南開大學
Nankai University

计算机学院

深度学习期末实验报告

显著性目标检测

学号：2112529 姓名：赵廷枫

学号：2112380 姓名：黄昶玮

学号：2111033 姓名：艾明旭

2024 年 6 月 23 日

目录

1 概要

1.1 实验内容

在本次实验中，我们主要针对显著性目标检测任务，选择了三篇基准论文作为研究对象，分别是EGNet、PoolNet 和 F³Net。实验的目的是复现这些方法的结果，并在此基础上进行改进，以提升显著性目标检测的性能。

首先，我们对 EGNet、PoolNet 和 F³Net 的原始实现进行了复现，使用相同的数据集和实验配置，得到了各自的基准结果。随后，我们对 EGNet 进行了多个改进实验，包括优化显著性损失函数、引入结构化损失、以及融合 CFM 模块。这些改进的目标是通过更好地利用显著性目标信息和边缘信息，提升模型的检测精度和鲁棒性。

接下来，我们对 F³Net 进行了实验，除了复现原始方法外，还引入了显著性边缘信息作为监督信号，目的是进一步提升模型在边缘细节处的检测性能。我们通过将显著性边缘信息与目标特征融合，设计了一个增强版的 F³Net 模型，并对其进行了充分训练和验证。

最后，我们进行了消融实验，评估了各个改进模块对整体模型性能的贡献。通过对比不同配置下的实验结果，我们分析了各模块的有效性，并总结了最佳的改进方案。同时，我们还探讨了未来的研究方向，包括引入 Res2Net 增强多尺度特征表示，以及探索深度信息在显著性目标检测中的应用。

我们的代码开源在：DeepLearningFinal.

1.2 评价指标

我们采取了两个评价指标来评估我们的模型效果，一个是 MAE. 另外一个是 F-Measure，他们的计算公式分别为：

$$\mathrm{MAE} = \frac{1}{W \times H} \sum_{x=1}^{W} \sum_{y=1}^{H} |S(x,y) - \mathrm{GT}(x,y)|$$

$$F_\beta = \frac{(1+\beta^2) Precision \times Recall}{\beta^2 Precision + Recall}$$

1.3 分工

赵廷枫同学：

- 阅读 EGNet 论文，理解原理并复现 EGNet 的 baseline 实验结果
- 对 EGNet 进行了三项改进，包括：优化显著性损失函数、使用 CFM 模块优化特征融合、尝试深度信息融合
- 对 EGNet 的多项改进进行消融实验，整理实验结果并进行分析

黄昶玮同学：

- 阅读 F3Net 论文，理解原理并复现 F3Net 的 baseline 实验结果
- 对 F3Net 进行改进，使用边缘信息监督模型的训练
- 对 F3Net 相关改进进行消融实验，整理实验结果并进行分析

艾明旭同学：

- 阅读 PoolNet 论文，理解原理并复现 PoolNet 的 baseline 实验结果
- 阅读前沿论文并定期分享，为 PoolNet、EGNet、F3Net 的改进提供理论支持

2 相关工作

2.1 显著性目标检测

显著性目标检测是计算机视觉领域的一项重要任务，旨在从图像或视频中准确地识别和定位出显著性目标。显著性目标通常是指在图像中引起人眼注意的突出目标，如人物、车辆、动物等。显著性目标检测可以帮助计算机系统更好地理解图像内容，从而在图像处理、图像搜索、智能监控等领域发挥重要作用。

显著性目标检测算法通常基于以下原理: 首先，通过计算图像中每个像素的显著性值，来衡量其在整个图像中的重要程度。然后，根据这些显著性值，将图像分割成显著性目标和背景两部分。最后，通过进一步的处理和分析，确定显著性目标的位置和边界。

显著性目标检测算法可以基于不同的特征和方法进行实现，包括传统的基于颜色、纹理、边缘等低级特征的方法，以及基于深度学习的方法。近年来，深度学习方法在显著性目标检测中取得了显著的进展，通过使用深度神经网络可以更准确地提取图像特征和进行目标定位。

2.2 传统方法

传统显著性目标检测方法通常基于低级视觉特征，如颜色、纹理和边缘等。早期的方法如 Itti 等人提出的基于生物学视觉模型的方法，通过模拟人类视觉系统的注意机制来检测显著性目标。这类方法通常包括三个步骤：计算图像中每个像素的显著性值，分割图像以区分显著性目标和背景，最后通过进一步处理确定显著性目标的位置和边界。然而，这些方法在复杂场景下的表现受限，难以应对多样化的背景和目标。

2.3 深度学习方法

近年来，深度学习在显著性目标检测中取得了显著进展。深度学习方法通过深度神经网络能够更好地提取图像特征，实现更精确的目标定位。典型的深度学习方法包括基于卷积神经网络（CNN）的多尺度特征融合方法和基于生成对抗网络（GAN）的端到端训练方法。近年来，结合边缘信息的显著性目标检测方法受到了广泛关注。显著性目标的边缘通常具有较高的梯度变化，通过引入边缘信息，可以帮助模型更准确地定位显著目标的边界。

3 Baseline 原理

本次实验，我们选取了三篇文章作为我们的 Baseline，分别为

1. EGNet[6]: Edge guidance network for salient object detection
2. PoolNet[2]: A simple pooling-based design for real-time salient object detection
3. F^3Net[3]: fusion, feedback and focus for salient object detection

这三篇文章分别设计了不同的模型和方法来完成显著性目标检测这一任务，并通过 F-measure, MAE, S-measure 这些评价指标来系统地评估了模型的效果。

其中 EGNet 通过融合显著边缘信息和显著目标信息来提升边界检测的准确性。在一个单一的网络中，通过逐层提取和整合这些互补的信息，增强了显著目标的边缘细节，显著提升了检测效果。PoolNet 利用基于池化的设计来高效地整合多尺度上下文信息。这种设计通过简化网络结构，实现了实时的显著目标检测，同时保持了高准确性。F^3Net 采用融合、反馈和聚焦机制，通过迭代地细化特征提取过程并强调显著区域，同时抑制非显著区域，从而大幅提高了显著目标检测的性能。

