I-021

LSTM による JPEG/HEVC ビットストリームの 画像認識精度に関する考察
A Study on Image Recognition Accuracy for JPEG/HEVC Bitstream using LSTM

冨田 直生† 八島 由幸†
Naoki Tomita Yoshiyuki Yashima

1. はじめに

近年，画像認識の分野において深層学習を用いた画像認識手法が多く提案され，従来よりも優れた認識性能を発揮することが示されている．代表的な例として，画像のピクセルデータを入力し畳み込みニューラルネットワーク（CNN）を設計する手法が挙げられる．一方，我々が日常的に扱う画像は JPEG，H.265/HEVC（以下 HEVC）等の圧縮されたビットストリーム状態で存在することが多い．ビットストリームから直接画像認識を行う試みとして，ビットエラーが混入した圧縮ストリームを直接認識する試み[1]や，JPEG 符号化における符号化パラメータと認識精度に関する考察[2]が挙げられるが，符号化方式と圧縮ビットストリームを得る際のパラメータにより認識制度がどのように変化するかは明らかになっていない．本検討では，JPEG，HEVC で圧縮符号化された画像のビットストリームから直接画像認識を行うことを試み，JPEG，HEVC における符号化パラメータと画像認識精度の関係を明らかにする．

2. 圧縮バイナリデータの学習

2.1 訓練データ

本検討では，代表的な画像圧縮符号化方式である JPEG と HEVC に焦点を当て実験を行う．データセットは MNIST を用いた．データの前処理として，画像の幅に対し 20%の範囲で水平方向へシフトする処理と水平反転を施した．これを JPEG，HEVC で符号化し，バイナリデータを 8bits ごとにまとめたものを訓練データとする．圧縮ビットストリームの例を図 1 に示す．HEVC は JPEG と比較し圧縮効率に優れているため、圧縮後の有効ビット長が短くなっている．符号化には Python の画像処理ライブラリである Pillow[3]と，HEVC のリファレンスソフトウェアである HM(HEVC Test Model)[4]を用いた．また，訓練の際はバッチごとのビットストリーム長が等しくなるよう，最長のものにあわせてパディングを施した．

2.2 ネットワーク

JPEG，HEVC によって得られるビットストリームはブロックごとにラスタースキャン順に取得され，いずれも可変長となる．符号化前の画像の画素間には空間的な相関があることから，得られた圧縮データには時系列的な特徴が存在すると考えられる．また，コンテキスト適応型符号化である HEVC では，時系列上で先行する符号化状態が現在の符号化に反映される．このような観点から，圧縮バイナリデータの学習に用いるニューラルネットワークは，図 2 に示すように時系列データに適した LSTM (Long short-term memory)とする．

