

# EarthBridge: A Solution for 4th Multi-modal Aerial View Image Challenge

## Translation Track

### Supplementary Material

#### A. Competition test set: NFE selection and SAR→IR model choice

Tab. 2 in the main paper reports per-task NFE that may appear inconsistent with “few-step” DBIM sampling (e.g., 500–1000 for some tracks vs. 5 for RGB→IR). DBIM is *capable* of very low-NFE sampling everywhere; the discrepancy reflects a *task-dependent* choice of step count under the challenge’s composite metric, not a limitation of the sampler. Tab. S1 reports **official MAVIC-T competition scores on the challenge test set**—the same evaluation pipeline and composite metric as the leaderboard (FID uses normalized atan), with each row corresponding to a distinct evaluated configuration submitted to the server. **Light-blue rows** indicate the best composite score within each task block.

Table S1. MAVIC-T **test-set** competition scores (all tasks). **Light-blue rows** mark the best composite score per task block (full-row fill). ↓ lower is better. DDBM uses a numerical bridge sampler with the listed steps. CUT rows list generator capacity in the Method column and use NFE=1 (single feed-forward pass). *Nano Banana Pro* is an external image-generation baseline; we omit NFE (—) when the system does not expose an equivalent step count.

Task	Method	NFE	FID↓	LPIPS↓	L1↓	Score↓
<i>RGB→IR</i> (NFE sweep)						
RGB→IR	DBIM	1	0.786	0.293	0.093	0.390
RGB→IR	DBIM	2	0.342	0.195	0.091	0.209
RGB→IR	DBIM	5	0.357	0.151	0.090	<b>0.200</b>
RGB→IR	DBIM	10	0.380	0.150	0.090	0.206
RGB→IR	DBIM	20	0.601	0.468	0.076	0.382
RGB→IR	DBIM	100	0.478	0.188	0.108	0.258
RGB→IR	DBIM	1000	0.474	0.225	0.110	0.270
RGB→IR	DDBM	100	0.709	0.366	0.154	0.410
<i>SAR→EO</i> (NFE sweep)						
SAR→EO	DBIM	20	0.737	0.464	0.074	0.425
SAR→EO	DBIM	100	0.451	0.477	0.077	0.335
SAR→EO	DBIM	500	0.222	0.504	0.080	<b>0.269</b>
SAR→EO	DBIM	1000	0.233	0.511	0.079	0.274
<i>SAR→IR</i> (CUT vs. diffusion bridges)						
SAR→IR	CUT (huge)	1	0.645	0.603	0.145	<b>0.464</b>
SAR→IR	CUT (large)	1	0.825	0.610	0.157	0.531
SAR→IR	DDBM	100	0.874	0.616	0.244	0.578
SAR→IR	DBIM	100	0.956	0.706	0.256	0.640
SAR→IR	DBIM	1000	0.958	0.630	0.194	0.594
<i>SAR→RGB</i> (NFE vs. baselines)						
SAR→RGB	DBIM	1000	0.878	0.636	0.214	<b>0.576</b>
SAR→RGB	DDBM	100	0.894	0.656	0.236	0.595
SAR→RGB	Nano Banana Pro	—	0.957	0.740	0.208	0.635

For RGB→IR, the composite score is best at  $N=5$  steps and degrades for both very small and very large  $N$  (Tab. S1, RGB→IR block); this matches the  $N=5$  entry in Tab. 2. For SAR→EO, scores improve sharply between  $N=100$  and  $N=500$  and are near-optimal at 500–1000 steps (Tab. S1, SAR→EO block), which motivates the high-NFE settings used for that track despite DBIM’s support for fewer steps. For SAR→RGB, DBIM at  $N=1000$  outperforms DDBM at  $N=100$  and the Nano Banana Pro image-generation baseline on the **test set** (Tab. S1, SAR→RGB block). For SAR→IR, CUT clearly beats DBIM/DDBM on task score on the **test set** (Tab. S1, SAR→IR block); we therefore submit CUT for that track.

**Relation to the main paper.** Tab. S1 reports official MAVIC-T **test-set** competition scores under alternate step counts and model choices; it complements Tab. 2 by showing how per-task NFE was selected under the composite metric and how CUT

compares to diffusion-bridge baselines for SAR→IR. Sec. 7 states training-resolution and channel-handling limitations of our submission.