



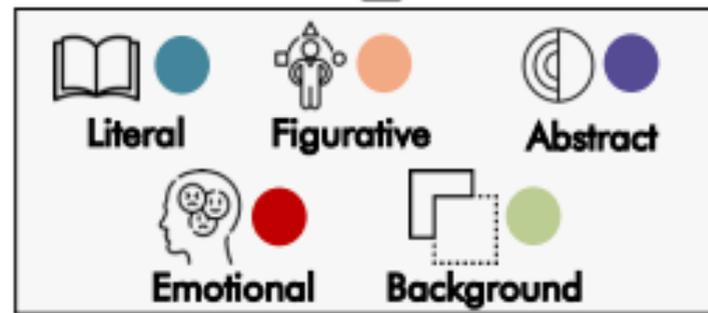
### Baseline (single literal caption)

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

One Image  $\neq$  One Meaning

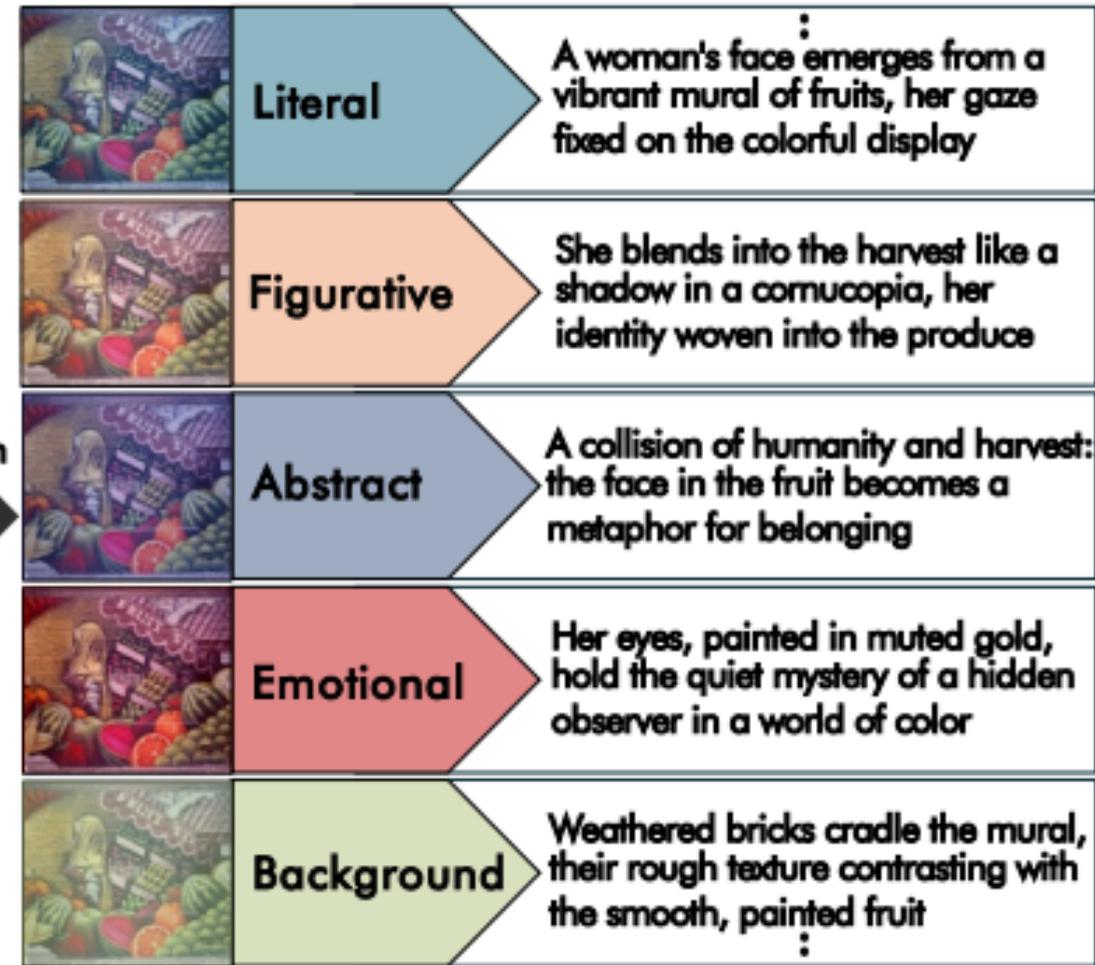


Five Guided Lenses



Shared Lens Embeddings

Caption



Grounded Captions by Perspective

### A) Multi-Lens Views



### B) Category-Aligned Set Matching

