

基于 PyTorch 的 FCN 语义分割网络构建与性能分析

梨花先雪

1. 语义分割算法

1.1 语义分割网络介绍

本实验采用 FCN 语义分割网络。

FCN 的整体网络结构可以分为两部分：**全卷积部分与反卷积（上采样）部分**。其中，全卷积部分基于经典的卷积神经网络结构（VGG），将原有的全连接层替换为卷积层，用于提取图像特征并生成低分辨率的特征图；反卷积部分则通过逐步上采样操作，将特征图还原至与输入图像相同的尺寸，从而得到最终的语义分割结果。

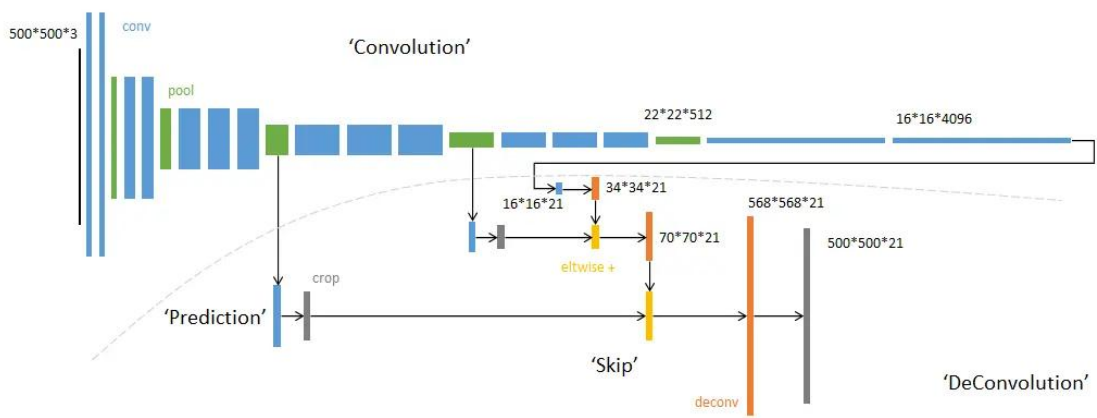


图 1 FCN 网络结构

1. 输入和输出

网络的输入可以为任意尺寸的彩色图像；输出与输入尺寸相同，通道数为： n （目标类别数）+1（背景）。

2. 全卷积

网络在 CNN 卷积部分不用全连接而是替换成 1×1 卷积的目的是允许输入的图片为超过某一尺寸的任意大小。

3. 上采样 Upsampling

在卷积操作后，热力图（heat map）的尺寸通常会变得较小（例如，长宽缩

小至原图的某个比例)。为了恢复到与原图相同的尺寸并获得稠密的像素级预测，通常需要进行上采样。一个直观的解决方案是使用双线性插值，而双线性插值可以通过反向卷积（backward convolution）来实现，具体来说，就是使用固定卷积核的方式进行插值。然而，近年来的研究表明，反向卷积实际上是可以训练的卷积层，通常被称为**转置卷积**（transposed convolution）。在实际应用中，作者并没有采用固定的卷积核，而是将卷积核作为可学习的参数，以便模型能够自动优化这些参数，获得更精确的上采样效果。

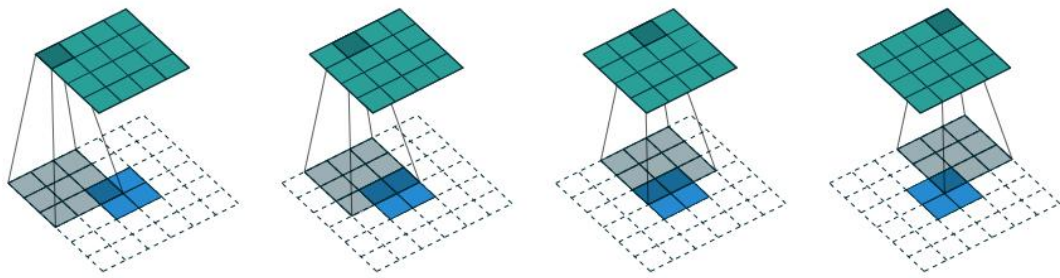


Figure 4.1: The transpose of convolving a 3×3 kernel over a 4×4 input using unit strides (i.e., $i = 4$, $k = 3$, $s = 1$ and $p = 0$). It is equivalent to convolving a 3×3 kernel over a 2×2 input padded with a 2×2 border of zeros using unit strides (i.e., $i' = 2$, $k' = k$, $s' = 1$ and $p' = 2$).

