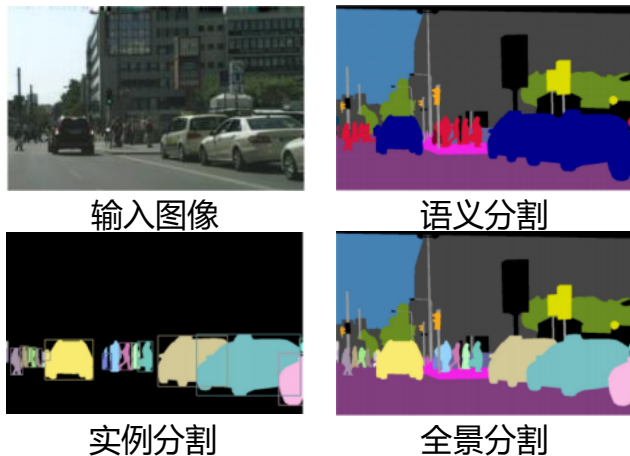


语义分割算法技术分享

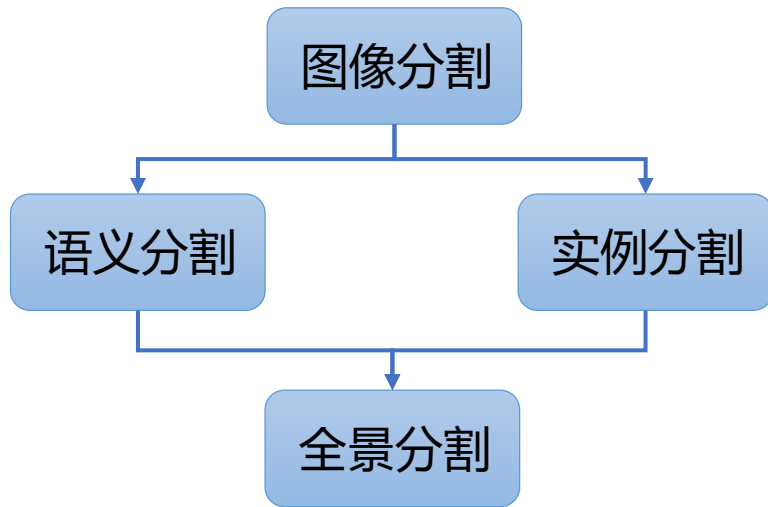
2021-5-28

前言：什么是语义分割？



语义分割的难点：

- 准确的语义
- 精细且准确的分割边缘
- 小目标物体分割
- 良好的泛化性



语义分割的评测指标：

- 采用mIoU进行综合性评测。
- 采用Boundary Displacement Error (BDE) 评测分割边界精度。
- 采用 F_{score} 度量给定阈值下准确率和召回率的谐波平均值。
- 采用Pixel Accuracy (PA)计算正确分类的像素数量与其总数之间的比率

● 数据集成果

- ✓ 统一的、规范化的数据格式
- ✓ 详实的业务数据特性分析
- ✓ 揭露工业质检难点

● 算法成果

- ✓ 复现7种语义分割SOTA算法
- ✓ 三个流行学术数据集上与mmseg的性能对齐
- ✓ 总结与比较各SOTA算法在长尾数据、小目标物体、轻微缺陷等特定数据上的性能

● 代码成果

- ✓ 自研模块化搭建的基于Pytorch的语义分割框架freshSegmentation
- ✓ 支持多种主流SOTA语义分割算法
- ✓ 支持多种学术及工业业务数据集
- ✓ 支持DDP分布式训练

