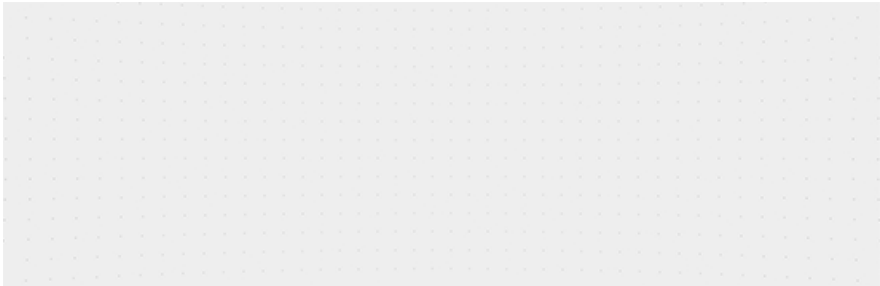


一个RGB-T工作的小小感悟与总结

原创 CV开发者都爱看的 极市平台 2021-10-19 22:00:00 手机阅读 囧

收录于话题
#计算机视觉

↑ 点击蓝字 关注极市平台



作者 | 战斗系牧师@知乎 (已授权)
来源 | <https://zhuanlan.zhihu.com/p/421925918>
编辑 | 极市平台

极市导读

作者从研究背景、原理出发详细介绍了他们中稿IROS-2021的工作：为 RGB-T 语义分割任务提出了一个两阶段的特征增强注意网络—FEANet。在实验方面，FEANet在客观指标和主观视觉比较方面优于其他最先进的 (SOTA) RGB-T方法。>>加入极市CV技术交流群，走在计算机视觉的最前沿

首先，非常激动的和大家说一个本应该三个多月前就激动的一个消息，就是我们的FEANet的工作在IROS-2021上被接收了，当然这绝对离不开老师和师兄，所以在开头先感激一下老师和师兄。也应各位很想知道我干了啥的同学的需求，如果有机会的话，可以引用一下这篇文章。

FEANet: Feature-Enhanced Attention Network for RGB-Thermal Real-time Semantic Segmentation

Fuqin Deng^{1,3,4*}, Hua Feng^{1*}, Mingjian Liang¹, Hongmin Wang¹, Yong Yang⁴, Yuan Gao³, Junfeng Chen³, Junjie Hu³, Xiyue Guo³, and Tin Lun Lam^{2,3,†}

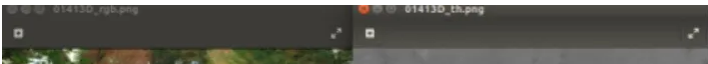
原文链接: <https://arxiv.org/abs/2110.08988>

第一部分

我们第一部分直观的通过2个问题快速介绍一下我们的研究背景，先让大家了解一下这篇文章的工作做了些什么。

1、RGB-T是干什么的？

答：在纹理相似、背景暗光，复杂的场景下，RGB图像往往并不能为模型训练提供更多更具有区分度的信息，因此常常会导致预测结果不准确，或者没有办法识别出物体。近年来随着热成像相机的普及，我们发现热红外信息对于照明条件差产生的识别模糊非常有效，例如在城市街景的语义分割任务中就取到了很好的效果，（RGB-T城市街景数据集链接: https://www.mi.t.u-tokyo.ac.jp/static/projects/mil_multispectral/）。因此，可以将热像仪生成的热红外图像作为重要的信息补充。



