

類似性に着目した GUI 画像検出に関する考察

2019SE064 渡部祐士

指導教員：張漢明

1 はじめに

近年 CNN 関連分野の発展により機械学習による物体検出が広く利用されているようになった。GUI 画像の検出もその内の 1 つである。GUI とはグラフィカルユーザインタフェースの略称であり、ウェブサイトやアプリケーション等の操作をユーザが行う際の要となる情報を媒介する視覚的要素の事である。機械学習による GUI の検出は、ソフトウェア自動テスト技術 [1]、テキストリーダーアプリの補助 [2]、アプリ実行環境の変化によるレイアウト変化の検知 [3] 等に用いられる。これらは一般的な物体検出モデルと同様の手法で行われているものの、Chen らの研究 [1] では GUI がもつ 4 つの固有の特性が機械学習による GUI 画像検出の精度の低下を招いているとしている。そのため、機械学習による GUI 検出精度の向上には、モデルの性能とは別に GUI の特性に着目したアプローチが求められる。

本研究の目的は、GUI 固有の特性の 1 つである類似性の高さに着目したラベル付けと GUI 検出精度を考察する事である。GUI の類似性に基づいてラベルの種類や数を変化させることで GUI の検出に起こる影響を、機械学習モデルを用いた検証によって明らかにしていく。

2 関連研究

2.1 物体検出モデル

物体検出技術モデルでは一般的に CNN(Convolution Neural Network) をベースとしたモデルが多数存在しているが、今回は YOLO[4] を用いる。YOLO は完全畳み込みネットワークを用いた検出と認識を並行して行うモデルである。矩形による物体範囲の検出、他の機械学習モデルに比べて処理速度が速い、検出した物体に対して confidence (信頼度) という検出領域の正確さとクラス予測確率を示す指標が示される、等の特徴がある。矩形による検出と confidence という指標の存在、先行研究で GUI 検出に一定の成果が認められている点を、GUI 検出の評価をより適切に行える要因と考え YOLO を検証に採用する。

2.2 評価手法

物体検出の精度を評価する際、再現率 (Recall) と適合率 (Precision) という指標が用いられる。再現率は正解が正例であるもののうち、モデルが正例だと判定したものの割合、適合率はモデルが正例だと判定したうち、本当に正例だったものの割合を指す。この 2 値はトレードオフの関係にあり、グラフにとる事で、クラスの精度を示す指標である AP(平均適合率) を求める事ができる。またモデル全体の精度は各クラスの AP を平均化した mAP で評価され

る。AP, mAP はどれも ($0 < n < 1$) の範囲で表せられ、1 に近い程優秀である事を示す。本研究ではこの指標を中心に精度の評価を行う。

2.3 GUI 画像固有の性質

Chen らの研究 [1] では、機械学習モデルの GUI 検出時に障害となりうる 4 つの GUI 固有の特性を提示している。1 つの GUI が複数の要素 (パーツ) で構成されていると捉えた場合、以下の様な特性が考えられる。

特徴 1 Large in-class variance

クラス内の分散性の高さ。同クラスに設定されている GUI 同士でも、構成要素の外形やサイズに差異がある事。

特徴 2 High cross-class similarity

クラス間の類似性の高さ。別クラスに設定されている GUI 同士でも、構成要素の外形に類似性が見られる特徴。

特徴 3 Mix of heterogeneous objects

様々な要素の組み合わせ。GUI の構成要素は図形だけで無く、画像やテキストが用いられるという特徴。GUI の一部の要素を別個体の GUI として捉えてしまう誤謬となる。

特徴 4 Packed scene and close-by elements

まとめられたシーンと近接した要素。多くの GUI が他の GUI と隣接して整列している事。そのため、個別の GUI を 1 つの GUI として捉えてしまう誤謬となる。

また、一般的な物体検出と共通する事項ではあるがデータセット数の多さが精度の高さに影響する事も、検証結果より示されている。

2.4 GUI 画像検出を利用した先行研究事例

丹野の研究 [3] で提案された GUI 検出モデルでは、意図していない広いサイズで GUI 領域が切り取られている場合が確認されており、先述の特徴 3,4 による影響だと考えられる。

