Teach the way to deform: 教師モデルが持つ不変性の転移

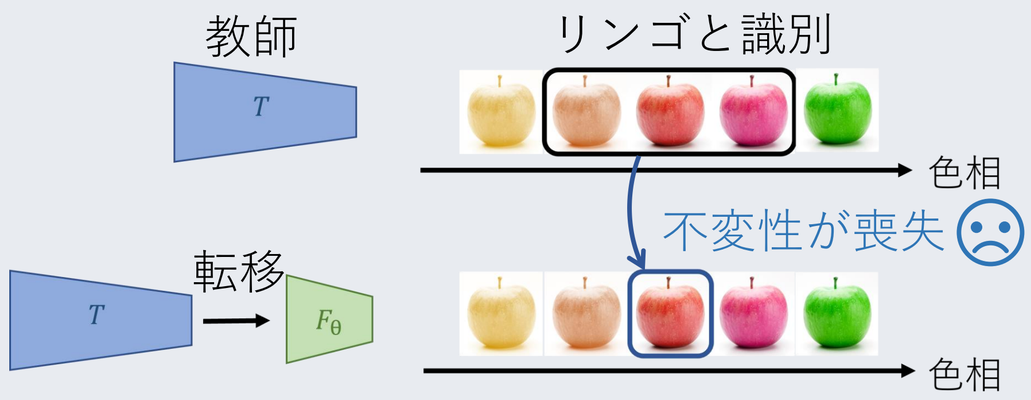
栗岡 保[1], 鈴木 哲平[2], 川上 玲[1], 佐藤 育郎[1,2]

[1]東京工業大学, [2]デンソーITラボラトリ

背景と本研究の目的

背景

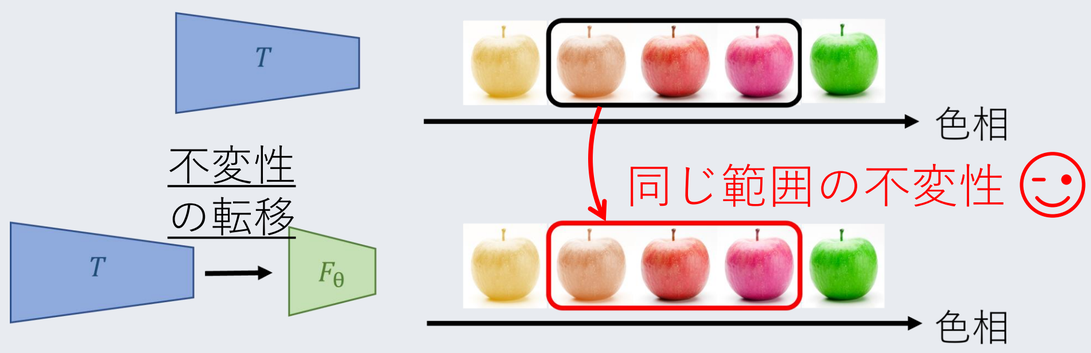
- 転移学習(Fine-tuning)は教師モデルが持つ不変性を必ずしも転移できない [Yamada et al., CVPR 2022]



- 無作為なデータ拡張は過剰な不変性をモデルに与えるおそれがある

目的

教師モデルが不変性を示す範囲でターゲットモデルに不変性を持たせる転移学習方法を開発する



提案する学習方法

学習指針

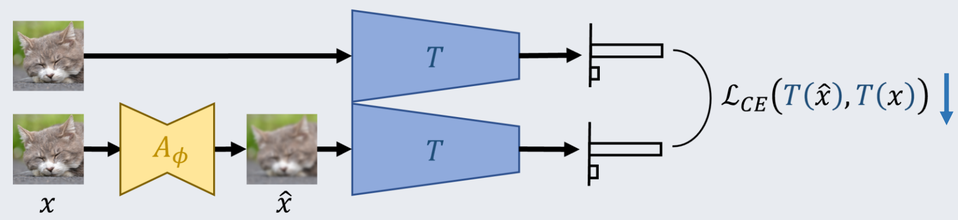
データ拡張器A_ϕとターゲットモデルF_θを同時最適化

$$\begin{cases} \phi^\star = \arg\min_\phi \left(\mathcal{L}_{CE}\big(T(\hat{x}), T(x)\big) - \mathcal{L}_{CE}\big(F_{\theta^\star}(\hat{x}), y\big)\right) \\ \theta^\star = \arg\min_\theta \mathcal{L}_{CE}\big(F_\theta(A_{\phi^\star}(x)), y\big) \end{cases}$$

