摘要

图像视频压缩能有效减少其传输与存储代价，从而成为大多数视觉应用中必不可少的技术环节。然而图像视频经过有损压缩后会产生失真，不仅影响人的视觉观看体验，还会降低机器视觉分析的准确性。因此，如何在信号与特征层面上对压缩图像视频进行增强，从而降低压缩失真带来的影响，同时满足人类视觉与机器视觉的需求，是一个亟待探索的科学问题。压缩图像视频的增强技术无需改变现有采集设备上的压缩模块设置，可用于大量不同应用场景，具有很强的应用价值。本文围绕“压缩图像视频的信号与特征增强”这一目标，在信号增强、特征增强、特征信号联合增强三个层面开展研究，并构建了基于压缩视频增强的端边云协同处理系统。本文的主要创新工作包括：

第一，为降低压缩失真对人类视觉的影响，提出一种基于解码残差约束的压缩视频信号增强方法。本方法通过分析压缩辅助信息与信号增强的内在关联，将压缩辅助信息显式地应用在对原始视频信号的预测中，使用解码残差约束信号增强过程。在此基础上，本方法设计了一种双通道激活函数来对特征进行双重激活，从而进一步地减少信号增强网络中的信息丢失。在 HEVC 国际标准通测序列上的实验结果表明，本方法在发表时比 4 种相关国际先进方法的率失真性能提升 3.3% 以上。

第二，为降低压缩失真对机器视觉的影响，提出一种基于量化失真建模的压缩图像视频特征增强方法。本方法首先分析了原始与压缩数据在特征学习中的差异，进而使用量化步长来建模量化失真对特征学习的影响。在此基础上，提出基于量化步长的样本置信度来为不同压缩等级的图像视频分配训练权重，从而减轻量化失真对参数更新的影响，同时设计基于元学习的量化失真特征标准化方法来修正量化失真造成的特征偏移。在常用国际基准数据集 ImageNet 上的实验结果表明，本方法在发表时比 8 种相关国际先进方法的图像分类准确率提升 5.5% 以上。

第三，为同时降低压缩失真对人类视觉与机器视觉的影响，提出一种基于压缩解耦学习的特征信号联合增强方法。本方法首先分析了特征与图像视频压缩的关系，进而提出压缩不敏感与压缩敏感特征的解耦学习框架。在此基础上，利用解耦特征生成变换参数，对信号增强网络的中间特征进行变换，从而根据解耦特征中的信息对联合增强网络学习到的映射进行自适应调整；同时提出非局部特征融合模块将压缩不敏感特征融合至联合增强网络中，进一步提升了特征保真度。在常用国际基准数据集 MSCOCO 上的实验结果表明，本方法在发表时比 8 种相关国际先进方法的平均峰值信噪比提升 0.21 dB 以上，目标检测均值平均正确率提升 2.5% 以上，语义分割平均交并比提升 2.6%

以上，即同时增强了面向人类视觉的感知质量与面向机器视觉的分析识别性能。

第四，构建了基于压缩视频增强的端边云协同处理系统，以验证所提出方法的有效性。针对压缩造成的视频特征质量低的问题，本文设计并实现了基于特征增强的边子系统，使用基于量化失真建模的特征增强方法提升系统检索性能；针对压缩视频信号质量低的问题，本文设计并实现了基于信号增强的云子系统，使用基于解码残差约束的信号增强方法和基于压缩解耦学习的特征信号联合增强方法提升视频质量。实验结果表明，本文提出的特征增强方法可以将系统行人与车辆检索性能提升 5.0% 以上，信号增强方法可以将按需调取压缩视频的峰值信噪比提升 1.00 dB 以上。

综上所述，本文从信号增强、特征增强、特征信号联合增强三个层面对压缩图像视频进行增强，并实现了一种基于压缩视频增强的端边云协同处理系统。本文的研究可以为压缩图像视频增强的后续研究工作奠定基础。

