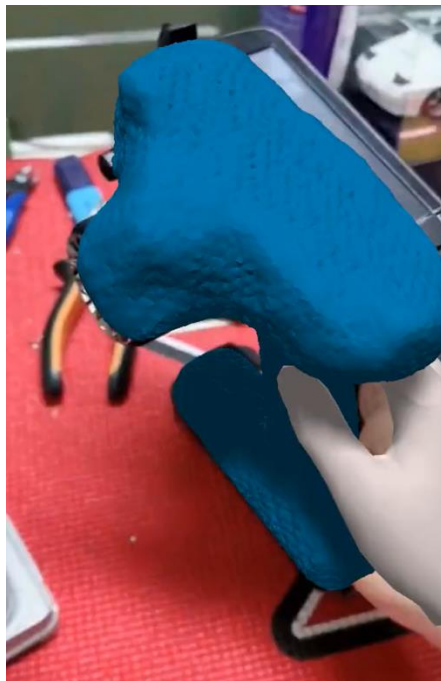
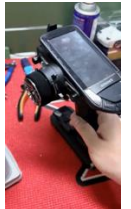
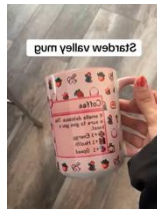
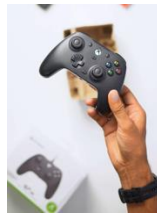
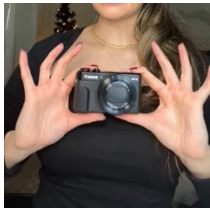
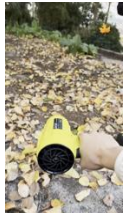
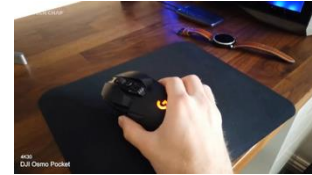


MagicHOI: Leveraging 3D Priors for Accurately Reconstruct Hand-object Interaction from Short Monocular Video Clips

