Advances in Applied Mathematics 应用数学进展, 2023, 12(8), 3587-3597
Published Online August 2023 in Hans. https://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2023.128357

Hans汉斯

基于神经网络架构搜索的语义分割方法

朱　烜，马中华*

天津职业技术师范大学理学院，天津

收稿日期：2023年7月18日；录用日期：2023年8月8日；发布日期：2023年8月16日

摘　要

神经网络架构搜索旨在使用搜索策略在给定的搜索空间上让算法自动搜索出网络结构模型以减少人工设计网络的任务量，拓展神经网络架构搜索在语义分割领域的应用对自动化深度学习领域的研究有重要意义。通过设计U型搜索空间，将可微分神经网络架构搜索策略应用于语义分割模型。实验结果显示，在The Oxford-IIIT Pet数据集搜索得到的网络与基准网络UNet相比，搜索出的网络模型mIOU提高了14.1%，分割的效果更加显著，轮廓边界更加清晰。将搜索出来的网络迁移到Camvid数据集上进行测试，比基准网络实验精度提升了20.5%。研究表明，神经网络架构搜索与语义分割的结合在自动化深度学习领域的研究中具有重要意义，能够使语义分割模型获得更优秀的性能。

关键词

神经网络架构搜索，语义分割，自动化深度学习

Semantic Segmentation Method Based on Neural Architecture Search

Xuan Zhu, Zhonghua Ma*

College of Science, Tianjin University of Technology and Education, Tianjin

Received: Jul. 18th, 2023; accepted: Aug. 8th, 2023; published: Aug. 16th, 2023

Abstract

The objective of Neural Architecture Search (NAS) is to use a search strategy to automatically find network structure models within a given search space, thereby reducing the task load of manually designing networks. Expanding the application of NAS in the field of semantic segmentation bears significant importance for research in automated deep learning. A U-shaped search space was designed, and a differentiable NAS strategy was applied to a semantic segmentation model. Experimental results showed that the network found on The Oxford-IIIT Pet dataset outperformed the

*通讯作者。

文章引用: 朱烜, 马中华. 基于神经网络架构搜索的语义分割方法[J]. 应用数学进展, 2023, 12(8): 3587-3597.
DOI: 10.12677/aam.2023.128357

benchmark UNet network model, with a Mean Intersection over Union (mIOU) increase of 14.1%, and produced more prominent segmentation results with clearer contour boundaries. When the discovered network was transferred to the Camvid dataset for testing, it surpassed the benchmark network experimental accuracy by 20.5%. This study demonstrated that the integration of NAS and semantic segmentation holds significant importance in the field of automated deep learning research. This approach enables semantic segmentation models to achieve superior performance.

Keywords

Neural Architecture Search, Semantic Segmentation, Automated Deep Learning

1. 引言

近年来，深度学习领域的不断发展带动了神经网络模型的研究，出现诸如 VGG [1]、ResNet [2]、UNet [3]以及 DeepLab [4]等优秀的传统网络模型，前两者用于图像分类，后两者在语义分割领域大放异彩。但是设计这些出色并且高效的神经网络模型需要依赖于专家经验和知识，并且需要花费大量的时间和精力，因此如何针对特定任务自动寻找一个神经网络模型越来越受研究学者的注重。这一课题在深度学习中被称为神经网络架构搜索(Neural Architecture Search, NAS) [5]。

NAS 的关键组成部分是：搜索空间、搜索策略、评价策略。搜索空间定义了有哪些可用的架构作为候选，若引入关于适合某项任务的结构属性的先验知识，可以减小搜索空间的大小并简化搜索；搜索策略定义如何在搜索空间中进行搜索；性能评估策略在神经网络结构搜索中用来估计采样到的神经网络结构的泛化性能。由 Google 公司提出的 NAS-Net [6]是第一篇 NAS 工作，NAS-Net 使用 RNN 作为控制器用于生成随机的网络结构，并用目标网络在验证集上的误差作为损失函数训练控制器，因而 NAS-Net 每一个生成的网络都需要进行完整的训练，这大大降低了 NAS-Net 的搜索速度。以 NAS-Net 在 CIFAR10 数据集上进行图像分类的任务表现为例，其需要 500 张 GPU 连续训练 28 天才取得与当时 SOTA 网络相近的实验结果。NAS 任务因而暴露出搜索时间太长、搜索成本高以及计算资源消耗过大等问题，可微分架构搜索 DARTS [7]的出现在一定程度上缓解了这些缺点。DARTS 是一种能够将网络结构变为可微分参数的方法，这意味着网络的训练和架构的调优可以同时进行，且可一边训练一边评价直至收敛。在 CIFAR10 数据集上，DARTS 可以在 4 小时以内搜索出性能超过 ResNet 的网络。因此，DARTS 可以极大地加速神经网络的设计和优化过程，并且具有非常高的效率和性能。

当前的 NAS 大多应用于图像分类任务，为了拓展 NAS 领域的研究深度，本文将 DARTS 应用于语义分割领域。本文的主要工作如下：

1) 将可微分神经网络结构搜索应用于语义分割领域，建立并设计了 U 型搜索空间，对分割网络的 backbone 进行多尺度搜索，通过多尺度结构有效地提取图像在不同分辨率上的特征，通过跳跃连接加快搜索和训练速度。

