Optimized Based on Dexined: Edge-Connect'S Two-Stage Face Image Restoration

Jinxian Chen

School of computer science, University Science Malaysia, Penang, Malaysia

Abstract. Image restoration is an important research direction in the field of computer vision. Especially in the task of face image restoration, how to generate structurally complete and visually realistic image content in the missing area has always been a challenging problem. Aiming at the problem of low edge guidance quality in traditional methods, this paper proposes a face image restoration method based on a two-stage architecture. In the first stage, the deep edge detection network DexiNed is used to replace Canny operator, and the masked image is used as input. The edge map generated by the complete image is used as the supervision signal to train a model with complete edge prediction ability. In the second stage, under the guidance of the predicted complete edge map, the generative adversarial network (GAN) is used to complete the content completion and detail restoration of the final image. Experimental results on the CelebA dataset show that the proposed method is superior to existing methods such as Edge-Connect in both PSNR and SSIM mainstream evaluation indicators. The PSNR improvement can reach up to 1.2dB, and the SSIM index is improved by about 0.14%, which is superior in terms of structure restoration and visual consistency.

1 Introduction

Image inpainting is a basic but challenging task in the field of computer vision. Its main goal is to fill in the missing or damaged areas in the image with appropriate content, so that the repaired image is visually coherent and natural, and semantically reasonable. Among various image inpainting scenarios, face image inpainting has received increasing attention in practical applications such as photo editing, face restoration, and identity protection.

However, due to the complexity of the human face structure, this task still faces challenges. Currently, there are two main difficulties in face restoration. The first is the complex facial structure. The face contains multiple key structures such as eyes, nose, and mouth. The shapes and relative positions of these parts are very important. Any slight deviation in the restoration will seriously affect the authenticity of the overall image. The second is the high requirement for detail consistency. Facial image restoration must not only fill in the missing areas, but also make the reconstructed areas consistent with the

chenjinxian443357848@student.usm.my

original image in terms of texture, lighting, style, etc. Otherwise, it is easy to produce an obvious "sense of splicing" or blur distortion.

To address the above issues, this paper proposes a two-stage face image restoration method based on Dense Extreme Inception Network (DexiNed) [2] and Edge-Connect. This method uses a face image with a random mask as input to guide DexiNed to predict the edge information of the missing area. At the same time, the edge map generated by the corresponding complete image is used as a supervisory signal, so that the model can learn to reasonably infer the edge structure in the occluded area when only part of the image content is visible. Mask information is introduced during the training process to clarify the boundary between known and unknown areas and avoid unnecessary modifications to the known area. The predicted edge image is then input into the second-stage guided generative adversarial network (GAN) of Edge-Connect to complete the final image completion. The goal of this research is to improve the edge detection quality of the first stage of Edge-Connect so that the edge completion network can make predictions based on more accurate structural information, thereby improving the final restoration quality.

This paper has two main contributions: First, this paper proposes a high-quality edge prediction scheme based on DexiNed. By introducing the deep edge detection network DexiNed to replace the traditional Canny operator, the clarity, continuity and structural integrity of the edge map are significantly improved, providing more reliable structural information for subsequent image restoration. Second, this paper constructs a weakly supervised training mechanism to enhance the edge completion capability of DexiNed. Using masked face images as input and combining them with edge maps generated by complete images for supervision, DexiNed has the ability to predict missing edges based on context, thereby improving the structural reasoning effect in occluded areas.

2 Research Methods

2.1 Overall process

This research method is divided into two stages: edge prediction and image restoration. In the first stage, the DexiNed network is weakly supervised and trained using masked images, so that it has the ability to infer edge structures in missing areas. After training, DexiNed can generate an edge map with complete structural information, which includes not only the real edges of known areas, but also the inferred edges of unknown areas. In the second stage, the predicted edge map and the original image with defects are input into the Edge-Connect image restoration network, and the image is reconstructed with consistent structure and natural texture through edge guidance, and finally a restoration result with realistic visual effects is generated. The two-stage process is shown in Figure 1.

