

# 利用可微运算的重要性 神经架构搜索

袁州

, 高级会员, IEEE, 谢旭凯

, 和 Sun-Yuan Kung, IEEE 终身院士

**摘要** 最近,可微分神经架构搜索 (NAS) 方法在降低 NAS 的计算成本方面取得了重大进展。现有方法通过选择具有更高架构权重的候选操作来搜索最佳架构。然而,架构权重不能准确反映每个操作的重要性,即权重最高的操作可能与最佳性能无关。为了避免这一缺陷,我们提出了一种新的指标,可以充分代表操作的重要性,从而作为指导模型搜索的有效指标。基于这个指标,我们进一步开发了一个 NAS 方案,用于“利用有效 NAS 的操作重要性”(EoiNAS)。更准确地说,我们提出了一种基于高阶马尔可夫链的策略来缩小搜索空间,进一步提高搜索效率和准确性。为了评估所提出的 EoiNAS 的有效性,我们将我们的方法应用于两个任务:图像分类和语义分割。

对这两项任务的广泛实验提供了强有力的证据,证明我们的方法能够发现高性能架构,同时保证搜索过程中必要的效率。

**索引词** 高阶马尔可夫链,图像分类,神经结构搜索 (NAS),语义分割。

## 一、引言

为特定问题找到合适的网络架构是一项具有挑战性的任务。更好的网络架构通常会带来显著的性能提升。近年来,神经架构搜索 (NAS) 技术 [1]-[8] 已证明在自动化神经架构设计方面取得了成功。许多由 NAS 方法产生的架构在任务中取得了比手动设计的架构更高的精度,例如图像分类 [1]、[9]、超分辨率 [10]、语义分割 [11]-[13] 和对象检测 [14]。NAS 方法不仅可以提高模型性能,还可以将人类专家从调整架构的繁琐任务中解放出来。

到目前为止,NAS 主要使用了三个主要框架:基于进化算法 (EA) 的 NAS [7]、

手稿于 2020 年 3 月 17 日收到; 2020 年 10 月 6 日和 2021 年 3 月 28 日修订; 2021 年 3 月 31 日接受。出版日期 2021 年 5 月 17 日;当前版本的日期为 2022 年 10 月 28 日。这项工作部分得到了中国国家自然科学基金资助 U2006211 的支持,部分得到了国家重点研发项目资助 2020YFC1523204 的支持。(通讯作者:袁州。)

Yuan Zhou 和 Xukai Xie 来自天津大学电气与信息工程学院,天津 300072,中国(电子邮件:zhouyuan@tju.edu.cn;xxie@tju.edu.cn)。

Sun-Yuan Kung 就职于普林斯顿电气工程系  
美国新泽西州普林斯顿大学 08540 (电子邮件:kung@princeton.edu)。  
本文中一个或多个图形的彩色版本可在  
<https://doi.org/10.1109/TNNLS.2021.3072950>。  
数字对象标识符 10.1109/TNNLS.2021.3072950

2162-237X © 2021 IEEE。允许个人使用,但再版/再分发需要 IEEE 许可。

有关详细信息,请参阅 <https://www.ieee.org/publications/rights/index.html>。

[16]、[17]、基于强化学习 (RL) 的 NAS [1]、[2]、[18] 和基于梯度的 NAS [6]、[19]、[20]。在基于 EA 和基于 RL 的方法中,搜索过程需要大量架构候选者的验证准确性,这在计算上是昂贵的。例如,RL 方法 [1]、[2] 在四天内跨 500 个 GPU 训练和评估了 20000 多个神经网络。这些方法使用大量的计算资源,效率低下且负担不起。

为了消除这种缺陷,最近开发了基于梯度的 NAS 方法 [6]、[15]、[19]、[20],例如 DARTS [6] 和 GDAS [20]。他们构建了一个超级网络并通过为候选操作分配架构权重来放松架构表示。在 DARTS 中,搜索计算单元作为最终架构的构建块,每个单元表示为由 N 个节点的有序序列组成的有向无环图 (DAG)。

然后将具体的搜索空间放松为一个连续的搜索空间,以便可以通过梯度下降在其验证集性能方面优化架构。它实现了与基于 EA 的 [16] 和基于 RL 的 [1] 方法相当的性能,同时只需要几个 GPU 天的搜索成本。但是,DARTS 需要大量内存,因为它在每次迭代时都会优化整个超级网络。

为了减少搜索过程中的内存成本,提出了 GDAS [20]。它根据架构权重对子图进行采样,并且在每次训练迭代时只优化一个子图。

现有的基于梯度的 NAS 方法根据其架构权重选择候选操作。具有更高权重的候选操作更有可能被选择来导出目标架构。然而,我们发现架构权重并不能准确反映每个操作的重要性。为了证明这个问题,我们通过用所有其他候选操作替换最终搜索架构的单元格中的选定操作,获得具有不同候选操作的独立架构,充分训练它们直到收敛,并计算每个支架的准确性-单独模型。精度与其相应架构权重的关系如图 1 所示。

可以看出,架构权重最高的操作并没有达到最佳精度。图 2 说明了基于梯度的 NAS 方法的整个过程。

鉴于架构权重的限制,很自然地会问这样一个问题:如果我们应用更有效的指标来指导模型搜索,我们是否能够提高架构搜索性能?为此,我们提出一个简单的

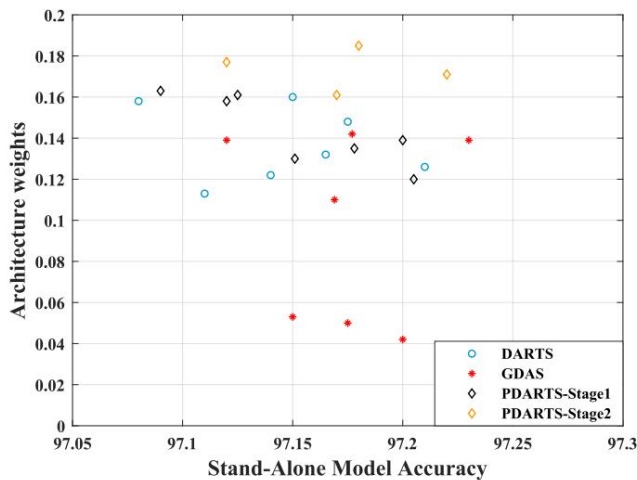


图 1. 独立模型精度与架构权重之间的相关性。为了获得具有不同候选操作的独立架构,我们分别用所有其他候选操作替换最终架构的一个单元中的选定操作,并对它们进行全面训练,直到收敛。PDARTS-Stage1 和 PDARTS-Stage2 代表 PDARTS [15] 搜索过程的不同阶段。

NAS 的有效解决方案,称为 Exploring Operation Importance for Effective NAS (EoiNAS)。我们的方法概要如下。

- 1)在[21]中已经证明,如果 A 在搜索过程中具有较少的训练时期和较高的验证准确性,则操作 A 优于操作 B。

基于这个标准,提出了一个新的指标来充分利用操作重要性并指导模型搜索。与使用训练时期和验证精度的差异来更新架构参数的现有方法(例如 MDENAS [21])相比,我们计算验证精度与训练时期的比率作为判断操作重要性的附加条件。

- 2)为了进一步提高搜索效率和准确性,我们提出了一种基于高阶马尔可夫链的策略来缩小搜索空间。现有方法在操作剪枝过程中仅考虑当前信息来判断操作重要性,导致判断不准确。相比之下,我们使用高阶马尔可夫链将操作的历史重要性与其当前指标相结合,以更准确地确定操作的重要性。

然后根据马尔可夫链计算的重要性修剪最差的操作。这个过程一直持续到只剩下一个操作;这个操作可以看作是推导出最终架构的最佳操作。由于这种策略,我们的超级网络表现出比最先进的 NAS 方法更好的收敛性。

EoiNAS 的有效性在两个任务上得到验证:图像分类和语义分割。对于图像分类任务,在 CIFAR-10 上执行搜索,并在 CIFAR-10/100 和 ImageNet 数据集上执行评估。EoiNAS 可以在几个 GPU 小时内完成一个搜索过程,并发现了一个强大的神经网络

测试误差为 2.50% 的网络。此外,在 CIFAR-10 上发现的网络可以成功转移到 ImageNet,分别达到 25.6% 和 8.3% 的前一和前五错误。我们还展示了我们的搜索算法对语义分割任务的有效性,即使没有 ImageNet 预训练也能实现高性能。

本文的其余部分组织如下。

在第二节中,我们回顾了最近有关 NAS 算法的工作。在第 III 节中,我们描述了我们提出的操作重要性指标 (OII) 和基于高阶马尔可夫链的搜索空间缩减 (SSS) 策略。第四节介绍了我们提出的方法在分类任务上的实验结果,并通过与以前的 NAS 算法进行比较来分析它们的有效性。第五节展示了我们的搜索算法在语义分割任务上的有效性。最后,我们的结论在第六节中提出。

## 二. 相关作品

随着深度学习的快速发展,为广泛的计算机视觉问题带来了显著的性能提升,例如图像分类和语义分割,其中大部分归功于手动设计的网络架构 [22]-[30]。最近,一个名为 NAS [1]-[5],[31],[32] 的新研究领域引起了越来越多的关注。NAS 的目标是找到设计神经架构的自动化方法,以取代传统的手工架构。根据探索大型架构空间的启发式方法,现有的 NAS 方法可以大致分为三类,即基于 EA 的方法 [7],[16],[17]、基于 RL 的方法 [1],[2],[18],[33] 和基于梯度的方法 [6],[19],[20]。

### A. 基于强化学习的 NAS

Zoph 和 Le [1] 以及 Zoph 等人提出了一种基于 RL 的方法。[2] 对于 NAS。他们使用循环网络作为控制器来生成指定用于给定任务的子神经网络的模型描述。由此产生的架构 (NASNet) 改进了当时现有的手工网络模型。

### B. 基于进化算法的 NAS

Real 等人提出了另一种搜索技术。[7] 其中使用进化 (遗传) 算法来找到为给定任务量身定制的神经架构。

进化的神经网络 (AmoebaNet) 的性能优于 NASNet。尽管这些工作在各种分类任务上取得了最先进的结果,但它们的主要缺点是它们需要大量的计算资源。

### C. 基于梯度的 NAS

与将架构搜索视为黑盒优化问题相比,基于梯度的 NAS 方法 [6],[15],[19]-[21],[34]-[38] 利用在训练中获得梯度优化神经结构的过程。飞镖 [6] 放松

