

DCT係数のヒストグラムの偏りを利用した JPEG画像の編集履歴解析

小浦 啓太郎^{1,a)} 栗林 稔¹ 船曳 信生¹

概要: デジタルカメラで撮影された画像は通常、メモリデバイスに保存する際に JPEG 圧縮が 1 回行われている。一方、改ざんされた画像は編集後に再度圧縮保存されることから、少なくとも 2 回の圧縮となる。そこで、画像から 2 重圧縮によって生じた特徴を解析することによって、画像の改ざんを検知する方法が研究されている。その特徴は、1 回目と 2 回目の JPEG 圧縮時における量子化テーブルの相違により、DCT 係数値の統計的なヒストグラムの偏りとして現れる。本研究では、量子化テーブルの特徴を考慮した 2 重 JPEG 圧縮を検知する方式を提案する。1 回目の圧縮時に量子化する際の品質パラメータ (QF) が 2 回目の圧縮時より大きい場合、DCT 係数のヒストグラムの値が偏る。その特徴を検知するために、量子化テーブルにおいて収集しやすい DCT 係数値を複数扱う。収集した DCT 係数値のヒストグラムを入力とし、畳み込みニューラルネットワーク (CNN) に学習させることで、1 回圧縮と 2 重圧縮を分類する。

キーワード: JPEG 圧縮, クオリティファクタ, 量子化テーブル, ヒストグラム

Analysis on Editing History of JPEG File Based on the Bias in Histogram of DCT Coefficients

KEITARO OURA^{1,a)} MINORU KURIBAYASHI¹ NOBUO FUNABIKI¹

Abstract: Detection of double JPEG compression plays an important role in digital forensics. As we can find some clues from a given image about the history of JPEG compression, it is possible to classify the image if it is tampered or not. The characteristic of double compression appears at the statistical distribution of histogram in DCT coefficients because of quantization with different QFs. Unfortunately, such characteristics in the histogram are distorted by the rounding operation occurred at the process of decompression. In this study, considering the value of quantization table, we select some of them to collect useful statistical information as possible. We propose a CNN-based classifier suitable for the analysis of this histogram in order to classify a given image into single compression or double compression.

Keywords: JPEG compression, quality factor, quantization table, histogram

1. まえがき

JPEG は画像符号化の国際標準であり、Joint Photographic Experts Group によって考案された。JPEG は人の視覚特性を利用し、画像の一部のデータを省略することで圧縮符号化を行っている [1], [2]。そのため、JPEG は符

号化データから元の画像を完全に復号することができない非可逆圧縮である。JPEG 画像の改ざんを行う場合、すでに JPEG 圧縮が行われた画像に対してまずは伸張してから改ざん処理を行い、その後再度圧縮を行うため、少なくとも 2 回以上の圧縮が行われた画像となっている。複数回行われた JPEG 圧縮処理によって生じる歪みの特徴を検知することで、改ざんを検知することが可能となる。JPEG 圧縮の過程において処理が不可逆となる要因は、量子化処理と丸め込み処理である。これらによって 2 重圧縮された画

¹ 岡山大学大学院
Okayama University

a) oura.keitaro@s.okayama-u.ac.jp

像には、これらの不可逆な処理によって生じる統計的な特徴が現れる [3].

本研究では、量子化テーブルの特徴を考慮し、畳み込みニューラルネットワークを用いて、2重 JPEG 圧縮を検知すること目的としている。その際、2重圧縮される間に画像に対して加工処理が行われていない場合、メディアンフィルタを適用した場合の2種類を考える。提案方式では、対象画像に対して、量子化テーブルの特徴を考慮した DCT 係数のヒストグラムを作成し、畳み込みニューラルネットワークを用いて分類する。

