CVPR SEATTLE WASHINGTON JUNE 16-18 2020

A Shared Multi-Attention Framework for Multi-Label Zero-Shot Learning

Dat Huynh and **Ehsan Elhamifar**

Khoury College of Computer Sciences

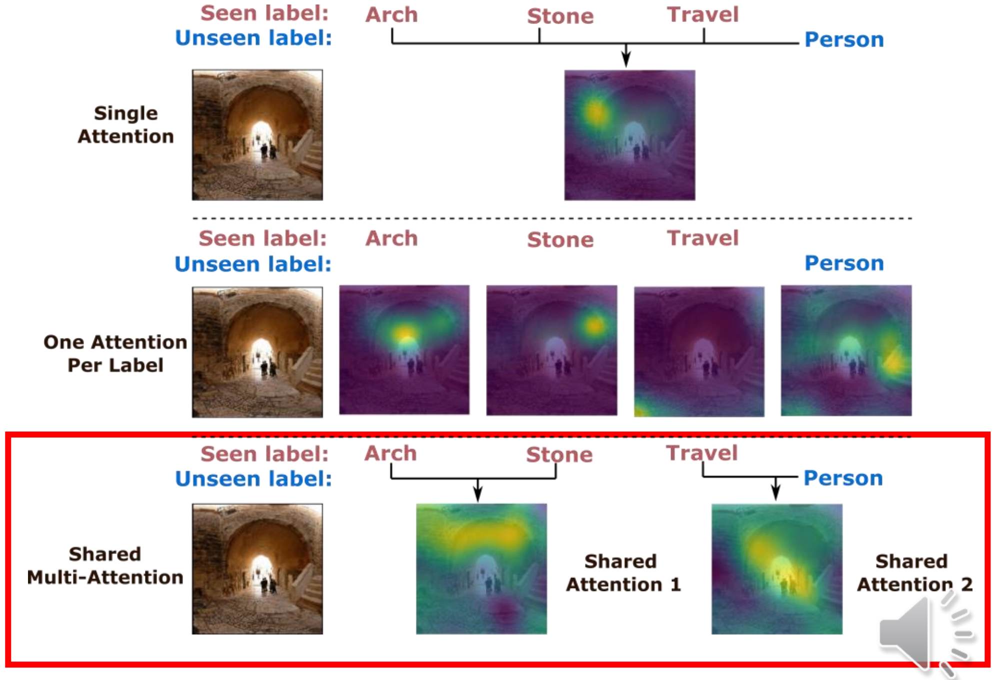
Northeastern University

Motivation

- **Multi-label Learning**:
  - Recognize all labels in an image
  - Require large costly annotations
- **Multi-label Zero-Shot Learning**:
  - Recognize both seen and unseen labels
  - Annotations for only seen labels
- Few work on multi-label ZSL
  - Holistic feature ⟶ cannot encode all labels
  - Ignore labels from small regions
- **Contributions**:
  - **Shared multi-attention** features for ZSL
  - **Transfers knowledge** between seen/unseen

Arch
Stone
Travel
Person

Seen label	Unseen label
Arch	
Stone	Person
Travel	



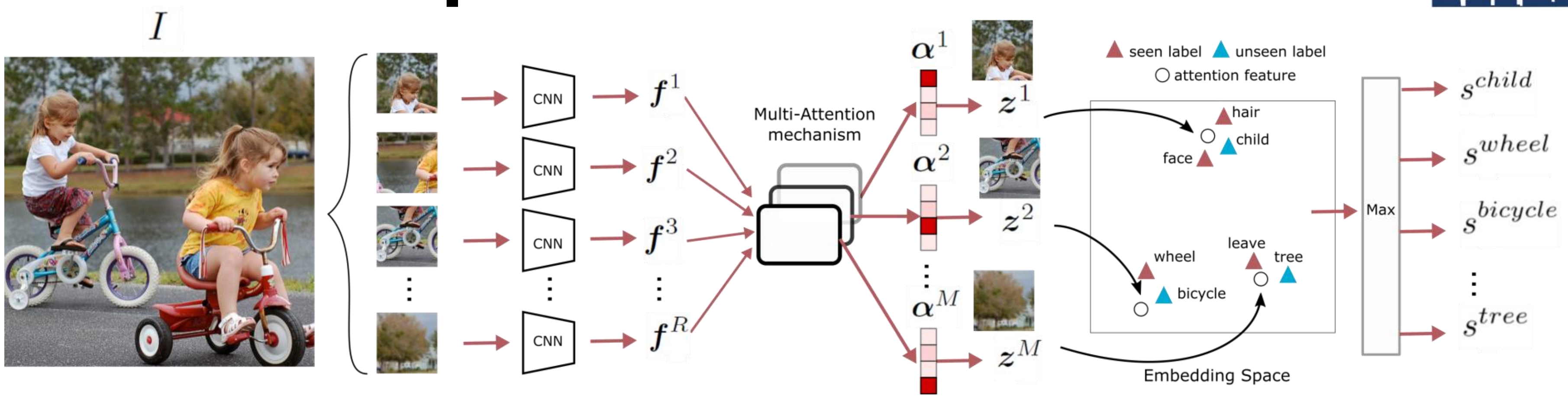
Seen label: Arch Stone Travel
Unseen label: Person
Single Attention

