

# 基于神经网络结构搜索的目标识别方法

卞伟伟, 邱旭阳, 申 研

(北京机械设备研究所, 北京, 100854)

**摘要** 针对目标识别需求,对基于神经网络的深度学习方法展开研究。由于深度学习模型中包含了对数据的先验假设,因此人工设计神经网络需要领域内专家丰富的先验知识,且具有劳动密集与时间成本高的缺点。为了获得超越专家个人经验、表现更好的网络,采用一种可微神经网络搜索的高效结构搜索方法,将搜索空间放宽为连续的空间,然后通过梯度下降来优化体系结构的验证集性能,从而找到面向目标识别的最优神经网络结构。仿真实验结果表明,将基于神经网络结构搜索的目标识别方法应用于“低慢小”类目标识别是可行的。

**关键词** 目标识别;卷积神经网络;神经网络结构搜索;深度学习

**DOI** 10.3969/j.issn.1009-3516.2020.04.014

**中图分类号** TN911.73 **文献标志码** A **文章编号** 1009-3516(2020)04-0088-05

## A Target Recognition Method Based on Neural Network Structure

BIAN Weiwei, QIU Xuyang, SHEN Yan

(Beijing Institute of Mechanical Equipment, Beijing 100854, China)

**Abstract** In view of the requirements of target recognition, the deep learning methods based on neural network are set off. Generally, there is a priori hypothesis of data contained in the in-depth learning model, the artificial design aimed at neural network for data needs an abundance of priori knowledge for experts in the field, and has the disadvantages of labor-intensive and high time cost. In order to obtain better network performance beyond the personal experience of network design experts, an efficient structure search method, i. e. Differentiable Architecture Search, is adopted. In this method, the search space is broadened to be continuous, and then the performance of the verification set of the architecture is optimized by gradient descent, finding the optimal neural network structure for target recognition. The simulation results show that it is feasible to apply the target recognition method based on neural network structure search to the LSS target recognition.

**Key words** target recognition; convolution neural network; neural network structure search; deep learning

城市环境中“低慢小”无人机目标体积小,颜色与复杂的背景环境相接近,使得对其进行探测与识别已成为一个世界性难题<sup>[1]</sup>,将计算机视觉与图像

处理技术相结合成为目标检测与识别的必然选择。神经网络深度学习<sup>[2]</sup>的成功很大程度上归功于特征工程的自动化,即利用分层特征提取器以端到

**收稿日期:** 2020-04-20

**基金项目:** 国防基础科研计划(JCKY2016204A601)

**作者简介:** 卞伟伟(1984—),男,江苏灌云人,高级工程师,博士,主要从事“低慢小”目标防控技术研究。E-mail:karovie@163.com

**引用格式:** 卞伟伟, 邱旭阳, 申研. 基于神经网络结构搜索的目标识别方法[J]. 空军工程大学学报(自然科学版), 2020, 21(4): 88-92. BIAN Weiwei, QIU Xuyang, SHEN Yan. A Target Recognition Method Based on Neural Network Structure[J]. Journal of Air Force Engineering University (Natural Science Edition), 2020, 21(4): 88-92.

端的方式从数据而不是手动设计中学习,故深度学习的模型效果依赖于数据,在以某一特定数据集进行训练时,模型提取该类数据的特征,那么将该模型运用在该类数据集上将会获得优异的表现。在影像分类中运用深度学习模型,除了借助已经存在的模型,也可以重新设计并训练新的模型。随着数据的复杂度不断增加,在设计相应的神经网络时需要更多的知识,不同类型的数据需要不同的先验知识,如果训练的是大规模的网络,将有百万个参数需要设计与学习,从而带来巨大的挑战;如果训练小规模的网络,获得的神经网络只是面向当前影像数据集,难以在其他种类的影像数据集上获得优异的表现。除此以外,也不能保证得到的模型优于自然图像集训练的模型以及根据自然图像集模型微调的模型。神经网络的设计是一个长耗时过程,如何在较短的时间内自动化设计一个高效的面向影像的神经网络成为了提升影像分类精度的关键,关系着目标识别概率的高低。

