

# UAV Aerial Image Recognition Based on Deep Residual Net

Wenbo Wang, Dalei Chen, Xiance Dai

PLA Armored Force Academy Bengbu Campus, Bengbu Anhui  
Email: 13083272717@163.com

Received: Oct. 11<sup>th</sup>, 2018; accepted: Oct. 22<sup>nd</sup>, 2018; published: Oct. 29<sup>th</sup>, 2018

---

## Abstract

In order to effectively improve the recognition accuracy of UAV aerial image, this paper puts forward a kind of convolutional neural network model based on deep residual net. This model based on deep residual net uses the randomized ReLU excitation function, which can make the network have more rapid convergence speed. For participation of some parameters in the spread of gradient descent process which is not high in deep nets, the model uses random dropout to reduce the amount of calculation, improving training speed of network. The image recognition model proposed in this paper is verified having higher identification accuracy comparing with several classical deep convolution neural network model by experiment.

## Keywords

Aerial Images, Deep Residual Net, Random Dropout

---

# 基于深度残差网络的无人机航拍图像识别

王文博, 陈大雷, 戴宪策

中国人民解放军陆军装甲兵学院蚌埠校区, 安徽 蚌埠  
Email: 13083272717@163.com

收稿日期: 2018年10月11日; 录用日期: 2018年10月22日; 发布日期: 2018年10月29日

---

## 摘要

为了有效提高无人机航拍图像的识别准确率, 本文提出了一种基于深度残差网络的深度卷积神经网络模型。该模型在深度残差网络的基础上使用了随机化ReLU激励函数, 能够使网络拥有更加快速的收敛速度, 同时针对深度残差网络层数较深, 且部分参数在梯度下降的传播过程中参与度不高的特点, 利用随机失

活有效降低整个网络训练量，进而提高网络的训练速度。经过实验验证，本文提出的图像识别算法相比几种经典深度卷积神经网络模型拥有更高的识别准确度。

## 关键词

航拍图像，深度残差网络，随机失活

Copyright © 2018 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

航拍图像是通过飞行器或卫星等设备对目标区域拍摄的图像，在军事目标侦察和民用方面具有重要作用。近年来随着无人机的快速发展，依托无人机航拍图像进行目标侦察或民用测绘作业相比卫星航拍图具有经济、高效、方便等多方面优势。而在进行航拍图像识别作业时，人工方式效率偏低，且需要进行一定的培训才能进行识别工作，成本较高。传统的算法识别通常采用特征提取加分类器选择的方式，例如 Gabor 特征、Gradient 特征结合贝叶斯分类器或支持向量机分类器的方式，虽然相对人工方式效率较高但识别率偏低，容易造成误判、错判。

近年来，随着机器学习算法的快速发展，利用深度学习技术构建图像分析识别模型，进而应用训练好的模型进行识别正越来越广泛[1] [2] [3] [4]。此种方式能够在识别正确率与人工识别率上不相上下，而且在识别速度与识别规模上优势很大。2006 年 Hinton 等人[5]提出由一组受限玻尔兹曼机组成的深度置信网络，通过逐层预训练的方式有效完成了模型训练过程，证明了神经网络模型在图像识别等领域的巨大优势。2012 年 Alex 等人[6]以五层卷积层和三层全连接层构建了 Alex-Net 卷积神经网络，利用 GPU 并行训练在 ImageNet 竞赛中取得高于第二名 10.9 个百分点的识别结果。2014 年 Simonyan 等人[7]利用小卷积核提出了 VGG-Nets，相比 Alex-net，VGG-Nets 网络在训练时使用了更多的网络层次，且 VGG-Nets 中的卷积层都使用了  $3 \times 3$  的滤波器。Szegedy 等人[8]提出的 GoogLeNet 同样采用了较深的网络层次，而且使用了  $3 \times 3$  的卷积层进行升维和降维。

