中山大學
SUN YAT-SEN UNIVERSITY

中文题目

The English Title
Title...

学位申请人　陈胜杰
专 业 名 称　工程硕士 (软件工程)
答 辩 时 间　May 13, 2018

目录

引言

选题背景与意义

是什么，为什么？

国内外研究现状

主流方法是什么？

本文的工作

我提出的方法是什么，有什么不同？

图像语义分割问题的定义

图像语义分割 (Semantic Image Segmentation) 是根据物体类别把图像分成若干个有意义的区域, 并为不同的区域标注其所属标签的视觉任务。

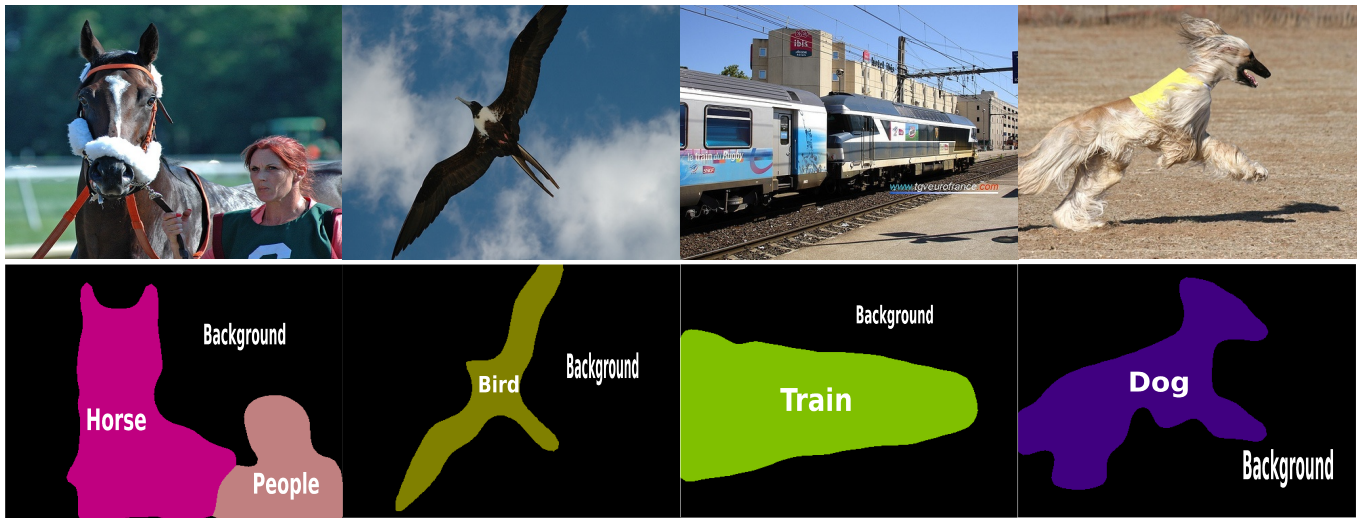


Figure 1: 本文模型在 VOC 2012 验证集上的语义图像分割例子

选题的背景

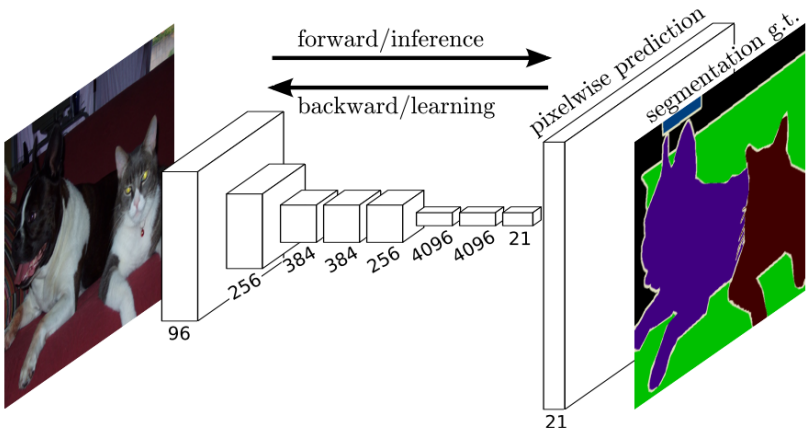
- 手工设计特征的局限性 (SIFT, HOG)
- 深度学习技术的兴起
  - 基于 GPU 的并行化计算
  - 大型训练集的标注
- 大数据与智能时代正在来临

