AudioVisual Retrieval (AVR)



Query: With the sound of the wind and the sound of birds, a man in white tried to row in the water.

Rank top-3 candidates by AVLNet:







Rank top-3 candidates by VALOR(Ours):







AudioVisual Captioning (AVC)









<u>GT:</u> With the sound of the wind and the sound of birds, a man in white tried to row in the water.

Generated by SMPFF: With the sound of the wind and the sound of the water a man in a white coat sat on a boat on the water.

<u>Generated by VALOR(Ours)</u>: With the wind and the sound of birds a man in a white coat paddled a white boat over the water.

<u>GT:</u> The English letters rolled up in the sound of firecrackers.

Generated by SMPFF: In the dark room a man was beating the fireworks making a crackling sound.

Generated by VALOR(Ours): In the room, fireworks were set off in the air, making a crackling sound.

<u>GT:</u> As the noise rang, a yellow train slowly turned, the picture turned, and a white train started slowly from the stillness with the noise.

Generated by SMPFF: A train was slowly moving along the track.

Generated by VALOR(Ours): A train was slowly moving along the track with the roar of the engine and the voice of people.