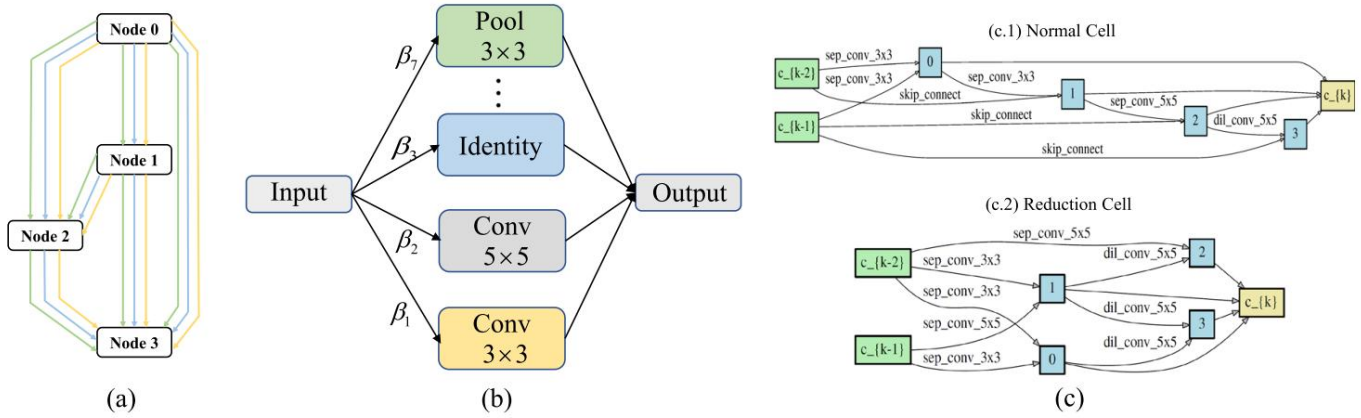


图 2. 基于梯度的 NAS 过程的图示。(a) 细胞结构由 DAG 表示。(b) 每两个节点之间的候选操作被分配架构权重以表示它们的重要性。(c) 在每条边上,选择权重最高的候选操作来导出目标架构 (selected architectures)。

搜索空间是连续的,这样就可以通过梯度下降优化架构的验证集性能。因此,基于梯度的方法成功地加速了架构搜索过程,并且只需要几个 GPU 日。由于 DARTS 在搜索过程中优化了整个超级网络,因此可能会遇到连续架构编码与派生的离散架构之间存在差异的问题。GDAS [20] 已被建议作为减轻这种差异的替代方法。GDAS 将搜索问题处理为从架构分布中抽样,其中分布本身以连续的方式学习。分布通过松弛软化的单热变量表示,这些变量乘以操作并使采样过程可微分。SNAS [19] 应用了类似的技术来将架构参数限制为 one-hot,以解决搜索和评估场景之间优化目标的不一致问题。

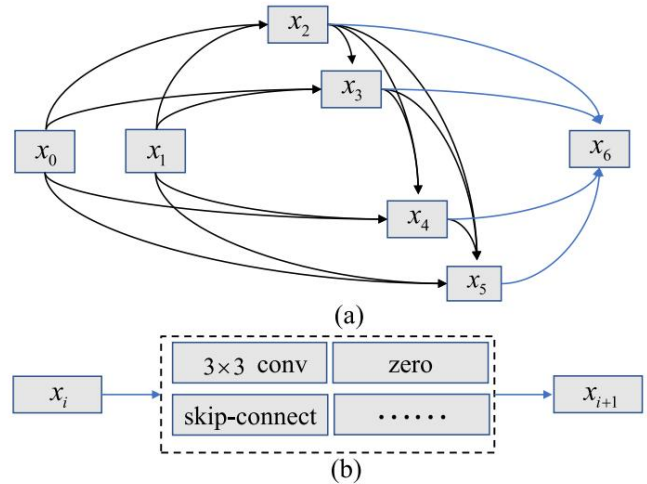


图 3. 搜索空间。(a) Cell 包含七个节点、两个输入节点、四个对输入节点和上层节点应用采样操作的中间节点,以及一个连接四个中间节点的输出的输出节点。(b) 两个节点之间的边表示在搜索空间 (候选操作集) 中根据离散概率分布  $P\{O = on\}$  采样的可能操作。

此外,MdeNAS [21] 提出了一种用于极其有效的 NAS 的多项分布学习方法,该方法将搜索空间视为联合多项分布。分布被优化以具有高性能期望。SGAS [37] 贪婪地选择和修剪可以使用启发式标准来确定操作,这些标准考虑了边缘重要性、选择确定性和选择稳定性。这种方法减轻了退化的搜索-评估相关性问题的影响,并反映了架构的真实排名。StacNAS [38] 指出相似运算符的相关性是影响排名问题的因素,然后通过运算符聚类向可区分 NAS 引入了一种新的层次搜索算法。

### 三.方法

我们首先在第 III-A 节中以一般形式描述我们的搜索空间和连续松弛,其中架构的计算过程表示为 DAG。然后,我们提出了一个新指标,以充分利用第 III-B 节中每个操作的重要性。最后,我们设计了一种基于高阶马尔可夫链的 SSS 策略,使超级网络在第 III-C 节中表现出快速收敛和高训练精度。

#### A. 搜索空间和持续放松

在这项工作中,我们利用 GDAS [20] 作为我们的基线框架工作。我们的目标是搜索最佳单元并构建包含  $L$  个单元的网络。Cell 定义为  $N$  个节点的 DAG,其中每个节点是一个网络层,即执行特定的数学功能。在这项工作中,  $N$  被设置为 7。如图 3 所示,一个单元格包含七个节点,包括两个输入节点,四个对输入节点和上层节点应用采样操作的中间节点,以及一个连接四个中间节点的输出。我们将操作空间表示为  $O$ ,其中每个元素代表一个候选操作  $o(\cdot)$ 。我们将节点  $i$  的特征图表示为  $x_i$ ;一个单元格中  $N$  个节点的特征图可以表示为  $\{x_0, x_1, \dots, x_{N-1}\}$ 。边  $f_{i,j}$  表示连接节点  $i$  和  $j$  的信息流,它由一组由架构权重  $\beta_{i,j}$  加权的操作组成,并且是,

因此,公式为

f\_{i,j}(x\_i) = \frac{\exp(\alpha\_{i,j} \beta\_{q,j}(x\_i))}{\sum\_{o \in O} \exp(\alpha\_{i,j} \beta\_{q,j}(x\_i))} \tag{1}

\beta\_{q,j} = \frac{\exp(\alpha\_{i,j})}{\sum\_{o \in O} \exp(\alpha\_{i,j})} \tag{2}

其中  $\alpha_{i,j}$  是  $|O|$  维可学习向量  $\alpha$  的第一个元素,  $j \in R|O|$  编码节点  $i$  和  $j$  之间函数的采样分布,如下所述。  $\beta_{q,j}$  和  $\beta_{i,j}$

直观地,一个熟知的  $\beta = \{\beta_{o,i,j}\}$  表示操作  $o$  对于将节点  $i$  的特征图转换为节点  $j$  的特征图的相对重要性。与 GDAS 类似,在节点  $i$  和  $j$  之间,我们根据离散概率分布  $P\{O = o(\cdot)\} = \beta_{o,i,j}$  从  $O$  中采样一个操作。在  $i,j$ ,

搜索,我们计算单元格中每个节点的特征图作为

x\_j = f\_{i,j}(x\_i) \tag{3}

其中  $f_{i,j}$  是从候选操作集  $O$  中采样得到的。

因为操作  $f_{i,j}$  是根据离散概率分布进行采样的,所以我们不能反向传播梯度来优化  $\alpha_{i,j}$ 。

为了允许反向传播,我们使用 Gumbel-Max 技巧 [27],[39] 和 softmax 函数 [40] 重新制定 (3) 和 (4),这提供了一种从离散概率分布中抽取样本的有效方法可微方式

x\_j = \sum\_{o=1}^{j-1} \beta\_{o,i,j} f\_{o,i,j}(x\_i) \tag{4}

st\_{ho,i,j} = \frac{\exp(\alpha\_{i,j} + \tau \log \beta\_{o,i,j})}{\sum\_{o=1}^{|O|} \exp(\alpha\_{i,j} + \tau \log \beta\_{o,i,j})} \tag{5}

这里,  $\tau$  是从  $\text{Gumbel}(0,1)$  中抽取的 iid 样本,  $f_{o,i,j}$  表示  $O$  中的第  $o$  个函数,  $h_o$  是  $O$  中的第  $o$  个元素,  $j$  是  $f_{i,j}$  的权重  $h_{i,j}$ ,  $w_{i,j}$  在参数  $\alpha$  和  $\beta$  之间,  $\text{Gumbel}$  温度参数  $\tau$  有控制等可微 softmax 分布。在我们的搜索过程中,  $\tau$  从 5.0 退火到 0.0。

我们使用与 [20] 中相同的候选函数集 ( $O$ ) 进行公平比较。我们的候选操作集  $O$  包含以下八个操作: 1) 身份; 2) 零; 3)  $3 \times 3$  可分离卷积; 4)  $3 \times 3$  扩张的可分离卷积; 5)  $5 \times 5$  可分离卷积; 6)  $5 \times 5$  扩张的可分离卷积; 7)  $3 \times 3$  平均池化; 8)  $3 \times 3$  最大池化。我们搜索了两种细胞,即正常细胞和还原细胞。在搜索正常单元格时,  $O$  中的每个操作的步幅为 1。对于 reduction cell,操作的步幅在输入节点上设置为两个,在其他节点上设置为一个。

为了获得最佳单元格,我们通过分别对训练集和验证集执行交替梯度下降来优化网络参数  $w$  和架构参数  $\alpha$ 。这个优化过程被定义为一个双层

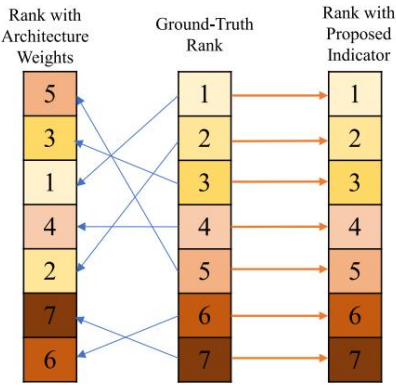


图 4. 我们通过颜色和数字可视化每个操作的相应重要性。架构权重排名和地面实况之间发生排名变化,我们提出的指标可以准确反映操作重要性。杨等。 [41] 已经表明,对于许多 NAS 方法,基于单元的搜索空间具有非常窄的精度范围,因此种子对架构排名具有相当大的影响。因此,我们使用不同的随机种子运行 EoiNAS 四次,以获得四种不同的模型,并根据其验证性能选择最佳模型。

