SegNet：

A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation

分享人：王嘉恒

目录

1.写作背景

SegNet的设计动机是需要设计一种用于理解道路和室内场景的有效架构。SegNet是用于端到端训练的语义分割网络，它是编码器-解码器结构，作者的创新在于使用了一种池化索引的上采样方法。

2.网络结构

- 编码器网络
- 相应的解码器网络
- 按像素分类层

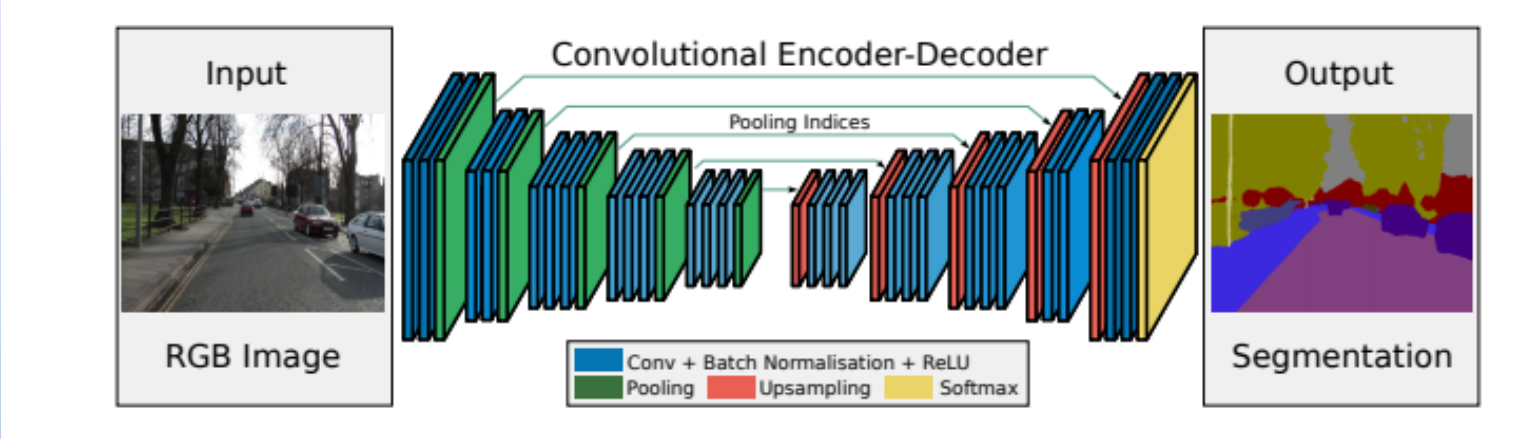

图1. SegNet网络结构

编码器结构是在VGG16的基础上进行的改进，去掉了后3层的全连接层，只取前13层以减少参数量。进行最大池化操作时把滑动窗口最大特征的位置记录下来。

2.网络结构

解码器：前面编码器部分通过池化缩小尺寸，解码器部分使用上采样来恢复尺寸，SegNet中用了池化索引，将最大值的位置记录了下来。
直接将数据放回原位，后面再接卷积核进行学习。

