

用于神经元图像的细节敏感的外廓分割方法

队伍名: deadline

成员名: 杨丽鹤

南京大学 liwe.yang.cs@gmail.com

简单赛道第二名, 得分: 0.56213

复杂赛道第五名, 得分: 0.60066

摘要

比赛的任务是根据神经元的电镜图片，像素级地标注每个神经元的外廓，主要涉及到计算机视觉中的边缘检测任务。因此，主要参考了一些边缘检测的技术来完成该任务，通过搭建深度学习网络并利用标注好的数据来训练它。简单赛道由于训练数据很少，采用的网络结构是以 ResNet-50[4]作为骨干网络的 CASENet[1]，主要思想是利用 ResNet 作为特征提取器，在提取出来的多个层次的特征上分别加一个分支结构，用来检测边缘。其优点主要在于可以融合多个尺度的特征，从而能够识别不同尺度的神经元并进行边缘检测。复杂赛道训练数据较多，且尺寸很大，需要先做裁剪的预处理，采用的网络结构是以 ResNet-152[4]作为骨干网络的 DFF[2]，DFF 是对 CASENet 的一种改进，CASENet 对特征图每个位置的特征采用了相同的权重进行融合，而 DFF 认为每个位置是不一样的，应该对每个位置采用动态学习的权重来融合多个层次的特征。在测试阶段，采用测试时增强（test-time augmentation）策略，融合多个增广方式下的预测结果，增强结果的可靠性。

关键词：

边缘检测、损失函数、CASENet、DFF、多尺度

介绍

本次比赛属于医学图像的边缘检测任务。医学图像和自然图像的区别主要表现在数量、图像尺寸、多/单模态等方面。其中，关于数据量和图像尺寸这两方面在本次比赛中都得到了体现。简单赛道只包含 30 张 1024x1024 的训练图像，图像尺寸不大并且数据量很小，因此需要选择大小合适的模型，并且合理利用尽可能多的数据增广方式，避免过拟合。而复杂赛道中，虽然训练数据只有 20 张图像，但是每张图像尺寸是 9958x9959，可以裁剪得到很多的子图像，因此需要注重模型的学习能力以及设计合适的裁剪策略。

在模型方面，U-Net[7]是常用的医学图像分割模型，因为其下采样和上采样之间的 skip connection 机制，使得其能够恢复出相对较精确的边缘。而 CASENet[1]和 DFF[2]是近年被提出的用于解决语义边缘分割的模型，由于不包含连续多次的反卷积，因此复杂度小于 U-Net，在实验对比后，选择用 DFF 和 CASENet 分别作为简单赛道和复杂赛道的模型。不同的 backbone 甚至相同 backbone 不同的预训练参数也会对最终模型的收敛位置有很大的影响，实验中主要对比了 ResNet[4]系列及其一些变种在本次数据集上的表现，最终选择了 ResNet-50 和 ResNet-152 分别作为简单赛道和复杂赛道的 backbone。

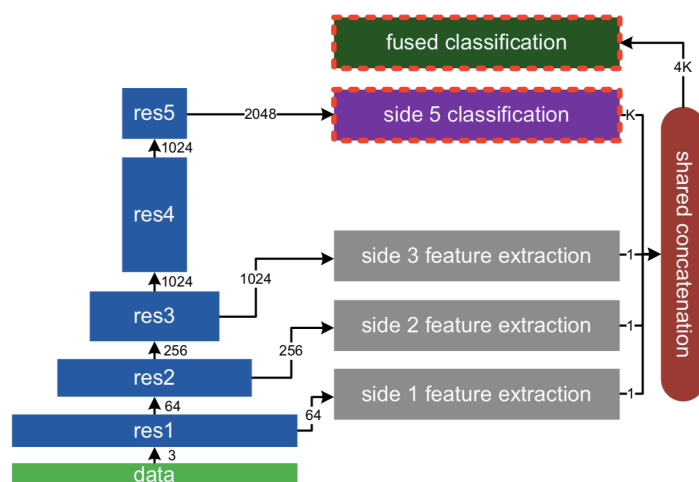
方法说明

简单赛道

数据预处理

对于 groundtruth label 进行二值化处理，以 122 作为阈值；对输入图像进行归一化。

网络结构



CASENet 结构图

以 ResNet-50 作为骨干网络的 CASENet[1]。CASENet 是在语义边缘检测任务中被提出的，可以识别多类别的边缘，此任务是二类别的边缘检测，不需要进一步识别出边缘的语义信息，但对于边缘细节的准确性要求很高，因此对 ResNet-50 作了少许改进，将第一层卷积层的 stride 设置为 1，保留 max pooling 层，并将 layer1-layer4 的 stride 分别设为 1、1、2、2，因此最终的 output_strides 为 8，主要目的是尽量延迟 downsample，可以让网络前部分得到的特征图上保留更多的细节信息。

ResNet 共有 5 个阶段，构成 5 个层次的特征图，原 CASENet 网络选择了其中四个层次的特征图进行融合，没有利用倒数第 2 层次的特征图，本次比赛中采用的 CASENet 利用了所有的 5 个层次的特征图，以此获得更加丰富的特征。