优化问题

\min\_{\alpha} L\_{\text{valid}}(\alpha, w(\alpha)) \tag{6}

st w(\alpha) = \operatorname{argmin}\_w L\_{\text{train}}(\alpha, w) \tag{7}

其中  $L_{\text{valid}}$  是验证损失,  $L_{\text{train}}$  是训练损失。一旦我们发现了最佳的正常细胞和还原细胞,我们就会堆叠这些最佳细胞的副本以构成神经网络。

B. Operation Importance Indicator在现有的

基于梯度的 NAS 方法中,操作重要性直接由架构权重  $\beta$  排序,它应该表示候选操作相对于其他操作的相对重要性。当搜索过程完成后,选择最重要的操作,并根据体系结构权重的值修剪其他次要操作。

然而,架构权重不能准确反映每个操作的重要性,如第一节所述,图 4 说明了这个问题:架构权重排名与真实排名之间发生了排名变化。

1) 建议的指标: [21] 中已经证明,如果 A 在搜索过程中具有较少的训练时期和较高的验证精度,则操作 A 优于操作 B。每个操作的绝对验证准确率并不能真实反映操作的重要性,因为如果一个操作经过更多的迭代训练可以获得更好的性能。因此,训练时期需要结合验证精度来计算相对精度参数。具体来说,accuracy 参数定义为验证准确率与训练 epochs 的比值,作为判断操作重要性的附加条件。精度参数表示为

C\_{o,i,j} = \frac{\alpha\_{i,j}}{E\_{o,i,j}} \tag{8}



其中 $A_{o,i}$ 是每条边 $i$ 的精度权重,在epoch $t$ 我们还使用softmax对精度权重进行归一化,从而得到每个操作 $o$ 的精度权重 $\beta_{o,i}$ 。

$$\beta_{o,i} = \frac{\exp(C_{o,i})}{\sum_{o \in O} \exp(C_{o,i})} \quad (9)$$

准确度权重作为附加的操作重要性判断条件。基于架构权重 $\beta_o$ 和准确度权重 $\beta_{o,i}$ 我们得到更多的 $i, j, j$ , 有效指标 $l$ , 如式 (10) 所示, 可以充分挖掘每个操作的重要性

$$l_{o,i,j} = \beta_{o,i} + \lambda \beta_{o,j} \quad (10)$$

其中 $\lambda$ 是平衡这两个贡献的参数,在这项工作中设置为0.5。

2)分析比较:如图4所示,与DARTS [6]和GDAS [20]直接通过架构权重对操作重要性进行排序相比,我们提出的指标可以准确反映操作重要性,有助于选择最佳操作,从而达到最高精度。与根据三个选择标准对操作进行排序的SGAS [37]相比,我们提出的指标可以更直接地利用操作重要性并指导模型搜索。与StacNAS [38]相比,我们的搜索策略在对操作重要性进行排序时可以结合操作的历史重要性,因此搜索结果是稳健的。

然后我们分析了我们的搜索方法和MdeNAS [21]之间的区别。在MdeNAS的学习过程中,通过训练时期和准确率的差异来更新架构参数,从而将错误操作的概率转移到更好的操作。在我们的工作中,架构参数是通过随机梯度下降更新的;将准确率与训练epochs的比值确定的准确率参数作为附加的运算重要性判断条件,更好地反映运算重要性。

基于这个新的指标和高阶马尔可夫链模型,在搜索过程中提出了SSS策略,以进一步提高搜索效率和准确性,我们将在下面讨论。

### C. 基于高阶马尔可夫链的搜索空间瘦身策略

为了进一步提高搜索效率和准确性,我们提出了一种基于高阶马尔可夫链的策略来缩小搜索空间。现有的操作剪枝算法使用架构参数来判断训练过程中候选操作的重要性[15]。然而,当超网尚未稳定收敛时,架构参数总是会波动;在这种情况下,架构参数无法准确反映操作重要性。

为此,我们提出了一种新的机制来更好地反映操作的重要性,以使操作剪枝更加鲁棒。更具体地说,我们使用高阶马尔可夫链来综合来自

之前的迭代。与以往在操作剪枝过程中仅考虑当前信息来判断操作重要性的方法[15]相比,我们基于高阶马尔可夫链的SSS策略可以将操作的历史重要性与其当前指标相结合,从而确定操作更准确和健壮性的重要性。

我们将 $k$ 个时期的训练表示为一个步骤。在每一步中,我们使用一个 $n$ 阶马尔可夫链模型结合历史信息来更准确地判断操作重要性。具体来说,我们为每两个节点之间的每个候选操作定义两个状态,即Live(L)和Die(D)。每个候选操作都有一个状态转移矩阵 $Q$ 来更新其对应的状态。确定的 $n$ 阶马尔可夫链模型在(11)中定义

$$X_{i+1} = \sum_{m=1}^n X_i - m Q_i - m \quad (11)$$

$$\text{st } \sigma_m \geq 0, \quad \sigma_m = 1 \quad (12)$$

其中 $X_i = (l_i, d_i)$ 为第 $i$ 次迭代时的运行状态, $l_i$ 和 $d_i$ 分别表示L和D的概率。 $q_{L,L}$ 和 $q_{D,L}$ 是对应的状态转移矩阵,其中 $q_{L,L}$ 和 $q_{D,L}$ 表示停留在同一状态的概率, $q_{L,D}$ 表示从L状态移动到D状态的概率, $q_{D,D}$ 表示从D状态转变为L状态。

我们设置初始操作状态 $X_0 = (l_0, d_0) = (0.5, 0.5)$ 和初始状态转移矩阵 $Q_0 = 0.5 \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}$ 。我们定义更新状态转移矩阵 $Q$ 的规则。随着重要性指标的增加,我们通过 $\epsilon_1$ 增加 $q_{L,L}$ 和减少 $q_{L,D}$ ,同时增加 $q_{D,L}$ 和减少 $q_{D,D}$ 。相反,当指标下降时,我们进行相反的更新。更新后D概率最高的操作被认为是最差的操作。在每一步,我们判断剩余操作中最差的一个并修剪它,这个过程重复直到收敛到最终操作。

我们的搜索算法如算法1所示。在前20个时期,我们仅对网络参数执行基于梯度下降的优化。它有助于获得能够平衡参数化操作(例如,卷积操作)和非参数化操作(例如,跳过连接操作)的体系结构权重。然后,我们通过具有一阶近似的随机梯度下降以交替方式联合优化架构参数 $\alpha$ 和网络参数 $w$ 。具体来说,以迭代的方式,我们通过在训练集上降低 $w$ 来优化网络参数,并通过在验证集上降低 $\alpha$ 来优化架构参数。记录训练时期的数量和验证的准确性,以计算每个操作的重要指标。那么操作状态 $X$ 是



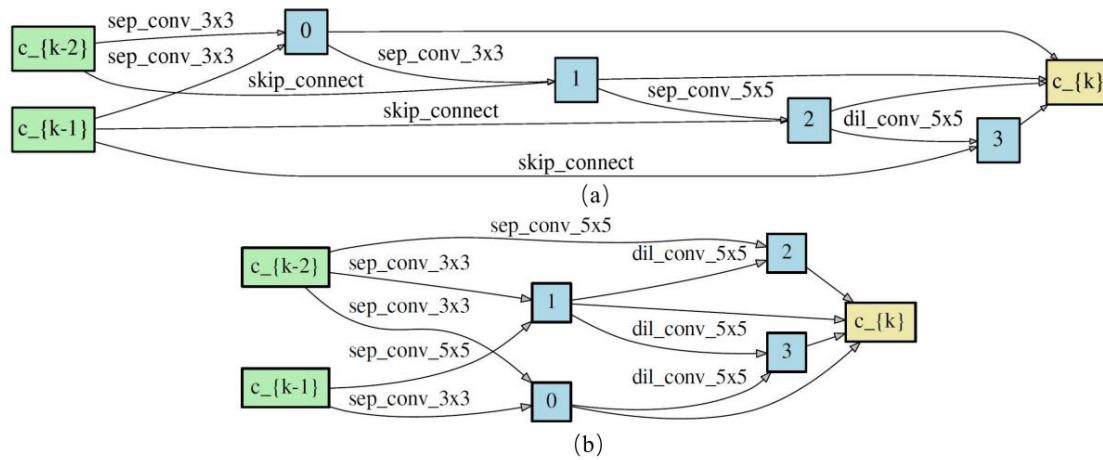


图 5. 我们的 EoiNAS 在 CIFAR-10 上发现的最佳细胞的详细结构。(a) 正常细胞。(b) 还原池。边缘操作的定义在第 III-A 节中描述。在正常的单元格中,每个操作的步幅是一个。对于 reduction cell,操作的步幅在输入节点上设置为两个,在其他节点上设置为一个。

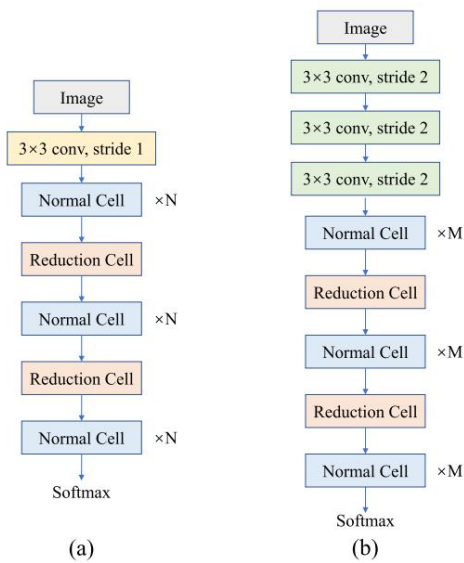


图 6. 设计策略 (a) CIFAR-10/100 架构和 (b) 基于已发现的正常单元和缩减单元的 ImageNet 架构。

重量 0.4 [1] 被应用。使用了权重衰减为 0.0003 和动量为 0.9 的标准 SGD 优化器。初始学习率为 0.025,按照余弦法则衰减为 0。

表一总结了评估结果和与最先进方法的比较。如表一所示,EoiNAS 在 CIFAR-10 和 CIFAR-100 上的测试错误率分别为 2.50% 和 17.3%。搜索成本为只有 0.6 个 GPU 天。为了获得相同的性能,AmoebaNet [7] 花费了四个数量级的计算资源 (0.6 GPU 天与 3150 GPU 天)。

