

1. はじめに

深層学習モデルを学習する場合、データを事前に収集・蓄積して学習するオフライン学習と、蓄積せずに逐次データを入力して学習するオンライン学習の2つのアプローチがある。オフライン学習は、蓄積したデータセットを繰り返し用いて学習するため高い精度を達成しやすい一方で、データの蓄積に伴うストレージコストが課題となる。オンライン学習は、収集したデータを即座に学習し、学習後はそのデータを破棄するため、ストレージコストを大幅に削減できる。しかし、収集したデータに対してリアルタイムでラベル付けして学習するのは困難である。

これに対して、自己教師ありオンライン継続学習は、逐次入力するデータに対して自己教師あり学習を行うことで、ラベルを付与するコストを低減できる。しかし、従来の自己教師ありオンライン継続学習手法は、2つの課題がある。(i) パラメータ更新時の勾配に相関が発生し、モデルが特定のデータに過度に適合することで汎化性能が低下する。(ii) 自己教師あり学習の収束が遅く、収集したデータを破棄するまでの短い時間で十分に学習することが困難である。

そこで本研究では、コサイン類似度を使用した学習データの選択によって勾配の相関を抑制し、マルチクロップ対照損失によって自己教師あり学習の収束速度を改善する自己教師ありオンライン継続学習手法を提案する。実験により、提案手法が従来手法と比較して分類精度を改善することを示す。

2. 自己教師ありオンライン継続学習

自己教師ありオンライン継続学習 (Self-Supervised Online Continual Learning: SSOCL) は、逐次入力されるラベルのないデータであるデータストリームを用いて継続的に学習するアプローチである。データストリームは、連続するデータ間に強い相関を持つと共に、時間の経過に伴いデータ分布が変化する非定常性という2つの特性がある。このような特性を持つデータストリームで学習するSSOCLには2つの課題がある。

課題1：パラメータ更新時の勾配の相関。 図1に t 回目のイテレーションと $t+1$ 回目のイテレーションにおけるパラメータ更新時の勾配のコサイン類似度を示す。従来の深層学習モデルは、サンプル間に相関がないデータで学習を行うことを仮定しており、勾配の類似度は0に近い値となる。一方で、サンプル間に相関のあるデータで学習を行うと、勾配の類似度が1.0に近い値になる。勾配の類似度が高いと、モデルのパラメータを特定のデータ分布に過度に適合する方向へ更新するため、汎化性能が低下する。

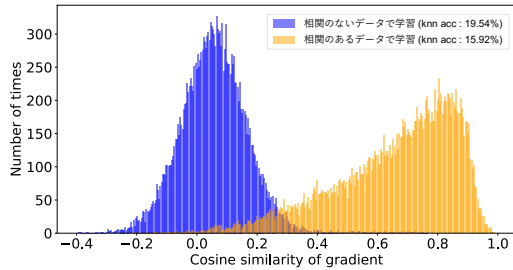


図1: パラメータ更新時の勾配の類似度

課題2：自己教師あり学習の収束の遅さ。 図2に教師あり学習と自己教師あり学習の収束速度を示す。図2より、自己教師あり学習は、教師あり学習と比較すると学習収束が遅いことがわかる。これは、データ分布が時間と共に変化する実世界において、自己教師あり学習の学習が不足する可能性を示している。

3. 提案手法

本研究では、SSOCLにおける2つの課題に対処する手法を提案する。提案手法は、勾配の相関に対処するためコサイン類似度を用いたデータ選択を行い、自己教師あり学習

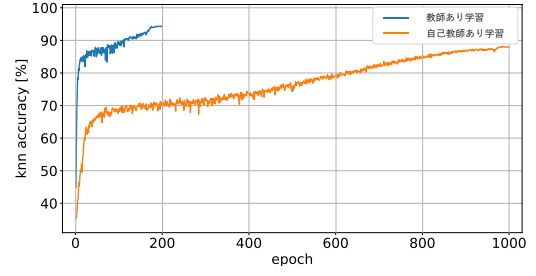


図2: 教師あり学習と自己教師あり学習の収束速度

の学習収束の遅さに対処するため Multi-Crop Contrastive Loss (MCC Loss) を導入する。

3.1. 学習プロセス

提案手法の学習プロセスを図3に示す。図3に示すように、提案手法は、データストリームで観測されたサンプルを固定サイズのバッファに追加する。その後、バッファからランダムにサンプリングしたデータで K 個のミニバッチを作成し、学習に使用する。このとき、自己教師あり学習の収束の遅さに対処するため、損失関数に MCC Loss を導入する。従来の Contrastive Loss は、1枚の画像に異なる2種類のデータ拡張を加えて得た2つのクロップに対し、それらの特徴量を1対1で近づけ、異なる画像から得たクロップは、遠ざけるように学習を行う。これに対して、提案手法で導入する MCC Loss は、3種類以上の異なるデータ拡張を適用し、得られた各クロップの特徴量をそれらの平均特徴量へと同時に近づける。クロップ数を増加させることで、1回のパラメータ更新においてより多くの情報を効率的に学習できるため、自己教師あり学習の収束の遅さを改善することが可能である。

