

文章编号: 2095-2163(2019)05-0001-07

中图分类号: TP391.41

文献标志码: A

基于卷积神经网络的视频软广播

尹文斌, 范晓鹏

(哈尔滨工业大学 计算机科学技术学院, 哈尔滨 150001)

摘要: 随着信息技术和互联网技术的不断发展, 无线视频广播越来越受到人们的欢迎, 成为流行的多媒体应用之一。然而, 传统的数字编码和传输方法很难适应于向多个具有不同信道质量的用户同时发送视频的场景, 通常会遭遇悬崖效应。近期, 一种新颖的无线视频广播方法称为 SoftCast 被提出, 其保存在信道中传输的信号与视频像素值之间所具有线性关系并利用有效的能量分配方法, 使得视频重构质量随着信道噪声的增加而平缓下降。在本文中, 提出了一种新型的无线视频广播方法, 其利用深度卷积网络和基于图像组的稀疏表示模型, 通过解码端估计的信道质量, 优化视频的解码过程并减轻多种由信源编码和信道噪声造成的视觉失真。通过视频软传输技术, 本文提出的方法具有出色的视频广播质量可伸缩性并避免了悬崖效应的发生, 同时还能提供视觉友好的主客观重构质量。实验结果表明, 本文提出的方法在视频广播场景下能够获得优于传统 SoftCast 最高 1.2 dB 的重建质量。

关键词: 无线视频广播; 卷积神经网络; 基于图像组的稀疏表示; 视频软广播

Convolutional Neural Networks based soft video broadcast

YIIN Wenbin, FAN Xiaopeng

(School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)

[Abstract] With the continuous development of information technology and Internet technology, video broadcasting is becoming more and more popular in wireless networks. However, the existing digital coding and transmission approaches can hardly accommodate users with diverse channel conditions, which is called the cliff effect. Recently, a novel video broadcasting method called SoftCast has been proposed. It achieves graceful degradation with increasing noise by making the magnitude of the transmitted signal proportional to the pixel value and using a novel power allocation scheme. This paper proposes a novel video broadcast method that exploits deep convolutional networks and group based sparse representation. They utilize the channel condition information generated from decoder to optimize the decoding process and reduce the various artifacts caused by source and channel coding. By utilizing soft video broadcast transmission, it achieves good broadcasting performance, avoids the cliff effect, and also can provide visually friendly subjective and objective reconstruction quality. The experimental results show that the proposed scheme provides better performance compared with the traditional SoftCast with up to 1.2 dB coding gain.

[Key words] wireless video broadcasting; Convolutional Neural Networks; group based sparse representation; soft video broadcast

0 引言

随着科技的不断发展, 人们开始更多地使用图片或者视频来交流和分享信息。在无线通信技术进步的推动下, 关于无线视频广播技术研究已然成为当前学界热点, 其研发成果也在陆续涌现。与此同时, 3G、4G 技术的逐渐应用和近年来智能手机与平板电脑的处理能力越来越强而且日趋普及, 人们也越来越青睐使用这些移动终端来观看视频, 因为这样做更方便、更智能、也更快捷。

传统数字视频广播标准^[1]中主要包含 2 部分。一部分是分层传输方法^[2-3], 另一部分则是可伸缩视频编码技术 (SVC)^[4-5]。其中, 可伸缩编码技术是指编码端将视频信号编码为一个基本层 (BL) 和

多个增强层 (EL)。分层调制 (HM)^[6]可用于将基本层和增强层的比特流叠加到一个需要传输的无线信号中, 如此一来也就实现了同一编码端对于信道质量不同的用户进行视频广播的目的。传统伪模拟传输应用 Softcast^[7-8]软广播技术的主要贡献是将所要传输的信号的线性变换直接在模拟信道上进行传输, 这些信号只需要执行能量分配而不需要进行量化、编码和调制。因此信道噪声也就直接转化为了重构噪声, 具有质量可伸缩性。为了缓解块效应, 学者们提出了很多去块效应的后处理方案, 大体上可分为 2 类^[9-10], 即: 基于图像增强的去块效应方法和基于图像恢复的去块效应方法。总地来说, 对于图像增强类的方法, 其基本思路是将去块效应视为一种图像增强过程, 通过在空域和频域进行滤波来平

