Extreme Face Inpainting with Sketch-Guided Conditional GAN

Nilesh Pandey and Andreas Savakis; Rochester Institute of Technology; Rochester, NY, USA.

Abstract

Recovering badly damaged face images is a useful yet challenging task, especially in extreme cases where the masked or damaged region is very large. One of the major challenges is the ability of the system to generalize on faces outside the training dataset. We propose to tackle this extreme inpainting task with a conditional Generative Adversarial Network (GAN) that utilizes structural information, such as edges, as a prior condition. Edge information can be obtained from the partially masked image and a structurally similar image or a hand drawing. In our proposed conditional GAN, we pass the conditional input in every layer of the encoder while maintaining consistency in the distributions between the learned weights and the incoming conditional input. We demonstrate the effectiveness of our method with badly damaged face examples.

Introduction

Image inpainting is a popular and challenging problem in computer vision with applications in image editing and restoring of damaged image regions [1], [2], [3], [4], [5], [6]. The damaged image regions may range from rectangular patches of various sizes to irregular regions or entire objects. In image editing, some regions are tactically damaged so that they can be inpainted with different content.

In this paper we deal with the problem of extreme inpainting, where large regions of faces have been damaged with the application of an irregular mask. Initial works dealt with small square patches, while more recent approaches take on the more challenging problem of inpainting large irregular regions [3], [4]. Our face inpainting method, called FIpoly-GAN, is based on the Poly-GAN conditional Generative Adversarial Network (GAN) architecture [7], that was initially proposed for fashion synthesis and is suitable for various tasks, e.g. geometric transformations, image stitching and inpainting. To recover the facial structure that is lost after significant image regions have been masked, we utilize an edge sketch as a condition that guides the inpainting process by providing structure in the damaged regions. The proposed FIpoly-GAN has shown ability to generalize on faces obtained in the wild from various sources unrelated to the training data. A representative result is shown in in Fig. 1. The main contributions of this paper can be summarized as follows:

- We propose FIpoly-GAN, a conditional GAN architecture that leverages edge information for extreme face inpainting.
- Our FIpoly-GAN architecture uses the edge map, from a sketch or a similar image, as a condition that is fed at various levels of the network to learn the structure of the image for inpainting color and texture in the missing region.
- FIpoly-GAN achieves state of the art results on standard face datasets and outperforms other methods on images that are unrelated to the training dataset.

Figure 1. *Example of sketch-conditioned face inpainting with the proposed FIpoly-GAN network. From left to right: original, damaged image with large irregular mask, hand-drawn sketch, inpainted result.*

Related Work

Early inpainting methods operated under the assumption that the damaged region is a standard shape, i.e. square or rectangle. Free form inpainting, where the damaged regions are large and irregular, poses a challenge to many methods due to lack of relevant information that can be used from nearby regions. A popular approach is the Patch-based image inpainting network [1], which progressively learns to associate nearby image contect to the missing region. However, this approach is not as effective in the case of free form inpainting.

Image Inpainting using Partial Convolution [2] is one of the first methods to tackle free form inpainting on images with irregular damage. It utilizes a neural network with a partial convolution layer and performs an automatic update on the binary mask. Given a binary mask, the partial convolution utilizes non-damaged information near damaged region to estimate missing content. The next step in the method is to update the binary mask according to the output of the partial convolution, so that if the output is positive for the damaged region then the mask is updated accordingly.

Free-form image inpainting with Gated Convolution (SN-patchGAN) [3] is a state-of-the-art method for inpainting images with irregular damage. This method introduces a new convolution layer named gated convolution. The input features in the model are first used to compute the gating value, $g = sigmoid(W_g * x)$, where x denotes the input features. The final output of the layer is multiplication of learned features and gating values, $y = \phi(W * x) \odot g$. The architecture includes a contextual attention module with the refinement network, with both having a gated convolution layer. Gated Convolution has inspired several other methods for image inpainting.

Image inpainting via Generative Multi-column Convolutional Neural Networks (GMCNN) [4] The GMCNN network consists of three sub-networks, one generator, one global and one local discriminator for adversarial training. Only the generator is deployed when testing on a new image.

