CH-013

畳み込みニューラルネットワークを用いた改ざん JPEG 画像の検出
Detecting doctored JPEG image using Convolutional Neural Network

多谷邦彦 * 黒木修隆 † 竹田直人 ‡ 宿院康昭 ‡ 小林正 *
Kunihiko Taya **Nobutaka Kuroki** **Naoto Takeda** **Yasuaki Shukuin** **Tadashi Kobayashi**

1 はじめに

デジタルカメラやカメラ付きスマートフォンの普及により多くの人が日常的にデジタル写真を撮影しており、それと共に誰でも簡単に画像処理を行えるアプリケーションも数多く流通している。これらのアプリケーションは購入したデジタルカメラに付属されていたり、スマートフォンや PC にあらかじめインストールされていることも多く、画像処理に関する専門的な技術や機器を持っていなくても、簡単に画像処理を行える環境となっている。デジタル写真は被写体をありのまま保存するため、犯罪捜査や事故状況の確認において客観証拠としての価値が非常に高い。そのため、デジタル写真に記録された内容を真実をありのまま捉えたものとして正しく活用するためには、画像に何者かによる悪意ある作為的行為、すなわち改ざんが加えられていないかを確認する作業が必要である。

改ざん手法にはその目的に応じていくつかの手法が考えられ、様々な検出方法に関する研究が報告されている [1]。画像内のコピー&ペーストの検出 [2, 3]、カメラのカラーフィルタ配列の不連続性に着目した検出 [4]、ノイズ特性の推定による検出 [5] などがある。デジタルカメラで撮影された画像の多くは JPEG 形式であることから、著者らは JPEG 画像に対象を限定して改ざん検出の研究を行ってきた。JPEG 画像を対象とすることは実用的であり、著者らと同じく JPEG 画像を対象とした改ざん検出手法としては J.He[6]、T.Pevny[7]、T.Bianchi[8] の報告がある。特に J.He らの手法は、改ざん領域の形状や画素数に限定がなく、改ざんの有無を判定するだけでなく改ざん領域の自動特定まで行っていることから、汎用的で有効な手法であると言える。著者らは、改ざん部にはブロックノイズによる影響が生じるため、DCT 係数の高周波成分に注目することなどにより、J.He らの手法と比較してより高い精度で改ざん画像を検出する手法 (MDBD 法) について既に報告した [9]。同手法は DCT 係数の 64 成分のうち 19 成分を用いているがその成分の選定は経験的なものであったため、選定する成分をさらに吟味することで検出率が向上する可能性が期待されていた。

本研究では、DCT 係数の 64 成分すべてを入力信号とし、畳み込みニューラルネットワーク (CNN) を用いて膨大なパラメータ計算を行わせることで、従来手法よりも高い検出精度を得ることができたので報告する。

2 従来手法の問題点

本章では、J.He らの手法と著者らが報告した手法について、その概要と問題点を述べる。

2.1 J.He らの手法 (従来手法 1)

J.He らは、Double JPEG に注目した改ざん検出手法を提案した。JPEG 画像はその国際標準規格 [10] により 8×8 画素のブロックに区分けして離散コサイン変換 (DCT) されたのち、量子化テーブルにより量子化される。その後、画像編集ソフトにより JPEG 画像を展開し、改ざん処理等を行った後に JPEG 画像として保存した場合、画像編集ソフトが保有する量子化テーブルにより量子化が行われる。この 2 つの量子化テーブルが異なる場合、量子化された DCT 係数のヒストグラムは図 1(a) のように櫛状となる。一方、原画像のヒストグラムは図 1(b) となる。これが、Double JPEG と呼ばれる現象である。

ある画像の一部を切り取り、別の JPEG 画像に貼り付けた場合、非改ざん領域は同じ 8×8 画素メンバーで 2 度量子化されるため Double JPEG が生じるが、改ざん領域については原画像とは異なる 8×8 画素メンバーで DCT されたのち量子化が行われるため、Double JPEG が生じない。

