

CNNを用いた予測に有効なナンバープレート写真用 トレーニングデータの検討

鈴木 友哉^{1,a)} 宇田 隆哉¹

受付日 2020年5月14日, 採録日 2020年11月5日

概要: 犯罪捜査などで自動車を特定するため、ナンバープレートの数字を読み取ることは重要である。しかし、自動車が画像または映像でとらえられていても、そのナンバープレートの数字を人間が読み取ることは難しい場合がある。とくに、監視カメラは、メモリ容量を節約したり、肖像権またはプライバシーの侵害リスクを低減させたりするために、一般的に解像度が低く設計されている。また、画像は、ボケ、ノイズ、ヘイズ、光量の偏差によって劣化する。そこで我々は、人間が読むことが難しいナンバープレートの数字を、CNNによって推測する方法を提案し、実験を行った。CNNによる画像の分類を行う場合、トレーニング用の画像として実際の写真を大量に集めることが一般的である。しかし、角度や明るさなど、あらゆる条件を合わせた大量の写真をそのつど用意することは困難である。そこで、我々はトレーニング用の画像をCGを用いて作成した。これにより、異なるものを何枚でも用意できるが、計算コストに対して精度を高める効率の悪いものや、加えることで精度を下げてしまうものが存在する可能性がある。そこで、本論文では、どのような画像を追加することで、効率良く精度を高められるのか評価を行った。実験では、明るくした画像や暗くした画像の追加が有効であり、画像の一部を切り出した画像の追加はあまり有効でないという結果が得られた。また、人間が確信を持って判読できない数字の1つを、コンピュータが最大平均98.6%の予測値で推測可能であることが示された。本論文の技術は、デジタルフォレンジックや捜査対象を絞り込む際の補助的な情報として利用可能である。

キーワード：ナンバープレート、機械学習、デジタルフォレンジック

Examination of Choosing License Plate Photographs for Effective Training Data of CNN

TOMOYA SUZUKI^{1,a)} RYUYA UDA¹

Received: May 14, 2020, Accepted: November 5, 2020

Abstract: Reading numbers on license plates of cars are one of the important methods for criminal investigation. However, it is sometimes difficult to read the numbers on images or video by humans. Especially, resolution of video by surveillance cameras is usually low due to saving recording memory space or reducing infringement risks of portrait right or privacy. Moreover, images become deteriorate by blur, noise, haze and quantity of light. Therefore, we proposed a method to read numbers on license plates of cars, which are difficult to be read by humans, by CNN and evaluated it. When images are classified by CNN, quite many real images for training data must be collected. However, it is difficult to collect such many real images which match any environment such as angles, lights, etc. for only one image. Therefore, we created the images by computer graphics. The creation enables to prepare any numbers of different images, but there may be ineffective high cost images or those decrease classification accuracy. Therefore, in this paper, we investigated which kind of images can enhance the accuracy effectively. As a result, to add lighter and darker images was effective and to add cropped images was not effective. In addition, the result shows that computers can read a number with 98.6% predicted accuracy in maximum average, which cannot be read by humans. The proposed method can be used for digital forensics and narrowing down the candidates in criminal investigation.

Keywords: license plate, machine learning, digital forensics

1. はじめに

ナンバープレートは自動車に一意に割り振られているものであり、犯罪捜査においてその特定に有用である。また、監視カメラや車載カメラの普及により、様々な時間や場所でナンバープレートが撮影されている。吉浦らは、防犯用の監視カメラは犯罪捜査のために不可欠であると指摘している [1]。吉浦らによると、2014 年度からは、日本の地方自治体の地方道路に防犯用の監視カメラが導入される動きがあった。また、東京では、公立小学校 1,300 校の学校道路に計 6,500 台の防犯用の監視カメラ (1 校あたり 5 台) が導入される計画が発表された。しかし、防犯用の監視カメラの多くは、メモリ容量を節約したり、肖像権やプライバシーの侵害リスクを低減させたりするために、解像度が低く設計されている。また、画像は、ボケ、ノイズ、ヘイズ、光量の偏差によって劣化する。よって、場合によってはナンバープレートの画像から、人間が数字を読み取ることが困難である。そこで、本研究では、人間に数字の判別が困難なナンバープレートの画像から、コンピュータを用いてその数字を読み取る手法を提案し、実験を行った。

本研究では、機械学習によりナンバープレートの数字を読み取る方法でアプローチした。画像を機械学習で判別する場合、トレーニング用の画像が大量に必要となる。しかし、その方法では、あるナンバープレートの写真から数字を読み取ろうとする場合、類似の角度や距離、明るさなどの条件を合わせたうえで、異なる数字を持つナンバープレートの写真を大量に用意しなければならない。読み取るナンバープレートの写真が複数であれば、その数だけ同様の作業を行う必要があり、困難である。

そこで、我々は、トレーニング用の画像を CG を用いて作成する方法を考案した [2]。日本において、ナンバープレートに用いられている数字のフォントは一意で位置も固定されており、ナンバープレートそのものも剛性の高い素材で作成されている。すなわち、扱う対象は手書き文字のようにサンプルによって形が異なるものや、人間の顔や動物のように変形を前提とするものではないということである。これにより、異なる画像を何枚でも用意できるが、計算コストに対して精度を高める効率の悪いものや、加えることで精度を下げてしまうものが存在する可能性がある。本論文では、トレーニング用画像に加える画像の種類と精度の変化について評価する。

本論文で述べる技術の目的は、人間に数字の判別が困難なナンバープレートの画像から、裁判の証拠や容疑者の自動車の絞り込みに必要な精度をもって、コンピュータが自

動的に数字を分類できる可能性を示すことである。なお、その際の機械学習においてトレーニング用の実画像を必要としないことを特徴とする。他の研究と根本的に異なるのは、ナンバープレートの数字をいかに高い精度で読み取るかを目的とはしていないことである。たとえばある数字が「1」であると推測された場合、他の数字である確率は 50%未満であったとしても、50%の確率で 1 であるのと、99%の確率で 1 であるのは、裁判の証拠能力としては大きく異なる。よって、本研究においては、ある数字を正しく分類できたという精度ではなく、その数字がその数字である確率が問題となる。本論文では、これを予測値として扱う。もちろん、ナンバープレートの数字を高い精度で読み取れば本研究の目的にも叶うが、非常に精度の低いものの精度が少し向上しても意味はない。すべての数字を可能な限り正確に読み取って証拠として扱うのではなく、人間にとって識別が困難な数字の中に、コンピュータが正確に読み取って証拠として扱えるものがあることを示すことが本論文の主旨である。

2. 関連研究

ナンバープレートを認識するという研究には分野があり、Patel らのサーベイ [3] および Sonavane らのサーベイ [4] によると、この分野は (1) 自動車画像のキャプチャ、(2) ナンバープレート領域の抽出、(3) 文字 (数字) のセグメンテーション、(4) 文字 (数字) の認識の 4 つに分けられる。本研究はこのうち (4) に該当する。なお、(4) のうち、人間に判読可能なナンバープレートの数字の自動識別と人間に判読不可能なナンバープレートの数字の自動識別があるので、本章ではこれらを分けて扱う。

2.1 自動車画像のキャプチャ

Albiol らは、道路に駐車された車を検出するための、映像ベースのアプローチを示している [5]。彼らの手法では、低レベルの特徴点から情報が抽出され、そのシーンで何が起きているかを説明する時空間マップが作成されている。この手法は、照明の変化、カメラの振動、他の車両による一時的な遮蔽、高いノイズレベルなどの問題に対して堅牢であると述べられている。

Fan は、遮蔽された車両やカメラの死角の問題に対処するために、マルチカメラ結合に基づく車両識別方法を提案している [6]。彼の手法では、異なるカメラから撮影された同一の車両に対して、その画像を切り出す共通の境界領域を抽出している。さらに、この手法に基づいて、投票メカニズムとバイリニアネットワークを組み合わせた車種検出手法を提案している。様々な角度からの画像を使用して車両情報を取得しており、この方法がカメラによる死角の問題を効果的に解決できることが示されている。

¹ 東京工科大学大学院バイオ・情報メディア研究科
Tokyo University of Technology Graduate School, Hachioji,
Tokyo 192-0982, Japan

a) g21190111c@edu.teu.ac.jp

2.2 ナンバープレート領域の抽出

Wang らはニューラルネットワークの概念を適用したナンバープレート認識システムを提案している [7]. このシステムは, ナンバープレートの位置を画像の中から検出するものである.