本文通过网络收集与实际拍摄,以多旋翼无人机、固定翼无人机、城市飞鸟为目标,采用一种可微神经结构搜索的高效结构搜索方法<sup>[3]</sup>,对目标识别问题进行研究。

## 1 问题描述

在目标识别分类任务中,需要使用大量带有标签的数据来训练模型,但是在实际中,训练影像的人工标记需要花费大量的人力和物力,同时需要大量的先验知识,导致影像数据集规模较小,可用于模型训练的数据较小,而为了让影像分类取得优异的效果,设计的深度学习模型包含的神经网络层数多,可能产生大量的参数,在训练时会出现过拟合现象,造成分类精度过低。模型迁移是将一个样本数据集上学习到的模型通过特定的方式运用到另一个具有相同或者相似学习任务的样本数据集上<sup>[4]</sup>,这样可以提高模型的实用性,增加深度学习模型的泛学习能力。但是由于当前获取影像的方式众多,不同的仪器采集的影像具有不同的特点,故影像数据的成像几何形状和内容不同,同时随着空间分辨率、光谱分辨率的增加,影像包含目标以及通道信息越来越丰富,因此当直接使用已存在的模型对其他种类的影像数据集进行训练时,实际取得的分类果不佳,为了使该模型可以在其他数据集上取得优异的结果,便需要对当前模型添加特定的任务知识,使其充分提取当前影像的特征,提高分类精度。

可微神经结构搜索方法将搜索空间放宽为连续的候选体系结构,以便通过梯度下降来优化体系结构的验证集性能。与低效的黑盒搜索相比,基于梯度的优化数据效率允许可微神经结构搜索使用数量级更少的计算资源来实现与现有技术的竞争性能。同时可微神经结构搜索方法相比许多现有方法更简单,因为它不涉及任何控制器或性能预测器,并且可通用搜索卷积和循环体系结构。另外,可微神经结构搜索方法能够在丰富的搜索空间内发现具有复杂图形拓扑的高性能结构,而限于任何特定的结构系列,并且能够发现卷积网络和循环网络。

## 2 深度神经网络结构搜索

对于搜索空间而言,以搜索计算单元作为最终结构的构建块,学习的单元可以堆叠形成卷积网络或者递归地连接以形成循环网络<sup>[5]</sup>。每个单元是有向无环图,由  $N$  个节点的有序序列组成。每个节点  $x^{(j)}$  是潜在表示(例如卷积网络中的特征映射),每个有向边  $(i, j)$  都与每个节点的  $x^{(j)}$  某些  $o^{(i,j)}$  转换操作相关联,假设每个单元有 2 个输入节点和单个输出节点。对于卷积单元,输入节点在前 2 层中定义了单元输出,通过对所有中间节点应用缩小操作(例如连接)来获得单元的输出。每个中间节点的计算都是基于它的所有前导节点计算的:

$$x^{(j)} = \sum_{j < i} o^{(i,j)} x^{(i)} \quad (1)$$

为了将离散的搜索空间转为连续的,首先定义一组候选操作  $O$ ,例如卷积、最大池化,表示应用于  $x^{(j)}$  的一些函数  $o(\cdot)$ ,为了使搜索空间连续,将特定操作的分类选择放宽为所有可能操作的输出层的激励函数(SoftMax):

$$\bar{o}^{(i,j)}(x) = \sum_{o \in O} \frac{\exp(\alpha_o^{(i,j)})}{\sum_{o' \in O} \exp(\alpha_{o'}^{(i,j)})} o(x) \quad (2)$$

式中:一对节点  $(i, j)$  的操作混合权重由维度  $|O|$  的向量  $\alpha^{(i,j)}$  参数化,在放宽操作后,结构搜索任务就变为了学习一组连续变量:

$$\alpha = \{\alpha^{(i,j)}\} \quad (3)$$

在搜索结束时,通过替换每个混合运算  $\bar{o}^{(i,j)}$  中最可能的运算获得离散结构,即:

$$o^{(i,j)} = \arg \max_{o \in O} \alpha_o^{(i,j)} \quad (4)$$

式中: $\alpha$  为(编码)体系结构。

放宽之后,目标就是在所有混合操作(例如卷积滤波器)的权重内共同学习结构  $\alpha$  和权重  $\omega$ ,类似于使用强化学习或进化方法的结构搜索将验证集的性能作为奖励,可微神经结构搜索的目标是