SC-FEGAN [5] and Face-Shop [6] are state-of-the-art image inpainting methods based on image editing. They work by providing the network a simple sketch of the face. Both methods use a similar approach to generate sketch images of the face, which can be edited and passed to the model as conditional input. The Edge-
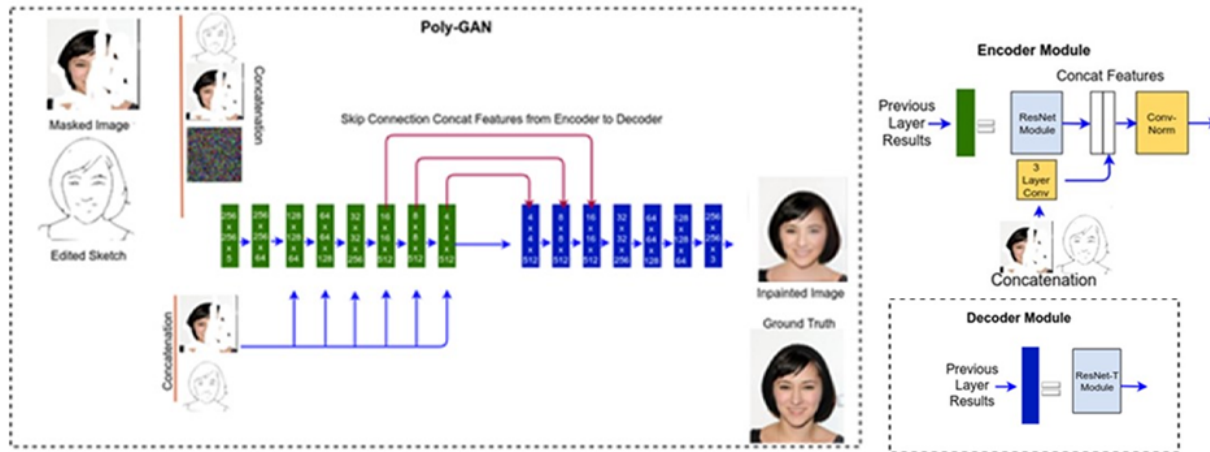
Figure 2. *The FIpoly-GAN architecture used for edge-guided face inpainting. The left side block shows the overall architecture and the right side shows more details of the encoder and decoder layers.*

Connect method [8] uses an edge generator to estimate structural information in the missing regions.

The SC-FEGAN trains both the generator and discriminator to generate a (512x512) image. The SC-FEGAN generator uses gated convolution in all layers, except the initial input layer in the encoder and final output layer in the decoder, followed by a local signal normalization layer, as in [10]. The SC-FEGAN discriminator is similar to that in SN-patchGAN [3].

Proposed Inpainting GAN Architecture
Generator

In the proposed edge-guided inpainting approach, we use the FIpoly-GAN architecture shown in Figure 2, which is based on the Poly-GAN network initially used for fashion synthesis [7]. The main differences between the FIpoly-GAN inpainting architecture and the original Poly-GAN architecture are due to different inputs and training methodology. FIpoly-GAN is a conditional GAN based on encoder-decoder network configuration with two additional modules: the Conv-Module and the Conv-Norm Module, illustrated in Fig. 2. We present the architecture, training methodology and results in the following sections.

Encoder

The FIpoly-GAN encoder is an extension of deep learning conditional GAN encoders with some variations in the network. The most important aspect of the FIpoly-GAN encoder is that the conditional input is fed to multiple layers. We found that if the conditional input is provided only in the first layer of the encoder, then its effect diminishes as the features propagate deeper in the network.

