Kobe University Repository : Kernel

PDF issue: 2026-11-21

法科学における画像解析技術の高度化に関する研究

多谷，邦彦

(Degree)
博士（工学）
(Date of Degree)
2020-09-25
(Date of Publication)
2021-09-01
(Resource Type)
doctoral thesis
(Report Number)
甲第7879号
(URL)
https://hdl.handle.net/20.500.14094/D1007879

※ 当コンテンツは神戸大学の学術成果です。無断複製・不正使用等を禁じます。著作権法で認められている範囲内で、適切にご利用ください。

博 士 論 文

法科学における画像解析技術の高度化に関する研究

2020 年 7 月

神戸大学大学院工学研究科

多谷邦彦

内容梗概

近年、ディジタルカメラやスマートフォンの普及と使い方の手軽さにより、多くの人が日常的にディジタル写真を撮影している。また、防犯意識の高まりにより、商業施設や公共機関など街中のあらゆる場所に防犯カメラや監視カメラが設置されており、これらは 24 時間 365 日稼働して付近の様子を記録している。事故や事件が発生した場合、当事者がすぐにディジタルカメラでその状況を撮影することができるし、近くに防犯カメラが設置されていれば当時の状況が記録されていることもある。ディジタル写真や映像は、犯罪捜査や事故状況の確認などの警察活動において極めて有用な資料であり、いつ、どこで、何があったのかを正確に把握するための重要な手段の一つとなる。

ディジタル画像資料をありのままの真実を捉えたものとして正しく活用するためには、その画像に悪意ある何者かの加工、すなわち改ざんが施されていないかを検証することが必要不可欠である。このような考えは法科学（フォレンジック・サイエンス、Forensic Science）において重要な意味を持つ。特に、ディジタル画像資料を扱う分野はディジタル・フォレンジックと呼ばれており、国内外を問わず研究が活発に行われている。

本論文は、改ざんが加えられたディジタル静止画やディジタル動画の検出に焦点を当て、「法科学における画像解析技術の高度化に関する研究」としてまとめたものである。

第 2 章では、ブロックノイズ解析と Double JPEG 解析の 2 つの側面から改ざん画像の解析を行い、改ざん画像と非改ざん画像を判別する手法についての研究をまとめている。ブロックノイズ解析はさらに DCT 高周波成分解析法とブロックノイズ位置推定法の 2 つの解析法に分かれる。ブロックノイズ解析と Double JPEG 解析から得られるパラメータを定量的に算出して SVM に入力することで改ざんの有無を総合的に判別し、さらに改ざん領域を可視化する。提案手法では He らの従来手法よりも誤検出を抑えることで全体の検出精度を高めることに成功し、改ざん領域を視覚的に示すことにより本手法の信頼性を高めることを実現している。

第 3 章では、畳み込みニューラルネットワーク（Convolutional Nueral Networks, CNN）を用いることで、改ざんされた JPEG 画像における改ざん領域を検出する手法についての研究をまとめている。第 2 章での研究成果から、JPEG 画像における改ざん検出には、DCT 係数の解析が有効であること、また、改ざんが疑われる画素同士の空間的連結性を考慮する必要があることが明らかとなった。そこで、DCT 係数を二次元画像として表示して CNN への入力画像とし、出力結果として改ざん領域が白、非改ざん領域が黒の 2 値化画像を出力する改ざん領域検出手法を提案する。本研究ではネットワークの最適化を行うために 2 つの予備実験を行う。予備実験 1 では n 層（$n = 3$〜6）の CNN を用いて入力の参照範囲を決定し、次に予備実験 2 で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う。最後に、予備実験 2 で決定したネットワークを用いた提案手法と従来手法 (He らの手法、MDBD 法) について、改ざん領域の検出

結果を比較する。提案手法では、F 値が He らの手法の 7.1 倍、MDBD 法の 2.3 倍を達成しており、改ざん領域の検出精度を飛躍的に向上させることに成功している。

第 4 章では、あるシーンに写っている被写体が改ざん処理により消去された場合において、改ざん領域を検出する手法についての研究をまとめている。動画が静止画の時間的連続体であることに着目し、再帰型ニューラルネットワーク（Recurrent Neural Network, RNN）と CNN を併用して時間と空間の両面から解析を行う。時間方向には RNN の中で使われる機構の一つである LSTM(Long short-term memory) を使用し、検査対象画素に前後 15 フレームを含めた 31 個のデータを利用して解析を行う。空間方向には U-Net 構造の CNN を使用し、改ざん領域の全体的な位置情報と局所的特徴を合わせた解析を行う。評価実験として 4 通りの検出手法を比較する。1. 各フレームをそのまま入力画像とする、2. 絶対値差分画像を用いる、3. 絶対値差分画像を LSTM の入力画像とする、4.LSTM の出力結果を入力画像として U-Net 解析を行う (提案手法) である。提案手法では、改ざん領域が画像内を空間方向に移動・変形した動画であっても F 値 0.98 を達成しており、動画改ざん領域の検出手法として有用であることを示している。

第 5 章は、まとめとして本研究の結論と今後の展望について述べる。

目次

1 緒論 1

2 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 5

2.1 概要 . . . 5
2.2 背景 . . . 5
2.3 先行研究 . . . 6
2.4 先行研究の問題点 . . . 6
2.5 Double JPEG について . . . 7
2.6 He らの従来手法 . . . 7
2.7 研究の目的 . . . 10
2.8 提案手法の流れ . . . 11
2.8.1 ブロックノイズ解析 . . . 11
2.8.2 Double JPEG 解析法 . . . 16
2.9 実験方法 . . . 18
2.10 実験結果 . . . 19
2.11 考察 . . . 22
2.12 まとめ . . . 22

3 畳み込みニューラルネットワークを用いた JPEG 画像における改ざん領域の検出 25

3.1 概要 . . . 25
3.2 背景および目的 . . . 25
3.3 提案手法の流れ . . . 25
3.3.1 改ざん画像および教師画像の準備 . . . 28
3.3.2 入力参照範囲の決定 . . . 28
3.3.3 ネットワーク構造の決定 . . . 29
3.4 実験方法 . . . 34
3.5 実験結果 . . . 34
3.6 考察 . . . 42
3.7 まとめ . . . 42

4 LSTM 及び U-Net を用いた時空間解析による動画改ざん領域の検出 43

4.1 概要 . . . 43

4.2 背景 43
4.3 従来手法と問題点 44
4.4 研究の目的 45
4.5 検査対象とする改ざん動画 45
4.6 提案手法の流れ 48
4.6.1 LSTM による時間方向の解析 48
4.6.2 U-Net による空間方向の解析 50
4.7 改ざん動画の作成 51
4.8 実験方法 55
4.9 実験結果 58
4.10 考察 62
4.11 まとめ 65

5 結論 67

謝 辞 69

参考文献 70

研究業績 73

付録 75

図目次

2.1 Double JPEG の発生条件 8
2.2 DCT 係数のヒストグラム 8
2.3 Double JPEG 領域 (白) 10
2.4 DCT 係数のヒストグラム (再圧縮のみ) 10
2.5 MDBD 法の流れ 12
2.6 ブロックノイズの移動 13
2.7 ブロックノイズ解析結果 14
2.8 DCT 高周波成分解析に用いる 19 成分 15
2.9 block(m,n) の DCT 係数 dct(x,y) 15
2.10 DCT を行うブロック内の画素 15
2.11 改ざん画像（左列）と検出画像 $\boldsymbol{BN}$（右列） 17
2.12 $h(s)$ の定義 18
2.13 検出結果 20
2.14 実験画像と検出画像 (a)〜(d) 20
2.14 実験画像と検出画像 (e)〜(j) 21
3.1 従来手法の検出画像 26
3.2 CNN に入力する 2 値化画像 27
3.3 提案手法の CNN 構造 27
3.4 参照範囲（層数 6 ） 29
3.5 参照範囲を考慮した入力画像 30
3.6 予備実験 2 の結果 31
3.7 バリデーション精度 31
3.8 各グループのバリデーション精度（No.1〜6） 35
3.8 各グループのバリデーション精度（No.7〜10） 36
3.9 バリデーション精度 (10 グループの平均値) 36
3.10 提案手法及び各従来手法の検出結果 (Precision-Recall) 39
3.11 提案手法及び各従来手法の検出結果 (50 画像の平均値) 39
3.12 検出結果画像 (No.10, 16) 40
3.12 検出結果画像 (No.17, 36) 41
4.1 オリジナル動画 46
4.2 改ざん方法 46
4.3 改ざん領域の例 47
4.4 LSTM を用いた解析の流れ 49

4.5 U-Net 構造の CNN ネットワーク 50
4.6 改ざん動画 (No.1〜4) 52
4.6 改ざん動画 (No.5〜8) 53
4.6 改ざん動画 (No.9〜12) 54
4.7 比較実験を行う TypeA〜D 56
4.8 TypeA,C,D のバリデーションロス 57
4.9 推論動画 No.7 の検出結果 59
4.10 推論動画 No.10 の検出結果 60
4.11 推論動画 No.12 の検出結果 61
4.12 改ざん検出画素比 (推論動画 No.7) 63
4.13 改ざん検出画素比 (推論動画 No.10) 63
4.14 改ざん検出画素比 (推論動画 No.12) 64
4.15 P-R 曲線 64
A.1 DCT 係数のヒストグラム 77

表目次

2.1 検出結果 19
3.1 予備実験 1 の結果 29
3.2 CNN モデルとその精度 32
3.2 CNN モデルとその精度（続き） 33
3.3 提案手法及び各従来手法の検出結果 (No.1〜30) 37
3.3 提案手法及び各従来手法の検出結果 (No.31〜50) 38
4.1 検出結果 58

1　緒論

近年、ディジタルカメラやスマートフォンの普及により多くの人が日常的にディジタル写真を撮影することができる。ポケットから取り出して数秒も経たずにシャッターを切れるという手軽さゆえ、観光地やイベントでの撮影のみならず、ふと目についた道端の花々や販売店のディスプレイ、家族や友人との何気ない生活風景など被写体は様々である。オートフォーカス機能や自動コントラスト調整などの機能も備わり、カメラを専門的に扱わない人であってもきれいな写真を撮影することができる。

また、防犯意識の高まりにより、商業施設や店舗、商店街や公共機関、さらには一般民家など、街中のあらゆる場所に防犯カメラや監視カメラが設置されている。これらは 24 時間 365 日稼働しており、常に付近の様子を記録している。記録された映像は店舗の機器に一定期間保存されており、必要となったときには該当する時間帯の映像を確認することができる。

不幸にして思いもよらない事故や事件に巻き込まれたとき、当事者や周囲に居合わせた人は手に持ったディジタルカメラでその状況を撮影することができるし、その場に防犯カメラが設置されていれば、当時の状況が記録されていることもある。あるいは、関係のない第三者の撮影写真に偶然にも事故や事件現場が写りこんでいることはあり得るし、離れた場所に設置された防犯カメラ映像であっても有力な情報が得られることがある。ディジタル写真や映像は当時の状況をありのまま記録しているため客観証拠としての価値が非常に高く、犯罪捜査や事故状況の確認などの警察活動において極めて有用な資料であり、いつ、どこで、何があったのかを正確に把握するための重要な手段の一つとなる。

一方で、ディジタルカメラの普及に伴いディジタル写真の加工や編集が行えるアプリケーションも数多く流通している。購入したディジタルカメラに画像編集ソフトが付属されていたり、インターネットからフリーソフトをダウンロードすることができる。また、家電量販店などではディジタルカメラとともに多くのソフトウェアが販売されている。画像処理に関する専門的な知識がなくてもインターネット検索により自分が行いたい加工方法を調べられるため、PC さえあれば誰でもディジタル写真を加工することができる、

ディジタル画像を取り巻くこのような昨今の状況を考慮すると、ディジタル画像資料をありのままの真実を捉えたものとして正しく活用するためには、その画像に悪意ある何者かの加工、すなわち改ざんが施されていないかを検証することが必要不可欠である。改ざん画像であればそれを看破して検出し、改ざんされていない画像についてはそれが真正であることを確認する作業が重要となる。このような考えは法科学（フォレンジック・サイエンス、Forensic Science）において重要な意味を持つ。法科学とは、事件の

解決や法廷での立証のために用いられる科学であり、公正かつ公平で適切に手続きや分析、結果の解釈を行うことが求められている。我が国では日本法科学技術学会、海外では Forensic Science International などの機関が設立されており、あらゆる科学部門において日々研究が行われている。H.Farid[1] はディジタル画像の改ざん検出について幅広く研究しており、自身の lecture notes において様々な手法を説明している。

本論文は、改ざんが加えられたディジタル静止画やディジタル動画の検出に焦点を当て、「法科学における画像解析技術の高度化に関する研究」としてまとめたものである。

第 2 章では、改ざんされた JPEG 画像とされていない JPEG 画像を判別し、改ざん画像と判別した画像についてはその領域を表示する手法についての研究をまとめている。静止画像の改ざんには様々な手法が考えられ、一般的な例として画像内の一部をコピーして異なる部分へペーストすることが挙げられる。検出手法として、SIFT 特徴点や DCT 係数、被写体の光の当たり方に着目した解析などが提案されている [2]-[6]。また、撮影されたカメラの特性推定から改ざんを検出する手法も報告されている [7]-[10]。今日のディジタル静止画の多くは JPEG 形式であり、多くの撮影デバイスで採用されていることから、JPEG 画像に特化した検出手法の研究も盛んに行われている [11]-[14]。

He らの手法 [14] は JPEG 画像を対象とした実用的なものであり、改ざん画像であるか否かの判別だけでなく、改ざん領域の特定まで行う優れた手法であった。しかしながら、同手法では画像を単に再圧縮したり全体の明るさを補正するだけの処理に対しても改ざんと判断することがあった。撮影したディジタル写真の再圧縮や明るさ補正などは誰もが容易に行うことができることから、これを「改ざん」として判断することが妥当でない場合がある。

このような課題を克服するため、本章では 2 つアプローチを併用した解析手法を行い、コピー＆ペーストの改ざん処理を施された画像を改ざん画像として判別する手法を提案する。 1 つはブロックノイズに着目した解析である。JPEG 画像は画像データが離散コサイン変換（DCT）され、低周波成分から高周波成分までが分けられて量子化される [15]。DCT は 8×8 画素で行われるためブロックノイズと呼ばれる 8×8 画素サイズのタイル状の歪みが生じる。このブロックノイズを改ざんの痕跡を探る端緒とするものである。もう 1 つは JPEG 画像が改ざんにより 2 度目の圧縮が行われることにより生じる痕跡（「Double JPEG」と呼ばれる。）に着目した解析である。2 つの解析から得られる特徴量を定量化して Support Vector Machine（SVM,[16]）に入力することで異なるアプローチからの結果を総合的に判断することが可能となり、従来手法よりも高い判別精度が得られる。

第 3 章では、改ざんされた JPEG 画像において、畳み込みニューラルネットワーク（CNN）を用いることで、改ざん領域の検出精度を向上させる手法についての研究をまとめている。第 2 章で述べた提案手法では判別精度が向上したものの、「改ざん領域」を正確に検出する点において課題が残った。改ざん領域について適切に表示させることがで

きれば改ざん画像であると判別した結果の信頼度が高まり、改ざん領域を誤って表示していれば判別結果の信頼度が低下することとなる。第 2 章で述べた成果から、改ざん領域の検出には DCT 係数の高周波成分を解析することが有効であり、加えて、改ざんが疑われる領域の空間的連結を考慮することが重要であるという有用な知見が得られた。そこで、改ざん領域の空間的精度を向上させるために CNN を適用する手法を提案する。

近年、画像認識分野で注目を浴びている CNN などのディープラーニング技術は、画像改ざん検出においても利用されている。画像に対して施した処理（メディアンフィルタなど）を識別する手法 [17]、カメラで撮影した画像か CG かを識別する手法 [18]、CNN を用いてカメラ特性を推定することで改ざんを検出する手法 [19] などがある。また、LSTM を用いて改ざん領域境界の不連続性を検出する手法 [20] や、CNN を用いて DCT 係数のヒストグラムを解析することで改ざん領域を検出する手法 [21] が報告されている。

本章で提案する手法は、DCT 係数に処理を加えた 2 次元画像を 5 層で構成される CNN に入力し、改ざん領域の検出画像を出力するものである。改ざん領域の検出漏れを低減するとともに非改ざん部の誤検出を抑えることに成功し、従来手法よりも高い検出精度を達成することができる。

第 4 章では、改ざんが加えられたディジタル動画に対する改ざん領域の検出手法についての研究をまとめている。ディジタル静止画と同様にディジタル動画も撮影デバイスや編集ツールが普及しており、動画においても改ざん検出手法の研究が行われている。JPEG 画像の改ざん検出に用いられる Double JPEG 技術を MPEG 動画に応用したもの [22],[23]、HOG 特徴量を用いて動画内のコピー＆ペーストを検出するもの [24]、カメラのノイズ特性から改ざん領域を検出するもの [25]、クロマキー撮影により作成した動画を対象とするもの [26]-[28] などが報告されている。P.Bestagini ら [29] は、背景画像を用いて被写体を消去した改ざん動画の検出手法を提案しており、これは動画に写っている不都合な事実を隠蔽するという状況を想定しているものの、改ざん領域が空間的に固定されているため検査対象が限られるものであった。

本章で述べる提案手法は、P.Bestagini らと同様に動画内の被写体を消去している改ざん動画を対象としているが、より実用的なシチュエーション、例えば画面手前から奥へ人が歩いている状況などを想定して改ざん領域の移動及び変形に対応している。動画は静止画の連続体であり、改ざんされた静止画像が時間方向へ複数フレーム続いているものと考えることができるため、空間的に解析を行う CNN に加えて、時間方向に解析を行う RNN を用いる。検査対象とするフレームと前後 30 フレームを RNN で解析することにより、改ざん領域の時間的ゆらぎを特徴量として捉えることができる。また、静止画と同様に改ざん領域の空間的連結性を解析するために検査対象フレームに対して CNN 解析を用いる。動画はエンコード時の圧縮の影響により、静止画と比較して改ざん痕跡が残りにくい傾向があるが、時空間解析を行うことによりわずかな痕跡を抽出することが可能となる。CNN のみ、RNN のみの検出結果と比較し、両者を用いる時空間解析の有効

性を評価する。

第 5 章は、まとめとして本研究の結論と今後の展望について述べる。

2 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

2.1 概要

本章では、ブロックノイズ解析と Double JPEG 解析の 2 つの側面から画像の解析を行い、改ざん画像と非改ざん画像を判別する手法を提案する。ブロックノイズ解析はさらに DCT 高周波成分解析法とブロックノイズ位置推定法の 2 つの解析法に分かれる。ブロックノイズ解析と Double JPEG 解析から得られるパラメータを定量的に算出して SVM に入力することで改ざんの有無を総合的に判別し、さらに改ざん領域を可視化する。提案手法では He らの従来手法よりも誤検出を抑えることで全体の検出精度を高めることに成功し、改ざん領域を視覚的に示すことにより本手法の信頼性を高めることを実現している。

2.2 背景

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、インターネットから簡単にダウンロードすることができ、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、誰でも画像処理を行える環境となっている。一方で、デジタル写真は被写体をありのまま記録するため、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。

ここで、本章で述べる研究で扱う改ざん画像について定義する。撮影したディジタル写真に対して、局所的に被写体の追加や被写体の削除を行うことを改ざんと呼ぶ。具体的には、本来写っていない人物を他の画像からコピー＆ペーストすることにより追加したり、また、本来写っている人物を消すために、背景画像の一部をコピー＆ペーストすることなどである。しかしながら、両者の実際に行う画像処理の過程は本質的には同じである。

2.3　先行研究

ディジタル画像の改ざん検出手法はこれまでも国内外において研究されている。同じ画像内のコピー＆ペーストを行う改ざんに対して SIFT 特徴点や DCT 係数類似性を解析する手法、対象物に対する光の当たり方を推定する手法、カメラのノイズ特性を解析する手法など様々なアプローチが報告されている。