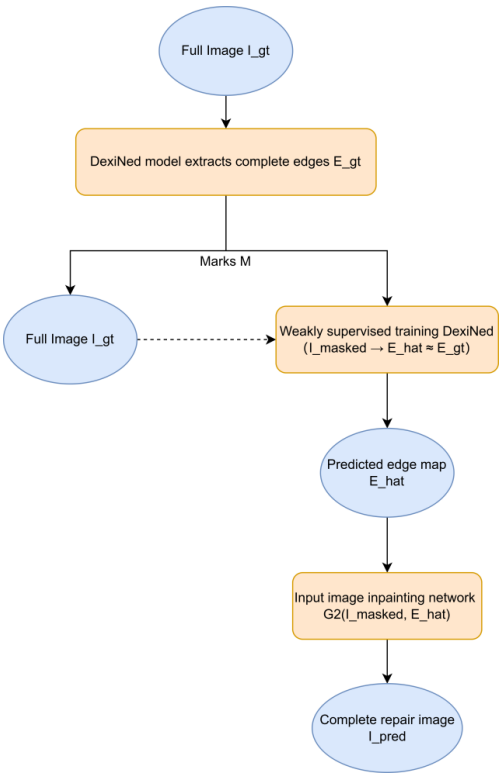

Figure 1 Overall flow chart of this research plan（Picture credit: Original）

2.2 DexiNed Weakly Supervised Edge Inference

To improve the effect of edge-guided face image restoration, this research introduced the deep learning-based edge detection network DexiNed in the edge prediction stage of Edge-Connect. DexiNed has stronger edge perception capabilities and can generate smoother, clearer and more coherent edge maps, effectively overcoming the limitations of traditional edge detection methods (such as the Canny operator) in terms of noise, fractures and detail representation.

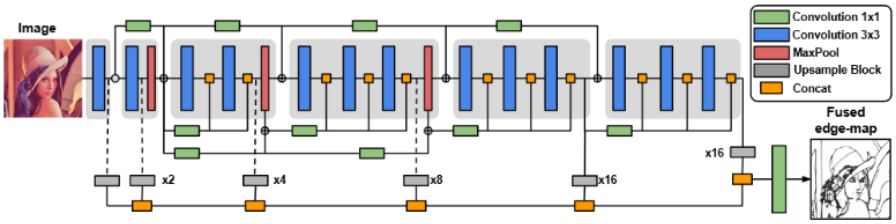

Figure 2 DexiNet model structure diagram [2]

The DexiNet model structure is shown in Figure 2. The whole model consists of multiple 3×3 convolution modules, downsampling (MaxPool), upsampling modules, and feature fusion layers. DexiNed extracts edge features at different scales (×2, ×4, ×8, ×16) and fuses them using jump connections, which effectively improves the continuity and accuracy of edge detection. The final output edge map is generated by the fusion of

multiple scale features through 1×1 convolution, with high resolution and structural integrity.

In order to enable DexiNed to have the ability to predict complete structural edges in missing areas, this research designed a training strategy based on weak supervision. By constructing occluded inputs and complete edge maps as pseudo-labels, the model is guided to learn to infer edge information of occluded areas.

2.2.1 Input and output data design

First is the input image. For each complete face image, this paper constructs a binary mask M of the same size, randomly blocking a part of the image (usually the center of the image, with a blocking ratio of about 10%–15% of the image area) to simulate the situation of image missing. Based on the mask information, this research constructs the missing image as follows:

$$\mathrm{I}_{\mathrm{masked}} = \mathrm{I}_{\mathrm{gt}} \odot (1 - \mathrm{M}) \quad (1)$$

Among them, $\mathrm{I}_{\mathrm{masked}}$represents the face image with a mask added, $\odot$ represents the pixel-by-pixel multiplication operation, the position where the pixel value of M is 1 represents the occluded area, and 0 represents the visible area.

The second point is the supervisory signal (complete edge image). This research uses the DexiNed model to run on the unoccluded complete image I_{gt} to obtain a clear and complete edge image as a pseudo label for training (i.e., supervisory signal):

$$E_{gt} = DexiNed(I_{gt}) \quad (2)$$

Among them, E_{gt}represents the edge image generated by the DexiNed model based on the complete image I_{gt}.

