

テストデータの特徴と計測誤差を用いた判読困難なナンバープレートの自動判別

松澤 南到¹ 宇田 隆哉¹

概要: 本論文では, 人間に判読困難なナンバープレートの数字を自動で判別もしくは候補を絞ることを目的としている. 人間に判読困難なナンバープレートを対象とした研究には, 実用性に欠けていることや, トレーニングデータにテストデータの画像と特徴が異なる画像が多く含まれていることにより, 判別精度が低いなどといった問題がある. そこで本論文では, 人間に判読困難なナンバープレートが写った車両の画像(対象画像)から得られる角度や明度など, 画像全体から得られる特徴も利用してトレーニングデータを作成することで, トレーニングデータが網羅的になることを防ぎ, 対象画像に適したデータセットの作成を行う手法を提案した. その結果, 数字によって判別精度に差があるものの, 精度向上に対して有効に働いていることが分かった. また, 判読困難であることを客観的に評価し, 判読困難な画像の目視による判断の危険性についても示唆した.

Automatic Classification of Unreadable License Plates Using Test Data Features and Measurement Errors

MINATO MATSUZAWA¹ RYUYA UDA¹

1. はじめに

現在, 世界各国の様々な場所に防犯カメラが設置されており, カメラに映った車のナンバープレートから数字などを読み取り, 車を特定するという犯罪捜査が行われている. しかし, 例え防犯カメラにナンバープレートが映っていたとしても, 画像のボケや光量の偏差, 防犯カメラの設計などが原因で, ナンバープレートが人間に判読困難な状態になっている場合がある. また, ナンバープレート認識に関する研究は世界各国で行われているが, 多くの研究ではこのようなナンバープレートは対象外とされている. この問題を解決するために, 本研究では人間に判読困難なナンバープレートを対象として, ナンバープレートの数字を判別もしくは候補を絞ることを目的としている.

2. 関連研究

2.1 人間に判読可能なナンバープレートの文字の自動識別

Wang らの研究 [1] では, 中国の車両のナンバープレートに対して, 自動的にセグメンテーションと認識を行う一連のアルゴリズムを提案している. 彼らの研究では, 人間がナンバープレートの文字を読める程度の環境で画像を取得することが挙げられている. アルゴリズムは, 画像の前処理, ナンバープレートの位置特定, セグメンテーション, 文字認識の4つに分かれている. 文字のセグメントの際には, 中国のナンバープレートの形式が決まっていることを利用し, ナンバープレートの標準形式を垂直投影することでセグメントしている. また, トレーニングモデルの生成には BPNN を使用している. このアルゴリズムに基づく実験では, 非常に高い文字認識の正解率が得られたことを示している.

Xing らの研究 [2] では, 改良されたラドン変換を使用したナンバープレートの認識について示している. 彼らの研究では, 傾斜があるナンバープレートの認識失敗を防ぐた

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

めに、ラドン変換によって傾斜補正を行っている。傾斜があるナンバープレートに対して、改良されたラドン変換を使用して傾斜補正を行ったうえで、文字のセグメンテーションや文字認識を行うことで、認識精度が90%から95%まで高くなったことを示している。

Zhangらの研究[3]では、実用性を目指して、機械学習に基づく堅牢なナンバープレートの認識システムを提案している。彼らの研究では、ナンバープレートの検出時に統計的学習理論に基づく機械学習アルゴリズムであるサポートベクターマシン(SVM)を使用して、画像ヒストグラム特徴からナンバープレートの面積を判断している。また、ナンバープレートの文字認識には、畳み込みニューラルネットワーク(CNN)モデルであるLeNet-5ネットワークが使用されている。実験では、正常状態(晴れなどの理想的な条件)と過酷条件(雨、夜、霧など)で分けて行い、認識精度は正常状態で100%、過酷状態で90.5%と高い値が得られている。また、ナンバープレートの認識速度が速く、システム効率が高いことから実用性の目的を達成していることを示している。

Ismailらの研究[4]では、夜間および雨天時に動いている車両のナンバープレートを検出して認識する方法を提案している。彼らの研究では、ビデオデータを対象に夜間や雨天時のナンバープレート認識を行っている。提案されている手法には、ローカルバイナリパターン(LBP)を使用したカスケード分類器によるナンバープレート検出、画像強調による文字検出、光学式文字認識(OCR)とプレートマッチング(TM)による文字認識といった機械学習に頼らない方法を用いている。また、画像強調にはトプハット変換やアスペクト比などの画像処理技術を実装しており、様々な照明条件で画像を解くことができている。それにより実験結果も画像強調なしの場合に比べ、45.16%増加していることを示している。

これらの研究では、人間に判読可能なナンバープレートのみを対象としており、ボケやノイズ、光などといった様々な要因により人間に判読困難となったナンバープレートでは、これらの手法は有効ではない。本論文では、人間に判読困難なナンバープレートに着目し、それを判別する。

2.2 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

神宮らの研究[5]では、防犯カメラで撮影された低解像度のナンバープレートから数字を識別する方法を提案している。この研究では、人間には判読不可能なナンバープレート画像を対象としている。この研究の手法は、実際のナンバープレートと同じフォント・数字形式を用いたリファレンスプレートを作成し、対象画像のナンバープレートと同一の位置に置き、同一のカメラで撮影している。これにより文字テンプレートを作成し、対象画像と文字テンプレ

ートを比較し、最も類似する文字テンプレートの数字を対象画像のナンバープレートの数字としている。この手法の精度としては、人間が判読困難な解像度でも3番目(第3候補)までには9割の正解が含まれているという結果になり、犯罪捜査にも使用可能な精度である。

しかし、この研究で提案されている手法には無理がある。神宮らの実験では、対象画像とリファレンスプレートの撮影の日にちを分けているなどといった記述がないため、同じ日に同じ位置で車両を移動させずに、リファレンスプレートのみを変えて撮影しているように読める。また、同じ日に同じ位置で撮影されていると考えられるため、天候や明るさなどの条件もほぼ揃えられていると考えられる。つまり、神宮らの手法を再現するには、対象画像とリファレンスプレートの撮影の際に、車両の位置や向き、天候条件や明るさなどといった条件を全て揃える必要がある。しかし、実際の犯罪捜査などでは、車両を対象となる画像と1ミリもずれずに同じ位置・同じ向きに停車させることは困難である。また、天候や明るさなどの条件もそろえる必要があるが、別の日にリファレンスプレートの撮影を行うということは、天候や明るさが同条件にはならない。そのため、神宮らの手法を実際の犯罪捜査等で使用するのには困難である。