3.1 PoolNet

基础模块：GGM 模块和 FAM 模块。如图 1 所示，GGM 模块由金字塔池化（PPM）和一系列的 GGFs 组成。本文的 GGM 都是独立模块，PPM 位于自顶向上路径的最高层，这样做的目的是得到全局引导信息。通过引入 GGFs，将 PPM 提取的高层次语义信息送到每层金字塔层的每个特征图中，从而弥补 U 型网络自上而下的信号逐渐被稀释的缺点。考虑到 GGFs 中的粗糙特征图与金字塔不同尺度提取出来的特征图的融合问题，提出一个模块叫 FAM 模块。FAM 模块输入为融合后的特征图，FAM 模块首先将融合的特征图转换为多个特征空间，以捕获不同尺度的局部上下文信息，然后组合信息以更好地权衡融合输入特征图的组成。作为这项工作的延伸，我们还为我们的架构配备了一个边缘检测分支，通过与边缘检测联合训练我们的模型来进一步锐化显著目标的细节。

GGM 模块：CNN 的经验感受野远小于理论上的感知域，因此整个网络的感受域不足以捕获输入图像的全局信息。对此的直接影响是只能发现部分显著物体，如图中所示。GGM 模块中 PPM 模块，如下图所示，包含四个分支，第一个是恒等映射层，最后一个是全局平均池化层，中间两个是自适应平均池化层，为了保证输出的特征图是 3x3 和 5x5。我们的 GGM 独立于 u 形结构。通过引入一系列全局引导流 (身份映射)，可以很容易地将高级语义信息传递到各个级别的特征图中。

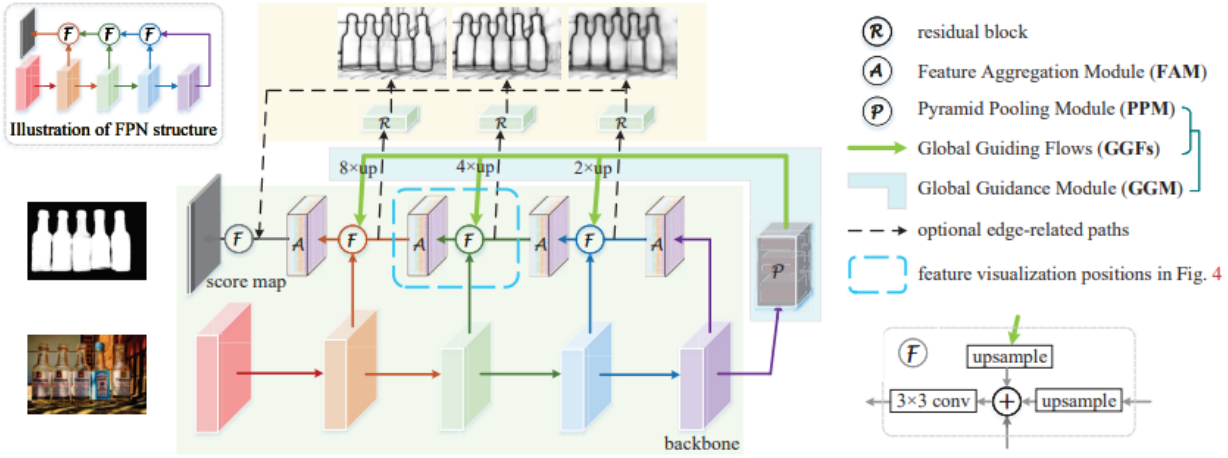


Figure 1. The overall pipeline of our proposed approach. For clarity, we also place a standard U-shape FPN structure [22] at the top-left corner. The top part for edge detection is optional.

图 3.1: PoolNet 结构

FAM 模块: 前向过程中，仍然使用了金字塔池化的思想，对输入进行四个分支的平均池化，再通过 3x3 卷积进行上采样，最后再将四个分支拼接，拼接完加 3x3 卷积。首先，它帮助我们的模型减少了上采样的混叠效应，特别是当上采样率很大时。此外，它允许每个空间位置在不同的尺度空间上查看本地上下文，进一步扩大了整个网络的接受场。

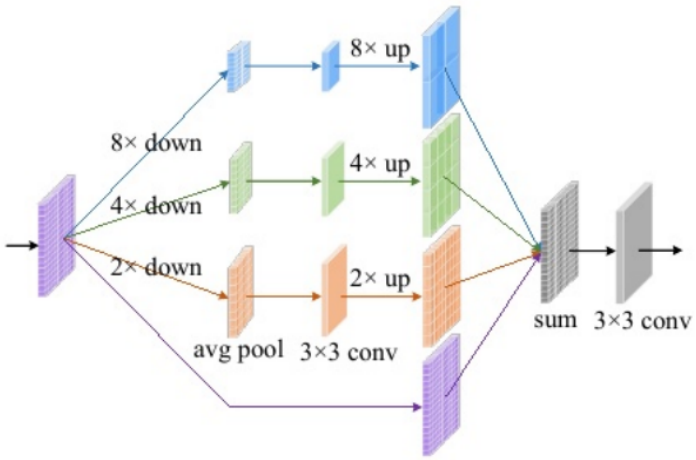


图 3.2: FAM

3.2 EGNet

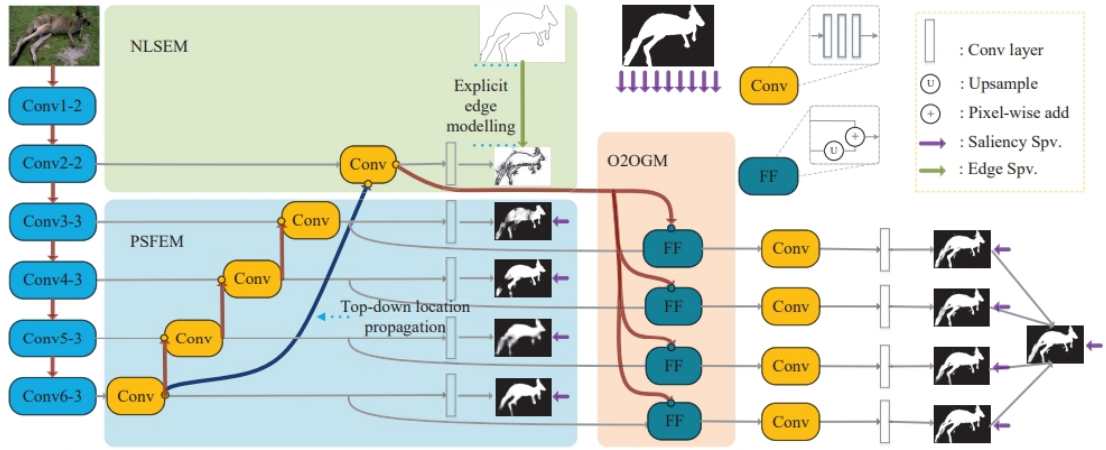


Figure 2. The pipeline of the proposed approach. We use brown thick lines to represent information flows between the scales. PSFEM: progressive salient object features extraction module. NLSEM: non-local salient edge features extraction module. O2OGM : one-to-one guidance module. FF: feature fusion. Spv.: supervision.

