ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

Image Inpainting with Contrastive Relation Network

Xiaoqiang Zhou[1], Junjie Li[1], Zilei Wang[1*], Ran He[2,3], Tieniu Tan[2,3]

[1]University of Science and Technology of China
[2]Center for Research on Intelligent Perception and Computing, CASIA
[3]School of Artificial Intelligence, University of Chinese Academy of Sciences

Introduction

- Definition

  filling missing regions of an image with plausible contents

Learning Pyramid-Context Encoder Network for High-Quality Image Inpainting, CVPR 2019

Introduction

- Definition

  filling missing regions of an image with plausible contents

- Challenge
  - diverse scenes, arbitrary corrupted regions

Introduction

- Definition

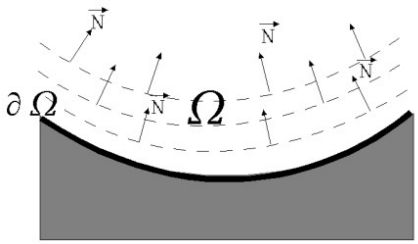
  filling missing regions of an image with plausible contents

- Challenge
  - diverse scenes, arbitrary corrupted regions
- Application
  - image restoration, object removal, privacy protection, entertainment

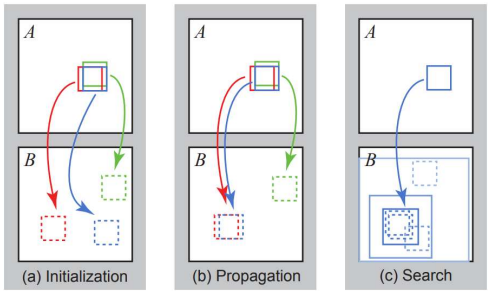
Related Work

- Diffusion-based or patch-based methods

diffusion-based [1]



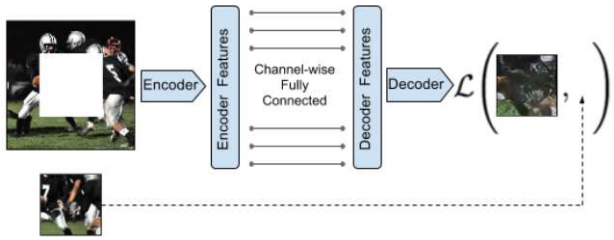
patch-based [2]



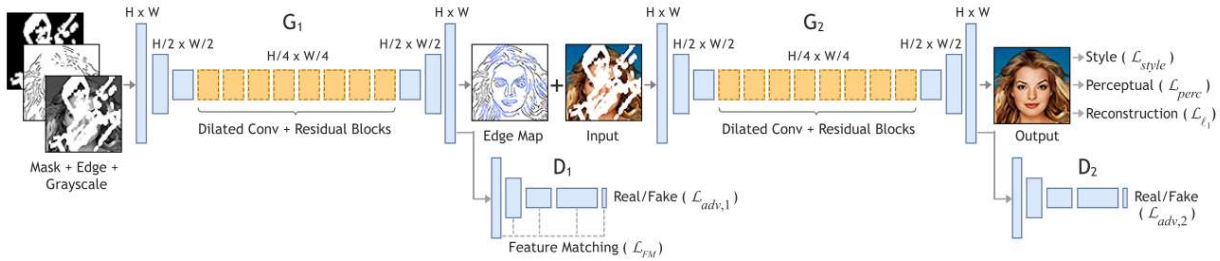
[1]: *Image inpainting, 2000*
[2]: *Patchmatch: A randomized correspondence algorithm for structural image editing, 2009*

Related Work

- Diffusion-based or patch-based methods
- Deep learning related methods



Context Encoder [3]



