The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019

高環境耐性なCNN回帰モデルのメモリ削減検討

Application of Memory Reduction into Environmentally Invariant Regressor CNN

鈴木 まり　　増田 誠
Mari Suzuki　　Makoto Masuda

沖電気工業株式会社 経営基盤本部　研究開発センター
Corporate Research and Development Center, Oki Electric Industry Co., Ltd.

In recent years, the need for computer vision applications such as object/scene classification has grown rapidly. For example, tracking of vehicles and recognizing number plates can be used to automatically estimate traffic flow. These applications require high environmental invariance （due to weather conditions, illumination changes, etc.) as well as compact design and fast rocessing speed in order to be deployed into edge devices.

In this paper, we consider the regression problem for vehicle's number plate recognition. It is shown in [1] that recognition accuracy increases if letters/digits are geometrically transformed beforehand to fix distortions or mis-alignments. However, using CNN to realize such regression task requires a large nunber of parameters.

We propose to use low-rank approximation and fine-tuning to reduce the size of such CNNs. Experiments show that under various environments, our method is effective in greatly reducing memory consumption without significantly degrading performance.

1. はじめに

近年、ディープラーニング技術の発展が目覚しく、画像センシングにおけるシーン認識や物体認識等の認識精度が格段に向上しており、多分野で画像認識ソリューションの需要が拡大している。特に交通インフラの分野においては、路側に設置されたカメラを用いて車両や人物の検出・追跡、車種判別、ナンバープレート（以下、NP）認識等により、交通量計測や異常事象検知、車両捜査等の監視機能高度化・自動化を行うことが期待されている。ここで、交通インフラは屋外での運用となるため、天候や昼夜の明るさの変化に対応する環境耐性が必要となる。また、近年カメラの高解像度化が進み、監視カメラ設置台数も増加傾向を示しており、撮影した大量の画像をクラウドに集約してリアルタイム処理することが困難となるため、カメラ等の各エッジで画像認識することが望まれている。しかしながら、エッジで処理する場合、クラウドと比べて計算リソースが制限されるため、軽量高速な認識技術が必要となる。

我々はNP文字認識を対象とし、高環境耐性な畳み込みニューラルネットワーク (以下、CNN) 回帰モデルのメモリ削減について検討する。文字認識を行う際、事前に歪みやずれを補正して正面に向ける幾何学変換を行うことで、その後の文字認識精度が向上することが報告されており [1]、NP に対しても有効であることを確認している。歪み補正を高精度に実現するためには CNN の適用が考えられるが、CNN はパラメータ数が増えるため、実行メモリが大きくなることが課題となる。

そこで本稿では、歪み補正の幾何学変換パラメータの推定を行う CNN に対し、様々な環境下での推定精度の劣化を最小限に実行メモリを大幅に軽量化することを目的とし、低ランク近似 [2] および fine-tuning を用いた軽量化の検討結果について報告する。

連絡先: 鈴木 まり　沖電気工業株式会社
経営基盤本部　研究開発センター
埼玉県蕨市中央 1-16-8，suzuki512@oki.com

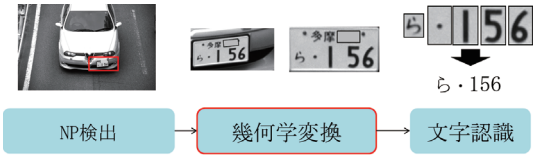

図 1: NP 文字認識の処理フロー

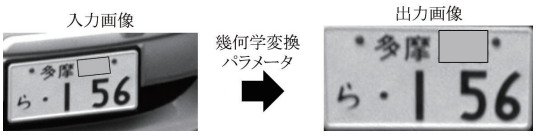

図 2: NP の歪み補正の変換

2. 背景と問題設定

我々は、交通インフラの中でも NP 文字認識を題材とした軽量化の検討を進めた。交通インフラで使用されるカメラは屋外に設置されているため、天候や時間帯の違いによって取得できる画像の見え方が大きく変化する。照明やカメラ制御である程度対応可能であるが、刻々と変わる環境に対応し、高精度な認識をするには限界があるため、環境耐性のある認識技術の開発が必要となる。また、リアルタイム処理を行うには路側に設置したカメラ等のエッジでの処理が要求される。

