

# 基于 AutoML 的保护区物种识别<sup>①</sup>



刘耀<sup>1,2</sup>, 罗泽<sup>1</sup>

<sup>1</sup>(中国科学院 计算机网络信息中心, 北京 100190)

<sup>2</sup>(中国科学院大学, 北京 100049)

通讯作者: 罗泽, E-mail: [luoze@cnic.cn](mailto:luoze@cnic.cn)

**摘要:** 随着我国在生态保护上的投入加大, 红外相机技术在我国各级自然保护区的应用发展迅猛, 在如何充分挖掘照片的信息方面, 物种识别显得尤为重要, 是其他工作的前提. 在图像识别方面, 随着深度学习的爆发, 给图像识别带来了革命性的提升, 以卷积神经网络为代表的网络结构在准确率上几乎完胜传统方法. 然而, 由于网络结构对最终图像识别准确率的影响巨大, 人们在实际应用中往往都是使用一些经典的网络结构, 比如 VGG16、VGG19、ResNet50 等, 从中选择一个适合自己的数据集的网络结构, 同时对于不同的数据集, 可能需要重新选择. 因此, 在保护区红外相机物种的识别中, 本文提出了基于 AutoML 的自动构建网络结构技术, 针对不同的保护区的数据集, 自动构建合适的网络结构, 避免人工选择, 同时达到了与人工选择网络相当的准确率.

**关键词:** 物种识别; 自动机器学习; 自动构建网络结构

引用格式: 刘耀, 罗泽. 基于 AutoML 的保护区物种识别. 计算机系统应用, 2019, 28(9): 147-153. <http://www.c-s-a.org.cn/1003-3254/7038.html>

## Species Recognition of Protected Area Based on AutoML

LIU Yao<sup>1,2</sup>, LUO Ze<sup>1</sup>

<sup>1</sup>(Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China)

<sup>2</sup>(University of Chinese Academy of Sciences, Beijing 100049, China)

**Abstract:** With the increase of investment in ecological protection, the application of infrared camera technology in natural reserves has developed rapidly. Species recognition, which is particularly important in how to fully mine photo information, is the premise of other work. In image recognition, with the outbreak of deep learning, the image recognition has been revolutionized. Convolutional neural network as the representative network structure almost completely overcomes the traditional method in accuracy. However, due to the huge impact of the network structure on the accuracy of the final image recognition, people often choose a network structure suitable for their own dataset from some classic network structures, such as VGG16, VGG19, ResNet50, and so on, in practical applications. Nevertheless, it may need to re-select network structure for different datasets. Therefore, in the species recognition of protected area, this study proposes an automatic construction network structure technology based on AutoML. The technology can automatically build appropriate network structures for different datasets of protected area to avoid manual selection of network structures. At the same time, the technology achieves an accuracy comparable to manual selection of network structures.

**Key words:** species recognition; auto machine learning; auto construct network structures

① 收稿时间: 2019-02-27; 修改时间: 2019-03-15; 采用时间: 2019-03-18; csa 在线出版时间: 2019-09-05

近年来,红外相机技术因为其不受环境限制,全天不间断的特点,在各级保护区得到了广泛应用<sup>[1-3]</sup>。目前有大量的照片数据,如何充分利用这些数据,从中挖掘有效的信息,成为了人们研究的重点。其中物种识别技术是很多工作的前提,比如物种行为分析、物种多样性检测等,都得先从图片中识别出物种。

