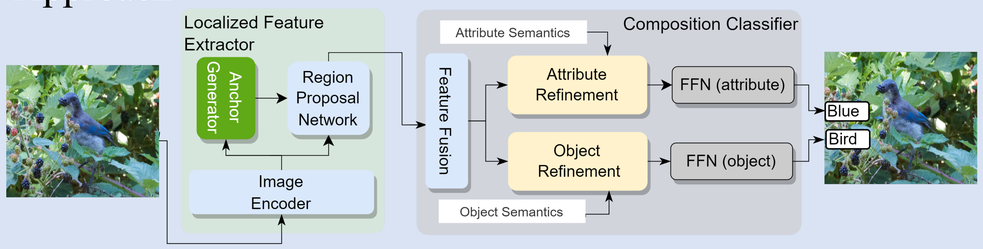
LOCL : Learning Object-Attribute Composition using Localization

Background

The problem of unseen O-A associations has been well studied in the field of Composition Zero shot Learning (CZSL); however, the performance of existing methods is limited in challenging scenes.

LOCL generalizes CZSL to objects in cluttered/more realistic settings.

Approach



Localized Feature Extractor (LFE) generates proposals that are likely to contain objects → The composition Classifier takes the proposal features and refine with object-attribute semantics → The refined featured are used to make the final object attribute prediction

Results

Methods	MIT-State			UT-Zappos			CGQA		
	Seen	Unseen	AUC	Seen	Unseen	AUC	Seen	Unseen	AUC
Attop	14.3	17.4	1.6	59.8	54.2	25.9	11.8	3.9	0.3
LabelEmbed	15	20.1	2.0	53.0	61.9	25.7	16.1	5	0.6
TMN	20.2	20.1	2.9	58.7	60.0	29.3	21.6	6.3	1.1
SymNet	24.2	25.2	3.0	49.8	57.4	23.4	25.2	9.2	1.8
CompCos	25.3	24.6	4.5	59.8	62.5	28.1	28.1	11.2	2.6
ProtoProp	-	-	-	62.1	65.5	34.7	26.4	18.1	3.7
BMP-Net	**38.6**	21.7	6.0	**87.3**	64.5	**49.7**	-	-	-
CGE	32.8	28.0	6.5	64.5	71.5	33.5	**31.4**	14	3.6
LOCL (Ours)	35.3	**36.0**	**7.7**	68.0	**76.7**	37.9	29.6	**26.4**	**4.2**

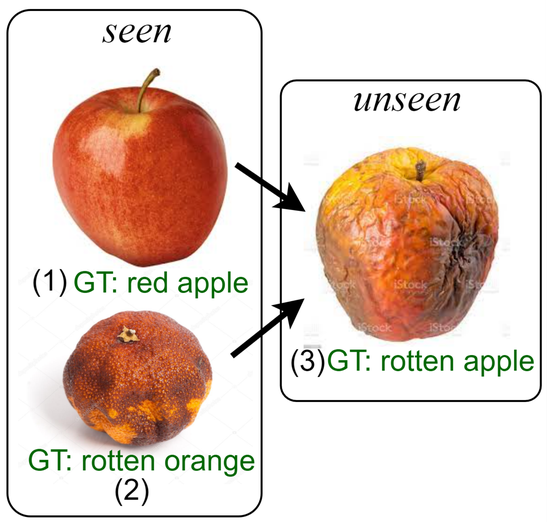
Performance comparison with SOTA methods on simple datasets (MIT-states & UT-Zappos) and challenging dataset (CGQA). Table below shows effectiveness of the Localized Feature Extractor (LFE)

Methods	Our BB	LFE	CGQA			MIT-States		
			Seen	Unseen	AUC	Seen	Unseen	AUC
	✗	✗	25.2	9.2	1.8	24.2	25.2	3.0
SymNet	✓	✗	25.3	9.3	1.8	26.6	26.1	3.5
	✓	✓	27.7	13.5	2.0	28.7	27.7	3.8
	✗	✗	28.1	11.2	2.6	25.3	24.6	4.5
CompCos	✓	✗	28.4	13.5	2.8	25.6	24.8	4.5
	✓	✓	28.9	16.7	2.9	27.9	26.7	5.1
	✗	✗	31.4	14.0	3.6	32.8	28	6.5
CGE	✓	✗	31.4	19.3	3.8	33.3	28	6.5
	✓	✓	**31.9**	26.1	4.1	**36.3**	29.8	6.6
LOCL	✓	✓	29.6	**26.4**	**4.2**	35.3	**36.0**	**7.7**

Object Localization leads to Right object-attribute association in a cluttered environment.

Correct
Misclassified

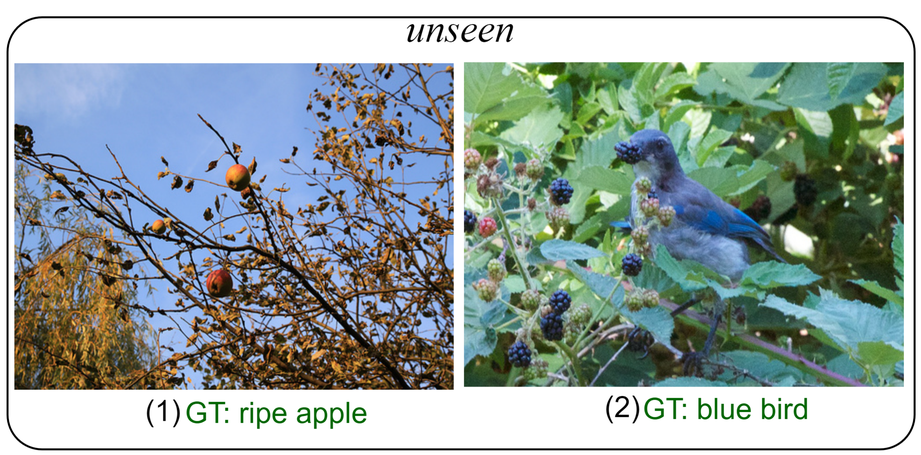
Simple Scenarios



Predictions {All Methods : (1) red apple, (2) rotten orange, (3) rotten apple}

A

Challenging Scenarios



Predictions {SymNet: small bird, CGE:clear tree, CompCos: blue berries, **LOCL:** {red apple, ripe apple}}

Predictions {SymNet: green bird, CGE:lying bird, CompCos: cooked bird, **LOCL:** blue bird}

B

AUTHORS: SATISH KUMAR, ASM Iftekhar, Ekta Prashnani, B.S. Manjunath

UC Santa Barbara

This research was in parts supported by NSF SI2-SSI award #1664172 and US Army Research Laboratory (ARL) under agreement # W911NF2020157