选题的意义

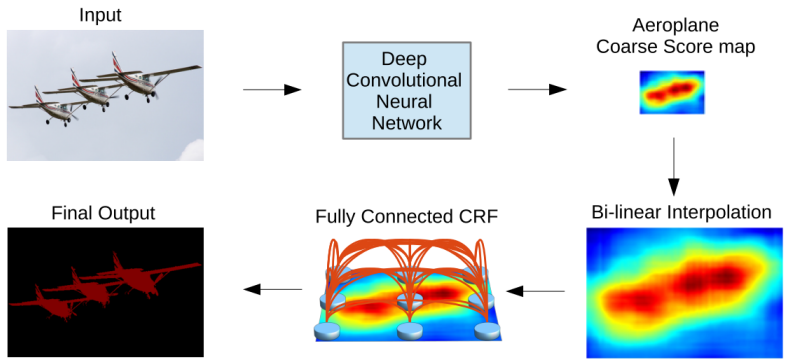
† 图像语义分割解决了图像里面有什么，物体的形状与位置的问题

- 图像检索
- 图像编辑
- 现实增强
- 机器导航

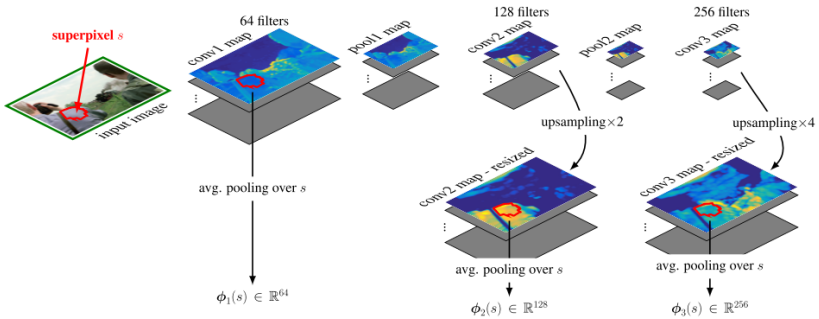
主流方法中的代表性工作



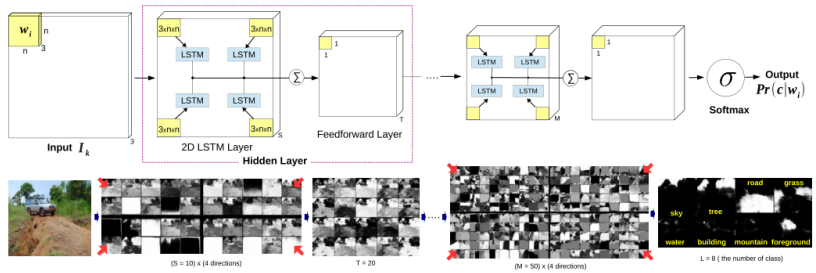
a. 全卷积网络 [Long et al, CVPR 2015]



b. 全卷积网络 + 概率图模型 [Chen et al, ICLR 2015]



c. 卷积网络 + 高低层次特征融合 [Mostajabi et al, CVPR 2015]



d. 二维长短记忆网络 [Byeon et al, CVPR, 2015]

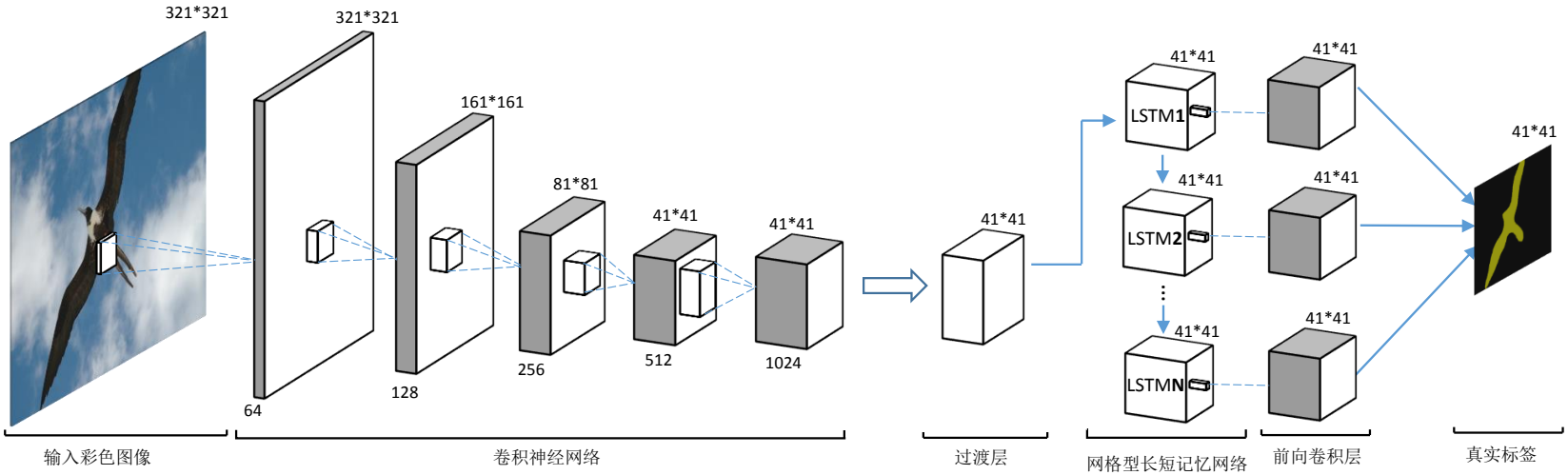


Figure 2: 网络整体结构图

目标与思路

- 充分利用全卷积网络强大的特征学习能力
- 借助长短记忆网络对特征整体与局部建模的良好能力
- 使用随机梯度下降法进行端到端的训练
- 在主流数据集验证模型有效性

