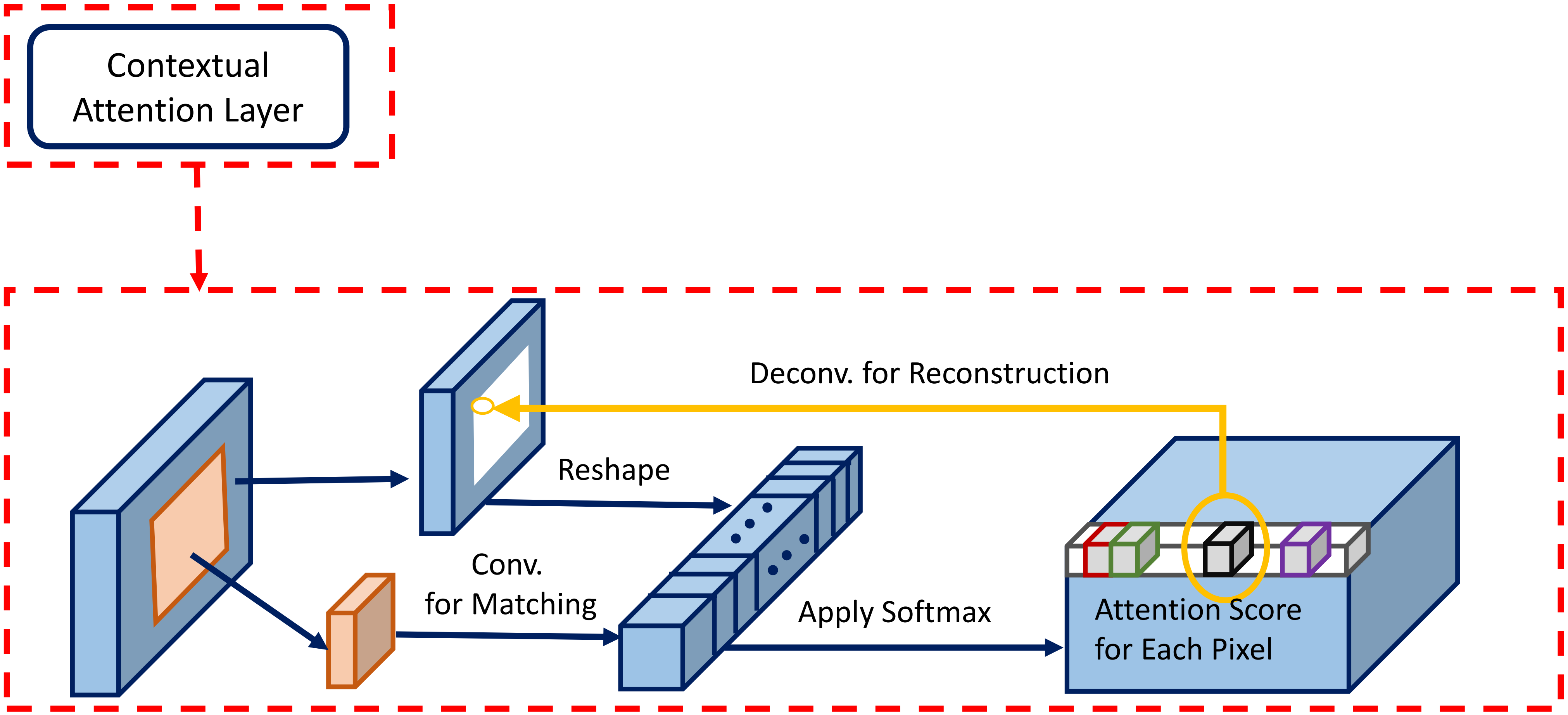
Generative Image Inpainting with Contextual Attention

