Edge-aware Context Encoder for Image Inpainting

Liang Liao, Ruimin Hu*, Jing Xiao, Zhongyuan Wang

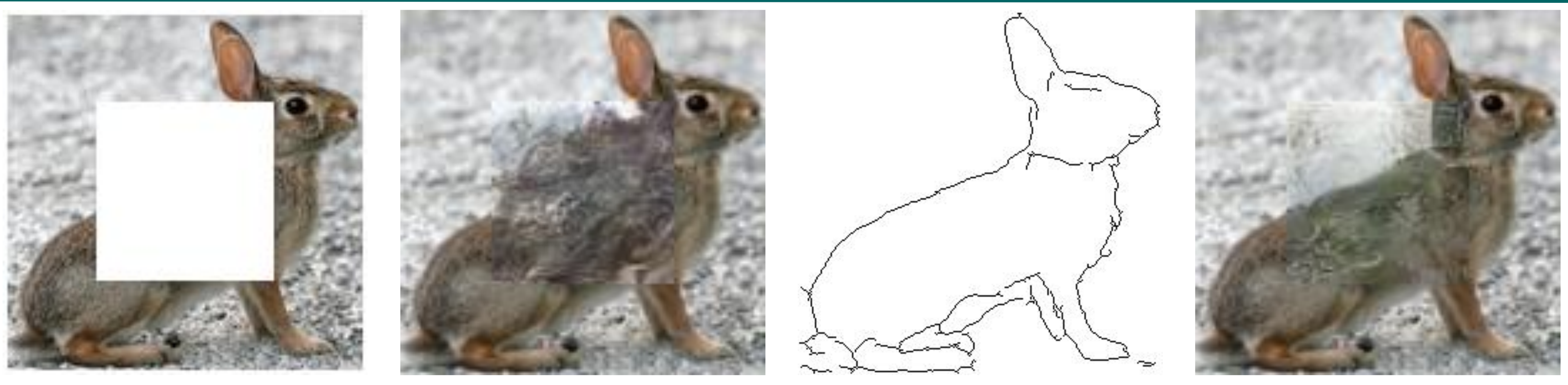
National Engineering Research Center for Multimedia Software, School of Computer, Wuhan University, Wuhan, 430072, China

Introduction

We present edge-aware context encoder (E-CE): an image inpainting model which takes scene structure and context into account. Unlike previous CE which predicts the missing regions using context from entire image, E-CE learns to recover the texture according to edge structures, attempting to avoid context blending across boundaries. In our approach, edges are extracted from the masked image, and completed by a full-convolutional network. The completed edge map together with the original masked image are then input into the modified CE network to predict the missing region. The experiments demonstrate that E-CE can generate images with better shapes and structures than CE. Our contributions are three-fold:

1) we demonstrate how structures can be employed for image inpainting and its effectiveness; 2) we develop a edge completion network to allow the prediction of context structures；3）we develop a inpainting network to fill in textures according the high level structure constraints and the low level texture constraints.

Edge-aware Image Inpainting Strategy



(a) (b) (c) (d)

Fig.1 Visual illustration of the task. (a) input masked image; (b) inpainting result from CE; (c) binary edges; (d) edge-aware inpainting result.

➢ **Problem analysis**

Context Encoder (CE) [1] has shown promise as prediction model of high-level context for completing images with large holes. While natural images contain composite structures and textures, and the textures are usually regions with homogenous patterns bounded by structures. Uniformed processing the results in clutter structures and mixing textures (Figure 1 (b)).

Structure has been proven to have high representation ability for the context. On one hand, compared to images with high frequency textures, the structure serves as a very good low dimensional embedding for scene context, which might be easier to complete than the original image. On the other hand, the structure is a good constraint for the propagation of textures. Moreover, in the creation of paintings, the artists normally start with sketch, and color the picture in the later step.

➢ **Generation of edge image**

Firstly, edges are firstly extracted using already trained HED model. Then we adopt standard non-maximum suppression and edge thinning for the post-processing.

