

Enriching Variety of Layer-wise Learning Information by Gradient Combination

Chien-Yao Wang¹, Hong-Yuan Mark Liao¹, Ping-Yang Chen², and Jun-Wei Hsieh³

¹Institute of Information Science, Academia Sinica, Taiwan

²Department of Computer Science, National Chiao Tung University

³College of Artificial Intelligence and Green Energy, National Chiao Tung University

x102432003@yahoo.com.tw, liao@iis.sinica.edu.tw,

pingyang.cs08g@nctu.edu.tw, and jwhsieh@nctu.edu.tw

Abstract

This study proposes to use the combination of gradient concept to enhance the learning capability of Deep Convolutional Networks (DCN), and four Partial Residual Networks-based (PRN-based) architectures are developed to verify above concept. The purpose of designing PRN is to provide as rich information as possible for each single layer. During the training phase, we propose to propagate gradient combinations rather than feature combinations. PRN can be easily applied in many existing network architectures, such as ResNet, feature pyramid network, etc., and can effectively improve their performance. Nowadays, more advanced DCNs are designed with the hierarchical semantic information of multiple layers, so the model will continue to deepen and expand. Due to the neat design of PRN, it can benefit all models, especially for lightweight models. In the MSCOCO object detection experiments, YOLO-v3-PRN maintains the same accuracy as YOLO-v3 with a 55% reduction of parameters and 35% reduction of computation, while increasing the speed of execution by twice. For lightweight models, YOLO-v3-tiny-PRN maintains the same accuracy under the condition of 37% less parameters and 38% less computation than YOLO-v3-tiny and increases the frame rate by up to 12 fps on the NVIDIA Jetson TX2 platform. The Pelee-PRN is 6.7% mAP@0.5 higher than Pelee, which achieves the state-of-the-art lightweight object detection. The proposed lightweight object detection model has been integrated with technologies such as multi-object tracking and license plate recognition, and is used in a commercial intelligent traffic flow analysis system as its edge computing component. There are already three countries and more than ten cities have deployed this technique into their traffic flow analysis systems.

1. Introduction

Since AlexNet [6] won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [1] in 2012, the era of

Deep Convolutional Networks (DCN) has begun. After that, the architecture of Plain networks (PlainNet) makes VGG-16 and VGG-19 [16] achieve excellent performance. But the researchers also found that when a DCN reaches a certain depth, the accuracy of PlainNet begins to decrease as the depth increases. This problem is uneasy to solve, even if GoogLeNet [18] adopts the strategy of auxiliary loss, it still cannot solve the problem. In the design of Residual networks (ResNet), He et al. [2] introduce the concept of identity shortcut connection, which attempts to make gradients more efficiently propagate to all layers. Using this concept, they successfully built a number of very deep architectures, such as ResNet-50, ResNet-101, and ResNet-152. The highway networks [17] published before ResNet and the stochastic depth [5] developed after ResNet all adopted the shortcut concept. Also want to let gradients quickly propagate to the various layers, Larsson et al. [7] used the characteristics of fractal to achieve the goal. They also proposed the concept of drop path to increase the efficiency of information propagation. In [4] and [3], Huang et al. proposed dense networks (DenseNet) and condense networks (CondenseNet), respectively, to directly let the loss layer link to all layers. However, the sparse networks (SparseNet) proposed by Zhu et al. [24] have confirmed that DCN can achieve better learning results under a designed topology architecture.

From the above state-of-the-art work, we found that the strategy to improve DCN performance is nothing more than how to combine features and propagate to subsequent layers, and how to make gradients more efficiently propagate to all layers. After an in-depth examination of the learning process of the above mentioned models, we propose a new perspective, i.e., "How to combine gradients of each layer in the training process to achieve better learning results?" In this study, we propose the concept of partial residual networks (PRN). We converted the residual connection into a path that produces the combinations of gradients and designed several PRN-based models to verify the proposed concept. Since the strategy for designing PRN is no longer to prop-

