

GhostNet with Selective-MCA

日大生産工 ○小林 遼翔 日大生産工 山内 ゆかり

1. まえがき

現在, エッジデバイスやモバイルデバイスにおいて, 高効率かつ高性能なディープニューラルネットワークモデルの開発がますます重要視されている. これに応じて, Han らによって提案された GhostNet シリーズは, 計算コストを大幅に削減しつつ, 性能を向上させるための有効なアプローチを提供してきた. 特に, GhostNet v1[1]では, 特徴マップの冗長性を活用して計算効率を最適化し, 少数の基礎的な特徴マップを生成し, それを安価な操作で拡張する手法を導入した. この手法により, 計算コストを削減しつつ精度を維持することが可能となった.

本研究では, GhostNet を基に, Selective Multi-scale Convolutional Attention GhostNet という新たなアプローチを提案する. この手法は, 従来の GhostNet モデルと比較して, 計算コストをさらに抑えつつ, 高精度な特徴抽出を可能にする. 具体的には, MCA (Multi-scale Convolutional Attention)[2]を全ての層に適用するのではなく, 特定の層に限定して適用することで, モデルの効率と精度を両立させる.

2. 従来研究

2-1. CNN

CNN (畳み込みニューラルネットワーク) は, 畳み込み層, プーリング層, 全結合層から構成され, 主に画像認識に使われる深層学習アルゴリズムである. 畳み込み層で特徴を抽出し, プーリング層で不要な情報を省き, 全結合層で最終的な予測や分類を行う.

2-2. GhostNet

GhostNet は, 既存の CNN モデルに Ghost Moduleを統合させたものであり, 既存の CNN の精度を維持させ, モデルの大幅な軽量化が期待できるものである.

Ghost Moduleは, 通常の畳み込み層を2段階に分けて処理する. まず, 畳み込みで m 個の特徴マップを生成し, それに単純な線形演算を加えて s 個の特徴マップを得る. これらを組み合わせることで, 冗長性を持つ n ($n=s \times m$) 個の特徴マップを生成する. GhostNetは, このModuleによ

り従来のCNNモデルと比べて, 計算速度とモデル圧縮率が s 倍向上するとされている. 式(1)では高速化率, 式(2)では圧縮率を示す.

$$r_s = \frac{c \times k \times k}{\frac{1}{s} \times c \times k \times k + \frac{s-1}{s} \times d \times d} \quad (1)$$

$$\approx \frac{\frac{s \times c}{s+c-1}}{\frac{s \times c}{s+c-1}} \approx s$$

$$r_c = \frac{n \times c \times k \times k}{\frac{n}{s} \times c \times k \times k + (s-1) \times d \times d} \quad (2)$$

$$\approx \frac{\frac{s \times c}{s+c-1}}{\frac{s \times c}{s+c-1}} \approx s$$

Ghost bottleneckはResNetの残差ブロックに似ており, 複数の畳み込み層とショートカットが統合されている.

主に2つのGhost Moduleから構成され, 1つ目はチャンネルを増やす拡張層として, 2つ目はショートカットに合わせてチャンネルを減らす役割を果たす. ショートカットはこれらのGhost Moduleの入力と出力を結び, Batch正規化とReLU関数が各層後に適用されますが, 2つ目のGhost Module後にはReLUは適用されない. 下にGhostNet v1のGhost bottleneck図を示す.

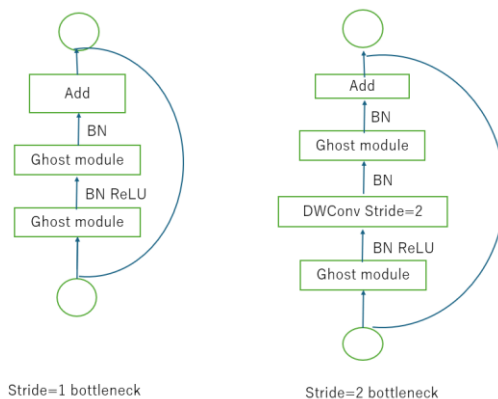


Figure 1 GhostNet v1 の Ghost bottleneck

GhostNet 全体の構成は, 基本的には MobileNetV3 の bottleneck ブロックを Ghost bottleneckに置き換える形で構成されている.

