

PSPNet の質感に対するクラス分類精度の評価と考察

岡村一矢*¹ 松村遼*² 北風裕教*²

Evaluation and discussion of classification accuracy for texture of PSPNet

Kazuya OKAMURA, Ryo MATSUMURA and Hironori KITAKAZE

Abstract

In recent years, damage to fisheries caused by *Phalacrocorax carbo* occurs throughout Japan. Our research group have been developing system for driving away verminous birds by using drone. In previous research, a method that PSPNet and original classifier is proposed. But there are problems such as the inability to cope with an increase in number of target objects. Therefore, in this paper, confirmation that whether PSPNet can classify with high accuracy and discuss classification accuracy for texture. As a result, it could be confirmed that classifying with high accuracy in 4 classes airplane, car, crow, and boar.

Keywords: Deep Learning, Pyramid Scene Parsing Network, Semantic Segmentation

1 まえがき

カワウによる漁業被害は日本各地で発生している。主に河川や湖沼等の内水面や養殖場においてアユやヤマメ、コイなどを大量に捕食するため、その対策は急務である。全国内水面漁業協同組合連合会が発表しているカワウによる捕食被害の推定額¹⁾は、平成 29 年で約 55 億円となっており、漁師をはじめ漁業関係者を苦しめている。この問題解決のために昨今では、カワウ対策として水産庁等の機関がドローンを活用した取り組みをはじめている²⁾。しかし、ドローンの操縦には飛行条件を満たした機体と操縦者が必要であり、長時間の対応は操縦者に大きな負担を与える。さらに、ドローンのバッテリー駆動時間の限界から長時間にわたる運用が困難である。

我々の研究グループでは、小水力発電システムから供給される電力でドローンを自動飛行させ、音や光で害鳥を追い払うシステムの開発を試みている³⁾。この核となる害鳥の識別器は、ドローンに設置のカメラ映像を深層学習で学習することで構築する。この技術により、ドローンの無人制御が可能となり、操縦者確保の問題解決と 24 時間の継続的な対策を実施することができる。ここで、ドローンの自動制

御を実現するためには、カメラ映像から対象となるカワウを高精度に認識できなければならない。著者らのこれまでの研究において、『カワウ』、『カラス』、『カモ』を対象として、深層学習を用いた物体検出アルゴリズム You Only Look Once v3 (YOLOv3) を用いた手法と、ピクセル単位でクラス分類を行う画像セグメンテーションアルゴリズム Pyramid Scene Parsing Network (PSPNet) に独自の分類器を用いた 2 段階の手法の比較実験を試みた結果、PSPNet と独自の分類器を用いた手法⁴⁾の方が優れた認識結果が得られることを確認している。しかし、この手法で用いている独自の分類器は、HSV 色空間に各鳥の特徴をもとにしきい値を設けて分類を行っており、対象とする物体が酷似している場合や対象の数が増加した場合に対応できない問題が課題となっていた。この問題を解決することができれば、クラス数の増加にも対応できると考えられる。

そこで本研究では、複数クラスの分類において PSPNet が高精度に多クラス分類を可能にするか確認を行い、画像の質感の精度が向上した場合において、認識精度も向上するのかについて検討を行う。PSPNet の多クラス分類においては、①飛行機、②カ

* 1 電子・情報システム工学専攻 * 2 情報工学科



図1 学習対象の物体

ラス、③車、④イノシシの4種と⑤背景とする。これらの評価は物体の重なり割合から算出される評価指標のIoU (*Intersection over Union*) を用いる。ここで、④イノシシの3DCGモデルを準備し、その質感の有無に対して、PSPNetにおける多クラス分類実験を行った結果、『質感なし』のイノシシにおける多クラス分類の結果に比べて、『質感あり』のイノシシの多クラス分類の結果が各IoUにおいて認識精度が向上し、mIoU (*mean IoU*) においても認識精度の向上が見られたので考察と合わせて報告する。

2 学習用データセットの準備

本実験では、認識対象の形状が極端に近い害鳥の識別精度について検討するのではなく、基礎研究として特徴が明らかに異なる対象物体間のクラス分類について明確にした状態で、質感の有無における他クラス分類による認識精度の確認を行う。

