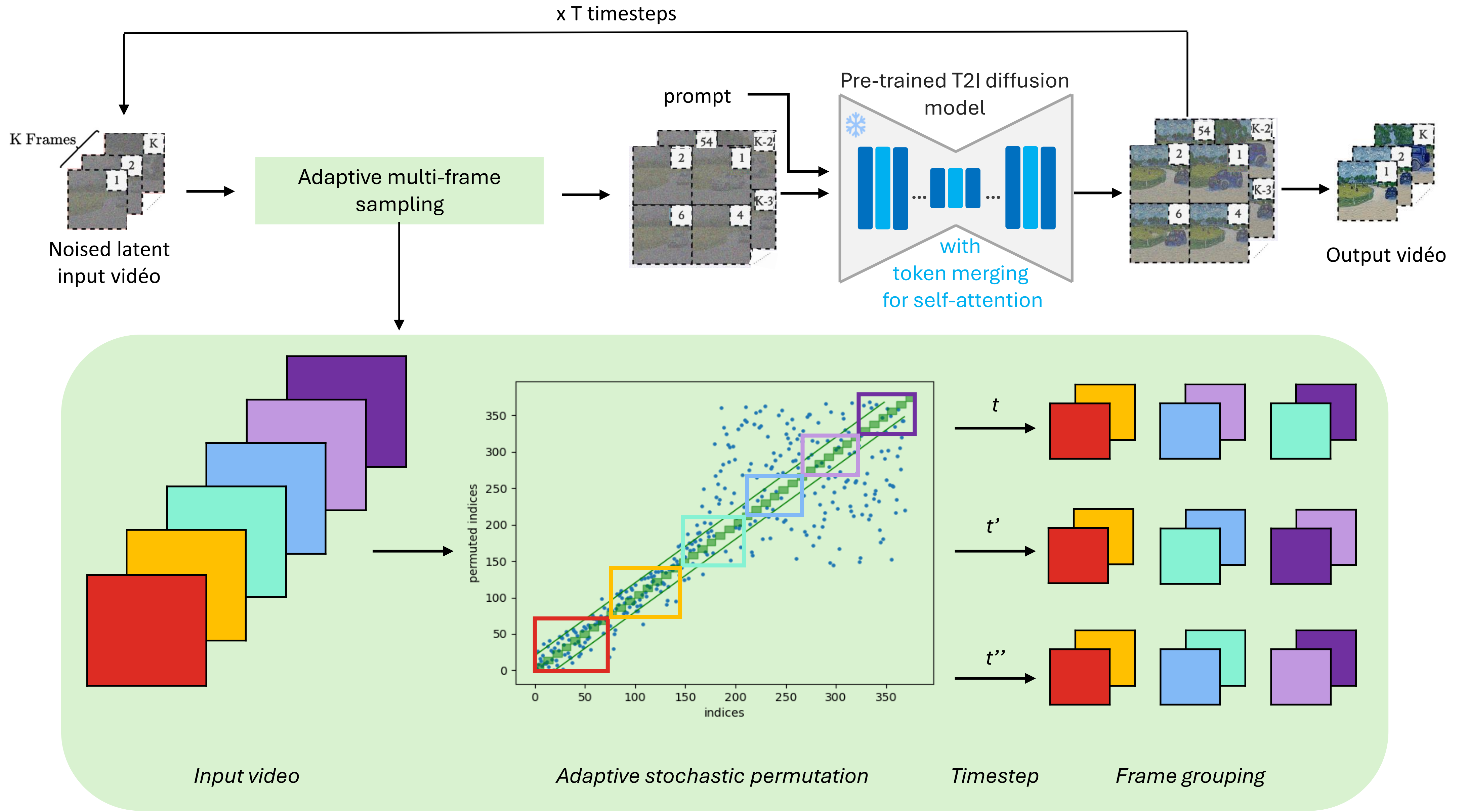
x T timesteps
K Frames
K
2
1
Noised latent input vidéo
Adaptive multi-frame sampling
prompt
54
K-2
K-3
6
4
Pre-trained T2I diffusion model
with token merging for self-attention
Output vidéo
permuted indices
0
50
100
150
200
250
300
350
indices
t
t'
t''
Input video
Adaptive stochastic permutation
Timestep
Frame grouping