2. 基礎技術

2.1 JPEG 画像

JPEG 圧縮符号化は、まず入力画像を輝度成分と色差成分に変換し、それぞれの成分に対して処理を行う。人間の視覚的な特徴を考慮して、色差成分は2分の1または4分の1に間引くことで大幅な情報削減がなされる。それぞれの成分の符号化の際には、 8×8 画素のブロックごとに離散コサイン変換 (Discrete Cosine Transform: DCT)、量子化、エントロピー符号化を行い、符号化データを生成する。復号においては、符号化データに対してエントロピー復号化、逆量子化、IDCT (Inverse DCT) を行い、復元画像を生成する。

2.1.1 量子化

DCT によって出力された DCT 係数のブロック内において、より右下に位置する係数ほど高い周波数成分である。この高周波成分は精度を落として情報を削減しても、画質の劣化が目立ちにくい。そのため、高周波成分ほど大きい量子化ステップとなるように係数の位置ごとに異なる量子化ステップを用いて量子化することで、情報を削減する。このような位置ごとに異なる量子化ステップを設定した行列を量子化テーブルと呼ぶ。多くのアプリケーションでは JPEG の規格書に掲載されているテーブルを用いている。以下に輝度成分、色差成分のそれぞれに対する量子化テーブルの例を示す。ユーザーが任意に圧縮率を変えたい場合、

表 1: 輝度成分に対する量子化テーブル

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

この量子化テーブルを基本のテーブルとして、以下の計算式を用いて実際に量子化を行う量子化テーブルを算出する。

$$Q_{QF}(i, j) = \left\lfloor \frac{S \cdot Q_b(i, j) + 50}{100} \right\rfloor \quad (1)$$

$$S = \begin{cases} \frac{5000}{QF} & (QF < 50) \\ 200 - 2 \cdot QF & (QF \geq 50) \end{cases} \quad (2)$$

ただし、 QF は $0 \leq QF \leq 100$ を満たすユーザーが任意に設定できる品質パラメータであり、 $Q_b(i, j)$ は量子化テーブルである。JPEG 圧縮における量子化では、DCT 係数のブロック $F(i, j)$ を量子化テーブル $Q_{QF}(i, j)$ で割り、整数値に四捨五入した値を出力する。

$$F_q(i, j) = \text{round} \left(\frac{F(i, j)}{Q_{QF}(i, j)} \right) \quad (3)$$

逆量子化では量子化の際に使用した量子化テーブル $Q_{QF}(i, j)$ を量子化データ F_q にかけることで $F'(i, j)$ を復元する。

$$F'(i, j) = Q_{QF}(i, j) \cdot F_q(i, j) \quad (4)$$

2.2 畳み込みニューラルネットワーク

畳み込みニューラルネットワーク (Convolutional Neural Network: CNN) はニューラルネットワークの1種であり、主に画像認識で用いられ、その精度の高さから注目されている。CNN は一般的な順伝播型のニューラルネットワークと異なり、全結合層だけでなく、画像の局所的な特徴抽出を担う「畳み込み層」と、局所ごとに特徴をまとめ上げる「プーリング層」から構成される。CNN の畳み込み層で行われる畳み込み処理は図 1 に示すように、任意の小さなフィルタを移動させながら画像の左上から右下まで各要素の積の和を算出することである。

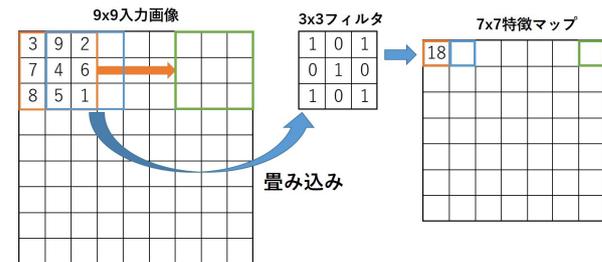


図 1: 畳み込み

プーリング層は基本的に、畳み込み層と次いで利用され、畳み込み層で出力された特徴マップを縮小する。その際、注目領域を設定し、その領域の特徴マップから新たに縮小された特徴マップを求める。代表的なプーリングの手法として、最大値プーリング、平均値プーリング、 L_p プーリングがあり、最も一般的に利用されるのは最大値プーリングである。最大値プーリングは注目領域の中の最大値を抽出し、新たな特徴マップとする。このプーリング処理は画像を縦横方向に間引いて行うため、1行1列ずつ注目領域

を移動させるのではなく、2行2列以上ずつスライドさせる。CNNでは以上により得られた特徴マップを全結合層の入力として学習させる [6].