The third point is the output image (predicted edge image). Next, the occluded image $\mathrm{I}_{\mathrm{masked}}$ is input into the trained DexiNed model and the predicted edge map is output:

$$\hat{E} = DexiNed(I_{masked}) \quad (3)$$

Among them, $\hat{\mathrm{E}}$represents the edge image of I_{gt} generated by the DexiNed model based on the occluded image.

The training goal is to minimize the difference between the predicted edge map $\hat{\mathrm{E}}$ and the pseudo-label edge map E_{gt}, so that the model can generate as complete and well-structured edge information as possible in the missing area:

$$\hat{E} \approx E_{gt} \quad (4)$$

This weakly supervised edge inference process not only avoids the dependence on real edge annotation data, but also enhances the structure prediction ability of the DexiNed model in occluded scenes, providing a more stable and reliable structure prior to the subsequent image restoration stage.

2.2.2 Mask generation strategy

The mask generation strategy plays a crucial role in DexiNed model training, because the design of the mask directly affects the accuracy of edge prediction and the training effect. Simply using a regular rectangular mask will make the edge prediction task too simple and it is difficult to effectively guide the model to learn complex structural features. Therefore,

this study uses the Quick Draw Irregular Mask Dataset (QD-IMD) strategy to generate masks that are more in line with actual scenarios.

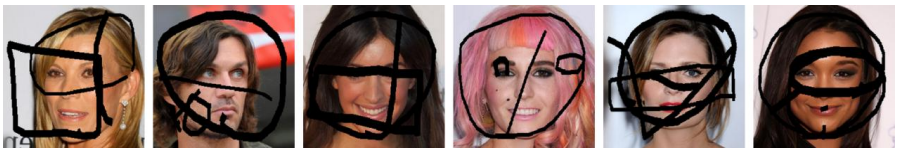

Figure 3 Example of a mask generated by the QD-IMD strategy [3]

QD-IMD is built on the Quick Draw dataset (a collection of 50 million human-drawn images). The specific masking effect is shown in Figure 3. Researchers have found that the combination of manually drawn strokes can generate highly irregular mask patterns, which are more suitable for training models to infer complex edge structures [3]. Using this strategy, this study generates masks for the CelebA [4] dataset. The mask area is set to 10%–15% of the total image area. This ratio ensures the difficulty of the occluded area without excessively affecting the ability to understand the context, which meets the requirements of the structure completion task.

The generated mask map M is shown in Figure 4. It is not only used to block part of the image, but also serves as the basis for calculating the loss. During the training process, the mask is passed into the model together with the input image to help the network learn how to infer the edge information of the missing area and further improve the prediction accuracy and visual consistency of the restoration stage. Through this irregular mask generation strategy, the model can obtain more realistic structural information when processing the occluded area, thereby improving the quality of image restoration.

Figure 4 Example of mask generated by QD-IMD model [3]

2.2.3 DexiNed model training process

During the training process, the DexiNed model takes the occluded image I_{masked} and the mask image M as input. The goal is to predict a complete edge map $\hat{E}$ with a coherent structure, making it as close as possible to the pseudo-label edge map E_{gt} generated by the complete image. First, I_{masked} and M are concatenated into multi-channel inputs so that the model can clearly identify the location of the missing area, thereby focusing on predicting the structural information of the occluded area. Subsequently, forward propagation is performed through the DexiNed network to generate an edge prediction map. During the training process, the Adam optimizer [5] is combined for backpropagation and parameter update to continuously improve the model's edge restoration ability and robustness in occluded scenes.

2.2 Edge Linking and Image Completion

This part will be roughly consistent with the image completion network in the EdgeConnect project.

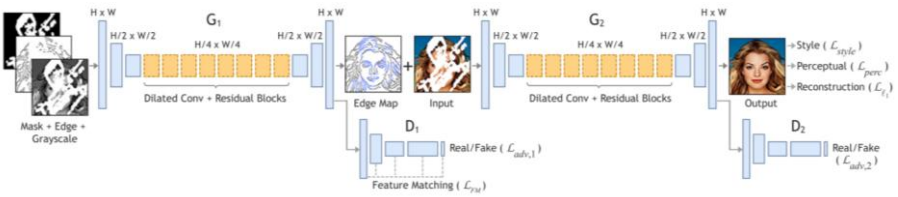

Figure 5 Image completion network flow chart of EdgeConnect project [3]

As shown in Figure 5, the image completion network uses the incomplete color image I_{masked} as input and completes it using the complete edge image derived above. The image completion network returns a color image I_{pred} with the same resolution as the input image, where the missing parts have been filled:

$$I_{pred} = G_2(I_{masked}, \hat{E}) \quad (5)$$

Among them, G_2represents the generator of the image completion network stage in the Edge-Connect project.

