西南财經大學
SOUTHWESTERN UNIVERSITY OF FINANCE AND ECONOMICS

基础知识：残差神经网络

相关论文：Deep Residual Learning for Image Recognition等

主讲人：阎根伟

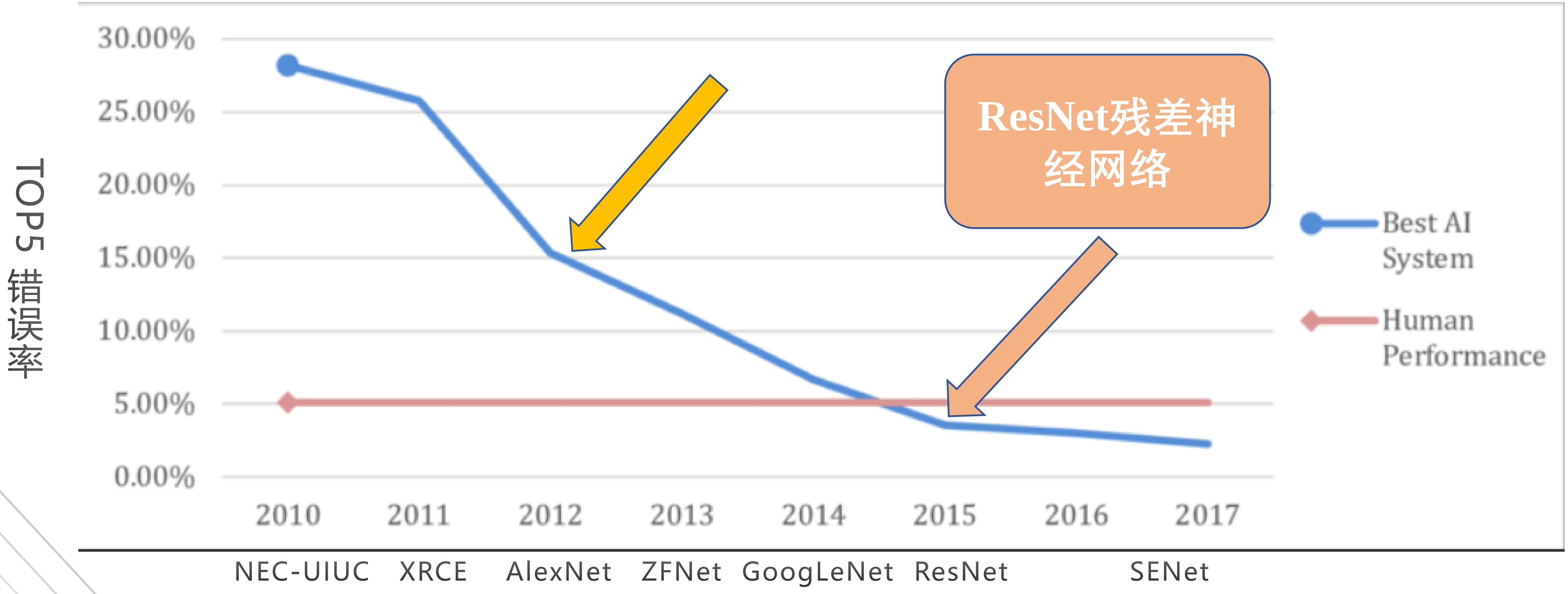
2024.4.17

研究背景（ILSVRC-ImageNet竞赛）

两个关键时间点：

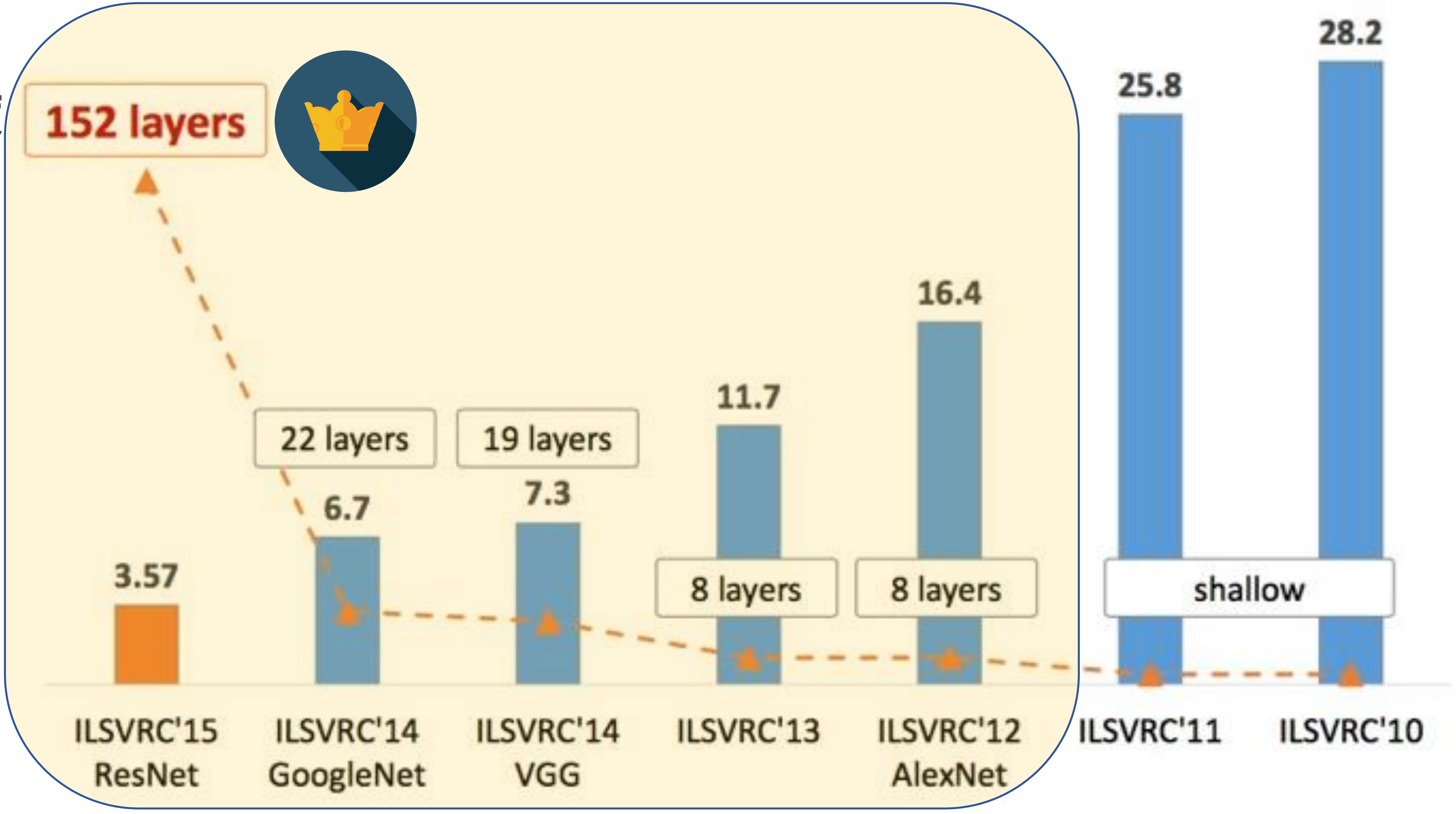
2012年Alexnet将深度学习和卷积神经网络用在了计算机视觉

2015年深度残差网络Resnet超过人类的水平