Amirgaliyev らは ALPR (Automatic License Plate Recognition) の様々なアプローチを検討し, 彼らが評価した検証方法に関して述べている [8]. 彼らは, ナンバープレートを画像の中から検出する研究を行っており, ナンバープレートに類似しているが, ナンバープレートではないものが偽陽性として確認されるとしている.

Ashtari らは, テンプレートマッチング技術を目標色画素の分析を用いて改善し, 車両のイタリアのナンバープレートの位置を, 最大 96% で検出する手法を示している [9].

Rashid は, エジプトのナンバープレートの高速な自動検出システムのアルゴリズムについて説明している [10]. 彼のアルゴリズムでは, 高価なハードウェアを用いた高品質画像を必要とすることなく, 小さなデータセットで約 96% の検出率を達成している.

Yang らは, カラーセグメンテーションに基づいたナンバープレートの位置特定のアルゴリズムを提案している [11]. 具体的には, 前処理段階での画像強調アルゴリズムと, ポジショニング段階での色や形のライセンス情報を活用したロケーションアルゴリズムである. 彼らの手法は, グレースケールセグメンテーションに基づいてナンバープレートを見つけるアルゴリズムよりも速く正確であると述べられている.

2.3 文字のセグメンテーション

羽田らはナンバープレートの数字を抽出する手法について検討している [12]. 彼らの研究では, ナンバープレートの数字が書かれている領域から, 数字の形を切り出している.

Cui らは, 移動中の車両のナンバープレート上の文字を抽出するアプローチを提案している [13]. 彼らの手法では, 空間情報と時間情報を同時に利用し, 文字の抽出をマルコフランダムフィールド (MRF) としてモデル化している. 彼らの手法は既存の手法よりも高いパフォーマンスを発揮すると述べられている.

2.4 人間に判読可能なナンバープレートの数字の自動識別

Wang らは, 中国の自動車のナンバープレートに関して, 文字認識を行うアルゴリズムを示している [14]. 彼らの研究の前提条件として, 人間もその文字をはっきりと読める程度の環境で画像を取得することがあげられる. 用途は, 特定の場所を通過する自動車のナンバープレートを自動的に認識するというものである.

Xing らは, 改良されたラドン変換を用いて, ナンバープ

レート領域の位置を特定したうえで, 文字のセグメンテーションと認識を行う手法について述べている [15].

Nagare らは, 高性能ナンバープレート認識システム (LPRS) を提案している [16]. 提案されている LPRS は, プレート領域の決定, 文字のセグメンテーション, 文字の認識の 3 段階に分かれている. 文字はプレート領域の垂直投影を使用してセグメント化され, 逆伝播学習アルゴリズムを使用した 3 層ニューラルネットワークにより文字が決定される.

Damak らは, Number Plate Localization (NPL) と Characters Segmentation (CS) のコンピュータビジョンアルゴリズムを提案している [17]. さらに, 深層学習による光学式文字認識 (OCR) の手法も提案している. NPL および CS ステップ後に, 検出されたプレートの数字を特定するために CNN が用いられており, モデルのトレーニングには Jetson TX2 NVIDIA ターゲットの数値画像データベースが使用されている. 検出精度は 95.84% と述べられている.

2.5 人間に判読不可能なナンバープレートの数字の自動識別

神宮らは, カメラで撮影したナンバープレートから数字を特定する方法を提案している [18]. 彼らが対象としているのは, 人間にも判読できない画像であり, 目的は完全に我々と同一である. 彼らの実験の精度としては, 文字のサイズが 5×8 画素の 72 枚のプレートに対して数字を推定した場合, 第 1 候補以内が 95.0%, 第 2 候補以内が 100.0%, 第 3 候補以内が 100.0% であり, 文字のサイズが 4×6 ピクセルの 56 枚のプレートに対して数字を推定した場合, 第 1 候補以内が 57.5%, 第 2 候補以内が 77.5%, 第 3 候補以内が 90.0% である. この精度であれば, デジタルフォレンジックとして使用可能であり, 有用な技術である.

しかし, 神宮らの手法には多くの問題がある. まず, 彼らは原画像と完全に同一の位置に, リファレンスプレートと呼ばれるトレーニング用の画像を置いている. 実際には, 任意に作成したナンバープレートを同一の自動車の同一の位置に貼り付けて撮影を行っている. さらに, 同一のカメラを使用している. また, 論文中に記述がないため著者らの推測によるものであるが, カメラの設定が同一で変更しておらず, 撮影位置も同一で, 原画像を撮影してから動かしていないと思われる.

仮に, 警察が監視カメラの映像から同様の実験を行う場合, 寸分違わぬ位置に同じ車種の自動車を置いて同じカメラで撮影する必要がある. さらに, 天候や明るさなどの条件も揃える必要があると思われ, 屋外であった場合には困難をきわめる. もし, 車載カメラの映像であった場合, 現場で同様の再現を行うのはさらに困難となる. そして, 数字を識別したい 1 つのナンバープレートの画像につき, この撮影を行わなければならないことになる. 一方で, 我々

の手法では、トレーニング用画像を CG によって作成しているため、これらの撮影を必要としない。

Spanhel らは、ナンバープレートの文字認識を行っており、その一部として Svoboda らのデータセットを使用している [19]。これは交通監視システムが撮影したナンバープレートの中で、モーション・ブラー（高速な移動によるブレ）があるもので、Svoboda らが CNN を用いてデブラー（ブラーを解除）したものである。Spanhel らの手法では、Svoboda らのデータセットの文字識別を 93.1% の精度で行っている。我々の研究が対象としているものは、人間に判別が困難なナンバープレートであるため、このようなモーション・ブラーも対象となる。ただし、彼らの研究では、ブラーがなければはっきりと人間が文字を読める大きさの画像に対するモーション・ブラーのみを対象としており、解像度が低くて読めない画像には同一手法を単純に適用できない。また、彼らはトレーニング用画像に実データを使用している。一方、本論文では、解像度が低くて読めないナンバープレートを対象とし、実データをトレーニング用画像に使用せずに識別している。

3. 提案手法

3.1 提案概要

我々は、トレーニングデータの画像を CG で作成することにより、実際に撮影されたナンバープレートの数字を、機械学習を用いて識別する手法を考案した。

本論文におけるナンバープレートの数字とは、一連指定番号と呼ばれる、1 から 4 桁のアラビア数字である。なお、3 桁以下のときは「・」の記号（中黒）で空白になる桁が埋められる。ナンバープレートの配色にはいくつかの種類が存在し、図柄入りのものもあるが、本論文では、自家用自動車に取り付けられる、白地に緑文字のナンバープレートを対象とした。

トレーニング用のナンバープレートは CG で作成しているため、実際にナンバープレートを撮影するのは異なり、容易に何枚でも異なる画像を作成できる。本論文では、どのような種類の画像をトレーニングデータに加えれば、効率良く数字の判別が行えるかについて評価する。

3.2 トレーニングデータ

3.2.1 基本画像

トレーニングデータは CG によって作成される。この CG は、ナンバープレートの形状の平面によって作成され、ナンバープレートの画像をテクスチャとして貼り付けている。日本において、ナンバープレートのフォントは公開されていない。私的に再現した画像はあるが、信頼できる根拠はない。そこで、我々は、テクスチャの画像を、十分な解像度で撮影された写真から、各数字および中黒につき 1 枚ずつ（合計 11 枚）作成した。本論文では、これらの画像

を基本画像と呼称する。また、以降は「各数字」という表現には中黒も含むものとする。

基本画像の作成のため、我々はナンバープレートの撮影を行った。カメラは、iPhone 7 に備え付けられているものを用いた。iPhone で撮影された写真は HEIC 形式で保存されるため、これを iMazing HEIC Converter を利用して PNG 形式に変換している。以後、本論文中で示されている iPhone で撮影された写真は、すべて同様の変換が行われたものである。目視により汚れや塗装ハゲのないナンバープレートを選択し、ナンバープレート全体が収まるように撮影した。その後、写真からナンバープレートのみを切り出し、手で傾きを修正した。以上の処理を経た画像の例を図 1 に示す。なお、東京工科大学の指示により、一連指定番号以外を隠して表示したが、研究においては隠す作業は行っておらず、評価には影響しない。