2) 在 The Oxford-IIIT Pet 数据集搜索得到的网络最优模型命名为 SEARCH-Net，该模型与基准网络 UNet 相比，mIOU 提升 14.1%，直接迁移到 Camvid 数据集上进行测试，比基准网络实验精度提升了 20.5%。

2. 相关工作

语义分割是计算机视觉领域的一个重要研究方向，其目的是将图像中的每个像素分配到不同的语义类别中。该技术在许多应用中都有广泛的应用，如自动驾驶、医学图像分析等。在语义分割的发展历程中，最早的方法是基于手工设计的特征提取器和分类器，如基于 SIFT [8]和 HOG [9]的方法。然而，这些方法需要大量的人工设计和调整，且难以适应不同的场景和任务。随着深度学习技术的发展，基于卷积神经网络(CNN)的语义分割方法逐渐成为主流。其中，最早的方法是基于全卷积网络(FCN) [10]的方法，该方法将传统的卷积神经网络中的全连接层替换为卷积层，从而实现了端到端的像素级别的语义分割。后来，又出现了一系列的改进方法，如 U-Net、SegNet [11]、DeepLab 等。尽管这些方法在许多任务中都取得了很好的效果，但它们仍然存在一些缺点。例如，FCN 等方法没有考虑到不同尺度的特征对语义分割的影响，导致分割结果不够精细；U-Net 等方法虽然考虑了多尺度特征，但其上下采样的方式容易导致信息丢失；DeepLab 等方法虽然采用了空洞卷积来扩大感受野，但其计算复杂度较高，难以应用于实际场景中。

神经网络架构搜索(NAS)把自动化超参数搜索技术引入深度神经网络结构的设计过程当中，通过启发式方法自动化地搜索出比人类研究者手工设计的神经网络结构拥有更高性能的神经网络结构。NAS 的出现为探索更加高效和精确的语义分割方法提供了良好的思路和方向，因此可采用采用 NAS 方法针对语义分割任务自动构建模型。可微分架构搜索 DARTS 可将搜索空间表示成有向无环图的形式，接着通过梯度下降的方式去寻找最优的网络结构。本文设计了不同于 DARTS 的 U 型搜索空间，对分割网络的 backbone 进行多尺度的搜索，缓解传统网络信息丢失的问题。

3. 基于 DARTS 的图像语义分割方法

3.1. 可微分神经网络架构搜索

可微分神经网络架构搜索(Differentiable Architecture Search，简称 DARTS)是一种自动化神经网络架构搜索方法。它通过在训练过程中使用梯度信息来搜索最优的神经网络结构。

DARTS 的核心思想是将神经网络的结构搜索问题转化为一个优化问题，DARTS 通过引入超图来进行这一转化。即把神经网络架构设计表示成一个有向无环图(超图)，超图中的节点代表神经元或特征图，超边代表这些元素之间的连接，这些连接代表神经元之间可能的操作，如卷积、池化或者恒等映射等。而且不同于传统的神经网络结构搜索方法需要在离散的搜索空间中进行搜索，DARTS 通过引入可微分的操作来实现连续的搜索空间。这样一来，DARTS 可以使用梯度下降等优化算法来搜索最优的网络结构。

DARTS 通过首先在当前超图结构上展开整个网络，得到具有权重参数的完整网络，然后再在该网络上进行端到端训练，获取超图结构的梯度信息。超图中每个节点对应的特征图或神经元可以表示为 x_i，每个超边连接表示为 $\epsilon_{i,j}$，从而超边结构包括 n 种可能的操作和 n 个结构参数 $\alpha_i\left(i=1,\cdots,n\right)$，对于每个超边，如果它存在连接，则 $\epsilon_{i,j}=1$，否则 $\epsilon_{i,j}=0$。因此 DARTS 的目标便可表述成寻找最优的超边结构组合，所以在搜索过程中，需要对该超图结构进行优化，得到该结构的连接概率并且保证搜索得到的网络架构可导，以控制网络架构的稀疏性。为了实现这一目的，DARTS 使用一个 softmax 函数来将离散选择边的操作弱化为连续空间，这样就保证了网络结构的可微分性。设 c_i 为该超边上的连接概率，DARTS 的连续松弛化操作表示为：

$$c_i=\frac{\exp\left(\alpha_i\right)}{\sum_{j=1}^{n}\exp\left(\alpha_j\right)},\sum_{i=1}^{n}c_i=1 \tag{1}$$

最终，DARTS 的目标函数可以表示为：

$$L_{train}(w,\alpha)=\frac{1}{\left|D_{train}\right|}\sum_{(x,y)\in D_{train}}L\left(f_{\alpha,w}^{c}(x),y\right) \tag{2}$$