J.He らは図 1(a) のヒストグラムから各 8×8 画素ブ

* 京都府警察科学捜査研究所
† 神戸大学大学院工学研究科
‡ (株) 扶桑プレシジョン

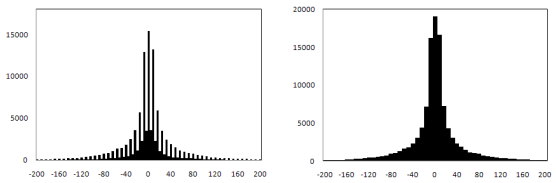
ロックが改ざん領域である確率を導出し、その確率分布から画像内を改ざん領域と非改ざん領域に分類し、最後にサポートベクトルマシン (SVM) により改ざん画像か非改ざん画像かの識別を行った。

J.He らの手法は SVM により改ざん画像か非改ざん画像かについては高精度で判定することができるが、改ざん領域を示した結果画像の精度は低く、図 2(c) のように多くの非改ざん領域についても改ざん領域として誤って表示することに課題が残っていた。改ざん画像として正しく判定できていたとしても改ざん領域の表示が不正確であれば、その判定結果の真偽について疑われるおそれがある。

2.2 MDBD 法 (従来手法 2)

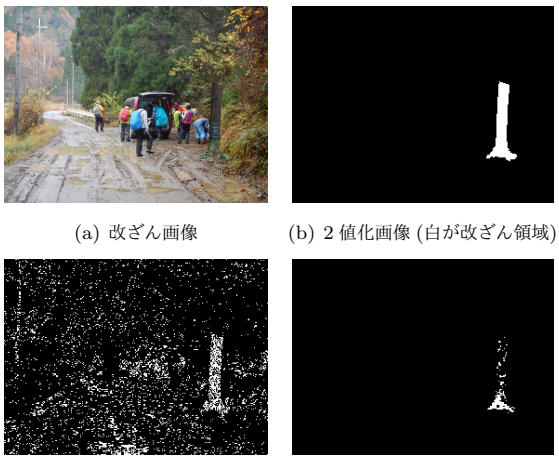
著者らは、J.He らの手法の課題を解決するため、ブロックノイズ解析と Double JPEG 解析を併用する MDBD 法を用いることで、改ざん検出性能の向上を実現した。

ある画像の一部を切り取り、別の JPEG 画像に貼り付けた場合、改ざん領域と背景画像のブロックノイズの境界は多くの場合一致しない。そのため、改ざん領域に



(a) 改ざん画像　(b) 原画像

図 1　改ざん画像と原画像の DCT 係数のヒストグラム



(a) 改ざん画像　(b) 2 値化画像 (白が改ざん領域)

(c) J.He らの手法 (従来手法 1) による検出結果　(d) MDBD 法 (従来手法 2) による検出結果

図 2　従来手法による検出結果

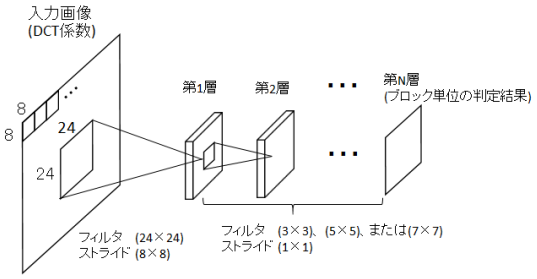


図 3　提案手法のネットワーク構造

はブロックノイズのずれに由来する高周波成分が存在する。各ブロックの DCT 係数 64 成分のうち、19 成分を抽出して高周波成分解析を行うと共に、ブロックノイズ自体の位置を推定し、これらを基に改ざんの疑わしさを定量的に表した。さらに Double JPEG の影響についても数値で評価し、これらを SVM に入力して改ざん画像と非改ざん画像の識別を行った。

その結果、J.He らの手法よりも識別結果が高く、さらに改ざん領域を示した結果画像についても精度が向上した (図 2(d))。J.He らの手法と比べて非改ざん領域の誤検出については大幅に削減できているが、その反面、改ざん領域の検出漏れが生じたため、依然として検出結果画像の精度に課題が残っていた。

3　提案手法

提案手法は、改ざん画像の DCT 係数を画像化し、これを入力画像として CNN を用いることで改ざん領域を検出するものである。図 3 に提案する CNN の構造を示す。一般に CNN はフィルタサイズが大きくなる程、入力の広い範囲を出力に反映できるが、それに応じて畳み込み処理後の出力サイズが小さくなる。そこで、まず予備実験 1 で入力の参照範囲を決める実験を行い、次に予備実験 2 ではその参照範囲を満たす 45 種類のネットワーク構造について比較検討を行った。