3. 先行研究

本章では JPEG の 2 重圧縮の特徴についての先行研究を以下に説明する [3], [4]. 以下の説明における量子化は JPEG 圧縮の過程においては量子化後、逆量子化を行った処理を表す.

3.1 2重JPEG圧縮

デジカメ等の撮影機器で撮られた一般的な画像は、撮影時に1回のJPEG圧縮が行われている。その画像を改ざんする場合、圧縮ファイルを展開し、何らかの加工を行った後に再度JPEG圧縮が行われる。そのため、改ざん画像は少なくとも2回のJPEG圧縮が行われている。以上より、画像から2重のJPEG圧縮が行われていると判断できる場合、その画像は改ざんされていると考えられる。

先行研究では2重圧縮された画像を対象とし、そのDCT係数のヒストグラムにみられる特徴を利用して1回目の量子化テーブルの推定を目的としている。2重圧縮の特徴は1回目の圧縮時と2回目の圧縮時における量子化ステップの差によって現れるため、DCT係数のヒストグラムから1回目の量子化テーブルを推定することが可能である。一般的に、1回圧縮された画像を再度圧縮する際には、画像の歪みを抑えるために QF を高く設定する。そのため、 $Q_{QF}^1(i, j)$ は $Q_{QF}^2(i, j)$ よりも大きい値になる。次に、2重圧縮される過程においてDCT係数に生じる変化を考える。1回圧縮された画像においてDCT係数 $D^1(i, j)$ は1回目の量子化テーブル $Q_{QF}^1(i, j)$ によって量子化されている。つまり、ブロック内の $D^1(i, j)$ の値は対応する $Q_{QF}^1(i, j)$ の値の倍数になる。さらに、2回目の圧縮において1回目の圧縮時よりも小さい量子化ステップで量子化すると、ヒストグラムにおいて本来値が存在するはずの $Q_2(i, j)$ の倍数頻度が0となる。以下ではこれを0値と呼ぶ。

$Q_{QF}^1(i, j) = 4, Q_{QF}^2(i, j) = 3$ で2回の量子化を行ったDCT係数のヒストグラムを図2に示す。図2では、 $D^2(i, j) = 6$ の場合において、頻度が0となっている。これは2回目の量子化において、 $D^1(i, j) = \{5, 6, 7\}$ が $D^2(i, j) = 6$ に量子化されるが、1回目の量子化によって、 $D^1(i, j) = \{5, 6, 7\}$ の頻度が0となっているために $D^2(i, j) = 6$ の場合に頻度が0となる。このように $Q_{QF}^2(i, j) < Q_{QF}^1(i, j)$ の条件下では、ヒストグラムにおいて特定のDCT係数に対して0値が生じる。しかし、 $Q_{QF}^1(i, j)$ が $Q_{QF}^2(i, j)$ の倍数である場合、 $Q_{QF}^2(i, j) < Q_{QF}^1(i, j)$ の条件を満たしているにもかかわらず、DCT係数のヒストグラムに特徴がみられない。具体例として、 $Q_{QF}^1(i, j) = 6, Q_{QF}^2(i, j) = 3$ で量子化されたデータ系列のヒストグラムを図3に示す。図3を見る

