



Generative Perception of Shape and Material from Differential Motion

Xinran Nicole Han, Ko Nishino, Todd Zickler



Check our project page for videos!

Motivation

image(s) \longrightarrow shape, materials

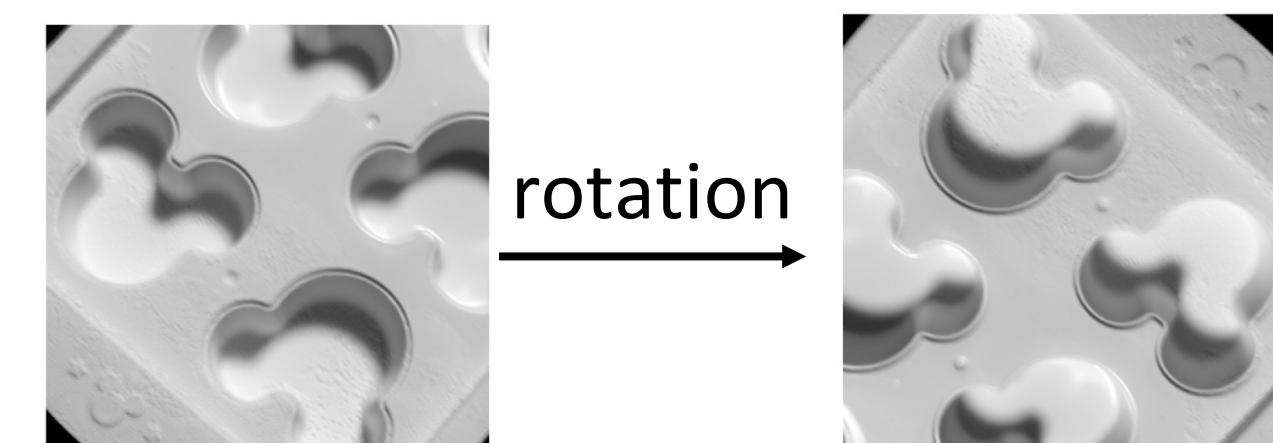
Perceiving shape and material from a single image is inherently ambiguous and requires joint reasoning.

Ambiguous shape and material perception



Doerschner et al., 2011

Ambiguous shape perception



Han, Zickler & Nishino, NeurIPS 2024



Image generated by Nano Banana

Humans often resolve ambiguity by moving the object around or rotating our head slightly.

What's missing in existing methods?

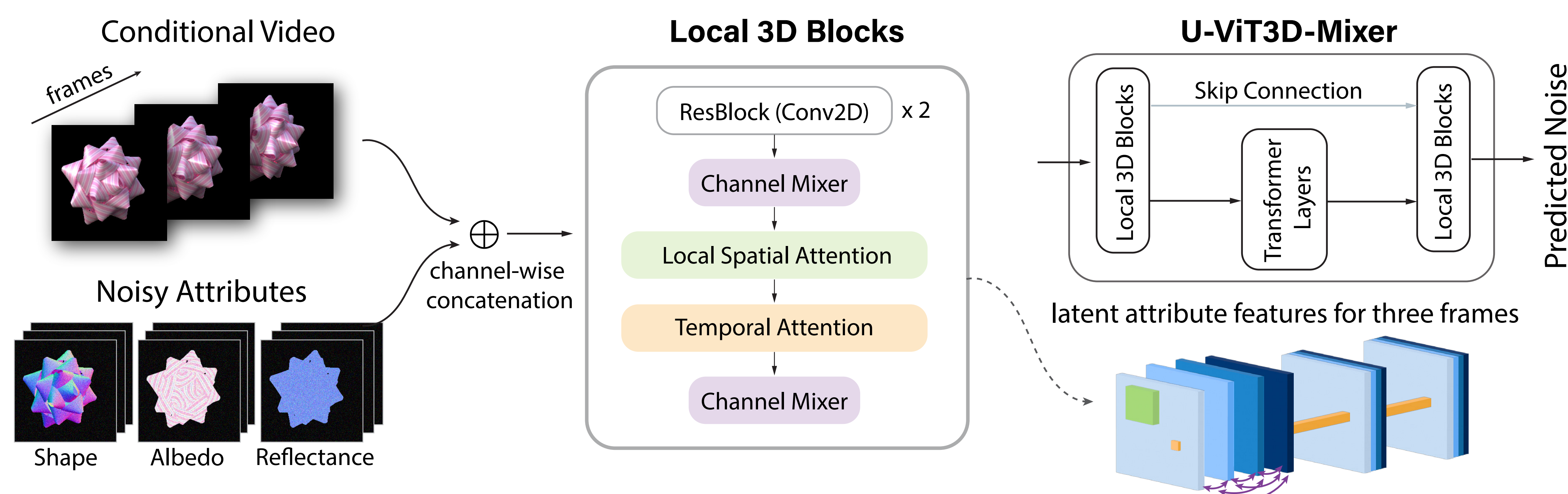
- Ability to sample diverse explanations / ambiguity-awareness
- Simultaneous disentanglement of shape and materials
- Ability to exploit differential motion to reduce ambiguities