月发文数目: **
月平均阅读: **

文章工具

已发文

采集图文

合成多

采集样式

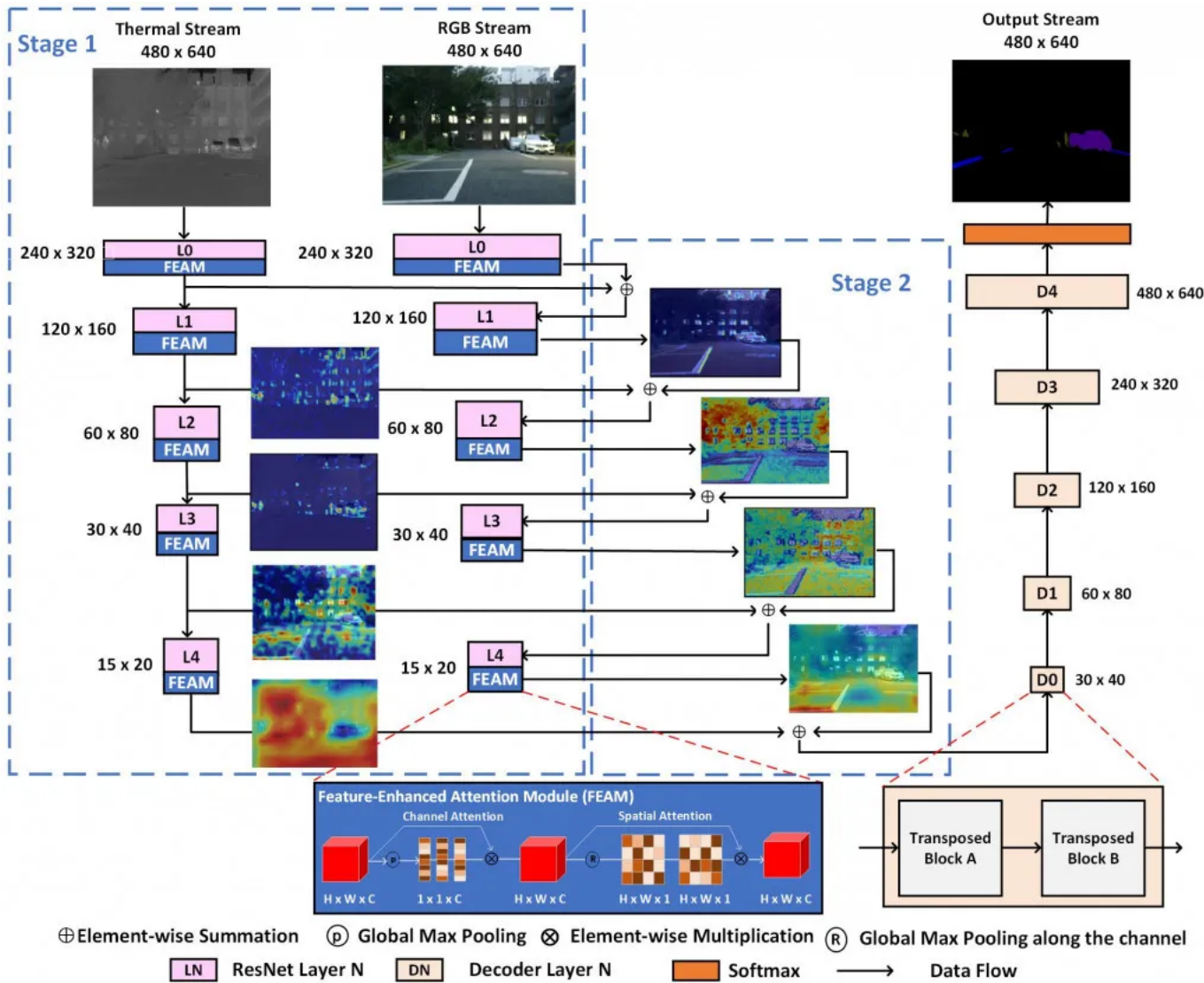
查看



被限速牌挡住的人，会在红外图像中非常直观的展现出来

2、FEANet是干什么的？

答：我们为 RGB-T 语义分割任务提出了一个两阶段的特征增强注意网络 (FEANet)。具体来说，就是我们引入了一个特征增强注意力模块 (FEAM) 从通道和空间的两个方向去提高模型的信息挖掘能力和增强模型的多级特征的提取和整合能力。



