Free-Form Image Inpainting with Gated Convolution

Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, Thomas Huang, 2019

Isa-Ali Kirca, Venkat Mohit Sornapudi, Marten Rozema, Juno Prent

UNIVERSITY
OF AMSTERDAM

Outline

- Introduction
- Previous work
- Key contributions
- Approach
- Results
- Strengths and weaknesses
- Applications
- Later work

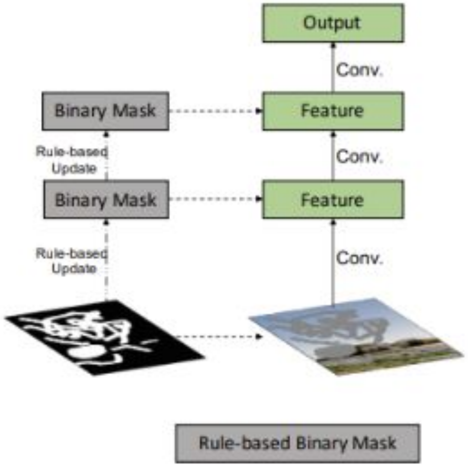
Introduction

- Image inpainting (image completion or image hole-filling)
    - Synthesizing alternative contents in missing regions

UNIVERSITY
OF AMSTERDAM

Previous work

- Patch matching
- Deep generative models based on vanilla convolutions are naturally ill-fitted for image hole-filling
  - Visual artifacts like color discrepancy, blurriness etc.
- Partial convolution
  - Ignores important information regarding spatial locations
  - No user guided image inpainting
  - Disappearance of “invalid” pixels layer by layer



$$O_{y,x} = \begin{cases} \sum\sum W \cdot (I \odot \frac{M}{sum(M)}), & \text{if sum(M)} > 0 \\ 0, & \text{otherwise} \end{cases}$$

$$m'_{y,x} = 1, \text{ iff } \text{sum(M)} > 0$$

References:

1. Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. Generative image inpainting with co textual attention. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5505– 5514,2018.
2. Guilin Liu, Fitsum A Reda, Kevin J Shih, Ting-Chun Wang, Andrew Tao, and Bryan Catanzaro. Image inpainting for irregular holes using partial convolutions. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 85–100, 2018.

Key contributions/solutions

1. Gated convolution to learn a dynamic feature selection mechanism
2. A more practical patch-based GAN discriminator (SN-PatchGAN)
3. Interactive inpainting model enabling user sketch as guidance
4. Higher quality free-form inpainting than previous state-of-the-arts

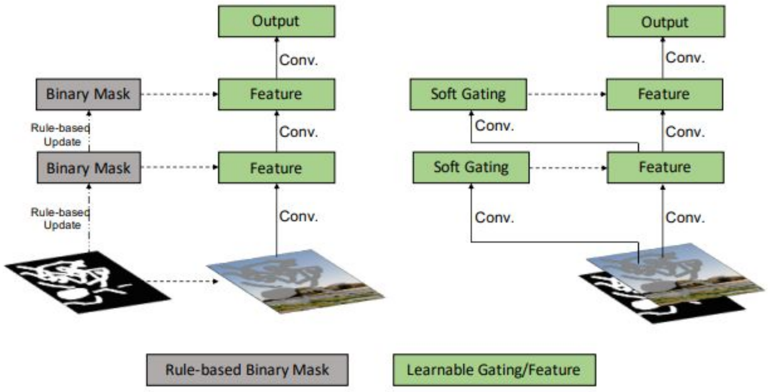
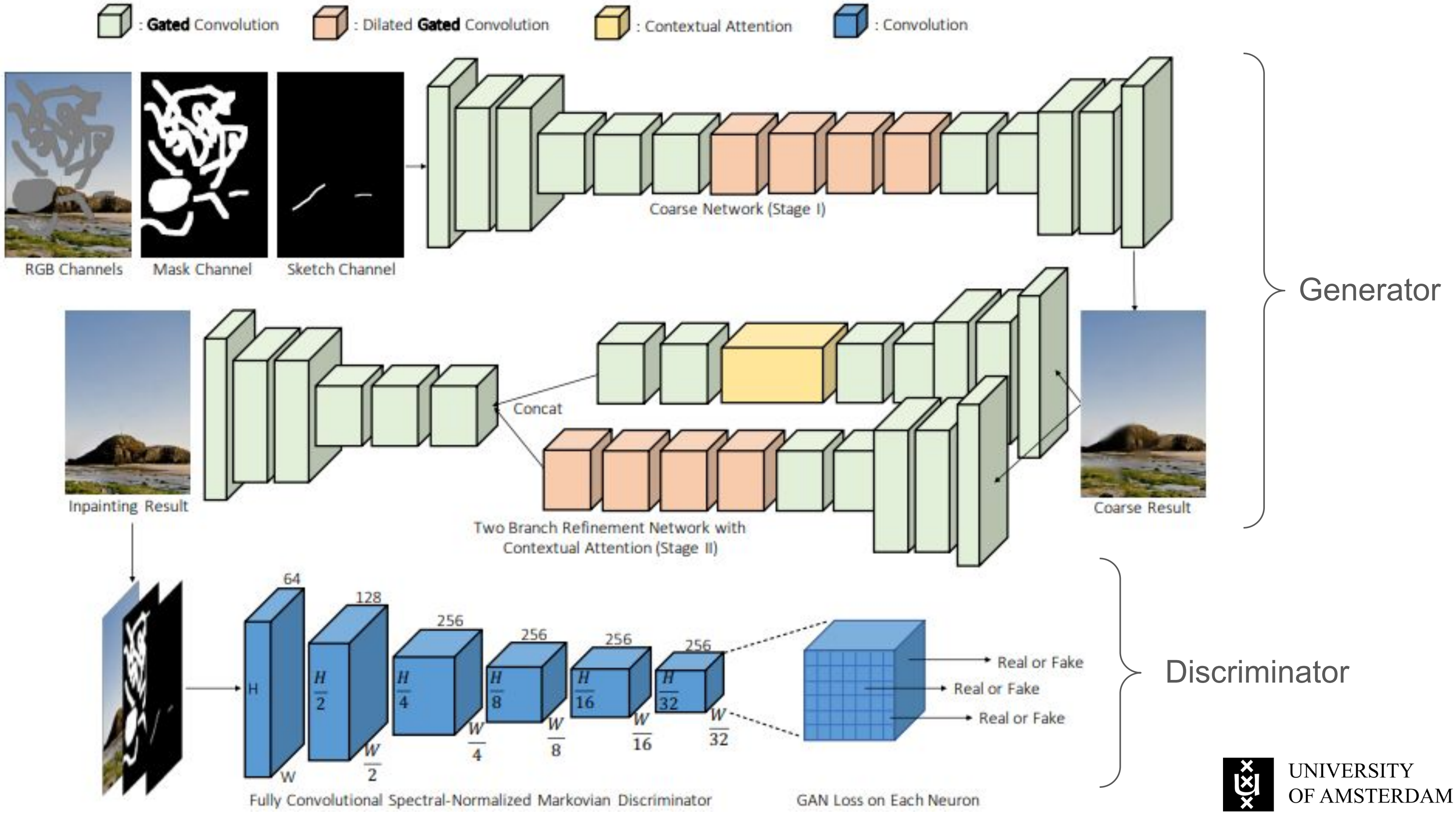


