

**显著性物体检测**

院 系 信息工程学院

专　 　 业 电子信息工程

年 级 2019级

授 课 教 师 刘铁

论 文 作 者  第一组成员

完 成 日 期 2022年6月14日

目录

**[一、 研究背景介绍 3](#_Toc7616)**

**[二、 研究问题定义 3](#_Toc8903)**

**[三、 研究数据介绍 3](#_Toc28491)**

**[四、 研究算法介绍 5](#_Toc28266)**

[1. 算法解决的问题 5](#_Toc7357)

[2. 核心算法介绍 5](#_Toc16803)

[3. CFM模块 6](#_Toc27266)

[4. CFD模块 7](#_Toc29605)

[5. PPA损失函数 8](#_Toc11691)

**[五、 算法结果展示与分析 9](#_Toc7483)**

[1. 视觉直观分析 9](#_Toc23504)

[2. 基于四种评价指标分析 10](#_Toc1460)

[3. 消融实验 11](#_Toc28370)

**[六、 小组成员贡献 12](#_Toc30731)**

# 研究背景介绍

显著性物体检测最早可以追溯到1998年，Itti等发表了第一个基于显著度的视觉主义计算模型，使用多种特征的融合获得最终的显著图，视觉注意力机制模型正式引入到计算机视觉领域，从此显著性物体检测算法得到了蓬勃发展。现在显著性物体检测可以分为两种传统的显著性物体检测方法和基于深度学习的显著性物体检测方法。其中传统显著性物体检测算法还可以继续划分为基于对比度的显著性检测算法、基于频域的计算方法、基于稀疏理论的计算方法、基于图模型的计算方法、基于背景和前景先验的计算方法。在这其中例如Cheng等人提出全局对比度算法HC和RC等。在传统的方法使用空域或领域中大量的特征对比或者利用先验信息等进行图像显著性检测，其针对性较强，一般来说只适用于主题明确、物体颜色及结构较为单调的场景。基于深度学习的显著性物体检测方法又可以分为基于传统卷积神经网络的方法和基于完全卷积神经网络的方法。基于传统卷积神经网络的方法是将图像块的区域大小调整到固定的尺寸或进行多尺度的处理，然后进行卷积神经网络的特征提取，最后融合得到显著图。基于完全卷积神经网络的方法主要是利用VGGNet骨干网络架构和ResNet骨干网络架构。在基于深度学习的显著性物体检测方法中主要是利用卷积、下采样和上采样，会丧失显著物体的一些细节，在此基础上基于复杂的特征融合网络结构被提出。以下是显著性物体检测的发展历程图：

日程表

描述已自动生成

图 1 发展历史

# 研究问题定义

显著性物体检测可以分为静态物体检测和动态物体检测，静态物体检测主要是针对一幅图像中的显著物体进行检测，而动态显著物体检测主要是对视频场景中的显著物体进行检测。一般来说，对于视频中的运动物体，其显著性检测依赖于每一帧图像，当存在物体消失和出现时，其显著性检测也依赖于前后帧图像的对比。因此，图像的显著性物体检测是基础，所以本次大作业对显著性物体检测方法是在静态图像的显著性物体检测上。对于静态图像的显著性物体检测的研究问题是根据任意给定一张图片，检测出图片中的明显物体，然后就是检测出的显著图应具有高分辨率，同时能够准确定位并保留原始图像的信息，最后就是如何快速检测出显著物体。

# 研究数据介绍

显著性物体检测近年来得到快速发展的重要原因是其具有很多公开数据集，下表主要列举了数据集的名称的名称，容量、数据集中物体的属性、特点，以及相关文献和发表时间等信息。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 数据集名称 | 图片数量 | 物体属性 | 物体特点 | 发表年份 |
| MSRA-B | 5000 | 多为单目标，大小适中 | 简单 | 2007 |
| MSRA-10K | 10000 | 多为单目标，目标较大 | 简单 | 2007 |
| SED1 | 100 | 单目标，大小适中 | 简单 | 2007 |
| SED2 | 100 | 双目标，大小适中 | 简单 | 2007 |
| ASD | 1000 | 多为单目标，大小适中 | 简单 | 2009 |
| SOD | 300 | 多目标，大小适中 | 复杂 | 2010 |
| CSSD | 200 | 多为单目标，目标较大 | 简单 | 2013 |
| DUT-OMRON | 5168 | 多目标，大小不一 | 复杂 | 2013 |
| Thurl5k | 15000 | 单目标，目标较大 | 简单 | 2014 |
| PASCAL-S | 850 | 多目标，大小适中 | 复杂 | 2014 |
| HKU-IS | 4447 | 多目标，大小适中 | 复杂 | 2015 |
| ECSSD | 1000 | 多为单目标，目标较大 | 简单 | 2015 |
| DUTS | 15572 | 多目标，大小不一 | 复杂 | 2017 |
| XPIE | 10000 | 单目标，大小适中 | 复杂 | 2017 |
| SOC | 6000 | 多目标、大小不一，有非显著样本 | 复杂 | 2018 |