但是随着新的卷积神经网络模型中网络层次的逐渐加深，有时会出现训练误差不降反升的现象。为了有效解决此类问题，He 等人[9]提出了深度残差网络，通过残差网络中的残差块结构有效预防了网络中训练损失值不再出现随着训练逐渐升高的现象。本文基于深度残差网络提出一种改进的深度卷积神经网络模型，利用随机化 ReLU 函数帮助收敛，利用随机失活降低训练过程中的参数规模，降低网络中参数规模，提高整个神经网络模型的泛化性，能够有效实现无人机航拍图像的精确识别。

## 2. 深度卷积神经网络的基本原理

深度卷积神经网络通常由卷积层、池化层和全连接层等多个网络组成。在训练时，算法将误差由最后一层逐层向前反馈，更新每层参数，并在更新参数后再次前馈，如此往复，直到深度卷积神经网络模型收敛，从而达到模型训练的目的。通过训练得到的模型深度卷积神经网络进而可以进行样本预测，也就是进行图像识别，目标检测等工作。图 1 为卷积神经网络 LeNet-5 [10]的网络结构，该网络由多个卷积层、池化层(下采样层)以及数个全连接层构成。

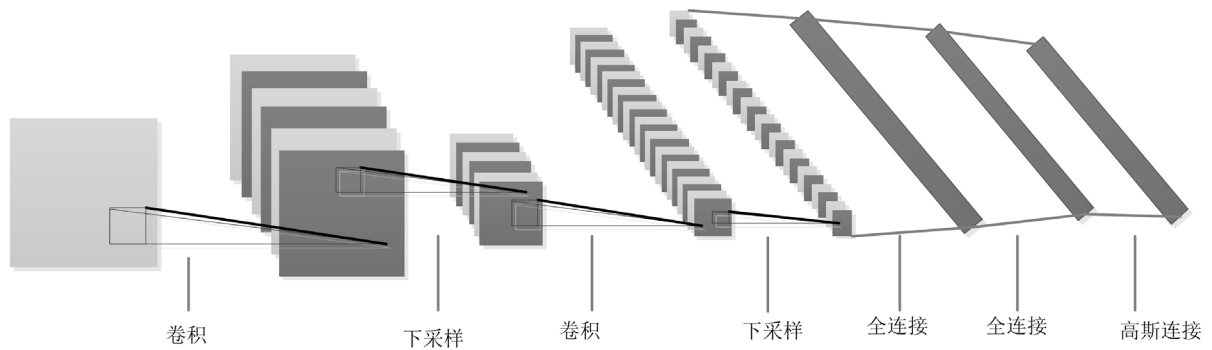


Figure 1. Structure of deep convolution neural network LeNet-5

图 1. 深度卷积神经网络 LeNet-5 的网络结构

### 2.1. 卷积层

卷积层主要用于对输入的数据进行卷积运算操作，从而提取出输入数据的特定特征[11]。卷积运算通过卷积核与输入数据逐位相乘后再整体相加实现的，其实质是实现输入数据从低维度到高维度的特征提取。

设输入数据为  $x \in R^{M \times N}$ ，卷积核为  $f \in R^{M \times N}$ ，则形式化的卷积操作可表示为公式(1):

$$y_{uv} = \sum_{i=1}^M \sum_{j=1}^N f_{ij} \cdot x_{u-i+1, v-j+1} \quad (1)$$

通常在卷积操作时加入偏置项  $b_k$ ，可得公式为:

$$y_{uv} = \sum_{i=1}^M \sum_{j=1}^N f_{ij} \cdot x_{u-i+1, v-j+1} + b_k \quad (2)$$

卷积层中的卷积核可以看做为不同的滤波器，其对输入数据的特征提取就是对不符合滤波器特性的目标信息进行处理，此过程通过多轮卷积方能实现，在前几层通常抽取纹理、颜色、局部形状、线条等低层次特征，进而逐步过渡到目标轮廓等高层次特征。

### 2.2. 池化层

池化操作通常在卷积操作之后，也即池化层的输入通常为卷积层数据。池化操作也叫下采样操作，通常包括平均值池化和最大值池化两种方式。池化通常也需要进行多轮操作，其目的是通过特征降维来获得输入数据的不变性特征。池化层的降维操作减少了参数个数和计算量，同时也可以在一定程度上预防过拟合，增加模型的泛化度。

