

根据学校课堂纪律的要求



请同学们坐在前五排





数字媒体技术基础

Meng Yang

www.smartllv.com

SUN YAT-SEN University



**机器智能与先进计算教
育部重点实验室**



**智能视觉语言
学习研究组**



12 新媒体压缩



随着新媒体技术的快速发展对音视频的要求也越来越高



12 智能新媒体压缩技术

传统压缩技术够用了吗？

❑ YouTube下调视频默认清晰度



新冠疫情爆发后，随着数亿人留在家中躲避疫情，流媒体服务的用量出现激增。2020年3月25日，Google旗下的全球最大的在线视频平台YouTube宣布将在全球范围内降低视频的默认清晰度，以缓解互联网的拥堵现象，但却引起了用户的强烈不满

<https://tech.sina.com.cn/i/2020-03-25/doc-imxxsth1558187.shtml>

这张图存储空间多大？

问题？



仅有41KB



- ❑ 第12章 智能新媒体压缩技术
- ❑ 12.1 基于深度学习的图像压缩技术
 - 12.1.1 基于CNN的图像压缩
 - 12.1.2 基于RNN的图像压缩
 - 12.1.3 基于GAN的图像压缩
- ❑ 12.2 基于深度学习的视频压缩技术
 - 12.2.1 基于深度模型替代传统模块
 - 12.2.2 基于端到端学习

第一部分

12.1 基于深度学习的图像压缩技术



12.1 基于深度学习的图像压缩

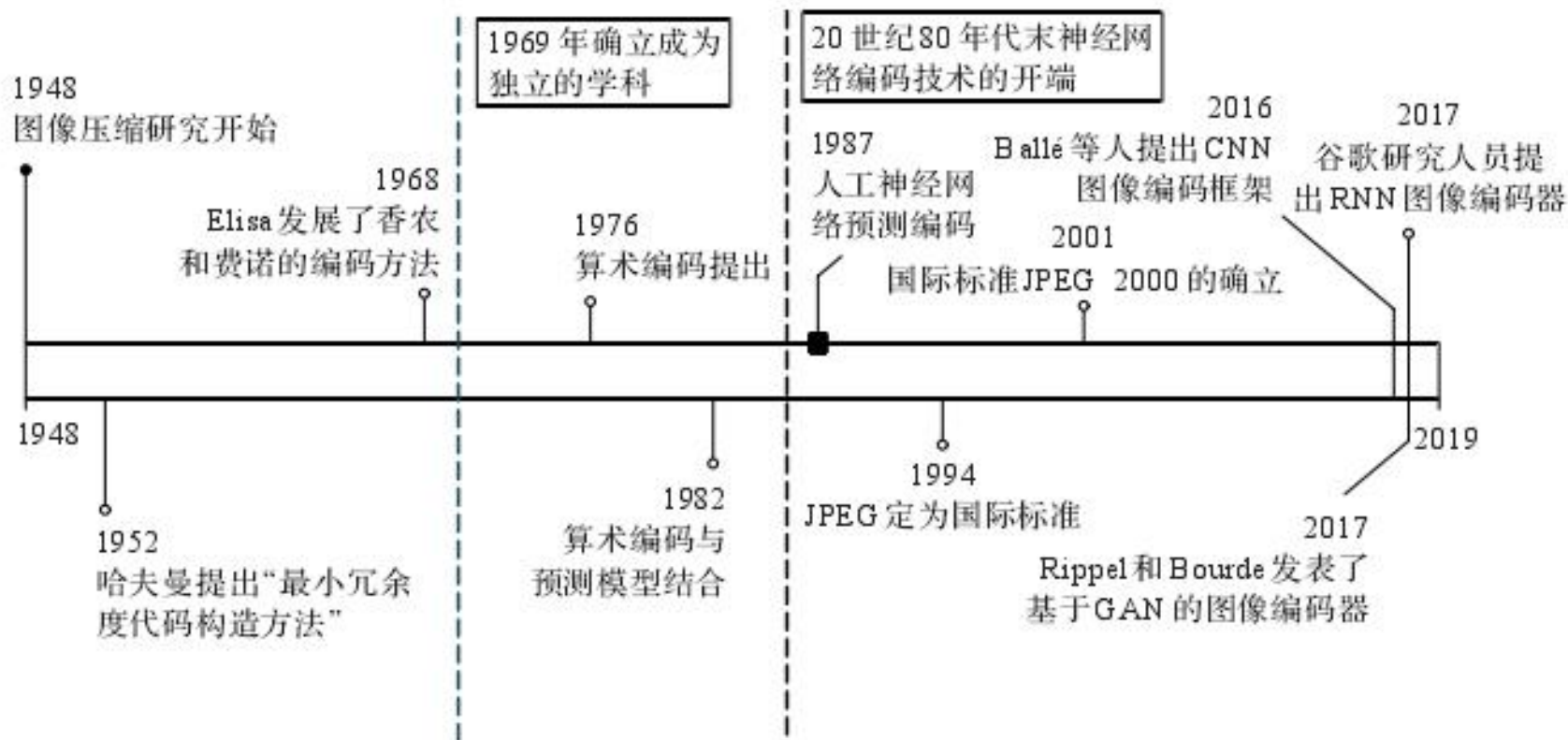
深度学习在图像压缩中的潜力

- ❑ 传统图像压缩算法的局限性.
 - 压缩后图像的分辨率相对较低
 - 图像压缩的比特率有待提升
 - 图像变换方式固定
- ❑ 深度学习方法的优势.
 - 图像特征提取、表达能力强
 - 高维数据处理能力强
 - 支持端到端训练得到更优的图像变换方式



12.1 基于深度学习的图像压缩

图像压缩发展历程



[1] 于恒, 梅红岩, 许晓明, 等. 基于深度学习的图像压缩算法研究综述.



12.1 基于深度学习的图像压缩

发展现状

❑ 图像压缩算法分类.

- 无损压缩（对编码信息可以恢复）
- 有损压缩（大多数基于深度学习的方法属于此类）

❑ 图像压缩算法性能.

- 目前压缩性能已经超越JPEG
- 对传统压缩算法的性能优势逐步扩大





12.1 基于深度学习的图像压缩

发展现状

- ❑ CVPR Workshop and Challenge on Learned Image Compression (CLIC)
 - 自CVPR 2018起，每年举办1届，今年为第4届

4th Workshop and Challenge on Learned Image Compression

Sign up

Introduction

Our workshop aims to gather publications which will advance the field of image and video compression using machine learning and computer vision. We want to encourage the development of novel encoder/decoder architectures, novel ways to control information flow between the encoder and the decoder, new perceptual losses, and new ways to learn quantized representations.

At the workshop we will also present the winners of our annual compression challenge. Like the workshop, the challenge is designed to encourage the development of new learned codecs. But it is also an opportunity to evaluate and compare end-to-end trained approaches against classical approaches and every submission is welcome.

Location

The workshop will be part of [CVPR 2021](#) and will be entirely virtual. The conference will take place between June 19 and June 25.



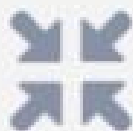
12.1 基于深度学习的图像压缩

发展现状

□ CVPR-CLIC

- 2018/2019年两届冠军方案团队

图鸭科技



高压缩比

比jpeg节省70%带宽流量



视觉效果好

在同样压缩比下PSNR值、ssim值保持领先，视觉效果好

12.1 基于深度学习的图像压缩

发展现状

□ CVPR-CLIC

- 2018/2019年两届冠军方案团队

图鸭科技



原始图片



JPEG

Rate: 0.174bpp
PSNR: 23.82134
MS-SSIM: 0.7442
MS-SSIM-DB: 5.921



TNG

Rate: 0.101bpp
PSNR: 33.52165
MS-SSIM: 0.9623
MS-SSIM-DB: 14.234

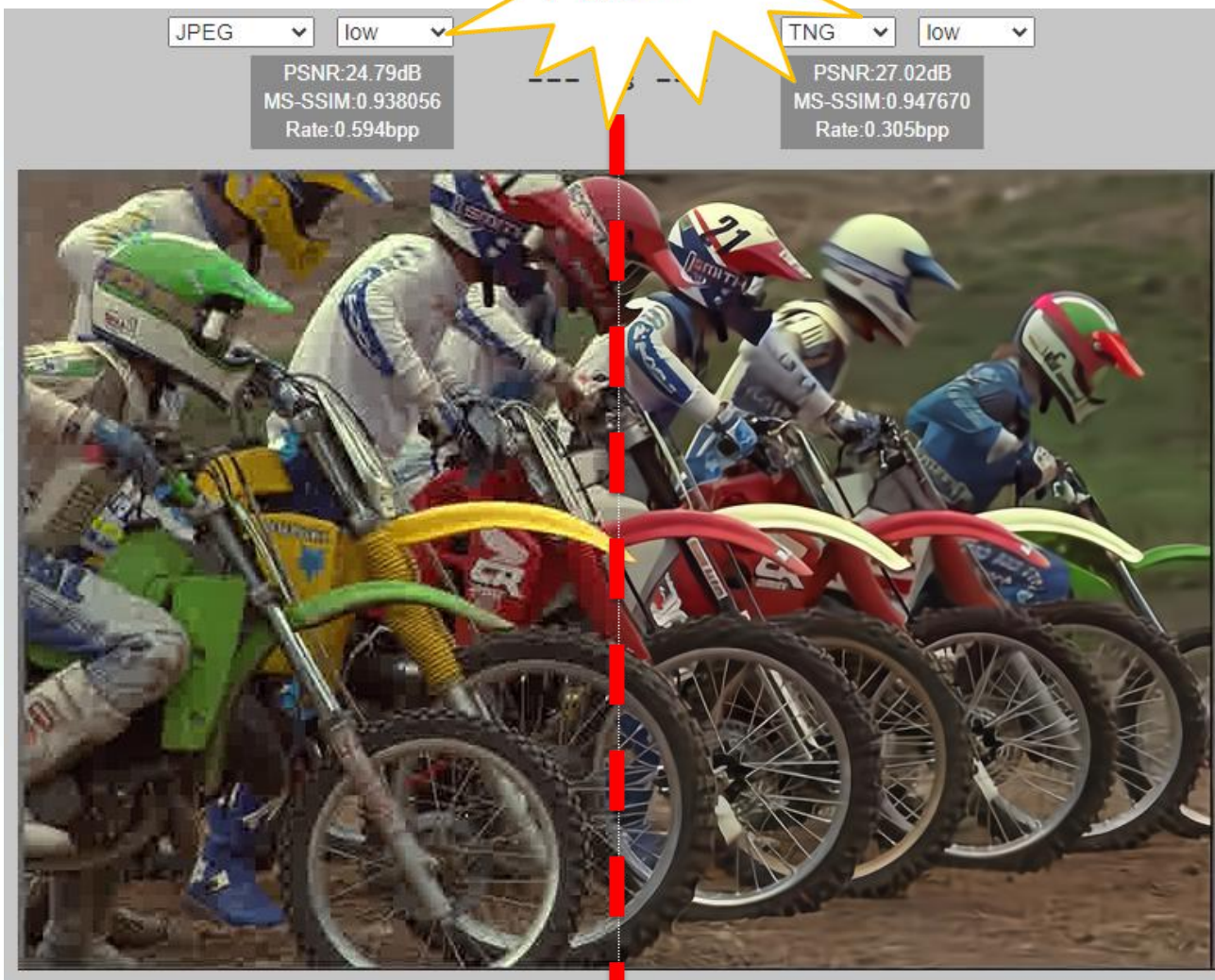
12.1 基于深度学习的图像压缩

Demo

JPEG

问题?

