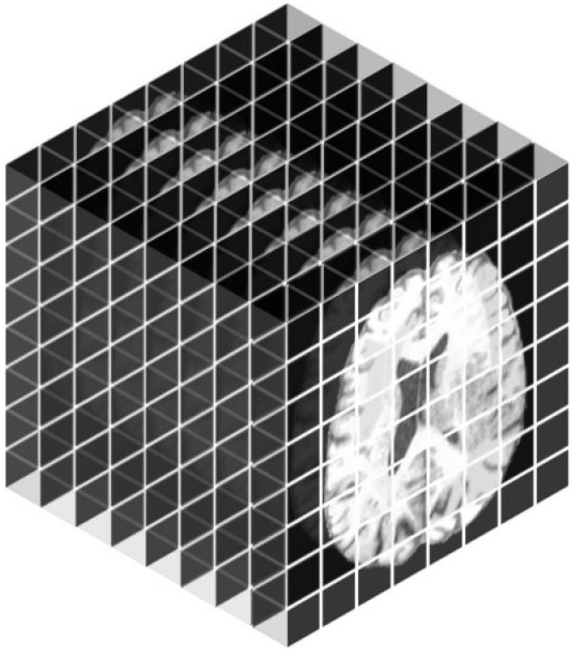
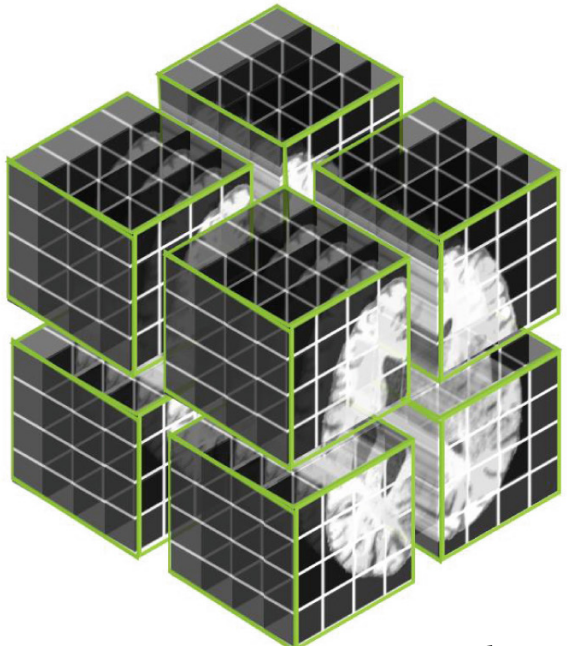


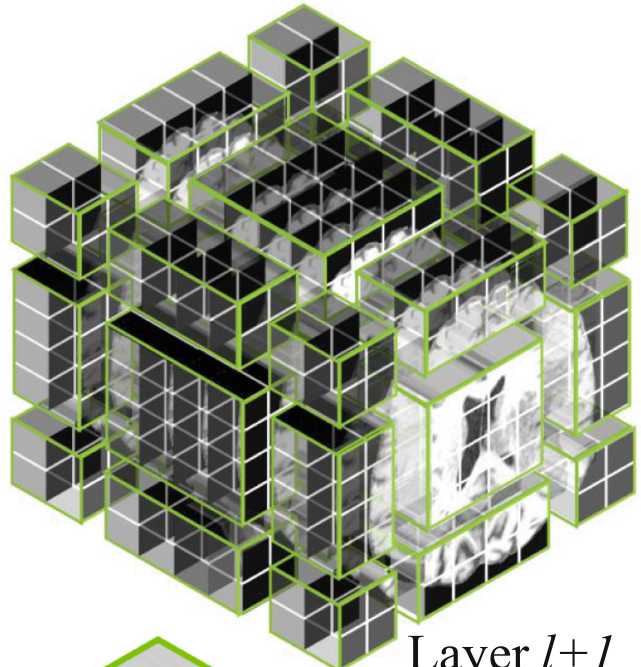
3D Tokens: 8× 8 × 8

Window size: 4× 4 × 4



Layer *l*

Number of windows: 8



Layer *l*+*1*

Self-attention Unit