I-017

DNN に基づく変換行列を用いたフレーム補間性能の符号化雑音依存特性

Correlation between Image Compression Noise and Frame Interpolation Performance using DNN-based Transformation Matrices

神保 悟† 王 冀† 八島 由幸†
Satoru Jimbo Ji Wang Yoshiyuki Yashima

1. はじめに

近年，動画像符号化へ深層学習を応用する様々な検討が行われており，PU/CU 分割の決定手法やループ内フィルタ，デコード画像の画質推定など，符号化効率の改善や計算量の削減を狙う手法等が提案されている[1][2][3]．筆者らは，フレーム間予測効率向上を目的として，畳み込みニューラルネットワーク（CNN）を用いて 2 つの参照フレームからフレーム補間のための変換行列を推定し，参照フレームにその変換行列を乗算することで中間フレームを生成する手法を提案した[4]．この手法は任意精度の平行移動，拡大縮小に加え，参照フレームにフィルタの畳み込みを行った補間できるため，参照フレームに雑音が生じていても効果的な補間が可能であると考えられる．本検討では，H.265/HEVC による符号化雑音が生じた画像を参照フレームとして，[4]で提案した手法に基づいて中間フレームを生成した場合の予測特性を評価したので報告する．

2. 提案手法

2.1 DNN に基づく変換行列を用いたフレーム補間

いま，予測対象フレームを I，時間的に前の参照フレームを F，時間的に後ろの参照フレームを G とし，F と G を用いて I を予測することを考える．HEVC 等で用いられる動き補償フレーム予測では，エンコーダ側で F,G と I の位置のずれを動きベクトルとして算出してデコーダに送信し，デコーダ側で動きベクトルに基づいて I を生成する．これに対し，本検討では，デコーダ側で CNN を用いて F,G から 4 つの変換行列M_1, M_2, M_3, M_4を推定し，変換行列を F,G に式(1)のように乗算することによってエンコーダ側から情報を送信することなく I を生成する．

$$I = M_1(F-\mu)M_2 + M_3(G-\mu)M_4 + \mu \quad (1)$$

$$\mu = \frac{\sum F + \sum G}{2n} \quad (2)$$

このとき，画素値の平均 μ を分離して乗算を行い，最後に加算している．n は画素数を表している．

この手法の特徴として，単純な形の行列を用いることで任意の精度の平行移動や拡大縮小が表現できることが挙げられる．たとえば，図 2(a)をM_1, M_3として乗算を行うと下に 1/4 だけ平行移動した画像が得られ，図 2(b)をM_1, M_3として乗算を行うと垂直方向に 1.2 倍拡大した画像が得られる．また，図 2(a),(b)の行列を乗算した行列をM_1, M_3として用いることで，両者の処理を同時に行ったフレームを生成できる．

† 千葉工業大学大学院情報科学研究科, Graduate School of Information and Computer Science, Chiba Institute of Technology

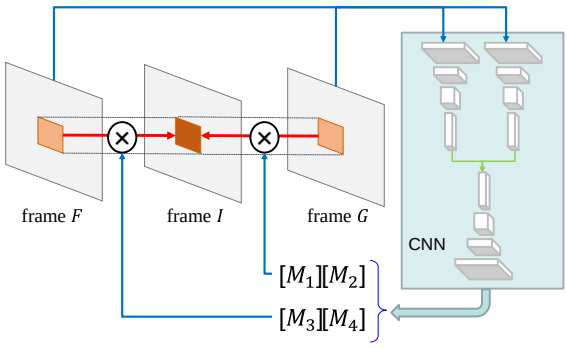


図 1 変換行列によるフレーム補間

$$\begin{bmatrix} 3/4 & 1/4 & 0 & 0 & \cdots \\ 0 & 3/4 & 1/4 & 0 & \\ 0 & 0 & 3/4 & 1/4 & \\ 0 & 0 & 0 & 3/4 & \\ \vdots & & & & \ddots \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots \\ 1/6 & 5/6 & 0 & 0 & \\ 0 & 2/6 & 4/6 & 0 & \\ 0 & 0 & 3/6 & 3/6 & \\ \vdots & & & & \ddots \end{bmatrix} \quad \begin{bmatrix} 1/4 & 1/2 & 1/4 & 0 & \cdots \\ 0 & 1/4 & 1/2 & 1/4 & \\ 0 & 0 & 1/4 & 1/2 & \\ 0 & 0 & 0 & 1/4 & \\ \vdots & & & & \ddots \end{bmatrix}$$