基本画像から、一連指定番号のみを手作業で切り出した。作成された各数字の画像を図 2 に示す。

切り出された各数字は、 30×60 ピクセルの画像として保存した。背景は白で塗りつぶした。数字によって形状が異なるため厳密ではないが、上下の空白が 1 から 2 ピクセルの範囲に収まるように配置した。ただし、中黒は中央に配置した。

画像サイズを 30×60 ピクセルとした理由は、この時点で数字が読めなくなるほどに劣化させないため、 64×64 ピクセルより小さい画像とすることを前提として、3 や 5 などの奇数で割り切れる値としたかった（フィルタによる処理が行われるため）ためや、評価を行ううえで 10 の倍数が都合がよいと考えたためである。

3.2.2 基本トレーニングデータ

提案手法は、正面から撮影されたナンバープレートでなくても対応は可能であるが、本論文の評価においてはおおむね前方から撮影されたものを対象とした。また、対象となる画像の数字は人間が読めないことを前提としている。そこで、我々は、完全な正面から見た基本画像に対して、ある程度の 3 次元回転とボケを付加した画像をトレーニングデータとして用意した。本論文では、これを基本トレーニングデータと呼称する。

まず、完全な正面から見た基本画像が 3 次元回転によって新たな画像となる。実験においては、図 3 に示すとおり軸の向きで、正面の値を π とし、X、Y、Z 軸方向それぞれ、 $-\pi/4$ 、 $-\pi/8$ 、 π （正面）、 $\pi/8$ 、 $\pi/4$ の 5 段階の回転が行われるものとした。回転処理について図 4 に示す。X、Y、Z 軸の向きは図 3 に示したとおりである。X 軸上で負の位置から正の位置を向いた場合に、時計回りに回転することを X 軸方向に正回転、反時計回りに回転することを X 軸方向に負回転という。Y 軸、Z 軸についても同様である。図 3 の X 軸回転では、正面から見た「6」（ π の角度）を、正と負に $\pi/4$ と $\pi/8$ ずつ回転させている。Y 軸回転と Z



図 1 基本画像に用いたナンバープレート画像の例

Fig. 1 Example of license plate image for training samples.

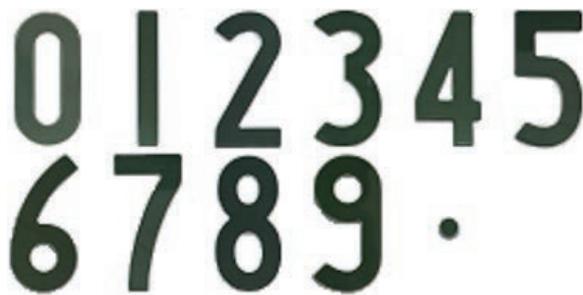


図 2 切り出された数字の画像

Fig. 2 Images of cut out numerals.

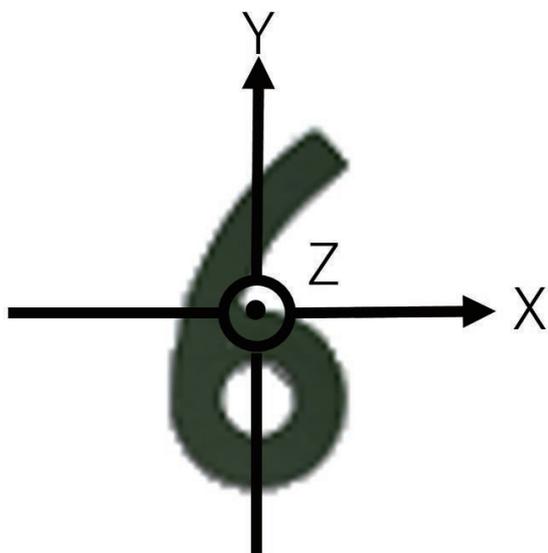


図 3 ナンバープレートと軸の向き

Fig. 3 License plate and direction of axes.

軸回転も同様である。なお、正面から見た画像を 5 段階に X 軸回転させ、5 枚にした画像を、それぞれ正面から見た新たな画像として Y 軸方向に 5 段階に回転させ、25 枚にする。さらに、それらをそれぞれ正面から見た新たな画像として Z 軸方向に 5 段階に回転させる。したがって、基本画像は各数字につき 125 枚、合計 1,375 枚に拡張されることとなる。5 段階でなければならない理由はないが、3.2.3 項に示されたトレーニングデータ数に帰着させることを考慮すると、これが適当である。

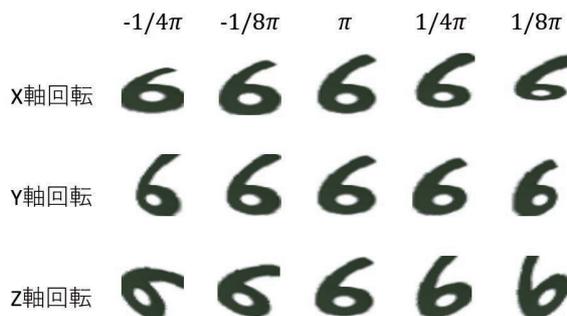


図 4 X, Y, Z 方向へのトレーニングデータの回転

Fig. 4 Rotation of training data around X, Y and Z axes.

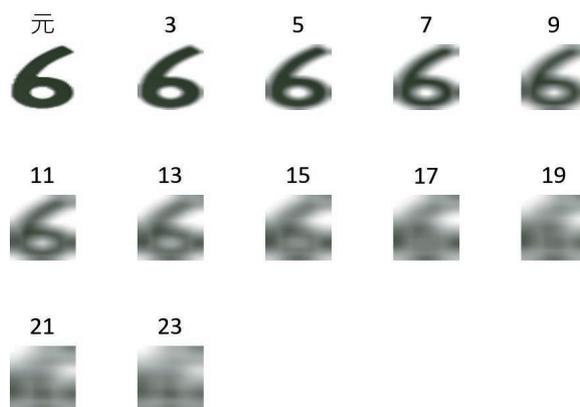


図 5 トレーニングデータの平滑化

Fig. 5 Smoothing of training data.

さらに、これらすべての画像に対しボケの付加が施された画像を作成した。具体的には、平均値フィルタによる平滑化によって処理を行う。平滑化とは、注目画素のその周辺の輝度値を用いて、輝度値の平均を算出し、処理後画像の輝度値とするフィルタリング処理である。カーネルは 3 から 23 の範囲のすべての奇数のものを用いた。平滑化処理について図 5 に示す。図のように元の画像 1 枚につき、カーネルが 3 から 23 の範囲の奇数のものに対応する画像 11 枚が作成される。図中の「6」の上に示す数字がカーネルの値である。すなわち、各数字につき $125 \times 11 = 1,375$ 枚、合計 15,125 枚の画像が作成されることとなる。カーネルサイズは、実際に処理された画像を我々が目視で確認することで、各数字の差異が認識できる程度の上限として決定されたものである。

以上の手順で作成された $1,375 + 15,125 = 16,500$ 枚の画像は、 64×64 ピクセルにリサイズされる。これは、5.1 節に示すテストデータに規格を合わせるためである。リサイズには、バイリニア補間法を用いた。この補間法は、周囲の 4 つの画素を用いるもので、画質の劣化を抑えることに優れている。

なお、角度については、5.1 節で述べるテストデータの角度がおよその範囲に入っていると思われたので、これら 5 段階としている。テストデータとなるナンバープレ-



図 6 トレーニングデータの明度上昇

Fig. 6 Increasing brightness of training data.



図 7 トレーニングデータの明度下降

Fig. 7 Decreasing brightness of training data.

