3D Brain MRI Generation with a Clinically-Conditioned VAE-GAN and Diffusion-Driven Feature Sampling

Najmeh Mashhadi, Emmanouil Nikolakakis, Razvan Marinescu[*]
Department of Computer Science
University of California, Santa Cruz

Supplementary

Table S1: Training and inference runtimes for framework components.

Model	Batch Size (train)	Learning rate	Epochs	Training Time	Inference Time Per Sample (batch=1)
VAE-GAN	4	1×10^{-4}	2000	~ 62 h	164 ms
AD/CN Classifier	8	1×10^{-4}	51	~ 1 h	151 ms
Volume Regression	8	1×10^{-4}	56	~ 1 h	151 ms
Diffusion Model	8	5×10^{-4}	1000	< 1 h	450 ms (500 steps)

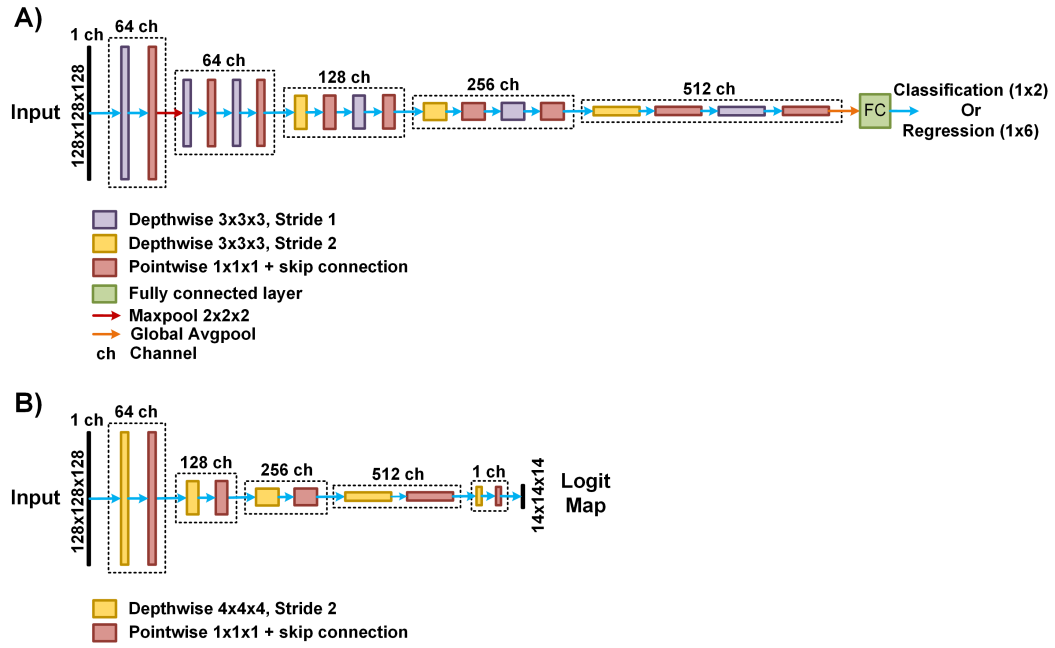[*]R. Marinescu is the corresponding author (`ramarine@ucsc.edu`).

Figure S1: A) Schematic of the 3D depth-wise–separable ResNet head used for clinical supervision: stacks of depth-wise plus point-wise residual blocks with max-pool and global average pooling; A final FC outputs either two logits (AD/CN) or a 6-value volume vector. B) Schematic of the 3D PatchGAN discriminator: depth-wise blocks with residual connections down-sample the 128^3 input to a $14 \times 14 \times 14$ logit map, so it checks realism patch by patch (one score per patch) across the whole 3D scan.