

- [1.大数据导论作业mission1（部分整理）](#)
 - [（1）数据下载完成：](#)
 - [（2）代码解释：](#)
 - [（3）padding、kernel_size和stride 的区别：](#)
 - [（4）池化后的输出层：](#)
- [2.Accumulate Hydrothermal&Cold seeps Sites Distribution（部分整理）](#)
 - [2.1 Cold seeps](#)
 - [2.2Hydrothermal vents](#)
 - [2.3文献整理](#)
- [3.Probabilistic Prediction of Cold Seeps and associated SEAFLEA](#)
 - [3.1 Notification](#)
 - [3.1.1Best predictors of SEAFLEAs](#)
 - [3.1.2Method](#)
 - [3.1.3Discussion](#)
 - [3.1.4Conclusion:](#)
 - [3.2 Vocabulary](#)
- [4本周遇到的问题](#)
 - [4.1 Cold Seeps Sites（未解决）](#)
 - [4.2 Hydrothermal Sites（未解决）](#)

1.大数据导论作业mission1（部分整理）

（1）数据下载完成：



The screenshot shows a JupyterLab interface with a notebook titled 'mission1'. The code in the notebook includes:

```
print("训练集大小:", len(trainset))
print("测试集大小:", len(testset))
num_classes = 10
#创建10行x10列的子图网格，总大小15x15英寸；axes是一个10x10的数据，每个元素是一个子图对象
fig, axes = plt.subplots(num_classes, 10, figsize=(15, 15))
#提取标签
labels = np.array([t[1] for t in trainset])
#遍历每个类别
for i in range(num_classes):
    indice = np.where(labels == i)[0]
    #为每个类别显示10个样本
    for j in range(10):
        axes[i][j].imshow(trainset[indice[j]][0].permute(1, 2, 0).numpy())
        axes[i][j].set_xticks([]) #移除x轴刻度：为了减少视觉干扰、节省空间
        axes[i][j].set_yticks([])
plt.show()
```

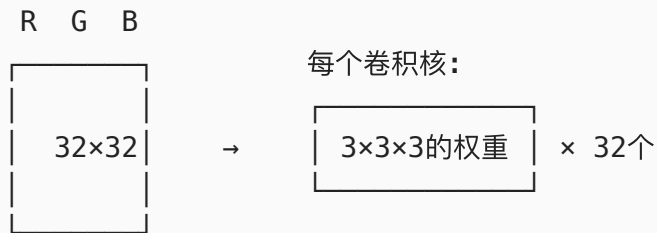
Below the code, there is a status bar showing '100%' completion and a progress bar. At the bottom, a code cell shows the definition of a CNN class:

```
class CNN(nn.Module):
    def __init__(self, num_classes=10):
        super().__init__()
        self.num_classes = num_classes
        #Conv2d: 二维卷积层; in: 输入数据的通道数，CIFAR有3个通道（RGB图像）；out: 输出通道数；ker: 卷积核大小；padding: 在输入图像四周填充的像素层数
        self.conv1 = nn.Conv2d(in_channels=3, out_channels=32, kernel_size=3, padding=1)
```

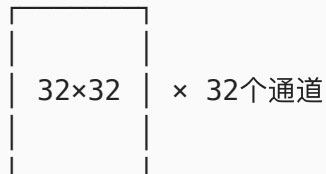
（2）代码解释：

```
self.conv1 =  
nn.Conv2d(in_channels=3,out_channels=32,kernel_size=3,padding=1)
```

输入图像（3通道，32×32） 卷积核（32个，每个3×3×3）



输出特征图（32通道，32×32）



(3) padding、kernel_size和stride 的区别：

参数	含义	比喻	示例
kernel_size	卷积核大小	你的手掌大小	3×3的手掌
stride	步长	你每次走几步	走1步或2步
padding	填充	在图片外加边框	加1圈空白

(4) 池化后的输出层：

输出高度 = (输入高度 - kernel_size) / stride + 1

输出宽度 = (输入宽度 - kernel_size) / stride + 1

2.Accumulate Hydrothermal&Cold seeps Sites Distribution（部分整理）

2.1 Cold seeps

1、Pangaea：

“PANGAEA 中的 ‘data points’ 指的是数值数据单元数量，而非空间站位数量。实际可用于 GIS 的冷泉站位约为一百多个。”