深度神经网络

前馈神经网络

传统的人工神经网络

卷积神经网络

目前最为流行的，广泛应用于视觉任务的神经网络

长短记忆网络

与卷积网络相比，更适用于处理时序信号

- 有向无环图的结构
- 输入层 (数据特征)
- 隐含层 (映射后的特征)
- 输出层 (预测结果)
- 反向传播算法 (训练方法)

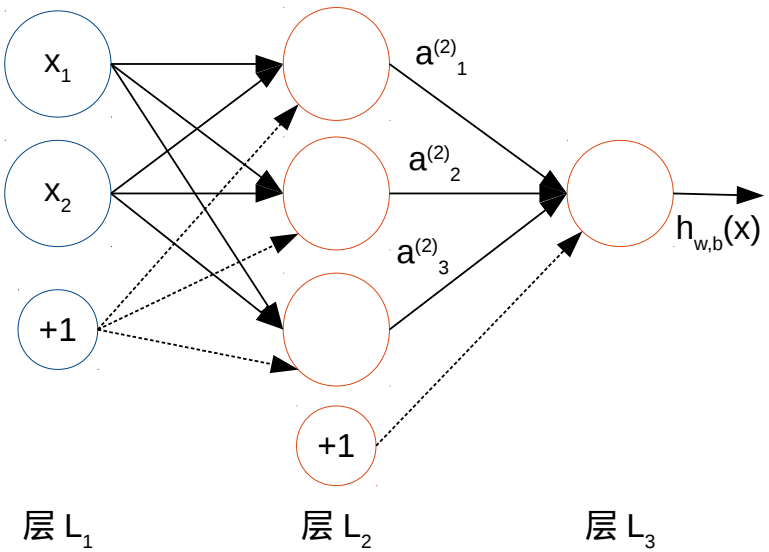


Figure 3: 前馈神经网络模型示意图

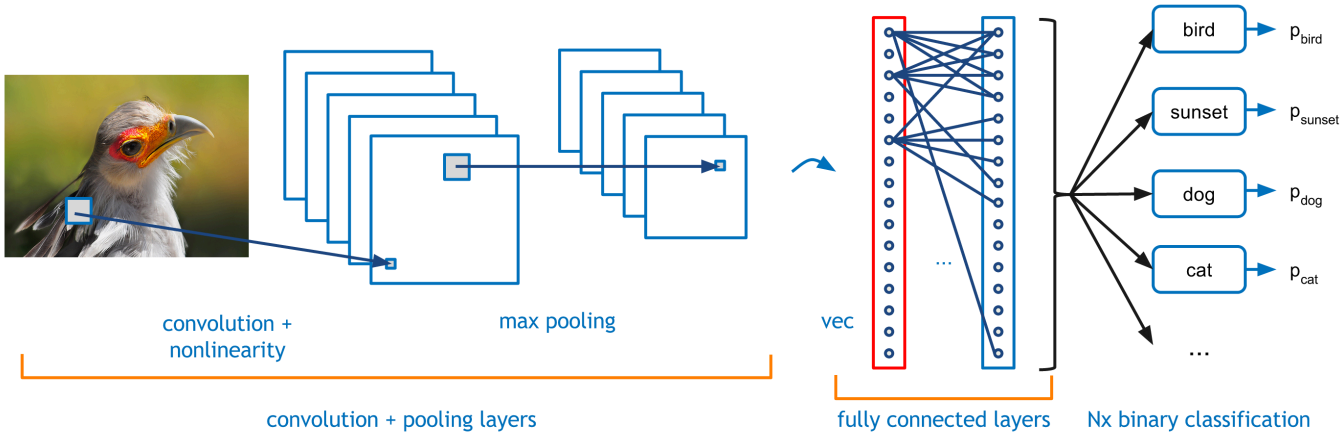
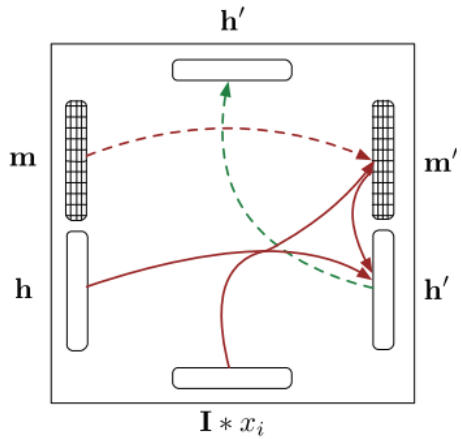


Figure 4: 卷积网络模型示意图

与前馈神经网络的区别

- 直接作用于二维图像，无需特征设计阶段
- 卷积层，池化层
- 局部感知域，权重共享



$$
\begin{aligned}
\mathbf{g}^u &= \delta(\mathbf{W}^u * \mathbf{H}) \\
\mathbf{g}^f &= \delta(\mathbf{W}^f * \mathbf{H}) \\
\mathbf{g}^o &= \delta(\mathbf{W}^o * \mathbf{H}) \\
\mathbf{g}^c &= \tanh(\mathbf{W}^c * \mathbf{H}) \\
\mathbf{m}' &= \mathbf{g}^f \odot \mathbf{m} + \mathbf{g}^u \odot \mathbf{g}^c \\
\mathbf{h}' &= \tanh(\mathbf{g}^o \odot \mathbf{m}') \\
\mathbf{H} &= \begin{bmatrix} I * \mathbf{x}_i \\ \mathbf{h} \end{bmatrix}
\end{aligned} \tag{1}
$$