其中 D_{train} 是训练集，$f_{\alpha,w}^{c}$ 表示带有超图结构 α 的神经网络，w 是权重参数。L 表示损失函数，该函数可以根据不同的任务而更改。在训练过程中，DARTS 同时使用梯度下降的方式优化结构参数 α 和权重参数 w，使 DARTS 能够在不需要重新搜索结构的情况下适应于不同的训练任务。通过不断迭代搜索和优化的过程，DARTS 可以自动地搜索出最优的神经网络结构。相比传统的手动设计网络结构的方法，DARTS 具有更高的效率和准确性。它可以在保持网络性能的同时，大大减少了人工设计的工作量。

DARTS 的搜索流程可用流程图表示，见下图 1：

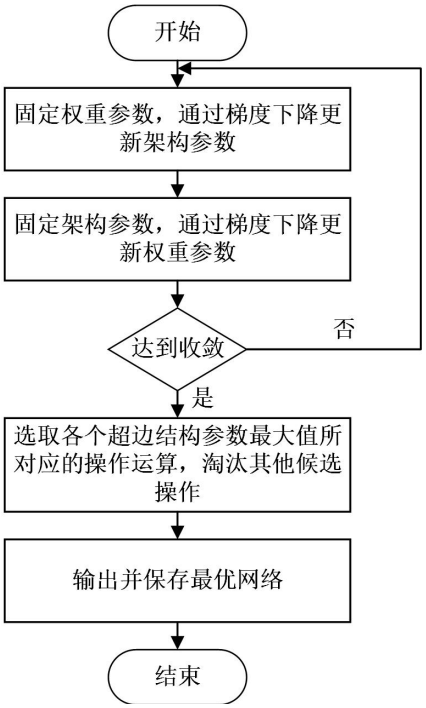


Figure 1. The process of DARTS search network architecture
图 1. DARTS 搜索网络架构流程

3.2. 搜索空间

有了式(1)和式(2)的理论基础，本文设计了如下搜索空间，如图 2 所示。

如上图所示，C1 到 C8 表示 8 个搜索单元，搜索单元下的候选操作为：

1) Conv_3 × 3，3 × 3 卷积

2) Conv_5 × 5，5 × 5 卷积

3) Dilconv_3 × 3，3 × 3 膨胀卷积

4) Dilconv_5 × 5，5 × 5 膨胀卷积

5) MBconv_3 × 3，3 × 3 深度可分离卷积

6) MBconv_5 × 5，5 × 5 深度可分离卷积

7) 恒等连接

与 DARTS 原本的候选操作集不同，本文新增了 MBconv [12]带有 SE 注意力机制块的深度可分离卷积，SE (Squeeze-and-Excitation)注意力机制块能够方便地嵌入到神经网络结构中，可以直接放置在卷积层之后，通过全局特征挖掘来调整每个通道的重要性，无需增加过多的参数。此外，SE 注意力机制块能够对特征图进行自适应的加权操作，能够更好地融合特定的位置信息和特征信息，从而提高模型的性能和泛化能力。

