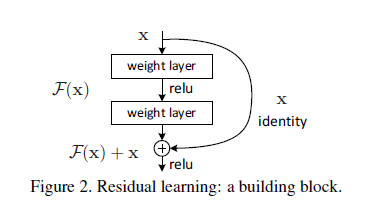
Deep Residual Learning for Image Recognition

1. Introduction

When deeper networks are able to start converging, a degradation problem has been exposed: with the network depth increasing, accuracy gets saturated (which might be unsurprising) and then degrades rapidly.



1. Related Work

Residual Representations.

Shortcut Connections.

1. Deep Residual Learning
   1. Residual Learning

If the optimal function is closer to an identity mapping than to a zero mapping, it should be easier for the solver to find the perturbations with reference to an identity mapping, than to learn the function as a new one.

* 1. Identity Mapping by Shortcuts
  2. Network Architectures

Plain Network.

(i) for the same output feature map size, the layers have the same number of filters;

and (ii) if the feature map size is halved, the number of filters is doubled so as to preserve the time complexity per layer.

Residual Network.

* 1. Implementation

1. Experiments
   1. ImageNet Classification
   2. CIFAR10 and Analysis

4.3. Object Detection on PASCAL and MS COCO