表 1 数据集说明

在这些数据集中MSRA-10K，DUT-OMRON，HKU-IS，ECSSD，DUTS和SOC被广泛应用于深度显著性检测模型的训练，下图展示了这六个常用数据集的图片和对应的真值。

图形用户界面, 应用程序

描述已自动生成

图 2 六个常用数据集图片与真值

在本次大作业中我们利用的数据集是DUTS，该数据集是Wang等于2017年建立的，该数据集包含10553张训练数据和5019张测试数据，总共15572张图片，适用于基于深度学习的检测方法，需要大量标签的数据。该数据集的图片中具有多个目标，并且目标与背景相似度高，被广泛应用于基于深度学习的方法当中。

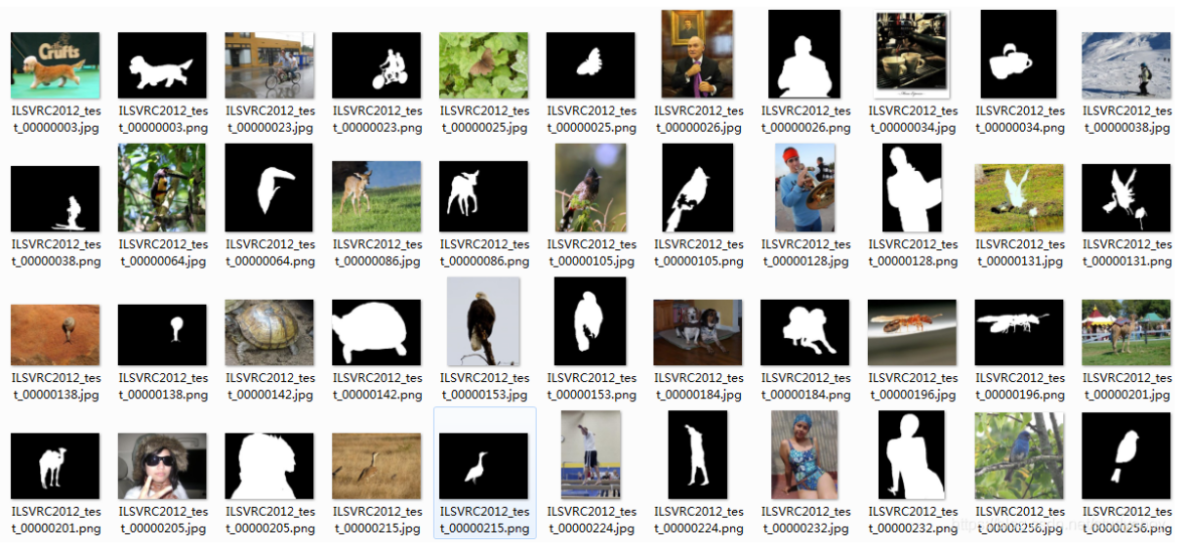


图 3 DUTS图片与真值

# 研究算法介绍

本次大作业的算法模块的实现是基于《F3Net:Fusion, Feedback and Focus for Salient Object Detection》这篇论文来完成的，此篇论文提出了一个F3Net结构来进行显著性物体目标检测，该结构是由交叉特征模块和通过最小化像素位置感知损失函数的级联反馈解码器组成。

## 算法解决的问题

现有的显著性目标检测模型大多是将卷积神经网络提取的多层次特征进行聚合，取得了较大的进展。但是，由于不同卷积层的接受域不同，这些层产生的特征之间存在较大的差异。常见的特征融合策略（添加或拼接）忽略了这些差异，可能导致次优解。此外显著性物体检测的损失函数一般是BCE，但是这种损失函数仅仅关注边界的像素明显不够。本次大作业中提出的算法能够很好解决上述问题，获得一个比较好效果的显著图。

## 核心算法介绍

本次大作业中算法提出了F3Net，该F3Net主要由交叉特征模块（CFM）和通过最小化像素位置感知损失（PPA）训练的级联反馈译码器（CFD）组成，下图为该算法的结构流程图。