损失函数

结合了 2 种损失函数，分别为 F1-score loss 和 focal loss[3]。由于评测指标是边缘的 F1-score，因此将 F1-score 的计算方式转换为了对应的损失函数的计算；由于边缘检测的正负样本的数量极度不平衡，无法利用常规的 binary cross entropy loss，一种改进方式是计算正负样本的比例，从而在 loss 上分配不同的权重，另一种改进方式是通过网络的预测值采用动态的权重，具体的实现就是 focal loss，本任务中采用的是后者，即 focal loss。

CASENet 的论文中，在最顶层的 feature map 上也加了损失函数，但是根据 DFF 文章中的实验发现，这是没有必要的，因此在本次比赛中移除了顶层的损失函数。

超参数

discriminative learning rate: pretrained backbone: 0.0014, others: 0.014

batch size: 4 (7 iterations per epoch)

training epochs: 200

crop size: 960

gamma of focal loss: 2

alpha of focal loss (edge): 0.7

optimizer: Adam

learning rate scheduler: $lr = \text{baselr} * (1 - \text{iter} / \text{total_iter})^{**} \text{power}$, power = 0.9

weight of F1-score loss: 0.5

weight of focal loss: 0.5

threshold of post-processing: 127.5

数据增广

主要包括水平/垂直翻转、90/180/270 度旋转、对比度变换、随机裁剪。

对于随机的 $p \sim U(0, 1)$ ，若 $p < 2/8$ ，水平翻转；若 $2/8 < p < 3/8$ ，垂直翻转；若 $3/8 < p < 4/8$ ，旋转 90 度；若 $4/8 < p < 5/8$ ，旋转 180 度，若 $5/8 < p < 6/8$ ，旋转 270 度；若 $6/8 < p < 8/8$ ，无操作。

对于随机的 $p \sim U(0, 1)$ ，若 $p < 1/2$ ， $\text{image} = \text{image} * c$, $c \sim U(0.9, 1.1)$ ；若 $p > 1/2$ ，无操作。

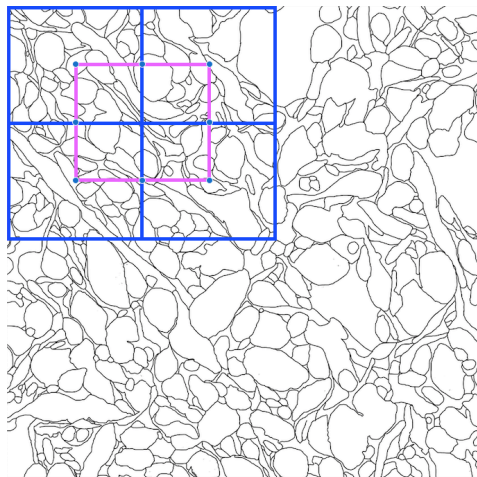
对于每张图像都按照 crop size 进行裁剪。

测试时增广

对原图、水平/竖直翻转、旋转 90/180/270 度这 6 个输入的预测结果做平均，得到最终的预测结果。

复杂赛道

数据预处理



裁剪示意图

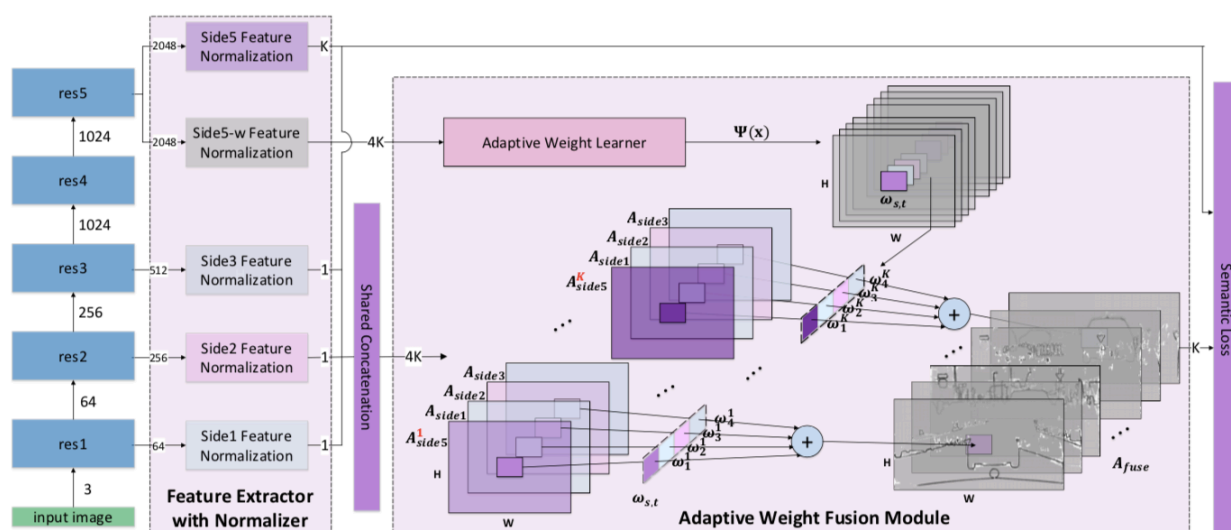
每张图像的尺寸是 9958x9959，因此需要进行裁剪，裁剪出的每个 tile 的尺寸是 1990；并对四个相邻 tile 的中心进行裁剪，具体如上图所示，可以从四个蓝色矩形框中裁出 4 个 tile，但是如果仅以这种方式裁剪，会导致原图中的一些区域无法作为一个整体输入到网络中，因此对于这四个 tile 的中心区域（如红色矩形框所示）也进行裁剪；