剖数据，析问题：

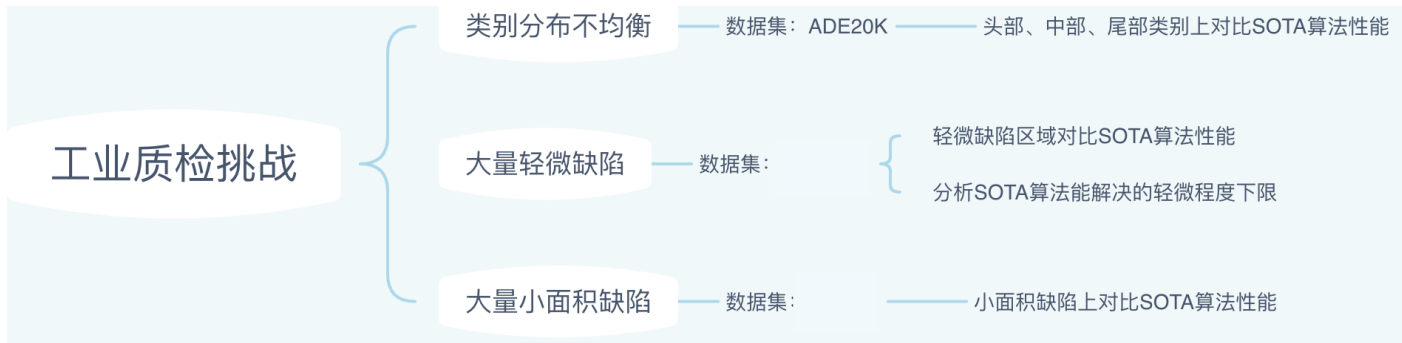
● 剖数据：

➤ 学术数据集

- ✓ Cityscapes
- ✓ ADE20K
- ✓ PASCAL VOC 2012
- ✓ PASCAL-Context

➤ 业务数据集

● 析问题：



1.1 数据集 - 总览

1.1.1 学术数据集

Cityscapes

- 训练集2975, 验证集500, 测试集1525
- 30类语义分割
- 8类实例分割



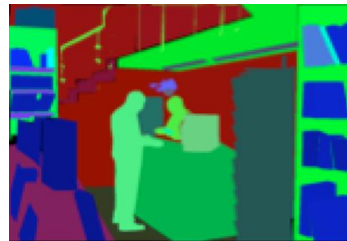
PASCAL VOC 2012

- 训练集1464, 验证集1449, 测试集1456
- 20类语义分割



ADE20K

- 训练集25574, 验证集2000
- 150类语义分割



PASCAL-Context

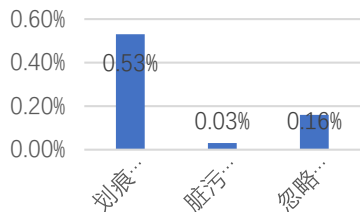
- 训练集10103, 测试集9647张图
- 540类语义分割



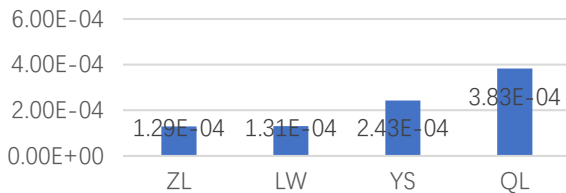
1.2 工业数据集特性——类别不均衡

1、各类别像素占比不均衡

各类别缺陷像素级占比



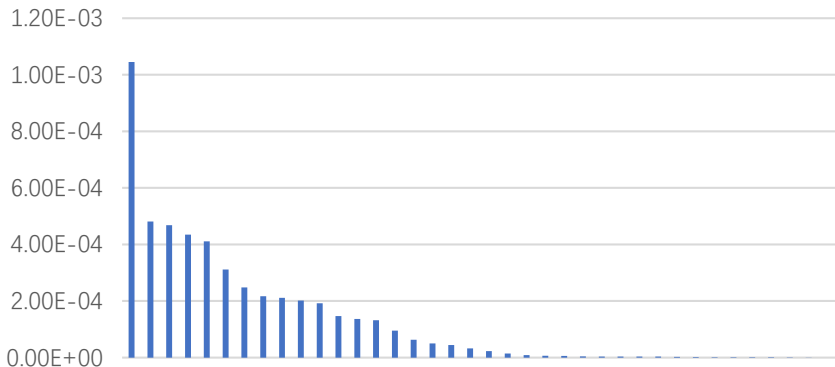
各类别缺陷像素级占比



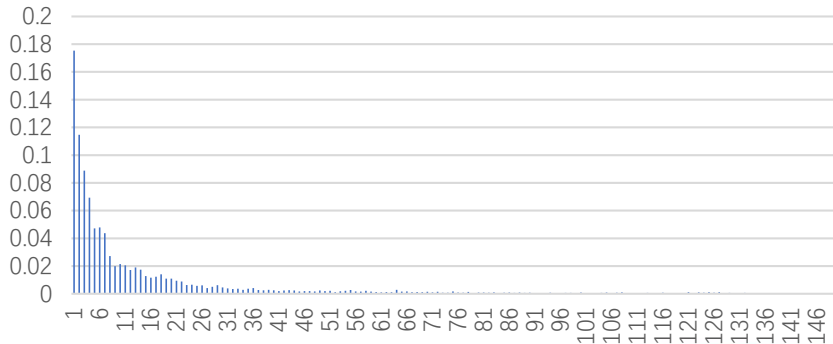
当前主流SOTA算法在头部、中部、尾部类别性能如何？

各业务数据集普遍存在类别分布不均衡的问题。

各类别缺陷像素级占比



各类别缺陷像素级占比

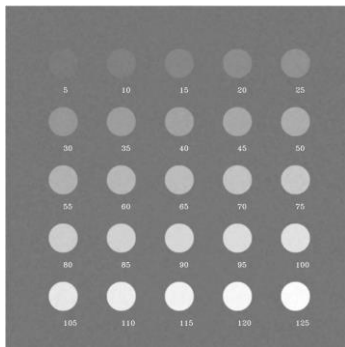
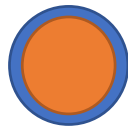


