

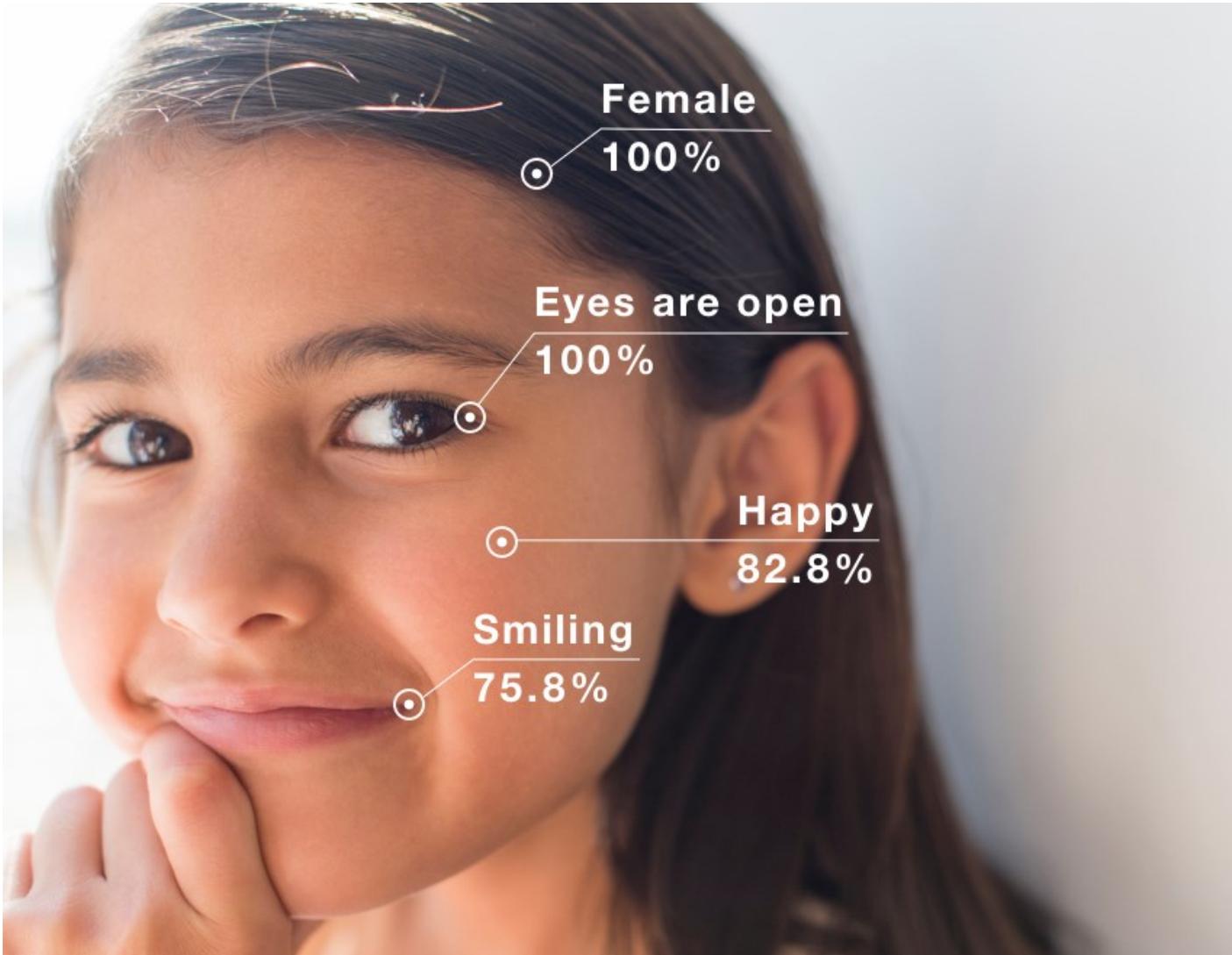
NANYANG  
TECHNOLOGICAL  
UNIVERSITY  
SINGAPORE