Figure 2: Illustration of partial convolution (left) and gated convolution (right).

$$Gating_{y,x} = \sum \sum W_g \cdot I$$
$$Feature_{y,x} = \sum \sum W_f \cdot I$$
$$O_{y,x} = \phi(Feature_{y,x}) \odot \sigma(Gating_{y,x})$$

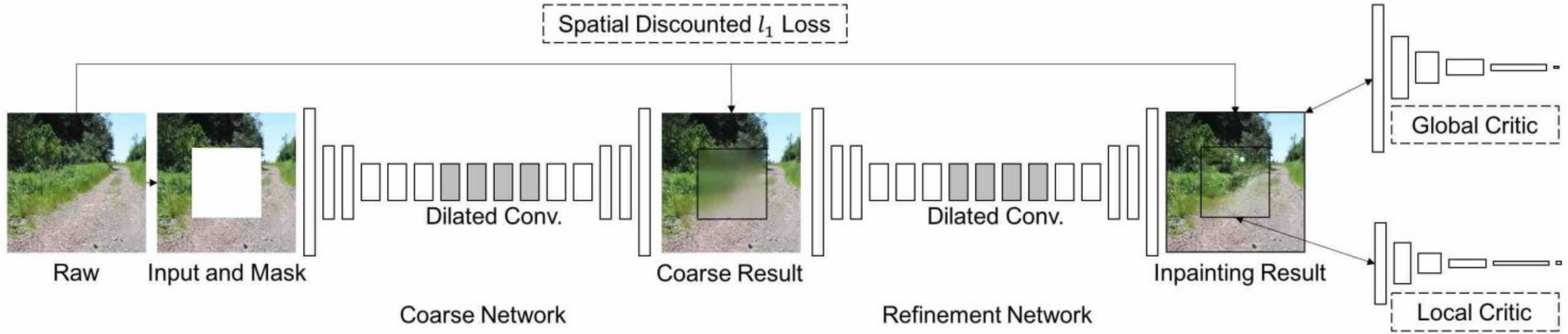


UNIVERSITY OF AMSTERDAM

GatedConv/ Deepfill v2: Inpainting Network

The model is a customized version of ContextAttention model proposed in

“*Generative Image Inpainting with Contextual Attention*” paper

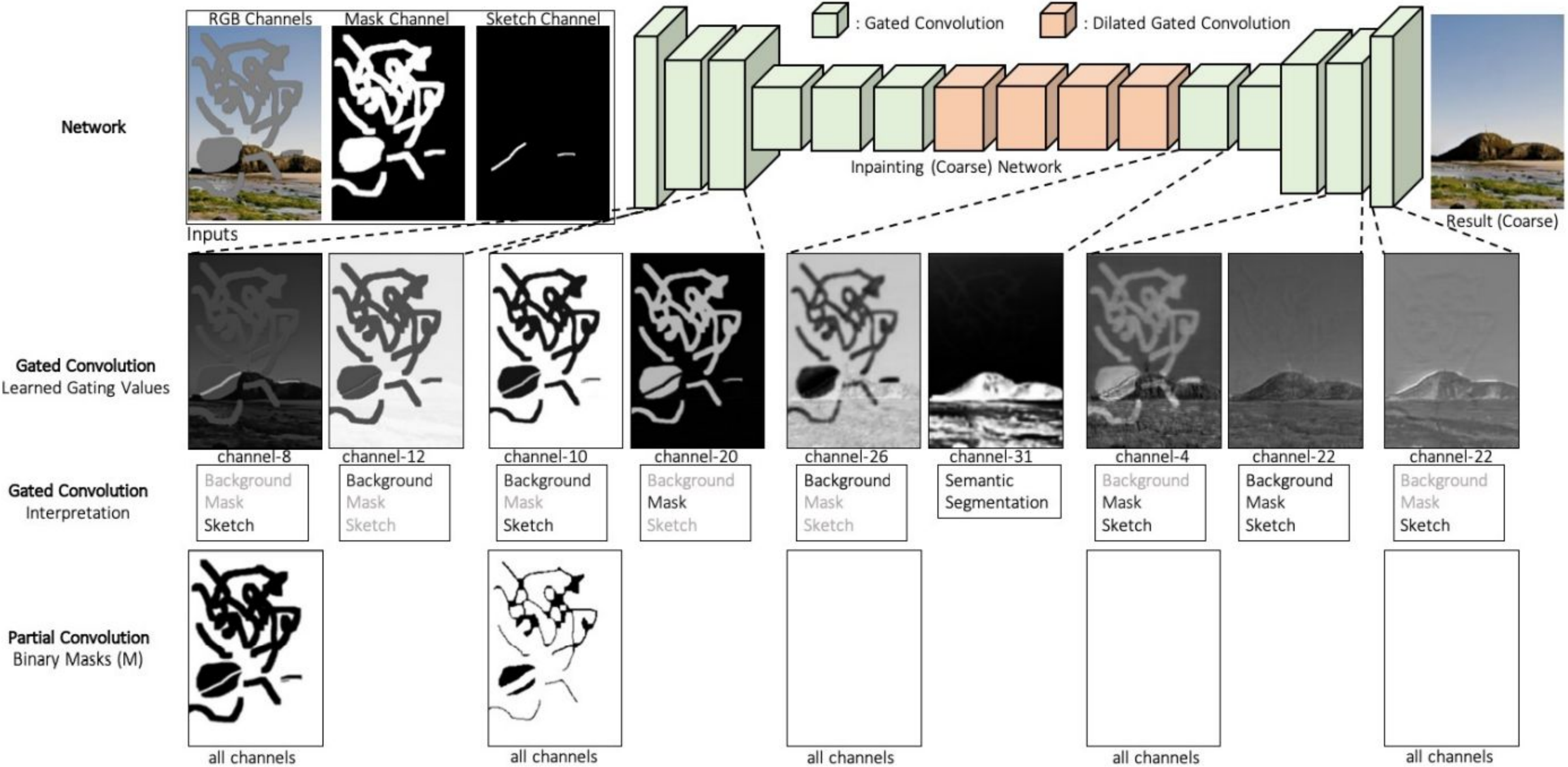