研究背景（ILSVRC-ImageNet竞赛）

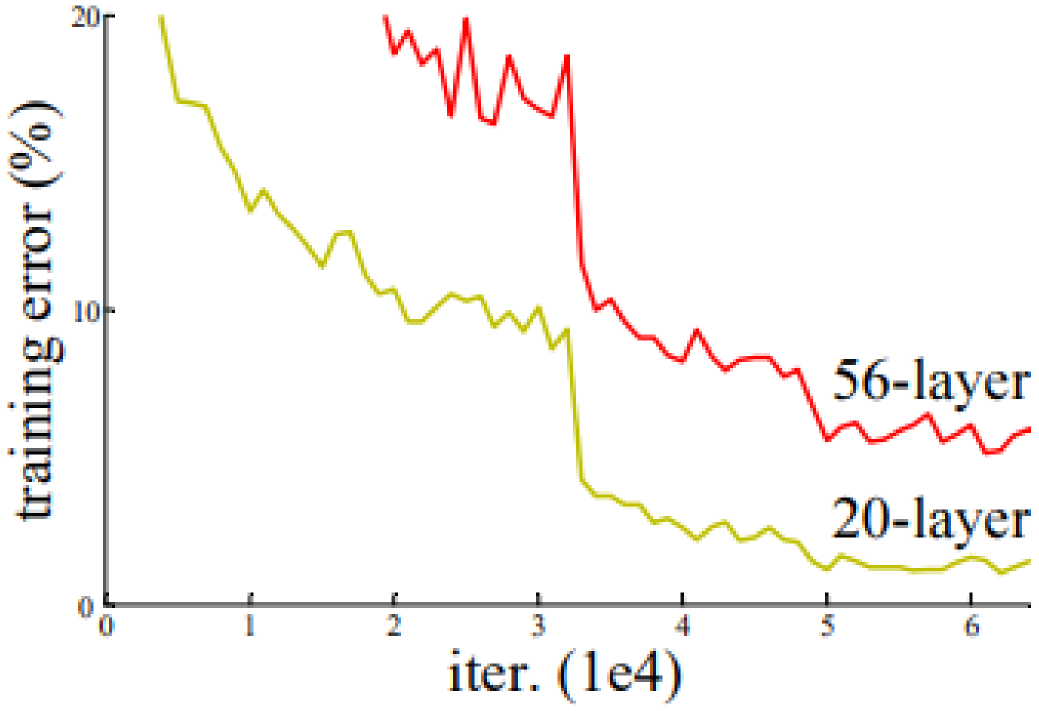
深度



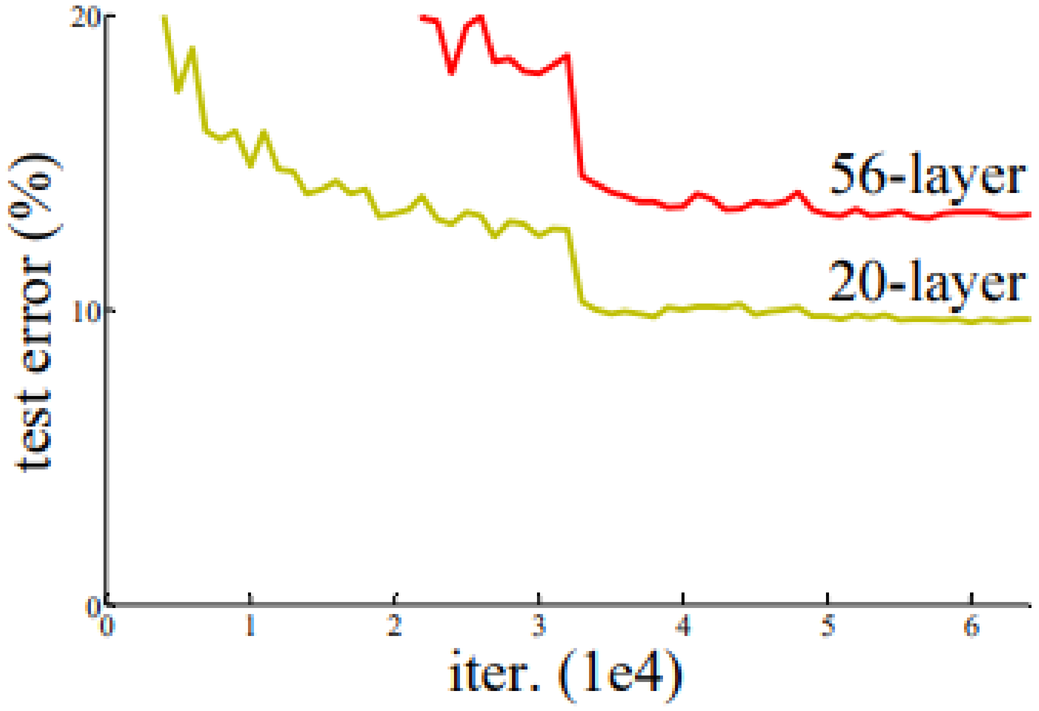
TOP5错误率

ResNet核心解决问题

网络退化现象：网络变深以后，性能还不如浅层的网络

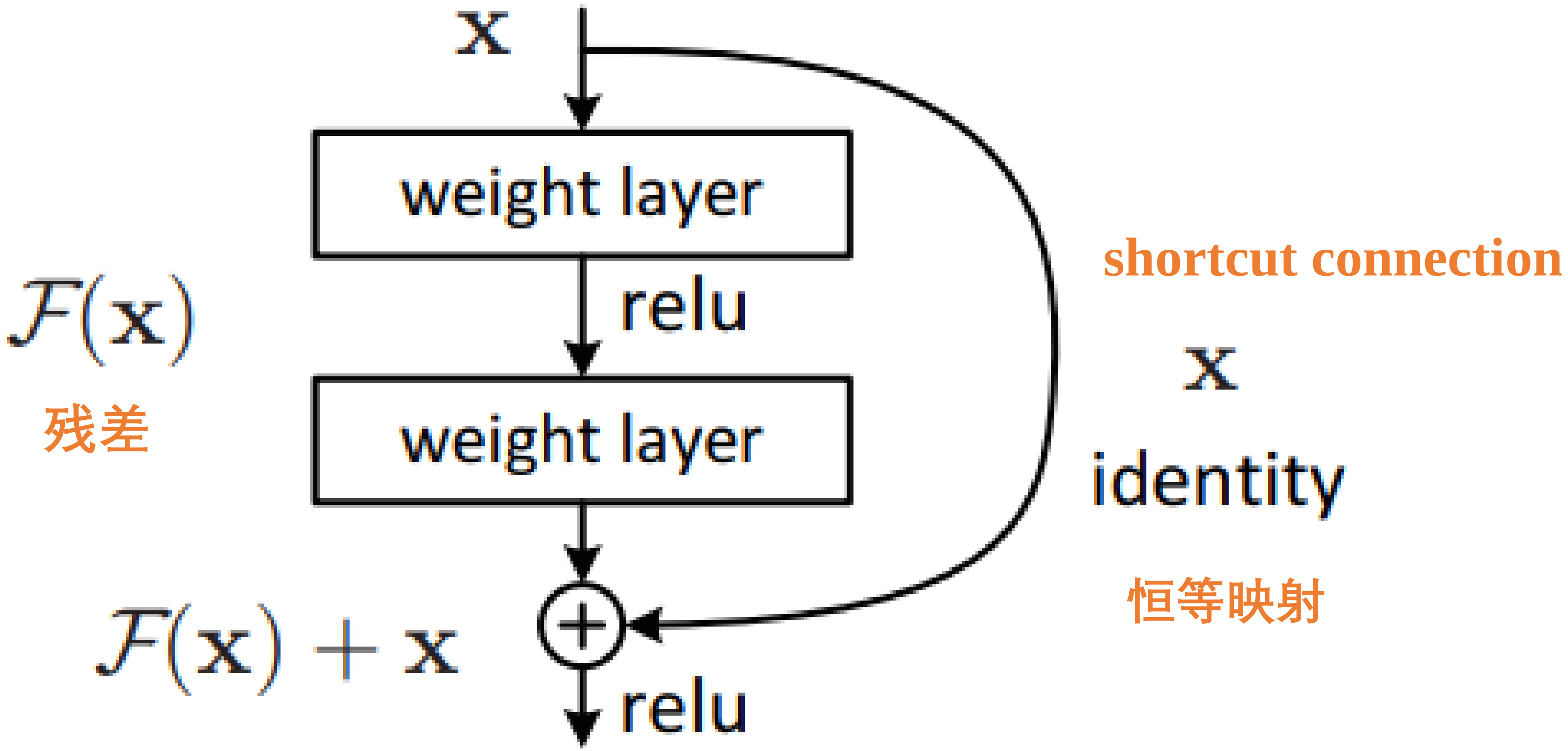


训练集误差

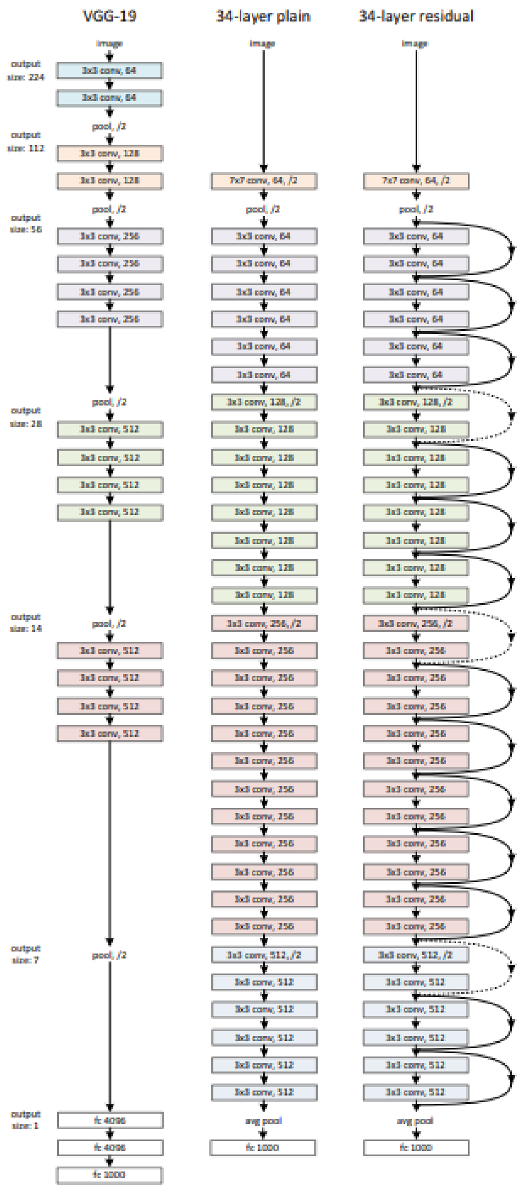


测试集误差

ResNet核心解决问题



ResNet核心解决问题



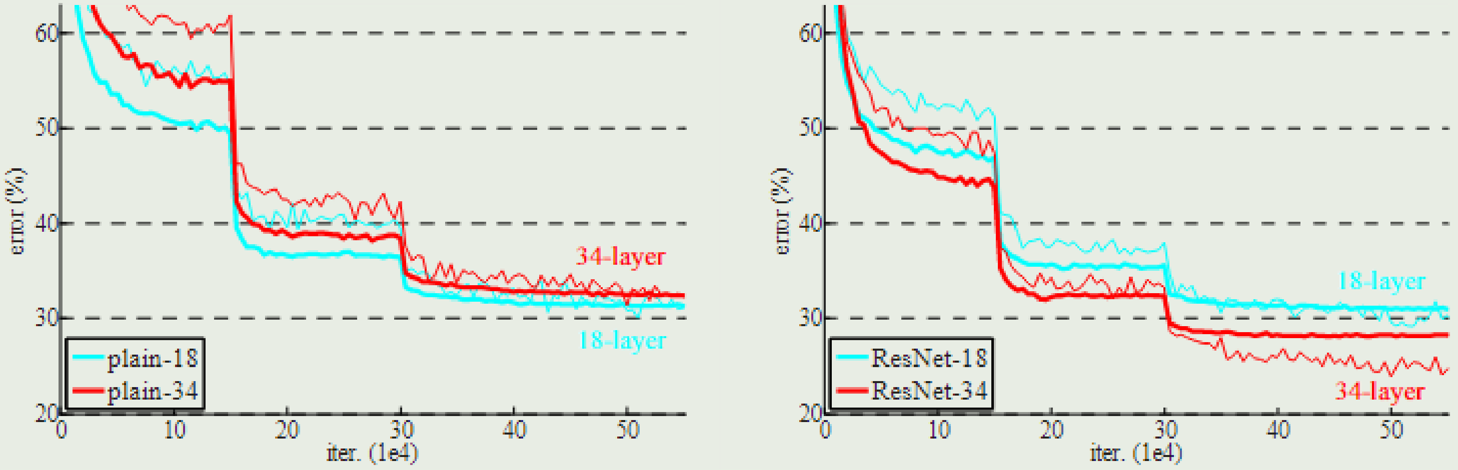


Figure 4. Training on **ImageNet**. Thin curves denote training error, and bold curves denote validation error of the center crops. Left: plain networks of 18 and 34 layers. Right: ResNets of 18 and 34 layers. In this plot, the residual networks have no extra parameter compared to their plain counterparts.

