作业三

概要

原文

Magnetic resonance imaging is a fundamental tool to reach a diagnosis of multiple sclerosis and monitoring its progression. Although several attempts have been made to segment multiple sclerosislesions using artificial intelligence, fully automated analysis is not yet available. State-of-the-art methods rely on slight variations in segmentation architectures (e.g. U-Net, etc.). However, recent research has demonstrated how exploiting temporal-aware features and attention mechanismscan provide a significant boost to traditional architectures. This paper proposes a framework that exploits an augmented U-Net architecture with a convolutional long short-term memory layer and attention mechanism which is able to segment and quantify multiple sclerosis lesions detected in magnetic resonance images. Quantitative and qualitative evaluation on challenging examples demonstrated how the method outperforms previous state-of-the-art approaches, reporting an overall Dice score of 89% and also demonstrating robustness and generalization ability on never seen new test samples of a new dedicated under construction dataset.

翻译

核磁共振成像技术是诊断多发性硬化症及监测情况的一项基本工具。虽然已经可以借助人工智能将多发性硬化症进一步划分，但还没有全自动的分析工具。目前最先进的方法，例如U-net也只是轻微改变了分割的架构。不过我们最近发现，利用时间感知特征和注意力机制可以显著推动传统架构的发展。本文提出了具有长短时记忆卷积层和注意力机制的U-net升级版框架，该框架可以分割核磁共振影像中的病变组织并量化多发性硬化症程度。使用具有挑战性的例子对框架进行定量及定性分析后，得到DICE得分为89%，此外，在从未见过的数据上该框架也体现出了足够的健壮性和泛化能力，充分说明该框架优于以往的方法。

介绍

原文

Multiple Sclerosis(MS) is a chronic inflammatory demyelinating disease of the Central Nervous System (CNS), with neuropathologic features characterized by focal areas of inflammation with myelin and axonal loss. MS lesions may be detected in vivo by Magnetic

Resonance Imaging (MRI) in different areas of the brain and the spinal cord and they accumulate over time. Selective localization of lesions on MRI (periventricular, cortical/iuxtacortical, brain stem/cerebellar, and spinal cord) is also relevant for the diagnosis of MS and detection of new or enlarging lesions at follow-up, and is routinely used in the evaluation of therapeutic response and disease progression. Manual annotation of MS lesions on MRI scans is a time consuming task and requires substantial efforts by specialized experts. Moreover, inter and intra operator variability is unavoidable and may affect accuracy and reproducibility of lesion segmentation Thus, there is an increasing interest today in automation of MRI reading and evaluation to avoid the bias introduced by human raters and to make this information available for routine clinical practice. Typically, the longitudinal brain MRI protocol involves distinct kinds of sequences, which generate different types of images that vary according to the contrast of the various tissues that compose the brain. The most common MRI sequences used to detect MS lesions are the Fluid Attenuated Inversion Recovery (FLAIR), T1-weighted, T2-weighted, and PD-weighted images. In the T1-weighted sequence, white matter appears lighter than gray matter, and cerebrospinal fluid (CSF) appears dark. In the T2-weighted sequence, the white matter appears darker than the gray matter, while the CSF appears bright. FLAIRs images are like T2s, except that CSF is suppressed. MS lesions appears hypointense in T1-w and hyperintense in T2-w, PD-w and FLAIR sequences, with respect to normal tissue intensities. Lesions are most detectable in the FLAIR images, where they appear hyperintense and usually well distinguishable from surrounding tissues. Fig. 1 shows four MRI brain images for each different acquisition types with MS lesions (MS lesions are pointed by red circle). In our method, similarly to other works in the field, the most discriminating MRI sequence (FLAIR) was exploited.

The starting point of our study is one of the most widely used networks in the state of the art for this purpose, the U-Net architecture , which is widely used not only in medical image segmentation, but also in general segmentation tasks. In this work, an extended Fully Convolutional DenseNet (FC-DenseNet) for MS lesion segmentation is proposed; it follows the U-Net structure with the addiction of Long Short-Term Memory (LSTM) layer and extensive usage of attention mechanisms to detect FLAIR-w MS lesions in longitudinal brain MRI. Attention is a technique that aims to mimic the cognitive attention of humans by enforcing neural networks to pay greater attention to most informative input data and ignore the rest. Attention mechanisms have been shown to be effective in capturing global dependencies and have become an integral part of semantic segmentation tasks. The FC-DenseNet has been properly extended with an attention mechanism based on the usage of squeeze and attention blocks, in order to accentuate the group of pixel from the same classes employing different spatial scales. Squeeze and Attention blocks (SA) represent a component that can be easily incorporated within the backbone, able to improve network performance through operations applied on both local and global level. Moreover, the space propagation of the lesions with a similar shape between adjacent images suggested to introduce a Long short-term memory (LSTM) layer ; it permits to preserve spatial information between longitudinal axis of data.

The performance of the proposed architecture was evaluated employing a cross-validation scheme in patients with lesions on follow-up scans. The architecture described in Fig. 4 represents the best results of some tested models described in the ablation studies.

The training phase of a Deep Neural Network architecture typically requires a large amount of labeled images. A relevant issue in MRI lesions segmentation is the presence of just few small example in each dataset available and the lack of homogeneity between different repositories, due to the usage of different scanners and/or acquisition protocols. This makes the segmentation challenging, raising concerns about the results obtained from the different methods, which are difficult to compare and generalize to other datasets. For these reasons, we are actually working on the generation of a new labeled MRI dataset as part of the *"In Silico World (ISW): Lowering barriers to ubiquitous adoption of In Silico Trials"* (Grant agreement ID: 101016503, PROGRAMME: H2020-EU.3.1. - SOCIETAL CHALLENGES - Health, demographic change and well-being, CALL: H2020-SC1-DTH-2018-2020). It will be larger than the actual ones, with heterogeneous samples (patients with different stages of disease) and with labeled MS lesions validated by employing different experts. The proposed method, its future extensions and the under construction labeled dataset will be included into the Universal Immune System Simulator (UISS). UISS is a multi-compartment, multi-scale, polyclonal, stochastic, and patient-specific agent-based model (ABM) that is able to simulate immune system dynamics both in physiological and pathological scenarios .UISS simulator framework has been extended to model MS pathogenesis and its interaction with the host immune system , taking into account both cellular and molecular entities. Particularly, UISS- MS takes into account B cells, T helper (CD4+ T cells), T cytotoxic (CD8+ T cells), conventional dendritic cells (DCs), macrophages (M), plasma B cells (P cells), immunocomplexes(IC), oligodendrocytes(ODC), interferon-gamma (IFN-G), interleukins of type x (IL-x), transforming growth factor beta(TGFB), myelin basic proteins(MBP), immunoglobulins class G (IgG) and chemokines(as generic chemokines). For each modeled patient, the age at MS onset, baseline MRI lesion load, oligoclonal bandsstatus, and the administered treatment are usually considered.

A limit of the current UISS-MS framework is that only qualitative data about MRI lesion load have been inserted. For this reason, the quantitative data about the MRI lesion load obtained with the framework here proposed will be integrated into the UISS framework, with the aim to represent and predict the disease progression of MS patients as well as to more realistically simulate the immune response to specific treatments.

翻译

多发性硬化症MS是慢性中枢神经系统炎症性脱髓鞘的一种，其神经病理学特征为局部性炎症、骨髓磷脂与轴突的丢失。MS在体内的病变可通过核磁共振成像MRI在脑和脊髓的不同区域检测到，且其会逐渐积累。MRI中病变的选择性定位也与MS的诊断、随访时对新发或扩大病灶的检查有关，并且经常用于评估治疗反应和疾病进展。在MRI上手动注释MS病变是一项耗时的工作，需要专专业医师的大量努力。此外，操作者与被操作者的特异性是不可避免的，可能会影响病灶

分割的准确性和可重复性。因此，为了避免人类评分者引入的误差，人们对MRI读取和评估的自动化越来越感兴趣，并将其应用与常规临床实践。通常，纵向脑MRI协议涉及不同类型的序列，这些序列会产生不同类型的图像，这些图像根据组成大脑的各种组织的对比度而变化。用于检测 MS 病变的最常见 MRI 序列是流体衰减反转恢复 （FLAIR）、T1 加权、T2 加权和 PD 加权图像。在 T1 加权序列中，白质看起来比灰质亮，脑脊液（CSF） 看起来很暗。在 T2 加权序列中，白质看起来比灰质更暗，而脑脊液看起来很亮。FLAIR图像类似于T2，只是CSF被抑制。相对于正常组织强度，MS 病变在 T1-w 中为低信号，在 T2-w、PD-w 和 FLAIR 序列中为高信号。病变在FLAIR图像中最容易检测到，它们为高信号，通常与周围组织区分开来。

我们研究的起点是应用与此最广泛的架构之一———U-net架构。它不仅应用与医学图像分割，广义上也可用于图像分割。在这项工作中，提出了一种用于MS病变分割的扩展全卷积密集网（FC-DenseNet）。它遵循U-Net结构，具有长短期记忆（LSTM）层的成瘾性，并广泛使用注意力机制来检测纵向脑MRI中的FLAIR-w MS病变。注意力是一种旨在通过强制神经网络更加关注信息量最大的输入数据而忽略其余数据来模仿人类的注意力的技术。注意力机制已被证明在捕获全局依赖关系方面是有效的，并已成为语义分割任务的一个组成部分.FC-DenseNet已经适当地扩展了基于挤压和注意力块的注意力机制，以突出来自采用不同空间尺度的相同类别的像素组。挤压和注意力块（SA）代表了一个可以轻松合并到骨干中的组件，能够通过在本地和全球层面应用的操作来提高网络性能。此外，相邻图像之间形状相似的病变的空间传播表明引入了长短期记忆（LSTM）层;它允许在数据的纵轴之间保留空间信息。

在随访扫描中，采用交叉验证方案对病变患者评估所提出的架构的性能。

深度神经网络架构的训练阶段通常需要大量的标记图像。病变分割的一个相关问题是，每个可用的数据集中只有几个小例子。并且由于使用不同的扫描仪和/或采集协议，不同存储库之间缺乏同质性，导致这些结果难以比较和推广到其他数据集。这使得分割具有挑战性，引起了人们对从不同方法获得的结果的担忧。出于这些原因，我们实际上正在开发一个新的标记MRI数据集：计算机世界（ISW）。