トの撮影時に、角度も記録していれば、トレーニングデータの角度を変えて分類精度の評価を行うことも可能であったが、撮影許可の都合により角度の記録が行えなかったため、トレーニングデータの範囲をこの5段階と固定した。

3.2.3 明度を上げたトレーニングデータ

本研究においてトレーニングデータが何個程度用意されるべきかについては、根拠となるものがなく、また本研究の主な評価の対象とするものでもない。そこで、我々は、トレーニング用 60,000 個、検証用 10,000 個とする MNIST を参考に、各数字について基本トレーニングデータを4倍に拡張した 66,000 個を加え、82,500 個を用意することとした。

まず、拡張されたトレーニングデータとして明度を上げたものを作成した。実際に撮影されたナンバープレートでは白飛びしていることが想定されるため、明度を上げることで白飛びを疑似的に再現するトレーニングデータを考慮する必要がある。しかし、白一色の画像が作成されてしまうと、数字としての特徴を持たない質の悪いデータになってしまう。もちろん、現実の写真が白一色につぶれてしまうことも考えられるが、そもそもその数字は読めない。したがって、白一色の画像が作成されないよう、輝度ではなく明度を用いた。本実験で対象とするナンバープレートは白と緑の2色によって構成されるため、明度特有の、色が異なる部分が同化してしまうという問題は、少なくともトレーニングデータの作成過程においては発生しない。すべての基本トレーニングデータに対して、全画素につき元の明度と 51, 102, 153, 204 との和を新たな明度とした4つの画像を新たに作成した。ただし、明度が 255 を上回った画素は、255 とした。図 6 に明度を上げたトレーニングデータについて示す。図中の「6」の上に示す数字が元の明度に加える値である。

3.2.4 明度を下げたトレーニングデータ

次に、拡張されたトレーニングデータとして明度を下げたものを作成した。実際に撮影されたナンバープレートでは黒つぶれしていることが想定されるため、明度を下げることで黒つぶれを疑似的に再現するトレーニングデータを考慮する必要がある。輝度ではなく明度とした理由は 3.2.3 項で述べたものと同様である。すべての基本トレーニングデータに対して、全画素につき元の明度から 51, 102, 153, 204 を引いたものを新たな明度とした4つの画像を新たに作成した。ただし、明度が 0 を下回った画素は、0 とした。図 7 に明度を下げたトレーニングデータについて示す。図中の「6」の上に示す数字が元の明度から引かれる値であ



図 8 トレーニングデータの切り出し

Fig. 8 Cropping of training data.

る。なお、明度は 0 が下限であるため、それを下回った画素は黒い画素として見える。

3.2.5 一部分を切り出したトレーニングデータ

続いて、拡張されたトレーニングデータとして画像の一部分を切り出したものを作成した。実際に撮影されたナンバープレートでは数字の一部が欠損していることが想定されるため、画像の一部分を切り出すことで、白飛びやヘイズによって数字の一部が欠損した状態を疑似的に再現するトレーニングデータを考慮する必要がある。すべての基本トレーニングデータに対して、左上が 48×48 ピクセルのサイズで切り出された画像を新たに作成した。右上、左下、右下に対しても同様に画像を作成した。これら4種類の画像は、バイリニア補間法によって 64×64 ピクセルにリサイズしている。図 8 に一部分を切り出したトレーニングデータについて示す。図中の「6」の上に、元の画像のどの部分を切り出したか示した。

3.2.6 数字を縮小したトレーニングデータ

最後に、拡張されたトレーニングデータとして数字を縮小したものを作成した。数字の縮小とは画像全体に占める数字部分の割合を小さくすることである。実際に撮影されたナンバープレートでは数字の周囲の余白の大きさにむらがあることが想定されるため、数字を縮小することで、分類対象の画像の切り出しにおける非厳密性を許容したトレーニングデータを考慮する必要がある。ここでの非厳密性とは、数字の周囲の余白が数字に対して同一の割合になるように切り出そうとしても、かなり縮小された数字の周囲の画像を 1 ピクセル単位で切り出さねばならないため、その割合の差は大きくなってしまいうことである。同じ写真であっても、切り出し方の違いによって精度に差が生じることが示されている [2] ため、有効であると考えた。なお、この処理が平滑化の後に行われてしまうと、平滑化された数字の周りに空白が存在するという、現実に即さない画像ができてしまう。そこで、このトレーニングデータの作成のみ、3D 回転による拡張の後に、それらすべてに対して数字を縮小した4種の画像を作成し、それらすべてに対して平滑化を適用した画像を 11 枚作成するという手



図 9 トレーニングデータの縮小
Fig. 9 Shrinking of training data.

順を踏んだ。これらすべての画像は、バイリニア補間法によって 64×64 ピクセルにリサイズしている。作成されるデータ数は他の拡張と同様に 82500 個である。具体的な縮小の割合は、28/30, 26/30, 24/30, 22/30 の 4 種とした。これは、数字が画像の 2/3 以下しか占めないという状況は起こりえないと想定し、その間の大きさを均等に区切ったものである。また、画像サイズの変更の際に端数が生じない値となるようにも考慮している。図 9 に数字を縮小したトレーニングデータについて示す。図中の「6」の上に示す数字が縮小の割合である。図のように数字部分を小さくすると余白が大きくなる。この作業は、テストデータとなる画像をナンバープレートの写真から切り出す際に、人間にとって判別が困難な数字であるということは、数字と余白の境界も曖昧である可能性が高いことに起因している。言い換えると、写真に小さく写っている数字が、ボケて大きく見えている可能性があるということである。よって、この作業で小さいサイズの数字を追加で作成した後に、3.2.2 項で述べた平滑化の処理を行ってトレーニングデータの画像をボケさせている。

3.3 CNN

我々は、機械学習のネットワークとしては、画像分類によく用いられている畳み込みニューラルネットワーク (CNN) を選択し、評価のための CNN に ResNet を使用した。ResNet は、2015 年に He らによって提案された [20]。これは、誤差が上層から下層へ伝播するにつれて誤差の値が 0 に近づいてしまい、下層のパラメータのトレーニングが進まなくなるという問題を改善したものである。ResNet は画像分類の分野において広く使用されているネットワークであり、これをベースとして改良したネットワークに関する論文も発表されている [21], [22], [23]。本研究においては、ネットワークではなくトレーニングデータを主な評価対象とするため、最も一般的な ResNet50 を用いることとした。実際には、トレーニングデータのうち 9 割がトレーニングに、1 割が検証に使用される。また、バッチサイズは我々の動作環境で設定できる最大値であった 256 とした。ソフトマックス層によって出力されるユニット数は、0 から 9 の数字および中黒を対象とするため 11 である。数字に対応したラベルを与える際には、中黒のラベルは「10」とした。エポック数は上限を設けず、アーリーストッピングを導入した。これは、過学習が進む前にトレーニングを中断させることを目的とするものである。本評価において

は、検証時の損失関数の値 (Loss) を監視し、下降が認められなくなった場合に学習が停止されることとした。具体的には、Patience を 10 エポックとし、10 エポック以上 Loss の下降が認められない場合に、その時点で Loss が最小だったエポックのモデルが採用される。

4. 実装

4.1 トレーニングデータ

3.2 節に示したほぼすべての画像処理は、Python3 によって実装した。画像の読み込みや保存、変換などの処理には、OpenCV を利用している。三次元回転には Java 3D を用いた。

4.2 CNN

CNN の実装には、機械学習のためのソフトウェアライブラリである TensorFlow と、TensorFlow 上で実行可能なニューラルネットワークライブラリの 1 つである Keras を用いた。TensorFlow のバージョンは 1.12.0, Keras のバージョンは 2.2.4 である。Keras の Model クラス API の中に定義されている predict() メソッドによって、テストデータに対するラベルごとの予測値が確率として出力される。

5. 評価手法

5.1 テストデータの作成

評価のためのテストデータとして、各数字につき 100 個、合計 1100 個を用意した。この数としたのは、百分率で分析する際に刻み幅が 1% ずつになるので、計算・比較がしやすいからである。当然であるが、すべてのテストデータは異なるものあり、同じ写真の同じ部分から複数個のテストデータを作成していない。

本手法は、人間に数字の判別が困難なナンバープレートの画像から、コンピュータを用いてその数字を読み取るものである。これは、日本のナンバープレートが、フォントが固定された数字 4 桁で表現されているという特殊な事情を考慮してのものである。該当するものが数字 10 通りおよび中黒のみであり、フォントも 1 種類しかないため、画像の状態によっては高い精度で予測できる場合がある。利用目的は、裁判における証拠や容疑者の自動車を特定する手段である。警察が集めるナンバープレートの画像の中で、人間に数字の判別が困難なものとしては、防犯カメラに映ったもの、車載にカメラに映ったもの、スマートフォンで撮影された写真の背景に映ったものなどが考えられる。

大前提として、テストデータの撮影は日本のナンバープレートが装着された車両に対して行う必要がある。そして、大学による撮影許可の都合により、本実験を通して撮影可能な車両はすべて本学園のスクールバスである。そのため、バスやその他の活動に支障のないように迅速に撮影しなければならないという制限が存在した。防犯カメラで