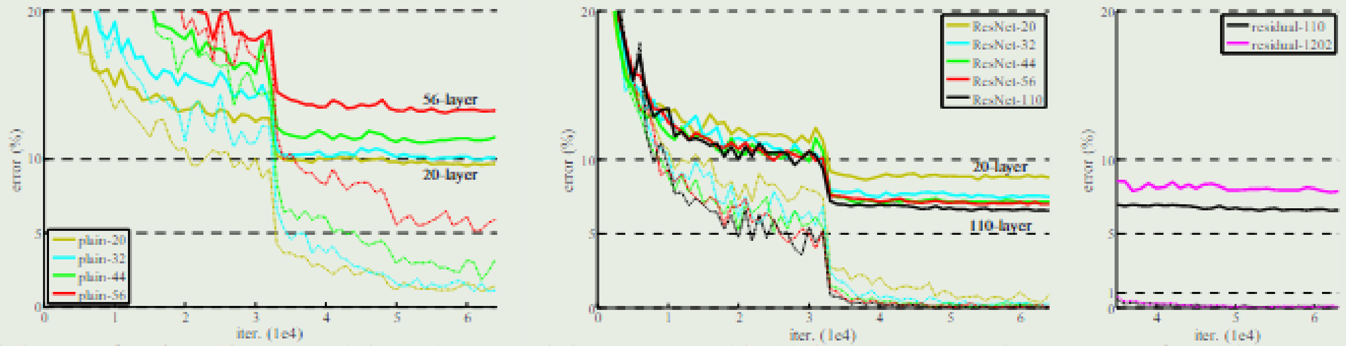
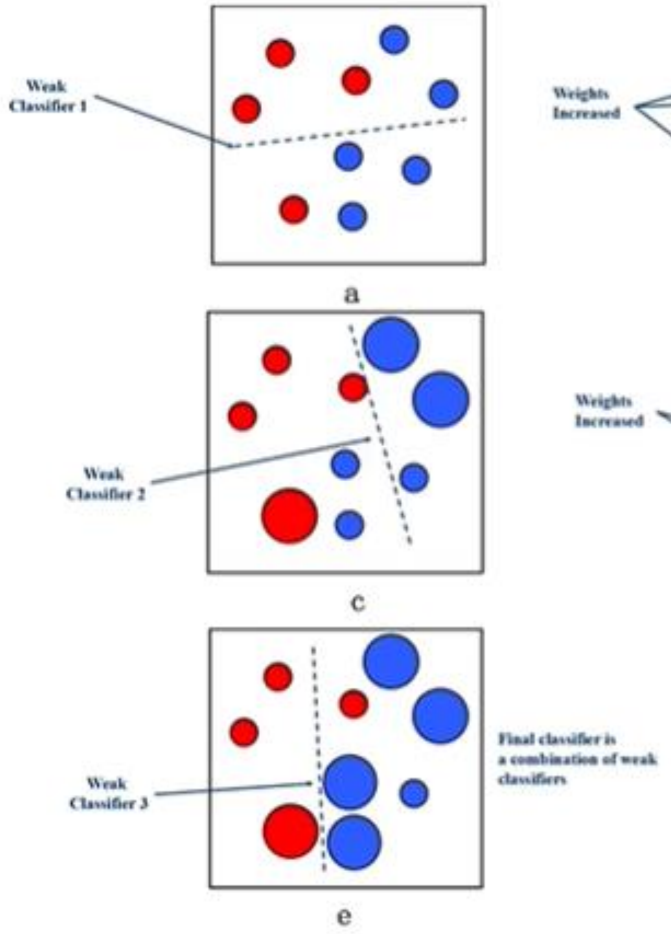
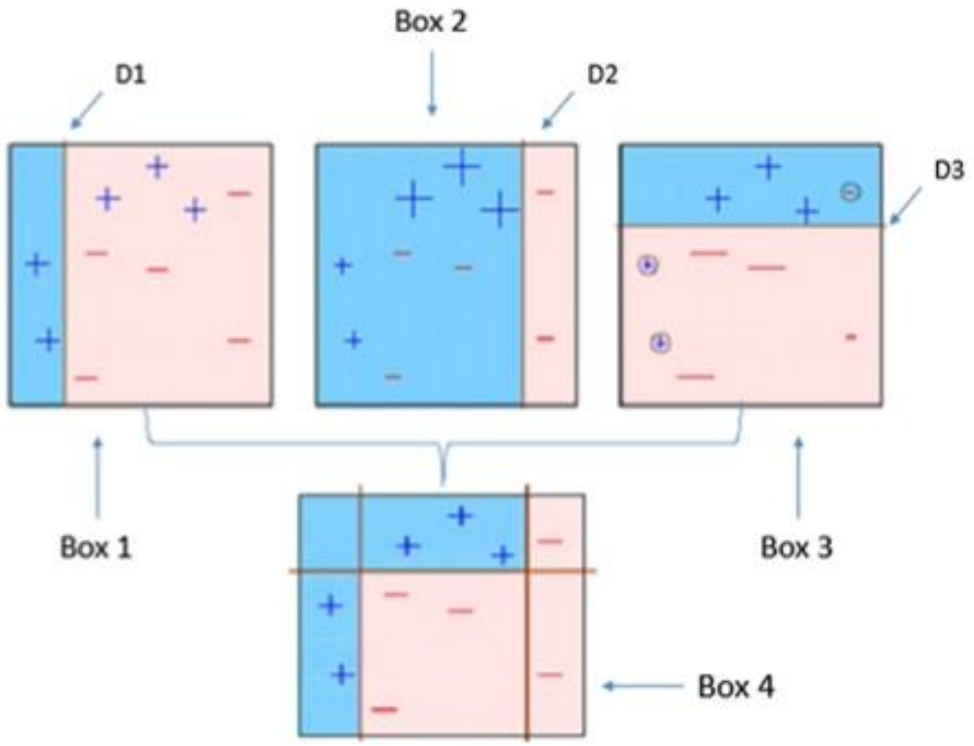


Figure 6. Training on **CIFAR-10**. Dashed lines denote training error, and bold lines denote testing error. **Left**: plain networks. The error of plain-110 is higher than 60% and not displayed. **Middle**: ResNets. **Right**: ResNets with 110 and 1202 layers.

ResNet如何解决网络退化的机理

类似某些集成学习方法、或长短时记忆神经网络LSTM的遗忘门





ResNet如何解决网络退化的机理

传统线性神经网络很难做到恒等映射

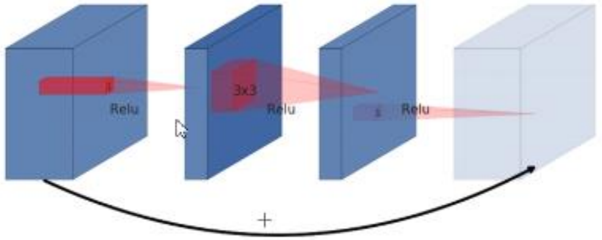
MobileNetV2: Inverted Residuals and Linear Bottlenecks

Mark Sandler Andrew Howard Menglong Zhu Andrey Zhmoginov Liang-Chieh Chen
Google Inc.
{sandler, howarda, menglong, azhmogin, lcchen}@google.com

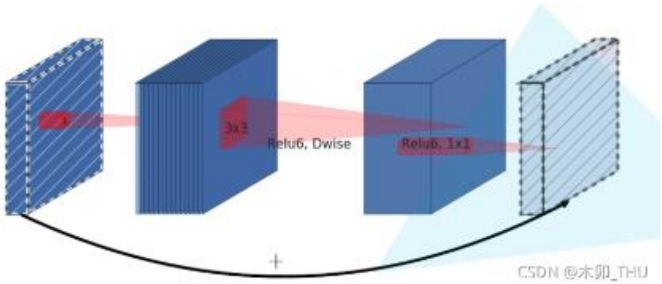
① 1x1 卷积降维
② 3x3 卷积
③ 1x1 卷积升维

① 1x1 卷积升维
② 3x3 卷积 DW
③ 1x1 卷积降维

(a) Residual block



(b) Inverted residual block



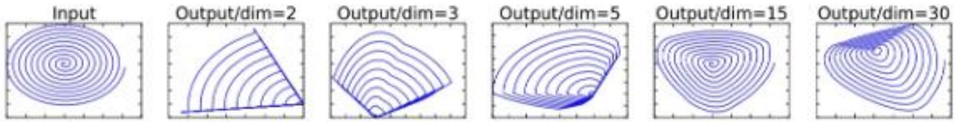


Figure 1: Examples of ReLU transformations of low-dimensional manifolds embedded in higher-dimensional spaces. In these examples the initial spiral is embedded into an n-dimensional space using random matrix T followed by ReLU, and then projected back to the 2D space using T^{-1}. In examples above $n = 2, 3$ result in information loss where certain points of the manifold collapse into each other, while for $n = 15$ to 30 the transformation is highly non-convex.

