The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

U-Net による手書き文字画像内のノイズ除去

Removing Noise from Handwritten Character Images using U-Net

小松里奈
Rina Komatsu
上智大学
Sophia University

ゴンサルベス タッド
Tad Gonsalves
上智大学
Sophia University

Offline handwritten character recognition still remains a tough challenge for AI techniques and algorithms. This is because handwritten documents frequently introduce some amount of noise in the images during the scanning procedures. The presence of noise in the scanned images make them murky and/or blurred and therefore hard to read. In this study, we tried using the CNN architecture named “U-Net” to analyze 607,200 sample images consisting of 3,036 Japanese characters. Our results indicate that the “U-Net” has sufficient ability to get rid of noise from characters and enhance the parts of strokes even though there are a huge variety of handwritten styles.

1. 概要

ある手書きの文書をスキャンし、コンピューターへ文書の内容を出力した際、スキャン時で生じたぼやけ・かすれなどの「ノイズ」が文書内の文字の一部にかかってしまい、読み取りおよび認識の障害となる場合がある。

ノイズのある手書き文字画像から文字という有用なピクセル情報領域を取り出すのに「セグメンテーション」という手段が存在する。領域は、色、エッジ、近傍との類似性などの特徴量を使って構成が行われる[Jan Erik Solem 2013]。しかし、特徴量をとりだすためのパラメーターが一定値で決められている場合、強く表れるノイズが特徴量となり誤って文字領域として抽出される場合もある。そのことから、すべての書体のパターンをカバーできるとは限らず、手書き文字の特徴量が失われてしまうこともある。

本研究では、ノイズ入りの入力画像に対しノイズをニューラルネット側で除去し、人間にとって読み取りやすい結果画像を出力できるモデル作成を U-Net を用いて試みを行ったものである。

2. U-Net の概要

U-Net は論文”U-Net: Convolutional Networks for Biomedical Image Segmentation”にて紹介された、入力された画像に対しセグメンテーションを行う CNN アーキテクチャである[Olaf Ronneberger 他 2015]。

また論文”Image-to-Image Translation with Conditional Adversarial Networks”では,出力画像を生成するにあたってこの U-Net の構造を用いており、線画のみの画像に色付け、衛星写真に対し地形のセグメンテーションを行うなどの成果が示されている[Phillip Isola 他 2017]。

本研究では、U-Net を用いればノイズの入った手書き文字画像にて、文字部分のみをセグメンテーションとして取り出せるのではないかと推測した。以下では、U-Net の構造と本実験で作成した U-Net の構造について紹介をする。

2.1 U-Net の構造について

U-Net は構造内に、コンテキストを取得するための縮小パスと正確な局所情報を可能にする拡張パスが含まれている [Olaf Ronneberger 他 2015]。 以下の図 1 に U-Net の構造を表した図を示す。縮小パスは図 1 での左側、コンボリューションネットワーク層で入力画像の畳み込みを行っている部分にあたる。一方で拡張パスは図 1 の右側、縮小パスによって畳み込まれた情報をアップサンプリングしている部分にあたる。

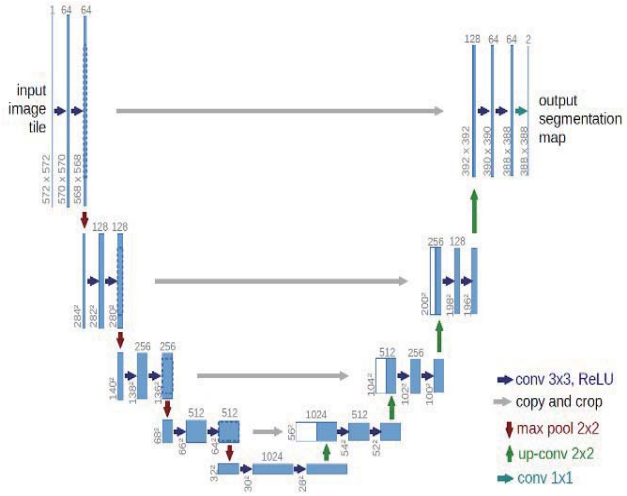


図 1. U-Net の構造 ([Olaf Ronneberger 他 2015])

U-Net では、縮小パスでコンボリューション層からの出力を、アップサンプリングにてデコンボリューションするときにマージを行っているのが特徴となっている。

2.2 実験で作成した U-Net について

図 1 での構造では、入力画像のサイズと出力画像のサイズが異なっている。本研究では、同じサイズでの入力画像、出力画像、教師用画像を可視化し比較を行えるようにするため、U-Net の構造を以下の図 2 のように構成した。

本研究で構成したものでは、縮小パスでのコンボリューション時にパラメータにフィルタサイズ=3×3、ストライド=1、パディング=1、拡張パスでのデコンボリューション時にフィルタサイズ=4×4、ストライド=2、パディング=1 と、コンボリューションおよびデコンボリューションの後で入力と出力でのサイズが異ならないように値を指定した。

