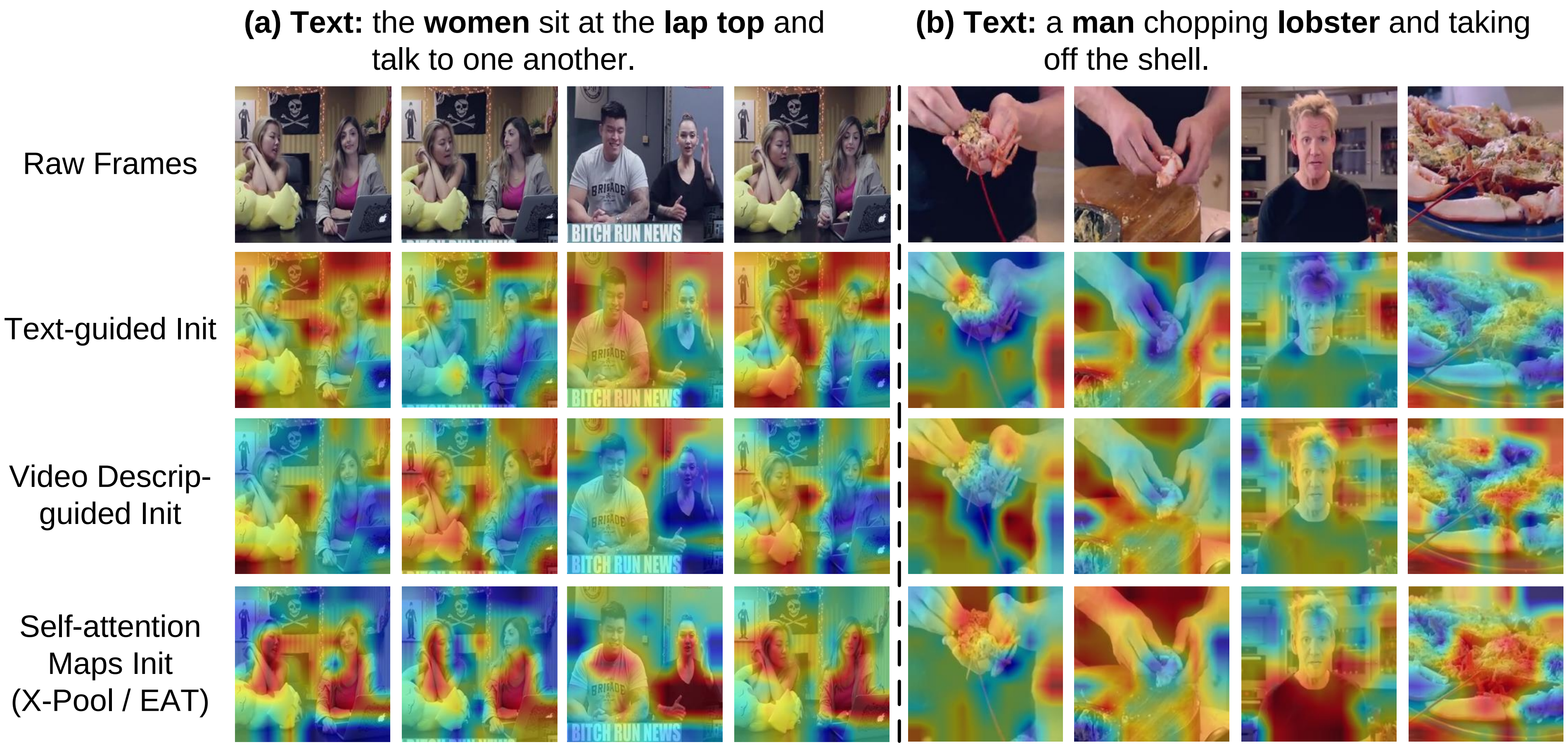

(a) Text: the **women** sit at the **lap top** and talk to one another.

(b) Text: a **man** chopping **lobster** and taking off the shell.

Raw Frames
Text-guided Init
Video Descrip-guided Init
Self-attention Maps Init (X-Pool / EAT)
BITCH RUN NEWS