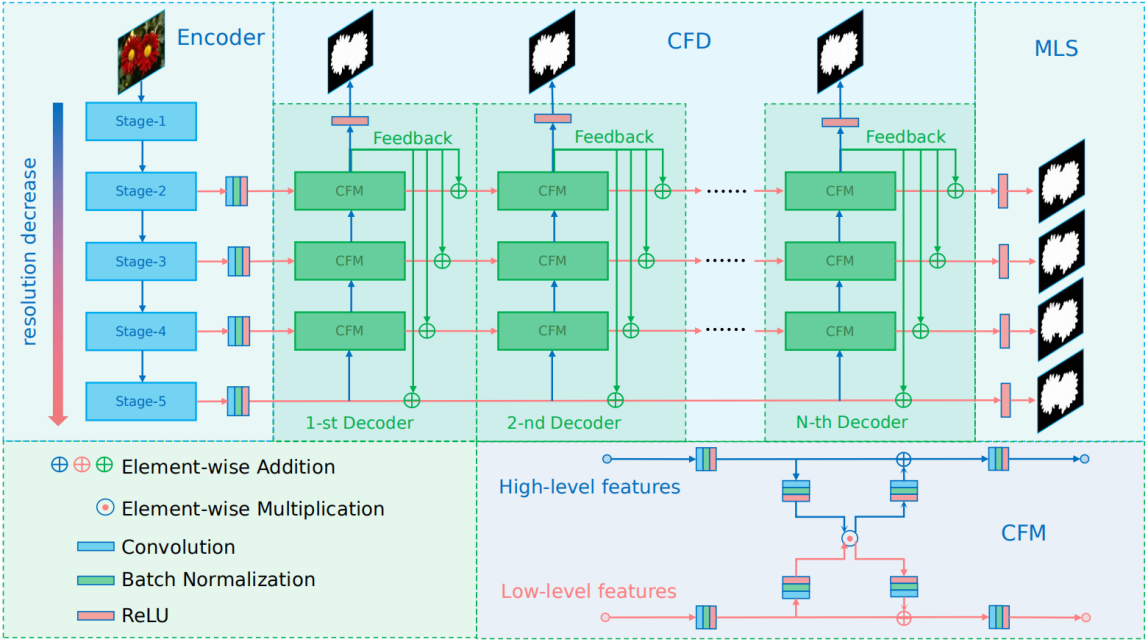


图 4 算法完整结构

针对该算法，我们首先来简单介绍一下该算法的实现过程。首先，我们输入一张图片，通过不同层之间进行卷积，然后得到不同层之间的卷积特征，对于这些得到的卷积特征，我们通过CFM这个模块来融合相邻两个层之间的卷积特征，也就是融合低层和高层之间的卷积特征。然后在通过CFD这个模块来生成我们最终想要的显著图，然后由图可以知道，一个CFD模块包括很多个子解码器，其中每一个子解码器包括很多个CFM，通过图上这种关系，一是自顶向下融合低层和高层的卷积特征，得到一个粗糙显著图，然后又将得到的显著图的特征反馈之前层来帮助提高得到显著图的效果，通过这个不断的迭代来获得一个效果优异的显著图。

## CFM模块

CFM模块执行特征交叉来缓解不同卷积层提取的特征的差异，其过程是高层和底层特征卷积、BN和ReLU后，元素相乘进行特征融合来提取特征的公共部分，然后分别和原来的特征进行元素加法进行特征细化下，图是CFM模块的结构显示。

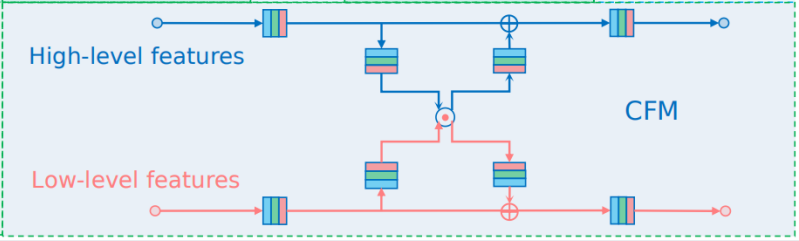


图 5 CFM模块

CFM模块，此模块的设计就是为了消除不同层之间的差异。首先由图可以知道CFM模块的流程是高层和底层先进行特征卷积、Batch Normalization和ReLu之后，我们将高层和底层得到的特征进行相乘，从而提取两层特征之间的公共部分，然后再分别和我们原来的特征进行元素加法进行特征细化，从而缓解不同卷积层提取的特征的差异，整个过程如下所示。



然后我们看一幅图（下图所示），由图我们可以知道底层特征提取能够看到图像的细节部分，但是有很大的背景噪音，高层特征提取能够很好提取图像结构和轮廓特征，同时消除噪音，但是比较模糊丢失了图像的细节部分，然后我们通过我们设计的CFM模块能够很好消除背景噪音和得到图像结构和轮廓特征，已经非常贴近我们的标签。