対象物体は、空を飛ぶ物体である①飛行機と②カラス、地上に存在する物体である③車と④イノシシとする(図1)。③車においては、深層学習用に準備されたデータベースから車に関する実写画像を大量に準備することができるため、それらを用いる。しかし、①飛行機、②カラス、④イノシシに関しては深層学習用の学習データを大量に揃えることが難しいため、3DモデルからData Augmentationによって大量に生成した画像を準備する。ここで、取得した④イノシシは、体を覆う体毛マテリアルを追加することで質感を表し、『質感なし』の画像と『質感あり』の画像を準備する。詳細について以下に示す。

2.1 ラベル画像の取得について

PSPNetのような画像セグメンテーションタスクでは、RGBのカラー画像に加え、カラー画像と同じ解像度を持ち且つ対応した各ピクセルがどのクラスに属するかを示すラベル画像の2種類の画像データ



図2 カラー画像(左)とラベル画像(右)

表1 ラベル画像の対応情報

クラス	インデックス番号	RGB情報
背景	0	(0, 0, 0)
飛行機	1	(128, 0, 128)
車	2	(220, 128, 0)
カラス	3	(0, 128, 128)
イノシシ	4	(0, 200, 0)

が必要となる(図2)。図2のラベル画像は、カラーパレット形式でインデックス番号とRGB情報が対応している(表1)。通常、このラベル画像はカラー画像をもとにアノテーションツールを用いて手動で行うが、本実験では以降説明するData Augmentationおよび学習済みモデルの利用により取得する。

2.2 飛行機、カラス、イノシシの画像準備

2.2.1 従来のData Augmentation

①飛行機、②カラス、③イノシシの学習用データは3Dモデルを用いたData Augmentationによりデータセットを準備する。我々の研究グループでは独自のData Augmentation手法³⁾を提案しており、対象物体の2D画像から3Dモデルを作成し、3Dリアルタイム開発プラットフォーム(Unity)を用いて様々な方向から2D画像をVR空間上で再取得する手法を実現している。この独自のData Augmentation手法は、物体検出タスクを想定したアノテーションにおいて実現しており、この仕組みを基礎とした手法で画像セグメンテーションに対応したData Augmentationを実現していた。したがって3Dモデルを撮影する際、3Dモデル以外の領域を白色に設定して一度保存した後、各ピクセルをチェックし、白色のピクセルは背景であるとみなし事前に用意した背景画像のピクセルと置き換えることでカラー画像を生成する。その際、対応するインデックス番号を設定したラベル画像も同時に生成する。しかし、この方法では保存時のJPEG形式での不可逆圧縮の処理により、3Dモデルと白背景の境界が曖昧になるため、白い輪郭が残る問題に加え3Dモデルに白色が含まれた場合に背景としてみなされてしまう可能性があり、ピクセ



図3 ピクセルチェック

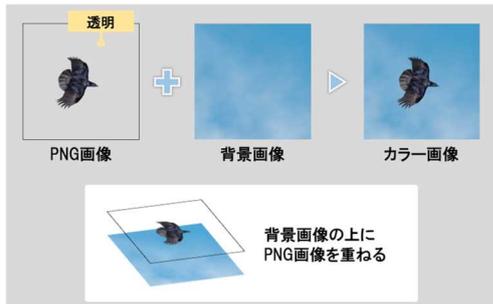


図4 カラー画像の合成手法

ル単位でクラス分類を行う画像セグメンテーションタスクにおいては認識精度に大きくかわる問題である。

2. 2. 2 改良後の Data Augmentation

先行研究で用いた Data Augmentation プログラムを改良し、より精密な学習用データの生成を行う。Unity の撮影処理では通常カメラに映るすべてのオブジェクトが撮影されるが、保存前にビューポート情報をテクスチャとして取得し、PNG 形式にエンコードする処理を加えることで特定の 3D モデル以外を透明（アルファ値=255）とした PNG 形式で保存することができる。これにより、ピクセルチェックを行う際に RGB 値に依存せず 3D モデル（アルファ値≠255）と背景領域（アルファ値=255）を見分けることができ、正確なラベル画像が生成可能となる（図3）。さらに、背景画像の上に PNG 画像を重ねて合成することでカラー画像が生成可能である（図4）。

