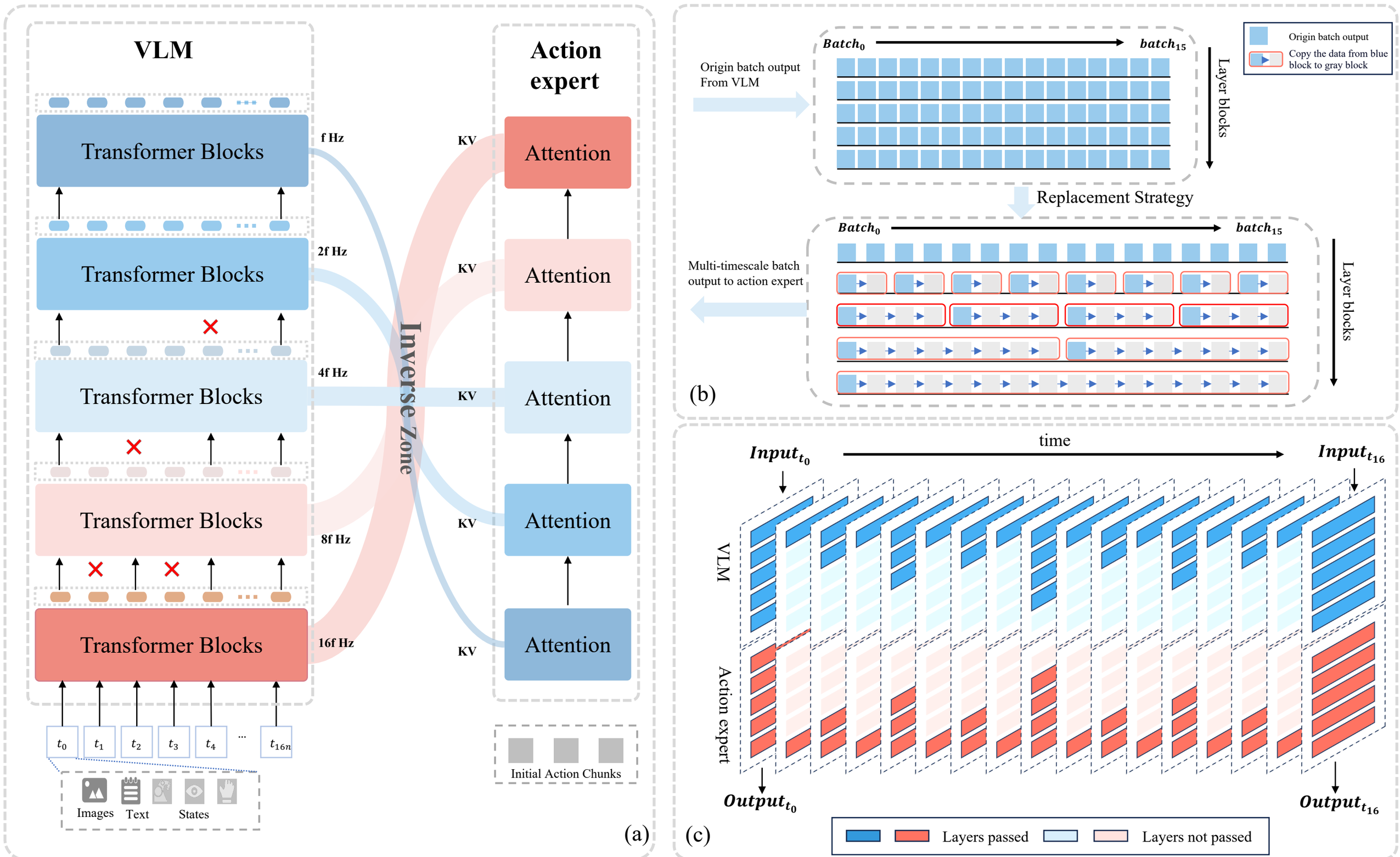
VLM
Action expert
Transformer Blocks
Transformer Blocks
Transformer Blocks
Transformer Blocks
Transformer Blocks
f Hz
2f Hz
4f Hz
8f Hz
16f Hz
KV
KV
KV
KV
KV
Inverse Zone
Attention
Attention
Attention
Attention
Attention
t_0 t_1 t_2 t_3 t_4 ... t_{16n}
Images
Text
States
Initial Action Chunks
(a)
Origin batch output From VLM
$Batch_0$
$batch_{15}$
Layer blocks
Origin batch output
Copy the data from blue block to gray block
Replacement Strategy
Multi-timescale batch output to action expert
$Batch_0$
$batch_{15}$
Layer blocks
(b)
time
$Input_{t_0}$
$Input_{t_{16}}$
VLM
Action expert
$Output_{t_0}$
$Output_{t_{16}}$
Layers passed
Layers not passed
(c)