它将比实际样本大，具有异质性样本（不同疾病阶段的患者）并且通过雇用不同专家标记MS病变。所提出的方法，其未来的扩展和正在建设的标记数据集将被纳入通用免疫系统模拟器（UISS）。UISS是一种多室、多尺度、多克隆、随机和患者特异性药物模型（ABM），能够在生理和病理情况下模拟免疫系统动力学。UISS模拟器框架已扩展到模拟MS发病机制及其与宿主免疫系统的相互作用，同时考虑了细胞和分子实体。特别是，UISS-MS考虑了B细胞，T辅助细胞（CD4 + T细胞），T细胞毒性（CD8 + T细胞），常规树突状细胞（DC），巨噬细胞（M），血浆B细胞（P细胞），免疫复合物（IC），少突胶质细胞（ODC），干扰素-γ（IFN-G），x型白细胞介素（IL-x），转化生长因子β（TGFB），髓磷脂碱性蛋白（MBP），G类免疫球蛋白（IgG）和趋化因子（作为通用趋化因子）。对于每个建模患者，通常考虑MS发病年龄，基线MRI病变负荷，寡克隆带状态和给药治疗。

目前UISS-MS框架的一个局限性是只插入了关于MRI病变负荷的定性数据。出于这个原因，使用此处提出的框架获得的有关MRI病变负荷的定量数据将被整合到UISS框架中，旨在表示和预测MS患者的疾病进展，以及更真实地模拟对特定治疗的免疫反应。

复现

详细复现过程

> 包含复现时踩的坑以及详细改动

本地克隆

首先将文章对应的仓库克隆至本地，大致阅读文件后发现仓库拥有非常工整的结构，最主要的train.py内也为了方便训练添加了许多注释，因此在文件结构上没有做太大改动。

数据准备

文章数据集采用医学图像分割领域一项公开的挑战ISBI中的子集ISBI2015，下载至本地后按照代码内部OS库相关代码推导出应有的结构，并编写了一个简单的程序处理数据，处理好的数据集结构如下：

```
dataset
└──ISBI_2015
    ├──test1
    │   └──Patient1
    ├──test1annot
    │   └──Patient1
    ├──...
    ├──train1
    │   ├──P1_T1
    │   ├──...
    │   ├──...
    │   └──P5_T4
```

> 在处理完数据集后准备阅读作者所留训练方法时，发现作者其实已经上传了整理好的数据集……

数据集内主要分为训练集、面具、验证集、测试集，约170MB。内容主要为核磁共振成像不同切片，面具是为了避免成像中过多的黑色造成干扰引入。

> 仔细阅读代码后，工整的结构和精巧的封装让我感叹大佬就是大佬，反观个人代码犹如一团乱麻。

本地测试

得益于作者的代码健壮性和可移植性极强，训练过程没有任何的阻碍。

本地环境为RTX 3060, pytorch==2.0, cuda==11.7

> 训练时唯一的问题出在了最不应该出问题的anaconda的环境配置和包上，网络错误与solving environment卡顿消耗了大部分时间，最后使用网络代理全盘解决。

使用epoch=1 batch_size=1进行测试后没有出现问题，但是时间过长，五个文件夹共五个epoch耗费大约半个小时，GPU资源并未吃满，但奈何笔记本散热不佳，根据作者所用200epoch来训练大概率要出问题，于是转在gitahub上训练。

线上训练

由于没有租赁实例进行训练的经验，复现时大部分时间都耗在了这一部分。

1. 代码上传

代码按照网站所说上传成功，但是数据集在500MB的限制内，却无法上传成功，经常性死机加网络卡顿失败。个人猜测是图片数量过多导致上传速度过慢所以无响应。

尝试上传压缩包后成功，便在代码文件中加入run.sh，内部包含启动所需命令及解压命令，但解压后训练文件找不到数据集，ssh查看实例文件结构也失败，端口不知为何变成了10.11.1.3，无法连接。

于是重开了一个项目使用ssh与Xftp方式上传，上传的数据集随便选了一张图片，之后开启一个debug任务，并使用xftp上传本地数据集内容进行替换。

> 实例似乎存在限制，无法删除数据集文件夹，只能删除内部内容，所以将上传的文件夹改为和本地同名

2. 在线训练

网站内实例只有对应框架，所以在启动命令中额外安装了scipy, matplotlib两个库，之后犯了一些很蠢的错误，启动命令的改动如下：

- 无法找到数据集，加入 --dataset-path /data/kinox/ISBI_2015
- 训练完毕但是没结果，加入 --results-path /output
- 为了更快测试加入对epoch数的更改
- 训练成功完成，但是输出只有曲线图，没有模型，加入-wp /output

第一次训练完成后在保存曲线图部分出错，检查后发现代码将图片保存至代码文件夹，而这一部分貌似是不可写的，于是将代码中WEIGHT_PATH加上/output，这样可以在输出结果中查看模型与曲线图。

最终的启动命令为

```
pip install matplotlib scipy;python attention-cnn-MS-segmentation/train.py --dataset-path /data/kinox/ISBI_2015 --results-path /output --num-epochs 40 --folders 5 -b 4 -wp /output
```

复现结果

文章中提到在200epoch时达到最佳结果，但作者选用了5个文件夹交叉训练的方式，不确定这200epoch是每个文件夹分开算还是合起来算，所以用不同的epoch数进行训练。

- 1epoch

以fold5为例，训练结果为

```
Epoch 1
Train - Loss: 0.6627, Acc: 0.9876, Dice: 0.2912, Sens: 0.4720, Spec: 0.9905
Train Time 9m 27s
Val - Loss: 0.5820, Acc: 0.9955, Dice: 0.6595, Sens: 0.5789, Spec: 0.9987
Total Time 11m 5s
```

和文章所给结果对比得到

	acc	dice	sens	spec
std	0.9987	0.8190	0.8442	0.9992
mine	0.9955	0.6595	0.5789	0.9987

可见dice, sens值偏差较大，以这两个值作为主要判断可以更好的分析，作者也是主要使用了这两个值。

- 40 epoch

以fold1为例，结果为

	acc	dice	sens	spec
std	0.9971	0.8448	0.8601	0.9983

	acc	dice	sens	spec
mine	0.9986	0.8693	0.8699	0.9993
best	0.9992	0.8900	0.9071	0.9997

