Supplementary Materials

A. Supplementary experiments on nuScenes

We conduct comprehensive supplementary experiments to further validate our design choices. First, we investigate the impact of different feature aggregation architectures: comparing MLPs with deformable attention layers (defo.), with results summarized in Table A.1. We observe that as long as the key coupling layer is present, performance remains stable, and architectures equipped with deformable attention achieve even stronger results. In contrast, removing the coupling layer leads to a substantial performance drop. These findings further confirm that our invertible transformations provide reliable alignment of feature distributions across heterogeneous tasks while effectively mitigating dimensionality collapse.

We also evaluate two geographically partitioned supervision settings. The Standard Geo-Split (S-Geo) setup follows our main protocol, where mapping supervision is available only in **Boston** and tracking supervision only in **Singapore**. We additionally introduce a Reversed Geo-Split (R-Geo) configuration that swaps the supervision visibility across regions. This complementary variant allows us to assess the consistency of our method under geographically perturbed supervision, with results reported in Table A.2.

Table A.1. **Supplementary experiments on nuScenes (Architecture).** We evaluate different settings on nuScenes dataset with different feature aggregation architectures.

Method	Multi-object Tracking				Online Mapping (IoU %)			
	AMOTA↑	AMOTP↓	Recall↑	IDS↓	Lanes↑	Drivable↑	Divider↑	Crossing↑
Baseline	0.289	1.488	0.362	1025	27.1	62.7	22.6	14.1
Ours (MLP w/o Inv)	0.214	1.507	1.355	731	32.3	56.8	27.7	21.9
Ours (MLP + Inv)	0.316	1.250	0.397	879	36.2	62.31	28.80	20.75
Ours (defo. + Inv)	**0.329**	**1.322**	**0.428**	**690**	**37.1**	**64.5**	**30.0**	**22.8**

Table A.2. **Supplementary experiments on nuScenes (Geo-Split).** We evaluate two geographically partitioned supervision settings on nuScenes dataset.

Method	Multi-object Tracking				Online Mapping (IoU %)			
	AMOTA↑	AMOTP↓	Recall↑	IDS↓	Lanes↑	Drivable↑	Divider↑	Crossing↑
Baseline	0.289	1.488	0.362	1025	27.1	62.7	22.6	14.1
Ours (R-Geo)	**0.332**	**1.304**	**0.435**	**653**	36.6	63.8	29.7	22.5
Ours (S-Geo)	0.329	1.322	0.428	690	**37.1**	**64.5**	**30.0**	**22.8**

B. Supplementary experiments on NYU-V2

We conduct additional supplementary experiments to validate our design choices on dense prediction tasks. We first examine the impact of different feature aggregation architectures by comparing MLP-based aggregators under the same invertible-layer configuration (3 layers), as shown in Table B.1. We then evaluate the complementary setting in which the feature aggregator is fixed (MLP, 3 layers) while varying the number of invertible layers, with results reported in Table B.2.

Table B.1. **NYU-V2 Results Part 1.** We evaluate settings on NYU-V2 dataset with different MLP layers in feature aggregator.

Method	Seg. (IoU) ↑	Depth (aErr) ↓	Norm. (mErr) ↓
Baseline	26.75	0.6511	35.17
1 MLP	28.57	0.6841	34.54
2 MLP	29.46	0.6311	33.42
3 MLP	**31.70**	**0.6055**	**31.88**
4 MLP	29.50	0.6404	33.39
5 MLP	29.78	0.6268	33.12
6 MLP	29.85	0.6258	33.05

Table B.2. **NYU-V2 Results Part 2.** We evaluate settings on NYU-V2 dataset with different Invertible layers.

Method	Seg. (IoU) ↑	Depth (aErr) ↓	Norm. (mErr) ↓
Baseline	26.75	0.6511	35.17
1 Inv	28.20	0.6300	33.50
2 Inv	29.50	0.6200	33.10
3 Inv	29.90	0.6250	33.00
4 Inv	**31.70**	**0.6055**	**31.88**
5 Inv	29.80	0.6280	33.20
6 Inv	27.10	0.6400	34.00