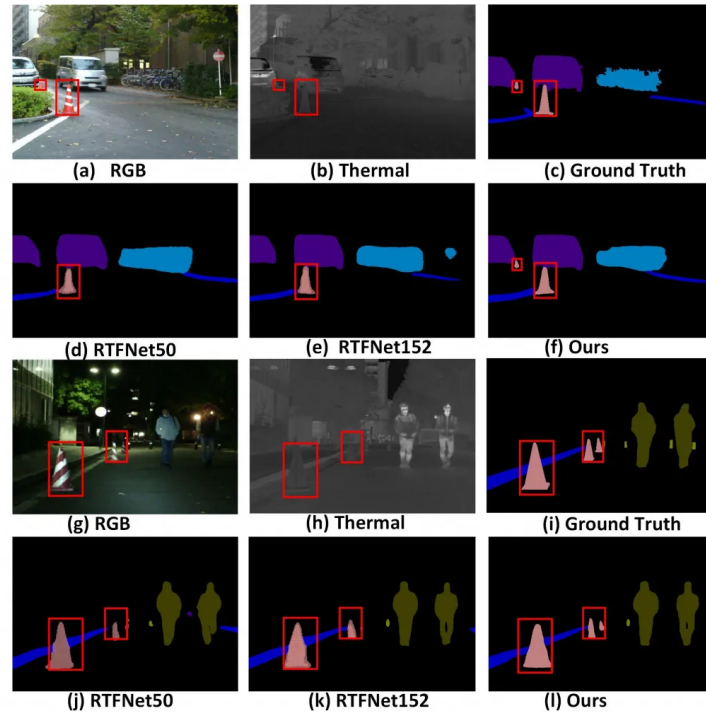
FEANet的网络结构，中间长得不是很好看的是 CAM可视化对于FEAM效果的一个直观的表达，可以看出FEAM结构能够很好的整合Thermal和RGB图像上的特征信息，并且也加强了RGB和Thermal图像的一个细节特征的捕获能力，以及部分物体的长距离建模都有了明显的改善。

第二部分

FEANet为什么可以？在这一部分我打算把文章网络结构中的几个idea由来的始末给大家介绍一下。

Idea1: 双encode, skip-connect结构

我们的FEANet的工作起初是基于RTFNet上进行改进的，在RTFNet的文中的消融实验结构充分表明了双encoder， skip-connect 结构的强大之外，RTFNet其实是具有很强的可加工性的。没错其实主要的原因就是它encoder部分太像Unet了，就让人忍不住想魔改它，就这样RTFNet就成了我们这次的baseline。然后，我们基于RTFNet在数据集上的不足点进行改进，同时，也与最新的（RTFNet的同一作者）的另一篇FuseSeg进行一个指标上的竞技。通过与RTFNet的结果进行研究，我们很快就发现了问题。