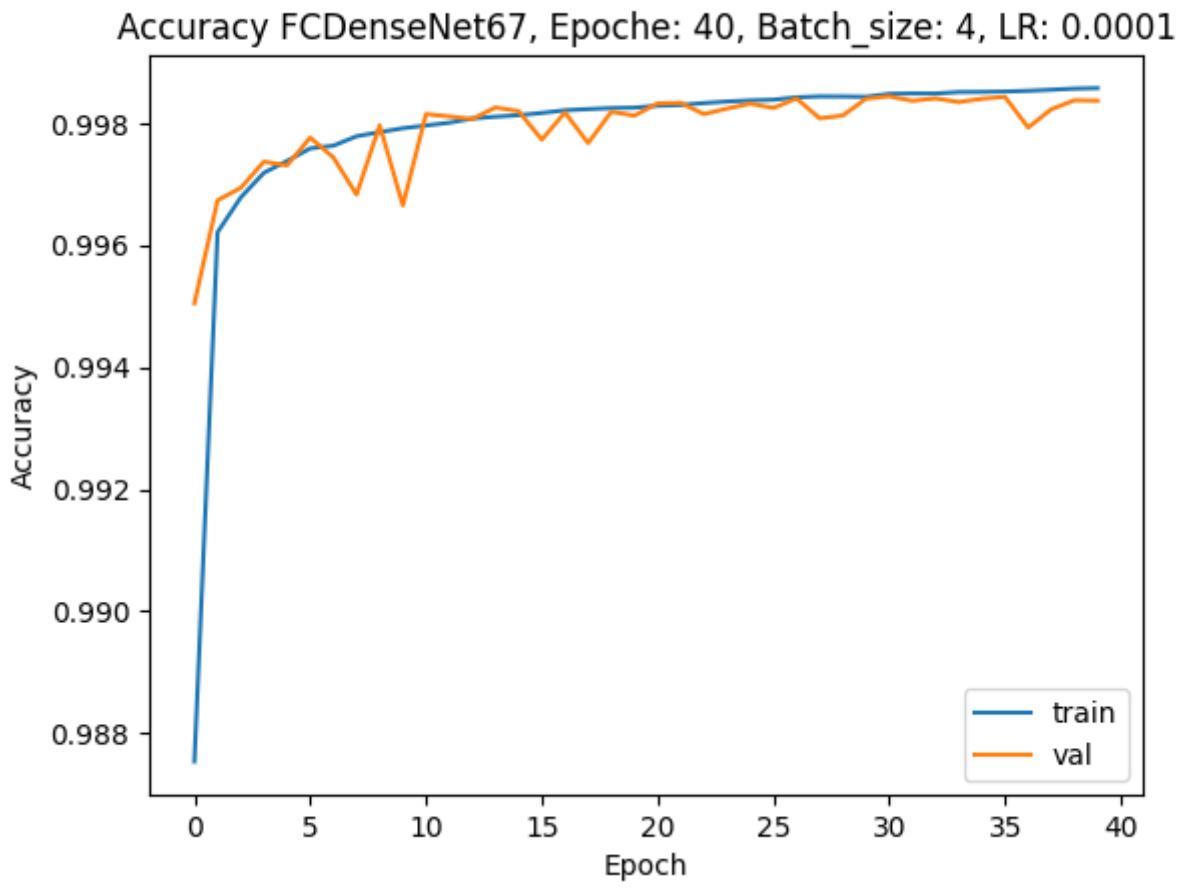
训练结果均好于作者所给对应文件夹训练结果，但均小于5个文件夹内最佳值。

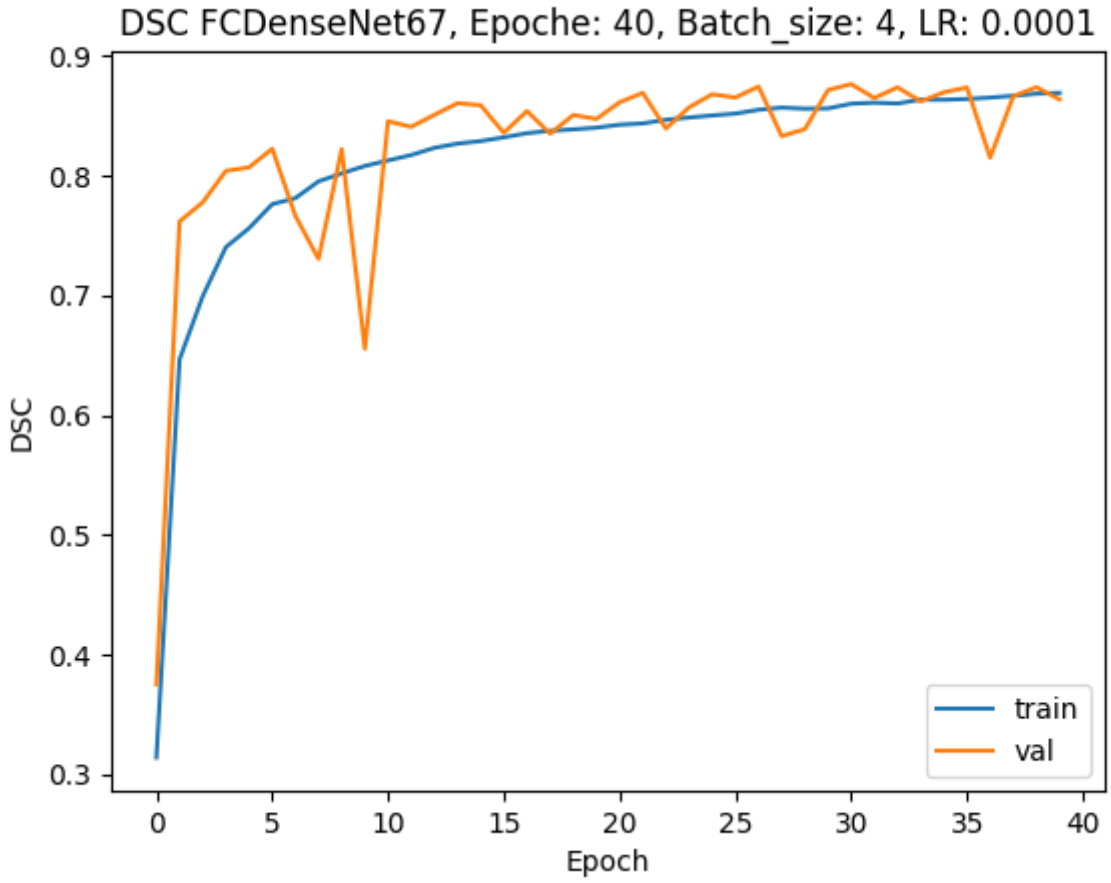
文章中还给出了其他研究方法所得到的结果，其中最佳值为dice==0.81(U-net),sens==0.79(sens)。

