论文题目

Auto-DeepLab: Hierarchical Neural Architecture Search for Semantic Image Segmentation

论文作者

Chenxi Liu, Liang-Chieh Chen, Florian Schroff, Hartwig Adam, Wei Hua,Alan Yuille, Li Fei-Fei
Johns Hopkins University
Google
Stanford University

代码地址

https://github.com/tensorflow/models/tree/master/research/deeplab

论文的前提

NMS(Neural Architecture Search)是 AutoML 中得一种,目前的模型都是要工程师通过自身得学识和经验设计出网络架构,而 NMS 可以通过在 search space 中通过 DQN 或是 EA 等方式得到适合当前数据的网络结构
目前 NMS 在分类上的效果是比较好的了,但是仍旧没有在语义分割上有应用

论文观点

将 NAS 用于语义分割任务上

论文做法

1.cell level(网络的基本单元,包括更小的 block)是搜索出来的,network level(空间分辨率的变化)也是搜索出来的

cell level 的 search space

fully cnn 单元,内部包含 block,每一个 block 是一个 two-branch architecture,每一个 block 包含 5 个参数(I1[第一个 branch 的输入], I2, O1[第一个 branch 的输出], O2, C[两个 branch 合并的方法,目前仅支持两个分支逐元素相加])

I->O 的过程:

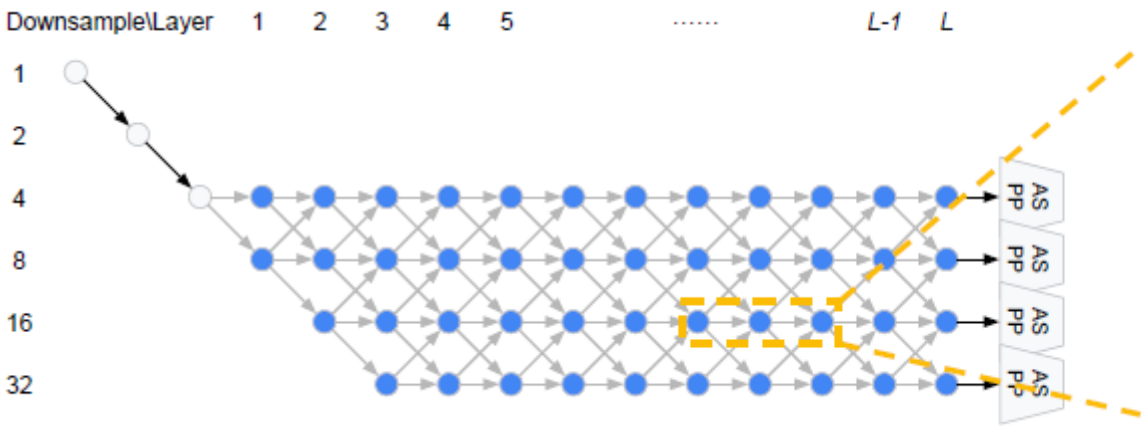
- 3×3 depthwise-separable conv
- 5×5 depthwise-separable conv
- 3×3 atrous conv with rate 2
- 5×5 atrous conv with rate 2
- 3×3 average pooling
- 3×3 max pooling
- skip connection
- no connection (zero)