深度学习技术的出现给各领域都带来了革命性的变化。在图像领域中,以卷积神经网络(Convolutional Neural Network, CNN)为代表的深度学习网络结构,在图像识别<sup>[4]</sup>、目标检测<sup>[5]</sup>、图像分割<sup>[6]</sup>等各项工作中,都以绝对的优势战胜了传统方法,带来了显著的提升。传统方法利用手工提取特征,包括纹理、颜色、灰度、梯度等特征,然后将这些手工提取的浅层特征输入到分类器中,进行识别。深度学习可以学习深层特征,而且理论上讲,网络越深,学习到的特征越高级,是低阶特征的组合,同时,深度学习还是端到端的,特征提取与分类一起进行。但是网络结构并不是越深越好,随着网络结构的加深,会出现梯度爆炸、梯度消失<sup>[7]</sup>、过拟合等问题,虽然现在有一些技术,比如加BN<sup>[8]</sup>(Batch Normalization)、ReLU<sup>[9]</sup>(Rectified Linear Units)、Dropout<sup>[10]</sup>、Data Augmentation<sup>[11]</sup>等,可以在一定程度上缓和这些由网络结构加深带来的问题,但是也只是缓和,并不能从根本上解决。所以卷积神经网络在发展的过程就是朝着越来越深的方向发展,同时会有一些新的结构单元出现,比如VGG<sup>[12]</sup>、Inception<sup>[13]</sup>、ResNet<sup>[14]</sup>、DenseNet<sup>[15]</sup>等。在实际应用中,人们往往不会自己设计一个网络结构,而是使用一些已经证明有效的网络结构,比如VGG16、VGG19、ResNet50等。没有一个网络结构可以保证在任何数据集上都优于其它网络结构,对于一个特定的数据集,我们需要进行实验,根据实验结果选择一个性能最好的网络结构。所以,当拿到一个新的数据集的时候,我们需要重新进行实验选择。

这两年自动机器学习(Auto Machine Learning, AutoML)技术正在兴起,虽然刚刚起步,但已经取得了不错的成绩。AutoML技术就是一种让机器学习怎么学习的算法,在传统机器学习就主要表现在自动调参<sup>[16-18]</sup>,在深度学习上就主要表现在网络架构搜索<sup>[19-22]</sup>(Neural Architecture Search, NAS),因为在深度学习最重要的参数就是网络结构的层数,每层神经元的个数等。

因此,本文将基于AutoML的图像识别技术应用

到保护区的红外相机物种识别上,对于不同保护区的物种图片数据,自动设计适合该数据集的网络结构,而不需要自己手动选择网络结构,在保证准确率的前提下,节省了人力。

## 1 相关理论

### 1.1 贝叶斯优化

贝叶斯优化是AutoML中的常用的方法<sup>[23-25]</sup>,它是一种典型的黑盒优化算法,就是只能观察到输入和输出,无法知道输入和输出之间的关系以及梯度信息,贝叶斯优化通常用高斯过程模型来捕捉输入和输出之间的关系<sup>[26-27]</sup>。

高斯过程是假设你的输入之间服从高斯分布,通过已知的输入和输出 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ 来拟合高斯分布模型,当你给出一个新的输入 $x$ 时,高斯过程模型能给出输出 $y$ 的概率分布,因为假设输出 $y$ 是服从高斯分布的,也就是能给出输出的均值 $\mu$ 和方差 $\sigma$ 。得到均值和方差后,通常还需要一个采集函数(Acquisition Function, AC),以该采集函数的值作为准则来判断这个输入 $x$ 的好坏。

在AutoML中,贝叶斯优化通常被用来捕捉参数和模型性能之间的关系,通过贝叶斯优化器给出下一个最有“潜力”的参数组合,训练完后返回模型性能用来优化高斯过程模型,反复迭代,快速找到最优的参数组合。

使用高斯过程模型来捕捉输入和输出关系的贝叶斯优化算法过程:

- (1) 假设输入之间服从高斯分布模型  $M$ 。
- (2) 从  $M$  中选择下一个采集函数值较高的输入  $x$ 。
- (3) 观察输入  $x$  的输出  $y$ , 如果  $y$  满足要求, 那么结束算法; 否则, 返回  $(x, y)$  修正高斯分布模型  $M$ , 返回第(2)步。

