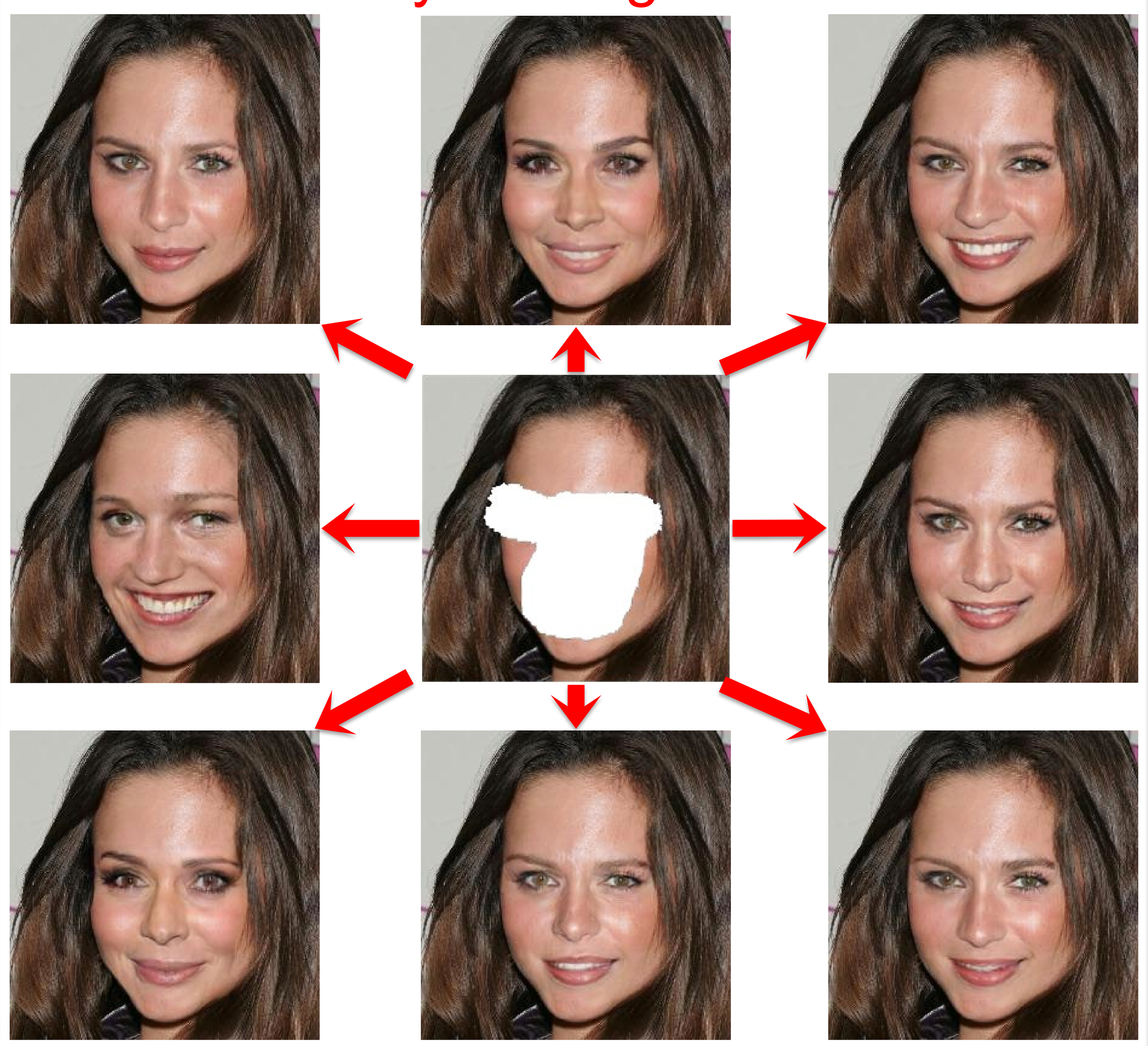
Pluralistic Image Completion

Chuanxia Zheng, Tat-Jen Cham, Jianfei Cai
School of Computer Science and Engineering
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

CVPR LONG BEACH CALIFORNIA June 16-20, 2019

Motivation

What would you imagine to be filled?

