一般社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report
LOIS2020-8,IE2020-20,EMM2020-32(2020-09)

CTPN を使用した植物標本画像のラベル自動マスキング方法の検討

張　　徳鵬†　檜垣　泰彦††,†††　須貝　康雄†††

† 千葉大学大学院融合理工学府　〒263–8522 千葉市稲毛区弥生町 1–33
†† 千葉大学アカデミック・リンク・センター　〒263–8522 千葉市稲毛区弥生町 1–33
††† 千葉大学大学院工学研究院　〒263–8522 千葉市稲毛区弥生町 1–33
E-mail: ††higaki.yasuhiko@faculty.chiba-u.jp

あらまし　萩庭植物標本画像データベースの c-arc への登録に必要な画像処理を検討した．萩庭植物標本は珍重な植物標本を含んでいるため，採集地などの重要な情報を公開できず、マスキングする必要がある．これまでは，文字認識を活用したラベルの位置特定法で実験したが十分な精度が得られていない．これを解決するため，ディープラーニングに基づいた画像自動処理法を検討した．本研究ではテキスト検出ネットワーク CTPN を使って，ラベル上のテキストを検出してラベルの位置を試み，この方法が有効であるという結果を得た．
キーワード　植物標本画像，ディープラーニング，自動マスキング，テキスト検出，CTPN

Study on automatic label masking method for plant specimen images using CTPN

Depeng ZHANG†, Yasuhiko HIGAKI††,†††, and Yasuo SUGAI†††

† Graduate School of Science and Engineering, Chiba University　1–33 Yayoi-cho, Inage-ku, Chiba-shi, 263–8522 Japan
†† Academic Link Center, Chiba University　1–33 Yayoi-cho, Inage-ku, Chiba-shi, 263–8522 Japan
††† Graduate School of Engineering, Chiba University　1–33 Yayoi-cho, Inage-ku, Chiba-shi, 263–8522 Japan
E-mail: ††higaki.yasuhiko@faculty.chiba-u.jp

Abstract　In this study, the image processing required for registration of the Haginiwa plant specimen image database in c-arc. Since Haginiwa plant specimens include rare plant specimens, important information such as the collection site cannot be disclosed and must be masked. Up to now, we have been experimenting with a label position identification method that utilizes character recognition, but sufficient accuracy has not been obtained. In order to solve this problem, we consider an automatic image processing method based on deep learning. In this research, we use the text detection network CTPN, the text on the label was detected and the position of the label was tried, and this method obtained a valid result.
Key words　Plant specimen image, Deep learning, Automatic masking, Text detection, CTPN

1. はじめに

「萩庭植物標本画像データベース」は，千葉大学名誉教授であった萩庭丈壽（HAGINIWA Joju， 1917-1996）が生涯にわたり採集・収集したさく葉標本のデータベースである [1]．北海道，琉球，小笠原の各諸島を含む日本全国から台湾，タイなど海外の植物も採集している．日本全土の顕花植物の約 95% を含むと言われ，中にはすでに絶滅した植物・絶滅危惧種の植物が 1000 種余も含まれており（環境省レッドリスト 2018 による），質・量ともに日本の自生顕花植物のさく葉標本として比類がないものである．

5 万点を超える膨大な植物標本は，標本の整理，デジタル写真撮影，標本ラベルのデータ入力，植物名・採取地名の補完などの作業が行われ，2002 年の千葉大学薬学部ウェブサイトでの試験公開を経て，2008 年にデータベースが完成した．標本の現物は保存と研究のために，2005 年に国立科学博物館へ移管された．