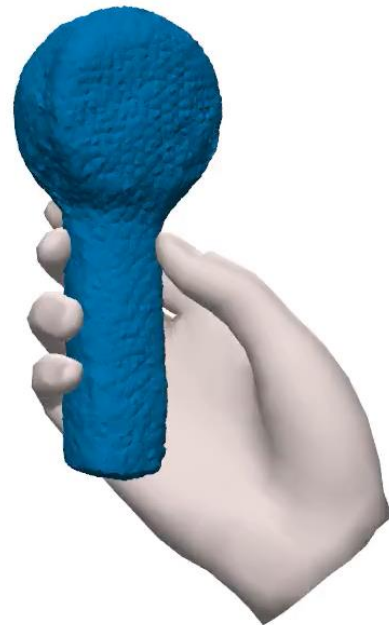


Real-world: Occluded Views



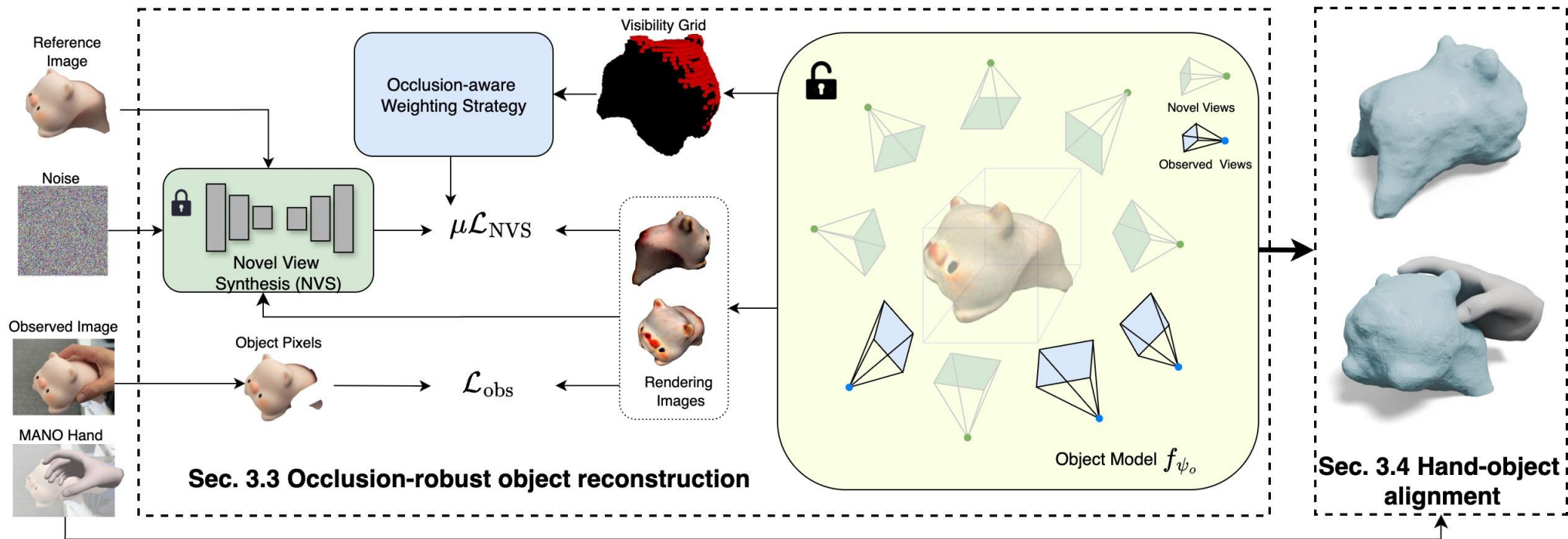
