Deep Learning を用いた画質劣化にロバストな手書き文字認識

2016SC087 高橋飛翔
指導教員：河野浩之

1　はじめに

手書き文字の認識は, 銀行, オフィスや家庭で利用する機会がある. 銀行では種類の多い手書きの伝票や印刷された伝票の文字をスキャナで読み取りそのままデータとして取り扱うことが可能である. オフィスや家庭では, 手書きのメモから書類まで様々な紙に書かれた文字をスキャナで読み取って文字認識しテキストデータとして管理ができる[*1]. しかし, 手書き文字は筆記者によって文字の形状が大きく異なり, 文字の記入位置, 文字の大きさが大きく変化する, 線分の切断やかすれが多いなどの問題点が存在し, 文字の認識を難しくしている [5].

前述のとおり, ノイズのある手書き文字は身近に存在し, そのような文字に対する研究も存在する. 文献 [6] では, Convolutional Neural Network (CNN) を用いているが, Max pooling はノイズのある手書き文字に有効ではないことが示されている. そこで, 本研究では Average pooling を使用する.

2　手書き文字認識の先行研究

Huang らの研究 [2] では, Deep Learning はネットワークの層が深くなるほどネットワークの劣化や勾配消失問題に対し, ResNet[1] をベースとして提案したネットワークを用いることで精度の向上を行った. 結果として回転した手書き漢字を対象とし, AlexNet[3] と提案手法の 2 種を用いて 4 方向 $\{0^\circ, 90^\circ, 180^\circ, 270^\circ\}$ のいずれかに回転した手書き漢字の方向検出精度は提案手法では 98.4% であり, AlexNet[3] は 96.7% であった. また, 回転文字の回転角度の検出誤差を比較した結果, ResNet[1] をベースとしたネットワークでは 0.6°であり AlexNet[3] より 1.7°精度が向上した.

Xu らの研究 [6] では, Gaussian ノイズまたは Salt-and-pepper ノイズを付与した手書き数字画像に対し, CNN により認識精度の向上を行った. CNN は入力層と出力層を含む 2 つの畳み込み層, 2 つの Max プーリング層からの 7 層である. 提案 CNN と三層パーセプトロンで認識精度の比較結果, Gaussian ノイズでは, PSNR が 15.8034dB では CNN は 0.7241, 三層パーセプトロンは 0.6901. 12.8856dB の場合, CNN は 0.5598, 三層パーセプトロンは 0.6305, Salt-and-pepper ノイズでは PSNR が 16.2916dB では CNN は 0.9718, 三層パーセプトロンでは 0.7188. 13.2293dB の場合, CNN は 0.9171, 三層パーセプトロンは 0.6754 の認識精度であった.

3　ノイズの分類

Xu らの研究 [6] では 2 種類のノイズを使用していたが, ノイズはガウス分布従った加法性ノイズである Gaussian ノイズ, 白と黒の斑点模様の Salt and Pepper ノイズ, 光子量の変化による Poisson ノイズ, ガウス分布従った乗法性ノイズである Speckle ノイズに分けられる [4].

4　Average pooling を用いた CNN の提案

本章では, Xu らの研究 [6] で Max pooling を用いた CNN はノイズのある手書き文字に適していないことから, Average pooling を用いた CNN を提案する. 図 1 に提案した CNN の構造を示す. CNN は入力層と出力層を含む 7 層のネットワークからなり, 1 層目は入力層, 2 層目は畳み込み層であり (5×5) サイズの畳み込みを行い 6 個の特徴マップを作成し, 活性化関数は Sigmoid 関数を使用する. 3 層目はプーリング層であり (2×2) サイズの Average pooling を行う, 4 層目も畳み込み層であり 2 層目と同様に (5×5) サイズの畳み込みを行い, 12 個の特徴マップを作成し活性化関数は Sigmoid 関数を使用する. 5 層目もプーリング層であり (2×2) サイズの Average pooling を行う, 6 層目では全結合を行うので Flatten を行い, 7 層目で全結合を行い, 活性化関数を Softmax 関数とし, 出力とする.

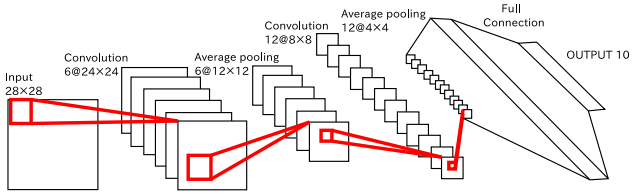

図 1　提案した Convolutional Neural Network

5　ノイズのある手書き文字の認識実験

本章では, 4 章で提案した CNN を用いてノイズのある手書き文字の認識をし, 文献 [6] で使用された CNN, MLP との認識精度の比較をする. ノイズのある手書き文字は, 手書き文字データセットである MNIST[*2]を使用し, ノイズを手動で生成する. 使用するノイズは 3 章で説明した 4 種類のノイズである.

本実験では, プログラミング言語は python 3.6.8, ノイズの生成は Scikit-image , Deep Learning の構築には Keras2.2.5, OS は Windows 10 Home, GPU は Google Colaboratory を使用するので NVIDIA® Tesla® K80 である.

[*1] 富士通研究所, https://www.fujitsu.com/jp/group/labs/resources/tech/techguide/list/character-recognition/p07.html, Accessed, Dec, 31, 2019.

[*2] THE MNIST DATABASE of handwritten digits, http://yann.lecun.com/exdb/mnist/, Accessed, Nov, 28, 2019.