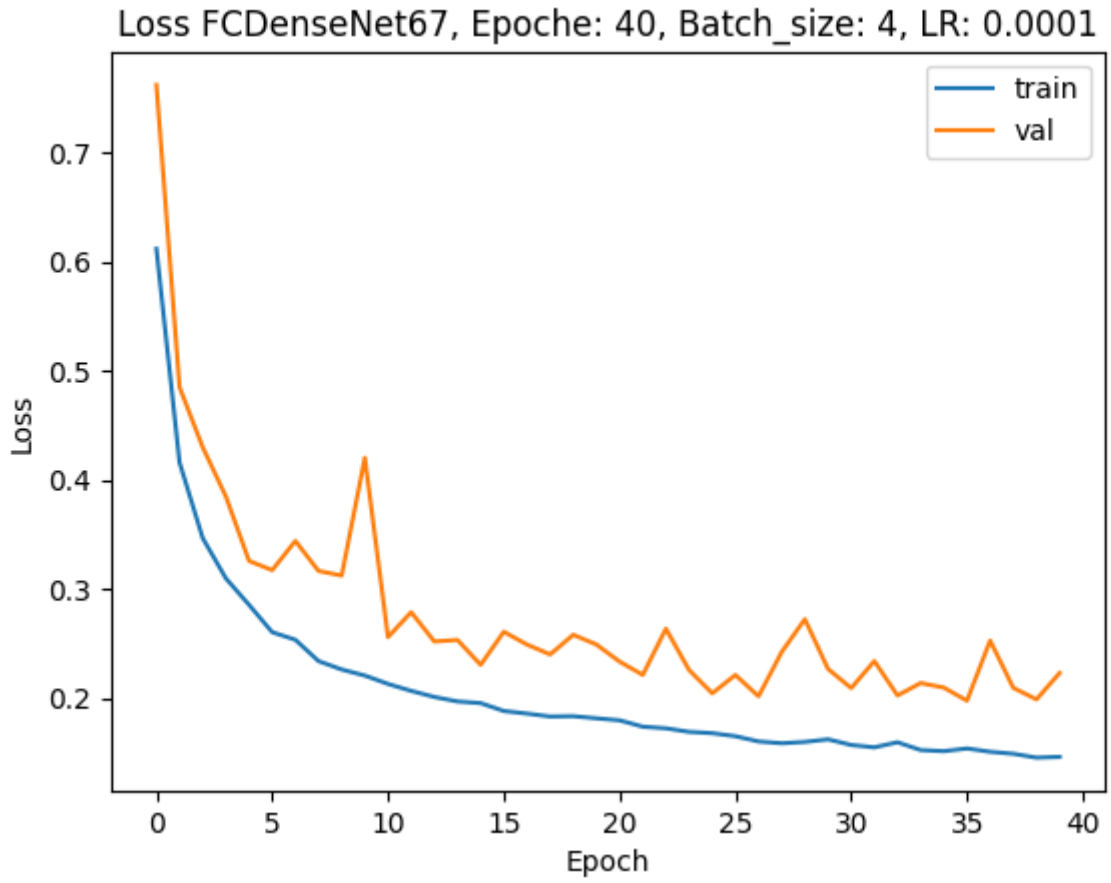
可以发现40epoch时虽然没有达到作者的最好结果，但已经体现出了优越性

以下为评估指标折线图

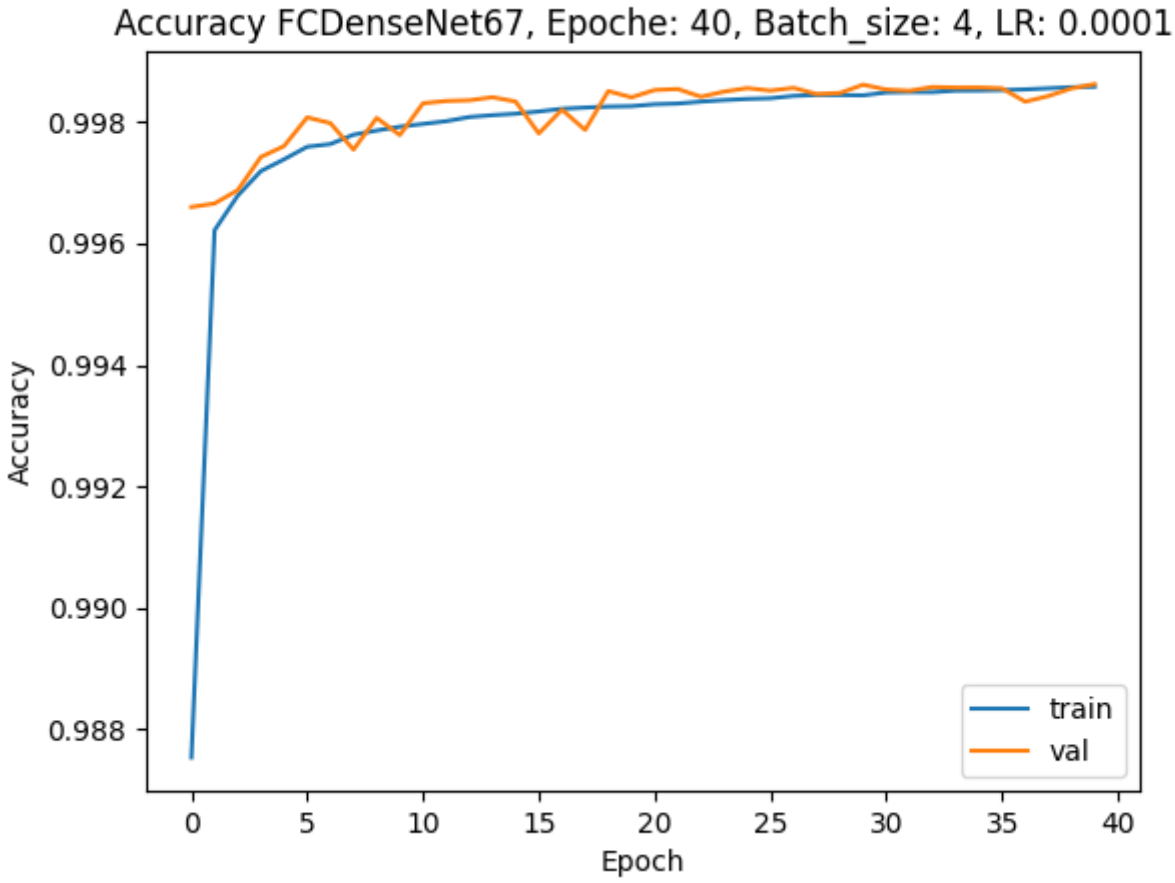
fold1

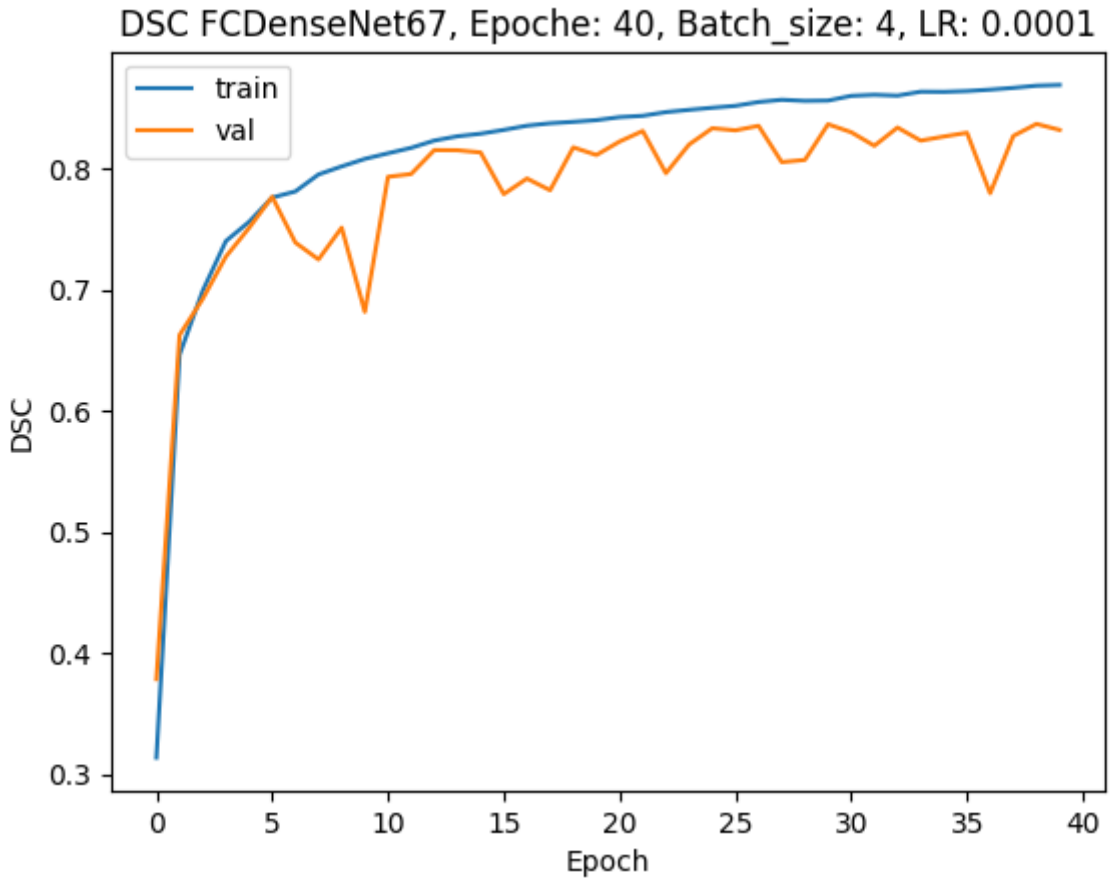


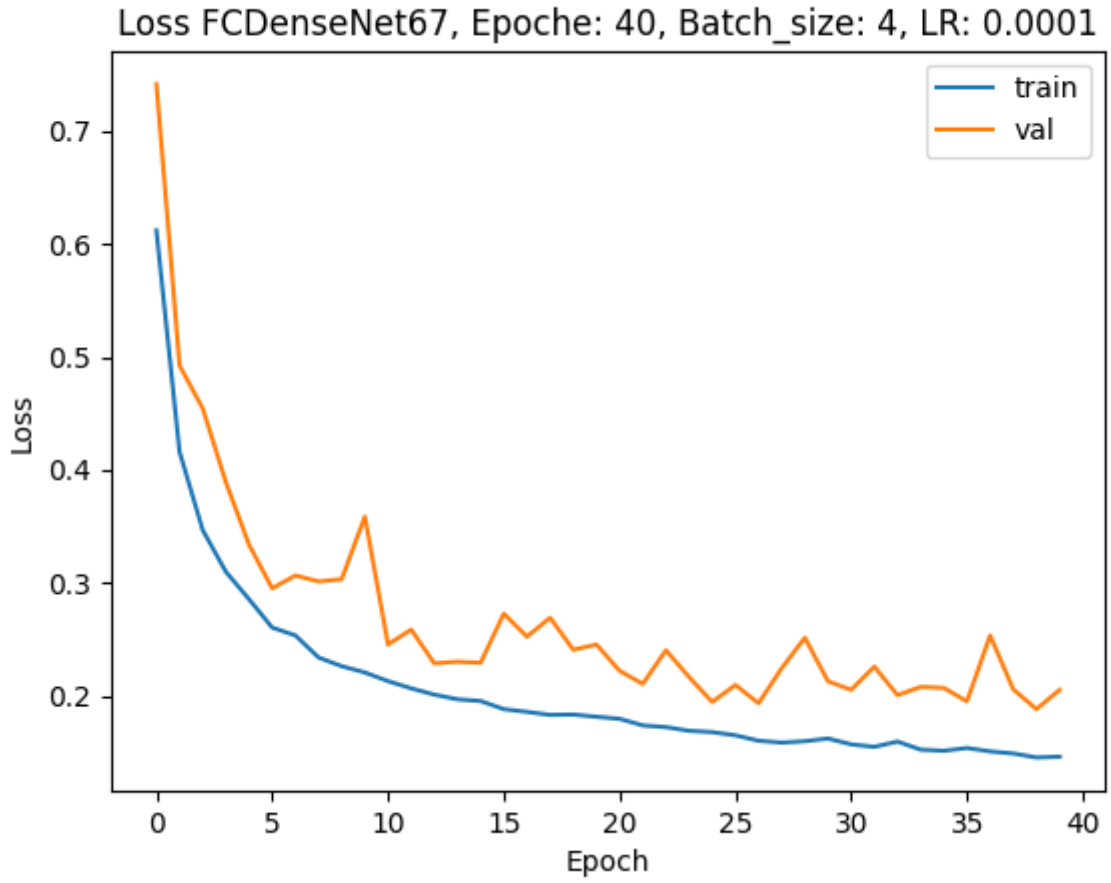

DSC FCDenseNet67, Epoche: 40, Batch_size: 4, LR: 0.0001
DSC
Epoch
train
val


Loss FCDenseNet67, Epoche: 40, Batch_size: 4, LR: 0.0001
Loss
Epoch
train
val

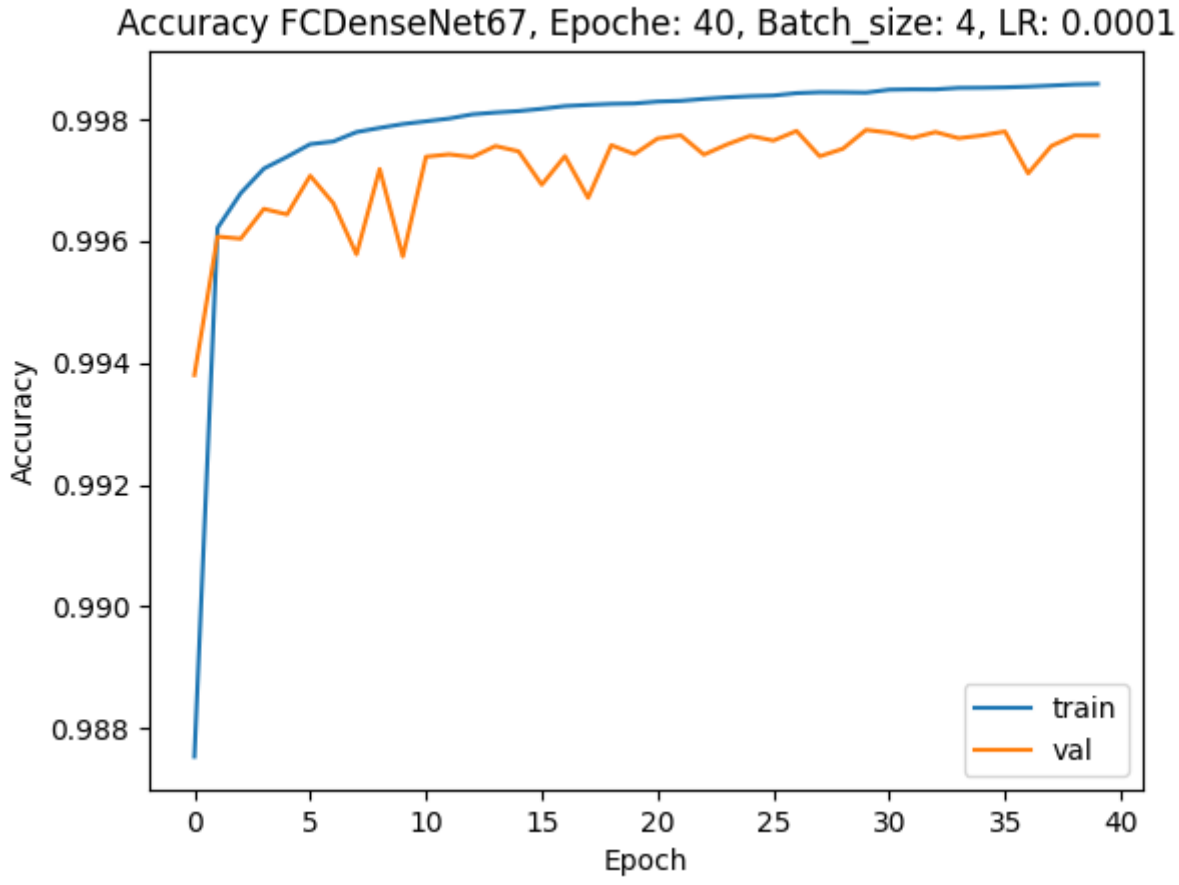
fold2

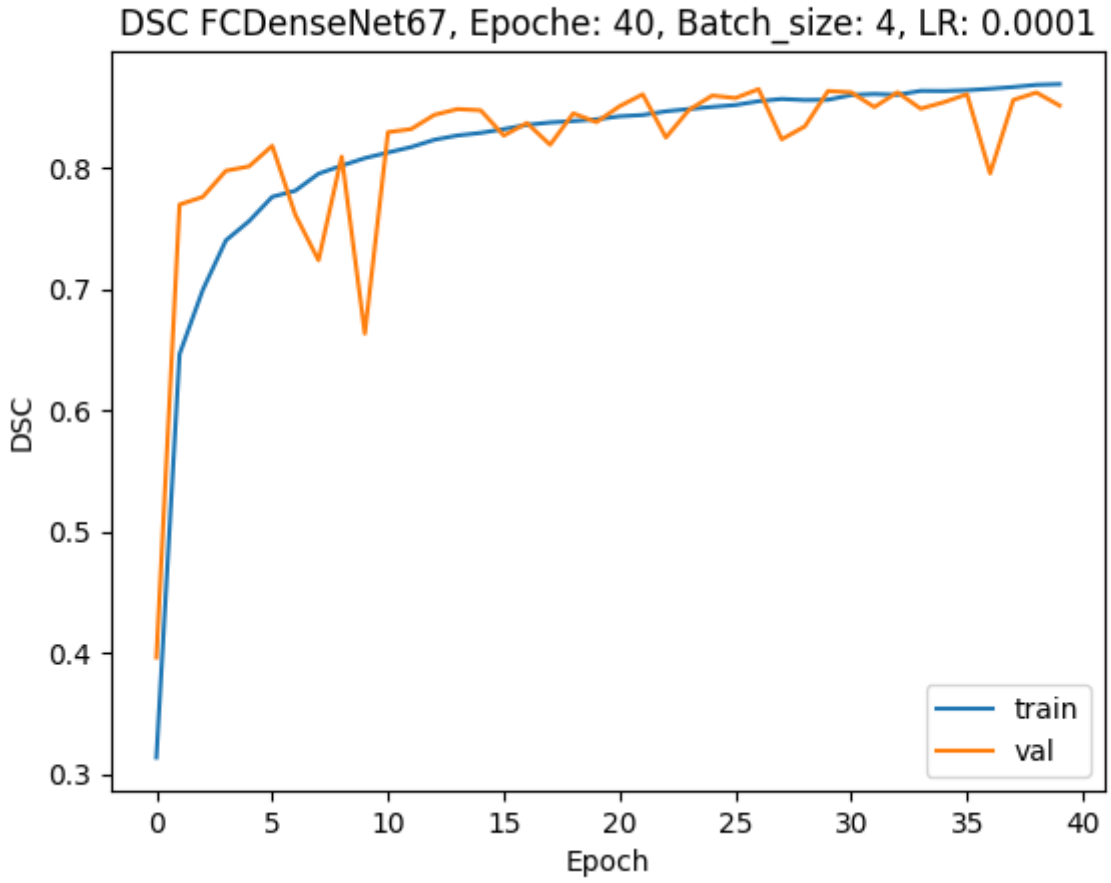