In our encoder, each layer is like a ResNet module [9] consisting of two spectral normalization layers followed by a ReLU activation function. The encoder in Figure 2 has modules to handle the incoming conditional inputs in every layer using a Conv Module and a Conv-Norm Module, shown in Fig. 2. In the **Conv Module**, we pass the conditional inputs to every convolution layer starting with the third layer, followed by a ReLU activation function. The learned features from the previous layer have different
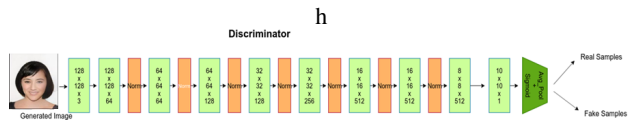

Figure 3. *Discriminator network using in FIpoly-GAN.*

statistics than the learned features from Conv Module, so we concatenate the features for input through the Conv-Norm Module. The Conv-Norm Module consists of two convolutions each followed by an instance normalization and an activation function. It is important to note that the initial input to the encoder consists of conditional input with Gaussian noise.

Decoder

The FIpoly-GAN decoder, shown in Figure 2 is the same as the one in Poly-GAN [7]. The layers in the decoder consist of a ResNet module with two spectral normalization layers followed by a ReLU and a transposed convolution to upsample the learned features. The decoder learns to generate the desired output using the input from the encoder. To increase the effectiveness of the decoder, we utilize skip connections from low resolution layers of the encoder to the corresponding resolution layers of the decoder, as illustrated in Fig. 2. It is known from previous works, such as Progressive GAN [10], that the coarse layers learn structural information which is easy to manipulate through the conditional input. Our experiments determined that only three low resolution skip connections are needed. These skip connections help the decoder enforce the structure of the conditional inputs fed to consecutive layers of the encoder.

Discriminator

The FIpoly-GAN discriminator network shown in Figure 3 is the same as the one in [7]. The discriminator architecture is similar to the one in Cycle GAN [11], as it has been found to be effective in various applications.

Loss

The loss function used for training consists of three components: adversarial loss L_{adv}, GAN loss L_{gan} and identity loss L_{id}. The total loss and its components are presented below, where D is the discriminator, G is the generator, $x_i, i = 1, ..., N$ represent N input samples from different distributions $p_i(x_i), i = 1, ...N$, t is the target, F are fake labels for generated images and R are real labels for ground truth images.

$$\min_G L_{GAN} = \lambda_3 E_{x_1 \sim p_1(x_1), ..., x_N \sim p_N(x_N)} \tag{1}$$
$$\|D(G(x_1, .., x_N)) - R\|_2^2$$

$$\min_D L_{Adv} = \lambda_1 E_{t \sim p_d(t)} \|D(t) - R\|_2^2 +$$
$$\lambda_2 E_{x_1 \sim p_1(x_1), .., x_N \sim p_N(x_N)} \tag{2}$$
$$\|D(G(x_1, .., x_N)) - F\|_2^2$$

$$L_{Id} = \lambda_4 E_{t \sim p_d(t), x_1 \sim p_1(x_1), .., x_N \sim p_N(x_N)} \tag{3}$$
$$\|G(x_1, .., x_N) - t\|_1$$

In the above equations, λ_1 to λ_4 are hyperparamters that are tuned during training. Similarly to [12], we use the L_2 loss as the adversarial loss of our GAN in Eqs. (1) and (2). We use the L_1 loss for the identity loss, which helps reduce texture and color shift between the generated image and the ground truth. We considered adding other loss functions, such as the perceptual loss [13] and SSIM loss, but they did not improve our results.

For each epoch, we first train the generator with the GAN loss and identity loss in Eqs. (1) and (3) respectively. Then we train the discriminator with the adversarial loss in Eq. (2). This process was repeated for several epochs until a satisfactory result was obtained.

Experiments and Results

For our experiments, we used standard datasets, specifically Labeled Faces in the Wild (LFW) [14] and Large-scale Celeb-Faces Attributes (CelebA) [15]. The LFW dataset contains more than 13,000 images of faces collected from the web. Each face is labeled with the name of the person pictured and 1680 people have two or more distinct photos in the dataset. The version of LFW dataset we uses has been aligned using the deep funneling method. The CelebA dataset has more than 200k celebrities faces in different scenes, poses, and with and without jewelry. The dataset is pre-processed and aligned. A key similarity between the datasets is the alignment of facial features, e.g. all images are aligned so that the tip of the nose is at the center of the image. Training of FIpoly-GAN was done using free form irregular masks.

