

1. はじめに

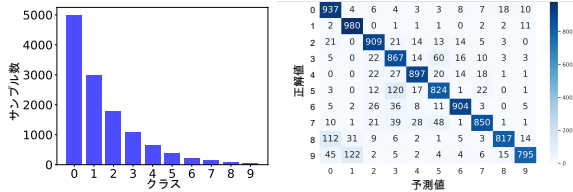
Long-tailed 物体認識は、クラス毎のサンプル数に偏りがあるデータを対象とする。代表的な手法である GLMC[1]は、教師あり学習と表現学習のマルチタスク学習によって、Long-tailed 物体認識の精度を向上した。表現学習では、mixup と CutMix を施した画像同士の特徴量の類似度を最大化する。これにより、混合したクラスの特徴表現の一貫性を向上する一方で、異なるクラス間の特徴分布が近くなる。特に、混合する頻度が少ない少数派クラスに対して表現学習が十分に発揮されず、多数派クラスの認識精度と大きな差が生じる。そこで、本研究では、対照学習を用いた特徴表現を向上する学習法を提案する。

2. 従来手法 (GLMC)

GLMC [1] は、入力画像に mixup と CutMix というデータ拡張を施し、それらに対して教師あり学習と表現学習を行う手法である。教師あり学習では、通常のカテゴリ損失とサンプル数を考慮した損失の割合をエポック毎に調整し、多数派クラスに対する過度な学習を緩和する。表現学習では、混合した画像同士の特徴量のコサイン類似度を最大化し、クラス毎の一貫した特徴表現を獲得する。GLMC は、この 2 つのタスクを同時に学習することで、Long-tailed 物体認識における高精度化を実現した。

3. GLMC におけるモデルの予測傾向の調査

GLMC が高精度化を達成した一方で、多数派クラスと少数派クラスの精度に大きな差が存在する問題がある。GLMC で CIFAR-10-LT を評価した際の評価結果を図 1 に示す。図 1(a) は、学習データの分布である。図 1(b) より、少数派クラスを多数派クラスと多く誤認識する傾向があることが分かる。これは、mixup と CutMix が、混合したクラスの中間的な特徴を学習して識別能力を向上させる一方で、混合したクラスの特徴分布が近くなるからである。また、各クラスの混合する頻度は、クラスのサンプル数に依存する。そのため、少数派クラスと多数派クラス間での識別境界が学習しきれず、特徴分布同士が近づくことで誤分類を誘発したと考える。以上より、混合した画像の特徴量の類似度を最大化する表現学習だけでは不十分であると考える。



(a) 学習に使用したデータ分布 (b) 評価データに対する混同行列

図 1: 傾向調査の結果

4. 提案手法

本研究では、少数派クラスの精度向上を目的とし、特徴表現を向上する手法を提案する。提案手法では、3 節で述べた問題を解決するために、混合前後の画像の特徴量に対する対照学習を GLMC に導入する。この時、混合前後の画像の特徴量に対する類似度を混合比率に基づいて最大化し、それ以外の特徴量に対する類似度は最小化する。これにより、混合前後の画像の特徴関係を考慮した学習と、混合前の画像を基準としたクラス毎の特徴分布を離す学習を可能にする。提案手法の流れを図 2 に示す。具体的には、各イテレーション毎に 3 つのステップを行う。

Step1. 混合画像の生成

mixup, CutMix を混合比率 λ に基づいて異なるミニバッチ間で行い、混合画像を生成する。

Step2. 特徴量を抽出

モデルに混合前後の画像を入力し、特徴量を抽出する。ここで、混合前の画像の特徴量を $z^{(1)}$ と $z^{(2)}$ 、mixup と CutMix した画像の特徴量を $\tilde{z}^{(m)}$ 、 $\tilde{z}^{(c)}$ とする。

