

CNNを用いた低解像度ナンバープレート数字の識別

Discriminating Car License Plate Numbers on Low Resolution using CNN

松岡剛史*

Tsuyoshi MATSUOKA

藤田和弘†

Kazuhiro FUJITA

四宮康治‡

Koji SHIOMIYA

1 はじめに

警察の捜査等において要望のある防犯カメラなどの車両の自動車登録番号票(ナンバープレート)は小さく、車両登録番号(ナンバープレート数字)の画素数が十分でなく、そのまま、人間が数字として読みとることは難しい。具体的には、数字当たり縦4, 横2画素程度の画素数しかない場合があり、このような数字を読み取ることは困難である。また、従来の十分な解像度と撮影条件を前提とした識別方法[1]では、低解像度であるために数字本来の特徴が弱くなっているため、識別は困難である。

そのような低解像度ナンバープレート数字の識別方法として、吉川らは、ファジー推論を用いた識別方法を提案している [2]。この吉川らの方法は、低解像度では識別の難しい「0」、「1」、「8」の識別を行わず、「0,1,8」という同一のカテゴリとして識別を行うものであり、警察の捜査等からの要望を十分に満たしている識別方法ではない。また、四宮らは、低解像度ナンバープレート数字の幾何学的な特徴として縦横両方向の3,4次のモーメント特徴量および縦横方向の標準偏差の比の5つの特徴量を用いてベイズ識別する方法を提案している [3]。この四宮らの方法は、特徴量が少ないために低解像度化シミュレーション画像に対する識別率が高くなく、十分な識別結果が得られていない。つぎに、四宮らは、低解像度化シミュレーション画像に主成分分析(PCA)を行い、その固有画像による展開係数を特徴量としてベイズ識別する方法を提案している [4]。この四宮らの方法は、実際の画像に適用した結果では、間違っただ数字に対する確率を高く識別することがあり、十分な識別結果を得られていない。また、大江らは、主成分分析(PCA)の代わりに、スパース主成

分分析(Sparse PCA)を用いた手法を提案している [5]。この大江らの方法は、主成分分析が画像全体の特徴を捉える固有画像となっている点に注目し、スパース主成分分析による固有画像を用いることで、画像の局所的な特徴を捉えることができるということを用いた手法であり、主成分分析による固有画像を用いた手法と比較して、実際の観測画像に対する識別結果が良好である。

ひとつの識別手法では、間違っただ数字に対して高い確率となる問題が発生してしまうことがあるために、複数の手法を統合する手法が、大江らにより提案されている [6]。この大江らの方法では、低解像度ナンバープレート数字のモーメント特徴量をベイズ識別した結果、スパース主成分分析(SparsePCA)による固有画像の展開係数を特徴量としてベイズ識別をした結果、CNN(Convolutional Neural Network)による識別結果の計3つの出力に対してSoft Votingを行っている。3つの手法の特徴を総合的に利用しようとするものであり、上記の問題に対する解を提供する提案手法である。

本研究の目的も、上記の既存研究と同様に、低解像度ナンバープレート数字の識別を行うものであるが、以下の点において単なるCNNによる低解像度ナンバープレート数字の識別手法ではない。まず、CNNをベースとして、上記の問題に対する解として、MCDropoutにより解決を試みている点、つぎに、ブラックボックスとも考えられるCNNの識別結果に説明性を持たせるために、Occlusion解析により対象の低解像度ナンバープレート数字のどの部分が識別結果に影響しているかを図示している点、また、どのような画像ならばその数字と識別するかという目的で、フェイク画像の生成を行っている点に特徴がある。

2 ナンバープレート数字の低解像度化モデル

低解像度化画像 $\{\hat{x}_{m,n}\}$ は、原画像 $\{x_{m,n}\}$ をピクセライズした画像 $\{\tilde{x}_{m,n}\}$ に対して平滑化したものと仮定する[3]。ナンバープレート数字の原画像を、次式によりピクセ