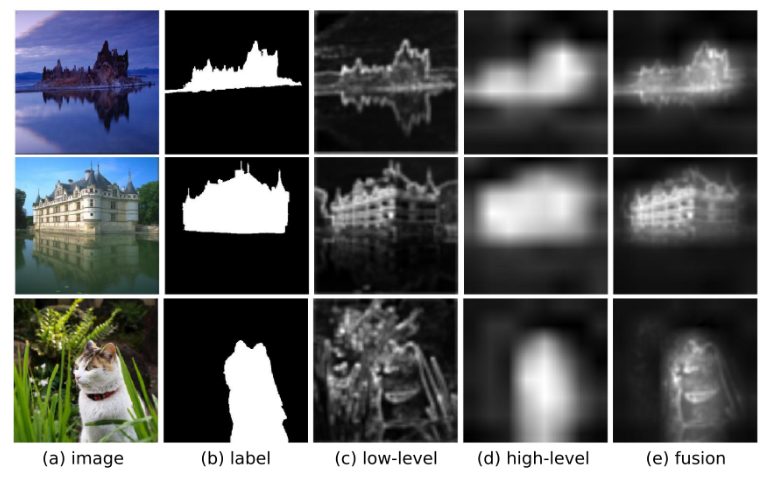


图 6 各层特征与聚合特征同图片和标签的横向对比

## CFD模块

CFD模块是建立在CFM的基础上，细化多层次特征，反馈迭代生成显著图。其过程可以理解为多个decoder连接迭代获得显著图。图为CFD模块结构图。

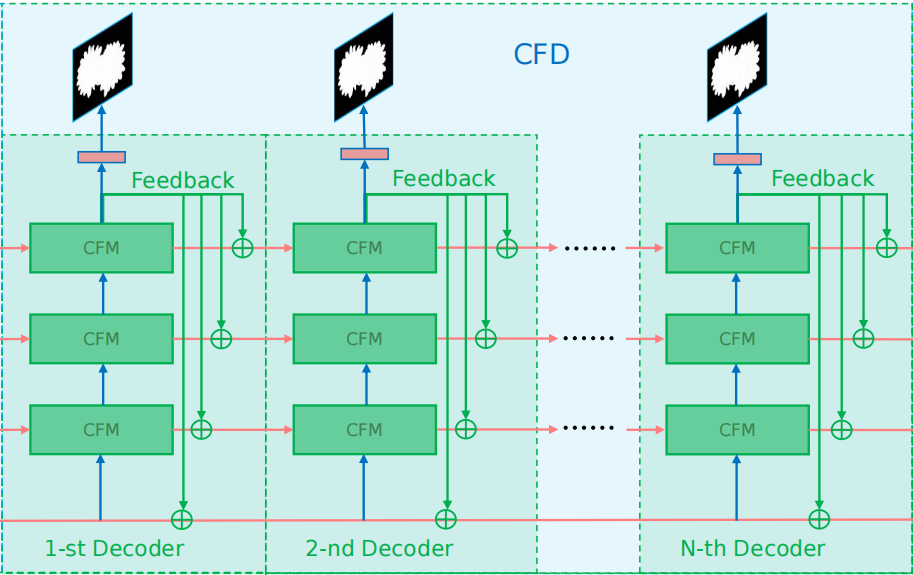


图 7 CFD模块

每个decoder包含两个操作。一个自底向上，另一个自顶向下。对于每次自底向上的操作，CFD模块使用CFM方法聚合多层次特征最后得到一个融合了多层特征的feedback。对于每次的自顶向下操作，CFD模块将feedback以CFM的方法重新融合到各层的输出图中作为反馈补充。以此完成循环迭代。如果以{f2,f3,f4,f5}表示网络的最后五层，Dei(·)表示第i个子decoder，Dsi(·)表示下采样操作，那么该算法流程可用如下伪代码表示：

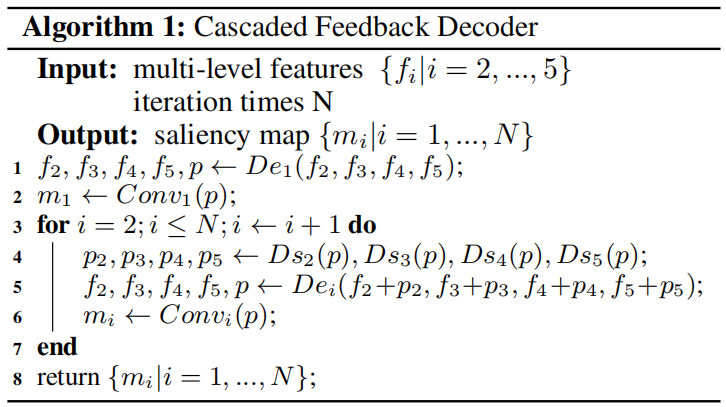
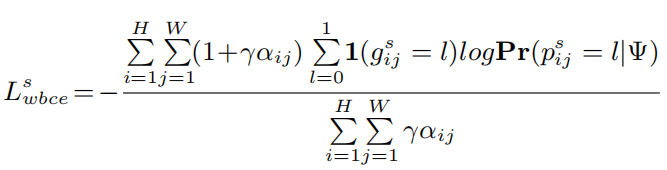


