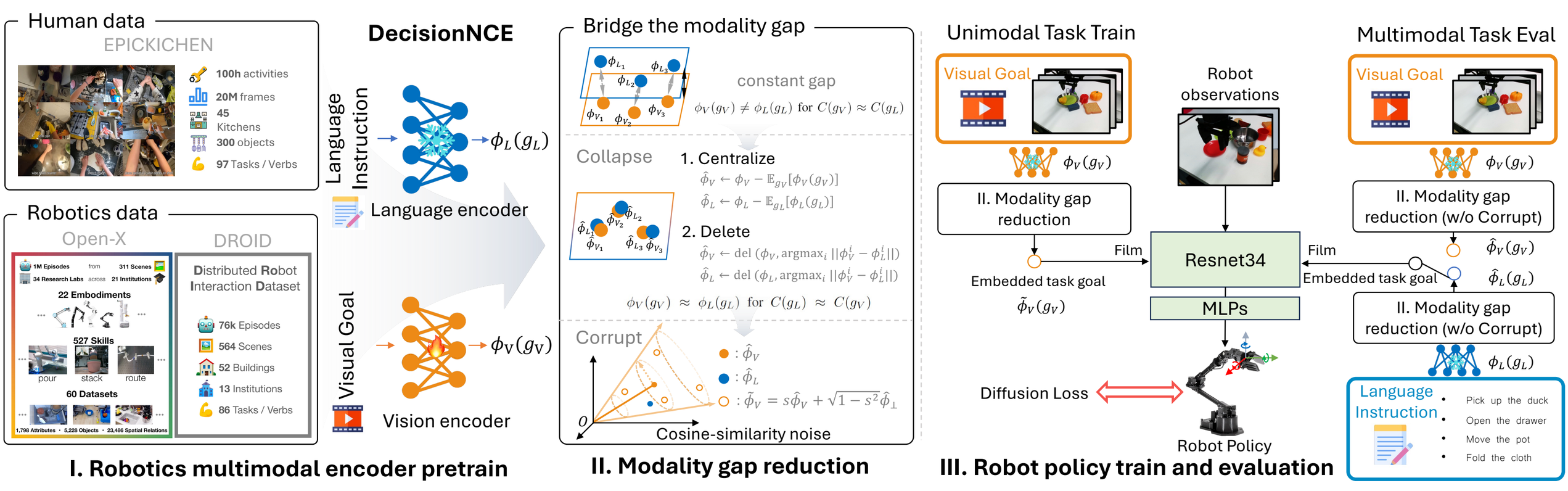
Human data
EPICKICHEN
100h activities
20M frames
45 Kitchens
300 objects
97 Tasks / Verbs
Robotics data
Open-X
1M Episodes from 311 Scenes
34 Research Labs across 21 Institutions
22 Embodiments
527 Skills
pour
stack
route
60 Datasets
1,798 Attributes · 5,228 Objects · 23,486 Spatial Relations
DROID
Distributed Robot Interaction Dataset
76k Episodes
564 Scenes
52 Buildings
13 Institutions
86 Tasks / Verbs
DecisionNCE
Language Instruction
Language encoder
Visual Goal
Vision encoder
Bridge the modality gap
constant gap
Collapse
1. Centralize
2. Delete
Corrupt
Cosine-similarity noise
Unimodal Task Train
Visual Goal
II. Modality gap reduction
Embedded task goal
Film
Resnet34
MLPs
Robot observations
Diffusion Loss
Robot Policy
Multimodal Task Eval
Visual Goal
II. Modality gap reduction (w/o Corrupt)
Film
Embedded task goal
II. Modality gap reduction (w/o Corrupt)
Language Instruction
Pick up the duck
Open the drawer
Move the pot
Fold the cloth
I. Robotics multimodal encoder pretrain
II. Modality gap reduction
III. Robot policy train and evaluation