* 龍谷大学大学院, Ryukoku University

† 龍谷大学, Ryukoku University

‡ 兵庫県警察本部科学捜査研究所, Hyogo Police Forensic Laboratory



図1 低解像度化したナンバープレート数字画像

ライズを行う。次式中の W は解像度を何分の一にするかというパラメータであり、 m_0, n_0 はピクセライズを行う際のオフセットのパラメータである。

$$\tilde{x}_{m,n} = \frac{1}{W^2} \sum_{k=0}^{W-1} \sum_{l=0}^{W-1} x_{\lfloor \frac{m-m_0}{W} \rfloor W+k, \lfloor \frac{n-n_0}{W} \rfloor W+l} \quad (1)$$

ピクセライズした後、以下の式により平滑化を行う。以下の式における r は平滑化のパラメータである。

$$\hat{x}_{m,n} = \sum_{k,l} b_{k,l} \tilde{x}_{m-k,n-l} \quad (2)$$

$$b_{k,l} = \int_{k-\frac{1}{2}}^{k+\frac{1}{2}} \int_{l-\frac{1}{2}}^{l+\frac{1}{2}} b(u,v) du dv \quad (3)$$

$$b(u,v) = \begin{cases} \frac{1}{\pi r^2}, & u^2 + v^2 \leq r \\ 0, & \text{elsewhere} \end{cases} \quad (4)$$

原画像(縦256画素x横128画素)を、 $W = 32, m_0 = 0, n_0 = 0$ として式(1)によりピクライズし、 $r = 16$ として上式により平滑化を行った数字画像を、図1に示す。

3 CNN (Convolutional Neural Network)による識別

識別を行うCNN (Convolutional Neural Network)の全体構造を、図2に示す。対象となるクラス数は、ナンバープレート数字画像であるので10であり、また、対象画像は低解像度化画像であるために、深いCNN構造でなく、ConvolutionとMaxPoolingの組み合わせが4層の比較的浅いCNN構造としている。また、MCdropoutを行うためのドロップアウト層は、ConvolutionとMaxPooling層と全結合層(Dense)との間に設置している。また、最終全結合層(Dense)において、係数にL1正則化を行っている。

3.1 MCdropout

識別対象であるナンバープレート数字0～9に対するCNNの出力である確率を、 $\{P_0, P_1, P_2, \dots, P_9\}$ とする。

	kernel / pooling	activation
Convolution	7×7	elu
MaxPooling	4×4	
Convolution	7×7	elu
MaxPooling	4×4	
Convolution	3×3	elu
MaxPooling	2×2	
Convolution	3×3	elu
MaxPooling	2×2	
Dropout	(ratio = 0.3)	
Dense	(filters=16)	tanh
Dense	(filters=10)	softmax
	(L1正則化)	

図2 CNNの構造

通常は学習時のみドロップアウトを行うが、対象画像を推論する場合にも、ドロップアウトを行い、 N 回推論を行った結果の出力の平均値を最終の確率とする。つまり、 n 回目のCNNの出力を、 $\{P_0^{(n)}, P_1^{(n)}, P_2^{(n)}, \dots, P_9^{(n)}\}$ とすると、最終の確率 \hat{P}_i は次式となる。

$$\hat{P}_i = \frac{1}{N} \sum_{n=0}^{N-1} P_i^{(n)}, \quad i = 0, 1, 2, \dots, 9 \quad (5)$$

4 Occlusion解析

対象画像 $\{x_{m,n}\}$ の一部、サイズ $W \times W$ の領域を、画素値の最大値1(白)でマスクし、その推論結果 P_i を画像化することにより、対象画像 $\{x_{m,n}\}$ のどの部分が推論結果に影響するのかを図示する。つまり、マスク画像 $\{x_{m,n}^{(k,l)}\}$ を次式で定義する。ここで、 (k,l) はマスク領域の位置を表す。

$$x_{m,n}^{(k,l)} = \begin{cases} 1, & kW \leq m < (k+1)W, \\ & lW \leq n < (l+1)W \\ x_{m,n}, & \text{elsewhere} \end{cases} \quad (6)$$

このマスク画像 $\{x_{m,n}^{(k,l)}\}$ に対するCNNによる推論結果を $P_i^{k,l}$ とし、画像化は次式により行う。

$$y_{m,n}^{(i)} = P_i^{(k,l)}, \quad \begin{aligned} & kW \leq m < (k+1)W \\ & lW \leq n < (l+1)W \end{aligned} \quad (7)$$