作者简介: 尹文斌(1989-), 男, 博士研究生, 主要研究方向: 无线视频传输、图像编码; 范晓鹏(1978-), 男, 博士, 教授, 博士生导师, 主要研究方向: 图像/视频处理、模式识别。

收稿日期: 2019-04-02

哈尔滨工业大学主办 ◆ 学术研究与应用

滑可见的失真效应。对于图像复原类的方法,去块效应通常被表述为一个病态图像优化问题并利用一些图像先验知识和观测数据进行求解。全变差^[11]、基于块的稀疏表示^[12-14]以及马尔科夫随机场(MRF)均被作为图像先验模型用于寻找原始图像的MAP估计。在各类研究中,文献[15]将量化失真作为高斯噪声,使用FoE作为图像先验来建立图像去块效应最优化问题。深度神经网络在图像处理、视频分析、自然语义理解等方面取得了可观进展。深度神经网络是一种多层的神经网络,通过网络学习,从原始数据中提取不同层级的抽象信息。这种方法自然地体现了底层视觉特征到高层语义特征的演变。使用深度学习方法可以自适应地捕获到目标的多层次表示特征,相比于人工设计的特征,通常有着更好的应用性能。以图像去噪为例,文献[16]提出了一种基于卷积神经网络的图像去噪方法并证明卷积神经网络具有对马尔科夫随机场相近、甚至更高的表达能力。文献[17]成功地将多层感知器应用于图像去噪问题。文献[18]利用稀疏去噪自编码器来处理高斯噪声去除问题,并取得了与K-SVD相近的结果。文献[19]中,提出了一个可训练的非线性映射传递模型,而且可以通过一个前馈神经网络得以实现。

1 提出的基于卷积神经网络的无线视频广播方案

时下,无线视频广播面临3个主要问题就是可伸缩性、鲁棒性和重建视频质量。传统的SoftCast无线视频广播系统虽然能够避免悬崖效应的发生,但是由于其采用基于块的编码方式且信道噪声直接叠加在传输信号上,所以其重构视频中不可避免地含有编码失真和传输失真,极大地降低了重构视频的主客观质量,因此如何去除编码与传输失真是亟待解决的研究课题。针对图像复原问题,稀疏表示理论利用先验知识和重构图像进行稀疏表示优化求解;深度网络模型可以通过强大的特征提取能力对自然图像的深层次特征进行学习,提取降质图像中的有效信息。根据稀疏表示理论,本文拟利用视频帧的局部稀疏性和非局部自相似性,通过基于组的稀疏表示模型减轻信重构视频中的编码失真。基于深度神经网络特性,利用卷积神经网络对视频帧进行紧致而高效的表达,从而区分重构帧中的信道噪声与有效视频信息,以便于提升重构帧质量。实验结果表明,本文提出的方案不仅具有良好的视频广

播可伸缩性,还能提供视觉友好的主客观重构质量。

1.1 编码框架

基于卷积神经网络的无线视频广播框架如图1所示。本文提出的基于深度神经网络的无线视频传输方案的编码框架主要包含:视频压缩、视频软传输、基于图像组的稀疏表示以及深度卷积网络。