TNG



<https://www.tucodec.com/picture/comparison>



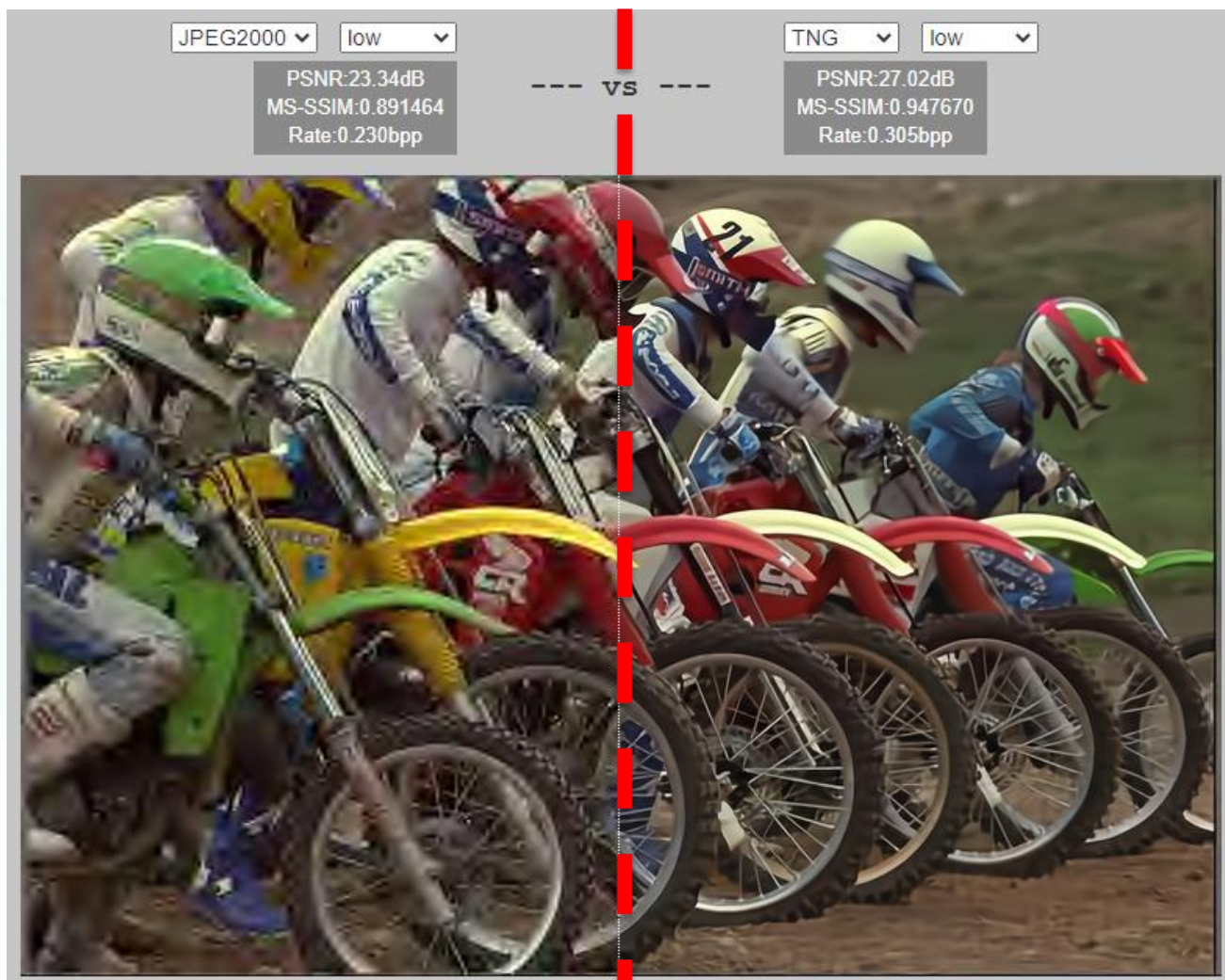
12.1 基于深度学习的图像压缩



Demo

JPEG2000

TNG



<https://www.tucodec.com/picture/comparison>



12.1 基于深度学习的图像压缩

发展现状

□ 应用场景





12.1 基于深度学习的图像压缩

训练数据集

名称	图像数量及简介	备注
ImageNet ^[41]	1 500 多万张图像,2 万多类图像并且带有标签	ImageNet 通常用于图像分类、目标检测等任务,由于其图像数据量大,因此在实际实验中可选用需要的类型即可
Yahoo Flickr	图像存储及视频托管网站,存储超过 6 亿张图片	Flickr 图像均为用户上传图像,其图像分辨率较高,图像类型多样,图像尺寸不同
Open Images V4 ^[42]	具有对象位置注释数据集,约包含 900 万张图像	该数据集源自图像挑战赛,其图像类型多样,图像内容场景复杂,图像中包含对象较多,并对图像中对象位置有注释
DIV2K ^[43]	1 000 张高清图,其中 800 张作为训练,100 张作为验证,再 100 张作为测试	数据集拥有超高的分辨率,并且有较强的多样性,是 NTIRE2018 指定的训练数据集,多用于图像超分辨率任务中
Cityscapes ^[44]	5 000 幅图像,500 张验证图和 1 525 张测试图	包含 50 个城市的一年四季的图像,有高质量的图像注释,从 27 个城市手动选择 5 000 幅图像进行了密集的注释
Place365 ^[45]	包括 365 个场景的 180 万张图像,每个场景最多包含 5 000 张图像	Places365 有两个版本:Places 365- standard 和 Places365- challenge,该数据集标有场景的语义类别,多用于场景分类任务中
COCO	包括 80 类超过 33 万张图像,其中 20 万张图像附有标注	COCO 多用于目标识别、物体检测、对象分割等领域,其单一类别图像数据量大,并且类别丰富
Cifar-10	32×32 彩色图像包含 5 万个训练样本,10 个类别	经典的图像分类、图像识别数据集,图像分辨率不高,类别较少
LSUN ^[46]	包含 10 个场景类别 20 个对象类别,约 100 万个标记图像	主要包含了卧室、客厅、教室等 10 类场景,涵盖了基本的生活场景。可用于大规模场景理解





12.1 基于深度学习的图像压缩

算法性能评价指标

使用被压缩后的重建图像进行图像压缩性能的评估

均方误差 (Mean Square Error, MSE)

$$MSE = \frac{1}{MN} \sum_{y=1}^M \sum_{x=1}^N [I(x, y) - I'(x, y)]^2$$

$I(x, y)$ 和 $I'(x, y)$ 分别表示待评价图像与原始图像

$I(x, y)$ 表示在 (x, y) 位置的像素值



12.1 基于深度学习的图像压缩

算法性能评价指标

均方误差 (Mean Square Error, MSE)

$$MSE = \frac{1}{MN} \sum_{y=1}^M \sum_{x=1}^N [I(x, y) - I'(x, y)]^2$$

□ 局限性.

- 只能说明两幅图像的差异，不考虑到图像高频像素和低频像素分量，而图像压缩就是需要能更大程度上保留和恢复低频像素

12.1 基于深度学习的图像压缩

算法性能评价指标

峰值信噪比 (Peak Signal to Noise Ratio, PSNR)

$$PSNR = 10 \lg \frac{R^2}{MSE}$$

R 表示输入图像的最大值

- 大于 40 dB 说明图像质量接近原图,
- 在 30 dB 与 40 dB 之间时图像存在失真,
- 20 dB 到 30 dB 说明图像质量不好,
- 低于 20 dB 时说明图像质量差



12.1 基于深度学习的图像压缩

算法性能评价指标

峰值信噪比 (Peak Signal to Noise Ratio, PSNR)

$$PSNR = 10 \lg \frac{R^2}{MSE}$$

❑ 局限性.

- 是一种客观的评价指标不会考虑到人的视觉特性，因此可能存在出现评价结果与人的主观感受不一致的情况





12.1 基于深度学习的图像压缩

算法性能评价指标

结构相似性 (Structural Similarity Index, SSIM)

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1) + (2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

x 、 y 分别代表两幅图像

μ_x 、 μ_y 代表两幅图像的平均像素值

σ_{xy} 是图像 x 和图像 y 的协方差

σ_x^2 、 σ_y^2 分别代表图像 x 和图像 y 的方差





12.1 基于深度学习的图像压缩

算法性能评价指标

结构相似性 (Structural Similarity Index, SSIM)

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1) + (2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

□ 优势.

- 考虑到了图像中的可见结构，可以很好地评估图像压缩前与图像压缩后的图像质量





12.1 基于深度学习的图像压缩

算法性能评价指标

多层结构相似性 (Multi-Scale, MS-SSIM)

$$MS-SSIM(X, Y) = [l_M(X, Y)]^{\alpha_M} \cdot \prod_{j=1}^M [c_j(X, Y)]^{\beta_j} [s_j(X, Y)]^{\gamma_j}$$

$l_M(X, Y)$ 表示在 M 尺度下亮度比较

j 表示在第 j 尺度下进行比较

$c_j(X, Y)$ 表示对比度比较

$s_j(X, Y)$ 表示结构比较

α_M 、 β_j 和 γ_j 取经过实验得出的经验值



12.1 基于深度学习的图像压缩

算法性能评价指标

多层结构相似性 (Multi-Scale, MS-SSIM)

$$MS-SSIM(X, Y) = \left[\ell_M(X, Y) \right]^{\alpha_M} \cdot \prod_{j=1}^M \left[c_j(X, Y) \right]^{\beta_j} \left[s_j(X, Y) \right]^{\gamma_j}$$

□ 优势.

- 在融合图像分辨率和观测条件变化方面比单尺度方法具有更大的灵活性
- 比SSIM 更贴近人的主观视觉



12.1 基于深度学习的图像压缩

算法性能评价指标

小结

❑ MSE和PSNR

- 在最简单和最广泛使用的全参考质量度量
- 计算简单，物理意义明确，评价侧重客观性

❑ SIM和MS-SSIM

- 更加贴近人的主观视觉
- 往往结合客观评价指标综合考量重建图像质量





12.1.1 基于CNN的图像压缩

经典CNN图像压缩模型

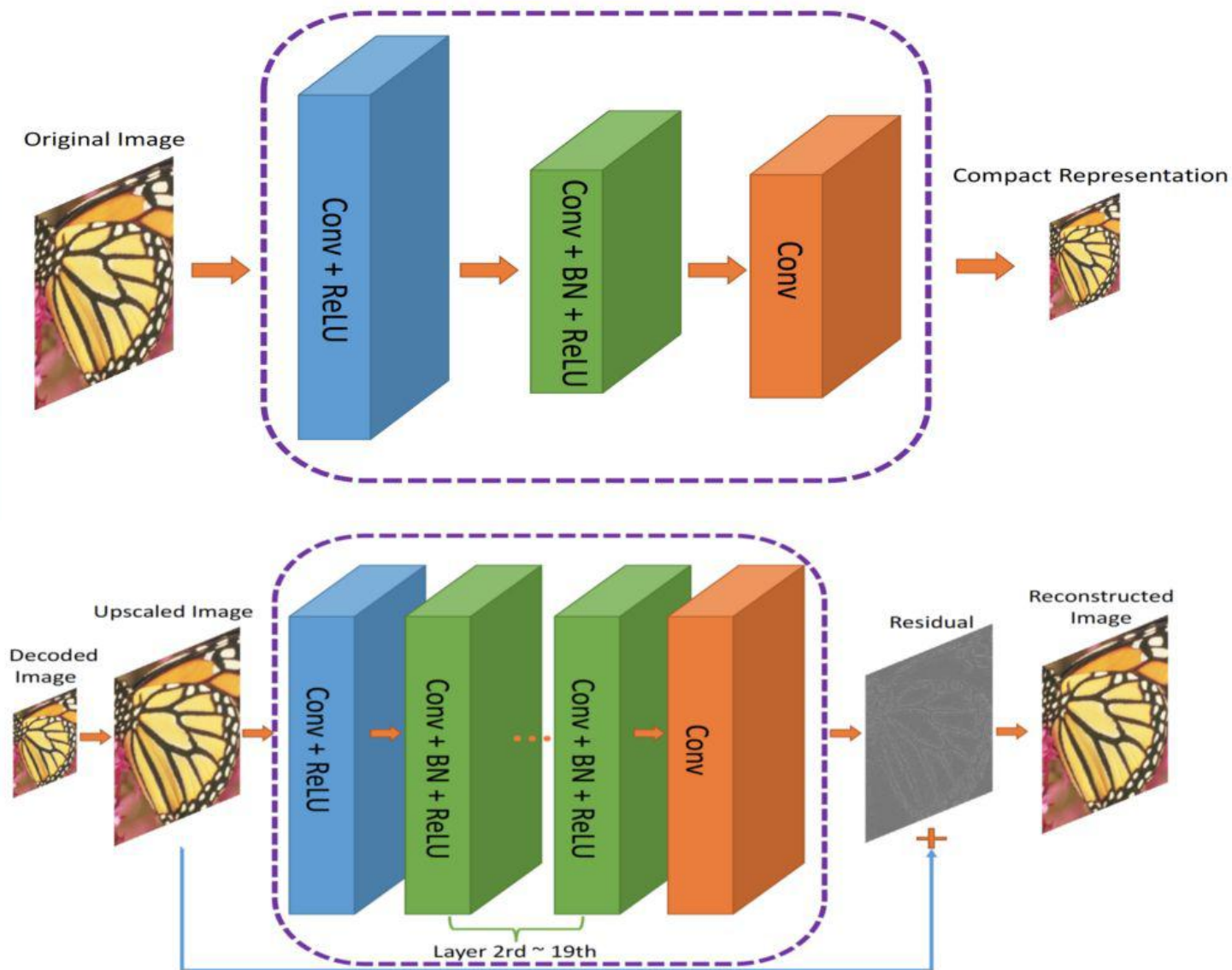
基于广义分歧归一化的卷积神经网络图像编码框架

- 经典的图像压缩如 JPEG、JPEG2000 通常是将变化、量化、熵编码三个部分分别手动优化，图像码率经过量化计算后为离散系数，而基于 CNN 端到端优化采用梯度下降时要求函数全局可微
- 采用卷积神经网络图像编码框架，使线性卷积和非线性更灵活的转换
- 将卷积层分为两个部分，一部分负责分析图像的紧凑表示，另一部分负责重建和逆过程，使用广义分歧归一化函数作为激活函数
- 媲美 JPEG2000 的编码性能

Ballé J, Laparra V, Simoncelli E P. End-to-end optimized image compression [J]. arXiv: 1611.01704, 2016.