しかし、この手法で作成可能な学習用データは、撮影時に必ず 3D モデルが中心に存在するため物体位置の変化パターンが乏しい。これを解決するために合成処理を施す際、重ねる PNG 画像のアンカーを変化させランダムな位置に物体位置を変更する。通常、PNG 画像と背景画像の座標 (0, 0) の位置、すなわち画像の左上同士をアンカーとして合成する。PNG 画像におけるアンカーを X 座標 Y 座標それぞれ



図5 物体位置のランダム化

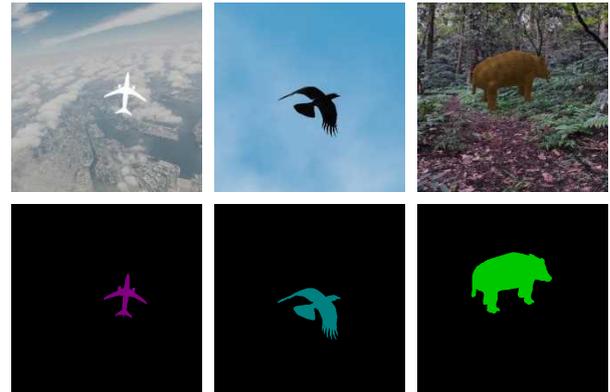


図6 3D モデルにより作成された学習用データ

れランダムに決定した数値だけ移動し、同じ画像サイズで切り取りを行うことで物体位置の変化を実現できる（図5）。これらの Data Augmentation を行うことで本実験用として、①飛行機と②カラスは各 4,080 枚、④イノシシは 2,880 枚（質感あり）、2,880 枚（質感なし）の学習用データを取得する（図6）。

2. 3 車画像の準備

2. 3. 1 実写画像の取得

③車は形状や色が異なる車種が多様多様に存在するため、2.2.2 の 3D モデルを用いて Data Augmentation を行う手法は、大量の 3D モデルやマテリアルが必要となり非効率である。一方、様々な機関からデータセットとして車の実写画像が公開されているため、それらの画像を利用して学習用データを準備する方が効率的である。よって本実験では、スタンフォード大学が研究目的に公開する Cars Dataset⁹⁾ の実写画像を利用する。データセット内の画像は正方形にトリミングし、473×473 にリサイズする。その際、トリミング領域に車全体が納まらない場合は、画像領域以外を白ピクセルで正方形に補完する。画像サイズが小さい 300×300 以下の画像は除外し、これにより 4,271 枚の実写画像を取得する。