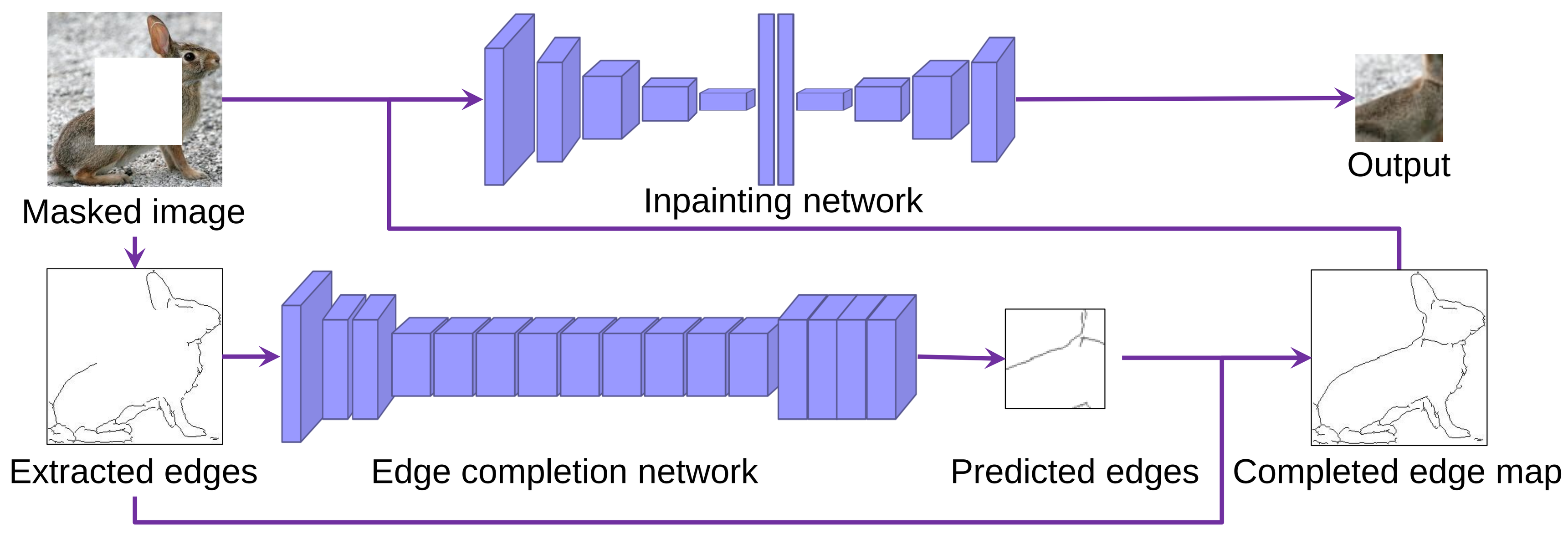


Fig.2 Framework of edge-aware context encoder for image inpainting.

Then, we adopted a fully convolutional network to recover the connectivity between edges. The network architecture follows an encoder-decoder structure, and decreases the resolution using strided convolutions to reduce the memory usage and computational time.