经典CNN图像压缩模型



12.1.1 基于CNN的图像

经典CNN图像压缩模型

基于 CNN 的图像端到端压

- 从图像的编码器端和解码器端
- 网络将编码器与解码器
- 采用统一优化方法进行

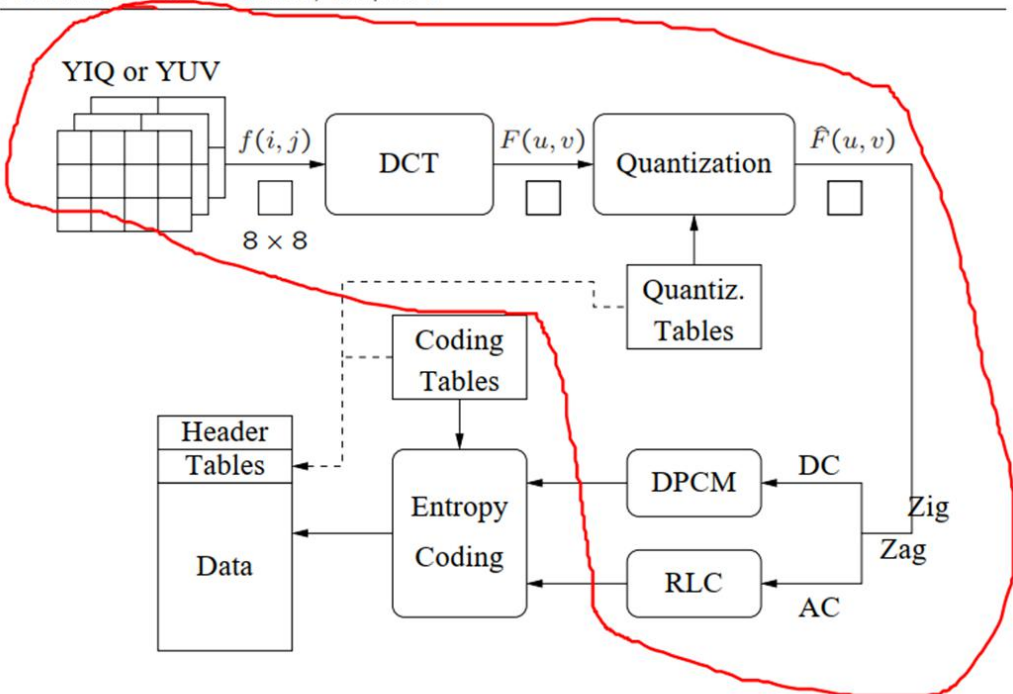
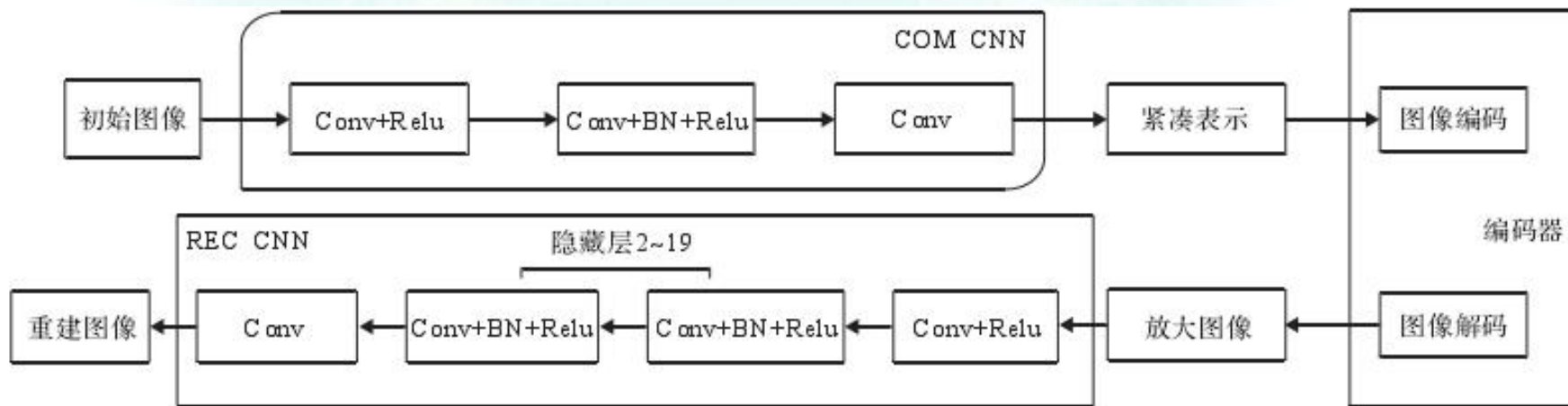


Fig. 9.1: Block diagram for JPEG encoder.



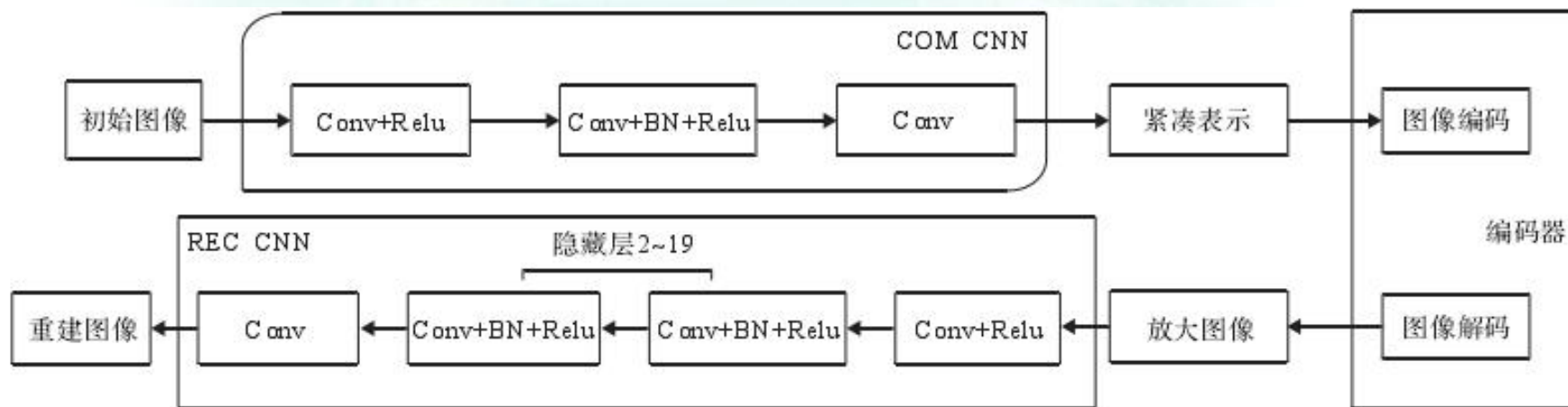
Jiang F, Tao W, Liu S, et al. An end-to-end compression framework based on convolutional neural networks[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2017, 28 (10) : 3007-3018.

12.1.1 基于CNN的图像压缩

经典CNN图像压缩模型

基于 CNN 的图像端到端压缩框架

- 在编码器端使用一个 CNN 用于对图像进行紧凑表示后，在通过编码器进行编码，在解码器端使用一个 CNN 对解码后的图像进行高质量的复原
- 两个网络同时作用，通过卷积采样代替传统图像压缩以图像块为单位的变换计算，其块效应与JPEG相比有明显提升

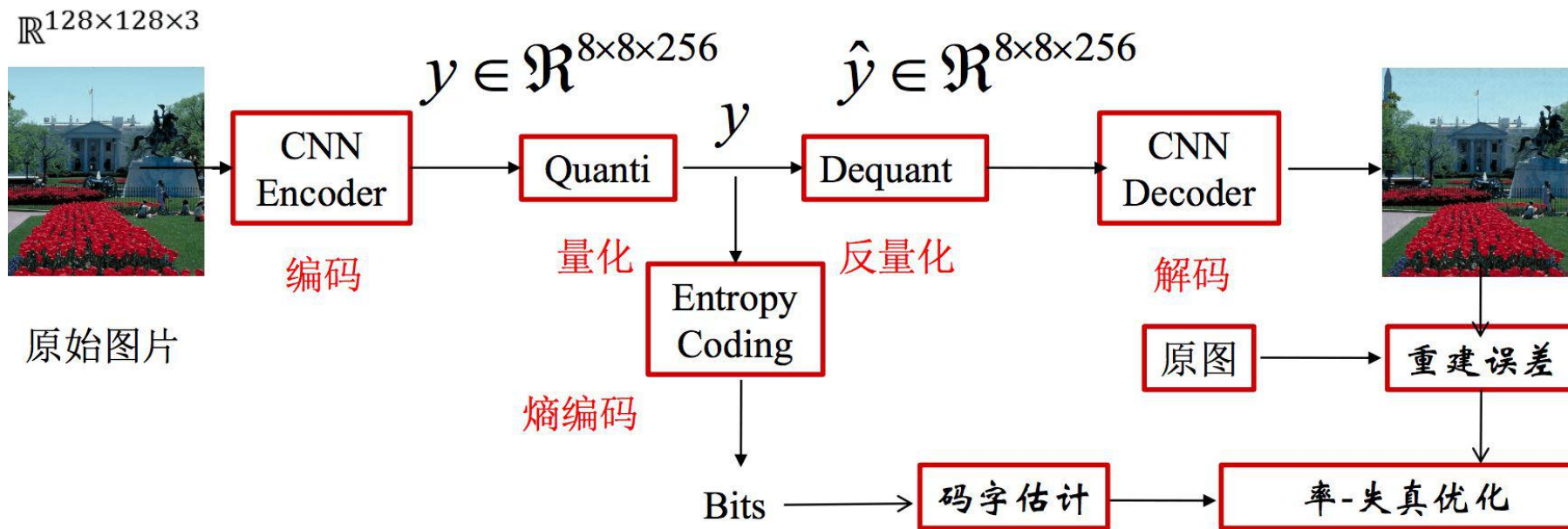


Jiang F, Tao W, Liu S, et al. An end-to-end compression framework based on convolutional neural networks[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2017, 28 (10) : 3007-3018.

12.1.1 基于CNN的图像压缩

CNN图像压缩基本流程

- 以端到端的方式进行图像压缩，通过 CNN 设计编码端与解码端
- 通过大量图像数据以及优化网络方式，获得高性能的压缩框架



12.1.1 基于CNN的图像

经典CNN图像压缩模型

基于卷积自动编码器CAE的

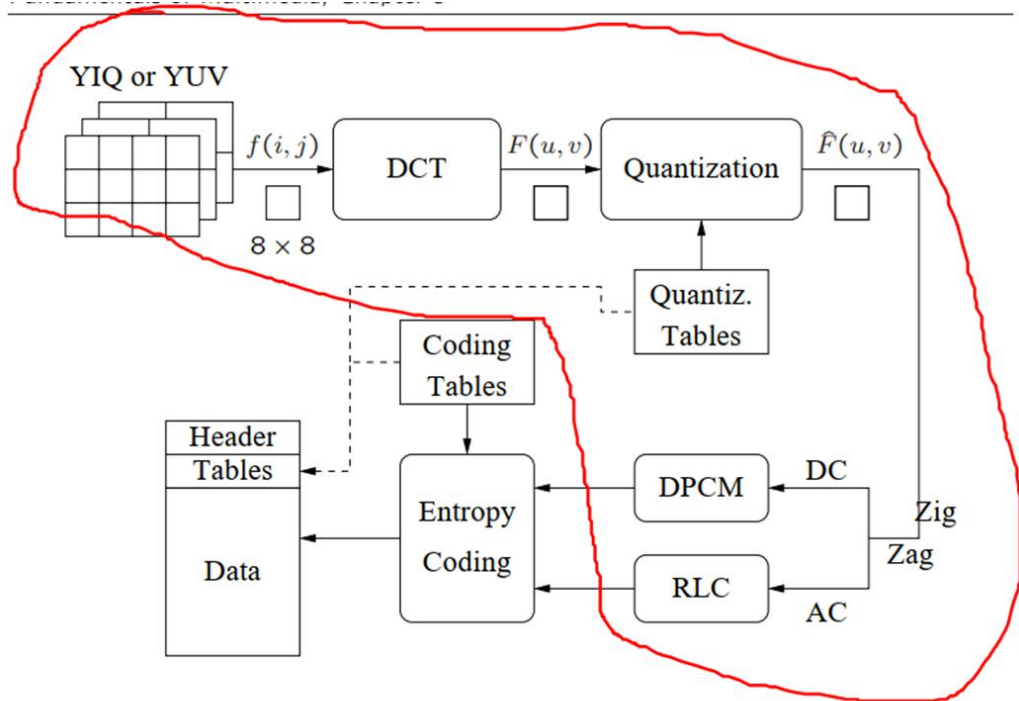
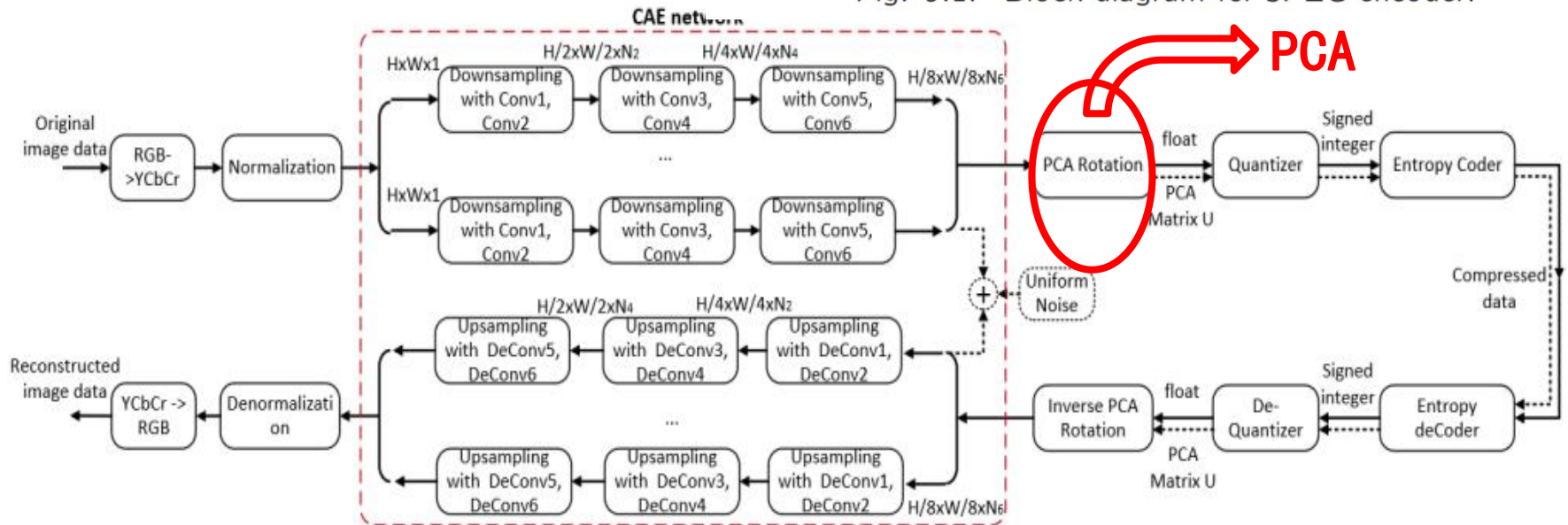


Fig. 9.1: Block diagram for JPEG encoder.