(a)平行移動 (b)拡大縮小 (c)平滑化+平行移動

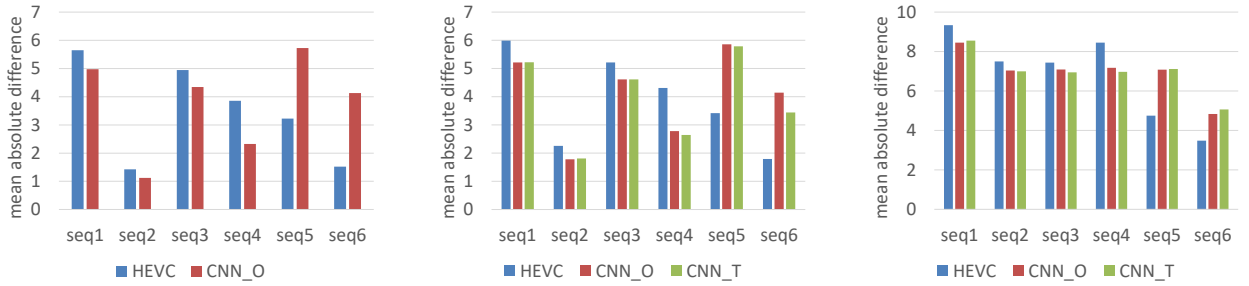
図 2 変換行列の例

さらに，この手法では参照フレームに対してフィルタの畳み込みなどの様々な処理を加えた補間が行える．図 2(c)はその一例で，M_1, M_3として乗算を行うと参照フレームに対して垂直方向の平滑化処理を行いつつ下に 1 画素平行移動した画像が得られる．これによって，圧縮符号化によって参照フレームに雑音が生じている場合でも，ループ内フィルタを施すように参照フレームの雑音を軽減したフレーム補間が可能になると予想される．

2.2 CNN による変換行列の推定

変換行列の推定には 3 個の畳み込み層からなる ResidualBlock を 8 層と最終的な出力を行う畳み込み層 1 層で構成される計 25 層の CNN を用いる[4]．なお，CNN の最終層以外の畳み込み層では ReLU 関数による活性化，バッチ正規化を行っている．CNN におけるフィルタのサイズはすべて 3×3 である． 2 つの参照フレーム F，G を CNN に入力する時には，同じ空間位置にある$N_1 \times N_1$サイズのブロックB_F, B_Gを参照ブロックとして取り出す．これらを (1)式に基づいて，出力された変換行列と参照ブロックとを乗算して予測ブロック$\widehat{B_I}$を生成し，$\widehat{B_I}$の中央の$N_2 \times N_2$領域（$N_2 \leq N_1$）と正解ブロックB_Iの対応する領域の mean absolute difference(MAD)を求め，MAD を損失関数として CNN の更新を行う．実際の予測値には CNN 出力として得られる$N_1 \times N_1$サイズのブロックの中央の$N_3 \times N_3$（$N_3 \leq N_2$）領域内の画素を用いる．フレームの端を入力するときはゼロパディングを行っている．CNN の最適化手法として Adam を使用し，訓練回数は 500 万回，学習率は 0.0001 とした．

Copyright © 2018 by
The Institute of Electronics, Information and Communication Engineers and
Information Processing Society of Japan All rights reserved.



(a) 参照フレーム： 原画像　(b) 参照フレーム：HEVC-Intra(QP22)　(c) 参照フレーム：HEVC-Intra(QP37)

図 3　予測画像の mean absolute error

3. 実験と考察

3.1 実験条件

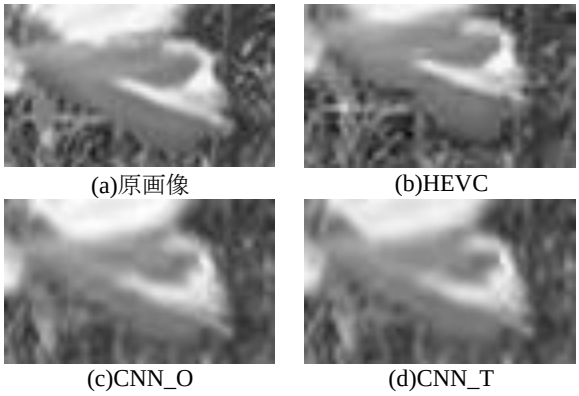
原画像，H.265/HEVC イントラ予測符号化画像（QP=22 および QP=37）で符号化した 3 種の画像を参照フレームとして用いる．深層学習ベース雑音除去フィルタの研究[5]によると，全ての QP に適応した CNN を訓練することは難しく，[6]では，QP ごとに別々の CNN を訓練し，入力する画像の QP によって CNN を使い分けている．そこで，原画像を用いて訓練を行った CNN（CNN_O），性能評価時と同じ QP で符号化した画像を用いて訓練を行い， それぞれの QP にチューニングした CNN（CNN_T）の 2 つの条件で実験を行った．