Courtesy: Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. Generative image inpainting with contextual attention. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 5505– 5514, 2018

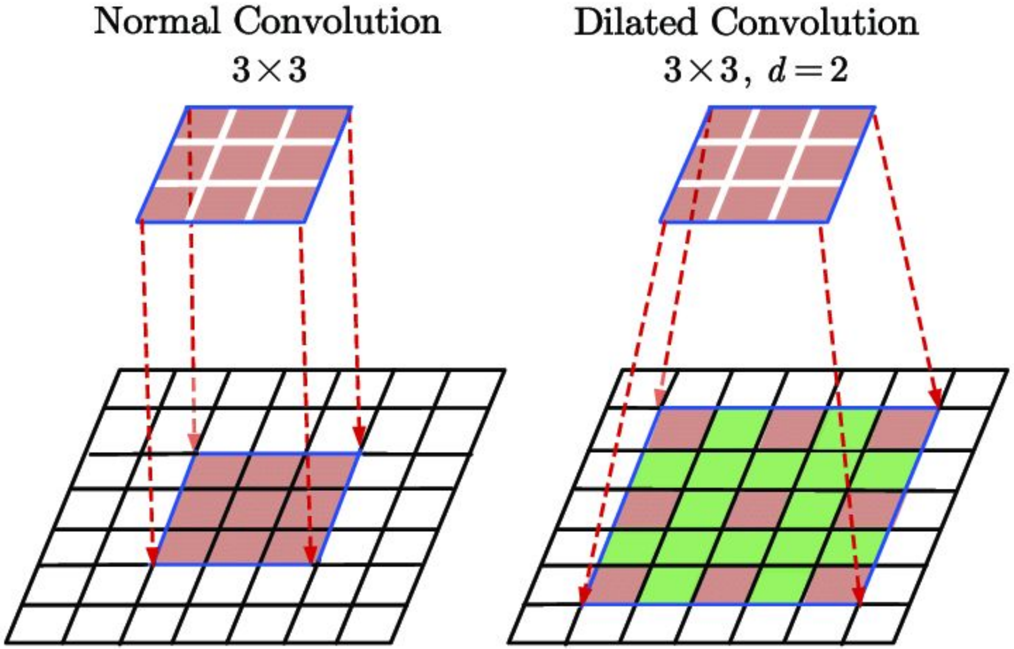
What’s new?? 1. Gated convolution 2. SN-PatchGAN loss (replacement)

UNIVERSITY OF AMSTERDAM

Coarse Network



Dilated convolution



Courtesy: Du, Jinglong & Wang, Lulu & Liu, Yulu & Zhou, Zexun & He, Zhongshi & Jia, Yuanyuan. (2020). Brain MRI Super-Resolution Using 3D Dilated Convolutional Encoder–Decoder Network. IEEE Access. PP. 1-1. 10.1109/ACCESS.2020.2968395.