c-arc（千葉大学学術リソースコレクション）[2] とは Chiba

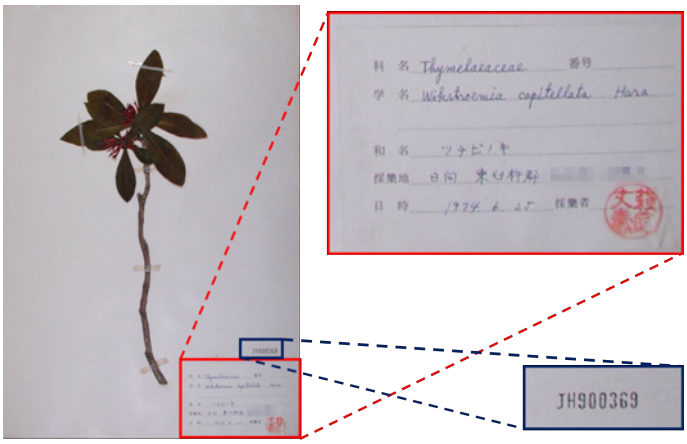

図 1　標本画像と標本ラベルの拡大図

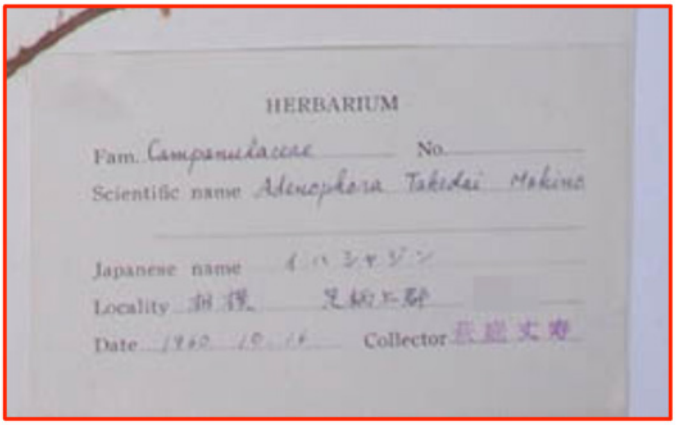

図 2　英語ラベルの例

University Academic Resource Collections の略で，千葉大学附属図書館がウェブ上で公開・提供するコンテンツを学術リソースとして広く使ってもらうためにコレクションとしてまとめたものである． 画像データについては，International Image Interoperability Framework (IIIF) 技術を用いて，学術リソースとして簡単に利活用できる環境を提供している． c-arc は，デジタルコンテンツを活用した研究・教育・学習を実現するための新しい教育研究基盤である「デジタル・スカラシップ」構築の一環として，アカデミック・リンク・センターが制作した．

このような背景もあり，萩庭植物標本画像についても，前紹介した c-arc での公開，つまり広く相互利用可能な形での公開が予定されている．しかし，萩庭植物標本画像データベースには，いま判明しているだけでも 8,493 件（うち絶滅 5 件）と決して少なくない数の絶滅危惧種が含まれており，さらに標本画像によっては非常に細かい粒度の採取地が記載されている場合もある．これをそのまま広く容易に利用できる形で公開してしまうのは，資源保護の観点から望ましいとは言い難い面がある．対策として採取地の詳細をマスキングすることが考えられるが，画像点数は 5 万点を超えており，マスキングを人手で行うには相当の時間を要することが予想される．

本研究では c-arc での公開に向け，標本画像に含まれる採取地表記部分を検出し，マスキングをかけた画像を自動的に作成することを目的とする．

2.　植物標本画像

本研究では図 1 のように標本画像中にある標本情報（科名・学名・和名・採集地・採集日など）が記載されている領域を「標本ラベル」と称する．図 1 では詳細な採集地に相当する箇所にはモザイク加工を施している．青い枠の JH で始まる記号は「JH 番号」であり標本のユニークキーである．また，標本ラベルの見出し部分が図 1 のように「日本語であるものを「日本語ラベル」，英語であるものを「英語ラベル」と称する．英語ラベルの一例を図 2 に示す．ラベルの各項目の内容は手書きである．

