(a) Overall Architecture of Implicit-Explicit Diffusion Model

Input (x, mask, t) → Input Projection, Mask Embedding → + → Implicit Branch: Implicit Module (Dilated Conv); Explicit Branch: Explicit Module (S4 Layer ×2) → Gated Fusion → + ← Time Embed → Residual Blocks ×6 → Output Proj → Output