# Sensing, Understanding, and Synthesizing Humans in an Open World

Ziwei Liu

Nanyang Technological University

# Human-Centric AI



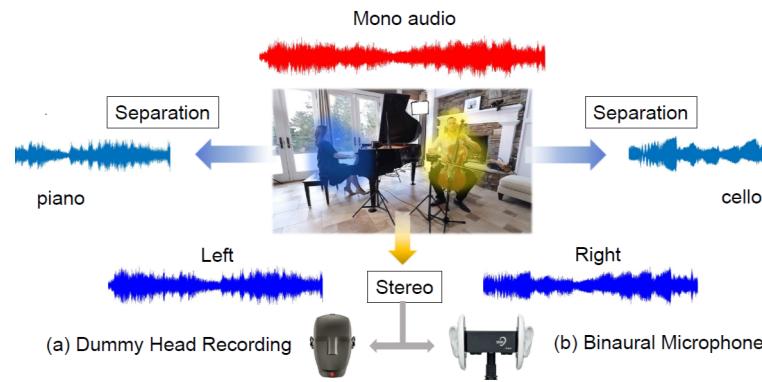
# Human-Centric AI





## Robust Sensing

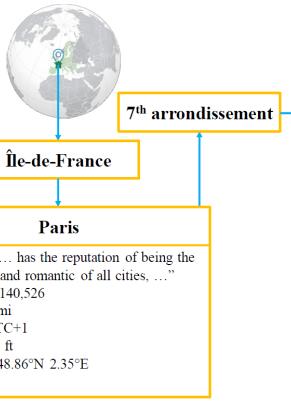
CelebA-Spoof



## Synthesizing across Modalities

Sep-Stereo

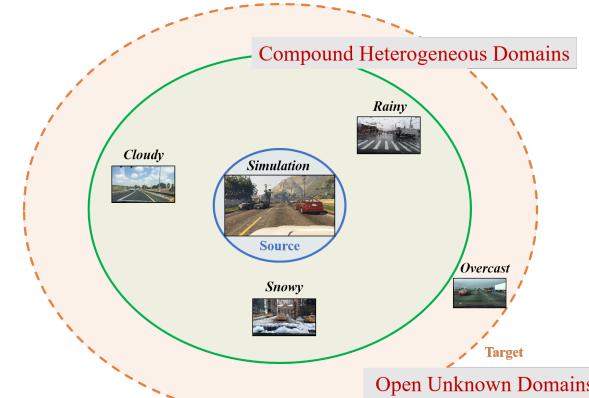
Musée du Louvre
Street-address: Place du Carrousel
Function: See
Category: Museums
Description: "Its most famous exhibit, .... is Leonardo da Vinci's painting of the Mona Lisa, ...."
Coordinates: 48.86°N 2.34°W
Opening Hours: W-M 09:00-18:00, closed public holidays ...
Price: €12-16; under 18, free ...
...



La Tour Eiffel
Nickname: The Eiffel Tower
Function: See
Category: Landmarks
Description: "A symbol of Paris and one of the most famous landmarks in the world. ..."
Coordinates: 48.86°N 2.29°W
Opening Hours: 09:30-23:45; Jun 21-Sep 02:09:00-00:45; Jul 14-Jul 15 off
Price: 25€ (12€ for age 12-24) ...
...

