

(c) Multimodal transformer with diffusion head