5 Fake画像の生成

学習済CNNを識別器として、特定数字に対応する確率が高くなるように、生成器を学習させることにより、どの

ような画像を識別器がその数字と識別するかを可視化する。具体的なフェイク画像の生成方法を、図3に示す。

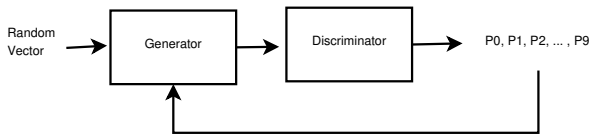


図3 フェイク画像の生成

6 計算機実験結果

0～9のナンバープレート数字画像(縦256×横128画素)から、 $W = 32$, $m_0 = 0, 1, 2, \dots, 31$, $n_0 = 0, 1, 2, \dots, 31$ として、各数字あたり1,024枚、合計10,240枚の低解像度画像を生成し、その半分5,120枚を学習用画像、残りの半分5,120枚を検証用画像として、CNNを学習させた際の学習曲線を、図4に示す。なお、学習用画像に対する識別率および検証用画像に対する識別率は、両方とも1であった。

6.1 実際の対象画像の識別結果

図5に示す実際の低解像度ナンバープレート数字画像に対する識別結果を、表1に示す。なお、推論時には、試行回数100回のMCdropoutおよび重み係数0.001のL1正則化を用いている。

実際の低解像度ナンバープレート数字画像に対する識別結果では、「9」は第1候補、「1」は第1候補、「7」は第2候補、「3」は第1候補、「4」は第2候補、「6」は第3候補、「8」は第1候補、「5」は第1候補として、識別できており、一定の精度を満たしていると判断できる。しかし、上記のとおり、すべて第1候補での識別ではない。そのため、識別機がどのような特徴に従い識別しているかOcclusion解析を行い、識別機が、どのような特徴をもとに、誤識別したのかを確認した。

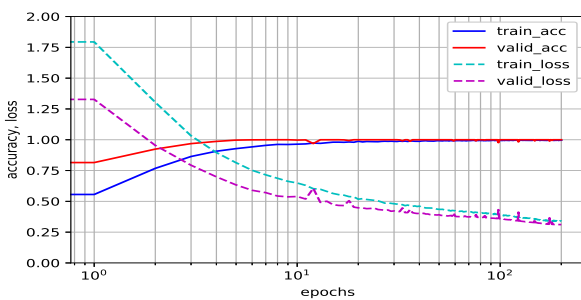


図4 CNNの学習曲線

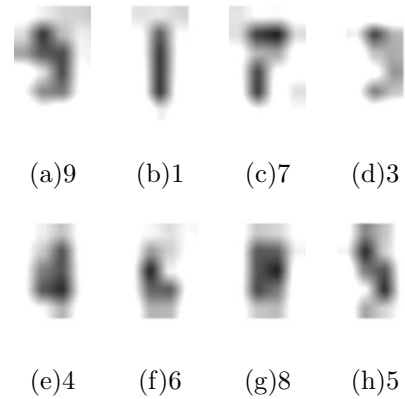


図5 実際の低解像度ナンバープレート数字画像

表1 実際の低解像度ナンバープレート数字画像の識別結果

	9	1	7	3	4	6	8	5
P_0	0.009	0.008	0.004	0.096	0.011	0.011	0.009	0.005
P_1	0.009	<u>0.874</u>	0.026	0.094	0.029	<u>0.761</u>	0.009	0.005
P_2	0.009	0.007	<u>0.816</u>	0.093	0.011	0.043	0.009	0.005
P_3	0.017	0.008	0.012	<u>0.155</u>	0.011	0.016	0.009	0.012
P_4	0.009	0.008	0.004	0.092	0.045	0.011	0.021	0.005
P_5	0.063	0.051	0.036	0.093	0.011	0.078	0.009	<u>0.940</u>
P_6	0.009	0.020	0.004	0.094	<u>0.848</u>	0.049	0.009	0.013
P_7	0.009	0.010	0.089	0.098	0.011	0.011	0.011	0.005
P_8	0.009	0.007	0.004	0.094	0.011	0.010	<u>0.901</u>	0.005
P_9	<u>0.857</u>	0.007	0.004	0.092	0.011	0.010	0.012	0.006