ADE20K

1.2 工业数据集特性——大量轻微缺陷

2、存在大量轻微缺陷

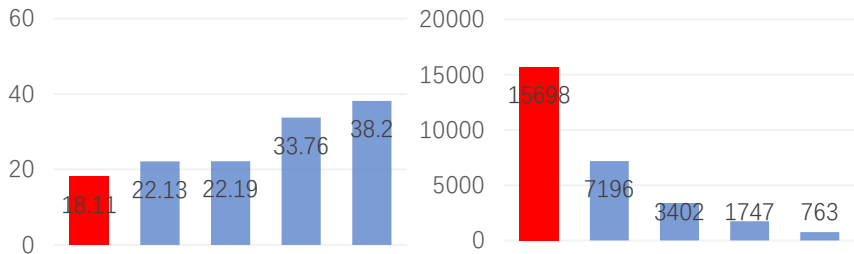
定义：缺陷轻微值 = | 缺陷区域RGB均值 - 缺陷周围区域RGB均值 |



就轻微缺陷而言，SY V23数据集最具代表性

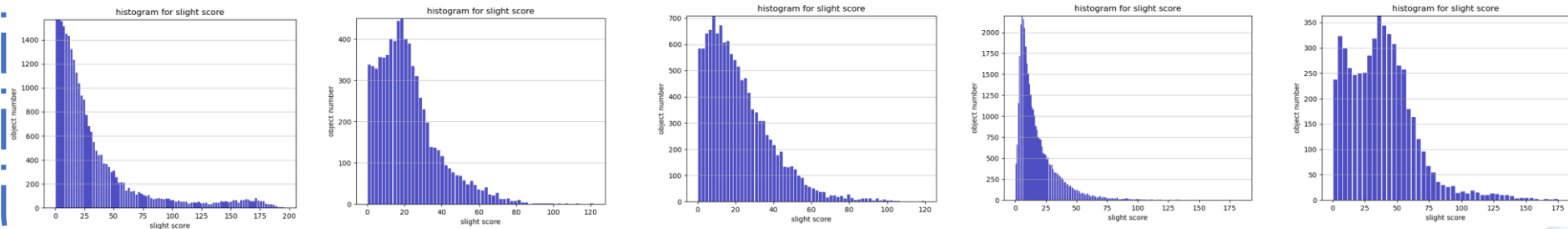
缺陷轻微值均值

轻微缺陷个数



当前主流
SOTA算法
在能否准
确定位到
轻微缺陷？

缺陷轻微值直方图——工业质检中存在大量轻微缺陷

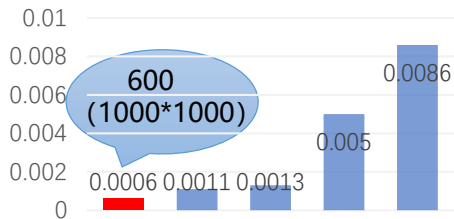


1.2 工业数据集特性——大量小面积缺陷

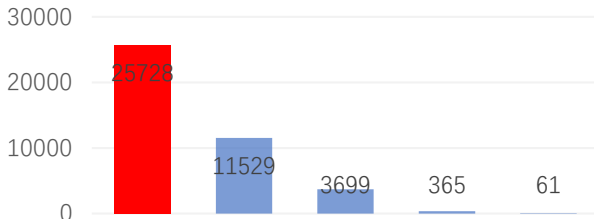
3、存在大量小面积缺陷

缺陷相对面积 = 缺陷区域面积/图片面积

相对面积均值



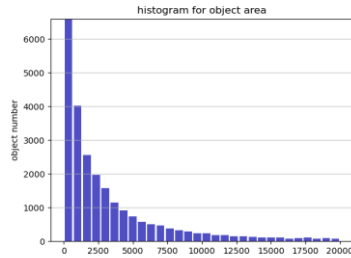
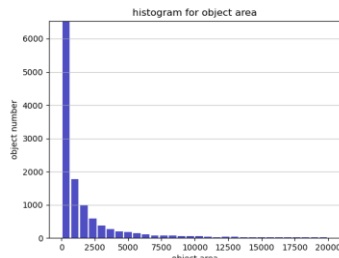
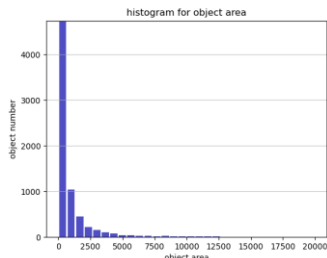
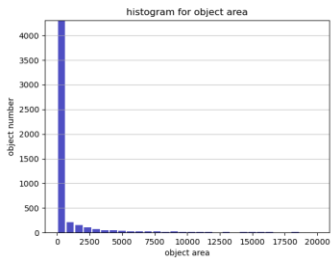
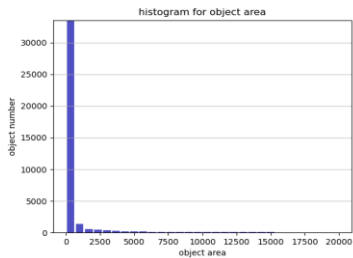
小目标数目