Refinement Network

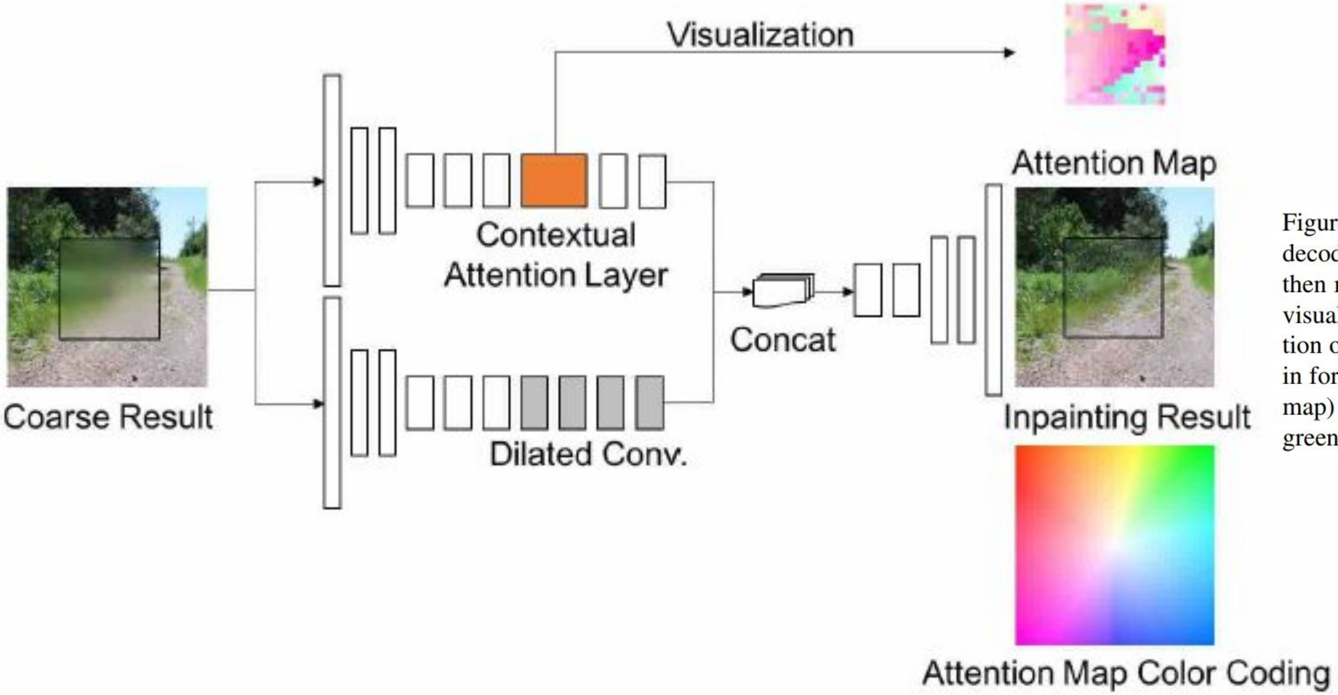


Figure 4: Based on coarse result from the first encoder-decoder network, two parallel encoders are introduced and then merged to single decoder to get inpainting result. For visualization of attention map, color indicates relative location of the most interested background patch for each pixel in foreground. For examples, white (center of color coding map) means the pixel attends on itself, pink on bottom-left, green means on top-right.

Courtesy: Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. Generative image inpainting with contextual attention. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 5505– 5514, 2018

Contextual attention layer

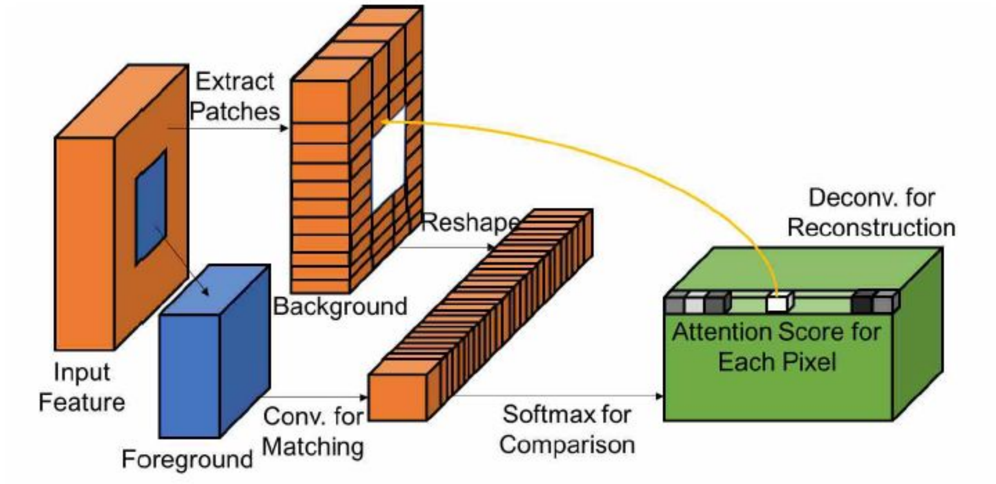


Figure 3: Illustration of the contextual attention layer. Firstly we use convolution to compute matching score of foreground patches with background patches (as convolutional filters). Then we apply softmax to compare and get attention score for each pixel. Finally we reconstruct foreground patches with background patches by performing deconvolution on attention score. The contextual attention layer is differentiable and fully-convolutional.