Accuracy FCDenseNet67, Epoche: 40, Batch_size: 4, LR: 0.0001
Accuracy
Epoch
train
val
0.998
0.996
0.994
0.992
0.990
0.988
0
5
10
15
20
25
30
35
40

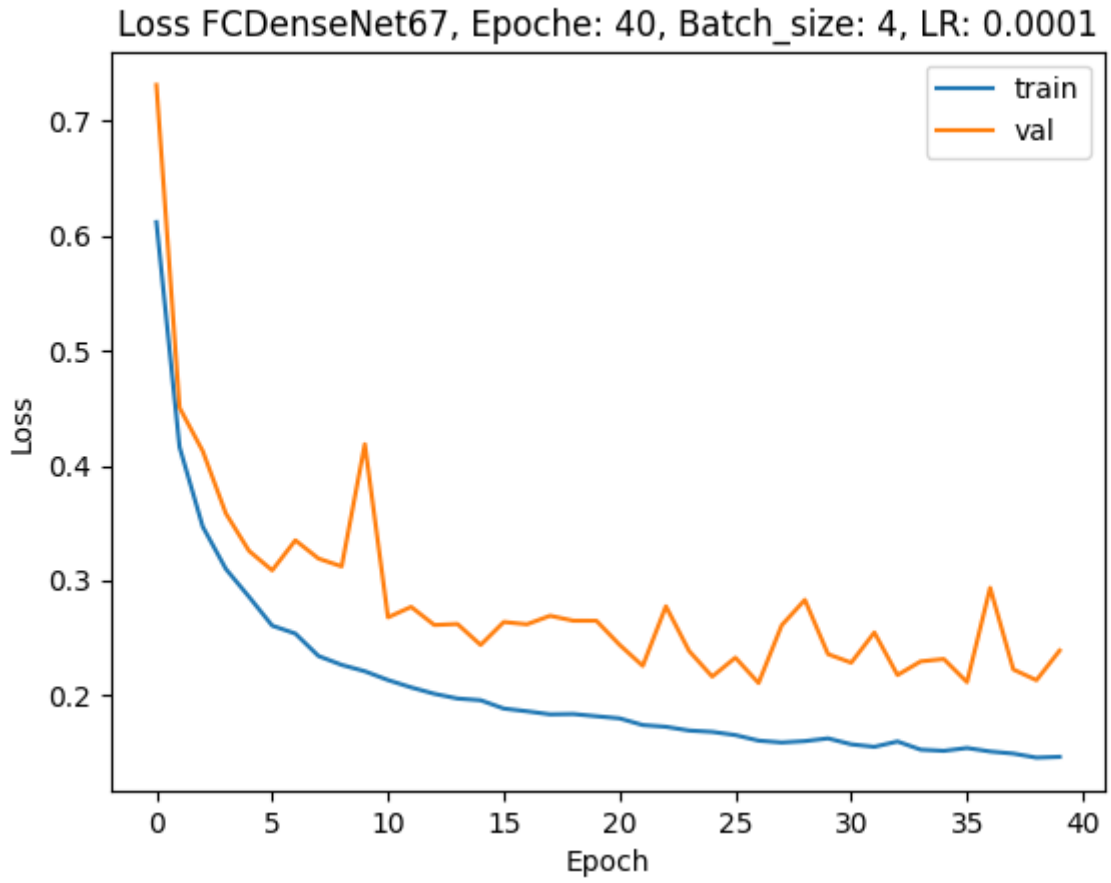

DSC FCDenseNet67, Epoche: 40, Batch_size: 4, LR: 0.0001
DSC
Epoch
train
val
0.8
0.7
0.6
0.5
0.4
0.3
0
5
10
15
20
25
30
35
40


Loss FCDenseNet67, Epoche: 40, Batch_size: 4, LR: 0.0001
train
val
Loss
Epoch
0.7
0.6
0.5
0.4
0.3
0.2
0
5
10
15
20
25
30
35
40

