

A. Prompt card and decoding details

Prompt template We use a single RGB multiple-choice prompt for all models and experiments. The template below is instantiated with the BigEarthNet-S2 class list and a single input image.

Decoding and label extraction At inference time the model returns a free-form text response. We convert this response into a set of predicted labels as follows:

1. Run a case-insensitive regular expression that extracts all substrings of the form $(\$k)$ where $\$k$ is an integer between 1 and C .
2. Map each extracted index k to the corresponding BigEarthNet-S2 class.
3. Discard duplicates and out-of-range indices.

The same parsing rule is applied to all models (zero-shot Qwen3-VL, Gemini 2.5-Flash, and fine-tuned Qwen3-VL), and no additional thresholding or calibration is used. If no valid indices are extracted, the prediction is treated as the empty set.

B. Extended quantitative results

Full metrics for zero-shot and SFT models Table 4 extends the main-text results by reporting micro-averaged precision, recall, and F1 together with macro-F1 and subset accuracy on the 1k-image class-balanced BigEarthNet-S2 test set for all backbones in both zero-shot and SFT settings.

C. Per-class behaviour and qualitative analysis

Table 5 reports per-class precision, recall and F_1 on the BigEarthNet-S2 test set before and after class-balanced fine-tuning of Qwen3-VL-4B with 4-bit LoRA ($r=128$). The baseline model achieves a macro-averaged F_1 of 0.26 and a weighted F_1 of 0.30. After fine-tuning on the class-balanced subset, macro- F_1 increases to 0.61 and weighted F_1 to 0.65, indicating that the gains are not limited to a few dominant classes.

The largest F_1 improvements are observed for under-represented vegetation categories. For example, *Agroforestry areas* improves from 0.00 to 0.70, *Coniferous forest* from 0.03 to 0.69, and *Mixed forest* from 0.05 to 0.68. *Transitional woodland, shrub*, *Natural grassland and sparsely vegetated areas*, and *Permanent crops* also benefit strongly from the balanced training scheme (all with $\Delta F_1 \geq 0.37$). Classes that were already comparatively easy for the baseline, such as *Arable land*, *Marine waters* and *Urban fabric*, show smaller but still positive gains (e.g., $0.70 \rightarrow 0.79$, $0.55 \rightarrow 0.74$, and $0.58 \rightarrow 0.78$). Notably, all classes improve after fine-tuning; no category experiences a drop in F_1 .

Figure 6 shows qualitative examples of images where the class-balanced fine-tuned model corrects errors of the

baseline. For each patch, we display the ground-truth labels together with baseline and fine-tuned predictions, as well as the per-example multi-label F_1 before and after fine-tuning. The selected examples cover a diverse set of land-cover types (forests, agricultural land, grassland, water and wetlands) and consistently show large F_1 improvements (up to $\Delta F_1 = 1.0$). In many cases, the baseline either misses all relevant labels or predicts overly generic categories, whereas the fine-tuned model recovers the full set of ground-truth labels, illustrating the benefit of class-balanced training beyond the aggregate metrics in Table 5.

D. Implementation details

LoRA target modules For all Qwen3-VL backbones, we attach LoRA adapters to:

- All self-attention projections (q_proj , k_proj , v_proj , o_proj) in the language module.
- All MLP projections in the language module.
- The corresponding attention and MLP projections in the vision module.

We verified that all of these modules are marked as trainable during SFT. As shown in Section 5.2, however, the learned LoRA energies concentrate almost entirely in the language-side MLP and attention blocks, with negligible energy in the vision modules.

Instructions: Answer the question asked after the given image. Output format: Output the option numbers corresponding to the correct answer in the format “(X)” where X is the correct digit choice (among 1 to 19). In case of multiple correct answers, output the answer choices in the same format separated by commas. For example, if the correct answers are choices 1 and 3, output “(1),(3)”. Only output the option numbers in the above format and DO NOT OUTPUT the full answer text or any other extra words.

Question: To which of the following class does the given image belong to? Select all that apply.

