

知能システム学特論レポート

(DL2 班) Caffe on Ubuntu

2015 年 7 月 2 日

1 報告者

15344203	有田 裕太
15344206	緒形 裕太
15344209	株丹 亮
12104125	宮本 和

2 進行状況

- 理論研究
- 中間層の出力

3 理論研究

3.1 畳込みニューラルネットワーク

畳込みニューラルネットワークは主に画像認識に応用される順伝播型ニューラルネットワークである。順伝播型ネットワークは隣接層間のユニットの全てが結合されているのに対し、畳込みニューラルネットワークでは特定のユニットのみが結合を持つ特別な構造を持つ。そして、畳込みとプーリングという画像処理の基本的な演算を行う。これは生物の脳の視覚野に関する神経科学の知見をヒントにしている。神経細胞はその受容野に刺激が入った場合にのみ反応し、受容野の外側にいくら刺激を与えても、細胞を発火しないという特徴を持つ。

3.2 単純型細胞と複雑型細胞

単純型細胞と複雑型細胞のモデルを Fig. 3 に示す。(a), (b) は 4×4 の入力層が中間層のそれぞれと結合していることを示しており、(c),(d) は入力パターンの位置変化に伴う中間層および出力層の状態の変化を示している。

中間層は単純型細胞、出力層は複雑型細胞のモデルとなっており、中間層は位置選択性が厳密であるが、出力層は入力パターンが少しずれても活性化する。全体への入力 (c) から (d) のように変化すると、中間層で

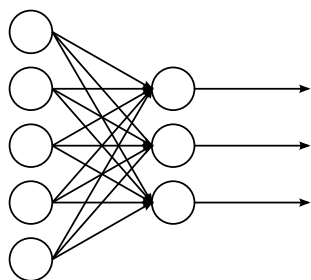


Fig.1 順伝播型ニューラルネットワーク

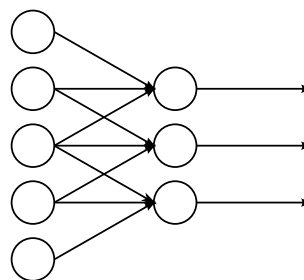


Fig.2 畳込みニューラルネットワーク

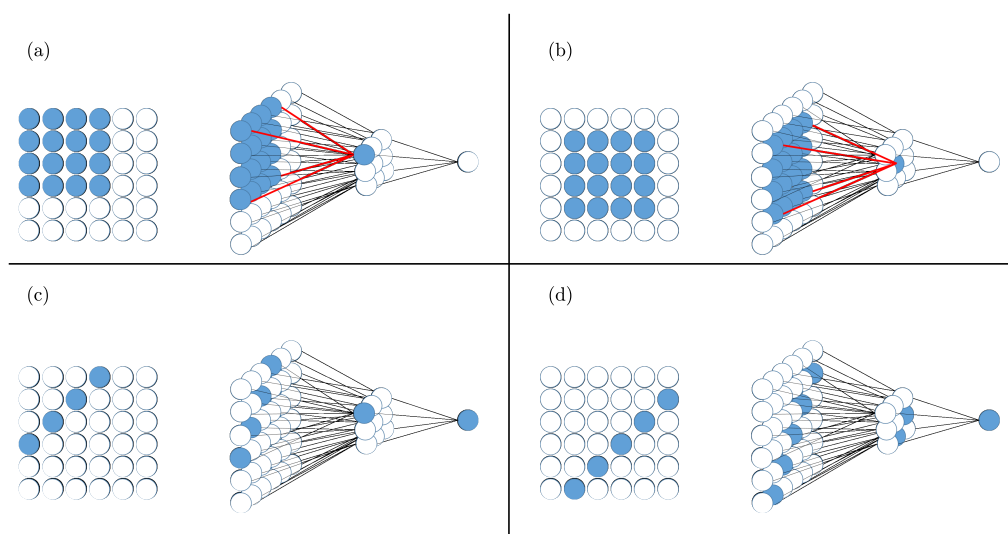


Fig.3 単純型細胞と複雑型細胞のモデル

も図のように活性化するユニットが変化するが、出力層のユニットは活性化したままである。これは出力層は中間層のユニットが一つでも活性化されていれば活性化するためである。