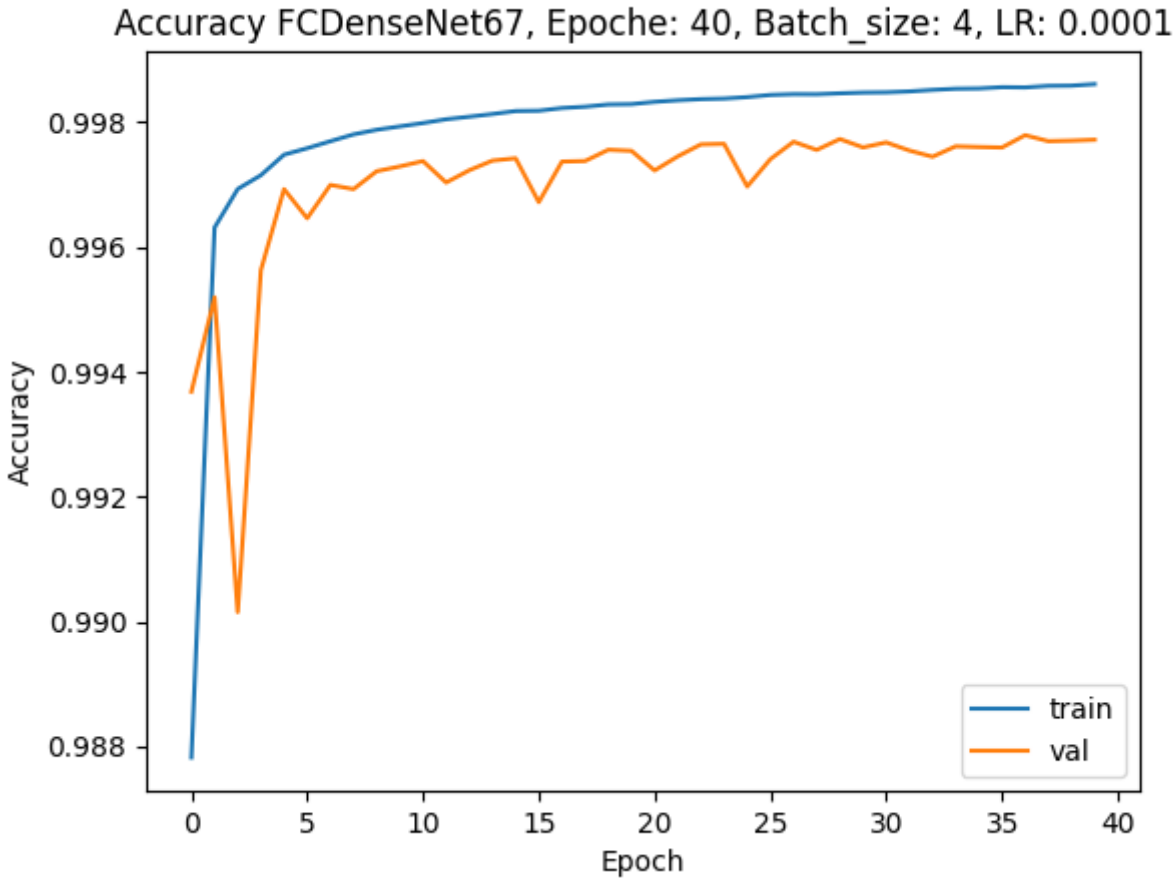
fold3

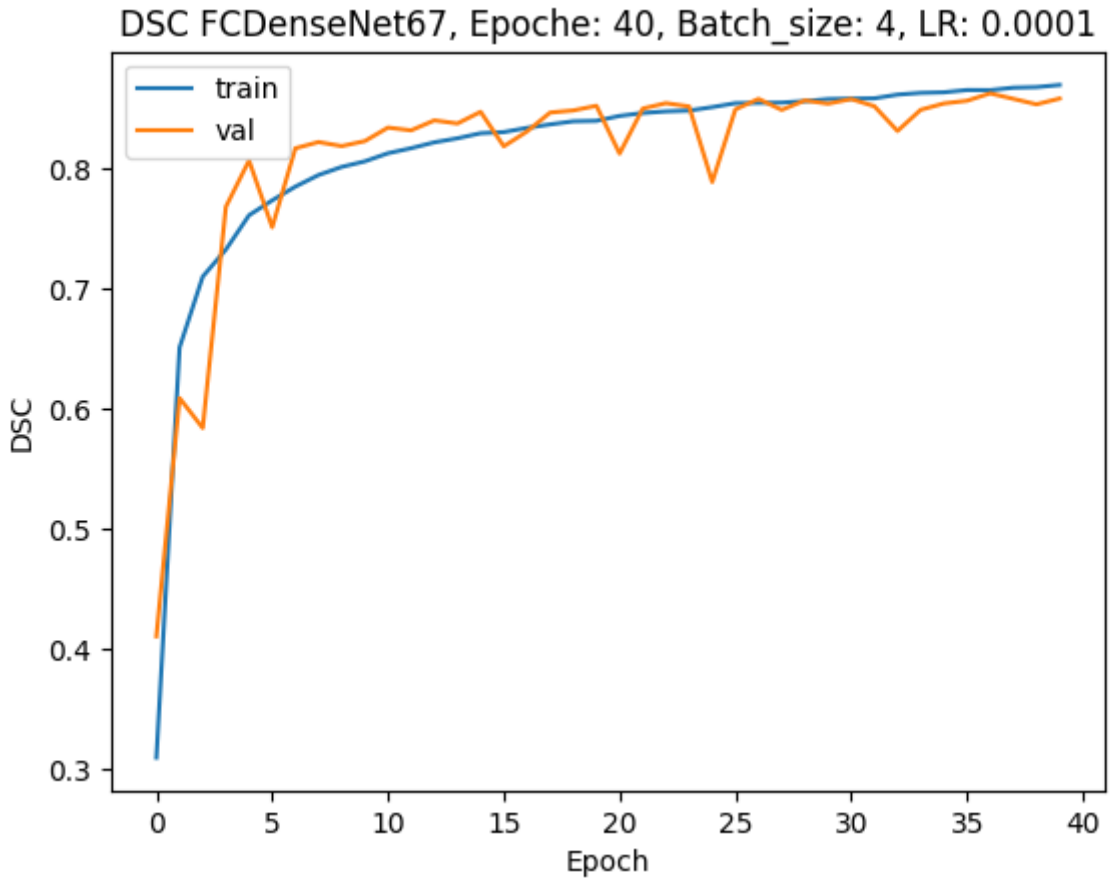

Accuracy FCDenseNet67, Epoche: 40, Batch_size: 4, LR: 0.0001
train
val
Accuracy
Epoch
0.998
0.996
0.994
0.992
0.990
0.988
0
5
10
15
20
25
30
35
40

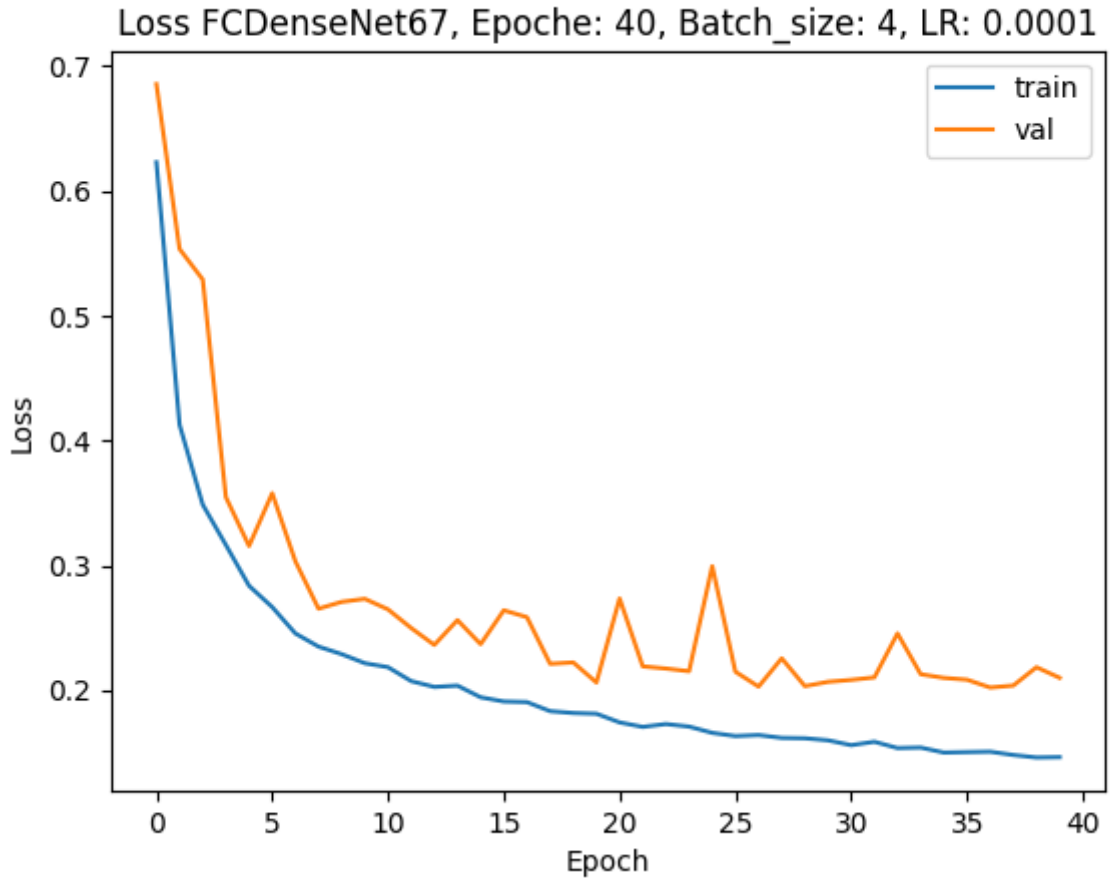

DSC FCDenseNet67, Epoche: 40, Batch_size: 4, LR: 0.0001
DSC
Epoch
train
val
0.3
0.4
0.5
0.6
0.7
0.8
0
5
10
15
20
25
30
35
40


Loss FCDenseNet67, Epoche: 40, Batch_size: 4, LR: 0.0001
Loss
Epoch
train
val
0.2
0.3
0.4
0.5
0.6
0.7
0
5
10
15
20
25
30
35
40

