

Image Compressor

Input Range Image x

CurveConv

Downsample

Downsample

Encode

Decode

Upsample

Upsample

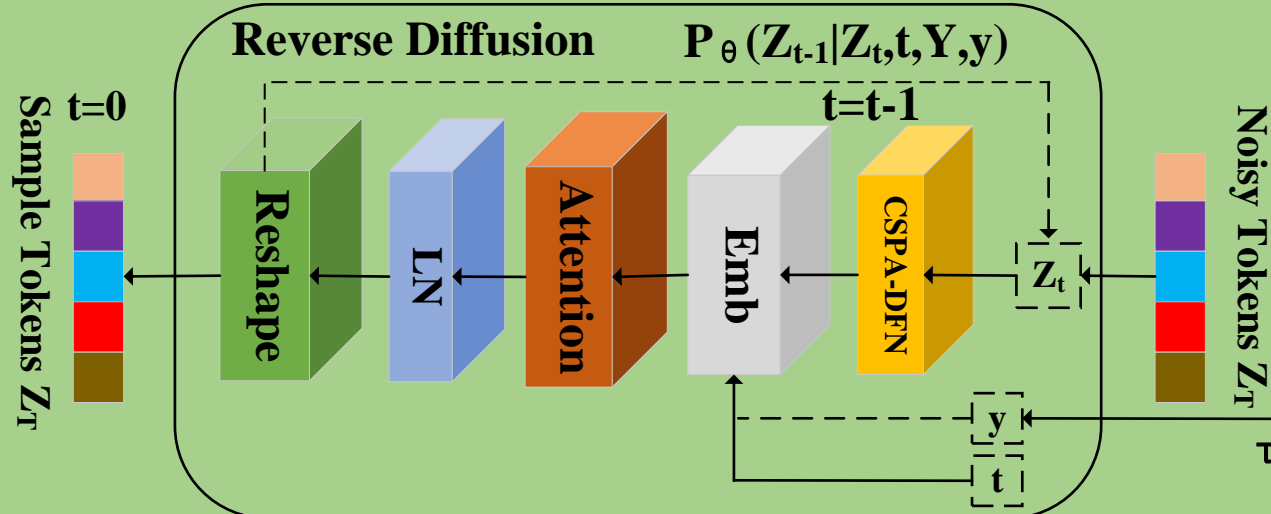
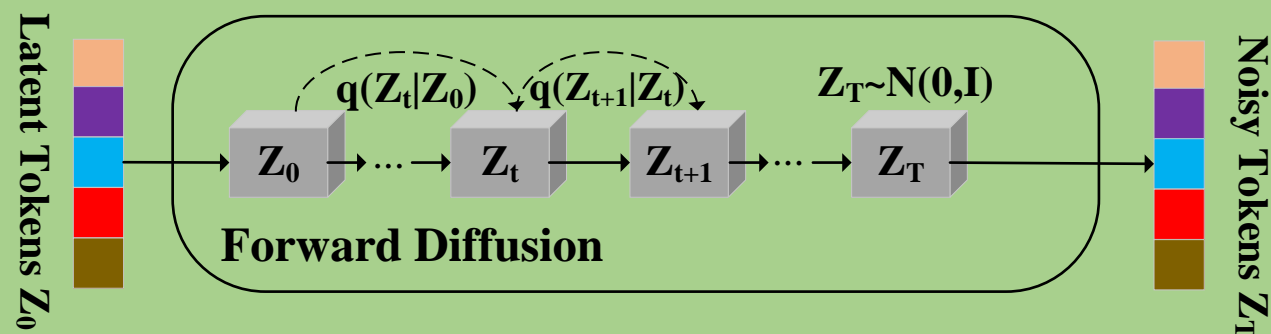
CurveConv

Postprocess

Output Range Image x

Diffusion Generator

CSPA-DFN: Channel-Spatial Parallel Attention and Dilation Fusion Network



Emb: Token Embedding

LN: LayerNorm

Attention: Global Self-attention

CurveConv: Curve-wise Convolution

Postprocess: ReLU Linear CrossAttention between range image and voxel features

Condition Injection

Semantic Map

Camera Images

Text

Condition Network