

华中科技大学

课程设计报告

In Context Matting和Robust Video Matting对比

姓名 : 任雨松
学号 : U202115207
专业班级 : 人工智能本硕博2101班
课程 : 视觉认知工程
指导教师 : 陆昊
院(系,所) : 人工智能与自动化学院

华中科技大学

2024 年 6 月 30 日

目 录

1	数据标注过程	1
1.1	数据来源	1
1.2	蒙版标注	1
1.3	生成Trimap	1
2	实验结果	1
2.1	使用模型生成预测蒙版	1
2.2	性能参数分析	2
3	代码说明	3
4	参考文献	3

1 数据标注过程

1.1 数据来源

本次实验中，需要选择一部电影并从中截取一段5秒长度的浅景深视频片段，且聚焦主体应有一定的运动变化。在此，我选择了电影《哈利·波特与魔法石》作为数据来源，截取了其中的片段。



图 1.1.1 电影片段截图

将5秒片段分割为120帧后，进行下一步处理。

1.2 蒙版标注

使用Photoshop 2023对截取的每一帧进行蒙版标注，在借助辅助标注的同时人为调整细节，生成相应的alpha蒙版图片。对全部120帧图像都进行这一过程，得到120帧的蒙版图像，作为验证使用。

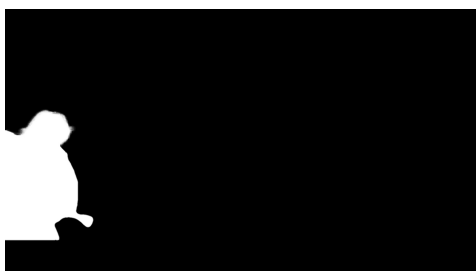


图 1.2.2 第一帧的alpha蒙版

1.3 生成Trimap

Trimap通过对alpha蒙版进行一系列膨胀腐蚀操作生成，分别使用三值灰度对背

景、未知区域、前景进行标注，trimap可以起到辅助模型进行推理的作用。

生成trimap的程序为 `./datasets/HP/tri.py`。



图 1.3.3 第一帧的trimap

2 实验结果

2.1 使用模型生成预测蒙版

分别使用In-Context Matting[1]和 RobustVideoMatting[2]模型预测蒙版，某一帧的预测结果大致如下：



图 2.1.4 标注的蒙版



图 2.1.5 ICM预测的蒙版

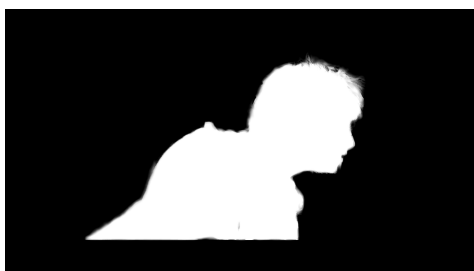


图 2.1.6 RVM预测的蒙版

在此帧上，两模型的表现均较为优秀，当然这是由于视频中人物与背景的景深非常明显，人物又占据画面主体导致的。如果是另一帧，效果如下：

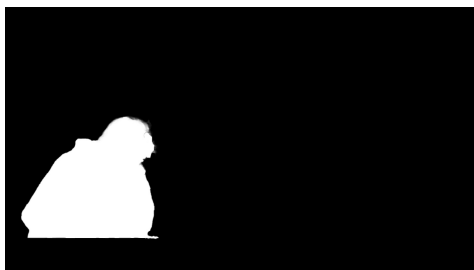


图 2.1.7 标注的蒙版



图 2.1.8 ICM预测的蒙版



图 2.1.9 RVM预测的蒙版

此时，ICM的预测性能大打折扣，这是由于图片中背景建筑较为鲜艳，容易被识别为主体部分，而RVM的预测结果仍然优秀。由此可以定性看出RVM的鲁棒性更为优秀。

2.2 性能参数分析

在本次实验中，采用 RobustVideoMatting 的文章中使用的五种评价指标，分别为平均绝对误差（MAD）、均方误差（MSE）、Grad（spatial gradient）、Conn（connectivity）和dtSSD。

由于原图像为1920*1080，而ICM生成的图像为768*768，故计算ICM模型预测结果的相关参数时，首先将标注图像居中裁剪为1080*1080，再等比例缩小进行比对。

与文章中的做法类似，为了使结果更为直观，MAD与MSE值乘1000后显示，dtSSD值乘100后显示。

结果如下表所示：

列1	MAD	MSE	Grad	Conn	dtSSD
ICM	16.5	9.33	0.525	0	1.71
RVM	11.72	8.34	0.289	11	0.672

图 2.2.10 模型性能比较

由数据定量可知，在样本视频片段中，RVM的综合预测性能优于ICM，这主要是由于ICM预测结果中容易误将背景中移动的建筑物判断为主体的一部分，导致在某些帧上与真实值严重不符。

3 代码说明

https://github.com/Lumos-s/ICM_Assignment

本次实验所有代码以及所用的数据集
可以在以下github仓库中获取：

使用方法参考仓库中的readme.md

2

4 参考文献

[1] GUO H, YE Z, CAO Z, 等 . In-Context Matting[C/OL]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 3711-3720[2024-06-28].

[2] LIN S, YANG L, SALEEMI I, 等. Robust High-Resolution Video Matting With Temporal Guidance[C/OL]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2022: 238-247[2024-06-28].