この2つの細胞をモデル化した二層構造のペアを繰り返す構造が畳み込みニューラルネットに用いられている。物体カテゴリ認識は長年コンピュータには難しいとされてきたが、近年畳み込みニューラルネットによってそれも解決されつつある。神経科学の分野では、多層の畳み込みニューラルネットが霊長類の脳の高次視覚野と似た振る舞いを示すことが示されている。

3.3 全体の構造

物体の認識などの画像認識でよく用いられる畳み込みネットの典型的な構造を Fig. 4 に示す。入力側から出力側に向けて、畳込み層 (convolution layer) とプーリング層 (pooling layer) がペアでこの順に並び、このペアが複数回繰り返される。特に畳込み層とプーリング層のペアの一回操作で行われる手順の一例を Fig. 5 に示す。また畳込み層とプーリング層の後に、局所コントラスト正規化 (local contrast normalization, LCN) 層を挿入する場合がある。

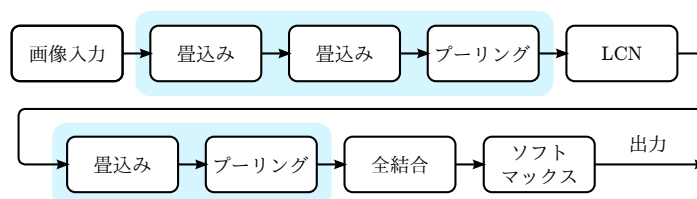


Fig.4 畳込みネットの構造

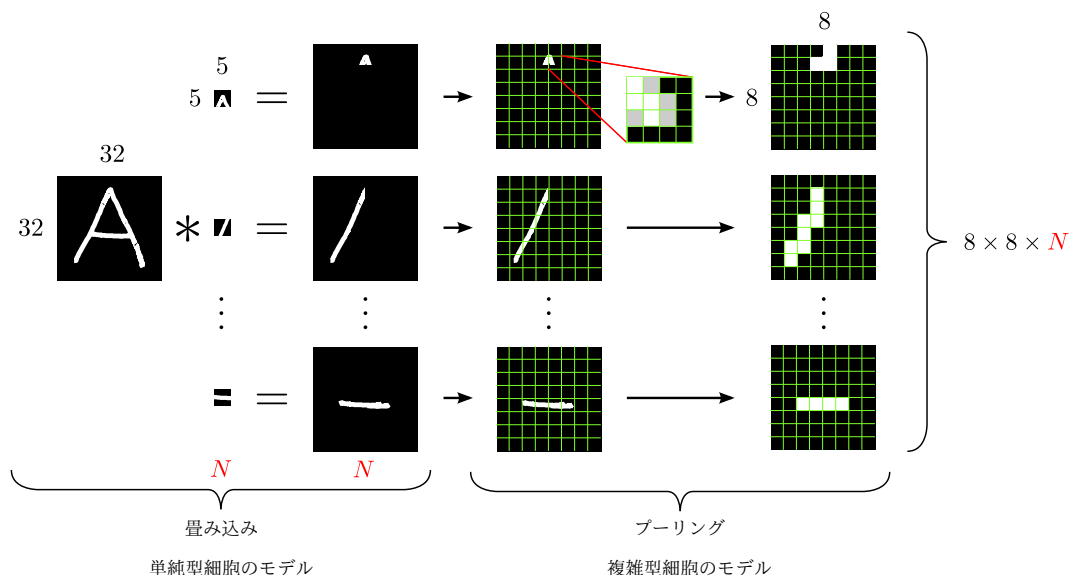


Fig.5 畳込み層とプーリング層の操作

そして畳込み層とプーリング層の繰り返しの後には、隣接層間ユニットが全結合した層が配置される。これを全結合層（fully-connected layer）と呼ぶ。全結合層は順伝播型ネットワークの各層と同じで、目的がクラス分類であれば最終的な出力はソフトマックスを適用する。