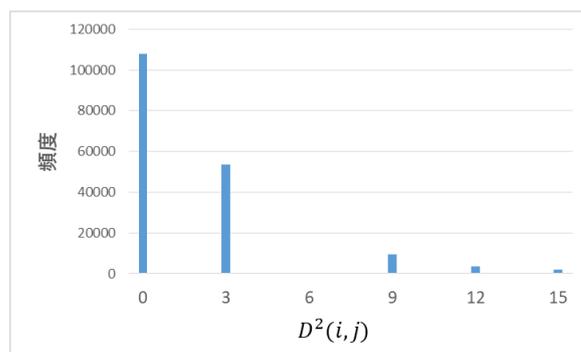


図 2: 2回量子化を行った係数のヒストグラム

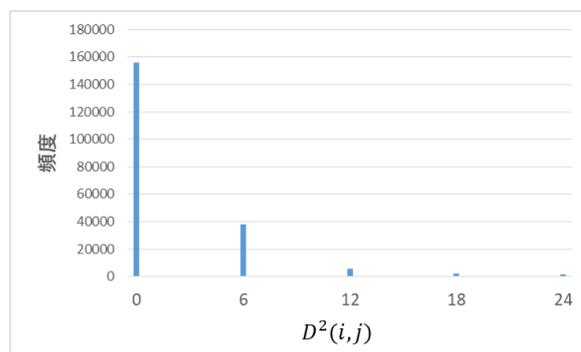


図 3: $Q_{QF}^1(i, j) = 6, Q_{QF}^2(i, j) = 3$ で量子化された係数のヒストグラム

と特定のDCT係数値において、頻度が0となる特徴が生じないことがわかる。この場合では、量子化ステップ6で1回圧縮した画像であると判断されるため、2重圧縮を検知できない。

これらを踏まえ、先行研究 [4] ではヒストグラムをCNNによって学習し、2重圧縮検知を行っている。

3.2 問題点

先行研究では 8×8 ブロック内の左上の9か所(0,1), (1,0), (2,0), (1,1), (0,2), (0,3), (1,2), (2,1), (3,0)の座標において、 -5 から $+5$ までをレンジとするヒストグラムを座標毎に作成し、学習している。JPEGファイルの伸長の際に、IDCT後のYCbCr成分値はRGB成分値に変換され、範囲 $[0, 255]$ の整数値に丸め込まれるため、これによる誤算影響により、従来手法では推測精度が低下する問題がある。さらに、ヒストグラムのレンジが小さすぎる場合、小さい QF において、その間に2重圧縮の特徴が見られず推定精度が低下する問題もある。

4. 提案方式

1回圧縮と2重圧縮をCNNを用いて分類する方式を提案する。以下では、1回目の品質パラメータを QF_1 、2回目の品質パラメータを QF_2 とする。

以下に提案方式1の具体的な処理の流れについて説明

し、図4に示す。

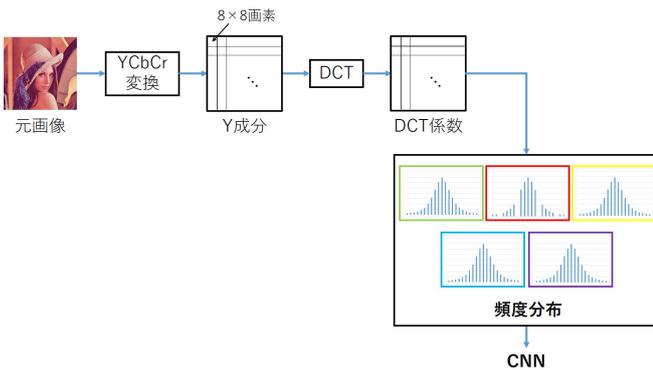


図4: 提案方式の流れ

- (1) フルカラー画像について RGB 成分から YCbCr 成分に変換する。
- (2) Y 成分について 8 × 8 画素単位でブロックに分割し、DCT 係数を求める。
- (3) ブロック内の任意の座標の係数のみを用いてヒストグラムを数種類作成する。この際、表1の量子化テーブル内において値が小さく、複数回出現しているものを選択する。
- (4) 作成したヒストグラムを結合したベクトルを図5に示す CNN モデルに学習させ、分類する。

