基盤研究（A）

「平城宮・京跡出土木簡とその歴史環境のグローバル資源化」成果報告書

東京農工大学

中川正樹

１．　はじめに

表記の科学研究費補助金を分担させていただき，章末の文献欄に示した発表をすることができた．ここでは，主要な成果を要約する．

２．　発表成果

主要な成果を要約する．

（１）劣化した木簡文字の画像改善［7］

深層ニューラルネットワークにより，ノイズが少ない文字パターンの単独文字認識率は飛躍的に向上したが，歴史文書のように画像の劣化が激しいと文字認識率は非常に下がる．また，歴史文書のように学習パターンが不足し，正解コードを付与するコストが過大になる対象では，深層ニューラルネットワークの弱点が露呈する．

本研究は，歴史文書で画像が著しく劣化しノイズが多い文字画像を復元するために，CAGAN と命名したアテンション付きの敵対的生成ネットワークを提案した．これにより，文字認識率が改善し，考古学者の解読を助けることが期待できる．深層ニューラルネットワークはスキップ接続を含んだ U-Net に基づき，敵対的損失（大域的損失）と階層的文字アテンション損失（局所的損失）を縮小させるべく学習させる．奈良の平城京から出土した木簡から切り出した汚損の著しく，かつ，よく使われる漢字 118 字種に対して実験を行ったところ，文字形状は適切に復元され，文字認識率が有意に改善することが確認された．

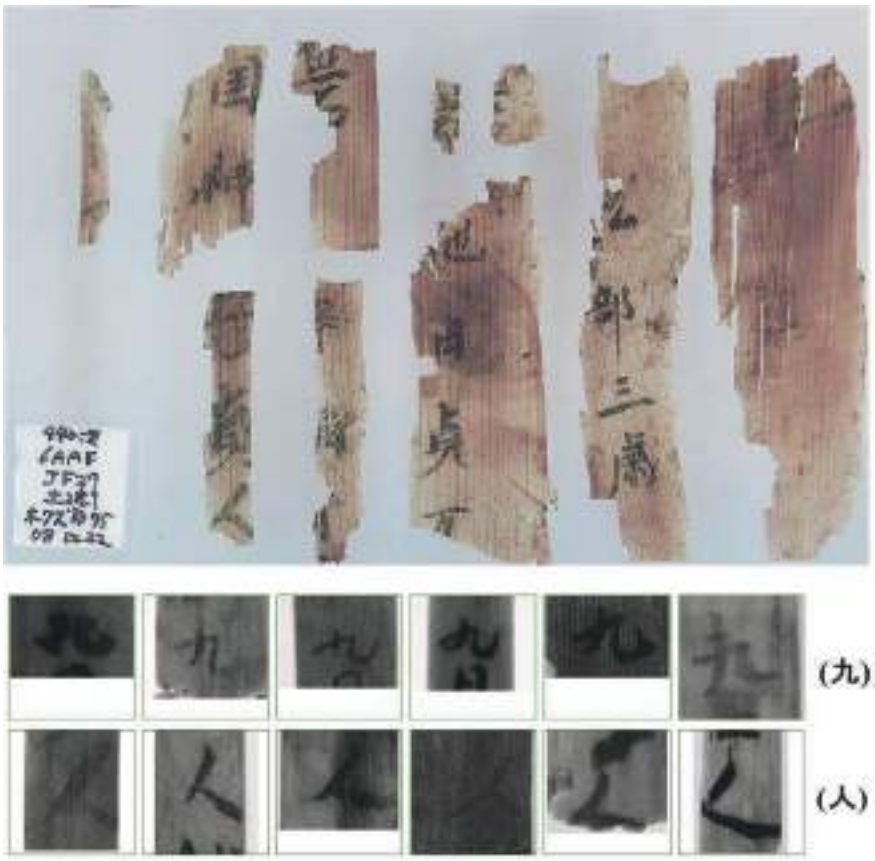


図１－１．木簡とその破片，「九」と「人」の文字画像.

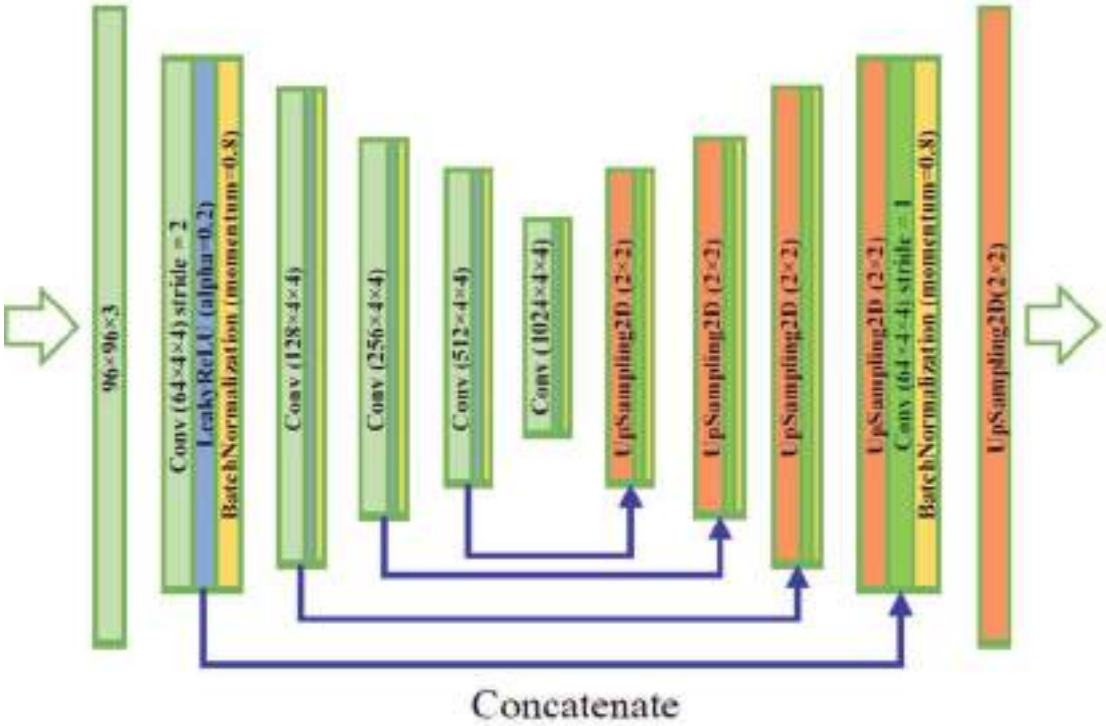


図１－２．スキップ接続を含んだU-Net.

図１－３．文字画像の復元例.

（２）変体仮名の認識［2］

本研究では，PRMU コンテストでの日本語歴史文書のくずし字認識モデルを示した．PRMU コンテストは３つのレベルがある：レベル 1：単文字認識；レベル 2：縦方向の 3 文字認識；レベル 3：縦横方向の 3 文字以上の文字列である．本研究では，PRMU コンテストで優勝したレベル 2 とレベル 3 の方法に注目する．基本的に，文字切り出しをしない手法を採用し，特徴抽出のための CNN；フレーム予測のための双方向長短期記憶（BLSTM）；テキスト認識のための CTC からなる．これを Deep Convolutional Recurrent Network（DCRN）と命名した．また，DCRN モデルにおいて，事前学習された CNN アプローチと End-to-End アプローチをレベル２のデータセットで比較した．次に，レベル 3 では DCRN を適用する前に，行切り出しと複数行連結の方法を提案した．さらに，レベル３に対し，2 次元の BLSTM（2DBLSTM）の方法も評価した．そして，相互検証（Cross-validation）による評価を示した．言語モデルなしで，レベル２の３文字列の認識で 89.10%の精度を達成し，レベル３の縦横方向の 3 文字以上の文字列認識で 87.70%の精度を達成した．

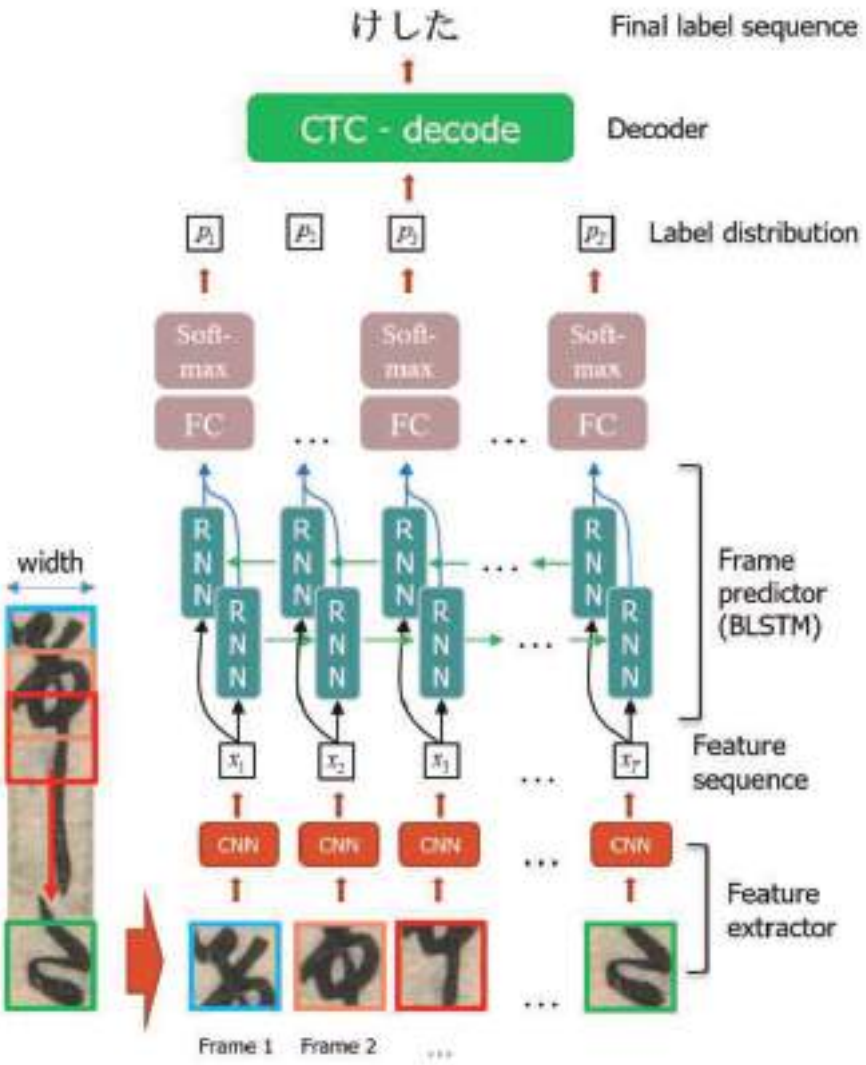


図２－１．DCRNのネットワーク構造.