### 1.2 模拟退火

模拟退火算法是一种常见的搜索算法,常用在一定时间内在一个很大的搜索空间中寻找近似最优解<sup>[28]</sup>。

模拟退火算法的思想借鉴了固体的退火原理,当固体的温度很高的时候,内能比较大,固体的内部粒子处于快速无序运动,当温度慢慢降低的过程中,固体的内能减小,粒子的慢慢趋于有序,最终,当固体处于常温时,内能达到最小,此时,粒子最为稳定。模拟退火算

法便是基于这样的原理设计而成. 模拟退火算法从某一较高的温度出发, 这个温度称为初始温度, 伴随着温度参数的不断下降, 算法中的解趋于稳定, 但是, 可能这样的稳定解是一个局部最优解, 此时, 模拟退火算法中会以一定的概率跳出这样的局部最优解, 以寻找目标函数的全局最优解. 这个概率就与温度参数的大小有关, 温度参数越大, 概率越大, 反之则越小.

模拟退火的算法过程:

(1) 初始化: 初始温度  $T$  (充分大), 初始解状态  $S$  (是算法迭代的起点), 每个  $T$  值的迭代次数  $L$ .

(2) 对  $k = 1, 2, \dots, L$  做第 (3) 至第 (6) 步.

(3) 产生新解  $S'$ .

(4) 计算增量  $\Delta T = C(S') - C(S)$ , 其中  $C(S)$  为评价函数.

(5) 若  $\Delta T < 0$ , 则接受  $S'$  作为新的当前解, 否则以概率  $e^{-\Delta T/T}$  接受  $S'$  作为新的当前解.

(6) 如果满足终止条件则输出当前解作为最优解, 结束程序.

终止条件通常取为连续若干个新解都没有被接受时终止算法.

(7)  $T$  逐渐减少, 且  $T \leftarrow 0$ , 然后转第 2 步.

## 2 本文方法

本文使用的基于 AutoML 的图像识别技术, 能够针对特定的数据集自动设计网络结构, 不需要人工干预, 主要思想如图 1 所示.

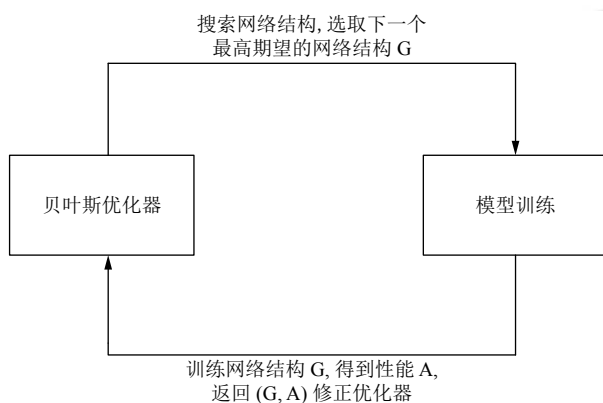


图 1 AutoML 技术

贝叶斯优化器模块主要负责的是网络结构的自动构建与搜索任务, 从旧的网络结构构建出新的网络结构, 然后搜索出下一个网络结构用于训练, 得到的性能

用来修正这个优化器. 模型训练模块主要负责训练, 给定一个网络结构, 返回在验证集合上的准确率.

### 2.1 经典的网络结构组件

经典的网络结构组件主要有 3 种: CNN (Convolutional Neural Network)、MLP (Multi-Layer Perceptron)、ResNet (Residual Network), 这是构成网络结构的基本组成部分, 比如 VGG、Inception、ResNet、DenseNet 等网络结构都是由这 3 种基本组件组成, 只是组件个数与组成方式不同而已.

