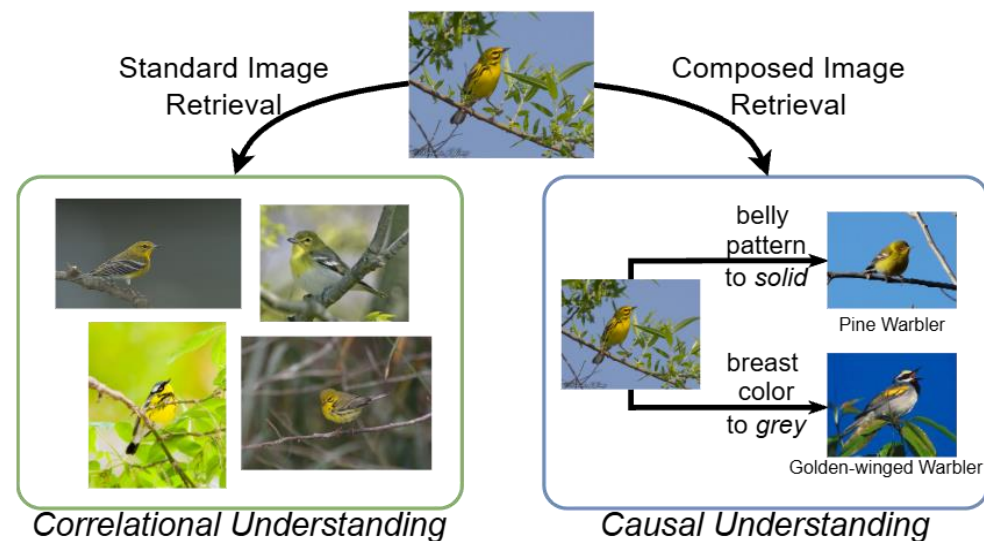

Retrieving Counterfactuals Improves Visual In-Context Learning

Guangzhi Xiong, Sanchit Sinha, Zhenghao He, Aidong Zhang

University of Virginia

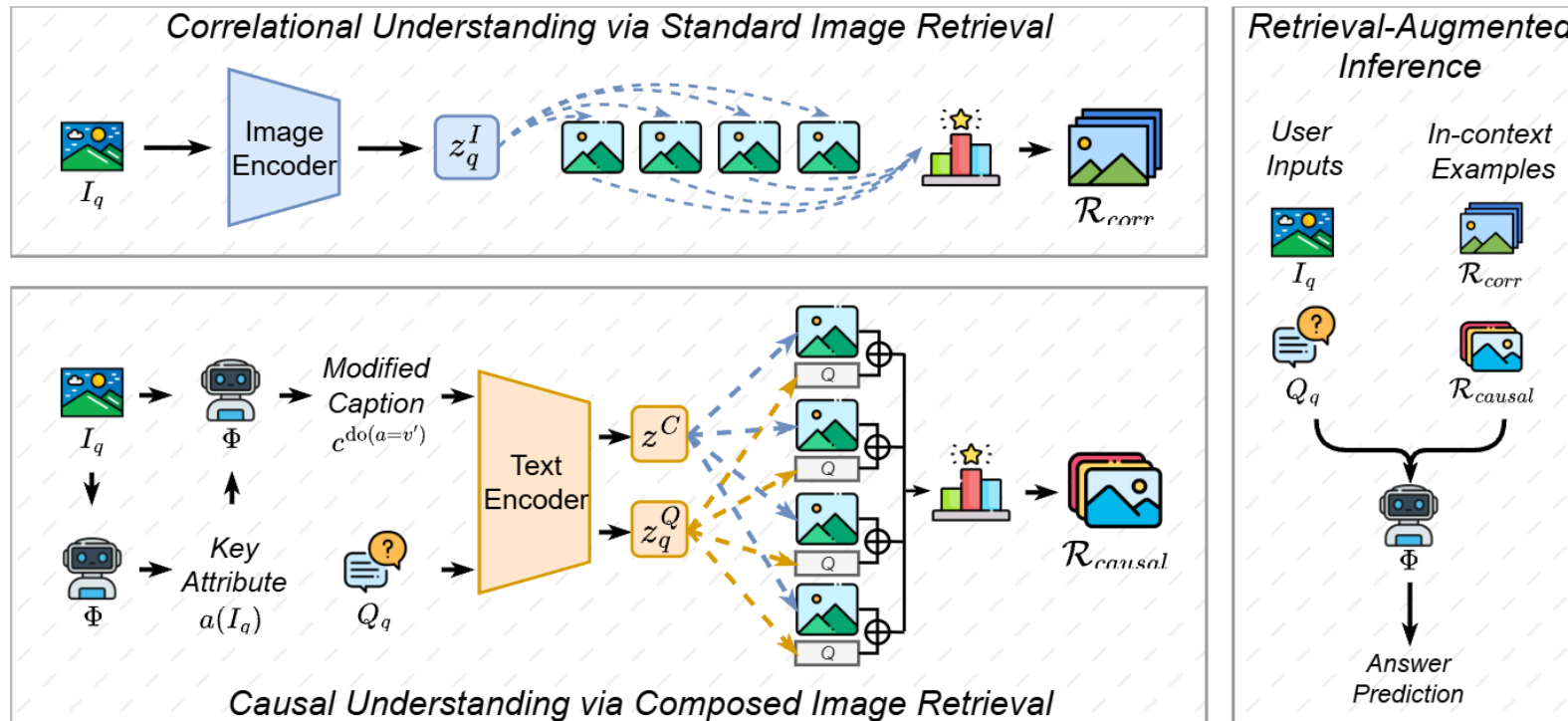
Motivation

- The effectiveness of visual ICL critically depends on the selection of examples.
- Existing retrieval-augmented approaches rely on passive similarity-based retrieval.
- The selection of correlated but non-causal examples may amplify spurious associations.



CIRCLES: Composed Image Retrieval for Causal Learning Example Selection

