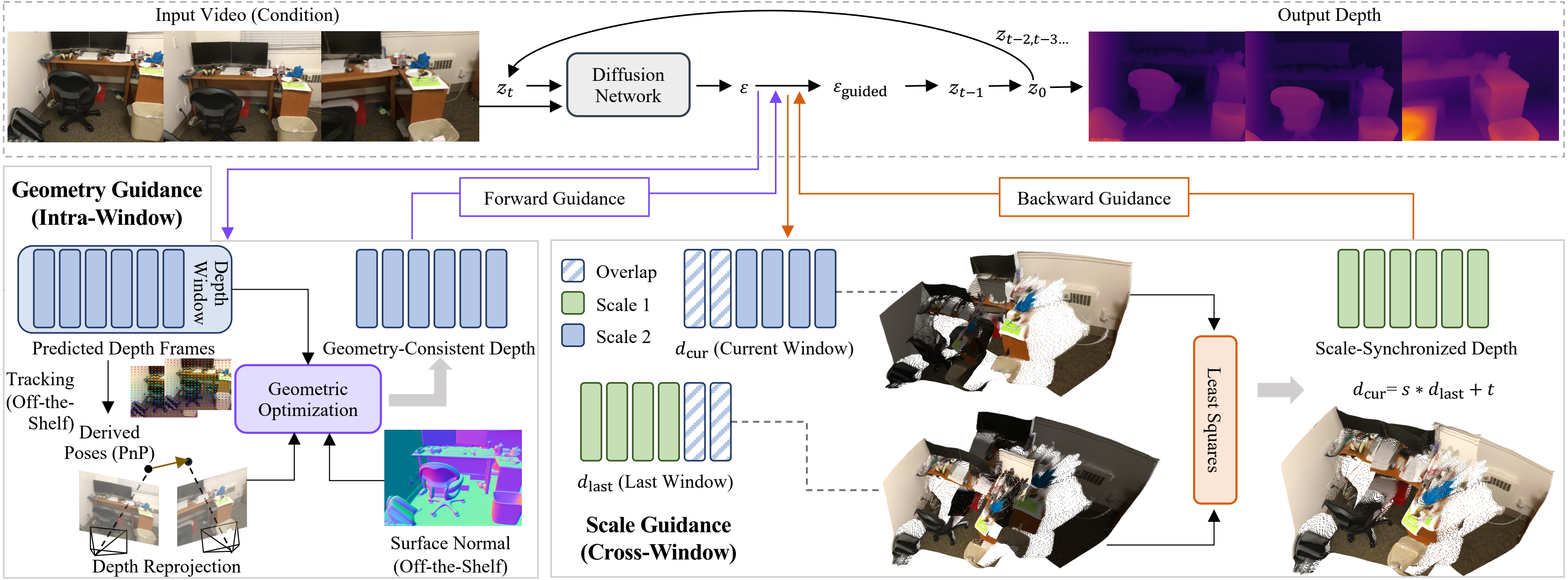
Input Video (Condition)
Output Depth
$z_t$
Diffusion Network
$\varepsilon$
$\varepsilon_{\text{guided}}$
$z_{t-1}$
$z_0$
$z_{t-2,t-3...}$
Geometry Guidance (Intra-Window)
Forward Guidance
Backward Guidance
Depth Window
Predicted Depth Frames
Geometry-Consistent Depth
Tracking (Off-the-Shelf)
Derived Poses (PnP)
Geometric Optimization
Depth Reprojection
Surface Normal (Off-the-Shelf)
▨ Overlap
🟩 Scale 1
▦ Scale 2
$d_{\text{cur}}$ (Current Window)
$d_{\text{last}}$ (Last Window)
Least Squares
Scale-Synchronized Depth
$d_{\text{cur}} = s * d_{\text{last}} + t$
Scale Guidance (Cross-Window)