

基于量化误差估计模型的视频超分辨率重建算法

王春萌

(金陵科技学院计算机工程学院, 江苏 南京 211169)

摘要:提出了一种量化噪声的综合分布模型来估计量化过程带来的误差,并结合深度学习进行超分辨率重建。首先对数据库训练样本进行矫正量化误差的预处理,然后通过训练学习误差矫正后的低分辨率视频帧与高分辨率视频帧之间的特征,获得更加准确的映射关系,减少了量化过程带来的误差和质量损失。实验结果表明,本实验的算法不管在主观体验还是客观指标上都优于以前的算法。

关键词:超分辨率重建;量化误差;深度学习;卷积神经网络;视频编码

中图分类号:TP391.41 文献标识码:A 文章编号:1672-755X(2020)01-0010-06

Video Super-resolution Reconstruction Method Based on Quantization Error Estimation Model

WANG Chun-meng

(Jinling Institute of Technology, Nanjing 211169, China)

Abstract: This paper proposes a comprehensive distribution model for quantization noise to estimate the quantitative error and do the super-resolution reconstruction with deep learning. First, we make the preprocessing for all the frames in the training database to correct the quantization error, and then by training features between the low resolution frames and the high resolution frames after error correction, we acquire more accurate mapping relationship, and reduce the error and quality loss brought by quantitative process. Experimental results show that our algorithm is superior to the previous algorithms in both subjective experience and objective index.

Key words: super-resolution reconstruction; quantization error; deep learning; convolutional neural network; video coding

超分辨率(Super-resolution)重建是指将较低分辨率的图像或视频提升到较高分辨率的图像或视频,并且尽可能的恢复和重建低分辨率图像或视频所不能展示的细节信息。超分辨率重建问题是图像处理领域典型的逆问题(Inverse problem),其重建解是不唯一的。所以,需要最大化的利用本幅图像或样本库中其它图像的特征信息进行重建。

当前,超分辨率重建技术基本都是针对单幅图像进行的,而对压缩视频的研究较少。视频的超分辨率重建与单幅图像不同,在视频编码压缩过程中已经损失了很多原始数据,这个损失数据的过程被称为降质过程,其中量化是降质过程中很重要的一步,因为视频编码时的量化过程造成的失真和误差是不可逆的。

收稿日期:2020-01-22

基金项目:金陵科技学院高层次人才科研启动基金(jit-b-201802);江苏省高等学校自然科学研究面上项目(19KJB520007)

作者简介:王春萌(1987—),男,山东潍坊人,讲师,博士,主要从事计算机图像处理、数字媒体方面的研究。

本文针对基于深度学习的压缩视频超分辨率重建训练样本存在的量化降质问题,提出了一种量化噪声的综合分布模型来估计训练库视频帧在量化过程中产生的误差,通过训练学习误差矫正后的输入视频帧与输入高分辨率视频帧之间的特征,更加准确的得到高低分辨率样本之间的非线性映射关系,减少了量化过程带来的质量损失,使得重建后的图像质量优于以前的算法。

1 相关工作

单幅图像超分辨率重建技术已经成为当前学术界的一个研究热点。单幅图像的超分辨率重建方法主要有以下几类:

1) 插值法^[1]。包括邻近插值、双线性插值、双立方插值等。此类算法虽然简单,但效果较差,图像中的高频细节信息丢失严重。

2) 图像模型约束法。结合图像的先验知识进行还原重建。此类方法包括迭代反向投影法(Iterative back projection)^[2]、凸集投影法(POCS: Projection onto convex set)^[3]和最大后验概率法(MAP: Maximum a posteriori)^[4]等。这些方法一般存在迭代收敛慢,重建的高分辨率图像细节保持不够。另外,还有可能出现解不唯一的情况。

