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
     \[
     \log p(I) \geq -KL(q_k(x_k|I_k)||p_k(x_k)) + \mathbb{E}_{x_k \sim q_k(x_k|I_k)} \log p_k(x_k|I_k)
     \]
   - Couple prior-conditional lower bound
     \[
     \log p(I|I_m) \geq -KL(q_k(x_k|I_m)||p_k(x_k|I_m)) + \mathbb{E}_{x_k \sim q_k(x_k|I_m)} \log p_k(x_k|I_m)
     \]
   - Reconstruction vs Creative Generation
     \[
     \log p(I|I_m) \geq \lambda \mathbb{E}_{x_k \sim q_k(x_k|I_m)} \log p_k(x_k|I_m) - KL(q_k|p_k)
     \]
   - Joint unconditional and conditional lower bounds
     \[
     B = \beta B_1 + B_2
     \]
     \[
     = -\lambda \mathbb{E}_{x_k \sim q_k(x_k|I_m)} \log p_k(x_k|I_m) + (\beta + \lambda) \mathbb{E}_{x_k \sim q_k(x_k|I_m)} \log p_k(x_k|I_m)
     \]
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.

Overall Comparison with Other Methods

Pluralistic Image Completion Network (PICNet)

Analysis: Attention

Analysis: Diversity

Visual Results for Buildings

Visual Results for Natural Scenes

More Source: Project and Code