4 プログラム

4.1 caffeNet の構造

caffe では訓練済みのニューラルネットワークをダウンロードし、使用することができる。先日、示したサンプルの画像認識に用いた CNN の構造を Fig.6 に示す。赤色は畳込み層、オレンジ色はプーリング層、青色は正規化層、紫色は全結合層を表している。また、活性化関数にはソフトマックス関数を用いている。すなわち、この CNN は 5 つの畳込み層、3 つのプーリング層、2 つの正規化層、そして 3 つの全結合層からできている。

Fig.6 を出力するには、caffe のディレクトリ直下で、次のコマンドを実行する。

```
python/draw_net.py models/bvlc_reference_caffenet/train_val.prototxt caffeNet.png --
rankdir BT
```

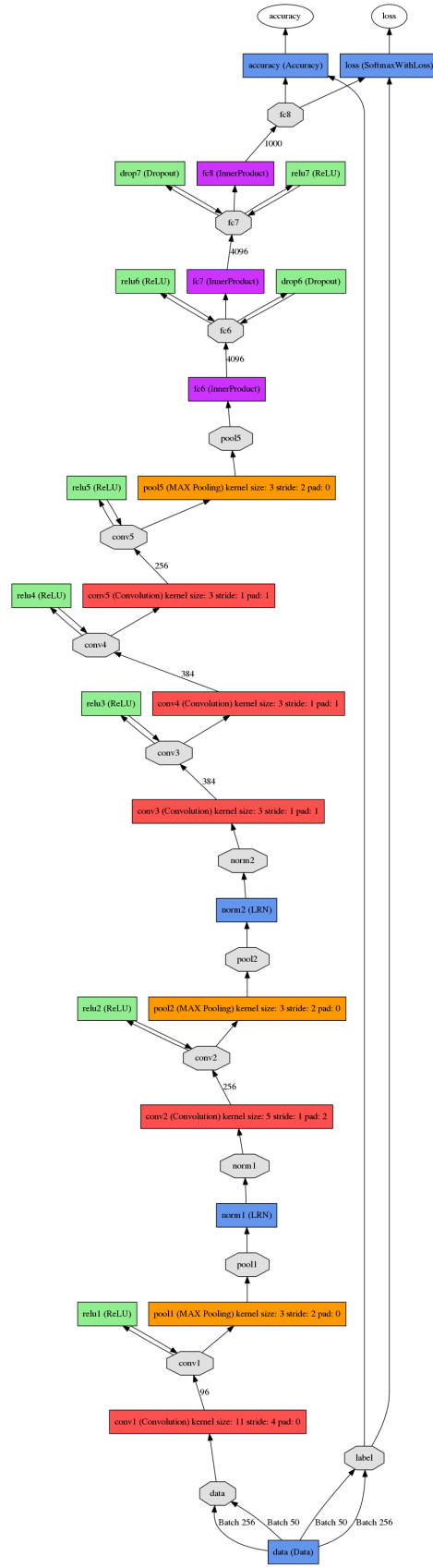


Fig.6 caffeNet

4.2 中間層の出力

畳込み 1 層目のフィルタと Fig. に示す入力画像を入力したときの出力を得ることができた。1 層目のフィルタを Fig.8, その出力を Fig.9 に示す。畳込み 1 層目は全部で 96 個の 11×11 のフィルタから構成されており、ストライドが 4 である。従って、出力される画像のサイズは 55×55 であり、96 個の画像が出力される。中間層を表示するにあたり ipython notebook を用いた。

5 今後の課題

- CNN の詳細な調査
- 畳み込み、プーリングの理論
- データセットの作成

参考文献

[1] 岡谷貴之, “機械学習プロフェッショナルシリーズ 深層学習”, 講談社, 2015.