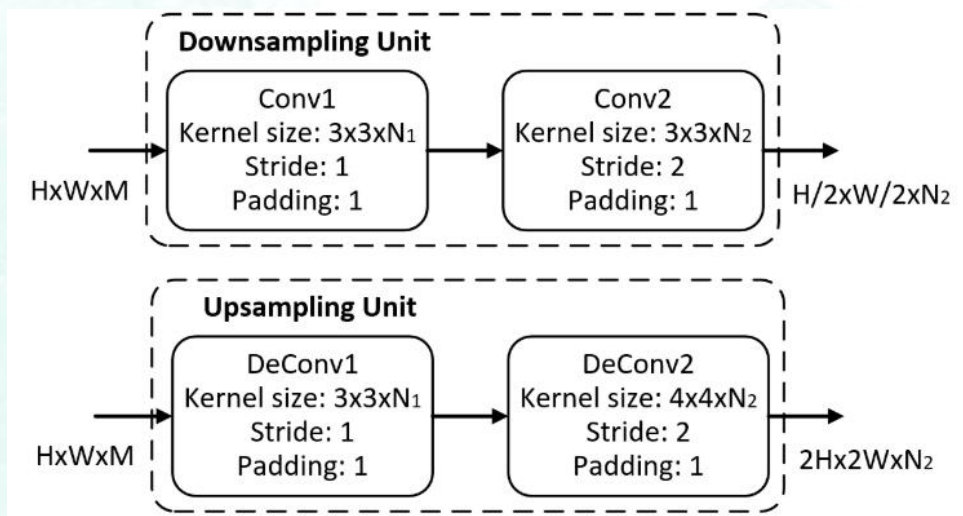
Cheng Z, Sun H, Takeuchi M, et al. Deep convolutional autoencoder-based lossy image compression[C]//2018 Picture Coding Symposium (PCS). IEEE, 2018: 253-257.

12.1.1 基于CNN的图像压缩

经典CNN图像压缩模型

基于卷积自动编码器CAE
的图像压缩框架

□ CAE



- 可视为一个编码的变换函数和一个解码的变换函数，卷积编码的过程是下采样处理，反卷积解码的过程是上采样处理，在编码压缩过程中，CAE的最终输出是图像的低维度特征图



12.1.1 基于CNN的图像压缩

经典CNN图像压缩模型

基于卷积自动编码器CAE
的图像压缩框架

□ PCA

- 目的：为了进一步提升压缩比，采用了PCA主成分分析法降维，使得特征图中生成更多零值。提升编码效率，便于图像压缩后续的量化和熵编码操作
- 主要步骤：
 - 计算CAE输出特征图的协方差
 - 计算这个协方差矩阵的特征向量，根据特征值的大小，按照列方向降序排列
 - 将特征图 y 乘以特征矩阵的转置，得到的输出可以更好的进行对角和横向扫描，并且右下角会产生更多的零，这有利于后面的熵编码器实现较大的压缩比





12.1.1 基于CNN的图像压缩

经典CNN图像压缩模型

模型	压缩程度	重建图像质量	方法创新点	引用
Generalized Normalization	一般	一般	首次将广义分歧归一化的方法引入到了端到端的图像压缩	【1】
ComCNN +RecCnn	一般	较好	模型较为复杂，训练难度较大，时间较长	【2】
End-to-End Learning	一般	较好	基于ResNet的架构基础上，引入了感知损失和对抗损失以便生成更清晰的重建图像	【3】
Virtual Codec	一般	较好	为了增加编码端与解码端的关联程度，引入了虚拟CNN进行训练	【4】
Residual CNN	一般	一般	使用CNN进行图像特征提取，在重构图像时使用SRGAN提高重建图像质量	【5】
CNNC	可控	较好	利用卷积与反卷积实现图像的压缩与重建，通过控制卷积层和池化层的层数控制压缩比	【6】





12.1.1 基于CNN的图像压缩

经典CNN图像压缩模型

- [1] Ballé J, Laparra V, Simoncelli E P. End-to-end optimized image compression[J]. arXiv: 1611.01704, 2016.
- [2] Jiang F, Tao W, Liu S, et al. An end-to-end compression framework based on convolutional neural networks[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2017, 28 (10) : 3007–3018.
- [3] Liu H, Chen T, Shen Q, et al. Deep image compression via end-to-end learning[C]//IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018: 2575–2578.
- [4] Zhao L, Bai H, Wang A, et al. Learning a virtual codec based on deep convolutional neural network to compress image[J]. Journal of Visual Communication and Image Representation, 2019, 63: 102589.
- [5] Deshmukh K, Pollett C. Residual CNN image compression[C]//International Symposium on Visual Computing. Cham: Springer, 2019: 371–382.
- [6] 崔建良, 李建飞, 陈春晓, 等. 基于 CNNC 的卷积神经网络图像的压缩方法[J]. 生物医学工程研究, 2019, 38 (4) : 415–419.

12.1.1 基于CNN的图像压缩

面临挑战

❑ 优化问题

- 在传统编码器的两端加入CNN，这两个CNN都需要通过训练来达到图像压缩和图像重建的目的，同时让两端进行联合优化，难度相对较高，训练不稳定

❑ 固定比率的图像压缩问题

- 传统的图像压缩方法往往能够定量地对图像压缩，如 JPEG 可以对图像进行 50 : 1 的压缩
- 基于 CNN 的图像压缩很少能够对图像进行固定比率的图像压缩

12.1.1 基于CNN的图像压缩

面临挑战

❑ 压缩图像分辨率问题

- 输入压缩图像的分辨率与训练时网络预定义的输入分辨率不一致时（如将 1024×1024 的图像输入 128×128 的训练模型），压缩后重建图像的效果往往不佳
- 为了适配更多不同分辨率的图像，往往需要训练不同的模型，或者提升网络预定义的分辨率大小，加深网络，但也增加了计算资源和时间的要求

12.1.2 基于RNN的图像压缩

RNN利于图像压缩的特性

□ 时序信息

- CNN 的参数共享是空间上的，而 RNN 则是时间上的，也就是序列上的，这使得 RNN 对于之前的序列信息有了“记忆”

□ 在压缩中的可控性和灵活性

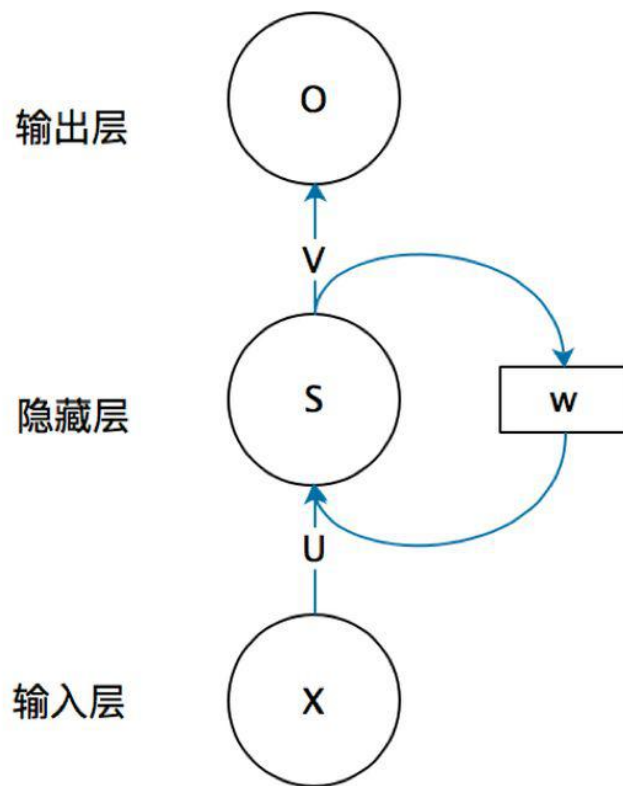
- 可以提高数据的压缩程度，也可以通过迭代的方式来控制图像的码率，都可以提高图像的压缩性能。因此应用 RNN 的图像压缩在对全分辨率图像压缩和通过码率来控制压缩比都取得了较为不错的成果

12.1.2 基于RNN的图像压缩

RNN

基本原理

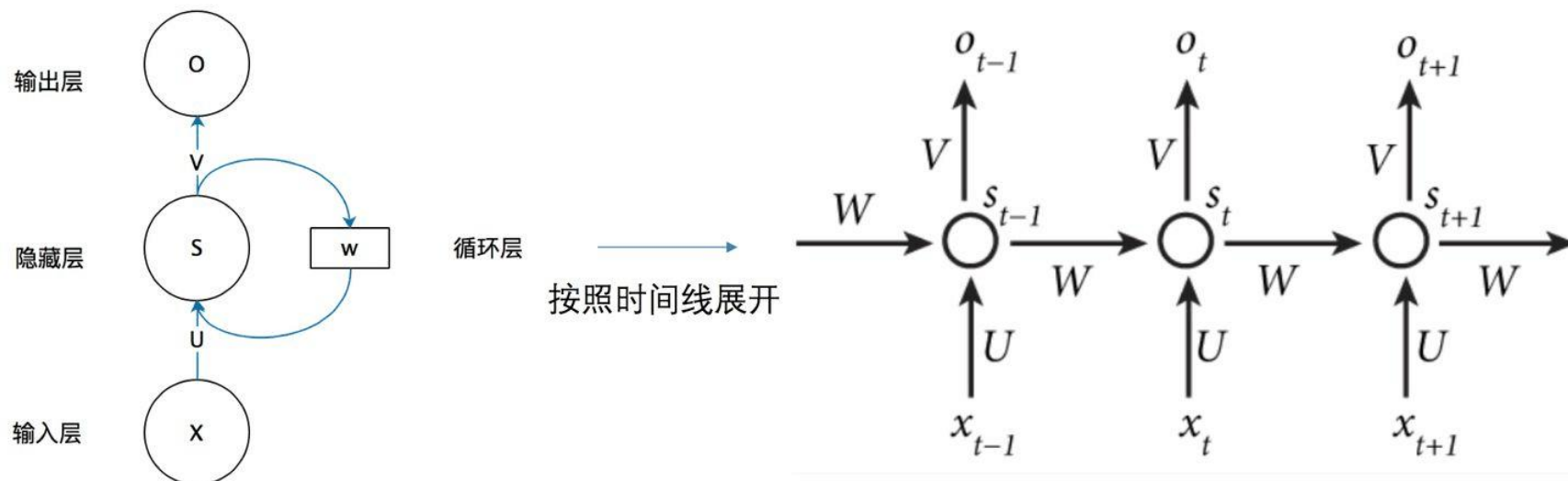
- 向量 x 表示输入层的；向量 s 表示隐藏层的值（可以是多个节点，节点数与向量 s 的维度相同）
- U 是输入层到隐藏层的权重矩阵，向量 o 表示输出层的值； V 是隐藏层到输出层的权重矩阵
- RNN隐藏层的值 s 不仅仅取决于当前这次的输入 x ，还取决于上一次隐藏层的值 s 。
权重矩阵 W 就是隐藏层上一次的值作为这一次的输入的权重
- 特点：可以处理序列信息



12.1.2 基于RNN的图像压缩

RNN

□ 基本原理



网络在 t 时刻接收到输入 x_t 之后，隐藏层的值是 S_t ，输出值是 O_t 。 s_t 的值不仅仅取决于 x_t ，还取决于 S_{t-1}

$$O_t = g(V \cdot S_t)$$

$$S_t = f(U \cdot X_t + W \cdot S_{t-1})$$

S_t 的值不仅仅取决于 X_t ，还取决于 S_{t-1}



12.1.2 基于RNN的图像压缩

经典RNN图像压缩模型

❑ 全分辨率的图像压缩

○ RNN引入的作用

- 模型迭代式地细化原始图像的重建，编码器和解码器都使用了RNN，使得更多的信息可以从一次迭代传递到下一次迭代。
- 每一次迭代都会给编码增加更多比特，从而实现更高质量的重建



12.1.2 基于RNN的图像压缩

经典RNN图像压缩模型

■ 全分辨率的图像压缩

○ RNN模块前向传播流程

- (1) 初始残差 $R[0]$ 对应于原始图像 I : $R[0]=I$
- (2) 第一次迭代, 设 $i=1$
- (3) 第 $[i]$ 次迭代以 $R[i-1]$ 作为输入, 并运行编码器和二值化器 (binarizer) 将该图像压缩成 $B[i]$
- (4) 第 $[i]$ 次迭代在 $B[i]$ 上运行解码器以生成一个重建出的图像 $P[i]$
- (5) 计算出第 $[i]$ 次迭代的残差: $R[i] = I - P[i]$
- (6) 设 $i=i+1$, 然后转到第 (3) 步 (直到达到所需的迭代次数)