(a) 正しく認識された例.

(b) 誤認識例（赤字が誤認識箇所).

図２－２．DCRN-o_12_Lv3による認識例.

（３） 日本語古文書の認識 [8,4]

本研究では，行切り出しなしで日本語歴史文書の複数行のテキスト画像を認識するための，注意メカニズムによる畳込み Sequence to Sequence（ACseq2seq）モデルを提案した．本モデルは，１）CNNによって入力画像から特徴の列を抽出する特徴抽出器；２）BLSTM（双方向の長短期記憶）によって特徴の列をエンコードするエンコーダ；３）注目メカニズムを利用したLSTMによって最終的な文字列を生成するデコーダ，の3つのからなる．また，デコーダの注目ベクトルとソフトマックス層の間にResidual LSTMネットワークを導入した．本モデルは，標準のクロスエントロピー損失関数によってエンドツーエンドで学習させた．実験では，PRMUコンテストのくずし字データセットでACseq2seqモデルを評価し，実験結果は提案モデルが最先端のパフォーマンスを達成することを示した．

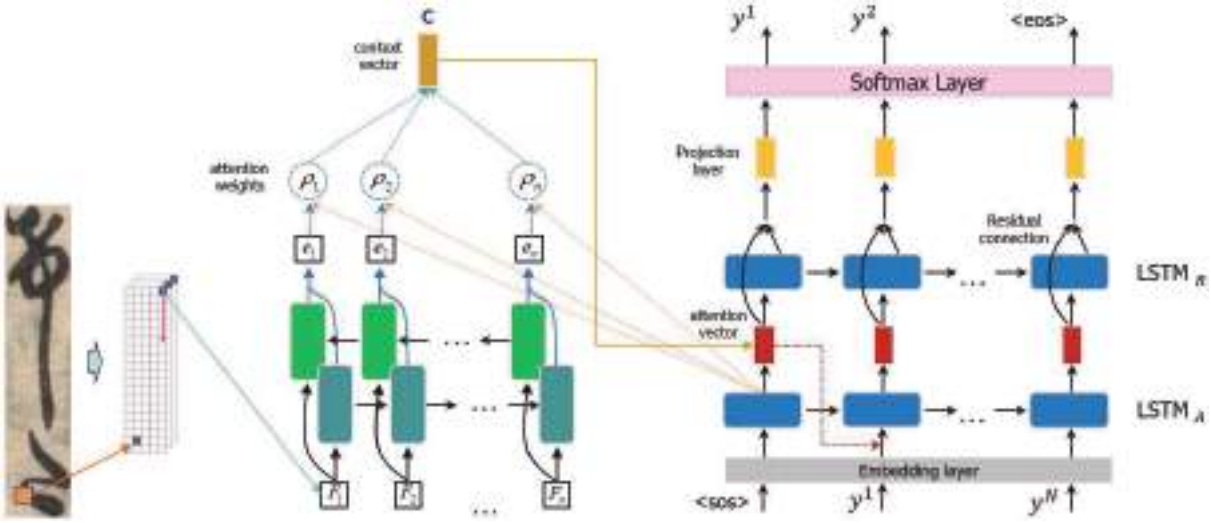


図３－１．ACseq2seqのネットワーク構造.

図３－２．ACseq2seqにより透けにも関わらず正認識された例.

図３－３．ACseq2seqによる誤認識例（赤字が誤認識箇所）.

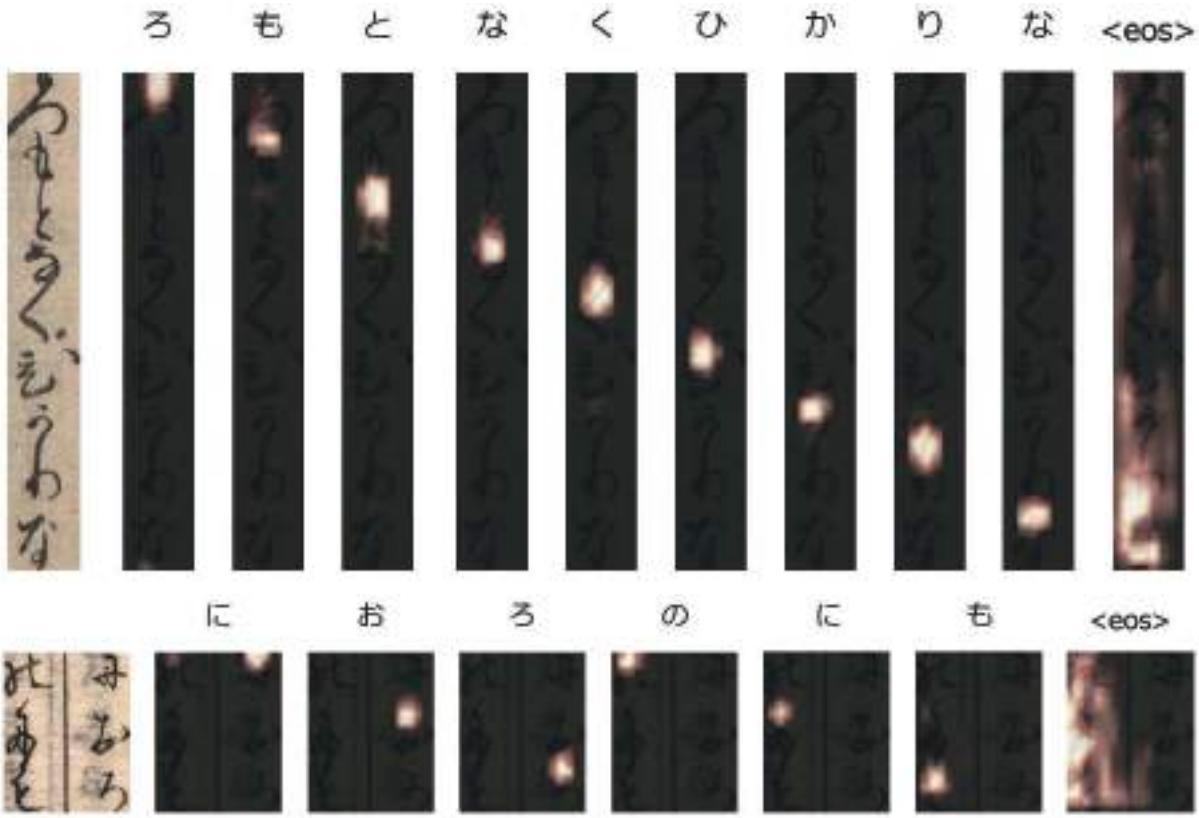


図３－４．縦書きの１行と複数行を認識した際のアテンションの視覚化.

引き続き，行切り出しなしで日本の歴史文書の複数行テキストを認識するために，注意メカニズムによる縦横エンコーダデコーダ（ARCED）モデルを提案した．本モデルは，特徴抽出器，横縦-エンコーダ，および注意メカニズムによるデコーダの 3 つの部分がある．エンコーダに横縦 BLSTM を導入し，デコーダに Residual LSTM を導入した．本モデルを，標準のクロスエントロピー損失関数によってエンドツーエンドで学習させた．実験では，PRMU コンテストのくずし字データセットで ARCED モデルを評価した．実験結果は提案モデルが最先端のパフォーマンスを達成することを示した．さらに，エンコーダの横縦 BLSTM とデコーダの Residual LSTM が，ARCED モデルの精度向上に寄与することを示した．

