计算机视觉课程报告

学号： 10185102144

姓名： 董辰尧

专业名称： 计算机科学与技术

学生年级： 2018 级本科生

课程性质： 专业选修

研修时间： 2020～2021 学年第 2 学期

计算机科学与技术学院

2021 年 6 月

课程内容统计

- 请自评你的项目完成情况，在表中相应位置划√。

课程学习自我评价

内容\评价	阅读文献 0—5 篇	阅读文献 5—10 篇	阅读文献 10 篇以上	代码 实现
第 13 章　深度学习基础	√			√

总体课程学习情况自我评价

完成情况	尚 未 完 成	基 本 完成	较 好 完 成	圆 满 完 成
总体情况		√		

第 13 章 智能图像分割

一、这一章学习中你的工作

这一章我阅读了智能图像分割的相关文献，并且尝试实现了其中的算法。

二、查阅文献清单

[1] Li H , Xiong P , Fan H , et al. DFANet: Deep Feature Aggregation for

Real-Time Semantic Segmentation[J]. 2019.

[2] Lin Z , Zhang Z , Chen L Z , et al. Interactive Image Segmentation With First Click Attention[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020.

[3] Zhang S , Liew J H , Wei Y , et al. Interactive Object Segmentation With Inside-Outside Guidance[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020.

三、文献解读

1. 文献 1

（a）文献名：Li, H. , Xiong, P. , Fan, H. , & Sun, J. . (2019). Dfanet: deep feature aggregation for real-time semantic segmentation.

（b）主要创新思想

本论文介绍了一种高效的 CNN 结构：DFANet 用于在资源限制下的语义分割。

我们提出的网络从单个轻量级骨干网开始，分别通过子网和子级级联聚合判别特征。基于多尺度特征传播，DFANet 大大减少了参数的数量，但仍然获得了足够的感受野，提高了模型学习能力，在速度和分割性能之间取得了平衡。

(c) 主要原理剖析及说明

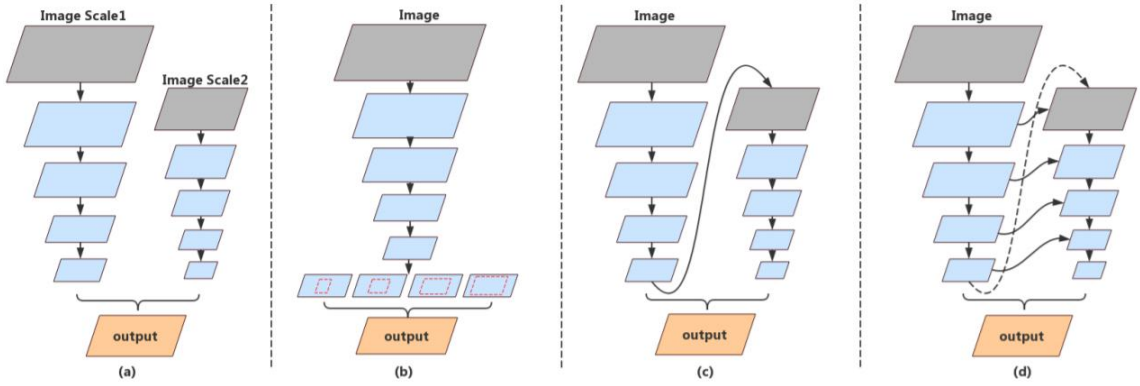


Figure 2. Structure Comparison. From left to right: (a) Multi-branch. (b) Spatial pyramid pooling. (c) Feature reuse in network level. (d) Feature reuse in stage level. As a comparison, the proposed feature reuse methods enrich features with high-level context in another aspect.

用了空间金字塔和多尺度池化两个方案的融合

(d) 主要实验结果（现有原文章中的）

Model	InputSize	FLOPs	Params	Time(ms)	Frame(fps)	mIoU(%)
PSPNet[34]	713 × 713	412.2G	250.8M	1288	0.78	81.2
DeepLab[4]	512 × 1024	457.8G	262.1M	4000	0.25	63.1
SegNet[1]	640 × 360	286G	29.5M	16	16.7	57
ENet[22]	640 × 360	3.8G	0.4M	**7**	**135.4**	57
SQ[25]	1024 × 2048	270G	-	60	16.7	59.8
CRF-RNN[35]	512 × 1024	-	-	700	1.4	62.5
FCN-8S[19]	512 × 1024	136.2G	-	500	2	63.1
FRRN[24]	512 × 1024	235G	-	469	0.25	71.8
ICNet[33]	1024 × 2048	28.3G	26.5M	33	30.3	69.5
TwoColumn[27]	512 × 1024	57.2G	-	68	14.7	72.9
BiSeNet1[29]	768 × 1536	14.8G	5.8M	13	72.3	68.4
BiSeNet2[29]	768 × 1536	55.3G	49M	21	45.7	**74.7**
DFANet A	1024 × 1024	**3.4G**	7.8M	**10**	**100**	71.3
DFANet B	1024 × 1024	**2.1G**	4.8M	**8**	**120**	67.1
DFANet A'	512 × 1024	**1.7G**	7.8M	**6**	**160**	70.3

Table 5. Speed analysis on Cityscapes *test* dataset. "-" indicates that the corresponding result is not provided by the methods.