图 3.3: EGNet[6] 模型框架

本文提出 EGNet 来显式地将显著目标信息和显著边缘信息进行融合，以优化分割结果的显著性以及显著目标的边缘特征。概要而言：

1. EGNet 使用渐进式显著目标特征提取模块（PSFEM）提取出显著的目标特征。
2. EGNet 使用非局部显著边缘特征提取模块（NLSEM）将局部边缘信息和全局位置信息相结合，从而得到显著的边缘特征。
3. EGNet 通过一对一指导模块（O2OGM）将相同的显著边缘特征与不同分辨率下的显著目标特征相结合。

通过上述三个模块，使用 VGG 或者 ResNet 提取出来的图像特征被充分地利用起来，图像中的显著目标信息和显著边缘信息进行了充分的互补监督和融合，最终显著性目标检测的效果也有了显著的提升。

3.2.1 PSFEM

渐进式显著目标特征提取模块（PSFEM）是为了提取出图像中的显著目标信息。因为高层的特征对显著目标的位置信息更加关注，所以使用的是 Backbone 处理得到的 3、4、5、6 层的特征。

为了获得更稳健的显著性目标特征，将每个边路径经过三个卷积层，并在每个卷积层后面添加了 ReLU 激活函数层以保证非线性。这里不同层的特征进行融合之前，因为通道数会有不同，所以会先经过一个转换层使得通道数一致融合的时候，直接逐像素相加就得到了融合之后的特征，然后传递给上一次，从而能够以一种渐进式的方式提取显著目标信息。

为了更好地监督这一过程，每一层经过前面的融合和处理之后，采用一个卷积层将特征映射转换为单通道预测映射，最终输出四个显著目标特征提取映射图，这个过程中设计的损失函数是：

$$\begin{aligned}\mathcal{L}^{(i)'}(\hat{G}^{(i)}; W_{D'}^{(i)}) = &- \sum_{j \in Y_+} \log Pr(y_j = 1 | \hat{G}^{(i)}; W_{D'}^{(i)}) \\ &- \sum_{j \in Y_-} \log Pr(y_j = 0 | \hat{G}^{(i)}; W_{D'}^{(i)}), i \in [3, 6].\end{aligned}$$

其中，Y_+ and Y_- 分别是显著区域像素集和非显著区域像素集，$Pr(y_j = 1 | \hat{F^{(i)}}; W_D^{(i)})$ 表示的是预测图中对该像素的置信度，F 表示的是提取并处理之后的特征，W 表示的是转换层的参数。

3.2.2 NLSEM

非局部显著边缘特征提取模块（NLSEM）的目的是获取显著目标的边缘信息，因为第二层的感受野比较小可以关注更多的细节信息，所以以 Backbone 第二层的提取的特征为基础来获取显著边缘信息，但是因为第二层的感受野过小导致缺失高级语义信息或位置信息，所以 EGNet 采用了一种自顶向下的位置传播，将顶层位置信息传播到侧路径 S（2）以抑制不显著的边缘，然后进行特征的融合，获得显著边缘特征。

这里为了更好地对这一过程进行监督，EGNet 也是使用一个卷积层将特征映射转换为单通道预测映射，最终输出显著边缘图，然后使用 groundtruth 处理得到的边缘进行监督，这一过程中的损失函数设计如下：

$$\begin{aligned}\mathcal{L}^{(2)}(F_E; W_D^{(2)}) = &- \sum_{j \in Z_+} \log Pr(y_j = 1 | F_E; W_D^{(2)}) \\ &- \sum_{j \in Z_-} \log Pr(y_j = 0 | F_E; W_D^{(2)}),\end{aligned}$$

其中 Z_+ and Z_- 表示显著目标像素集以及背景像素集。

3.2.3 O2OGM

一对一指导模块（O2OGM）的作用是利用显著边缘特征来一对一地引导显著目标特征，以更好地进行显著性目标的分割和定位。具体的做法就是将上述两个模块获得的显著边缘信息和显著目标信息进行融合。

EGNet 在上述两个模块的处理过程中都加入了旁路，从而将得到的特征在输出为图像之前，输入到一对一指导模块（O2OGM）进行高效融合。这样就在先前得到的四个显著目标图的基础上得到了显著边缘信息加强之后的结果，之后在对四个结果进行融合，从而得到最后的显著性目标分割结果。

在这个过程中，EGNet 也设计了专门的损失函数来监督这一过程，每一条分支的损失计算公式如

下：

$$\mathcal{L}^{(i)'}(\hat{G}^{(i)}; W_{D'}^{(i)}) = -\sum_{j \in Y_+} \log Pr(y_j = 1 | \hat{G}^{(i)}; W_{D'}^{(i)}) - \sum_{j \in Y_-} \log Pr(y_j = 0 | \hat{G}^{(i)}; W_{D'}^{(i)}), i \in [3, 6].$$

针对融合过程的损失设计如下：

$$\mathcal{L}'_f(\hat{G}; W_{D'}) = \sigma(Y, \sum_{i=3}^{6} \beta_i f(\hat{G}^{(i)}; W_{D'}^{(i)}))$$