今日のディジタル静止画の多くは JPEG 形式であることから、JPEG 画像に特化した検出手法の研究も盛んに行われている。JPEG 画像は撮影時にカメラにより JPEG 圧縮が行われ、画像処理ソフトで改ざんを施して保存するときには、そのソフトにより再び JPEG 圧縮が行われる。カメラが保有する量子化テーブルとソフトが保有する量子化テーブルは一般に異なることから、二度の JPEG 圧縮を受けた JPEG 画像には量子化テーブルの違いに由来する特異な痕跡が残る。これは「Double JPEG」と呼ばれ、この痕跡を端緒に改ざんを検出する手法が報告されている。

T.Bianchi ら [11] は、ある画像の一部をコピーして他の画像へペーストした場合に、改ざん領域と他の領域では Double JPEG の影響から DCT 係数の周期性が異なるとした検出手法を提案した。また、T.Pevny ら [12] は、steganography（電子迷彩技術）を用いて JPEG 画像内に秘密情報を埋め込んだ場合に、DCT 係数の低周波数領域のヒストグラムを解析することにより、steganography を使った二重圧縮か、または単なる二重圧縮であるかを識別する方法を提案した。

J.He ら [14] は、Doubel JPEG が与える影響を改ざん画像内の各エリアにおいて解析を行い、各エリアがもつ特徴量を分類することで改ざん画像であるかオリジナル画像であるか、さらには、改ざん画像と判定した場合には改ざん領域を特定する手法を提案した。コピー＆ペーストによる改ざんだけでなく、画像内の一部の塗りつぶしにも対応しており、改ざん領域の形状や大きさに制限を与えないため、非常に強力で有効な検出手法の一つである。

2.4　先行研究の問題点

改ざん検出手法は様々提案されているものの、改ざん手法も多種多様であるため、どれか一つを用いればあらゆる改ざんを検出できるというものはない。同一画像内のコピー＆ペーストの精度や頑強性を高めても、他の画像からコピーした領域をペーストする改ざん画像には対しては無力であるし、画像 1 枚だけを検査する場合にはカメラのノイズ特性を正しく推定することは困難である。JPEG 画像に特有の Double JPEG を利用する手法は効果的であるが、JPEG 以外の画像では同手法を適用することができない。ここに、改ざん検出の難しさが顕著に表れているといえる。一つの検出手法にのみ頼る

と想定外の改ざん画像を検査した場合に改ざんを見落とす可能性がある。

これまで提案されてきた手法の中でも、特に He らの手法 [14] は JPEG 形式の静止画像を対象とした実用的なものであり、改ざん画像であるか否かの判別だけでなく、改ざん領域の特定まで行う優れた手法であった。しかしながら、Double JPEG のみに着目する解析手法であるため、二重圧縮された画像であれば、改ざん処理を施していなくても改ざん画像として誤検出することがあった。昨今では日常的にディジタル画像を扱い、だれでも手軽に画像処理ソフトで明るさ調整やトリミングを行えることから、改ざんの意思が無くても異なる量子化テーブルで二度の JPEG 圧縮を行ってしまうことは十分にあり得る。これらの画像をすべて改ざん画像と判別することが妥当でない場合がある。

2.5 Double JPEG について

本章で述べる提案手法や He らの従来手法では、JPEG 画像に特有の Double JPEG と呼ばれる現象の解析が鍵となるため、ここで Double JPEG について説明する。

JPEG 画像はその国際標準規格 [15] により 8×8 画素のブロックに区分けした画素ごとに離散コサイン変換（DCT）され、撮影したデジタルカメラが保有する量子化テーブル A の値により量子化される。これが一度目の量子化であり、どのような JPEG 画像であってもこの過程を経ている。その後、画像編集ソフトにより JPEG 画像を展開し、画像に対して改ざん処理もしくは明度調整など行い 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 B の値により量子化が行われる。これが二度目の量子化であり、ここで Double JPEG が生じる (図 2.1)。

Double JPEG の有無は DCT 係数のヒストグラムを表示すると視覚的に認識しやすい。図 2.2 にオリジナル画像と Double JPEG が生じた改ざん画像の DCT 係数のヒストグラムを示す。図 2.2(b) のようにヒストグラムが二重に重なっているように見えるのは、一度目の量子化テーブルと二度目の量子化テーブルの値が異なることが原因である（付録参照）。

2.6 He らの従来手法

JPEG 画像を対象とした改ざん画像の検出手法として、He らは Double JPEG に注目した検出手法を提案した。画像の一部をコピーし別の画像にペーストして改ざんを行った場合、非改ざん領域に属するブロック（白）は一度目の量子化と同じ 8×8 画素のブロックの画素メンバーで DCT された後に二度目の量子化が行われるため Double JPEG が生じる (図 2.3)。一方、改ざん領域に属するブロック（黒）は一度目の量子化とは異なる画素メンバーで DCT された後に二度目の量子化が行われるため、Double JPEG が生じない。

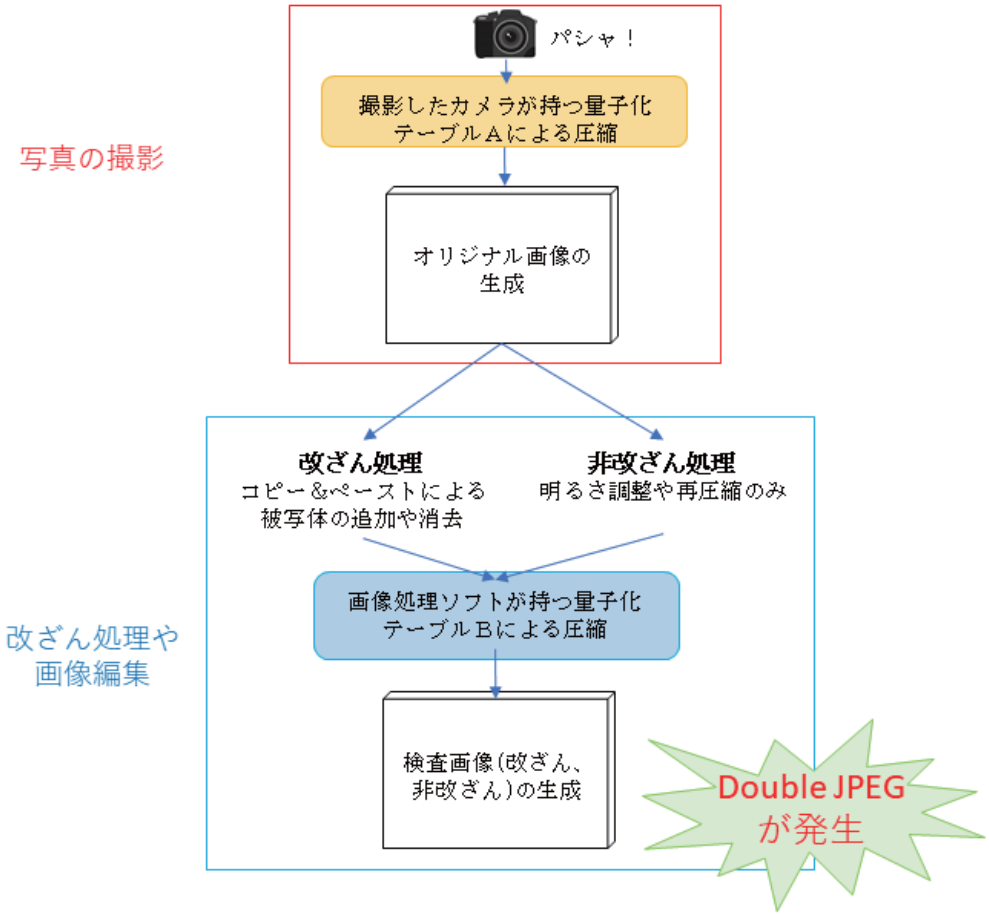
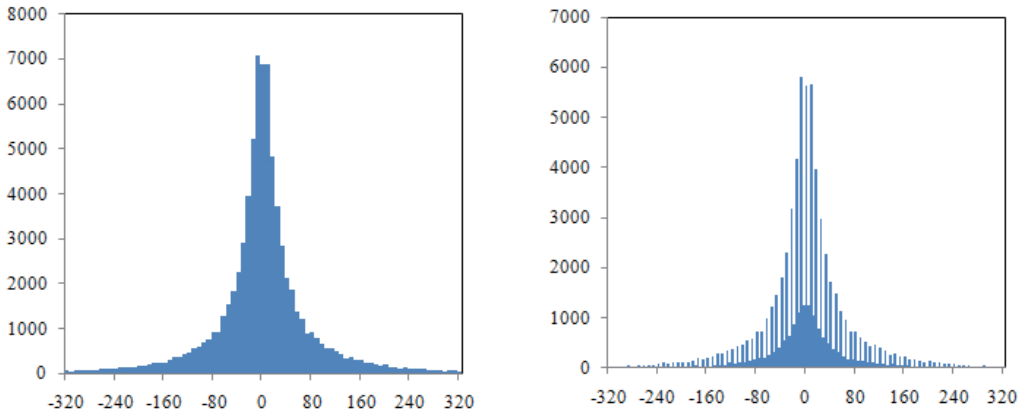


図 2.1　Double JPEG の発生条件



(a) オリジナル画像　　(b) 改ざん画像

図 2.2　DCT 係数のヒストグラム

He らはまず図 2.2(b) のような櫛状のヒストグラムの周期 p を式 (2.1) より算出し、$p = 1$ のとき、つまり図 2.2(a) のときは非改ざん画像と判断した。

$$p = \underset{p}{\operatorname{argmax}} H(p) \tag{2.1}$$

$$H(p) = \frac{1}{i_{max} - i_{min} + 1} \sum_{i=i_{min}}^{i_{max}} [h(i \cdot p + s_0)] \tag{2.2}$$

p >1 のとき、改ざん領域では Double JPEG が生じないため図 2.2(b) の小さいヒストグラムに寄与し、非改ざん領域では Double JPEG が生じるため図 2.2(b) の大きいヒストグラムに寄与する確率が高いとして 8×8 画素のブロックごとに改ざん領域である確率を導出し、その確率分布から各ブロックを改ざん領域クラス C_0 と非改ざん領域クラス C_1 に分類した。それぞれのクラスの分散 σ_0、σ_1 から σ を式 (2.3) より、C_0 と C_1 を分ける閾値 T_{opt} を式 (2.4) より求めた。

$$\sigma = \sigma_0^2 - \sigma_1^2 \tag{2.3}$$

$$T_{opt} = \underset{T}{\operatorname{argmax}}(\sigma / (\sigma_0 + \sigma_1)) \tag{2.4}$$

さらに、C_0 クラスに該当する N_0 個のブロックのうち i 番目のブロックの周辺 4 近傍の連結個数を e_i とするとき、連結性 K_0 を式 (2.5) と定義した。

$$K_0 = \sum_i \frac{max(e_i - 2, 0)}{N_0} \tag{2.5}$$

以上より導出した 4 変数、T_{opt}、σ、σ_0+σ_1、K_0 を SVM に入力し、改ざん画像か非改ざん画像かを判別した。

しかしながら、コピー＆ペーストなどの改ざん行為を行わず、異なる量子化テーブルを持つフォトビューアソフトなどで上下の向きを変えたり画像全体の明るさを変えて保存するなど悪意なく画像に処理を加えた場合にも画像の全領域において Double JPEG が生じるため、DCT 係数のヒストグラムは図 2.4 のようになり、p >1 となる。そのため、He らの手法ではこの画像を改ざん画像の可能性があるものと認識して画像内を C_0 と C_1 に分類し、結果として改ざん画像として誤判定することがある。簡単に画像処理の行えるソフトが普及している今日の状況を考えると、これらの処理を改ざんとして検出することは誤判定となってしまう。

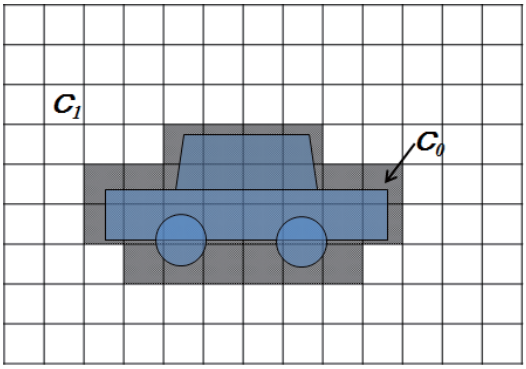


図 2.3　Double JPEG 領域 (白)

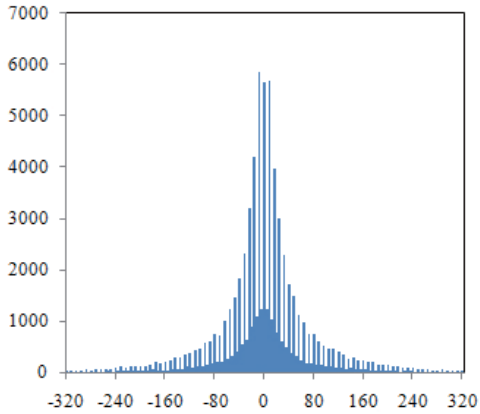


図 2.4　DCT 係数のヒストグラム (再圧縮のみ)

2.7　研究の目的

本研究では、He らと同じく JPEG 形式の改ざん静止画像の検出を目的としているが、“コピー＆ペーストにより画像内の被写体を加工した改ざん画像” を的確に検出し、改ざんでない画像処理を行ったものを非改ざん画像として判断することを目的としている。複数の解析を行って複合的に判断することにより、He らの手法で生じた誤検出の割合を低減させ、従来手法よりも高い精度で改ざん画像と非改ざん画像の判別を行う手法を提案する。

2.8　提案手法の流れ

提案する改ざん検出法（「MDBD 法、Multiple Detection using Block Noise and Double JPEG」と呼ぶ。）は 2 つの解析から改ざんの有無を判定する手法であり、検出過程の概略を図 2.5 に示す。本手法は、画像内のブロックノイズに注目したブロックノイズ解析法と、Double JPEG の度合いを定量的に評価する Double JPEG 解析法で構成される。ブロックノイズ解析法は、さらに DCT 高周波成分解析法とブロックノイズ位置推定法に分かれる。以下に、各解析方法の詳細を述べる。

2.8.1　ブロックノイズ解析

JPEG 画像は 8×8 画素ごとに区分けして DCT を行い量子化されるため、JPEG 画像には 8×8 画素ごとに規則的にタイルが並んだようなノイズがみられる。これはブロックノイズ（ブロック歪み）と呼ばれ JPEG 画像特有のものであり、圧縮が大きいほど、つまり量子化係数が大きいほど顕著に見られる。ブロックノイズが生じると極めて目障りであるため、通常のデジタルカメラで撮影された画像はブロックノイズが視認できない程度の圧縮レベルで抑えられているが、JPEG 画像である以上はブロックノイズの発生を避けることはできない（図 2.6(a) の実線）。

ある画像の一部を切り取り、これを同一の画像または他の画像に張り付けるとき、このブロックノイズの位置はほとんどの場合、背景となる画像の 8×8 画素のブロック境界とずれてしまう（図 2.6(b) の点線）。2 つのブロックが正確に一致する確率はわずか 1/64=1.6% である。また、貼り付け部の形状は種々多様であり、矩形であることは現実的にあり得ない。したがって、貼り付け部境界に位置するブロックには、貼り付け部の境界が入り込む。このため、改ざん後の再圧縮時に改ざん部及び改ざん部境界に位置する 8×8 画素ブロックが DCT されると、改ざん部のブロックノイズの境界及び改ざん部の境界が新たなテクスチャとして加わったことになるため高周波成分が生じる。改ざん部以外の領域は撮影時の一度目の量子化により高周波成分が一様にカットされており、改ざん後も DCT されるブロックの画素メンバーに変化がないため、新たに高周波成分が生じることはない。よって、画像内で比較的高周波成分が集中して残っている領域の抽出により改ざん領域を検出することがブロックノイズ解析法の目的となる。

画像にはさまざまな情景が写っており、細かいテクスチャが多く含まれる領域も存在する。例えば、木々が写っている山中の写真や細かな模様の絨毯などである。単に高周波成分が残っている領域を抽出するだけでは、これらの本来から写っている細かなテクスチャ領域までも抽出してしまう。ブロックノイズ解析法では、検出結果が高周波成分のみに依存することを避けるために、DCT 高周波成分解析法とブロックノイズ位置推定

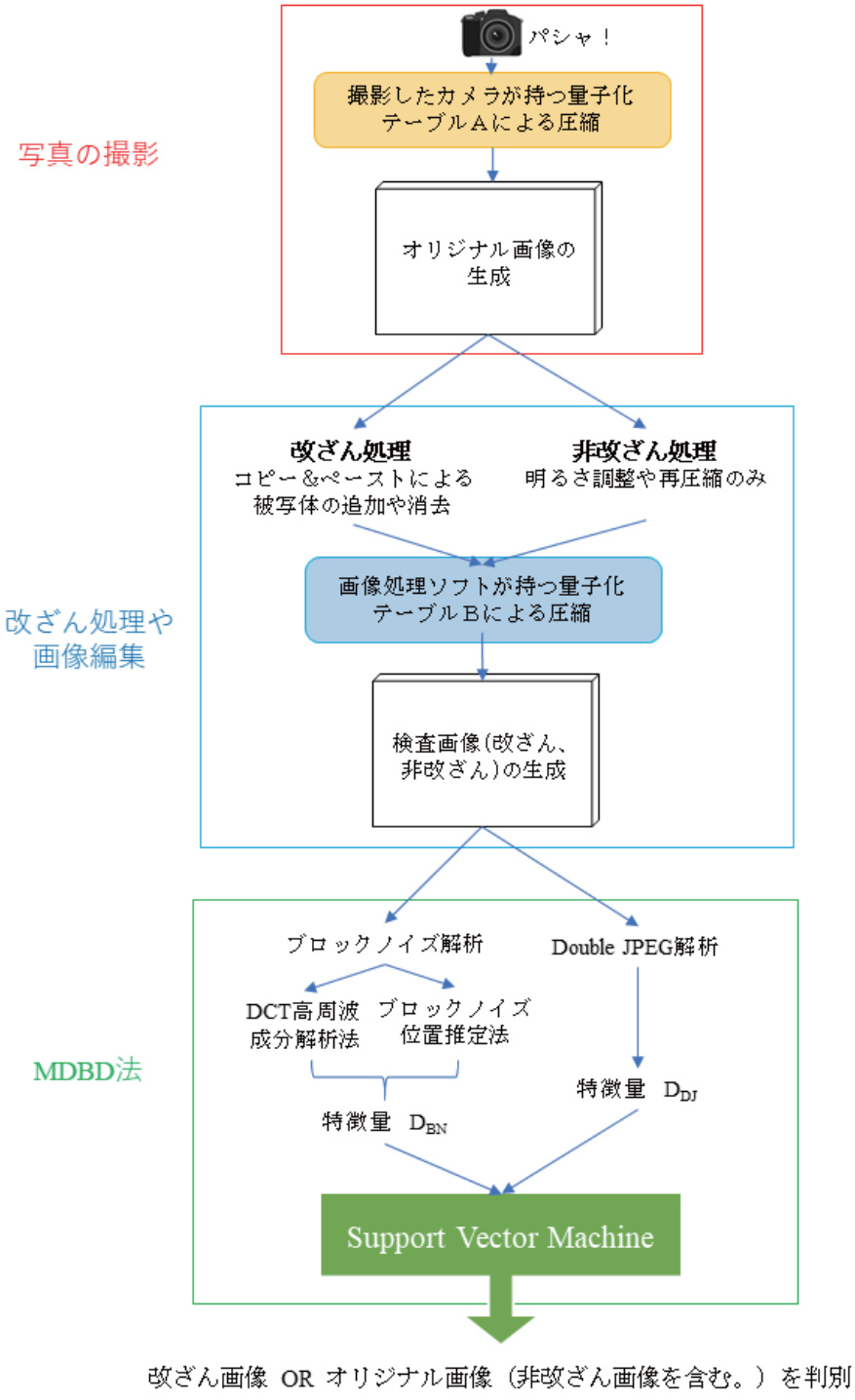
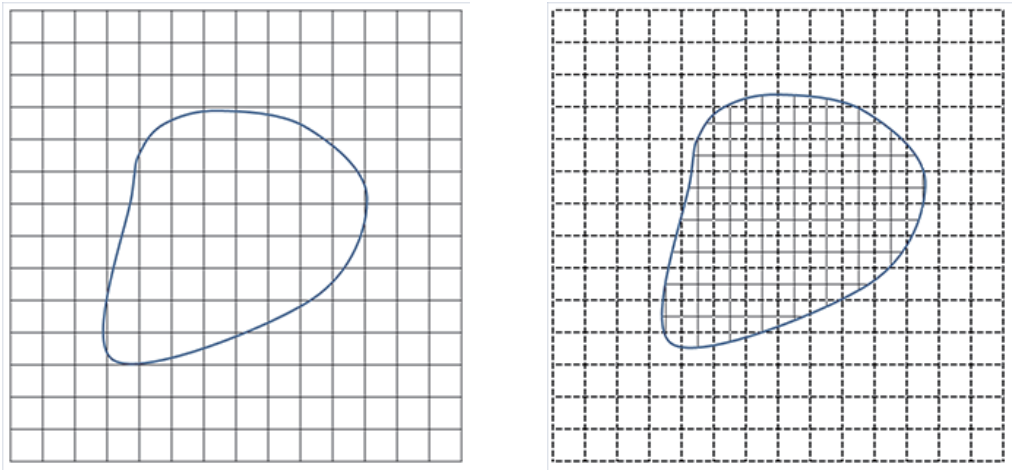


