Training Phase

Skeleton

ϵ

Frequency-aware Diffusion Model

Skeleton Encoder

z_0

Add Noise

SG-SRM

ϵ^θ

$\hat{z_0}$

Diffusion Loss

ϵ

ϵ^θ

Curriculum-based Semantic Abstraction

Action Label: Yawning

LLM description: move the open hand up to the front of the head then look up and lean the torso back slightly

Text Encoder

z_0

$\hat{z_0}$

Timestep-Adaptive Spectral Loss

Inference Phase

Skeleton

ϵ

Frequency-aware Diffusion Model

Skeleton Encoder

z_0

Add Noise

SG-SRM

ϵ^θ

$\hat{z_0}$

ϵ

ϵ^θ

Candidate Action Label 1: Yawning
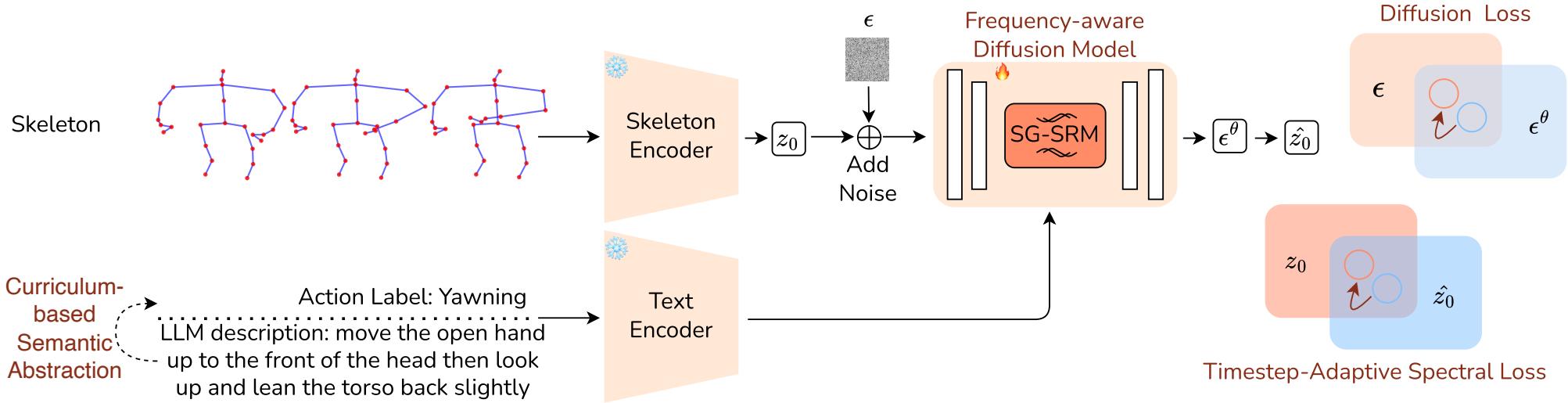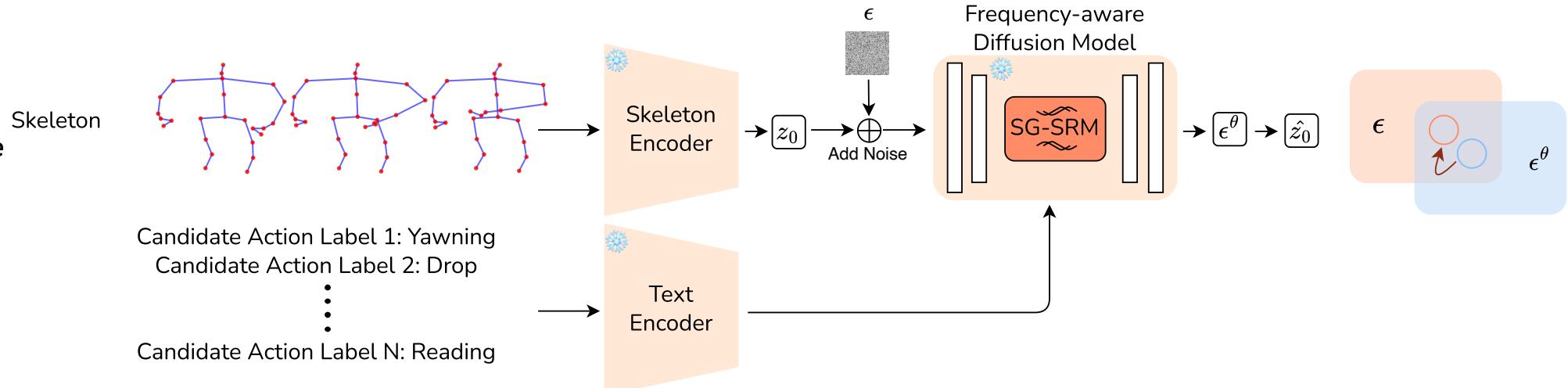Candidate Action Label 2: Drop
⋮
Candidate Action Label N: Reading

Text Encoder

SG-SRM Semantic-Guided Spectral Residual Module

z_0 skeleton latent

$\hat{z_0}$ estimated skeleton latent