图 8 CFD算法

## PPA损失函数

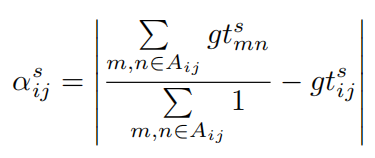
在显著物体检测领域，BCE是应用最为广泛的损失函数。然而，BCE有三个问题存在。首先，它独立地计算每个像素的损失，而忽略了图像的全局结构。其次，在背景占主导地位的图片中，前景像素的丢失将会被稀释。第三，它平等地对待所有像素，而没有对边缘和杂乱点的像素更为看重。所以首先，我们提出了一种加权的BCE损失函数WBCE。



公式 1 wbce损失函数

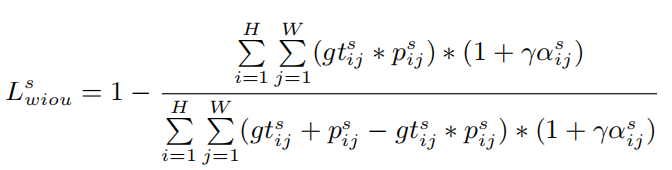
其中1(·)是指标函数，γ是超参数，l∈{0, 1}指示了两个不同的标签。p和g分别表示prediction预测值和ground-truth真值。Ψ表示模型的所有参数，而Pr则以probability可能性表示着模型的预测结果。

而其中的α则表示着权重，α可以被视为像素重要性的指标，是根据中心像素与周围像素的差值计算的。其中Aij表示表示在像素(i,j)周围的区域。



公式 2 权重计算公式

Wbce使损失函数在局部结构信息上的关注增加了，而为了使损失也能够更好的关注全局信息，我们还提出了加权IoU损失函数wIoU。



公式 3 wIoU损失函数

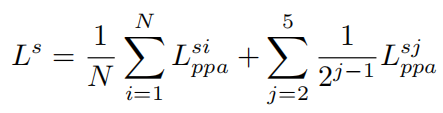
IoU损失函数被广泛的应用于图像分割领域，其目的是优化全局结构，而不是专注于单个像素，且不受不平衡分布的影响。今年来被应用到显著物体检测领域来弥补BCE损失函数的不足。但它依然是平等的对待所有像素，因此这里同样赋予了权重α。

基于以上的讨论，我们得到了PPA，它综合了局部信息，给像素加权，并引入局部限制WBCE和全局限制WIoU，这使它能够更好的进行网络学习，生成更清晰的预测图。



公式 4 PPA损失函数

另外，多层次监督MLS同样被我们引入来完善损失函数并得到更好的训练结果。N为CFD中子decoder的数量。5为网络层数。



公式 5 引入MLS的PPA损失函数

第一项对应所有子decoder损失的平均值，第二项对应引入MLS的损失加权和，其中高层损失由于误差较大而权重较小。

# 算法结果展示与分析

## 视觉直观分析

首先，我们从视觉上直观的来分析结果，其中ours是我们的算法，从算法间的横向比较来说，从视觉上可以看到，我们的算法在表现上十分不错，接下来，我们将从具体的评价指标上进行分析。

