超解像とMPEG NNR 深層学習メタデータを利用した符号化映像伝送

Coded Video Transmission using Super-Resolution and MPEG-NNR Deep Learning Metadata

鈴木 拓矢† 佐々木 瑛一† 中條 健† 猪飼 知宏† 渡辺 裕‡

Takuya Suzuki† Eiichi Sasaki† Takeshi Chujoh† Tomohiro Ikai† Hiroshi Watanabe‡

シャープ株式会社† 早稲田大学‡

Sharp Corporation† Waseda University‡

Abstract: 地上デジタル TV 方式の高度化に向けて、4K/8K 映像放送の実現を目指すために、超解像を用いた方式を検討している。超解像を用いた伝送方式は、低レートで高精細な映像を得ることが可能である。一方で、学習していない未知の映像を超解像した場合、映像が劣化してしまう可能性があった。そこで、エンコード側では実際に伝送する映像で学習を行い、その超解像ネットワークパラメタを MPEG NNR 方式で圧縮して符号化映像と共に伝送し、デコード側では受信したパラメタを用いて符号化映像を超解像することで映像品質を向上させる手法を提案する。従来手法と提案手法を ITE 標準動画を用いて評価した結果、提案手法は、符号化レートを抑えつつ、高品質であることが示された。

1 はじめに

近年では、超高精細度テレビジョン放送サービスの高度化が取り組まれている。その一つである知的画像処理（超解像処理）を用いる方式は、低レートで高精細な映像伝送が可能であると報告されてきた。超解像処理方法とは、低解像度映像を符号化し、デコード側の超解像ネットワークで高解像度化することである。この方式は殆どの映像で有効である一方、デコード側の超解像ネットワークをどれだけ高性能化しても、未知の映像を高解像度化する際には映像品質が十分ではない、最悪、劣化してしまう可能性がある。

そこで、未知の映像に対しても高精細な映像を伝送可能とする MPEG NNR 深層学習メタデータを利用した超解像方式を検討する。

2 提案手法

提案手法では、事前学習した超解像用の Convolutional Neural Network（CNN）に対し、未知の映像でファインチューニングする。そのネットワークを MPEG の Neural Network Representations（NNR）[1] を用いて圧縮・伝送し、再構成した新たなモデルによって低解像度映像を高解像度化する。これにより、符号化レートを抑えつつ、映像品質を向上することが可能である。

3 実験

実験では、従来の伝送方式で符号化・復号した 8K 映像と、VVC の Reference Picture Resampling（RPR）で 1/2 縮小圧縮し、VVC の備える線形フィルタで拡大して得られた 8K 映像、RPR 縮小で圧縮し、復号映像を従来の超解像方式で拡大した 8K 映像、同じく RPR 縮小圧縮後、提案の NNR を利用した超解像方式で拡大した 8K 映像との符号化レートと映像品質を比較した（以降、従来の超解像方式をベースモデル、NNR を利用した超解像方式をファインチューニングモデルと呼ぶ）。それぞれ

の符号化・復号化には Versatile Video Coding（VVC）方式のテストモデル VTM10.0 [2]を利用した。超解像手法は、Residual 構造を取り入れた小規模な CNN を構築し、ベースモデル及びファインチューニングモデルで使用した。超解像ネットワークの学習では、SJTU 4K Video Sequences（SJTU）[3]と Ultra Video Group（UVG）[4]のデータセットを用いた。また、従来手法、提案手法の評価では、ITE の超高精細・広色域標準動画像の一部のデータを用い、伝送及び精度の検証をした。

ベースモデルの学習では、SJTU と UVG のデータセットを用いて、各種の量子化パラメタ（QP）で符号化したデータを使用した。具体的なデータの作成方法としては、QP は 22、27、32、37、42、47 とし、RPR を利用して 1/2 縮小して符号化、復号した。ファインチューニングモデルの学習時には、ファインチューニングモデルの重みを学習済みのベースモデルの重みで初期化し、伝送する映像を学習した。

ベースモデルとファインチューニングモデルの損失関数には L1、最適化関数には Adam を用いた。

復号された評価映像の品質は、映像配信事業会社である Netflix が提案した客観値 Video Multi-Method Assessment Fusion（VMAF）によって評価した。VMAF は動画の品質を評価するための指標であり、人の主観に近い評価が可能であるため、本指標を利用した。

4 結果・考察

提案手法において、NNR で圧縮したネットワークパラメタのデータ量は、約 10 秒に 1 回伝送とした場合 90kbps となった。実験によって得られた符号化レートと VMAF の関係例を図 1 に示す。NNR を用いた提案手法は、比較手法と比べて高精度であった。また、提案手法は符号化した縮小映像に加えて、NNR によって圧縮された超解像ネットワークのパラメタを伝送しているが、その符号化レートは低く、効率的に映像を高解像度化できていると考えられる。これらのことから本手法の有用性が示された。

図 1: 提案手法と比較手法の Bit-rate と VMAF の関係例
ITE 標準動画 A シリーズ No.7 の桂川(River)の結果

5 おわりに

超解像を利用した符号化映像伝送において、超解像ネットワークパラメタを NNR により圧縮して符号化映像と共に伝送する方法により、映像品質を向上させる手法を提案した。ITE 標準映像の一部のデータを用いて評価した結果、従来よりも効率的に高品質な映像を伝送可能なことを確認できた。

参考文献

[1] Test Model 3 of Compression of Neural Networks for Multimedia Content Description and Analysis, ISO/IEC JTC 1/SC 29/WG 11 N18993, Jan. 2020

[2] VVC test model (VTM), https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM

[3] L. Song, et al.: “The SJTU 4K Video Sequence Dataset,” the Fifth International Workshop on Quality of Multimedia Experience (QoMEX2013), July 3rd-5th, 2013

[4] A. Mercat, et al.: “UVG dataset: 50/120fps 4K sequences for video codec analysis and development,” in Proc. ACM Multimedia Syst. Conf., June 2020

シャープ株式会社 研究開発事業本部 通信・映像技術研究所 第３研究室
〒261-8520 千葉県千葉市美浜区中瀬 1–9–2
Phone: 050-5370-8584
E-mail : takuya.suzuki@sharp.co.jp