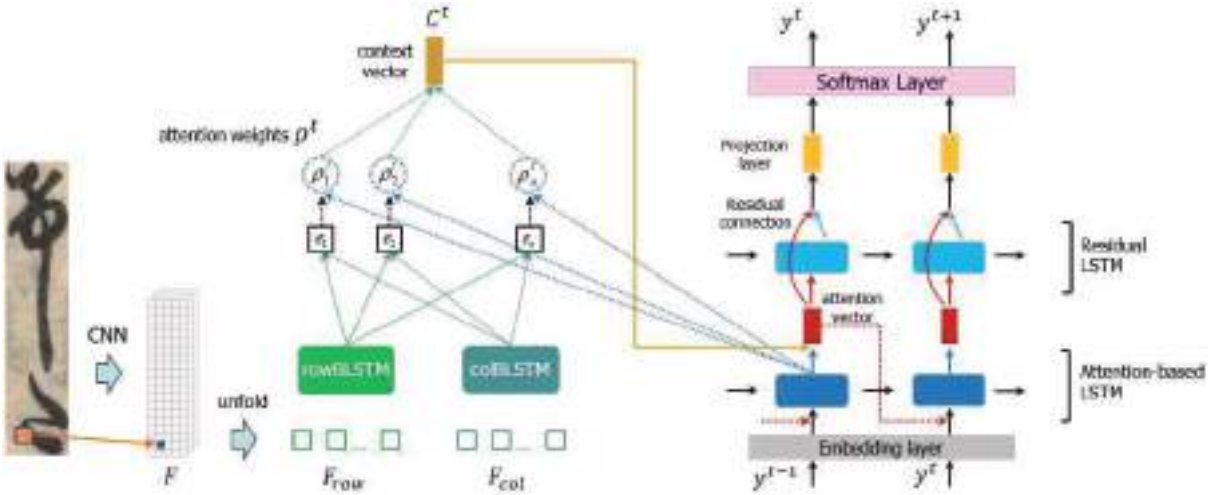


図３－５．ARCED のネットワーク構造（パラメータ $\rho t, Ct, yt$ は，時刻 t における，注意の重み，文脈ベクトル，そして，出力を示す）．

（４）古文書画像の領域分割［15］

日本の歴史的文書画像からの領域分割に，完全畳み込みネットワーク（FCN）を使用することを提案した．学習されたFCNモデルは，様々な背景と様々なサイズの文書画像からテキスト部分を切り出す．ただし，FCNの欠点は，画素レベルで正解ラベルを付与しなければならないことであり，特に歴史的文書の場合は費用がかかる．そこで，文字領域ごとに大津の二値化法を採用することにより，日本古典籍（PMJT）データセットに属する全部ドキュメント画像の画素にラベルを準備した．他の問題に，背景画素数とテキスト画素数の不均衡がある．そこで，学習中には背景画素数とテキスト画素数の比率に基づいて，重み付きパ

ラメータを勾配に乗算する．実験の結果によると，画素レベルのセグメンテーションの精度は約92%である．訓練されたモデルはより少ない労力で多数の画像を処理するために使用でき，歴史的な文書処理分野の研究者にとって有用である．

a) Input image　　b) Prediction

図４－１．表形式の歴史文書のセグメンテーション例.

a) Input image.　　b) Prediction.

図４－２．図を含んだ文書のセグメンテーション例.

（５）歴史文書における筆者識別［12］

本研究は，歴史文書における筆者識別ための注意層に基づくニューラルネットワークを提案した．このネットワークは，二値化またはセグメンテーションなどの前処理段階を必

要としない．これは，畳み込みニューラルネットワーク（CNN）を使用した特徴抽出器，キーポイントを選択するための注意層，そして，抽出されたキーポイントを集約することによって代表ベクトルを形成するための一般化された深部神経VLADモデル，の3つの主要な部分からなる．ネットワーク全体は，交差エントロピー損失関数とTriplet 損失関数の組合せによってエンドツーエンドで学習される．多く筆者からの約120,000枚の歴史文書破片を集めたHisFragIR20データセットを用いて評価したところ，最新の結果と比較して，MAP（mean average precision)と一位候補が正解である率（第一位候補正解率）が優れていることを示した．

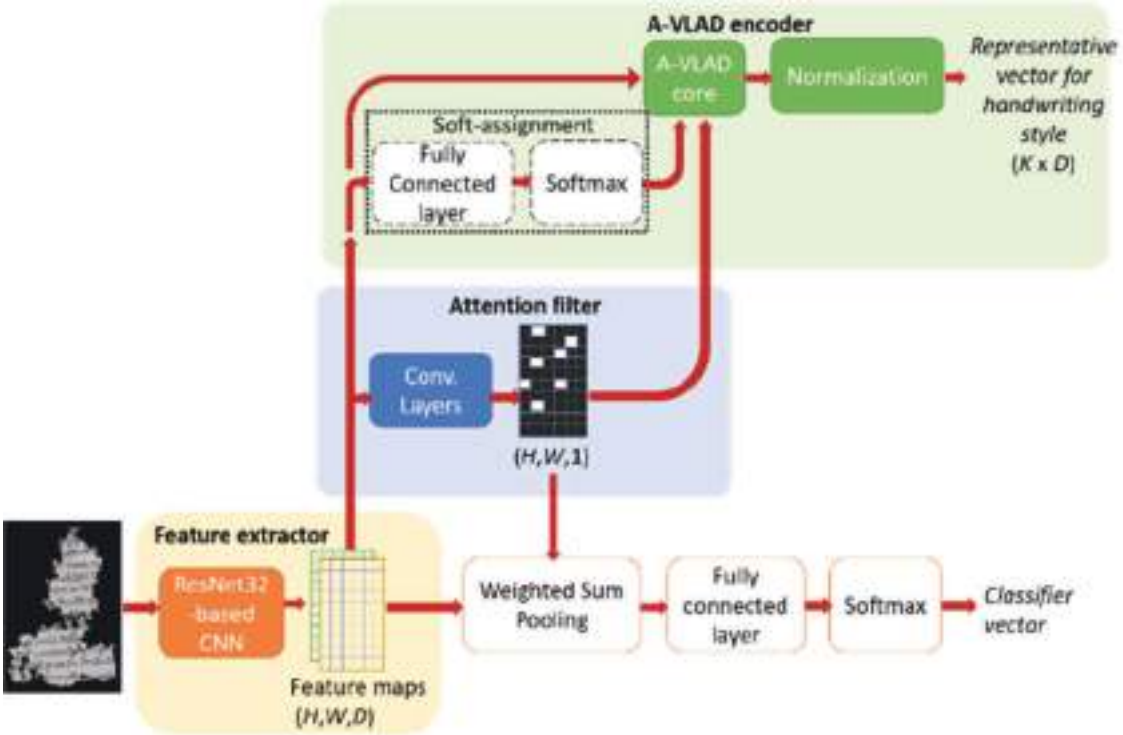


図５－１．A-VLADのネットワーク構造.

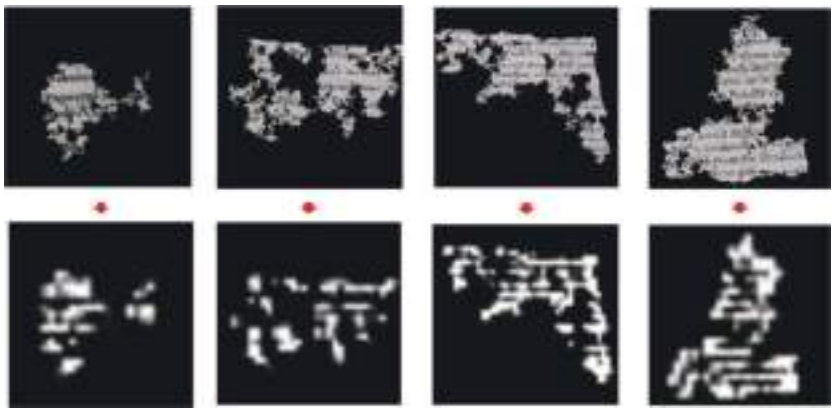


図５－２．入力画像と注意フィルター.

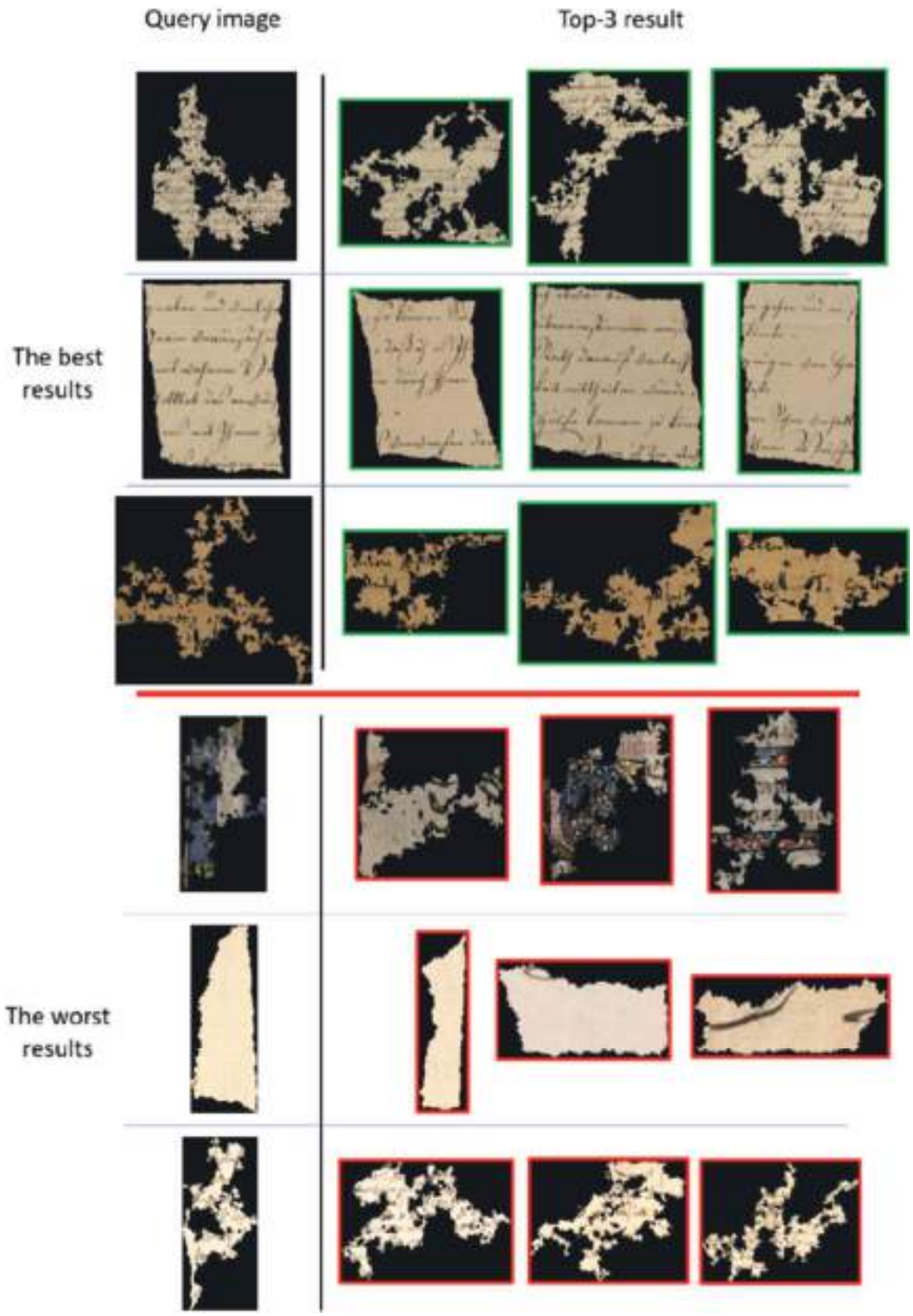


図５－３．HisFragIR20 テストセットにおける最良結果と最悪結果.

