

IB Focal 損失に基づく教師なし領域分割手法 PiCIE の精度向上 Improvement Method for PiCIE Based on IB Focal Loss

林 正隆[†] 大野 将樹[‡] 獅々堀 正幹[‡]
Masataka Hayashi Masaki Oono Masami Shishibori

1. はじめに

領域分割 (Semantic Segmentation) とは、画像をピクセル単位でクラス分類する技術である。教師あり学習手法では、ピクセル単位で正確なクラスラベルを手手でアノテーションする必要があるため、作成コストが高い。それに対し、教師なし学習に基づく領域分割は、クラスタリング後のデータ分布から各ピクセルのクラスを推定し、得られたラベルを教師データ (疑似ラベル) とみなすことで、教師データなしでピクセルごとにクラス推定を行い学習する手法である。教師あり学習手法と比べて教師データを作成する必要がないため、アノテーションコストを抑えて学習出来るという利点がある。教師なし領域分割手法の 1 つに PiCIE (Pixel-Level Feature Clustering using Invariance and Equivariance) [1] がある。PiCIE は深層学習に基づくクラスタリング手法 DeepCluster [2] に基づいて、疑似ラベルを作成し、作成した疑似ラベルを教師データとして、モデルの学習を行う。しかし、一般的なデータセットでは、各クラスのサンプル数が異なるため、モデルがサンプル数の多いクラスに過度に適応する過学習 (Overfitting) が起こる問題がある。これにより、未知データに対するクラスの識別性能の低下が生じ、クラス間で精度のばらつきが大きくなる。本研究は、PiCIE が過学習を起こしやすい点に着目し、過学習を抑える IB Focal 損失を導入することでこの問題を解決し、精度を高める。

2. PiCIE の概要

PiCIE [1] は DeepCluster [2] を領域分割に応用し、k-means 法に基づいてピクセル単位でクラスタリングを行う手法である。各ピクセルに正しいラベルを付与するために、フォトメトリック変換 (Invariance to photometric transformations) に対する不変性とジオメトリック変換 (Equivariance to geometric transformations) に対する等変性の 2 つの制約を加える。これにより、輝度の変化や角度の変化に対して影響を受けにくくなる。そのため、一般的な教師なし領域分割手法と比較して、モデルが特定の照明条件や視点に過度に適応することなく、モデルの学習を行うことができる。

2.1 PiCIE の問題点

PiCIE を用いた領域分割の精度 (Accuracy, mIoU) を表 1 に示す。表中の *Thing* は物体クラス、*Stuff* は背景クラスを意味する。

表 1 : PiCIE の精度 (All: 全体 *Thing*: 物体 *Stuff*: 背景)

	Accuracy	mIoU	Accuracy	mIoU	Accuracy	mIoU
	All	All	<i>Thing</i>	<i>Thing</i>	<i>Stuff</i>	<i>Stuff</i>
PiCIE	49.99	14.36	69.39	23.83	74.56	17.32

[†] 徳島大学大学院 創成科学研究科 Graduate School of Sciences and Technology for Innovation, Tokushima University

[‡] 徳島大学大学院 社会産業理工学研究部 Graduate School of Technology, Industrial and Social Science, Tokushima University

表 1 より、物体クラス (*Thing*) と背景クラス (*Stuff*) の Accuracy を比較すると、背景クラスが物体クラスよりも精度が約 5 ポイント高いことがわかる。理由は、背景クラスは単純で一般的な特徴が多く、モデルが容易に学習しやすいのに対し、物体クラスは複雑な形状をしたものが多いため、モデルの学習が困難であるからだと考えられる。また、COCO-*Stuff* データセットは、日常に近い風景の画像が多く含まれているため、背景クラスが物体クラスに比べて面積が広く、データ数が多くなりやすいため、背景クラスに学習が偏っていると考えられる。そのため、PiCIE では物体クラスにおいて、過分割が発生しているサンプルが多い。

