



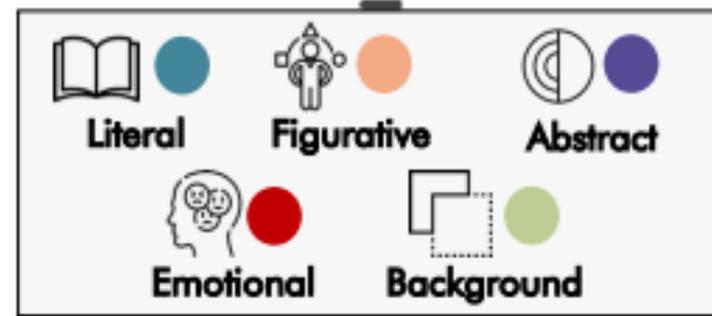
### One Image $\neq$ One Meaning

**Literal Caption:** A tapestry of watermelons, oranges, and grapes spills across the wall, each slice a brushstroke of summer.

① Baseline (single literal caption)

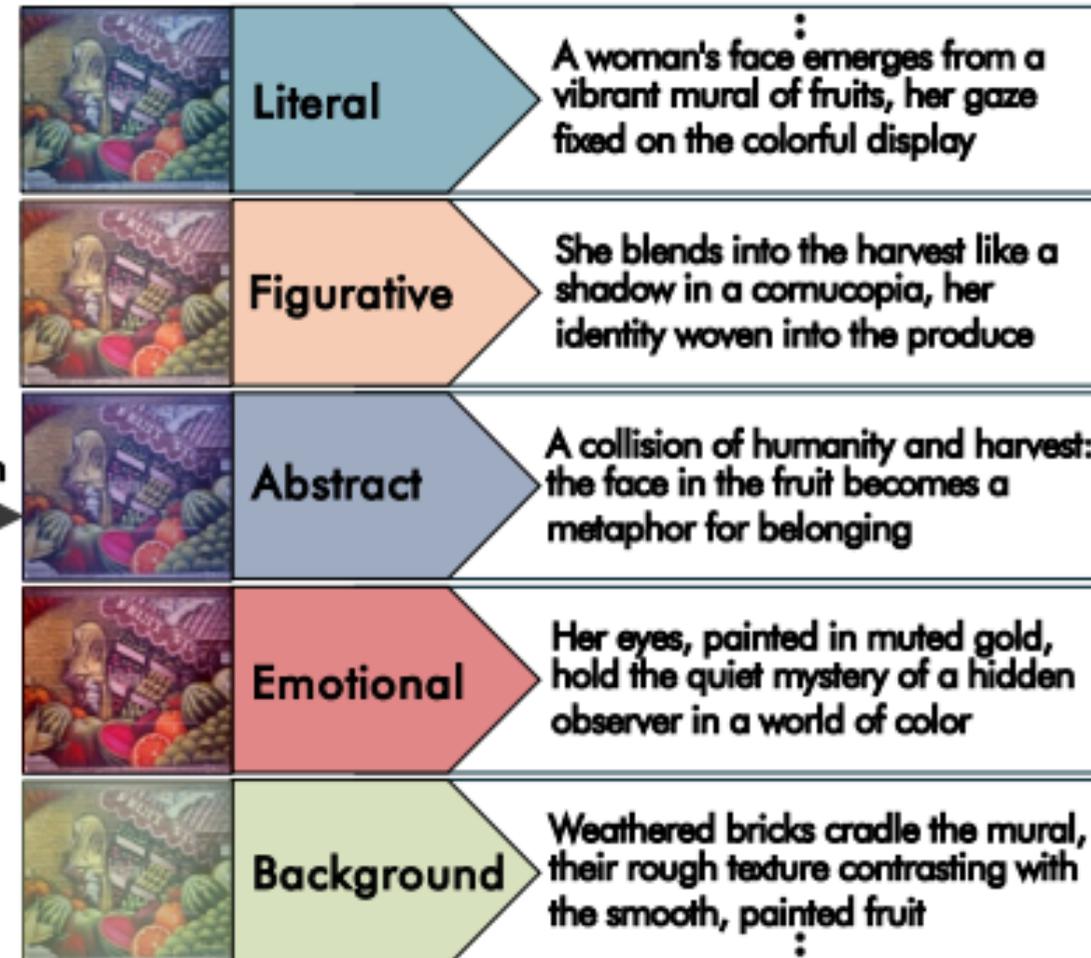


② Five Guided Lenses



③ Shared Lens Embeddings

Caption

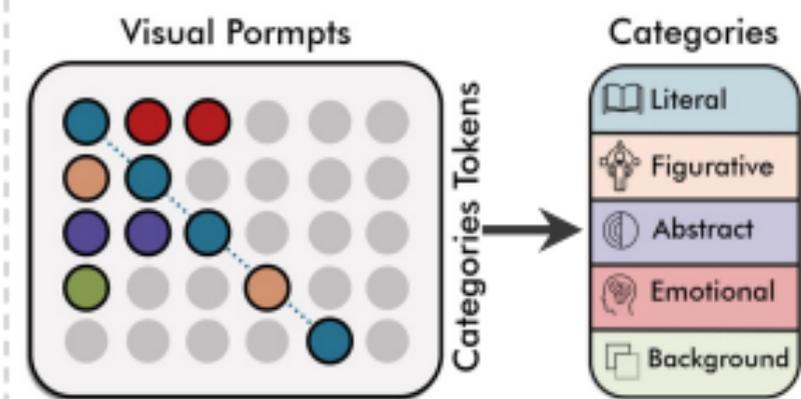


④ Grounded Captions by Perspective

### A) Multi-Lens Views



### B) Category-Aligned Set Matching



⑤ Multi-Perspective Views & Alignment