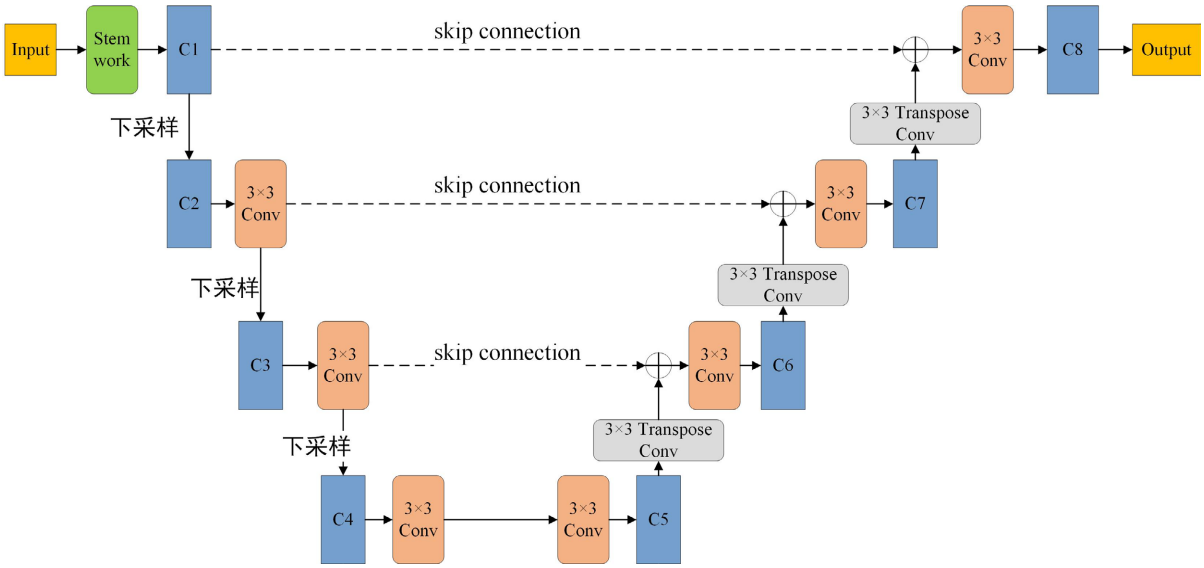


Figure 2. U-shaped search space architecture
图 2. U 型搜索空间框架

为了能够获取图像更多的特征信息，基于分治的思想，在搜索空间中设置 8 个搜索单元 4 个分辨率尺度，搜索单元的结构如下图 3：

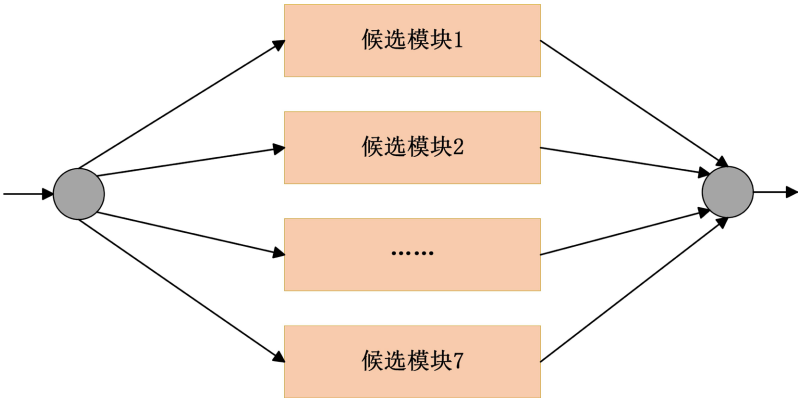


Figure 3. Search unit architecture
图 3. 搜索单元结构

在原始分辨尺度设置 C1 和 C8 单元，单元中包含 64 个特征图，C1 是下采样过程中的搜索单元，C8 是上采样过程中的搜索单元。同理，在二级分辨尺度上，下采样单元 C2 和上采样单元 C7 包含 128 个特征图；在三级分辨尺度上，下采样单元 C3 和上采样单元 C6 包含 256 个特征图；在四级分辨尺度

上，下采样单元 C4 和上采样单元 C5 包含 512 个特征图。虽然这 8 个搜索单元共享统一的候选操作，但经过 DARTS 搜索过后，由于每个超图的结构参数和网络权重不同，因此每个搜索单元得到的结构也会千差万别。在相同分辨率尺度上，使用与 UNet 一样的跳跃连接，实现了更多细节信息的传递和融合。

4. 实验

4.1. 实验数据集

为了客观的评价本文的方法，本文使用以下两个公开数据集进行测试：Oxford-IIIT Pet 数据集[13]和 CamVid 数据集[14]。前者常用于图像分类、目标检测和分割等任务，它由牛津大学和印度理工学院创建，包含了来自不同品种的猫和狗的图像。该数据集包含了 37 个不同的宠物动物类别，每个类别代表一种特定的猫或狗的品种。每个类别大约有 200 张不同的图像，总共超过 7000 张图像。并且数据集中的图像展示了宠物以不同的姿势、背景和光照条件出现，这种多样性有助于训练更健壮、能够较好地泛化到未见过数据的模型。与此同时，数据集中的每个图像都附带有准确和详细的注释，如对象边界框、真实分割掩模和类别标签。这些注释有助于进行目标检测和分割等任务。

后者由剑桥大学的研究人员在 2007 年创建。该数据集包括 700 多张精准标注的图片用于强监督学习，可分为训练集、验证集、测试集。这些图像数据通过在剑桥市区的汽车上安装的摄像头拍摄的。CamVid 数据集总共包含 32 个不同的语义类别，包括道路、行人、汽车、建筑物、树木等，因而该数据集的图像变异性较大，包含了在不同季节、不同天气、不同时间和不同地点拍摄的图像。这种多样性使得模型需要具备较强的泛化能力，能够在多样的道路场景中进行准确的语义分割，并且每个像素都被标记为对应的语义类别，这使得 CamVid 数据集适用于语义分割任务的训练和评估。

4.2. 实验设置

本文选取 Dice Loss [15]作为损失函数，Dice Loss 是一种用于语义分割任务的损失函数，它的基本思想是计算预测结果和真实结果的重叠部分，通过最小化两者的差异来优化模型，Dice Loss 越小表示模型的预测结果与真实标签越相似。它的计算公式如下：

$$\text{Dice Loss} = 1 - \frac{2\sum_{i=1}^{N} p_i * q_i}{\sum_{i=1}^{N} p_i^2 + \sum_{i=1}^{N} q_i^2} \tag{3}$$

其中，N 是像素总数，p_i 是模型预测的第 i 个像素的值，q_i 是真实标签的第 i 个像素的值。Dice Loss 的取值范围在 0 到 1 之间，当预测结果完全匹配真实标签时，Dice Loss 为 0；当预测结果与真实标签没有重叠部分时，Dice Loss 为 1。Dice Loss 相比于其他损失函数，如交叉熵损失函数，更适合处理图像分割任务中不均衡的类别问题。它将像素级别的误差考虑在内，能够对小目标和边缘等关键区域进行更精确的预测。在训练过程中，通过最小化 Dice Loss 来优化模型参数，使得预测结果与真实标签之间的相似度最大化。这样可以帮助模型更好地分割图像，并提供更准确的边界。