3.3 F3Net

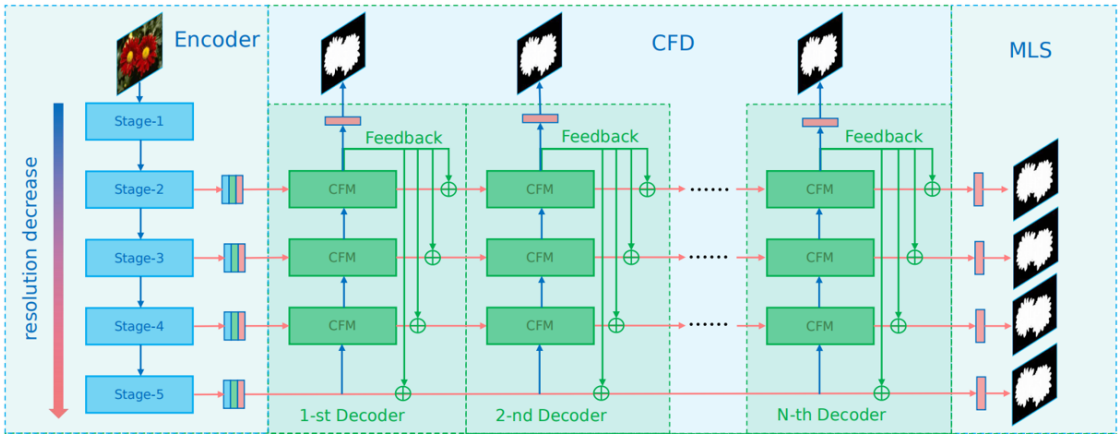


图 3.4: F^3net 网络框架 [3]

CFM 和 CFD: 首先，为了减轻特征之间的差异，F3Net 设计了交叉特征模块 (CFM)，该模块通过元素乘法融合不同层次的特征。与加法和拼接不同，CFM 采用选择性融合策略，剔除冗余信息，避免特征之间的污染，重要特征相互补充。与传统的融合方法相比，CFM 能够去除背景噪声和锐化边界。其次，由于下采样，高电平特征可能遭受信息丢失和失真，这是 CFM 无法解决的。因此，F3Net 使用了级联反馈解码器 (CFD) 来迭代地改进这些特征。CFD 包含多个子解码器，每个子解码器都包含自底向上和自顶向下的过程。对于自底向上的过程，多级特征由 CFM 逐步聚合。对于自顶向下的过程，聚合的特性被反馈到以前的特性中以改进它们。

PPA: F3Net 提出了像素位置感知损失 (PPA) 来改进常用的对所有像素都平等对待的二值交叉熵损失。事实上，位于边界或拉长区域的像素更加困难和区分。对这些硬像素的关注可以进一步提高模型的泛化能力。PPA 损耗对不同像素点赋予不同的权值，扩展了二值交叉熵。每个像素的权重由其周围的像素决定。硬像素将获得更大的权重，而容易像素将获得更小的权重。损失函数这里指出了 BCE 的三个缺点：

像素级损失: 首先，它独立计算每个像素的损失，忽略图像的全局结构。易受大的区域的引导: 其次，在背景占主导地位的图片中，前景像素的损失将被稀释。平等对待每个像素: 第三，它平等对待所有像素。事实上，位于杂乱或细长区域 (如杆和角) 的像素容易出现错误预测，值得更多关注，而位于天空和草地等区域的像素则不值得关注。最终使用位置重加权的方式，结合了像素级损失 BEL 和区域级损失 IOU 来进行监督：每一个被计算的像素都被赋予了一个权重，事实上，位于杂乱或细长区域 (如杆和角) 的像素容易出现错误预测，值得更多关注，而位于天空和草地等区域的像素不值得关注。

权重的计算公式：

$$\alpha_{ij}^{s} = \left| \frac{\sum_{m,n \in A_{ij}} gt_{mn}^{s}}{\sum_{m,n \in A_{ij}} 1} - gt_{ij}^{s} \right|$$

4 PoolNet 实验及改进

本次实验，我们主要利用 windows 11 系统配置相关环境完成。在基础的 pytorch 之外，还要配置相应的 torchvision，已完成相应的目标训练。

PoolNet 的具体实现方式如下：

- 首先确保安装了 pytorch 和 torchvision 两个必要的库。
- 接下来在 github 上 clone 相关的代码
- 之后下载主干网的预训练模型
- 然后设置训练集和测试集，这里设置为题目要求的 700 张图片和 300 张图片
- 选择训练方式并进行训练。

图 4.5: PoolNet 结果示例

一张输出的预测结果图如上图所示，可以看到预测的结果能够较好的预测分析显著性边缘。

最终的实验结果如下：

基本的实验配置：

num_thread	1
batch_size	32
learning_rate	5e-5
num_workers	config.num_thread
iter_size	10
decay	5e-4
epoch	24

结果：

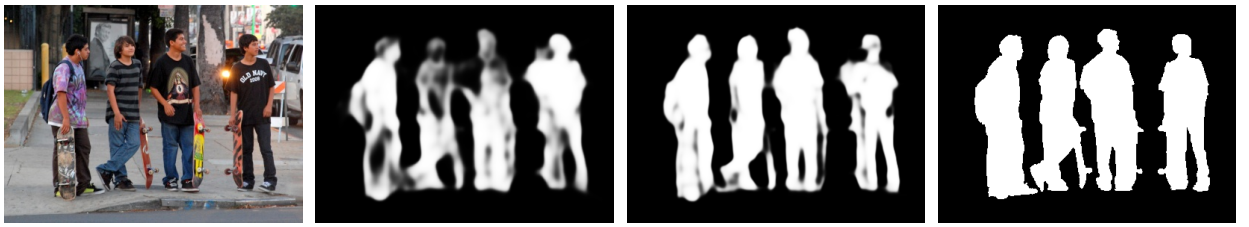
mean_mae	0.0541
mean_fb	0.9671

由于 baseline 的实现结果并不十分优秀，并且我们尝试的优化方向，诸如损失函数优化，引入边缘性，深度性模块的方法也不甚理想。因此我们的实验最终选择了放弃优化此网络结构，专注于优化另外两种 baseline 以求得较好的实验结果。

