

画像処理 DNN の領域重要度に基づく動的な計算効率化

Grad-CAM重要度予測による動的な画像処理領域の削減

- ▶ 深層学習モデル（ViT）を用いた画像処理において、効率化が課題
- ▶ 画像の各領域の重要度を動的に推定、重要でない領域を削除して効率化
- ▶ 不要領域の処理を削減した際の精度変化を評価・検証

1. 背景：Vision Transformerの計算コスト問題

- Vision Transformer(ViT)[1] は高精度な画像認識を実現する一方で、計算コストが非常に高く、実行時の膨大なエネルギー消費が課題である。
- Grad-CAM[3]は、生成された予測から生成結果における画像の各領域の寄与度を定量化するため、この手法を用いて予測の生成に大きく寄与しない画像の領域を処理から除去できる可能性がある。
- 本研究では、Grad-CAMの寄与度を実行時に予測し、画像の不要領域を削除するScorerを構築することで、課題の軽減を図る。

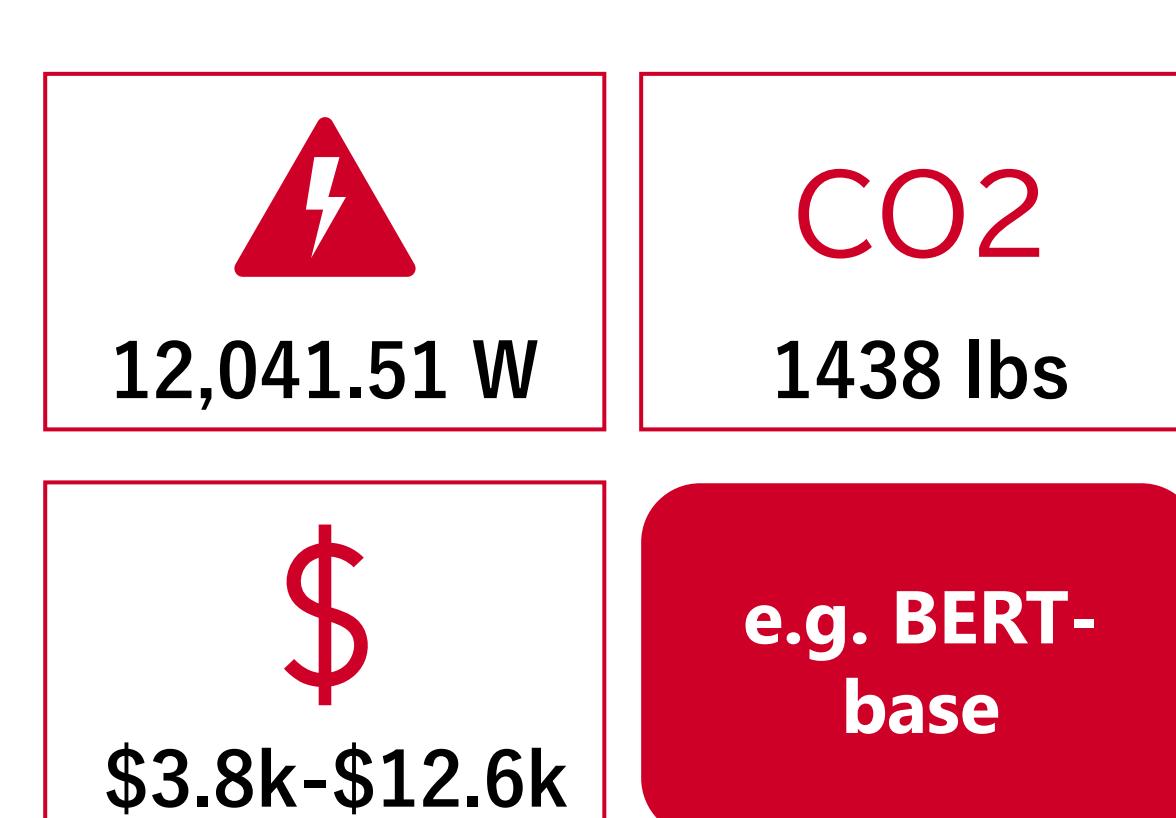


Fig.1: DNN のエネルギー消費量の例([2]をもとに作成)



Fig.2: Grad-CAM ([3]より引用)