Deep Learning では, 勾配法は Stochastic Gradient Descent(SGD), パラメータ学習時の学習率, バッチサイズ, エポック数は 0.7, 60, 300 とし, 損失関数は平均二乗誤差を適用する. 3 種類のネットワークに対し, MNIST の訓練用データセット 60,000 個使用し訓練を行った. ノイズの生成では, Gaussian ノイズと Speckle ノイズは分散を 0.05 から 0.5 の間で 0.05 ずつ増加させる. Salt and Pepper ノイズはノイズ量である amount を 0.0 から 0.5 の間で 0.05 ずつ増加させる. ノイズを加えるデータは, MNIST のテスト用データ 10,000 個である.

表 1, 表 2, 表 3, 表 4 に 4 種類のノイズを 3 種類のネットワークで認識させた結果を示す. CNN 同士で最も精度に差があったのは, Gaussian ノイズの場合, Average Pooling を適用した CNN は Max Pooling を適用した CNN より認識精度が 56.82% から 73.41%, Salt and Pepper ノイズの場合, 43.48% から 69.52%, Poisson ノイズの場合, 98.25% から 98.78%, Speckle ノイズの場合, 50.20% から 56.56% になる. また, Peak Signal to Noise rate (PSNR) を用いることでもとのデータとノイズのあるデータとの不一致を数値化を行う.

表 1 Gaussian ノイズ

variance	PSNR [dB]	CNN(Average Pooling) [%]	CNN(Max Pooling) [%]	MLP [%]
0.05	15.8039	95.97	95.50	82.89
0.1	12.8893	73.41	56.82	62.71
0.15	11.2499	41.56	23.08	41.84
0.2	10.1667	22.44	14.08	28.90
0.25	9.3757	14.92	11.75	20.94
0.3	8.7921	11.44	10.39	16.00
0.35	8.3208	10.72	9.900	13.09
0.4	7.9488	10.45	10.15	11.39
0.45	7.6364	10.04	10.70	10.51
0.5	7.3735	9.73	9.76	10.03

表 2 Salt and Pepper ノイズ

amount [%]	PSNR [dB]	CNN(Average Pooling) [%]	CNN(Max Pooling) [%]	MLP [%]
0	−	98.80	98.42	88.61
0.05	16.2950	96.34	95.08	86.72
0.1	13.2176	87.40	76.71	80.84
0.15	11.4441	69.42	43.48	68.98
0.2	10.1968	47.55	22.82	53.43
0.25	9.2149	28.64	14.43	38.93
0.3	8.4215	18.80	12.18	28.13
0.35	7.7465	13.98	11.12	20.79
0.4	7.1630	11.97	10.67	15.88
0.45	6.6500	10.60	10.16	12.73
0.5	6.1946	10.27	10.12	11.50

表 3 Poisson ノイズ

PSNR [dB]	CNN(Average Pooling) [%]	CNN(Max Pooling) [%]	MLP [%]
30.3823	98.78	98.25	88.50

表 4 Speckle ノイズ

variance	PSNR [dB]	CNN(Average Pooling) [%]	CNN(Max Pooling) [%]	MLP [%]
0.05	25.0860	98.64	98.16	88.05
0.1	22.2134	98.38	97.70	87.44
0.15	20.5686	97.42	96.04	86.36
0.2	19.4821	95.63	92.96	85.22
0.25	18.6912	92.07	87.53	83.74
0.3	18.0839	85.87	81.35	81.16
0.35	17.6123	77.77	72.85	78.82
0.4	17.2257	70.61	65.35	75.64
0.45	16.9022	63.17	57.37	73.48
0.5	16.6383	56.56	50.20	70.41

文字部分の画素値は 255 の白色, 背景部分の画素値は 0 の黒色である. 背景部分に画素値 255 の白色が混入した場合, Max pooling を行うと画素値 255 が反映され, 背景であると誤認識してしまう. しかし, Average pooling であれば, 画素値の平均をとるので誤認識する可能性を防ぐことができる可能性がある.

6 むすび

本研究では, 4 種類のノイズのある手書き文字に対し, Average pooling を用いた CNN で文字認識をした. 5 章で示した通り, ノイズ量が少ない場合, Average pooling を使用した CNN は Max pooling を使用した CNN より精度が上昇した.

今後の課題は, Max pooling, Average pooling のどちらも大量のノイズのある文字に対して効果に差が無いことから, pooling 操作をなくすことで認識精度が上昇するか確かめることである.

参考文献

[1] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” Proc. 2016 IEEE Conference on CVPR, 2016.

[2] Z. Huang, Q. Zhang, “Skew correction of handwritten chinese character based on resnet,” Proc. 2019 International Conference on HPDB&IS, pp. 223-227, 2019.

[3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” Proc. the 25th International Conference on NeurIPS, Vol. 1, pp. 1097-1105, 2012.

[4] H. R. Mamatha, S. Madireddi, and M. K. Srikanta, “Performance analysis of various filters for de-noising of handwritten kannada documents, ” Article in IJCA, Vol. 48, No.12, pp. 29-37, 2012.

[5] 村上伸一, “7.2 手書き文字認識技術”, 画像処理工学, pp. 123-128, 東京電機大学出版局, 2004.

[6] Z. Xu, Y. Terada, D. Jia, Z. Cai, and S. Gao, “Recognition effects of deep convolutional neural network on smudged handwritten digits,” Proc. 2018 5th International Conference on ICISCE, pp. 412-416, 2018.