Spanhelらの研究[6]では、CNNを使用した低品質ナンバープレートの文字認識を行っている。この研究では、文字のセグメンテーションを行わず、ナンバープレートの画像を全体的に処理するためにCNNが使用されている。また、対象とされているのはヨーロッパのナンバープレートだが、十分な量のトレーニングデータがあれば、様々なスタイルのナンバープレートに適用できると示唆している。この研究では、一部にはSvobodaらのデータセットが使用されている。このデータセットは、交通監視システムの監視カメラによって撮影されたモーション・ブラー付きのナンバープレート画像である。それをSvobodaらがCNNを使用してデブラーしたものである。本研究でも、人間に判読困難なナンバープレートが対象にしているため、このようなモーション・ブラーも研究の対象となる。しかし、この研究で使用されているSvobodaらのデータセットは、ブラーを除去したときに人間に文字が読めないナンバープレート画像は削除されている。つまり、このような解像度が低くて読めないナンバープレート画像に対しては正解ラベルが不明であるため、教師有りのトレーニングができない。そのため、同様の手法を適用することができない。

鈴木らの研究[7]では、人間による判読が困難なナンバープレートをCNNを用いて自動的に判別する際に、テストデータとなるナンバープレートの画像に対して、高解像度化処理を行うことで機械学習の予測値への影響を調査している。この研究では、日本のナンバープレートのフォントや配置が一意であることを利用し、コンピュータグラ

フィックスを用いてボケやノイズなどを再現し、機械学習のためのトレーニングデータを大量に作成している。さらに、テストデータとなる実際に撮影したナンバープレートに対して、RAISR とバイリニア補間を用いて、人間の視認性を向上させるための高解像度化処理を施し、判別精度の変化について確かめている。その結果、高解像度化による大きな改善は見られなかったが、画像によっては高解像度化による効果が大きいものがあり、高解像度化が有効に働く可能性を示唆している。

しかし、鈴木らが作成したトレーニングデータの中には、テストデータとして用いているナンバープレートの画像と比べて、明度やボケ、ノイズなどの入り方が全く異なるものが多く含まれており、それが判別の精度を下げている可能性がある。ナンバープレートの場合、ナンバープレートの周囲の画像情報から明るさや車両の傾きなど、いくつかの条件を推測することが可能である。推測可能な条件を用いてトレーニングデータの作成を行えば、判別精度を向上できる可能性がある。

また、鈴木らが使用しているテストデータは鈴木らが独自に作成したものであり、誰から見ても判読困難であるかという検証が行われていない。つまり、鈴木らの主観で判読困難と判断されたものであり、客観的に判読困難であるかは不明である。さらに、論文に掲載されている RAISR やバイリニア補間を行った後の画像の中には、判読困難なものも含まれているが、反対に非常にはっきりと数字が判断できるものも含まれている。そのため、高解像度化処理を行う前のテストデータ画像の中には、人間に判読可能な画像が含まれている可能性がある。

また、鈴木らの研究では機械学習において、研究当時の SoTA メソッドを用いていないことや、ハイパーパラメータのチューニングも行われていないため、判別精度が最善のものでない可能性もある。

3. 提案手法

本研究では、人間に判読困難なナンバープレートが写った車両の画像（対象画像）から得られる角度や明度など、画像全体から得られる特徴も利用してトレーニングデータを作成することで、トレーニングデータが網羅的になることを防ぎ、対象画像に適したデータセットの作成を行う。また、鈴木らの研究や一般的なナンバープレート認識に関する研究では、1つのデータセットで機械学習を行い、その学習結果で様々なテストデータの予測値を出しているが、本研究では対象画像（テストデータ）ごとにトレーニングデータを作成し、それぞれ学習を行うことで精度の向上を目指す。なお本研究で対象とする人間に判読困難なナンバープレートは、日本で使用されているものとする。日本のナンバープレートは、一連指定番号と呼ばれる1~4桁のアラビア数字から構成され、4桁に満たない場合は中

黒と呼ばれる黒点が使用される。本論文では、以降「数字」という表現には、0~9のアラビア数字と中黒を含むものとし、この数字の判別を行う。

3.1 対象画像

本研究で対象とするナンバープレートが写った画像は、鈴木らの研究で使用されている画像と、我々が鈴木らが撮影した手法と同様に撮影した本学のスクールバス^{*1}の画像を使用する。鈴木らの論文では、スクールバス（被写体）との具体的な距離は明記されていないが、鈴木らが撮影した画像と同じ場所から撮影した結果、天候などの光の条件にもよるが、70m 前後の距離でスクールバスのナンバープレートの数字が人間に判読困難な状態になった。なお、鈴木らが使用したカメラは Apple の iPhone 7 に搭載されているカメラであるが、我々が使用したカメラは、Apple の iPhone X に搭載されているカメラである。鈴木らが使用している画像と我々が撮影した画像の中から、拡大してもナンバープレートの数字が判読困難な画像を選別し、それらを本研究の対象画像とする。

なお、本研究で「判読困難」であるとしている画像は以下の(3)の条件を満たすものとしている。

- (1) 100%の確信を持って数字を言い当てられないもの
- (2) その画像を拡大しても上記にあてはまるもの
- (3) 複数人が見たときに、過半数の人が上記2つに当てはまると判断したもの

また、本章冒頭でも記述したように、対象画像ごとに適したデータセットの作成を行う。そのため、本章の以降の作業は対象画像1枚につき1回ずつ行い、それぞれのトレーニングデータ及びテストデータの作成を行う。