3) 稀疏表示法。这类算法基于如下思想:假定输入的低分辨率图像块可以被一个过完备的图像块字典的基元素线性稀疏表示,基于训练好的高低分辨率图像块字典,低分辨率图像块的稀疏表示可以应用到对应的高分辨率图像块字典来生成高分辨率图像块。典型的方法如 Yang J^[5]、曹翔等^[6]从压缩感知的角度实现超分辨率重建, Timofte R 等^[7]结合稀疏表示和最近邻域嵌入思想,提出了基于锚定邻域回归的超分辨率算法。此类方法重建效果较好,但是使用线性规划求解,速度较慢;而且重建的图像边缘细节质量不高。如何从庞大的样本集中训练出通用性更好的字典也是研究的难点。

4) 基于学习的超分辨率图像方法。比如基于卷积神经网络(CNN)的深度学习超分辨率算法可以直接学习低分辨率图像到高分辨率图像的端到端的映射函数。Dong C 等^[8]通过三层卷积神经网络 SRCNN 训练低分辨率和高分辨率图像块之间的特征得到它们之间的非线性映射,然后通过此映射关系进行重建得到高分辨率图像。此方法首次将 CNN 引入了超分辨率领域,取得了很好的效果。但是训练网络只有三层,重建图像质量不理想,而且训练过程需要长达一周的时间。后来提出的诸多基于深度学习的算法,如 Kim J 等^[9]提出采用非常深的网络进行学习的超分辨率重建方法(VDSR), Shi W 等^[10]提出在低分辨率图像上直接计算卷积得到高分辨率图像的方法(ESPCN), Ledig C^[11]提出的基于生成对抗网络(Generative adversarial networks, GAN)的超分辨率方法(SRGAN)等,这些算法都是受到 SRCNN 算法的启发,在此算法基础上的改进。本文的算法也是基于深度学习框架的超分辨率算法。

但是,当前的超分辨率重建技术基本是针对单幅图像进行的,而对压缩视频的研究较少。文献[12]将粒子滤波应用到视频超分辨率问题中,针对粒子匮乏问题改进基本粒子滤波算法,用修正匹配误差。文献[13]虽然也分析了视频超分辨率重建中量化噪声的分布,但并未提出统一的量化误差估计模型,也并未与深度学习相结合。而本文算法是从量化误差恢复的角度,并结合深度学习来改善超分辨率重建的质量。

2 算法模型描述

2.1 算法基本思想

以往的基于深度学习的图像超分辨率图像重建方法中,都是通过训练单幅低分辨率图像像素和对应的高分辨率图像像素直接的多层特征,得到两者之间的映射关系。对于输入的低分辨率图像,根据学习得到的关系预测得到高分辨率图像。对于压缩视频流的超分辨率重建,如 MPEG、H. 26X 等,输入的并不是单幅由像素组成的图像,而是压缩的视频比特流。压缩流中包含着量化后的变换系数,而量化会带来图像的降质。所以需要考虑不可逆量化过程带来的信息损失。

视频中每一帧的像素值都是通过解码过程得到的,其中解码中的反量化过程是根据码流中的量化因子等信息完成的,一旦完成反量化这一步,单凭每一帧图像的像素值无法恢复因为量化所丢失的信息,而

当前的基于空间域像素值的学习方法都没有利用码流中包含的量化信息。

为了恢复在编码量化过程中丢失的信息,本文提出了一种量化估计模型,根据压缩视频码流中包含的量化后的变换系数和量化因子 QP 值得到量化误差的分布模型,然后根据量化估计模型对训练数据库样本进行预处理,得到更加精确的反离散余弦变换(IDCT)后的空间域系数值,通过三层卷积神经网络重新训练学习高低分辨率精确的空间域系数值得到更加准确的映射关系,得到的重建图像更加准确。

2.2 数据库样本的精确建立

数据库样本中包含很多组低分辨率视频帧和对应的高分辨率视频帧,本文利用视频码流中的量化因子(QP)和离散余弦变换(DCT)系数等信息估计量化噪声的分布模型,然后根据分布模型得到更加准确的反 DCT 变换后的空域系数并保存,来减少量化过程带来的质量损失。对每一组高低分辨率视频都做同样的噪声估计过程。下面阐述量化噪声的估计过程。