3.1 実験条件

実験には一般的な家庭用デジタルカメラ 5 機種 (CASIO:EX-ZR200, SONY:NEX-3N, Canon:IXY510, OLYMPUS:FE220, NIKON:AW210) で撮影した 50 枚の JPEG 画像を使用した。撮影画像の被写体は自然風景、人物、建造物など様々である。ある画像の一部を切り取り、同機種で撮影した他の画像に貼り付け、JPEG 画像で保存することで改ざん画像を作成した。使用したアプリケーションは PhotoshopCS であり、画像保存時

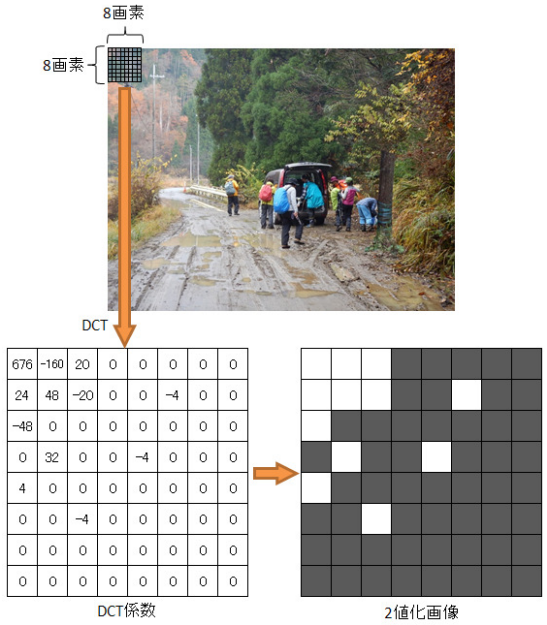


図 4　2 値化画像作成手順

の設定は最高画質 (最低圧縮) である。

入力画像は、8×8 ブロックの DCT 係数 64 成分それぞれについて 0 か非 0 かに応じて 0 又は 255 に 2 値化した画像を使用し、この 2 値化画像を元の 8×8 画素と対応させる (図 4)。画像内の全てのブロックにおいてこの 2 値化画像を作成するため、原画像と同じサイズの 2 値化画像が作成されることとなる。教師画像には、図 2(b) のように改ざん領域を 255(白)、非改ざん領域を黒 (0) とした 2 値化画像を使用する。しかし、原画像はデジタルカメラで撮影した 4M〜16M 画素の非常に大きい画像であるため、予備実験 1 では各画像を 256×256 画素に分割して入力画像としている。従って、改ざん画像 50 枚から得られる入力画像は 6489 画像である。作製した改ざん画像 50 枚 (6489 画像) のうち、40 枚 (5635 画像) を学習画像、5 枚 (427 画像) をバリデーション画像、5 枚 (427 画像) をテスト画像とする。また、学習時のエポック数を 100 回とし、テスト画像を用いた推定には、最もバリデーション精度の高かったモデルを用いる。

著者らの先行研究により、改ざん JPEG 画像の検出には 8×8 画素ブロック単位での解析が有効であること、さらに改ざんの疑いがあるブロックの上下左右の連結性が重要であることが明らかになっていた。そこで、図 3 のように 1 層目のフィルタサイズを 24、ストライドを 8 とすることで、連結性を考慮したブロック単位の解析を行う。ストライドを 8 とするため教師画像は入力画像の 8 分の 1 のサイズである 32×32 画素の画像が必要となる。よって、2 値化画像を入力画像に合わせて 256×256 画素で分割し、さらにバイキュービック法で 32×32 画素に縮小した画像を教師画像として用いる。