Possible answer choices:

(1) Agro-forestry areas (2) Arable land (3) Beaches, dunes, sands (4) Broad-leaved forest (5) Coastal wetlands (6) Complex cultivation patterns (7) Coniferous forest (8) Industrial or commercial units (9) Inland waters (10) Inland wetlands (11) Land principally occupied by agriculture, with significant areas of natural vegetation (12) Marine waters (13) Mixed forest (14) Moors, heathland and sclerophyllous vegetation (15) Natural grassland and sparsely vegetated areas (16) Pastures (17) Permanent crops (18) Transitional woodland, shrub (19) Urban fabric

Answer:

Figure 4. Prompt template used for all zero-shot and fine-tuned models with the 19 BigEarthNet-S2 land-cover classes.

Table 4. Full evaluation metrics on the 1k-image class-balanced BigEarthNet-S2 test set. We report micro precision (P), recall (R), and F1, along with macro-F1 and subset accuracy. SFT rows correspond to the LoRA configurations highlighted in the main text (for Qwen3-VL-4B we show the full rank sweep).

Model	Setting	Micro-P	Micro-R	Micro-F1	Macro-F1	Subset Acc.
Qwen3-VL-2B	zero-shot	0.350	0.273	0.306	0.166	0.001
Qwen3-VL-2B	SFT (r=128)	0.623	0.655	0.639	0.598	0.135
Qwen3-VL-4B	zero-shot	0.450	0.310	0.367	0.253	0.014
Qwen3-VL-4B	SFT (r=4)	0.574	0.595	0.584	0.503	0.107
Qwen3-VL-4B	SFT (r=16)	0.597	0.622	0.609	0.557	0.116
Qwen3-VL-4B	SFT (r=64)	0.626	0.655	0.640	0.594	0.126
Qwen3-VL-4B	SFT (r=128)	0.636	0.667	0.651	0.612	0.133
Qwen3-VL-4B	SFT (r=256)	0.628	0.650	0.639	0.599	0.129
Qwen3-VL-8B	zero-shot	0.430	0.163	0.237	0.176	0.009
Qwen3-VL-8B	SFT (r=128)	0.603	0.636	0.619	0.569	0.115
Qwen3-VL-32B	zero-shot	0.480	0.267	0.344	0.264	0.045
Qwen3-VL-32B	SFT (r=128)	0.608	0.618	0.613	0.582	0.132
Gemini 2.5-Flash	zero-shot	0.591	0.240	0.342	0.270	0.038
Gemini 2.5-Flash	thinking 1024	0.511	0.393	0.444	0.380	0.017
Gemini 2.5-Flash	thinking 4096	0.525	0.389	0.446	0.372	0.024

Table 5. Per-class precision, recall and F_1 on the BigEarthNet-S2 test set before (Base) and after class-balanced fine-tuning (FT) of Qwen3-VL-4B with 4-bit LoRA ($r=128$). ΔF_1 denotes FT-Base for each class.

class_id	class_name	P_base	R_base	F1_base	P_ft	R_ft	F1_ft	Δ_{F1}	support
1	Agro-forestry areas	0.00	0.00	0.00	0.64	0.78	0.70	0.70	128
2	Arable land	0.69	0.71	0.70	0.76	0.83	0.79	0.09	457
3	Beaches, dunes, sands	0.20	0.09	0.13	0.75	0.28	0.41	0.28	54
4	Broad-leaved forest	0.49	0.17	0.26	0.59	0.71	0.65	0.39	343
5	Coastal wetlands	0.21	0.12	0.15	0.49	0.46	0.48	0.33	52
6	Complex cultivation patterns	0.38	0.62	0.47	0.59	0.64	0.61	0.14	273
7	Coniferous forest	0.44	0.02	0.03	0.68	0.70	0.69	0.66	248
8	Industrial or commercial units	0.49	0.39	0.44	0.75	0.57	0.64	0.21	69
9	Inland waters	0.64	0.43	0.51	0.87	0.57	0.69	0.17	222
10	Inland wetlands	0.16	0.19	0.17	0.45	0.46	0.45	0.28	81
11	Land principally occupied by agriculture, with significant areas of natural vegetation	0.35	0.82	0.49	0.58	0.61	0.59	0.10	317
12	Marine waters	0.55	0.54	0.55	0.66	0.86	0.74	0.20	83
13	Mixed forest	0.50	0.03	0.05	0.68	0.68	0.68	0.63	337
14	Moors, heathland and sclerophyllous vegetation	0.24	0.06	0.10	0.51	0.53	0.52	0.42	81
15	Natural grassland and sparsely vegetated areas	0.01	0.02	0.01	0.57	0.63	0.60	0.59	63
16	Pastures	0.44	0.11	0.17	0.54	0.62	0.57	0.40	229
17	Permanent crops	0.06	0.01	0.02	0.35	0.43	0.39	0.37	113
18	Transitional woodland, shrub	0.73	0.03	0.06	0.64	0.66	0.65	0.59	355
19	Urban fabric	0.76	0.47	0.58	0.75	0.81	0.78	0.20	180

Table 6. Effect of layer pruning on multi-label classification performance. LLM-All retains all layers, MLP-only keeps only MLP layers, and Attn-only keeps only attention layers.