3.1.1 ナンバープレートと数字のトリミング

特徴の読み取りやテストデータの作成のために、対象画像からナンバープレートとそのナンバープレートの数字のトリミングを行う。図1は、本研究の対象画像の例である。図1には、小さいがスクールバスが写っていることが確認できる。このスクールバスからナンバープレートのトリミングを行う。トリミングの際には、スクールバスが低画素になっており、ナンバープレート部分の堺の判断が難しいため、確実にナンバープレート部分と判断できる部分のみをトリミングする。図1からトリミングしたナンバープレートを図2に示す。なお、1枚の対象画像からトリミングするナンバープレートは1つのみとする。

図2は、トリミングした23×11ピクセルのナンバープレートを本論文のサイズに合わせて、表示させている。さらに、このナンバープレート画像から4つの数字をそれぞれトリミングする。数字をトリミングする際には、数字の

^{*1} 許可を得て撮影・使用できる車両が本学のスクールバスのみであったため。なお、本学の学務課から許可を得た。



図 1: 対象画像の例

シルエットから 1 ピクセルの間隔をあけ、4 つの数字が同じサイズになるようにトリミングを行う。また、ボケや白飛びなどにより数字の位置の判断が難しい場合があるが、日本のナンバープレートは道路運送車両法施行規則により、数字ごとの間隔や幅が決められているため、それを基に位置を判断してトリミングを行う。図 2 からトリミングした各数字を図 3 に示す。

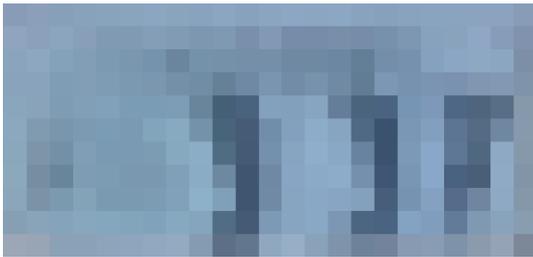


図 2: 切り出したナンバープレート

図 3 の a に示すように、数字のシルエットが確認できない場合でも、数字ごとの間隔や幅から位置を決め、他の数字と同様のサイズでトリミングを行う。

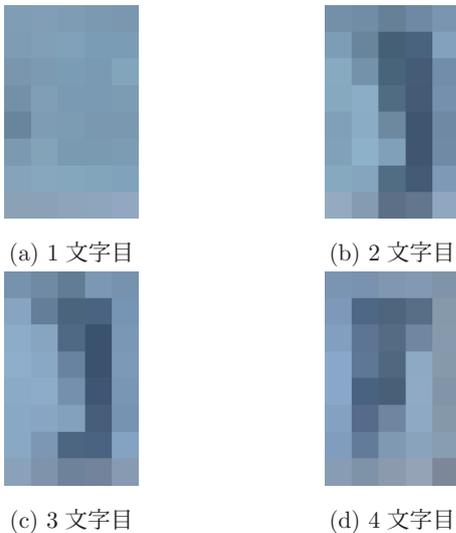


図 3: ナンバープレートから切り出した数字

3.1.2 傾きの読み取り

対象画像のナンバープレートの傾きを再現するために、対象画像から車両の傾きを読み取る。傾きの読み取りは、画像編集ソフトを用い、図 1 のような対象画像に対して行う。また、事前に対象となる車両の寸法（全長・全幅・全高）を調査しておくことが必要である。対象画像の車両の寸法（ピクセル数）を基に 3 次元回転の角度を算出する。回転軸ごとの回転角度の算出方法については以下に記す。

Y 軸回転については、対象画像に写っている車両の側面の長さ (X_y) と実際の車両の側面の長さ (Z_y) から回転角度 (θ_y) を求める。側面の長さ (Z_y) は画像のピクセル数とする。側面の長さ (X_y) は、車両の寸法と図 4 に示す画像における車両前面の高さ H から比を求め算出する。これらの値から Z_y を斜辺とする直角三角形ができ、三角比により回転角度 (θ_y) を求める。実際に求める直角三角形は、図 5 に示すように車両を俯瞰した視点から得られる三角形である。回転角度 (θ_y) は次の式から求めることができる。

$$\theta_y = \arccos \frac{X_y}{Z_y} \quad (1)$$

Z 軸回転については図 6 に示すように、画像の縦長さ (Y_z) と、車両の横幅の中心位置を延長した直線を斜辺とする直角三角形を描き、もう 1 つの辺の長さ (X_z) から回転角度 (θ_z) を求める。なお、これらの長さは画像のピクセル数とする。回転角度 (θ_z) は次の式から求めることができる。

$$\theta_z = \arctan \frac{X_z}{Y_z} \quad (2)$$

また、本研究で使用する対象画像は、平らな場所で人の目線の高さから車両を撮影しており、X 軸については大きく回転していない。そのため、X 軸回転については考慮しないものとする。



図 4: Y 軸の回転角度

またピクセル数を数える際、ピクセル数の読み取りが手動であることや車両の角や面が丸いこと、対象画像の車両も低画質であることが原因で、車両の寸法のピクセル数を

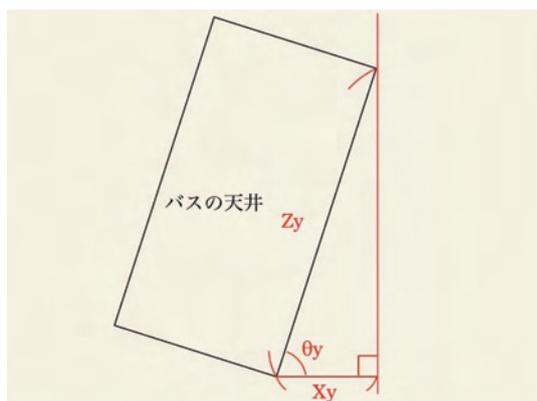


図 5: Y 軸の回転角度 (俯瞰)



図 6: Z 軸の回転角度

正確に読み取るとは困難であり、画像上の実際の寸法と誤差が生じると考えられる。そのため、そのピクセル数を数える際の誤差（計測誤差）についても考慮する必要がある。計測誤差は、対象画像の車両の丸みなどを見て判断する。例えば、図 7 の赤線は、実際に画像のバスの側面の長さ X_y を測った線である。このバスは、赤線の左端の白いラインからフロントガラスの湾曲が始まっており、丁度赤線の長さが確実に画像から読み取れるバスの側面の長さとなる。これに計測誤差として、フロントガラスが湾曲している分の厚みを考慮する。図 7 の場合、計測誤差は $+0 \sim 3$ ピクセルと判断できる。