Courtesy: Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. Generative image inpainting with contextual attention. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 5505– 5514, 2018

Spectral-Normalized Markovian Discriminator (SN-PatchGAN)

Hinge loss as objective function for generator and discriminator:

$$\mathcal{L}_G = -\mathbb{E}_{z\sim\mathbb{P}_z(z)}[D^{sn}(G(z))]$$

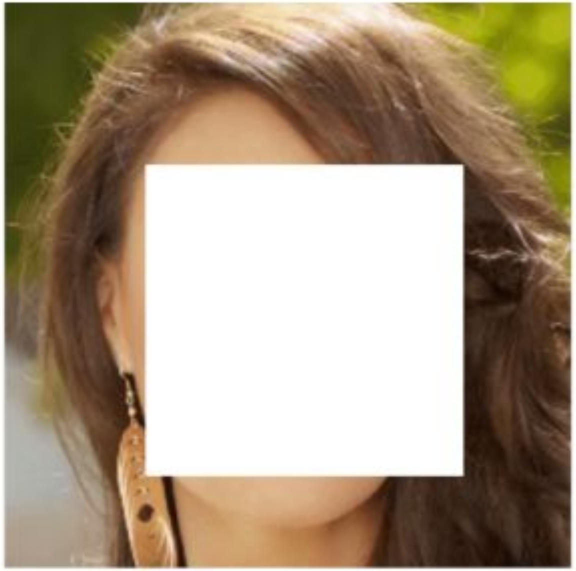
$$\mathcal{L}_{D^{sn}} = \mathbb{E}_{x\sim\mathbb{P}_{data}(x)}[ReLU(\mathbb{1} - D^{sn}(x))] + \mathbb{E}_{z\sim\mathbb{P}_z(z)}[ReLU(\mathbb{1} + D^{sn}(G(z)))]$$

where D^{sn} represents spectral-normalized discriminator (using default fast approximation algorithm of spectral normalization described in SN-GANs [1]), G is image inpainting network that takes incomplete image z

References:
1. Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. Spectral normalization for generative adversarial networks. arXiv preprint arXiv:1802.05957, 2018.

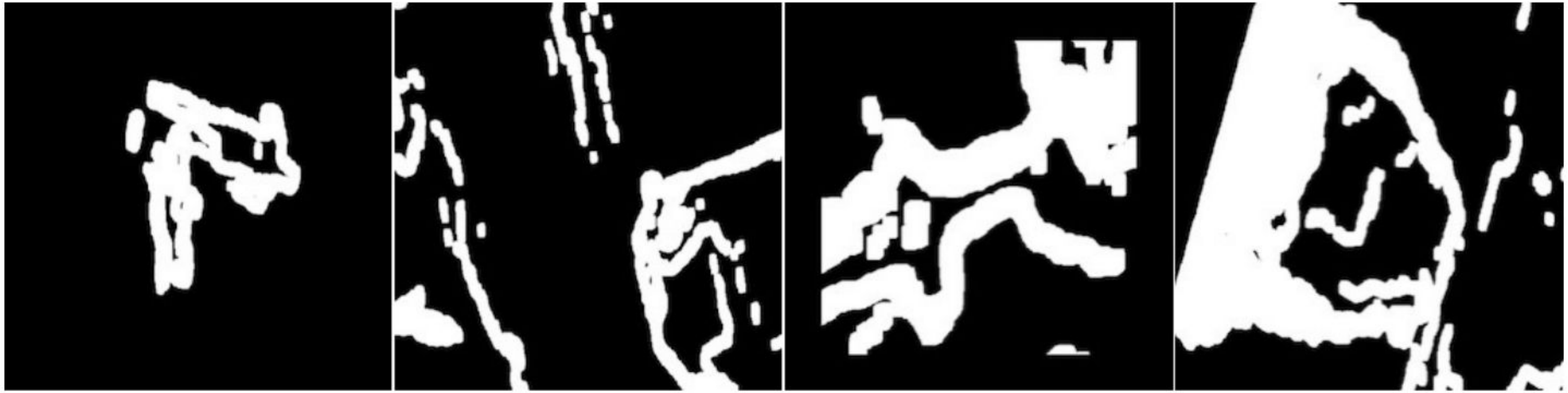
Free Form Mask Generation

- Similar to masks drawn in real use-cases
- Diverse to avoid over-fitting
- Efficient in computation and storage
- Controllable and flexible