对输入图像进行归一化。

网络结构

以 ResNet-152 作为骨干网络的 DFF[2]。类似于简单赛道的做法，对 ResNet-152 作了少许改进，将第一层卷积层的 stride 设置为 1，保留 max pooling 层，并将 layer1-layer4 的 stride 分别设为 1、2、2、1，output_strides 仍为 8，但这样延迟下采样操作可以让网络前部分得到的特征图上保留更多的细节信息。

ResNet 共有 5 个阶段，构成 5 个层次的特征图，原 DFF 网络选择了其中四个层次的特征图进行融合，没有利用倒数第 2 层次的特征图，本次比赛中采用的 DFF 利用了所有的 5 个层次的特征图，以此获得更加丰富的特征。



DFF 示意图

损失函数

结合了 3 种损失函数，分别为 F1-score loss、focal loss 和自己设计的 edge loss，前两个 loss 的设计和简单赛道相同，最后的 edge loss 主要目的是解决检测出来的边缘偏粗的问题，做法是对于 groundtruth 边缘附近的像素，计算他们的 binary cross entropy loss 并给予更大的权重，具体示例如下图，黑色线条是 groundtruth 所标注出的边缘，红色线条是真实边缘附近的一些像素，这些像素容易被错分成边缘，因此为这些位置计算 edge loss，使得网络能够更加关注这些难的区域。



edge loss 区域示意图

超参数

discriminative learning rate: pretrained backbone: 0.0014, others: 0.014

batch size: 4 (203 iterations per epoch)

training epochs: 45

crop size: 1280

gamma of focal loss: 2

alpha of focal loss (edge): 0.7

optimizer: Adam

learning rate scheduler: $lr = \text{baselr} * (1 - \text{iter} / \text{total_iter})^{**} \text{power}$, power = 0.9

weight of F1-score loss: 0.5

weight of focal loss: 0.5

weight of edge loss: 0.4

threshold of post-processing: 127.5

数据增广

主要包括水平/垂直翻转、90/180/270 度旋转、对比度变换、随机裁剪。

对于随机的 $p \sim U(0, 1)$, 若 $p < 2/8$, 水平翻转; 若 $2/8 < p < 3/8$, 垂直翻转; 若 $3/8 < p < 4/8$, 旋转 90 度; 若 $4/8 < p < 5/8$, 旋转 180 度, 若 $5/8 < p < 6/8$, 旋转 270 度; 若 $6/8 < p < 8/8$, 无操作。

对于随机的 $p \sim U(0, 1)$, 若 $p < 1/2$, $\text{image} = \text{image} * c$, $c \sim U(0.9, 1.1)$; 若 $p > 1/2$, 无操作。

对于每张图像都按照 crop size 进行裁剪。

测试时增广

对原图、水平/竖直翻转、旋转 90/180/270 度这 6 个输入的预测结果做平均，得到最终的预测结果。

结论

对于简单赛道，个人认为最大的挑战在于数据量，只有 30 张 1024x1024 的训练图像，合适的数据增广策略和模型的选择至关重要。并且，由于实验任务是边缘检测，因此需要尽量不损失细节，不同 stride 的设计也会产生很大的区别。

对于复杂赛道，训练数据比较充分，模型的复杂度可以相应提高，模型的设计以及 loss 函数的设计对于最终结果会产生不少影响，最终采用的是 ResNet-152 的结构作为特征提取器，模型复杂度比较高，内存开销也比较大，尝试过其他几种 ResNet 的变种，比如 ResNeXt[6]，Res2Net[5]，但最终的表现都没有原始的 ResNet 表现好。

参考文献

- [1] Yu Z, Feng C, Liu M Y, et al. Casenet: Deep category-aware semantic edge detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 5964-5973.
- [2] Hu Y, Chen Y, Li X, et al. Dynamic feature fusion for semantic edge detection[J]. arXiv preprint arXiv:1902.09104, 2019.
- [3] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2980-2988.
- [4] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
- [5] Gao S, Cheng M M, Zhao K, et al. Res2net: A new multi-scale backbone architecture[J]. IEEE transactions on pattern analysis and machine intelligence, 2019.
- [6] Xie S, Girshick R, Dollár P, et al. Aggregated residual transformations for deep neural networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1492-1500.
- [7] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015: 234-241.