

# Learning Pyramid-Context Encoder Network for High-Quality Image Inpainting

Yanhong Zeng<sup>1,2\*</sup>, Jianlong Fu<sup>3</sup>, Hongyang Chao<sup>1,2</sup>, Baining Guo<sup>3</sup>

<sup>1</sup>School of Data and Computer Science, Sun Yat-sen University, Guangzhou, P.R. China

<sup>2</sup>The Key Laboratory of Machine Intelligence and Advanced Computing (Sun Yat-sen University),  
Ministry of Education, Guangzhou, P.R. China

<sup>3</sup>Microsoft Research, Beijing, P.R. China



中国科学院大学

University of Chinese Academy of Sciences

# 目录

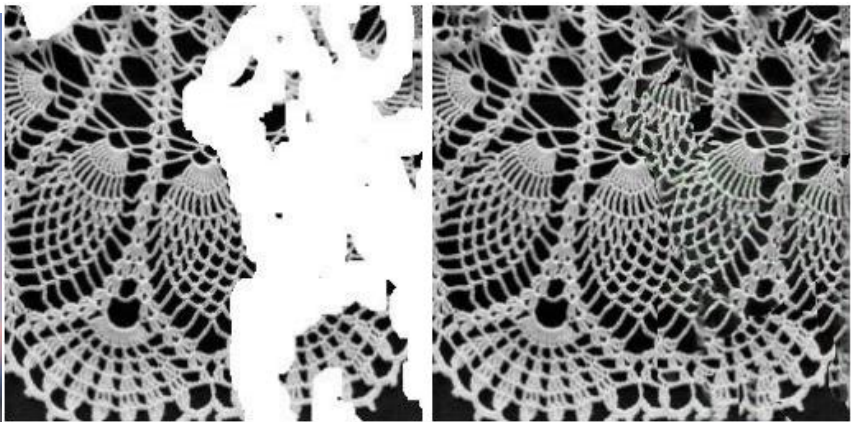
- 引言
- 相关工作
- 主要贡献
- 实验结果
- 总结



# 引言

## ■ 图像补全任务的应用

- 旧照片的恢复
- 去除水印
- 视频特效
- ...



# 引言

## ■ 存在的问题

- 语义不够合理
- 纹理不够清晰
- ...



中国科学院大学  
University of Chinese Academy of Sciences



# 相关工作

## ■ Image inpainting by patch-based methods

- 从无需修复的区域中找到合适的patch或者像素进行填充（cvInpaint Telea模式、PS CS5中的content aware fill功能）
- 优势：能得到与上下文相似的清晰结果。
- 劣势：由于缺乏对图像整体的高级理解，导致无法生成语义上可行的结果。



# 相关工作

## ■ Image inpainting by deep generative models

- 用于图像修复的深度生成模型通常将图像编码为潜在特征，在特征级填充缺失区域，并将特征解码回图像。
- 优势：利用高级语义信息，深度生成模型能得到语义信息一致的结果。
- 不足：生成视觉逼真的结果的能力，还有提高的空间