あれば、通常は高い位置から撮影されていて、画像の圧縮率が高い。車載カメラやスマートフォンであれば高さは被写体より少し上程度である。今回撮影した画像は、撮影許可の都合もあり、被写体との距離や角度は厳密に設定されていないが、おおむね車両の正面に対して、スマートフォンで左右45度ずつの範囲に収まるように撮影した。よって、上下の傾きについては考慮することができなかったが、左右の傾きは考慮できており、傾きの評価はこれで代用することとした。また、画像サイズはそれほど小さくないが暗すぎて読めないものや、同じく画像サイズはそれほど小さくないが監視カメラによる画像圧縮によって劣化しすぎて読めないものも考えられるが、今回の評価対象は、小さく写っているものとした。カメラは、iPhone 7に備え付けられているものを用いた。多くの場合、監視カメラにフラッシュやその他の光源が装備されていないため、フラッシュは使用していない。

撮影において、連写機能などは用いず、1枚1枚の写真は、被写体との距離や角度を少しずつ変更しながら撮影している。なお、撮影されたバスのナンバープレートの仕様は、一般の乗用車のものと同様である。また、本実験においては、遮蔽物による影響を排除するため、撮影日として降水のない日を選んでおり、それ以外の制限は設けていない。すなわち、各数字について、撮影時の光量や光源の位置を揃えるための工夫はしていない。スクールバスは使用される車両が日によって変わるため、およそ20台のバスを数日かけて撮影した。撮影時の距離および角度については、撮影許可の関係で取得することができなかった。

本評価の目的は、実際の監視カメラや車載カメラで撮影されたナンバープレートの数字を高精度で読めることを示すことではなく、人間に数字の判別が困難なナンバープレートの画像から、裁判の証拠や容疑者の自動車の絞り込みに必要な精度をもって、コンピュータが自動的に数字を分類できる可能性を示すことである。また、トレーニングにおいて実画像のデータセットを必要としないことを特徴としているため、どのような画像をトレーニングデータに加えることで、その精度がどのように変わるのか示すことである。そのため今回は、実際に設置されている防犯カメラの使用、動画データからのテストデータの抽出、走行中の撮影などについては評価の対象外とし、スマートフォンで撮影された際に小さすぎて読めないもののみを扱うこととした。

撮影した写真の例を図10に示す。図10のような写真から、ナンバープレートを目視で確認し、一連指定番号の各数字を手動でトリミングした。トリミングの範囲は目視であり、図2に示したトレーニングデータと極力同じになるよう、範囲を調整してトリミングしている。なお、遮蔽物によって数字の半分以上が隠蔽されている写真や、過剰な改変が行われた写真は本実験において使用していない。



図10 テストデータに使用される写真の例
Fig. 10 Example of pictures for test data.

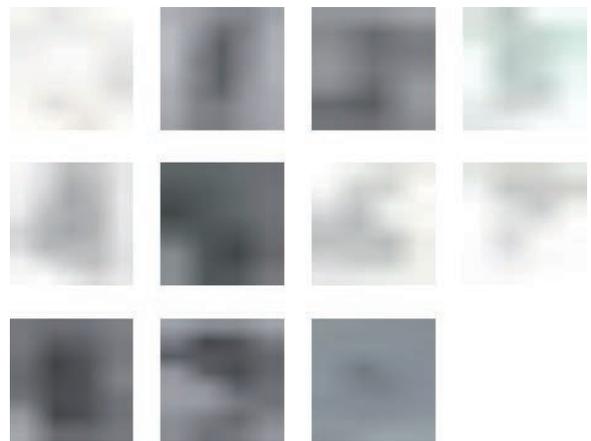


図11 テストデータの例
Fig. 11 Examples of test data.

トリミングされた画像の最小サイズは2×5ピクセル、最大サイズは27×47ピクセル、平均サイズは5.03×9.11ピクセルであった。作成されたテストデータは、ネットワークに入力する前に、サイズ統一のためにバイリニア補間法を用いて64×64ピクセルにリサイズしている。リサイズされたテストデータの例を図11に示す。トリミングされた時点の各数字の画像サイズが様々であるため、人間が数字を推測可能な画像も含まれてはいるが、図11の画像のように基本的には人間が確信を持って判読できない状態である。本研究の主目的は、ナンバープレートの各数字を確実に判別することではなく、どのようなトレーニングデータセットを用意すれば効率良く精度を上げられるかであるため、テストデータの画像サイズをこのようにした。ただし、デジタルフォレンジックに使用することを最終的な目標としているため、条件によっては良い精度となることを期待している。

5.2 予測値の確率の算出

テストデータのセットについてテストが行われ、ラベルごとの予測値が確率として出力される。つまり、どのラベ

ルの数字にどれくらいの確率で分類されるかが各数字について出力される。

以上は、すべてのトレーニングデータセット（基本トレーニングデータ・明度を上げたもの・明度を下げたもの・一部分を切り出したもの・数字を縮小したもの）について行われる。トレーニングおよびテストはそれぞれ5回ずつ行い、各回の平均値と標準偏差を求めている。

6. 評価結果

紙面の都合上、以降に示される表では、基本トレーニングデータ、明度を上げたもの、明度を下げたもの、一部分を切り出したもの、数字を縮小したものを、順に basic, bright, dark, cropped, downsized と表記した。また、中黒は 3.3 節で述べたラベルに合わせ「10」と表記している。

6.1 基本トレーニングデータ

基本トレーニングデータに関する予測値の確率の平均値を表 1 に、その標準偏差を表 2 に示した。これらを見ると、平均値が 0.4 を上回るものが存在しない。なかでも、「0」に関しては 5 回とも 0.1 を下回っており、標準偏差も他の数字と比べて小さい。「6」「8」「10」の 3 つが比較的高い平均値を示している。

6.2 明度を上げたトレーニングデータ

明度を上げたトレーニングデータに関する予測値の確率の平均値を表 3 に、その標準偏差を表 4 に示した。平均値は、basic と比べて上昇が認められる。ただし、「10」は値が低くなっている。「0」「1」「7」が比較的高い値を示しており、basic と比べて傾向の変化が認められる。

6.3 明度を下げたトレーニングデータ

明度を下げたトレーニングデータに関する予測値の確率の平均値を表 5 に、その標準偏差を表 6 に示した。平均値は、basic と比べて上昇が認められる。「0」「1」「7」が比較的高い値を示しており、この点では bright と同様である。しかし、「3」や「10」が比較的高い値を示している点は異なる。

6.4 一部分を切り出したトレーニングデータ

一部分を切り出したトレーニングデータに関する予測値の確率の平均値を表 7 に、その標準偏差を表 8 に示した。basic と比べた場合、平均値の上昇は認められない。平均値、標準偏差ともに傾向の変化も認められない。

6.5 数字を縮小したトレーニングデータ

数字を縮小したトレーニングデータに関する予測値の確率の平均値を表 9 に、その標準偏差を表 10 に示した。「1」「5」「6」について、basic から平均値の上昇が認められる。

表 1 予測の平均値 (basic)

Table 1 Mean values by prediction (basic).

	1 回目	2 回目	3 回目	4 回目	5 回目
0	0.000332719	1.33E-13	0.002856221	2.17E-06	0.017160163
1	0.138121172	0.129007437	0.144055518	0.168185675	0.321274635
2	0.112781292	0.051589367	0.089266589	0.09087206	0.197771501
3	0.167825945	0.123129299	0.060612833	0.164653214	0.295126618
4	0.125979559	0.161563842	0.164938149	0.124594522	0.178838672
5	0.146576356	0.06984486	0.121321938	0.118633574	0.134657351
6	0.3289646	0.235260406	0.279008037	0.222636786	0.162583976
7	0.131098889	0.065123275	0.166317733	0.065845375	0.111641723
8	0.356742873	0.398505299	0.371152433	0.364157639	0.35436383
9	0.150802685	0.14165956	0.127284232	0.135929297	0.158239589
10	0.291618808	0.278649824	0.226113247	0.266476549	0.198340971

表 2 予測の標準偏差 (basic)

Table 2 Standard deviation values by prediction (basic).