最初のレイヤーは16のフィルターを持つ標準的な畳み込みレイヤーで、その後、徐々にチャンネルを増やした一連のGhost bottleneckが続く。

最後に、特徴マップを特徴ベクトルに変換するためにGlobal Average Pooling(GAP)と畳み込み層が利用される。

2-3. MCA

MCA (Multi-scale Convolutional Attention) は、画像認識や物体検出などのコンピュータビジョンタスクにおいて、異なるサイズの畳み込みフィルターを通してマルチスケールで特徴を抽出する手法である。この特徴抽出プロセスでは、小さなフィルターが詳細な局所的特徴を捉え、大きなフィルターが広い範囲のコンテキスト情報を捉えることで、異なるスケールの情報を並行して取得する。

まず、クエリ、キー、バリュー行列を得るために 3 つの 1×1 畳み込みが使用される。従来の Transformer の Attention Mechanism[3]に従い、クエリとキーの内積が計算され、softmax 関数が適用されて注意行列「Atten_conv」が導かれる。

$$Atten_{conv} = softmax(query \times value^T) \quad (3)$$

$$f_{atten} = F + Drop(Conv_{1 \times 1}(Atten_{conv} \times value)) \quad (4)$$

最終的に、異なるスケールで抽出された特徴を統合し、強化された特徴マップを生成する。この統合により、小さなディテールから大きな構造まで、幅広い情報がバランス良く反映された特徴表現が得られる。

3. 提案手法

GhostNet with Selective MCA モデルは、最初に CIFAR-10 データセットを読み込み、正規化した後、バッチ分割し効率的に処理する。浅い層では Ghost Module を用いて基本的な特徴を抽出し、中間層に MCA を適用することでマルチスケールの特徴を捉える。前進計算後、クロスエントロピー損失関数を使って誤差を計算し、逆伝播でパラメータを更新する。トレーニング後、テストデータでモデルを評価し、CNN と GhostNetv1 と全ての bottleneck 層にて MCA

を適用する場合と Selective MCA の効率や精度を比較し最適化する。

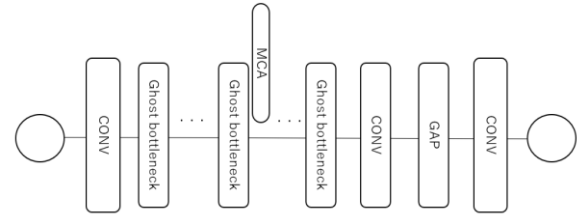


Figure 2 GhostNet with Selective-MCA

4. 実験および検討

本提案では、 32×32 ピクセルのカラー画像データであるCIFAR-10データセットを用いて選択的MCA GhostNetモデルを構築する。精度、効率、速度をCNNやGhostNetv1、全てのbottleneck層にてMCAを適用するモデルと比較し、提案手法であるSelective MCAの効果を評価する。また、正則化を活用して過学習を防ぎ、結果に基づいて最適化を行う。これにより、パフォーマンス向上を図り、モデルの効率性と精度のバランスを検討する。

5. まとめ

本研究では、異なるスケールの特徴を同時に抽出し、畳み込み自己注意によって重要な情報を強調・統合することで、効率的かつ高精度な特徴マップを生成する。この手法により、画像認識や物体検出の精度が大幅に向上し、さまざまなスケールの情報をバランスよく捉えることができ、計算コストが削減することを目指す。

参考文献

- 1) Kai Han, Yunhe Wang, Qi Tian, Jianyuan Guo, Chunjing Xu, Chang Xu, "GhostNet: More Features from Cheap Operations", Proc. CVPR (2020)
- 2) Xiaohang Hu, Seungjun Yang, Kyungeun Cho, "CAGNet: A Multi-Scale Convolutional Attention Method for Glass Detection Based on Transformer", Mathematics 11(19) 4084 (2023)
- 3) Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin, "Attention Is All You Need", Proc. NIPS (2017)