3.1.3 明度の読み取り

対象画像のナンバープレートの位置に含まれる画素から明度を読み取る。明度の読み取りには OpenCV を用いて行い、図 2 のようなナンバープレートの画像と図 3 のような各数字画像に対して行う。

明度の読み取りの際には、数字部分の画素と背景の画素を分けて明度を読み取る。数字部分の画素は、画像の画素の中で最も濃い画素数字部分の明度として取得する。背景の画素は、ボケによって数字部分と背景の色が混ざった中間色などが含まれるため、数字部分の明度より 25% 以上明るい画素を背景の画素とする。その際、対象となる画素の明度が 255（明度の最大値）を超えてしまう場合を考慮し、その場合には 250 以上の明度の画素を対象とする。また、



図 7: 計測誤差を判断するの例

これ以外の理由により、数字部分の明度より 25% 以上明るい画素が 1 つも無い場合には、数字部分の明度より 10% 以上明るい画素を背景の画素とする。さらにこれらに当てはまらない場合には、9%, 8% と 1 ずつ下げていながら、背景の画素を探す。なお、このような状況は白飛びや黒つぶれによって数字と背景の色に差が小さい場合に見られる。このようにして指定された背景の画素から明度の値を取得し、それらの平均値を背景の明度とする。

3.1.4 中間色の量の読み取り

対象画像のナンバープレートは低解像度であるため、数字の濃緑と背景の白色が混ざり合ったような中間の色がボケとして出ている。このボケを再現するために、この中間色のピクセル数を読み取る。中間色の量の読み取りにおいても、ナンバープレートの画像と各数字画像に対して行う。中間色はナンバープレートや数字画像の全体の明度の平均値を中間色とし、この平均値を中心として 10% の範囲に含まれる明度を持つピクセル数を読み取る。

3.2 テストデータ

3.2.1 判読可否の評価

3.1.1 項で切り取った数字画像が客観的に見ても判読困難であると証明するために、判読可否の評価を行う。評価にはアンケートを用い、被験者を 10 名以上用意する。切り取った 110 枚の数字画像をそれぞれバイキュービック法で 80 倍に拡大し、それら 1 つ 1 つを見て「100% の確信を持って数字が分かる」と「分からない」のどちらかを選択してもらう。また、「100% の確信を持って数字が分かる」を選んだ際には、その分かった数字を記述してもらう。これにより、「分からない」を選択された票と「100% の確信を持って数字が分かる」と選択されたが、記述された数字が画像の実際の数字と異なる場合の票の総数が過半数以上だったものを客観的にも判読困難である数字画像とする。そして、この数字画像（拡大前）をテストデータとする。

3.2.2 リサイズ

機械学習の際には、トレーニングデータとの画像サイズを揃える必要があるため、テストデータに対してリサイズを行う。リサイズの方式にはバイキュービック法を用いて、 64×64 のサイズにリサイズする。

3.3 トレーニングデータ

3.3.1 基本画像の作成

トレーニングデータを作成するために、各数字の基となる画像（基本画像）を作成する。本過程で作成される各数字の基本画像を図 8 に示す。日本のナンバープレートは、フォントが公開されていないため、実際に使用されている車両のナンバープレートを正面から全体が収まるように撮影し、その画像を使用して基本画像を作成する。撮影には複数枚のナンバープレートを撮影する際に、カメラ内部の撮影条件を等しくするため、絞り (F 値)・シャッタースピード・ISO が手動で設定可能なカメラを使用する。絞り・シャッタースピード・ISO はその場の明るさなどに合わせて、ノイズが出ず、数字が明るくはっきりと写るように設定する。また、撮影条件を等しくするため、三脚を用いてカメラを固定し、撮影するごとにナンバープレートとカメラセンサーまでの距離・角度が極力同じになるように撮影する。また、照明条件も等しくなるよう、同じ日・同じ時間帯に撮影したものをを用いる。撮影した画像は、繰り返し加工や保存を行う際に画質の劣化が無いよう、PNG 形式に変換し、カラーのまま保存する。なお、本研究で作成する画像は全て PNG 形式で保存する。

撮影した画像からは、各数字を手作業でトリミングする。トリミングした数字は、背景を白で塗りつぶして保存する。なお図 8 のそれぞれの基本画像は、トリミングした各数字の上下の空白が 1~2 ピクセルに収まるように配置し、中黒は中央になるように配置している。



図 8: 基本画像

これらの画像に対して、3.1 節で読み取った特徴を基に変化を加えていくことで、テストデータの特徴と類似したトレーニングデータを作成する。

3.3.2 傾きの変更

基本画像に 3.1.2 項で読み取った Y 軸と Z 軸の回転を行うことで、基本画像をテストデータに近づける作業を行う。対象画像の車両の傾き方によっては、ナンバープレートの 4 つの数字の大きさに差ができることが考えられる。そのため、図 9 に示すように同じ数字を実際のナンバープレートと同じように並べたプレート画像を作成する。このプレート画像は数字ごとに作成する。作成した 11 枚の画像に対して、Java の拡張 API である Java3D を用いて 3 次元回転を行う。

また回転を行う際、3.1.2 項の回転角度に計測誤差を考慮して回転を行う。計測誤差はピクセル数の誤差を考慮したものであるため、3.1.2 項の (1),(2) 式の分母と分子にそれぞれ誤差の数を加える。このとき、全ての分子分母に誤差の数をそのまま加えるのではなく、総当たりの様に 4 つの分子や分母に 1 ずつ増やしていく。例えば計測誤差が +0~3 だとすれば、 4^4 通りの回転角度ができる。プレート画像 1 枚につき、 n^4 通り (n は 0 以上の自然数) の回転角度を適用した画像を作成する。これには、対象画像の回転をより正確に再現するだけでなく、機械学習を行う際のデータ不足を解消する狙いもある。