K 個のミニバッチで学習後、バッファ内のサンプル数が一定数を超えている場合、コサイン類似度に基づいて多様なサンプルのみをバッファに保持し、冗長なサンプルを削除する。これにより、提案手法は、データストリームが持つ相関のある冗長なデータでの学習を防止し、勾配の相関に対処する。その後、データストリームで観測したサンプルに対して同様の処理を繰り返す。以下では、MCC Loss とコサイン類似度によるデータ選択について詳細に説明する。

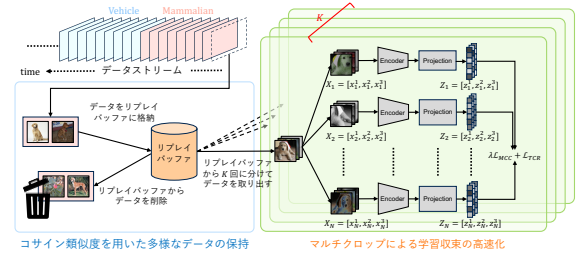


図3: 提案手法の概要

3.2. Multi-Crop Contrastive Loss

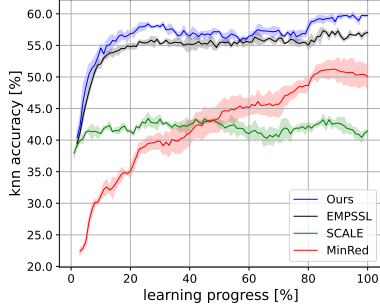
自己教師あり学習の収束速度は、クロップ数の増加によって高速化可能であることが知られている [3]。提案手法は、従来の2クロップのみを対象とした Contrastive Loss を3クロップ以上に拡張した Multi-Crop Contrastive Loss (MCC Loss) を導入する。MCC Loss を式 (1) に示す。

$$\mathcal{L}_{MCC} = \frac{1}{Nb} \sum_{i=1}^N \sum_{j=1}^b \left(-\log \frac{\exp(\bar{\mathbf{z}}^j \cdot \mathbf{z}_i^j / \tau)}{\sum_{k=1}^N \sum_{l=1}^b \exp(\bar{\mathbf{z}}^j \cdot \mathbf{z}_k^l / \tau)} \right) \quad (1)$$

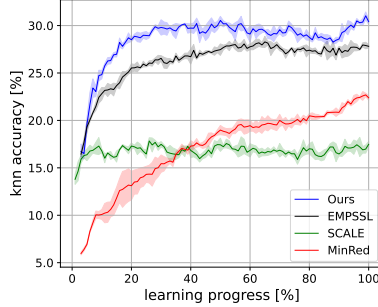
ここで、 N はクロップ数、 b はバッチサイズ、 \mathbf{z} は各サンプルの特徴量、 τ は温度パラメータ、 $\bar{\mathbf{z}}^j$ はサンプル \mathbf{x}^j の

表 1: 学習終了時の kNN 分類精度 [%]

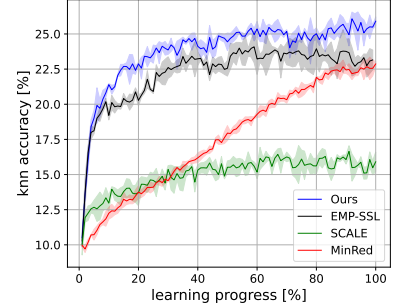
	CIFAR10			CIFAR100			ImageNet100		
	Seq	Seq-bl	Seq-im	Seq	Seq-bl	Seq-im	Seq	Seq-bl	Seq-im
MinRed[2]	50.04	51.18	46.41	22.38	23.20	21.26	22.87	22.71	20.46
SCALE[1]	41.41	40.85	41.31	17.49	16.93	17.03	15.46	15.41	15.77
EMP-SSL[3]	57.02	57.32	57.31	27.81	28.40	27.90	22.79	22.01	22.99
Ours	59.71	59.96	58.67	30.41	30.32	30.00	25.81	25.64	25.24



(a) Seq-CIFAR10



(b) Seq-CIFAR100



(c) Seq-ImageNet100

図 4: 学習過程における kNN 分類精度 [%]

平均特徴量を示し、 $\bar{\mathbf{z}}^j$ は式 (2) で求める。

$$\bar{\mathbf{z}}^j = \frac{1}{N} \sum_{i=1}^N \mathbf{z}_i^j. \quad (2)$$

MCC Loss は、1 枚の画像に対して N 種類の異なるデータ拡張を加え、各サンプルの平均特徴 $\bar{\mathbf{z}}^j$ を計算する。そして、各データの特徴量 \mathbf{z}_i^j をその平均特徴量 $\bar{\mathbf{z}}^j$ に同時に近づける。また、異なるデータの平均特徴 $\bar{\mathbf{z}}^j$ から各データの特徴量 \mathbf{z}^l を遠ざけるように学習する。