我们的 EoiNAS 也大大优于 GDAS [20] 和 SNAS [19]。此外,我们将我们的方法与随机搜索 (RS) [6] 进行了比较,后者被认为是非常强大的基线。请注意,EoiNAS 搜索的模型准确率比 RS 高 0.7%。

#### D. ImageNet 上的结果

ImageNet 数据集用于测试在 CIFAR-10 上发现的架构的可迁移性。我们采用与 GDAS [20] 相同的网络配置。如图 6(b) 所示,在单个 NVIDIA 1080Ti GPU 上构建了一个由 14 个单元组成的网络并训练了 250 个批次大小为 128 的 epoch,使用 PyTorch [58] 实现耗时 12 天。通过使用 SGD 优化器优化网络参数,初始学习率为 0.1,动量为 0.9,权重衰减为  $3 \times 10^{-5}$ 。在训练期间应用了其他增强功能,包括标签平滑和辅助损失塔。

表二总结了评估结果和与最先进方法的比较。在 ImageNet 上,网络的前一错误和前五错误分别达到了 25.6% 和 8.3%。EoiNAS 发现的架构在分类精度和模型大小方面大大优于 GDAS 发现的架构。这证明了所发现架构从小型数据集到大型数据集的迁移能力。

#### E. 消融研究

此外,我们进行了一系列消融研究,验证了所提出的基于高阶马尔可夫链的 SSS 策略和 EoiNAS 设计中包含的 OII 的重要性和有效性。

1) 搜索空间瘦身策略:在表 III 中,我们展示了 CIFAR-10 的消融研究。SSS 指的是基于高阶马尔可夫链的 SSS 策略,OII 是提议的 OII。所有架构都经过 600 个 epoch 的训练。

如结果所示,由于基于高阶马尔可夫链的 SSS 策略,我们的超级网络表现出快速收敛和高训练精度。

2) 运营重要性指标:表 III 也展示了提议 OII 的有效性。所提出的指标可以更好地判断每个操作的重要性,达到更高的准确率。这些结果揭示了 OII 的必要性。此外,具有相同 OII 的操作同样

表一  
我们的 EoiNAS 的分类错误以及 CIFAR - 10 和 CIFAR - 100 的基准

Type	Architecture	GPUs	Times (days)	Params (million)	Test Error		Search Method
					C10(%)	C100(%)	
Human expert	ResNet + CutOut [22]	—	—	1.7	4.61	22.10	manual
	DenseNet [23]	—	—	25.6	3.46	17.18	manual
	SENet [42]	—	—	11.2	4.05	—	manual
Neural architecture search	MetaQNN [43]	10	8-10	11.2	6.92	27.14	RL
	NAS [1]	800	21-28	7.1	4.47	—	RL
	NAS + more filters [1]	800	21-28	37.4	3.65	—	RL
	NASNet-A [2]	450	3-4	3.3	3.41	—	RL
	NASNet-A + CutOut [2]	450	3-4	3.3	2.65	—	RL
	ENAS [18]	1	0.45	4.6	3.54	19.43	RL
	ENAS + CutOut [18]	1	0.45	4.6	2.89	—	RL
	Net Transformation [3]	5	2.0	19.7	5.70	—	RL
	SMASH [44]	1	1.5	16.0	4.03	—	RL
	AmoebaNet-A + CutOut [7]	450	7.0	3.2	3.34	18.93	evolution
	AmoebaNet-B + CutOut [7]	450	7.0	2.8	2.55	—	evolution
	Hierarchical NAS [16]	200	1.5	61.3	3.63	—	evolution
	CARS + CutOut [45]	1	0.4	3.0	2.86	—	evolution
	NSGANet [46]	1	8.0	3.3	3.85	—	evolution
	Progressive NAS [5]	100	1.5	3.2	3.63	19.53	SMBO
	DARTS (1st) + CutOut [6]	1	0.38	3.3	3.00	17.76	gradient-based
	DARTS (2nd) + CutOut [6]	1	1.0	3.4	2.82	17.54	gradient-based
	BayesNAS + cutout [47]	1	0.2	3.4	2.81	—	gradient-based
	SNAS + CutOut [19]	1	1.5	2.9	2.98	—	gradient-based
	GDAS [20]	1	1.0	3.4	3.87	19.68	gradient-based
	GDAS + CutOut [20]	1	1.0	3.4	2.93	18.38	gradient-based
	GDAS(FRC) + CutOut [20]	1	0.8	2.5	2.82	18.13	gradient-based
	GHN + CutOut [48]	1	0.84	5.7	2.84	—	gradient-based
	DATA + cutout [49]	1	1.0	3.4	2.59	—	gradient-based
	NAONet [8]	200	1.0	10.6	3.18	—	NAO
	MdeNAS + CutOut [21]	1	0.16	3.61	2.55	—	MDL
	Random Search + CutOut [6]	1	4.0	3.2	3.29	—	random
	EoiNAS	1	0.6	3.4	3.42	18.4	gradient-based
	EoiNAS + CutOut	1	0.6	3.4	2.50	17.3	gradient-based

表二  
与 IMAGENET 上最先进的网络进行比较。在 CIFAR - 10 上搜索所有 NAS 网络,然后直接传输到 IMAGENET

Type	Architecture	GPUs	Times (days)	Params (million)	MAdds (million)	Test Error (%)		Search Method
						Top-1	Top-5	
Human expert	Inception-v1 [27]	—	—	6.6	1448	30.2	10.1	manual
	MobileNet-V1 [52]	—	—	4.2	569	29.4	10.5	manual
	MobileNet-V2 [53]	—	—	3.4	300	28.0	9.0	manual
	ShuffleNet-V1 [54]	—	—	5.0	524	29.1	9.2	manual
	ShuffleNet-V2 [30]	—	—	5.0	524	26.3	—	manual
	CondenseNet [29]	—	—	4.8	529	26.2	—	manual
Neural architecture search	NASNet-A [2]	450	3-4	5.3	564	26.0	8.4	RL
	NASNet-B [2]	450	3-4	5.3	488	27.2	8.7	RL
	NASNet-C [2]	450	3-4	4.9	558	27.5	9.0	RL
	AmoebaNet-A [7]	450	7.0	5.1	555	25.5	8.0	evolution
	AmoebaNet-B [7]	450	7.0	5.3	555	26.0	8.5	evolution
	AmoebaNet-C [7]	450	7.0	6.4	570	24.3	7.6	evolution
	CARS [45]	1	0.4	4.4	510	26.3	8.4	evolution
	Progressive NAS [5]	100	1.5	5.1	588	25.8	8.1	SMBO
	MdeNAS [21]	1	0.16	6.1	596	25.5	7.9	MDL
	GHN [48]	1	0.84	6.1	569	27.0	8.7	gradient-based
	DARTS (2nd) [6]	1	1.0	4.9	595	26.9	9.0	gradient-based
	PARSEC [36]	1	1.0	5.6	548	26.0	8.4	gradient-based
	BayesNAS [47]	1	0.2	3.9	—	26.5	8.9	gradient-based
	SNAS [19]	1	1.5	4.3	522	27.3	9.2	gradient-based
	GDAS [20]	1	1.0	5.3	581	26.0	8.5	gradient-based
	GDAS(FRC) [20]	1	0.8	4.4	497	27.5	9.1	gradient-based
	SETN [55]	1	1.8	5.2	597	26.7	8.6	gradient-based
	EoiNAS	1	0.6	5.0	570	25.6	8.3	gradient-based