## Understanding beyond Recognition

Placepedia



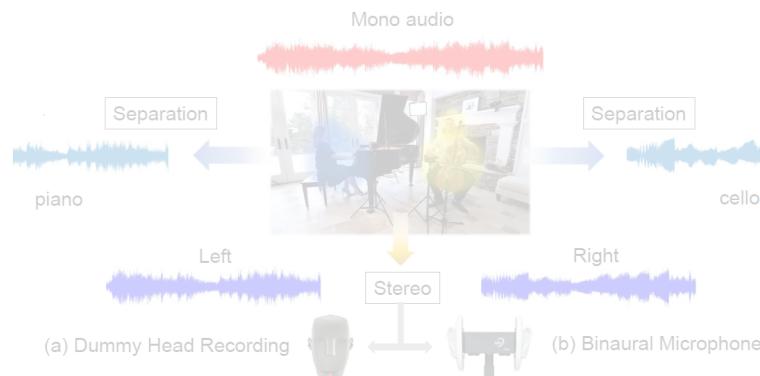
## Open World Learning

Open Compound Domain Adaptation



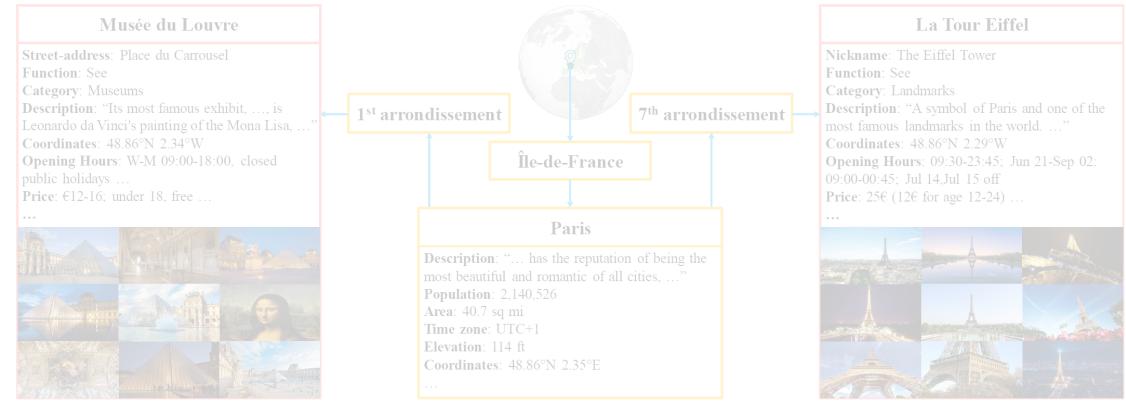
# Robust Sensing

## CelebA-Spoof



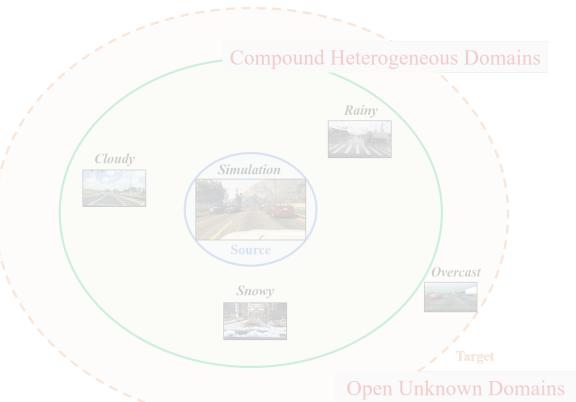
# Synthesizing across Modalities

## Sep-Stereo



# Understanding beyond Recognition

## Placepedia



# Open World Learning

## Open Compound Domain Adaptation



# CelebA-Spoof: Large-Scale Face Anti-Spoofing Dataset With Rich Annotations

Yuanhan Zhang<sup>\*1,2</sup> Zhenfei Yin<sup>\*2</sup> Yidong Li<sup>1</sup> Guojun Yin<sup>2</sup> Junjie Yan<sup>2</sup> Jing Shao<sup>2</sup> Ziwei Liu<sup>3</sup>

<sup>1</sup>Beijing Jiaotong University