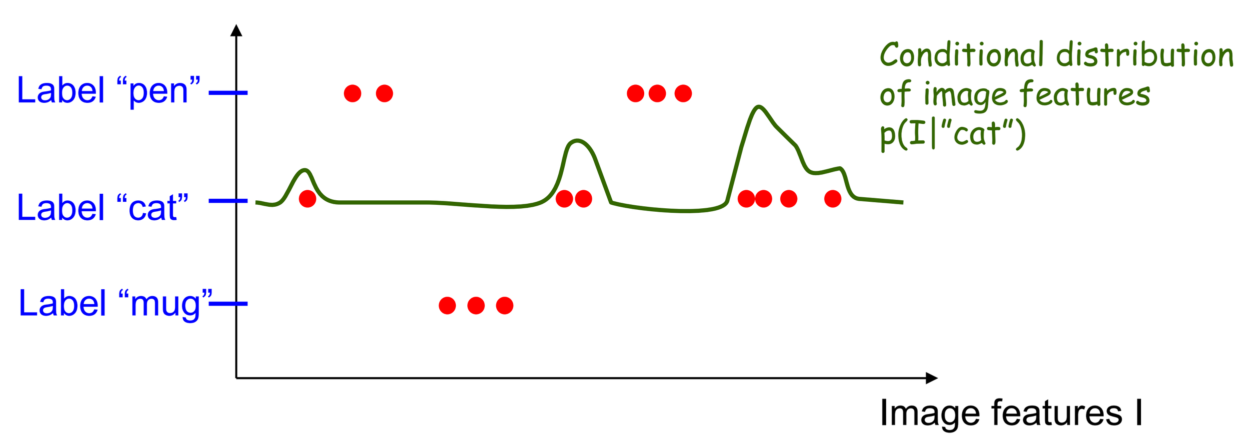


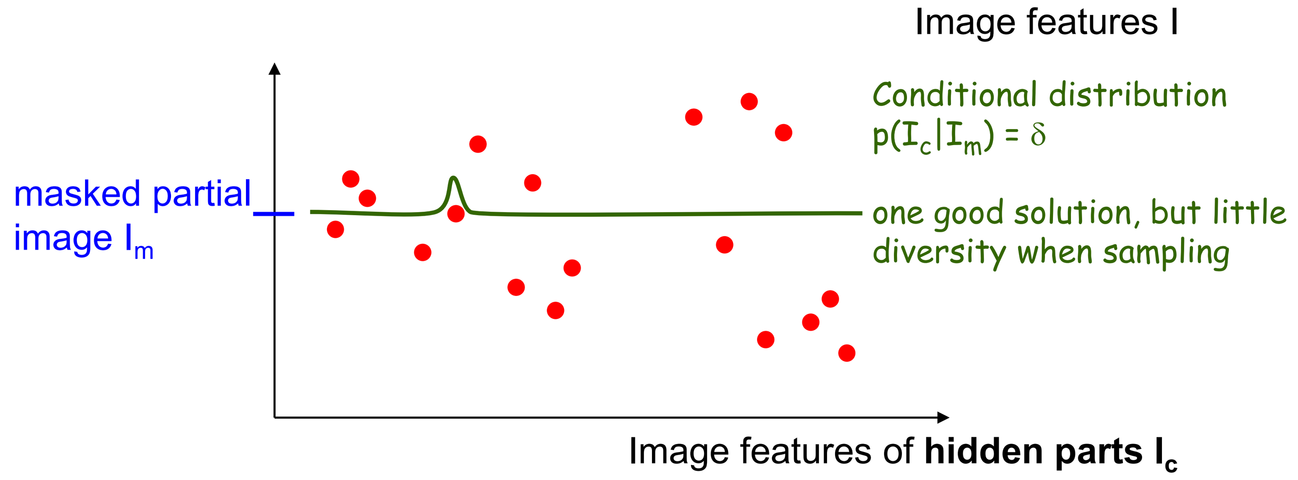
Problem:

1. Only one "optimal" result is typically generated in existing image completion work
2. Most methods focus on reconstructing the original image during the training.