# 主要贡献

## ■ Cross-layer attention transfer

■ 在编码器中从高级特征图学习区域亲和力图，用于指导相邻的低级特征图的缺失区域的填充。

## ■ Pyramid filling

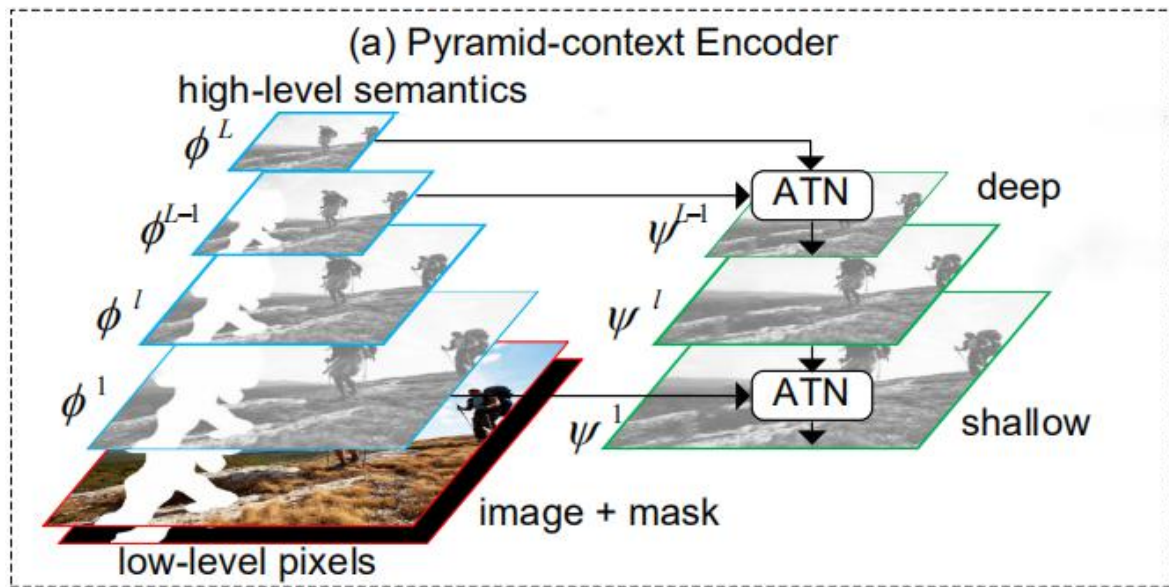
■ 在解码的时候，从深到浅重复使用ATNS,多次填充mask区域(取决于编码器的深度)，从而可以生成细粒度更高的细节图像。





# 主要贡献

## ■ Pyramid context ENcoder Network (PEN-Net)



$$\psi^{L-1} = f(\phi^{L-1}, \phi^L),$$

$$\psi^{L-2} = f(\phi^{L-2}, \psi^{L-1}),$$

$\dots$ ,

$$\psi^1 = f(\phi^1, \psi^2) = f(\phi^1, f(\phi^2, \dots f(\phi^{L-1}, \phi^L)))$$

$\phi^L, \phi^{L-1}, \dots, \phi^1$ : 表示特征图由深到浅

$f$ : Attention Transfer Network(ATN)操作





# 主要贡献

## ■ Attention Transfer Network (ATN)

step1: 计算 mask 区域内外的余弦相似度

$$s_{i,j}^l = \left\langle \frac{p_i^l}{\|p_i^l\|_2}, \frac{p_j^l}{\|p_j^l\|_2} \right\rangle$$

step2: 对每个 patch 的余弦相似度用 softmax 计算 attention score

$$\alpha_{j,i}^l = \frac{\exp(s_{i,j}^l)}{\sum_{i=1}^N \exp(s_{i,j}^l)}$$

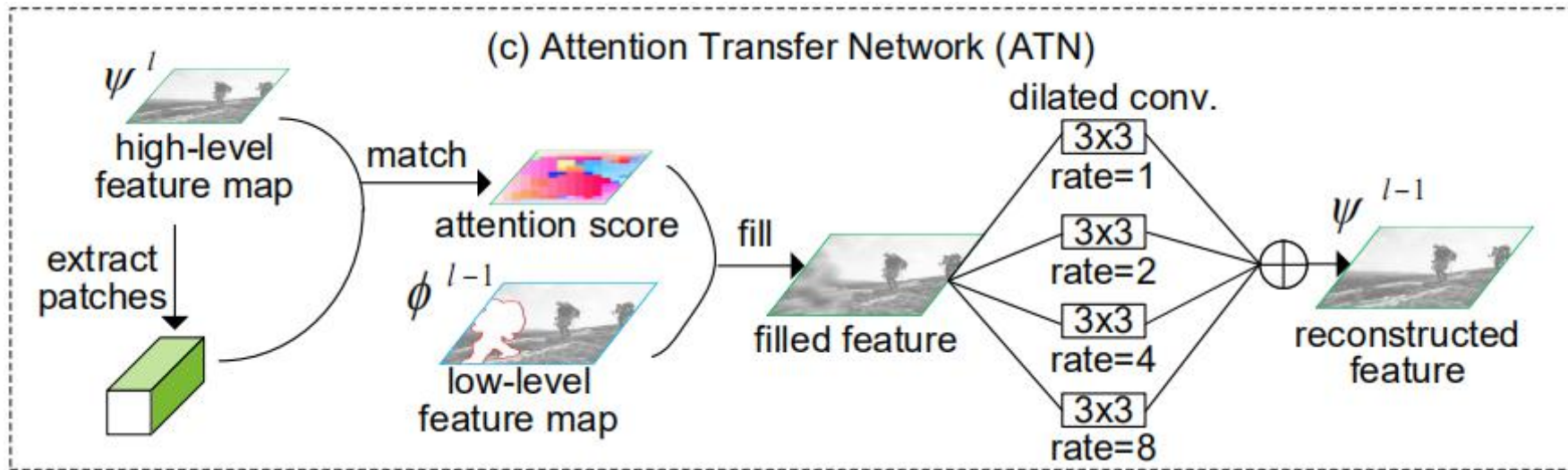
step3: 利用高级特征得到的 attention score, 去填充相邻的低级特征图的缺失区域