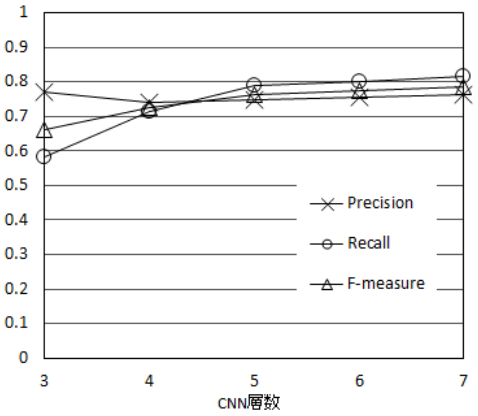


図 5　層数による精度比較

実験に用いた計算機器は、CPU:Intel Xeon W3670、GPU:GeForce GTX1060 6GB、フレームワーク:Theano(Keras) である [11]。

3.1.1　予備実験 1　入力の参照範囲の決定

次のネットワーク構造 (フィルタサイズ f, マップ数 n, ストライド s, 全層ゼロパディング) を用いて層数を変えながら検出精度を計測した。

1 層目　f=24, n=32, s=8
2 層目〜6 層　f=3, n=32or16, s=1
最終層　f=3, n=1, s=1

層数 N に対する結果を図 5 に示す。ここで、P は適合率 (Precision)、R は再現率 (Recall)、F は P と R の調和平均 (F-measure) であり、次式で定義する。また、T_P は検出した改ざん画素数、F_N は検出できなかった改ざん画素数、F_P は改ざん画素として誤検出した非改ざん画素数である。

$$P = \frac{T_P}{T_P + F_P} \tag{1}$$

$$R = \frac{T_P}{T_P + F_N} \tag{2}$$

$$F = \frac{2 \cdot P \cdot R}{P + R} \tag{3}$$

概ね層数が 6、つまりフィルタサイズが (24,3,3,3,3,3) のところで F が高止まりしていることから、ネットワーク構造は最大で 6 層とし、出力 1 画素あたりの入力の参照範囲を 104×104 とすることとした。

3.1.2 予備実験 2　ネットワーク構造の決定

入力の参照範囲が 104×104 となる条件の下で、マップ数とフィルタ数を変化させ、ネットワーク構造の比較検討を行った。予備実験 1 では全ての層においてゼロパディングをしたが、その結果、出力画像の境界が不連続な値となった。そこで、予備実験 2 ではゼロパディングを行わないこととした。その場合、256×256 の画像に対しては参照領域が上下左右で 48 画素ずつはみ出すため、入力画像のサイズを 352×352 画素に変更した。教師画像は予備実験 1 と同じ 32×32 画素である。

マップ数 n(16, 32, 64, 128)、フィルタサイズ f(3, 5, 7)、ネットワーク層数 (3, 4, 5, 6) を組み合わせた合計 45 通りのネットワーク構造を作成した。なお、1 層目は (f=24, s=8) に、最終層は n=1 に固定する。学習画像、バリデーション画像、テスト画像の組み合わせは、すべて予備実験 1 と同じである。表 1 に各ネットワークの推定精度を示す。ただし、改ざんブロックを全く検出しない場合、$T_P = 0$、F_P=0 となるため、– と記入している。表中の 5 層のネットワークの中に、P=0.95、R=0.81、F=0.86 に達したものがあった。よって、このネットワーク構造、すなわち (f,n) が 1 層目から順に (128, 24),(128, 5),(64, 3),(32, 3),(1, 3) の 5 層を用いて、次節で従来手法と比較する。

3.2 従来手法との比較

50 画像を 10 グループに分割し、前節で導出したネットワークを用いてクロスバリデーション行った。すべての組み合わせにおいて、40 画像が学習、5 画像がバリデーション、5 画像がテストである。ここで、学習を安定化させるため各層間に 0.5 の Dropout 層を設けた。各グループにおいてエポック数は 100 回であり、トレーニング画像数は 4821〜5635 であるので、バックプロパゲーションは約 48 万〜56 万回となる (図 6)。各グループごとにバリデーション精度が最大となるモデルを用いてテストを行った。本手法及び各従来手法の推定結果を表 2 に、それらの平均値を図 7 に、また検出例を図 8 に示す。

提案手法では P と R の両者が高いことから、非改ざん部の誤検出を抑えながらも改ざん部の検出が向上していることがわかる。そのことは図 8 からも見て取れ、従来手法 1、2 と比較して良好な検出結果画像を出力している (No.10,16,32,48)。ただし、50 画像のうち 2 画像において、改ざん部を検出しない場合があった (No.18,36)。No18 は空の一部を切り抜いて塔の左側を隠したものである。従来手法 2 でも検出できておらず、改ざん部が色変化の少ないテクスチャであるため周波数成分に改ざんの特徴がでていなかったと考えられる。No36 は右上に花一輪を付け足したものである。従来手法 2 では改ざん部の一部を検出できていたが、提案手法では検出でき

