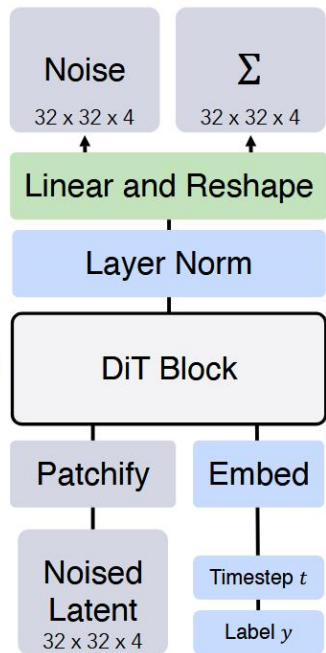
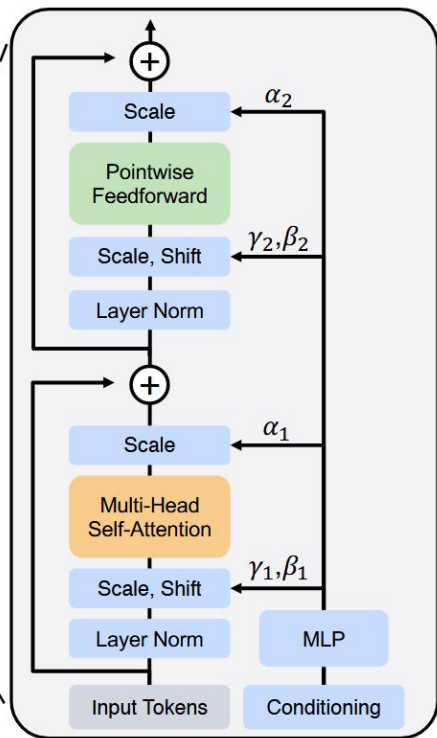


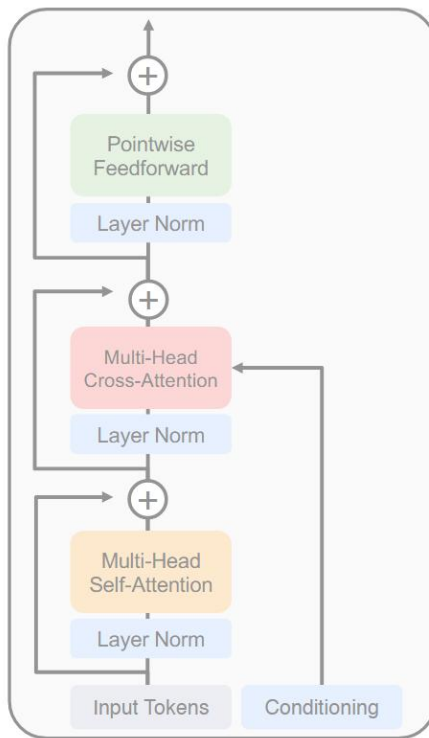
$N \times$



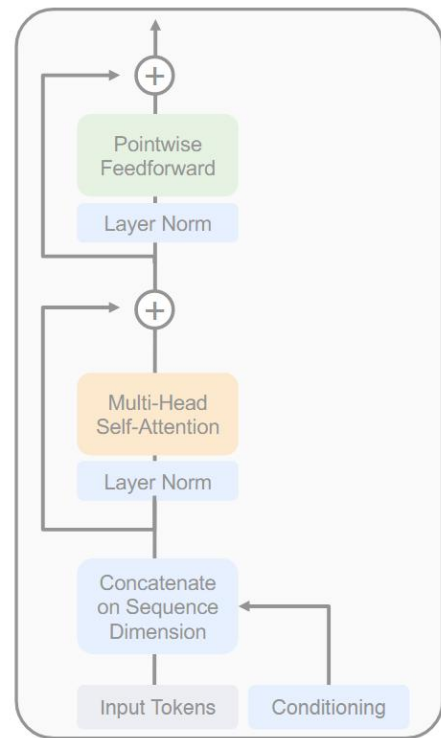
Latent Diffusion Transformer



DiT Block with adaLN-Zero



DiT Block with Cross-Attention



DiT Block with In-Context Conditioning