

1. はじめに

深層学習モデルの学習において、データ拡張は認識精度向上に不可欠なテクニックである。mixup などの複数データを混合して生成するデータ拡張手法は、学習データに含まれない新たな多様性を獲得できる。このとき、混合するデータの選択は、ランダムサンプリングのため、学習経過において適切なクラスやデータが選択されない場合があり、学習に有効とは限らない。本研究では、特徴ベクトルに基づいてクラス間距離を算出し、学習過程で混合対象として適したクラスからデータを選択して混合するデータ拡張法を提案する。評価実験により、提案手法が一般的な物体認識データセット及び Long-tailed 物体認識データセットで認識性能を改善することを示す。

2. 複数データの混合によるデータ拡張

複数データを混合するデータ拡張手法は、単一データに対する幾何変化と比べ、学習データの見えの多様性を高めることができる。mixup[1] は、ランダムにサンプリングした画像と対応するラベルを混合することで、新たに混合データを生成する手法である。混合データは、2 枚の画像と正解ラベルを β 分布に基づいてサンプリングした比率を用いて線形補間して求める。CutMix は、画像の一部を別の画像に置き換える手法である。

これらの複数データを混合する手法は、クラス毎のサンプル数に偏りのあるデータセットを対象とする Long-tailed 物体認識にも用いられている。代表的な手法である Uniform-mixup (UniMix)[2] は、クラス毎のサンプル数を考慮して、ミニバッチ生成時のデータのサンプリング傾向と混合する比率を調整する。

これらの従来手法は、データ混合処理の改良やデータバランスを考慮した事前調整によって性能を向上させている。しかしながら、クラス間の関係性を考慮して混合データを生成するには至っていない。

3. 提案手法

本研究では、特徴ベクトルに基づくクラス間距離を考慮して、学習過程で混合するクラスを動的に選択する新たなデータ拡張法を提案する。提案手法は、学習中にネットワークが出力する特徴ベクトルを用いて、クラス間のコサイン距離を算出する。このクラス間距離を考慮し、学習過程のネットワークの出力傾向に適した混合クラスを選択する。ここで、混合クラスは、混合対象データをサンプリングするクラスを指す。図 1 に提案手法によるデータサンプリングの流れを示す。提案手法は、特徴ベクトル空間におけるクラス間距離に基づいて、混合クラスを選定する。

3.1 クラス間のコサイン距離の算出

クラス間のコサイン距離は、エポック毎に更新する。コサイン距離の算出手順を以下に示す。

1. $t-1$ エポックまで学習したネットワーク f_θ に、学習サンプル x_i を入力し、特徴ベクトル v_i を出力する。これを全ての学習サンプルに対して行う。
2. 同一クラス (図 1 では *ClassA*) に属する学習サンプルに対するクラス確率をもとに、平均特徴ベクトル \bar{V}_A を求める。これを全てのクラスに対して行う。
3. 基準クラス (*ClassA*) と他クラス (*ClassB*, *ClassC*, ...) とのコサイン距離 $d_{A \rightarrow B}$, $d_{A \rightarrow C}$, ... を算出する。これを、基準クラスを変更し、全てのクラスについて行う。*ClassA* と *ClassB* 間のコサイン類似度を式 (3)、コサイン距離を式 (2) に示す。

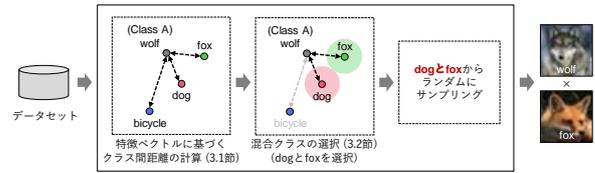


図 1：混合対象データのサンプリング処理

$$\begin{aligned} \cos(\bar{V}_A, \bar{V}_B) &= \frac{\bar{V}_A \cdot \bar{V}_B}{\|\bar{V}_A\| \|\bar{V}_B\|} \\ &= \frac{\sum_{k=1}^{dim} \bar{V}_{A_k} \bar{V}_{B_k}}{\sqrt{\sum_{k=1}^{dim} \bar{V}_{A_k}^2} \sqrt{\sum_{k=1}^{dim} \bar{V}_{B_k}^2}} \end{aligned} \quad (1)$$