連絡先：小松里奈，上智大学，r_komatsu@outlook.com

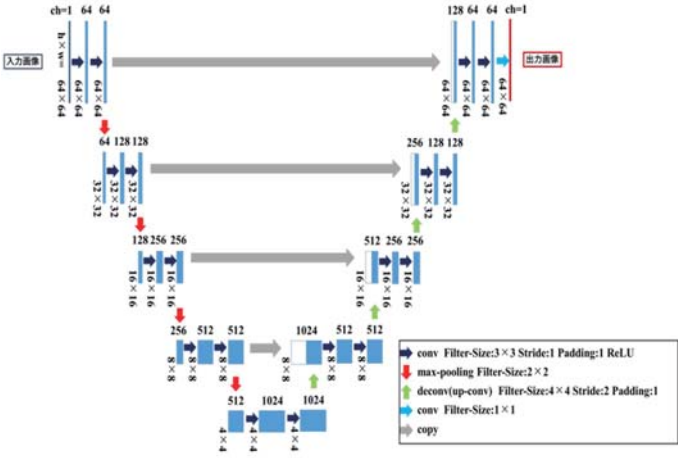


図 2. 本研究で作成した U-Net の構造

3. 実験

3.1 実験で使用した手書き文字画像データについて

本研究では、電子総合技術研究所および独立行政法人産業技術総合研究所(電子技術総合研究所の後継組織)が提供する ETL 文字データベースから、ETL-9G Character Database [電子技術総合研究所 1973-1984]のデータセットを使用した。

このデータセットについて、対象とした文字種は JIS 第一水準漢字 2965 種、ひらがな 71 種の合計 3036 種類であり、筆記者数延べ 4000 人に 1 文字ずつ OCR シート上に漢字を書いてもらい、全サンプル数 607200 を収集したセットとなっている。

3.2 実験環境について

本研究ではプログラム言語は python、画像処理を行うライブラリは openCV、深層学習を行うライブラリは chainer を用いて、プログラム作成を行った。

3.3 実験手順

手書き文字画像データをグレースケール形式で読み取ったものに対し、以下の手順で実験を行った。

(1) トレーニング・テスト用画像の作成および分割

トレーニング・テスト用画像の作成にあたって、3.1 にて記した手書き文字データの各サンプルに対し、(高さ,幅)=(64,64)にリサイズした後、ガウシアンノイズをかけた上にぼかしを加えた画像処理を行った。一例として以下の図 3 に、画像処理前の手書き文字画像、処理後の手書き文字画像を並べたものを示す。分割では全サンプル数 607200 に対し、トレーニング用画像:テスト用画像=7:3 となるように行った。

図 3. 画像処理前・画像処理後の手書き文字画像

(2) U-Net のトレーニング実施

実験手順(1)で生成したトレーニング用画像を、図 2 で示した構成を取り入れた U-Net でのインプットとして、トレーニングを行った。

U-Net へトレーニング画像をインプットする際は、トレーニング画像全体の画素値を 255 で割り正規化を行った。

学習の際は、U-Net により生成されたノイズ除去画像と実験手順(1)にて画像処理を行う前の画像との誤差 Loss を平均絶対誤差関数で計算。その値をもとにモデル内のパラメータの更新を行った。

トレーニングでの学習方式はミニバッチ学習を採用した。このとき、ミニバッチのサイズは 1 つにつきトレーニング用画像 200 枚分に設定した。

(3) トレーニング済の U-Net によるテスト用画像のノイズ除去

ある程度のエポック数をこなした U-Net モデルを用いて、実験手順(1)で生成したテスト用画像をインプットとしたノイズ除去を行った。

4. 結果

4.1 トレーニング・テスト実施での Loss の評価

トレーニング・テスト実施時にて、各エポックでの Loss の値を記録した結果をグラフにまとめたものを図 3 に示す。エポック数が 10 を超えたあたりから、Loss の値が 0.01 前後に収まりグラフ上で平坦な線を描いていった。

