

センター試験のイラスト問題におけるパーツ単位での物体検知 Object detection by parts in the illustration questions of english exam.

向後 ジェフリー¹⁾
Jeffrey Kougo

大和 淳司¹⁾ 杉山 弘晃²⁾
Jyunji Yamato Sugiyama Hiroaki

1. はじめに

「ロボットは東大に入れるか」(以下「東ロボ」)[1]は、大学入試問題を解く総合的情報処理を題材としたAIプロジェクトである。東ロボの英語班は研究の一つとしてセンター試験英語リスニング問題のイラスト問題に着目し取り組んでいる。解答のための基本的な方策として、各選択肢のイラストに対して英語キャプションを行うことで文章問題に変換することを想定している。しかしセンター試験のイラスト問題では選択肢が類似していることが多く、単純なキャプションでは異なる選択肢に同一の英文が付与されてしまう問題が発生しやすい。そこで本研究では、選択肢間の差異を反映した表現を獲得することを目的とした画像認識を行うことを目的として問題の分析と初期検討を行った。選択肢間の差異の情報として画像中のオブジェクト同士の位置関係や特定種類のオブジェクトの有無、オブジェクトの属性情報などが挙げられる。そこで、事例を用いて選択肢間から得られる差異の情報がイラスト問題の正答に与えられるかの検討を行った。

2. イラスト問題

2.1 出題問題の分析

イラスト問題を解く上で必要な選択肢間の差異のパターンを調べるため、実際に出題された過去のイラスト問題の分析を行った。その結果、イラスト問題を解答する上で必要な選択肢間の差異は以下のパターンに分類することができた。

- (1) 特定種類のオブジェクトの有無
- (2) 同一クラスのオブジェクト数
- (3) オブジェクト同士の位置関係
- (4) オブジェクトの属性情報

例えば、(1)は、選択肢1が犬の画像、選択肢2が猫の画像といった差異のパターンである。各選択肢画像に対して物体検出を行うことで必要な差異情報を獲得可能である。次に(2)は、選択肢1が一匹の犬の画像、選択肢2が二匹の犬の画像、という差異のパターンである。この場合も各選択肢画像に対して物体検出を行い、オブジェクトの数を正確に取得できれば選択肢間の差異を獲得できる。次に(3)では、選択肢1が犬と猫が横に並んでいる画像、選択肢2が犬と猫が縦に並んでいる画像、というパターンである。それぞれの選択肢画像のオブジェクトに対し二次元平面上の位置情報を取得し、犬のオブジェクトのx, y座標と猫のオブジェクトのx, y座標を取得することで選択肢の差異を獲得可能である。最後に(4)は、選択肢1が白い犬の画像、選択肢2が黒い犬の画像、という差異パターンである。この時どちらの画像も物体

検出結果としては「犬」のラベルであるが、オブジェクトの属性としてそれぞれ「白い」、「黒い」を正確に認識できれば正しく選択肢間の差異を検出できる。

以上の分類結果が可能であり、またここで列挙した順に難易度が高い問題となっていることが予想できる。そこで本論文ではまず(1),(2)にターゲットを絞り、複数オブジェクトを含む選択肢画像から「特定種類のオブジェクトの有無」と「同一クラスのオブジェクト数」の取得を行うこととした。

2.2 差異情報を用いたイラスト問題解法の事例

2010年度本試で実際に出題されたイラスト問題[2]を用いて、解答の処理の流れの例を示す。図1に選択肢、図2に設問を示す。図1、図2から「特定種類のオブジェクトの有無」と「同一クラスのオブジェクト数」が正答する上で必要であるとわかる。まず各選択肢の画像それぞれに対して物体検出を行う。適切な物体検出を行うことができれば、画像ごとのクラス名をカウントすることで表1を得ることができる。表1の「swing」、「tennis court」の列が正答に必要な差異であるため、正解の選択肢は「①」とであると認識できる。

問 4 Which park are the man and woman talking about? 4

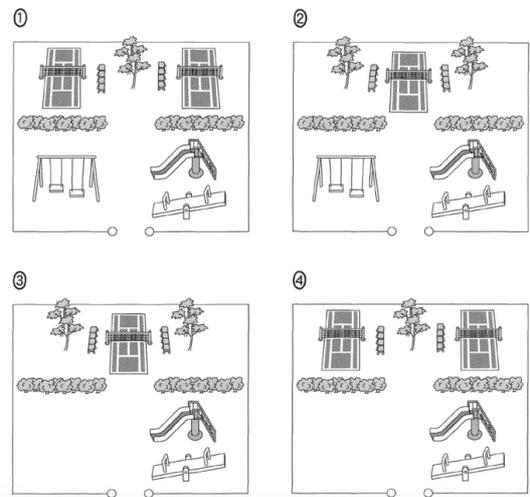


図 1: 2010 年度センター試験イラスト問題の選択肢

M: How's the new park?

W: It's nice, because it has swings, a slide and seesaw for kids.

M: Really? Can we play sports there, too?

W: Well, it has two tennis courts.