$$p_j^{l-1} = \sum_{i=1}^N \alpha_{j,i}^l p_i^{l-1}$$



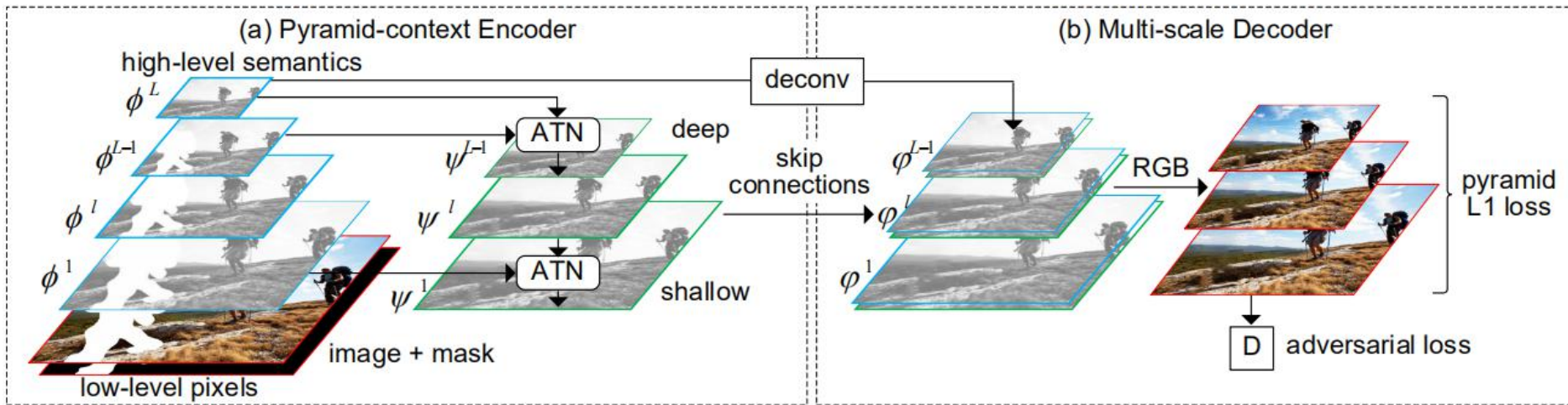
# 主要贡献

## ■ Attention Transfer Network (ATN)



# 主要贡献

## Multi-scale decoder



$\varphi^{L-1}, \varphi^{L-2}, \dots, \varphi^1$ : 表示多尺度解码器生成的由深到浅的特征图

$g$ : 表示转置卷积操作

$$\varphi^{L-1} = g(\psi^{L-1} \oplus g(\phi^L)),$$

$$\varphi^{L-2} = g(\psi^{L-2} \oplus \varphi^{L-1}),$$

$\dots,$

$$\varphi^1 = g(\psi^1 \oplus \varphi^2),$$



# 主要贡献

## ■ Pyramid L1 losses

对每个尺度的预测结果和ground truth  $x^l$  做L1距离

