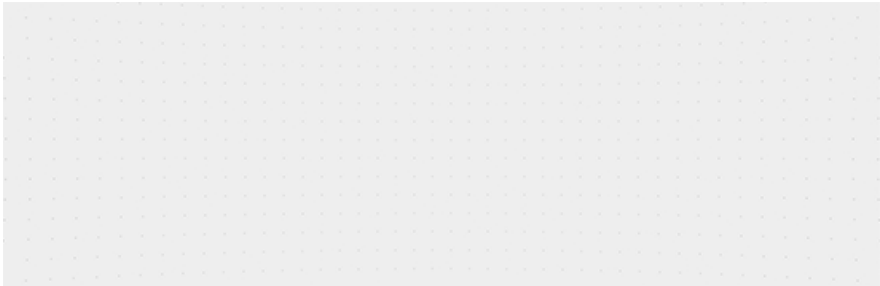


# CVPR 2021 Short-video Face Parsing Challenge 季军方案分享

原创 CV开发者都爱看的 极市平台 2021-07-14 22:00:00 手机阅读 眼

收录于话题  
#视觉顶会

↑ 点击蓝字 关注极市平台



作者 | 刘潇, 司晓飞, 谢江涛  
编辑 | 极市平台

## 极市导读

CVPR 2021 Short-video Face Parsing Challenge 大赛已于6月21日圆满结束，为cvpr workshop的竞赛单元赛事。本文为来自大连理工大学的大赛季军的方案分享：训练一个人脸解析模型，能够对短视频人脸进行解析分割。 >>加入极市CV技术交流群，走在计算机视觉的最前沿

近日，CVPR 2021 Short-video Face Parsing Challenge 大赛已于6月21日圆满结束。为cvpr workshop的竞赛单元赛事。大赛由PIC，马达智数和北京航空航天大学联合主办，极市平台提供技术支持的国际性赛事。该赛事主要关注可以应用到各场景中的短视频人脸解析技术。

本次大赛报名人数吸引海内外共300多支团队参与大赛，相比往期cvpr大赛，报名人数有较大增长。本次大赛的前三名分别由腾讯、北邮模式识别实验室及大连理工大学获得。我们邀请到了本次大赛的第三名获奖团队分享他们的工作。

## 赛题介绍

短视频涵盖技能分享、幽默风趣、时尚潮流、社会热点、街头采访、公益教育、广告创意等话题，使得其在社交媒体平台迅速流行起来。人脸解析是像素级的在人脸肖像中提取语义成分(例如嘴巴、眼睛和鼻子)。人脸解析可以提供比人脸轮廓和人脸关键点更加精细的信息，是一项更具挑战性的任务。

在本次比赛中，我们将训练一个人脸解析模型，能够对短视频人脸进行解析分割。评价指标为 Davis J/F score and temporal decay。

## 赛题分析

本次比赛的数据集由马达智数从数据库中挑选了总计1890段视频，每段视频约为20张图像。每段视频每秒抽取一帧，并对所有图像进行了采样和标注，划分为训练集、验证集和测试集三部分。需要划分的类别总共有17类(不含背景类)。经过对图像数据的分析可以总结出该数据集有以下特点：

1. 图像人像区域占比较小

壹伴图

极市平台  
extreme

月发文数目: \*\*  
月平均阅读: \*\*

文章工具

已发文

采集图文 合成多

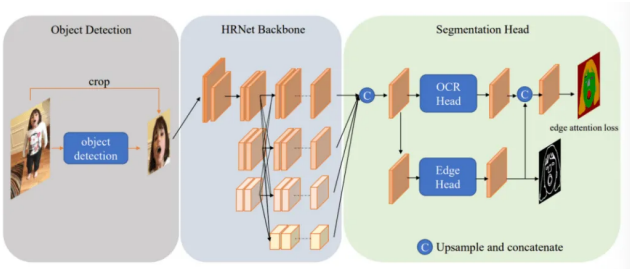
采集样式 查看

部分图像中的目标区域在整体图像中的占比较小，直接进行分割的话，一方面会引入过多的背景信息，另一方面随着模型的下采样，目标区域的信息会造成严重的丢失。

2. 一些图像只有部分人脸 数据集的场景覆盖范围较广，部分图像中存在只有部分人脸的情况。
3. 数据存在一定的标注错误 数据集中对于人脸的左右眼，左右眼影，左右耳，左右眉毛分别标注，数据集中存在少量的标注错误。
4. 部分图像质量较差 部分数据存在分辨率较低，亮度度过高/低，运动模糊和失焦等问题。

## 解决方案

考虑到人像区域占比小的问题，我们先将人脸区域裁剪出来然后进行分割，减少背景带来的噪声。因此我们采用了一个目标检测加语义分割的两阶段的做法来完成此次任务，使用PaddleDetection和PaddleSeg作为codebase。



### 第一阶段：目标检测裁剪人脸

我们训练一个目标检测器来裁剪人脸，因为我们只需要能把人脸检测出来，不需要检的很准，只需要尽可能的保证人脸都能够检测出来就可以，即需要一个高召回率。

我们利用数据集的标注mask，构造出人脸区域的边界框数据集，检测模型采用Cascade-RCNN-DCN-ResNet50-FPN，采用1x训练策略，数据增强使用水平翻转，其他都采用常规设置，没有任何trick。