図 2.5　MDBD 法の流れ

法の 2 つの解析手法に分ける。それぞれの検出結果画像を $\boldsymbol{D}$、$\boldsymbol{L}$ とし、ブロックノイズ解析法の最終結果画像 $\boldsymbol{BN}$ を $\boldsymbol{BN} = \boldsymbol{D} \cap \boldsymbol{L}$ で求める（図 2.7）。



(a) オリジナル画像　　(b) 改ざん画像

図 2.6　ブロックノイズの移動

2.8.1.1　DCT 高周波成分解析法

画像内の 8×8 画素ブロックの DCT 係数について、図 2.8 に示す 19 成分 (6,0)、(6,1)、(0,6)、(1,6)、(7,0)、(7,1)、(7,2)、(7,3)、(7,4)、(7,5)、(7,6)、(7,7)、(0,7)、(1,7)、(2,7)、(3,7)、(4,7)、(5,7)、(6,7) のうち 0 でない値を持つ成分の個数を N とし、$N \geq 7$ となるブロックが右に連続している場合は改ざん候補ブロックとする。 改ざん候補ブロックを 1、非改ざん候補ブロックを 0 とした二値化画像を $\boldsymbol{D}$ とする（図 2.7(b)）。

2.8.1.2　ブロックノイズ位置推定法

改ざん領域はブロックノイズの境界がずれている確率が 98.4% と非常に高く、非改ざん領域はブロックノイズの位置に変化がない。そこで、ブロックノイズの境界位置を推定することで改ざん候補となるブロックを検出する。

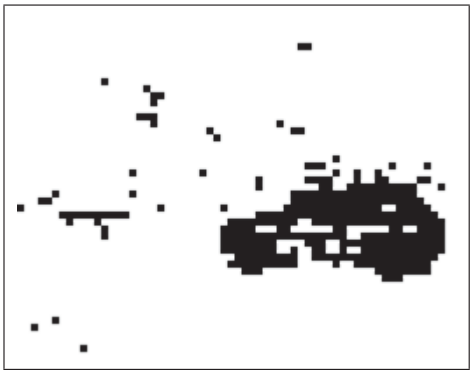
画像全体を 8×8 画素のブロックに分割したとき、(m, n) 番目のブロックに対して DCT を行い、DCT 係数の (x, y) 成分を $dct(x, y)$ とする（図 2.9）。$x + y$>8 となる $dct(x, y)$ の絶対値の総和を $S_{(m,n)}$ とする。

$$S_{(m,n)} = \sum_{x+y>8} |dct(x, y)| \tag{2.6}$$

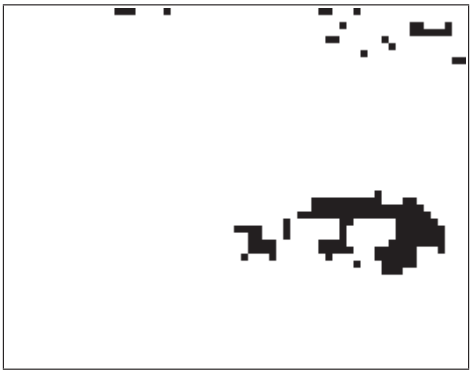
次に、DCT を行うブロックを右に i 画素、下に j 画素ずらして DCT を行う。ここで $0 \leq i, j \leq 7$ であり、$S_{(m,n)}$ を $S_{(m,n)(i,j)}$ と表すと、式 (2.6) の $S_{(m,n)}$ は $S_{(m,n)(0,0)}$ と



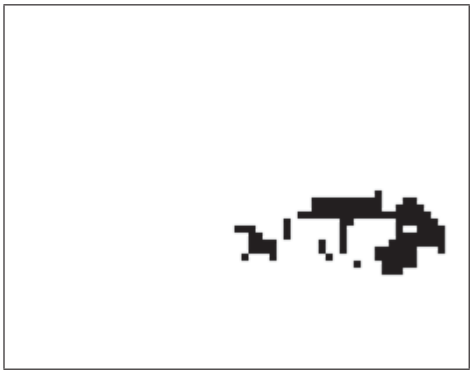
(a) 改ざん画像



(b) $\boldsymbol{D}$：DCT 高周波成分解析



(c) $\boldsymbol{L}$：ブロックノイズ位置推定法



(d) $\boldsymbol{BN}$：最終結果画像

図 2.7　ブロックノイズ解析結果

表せる（図 2.10）。DCT を行うブロック内にブロックノイズの境界が含まれると高周波成分が大きくなるため、$S_{(m,n)(i,j)}$ が大きな値となる。したがって、$S_{(m,n)(i,j)}$ のうち、$S_{(m,n)(0,0)}$ が最小となった場合のみ非改ざんブロックとし、$S_{(m,n)(0,0)}$ 以外で最小となった場合は改ざん候補ブロックとする。

$$block(m,n) = \begin{cases} \text{undoctored} & \text{if } \ S_{(m,n)(0,0)} = \min\limits_{(i,j)} S_{(m,n)(i,j)} \\ \text{doctored} & \text{else} \end{cases}$$

改ざん候補ブロックを 1、非改ざんブロックを 0 とした二値化画像を作成し、膨張処理を行うことで孤立ブロックを除去した画像を $\boldsymbol{L}$ とする（図 2.7(c)）。

2.8.1.3　ブロックノイズ解析法による改ざん度の算出

ブロックノイズ解析法により求めた改ざん検出画像 $\boldsymbol{BN}$ の疑わしさを定量的に表すため、改ざん度 D_{BN} を次式で定義する。

(0, 0)	(1, 0)	(2, 0)	(3, 0)	(4, 0)	(5, 0)	(6, 0)	(7, 0)
(0, 1)	(1, 1)	(2, 1)	(3, 1)	(4, 1)	(5, 1)	(6, 1)	(7, 1)
(0, 2)	(1, 2)	(2, 2)	(3, 2)	(4, 2)	(5, 2)	(6, 2)	(7, 2)
(0, 3)	(1, 3)	(2, 3)	(3, 3)	(4, 3)	(5, 3)	(6, 3)	(7, 3)
(0, 4)	(1, 4)	(2, 4)	(3, 4)	(4, 4)	(5, 4)	(6, 4)	(7, 4)
(0, 5)	(1, 5)	(2, 5)	(3, 5)	(4, 5)	(5, 5)	(6, 5)	(7, 5)
(0, 6)	(1, 6)	(2, 6)	(3, 6)	(4, 6)	(5, 6)	(6, 6)	(7, 6)
(0, 7)	(1, 7)	(2, 7)	(3, 7)	(4, 7)	(5, 7)	(6, 7)	(7, 7)

図 2.8 DCT 高周波成分解析に用いる 19 成分

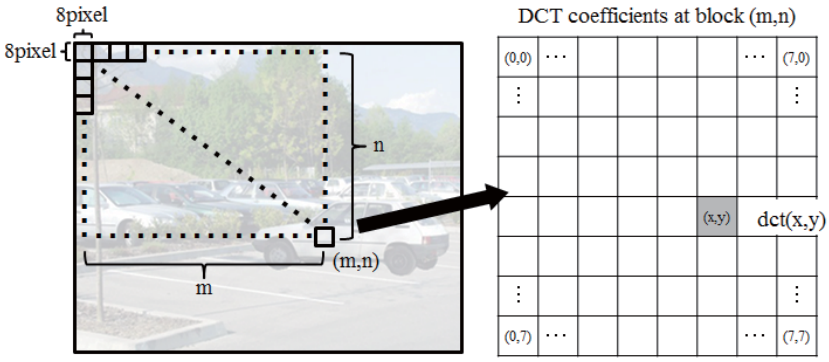


図 2.9 block(m,n) の DCT 係数 dct(x,y)

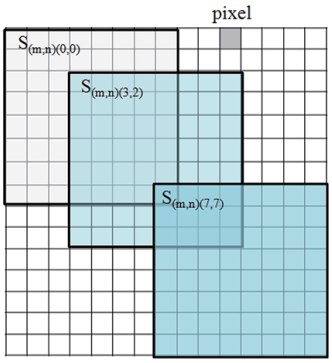


図 2.10 DCT を行うブロック内の画素

$$D_{BN} = \log(\frac{P \cdot w \cdot h}{P_{ed}} + 1) \tag{2.7}$$

$$P = \sum_{BN(m,n)=1} \frac{\sum_{i=-1}^{1}\sum_{j=-1}^{1} 255 \times BN(m+i, n+j)}{3 \times 3} \tag{2.8}$$

$$P_{ed} = \sum_{x,y} |f(x,y) * g(x,y) - f(x,y)| \tag{2.9}$$

ここで、P は改ざん候補ブロックを中心とする 3×3 ブロックの領域ごとに求めた密度の総和、w は対象画像の横サイズ、h は縦サイズである。また、P_{ed} は対象画像 $f(x,y)$ にガウシアンフィルタ $g(x,y)$ を畳み込みしたものと $f(x,y)$ との差分の絶対値の総和である。

画像の一部をコピーし異なる画像にペーストし、その輪郭にスムージング処理を行った改ざん画像の改ざん度 D_{BN} を算出した例を示す（図 2.11）。貼り付け領域のブロックノイズの境界と背景画像のブロックノイズの境界が一致するとき、DCT を行うブロック画素のメンバーに変化はない。従って、高周波成分にブロックノイズの影響は生じずブロックノイズの位置も変わらないためブロックノイズ解析法では検出できない（図 2.11(a): ブロックノイズの位置が (8×k, 8×l)）。貼り付け領域内のブロックノイズの左上画素の x,y 座標のどちらかが 8 の倍数でないときは、高周波成分にブロックノイズの影響が生じ、かつブロックノイズの位置が変化するため、ブロックノイズ解析法で検出可能である（図 2.11(c):(8×k, 8×l+4)、(e):(8×k+4, 8×l)、(g):(8×k+4, 8×l+4）)。ここで、k、l は共に整数である。

2.8.2　Double JPEG 解析法

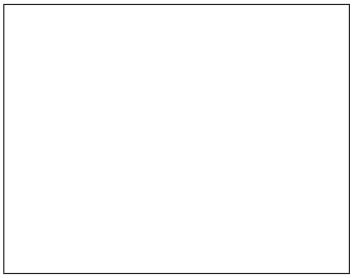
異なる量子化テーブルにより二度の量子化（Double JPEG）が行われた JPEG 画像では、DCT 係数のヒストグラムが櫛歯状になる。この櫛歯状のヒストグラムは実験的に (2,1) 成分に顕著にみられた。(2,1) 成分のヒストグラム h を作成し、s 番目の頻度の値を $h(s)$ とする。$h(s)$ が最大となるインデックスを s_0、$h(s_0)$ の次に大きい h(s) のインデックスを s_1 としたとき、Double JPEG の度合を示す値 D_{DJ} を式 (2.10) で定義する。D_{DJ} が小さいほど Doube JPEG の影響が大きいと判断する (図 2.12)。

$$D_{DJ} = \frac{h(s_0 - 1)}{h(s_1)} \tag{2.10}$$

以上により導出した 2 変数 D_{BN} 及び D_{DJ} を SVM に入力し、改ざん画像と非改ざん画像を判別する。



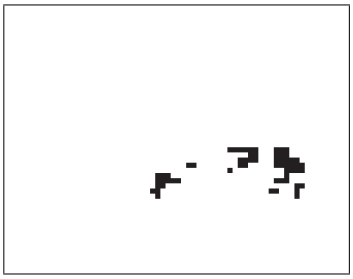
(a) 改ざん画像



(b) D_{BN}=0



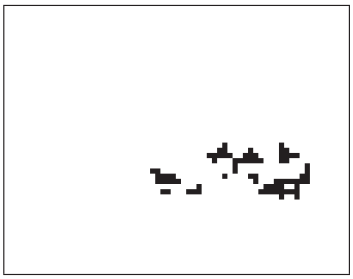
(c) 改ざん画像



(d) D_{BN}=3.4



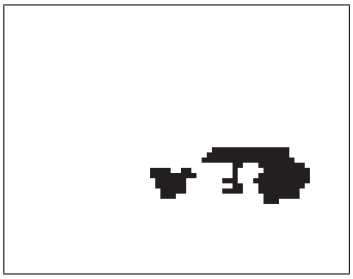
(e) 改ざん画像



(f) D_{BN}=3.5



(g) 改ざん画像



(h) D_{BN}=3.8

図 2.11 改ざん画像（左列）と検出画像 $\boldsymbol{BN}$（右列）

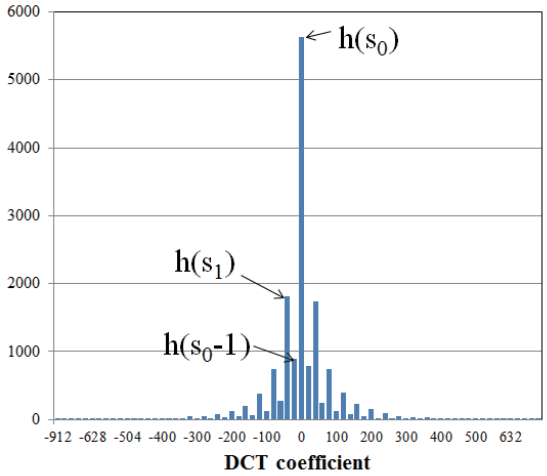


図 2.12　$h(s)$ の定義

2.9　実験方法

実験に用いるディジタル画像には、一般的な家庭用ディジタルカメラ 5 機種（CASIO:EX-ZR200, SONY:NEX-3N, Canon:IXY510, OLYMPUS:FE220, NIKON:AW210）で撮影した計 50 枚の画像を使用する。撮影画像の被写体は、自然風景、人物、建物など様々である。改ざん画像の作製方法として、まず画像内のある被写体を輪郭に沿って切り取り（コピー）、次に同機種で撮影した他の画像に貼り付け（ペースト）、JPEG 画像で保存する。保存時の設定は最高画質（最低圧縮）とする。改ざん画像 50 枚、非改ざん画像 50 枚を用意し、非改ざん画像の内訳は、まったく手を加えていないオリジナル画像 30 枚、画像処理ソフトで上書き保存 (再圧縮) のみを行った画像 10 枚、画像全体の明度補正のみを行った画像 10 枚とする。以上の処理には、Photoshop CS を使用する。計 100 枚の画像に対して SVM(libsvm-3.20[16]) を使用した分類実験を行い、その方法は、100 枚の画像のうち 90 枚を学習画像、残りの 10 枚をテスト画像とした分類実験を 10 回繰り返す交差検定（10-fold cross validation）である。SVM に入力する変数は、MDBD 法では D_{BN} と D_{DJ} であり、He らの手法では、T_{opt}、σ、$\sigma_0+\sigma_1$、K_0 である。なお、T_{opt} の導出にあたっては、式 (2.4) を用いず、より正確に C_0 と C_1 をクラス分けできる判別分析法を用いる。

2.10 実験結果

改ざん画像 50 枚のうち検出した画像数を T_P、検出できなかった画像数を F_N、非改ざん画像のうち誤検出した画像数を F_P とするとき、適合率 P(Precision)、再現率 R(Recall)、P と R の調和平均 F(F-measure) を次式で定義する。

$$P = \frac{T_P}{T_P + F_P} \tag{2.11}$$

$$R = \frac{T_P}{T_P + F_N} \tag{2.12}$$

$$F = \frac{2 \cdot P \cdot R}{P + R} \tag{2.13}$$

P、R、F の平均値を表 2.1、図 2.13 に、検出結果画像の一例を図 2.14 に示す。図 2.14 では、改ざん領域の輪郭を赤線で示しており、結果画像では改ざん候補ブロックを赤点で示し、候補ブロックが集中しているほど濃色で表している。

表 2.1 に示すとおり、He らの手法と比較して P、R、F の全ての値が向上しており、MDBD 法の優位性が示されている。特に P は大幅に向上している。これは非改ざん画像に分類している画像のうち、上書き保存のみまたは明度補正のみを行った画像に対する結果が良好であることが大きく寄与している。上書き保存のみまたは明度補正のみを行った 20 画像に対する判定では、He らの手法では 16 画像を改ざん画像として誤検出したのに対し、MDBD 法ではそれを 7 画像にまで抑えている。

表 2.1 検出結果

Method	Precision	Rrecall	F-measure
提案手法	0.89	0.94	0.90
He らの手法	0.75	0.92	0.82

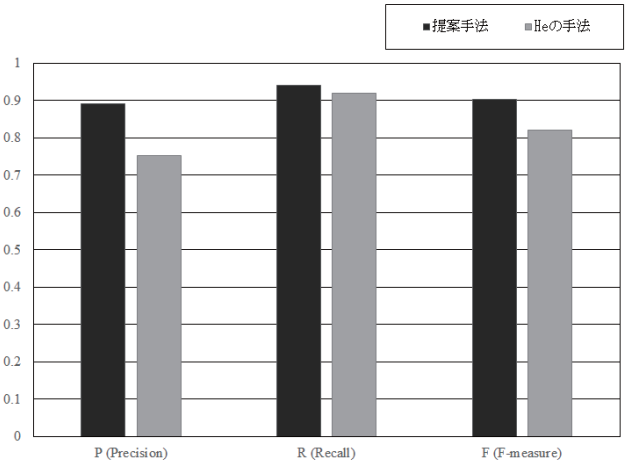


図 2.13　検出結果



(a) 改ざん画像



(b) 改ざん画像と判断
($D_{BN} = 4.2, D_{DJ} = 11$)



(c) 非改ざん画像

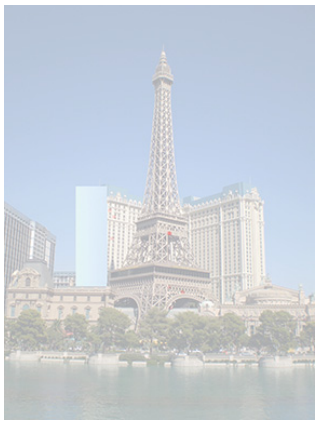


(d) 非改ざん画像と判断
($D_{BN} = 2.3, D_{DJ} = 21$)

図 2.14　実験画像と検出画像 (a)〜(d)



(e) 改ざん画像



(f) 非改ざん画像と誤判断
($D_{BN} = 3.3, D_{DJ} = 32$)



(g) 非改ざん画像



(h) 改ざん画像と誤判断
($D_{BN} = 4.7, D_{DJ} = 31$)



(i) 改ざん画像



(j) 非改ざん画像と誤判断
($D_{BN} = 3.8, D_{DJ} = 54$)

図 2.14 実験画像と検出画像 (e)〜(j)