Group	Keep Fraction	Micro			Macro			Subset Acc.
		P	R	F1	P	R	F1	
LLM-All	1.00	0.577	0.583	0.580	0.552	0.515	0.505	0.112
	0.75	0.586	0.580	0.583	0.558	0.510	0.510	0.104
	0.50	0.575	0.550	0.562	0.534	0.469	0.479	0.090
	0.20	0.512	0.437	0.472	0.501	0.373	0.387	0.057
	0.10	0.472	0.388	0.426	0.438	0.336	0.337	0.039
	0.05	0.453	0.343	0.390	0.392	0.290	0.289	0.028
	0.02	0.457	0.327	0.381	0.402	0.272	0.274	0.023
MLP-only	1.00	0.576	0.527	0.550	0.545	0.463	0.478	0.089
	0.75	0.570	0.496	0.530	0.544	0.442	0.462	0.073
	0.50	0.547	0.455	0.497	0.521	0.399	0.419	0.059
	0.20	0.477	0.385	0.426	0.447	0.337	0.344	0.047
	0.10	0.465	0.341	0.394	0.421	0.290	0.296	0.036
	0.05	0.453	0.325	0.379	0.404	0.275	0.278	0.028
	0.02	0.459	0.321	0.378	0.406	0.265	0.270	0.021
Attn-only	1.00	0.505	0.455	0.479	0.471	0.376	0.381	0.058
	0.75	0.491	0.430	0.458	0.463	0.360	0.363	0.054
	0.50	0.480	0.402	0.437	0.450	0.338	0.338	0.052
	0.20	0.453	0.360	0.401	0.394	0.307	0.298	0.032
	0.10	0.448	0.334	0.383	0.388	0.280	0.277	0.022
	0.05	0.451	0.326	0.379	0.387	0.269	0.269	0.020
	0.02	0.454	0.318	0.374	0.394	0.260	0.263	0.018

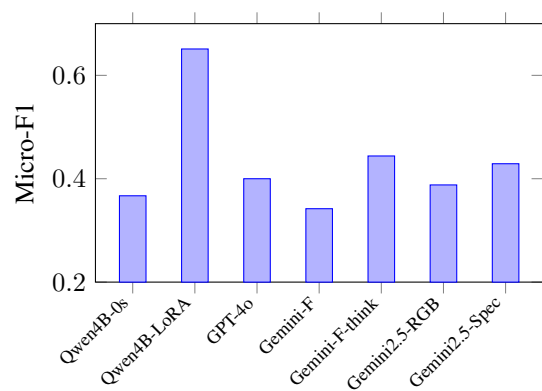


Figure 5. Micro-F1 comparison across model families and modes on BigEarthNet-S2 (1k balanced test set). LoRA-adapted Qwen3-VL-4B (4-bit, $r=128$) surpasses zero-shot proprietary models and previously reported Gemini 2.5 variants using multispectral inputs.

Sample 994 | $\Delta F1=1.00$



GT: Agro-forestry areas, Arable land, Pastures

Base: Inland waters, Inland wetlands, Land principally occupied by agriculture, with significant areas of natural vegetation (F1=0.00)

FT: Agro-forestry areas, Arable land, Pastures (F1=1.00)

Sample 974 | $\Delta F1=1.00$

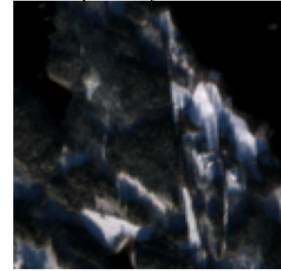


GT: Marine waters

Base: Land principally occupied by agriculture, with significant areas of natural vegetation (F1=0.00)

FT: Marine waters (F1=1.00)

Sample 93 | $\Delta F1=1.00$



GT: Coniferous forest, Inland waters, Mixed forest

Base: Natural grassland and sparsely vegetated areas (F1=0.00)