### 2.2 自动构建网络结构

网络结构的构建是从小到大的过程, 所以网络结构会越来越大. 初始的网络结构有 3 个, 分别由 CNN、MLP、ResNet 基本组件组成, 然后由这 3 个网络结构开始扩增. 有一个候选队列  $Q$ , 里面是候选的网络结构, 从队列头取出网络结构  $G$ , 对  $G$  进行扩增, 扩增的方法如图 2 所示, 主要有 4 种: 1)  $deep(G, u)$ , 加深网络结构, 随机选择一层  $u$ , 在层  $u$  处加入新的层使网络结构变深; 2)  $wide(G, u)$ , 扩宽网络结构, 随机选择一层  $u$ , 在层  $u$  处加入新的节点使网络结构变宽, 这里的加入新的节点指的是卷积层中的卷积核, 也就是增加卷积层  $u$  中卷积核的个数; 3)  $add(G, u, v)$ , 随机选择一层  $u$  和  $v$ , 增加跳跃连接, 给层  $u$  和层  $v$  增加一条跳跃连接, 使用 ResNet 的思想; 4)  $concat(G, u, v)$ , 随机选择一层  $u$  和  $v$ , 将层  $u$  和层  $v$  合并;

主要步骤为:

(1) 从候选队列头取出网络结构  $G$ .

(2) 将  $G$  进行上述 4 种扩展操作, 得到 4 个新的网络结构, 将新的网络添加到候选队列中.

### 2.3 搜索网络结构

对于自动构建的网络结构, 并不是每个新网络结构都训练, 而是挑选最有“潜力”的网络结构, 也就是最有可能获得最好性能的网络. 这里的网络结构可以看作输入, 而模型性能可以看作输出, 这是一种典型的黑盒优化问题. 关于黑盒优化算法, 典型的有网格搜索、随机搜索、贝叶斯优化. 网格搜索就是遍历所有的输入, 观察每个相应的输出, 最后选取最好输出对应的输入作为问题的解. 随机搜索就是随机选择一个输入, 观察相应的输出, 最后选取最好输出对应的输入作为问题的解. 可以看到网格搜索相对耗时, 但可以得到问题的最优解, 而随机搜索虽然不耗时, 但具有随机性, 解的方差较大. 贝叶斯优化既不用遍历所有输入, 又可以

获得有一定置信度的解。

将贝叶斯优化应用到网络结构的搜索,也就是在训练了 $n$ 个网络后,观察到了这 $n$ 个网络结构的性能,就可以更新贝叶斯优化控制器.当给定一个新的网络 $f$ ,

可以得到新的网络结构性能的均值 $\mu(f)$ 和方差 $\sigma(f)$ . $\mu(f)$ 越大,说明网络结构 $f$ 性能均值越好,也就是期望越高, $\alpha(f)$ 越大,说明网络结构 $f$ 的性能方差越大,但说明不确定性也越大。

