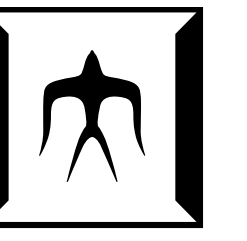


摂動に対するスパース混合エキスパートモデルの頑健化



東京工業大学
Tokyo Institute of Technology

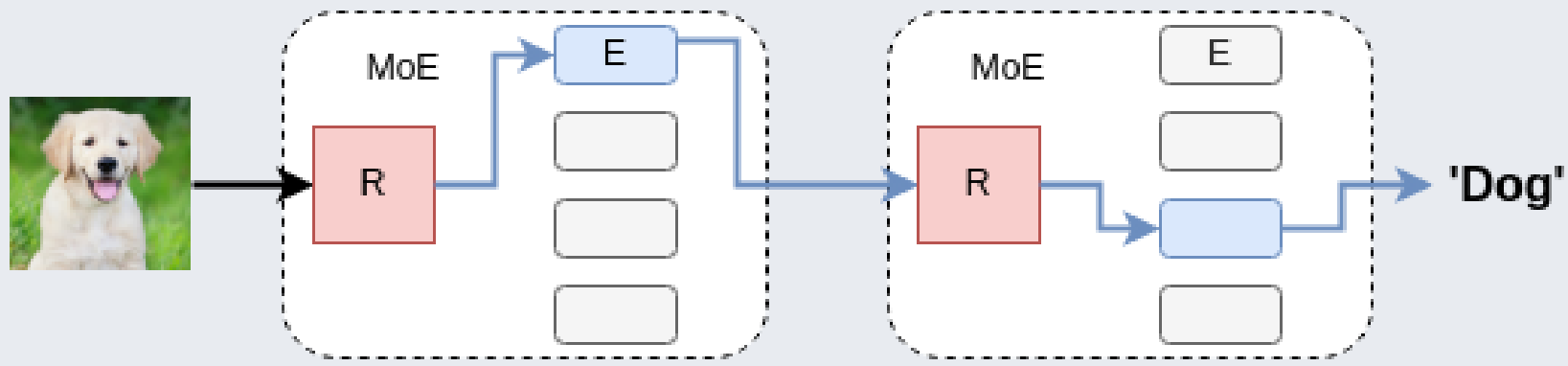
加太 将弘¹, 吉橋 亮太¹, 池畑 諭^{1,2}, 川上 玲¹, 佐藤 育郎^{1,3}

¹東京工業大学, ²国立情報学研究所, ³デンソー IT ラボラトリ



1. 背景: Sparse Mixture of Experts (MoE)

