

HEVC への深層学習によるフレーム補間の適用

清水盛偉^{†1} 程 正雪^{†1} 孫 鶴鳴^{†1†2} 竹内健^{†1} 甲藤二郎^{†1}

概要: 近年、深層学習によるフレーム補間技術が進歩している。本研究では、動画圧縮の規格 HEVC に深層学習ベースのフレーム補間を組み合わせた新たな圧縮の手法を提案する。具体的にはまず、奇数フレームのみを HEVC で符号化し、復号の際に偶数フレームを補間する。また、補間されたフレームと実際のフレームの差分情報の圧縮も検討する。動画によっては提案手法が HEVC のみによる圧縮をレートひずみ特性として超える事が確認できた。

キーワード: 動画圧縮, フレーム補間, 深層学習

Application of Deep Learning Based Frame Interpolation to HEVC Video Coding

JOI SHIMIZU^{†1} ZHENGXUE CHENG^{†1} HEMING SUN^{†1†2}
MASARU TAKEUCHI^{†1} JIRO KATTO^{†1}

Abstract: In recent years, we have seen great progress in frame interpolation by deep learning. In this paper, we propose a novel video compression method which combines deep learning based frame interpolation and HEVC, which is one of the video compression standards. We only compress the odd frames by HEVC and interpolate the even frames by frame interpolation. Furthermore, we also compress the residual frames between original frames and interpolated frames. Results show that our method can outperform HEVC in some scenarios.

Keywords: video compression, interpolation, deep learning

1. はじめに

動画の通信は年々増加しており、2022 年までに動画のトラフィックは IP トラフィック全体の 80%以上を占めると予測されている[1]。したがって、動画の保存や通信のためにはより高効率な圧縮符号化技術が必要とされている。

近年、[2]や[3]のような深層学習をベースとした画像圧縮の研究が進歩している。深層学習は[4]や[5]の様な動画圧縮にも利用されている。[4]では深層学習ベースの画像圧縮とフレーム補間を組み合わせ、動画圧縮を実現している。本稿では動画圧縮規格 HEVC とフレーム補間を組み合わせ、より高効率な圧縮を実現する。

2. 関連研究

従来の圧縮手法に JPEG2000[6]や HEVC[7]がある。これらは画像をブロック単位に分け、離散コサイン変換、量子化、エントロピー符号化を経て圧縮される。一方、[2]や[3]は深層学習をベースとした画像圧縮である。これらは共に VAE (Variational AutoEncoder) を変形し、非線形変換に加え、算術符号器の確率モデルを更新するネットワークを導入している。

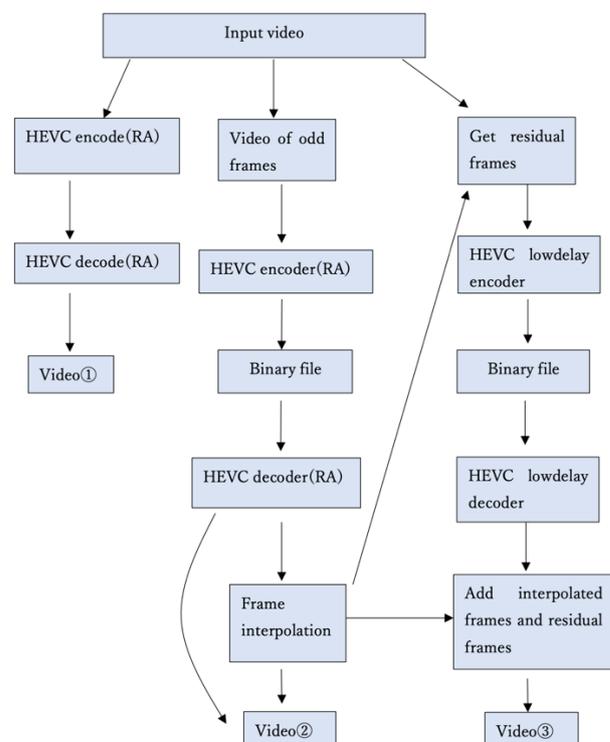


図1 提案手法の概要図

^{†1} 早稲田大学
Waseda University
^{†2} JST, PRESTO

深層学習は[4]や[5]の様な動画圧縮にも利用されている。[4]では深層学習ベースの画像圧縮と両方向予測によるフレーム補間を組み合わせ、動画圧縮を実現している。[5]は深層学習ベースの画像圧縮とオプティカルフローを組み合わせ、low delay 圧縮を実現した。