5 EGNet 实验及改进

5.1 基础实验结果

将论文中实现的代码在本地重新训练之后得到 baseline 结果，将我们的 EGNet 结果和 PoolNet 结果进行对比可以得到：



Origin PoolNet EGNet GroundTruth

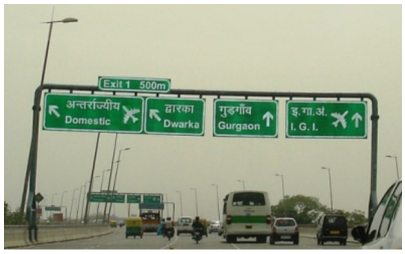
Origin PoolNet EGNet GroundTruth

图 5.6: EGNet 与 PoolNet 结果对比

从图中的实验结果可以看出，EGNet 通过将显著边缘信息和显著目标信息有效地融合，最终的分割结果无论是在显著性上还是边缘的细节上都要优于 PoolNet。

5.2 优化显著性损失函数

我们在观察实验结果的过程中发现，很多分割结果的显著性损失较大，很多非显著性的物体也被当做显著性物体分割了出来：



Origin

EGNet

GroundTruth

Origin

EGNet

GroundTruth

其实 EGNet 中使用的损失函数其实就是显著目标分割任务中最常用 BCE 损失函数，但是该损失函数存在三个不足 [3]：

1. 它独立计算每个像素的损失，忽略了图像的全局结构。
2. 在背景占主导地位的图片中，前景像素的损失将被稀释。
3. 它平等地对待所有像素。

所以我们借鉴了 F3Net 中的两个损失函数来优化 EGNet 中针对显著性设计的损失函数：一个是带权重的 BCE 损失函数

$$L^s_{wbce} = -\frac{\sum_{i=1}^{H}\sum_{j=1}^{W}(1+\gamma\alpha_{ij})\sum_{l=0}^{1}\mathbf{1}(g^s_{ij}=l)log\mathbf{Pr}(p^s_{ij}=l|\Psi)}{\sum_{i=1}^{H}\sum_{j=1}^{W}\gamma\alpha_{ij}}$$

其中，$\mathbf{1}(\cdot)$ 是一个指示函数，l 是标签，标记当前像素是目标还是背景；γ 是超参数，p^s_{ij} 是预测图中的像素值，g^s_{ij} 是 groundtruth 中的像素值，$\mathbf{Pr}(p^s_{i,j}=l|\hat{\Psi})$ 是每一个像素被预测的概率，α 是赋予给每一个像素的权重，该权重的计算方式为：

$$\alpha^s_{ij} = \left|\frac{\sum_{m,n\in A_{ij}} gt^s_{mn}}{\sum_{m,n\in A_{ij}} 1} - gt^s_{ij}\right|$$

其中，A_{ij} 表示的当前像素的邻域，gt 表示的是 groundtruth 中的显著性值。

另外一个是带权重的 IOU 损失函数：

$$L^s_{wiou} = 1 - \frac{\sum_{i=1}^{H}\sum_{j=1}^{W}(gt^s_{ij} * p^s_{ij}) * (1 + \gamma\alpha^s_{ij})}{\sum_{i=1}^{H}\sum_{j=1}^{W}(gt^s_{ij} + p^s_{ij} - gt^s_{ij} * p^s_{ij}) * (1 + \gamma\alpha^s_{ij})}$$

使用新的损失函数重新训练之后得到的结果如下：



EGNet EGNet-StructureLoss GroundTruth



EGNet EGNet-StructureLoss GroundTruth

从图中的分割结果可以看出，使用优化后的损失函数训练之后，模型分割出来的结果显著性有了明显的改进。

5.3 CFM 模块优化融合

在对 baseline 实验结果的观察中我发现：不同层次特征融合的过程中采用的是简单的逐像素加法，这种方法引入了多余的低位噪声，导致很多分割结果受噪声的影响而分割结果不佳。

为了减少噪声的影响，我借鉴了 F3Net[3] 中的 CFM 模块中的乘法操作来代替原先的加法操作，从而防止多余的噪声的引入。

修改模型结构之后，我重新对模型进行了充分的训练，得到的一些结果如下：

Origin EGNet EGNet-CFM GroundTruth

图 5.7: 使用 CFM 优化 EGNet 效果对比

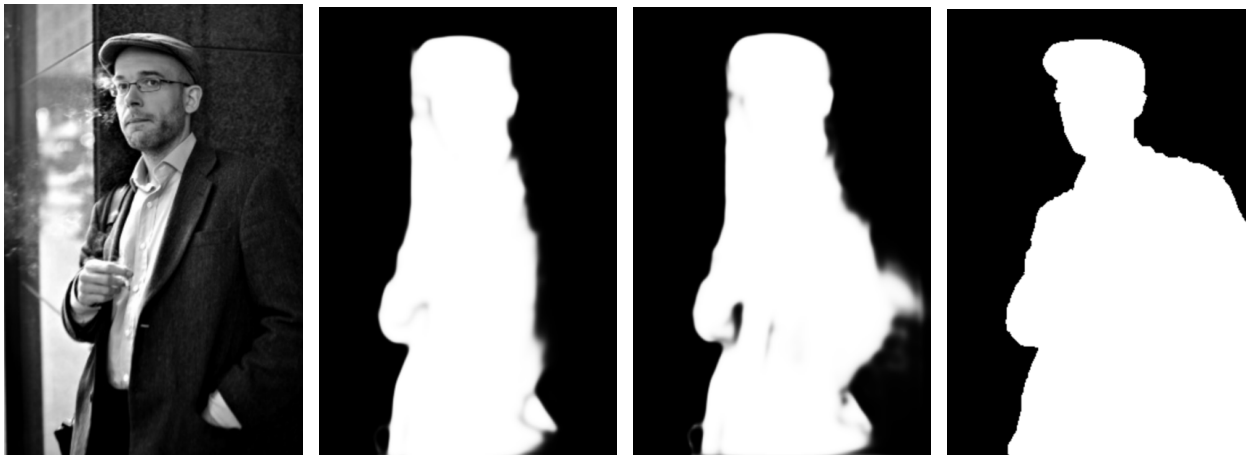
从图中可以看出，加入 CFM 模块之后的 EGNet 能够防止低维噪声的印象，从而呈现出更好的分割结果。

5.4 深度信息融合

在这一部分，我尝试将深度信息融合进入模型中，从而使得模型可以从深度信息中提取出显著性信息。

我首先使用 Depth-Anything[4] 离线处理得到了所有的深度图，之后我尝试借鉴了 BBSNet[5] 中处理深度信息的方式来将深度信息加入到模型之中。

但是最终的结果差强人意，导致模型测试的 MAE 求和比 Baseline 还要高，最终消融实验的时候也是舍弃掉该部分，但是还是存在一些分割结果是要比 Baseline 的结果优秀：



