

# LeNet-5 の平行移動に対するロバスト化

東北大学大学院医学系研究科 博士課程 田中 真一

2024.04.23

## 序

前回レポート (2024.04.20) では、LeNet-5<sup>1)</sup> でみられる文字の平行移動に対する軽度のロバスト性について MNIST<sup>2)</sup> を用いて検討した。そこでは、この軽度のロバスト性は教師データが整っていることを前提とし、畳み込みにより平行移動前の「面影」が残ることに依存していることが示唆された。また、LeNet-5 は構造上、図形の形態を解釈するような機能を持っていないにもかかわらず、インターネット上などには畳み込みやプーリングによりロバスト性が得られるかのような、不正確な記述が溢れていることを指摘した。

今回は LeNet-5 に改良を加え、平行移動に対するロバスト性を備えた畳み込みニューラルネットワーク (Convolutional Neural Network; CNN) のモデルを検討する。同様のモデルはおそらく既に誰かが提唱・命名しているであろうが、著者は把握していない。便宜上、ここではロバスト化 LeNet-5 と称することにする。このロバスト化モデルは、深層学習による領域分割などの技法に用いられる U-Net<sup>3)</sup> と類似の upsampling を採用している一方、単純な構造をしているため、U-Net の理論的背景を理解する助けになると考える。

## モデルの概要

LeNet-5 の構造を図 1 に示す。LeNet-5 の 3 回目の畳み込み<sup>\*1</sup>の結果として得られる 120 個のノードは、数字が画像中央に「行儀良く」配置されている場合に限り、MNIST の手書き数字を精度良く判定することができる。そこでロバスト化 LeNet-5 では、数字の平行移動が行われた際の判定精度を上げるため、この 120 個のノードに対し upsampling を行い、1 回目の畳み込み層と連結する。その後は LeNet-5 の後半と同様に全結合層によって 10 個のノードにまとめる。これは本質的には U-Net で採用されているものと同様の手法である。ロバスト化 LeNet-5 と LeNet-5 との相違を図 2 に示す。

このロバスト化の基本的な考え方は、次のようなものである。入力画像のどこかに数字が描かれているはずであるが、どこに描かれているかは未知である。とはいえ、数字が描かれている部分は「白く」なっているはずであるから、そのような「白い」部分に対して LeNet-5 を適用すれば、精度良く数字を判定できるであろう。そこで、LeNet-5 のエッセンスを含む上述の 120 個のノードを upsampling してプーリング前の層と連結し、さらに  $1 \times 1$  カーネルでの畳み込み (いわゆる pointwise convolution) を行う。文字が描かれていない部分では、この畳み込みの結果は概ね 0 になるので、結果として文字が描かれている部分だけを使って LeNet-5 を適用したことになる。これにより、文字が描かれている位置に依存せず数字の判定を行うことができる。

このロバスト化 LeNet-5 においては、全ての教師データの数字が画像中央に「行儀良く」描かれていると、かえって最後の  $1 \times 1$  カーネルによる畳み込み以降の部分の学習が正しく行われず、数字が描かれていない黒

---

\*1 これを全結合層として表現している文献もあるが、ここではオリジナルの Lecun の表現に従って畳み込み層として扱う。

い部分に対応する重み係数が不適切に設定される恐れがある。そのため、教師データを上下左右にランダムに平行移動させてから入力を行った。この平行移動は、エポック毎にやり直すこととした。また、この教師データの平行移動自体にもロバスト化の効果があるため、通常の LeNet-5 を用いた場合や、教師データの平行移動を行わない場合も含めて比較検討した。

損失関数には交差エントロピー誤差関数を、最適化には確率的勾配降下法を用いた。計算には Python 3.11.2<sup>4)</sup> + PyTorch 1.13<sup>5)</sup> を使用した。計算時間の都合上、エポック数や学習率の最適化は行っておらず、学習率 0.05, エポック数 20 で固定した。