3. 提案手法

提案手法の概要を図1に示す。始めに提案手法と比較するために、入力動画を HEVC で圧縮する。コンフィギュレーションには random access で GOP が 8 のものを用いた。HEVC でエンコードし、デコードによって得られた動画を Video①とする。次に入力動画の奇数フレームのみからできた新たな動画を準備する。この新たな動画を Video①と同様に圧縮する。デコード後に得られたフレームの間を、深層学習をベースとしたアルゴリズムで補間することで、偶数フレームを得る。ここで得られる動画を Video②と呼ぶ。圧縮の精度向上のために差分フレームを活用し、元の偶数フレームのピクセル値から補間されたフレームのピクセル値を引く。この差分フレームの圧縮には HEVC の low delay を用いた。最後に復号された差分フレームを補間フレームに足し合わせ、新たな動画、Video③を得る。

4. 評価実験

4.1 偶数フレームの補間

はじめに Video①と Video②の比較を行う。Video①では GOP を 8 とした random access が用いられた。QP は 27, 32, 37, 42 とした。奇数フレームの動画でも同じ QP 値を用いた。偶数フレームの補間には Super Slomo[8], DAIN[9]の2種類を用いた。本稿では Johnny と BasketballDrill の2つのシーケンスの結果を示す。Johnny のフレームのサイズは 1280*720, BasketballDrill は 832*480 である。

この実験の結果は図2と図3に示す。青い線は HEVC, 水色の線は Super Slomo による補間, 黄緑は DAIN による補間を用いた結果を示す。

どちらのシーケンスでも Super Slomo より DAIN の方が補間の結果が良いことが分かる。また、提案手法は BasketballDrill より Johnny での方が良い精度を示している。これはシーケンス内での動きの大きさに関係がある。図4は連続フレーム(フレーム 10 とフレーム 11)のピクセル値の違いを示す。

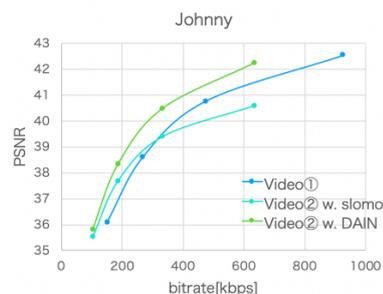


図2 Johnny の Video①と Video②の比較

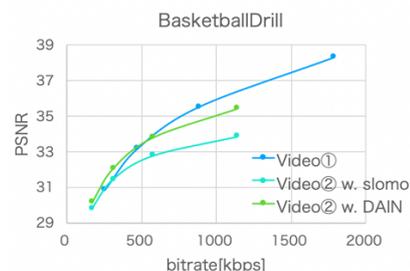


図3 BasketballDrill の Video①と Video②の比較

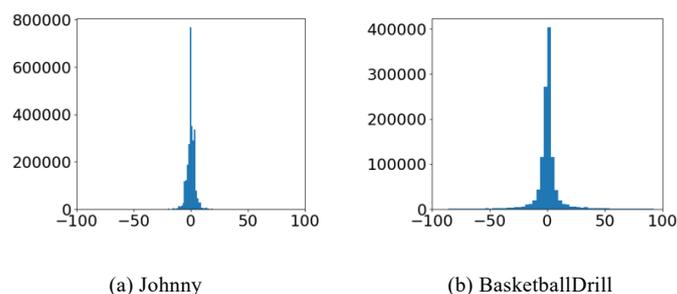


図4 連続フレームのピクセル値の差

図4から Johnny の方がピクセル値の違いが0に集中していることが分かる。つまり、Johnny の方が動画内での動きが少ないことを示している。Johnny では HEVC による符号化よりも DAIN による補間を用いた手法が精度で上回っている事が分かる。