就轻微缺陷而言，SY V23数据集最具代表性

当前主流SOTA算法在在小目标上性能如何？

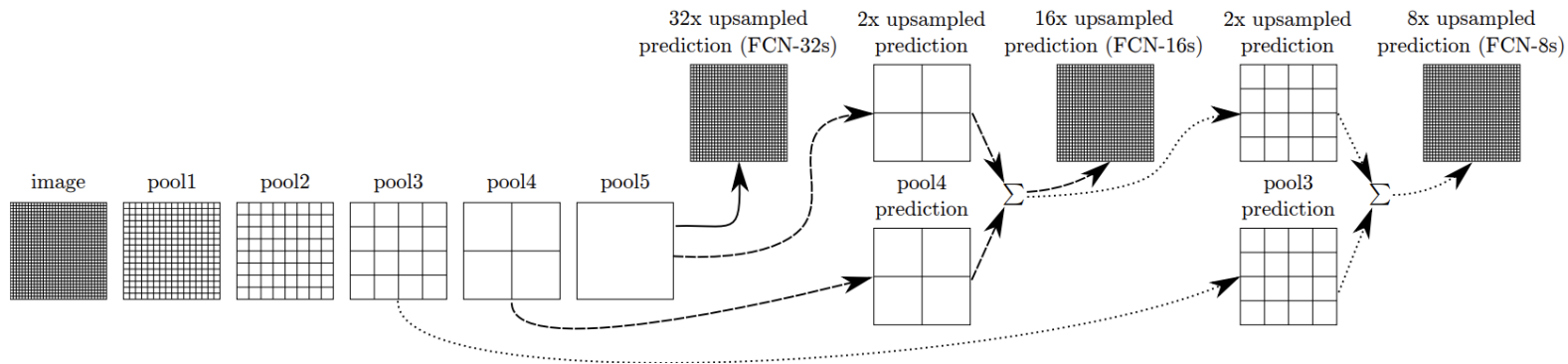
缺陷（绝对）面积直方图——工业质检中存在大量小面积缺陷



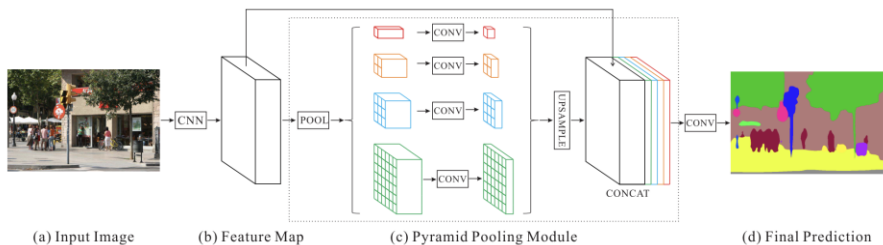
2.1 算法 – 总览

方法	简介	小目标	边缘精细	轻微缺陷	长尾——中部类别	长尾——尾部类别
PSPNet	提出金字塔池化模块聚合上下文信息				✓	✓
CCNet	提出recurrent criss-cross attention module，对GPU显存友好				✓	✓
DANet	双重注意力网络以集成局部特征和全局依赖				✓	✓
FCN	图像级到像素级开山之作		✓	✓		
Deeplab V3	提出ASPP模块挖掘不同尺度卷积特征			✓		
FPN	增加极小的计算量的情况下，处理好多尺度变化	✓	✓	✓		
Upernet	多任务框架	✓	✓			
HRNet	高分辨率	✓	✓			
Gated-SCNN	Shape Steam专注学习边缘		✓			