优化验证损失,但是需要使用基于梯度下降的方法来计算。

分别用  $L_{\text{train}}$  和  $L_{\text{val}}$  表示训练损失和验证损失,这2种损失不仅由结构  $\alpha$  确定,而且由网络中的权重  $\omega$  确定,结构搜索的目标是找到使验证损失  $L_{\text{val}}(\omega^*, \alpha^*)$  最小化的  $\alpha^*$ ,其中通过最小化训练损失来获得与结构相关联的权重  $\omega^*$ ,即:

$$\omega^* = \arg \min_{\omega} L_{\text{train}}(\omega, \alpha^*) \quad (5)$$

这意味着一个双层优化问题,  $\alpha$  为上层变量,  $\omega$  为下层变量:

$$\min_{\alpha} L_{\text{val}}(\omega^*(\alpha), \alpha) \quad (6)$$

$$\text{s. t. } \omega^*(\alpha) = \arg \min_{\omega} L_{\text{train}}(\omega, \alpha) \quad (7)$$

嵌套公式也出现在基于梯度的超参数优化中,尽管连续体系结构  $\alpha$  维数远高于标量值超参数(如学习率),并且更难以优化,但是在某种意义上,连续体系结构  $\alpha$  也可以被视为一种特殊类型的超参数。

完全解决双层优化是非常困难的,因为无论  $\alpha$  发生任何变化,都需要通过求解式(7)来重新计算  $\omega^*(\alpha)$ 。因此,可以使用近似迭代优化的方法,其中  $\omega$  和  $\alpha$  分别通过在权重和结构空间中的梯度下降步骤之间交替优化。在步骤  $k$ , 给定当前结构  $\alpha_{k-1}$ , 通过在最小化训练损失  $L_{\text{train}}(\omega_{k-1}, \alpha_{k-1})$  的方向上移动  $\omega_{k-1}$  来获得  $\omega_k$ , 然后,在保持权重  $\omega_k$  不变的情况下对结构进行更新,在权重梯度下降一步后最小化验证损失:

$$L_{\text{val}}(\omega_k - \xi \nabla_{\omega} L_{\text{train}}(\omega_k, \alpha_{k-1}), \alpha_{k-1}) \quad (8)$$

式中:  $\xi$  是该虚拟梯度步骤的学习率。

式(8)的目的是找到一个结构,当它的权重通过(单步)梯度下降优化时具有低验证损失,其进一步展开的权重用作  $\omega^*(\alpha)$  的替代。值得注意的是,当前动态迭代算法定义了  $\alpha$  的优化器(领导者)和  $\omega$  的优化器(跟随者)之间的斯塔克伯格(Stackelberg)博弈,这通常要求领导者预测跟随者的下一步移动以达到平衡,虽然目前还没有意识到优化算法的收敛保证,但实际上它能够选择合适的  $\xi$  收敛。除此以外,当冲量被用于权重优化时,一步前进学习目标(8)被相应地修改,并且上述分析仍然适用。

通过对式(8)中的  $\alpha$  求微分来求出体系结构梯度(为了简便起见,省略步骤索引  $k$ ):

$$\nabla_{\alpha} L_{\text{val}}(\omega', \alpha) - \xi \nabla_{\alpha, \omega}^2 L_{\text{train}}(\omega, \alpha) \nabla_{\omega}' L_{\text{val}}(\omega', \alpha) \quad (9)$$

其中,

$$\omega' = \omega - \xi \nabla_{\omega} L_{\text{train}}(\omega, \alpha) \quad (10)$$