Goal: *multiple* and *diverse* plausible results

Challenge: only one "ground truth" is available during the training

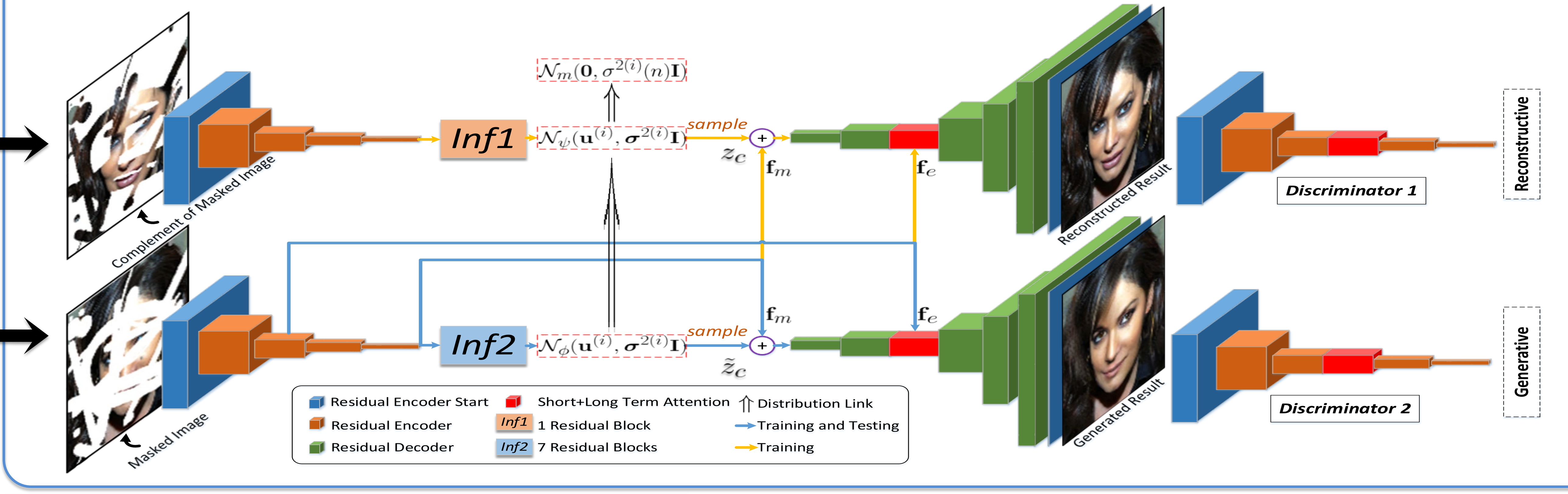




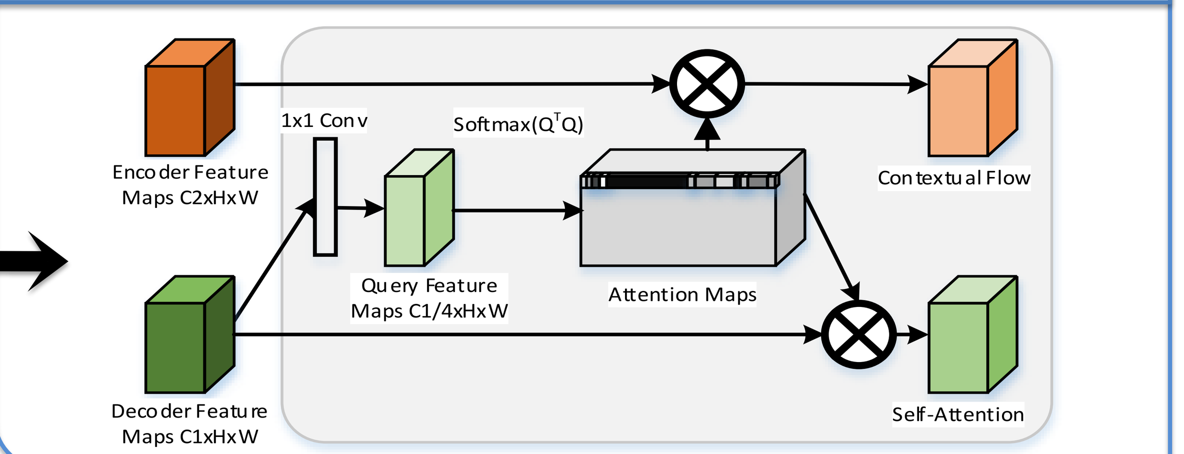
Key Ideas

1. A probabilistically principled framework
   - Assume missing partial images (patches) belong to a prior distribution
   $$\log p(\mathbf{I}_c) \geq -\mathrm{KL}(q_\psi(\mathbf{z}_c|\mathbf{I}_c)||p(\mathbf{z}_c)) + \mathbb{E}_{q_\psi(\mathbf{z}_c|\mathbf{I}_c)}[\log p_\theta(\mathbf{I}_c|\mathbf{z}_c)]$$
   - Couple prior-conditional lower bound
   $$\log p(\mathbf{I}_c|\mathbf{I}_m) \geq -\mathrm{KL}(q_\psi(\mathbf{z}_c|\mathbf{I}_c)||p_\phi(\mathbf{z}_c|\mathbf{I}_m)) + \boxed{\mathbb{E}_{q_\psi(\mathbf{z}_c|\mathbf{I}_c)}}[\log p_\theta(\mathbf{I}_c|\mathbf{z}_c, \mathbf{I}_m)]$$
   - Reconstruction vs Creative Generation
