

## 分岐構造の CNN による高速な多クラス分類システムに関する研究 High-Speed Classification System using Branched CNN

村岡 雄太<sup>†</sup> 黒木 修隆<sup>†</sup> 廣瀬 哲也<sup>‡</sup> 沼 昌宏<sup>†</sup>  
Yuta Muraoka Nobutaka Kuroki Tetsuya Hirose Masahiro Numa

### 1. はじめに

近年, SNS 上の写真の自動タグ付けや携帯電話における顔認証など, 身近な所で画像認識技術が用いられている. また, 将来の自動運転に向けて画像中の全ての物体を分類するセグメンテーションの技術が研究されている. いずれの場合も画像認識は高速かつ高精度であることが求められている.

画像認識の分野では, 脳の視覚野の構造を基に構成された畳み込みニューラルネットワーク (Convolutional Neural Network) [1]が大きな成果を挙げている. CNN を用いて画像分類をする場合, 一般に層構造を深くする程性能が上がる傾向がある. しかし, 単純に層構造を深くすると, 処理速度が低下するというトレードオフの関係性が存在する. そのため, 近年では認識精度を維持しつつ, 計算量を削減する技術が必要とされている.

### 2. 提案手法

#### 2.1 提案手法の概要

画像の中にはクラス分類が容易なものとうでないものが混在すると考えられる. それらを一律に同じ深さの CNN で推論することは非効率的と考えられる. そこで, 提案手法では中間層の途中で一旦推論結果を出力し, その尤度が低い場合のみ, さらに深い層へと分岐するようなネットワーク構造を提案する.

#### 2.2 分岐構造の CNN

図 1 に, 提案手法である分岐構造の CNN を示す. ネットワーク構造は VGGNet [2] を模しているが, 上段に分岐すれば 6 層構造, 下段に分岐すれば 11 層構造となる. なお, 分岐の条件については後に述べる.

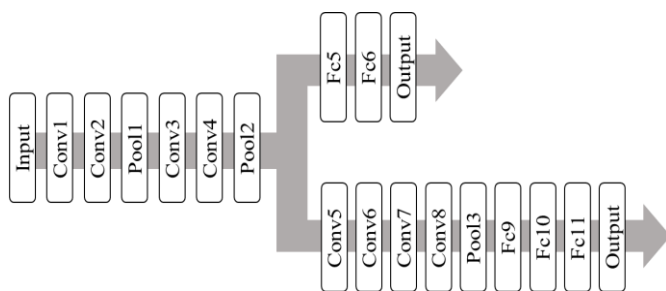


図 1 分岐構造の CNN

#### 2.2.1 学習方法

図 1 に示した分岐構造の CNN の学習方法について説明する. まず, 6 層と 11 層の独立した CNN を構築し, 先に 6 層の CNN の学習を行う. 得られた畳み込み層のパラメータを 11 層の CNN の前段に与えて固定し, 後段のみ学習を行う. このようにして得られた 6 層, 11 層の CNN について前段を共有することで, 分岐構造の CNN を構築している.

#### 2.2.2 推論時の分岐条件

分岐先の決定方法について説明する. まず 6 層 CNN の出力を用いて, 各クラスに属する確率の最大値を求める. そして, その値がある閾値より大きい場合, 尤度が高いと判断し, Conv5 以降の処理を打ち切り認識結果を出力する. そうでない場合, Conv5 以降の処理を行ったあとに 11 層 CNN の認識結果を出力する. このように, クラスの確率によって CNN の経路を分岐させることで, 畳み込み処理による過剰な特徴量計算が削減され処理時間の短縮が見込まれる.

#### 2.2.3 閾値の決定方法

本研究では, 認識率を維持しつつ推論時間を短くすることを目的としている. そのため, 認識率が従来手法と比較して低下しない範囲内で最も推論時間が短くなる閾値を目標値とする.

閾値  $\beta$  に対する認識率と推論時間の一般的な関係を図 2 に示す. 図 2(a) に示したように,  $\beta$  を小さくすると多くの画像が 6 層側に流れるため推論時間は短くなる. しかし, 図 2(b) に示したように,  $\beta$  を小さくしすぎると認識率は 11 層の手法より低くなってしまふ. そのため認識率が従来手法と同じ値になる  $\beta_0$  を目標値とする. なお,  $\beta$  が 1.0 のときは単に 11 層の CNN と認識率が同じ値になるのは明らかであり推論時間も短縮できない.

### 3. 評価実験

#### 3.1 実験内容

従来手法と提案手法を用いて, 一般物体の分類とその評価を行った. 従来手法は分岐を行わない 11 層の CNN であり, ネットワーク構造は VGGNet を模している. 提案手法における分岐条件の閾値は 2.2.3 で述べた手法を用いて算出する. 実験に用いるデータセットは, インターネット上に公開されている CIFAR-10 および CIFAR-100 を用いた. CIFAR-10 は, 10 個のクラスが, CIFAR-100 は 100 個のクラスがラベリングされており, それぞれ  $32 \times 32$  pixel のカラー画像である. 図 3 に画像の一例を示す. 学習用画像は 40,000 枚, 評価用画像は 10,000 枚, 学習回数は 200,000 回とする. 評価指標は認識率と推論時間とする.