Origin EGNet EGNet-Depth GroundTruth

图 5.8: 使用深度信息优化 EGNet 效果对比

5.5 实验结果汇总

我将 EGNet 相关改进得到的实验结果汇总如下：

原始 EGNet 的 Overall-MAE 为 14.72968547028606，对应的 Mean-MAE 为 0.04909895。这是我们的基准性能，用于评估其他改进方法的效果。

Method	Overall-MAE	Mean-MAE
EGNet	14.72968547028606	0.04909895
EGNet-StructureLoss	13.208199977642929	0.04402733
EGNet-CFM	14.832710405610996	0.04944237
EGNet-Depth	16.13233493761359	0.05377445
EGNet-CFM+StructureLoss	13.274288225342877	0.04424763

表 1: EGNet 相关方法的实验结果

引入结构化损失后的 EGNet-StructureLoss，其 Overall-MAE 降低至 13.208199977642929，对应的 Mean-MAE 为 0.04402733。这一结果表明结构化损失显著提升了模型的检测性能。理论上，结构化损失通过更好地捕捉全局结构信息，减少了单个像素损失的局限性，从而提高了显著性目标的检测精度。

引入 CFM 模块的 EGNet-CFM，其 Overall-MAE 为 14.832710405610996，对应的 Mean-MAE 为 0.04944237。相比原始 EGNet，CFM 模块的改进效果不显著。这可能是因为 CFM 模块的多尺度特征融合并未能充分发挥作用，或是融合过程中出现了一些特征污染，未能显著提高整体性能。

引入深度信息的 EGNet-Depth，其 Overall-MAE 为 16.13233493761359，对应的 Mean-MAE 为 0.05377445。相较于原始 EGNet，深度信息的引入反而增加了误差。这可能是因为深度信息的融合方式不够优化，导致额外的噪声或不一致性，从而降低了模型的检测性能。

结合 CFM 和结构化损失的 EGNet-CFM+StructureLoss，其 Overall-MAE 为 13.274288225342877，对应的 Mean-MAE 为 0.04424763。尽管这一方法的性能优于原始 EGNet 和 EGNet-CFM，但与单独引入结构化损失的 EGNet-StructureLoss 相比，改进效果略逊一筹。这表明结构化损失的引入在提升模型性能方面更为有效，而 CFM 模块的效果在该组合中未能充分体现。

综上所述，实验结果表明结构化损失是最有效的改进方法，通过捕捉全局结构信息显著提高了显著性目标检测的精度。而 CFM 模块和深度信息的引入在现有实现中未能表现出显著的优势，可能需要进一步优化融合策略和网络结构。未来的工作可以探索更为有效的特征融合方法和深度信息的利用方式，以进一步提升显著性目标检测的性能。

6 F3Net 实验及改进

6.1 基础实验结果

我们以原论文提出的网络作为 baseline，使用 ECSSD 作为数据集。随机划分 700 张作为训练集，300 张作为测试集，经过训练后得到结果。

基本的实验配置：

optimizer	SGD
batch_size	32
learning_rate	0.05
momentum	0.9
nesterov	true
decay	5e-4
epoch	32

结果：

mean_mae	0.04037
mean_fb	0.98

6.2 使用边缘信息监督

在训练过程中我们发现，原始的 F3Net 网络对部分物体的边缘信息感知不够，生成的图像在边缘附近会比较模糊。因此，我们提出对 F3Net 的优化，即引入边缘图片对网络进行监督训练。

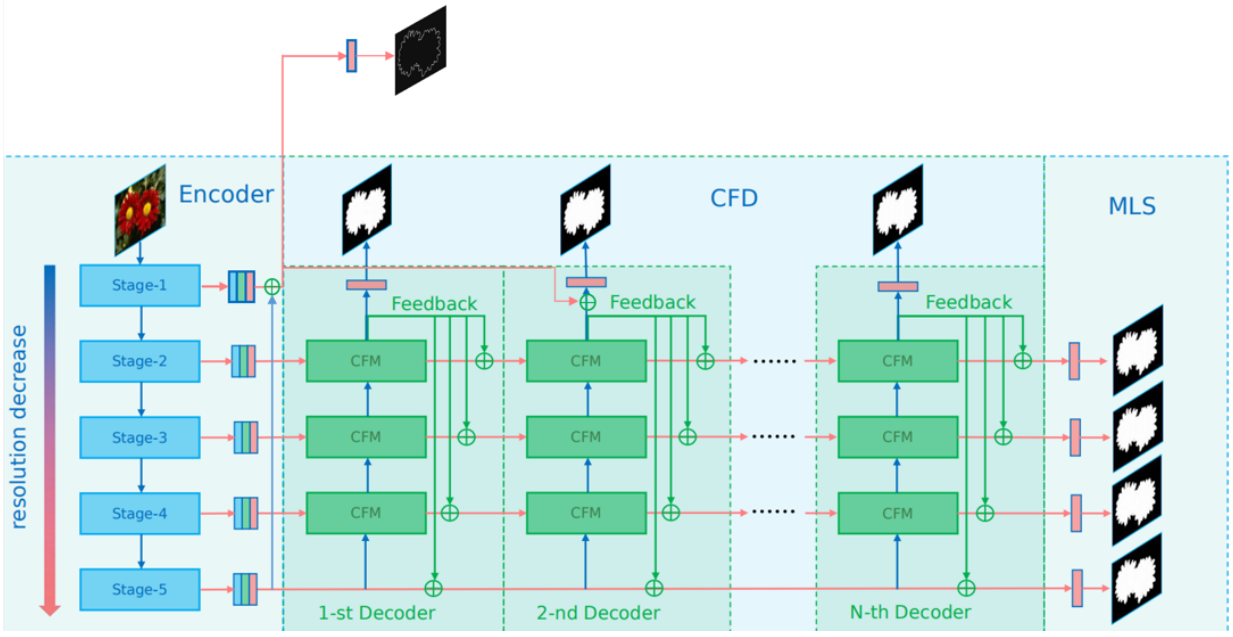


图 6.9: 引入边缘监督的 F^3net 网络框架

如图 6.9 所示，我们利用了 backbone 的第一层特征图 (拥有精确的位置信息) 和最高层的特征图 (拥有最丰富的语义信息) 融合后，使用预生成的边缘图进行监督。

