



Large Multimodal Models as General In-Context Classifiers

Marco Garosi¹, Matteo Farina¹, Alessandro Conti¹, Massimiliano Mancini¹, Elisa Ricci^{1,2}

¹ University of Trento, ² Fondazione Bruno Kessler

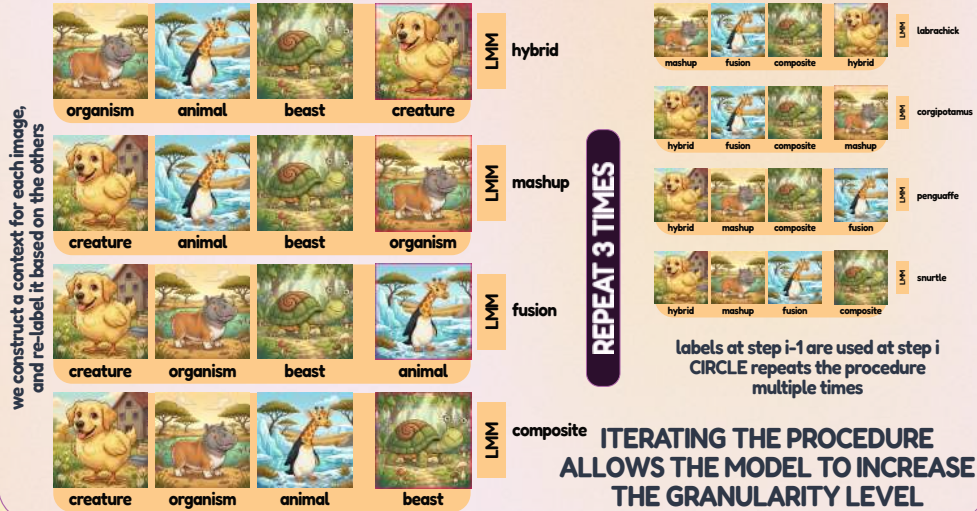
INPUT: N UNANNOTATED IMAGES



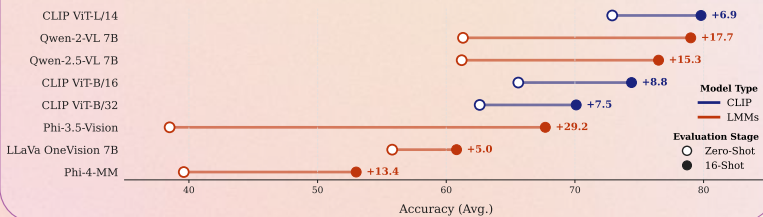
CIRCLE: INITIAL PSEUDO-LABEL



CIRCLE: CONTEXTUAL LABEL REFINEMENT



CLOSED WORLD: CONTEXT IMPROVEMENTS



OUTPUT: REFINED CONTEXT



OPEN WORLD: CIRCLE vs ALL, ALL MODELS

Model	Method	Prototypical				Non-prototypical				Fine-grained				Very fine-grained			
		U	SS	bCS	mCS	U	SS	bCS	mCS	U	SS	bCS	mCS	U	SS	bCS	mCS
Contrastive Vision-Language Models																	
CaSED (ViT-L/14)		46.3	58.9	59.8	58.8	18.6	41.8	42.4	41.8	46.6	60.7	61.6	60.7	47.1	38.5	38.5	38.5
CLIP retrieval (ViT-L/14)		81.1	65.6	76.2	44.4	61.6	43.4	49.4	23.8	53.1	40.2	64.4	30.7	49.1	38.9	49.1	38.6
Large Multimodal Models																	
Qwen2-VL 7B	Zero-Shot	78.7	51.9	76.0	43.7	42.6	30.8	49.8	29.2	64.0	39.2	62.9	31.9	63.0	34.5	43.4	33.1
	Random Ctx	24.4	41.4	52.7	39.7	17.1	23.4	41.3	23.0	31.7	34.4	44.8	34.6	31.1	29.2	34.1	27.9
	Pseudo ICL	81.1	53.4	76.2	44.4	42.8	31.2	50.1	26.9	53.1	40.2	64.4	30.7	49.1	38.9	49.1	38.6
	CIRCLE	91.5	65.6	74.3	63.5	61.6	41.9	49.4	40.5	87.3	61.1	72.0	57.3	91.5	42.5	50.2	39.8
Qwen2.5-VL 7B	Zero-Shot	82.9	47.9	79.9	31.1	45.9	30.5	54.0	24.8	73.8	47.0	78.9	29.5	69.0	45.8	68.6	27.1
	Random Ctx	14.0	29.3	36.7	29.6	8.6	26.0	39.3	24.1	21.0	33.2	35.8	33.4	34.7	52.4	70.2	42.0
	Pseudo ICL	80.6	49.3	78.6	31.0	42.7	31.6	53.2	24.2	63.8	45.1	74.8	28.4	39.7	46.0	63.4	25.5
	CIRCLE	94.9	67.7	68.1	67.2	67.6	42.6	45.1	42.3	86.3	60.1	60.9	59.7	93.6	36.4	36.6	36.5
LLaVa OV 7B	Zero-Shot	53.2	56.2	62.0	53.4	28.1	31.6	43.8	30.2	40.4	39.0	43.9	37.2	76.7	31.8	32.3	30.9
	Random Ctx	14.0	29.3	36.7	29.6	8.6	26.0	39.3	24.1	21.0	33.2	35.8	33.4	34.7	52.4	70.2	42.0
	Pseudo ICL	19.7	31.2	41.1	30.5	3.3	18.3	31.5	19.7	20.3	35.5	40.0	34.7	70.4	30.6	31.1	29.8
	CIRCLE	72.2	74.0	74.0	74.0	61.7	55.3	55.3	55.3	55.1	46.0	46.0	46.0	74.2	32.9	32.8	32.9
Phi-3.5-V	Zero-Shot	60.7	48.2	65.6	46.1	28.7	24.9	36.7	24.1	50.7	32.1	47.2	31.3	54.2	29.5	36.3	29.8
	Random Ctx	41.1	48.3	54.4	48.4	10.7	26.3	36.4	27.4	25.2	26.3	36.4	27.4	59.2	27.1	31.7	26.3
	Pseudo ICL	54.1	44.1	61.7	40.1	23.7	22.8	35.1	21.5	43.1	33.0	48.6	29.7	24.9	32.4	41.8	32.5
	CIRCLE	92.1	59.7	63.2	60.3	58.3	30.0	35.2	31.7	88.1	39.2	45.7	42.1	99.6	33.0	33.6	33.1
Phi-4-MM	Zero-Shot	49.8	57.4	58.7	57.2	21.2	29.2	32.7	29.2	37.7	39.2	39.2	39.1	73.6	31.6	31.7	31.6
	Random Ctx	11.6	32.8	32.8	32.8	5.9	28.5	28.6	28.5	31.3	37.9	37.9	37.9	74.6	30.8	30.7	30.6
	Pseudo ICL	51.9	61.5	62.0	61.5	15.1	26.6	31.2	26.7	37.7	41.6	41.7	41.5	72.8	31.9	31.9	31.9
	CIRCLE	91.5	65.5	70.1	66.4	67.6	43.2	46.1	43.4	79.1	53.3	55.5	53.0	75.2	40.2	42.5	37.9

INFERENCE