FT: Coniferous forest, Inland waters, Mixed forest (F1=1.00)

Sample 188 | $\Delta F1=1.00$



GT: Coniferous forest, Mixed forest, Transitional woodland, shrub

Base: Broad-leaved forest, Land principally occupied by agriculture, with significant areas of natural vegetation (F1=0.00)

FT: Coniferous forest, Mixed forest, Transitional woodland, shrub (F1=1.00)

Sample 418 | $\Delta F1=1.00$

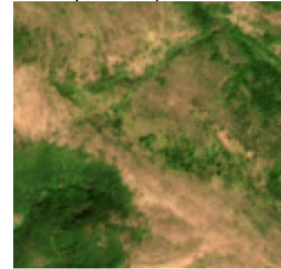


GT: Pastures

Base: Arable land, Complex cultivation patterns, Land principally occupied by agriculture, with significant areas of natural vegetation (F1=0.00)

FT: Pastures (F1=1.00)

Sample 930 | $\Delta F1=1.00$

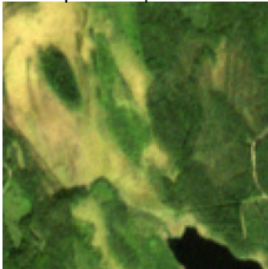


GT: Broad-leaved forest, Natural grassland and sparsely vegetated areas, Transitional woodland, shrub

Base: Land principally occupied by agriculture, with significant areas of natural vegetation (F1=0.00)

FT: Broad-leaved forest, Natural grassland and sparsely vegetated areas, Transitional woodland, shrub (F1=1.00)

Sample 899 | $\Delta F1=0.89$



GT: Coniferous forest, Inland waters, Inland wetlands, Mixed forest, Transitional woodland, shrub

Base: Broad-leaved forest, Land principally occupied by agriculture, with significant areas of natural vegetation (F1=0.00)

FT: Coniferous forest, Inland waters, Mixed forest, Transitional woodland, shrub (F1=0.89)

Sample 44 | $\Delta F1=0.86$



GT: Agro-forestry areas, Arable land, Complex cultivation patterns, Permanent crops

Base: Natural grassland and sparsely vegetated areas (F1=0.00)

FT: Agro-forestry areas, Arable land, Permanent crops (F1=0.86)

Sample 497 | $\Delta F1=0.86$



GT: Broad-leaved forest, Mixed forest, Transitional woodland, shrub

Base: Land principally occupied by agriculture, with significant areas of natural vegetation (F1=0.00)

FT: Broad-leaved forest, Land principally occupied by agriculture, with significant areas of natural vegetation, Mixed forest, Transitional woodland, shrub (F1=0.86)

Sample 484 | $\Delta F1=0.80$

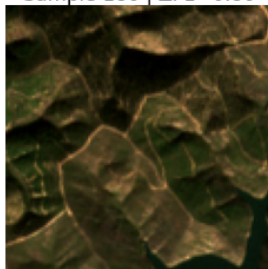


GT: Arable land, Broad-leaved forest, Coastal wetlands, Mixed forest, Pastures

Base: Land principally occupied by agriculture, with significant areas of natural vegetation (F1=0.00)

FT: Arable land, Broad-leaved forest, Coastal wetlands, Marine waters, Pastures (F1=0.80)

Sample 180 | $\Delta F1=0.80$

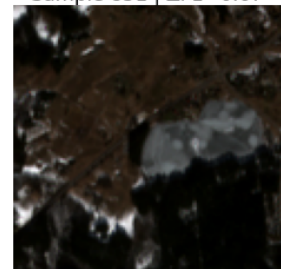


GT: Inland waters, Moors, heathland and sclerophyllous vegetation, Transitional woodland, shrub

Base: Land principally occupied by agriculture, with significant areas of natural vegetation (F1=0.00)

FT: Moors, heathland and sclerophyllous vegetation, Transitional woodland, shrub (F1=0.80)

Sample 851 | $\Delta F1=0.67$



GT: Arable land, Complex cultivation patterns, Coniferous forest, Mixed forest, Transitional woodland, shrub, Urban fabric

Base: Land principally occupied by agriculture, with significant areas of natural vegetation (F1=0.00)

FT: Coniferous forest, Mixed forest, Transitional woodland, shrub (F1=0.67)