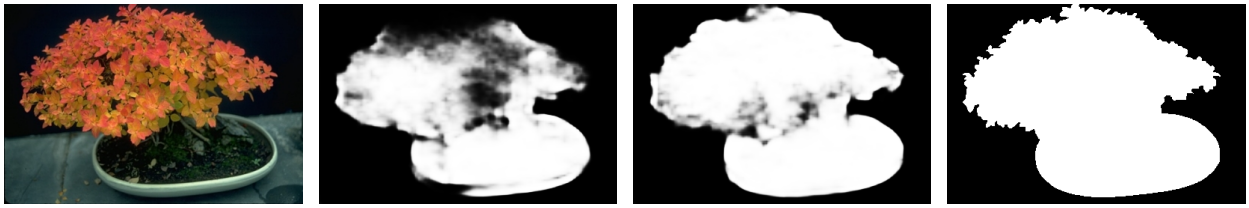
引入显著性边缘信息的监督有多方面的好处：

1. 显著性物体的边缘通常比内部区域具有更高的梯度变化，引入这些边缘信息可以帮助模型更好地区分显著性物体与背景。
2. 模型在训练过程中，如果能够获得准确的边缘信息，就能减少把背景误判为显著性物体的概率，这增强了模型的鲁棒性。
3. 这种多任务学习能够共享不同任务之间的信息和特征，能提高模型的泛化能力和整体性能。

我们对优化后的模型使用与 baseline 同样的配置进行充分训练，得到的结果：

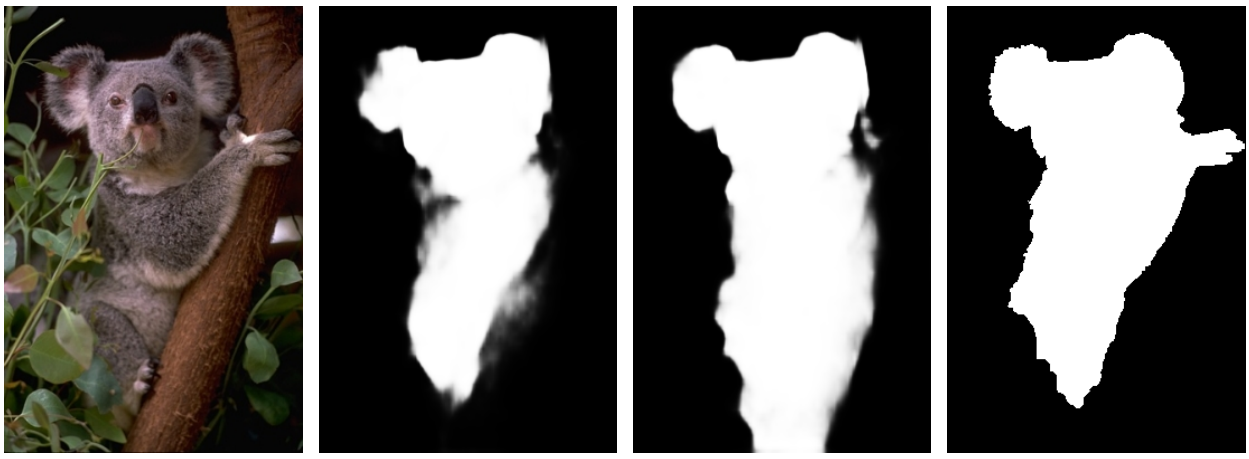
mean_mae	0.03847
mean_fb	0.986

效果对比图：



Origin F3Net F3Net-Edge GroundTruth

图 6.10: Comparison between two nets



Origin F3Net F3Net-Edge GroundTruth

图 6.11: Comparison between two nets

实验结果表明良好的显著性边缘检测结果可以帮助显著性目标检测任务进行分割和定位。

7 消融实验

7.1 EGNet 消融实验

Method	Overall-MAE	Mean-MAE	Mean-Fb	Max-Fb
EGNet	12.760289	0.04253	0.9855	0.9997
EGNet-StructureLoss	11.653789	0.03885	0.9863	0.9993
EGNet-CFM	12.553570	0.04183	0.9858	0.9997
EGNet-CFM+StructureLoss	12.129940	0.04043	0.9851	0.9996

表 2: EGNet 相关方法的实验结果

首先，对比原始 EGNet（Overall-MAE 为 12.760289，Mean-MAE 为 0.04253）和引入结构化损失的 EGNet-StructureLoss（Overall-MAE 为 11.653789，Mean-MAE 为 0.03885），可以看出引入结构化损失显著降低了误差。这表明结构化损失能够更好地监督显著性目标的整体结构，提高检测精度。理论上，结构化损失通过引入全局结构信息，减少了独立像素损失的局限性，从而提升了模型的整体性能。

其次，对比原始 EGNet 和引入 CFM 模块的 EGNet-CFM（Overall-MAE 为 12.553570，Mean-MAE 为 0.04183），可以看到 CFM 模块也对模型性能有所提升。CFM 模块通过元素乘法融合不同层次的特征，剔除冗余信息，避免特征污染，从而提高了显著目标的边界检测能力。理论上，这种多尺度特征的选择性融合能够减少背景噪声，提高边界细化效果。

最后，结合结构化损失和 CFM 模块的 EGNet-CFM+StructureLoss（Overall-MAE 为 12.129940，Mean-MAE 为 0.04043），进一步验证了这两种改进的有效性。虽然 EGNet-CFM+StructureLoss 的 Mean-MAE 略高于 EGNet-StructureLoss，但总体性能仍优于原始 EGNet。这说明结构化损失和 CFM 模块的结合能够综合提升模型的检测精度和鲁棒性。

7.2 F3Net 消融实验

Method	Overall-MAE	Mean-MAE	Mean-Fb	Max-Fb
F3Net	12.113106	0.04038	0.9862	0.9995
Edge-F3Net	11.541758	0.03847	0.9863	0.9996

表 3: F3Net 相关方法的实验结果

从结果中可以看出，引入显著性边缘信息的 Edge-F3Net（Overall-MAE 为 11.541758，Mean-MAE 为 0.03847）在各项指标上均优于原始 F3Net（Overall-MAE 为 12.113106，Mean-MAE 为 0.04038）。这表明显著性边缘信息能够有效地增强模型在边缘细节处的检测性能，减少误检和漏检现象。