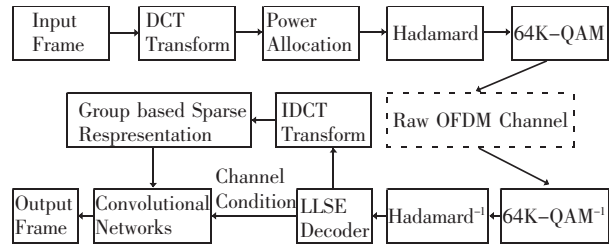


图1 基于卷积神经网络的无线视频广播框架图

Fig.1 Framework of CNN based video broadcasting method

在编码端,通过BDCT(Block based DCT)去除视频帧中的冗余,对视频进行压缩处理。利用对视频的频域系数的伸缩来最小化信号在信道传输中的总体失真。编码后的信号直接经过稠密的64K-QAM星座图发送给具有不同信道质量的多个用户。传统的数字视频传输方案中,悬崖效应严重影响了各用户的解码体验。在本文提出的方案中,通过软传输的方式直接将编码后的信号发送给客户端,为用户提供具有良好的视频质量可伸缩性。在解码端,其利用LLSE对接收到的信号进行解码。基于图像组的稀疏表示模型能够同时利用视频帧的局部稀疏性和非局部自相似性,本文利用基于图像组的稀疏表示模型降低由BDCT编码所造成的块效应。由于深度神经网络可以通过学习的方式提取不同程度的信道噪声特征,在获得具有平滑属性的解码视频帧后,本文利用卷积神经网络优化由软传输中信道噪声引起的失真。

1.2 基于块的变换

由于视频帧内通常具有较强的空域相关性,本文的方法利用这一性质,通过BDCT变换的方式对视频帧内信息进行紧致表达。传统的视频编码方法需要已知信道条件,根据信道条件选择码率并对频域系数进行量化。这类的量化方案会迫使所有的用户观看质量一致的解码视频。

本文提出的方案将视频帧划分为图像块,再利用BDCT变换将视频帧由空域转换到频域。通常来说,DCT系数具有能量集中的特性,也就是具有较高重要性的低频系数集中的变换系数矩阵的左上角,而具有较低重要性的高频系数集中在系数矩阵的右下角,数值通常是接近或等于零。由于高频

DCT 系数对于视频帧内信息的影响较轻微,所以采用丢弃 DCT 系数中的零值的方式对视频数据进行压缩。当带宽受限时,本文的方案会根据 DCT 系数的分布特性,根据带宽要求进一步丢弃当前最不重要的 DCT 系数。然而这样的方式需要面临的问题是编码端不得不发送大量的元数据来标识丢弃的 DCT 系数所在的位置。

为了减少传输被丢弃频域系数位置所需的元数据,本文的方法将不同块的频域系数按照频率划分为 band,以 band 为单元对视频进行压缩。具体来说,可将不同块的同一位置系数放入一个 band。而后根据压缩率及带宽需求,以 band 为单位判别是否丢弃其中的 DCT 系数。不同图像块变换系数的高频信息通常处于接近或者一致的区域,所以对 band 进行丢弃操作与对独立的 DCT 系数进行丢弃操作具有接近的压缩性能,却大幅减少了元数据规模。

1.3 能量分配与传输

能量分配是对不同频率的 DCT 系数的保护方法。假定总体传输能量为 P , DCT 系数为 R ,其中频率 i 的 DCT 系数为 R_i , g 表示对应频率的伸缩系数。根据最优线性编码理论, g 可以由以下公式求得:

$$g_{R_i} = \left(\frac{P}{\sqrt{\lambda_i} \sum_k \sqrt{\lambda_k}} \right)^{1/2}, \quad (1)$$

其中, λ_i 表示频率 i 的 DCT 系数的方差, K 表示总体频率数量。定义对角矩阵 $\mathbf{G} = \text{diag}\{g_{R_1}, g_{R_2}, \dots, g_{R_K}\}$, 经过能量分配的信号 M 可以表示为 $M = \mathbf{G} \cdot R$ 。各个频率系数的能量 λ_i 需要作为元数据发送给解码端。

经过能量分配,为了增强系统的抗丢包能力,研究利用哈德玛变换将数据转化为具有相同重要性的数据包,其数学公式可表示为:

$$\mathbf{U} = \mathbf{H} \cdot \mathbf{M} = \mathbf{H}\mathbf{G} \cdot \mathbf{R} = \mathbf{C} \cdot \mathbf{R}. \quad (2)$$