关键词：图像视频压缩，信号增强，特征增强，特征信号联合增强，端边云协同视频处理

The Signal and Feature Enhancement of Compressed Images and Videos

Li Ma (Computer Application Technology)
Directed by Prof. Yonghong Tian

ABSTRACT

Lossy compression is an indispensable technology in visual applications, consequently reducing the transmission and storage cost of images and videos effectively. However, due to the inevitable information loss in the compression operations, compression distortion is accordingly brought into images and videos, which reduces the visual quality and the performance of downstream parsing tasks. Therefore, how to enhance the compressed images and videos at the level of signals and features, thereby reducing the impact of compression distortion and meeting the needs of both human vision and machine vision, is one of the most important scientific problems that need to be explored urgently. Moreover, compressed image and video enhancement could be used in visual applications without changing the configuration of the compression module in existing cameras, hereby having an important application value. Therefore, with the goal of "enhancing the signal and feature of compressed images" in mind, this thesis explores the problem from three perspectives: signal enhancement, feature enhancement, and feature-signal joint enhancement. Moreover, a compressed video enhancement-based end-edge-cloud collaborative processing system is developed. The main contributions of this thesis are summarized as follows:

Firstly, to reduce the impact of compression distortion on human vision, this thesis proposes a signal enhancement method with the decoded residual constraint for compressed videos. The proposed method first analyzes the connection between side information and signal enhancement. On this basis, side information is explicitly applied to predict the original signal. Specifically, the proposed method utilizes decoded residual frames to constrain the signal enhancement process. Moreover, a dual-channel activation function is designed to reduce information loss in the signal enhancement network. The experimental results on HEVC test sequences show that the proposed method improves the BD-rate by more than 3.3% compared with 4 state-of-the-art methods at the time of publication.

Secondly, to reduce the impact of compression distortion on machine vision, this thesis proposes a feature enhancement method via quantization distortion modeling for compressed images and videos. The proposed method first analyzes the difference between original and compressed data in feature learning. On this basis, the quantization steps are employed to model the influence of quantization distortion on feature learning. Specifically, the proposed method designs a quantization steps-based confidence as sample weights to reduce quantization distortion's influence on the parameters' updating. Moreover, a meta-learning-based quantization distortion-related batch normalization is designed to alleviate the variance of feature distribution. The experimental results on ImageNet show that the proposed method improves the accuracy of image classification by more than 5.5% compared with 8 state-of-the-art methods at the time of publication.

Thirdly, to reduce the impact of compression distortion on human vision and machine vision analysis simultaneously, this thesis proposes a feature-signal joint enhancement method via sensitivity decouple learning. The proposed method first analyzes the relationship between features and lossy compression. On this basis, a sensitivity decouple learning is proposed to extract compression insensitive and sensitive features from compressed images and videos. Then, the proposed method transforms the intermediate features of the signal enhancement network to adjust the mapping learned by the joint enhancement network according to the parameters generated from the decoupled features. Meanwhile, a non-local feature fusion module is proposed to fuse the compressed insensitive features into the joint enhancement network to improve the feature fidelity further. The experimental results on MSCOCO show that the proposed method could improve the average PSNR by more than 0.21 dB, the average mAP of object detection by more than 2.5%, and the average mIoU of semantic segmentation by more than 2.6% compared with 8 state-of-the-art methods at the time of publication.

Fourthly, an end-edge-cloud collaborative processing system based on compressed video enhancement is developed to verify the effectiveness of the proposed methods. This system consists of the end, edge, and cloud subsystems. The edge subsystem utilizes the feature enhancement method via quantization distortion modeling to address the features' low quality caused by compression, thereby improving the retrieval performance of the system. Meanwhile, the cloud subsystem employs the signal enhancement method and the feature-signal joint enhancement method to address the signals' low quality caused by compression, thereby giving users a better visual experience. Experimental results show that the proposed methods could improve the system's pedestrian and vehicle retrieval performance by more than 5.0%

and the PSNR of compressed videos by more than 1.00 dB.

In conclusion, this thesis enhances compressed images and videos from three perspectives: signal enhancement, feature enhancement, and feature-signal joint enhancement, and implements a compressed video enhancement-based end-edge-cloud collaborative processing system. The research in this thesis could build a foundation for subsequent research work on the enhancement of compressed images and videos.

KEY WORDS: Image and video compression, Signal enhancement, Feature enhancement, Feature-signal joint enhancement, End-edge-cloud collaborative video processing