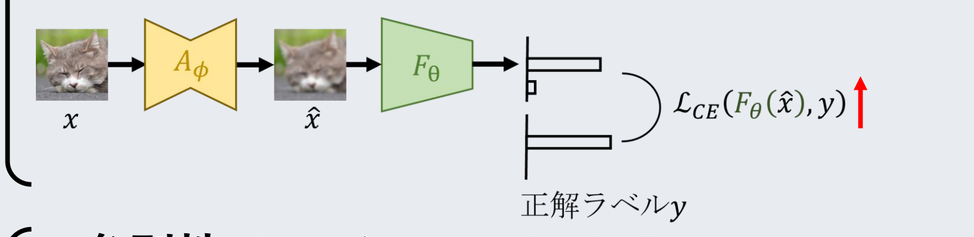
1. 不変性　2. 敵対性　3. 弁別性

データ拡張器への要求

1. 不変性 教師の出力の変化を最小化

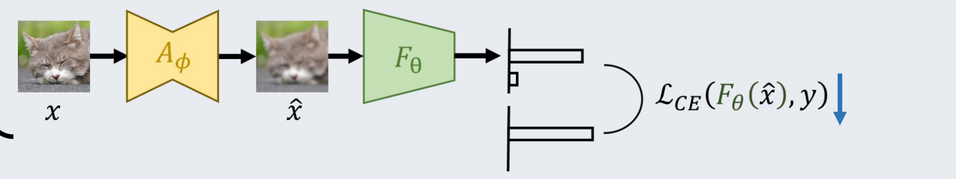


2. 敵対性 ターゲットモデルの識別損失を最大化



ターゲットモデルへの要求

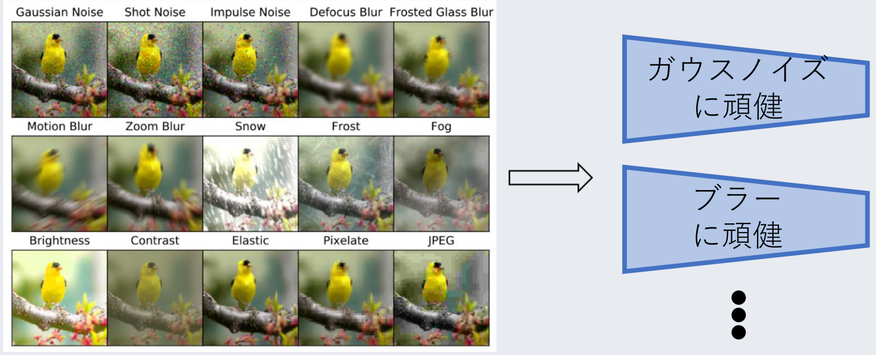
3. 弁別性 変形データの識別損失を最小化



不変性転移の評価実験

実験設定

教師モデル： 15種類の画像変形に対してそれぞれ頑健な教師モデルを生成



CIFAR-10- c

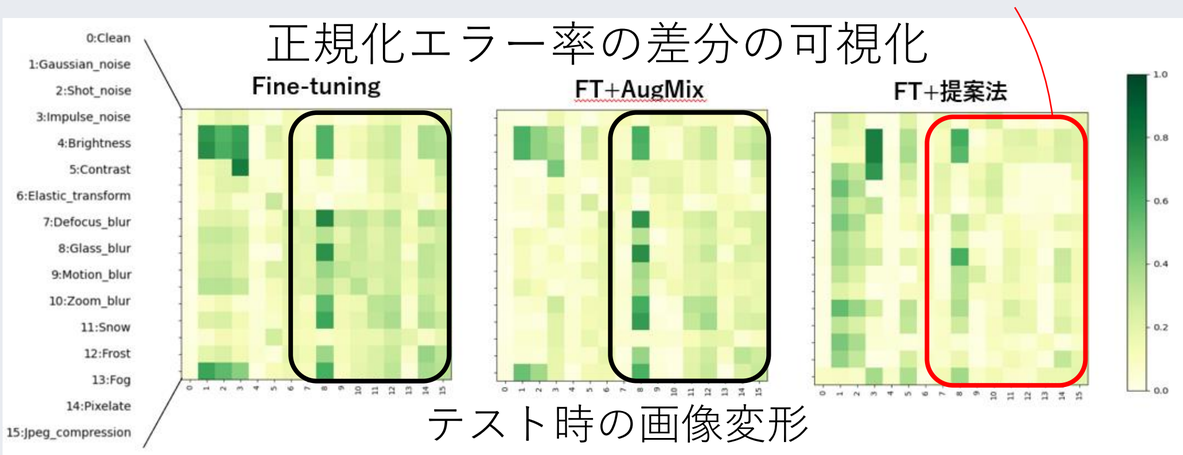
データ拡張器： 色,幾何,ガウスノイズ,ブラーの4種 適用確率と変換の強さを学習

ベースライン： 各教師モデルからのfine-tuning(FT), FT + AugMix

評価指標： 15種のモデルのCIFAR-100-Cでの各変形に対するエラー率を正規化 教師とターゲットの正規化エラー率差分の大きさを比較

実験結果

提案法で差分が小さい

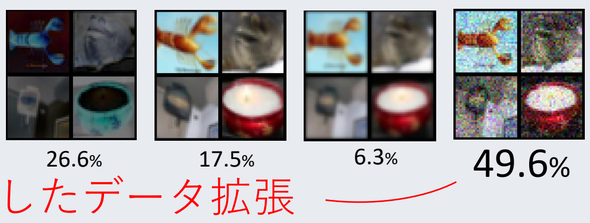


差分(L2ノルム)の値

FT	FT+AugMix	FT+提案法
0.972	0.778	**0.664**

提案法で差分が小さい
教師の不変性の転移が出来ている

例）Shot noiseに強い教師で得られたデータ拡張



shot noiseを反映したデータ拡張

本研究は東京工業大学「DENSO IT LAB 認識・学習アルゴリズム共同研究講座」における研究プロジェクト「機械学習の実応用に向けた品質基盤開発」の成果である.