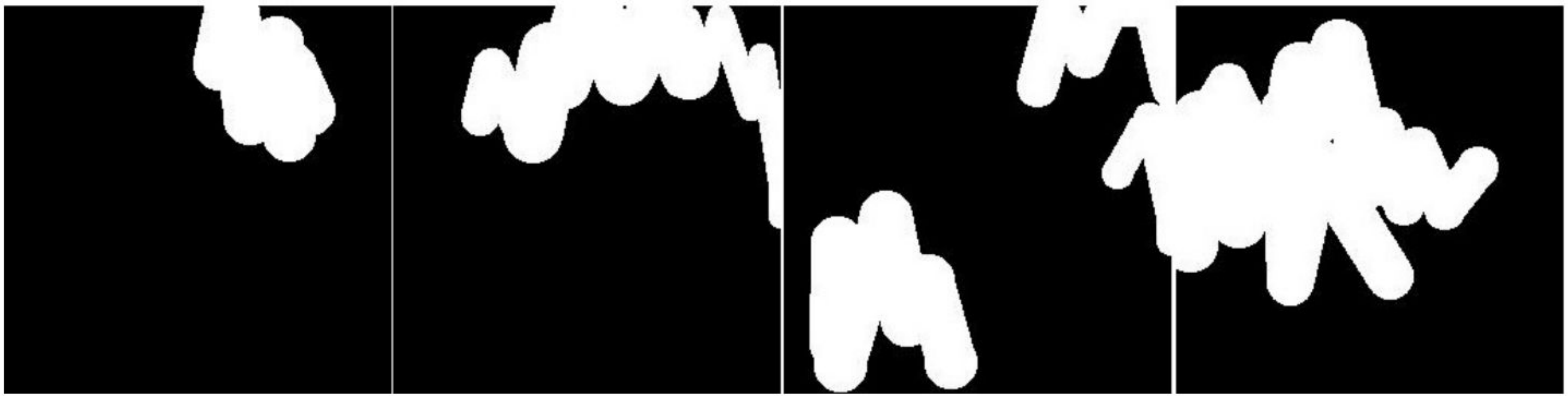
Free Form Mask Generation

PartialConv



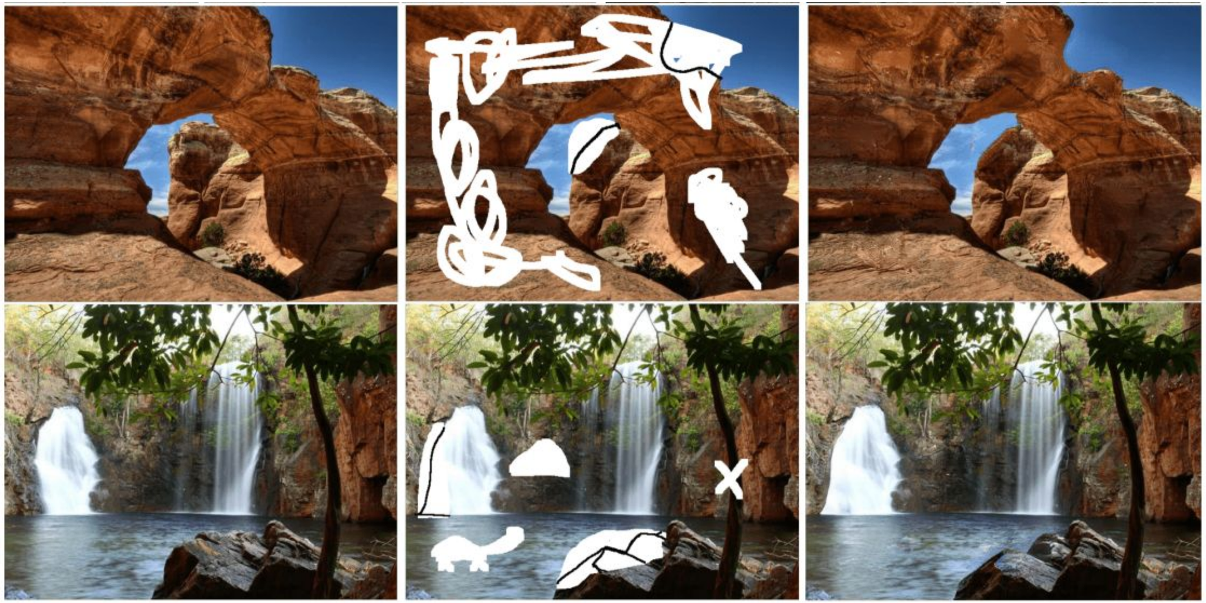
Free Form Mask Generation

DeepFill v2



Extension to User-Guided Inpainting

- Holistically-nested Edge Detector (HED)
- Faces and Landscapes



UNIVERSITY
OF AMSTERDAM

Extension to User-Guided Inpainting

Training and Testing

- Places2
- CelebA-HQ

UNIVERSITY
OF AMSTERDAM

Training and Testing

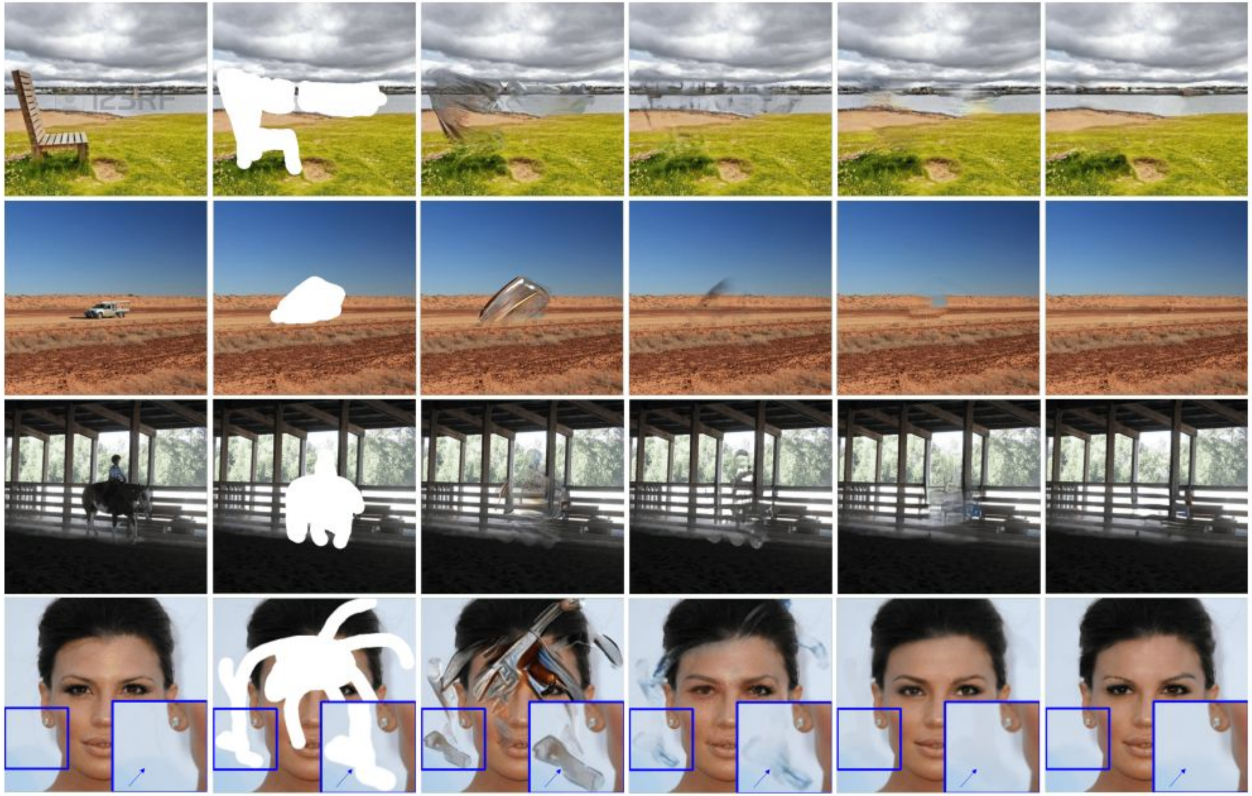
Testing on images of 512 x 512 resolution of Places2 validation set