式(10)表示一步正演模型的权重,梯度(9)在其第2项中包含矩阵向量乘积,其计算成本高,使用有

限差分近似可以大大降低复杂性。假设  $\epsilon$  是一个很小的实数的平方,有:

$$\omega^+ = \omega + \epsilon \nabla_{\omega}' L_{\text{val}}(\omega', \alpha) \quad (11)$$

和

$$\omega^- = \omega - \epsilon \nabla_{\omega}' L_{\text{val}}(\omega', \alpha) \quad (12)$$

则:

$$\frac{\nabla_{\alpha, \omega}^2 L_{\text{train}}(\omega, \alpha) \nabla_{\omega}' L_{\text{val}}(\omega', \alpha) \approx \nabla_{\alpha} L_{\text{train}}(\omega^+, \alpha) - \nabla_{\alpha} L_{\text{train}}(\omega^-, \alpha)}{2\epsilon} \quad (13)$$

评估有限差分只需要2个向前传递的权重和2个向后传递的  $\alpha$ , 并且复杂性  $O(|\alpha| |\omega|)$  降低到  $O(|\alpha| + |\omega|)$ 。

在获得连续的体系结构编码  $\alpha$  后,离散结构通过以下方式生成:

1) 保留每个中间节点的  $k$  个最强的前导,其中边缘的强度定义为:

$$\max_{o \in O, o \neq \text{zero}} \frac{\exp(\alpha_o^{(i,j)})}{\sum_{o' \in O} \exp(\alpha_{o'}^{(i,j)})} \quad (14)$$

为了生成的结构与现有工作中的结果相当,对卷积网络使用  $k=2$ 。

2) 通过采用函数  $\arg\max$  将每个混合操作替换为最可能的操作。

## 3 实验设计

### 3.1 实验参数设置

实验采用由重复单元组成的神经网络结构,其中重复单元有正常单元和缩小单元,生成的神经网络的层数即为重复单元的个数,完整的实验包括2个阶段:结构搜索和结构评估。在结构搜索阶段,使用可微神经结构搜索对多个结构进行研究,并根据其验证性能确定最佳单元;在结构评估阶段,使用这些单元构建更大的结构,从头开始训练并在测试集上测试它们的性能。

实验中,上文提及的一组候选操作  $O$  中包括以下操作:  $3 \times 3$  和  $5 \times 5$  可分离卷积、 $3 \times 3$  和  $5 \times 5$  扩张可分离卷积、 $3 \times 3$  最大池化、 $3 \times 3$  平均池化、Identity 和 Zero,所有操作的步长都是1(如果适用),并且填充卷积特征图以保持其空间分辨率,使用 ReLU-Conv-BN 顺序进行卷积运算,并且每个可分离卷积总是应用2次。

每个卷积单元由  $N=7$  个节点组成,其中输出节点被定义为所有中间节点(排除的输入节点)的深度级联。其余的设置遵循 Zoph 等<sup>[6]</sup>、Liu 等<sup>[7]</sup> 和 Real 等<sup>[8]</sup>, 然后通过将多个单元堆叠在一起形成网络。

卷积单元  $k$  的第 1 和第 2 节点分别设置为卷积单元  $k-2$  和卷积单元  $k-1$  的输出,并且根据需要插入  $1 \times 1$  个卷积。位于网络总深度的  $1/3$  和  $2/3$  处的小区域是缩小单元,其中与输入节点相邻的所有操作步长都是 2。因此,体系结构编码是  $(\alpha_{normal}, \alpha_{reduce})$ ,其中  $\alpha_{normal}$  由所有正常单元共享, $\alpha_{reduce}$  由所有缩小单元共享。

### 3.2 结构搜索

将训练集中 40% 的数据作为验证集,其余部分作为训练集。共进行 2 次结构搜索,第 1 次网络层数为 8,训练周期 50 个,第 2 次网络层数为 12。由于 GPU 显存有限,而数据集原始图像分辨率为  $256 \times 256$ ,为了保证输入图片可以在 GPU 上进行运算,在加载数据集时,将数据分辨率压缩至  $32 \times 32$ 。