对于压缩视频,由于 DCT 系数已经被量化,DCT 系数的精确值是未知的,但是量化间隔是已知的。量化是多对一的映射过程,将第 (k, l) 个原 DCT 系数记为 $Y[k, l]$,则量化过程把实值量化间隔 $[q_i^{k,l}, q_{i+1}^{k,l}]$ 映射到单一的实数量化后的系数 $Y_q[k, l] = Q[Y[k, l]]$ 。一般来说,DCT 量化误差在每一频率处是不相等的,对于平滑的信号,低频系数汇聚了信号的大部分能量,量化噪声较大;而高频系数携带较少的能量,大部分高频系数被量化为 0,因此量化噪声较小。由于 DCT 过程具有相关性,变换域的量化噪声可认为是不相关的随机变量。

在高码率时,由于量化因子 QP 较小,压缩程度较低,此时,可以认为量化噪声服从均匀分布,并且与输入的 DCT 信号不相关,即量化误差均匀的分布在量化边界内,那么此时第 (k, l) 个 DCT 系数的量化误差可以表示为:

$$\sigma_1[k, l] = \pm \left| \frac{q_{i+1}^{k,l} - q_i^{k,l}}{QP} \right| \quad (1)$$

其中 $\sigma_1[k, l]$ 的符号由 $q_{i+1}^{k,l}$ 和 $q_i^{k,l}$ 的大小决定。

在低码率时,由于量化因子 QP 较大,压缩程度较高,量化间隔较大,此时需要把 DCT 系数的先验知识加入到模型中进行噪声的精确估计,DCT 系数是频域信号,根据 Lam E Y 等^[14]对 DCT 系数的分布进行的研究,此时的系数可近似为拉普拉斯模型,那么第 (k, l) 个 DCT 系数的量化误差可表示为:

$$\sigma_2[k, l] = \gamma \int_{q_i^{k,l}}^{q_{i+1}^{k,l}} (Y_q[k, l] - y) p_{Y[k, l]}(y) dy \quad (2)$$

其中, γ 是保证分布归一化的常数, $p_{Y[k, l]}(y)$ 是拉普拉斯分布函数。

本文以 H. 264 编码格式为例,当量化参数 QP 取最小值 0 时代表最精细的量化,码率最高;当 QP 取最大值 51 时代表最粗糙的量化,码率最低。本文将两种码率下的模型通过一个权重函数统一起来。令 $\lambda = QP/51$,第 (k, l) 个 DCT 系数的量化误差可以归纳为:

$$\sigma[k, l] = (1 - \lambda)\sigma_1[k, l] + \lambda\sigma_2[k, l] \quad (3)$$

估计出每个 DCT 系数的量化误差后,则可以在原始量化后的 DCT 系数基础上加上量化误差来计算精确的 DCT 系数,公式如下:

$$\hat{Y}[k, l] = Y_q[k, l] + \sigma[k, l] \quad (4)$$

其中, $\hat{Y}[k, l]$ 为经过量化误差修正的更加精确的 DCT 系数。

最后,对进行过量化模型估计的精确的 DCT 系数进行反 DCT 变换,将新的系数从频域值转换为空域值,即解码得到精确的空间域像素值。采用新的像素值更新原始的深度学习数据库中每帧图像的像素值,采用更新后的像素值进行后续的学习和重建过程。

2.3 学习训练过程

本文采用的 Dong C 等^[8]的方法进行学习训练过程,学习的目标是建立一个映射函数 $F(Y)$ 来重建 Y 使其尽可能接近理想的高分辨率图像 X 。 Y 是原始低分辨率图像采用双三次插值得到的中间图像,此图像与高分辨率图像 X 具有相同的尺寸。映射函数 $F(Y)$ 通过基于三层卷积神经网络(CNN)的训练模型学

习得到,三层网络分别为特征提取层、非线性映射层和重建层。三个卷积层的卷积核尺寸分别为每个卷积层跟随一个激活函数,本文采用 ReLU 函数作为激活函数,采用此激活函数的收敛效果更好而且训练过程较快。

