SIGMA实验室本科生入组培训方案

## 前言

培训内容包含从编程入门到科研入门，涵盖独立完成科研任务的全部内容。[Sigma主页](http://sigma.whu.edu.cn/index.php)，**组里涵盖方向很多，一定能找到你感兴趣的方向**。进入实验室没有时间限制，我们欢迎任何年级和任何专业的同学前来学习。

## 培养目标

本方案以新进组本科生快速熟悉科研任务为主要培养目标，希望能在培训任务的指引下，带领你掌握以下能力：

* **熟悉深度学习/机器学习的基本原理**：重点掌握计算机视觉方向的常见任务。比如：了解诸如分类模型中的AlexNet、ResNet等经典卷积神经网络模型；了解以Transformer技术为基础的视觉模型ViT及其后续衍生，能够简单、基本的视觉任务训练及测试；了解自监督学习概念，对比学习技术；了解模型压缩的概念以及知识蒸馏技术；了解大语言模型的发展现状和模型微调。
* **具备深度学习代码的基本功**：对Python、PyTorch有基本的了解，清楚Anaconda的作用，能使用Linux系统的远程服务器进行深度学习任务；能在给出GitHub仓库的情况下，根据论文内容跑通代码；能对已有代码进行改进；
* **熟练常用的科研工具**：了解什么是arXiv、如何进行文献管理和阅读(Zotero)、如何使用Overleaf进行写文章、如何利用VScode进行远程代码调试、如何利用PPT做论文调研报告和科研汇报等。

在具备以上技能的基础上，努力探索创新性idea并实现，争取在人工智能领域顶级期刊/会议上发表学术论文。

## 培训内容

### 数学基础（1周，选看）

这部分主要是针对尚未学习过高等数学的大一或大二本科生。因为微积分和线性代数的基础是必须要掌握的，不然对于理解学习算法的原理会有困难。如果已经有一定的数学基础，可以先跳过这一部分，简单浏览即可，需要的时候再回来补。

* [微积分（导数、偏导数等）](https://github.com/loveunk/machine-learning-deep-learning-notes/blob/master/math/calculus.md)
* [线性代数（包括向量、矩阵等）](https://github.com/loveunk/machine-learning-deep-learning-notes/blob/master/math/linear-algebra.md)
* [矩阵理论（包括方差、协方差、积等）](https://github.com/loveunk/machine-learning-deep-learning-notes/blob/master/math/pca.md)

### Python编程和Linux基础（2-3周，部分必看）

#### Python和机器学习库（选看）

这部分主要是针对没有面向对象编程基础的本科生，推荐学习[Python](https://github.com/loveunk/machine-learning-deep-learning-notes/blob/master/python/python-basic#tdsub)或[Python基础教程](https://www.bilibili.com/video/BV1wW411Y7ai)。如果有比较好的Python和机器学习相关Library的知识基础，这部分可以暂时跳过，等后面用到再来针对性学习。

#### Linux和Git学习（必看）

对Linux不熟练的，学习[Linux基础教程](https://www.bilibili.com/video/BV1zx411E7KH)。对Anaconda不熟悉的，可以参考[如何在Linux服务器上安装Anaconda](https://blog.csdn.net/wyf2017/article/details/118676765)。对Git不熟悉的，学习[一小时Git教程](https://www.bilibili.com/video/BV1HM411377j/)。对Visual Studio Code如何连接远程服务器进行代码调试，学习[利用VScode连接远程服务器进行代码调试及可视化界面](https://blog.csdn.net/qq_31347869/article/details/121672295)。

### 深度学习的基本知识（2-3周，部分必看）

#### 参考[李宏毅的机器学习课程](https://speech.ee.ntu.edu.tw/~hylee/ml/2022-spring.php#tdsub)，学习机器学习的基本原理（必看）

只需学习2/18，2/25，3/04，3/25，4/22，4/29，5/06这几日的课程，其余课程内容目前不用学习。在听课程内容时，优先听Class Material，其余的Preparation和Extra Material作为补充；4/22，4/29，5/06的课程只听Preparation；

**注：**可以参考[这个CS231N专栏](https://space.bilibili.com/216720985/channel/seriesdetail?sid=1542349#tdsub)，学习计算机视觉的基本原理。这个教程和上面课程有一些重复，因此可以主学一个，然后快速过一遍这个专栏来查漏补缺，强烈推荐看完9、12、13和16；另外，对于一些难以理解的地方，也可以参考李沐老师主讲的一系列深度学习视频[动手学习深度学习](https://github.com/MLNLP-World/DeepLearning-MuLi-Notes)（文字）和[B站视频](https://space.bilibili.com/1567748478/channel/seriesdetail?sid=358496)，该视频讲的更细一些。

