



中国科学技术大学

University of Science and Technology of China

# Conditional Probability Models for Deep Image Compression



- 作者介绍
- 文章概述
- 研究方法
- 实验效果
- 总结



- 作者介绍
- 文章概述
- 研究方法
- 实验效果
- 总结

- Fabian Mentzer: 苏黎世联邦理工学院博士生  
研究方向: 图像压缩
- Eirikur Agustsson: 苏黎世联邦理工学院博士生  
研究方向: 图像超分辨/图像压缩/人工智能/生成模型
- Michael Tschannen: Google Research (Brain team) 博士后研究员  
研究方向: 生成模型/图像压缩
- Radu Timofte: 苏黎世联邦理工学院讲师  
研究方向: 图像超分辨/深度学习/目标检测/识别/分割/稀疏表示
- Luc Van Gool: 苏黎世联邦理工学院教授  
研究方向: 计算机视觉/机器学习/人工智能/自动驾驶



- 作者介绍
- 文章概述
- 研究方法
- 实验效果
- 总结

## □ 背景

- 深度学习用于图像压缩
- 两个挑战：处理量化和rate-distortion平衡

## □ 目的

- 引导图像压缩自编码器中的rate-distortion平衡

## □ 方法

- 利用上下文模型（3D卷积）来学习自编码器潜在分布的条件概率模型



- 作者介绍
- 文章概述
- 研究方法
- 实验效果
- 总结

- 给定中心  $\mathcal{C} = \{c_1, \dots, c_L\} \subset \mathbb{R}$ ，使用最近邻分配计算：

$$\hat{z}_i = Q(z_i) := \arg \min_j \|z_i - c_j\|$$

但是在网络的反向传播中使用可微分的软量化：

$$\tilde{z}_i = \sum_{j=1}^L \frac{\exp(-\sigma \|z_i - c_j\|)}{\sum_{l=1}^L \exp(-\sigma \|z_i - c_l\|)} c_j$$

来计算梯度。



## □ 采用PixelRNN的方法：

- $p(\hat{\mathbf{z}})$  的分布理解为条件分布的乘积：
$$p(\hat{\mathbf{z}}) = \prod_{i=1}^m p(\hat{z}_i | \hat{z}_{i-1}, \dots, \hat{z}_1)$$
- 3D特征图  $\hat{\mathbf{z}}$  按照光栅扫描顺序索引。使用一个神经网络  $p(\hat{\mathbf{z}})$  (上下文模型) 来估算每一项  $P_{i,l}(\hat{\mathbf{z}}) \approx p(\hat{z}_i = c_l | \hat{z}_{i-1}, \dots, \hat{z}_1)$ 。

## □ 因果约束

- 在网络的每一层对权重进行适当的masking

- 对  $\hat{z}_i$  的索引利用交叉熵损失函数进行分类

$$CE := \mathbb{E}_{\hat{\mathbf{z}} \sim p(\hat{\mathbf{z}})} \left[ \sum_{i=1}^m -\log P_{i,I(\hat{z}_i)} \right]$$

- 可以将CE看作是  $H(\hat{\mathbf{z}})$  的估计

$$\begin{aligned} H(\hat{\mathbf{z}}) &= \mathbb{E}_{\hat{\mathbf{z}} \sim p(\hat{\mathbf{z}})} [-\log(p(\hat{\mathbf{z}}))] \\ &= \mathbb{E}_{\hat{\mathbf{z}} \sim p(\hat{\mathbf{z}})} \left[ \sum_{i=1}^m -\log p(\hat{z}_i | \hat{z}_{i-1}, \dots, \hat{z}_1) \right] \\ &\approx \mathbb{E}_{\hat{\mathbf{z}} \sim p(\hat{\mathbf{z}})} \left[ \sum_{i=1}^m -\log q(\hat{z}_i | \hat{z}_{i-1}, \dots, \hat{z}_1) \right] \\ &= \mathbb{E}_{\hat{\mathbf{z}} \sim p(\hat{\mathbf{z}})} \left[ \sum_{i=1}^m -\log P_{i,I(\hat{z}_i)} \right] \\ &= CE \end{aligned}$$

- 可以通过交叉熵CE来间接地对  $H(\hat{\mathbf{z}})$  进行最小化。使用  $C(\hat{\mathbf{z}}) := \sum_{i=1}^m -\log P_{i,I(\hat{z}_i)}$  作为潜在图像表示的编码成本。

□ 给定一个自编码器(E, D):

- 可以通过最小化  $CE := \mathbb{E}_{\hat{\mathbf{z}} \sim p(\hat{\mathbf{z}})} \left[ \sum_{i=1}^m -\log P_{i,I(\hat{z}_i)} \right]$  训练  $\mathbf{P}$  来对  $\hat{\mathbf{z}}$  的依赖关系进行建模