<sup>2</sup>SenseTime Research

<sup>3</sup>The Chinese University of Hong Kong

<sup>\*</sup>Equal contribution

# Introduction

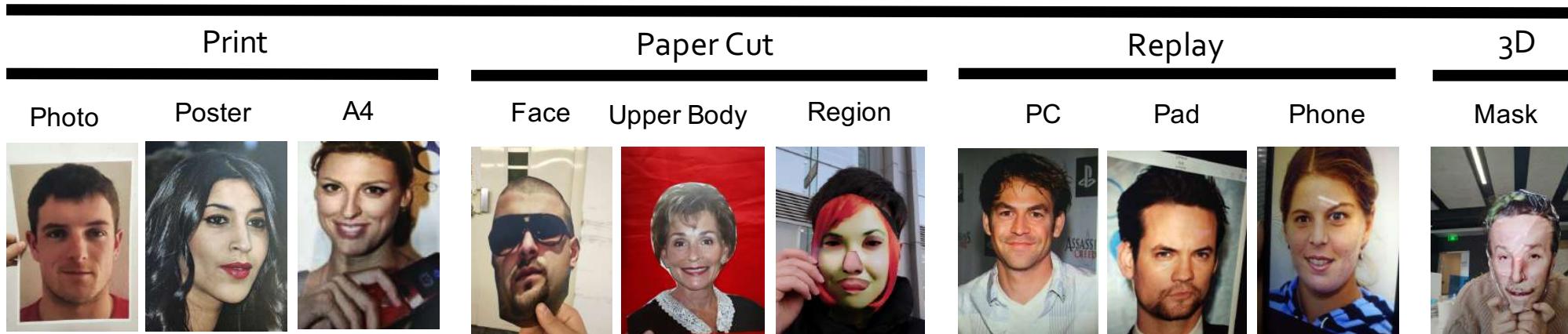
# CelebA-Spoof Dataset

1. Lack of Diversity
2. Lack of Annotations

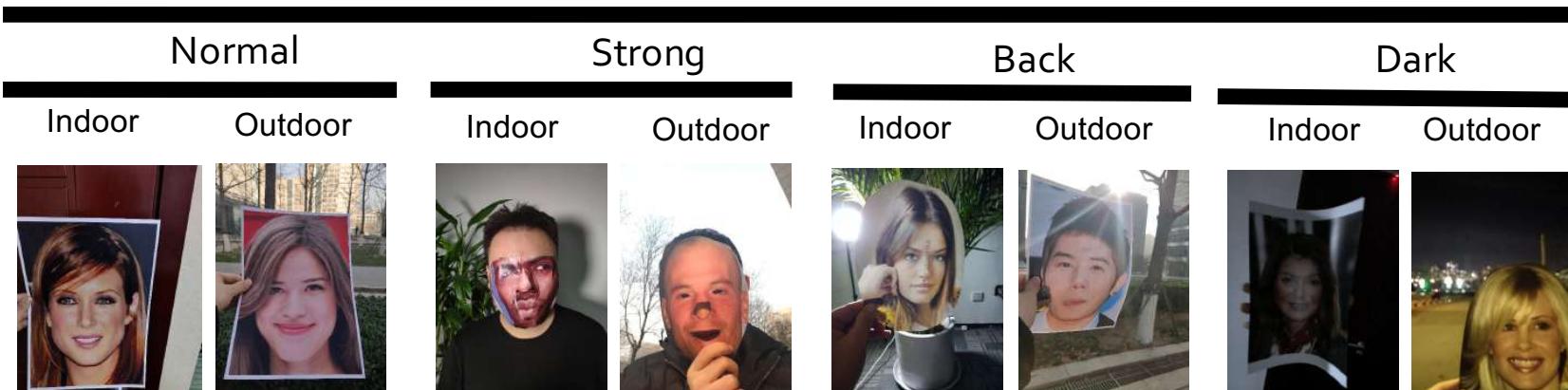
Dataset	Year	#Subjects	#Data(V/I)	#Annotation
Replay-Attack	2012	50	1,200 (V)	
CASIA-MFSD	2012	50	600 (V)	
3DMAD	2014	14	255 (V)	
MSU-MFSD	2015	35	440 (V)	
Mspoof	2015	21	4,704 (V)	
HKBU-MARs V2	2016	12	1,008 (V)	
MSU-USSA	2016	1,140	10,260 (I)	1
Oulu-NPU	2017	55	5,940 (V)	
SiW	2018	165	4,620 (V)	
CASIA-SURF	2018	1,000	21,000 (V)	
CSMAD	2018	14	260 (V), 17 (I)	
HKBU-MARs V1 +	2018	12	180 (V)	
SiW-M	2019	493	1,628 (V)	
CelebA-Spoof	2020	10,177	625,537 (I)	43