$$d_{A \rightarrow B} = 1 - \cos(\bar{V}_A, \bar{V}_B) \quad (2)$$

dim は特徴ベクトルの次元数である。

3.2 クラス間距離を考慮した混合クラスの選定

算出したクラス間距離を用いて、混合クラスを選定する。処理の概要を図 2 に示す。 t エポックにおいて学習サンプルを混合する際に、1 枚目のサンプルが属するクラス (図 2 では *ClassA*) とのコサイン距離に応じて、2 枚目のサンプルを選出する混合クラスを決定する。このとき、 $t-2$ と $t-1$ エポック間における認識精度の変化で、選出するための距離関係 dr と混合クラス数 n を、以下の流れで基準とするクラス毎に決める。距離関係 dr は、混合クラスを選ぶ際にクラス間距離を近距離または遠距離のいずれで並び替えるかを表す。

- 1 枚目サンプルのクラスの精度が停滞・向上した場合
 - (a) 距離関係 dr は変更しない
 - (b) 混合クラス数 n は l クラス増加
- 1 枚目サンプルのクラスの精度が低下した場合
 - (a) 距離関係 dr は $t-1$ エポックが近距離の場合は遠距離、遠距離の場合は近距離に変更
 - (b) 混合クラス数 n は l クラス減少

混合クラス数 n は初期値を下限とし、1 エポックにおける増減数 l はあらかじめ定義しておく。

3.3 混合クラスの選定による学習傾向の変化

遠距離クラス混合時に精度が低下する場合、近距離クラスによって偽陽性が生じていると考えられる。そのため、 t エポックでは近距離クラスを混合対象にしてクラス間距離を遠ざけるように学習する。一方、近距離クラス混合時に精度が低下する場合、距離が遠いことにより識別境界付近のデータが少ないことが要因と考えられる。そのため、 t エポックでは遠距離クラスを混合対象にして識別境界を学習しやすくする。また、精度が低下した場合は精度低下の原因となるクラスに絞るため、混合クラス数を減らす。一方、精度が向上した場合は見えの多様性を向上させるために、混合クラス数を増やす。

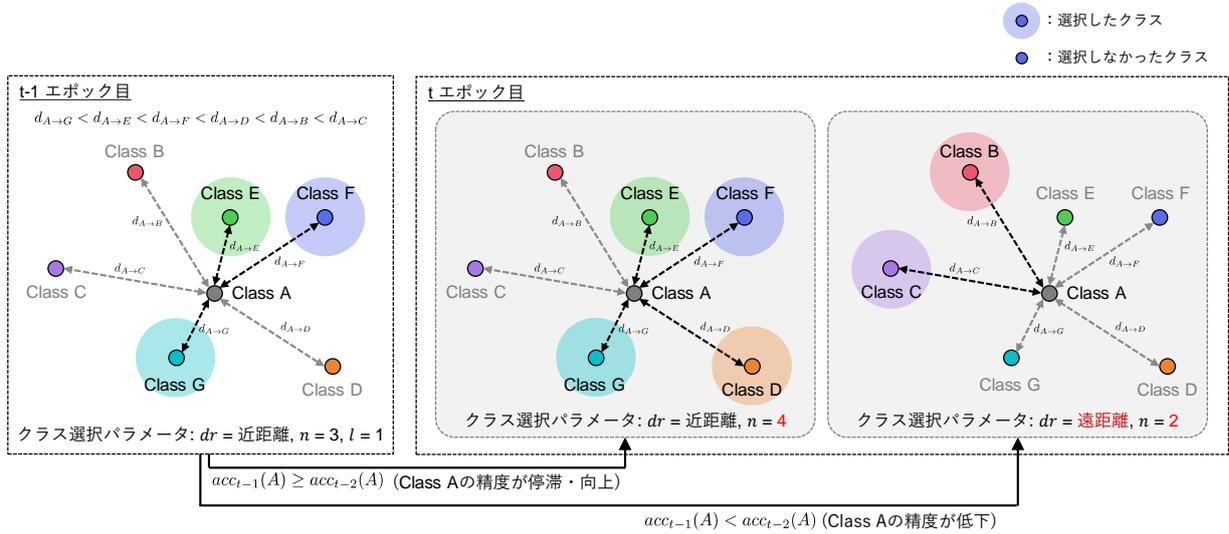


図 2 : クラス間距離を考慮した混合クラスの選定処理

表 1 : CIFAR-10 における精度 [%]