为了反应模型分割结果的质量，选用 mIOU (Mean Intersection over Union)作为评价指标，其可以用于衡量模型在图像分割任务中的性能。mIOU 是通过计算所有类别的交并比的平均值来评估模型的性能。具体而言，对于每个类别，mIOU 计算该类别的预测区域与真实区域的交集面积除以它们的并集面积，然后将所有类别的交并比求平均。这样可以得到一个介于 0 和 1 之间的值，其中 0 表示完全不匹配，1 表示完全匹配。mIOU 的计算公式如下：

$$\text{mIOU} = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_i}{FN_i + FP_i + TP_i} \tag{4}$$

其中，TP_i 表示第 i 类别的真正例数(即正确分类为该类别的像素数)，FP_i 表示第 i 类别的假正例数(即错误分类为该类别的像素数)，FN_i 表示第 i 类别的假负例数(即未正确分类为该类别的像素数)，N 表示总类别数。

基于上述的损失函数和评价指标，本文实验在搭载 GeForce RTX 3090 显卡的服务器上进行，使用 Pytorch1.10 框架。在模型架构搜索阶段，每一次超网训练轮数为 300 轮，设定初始学习率为 0.001，采用 Adam 优化器优化；在对搜索出来的架构重新训练阶段，训练轮数为 200 轮，初始学习率为 0.001，采用余弦退火算法自动修正，冲量参数 0.9，权重衰减率为 0.0005。

4.3. 模型评估

在 Oxford-IIIT Pet 数据集进行网络搜索，经过 300 轮的搜索后，把最优的网络保存并命名为 SEARCH-Net，将 SEARCH-Net 在该数据集上重新进行训练，训练的 loss 和 mIOU 收敛曲线如下图 4 和图 5，loss 损失收敛于 0.08，评价指标 mIOU 最后在 0.953 上下波动。为了衡量 SEARCH-Net 的网络性能，选取 UNet、FPN [16]、UNet++ [17]、SETR [18]等方法在同一数据集上进行分割效果的比较，分割结果如表 1 所示。

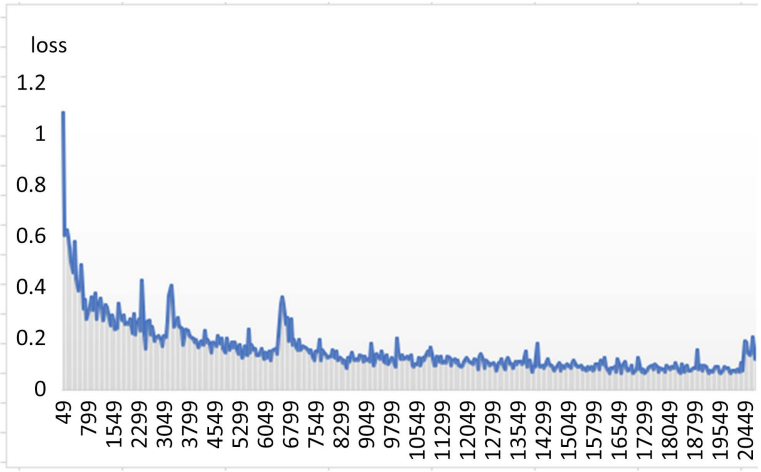


Figure 4. Accuracy variation chart of loss for SEARCH-Net
图 4. SEARCH-Net 的 loss 损失精度变化图

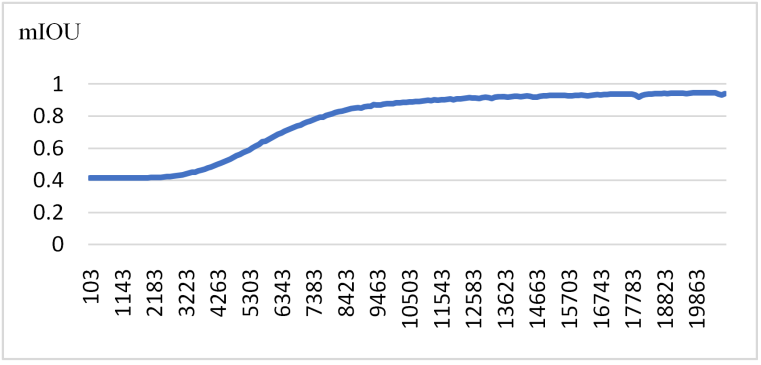


Figure 5. Convergence curve of mIOU for SEARCH-Net
图 5. SEARCH-Net 的 mIOU 收敛曲线

Table 1. Comparison with other segmentation methods on the Oxford-IIIT
表 1. Oxford-IIIT Pet 数据集实验结果

	Loss	mIOU/%
FPN [16]	0.39	74.3
UNet [3]	0.41	81.2
DeepLabV3 [19]	0.13	92.7
UNet++ [17]	0.11	97.3
SETR [18]	0.13	95.6
SEARCH-Net (本文方法)	0.08	95.3