ResNet如何解决网络退化的机理

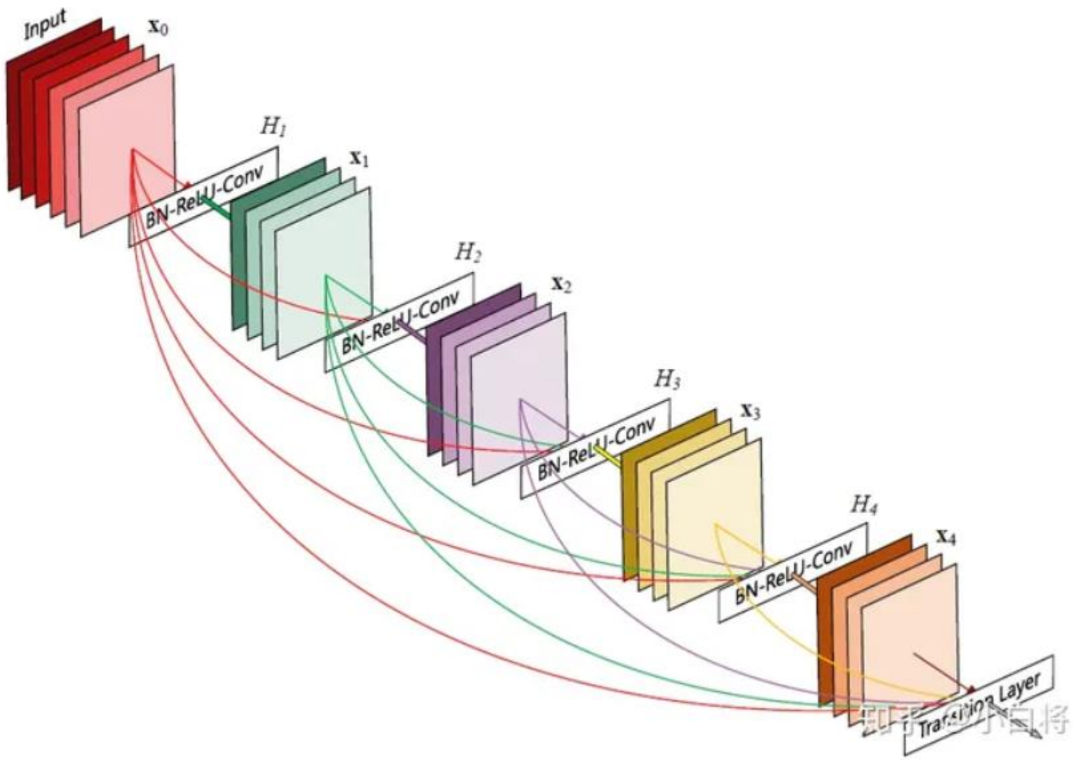
skip connection可以实现不同分辨率特征组合

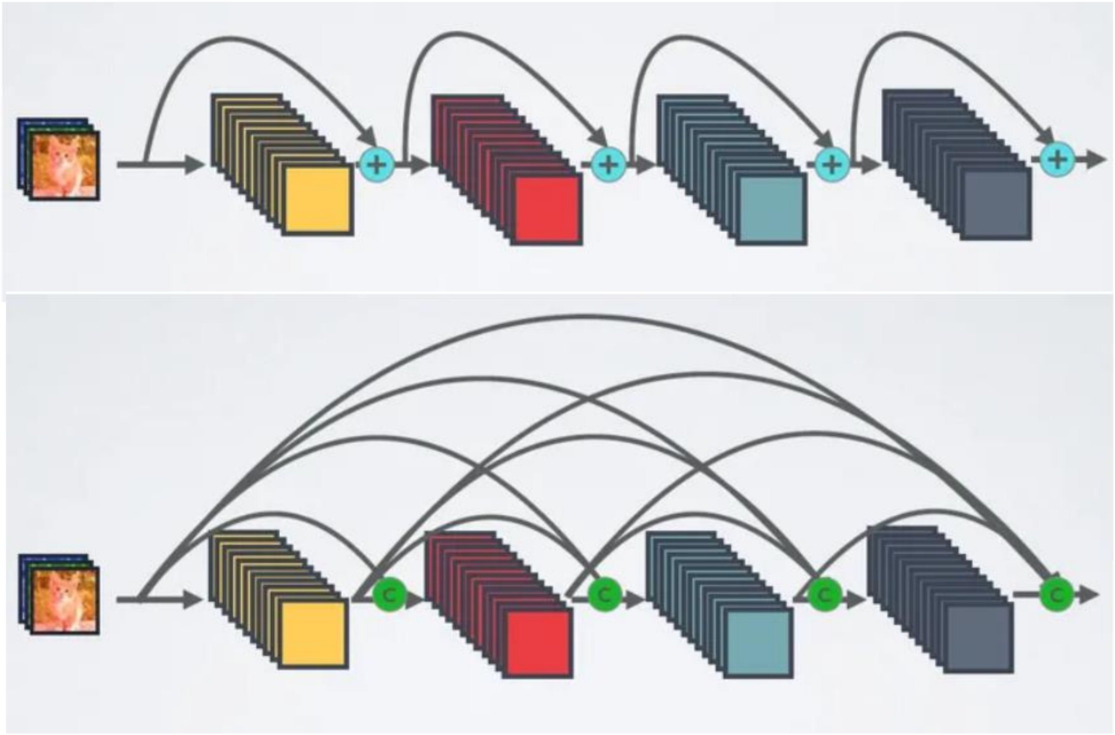
[Submitted on 25 Aug 2016 (v1), last revised 28 Jan 2018 (this version, v5)]

Densely Connected Convolutional Networks

Gao Huang, Zhuang Liu, Laurens van der Maaten, Kilian Q. Weinberger

Recent work has shown that convolutional networks can be substantially deeper, more accurate, and efficient to train if they contain shorter connections between layers close to the input and those close to the output. In this paper, we embrace this observation and introduce the Dense Convolutional Network (DenseNet), which connects each layer to every other layer in a