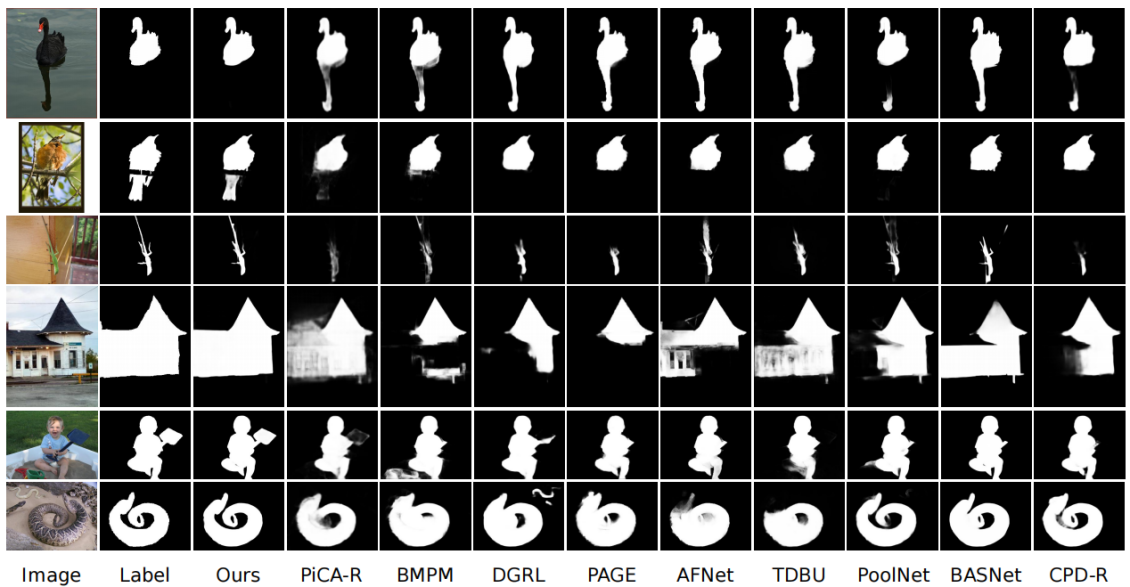


图 9 算法间横向比较

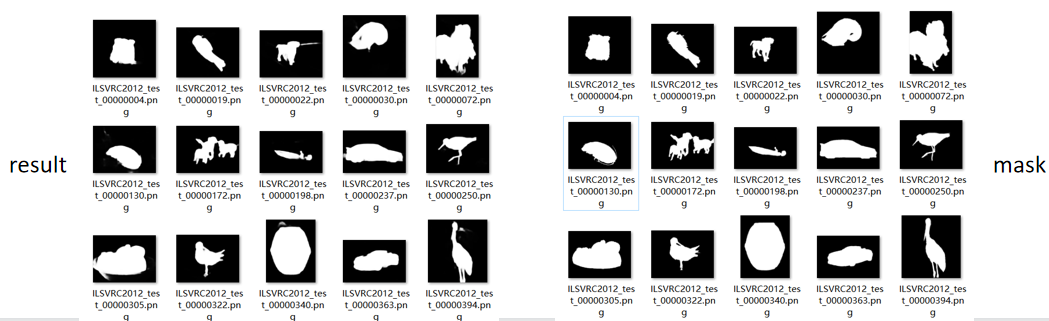


图 10 左图为小组自我训练推理结果，右图为标签

## 基于四种评价指标分析

我们一共使用四个评价指标来评估结果的好坏：

MAE（Mean Absolute Error越小越好）、mF（F-measure越大越好）、Sα（S-measure越大越好）、Eξ（E-measure越大越好）。

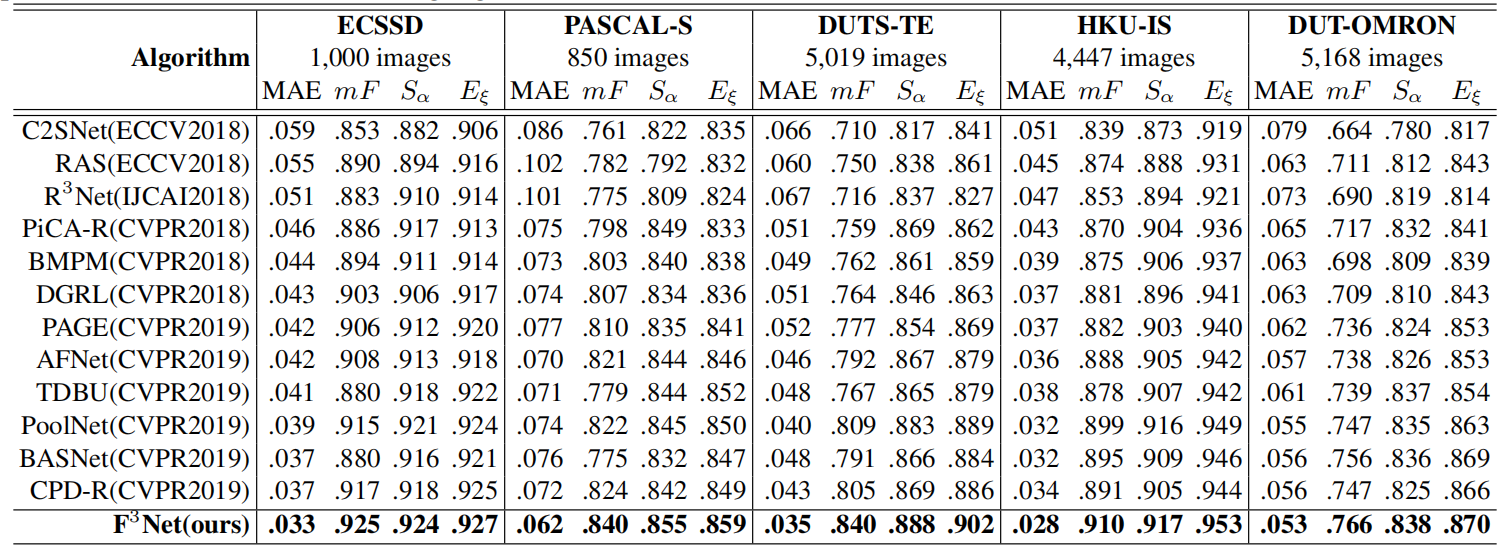


图 11 不同算法与数据集间的横向比较

对于不同数据集，我们可以发现，在ECSSD上表现最好，在其他数据集上也同样有良好的表现。对于不同算法，经过比较指标我们可以发现，在这些数据集上，F3Net都是表现最良好的。