Example Videos From TikTok and YouTube

Our Reconstruction Results

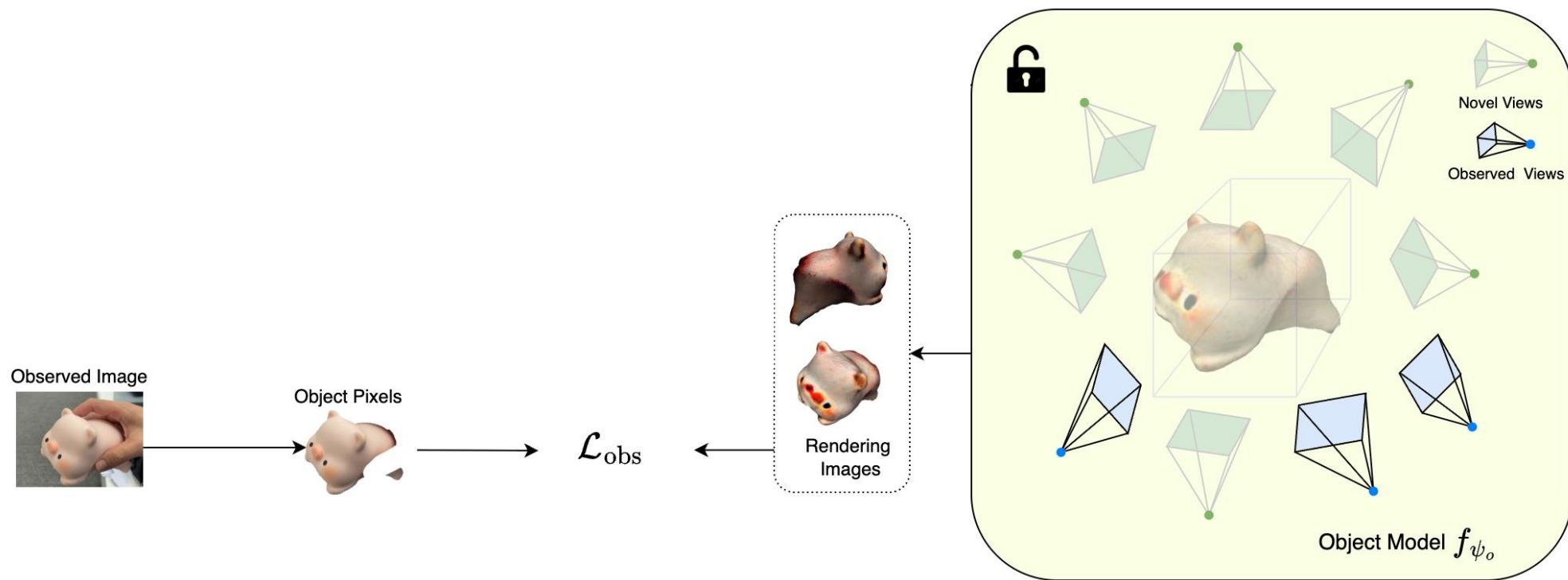