回転させたプレート画像の例を図 10 に示す。次に図 10 のように回転させたプレート画像から 4 つの位置の数字をそれぞれトリミングする。また、3.2 節で作成したテストデータの数字は判読困難な画像であるため、正確に数字が画像の中心に位置するようにトリミングができていないとは限らない。そのため、回転させた画像から各数字をトリミングする際には、これを考慮し、画像内に数字部分が完全に含まれることを前提として、トリミングする 4 つの数字が全て同じ位置にならないようにトリミングを行う。また、トリミングの際に背景の黒色の部分が入ってしまう場合は、その部分を白で塗りつぶす。



図 9: プレート画像

3.3.3 リサイズ

機械学習を行うために、テストデータとの画像サイズを揃える作業を行う。計測誤差にもよるが前項によって大量の回転画像が作成されるため、これには今後の作業の処理時間を短縮させる狙いもある。リサイズのサイズはテストデータと同じく 64×64 の大きさにリサイズする。リサイ



図 10: 回転させたプレート画像

ズの方式についても、同じくバイキュービック法を用いる。

3.3.4 明度の変更

次に、前項で作成された画像に対して、3.1.3 項で取得した明度を適用する。明度を適用した画像の例を図 14 に示す。明度を変更する際、ナンバープレートの数字部分と背景を区別して変更する。再現画像は背景を白で塗りつぶしているため、背景の明度が 255 となっている。そのため、再現画像の明度が 255 の画素を背景の明度の値に、それ以外の画素を数字部分の明度の値に変更する。明度の変更を行った画像は新規で保存する。なお、3.1.3 項で取得した明度は、ナンバープレート全体と各数字の計 5 種類である。そのため、前項で作成した画像 1 つにつき、5 枚の画像が作成される。

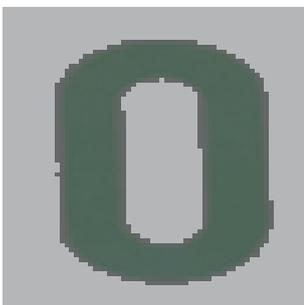


図 11: 明度を変更した数字画像

3.3.5 平滑化

対象画像のナンバープレートは、含まれる画素が少ないことなどから、数字部分がぼやけたような画像になっている。これを再現するために、前項までに作成された画像に対して平滑化処理を行う。前項で作成した数字画像に平滑化を行った例を図 12 に示す。平滑化では、3.1.4 項で取得した対象画像の中間色量を基準値とする。そして平滑化後の画像の中間色量が基準値と近い値を取るよう平滑化を行う。平滑化には平均値フィルタを使用し、3 から再現画像の横の長さの半分である 32 までの値のうち、奇数の値をカーネルの一辺のサイズとし、これらのカーネルで平滑化を行ったときに、最も中間色量が基準値と近い値を取ったものを採用する。また、さらに同じカーネルサイズで平

滑化を 10 回行い、そのなかでより中間色量が基準値と近い値を取った場合にはそれを採用する。なお、この作業でも変更を行った画像は新規で保存する。また、3.1.4 項で取得した中間色は、ナンバープレート全体と各数字の計 5 種類であるため、前項で作成した画像 1 つにつき、5 枚の平滑化された画像が作成される。

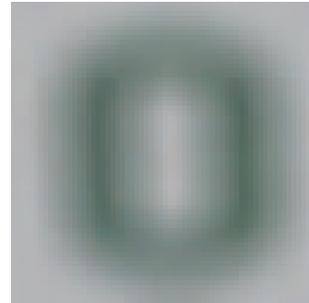


図 12: 平滑化を行った数字画像

3.4 機械学習による判別

本研究では、人間に判読困難な数字を対象としているため、人の目でそれらの画像を見て数字を判別することは不可能である。そこで、機械学習を用いて判別を行う。機械学習には、以下の 2 つを用いる。

3.4.1 RandomForest による判別

1 つは、決定木によって分類するランダムフォレストを用いる。また、ランダムフォレストは複数の決定木を作成し、それらを集団学習させて分類するため、欠損値の多いデータでも精度を維持することが期待できる。なお、ランダムフォレストのライブラリには、Python の機械学習ライブラリである Scikit-learn を用いる。また、ランダムフォレストで設定するハイパーパラメータは、グリッドサーチを用いて最適な値を自動的に選択させる。

3.4.2 CNN による判別

もう 1 つは、CNN を用いる。CNN のモデルには、2019 年に Mingxing らが考案した EfficientNet[8] を用いる。EfficientNet は、Compound Coefficient という係数を用いたスケーリングモデルを使用しており、それにより最適なパラメータ数を決定する。そのため、小さなモデルでも効率良く高い精度が達成でき、2019 年には、5 つのデータセットで SoTA になっている。このことから、本研究では EfficientNet モデルを採用する。

4. 実装

提案した手法の実装には、基本的に Python3 を使用し、3.3.2 項の画像の傾きの変更のみ、Java と Java の拡張 API である Java 3D を使用して実装を行った。画像のトリミングやピクセル数を数える際などには、本研究では画像編集ソフト「GIMP」を使用している。また、アンケートは

Google Form を使用して作成した。

4.1 機械学習

4.1.1 RandomForest

3.4.1 項で示している RandomForest の実装には、機械学習用のライブラリである Scikit-learn を使用した。学習の際には GridSearch を行い、自動でハイパーパラメータの設定を行う。RandomForest で使用する GridSearch のハイパーパラメータの組み合わせを以下に示す。

- max_depth : 3, 5, 10, 20
- min_samples_leaf : 1, 2, 3, 4, 5
- min_samples_split : 2, 3, 5, 10, 20
- n_estimators : 10, 20, 30, 50, 70, 90, 100, 200
- random_state : 0

このパラメータで学習を行い、12 枚ほどの対象画像の GridSearch によって選ばれたハイパーパラメータを追ったところ、min_samples_leaf が 1 または 2, min_samples_split は 2 のみ出現していた。そこで、学習時間の短縮のために 13 枚目以降の学習では、この 2 つのパラメータを以下のよう