为了更加直观的比较各种方法语义分割的效果，选取 UNet 与本文方法以及 SETR 和本文方法的视觉对比图作为示例：

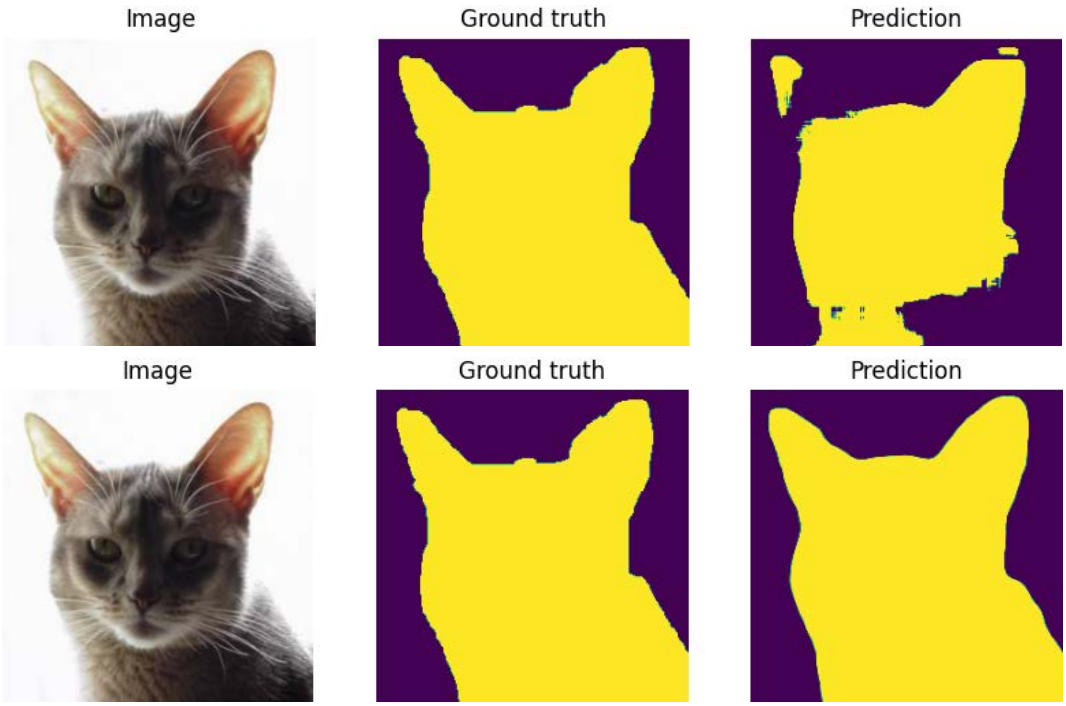


Figure 6. Visual effects of UNet segmentation (Top) and SEARCH-Net segmentation (Bottom)
图 6. UNet 分割视觉效果(上)和 SEARCH-Net 分割视觉效果(下)

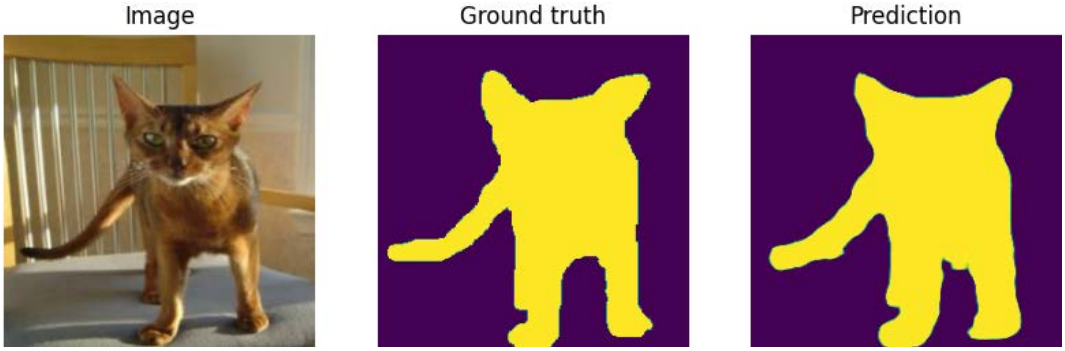


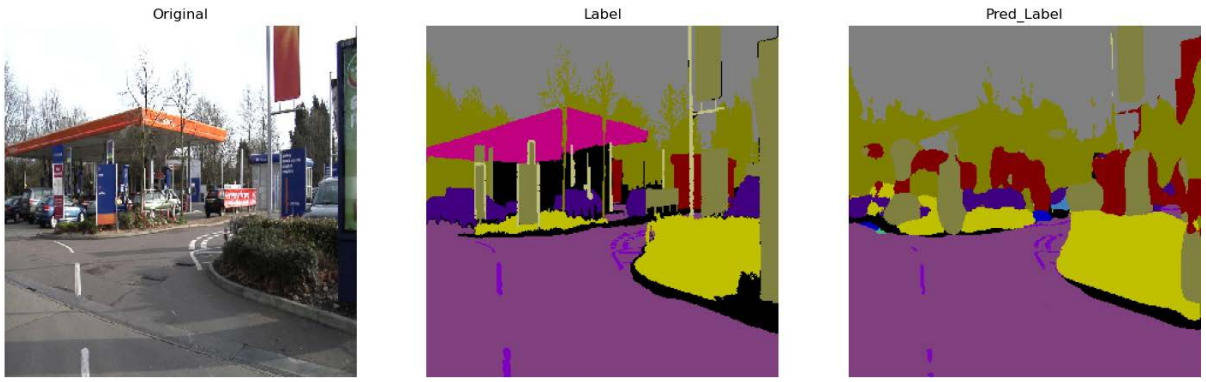
Figure 7. Visual effects of SETR segmentation (Top) and SEARCH-Net segmentation (Bottom)
图 7. SETR 分割视觉效果(上)和 SEARCH-Net 分割视觉效果(下)