压缩图像被表示成了 $B[1]$ 到 $B[N]$ 的级联。对于更大的 N 值, 解码器可在减少误差和生成更高质量的原始图像的重建上获得更多的信息

12.1.2 基于RNN的图像压缩

经典RNN图像压缩模型

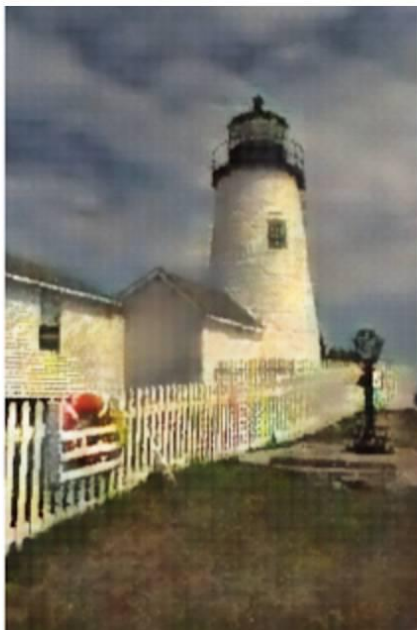
□ 全分辨率的图像压缩

○ RNN模块前向传播流程

残差图像被用作该网络的输入，以便为压缩图像的下一个版本移除压缩误差



原始图像 $I=R[0]$



重建的图像 $P[1]$



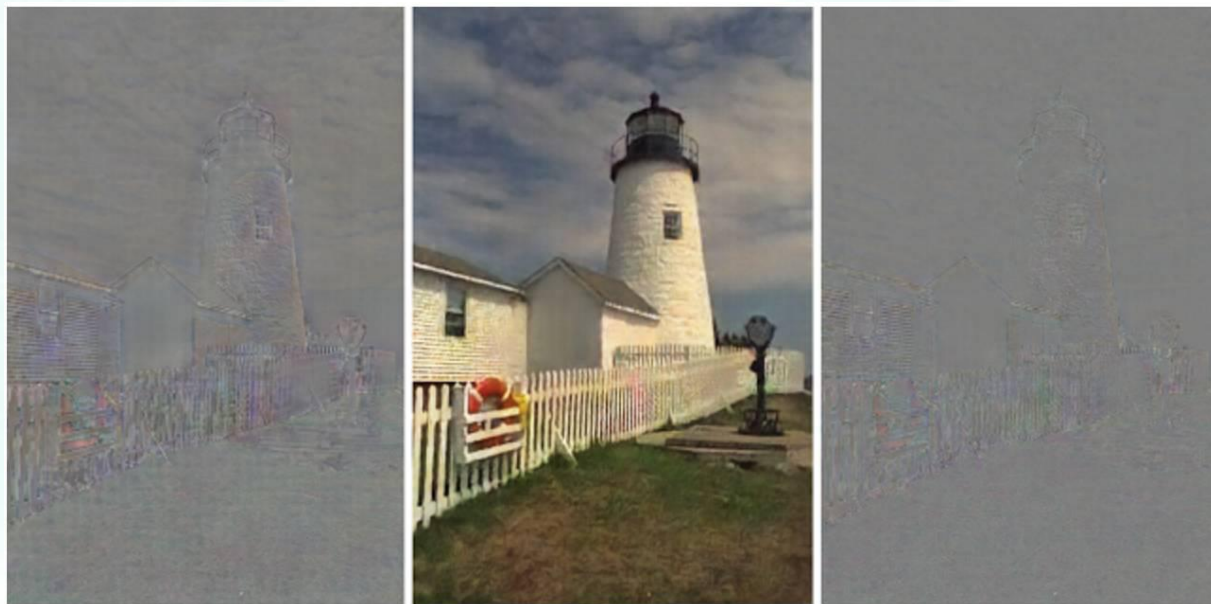
残差 $R[1]$ ，其表示
压缩中所产生的误差

12.1.2 基于RNN的图像压缩

经典RNN图像压缩模型

- 全分辨率的图像压缩
 - RNN模块前向传播流程

在更进一步的每次迭代中，该网络都会获得更多有关压缩中所引入的误差的信息。如果其可以使用这些信息来预测残差则可以得到更好的重建效果



用作输入的 $R[1]$ 更高质量的重建 $P[2]$ 原始图像减去 $P[2]$ 所得到的更小的残差 $R[2]$



12.1.2 基于RNN的图像压缩

经典RNN图像压缩模型

❑ 可变比特率的端到端学习图像压缩

- 实现了基于LSTM的端到端学习图像压缩，是利用RNN进行图像压缩具有代表性的方法之一
- 验证了任意的输入图像，在给定图像质量的情况下都能得到比目前最优压缩率更好的重建图像质量效果

Toderici G, O'Malley S M, Hwang S J, et al. Variable rate image compression with recurrent neural networks[J]. arXiv: 1511.06085, 2015.





12.1.2 基于RNN的图像压缩

经典RNN图像压缩模型

□ 全分辨率的图像压缩

- 基于残差块的剩余编码器和一个熵编码器，不仅能够捕捉图像中补丁之间的长期依赖关系并结合两种可能的方法来提高给定质量的压缩率
- 利用 RNN 梯度下降的训练方式，提出了一种基于全分辨率的有损图像压缩方法

Toderici G, Vincent D, Johnston N, et al. Full resolution image compression with recurrent neural networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 5306–5314.

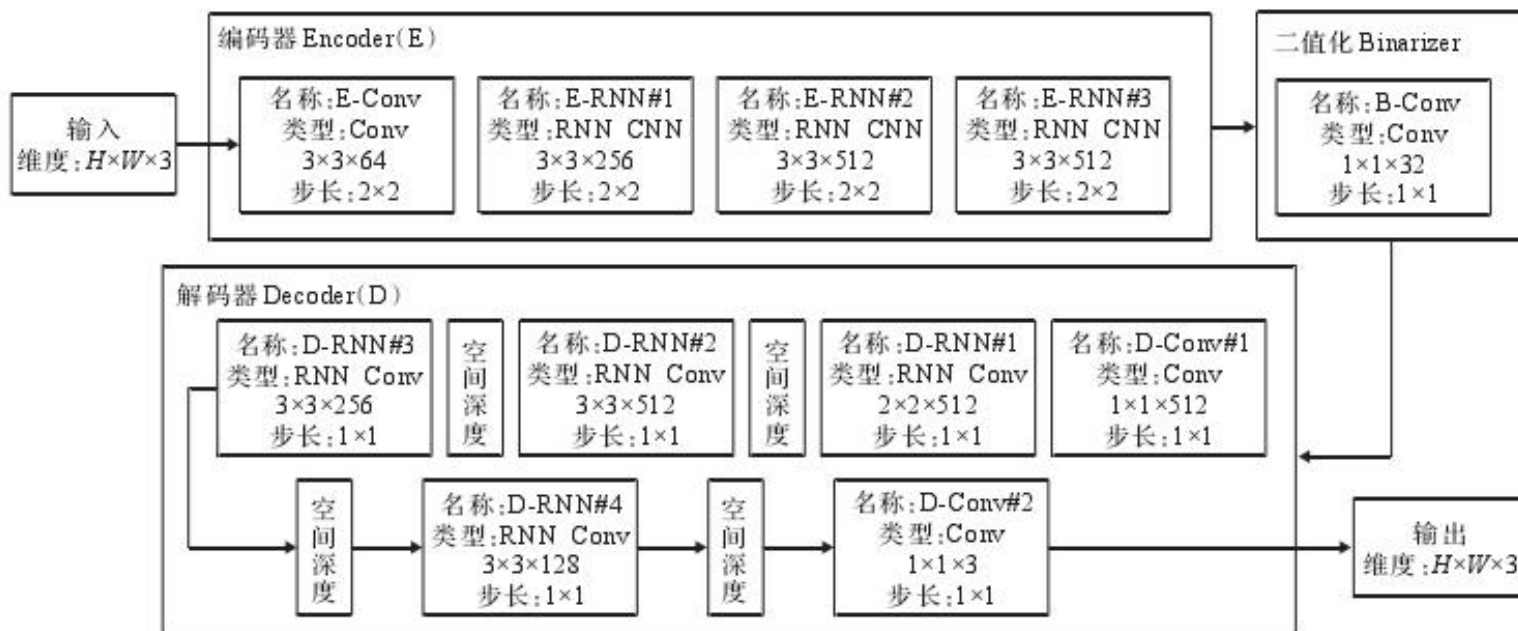


12.1.2 基于RNN的图像压缩

经典RNN图像压缩模型

全分辨率的图像压缩

- Decoder 解码部分使用卷积-循环网络的结构对信号进行迭代来恢复原图像，在迭代的过程中权值共享，并且每次迭代都会产生一个二值化的比特数，同时**在每次迭代中，网络从当前的残差中提取新的信息**，并将其与存储在循环层的隐藏状态中的上下文相结合，通过这些信息实现图像的重建





12.1.2 基于RNN的图像压缩

经典RNN图像压缩模型

模型	压缩程度	重建图像质量	方法创新点	引用
Variable Rate	可控	一般	首次提出了基于LSTM的端到端图像压缩，并且可以拓展到视频中	【1】
Full resolution	可控	优秀	实现了全分辨率的RNN图像压缩，模型结构复杂，训练较为困难	【2】

[1] Toderici G, O'Malley S M, Hwang S J, et al. Variable rate image compression with recurrent neural networks[J]. arXiv: 1511.06085, 2015.

[2] Toderici G, Vincent D, Johnston N, et al. Full resolution image compression with recurrent neural networks[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 5306–5314.





12.1.2 基于GAN的图像压缩

GAN利于图像压缩的特性

□ 对抗训练

- 通过对生成器和判别器的不断对抗，网络架构得到优化。GAN 根据这一特性，通过生成器的生成图像来不断“愚弄”判别器，使得最后得到的输出图像有更加清晰的纹理，更好的视觉感官效果

□ 模型优势

- 可以对全分辨率图像进行压缩，有较好的适用性
- 能够实现极限码率的图像压缩
- 重建图像的分辨率和清晰度高





12.1.2 基于GAN的图像压缩

经典GAN图像压缩模型

□ 基于GAN的实时自适应图像压缩

- 首次将 GAN 引入到图像压缩中，并且该算法在低码率条件下生成的文件要比传统的 JPEG 小 2.5 倍
- 该算法在率失真目标函数加入了一个多尺度对抗训练模型，使得重建图像与真实图像更加接近，即使在低比特率的情况下也能产生更清晰的图片
- 为基于GAN的图像压缩奠定了基础

Rippel O, Bourdev L. Real-time adaptive image compression[C] // Proceedings of the 34th International Conference on Machine Learning-Volume 70, 2017: 2922-2930.





12.1.2 基于GAN的图像压缩

经典GAN图像压缩模型

□ 生成压缩模型

- 将合成变换训练成模型，替代图像重建的优化
- 不仅能通过GAN生成高质量的图像，同时也与编码器进行了很好的结合，在编码器中加入GAN，通过不断优化网络结构得到更高质量的重建图像
- 生成图像有着极大的不稳定性，在生成图像时有可能生成的图像具有清晰的纹理，很好的视觉效果，很高的分辨率和清晰度，但与原图对比却可能存在明显差异，这也就形成一种欺诈性的清晰与高分辨率

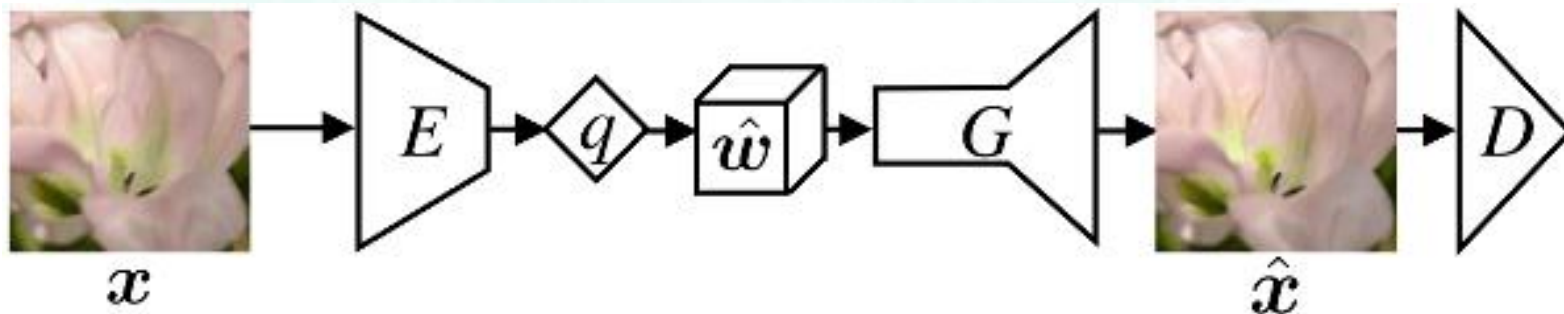


12.1.2 基于GAN的图像压缩

经典GAN图像压缩模型

基于语义标签映射的方法

- 在全分辨率的前提下实现了超低码率的极限压缩，同时也实现了在低码率时的高分辨率重建图像



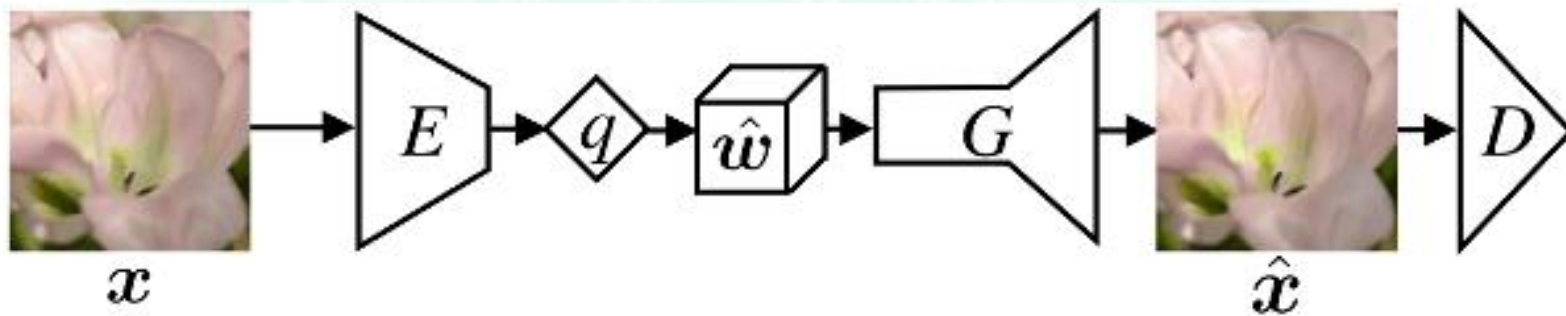
E 和 q 分别表示编码器和量化，
 \hat{w} 则代表一个压缩表示，
 G 和 D 分别为生成器与判别器