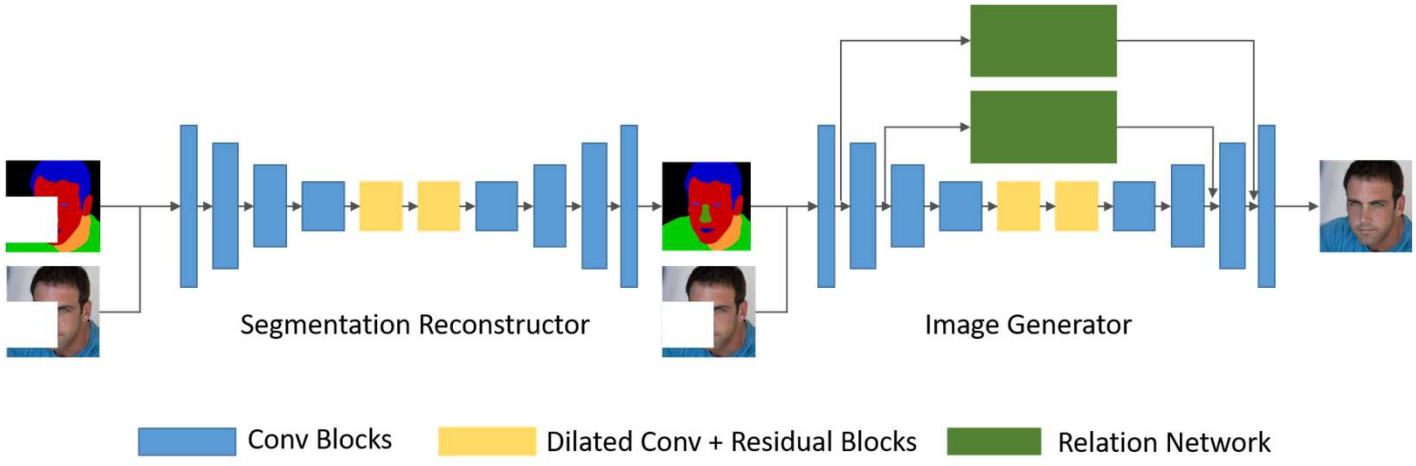
Edge Connect [4]

[3]: *Context encoders: Feature learning by inpainting, CVPR, 2016*
[4]: *Edgeconnect: Generative image inpainting with adversarial edge learning, arXiv 2019*

Motivation

- Complete corrupted segmentation map
  - Easier to complete
  - Provide a global structure guidance
- Exploit relationships under the guidance of segmentation map
  - intra-relationship for pixels in the same semantic region
  - inter-relationship between different regions

Approach

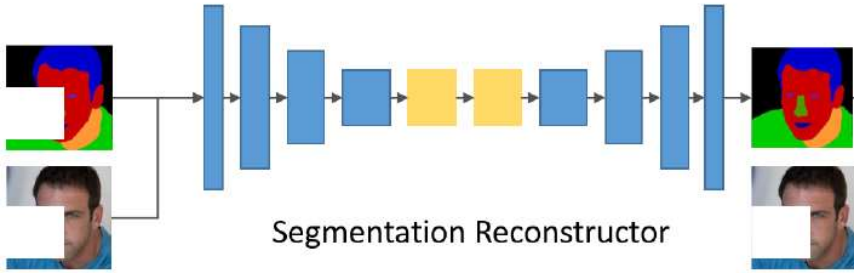


Overall Framework

Visual Illustration

Approach

- Segmentation Reconstructor



$$\mathrm{S_{init}} = UNet(\mathrm{I_{in}})$$

$$\mathrm{S_{in}} = \mathrm{S_{init}} \cdot (1 - \mathrm{M}) + \mathrm{P} * \mathrm{M}$$