- 0.21 seconds single NVIDIA Tesla V100 GPU
- 1.9 seconds on Intel(R) Xeon(R) CPU at 2 GHz

Quantitative Results

	rectangular mask		free-form mask	
Method	ℓ_1 err.	ℓ_2 err.	ℓ_1 err.	ℓ_2 err.
PatchMatch [3]	16.1%	3.9%	11.3%	2.4%
Global&Local [15]	9.3%	2.2%	21.6%	7.1%
ContextAttention [49]	**8.6%**	2.1%	17.2%	4.7%
PartialConv* [23]	9.8%	2.3%	10.4%	1.9%
Ours	**8.6%**	**2.0%**	**9.1%**	**1.6%**

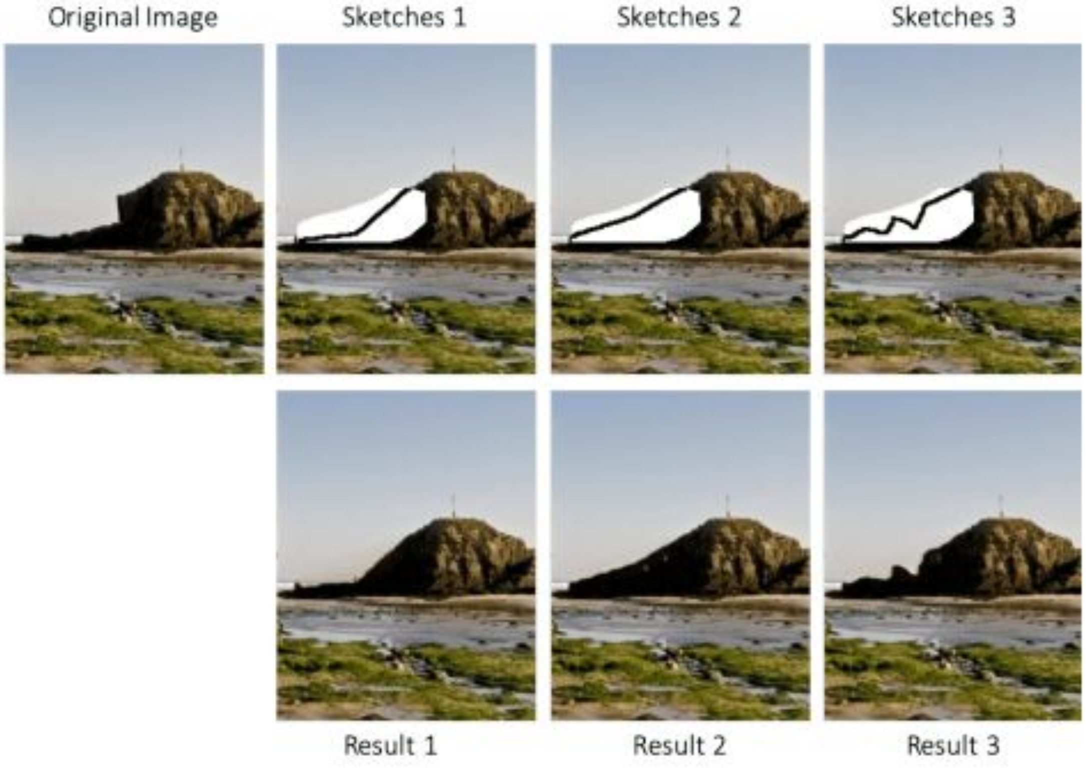
Qualitative Results



Original Image | Input | Global&Local | ContextAttention | Official PartialConv via Online Demo | GatedConv (Ours)

UNIVERSITY OF AMSTERDAM

Qualitative Results



UNIVERSITY OF AMSTERDAM

User study

- 30 images, 104 users
- Naturalness and quality:
  - Ground truth: 9.89
  - DeepFill v2: 7.72
  - re-imp. PartialConv: 7.07
  - PartialConv: 6.54
- Pairwise comparison with PartialConv: 79.4% prefers DeepFill v2

