**4.Datasets （这一块具体病人的数据是不是需要医生去写，我们就写关于划分的一些，另外其实可以写一下处理的方式和尺寸之类的）**

**4.1Internal dataset**

For the Internal dataset,we collected 550 patients from Shanghai Tenth People's Hospital as an internal dataset, and divided them into training and validation sets at a 4:1 ratio. The ratio of T0 and T1 stage prostate cancer patients in both the training and validation sets is close to 1:1, ensuring a relatively balanced distribution.

**4.2 External validation datasets**

For the External validation datasets.....

|  |  |  |  |
| --- | --- | --- | --- |
| datasets | Internal datasets  （train） | Internal datasets  （validation） | External validation |
| All patients | 440 | 110 | ... |
| T0 stage patients | 230 | 58 | ... |
| T1 stage patients | 210 | 52 | ... |

Table1. Composition of the Internal datasets and External validation datasets

**5.Implementation**

**5.1 Pretraining**

We implemented our experiments using PyTorch,We used Kinetics-710 pretrained weights for MAE weight initialization. Although there is a domain gap in natural and medical image data, studies show that pretraining on natural image data improves network performance on medical imaging tasks.（这里不需要说那么详细，对于我们的任务来说这是一个trick，就说我们装载了已有的模型作为我们的初始化参数，然后使用我们的数据进行预训练和微调）

We chose ViT-Base as the back bone.We set the patch size as 2x3x16x16,which means our token generation will give out tokens for an input video clip for 16x3x224x224.We set the mask ratio as 0.9,resulting in 1568\*(1-0.9) =157 tokens that are visible for vision transformer encoder. The pretraining phase is trained with an AdamW optimizer with LR 0.0004,for minimizing the MSE loss over 500 epochs. The batch size is 4. Warm-up was done for 5 epochs with LR 0.001.（时态保持一致，然后计算过程省略，最好是前面方法中提到的符号在这里给符号幅值，比如说之前如果说mask ratio为γ，则这里直接写γ equals to 0.9等等）

**5.2 Fintuning**

As mentioned in Sec. 3,~~As our method in sec 3 said,~~we divide the video into 16 frames to constitute a clip.From each video, we sampled 5 clips uniformly. During inference, we predict the labels for each of the clips. If any of the clips is predicted as malignant, the entire video is labelled as malignant.We minimized a cross entropy loss using an AdamW optimizer with LR 0.0004.

We process our experiments with 2 NVIDIA GeForce RTX 3090 GPUs for both pretraining and fintuning.

**6.Experiments and Results**

**6.1 Performance of our Segment guided mask and random mask**

As Quantitative Analysis,We present the final accuracy and ROC curve based on the random masking pretraining method, as well as the final accuracy and ROC curve based on the segmentation region masking method in the Figure 3 and Table 2. (表格这里的method别写mask策略，策略相当于消融实验，写不同的模型方法)

|  |  |  |
| --- | --- | --- |
| Method | Accuracy | AUC |
| Random Mask | 0.8207 | 0.875 |
| Segment Guided Mask(our work) | **0.9236** | **0.963** |

Table 2.Accuracy of Random Mask and Segment Guided Mask(our work)

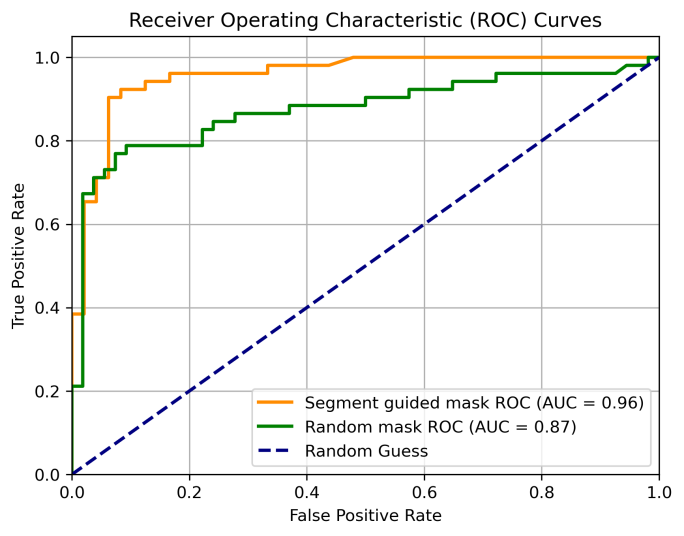
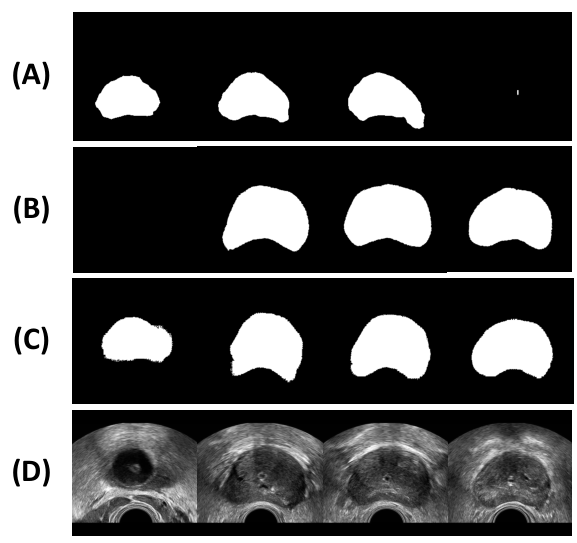


Figure 3.Performance of our work and random mask in AUC.(Internal datasets)

**6.2 Visualization of the Multi-export Segment-Generation（讲实验的时候就直接从实验结果进行分析，这里说的分割图就描述分割的效果本身 ，但是有个疑问，就是分割不属于我们任务要做的，这个是不是更适合放在前面讲方法的部分，即说明我们方法的时候佐证为什么要这么做）**

****

**Figure 4.The visualization of the Multi-export Segment-Generation.(A)the deeplabv3 network’s segment output.(B)the nnUNet2++ network’s segment output.(C)the MedSAM2 network’s segment output.(D)The raw frames of the prostate ultrasound video.**

By using the segmentation images from three pretrained medical segmentation models, we found that because the models were not fine-tuned on our data, their segmentations often contained errors As the fig4. However, the likelihood of all three models making errors simultaneously was relatively low. We randomly sampled several segmentation results from the training set along with their original prostate ultrasound images and found that in many cases, all three models were able to provide a basic segmentation. In some cases, the ultrasound image did not contain any lesions, and all three models generated completely black segmentation images. This can be explained: non-invasive scans may not have covered the prostate region in certain segments, resulting in the absence of lesions.