□ 利用  $\mathbf{P}$  模型可以得到熵的估计值  $CE$ :

- 调整编码器和解码器
- 使得  $d(\mathbf{x}, D(Q(E(\mathbf{x})))) + \beta H(\hat{\mathbf{z}})$  减小 (平衡 rate-distortion)

## □ 简化设置:

- 使用编码器 $\mathbf{E}$ 的最后一层，并添加第二个单通道输出  $\mathbf{y} \in \mathbb{R}^{\frac{W}{8} \times \frac{H}{8} \times 1}$
- 将单通道输出 $\mathbf{y}$ 扩展成一个与 $\mathbf{z}$ 具有相同维度的掩膜  $\mathbf{m} \in \mathbb{R}^{\frac{W}{8} \times \frac{H}{8} \times K}$

$$m_{i,j,k} = \begin{cases} 1 & \text{if } k < y_{i,j} \\ (y_{i,j} - k) & \text{if } k \leq y_{i,j} \leq k + 1 \\ 0 & \text{if } k + 1 > y_{i,j} \end{cases}$$

- 使用 $\mathbf{m}$ 的二值化点乘来对 $\mathbf{z}$ 进行masking操作  $\mathbf{z} \leftarrow \mathbf{z} \odot [\mathbf{m}]$

- 可以通过计算每一列  $\mathbf{z}_{i,j,:}$  最后的连续零符号来从  $\hat{\mathbf{z}}$  中恢复  $\lceil \mathbf{m} \rceil$ :

$$\begin{aligned} H(\hat{\mathbf{z}}) &= H(\lceil \mathbf{m} \rceil | \hat{\mathbf{z}}) + H(\hat{\mathbf{z}}) \\ &= H(\hat{\mathbf{z}}, \lceil \mathbf{m} \rceil) \\ &= H(\hat{\mathbf{z}} | \lceil \mathbf{m} \rceil) + H(\lceil \mathbf{m} \rceil) \end{aligned}$$

- 把  $H(\lceil \mathbf{m} \rceil)$  当作常数，可以通过  $H(\hat{\mathbf{z}} | \mathbf{m})$  来间接最小化  $H(\hat{\mathbf{z}})$

- 为了估计  $H(\hat{\mathbf{z}} | \mathbf{m})$ ，使用与  $p(\hat{\mathbf{z}}) = \prod_{i=1}^m p(\hat{z}_i | \hat{z}_{i-1}, \dots, \hat{z}_1)$  中相同的  $\mathbf{p}$  因式分解，但由于  $\lceil \mathbf{m} \rceil$  已知，在  $\hat{\mathbf{z}}$  中确定位置有  $p(\hat{z}_i = c_0) = 1$ ，熵估计中对应项的对数为0。剩下的项可以由相同的上下文模型建模：

$$H(\hat{\mathbf{z}} | \lceil \mathbf{m} \rceil) \approx E_{\hat{\mathbf{z}} \sim p(\hat{\mathbf{z}})} \left[ \sum_{i=1}^m -\lceil m_i \rceil \log P_{i,I(\hat{z}_i)} \right]$$

- 自编码器和量化器的率-失真平衡：

$$\mathcal{L}_{E,D,Q} = \frac{1}{B} \sum_{j=1}^B d(\mathbf{x}^{(j)}, \hat{\mathbf{x}}^{(j)}) + \beta MC(\hat{\mathbf{z}}^{(j)})$$

