

CNNにおける畳み込み層の重層化と特徴抽出の関係に関する検討

Investigation on the Relationship between Stacking Convolutional Layers and Feature Extraction in CNN

外川 宙¹⁾ 神野 健哉¹⁾
Sora Togawa Kenya Jin'no

概要

畳み込みニューラルネットワーク (CNN) は、畳み込み層に含まれるカーネルによって、入力画像の特徴を抽出している。一般的に、カーネルのサイズは入力画像よりも小さく設定される。また、CNN は畳み込み層を複数重ねることで、入力画像の特徴抽出範囲を広げることができ、その結果としてモデルの性能が向上すると言われている。本研究では、特徴抽出範囲を統一し、各モデルで採用するカーネルサイズが異なる CNN を学習し、畳み込み層の重ね方や活性化関数と特徴抽出やモデルの性能との関係について検討を行った。実験の結果、畳み込み層を重層化することで、画像特有の特徴抽出に貢献していることが示唆された。また、非線形性は単純に増やすだけでは、性能に良い影響がないことを確認した。非線形性を追加する適切な数や位置を検討するために、活性化関数の直前層の出力を詳細に分析する必要がある。

1 まえがき

CNN は画像処理分野において優れた性能を発揮するニューラルネットワークの一種である。CNN の畳み込み層では、カーネルを適用して畳み込み演算を行うことで画像の特徴を抽出しているが、カーネルの大きさはモデルの性能に大きく影響を与える。また、畳み込み層や活性化関数の数や位置も同様に、性能に大きく影響するため、適切なハイパーパラメータや構造を模索する研究が行われている。Simonyan らの論文 [1] では、畳み込み層で大きなサイズのカーネルを採用する代わりに、 3×3 のカーネルを採用し、このようなカーネルを持つ畳み込み層を何層にも重ねた深いモデルである VGG が提案されており、ネットワークの深さがモデルの性能において重要であることが示されている。また、Szegedy らも、深い CNN アーキテクチャを持つ GoogLeNet [2] を提案し、深いネットワークが高い性能を発揮することを示している。これらのモデルが評価される要因は、畳み込み層を深く重ねたことによる、特徴抽出範囲の拡大や、活性化関数による非線形性の増加であると言及されている。しかし、深い構造である故に、モデル構造の小規模な要素が性能に与える影響を言及するのが困難である。

特徴抽出範囲の拡大について、我々は小さいカーネルを持つ非常に浅いネットワークにおいても、画像の大局的な特徴を抽出しているという知見を得ている [3]。そこで、本研究では、同様に浅いネットワークを用いて、モデルの小規模な要素として、畳み込み層の重層化によるモデルの性能への影響について検討する。

2 データセット

実験には、グレースケールに変換した CIFAR-10 [4] に、白黒の市松模様を加えた画像データセットを用いる。市松模様を加えたのは、市松模様の周期とカーネルサイズの異なる CNN の分類精度に関係があるのかを検討するためである。CIFAR-10 は、画像サイズが 32×32 、全 10 クラスのデータセットであり、トレーニングデータ 50,000 枚、テストデータ 10,000 枚で構成され、画素値は 0 から 255 である。市松模様の加算は、グレースケール化した CIFAR-10 のスケールを圧縮し、市松模様を加算後のスケールが 0 から 255 になるように、市松模様のスケールを設定した。例えば、CIFAR-10 を 0 から 127 に圧縮した場合、市松模様のスケールは 0 から 128 に設定した。また、市松模様の周期も設定可能である。今回使用するデータセットの一例を図 1 に示す。各画像のタイトルは「CIFAR-10 の圧縮率 (%) - 市松模様の半周期」のデータセットを意味している。今回は、50-1, 50-2, 50-3, 70-1, 70-2, 70-3, 90-1, 90-2, 90-3 の 9 種類のデータセットでそれぞれ実験を行った。

