

Supplementary Table S1. Detailed layer architecture of the SwinUNETR model

Layer (type)	Output Shape	Param #
Conv3d-1	[-1, 48, 48, 48, 48]	1,200
PatchEmbed-2	[-1, 48, 48, 48, 48]	0
Dropout-3	[-1, 48, 48, 48, 48]	0
LayerNorm-4	[-1, 48, 48, 48, 48]	96
Linear-5	[-1, 343, 144]	7,056
Softmax-6	[-1, 3, 343, 343]	0
Dropout-7	[-1, 3, 343, 343]	0
Linear-8	[-1, 343, 48]	2,352
Dropout-9	[-1, 343, 48]	0
WindowAttention-10	[-1, 343, 48]	0
Identity-11	[-1, 48, 48, 48, 48]	0
LayerNorm-12	[-1, 48, 48, 48, 48]	96
Linear-13	[-1, 48, 48, 48, 192]	9,408
GELU-14	[-1, 48, 48, 48, 192]	0
Dropout-15	[-1, 48, 48, 48, 192]	0
Linear-16	[-1, 48, 48, 48, 48]	9,264
Dropout-17	[-1, 48, 48, 48, 48]	0
MLPBlock-18	[-1, 48, 48, 48, 48]	0
Identity-19	[-1, 48, 48, 48, 48]	0
SwinTransformerBlock-20	[-1, 48, 48, 48, 48]	0
LayerNorm-21	[-1, 48, 48, 48, 48]	96
Linear-22	[-1, 343, 144]	7,056
Softmax-23	[-1, 3, 343, 343]	0
Dropout-24	[-1, 3, 343, 343]	0
Linear-25	[-1, 343, 48]	2,352
Dropout-26	[-1, 343, 48]	0
WindowAttention-27	[-1, 343, 48]	0
Identity-28	[-1, 48, 48, 48, 48]	0
LayerNorm-29	[-1, 48, 48, 48, 48]	96
Linear-30	[-1, 48, 48, 48, 192]	9,408
GELU-31	[-1, 48, 48, 48, 192]	0
Dropout-32	[-1, 48, 48, 48, 192]	0
Linear-33	[-1, 48, 48, 48, 48]	9,264
Dropout-34	[-1, 48, 48, 48, 48]	0
MLPBlock-35	[-1, 48, 48, 48, 48]	0
Identity-36	[-1, 48, 48, 48, 48]	0
SwinTransformerBlock-37	[-1, 48, 48, 48, 48]	0
LayerNorm-38	[-1, 24, 24, 24, 384]	768
Linear-39	[-1, 24, 24, 24, 96]	36,864
PatchMerging-40	[-1, 24, 24, 24, 96]	0
BasicLayer-41	[-1, 96, 24, 24, 24]	0
LayerNorm-42	[-1, 24, 24, 24, 96]	192
Linear-43	[-1, 343, 288]	27,936
Softmax-44	[-1, 6, 343, 343]	0
Dropout-45	[-1, 6, 343, 343]	0
Linear-46	[-1, 343, 96]	9,312
Dropout-47	[-1, 343, 96]	0