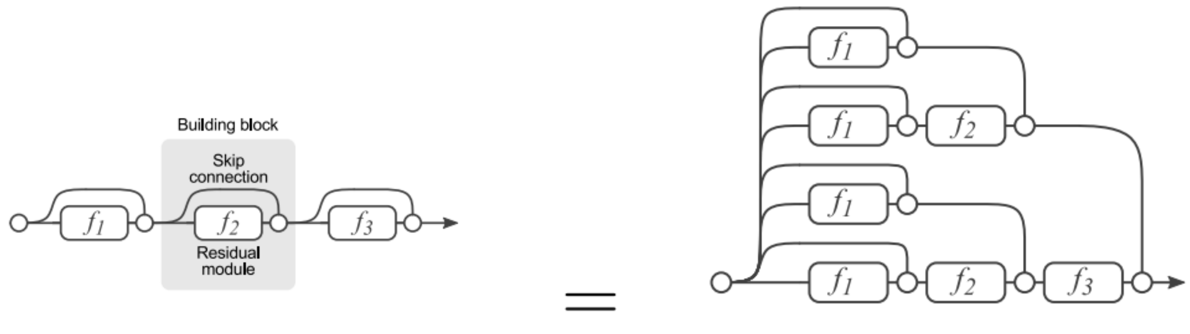


ResNet如何解决网络退化的机理

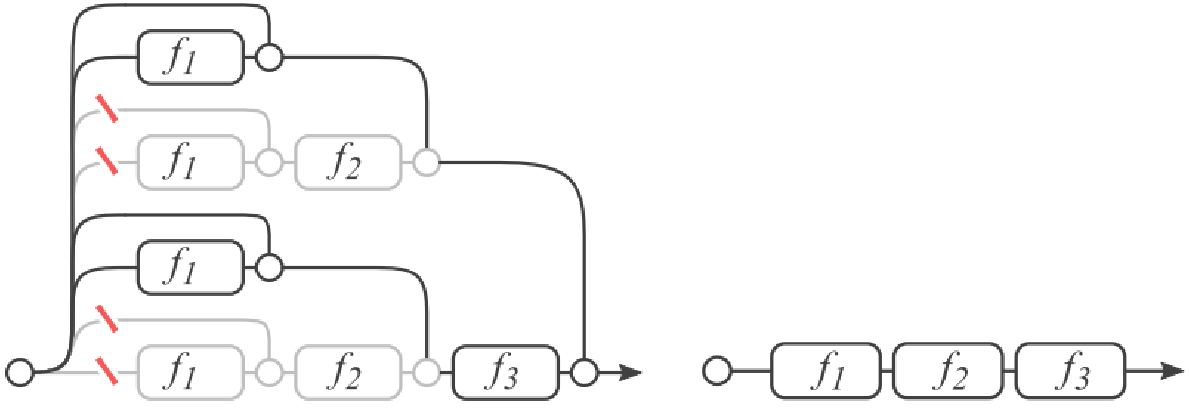
ResNet相当于几个浅层网络的集成

Residual Networks Behave Like Ensembles of Relatively Shallow Networks

Andreas Veit **Michael Wilber** **Serge Belongie**
Department of Computer Science & Cornell Tech
Cornell University
{av443, mjw285, sjb344}@cornell.edu



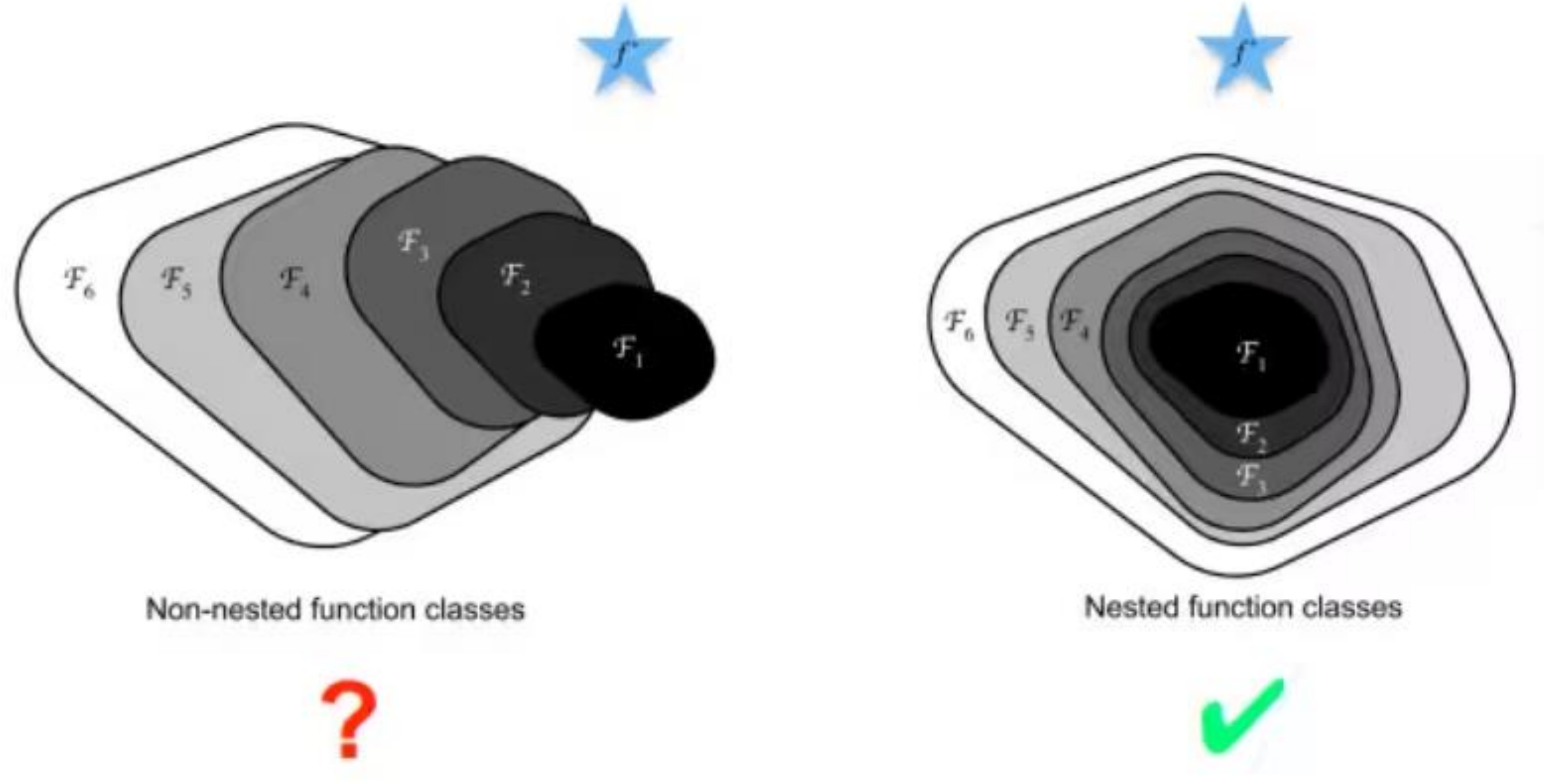
(a) Conventional 3-block residual network (b) Unraveled view of (a)



(a) Deleting f_2 from unraveled view (b) Ordinary feedforward network

ResNet如何解决网络退化的机理

数学角度



ResNet如何解决网络退化的机理

物理动力系统角度：从动力系统角度解释Resnet：Weinan E. A Proposal on Machine Learning via Dynamical Systems【J】. Communication in Mathematics and Statistics, 2017.

where z_l and z_{l+1} are the input and output of the l-th layer, y_l is an auxiliary variable for the l-th layer, h and g are some mappings which could in principle be nonlinear. A main result of [11], found through numerous numerical experiments, is that for very deep networks (hundreds or thousands of layers), training is the easiest if both g and h are the identity map [11]. This is quite expected from the viewpoint of dynamical systems. In fact, denote by G the inverse map of g, we can then write the above dynamical system as:

$$z_{l+1} = G(h(z_l) + \mathscr{F}(z_l, W_l)) \tag{2.3}$$

In order to have a stable behavior (nonvanishing or exploding), the gradient of the right hand side should be close to an identity map. Assuming that $\mathscr{F}$ is a small perturbation, we then need

$$\nabla G \nabla h \sim I \tag{2.4}$$

This is certainly fulfilled when both g and h are identity maps. In general, one has

$$z_{l+1} \sim G(h(z_l)) + \nabla G \cdot \mathscr{F}(z_l, W_l) \tag{2.5}$$

This means that the leading order behavior is dominated by $G(h)$. The added flexibility from choosing more general g and h does not offer much real improvement if one wants to use a large number of layers.

