YOLO Poseを用いた フィギュア専用ランドマーク推定モデルの構築

佐藤 一樹　湯村 翼 *
(北海道情報大学)

1　はじめに

拡張現実（Augmented Reality：AR）技術の普及により，現実空間にデジタル情報を重ね合わせることで，従来とは異なるコンテンツの閲覧や体験が可能となっている．特にスマートフォンなどのカメラを利用した ARでは，現実の物体の位置や姿勢を認識し，その位置に応じてデジタルコンテンツを表示することで，現実空間とデジタルコンテンツを組み合わせた表現が可能である．

本研究では，フィギュアを対象として，フィギュアの位置や向き，さらに角や羽，しっぽなどの特徴的な部位を認識し，その情報を利用して AR エフェクトや背景を表示するシステムの構築を目指す．フィギュアを単に鑑賞するだけでなく，AR を組み合わせることで，フィギュアの新たな楽しみ方を提供することを目的としている．研究全体としては，フィギュアを認識した位置や姿勢に応じて AR エフェクトを生成するシステムを想定している．

初期検討では，人体の姿勢推定に利用される MediaPipe を用いた認識を行った．MediaPipe の Pose Landmarker は画像中の姿勢ランドマークを検出する機能を提供している [1]．しかし，フィギュアの角を手として認識するなど，フィギュア特有の形状によって誤認識が発生する問題が確認された．また，画像が少し暗くなると認識できなくなる場合もあった．このことから，人体を対象とした既存の姿勢推定モデルをそのままフィギュアに適用するのではなく，フィギュア特有の形状を考慮したモデルが必要であると考えた．

そこで本研究では，物体検出と姿勢推定を同時に行うことができる YOLO Pose[2] を利用し，フィギュア専用のランドマーク推定モデルを構築する．YOLO の Pose タスクでは，対象物上のキーポイントをランドマークとして推定できる [3]．フィギュアの人体部分だけでなく，角やキャラクターの耳，羽，しっぽなどの特徴的な部位をランドマークとして追加することで，フィギュアに適した姿勢推定を行うことを目指す．

2　フィギュア専用ランドマーク推定システム

2.1　システム概要

本研究では，フィギュアの画像を収集し，画像中のフィギュアに対してランドマークを設定したデータセットを作成する．その後，作成したデータセットを用いて YOLO Pose の学習を行い，学習済みモデルを用いて新たな画像からフィギュアのランドマークを推定する．



Fig. 1: モデル構築の手順.

システムの構成を図 1 に示す．画像の収集後，オープンソースの画像アノテーションツール CVAT[4] を用いてフィギュアの各部位にランドマークを設定する．CVAT は画像に対する点やスケルトンなどのアノテーションに対応しており，本研究ではフィギュアの各部位を手動でアノテーションした [4]．作成したアノテーションデータを YOLO 形式のデータセットとして整理し，YOLO Pose による学習を行う．学習後は，入力画像に対してフィギュアの検出およびランドマーク推定を行う．

現段階では，YOLO Pose によるランドマーク推定結果を AR 表示へ接続するところまでは実装していない．推定したランドマークの位置情報は，将来的に AR エフェクトなどの表示位置として利用することを想定している．なお，AR 表現に関する先行的な試作として，MediaPipe Hands と p5.js を使用し，手の位置に炎のエフェクトを表示するプロトタイプを作成している．このプロトタイプと本研究の YOLO Pose モデルは，現時点では別々の実装である．

2.2　フィギュア用ランドマークの設計

YOLO Pose では対象物の姿勢を表すキーポイントを利用できる．しかし，フィギュアには人体には存在しない角や羽，しっぽなどの特徴的な部位が存在する．本研究では，まずラプラス・ダークネスを対象とした実験として，フィギュア固有の 4 点をランドマークとして設定した．

4 点は，左角の先端，右角の先端，しっぽの付け根，しっぽの先端で構成する．左右はキャラクター自身から見た左右とする．角については角の先端，しっぽについては身体から出ている部分をしっぽの付け根，終端部分をしっぽの先端として設定する．

4 点のランドマークの定義を図 2 に示す．

角については角の先端付近，キャラクターの耳や羽についてはそれぞれの特徴的な位置にランドマークを設定

* yumu@yumulab.org



図 2　フィギュア用 25 点ランドマーク定義（正面図）

Fig. 2: ラプラス・ダークネス用 4 点ランドマークの定義.