#### 深度学习框架 Pytorch（部分选看）

PyTorch无论是在工业界还是学术界，都已经碾压了其他的框架，例如TensorFlow、Keras，所以建议直接学PyTorch就好了。对于PyTorch的学习，建议在掌握理论知识，听完了李宏毅老师的课程的基础上，学习[PyTorch官方文档](https://pytorch.org/tutorials/beginner/basics/intro.html)给出的教程和几个示例。同时，强烈推荐学习[小土堆的教程](https://www.bilibili.com/video/BV1hE411t7RN/)**（如果想快点上手，重点推荐这个教程）**。下面是Ptrorch的中文手册和官方说明文档，可能直接看文档比较枯燥，建议在编程过程中结合各种博客来学习。

* [PyTorch 中文手册](https://github.com/zergtant/pytorch-handbook#tdsub)（选看）
* [PyTorch 官网的Tutorial](https://pytorch.org/tutorials/#tdsub)（选看）

另外，还有一些其他的深度学习的库和包可以作为了解，后面用到的时候再来深入学习。

* [Pandas](https://github.com/loveunk/machine-learning-deep-learning-notes/blob/master/python/pandas#tdsub)是Python中的数据处理和分析库，尤其用于处理结构化数据（选看）。
* [NumPy](https://github.com/loveunk/machine-learning-deep-learning-notes/blob/master/python/numpy#tdsub)是Python中最基础的科学计算库，提供了高效的数组操作和线性代数功能。（选看）
* [Matplotlib](https://github.com/loveunk/machine-learning-deep-learning-notes/blob/master/python/Matplotlib#tdsub)是一个Python绘图库，提供了丰富的二维图形绘制功能，广泛应用于数据可视化。（选看）
* [Scikit-Learn](https://github.com/loveunk/machine-learning-deep-learning-notes/blob/master/python/Sklearn#tdsub)是一个Python机器学习库，提供了许多标准机器学习算法，如分类、回归、聚类等。（选看）

### 经典深度学习算法（3-4周，必看）

这部分列出了深度学习中最具有代表性的几种网络，包括卷积神经网络CNN、生成式对抗网络GAN、残差网络ResNet、Transformer及其变型Vision Transformer、对比学习等等，还有很多经典的深度学习网络。另外，这里还为大家推荐了一些讲的比较通俗易懂的学习资料，仅作为参考性案例。

* **卷积神经网络CNN*（必看）***：

它广泛应用于计算机视觉和图像处理任务，如图像分类、目标检测、图像分割等。CNN的核心思想是利用卷积操作来提取图像的局部特征，并通过多层网络进行特征的逐步抽象和表达。CNN网络上可学习的资源很多，可参考学习[卷积神经网络CNN详解](https://github.com/loveunk/machine-learning-deep-learning-notes/blob/master/deep-learning/4.convolutional-neural-network-1.foundations-of-cnn.md)。

* **生成对抗网络GAN*（必看）***：主要用于生成新数据，它由两部分组成：生成器（Generator）和判别器（Discriminator）。这两部分通过对抗过程共同训练，从而使得生成器能够生成尽可能接近真实数据的样本。推荐学习李沫老师主讲的[GAN论文逐段精读](https://www.bilibili.com/video/BV1rb4y187vD)。
* **残差网络ResNet*（必看）***：通过引入残差连接（skip connections）解决了随着网络深度增加，训练变得越来越困难的问题。它的核心思想是利用残差学习，简化了深层网络的训练过程，允许构建非常深的神经网络。ResNet网络上可学习的资源很多，可参考学习李沫老师主讲的[ResNet论文逐段精读](https://www.bilibili.com/video/BV1P3411y7nn/?spm_id_from=333.999.0.0&vd_source=b01387a75c5baa957d2a119ec1bc816e)。
* **Transformer*（必看）***：一种用于处理序列数据的深度学习模型，首次提出于2017年的论文《Attention is All You Need》。与传统的循环神经网络（RNN）和长短时记忆网络（LSTM）不同，Transformer 完全基于自注意力机制（Self-Attention），通过并行化处理序列数据，显著提高了训练效率，并在多个任务中实现了卓越的表现，尤其在自然语言处理（NLP）领域。推荐学习李沫老师主讲的[Transformer论文逐段精读](https://www.bilibili.com/video/BV1pu411o7BE?spm_id_from=333.788.videopod.sections&vd_source=b01387a75c5baa957d2a119ec1bc816e)，介绍的特别详细。
* **Vision Transformer (ViT)*（必看）***：将Transformer模型应用于计算机视觉领域的一个变型。它摒弃了卷积层，直接将图像划分为块并处理，能够在大规模数据集上表现出色，尤其适用于图像分类、目标检测和语义分割等任务。ViT网络上可学习的资源很多，推荐学习[Vision Transformer详解](https://blog.csdn.net/qq_37541097/article/details/118242600)。
* **Contrastive Learning*（必看）***：是一种无监督学习的方法，旨在通过学习样本之间的相似性和差异性来学习有效的特征表示。主要思想是通过构造正样本对（相似的样本）和负样本对（不相似的样本），使得模型能够在特征空间中将相似的样本拉近距离，将不相似的样本推远，从而学习到具有判别性的特征表示。推荐学习[对比学习论文综述](https://www.bilibili.com/video/BV19S4y1M7hm)来了解最典型的对比学习论文。
* **Diffusion Model*（必看）***：是近年来在生成模型领域崭露头角的技术，因其强大的生成能力和灵活性，迅速成为计算机视觉和其他任务中的研究热点。首先学习DDPM扩散和去噪过程的理论推导，明白DDIM如何优化采样对DDPM进行加速的。可结合Stable Diffusion了解文本到图像生成（Text-to-Image Generation） 任务。
* 其他待完善。。。。