We used publicly released code or online demos for comparisons with other methods. We experimented with image inpainting for irregular holes using the method of image inpainting via Generative Multi-column Convolutional Neural Networks [4], image inpainting Using Gated Convolution [3], and image inpainting using SC-FEGAN [5], due to the availability of code or online demo.

We next present our experiments with conditional inputs to our network, passing the structural information such as edges, edges of similar-looking images, and edited edges.
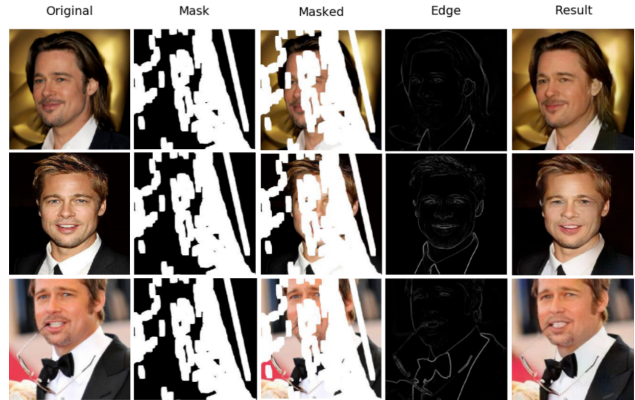

Figure 4. *Results inpainting with FIpoly-GAN when passing the edge map as conditional input. These examples use images from the internet to illustrate that FIpoly-GAN is able to generalize to unknown images from random sources.*
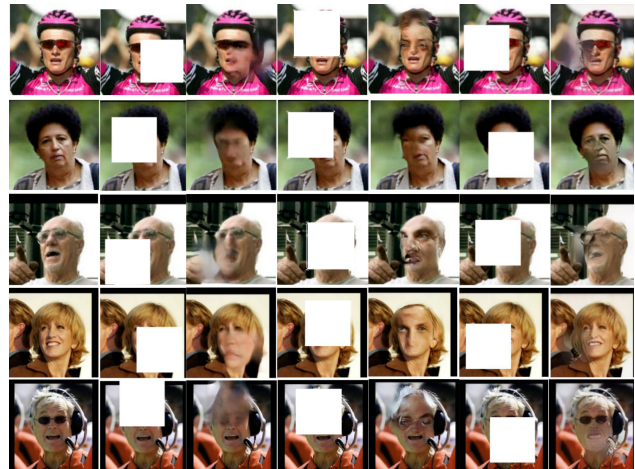

Figure 5. *Results comparison of edge-based FIpoly-GAN with Generative Multi-column convolutinal nets (GMCNN) [4] and Gated Convolution method [3]. Left to right columns show: Original, masked, GMCNN result, masked, GatedConv result, masked, edge-based FIpoly-GAN result. Best viewed in color and 4× zoom.*

Edge-based Face Inpainting

Edge-based face inpainting operates under the assumption that the edges of the masked regions are available. In our experiments we obtained the edge map from the original image, operating under a best case scenario. Alternatively, the edges can be obtained from a similar image or a sketch drawing. We utilized the processing pipeline as an encoder-decoder style conditional GAN illustrated in Figure 2. For the training data, we randomly sampled nearly 30 thousand images from the CelebA [15] dataset. We created edge maps for the sampled images using the Sobel filter and thresholded the filter response to obtain the edges that are used as conditions. We used an NVIDIA 1080TI GPU for training.

We display representative results from our edge-based inpainting in Figure 4 and comparisons with other methods in Figure 5. Our method is trained using free-form irregular masks

Method	SSIM	MSE	SNR
Edge-based FIpoly-GAN	0.824	**27.533**	**1.678**
GMCNN [4]	**0.845**	36.041	1.568
Gated Conv [3]	0.834	37.261	1.540

Quantitative results comparison of edge-based FIpoly-GAN with other methods on randomly sampled images from the LFW dataset. Bold shows best performance.

on CelebA, while GMCNN is pre-trained on **rectangular** shaped mask on CelebA. The test images are randomly sampled from the LFW dataset. From the results of Fig. 5 it is evident that our edge-conditioned FIpoly-GAN significantly outperforms the other methods when using images from unknown sources. In contrast to the other methods, our network has the ability to generalize and recover the masked image content consistently close to the original image.