其他超参数设置如下:每次训练数据输入批量大小为 64(对于训练集和验证集),初始通道数为 16,使用动量 SGD 来优化权重  $\omega$ ,初始学习速率  $\eta_{\omega}=0.025$ (在余弦调度(cosine schedule)之后衰减到 0)、动量大小为  $\beta=0.9$  和权重以  $\omega=3 \times 10^{-4}$  的大小衰减,使用 Adam 作为结构变量的优化器(即正常和缩小单元中的  $\alpha$ ),初始学习率  $\eta_{\alpha}=3 \times 10^{-4}$ ,动量  $\beta=(0.5, 0.999)$ ,权重以  $\omega=3 \times 10^{-3}$  的大小衰减。

网络层数为 8 的小网络在单个 GPU 上进行结构搜索花费了大约 40 min,网络层数为 12 的小网络花费了 1 h。

### 3.3 结构验证

对一个有 20 层的大型网络进行训练,训练时间为 600 个时期,批量数据输入大小为 96,图片大小仍然为  $32 \times 32$ ,其他超参数与结构搜索的超参数保持相同,增加其他增强功能包括图片剪切、随机失活概率为 0.3 和权重为 0.4 的辅助节点。由 8 层小网络卷积单元组成的大网络在单个 GPU 上进行结构验证花费了 1 h 24 min,由 12 层小网络卷积单元组成的大网络花费了 1 h 15 min。

## 4 结果与分析

训练数据对基于深度学习的目标识别至关重要。针对城市中常见的飞行物类型,本文通过网络收集与实际拍摄,收集了丰富的目标影像数据,包括 3 类目标:固定翼无人机、多旋翼无人机与飞鸟,各 1 000 张。影像分辨率固定为  $256 \times 256$ ,以 jpg 格式保存。数据集实例见图 1。



图 1 数据集实例

### 4.1 结构搜索得到的正常单元和缩小单元

“低慢小”目标数据集共进行了 2 次实验,共得到 4 种基本单元,见图 2~5。在每个单元中  $c_{k-1}$  和  $c_{k-2}$  表示前 2 个单元的输出作为本单元的输入, $c_k$  表示本单元的输出。从结果来看,不同层数的网络,其基本的卷积单元不同,同时每个卷积单元每个节点处有多个输入和输出,构建了一个复杂的网络结构。

