Why Synthetic Teacher?

- Real-world depth (LiDAR/structured light) is often **sparse, noisy, and incomplete** (see Fig. 4).
- Synthetic data provides **dense, clean geometry** with perfect ground truth.

HyperSim | Virtual KITTI | TartanAir | ScanNet++ (Noisy Target)

The Teacher Model (DA3-Teacher)

- **Architecture:** Monocular DINOv2 + DPT decoder (same backbone class).
- **Target:** Scale-shift-invariant *exponential* depth (better for near-field).
- **Losses:** Gradient + Global-Local (ROE) + Surface Normal + Sky/Obj Masks.

Robust Alignment Strategy

The teacher's relative depth $\tilde{D}$ is aligned to noisy sparse real measurements D_p via robust RANSAC least squares.

$$(s,t) = \text{argmin}_{s>0,t} \sum_{p\in\Omega} m_p(s\tilde{D}_p + t - D_p)^2$$

$$D_{aligned} = \hat{s}\tilde{D} + \hat{t}$$

RANSAC Benefit

Filters out gross outliers in real sensor data using Median Absolute Deviation (MAD) thresholding, preventing teacher degradation.

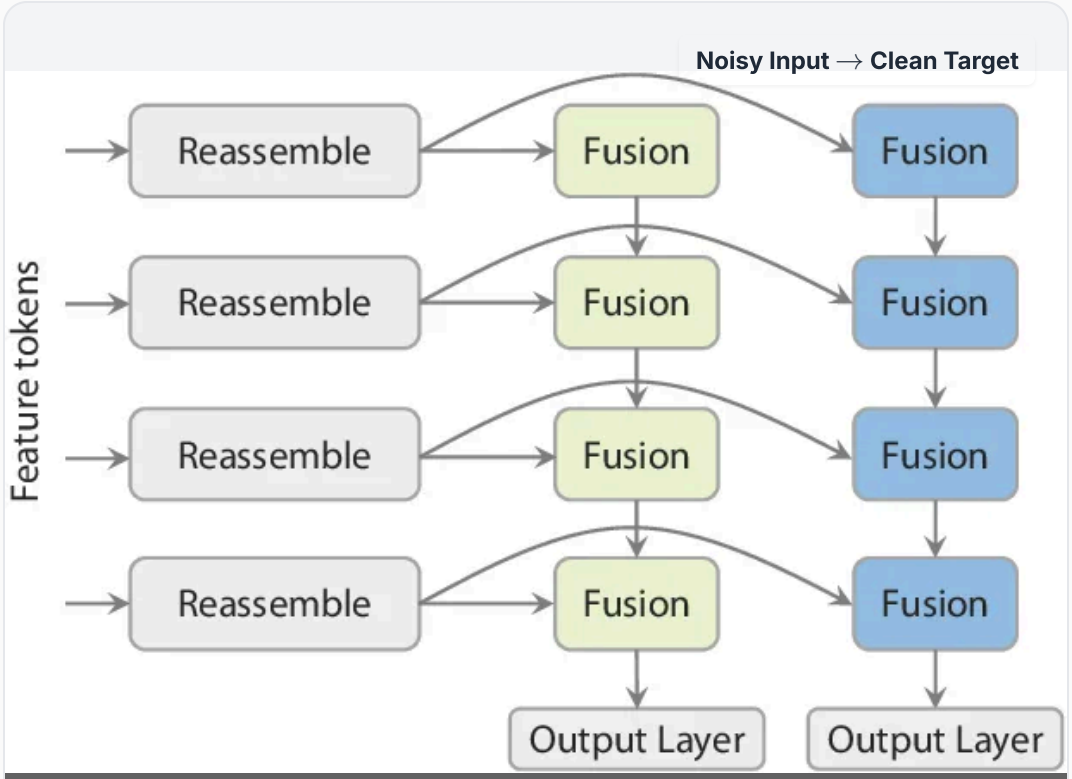

Fig 4: Data Quality & Alignment Sparse Real vs. Dense Pseudo-Label