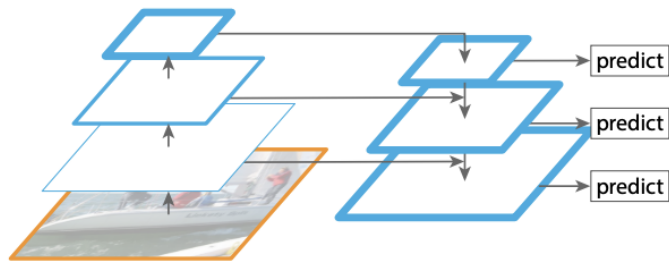
2.1 算法 – 总览



(a) FCN

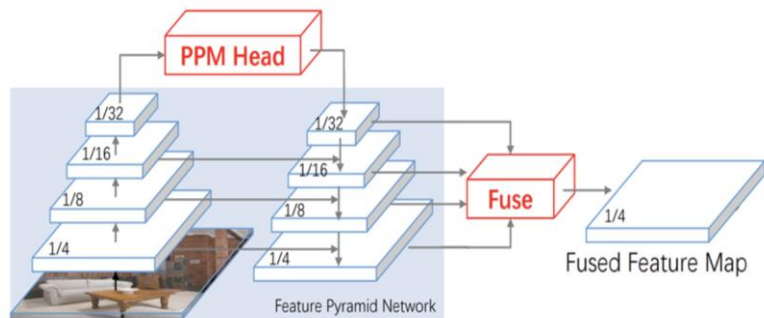


(b) PSPNet

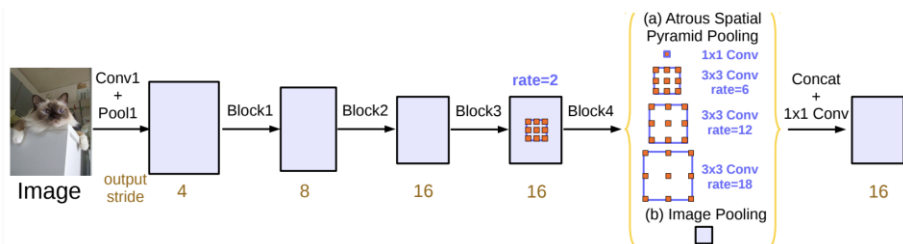


(c) FPN

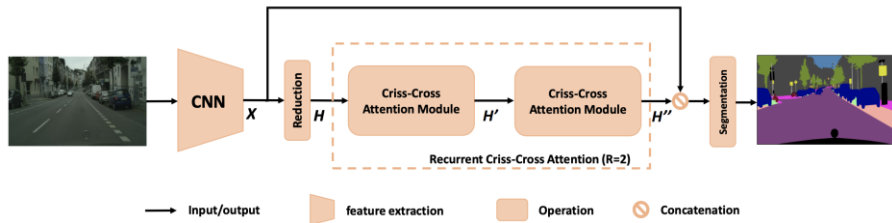
2.1 算法 – 总览



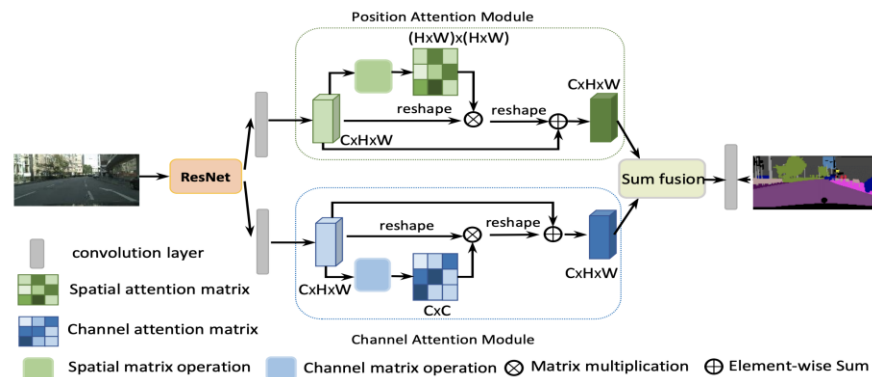
(d) UperNet



(e) Deeplab V3



(f) CCNet



(g) DANet

2.2 算法-小目标 (算法: FPN, UperNet)

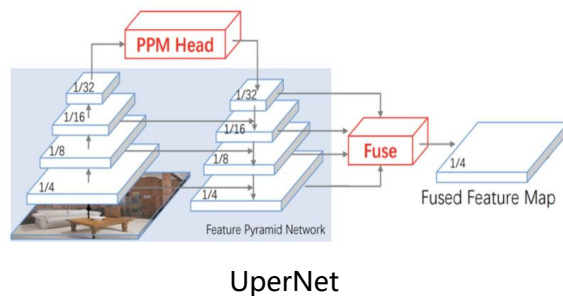
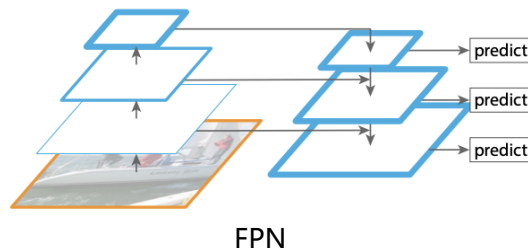
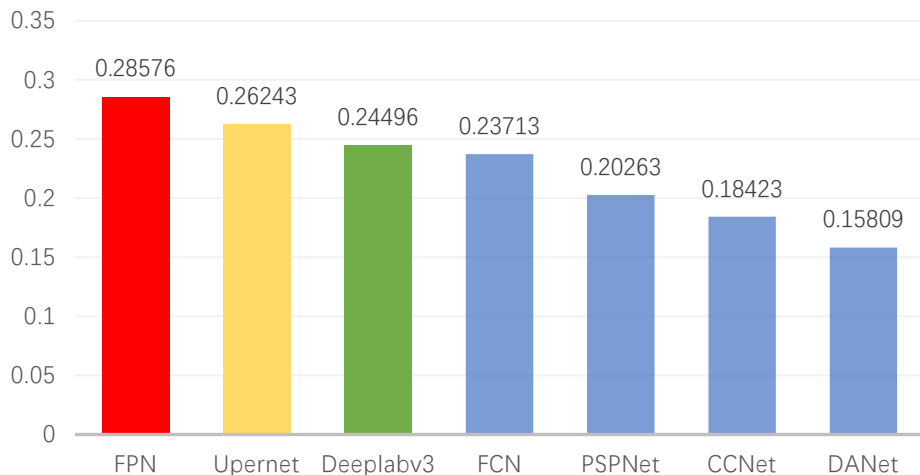
结论: 多尺度的特征金字塔类型网络结构, 在小目标分割上性能最优。

分析:

- ✓ 充分利用底层信息有助于提升小目标分割效果;
- ✓ 利用空洞卷积减少池化操作有助于提升小目标分割效果;
- ✓ 在网络高层, 小目标相关信息大量丢失, 因此增加全局上下文信息对小目标分割无益。

