物体検出モデルを用いた アクションゲームにおける対戦分析

高橋一樹 [1]　尾関智子 [1]

概要：アクションゲームで勝率を上げるには，自分と相手プレイヤーの特徴を理解し，戦術を練ることが重要である．そのための手段として，自分の対戦動画や強いプレイヤーの対戦動画を分析する方法が有効だが，分析に大量の時間がかかる点が課題である．本論文では，技の使用回数を自動でカウントをするために，物体検出モデルによって技のモーション中の特徴的なフレーム画像を検出する手法を提案する．提案手法高精度に技を検出し，対戦動画の分析に有効であることを示す．

キーワード：アクションゲーム解析，物体検出，ゲーム技術支援

Using Object Detection Models Battle Analysis in Action Games

KAZUKI TAKAHASHI[1]　TOMOKO OZEKI[1]

Abstract: To increase the odds of winning in action games, it is important to understand the competitive style of the players and their opponents, and to develop tactics. One effective way to achieve this is to analyze video clips of the player's matches and those of strong players, but this analysis is time consuming. In this paper, we propose a method to detect characteristic frame images during the motion of a technique by using an object detection model to automatically count the number of times the player uses the technique. The proposed method detects techniques with high accuracy and is effective for analyzing playing styles in action games.

Keywords: Action Game Analysis, Object Detection, Game Skill Support

1. はじめに

近年，競技性の高いコンピュータゲームをスポーツ競技として捉えた e-Sports が話題になり，コンピュータゲーム業界に注目が集まっている．しかし，業界全体が抱える課題として，既存プレイヤーとの技術差による参入ハードルの高さや競技シーンにおけるトップ層との差が挙げられる．

対戦アクションゲームにおいて，強いプレイヤーになるために重要なことが 2 つある．1 つ目は練習の量を増やすことである．長い時間をかけて練習を積み重ねることでキャラクターを操作する精度が上がり，理想的な動きを無意識に選択できるようになる．2 つ目は練習の質を上げることである．対戦アクションゲームは人と人との対戦であるため，闇雲に練習をするのではなく，キャラクターの強みを理解し，相手プレイヤーに対する戦術を考えることで効率的に上達できる．そのために有効な方法として，動画共有サイトに投稿されている強いプレイヤーの対戦動画や自分の対戦を分析することが挙げられる．しかし、動画の分析には大量の時間がかかる点が問題である．

本研究では，Nintendo Switch の大乱闘スマッシュブラザーズ SPECIAL（スマブラ）[1]を題材に，対戦動画内で使用されたキャラクターの技の数を自動でカウントすることを目的とする．キャラクターが技を使用する際の特徴的なモーションのフレーム画像を学習し，カウントすることで分析による熟達を支援する．

1 東海大学大学院工学研究科電気電子工学専攻
Tokai University Graduate School of Engineering of Electrical and Electronic Engineering

2. 関連研究

三ツ井ら[2]は本研究と同様にスマブラの熟達支援を題材に，ステージや背景を無視してキャラクターの位置関係やポーズが類似した画像を検索する手法を提案した．三ツ井らの研究では，キャラクターの位置関係のみが似ているシーンについては 8 割の精度で検索を行えることが示された．しかし，キャラクターのエフェクトが似ているシーンに関しては検索精度が 4 割，キャラクターのポーズが似ている画像については背景が単調なステージにおいて 2 割，その他のステージでは 2 割以下の検索精度であった．また，三ツ井らは動画からキャラクターのみを抽出するため，物体を追跡しマスクを作成する SiamMask[3]と STM[4]を試したが，エフェクトやキャラクター同士の衝突によって失敗すると述べている．

梶並ら[5]は，OpenCV を用いたテンプレートマッチによりキャラクターの位置を検出し，位置関係に応じて注釈を行うことで対戦の観戦を支援するシステムを提案した．梶並らが題材としたゲームは，カメラのアングルとズーム度合いが一定であったためキャラクターの衣装に対してテンプレートマッチを適用することでキャラクターの位置を検出することができた．しかし，スマブラでは相手キャラクターとの位置関係によってカメラのアングルとズーム度合いが変わる（図 1）．また，爆発などのエフェクトによりキャラクターの色合いが変わる問題や，アイコンや相手との重なりによってキャラクターが隠れる問題がある（図 2）．そのため，既存の検出手法の適応は困難である．そこで，本研究ではより高精度でキャラクターの技を検出するため，物体検出モデルである YOLO[6]を用いて実験を行う．

