System Architecture: Domain-Adaptive Zero-Shot Classification

Input Image + Domain Hint

Domain Inference

Domain-Aware Prompt Generation
- Natural templates
- Medical templates
- Satellite templates

CLIP Image Encoder (ViT-L/14)

Class Prototypes (Adaptive)

CLIP Text Encoder (ViT-L/14)

Cosine Similarity Computation

Caption Generator (BLIP-2)

Top-K Selection (K=5)

LLM Reasoning (Gemini 1.5)
Combines caption + top-k + domain context

Confidence Check
threshold > 0.15?

Adaptive Update
EMA:
$q \leftarrow (1-\alpha)q + \alpha v$

Yes

Final Prediction + Explanation