In the Edge-Connect project, perceptual loss (VGG extracted feature difference), style loss (Gram matrix), and adversarial loss (PatchGAN discriminator) were designed for the repaired images, which greatly improved the authenticity of the repaired facial images [1].

3 Experiment

3.1 Dataset

This study uses the CelebA dataset. The CelebFaces Attributes Dataset (CelebA) is a large-scale face image dataset proposed by the Multimedia Laboratory of the Chinese University of Hong Kong, China [4]. This dataset is widely used in computer vision tasks such as face recognition, face attribute prediction, face generation and restoration, and is regarded as an important benchmark dataset for face-related research due to its rich annotation information and diversity.

The CelebA dataset contains more than 200,000 color face images of different identities. The images are cropped and aligned, with the face-centered, and the resolution is 178 × 218. Each image is accompanied by 40 binary face attribute labels (such as gender, hairstyle, whether wearing glasses, smiling, etc.) and five facial landmark positions (eyes, nose tip, mouth corners), providing strong support for structured analysis and supervised learning.

This study selected images from the CelebA dataset as training and test samples, and combined irregular masking strategies (such as QD-IMD) to construct missing images for evaluating the performance of face image restoration methods, especially in terms of structure restoration and detail reconstruction.

3.2 Loss Function

In order to train the edge completion capability of the DexiNed model, this study adopted the full-image L1 loss function[6], with the aim of allowing the model to restore the complete edge structure even when faced with occluded images.

Let the predicted edge map be $\hat{E}$ and the pseudo label (complete edge map) be E_{gt}, both of which are of size H * W, then the full-image L1 loss is defined as:

$$L_{edge} = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W}\left|\hat{E}_{i,j} - E_{gt_{i,j}}\right| \quad (6)$$

Among them, $\hat{E}_{i,j}$ represents the value of the predicted edge map at pixel (i, j); $E_{gt_{i,j}}$represents the true edge value of the pseudo-label edge map at the same position; all pixels are included in the calculation (that is, the occlusion area and the visible area are not distinguished);

This paper chooses L1 because of its strong robustness and suitability for processing sparse and uneven feature images such as edges. Compared with adversarial loss or perceptual loss, L1 focuses more on the direct difference between image pixels, which is conducive to learning the accurate position of edge lines [7]. In order to enable the model to infer the edges of occluded areas and make training in occluded areas more stable, the full-image L1 loss provides a dense and stable supervision signal.

3.3 Comparison of this study and the original EdgeConnect

Table 1 Performance comparison table of this research solution and the original EdgeConnect

Item	Edge-Connect	Research plan for this project
Edge extraction	Canny + Edge Generator	DexiNed
Accuracy	Easy to break, high noise	Continuous, clear and detailed
Trainability	The generator needs to be trained separately	Only one unified model is trained
Training Method	Strong Supervision + Canny	Weak Supervision

Figure 6 shows the specific restoration results of this research scheme through the original image, mask image, predicted edge map, and restored image:

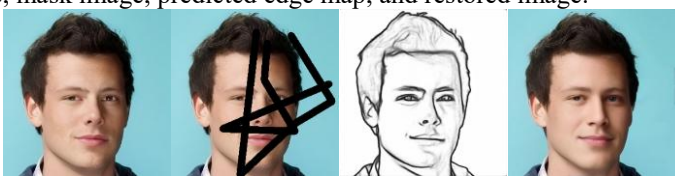

Figure 6 The original image, mask image, predicted edge map, and restored image of this research project (Picture credit: Original)

In order to verify the image inpainting performance of the proposed method, this paper conducted quantitative comparisons with several representative classic image inpainting methods, including the non-learning method PatchMatch [8], the early deep learning method Context Encoder (CE) [9], and subsequent improved models such as GL [10], PConv [11], and EdgeConnect. The experimental results are shown in the Table 2.