Method

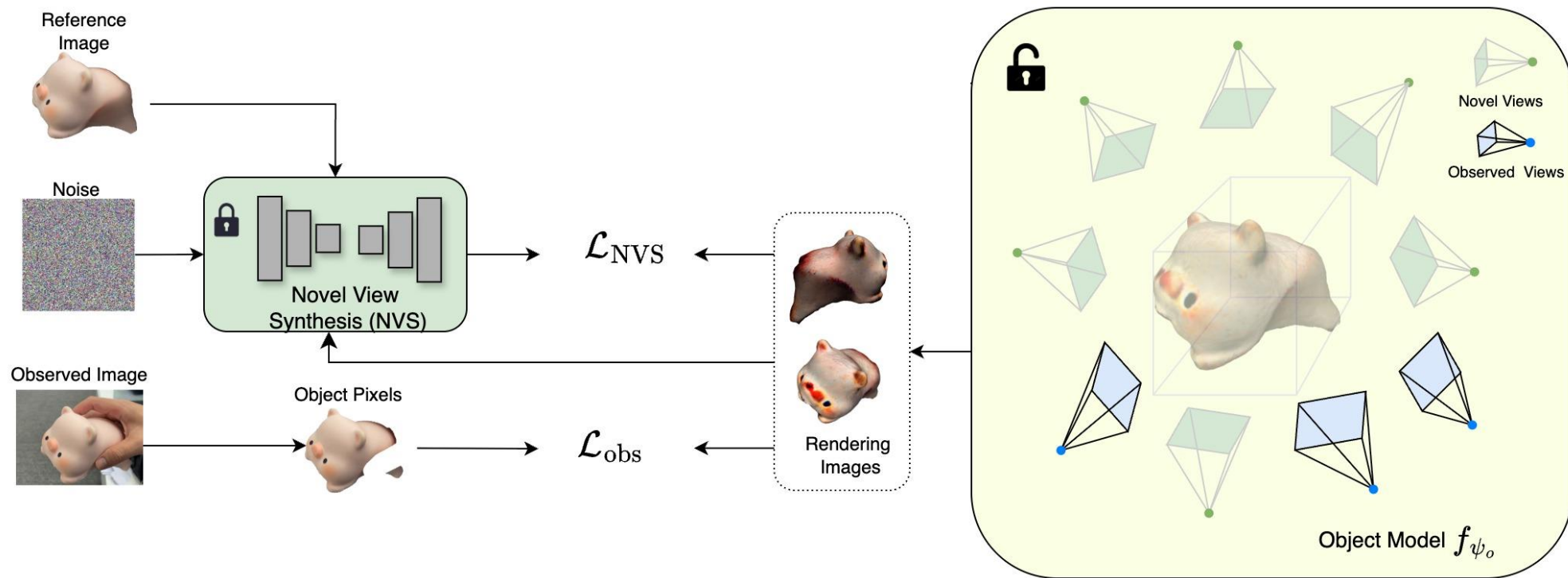
Method



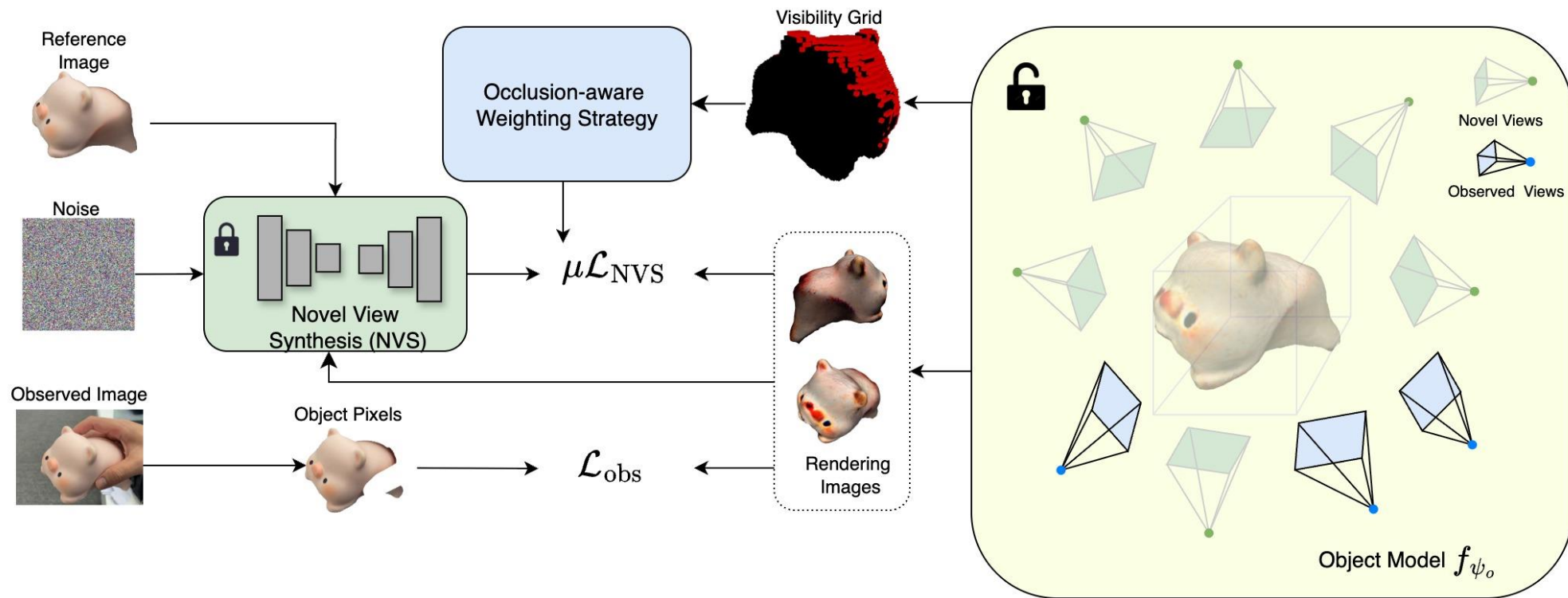