概要 特定のクラスの処理に特化した**エキスパート**と少数のエキスパート選択を行う**ルータ**を含むニューラルネットワーク

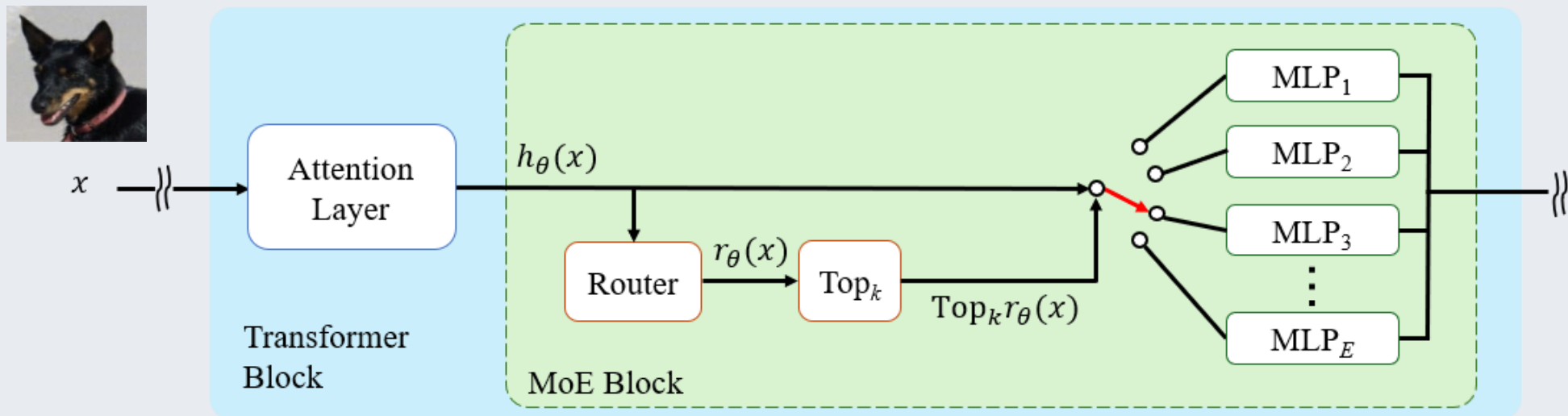


利点 エキスパート数を増やすことにより推論時間を一定に保ちつつモデルサイズを拡大可能

2. 従来手法: Vision MoE (V-MoE)

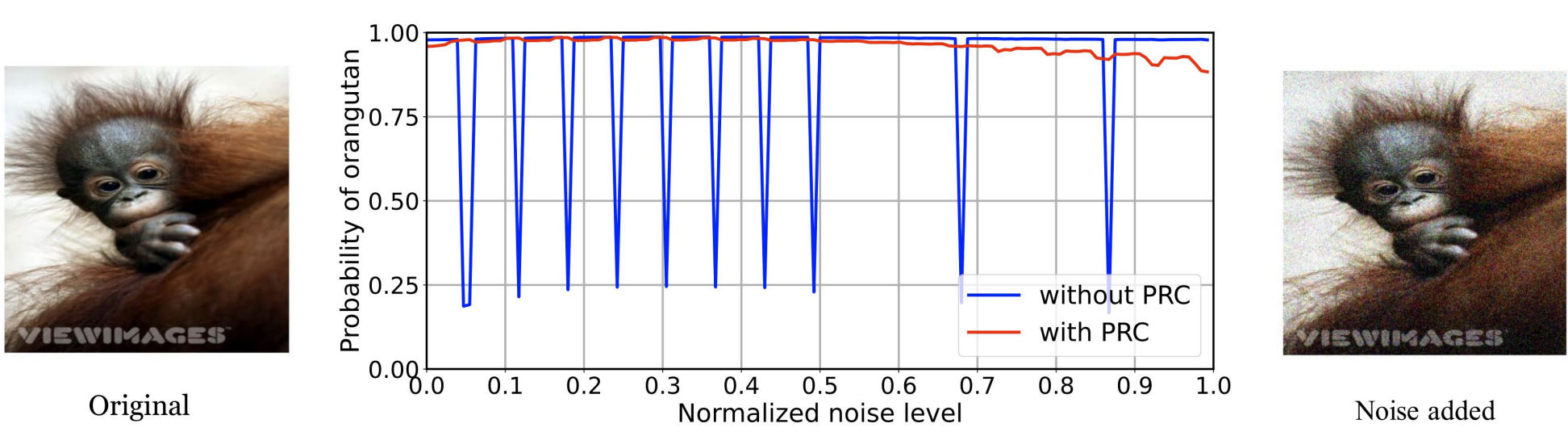
[C. Riquelme+, NeurIPS2021]