2. 3. 2 学習済みモデルによるラベル画像生成

2.3.1 で得られた実写画像は、イメージ内の全ピク

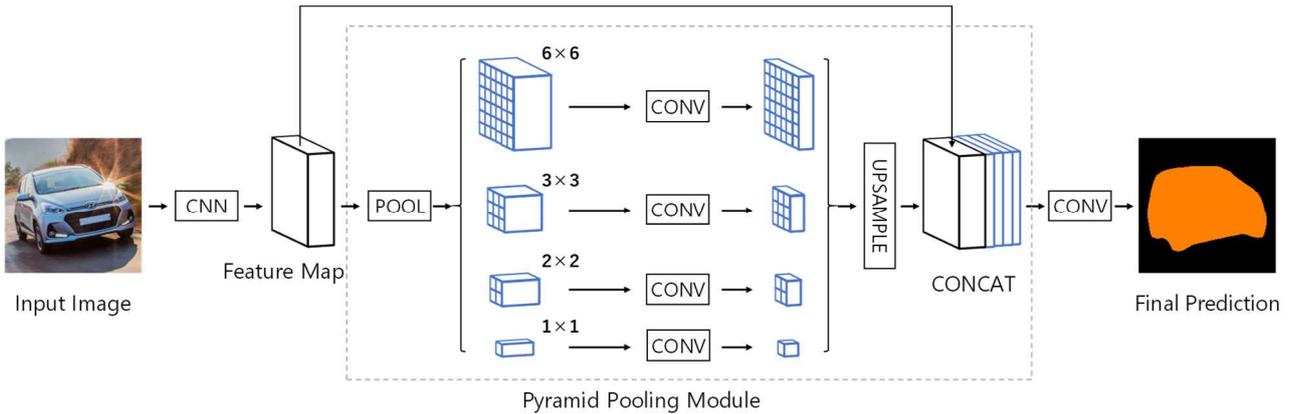


図7 PSPNet のネットワーク構造

表2 ハイパーパラメータ

パラメータ名	値
入力サイズ	473 × 473
エポック数	14
バッチサイズ	16
初期学習率	0.01
Momentum	0.9

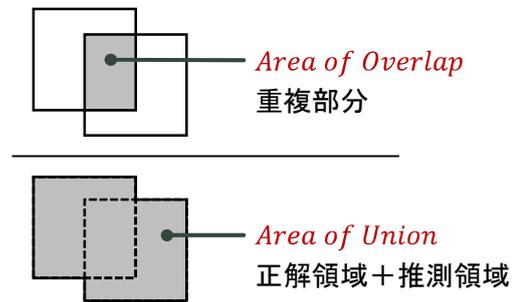


図8 評価指標 IoU

表3 認識実験の結果 (14 epoch)

クラス	質感なし	質感あり
背景	0.936	0.995
飛行機	0.669	0.726
車	0.941	0.953
カラス	0.841	0.919
イノシシ	0.543	0.871
mIoU	0.749	0.869

セルに対して『車』や『背景』などの意味のあるラベルを割り当てるセマンティックイメージセグメンテーションを行う。この処理には、車を事前に学習した学習済みモデル DeepLab v3⁶⁾に投入しその出力画像をラベル画像として利用する。DeepLab v3 は、PASCAL VOC2012⁷⁾データセットを学習したモデルであり、人や動物、乗り物など 20 クラスの物体を分類可能な特徴をもつ。ただし、出力されたラベル画像の中で、正確な形状を認識できない画像も一部存在するため、これらについては目視でデータを確認して事前に除外する。これにより、2,929 枚の学習用データを取得する。

3 PSPNet による学習

本実験では、①飛行機、②車、③カラス、④イノ

シシの 4 クラスを対象に、イノシシの『質感あり』と『質感なし』それぞれのデータセットを用いて PSPNet による学習および認識実験を行う。学習時に設定した主要なハイパーパラメータを表 2 に示す。

PSPNet は図 7 に示すネットワーク構造により実現されている。3 つのモジュールのうち特に、Pyramid Pooling Module では大きさの異なる特徴マップを生成、畳み込み、結合を行うことで、画像中の対局的な情報を取得でき、対象物周りの領域情報（コンテキスト）を利用することが可能となっている。

4 評価方法

PSPNet による学習を行った後、テスト用に各クラス 20 枚の実写画像を用いて認識率の評価を行う。テスト用画像は、出力結果（推測領域）と物体の正解領域を照らし合わせるために事前に対応するラベル画像を作成しておく。本実験では評価の指標に (1) 式の IoU (図 8) および (2) 式の mIoU を用いる。

$$IoU = \frac{\text{Area of Overlap}}{\text{Area of Union}} \quad (1)$$

$$mIoU = \frac{1}{n} \sum_{i=1}^n IoU_i \quad (2)$$

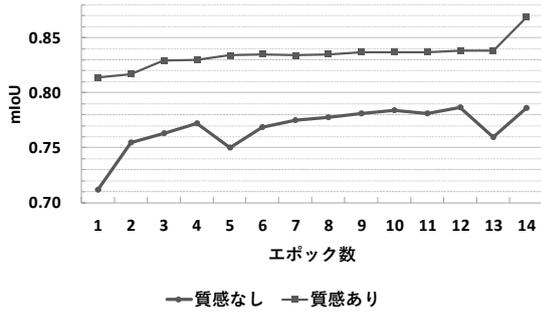


図9 認識率の推移

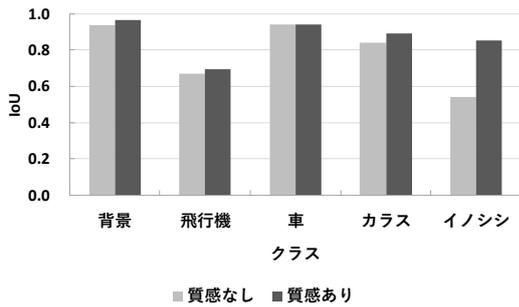


図10 各クラス IoU の比較結果 (14 epoch)