<https://doi.pangaea.de/10.1594/PANGAEA.893300?format=html#download>

2、UNEP-WCMC：

[Global Distribution of Cold Seeps \(2010\) - Overview](#)

3、NOAA（美国大西洋边缘）：

https://noaa-wcsd-pds.s3.amazonaws.com/index.html#data/processed/US_ATLANTIC_MARGIN_COLD_SEEPS_DATABASE_2011-2016/data/

2.2 Hydrothermal vents

文件夹名称：找到数据的网站

数据名称：数据原始来源

1、Hydrothermal Vents GIS Data：（2020年）

就是interridge3.4，2020版本

（shp格式）

https://nauru-data.sprep.org/resource/hydrothermal-vents-gis-data-interridge-global-database-active-submarine-hydrothermal-vent?utm_source=chatgpt.com

（csv格式）

https://tuvalu-data.sprep.org/resource/hydrothermal-vents-tabular-data-interridge-global-database-active-submarine-hydrothermal?utm_source=chatgpt.com

2、interridge：（2015年）

（3.3）kml格式

<http://vents-data.interridge.org/>

3、NOAA（未标明年份）

[Global Hydrothermal Vents \(text version\)](#)

4、（不确定）MARHYS Database (MARine HYdrothermal Solutions)

[迪尔，A;巴赫，W（2024）：MARHYS数据库4.0](#)

数据是所有热液喷口流体成分的坐标，并不是点位的坐标

热液流体的化学数据库，有采样点坐标，但不是完整的全球分布图。

下载的数据是所有样本的测量记录（在一个点位可能测量了很多次），所以需要去重：

使用数据透视功能，字段选择Vent，值选经纬度，字段设置选择取平均值。

5、Stanford University（2010年）

[Global Hydrothermal Vent Locations](#)

2.3 文献整理

Cold seep：

Hydrothermal：

1、An authoritative global database for active submarine hydrothermal vent fields：

（数据来源于interridge）

<https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2013GC004998>

2、A Global Probabilistic Prediction of Cold Seeps and Associated SEAFloor FLuid Expulsion Anomalies (SEAFLEAs)

3. Probabilistic Prediction of Cold Seeps and associated SEAFLEA

3.1 Notification

3.1.1 Best predictors of SEAFLEAs

geologic, biologic, and chemical predictors

3.1.2 Method

Define and assemble a database of existing anomalies for quantifying the observed distribution of seepage and to predict the probability of FLEAs.

3.1.3 Discussion

For each predictor: 10% of data is removed and predicted using the remaining 90 as training set.

This random validation approach can show the robustness of NNL single-class method.

3.1.4 Conclusion:

There is no distinction in the likelihood of encountering fluid expulsion between passive and active margins.

Future work should expand SEAFLEA database and, where possible, collect absence data to refine local-scale flux estimates.

3.2 Vocabulary

1 “SEAFloor FLuid Expulsion Anomalies” (Phrampus 等, 2020, p. 1) 海底流体喷出异常。

2 “anomalies” (Phrampus 等, 2020, p. 1) n. 异常; 畸形物; 网络释义: 异常现象; 异象; 反常

3 “global continental margins” (Phrampus 等, 2020, p. 1) 全球大陆边缘。

4 “bathymetry” (Phrampus 等, 2020, p. 6) n. 海洋测深学[术]; 海洋生物分布学; 网络释义: 水深测量; 测深法; 测深术

5 “Geospatial Validation” (Phrampus 等, 2020, p. 7) 地理空间验证。

4 本周遇到的问题

4.1 Cold Seeps Sites (未解决)

整理的冷泉点位投到QGIS中, 添加底图后显示在内陆, 但是在Google earth中显示是正常的。目前怀疑是底图或者投影坐标系有问题。

4.2 Hydrothermal Sites (未解决)

在SEAFLEAs文献中发现有公开的冷泉点位数据集，但是网站需要内部权限才能进入，另一个链接地址也打不开。