Image Inpainting via Generative Multi-column Convolutional Neural Networks

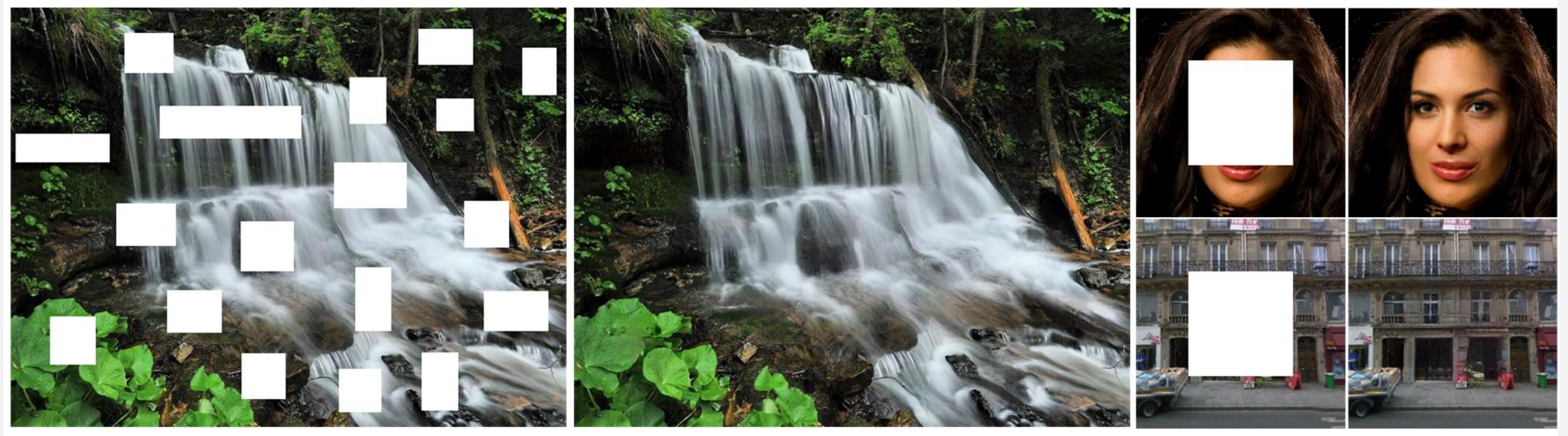
Yi Wang[1] Xin Tao[1,2] Xiaojuan Qi[1] Xiaoyong Shen[2] Jiaya Jia[1,2]

[1]The Chinese University of Hong Kong [2]YouTu Lab, Tencent

Introduction

Target

- Estimating suitable pixel information to fill holes in images.