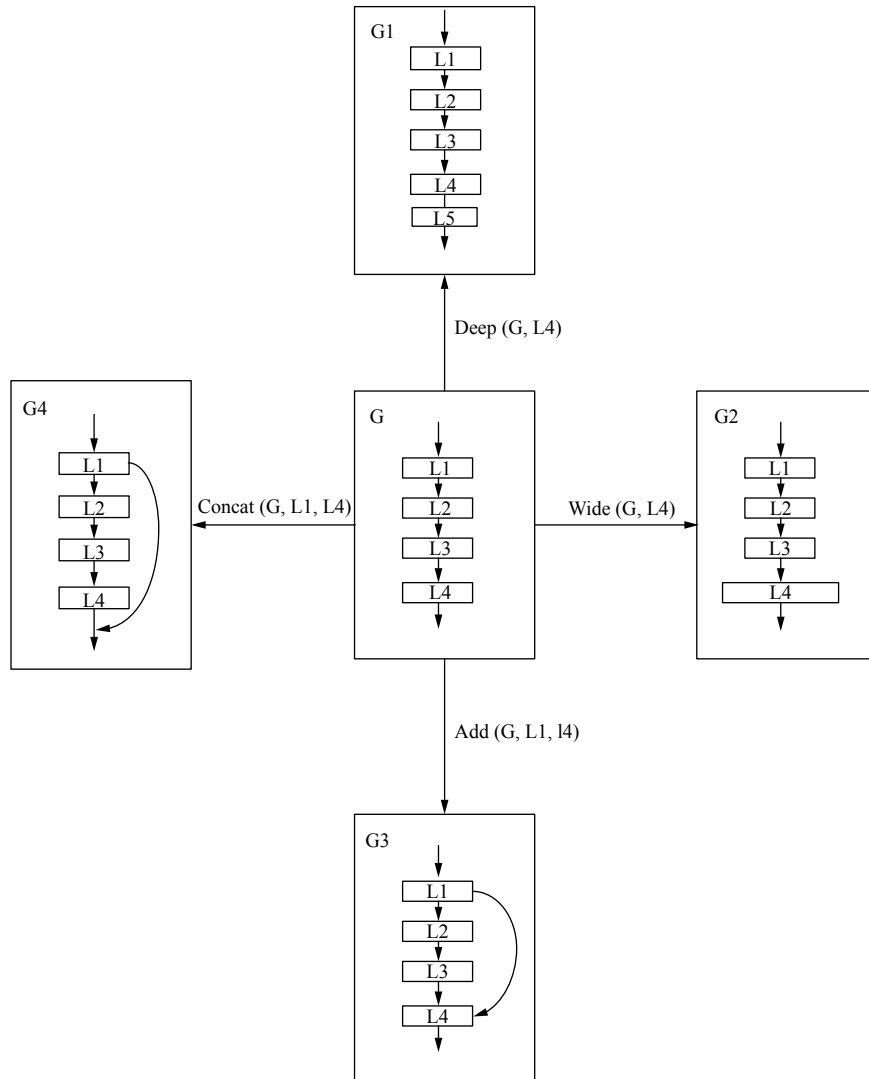


图2 自动构建网络结构

选择网络结构时,我们要平衡探索与利用,探索指的是倾向于选择那些不确定性大的网络结构,利用指的是倾向于选择那些性能期望高的网络结构,最终网络结构 $f$ 性能的采集函数是

$$\alpha(f) = \mu(f) + \beta\sigma(f) \quad (1)$$

$\beta$ 是一个平衡因子,来平衡探索与利用.得到每个网络结构的 $\alpha$ 值之后,通过模拟退火算法来选取下一个网络结构作为备训练的网络结构.

搜索算法的算法过程:

- (1) 初始化模拟退火的温度衰减率 $r$ ,温度参数 $T$ 以及最低温度阈值 $T_{low}$ ,最高历史模型性能值 $c_{max}$ ,最优网络结构 $f_{max}$ ,优先级搜索队列 $Q$ .
- (2) 取出队列头网络结构 $f$ ,对该网络结构做上述四种操作进行扩增,得到四个新的网络结构,对于每个新网络结构 $f'$ ,如果 $e^{\frac{\alpha(f') - c_{max}}{T}} > Rand()$ ,将该网络加入到优先级搜索队列中,否则不加入队列.如果 $c_{max} <$

$\alpha(f')$ , 那么  $c_{\max} \leftarrow \alpha(f')$ ,  $f_{\max} \leftarrow f'$ . 同时衰减温度  $T \leftarrow T \times r$ .

(3) 如果队列不为空, 并且  $T > T_{\text{low}}$ , 返回第二步; 否则, 返回最优网络结构  $f_{\max}$ .

## 2.4 算法流程

算法流程图如图3所示.