表 1　OCR でマスキングした画像の結果（先行研究）

total	masked	ng	proportion
55,845	14,039	41,806	25%

3.　先 行 研 究

文献 [3] では萩庭植物標本画像に対し以下の手法およびパラメータの組み合わせでマスキング処理を行った．

- OCR の入力に用いる画像

無加工の標本画像と，OpenCV の読み込みオプションでグレースケールを指定し，保存した画像（グレースケール画像），適用的ヒストグラム平坦化を施した画像の三種類を用いた．

- OCR エンジン

OCR エンジンは Google 提供の Cloud Vision API を利用した．バージョン指定は v1p3beta で，それに対応したクライアントライブラリを使用した．

- OCR のパラメータ

パラメータ Language Hint が日本語ラベルの場合は日本について語を意味する “jp” を，英語ラベルの場合，ラテン語系手書き文字を意味する “mul-Laten-t-i0-handwrit” を指定した．

その際の処理の流れは図 3 の通りである．この方法では，一つの標本に対して複数のマスキング済み画像が生成されるが，座標ずれなどに関して，マスキングの段階で検証する術に乏しいため，最終的な取捨選択の判断については，現時点では目視ベースで行うものとした．

この方法で実験した画像を評価した結果を表 1 に示す．

先行研究では標本画像に記載された採集地を検出し，マスキング範囲の決定・実行までの一連の操作を自動的に行う方法について検討を行った．複数手法を試したところ，採集地の直接的検出は難しいものの，光学的文字認識を活用することで，間接的に位置を推定するアプローチが有効との結果が得られた．表 1 の結果を見ると，合計 55,845 点の画像の中に、マスキングした数は 14,039 点で、そのうちの 25% は公開できない程度だと考えられる．また，Cloud Vision API 自体に手を入れる

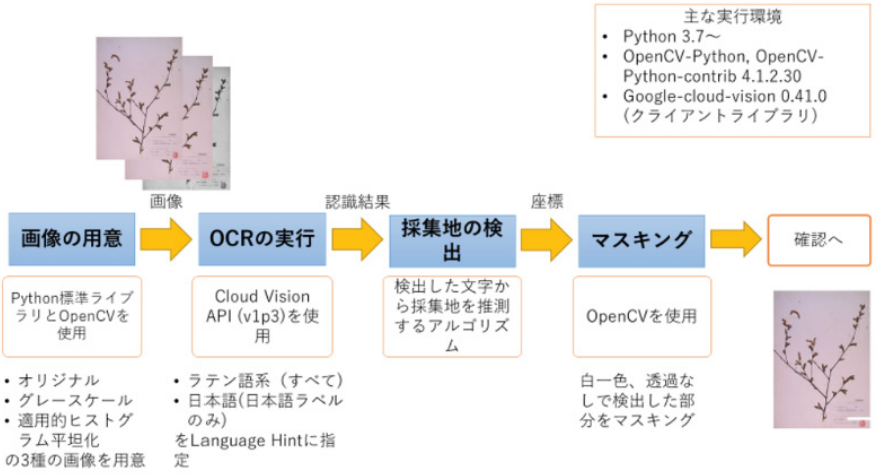

図 3　先行研究の流れとパラメータ [3]

ことはできないため，入力画像やパラメータで調整することしかできず，これ以上マスキング率を向上するのが難しいと考えられる．

4.　CTPN ネットワーク

文字認識を活用したラベルの位置特定法では十分な精度が得られていない．これを解決するため，ディープラーニングに基づいた画像自動処理法を検討する．本件に適したネットワークが 3 つある．

ターゲット検出ネットワーク Faster R-CNN [4] は画像としてテキスト検出することで，各テキストが画像に現れる場所を見つける．テキストありかテキストなしの 2 種類しか判別できないが．テキストのみという単純な単一カテゴリのターゲット検出タスクなので，Faster R-CNN の利用が有効であると考えられる．