   $$\log p(\mathbf{I}_c|\mathbf{I}_m) \geq \lambda \left\{\mathbb{E}_{q_\psi}[\log p_\theta^r(\mathbf{I}_c|\mathbf{z}_c, \mathbf{I}_m)] - \mathrm{KL}(q_\psi||p_\phi)\right\} + (1-\lambda)\,\mathbb{E}_{\boxed{p_\phi}}[\log p_\theta^g(\mathbf{I}_c|\mathbf{z}_c, \mathbf{I}_m)]$$
   - Joint unconditional and conditional lower bounds
   $$\begin{aligned}\mathcal{B} &= \beta\,\mathcal{B}_1 + \mathcal{B}_2 \\ &= -\left[\beta\mathrm{KL}(q_\psi||p_{z_c}) + \lambda\mathrm{KL}(q_\psi||p_\phi)\right] \\ &\quad + (\beta+\lambda)\mathbb{E}_{q_\psi}\log p_\theta^r + (1-\lambda)\mathbb{E}_{p_\phi}\log p_\theta^g\end{aligned}$$
2. A novel self-attention layer to exploit **short + long** term context information
   - Attend to the finer-grained features in the encoder or the more semantically generative features in the decoder

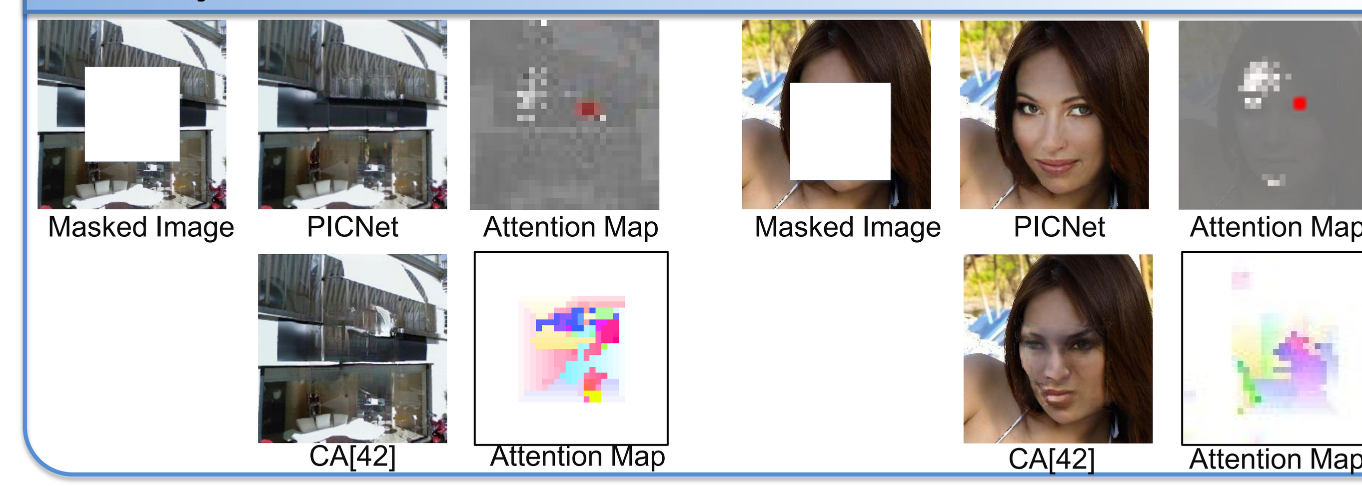
Pluralistic Image Completion Network (PICNet)