概要 Vision Transformer の MLP 層を MoE 化。ルータ出力 (softmax) の Top-k 個のエキスパートを選択する。



※ 画像パッチごとにルーティングを行う。

欠点 摂動に対して出力が不連続に変化する

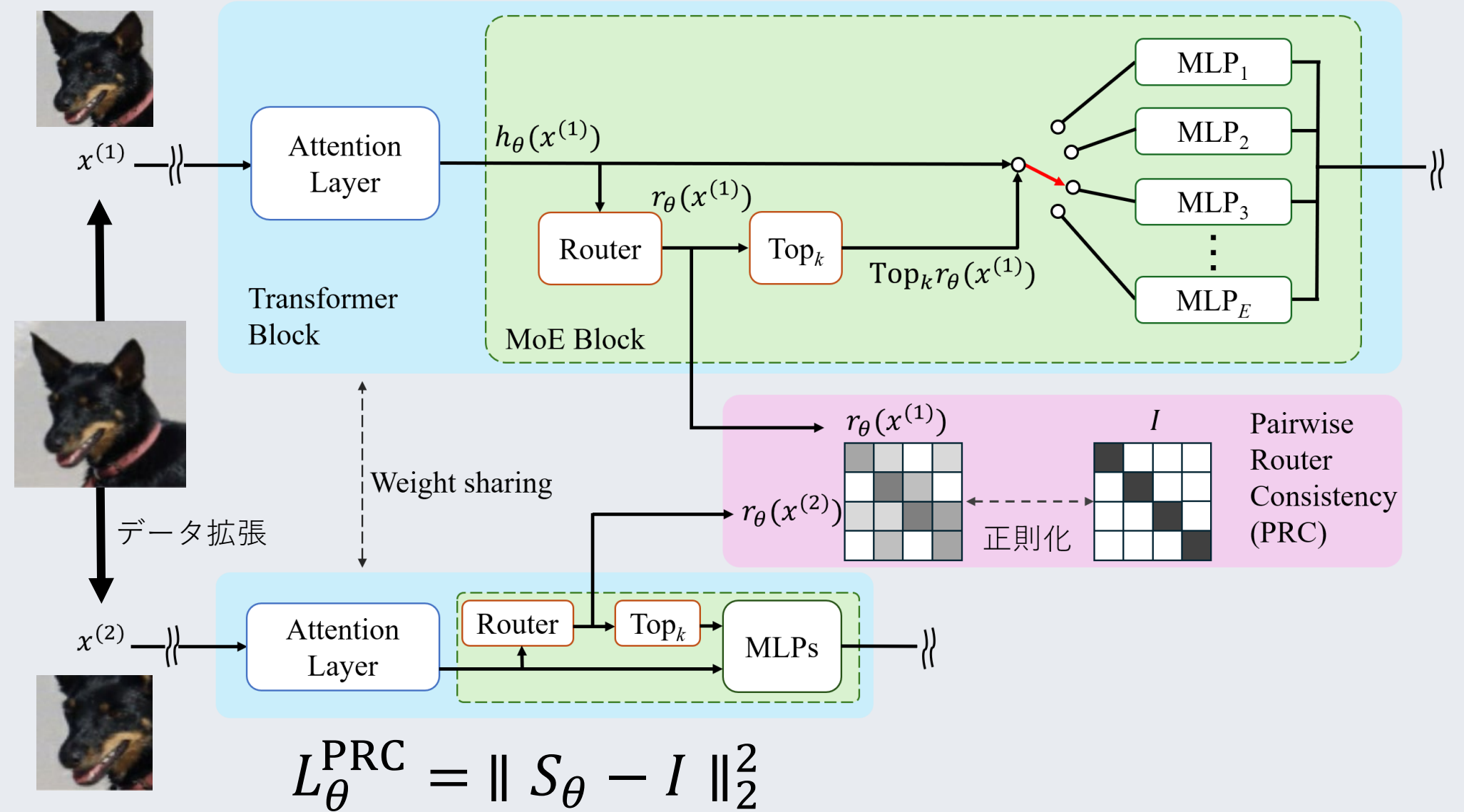


3. 提案手法: PRC

PRC: Pairwise Router Consistency

Pairwise Router Consistency (PRC)