接下来，我们又从PR曲线与不同阈值的mF的角度进行了分析。

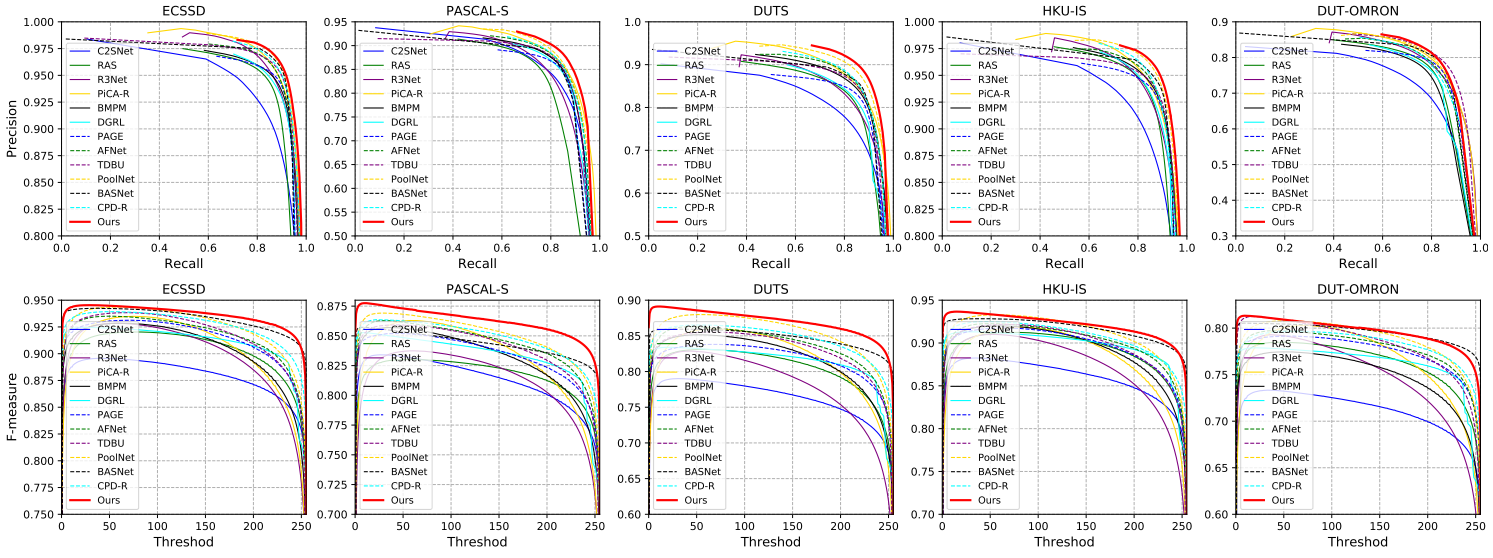


图 12 第一行为PR曲线；第二行为不同阈值的mF

对于第一行，PR曲线，我们通常可以看P = R时的值来判断结果的好坏，显然有F3Net表现最好。针对第二行的F-measure，在改变阈值的情况下，F3Net的值一直处于较高的水平，说明性能更好。

## 消融实验

对于消融实验，我们设计了自己的消融实验内容。实验对象包括之前提到的CFM、CFD、PPA，以及BCE、IoU、MLS共六个模块。

注意到BCE、IoU、PPA实质上都属于loss函数一类，其中PPA为本文提出的方法。而CFM、CFD、MLS属于不同的方法。我们将消融实验区分为如下几场：

**·without feedback in CFD**

注意到Cascaded feedback decoder，也就是CFD中，最关键的部分就是feedback部分，我们将feedback屏蔽，以此实现消融CFD模块进行实验。

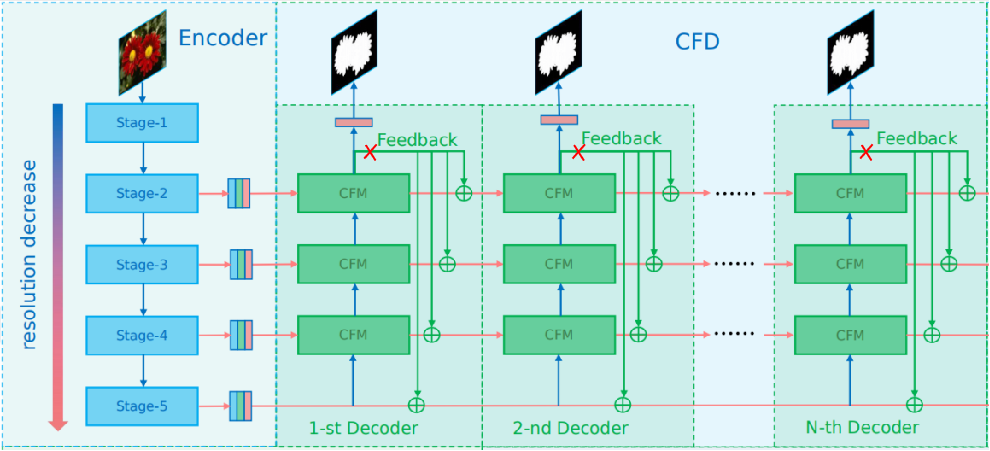


图 13 屏蔽feedback的CFD模块

**without CFM**

为了保证 Cross feature module，也就是CFM模块的消融不会影响算法完整性，我们选择了direct addition feature的方式作为替代，且为了与原算法前后保持相同的分布、避免可能的梯度消失或梯度爆炸，我们通过了与原算法类似的convert、batchnorm层和relu激活。