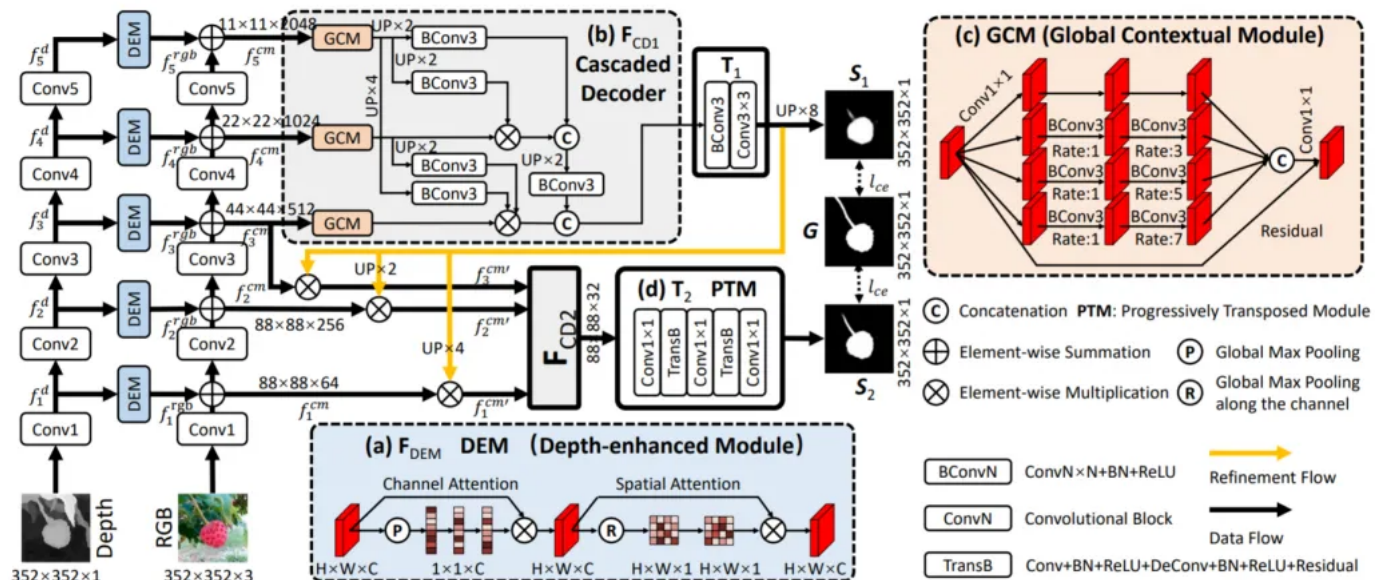


结果对比图

RTFNet在物体的细节上的分割非常的粗糙以及没有办法识别出小物体（比如：色锥），结合当时注意力机制模块对与细小目标上的良好表现来看，是否可以通过引入注意力机制模块来改善上述的不足之处呢？

Idea2: FEAM结构（注意力机制模块）

其实FEAM是受到另一篇文章BBSNet（基于RGB-D数据所开发的网络）的启发构建的，BBSNet是一篇成功将当时非常红的CBAM模块（图中的DEM结构）植入了网络的encode部分并取得SOTA效果的网络，那么本着RGB-D和RGB-T都是多光谱的图像会不会可以相互借鉴的好奇心态，我们设计了一个FEAM模块，使用注意力机制从融合数据中学习特征，然后改进网络的预测结果。



BBSNet的网络结构

Idea3: 为什么每层都添加

继续本着RGB-D和RGB-T都是多光谱的图像会不会可以相互借鉴的好奇心态，我们是否可以按照，RGB-FEAM-Thermal的结构模式去改造RTFNet呢？然后就有了我们第一版实验的。实验结果当时记得是：mAcc是69.1，mIou是54.1。其实从实验结果上看，我们已经高出了RTFNet一大截有多了，当时我都兴奋的蹦起来了，但是后面师兄看到结果后只是默默的给我发来FuseSeg这篇文章，我当时第一时间看了网络结构，哦原来是用FuseNet161换了ResNet152而已，但是当后面翻到结果一看，上面两个指标都高于目前我们设计的FEANet的时候，人都傻了。

还记得那天晚上组会的时候老师和我们说，现在指标都高于RTFNet就先和RTFNet做比较，就用田忌赛马的策略，如果故事讲得好说不定还有中的机会。我也在那时候真正的明白到了指标竞争的残酷之处。如果故事到这里的话也许就看不见现在的每层都加的FEANet了。

记得当时还希望说能不能通过调参数的方法打倒这两个指标，在调了两天参数后的一个晚上，我看到了之前自己做的CAM可视化的一个代码，然后就试着可视化了当时的添加FEAM后每一层网络。可视化的过程我突然想到，其实Thermal本质上不是想利用热力信息对RGB中明显可以看出的物体进行补充，而是给RGB看不见的信息进行一个补充，所以两张图关注到的对象不一样。可视化后在Thermal上添加了FEAM结构后，更加坚定了每层都加的想法，所以抱着试一试的心态进行了，每层都加的实验，最后实验出来了，mAcc：71.2，mIou：54.3，然后就有你们现在看到的FEANet网络结构了。

Idea4：loss函数不是传统的交叉熵

这个损失函数组合是通过一次天池上的语义分割比赛上看到并记下来的，DiceLoss 和 SoftCrossEntropy组合。本质上就是针对样本不平衡进行优化的，还记得当时师兄进行实验的时候，有一个有经验的博士说过可以通过改变损失函数对于样本不平衡的现象进行优化，只是当时实验出来的结果就有种拆东墙补西墙的感觉，但是没有想到组合后的损失函数也能够顺利下降并且对各个指标有有了提升，最后就有了我们的结果：