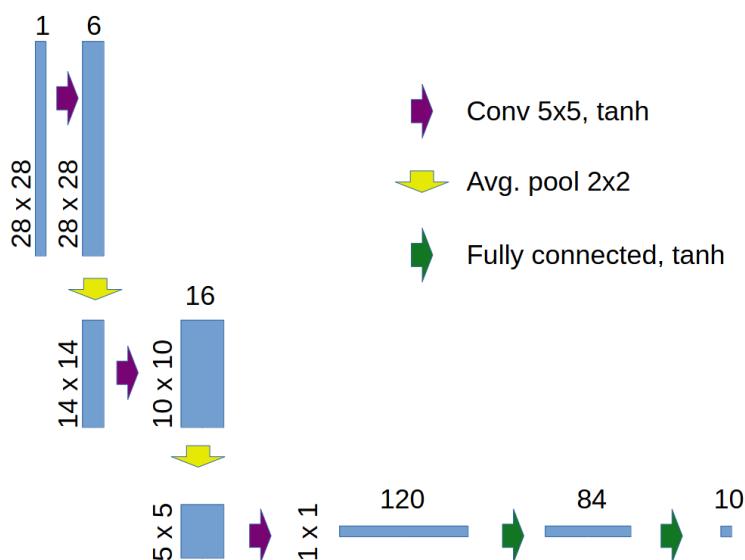


図1 LeNet-5 の構造

## MNIST を用いた評価結果

MNIST test データセットを用いた手書き文字の認識推定結果を表 1 に示す。教師データの平行移動なし LeNet-5 は、前回レポート (2024.04.20) の際に計算したものを再掲する。また計算時間の都合上、教師データの平行移動なしテストデータ平行移動ありのロバスト化 LeNet-5 については、MNIST test dataset の最初の 1000 例 (平行移動を含め 55000 通り) についてのみ計算した。

テストデータに平行移動を加えない場合、いずれのモデルを用いても誤答は 2-3% であった。一方、テストデータに平行移動を加えた場合は、LeNet-5 では誤答率が 40% となったが、ロバスト化 LeNet-5 や教師データに平行移動を加えた場合には誤答率は 2-5% に抑えられた。

## 考察

教師データ、テストデータともに平行移動なしの場合、LeNet-5 の誤答率が 3% であるのに対しロバスト化 LeNet-5 では誤答率 1.8% であった。これは、入力データに含まれる微小な平行移動の成分を適切に評価できたためである。テストデータに平行移動がある場合には、LeNet-5 に比して軽度であるとはいえ、ロバスト化 LeNet-5 でも著明な誤答率の上昇がみられた。これは先述のように、教師データが中央に「偏って」いるため

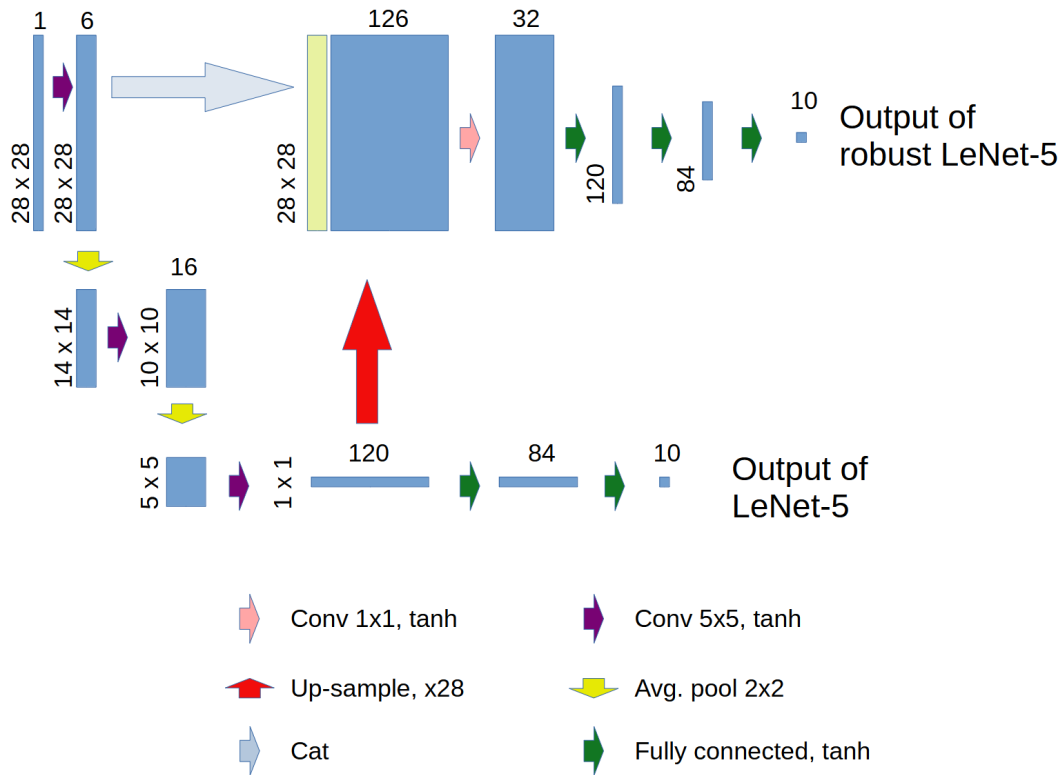


図2 ロバスト化 LeNet-5 の構造

表1 MNIST を用いた手書き数字画像の認識結果

モデル	教師データ 平行移動	テストデータの平行移動なし			テストデータの平行移動あり		
		正答	誤答	誤答率	正答	誤答	誤答率
LeNet-5	なし	9696	304	3.0%	331140	218860	40%
ロバスト化 LeNet-5	なし	9822	178	1.8%	38800*	16200*	29%
LeNet-5	あり	9658	342	3.4%	523358	26642	4.8%
ロバスト化 LeNet-5	あり	9771	229	2.3%	534034	15966	2.9%