设池化核为  $p \in R^{M \times N}$ ，则平均池化的公式可表示为:

$$y_{ij} = \frac{1}{MN} \sum_{1 \leq i \leq M, 1 \leq j \leq N} x_{i \times M, j \times N} \quad (3)$$

最大值池化的公式可表示为:

$$y_{ij} = \max_{1 \leq i \leq M, 1 \leq j \leq N} x_{i \times M, j \times N} \quad (4)$$

从公式(3)和公式(4)中可以看出，池化操作明显对输入的卷积层数据实现了降维，减少了神经网络的整体训练量。

### 2.3. 全连接层

在进行多层卷积和池化操作之后，神经网络通常连接有两到三层的全连接层。全连接层中的神经元

与上一层网络是全连接关系,通过全连接层可实现对网络特征映射到样本空间的预测,也就是分类作用。

### 3. 改进的深度残差网络

2015 年 He 等人提出了主要由  $3 \times 3$  的卷积核和  $2 \times 2$  的最大池化层反复堆叠构建而成的深度残差 (Residual Neural Network, ResNet) 网络。相较于其他网络结构,深度残差网络模型更为简单清楚,其核心贡献是通过残差块结构有效解决了卷积神经网络在网络层数加深时因为梯度弥散导致的退化问题,以此可以构建网络层数更深的神经网络,提高准确度,并且在网络层数加深时还能够保持良好的泛化性。

#### 3.1. 深度残差网络的残差块结构

神经网络在深度不断加深时,训练集的损失函数值会先降低趋于平缓,如果网络深度继续增加损失函数值反而会逐渐上升,这一现象的产生可能是是由于反向传播过程中梯度弥散导致了整个神经网络的退化。深度残差网络正是为了解决这一问题而产生,通过引入残差网络结构,可以在输入层与输出层之间建立一个捷径连接,在捷径连接上通过恒等映射保证输出层误差不高于输入层,以此实现网络的进一步加深,取得更好的识别效果。深度残差网络通常由多个残差块结构重叠形成,残差块结构如图 2 所示。

从图 2 中可以看出,在残差块结构中,通过把输入  $x$  传到输出作为结果,可得公式:

$$y = F(x) + x \tag{5}$$

则深度残差网络的学习函数变为了残差项  $y - x$  :

$$F(x) = y - x \tag{6}$$

残差块结构中还引入了恒等映射 ( $y = x$ ), 以保证在网络损失函数值下降趋势平缓时,该捷径连接能够保证损失值不再提高,进而能够继续在网络中实现增加网络层数的目的。最后再经过非线性变换 ReLU 激活函数处理,形成残差块结构。

#### 3.2. 随机化 ReLU 激励函数

卷积神经网络的若干线性操作层的重叠只能起到线性映射作用,要想实现非线性映射,增强整个网络的表达能力,往往需要在神经网络不同层之间增加激励函数。经典的神经网络激励函数包括 Sigmoid 函数、tanh 函数和 ReLU 函数等。

Sigmoid 函数是常用的非线性激励函数,也称为 Logistic 函数,其数学形式为:

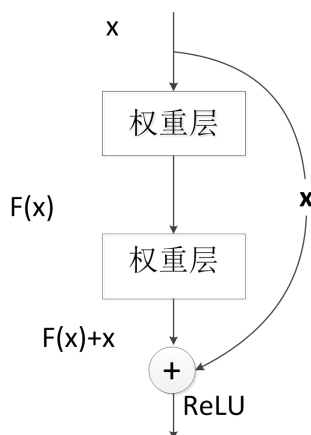


Figure 2. Structure of residual block  
图 2. 残差块结构

$$f(x) = \frac{1}{1 + e^{-x}} \quad (7)$$

从公式中可以看出, Sigmoid 函数能够将输入值转化为 0 到 1 之间的输出值。在输入值为很大的负值和很大的正值时, Sigmoid 函数会趋近于 0 和 1, 很适合二分类问题的使用。但是 Sigmoid 函数在卷积神经网络的梯度反向传播时可能导致梯度饱和和梯度消失问题。