We propose “**generative perception**” where we infer a distribution of plausible explanations from static or motion observations:

$$s(t), m(t) \sim p_{\theta}(S(t), M(t) \mid I(t))$$

Model Architecture

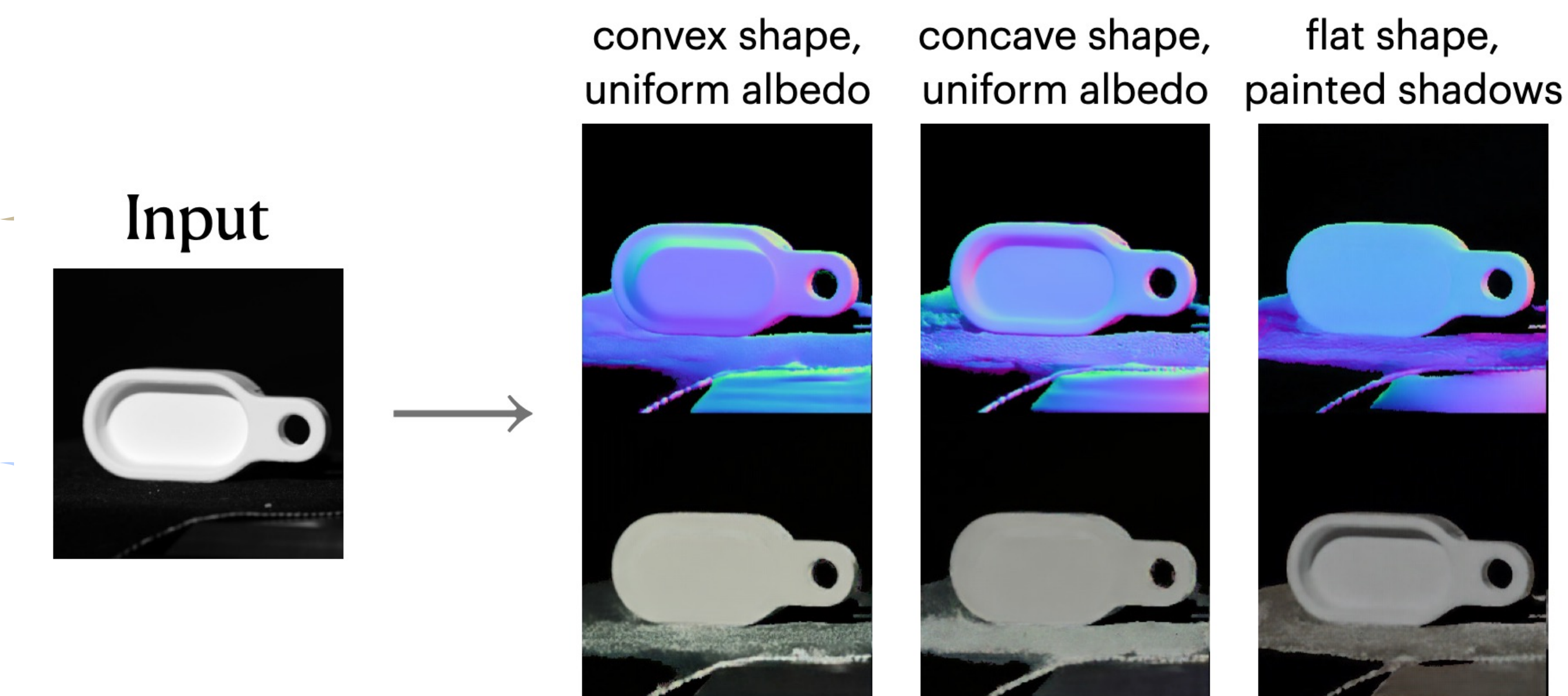
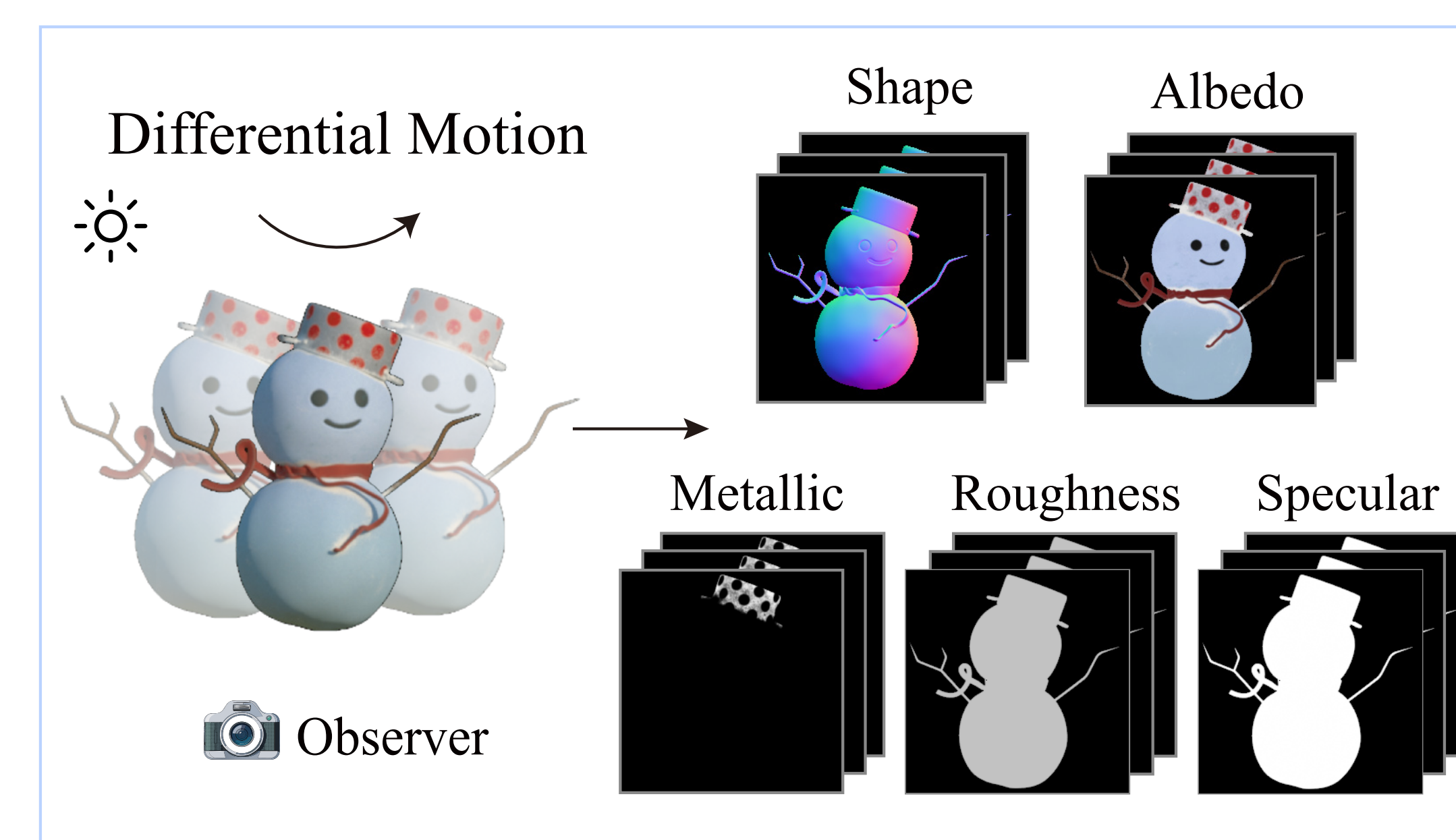
Pixel-space (conditional) Video Diffusion for Joint Disentanglement