2. 提案手法：Grad-CAMを用いた画像領域削除

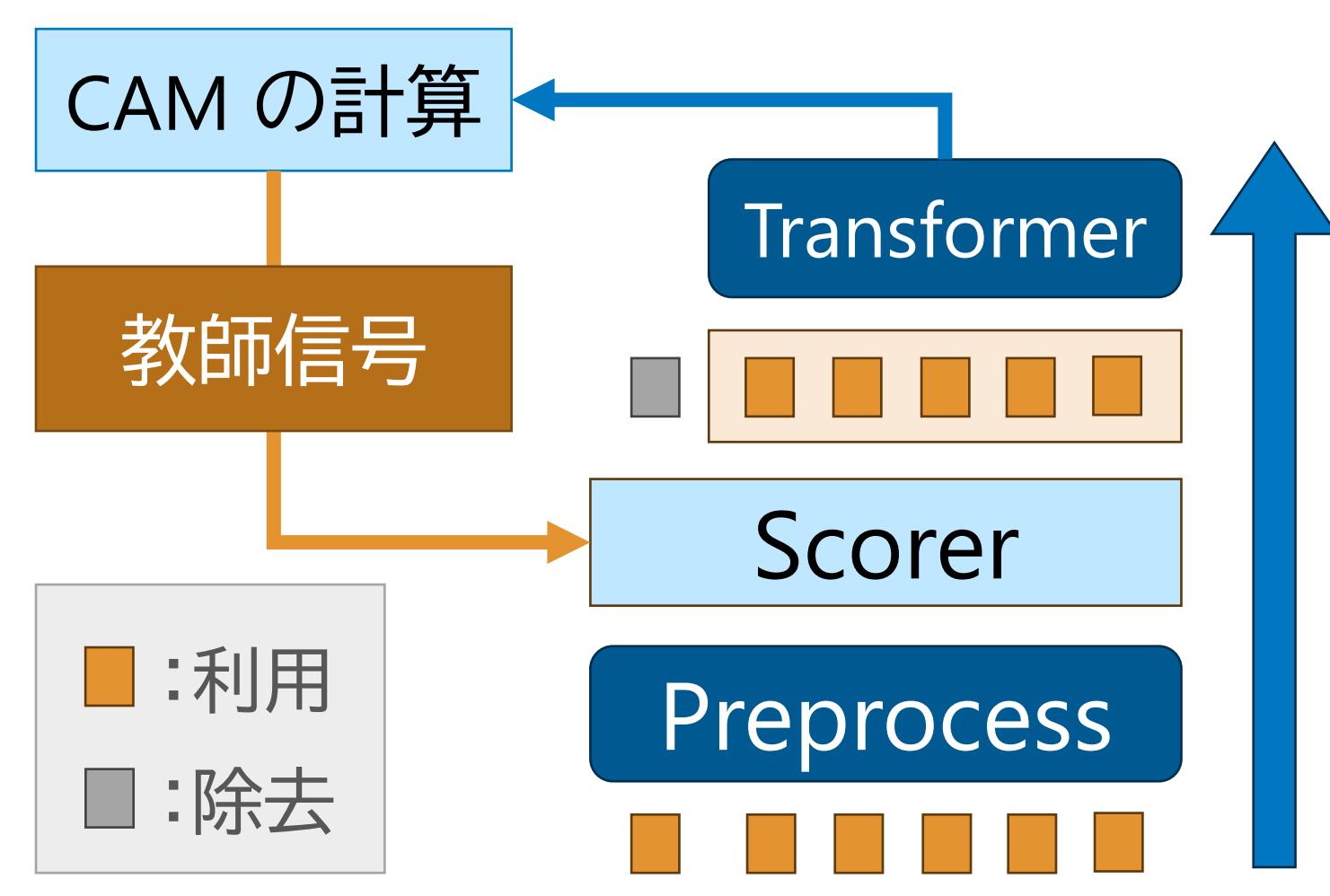


Fig. 3: 提案手法のモデル構成([5]より作成)

- Grad-CAM[3]により各領域の重要度を定量的に評価し、重要性の低い領域を削除することで計算効率を向上。
 - 学習フェーズでは、各領域のGrad-CAMの値を教師信号として学習することで、画像領域の重要度予測モデル:Scorerを構築する。
 - 推論フェーズではScorerが予測したGrad-CAMの値を利用し、重要性の低い領域を削除することで、Scorer以降の処理を軽量化する。

3. 実験と考察：精度とScorerによる削減結果の可視化

- Caltech101データセット[4]での実験により、削減率が高い場合において本手法の方がランダムな削減よりも高い精度を示した[Fig.4 左]
- 削除対象領域の可視化により、背景など不要な領域が効果的に除外され、意味的に重要な領域が保持されていることを確認[Fig.4 右]

人間社会拡張における意義

視覚的データを処理する上で重要性の高い領域を特定することで複数の画像や動画を一度に処理できる可能性を広げ、ウエアラブルデバイス等を用いた日常生活における非言語的な情報の利活用を促進したい

