Residual Networks of Residual Networks: Multilevel Residual Networks

IEEE Transactions on Circuits and Systems for Video Technology, 2017

Ke Zhang, Miao Sun, Tony Han, Xingfang Yuan, Liru Guo, Tao Liu

Department of Electronic and Communication Engineering, North China Electric Power University, Baoding, Hebei, China

Residual Networks of Residual Networks: Multilevel Residual Networks

Introduction

A residual-networks family with hundreds or even thousands of layers dominates major image recognition tasks, but building a network by simply stacking residual blocks inevitably limits its optimization ability. This paper proposes a novel residual-network architecture, Residual networks of residual networks (RoR), to dig the optimization ability of residual networks. ResR subnetworks optimizing residual mapping of residual mapping for optimizing original residual mapping. In particular, RoR adds level-wise shortcut connections upon original residual networks to promote the learning capability of residual networks. More importantly, RoR can be applied to various kinds of residual networks (ResNets, Pre-ResNets and WRN) and significantly boost their performance. Our experiments demonstrate the effectiveness and versatility of RoR, where it achieves the best performance in all residual-network-like structures. Our RoR-3-WRN58-4+SD models achieve new state-of-the-art results on CIFAR-10, CIFAR-100 and SVHN, with test errors 3.77%, 19.73% and 1.59%, respectively. RoR-3 models also achieve state-of-the-art results on ImageNet data set.

Methods

Architectures of RoR

RoR is based on a hypothesis: To dig the optimization ability of residual networks, we can optimize the residual mapping of residual mapping. So we add shortcuts level by level to construct RoR based on residual networks.

Optimization of RoR

- Shortcut level number of RoR

It is important to choose a suitable number of RoR levels for a satisfying performance. The more shortcut layers chosen, the more branches and parameters are added. The overfitting problem will be exacerbated, and the performance may decrease. However, RoR improvements will be less obvious if the number of levels is too small. So we must find a suitable number to keep the balance, So we choose \( n \).

- Identity Mapping Types of RoR

We all used Type A in the first short level, and Type B in the other shortcut levels.

- Maximum Epoch Number of RoR

In this paper, we choose 500 as the maximum epoch number for RoR and Steepness of Depth.

- Drop Path by Steepest Depth

Overfitting can be a critical problem for the CIFAR-100 data set. Adding extra shortcuts to the original residual network can cause the overfitting problems to be even more severe. So in this paper we use the steepest drop path method in our RoR except for the ImageNet data set, and it can significantly alleviate overfitting, especially on the CIFAR-100 data set.

Results

CIFAR-10 and CIFAR-100 Classification by RoR

<table>
<thead>
<tr>
<th>Depth</th>
<th>CIFAR-10 Pre-RoR-3</th>
<th>CIFAR-10 Pre-RoR-3+SD</th>
<th>CIFAR-10 Pre-RoR-3+SD</th>
<th>CIFAR-10 Pre-RoR-3+SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>15-laver</td>
<td>3.43</td>
<td>3.37</td>
<td>3.33</td>
<td>3.32</td>
</tr>
<tr>
<td>64-laver</td>
<td>4.51</td>
<td>22.94</td>
<td>22.94</td>
<td>22.94</td>
</tr>
<tr>
<td>128-laver</td>
<td>4.49</td>
<td>20.64</td>
<td>20.64</td>
<td>20.64</td>
</tr>
</tbody>
</table>

Depth and Width Analysis

The performance can be improved by increasing depth or width.

Depth

<table>
<thead>
<tr>
<th>Depth</th>
<th>CIFAR-10 Pre-RoR-3</th>
<th>CIFAR-10 Pre-RoR-3+SD</th>
<th>CIFAR-10 Pre-RoR-3+SD</th>
<th>CIFAR-10 Pre-RoR-3+SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>15-laver</td>
<td>3.43</td>
<td>3.37</td>
<td>3.33</td>
<td>3.32</td>
</tr>
<tr>
<td>64-laver</td>
<td>4.51</td>
<td>22.94</td>
<td>22.94</td>
<td>22.94</td>
</tr>
<tr>
<td>128-laver</td>
<td>4.49</td>
<td>20.64</td>
<td>20.64</td>
<td>20.64</td>
</tr>
</tbody>
</table>

Conclusions

This paper proposes a novel residual networks architecture (RoR), which was proved capable of obtaining a new state-of-the-art performance on CIFAR-10, CIFAR-100, and ImageNet with ImageNet for image classification. Through empirical studies, this work not only significantly advanced the image classification performance, but also provided an effective complement to the residual-networks family in the future. In other words, any residual network can be improved by RoR. Hence, RoR has a good prospect of successful application on various image recognition tasks.