For quantitative comparison with other methods, we trained on the LFW dataset. We used three quantitative metrics to compare our results against state of the art methods: i) Structural Similarity Index (SSIM); ii) Mean Squared Error (MSE); iii) Signal to Noise Ratio (SNR). SSIM should be high for similar images, while MSE should be low for similar images. Higher SNR values represent better quality of image recovery. The results indicate that the edge-based FIpoly-GAN scores highest in two of the metrics. Most importantly, our method is able to generalize much better than the other methods on previously unseen images, as illustrated in Fig. 5.

Sketch-based Face Inpainting

Our edge-based inpainting experiments demonstrated that even if the image is badly damaged, we can recover the original content by conditioning on the edge information. However, edge information is not readily available when inpainting. As an alternative, we consider a hand-drawn sketch that provides edge information for the damaged image regions. We created a dataset consisting of sketches of CelebA faces following the instructions provided in FaceShop [6]. We used Potrace to generate a vectorized representation of the image edges, which is used for training purposes. We randomly sampled 150k images for training and testing purposes from CelebA. The model was trained on Ubuntu 16.04 with an NVIDIA 1080TI GPU for nearly 18 hours.

Representative results generated by our sketch-based FIpoly-GAN method are shown in Figure 6. The results illustrate that the network can generate realistic images by inpainting conditioned on a rough sketch.

Conclusion

In this paper, we presented FIpoly-GAN, a conditional GAN for extreme face inpainting. Our work demonstrated the importance of edge information in the recovery of badly damaged regions for image restoration or editing. The FIpoly-GAN approach performs equally well or better than state-of-the-art methods on standard datasets and outperforms other methods on images unrelated to the training data.

Acknowledgments

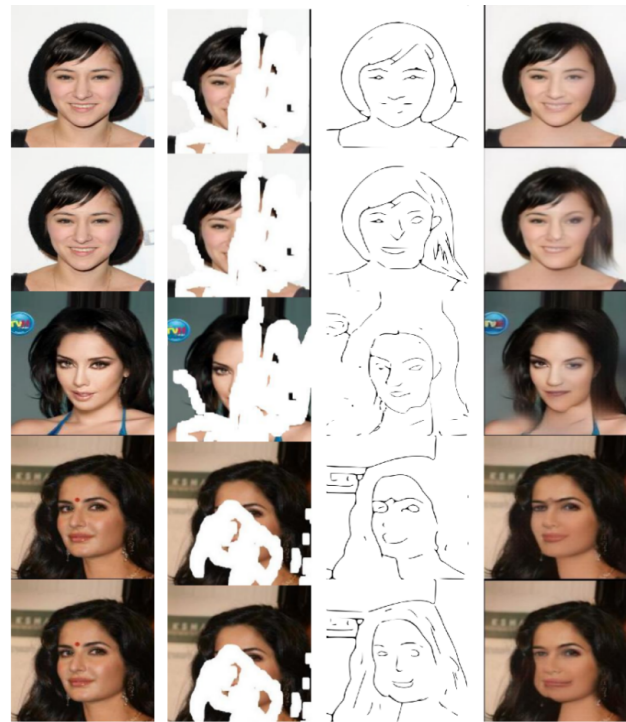This research was supported in part by the Air Force Office of Scientific Research (AFOSR) under Dynamic Data Driven Applications Systems (DDDAS) grant FA9550-18-1-0121 and the Center for Emerging and Innovative Sciences (CEIS), an Empire State Development-designated Center for Advanced Technology..

Figure 6. *Results obtained with sketch-based FIpoly-GAN. From left to right: original, damaged, sketch, inpainted image.*

References

[1] H. Zhang, Z. Hu, C. Luo, W. Zuo, and M. Wang, "Semantic image inpainting with progressive generative networks," In Proceedings of the 26th ACM International Conference on Multimedia, pp. 1939-1947, 2018.