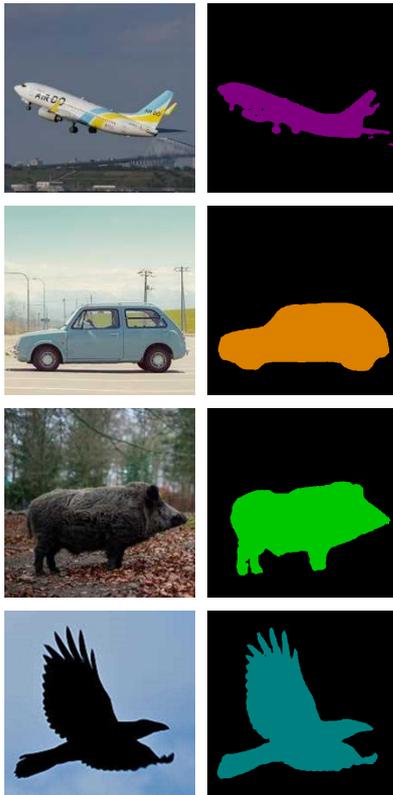


図11 質感あり学習モデルの出力結果例

5 実験結果と考察

『質感あり』、『質感なし』それぞれを学習させたモデルに対して、テスト用画像の認識実験を行った。クラスごとの mIoU を表3に示す。また、エポック

ごとの認識率の推移を図9、14エポックにおける各クラス IoU の比較結果を図10に示す。

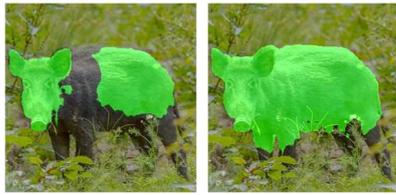
表3から、両学習モデルにおいて各クラスの IoU は、0.5以上を示しており PSPNetが複数クラスの分類が可能であることが確認できた。『質感あり』学習モデルでは、各クラスの IoU が 0.7以上を示していること、さらに図11の出力結果例から高精度にクラス分類ができていていると考えられる。

図9から、各学習用データセットにおける mIoU は学習過程の全エポックで mIoU が 0.5以上の差が生じており、質感の有無によって認識精度が大きく異なることが明らかである。特に14エポック時には、『質感あり』学習モデルの 0.869 に対し『質感なし』学習モデルは 0.749 を示しており、質感を有した方が高いクラス分類精度を得ることができた。また、クラス IoU の比較では、イノシシの IoU が質感を与えることで大幅に向上していることから、質感が大きく影響していることが明らかであり、イノシシ以外のクラスにおいても IoU の向上が見られる(図10)。

各学習モデルにおける出力結果画像を図12、図13に示す。『質感なし』学習モデルでは認識できていなかった領域において、『質感あり』学習モデルではイノシシとして正確に認識できている(図12)。さらに背景の模様をイノシシとして誤認識されていた領域においても改善されている(図13)。これらより、毛の質感によって生じた光の反射の変化や毛並みを特徴として学習され、精度が大きく向上したと考えられる。しかし、図12からイノシシの足が草葉に隠れている領域を認識できておらず、本実験で用いた Data Augmentation 手法では対象物体の前面に物体が存在しないことが原因となり認識できなかったと考えられる。そのため、体の一部が隠れてしまう場合など、生息場所の状況を考慮する必要があると考えられる。

また、誤認識が減少したことにより、これまでイノシシと認識されていた領域が正解クラスの領域として認識されたことで、イノシシ以外の他クラスにも IoU の向上が見られたと推測される。

『質感あり』学習モデルでは、飛行機の IoU が最も低い結果となっている。この要因として、飛行機は図14に示すクラスと誤認識する例が多く見られ、飛行機の胴体部分は正しく認識できているが、影で黒くなった羽や尾翼の部分はクラスと誤認識された。これは、空を飛ぶ黒い物体であるクラスと認識されたと考えられる。しかし、図15に示す黒い飛行機での結果では、部分的に領域が欠落しているが飛行



