

一、实验方法

1.1 数据集

我们使用了一个包含多种动物图像的数据集。每张图像包含一个或多个动物，标签图像标注了每个动物的像素级别区域。数据集分为训练集和测试集，其中训练集包含 167(标记的图像: 40) 张图像，测试集包含 40 张图像。

1.2 数据预处理

数据预处理主要包括以下几个步骤：将图片先进行标签的处理，即换成 *jpg* 格式；再为了保证输入图片的尺寸的一致，且为了不失真，将图片贴在一个 *mask* 矩形上，即完成等比缩放，再用 *transform* 进行归一化。

1.3 模型架构

我们使用 U-Net 网络进行图像的语义分割。U-Net 是一种基于卷积神经网络的图像分割模型，具有对称的编码器和解码器结构，能够有效捕捉图像中的上下文信息和细节信息。U-Net 的结构如下图，其中具体内容如下：

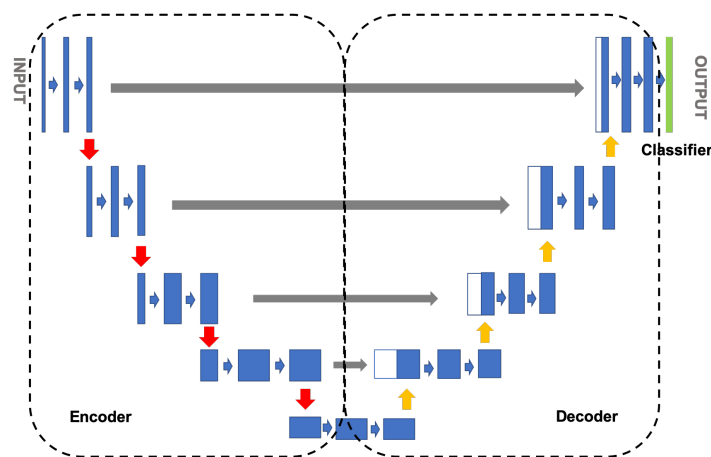


图 1-1: Unet 网络架构

- Encoder: 对输入图像数据进行两次卷积后，再进行下采样，以不断实现特征提取，同时在过程中获得五个特征层。
- Decoder: 对 Encoder 得到的图像数据进行特征融合，即将 Encoder 得到的五个特征层依次进行上采样并与另一个特征层进行堆叠，再进行两次卷积以逐步恢复图像大小。

1.4 模型训练

在使用 *DataLoader* 进行数据载入之后，将图片输入网络后与样本数据得到交叉熵损失，之后进行反向传播，采用 *Adam* 优化器更新梯度。最后输出损失，并每隔 5 个批次保存权重。

二、实验结果

2.1 分割结果

以下是模型在测试集上进行语义分割的部分结果。

(运行失败了，这里就没有有效图片了，哭)