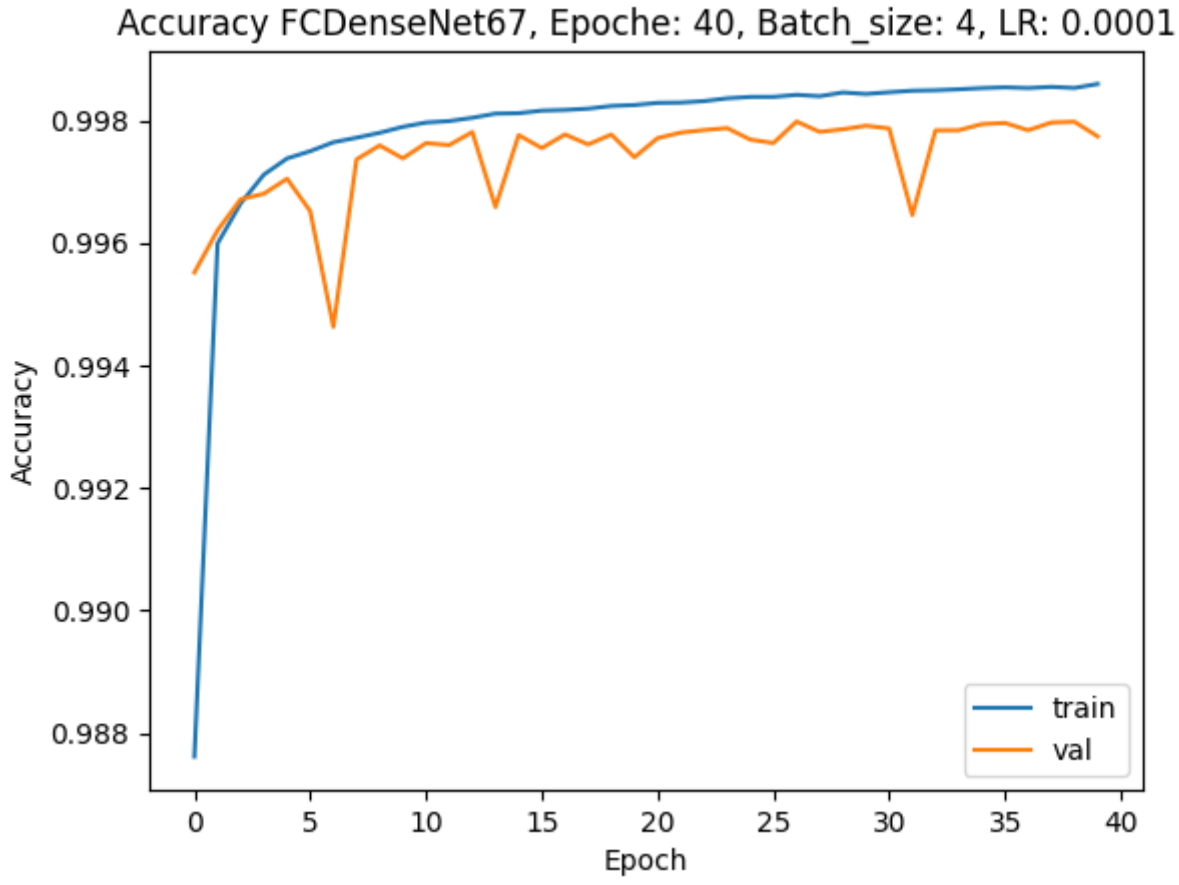
fold4

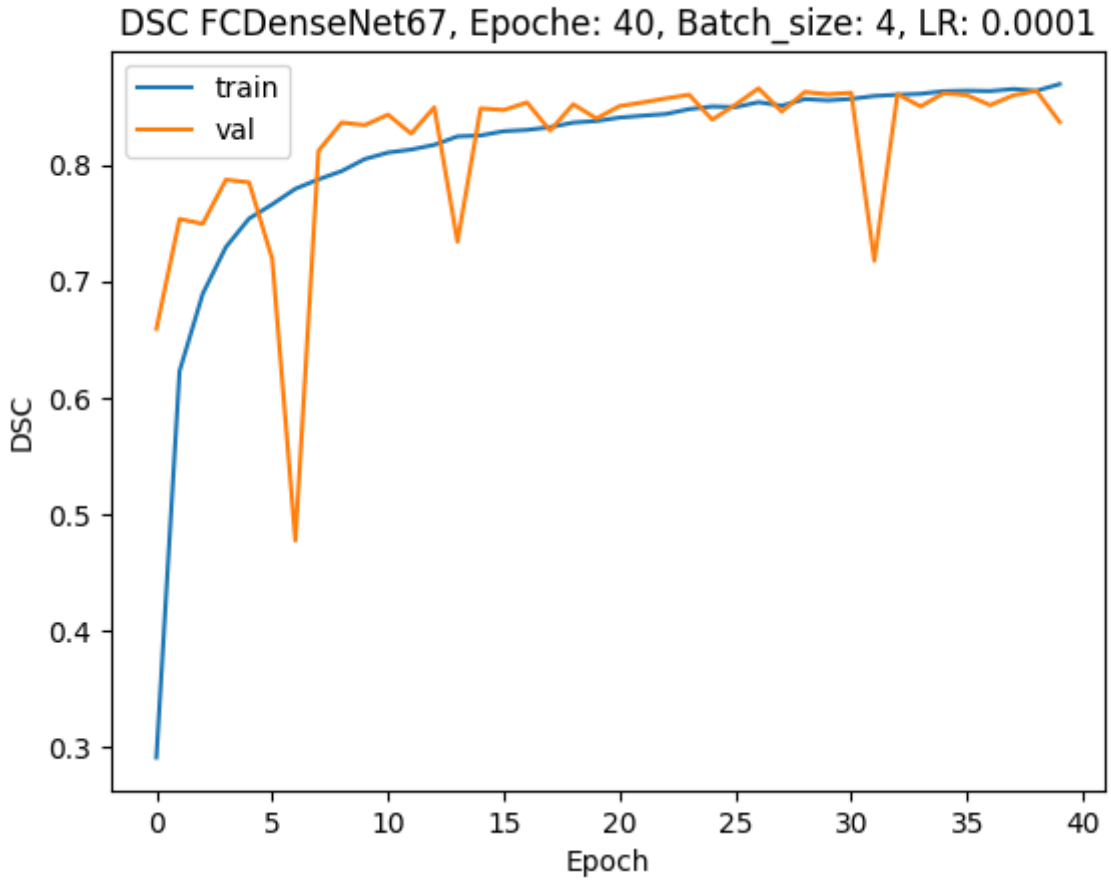

Accuracy FCDenseNet67, Epoche: 40, Batch_size: 4, LR: 0.0001
Accuracy
Epoch
train
val
0.998
0.996
0.994
0.992
0.990
0.988
0
5
10
15
20
25
30
35
40

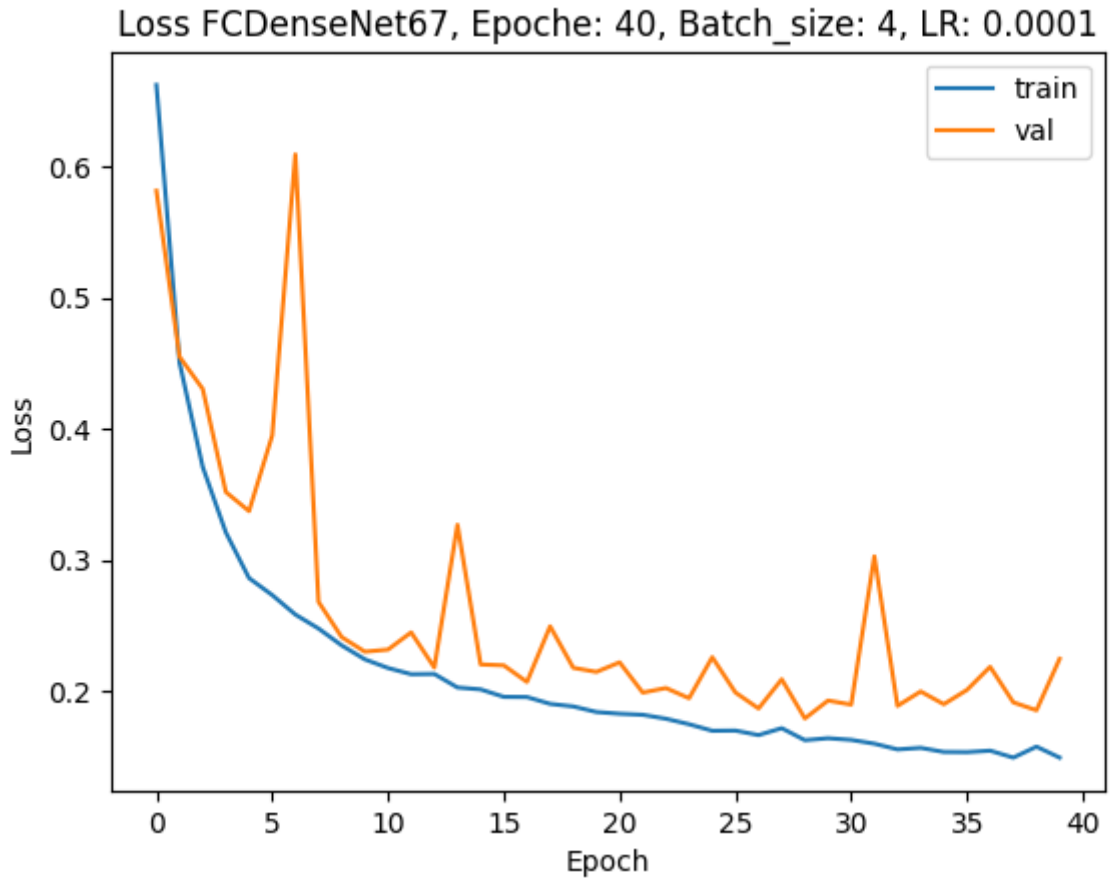

DSC FCDenseNet67, Epoche: 40, Batch_size: 4, LR: 0.0001
DSC
Epoch
train
val
0.8
0.7
0.6
0.5
0.4
0.3
0
5
10
15
20
25
30
35
40


Loss FCDenseNet67, Epoche: 40, Batch_size: 4, LR: 0.0001
train
val
Loss
Epoch
0.7
0.6
0.5
0.4
0.3
0.2
0
5
10
15
20
25
30
35
40

fold5


Accuracy FCDenseNet67, Epoche: 40, Batch_size: 4, LR: 0.0001
train
val
Accuracy
Epoch
0.998
0.996
0.994
0.992
0.990
0.988
0
5
10
15
20
25
30
35
40

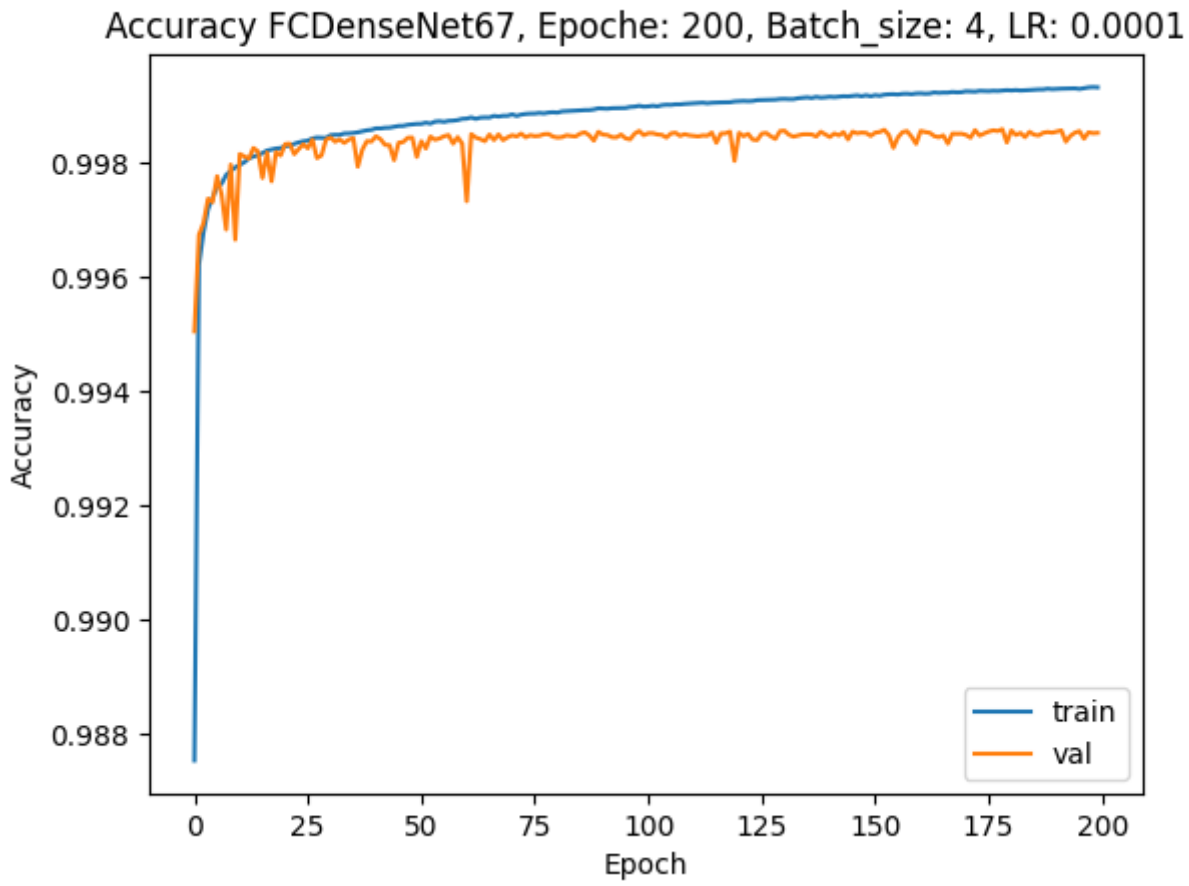

DSC FCDenseNet67, Epoche: 40, Batch_size: 4, LR: 0.0001
train
val
DSC
Epoch
0.3
0.4
0.5
0.6
0.7
0.8
0
5
10
15
20
25
30
35
40