7107053114 沈佳詠

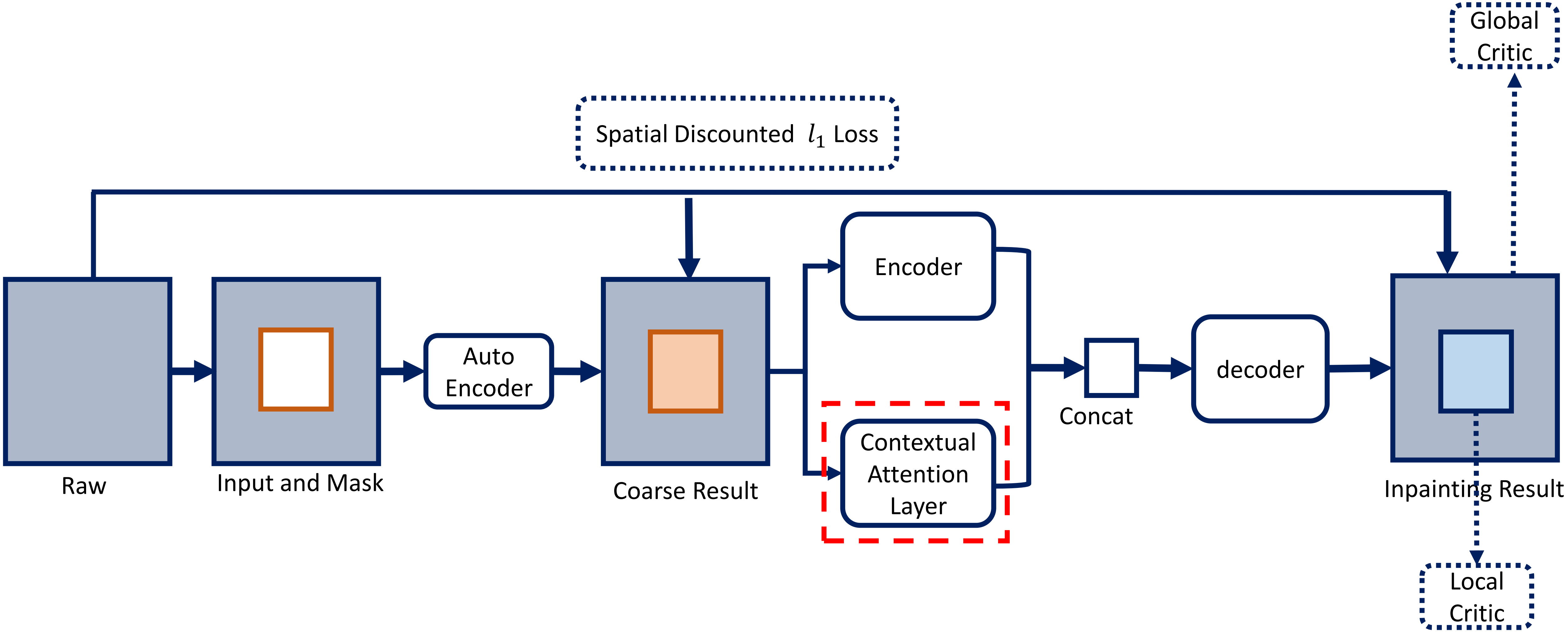
Abstract

- **Approach**： a new deep generative model-based approach
    - Filling missing pixels of an image
    - Utilize surrounding image features as references during network training
- **Model**： a feed-forward fully convolutional neural network
    - Process image with multiple holes at arbitrary locations and with variable sizes

Contextual Attention Layer



Improved generative inpainting Network



Parameter

- x = real image
- z = input image, $x \odot m$
- $\tilde{x}$ = inpainting output, $G(\mathrm{z})$
- $\hat{x} = (1 - \mathrm{t})x + \mathrm{t}\tilde{x}$, $\mathrm{t} \sim \mathrm{U}[0, 1]$
- G ： generator
- D ： discriminator
- $\mathcal{D}$ ： the set of 1-Lipschitz function
- P_r ： the model distribution defined by x
- P_g ： the model distribution implicitly defined by $\tilde{x}$

Parameter

- λ ： set to 10
- m ： input and mask

$$\begin{cases} 0, & \text{for missing pixels} \\ 1, & \text{for elsewhere} \end{cases}$$

- $P_{\hat{x}}$ ： the model distribution defined by $\hat{x}$
- $\nabla_{\hat{x}} D(\hat{x})$ ： the gradient penalty apply to pixels inside the holes

Improved WGAN

- $\min_{G} \max_{D \in \mathcal{D}} E_{x \sim P_r}[D(x)] - E_{\tilde{x} \sim P_g}[D(\tilde{x})]$

$$+\lambda E_{\hat{x} \sim P_{\hat{x}}}(\|\nabla_{\hat{x}} D(\hat{x}) \odot (1-m)\|_2 - 1)^2$$

- m : $\begin{cases} 0, \text{for missing pixels} \longrightarrow Local\ Critic \\ 1, \text{for elsewhere} \longrightarrow Global\ Critic \end{cases}$

Results

Places2

Method	l_1 Loss	l_2 Loss	PSNR	TV Loss
Patch Match	16.1%	3.9%	16.62	**25.0%**
Baseline model	9.4%	2.4%	18.15	25.7%
Our method	**8.6%**	**2.1%**	18.91	25.3%