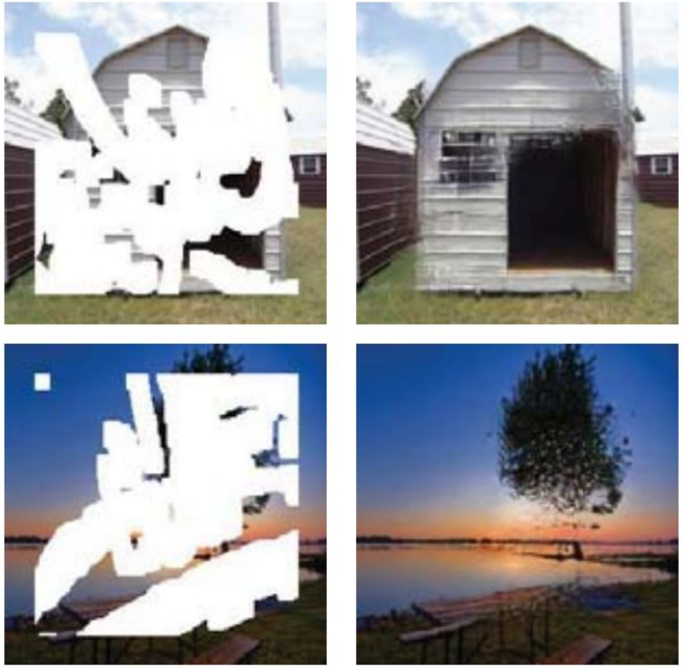
Ablation study

Figure 8: Ablation study of SN-PatchGAN. From left to right, we show original image, masked input, results with one global GAN and our results with SN-PatchGAN.

Strengths and Weaknesses

- Gated Convolution
- Sketch
- Stable training
- Less loss functions used

- Encoder-Decoder Structure
- Many Parameters
- Four networks needed for training



UNIVERSITY
OF AMSTERDAM

Applications

- Object Removal and Creative Editing
- Potential everyday use cases
- Able to manipulate images through user guidance
    - Works for either adding, removing or altering existing parts of an image

UNIVERSITY
OF AMSTERDAM

Building upon this paper

- Cited 853 times since being published in 2019
- Yang et al., “**Deep Face Video Inpainting via UV Mapping**”, 2021 (6 citations)
- Chung et al., “**Come-Closer-Diffuse-Faster: Accelerating Conditional Diffusion Models for Inverse Problems through Stochastic Contraction**” (6 citations)
- R. Suvorov et al., "**Resolution-robust Large Mask Inpainting with Fourier Convolutions**", 2022 (4 citations)

Resolution-robust Large Mask Inpainting with Fourier Convolutions

- Suvorov et al., 2022
- Improvement upon main weakness of the DeepFill v2 performance
- Uses fast Fourier convolutions to more accurately fill in large masks

Courtesy: R. Suvorov *et al.*, "Resolution-robust Large Mask Inpainting with Fourier Convolutions," *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2022, pp. 3172-3182, doi: 10.1109/WACV51458.2022.00323.

UNIVERSITY OF AMSTERDAM

Resolution-robust Large Mask Inpainting with Fourier Convolutions

Method	# Params $\times 10^6$	Places (512 × 512)						CelebA-HQ (256 × 256)			
		Narrow masks		Wide masks		Segm. masks		Narrow masks		Wide masks	
		FID ↓	LPIPS ↓	FID ↓	LPIPS ↓	FID ↓	LPIPS ↓	FID ↓	LPIPS ↓	FID ↓	LPIPS ↓
LaMa-Fourier (ours)	27	0.63	0.090	2.21	0.135	5.35	0.058	7.26	0.085	6.96	0.098
CoModGAN [64]	109▲	0.82▲30%	0.111▲23%	1.82▼18%	0.147▲9%	6.40▲20%	0.066▲14%	16.8▲131%	0.079▼7%	24.4▲250%	0.102▲4%
MADF [67]	85▲	0.57▼10%	0.085▼5%	3.76▲70%	0.139▲3%	6.51▲22%	0.061▲5%	—	—	—	—
AOT GAN [60]	15▼	0.79▲25%	0.091▲1%	5.94▲169%	0.149▲11%	7.34▲37%	0.063▲10%	6.67▼8%	0.081▼4%	10.3▲48%	0.118▲20%
GCPR [17]	30▲	2.93▲363%	0.143▲59%	6.54▲196%	0.161▲19%	9.20▲72%	0.073▲27%	—	—	—	—
HiFill [54]	3▼	9.24▲1361%	0.218▲142%	12.8▲479%	0.180▲34%	12.7▲137%	0.085▲49%	—	—	—	—
RegionWise [30]	47▲	0.90▲42%	0.102▲14%	4.75▲115%	0.149▲11%	7.58▲42%	0.066▲14%	11.1▲53%	0.124▲46%	8.54▲23%	0.121▲23%
DeepFill v2 [57]	4▼	1.06▲68%	0.104▲16%	5.20▲135%	0.155▲15%	9.17▲71%	0.068▲18%	12.5▲73%	0.130▲53%	11.2▲61%	0.126▲28%
EdgeConnect [32]	22▼	1.33▲110%	0.111▲23%	8.37▲279%	0.160▲19%	9.44▲76%	0.073▲27%	9.61▲32%	0.099▲17%	9.02▲30%	0.120▲22%
RegionNorm [58]	12▼	2.13▲236%	0.120▲33%	15.7▲613%	0.176▲31%	13.7▲156%	0.082▲42%	—	—	—	—

MADF: Mask-Aware Dynamic Filtering

CoModGAN: Co-modulated GAN

Courtesy: R. Suvorov *et al.*, "Resolution-robust Large Mask Inpainting with Fourier Convolutions," *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2022, pp. 3172-3182, doi: 10.1109/WACV51458.2022.00323.

Questions?