図 1　位置関係によるカメラ位置の変化

図 2　エフェクトやシステムによる影響の例

3. 物体検出モデル

3.1 物体検出

物体検出とは，画像や動画の中から特定の物体の位置を検出する技術のことである．物体検出は 1.対象物が画像のどこにあるかを絞り込み Bounding Box を表示する，2.対象物が何であるかカテゴリー分類する，の 2 ステップで行われる．物体検出はディープラーニングを基にした手法が使われることが多く，有名なモデルとして YOLO[6]，Fast R-CNN[7]，SSD[8]が挙げられる．

3.2 YOLO

YOLO(You Only Look Once)[6]はディープラーニングに基づいた物体検出モデルである．YOLO では，まず入力画像を $S \times S$の正方形で細かく領域分けし，物体を検出できるよう学習する．それぞれのグリッドセルは B個の Bounding Box を持ち，以下の式で表される信頼度スコア(confidence)を推定する．

$$\text{confidence} = \mathrm{P_r}(\text{Object}) * \mathrm{IOU}_{\text{pred}}^{\text{truth}} \quad (1)$$

IOU(Intersection over Union)とは正解である範囲からの物体の近さを表す指標であり Bounding Box に物体が近いほど信頼度スコアは高くなる．それと同時に各グリッドセルは C個のクラスに対する条件付きクラス確率$\mathrm{P_r}(\mathrm{Class_i}|\text{object})$を予測する．さらに，計算された条件付きクラス確率と個々の Bounding Box の信頼度スコアを掛け合わせることで，Bounding Box 毎のクラスに対する信頼度スコアを得る．

$$\mathrm{P_r}(\mathrm{Class_i}|\text{object}) * \text{confidence} = \mathrm{P_r}(\mathrm{Class_i}) * \mathrm{IOU}_{\text{pred}}^{\text{truth}} \quad (2)$$

YOLO は，この信頼度スコアに基づいてどの Bounding Box で正解の物体を検出しているかを判断しており， 特徴としてリアルタイム処理が可能な処理速度，背景と物体の判別精度の高さが挙げられる．

3.3 YOLOv5

YOLO は 2023 年 10 月現在 v1 から v8 の 8 つのシリーズが開発されている．YOLOv1 は SSD よりも精度の面で劣っていたが，v2 以降は精度，処理速度ともに上回っている．本論文執筆時点では v8 が最新モデルであるが， Glenn Jocher が 2020 年 6 月に公開した YOLOv5[9]を使用する．

4. 提案手法

本研究では，YOLOv5 を用いて対戦動画からキャラクターA の 6 つの技のモーション中の特徴的なフレーム画像（図 3）を検出することで，技の使用回数をカウントする．技のモーション全体を動画として検出対象にすると，技が相手に当たった場合や相手にガードされた場合，そもそも当たらなかった場合，技が途中で途切れた場合，物体が途中で重なった場合など，学習のパターンが無数になる問題がある．また，処理時間が大幅に長くなることが予想されるため，モーション中のフレーム画像のみを検出対象とする．

5. 実験

5.1 学習画像の収集

スマブラでは攻撃が相手に当たると大きなエフェクトが発生するため，学習に使用する画像は他のモーションと被らず，攻撃判定が存在しないフレーム画像を選定する．さらに，どのような状況でも検出を行うため，学習用画像は実際の対戦動画から収集を行う．各技について 120 枚、計 720 枚収集した．

5.2 アノテーション

アノテーションとは，機械学習においてデータに情報を付加するプロセスのことを指す．YOLOv5 の学習には，学習対象の画像データに加えて，学習対象の位置を示す Bounding Box の座標と正解のクラスを記載したアノテーションファイルが必要である．本研究では，microsoft の提供するアノテーションツールである VoTT[10]を用いてアノテーションファイルを生成する．また，VoTT で出力される.json ファイルは YOLOv5 で読み込むことができないため，コンピュータビジョンプラットフォームである roboflow[11]を用いて変換し，YOLOv5 形式のデータセットを作成する．