重要的。我们进行了相应的实验,发现如果两个候选操作的 OII 相同,我们可以通过修剪其中一个来获得相似的精度。

我们进一步显示了图 1 和 2 中搜索过程中的损失和准确度曲线。7和8.图7证明了SSS的有效性,而图8证明了OII的有效性。

如上所述,SSS 和 OII 可以大大加快



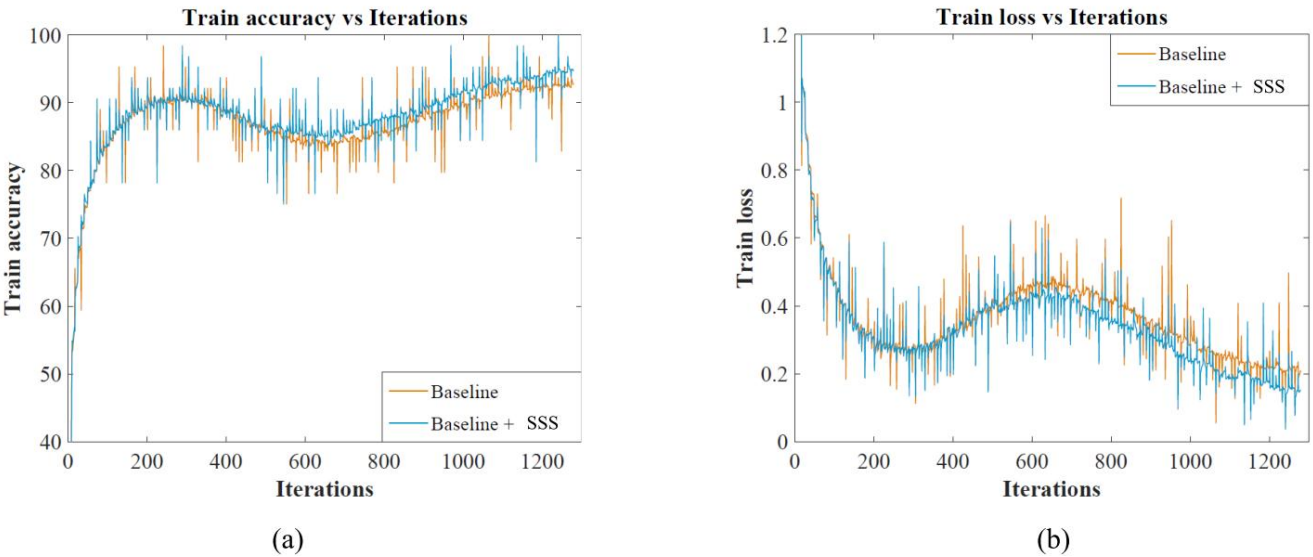


图 7. 使用和不使用 SSS 时的精度和损耗曲线。该图比较了 (a) 训练精度和 (b) 在搜索阶段使用和不使用 SSS 时的训练损失。

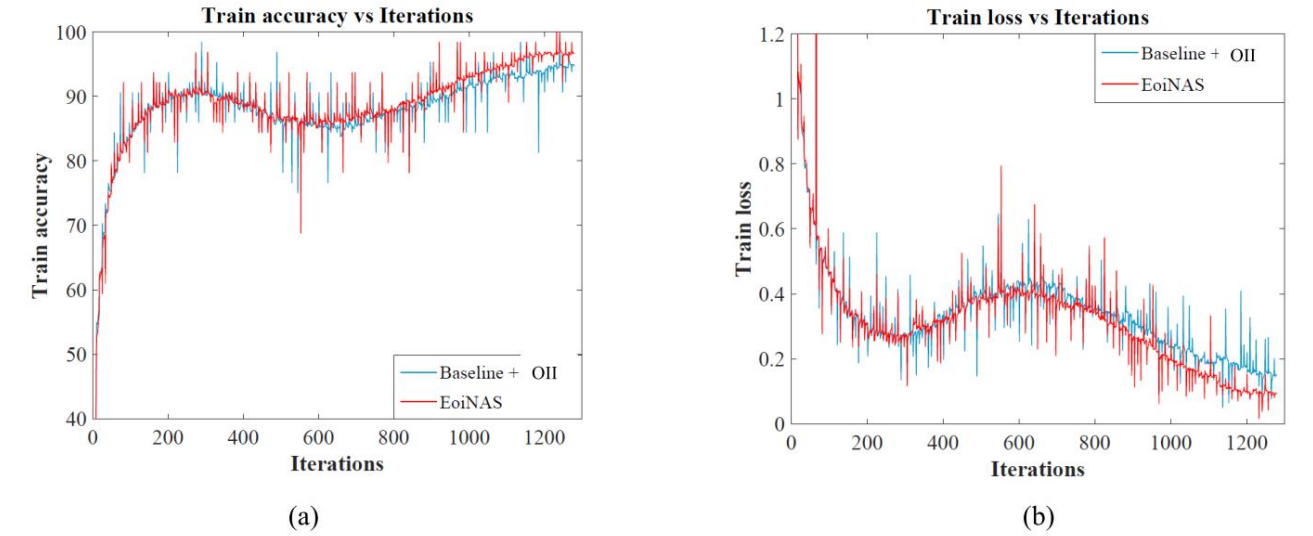


图 8. 使用和不使用 OII 时的精度和损耗曲线。该图比较了 (a) 训练精度和 (b) 在搜索阶段使用和不使用 OII 时的训练损失。

表三

CIFAR-10 的消融研究。SSS 是指基于高阶马尔可夫链的 SSS 策略；OII 是指拟议的 OII。所有架构都经过 600 个 EPOCHS 训练

Architecture	SSS	OII	Times (days)	Params (million)	Error (%)
EoiNAS without SSS and OII	×	×	1.0	3.4	2.93
EoiNAS without OII	✓	×	0.6	3.4	2.72
EoiNAS	✓	✓	0.6	3.4	2.50

在搜索阶段收敛,与基线相比获得更高的准确性。

F. Searched Architecture Analysis在

可微NAS方法中,架构权重不能准确反映每个操作的重要性,

正如第一节和第三节所讨论的,因为完全训练的独立模型的准确性与其相应的体系结构权重具有低相关性。所提出的 OII 可以更好地决定哪些操作应该保留在每条边上,哪些边应该作为每个节点的输入,尤其是对于 skip-connect 的选择。

跳过连接操作在单元结构中起着重要作用。正如 [61] 和 [62] 中研究的那样,包括合理数量和位置的跳跃连接将使梯度流更容易并优化深度神经网络更稳定。比较图中的搜索结果。如图 5 和 9 所示,EoiNAS 在 CIFAR-10 上发现的架构倾向于以分层方式保留跳跃连接操作,这可以促进梯度反向传播并使网络具有更好的收敛性。

此外,与无花果相比。5和9,我们可以看到EoiNAS鼓励cell中的连接级联更多

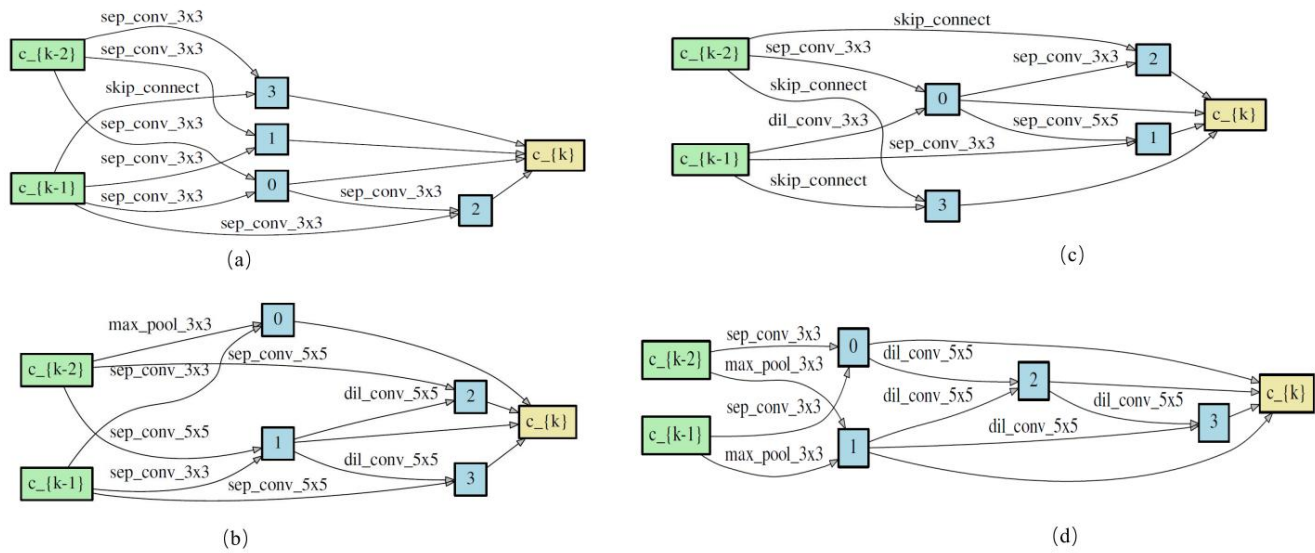


图 9. 通过基线和基线 + SSS 在 CIFAR-10 上发现的最佳细胞的详细结构。(a) 和 (b) 基线在 CIFAR-10 上发现的最佳细胞。(c) 和 (d) 基线 + SSS 发现的最佳细胞。(a) 和 (c) 正常细胞。(b) 和 (d) 还原池。

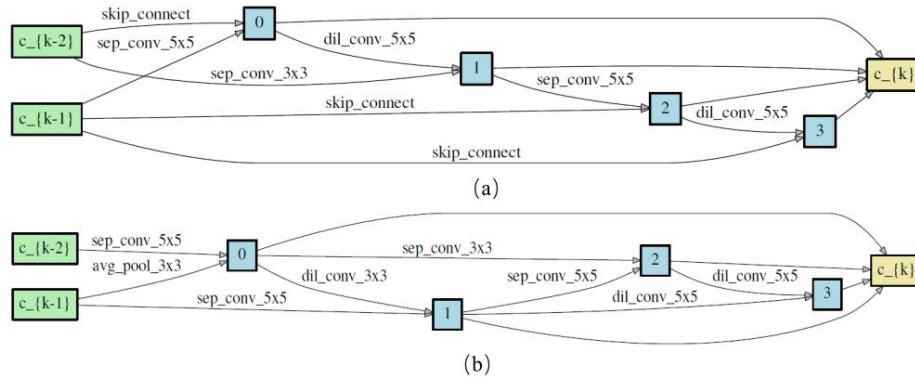


图 10. 我们的 EoiNAS 在 Cityscapes 上发现的最佳细胞的详细结构。(a) 正常细胞。(b) 还原池。边缘操作的定义在第 III-A 节中。在正常的单元格中,每个操作的步幅是一个。对于 reduction cell,操作的步幅在输入节点上设置为两个,在其他节点上设置为一个。

级别;换句话说,单元中有更多的层,使评估网络更深,并获得更好的分类性能。

最后,将 OII 与 SSS 策略相结合可以相互促进两者。该指标能够准确地表示操作的重要性并确定剩余和剪枝操作。同时,通过逐步修剪劣质操作,我们可以获得更准确的指标。

##### 五、语义分割任务实验

在本节中,我们进一步证明了 EoiNAS 在语义分割任务中的有效性。建筑搜索和评估是在 Cityscapes [61] 数据集上进行的。

##### A. 数据集

Cityscapes [61] 是一个流行的城市场景语义理解数据集。它包含从街道场景中收集的 5000 张图像的高质量像素级注释

50 个不同的城市。图像分辨率为  $1024 \times 2048$ 。

在分割任务中遵循标准协议 [61],使用 19 个语义标签进行评估。训练集、验证集和测试集分别包含 2975、500 和 1525 张图像。该数据集还提供了一组额外的 23 473 张粗略注释图像。与分类任务类似,训练集被平均分为两个子集:一个用于更新网络参数,另一个用于更新架构参数。

##### B. 实施细节

与分类任务类似,我们的实验包括两个阶段。首先,将 EoiNAS 应用于 Cityscapes 数据集以搜索构成架构的正常/还原单元。然后,对该架构进行再培训以进行评估。

1) 用于搜索的网络配置:按照 [12] 的方法在 Cityscapes 上搜索 CNN 的神经细胞。候选函数集  $\mathcal{O}$  具有与分类任务相同的八个不同的候选操作。