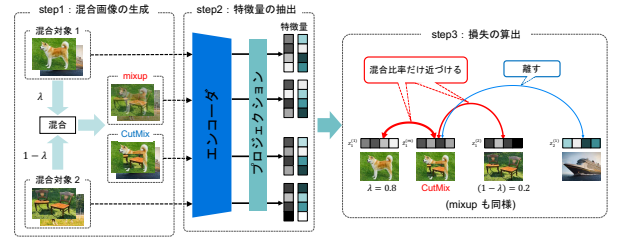


図 2: 提案手法の概念図

Step3. 損失の算出

まず、各特徴量のコサイン類似度を算出後、式 (1) に示す simCLR の NT-Xent Loss に基づいて混合前後の損失を算出する。ここで、 $\text{sim}(z, \tilde{z})$ は z と $\tilde{z}$ のコサイン類似度、 τ は温度パラメータ、 N はバッチサイズを表す。

$$L(z, \tilde{z}) = \frac{1}{2N} \sum_{i=1}^{2N} \left(-\log \frac{\exp(\text{sim}(z_i, \tilde{z}_i)/\tau)}{\sum_{k=1}^{2N} \exp(\text{sim}(z_i, \tilde{z}_k)/\tau)} \right) \quad (1)$$

式 (1) と混合比率 λ を使用して、mixup した画像及び CutMix した画像に対する損失 L_{mixup} 、 L_{cutmix} を算出する。損失 L_{mixup} と L_{cutmix} を式 (2) に示す。

$$\begin{aligned} L_{\text{mixup}} &= \lambda L_{(z^{(1)}, \tilde{z}^{(m)})} + (1 - \lambda) L_{(z^{(2)}, \tilde{z}^{(m)})} \\ L_{\text{cutmix}} &= \lambda L_{(z^{(1)}, \tilde{z}^{(c)})} + (1 - \lambda) L_{(z^{(2)}, \tilde{z}^{(c)})} \end{aligned} \quad (2)$$

最後に、これらの平均を求めて、提案手法の総損失を算出する。提案手法の総損失 L_{CLR} を式 (3) に示す。

$$L_{\text{CLR}} = (L_{\text{mixup}} + L_{\text{cutmix}})/2 \quad (3)$$

5. 評価実験

Long-tailed 物体認識における提案手法の有効性を評価する。

5.1. 実験概要

本実験で使用するデータセットは、不均衡率 ρ を [100, 50] とした CIFAR-100-LT である。エンコーダには、ResNet-32 を使用する。ミニバッチサイズは 128、エポック数は 200 とする。比較手法は、GLMC、提案手法とする。評価指標は、学習サンプル数に基づいて 3 つのクラスに分割した Many-shot, Med-shot, Few-shot グループの Accuracy とする。また、各評価指標は seed 値を 5 回変更して学習を行った結果の平均値とする。

5.2. 実験結果

比較結果を表 1 に示す。表 1 より、どちらの不均衡率でも Few-shot, Med-shot の精度が向上したことが分かる。この結果より、提案手法は少数派クラスの精度改善が可能であると考えられる。一方、Many-shot の精度低下は、提案手法による同じクラスに属する画像の特徴量も離す作用が、多数派クラスに大きく影響するためと考える。

表 1: CIFAR-100-LT における各精度 [%]

ρ	手法	Many-shot	Med-shot	Few-shot	平均
100	GLMC	74.55	57.36	37.96	57.34
	提案手法	71.41	58.01	41.04	57.44
50	GLMC	75.14	57.58	42.55	61.92
	提案手法	74.92	58.19	43.14	62.19

6. おわりに

本研究では、混合比率を用いた対照学習による表現学習を向上する手法を提案し、少数派クラスの認識精度が向上することを確認した。今後の展望として、Many-shot の精度低下を抑制するために、特徴空間で離す対象を学習データのラベル情報を用いて選択するように改善する。

参考文献

- [1] F. Du, et al., “Global and Local Mixture Consistency Cumulative Learning for Long-tailed Visual Recognitions”, CVPR, 2023.