5.3 データ拡張

データ拡張(Data Augmentation)とは，学習用の画像データに対して変換を与えることでデータの数を水増しする手法である．機械学習には通常多くの画像データが必要であるが，データ拡張によってデータ数を増加させることで少ない画像データからでも学習を行うことや精度を高めることが可能である．また，データ拡張によってデータのバリエーションを増やすことで過学習を防ぐ効果もある．

本実験では，通常の 720 枚のデータセットに加えて，データ拡張を行ったデータセットを用意する．コンピュータビジョンプラットフォームである roboflow を用いて，カメラの移動とエフェクトの影響に対応する狙いから，画像の拡大率を 0%から 30%，輝度とコントラスト 25%から−25%の間で変化させて 3 倍の 2160 枚に拡張したデータセットを作成した（図 4）．

5.4 YOLOv5 の学習

YOLOv5 の学習において，エポック数は 200，画像サイズは 640×640，バッチサイズは 8 とした．本実験では，画像の枚数とクラスが少なく，実際の動画データを用いてより正確に評価を行うため，学習画像は全て train データに割り振って学習を行う．

図 3 技名と選定したモーション画像

図 4 元の画像データの例（上）
データ拡張後の画像データの例(下)

5.5 性能評価指標

本実験では，性能評価の指標として，再現率(recell)と適合率(precision)，F 値(F-Measure)を用いる．再現率とは正例データをどれだけ正例と予測できたかを示す確率であり， 適合率は正例と予測したもののうち，どれだけ正解であるかを示す．そして，F 値は再現率と適合率の調和平均を示し，総合的な評価を行う指標である．

$$recell = \frac{TP}{TP + FN} \tag{3}$$

$$precision = \frac{TP}{TP + FP} \tag{4}$$

$$F = \frac{2(precision \times recall)}{precision + recall} \tag{5}$$

本研究では，YOLOv5 で検出を行った回数と，実際に使用した技の回数に対する再現率と適合率, F 値を求めることで性能を評価する．また，検出対象のフレーム画像の前後には類似したモーションがあり連続して検出を行うことが予想されるため, 適合率と F 値は 2 パターンを用いて評価する．検出対象のフレーム画像のみを正例とした場合を適合率 1，F 値 1，検出対象のフレーム画像の前後のフレーム画像も正例とした場合を適合率 2，F 値 2 とする．

5.6 実験方法

実験では，図 5 に示す 3 つのステージで各技を 12 回ずつ，計 36 回使用した，学習には用いていない新規の動画データを用いて性能を評価する．技 1～5 は右を向いて技を当てた場合と外した場合，左を向いて技を当てた場合と外した場合で 4 分類し 9 回ずつ使用している．技 6 のみ右を向いて地上で技を使用した場合と空中で技を使用した場合，左を向いて地上で技を使用した場合と空中で技を使用した場合で 4 分類し 9 回ずつ使用している．また，動画データにはキャラクターが歩くモーションやジャンプするモーション，技を使用する前後のモーションなどが含まれている．

検出の際は，不要な検出を防ぐため信頼度スコアに一定の閾値を設ける．閾値を 0.95，0.96，0.97，0.98 に設定して実験を行った後，データ拡張前のデータに 0.975，データ拡張後のデータに 0.955 の閾値を設けて追加実験を行った．

図 5　実験に使用したステージ

5.7 実験結果

実験の結果を図 6 と表 7，表 8 に示す。図 6 より，検出対象のフレーム画像のみを正例とした F 値 1 が最も高いのはデータ拡張後の閾値 0.96 の 0.492 であった。再現率は 0.787，適合率 1 は 0.497，適合率 2 は 0.985，F 値 1 は 0.492，F 値 2 は 0.875 である（表 1）．

F 値 2 が最も高いのはデータ拡張後の閾値 0.96 の 0.915 であった. 再現率は 0.870, 適合率 1 は 0.444, 適合率 2 は 0.968, F 値 1 は 0.482，F 値 2 は 0.915 である（表 2）．適合率は低下したものの，閾値を下げたことで再現率が上がっている．

