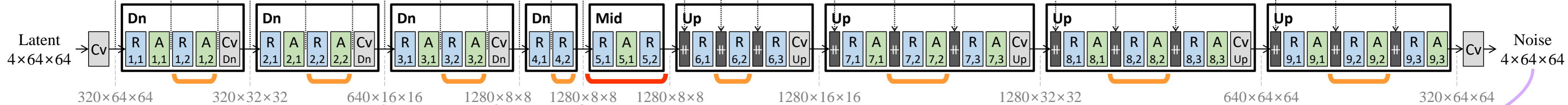


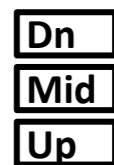
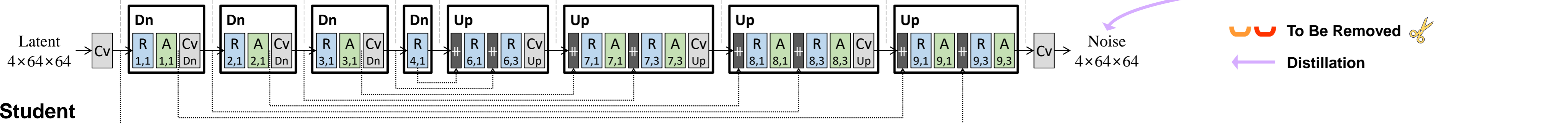
# Teacher

Original  
SDM

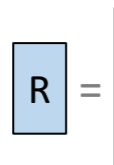


# Student

Compact  
BK-SDM-Small

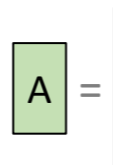


Down,  
Mid-,  
Up Stages



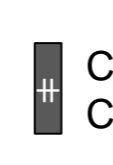
ResBlock  
(time-embedding  
conditioned)

ResBlock  
(time-embedding  
conditioned)

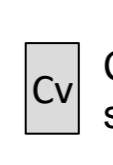


CrossAttentionBlock  
(text-embedding  
conditioned)

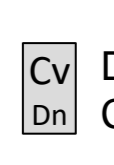
CrossAttentionBlock  
(text-embedding  
conditioned)



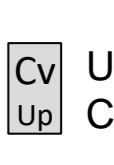
Channel  
Concatenation



Conv (spatial  
size preserved)



Down-sizing  
Conv



Up-sizing  
Conv