理论上，显著性边缘信息通过提供额外的边界约束，使得模型能够更准确地识别显著目标的边缘。显著目标的边缘通常具有较高的梯度变化，引入这些信息可以帮助模型更好地区分显著性目标和背景，从而提高检测精度。此外，边缘信息的引入还能够增强模型的鲁棒性，使其在复杂背景下仍能保持良好的检测效果。

8 未来展望

在接下来的工作中，我们打算从两方面对现有的模型进行改进：

一方面，我们打算使用 Res2Net[1] 来替换现有的 ResNet50，Res2Net 通过在单个残差块内构建分层的类似残差连接，从而可以在细粒度层面上表示多尺度特征，增加了每个网络层的感受野范围。

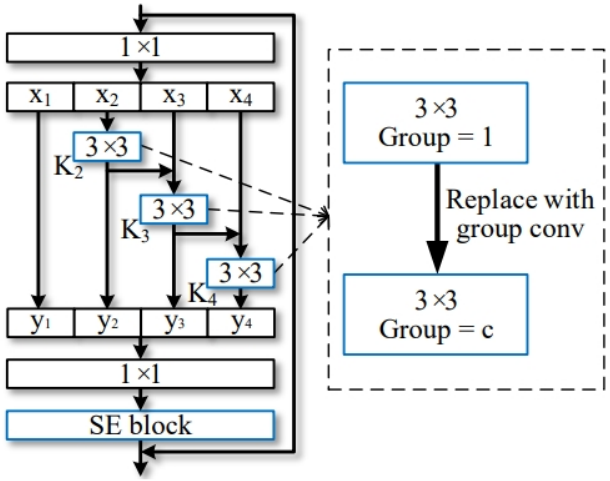


图 8.12: Res2Net[1] 核心模块

我们认为这种多尺度的信息对模型区分显著目标和背景有很大的增益，我们发现已经有工作Res2Net-PoolNet将 Res2Net 应用在显著目标检测领域并取得了不错的效果，我们接下来会将 Res2Net 应用在 EGNet 中从而提升 EGNet 检测显著性目标的能力。

另一方面，我们认为深度信息在显著性检测的过程中可能发挥着至关重要的作用，我们之前已经在 EGNet 的基础上探索了将深度信息融合进来，主要是借鉴了 BBSNet[5] 中对于深度信息处理和融合的 DEM 模块，并在实验结果中发现有一定的效果。

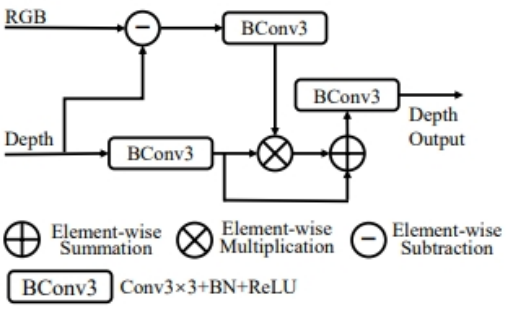


Fig. 4. **Architecture of the depth adapter module (DAM).**

图 8.13: BBSNet[5] 中深度信息和 RGB 信息融合模块

所以我们将在未来的工作中继续深入探索如何高效获得图像的深度信息以及如何高效地将深度信息融合到模型之中，从而能够更好地进行显著性目标检测/分割。

9 实验总结

在本次实验中，我们探讨了显著性目标检测这一计算机视觉领域的重要任务，重点研究了三篇基准论文——EGNet、PoolNet 和 F³Net，并在此基础上进行了进一步的实验和改进。显著性目标检测通过识别和定位图像中引人注目的目标，广泛应用于图像处理、图像搜索和智能监控等领域。

首先，我们复现并对比了 EGNet 和 PoolNet 的实验结果，发现 EGNet 在融合显著目标信息和边缘信息方面表现出色，而 PoolNet 通过基于池化的设计实现了高效的多尺度上下文信息整合。基于此，我们提出了改进方案，包括优化显著性损失函数、引入结构化损失、和 CFM 模块的优化融合等，实验结果显示这些改进有效提升了模型的检测性能。

接着，我们探讨了 F³Net 的基础实验结果，并提出通过引入显著性边缘信息进行监督来优化模型的性能。实验结果表明，引入边缘信息可以显著提升模型的鲁棒性和泛化能力，进一步增强了显著性目标的边缘检测效果。

最后，我们进行了消融实验，验证了不同模块对整体模型性能的贡献。结合实验结果，我们提出了未来改进的方向，包括引入 Res2Net 以增强多尺度特征表示，以及探索深度信息在显著性检测中的应用。我们相信，通过这些改进，显著性目标检测模型的性能将进一步提升，助力更多实际应用场景。

整体而言，本次实验不仅验证了现有显著性目标检测方法的有效性，还通过提出改进方案和未来展望，展示了该领域的巨大潜力和发展方向。收获颇丰。

我们的代码开源在：DeepLearningFinal.

参考文献

[1] Shang-Hua Gao, Ming-Ming Cheng, Kai Zhao, Xin-Yu Zhang, Ming-Hsuan Yang, and Philip Torr. Res2net: A new multi-scale backbone architecture. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(2):652–662, February 2021.

[2] Jiang-Jiang Liu, Qibin Hou, Ming-Ming Cheng, Jiashi Feng, and Jianmin Jiang. A simple pooling-based design for real-time salient object detection, 2019.

[3] Jun Wei, Shuhui Wang, and Qingming Huang. F3net: Fusion, feedback and focus for salient object detection, 2019.

[4] Lihe Yang, Bingyi Kang, Zilong Huang, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao. Depth anything: Unleashing the power of large-scale unlabeled data, 2024.

[5] Yingjie Zhai, Deng-Ping Fan, Jufeng Yang, Ali Borji, Ling Shao, Junwei Han, and Liang Wang. Bifurcated backbone strategy for rgb-d salient object detection. *IEEE Transactions on Image Processing*, 30:8727–8742, 2021.

[6] Jia-Xing Zhao, Jiangjiang Liu, Den-Ping Fan, Yang Cao, Jufeng Yang, and Ming-Ming Cheng. Egnet:edge guidance network for salient object detection, 2019.