数据集: SY V23

mIoU on Small Object



2.3 算法 – Boundary （算法：UperNet, FPN, FCN）

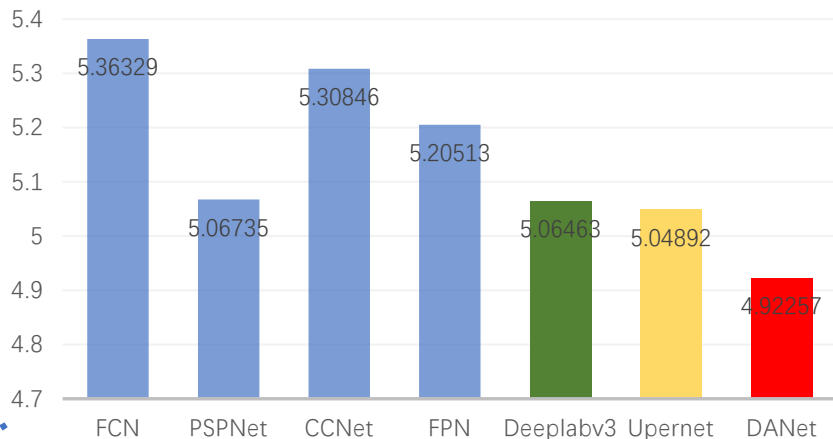
结论：

- ✓ 结合网络底层细节特征有助于提升边缘分割精度；
- ✓ 当任务难度加大时，若目标面积不是特别小，增加全局上下文信息有助于提升分类准确性，BDE指标也更佳。

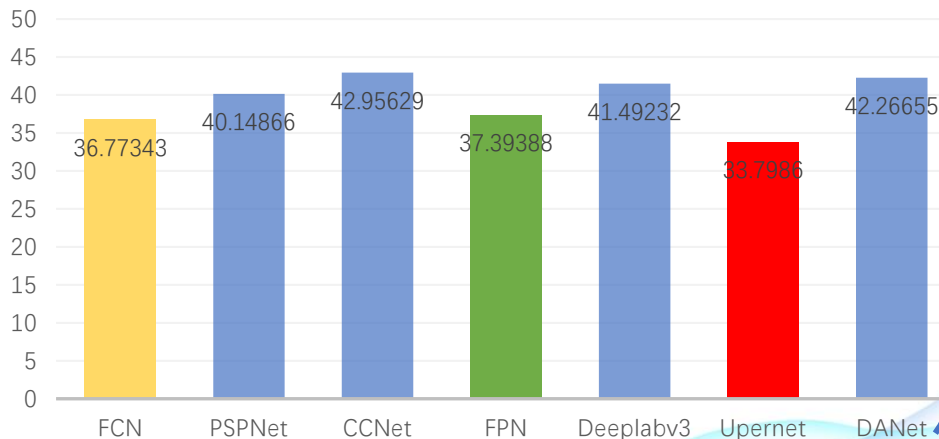
数据集：ADE20K（类别多达150）、SY V23（小目标多、二分类）

$$\text{指标: } \text{BDE}(X, Y) = \frac{1}{2} \left[\frac{1}{N_X} \sum_x \inf_{y \in Y} d(x, y) + \frac{1}{N_Y} \sum_y \inf_{x \in X} d(x, y) \right]$$

BDE on ADE20K



BDE on SY V23



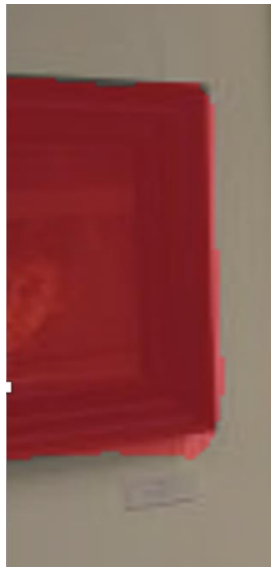
2.2 算法-小目标 (算法: FPN, UperNet)



(a) UperNet



(b) FPN



(c) PSPNet



(d) CCNet

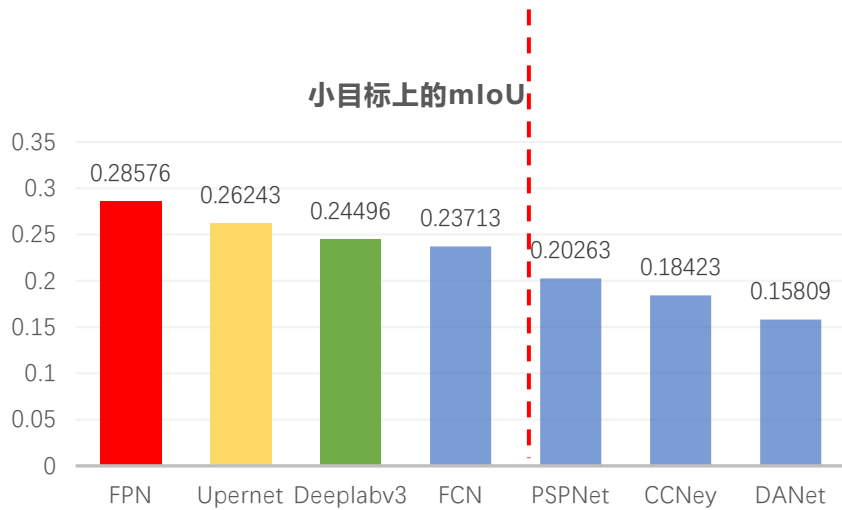
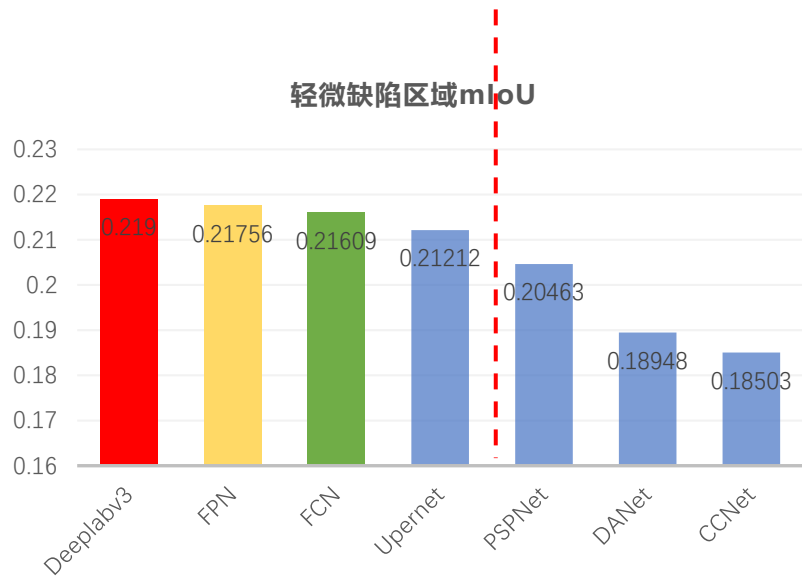
2.4 算法-轻微缺陷（算法：Deeplab V3）

