

## 3D モデルを用いた Data Augmentation におけるモデルの高品質化が検出精度へ与える影響 Effect of Model Quality Improvement on Detection Accuracy in Data Augmentation using 3D Models

末岡 祐馬<sup>†</sup> 北風裕教<sup>†</sup> 松村遼<sup>†</sup>  
Yuuma Sueoka Hironori Kitakaze Ryo Matsumura

### 1. はじめに

我々は害獣検出を用いた害獣忌避システムの開発に取り組んでいる。検出には深層学習を用いているが、その学習データを効率よく収集する手段に限られている。この問題に対して害獣の 3D モデルを作成し、これを学習データの生成に用いる手法が提案されている[1]。文献[1]における害獣の 3D モデルは Smoothie-3D [2] という Web サービスを利用して作成しているが、頂点の配置やトポロジーの流れが歪であるため(Fig. 1)、これが学習に悪影響を及ぼしている可能性がある。また、学習用画像の生成には、3D モデルをレンダリングしたアルファ付き画像と実写背景画像を組み合わせて生成しているため、背景画像の画角に対してレンダリング画像の画角が不自然であることがあった。

そこで、本研究では高精度な害獣や背景オブジェクトの 3D モデルを作成してレンダリングを行い、学習用データセットを生成することを目的とする。また、害獣モデルにアニメーションを付与したり、回転や拡大縮小を行うことで、大量の学習用画像を生成することを目指す。

本稿では、イノシシの 3D モデルを作成し、これと Smoothie-3D を用いて作成したイノシシのモデルで生成した学習用画像、それぞれで学習した場合の害獣検出性能を比較し、モデルの高品質化が検出精度へ与える影響の検証を行った。

### 2. 害獣 3D モデルの作成および学習用画像生成

#### 2.1 モデル作成

本稿での検出対象はイノシシであるため、実際のイノシシの画像を参考に Blender を用いて 3D モデルを作成した。イノシシの形状に沿って平面ワイヤーフレームを作成、それにソリッド化処理を施して、立体とした。次にスカルプトツールで外形を整え、蹄や耳などのパーツを追加した。

作成したイノシシの 3D モデルは、Smoothie-3D で作成したものとは比べ、膨大な頂点数を持っているため、レンダリング時の演算時間も膨大なものになる。そこで、モデルデ

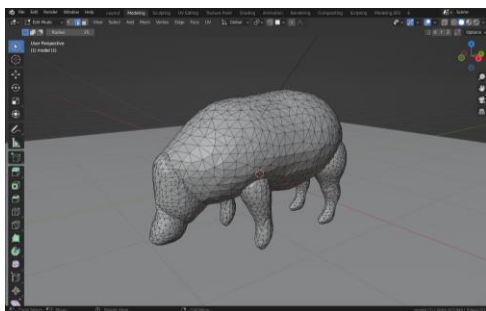


Fig. 1: Smoothie-3D で作成された 3D モデル

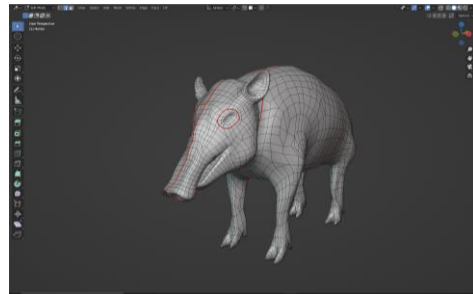


Fig. 2: リメッシュ後の 3D モデル

Table 1: 各 3D モデルのデータ

	Smoothie-3D	リメッシュ前モデル	リメッシュ後モデル
頂点数	3,944	35,666	3,666
辺数	11,771	71,322	7,316
面数	7,848	35,660	3,648
三角面数	7848	71,228	7,314
Frame あたりのレンダリング時間	00:06.42	00:07.05	00:06.01

ータの軽量化のためリメッシュを行い、頂点数を減らした後にトポロジーの調整を行った(Fig. 2)。その結果、元モデルの頂点数の約 9.7%まで削減することができた (Table 1)。

#### 2.2 レンダリング

それぞれのモデルに対して、カメラを対象の周りで周回させ 3.6 度ごとに 1 枚の画像をレンダリングさせた。また、レンダリングの設定は Cycles で画素値を決めるレイの最大本数を 128、レイがオブジェクトに反射する最大数を 12 回、レンダリング画像の画質を 800×800 とした。

##### 2.2.1 Cycles レンダー

Blender には標準レンダリングエンジンとして Cycles と Eevee が存在する。前者はパストレーシング法を使用しており演算速度は遅いが現実に近い画像を取得できる。後者はラスタライズ法を使用しており、演算速度は速いが AO(Ambient Occlusion)の精度や透過、屈折などの表現に欠陥がある。Eevee の欠陥はコンポジット等を用いると解決するが、それには複雑なノードを構築しなければならない。

