法政大学学術機関リポジトリ
HOSEI UNIVERSITY REPOSITORY

PDF issue: 2025-07-12

A Research on Enhancing Reconstructed Frames in Video Codecs

PHAM DO, Kim Chi

(開始ページ / Start Page)
1
(終了ページ / End Page)
122
(発行年 / Year)
2022-09-15
(学位授与番号 / Degree Number)
32675甲第554号
(学位授与年月日 / Date of Granted)
2022-09-15
(学位名 / Degree Name)
博士（工学）
(学位授与機関 / Degree Grantor)
法政大学 (Hosei University)
(URL)
https://doi.org/10.15002/00025871

博士学位論文
論文内容の要旨および審査結果の要旨

論文題目	A Research on Enhancing Reconstructed Frames in Video Codecs
氏名	PHAM DO　Kim　Chi
学位の種類	博士（工学）
学位番号	第 554 号
学位授与年月日	2022 年 9 月 15 日
学位授与の要件	法政大学学位規則第 5 条第 1 項第 1 号該当者（甲）
論文審査委員	主　査　周　金佳　　准教授
	副　査　尾川　浩一　　教授
	副　査　彌冨　仁　　教授

１．論文内容の要旨

To relieve the burden on video storage, streaming and other video services, researchers from the video community have developed a series of video coding standards for compressing the videos with higher quality and lower bitrates. While traditional coding algorithms have been continuously enhanced, deep learning based video coding have been receiving significant attention in recent years. This thesis proposes several deep learning methods to improve the performance of existing video codecs. This thesis includes the following chapters.

Chapter 1 [Introduction] describes a comprehensive review of video coding, then an analysis of the drawbacks of state-of-the-art methods.

Chapter 2 [Enhancing reference-frame interpolation for video encoder] presents the deep learning-based fractional interpolation method for generating the half-pixel and quarter-pixel of the reference samples. Motion-compensated prediction is one of the essential methods for reducing temporal redundancy in inter coding. The target of motion-compensated prediction is to predict the current frame from the list of reference frames. Recent video coding standards commonly use interpolation filters to obtain sub-pixel for the best matching block located in the fractional position of the reference frame. However, the fixed filters are not flexible to adapt to the variety of natural video contents. This chapter describes a novel CNN-based fractional interpolation in motion-compensated prediction to improve the coding efficiency. The proposed interpolation filters are designed to replace the hand-crafted interpolation filters of video coding standards, extending the ability to deal with the diversity of video content. Only

one model is trained for all the fractional positions, enabling the flexibility to deal with other video coding standards with the least modifications. Moreover, two syntax elements indicate interpolation methods for the Luminance and Chrominance components, which have been added to bin-string and compressed by an entropy encoder. As a result, the proposal gains 2.9%, 0.3%, 0.6% Y, U, V BD-rate reduction, respectively, under low delay P configuration.

Chapter 3 [Compressive sensing image enhancement at video decoder] describes a deep learning-based compressive sensing (CS) enhancement technology using multiple reconstructed images for enhancing decoded images. Different from the other image compression standards, CS can get various reconstructed images by applying different reconstruction algorithms to coded data. Using this property, it is the first time to propose a deep learning based compressive sensing image enhancement framework using multiple reconstructed signals. Firstly, images are reconstructed by different CS reconstruction algorithms. Secondly, reconstructed images are assessed and sorted by a no-reference quality assessment module before input to the quality enhancement module by order of quality scores. Finally, a multiple-input recurrent dense residual network is designed for exploiting and enriching the useful information from the reconstructed images. Experimental results show that the proposal obtains 1.88–8.07dB PSNR improvement while the state-of-the-art works achieve a 1.69–6.69 dB PSNR improvement under sampling rates from 0.125 to 0.75.

Chapter 4 [In-loop filtering image enhancement for video encoder-decoder] presents a deep learning-based in-loop filtering framework for the latest video coding standard, Versatile Video Coding (VVC). The existing deep learning-based VVC in-loop filtering enhancement works mainly focus on learning the one-to-one mapping between the reconstructed and the original video frame, ignoring the potential resources at encoder and decoder. This work proposes a deep learning-based Spatial-Temporal In-Loop filtering (STILF) that takes advantage of the coding information to improve VVC in-loop filtering. Three filtering modes are applied, including VVC default in-loop filtering, self-enhancement convolutional neural network with coding unit map, and the reference-based enhancement CNN with the optical flow. To further improve the coding efficiency, this work proposes a reinforcement learning-based autonomous mode

selection (AMS) approach. The agent is trained to predict the trend of splitting and filtering mode in each coding unit. By predicting filtering mode and allowing the coding unit to be split more, STILF-AMS requires zero extra bit while ensuring the quality of reconstructed images. As a result, this work outperforms the latest VVC standard and the state-of-the-art deep learning-based in-loop filtering algorithms. Remarkably, up to 18% and an average of 5.9% bitrate savings have been.

Chapter 5 [Conclusion] concludes the contributions of this dissertation.

As mentioned above, this dissertation proposes several deep learning based techniques for enhancing reconstructed frames in video encoders and decoders. As a result, the compression ratio and video quality are greatly improved.

２．審査結果の要旨

With the growing demand for video content and video resolution, efficient video coding techniques are required to compress the vast video data while keeping the good video quality. This dissertation develops hybrid video codecs (encoder and decoder), taking advantage of both traditional video coding technologies and deep learning techniques to increase the video compression ratio and improve the quality of the decompressed video. The novelty and effectiveness of this dissertation were confirmed in the following points.

1. In video encoder, a deep learning based interpolation algorithm is proposed to replace the hand-crafted interpolation filters of mainstream video coding standards. Recent video coding standards commonly use hand-crafted interpolation filters to obtain sub-pixel for the best matching block located in the fractional position of the reference frame. However, the fixed filters are not flexible to adapt to the variety of natural video contents. This dissertation proposes a novel learning based fractional interpolation in motion-compensated prediction to improve the coding efficiency by handling the diversity of video content.

2. In video decoder, a deep learning based frame enhancement technology is proposed to increase the quality of the decoded images and videos. Different from the other compression standards, compressive sensing based compression can get various reconstructed frames by applying different reconstruction algorithms to coded data. Using this property, it is the first time to propose a deep learning based frame

enhancement framework using multiple reconstructed signals.

3. In both video encoder and decoder, a deep learning based in-loop filtering framework is proposed to enhance the reference frames. This dissertation proposes three filtering modes and a reinforcement learning-based autonomous mode selection approach. It is applied in both video encoder and decoder to improve the compression efficiency and improve the visual quality of the reconstructed videos significantly.

Based on all of these, this examination committee is unanimous that the submitted doctoral thesis is fully qualified as a Doctor of Philosophy (Engineering).

（報告様式Ⅲ）