	質感なし	質感あり
IoU	0.468	0.795



	質感なし	質感あり
IoU	0.274	0.812

図 1 2 イノシシの出力結果の比較

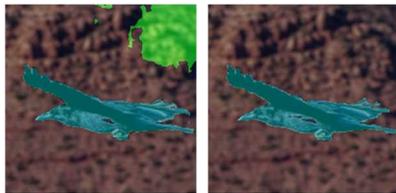


図 1 3 背景の誤認識改善例
(左：質感なし、右：質感あり)



図 1 4 カラスと誤認識された飛行機



図 1 5 黒い飛行機の認識例

機と認識されている。このため、色と形状が似ている場合においても飛行機の胴体とカラスの首の長さが異なるといった細部の輪郭や羽毛の質感を見て判断されていると推測できる。

本実験での対象物体 4 クラスは、地上と空中に存在する物体で分け、色や形状が酷似していないことから比較的分類が容易であると推測される。これに対し、鳥類の分類を行う際には、形状が酷似してお

り高い精度を得ることは困難であると予想される。しかし、前述したイノシシの質感および黒い飛行機の結果(図 1 5) から PSPNet は色や形状だけではなく、細部の輪郭や質感においても目が注目している。よって、これらの特徴を詳細に捉えることが、識別精度を向上させる重要な鍵となると考えられる。したがって、くちばしの色や首の長さなど局所的な部位の特徴や羽の光沢など、より詳細な情報をもたせることで、鳥類を対象とした識別器の実現が可能となると考えられる。

6 あとがき

本研究は、独自に開発しているドローンを用いた害鳥追い払いシステムに搭載する識別器において、先行研究で用いられた PSPNet が高精度に多クラス分類が可能であるか確認を行うこと、および画像の質感の精度が向上した場合における認識精度の向上について検討を行うことを目的とした調査研究である。実験では対象物体を、空を飛ぶ物体である飛行機とカラス、地上に存在する物体である車とイノシシの 4 クラスに定め、イノシシにおいては 3D モデルに質感の有無を与えたデータセットを用意し、それぞれ学習を行い、認識実験を行った。学習用データセットの作成には、3D モデルを用いた従来の Data Augmentation 手法を改善した高精度な学習用データ、および学習済みモデルによる出力画像を用いた。

実験結果は PSPNet が複数クラスの物体を高い精度で分類可能であると確認できた。また、イノシシの質感の有無に対して多クラス分類を行った結果、『質感なし』の結果に比べ、『質感あり』の結果が各 IoU において認識精度が向上し、mIoU においても大幅に認識精度が向上していることが確認できた。この要因から、毛の質感によって生じた光の反射の変化や毛並みの特徴として学習され、精度の向上に大きな影響を与えたと考察できた。さらに、飛行機の影になった部分をカラスと誤認識される例が見られたが、黒い飛行機においては飛行機として認識することができており、色や形状が似ている場合においても細部の輪郭や質感の情報が識別に関係していると考えられた。鳥類の分類を行うには、形状が酷似しており、本実験での対象物体より分類が困難であると予想されるが、色や形状だけではなく細部の輪郭や質感などの特徴を詳細に捉えることで識別器の実現が可能になると結論付けた。

今後は、害鳥追い払いシステムの実システム化に向け、鳥類の多クラス分類が可能な識別器の実現を目指す。

参考文献

- 1)<http://www.naisuimen.or.jp/jigyoku/kawau/kawaur1-12.pdf>
(2022.9.18 現在)
- 2)<https://www.jfa.maff.go.jp/j/enoki/attach/pdf/naisuimeninfo-23.pdf> (2022.9.18 現在)
- 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.q10-16 (2020)
- 4)飯田紘崇: 小水力発電システムと連携するドローン搭載型野鳥認識システムの Data Augmentation による深層学習認識率向上の検討
- 5)http://ai.stanford.edu/~jkruse/cars/car_dataset.html
(2022.9.24 現在)
- 6)Liang-Chieh Chen, George Papandreou, Florian Schroff, Hartwig Adam: Rethinking Atrous Convolution for Semantic Image Segmentation, arXiv preprint arXiv:1706.05587 (2017)
- 7)<http://host.robots.ox.ac.uk/pascal/VOC/voc2012> (2022.9.28 現在)