

YOLOv5 を用いたカメラ画像からのキムワイプ検出と 入カインターフェースへの応用

小嶋明

東京大学キムワイプ卓球会

a.kojima8924@gmail.com

1 はじめに

東京大学キムワイプ卓球会は 2017 年から東京大学の五月祭において研究会を開催している。過去の研究会では、機械学習を利用して画像のキムワイプらしさを定量化して判定する研究 [1] や、深層学習によってキムワイプに類似する画像を生成する研究 [2] が行われた。これらの研究ではキムワイプの画像に対して機械学習の手法を適用することの有効性が示されたが、キムワイプ卓球との関連性は高くない。また、研究会では VR 空間内にキムワイプ卓球を再現する VR キムワイプ卓球が提案された [3]。VR キムワイプ卓球は VR ゲームの手法によって新しいキムワイプ卓球の体験をもたらす可能性をもっているが、VIVE コントローラーを用いて VR 空間のキムワイプを操作するため、キムワイプを手を持ってピンポン球を打つキムワイプ卓球本来の打球感が得られないという指摘がされている。

本研究では、コンピュータゲームとしてキムワイプ卓球を再現し、画像中のキムワイプの検出を深層学習の手法で実現し、カメラの前のプレイヤーのキムワイプの操作に連動するキムワイプ卓球ゲームを提案する。本研究は、新型コロナウイルスによって対面でのキムワイプ卓球が困難な状況において、遠隔地のプレイヤーが各自のキムワイプを持って対戦するリモートキムワイプ卓球の実現に貢献する。

2 方法

提案するシステムの構成を図 1 に示す。実験には Web カメラを接続したコンピュータとキムワイプを用いる。まず、プレイヤーは Web カメラの前に

キムワイプを持って構える。Web カメラで撮影された画像は、後述する物体検出手法である YOLOv5 [4] のニューラルネットワークに入力され、画像の中に存在するキムワイプの縦横の位置と幅・高さが出力される。キムワイプの位置と大きさの情報は、Unity で構成されたキムワイプ卓球ゲームに入力され、ゲーム内のキムワイプの 3 次元の位置に反映される。画像の撮影からゲームへの反映を約 18 FPS の速度で繰り返し実行し、プレイヤーが手で操作するキムワイプとゲーム内のキムワイプの位置を連動させる。ただし、現時点ではキムワイプ卓球のゲーム部分は未完成であり、キムワイプの検出と Unity 内のキムワイプの操作のみ実装されている。

システムにおいて最も重要な部分は入力画像からキムワイプを検出するプロセスである。本研究では、CNN を用いる 1 ステージ型の物体検出手法である YOLOv5 を利用する。YOLOv5 は高速で実行できる長所をもち、リアルタイムの物体検出を十分な精度で達成することができる上に、導入が容易であり、独自のデータセットを用いて学習させることで様々な物体を対象とする検出器を作成できる。YOLOv5 には大きさが異なり、検出精度と速度に差がある 4 つのモデルが実装されているが、本研究では 2 番目に大きいモデルである YOLOv5l を採用する。また、各モデルには COCO データセットで事前学習された重みデータが配布されており、利用可能である。学習を高速化するため、YOLOv5l のモデルに事前学習の重みを適用し、画像から特徴を抽出する Backbone 部分の 10 層の重みを固定して転移学