3. 提案手法

図 1 に提案手法のシステム概要を示す。

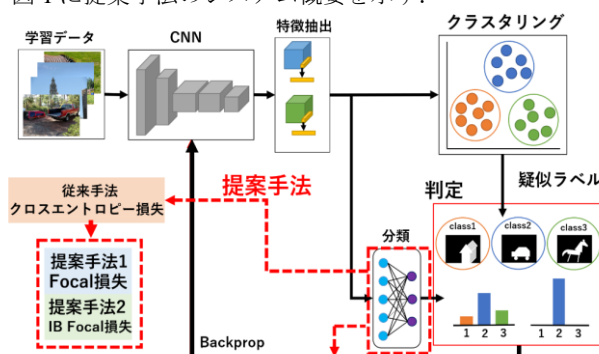


図 1 : 提案手法のシステム概要

従来手法では、クラスタリングで推定された疑似ラベルを参照し、softmax 関数でクラス分類を行う。その後、クロスエントロピー損失を用いて、予測したラベルと同じ特徴量から、生成した疑似ラベルを一致させるために、セントロイドへの距離に基づいて各ピクセルの損失を求める。

提案手法では、PiCIE の問題点である物体クラスの過分割を改善する手法として、損失関数をクロスエントロピー損失から Focal 損失、IB Focal 損失へと変更する。

3.1 Focal 損失に基づく手法 (提案手法 1)

Focal 損失 [3] は Lin らによって提案された損失関数である。Focal 損失の特徴は、クロスエントロピー損失の計算式に、softmax 関数における正解クラスの確率値に応じて重み付けを追加していることである。クロスエントロピー損失の計算式を式 (1)、Focal 損失の計算式を式 (2) に示す。ここで、 $S(x_j)$ は softmax 関数の出力値が、最大値のクラスを正解クラスと予測した場合の確率値を示している。

$$CrossEntropyLoss = -\log(S(x_j)) \quad (1)$$

$$FocalLoss = -(1 - S(x_j))^2 \log(S(x_j)) \quad (2)$$

式(2)より, Focal 損失はクロスエントロピー損失と比較し, クラス予測がしやすいサンプル($S(x_j)$ の値が 1 に近い)に対しての損失値が低下する. よって, 従来手法よりクラス予測が困難なサンプルに対して重点的に学習を行えることが考えられる. これにより, 複雑な形状が多いと推察される物体クラスの精度向上が期待できる.

3.2 IB Focal 損失に基づく手法 (提案手法 2)

IB 損失 (Influence-Balanced loss) [4] は, Park らによって提案された損失関数である. クロスエントロピー損失で学習をすると, 全てのサンプルに対して重みが一定である. そのため, クラスごとのサンプル数に違いがある場合, モデルが多数派のクラスに過度に適応する問題がある. よって, クラスを分ける決定境界が多数派のクラスのサンプルに依存するため, 少数派クラスの決定境界の学習が適切に行えない. IB 損失では, クラスを分ける決定境界付近の重みを下げることで, 境界の平滑化を行い, 決定境界の過学習を抑制する. IB 損失の計算式に含まれるクロスエントロピー損失を Focal 損失に変更したものを, IB Focal 損失とする. IB 損失は決定境界の過学習に焦点を当てるため, 決定境界を学習させてから損失関数を変更する必要がある. よって, 本手法ではクロスエントロピー損失で学習中に, 損失が最小値に到達したエポックで, クロスエントロピー損失から IB Focal 損失に切り替える.

4. 実験

4.1 実験環境とデータセット

Docker 環境を用いて, NVIDIA RTX A5000 VRAM 24GB の GPU を用いて実行した. また, 従来手法と同条件で実験するため, データセットを COCO-Stuff (2017) とし, 学習データ, 検証用データ, テストデータともにクラス数を 27 (*Thing*-12, *Stuff*-15), バッチサイズを 256, 入力画像をすべて 320*320 サイズに変換し, 実験を行った. 損失関数を変更すると学習速度に違いが表れるため, 提案手法 1 ではエポック数を 15, 提案手法 2 ではエポック数を 20 と設定した.