Methods	Car		Person		Bike		Curve		Car Stop		Guardrail		Color Cone		Bump		mAcc	mIoU
	Acc	IoU	Acc	IoU	Acc	IoU	Acc	IoU	Acc	IoU	Acc	IoU	Acc	IoU	Acc	IoU		
FRRN(4c)	81.9	74.7	66.2	60.8	62.8	50.3	41.2	35.0	12.5	11.5	0.0	0.0	37.2	34.0	35.2	34.6	48.5	44.2
FRRN(3c)	80.0	71.2	53.0	46.1	65.1	53.0	34.0	27.1	21.6	19.1	0.0	0.0	34.7	32.5	36.2	30.5	47.1	41.8
BiSeNet(4c)	89.7	84.1	72.0	63.2	74.1	60.1	45.1	36.7	34.2	25.3	18.2	5.0	47.4	42.2	39.8	35.9	57.7	50.0
BiSeNet(3c)	90.0	84.5	65.0	54.3	75.0	61.4	32.1	25.7	32.3	26.2	3.2	0.9	49.6	43.3	48.1	40.5	54.9	48.2
DFN(4c)	90.0	84.4	73.2	65.0	75.5	60.9	54.0	40.4	38.9	25.7	10.2	2.7	48.3	42.5	55.8	47.4	60.5	52.0
DFN(3c)	90.7	81.4	67.7	52.8	71.5	57.5	49.2	34.9	35.1	23.8	4.1	1.4	44.2	31.0	54.6	47.5	57.3	47.5
SegHRNet(4c)	92.8	87.6	79.3	71.0	78.3	63.4	59.8	42.5	25.7	19.1	18.8	0.0	56.5	49.8	63.5	44.5	63.7	53.2
SegHRNet(3c)	92.2	86.6	73.1	59.8	74.9	61.3	47.0	33.2	23.8	28.7	7.3	0.0	54.6	47.2	61.5	46.2	60.9	51.3
MFNet	77.2	65.9	67.0	58.9	53.9	42.9	36.2	29.9	19.1	9.9	0.1	8.5	30.3	25.2	30.0	27.7	45.1	39.7
FuseNet	81.0	75.6	75.2	66.3	64.5	51.9	51.0	37.8	28.7	15.0	0.0	0.0	31.1	21.4	51.9	45.0	52.4	45.6
DepthAwareCNN	85.2	77.0	61.7	53.4	76.0	56.5	40.2	30.9	9.9	29.3	22.8	6.4	32.9	30.1	36.5	32.3	55.1	46.1
RTFNet-50	91.3	86.3	78.2	67.8	71.5	58.2	69.8	43.7	32.1	24.3	13.4	3.6	40.4	26.0	73.5	57.2	62.2	51.7
RTFNet-152	93.0	87.4	79.3	70.3	76.8	62.7	60.7	45.3	38.5	29.8	0.0	0.0	45.5	29.1	74.7	55.7	63.1	53.2
FuseSeg-161	93.1	87.9	81.4	71.7	78.5	64.6	68.4	44.8	29.1	22.7	63.7	6.4	55.8	46.9	66.4	47.9	70.6	54.5
FEANet(Ours)	93.3	87.8	82.7	71.1	76.7	61.1	65.5	46.5	26.6	22.1	70.8	6.6	66.6	55.3	77.3	48.9	73.2	55.3

实验结果

添加了FEAM结构后并没有引入了很大的参数量，也没有造成推理速度的下滑，所以

FEANet: Feature-Enhanced Attention Network for RGB-Thermal Real-time Semantic Segmentation