2016 年に，テキストと他のターゲットを区別できるよう，これを改善した CTPN というアルゴリズム [5] が提案された．今まで，このネットワークフレームワークは，OCR システムのテキスト検出に一般的に使用されていたネットワークであり，後続のテキスト検出アルゴリズムの方向に大きな影響を与えている．

CTPN にも明らかな欠点があり，非水平テキストの検出効果がよくない．論文に記載されているテキスト検出結果はすべて水平方向のものだが，多方向のテキスト検出については詳しく述べられていない．2017 年に，任意の角度でテキストを検出できるよう改善した SegLink というアルゴリズム [6] が考案された．萩庭植物標本のラベルテキストは全部水平なので，テキスト検出に CTPN を適用できる．しかし，CPTN の論文では手書き文字における効果は検証されていない．手書き文字の検出は期待できないと考えられる．

図 4 に CTPN のアーキテクチャを示す．VGG16 モデル [7] の最後の畳み込みマップ (conv5) を介して 3 × 3 の空間窓を密にスライドさせる．各行の連続したウィンドウは，双方向

表 2　テキスト検出した画像の評価基準

standard1	テキストを全部検出できた
standard2	一部テキストを検出できた
standard3	テキスト検出されてない画像やテキストではない部分が検出された画像
NG	テキストを全然検出できなかった

LSTM (BLSTM) [8] によって再帰的に接続され，各ウィンドウの畳み込み特徴（3 × 3 × C）が 256D BLSTM (2 つの 128D LSTM を含む) の入力として使用される．RNN 層は 512D の完全連結層に接続され，続いて出力層が，テキスト/非テキストスコア，y 軸座標，k 個のアンカーのサイドレフィンメントオフセットを共同で予測する．図 5 に示すように CTPN は，固定幅の細かいテキスト提案を順次出力する．各ボックスの色は，テキスト/非テキストスコアを示す．スコアが正のボックスのみを表示している．

5.　テキスト検出実験

実験環境は以下の通りである．

- OS ： Ubuntu 16.04 LTS
- GPU ： GeForce GTX 1080 Ti
- Python：3.6
- Tensorflow：1.3

5 万枚の計算に約 12 時間かかった．テキスト検出した結果の画像を以下に示す．

（1） 全部検出された画像 (図 6)

（2） 一部テキスト検出された画像 (図 7)

（3） テキスト検出されてない画像やテキストではない部分検出された画像 (図 8)

表 2 に評価基準を示す．この評価基準で評価した結果を表 3 に示す．

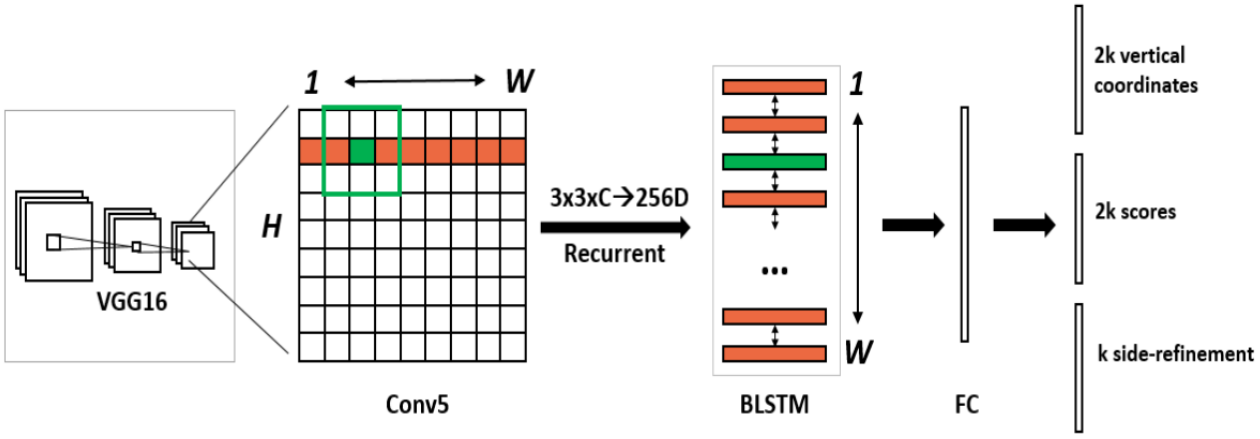

図 4　CTPN のアーキテクチャ [5]