- min_samples_leaf : 1, 2, 3
- min_samples_split : 2, 3

4.1.2 CNN

3.4.2 項で示している CNN の実装には、機械学習のライブラリである TensorFlow と、この TensorFlow 上で動作する Keras を使用して実装した。また、TensorFlow のバージョンは 2.5.0 のものを使用し、CNN のモデルには、EfficientNet を使用した。EfficientNet にはベースモデルとして、EfficientNetB0~B7 までであるが、作成した 1 つのデータセットで試した結果、EfficientNetB3 の精度が良かったため、これをベースモデルとして選択した。

また、トレーニングデータは 8 割をトレーニング用に、2 割を検証用に分けて学習を行った。エポック数は 50 とし、EarlyStopping を導入することにより、3 エポック (Patience) の間に検証時の損失関数の値 loss に下降が認められなくなった場合、学習が停止されるようにした。バッチサイズは 32, 64, 128 の値を試した結果、64, 128 ではエポック間の loss の変動が大きかったため、32 とした。

出力には、0~9 及び中黒に判別できるように 11 値分類とした。また、学習によって作成されたモデルは、evaluate() 関数にテストデータを指定することで正解率の算出を行い、predict() 関数でテストデータに対するラベルごとの予測値の確率が算出できるように実装した。

5. 評価

本論文では、機械学習による判別において中黒のラベルを 10 とした。そのため、以降の表などの「10」という値は、中黒を示している。

表 1: アンケート結果

	10 人中					合計 (枚)
	10 人	9 人	8 人	7 人	6 人	
0	8	2				10
1	9	1				10
2	9	1				10
3	7	3				10
4	8	1	1			10
5	8		1		1	10
6	9	1				10
7	9	1				10
8	9	1				10
9	10					10
10	10					10

5.1 テストデータ

評価は、各数字 10 枚ずつ判別が行えるように対象画像を選別した。それにより、対象画像は全てで 53 枚となった。

それぞれの対象画像から数字を切り出し取り、3.2.1 項の判読可否について評価する。21~23 歳の学生及び一般人 10 名を対象に、アンケートを行った。その結果を表 1 に示す。表 1 には、「分からない」を選択した人の数に対する各数字の枚数を示している。例えば 0 の数字画像でひよは、10 枚中 8 枚が 10 人全員に「分からない」と選択され、2 枚は 9 人に「分からない」と選択されている。なお、「100%の確信を持って数字が分かる」と選択されたが、記述された数字が画像の実際の数字と異なる場合も「分からない」の票としている。

表 1 より、110 枚全ての数字画像が過半数以上の「分からない」の票を得ており、3.1 節の条件を満たしていることから、全ての数字画像が判読困難な画像であると判断できる。そのため、全ての画像をテストデータとして用いた。また、「100%の確信を持って数字が分かる」と選択されたが、記述された数字が画像の実際の数字と異なる画像は 8 枚あった。

5.2 計測誤差

本研究で使用した対象画像は、全て 3.1 節に記述した条件で撮影した画像を用いている。本学のスクールバスは、スクールバスによってメーカー・車種が異なるが、ほとんど類似した形の車両である。また撮影条件も統一しているため、いくつかの対象画像を確認したところ、車両のボケも同程度であることが分かった。そのため、対象画像全てに対して一律で計測誤差を +0~3 とした。それにより、対象画像 1 枚につき、テストデータは計 110 枚、トレーニングデータは計 123,904 枚となった。

5.3 RandomForest

3.4.1 項の RandomForest の結果について評価する。RandomForest により判別された各数字 10 枚の予測値の確率の平均値とその標準偏差を表 2 に示す。また RandomForest による判別では、出力されるときに 0~9 及び中黒のラベルごとに予測値の確率が出力される。そこで、最も予測値の確率が高いラベルを第 1 候補, 2 番目に高いものを第 2 候補・・・とし、正解であるラベル(数字)が第何候補に出現するかを確かめた。各数字 10 枚における候補の順位に対する枚数をまとめたものを表 4 に示す。

表 2 より、「8」や「10」の正解率が比較的高く、特に「8」は標準偏差も比較的低く、判別精度が高いことが分かる。反対に、「0」や「3」の正解率が低く、特に「0」の標準偏差が非常に低いことから、「0」の判別精度が低いことが分かる。実際に、表 3 を見ても、「0」は 10 枚のテストデータの中で第 1 候補となった画像が無く、正しく判別されていないが、「8」は 10 枚全てが正解していることが分かる。また、他にも「2」「6」「7」「10」において、正解のラベルが第 3 候補以内に入っている割合が高いことが分かる。

表 2: RandomForest による判別結果

	平均値	標準偏差
0	0.048173481	0.075256958
1	0.536820906	0.396597908
2	0.314370861	0.299344349
3	0.17260203	0.213088663
4	0.340860505	0.30867656
5	0.262414106	0.217486354
6	0.527993149	0.332893676
7	0.46919238	0.251266965
8	0.838711726	0.121518168
9	0.306250142	0.347187515
10	0.736952381	0.39075479

表 3: RandomForest の判別候補

	候補				
	第 1	第 2	第 3	第 4	第 5 位以下
0		2	1	2	5
1	6	1		2	1
2	5	1	3		1
3	2		3	2	3
4	5		2	2	1
5	2	3	2	1	2
6	7	2			1
7	7	1	2		
8	10				
9	5			1	4
10	8	1			1

5.4 CNN

3.4.2 項の CNN の結果について評価する。CNN により判別された各数字 10 枚の予測値の確率の平均値とその標準偏差を表 4 に示す。また RandomForest と同様に、各数字 10 枚における候補の順位に対する枚数をまとめたものを表 5 に示す。

表 4 より、CNN においても「8」や「10」の正解率が比較的高いことが分かる。また、数字によっては RandomForest の正解率よりも高い精度の数字もある。しかし、全体的に標準偏差の値が大きく、どの数字においてもばらつきが大きいことが分かる。また、「8」の正解数が多いが、全体的に正解のラベルが第 3 候補以内に入っている割合が低く、RandomForest と比べても精度は良くないと言える。

6. 考察

6.1 判別精度について

機械学習による判別精度について考察する。なお、ここでの考察は、主に判別精度が高かった RandomForest についてのものである。始めに「8」や「10」の精度が高かったことについて考える。8 や中黒は、数字の中でも画像上