Figure 5: 长短记忆网络区块示意图

缩写形式

$$(\mathbf{h}', \mathbf{m}') = \mathrm{LSTM}(\mathbf{H}, \mathbf{m}, \mathbf{W})$$

其中 $\mathbf{W}$ 包含了四个门权值矩阵 $\mathbf{W}^u, \mathbf{W}^f, \mathbf{W}^o, \mathbf{W}^c$。

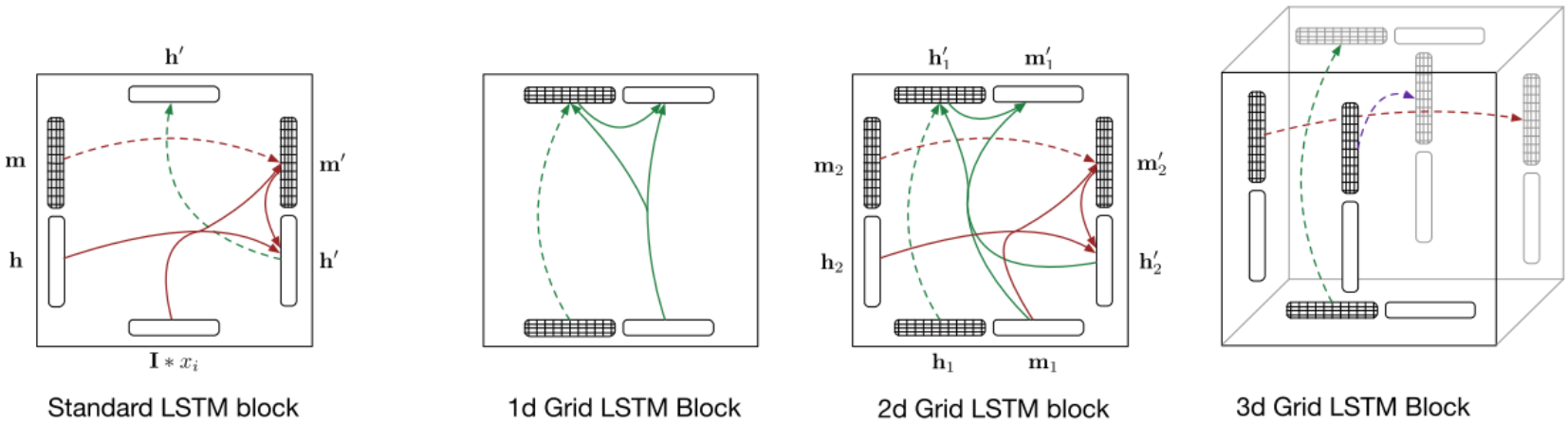


Figure 6: 网格型长短记忆网络区块示意图 [Kalchbrenner et al, Grid LSTM, ICLR 2016]

网格型长短记忆网络更新过程

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_i \\ \vdots \\ \mathbf{h}_N \end{bmatrix} \tag{2}$$

$$\begin{aligned} (\mathbf{h}'_1, \mathbf{m}'_1) &= \text{LSTM}(\mathbf{H}, \mathbf{m}_1, \mathbf{W}_1) \\ &\vdots \\ (\mathbf{h}'_N, \mathbf{m}'_N) &= \text{LSTM}(\mathbf{H}, \mathbf{m}_N, \mathbf{W}_N) \end{aligned} \tag{3}$$

