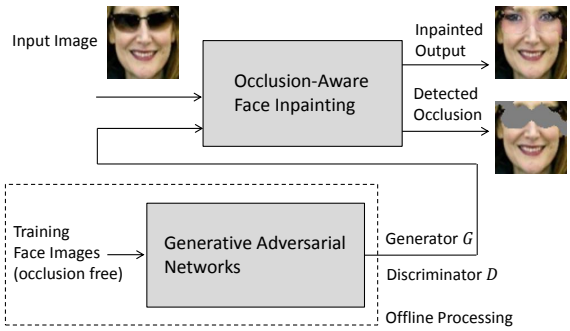

Fig. 1. Flowchart of our proposed framework.

max game with the following objective function:

$$\min_G \max_D V(D,G) = \mathbb{E}_{\mathbf{d}\sim p_{data}(\mathbf{d})}[\log D(\mathbf{d})] + \mathbb{E}_{\mathbf{z}\sim p_{\mathbf{z}}(\mathbf{z})}[\log(1-D(G(\mathbf{z})))], \quad (1)$$

where $\mathbf{d} \in \mathbb{R}^{m\times n\times 3}$ is viewed as a color image of size $m \times n$ pixels sampled from the training set with a particular data distribution, while $\mathbf{z} \in \mathbb{R}^p$ is considered as a randomly sampled vector from a prior distribution $p_{\mathbf{z}}$. From (1), it can be seen that the goal of GAN is to have the generator G with the ability to synthesize/recover images in $\mathbb{R}^{m\times n\times 3}$ lying within the distribution of the training ones, while the discriminator D is to discriminate between the images with the same and different distributions as that of the training ones.

2.2. Occlusion-Aware GAN

2.2.1. Occlusion-Aware Face Inpainting

Given a corrupted/occluded face image $\mathbf{y} \in \mathbb{R}^{m\times n\times 3}$, our proposed method aims to restore the pixles of $\mathbf{y}$ without knowing the exactly which of them are corrupted or occluded in advance. The flowchart of our proposed method is shown in Fig. 1.

To perform this task of occlusion-aware face inpainting, we denote the reconstructed image output as $\mathbf{y}_r \in \mathbb{R}^{m\times n\times 3}$. More specifically, we reconstruct $\mathbf{y}_r$ by:

$$\mathbf{y}_r = \mathbf{W} \odot \mathbf{y} + (1-\mathbf{W}) \odot G(\mathbf{z}), \quad (2)$$

where $\odot$ indicates pixel-wise multiplication, and $G(\mathbf{z})$ is the generator of GAN which predicts the pixel outputs from occluded/corrupted regions. We note that, $\mathbf{W} \in \mathbb{R}^{m\times n\times 3}$ in (2) is a binary matrix, which can be viewed as a mask with each entry equal to 1 only if the associated pixel is occlusion free. As detailed in the next subsection, our algorithm will learn and derive both the input vector $\mathbf{z} \in \mathbb{R}^p$ and the binary matrix $\mathbf{W}$ automatically, without the prior knowledge of type of corruption (or the corresponding pixel locations).

It is worth pointing out that, while a recent work in [8] utilized a similar formulation as (2) for GAN-based face inpainting, they only focused on the recoverability of the occluded image region. Moreover, they required the prior knowledge of the binary matrix $\mathbf{W}$. In other words, if the type of image corruption is not seen during training, they cannot provide satisfactory performance (see examples shown in [8]).

2.2.2. Our Proposed Formulation

We now discuss how our proposed GAN model performs occlusion-aware face inpainting without prior knowledge on the type of image corruption. In our work, we propose the following objective function for learning the vector $\mathbf{z} \in \mathbb{R}^p$ and the binary matrix $\mathbf{W} \in \mathbb{R}^{m\times n\times 3}$:

$$\min_{\mathbf{z},\mathbf{W}} \|\mathbf{W} \odot G(\mathbf{z}) - \mathbf{W} \odot \mathbf{y}\|_1 + \lambda \log(1 - D(\mathbf{W} \odot \mathbf{y} + (1-\mathbf{W}) \odot G(\mathbf{z}))). \quad (3)$$

From (3), we can see that the first term focuses on reconstructing image regions without corruption by imposing the derived mask $\mathbf{W}$, while the second term enforces the recoverability of face images via our GAN model (and is regularized by the parameter λ). To further simplify the above equation, we can rewrite the second term in (3) as $\lambda \log(1 - D(\mathbf{y}_r))$, which indicates that the reconstructed image $\mathbf{y}_r$ would lie within the distribution of training image data.