Zhang らの研究 [2] で提案された GUI 検出モデルでは、checkbox の GUI の検出率が想定的に低いという結果がでている。この原因にデータセット数の少なさと icon の GUI と誤認を挙げており、特徴 2 による影響だと考えられる。

3 基本的なアイデア

本研究では単一の機械モデルによる GUI 検出精度が、ラベルの分類によってどう変化するかを検証で確かめる。

ラベルの対象は、**特徴 1,2** が顕著に表れかつ使用頻度も多い Button の GUI を対象に行う。ラベルの分類は 2 パターン行う。(1) **図 1** に示されるようなクラス間の類似性が高い 5 種ラベル, (2) **図 1** の 5 種を Button クラスとして統合したクラス内の分散が大きい 1 種ラベル。(1) は**特徴 1** から GUI 同士の誤分類が多くなると考えられる, (2) **特徴 2** から対象外の GUI の誤認識が多くなると考えられる。検証手順を以下に示す。

1. ウェブページや UI デザインのスクリーンショット 357 枚をデータセットにラベリングツール Vott でラベル付けを施す。
2. 学習は Google Colaboratory 環境で構築した YOLOv5s で行う。
3. テストデータからモデルの精度が算出される。
4. モデルの作成後、追加で検出テストを行う。テストでは類似性の高い GUI を含む画像 11 枚を共通で扱う。



図 1 ラベルの分類 5 種とその例

4 結果

モデルの検出精度と追加テストの結果について確認できた特徴をまとめる。

結果 1 mAP 値では 1 種ラベル > 5 種ラベルとなった。

結果 2 5 種ラベル内では、AP の値は contained が最も高く radio が最も低かった。

結果 3 分散の大きい button クラス、5 種ラベルの contained、icon クラスでは対象外の GUI を対象と誤認識するケースが多く見られた (**図 2**)。

結果 4 類似性の高い 5 種ラベル内のクラス同士で、同一の GUI に対して複数のクラスを重複して誤認識しているケースが多く見られた (**図 3**)。

button クラスや contained クラスはデータセット数が他クラスよりも多かった事から**結果 1**・**結果 2** のような精度の高さに繋がったと考えられる。また、クラス内分散が大きいクラスでは**結果 3**、クラス間類似性の高いクラスでは**結果 4** がより顕著に表れる事が確認できた。

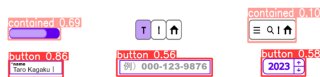


図 2 対象外と誤認する例



図 3 ラベルが重複した例

5 考察

以上の結果から精度の向上をおこなう手段として以下の作業が考えられる。

- 1 誤検出・分類防止のためしきい値を高めを設定する。
- 2 クラスの対象を広く設定し、モデル内・クラス間の類似性を低くする。
- 3 各クラスの種類は類似性が低くする事を優先し、分散の大きさによる精度の低下はデータセットの向上で補う。
- 4 以上に着目し、分散と類似性の均衡が保たれる分類を考える。

6 おわりに

本研究の検証によって GUI 固有の性質とラベル付けが与える影響の一部を考察する事が出来た。一方で対象とする GUI の数及び種類の増加, 使用モデルの変更, 他の GUI 特性への着目等, 異なる条件での研究も必要だと感じた。

今後も機械学習モデル精度と GUI 特性の関係性を解明する事で, GUI 自動検出技術の幅広く実用的な利用に貢献に期待したい。

参考文献

- [1] Jieshan Chen, Mulong Xie, Zhenchang Xing, et al. : Object Detection for Graphical User Interface: Old Fashioned or Deep Learning or a Combination?, Proc.ESEC/FSE 2020, pp.1202-1214, ACM, 2020.
- [2] Xiaoyi Zhang, Lilian de Greef, Amanda Swearngin, et al. : Screen Recognition: Creating Accessibility Metadata for Mobile Applications from Pixels, Proc.CHI '21 pp.1-15, 2021.
- [3] 丹野 治門: 画像処理を活用した UI レイアウト崩れ検出支援手法の提案, 研究報告ソフトウェア工学 (SE), Vol.2016-SE-194, No.9, pp.1-8, <http://id.nii.ac.jp/1001/00175664/>, 2016.
- [4] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi : You Only Look Once: Unified, Real-Time Object Detection, Proc.CVPR, pp.779-788, IEEE, 2021.