2.11　考察

図 2.14(a) は改ざん画像であり、He らの手法では検出できなかったが MDBD 法では検出できている（図 2.14(b)）。D_{BN} が大きく D_{DJ} が小さい値をとっている。図 2.14(c) は上書き保存のみを行った非改ざん画像であり、He らの手法では改ざん画像として誤検出しているが、MDBD 法では誤検出せず非改ざん画像に分類している（図 2.14(d)）。再圧縮により D_{DJ} が小さい値であるが D_{BN} も小さい値であるため、誤検出を防いでいる。図 2.14(e) は改ざん画像であり、MDBD 法では検出できていないが、He らの手法では検出できている（図 2.14(f)）。D_{DJ} は小さい値であるが、D_{BN} も小さい値であるため検出には至っていない。D_{BN} が小さい原因として、ペーストした改ざん領域が色変化の少ないテクスチャであり、ブロックノイズに由来する高周波成分が少ないことが考えられる。

図 2.14(g) は明度補正のみを行った画像であり、He らの手法、MDBD 法で共に改ざん画像として誤検出している（図 2.14(h)）。これは、再圧縮により D_{DJ} が小さい値であり、かつ画像本来のテクスチャに由来する高周波成分のため D_{BN} が高い値となっているからである。図 2.14(i) は改ざん画像であり、He らの手法、MDBD 法で共に検出できていない。MDBD 法では改ざん部にわずかに反応しているものの、全体として改ざんとの判断には至っていない（図 2.14(j)）。これは、改ざんによる D_{BN} の値が比較的小さい値であり、かつ Double JPEG の影響が小さい（D_{DJ} の値が高い）ためである。

2.12　まとめ

提案手法の優れた点は、異なる着眼点から検査を複合的に行うことで高い精度で改ざん画像を検出することである。He らの手法は Double JPEG のみに依存しているため、上書き保存のみや明度補正のみであっても異なる量子化テーブルで保存されてしまうと、改ざん画像として判定してしまうことがある。He らの手法ではこれらの画像を検査対象とはしていないが、検査する画像にどういう処理が施されているかは本来不明であること、また、デジタル画像を取り巻く昨今の状況から明度補正等は誰でも容易に行えることを考慮すると、He らの手法では不十分である。

本章で述べた評価実験では、提案手法が適合率 P、再現率 R、調和平均 F のすべてにおいて He らの手法よりも良好な結果であり、特に上書き保存のみの画像や明度補正のみを行った非改ざん画像に対しては分類精度が大幅に向上している。改ざん画像として判断した場合における改ざん領域の視覚的表示については、適切にできた画像もあれば、不十分な画像もあり、精度の向上が必要である。改ざん領域の視覚的表示を的確に行うことが出来れば、判断結果の信頼性が高まり説得力を与えることができる。この課題を克

服するため、画像認識に用いられれる畳み込みニューラルネットワークを用いて改ざん領域の検出精度向上を行った。その手法及び実験結果について次章で述べる。

3　畳み込みニューラルネットワークを用いた JPEG 画像における改ざん領域の検出

3.1　概要

本章では、畳み込みニューラルネットワーク（Convolutional Nueral Networks, CNN）を用いることで、改ざんされた JPEG 画像における改ざん領域を検出する手法を述べる。これまでの研究結果から、JPEG 画像における改ざん検出には、DCT 係数の解析が有効であること、また、改ざんが疑われる画素同士の空間的連結性を考慮する必要があることが明らかとなった。そこで、DCT 係数を二次元画像として表示して CNN への入力画像とし、出力結果として改ざん領域が白、非改ざん領域が黒の二値化画像を出力する改ざん領域検出手法を提案する。本研究ではネットワークの最適化を行うために 2 つの予備実験を行う。予備実験 1 では n 層（$n = 3$〜6）の CNN を用いて入力の参照範囲を決定し、次に予備実験 2 で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う。最後に、予備実験 2 で決定したネットワークを用いた提案手法と従来手法 (He らの手法、MDBD 法) について、改ざん領域の検出結果を比較する。

提案手法では、F 値が He らの手法の 7.1 倍、MDBD 法の 2.3 倍を達成しており、改ざん領域の検出精度を飛躍的に向上させることに成功している。

3.2　背景および目的

前章では、改ざん画像であるか非改ざん画像であるかの判別において、MDBD 法が He らの従来手法よりも高い精度であることを示した。He らの手法や MDBD 法は、改ざん画像の判別と同時に改ざん領域の検出も行うことができる。しかしながら、He らの手法は非改ざん領域の誤検出が多く、一方で MDBD 法は改ざん領域の未検出が多いという欠点があり、改ざん領域の検出精度を向上させる必要がある（図 3.1）。改ざん領域についても適切に表示させることができれば、改ざん画像であると判別した結果の信頼度が高まり、逆に、改ざん領域を誤って表示していれば、判別結果の信頼度が低下するからである。本研究では、画像の一部をコピー＆ペーストにより改ざんした JPEG 画像を対象として、その改ざん領域を検出する精度の向上を目的としている。

3.3　提案手法の流れ

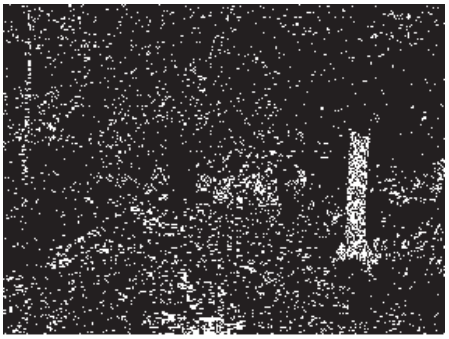
提案手法は、改ざん画像の DCT 係数を画像として二次元表示し、CNN に入力することで改ざん領域を検出するものである。まず、図 3.2 のように検査対象画像内の 8×8 画



(a) 改ざん画像



(b) 白色が改ざん領域を示す。



(c) He らの手法



(d) MDBD 法

図 3.1　従来手法の検出画像

素ごとに DCT 係数 64 成分を抽出し、各成分が 0 か 0 でないかに応じて画素値が 0 か 255 に 2 値化した 8×8 画素画像を作成する。この 2 値化画像は元の 8×8 画素と対応するため、画像全領域において同様の処理を行うことで原画像と同じサイズの 2 値化画像が入力画像として作成されることとなる。次に CNN の構造について図 3.3 を用いて説明する。第 2 章で述べた先行研究により、改ざん JPEG 画像の検出には 8×8 画素ブロック単位の解析が有効であること、改ざんの疑いがあるブロックの x、y 方向の空間的連結性を考慮することが重要であることが明らかとなっていた。そこで、1 層目のフィルタサイズを 24 とすることで、3×3 個のブロックの連結性を考慮した解析を行う。また、1 層目のストライドを 8 とすることで、ブロックの境界とフィルタの開始位置を同期させる。これにより、8×8 の周期で現れる DCT 係数の特徴を効率よく学習できると考える。2 層目から出力層まではストライドを 1 とし、プーリングは行わない。これにより、出力画像のサイズは入力画像の 8 分の 1 となる。出力画像の各画素値はブロック単位の改ざんの有無を表す。

それぞれの特徴マップは、

$$F_1(Y) = max(0, Y * W_1 + B_1) \tag{3.1}$$

$$F_2(Y) = max(0, F_1 * W_2 + B_2) \tag{3.2}$$

$$F_n(Y) = F_{n-1} * W_n + B_n \tag{3.3}$$

$$F(Y) = \begin{cases} 1 & (F_n \geqq 0.5) \\ 0 & (F_n < 0.5) \end{cases} \tag{3.4}$$

で求められる。ここで、Y は入力画像、W_n と B_n は n 層目のフィルタ及びバイアスであり、$F(Y)$ が出力画像である。

一般に CNN はフィルタサイズが大きいほど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、予備実験 1 では n 層（$n = 3$〜6）の CNN を用いて入力の参照範囲を決定し、次に予備実験 2 で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う。

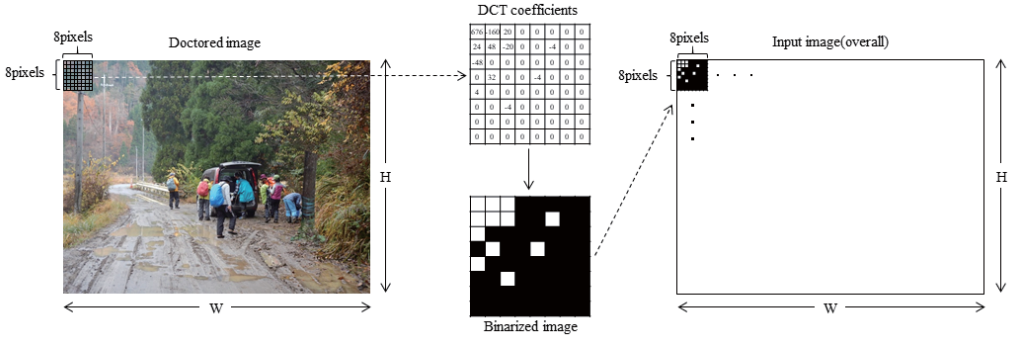


図 3.2 CNN に入力する 2 値化画像

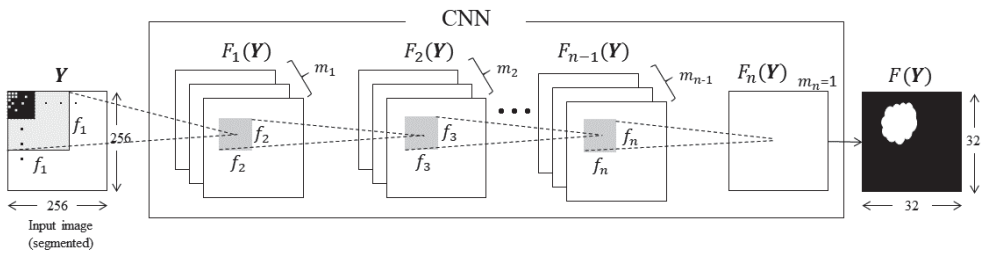


図 3.3 提案手法の CNN 構造

3.3.1　改ざん画像および教師画像の準備

実験には一般的なデジタルカメラ 5 機種 (CASIO : EX-ZR200,SONY : NEX-3N, Canon : IXY510, OLYMPUS : FE220,NIKON : AW210) で撮影した 50 枚の JPEG 画像を使用する。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。改ざん画像は、撮影した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た後 (コピー&ペースト)、JPEG 形式で保存することで作成する。本実験では、異なる画像の一部をコピーし対象画像にペーストすることで改ざん画像を作成するが、これは同一画像内のコピー&ペーストによる改ざん画像においても解析を行う上で相違点はない。使用したアプリケーションは Photoshop CS であり、画像保存時の設定は最高画質 (最低圧縮) である。また、教師画像には図 3.1(b) のように改ざん領域の画素値を 255(白)、非改ざん領域を 0(黒) とした 2 値化画像を使用する。

対象画像はデジタルカメラで撮影した 4M〜16M 画素であり、作成される入力画像も 4M〜16M 画素の非常に大きいサイズとなる。そこで、予備実験 1 では各画像を 256×256 画素に分割したものを入力画像としており、改ざん画像 50 枚から得られる入力画像は 6489 画像となる。作成した改ざん画像 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、最もバリデーション精度の高かったモデルを用いてテストを行う。1 層目でストライドを 8 とするため教師画像は入力画像の 8 分の 1 の 32×32 画素となる。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、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。

3.3.2　入力参照範囲の決定

層数 n における 5 枚 (427 画像) のテスト画像に対する検出結果の平均値を表 3.1 に示す。ここで、P、R、F は式 2.11〜2.13 と同じく下記で定義されるが、T_P は検出した改ざん画素数、F_N は未検出の改ざん画素数、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点が異なる。

$$P = \frac{T_P}{T_P + F_P} \tag{3.5}$$

$$R = \frac{T_P}{T_P + F_N} \tag{3.6}$$

$$F = \frac{2 \cdot P \cdot R}{P + R} \tag{3.7}$$

予備実験 1 では層数 $n = 6$ で F が概ね高止まりする結果が得られた。層数 6 の時のネットワーク構造から出力 1 画素あたりに対する参照範囲を計算する。まず 1 層目では x、y 方向とも入力画像がそのままのサイズで入力されるため 24 画素が必要となる。次に、n 層目 $(n = 2, \cdots, 6)$ では 1/8 サイズの出力マップに対して f_n=3 のフィルタを使用することから、入力サイズに換算すると x、y 方向に 8 倍した 8 画素ずつが拡張される。よって、全体での参照範囲は 24+8×5×2=104 画素が必要となる（図 3.4）。

表 3.1　予備実験 1 の結果

n	$\{f_1,\cdots,f_n\}$	$\{m_1,\cdots,m_n\}$	P	R	F
3	{24,3,3}	{32,16,1}	0.77	0.58	0.66
4	{24,3,3,3}	{32,16,16,1}	0.74	0.72	0.73
5	{24,3,3,3,3}	{32,32,16,16,1}	0.75	0.79	0.76
6	**{24,3,3,3,3,3}**	**{32,32,16,16,16,1}**	**0.76**	**0.80**	**0.77**
7	{24,3,3,3,3,3,3}	{32,32,16,16,16,16,1}	0.76	0.82	0.79

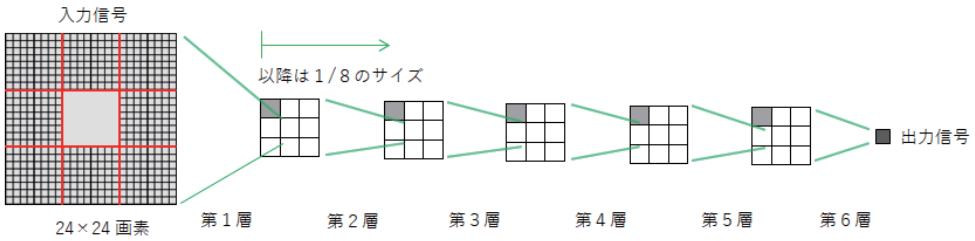


図 3.4　参照範囲（層数 6 ）

3.3.3　ネットワーク構造の決定

入力の参照範囲が 104×104 画素となる条件の下で、フィルタ数とマップ数を変化させネットワーク構造の比較検討を行う。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。そのため予備実験 2 ではゼロ

パディングを行わないこととする。この場合、256×256 画素の画像に対しては参照領域が x、y 方向で 48 画素ずつはみ出すため、入力画像のサイズを 352×352 画素に変更する (図 3.5)。教師画像は予備実験 1 と同じく 32×32 画素である。ネットワーク層 n、フィルタサイズ f_n、マップ数 m_n 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検出精度を測定した結果を図 3.6 及び表 3.2 に示す。予備実験 1 と同様に f_1=24、m_n=1 であり、1 層目のストライドのみ 8 である。図 3.6 及び表 3.2 に示す値は、5 枚 (427 画像) のテスト画像に対する検出結果の平均値である。　ただし、改ざんブロックを全く検出しない画像があった場合、T_P=0、F_P=0 となり P 及び F が算出できないため－と表記している。最も F の高いネットワーク構造は $(n, \{f_n\}, \{m_n\})$=(5,{24,5,3,3,3},{128,128,64,32,1}) であり、P=0.95、R=0.81、F=0.86 である。よって、このネットワーク構造を用いて次節で従来手法と比較する。同ネットワークのエポック数におけるバリデーション精度の推移を図 3.7 に示す。

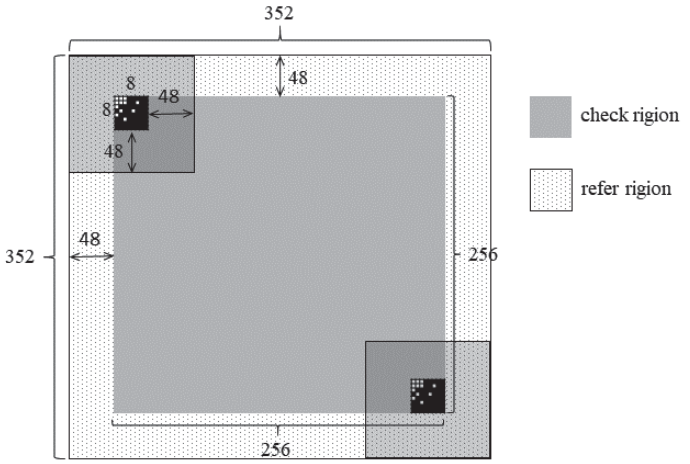


図 3.5　参照範囲を考慮した入力画像

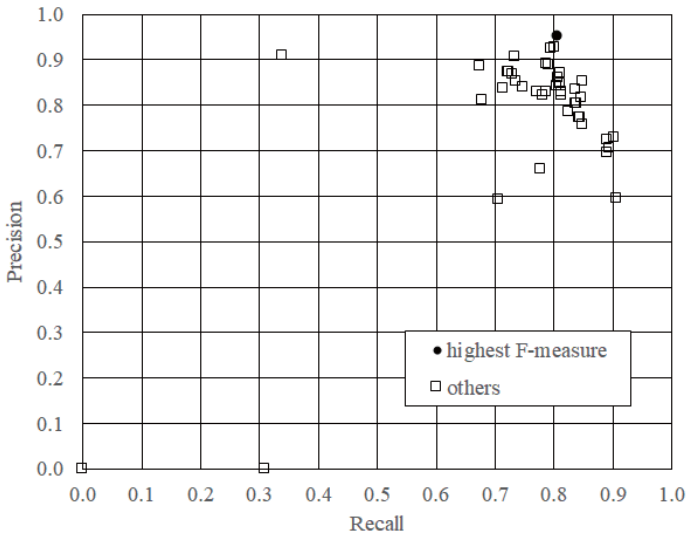


図 3.6 予備実験 2 の結果

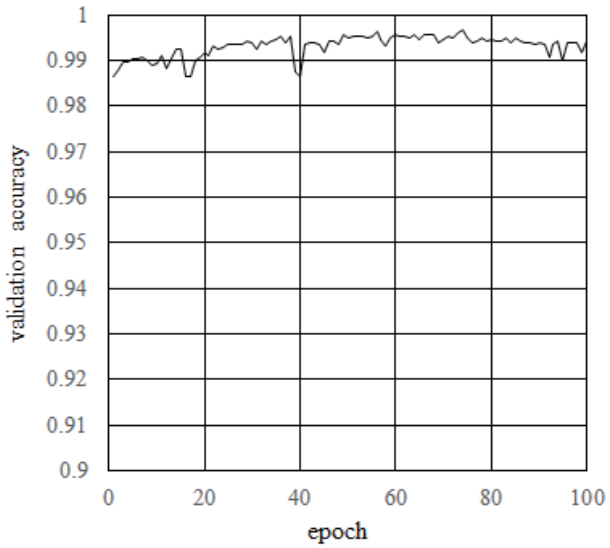


図 3.7 バリデーション精度

表 3.2 CNN モデルとその精度

n	filter size $\{f_1,\cdots,f_n\}$	map $\{m_1,\cdots,m_n\}$	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

表 3.2 CNN モデルとその精度（続き）

n	filter size $\{f_1,\cdots,f_n\}$	map $\{m_1,\cdots,m_n\}$	P	R	F
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

3.4　実験方法

本実験は、改ざん領域を検出する精度の比較が目的とするため改ざん画像のみを使用し、改ざん方法はコピー&ペーストのみである。改ざんした 50 画像を 10 グループに分割し、前節で導出した 5 層構造のネットワーク $(n, \{f_n\}, \{m_n\})$=(5,{24,5,3,3,3},{128,128,64,32,1}) を用いてクロスバリデーション行う。すべての組み合わせにおいて 40 画像を学習、5 画像をバリデーション、5 画像をテストに用いる。

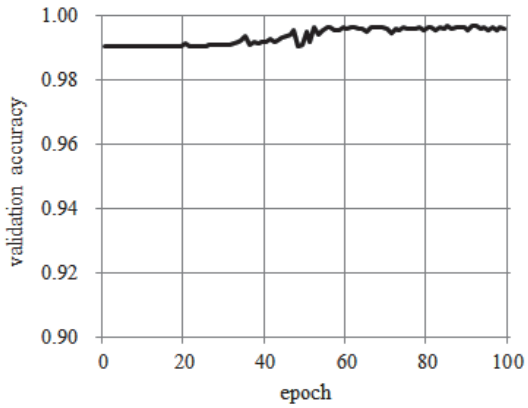
提案手法では、学習を安定化させるため各層間でドロップアウト (0.5) を行う。エポック数は 100 回、トレーニング画像数は 4821〜5635 であり、バックプロパゲーションは約 48 万〜56 万回となる。