図 5　CTPN のアーキテクチャ [5]

表 3　テキスト検出した画像の結果

total	standard1	standard2	standard3	ng
51,819	6,200	44,353	1,266	0

6. ラベルの検出

検出されたテキスト位置に基づいてラベルの位置を特定する．

Step 1　テキスト検出された各矩形の中心を求める（図 9）．中心は各座標の平均値だと定義する $(\overline{X}, \overline{Y})$．

$$\overline{X} = \sum_{i=1}^{n} x_i, \overline{Y} = \sum_{i=1}^{n} y_i \quad (1)$$

Step 2　求めた各中心の中心を求める（図 10）．

Step 3　得た中心に基づいて正方形を描く（図 11）．

ラベル位置を特定できた画像の評価結果を表 4 に示す．

7. 結　　果

検出エラーの原因を以下に示す（図 12）．

（1） 極端値（テキスト検出が間違ってるところ）に影響を受けた（図 12(a)）．

（2） JH 番号とラベルをかなり離れている（図 12(b)）．

（3） JH 番号しか検出してない（図 12(c)）．

（4） JH 番号しかない（ラベルがない）（図 12(d)）．

本研究では標本画像のラベルを検出して，ラベルのすべてをマスキングするのではなく，将来における研究でのデータ利用も考慮し，マスキング範囲は最小限にするよう努め，科名，和名や採集日などを隠してしまわないようにする予定である．

表 4　ラベル位置した画像の結果

total	OK1	NG2	proportion
51,819	45,560	6,259	88%

8. 考　　察

CTPN ネットワークを使って萩庭植物標本画像のテキスト検出を行った結果，全部のテキストを検出できたのは 1 割で，一部分のテキストを検出できたのは 9 割である．ある程度の手書き文字を検出できた一方，印刷書き文字を十分に検出できておらず，CTPN ネットワークにおいて，手書き文字検出は可能であると考えられる．また印刷書き文字の検出率が低い原因としては，萩庭植物標本画像の解像度が低いことやノイズのためであると考えられ，適当な画像前処理が必要である．

CTPN ネットワークを使って，萩庭植物標本画像のテキスト検出率を向上きる可能性があるが，できるだけ早期に公開するために，標本ラベル全体のマスキングを第一段階の目標にする．テキスト検出結果に基づいて，標本ラベル位置を特定できる画像は約 88% である．各矩形の中心に基づいてラベル位置を特定する方法は有効であった．極端値の影響を減少するのが今後の課題である．

また，萩庭植物標本画像に向けて，CTPN ネットワークを調整するのも今後の課題である．標本上のテキストを全部検出できるのは次の目標である．検出できた画像から光学的文字認識を使用して，マスキング範囲を最小限にすることを最終目的とする．

