模型步骤二详细:

1. 输入，输入的eeg数据是15921，train为9552，val为3185，时间长度不固定，74个电极个数。
2. Transform，eeg时间维度进行截取知只取512，电极维度用复制的方法复制成128，最终的格式是8\*128\*512，图片按rgb读取为8\*512\*512\*3
3. 图片变成8\*3\*512\*512，经过了一个特征提取器(未知s)之后变成了8\*4\*64\*64
4. 图片经过一个vgg网络得到了8\*3\*224\*224，
5. Eeg数据经过第一步骤的编解码器，变成8\*77\*768和8\*128\*1024两部分数据(跟编解码器的设计有关)
6. 计算loss1，个人理解，按图片8\*4\*64\*64的格式生成一个噪声数据，把eeg数据（8\*77\*768）跟图片数据跟噪声数据输入到diffusion模型(?)得到8\*4\*64\*64的输出，用这个模型的输出跟eeg数据进行求l2loss，得到的loss为batchsize大小
7. Loss2，计算了loss1之后，利用随机函数得到1-1000中随机的batchsize的个数的数字，之后在lvlb\_weights中选择这batchsize的权重成以loss1，得到loss2(作者把loss2的权重设置为0)
8. Loss3把8\*3\*224\*224embedder成8\*756，把eeg数据（8\*128\*1024）形变成8\*754，计算余弦相似度。