结论：在舜宇数据集的轻微缺陷上，Deeplab V3性能最优——语义与细节兼顾

分析：

- ✓ 舜宇数据集轻微缺陷大部分也是小目标，轻微缺陷性能整体趋势与小目标相似；
- ✓ 轻微缺陷要求网络对缺陷/非缺陷有强语义区分，Deeplab V3引入空洞卷积及ASPP，增大了感受野，对轻微缺陷检测有益。

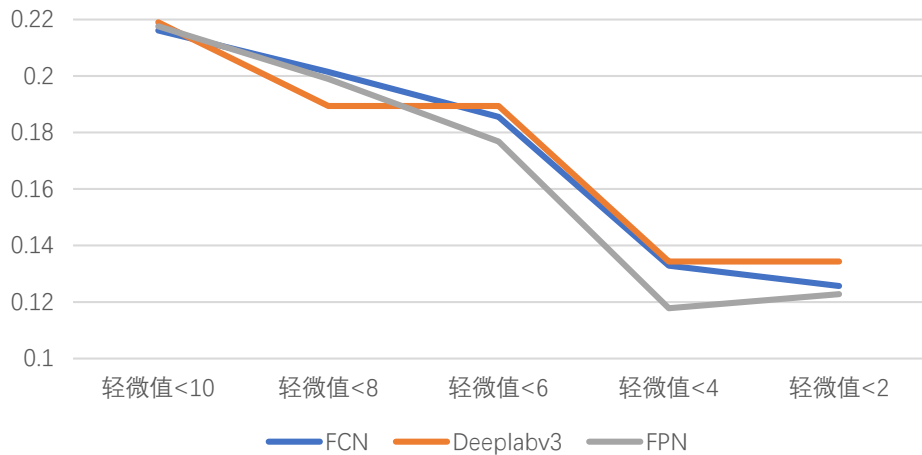
数据集：实际业务数据集上进行实验比较。



2.4 算法-轻微缺陷（算法：Deeplab V3）

轻微缺陷检测极限

mIoU随缺陷轻微程度变化曲线



结论：当缺陷的轻微值低于 4 时，算法性能将急剧下降。

2.5 算法-长尾 (算法: DANet, CCNet, PSPNet)

结论:

- ✓ 长尾数据上对网络的语义泛化能力要求较高, 因此高层上下文信息利用较强的网络更胜一筹。
- ✓ CCNet引入了non-local的全局上下文信息, 学到的语义特征更强, 泛化性更好, 因此在尾部数据表现优异。

数据集: ADE20K

我们按照像素占比排序分出头部数据 (50类)、中部数据 (50类) 和尾部数据 (50类), 分别计算各算法在各类数据上分割性能, 并进行对比。

算法	backbone	val-set miou	iou_major_cate (头部类别)	iou_median_cat e (中部类别)	iou_minor_cate (尾部类别)
FCN	resnet101-s8	0.3923	0.69252	0.38207	0.29843
PSPNet	resnet101-s8	0.4193	0.69919	0.40380	0.32436
CCNet	resnet101-s8	0.42850	0.68877	0.40236	0.32692
FPN	resnet101-s8	0.39279	0.68532	0.38046	0.28649
Deeplabv3	resnet101-s8	0.41115	0.69563	0.38390	0.29184
Upernet	resnet101-s32	0.40072	0.68964	0.39321	0.29548
DANet	resnet101-s8	0.41999	0.70043	0.40318	0.32584

2.6 算法 – 颜色校准

结论：

- ✓ 对于HX这样颜色差异巨大的数据集，颜色校准至关重要
- ✓ 对于SY这样由差异不大的灰度图像构成的数据集，颜色校准并不能发挥其优势和作用

颜色校准：

1. 计算训练集每张图片的 BGR 三个通道均值，对均值做聚类，找到聚类中心，选取一个图片量比较大，并且颜色偏暗的聚类中心，作为参考色，将所有训练图片的 BGR 均值都校准到参考色。
2. 网络训练。
3. 推理阶段，先对图片用参考色做校准，再做后续计算

2.6 算法-颜色校准

舜宇数据集

2、实验结果

训练数据集	是否进行颜色校准	mIoU on V23	Acc on V23	mIoU on 01B	Acc on 01B
V23	否	0.7607	0.9890	-	-
	是	0.7583	0.9889	-	-
01B	否	-	-	0.7627	0.9963
	是	-	-	0.7589	0.9963
V23+01B	否	0.7574	0.9888	0.7851	0.9966
	是	0.7564	0.9891	0.7678	0.9964

结论：对于颜色差异巨大的数据集，颜色校准至关重要；而对于差异不大的灰度图像构成的数据集，颜色校准并不能发挥其优势和作用

2.7 算法 – pretrain model影响

结论:

- ✓ 在ImageNet上 预训练backbone性能最优;
- ✓ HX数据集上 pretrain整个model较仅pretrain backbone性能提升较大。

Model	Dataset	no pretrain	Imagenet pretrain backbone	HX pretrain backbone	HX pretrain model fine-tune
Upernet-Se-Res50-D16	V23	0.68926	0.75109	0.69194	0.71908
Upernet-Se-Res50-D16	01B	0.69332	0.70168	0.66829	0.68801