- We propose CIRCLES, an ICL framework that enriches demonstration sets by retrieving counterfactual examples, moving beyond standard similarity-based retrieval.



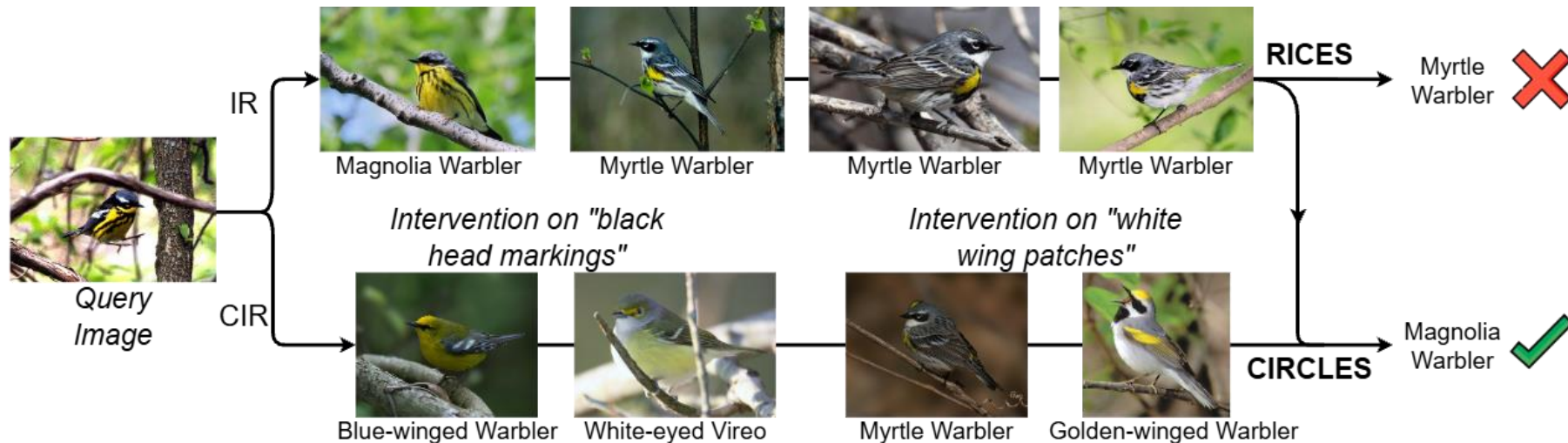
Main Results: Quantitative Analysis

- CIRCLES outperforms zero-shot prompting, random example selection, and prior retrieval-based baselines across classification and VQA tasks.
- The improvements are especially notable for smaller vision-language models, showing that better demonstrations compensate for weaker internal knowledge.