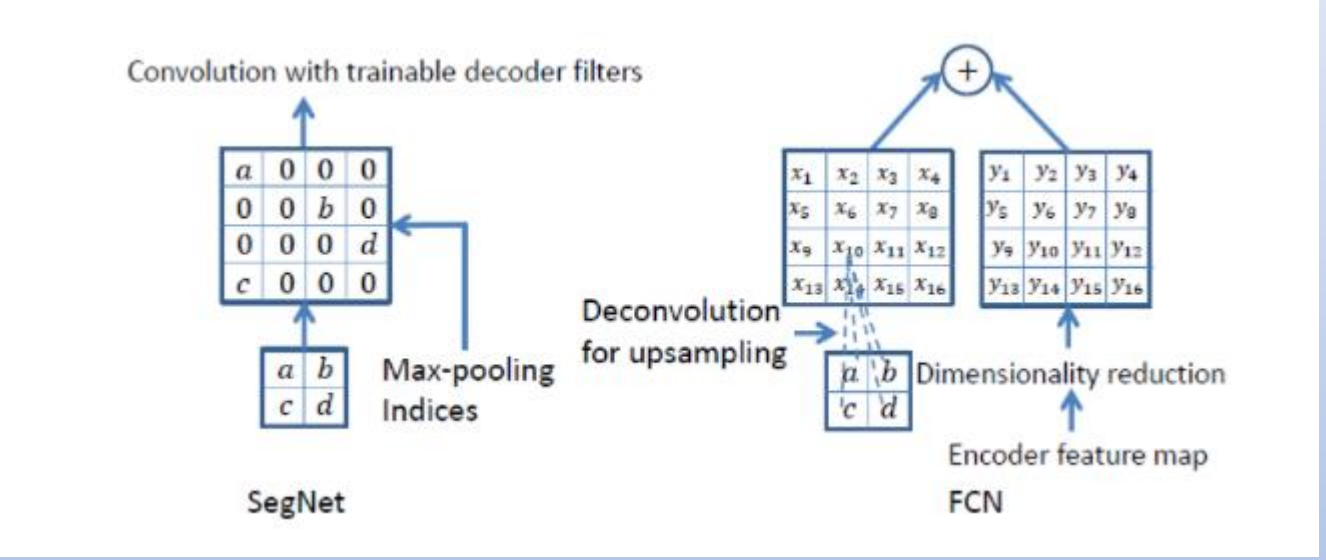

图2. SegNet和FCN的上采样方式

最后是一个Sofmax分类层，但是此处的softmax是对每个像素都进行计算。

2.网络结构

SegNet相较于FCN的优势：

- 可以使特征图边界清晰
- 减小可训练的参数
- 可以广泛使用于编码-解码的语义分割网络，提高其性能

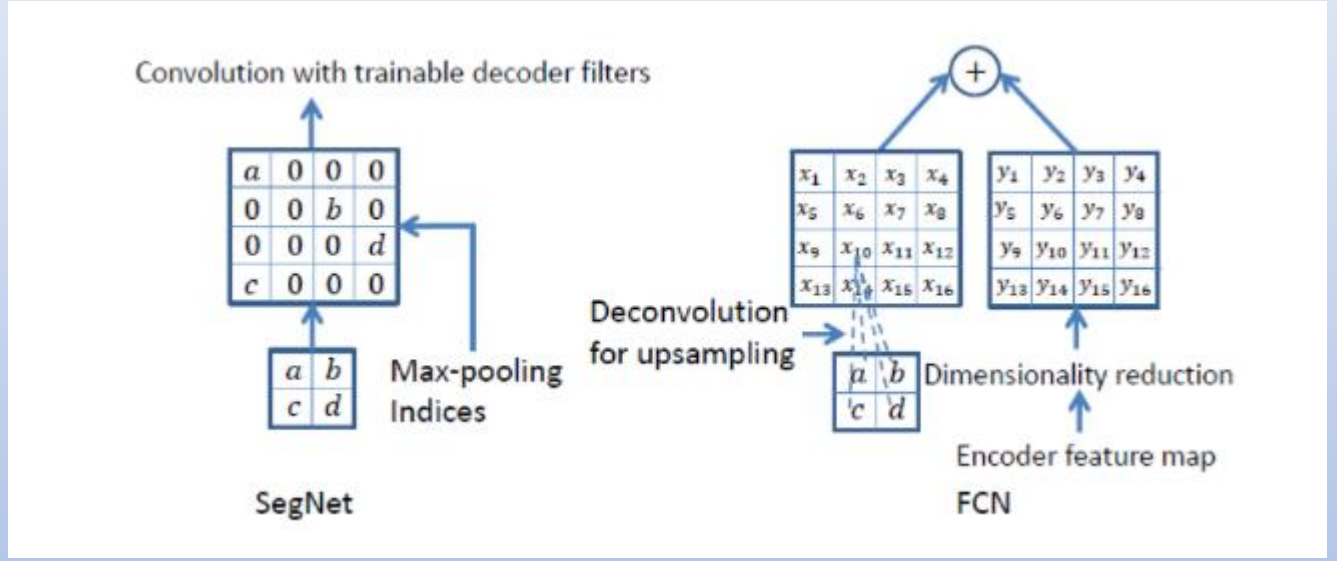

图2. SegNet和FCN的上采样方式

2.网络结构-解码器变体

网络名称	备注
Bilinear-Interpolation	双线性插值上采样固定权值
SegNet-Basic	只有4个编码器和4个译码器
SegNetBasic-SingleChannelDecoder	译码器采用单通道
SegNet-Basic-encoderAddtion	保留编码器的特征图和译码器上采样后的特征图相加
FCN-Basic	相同的编码器，采用FNC的解码方式
FCN-Basic-NoAddition	去掉特征图相加的步骤
FCN-Basic-NoDimReduction	不对编码器的输出进行降维