Different from [8] which adopts a similar objective function as (3) but replaces $\lambda \log(1 - D(\mathbf{y}_r))$ with $\lambda \log(1 - D(G(\mathbf{z})))$, our formulation would enforce the recoverability of output image $\mathbf{y}_r$ instead of the inpainted region $G(\mathbf{z})$ only (as did in [8]). More importantly, the proposed formulation in (3) allows one to identify the corrupted image regions (via $\mathbf{W}$). We now detail how we solve both $\mathbf{z}$ and $\mathbf{W}$ in our GAN-based face inpainting framework.

Learning z for Occlusion-Aware GAN

We apply the technique of alternative optimization to derive $\mathbf{z}$ and $\mathbf{W}$ for optimizing (3). To derive and update the vector input $\mathbf{z}$ for GAN, we fix the binary matrix $\mathbf{W}$. As a result, we rewrite and simplify the original objective function to be optimized as follows:

$$\min_{\mathbf{z}} \|\mathbf{W} \odot G(\mathbf{z}) - \mathbf{W} \odot \mathbf{y}\|_1 + \lambda \log(1 - D(\mathbf{W} \odot \mathbf{y} + (1-\mathbf{W}) \odot G(\mathbf{z}))). \quad (4)$$

Following [8], we adopt pre-trained GANs and apply the resulting generator G and discriminator D. Hence, (4) can be solved via gradient descent optimization algorithms. In particular, adaptive moment estimation (Adam) [11] is utilized for solving (4) due to its computational efficiency.

Learning The Occlusion Mask W

We note that, the binary matrix $\mathbf{W} \in \mathbb{R}^{m\times n\times 3}$ has three channels corresponding to those of a color image, with the masks for each channel are identical to each other. To learn the binary matrix $\mathbf{W}$ with a fixed $\mathbf{z}$, we can simply focus on the the third dimension in the objective function for the ease of both presentation and derivation.

Algorithm 1 Face Inpainting via Occlusion-Aware GAN

Input: Input impage $\mathbf{y}$
  while not converged **do**
    Update vector $\mathbf{z}$
      Solving (4) using back-propagation
    Update binary matrix $\mathbf{W}$
      Calculate $\mathbf{W}$ based on (5)
  end while
  Compute $\mathbf{y}_r$ as in (2)
Output: The reconstructed image $\mathbf{y}_r$

In order to identify the corrupted image regions automatically, we uniquely propose to learn $\mathbf{W}$ as follows:

$$\mathbf{W}(i,j) = \rho(\mathbf{e}(i,j)),\ \mathbf{e} = \mathbf{y} - \mathbf{y}_r,$$
$$\rho(e) = \frac{\exp(-\mu e^2 + \mu\delta)}{1 + \exp(-\mu e^2 + \mu\delta)}, \tag{5}$$

where $\mathbf{W}(i,j)$ denotes the (i,j)-entry of $\mathbf{W}$, and matrix $\mathbf{e}$ calculates the pixel value difference between the input image $\mathbf{y}$ and the reconstructed output $\mathbf{y}_r$.

It is worth noting that, the residual function $\rho : \mathbb{R} \to \mathbb{R}$ in (5) aims to output 0 for large input values; on the other hand, its output would be close to 1 when the input value is small. That means, if the (i,j)-pixel is poorly reconstructed by our GAN model (i.e., potentially a corrupted pixel), we have $\mathbf{W}(i,j) = \rho(\mathbf{e}(i,j)) \approx 0$; if the (i,j)-pixel does not observe significant reconstruction error, then $\mathbf{W}(i,j) \approx 1$ can be expected. This strategy is inspired by our previous work for robust face recognition via sparse representation [12], while now we advance and extend such techniques for the purpose of occlusion-aware facial image inpainting.

2.2.3. Algorithm

With the above derivation details, the inpainting process of our proposed method can be summarized in Algorithm 1. Although the use of alternative optimization requires one to iterate between the learning stages of $\mathbf{z}$ and $\mathbf{W}$, we observe that the proposed GAN-based inpainting algorithm would converge within 3 to 5 iterations. Fig. 2 shows example results (i.e., the predicted corrupted region and the recovered pixels) using our proposed model at different iteration stages.