<sup>†</sup> 神戸大学大学院工学研究科

Graduate School of Engineering, Kobe University

<sup>‡</sup> 大阪大学大学院工学研究科

Graduate School of Engineering, Kobe University

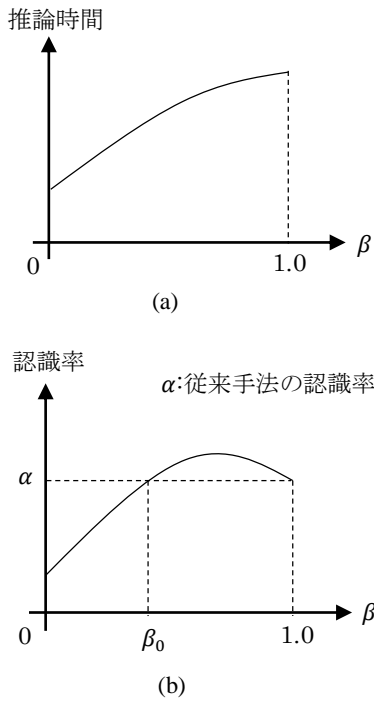
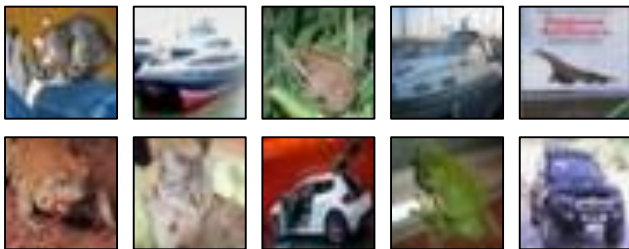
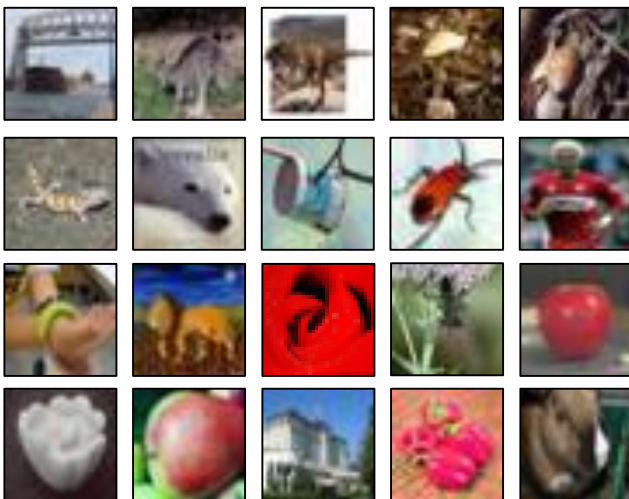


図2 閾値 $\beta$ に対する認識率と推論時間の一般的な傾向



(a) CIFAR-10



(b) CIFAR-100

図3 画像の一例

表1 認識率

データセット	CIFAR-10		CIFAR-100	
	従来	提案	従来	提案
認識率 [%]	86.30	<b>86.38</b>	57.64	<b>57.69</b>

表2 推論時間

データセット	CIFAR-10		CIFAR-100	
	従来	提案	従来	提案
推論時間 [s]	21.31	<b>17.34</b>	20.83	<b>18.99</b>

表3 分岐割合

データセット	CIFAR-10		CIFAR-100	
	6層	11層	6層	11層
画像の割合 [%]	69.30	30.70	59.26	40.74

### 3.2 実験結果と考察

提案手法を用いて閾値を求めた結果、CIFAR-10は $\beta = 0.9692$ 、CIFAR-100は $\beta = 0.8170$ となった。表1に認識率を示す。認識率はどちらのデータセットも、従来手法とほぼ同じ値を示した。このことから一部の画像は6層のCNNで正しく分類が行えていると考えられる。表2に推論時間を、表3に分岐割合を示す。推論時間は、提案手法が従来手法に比べCIFAR-10においては3.97s短い17.34sに、CIFAR-100においては1.84s短い18.99sになった。これは、表3に示したように、評価用画像の半分以上が6層側へ分岐し、Conv5以降の特徴抽出処理を省略できたためである。

### 4. おわりに

本論文では、認識精度を維持しつつ、高速に画像分類を行うCNNを構築することを目的とし、CNNの中間層で分岐を行う手法を提案した。

本手法を評価するために、一般物体認識の実験を行った結果、提案手法では従来手法と比べて、認識率が低下することなく、推論時間においてはCIFAR-10が3.97s短い17.34sに、CIFAR-100が1.94s短い18.99sになった。これらの結果から、分岐構造のCNNの有効性が確認できた。

今後は、さらに分岐数を増やしたネットワーク構造を検討する予定である。

### 参考文献

- [1] Y.Lecun, L.Bottou, P.Haffner, "Gradient-based learning applied to document recognition", *IEEE International Symposium on Intelligent Signal Processing*, pp.2278-2324, 1998.
- [2] K.Simonyan, A.Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition", *International Conference on Learning Representations*, 2015.