Image Encoder	ViT-L/14	ViT-g/14
Fine-tuning epochs	5	
Warmup steps	1000	
Learning rate	5e-6	1e-5
Batch size	224	
AdamW β	(0.9,0.98)	(0.9,0.999)
Weight decay	0.05	
Drop path	0	
Image resolution	364	
Layer-wise learning rate decay for ViT	1	0.95