12.1.2 基于GAN的图像压缩

经典GAN图像压缩模型

□ 基于语义标签映射的方法

- 生成压缩用于保留图像的整体结构，以及不同尺度的图像结构
- 选择性地生成压缩用于从语义标签映射中完全生成图像的各个部分，同时保留用户定义的具有高度细节的区域
- 在两种方式共同作用下，保证重建图像的分辨率





Kodak Image 21



Ours (0.036bpp)



BPG (0.036bpp)



Kodak Image 22



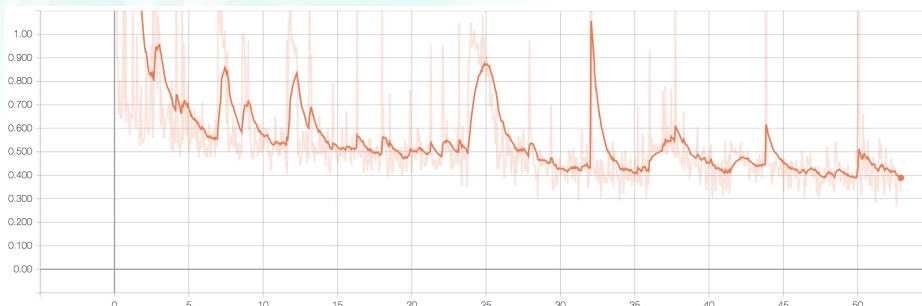
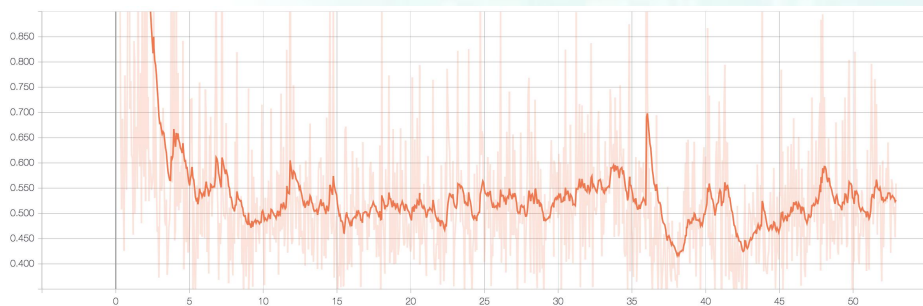
Ours (0.036bpp)



BPG (0.043bpp)

生成器损失

判别器损失





12.1.2 基于GAN的图像压缩

经典GAN图像压缩模型

模型	压缩程度	重建图像质量	方法创新点	引用
Real-time	较高	优秀	第一个采用 GAN 的多尺度实时性图像压缩	【1】
Generative Model	极高	优秀	首次使用 GAN 对图像进行极低比特率的图像压缩	【2】
Generative compression	较高	优秀	尽管有清晰的生成图像但是生成图像单一、不稳定，有可能与真实图像存在偏差，训练时优化困难	【3】
Extreme Learned	极高	优秀	可以通过语义提高重建图像分辨率	【4】
Extremely Low Bitrates	极高	一般	将自编码器变体与 GAN 结合，实现了极低比特率下的图像压缩	【5】

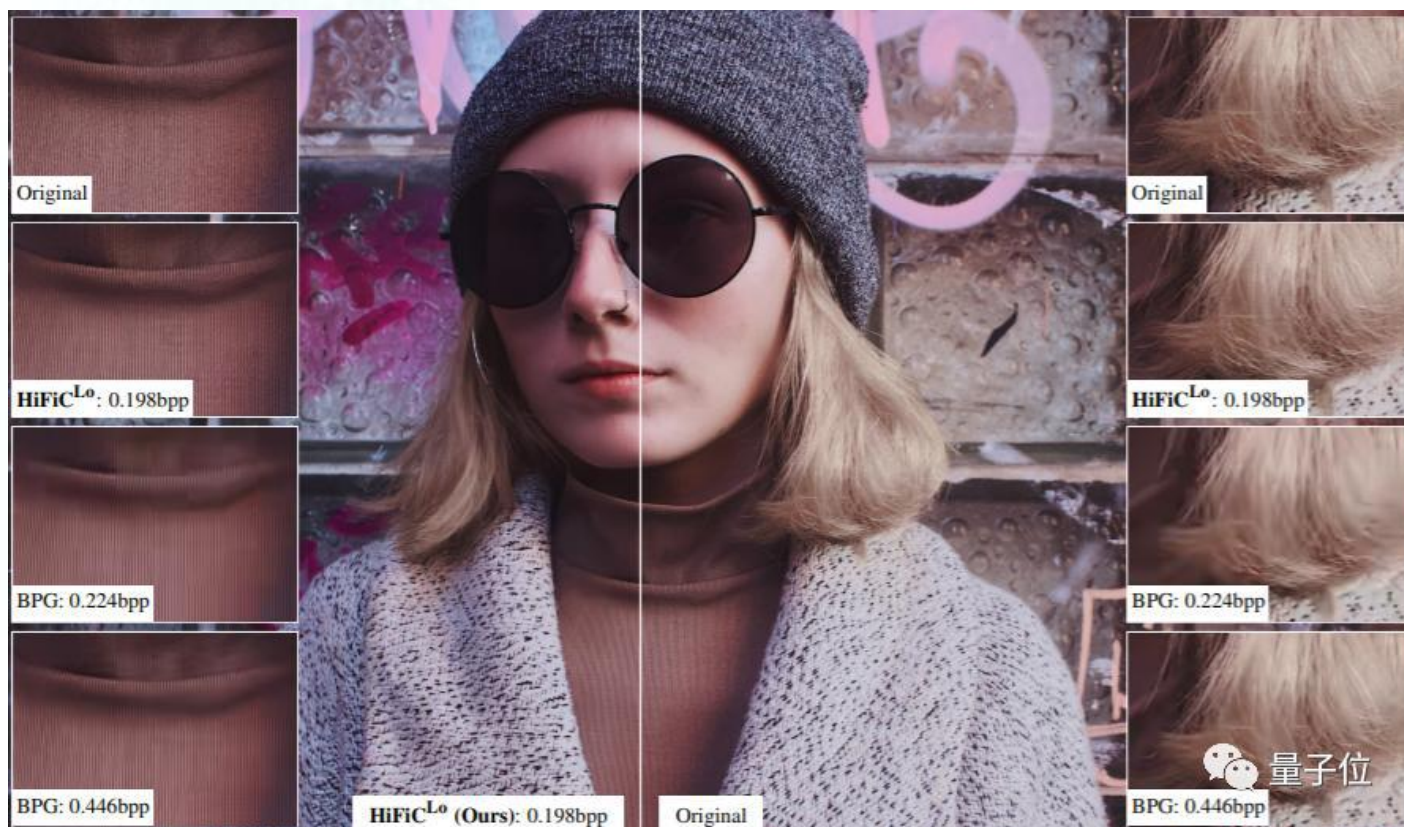


基于GAN的图像压缩方法



High-Fidelity Generative Image Compression

□ 工业界不断推出深度学习压缩新算法



左：压缩后图像 右：高清原图

Demo: <https://hific.github.io/>



74KB图片也高清，谷歌用神经网络打造图像压缩新算法

2020-09-06 20:38 来源：澎湃新闻·湃客

字号

量子位QbitAI | + 关注

□ 工业界不断推出深度学习压缩新算法



Samson Ootoovak @ootoovak · Jul 4

Replying to @eoinkelly

😲 Wow! If they can eventually do the same for video I'm sure Netflix and YouTube will be all over that.



推特网友评论：若能在视频压缩中实现类似效果，Netflix和YouTube等流媒体平台将大大获益

Demo: <https://hific.github.io/>





12.1.2 基于GAN的图像压缩

经典GAN图像压缩模型

[1] Rippel O, Bourdev L. Real-time adaptive image compression[C] // Proceedings of the 34th International Conference on Machine Learning–Volume 70, 2017: 2922–2930.

[2] Tschannen M, Agustsson E, Lucic M. Deep generative models for distribution-preserving lossy compression[C] // Advances in Neural Information Processing Systems, 2018: 5929–5940.

[3] Santurkar S, Budden D, Shavit N. Generative compression[C] // 2018 Picture Coding Symposium (PCS), 2018: 258–262.

[4] Agustsson E, Tschannen M, Mentzer F, et al. Generative adversarial networks for extreme learned image compression[C] // Proceedings of the IEEE International Conference on Computer Vision, 2019: 221–231.

[5] Raman S K, Ramesh A, Naganoor V, et al. CompressNet: generative compression at extremely low bitrates[C] // The IEEE Winter Conference on Applications of Computer Vision, 2020: 2325–2333.



