Synthesize the training image-text pairs

Caption on facial actions

The face shows the eyelids tightening and the cheeks rising. … Additionally, the chin may appear wrinkled.
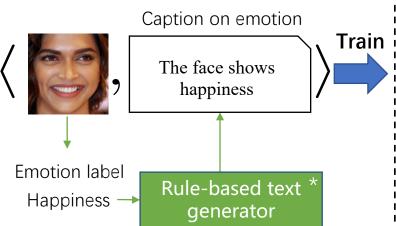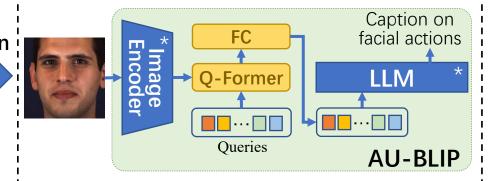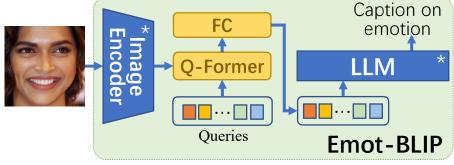
GPT-3.5 *

AU label
0,0,⋯,1,0,0

Prompt

Caption on emotion

The face shows happiness

Emotion label
Happiness

Rule-based text generator *

Train AU-BLIP and Emot-BLIP to describe on single aspect

Train

Image Encoder *

FC

Q-Former

Caption on facial actions

LLM *

Queries

AU-BLIP

Train

Image Encoder *

FC

Q-Former

Caption on emotion

LLM *

Queries

Emot-BLIP

Train Exp-BLIP to describe on both aspects

Caption on emotion

AU-BLIP

Caption on facial actions

Caption on facial actions

Emot-BLIP

Caption on emotion

⊕

Prompt

GPT-3.5 *

Caption on facial expressions

⊕

Prompt

GPT-3.5 *

Caption on facial expressions

This face is happy as the corners of the lips are pulled back, …

With a wide smile, … , infers a sense of joy and satisfaction.

Train

Image Encoder *

FC

Q-Former

Caption on facial expressions

LLM *

Queries

Exp-BLIP