图 2 转置卷积

4. 跳级结构 Skips

在对最后一层的特征图进行上采样以恢复到原图大小时，虽然能够获得粗略的语义分割结果，但由于该特征图的空间分辨率较低，往往会导致细节信息的严重丢失。为了解决这一问题，作者引入了 **跳跃连接**（Skip Connections）的设计。该结构通过将浅层特征（包含更多局部细节）与深层特征（包含更丰富的全局语义）进行融合，从而在保留全局语义信息的同时，增强了对图像边缘与细节的刻画能力，使得最终的分割结果更加精确。

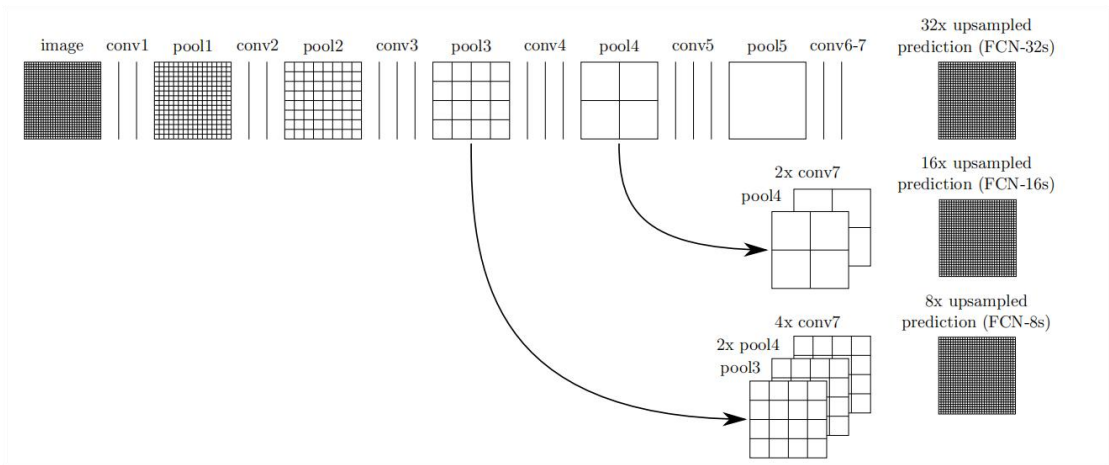


图 3 Skip connection

我们首先对来自底层（stride 32）生成的预测结果（FCN-32s）进行 2 倍上采样，以恢复到原图尺寸，并将其与来自 pool4 层（stride 16）得到的预测进行融合（相加）。这一部分的网络结构被称为 FCN-16s。接着，将 FCN-16s 的预测结果再进行一次 2 倍上采样，并与来自 pool3 层的预测进行融合，形成最终的网路结构，称为 FCN-8s。

本实验将在相同批次大小（batch size=8）条件下，探讨不同学习率（learning rate=1e-3,5e-4,1e-4,5e-5）和训练轮数（epoch=50,100,150）对结果的影响。并从定性（分割可视化结果）和定量（各项评价指标）两个角度，对实验结果进行分析。