由此诞生

第三部分

论文成功接受后的复盘，通过下面五个问题进行一个总结复盘

1、FEAM结构会不会过时了？

答：会，当时论文发的时候Transformer已经把CNN按在地上了，整个朋友圈CV清一色的Transformer，但是我们在设计的时候由于Transformer还没有蔓延到下游任务中，都是在围绕着ViT图像分类中进行的，ICCV2021的best paper SWIN-Transformer在各类比赛中作为backbone碾压各种CNN结构的网络，也许是时候了。看着指标做来去做都是60~70,50~55，当时就连评审都说只是高了0.8%而已，所以可能只是可能啊，更好的预训练模型会不会带来更好的结果呢？

2、BBSNet后续的cascade结构为啥没有延用呢？

答：实验证明，FEANet可能真的不适合这种结构，从因为与Deep图像补充RGB同一物体信息不同，Thermal图像是补充与RGB不同物体信息的，也许过分的级联会导致特征与特征间的语义对冲，从而会导致反作用，从结果上看也确实是的，并不适合。但是会不会有更合适的结构呢？别急，下篇工作告诉你。

3、实时性只在RTX 2080TI上反映靠谱吗？

答：见人见智的问题，以前他们都是在RTX 2080TI上进行对比实验的所以我们这么对比也是为了更加直观的反映我们比他们好，所以最开始在RTX 2080TI提实时性的这个文章可能责任全在它身上了，但是我相信很快部署在例如Jetson的轻量化RGB-T很快就会到来了，别细问，问就是下一篇文章的内容。

4、skip-connect真的可靠吗？

实验表明是可靠的，但是换个角度看现在也许是可靠了，但是如果这个是RGB-T-D这种多多模态的任务呢？如果是RGB-1-2-3-4-5-6呢？还这样连吗？skip-connect我感觉还是过于粗暴了，其实可以从可视化后的两个结构融合的图像上，确实出现了对于同一物体的不同识别结果的语义对冲，那有没有更好的方式组合，能够更加降低两种特征融合发生丢失，或者冲突的问题呢？会有，也许就是下一篇文章。

第四部分

感谢

我还记得这篇文章是寒假一个月时间赶出来的，为了投当时的3月的IROS，其实回想当时，过年的时候我们都还在拼命的研究和工作，老师和师兄都付出了非常多的心力，论文改又改，我的结构图也是画了又画，都忘记是第40几版了，非常庆幸的是我们的辛苦工作最终得到了认可，文章被接受了，但是更值得庆幸的是，我能够认识老师和师兄。最后再次的感谢各位帮助过我的师兄，还有同伴们，我们下一次顶会见。

如果觉得有用，就请分享到朋友圈吧！



极市平台

专注计算机视觉前沿资讯和技术干货，官网：www.cvmart.net

624篇原创内容

公众号

▲点击卡片关注极市平台，获取最新CV干货

公众号后台回复“CVPR21检测”获取CVPR2021目标检测论文下载~

极市干货

神经网络：视觉神经网络模型优秀开源工作：timm库使用方法和最新代码解读

技术综述：综述：神经网络中 Normalization 的发展历程 | CNN轻量化模型及其设计原则综述

算法技巧 (trick)：8点PyTorch提速技巧汇总 | 图像分类算法优化技巧

CV技术社群邀请图

△长按添加极市小助手

添加极市小助手微信 **(ID : cvmart4)**

备注：**姓名-学校/公司-研究方向-城市**（如：**小极-北大-目标检测-深圳**）

即可申请加入极市**目标检测/图像分割/工业检测/人脸/医学影像/3D/SLAM/自动驾驶/超分辨率/姿态估计/ReID/GAN/图像增强/OCR/视频理解**等技术交流群

每月大咖直播分享、真实项目需求对接、求职内推、算法竞赛、干货资讯汇总、与 **10000+**来自港科大、北大、清华、中科院、CMU、腾讯、百度等名校名企视觉开发者互动交流~

觉得有用麻烦给个在看啦~

阅读原文

喜欢此内容的人还喜欢

学术报告 | 阵列雷达波前调制前视成像技术（视频）
雷达学报

英伟达推出专属GeForce RTX键帽，将作为活动奖品送出
超能网

linux 之 cp, mv, rm 参数过长 Argument list too long
Linux就该这么学