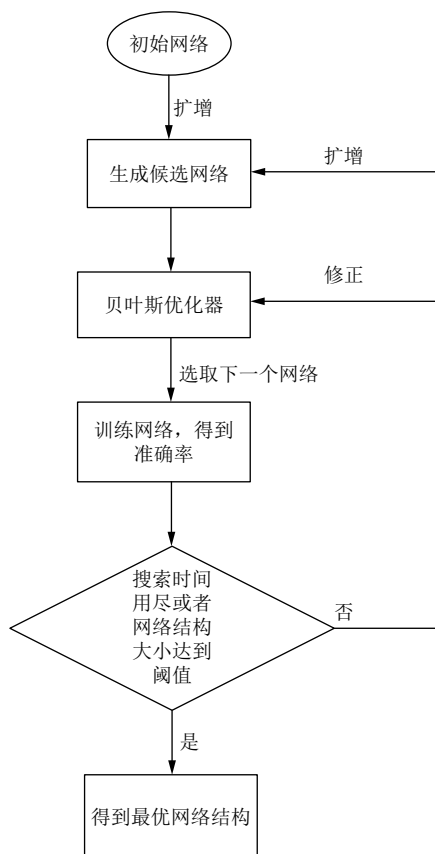


图3 算法流程图

## 3 实验分析

### 3.1 数据集

本文在保护区的红外照片数据集上验证了基于 AutoML 的图像识别方法. 本文从两个保护区的数据图片中制作了四个数据集, 为了使不同数据集之间具有可比性, 尽量保持不同数据集之间相似性, 比如 1) 数据图片都是鸟类图片, 不同类别是不同的鸟类; 2) 每个数据集类别数相同, 都是十类; 3) 样本数量相差不大, 每个类别大概 900 张左右; 4) 使用相同的方式划分训练数据和测试数据, 测试数据是从训练数据的每类中随机抽取 20 张得到的, 测试数据一共 200 张, 可以保证测试数据中不会因为类别的不均衡导致某些类

别测试样本过少.

数据处理方面主要有两点: 1) 数据归一化, 图片像素值的范围是  $[0, 255]$ , 将数据乘以  $1 / 255.0$  缩放到  $[0, 1]$ ; 2) 数据增强, 为了增强数据以得到更好的性能, 本文通过将数据水平翻转、缩放、随机裁剪、增加噪声等技术对数据进行增强.

### 3.2 模型训练

对于自动构建的网络结构和经典的网络结构 VGG16、VGG19、ResNet50, 都使用相同的训练参数, 来验证自动构建网络的有效性, 训练参数如表所示. 训练过程都使用了一些深度学习的训练技巧, 比如在 Dense 层加入 Batch Normalization 来防止过拟合, 同时防止梯度消失, 使得训练更深的网络成为可能.

在使用经典网络时, 人们往往都会使用 fine-tune 的方式, fine-tune 的一般步骤为: 1) 使用公开的权重来初始化参数, 先固定卷积层和池化层, 保持它们的参数不更新, 只更新全连接层的参数; 2) 在训练一定的轮次后, 不再固定卷积层和池化层的参数, 调小学习率, 再训练整个网络. 当你的数据量不足时, 这样做往往能取得不错的效果. 对于经典的网络结构, 公开的权重是别人使用大量的数据训练得到的, 这样的权重提取到的特征对于大部分任务来说, 往往是有效的, 如果使用公开的权重来初始化, 可以省去你的时间. 相当于使用公共的卷积核来提取低级特征, 因为很多数据集的低级特征类似, 比如一些边缘特征、纹理特征等, 训练全连接层, 也就是组合低级特征, 不同的数据集组合方式可能不同, 所以第一步中要固定卷积层和池化层. 第二步再更新整个网络, 调小学习率, 针对这个数据集来进行微调. 在本文中, 经典网络使用 fine-tune 的方式, 而自动构建的网络没有使用这种方式, 因为经典网络结构是固定的, 参数个数与网络形状相同, 可以使用公开的权重来初始化, 而自动构建的网络可能从来都没有见过, 自然也就没有公开的权重来初始化.

