Deep Residual Learning for Image Recognition

A critical review

Abhishek Ranjan

March 9, 2020

Toronto, ON

Table of contents

Problem Statement

Deep NN Issues

1. Vanishing/exploding gradients, mostly addressed by:
    - BN in forward propagation
    - ReLU/PReLU + weight initialization
2. Training degradation, empirically observed by:
    - Increasing depth $\not\Longrightarrow$ High accuracy
    - Difficulty in learning Identity functions

Can we increase depth without degrading accuracy?

Proposed solution

How about forcing Identity shortcuts to blocks?

- Add a direct path (Residual connection) for x to reach the output
- Let the original block learns $H(x)$, with the shortcut it learns $F(x) = H(x) - x$

Benefits of residuals

- Residuals are easier to learn (evidenced by on learnings from other areas)
- Solver can move the weights to zero more easily than to identity mapping

Architecture

Plain network

Residual connections [1, 2]

Training parameters

Experiments and Results

Accuracy

Layer responses

References i

- He, Kaiming, et al. "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

- He, Kaiming, et al. "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

Residuals as shortcuts

An ensemble of shallow networks

ResNet as feature boosting