

Wide Residual Networks

Gobind Puniani
MATH 190: Fall 2019
10/24/19

Wide vs. Deep

- “Widening” residual networks can potentially improve performance better than deepening
 - Fewer layers, less training time
 - Doubling number of parameters is better than adding a thousand more layers
- Deepening networks yields diminishing accuracy gains
 - Longer time to train
 - Additional identity mappings may not enhance learning

Experiment

- Even a 16-layer WRN can outperform thousand-layer deep networks in accuracy and efficiency
- Record performances on CIFAR, SVHN, and COCO
 - 16-layer WRN with dropout achieves 1.64% error on SVHN
 - Improvements for ImageNet

Important Equations

- Residual block with identity mapping:
$$x_{l+1} = x_l + F(x_l, W_l)$$

Conclusion

- Widened architecture improved accuracy and efficiency
 - State-of-the-art results
- Dropout in identity part of ResNet block may have adverse effects (e.g. overfitting)
- Dropout should be inserted between convolutional layers

Questions

- What is the difference between a residual function and an activation function?