关于训练停止的策略, 都是用 EarlyStopping 策略, 也就是网络的验证性能在  $n$  个 Epoch 后没有提升就停止训练, 验证使用的是验证数据集, 保留验证性能最高的 Epoch 的权重, 将测试性能记作该网络结构的最终性能指标. 在基于 AutoML 的方法中, 选择性能指标最高的网络结构作为最终的网络结构.

### 3.3 实验结果

在相同的 4 组数据集上, 本文做了一组对比实验.

训练集和测试集相同, 图像增强技术只作用于训练数据。

表格中 AutoML 后的 2 h、12 h、36 h 指的是网络结构搜索的小时数为 2 小时、12 小时、36 小时, 本文做实验的机器配置: CPU 为双核 4 线程, 显卡内存为 16 GB, 所以这里的搜索时间与机器配置也有关系。对于 AutoML 技术来说, 主要是网络结构搜索消耗时间, 所以 AutoML 技术的整个流程消耗时间基本上为 2 小时、12 小时、36 小时; 对于经典的网络结构来说, 由于有训练过程、调优过程以及反复的试错过程, 消耗时间均为为 3 小时左右, 其中调优过程最为耗时。由表 1 可知, 对于经典网络结构来说, 没有一个网络结构在所有数据集上性能比其他网络结构好, 这就是我们每次给定一个数据集需要选择网络结构的原因。而对于 AutoML 技术来说, 只要有足够的时间来搜索网络结构, 一定会获得较好的性能, 而且在一定的时间范围内, 搜索时间越长, 搜索到的网络性能越好。

表 1 AutoML 技术与经典网络结构的精度对比 (%)

网络结构	数据集 1	数据集 2	数据集 3	数据集 4
VGG16	74.11	70.04	79.99	82.49
VGG19	75.44	72.50	76.50	82.07
ResNet50	89.45	68.54	<b>91.51</b>	80.33
AutoML(2 h)	68.51	64.38	67.72	72.35
AutoML(12 h)	84.54	85.90	87.31	89.76
AutoML(36 h)	<b>89.62</b>	<b>88.21</b>	90.43	<b>92.59</b>

#### 4 结论与展望

本文将基于 AutoML 的图像识别技术应用到保护区的物种识别上, 解决了目前对于不同数据集需要重新训练和选择网络的问题。通过实验结果的分析, 可以看出基于 AutoML 的方法能够自动构建网络结构, 同时保证了与人工选择的经典网络相持的准确率。然而在性能上还有待提高, 从实验结果来看, 可能还需要 12 小时的网络结构搜索时间, 才能达到不错的准确率, 可以考虑在训练模型的时候, 采用适当的方法来减少模型训练时间, 比如提前终止训练, 都训练 50 步, 因为我们不需要知道最终的性能, 只需要知道相对性能, 如果一个网络在 50 步的性能比另一个网络高, 那么证明这个网络比另一个网络有“潜力”。此外, AutoML 是近年兴起的热门研究领域, 基于 AutoML 的图像识别方法有望设计出更好的网络结构, 在准确率上大幅超过原始的经典网络。