摂動に対してルータ出力が不変となるようにデータ拡張された標本ペアのルータ出力を正則化

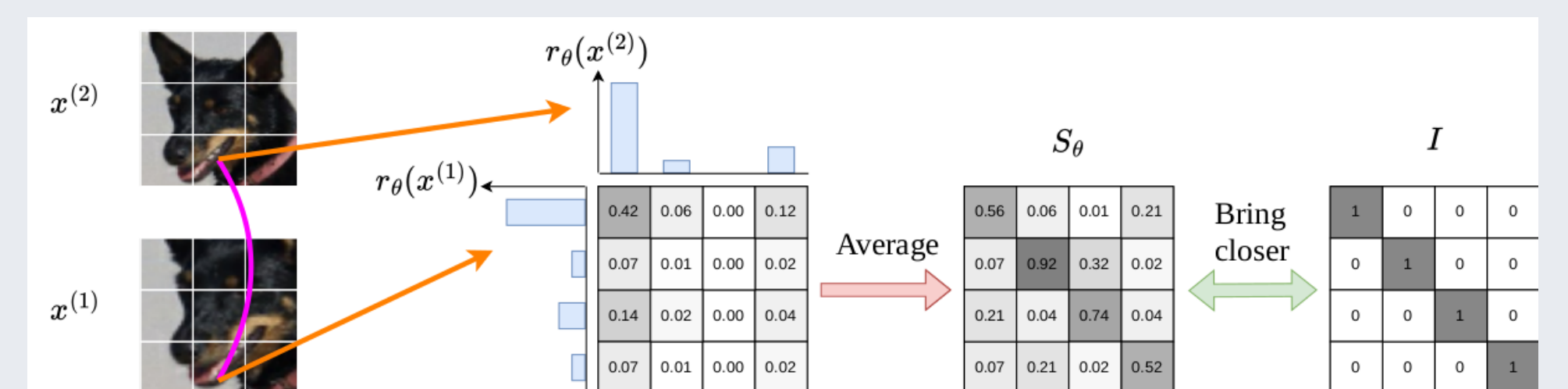


$$S_{\theta} = C \sum_{x \in X} r_{\theta}(x^{(1)}) r_{\theta}(x^{(2)})^T$$

(X: データセット, C: 正規化係数)

V-MoE への適用

データ拡張した 2 画像間でパッチ画像の対応を取り、対応するルータ出力を近づける



4. 実験

画像分類精度の評価結果

- PRC は従来手法の精度を上回る
- 選択エキスパート数 k を 1 とした PRC が, k を 2 とした従来手法の精度を上回った

ルーティングの解析結果

データ拡張の下で, PRC は従来手法と比較し, ルータ出力がより頑健に
→ 入力の微細な変化に対して出力が離散化する問題が緩和

表1. ImageNet-1K, CIFAR の画像分類精度

	ImageNet-1K	CIFAR-100	CIFAR-10
従来手法 (k = 1)	75.23 ± 0.01 %	81.18 ± 0.07 %	94.81 ± 0.06 %
提案手法 (k = 1)	75.92 ± 0.02 %	82.12 ± 0.18 %	95.12 ± 0.16 %
従来手法 (k = 2)	75.84 ± 0.04 %	81.38 ± 0.03 %	95.20 ± 0.07 %
提案手法 (k = 2)	76.27 ± 0.01 %	82.27 ± 0.10 %	95.36 ± 0.02 %

表2. x¹, x² 間での Top-k 個の index の要素の一致割合

	Top-1	Top-2	Top-2 (順不同)
従来手法	63.04 %	34.77 %	45.44 %
提案手法	75.78 %	48.54 %	58.96 %

5. 貢献

- データ拡張に対してルータの出力を頑健化する **Pairwise Router Consistency (PRC)** を提案
- PRC の適用によりルータが入力の変形の影響を受けにくくなることを確認
- ImageNet-1K, CIFAR の画像分類タスクにおいて従来手法と比較し精度を改善