よって本稿ではライトパスバウンズ数やサンプリング数、光源の光量を設定するだけで写実性の高いレンダリング画像の出力が可能な Cycles エンジンを使用した。



(a) データセット A (b) データセット B (c) データセット C

Fig. 3: 学習画像 (イノシシ) の例

### 2.3 レンダリング時における円運動カメラ

これまで、アルファ付きレンダリング画像の出力は、Unity のシーン上のカメラで対象のモデルを中心に、一定の距離で円周運動をさせながらキャプチャすることで行っていた。本稿ではモデルに対してレイトレーシングを行い、レンダリングするため、画像出力までを Blender 内で完結させる必要がある。また、画角の変化による出力画像の多様性を実現させるため、NURBS 曲線をパスとしたカメラトラッキングシステムを作成した。NURBS 曲線は 3D モデルであるため、ノーコードで作成でき、またカメラの位置や対象のモデルからの距離などを直感的に変更できる。

## 3. 比較検証実験

### 3.1 実験条件

本実験では 3D モデル品質の差による検出精度比較を行い、その検出性能への影響を調査する。学習用画像の生成に用いたアルファ付きレンダリング画像は、リメッシュ後モデルをレンダリングしたもの、これの明度を下げたもの、および Smoothie-3D で作成されたモデルをレンダリングしたものをそれぞれ 72 枚使用した。そして、これらの画像を実写背景画像と合成し、それぞれ 10,000 枚の画像を生成して学習用データセットとした。これら 3 つのデータセットをそれぞれデータセット A,B,C とする。生成した学習用画像の例を Fig. 3 に示す。学習用データセットは生成したイノシシ画像とその他の画像の 2 クラスで構成される。また、評価用のテスト画像データセットのイノシシは実写画像であり、学習用データセットと同様に 2 クラスで構成される。

学習には ImageNet で学習済みの Alexnet を用いて、転移学習を行った。学習条件は epoch 数を 10、学習係数を 0.00001、最適化手法にモーメンタム SGD を用い、モーメンタムを 0.9、バッチサイズを 16 に設定した。

### 3.2 実験結果

実験結果を Table 2 に示す。Fig. 4 には、それぞれの学習用データセットで学習したモデルの Grad-CAM [3]出力を示す。ここで、データセット A で学習したモデルを学習済みモデル A、B で学習したモデルを学習済みモデル B、C で学習したモデルを学習済みモデル C とする。

Table 2 より 学習済みモデル A と C を比較すると、A のほうが検出率が高く、3D モデルの品質向上は精度向上につながると思われる。また、学習済みモデル A と B を比較すると、明るさ要素は検出精度に大きな影響を与えることがわかる。

Table 2: 各学習済みモデルの検出率

	学習済み モデル A	学習済み モデル B	学習済み モデル C
検出率	95.72 %	84.04 %	95.55 %



(a) 入力画像



(b) 学習済みモデル A



(c) 学習済みモデル B



(d) 学習済みモデル C

Fig. 4: Grad-CAM 出力

また、Fig. 4 より (b)と(c)を比較すると、頭部と脚部は類似した注目を見せているが、(c)では胴体部の体毛色の淡い部分に反応がほぼ無いため、明度変更後のモデルのテクスチャは検出にはあまり適していないと推測できる。(b)と(d)を比較すると、(b)は(d)より頭部先端と脚部に強い注目を示しているため、3D モデルにおいて起伏が細かいパーツの品質が、検出精度に大きく影響すると思われる。

## 4. まとめ

本稿では、3D モデルの高精度化による検出性能への影響について調査を行った。実験結果より、3D モデルの品質の差異が検出性能へ影響を及ぼすことが確認できた。また、同じ 3D モデルでもテクスチャデータの差異で性能に影響が出ることも確認できた。今後の課題として、テクスチャの改良を行うことが挙げられる。また、背景の 3D データを作成し、3D モデルだけで生成した学習用画像と実写画像のみで作成した学習用画像、およびそれらを互い違いに合成した画像で、検出性能にどのような影響を及ぼすか比較実験を行うことも課題となる。

### 参考文献

- [1] Naoya Ryoke, Hironori Kitakaze, and Ryo Matsumura, "Data Augmentation with 3DCG Models for Nuisance Wildlife Detection using a Convolutional Neural Network," Proceedings of the 8th IIAE International Conference on Intelligent Systems and Image Processing 2021, pp.179-185, 2021.
- [2] Smoothie-3D, [https://smoothie-3d.com/site/page\\_index.php](https://smoothie-3d.com/site/page_index.php) (2022 年 6 月 23 日確認)
- [3] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D., "Grad-cam: Visual explanations from deep networks via gradient-based localization," International Journal of Computer Vision, 128, pp.336-359, 2020.