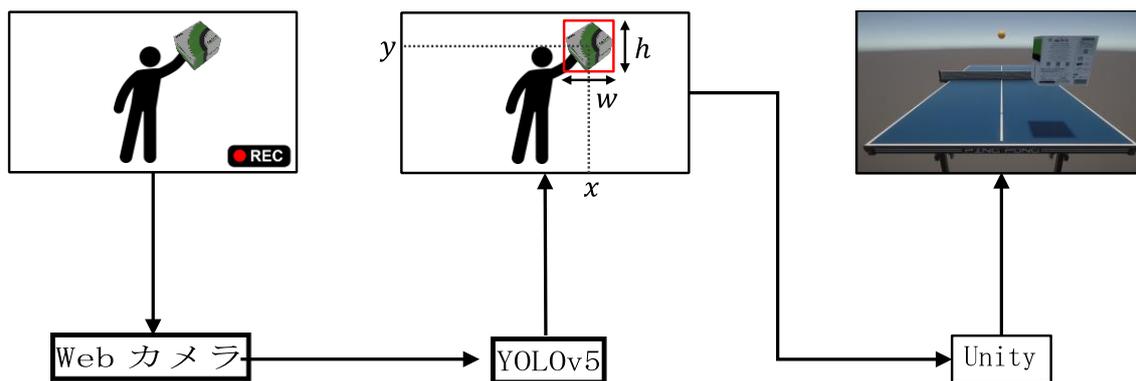


図1 システムの構成図

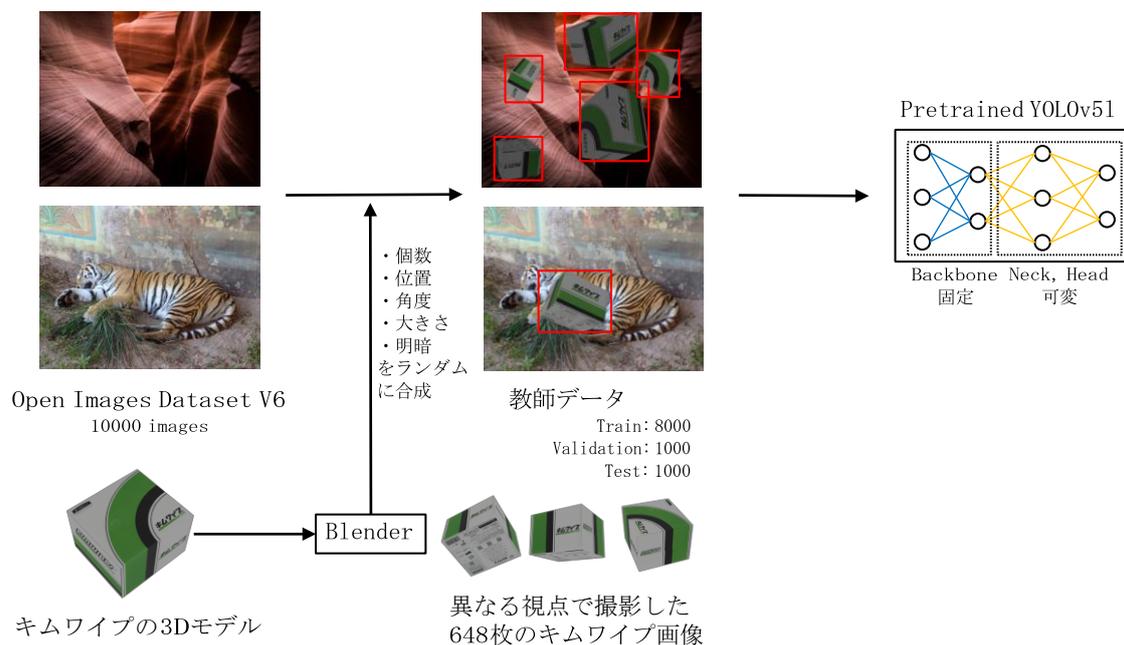


図2 YOLOv5 の学習

習を行う。

重みの学習には、十分な量の教師データが必要である。教師データはキムワイブを含む画像を撮影・収集し、キムワイブの存在する範囲を矩形で指定することでも作成できるが、人力でのデータセットの作成には膨大な時間を要する。そのため、本研究ではキムワイブの3Dモデルを背景に合成し、教師データを生成する [5-7]。学習の手順を図2に示す。まず、キムワイブの各面をスキャンしてキムワイブの3Dモデルを作成し、Blender上でカメラの位置を変えながら648枚の画像を撮影する。さらに、汎用的な画像データセットであるOpen Images V6から10000枚の画像を入手し、個数、位置、角度、

大きさ、明暗をランダムに設定したキムワイブ画像を合成してキムワイブの存在領域のアノテーションを付加し、訓練データ、検証データ、テストデータにそれぞれ8000枚、1000枚、1000枚を割り当てる。訓練データを用いて、学習を300エポック繰り返す。学習させた検出器に合成画像と実画像のテストデータを入力し、性能を評価する。

