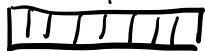


spatial attention



attention between frames

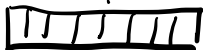
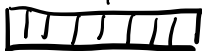


image model

video model