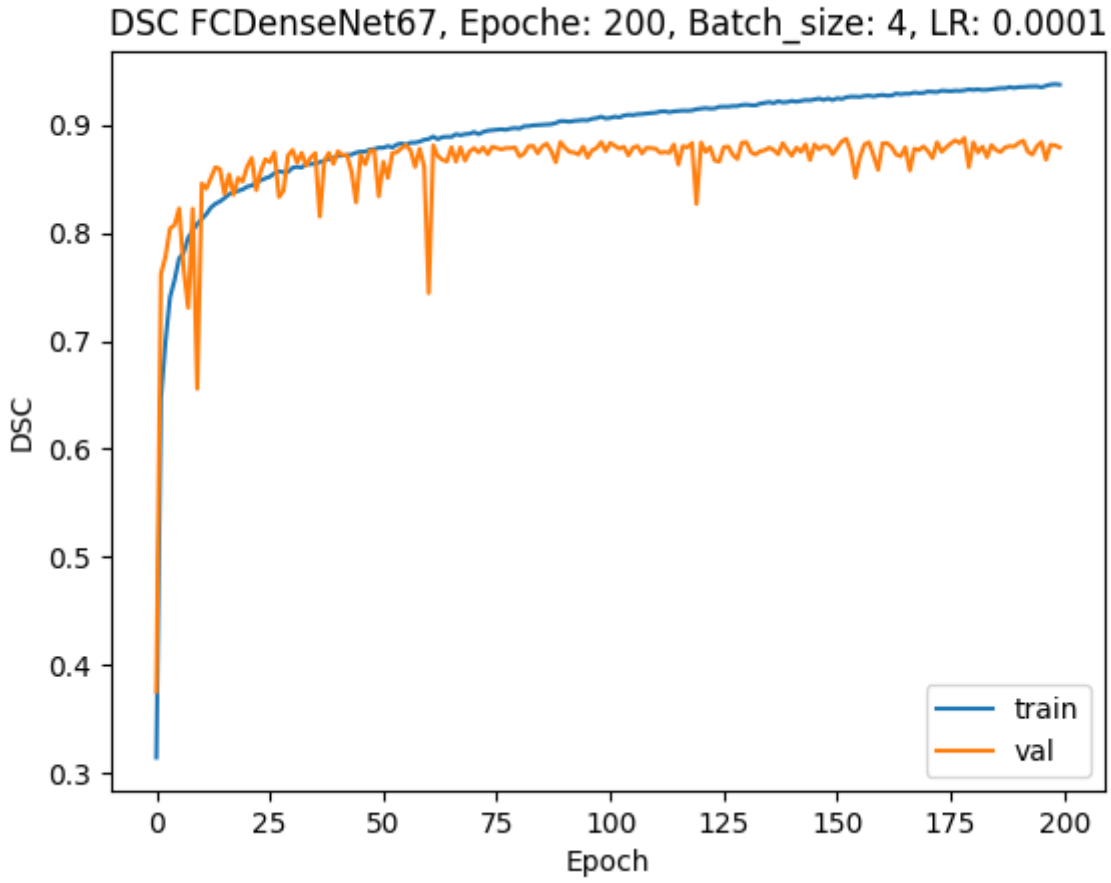

Loss FCDenseNet67, Epoche: 40, Batch_size: 4, LR: 0.0001
train
val
Loss
Epoch
0.2
0.3
0.4
0.5
0.6
0
5
10
15
20
25
30
35
40

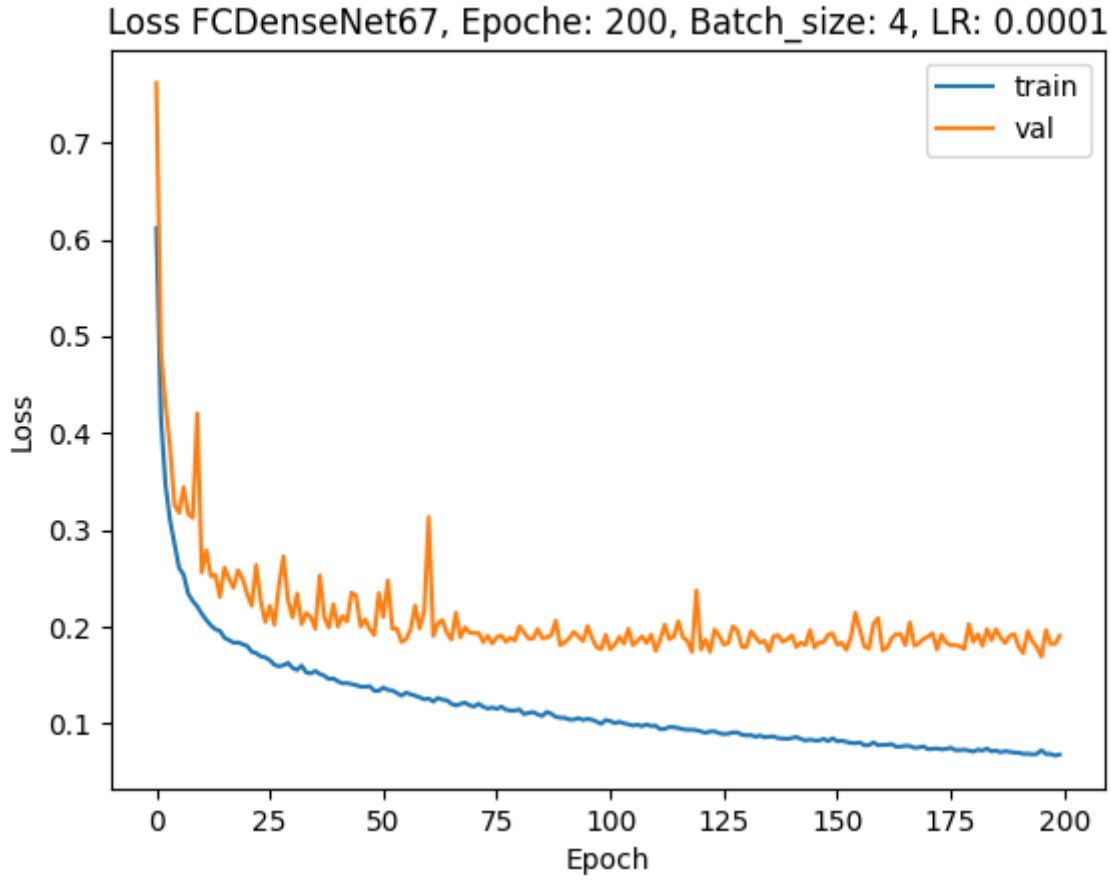
- 200epoch

5个fold全部训练完毕大约需要六天，所以只以fold1作为例子。

	acc	dice	sens	spec
std	0.9971	0.8448	0.8601	0.9983
mine	0.9985	0.8879	0.9001	0.9995
best	0.9992	0.8900	0.9071	0.9997







可见在200epoch时可以达到文章中所说最佳结果。

总结

> 复现结果分为图片和Gitahub上的日志，与数据集一起均放在代码库内

本次的复现过程我没有进行任何的创新，仅仅按照作者的流程走了一遍，顶多做了一点本地化改动，但是依旧消耗了大量时间，这让我深感科研不易。

作为一名本科生，没有接触过太多与科研相关的东西，所以我认为收获最大的部分是阅读文献阶段。从浩如烟海般的文章中找出一篇合适的并没有我原先想象的那么容易，许多文章已经花了几个小时了解的差不多了，最后可能因为某一个部分的不合适就只能抛弃，不过好处是英文阅读水平大幅提升。

说来惭愧，在计科的这两年我竟然没怎么接触过别人的项目，一直都是闭门造车。这次阅读了许多别人的代码，大佬们扎实的功底也让我大开眼界，虽然说我写的python代码也不少，但是两相比较发现自己走的仍然是C的老路子，完全没有发挥python的特性与优势。以本次代码为例，作者提供了丰富的接口改变内部参数，这为线上训练提供了许多方便，内部框架拥有良好的对称性，函数的复用度也很高，最后甚至还贴心的附带了plot绘制指标折线图的功能，本来这一部分我是想自己加上去的。内自省也，我的代码基本功还是需要提升和大量训练。

除了找到方便复现的文章，阅读过程中也接触到了许多前沿项目，才发现深度学习已经渗透到了几乎每个方面，发展的速度超乎我的想象。而这些项目都是触手可得的，代码也都无私的被贡献了出来。阅读别人的项目，自己动手复现，进行一些小改动，确实是是了解这些前沿知识的最好途径。

最后是这门课程带给我的感触，不论什么课，在上课的过程中我总是想接触一些实际生活中的的确确能应用的东西，或者是最近大伙正在研究的最新方面。我认为这门课给了我一个很好的机会，在课上总是能听见老师以自己最近的项目作为例子进行说明，也总是惊讶于这个方面的内容竟然已经发展的如此迅速（虽然大部分时间听不懂）。模型这个黑盒子到底是怎么通过深度学习搭建起来的，我也对这个问题很感兴趣，一个学期下来我也大致得到了自己想要的答案。如此这般，一边叙述前沿技术，一边教授底层原理，我认为在这门课上收获还是很大，不过说实话我并不是一个很好的学生，成绩也不是很高，第二次作业甚至没有完全完成，本次复现相信助教也能看出来，解决的问题都没有什么技术含量，与班上大佬的差距不是一星半点儿，但是我依然对自己在这门课上学到的知识心满意足，毕竟也算正式开启了这一方面的学习。

以上，感谢老师们幽默风趣干货满满的教学，感谢助教的辛勤付出和认真解答。

PB21111626

辛迦诺