Method



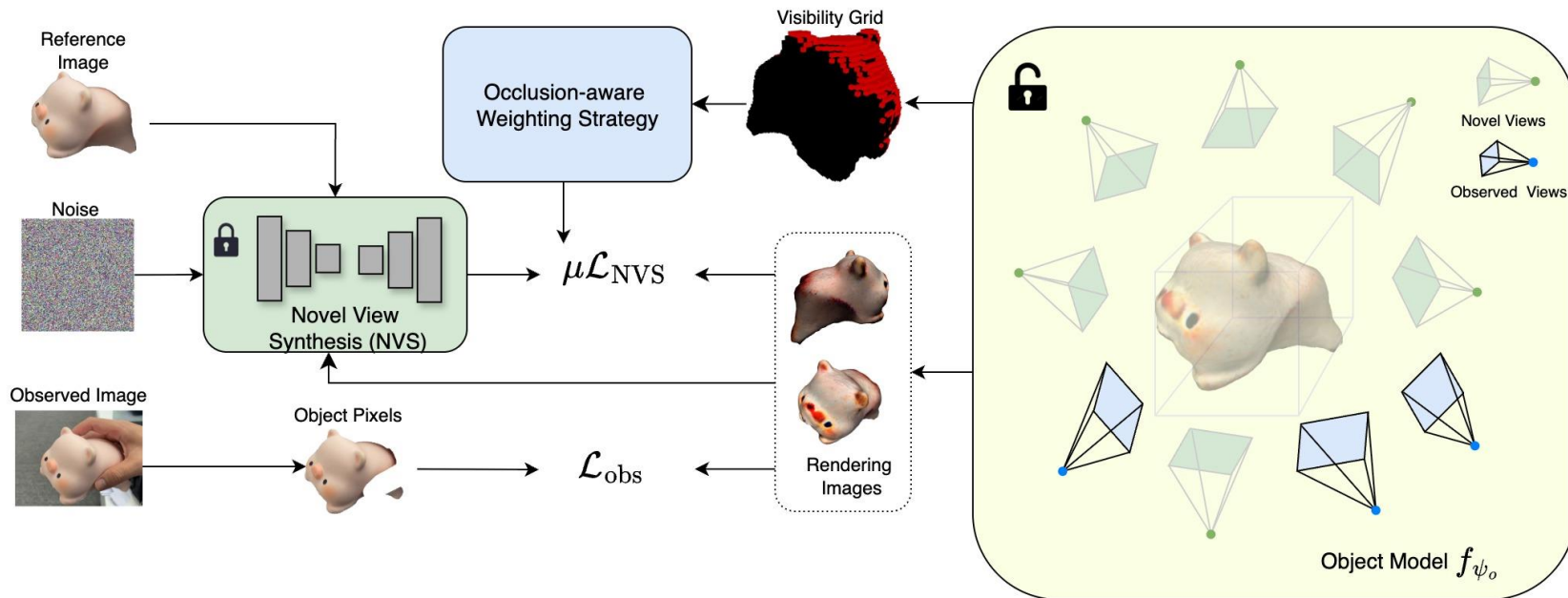
Method



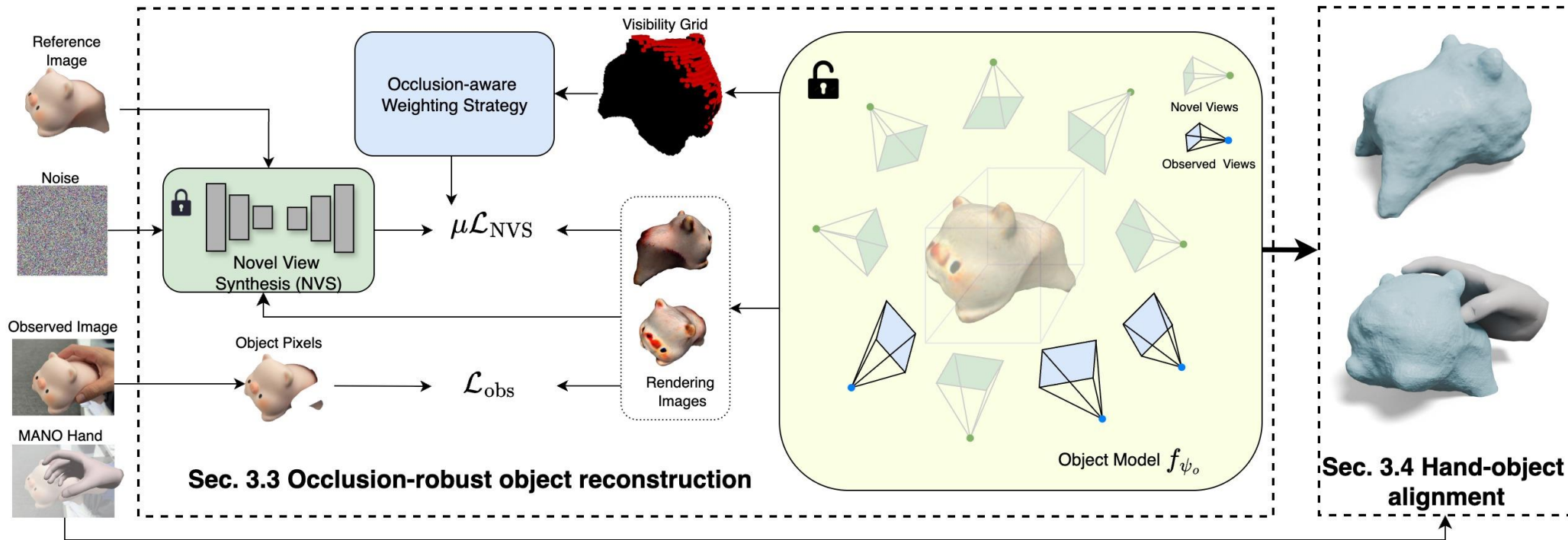
Method