図 6　全部テキスト検出された画像（例）

図 7　一部テキスト検出された画像（例）

図 8　テキスト検出されてない画像やテキストではない部分検出された画像（例）

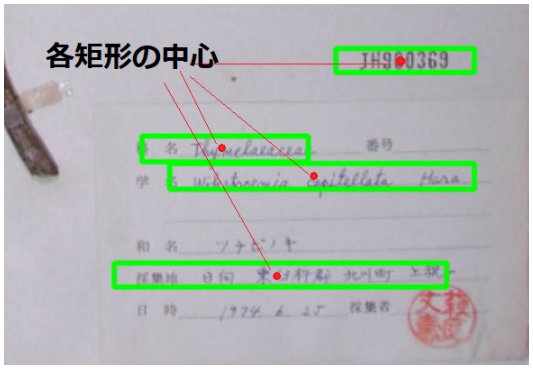

図 9　ラベル位置特定の Step1

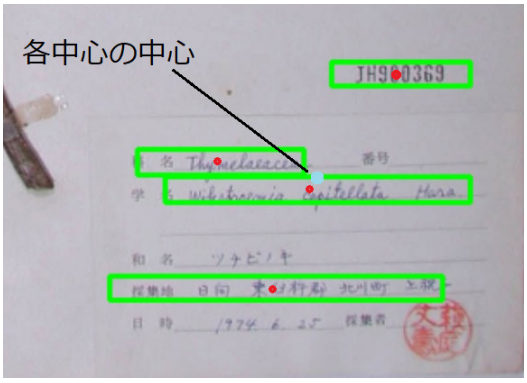

図 10　ラベル位置特定の Step2

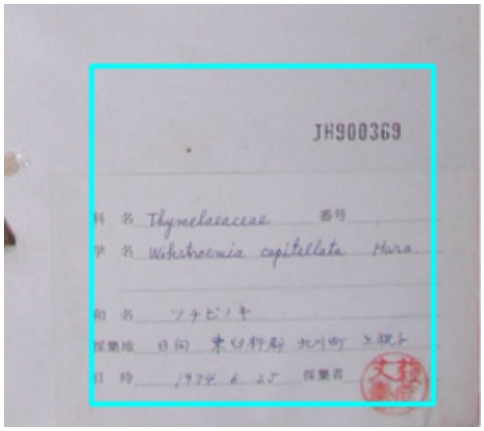

図 11　ラベル位置特定の Step3

9. ま と め

本研究では標本画像における公開できないデータが記載されてるラベルを検出し，マスキング範囲の決定・実行までの一連の操作を自動的に行う方法について検討を行った．先行研究うを含め複数手法を試したところ，ラベルの直接的検出は難しいものの，テキスト検出ネットワーク CTPN を使用することで，間接的に位置を推定するアプローチが有効と言える結果が得られた．まだ調整を要する部分はあるものの，標本画像に記載された採集地など貴重なデータだけ自動マスキングする方法のめどが得られ，OCR の活用やデジタル・スカラシップ開発の上でも，本研究では一定の成果および示唆が得られたと考えられる．

(a)　(b)

(c)　(d)

図 12　検出エラーの例

文　献

[1] 萩庭植物標本画像データベース作成協力会 , 萩庭植物標本画像データベース作成プロジェクト総括報告書. 千葉市稲毛区 : 萩庭植物標本画像データベース作成協力会, 2008. p. 17. 第 1 巻.

[2] 檜垣泰彦, ほか.“千葉大学学術リソースコレクション (c-arc) ～大学図書 館 における情報システム開発事例～”. 一般社団法人電子情報通信学会, 信学技報, vol. 118, no. 420, LOIS2018-51, pp.51–56, 2019 年 1 月.

[3] 日野遥，檜垣泰彦．（２０２０）．光学的文字認識を活用した植物標本画像のラベル自動マスキング方法の検討．電子情報通信学会技術研究報告，vol. 119, no. 477，pp.57–62，2020 年 3 月.

[4] Ren, S., He, K., Girshick, R., & Sun, J.. Faster r-cnn: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems, pp. 91–99 (2015).

[5] Tian, Z., Huang, W., He, T., He, P., & Qiao, Y. Detecting text in natural image with connectionist text proposal network. In European conference on computer vision, pp. 56–72. Springer, Cham (2016, October).

[6] Shi, B., Bai, X., & Belongie, S. . Detecting oriented text in natural images by linking segments. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2550–2558 (2017).

[7] Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition , in International Conference on Learning Representation (ICLR) (2015).

[8] Graves, A., Schmidhuber, J.: Framewise phoneme classification with bidirectional lstm and other neural network architectures. Neural Networks 18(5), pp. 602-–610 (2005).