其中, \mathbf{H} 表示哈德玛变换矩阵, $\mathbf{C} = \mathbf{H}\mathbf{G}$ 表示编码矩阵。

在物理层中,元数据与 DCT 系数经过不同的方式发送给解码端。由于元数据需要无损地传输至解码端,这里利用传统的数字编码方式发送元数据。编码端对元数据进行 8-bit 量化,并通过变长编码 (VLC) 对量化结果进行压缩。压缩后的码流通过含有前向纠错编码和调制的 802.11 物理层进行传输。为了妥善保护元数据,研究使用 1/2 卷积码作为保护编码,采用 BPSK 调制类型。

与元数据不同, DCT 系数信号由实数值组成。

在物理层,这些实数值通过 64K-QAM 星座图直接映射为复数符号。这种星座图是一种经典的 N -QAM 星座图,其中,每个输入的实数值都通过统一的量化器量化成为 8-bit 整数。量化器的动态范围由输入值的上下限决定。在量化后,每两个整数组成一个复数作为 64K-QAM 的输出。IFFT 对每个数据包中的符号,得出一系列复数时域样本。这些样本通过标准方式正交混合到频带中。实数和虚数部分首先通过 D/A 转换器变换到模拟信号。这些模拟信号组成最终的传输信号。

1.4 线性最小平方估计

定义 N 为信道噪声,解码端接收到的信号可以表示为:

$$\hat{\mathbf{U}} = \mathbf{U} + \mathbf{N} = \mathbf{H}\mathbf{G} \cdot \mathbf{R} + \mathbf{M}, \quad (3)$$

接收的信号可以通过线性最小平方估计 (LLSE) 进行恢复,解码出的频域系数可以表示为:

$$\hat{\mathbf{R}} = \sum_R \mathbf{C}^T (\mathbf{C} \sum_R \mathbf{C}^T + \sum_N)^{-1} \hat{\mathbf{U}}, \quad (4)$$

其中, \sum_R 和 \sum_N 表示信号 R 和 N 的协方差矩阵。

当 CSNR 较高时,信道噪声 N 中的元素接近或等于零,根据 LLSE 的性质,则:

$$\mathbf{R}_{LLSE} \approx \mathbf{C}^{-1} \mathbf{U}. \quad (5)$$

可以看出在 CSNR 较高时,LLSE 通过编码端的逆变换就可以获得解码数据。主要原因在于 CSNR 较高时,研究可以充分信任接收的信号而不需要其分布和统计特性,例如 DCT 系数的方差。相反地,当 CSNR 较低时噪声能量较大,研究无法完全信任接收到的信号。因此,根据各个 band 中 DCT 系数的统计特性对估计进行调整是更为有效的解码方式。

传统 802.11 系统对于丢包非常敏感,丢包会导致失真增大、甚至解码失败。但是当本文的方法面临丢包问题时,由于经过哈德玛变换,所有数据包具有同等重要性,不会发生 MPEG 系统中出现的某些特殊数据包丢失导致解码失败的情况。当出现丢包时,相当于信号 U 中的某行缺失,将矩阵 \mathbf{C} 和 \mathbf{N} 以及协方差矩阵 \sum_R 和 \sum_N 对应行或列去除,就可以在丢包情况下完成解码。

1.5 稀疏表示去噪

受到 BDCT 与能量分配的影响,本文解码的视频中通常含有严重的块效应。前文中提到,稀疏表示在去除块效应问题中具有出色的性能,为基于 BDCT 编码的视频提供良好的视觉效果。在本文的

方法中,研究将基于组的稀疏表示(GSR)去块效应算法转化为最大后验概率(MAP)问题。

假设输入帧为 x , 其解码端重构帧为 \hat{x} , 那么优化后的重帧可以通过如下方式获得:

$$y = \arg \max \log(p(\hat{x} | x)) + \log(p(x)), \quad (6)$$

其中,第一项表示数据保真约束,第二项表示视频帧先验知识。受到基于图像组的稀疏表示在图像处理问题中取得成果的启发,公式(6)中的MAP最优化问题可以表示为:

$$y = \arg \min \log(p(\hat{x} | x)) + \log(p_{GSR}(x)) + \log(p_{CC}(x)), \quad (7)$$