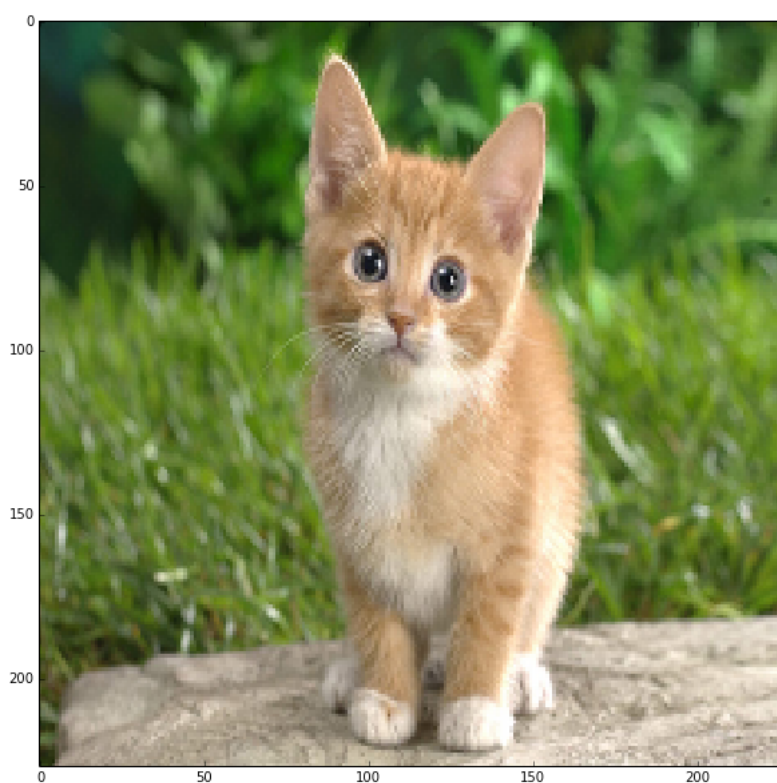


Fig.7 入力画像

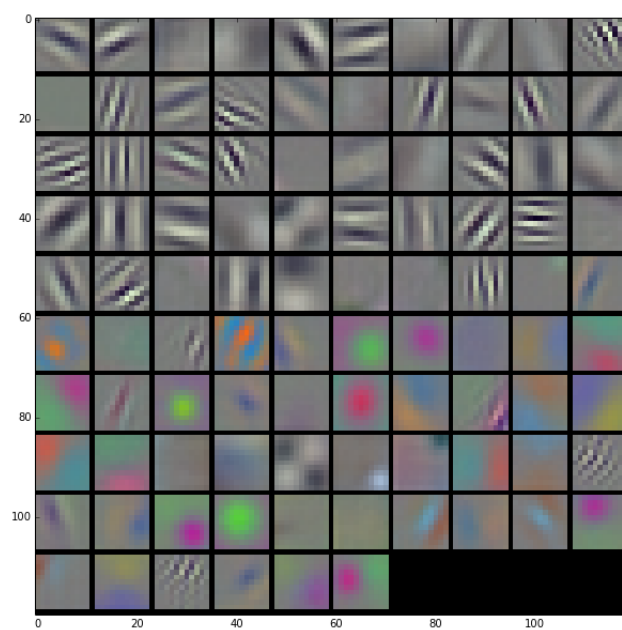


Fig.8 畳込み層 1 層目のフィルタ

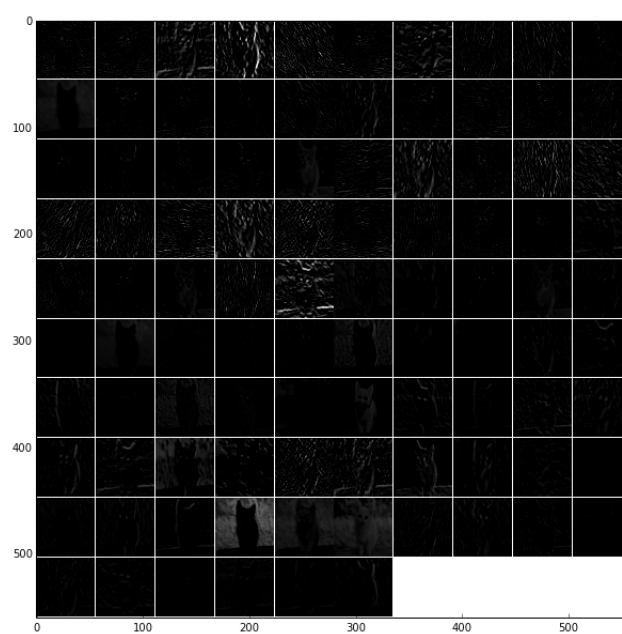


Fig.9 畳込み層 1 層目のフィルタ出力