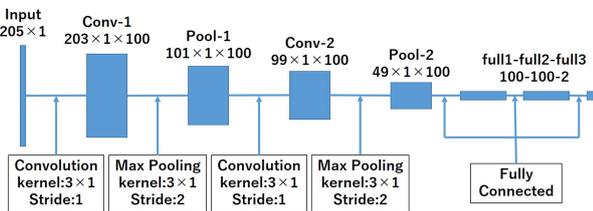


図5: 提案する CNN モデル

提案方式では DCT 係数のヒストグラムを CNN を用いて学習し、1 回圧縮と 2 重圧縮を分類することを目的としている。DCT 係数のヒストグラムを観測する際、JPEG 圧縮時に用いられる量子化テーブルを考慮したヒストグラムを作成することで精度が向上する。

8 × 8 ブロック内のすべての DCT 係数を用いたヒストグラムは、様々なステップによって量子化されているため特徴を抽出することが難しい。そのため、量子化テーブルを参照し、1つのヒストグラム内には同じ値で量子化された DCT 係数のみが存在するように調整する。しかし、ブロック内の各座標におけるヒストグラム全 64 個を作成するのは計算量が多くなる。ここで、表1に示した量子化テーブルを参照すると、いくつかの値が複数回出現することが確認できる。そこで提案方式では、その中から値が比較的小さく、出現回数が多い量子化ステップについて個々にヒス

トグラムを観測する。

ブロック内右下のようにステップが大きい位置の DCT 係数は 0 に量子化されてしまうため有用ではない。出現回数が多い値を選択する理由は、ヒストグラムにおける特徴は、JPEG ファイルが復号された時に行われる丸め処理によって歪められることから、できるだけ多くの有用な統計情報を収集するためである。提案方式では、図6において色付けされた量子化テーブルの値が 12, 14, 16, 22, 24 の5種類の位置に対応するヒストグラムを連結させ、CNN を用いて学習、分類する。

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

図6: 選択係数

提案方式で用いた CNN モデルは、図5に示す通り、8 層からなる。1次元のベクトルを入力とし、畳み込み、最大値プーリングを交互に2回ずつ行ったのち、全結合層2つを経て出力層となる。本モデルの畳み込み層では 100 種類の 3 × 1 フィルタを1行ずつスライドさせ特徴マップを得る。その後、3 × 1 のサイズで2行ずつスライドさせ最大値プーリングを行い、情報を圧縮する。全結合層においては、各層 100 個のノードを持ち、活性化関数として、多くの CNN 構造で精度が向上した結果が得られている elu 関数を用いる [5]。最後に、本方式は 1 回圧縮か 2 重圧縮かの 2 値分類であるため、活性化関数に 2 値分類に適する sigmoid 関数を使用した。

5. シミュレーション

提案方式の精度を検証するため、2種類の計算機シミュレーションを行った。1つは、1回目の圧縮と2回目の圧縮の間に処理を一切行わない場合。もう一方は、実際の画像編集処理を想定し、1回目の圧縮と2回目の圧縮の間にメディアンフィルタを適用した場合である。メディアンフィルタは画像のノイズを除去する手法の一つで、ある画素を周りの画素値の中央値に変換する。例として、メディアンフィルタの窓サイズが 3 × 3 の場合、注目画素の近傍 9 画素において画素値の中央値を選択する。メディアンフィル

タは画像内の画素値において不連続な数値を除去する。そのため、JPEG 圧縮のようにブロック処理を行う圧縮の特徴を除去するため、圧縮履歴の改ざんに用いられる。

以下ではそれぞれのシミュレーションにおける実験条件、実験結果、考察について述べる。

5.1 実験条件

実験で用いる画像は、“Raise Database”^{*1}から取得した。これは、主に画像の改ざん検知アルゴリズムを評価するために開発されたデータセットで、圧縮や加工などの手が加えられていない高解像度の 8156 枚の RAW 画像で構成されており、屋内外の幅広いシーンが含まれている。本研究ではデータセットとして、このうち 1097 枚の非圧縮画像を取得した。それらを、 512×512 画素のブロックに分割し、ImageMagick^{*2}の convert コマンドを用いて 1 回圧縮画像を作成。その後、再度 convert コマンドによって 2 重圧縮画像、メディアンフィルタを適用した 2 重圧縮画像を作成する。その際に使用した、1 回目と 2 回目の QF の組み合わせは、表 2、表 3 に示す。