3. EXPERIMENTAL RESULTS

3.1. Implementation

In our work, we utilize the architecture of deep convolutional GAN (DCGAN) [13] to train the generative model G and the discriminative model D. The generator input $\mathbf{z}$ is a vector in $\mathbb{R}^{100}$ drawn from the uniform distribution on [-1,1]. The generator output is of size $64 \times 64 \times 3$. To train the DCGAN, we take the CelebFaces Attributes (CelebA) dataset [14], which consists of 10,177 identities with 202,599 face images. We note that, face images with the attribute of Eyeglasses in CelebA are excluded, since we only require occlusion-free images in the training set. After the optimizion of our proposed algorithm is complete (i.e., convergence of Algorithm 1), we apply morphological processing of image closing and opening to refine the inpainted output regions.

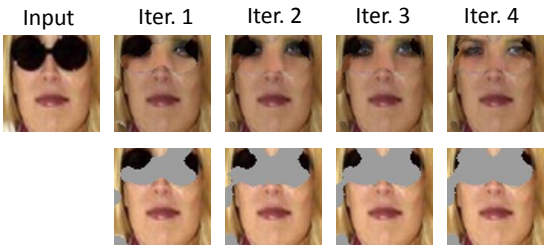

Fig. 2. Our example inpainting outputs (top) and the detected binary matrix $\mathbf{W}$ (bottom) through the iterative optimization process.

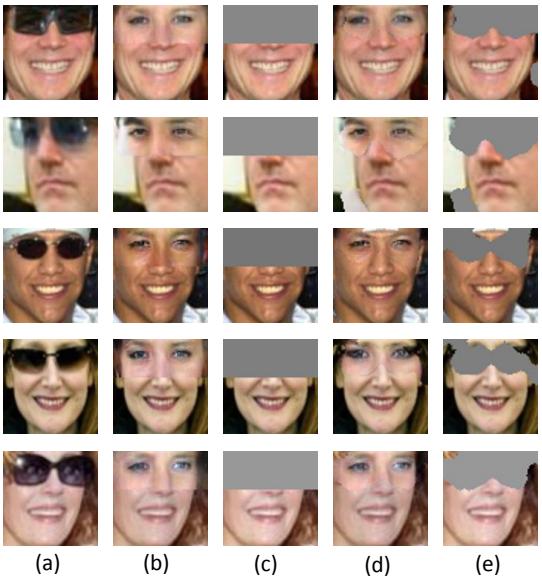

Fig. 3. Example results: (a) LFW input images, (b) inpainted outputs of [8], (c) pre-defined masks of [8], (d) our inpainted outputs, and (e) occlusion masks automatically detected by our method.

3.2. Qualitative Evaluation

For evaluating the performance of face inpainting, we consider the Labeled Faces in the Wild (LFW) database [9]. While LFW contains 5,749 identities with 13,233 images, we only select the images with either sunglasses or occluded by synthetic masks as test inputs (see first columns in Figures 3 and 4 for example)). We compare our method with a recent face inpainting method [8], which is also based on GAN.

Example inpainting results are shown and compared in Figures 3 and 4. We can see that, since the method of [8] required to predefine the location/size of occluded image regions, it was not able to produce satisfactory results if such regions contain occlusion-free pixels (e.g., the nose region of the first input image in Fig. 3). Unlike [8], our method was able to detect occluded regions automatically, as shown in the

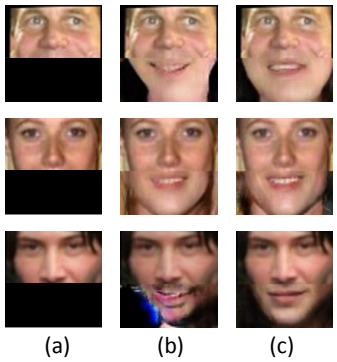

Fig. 4. (a) Example input images with synthetic occlusion regions, and the inpainted outputs of (b) [8] and (c) our approach.

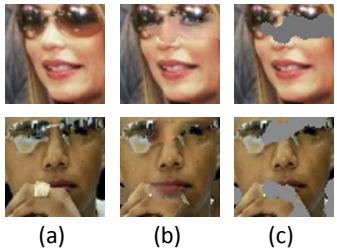

Fig. 5. Example failure samples. The input images are shown in (a), while the inpainted outputs and the detected corrupted regions are shown in (b) and (c), respectively

fifth column of Fig. 3. As for Fig. 4, we do not perform occlusion detection, since the lower part of the input image does not contain any useful information. It can be seen from Fig. 4 that our method still produced natural-looking inpainted results, and thus would be preferable over the use of [8].

While our method leads to promising inpainted results, it still has its limitations. As shown in Fig. 5, if the pixel values of the corrupted image regions are not significantly different from the occlusion-free ones (e.g., reflections in sunglasses), our method would not correctly detect the occlusion, and thus result in partially inpainted outputs.