3.3. コサイン類似度による多様なデータの選択

コサイン類似度を用いたデータ選択の目的は、データストリームから多様なデータを選択して学習に利用することである。コサイン類似度の計算は、バッファ内サンプル \mathbf{x}_i と同時に保存した代表的な特徴量 $\bar{\mathbf{z}}_i^*$ を用いて計算する。提案手法のデータ選択は、式 (3) で定式化できる。

$$\mathbf{x}_i^* = \arg \min_{\mathbf{x}_i \in \mathcal{M}} \min_{\mathbf{x}_j \in \mathcal{M}} \text{Sim}(\bar{\mathbf{z}}_i^*, \bar{\mathbf{z}}_j^*) \quad (3)$$

$$\bar{\mathbf{z}}_i^* \leftarrow \alpha \bar{\mathbf{z}}_i^* + (1 - \alpha) \bar{\mathbf{z}}_i \quad (4)$$

ここで、 \mathcal{M} はバッファ、 \mathbf{x}_i^* はバッファに保存するデータ、 $\bar{\mathbf{z}}_i^*$ は平均特徴 $\bar{\mathbf{z}}_i$ の指数移動平均であり、 \mathbf{x}_i^* と $\bar{\mathbf{z}}_i^*$ をバッファに保存する。リプレイバッファ内で類似度が高いデータを削除し、類似度が低く多様なデータを優先してリプレイバッファに保持する。

4. 評価実験

データストリームで学習した各手法のクラス分類における分類精度を評価する。

4.1. 実験条件

評価には CIFAR10/100, ImageNet100 の 3 つのデータセットを用いて、データストリームの構築を行う。各データセットを従来研究 [1] に従って、Seq, Seq-bl, Seq-im データストリームを構築する。Seq は、クラスごとのデータ数を統一し、データ分布が一定のタイミングで変化する模擬的なデータストリームである。また、より現実的なデータストリームで評価を行うため、Seq-bl はデータ分布の変化境界を曖昧にし、Seq-im はクラス毎のデータ数を不均衡にする。これにより、Seq-bl は現実世界で発生する環境の滑らかな変化を、Seq-im はクラス毎の出現頻度の偏りという現実的な不均一性を再現する。

4.2. 実験結果

各手法の学習終了時の分類精度を表 1 に示す。表 1 より、提案手法は、CIFAR10 で最大 19.11pt, CIFAR100 で最

大 13.39pt, ImageNet100 で最大 10.35pt の精度向上を確認できる。また、MinRed [2] は、Seq-im での精度低下が確認できる。これは、クラス毎の出現頻度を不均一にすることで、観測回数の少ないクラスに対して学習が収束しなかったためだと考えられる。一方で、提案手法は、Seq-im の精度は他のデータストリームと同程度であり、これはより現実的なデータストリームにおいても学習が可能であることを示している。

次に、学習過程における分類精度の比較を行う。各手法の学習過程における分類精度の推移を図 4 に示す。図 4 より、提案手法の kNN 分類精度は、学習進捗が 20% の時点において、Seq-CIFAR10 で約 56.0%, Seq-CIFAR100 で約 28.0%, Seq-ImageNet100 で約 23.0% である。これは、他手法の学習終了時点での精度と同等かそれ以上である。これは、提案手法の Multi-Crop Contrastive Loss によって学習収束を高速化したことに起因すると考えられる。

5. おわりに

本研究では、自己教師ありオンライン継続学習における勾配の相関による性能劣化と学習収束の遅さに対処する手法を提案した。提案手法は、コサイン類似度によるデータ選択によって、バッファ内に多様なデータを保持し学習に用いることで勾配の相関に対処し、MCC Loss を導入することで自己教師あり学習の収束の遅さに対処した。実験結果より、複数のデータセットにおいて提案手法は、従来手法よりも高い分類精度を達成することを確認した。今後は、ImageNet21K などより大規模なデータセットを使用して、より実世界に近いデータストリームを構築し、その有効性を評価する予定である。

参考文献

- [1] Yu *et al.*, “SCALE: Online Self-Supervised Lifelong Learning Without Prior Knowledge”, CVPRW, 2023.
- [2] Purushwalkam *et al.*, “The challenges of continuous self-supervised learning”, ECCV, 2022.
- [3] Tong *et al.*, “Emp-ssl: Towards self-supervised learning in one training epoch”, arXiv, 2023.

研究業績

- [1] Imai *et al.*, “Faster convergence and Uncorrelated gradients in Self-Supervised Online Continual Learning”, ACCV, 2024. (他 3 件)