网络整体结构

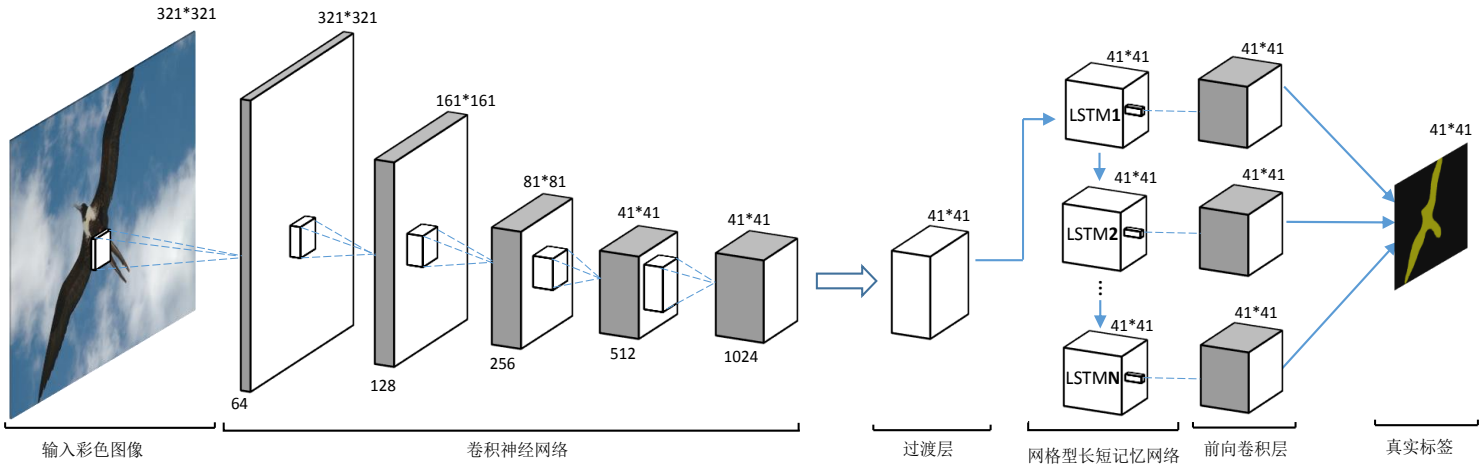


Figure 7: 网络整体结构图

- 四个组成部分: **卷积网络部分**，过渡层，**网格型长短记忆网络部分**，前向卷积层
- 核心思想：在卷积网络后堆叠多层网格型长短记忆层

- 基于 VGG_{16} 模型[1]，含有 16 层卷积层
- 使用了“孔算法”，在不损失精度的情况下将模型参数减少了 6.5 倍[2]

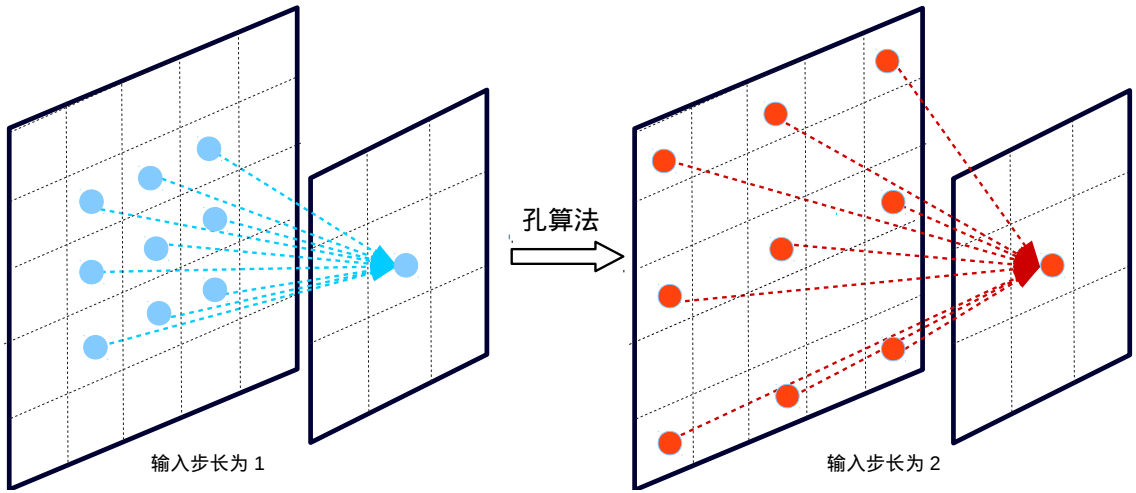
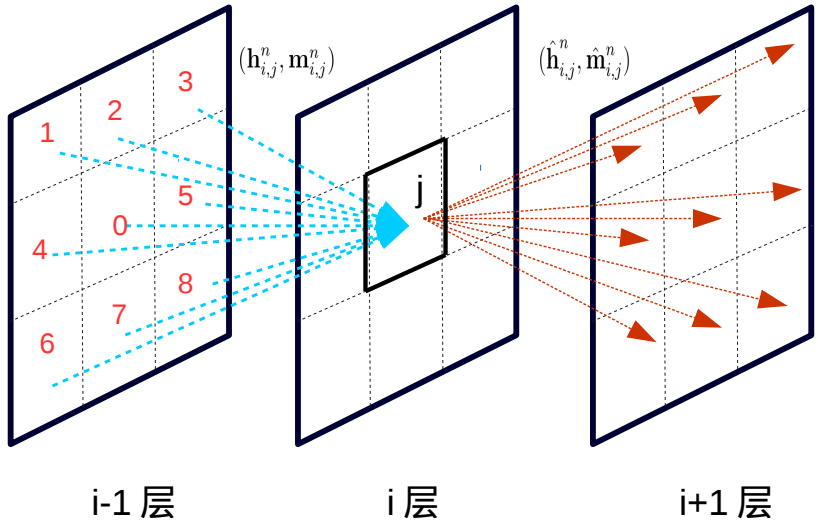


Figure 8: ” 孔算法” 示意图

[1] Simonyan & Zissermanet, Very deep Convolutional Networks For Large-scale Image Recognition, ICLR 2015