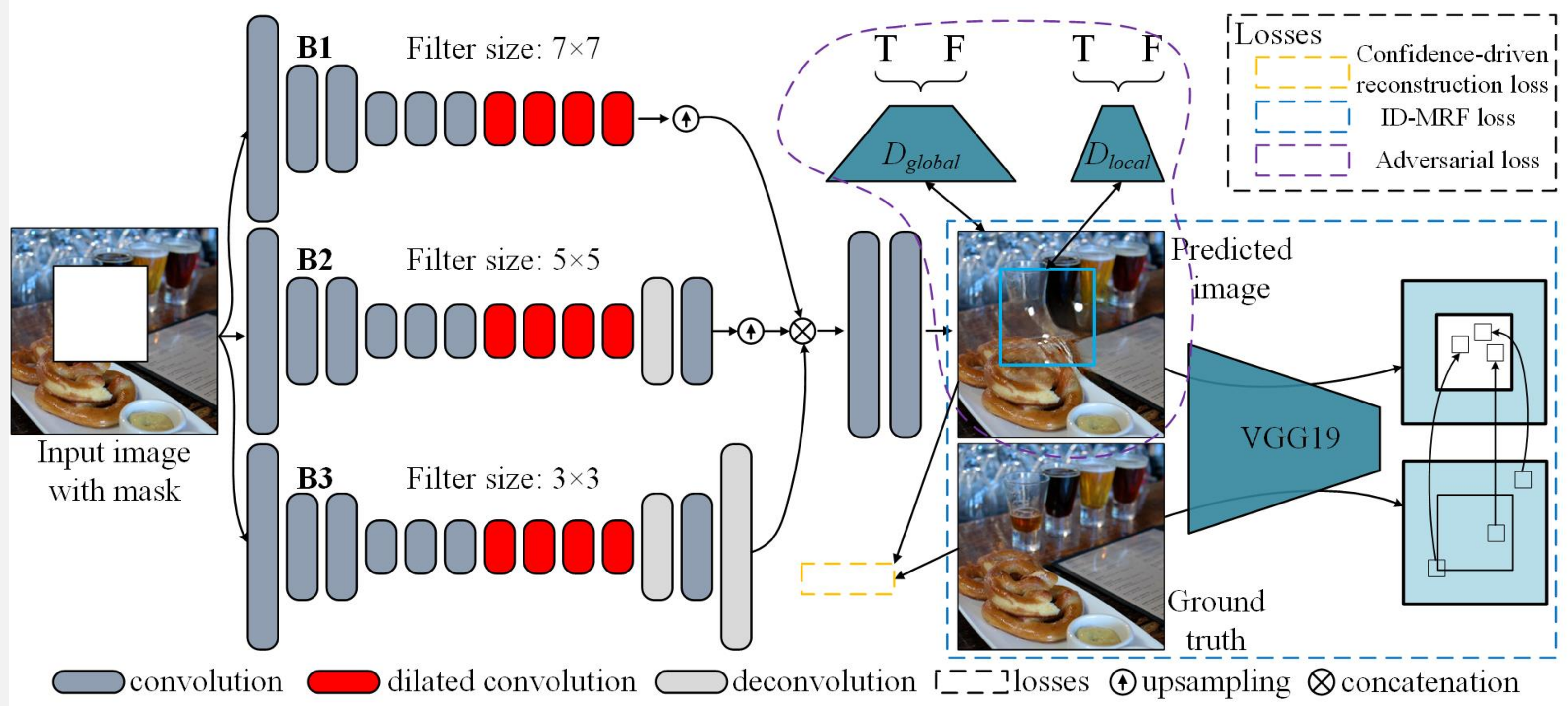
Motivations

A typical inpainting method focuses on

- *Features for generation*: to consider features as a group of different components and combine both *global semantics* and *local textures*.
- *Reliable similar patches*: using neighboring patches to regularize the generated ones.
- *Spatial-variant constraints*: pixels close to area boundary are with few choices, while the central part can be less constrained.

Our Model

Overall framework



Our Model

ID-MRF regularization

To take MRF-like regularization only in the training phase. For neural patches $\mathbf{v}$ and $\mathbf{s}$ extracted from $\hat{\mathbf{Y}}_g^L$ and $\mathbf{Y}^L$ respectively, their *relative similarity* is: $\mathrm{RS}(\mathbf{v},\mathbf{s}) = \exp\left(\frac{\mu(\mathbf{v},\mathbf{s})}{h\times(\max_{\mathbf{r}\in\rho_{\mathbf{v}}(\mathbf{Y}^L)}\mu(\mathbf{v},\mathbf{r})+\epsilon)}\right)$, and it is further normalized as:

$$\overline{\mathrm{RS}}(\mathbf{v},\mathbf{s}) = \frac{\mathrm{RS}(\mathbf{v},\mathbf{s})}{\sum_{\mathbf{r}\in\rho_{\mathbf{v}}(\mathbf{Y}^L)}\mathrm{RS}(\mathbf{v},\mathbf{r})}.$$

The final ID-MRF loss:

$$\mathcal{L}_{mrf} = \mathcal{L}_M(conv4\_2) + \sum_{\mathbf{t}=3}^{4} conv\mathbf{t}\_2,$$

where $\mathcal{L}_M(L) = -\log\left(\frac{1}{Z}\sum_{\mathbf{s}\in\mathbf{Y}^L}\max_{\mathbf{v}\in\hat{\mathbf{Y}}_g^L}\overline{\mathrm{RS}}(\mathbf{v},\mathbf{s})\right)$.

Spatial variant reconstruction loss

To propagate the confidence of known pixels to unknown ones,

$$\mathbf{M}_w^i = \left(g * \overline{\mathbf{M}}^i\right) \odot \mathbf{M},$$

where $\overline{\mathbf{M}}^i = \mathbf{1} - \mathbf{M} + \mathbf{M}_w^{i-1}$ and $\mathbf{M}_w^0 = \mathbf{0}$.

Our confidence-driven loss is:

$$\mathcal{L}_c = ||(\mathbf{Y} - G([\mathbf{X},\mathbf{M}];\theta)) \odot \mathbf{M}_w||_1$$

Adversarial loss

$$\mathcal{L}_{adv} = -E_{\mathbf{X}\sim\mathbb{P}_{\mathbf{X}}}\left[D\left(G(\mathbf{X};\theta)\right)\right]$$
$$+\lambda_{gp}E_{\hat{\mathbf{X}}\sim\mathbb{P}_{\hat{\mathbf{X}}}}[(||\triangle_{\hat{\mathbf{X}}} D(\hat{\mathbf{X}}) \odot \mathbf{M}_w||_2 - 1)^2]$$

Experiments

User studies

	Paris street view	ImageNet	Places2	CelebA	CelebA-HQ
GMCNN > CE	98.1%	88.3%	-	-	-
GMCNN > MSNPS	94.4%	86.5%	-	-	-
GMCNN > CA	84.2%	78.5%	69.6%	99.0%	93.8%

GMCNN: generative multi-column convolutional neural network; **CE**: context encoder; **MSNPS**: multi-scale neural patch synthesis; **CA**: contextual attention.

Quantitative Evaluation