表 2: 1 回圧縮における品質パラメータ

QF_1				
70	75	80	85	90

表 3: 2 重圧縮における QF_1 と QF_2 の組み合わせ

QF_1, QF_2				
(70, 75)	(70, 80)	(70, 85)	(70, 90)	(70, 95)
(75, 80)	(75, 85)	(75, 90)	(75, 95)	(80, 85)
(80, 90)	(80, 95)	(85, 90)	(85, 95)	(90, 95)

以上によって得られた各 75000 枚の 1 回圧縮画像、2 重圧縮画像、メディアンフィルタを適用した 2 重圧縮画像に対して、4 節に記したようにヒストグラムを作成した。そのうち 1 回圧縮のブロックのヒストグラムと 2 重圧縮のブロックのヒストグラムをそれぞれランダムに 30000 個分を選出し、計 60000 個のヒストグラムを用いて CNN の入力とする。その際、トレーニング用データと検証用データの比率は 8 : 2 とした。

5.1.1 実験結果

原画像、1 回圧縮画像、2 回圧縮画像、メディアンフィルタを適用した画像の例を図 7 に示す。図 8 から図 11 は、図 7 を例に示すデータセットの分類におけるトレーニング結果と検証結果を示す。Accuracy は正確性、Loss は正解とどれほど離れているかを表す。表 4、表 5 に最終分類結果を示す。

*1 <http://loki.disi.unitn.it/RAISE/>

*2 <https://imagemagick.org>



図 7: 512×512 画素のサンプル画像

表 4: 1 回圧縮画像と単なる 2 重圧縮画像の最終分類結果

	Final Accuracy	Final Loss
training	0.9850	0.0312
validation	0.9833	0.0365

表 5: 1 回圧縮画像とメディアンフィルタを適用した 2 重圧縮画像の最終分類結果

	Final Accuracy	Final Loss
training	0.9845	0.0365
validation	0.9872	0.0325

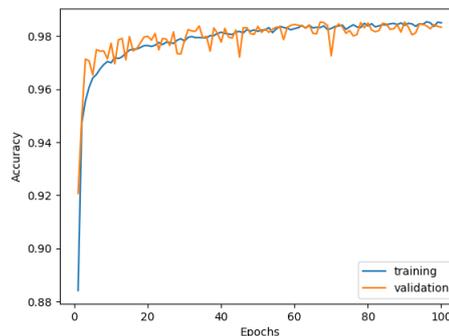


図 8: 1 回圧縮画像と単なる 2 重圧縮画像の分類結果 (Accuracy)

5.2 考察

シミュレーション結果より、提案方式において、1 回目の QF が 2 回目より小さい場合に 2 重圧縮を高い精度で検知することができた。CNN を用いて分類することで、ヒトの視覚情報より細かい特徴を抽出することができると考

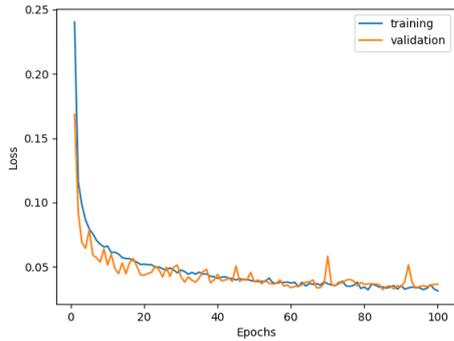


図 9: 1 回圧縮画像と単なる 2 重圧縮画像の分類結果 (Loss)

