

1. はじめに

SimCLR [1] は、対照学習を用いた自己教師あり学習である。対照学習では、基準となるアンカーと正例もしくは負例との 2 点間のみの距離関係を学習するため、特徴空間内で構造的な関係を考慮することが難しい問題がある。また、SimCLR の学習時には色変換、クロップ、リサイズ、画像反転といったデータ拡張を用いる。この色変換は、自然画像には存在しない不自然な色情報への変換を行うことがあるため、画像に含まれる重要な色情報を無視した学習となる。そこで、本研究では SimCLR のデータ拡張に mixup を加え、トリプレット損失を導入した学習法を提案する。mixup は 2 枚の画像の合成比率を調整することで不自然な色変換を抑制し、従来の色変換と比較して色情報を保持した画像の生成が可能である。提案手法では、mixup より色情報を保持した画像をアンカー、従来の色変換を行った画像を正例と負例にすることで、色情報を活用した構造的な特徴空間の構築を行い、特徴表現の獲得を促進する。

2. SimCLR

SimCLR[1] は、様々なデータ拡張を学習用画像に適用し、異なる変換が適用された画像ペアを生成する。損失関数 $\mathcal{L}_{NT-Xent}$ を用いて、同じ画像から拡張された正例ペアの類似度を最大化し、異なる画像から拡張された負例ペアの類似度を最小化することで、様々なタスクにおけるクラス間の分類に有用な特徴表現の学習を促進する。

3. 提案手法

提案手法のアーキテクチャを図 1 に示す。まず、入力画像 X に色変換以外のデータ拡張を施した 2 枚の画像 X_1, X_2 を生成する。そして画像 X_1, X_2 に比率 λ で mixup を適用し、色情報を保持した画像 X_{mix} を生成する。 λ は Beta 分布で決定される。次に、 X_1, X_2 にはランダムな色変換を適用し、色変換画像 X'_1, X'_2 を生成する。 X_{mix} をアンカー、 X'_1, X'_2 を正例候補とする。損失関数は式 (1) に示すように、 A と B のトリプレット損失で 2 つの正例候補を用いて損失を計算し、それらの平均とする。損失関数に 2 つのトリプレット損失を組み合わせることで、異なるデータ拡張が適用された 2 つの正例を考慮した構造的な特徴空間の構築が可能となり、色情報を活用した学習を促進する。

$$\mathcal{L}_{\text{triplet}} = \frac{1}{N(2N_n)} \sum_{k=1}^N (A + B)$$

$$A = \sum_{l=1}^{N_n} \left(pd_k^{X'_n} - nd_l^A + \alpha \right), \quad B = \sum_{l=1}^{N_n} \left(pd_k^{X'_f} - nd_l^B + \alpha \right) \quad (1)$$

$pd_k^{X'_n}, pd_k^{X'_f}$ はアンカー・各正例間の類似度、 nd_l^A, nd_l^B はアンカー・各負例間の類似度、 α はマージンである。正例候補 X'_1, X'_2 のうち、比率 λ が大きい方を X'_n 、小さい方を X'_f とする。 A と B の負例は、他の入力画像に色変換を適用した画像群から同一の画像がないように、それぞれ N_n 個ずつサンプリングする。

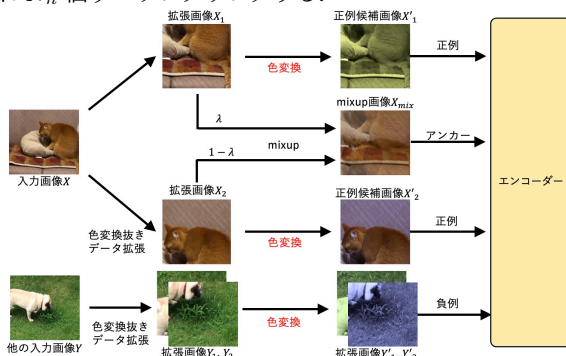


図 1: 提案手法のアーキテクチャ

4. 調査実験

線形評価を用いて、事前学習済みモデルの重みを固定し、その特徴表現を入力とする線形分類器を訓練する。分類精度を指標として提案手法の有効性を評価する。データセットには CIFAR-10, CIFAR-100, Flowers102, Food101 を用いる。比較対象は SimCLR とする。また、提案手法が色情報を活用して学習しているかを確認するため、評価用画像をグレースケール化した場合の精度を調査する。さらに、UMAP を用いて特徴分布を可視化し、グレースケール化がモデルの特徴表現に与える影響を分析する。

4.1. 分類精度比較

分類精度を表 1 に示す。表 1 より、提案手法は SimCLR よりも CIFAR-10 で 2.79pt, Food101 で 8.54pt 上回る。一方で、他のデータセットでは SimCLR を下回った。以上の結果から、提案手法による色情報を活用した学習は、特定のデータセットの性能向上に寄与する一方、色に依存した学習に課題が残ることがわかった。また、評価用画像をグレースケール化した際の分類精度は、SimCLR では通常評価時と比較して 1.1pt 低下し、提案手法では 5.52pt 低下した。グレースケール化による色の喪失の影響が SimCLR と比較して提案手法の方が大きいことから、提案手法は色情報を活用した学習を行なっていることがわかる。

表 1: 提案手法と SimCLR におけるデータセット別の精度比較

| データセット | Gray Scale | Ours Acc.(%) | SimCLR Acc.(%) |
|------------|------------|--------------|----------------|
| CIFAR-10 | - | 84.70 | 81.91 |
| | ✓ | 79.18 | 80.81 |
| CIFAR-100 | - | 50.57 | 64.87 |
| Flowers102 | - | 75.29 | 95.21 |
| Food101 | - | 53.49 | 44.95 |

4.2. 特徴分布比較

UMAP による特徴分布の可視化結果を図 2 に示す。SimCLR は通常評価時とグレースケール評価時で分布に変化が見られない一方、提案手法は特徴空間の分布に明確な変化が見られる。この結果から、提案手法は色情報を活用した学習を行なった結果、SimCLR とは異なる表現を学習したことが考えられる。

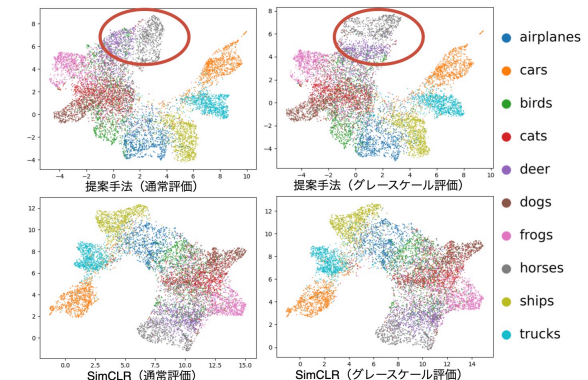


図 2: CIFAR-10 における特徴空間上の分布比較

5. おわりに

本稿では、データ拡張に mixup を加えたトリプレット損失による自己教師あり学習法を提案した。提案手法は、CIFAR-10 および Food101 にて SimCLR より精度向上を確認した。また、評価画像をグレースケール化した際の精度の変化と UMAP による特徴空間の可視化結果から、提案手法が色情報を活用して学習していることを示した。今後は、mixup によるデータ拡張とトリプレット損失の見直しにより、様々なデータセットでの精度向上を目指す。

参考文献

- [1] Ting Chen, *et al.* "A simple framework for contrastive learning of visual representations", ICML, 2020.