全体を通して先行研究と比べて大幅に検出精度が向上しており，スマブラのキャラクターの特定のモーション検出に物体検出モデルが適しているといえる．

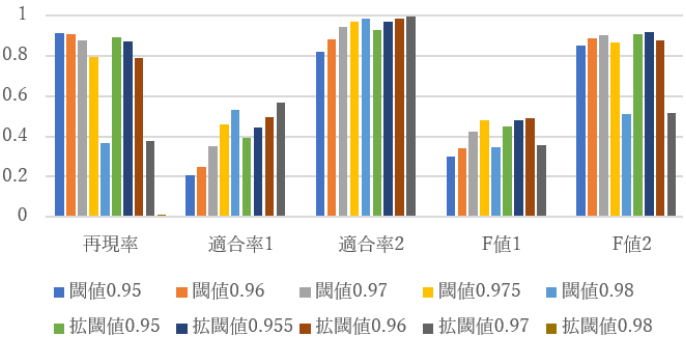

図 6　閾値毎の評価値

表 1　データ拡張後閾値 0.96 の評価値

技名	再現率	適合率 1	適合率 2	F 値 1	F 値 2
技 1	26/36	26/31	31/31	0.776	0.839
技 2	28/36	28/30	29/30	0.848	0.862
技 3	30/36	30/31	31/31	0.896	0.909
技 4	28/36	28/354	339/354	0.144	0.858
技 5	29/36	29/1066	1053/1066	0.053	0.887
技 6	29/36	29/210	210/210	0.236	0.892
平均	0.787	0.497	0.985	0.492	0.875

表 2　データ拡張後閾値 0.955 の評価値

技名	再現率	適合率 1	適合率 2	F 値 1	F 値 2
技 1	28/36	28/35	35/35	0.789	0.875
技 2	33/36	33/41	39/41	0.857	0.934
技 3	33/36	33/38	38/38	0.892	0.957
技 4	33/36	33/499	452/499	0.123	0.911
技 5	32/36	32/1346	1322/1346	0.046	0.933
技 6	29/36	29/282	273/282	0.182	0.879
平均	0.870	0.444	0.968	0.482	0.915

ステージ 1 は背景が非常に複雑で大きく動き，ステージ 2 は背景が複雑だがあまり動かない，ステージ 3 は背景が単調で動かない特徴を持つ（図 5）．実験ではステージの違いによる評価値への影響は見られず（図 7），汎用的な性能を持たせることに成功したといえる．さらに，データ拡張前よりもデータ拡張後のデータの方が F 値 1，F 値 2 の最大値が高く，データ拡張の有効性を示した．

しかし，キャラクターの一部が隠れた場面やカメラが離れた場面で検出に失敗した．加えて，検出対象のフレーム画像と類似したフレーム画像を検出してしまう問題があり，類似したモーションが多い技 4～6 では検出回数が多くなっている（図 8）．これにより，実際の技の使用回数と検出回数の割合が大きく異なる問題がある（図 9）．

