

Domain-Specific Performance Breakdown (Our Full Method)

Domain	Top-1	Top-5	ECE	Avg Conf	Time (ms)	Samples
Natural Images (ImageNet)	74.1 ± 0.3	91.2 ± 0.2	0.047	0.82	156	1000
Medical Images (ChestX-ray)	52.7 ± 0.5	78.9 ± 0.4	0.092	0.68	163	1000
Satellite Images (EuroSAT)	63.4 ± 0.4	85.3 ± 0.3	0.063	0.75	151	1000
Anime (Danbooru)	69.8 ± 0.4	88.5 ± 0.3	0.055	0.78	148	800
Sketches (Sketchy)	63.1 ± 0.5	83.7 ± 0.4	0.071	0.72	145	800
Average	64.6 ± 0.4	85.5 ± 0.3	0.066	0.75	153	920

ECE: Expected Calibration Error (lower is better) | Avg Conf: Average prediction confidence | Time: Average inference time per image