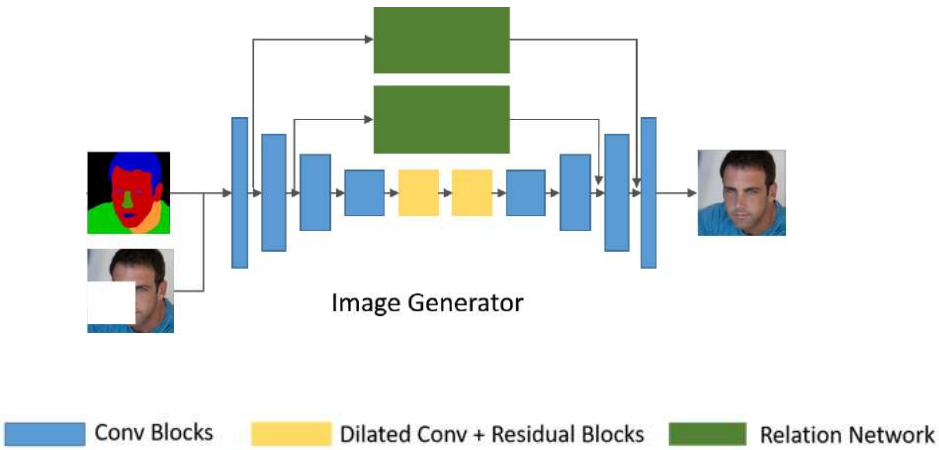
In M, 1 indicates the hole

$$\mathrm{S_{pred}} = G_s([\mathrm{S_{in}}, \mathrm{I_{in}}, \mathrm{M}])$$

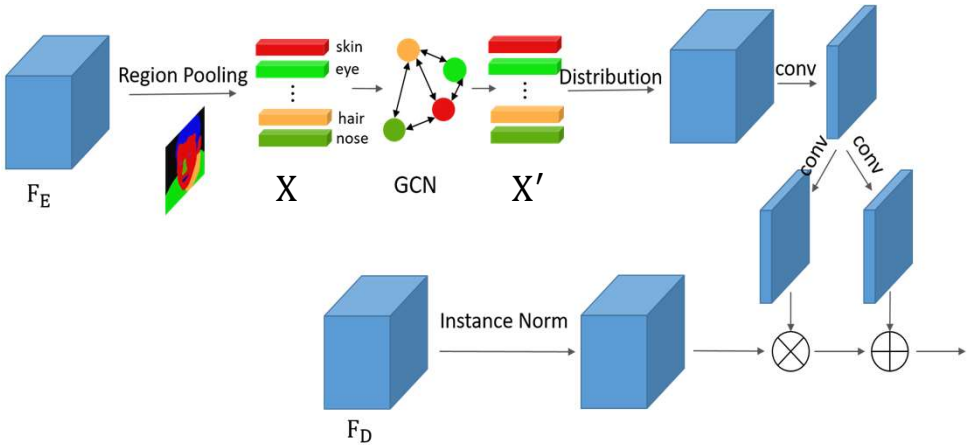
$$L_{parse} = CrossEntropy(\mathrm{S_{perd}}, \mathrm{S_{gt}})$$

Approach

- Image Generator



$$I_{pred} = G_I([S_{pred}, I_{in}, M])$$



$$R_k = S'^{k}_{pred} \cdot (1 - M')$$

$$N_k = \mathrm{AvgPool}(S), \text{where } S = \{F_E(i,j) | R_k(i,j) \neq 0\}$$

$$X' = \sigma\left(D^{-\frac{1}{2}}\widehat{A}D^{-\frac{1}{2}}XW\right), \text{ where } \widehat{A} = A + I$$

$$F'_D = \gamma \cdot \mathrm{IN}(F_D) + \beta$$

$$L_{contrast} = \sum_a \max(0, \min(d(a,P)) - \max(d(a,N)) + \delta)$$

P: samples with same groundtruth but different masks

Experiments

		PSNR	SSIM	FID
CelebA-HQ	GMCNN [3]	27.35	0.8839	8.81
	EdgeConnect [4]	26.60	0.8724	8.38
	GatedConv [8]	27.12	0.8814	7.74
	StructureFlow [9]	27.48	0.8885	7.21
	Ours	**28.61**	**0.9013**	**5.94**
DeepFashion	GMCNN [3]	22.40	0.8074	9.41
	EdgeConnect [4]	22.79	0.8162	10.56
	GatedConv [8]	23.32	0.8175	8.24
	StructureFlow [9]	23.20	0.8166	8.00
	Ours	**24.16**	**0.8197**	**7.37**

TABLE I

QUANTITATIVE COMPARISONS IN TERMS OF PSNR, SSIM AND FID OVER CELEBA-HQ AND DEEPFASHION DATASETS. BOLD TYPE INDICATES THE BEST PERFORMANCE

Comparisons with SOTA methods

	PSNR	SSIM	FID
w/o segmentation map	26.04	0.8594	10.12
w/o relation network	27.20	0.8814	8.64
w/o contrast loss	28.47	0.9004	6.23
Ours	**28.61**	**0.9013**	**5.94**

TABLE II

ABLATION STUDIES OF DIFFERENT COMPONENTS IN OUR METHOD OVER DATASET CELEBA-HQ.

Ablation studies

Experiments

Summary

1. A two-stage framework
2. Relation network: intra-relationship & inter-relationship
3. Contrastive loss.
4. Superiority of our method is demonstrated by qualitative results and quantitative comparisons

References

- [1] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, “Image inpainting,” in Proceedings of the Conference on Computer Graphics and Interactive Techniques, 2000, pp. 417–424.
- [2] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman, “Patchmatch: A randomized correspondence algorithm for structural image editing,” ACM Transactions on Graphics, vol. 28, no. 3, pp. 24–33, 2009.
- [3] 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, 2016, pp. 2536–2544.
- [4] K. Nazeri, E. Ng, T. Joseph, F. Z. Qureshi, and M. Ebrahimi, “Edgeconnect: Generative image inpainting with adversarial edge learning,” arXiv preprint arXiv:1901.00212, 2019.
- [5] 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, 2018, pp. 85–100.
- [6] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, “Generative image inpainting with contextual attention,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 5505–5514.
- [7] Y. Ren, X. Yu, R. Zhang, T. H. Li, S. Liu, and G. Li, “Structureflow: Image inpainting via structure-aware appearance flow,” in Proceedings of the IEEE International Conference on Computer Vision, 2019, pp. 181–190.
- [8] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, “Free-form image inpainting with gated convolution,” in Proceedings of the IEEE International Conference on Computer Vision, 2019, pp. 4471–4480

Thanks