NP 文字認識の処理フローを図 1 に示す。認識は、車両が写る画像から NP 位置を見つけたのち、文字認識を行う。環境耐性のある高精度な認識を行うため、 Jaderberg らの研究 [1] を参考にし、図 2 に示すような NP の歪み補正を行う幾何学変換パラメータを求め、NP を正面に向けた後文字認識を行う。本稿で推定する幾何学変換パラメータは、画像の移動、回転、拡大、歪みの変換を行うことができる射影変換パラメータを推定する。さらに、射影変換パラメータは CNN を用いた推定を行う。

射影変換パラメータは、入力画像を $I(x, y)$、変換後の画像

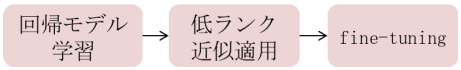

図 3: 軽量化アプローチ方法

を $T(x^{'}, y^{'})$、変換行列を $H(h_{11} \cdots h_{33})$ とすると、

$$\begin{pmatrix} x^{'} \\ y^{'} \\ 1 \end{pmatrix} \sim \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \tag{1}$$

と表すことができる。

式 (1) によって $T(x^{'}, y^{'})$ を求めると、

$$\begin{aligned} x^{'} &= \frac{h_{11}x + h_{12}y + h_{13}}{h_{31}x + h_{32}y + h_{33}} \\ y^{'} &= \frac{h_{21}x + h_{22}y + h_{23}}{h_{31}x + h_{32}y + h_{33}} \end{aligned} \tag{2}$$

となる。変換行列 H を求めるには、変換前と変換後の 4 組以上の座標点を使い最小二乗法で求めることができる。

本稿では、変換行列 H は CNN を用いた推定を行う。CNN を用いることで、高精度な推定が期待できる他、例えば NP のコーナーを検出し、特徴点を用いたパラメータ推定の計算が不要となり、入力画像から直接変換行列 H を推定することが可能となる。CNN の学習と評価の正解行列は、事前に最小二乗法の計算によって求めた行列を用いる。

3. 軽量化アプローチ

一般的に認識性能が高い CNN モデルはパラメータ数が増えてしまうため、エッジでの実行メモリが課題となる。近年 CNN の軽量高速化手法について研究が進んでおり、枝刈りや量子化、知識の蒸留や低ランク近似などがよく知られている [4]。その中でも本稿では、低ランク近似を用いた軽量化を検討した。本稿の軽量化アプローチの流れを図 3 に示す。はじめに、幾何学変換パラメータを予測する回帰モデル CNN の学習を行う。本稿で用いた CNN は、AlexNet[3] を一部改良したモデルを使用している。次に学習結果に対し低ランク近似を適用する。その後、fine-tuning により精度回復し、精度劣化を最小限に抑えた大幅なメモリ削減を行う。

3.1 低ランク近似

低ランク近似とは、特異値分解を利用して行列を次元削減し、軽量高速化を行う手法である [2]。例えば行列 A は特異値分解により、

$$A_{m\times n} = U_{m\times n} \times \ \Sigma_{n\times n} \times \ V^{\mathrm{T}}_{n\times n} \tag{3}$$

と表すことができる。Σ は対角行列になるように定義されており、Σ の対角成分は行列 A の特異値となる。Σ の特異値の上位 k 個を使い、

$$\begin{aligned} A^{(k)} &\simeq U_{m\times k} \times \ \Sigma_{k\times k} \times \ V^{\mathrm{T}}_{k\times n} \\ &= U_{m\times k} \times \ W_{k\times n} \end{aligned} \tag{4}$$

のように書き換えることができる。(但し、$W_{k\times n} = \Sigma_{k\times k} \times V^{\mathrm{T}}_{k\times n}$) このように、行列 A を小さな行列 U と W に分解することで、パラメータ数は mn から $k(m+n)$ になるため、$mn > k(m+n)$ の条件を満たせば、パラメータ数を削減し、軽量化できる。

図 4: 学習および評価データの例

低ランク近似を選択した理由は 3 つある。一つ目は、学習済みモデルを使用するため、追加学習が必要ないこと、二つ目は、どの程度近似するかを選択できること、三つ目は fine-tuning により精度回復が見込めることである。