1.2 结果分析

1. 定性分析

以下是在不同学习率（learning rate）和训练轮数（epoch）条件下获取的三个样本，以<原图，预测图，真实标签图>顺序呈现。

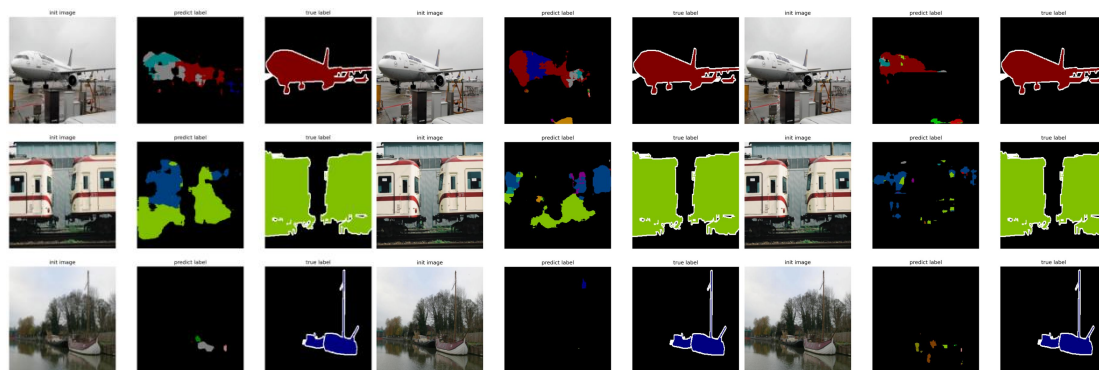


图 4 learning rate=1e-3, epoch=50,100,150

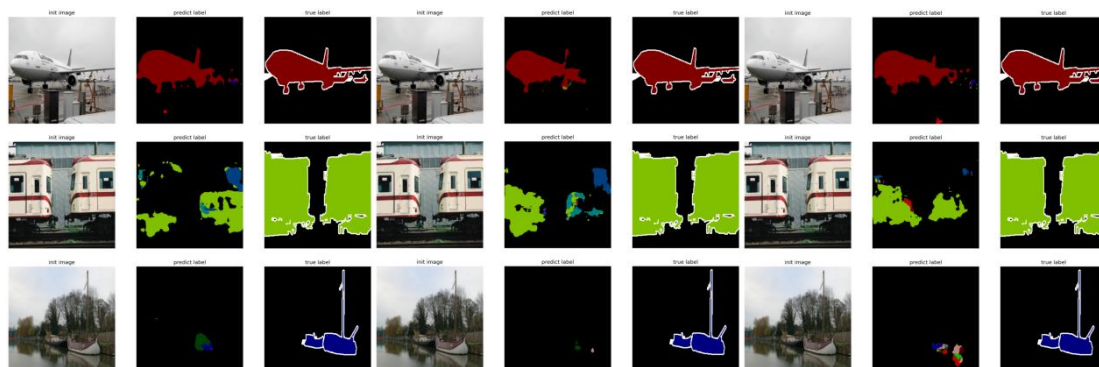


图 5 learning rate=5e-4, epoch=50,100,150

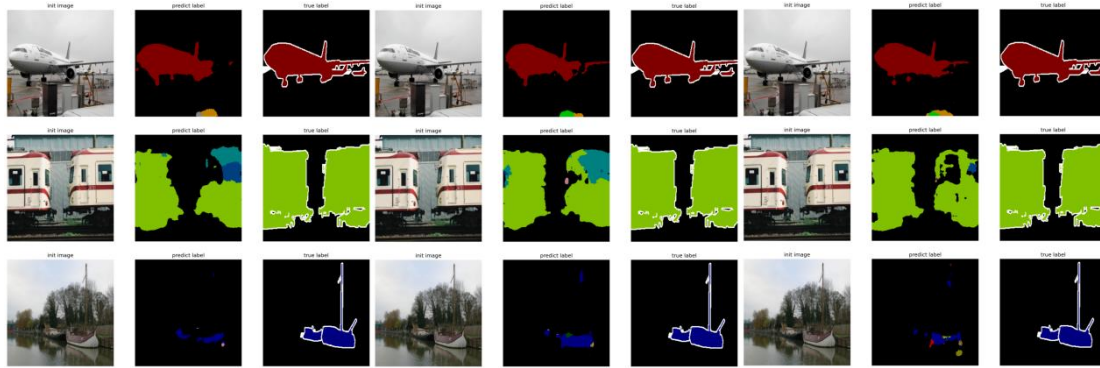


图 6 learning rate=1e-4, epoch=50,100,150

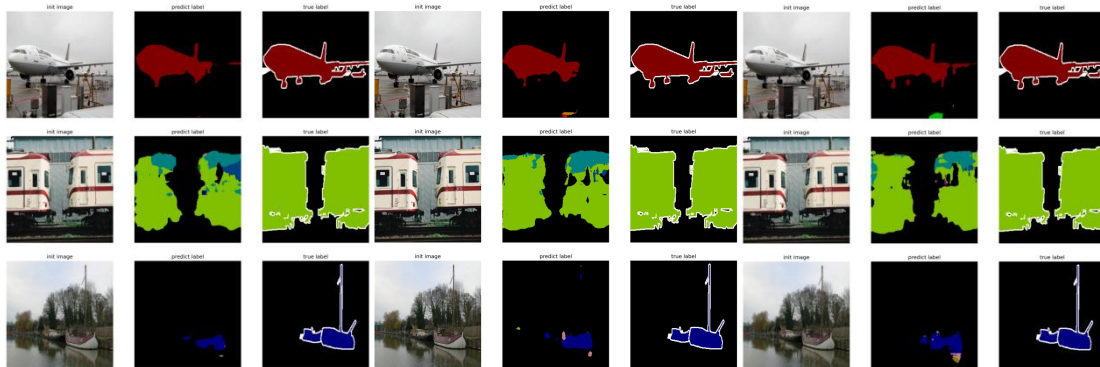


图 7 learning rate=5e-5, epoch=50,100,150

由上图可得，学习率在大于 $5e-4$ 时，随训练轮数增加，预测效果越来越好；学习率在小于 $5e-4$ 时，随训练轮数增加，预测效果在 epoch=100 附近的效果最好。