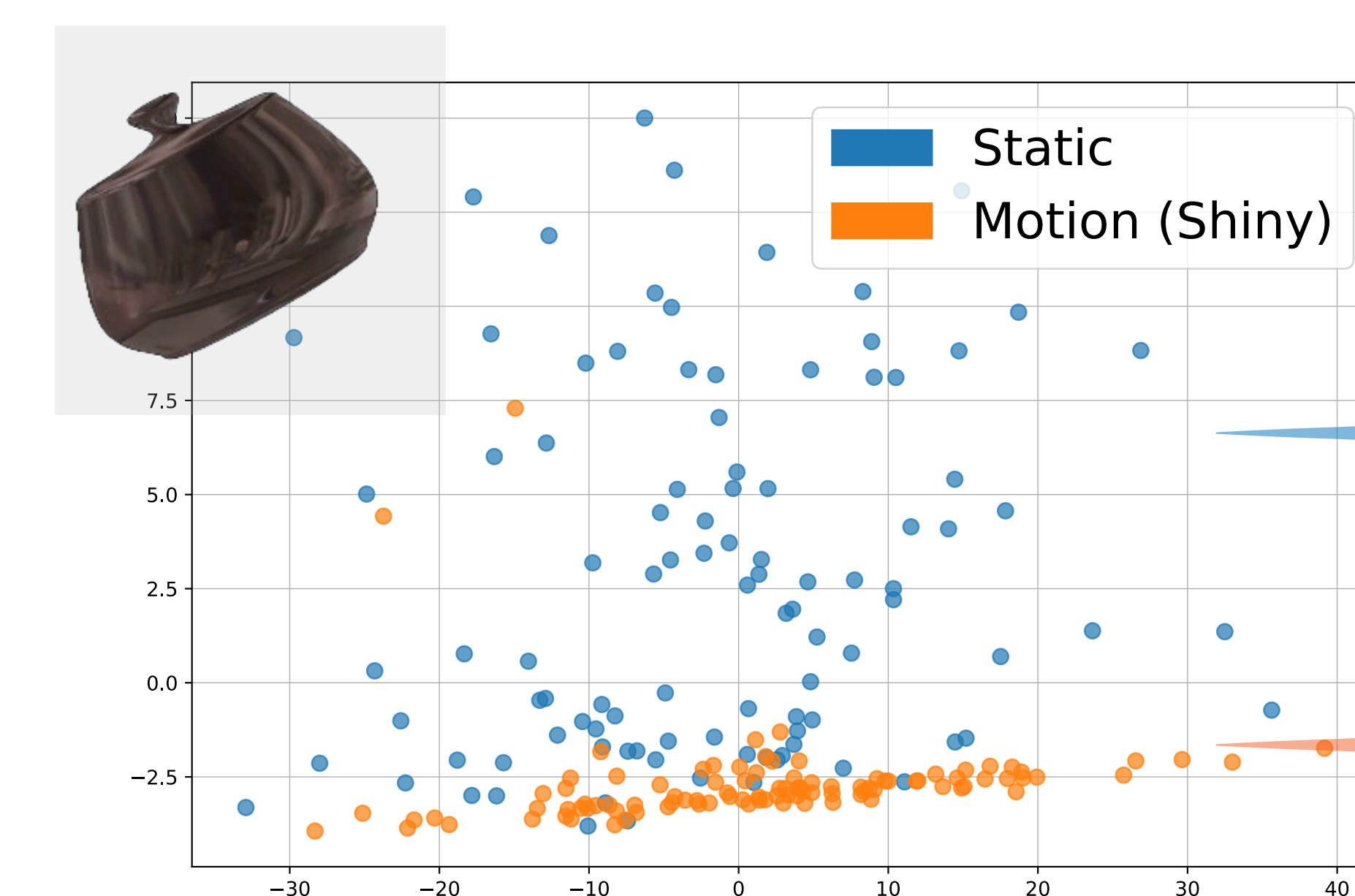
Properties: Local shift-invariant spatio-temporal attention + global 3D attention