[2] Chen et al, DeepLab-LargeFOV, ICLR 2015



$$
\begin{aligned}
(\hat{\mathbf{h}}^0_{i,j}, \hat{\mathbf{m}}^0_{i,j}) &= \mathrm{LSTM}(\mathbf{H}_{i,j}, \mathbf{m}^0_{i,j}, \mathbf{W}_i) \\
(\hat{\mathbf{h}}^1_{i,j}, \hat{\mathbf{m}}^1_{i,j}) &= \mathrm{LSTM}(\mathbf{H}_{i,j}, \mathbf{m}^1_{i,j}, \mathbf{W}_i) \\
&\vdots \\
(\hat{\mathbf{h}}^N_{i,j}, \hat{\mathbf{m}}^N_{i,j}) &= \mathrm{LSTM}(\mathbf{H}_{i,j}, \mathbf{m}^N_{i,j}, \mathbf{W}_i) \\
\mathbf{H}_{i,j} &= [\mathbf{h}^0_{i,j}\ \mathbf{h}^1_{i,j}\ \ldots\ \mathbf{h}^N_{i,j}]^T
\end{aligned} \tag{4}
$$

Figure 9: 九维网格型长短记忆网络层之间的通信示意图

九维网格型长短记忆网络

- 每个位置的预测会受到上一层相邻八邻域特征的影响
- 随着层数的堆叠，每一位置将会有更大的感知域。
- 网格型长短记忆网络的层数通过实验来确定

实验与结果

数据集

准确率度量方式

VOC 2012 实验结果

SIFT FLOW 实验结果

Pascal VOC 2012 & SIFT FLOW 数据集

Figure 10: VOC 2012 数据集：10582 张训练样本，1464 张验证样本和 1456 张测试样本，21 个类别

Figure 11: SIFT FLOW 数据集：2488 张训练样本，200 张测试样本，33 个类别

图像预处理

- 训练时图像均缩放为 321*321
- 随机选取训练图像，随机取镜像, 数据白化

准确率度量方式

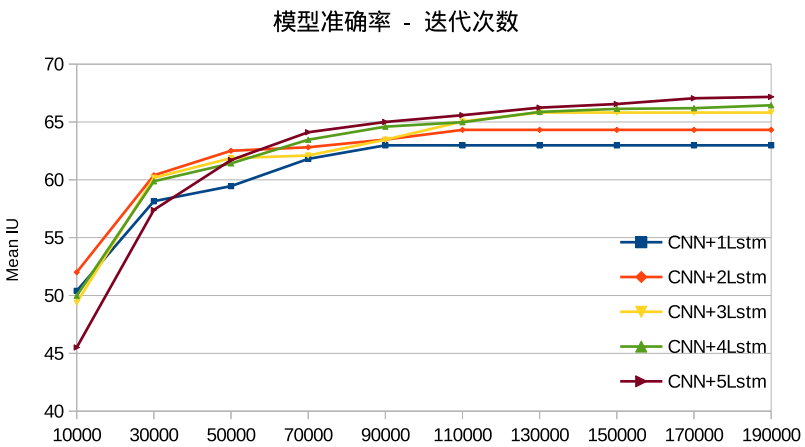
设 n_{ij} 为真实值属于类别 i 但被分类为类别 j 的像素个数，n_{cl} 表示有多少种不同的标签，$t_i = \sum_{j=1}^{n_{cl}} n_{ij}$ 为所有真实值为类别 i 的像素个数。

$$\text{像素准确率} = \sum_{i=1}^{n_{cl}} n_{ii} / \sum_{i=1}^{n_{cl}} t_i$$

$$\text{平均像素准确率} = \frac{1}{n_{cl}} \sum_{i=1}^{n_{cl}} (n_{ii}/t_i) \tag{5}$$

$$\text{Mean IU} = \frac{1}{n_{cl}} \sum_{i=1}^{n_{cl}} \frac{n_{ii}}{t_i + \sum_j^{n_{cl}} n_{ji} - n_{ii}}$$