Method



Method



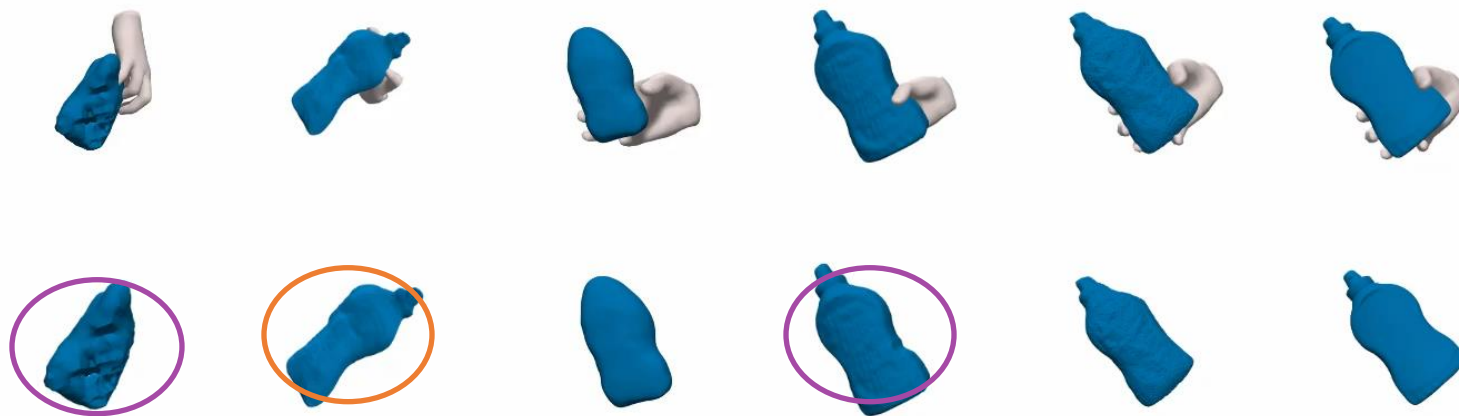
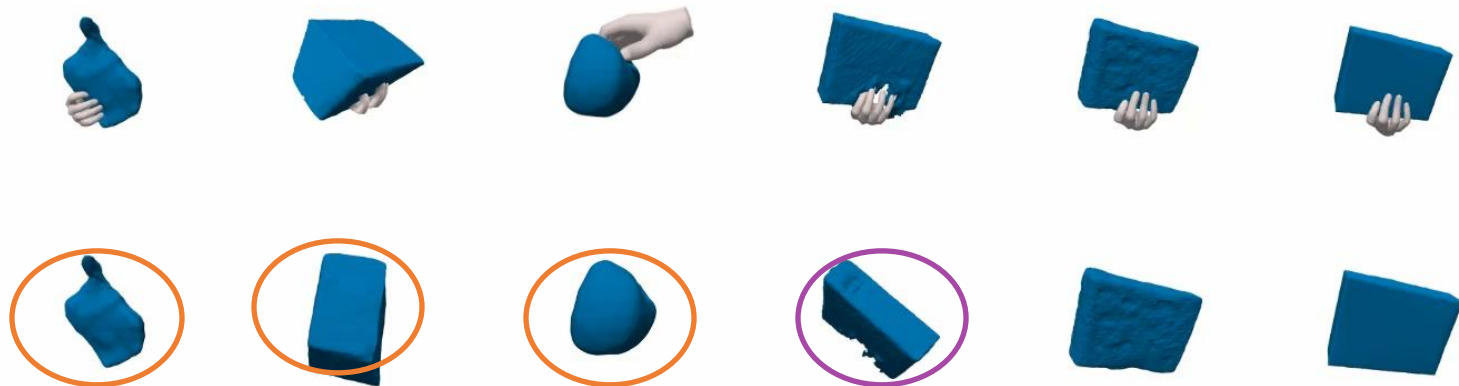
Results