#### 参考文献

- 1 刘雪华, 武鹏峰, 何祥博, 等. 红外相机技术在物种监测中的应用及数据挖掘. 生物多样性, 2018, 26(8): 850–861. [doi: 10.17520/biods.2018053]
- 2 唐卓, 杨建, 刘雪华, 等. 基于红外相机技术对四川卧龙国家级自然保护区雪豹 (*Panthera uncia*) 的研究. 生物多样性, 2017, 25(1): 62–70.
- 3 兰慧, 金崑. 红外相机技术在北京雾灵山自然保护区兽类资源调查中的应用. 兽类学报, 2016, 36(3): 322–329.
- 4 Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe, NV, USA. 2012. 1097–1105.
- 5 Ren SQ, He KM, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks. Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal, Canada. 2015. 91–99.
- 6 Zelener A, Stamos I. CNN-based object segmentation in urban LIDAR with missing points. Proceedings of the 4th International Conference on 3D Vision. Stanford, CA, USA. 2016. 417–425.
- 7 Hochreiter S. The vanishing gradient problem during learning recurrent neural nets and problem solutions. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 1998, 6(2): 107. [doi: 10.1142/S0218488598000094]
- 8 Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. Proceedings of the 32nd International Conference on Machine Learning. Lille, France. 2015.
- 9 Agarap AF. Deep learning using rectified linear units (ReLU). arXiv preprint arXiv: 1803.08375, 2018.
- 10 Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: A simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research, 2014, 15(1): 1929.
- 11 Inoue H. Data augmentation by pairing samples for images classification. arXiv preprint arXiv: 1801.02929, 2018.
- 12 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv: 1409.1556, 2014.
- 13 Szegedy C, Liu W, Jia YQ, et al. Going deeper with convolutions. Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA. 2015. 1–9.

- 14 He KM, Zhang XY, Ren SQ, *et al.* Deep residual learning for image recognition. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. 2016. 770–778.
- 15 Huang G, Liu Z, van der Maaten L, *et al.* Densely connected convolutional networks. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA. 2017. 2261–2269.
- 16 Bergstra J, Komer B, Eliasmith C, *et al.* Hyperopt: A Python library for model selection and hyperparameter optimization. Computational Science & Discovery, 2015, 8(1): 014008.
- 17 Hutter F, Hoos H, Leyton-Brown K. An evaluation of sequential model-based optimization for expensive blackbox functions. Proceedings of the 15th Annual Conference Companion on Genetic and Evolutionary Computation. Amsterdam, The Netherlands. 2013. 1209–1216.
- 18 Hutter F, Hoos HH, Leyton-Brown K, *et al.* Time-bounded sequential parameter optimization. Proceedings of the 4th International Conference on Learning and Intelligent Optimization. Venice, Italy. 2010.
- 19 Jin HF, Song QQ, Hu X. Auto-keras: An efficient neural architecture search system. arXiv preprint arXiv: 1806.10282, 2018.
- 20 Casale FP, Gordon J, Fusi N. Probabilistic neural architecture search. arXiv preprint arXiv: 1902.05116, 2019.
- 21 Pham H, Guan MY, Zoph B, *et al.* Efficient neural architecture search via parameter sharing. arXiv preprint arXiv: 1802.03268, 2018.
- 22 Cai H, Chen TY, Zhang WN, *et al.* Efficient architecture search by network transformation. arXiv preprint arXiv: 1707.04873, 2017.
- 23 Swersky K, Snoek J, Adams RP. Freeze-thaw Bayesian optimization. arXiv preprint arXiv: 1406.3896, 2014.
- 24 Nickson T, Osborne MA, Reece S, *et al.* Automated machine learning on big data using stochastic algorithm tuning. arXiv preprint arXiv: 1407.7969, 2014.
- 25 Klein A, Falkner S, Bartels S, *et al.* Fast bayesian optimization of machine learning hyperparameters on large datasets. arXiv preprint arXiv: 1605.07079, 2016.
- 26 Snoek J, Larochelle H, Adams RP. Practical Bayesian optimization of machine learning algorithms. Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe, NV, USA. 2012. 2951–2959.
- 27 Wistuba M, Schilling N, Schmidt-Thieme L. Scalable Gaussian process-based transfer surrogates for hyperparameter optimization. Machine Learning, 2018, 107(1): 43–78. [doi: [10.1007/s10994-017-5684-y](https://doi.org/10.1007/s10994-017-5684-y)]
- 28 Steinbrunn M, Moerkotte G, Kemper A. Heuristic and randomized optimization for the join ordering problem. The VLDB Journal, 1997, 6(3): 191–208. [doi: [10.1007/s007780050040](https://doi.org/10.1007/s007780050040)]