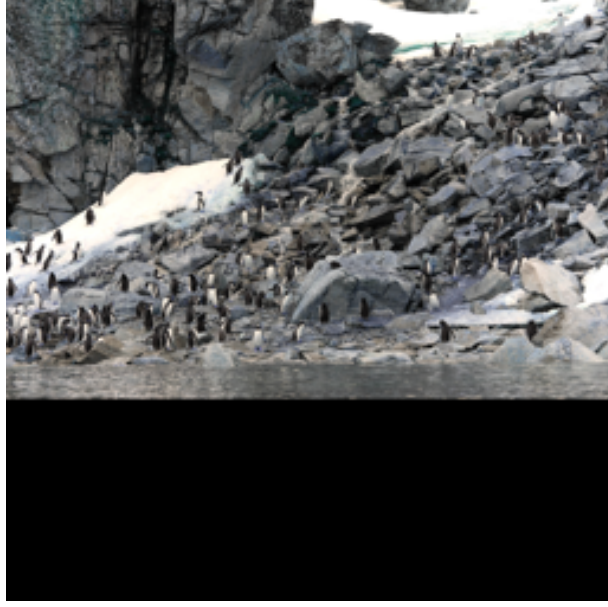


图 2-2: result1

2.2 评价指标

为了评估模型的性能，可采用以下的评价指标：

- **准确率 (Accuracy):** 计算正确预测的像素与总像素数的比值：

$$PixelAccuracy = \frac{N_{Right}}{N_{All}}$$

其中， N_{Right} 为正确预测的像素数，而 N_{All} 为全部像素数；

- **交并比 (IoU):** 计算预测区域与真实区域的交集与并集的比例：

$$IoU = \frac{S_{Inter}}{S_{Union}}$$

$$mIoU = \frac{1}{N} \sum_{i=1}^N IoU_i$$

其中， S_{Inter} 为预测区域与真实区域重合部分， S_{Union} 为预测区域与真实区域的并集，且 N 为所有类型区域的总数，而 IoU_i 则为类别 i 的 IoU，mIoU 为所有类别的平均 IoU。

- **Dice 系数 (Dice Coefficient):** 衡量预测区域与真实区域的相似度，值越大表示越相似：

$$Dice = \frac{2S_{Inter}}{S_{Pred} + S_{Truth}}$$

$$mDice = \frac{1}{N} \sum_{i=1}^N Dice_i$$

其中, S_{Inter} 为预测区域与真实区域重合部分面积, $S_{Pred} + S_{Truth}$ 为预测区域与真实区域的面积和, 且 N 为所有类型区域的总数, 而 $Dice_i$ 则为类别 i 的 Dice, $mDice$ 为所有类别的平均 Dice。

- **F1-score:** 计算精确率和召回率的调和均值:

$$P_{Prec} = \frac{S_{Truth1}}{S_{Pred}}$$

其中, P_{Prec} 即为计算精确率, S_{Pred} 为所有预测为某一类别的像素数, S_{Truth1} 为在预测为对应类别像素中真正属于该类别的像素数;

$$P_{Recall} = \frac{S_{Truth2}}{S_{Truth}}$$

其中, P_{Recall} 即为召回率, S_{Truth} 为所有真实属于某一类别的像素数, S_{Truth2} 为真实属于该类别的像素中, 别正确预测的像素数;

$$P_{Average} = \frac{2 \times P_{Prec} \times P_{Recall}}{P_{Prec} + P_{Recall}}$$

其中, $P_{Average}$ 即为 F1-score, 可以综合考虑精度和召回率。

三、 讨论与总结

3.1 讨论

从实验结果来看, 模型并没有能够成功完成语义分割任务, 可能是由于数据集标记效果较差或是网络本身对特征信息的学习能力较差导致。

3.2 模型优缺点

- **优点:**
 - U-Net 模型本身能够较好地捕捉图像中的全局和局部特征。
- **缺点:**
 - 对于小物体的分割效果较差, 可能需要更多的数据增强策略来提升性能。在训练过程的效果图中会发现对小物体的特征学习效果很差。

3.3 改进方向

为进一步提高模型性能, 可以考虑以下改进:

- 使用更大的数据集或更丰富的标签信息进行训练。
- 在模型中加入更多的多尺度特征提取模块, 以提升对小物体的分割能力。

四、 结论

本文采用 U-Net 模型进行语义分割实验, 评估了模型在多种动物图像上的表现。但由于结果并未取得较好成果, 评价指标也无法开展工作。因此, 未来的工作可以通过数据增强和多尺度特征提取, 以及对网络本身进行优化等方法进一步提高模型的分割效果。

参考文献

- [1] 深度学习麋了鹿, 图像分割 *UNet* 硬核讲解 (带你手撸 *unet* 代码)
https://www.bilibili.com/video/BV11341127iK/?p=3&share_source=copy_webvd_source=3c088cc2b450dc56f0ec41bdcd51b772.
- [2] Bubbliiiing,[Pytorch 搭建自己的 Unet 语义分割平台 (Bubbliiiing 深度学习教程)]
https://www.bilibili.com/video/BV1rz4y117rR/?share_source=copywebvdsouce=3c088cc2b450dc56f0ec41bdcd51b772

A 代码部分由 GPT 生成与修改部分

脚本部分已标注 GPT 生成部分, 其余代码中部分参考网页资料得到, 并在此基础上自己进行修改, GPT 修改部分已标注: 其中注释部分几乎均为 GPT 添加。

B 论文部分由 GPT 生成与修改部分

评价指标和模型优缺点由 GPT 生成, 以及改进方向中的第二点也由 GPT 生成