从表 1 分析可知，基于可微分架构搜索的语义分割技术在实践层面得到验证，在公开数据集 Oxford-IIIT Pet 上的表现优于传统的语义分割方法 FPN 和 UNet 网络，评价指标 mIOU 相比于基线网络 UNet 提升了 14.1%，与基于 transformer 的 SETR 先进网络性能相当。从图 6 和图 7 可以看出，基于可微分架构搜索搭建出的网络结构有良好的视觉体验效果。在图 6 中，UNet 网络对猫的分割效果较差，分割轮廓不完整，猫耳朵细节没有完全分割出来，猫的身体部分也有缺失，原因在于该网络在无法完整提图片数据的前景特征信息；本文方法 SEARCH-Net 对猫的分割效果较好，猫的整体轮廓分割清晰，分割的效果更加显著，说明本文网络提取图像细节特征的能力更强，这归功于包含 SE 注意力机制的 MBconv 块，它能够更好地融合特定的位置信息和细节特征信息，从而提高了模型的性能。虽然在数据指标上基于 transformer 方法的 SETR 网络效果与 SEARCH-Net 的分割性能接近，但从图 7 可以看出，前者的分割效果略逊于本文方法的分割效果，原因在于 SETR 架构具有大量的自注意力机制和多头注意力机制，这导致了对图像信息过度处理，将椅子靠背底部边缘分割出来，从而导致对猫尾部分的分割出现误差；而本文方法清晰的将猫咪轮廓从椅子上分割出来，有较好的分割效果。

为了更好的体现搜索网络 SEARCH-Net 的泛化性，将其与基线网络 UNet 在 CamVid 数据集进行对比，结果如下表 2：

Table 2. Comparison results on the CamVid Dataset
表 2. CamVid 数据集实验结果

	Loss	mIOU/%
UNet [3]	0.23	46.2
SEARCH-Net (本文方法)	0.14	66.7



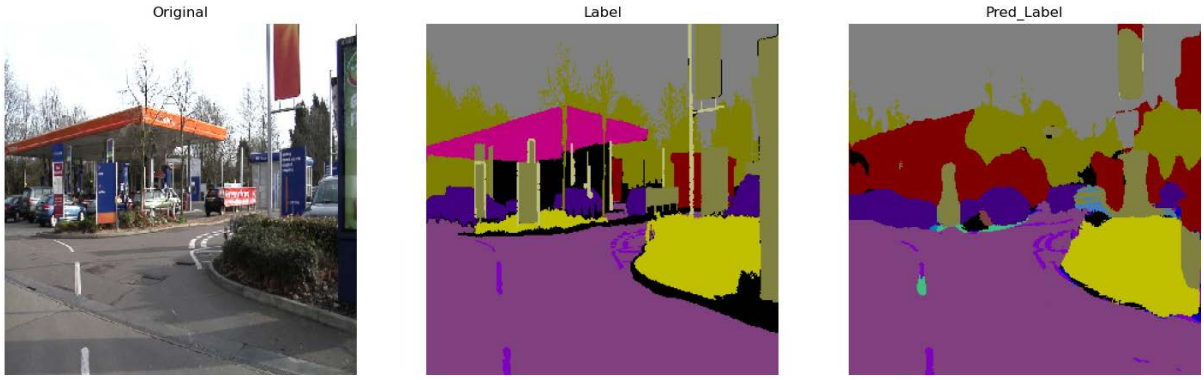


Figure 8. UNet (Top) and SEARCH-Net (Bottom) segmentation results on the CamVid Dataset
图 8. 在 CamVid 数据集 UNet 分割结果(上)和 SEARCH-Net (下)分割结果

由表 2 中的数据以及分割效果图 8 可知，SEARCH-Net 具有良好的泛化性，经过重新训练后能适用于其他语义分割数据集，使用 NAS 做语义分割网络搜索具有良好的可行性。

5. 总结

本文将可微分网络架构搜索引入到语义分割当中，提出 U 型搜索空间以及新的搜索候选操作，并且在不同尺度上进行网络结构的搜索，不仅降低了网络搜索的计算难度，还可以提升网络模型的精度。搜索出来的网络架构 SEARCH-Net 在 Oxford-IIIT Pet 数据集上表现良好，与基线网络 UNet 相比，mIOU 提升了 14.1%，并且与先进网络 SETR 相比，网络性能相当。把 SEARCH-Net 迁移到 CamVid 数据集上后，网络显示出良好的泛化性能。将可微分网络架构搜索引入到语义分割网络的建设当中，极大地减少了人工设计网络的难度，能使算法自动设计出适应分割任务的网络结构，提升了网络设计的效率。然而本文方法也存在着局限性，主要是在实现过程当中网络搜索需要大量的计算资源，并且推理速度较慢，同时结构也受搜索空间的限制，设计不同的搜索空间会对搜索结果产生影响。但初始的 NAS 神经网络架构搜索大多只局限于图像分类任务，本文的工作将 NAS 拓展到语义分割领域，并验证了其可行性，丰富了 NAS 的任务领域。在接下来的研究当中，计划将剪枝等技术与 NAS 结合，争取设计出复杂度更低、推理速度更快的语义分割模型。