Key Ideas

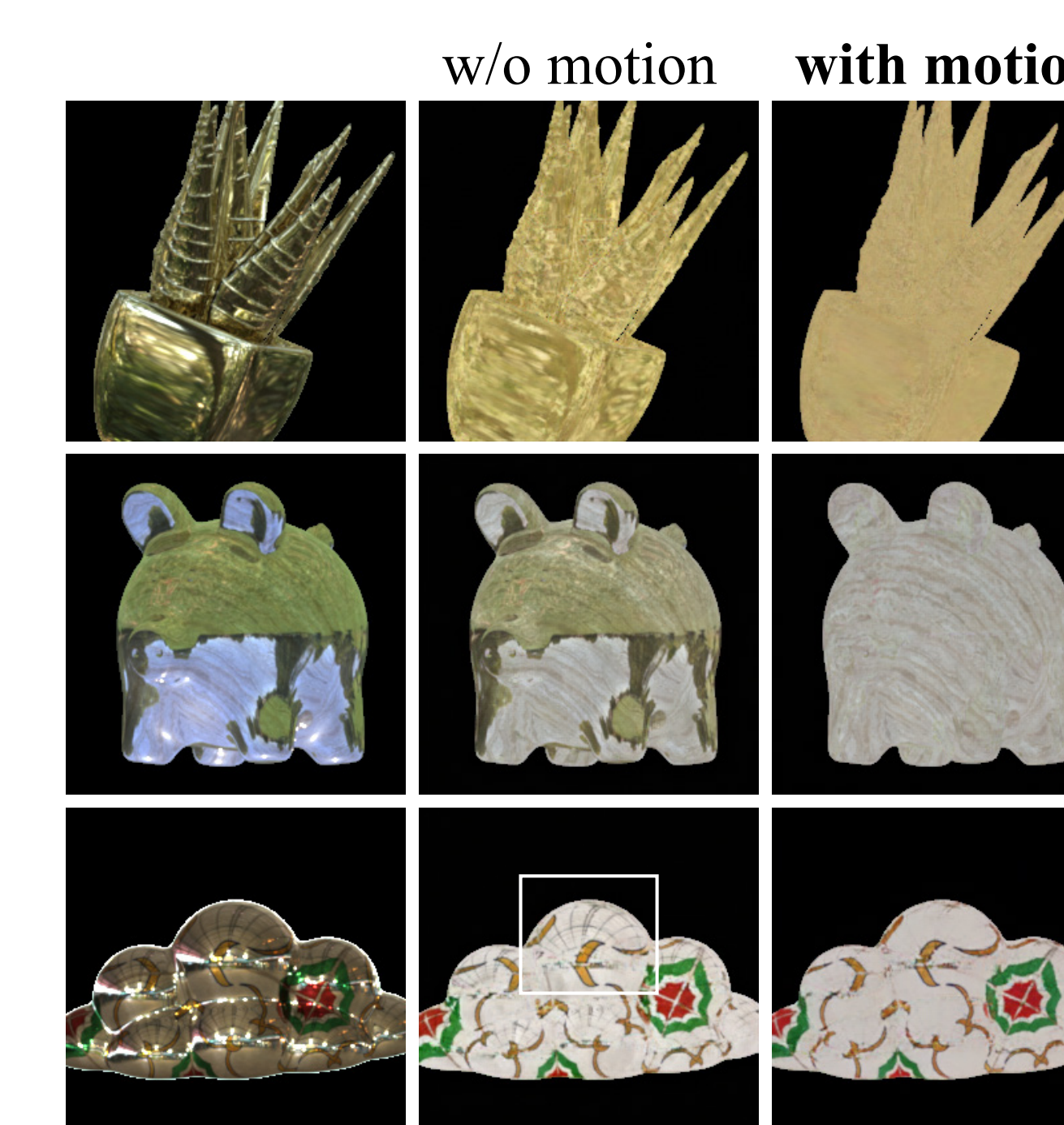
1. Joint estimation of multiple physical attributes from **static or moving** objects.
2. **Ambiguity awareness**; diverse samples when solution space is multi-modal



3. Leverage **motion cues** if they are available!



Input video from Hartung & Kersten, "Shiny or Matte?"



4. For **less ambiguous objects**, generate **accurate**, veridical shape and material predictions.