本稿で用いた CNN は、全結合層部分のパラメータ数が全体の 9 割を超えているため、全結合層の重み行列に対して低ランク近似を適用し、軽量化を行う。低ランク近似を用いる場合、一般的には累積寄与率等を考慮し次元削減を行うが、本稿では実行メモリの課題を解決するため、パラメータ数がどの程度削減されているかに注目し近似する。

3.2 fine-tuning

fine-tuning とは、学習済みのモデルを転用して、新たな学習モデルを生成する方法で、効率的に学習を行う事ができる。低ランク近似で fine-tuning を用いると、精度劣化が大きい場合にも精度回復が見込める [2]。そのため、本稿では低ランク近似適用後 fine-tuning を行う。fine-tuning は複数の再学習方法を検討し、比較を行う。精度向上のためには再学習が必須だと考えているが、どの部分をどの程度学習しなおせばよいか未知であるため、複数パターンで検証を行う。

4. 検証内容

学習および評価に用いたデータは、独自に撮影・正解入力を行った画像を使用する。検証は、NP の歪み補正を行う射影変換パラメータ H を CNN で学習し、複数パターンで低ランク近似を行い、軽量化を行う。低ランク近似の有効性確認のため、複数の削減パターンの精度と実行メモリの関係性について比較を行う。さらに、fine-tuning によって精度回復するか検証を行う。

4.1 学習および評価データ

独自に撮影・正解入力を行ったデータは、複数の環境のデータが含まれている。撮影環境は天候が 4 環境 (晴・曇・雨・雪)、時間帯が 2 環境 (昼・夜) の軸で合計 5 環境の画像を収集した。また、前方と後方の NP が見える車両に対して学習と評価を行っている。

使用した画像の例を図 4 に示す。学習データは、全ての環境を含む画像 6 万枚、評価データは環境別に用意した計 8056 枚の評価を行った。評価データの環境別の内訳は、表 1 に示す。

NP は大板・中板・小板および自家用・事業用が存在するが、収集データ数の関係上、評価データは中板の自家用車のみとなっている。

4.2 検証 1：低ランク近似の有効性確認

低ランク近似適用による、精度劣化と実行メモリ削減の関係性について検証を行う。使用する CNN は、AlexNet[3] か

表 1: 評価画像枚数

時間	天候	枚数
昼	晴	928
	曇	4838
	雨	1734
	雪	132
夜	曇	424

ら入力画像チャネルを 3ch から 1ch へ、出力層を識別から回帰へと変更したモデルを使用する。学習は、正解文字位置から最小二乗法によって求めた射影変換パラメータ H_0 とのユークリッド距離の誤差が最少となる学習を行う。学習は 6 万枚の画像に対し、300 エポックの学習を行っている。また、近似の基準となる CNN は、300 エポック学習終了時のパラメータを使用する。さらに本稿での精度は、正解射影変換パラメータ H_0 と予測射影変換パラメータ H のユークリッド距離の平均ロス値とする。なお、学習はロス値が十分に収束することを確認しており、評価画像が推定した射影変換パラメータ H によって概ね正しく画像変換が行えることを確認している。

まずは、低ランク近似による、ロス値と実行メモリの関係性について比較評価を行う。各全結合層のパラメータ数が基準モデル比率 75%、50%、25%、10%、1%、0.2%となる軽量化 12 パターンと、両全結合層に対してそれぞれ軽量化を行った 6 パターンの計 18 パターンの低ランク近似を行う。各削減パターンについて、1 層目は 01-XX、2 層目は 02-XX、両層は 03-XX のような通し番号を付けている。

評価は、まず天候曇り・時間昼の環境（以下、基本環境）について評価する。基本環境は、外光の影響を受け辛く、NP 周辺の明るさが一定になるため、低ランク近似の比較評価に有効であるといえる。その後、環境別の評価を行う。

4.3 検証 2：fine-tuning の有効性確認

fine-tuning によるロス値の回復について検証を行う。fine-tuning は、検証 1 の学習時と同様の学習を行い、学習画像や学習回数および学習率など全て同じ設定で行う。さらに、学習対象層を学習 1：全て、学習 2：畳み込み層のみ、学習 3：全結合層のみ、の 3 パターンの学習を行う。評価は、検証 1 の評価と同じ画像に対して、環境別評価を行う。