其中, $p_{GSR}(x)$ 和 $p_{CC}(x)$ 表示 GSR 先验概率和压缩约束 (compression constrict) 先验概率。

解码视频包含了传输噪声和压缩噪声。GSR 主要针对引起块效应的压缩噪声。利用高斯压缩噪声模型和压缩噪声方差 σ_{com}^2 , 数据保真约束项可以表示为:

$$\log(p(\hat{x} | x)) = -\frac{1}{2\sigma_{com}^2} \|x - \hat{x}\|_2^2, \quad (8)$$

基于图像组的稀疏表示模型证明,一组图像块可以通过稀疏系数与字典元子的线性组合进行表示。对于每个图像组的稀疏编码是为了寻找一个稀疏矢量 $x_{G_k} \approx D_{G_k} \alpha_{G_k}$, 其中 x_{G_k} 表示图像组集合、 D_{G_k} 表示字典、 α_{G_k} 表示图像组的稀疏系数。那么视频帧就可以稀疏地表示为以图像组为单元的稀疏系数 $\{\alpha_{G_k}\}$ 的集合。故而公式(7)中的第二项就可以表示为:

$$\log(p_{GSR}(x)) = -\eta \|\alpha_C\|_0, \quad (9)$$

其中, α_C 表示所有稀疏系数 α_{G_k} 的级联,并且要求编码 α_C 具有稀疏性。

为了描述 CC 先验概率,文中定义阈值 Ω 为:

$$\psi(x) = \begin{cases} 0, & \text{if } x \in \Omega \\ +\infty, & \text{if } x \notin \Omega \end{cases}, \quad (10)$$

其中, Ω 表示伸缩后 DCT 系数的范围。

公式(7)中的第三项就可以表示为:

$$\log(p_{CC}(x)) = -\psi(x), \quad (11)$$

利用上述先验知识,去块效应最优化问题就可以写作如下数学形式:

$$(\hat{\alpha}_C, \hat{D}_C) = \arg \min_{\alpha_C, D_C} \frac{1}{2\sigma_{noise}^2} \|D_C \circ \alpha_C - \hat{x}\|_2^2 + \theta \|\alpha_C\|_0 + \psi(D_C \circ \alpha_C), \quad (12)$$

在获得更新的字典和稀疏系数后,更新的视频帧可以通过 $y = \hat{D}_C \circ \hat{\alpha}_C$ 重建。在文中,研究利用

split Bergman 算法对公式(12)进行求解。定义 $f(x) = \frac{1}{2\sigma_{noise}^2} \|x - \hat{x}\|_2^2$, $g(\alpha_C) = \theta \|\alpha_C\|_0 + \psi(D_C \circ \alpha_C)$, 那么公式(12)的最优化问题可以等价地转为三步迭代,即:

$$y^{(t+1)} = \arg \min_y \frac{1}{2\sigma_{noise}^2} \|x - \hat{x}\|_2^2 + \frac{1}{2\beta^{(t)}} \|x - D_C^{(t)} \circ \alpha_C^{(t)} - b^t\|_2^2, \quad (13)$$

$$(\alpha_C^{(t+1)}, D_C^{(t+1)}) = \arg \min_{\alpha_C, D_C} \frac{1}{2\beta^{(t+1)}} \|x^{(t+1)} - D_C \circ \alpha_C - b^{(t+1)}\|_2^2 + \theta \|\alpha_C\|_0 + \psi(D_C \circ \alpha_C), \quad (14)$$

$$b^{(t+1)} = b^{(t)} - (x^{(t+1)} - D_C^{(t+1)} \circ \alpha_C^{(t+1)}), \quad (15)$$

以上分离的子问题可以获得有效解。进而研究还将推得如下数学公式:

$$\beta^{(t)} = \rho \cdot (\sigma_s^{(t)})^2, \quad (16)$$

$$\sigma_s^{(t)} = \delta \sqrt{\sigma_{noise}^2 - \|D_C^{(t)} \circ \alpha_C^{(t)} + b^{(t)} - \hat{x}\|_2^2}. \quad (17)$$

