

1. はじめに

半教師あり学習は、ラベルありデータとラベルなしデータの双方を用いた学習方法である。半教師あり学習手法として、2つのネットワーク間で知識を転移する Dual Student(DS) [1], DS のネットワーク数を4つ以上に拡張した Multiple Student(MS) [1] が提案されている。MS は DS に比べて高い精度であるが、全てのネットワーク間での知識転移を一度に行うことができないため、学習効率が悪い。そこで本研究では、効率的な知識転移方法による精度向上を目的として、1度に全てのネットワーク間で知識を転移する Refined Consistency を提案する。また、一貫性に基づいた半教師あり学習手法に利用されるデータ拡張手法を適用し、精度変化を評価する。

2. Dual Student

DS は、2つのネットワークを用いた半教師あり学習手法である。各ネットワークには、同じ入力に対して異なる摂動を加えた2つの入力を与える。そして、摂動に対して頑健になるよう学習する。さらに、安定性という指標を用いてネットワーク出力の一貫性を評価し、安定性が高いネットワークから優れた知識を転移させている。知識は多くのネットワークから転移させる方が高い精度を達成できる。MS は DS を拡張し、4つのネットワークを用いて、MS は、1度にランダムな2つのネットワークを用いて、DS と同様の方法で学習する。

3. 提案手法

本研究では、1度に全てのネットワーク間で知識を転移する Refined Consistency を提案する。MS と提案手法について、1イタレーションごとの知識転移の方向を図1に示す。提案手法の学習方法を図2に示す。

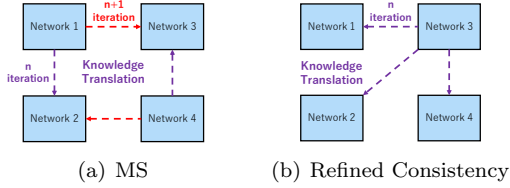


図 1: 知識転移方向の例 (ネットワーク数 4)

入力 x に対するネットワーク i の損失関数 $\mathcal{L}^i(x)$ を式 (1) に示す。

$$\mathcal{L}^i(x) = \mathcal{L}_{cls}^i + \lambda_1 \mathcal{L}_{con}^i + \lambda_2 \mathcal{L}_{sta}^i \quad (1)$$

ここで、 λ_1, λ_2 は重み係数であり、 $\mathcal{L}_{cls}$ は推測値とラベル間の交差エントロピーであり、 x がラベルありデータである場合のみ計算する。 $\mathcal{L}_{con}$ は一貫性損失、 $\mathcal{L}_{sta}$ は安定化損失を表す。安定化損失は、各ネットワークの一貫性損失を用いて計算する。

一貫性損失は、異なる摂動を入力に与えたときの事後確率間の誤差であり、値が小さいほどネットワークが摂動に頑健であることを表す。一貫性損失を式 (2) に示す。

$$\mathcal{L}_{con}^i(x) = \mathbb{E}_{x \in \mathcal{D}} \|f(\theta^i, x + \zeta_1) - f(\theta^i, x + \zeta_2)\|^2 \quad (2)$$

ここで、 $\mathcal{D}$ はデータセット、 ζ_1, ζ_2 はランダム性のある摂動、 $f(\theta, x)$ は重み θ をもつネットワークの事後確率を表す。

安定化損失は、一貫性損失を知識の指標として、最も一貫性損失が小さいネットワーク θ^s から他の全てのネットワークに知識を転移する。転移先のネットワークを θ^i とすると、知識は、事後確率を近づけるように学習することで転移させる。安定化損失を式 (3) に示す。

$$\mathcal{L}_{sta}^i(x) = \begin{cases} 0 & (\mathbf{A} = \emptyset) \\ \|f(\theta^i, x) - f(\theta^s, x)\|^2 & (\text{otherwise}) \end{cases} \quad (3)$$

ここで、 $\mathbf{A} = \{i \mid \mathcal{R}(f(\theta^i, x + \zeta_1), f(\theta^i, x + \zeta_2)) = \text{True}\}$ であり、 $\mathcal{R}_x^i$ は入力 x に対し、摂動を付与した際の出力

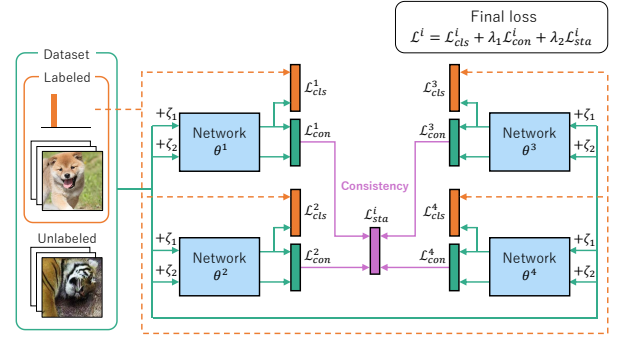


図 2: Refined Consistency の学習方法 (ネットワーク数 4)

$f(\theta^i, x + \zeta_1)$ と $f(\theta^i, x + \zeta_2)$ のクラスが一致した場合に真となる。

4. 評価実験

4.1. 実験概要

提案手法と MS の精度を比較する。データセットに CIFAR-100, ベースモデルに 13 層の CNN を使用する。学習回数は 350 epoch, ネットワーク数は 4, 8 とする。データ拡張として、左右反転, ランダムな画素数平行移動したのちに反射パディングを行う。CIFAR-100 は、クラス数が 100 であり、学習用に 50,000 枚、推論用に 10,000 枚の画像で構成される。本実験では、学習用画像のうち 10,000 枚をラベルありデータ、他 40,000 枚をラベルなしデータとする。

一貫性に基づいた半教師あり学習手法では、RandAugment(RA) [2] などの大きな摂動を付与するデータ拡張手法を扱うことで精度が向上する [3]。そこで、RA の有無による精度変化を評価する。

4.2. 実験結果

提案手法および MS の精度、各手法に対して RA を適用した精度を表 1 に示す。各結果は 5 回試行した時の正解率の平均値と標準偏差である。表 1 から、提案手法は MS と比べて、ネットワーク数 4 で 0.26 pt, ネットワーク数 8 で 0.31 pt 精度が高い。よって、提案手法における知識転移は有効であると言える。さらに、RA と提案手法を組み合わせることで、その効果は最大となり 68.18% まで向上した。

表 1: CIFAR-100 を用いた精度比較 [%]

手法	拡張	ネットワーク数	
		4	8
MS	-	66.70±0.15	66.93±0.09
	RA	66.43±0.14	66.34±0.10
提案手法	-	66.96±0.12	67.24±0.06
	RA	66.96±0.04	68.18±0.10

5. おわりに

本研究では、1度に全てのネットワーク間で知識を転移する Refined Consistency を提案した。評価実験では、提案手法が MS に比べて高精度であることを示した。今後は、知識転移グラフへの応用を行う予定である。

参考文献

- [1] K. Zhaghan, *et al.*, “Dual Student: Breaking the Limits of the Teacher in Semi-supervised Learning”, ICCV, 2019.
- [2] E. Cubuk, *et al.*, “RandAugment: Practical automated data augmentation with a reduced search space”, NeurIPS, 2019.
- [3] K. Sohn, *et al.*, “FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence”, NeurIPS, 2020.