

- Movies
- TV shows
- Advertisements
- Natural scenes images

**Multimodal
content**

Context

Local context
Visual scene context
Foreground context
Temporal context

Visual, Language, Audio
modalities

- ★ Vision-language models
- ★ Language models
- ★ Visual vocabularies
- ★ Screenplays
- ★ ASR

**External
knowledge**

✓ Human affect understanding

Instance-level

**Context
guided
attention**

Macro-level

- ✓ Topic
- ✓ Social relevance
- ✓ Tone transition
- ✓ Genre