➢ **Inpainting network**

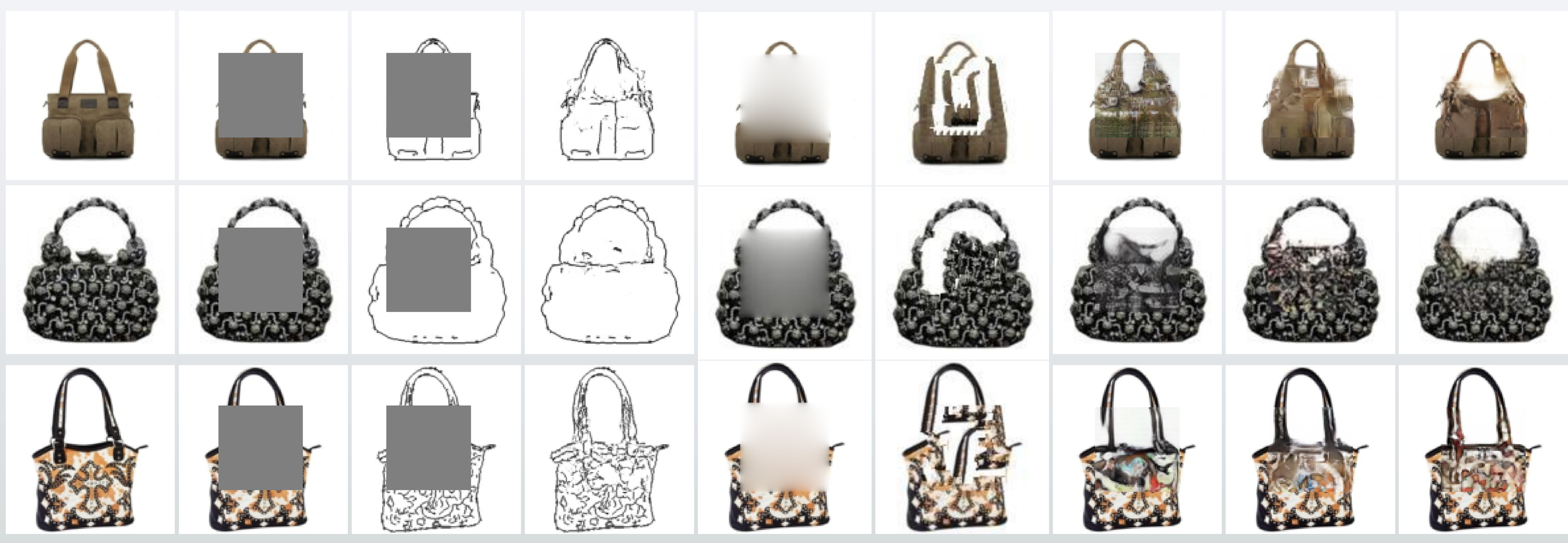
The inpainting network is based on the CE. The completed edge map, along with the masked input, is first mapped to hidden representations through the encoder. The decoder takes the latent feature representation to predict the missing region.

The inpainting network is trained by regressing to the ground truth content of the missing region. Reconstruction loss and adversarial loss are used together to handle the similarity in the overall structure and naturalness of the generated missing region. Rather than taking GAN loss, we use the Wasserstein GAN (WGAN) loss due to its ease of training and good results

Experimental Results

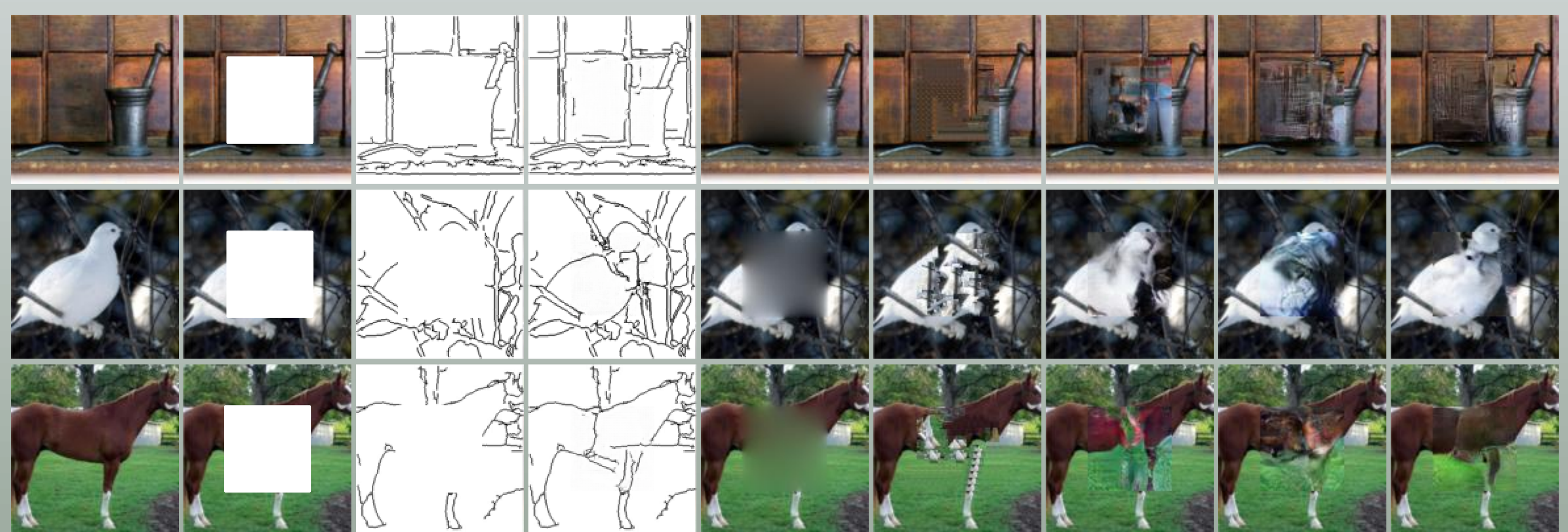
➢ **Datasets**. a) 137K Amazon Handbag images: contain simple and clear content and used to verify whether edges can formalize the structure of the completed objects. b) 100K-ImageNet dataset chosen at random in ImageNet: contain large variance in classes and background and used to test the overall performance of the proposed method.