5. 検証結果

5.1 検証 1 の結果と考察

検証 1 の基本環境の評価結果を表 2 に示す。03-05 ではロス値の劣化がほぼなく、実行メモリ 380MB から 72MB の実行メモリの削減を確認することができた。削減パラメータ数が大きくなるにつれて徐々にロス値が劣化していることが分かる。特に、01-06 と 02-06 および 03-06 では大幅にロス値の劣化が見られた。環境別評価は、基準と 03-05、および 03-06 の結果を示す。その結果を表 3 に示す。環境別にロス値に違いがあり、夜と雨のロス値が比較的大きい。

検証 1 の考察を述べる。パラメータ数が基準比率 0.2%になる 01-06 や 02-06 および 03-06 では、大幅な劣化が見られた。これは、パラメータ削減率がある閾値を超えると、基の行列を表現することができなくなり、計算結果が大幅に変わってしまうために起きることだと考えられる。

また、01-05 と 02-05、03-05 のロス値を比較すると、ロス値はほぼ同じであるが、実行メモリは 03-05 が一番小さい。この結果から、ロス値の劣化を最小限に抑え実行メモリを効率的に削減するには、複数の全結合層に対して低ランク近似を行ったほうがよいといえる。

表 2: 削減パターンの違いによるロス値とメモリ (基本環境)

パターン	全結合層 一層目 [%]	全結合層 二層目 [%]	全体パラ メータ数 [%]	ロス値 [$\times 10^{-3}$]	メモリ [MB]
基準	100	100	100	3.0486	380
01-01	75	100	83	3.0486	327
01-02	50	100	66	3.0486	273
01-03	25	100	49	3.0486	216
01-04	10	100	39	3.0486	184
01-05	1	100	33	3.0505	165
01-06	0.2	100	33	3.9616	164
02-01	100	75	92	3.0486	355
02-02	100	50	85	3.0486	330
02-03	100	25	78	3.0486	307
02-04	100	10	73	3.0486	294
02-05	100	1	70	3.0486	284
02-06	100	0.2	70	14.7693	260
03-01	75	75	76	3.0486	303
03-02	50	50	51	3.0486	223
03-03	25	25	27	3.0486	144
03-04	10	10	12	3.0487	88
03-05	1	1	4	3.0505	72
03-06	0.2	0.2	3	11.8500	67

表 3: 環境の違いによるロスの値 [$\times 10^{-3}$]

	晴	曇	雨	雪	夜	全体平均
基準	3.3483	3.0486	4.5517	3.2876	3.9695	3.6412
03-05	3.3503	3.0505	4.5546	3.2934	3.9667	3.6431
03-06	11.5730	11.8500	13.9007	10.2374	11.5348	11.8192

入力画像と射影変換後の画像を図 5 に示す。基準と 03-05 の変換後の画像を比較すると、見た目の変化はほとんどない。一方、ロス値が大幅に劣化する 03-06 では、大幅なずれが見られた。

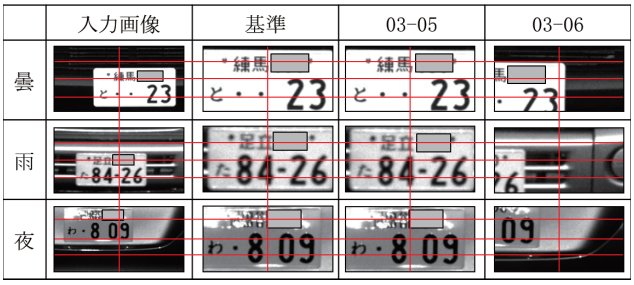

図 5: 削減パターンによる射影変換の比較

5.2 検証 2 の結果と考察

検証 2 ではロス値劣化が大きかった 03-06 に対して fine-tuning を行い、ロス値が回復をするか検証を行った。検証 2 の評価結果を表 4 および図 6 に示す。fine-tuning により、全ての環境においてロス値が回復している。3 つの学習方法を比較すると全環境平均から、学習 3 が最良であることが分かる。

検証 2 の考察を述べる。環境別評価から全ての環境に対してロス値の向上を確認することができたため、環境に依存しない軽量化ができたといえる。また、学習 3 が最良であったことから、本検証では特徴抽出部を変更しない方が効果的である。