ReLU 函数全称叫修正线性单元, 是目前最常用的激活函数, 其数学形式为:

$$\text{ReLU}(x) = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \quad (8)$$

从公式中可以看出, ReLU 函数是一种分段函数, 对于  $x \geq 0$  时, 输出值不变; 对于  $x < 0$  时, 输出值为 0, 形成单侧抑制状态。通过实验证明, ReLU 函数加快梯度下降速度效果明显。但是对于  $x < 0$  时, ReLU 函数输出值为 0, 会造成小于 0 的这部分结果无法影响到后续的网络训练。为了缓解这种现象, 本文中采用了随机化 ReLU 激励函数, 其数学形式为:

$$\text{RReLU}(x) = \begin{cases} x & \text{if } x \geq 0 \\ \alpha'x & \text{if } x < 0 \end{cases} \quad (9)$$

随机化 ReLU 激励函数中  $\alpha$  的取值在训练阶段服从均匀分布。随机化 ReLU 激励函数相比 ReLU 激励函数能够使训练网络的识别错误率更低。

### 3.3. 随机失活(Dropout)策略

随机失活是一种能够以一定概率使卷积神经网络中某层的神经元权重置为 0 的网络正则化方法[12]。神经元权重置为 0 也就意味着其处于失活状态, 不再参与后续的网络训练。随机失活的意义在于能够有效减少训练网络与训练集数据的匹配程度, 减少过拟合现象。随机失活意味着每次训练时训练网络失活的神经元都不一样, 这也就导致每次训练时训练网络都有所不同, 有效提高了训练过程的模型数量, 同时也能显著提升网络泛化能力。同时, 由于失活的神经元无法参与后续的网络训练, 随机失活对于整个网络也就意味着进行了剪枝处理, 移除了一些神经元, 简化了神经网络的参数规模, 能够有效提高网络的训练速度。

## 4. 仿真实验

为了验证本文所提出的基于深度残差网络的图像识别算法的识别准确率, 本文利用 TensorFlow 深度学习平台, 在同一计算机上进行了算法验证。计算机 CPU 型号为 Core-i7 双核, 内存为 8 GB, 显卡型号为 NVIDIA GeForce GTX 960M, 显存为 2 GB。

在 Tensorflow 平台上分别用 CNN、Alex-net、VGG-16 和本文算法对 MNIST 数据集进行识别, 在不同训练轮数下的损失函数值如图 3 所示。从表中可以看出, 本文算法能够有效实现 MNIST 手写数据集的手写数字识别, 而且随着训练轮数增加, 损失值在不断减小, 算法识别结果越来越精确。Alex-net 作为较早的卷积神经网络, 其收敛速度和收敛精度相比其他几种卷积神经网络有所降低。而本文算法相比 GoogleNet、VGG-16 等算法, 在前期的 400 个训练轮次, VGG-16 网络能够实现更好的训练效果, 但是随着训练轮数的增加, 在 400~2000 个训练轮次时, 本文算法有更明显的收敛优势。到达 2000 训练轮次时, 由于梯度达到局部最优, 三个算法达到相同的训练精度。在 2000~2800 训练轮次, 本文算法同样显示出更快的收敛速度, 且在 2800 个训练轮数时, 能够收敛到更高的训练精度。