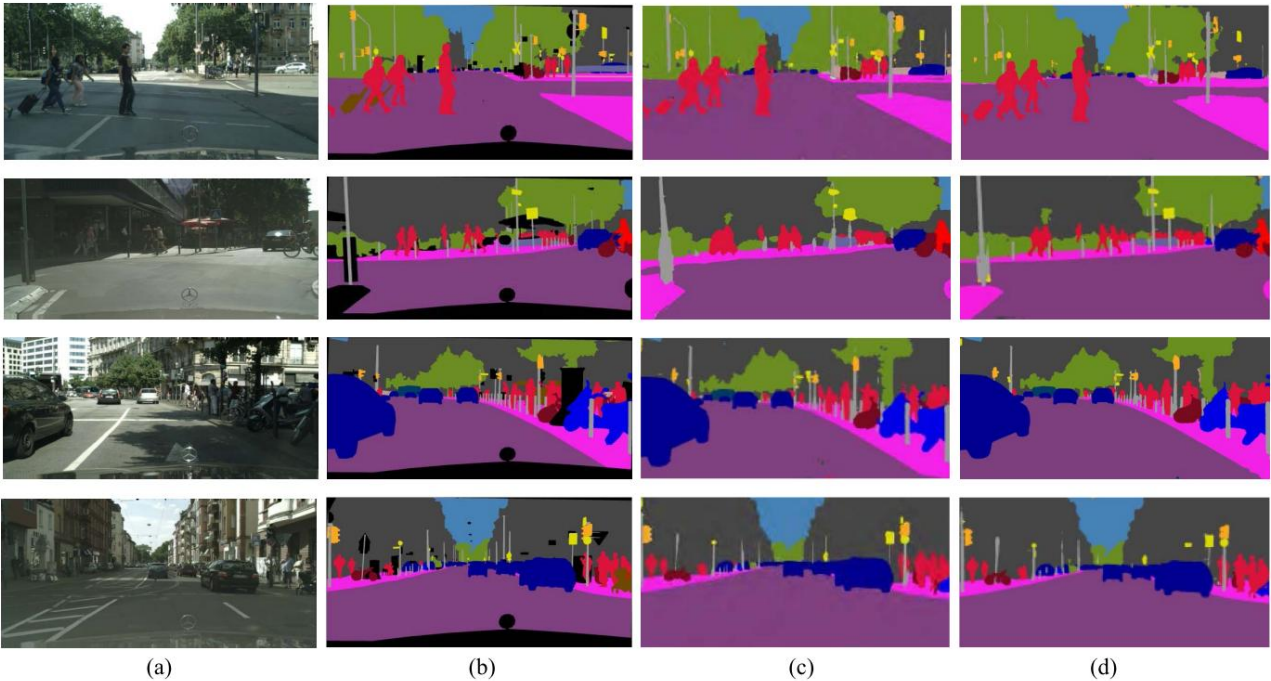


图 11. Cityscapes 数据集上的 EoiNAS 结果示例。(a) 图片。(b) 基本事实。(c) PSP 网络。(d) EoiNAS。

表IV

搜索阶段的详细网络结构,其中  
搜索正常和还原细胞

由我们的EoiNAS

Architecture	<i>s</i>	Feature map size	Output stride
Conv2d, 3 × 3	2	160 × 160	2
Conv2d, 3 × 3	1	160 × 160	2
Conv2d, 3 × 3	2	80 × 80	4
Normal cell × 3	1	80 × 80	4
Reduction cell × 1	2	40 × 40	8
Normal cell × 3	1	40 × 40	8
Reduction cell × 1	2	20 × 20	16
Normal cell × 2	1	20 × 20	16
ASPP	—	20 × 20	16

详细的网络结构如表四所示。三个卷积操作和十个单元构成了一个主干。位于网络总深度的 1/3 和 2/3 处的单元是缩减单元。空洞空间金字塔池化 (ASPP) [12] 模块被附加在最终单元的顶部,其输出被双线性上采样到原始分辨率以产生最终预测。默认情况下,我们总共对网络进行了 160 个 epoch 的训练。

在这里,我们阐明了为什么语义分割的网络结构不同于图像分类的网络结构。图像分类的任务侧重于整幅图像的语义聚合,这就需要一个联系特征图。因此,用于图像分类的骨干网络包含五个空间缩减,从而产生了原始图像 1/32 大小的小特征图。

相反,语义分割侧重于像素的语义聚合。然而,过多的空间缩减会导致像素空间信息的丢失,从而降低分割性能。为此,数

对于语义分割,应该减少空间缩减操作。在我们的工作中,最终特征图的空间分辨率设置为仅比输入图像分辨率小 16 倍,以平衡空间密度、语义和昂贵的计算。除了步幅为 2 的主干网络的第一和第三卷积层外,两个缩减单元还用于对特征图进行下采样。除还原细胞外,其余细胞均为未还原的正常细胞。

2)搜索参数设置:在学习网络参数w时,我们使用了动量为0.9的SGD优化器,余弦学习率从0.025衰减到0.001,权重衰减为0.0003。对于架构参数  $\alpha$ ,我们使用零初始化,这意味着对所有可能的操作都给予同等的关注。我们使用学习率为 0.0003、动量 (0.5; 0.999) 和权重衰减为 0.001 的 Adam 优化 [56]。为了控制 (5) 中 Gumbel softmax 的温度参数T,我们使用了指数衰减计划。T被初始化为 5,最后减少到 0。我们的 EoiNAS 在单个 NVIDIA 1080Ti GPU 上花费了大约 0.8 GPU 天来完成搜索过程。EoiNAS 搜索到的最佳小区如图 10 所示。

C. 城市景观的结果

为了在语义分割任务上评估我们搜索的架构,我们采用了编码器-解码器结构。具体来说,编码器的结构与搜索阶段的网络结构相同,如表IV所示,而解码器与 [46]中的相同,它通过利用低级特征恢复边缘信息下采样率为 4。我们首先利用 ImageNet 数据集对主干网络进行预训练



表五

CITYSCAPES数据集上的每类结果。我们未经 IMAGENET预训练训练的 EOINAS标有 “+”

Method	road	swalk	build.	wall	fence	pole	tlight	sign	veg.	terrain	sky	person	rider	car	truck	bus	train	mbike	bike	mIoU
ENet [62]	96.3	74.2	85	32.2	33.2	43.5	34.1	44	88.6	61.4	90.6	65.5	38.4	90.6	36.9	50.5	48.1	38.8	55.4	58.3
CRF-RNN [63]	96.3	73.9	88.2	47.6	41.3	35.2	49.5	59.7	90.6	66.1	93.5	70.4	34.7	90.1	39.2	57.5	55.4	43.9	54.6	62.5
FCN [64]	97.4	78.4	89.2	34.9	44.2	47.4	60.1	65.0	91.4	69.3	93.9	77.1	51.4	92.6	35.3	48.6	46.5	51.6	66.8	65.3
SiCNN+CRF [65]	96.3	76.8	88.8	40.0	45.4	50.1	63.3	69.6	90.6	67.1	92.2	77.6	55.9	90.1	39.2	51.3	44.4	54.4	66.1	66.3
DPN [66]	97.5	78.5	89.5	40.4	45.9	51.1	56.8	65.3	91.5	69.4	94.5	77.5	54.2	92.5	44.5	53.4	49.9	52.1	64.8	66.8
Dilation10 [67]	97.6	79.2	89.9	37.3	47.6	53.2	58.6	65.2	91.8	69.4	93.7	78.9	55.0	93.3	45.5	53.4	47.7	52.2	66.0	67.1
LRR [68]	97.7	79.9	90.7	44.4	48.6	58.6	68.2	72.0	92.5	69.3	94.7	81.6	60.0	94.0	43.6	56.8	47.2	54.8	69.7	69.7
DeepLabv2 [69]	97.9	81.3	90.3	48.8	47.4	49.6	57.9	67.3	91.9	69.4	94.2	79.8	59.8	93.7	56.5	67.5	57.5	57.7	68.8	70.4
Piecewise [70]	98.0	82.6	90.6	44.0	50.7	51.1	65.0	71.7	92.0	72.0	94.1	81.5	61.1	94.3	61.1	65.1	53.8	61.6	70.6	71.6
FRRN [71]	98.2	83.3	91.6	45.8	51.1	62.2	69.4	72.4	92.6	70.0	94.9	81.6	62.7	94.6	49.1	67.1	55.3	53.5	53.5	71.8
RefineNet [72]	98.3	83.7	91.8	47.2	52.3	62.4	69.8	73.1	92.9	70.2	94.9	82.5	64.8	94.8	64.2	73	58.4	59.6	64.5	73.6
TuSimple [73]	98.4	84.8	92.4	54.3	54.3	62.8	70.2	75.9	93.2	70.9	94.7	84.1	66.5	95.3	68.3	78	63.9	64.8	73.6	76.1
PSPNet [74]	98.6	86.2	92.9	50.8	58.8	64.0	75.6	79.0	93.4	72.3	95.4	86.5	71.3	95.9	68.2	79.5	73.8	69.5	77.2	78.4
DeepLabv3+ [75]	98.4	86.4	93.0	57.6	62.5	64.3	76.0	79.6	94.0	71.5	96.2	86.5	70.0	96.2	71.2	85.6	77.2	68.4	77.0	79.5
EoiNAS <sup>‡</sup>	97.9	81.5	91.4	50.5	52.7	59.4	69.6	72.7	92.5	70.1	95.0	81.3	60.1	94.3	51.2	67.7	61.3	59.5	71.2	72.6
EoiNAS	98.6	86.6	93.2	58.1	63.0	64.5	75.2	79.2	93.4	72.1	95.1	86.3	71.4	96.0	73.5	90.4	80.3	69.9	76.9	<b>80.2</b>

(三个卷积运算和十个单元格)。然后,整个编码器-解码器网络使用 ImageNet 预训练权重进行初始化,并在 Cityscapes 数据集上进行训练。我们的模型没有在粗略注释的图像上进行训练。

在训练期间,我们采用多项式学习率计划[75],[76],初始学习率为 0.007。裁剪尺寸为 769 × 769。由于 GPU 的限制,模型经过 90k 次迭代训练,批量大小为 4

记忆。

图 11 显示了 Cityscapes 数据集上 EoiNAS 结果的一些示例。分割结果列在表 V 中。我们的 EoiNAS 在 Cityscapes 数据集上实现了更高的 mIOU,即使没有 ImageNet 预训练。

情报,卷。33,2019,第 4780–4789 页。

[8] R. Luo,F. Tian,T. Qin,E. Chen 和 T.-Y. Liu,“神经架构优化”,Proc.进阶神经信息过程。系统,2018 年,第 7816–7827 页。

