## 实验结果和分析

本次实验将从实验数据和主观感知两个方面来评估实验结果。其中，实验数据是通过上述多个评价指标，如不同mIoU、准确率曲线来评判SegNet模型在训练CamVid数据集上所展现的效果。主观感知则是通过预测出来的图象直观地观察出SegNet网络分割的优劣之处。

在不停改变epochs的次数之后，我们发现，当epochs达到50时，训练模型的效果基本稳定并达到最好，再继续增大epochs值，评价指标mIoU的值基本上并没有太大变化，稳定在60%左右。这说明选择合适的 epochs 数量对于模型的训练效果至关重要。如果 epochs 太少，模型可能还没有充分学习训练数据中的模式，导致欠拟合（underfitting）。另一方面，如果 epochs 太多，模型可能会在训练数据上过度拟合（overfitting），即模型对训练数据中的噪声和细节过于敏感，导致在未见过的数据上表现不佳。为了避免过拟合，我们在 epochs为50 后，连续增加epochs 的值，发现mIoU值没有下降并保持稳定，并且Loss值也保持稳定，则可以停止训练并保留此时的模型权重。mIoU-epoches曲线如图3.2，Loss-epoches曲线如图3.3。

图 3.2 mIoU-epoches曲线

图 3.3 Loss-epoches曲线

在了解了训练最佳epochs之后，我们取epochs值为50，对其训练产生的各项指标的值进行导出，得到以下结果：

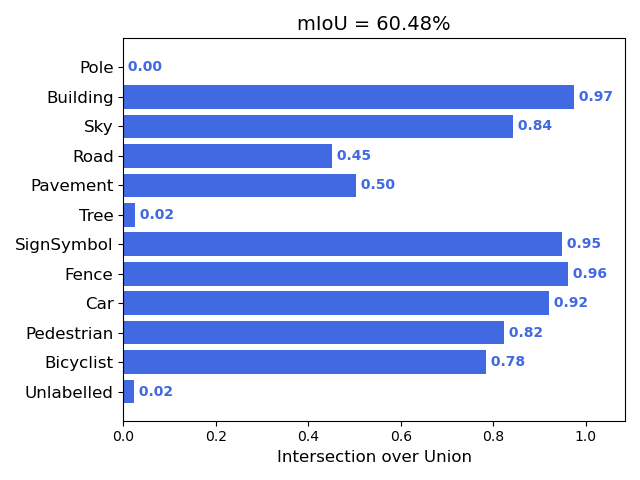


图 3.4 epochs=50时各个标签的Io

由图3.4可以看出，在CamVid数据集 12个分类中，绝大多数标签的交并比IoU是非常高的。其中，Building、SingSymbol、Fence、Car这四个分类IoU值甚至可以达到百分之九十以上；Sky、Pedestrain、Bicyclist这几个类别的交并比也能在百分之七十五以上；其余的Road和Pavement交并比也基本能达到百分之五十；而模型总体的均交并比mIoU达到百分之六十以上。一般约定，在计算机检测任务中，如果，那么结果是可以接受的，就说检测正确。由此表明，该模型的分类效果还是较为可观的。

但是，有几个标签类别的交并比值得我们关注，例如Pole和Tree这两个标签，它们的交并比非常小，Pole类别甚至为0，产生这样的结果主要有以下三个方面的原因：

第一，本次实验选取的数据集不大，在学习模型时发生数据集内标签类别不充分的情况，这就会导致无法学习到有些标签的类别，所以交并比就会很小甚至为零。

第二，在某些情况下，目标物体在图像中的尺寸本身就很小。由于预测框需要覆盖整个目标物体，因此即使预测框的定位相对准确，由于目标物体本身很小，预测框和真实框的交集面积也会很小，从而导致IoU值较低。

第三，模型在训练的过程中会偏向更高效便捷的归类方法，如果某一类在识别的时候只有很小的一部分，模型会更倾向于将这一小类归为更方便识别的大块类别，通俗的说，就是模型偶尔也会“偷懒”，这样也会导致某些类别的IoU值很低。