### 大模型（1-2周，部分必看）

#### LLM大语言模型*（选看）*

语言大模型（LLM）可以通过学习大量的语料来模拟人类语言处理的能力，如文本生成、翻译、问答等。建议借助网络，熟悉一下什么是语言大模型，当然也可以参考[语言大模型LLM介绍](https://www.bilibili.com/video/BV1Hj41177fb/?vd_source=b01387a75c5baa957d2a119ec1bc816e)by Andrej Karpathy。

#### LVM 视觉大模型*（必看）*

推荐学习一下2023年ICCV的Best Paper《Segment Anything》，推荐学习[Segment Anything论文详解（SAM）](https://blog.csdn.net/Zosse/article/details/130026210),[【论文阅读笔记】Segment Anything](https://blog.csdn.net/qq_46056318/article/details/136619295)，这里仅给出了两个CSDN博客，也可以在网络上检索其他学习资料。近期该团队又发不了SAM2: Segment Anything in Images and Videos，也可以关注一下。

#### MLLM 多模态大模型*（部分必看）*

随着语言大模型的发展，图文多模态领域从23年开始也获得了巨大的突破，也是当今研究热点和未来趋势。多模态大模型的代表包括：CLIP（图文对齐），DALLE（文本生成图片）、BLIP2（图文生成、图像描述、视觉推理），LLaVA（图文理解与生成、跨模态任务）、LLaMA3（大规模文本理解与生成）等。推荐学习[多模态论文串讲·上](https://www.bilibili.com/video/BV1Vd4y1v77v?vd_source=b01387a75c5baa957d2a119ec1bc816e&spm_id_from=333.788.videopod.sections)和[多模态论文串讲·下](https://www.bilibili.com/video/BV1fA411Z772?vd_source=b01387a75c5baa957d2a119ec1bc816e&spm_id_from=333.788.videopod.sections)学习***（选看即可）***。其中现有的多模态大模型都会利用CILP模型的Encoder和Decoder模块，建议重点学习CLIP模型，其他模型作为了解即可，推荐[多模态预训练模型CLIP论文详解](https://blog.csdn.net/h661975/article/details/135116957)，[CV大模型系列之：多模态经典之作CLIP，探索图文结合的奥秘](https://juejin.cn/post/7264503343996747830)***（必看）***。大模型发展日新月异，需要持续不断学习State-of-the-art的工作。

### 科研基础工具（1-2天，必看）

通过网络搜索，了解什么是[arXiv](https://arxiv.org/)、什么是[Overleaf](https://www.overleaf.com/)，并初步了解如何使用[Zotero](https://www.zotero.org/)等文献管理软件。如何做文献笔记和论文调研，了解[PaperWithCode](https://paperswithcode.com/), [Google Scholar](https://scholar.google.com/)；学习使用Powerpoint完成汇报PPT的制作；了解[BibTex](https://bibtex.eu/zh-cn/)，知道参考文献的几种写法和格式。对LaTeX不熟悉的同学请学习[LaTeX基础教程](https://www.bilibili.com/video/BV1Z24y157GM/)。网上资源很丰富，一般一天时间就能学会。

### 经验分享（1-2天，选看）

1. 关于如何快速上手读论文和判断自己的研究能否发paper，推荐看李沫老师主讲的[如何读论文](https://www.bilibili.com/video/BV1H44y1t75x/?spm_id_from=333.999.0.0&vd_source=b01387a75c5baa957d2a119ec1bc816e)和[如何判断（你自己的）研究工作的价值](https://www.bilibili.com/video/BV1oL411c7Us/?spm_id_from=333.999.0.0&vd_source=b01387a75c5baa957d2a119ec1bc816e)。
2. 关于论文写作方面的学习，可以参考北京大学施柏鑫教授在青源学术年会上做的[审稿视角下的计算机视觉论文——从投稿到接收](https://hub.baai.ac.cn/view/8659)。关于论文的Writing Tips，请参考南开大学李重仪教授的[How to Write a Paper](https://pi-lab.xyz/miscellaneous.html)。关于论文的投稿和返修也可参考组里叶茫教授的[学术论文投稿与返修（Rebuttal）分享](https://zhuanlan.zhihu.com/p/344008879)。