陈胤达2024.5.5 dreamdiffusion模型理解

Model1:encode and decode

输入batchsize\*电级个数\*时间

Transform成batchsize\*128\*512记作img

输入encode 输出latent8\*33\*1024，mask8\*128，ids\_restore8\*128

把latent和ids\_restore输入到decode得到predbatchsize\*128\*512

Loss计算，pred与img的差值的平方，之后按最后一维求均值得到8\*128记作loss，之后mask\*loss(mask值为0or1)，即计算mask中值为1的位置之后求均值。

指标cor(相似度)计算方式读取每个batchsize的每个电位数据进行求相关性(即把pred和img的最后一维512当作求相关性的数据)，之后求均值

Model2 diffusion

数据格式，是一个pth文件，里面的内容是有三个键分别是dataset，images和labesl(在本项目中没有作用)。dataset中包含了很多个元素，每个元素中有三个键，分别是eeg，image和label（在本项目中没有作用）,其中eeg为数据内容，其中image是images的缩影。images中包含了全部的刺激数据，但是是路径。

模型加载

1. 在model1中微调的编解码器模型记作pretrain\_mbm\_metafile
2. 加载dc\_lcm.models.diffusion.ddpm.LatentDiffusion模型
3. 别人已经训练好的模型v1-5-pruned参数

数据输入，将batchsize每个对应的刺激图片(4\*3\*512\*512)加载并输入到encoder模型中(没有decoder),输出z(4\*4\*64\*64),xc为eeg数据4\*128\*512输入到encoder模型中，得到了4\*77\*768的c和4\*128\*1024的re\_latent

Loss计算

1. 原图像跟diffusion模型输出的loss，输入图片经过encoder之后的数据4\*4\*64\*64，输出4\*4\*64\*64，采用l2loss，即做差平方，另一个加权的loss，最终loss1+loss2
2. eeg数据跟imgs的loss，经过encoder的eeg数据是4\*128\*1024，imgs输入是4\*3\*224\*224，

Imgs经过image\_embedder得到4\*768,eeg数据经过mapping得到4\*768，这两张都是通过神经网络结构进行二次特征提取的数据，最终计算两者的余弦相似度，用1-余弦相似度得到loss

指标:

1. mse（Mean Squared Error，均方误差）：是一种衡量两个图像之间差异的常见指标。它计算了原始图像与重构图像之间每个像素的差异的平方，并取平均值作为最终的误差值。
2. pcc（Pearson Correlation Coefficient，皮尔逊相关系数）：是一种衡量两个图像之间线性相关性的指标。它计算了原始图像和重构图像的像素值之间的相关性，值域在 -1 到 1 之间，越接近 1 表示两个图像之间的线性关系越强。
3. ssim（Structural Similarity Index，结构相似性指数）：是一种用于衡量两个图像之间结构相似性的指标。它考虑了亮度、对比度和结构三个方面的相似性，值域在 0 到 1 之间，越接近 1 表示两个图像之间的结构相似性越高。

4.psm（Peak Signal-to-Noise Ratio，峰值信噪比）：是一种衡量图像质量的指标。它计算了原始图像和重构图像之间的峰值信噪比，用于表示图像质量的好坏，值越高表示图像质量越好。