

# SPAR3D を用いた Image-to-3D によるデータ拡張を用いた物体検出 Object Detection Using Data Augmentation via Image-to-3D Conversion with SPAR3D

藤野 朋志<sup>†</sup>      大川 茂樹<sup>†</sup>  
Tomoshi Fujino      Shigeki Okawa

## 1. はじめに

昨今、機械学習やディープラーニングによる画像認識技術が発展しており、リアルタイムにおいても高い精度での物体検出が可能となっている。ディープラーニングによる物体検出では、データ数が多くなると精度が上がる傾向にあるが、高い精度を得るには数千から数万のデータが必要となる。通常の場合、十分なデータ数を集めるのは困難であるため、それに対処する手段としてデータ拡張があり、一般的な手法として、画像の回転、反転、拡大、縮小、ノイズ付与等がある[1]。

近年では、物体検出におけるデータ拡張の手法も多様化している。その中でも生成 AI である Stable Diffusion を使用しデータ拡張をおこない、物体検出での精度向上を目指した研究[2]がある。また、3D モデルを作成しモーションや背景合成を使用しデータ拡張をおこない、物体の検出精度向上を達成している[3]。

1 枚の入力画像から 3D モデルを生成する手法は技術的制約により実現されてこなかった。そのため 3D モデルを扱うデータ拡張の場合、3D モデルの自由度や再現性は低い傾向にあった。しかしながら SPAR3D[4]の登場により、多様な 3D モデルを生成することが可能になった。

生成 AI 等を使ったデータ拡張の場合、1 枚の入力画像から大量かつ多様な生成が得られることが多く、特徴に偏りが発生しやすい一般的な手法よりも高精度な拡張が可能である。我々は 3D モデルの生成によって本来不可視であるモデルの裏側をデータ拡張に利用することで、より物体検出の性能を高められると考えた。

本研究では生成 AI である SPAR3D[4]を活用し、1 枚の入力画像から 3D モデルを作成しデータ拡張をおこなうことによって、物体検出における精度の向上を目指すことを目的とする。また、検出難易度を考慮して検出対象を擬態生物であるカエルに限定した。

## 2. 方法

### 2.1 データセット

インターネット上から単体のカエルが映った画像を 200 枚収集し、これを基本データセットとし、学習に使用した。学習に使用しない Test データも別に 92 枚収集した。基本データセットには擬態の区別をせずカエルが映っている画像を収集し、Test データには擬態している状態のカエルを収集した。なお、収集した画像は著作権上問題ないと明記されたサイトから収集した。

### 2.2 SPAR3D[4]によるデータ拡張

SPAR3D (Stable Point Aware 3D)とは UIUC と Stability AI 社が共同研究によって開発した Image-to-3D の技術である。入力画像から Stable Diffusion の拡散モデルを物体の点群情報生成に特化させることによって、回帰モデルでは難しい曖昧度の高い裏面や見えない部分のモデル生成を可能にしている。

拡張方法は基本データセットを全て SPAR3D に入力し、カエルのモデルを 200 体分生成した。出力された 3D モデルを主観判断し、カエルと判断できるモデルのみを使用して拡張をおこなった。その際、モデルの判断基準として生成されたモデルの足の本数、胴体との接続位置、目の生成位置、テクスチャの崩れ具合を重視して判断をした。5 回ほど生成をおこない、カエルと判断できるモデルの頻度が高かったモデルを選定してデータ拡張に使用した。選定したモデルを Blender に入力し、モデルの傾きを変更せず 1 周回転する動画を作成し、フレームごとに分割してデータ拡張とした。Blender による拡張では動画フレーム数の問題から 1 動画から 120 枚の画像が得られる設定とし、背景画像を白単色背景のみで動画を作成した。

また、学習においては基本データセットの 200 枚に拡張データ 720 枚を加えて計 920 枚の学習データで学習をおこなった。

### 2.3 YOLOv8 による学習

物体検出の手法として YOLO (You Only Look Once)v8 を使用する。YOLO では CNN の出力層に入力するデータ数を工夫することによって高速かつ安定的な物体検出を可能としている。本研究ではバージョンの安定性と信頼性、推論の高速性を鑑みて YOLOv8 を使用した。