Let g and h both be identity maps. Then (2.3) becomes:

$$z_{l+1} = z_l + \mathscr{F}(z_l, W_l) \tag{2.6}$$

This can be viewed as a discretization of the dynamical system:

$$\frac{dz}{dt} = \mathscr{F}(z, W(t)) \tag{2.7}$$

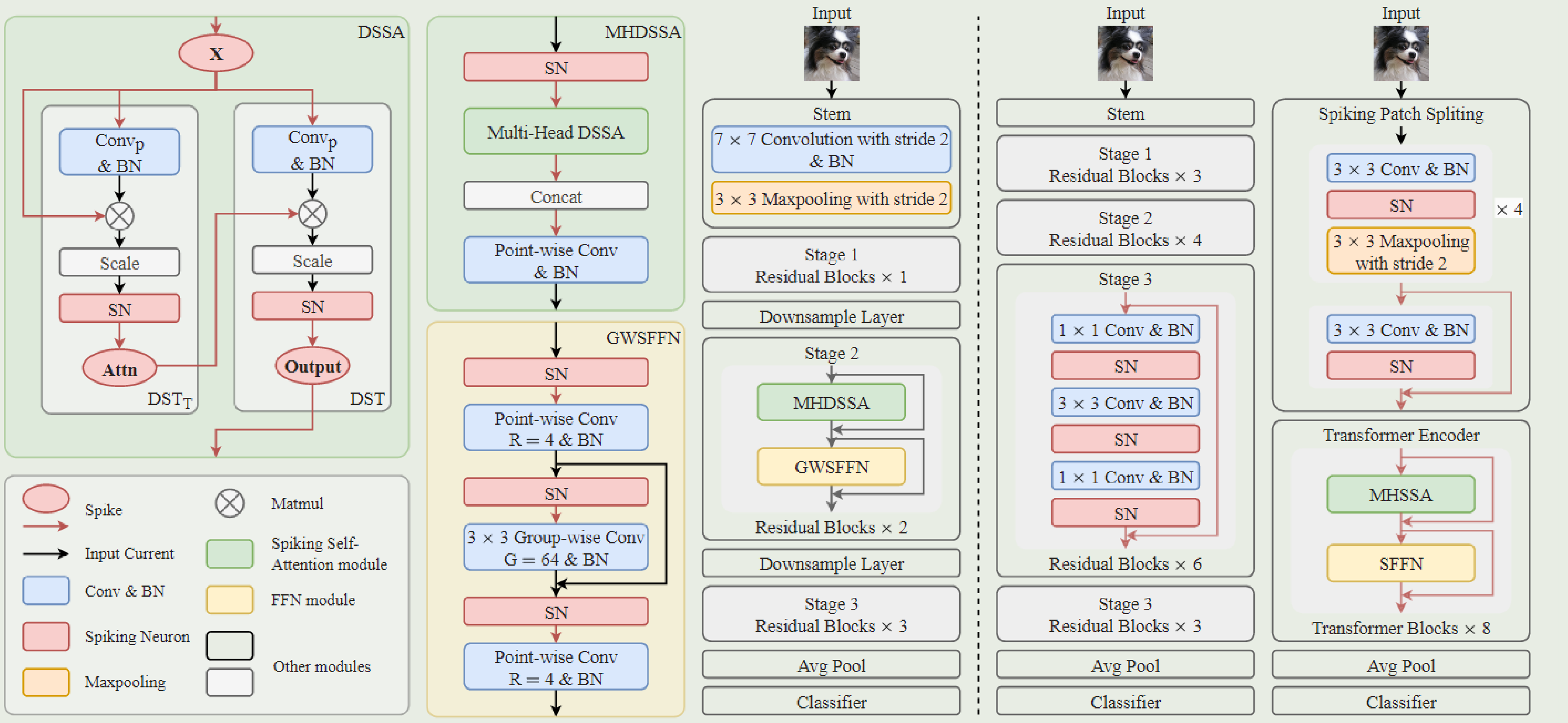
In fact the simplest discretization of (2.7) takes the form:

$$z_{l+1} = z_l + \Delta t_l \mathscr{F}(z_l, W_l) \tag{2.8}$$

实验结果得出

ResNet的应用

SpikingResformer: Bridging ResNet and Vision Transformer in Spiking Neural Networks



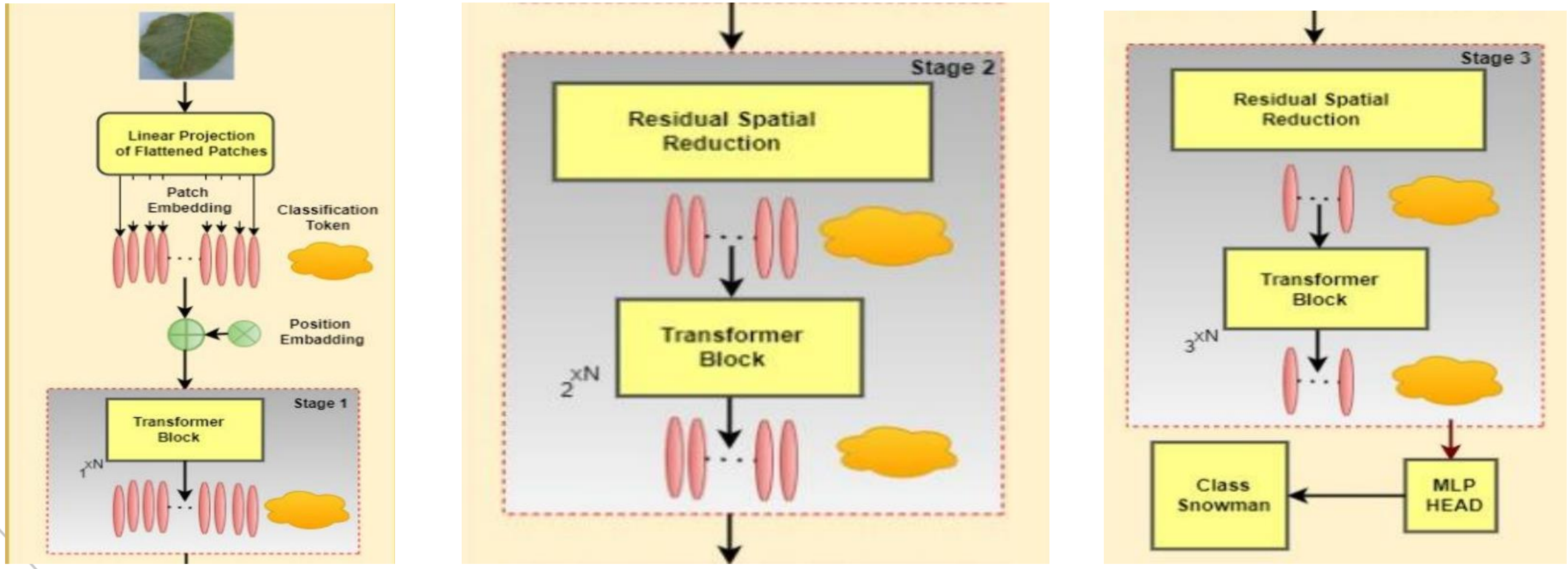
a) SpikingResformer b) SEW ResNet-50 c) Spikformer

ResNet的应用

Article

EfficientRMT-Net—An Efficient ResNet-50 and Vision Transformers Approach for Classifying Potato Plant Leaf Diseases