基金项目

天津市教委科研计划项目(2020KJ115)。

参考文献

[1] Sengupta, A., Ye, Y., Wang, R., Liu, C. and Roy, K. (2019) Going Deeper in Spiking Neural Networks: VGG and Residual Architectures. *Frontiers in Neuroscience*, **13**, Article 95. https://doi.org/10.3389/fnins.2019.00095

[2] Targ, S., Almeida, D. and Lyman, K. (2016) Resnet in Resnet: Generalizing Residual Architectures. ArXiv Preprint ArXiv: 1603.08029.

[3] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W. and Frangi, A., Eds., *Medical Image Computing and Computer-Assisted Intervention—MICCAI* 2015. *MICCAI* 2015. *Lecture Notes in Computer Science*, Vol. 9351, Springer, Cham, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28

[4] Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K. and Yuille, A.L. (2017) DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **40**, 834-848. https://doi.org/10.1109/TPAMI.2017.2699184

[5] Elsken, T., Metzen, J.H., Hutter, F. (2019) Neural Architecture Search: A Survey. *The Journal of Machine Learning*

Research, **20**, 1997-2017.

[6] Zoph, B. and Le, Q.V. (2016) Neural Architecture Search with Reinforcement Learning. ArXiv Preprint ArXiv: 1611.01578.

[7] Liu, H., Simonyan, K. and Yang, Y. (2018) DARTS: Differentiable Architecture Search. ArXiv Preprint ArXiv: 1806.09055.

[8] Ng, P.C. and Henikoff, S. (2003) SIFT: Predicting Amino Acid Changes That Affect Protein Function. *Nucleic Acids Research*, **31**, 3812-3814. https://doi.org/10.1093/nar/gkg509

[9] Pang, Y., Yuan, Y., Li, X. and Pan, J. (2011) Efficient HOG Human Detection. *Signal Processing*, **91**, 773-781. https://doi.org/10.1016/j.sigpro.2010.08.010

[10] Villa, M., Dardenne, G., Nasan, M., *et al*. (2018) FCN-Based Approach for the Automatic Segmentation of Bone Surfaces in Ultrasound Images. *International Journal of Computer Assisted Radiology and Surgery*, **13**, 1707-1716. https://doi.org/10.1007/s11548-018-1856-x

[11] Badrinarayanan, V., Kendall, A. and Cipolla, R. (2017) SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 2481-2495. https://doi.org/10.1109/TPAMI.2016.2644615

[12] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. and Chen, L.-C. (2018) Mobilenetv2: Inverted Residuals and Linear Bottlenecks. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 4510-4520. https://doi.org/10.1109/CVPR.2018.00474

[13] Parkhi, O.M., Vedaldi, A., Zisserman, A. and Jawahar, C.V. (2012) Cats and Dogs. 2012 *IEEE Conference on Computer Vision and Pattern Recognition*, Providence, 16-21 June 2012, 3498-3505. https://doi.org/10.1109/CVPR.2012.6248092

[14] Brostow, G.J., Fauqueur, J. and Cipolla, R. (2009) Semantic Object Classes in Video: A High-Definition Ground Truth Database. *Pattern Recognition Letters*, **30**, 88-97. https://doi.org/10.1016/j.patrec.2008.04.005

[15] Li, X., Sun, X., Meng, Y., *et al*. (2020) Dice Loss for Data-Imbalanced NLP Tasks. *Proceedings of the* 58*th Annual Meeting of the Association for Computational Linguistics*, Online, 5-10 July 2020, 465-476. https://doi.org/10.18653/v1/2020.acl-main.45

[16] Lin, T.-Y., Dollár, P., Girshick, R., *et al*. (2017) Feature Pyramid Networks for Object Detection. 2017 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, 21-26 July 2017, 936-944. https://doi.org/10.1109/CVPR.2017.106

[17] Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N. and Liang, J. (2019) UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation. *IEEE Transactions on Medical Imaging*, **39**, 1856-1867. https://doi.org/10.1109/TMI.2019.2959609

[18] Zheng, S., Lu, J., Zhao, H., *et al*. (2021) Rethinking Semantic Segmentation from a Sequence-to-Sequence Perspective with Transformers. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Nashville, 20-25 June 2021, 6877-6886. https://doi.org/10.1109/CVPR46437.2021.00681

[19] Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F. and Adam, H. (2018) Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In: Ferrari, V., Hebert, M., Sminchisescu, C. and Weiss, Y., Eds., *Computer Vision—ECCV* 2018. *ECCV* 2018. *Lecture Notes in Computer Science*, Vol. 11211, Springer, Cham, 833-851. https://doi.org/10.1007/978-3-030-01234-2_49