Comparison with state of the art

■ Incomplete shape

■ Hallucinatory shape

Input Video



iHOI

EasyHOI

DiffHOI

HOLD

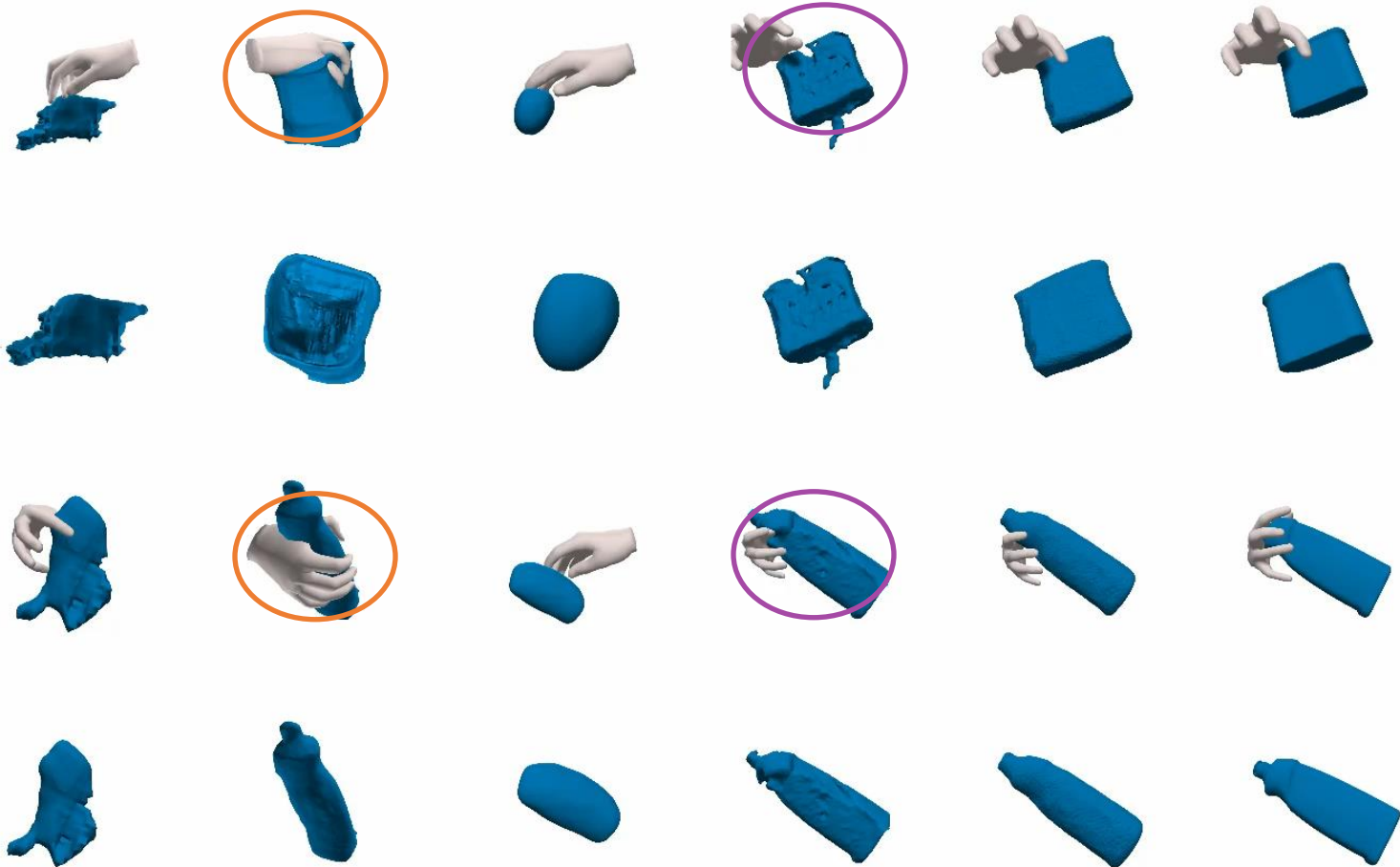
Ours

GT

Penetration

Big gap

Input Video



iHOI

EasyHOI

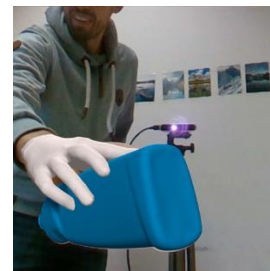
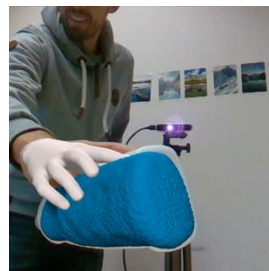
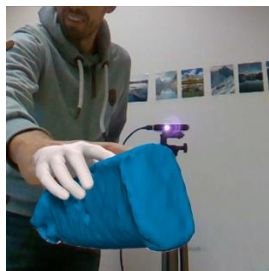
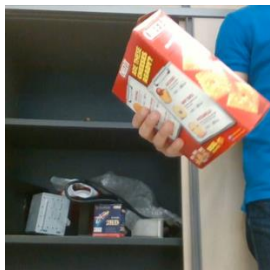
DiffHOI

HOLD

Ours

GT

Video comparison



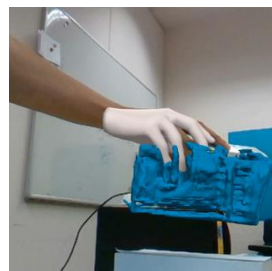
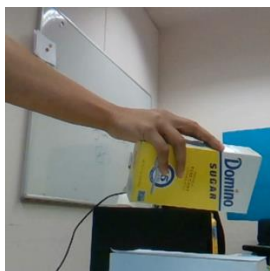
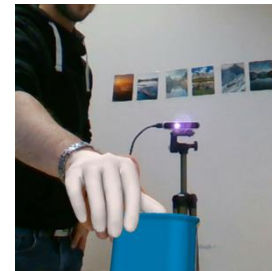
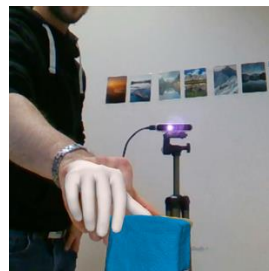
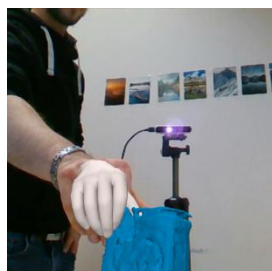
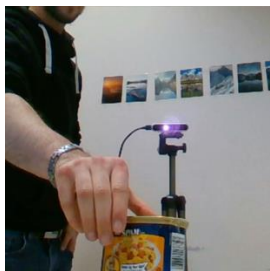
Video Input