Model	Method	CUB		Flowers		OK-VQA		VizWiz		Average	
		Acc	F1	Acc	F1	EM	F1	EM	F1	EM	F1
Gemma3 -4B	None	10.56	8.19	46.76	46.50	19.12	26.70	39.85	57.26	29.07	34.66
	Random	9.98	8.10	37.01	39.51	24.81	30.74	51.84	67.08	30.91	36.36
	RICES	65.40	67.62	86.70	87.43	26.65	32.72	56.08	70.40	58.71	64.54
	MUIER	65.21	67.42	86.39	87.28	26.87	32.82	56.59	70.75	58.77	64.57
	MMICES	13.95	10.98	38.09	38.20	26.24	32.41	52.72	68.21	32.75	37.45
	CIRCLES	71.97	72.39	93.32	93.49	31.27	36.89	57.61	71.35	63.54	68.53
Gemma3 -12B	None	30.34	25.02	67.51	64.33	25.47	33.89	56.82	71.26	45.04	48.62
	Random	29.27	25.82	71.28	71.56	33.59	39.56	70.13	80.54	51.07	54.37
	RICES	76.37	76.25	96.44	96.29	36.86	42.61	73.98	83.43	70.91	74.65
	MUIER	76.51	76.39	96.42	96.28	36.58	42.30	73.72	83.04	70.81	74.50
	MMICES	36.37	31.61	71.36	69.13	35.12	40.85	71.50	81.56	53.59	55.79
	CIRCLES	77.03	76.90	97.77	97.75	37.75	43.23	74.30	82.35	71.71	75.06
Qwen2.5 -VL-3B	None	7.09	5.97	22.57	24.04	42.29	46.12	75.83	78.58	36.95	38.68
	Random	15.81	15.63	41.76	44.57	41.54	45.30	74.48	77.40	43.40	45.73
	RICES	72.26	73.88	93.06	92.84	42.57	46.26	70.80	73.37	69.67	71.59
	MUIER	72.25	73.87	93.06	92.83	41.66	45.82	71.61	74.74	69.64	71.81
	MMICES	17.26	15.38	33.84	35.47	39.32	42.95	73.23	76.21	40.91	42.50
	CIRCLES	74.89	76.34	94.70	94.77	43.24	47.27	72.93	75.33	71.44	73.43
Qwen2.5 -VL-7B	None	14.83	11.79	43.47	42.14	33.31	38.33	68.70	73.30	40.08	41.39
	Random	26.27	24.42	47.15	49.74	41.66	46.20	76.11	78.34	47.80	49.67
	RICES	82.15	81.91	98.83	98.85	43.66	48.32	73.79	76.13	74.61	76.30
	MUIER	82.14	81.90	98.93	98.95	44.29	48.76	74.67	77.08	75.01	76.67
	MMICES	34.57	29.72	49.15	47.62	43.40	47.66	70.69	72.77	49.45	49.44
	CIRCLES	82.17	82.13	98.99	99.04	43.54	48.53	77.63	82.36	75.58	78.02

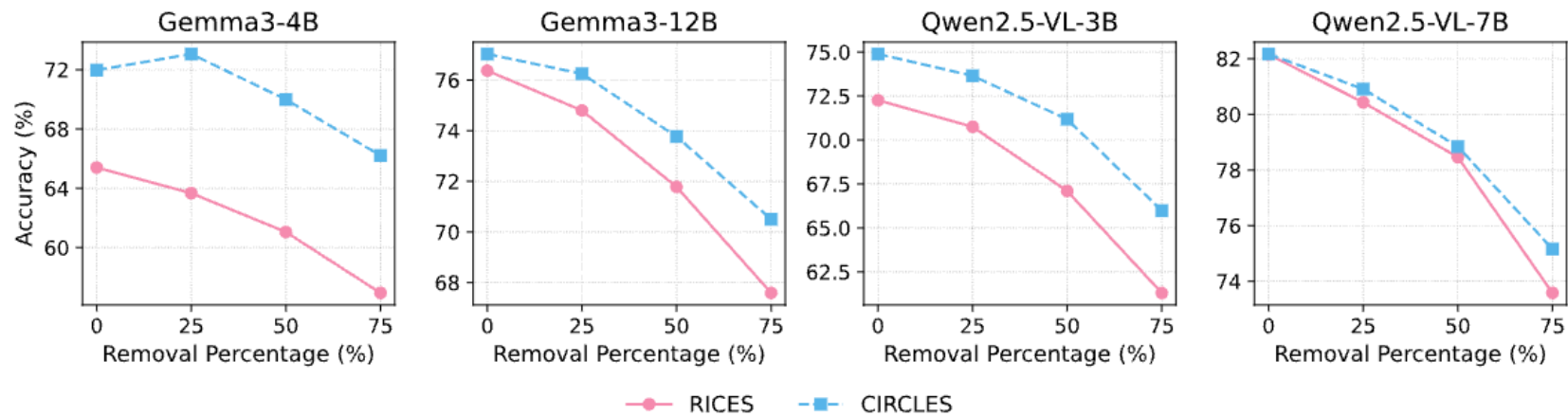
Main Results: Qualitative Analysis

- Standard retrieval returns lookalike neighbors that push the model toward the wrong species, while CIRCLES retrieves counterfactual examples that explicitly vary discriminative cues such as head markings and wing patches.