[9] Y. Chen,K. Zhu,L. Zhu,X. He,P. Ghamisi 和 J. Atli Benediktsson,“用于高光谱图像分类的卷积神经网络的自动设计”,IEEE Trans.地球科学。遥感,卷。57,没有。9,第 7048–7066 页,2019 年 9 月。

六.结论

在这篇文章中,我们介绍了 EoiNAS,一种简单而高效的卷积网络架构搜索算法,其中提出了一个新的指标来充分利用操作重要性来指导模型搜索。为了进一步提高搜索效率和准确性,提出了一种基于高阶马尔可夫链的搜索空间缩减策略。

为了验证有效性,我们将 EoiNAS 应用于两个视觉任务:图像分类和语义分割。EoiNAS 大大降低了计算消耗,同时实现了出色的模型精度,这优于人工设计的网络和其他最先进的 NAS 方法来完成这两项视觉任务。

参考

[1] B. Zoph 和 QV Le,“使用强化学习进行神经架构搜索”,2016 年,arXiv:1611.01578。[在线的]。可用: <http://arxiv.org/abs/1611.01578> [2] B. Zoph,V. Vasudevan,J. Shlens 和 QV Le,“学习可扩展图像识别的可迁移架构”,Proc. IEEE/CVF 会议。电脑。可见。模式识别,2018 年 6 月,第 8697–8710 页。

[3] H. Cai,T. Chen,W. Zhang,Y. Yu 和 J. Wang,“通过网络转换进行高效架构搜索”,Proc.第 32 届 AAAI 会议。神器。Intell.,2018 年,第 1–8 页。

[4] T. Elsken,J. Hendrik Metzen 和 F. Hutter,“神经架构搜索:一项调查”,2018 年,arXiv:1808.05377。[在线的]。可用: <http://arxiv.org/abs/1808.05377> [5] C. Liu 等人,“Progressive neural architecture search”,Proc.欧元。会议。电脑。可见。(ECCV),2018 年,第 19–34 页。

[6] H. Liu,K. Simonyan 和 Y. Yang,“DARTS:可微架构搜索”,2018 年,arXiv:1806.09055。[在线的]。可用: <http://arxiv.org/abs/1806.09055> [7] E. Real,A. Aggarwal,Y. Huang 和 QV Le,“图像分类器架构搜索的正则化演化”,Proc. AAAI 会议。神器。

[8] R. Luo,F. Tian,T. Qin,E. Chen 和 T.-Y. Liu,“神经架构优化”,Proc.进阶神经信息过程。系统,2018 年,第 7816–7827 页。

[9] Y. Chen,K. Zhu,L. Zhu,X. He,P. Ghamisi 和 J. Atli Benediktsson,“用于高光谱图像分类的卷积神经网络的自动设计”,IEEE Trans.地球科学。遥感,卷。57,没有。9,第 7048–7066 页,2019 年 9 月。

[10] X. Chu,B. Zhang,H. Ma,R. Xu 和 Q. Li,“使用神经架构搜索实现快速、准确和轻量级的超分辨率”,2019 年,arXiv:1901.07261。[在线的]。可用: <http://arxiv.org/abs/1901.07261> [11] L.-C. Chen 等人,“为密集图像预测寻找有效的多尺度架构”,Proc.进阶神经信息过程。系统,2018 年,第 8699–8710 页。

[12] C. Liu 等人,“Auto-DeepLab:用于语义图像分割的分层神经架构搜索”,Proc. IEEE/CVF 会议。电脑。可见。模式识别。(CVPR),2019 年 6 月,第 82–92 页。

[13] Y. Zhang,Z. Qiu,J. Liu,T. Yao,D. Liu 和 T. Mei,“语义分割的可定制架构搜索”,Proc. IEEE/CVF 会议。电脑。可见。模式识别。(CVPR),2019 年 6 月,p. 11。

[14] G. Ghiasi,T.-Y. Lin 和 QV Le,“NAS-FPN:学习用于对象检测的可扩展特征金字塔架构”,Proc. IEEE/CVF 会议。电脑。可见。模式识别。(CVPR),2019 年 6 月,第 7036–7045 页。

[15] X. Chen,L. Xie,J. Wu 和 Q. Tian,“Progressive differentiable architecture search: Bridging the depth gap between search and evaluation”,2019,arXiv:1904.12760。[在线的]。可用: <http://arxiv.org/abs/1904.12760>

[16] H. Liu,K. Simonyan,O. Vinyals,C. Fernando 和 K. Kavukcuoglu,“高效架构搜索的层次表示”,2017 年,arXiv:1711.00436。[在线的]。可用: <http://arxiv.org/abs/1711.00436> [17] L. Xie 和 A. Yuille,“Genetic CNN”,Proc. IEEE 译。会议。电脑。可见。(ICCV),2017 年 10 月,第 1379–1388 页。

[18] H. Pham,M.Y. Guan,B. Zoph,QV Le 和 J. Dean,“通过参数共享进行高效的神经架构搜索”,2018 年,arXiv:1802.03268。[在线的]。可用: <http://arxiv.org/abs/1802.03268> [19] S. Xie,H. Zheng,C. Liu 和 L. Lin,“SNAS:随机神经架构搜索”,2018 年,arXiv:1812.09926。[在线的]。可用: <http://arxiv.org/abs/1812.09926>

[20] X. Dong 和 Y. Yang,“在四个 GPU 小时内搜索强大的神经架构”,Proc. IEEE/CVF 会议。电脑。可见。模式识别。(CVPR),2019 年 6 月,第 1761–1770 页。

[21] X. Zheng,R. Ji,L. Tang,B. Zhang,J. Liu 和 Q. Tian,“用于有效神经架构搜索的多项分布学习”,2019 年,arXiv:1905.07529。[在线的]。可用: <http://arxiv.org/abs/1905.07529> [22] K. He,X. Zhang,S. Ren 和 J. Sun,“用于图像识别的深度残差学习”,Proc. IEEE 会议。电脑。可见。模式识别。(CVPR),2016 年 6 月,第 770–778 页。



[23] G. Huang,Z. Liu,L. Van Der Maaten 和 KQ Weinberger,“密集连接的卷积网络”, Proc. IEEE 会议.电脑.可见.模式识别. (CVPR), 2017 年 7 月,第 4700–4708 页。

[24] A. Krizhevsky,J. Sutskever 和 GE Hinton,“使用深度卷积神经网络进行 ImageNet 分类”, Proc.进阶神经网络信息过程.系统, 2012 年,第 1097–1105 页。

[25] K. Simonyan 和 A. Zisserman,“用于大规模图像识别的超深卷积网络”,2014 年, arXiv:1409.1556. [在线的]. 可用:http://arxiv.org/abs/1409.1556

[26] C. Szegedy等人, “深入卷积”, Proc. IEEE 会议.电脑.可见.模式识别. (CVPR), 2015 年 6 月,第 1-9 页。

[27] EJ Gumbel,极值统计理论和一些实际应用:一系列讲座,卷. 33. 华盛顿特区,美国:US 政府印刷局,1948 年。

[28] S.-H.高,M.-M. Cheng, K. Zhao,X.-Y.张,M.-H. Yang 和 P. Torr,“Res2Net:一种新的多尺度骨干架构”,2019 年, arXiv:1904.01169. [在线的].可用:http://arxiv.org/abs/1904.01169 [29] G. Huang,S. Liu,LVD Maaten 和 KQ Weinberger,“Con denseNet:使用学习组卷积的高效 DenseNet”, Proc. IEEE/CVF 会议.电脑.可见.模式识别, 2018 年 6 月,第 2752–2761 页。

[30] N. Ma, X. Zhang, H.-T. Zheng 和 J. Sun,“ShuffleNet V2:高效 CNN 架构设计实用指南”, Proc.欧元.会议.电脑.可见. (ECCV), 2018 年,第 116-131 页。

[31] D. Stamoulis等人, “单路径 NAS:在不到 4 小时内设计硬件高效的 ConvNet”,2019 年, arXiv:1904.02877. [在线的]. 可用:http://arxiv.org/abs/1904.02877

[32] Z. Guo等人, “具有均匀采样的单路经一次性神经网络搜索”,2019 年, arXiv:1904.00420. [在线的].可用:http://arxiv.org/abs/1904.00420

[33] M. Tan等人, “MnasNet:面向移动设备的平台感知神经架构搜索”, Proc. IEEE/CVF 会议.电脑.可见.模式识别. (CVPR), 2019 年 6 月,第 2820–2828 页。

[34] B. Wu等人, “FBNet:通过可微神经架构搜索进行硬件感知的高效 ConvNet 设计”, Proc. IEEE/CVF 会议.电脑.可见.模式识别. (CVPR), 2019 年 6 月,p. 10.