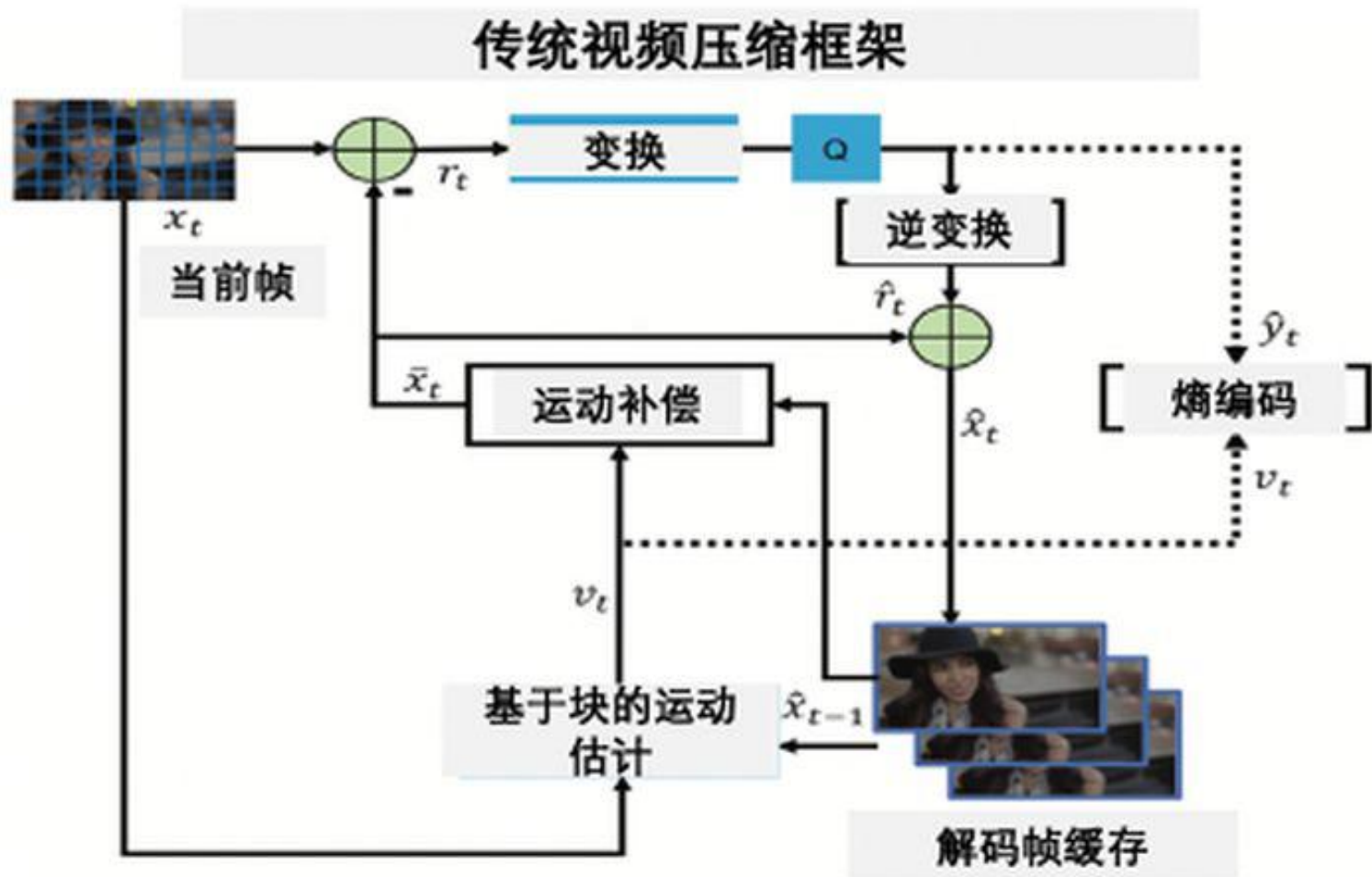
第二部分

12.2 基于深度学习的视频压缩技术



12.2 基于深度学习的视频压缩

传统视频压缩框架回顾



12.2 基于深度学习的视频压缩

传统视频压缩框架回顾

□ 编码器一侧的过程

输入帧 x_t 首先被分成一组相同大小的块（例如方形 8×8 块）

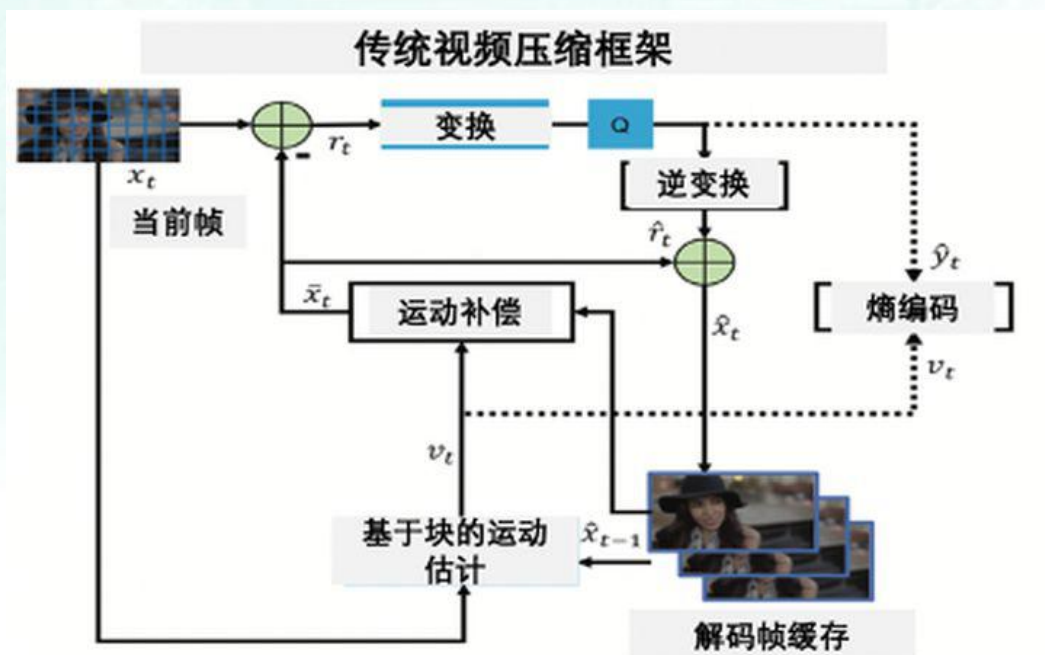
- 1) 运动估计。估计当前帧 x_t 与先前重构帧 x_{t-1} 之间的运动。获得每个块的对应运动矢量 v_t 。
- 2) 运动补偿。通过基于1) 步中的运动矢量 v_t 将先前重构帧的对应像素复制到当前帧来获得预测帧 \hat{x}_t 。获得原始帧 x_t 与预测帧 \hat{x}_t 之间的残差 r_t 。
- 3) 变换和量化。来自2) 步的残余 r_t 被量化为 y_t 。在量化之前使用线性变换（例如DCT）以获得更好的压缩性能。
- 4) 逆变换。3) 步的量化 y_t 逆变换得到重构残差 \hat{r}_t 。
- 5) 熵编码。1) 步的运动矢量 v_t 和3) 步的量化 y_t 通过熵编码成比特流送到解码器。
- 6) 帧重建。通过在2) 步添加 \hat{x}_t 并且在4) 步添加 \hat{r}_t 来获得重构帧 \hat{x}_t 。第 $(t + 1)$ 帧将使用重构帧进行1) 步的运动估计。

12.2 基于深度学习的视频压缩

传统视频压缩框架回顾

❑ 解码器一侧的过程

- 基于5) 步编码器提供的比特，执行2) 步的运动补偿，4) 步的逆量化，6) 步的帧重构，最后获得重建的帧 \hat{x}_t 。





12.2 基于深度学习的视频压缩

深度学习在视频压缩中的潜力

❑ 传统图像压缩算法的局限性.

- 传统的视频压缩工作需要手工设计很多模块
- 传统的视频压缩方法通常需要专门设计一些启发式的方法进行帧间预测

❑ 深度学习方法的优势.

- 能够实现更好的变换学习，从而取得更好的效果
- 端到端的深度学习模型能够自行学习
- 深度学习能够基于光流等进行预测，从而更好地减少帧间冗余





12.2.1 基于传统模块替换的方案

整体思路 采用深度学习替代传统视频编码中部分模块

□ 深度网络可替换模块

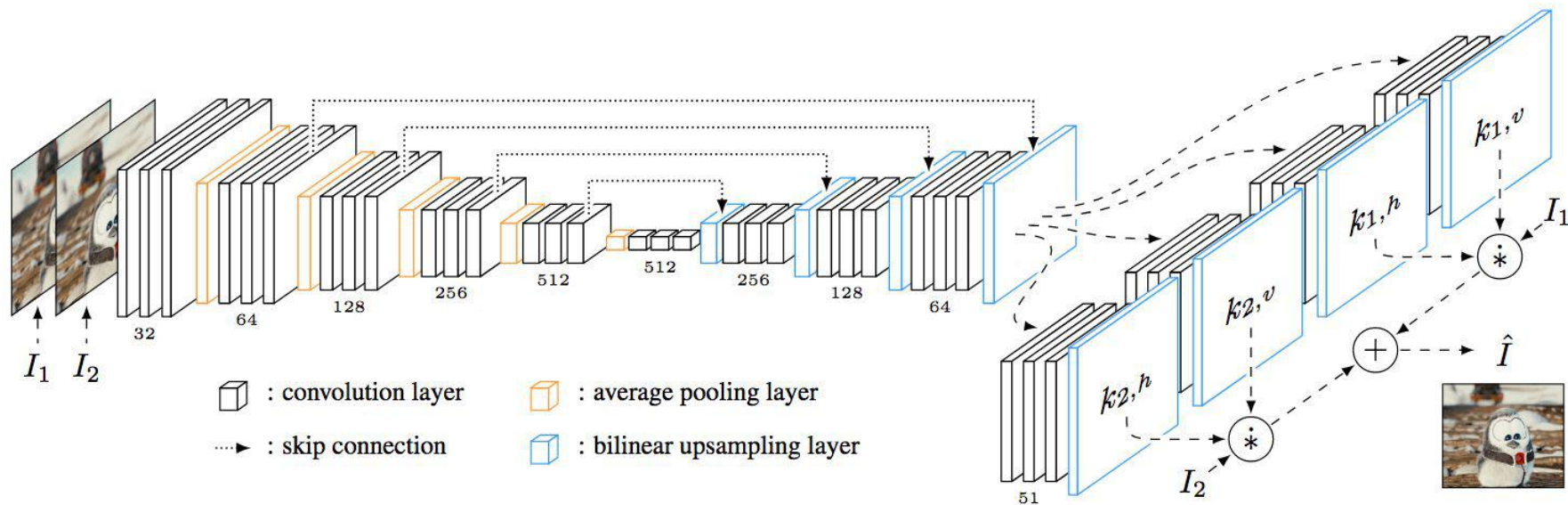
- 帧内/帧间预测
- 变换
- 上下采样
- 环路滤波
- 熵编码等



12.2.1 基于传统模块替换的方案

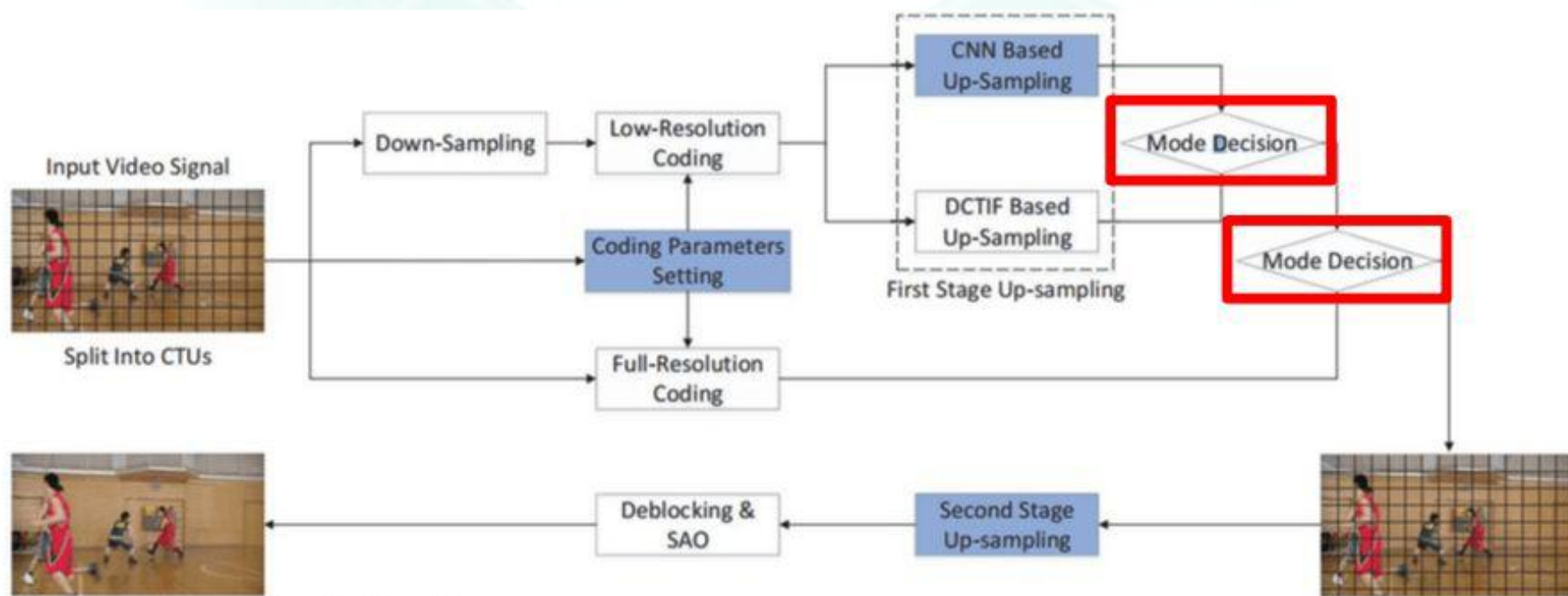
帧间预测

- 帧间预测能极大得减少帧间冗余，为参考帧和预测码字远小于每帧单独压缩的码字
- 帧间预测也是深度视频压缩方法与深度图像压缩方法的主要区别之一



下采样编码

- 对视频帧进行了分块处理，对适合进行下采样的块执行下采样操作，而对不适合进行下采样的块不执行下采样操作。

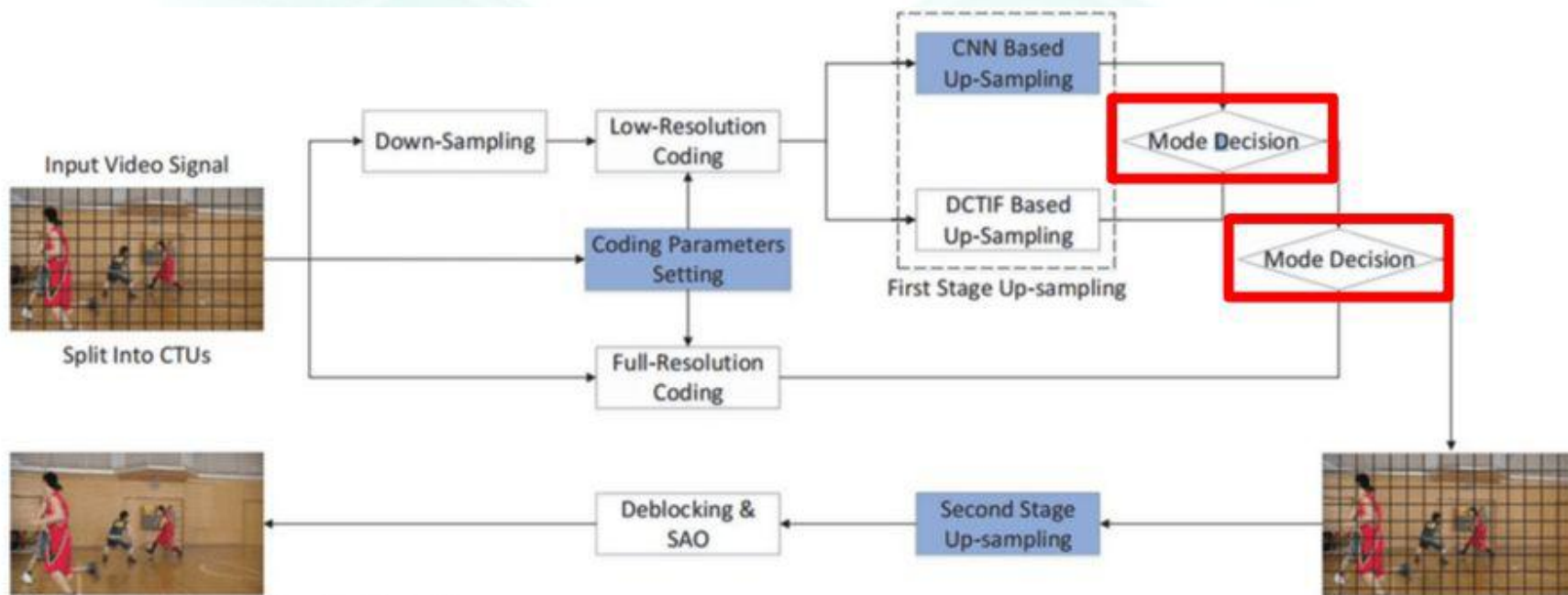


Y. Li et al., "Convolutional Neural Network-Based Block Up-Sampling for Intra Frame Coding," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 9, pp. 2316–2330, Sept. 2018, doi: 10.1109/TCSVT.2017.2727682.