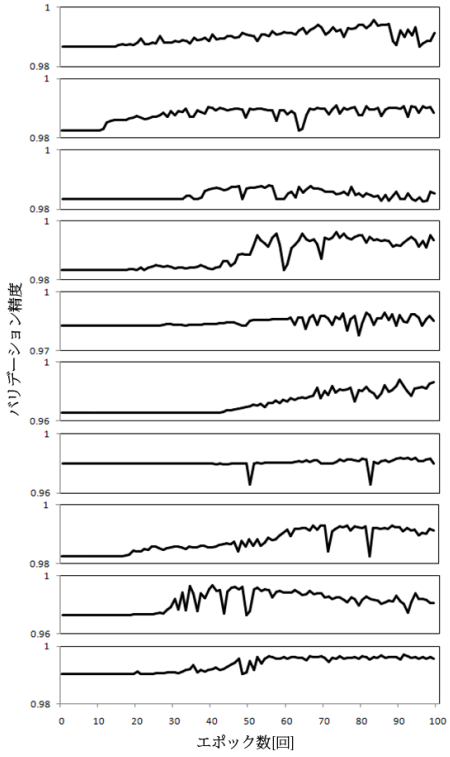


図 6　10 グループのエポック回数に対するバリデーション精度（下から順にグループ 1,2,…,10）

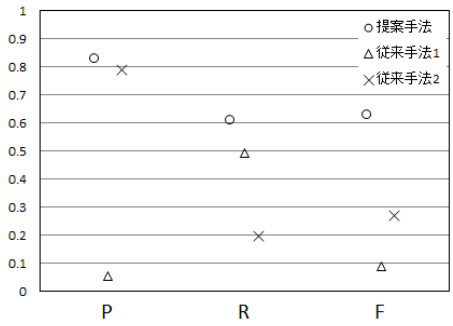


図 7　提案手法と従来手法の比較 (50 画像の平均値)

なかった。この理由についてはさらに検証を行う必要がある。

4 まとめ

本研究では、畳込ニューラルネットワークを用いた改ざん JPEG 画像の検出を行った。J.He らの手法 (従来手法 1) では改ざん部の検出漏れが少ないものの非改ざん部の誤検出が多く、逆に MDBD 法 (従来手法 2) では非改ざん部の誤検出を減らすことができたが改ざん部の検出漏れが多かった。提案手法では、非改ざん部の検出を抑えつつ改ざん部の検出を大幅に向上させ、F 値が J.He らの手法の 7.1 倍、MDBD 法の 2.3 倍を達成した。今後はさらに多くの入力画像に対してネットワーク構造の最適化を行い、改ざん検出率の向上を目指す。

参考文献

[1] H. Farid, "Digital Image Forensics", Dartmouth College, http://www.cs.dartmouth.edu/farid/, 参照 July, 2013.

[2] A.Kaur, R. Sharma, "Copy-Move Forgery Detection using DCT and SIFT", International Journal of Computer Applications, vol.70-no.7, pp.30-34, May 2013.

[3] Y. Huang, W. Lu, W. Sun, D. Long, "Improved DCT-based detection of copy-move forgery in images", Forensic Science International, vol206, pp.178-184, 2011.

[4] MK. Johnson, H. Farid, "Exposing Digital Forgeries by Detecting Inconsistencies in Lighting", MM&Sec 2005 Proceedings of the 7th workshop on Multimedia and security, pp.1-10, New York, USA, 2005.