[35] H. Cai,L. Zhu 和 S. Han,“ProxylessNAS:目标任务和硬件上的直接神经架构搜索”,2018 年, arXiv:1812.00332. [在线的]. 可用:http://arxiv.org/abs/1812.00332

[36] F. Paolo Casale,J. Gordon 和 N. Fusi,“概率神经架构搜索”,2019 年, arXiv:1902.05116. [在线的].可用:http://arxiv.org/abs/1902.05116

[37] G. Li,G. Qian,JC Delgadillo,M. Muller,A. Thabet 和 B. Ghanem,“SGAS:顺序贪婪架构搜索”, Proc. IEEE/CVF 会议.电脑.可见.模式识别. (CVPR), 2020 年 6 月,第 1620–1630 页。

[38] G. Li等人, “通过算子聚类进行分层神经架构搜索”,2019 年, arXiv:1909.11926. [在线的]. 可用:http://arxiv.org/abs/1909.11926 [39] CJ Maddison,D. Tarlow 和 T. Minka,“A\* 采样”,在Proc.进阶神经网络信息过程.系统, 2014 年,第 1-10 页。

[40] E. Jang,S. Gu 和 B. Poole,“使用 gumbel-softmax 进行分类重新参数化”,2016 年, arXiv:1611.01144. [在线的].可用:http://arxiv.org/abs/1611.01144 [41] A. Yang,PM Esperanca 和 FM Carlucci,“NAS 评估非常困难”,在Proc.诠释.会议.学习.代表, 2020 年,第 1-13 页。

[42] J. Hu,L. Shen 和 G. Sun,“Squeeze-and-excitation networks”,载于Proc. IEEE/CVF 会议.电脑.可见.模式识别, 2018 年 6 月,第 7132–7141 页。

[43] B. Baker,O. Gupta,N. Naik 和 R. Raskar,“使用强化学习设计神经网络架构”,2016 年, arXiv:1611.02167. [在线的].可用:http://arxiv.org/abs/1611.02167 [44] A. Brock、T. Lim,JM Ritchie 和 N. Weston,“SMASH:通过 HyperNetworks 进行一次性模型架构搜索”, 2017 年, arXiv:1708.05344. [在线的].可用:http://arxiv.org/abs/1708.05344 [45] Z. Yang等人,“CARS:高效神经架构搜索的持续进化”,2019 年, arXiv:1909.04977. [在线的].可用:http://arxiv.org/abs/1909.04977

[46] Z. Lu等人, “NSGA-net:使用多目标遗传算法进行神经结构搜索”,2018 年, arXiv:1810.03522. [在线的].可用:http://arxiv.org/abs/1810.03522

[47] H. Zhou,M. Yang,J. Wang 和 W. Pan,“BayesNAS:一种用于神经架构搜索的贝叶斯方法”, 2019 年, arXiv:1905.04919. [在线的].可用:http://arxiv.org/abs/1905.04919

[48] C. Zhang,M. Ren 和 R. Urtasun,“用于神经架构搜索的图形超网络”,2018 年, arXiv:1810.05749. [在线的].可用:http://arxiv.org/abs/1810.05749

[49] J. Chang等人, “数据:可微架构近似”, Proc.进阶神经网络信息过程.系统, 2019 年,第 874–884 页。

[50] A. Krizhevsky等人, “从微小图像中学习多层特征”,计算机系.科学.大学. Toronto, Toronto, ON, 加拿大, Tech.众议院 2009 年 7 号。

[51] O. Russakovsky等人, “ImageNet 大规模视觉识别挑战”, Int. J. 计算.可见,卷. 115, 没有. 3,第 211-252 页,2015 年 12 月。

[52] AG Howard等人, “MobileNets:用于移动视觉应用的高效卷积神经网络”,2017 年, arXiv:1704.04861. [在线的].可用:http://arxiv.org/abs/1704.04861 [53] M. Sandler、A. Howard,M. Zhu,A. Zhmoginov 和 L.-C. Chen,“MobileNetV2:反向残差和线性瓶颈”, Proc. IEEE/CVF 会议.电脑.可见.模式识别, 2018 年 6 月,第 4510–4520 页。

[54] X. Zhang,X. Zhou,M. Lin 和 J. Sun,“ShuffleNet:一种用于移动设备的极其高效的卷积神经网络”, Proc. IEEE/CVF 会议.电脑.可见.模式识别, 2018 年 6 月,第 6848–6856 页。

[55] X. Dong 和 Y. Yang,“通过自评估模板网络进行一次性神经架构搜索”, Proc. IEEE/CVF 诠释.会议.电脑.可见. (ICCV), 2019 年 10 月,第 3681–3690 页。

[56] DP Kingma 和 J. Ba,“亚当:一种随机优化方法”,2014 年,[在线]. http://arxiv.org/abs/1412.6980 [57] T. DeVries和G.W. Taylor,“EfficientNet:在数据效率和模型精度之间取得平衡”,arXiv:1808.08010,2018

[58] A. Paszke等人, “Pytorch 中的自动微分”, Proc.进阶神经网络信息过程.系统.研讨会,美国加利福尼亚州长滩,2017 年 12 月,第 1-4 页。

[59] K. He,X. Zhang,S. Ren 和 J. Sun,“深度残差网络中的身份映射”, Proc.欧元.会议.电脑.可见. Cham,瑞士:Springer,2016 年,第 630–645 页。

[60] KA Sankaraman,S. De,Z. Xu,W. Ronny Huang 和 T. Goldstein,“神经网络过度参数化对梯度混淆和随机梯度下降的影响”,2019 年, arXiv:1904.06963. [在线的]. 可用:http://arxiv.org/abs/1904.06963

[61] M. Cordts等人, “用于语义城市市场景理解的城市景观数据集”, Proc. IEEE 会议.电脑.可见.模式识别. (CVPR), 2016 年 6 月,第 3213–3223 页。

[62] A. Paszke,A. Chaurasia,S. Kim 和 E. Culurciello,“ENet:一种用于实时语义分割的深度学习神经网络架构”,2016 年, arXiv:1606.02147. [在线的].可用:http://arxiv.org/abs/1606.02147 [63] S. Zheng等人, “作为循环神经网络的条件随机场”,在Proc. IEEE 诠释.会议.电脑.可见. (ICCV), 2015 年 12 月,第 1529–1537 页。

[64] J. Long,E. Shelhamer 和 T. Darrell,“用于语义分割的全卷积网络”, Proc. IEEE 会议.电脑.可见.模式识别. (CVPR), 2015 年 6 月,第 3431–3440 页. c.J. Krapac 和 S. egvi’c,“语义分割的卷积尺度不变性”, Proc.德国会议.模式识别. Cham,瑞士:Springer,2016

[65] I. Kre o,D. Cau evi’ 年,第 64-75 页。

[66] Z. Liu, X. Li, P. Luo, C.-C. Loy 和 X. Tang,“通过深度解析网络进行语义图像分割”, Proc. IEEE 诠释.会议.电脑.可见. (ICCV), 2015 年 12 月,第 1377–1385 页。

[67] F. Yu 和 V. Koltun,“通过扩张卷积进行多尺度上下文聚合”,2015 年, arXiv:1511.07122. [在线的]. 可用:http://arxiv.org/abs/1511.07122

[68] G. Ghiasi 和 CC Fowlkes,“语义分割的拉普拉斯金字塔重建和细化”, Proc.欧元.会议.电脑.可见. Cham,瑞士:Springer,2016 年,第 519–534 页。

[69] L.-C. Chen,G. Papandreou,I. Kokkinos,K. Murphy 和 AL Yuille,“DeepLab:使用深度卷积网络.空洞卷积和完全连接的 CRF 进行语义图像分割”, IEEE Trans.模式识别.马赫.情报, 卷. 40,没有. 4,第 834–848 页,2018 年 4 月。

[70] G. Lin,C. Shen,A. van den Hengel 和 I. Reid,“用于语义分割的深度结构化模型的高效分段训练”, Proc. IEEE 会议.电脑.可见.模式识别. (CVPR), 2016 年 6 月,第 3194–3203 页。

[71] T. Pohlen,A. Hermans,M. Mathias 和 B. Leibe,“用于街景语义分割的全分辨率残差网络”, 载于Proc. IEEE 会议.电脑.可见.模式识别. (CVPR), 2017 年 7 月,第 4151–4160 页。

[72] G. Lin,A. Milan,C. Shen 和 I. Reid,“RefineNet:用于高分辨率语义分割的多路径细化网络”, Proc. IEEE 会议.电脑.可见.模式识别. (CVPR), 2017 年 7 月,第 1925–1934 页。

[73] P. Wang等人, “理解语义分割的卷积”, Proc. IEEE 冬季会议.申请电脑.可见. (WACV), 2018 年 3 月,第 1451–1460 页。

[74] H. Zhao,J. Shi,X. Qi,X. Wang 和 J. Jia,“金字塔场景解析网络”, Proc. IEEE 会议.电脑.可见.模式识别. (CVPR), 2017 年 7 月,第 2881–2890 页。

[75] L.-C. Chen,Y. Zhu,G. Papandreou,F. Schroff 和 H. Adam,“用于语义图像分割的具有空洞可分离卷积的编码器解码器”, Proc.欧元.会议.电脑.可见. (ECCV), 2018 年,第 801–818 页。

[76] L.-C. Chen,G. Papandreou,F. Schroff 和 H. Adam,“重新思考用于语义图像分割的空洞卷积”,2017 年, arXiv:1706.05587。 [在线的]。可用: <http://arxiv.org/abs/1706.05587>



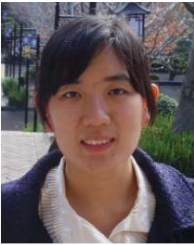
Xukai Xie获得工程学士学位。2018 年于中国天津河北工业大学获得学士学位。他目前正在攻读工程硕士学位。天津大学电气与信息工程学院, 天津。

他的研究兴趣是神经结构  
搜索和模型压缩。



Sun-Yuan Kung (IEEE 终身院士)目前是美国新泽西州普林斯顿市普林斯顿大学电气工程系的教授。

他撰写或合着了 500 多篇技术出版物和大量教科书,包括VLSI 阵列处理器 (Prentice-Hall,1988 年)、数字神经网络 (Prentice-Hall, 1993 年)、主成分神经网络 (Wiley,1996 年)、生物识别认证:一种机器学习方法 (Prentice-Hall,2004 年),以及内核方法和机器学习 (剑桥大学出版社,2014 年)。他目前的研究兴趣包括机器学习、数据挖掘和隐私、统计估计、系统识别、无线通信、VLSI 阵列处理器、信号处理和多媒体信息处理。



Yuan Zhou (高级会员,IEEE)获得工学学士、工学硕士和博士学位。分别于 2006 年、2008 年和 2011 年在中国天津天津大学电子工程和通信工程专业获得学位。

2011年起在天津大学电子信息工程学院任教,现任副教授。2013-2014年在澳大利亚昆士兰州图文巴南昆士兰大学机电工程学院做访问学者。2016-2017年在美国普林斯顿大学电气工程系做访问学者。她目前的研究兴趣包括计算机视觉和图像/视频通信。

Kung 教授是 IEEE 信号处理协会多个技术委员会的创始成员,并分别于 1984 年和 1991 年担任 IEEE TRANSACTIONS ON SIGNAL PROCESSING的超大规模集成电路领域的第一位副主编和神经网络的第一位副主编。1989年至1991年担任IEEE信号处理学会理事会成员,1992年因在信号处理的并行处理和神经网络算法方面的贡献获得IEEE信号处理学会技术成就奖,1996 年因在主成分神经网络方面的发表而获得 IEEE 信号处理协会最佳论文奖,2000 年获得 IEEE 第三千禧年奖章。1994 年成为 IEEE 信号处理协会特聘讲师。自 1990 年以来,他一直是 Journal of VLSI Signal Processing Systems主编。