# CelebA-Spoof Dataset

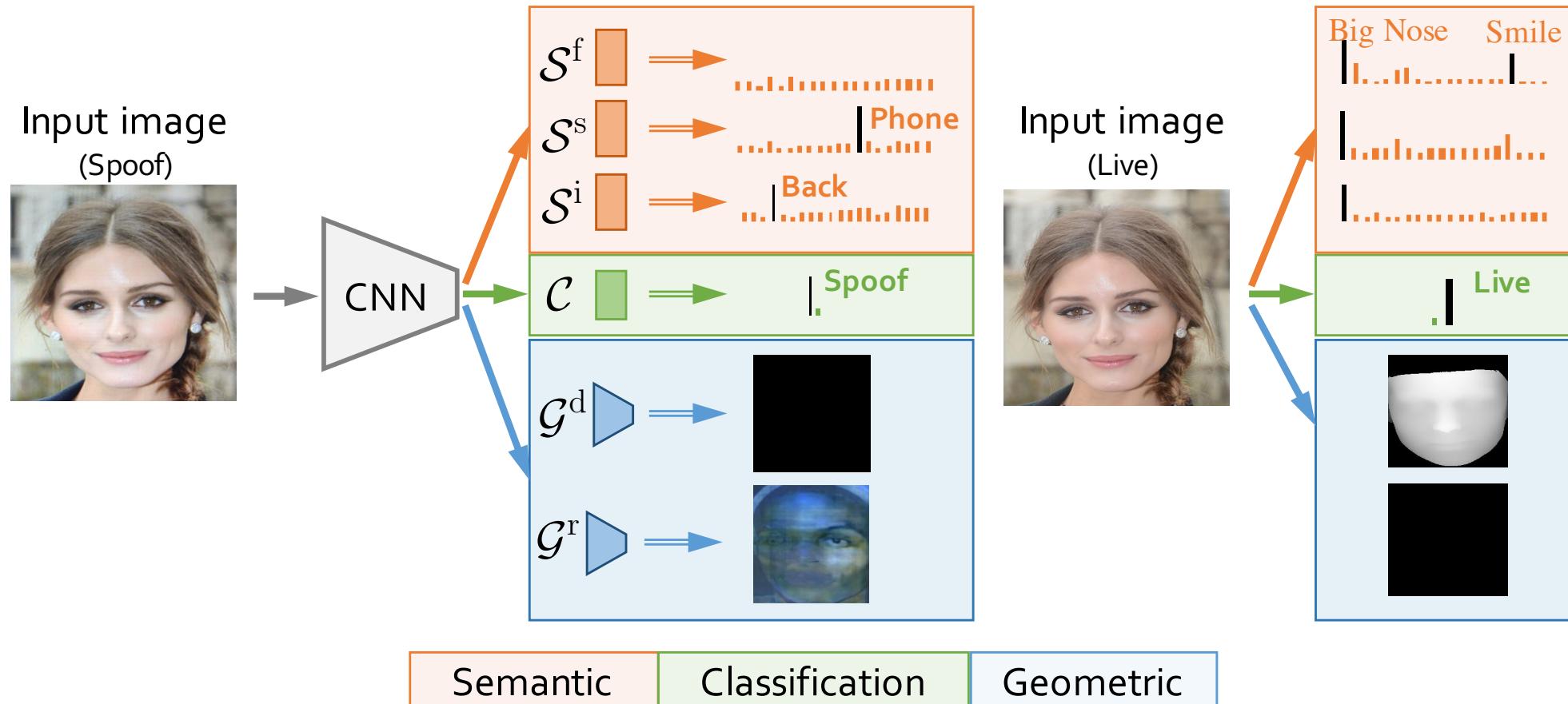
Presentation Attack Instrument



Illumination Condition and Environment



# Auxiliary Information Embedding Network (AENet)



# Observation

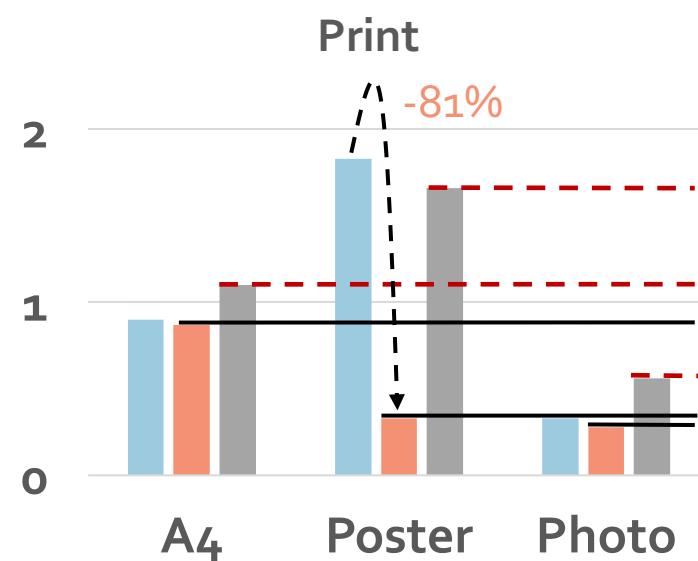
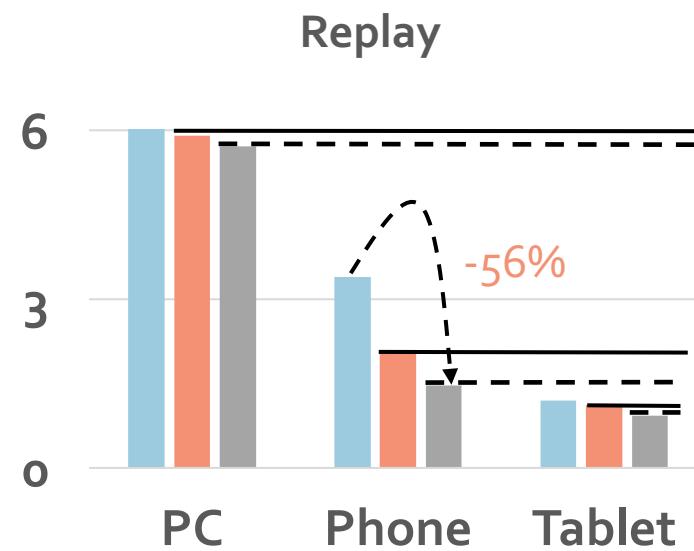
# Depth Maps are More Versatile

Smaller  $\text{APCER}_{\mathcal{S}^s}$  is better

----- Reflection map as auxiliary supervision

—— Depth map as auxiliary supervision

■ Baseline ■  $\text{AENet}_{\mathcal{C}, \mathcal{G}^W}/o\mathcal{G}^R$  ■  $\text{AENet}_{\mathcal{C}, \mathcal{G}^W}/o\mathcal{G}^D$