$$L_{pd} = \sum_{l=1}^{L-1} \|x^l - h(\varphi^l)\|_1$$

其中的h操作将  $\varphi^l$  解码为一个和  $x^l$  尺寸相同的RGB图

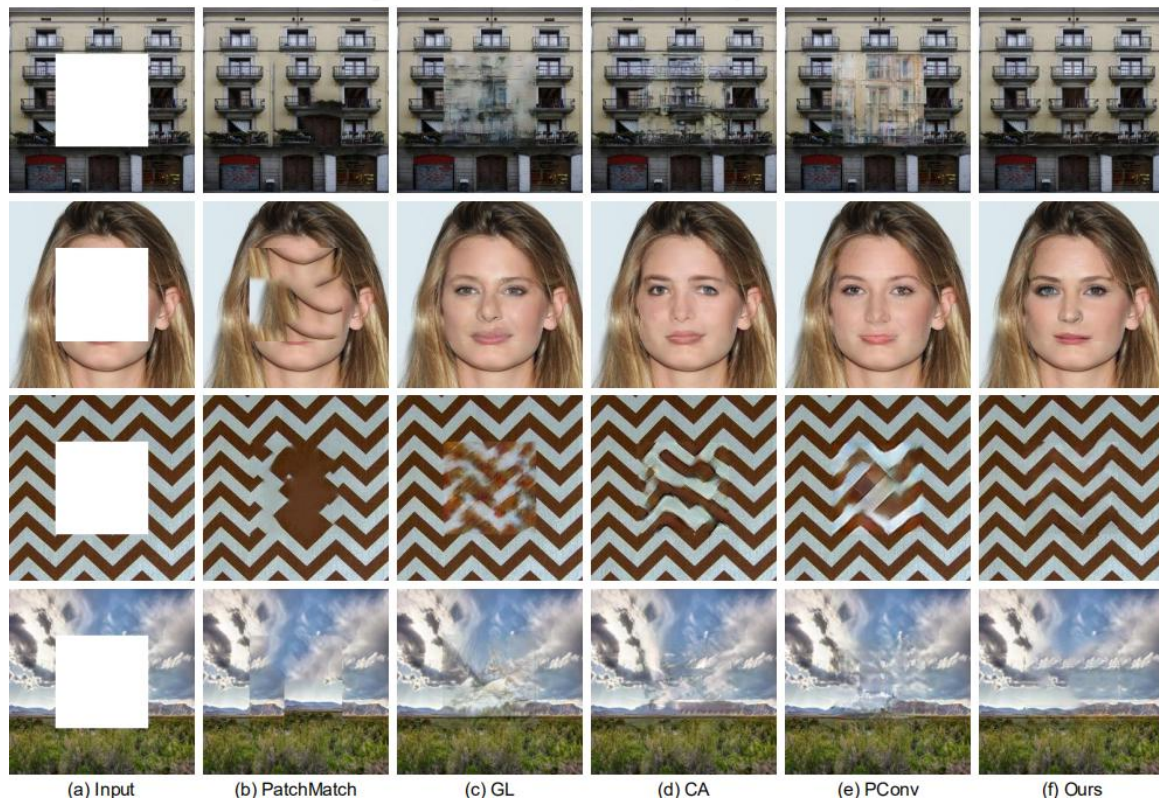


# 目录

- 引言
- 相关工作
- 主要贡献
- 实验结果
- 总结



# 实验结果



128\*128的mask



中国科学院大学  
University of Chinese Academy of Sciences





# 实验结果



随机的mask



中国科学院大学  
University of Chinese Academy of Sciences



# 目录

- 引言
- 相关工作
- 主要贡献
- 实验结果
- 总结



# 总结

- 提出了一个金字塔式的上下文编码器网络(PEN-net)生成语义合理和视觉逼真的图像修复结果。
- 提出了交叉层 的注意力迁移和金字塔式填充机制，用于保留更多细节信息
- 提高了U-net的编码和解码效率。