其中, ρ 表示伸缩系数, δ 表示控制方差估计的伸缩系数。

1.6 深度神经网络去噪

传统的软传输方案直接将信号通过 raw OFDM 信道进行发送,可以取得良好的可伸缩性能,为不同用户提供与其信道质量相一致的重构视频。但是由于信道噪声直接叠加在接收的噪声中,会导致重构视频中存在噪声模糊效应。借助于深度神经网络从原始数据中提取不同层级的抽象信息的能力。研究利用卷积神经网络提取由不同信道噪声所形成的特征,对解码端重构的视频进行复原。卷积神经网络的结构如图2所示。

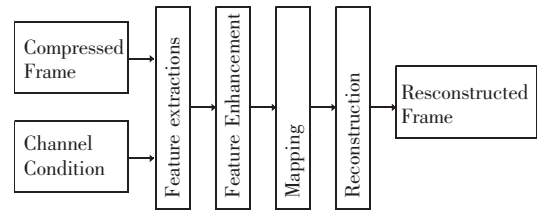


图2 卷积神经网络结构

Fig. 2 Convolutional Neural Network structure

复原网络中含有4层,分别是:特征提取层、特征增强层、映射层和重建层,每层都具有独立的目标。其中,特征提取层从重建帧中提取重叠的块,并利用高维向量表示每个块。特征增强层从特征提取层所获得的 n_1 个特征图中进一步提取特征并形成一个新的特征图集合。在特征增强层后,非线性映

射层通过高维矢量对高质量图像块进行表示。重建层利用映射后的高质量图像块对输入帧进行复原。卷积网络可以描述为:

$$F_i(y) = \max(0, W_i * y + B_i), i \in \{1, 2, 3\}, \quad (18)$$

$$\hat{y} = W_4 * F_3(y) + B_4,$$

其中, W_i 和 B_i 表示第 i 个频率下的滤波器以及偏置; F_i 表示输出的特征图; $*$ 表示卷积操作。 W_i 含有 n_i 个滤波器, 滤波器的尺寸为 $n_{i-1} \times f_i \times f_i$, n_0 表示输入帧的通道数。使用线性整流函数 (Rectified Linear Unit, RELU), $\max(0, x)$ 作为激活函数。

训练过程中, 研究定义了原始视频帧集合为 $\{x_i\}$, 解码端含有失真的视频帧集合为 $\{y_i\}$, 每个原始帧 x_i 都具有相对应的 y_i 。利用 MSE 作为损失函数, 网络训练的模型可以表示为:

$$L(\Omega) = \frac{1}{n} \sum_{i=1}^n \| F(y_i^{(c)}; \Omega) - x_i^c \|^2. \quad (19)$$

其中, $\Omega = \{W_1, W_2, W_3, W_4, B_1, B_2, B_3, B_4\}$, n 是训练样本的数量。通过随机梯度下降法和标准反向传播方式最小化网络的失真。

2 实验结果及分析

本节将通过实验来验证所提出方法的视频广播性能和压缩效率。实验中, 将相同的编码后信号发送给信道条件不同的多个用户端。在用户端比较最终解码出的视频质量。也就是将传统数字视频广播方案 (H. 264+802. 11^[14])、SoftCast 以及本文提出的方法的性能指标在同一条件下进行比较和对照。

2.1 实验设置

在实验中, 研究测试了本文提出的方法在视频单播与广播场景下的性能。对比本文方法、传统数字视频广播方案 (H. 264 信源编码与含有前向纠错编码与 QAM 调制的 802. 11 物理层) 以及 SoftCast 的广播性能。测试项目是向具有不同 CSNR 的多个用户广播同一视频序列。

研究使用 400 张尺寸为 180×180 的视频帧进行训练。视频帧被划分为 64×64 的图像块并通过 SoftCast 进行压缩和传输, 最终生成含有不同信道噪声强度的训练数据。在训练帧中按照补偿为 20 进行采样, 所有训练图像块的数量为 204 800。重建层与其它各层的网络学习率分别设置为 10^{-5} 和 10^{-4} 。卷积网络的参数设置如下: $f_1 = 9, f_1' = 7, f_2 = 1, f_3 = 5, n_1 = 64, n_1' = 32, n_2 = 16, n_3 = 1$ 。