通过学习端到端的映射函数 $F(Y)$ 来估计需要的网络参数 $\theta = \{W_1, W_2, W_3, B_1, B_2, B_3\}$, 其中 W_i 和 B_i 分别为每层深度网络的权重值和偏移量,映射函数的生成过程其实是这些参数的获得过程。数据库中采用我们的噪声估计方法得到的新的视频帧图像。对于损失函数的定义,本文采用均方误差(MSE)作为损失函数,有利于获得较高的峰值信噪比(PSNR)。PSNR 是当前使用最广泛的定量评价图像恢复质量的指标,它在一定程度上反映了图像的感知质量。损失函数如公式(5)所示。

$$L(\theta) = \frac{1}{n} \sum_{i=1}^n \|F(Y_i; \theta) - X_i\|^2 \quad (5)$$

其中, $L(\theta)$ 为损失函数, n 表示训练数据采样像素点的数量。 Y 为新的低分辨率图像值, X 为数据库中对应的高分辨率图像。

本文的训练数据库是 VideoSet 数据集^[15], 这个数据集包含了多种分辨率多种码率的不同场景的视频序列, 每个序列长度为 5 s。从中选取了 1 000 帧不同场景的图像作为训练集。对每幅帧图像进行高斯模糊, 然后进行宽高像素分别减半的下采样, 作为低分辨率输入图像。这样每个场景包含了一幅高分辨率图和对应的低分辨率图。

2.4 重建过程

输入一个低分辨率视频帧, 经过加入量化噪声估计模型校正的解码过程以后, 生成精确的低分辨率视频帧。将低分辨率视频帧使用双三次插值得到重建所需的高分辨率尺寸, 得到初始的插值后的中间图像 Y 。将 Y 输入到训练学习得到的非线性映射函数 $F(Y)$, 得到高分辨率的视频帧。

3 实验结果

本文的算法在主观和客观两个方面与当前主流的超分辨率算法, 包括双三次插值 Bicubic 算法和 SRCNN 算法^[8], 分别进行了对比实验, 并以原始高分辨率图像做参照。对于双三次插值算法, 本文采用了 FFMPEG 中封装的 Bicubic 上采样命令。对于 SRCNN 算法, 本文选择了原方法默认参数进行实验, 而且采用了与本文方法相同的训练数据库。

本文的超分辨率实验都是帧图像长和宽分别增倍(超分辨率倍数 $Scale=2$), 即输出分辨率大小是输入分辨率的 4 倍。本文对 10 个视频片段进行了实验对比, 这些视频片段取自多个专业视频数据库, 如 URL: <http://demo-uhd3d.com> 和 VideoSet 数据集。实验视频片段的输入输出分辨率分为 3 个档次, 其中输入视频帧的分辨率分别为 540×360 、 960×540 和 $1\ 920 \times 1\ 080$, 对应的输出视频帧的分辨率分别为 $1\ 280 \times 720$ 、 $1\ 920 \times 1\ 080$ 和 $3\ 840 \times 2\ 160$ 。

3.1 主观对比

主观对比我们主要采用了逐帧对比的方式, 即截取不同算法生成的同一帧图像, 对图像质量对比观察, 如图 1 所示。根据对比结果看出, Bicubic 插值算法的结果模糊现象很明显, 细节丢失较为严重; SRCNN 算法重建效果较好, 但是也丢失了一部分细节; 而本文的算法不仅保持了最多的细节, 而且视觉效果良好。

3.2 客观对比

本文对 10 个视频片段分别统计每帧的客观质量指标 PSNR 和 SSIM (Structural similarity index), 然后统计出每个视频所有帧的平均 PSNR 和平均 SSIM, 并与其它算法进行了客观数据对比。PSNR 和 SSIM 是当前衡量超分辨率质量应用最广泛的客观指标。表 1 和表 2 列出了分别采用不同超分辨率重建方法的平均 PSNR 和平均 SSIM 值, 对每个序列, 3 种算法的最高值采用了加粗显示。通过实验数据对比, 对于绝大部分序列来说, 本文算法的平均 PSNR 和平均 SSIM 值都比 Bicubic 算法和 SRCNN 算法要高。通过客观指标对比数据也证明了本文算法的质量优势。