[2] G. Liu, F.A. Reda, K.J. Shih, T.C. Wang, A. Tao, and B. Catanzaro, "Image inpainting for irregular holes using partial convolutions," In Proceedings of the European Conference on Computer Vision (ECCV), pp. 85-100, 2018.

[3] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, T.S. Huang, "Free-form image inpainting with gated convolution," In Proceedings of the IEEE International Conference on Computer Vision, pp. 4471-4480, 2019.

[4] Y. Wang, X. Tao, X. Qi, X. Shen, and J. Jia, "Image inpainting via generative multi-column convolutional neural networks," In Advances in Neural Information Processing Systems, pp. 331-340, 2018.

[5] Y. Jo and J. Park, "SC-FEGAN: Face Editing Generative Adversarial Network with User's Sketch and Color," In Proceedings of the IEEE International Conference on Computer Vision, pp. 1745-1753, 2019.

[6] T. Portenier, Q. Hu, A. Szabo, S.A. Bigdeli, P. Favaro, and M. Zwicker, "Faceshop: Deep sketch-based face image editing," ACM Transactions On Graphics, 2018.

[7] N. Pandey and A. Savakis, "Poly-GAN: Multi-conditioned GAN for fashion synthesis," Neurocomputing, 414, pp.356-364, 2020.

[8] K. Nazeri, E. Ng, T. Joseph, F. Qureshi, and M. Ebrahimi, "Edge-Connect: Structure guided image inpainting using edge predic-

tion," International Conference on Computer Vision Workshops (IC-CVW), 2019.

[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, pp. 770-778, 2016.

[10] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," arXiv preprint arXiv:1710.10196, 2017.

[11] J.Y. Zhu, T. Park, P. Isola, and A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," In Proceedings of the IEEE International Conference on Computer Vision, pp. 2223-2232, 2017.

[12] X. Mao, Q. Li, H. Xie, R.Y. Lau, Z. Wang, and S. P, Smolley, "Least squares generative adversarial networks," In Proceedings of the IEEE International Conference on Computer Vision, pp. 2794-2802, 2017.

[13] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," In European Conference on Computer Vision, pp. 694-711, 2016.

[14] G.B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," In Workshop on faces in 'Real-Life' Images: detection, alignment and recognition, 2008.

[15] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," In Proceedings of the IEEE International Conference on Computer Vision, pp. 3730-3738, 2015.

Author Biography

Nilesh Pandey completed his Bachelor in Electrical Engineering at Ramrao Adik Institute of Technology (2016) and received his Master in Computer Engineering from Rochester Institute of Technology (2019). His research interests include Generative Adversarial Networks (GAN), and applications for tracking and pose estimation architectures.

Andreas Savakis is Professor of Computer Engineering and Director of the Center for Human-aware Artificial Intelligence (CHAI) at Rochester Institute of Technology (RIT). He received his Ph.D. in Electrical and Computer Engineering from North Carolina State University. Prior to joining RIT, he was Senior Research Scientist at Kodak Research Labs. His research interests include computer vision, deep learning, machine learning, domain adaptation, object tracking, human pose estimation, and scene analysis. Dr. Savakis has coauthored over 12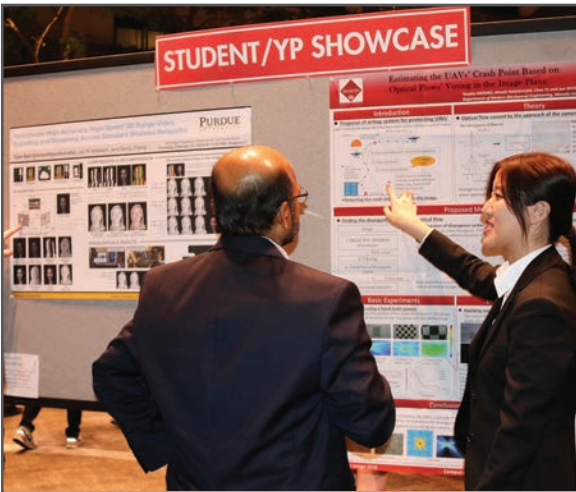0 publications and is co-inventor on 12 U.S. patents.