（６）手書き文字画像からの筆記過程の復元［11］

オフライン手書きの日本語漢字画像から動的なオンライン手書き軌跡を復元するための深層ニューラルネットワークを提案した．日本語の漢字は複数のストロークで構成されているため，これは難しい作業である．提案するモデルは，畳み込みニューラルネットワークのエンコーダ，注意層を備えた長短期記憶ネットワークのデコーダ，および，ガウス混合モデル（GMM）の3つの主要コンポーネントからなる．エンコーダは特徴抽出に焦点を合わせ，デコーダは抽出された特徴を参照し，GMMパラメータの時系列を生成する．注意層は軌跡回復の重要な要素である．GMMは提案されたモデルが学習したサンプルに過剰適合しないように，スタイルのバリエーションに堅牢性を提供する．提案手法を視覚的検証と手書き文字認識の両方によって評価した．オフラインの手書き認識精度を向上させるために復元されたオンラインの軌跡を使用する最初の試みである．視覚的検証によりいくつかの問題が明らかになったが，認識実験では復元された軌跡のオンライン認識を組み合わせたことで，手書き文字認識の精度が向上することを確認できた．

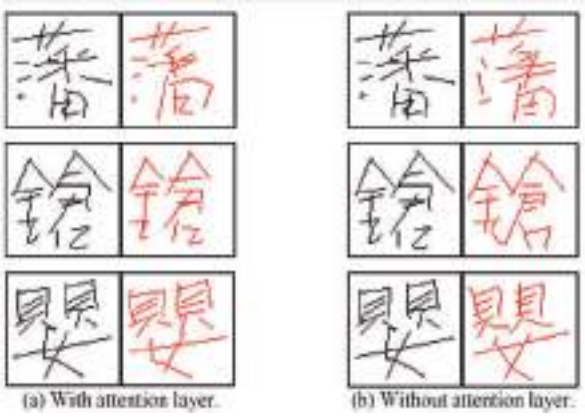


図６－１．注意層のあるなしに関わらず正しく復元できた例．

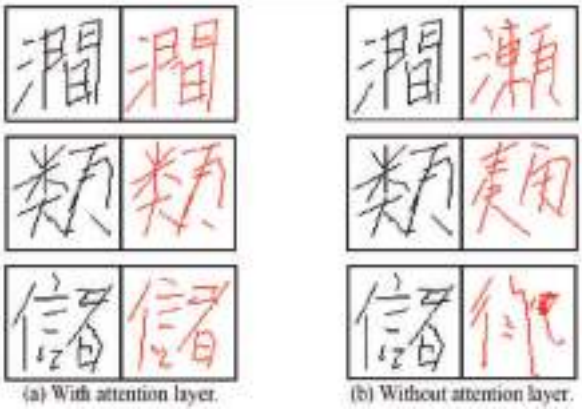


図６－２．注意層を用いることで成功した例．

（7）同一木簡からの木簡片の特定［9, 16］

本研究では，Siamese ニューラルネットワークを用いて，様々なサイズの木簡破片をマッチングする方式を提案した．この Siamese ネットワークは，大域平均プーリングと最適化された Resnet エンコーダ（GA-S-net）で構成される．また，大域平均プーリングを空間 Pyramid プーリング層に置き換えて，新しい Dense Absolute Different 層（SP-S-net）を追加し，より精度が高いネットワークも検討した．これらのネットワークで2つの破片が同じ木簡からかどうかを判断する実験のために，日本の奈良時代に使用された平城京遺跡から発掘された268個の完全な木簡から37,760破片のサンプルを生成した．テストデータセットでの両方のネットワークの結果は，ROC（Receiver Operating Characteristic）曲線の AUC（Area Under the Curve）で約90％とほぼ等しい．ただし，大きな破片に対する AUC では SP-S-net の方が97.1％と，GA-S-net の93.8％より優れている．

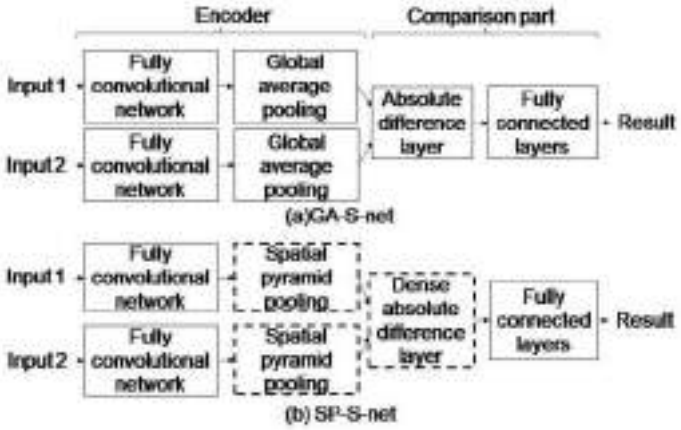


図７－１．GA-S-net (a) と SP-S-net (b).

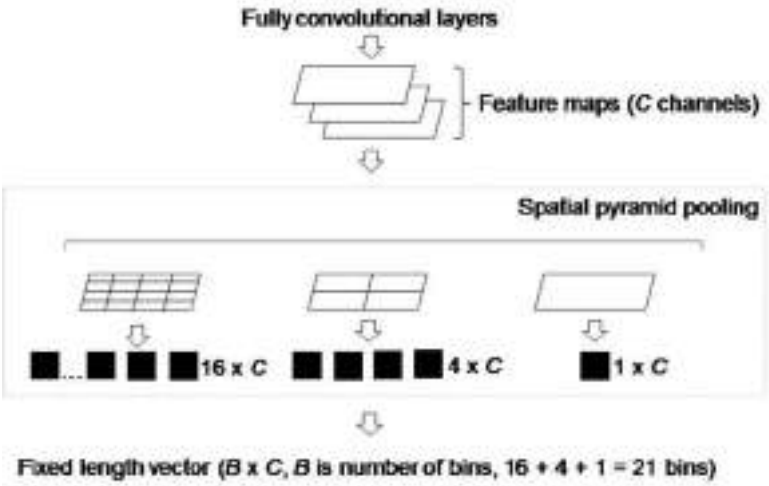


図７－２．空間 Pyramid プーリング層.

引き続いて，注意層を用いたニューラルネットワークA-VLADによって， 2つの様々なサイズ木簡破片が同じタブレットからのものであるかどうかを判定する方式を提案した．本手法は，二値化またはセグメンテーションなどの前処理段階を必要としない．畳み込みニューラルネットワーク（CNN）を使用した特徴抽出器，キーポイントを選択するための注意層，そして，抽出されたキーポイントを集約することによって代表ベクトルを形成する一般化された深部神経VLADモデルの３つの部分から構成される．ネットワーク全体は，交差エントロピー損失関数とTriplet 損失関数の組合せによってエンドツーエンドで学習される．実験のために，日本の奈良時代に使用された平城京遺跡から発掘された556個の完全な木簡を分解し，13,205破片のサンプルを生成した．A-VLADモデルは，木簡データセットの最先端の方法よりも75.5%のMAPと87.9%の第一位候補正解率を達成した．考古学者が木簡の断片を元の木簡に組み立てるのをサポートできる可能性がある．

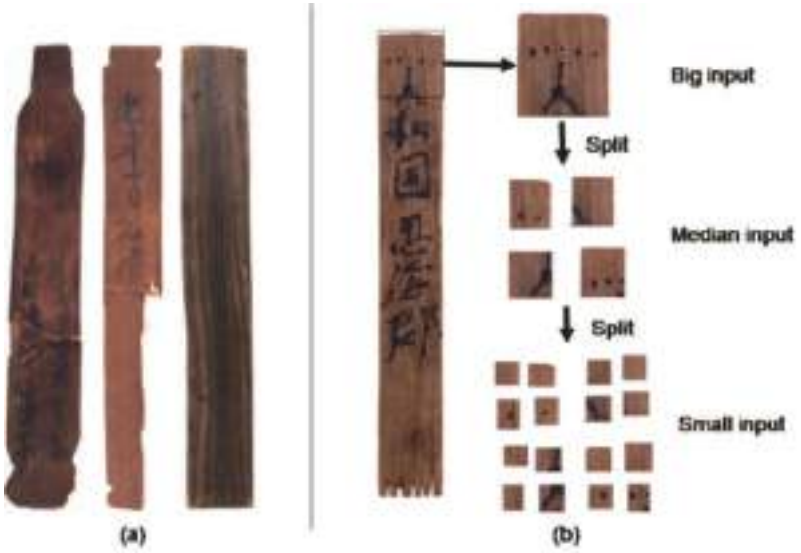


図７－３．元の木簡（a）と準備したサンプル（b）.