网格型长短记忆网络层数的选择



- 在一定范围内增加长短记忆层数可以有效提高网络效果
- 增加了 5 层网格型长短记忆网络之后，网络效果提升了 **7.5%**

Figure 12: 网格型长短记忆层数的不同对网络分割效果的影响方法

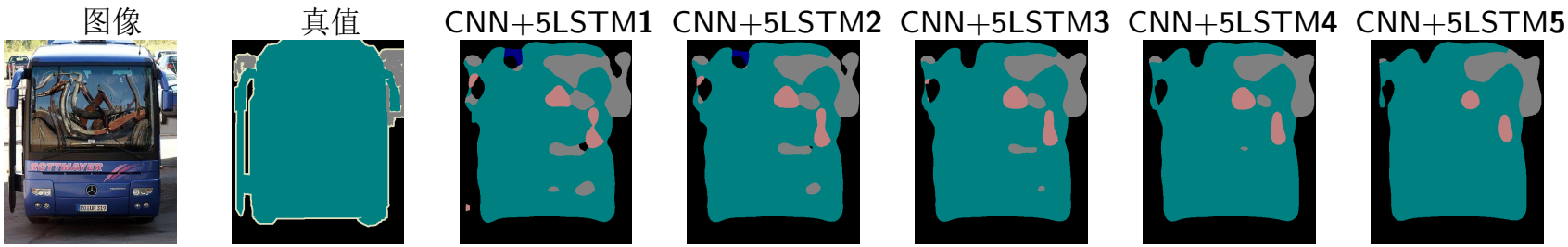


Figure 13: 网格型长短记忆网络层数增加对输出改善作用的可视化

与其它工作的定量对比

Method	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	shep	sofa	train	tv	mIoU.
SDS[3]	63.3	25.7	63.0	39.8	59.2	70.9	61.4	54.9	16.8	45.0	48.2	50.5	51.0	57.7	63.3	31.8	58.7	31.2	55.7	48.5	51.6
FCN-8s[4]	76.8	34.2	68.9	49.4	60.3	75.3	74.7	77.6	21.4	62.5	46.8	71.8	63.9	76.5	73.9	45.2	72.4	37.4	70.9	55.1	62.2
TTI-zoomout-16[5]	**81.9**	35.1	**78.2**	**57.4**	56.5	80.5	74.0	79.8	22.4	69.6	53.7	74.0	**76.0**	76.6	68.8	44.3	70.2	40.2	68.9	55.3	64.4
DeepLab-CRF[6]	78.4	33.1	**78.2**	55.6	**65.3**	81.3	75.5	78.6	**25.3**	69.2	52.7	**75.2**	69.0	**79.1**	**77.6**	**54.7**	78.3	45.1	73.3	56.2	66.4
CNN+**5**LSTM	80.2	**35.3**	74.1	54.4	64.4	**87.3**	**81.1**	**80.6**	22.7	**73.6**	**58.8**	73.9	73.7	78.7	77.4	50.2	**80.0**	**47.9**	**76.5**	**63.1**	**67.9**

Table 1: 模型在 VOC2012 测试集上的结果。

结论

† 模型比其他方法有更高的精确度，验证了模型的有效性

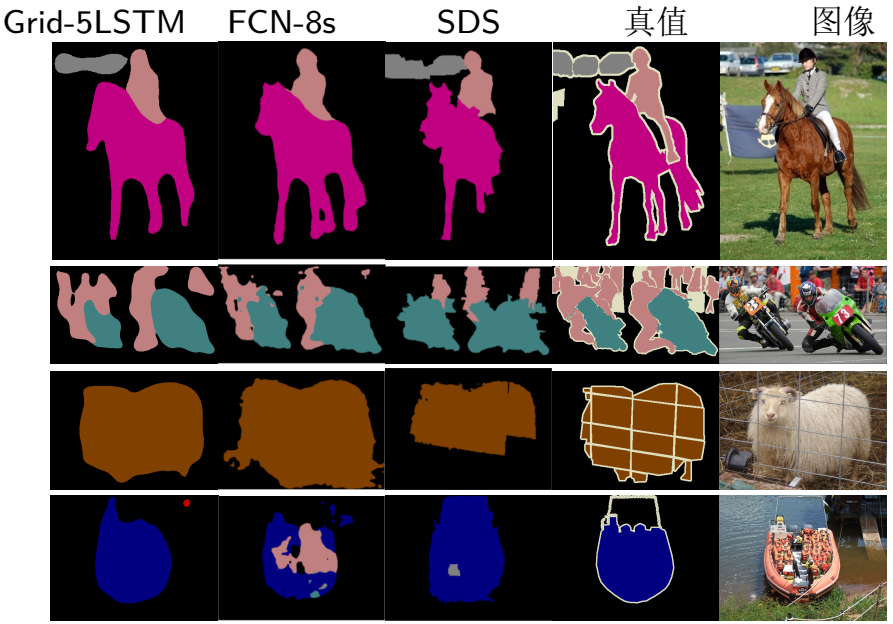
[3] Simultaneous Detection and Segmentation, ECCV 2014

[4] Fully convolutional networks for semantic segmentation, CVPR 2015

[5] Feedforward semantic segmentation with zoom-out features, CVPR 2015

[6] Semantic image segmentation with deep convolutional nets and fully connected crfs, ICLR 2015