将检测出的目标区域裁剪出来，作为下一阶段的输入数据。

### 第二阶段：人脸语义分割

- 数据增广

- 1)在做分割的时候大家都会加翻转，但是对于这个任务而言，人脸是镜像对称的，如果图像水平翻转后，mask也跟着翻转就会造成歧义。比如左眼翻转到右眼，但是由于镜像对称，它的label应该变成右眼，所以我们再水平翻转后再把镜像的部件翻转回来。
- 2)数据集中有部分人脸的情况，因此对完整人脸区域从上下左右四个方向，以1/3、2/3进行随机裁剪。
- 3)另外采用了随机小角度旋转和色彩抖动。

- 模型设计

在语义分割阶段，我们的Baseline为OCRNet-SEHRNet-w48，为了优化边缘分割结果引入了边缘监督，我们在Baseline的头部添加了一个和OCRHead并行的EdgeHead，利用边缘信息进一步加强模型的分割精度。

OCRHead就和原始OCR一致，我们尝试过加入Panotic-FPN，但效果不好，就直接用原始的OCR了。我们主要设计了一个EdgeHead，在原有基础上拉出一个分支，去分割边缘，再将edge feature 和 seg feature 融合送入Decoder分割。另外，我们还采用Edge Attention Loss来加强模型再边缘处的分割精度，Edge Attention Loss指的是在计算分割loss的时候，依据生成的边缘mask，只计算在边缘区域的多分类损失。我们还参考了Matting任务的思想，在最后的decoder输出，上采样到原尺寸后，在full resolution的尺寸上加入几层DenseLayer去refine分割结果。

• 损失函数设计

模型的损失构成如下式所示，总体分为两部分， $L_{aux}$  和  $L_{seg}$  为分割监督损失， $L_{edge}$  和  $L_{atte}$  为边缘监督损失。

$$L = \lambda_1 L_{aux} + \lambda_2 L_{seg} + \lambda_3 L_{edge} + \lambda_4 L_{atte}$$

其中， $L_{aux}$  为 HRNet 的输出进行粗分割的损失， $L_{seg}$ 为 EANet 头部输出进行分割的损失。分割监督损失为 CrossEntropy Loss 和 LovaszSoftmax Loss.  $L_{edge}$  为 BCE Loss,  $L_{atte}$  为 Edge Attention Loss, 如上述所示，只计算边缘区域的多分类损失。

• 模型集成 我们评估了将edge feature与seg feature的Add和Cat的融合方式，实验表明，Cat的融合方式性能更好。

我们简单的进行了三个模型的硬投票。三个模型分别是 OCRNet,EANet<sub>Add</sub> 和 EANet<sub>Cat</sub> 。其中 OCRNet 为我们的 Baseline, Add 和 Cat 表示不同的融合方式的 EANet。

• 一些Tricks

1) 使用大尺寸进行推理，我们在训练时采用448 \* 448的尺寸，在推理时采用480 \* 480的尺寸，小尺寸训练，大尺寸推理。

2) 我们使用PaddleSeg实现的模型，但是我们发现，PaddleSeg在推理阶段是先对模型输出的logit取argmax获得hard label然后再用最近邻插值恢复到原图，这样的推理模式在大尺寸图上会产生严重的锯齿，因此我们改进推理方式，先对logits双线性插值到原图获得soft logits然后再取argmax得到最终的分割结果，这个方式提升了一个点

3. 后处理上我们还采用GrabCut，利用得到的分割结果作为语义进一步refine。下图为使用GrabCut的对比图。

实验结果

Model	BackBone	Accuracy
OCRNet	SEHRNet-w48	83.42
EANet <sub>Add</sub>	SEHRNet-w48	84.51
EANet <sub>Cat</sub>	SEHRNet-w48	84.94
Ensemble	—	85.92
Ensemble+GrabCut	—	<b>86.16</b>



## 总结

除了上述的方案，我们还进行了尝试了Mutli-Scale, Pseudo-Labelling等，骨干模型上尝试过SwinTransformer + OCR等，分割范式尝试过MaskRCNN等，但并没有带来明显的性能提升。另外，我们尝试过利用帧间信息，将上一帧的预测mask与当前帧的图像拼接起来送入模型分割，该方法在val上有接近两个点的提升，但在test上并没有有效的提升，可能的原因是由于某一帧分割不准，造成误差传播到后续的帧上了。

## 团队介绍

团队成员均来自于大连理工大学，团队成员分别为：刘潇、司晓飞、谢江涛。

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](http://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

**YOLO教程：**一文读懂YOLO V5 与 YOLO V4 | 大盘点 | YOLO 系目标检测算法总览 | 全面解析YOLO V4网络结构

**实操教程：**PyTorch vs LibTorch：网络推理速度谁更快？ | 只用两行代码，我让Transformer推理加速了50倍 | PyTorch AutoGrad C++层实现

**算法技巧（trick）：**深度学习训练tricks总结（有实验支撑） | 深度强化学习调参Tricks合集 | 长尾识别中的Tricks汇总 (AAAI2021)

**最新CV竞赛：**2021 高通人工智能应用创新大赛 | CVPR 2021 | Short-video Face Parsing Challenge | 3D人体目标检测与行为分析竞赛开赛，奖池7万+，数据集达16671张！

添加极市小助手微信 (ID : cvmart2)

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](#) 文章已于2021/07/16修改

喜欢此内容的人还喜欢

## 15个目标检测开源数据集汇总