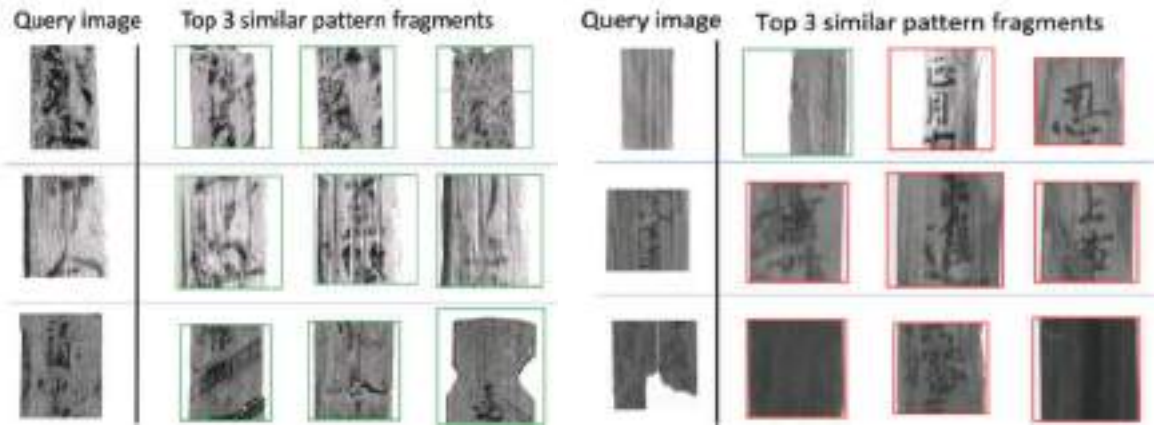
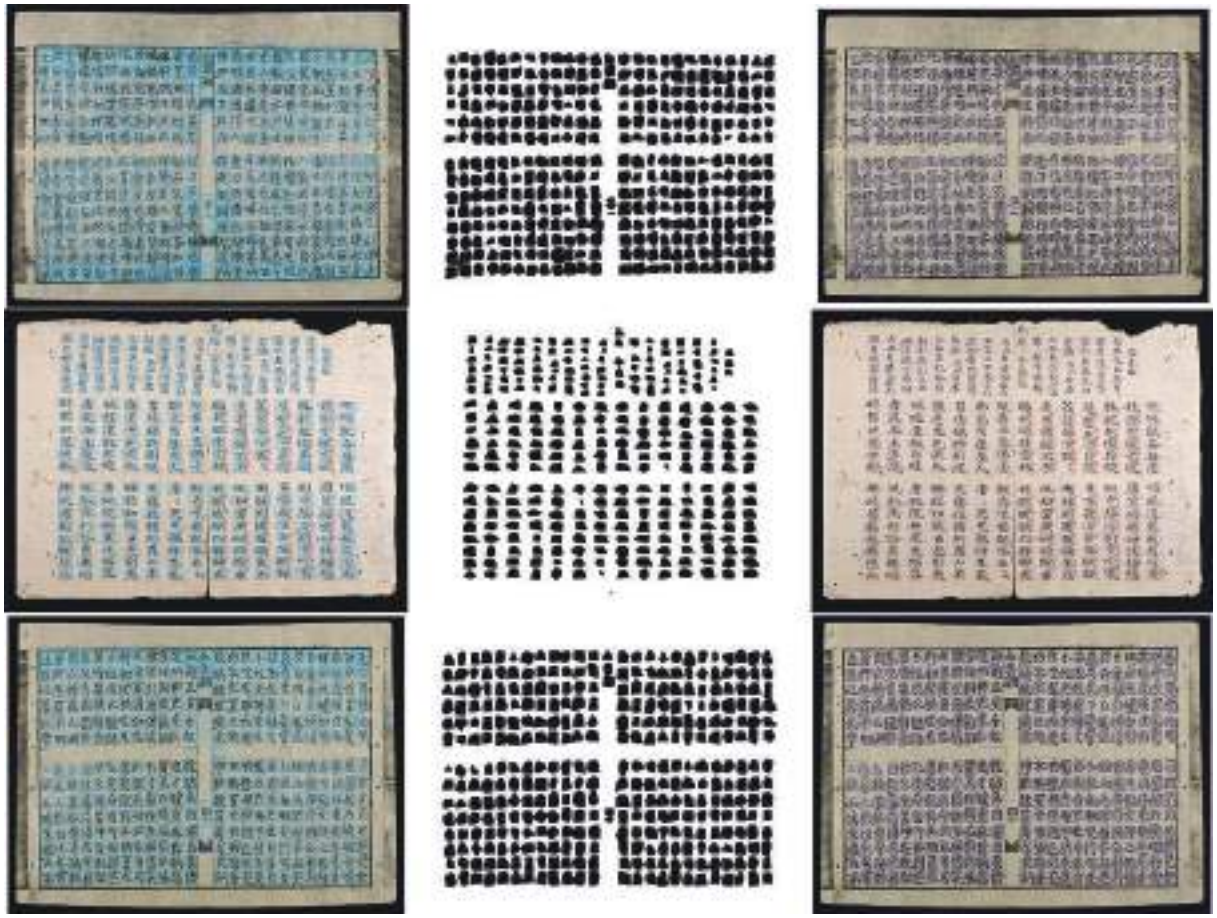


図７－４．テストセットに対する成功例と失敗例.

（8）ベトナム古文字文書の認識 [3]

ノムは，現代のアルファベットによる文字が使われるまで，千年にわたり使われたベトナムの古文字であり，それで記述された多くの書籍が残っている．しかし，書籍の経年的劣化とそれを読める人の減少により，ノム書籍の電子化が喫緊の課題である．本論文は，ノム文書を電子化するための畳み込みニューラルネットワークによるセグメンテーションと文字認識の方法を提示している．ノムのページを前処理し，文字に分割し，文字認識する．U-Netを用いてセグメンテーションマップを作製し，そこから文字を抽出する．そして，大分類と詳細分類を用いた文字認識を行う．そして，CTCによる言語モデルを用いて文章としてもっともらしい文字列を確定する．射影とボロノイ図によるセグメンテーションが81.23%のIoUであったことに対いて，本方式は　92.08%のIoUを達成し，文字認識においても，修正二次識別関数と学習ベクトル量子化を用いた従来手法の認識率を飛躍的に更新して，85.07%の認識率を達成した．大分類と詳細分類による二段階の認識，ごま塩ノイズを含めた学習パターンによる学習，アテンション層の採用が認識率の向上に貢献した．

図８－１．ノム文書の例.



射影とボロノイ図による分割　　本方式による分割　　文字領域の抽出

図８－２．ノム文書のセグメンテーション．

（９）人工パターンによるオフライン手書き日本語認識モデルの学習［5］

本研究では，オフライン手書き日本語テキスト認識のため，深層畳込み再帰ネットワークモデル（DCRN）を提案した．DCRNモデルでは，1）CNNによって入力画像から特徴の列を抽出する畳み込み特徴抽出器；２）BLSTMによって特徴の列から候補文字と確率の組の列を予測する再帰層；３）CTCによって候補文字と確率の組の列を文字列に変換する転写層，の3つの部分で構成される．本モデルの学習には大量の学習データが必要であるため，コーパスの文章と文字データセットから合成データ生成方法を提案した．実験では，DCRNモデルと合成データ生成方法の有効性を評価した．実験の結果は，DCRNモデルがTUAT Kondateデータセットの最先端の認識精度より高い精度を達成し，生成したデータなしで96.35%に対して，生成したデータありで98.05%の文字レベルの認識精度を達成した．