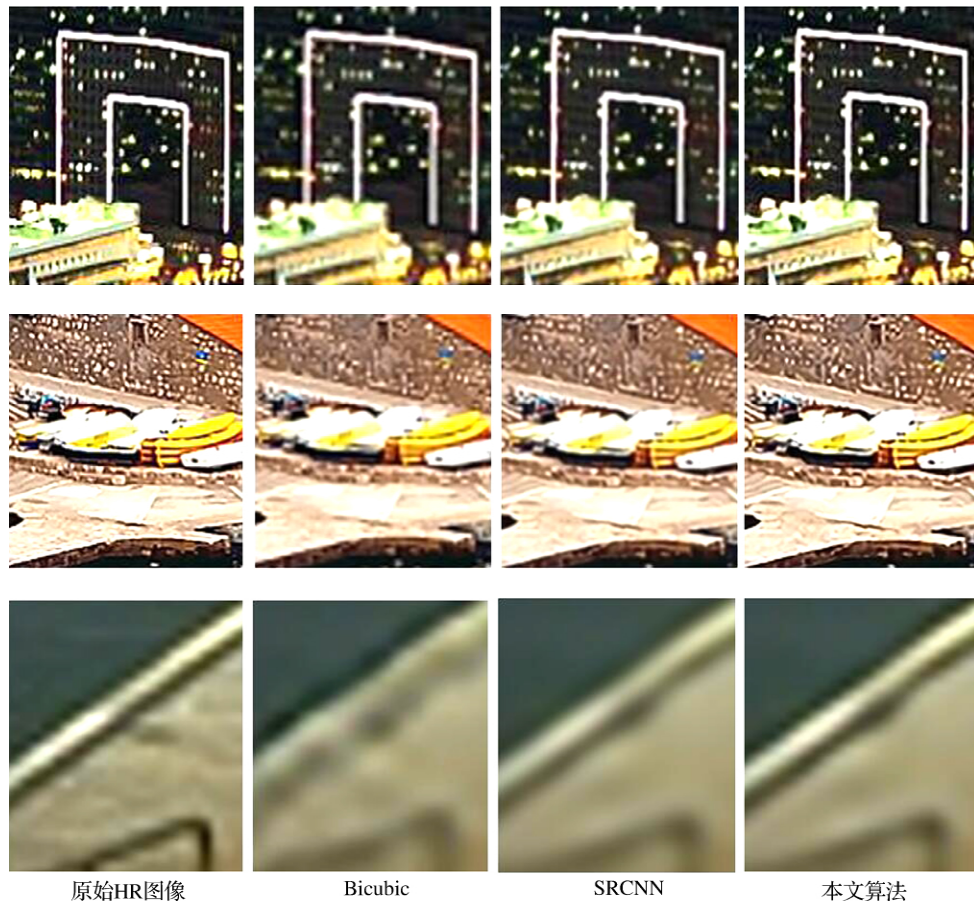


图1 三种超分辨率算法的对比结果

表1 不同方法的平均 PSNR 值对比结果

视频片段	Bicubic	SRCNN	本文算法
Video1	38.58	41.59	41.79
Video2	36.39	38.81	39.21
Video3	37.71	40.37	40.65
Video4	52.46	53.05	53.33
Video5	50.55	51.12	51.43
Video6	35.29	35.84	36.18
Video7	37.08	37.69	37.82
Video8	39.31	40.27	40.49
Video9	38.19	39.65	40.01
Video10	37.24	38.12	38.34

表2 不同方法的平均 SSIM 值对比结果

视频片段	Bicubic	SRCNN	本文算法
Video1	0.973 1	0.976 2	0.978 1
Video2	0.976 6	0.977 2	0.977 5
Video3	0.998 7	0.998 8	0.998 8
Video4	0.974 5	0.974 8	0.974 9
Video5	0.985 4	0.989 1	0.989 5
Video6	0.981 9	0.983 7	0.984 1
Video7	0.978 9	0.979 2	0.979 1
Video8	0.974 5	0.979 1	0.979 5
Video9	0.992 5	0.993 2	0.994 0
Video10	0.983 7	0.984 6	0.985 2