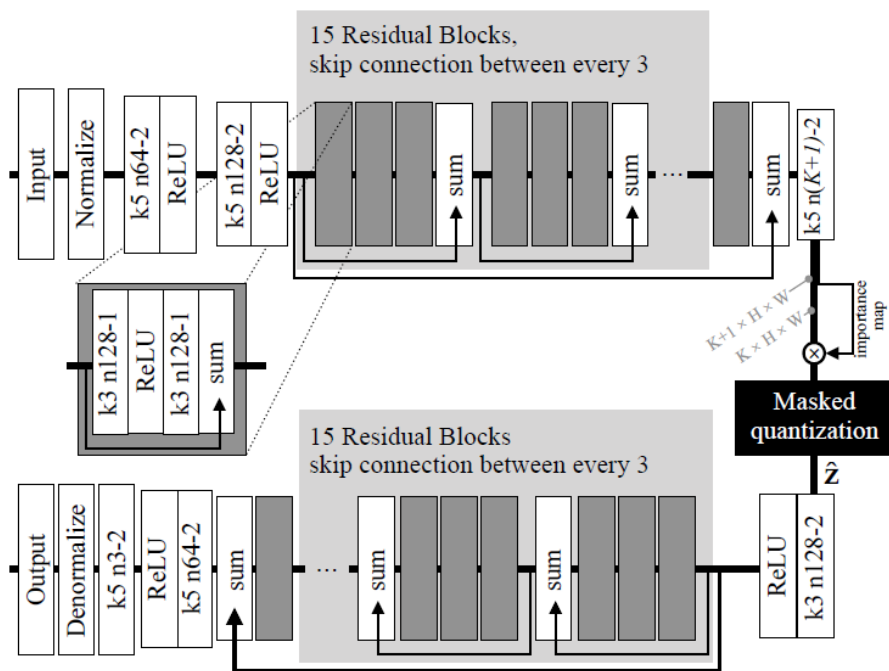
- 上下文模型的目标函数：

$$\mathcal{L}_P := \frac{1}{B} \sum_{j=1}^B d(\mathbf{x}^{(j)}, \hat{\mathbf{x}}^{(j)}) + \beta C(\hat{\mathbf{z}}^{(j)})$$

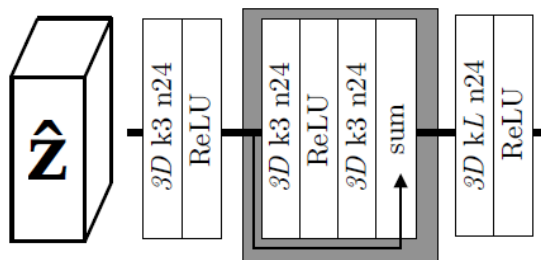
- 步骤：

- 从编码器中获得潜在表示 $\mathbf{z}$ 和重要性映射 $\mathbf{y}$ ；
- 将 $\mathbf{y}$ 扩展成掩膜 $\mathbf{m}$ ；
- 对 $\mathbf{z}$ 进行masking操作  $\mathbf{z} \leftarrow \mathbf{z} \odot [\mathbf{m}]$ ；
- 量化  $\hat{\mathbf{z}} = Q(\mathbf{z})$ ；
- 计算上下文  $P(\hat{\mathbf{z}})$ ；
- 利用解码器解码。

## 自编码器的网络结构:



## 上下文模型:

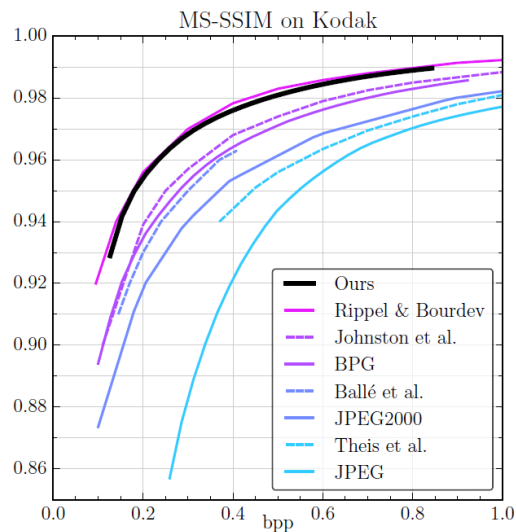




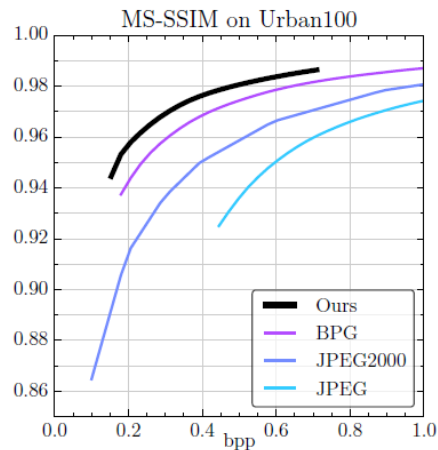
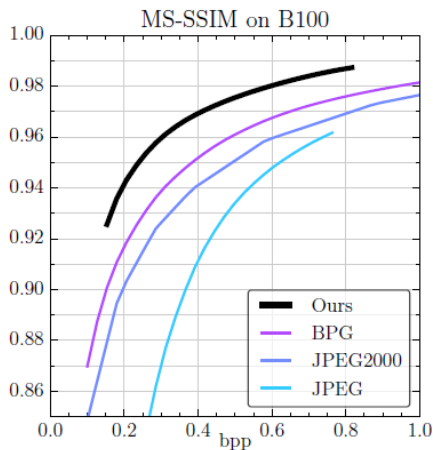
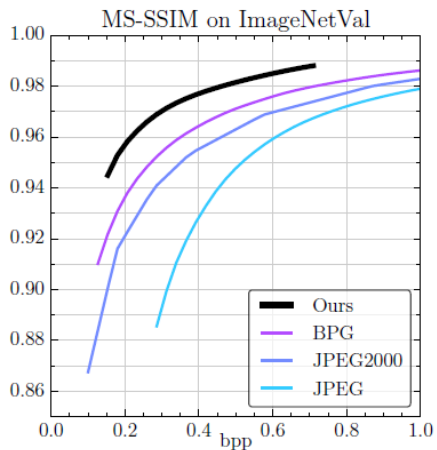
- 作者介绍
- 文章概述
- 研究方法
- 实验效果
- 总结