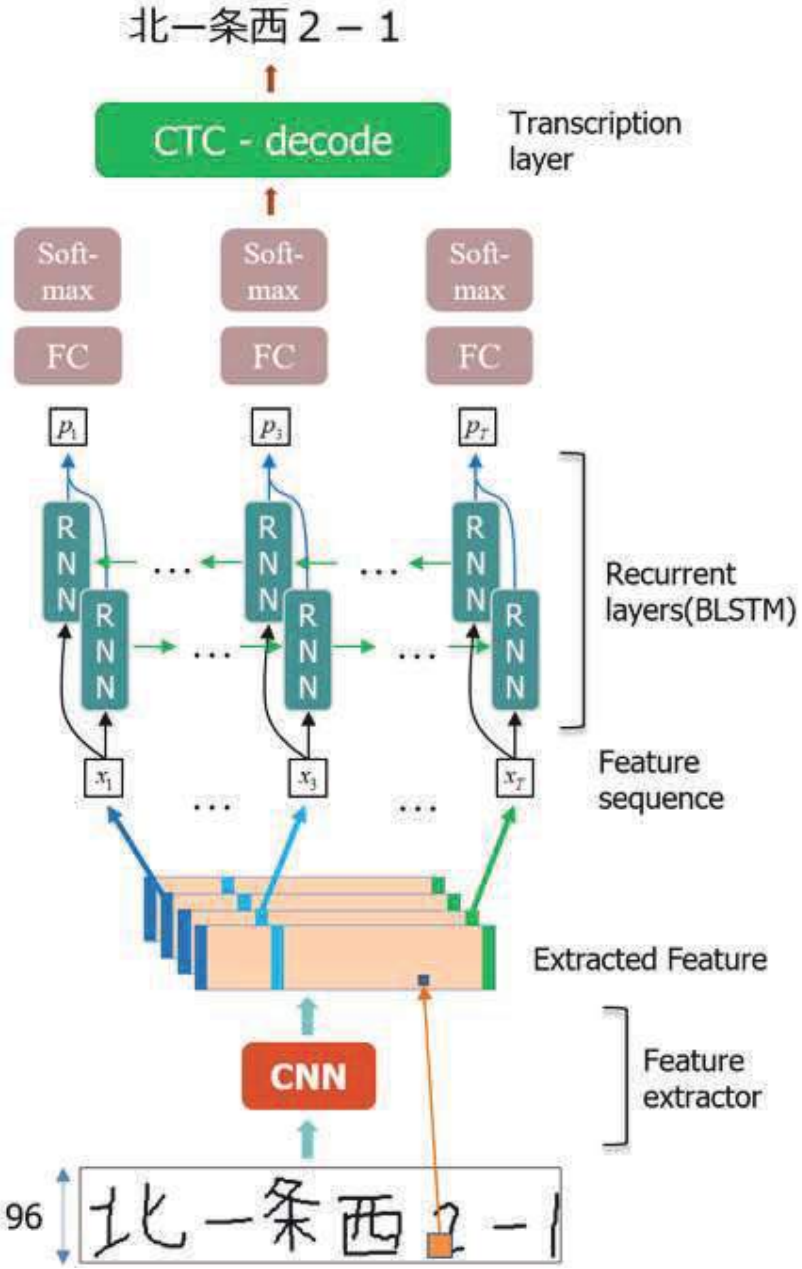
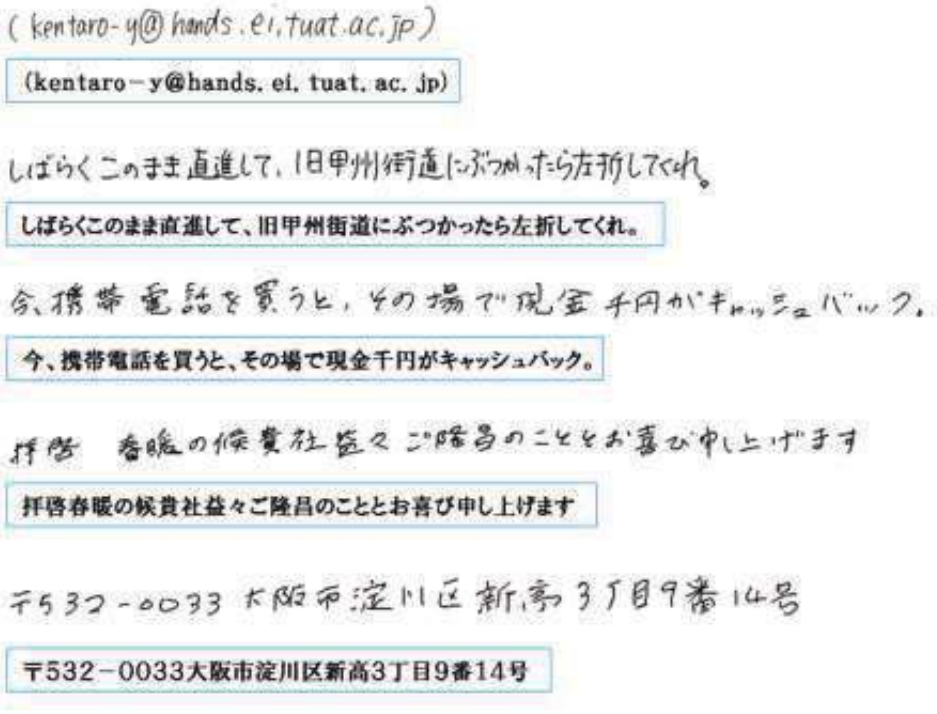
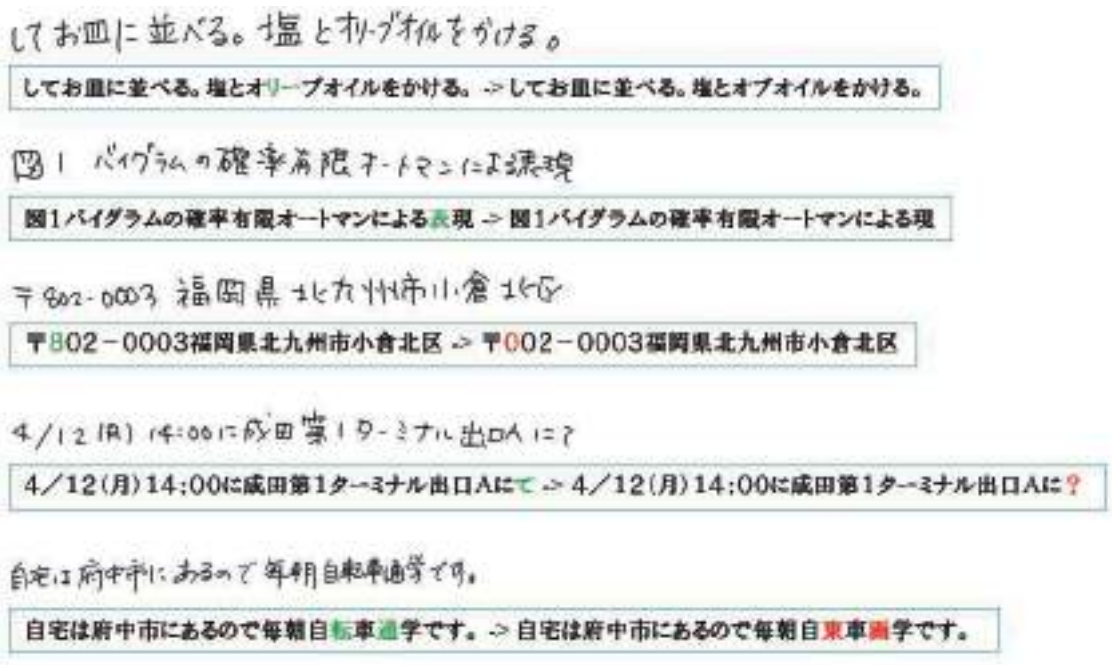


図９－１．DCRN モデルのネットワーク構造.



(a) 正しく認識された例.



(b) 誤認識例（赤字が誤認識箇所）.

図９－２．正認識例と誤認識例.

（１０）意味的セグメンテーションによる手書き日本語文章認識［10］

近年，畳み込みニューラルネットワークと再帰ニューラルネットワークを利用した事前のセグメンテーションをしない手書き文字列認識が注目を集めている．あるデータには良い結果を示しているが，課題も残っている．余程たくさんの学習パターンを与えない限り，文字間隔の違いや文字線の幅，字形などの手書きスタイルの違いや，文字の傾き，行のうねりなどに頑健でない．本論文では，従来の課題を克服したセグメンテーションに基づく日本語手書き文字列認識を提案する．本手法は，文字列を高い精度で文字に分割できる意味的セグメンテーションのモデルを採用している．意味的セグメンテーションモデルは，U-Net のエンコーダ・デコーダモデルに従い，その他の手法も併用している．U-Net より深い ResNet101 を使い，Atrous 畳み込みと空間ピラミッドプーリングを採用し，引き続いて CNN で文字画像を認識する．そして，結果の文字候補列を言語文脈で最適な文字列に確定する．提案する方法は，セグメンテーションの IoU を Atrous 畳み込みと空間ピラミッドプーリングをしない場合の 89.32%から 94.96%に高めた．セグメンテーションなしの手法が手書きスタイルの違いに極めて鋭敏であるのに対して本手法は頑健である．下図は，文字間隔や線幅，字形の変化に頑健であることを示している．

図１０－１．本手法によるセグメンテーション例．

（１１）2次元自己アテンションＣＮＮによる手書き認識［13］

オフライン手書き文字列認識は，様々な背景，ノイズ，多様な筆記スタイル，文字間の接触などの問題が今も挑戦的である．本研究では，手書きテキスト行を認識するための，2D 自己注意による畳込み再帰ネットワーク（2D-SACRN）のモデルを提案した． 2D-SACRN モデルは，1）入力画像から特徴の列を抽出する 2D 自己注意による畳み込み特徴抽出器，2）特徴の列から候補文字と確率の組の列に変換するエンコーダ，3）候補文字と確率の組の列から文字列に変換する CTC デコーダ，の 3 つの部分で構成される．このモデルでは，入力画像での遠い依存関係を取得するため，特徴抽出器に 2D 自己注意メカニズムを導入した．実験では，IAM（英語），Rimes（フランス語），および，TUAT_Kondate（日本語）の 3 つのデータセットで 2D-SACRN モデルを評価した．実験結果は，提案された 2D-SACRN モデルがすべてのデータセットの最先端のモデルより精度が高いことを示した．

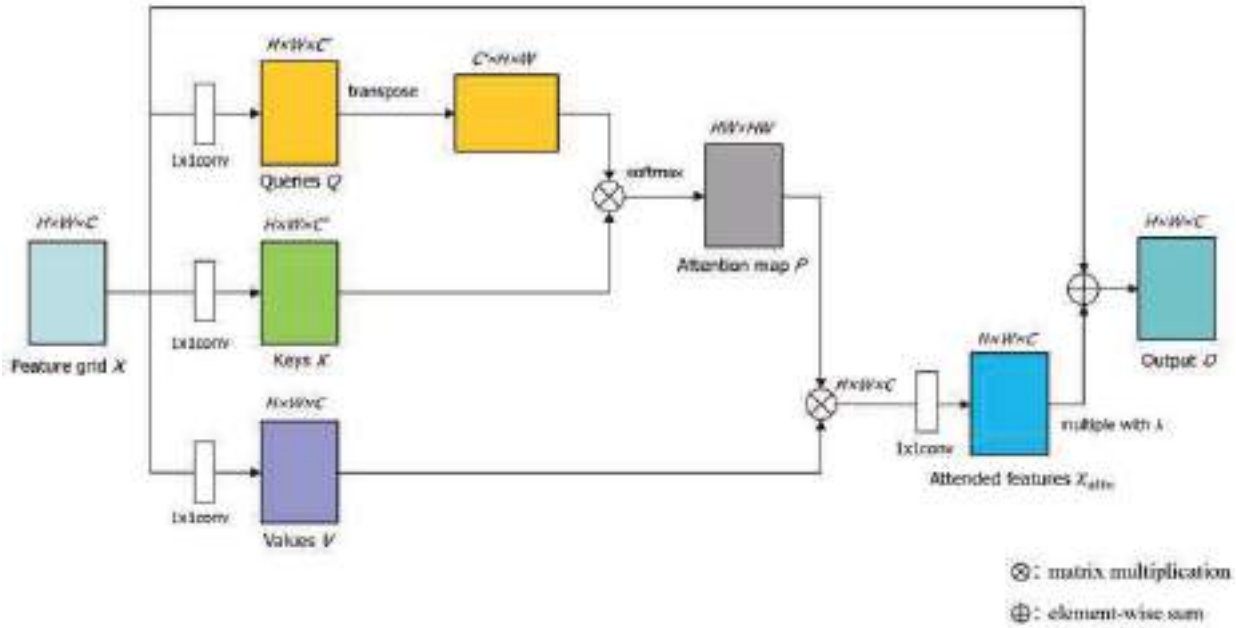


図１１－１．２次元自己アテンションＣＮＮの構造.