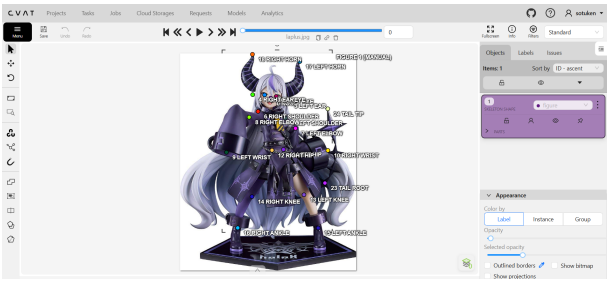


Fig. 3: CVAT を用いたフィギュア画像へのランドマークアノテーション.

する．しっぽについては，身体から出ている部分をしっぽの付け根，先端部分をしっぽの先端として設定する．左右はキャラクター自身から見た左右とする．また，フィギュアに存在しない部位や，画像上で確認できない部位については，位置を推測してランドマークを設定するのではなく，“outside” として扱う．特に身体の反対側に隠れている部位についても，見えていない場合には推測して設定しないようにした．

2.3　データセットの作成

ランドマークのアノテーションには CVAT を使用した．CVAT 上でフィギュア画像を読み込み，定義したランドマークを各部位に設定した．CVAT を用いたアノテーションの様子を図 3 に示す．

フィギュア画像の素材収集には，公開データセットを共有する Roboflow Universe の「Anime Figures Dataset」[5] を利用した．同データセットから本研究の対象や実験目的に合う画像素材を選択し，ラプラス・ダークネスのフィギュアを対象とするデータセットを作成した．収集した画像には，CVAT を用いて本研究で定義した 4 点のランドマークを手動でアノテーションした．ラプラス・ダークネスを対象として，14 枚の画像から学習を開始し，その後，異なる角度の画像を追加して合計 32 枚までデータセットを拡張した．32 枚のデータセットは学習用 25 枚，検証用 5 枚，テスト用 2 枚に分割した．各画像について，1 つのフィギュアのクラス，バウンディングボックス，4 点のランドマーク情報を YOLO Pose 形式で保存した．各ランドマークについては，画像上の x 座標，y 座標，および可視性情報を保持する．YOLO の Pose データセットでは，画像ごとに対応するラベルファイルを用意し，バウンディングボックスとキーポイントの座標を保存する形式が利用される [6].

Table 1: ラプラス・ダークネス 32 枚モデルの検証結果

評価指標	値
Box Precision	0.221
Box Recall	0.400
Box mAP@50	0.195
Box mAP@50–95	0.0354
Pose Precision	0.00781
Pose Recall	0.200
Pose mAP@50	0.0150
Pose mAP@50–95	0.0015

2.4　YOLO11n-Pose による学習

本研究では，Ultralytics YOLO11n-Pose をベースモデルとして使用した．YOLO11 では Pose/Keypoints タスクが提供されており，`yolo11n-pose.pt` などのモデルを利用できる [7]．通常の YOLO Pose では人体の 17 点を対象としているため，本研究のラプラス・ダークネス 4 点モデルでは，Pose Head の出力点数を 4 点に変更し，フィギュア固有のランドマークを出力できるようにした．

学習環境には Windows 11 Home，NVIDIA GeForce RTX 4050 Laptop GPU，Python 3.11.6 を使用した．また，Ultralytics 8.4.135，PyTorch 2.11.0+cu128 を使用した．ラプラス・ダークネスの学習では，画像サイズ 640，バッチサイズ 2，GPU を使用し，学習済みモデル `best.pt` を生成した．

2.5　推論結果

ラプラス・ダークネスの 14 枚モデルと 32 枚モデルによる推論結果を図 4 に示す．

ラプラス・ダークネスの 4 点モデルについて，14 枚モデルと 32 枚モデルを同じ正面画像に適用して比較した．両モデルともフィギュアのバウンディングボックスは検出したものの，角やしっぽのランドマークが正しい部位に配置されず，胴体付近に集中する傾向が確認された．この結果から，対象画像を増やすだけでは，フィギュア固有のランドマークを十分に推定できない場合があることが分かった．

3　評価

3.1　ラプラス・ダークネス 4 点モデルの評価

ラプラス・ダークネスを対象とした 32 枚モデルの検証では，検証画像 5 枚に対して評価を行った．結果を Table1 に示す．



(a) 14 枚モデルによる推論結果



(b) 32 枚モデルによる推論結果



(c) 同一正面画像による比較

Fig. 4: ラプラス・ダークネスに対する 4 点モデルの推論結果

この検証結果では，バウンディングボックスの検出性能およびランドマーク推定性能はいずれも十分とはいえない．特に Pose mAP@50 は 0.0150 であり，4 点のランドマークを正確な位置に推定する課題が残った．また，検証画像は 5 枚と少ないため，評価値は限定的な結果として解釈する必要がある．同一の正面画像を用いた 14 枚モデルと 32 枚モデルの定性的な比較でも，角としっぽの位置を正しく推定できず，明確な改善は確認できなかった．

