

## Web ページ画像からの UI パーツの抽出

加藤 聡太郎<sup>†</sup> 四十崎 航<sup>†</sup> 篠澤 佳久<sup>††</sup> 飯村 結香子<sup>‡</sup> 斎藤 忍<sup>‡</sup>慶應義塾大学大学院 理工学研究科<sup>†</sup> 慶應義塾大学 理工学部<sup>††</sup>日本電信電話株式会社 コンピュータ&データサイエンス研究所<sup>‡</sup>

## 1 はじめに

近年、深層学習の分野における技術の発展により、その適用領域は多様な分野に渡っている。そのような中で GUI 画像から基本的な要素 (UI パーツと呼ぶ) を深層学習の技術を用いて検出する試みがされている [1]。この試みが可能となった場合、UI パーツの再利用技術の確立につながることも期待できるためである。

そこで本研究においては、Web ページを対象として、検索したい UI パーツと類似した画像を Web ページから抽出することを試みる。特に深層学習による検索手法の確立について検討する。

## 2 提案

本研究における UI パーツ画像の検索は一般的な物体検出と同様の処理過程で可能と考える。そこで物体検出と同様に下記の手順にて UI パーツ画像の検索を行う (図 1)。

(1) 各 Web ページ画像 (検索先画像群) を対象として領域分割を行い、UI パーツ画像の候補領域を検出する (図 1 左図)。

(2) (1) の領域分割処理で検出された候補領域の中から、検索したい UI パーツ画像 (検索元画像) と類似した候補領域を検索する (図 1 右図)。

本稿では (1) の領域分割について報告する。

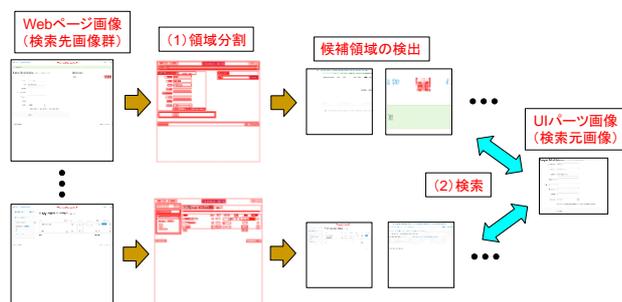


図 1 提案 (左: 領域分割, 右: 検索)

## 3 提案手法

先行研究 [1] においては、端点抽出処理のような画像処理による手法 (Old Fashioned)、深層学習による手法 (Deep Learning) を用いて UI パーツの検出を行っている。そこで本研究においては、下記の 3 種類の手法を用いて UI パーツ画像の候補領域を検出する。

手法①: Selective Search による領域分割

手法②: Pix2Pix を用いた領域分割

手法③: YOLO を用いた領域分割

## 3.1 Selective Search による領域分割 (手法①)

Selective Search [2] は色やテクスチャ特徴の類似した小領域を段階的に結合する領域分割の手法であり、R-CNN (Regions with CNN features) における領域分割としても利用されている。手法①の手順を以下に示す。

(1) 対象とする Web 画像を 1/4, 1/16 の大きさにした後、別個に Selective Search を行い、候補領域を検出する。

(2) 検出された全ての候補領域を二枚ごとに組み合わせ、IOU (Intersection over Union) を求める。そして算出された IOU が閾値 (0.85) よりも高い場合、一方の候補領域を削除する。以上の処理で残った候補領域を最終結果とする。

## 3.2 pix2pix を用いた領域分割 (手法②)

pix2pix [3] は敵対的生成ネットワーク (Generative Adversarial Network) の一手法であり、一对のペア画像からペア画像間の変換方法を学習することによって、一方の画像から対応する画像を生成する手法である。そこで Web ページ画像と対応する領域分割画像 (図 2 中の図①②) をペア画像として扱い、このペア画像を pix2pix により学習させることによって領域分割を行う。手法②の手順を以下に示す。

(1) Web ページ画像と領域分割画像から成るペア画像を学習データとして用意する。pix2pix により Web ページ画像から領域分割画像を生成する Generator、正解となる領域分割画像と Generator により生成された領域分割画像を区別して識別する Discriminator の学習を行う (図 3)。

Extraction of UI parts from web page images  
Sotaro Kato<sup>†</sup>, Wataru Aizaki<sup>†</sup>, Yoshihisa Shinozawa<sup>††</sup>, Yukako Iimura<sup>‡</sup>, Shinobu Saito<sup>‡</sup>