2. 文献 2

（a）文献名： Lin, Z. , Zhang, Z. , Chen, L. Z. , Cheng, M. M. , & Lu, S. P. . (2020). Interactive Image Segmentation With First Click Attention. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE.

（b）主要创新思想

在本文中，我们演示了第一次点击对于提供目标对象的位置和主体信息的关键作用。为了更好地利用这一特性，提出了一个名为 First Click Attention Network (FCA-Net)的深度框架。我们提出了考虑用户注释的点击损失和结构完整性策略，有助于交互式分割任务

(c) 主要原理剖析及说明

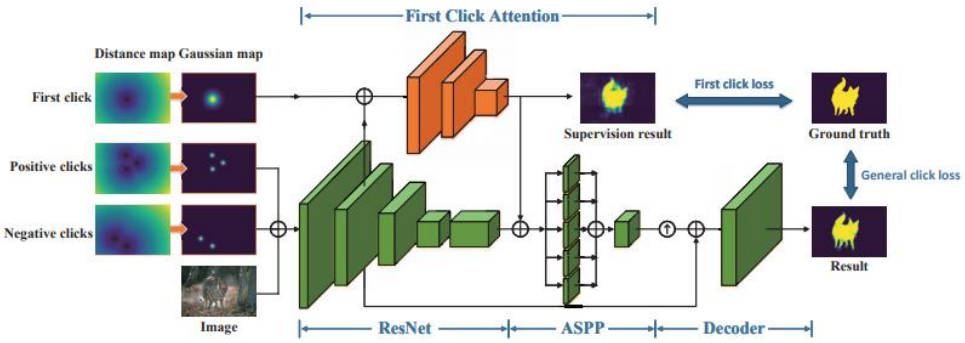


图 2. FCA-Net 的整体结果。绿色部分显示了基础网络，包括主干网络、空洞卷积池化金字塔模块和解码器模块。橙色部分显示初始交互点注意力模块。符号 "⊕" 和 "↑" 分别表示拼接和上采样操作。在节. 3.1中详细描述了细节。