HOLD

Ours

GT

Video comparison



Video Input

HOLD

Ours

GT

Ablation Study

Ablation Study: Weighting Strategy

WS=Weighting Strategy
OV=Observed Views
3D = 3D Prior

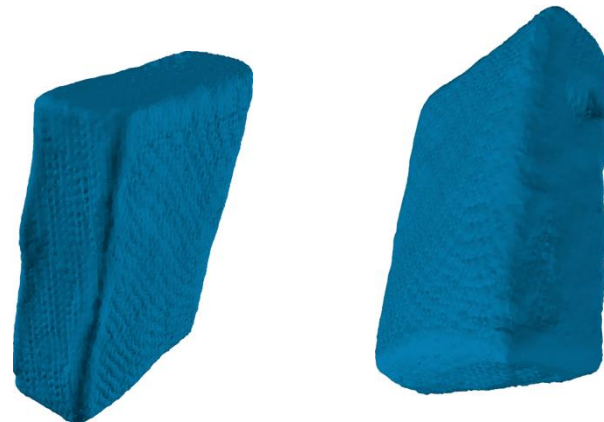
Ours (Full): WS+OV+3D



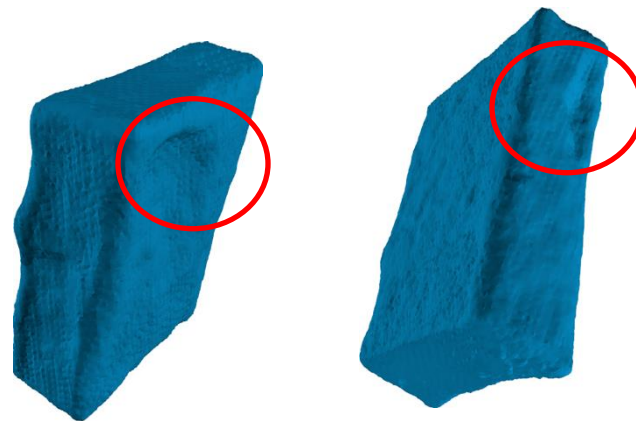
Input Video



3D Prior Input



Ours (w/o WS)



Ablation Study: Observed Views

WS=Weighting Strategy
OV=Observed Views
3D = 3D Prior

Ours (Full): WS+OV+3D



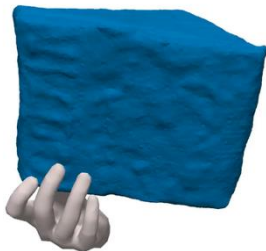
Input Video



3D Prior Input



Ours (w/o OV)



Ablation Study: 3D Prior

WS=Weighting Strategy
OV=Observed Views
3D = 3D Prior

Ours (Full): WS+OV+3D



Input Video



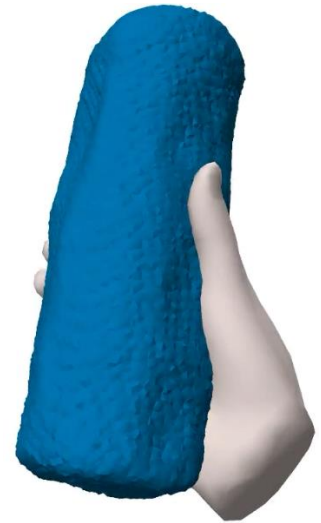
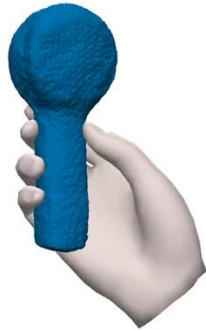
3D Prior Input



Ours (w/o 3D Prior)



Internet short-form video



Our in-the-wild short-form videos





Thanks for your listening!