Objective
Given a sequence of body and object motion:

- Accurately generate interacting-hand poses.

Why?
✓ Capture new datasets.
✓ Add hands to previous datasets.
✓ Refine generated/reconstructed hands.
✓ Retargeting grasp from one object to another.

Problem
Generated Motions & Datasets:
- Bodies in “isolation” without objects.
- Only hands without the body.
- Lack of accurate ground truth motions.

Limitations of Prior Work
- Not accurate hand grasping.
- Only hands without the body.
- Only refining hands, no generation.
- Slow (optimization).

Method
- Propose novel Spatio-Temporal Virtual Sensors, Ambient & Proximity Sensor.
- Novel latent temporal consistency (LTC) leads to smooth grasping motions.
- Two-stage inference for Arm and Hand.

Consistency Network (CNet)
- Generates consistent hand interactions based on object motion.
- Infers both hands separately.
- Uses LTC to ensures realistic hand-object interactions.

Latent Temporal Consistency (LTC)
- Defines consistency in the latent space for 2 successive frames.
  - A global latent code, \( z_2 \), and a relative one, \( z_{t+1} \).
  - Uses a shared decoder to further penalize the inconsistencies.

Figure 2: Latent Temporal Consistency (LTC)

Consistency Network (CNet)
\begin{align*}
X & \rightarrow D^C \\
& \rightarrow z_i \\
& \rightarrow \hat{y}_i
\end{align*}

Shared Decoder

\[ z_i \in \{t, t+1 \} \]

Spatio-Temporal Virtual Sensors
- Ambient Sensor
  - Continuous distance-based representation.
  - Obtains the object’s geometric features.
  - Spatial relation between the object & hands.

- Proximity Sensor
  - Fine-grained hand-object distance field.
  - Captures hand to object correspondence.
  - Helps with penetration and contact.

Arm Denoising Network (ANet)
- Denoises arm motions before using as input to CNet.
- Exploits the LTC with 5 frames in the latent space.

Refinement Network (RNet)
- Refines output of CNet to be more realistic.
- Preserves consistency generated by CNet.
- Recomputes proximity features to generate more accurate grasps and contact.

Figure 1: Core modules

References