3.训练过程

- 输入：大小360*480的RGB图像 并进行局部对比度归一
- Lr=0.1 动量=0.9 SGD
- 交叉熵损失
- 对损失进行加权， 中值频率权衡

3.训练

				Median frequency balancing							Natural frequency balancing						
				Test				Train			Test				Train		
Variant	Params (M)	Storage multiplier	Infer time (ms)	G	C	mIoU	BF	G	C	mIoU	G	C	mIoU	BF	G	C	mIoU
Fixed upsampling																	
Bilinear-Interpolation	0.625	0	24.2	77.9	61.1	43.3	20.83	89.1	90.2	82.7	82.7	52.5	43.8	23.08	93.5	74.1	59.9
Upsampling using max-pooling indices																	
SegNet-Basic	1.425	1	52.6	82.7	62.0	47.7	35.78	94.7	96. 2	92.7	84.0	54.6	46.3	36.67	96.1	83.9	73.3
SegNet-Basic-EncoderAddition	1.425	64	53.0	83.4	**63.6**	48.5	35.92	94.3	95.8	92.0	**84.2**	56.5	**47.7**	36.27	95.3	80.9	68.9
SegNet-Basic-SingleChannelDecoder	0.625	1	33.1	81.2	60.7	46.1	31.62	93.2	94.8	90.3	83.5	53.9	45.2	32.45	92.6	68.4	52.8
Learning to upsample (bilinear initialisation)																	
FCN-Basic	0.65	11	24.2	81.7	62.4	47.3	**38.11**	92.8	93.6	88.1	83.9	55.6	45.0	**37.33**	92.0	66.8	50.7
FCN-Basic-NoAddition	0.65	n/a	23.8	80.5	58.6	44.1	31.96	92.5	93.0	87.2	82.3	53.9	44.2	29.43	93.1	72.8	57.6
FCN-Basic-NoDimReduction	1.625	64	44.8	**84.1**	63.4	**50.1**	37.37	**95.1**	**96.5**	**93.2**	83.5	**57.3**	47.0	37.13	**97.2**	**91.7**	**84.8**
FCN-Basic-NoAddition-NoDimReduction	1.625	0	43.9	80.5	61.6	45.9	30.47	92.5	94.6	89.9	83.7	54.8	45.5	33.17	95.0	80.2	67.8

Bilinear-Interpolation表现最差，说明了在进行分割时，decoder学习的重要性。

encoder特征图全部存储时，性能最好。

当限制存储时，可以使用适当的decoder（例如SegNet类型）来存储和使用encoder特征图（维数降低，max-pooling indices）的压缩形式来提高性能。

更大的decoder提高了网络的性能。

4.基准测试（和其他深度学习网络比较）

（1）基于Camvid数据集

SegNet和DeconvNet的性能的性能最好

对于FCN，学习反卷积层而不是用双线性插值权值来固定，可以提高性能。

Network/Iterations	40K				80K				>80K				Max iter
	G	C	mIoU	BF	G	C	mIoU	BF	G	C	mIoU	BF	
SegNet	88.81	59.93	50.02	35.78	89.68	69.82	57.18	42.08	90.40	71.20	60.10	46.84	140K
DeepLab-LargeFOV [3]	85.95	60.41	50.18	26.25	87.76	62.57	53.34	32.04	88.20	62.53	53.88	32.77	140K
DeepLab-LargeFOV-denseCRF [3]	not computed								89.71	60.67	54.74	40.79	140K
FCN	81.97	54.38	46.59	22.86	82.71	56.22	47.95	24.76	83.27	59.56	49.83	27.99	200K
FCN (learnt deconv) [2]	83.21	56.05	48.68	27.40	83.71	59.64	50.80	31.01	83.14	64.21	51.96	33.18	160K
DeconvNet [4]	85.26	46.40	39.69	27.36	85.19	54.08	43.74	29.33	89.58	70.24	59.77	52.23	260K