In Fig. 6, we demonstrate the average convergence rate of our proposed method in determining the corrupted image regions. It can be seen that, based on our experiments, our algorithm would converge within 5 iterations for performing face inpainting via GAN.

3.3. Quantitative Evaluation

We now verify that the proposed face inpainting scheme could benefit the recognition of occluded face images. We randomly select 50 subjects with more than two images from the LFW database. For each subject, we randomly choose a pair of images, one as the gallery and the other as the probe, which results in 50 pairs of matched image pairs. We also form 200 pairs of mismatched images from the selected subjects. For each matched/mismatched pair, we apply a lightened convolutional neural network (CNN) [15] to extract features from

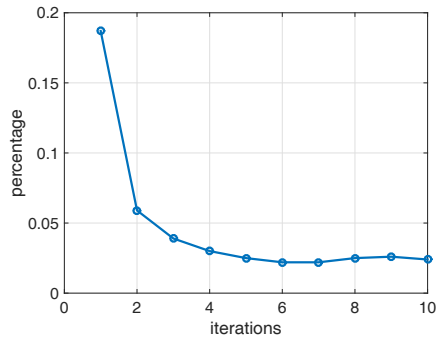

Fig. 6. Convergence of our face inpainting algorithm. The vertical axis shows the pixel difference in detected occlusion regions, while the horizontal axis denotes the iterations.

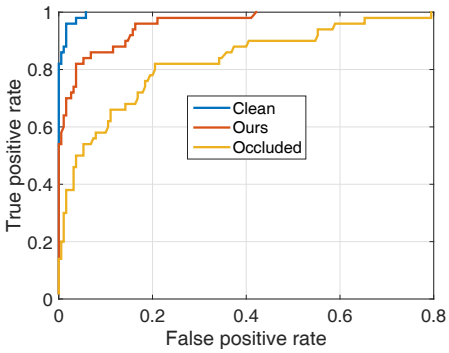

Fig. 7. ROC curves for face verification. Note that occlusion-free images are expected to result in the best performance.

each image, and calculate the cosine distance between the pair of extracted features.

For each probe image, we consider three scenarios: clean (occlusion-free), occluded (by imposing synthetic masks as in the first column of Fig. 4), and those inpainted by our method. The receiver operating characteristic (ROC) curves for the three scenarios are plotted in Fig. 7, in which the equal error rates (EER) were 0.98, 0.80, and 0.88 for images of Clean, Occluded, and Ours, respectively. From Fig. 7, we see that the recognition performance using occluded images degraded significantly, while our inpainting method improved the recognition performance (with EER from 0.80 to 0.88). Thus, the effectiveness of our approach for occluded face recognition can be successfully verified.

4. CONCLUSION

In this paper, we proposed a GAN based approach for occlusion-aware face inpainting. Our approach is able to automatically detect corrupted image regions and recover the associated pixels. As a result, natural-looking face images can be produced. With a major advantage of not requiring the prior knowledge of locations and types of image corruption, our method was shown to perform favorably against recent deep learning based approach on face inpainting, and benefit recognition tasks particularly on occluded face images.

5. REFERENCES

[1] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[2] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[3] J.-S. Park, Y. H. Oh, S. C. Ahn, and S.-W. Lee, "Glasses removal from facial image using recursive error compensation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005.

[4] Z. Zhou, A. Wagner, H. Mobahi, J. Wright, and Y. Ma, "Face recognition with contiguous occlusion using Markov random fields," in *Proceedings of International Conference on Computer Vision (ICCV)*, 2009.

[5] Y. Deng, Q. Dai, and Z. Zhang, "Graph Laplace for occluded face completion and recognition," *IEEE Transactions on Image Processing*, 2011.

[6] J. Xie, L. Xu, and E. Chen, "Image denoising and inpainting with deep neural networks," in *Advances in Neural Information Processing Systems (NIPS)*, 2012.

[7] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, "Context encoders: Feature learning by inpainting," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[8] 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.

[9] 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," Tech. Rep. 07-49, University of Massachusetts, Amherst, 2007.

[10] 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 (NIPS)*, 2014.

[11] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *International Conference on Learning Representations (ICLR)*, 2015.

[12] C.-P. Wei and Y.-C. F. Wang, "Undersampled face recognition via robust auxiliary dictionary learning," *IEEE Transactions on Image Processing*, 2015.

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

[14] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proceedings of International Conference on Computer Vision (ICCV)*, 2015.

[15] X. Wu, R. He, and Z. Sun, "A lightened CNN for deep face representation," *arXiv preprint arXiv:1511.02683*, 2015.