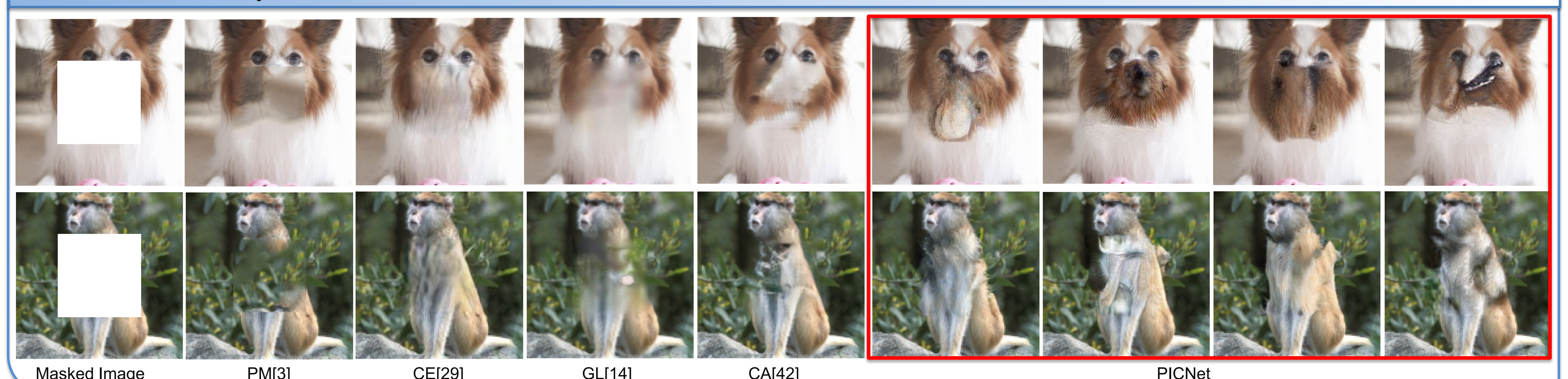
Short + Long Term Attention



Analysis: Attention



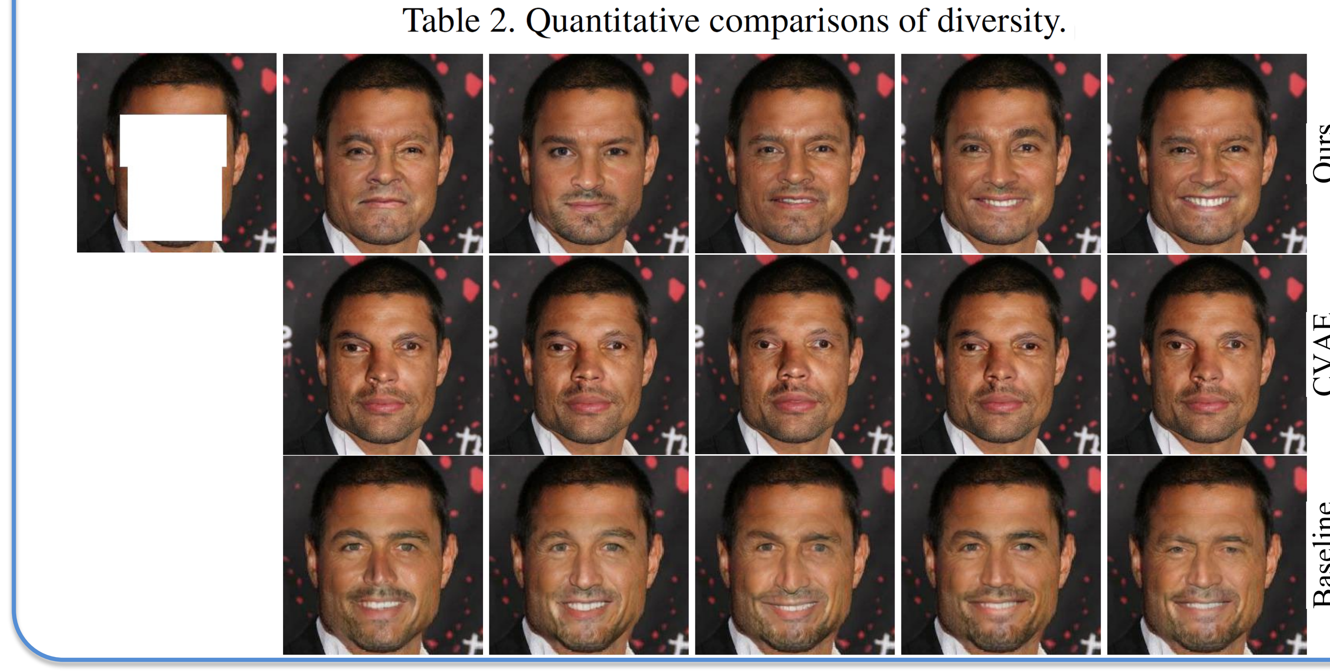
Overall Comparison with Other Methods



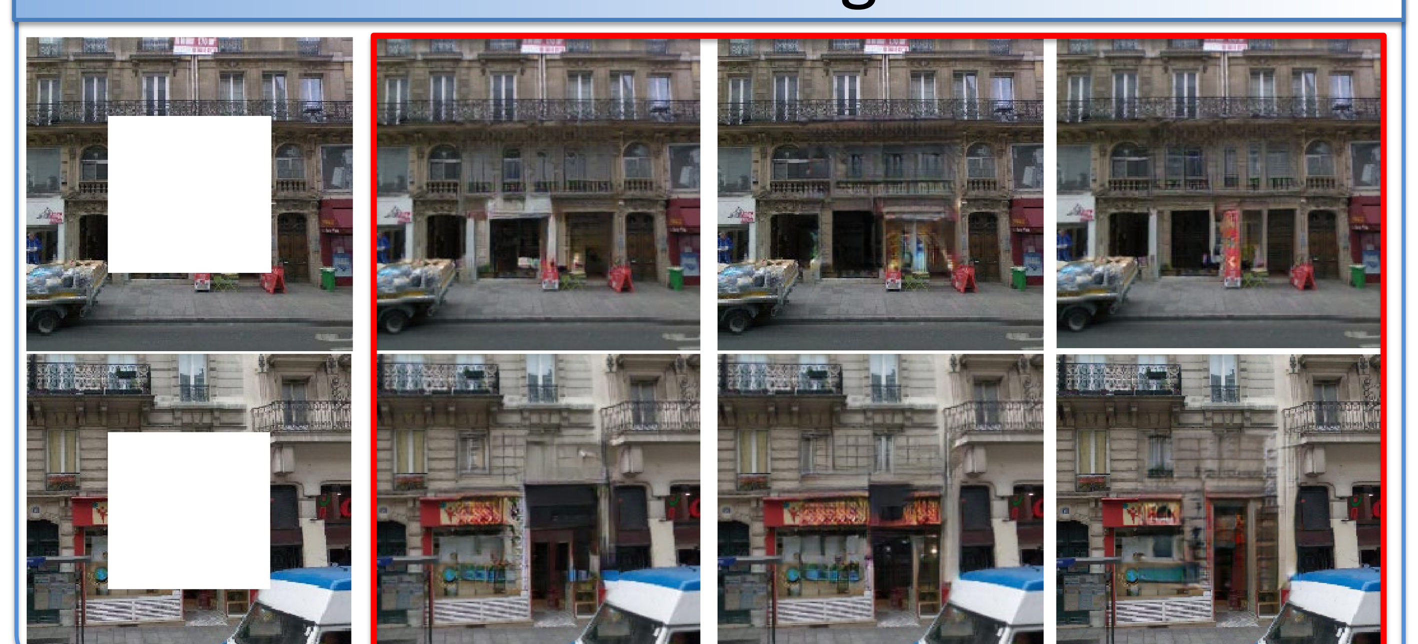