Seen label: Arch Stone Travel
Unseen label: Person
One Attention Per Label

Seen label: Arch Stone Travel
Unseen label: Person
Shared Multi-Attention
Shared Attention 1
Shared Attention 2

Proposed Architecture

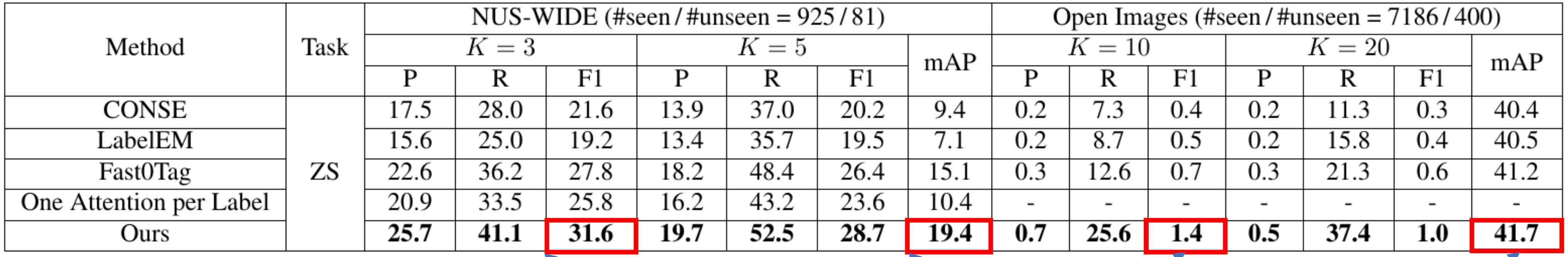
CVPR SEATTLE WASHINGTON JUNE 16-18 2020



- **Multiple Soft Attention NNs**:
  - Generating **multiple attention feature** for an image
- **Attention Selection** (label-agnostic):
  - For **each label**, **choose** attention feature maximizing **prediction score**
- Learning:
  - **Diversity Loss**: **Minimize overlap** between attention
  - **Relevance Loss**: Focus only on **regions improving prediction**
  - **Distribution Loss**: Effectively **use all attention** modules

Experiments

- **Recognition:** outperforms SOTA on **NUS-WIDE** and **Open Images**



Method	Task	NUS-WIDE (#seen / #unseen = 925 / 81)							Open Images (#seen / #unseen = 7186 / 400)						
		$K = 3$			$K = 5$			mAP	$K = 10$			$K = 20$			mAP
		P	R	F1	P	R	F1		P	R	F1	P	R	F1	
CONSE	ZS	17.5	28.0	21.6	13.9	37.0	20.2	9.4	0.2	7.3	0.4	0.2	11.3	0.3	40.4
LabelEM		15.6	25.0	19.2	13.4	35.7	19.5	7.1	0.2	8.7	0.5	0.2	15.8	0.4	40.5
Fast0Tag		22.6	36.2	27.8	18.2	48.4	26.4	15.1	0.3	12.6	0.7	0.3	21.3	0.6	41.2
One Attention per Label		20.9	33.5	25.8	16.2	43.2	23.6	10.4	-	-	-	-	-	-	-
Ours		**25.7**	**41.1**	**31.6**	**19.7**	**52.5**	**28.7**	**19.4**	**0.7**	**25.6**	**1.4**	**0.5**	**37.4**	**1.0**	**41.7**

+3.8% (F1@3)

+4.3% (mAP)

+0.7% (F1@10)

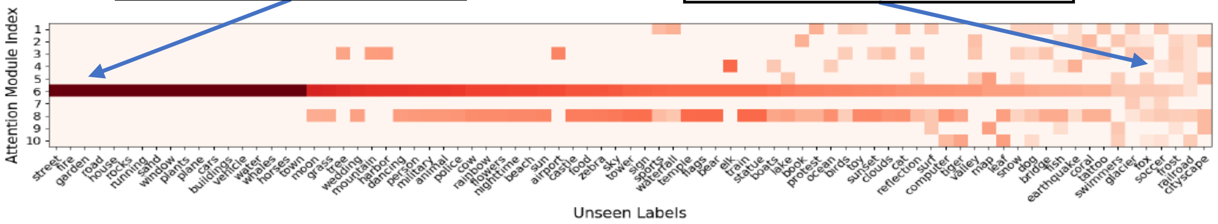
+0.5% (mAP)

- **Qualitative Results:**

Attention utility depends on **label complexity**

1 attention:
`street', `house, …

Multiple attentions:
`soccer', `railroad', …



Successfully attend **relevant image regions**