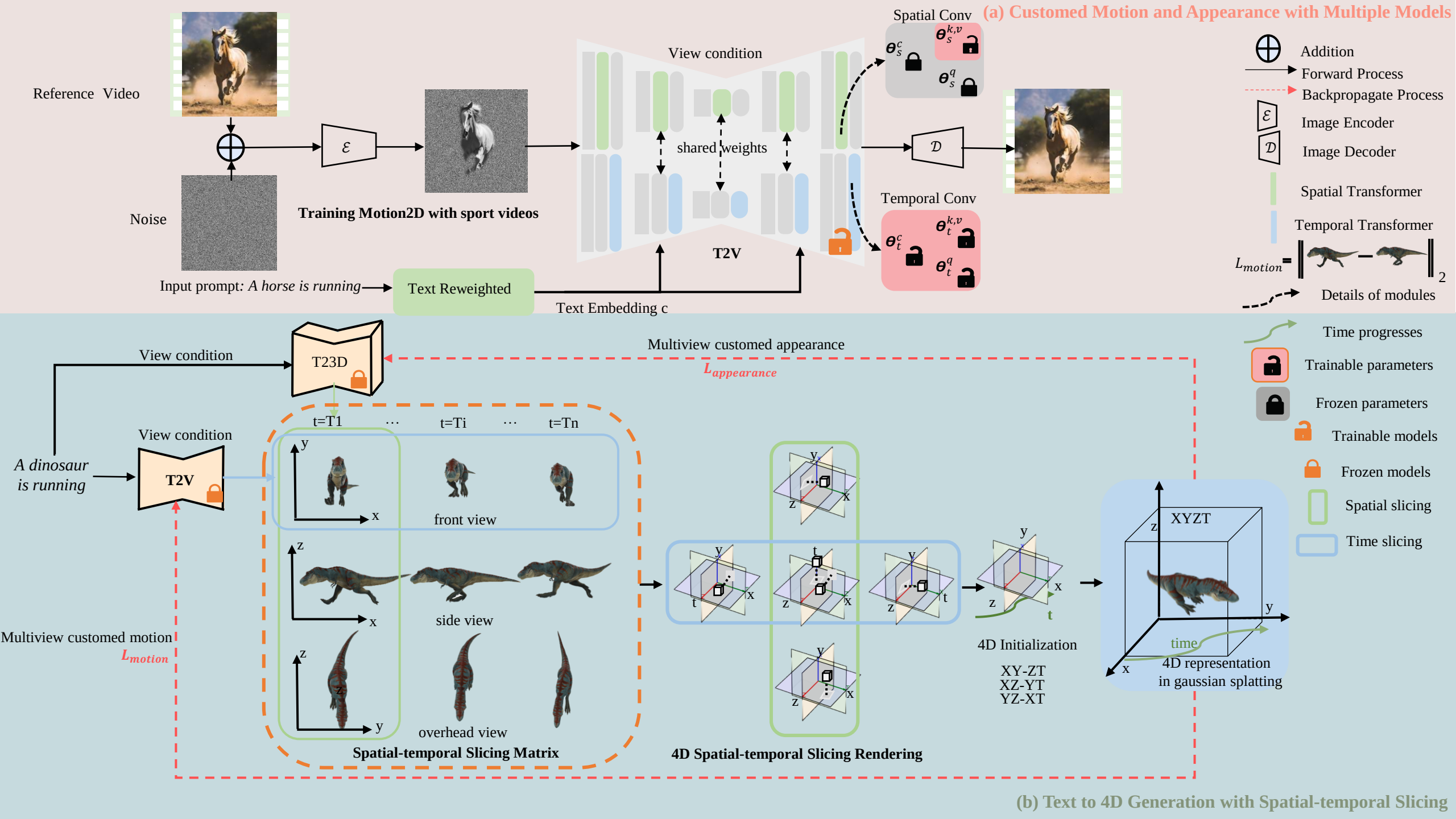
(a) Customed Motion and Appearance with Multiple Models
Reference Video
Noise
Training Motion2D with sport videos
View condition
shared weights
T2V
Spatial Conv
Temporal Conv
Input prompt: A horse is running
Text Reweighted
Text Embedding c
Addition
Forward Process
Backpropagate Process
Image Encoder
Image Decoder
Spatial Transformer
Temporal Transformer
Details of modules
Multiview customed appearance
T23D
View condition
A dinosaur is running
T2V
t=T1
t=Ti
t=Tn
front view
side view
overhead view
Spatial-temporal Slicing Matrix
Multiview customed motion
4D Spatial-temporal Slicing Rendering
4D Initialization
XY-ZT
XZ-YT
YZ-XT
XYZT
time
4D representation
in gaussian splatting
Time progresses
Trainable parameters
Frozen parameters
Trainable models
Frozen models
Spatial slicing
Time slicing
(b) Text to 4D Generation with Spatial-temporal Slicing