図 2: 2010 年度センター試験イラスト問題の設問

¹ 工学院大学大学院 Graduate School of Kogakuin University

² 日本電信電話 (株) NTT コミュニケーション科学基礎研究所 NTT Communication Science Laboratories

表 1: 選択肢に対するクラスごとのオブジェクト数

	slide	swing	seesaw	...	tennis_court
①	1	1	1	...	2
②	1	1	1	...	1
③	1	0	1	...	1
④	1	0	1	...	2

3. 実験

3.1 実験目的

本実験では、差異表現の認識として、画像に対する「特定種類のオブジェクトの有無」と「同一クラスのオブジェクト数」を取得し表 1 を適切に構築し提案解法の実現可能性を確認することを目的とする。

3.2 実験手順

本実験はデータの収集、データの拡張、モデルの学習、物体検知の手順で進めた。

データの収集として、まず 8 クラス分のイラスト画像を用意した。画像サイズは 320x320pixel である。またテストデータとして 1 画像中に 2 のオブジェクトを含む画像 A と画像 B を用意した。画像サイズは 640x640pixel でありオブジェクトのスケールは学習データと変わらない。

データ拡張の処理手順について説明する。収集したイラスト画像に対して 2pixel の間隔で水平方向と垂直方向それぞれ 10 パターンシフトさせる。この時点で各クラスに対し、100 枚の画像へ拡張される。次に、シフトによって拡張された画像に対しスケールの変化を行いさらにデータを拡張する。スケールを縮小させる場合、まず 320x320pixel 以下に画像サイズを縮小する。縮小した画像に対して 320x320pixel の画像サイズになるよう足りない画素に対し白埋めを行う。スケールを拡大させる場合、まず 320x320pixel 以下のサイズで画像の切り抜きを行う。切り抜いた画像に対して 320x320pixel のサイズになるようリサイズを行う。シフトに拡張された画像に対して、スケールの縮小とスケールの拡大をそれぞれ 2 パターン行うことで各クラス 500 枚の画像へと拡張を行った。

モデルの学習には SSD[2]を使用した。またエポック数は 10epochs バッチサイズは 16 として学習を行った。また学習したモデルを用いて、画像 A と画像 B に対し物体検知を行った。

3.3 実験結果

画像 A と画像 B に対する出力結果をそれぞれ図 3、図 4 に示す。また図 3、図 4 の出力結果から得られる画像 A と画像 B の差異表現を表 2 に示す。

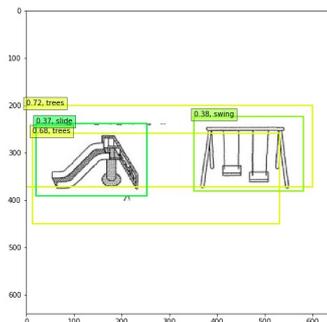


図 3:複数オブジェクトを含む画像 A の出力結果

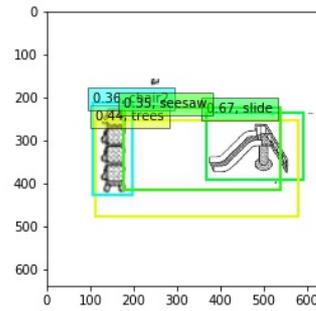


図 4:複数オブジェクトを含む画像 B の出力結果

表 2:画像 A と画像 B の差異表現

	slide	swing	seesaw	chair1	...	trees
画像A	1	1	0	0	...	1
画像B	1	0	1	1	...	1

図 3 で「tree」と認識されたバウンディングボックスが 2 つあるが、同一クラスのバウンディングボックスが重なっている場合オーバーラップの比率が閾値以上である場合そのバウンディングボックスのクラスはカウントされないため画像 A に対するクラス「tree」のオブジェクトの数は 1 となる。

4. 考察

図 3 と表 2 よりクラス「slide」とクラス「swing」に対してオブジェクト数の認識が適切に行えているとわかる。しかしクラス「trees」のオブジェクトが画像中に含まれていないのにも関わらず存在していると誤認識している。また図 4 と表 2 よりクラス「chair1」とクラス「slide」に対してオブジェクト数の認識は適切に行えているとわかる。しかしこちらも、画像内に含まれていないクラスのオブジェクトが存在すると誤認識してしまっている。線画に対する認識率が十分高くないため誤認識によって正しい差異表現の表を構築することが出来なかった。選択肢間の差異を正確に認識するためには、より検出器の精度をより高める必要がある。特に誤検出の抑制が必要である。

5. まとめ

イラスト問題の出題傾向の分析結果と解法の事例を示した。また、様々なテストデータの条件でイラスト画像の物体検知を行った。認識精度の向上のためにデータの増量や学習手法の改善など今後の課題は多い。引き続き、センター試験のイラスト問題の正答率向上を目指し必要な他の差異表現の取得を進めていく。

参考文献

- [1] 新井紀子, 松崎拓也, “ロボットは東大に入れるか”, 人工知能学会誌 27 巻 5 号 463-469, 2012
- [2] “センター試験英語 リスニング問題 過去問の全て 全 27 回分 大学入試センター試験 英語リスニング問題 千進 e 予備校”, 2020 年 5 月 1 日閲覧
- [3] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. Berg, “SSD:Single Shot MultiBox Detector”, Computer Vision -ECCV 2016 pp21-37, 2016.