图3.在Camvid上的结果

global accuracy (G)：数据集中正确分类的像素的百分比

class average accuracy (C)：所有类别预测准确率的平均值

boundary F1-measure (BF)：涉及计算边界像素的F1指标。

mean intersection over union (mIoU)：比类平均准确率更严格，因为它惩罚FP预测；

4.基准测试

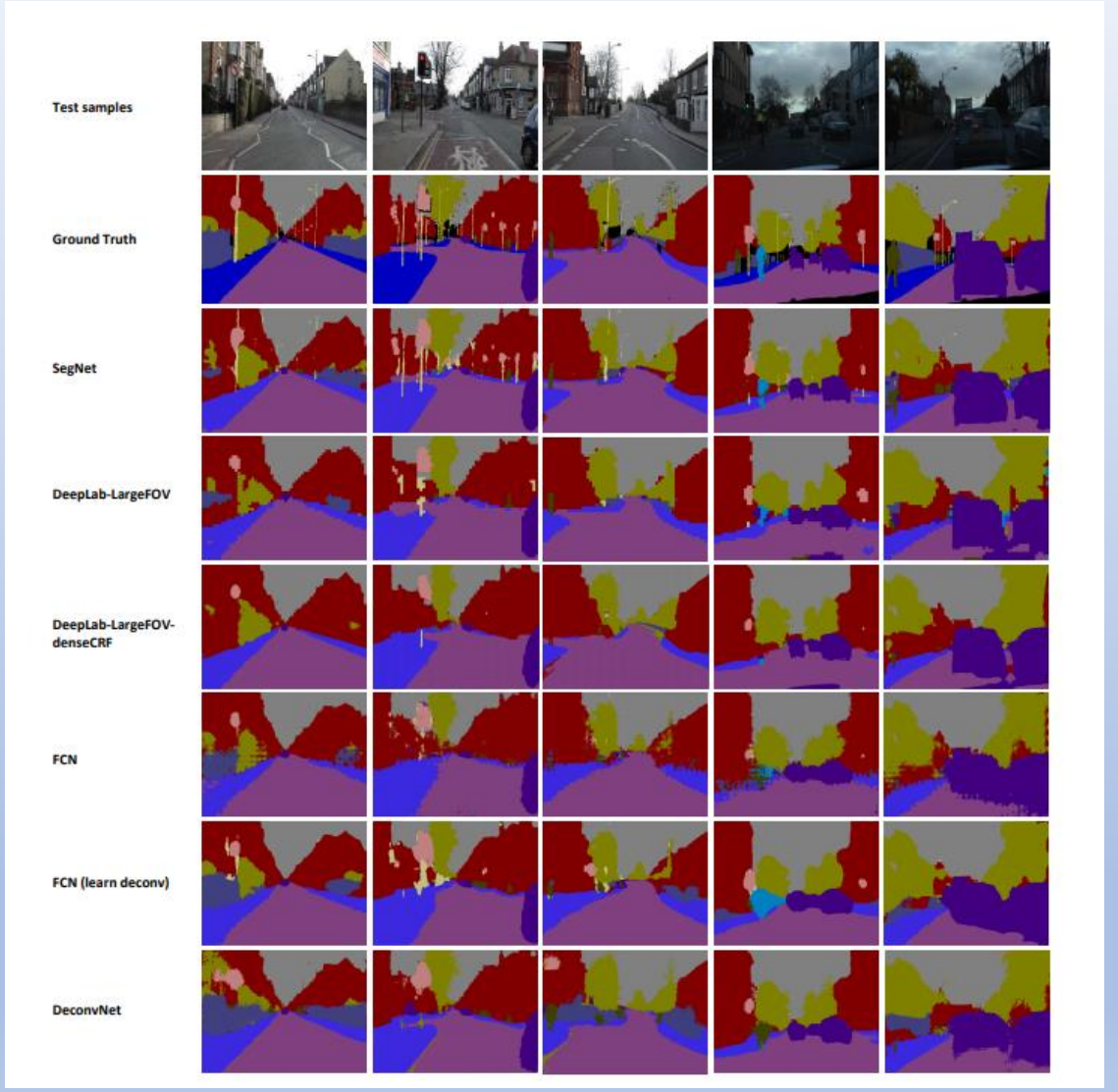

图4.在Camvid上的结果（定性分析）

4.基准测试

(2) 在SUNRGB-D上的结果

在室内环境数据集上各个网络的结果都很差，在较小的类别识别精度很低。

作者推测原因可能是深层架构在室内场景中缺乏可变性。

Network/Iterations	80K				140K				>140K				Max iter
	G	C	mIoU	BF	G	C	mIoU	BF	G	C	mIoU	BF	
SegNet	70.73	30.82	22.52	9.16	71.66	37.60	27.46	11.33	72.63	44.76	31.84	12.66	240K
DeepLab-LargeFOV [3]	70.70	41.75	30.67	7.28	71.16	42.71	31.29	7.57	71.90	42.21	32.08	8.26	240K
DeepLab-LargeFOV-denseCRF [3]	not computed								66.96	33.06	24.13	9.41	240K
FCN (learnt deconv) [2]	67.31	34.32	24.05	7.88	68.04	37.2	26.33	9.0	68.18	38.41	27.39	9.68	200K
DeconvNet [4]	59.62	12.93	8.35	6.50	63.28	22.53	15.14	7.86	66.13	32.28	22.57	10.47	380K