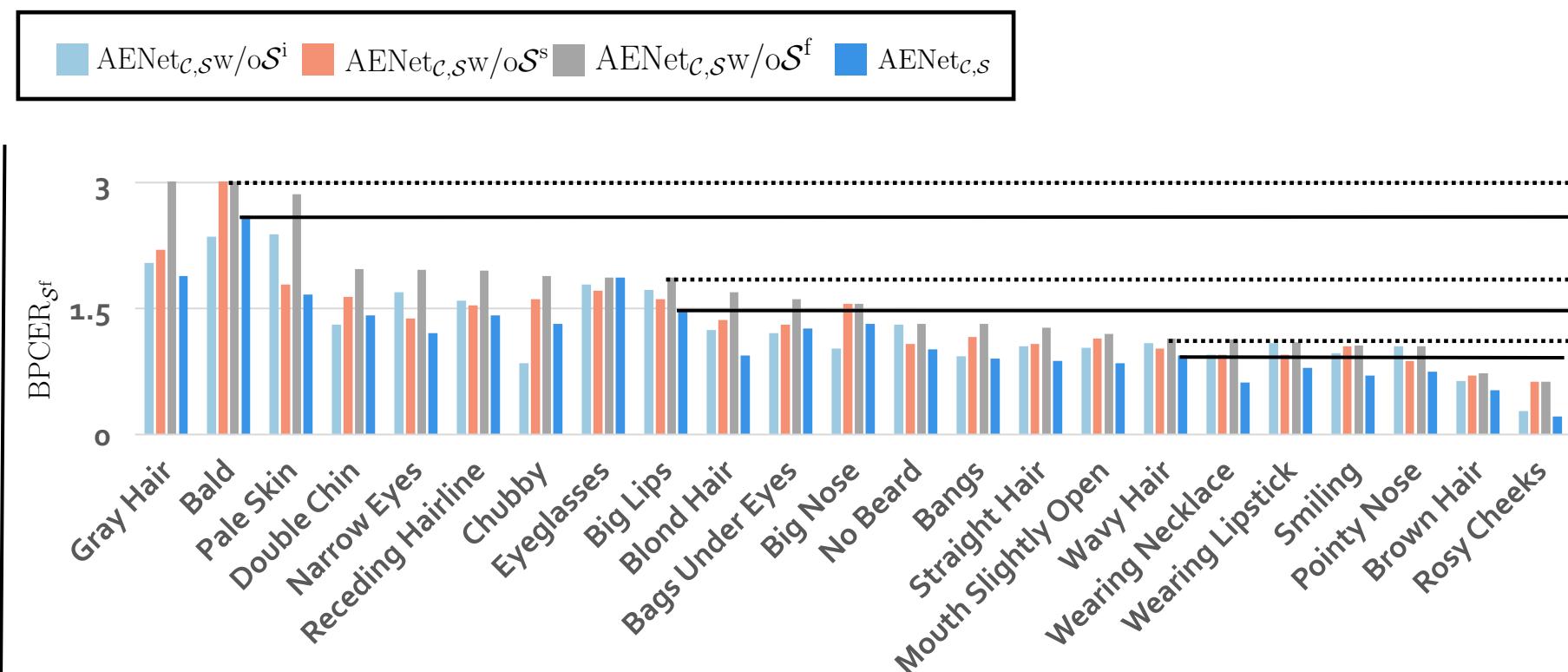
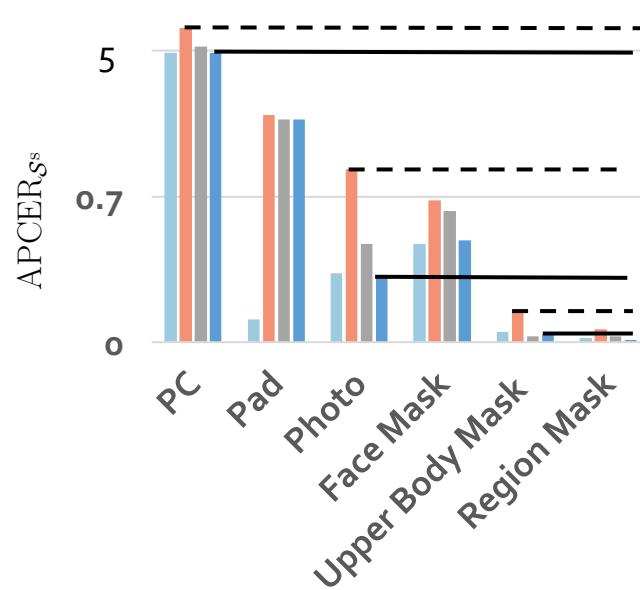


Depth Maps are More Versatile

# Semantic Attribute Matters

Smaller  $\text{APCER}_{\mathcal{S}^s}$  and  $\text{BPCER}_{\mathcal{S}^f}$  are better

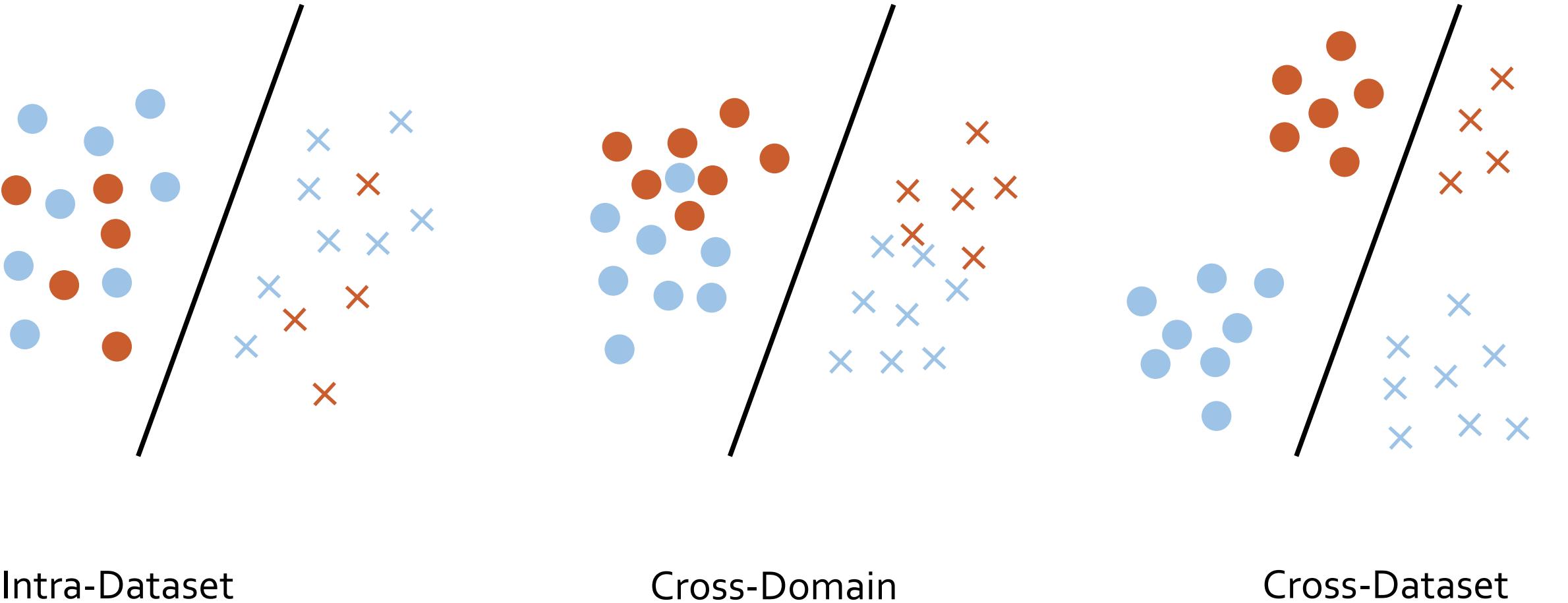
- AENet<sub>C,S</sub>
- AENet<sub>C,SW/oS<sup>s</sup></sub>
- - - AENet<sub>C,SW/oS<sup>f</sup></sub>