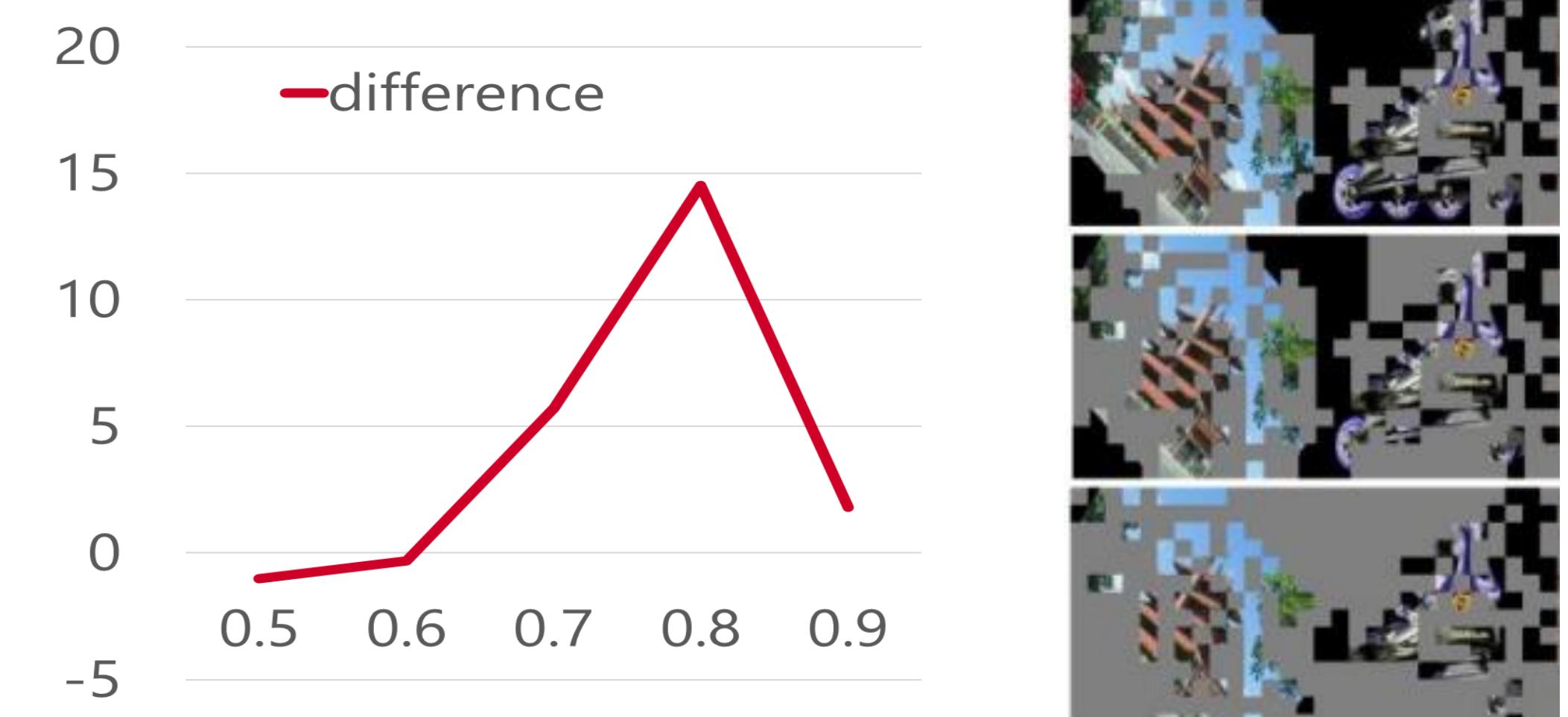


Fig. 4: 重要度に基づく処理領域削減—[左]削減率と予測精度増加量の変化(単位はポイント)[右]削減対象領域(グレイ部分)

【用語】

DNN: 深層ニューラルネットワーク(Deep Neural Network)の略
Vision Transformer: 画像処理用 DNN の一種

- [1] Dosovitskiy, Alexey. "An image is worth 16x16 words: Transformers for image recognition at scale." In arXiv, 2020.
- [2] Strubell, Emma, et al. "Energy and policy considerations for deep learning in NLP." Proceedings of the 57th annual meeting of the association for computational linguistics. 2019.
- [3] Selvaraju, Ramprasaath R., et al. "Grad-cam: Visual explanations from deep networks via gradient-based localization." Proceedings of the IEEE international conference on computer vision. 2017.
- [4] Fei-Fei, Li, et al. "Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories." 2004 conference on computer vision and pattern recognition workshop. IEEE, 2004.
- [5] Kiriyama, et al. "Token selection using Grad-CAM in Vision Transformer to reduce computational costs." Proceedings of the 2024 7th Artificial Intelligence and Cloud Computing Conference. 2024.