**without MLS**

正如前面提到的一样，Multi-level supervision，也就是MLS，实际上就是加权了多层SOD输出的loss，既然要消融它，那我们的loss就考虑选择接受了最完整fuse信息的feedback来计算，以保证其他影响最小化。

**not PPA but WIOU** & **not PPA but WBCE**

从前面我们知道，实际上PPA就是简单的WBCE与WIOU的累和，所以在消融时，只需要稍微修改损失函数的结构就可以达到效果。

最终，我们得到了如下结果：

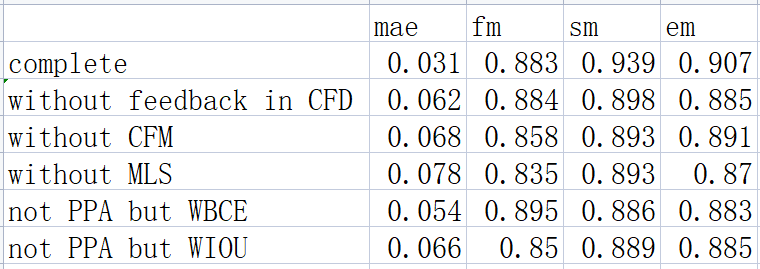


图 14 消融实验结果

可以直观看到的是，在mae、sm、em三个评价指标上，完整算法都有着最佳的结果，但在mf上的结果并不是很尽如人意，我们判断可能是在设计消融实验以及调试代码设置参数时造成的问题。

同时，如果我们将之展开成柱状图的形式，能够更直观的得到一些结论：

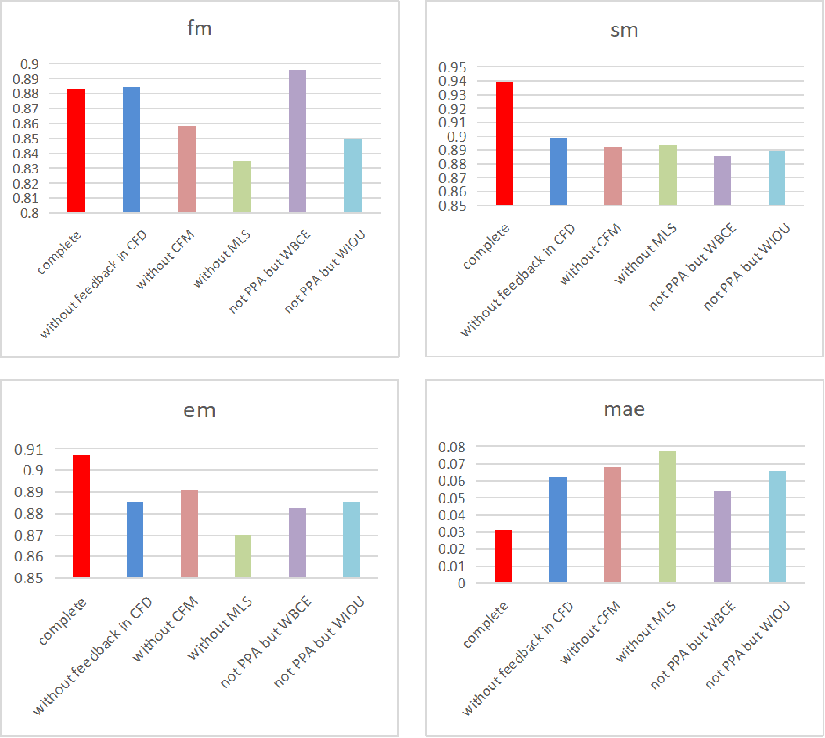


图 15 消融实验柱状图展开形式

MLS对于算法的贡献最大；在fm评价指标中，not PPA but WBCE反超了原算法，但在其他评价指标上，均能得出PPA对于算法有重大贡献的结论；在fm评价指标中，without feedback in CFD得到的评价比原算法略高，但在其他指标上，均能得出CFD对于算法有重大贡献的结论；在任意指标上，CFM都体现出了对于算法的贡献；对于CFD，CFM，PPA，三者对于算法的贡献基本在同一个层次上。

# 小组成员贡献

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 小组成员 | 曹博钧 | 褚思衡 | 郑昊文 | 王帝宬 | 赵凯 |
| 主要负责内容 | 算法实验复现与改善 | 实验论文研究、PPT制作 | 实验环境与算法调试 | 实验数据集收集和研究背景调研 | 算法实验复现与改善 |
| 贡献比 | 20% | 20% | 20% | 20% | 20% |