Semantic Attribute Matters

# Benchmark

# Benchmarks



# Benchmarks

## Intra-dataset Benchmark

Model	Backbone	Parm. (MB)	Recall (%)↑			AUC↑	EER (%)↓	APCER (%)↓	BPCER (%)↓	ACER (%)↓
			FPR = 1%	FPR = 0.5%	FPR = 0.1%					
Auxiliary* [23]	-	22.1	97.3	95.2	83.2	0.9972	1.2	5.71	1.41	3.56
BASN [16]	VGG16	569.7	98.9	<b>97.8</b>	<b>90.9</b>	<b>0.9991</b>	1.1	4.0	1.1	2.6
AENet <sub>c,s,g</sub>	ResNet-18	42.7	<b>98.9</b>	97.3	87.3	0.9989	<b>0.9</b>	<b>2.29</b>	<b>0.96</b>	<b>1.63</b>

## Cross-Domain Benchmark

Protocol	Model	Recall (%)↑			AUC↑	EER (%)↓	APCER (%)↓	BPCER (%)↓	ACER (%)↓
		FPR = 1%	FPR = 0.5%	FPR = 0.1%					
1	Baseline	93.7	86.9	69.6	<b>0.996</b>	2.5	5.7	2.52	4.11
	AENet <sub>c,g</sub>	93.3	88.6	<b>74.0</b>	0.994	2.5	5.28	2.41	3.85
	AENet <sub>c,s</sub>	93.4	89.3	71.3	<b>0.996</b>	2.4	5.63	2.42	4.04
2	AENet <sub>c,s,g</sub>	<b>95.0</b>	<b>91.4</b>	73.6	0.995	<b>2.1</b>	<b>4.09</b>	<b>2.09</b>	<b>3.09</b>
	Baseline	#	#	#	0.998±0.002	1.5±0.8	8.53±2.6	1.56±0.81	5.05±1.42
	AENet <sub>c,g</sub>	#	#	#	0.995±0.003	1.6±4.5	8.95±1.07	1.67±0.9	5.31±0.95
	AENet <sub>c,s</sub>	#	#	#	0.997±0.002	1.2±0.7	<b>4.01±2.9</b>	1.24±0.67	3.96±1.79
	AENet <sub>c,s,g</sub>	#	#	#	0.998±0.002	1.3±0.7	4.94±3.42	1.24±0.73	3.09±2.08

## Cross-Dataset Benchmark

Model	Training	Testing	HTER (%) ↓
FAS-TD-SF [34]	SiW	CASIA-MFSD	39.4
FAS-TD-SF [34]	CASIA-SURF	CASIA-MFSD	37.3
AENet <sub>c,s,g</sub>	SiW	CASIA-MFSD	<b>27.6</b>
Baseline	CelebA-Spoof	CASIA-MFSD	14.3
AENet <sub>c,g</sub>	CelebA-Spoof	CASIA-MFSD	14.1
AENet <sub>c,s</sub>	CelebA-Spoof	CASIA-MFSD	12.1
AENet <sub>c,s,g</sub>	CelebA-Spoof	CASIA-MFSD	<b>11.9</b>