† 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

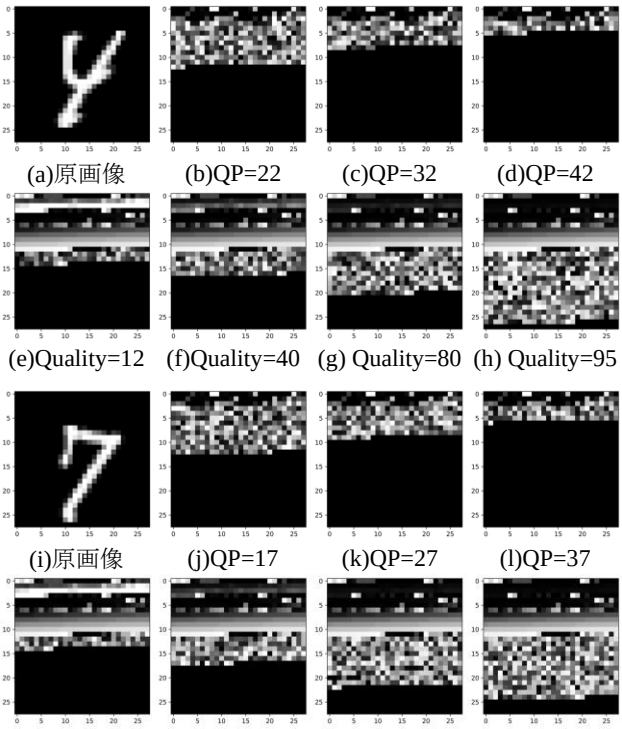


図 1 圧縮バイナリデータの例
（(b)～(d)，(j)～(l):HEVC, (e)～(h)，(m)～(p):JPEG）

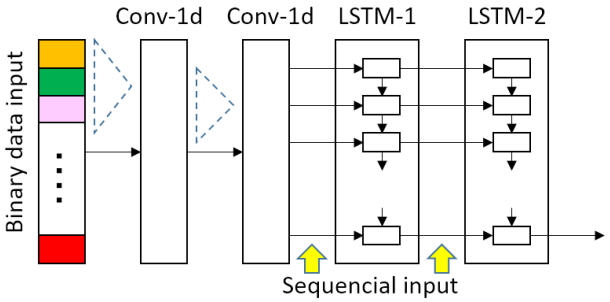
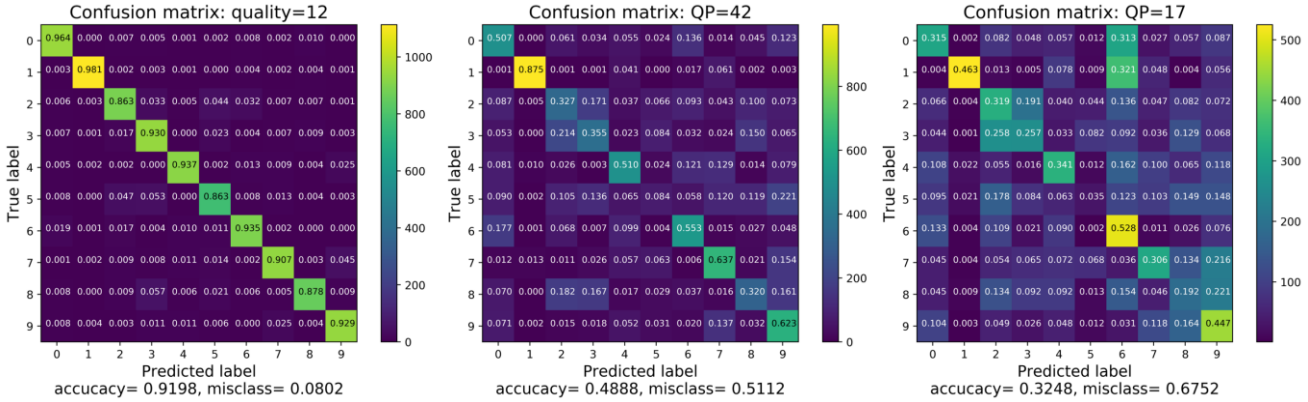


図 2 ネットワーク図

3. 実験

3.1 実験条件

JPEG 符号化では，DCT を行う際に画像を 8×8 ピクセルのブロックに分割し処理を行っている．このことから本検討では，HEVC 符号化における TU サイズを 8×8 ピクセルとした．HEVC の量子化パラメータ(QP)は 17, 22, 27, 32, 37,



(a)JPEG:Quality=12　(b)HEVC:QP=42　(c)HEVC:QP=17

図 3　各パラメータにおける混同行列の例

42 の 6 つの値について実験する．JPEG における品質パラメータ(Quality)は，QP=42 における HEVC ビットストリームと近い発生ビット数となる Quality=12 と，QP=27,32,37 における復号画像の PSNR が近い値となる Quality=90,80,60, および Quality=40 に設定する．訓練では，バッチサイズを 512 とし，損失関数として CrossEntropyLoss，最適化関数として Adam を用いた．また，学習率を 0.005 とした．推定時には訓練データと同じ量子化パラメータを使用し，認識精度の評価指標はテストデータに全体に対する認識正解率(Accuracy)とする．