以上より，フィギュア固有の部位のみを対象とした 4 点モデルでは，少数の画像による学習だけでは安定したランドマーク推定が難しいことが示唆された．ただし，この結果だけから 25 点モデルより 4 点モデルの方が性能が低いと結論づけることはできない．両者では対象ランドマーク，データセット，画像構成が異なるため，直接比較には同じ条件を用いた追加実験が必要である．

3.2 今後の評価

今後は，ラプラス・ダークネスについて，人体の基本姿勢情報を利用できる 25 点ランドマーク方式へ拡張する予定である．人体の 17 点とフィギュア特有の 8 点を設定し，既存の 4 点データをそのまま 25 点データとして扱うのではなく，追加のランドマークを CVAT 上でアノテーションする．自動推定を初期位置として利用する場合も，位置を確認・修正したデータを教師データとして用いる．

また，画像の向きや部位の見え方に偏りがないようにデータセットを確認し，学習用，検証用，テスト用の画像を分けて評価する．評価では，ランドマークの位置ずれ，誤検出，部位ごとの推定状況を確認し，必要に応じて画像数やアノテーションを見直す．

最終的には，YOLO Pose によって推定したランドマークの位置情報を AR システムと連携させ，フィギュアの特徴的な部位や姿勢に応じた AR エフェクトの表示を実現することを目指す．AR エフェクトの基本的な検討として，MediaPipe Hands と p5.js を使用し，フィギュアの手を認識した位置に炎のエフェクトを表示するプロトタイプを作成している．MediaPipe の Hand Landmarker は画像から手のランドマークを検出する機能を提供しており，p5.js は Web 上でインタラクティブな表現を作成できる JavaScript ライブラリである [8, 9]．ただし，このプロトタイプと YOLO Pose の 25 点モデルとの連携は今後の課題である．

4 おわりに

本研究では，フィギュアを対象とした専用ランドマーク推定モデルの構築を目的として，YOLO11n-Pose を用いた学習を行った．ラプラス・ダークネスを対象として，左右の角の先端，しっぽの付け根，しっぽの先端の 4 点をランドマークとして設定し，14 枚の画像から学習を開始した．その後，画像を追加して 32 枚までデータセットを拡張し，4 点モデルの学習と評価を行った．しかし，32 枚モデルの Pose mAP@50 は 0.0150 であり，同じ正面画像を用いた推論でも，角やしっぽを正しい位置に推定できなかった．この結果から，画像数を増やすことに加え，対象部位の見え方を考慮したデータ構成とアノテーションの確認が重要であると考えられる．

今後はラプラス・ダークネスに人体 17 点とフィギュア特有の 8 点を合わせた 25 点方式を適用し，CVAT によるアノテーションを整備した上で再学習を行う．さらに，学習用，検証用，テスト用のデータを分離して推定性能を評価し，最終的には AR システムとの連携を目指す．

参考文献

[1] Google AI for Developers: MediaPipe Pose Landmarker, https://ai.google.dev/edge/api/mediapipe/python/mp/tasks/vision/PoseLandmarker. 2026 年 9 月 23 日参照.

[2] Maji, D., Nagori, S., Mathew, M. and Poddar, D.: Yolo-pose: Enhancing yolo for multi person pose estimation using object keypoint similarity loss, *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, IEEE, pp. 2636–2645 (2022).

[3] Ultralytics: Pose Estimation with Ultralytics YOLO, https://docs.ultralytics.com/tasks/pose. 2026 年 9 月 23 日参照.

[4] CVAT: CVAT Documentation, https://docs.cvat.ai/. 2026 年 9 月 23 日参照.

[5] KMUSGROVE: Anime Figures Dataset, Roboflow Universe (2024). Accessed: 2026-10-10.

[6] Ultralytics: Pose Estimation Datasets, https://docs.ultralytics.com/datasets/pose. 2026 年 9 月 23 日参照.

[7] Ultralytics: Ultralytics YOLO11, https://docs.ultralytics.com/models/yolo11. 2026 年 9 月 23 日参照.

[8] Google AI for Developers: MediaPipe Hand Landmarker, https://ai.google.dev/edge/api/mediapipe/python/mp/tasks/vision/HandLandmarker. 2026 年 9 月 23 日参照.

[9] p5.js: p5.js, https://p5js.org/. 2026 年 9 月 23 日参照.