Layer (type)	Output Shape	Param #
WindowAttention-48	[-1, 343, 96]	0
Identity-49	[-1, 24, 24, 24, 96]	0
LayerNorm-50	[-1, 24, 24, 24, 96]	192
Linear-51	[-1, 24, 24, 24, 384]	37,248
GELU-52	[-1, 24, 24, 24, 384]	0
Dropout-53	[-1, 24, 24, 24, 384]	0
Linear-54	[-1, 24, 24, 24, 96]	36,960
Dropout-55	[-1, 24, 24, 24, 96]	0
MLPBlock-56	[-1, 24, 24, 24, 96]	0
Identity-57	[-1, 24, 24, 24, 96]	0
SwinTransformerBlock-58	[-1, 24, 24, 24, 96]	0
LayerNorm-59	[-1, 24, 24, 24, 96]	192
Linear-60	[-1, 343, 288]	27,936
Softmax-61	[-1, 6, 343, 343]	0
Dropout-62	[-1, 6, 343, 343]	0
Linear-63	[-1, 343, 96]	9,312
Dropout-64	[-1, 343, 96]	0
WindowAttention-65	[-1, 343, 96]	0
Identity-66	[-1, 24, 24, 24, 96]	0
LayerNorm-67	[-1, 24, 24, 24, 96]	192
Linear-68	[-1, 24, 24, 24, 384]	37,248
GELU-69	[-1, 24, 24, 24, 384]	0
Dropout-70	[-1, 24, 24, 24, 384]	0
Linear-71	[-1, 24, 24, 24, 96]	36,960
Dropout-72	[-1, 24, 24, 24, 96]	0
MLPBlock-73	[-1, 24, 24, 24, 96]	0
Identity-74	[-1, 24, 24, 24, 96]	0
SwinTransformerBlock-75	[-1, 24, 24, 24, 96]	0
LayerNorm-76	[-1, 12, 12, 12, 768]	1,536
Linear-77	[-1, 12, 12, 12, 192]	147,456
PatchMerging-78	[-1, 12, 12, 12, 192]	0
BasicLayer-79	[-1, 192, 12, 12, 12]	0
LayerNorm-80	[-1, 12, 12, 12, 192]	384
Linear-81	[-1, 343, 576]	111,168
Softmax-82	[-1, 12, 343, 343]	0
Dropout-83	[-1, 12, 343, 343]	0
Linear-84	[-1, 343, 192]	37,056
Dropout-85	[-1, 343, 192]	0
WindowAttention-86	[-1, 343, 192]	0
Identity-87	[-1, 12, 12, 12, 192]	0
LayerNorm-88	[-1, 12, 12, 12, 192]	384
Linear-89	[-1, 12, 12, 12, 768]	148,224
GELU-90	[-1, 12, 12, 12, 768]	0
Dropout-91	[-1, 12, 12, 12, 768]	0
Linear-92	[-1, 12, 12, 12, 192]	147,648
Dropout-93	[-1, 12, 12, 12, 192]	0
MLPBlock-94	[-1, 12, 12, 12, 192]	0
Identity-95	[-1, 12, 12, 12, 192]	0
SwinTransformerBlock-96	[-1, 12, 12, 12, 192]	0
LayerNorm-97	[-1, 12, 12, 12, 192]	384

Layer (type)	Output Shape	Param #
Linear-98	[-1, 343, 576]	111,168
Softmax-99	[-1, 12, 343, 343]	0
Dropout-100	[-1, 12, 343, 343]	0
Linear-101	[-1, 343, 192]	37,056
Dropout-102	[-1, 343, 192]	0
WindowAttention-103	[-1, 343, 192]	0
Identity-104	[-1, 12, 12, 12, 192]	0
LayerNorm-105	[-1, 12, 12, 12, 192]	384
Linear-106	[-1, 12, 12, 12, 768]	148,224
GELU-107	[-1, 12, 12, 12, 768]	0
Dropout-108	[-1, 12, 12, 12, 768]	0
Linear-109	[-1, 12, 12, 12, 192]	147,648
Dropout-110	[-1, 12, 12, 12, 192]	0
MLPBlock-111	[-1, 12, 12, 12, 192]	0
Identity-112	[-1, 12, 12, 12, 192]	0
SwinTransformerBlock-113	[-1, 12, 12, 12, 192]	0
LayerNorm-114	[-1, 6, 6, 6, 1536]	3,072
Linear-115	[-1, 6, 6, 6, 384]	589,824
PatchMerging-116	[-1, 6, 6, 6, 384]	0
BasicLayer-117	[-1, 384, 6, 6, 6]	0
LayerNorm-118	[-1, 6, 6, 6, 384]	768
Linear-119	[-1, 216, 1152]	443,520
Softmax-120	[-1, 24, 216, 216]	0
Dropout-121	[-1, 24, 216, 216]	0
Linear-122	[-1, 216, 384]	147,840
Dropout-123	[-1, 216, 384]	0
WindowAttention-124	[-1, 216, 384]	0
Identity-125	[-1, 6, 6, 6, 384]	0
LayerNorm-126	[-1, 6, 6, 6, 384]	768
Linear-127	[-1, 6, 6, 6, 1536]	591,360
GELU-128	[-1, 6, 6, 6, 1536]	0
Dropout-129	[-1, 6, 6, 6, 1536]	0
Linear-130	[-1, 6, 6, 6, 384]	590,208
Dropout-131	[-1, 6, 6, 6, 384]	0
MLPBlock-132	[-1, 6, 6, 6, 384]	0
Identity-133	[-1, 6, 6, 6, 384]	0
SwinTransformerBlock-134	[-1, 6, 6, 6, 384]	0
LayerNorm-135	[-1, 6, 6, 6, 384]	768
Linear-136	[-1, 216, 1152]	443,520
Softmax-137	[-1, 24, 216, 216]	0
Dropout-138	[-1, 24, 216, 216]	0
Linear-139	[-1, 216, 384]	147,840
Dropout-140	[-1, 216, 384]	0
WindowAttention-141	[-1, 216, 384]	0
Identity-142	[-1, 6, 6, 6, 384]	0
LayerNorm-143	[-1, 6, 6, 6, 384]	768
Linear-144	[-1, 6, 6, 6, 1536]	591,360
GELU-145	[-1, 6, 6, 6, 1536]	0
Dropout-146	[-1, 6, 6, 6, 1536]	0
Linear-147	[-1, 6, 6, 6, 384]	590,208

