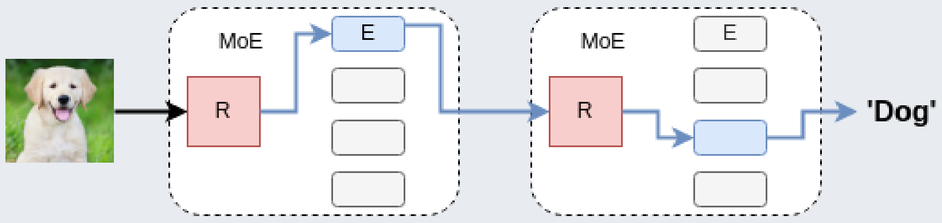
対照学習に基づく Mixture of Experts の経路表現学習

加太 将弘[1], 吉橋 亮太[1], 川上 玲[1], 池畑 諭[1,3], 佐藤 育郎[1,2]

[1]東京工業大学, [2]デンソー IT ラボラトリ, [3]国立情報学研究所

1. 背景: Deep Mixture of Experts (MoE)

概要 ルーターとエキスパートにより構成されるニューラルネットワーク



利点 エキスパート数を増やすことにより推論時間を一定に保ちつつモデルサイズを拡大可能

欠点 入力の微細な変化に対して出力が不連続に変化しうる

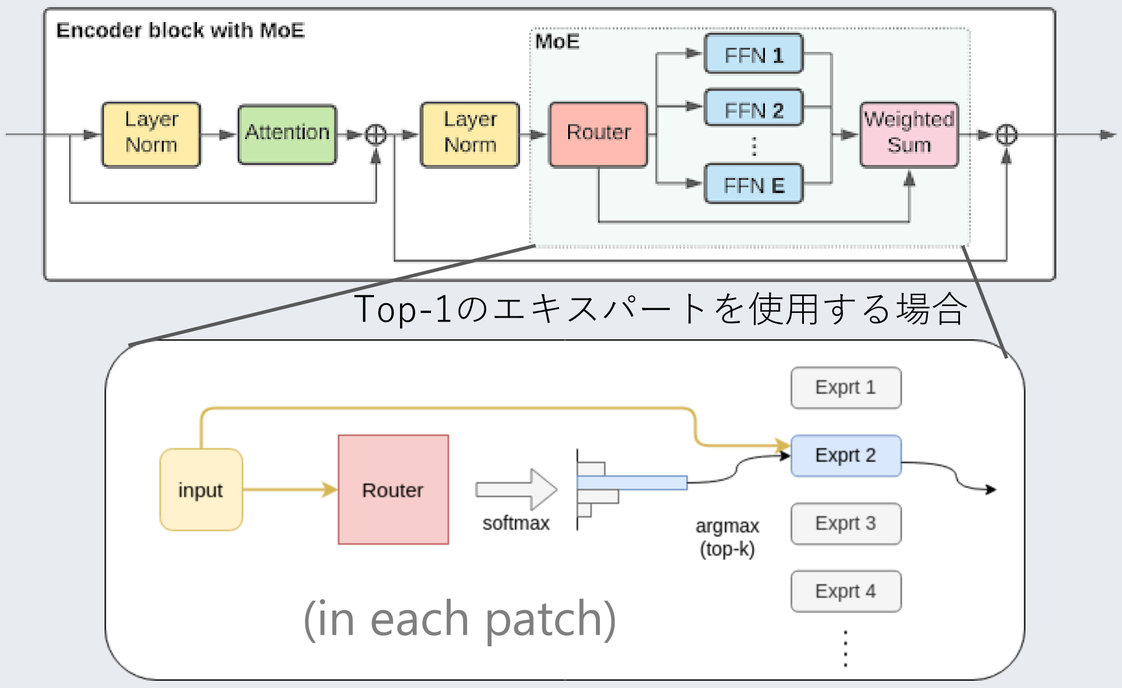
2. 貢献

- データ拡張に対してルーターの出力を頑健化する経路表現学習を提案
- Vision Transformer ベースの MoE の経路表現が入力の変形の影響を受けにくくなることを確認
- ImageNet-1K において従来手法と比較し精度を改善

3. 先行研究: Vision Mixture of Experts

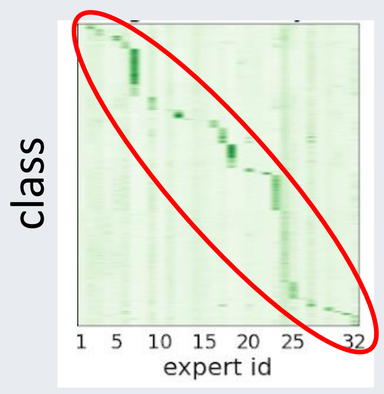
[C. Riquelme+, NeurIPS2021]