\*: 1000 例についてののみ

に、pointwise convolution 以後の処理において、辺縁部分に対応する重み係数を適切に設定できなかったためである。

LeNet-5 において教師データに平行移動を加えた場合、テストデータの平行移動なしにおける LeNet5 は誤答率 3.4% であり、教師データに平行移動を加えなかった場合に比して成績が悪化した。これは教師データの平行移動により学習の効率が低下したためである。テストデータの平行移動ありの場合は誤答率 4.8% であり、教師データの平行移動なしの場合に比して著明な改善がみられた。これは平行移動を含めて適切な学習が行われた結果である。テストデータの平行移動なしに比して平行移動ありの誤答率が高かったのは、数字が隅に描かれている場合に誤答が多かったためである。LeNet-5 で誤答した例を図 3 から図 5 に示す。隅に描か

れた数字について誤答しやすいのは、数字が隅に寄っている教師データの数が、数字が中央近傍に書かれている教師データに比べて少ないことによると考えられ、LeNet-5 自体の特性ではない。

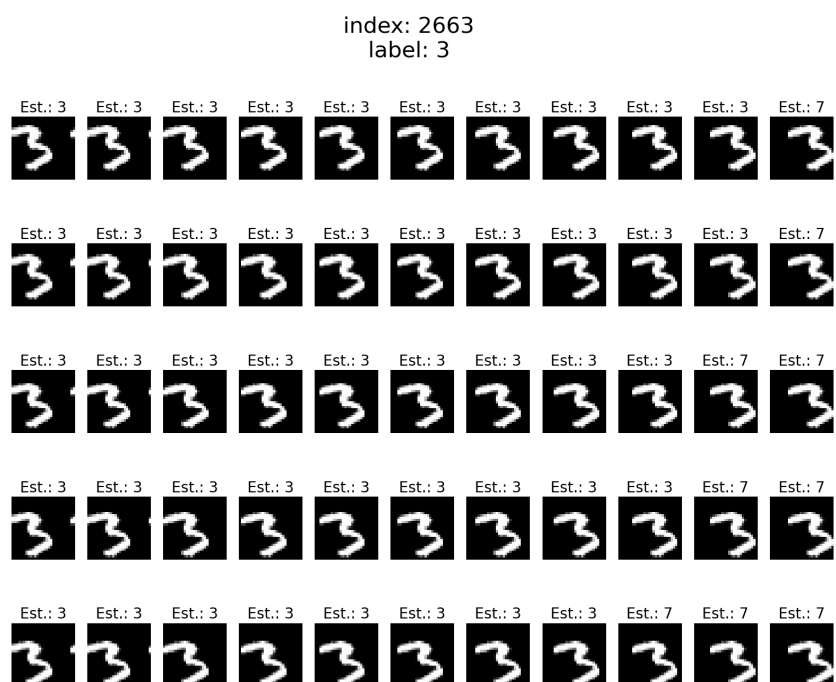


図3 LeNet-5, 教師データ平行移動あり (index 2663) の評価結果

ロバスト化 LeNet-5 では、テストデータの平行移動の有無にかかわらず、LeNet-5 に比して誤答率が低下した。図3から図5に示した問題について、ロバスト化 LeNet-5 で推定を行った結果を図6から図8に示す。図3と図6, および図4と図7を比較すると、LeNet-5 では隅に数字が描かれた場合に誤答した例についても、ロバスト化 LeNet-5 では正しく推定できている。これは教師データ、テストデータのいずれに対しても、数字が平行移動していることを考慮した学習および推定を行えたからである。一方、図8に示した例では、ロバスト化 LeNet-5 においても誤答が目立った。これは描かれた数字自体が2なのか3なのか判別し難いために誤答したものと考えられる。また、ロバスト化 LeNet-5 においても数字を隅に描いた場合に誤答しやすいことは、今回のモデルではロバスト化が不十分であることを示唆している。