Table 2 Performance comparison table of various image restoration models

Model	PSNR	SSIM	LPIPS
PatchMatch	~23.5	0.83	>0.10
ContextEnc	24.2	0.87	0.08–0.10
GL	25.0	0.88	0.06–0.08
PConv	26.2	0.89	0.05–0.07
EdgeConnect	27.5	0.91	0.044
Research plan for this project	27.61	0.9112	0.0577

From the results, it can be seen that early methods (such as PatchMatch and Context Encoder) have obvious deficiencies in numerical indicators, especially in the perceptual indicator LPIPS, which makes it difficult to effectively restore complex texture or structural information. With the gradual optimization of the model structure, such as the introduction of local discriminators in GL and partial convolutions in PConv, its PSNR and SSIM have been improved, but LPIPS is still difficult to reach a better level.

The EdgeConnect method further improves the consistency of image structure by guiding edges, and has achieved good results in all three indicators, becoming a widely used baseline method in recent years. Based on this method, this study improves the edge generation module and effectively improves the overall restoration quality by introducing a more robust edge prediction method.

Specifically, this research scheme is slightly higher than EdgeConnect in PSNR and SSIM, reaching 27.61 dB and 0.9112 respectively, indicating higher consistency at the pixel and structural levels; although LPIPS is slightly higher than EdgeConnect (0.0577 vs. 0.044), it is still far better than most early methods. Overall, it retains a certain degree of perceptual naturalness while maintaining the integrity of structural information.

In summary, the method has achieved a comprehensive improvement over early classic models in image restoration tasks and demonstrated competitive performance in the edge-guided restoration framework.

4 Conclusions

Focusing on the task of face image restoration, this paper proposes an image restoration method based on the DexiNed edge detection and improved EdgeConnect framework. By introducing high-quality edge information as a structural prior, the model achieves good restoration results on the CelebA dataset, reaching or exceeding the classic methods in terms of PSNR, SSIM and LPIPS indicators.

However, this method still has some limitations when there are large missing areas. When the missing area exceeds 20%, DexiNed is limited by insufficient edge context information and has large errors when reasoning about the edges. In addition, the model has only been trained and tested on face images and has not yet been generalized to complex scenes such as objects or buildings.

In future work, can consider combining DexiNed with other edge detection methods such as HED to enhance edge reasoning capabilities and improve the processing effect of large-scale missing areas. At the same time, introduce more public data sets for training and testing to further verify the versatility and robustness of the model.

In summary, the method proposed in this paper provides a new idea for structure-driven image restoration, which has certain theoretical value and practical application prospects, especially in the fields of face editing and image content completion.

References

1. Nazeri, Kamyar, et al. "Edgeconnect: Generative image inpainting with adversarial edge learning." arXiv preprint arXiv:1901.00212 (2019).
2. Poma, Xavier Soria, Edgar Riba, and Angel Sappa. "Dense extreme inception network: Towards a robust cnn model for edge detection." Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2020.
3. Liu, Guilin, et al. "Image in-painting for irregular holes using partial convolutions." U.S. Patent Application No. 16/360,895.

4. Liu, Ziwei, et al. "Deep learning face attributes in the wild." Proceedings of the IEEE international conference on computer vision. 2015.
5. Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980 (2014).
6. Zhao, Hang, et al. "Loss functions for image restoration with neural networks." IEEE Transactions on computational imaging 3.1 (2016): 47-57.
7. Li, Chen, et al. "Image inpainting using two-stage loss function and global and local markovian discriminators." Sensors 20.21 (2020): 6193.
8. Barnes, Connelly, et al. "PatchMatch: A randomized correspondence algorithm for structural image editing." ACM Trans. Graph. 28.3 (2009): 24.
9. Pathak, Deepak, et al. "Context encoders: Feature learning by inpainting." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
10. Iizuka, Satoshi, Edgar Simo-Serra, and Hiroshi Ishikawa. "Globally and locally consistent image completion." ACM Transactions on Graphics (ToG) 36.4 (2017): 1-14.
11. Liu, Guilin, et al. "Image inpainting for irregular holes using partial convolutions." Proceedings of the European conference on computer vision (ECCV). 2018.