ここで、入力画像と各学習の結果より射影変換した画像を図 7 に示す。学習前の画像は大幅なずれが見られたが、fine-

表 4: 学習の違いによるロスの値 [$\times 10^{-3}$]

	晴	曇	雨	雪	夜	全体平均
基準	3.3483	3.0486	4.5517	3.2876	3.9695	3.6412
03-05	3.3503	3.0505	4.5546	3.2934	3.9667	3.6431
03-06 学習前	11.5730	11.8500	13.9007	10.2374	11.5348	11.8192
03-06 学習 1	3.4969	3.2837	4.7992	3.3001	4.3357	3.8431
03-06 学習 2	5.6229	4.9533	7.1650	4.6248	6.5373	5.7807
03-06 学習 3	3.5611	3.1971	4.7752	3.2941	4.2582	3.8171

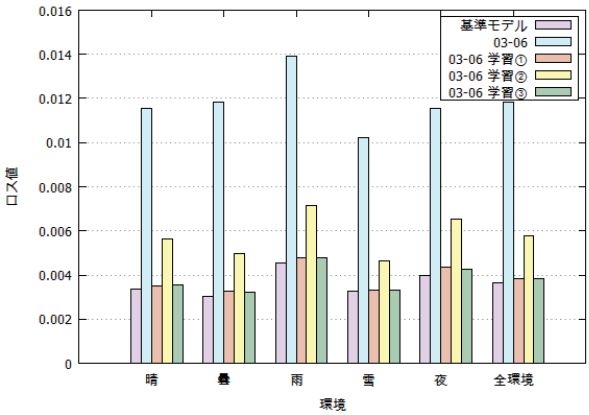

図 6: 学習別のロス値

tuning を行う事で正しく射影変換を行うよう学習できていることがわかる。学習 3 の射影変換結果と基準の結果を比較してみると、多少のずれは感じられるが概ね正しい変換ができている。

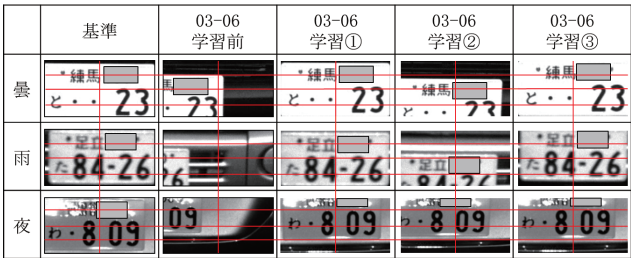

図 7: 学習パターンによる射影変換の比較

6. まとめ

本稿では、NP の歪みを補正する幾何学変換パラメータの推定を行う CNN に対して、低ランク近似と fine-tuning を用いた軽量化を行った。18 個の削減パターンで軽量化を検討した結果、精度劣化がほぼなく 380MB から 72MB の大幅な軽量化を確認した。劣化が大きかった削減パターンに対して、3 つの学習パターンで fine-tuning を検討し、全学習パターンで精度向上を確認することができた。最良学習パターンは、学習前と比較すると十分な精度回復を確認できた。環境別評価の結果から、全環境に対して精度向上を確認できたことから、低ランク近似と fine-tuning は環境に依存しない軽量化といえる。

今回は、全結合層に対して低ランク近似を行ったが、テンソル分解を使い畳み込み層に対して近似する手法も提案されている [5]。今後は畳み込み層に対しての軽量化手法について検討を進めていく予定である。

参考文献

[1] M. Jaderberg, K. Simonyan, A. Zisserman, *et al.* Spatial Transformer Networks networks, Advances in Neural Information Processing Systems (2015)

[2] J. Xue, J. Li, Y. Gong, Restructuring of deep neural network acoustic models with singular value decomposition, INTERSPEECH (2013)

[3] A. Krizhevsky, I. Sutskever, G. E. Hinton, ImageNet Classification with Deep Convolutional NeuralNetworks, In Advances in neural information processingsystems (2012)

[4] J. Cheng, P. Wang, G. Li, Q. Hu, H. Lu, Recent Advances in Efficient Computation of Deep Convolutional Neural Networks, Frontiers of Information Technology & Electronic Engineering (2018)

[5] E., Denton, W. Zaremba, J. Bruna, Y. LeCun, R. Fergus Exploiting linear structure within convolutional networks for efficient evaluation, Advances in Neural Information Processing Systems (2014)