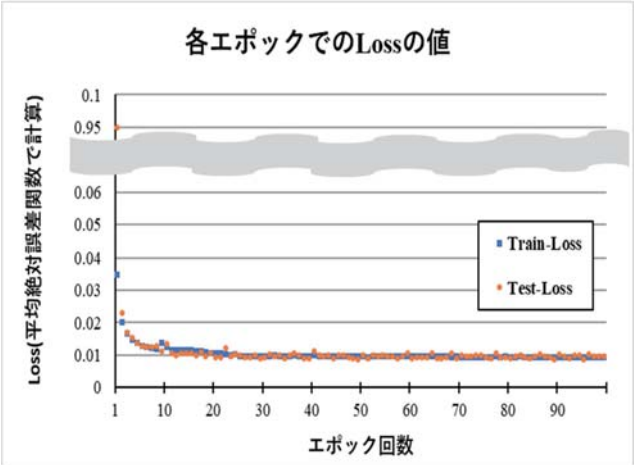


図 3. 各エポックでの Loss の値のグラフ

トレーニング・テストでの Loss の値は、エポック数を重ねるごとに小さくなっていった。エポック数を重ね、トレーニングでの Loss の値が小さくなる一方でテストでの Loss の値が増えていくという過学習の特徴[岡谷貴之 2016]がグラフ上で現れなかったことから、深層学習上での理想的なトレーニングおよびテストが行えたのを確認することができた。

4.2 テスト用画像の可視化による評価

4.1 にてエポック回数を増やすにつれ、トレーニングおよびテストでの Loss の値が減っていったのを確認することができたので、次に実際にどの程度ノイズを取り除くことができたかをテスト用画像をモデルのインプットとして用いて可視化を行った。

図 4 にて、エポック回数 0,10,30,50,100 をこなした U-Net モデルを用いて可視化を行った結果をまとめたものを示す。

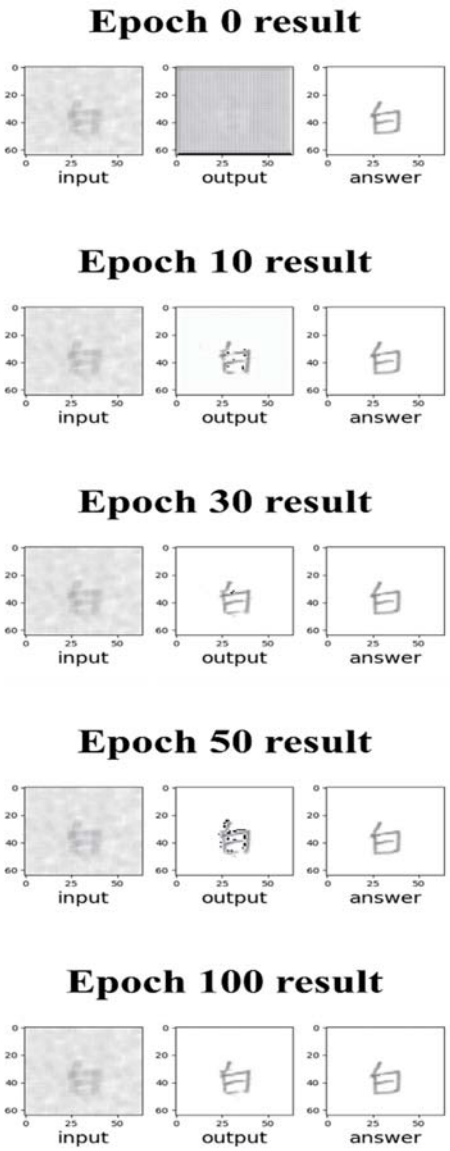


図 4. テスト用画像 input・出力画像 output・正解画像 answer を並べた図(上からエポック数 0,10,30,50,100 をこなした U-Net を用いての実施)

エポック数 0 の時点では、書かれた漢字がどの部分にあたるかをとらえきれず出力画像が真っ黒となっている。しかし、エポックを 10 回以上こなしたモデルの結果では、わずかに黒色のモザイクがあるものの、手書き文字のストローク部分が強調され、どのような漢字が書かれていたかを読むことが出来る程度でぼかしを除去できている。さらに、図 4 の例では「白」という文字を「臼」という字に間違えることなく出力を行うなど、正解画像にほぼ近い出力画像を U-Net を通じて得ることができた。

しかしながら、画像らの中には、どんなに多くのエポック回数をこなした U-Net を用いて、除去を行っても出力画像の結果にて文字の一部が欠けてしまったものも存在した。以下の図 5 にその一例を示す。

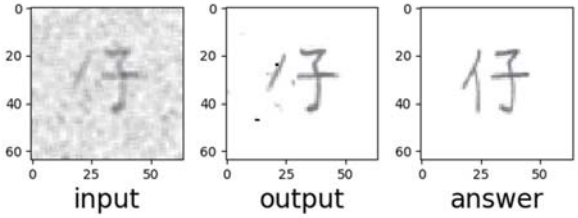


図 5.エポック回数 100 をこなした U-Net を用いてノイズ除去

全体を通して、U-Net はノイズを文字情報領域と誤ることがなく、手書き文字のストローク部分を強調し判読できる文字へと出力することを確認することができた。

5. 考察および今後の課題

ノイズのかかった手書き文字画像を判読できる文字画像に出力するのに、U-Net を利用することが有効手段の一つであることを確認することが出来た。

U-Net が手書き文字のストローク部分を強調することでノイズの除去ができた理由について、上記に記した通り、縮小パスではコンテキストを取得し、拡張パスでは正確な局所情報の出力を可能にしている [Olaf Ronneberger 他 2015] 。