3 結果と考察

3Dモデルから合成された1000枚の画像とWebクロウリングで収集し人力でアノテーションを付加した15枚の画像に対してキムワイブの検出を行った。検出結果の例を図3に示す。検出スコアは表1

に示される。合成画像では検出精度は非常に高いが、実画像では Recall がほぼ 100% であるため検出すべきほぼ全てのキムワイプを検出できているが、Precision はやや低いため過検出が発生している。

キムワイプ卓球ゲームでの利用においては、検出するキムワイプは 1 個で十分であり、カメラに向けるキムワイプの面を上面に限定するような制限を加えることで性能の向上が見込める。現状では検出精度を重視し、YOLOv5 の中で比較的大きなモデルを使用しているため十分な検出性能が得られているが、トレードオフとして計算負荷が大きい。そのため、高速化のためにより小さなモデルの使用も検討できる。

表 1 合成画像と実画像の検出スコア

	合成画像	実画像
Precision [%]	100	85.7
Recall [%]	99.9	99.5
AP ₅₀ [%]	99.5	97.5



図 3 合成画像と実画像に対する検出結果

4 まとめ

本研究では、キムワイプ卓球ゲームでの入力インターフェースへの利用を目的として、機械学習によるカメラの画像からのキムワイプの検出を試みた。キムワイプの 3D モデルから教師データを生成し、YOLOv5 のモデルを学習させてキムワイプ検出器を作成し、合成画像と実画像で性能を評価し、十分な検出性能が確認された。今後の展望として、キムワイプ卓球ゲームの完成とオンライン対戦機能の実

装、キムワイプの打球感を再現するためのウェアラブルデバイスの作成あるいはスマートウォッチ等の既存のデバイスの利用を目指す。参考文献

- [1] 崎下雄稀. SNS を活用したキムワイプの画像認識. 第 4 回キムワイプ卓球研究会, p. 12.
- [2] 田中匠, 中野堯雄, 妹尾梨子. KIMLIKES: 深層学習を用いたキムワイプ類似画像の生成. 第 6 回キムワイプ卓球研究会, pp. 2-3.
- [3] 濱田柚子. バーチャルキムワイプ卓球の実現及び VR 体験がもたらす可能性. 第 4 回キムワイプ卓球研究会, pp. 13-14.
- [4] Glenn Jocher, Ayush Chaurasia, Alex Stoken, Jirka Borovec, NanoCode012, Yonghye Kwon, TaoXie, Jiacong Fang, imyhxy, Kalen Michael, Lorna, Abhiram V, Diego Montes, Jebastin Nadar, Laughing, tkianai, yxNONG, Piotr Skalski, Zhiqiang Wang, Adam Hogan, Cristi Fati, Lorenzo Mammana, AlexWang1900, Deep Patel, Ding Yiwei, Felix You, Jan Hajek, Laurentiu Diaconu, and Mai Thanh Minh. ultralytics/yolov5: v6.1 - TensorRT, TensorFlow Edge TPU and OpenVINO Export and Inference, February 2022.
- [5] Xingchao Peng, Baochen Sun, Karim Ali, and Kate Saenko. Learning deep object detectors from 3d models. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 1278-1286, 2015.
- [6] Georgios Georgakis, Arsalan Mousavian, Alexander Berg, and Jana Kosecka. Synthesizing training data for object detection in indoor scenes. In *Proceedings of Robotics: Science and Systems*, Cambridge, Massachusetts, July 2017.
- [7] Pengzhou Tang, Yu Guo, Han Li, Zhen Wei, Guanguan Zheng, and Jun Pu. Image dataset creation and networks improvement method based on cad model and edge operator for object detection in the manufacturing

industry. *Machine Vision and Applications*,
Vol. 32, No. 5, p. 111, Aug 2021.