	1 回目	2 回目	3 回目	4 回目	5 回目
0	0.001375967	1.09E-12	0.015199872	7.62E-06	0.09350983
1	0.32268614	0.30706605	0.33086064	0.34659323	0.43830273
2	0.29753372	0.19636041	0.26328638	0.26233748	0.39143586
3	0.34228835	0.30256578	0.23060167	0.32976475	0.4425449
4	0.2977223	0.33691132	0.33626246	0.2952156	0.34738737
5	0.3386555	0.23265846	0.29409182	0.31073168	0.323083
6	0.44274908	0.39629427	0.4301967	0.40403008	0.3566395
7	0.29193258	0.21891019	0.34920532	0.21822684	0.29559076
8	0.46754855	0.47256142	0.47152954	0.46824408	0.47224775
9	0.33116898	0.30730334	0.30132648	0.31768712	0.32728204
10	0.43657795	0.43750653	0.4099981	0.4333548	0.3837281

表 3 予測の平均値 (bright)

Table 3 Mean values by prediction (bright).

	1 回目	2 回目	3 回目	4 回目	5 回目
0	0.885467774	0.781527108	0.782641565	0.767235802	0.888248329
1	0.704474855	0.477018792	0.608285357	0.389586142	0.461595825
2	0.448324852	0.418225562	0.480408213	0.388403543	0.625923645
3	0.461617079	0.525706646	0.564290301	0.489347755	0.691551974
4	0.475868968	0.74417568	0.518386482	0.540361696	0.496927078
5	0.597663704	0.692434776	0.601110427	0.606268826	0.58297348
6	0.512702284	0.694490342	0.660788603	0.654966148	0.598041687
7	0.707493472	0.554478508	0.515476549	0.374205447	0.557810138
8	0.411655726	0.58792543	0.688969108	0.779845142	0.68445396
9	0.525489048	0.858157177	0.730011419	0.829822566	0.598840877
10	0.065767627	0.001034315	0.039892278	0.049520494	0.015034643

表 4 予測の標準偏差 (bright)

Table 4 Standard deviation values by prediction (bright).

	1 回目	2 回目	3 回目	4 回目	5 回目
0	0.2599672	0.36381868	0.35551476	0.37913716	0.24857531
1	0.4016209	0.42325252	0.44455364	0.4201646	0.4446948
2	0.4290306	0.4086632	0.44626787	0.40079328	0.44348288
3	0.45437077	0.42299765	0.4217851	0.40218958	0.36654335
4	0.46226007	0.37937987	0.4356311	0.4324184	0.4141118
5	0.43981764	0.41475198	0.44892764	0.4430156	0.45082593
6	0.4293759	0.3882268	0.41319484	0.38679266	0.38957188
7	0.39192295	0.4415962	0.42867464	0.39677647	0.42035726
8	0.45284477	0.44882843	0.38964477	0.3605807	0.408657
9	0.41192657	0.31676704	0.37785718	0.3178779	0.42026174
10	0.2404232	0.007082158	0.1941223	0.19824432	0.10133448

表 5 予測の平均値 (dark)

Table 5 Mean values by prediction (dark).

	1 回目	2 回目	3 回目	4 回目	5 回目
0	0.793400613	0.355035829	0.266639441	0.447707308	0.266796451
1	0.877783054	0.986123254	0.973951237	0.968882421	0.690117931
2	0.257604779	0.088593325	0.098942396	0.107703573	0.205537392
3	0.538201717	0.49720198	0.252361015	0.818112616	0.373588796
4	0.286789276	0.286659931	0.362957385	0.27283929	0.350619173
5	0.178212405	0.183436555	0.143435147	0.226280229	0.176122075
6	0.480930463	0.159574001	0.249151047	0.349451057	0.268977783
7	0.621833387	0.322232112	0.26996318	0.823815752	0.498316111
8	0.182534529	0.308954438	0.228647869	0.220159849	0.33950193
9	0.05468643	0.243438131	0.237613873	0.170908499	0.185096292
10	0.429465	0.312802065	0.28158792	0.366861721	0.46416433

表 6 予測の標準偏差 (dark)

Table 6 Standard deviation values by prediction (dark).

	1 回目	2 回目	3 回目	4 回目	5 回目
0	0.30426076	0.39785329	0.36761007	0.4420777	0.36580023
1	0.28320956	0.09464898	0.13448146	0.16553174	0.448579
2	0.4111812	0.27044633	0.2713725	0.28483775	0.38506624
3	0.45413342	0.45967755	0.4072997	0.35100478	0.4479972
4	0.4179321	0.42920962	0.44955012	0.4154537	0.45090523
5	0.34213734	0.3543622	0.3090551	0.36521858	0.3566579
6	0.43148255	0.31587258	0.3896864	0.4012825	0.3750761
7	0.39583534	0.43226624	0.4177004	0.32937974	0.4322603
8	0.33161584	0.42050445	0.39666632	0.38488683	0.41755855
9	0.20181169	0.39734936	0.3510173	0.33941218	0.35786706
10	0.4531559	0.4260766	0.40929213	0.4476758	0.47586882

表 7 予測の平均値 (cropped)

Table 7 Mean values by prediction (cropped).

	1 回目	2 回目	3 回目	4 回目	5 回目
0	5.06E-06	1.29E-07	0.007859571	0.011500751	0.000154046
1	0.334113014	0.236126169	0.35418811	0.372251343	0.293588187
2	0.123004957	0.179458068	0.113968127	0.153955408	0.117047619
3	0.061110631	0.086909629	0.089643905	0.06066839	0.073146237
4	0.133150275	0.311681188	0.261536226	0.140498558	0.277841864
5	0.099576224	0.126583026	0.078376489	0.093455817	0.088198341
6	0.218526005	0.266339567	0.3718805	0.30408918	0.324255243
7	0.046308197	0.061208562	0.111684499	0.233477872	0.235912556
8	0.253128201	0.107455918	0.167896282	0.165600171	0.260048461
9	0.062344664	0.06403523	0.085774182	0.068138043	0.049366677
10	0.229969976	0.238758261	0.196308504	0.184462599	0.236795402

表 8 予測の標準偏差 (cropped)

Table 8 Standard deviation values by prediction (cropped).

	1 回目	2 回目	3 回目	4 回目	5 回目
0	5.02E-05	1.27E-06	0.06531637	0.09901109	0.001505317
1	0.4496415	0.39487326	0.4592101	0.4566526	0.4045133
2	0.3101996	0.36915758	0.2947911	0.34849772	0.3048112
3	0.2265467	0.25222328	0.26460373	0.2147295	0.23720257
4	0.3204767	0.41987	0.3825078	0.32331112	0.39212757
5	0.2709918	0.31839275	0.25553676	0.27256852	0.2664804
6	0.39741725	0.41923317	0.44394657	0.4063048	0.42958215
7	0.18333912	0.20463029	0.2781583	0.37683636	0.37386134
8	0.41009086	0.25506487	0.34600425	0.35057157	0.4269773
9	0.20588367	0.22331901	0.25867987	0.22526368	0.19840264
10	0.41394436	0.4225929	0.38373068	0.38314673	0.42061713

表 9 予測の平均値 (downsized)

Table 9 Mean values by prediction (downsized).

	1 回目	2 回目	3 回目	4 回目	5 回目
0	0.009941185	0.010265044	0.030494037	0.04773124	0.068630984
1	0.274128702	0.396219133	0.179849228	0.432022153	0.43919548
2	0.113271766	0.461084123	0.086118195	0.118248784	0.141015155
3	0.270230759	0.454476808	0.314263583	0.317938457	0.366077228
4	0.10710252	0.193883307	0.252045906	0.274418757	0.138902566
5	0.339580605	0.172609789	0.168368962	0.133386971	0.16554753
6	0.798401293	0.73495051	0.57183565	0.519326278	0.528090488
7	0.232672234	0.416489557	0.49120401	0.236417182	0.349492099
8	0.388457189	0.402980909	0.437837138	0.421004819	0.393950819
9	0.096049783	0.137257663	0.149403981	0.086845952	0.138023155
10	0.161272211	0.22798155	0.160883761	0.213014758	0.160855105

表 10 予測の標準偏差 (downsized)

Table 10 Standard deviation values by prediction (downsized).