与其它工作的定性对比



(a) 本文模型效果与其他工作的定性对比

(b) 其中第一行为图像，第二行为真值，第三行为 TTI-zoomout-16, 第四行为 DeepLab-CRF，第五行是 Grid-5LSTM 的结果

Figure 14: Grid-5LSTM 与其它模型在 VOC 2012 验证集上的定性比较

一些分割失败的例子

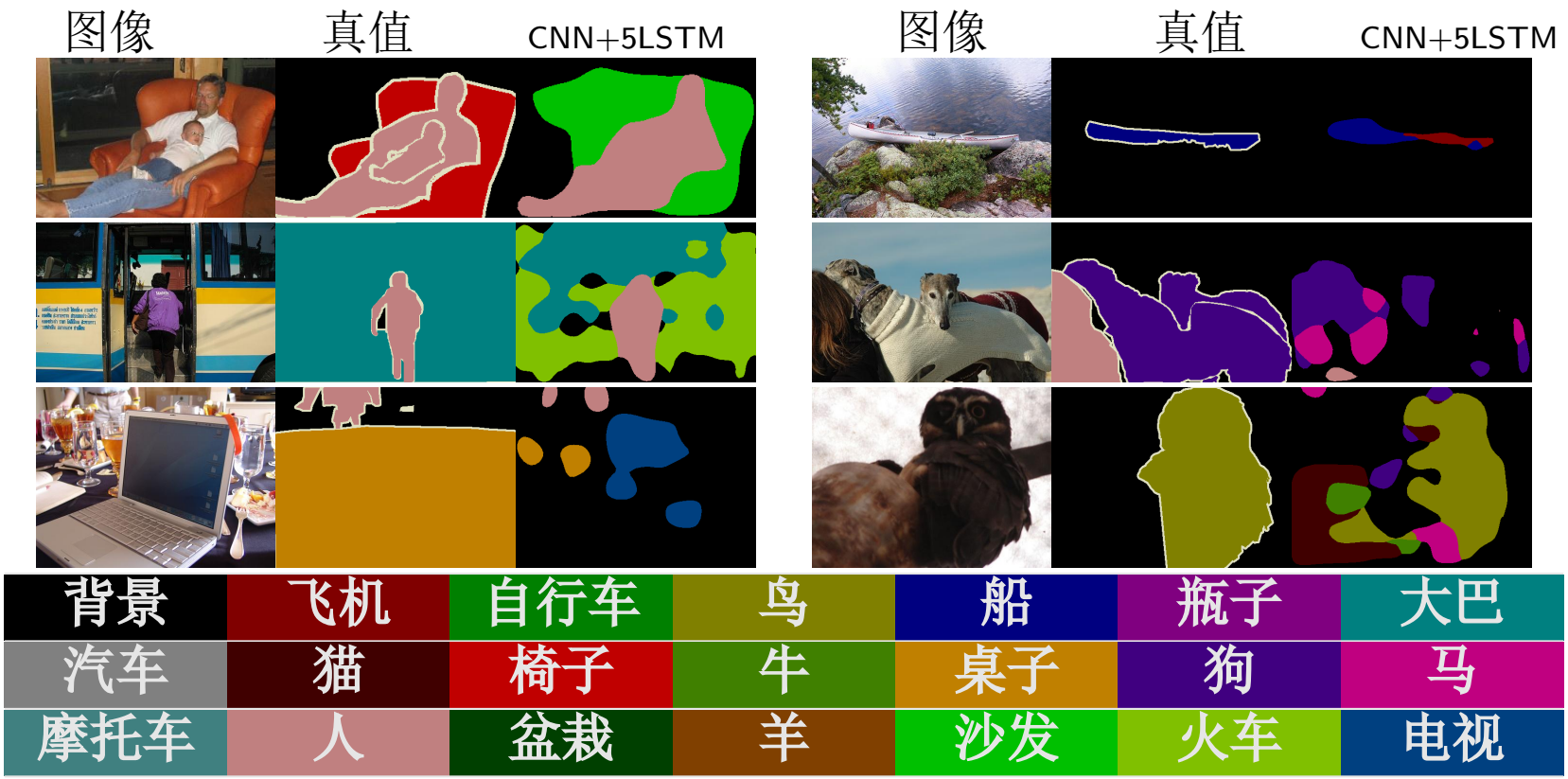


Figure 15: 一些 CNN+5LSTM 分类错误的例子

Method	Pixel Acc.	Mean Acc.	Mean IU.
Liu et al.[7]	76.7	-	-
Tighe et al.[8]	78.6	39.2	-
FCN-16s[9]	85.2	**51.7**	39.5
Deeplab-LargeFOV[10]	85.6	51.2	39.7
Grid-5LSTM	**86.2**	51.0	**41.2**

Table 2: 模型在 SIFT FLOW 上的结果。Tighe 等人的方法是用 SVM+MRF，Deeplab-LargeFOV 的结果是通过公开的源码实验得到的

[7] Sift flow: Dense correspondence across scenes and its applications, PAMI 2011

[8] Finding things: Image parsing with regions and per-exemplar detectors, CVPR 2013

[9] Fully convolutional networks for semantic segmentation, CVPR 2015

[10] emantic image segmentation with deep convolutional nets and fully connected crfs, ICLR 2015

工作总结

- 本文的模型结合了卷积网络的特征学习能力与长短记忆网络对整体局部建模的能力，相比于全卷积网络，大幅度地提高了了模型性能
- 大量的对比实验与结果分析证明了模型的有效性

展望

- 模型性能：提高网络的深度来学习更高层次的特征，提高模型效果 (He et al. ResNet, CVPR 2016)
- 模型大小：通过裁剪网络冗余部分 (Han et al. Deep Compression, ICLR 2016 Best Paper) 或使用二值网络减少模型参数 (Courbariaux et al. Binaryconnect, NIPS 2015)
- 训练数据：使用无监督或弱监督的方式训练网络 (Papandreou et al. Weakly-and semi-supervised learning, ICCV 2015)

致谢

感谢每一个帮助过我的人

- 首先要感谢的是我的指导老师的悉心指导
- 感谢师兄师姐、同学的帮助
- 感谢家人的支持
- 感谢答辩委员会的聆听和指导

Q & A

Questions?

Thank you!