

Cultural image (test set)

Stage 1

Step 1: VLM Spanish captioning

Qwen2.5-VL-72B-Instruct (4-bit) or
Qwen3-VL-8B
Prompt: culturally-aware, noun-first style

Spanish caption q (1 sentence per image)

Training

R

Dev

D

Stage 2

**Step 2: BM25 retrieval
+ dev exemplar selection**

Retrieve r nearest Es-LoRes pairs
Select d dev pairs (leave-one-out)
No reranking in frozen config

Step 3: Many-shot prompt assembly
System: "translate Es-LoRes, match example style"
 d gold dev pairs + r retrieved train pairs
~3-5K tokens per prompt

Step 4: Gemini 2.5 Flash translation
temperature=0.0, thinking=off, max_tokens=120
Learns register + vocabulary from
in-context examples

Step 5: Clean + format
Strip prefixes, normalize whitespace

**Target language caption
(submission JSONL)**

Evaluated by chrF++
+ human judges (top 5)