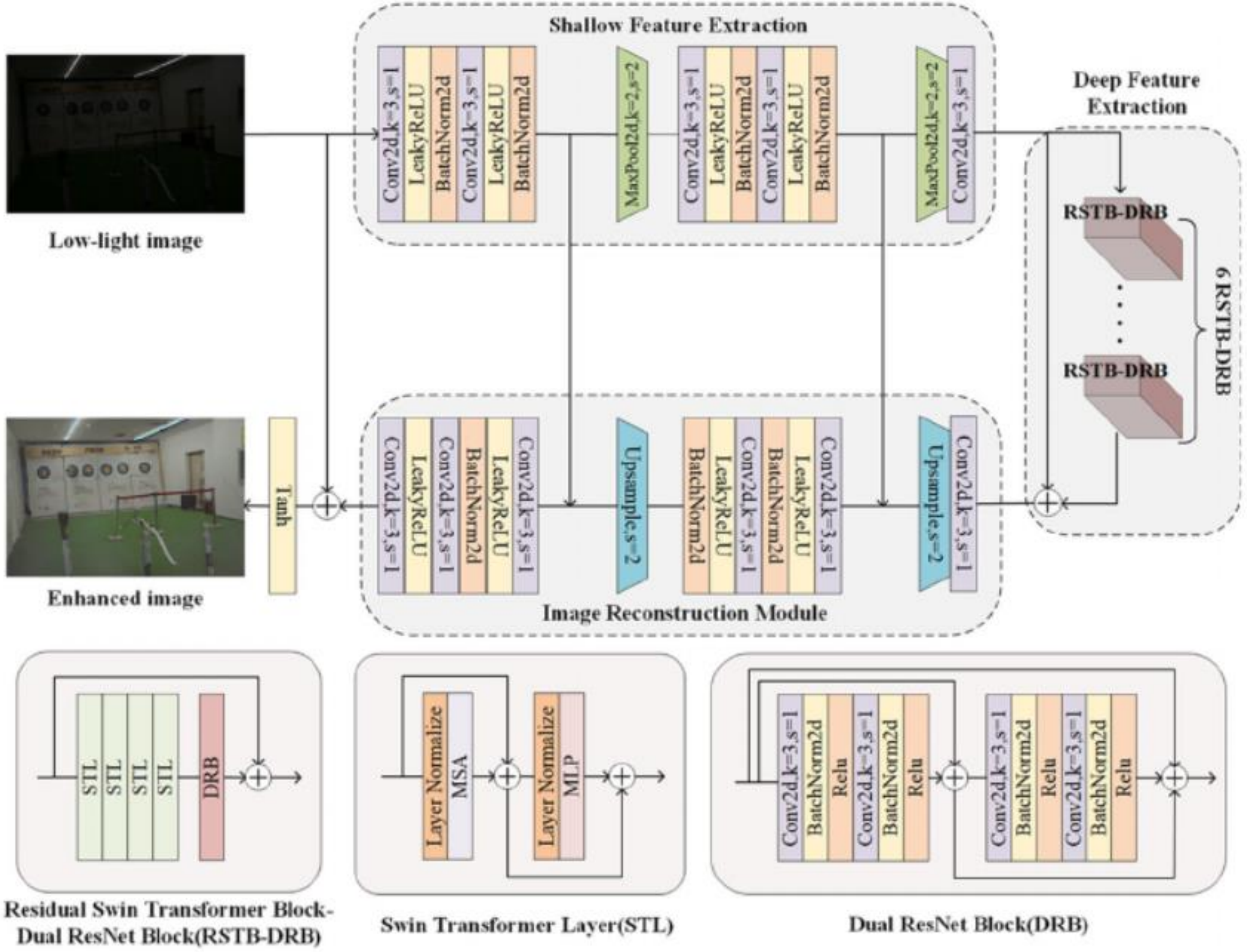
This paper addresses the issue of low classification accuracy in EfficientRMT-Net when applied to potato leaf images. To overcome this challenge, the authors propose a novel deep learning network called EfficientRMT-Net, which combines ResNet-50 with a Transformer architecture for improved speed and accuracy. The structure of EfficientRMT-



ResNet的应用

Swin transformer and ResNet based deep networks for low-light image enhancement

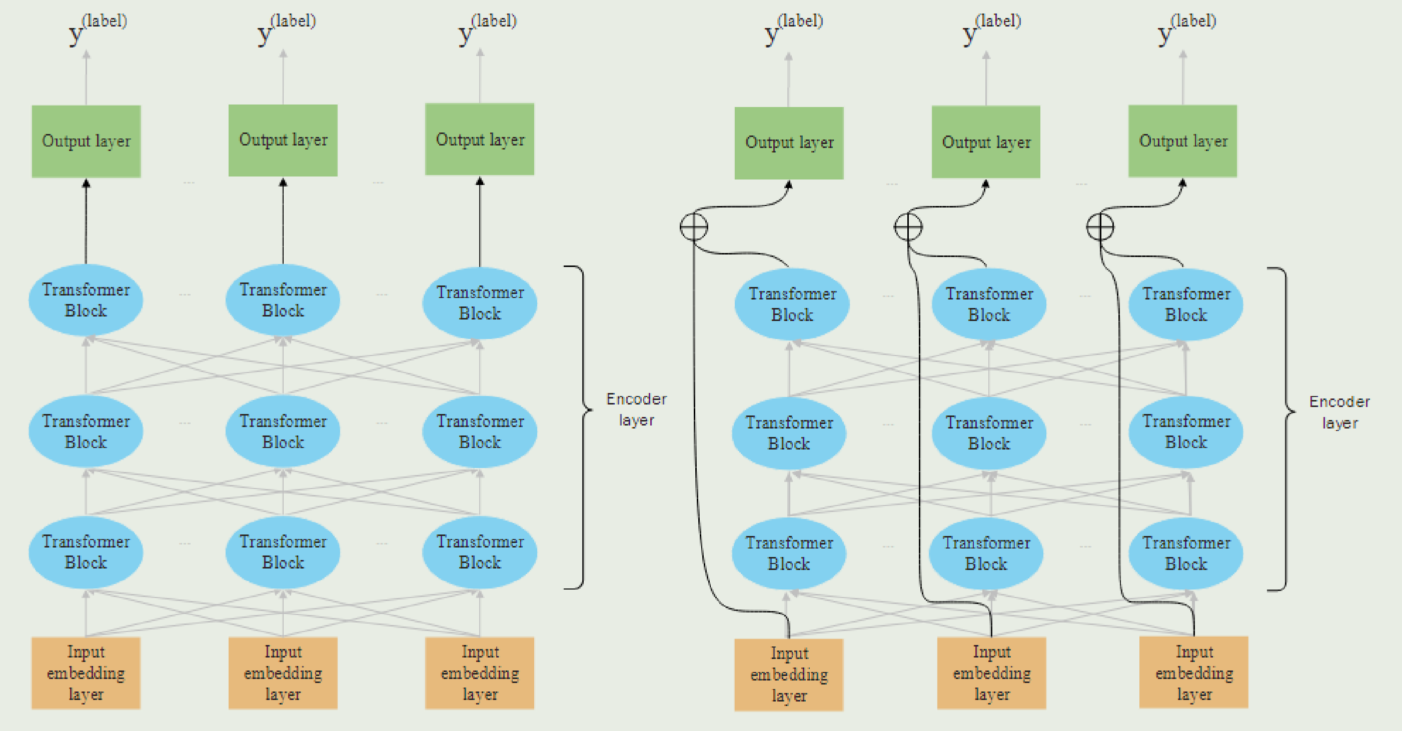
Lintao Xu[1] · Changhui Hu[1] · Bo Zhang[1] · Fei Wu[1] · Ziyun Cai[1]



ResNet的应用

Combining ResNet and Transformer for Chinese Grammatical Error Diagnosis

†‡Shaolei Wang, †Baoxin Wang, †Jiefu Gong, ‡Zhongyuan Wang, †Xiao Hu, †Xingyi Duan,



(a)BERT (b)ResBERT

3.1. ControlNet

ControlNet injects additional conditions into the blocks of a neural network (Figure 2). Herein, we use the term *network block* to refer to a set of neural layers that are commonly put together to form a single unit of a neural network, *e.g.*, resnet block, conv-bn-relu block, multi-head attention block, transformer block, *etc*. Suppose $\mathcal{F}(\cdot;\Theta)$ is such a trained neural block, with parameters Θ, that transforms an input feature map $\boldsymbol{x}$, into another feature map $\boldsymbol{y}$ as

$$\boldsymbol{y} = \mathcal{F}(\boldsymbol{x};\Theta). \tag{1}$$