[5] M.Kobayashi, T. Okabe, Y. Sato, "Detecting Forgery From Static-Scene Video Based on Inconsistency in Noise Level Functions", IEEE Trans. Information Forensics and Security, vol.5, pp.883-892, Sept. 2010, DOI:10.1109/TIFS.2010.2074194

[6] J. He, Z. Lin, L. Wang, X. Tang, "Detecting Doctored JPEG Images Via DCT Coefficient Analysis", Computer Vision-ECCV 2006, vol.3953 pp. 423-435, 2006

[7] T. Pevny, J. Fridrich, "Detection of double-compression in JPEG images for application in steganography", IEEE Trans. Information Forensics and Security, vol.3 pp. 247-258, June, 2008

[8] T. Bianchi, A. Piva, "Detection of Non-Aligned Double JPEG Compression Based on Integer Periodicity Maps", IEEE Trans. Information Forensics and Security, vol.7 pp. 842-848, April, 2012

[9] 多谷邦彦, 竹田直人, 小林正, 尾崎吉明, 黒木修隆, " ブロックノイズ解析と Double JPEG 解析に基づく改ざん JPEG 画像の検出", 電気学会論文誌, Vol.137 No.5, 2017

[10] ISO/IEC 10918:Digital compression and coding of continuous - tone still images

[11] 藤田一弥、高原歩, "実装ディープラーニング", オーム社, 2016

表 1　ネットワークの組み合わせと推定精度

層	filter size	map	P	R	F
6	24,3,3,3,3,3	16,16,16,16,16,1	0.80	0.84	0.82
		32,32,32,32,32,1	0.76	0.85	0.80
		64,64,64,64,64,1	0.73	0.90	0.79
		128,128,128,128,128,1	-	0	-
		32,32,32,16,16,1	0.71	0.89	0.77
		64,64,32,32,32,1	0.59	0.91	0.70
		128,128,128,64,64,1	0.77	0.84	0.80
		64,64,32,32,16,1	0.83	0.81	0.82
		128,128,64,64,32,1	0.93	0.80	0.84
		128,128,64,32,16,1	0.93	0.80	0.85
5	24,5,3,3,3	16,16,16,16,1	0.79	0.82	0.80
		32,32,32,32,1	0.85	0.81	0.82
		64,64,64,64,1	0.91	0.73	0.81
		128,128,128,128,1	0.82	0.81	0.81
		32,32,16,16,1	0.82	0.78	0.80
		64,64,32,32,1	0.72	0.89	0.79
		128,128,64,64,1	0.91	0.34	0.43
		64,64,32,16,1	0.70	0.89	0.77
		128,128,64,32,1	0.95	0.81	0.86
		128,64,32,16,1	0.87	0.72	0.79
4	24,5,5,3	16,16,16,1	-	0	-
		32,32,32,1	0.87	0.81	0.83
		64,64,64,1	0.84	0.81	0.82
		128,128,128,1	-	0	-
		32,32,16,1	0.89	0.79	0.83
		64,64,32,1	0.84	0.79	0.80
		128,128,64,1	0.87	0.73	0.79
		64,32,16,1	0.83	0.84	0.83
		128,64,32,1	0.89	0.79	0.82
4	24,7,3,3	16,16,16,1	0.85	0.74	0.79
		32,32,32,1	0.80	0.84	0.81
		64,64,64,1	-	0	-
		128,128,128,1	-	0.31	-
		32,32,16,1	0.83	0.77	0.80
		64,64,32,1	0.82	0.85	0.83
		128,128,64,1	0.84	0.71	0.77
		64,32,16,1	0.85	0.85	0.84
		128,64,32,1	0.86	0.81	0.83
3	24,7,5	16,16,1	0.81	0.68	0.72
		32,32,1	0.84	0.75	0.79
		64,64,1	0.89	0.67	0.76
		128,128,1	0.66	0.78	0.70
		32,16,1	0.77	0.85	0.80
		64,32,1	0.59	0.71	0.63
		128,64,1	0.87	0.72	0.78