Google Scholar^[2]を用いたオンラインサイエンティフィック・スポーツの模索

五百藏真聖

東京大学キムワイブ卓球会

ioroi-masato@g.ecc.u-tokyo.ac.jp

1. はじめに

2019年末にCOVID-19のパンデミックが始まって以降、東京大学キムワイブ卓球会はオンライン主体の活動へシフトしてきた。以降、キムワイブ卓球会においては非接触でのキムワイブ卓球会の実施を検討している。濱田によるVRキムワイブ卓球をひく非接触型のキムワイブ卓球が検討されてきたが、同時にサイエンティフィック・スポーツ全体での考慮も行った。12/21に実施された研究会において、Google Scholarの検索数を用いた単語数の類似性についての検討、word2vecを用いたコサイン類似度0に近づける検討などが提案された。これらはe-sportsとしてのサイエンティフィック・スポーツとして考えられた。ここでは前者について取り上げる。

サイエンティフィック・スポーツは国際キムワイブ卓球協会によると、「科学そのものではない科学界の方法」で「研究以外のこと」をすることであるとされている。

本研究では、オンラインにおいて手軽にできるサイエンティフィック・スポーツを提案する。パンデミックによって対

面活動が制限される中でサイエンティフィック・スポーツへの参加の敷居を下げ、研究者間の交流に資するものである。

2. 方法

方法に関して、以下の三点が提案された。

- ①単語を与えられ、その検索数に最も近い単語をあげる。
- ②単語を二つ上げ、最も検索数が近いものを挙げる。
- ③指定された検索数に最も近い単語を挙げる。

これらすべてGoogle Scholar等の検索数のわかる論文データベースにおける、学術的でない単語を検索したときに出るヒット数を比較するものである。ゲームマスター（以下GM）がお題を出して、参加者が検索することなく単語を上げていくものである。

- ①単語を与えられ、その検索数に最も近い単語をあげる。について。

参加者が三順の間に与えられた単語のヒット数を予測し、一巡に一単語ずつ上げ、最終的に最も近い単語を挙げた者の

勝利となる。12/21の部内の研究会においてこれが実施された。

②単語を二つ上げ、最も検索数が近いものを挙げる。について。

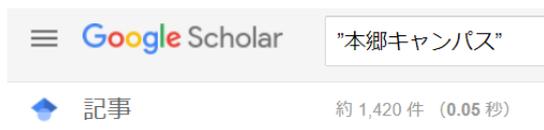
参加者が一巡の間に類似性の考えられない単語対を上げ、ヒット数の近い者の勝利となる。ここにおいてGMは不要であり、最少人数1人から実施可能である。

③指定された検索数に最も近い単語を挙げる。について。

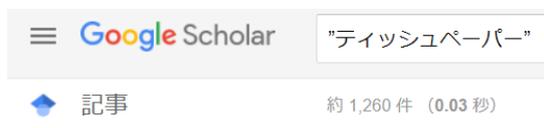
①の簡易版であり、参加者が三順の間に与えられた単語のヒット数を予測し、一巡に一単語ずつ上げ、最終的に最も近い単語を挙げた者の勝利となる。

3. 結果と考察

2021年12/21にオンラインで行った部内の研究会において①の形で実施。お題を1420件の“本郷キャンパス”と定め、最も近い単語は“ティッシュペーパー”であった。



[Google Scholar より作成](#)



[Google Scholar より作成](#)

まず、この営みをサイエンティフィック・スポーツと言えるかであるが、オン

ラインの論文データベースは紛れもなく「科学界の方法」であり、本質の「言葉遊び」は「研究ではない」を満たしているといえる。しかし、Google Scholarで「都市計画マスタープラン」と検索することは、研究的であることから、学術的な用語を避ける、又は「論文データベースで検索を行うこと」をサイエンティフィック・スポーツのサイエンティフィックの部分とすることで確固なものとすることができるであろう。

パンデミックとオンライン化の波でオンライン論文データベースはより重要になった。研究者同士の物理的な距離は近くなったと同時に、研究者間の様々な垣根を超えることはより求められるようになっていく。その需要にこれらは答えられる。

参考文献

- [1] 濱田柚子. バーチャルキムワイプ卓球の実現及びVR体験がもたらす可能性. 第4回キムワイプ卓球研究会, pp.13-14.
- [2] Google Scholar <https://scholar.google.com>
- [3] 国際キムワイプ卓球協会. サイエンティフィック・スポーツとは? <https://www.iktta.org/ss/index.ja.html> (2022年5月11日閲覧)