<sup>†</sup>Graduate School of Science and Technology, Keio University

<sup>††</sup>Faculty of Science and Technology, Keio University

<sup>‡</sup>Nippon Telegraph and Telephone Corporation

(2) 学習後の Generator に未知の Web ページ画像を入力し、領域分割画像を生成する (図 2 中の図③)。

(3) 生成された領域分割画像に対して、モルフォロジー変換による膨張処理、輪郭抽出処理による矩形検出、検出位置を線の太さの分修正する、などの後処理を行い最終的な領域分割画像を得る (図 2 中の図④)。

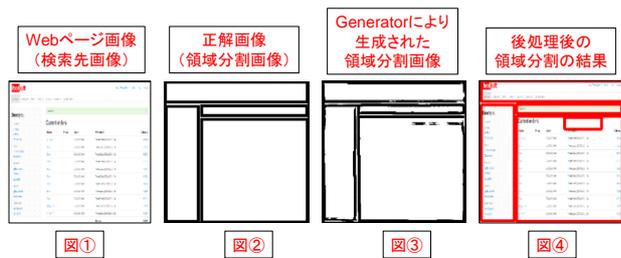


図 2 pix2pix における領域分割画像

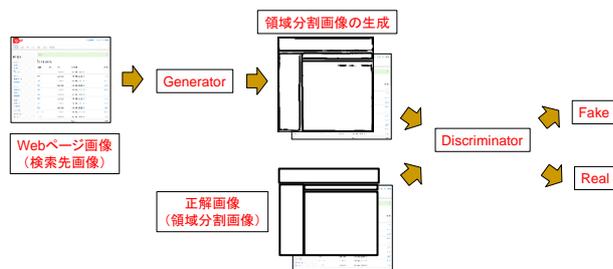


図 3 pix2pix を用いた領域分割画像の学習

### 3.3 YOLO を用いた領域分割 (手法③)

YOLO (You Look Only Once) [4]は、物体の検出と認識を同時に行う手法であるが、本研究においては、全ての UI パーツ領域の検出と認識を行うことによって領域分割を試みる (図 4)。なお本研究においては YOLO v5 を用いた。

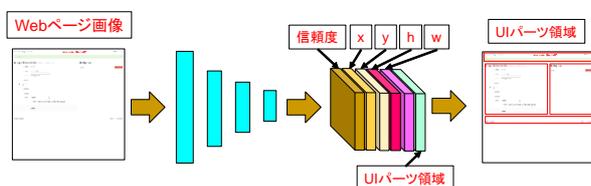


図 4 YOLO を用いた UI パーツ領域の学習

## 4 評価実験

評価実験のため、オープンソースソフトウェア (OSS) の Web システム上で 381 枚の Web ページ画像を収集した。そして UI パーツに起こり得る変化 (データ数や内容による変形、重畳など) を考慮した上で 56 種類 (ヘッダー、フッター、メッセージなど)、1,307 枚の UI パーツ

を手動で検出し、領域分割の各手法についての評価を行った。手法①においてはテストデータとして全データを用いる。一方、手法②③においては、学習データとテストデータの枚数比は 9:1 とした。手法③については予測時に、信頼度が 0.25 以上の候補領域を抽出した。

結果を表 1 に示す。表 1 には Web ページ画像一枚を領域分割後、検出された候補領域数の平均値、正解の UI パーツと検出された候補領域の中で IOU が最大となった領域を求め、それら最大 IOU の平均値を精度として示す。

表 1 評価実験の結果

	候補領域数の平均	精度
手法①	38.0 枚	0.739
手法②	6.9 枚	0.784
手法③	3.7 枚	0.894

表 1 より深層学習による手法 (手法②③) の方が、手法①よりも領域分割後の候補数が削減でき、さらには精度の向上も図れたことが分かる。一方で手法②③の場合、手法①と異なり学習データを必要とするため、事前に人手による学習データおよび正解ラベル付けが必要となる点について検討の必要性がある。

## 5 まとめ

本研究においては、Web ページ画像を対象として領域分割を行い、UI パーツ画像の候補領域を検出する手法の考案を試みた。今後は検出された候補領域の中から、検索したい UI パーツ画像と類似した候補領域を検索する手法の考案を行っていく予定である。

## 参考文献

[1] Jieshan Chen *et al.*, Object detection for graphical user interface: old fashioned or deep learning or a combination?, Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering, pp.1202-1214, 2020.  
 [2] J. R. Uijlings *et al.*, Selective Search for Object Recognition, International Journal of Computer Vision, 104(2), pp.154-171, 2013.  
 [3] Phillip Isola *et al.*, Image-To-Image Translation With Conditional Adversarial Networks, Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1125-1134, 2017.  
 [4] Joseph Redmon *et al.*, You Only Look Once: Unified, Real-Time Object Detection, Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.779-788, 2016.