BasketballDrill は動きが大きく、連続フレーム間の差分値も大きくなる。よって BasketballDrill でのフレーム補間は困難である。図3では低いビットレートでは HEVC と同程度の圧縮効率が得られるが、ビットレートが大きくなれば HEVC に比べて補間のみでは限界があることがわかる。

4.2 偶数フレームの補間

次に精度向上のために差分フレームの圧縮を検討する。偶数フレームの補間には DAIN を用いた。差分フレームは HEVC の low delay モードで符号化する。この際の QP の値は表1のように決定した。結果は図5, 図6に示す。

表 1 差分に用いられた QP

QP for the odd frames video (random access)	QP for the residual frames (low delay)
27	32
32	37
37	42
42	Residual frames were not used

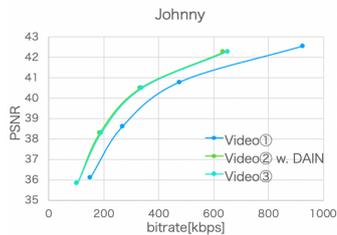


図5 Johnny の Video①, Video②, Video③の比較

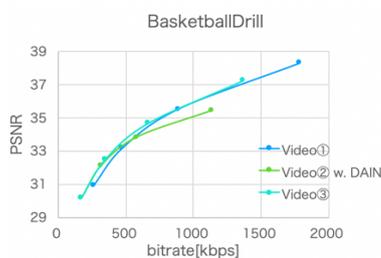


図6 BasketballDrill の Video①, Video②, Video③の比較

Johnny では Video②と Video③の精度に違いが見られなかった。これは差分がすでに 0 に近かったからであると考えられる。BasketballDrill では Video③の結果が HEVC の結果と同じ程度となった。このことから差分フレームの圧縮により、動きの激しい画像でも、提案手法は HEVC と同程度の圧縮効率を実現できる事が分かった。

5. まとめと今後の課題

本稿では深層学習によるフレーム補間を用いた新たな圧縮手法を提案した。偶数フレームの補間には Super Slomo よりも DAIN が適しており、提案手法によって HEVC を精度で上回るシーケンスがあることが確認できた。また、差分フレームの圧縮により提案手法の圧縮効率がより安定することが確認できた。BasketballDrill のように動きの大きいシーケンスでも HEVC に匹敵する精度が得られた。今後は VVC で同様の実験を行いたいと考えている。

参考文献

[1] Cisco Visual Networking Index, “Forecast and Trends 2017-2022 whitepaper”, 2019.
 [2] J. Balle, D. Minnen, S. Singh, S. J. Hwang, N. Johnston, “Variational Image Compression with a Scale Hyper-prior”, ICLR, 2018.

[3] Z. Cheng, H. Sun, M. Takeuchi, J. Katto, “Learning Image Compression with Discretized Gaussian Mixture Likelihoods and Attention Modules”, IEEE CVPR 2020.
 [4] Z. Cheng, H. Sun, M. Takeuchi, J. Katto, “Learning Image and Video Compression through Spatial-Temporal Energy Compaction”, IEEE CVPR 2019.
 [5] G. Lu, W. Ouyang, D. Xu, X. Zhang, C. Cai, Z. Gao, “DVC: An End-to-end Deep Video Compression Framework”, IEEE CVPR, 2019.
 [6] M. Rabbani, R. Joshi, “An overview of the JPEG2000 still image compression standard”, ELSEVIER Signal Processing: Image Communication, vol. 17, no. 1, pp. 3-48, Jan. 2002.
 [7] G. J. Sullivan, J. Ohm, W. Han and T. Wiegand, “Overview of the High Efficiency Video Coding (HEVC) Standard”, IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1649-1668, Dec. 2012.
 [8] H. Jiang, D. Sun, V. Jampani, M. Yang, E. Miller, J. Kautz, “Super SloMo: High Quality Estimation of Multiple Intermediate Frames for Video Interpolation”, IEEE CVPR, 2018.
 [9] W. Bao, W. Lai, C. Ma, X. Zhang, Z. Gao, M. Yang, “Depth Aware Video Frame Interpolation”, IEEE CVPR, 2019.