6.2 MCdropout

間違った数字に対して問題を回避するために、MCdropoutを導入しているが、その効果を検証するために、前述の実際の低解像度ナンバープレート画像(図5(c)「7」)に対する推論におけるMCdropoutの有無による確率の違いを、表2に示す。

MCdropoutにより、「7」を正しく識別するような変化はなかったが、「2」の確率が非常に高く、他の数字の確率が同等で非常に低い状態から、「7」の確率が増加している。

6.3 Occlusion結果

前述の実際の低解像度ナンバープレート数字画像(図5(c)「7」)に対する識別結果では、「7」を「2」と誤識別する結果となったが、その原因を調べるために、Occlusion解析を行った。図5(c)の「7」の画像に対して、Occlusion解析を行い、画像の一部をマスクした画像の例と、前述の $\{y_{m,n}^{(2)}\}$ を画像化したものを、図6に示す。

画像化した推論結果から、「7」右上部分が、識別結果に

表2 MCdropoutの効果(「7」)

	MCdropoutなし	MCdropout(試行回数100回)
P_0	0.005	0.004
P_1	0.005	0.026
P_2	0.955	0.816
P_3	0.005	0.012
P_4	0.005	0.004
P_5	0.005	0.036
P_6	0.005	0.004
P_7	0.005	0.089
P_8	0.005	0.004
P_9	0.005	0.004



(a)マスク画像例1 (b)マスク画像例2 (c)推論結果

図6 Occlusion解析結果

影響を与えていることがわかる。このような低解像度ナンバープレート数字画像において、右上部分が、識別結果に影響を与えていることを、確認するために、次節にて述べるFake画像の生成を行い、「7」、「2」と識別される画像の特徴を確認する。

6.4 Fake画像の生成結果

前述の例では、「7」を「2」と誤識別しているが、学習済CNNモデルはどのような画像を「7」と識別するのかを図示するために、前述のフェイク画像生成手法を使い、「7」と識別する複数のフェイク画像を生成した。その複数のフェイク画像の平均化画像を図7(a)に示す。つぎに、「2」と識別する複数のフェイク画像を生成し、その平均化画像を図7(b)に示す。

生成したフェイク画像を比較すると、両方、画像の中心から右上部分にかけて特徴があらわれている。この結果と、前節で述べたOcclusion解析の結果から、「7」を「2」と誤識別するのは、識別対象画像が右上部分に特徴を持っていることが要因と考えられる。



(a)「7」と識別するフェイク画像 (b)「2」と識別するフェイク画像

図7 フェイク画像

7 まとめ

CNNを用いた低解像度ナンバープレート数字の識別を行うと同時に、間違った数字に対して高い確率となる問題に対する解として、MCdropoutにより解決を試みた。また、ブラックボックスとも考えられるCNNの識別結果に説明性を持たせるために、Occlusion解析により対象の低解像度ナンバープレート数字のどの部分が識別結果に影響しているかを図解し、どのような画像ならばその数字と識別するかという目的で、フェイク画像の生成を行い、視覚的に識別機の特徴を示す事が可能であることを示した。

参考文献

- [1] C-N.E.Anagnostopoulos et al.: “ License Plate Recognition From Still Images and Video Sequences: A Survey ”, IEEE Trans. on Intelligent Transportation Systems, Vol.9, No.3, September 2008
- [2] 吉川歩他: “ 記号識別装置、記号識別方法、およびコンピュータプログラム ”, 特許出願2002-303032, 特許公開2004-139325, 特許登録4243941(平成21年1月26日)
- [3] Koji SHINOMIYA et al.: “ Discriminating Car License Plate Numbers on Low Resolution Using Moment Characteristics ”, ITE Transactions on Media Technology and Applications, Vol.1, No.4, pp.1-7 (2013-11)
- [4] 四宮康治他: “ EigenNumberを用いた低解像度ナンバープレート数字の識別 ”, FIT2013 (2013)
- [5] 大江凌太郎他: “ SparsePCAを用いた低解像度ナンバープレート数字の識別 ”, FIT2021 (2021)
- [6] 大江凌太郎他: “ 機械学習モデルのアンサンブルを用いた低解像度ナンバープレート数字の識別 ”, FIT2022 (2022)