测试序列包含 'foreman_cif.yuv'、'news_cif.yuv'、'mother_cif.yuv' 以及 'bus_cif.yuv'。视频帧率为 30 Hz。编码后的信号经过噪声类型为 AWGN 的 raw OFDM 信道进行传输。

2.2 噪声模型和衡量指标

信道噪声将会叠加在所有的采样值上, 噪声的强度由信道信噪比 (Channel Signal to Noise Ratio, CSNR) 进行衡量。其数学定义公式为:

$$CSNR = \|y\|_2^2 / \|e\|_2^2, \quad (20)$$

根据丢包率 p 可以计算出最终解码端得到的采样值数量为 $M = \lfloor (1 - P)N \rfloor$ 。

视频客观质量衡量指标采用峰值信噪比 (Peak Signal to Noise Ration, PSNR) 进行衡量。在运算时会用到如下数学公式:

$$PSNR = 10 \log_{10} \frac{\max_i^2}{MSE}, \quad (21)$$

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \|x(i, j) - \hat{x}(i, j)\|^2. \quad (22)$$

2.3 实验结果与分析

2.3.1 广播性能

研究对比了提出的方案、SoftCast 和基于 H. 264 的传统数字视频广播方案。对于传统方案, 文中采用 4 种推荐的 802. 11 信道编码与调制组合。根据 H. 264 编码器的带宽计算相应的比特率。各种方案在不同 CSNR 数值下的重建视频的 PSNR 则如图 3 所示。

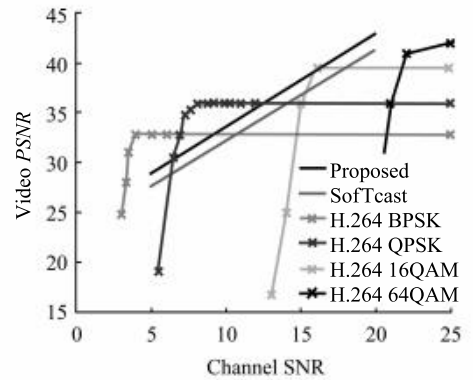


图3 广播性能对比

Fig. 3 Broadcasting performance comparison

由图 3 中可以看出, 所有基于 H. 264 的传统视频广播方案, 无论信道编码率如何, 都会遭遇严重的悬崖效应。举例来说, H. 264+BPSK 方案在 CSNR 为 3~5 dB 时表现良好, 但是当 CSNR 低于 3 dB 时, 会导致信道保护编码失效, 视频无法解码; 而当 CSNR 高于 10 dB 时, 由于信源编码率的限制, 导致

CSNR 上升时用户的视频质量无法进一步提升。

相反地,SoftCast 与本文的方案都较好地避免了悬崖效应,为用户提供了具有良好可伸缩性的平滑性能曲线。随着 CSNR 的提高,用户解码的视频质量也有相应的提升。但是由于本文利用 GSR 以及卷积神经网络对重构帧中的失真进行了优化,本文方案的重构质量在全部 CSNR 范围内都优于 SoftCast。

2.3.2 视觉质量

本文方法的主观质量对比如图 4 所示。在相同 CSNR 条件下,可以清晰看到本文提出的方案具有更好的重建质量。SoftCast 的重构帧中有较为明显的块效应和噪声干扰,而本文方案的重构帧中几乎没有块效应。实验证明,基于 GSR 的方法表现出优秀的去块效应性能。在相似 PSNR 条件下,本文方案的重构帧具有平滑的纹理以及清晰的边界,在主观视觉感受上更易于被接受。

本文方案与 SoftCast 在不同序列上的测试结果如图 5 所示。由于 GSR 需要在视频帧内寻找相似

块,使得当视频内容运动剧烈时,会在一定程度上影响去块效应结果,例如 bus. cif 中的测试结果。但是从图 5 中可以看出,本文的方案具有良好的鲁棒性,对于不同的视频序列都取得了高于 SoftCast 的重构质量。

