

## Modality-specific



Visual  
Encoder

SNN

Dynamic  
Threshold  
Modification

SNN

Visual  
Encoder

You get every  
right to a  
beautiful life.

Textual  
Encoder

## Modality-interactive

Spiking-scaled  
attention

Cross-Attention

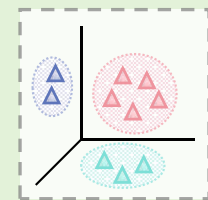
Spiking-scaled  
attention

Cross-Attention

## Label-Modality

Multi-label  
Classifier

Label-Modality  
Matching  
Attention



Label inputs