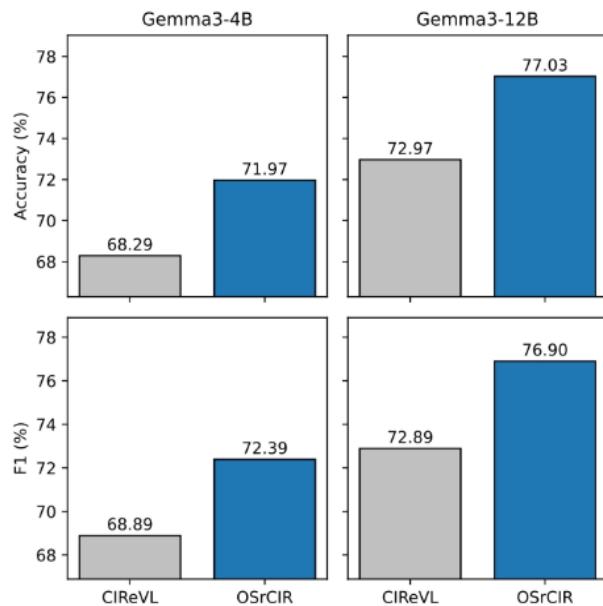
Discussions: CIRCLES under Information Scarcity

- The gap between CIRCLES and standard retrieval widens as the amount of available training data shrinks.



Discussions: Impact of CIR Implementation on CIRCLES

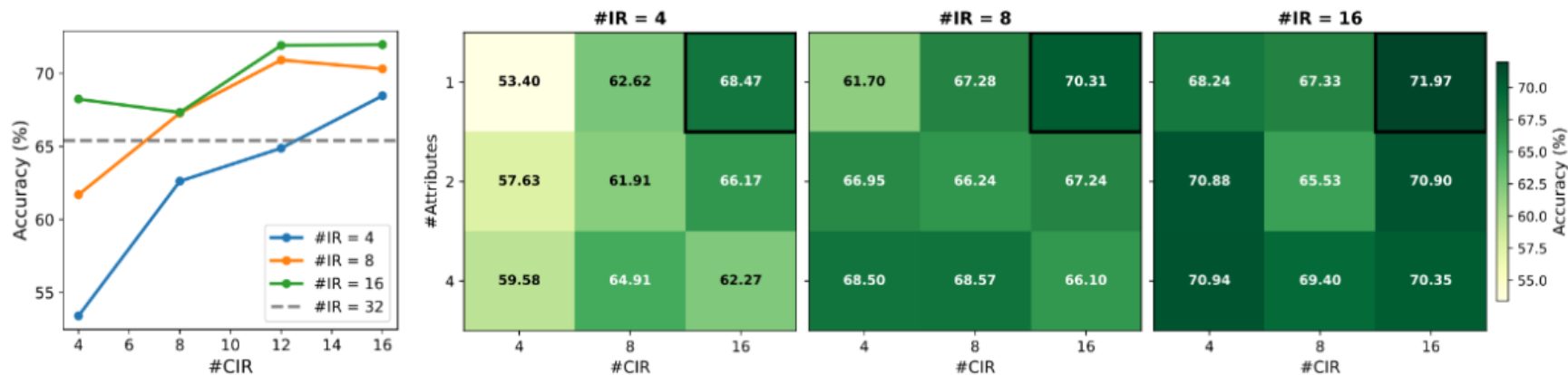
- The stronger CIR backend leads to better CIRCLES performance.
- Incorporating question-question similarity in CIR will lead to improved performance on both VQA datasets



Model	Setting	OK-VQA		VizWiz	
		EM	F1	EM	F1
Gemma-4B	w/o	27.72	33.64	57.40	71.65
	w/	31.27	36.89	57.61	71.35
Gemma-12B	w/o	33.02	39.07	74.37	83.20
	w/	37.75	43.23	74.30	82.35
Qwen-3B	w/o	41.12	45.33	73.28	75.94
	w/	43.24	47.27	72.93	75.33
Qwen-7B	w/o	40.80	46.22	77.29	81.72
	w/	43.54	48.53	77.63	82.36

Discussions: Retrieval Budget Analysis

- Increasing the number of composed retrieval examples (#CIR) consistently improves accuracy across all #IR configurations.
- As the CIR budget grows, the optimal allocation shifts from breadth (intervening on more attributes) to depth (focusing on fewer attributes).



Thank You!

Homepage: <https://gzxiong.github.io/CIRCLES>

Code: <https://github.com/gzxiong/CIRCLES>