表 2　提案手法と従来手法の比較

	提案手法			J.He らの手法 (従来手法 1)			MDBD 法 (従来手法 2)		
No	P	R	F	P	R	F	P	R	F
1	0.97	0.86	0.91	0.14	0.49	0.21	0.99	0.02	0.04
2	0.93	0.90	0.92	0.04	0.49	0.07	0.99	0.05	0.09
3	0.95	0.39	0.55	0.03	0.49	0.06	0.92	0.01	0.02
4	1.00	0.16	0.27	0.10	0.49	0.17	0.96	0.04	0.09
5	0.39	0.95	0.55	0.08	0.34	0.13	0.55	0.21	0.30
6	0.99	0.74	0.84	0.05	0.49	0.08	0.83	0.03	0.05
7	0.81	0.12	0.21	0.32	0.47	0.38	0.97	0.02	0.03
8	0.79	1.00	0.88	0.15	0.56	0.24	1.00	0.89	0.94
9	0.97	0.52	0.68	0.20	0.50	0.29	1.00	0.21	0.34
10	0.96	0.91	0.93	0.11	0.57	0.19	1.00	0.27	0.42
11	1.00	0.69	0.81	0.05	0.50	0.08	1.00	0.17	0.28
12	0.23	0.17	0.20	0.04	0.47	0.07	0.23	0.01	0.01
13	0.98	0.56	0.71	0.05	0.48	0.08	0.98	0.04	0.08
14	0.99	0.63	0.77	0.09	0.49	0.15	0.99	0.30	0.46
15	0.61	0.99	0.76	0.00	0.47	0.00	0.86	0.20	0.33
16	1.00	0.88	0.94	0.07	0.52	0.13	0.82	0.01	0.02
17	0.97	0.88	0.92	0.01	0.42	0.01	0.99	0.21	0.34
18	-	0.00	-	0.01	0.22	0.03	-	0.00	-
19	1.00	0.03	0.05	0.01	0.50	0.03	0.94	0.11	0.20
20	1.00	0.32	0.48	0.01	0.48	0.01	0.77	0.16	0.26
21	1.00	0.69	0.90	0.03	0.50	0.06	1.00	0.55	0.71
22	0.94	0.98	0.96	0.07	0.49	0.13	0.99	0.46	0.63
23	0.62	0.54	0.58	0.06	0.50	0.11	0.48	0.07	0.13
24	1.00	0.01	0.02	0.04	0.51	0.08	1.00	0.00	0.00
25	0.87	1.00	0.93	0.03	0.49	0.05	0.97	0.53	0.68
26	1.00	0.58	0.74	0.04	0.52	0.07	0.91	0.67	0.77
27	0.51	0.49	0.50	0.02	0.49	0.03	0.73	0.06	0.12
28	0.99	0.90	0.94	0.02	0.50	0.05	0.98	0.44	0.60
29	1.00	0.03	0.06	0.01	0.50	0.03	0.48	0.02	0.04
30	0.98	0.86	0.92	0.02	0.50	0.03	0.17	0.03	0.05
31	0.97	0.89	0.93	0.05	0.47	0.10	0.54	0.04	0.08
32	0.93	0.93	0.93	0.08	0.74	0.15	1.00	0.01	0.02
33	1.00	0.00	0.00	0.07	0.50	0.12	0.73	0.06	0.11
34	0.96	0.96	0.96	0.08	0.51	0.14	0.95	0.43	0.60
35	0.77	0.97	0.86	0.02	0.46	0.05	0.78	0.17	0.29
36	-	0.00	-	0.01	0.49	0.02	0.69	0.02	0.05
37	0.38	0.95	0.54	0.02	0.53	0.04	1.00	0.54	0.70
38	0.99	0.85	0.92	0.03	0.53	0.06	1.00	0.66	0.80
39	1.00	0.57	0.73	0.01	0.56	0.01	0.75	0.29	0.42
40	0.85	0.75	0.80	0.05	0.44	0.08	0.98	0.18	0.30
41	0.98	0.87	0.92	0.03	0.48	0.05	0.96	0.07	0.13
42	0.67	0.97	0.79	0.04	0.49	0.07	0.88	0.15	0.26
43	0.17	0.27	0.21	0.01	0.49	0.03	0.04	0.07	0.05
44	1.00	0.19	0.31	0.05	0.51	0.10	0.47	0.00	0.01
45	0.88	0.60	0.71	0.08	0.48	0.13	0.57	0.05	0.09
46	1.00	0.05	0.09	0.04	0.50	0.08	0.67	0.12	0.21
47	0.47	0.98	0.63	0.02	0.50	0.05	0.66	0.60	0.63
48	0.95	0.36	0.52	0.02	0.49	0.04	0.76	0.05	0.09
49	0.98	0.56	0.71	0.01	0.49	0.02	0.99	0.39	0.56
50	0.95	0.89	0.92	0.01	0.49	0.02	0.45	0.03	0.05