He らの手法や MDBD 法による検出結果と精度比較を行う。両手法には、特徴量の抽出部において最適化すべきパラメータは存在しない。また、改ざんの有無はそれらの特徴量に基づいて SVM によって判断されるため、人手によって調整する閾値はない。

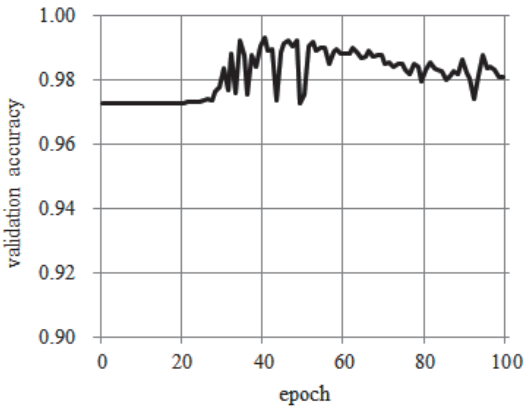
3.5　実験結果

実験の結果、提案手法では 10 グループすべてでバリデーション精度が 0.98 を超えている。各グループのエポック数におけるバリデーション精度の推移を図 3.8 に、10 グループの平均値を図 3.9 に示す。学習画像とテスト画像の組み合わせを変えても安定した精度が得られることから、同種の改ざん画像に対して汎用性が期待できる。

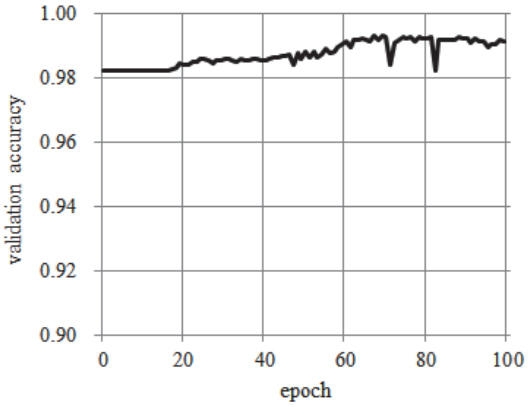
各グループごとにバリデーション精度が最大となるモデルを用いてテストを行い、提案手法及び各従来手法の検出結果を図 3.10 及び表 3.3 に、それらの平均値を図 3.11 に、また検出例を図 3.12 に示す。



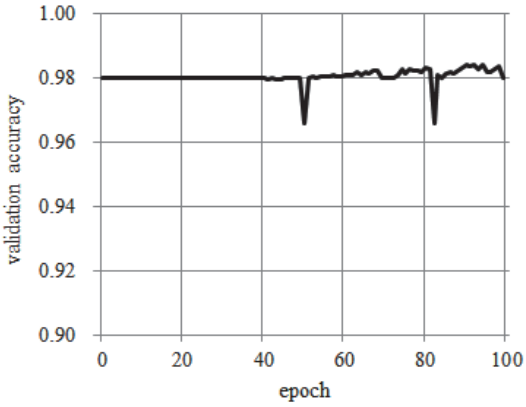
(a) group No.1



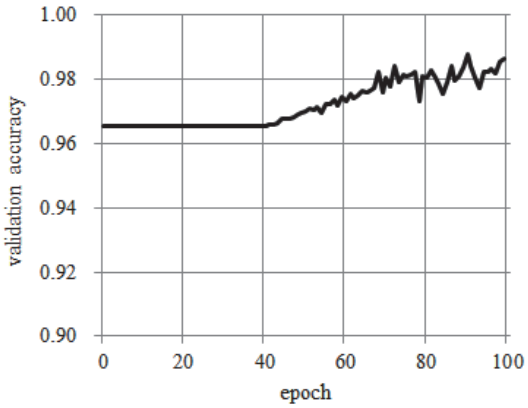
(b) group No.2



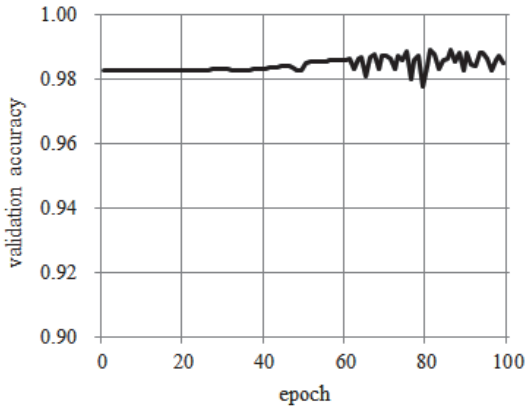
(c) group No.3



(d) group No.4

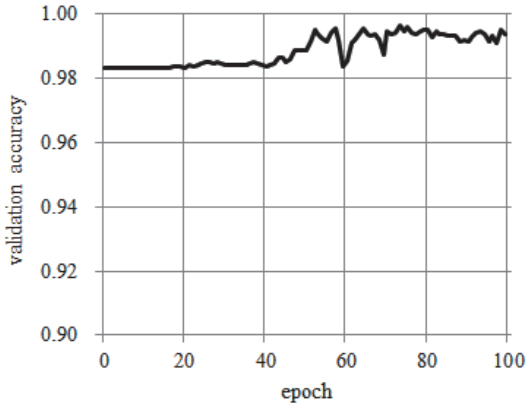


(e) group No.5

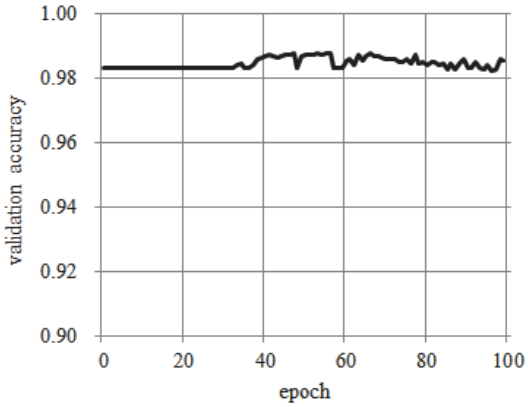


(f) group No.6

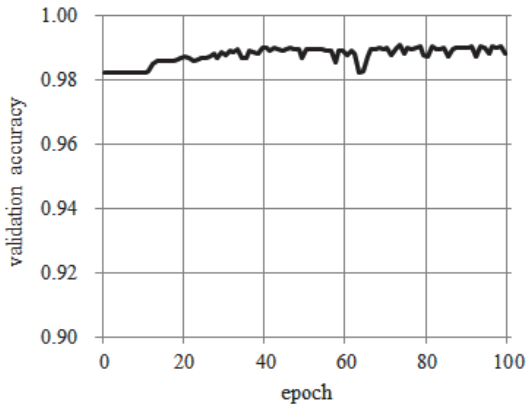
図 3.8 各グループのバリデーション精度（No.1〜6）



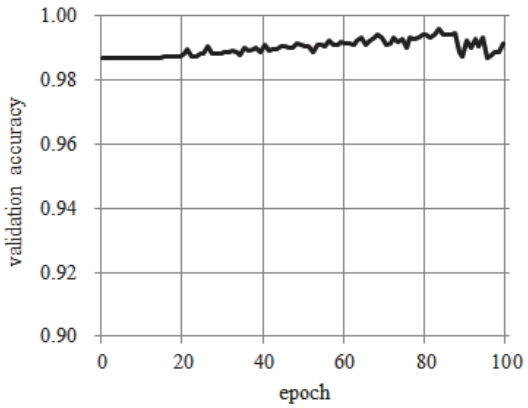
(g) group No.7



(h) group No.8



(i) group No.9



(j) group No.10

図 3.8　各グループのバリデーション精度（No.7〜10）

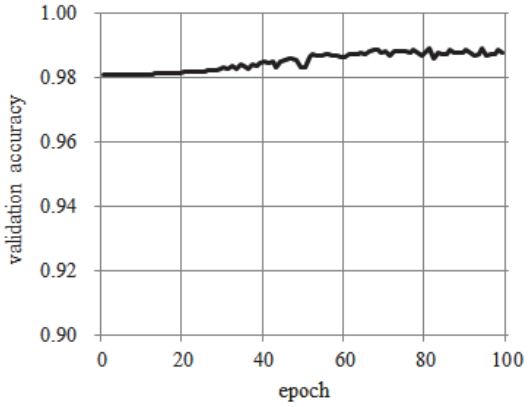


図 3.9　バリデーション精度 (10 グループの平均値)

表 3.3 提案手法及び各従来手法の検出結果 (No.1〜30)

	proposed method			He's method			MDBD method		
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

表 3.3 提案手法及び各従来手法の検出結果 (No.31〜50)

	proposed method			He's method			MDBD method		
No	P	R	F	P	R	F	P	R	F
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

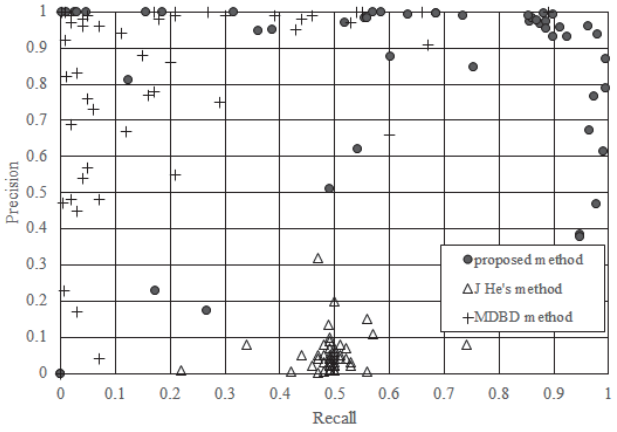


図 3.10 提案手法及び各従来手法の検出結果 (Precision-Recall)

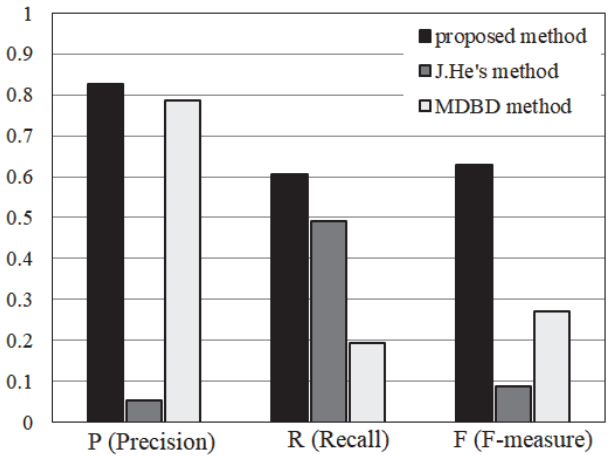
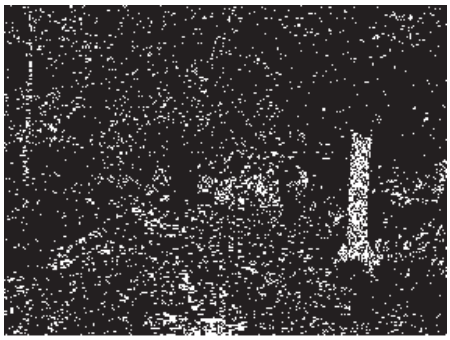


図 3.11 提案手法及び各従来手法の検出結果 (50 画像の平均値)



(a) 改ざん画像 (No.10)



(b) He らの手法



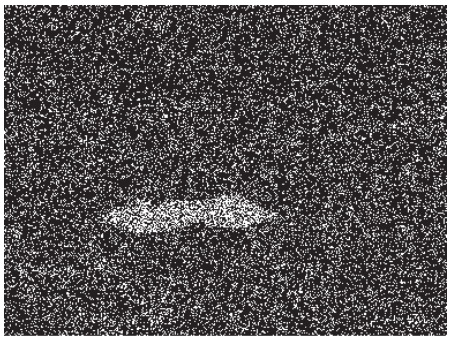
(c) MDBD 法



(d) 提案手法



(e) 改ざん画像 (No.16)



(f) He らの手法



(g) MDBD 法

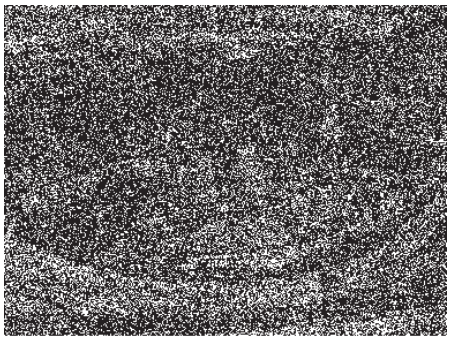


(h) 提案手法

図 3.12　検出結果画像 (No.10, 16)



(i) 改ざん画像 (No.17)



(j) He らの手法



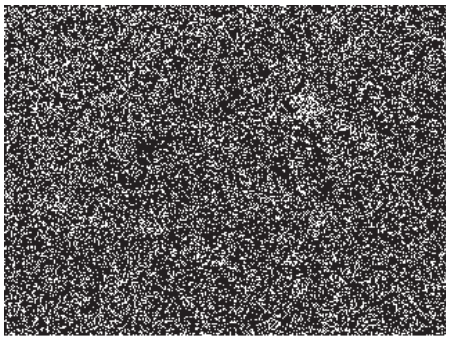
(k) MDBD 法



(l) 提案手法



(m) 改ざん画像 (No.36)



(n) He らの手法



(o) MDBD 法



(p) 提案手法

図 3.12　検出結果画像 (No.17, 36)

3.6　考察

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改ざん部の検出が向上していることがわかる。このことは図 3.12 からも見て取れ、従来手法と比較して良好な検出結果画像を出力している (No.10, 16,17)。No.36 の画像は右上に黄色い花一輪を付け足したものであるが、MDBD 法では改ざん部の一部を検出できていたにもかかわらず提案手法では検出できなかった。この原因についてはさらに検証を行う必要がある。

3.7　まとめ

本研究では、CNN を用いて改ざんされた JPEG 画像における改ざん領域の検出を行った。He らの手法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増加した。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 He らの手法の 7.1 倍、MDBD 法の 2.3 倍を達成した。本研究で提案したネットワーク構造は、コピー&ペーストによる改ざん画像の検出において学習グループを変えてもバリデーション精度 0.98 以上を維持したことから、同種の改ざん画像には汎用的に対応可能であると考えられる。しかしながら、被写体の色の変更など、本実験と異なる改ざん手法で作成した改ざん画像については、学習サンプルの追加とネットワークの再調整が必要である。また、解像度の低い画像や圧縮率の高い画像は改ざんの痕跡が残りづらくなることから、改ざん検出精度は下がる可能性がある。さらなる検出精度の向上には、様々な改ざん画像に応じたネットワーク構造の最適化が必要である。

4　LSTM 及び U-Net を用いた時空間解析による動画改ざん領域の検出

4.1　概要

本章では、ディジタル動画の法科学的検査を目的として、あるシーンに写っている被写体が改ざん処理により消去された場合において、改ざん領域を検出する手法を述べる。近年は機械学習の 1 つであるディープラーニング（深層学習）が注目を浴びており、自然言語処理においては時系列データに特化した RNN(再帰型ニューラルネットワーク)、画像認識においては空間的特徴量を抽出する CNN(畳み込みニューラルネットワーク) が用いられている。そこで、動画が静止画の時間的連続であることに着目し、RNN と CNN を併用して時間と空間の両面から解析を行う。時間方向には LSTM(Long short-term memory、長・短期記憶) を使用し、検査対象画素に前後 15 フレームを含めた 31 個のデータを利用して解析を行う。空間方向には U-Net 構造の CNN を使用し、改ざん領域の全体的な位置情報と局所的特徴を合わせた解析を行う。評価実験として 4 通りの検出手法を比較する。1. 各フレームをそのまま入力画像とする、2. 絶対値差分画像を用いる、3. 絶対値差分画像を LSTM の入力画像とする、4.LSTM の出力結果を入力画像として U-Net 解析を行う (提案手法) である。提案手法では、改ざん領域が画像内を空間方向に移動・変形した動画であっても F 値 0.98 を達成しており、動画改ざん領域の検出手法として有用であることを示している。

4.2　背景

スマートフォンやデジタルビデオカメラ、防犯カメラなどの技術進歩とそれに伴う低価格化により、私たちの身の周りには多くの動画撮影デバイスがある。それらの使い方は非常に簡単になってきており、誰でも、いつでも、どこにいてもすぐに動画を撮影することができる。ポケットからスマートフォンやデジカメを取り出し、写真や動画を撮影することは今では日常生活の一部となっており、これらを SNS などで他人と共有することも老若男女を問わず全世界で日々行われている。

ハードウェアの普及や進化はそれを扱うソフトウェアの成長も促しており、デジタル画像を処理・編集するソフトウェアも数多く販売されている。インターネットから無料でダウンロードできるものも多い。これらはパソコンやスマートフォンで簡単に処理できるように設計されており、画像処理に精通していない人であっても直感的に操作できるものも少なくない。

また、Youtube などでは真実か偽物なのか見た目では区別のつかない動画も多数アッ

プロードされおり、多くの人がその動画を見て楽しむことができる。これは、手品やイリュージョンショーのように娯楽目的であれば科学技術の進歩を享受できる最新映像と受け取ることができる。しかしながら、著名人のインタビューを加工して、話してもいない内容をあたかも当人が話しているかのように見せかける内容の動画も存在しており、このような動画が流布してしまうと大きな社会問題と成りうることが予想される。

一方で、デジタル写真や動画は被写体の状態をありのままに撮影し記録するため、犯罪捜査や事故状況の確認などの警察活動において非常に有用な客観証拠となりうる。記録された内容を適切に扱うためには、真実をありのまま記録しているかどうか、作為的な処理が加えられていないかを確認する作業が必要不可欠である。改ざん処理が施された動画に対しては、真実であるのか偽物であるかを判断するための高度な画像解析の知識や技術が必要であり、日々研究開発が行われている。

4.3 従来手法と問題点

W.Wang ら [22] は、MPEG 形式の動画を対象とした改ざん検出手法を提案した。エンコード時のデータ圧縮過程において MPEG 動画は JPEG 画像と類似した特徴を持っているため、JPEG 改ざん検出で効果的である Double JPEG 解析を応用したものである。MPEG 動画には I,P,B の 3 種類のフレームが存在するが、そのうちの I フレームに注目し Double JPEG を解析し改ざん領域を検出するものである。同手法を用いるためには同一フレームが I フレームとして再圧縮されなくてはいけないため全フレームを I フレームでエンコードしている。しかしながら、動画形式は MPEG 形式以外にも多数存在しており、また、時間方向のトリミングを行えば同一フレームが I フレームとして再圧縮されることはない。これでは検査対象となる動画が限定的であり汎用性が低いと言える。

J.Xu ら [26] や M.A.Bagiwa[27] ら、D.Dario ら [28] は、クロマキー撮影により合成した動画を対象とする改ざん検出手法を提案した。クロマキー撮影とは、背景にブルーやグリーンのスクリーンを設置して被写体を撮影し、背景となる画像にはめ込む技術であり、テレビのバラエティー番組や CM 撮影などで用いられている。しかしながら、クロマキー撮影をするためには大掛かりな撮影機材やソフトウェアが必要であり、街中を移動する人物や車両などの撮影範囲の大きな動画に対して行うことはできない。また、個人が所有する一般的な PC で簡単に作成することが困難である。これらは動画改ざんではなく CG 処理ともいえる。

P.Bestagini ら [29] は、背景画像を用いて被写体を消去した改ざん動画の検出手法を提案した。背景となるフレームを時間的に連続して貼り付けて被写体を消去する手法は誰もが思いつくものの一つである。本来写っている人物や車両を消して、あたかも当時はそこに存在しなかったかのような動画を作成することができる。しかしながら、同手法では改ざん領域が空間的に移動や変形をすることを想定していない。一般に、動画に写

る被写体領域は空間的に移動・変形しており、それに合わせて改ざん領域も移動・変形することが予想される。より実用的なシチュエーションを考えるのであれば、改ざん領域の移動・変形にも対応しなければならない。