表 4: CNN による判別結果

	平均値	標準偏差
0	0.601580194	0.487959914
1	0.49340448	0.478011469
2	0.36839323	0.433579305
3	0.328669024	0.447231309
4	0.451343774	0.458707481
5	0.480398807	0.458772166
6	0.478858033	0.436274408
7	0.453838698	0.46729461
8	0.82931134	0.346994563
9	0.194073354	0.386312811
10	0.606315757	0.479390638

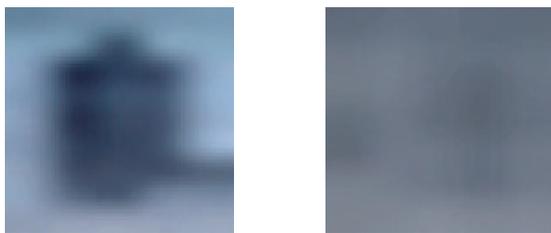
表 5: CNN の判別候補

	候補				
	第 1	第 2	第 3	第 4	第 5 位以下
0	6				4
1	5		2		3
2	3	2		3	2
3	3	1	1	1	4
4	5	1			4
5	5	1	1		3
6	5	1	1		3
7	5	1	1		3
8	8	1	1		
9	2			1	7
10	6		1		3

表 6: 数字が占めるピクセル数

	ピクセル数
0	14975
1	6450
2	11231
3	10440
4	12598
5	13492
6	11560
7	9490
8	14986
9	13030
10	825

の数字が占めるピクセル数が最大もしくは最小の数字である。このことが関係していると考えられる。このことについて、1枚の対象画像を用いて3.3.2項の作業によって、同じサイズにトリミングした各数字の背景を除いた数字部分のピクセルの合計を算出した結果を表6に示す。なお、対象画像の傾きやトリミングする位置によってピクセル数が変わるため、参考の値であることに注意して頂きたい。表6を見ても、実際に数字が占めるピクセル数が「8」は最大、「10」は最小になっていることが分かる。これにより、テストデータのボケやトレーニングデータの平滑化により、「8」の場合はさらに数字に近い色の画素が全体的に増え、「10」の場合は背景に近い色の画素が全体的に増えることになる。実際にテストデータとして使用した「8」と「10」の画像を図13に示す。図13からも分かるように、「8」では数字部分が一色で埋まっており、「10」では中黒が確認できないほど背景の色で占めている。こういった特徴が機械学習によるトレーニングの際に、他の数字にない特徴として得られたことにより、精度が高くなったと考えられる。



(a) 8 (b) 10

図 13: 「8」と「10」の画像

次に、「0」の判別精度が低いことについて考察する。「0」の判別精度が低い原因は、画像の明度が関係していると考えられる。対象画像を選出する際、「0」は形が単純であることにより、判読困難と言える画像が少なかった。そして、「0」において判読困難と言える画像は、ナンバープレート全体の明度が高いものや低いものが選ばれた。判別精度が高い「8」と比較しても、「8」はナンバープレート部分のピ

クセルの明度の平均が108~154であるのに対して、「0」は109以下もしくは238以上と黒つぶれや白飛びに近い画像が多い。このことから、数字と背景の明度の差が小さいと学習による特徴の分類精度が低くなることが考えられる。

次に、鈴木らの研究との比較を行った。鈴木らの研究では、CNNを用いてトレーニングを5回行っており、さらにそれぞれのトレーニング結果に対してテストを5回ずつ行った平均値をまとめている。そこから、我々がさらに平均値を算出し、まとめたものを表7に示す。そして、本研究のRandomForestの結果と表7の鈴木らの結果の平均値を比較したものを表8に示す。

表 7: 鈴木らの判別結果 (RAISR)

	1回目	2回目	3回目	4回目	5回目	平均値
0	0.09334953	0.044576733	0.041092712	0.059361862	0.390019188	0.125680005
1	0.279294158	0.261872314	0.362293305	0.379495648	0.390621731	0.334715431
2	0.058737888	0.067664178	0.044176263	0.129818671	0.118579555	0.083795311
3	0.199866788	0.143488232	0.108307166	0.3326466	0.423254089	0.241512575
4	0.179411467	0.185355518	0.202601813	0.184459714	0.155931706	0.181552044
5	0.236095842	0.202308147	0.033451719	0.309534851	0.42839626	0.241957364
6	0.107505549	0.171057447	0.039787422	0.340308961	0.306521643	0.193036204
7	0.14662847	0.098324641	0.159293538	0.200028152	0.234801079	0.167815176
8	0.029632299	0.019436201	0.0068915	0.032374157	0.078523471	0.033371526
9	0.051848277	0.007415475	0.02566524	0.025073094	0.048343174	0.031669052
10	0.726009459	0.876983691	0.525218854	0.639043499	0.273737982	0.608198697

表 8: 鈴木らの結果との比較

	RF(1)	鈴木ら (2)	(1)-(2)
0	0.048173481	0.125680005	-0.077506524
1	0.536820906	0.334715431	0.202105475
2	0.314370861	0.083795311	0.23057555
3	0.17260203	0.241512575	-0.068910545
4	0.340860505	0.181552044	0.159308461
5	0.262414106	0.241957364	0.020456742
6	0.527993149	0.193036204	0.334956945
7	0.46919238	0.167815176	0.301377204
8	0.838711726	0.033371526	0.8053402
9	0.306250142	0.031669052	0.27458109
10	0.736952381	0.608198697	0.128753684

表8では(1)-(2)という列において、正の値のものが本研究のRandomForestの結果の方が良いことを示している。これを見ると、ほとんどのラベルで本研究の結果が鈴木らの結果を上回っていることが分かる。特に「8」は0.8ポイント以上の向上が見られ、その他の数字でも7つで0.12~0.33ポイントの向上が確認できる。また、負の値(我々の精度の方が低いもの)では差が大きいものでも-0.077506524であるが、正の値の最大値が0.8053402であることと比べると、この負の値は小さな差と言える。また、本論文では鈴木らの研究で“テストデータに人間に判読・判別可能なものが含まれているのではないかと問題提起している。テストデータに人間に判読できそうな数字画像が含まれて

いれば、機械学習での判別がしやすくなり、判別精度が高くなると考えられる。これを考慮した上でも、我々の判別精度が同等以上であることから、本手法が人間に判読困難な画像に対して有効に働いていることが考えられる。