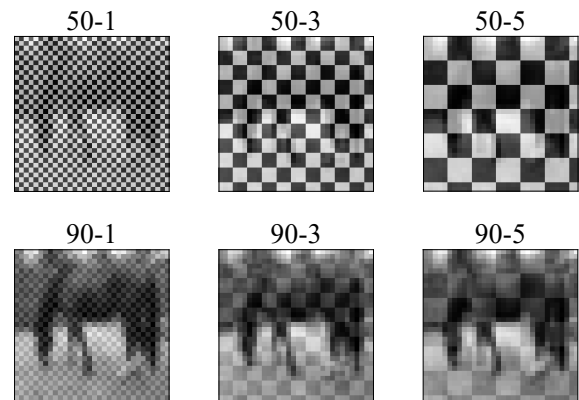


図 1 画像データの例

1) 東京都市大学 知能情報工学科 Department of Intelligent Systems, Tokyo City University

3 比較するモデル

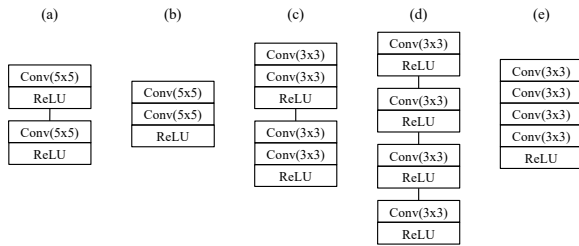


図 2 畳み込みブロック

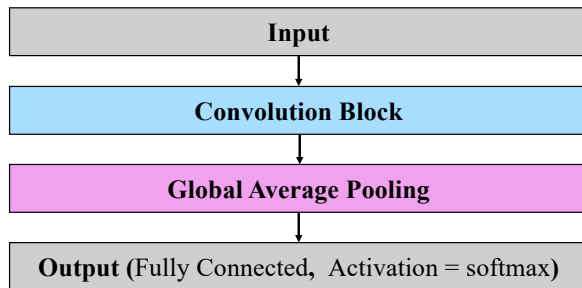


図 3 モデル全体の構造

比較する CNN モデルの畳み込みブロックの構造を図 2 に示しており、モデル全体の構造を図 3 に示す。畳み込みブロックは、畳み込み層と活性化関数の ReLU 関数から構成され、モデル全体は入力層の次に畳み込みブロック、Global Average Pooling 層 [5]、出力層と続く。図 2 の (a) から (e) の畳み込みブロックを持つモデルの特徴抽出範囲は、全て 9×9 であり、カーネルサイズと活性化関数の位置や数が異なる。各のモデルの分類精度から、畳み込み層の重ね方や活性化関数と特徴抽出やモデルの性能との関係について検討する。

4 実験結果

作成した 9 種類のデータセットを、2 の (a) から (e) の構造を持つ 5 種の CNN で学習した。表 1 に学習後のテストデータの分類精度を示す。

表 1 Classification Accuracy

	(a)	(b)	(c)	(d)	(e)
50v1	0.3533	0.3722	0.5302	0.4920	0.4929
50v3	0.3280	0.3898	0.4851	0.4107	0.4704
50v5	0.3133	0.3649	0.4116	0.3594	0.3856
70v1	0.3473	0.4277	0.5689	0.4901	0.5073
70v3	0.3787	0.4175	0.5307	0.4526	0.4857
70v5	0.3454	0.3861	0.4872	0.4080	0.4292
90v1	0.4472	0.4508	0.5922	0.4893	0.5141
90v3	0.4285	0.4415	0.5459	0.4827	0.4933
90v5	0.4204	0.4306	0.5468	0.4740	0.4859

実験の結果、全てのデータセットにおいて分類精度は (c) が最も高く、(a) が最も低くなった。(a) と (c) の構造は似ており、どちらも入力画像の特徴抽出範囲が、畳み込みによって 5×5 と 9×9 になった直後に ReLU 関数が組み込まれる構造で、畳み込みブロックに含まれる非線形変換の位置と数は同じである。このことから、非線形性の増加ではなく、特徴抽出範囲を複数の畳み込み層によって広げることには何かしらの利点があると考えられる。一方、カーネルサイズが 5×5 の (a) と (b)、 3×3 の (c) から (e) を比較すると、分類精度がそれぞれ最も高い (b) と (c) は、どちらも畳み込み層を 2 層重ねた後に ReLU 関数が組み込まれている。また、ReLU 関数の数は、(b) は (a) より少なく、(c) は (e) より多く (d) より少ない。このことから、非線形性の追加で重要なのは、単に数を増やすのではなく、適切な位置と数での追加であると言える。以上より、我々が今後分析すべきなのは、(a) と (c) における ReLU 関数の直前の出力の違いである。

また、市松模様の周期が大きいほど、分類精度が悪くなっていることが表 1 から確認できる。図 1 から、市松模様の周期が大きい画像 (50×50) では画像全体の特徴が見え難い。この結果は、我々の知見通り、小さいカーネルを持つ畳み込み層でも画像の大局的な特徴を捉えていると言える。市松模様の半周期が 5 である場合であっても、分類精度が最も高いのが、カーネルサイズが 3×3 の (c) であることから、畳み込み層の重層化は、各画像特有の特徴を抽出するのに貢献していることが示唆される。

5 まとめ

本研究では、特徴抽出範囲が同等であり、カーネルサイズと活性化関数の位置や数が異なる CNN の学習結果から、畳み込み層の重層化や非線形性が、モデル性能や特徴抽出に与える影響に関して検討を行った。今後は、活性化関数の直前層の出力から、畳み込み層の重層化が持つ影響について、より詳細な分析をする。

謝辞

本研究の一部は JSPS 科研費 23K11266, 23H03387, 東北大学電気通信研究所共同プロジェクト研究 東京都市大学重点推進研究未来知能ユニットの助成によるものです。

参考文献

- [1] K. Simonyan, A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," ICLR 2015.
- [2] C. Szegedy et al, "Going Deeper with Convolutions," CVPR 2014.
- [3] 外川 宙, 神野 健哉, "画像分類における CNN の学習の特徴に関する検討," 電子情報通信学会 2023 年 NOLTA ソサイエティ大会, NLS-42, 2023.
- [4] <https://www.cs.toronto.edu/~kriz/cifar.html>
- [5] M. Lin, Q. Chen, S. Yan, "Network In Network," CVPR 2014.