4.4 研究の目的

本研究は、個人レベルの機材を用いて比較的容易に作成できる改ざん動画に対して、その改ざん領域を検出することを目的としている。

そこで、動画内に写っている被写体 (人物や車両) が「消去されている」改ざん動画を対象とする。動画に不都合なシーンがあるときにそれを隠蔽するという状況を想定している。また、被写体は空間的に画面上を移動しており、その移動に合わせて改ざん領域も移動かつ変形を行う。被写体が写っていないシーンの領域の一部を背景画像としてコピーし、被写体が写っているシーンの被写体領域にペーストすることで作成する。このような処理は、一般的な PC やソフトウェアを用いることで実行可能である。作成した改ざん動画は一般的な形式 (H.264/AVC 形式) にエンコードされるが、この際にフレーム位置の指定は行わない。つまり、同一フレームを I フレームとして再エンコードするという制限を加えない。

見た目で判別できない程度の改ざん動画に対して、動画形式や改ざん領域の限定を減らし、より汎用的で実用的な検出手法を目指すものである。

4.5 検査対象とする改ざん動画

本研究で検査対象とする動画は、背景画像の一部の領域をコピー＆ペーストすることにより、本来写っている被写体を “消去した”改ざん動画である。ここで、改ざん領域は時間経過により空間方向へ変形及び移動することがあり、時間方向では途切れることなく連続している。

改ざん動画の一例を説明する。図 4.1 の原動画には階段を歩く 2 人が写っており、右側の人物は階段を下り、左側の人物は階段を上がっている。図 4.2 では背景画像の領域の一部をコピー＆ペーストして、この 2 人のうち左側の人物を消去している。図 4.3(a) は、背景画像からコピー＆ペーストする領域を 2 通りに分けて改ざん動画を作成した場合の改ざん領域を、時空間的に 3 次元で表示したものである。最も大きな直方体が改ざん動画全体を示しており、その中にある時間方向に連続した形の異なる直方体が改ざん領域を示している。図 4.3(b) は、各フレームを改ざん領域を白 (255)、非改ざん領域を黒 (0) で表したものを抜粋して示している。後述する実験では、この 2 値化した画像が各フレームの正解画像または教師画像 (ground truth) となる。

前節で述べたとおり、改ざん動画の検出に関する研究はクロマキーを用いた画像合成

に関するものが多く、被写体を消去した事例は少ない。また、その消去領域が時間によって空間的に変化したものに対する研究事例は見当たらない。本提案手法は、これまで研究発表されていなかった改ざん方法を用いた動画を検査対象としており、さらにニューラルネットワークを用いることで高い検出精度が得られたものである。

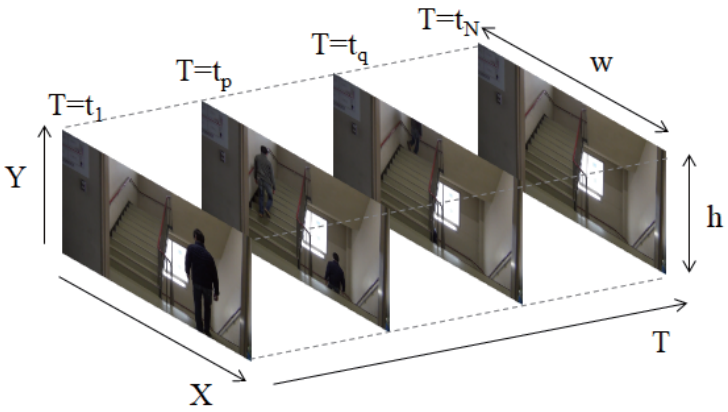


図 4.1 オリジナル動画

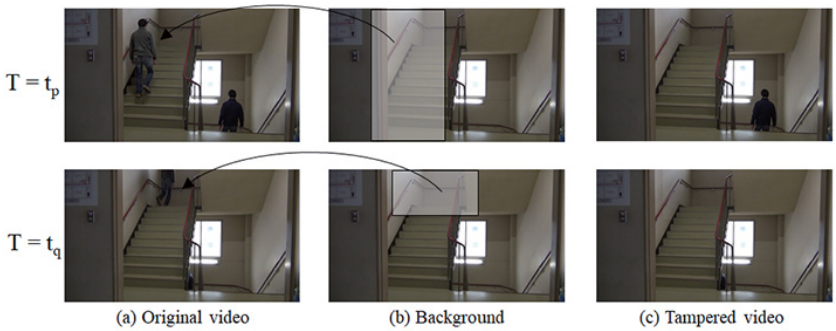
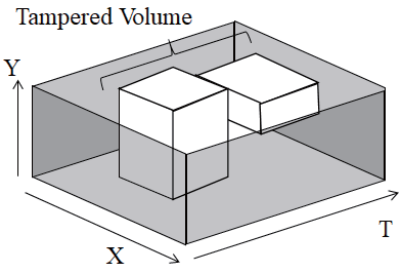
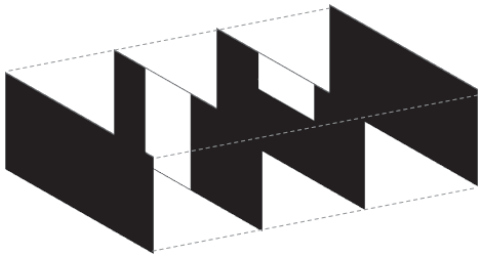


図 4.2 改ざん方法



(a) 改ざん領域の三次元表示



(b) 改ざん領域をフレームで図示したもの

図 4.3 改ざん領域の例

4.6 提案手法の流れ

4.6.1 LSTM による時間方向の解析

本提案手法は 2 つの解析に分かれており、1 つは LSTM を用いた時間方向の解析を行う。LSTM は RNN の中で使われる機構の一つであり、自然言語や時系列データなど連続性のあるデータを取り扱うことができる。動画には人物や構造物等の被写体が時々刻々と移動したり変化しながら記録されており、被写体と共に人が知覚できない細かなゆらぎ (ノイズ) も記録されていることが多く、これらノイズも時間的変動があると考えられる。そこで、ターゲットとなるフレーム及び前後 15 フレームのデータを LSTM に入力することにより時間方向の解析を行う。図 4.4 により、対象動画が横 X 画素、縦 Y 画素、N フレームの場合の解析方法を説明する。まず、時間的連続性の特徴に着目するため、(X,Y) サイズのフレーム間差分画像 $f_t = |F_{t+1} - F_t|(t : 1, \cdots, N-1)$ を作成する。次に、ターゲットとなる差分画像フレームを基準に前後 15 フレームを含めた 31 画素を LSTM に入力し、出力結果から各フレームの改ざん確率マップ $P_t(x,y)$ を作成する。$\mathbf{x}(t)$ をフレーム間差分画像 f_t の (x, y) 座標の画素値とすると、入力信号列は $\mathbf{X} = \{\mathbf{x}(t-15), \mathbf{x}(t-14), \cdots, \mathbf{x}(t), \cdots, \mathbf{x}(t+15)\}$ となり、この時の出力ベクトル $\mathbf{h}(t)$ は以下のように表される。

$$\mathbf{i}_t = \sigma(\mathbf{W}_i\mathbf{x}(t) + \mathbf{U}_i\mathbf{h}(t-1) + \mathbf{V}_i\mathbf{c}(t-1) + \mathbf{b}_i) \tag{4.1}$$

$$\mathbf{f}_t = \sigma(\mathbf{W}_f\mathbf{x}(t) + \mathbf{U}_f\mathbf{h}(t-1) + \mathbf{V}_f\mathbf{c}(t-1) + \mathbf{b}_f) \tag{4.2}$$

$$\mathbf{o}_t = \sigma(\mathbf{W}_o\mathbf{x}(t) + \mathbf{U}_o\mathbf{h}(t-1) + \mathbf{V}_o\mathbf{c}(t) + \mathbf{b}_o) \tag{4.3}$$

$$\mathbf{a}_t = \tanh(\mathbf{W}_a\mathbf{x}(t) + \mathbf{U}_a\mathbf{h}(t-1) + \mathbf{b}_a) \tag{4.4}$$

$$\mathbf{c}_t = \mathbf{i}_t \circ \mathbf{a}_t + \mathbf{f}_t \circ \mathbf{c}_{(t-1)} \tag{4.5}$$

$$\mathbf{h}_t = \mathbf{o}_t \circ \tanh(\mathbf{c}_t) \tag{4.6}$$

ここで、$\mathbf{i}_t$ は入力ゲート、$\mathbf{f}_t$ は忘却ゲート、$\mathbf{o}_t$ は出力ゲートであり、$\mathbf{c}_t$ は CEC(constant error carousel)、σ はシグモイド関数である。また、$\mathbf{W}$ 及び $\mathbf{U}$ は重み、$\mathbf{V}$ は覗き穴結合、$\mathbf{b}$ はバイアスである。$\mathbf{x}(t+15)$ までの入力に対する出力が時刻 t の改ざん確率となるため、$\mathbf{h}(t+15) = P_t(x, y)$ となる。対象フレームに対して前後 15 フレームを必要とするため、出力される P_t は f_{16}〜f_{N-15} に対応する $N-31$ フレームとなる。学習の最適化には Adam を使用し、バッチサイズ 500 とした。なお、$f_t(x, y)$ は 0〜255 までの値をとるため、これを 0〜1 へスケール変換した値が実際の $\mathbf{x}(t)$ となり、出力される $\mathbf{h}(t)$ を 0〜255 へスケール変換したものを出力画像として表示する。次節の U-Net においても同様のスケール変換を行っている。

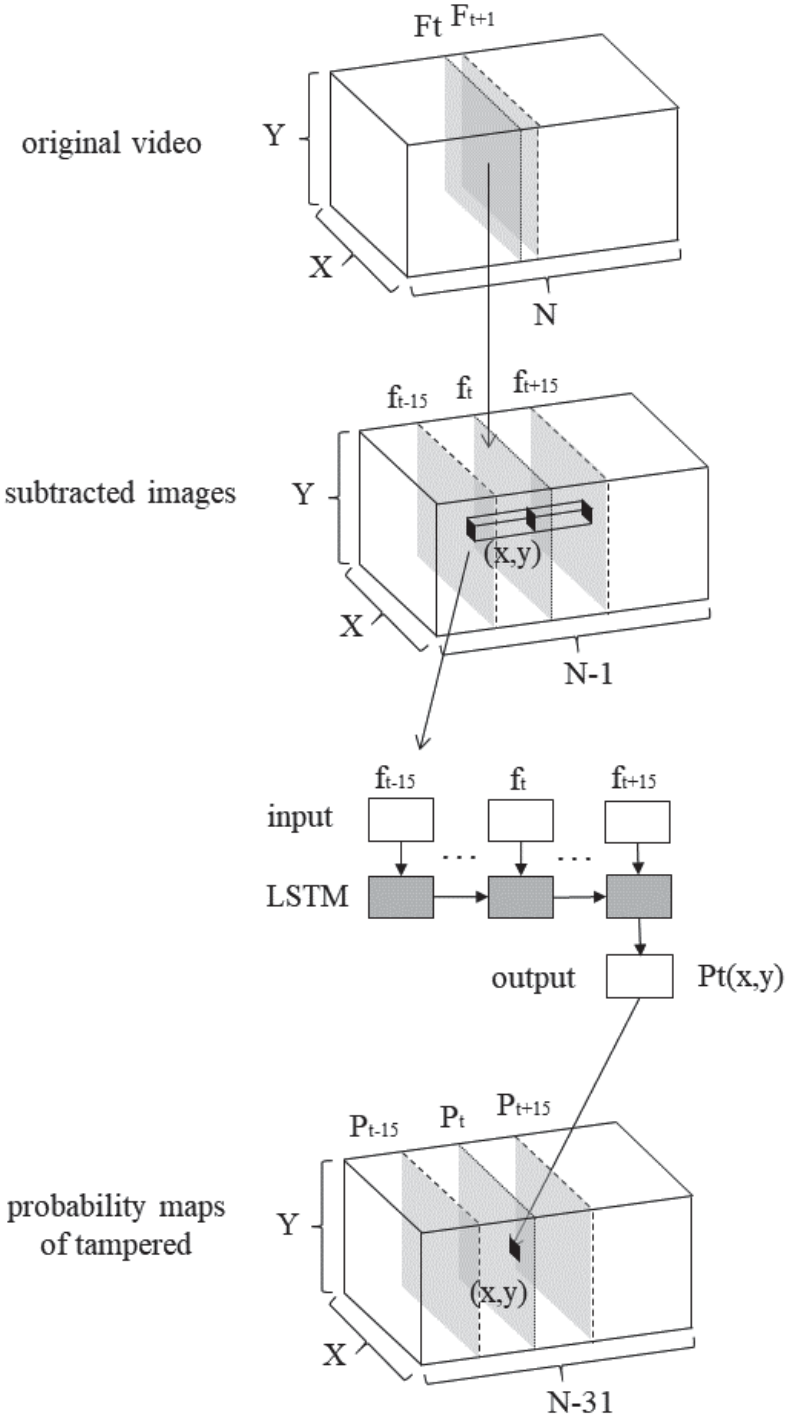


図 4.4 LSTM を用いた解析の流れ

4.6.2 U-Net による空間方向の解析

次に、前項で得られた $P_t(x, y)$ を U-Net(U 字型畳み込みニューラルネットワーク) に入力することで空間方向の解析を行う。U-Net は途中のプーリングで畳込み層の画素数を減らすことにより、入力データの抽象化を行い全体的な位置情報を学習する。一方で、プーリング前の画素数の多い層には局所的特徴が学習されている。アップサンプリングして 2 つの層の大きさを合わせ結合 (merge) することにより、局所的特徴と全体的な位置情報を学習することができる。本研究で扱う改ざん手法では改ざん領域が画像上で局所的に集中し、かつ、その位置が重要であるため U-Net が効果的である。LSTM のみでは非改ざん画素の誤検出が生じるため、U-Net を通すことによりこれらをノイズとして削減することができる。本研究では図 4.5 に示す 19 層の U-Net 構造を用いた。なお、計算機の性能上、画像を 288×288 画素に縮小して入力する。学習の最適化には Adam を使用し、バッチサイズ 16 とした。

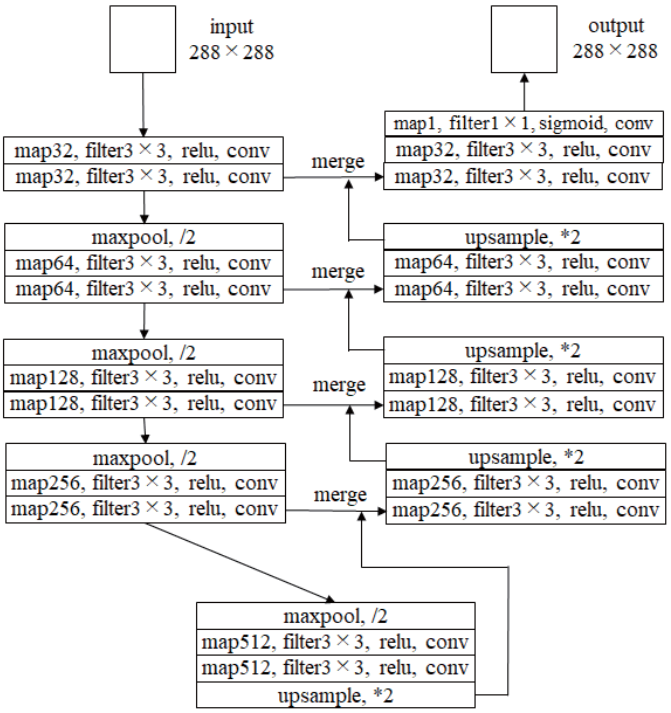


図 4.5 U-Net 構造の CNN ネットワーク

4.7　改ざん動画の作成

被写体を消去した改ざん動画を対象とする改ざん検出の研究例は少なく、正解画像 (ground truth) 付き改ざん動画の研究用データセットが公開されていない。そのため、次の手順で改ざん動画を作成する。まず、スマートフォン (NTT:MO-01J)、デジタルビデオカメラ (SONY:HDR-CX180) 及びデジタルカメラ (CASIO:EX-PRO100) を三脚に固定し、人物や車両が通過する状況を撮影して原動画とする。次に、全フレームを bmp 形式の静止画に分割し、消去対象の被写体が写っていない背景画像を適当に選定して、領域の一部をコピー＆ペーストすることで被写体を消去した改ざん画像を作成する。これらのフレームを動画変換ソフト (TMPGEnc Video Mastering Works 6) を用いて動画にエンコード (H.264/AVC 形式) して改ざん動画とする。各動画は 30fps で長さが 5〜35 秒であり、フレーム総数は 5490、改ざんフレームは 4118、非改ざんフレームは 1372 である。改ざん動画の一覧を図 4.6 に示す。左図がオリジナル動画であり、右図が改ざん動画を示す。改ざんにより消去する人物や車両を左図の矢印で示す。この被写体を含む領域を矩形で消去している。

No.1,2,7 は 640 × 480 画素、その他の動画は 1920 × 1080 画素である。No.1〜No.6 を学習に使用し、No.7〜No.12 を推論に使用する。No.9 は暗い通路の撮影動画であり被写体を視認できるように明るさを補正した図も示すが、実験では補正していない動画を使用している。



(a) 動画 No.1 (オリジナル)



(b) 動画 No.1 (改ざん)



(c) 動画 No.2 (オリジナル)



(d) 動画 No.2 (改ざん)



(e) 動画 No.3 (オリジナル)



(f) 動画 No.3 (改ざん)

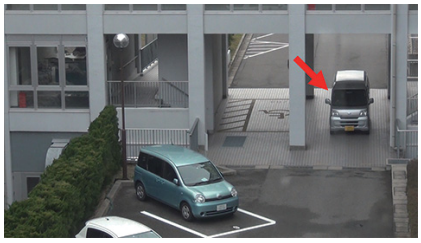


(g) 動画 No.4 (オリジナル)



(h) 動画 No.4 (改ざん)

図 4.6 改ざん動画 (No.1〜4)



(i) 動画 No.5 (オリジナル)



(j) 動画 No.5 (改ざん)



(k) 動画 No.6 (オリジナル)



(l) 動画 No.6 (改ざん)



(m) 動画 No.7 (オリジナル)



(n) 動画 No.7 (改ざん)



(o) 動画 No.8 (オリジナル)



(p) 動画 No.8 (改ざん)

図 4.6 改ざん動画 (No.5〜8)



(q) 動画 No.9 (オリジナル)



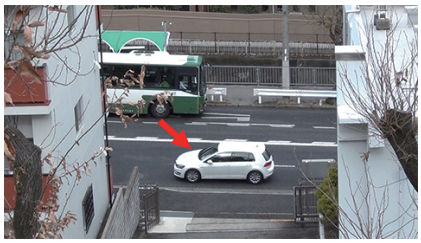
(r) 動画 No.9 (改ざん)



(s) 動画 No.9 (オリジナル、明るさ補正)



(t) 動画 No.9 (改ざん、明るさ補正)



(u) 動画 No.10 (オリジナル)



(v) 学習動画 No.10 (改ざん)



(w) 動画 No.11 (オリジナル)



(x) 動画 No.11 (改ざん)



(y) 動画 No.12 (オリジナル)



(z) 動画 No.12 (改ざん)

図 4.6 改ざん動画 (No.9〜12)

4.8　実験方法

本研究では、TypeA〜D の 4 通りの実験を行い推論精度を比較した (図 4.7)。

TypeA では、検出のための事前解析は何も行わず、表示された画像から改ざん領域が検出可能であるかを検証する。したがって、改ざん動画を 1 フレームずつに分割し、カラー画像のまま U-Net に入力する。計算機の性能上、オリジナル画像のサイズのまま入力することができないため、288×288 画素のサイズに縮小処理を行った画像を入力する。U-Net からの出力画像は 0〜1 までの値を有するため、閾値を 0.5 に設定して改ざん領域を白、非改ざん領域を黒で表示した 2 値化画像を最終出力画像とする。画像サイズは 288×288 画素である。用いるネットワークの構造は図 4.5 である。