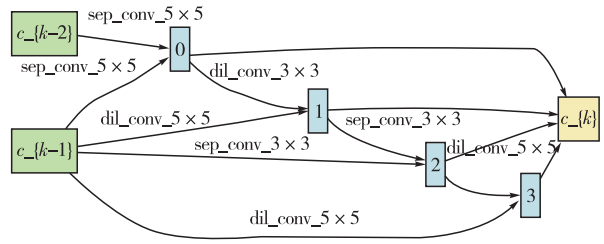


图 2 网络层数为 8 的正常单元

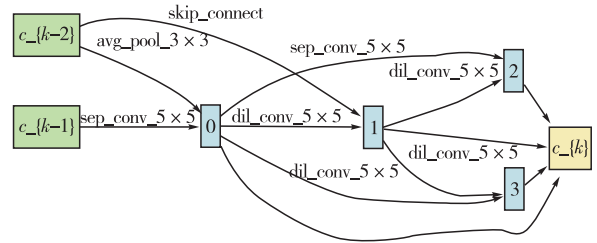


图 3 网络层数为 8 的缩小单元

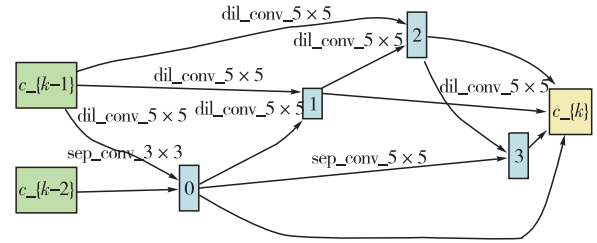


图 4 网络层数为 12 的正常单元

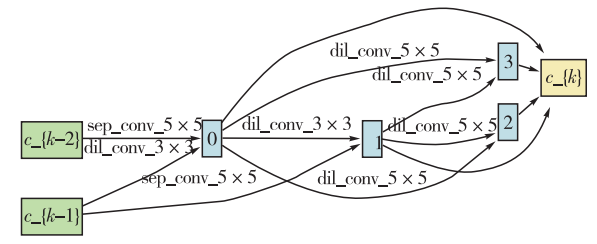


图 5 网络层数为 12 的缩小单元

### 4.2 目标数据集分类结果

通过对数据集进行实验,得到了当前数据集的最优神经网络结构,在找到结构后,通过在验证集上

验证,得到相应的测试结果,见表1。

表1 使用可微神经结构搜索的深度神经网络测试结果

| 网络层数 | 测试精度/% | 参数大小/<br>Mbytes | 结构搜索<br>时间/min | 结构评估<br>时间/min |
|------|--------|-----------------|----------------|----------------|
| 8    | 96.50  | 3.99            | 40             | 84             |
| 12   | 97.25  | 3.68            | 60             | 75             |

从表1可以得知,城市飞行物数据集的2个实验的分类精度分别为96.50%和97.25%,满足不小于95%的精度要求。

最优精度与人工设计的常用网络对比见表2。显然,本文方法具有更好的测试精度,在用于“低慢小”无人机类目标分类检测方面具有更强的适用性。

表2 不同结构神经网络在目标数据集上的测试结果

| 方法                     | 测试精度/% |
|------------------------|--------|
| AlexNet <sup>[9]</sup> | 91.00  |
| VGG16 <sup>[10]</sup>  | 93.25  |
| ResNet <sup>[11]</sup> | 94.75  |
| 本文                     | 97.25  |

## 5 结语

本文利用神经结构搜索可以在给定目标数据集上得到当前任务最优的深度神经网络,且在“低慢小”目标识别上取得了良好的效果,但这一方法仍然需要一定量的有标签数据。因此,在后续的目标识别方法研究过程中应结合当前在小样本数据进行分类的研究,将其运用在神经结构搜索中,提升样本数据量较小时的分类精度,防止过拟合现象的发生;同时,在面对无标签影像时,考虑是否可以通过结合神经结构搜索获取的高层语义特征与其他方法获取的有关图像的中、低层语义特征,从而提升分类精度。

## 参考文献

[1] 卞伟伟,邱旭阳,刘方.“低慢小”目标协同探测综合识别关键技术研究[C]//第五届中国指挥控制大会论文集,北京:中国指挥与控制学会,2017:264-269.

- [2] LE C Y, BENGIO Y, HINTON G. Deep Learning [J]. Nature, 2015, 521(7553): 436-444.
- [3] SHIN R, PACKER C, SONG D. Differentiable-Neural Network Architecture Search[C]//6th International Conference on Learning Representations. Vancouver, Canada: ICRL, 2018:1-4.
- [4] PAN S J, YANG Q. A Survey on Transfer Learning [J]. IEEE Transactions on Knowledge and Data Engineering, 2009, 22(10): 1345-1359.
- [5] AHMED K, TORRESANI L. Connectivity Learning in Multi-Branch Networks[Z]. ArXiv Preprint:1709.09582, 2017.
- [6] ZOPH B, VASUDEVAN V, SHLENS J, et al. Learning Transferable Architectures for Scalable Image Recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018: 8697-8710.
- [7] LIU C, ZOPH B, NEUMANN M, et al. Progressive Neural Architecture Search [C]//Proceedings of the European Conference on Computer Vision (ECCV). Munich Germany: IEEE, 2018: 19-34.
- [8] REAL E, AGGARWAL A, HUANG Y, et al. Regularized Evolution for Image Classifier Architecture Search[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Hawaii, USA: AAAI, 2019: 4780-4789.
- [9] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet Classification with Deep Convolutional Neural Networks [C]//Advances in Neural Information Processing Systems. Lake Tahoe, USA: NIPS, 2012: 1097-1105.
- [10] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition [Z]. arXiv Preprint:1409.1556, 2014.
- [11] HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016: 770-778.

(编辑:徐敏)