概要 Vision Transformer の MLP層を MoE 化



パッチごとにルーティングを行う．ルーター出力の softmax の argmax のエキスパートを選択する．

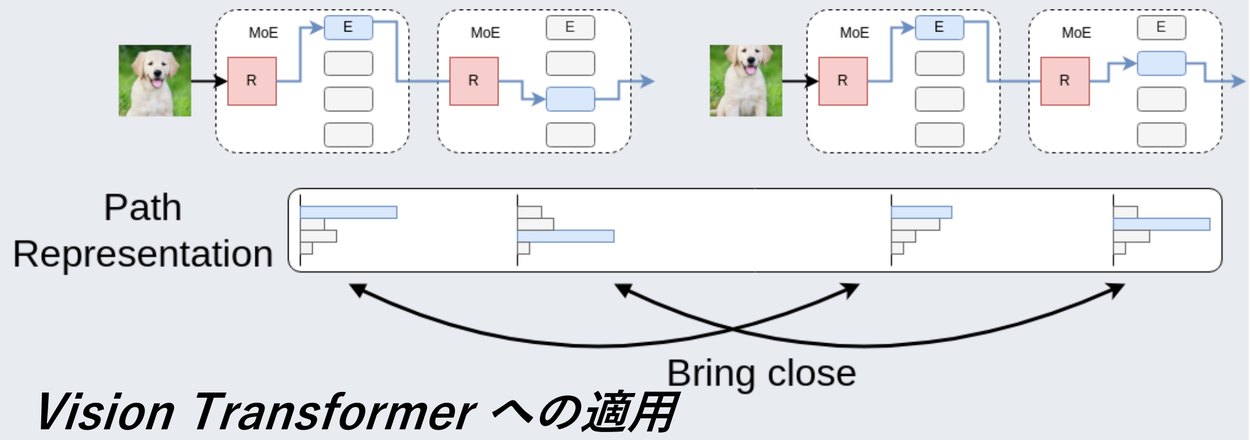
観察 エキスパートとラベルが概ね対応づいており，特定のラベルを主に処理するエキスパートが形成



4. 提案手法

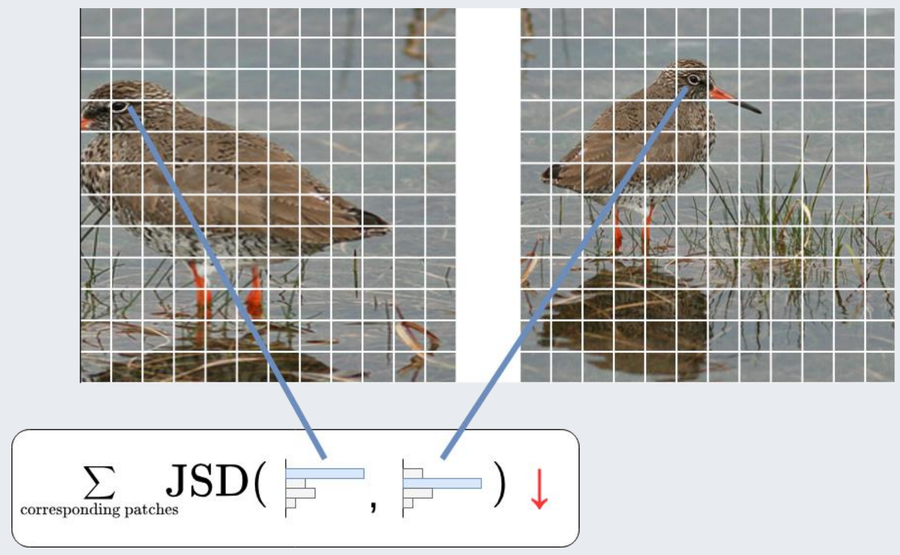
基本

データ拡張を用いて生成した 2 枚の画像のルーターの出力（経路表現）を近づける．



Vision Transformer への適用

2 画像間のパッチの対応を取り，ルーターの出力（経路表現）を近づける．



JSD: Jensen Shannon 距離

$$\mathrm{JSD}(p,q) = \frac{1}{2}\sum\left(p\log\frac{p}{M} + q\log\frac{q}{M}\right) \quad \text{, where } M = \frac{p+q}{2}$$

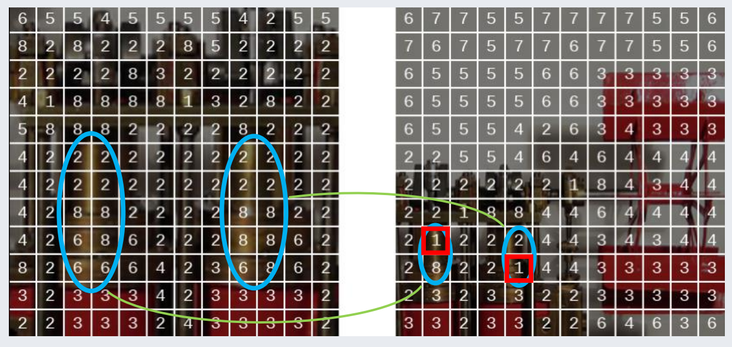
5. 実験

定性評価

従来手法と提案手法でルーティング結果を比較

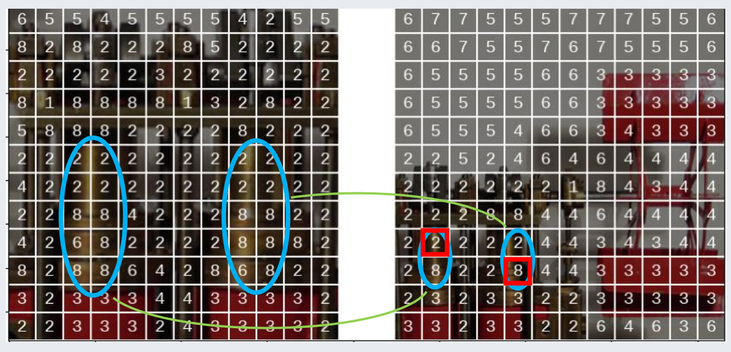
従来手法

対応するパッチ間でルーティングの出力が異なるものが存在



提案手法

対応するパッチ間でルーティングの出力が近づく



定量評価

[C. Riquelme+, NeurIPS2021] の結果を若干上回る

Dataset: ImageNet-1K

	Accuracy
従来手法	75.84%
提案手法	**75.89%**

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．