Basic Segmentation Network

我们采用了通用的 FCN 架构，其特定的结构类似于 DeepLab v3+。

它包含三个部分:骨干网、Atrous 空间金字塔池(ASPP)模块和解码器模块。

为了在交互分割中捕获多尺度对象，我们在 ResNet101 的最后阶段也采用了扩大卷积，而不是采用 stride 作为 2。

主干网的输入是将 RGB 图像与两个带正点和带负点的高斯图连接起来。高斯映射是根据欧氏距离映射计算的。

First Click Attention Module

为了利用第一次点击的引导信息，我们设计了一个简单的模块和基本的分割网络。

它以低级特征 F1 和以第一次点击为中心的高斯映射 Mf 作为输入。

将连接的特征(f1 ⊕M F)输入至 6 个 3 × 3 的卷积层。

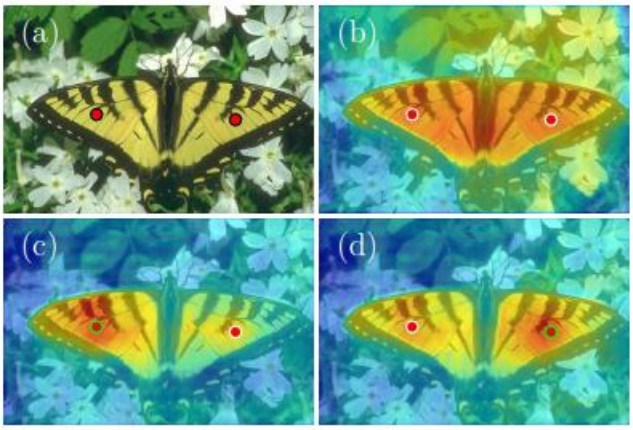


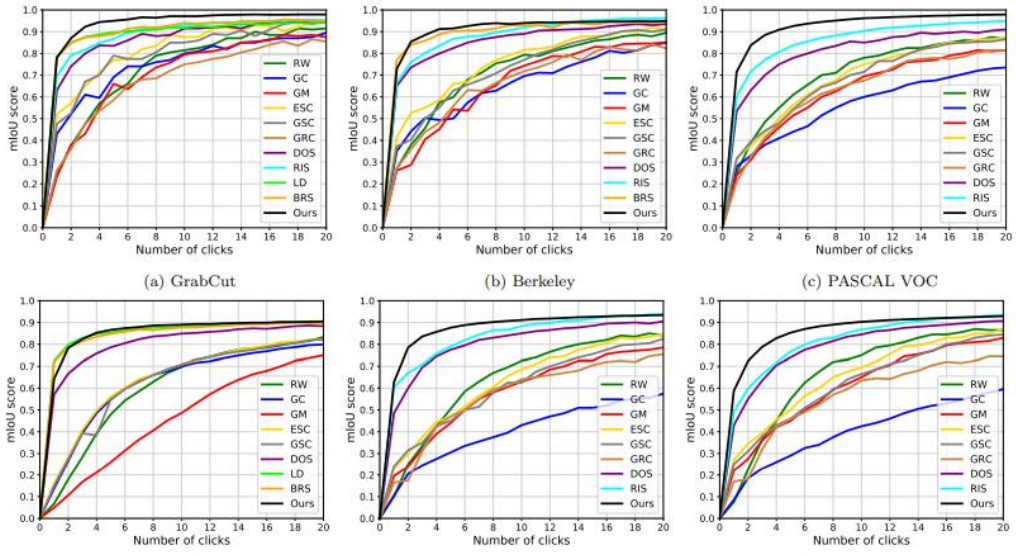
图 3. 初始交互点注意力的可视化。(b) 是有 FCA 模块的预测图; (c) 和 (d) 是初始交互点模块作用在不同位置的预测图。

Click Loss

对于监督最后预测图的损失函数，我们提出了一个考虑了所有交互点的全局交互点损失 ($\mathcal{L}_g$)，它的计算如下：

$$\mathcal{L}_g = \frac{1}{N} \sum_{p \in \mathcal{G}} \left(\widehat{w}_p \cdot \ell \left(p \right) \right). \tag{4}$$

(d) 主要实验结果（现有原文章中的）



(a) GrabCut (b) Berkeley (c) PASCAL VOC

(d) DAVIS (e) MSCOCO (seen) (f) MSCOCO (unseen)

图 5. FCA-Net 和其它 10 个方法在 5 个数据集 6 个子集上的交互点 vs. 平均交并比 (NoC-mIoU) 曲线。

Method	GrabCut @90%	Berkeley @90%	PASCAL VOC @85%	DAVIS @90%	MSCOCO (seen)@85%	MSCOCO (unseen)@85%
GC [5]$_{ICCV01}$	11.10	14.33	15.06	17.41	18.67	17.80
GRC [44]$_{POG05}$	16.74	18.25	14.56	N/A	17.40	17.34
RW [17]$_{PAMI06}$	12.30	14.02	11.37	18.31	13.91	11.53
GM [3]$_{IJCV09}$	12.44	15.96	14.75	19.50	17.32	14.86
ESC [18]$_{CVPR10}$	8.52	12.11	11.79	17.70	13.90	11.63
GSC [18]$_{CVPR10}$	8.38	12.57	11.73	17.52	14.37	12.45
DOS [46]$_{CVPR16}$	6.04	8.65	6.88	12.58	8.31	7.82
RIS [30]$_{ICCV17}$	5.00	6.03	5.12	N/A	5.98	6.44
LD [29]$_{CVPR18}$	4.79	N/A	N/A	9.57	N/A	N/A
BRS [24]$_{CVPR19}$	3.60	5.08	N/A	8.24	N/A	N/A
CMG [36]$_{CVPR19}$	3.58	5.60	3.62	N/A	5.40	6.10
FCA-Net	2.24	4.23	2.98	8.05	4.49	5.54
FCA-Net (SIS)	2.14	4.19	2.96	7.90	4.45	5.33
FCA-Net*	2.16	3.92	2.79	7.64	4.34	5.36
FCA-Net* (SIS)	2.08	3.92	2.69	7.57	4.08	5.01

表 2. 个数据集 6 个子集上的平均交互点数（mNoC）对比。SIS 表示使用了结构完整性策略进行后处理。FCA-Net* 表示该模型使用 Res2Net [15]作为主干网络。

3. 文献 3

（a）文献名：Zhang, S. , Liew, J. H. , Wei, Y. , Wei, S. , & Zhao, Y. . (2020). Interactive Object Segmentation With Inside-Outside Guidance. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE.