TypeB では、連続するフレーム間の差分処理のみを行うことで改ざん領域が検出可能であるかを検証する。実験に用いる改ざん手法は、被写体を消去するために背景画像をペーストするが、数フレーム間は同一の背景画像を使用している。そのため、エンコードによる影響を無視すれば、改ざん領域は差分処理により画素値が 0 になることが予想される。差分処理画像を閾値輝度 Vt=1 で 2 値化処理を行い、改ざん領域は黒、非改ざん領域は白の画像を作成する。閾値の妥当性は後述する実験により検討を行う。2 値化画像を階調反転し、最終出力画像は、改ざん領域が白、非改ざん領域が黒の 2 値化画像である。画像サイズはオリジナル画像と同一である。

TypeC では、TypeB で用いたフレーム間差分画像を入力として、さらに時間解析を加えるため LSTM を用いる。改ざん領域は時間的に連続しているため、フレーム間差分の時間的なつながりを解析に加えることで、検出精度を高めることを目的としている。これにより、あるフレーム間だけで偶然に差分が 0 となる画素を改ざん領域と誤検出してしまうことを低減することが期待される。LSTM では出力値が 0〜1 の値を持つため、閾値を 0.5 に設定して改ざん領域を白、非改ざん領域を黒で表示した 2 値化画像を最終出力画像とする。閾値の妥当性は後述する実験により検討を行う。画像サイズはオリジナル画像と同一である。

TypeD では、TypeC で得られた LSTM による出力結果画像を、図 4.5 に示す U-Net に入力する。改ざん領域は空間的に連結しているため、さらに改ざん領域を空間的に特定することができ、検出精度の向上が期待される。U-Net に入力する画像は LSTM の出力画像であるため 2 値化されておらず、0〜255 の値をもつ画像である。ここでも、計算機の性能上、U-Net に入力する画像のサイズは 288×288 画素に縮小処理した画像である。U-Net からの出力値は 0〜1 までの値を有するため、閾値を 0.5 に設定して改ざん領域を白、非改ざん領域を黒で表示した 2 値化画像を最終出力画像とする。閾値の妥当性は後述する実験により検討を行う。出力画像のサイズは 288×288 画素である。

TypeA と TypeD で用いた U-net 構造は共に図 4.5 に示すもので同一構造であり

Theano(Keras) を使用した。TypeC の LSTM には Tensorflow を使用した。また、学習用画像のうち 1 割をバリデーションに使用し、学習エポック数 100 のうち最もバリデーション精度の高かったモデルを用いて推論を行った (図 4.8)。

検出精度の算出には、次式で定義する適合率 P(Precision)、再現率 R(Recall) 及び Precision と Recall の調和平均 F(F-measure) を用いた。ここで、T_P は検出した改ざん画素数、F_N は未検出の改ざん画素数、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。

TypeB,C は出力画像がオリジナル画像と同サイズであるため、オリジナル画像の全画素で検出精度を算出する。一方、TypeA,D は出画像が 288×288 画素であるため、教師画像を 288×288 画素に縮小処理を行い、288×288 画素の全画素で検出精度を算出する。

$$P = \frac{T_P}{T_P + F_P} \tag{4.7}$$

$$R = \frac{T_P}{T_P + F_N} \tag{4.8}$$

$$F = \frac{2 \cdot P \cdot R}{P + R} \tag{4.9}$$

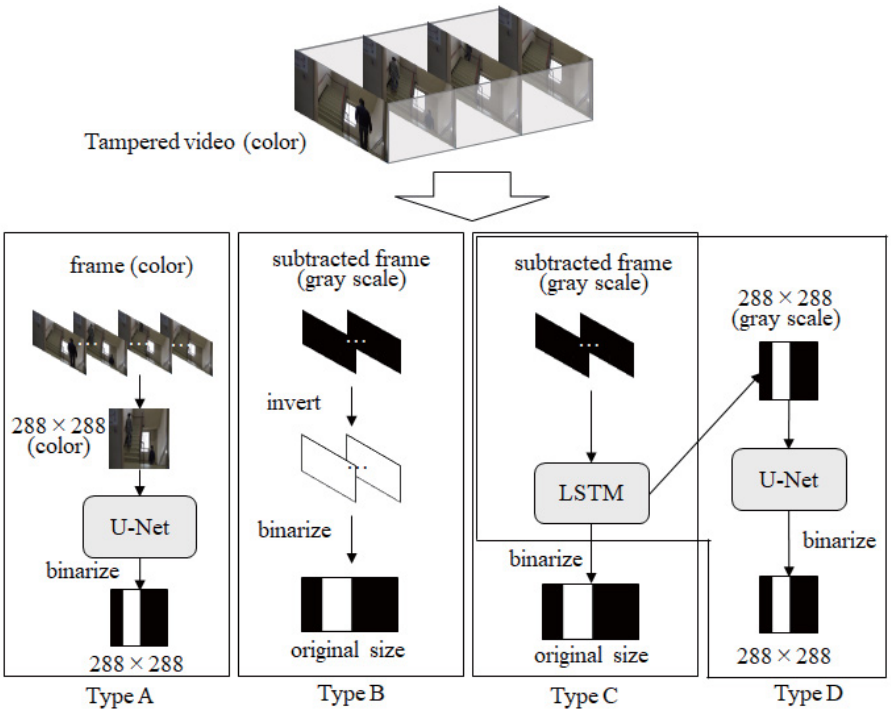


図 4.7 比較実験を行う TypeA〜D

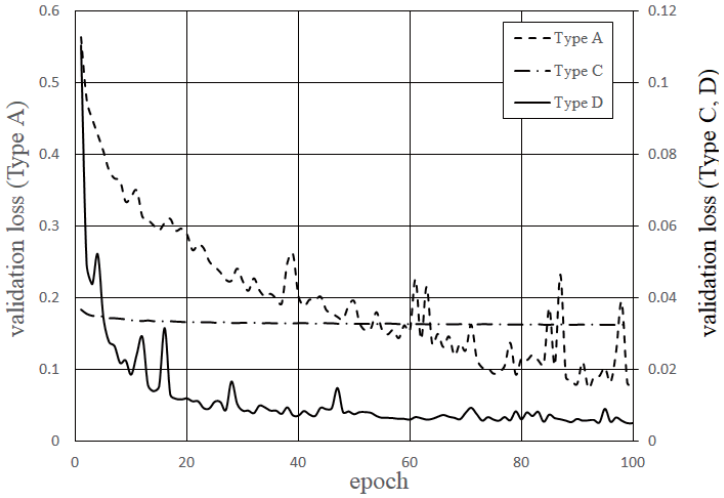


図 4.8 TypeA,C,D のバリデーションロス

4.9 実験結果

推論動画 No.7〜12 の全フレームの Precision、Recall 及び F-measure の平均値を表 4.1 に示す。また、推論動画 No.7,10,12 の推論結果画像を図 4.9〜4.11 に示す。なお、TypeA、D の出力画像は原画像と同じアスペクト比にリサイズして表示している。

TypeA では改ざん領域を正しく検出できておらず、F-measure は平均で 0.17 に留まった。改ざん領域に用いた背景画像は非改ざん領域と融合しており目視では認識しづらいため、U-Net のみでは検出できないことが示された。TypeB は動画によって Precision の値に差が出ている。隣接するフレーム間で時間的ゆらぎがない画素を改ざん画素として誤検出しているものと考えられる。TypeC では、31 の連続フレームから時間的ゆらぎを考慮するため Precision が改善されているが、依然として非改ざん領域の誤検出が細かく残っている。TypeD では U-Net を用いることにより、細かく残っていた誤検出画素を的確に除去することができており、F-measure が 0.94 以上の極めて高い値を得ることができた。

表 4.1 検出結果

	Type A			Type B		
video	*Precision*	*Recall*	$F-measure$	*Precision*	*Recall*	$F-measure$
No.7	0.22	0.25	0.23	0.25	0.98	0.39
No.8	0.62	0.14	0.23	0.30	0.98	0.45
No.9	0.00	0.00	-	0.68	0.99	0.80
No.10	0.07	0.38	0.11	0.04	0.92	0.08
No.11	0.00	0.00	-	0.61	0.99	0.75
No.12	0.33	0.07	0.11	0.49	0.99	0.65
ave.	**0.21**	**0.14**	**0.17**	**0.39**	**0.97**	**0.52**

	Type C			Type D		
video	*Precision*	*Recall*	$F-measure$	*Precision*	*Recall*	$F-measure$
No.7	0.47	0.97	0.63	0.98	0.99	0.98
No.8	0.48	0.97	0.64	0.89	0.99	0.94
No.9	0.85	0.99	0.91	0.99	1.00	1.00
No.10	0.12	0.87	0.21	0.99	0.96	0.98
No.11	0.78	0.99	0.87	0.99	1.00	1.00
No.12	0.68	0.98	0.80	0.99	1.00	0.99
ave.	**0.56**	**0.96**	**0.68**	**0.97**	**0.99**	**0.98**



(a) オリジナル動画



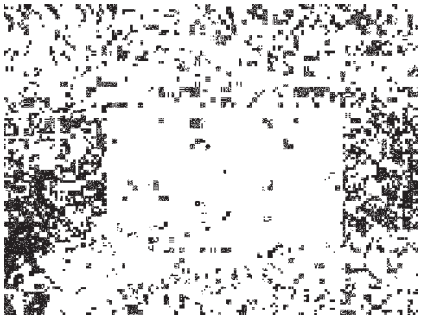
(b) 改ざん動画



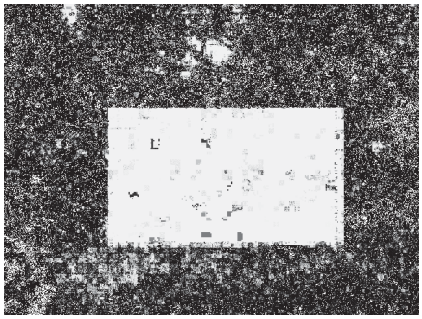
(c) ground truth



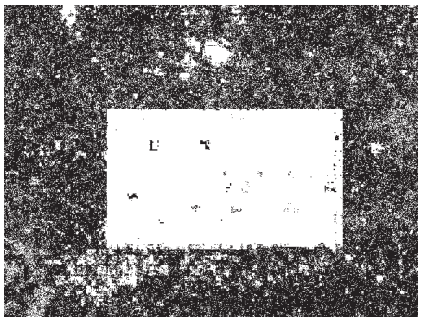
(d) TypeA の検出結果



(e) TypeB の検出結果



(f) LSTM の出力結果



(g) TypeC の検出結果



(h) TypeD の検出結果

図 4.9 推論動画 No.7 の検出結果



(a) オリジナル動画



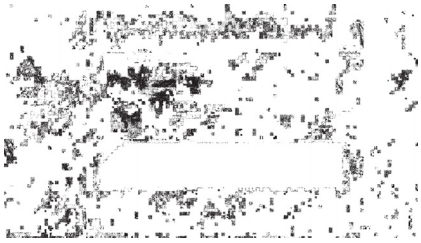
(b) 改ざん動画



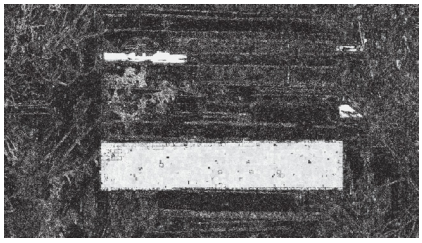
(c) ground truth



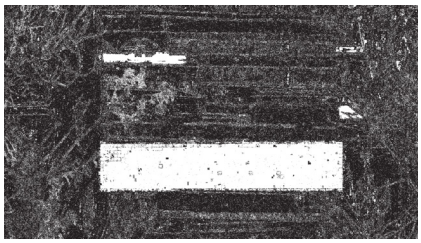
(d) TypeA の検出結果



(e) TypeB の検出結果



(f) LSTM の出力結果



(g) TypeC の検出結果



(h) TypeD の検出結果

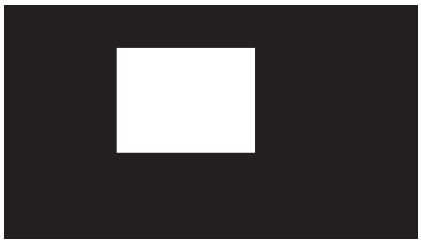
図 4.10 推論動画 No.10 の検出結果



(a) オリジナル動画



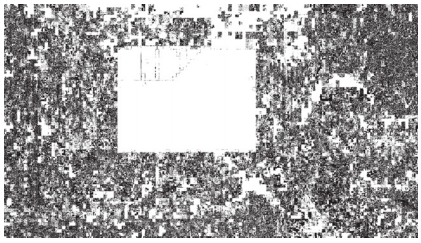
(b) 改ざん動画



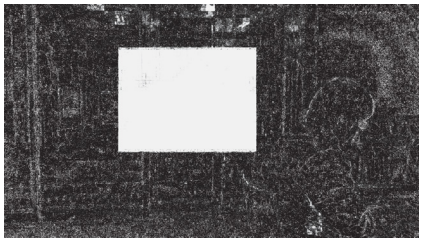
(c) ground truth



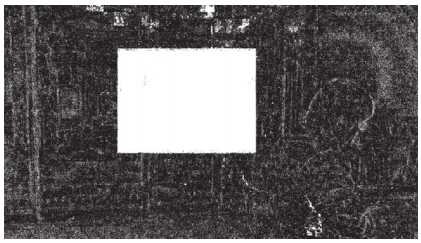
(d) TypeA の検出結果



(e) TypeB の検出結果



(f) LSTM の出力結果



(g) TypeC の検出結果



(h) TypeD の検出結果

図 4.11　推論動画 No.12 の検出結果

4.10 考察

推論動画 No.7,10,12 について、改ざん検出画素数の推移を図 4.12〜4.14 に示す。これは、各フレームにおいて改ざん検出画素数を全画素数で除したものである。検出した画素が真の改ざん画素であるか否かは考慮していないが、時間方向に改ざん検出画素数がどのように推移しているかを視覚的に把握することができる。

TypeA は一定数を常に検出している傾向があり、ground truth の改ざん面積の時間変化に追従していないことがわかる。

TypeB は ground truth に追従している場合もあるが検出数が多く、かつ、変動幅も大きい。これらのゆれは、MPEG Encorder における GOP の周期や樹木の揺れなどに敏感に反応しているものと思われる。

TypeC は ground truth に追従しており変動が落ち着いてきているが全体的に 0.1〜0.2 程度のバイアスが加わっている。これが細かく残る非改ざん領域の誤検出分であると考えられる。

TypeD は ground truth に追従しており、TypeC に見られたバイアスが無くなっている。U-Net を用いることで細かな非改ざん領域の誤検出を大幅に削減し、改ざん領域のみを検出することができている。さらに、TypeB,C が持つ時間的変動も取り除けており、安定した検出が維持できている。推論動画 No.7 の 340 フレーム目付近に検出数の高くなるピークが見られ、これは他の推論動画においても改ざん領域が遷移するタイミングで見られる。改ざん静止画を動画にエンコードする際に、GOP 内に改ざん領域の異なるフレームが含まれることが原因と考えられる。改ざん領域遷移時にも対応できるようにするには、LSTM と U-Net の両面からネットワーク構造を改良する必要がある。

次に、TypeB,C,D において 2 値化する閾値を変化させたときの P-R 曲線の推論動画 No.7〜12 の平均値を図 4.15 に示す。なお、TypeA は改ざん領域に応じた検出ができていないため本考察の対象外とする。TypeB は輝度閾値 Vt=1,2,…,5 の 5 通り、TypeC,D は出力層の閾値を 0.1,0.2,…,0.9 の 9 通りで Recall 及び Precision を求めており、黒塗りで示したマーカーは TypeB で輝度閾値 Vt=1、TypeC,D で出力層の閾値 0.5 のものである。TypeB では輝度閾値を大きくするほど改ざんとして検出する領域が増加するため、Recall が高くなる一方で Precision が低下しており、結果的に F-measure が低下する。TypeC,D を比較すると、閾値の調整次第で Recall の優劣が多少変化するものの、Precision の面では TypeD が圧倒的に有利であることが分かる。このことが F-measure の大きな改善につながっている。以上のように提案するネットワーク構造の有用性が示されている。

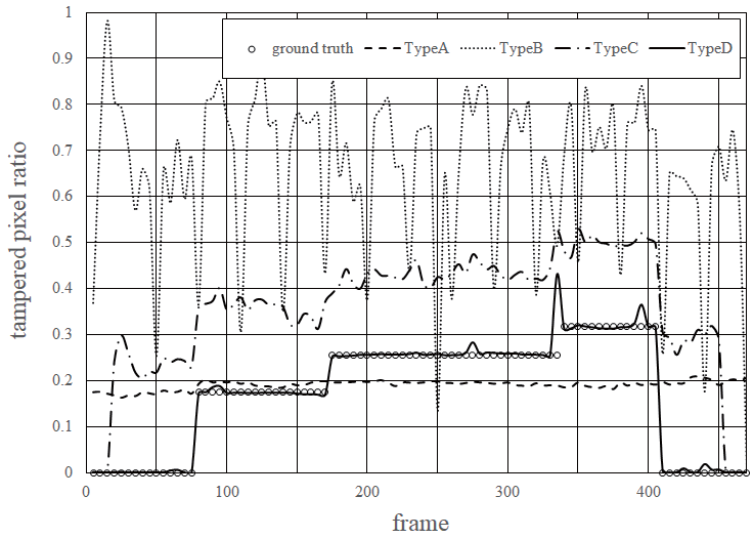


図 4.12 改ざん検出画素比 (推論動画 No.7)

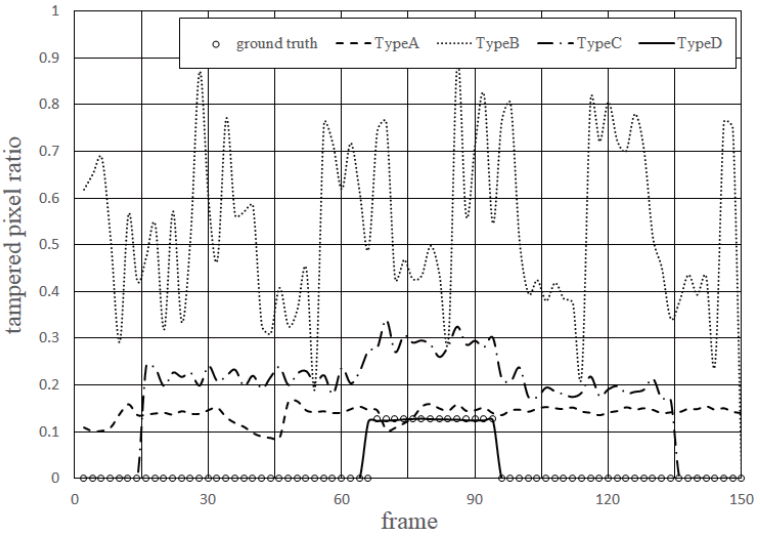


図 4.13 改ざん検出画素比 (推論動画 No.10)

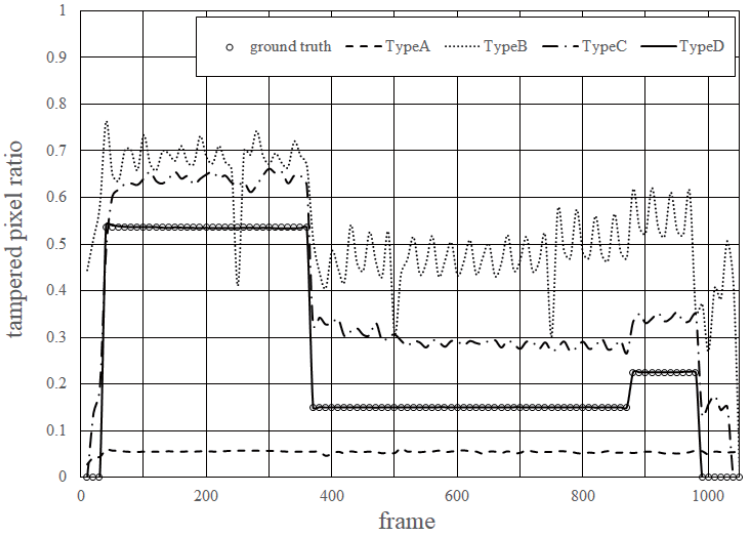


図 4.14 改ざん検出画素比 (推論動画 No.12)