	1 回目	2 回目	3 回目	4 回目	5 回目
0	0.09519938	0.060412176	0.14594738	0.18204905	0.2203964
1	0.42852262	0.46910253	0.34932932	0.47846004	0.4711788
2	0.28524595	0.43256748	0.2559107	0.29195327	0.31593624
3	0.42295587	0.4845064	0.3992455	0.42704073	0.44777906
4	0.2955713	0.36334097	0.3900415	0.40040025	0.30822936
5	0.43022794	0.34889346	0.31607297	0.31145903	0.34307313
6	0.36072764	0.40777877	0.4337681	0.44121787	0.46524322
7	0.39617833	0.428011	0.4426034	0.40503898	0.4278251
8	0.47972354	0.47222352	0.48177695	0.47967547	0.48204875
9	0.27049735	0.321887	0.34172574	0.26415357	0.32158816
10	0.3657388	0.3905135	0.36609492	0.38606623	0.3662857

それ以外は、平均値・標準偏差ともに改善は認められない。

7. 考察

本章では、6章に示した評価について考察を行う。

7.1 評価結果に関する考察

6.1節で、表1の値において、平均値が0.4を上回るものが存在しないと記した。本手法の結果を裁判における証拠や容疑者の自動車を特定する手段として用いる場合、予測値の値がいくつであるかは大きく影響する。他の証拠や証言と合わせるため、一概にいくつの値が境界値であるということはできないが、たとえば、40%で「1」と判別された数字が決定的な証拠になるとは考えにくい。表1の結果では、最大の数字でも平均値が0.4を上回るものが存在しないため、いずれの数字も前述の用途として使用するには不十分であると考えられる。

6.1節のbasicの結果において、「0」の平均値が低く「6」「8」「10」の平均値が高くなった理由に関して考察する。予測の傾向について確認するために、1回目のトレーニングについて、確率が最も高かった予測値と真値に関する混同行列を作成した。その混同行列を表11に示す。

表11を見ると、「10」であると予測される可能性が他の数字よりも高い。つまり、「10」が分類しやすい数字であ

表 11 混同行列 (basic)
Table 11 Confusion matrix (basic).

		真										
		0	1	2	3	4	5	6	7	8	9	10
予測	0	0	2	4	0	0	4	2	0	7	0	48
	1	0	14	0	0	15	0	1	2	26	0	0
	2	0	0	11	0	1	0	2	10	0	0	3
	3	0	1	0	15	11	0	0	6	0	48	0
	4	0	2	1	1	14	0	0	0	0	0	0
	5	0	0	0	1	0	14	0	0	6	2	3
	6	0	2	2	0	4	1	34	0	0	0	0
	7	0	2	0	1	5	0	0	12	9	0	0
	8	0	32	26	30	20	19	0	5	37	12	17
	9	0	1	1	14	0	4	0	0	0	16	0
	10	100	44	55	38	30	58	61	65	15	22	29

表 12 Wilcoxon の符号付順位和検定における p 値
Table 12 p-values of Wilcoxon signed-rank test.

basic	0.00
bright	0.00
dark	0.00
cropped	1.00
downsized	0.00

るというよりは、トレーニングデータが不十分または不適切である場合、中黒であっても他の数字であっても、「10」であると予測される傾向があるということである。「8」についても類似の傾向が見られ、また「0」は逆に低い可能性として同様の傾向が見られる。これらの原因は、CNN において「10」や「8」に分類されやすく「0」に分類されにくいモデルが生成されてしまったためであると推測され、トレーニングデータに改善が必要であることが示唆されている。一方、「6」にそのような傾向は認められない。これは、「6」という数字が他の数字と異なった特徴を持っており、同じ条件のトレーニングデータでも分類がうまくいくのではないかと推測される。

Shapiro-Wilk の正規性検定によって、それぞれの平均値の正規性を調べた。その結果、いずれも p 値は 0 であり、正規性が認められなかった。そこで、Wilcoxon の符号付順位和検定を用いて、basic 以外の 4 つの平均値が、それぞれ basic の平均値に対して有意に大きいかを確認した。basic 以外の 4 つの平均値が、それぞれ basic の平均値と有意差がないこと帰無仮説とし、basic 以外の 4 つの平均値が、それぞれ basic の平均値より大きくなることを対立仮説として片側仮説検定を行った。それぞれの p 値を表 12 に示す。有意水準を 0.05 とすると、bright, dark, downsized については帰無仮説が棄却され、basic に対して有意に平均値が大きいといえる。cropped については帰無仮説が採択され、basic の平均値との間に有意差は認められない。

basic と他 4 つを比較した場合に、主に有意な改善が認められたのは bright と dark であった。このことから、明度

表 13 正しく分類された明るい画像の割合
Table 13 Percentages of bright images predicted correctly.

	1 回目	2 回目	3 回目	4 回目	5 回目
basic	15.61%	14.15%	13.10%	12.57%	17.20%
bright	64.15%	68.78%	71.30%	67.33%	70.11%
dark	41.53%	28.57%	24.60%	39.55%	26.59%
cropped	10.98%	13.23%	17.33%	16.80%	15.74%
downsized	30.95%	35.19%	29.89%	30.56%	29.37%

表 14 正しく分類された暗い画像の割合
Table 14 Percentages of dark images predicted correctly.

	1 回目	2 回目	3 回目	4 回目	5 回目
basic	22.67%	18.90%	22.09%	22.09%	23.55%
bright	33.72%	38.08%	29.65%	29.94%	29.07%
dark	50.87%	46.80%	44.77%	53.20%	54.65%
cropped	22.09%	20.35%	14.53%	15.12%	22.97%
downsized	13.08%	27.91%	19.19%	14.83%	20.64%

の変更を施したトレーニングデータを加えることは有効であるといえる。逆にいえば、画像の明るさが CNN にとっての分類しやすさに与える影響が大きいことになる。

そこで我々は、この点について考察を深めるために、テストデータを「明るい画像」と「暗い画像」に分けて調べた。トレーニングデータ作成に使用された基本画像の平均明度を基準に、テストデータの画像のうち平均明度が高いものを「明るい画像」、小さいものを「暗い画像」とした。そのうえで、確率が最も高かった予測値が真値と一致していた場合を「正しく分類された」とし、その割合を算出した。明るい画像の結果を表 13 に、暗い画像の結果を表 14 に示す。

表 13 を見ると、bright が最も改善されている。dark と downsized に関しても改善が見られる。

また、表 14 を見ると、dark が最も改善されている。bright に関しても改善が見られるが、dark と bright のどちらも「明るい画像」の場合ほどの影響はない。downsized に関してはまったく改善が見られない。

以上の点からいえることは、トレーニングデータは bright, dark, downsized の順に有効であること、「暗い画像」は分類が比較的困難であることである。後者について、「明るい画像」は比較的単純な劣化であるのに対し、「暗い画像」は比較的原因が複雑であることが、理由として考えられる。具体的な原因として、画像全体の輝度の低下、ボケによる数字部分の膨張、ノイズによる劣化などがあげられる。

次に、ほとんど改善が見られなかった cropped について考察する。予測の傾向に変化があるかを確認するために、確率が最も高かった予測値と真値に関して、表 11 と同様に 1 回目のトレーニングに関する混同行列を作成した。その混同行列を表 15 に示す。

表 11 と表 15 を見比べると、数値が高い箇所はほぼ同じであり、「8」や「10」に誤分類されるものが他の数字よ

表 15 混同行列 (cropped)
Table 15 Confusion matrix (cropped).

		真										
		0	1	2	3	4	5	6	7	8	9	10
予測	0	0	0	3	0	0	0	0	2	1	0	49
	1	0	33	0	0	33	0	1	4	37	4	4
	2	0	1	13	0	0	1	0	2	0	1	2
	3	0	0	1	6	0	0	0	4	0	11	0
	4	0	17	1	2	13	5	0	0	17	0	0
	5	0	0	0	0	0	12	0	0	0	2	0
	6	0	0	0	0	1	1	21	0	2	0	0
	7	0	0	0	0	0	0	0	5	0	0	6
	8	0	18	27	38	16	22	0	5	26	16	16
	9	0	0	0	4	1	0	0	4	0	7	0
	10	100	31	55	50	36	59	78	74	17	59	23

り多い点も同様である。つまり、トレーニングデータ数の増加と、それにとまなうトレーニング時間の増加を考慮すると、一部分が切り出された画像をトレーニングデータに加えても効果は小さいと考えられる。

7.2 実際の運用における課題

本研究を実際に使用する場合、テストにおいて読み取りたい数字の予測値が高ければ、それはデジタルフォレンジックや捜査対象の絞り込みに使用できる。予測値が低い場合であっても、それにより誤認逮捕や冤罪につながるわけではなく、その画像の証拠能力が低かったというだけである。なお、確率は低いと思われるが、誤分類により、何人かの容疑者のうち、真犯人ではない容疑者の自動車のナンバープレートと偶然数字が高い予測値で一致してしまった場合には、冤罪が起ころう。このような冤罪を防ぐためには、類似の条件下の異なるテストデータを用いた場合に、予測値の平均が高くなければ証拠能力を低くするなどの工夫が必要である。