### 2.4 従来手法によるデータ拡張

SPAR3D におけるデータ拡張の有効性を調査するために従来手法である一般的な拡張手法との比較をおこなった。使用する手法として画像の回転、ガウシアンノイズ、メディアンフィルタを画像全体に付与したデータ拡張の 3 種類をおこなった。Fig. 1 にそれぞれの拡張手法の例を示す。拡張した画像の枚数は SPAR3D での拡張枚数と同じ枚数になるようランダムで抽出し拡張をおこなった。

<sup>†</sup> 千葉工業大学先進工学研究科未来ロボティクス専攻  
Chiba Institute of Technology

Table 1 物体検出時の学習データの比率と検出精度

学習データ	Train (枚)	Val (枚)	Test (枚)	Precision	Recall
拡張なし	140	60	20	0.329	0.400
SPAR3D による拡張 (提案手法)	644	276	92	0.936	0.350
回転による拡張	644	276	92	0.639	0.728
ガウシアンノイズによる拡張	644	276	92	0.719	0.684
メディアンフィルタによる拡張	644	276	92	0.704	0.717



Fig. 1 一般的な拡張手法 (左: 回転, 中: ガウシアンノイズ, 右: メディアンフィルタ)

### 3. 結果

#### 3.1.1 SPAR3D によるデータ拡張結果

Fig. 2 は入力された画像と生成されたモデルの画像例である。SPAR3D の出力は 5 回おこなったが、出力されたモデルの中で学習に使用できると判断したモデルは平均 15~20 モデルであり、高頻度のものは 12 モデルであった。他の拡張との兼ね合いのため 12 モデルの中から半分をランダムに抽出しデータ拡張に使用した。



Fig. 2 SPAR3D によるモデルの出力例 (左: 入力画像 中: 入力方向での 3D モデル, 右: 3D モデル上方)

#### 3.1.2 YOLOv8 による物体検出

学習データ全体を 7:3 の比率で Train データと Val データに分割して学習をおこなった。学習データの総量に対して 1 割になるように Test データ数を調整した。

学習した重みを用いて Test データ内の画像からカエルの検出をおこなった。モデルの評価方法として Precision と Recall を用いて評価をおこなった。Precision はモデルの正確さを、Recall はモデルの再現性を示す。

Table 1 は YOLOv8 で学習をおこなった項目と枚数の内訳と精度を示している。SPAR3D を用いた拡張の場合、一般的な拡張手法と比べて Precision が向上している結果が得

られた。反対に Recall の値が拡張なしの状態よりも低下したという結果が得られた。

### 4. 考察

Precision が向上した理由として、Blender でモデルの回転動画を得る際に背景を白単体の画像にしたことでモデル自体の輪郭や正確性が学習しやすくなったことに加えて、明確な表現が足りない特徴量に対して補う画像を供給できたことが考えられる。一方で、Test データに使用した画像には擬態している画像が多く含まれることや、背景情報が多彩な画像が多かったことにより、白背景中心で構成された学習データとの分布乖離が起きたため、Recall が低下したと考えられる。

### 5. おわりに

本研究では SPAR3D を用いた Image-to-3D によるデータ拡張による物体検出をおこなった。その結果 SPAR3D によるデータ拡張において有効性が示唆された。

今後の展望として生成されたモデルに骨格付与やアニメーションの付与をおこない拡張可能枚数を増やすことや、背景合成を通して擬態している画像を疑似的に生成することで擬態生物の検出精度をさらに向上させることが考えられる。

### 参考文献

- [1] 国立研究開発法人産業技術総合研究所, “機械学習品質評価・工場技術に関する報告書—機械学習品質マネジメントガイドライン付属文書 第 2 版—”, テクニカルレポート DigiARC-TR-2022-06a/CPSEC-TR-2022007, (2022 年 8 月).
- [2] Brandon Trabucco, Kyle Doherty, Max Gurinas, Ruslan Salakhutdinov, “Effective Data Augmentation With Diffusion Models”, arXiv:2302.07944, (2023).
- [3] 北風 裕教, 吉原 蓮人, 岡部 蒼太, 松村 遼, “オブジェクト検出 YOLO を用いた害鳥認識システムの開発”, 産業応用工学会論文誌, Vol. 8, No. 1, pp. 10-16 (Mar. 2020).
- [4] Zixuan Huang, Mark Boss, Aaryaman Vasishtha, James M. Rehg, Varun Jampani, “SPAR3D: Stable Point-Aware Reconstruction of 3D Objects from Single Images”, arXiv:2501.04689, (2023).