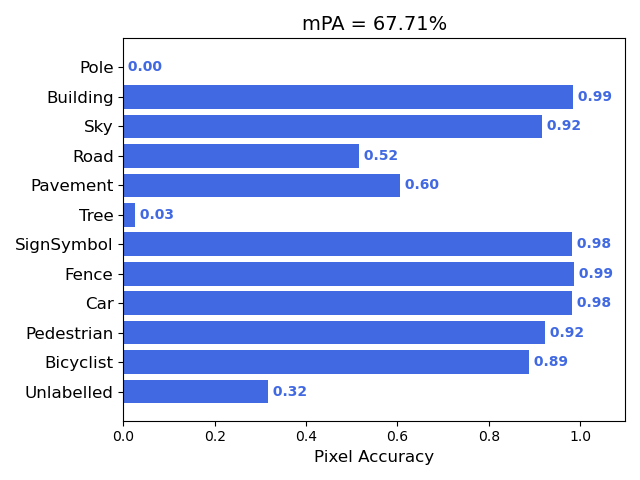


图 3.5 epochs=50时各个标签的PA值

PA值反映了模型预测正确的像素数占总像素数的比例。PA值越高，说明模型预测正确的像素数越多，模型性能越好。例如在图3.5中，“Building”、“Sky”、“SignSymbol”、“Fence“、“Car”、“Pedestrain”等几个分类的PA值都很高，在后面的预测实验中，可以看出，这些分类的物体的分割效果都较好。然而，PA值容易受到图像中类别不平衡的影响。例如，如果某个类别在图像中占据绝大多数像素，即使模型在该类别上的预测性能很差，PA值也可能很高。

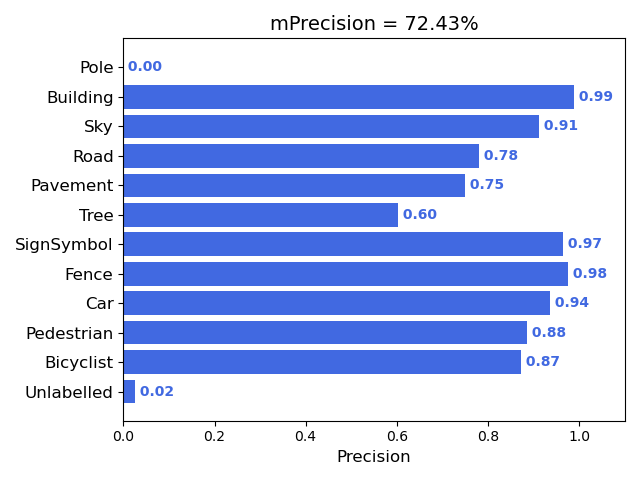


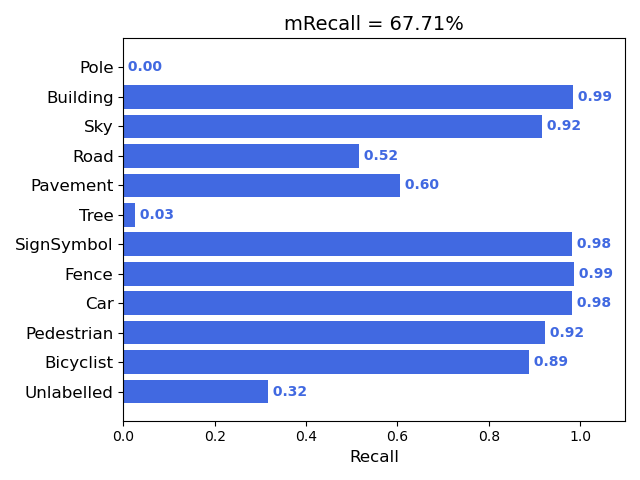
图 3.6 epochs=50时各个标签的Precision值 

图 3.7 epochs=50时各个标签的Recall值

Precision值越高，即精确率越高，说明模型的预测结果越理想。同时，精确率越高的模型，误报情况就越小。然而，在某些情况下，仅仅依赖精确率可能并不全面，因为当数据集中负例样本远大于正例样本时，模型可能将所有样本都预测为负例，从而获得非常高的精确率，但召回率（Recall）会很低。如图3.6，“Tree”这一分类，精确率是60%，但由图3.7可以看出，“Tree”这一标签类别的召回率非常低，只有3%。

Recall值越高，即召回率越高，说明模型对正例的识别能力越强，即模型越能够找出所有真正的正例。然而，在某些情况下，仅仅依赖召回率也可能并不全面，因为当模型过于宽松地将样本预测为正例时，虽然召回率很高，但精确率（Precision）可能会很低，导致预测结果中包含大量的误报。如图3.7，“Unlabelled”召回率是有32%的，但它的精确率却仅有2%。且还有几类样本的召回率是非常低的，如“Pole”、“Tree”，这可能受到了数据集中正、负样本数量差别太大的影响。而召回率高的样本分类也可能是误报的结果。