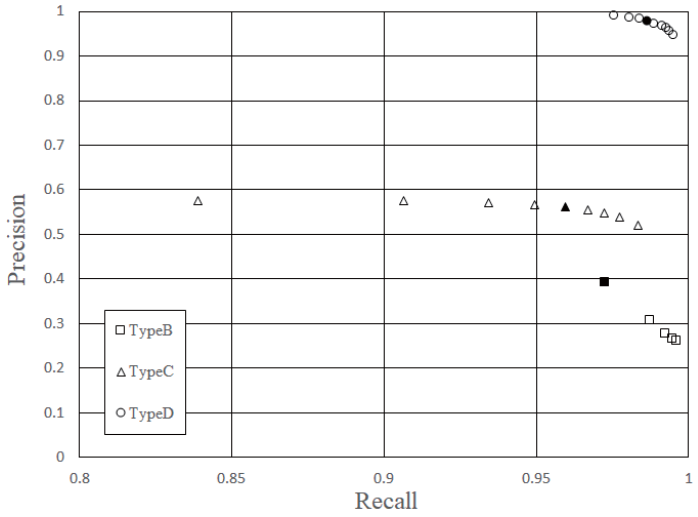


図 4.15 P-R 曲線

4.11 まとめ

本章では、改ざん動画における改ざん領域の検出手法について述べた。動画はエンコード時に画像に圧縮が加わるため改ざん痕跡が静止画ほどは残らず、1 フレームのみの検査では検出が困難である。そこで、動画では改ざん領域が時間的に連続する点に着目し改ざん領域の検出を試みた。動画を直方体の三次元空間と考え、時間方向に LSTM を、空間方向に U-Net を用いることより、わずかな改ざん痕跡であっても直方体内の不連続性を解析し、改ざん領域として識別することができた。同手法により、目視では認識できない改ざんであっても F-measure が 0.94 以上の高精度で検出することができ、改ざん動画の検出において本手法が極めて効果的であることを示した。本手法は動画形式を MPEG に限定するものではなく、また、改ざん領域の形状や位置が時間的に変化した場合においても適用できることから、汎用性の高い検出手法といえる。

5　結論

本論文では、法科学における画像解析技術の高度化を実現するために、次の 3 つのテーマに関して述べた。

1. 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（第 2 章）
2. 畳み込みニューラルネットワークを用いた JPEG 画像における改ざん領域の検出（第 3 章）
3. LSTM および U-Net を用いた時空間解析による動画改ざん領域の検出（第 4 章）

第 2 章では、対象画像が改ざん画像であるか否かを判別するために、ブロックノイズ解析及び Double JPEG 解析の二つ側面から解析を行った。それぞれの解析から得た定量化パラメータを SVM に入力する手法を提案し、従来手法よりも高い精度で改ざん画像か非改ざん画像かを判別することに成功した。

第 3 章では、改ざん画像において改ざん領域を検出する手法として、DCT 係数を二次元画像化し、これを畳み込みニューラルネットワークに入力する方法を提案し、従来手法よりも飛躍的に高い精度で検出することに成功した。

第 4 章では、改ざんが加えられたディジタル動画に対する改ざん領域を検出する手法を提案した。空間方向の解析（U-Net）に加え、時間方向の解析（LSTM）を行うことにより、静止画とは異なる動画特有の改ざん痕跡の時間変動を含めた特徴量抽出が可能となり、ディジタル動画を対象とした改ざん領域の検出手法を実現した。

画像処理技術は日進月歩であり AI を用いた改ざん（合成）技術なども研究開発されている。各改ざん方法には何らかの痕跡が残る可能性があるため、今後も最新改ざん方法に対応できる検出技術の研究を行う必要がある。

以上のように、本論文では法科学における画像解析能力を大きく向上させる 3 つの研究成果について述べた。法科学における画像解析は、多くの人々の日常生活とは縁がない技術であるが、一方で、事件や事故に巻き込まれた際には警察活動や公判において見聞きするであろう技術である。法科学においてディジタル画像やディジタル動画の客観性や真実性を適切に把握するために、本論文で提案した改ざん検出手法が一助となれば幸いである。

謝　辞

本論文をまとめるにあたり、終始熱心なご指導と貴重な助言を頂きました神戸大学大学院工学研究科 黒木修隆准教授に謹んで感謝の意を表します。

本研究の機会を与えていただき、貴重なご助言と熱心なご指導を賜りました神戸大学大学院工学研究科 沼昌宏教授、大阪大学大学院工学研究科 廣瀬哲也教授（元神戸大学大学院工学研究科准教授）に深く感謝いたします。

神戸大学大学院工学研究科 寺田努教授、増田澄男教授には、本研究を論文としてまとめるにあたって貴重な御教示とご助言を頂きましたことに、深く感謝いたします。

改ざん画像の検出技術について、基礎技術やノウハウ、解析ツールを惜しみなくご提供頂きました、株式会社扶桑プレシジョン代表取締役兼関西大学システム理工学部非常勤講師 竹田直人先生、同社開発部部長 宿院康昭氏には、心より厚く御礼申し上げます。

本研究の遂行にあたりお世話になりました京都府警察本部刑事部科学捜査研究所所長以下所員のみなさま、特に尾崎吉明氏、小林正氏、河瀬司氏には心から感謝いたします。

また、本研究に携わり、ディスカッションや貴重な改ざん動画データの作成に協力して頂きました、同研究室の原口俊樹君、吉田昇馬君、浦添和哉君、他学生諸君に心より感謝いたします。

最後に、社会人ドクターとしての博士後期課程への進学に理解を示し、生活を支えてくれた妻、子供たち、両親に心より感謝いたします。

参考文献

[1] Hany Farid, Digital Image Forensics, University of California, Berkeley, https://farid.berkeley.edu/

[2] Xunyu Pan, Siwei Lyu, “Region Duplication Detection Using Image Feature Matching,” IEEE Transaction on Information Forensics and Security, vol.5, No.4, Dec 2010.

[3] Amanpreet Kaur, Richa Sharma, “Copy-Move Forgery Detection using DCT and SIFT,” International Journal of Computer Applications, vol.70, No.7, 2013.

[4] Yanping Huang, Wei Lu, Wei Sun, Dongyang Long, “Improved DCT-based detection of copy-move forgery in images,” Forensic Science International, no.206, pp.178-184, 2011.

[5] Ashima Gupta, Nisheeth Saxena, S.K Vasistha, “Detecting Copy move Forgery using DCT,” International Journal of Scientific and Research Publications, vol.3, Issue 5, 2013.

[6] Micah K. Johnson, Hany Farid, “Exposing Digital Forgeries by Detecting Inconsistencies in Lighting,” ACM Multimedia and Security Workshop, New York, 2005.

[7] Alin C. Popescu, Hany Farid, “Exposing Digital Forgeries in Color Filter Array Interpolated Images,” IEEE Transaction on Signal Processing, vol.53, No.10, 2005.

[8] Yu-Feng Hsu, Shih-Fu Chang, “Camera Response Function for Image Forensics : An Automatic Algorithm for Splicing Detection,” IEEE Transactions on Information Forensics and Security, Vol.5, No.4, 2010.

[9] Zhouchen Lin, Rongrong Wang, Xiaoou Tang, Heung-Yeung Shum, “Detecting Doctored Images Using Camera Response Normality and Consistency,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.1087-1092, 2005.

[10] 藤本郷史, “工事写真における編集痕跡検出技術の要件分析および要件に適合した塗り潰し編集検出技術の開発と検証”, 日本建築学会構造系論文集, vol.77, no.674, pp.647-656, 2012.

[11] Tiziano Bianchi, Alessandro Piva, “Detection of Non-Aligned Double JPEG Compression Based on Integer Periodicity Maps,” IEEE Transaction on Information Forensics and Security, vol.7 pp.842-848, 2012.

[12] Tomas Pevny, Jessica Fridrich, “Detection of double-compression in JPEG im-

ages for application in steganography," IEEE Transaction on Information Forensics and Security, vol.3. pp.247-258, 2008.

[13] Hany Farid, "Exposing Digital Forgeries from JPEG Ghosts," IEEE Transaction on Information Forensics and Security, vol.4. pp.154-160, 2009.

[14] Junfeng He, Zhouchen Lin, Lifeng Wang, Xiaoou Tang, "Detecting Doctored JPEG Images Via DCT Coefficient Analysis," Computer Vision-ECCV 2006, vol.3953 pp.423-435, 2006.

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

[16] C.C.Chang, C.J.Lin, "LIBSVM : A Library for Support Vector Machines," National Taiwan Univ. https://www.csie.ntu.edu.tw/ cjlin/libsvm/

[17] Belhassen Bayar, Matthew C.Stamm, "A Deep Learning Approach To Universal Image Manipulation Detection Using A New Convolutional Layer," Proceedings of the 4th ACM Workshop on Information Hiding and Multimedia Security 2016.

[18] Edmar R.S.de Rezende, Guilherme C.S.Ruppert, Tiago Carvalho, "Detecting Computer Generated Images with Deep Convolutional Neural Networks," 30th Conference on Graphics, Patterns and Images SIBGRAPI 2017.

[19] Luca Bondi, Silvia Lameri, David Guera, Paolo Bestagini, Edward J.Delp, Stefano Tubaro, "Tampering Detection and Localization through Clustering of Camera-Based CNN Features," IEEE Conference on Computer Vision and Pattern Recognition Workshop 2017.

[20] Jawadul H.Bappy, Amit K.Roy-Chowdhury, Jason Bunk, Lakshmanan Nataraj, B.S.Manjunath, "Exploiting Spatial Structure for Localizing Manipulated Image Regions," Proceeding of the IEEE international conference on computer vision 2017.

[21] Qing Wang, Rong Zhang, "Double JPEG compression forensics based on a convolutional neural network," EURASIP Journal on Information Security 2016.

[22] W.Wang, H.Farid, "Exposing Digital Forgeries in Video by Detecting Double Quantization," Proceedings of the 11th ACM workshop on Multimedia and security, pp.39-48, Princeton NJ, USA, 2009

[23] Xinghao Jiang, Wan Wang, Tanfeng Sun, Yun Q.Shi, "Detection of Double Compression in MPEG-4 Videos Based on Markov Statistics," IEEE Signal Processing Letters, vol.20, No.5, 2013.

[24] A.V.Subramanyam, Sabu Emmanuel, "Video Forgery Detection using HOG Features and Compression Properties," IEEE 14th International Workshop on Multimedia Signal Processing, pp89-94, 2012.

[25] Michihiro Kobayashi, Takahiro Okabe, Yoichi Sato, "Detecting Forgery From

Static-Scene Video Based on Inconsistency in Noise Level Functions," IEEE Transaction on Information Forensics and Security, vol.5. 2010.

[26] J.Xu, Y.Yu, "Detection of Blue Screen Special Effects in Videos," Physics Procedia no.33, pp.1316-1322, ICMPBE International Committee 2012.

[27] M.A.Bagiwa, A.W.A.Wahab, "Chroma key background detection for digital video using statistical correlation of blurring artifact," Digital Investigation no.19 pp.29-43, 2016.

[28] D.Dario, C.Davide, "Autoencoder with recurrent neural networks for video forgery detection," Society for Imaging Science and Technology, pp.92-99, 2017.

[29] P.Bestagini, S.Milani, "Local tampering detection in video sequences," MMSP, Italy, 2013.

研究業績

学術掲載論文

1. 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 C, Vol.137, No.5, pp.742-749, 2017.
2. 多谷邦彦, 黒木修隆, 竹田直人, 宿院康昭, 小林正, “畳み込みニューラルネットワークを用いた JPEG 画像における改ざん領域の検出”, 電気学会論文誌 C, Vol.138, No.11, pp.1417-1424, 2018.
3. 多谷邦彦, 黒木修隆, 竹田直人, 廣瀬哲也, 沼昌宏, “LSTM および U-Net を用いた時空間解析による動画改ざん領域の検出”, 電気学会論文誌 C, Vol.140, No.4, pp.476-483, 2020.

国際会議

1. K.Taya, N.Kuroki, N.Takeda, T.Hirose, M.Numa, “Detecting tampered region in JPEG images via CNN,” 18th IEEE International NEWCAS conference, Canada, June. 2020.

学術講演

1. 多谷邦彦, 竹田直人, 宿院康昭, 小林正, 尾崎吉明, “JPEG 画像の改ざん検出ツールの開発及び研究”, 第 20 回日本法科学技術学会学術集会, D-38, 2014.
2. 多谷邦彦, 竹田直人, 宿院康昭, 小林正, 尾崎吉明, 川島成平, “DCT 成分解析等を用いた JPEG 画像の総合的改ざん検出ツールの開発”, 第 14 回情報科学技術フォーラム, I-006, 2015.
3. 多谷邦彦, 黒木修隆, 竹田直人, 宿院康昭, 小林正, “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”, 第 16 回情報科学技術フォーラム, CH-013, 2017.
4. 多谷邦彦, 黒木修隆, 竹田直人, “ディープラーニングを用いた法科学的画像解析の一例”, 第 23 回日本法科学技術学会学術集会, D-41, 2017.
5. 多谷邦彦, 黒木修隆, 竹田直人, 沼昌宏, “LSTM および CNN を用いた動画改ざん領域の検出”, 平成 31 年電気学会全国大会, 3-044, 2019.

受賞

1. 多谷邦彦, 日本法科学技術学会奨励賞（第 23 回法科学技術学会学術集会）, 2017 年 12 月 8 日.

付録

Double JPEG について

二度の JPEG 圧縮が生じる時、DCT 係数は次の過程を得て変化する。

1. 一度目の圧縮
   （a）DCT が行われる（変換後の係数の値を u とする。）。
   （b）一度目の量子化が値 q_1 で行われる。係数の値は $Q_{q_1}(u) = [u/q_1]$ となり、$[x]$ は x に最も近い整数である。
2. JPEG ファイルの展開
   （a）値 q_1 で逆量子化される。係数の値は $Q_{q_1}^{-1}(Q_{q_1}(u)) = [u/q_1]q_1$となる。
   （b）逆 DCT が行われる。
3. 二度目の圧縮
   （a）DCT が行われる。
   （b）二度目の量子化が値 q_2 で行われる。係数の値は $Q_{q_1q_2}(u) = [[u/q_1]q_1/q_2]$ となる。

次に、二度の JPEG 圧縮が行われた DCT 係数のヒストグラムには特有の性質が見られることを述べる。

一度目の量子化を行う前の DCT 係数の値を u_1、二度目の量子化の後の DCT 係数の値を u_2 とすると、

$$Q_{q_1q_2}(u_1) = \left[\left[\frac{u_1}{q_1}\right]\frac{q_1}{q_2}\right] = u_2 \tag{A.1}$$

と表される。したがって、

$$u_2 - \frac{1}{2} \leq \left[\frac{u_1}{q_2}\right]\frac{q_1}{q_2} < u_2 + \frac{1}{2} \tag{A.2}$$

そのため、

$$\left\lceil \frac{q_2}{q_1}\left(u_2 - \frac{1}{2}\right)\right\rceil - \frac{1}{2} \leq \frac{u_1}{q_2} < \left\lfloor \frac{q_2}{q_1}\left(u_2 + \frac{1}{2}\right)\right\rfloor + \frac{1}{2} \tag{A.3}$$

ここで、$\lceil x \rceil$ は天井関数であり、$\lfloor x \rfloor$ は床関数である。

q_1 が偶数の場合、

$$q_1\left(\left\lceil \frac{q_2}{q_1}\left(u_2 - \frac{1}{2}\right)\right\rceil - \frac{1}{2}\right) \leq u_1 < q_1\left(\left\lfloor \frac{q_2}{q_1}\left(u_2 + \frac{1}{2}\right)\right\rfloor + \frac{1}{2}\right) \tag{A.4}$$

となり、q_1 が奇数の場合は、

$$q_1\left(\left\lceil\frac{q_2}{q_1}\left(u_2-\frac{1}{2}\right)\right\rceil-\frac{1}{2}\right)+\frac{1}{2}\leq u_1<q_1\left(\left\lfloor\frac{q_2}{q_1}\left(u_2+\frac{1}{2}\right)\right\rfloor+\frac{1}{2}\right)-\frac{1}{2} \tag{A.5}$$

となる。

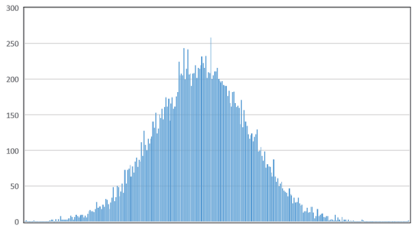
どちらの場合も、二度目の量子化後の DCT 係数のヒストグラム h_2 の bin u_2 に寄与する、オリジナルのヒストグラムの bin の数 $n(u_2)$ は、

$$n(u_2)=q_1\left(\left\lfloor\frac{q_2}{q_1}\left(u_2+\frac{1}{2}\right)\right\rfloor-\left\lceil\frac{q_2}{q_1}\left(u_2-\frac{1}{2}\right)\right\rceil+1\right) \tag{A.6}$$

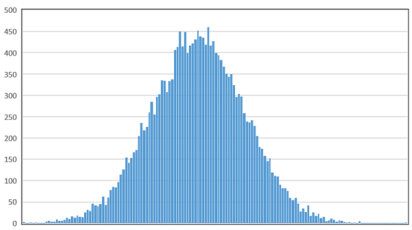
となる。

正規分布をもつ乱数から模擬的にオリジナル画像の DCT 係数を作成し、これを $q_1=2$ で量子化し、次に $q_2=5$ で量子化したもの、また、$q_1=5$ で量子化し、次に $q_2=2$ で量子化したものを図 A.1 に示す。$n(u_2)$ は q_1 の周期関数であり、これが二度目の量子化後のヒストグラムに周期性が現れる原因となっている (図 A.1 の (c)、(e))。

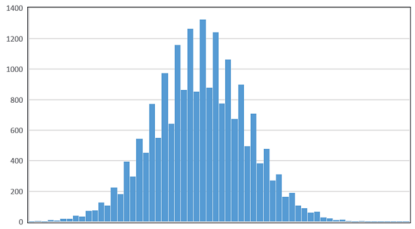
$q_1<q_2$ のとき、二度目の量子化後のヒストグラムには周期的な山と谷が現れる。一方、$q_1>q_2$ のときはヒストグラムの bin が周期的に 0 となる。図 A.1 の (d) に示す例は $q_1=5, q_2=2$ のとき、$n(5k+1)=0$ となるためである。このようなヒストグラムの周期的な特徴が Double JPEG effect と呼ばれる。



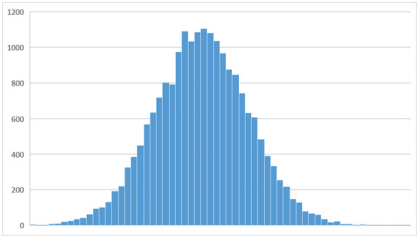
(a) オリジナル



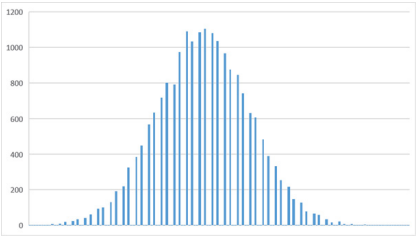
(b) (a) を $q_1 = 2$ で量子化したもの



(c) (b) を $q_2 = 5$ で量子化したもの



(d) (a) を $q_1 = 5$ で量子化したもの



(e) (d) を $q_2 = 2$ で量子化したもの

図 A.1 DCT 係数のヒストグラム

神戸大学博士論文「法科学における画像解析技術の高度化に関する研究」全 77 頁
提出日　2020 年 7 月 14 日
本博士論文が神戸大学機関リポジトリ Kernel にて掲載される場合、掲載登録日（公開日）はリポジトリの該当ページ上に掲載されます。
ⓒ　多谷　邦彦
本論文の内容の一部あるいは全部を無断で複製・転載・翻訳することを禁じます。