8. おわりに

我々は、人間にも判別が困難なナンバープレートの数字を、機械学習によって推測する方法を検討し、実験を行った。手法としては、トレーニング用の画像を大量に撮影する手間を省くため、CGを用いてこれを作成した。CGであるため、トレーニング用の画像を無制限に作り出すことが可能であるが、その中には効果の少ないものや逆効果になるものがあるのではないかと推測し、どのような種類の画像がトレーニング用の画像として適切であるか実験を行い評価した。結果として、明るくした画像や暗くした画像が最も有効であることが分かった。また、数字の大きさを変えた画像も有効に作用することが示唆された。一方で、画像の一部分を切り出した画像は、有効に働かない可能性が高いという知見が得られた。

明度を上げたトレーニングデータにおいては、「0」が最大平均 88.8%、「9」が最大平均 83.0%の予測値を示した。

また、明度を下げたトレーニングデータにおいては、「1」が最大平均 98.6%の予測値を示した。しかし、予測値が低いものも多かったため、精度の向上が今後の課題の1つである。また、実際の使用を想定した条件下で実験を行うことも求められる。さらに、トレーニングデータだけでなく、ネットワークやパラメータ、テストデータの加工などについても、最適な手法を探っていくことも必要であると考えている。今回、評価対象がスマートフォンで撮影され、小さくて読めないナンバープレートに限られたが、将来的には実際に防犯カメラで撮影した画像も評価したい。ただし、降雨、モーション・ブラー、ハードウェアの差など考慮すべき点は多々ある。

参考文献

- [1] 吉浦紀見, 加藤蒼悟, 田北啓洋, 太田直哉, 藤井雄作: 通学路への防犯カメラの導入に関するアンケート結果の分析, 情報処理学会論文誌, Vol.59, No.3, pp.1106–1118 (2018).
- [2] Suzuki, T. and Uda, R.: Classifying License Plate Numerals Using CNN, *Proc. 13th International Conference on Ubiquitous Information Management and Communication (IMCOM 2019)*, Lee, S., Ismail, R. and Choo, H. (Eds.), pp.1063–1075, Cham, Springer International Publishing (2019).
- [3] Patel, C., Shah, D. and Patel, A.: Automatic Number Plate Recognition System (ANPR): A Survey, *International Journal of Computer Applications (IJCA)*, Vol.69, pp.21–33 (online), DOI: 10.5120/11871-7665 (2013).
- [4] Sonavane, K., Soni, B. and Majhi, U.: Survey on Automatic Number Plate Recognition (ANR), *International Journal of Computer Applications*, Vol.125, pp.1–4 (online), DOI: 10.5120/ijca2015905920 (2015).
- [5] Albiol, A., Sanchis, L., Albiol, A. and Mossi, J.M.: Detection of Parked Vehicles Using Spatiotemporal Maps, *IEEE Trans. Intelligent Transportation Systems*, Vol.12, No.4, pp.1277–1291 (2011).
- [6] Fan, T.: A vehicle recognition method based on multi-camera binding, *2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, pp.1549–1553 (2019).
- [7] Wang, C.-M. and Liu, J.-H.: License plate recognition system, *2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, pp.1708–1710 (online), DOI: 10.1109/FSKD.2015.7382203 (2015).
- [8] Amirgaliyev, B.Y., Kenshimov, C.A., Kuvatov, K.K., Kairanbay, M.Z., Baibatyr, Z.Y. and Jantassov, A.K.: License plate verification method for automatic license plate recognition systems, *2015 Twelve International Conference on Electronics Computer and Computation (ICECCO)*, pp.1–3 (online), DOI: 10.1109/ICECCO.2015.7416892 (2015).
- [9] Ashtari, A.H., Nordin, M.J. and Fathy, M.: An Iranian License Plate Recognition System Based on Color Features, *IEEE Trans. Intelligent Transportation Systems*, Vol.15, No.4, pp.1690–1705 (online), DOI: 10.1109/TITS.2014.2304515 (2014).
- [10] Rashid, A.E.: A fast algorithm for license plate detection, *2013 International Conference on Signal Processing, Image Processing Pattern Recognition*, pp.44–48

- (online), DOI: 10.1109/ICSIIPR.2013.6497956 (2013).
- [11] Yang, Y., Gao, X. and Yang, G.: Study the Method of Vehicle License Locating Based on Color Segmentation, *Procedia Engineering*, Vol.15, pp.1324–1329 (online), DOI: <https://doi.org/10.1016/j.proeng.2011.08.245> (2011).
- [12] 羽田光太朗, 花泉 弘: ナンバープレート自動認識システムにおける数字抽出手法の検討, 第75回全国大会講演論文集, Vol.2013, No.1, pp.449–450 (2013).
- [13] Cui, Y. and Huang, Q.: Extracting characters of license plates from video sequences, *Machine Vision and Applications*, pp.308–320 (online), DOI: <https://doi.org/10.1007/s001380050081> (1998).
- [14] Wang, N., Zhu, X. and Zhang, J.: License Plate Segmentation and Recognition of Chinese Vehicle Based on BPNN, *2016 12th International Conference on Computational Intelligence and Security (CIS)*, pp.403–406 (online), DOI: 10.1109/CIS.2016.0098 (2016).
- [15] Xing, J., Li, J., Xie, Z., Liao, X. and Zeng, W.: Research and Implementation of an Improved Radon Transform for License Plate Recognition, *2016 8th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, Vol.01, pp.42–45 (online), DOI: 10.1109/IHMSC.2016.52 (2016).
- [16] Nagare, A. and Bhatia, S.: License Plate Character Recognition System using Neural Network, *International Journal of Computer Applications*, Vol.25, p.4 (online), DOI: 10.5120/3147-4345 (2011).
- [17] Damak, T., Kriaa, O., Baccar, A., Ben Ayed, M.A. and Masmoudi, N.: Automatic Number Plate Recognition System based on Deep Learning, *International Journal of Computer and Information Engineering*, Vol.14, pp.86–90 (2020).
- [18] 神宮彩実, 太田直哉: 防犯カメラで撮影された低解像度ナンバープレートの数字識別, 第73回全国大会講演論文集, Vol.2011, No.1, pp.527–528 (2011).
- [19] Spanhel, J., Sochor, J., Juránek, R., Herout, A., Marsík, L. and Zemčík, P.: Holistic recognition of low quality license plates by CNN using track annotated data, *2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp.1–6 (2017).
- [20] He, K., Zhang, X., Ren, S. and Sun, J.: Deep Residual Learning for Image Recognition, *CoRR*, Vol.abs/1512.03385 (online), available from (<http://arxiv.org/abs/1512.03385>) (2015).
- [21] Li, B. and He, Y.: An Improved ResNet Based on the Adjustable Shortcut Connections, *IEEE Access*, Vol.6, pp.18967–18974 (2018).
- [22] Hsiao, S. and Jang, J.R.: Improving ResNet-based Feature Extractor for Face Recognition via Re-ranking and Approximate Nearest Neighbor, *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp.1–8 (2019).
- [23] Wen, L., Zhao, Y., Li, S., Cheng, H. and Zhang, C.: MST-ResNet: A Multiscale Spatial Temporal ResNet for Steering Prediction, *2019 2nd China Symposium on Cognitive Computing and Hybrid Intelligence (CCHI)*, pp.246–251 (2019).



鈴木 友哉

2019年東京工科大学コンピュータサイエンス学部コンピュータサイエンス学科卒業。現在、同大学大学院バイオ・情報メディア研究科コンピュータサイエンス専攻博士前期課程在学中。情報セキュリティの研究に従事。



宇田 隆哉 (正会員)

1998年慶應義塾大学理工学部計測工学科卒業。2000年同大学大学院理工学研究科計測工学専攻前期博士課程修了。2002年同大学院理工学研究科開放環境科学専攻後期博士課程修了。博士(工学)。現在、東京工科大学コンピュータサイエンス学部講師。ネットワークセキュリティの研究に従事。2002年IFIP/SEC 2002 Best Student Paper Award受賞。電子情報通信学会会員。