3.2 実験結果

図 3 に，テストデータに対する各パラメータでの正解ラベルと推定ラベルの混同行列の例を示す．また，図 4 に，HEVC の量子化パラメータ QP と JPEG の画質パラメータ/Quality に対する，認識正解率を示す．HEVC では QP が大きい値ほど ビットストリーム長は短く低画質になり，JPEG では Quality が大きい値ほどビットストリーム長は長く高画質になる．

図 3 より，JPEG において誤認識した際の推定ラベルと正解ラベルの関係は対称に近いことがわかる．一方，HEVC ではばらつきが大きい．これは，MNIST のクラス分類において重要である数字のエッジ特徴を後者ではうまく認識できていないためと考えられる．

図 4 から，HEVC と JPEG の認識精度を比較すると，JPEG では高画質ほど認識精度が高くなる傾向がみられる．反対に，HEVC では低画質ほど認識精度が高くなる傾向がみられた．また，JPEG においては Quality の値にかかわらず 90%を超える正解率となっているが，HEVC では最も高い正解率で 50%程度にとどまる．

JPEG ではブロックごとに DCT によって周波数成分に変換された特徴量が，固定の符号化テーブルを用いて符号化され系列データとして出力されるが，HEVC ではコンテキスト適応型符号化によって，先行して符号化されたシンボルに基づき計算されたコンテキスト情報に応じて符号化テーブルを作成する．このため，出力されたビットストリームを 8bits ごとにまとめたシンボルの時系列的な特徴に，局所的な DCT 周波数成分情報が反映されづらく，本実験で用いたニューラルネットワークではうまく特徴を学習できていないと考えられる．また，文献[2]によると訓練時と推定時の量子化パラメータの差が小さいほど推定精度が高

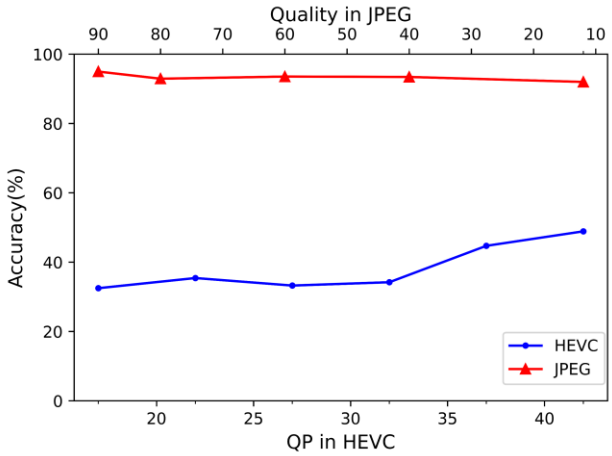


図 4　符号化パラメータの正解率の関係

く，量子化パラメータの差が大きくなるにつれ精度が低くなる傾向がある．さらに，MNIST において画質が認識正解率に与える影響は小さいと考えられる．このことから，認識正解率を低下させる要因の一つに生成されるビットストリーム長のばらつきが挙げられる．そのため，QP の値が大きくビットストリーム長が短いほどコンテキスト情報の影響やビットストリーム長のばらつきが小さくなり，認識正解率が高くなる傾向があると考えられる．

4. おわりに

本検討では，HEVC で符号化された画像の圧縮ビットストリームからの直接認識を試み，量子化パラメータと認識率との関係を明らかにするとともに，JPEG 符号化ビットストリームと比べて，認識精度が低下することを示した．今後はコンテキスト適応型符号化においても高精度で認識可能な訓練方法の模索や，異なるニューラルネットワークによる認識手法について検討を進める予定である．

なお本研究は，JSPS 科研費 JP18K11360 の助成を受けて行った．

参考文献

[1] https://github.com/Hi-king/compressed_image_recognition
[2] 冨田 直生, 八島 由幸, “RNN/LSTM を用いたビットストリームからの画像認識制度に関する考察”, 2019 年画像符号化／映像処理シンポジウム, P-2-10, (2019).
[3] https://pillow.readthedocs.io/en/stable/
[4] https://hevc.hhi.fraunhofer.de/