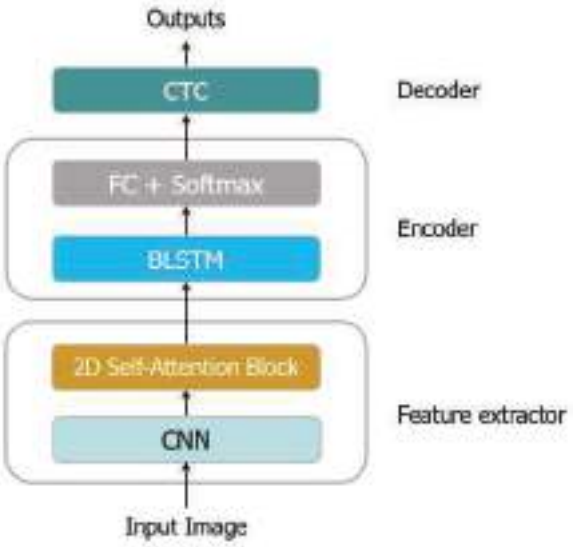


図１１－２．2D-SACRN のネットワーク構造.

（１２）再帰ニューラルネットワークによるオンライン手書き日本語認識［6］

本研究では，再帰ニューラルネットワーク（RNN）によって，オンライン手書き日本語の文字とテキスト行を認識を試みた．RNNを使用したオンライン中国語手書き認識に関するいくつかの成功した研究があるが，漢字や仮名などの様々な文字種のために，日本語の高精度を達成することは困難である．さらに，RNNの学習には文字クラスごとに多数のサンプルが必要である．したがって，元のサンプルに5つの異なる変換操作を適用して，いろいろな変形を伴う人工パターンを生成する．オンライン手書き日本語テキストには，Kondateデータセットを使用する，日本語の文字セット全体を網羅しているわけではない．そこで，コーパスの文とNakayosiおよびKuchibueデータセットの孤立した文字パターンを使用して，テキスト行を生成する．さらに，手書き文字パターンは異なる解像度の様々なデバイスで収集されたものを利用しているため，いくつかの前処理プロセスを適用し，認識に有用な特徴を抽出する．認識モデルでは，単独文字認識と手書きテキスト認識に，さまざまな双方向長短期記憶ネットワーク（BLSTMネットワーク）を実装した．単独文字認識のための最良モデルは，Nakayosiで97.91％，Kuchibueで97.74％の精度を達成した．手書きテキスト認識の最良モデルは，Kondateで86.31％，生成されたテキスト行で83.15％の文字認識率を達成した．

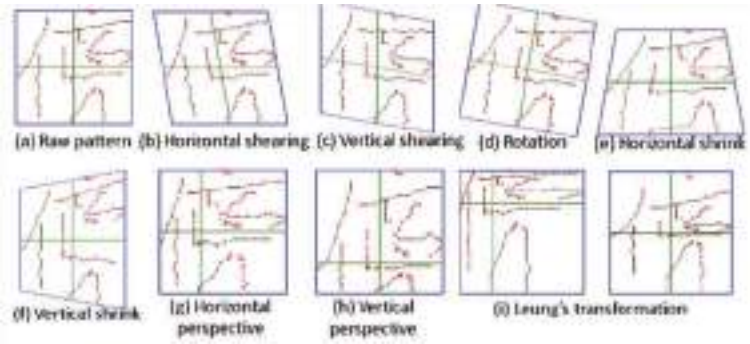


図１２－１．変換操作５種類とそれらを組わせた変換.

図１２－２．生成された手書きテキスト例.

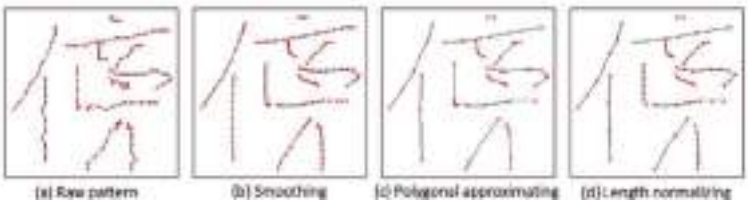


図１２－３．前処理の結果.

（１３）再帰ニューラルネットワークによる日本語・中国語手書き認識［14］

本研究では，日本語と中国語のオフライン手書きテキスト行画像を認識するための再帰型ニューラルネットワーク変換器（RNN-Transducer）モデルを提案した．オフライン手書きテキスト認識にRNN-Transducerモデルを採用した最初のアプローチである．本モデルは3つの主要な部分からなる．視覚的特徴エンコーダは，CNNによって入力画像から視覚的特徴を抽出し，BLSTMによって視覚的特徴をエンコードする．言語文脈エンコーダは，埋め込みレイヤーとLSTMによって入力画像からの言語特徴をエンコードする．ジョイントデコーダは完全に接続された層とソフトマックス層によって，視覚的特徴と言語的特徴を組み合わせて最終的なラベルシーケンスにデコードする．提案されたモデルは，入力画像からの視覚的情報と言語的情報の両方を利用する．実験では，KuzushijiデータセットとSCUT-EPTデータセットでモデルを評価した.実験結果は提案モデルが最先端のパフォーマンスを達成することを示す.

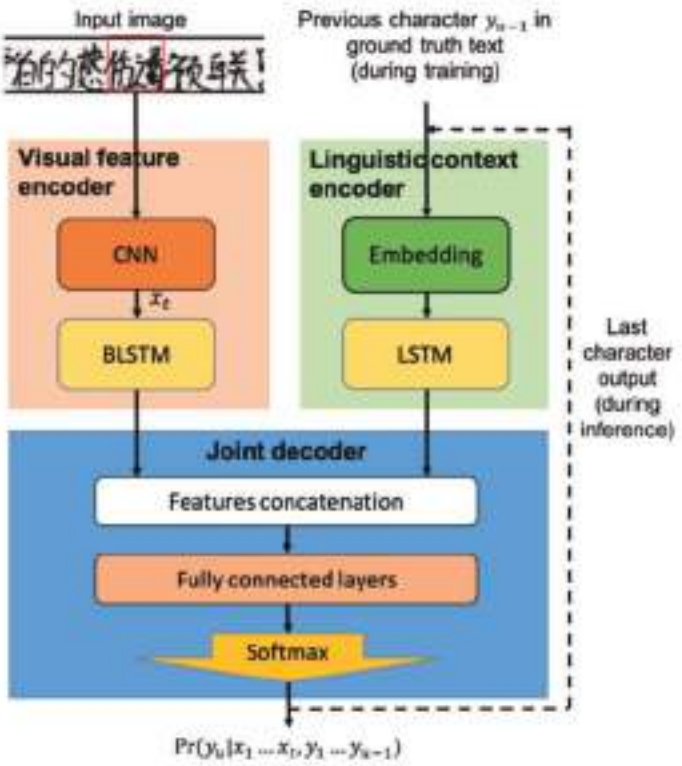


図１３－１．RNN-Transducerのアーキテクチャ.

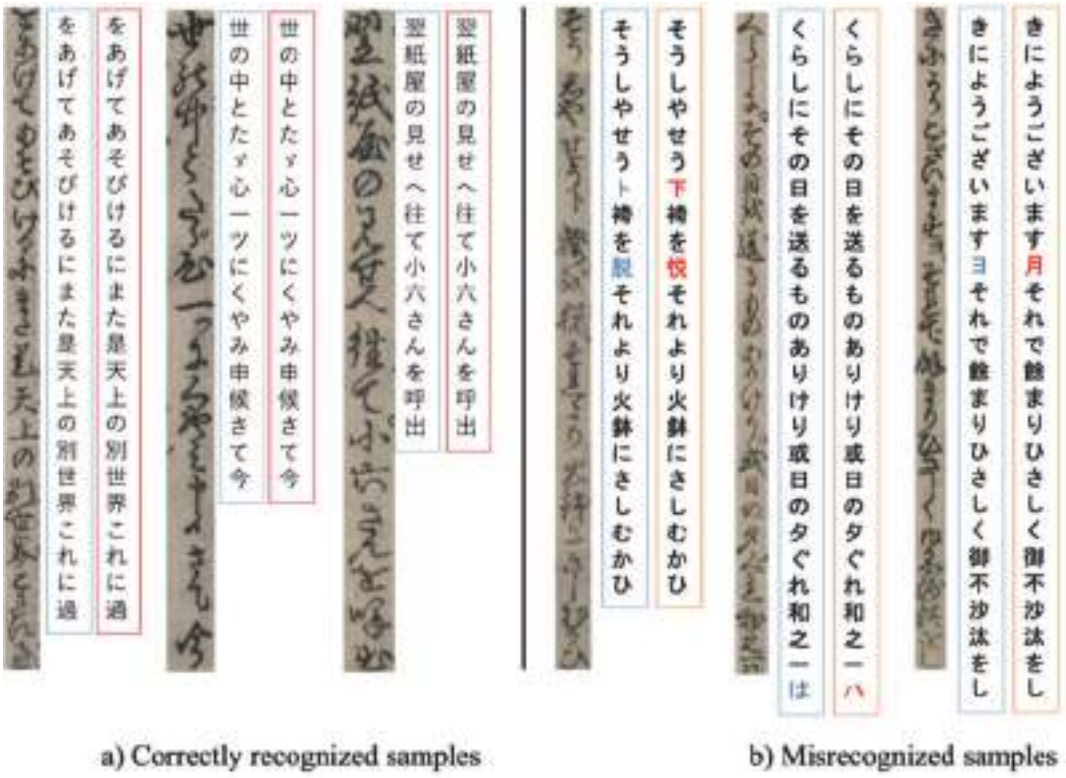


図１３－２．Kuzushijiデータセットにおける正認識例と誤認識例（赤字が誤認識）.