Layer (type)	Output Shape	Param #
Dropout-148	[-1, 6, 6, 6, 384]	0
MLPBlock-149	[-1, 6, 6, 6, 384]	0
Identity-150	[-1, 6, 6, 6, 384]	0
SwinTransformerBlock-151	[-1, 6, 6, 6, 384]	0
LayerNorm-152	[-1, 3, 3, 3, 3072]	6,144
Linear-153	[-1, 3, 3, 3, 768]	2,359,296
PatchMerging-154	[-1, 3, 3, 3, 768]	0
BasicLayer-155	[-1, 768, 3, 3, 3]	0
SwinTransformer-156	[[-1, 48, 48, 48, 48], [-1, 96, 24, 24, 24], [-1, 192, 12, 12, 12], [-1, 384, 6, 6, 6], [-1, 768, 3, 3, 3]]	0
Conv3d-157	[-1, 48, 96, 96, 96]	3,888
InstanceNorm3d-158	[-1, 48, 96, 96, 96]	0
LeakyReLU-159	[-1, 48, 96, 96, 96]	0
Conv3d-160	[-1, 48, 96, 96, 96]	62,208
InstanceNorm3d-161	[-1, 48, 96, 96, 96]	0
Conv3d-162	[-1, 48, 96, 96, 96]	144
InstanceNorm3d-163	[-1, 48, 96, 96, 96]	0
LeakyReLU-164	[-1, 48, 96, 96, 96]	0
UnetResBlock-165	[-1, 48, 96, 96, 96]	0
UnetrBasicBlock-166	[-1, 48, 96, 96, 96]	0
Conv3d-167	[-1, 48, 48, 48, 48]	62,208
InstanceNorm3d-168	[-1, 48, 48, 48, 48]	0
LeakyReLU-169	[-1, 48, 48, 48, 48]	0
Conv3d-170	[-1, 48, 48, 48, 48]	62,208
InstanceNorm3d-171	[-1, 48, 48, 48, 48]	0
LeakyReLU-172	[-1, 48, 48, 48, 48]	0
UnetResBlock-173	[-1, 48, 48, 48, 48]	0
UnetrBasicBlock-174	[-1, 48, 48, 48, 48]	0
Conv3d-175	[-1, 96, 24, 24, 24]	248,832
InstanceNorm3d-176	[-1, 96, 24, 24, 24]	0
LeakyReLU-177	[-1, 96, 24, 24, 24]	0
Conv3d-178	[-1, 96, 24, 24, 24]	248,832
InstanceNorm3d-179	[-1, 96, 24, 24, 24]	0
LeakyReLU-180	[-1, 96, 24, 24, 24]	0
UnetResBlock-181	[-1, 96, 24, 24, 24]	0
UnetrBasicBlock-182	[-1, 96, 24, 24, 24]	0
Conv3d-183	[-1, 192, 12, 12, 12]	995,328
InstanceNorm3d-184	[-1, 192, 12, 12, 12]	0
LeakyReLU-185	[-1, 192, 12, 12, 12]	0
Conv3d-186	[-1, 192, 12, 12, 12]	995,328
InstanceNorm3d-187	[-1, 192, 12, 12, 12]	0
LeakyReLU-188	[-1, 192, 12, 12, 12]	0
UnetResBlock-189	[-1, 192, 12, 12, 12]	0
UnetrBasicBlock-190	[-1, 192, 12, 12, 12]	0
Conv3d-191	[-1, 768, 3, 3, 3]	15,925,248
InstanceNorm3d-192	[-1, 768, 3, 3, 3]	0
LeakyReLU-193	[-1, 768, 3, 3, 3]	0
Conv3d-194	[-1, 768, 3, 3, 3]	15,925,248
InstanceNorm3d-195	[-1, 768, 3, 3, 3]	0
LeakyReLU-196	[-1, 768, 3, 3, 3]	0
UnetResBlock-197	[-1, 768, 3, 3, 3]	0