CNN の訓練にはハイビジョン・システム評価用標準動画像第 1 版から 450 フレームからなる 60 種類の動画を使用し，評価時には同第 2 版から 6 種のシーケンス seq1~seq6 を使用した．seq1~seq4 は動きが小さいシーケンスであり，seq5,6 はフレーム間の動きが大きいシーケンスである．ブロックサイズは，予備実験からN_1=80, N_2=32, N_3=16 とした．性能評価は，CNN による予測フレームと正解フレームとの MAD を用いた．比較のために，HEVC に基づくイントラ予測および動き補償フレーム間予測に基づくフレーム内挿結果も算出した．HEVC 予測では， 参照フレームにはデブロッキングフィルタ及び SAO による処理を施している．

3.2 実験結果

図 3 に MAD の測定結果を示す．図 3 より，動きが小さいシーケンスでは，参照フレームに雑音が生じていても，提案手法が優れていることが確認できる．一方，動きが大きいシーケンスでは，提案手法は HEVC 予測よりも性能が劣るが，QP が大きくなるにつれて両者の差が小さくなっていることが確認できる．図 4 は seq1 について，QP=37 で符号化された画像を参照フレームとして処理を行った例である．図 4 より，提案手法では HEVC 予測と比べて，よりブロックノイズの軽減が行われ，オブジェクトの輪郭を保持できている．

また， CNN と CNN_T では，補間フレームの MAD に大きな差がなく，深層学習ベース雑音除去フィルタで用いられるような QP に対するチューニングの必要性は低いことが確認できる．これは，雑音除去フィルタの方では CNN がフィルタそのものの役割をしているのに対し，提案手法では CNN は参照フレームからフィルタ処理を行う行列を生成する役割をしているため，雑音の量に柔軟に対応できているためだと考えられる．



(a)原画像　(b)HEVC　(c)CNN_O　(d)CNN_T

図 4　補間フレームの例（QP=37）

4. おわりに

CNN によって推定された変換行列に基づくフレーム補間は，動きが小さいシーケンスでは，符号化雑音が生じた画像を参照フレームに使用した場合でも，HEVC 予測よりも優れた性能を発揮することが確認された．

今後は，提案手法を拡張し，HEVC の低遅延モードや様々なランダムアクセスモードへの適用可能性や，MAD による評価のみでなく，復号画質と発生符号量を考慮した RD 最適化に基づく検討を行う予定である．

参考文献

[1] X. Yu, Z. Liu, J. Liu, Y. Gao, D. Wang, “VLSI Friendly Fast CU/PU Mode Decision for HEVC Intra Encoding: Leveraging Convolution Neural Network,” In IEEE International Conference on Image Processing (ICIP), pp.1285–1289, 2015.

[2] S. Bosse, D. Maniry, K. R. Müller, T. Wiegand, W. Samek, “Neural network-based full-reference image quality assessment,” In Picture Coding Symposium (PCS), pp. 1–5, 2016.

[3] W. S. Park and M. Kim, “CNN-based in-loop filtering for coding efficiency improvement,” In IEEE 12th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP), pp. 1–5, 2016.

[4] 神保悟, 王冀, 八島由幸, “深層学習による変換行列予測を用いたフレーム補間,” 2017 年映像情報メディア学会冬季大会, 2017.

[5] 神保悟, 王冀, 八島由幸, “量子化幅適応型ディープラーニングを用いた H.265/HEVC 符号化雑音除去,” 第 16 回情報科学技術フォーラム, I-011, 2017.

[6] C. Li, Li. Song, R. Xie, W. Zhang, “CNN based post-processing to improve HEVC,” In IEEE International Conference on Image Processing (ICIP), pp. 4577–4580, 2017.

Copyright © 2018 by
The Institute of Electronics, Information and Communication Engineers and
Information Processing Society of Japan All rights reserved.