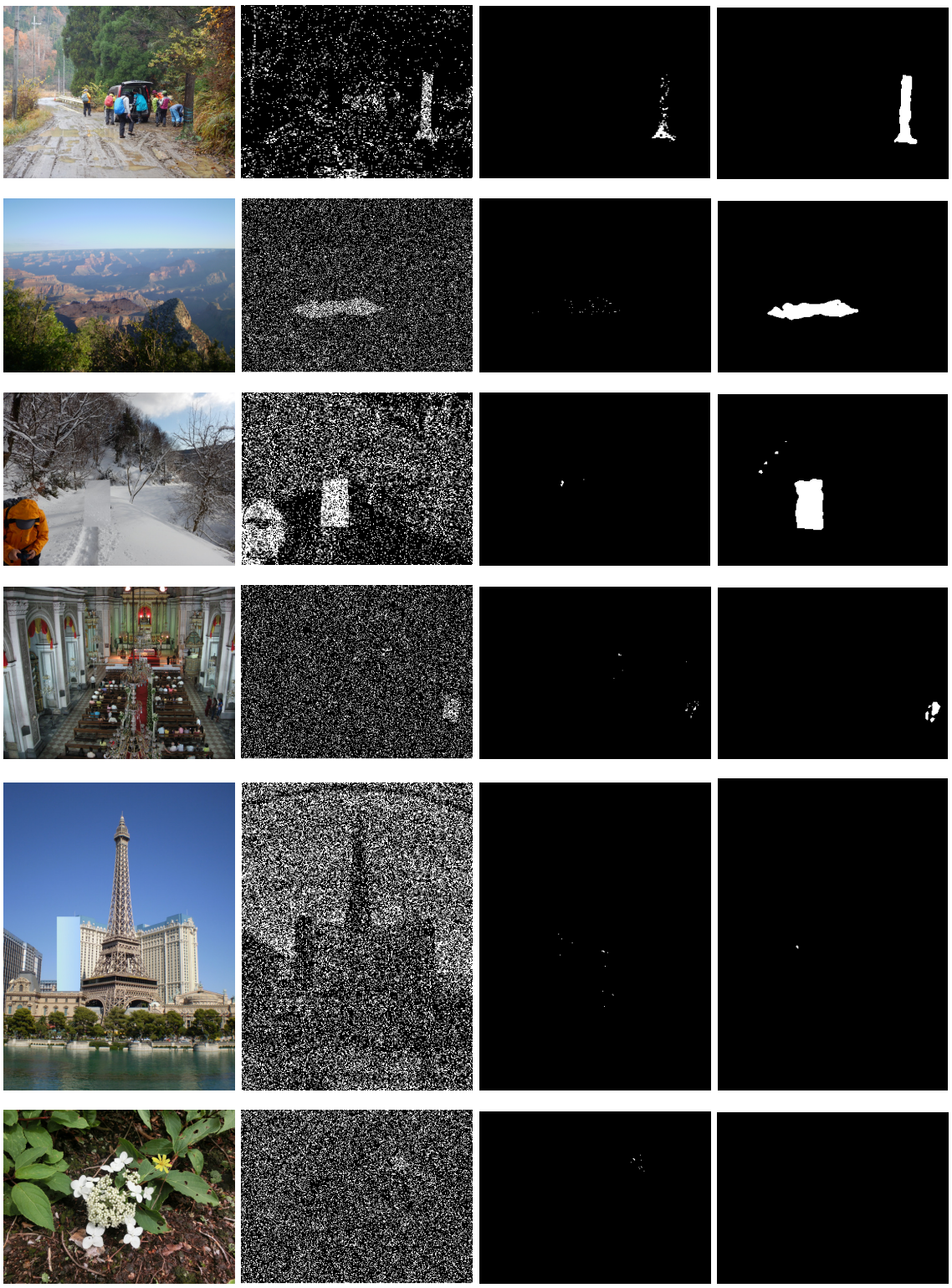


図 8　従来手法と提案手法による検出結果（上段から No10,16,32,48,18,36。左列から、改ざん画像、J.He らの手法 (従来手法 1)、MDBD 法 (従来手法 2)、提案手法の検出結果）