縮小パスに含まれていたコンボリューション、畳み込みの働きについて、フィルタにある特徴的な濃淡パターンと類似した濃淡パターンを入力画像のどこにあるかを検出する働きがあることが指摘されている [岡谷貴之 2016] 。 以下の図 6 にて、縮小パスでの最初のコンボリューション時に出力されていたチャンネルの一部分を可視化したものをこなしたエポック数別でそれぞれ示す。最初のコンボリューション時にて、エポック数 0 をこなした U-Net では入力画像そのもの、手書き文字部分に縁取りを行ったものが出力されている。一方で、エポック数 100 をこなしたものでは、手書き文字のとめ、はね、はらいやテクスチャが出力されている。

拡張パスに含まれていたデコンボリューションでの働きについては、特徴抽出後に画像全体で密な情報を保持しつつ画像の拡大を行うことが指摘されている [島田直希 他 2017] 。 以下の図 7 にて、最後のデコンボリューション時に出力されていたチャンネルの一部分を可視化したものをこなしたエポック数別でそれぞれ示す。こちらでは、エポック数 0 をこなした U-Net では白黒のモザイク状のものが出力されていた。エポック数 100 をこなしたものではすべてのチャンネルが図 7 のように真っ暗なものが出力されていた。

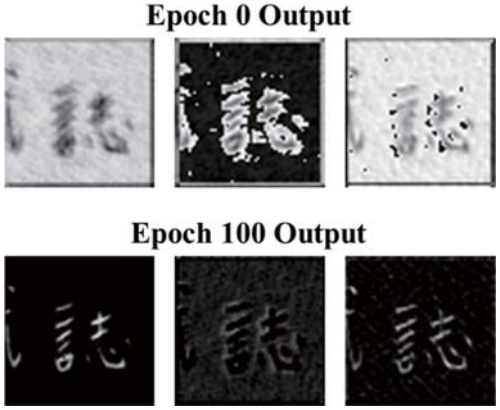
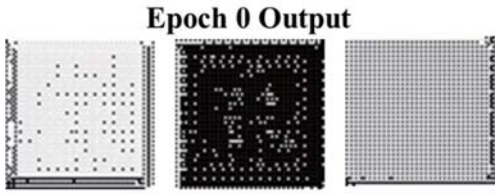


図 6.縮小パスでの最初のコンボリューション時に出力されたチャンネルの一部分



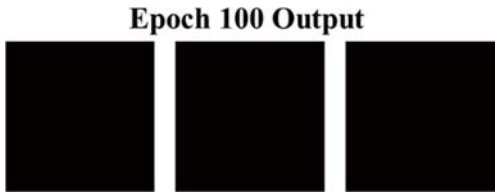


図 7.拡張パスでの最後のデコンボリューション時に出力されたチャンネルの一部分

図 6、図 7 の可視化結果から、縮小パスでは手書き文字にあったとめ、はね、はらい、テクスチャといった特徴量が、文字のストローク部分を伝えるコンテキストとなり、一方で拡張パスでは真っ黒な画像というノイズのない局所情報が出力され、「画像のストローク部分」と「ノイズのないバックグラウンド情報」がマージされることで、ノイズが除去された手書き文字を U-Net を通じて出力されたことが考えられる。

本研究では、手書き文字画像のノイズ除去を行えた。しかしながら、除去を行った対象は 1 文字の手書き画像であるので、手書き文字のドキュメントを対象に全体的にノイズの除去を行う場合は、ドキュメント内にて 1 文字 1 文字ノイズの除去を行うのは大変手間である。次のステップとして、ノイズの入った手書き文字のドキュメントをインプットとして、判読できるレベルでのドキュメント画像を出力できる U-Net の構築を行いたい。

参考文献

[Jan Erik Solem 2013] Jan Erik Solem: 実践 コンピュータービジョン, オライリー・ジャパン, 2013.

Olaf Ronneberger, Philipp Fischer, Thomas Brox : U-Net: “Convolutional Networks for Biomedical Image Segmentation”, Medical Image Computing and Computer-Assisted Intervention (MICCAI) Vol. 9351, pp234-241, Springer, LNCS, 2015.

Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros : “Image-to-Image Translation with Conditional Adversarial Networks”, CVPR 2017, 2017.

[電子技術総合研究所 1973-1984] 電子技術総合研究所: Japanese Technical Committee for Optical Character Recognition, ETL 文字データベース, 1973-1984.

[岡谷貴之 2016] 岡谷貴之: 深層学習, 講談社, 2015.

[島田直希 他 2017] 島田直樹, 大浦建志: Chainer で学ぶディープラーニング入門, 技術評論社, 2017.