图5. 在SUNRGB-D上的结果

4.基准测试

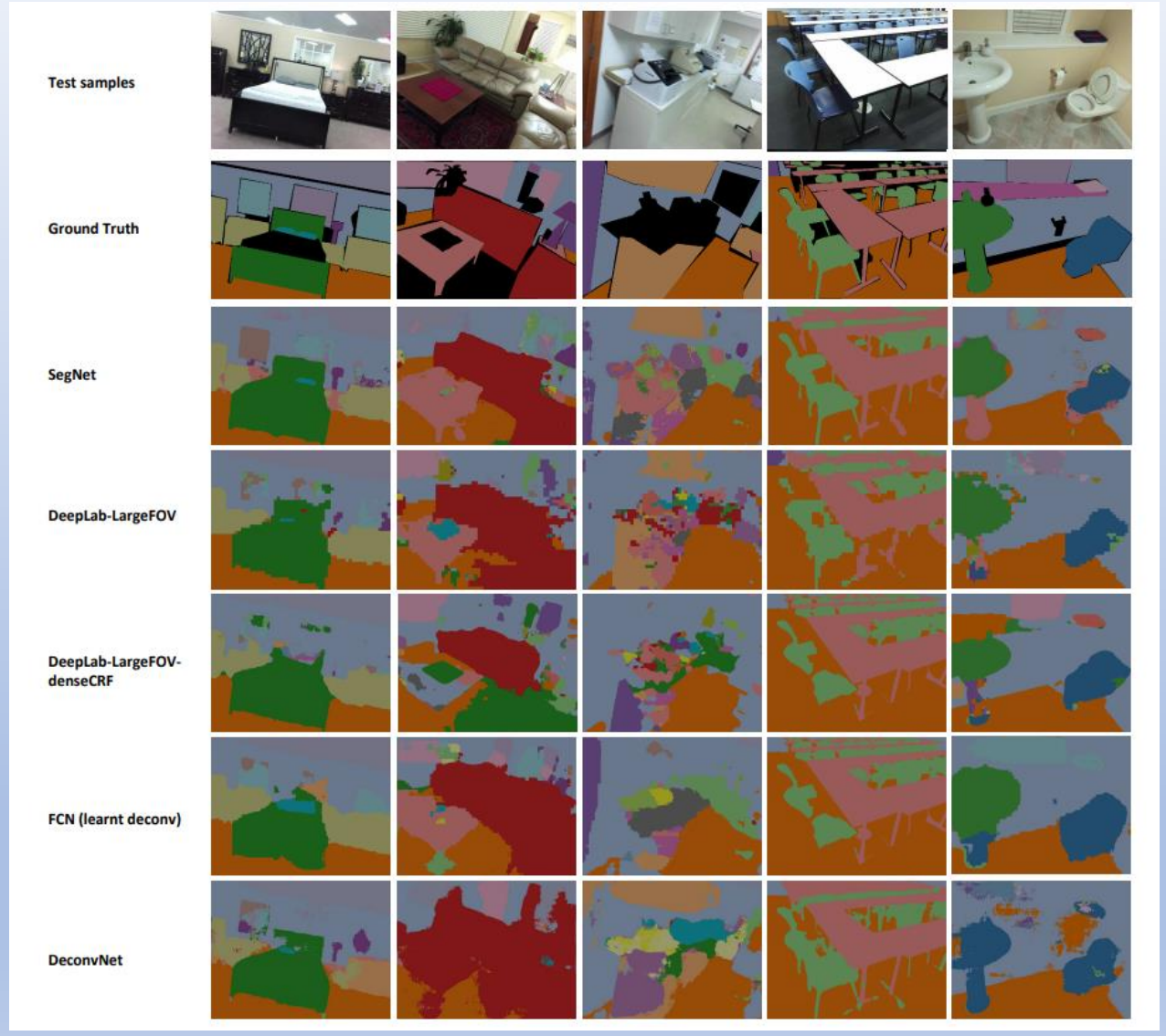

图6.在SUNRGB-D上的结果（定性分析）

5.评价

Network	Forward pass(ms)	Backward pass(ms)	GPU training memory (MB)	GPU inference memory (MB)	Model size (MB)
SegNet	422.50	488.71	6803	**1052**	117
DeepLab-LargeFOV [3]	**110.06**	**160.73**	**5618**	1993	83
FCN (learnt deconv) [2]	317.09	484.11	9735	1806	539
DeconvNet [4]	474.65	602.15	9731	1872	877

图7. 训练时间、内存比较

- 资源消耗和结果准确性上取得较好平衡（网络设计定位）
- 使用maxpooling index改善了边界划分清晰度
- 减少了训练量