これらのことから、本研究の結果は実用を考えるとまだ不十分ではあるが、鈴木らの研究よりも同等以上の判別精度が出せており、本手法が精度向上のために有効に働いていると考えられる。

6.2 実用に関して

6.2.1 目視による判別について

判読可否の評価のために行ったアンケート結果について考察する。アンケートの結果、表1にもある通り、全ての数字画像で過半数が「分からない」と回答しているが、中には「100%の確信を持って数字が分かる」と回答され、実際の画像の数字を当てられている画像もある。その中で、最も「100%の確信を持って数字が分かる」と回答されていた「5」の数字画像を図14に示す。



図 14: 被験者が正解した画像

図14は、10人中6人が「分からない」と回答し、4人が「100%の確信を持って数字が分かる」と回答して「5」と答えていた。「分からない」と回答した被験者にその理由を聞いたところ、「5の可能性が高いが上部が丸いため、白飛びによって一部が削れた9の可能性もあると考えた」といった返答をもらった。この画像の様に画像がボケていたり、判読困難であったりする場合では、比較的实际の数字が分かりやすい画像でも、他の数字の可能性も否定しきれない場合がある。実際の犯罪捜査では、正しい判断を行うために確実に正しい数字を判断しなくてはならない。

次に、「100%の確信を持って数字が分かる」と回答されていたが、実際の画像の数字とは異なった数字が記述されていた画像の中で、比較的シルエットがはっきりしていた2枚の画像を図15に示す。

図15の左側の画像は1人が「3」と回答したものであり、右側の画像は2人が「1」と回答したものである。実際の数字は左側が「2」、右側が「8」である。このように100%の確信があっても、実際は異なる数字である場合もある。

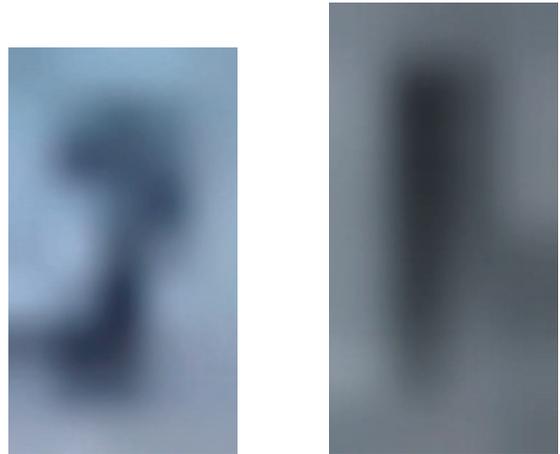


図 15: 被験者が誤った画像

こういった判断は、実際の犯罪捜査においては捜査の遅れや冤罪などの原因となり得る。そのため、ボケや白飛びなどによって判読困難となった画像を目視で判断することは、避けるべきであると考えられる。

6.2.2 数字の候補について

本研究では判別精度だけでなく、判別候補についても評価を行った。表2では、「8」や「10」以外の数字の判別精度は高いとは言えない。しかし、表4の判別候補についても考慮すると、「2」「6」「7」「8」「10」において第3候補以内に候補が出現する割合が大きい。また、第4候補まで考慮するとさらに「1」「4」「5」も候補が出現する割合が大きくなる。このことから、少なくとも第3候補まで考慮することにより、数字によっては高い確率で正解の数字を当てることができる。

多くの研究やMNISTなどの一般に公開されているデータセットを用いた学習では、単純な判別結果やその予測値の確率のみが注目される。しかし判読困難な画像に対してはそれらと違い、正解ラベルがない限りその判別結果が本当に正しいかを判断する術はない。また犯罪捜査においては、ナンバープレートの数字の候補が絞られるだけでも、捜査の時間の大幅な短縮に繋がる。そのため、実用においては、このように判別候補についても考慮して判別結果を確認することが重要であると考えられる。

7. おわりに

本論文では、人間が判読困難なナンバープレートの数字の判別を行うために、それらが写った対象画像全体から得られる特徴を基にしてトレーニングデータを作成することにより、判別精度の向上を狙った。

その結果、数字によって精度に差があるが鈴木らの結果と比較して精度が向上しており、本手法が精度向上に有効であることが確かめられた。また、画像上の数字が占めるピクセル数や明度の度合いが判別精度に関係するという考察が得られた。さらに、客観的に判読可否を評価し、判読

困難な画像の目視による判断の危険性についても示唆した。

また、今後はテストデータや平滑化後の画像に対してエッジ抽出などを行い、数字ごとの特徴を際立たせることにより、「8」や「10」以外の数字の判別精度の向上を目指したい。

参考文献

- [1] 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), IEEE, pp. 403–406 (2016).
- [2] 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. 1, IEEE, pp. 42–45 (2016).
- [3] Zhang, M., Yu, W., Su, J. and Li, W.: Design of License Plate Recognition System Based on Machine Learning, 2019 IEEE 4th International Conference on Image, Vision and Computing (ICIVC), pp. 518–522 (online), DOI: 10.1109/ICIVC47709.2019.8981074 (2019).
- [4] Ismail, M.: License plate Recognition for moving vehicles case : At night and under rain condition, 2017 Second International Conference on Informatics and Computing (ICIC), pp. 1–4 (online), DOI: 10.1109/IAC.2017.8280649 (2017).
- [5] 神宮彩実, 太田直哉ほか. 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, 情報処理学会, Vol. 2011, No. 1, pp. 527–528 (2011).
- [6] Špaňhel, J., Sochor, J., Juránek, R., Herout, A., Maršík, 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), IEEE, pp. 1–6 (2017).
- [7] 鈴木友哉, 宇田隆哉. 人間が判読困難なナンバープレートの自動判別における解像度化画像の応用. 情報処理学会, マルチメディア, 分散, 協調とモバイルシンポジウム (DICOMO), 情報処理学会, pp.118–125 (2020).
- [8] Tan, M. and Le, Q. V.: EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks, 2019 International Conference on Machine Learning (ICML), Vol. abs/1905.11946 (online), available from <http://arxiv.org/abs/1905.11946> (2019).