今回のロバスト化 LeNet-5 は、構造に無駄が多く、最適化されていない。たとえば pointwise convolution の後のチャンネル数はもっと少なくても問題ないと思われるし、その後に2層の全結合層を挟んでいるのも計算精度にはあまり寄与していないと思われる。また、upsampling 部分について U-Net に倣って二段階で行った方がモデルとして美しい。U-Net に倣うという意味では、最後の全結合層でも畳み込みを行うという考えもありえるが、この部分を畳み込むことの理論上の意義がはっきりしない。

index: 3287  
label: 9

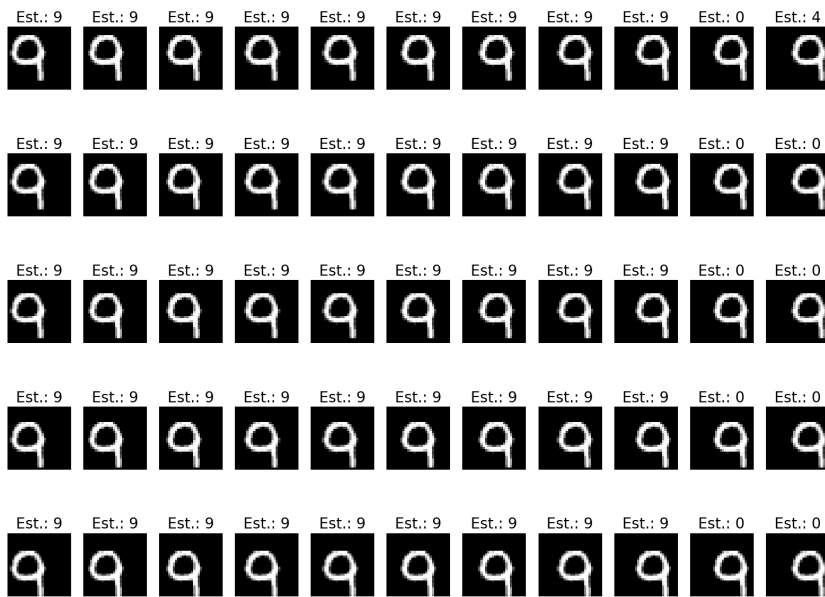


図4 LeNet-5, 教師データ平行移動あり (index 3287) の評価結果

index: 3962  
label: 3

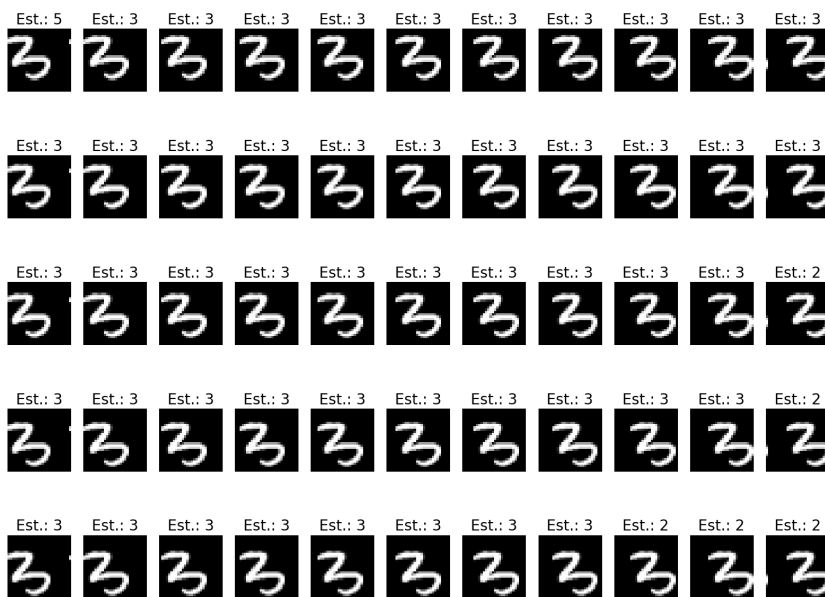


図5 LeNet-5, 教師データ平行移動あり (index 3962) の評価結果

index: 2663  
label: 3



図 6 ロバスト化 LeNet-5, 教師データ平行移動あり (index 2663) の評価結果

index: 3287  
label: 9



図 7 ロバスト化 LeNet-5, 教師データ平行移動あり (index 3287) の評価結果

index: 3962  
label: 3



図8 ロバスト化 LeNet-5, 教師データ平行移動あり (index 3962) の評価結果

## まとめ

LeNet-5 に、平行移動についてのロバスト化を加えることで、偏った位置に描かれた数字についても正しく判定することができた。このモデルは、U-Net など upsampling を伴うニューラルネットワークが位置についてロバスト性を有することの、理論的背景を理解する際に有用である。

## 参考文献

- 1) Lecun, Y., *et al.*, 'Gradient-Based Learning Applied to Document Recognition', Proc. IEEE **86**(11), 2278-2324 (1998).
- 2) The MNIST Database of Handwritten digits, <http://yann.lecun.com/exdb/mnist/> (2024.04.17 閱覽)
- 3) Ronneberger, O., *et al.*, 'U-Net: Convolutional Networks for Biomedical Image Segmentation', MIC-CAI 2015, Part III, LNCS 9351, 234-241 (2015).
- 4) Python.org, <https://www.python.org/> (2024.04.17 閱覽)
- 5) PyTorch, <https://pytorch.org/> (2024.04.17 閱覽)