➢ **Baselines**: a) local information-based approaches: the total variation approach (TV) [3] and the exemplar-based approach (EB) [2]. b) learning-based approaches: the CE and CE implemented with WGAN loss (CE-W).



(a) (b) (c) (d) (e) (f) (g) (h) (i)

Fig.3 Inpainting results for Handbag dataset. (a) original image; (b) masked image; (c) extracted edges; (d) completed edge map; (e) TV; (f) EB; (g) CE; (h) CE-W; (i) Our method.



(a) (b) (c) (d) (e) (f) (g) (h) (i)

Fig.4 Inpainting results for 100K-ImageNet dataset. (a) original image; (b) masked image; (c) extracted edges; (d) completed edge map; (e) TV; (f) EB; (g) CE; (h) CE-W; (i) Our method.

➢ **Results**: In general, classical methods (TV, EB) can only produce blurred or cracked results in the tasks of completing large missing region. Results from context based approaches are more semantically correct than classical methods.

a) Clear environment: our edge completion network is able to draw the boundary of the object (Fig. 3 (d)). We can also notice that the final completed images (Fig. 3. (i)) from our method can fit well with the edge maps, verified that the edge maps have strong effects on the texture prediction.

b) Complex environment: our proposed method can successfully avoid texture penetration of neighboring object (e.g. row 1-3 in Fig.4). We attribute it to the edge map, which recovered the boundary between objects in the edge completion step.

Table.1 Quantitative comparison on 100K-ImageNet dataset.

Method	Mean L1 loss	Mean L2 loss	PSNR	SSIM
CE	26.41%	14.91%	15.13 dB	0.495
CE-W	24.19%	13.16%	15.68 dB	0.515
Our method	**22.50%**	**12.19%**	**16.07** dB	**0.549**

Table 1 reports the quantitative results of completed region on 100K-ImageNet dataset, which reveals similar trend with the qualitative results. The PSNR value of our methods is about 0.39 dB higher than CE-W and about 0.94 dB higher than CE.

Conclusion

In this paper, we proposed an edge-aware image inpainting method to handle deformed shapes in the previous learning based approaches. Moreover, according to the characteristics of binary edges, we developed an edge completion network. Compared to CE, our method can obtain images with better structures and correctly located textures. Experimental results demonstrated its superior performance on challenging image inpainting examples.

[1] D. Pathak, P. Krähenbühl, J. Donahue, T. Darrell and A. A. Efros, "Context Encoders: Feature Learning by Inpainting," In *IEEE Conference on Computer Vision and Pattern Recognition*, pp: 2536-2544, 2016.
[2] A. Criminisi, P. Perez And K. Toyama, "Region Filling And Object Removal By Exemplar-Based Image Inpainting," In *IEEE Transactions on Image Processing*, Vol. 13, No. 9, pp. 1200-1212, 2004.
[3] T. F. Chan, J. Shen, "Mathematical Models for Local Nontexture Inpaintings," In *Siam Journal on Applied Mathematics*, Vol. 62, No. 9, pp. 1019-1043, 2001.