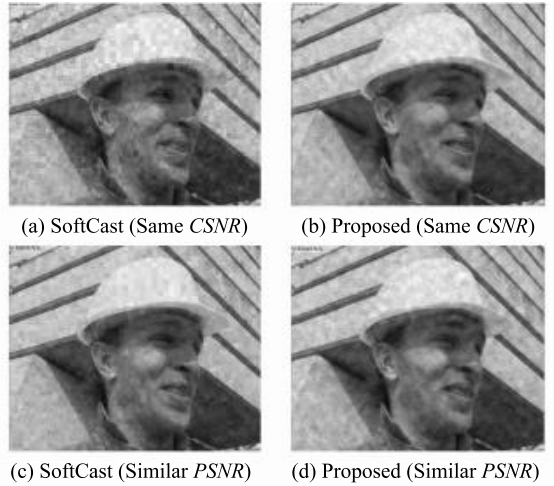


图 4 主观质量对比

Fig. 4 Visual quality comparison

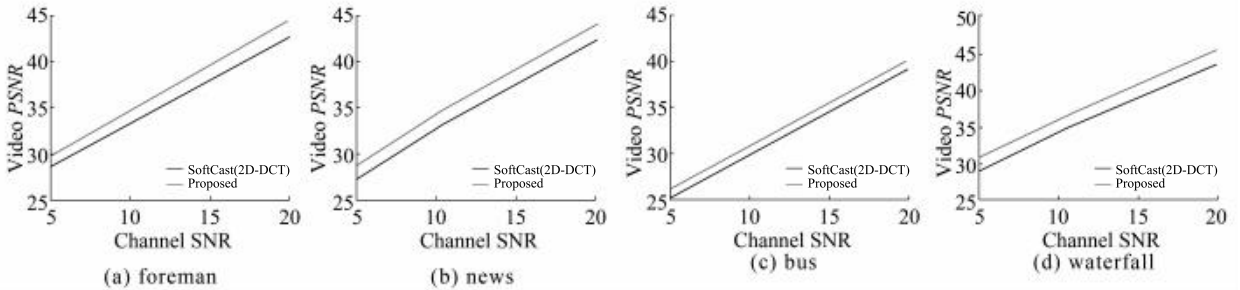


图 5 不同序列的广播性能

Fig. 5 Broadcast performance on different sequences

2.3.3 多播性能

文中使用 3 种方案服务一组客户(3 个具有不同信道质量的客户),每个客户的 CSNR 分别为 5 dB、10 dB 和 20 dB。传统的数字视频广播方案采用 H.264+BPSK 的组合。多播性能的效果对比即如图 6 所示。由于 3 个客户中信道最差的 CSNR 只有 5 dB,所以传统方案将必须采用 BPSK 进行调制,否则会导致 CSNR 为 5 dB 的客户无法做到正确解码。在 SoftCast 与本文方案的对比中,发送端可以同时适应多种信道条件。而在本文方案的测试结果中,虽然客户 1 的重构质量略低于传统方案,但是其余客户均获得了高于传统方案的视频质量。从图 6 中可以看出,本文较传统方案具有更好的可伸缩性,较 SoftCast 有着更好重构质量。

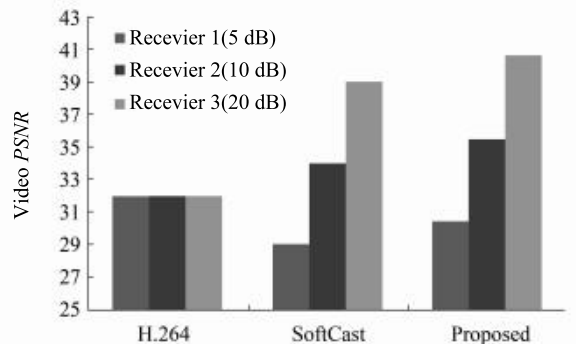


图 6 多播性能对比

Fig. 6 Multicast performance comparison

3 结束语

在本文中,提出了基于深度神经网络的无线视 (下转第 12 页)