2. 定性分析

注：HD 是 Hausdorff Distance 的缩写

Learning rate=1e-3

Epoch	mIoU	Dice	HD	Accuracy	Reall	F1 Score
50	0.1177	0.1544	7.6003	0.7419	0.1562	0.1544
100	0.1184	0.1547	7.8453	0.7433	0.1593	0.1547
150	0.1033	0.1360	7.6236	0.7349	0.1355	0.1360

表 1

Learning rate=5e-4

Epoch	mIoU	Dice	HD	Accuracy	Reall	F1 Score
50	0.1606	0.2040	7.0917	0.7803	0.2083	0.2040
100	0.1556	0.1979	7.3167	0.7770	0.2016	0.1979
150	0.1433	0.1844	7.5627	0.7657	0.1876	0.1844

表 2

Learning rate=1e-4

Epoch	mIoU	Dice	HD	Accuracy	Reall	F1 Score
50	0.2071	0.2540	6.3237	0.8204	0.2590	0.2540
100	0.2092	0.2570	6.4719	0.8191	0.2639	0.2570
150	0.2016	0.2495	6.6997	0.8136	0.2553	0.2495

表 3

Learning rate=5e-5

Epoch	mIoU	Dice	HD	Accuracy	Reall	F1 Score
50	0.2037	0.2511	6.2308	0.8194	0.2521	0.2511
100	0.2099	0.2585	6.4529	0.8198	0.2636	0.2585
150	0.2089	0.2569	6.4135	0.8208	0.2594	0.2569

表 4

在本实验中，我们在固定批次大小（batch size）条件下，探究了不同学习率（Learning Rate）和训练轮数（Epoch）对模型性能的影响，评估指标包括 mIoU、Dice、HD、Accuracy、Recall 和 F1 Score，具体结果如表 1 - 4 所示。

从实验结果来看，学习率对模型性能具有显著影响。较大的学习率（如 1e-3）训练过程中模型难以收敛，各项指标表现均较差，mIoU 在不同 Epoch 下均低于 0.12，且随着训练轮数的增加，模型性能反而有所下降，表明可能存在训练不稳定或发散的问题。而当学习率减小至 5e-4 时，模型性能有所提升，mIoU 最高达到 0.1606（在 Epoch 为 50 时），但随着 Epoch 增加，性能略有下降，可能存在一定程度的过拟合或震荡现象。

进一步减小学习率至 1e-4 后，模型在多个评价指标上取得最优结果。特别是在 Epoch 为 100 时，mIoU 达到 0.2092，Dice 和 F1 Score 均为 0.2570，Accuracy 高达 0.8191，整体表现稳定且优越。相比之下，学习率为 5e-5 时虽训练更加平稳，性能也接近 1e-4，但训练收敛速度稍慢。在 Epoch 为 100 时达到 mIoU 最高值 0.2099，F1 Score 达到 0.2585，略优于 1e-4。

综上所述，学习率为 1e-4 时模型整体性能最佳，在训练稳定性与精度之间取得了良好平衡。进一步减小学习率（如 5e-5）虽能带来小幅度性能提升，但训练时间相对较长，适合用于模型的后期微调。从训练轮数的影响来看，100 轮左右通常可以达到较优效果，继续增加轮数对性能提升有限，甚至可能带来轻微的下降。

2. 总结

通过本次图像分割实验，我深入理解了 FCN 网络的结构特点与工作机制，特别是全卷积网络在特征提取与像素级预测中的重要作用。通过设置不同的学习率和训练轮数，我观察到超参数对模型性能具有显著影响，这也让我更加体会到模型调参在深度学习实践中的关键性。

在实验过程中，我不仅掌握了如何基于语义分割任务评估模型性能（如 mIoU、Dice、Accuracy 等指标），也提升了对实验设计、数据可视化以及结果分析的能力。同时，通过调研转置卷积与跳跃连接等机制，我进一步加深了对图像分割模型在保持空间细节与全局语义之间平衡策略的理解。

总体而言，本次实验不仅加深了我对语义分割原理的理解，也锻炼了我在模型搭建、参数调优和结果分析等方面的综合能力。在反复实验和对比中，我体会到合理设置超参数对模型性能的影响，也认识到理论与实践结合对于掌握深度学习方法的重要性。这些收获为我后续在计算机视觉方向的学习和研究打下了坚实的基础。