4.2 提案手法に対する実験

表 2 : 提案手法の領域分割に対する評価結果

	Accuracy	mIoU	Accuracy	mIoU	Accuracy	mIoU
	All	All	Thing	Thing	Stuff	Stuff
PICIE	49.99	14.36	69.39	23.83	74.56	17.32
手法 1	46.21	16.34	77.17	33.68	61.27	22.68
手法 2	51.37	15.57	74.73	29.01	69.33	20.99

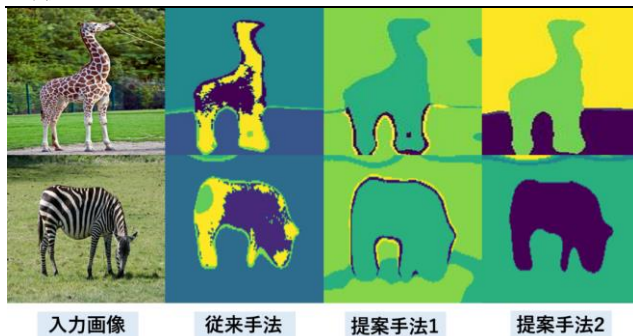


図 2 : 提案手法に対する領域分割結果の例

評価指標として, Accuracy と mIoU (mean Intersection over Union) を用いる. Accuracy はピクセル単位的全データ数に対し, 予想されたクラスが正解している割合 (%) を表している. また, mIoU は予測したクラス領域と, クラスの正解領域の面積の重なりを割合 (%) を全クラスに対して行い, 平均をとったものとする. 表 2 に提案手法 1, 2 の評価結果を示す. また, 図 2 に提案手法に対する領域分割画像の例を示す.

4.3 考察

提案手法 1 では表 2 より, 物体クラスに対して精度向上は見られたが, 背景クラスに対する精度が低下した. これは Focal 損失を用いて, 予測が難しいサンプルに対して, 損失を大きくしたため, 物体クラスに焦点を当てた学習を行えた一方で, 背景クラスに対する損失が十分に考慮されていないことが原因だと考察できる. 図 2 より, 従来手法と比較し, 物体の輪郭は捉えているが, 背景クラスに対する精度低下が顕著であることがわかる.

提案手法 2 では表 2 より, 全体の Accuracy, mIoU ともに従来手法より精度が向上することを確認した. これは物体クラス内や異なるクラス境界においての過分割が抑制されたからだとも図 2 より考察できる. IB 損失で誤分類しやすいサンプルの重みを低下させたことにより, クラスの特徴を間違えて学習する割合が減少したため, クラスごとの特徴量が学習しやすくなったからだとも考えられる. しかし, 提案手法 1, 2 ともに従来手法と比較し, 背景クラスに対する損失を小さくしたため, 背景クラスに対する精度低下が発生していると推察できる.

5. おわりに

本研究では, 従来手法の物体クラスが背景クラスにより精度が低いという問題に対し, 物体クラスの過分割が起こっていることに着目し, 損失関数を変更した. 提案手法 2 において, 全体の Accuracy, mIoU ともに従来手法より精度が向上することを確認した. 物体クラスにおいてクラスの過分割が改善されたことと, 異なるクラスの境界付近の精度が向上したことが要因だと考察できる. 今後の課題は, 物体クラスと背景クラスの両方に対してバランスよく損失を低下させる新しい損失関数を作成することである.

参考文献

- [1] Jang Hyun Cho, Utkarsh Mall, Kavita Bala, and Bharath Hariharan. Picie: unsupervised semantic segmentation using invariance and equivariance in clustering. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 16794–16804, June 2021.
- [2] Mathilde Caron, Piotr Bojanowski, Armand Joulin, and Matthijs Douze. Deep clustering for unsupervised learning of visual features. In Proceedings of the European Conference on Computer Vision (ECCV), September 2018.
- [3] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollar. Focal loss for dense object detection. In Proceedings of the - 2017.
- [4] Seulki Park, Jongin Lim, Younghun Jeon, and Jin Young Choi. Influence-balanced loss for imbalanced visual classification. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pages 735–744, October 2021.