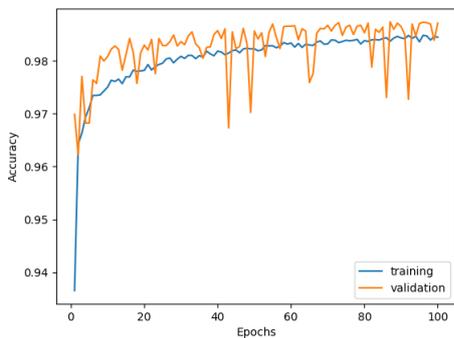


図 10: 1 回圧縮画像とメディアンフィルタを適用した 2 重圧縮画像の分類結果 (Accuracy)

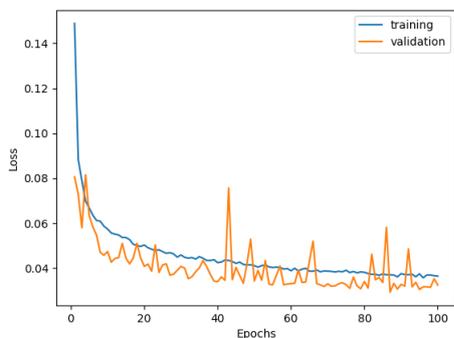


図 11: 1 回圧縮画像とメディアンフィルタを適用した 2 重圧縮画像の分類結果 (Loss)

えられる。また、メディアンフィルタを適用した場合には、編集処理を施さないものと比較すると、分類精度に関しては同等の結果が得られたが、Epoch 毎に精度のばらつきが見られたが、これは、トレーニングデータに対して、検証データ量が少量であったためだと考えられる。

6. むすび

本論文では JPEG 画像に対する改ざん検知方式を提案した。JPEG 画像の DCT 係数のヒストグラムを観測し、そのヒストグラム内のいくつかの DCT 係数においてヒスト

グラムに偏りが見られ、その特徴を CNN を用いて学習することで 2 重 JPEG 圧縮が行われているかを判断した。提案方式の精度検証結果、2 回目の圧縮時における QF が 1 回目の圧縮時より小さい場合には、高い精度で 2 重圧縮を検知でき、また、1 回目と 2 回目の圧縮の間にメディアンフィルタが適用されている場合にも高い精度で検知した。

今後の課題は、2 回目の圧縮時における QF が 1 回目の圧縮時より大きい場合においても、2 重圧縮を分類できるか検証する。また、メディアンフィルタ以外の編集処理にどの程度耐性があるかの検証が挙げられる。

謝辞 本研究は JSPS 科研費 JP19K22846 の助成を受けたものである。

参考文献

- [1] 小川文伸, 小野文孝, “静止画像国際標準符号化 (JPEG), ” *Medical Imaging Technology*, vol.14, no.3, pp.231–236, 1996.
- [2] 遠藤俊明, “カラー静止画像の国際標準符号化方式-JPEG アルゴリズム-, ” *テレビジョン学会誌*, vol.46, no.8, pp.1021–1024, 1992.
- [3] J. Lukas, J. Fridrich, “Estimation of Primary Quantization Matrix in Double Compressed JPEG Images, ” *Proc. Digit. Forensic Res. Workshop (DFRWS)*, pp. 5–8, 2003.
- [4] Q. Wang, R. Zhang, “Double JPEG compression forensics based on a convolutional neural network, ” *EURASIP Journal on Information Security 2016 (2016)* 23.
- [5] D.-A. Clevert, T. Unterthiner, and S. Hochreiter. “Fast and accurate deep network learning by exponential linear units (elus)”. *arXiv preprint arXiv:1511.07289*, 2015.
- [6] 山下 隆義 (2016), 「イラストで学ぶ ディープラーニング」 講談社.
- [7] Multimedia Signal Processing and Understanding Lab, (n.d), Introducing RAISE dataset, Retrieved November 1 2018, <http://loki.disi.unitn.it/RAISE/>