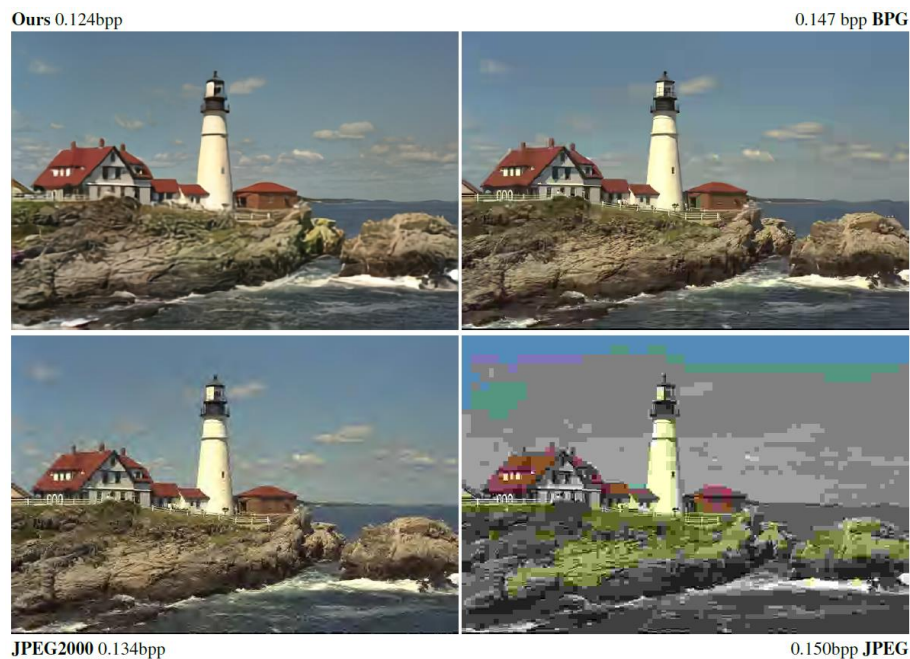
## 在Kodak数据集上与其他方法的对比



## 在其他测试集上与BPG、JPEG和JPEG2000的对比



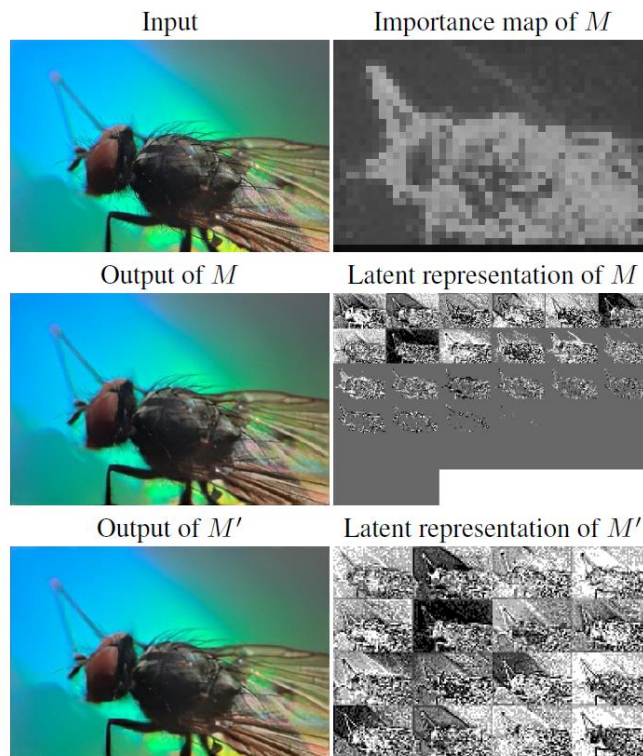
## □ 视觉效果对比



## □ 消融实验：上下文模型

| Model                 | rate      |
|-----------------------|-----------|
| Baseline (Uniform)    | 0.646 bpp |
| Zeroth order          | 0.642 bpp |
| First order           | 0.627 bpp |
| Our context model $P$ | 0.579 bpp |

## □ 重要性映射





- 作者介绍
- 文章概述
- 研究方法
- 实验效果
- 总结

## □ 文章贡献

- 提出了第一种同时学习图像压缩自编码器和轻量级上下文模型的方法
- 性能与目前最先进的深度图像压缩系统相当



中国科学技术大学  
University of Science and Technology of China

**Thanks!**