2.8 算法-超参搜索

采用机智自带的超参搜索功能，对学习率进行搜索。

设置的搜索的学习率区间为[0.0001-0.02]，初始学习率为0.016，设置的搜索次数为10次，下面为实验结果：

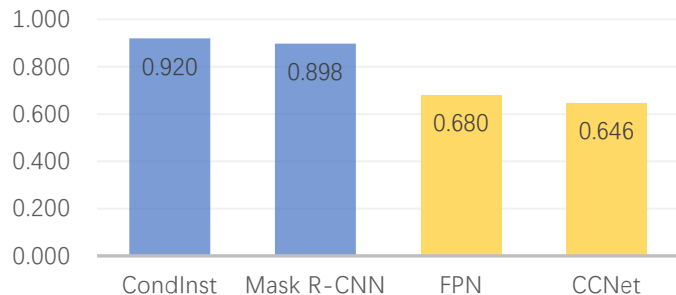
学习率LR	mIoU
0.00525	0.6633
0.00479	0.63594
0.0197	0.71203
0.01678	0.69248
0.016	0.69333
0.00972	0.6896
0.01568	0.71162
0.00789	0.67512

加速效果：搜索空间大小: 1990, 实际搜索次数: 10, 加速比: 199.00
收敛效果：相比人工调参，收敛性能提升2.70%

2.9 算法-对比检测与实例分割

实例分割

FC D52

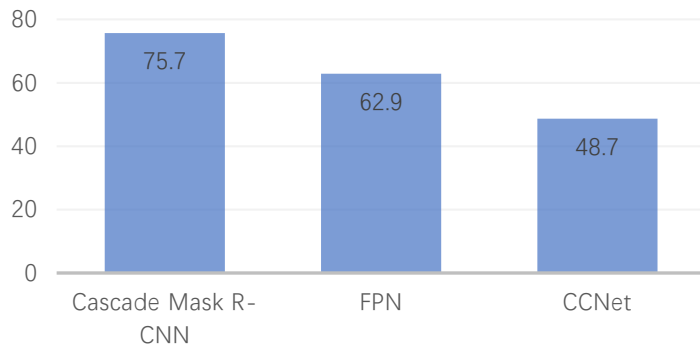


结论:

- ✓ 实例分割优于语义分割
- ✓ 目标检测优于语义分割

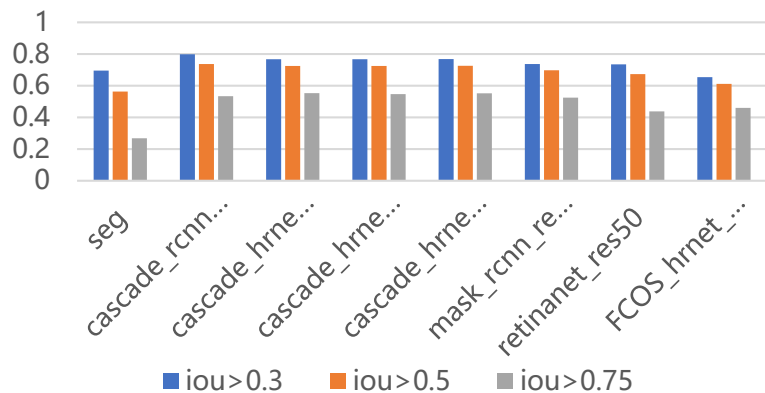


SY V23

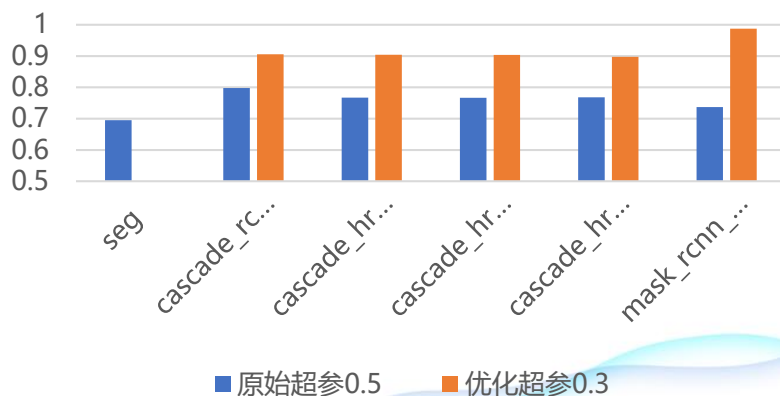


目标检测

Recall



Recall



Thanks

Reference:

- [1] Evolution of Image Segmentation using Deep Convolutional Neural Network: A Survey
- [2] Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.
- [3] Lin, Tsung-Yi, et al. "Feature pyramid networks for object detection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
- [4] Xiao, Tete, et al. "Unified perceptual parsing for scene understanding." Proceedings of the European Conference on Computer Vision (ECCV). 2018.
- [5] Zhao, Hengshuang, et al. "Pyramid scene parsing network." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
- [6] Chen, Liang-Chieh, et al. "Rethinking atrous convolution for semantic image segmentation." arXiv preprint arXiv:1706.05587 (2017).
- [7] Huang, Zilong, et al. "Ccnet: Criss-cross attention for semantic segmentation." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019.
- [8] Fu, Jun, et al. "Dual attention network for scene segmentation." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019.