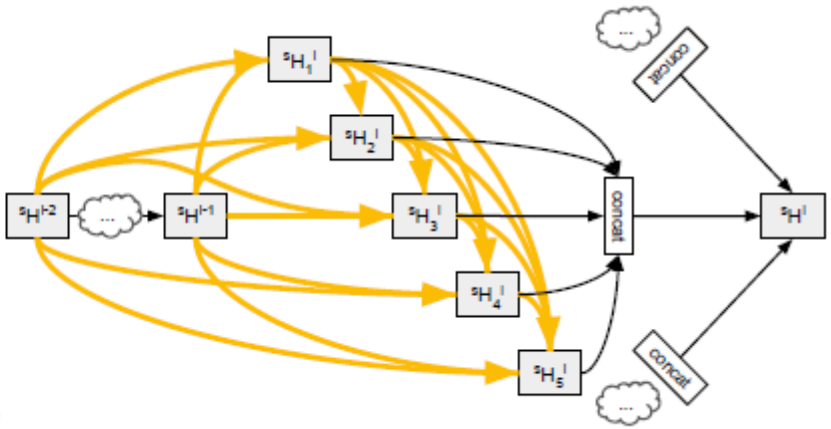
network level 的 search space

相邻两层的空间分辨率可以使 1/2 或是 2 倍或者不变的关系

最小的分辨率不能小于 1 / 32

开始的 stem 的 feature map 是原来的 1/4





2.将在离散的搜索空间转换到连续的搜索空间中,使得整个过程可以用 gradient descent 求解

cell level 搜索方式

cell level 我们要找出每个 block 的每个 branch 要使用哪个 input,以及对这些 input 使用上面 8 种转换方式的那些

让每个 input 使用全部的转换,再乘上 α,如果 input 的候选有 n 个,转换有 8 种,那应该每一个 block 有 8n 个 α

α 的限制是要大于零,且加总为 1

最后找到 α 最大的两个值就找到了在 cell level 上该 block 的两个 branch 使用了哪个 input,以及使用了哪两个转换

network level 搜索方式

某个分辨率的可能从三个地方转换到这里,一是比他两倍大的地方(缩小 1/2),一是比他两倍小的地方(放大 2 倍),三是分辨率和他一样大的地方

分别给予三个β权重

最后可以将β视为转移机率找到一条最大机率的路径(Viterbi)

最终效果

Method	MS	COCO	mIOU (%)
DropBlock [19]			53.4
Auto-DeepLab-S			71.68
Auto-DeepLab-S	✓		72.54
Auto-DeepLab-M			72.78
Auto-DeepLab-M	✓		73.69
Auto-DeepLab-L			73.76
Auto-DeepLab-L	✓		75.26
Auto-DeepLab-S		✓	78.31
Auto-DeepLab-S	✓	✓	80.27
Auto-DeepLab-M		✓	79.78
Auto-DeepLab-M	✓	✓	80.73
Auto-DeepLab-L		✓	80.75
Auto-DeepLab-L	✓	✓	82.04

Table 5: PASCAL VOC 2012 validation set results. We experiment with the effect of adopting *multi-scale* inference (**MS**) and COCO-pretrained checkpoints (**COCO**). Without any pretraining, our best model (Auto-DeepLab-L) outperforms DropBlock by 20.36%. All our models are not pretrained with ImageNet images.

Method	ImageNet	Coarse	mIOU (%)
FRRN-A [60]			63.0
GridNet [17]			69.5
FRRN-B [60]			71.8
Auto-DeepLab-S			79.9
Auto-DeepLab-L			80.4
Auto-DeepLab-S		✓	80.9
Auto-DeepLab-L		✓	82.1
ResNet-38 [82]	✓	✓	80.6
PSPNet [88]	✓	✓	81.2
Mapillary [4]	✓	✓	82.0
DeepLabv3+ [11]	✓	✓	82.1
DPC [6]	✓	✓	82.7
DRN_CRL_Coarse [91]	✓	✓	82.8

Table 4: Cityscapes test set results with *multi-scale* inputs during inference. **ImageNet:** Models pretrained on ImageNet. **Coarse:** Models exploit coarse annotations.

Method	ImageNet	COCO	mIOU (%)
Auto-DeepLab-S		✓	82.5
Auto-DeepLab-M		✓	84.1
Auto-DeepLab-L		✓	85.6
RefineNet [44]	✓	✓	84.2
ResNet-38 [82]	✓	✓	84.9
PSPNet [88]	✓	✓	85.4
DeepLabv3+ [11]	✓	✓	87.8
MSCI [43]	✓	✓	88.0

Table 6: PASCAL VOC 2012 test set results. Our Auto-DeepLab-L attains comparable performance with many state-of-the-art models which are pretrained on both **ImageNet** and **COCO** datasets. We refer readers to the official leader-board for other state-of-the-art models.

结论

没有在任何 pre-training 的前提下的语义分割任务获得比较先进的性能