手法	クラス間距離	Res-32	PreAct Res-18
データ拡張無し		93.23	95.11
mixup	✓	95.53 95.86	96.07 96.20
CutMix	✓	95.35 95.74	96.19 96.69

表 2 : CIFAR-100 における精度 [%]

手法	クラス間距離	Res-32	PreAct Res-18
データ拡張無し		74.64	75.50
mixup	✓	77.37 77.26	78.38 79.05
CutMix	✓	78.27 80.02	80.38 79.61

4. 評価実験

本章では、一般物体認識および Long-tailed 物体認識データセットを用いて、提案手法の有効性を評価する。

4.1 一般物体認識における精度比較

一般物体認識において、mixup, CutMix に提案手法を導入した際の性能を比較する。データセットは、CIFAR-10, CIFAR-100 を用いる。ネットワークモデルは、双方のデータセットにおいて ResNet-32, PreAct ResNet-18 を用いる。ミニバッチサイズは 128, エポック数は 200 とする。クラス間距離計算に用いる特徴ベクトルは、各ネットワークモデルにおける全結合層の一つ前の層から取得する。混合クラス選定の初期パラメータは、距離関係 dr を遠距離、混合クラス数 n を 5, エポック毎の n の増減数 l を 5 とする。

CIFAR-10, CIFAR-100 における各手法の認識精度をそれぞれ表 1, 表 2 に示す。表より、各条件において、提案手法を導入することによって概ね精度が向上している。よって、一般物体認識において、クラス間距離を考慮した混合クラスの選定は有効であるといえる。一部の精度低下の要因は、該当の条件でモデルが多様な混合データを十分に学習できるためであると推察する。

4.2 Long-tailed 物体認識における精度比較

Long-tailed 物体認識において、mixup, UniMix に提案手法を導入した際の精度を比較する。データセットは、CIFAR-10-LT, CIFAR-100-LT ($\rho = 50$) を用いる。 ρ は、各クラスのサンプル数の比率を調整する不均衡係数である。一般的には $\rho = 10 \sim 100$ が用いられ、 ρ が大きいほどサ

表 3 : CIFAR-10-LT における精度 [%]

手法	クラス間距離	Res-32	PreAct Res-18
データ拡張無し		74.94	75.77
mixup	✓	81.60 83.80	83.21 83.94
UniMix	✓	86.17 87.25	87.61 88.19

表 4 : CIFAR-100-LT における精度 [%]

手法	クラス間距離	Res-32	PreAct Res-18
データ拡張無し		44.02	45.53
mixup	✓	48.75 48.50	49.27 49.96
UniMix	✓	53.32 53.24	54.18 54.40

ンプル数が偏った構成となる。ネットワークモデルおよび学習設定は一般物体認識の実験と同様である。

CIFAR-10-LT, CIFAR-100-LT における各手法の認識精度をそれぞれ表 3, 表 4 に示す。表より、各条件において、提案手法により概ね精度が向上している。よって、Long-tailed 物体認識においても、提案手法は有効であるといえる。一部の精度低下の要因として、認識が困難なクラスの混合に制約が課されたことにより、認識が容易なクラスの混合パターンが減少したためであると推察する。

5. おわりに

本研究では、特徴ベクトルに基づくクラス間距離に応じて混合クラスを選択する手法を提案した。提案手法により、学習過程で性能向上に有効な混合クラスを選択することが可能となった。また、提案手法は、混合データを生成する従来手法に容易に導入でき、一般画像および Long-tailed 物体認識で有効であることを確認した。今後は、設計分析を継続し、データセットの構成による影響を更に抑制する設計を検討する。

参考文献

- [1] H. Zhang, *et al.*, "mixup: Beyond Empirical Risk Minimization", In ICLR, 2018.
- [2] Z. Xu, *et al.*, "Towards Calibrated Model for Long-Tailed Visual Recognition from Prior Perspective", In NeurIPS, 2021.

研究業績

- [1] S. Fujii, *et al.*, "Super-class Mixup for Adjusting Training Data", In ACPR, 2021.

(他 2 件)