Layer (type)	Output Shape	Param #
UnetrBasicBlock-198	[-1, 768, 3, 3, 3]	0
ConvTranspose3d-199	[-1, 384, 6, 6, 6]	2,359,296
Conv3d-200	[-1, 384, 6, 6, 6]	7,962,624
InstanceNorm3d-201	[-1, 384, 6, 6, 6]	0
LeakyReLU-202	[-1, 384, 6, 6, 6]	0
Conv3d-203	[-1, 384, 6, 6, 6]	3,981,312
InstanceNorm3d-204	[-1, 384, 6, 6, 6]	0
Conv3d-205	[-1, 384, 6, 6, 6]	294,912
InstanceNorm3d-206	[-1, 384, 6, 6, 6]	0
LeakyReLU-207	[-1, 384, 6, 6, 6]	0
UnetrResBlock-208	[-1, 384, 6, 6, 6]	0
UnetrUpBlock-209	[-1, 384, 6, 6, 6]	0
ConvTranspose3d-210	[-1, 192, 12, 12, 12]	589,824
Conv3d-211	[-1, 192, 12, 12, 12]	1,990,656
InstanceNorm3d-212	[-1, 192, 12, 12, 12]	0
LeakyReLU-213	[-1, 192, 12, 12, 12]	0
Conv3d-214	[-1, 192, 12, 12, 12]	995,328
InstanceNorm3d-215	[-1, 192, 12, 12, 12]	0
Conv3d-216	[-1, 192, 12, 12, 12]	73,728
InstanceNorm3d-217	[-1, 192, 12, 12, 12]	0
LeakyReLU-218	[-1, 192, 12, 12, 12]	0
UnetrResBlock-219	[-1, 192, 12, 12, 12]	0
UnetrUpBlock-220	[-1, 192, 12, 12, 12]	0
ConvTranspose3d-221	[-1, 96, 24, 24, 24]	147,456
Conv3d-222	[-1, 96, 24, 24, 24]	497,664
InstanceNorm3d-223	[-1, 96, 24, 24, 24]	0
LeakyReLU-224	[-1, 96, 24, 24, 24]	0
Conv3d-225	[-1, 96, 24, 24, 24]	248,832
InstanceNorm3d-226	[-1, 96, 24, 24, 24]	0
Conv3d-227	[-1, 96, 24, 24, 24]	18,432
InstanceNorm3d-228	[-1, 96, 24, 24, 24]	0
LeakyReLU-229	[-1, 96, 24, 24, 24]	0
UnetrResBlock-230	[-1, 96, 24, 24, 24]	0
UnetrUpBlock-231	[-1, 96, 24, 24, 24]	0
ConvTranspose3d-232	[-1, 48, 48, 48, 48]	36,864
Conv3d-233	[-1, 48, 48, 48, 48]	124,416
InstanceNorm3d-234	[-1, 48, 48, 48, 48]	0
LeakyReLU-235	[-1, 48, 48, 48, 48]	0
Conv3d-236	[-1, 48, 48, 48, 48]	62,208
InstanceNorm3d-237	[-1, 48, 48, 48, 48]	0
Conv3d-238	[-1, 48, 48, 48, 48]	4,608
InstanceNorm3d-239	[-1, 48, 48, 48, 48]	0
LeakyReLU-240	[-1, 48, 48, 48, 48]	0
UnetrResBlock-241	[-1, 48, 48, 48, 48]	0
UnetrUpBlock-242	[-1, 48, 48, 48, 48]	0
ConvTranspose3d-243	[-1, 48, 96, 96, 96]	18,432
Conv3d-244	[-1, 48, 96, 96, 96]	124,416
InstanceNorm3d-245	[-1, 48, 96, 96, 96]	0
LeakyReLU-246	[-1, 48, 96, 96, 96]	0
Conv3d-247	[-1, 48, 96, 96, 96]	62,208

Layer (type)	Output Shape	Param #
InstanceNorm3d-248	[-1, 48, 96, 96, 96]	0
Conv3d-249	[-1, 48, 96, 96, 96]	4,608
InstanceNorm3d-250	[-1, 48, 96, 96, 96]	0
LeakyReLU-251	[-1, 48, 96, 96, 96]	0
UnetResBlock-252	[-1, 48, 96, 96, 96]	0
UnetrUpBlock-253	[-1, 48, 96, 96, 96]	0
Conv3d-254	[-1, 1, 96, 96, 96]	49
UnetOutBlock-255	[-1, 1, 96, 96, 96]	0

Total params: 61,992,385

Trainable params: 61,992,385

Non-trainable params: 0

Input size (MB): 10.12

Forward/backward pass size (MB):
8796093032538.60

Params size (MB): 236.48

Estimated Total Size (MB):
8796093032785.21

Number of Parameters: 62190115

FLOPs: 330.96 GMac