（b）主要创新思想

1. 提出了一种内部-外部指导（IOG）方法。具体地说，我们利用在对象中心附近单击的一个内部点和包围目标对象的紧密边界框的对称角位置（左上角和

右下角或右上角和左下角）处的两个外部点。这将导致总共一次前景点击和四次背景点击进行分割。

2. IOG 有四个优点：1）两个外部点可以帮助消除来自其他对象或背景的干扰；2）内部点有助于消除边界框内不相关的区域；3）内部和外部点易于识别，减少了最先进的糊精标记引起的混淆一些极端的例子；4）我们的方法自然支持额外的点击注释，以便进一步修正。

(c) 主要原理剖析及说明

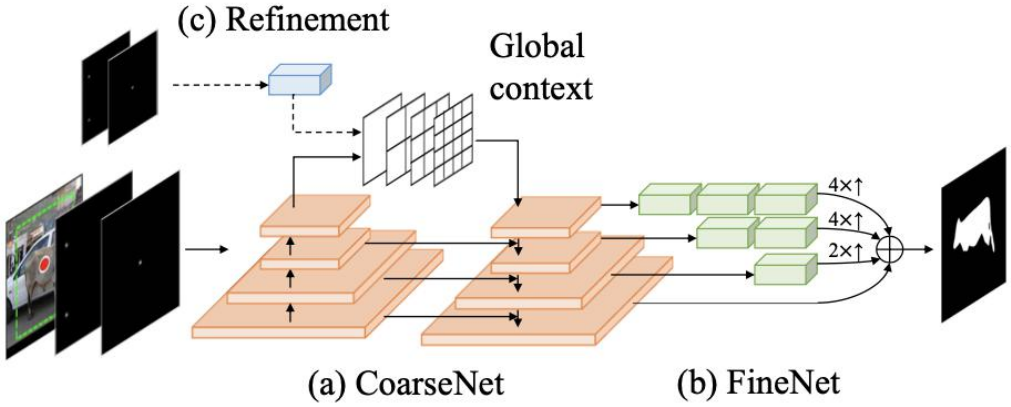


Figure 4. **Network Architecture.** (a)-(b) Our segmentation network adopts a coarse-to-fine structure similar to [14], augmented with a pyramid scene parsing (PSP) module [68] for aggregating global contextual information. (c) We also append a lightweight branch before the PSP module to accept the additional clicks input for interactive refinement

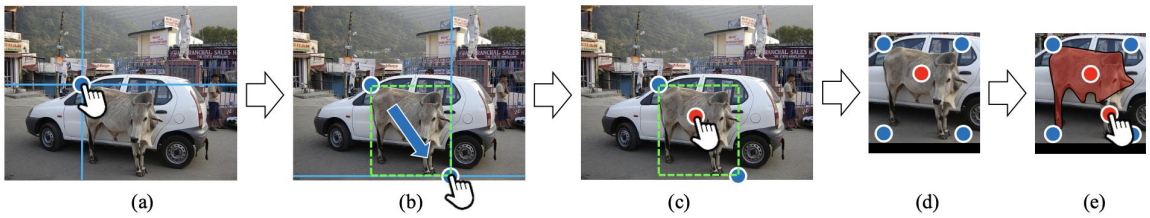


Figure 2. **Inside-Outside guidance.** (a) The vertical and horizontal guide lines are used to assist the user in clicking on the corner of an imaginary box enclosing the object. (b) A box is generated on-the-fly when the user moves the cursor. (c) An interior click is placed around the object center. (d) The box is relaxed by several pixels before cropping to include context. The interior click (red) with four exterior clicks (two clicked corners and two automatically inferred ones) (blue) constitute our Inside-Outside guidance that encode the foreground

2. 交互方式:一个内部点，两个边界点((either top-left and bottom-right or top-right and bottom-left)

3. Segmentation Network coarse-to-fine

(d) 主要实验结果（现有原文章中的）

Methods	Number of Clicks		IoU(%) @ 4 clicks	
	PASCAL@85%	GrabCut@90%	PASCAL	GrabCut
Graph cut [5]	> 20	> 20	41.1	59.3
Random walker [23]	16.1	15	55.1	56.9
Geodesic matting [2]	> 20	> 20	45.9	55.6
iFCN [66]	8.7	7.5	75.2	84.0
RIS-Net [38]	5.7	6	80.7	85.0
DEXTR [46]	4	4	91.5	94.4
Li et al. [37]	-	4.79	-	-
ITIS [45]	3.4	5.7	-	-
FCTSFN [28]	4.58	3.76	-	-
IOG-ResNet101 (**ours**)	3	3	93.2*	96.3*
IOG-ResNet101 (**ours**)	4	4	94.4	96.9

Table 1. Comparison with the state-of-the-art methods on PASCAL and GrabCut in terms of the number of clicks to reach a certain IoU and in terms of quality at 4 clicks. *denotes the IoU of our IOG given only 3 clicks.

三、本章学习小结

这一章我阅读了智能图像分割的相关文献，了解到很多智能分割领域的前沿知识，大概了解了整个领域的研究情况。