由此可见，精确率与召回率是两个相互制约的指标，通常，提高精确率可能会降低召回率，反之也是如此。因此，在实际应用中需要根据具体需求来权衡这两个指标。

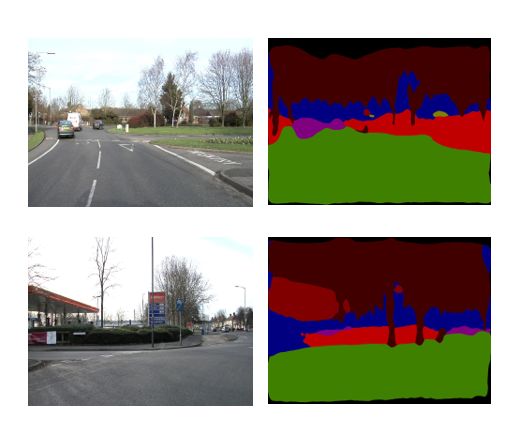


图 3.8 随机道路场景预测结果

由图3.8的结果可以观察出，对于道路场景的预测，SegNet模型的效果还是比较理想的，该模型能够较为清楚地分辨出道路、草丛、汽车等特征明显的类别，并且分割边缘较为平滑。相较SegNet最初提出时所呈现的分割的效果已经有很大的改善。

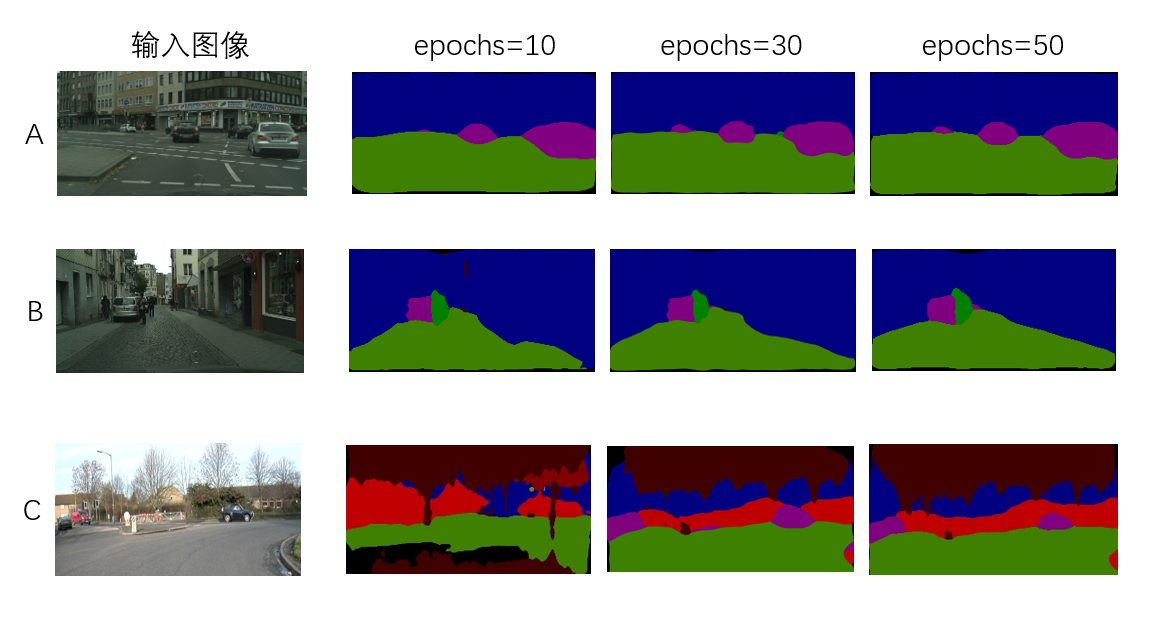


图 3.9 不同epochs下的预测结果

在本次实验中，在预测图像时导入了不同epochs值下的权重，可以发现，当epochs值很小的时候，该模型的预测图像还是较为抽象；分割边缘有很多毛刺的地方，不够清楚和细致；偶尔也会错误识别，比如将大树识别为了背景一类。但当我们适当增大epochs值之后，这种问题就会得到改善，在图3.9中，我们可以看到输入图像A中，当epochs值为10时，左边远处楼房前的小汽车几乎没有被识别到，边缘很模糊，识别出来的分割图像很小。当我们增大epochs到20时，发现模型能将小汽车较为清楚地识别出来，当epochs增大到50时，分割边缘就更加平滑、细致。我们用肉眼可以明显的看出：分割的效果随着epochs的增大而逐渐变好。图B也是如此，随着epochs值增大，路面与建筑之间的分割边缘就更加平滑。在图C中，可以看出epochs很小时，路面大范围被误识，树木也有很大部分未识别出来，汽车几乎一点都没有被识别，这时的分割效果特别差。但是当epochs增大至20、50时，这些很差的效果就在一点点变好，最终产生比较理想的分割图像。

最后，对该模型的训练效率进行分析。从时间上来看，训练一次的时间基本上控制在一小时之内，而通过查阅资料可知，相较PSPnet和FCN，该模型在时间效率上已经有了很大的提升，证明了SegNet模型的有效性。

本章首先介绍了SegNet网络的具体搭建过程，包括使用的框架、函数等具体的搭建思路。使用搭建好的网络进行训练实验，并对实验结果进行具体分析，得出的结论表明SegNet网络具有一定的优越性。

研究结果来自：唐欣雅

Findings from: Xinya Tang