Analysis: Diversity

Method	Diversity (LPIPS) $\mathbf{I}_{out}$	Diversity (LPIPS) $\mathbf{I}_{out(m)}$
CVAE	0.004	0.014
Instance Blind	0.015	0.049
BicycleGAN [46]	0.027	0.060
PICNet-Pluralistic	**0.029**	**0.088**

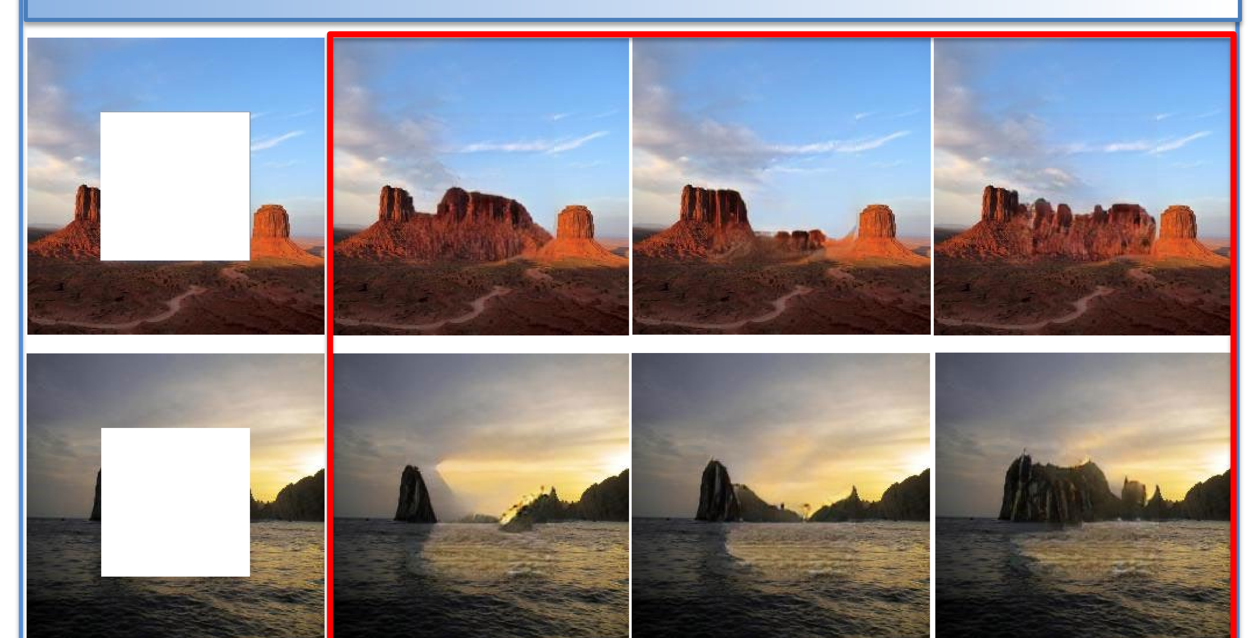
Table 2. Quantitative comparisons of diversity.



Visual Results for Buildings



Visual Results for Natural Scenes



More Source: Project and Code