从表 1 所示的最后精度对比, 相比几种经典的卷积神经网络, 在相同的实验条件下, 在对同一训练

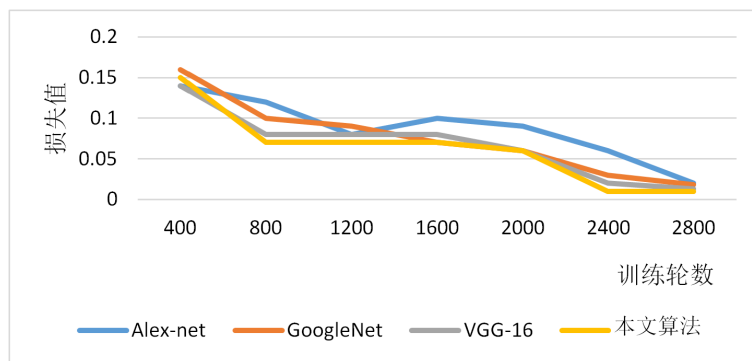


Figure 3. Loss value under different number in different neural network

图 3. 不同神经网络在不同轮数下的损失值

Table 1. Identification accuracy among several neural networks

表 1. 几种神经网络识别精度对比

训练模型	识别精度
Alex-net	98.0%
GoogleNet	98.2%
VGG-16	98.7%
本文算法	99.0%

集进行同样的训练轮数之后，本文的算法的识别精度更高，能够实现对 MNIST 数据集手写数字图片进行更精确的识别。

## 5. 总结

为了有效提高图像识别准确率，本文提出了一种改进的深度卷积神经网络模型。该模型基于深度残差网络建立，能够有效缓解深度卷积神经网络在网络层数较多时训练出现的损失函数值不降反升的现象，使用随机化 ReLU 激励函数增加了网络的非线性表示，利用随机失活策略在网络训练时有效减少了参数量和运算数据量，提高了整个网络的训练速度。经过对 MNIST 数据集进行识别证明，相比几种经典深度卷积神经网络模型，该模型能够在相同的训练轮数上更快地收敛，而且经过充分的训练轮次训练之后，网络模型的识别准确率更高，在无人机航拍图像识别上应用能够在更短的训练时间得到更好的训练效果，有效地提高无人机航拍图像识别的效率，降低识别成本。

## 参考文献

- [1] 袁培森, 黎薇, 任守纲, 等. 基于卷积神经网络的菊花花型和品种识别[J]. 农业工程学报, 2018, 34(5): 152-158.
- [2] 傅隆生, 冯亚利, Elkamil, 等. 基于卷积神经网络的田间多簇猕猴桃图像识别方法[J]. 农业工程学报, 2018, 34(2): 205-211.
- [3] 余鹏, 甘健侯, 文斌, 等. 经典深度卷积神经网络模型在手绘草图识别中的应用研究[J]. 云南师范大学学报(自然科学版), 2018(1): 29-34.
- [4] 顾婷婷, 赵海涛, 孙韶媛. 基于金字塔型残差神经网络的红外图像深度估计[J]. 红外技术, 2018(5).
- [5] Hinton, G.E. and Salakhutdinov, R.R. (2006) Reducing the Dimensionality of Data with Neural Networks. *Science*, 313, 504-507. <https://doi.org/10.1126/science.1127647>
- [6] Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2012) ImageNet Classification with Deep Convolutional Neural Networks. *International Conference on Neural Information Processing Systems*, Curran Associates Inc., 1097-1105.

- 
- [7] Simonyan, K. and Zisserman, A. (2014) Very Deep Convolutional Networks for Large-Scale Image Recognition. *Computer Science*.
- [8] Szegedy, C., Liu, W., Jia, Y., *et al.* (2015) Going Deeper with Convolutions. *IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 1-9.
- [9] He, K., Zhang, X., Ren, S., *et al.* (2015) Deep Residual Learning for Image Recognition. 770-778.
- [10] Lecun, Y., Bottou, L., Bengio, Y., *et al.* (1998) Gradient-Based Learning Applied to Document Recognition. *Proceedings of the IEEE*, **86**, 2278-2324. <https://doi.org/10.1109/5.726791>
- [11] 周飞燕, 金林鹏, 董军. 卷积神经网络研究综述[J]. 计算机学报, 2017, 40(6): 1229-1251.
- [12] 王忠民, 王希, 宋辉. 基于随机 Dropout 深度信念网络的移动用户行为识别方法[J]. 计算机应用研究, 2017, 34(12): 3797-3800.

#### 知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>  
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2161-8801, 即可查询
2. 打开知网首页 <http://cnki.net/>  
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>  
期刊邮箱: [csa@hanspub.org](mailto:csa@hanspub.org)