12.2.1 基于传统模块替换的方案

下采样编码

- 对下采样的块根据情况分别使用CNN或DCTIF进行上采样，以重建图像。为了达到更好的效果，亮度通道和色度通道使用了不同的网络架构。

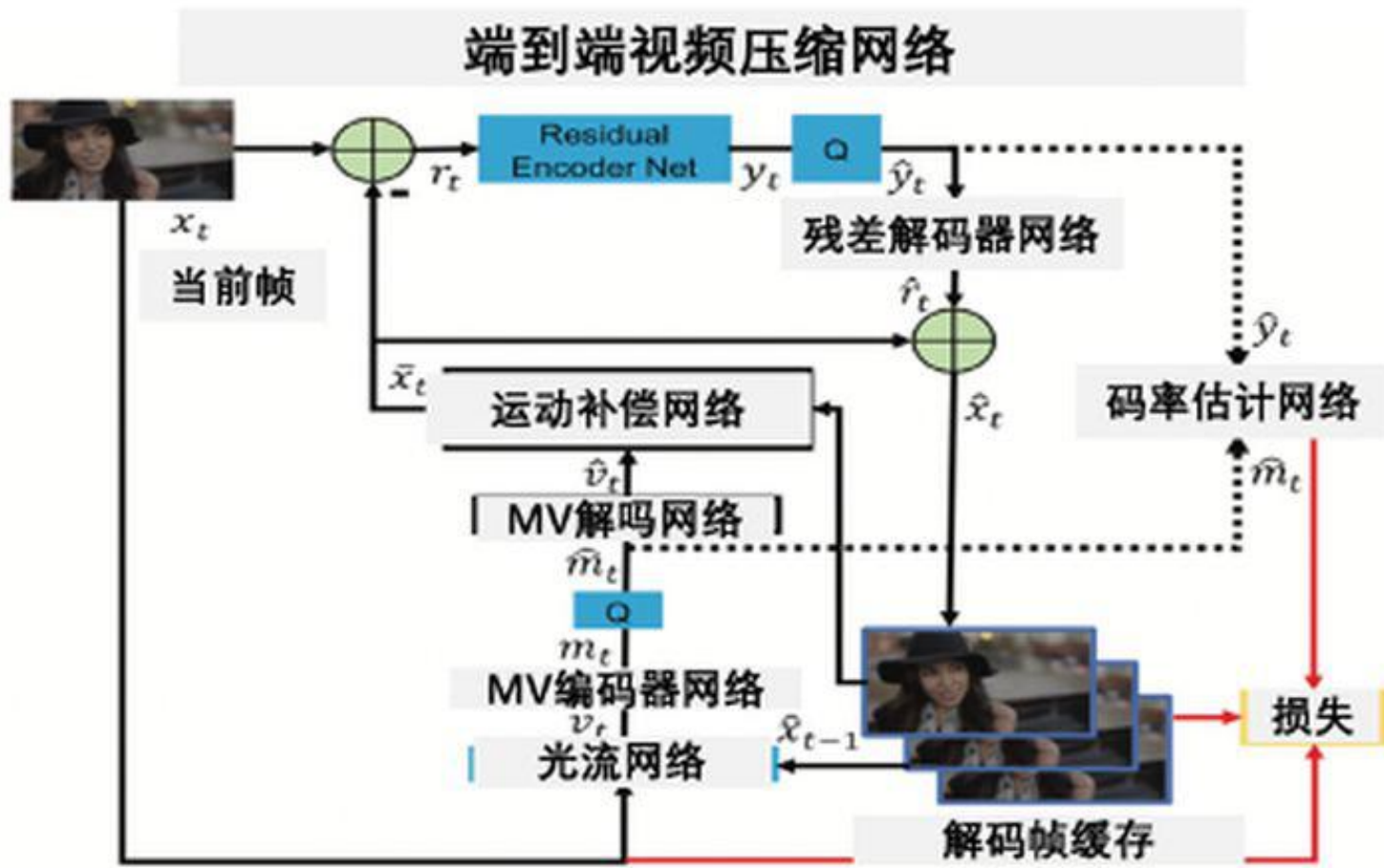


Y. Li et al., "Convolutional Neural Network-Based Block Up-Sampling for Intra Frame Coding," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 28, no. 9, pp. 2316-2330, Sept. 2018, doi: 10.1109/TCSVT.2017.2727682.

12.2.2 基于端到端学习的压缩方案

整体框架

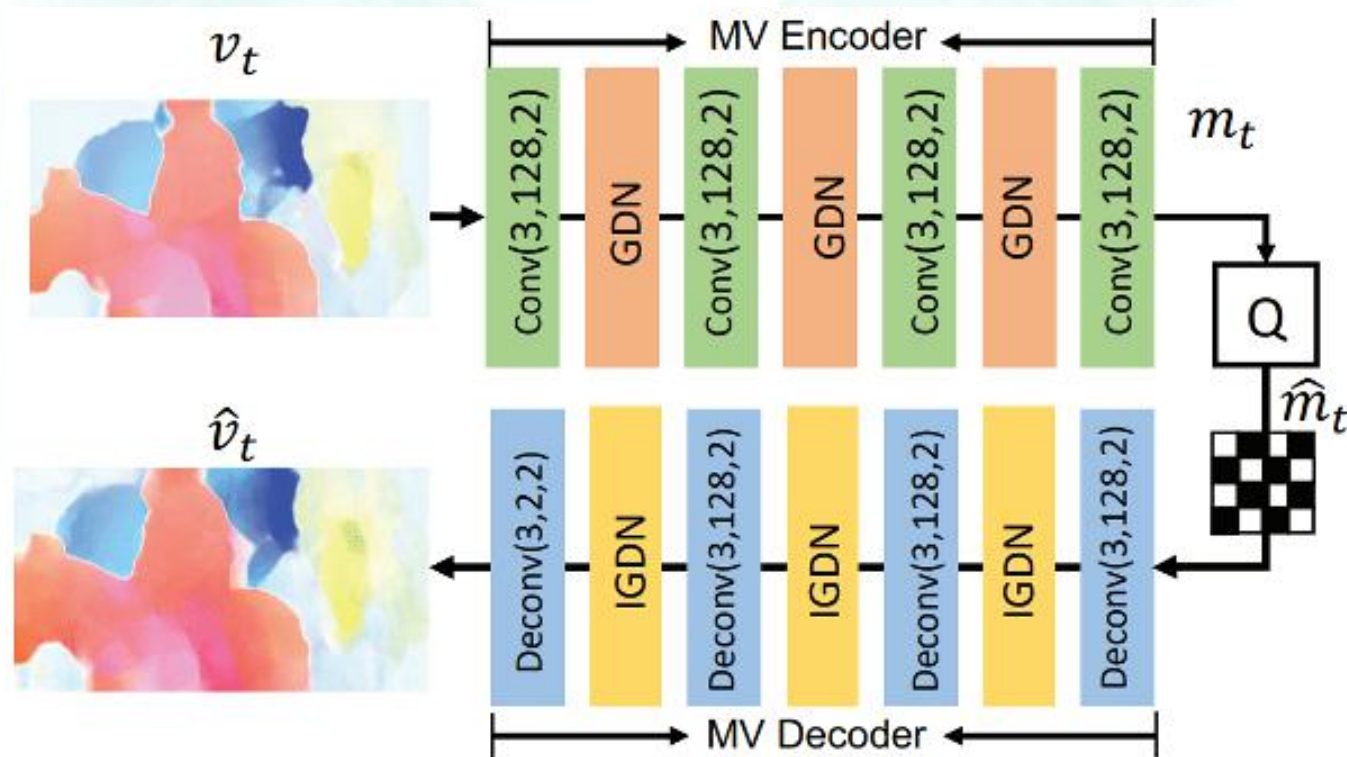
和传统方法是一一对应的



12.2.2 基于端到端学习的压缩方案

光流估计

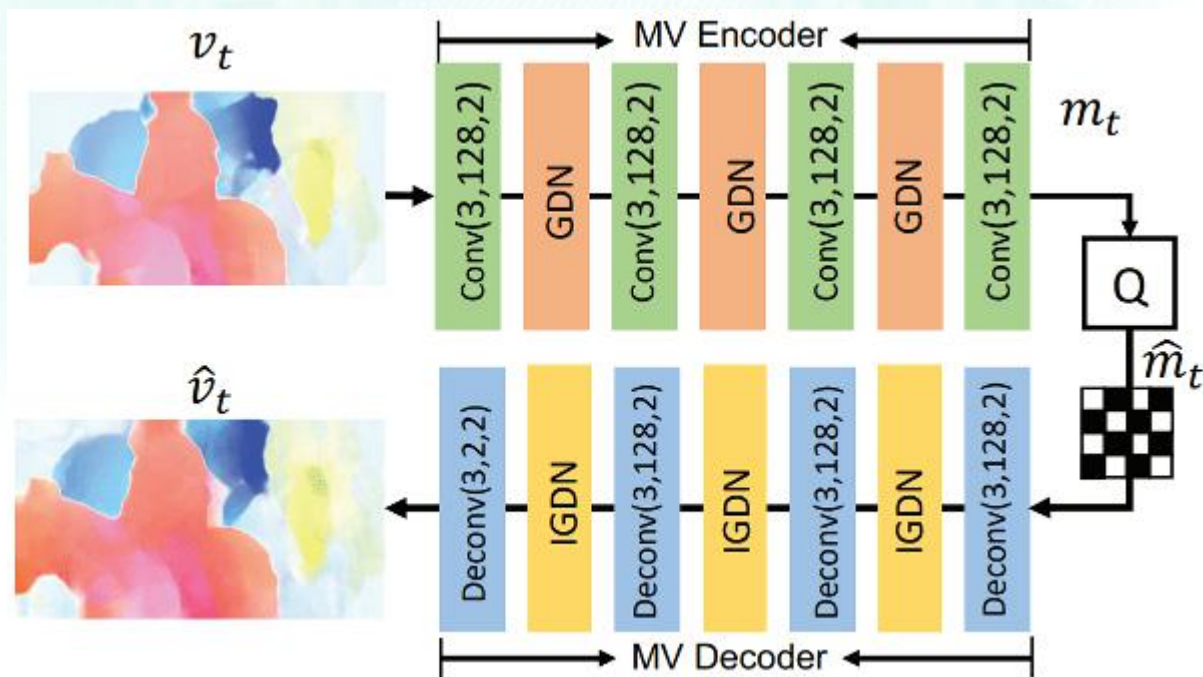
- 光流估计基于CNN模型，一个运动向量编码器-解码器网络（MV Encoder-decoder network）用来压缩和解码光流图



12.2.2 基于端到端学习的压缩方案

光流估计

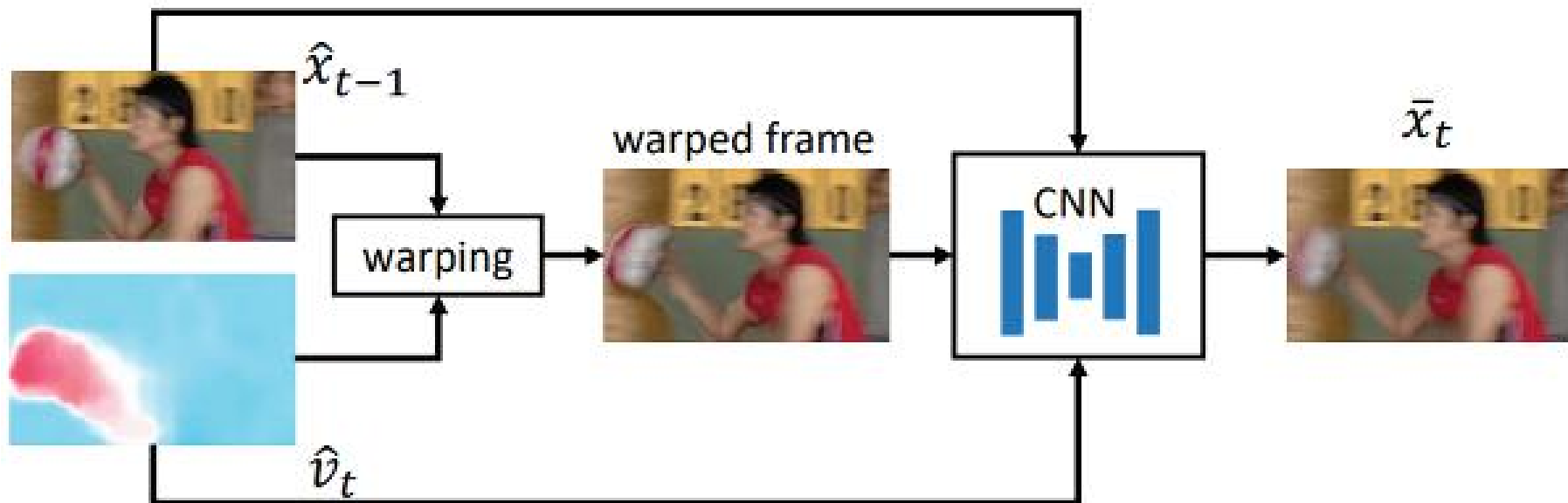
- 光流 v_t 被馈入一系列卷积运算和非线性变换。卷积（逆卷积）的输出通道数是128，除了最后一个逆卷积层的输出通道数等于2。给定 $M \times N \times 2$ 大小的光流 v_t ，MV编码器将生成 $M/16 \times N/16 \times 128$ 大小的运动表示 m_t 。然后 m_t 被量化为 \hat{m}_t 。MV解码器接收量化的表示并重建运动信息 \hat{v}_t 。另外，量化表示 \hat{m}_t 将用于熵编码



12.2.2 基于深度学习的视频压缩

运动补偿网络

- 给定先前重构的帧 \hat{x}_{t-1} 和运动向量 \hat{v}_t ，运动补偿网络获得预测的帧 \bar{x}_t ，期望尽可能接近当前帧 x_t ，其中CNN从输入的变形帧和光流产生预测帧



最后，高度非线性的残余编码器-解码器网络替换编码器一侧3) 步的线性变换（比如DCT），并且残差 r_t 非线性映射到表示 y_t 。然后， y_t 被量化为 \tilde{y}_t 。量化表示 \tilde{y}_t 被馈送到残余解码器网络获得重建残差 \tilde{r}_t