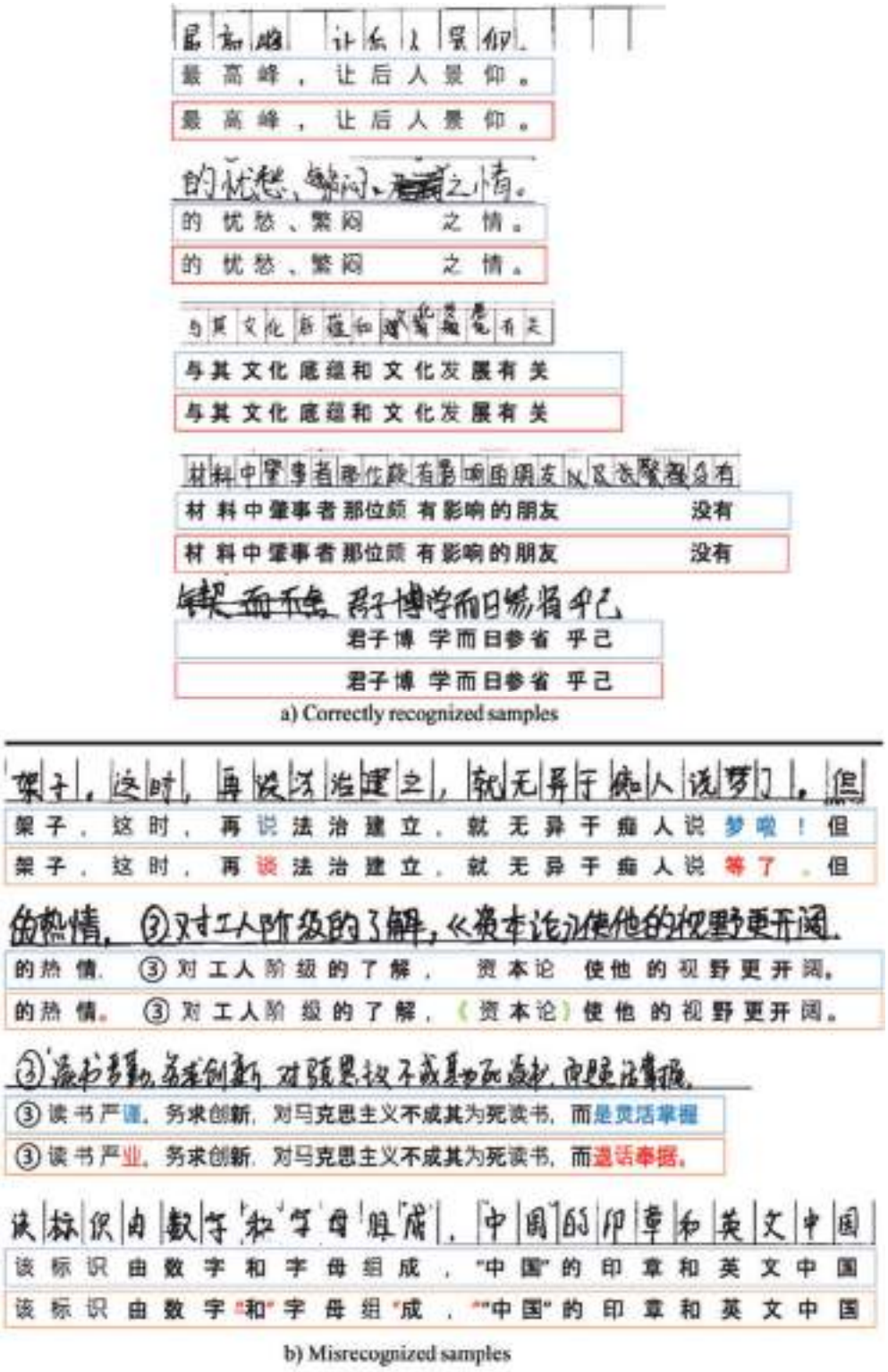


図１３－３．SCUT-EPT データセットに対する正認識例と誤認識例（赤字が誤認識）．

発表文献

（1） 査読付き論文誌論文

[1] Réjean Plamondon, Giuseppe Pirlo, Éric Anquetil, Céline Rémi, Hans-Leo Teulings, Masaki Nakagawa: "Personal Digital Bodyguards for e-Security, e-Learning and e-Health: A Prospective Survey," Pattern Recognition, Vol. 81, pp. 633-659 (2018.9).

[2] Nam Tuan Ly, Kha Cong Nguyen, Cuong Tuan Nguyen and Masaki Nakagawa: "Recognition of Anomalously Deformed Kana Sequences in Japanese Historical Documents," IEICE Trans. on Inf. & Syst., Vol.E102-D, No.8, pp.1554-1564 (2019.8).

[3] Kha Cong Nguyen, Cuong Tuan Nguyen and Masaki Nakagawa: Nom document digitalization by deep convolution neural networks, Pattern Recognition Letters, Vol. 133, pp.8-16 (2020.5).

[4] Nam Tuan Ly, Cuong Tuan Nguyen and Masaki Nakagawa: An attention-based row-column encoder-decoder model for text recognition in Japanese historical documents, Pattern Recognition Letters, Vol. 136, pp.134-141 (2020.8).

（2）査読のある国際会議

[5] Nam Tuan Ly, Cuong Tuan Nguyen and Masaki Nakagawa: "Training an End-to-End Model for Offline Handwritten Japanese Text Recognition by Generated Synthetic Patterns," International Conference on Frontiers in Handwriting Recognition, Buffalo, USA, pp.74-79 (2018.8).

[6] Hung Tuan Nguyen, Cuong Tuan Nguyen and Masaki Nakagawa: "Online Japanese Handwriting Recognizers using Recurrent Neural Networks," International Conference on Frontiers in Handwriting Recognition, Buffalo, USA, pp.345-440 (2018.8).

[7] Kha Cong Nguyen, Seiji Hotta and Masaki Nakagawa: "A Character Attention Generative Adversarial Network for Degraded Historical Document Restoration," Proc. of ICDAR 2019, Sydney, Australia, pp.420-425 (2019.9).

[8] Nam Tuan Ly, Cuong Tuan Nguyen and Masaki Nakagawa: "An attention-based end-to-end model for multiple text lines recognition in Japanese Historical Documents," Proc. of ICDAR 2019, Sydney, Australia, pp.629-634 (2019.9).

[9] Trung Tan Ngo, Cuong Tuan Nguyen and Masaki Nakagawa: A Siamese Network based

approach for matching various sizes of excavated wooden fragments, Proc. International Conference on Frontiers in Handwriting Recognition, (ICFHR), Dortmund, Germany, pp.307-312, (2020.9).

[10] Kha Cong Nguyen, Cuong Tuan Nguyen and Masaki Nakagawa: A Semantic Segmentation-based Method for Handwritten Japanese Text Recognition, Proc. International Conference on Frontiers in Handwriting Recognition, (ICFHR), Dortmund, Germany, pp.127-132, (2020.9).

[11] Hung Tuan Nguyen, Tsubasa Nakamura, Cuong Tuan Nguyen and Masaki Nakagawa: Online trajectory recovery from offline handwritten Japanese kanji characters, Proc. International Conference on Pattern Recognition, (ICPR), Milan, Italia, pp. 8320-8327 (2021.1).

[12] Trung Tan Ngo, Hung Tuan Nguyen and Masaki Nakagawa: A-VLAD: An End-to-End Attention-based Neural Network for Writer Identification in Historical Documents, Proc. 16th International Conference on Document Analysis and Recognition, Lausanne, Switzerland, pp.396-409 (2021.9).

[13] Nam Tuan Ly, Hung Tuan Nguyen and Masaki Nakagawa: 2D Self-Attention Convolutional Recurrent Network for Offline Handwritten Text Recognition, Proc. 16th International Conference on Document Analysis and Recognition, Lausanne, Switzerland, pp.191-204 (2021.9).

[14] Trung Tan Ngo, Hung Tuan Nguyen, Nam Tuan Ly and Masaki Nakagawa: Recurrent neural network transducer for Japanese and Chinese offline handwritten text recognition, Proc. 1st Workshop on Document Images and Language, Lausanne, Switzerland, pp.364-376 (2021.9).

（3）査読のないシンポジウム・研究会論文

[15] Hung Tuan Nguyen, Cuong Tuan Nguyen, Masaki Nakagawa, Asanobu Kitamoto：Text Segmentation for Japanese Historical Documents using Fully Convolutional Neural Network, じんもんこん 2019 論文集, Vol. 2019, pp. 253-260 (2019.12).

[16] Trung Tan Ngo, Hung Tuan Nguyen, Masaki Nakagawa: Neural Network-based Local Feature Descriptors for Matching Excavated Mokkan Fragments of Various Sizes, to appear in PRMU (2021.12).