4 结 语

本文将深度学习的训练数据库进行预处理,使得样本值更加精确。本文充分利用了数据库视频帧中所提供的量化信息,根据量化噪声的分布情况,采用一个统一的模型对不同码率的量化噪声进行估计,得出较为精确的量化前的 DCT 系数,然后经过反 DCT 变换得到更加精确的样本帧的空域像素值。在量化误差减少的更加精确的样本值上进行训练,使得学习得到的映射关系也更加精确,提高了超分辨率重建的

质量。后续试验将算法扩展到更高的超分辨率倍数,如 3 倍或 4 倍,并将本文的思想扩展到视频复原和视频去噪等领域,这需要同样的方法对训练数据库进行预处理。另外,我们开发适于 GPU 并行的深度网络架构,进行超分辨率重建的加速。

参考文献:

- [1] Lertrattanapanich S, Bose N K. High resolution image formation from low resolution frames using delaunay triangulation [J]. IEEE Transactions on Image Processing, 2002, 11(12): 1427-1441
- [2] Rasti P, Demirel H, Anbarjafari G. Improved iterative back projection for video super-resolution [C]// Signal Processing and Communications Applications Conference. State of Florida: Science and Technology Press, 2014: 552-555
- [3] Stark H, Oskoui P. High-resolution image recovery from image-plane arrays using convex projections [J]. Journal of the Optical Society of America A, 1989, 6(11): 256-261
- [4] Hardie R C, Barnard K J, Armstrong E E. Joint MAP registration and high-resolution image estimation using a sequence of undersampled images [J]. IEEE Transactions on Image Processing, 1997, 6(12): 1621-1633
- [5] Yang J, Wright J, Huang T S, et al. Image super-resolution via sparse representation [J]. IEEE Transactions on Image Processing, 2010, 19(11): 369-372
- [6] 曹翔, 陈秀宏, 潘荣华. 基于稀疏表示的快速图像超分辨率算法 [J]. 计算机工程, 2015, 41(6): 211-215
- [7] Timofte R, De V, Gool L V. Anchored neighborhood regression for fast example-based super-resolution [J]. IEEE International Conference on Computer Vision, 2013, 85: 1920-1927
- [8] Dong C, Chen C L, He K, et al. Image super-resolution using deep convolutional networks [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 38(2): 295-307
- [9] Kim J, Lee J K, Lee K M. Accurate image super-resolution using very deep convolutional networks [J]. IEEE Conference on Computer Vision & Pattern Recognition, 2016, 72: 1646-1654
- [10] Shi W, Caballero J, Huszar F, et al. Real-time single image and video super-resolution using an efficient subpixel convolutional neural network [J]. IEEE Conference on Computer Vision and Pattern Recognition, 2016, 78: 1874-1883
- [11] Ledig C, Theis L, Huszar F, et al. Photo-realistic single image super-resolution using a generative adversarial network [J]. IEEE Conference on Computer Vision and Pattern Recognition, 2017, 72: 105-114
- [12] 王爱侠, 赵越. 基于改进粒子滤波算法的视频超分辨率重建 [J]. 计算机工程, 2015, 41(4): 263-266
- [13] 徐忠强, 朱秀昌. 压缩视频超分辨率重建技术 [J]. 电子与信息学报, 2007, 29(2): 499-505
- [14] Lam E Y, Goodman J W. A mathematical analysis of the DCT coefficient distributions for images [J]. IEEE Transactions on Image Processing, 2000, 9(10): 1661-1666
- [15] Wang H, Katsavounidis I, Zhou J, et al. VideoSet: A large-scale compressed video quality dataset based on JND measurement [J]. Journal of Visual Communication & Image Representation, 2017, 46: 295-306

(责任编辑:谭彩霞)