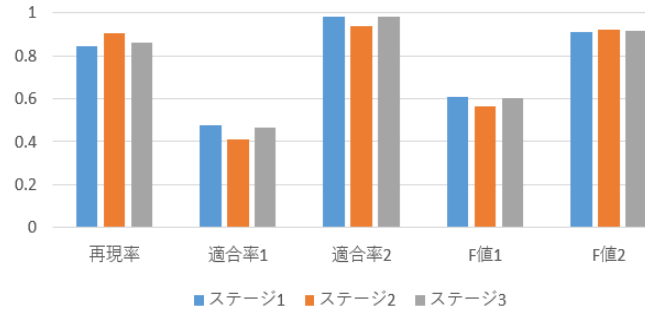

図 7　表 2 のステージ毎の評価値

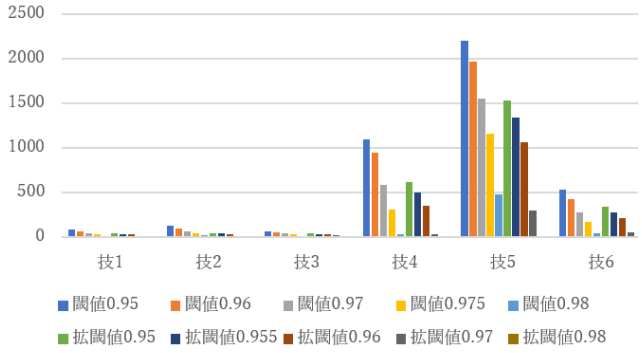

図 8　閾値毎の技の総検出数

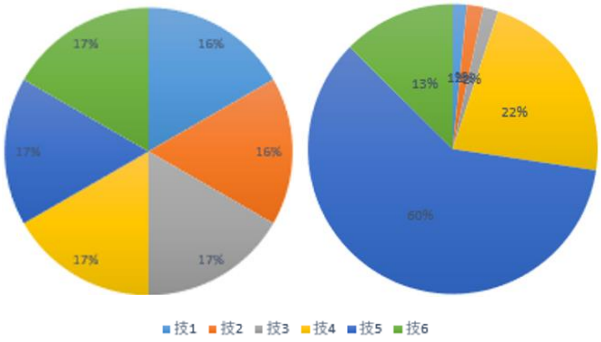

図 9　動画で技を使用した回数の割合（左）
YOLOv5 で技を検出した回数の割合（右）

6.　まとめと今後の課題

本研究では，YOLOv5 による物体検出がコンピュータゲームにおける特定のモーション検出に適用できることを示した．これにより，今まで手作業で行われていたプレイヤーの分析作業を効率化することができる．しかし，技を 1　回使用する毎に複数回検出を行う問題や，キャラクターを認識し辛い場面で検出に失敗する問題がある．そのため，実用化には精度の更なる向上と，連続した検出を防ぐプログラムの追加が必要である．また，本研究では実際の対戦動画のリプレイから手作業で学習画像を収集した．そのため，今後他の技や他のキャラクターへ対応するためには大量の画像データが必要となる点も課題である．

今後の展望としては，少量の画像からでも同様の精度を出すための効率的なデータセットの作成，学習するフレーム画像やデータ拡張の手法の調整による検出精度の更なる向上，他コンピュータゲームへの適用などが考えられる．

参考文献

[1] 大乱闘スマッシュブラザーズ SPECIAL, https://www.smashbros.com/ja_JP/,（参照 2023-10-14）.
[2] 三ツ井 慧太郎，岡部 誠：対戦アクションゲームにおけるプレイヤーの挙動観察のためのシーン検索，WISS, 2021（2021）.
[3] Q.Wang，L.Zhang，L.Bertinetto，W.Hu，P.H.S.Torr：Fast Online Object Tracking and Segmentation: A Unifying Approach，CVPR 2019（2019）.
[4] S.Wug Oh，J.-Y. Lee，N.Xu，S.Joo Kim：Video Object Segmentation using Space-Time Memory Networks”, ICCV 2019（2019）.
[5] 梶並 知記，長谷川 和也：キャラクタの位置情報に基づいた対戦型格闘ゲームの初心者向け観戦支援システム，情報処理学会論文誌 デジタルコンテンツ，Vol.6 ,No.1,pp 17–27 (Feb. 2018)（2018）.
[6] J.Redmon，S.Divvala，R.Girshick，A.Farhadi：You Only Look Once:Unified,Real-Time Object Detection，CVPR，2016（2016）.
[7] R.Girshick： Fast R-CNN，in IEEE International Conference on Computer Vision (ICCV) (2015).
[8] W.Liu，D.Anguelov，D. Erhan，C.Szegedy，S.Reed, C.Y.Fu，&A.C. Berg：SSD: Single shot multibox detector. In European conference on computer vision,pp. 21-37，(2016).
[9] Github：Ultralytics YOLOv5, 〈https://github.com/ultralytics/yolov5〉,（参照 2023-10-14）.
[10] robotflow：〈https://roboflow.com/〉,（参照 2023-10-14）.
[11] microsoft：VoTT,〈https://github.com/microsoft/VoTT〉,（参照 2023-10-14）.