# CelebA-Spoof

# Large-Scale







# Collection Dimensions

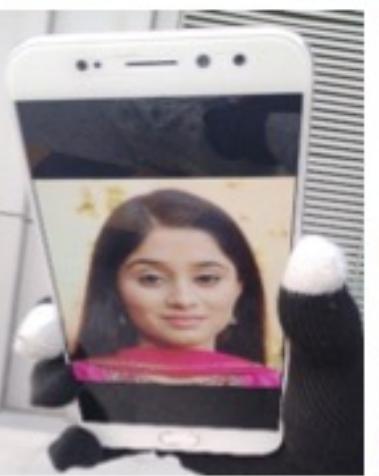
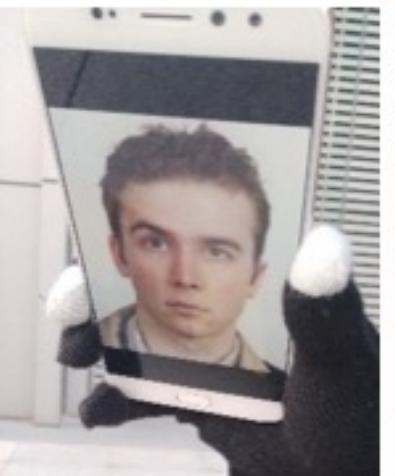
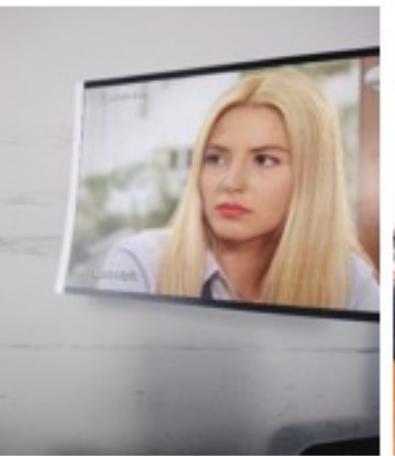
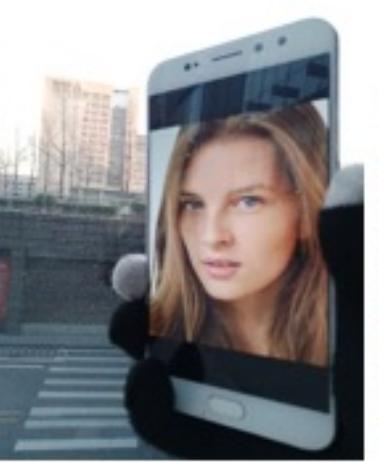
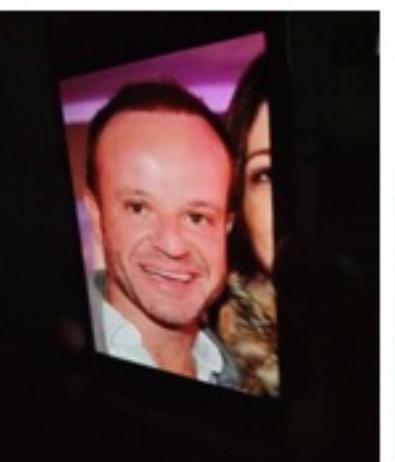
# Collection Dimensions

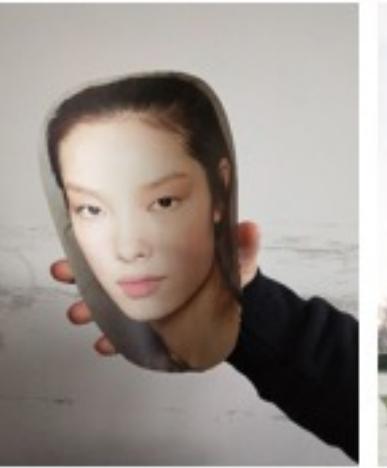
---

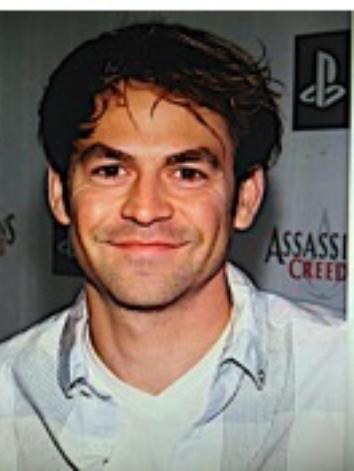
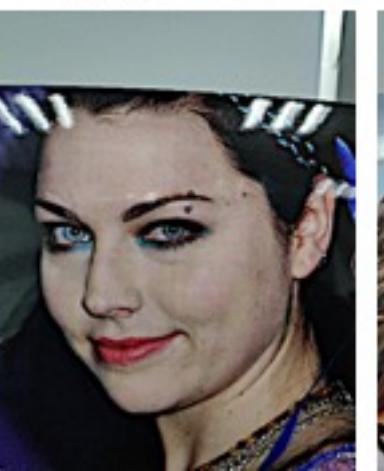
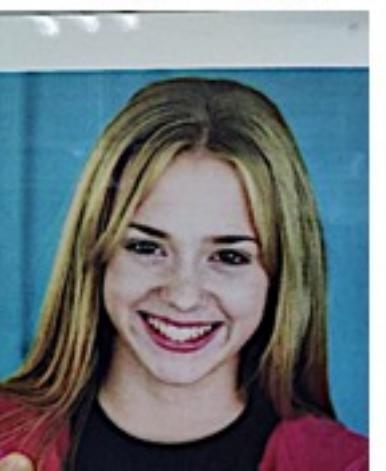
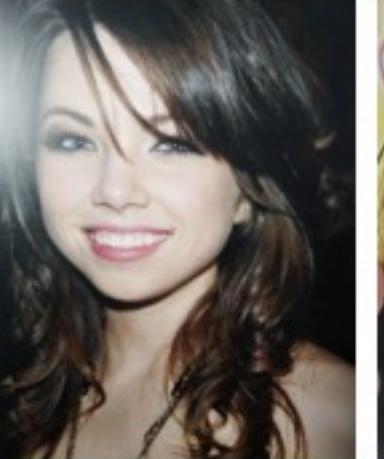
Angle

Input Sensor

Shape



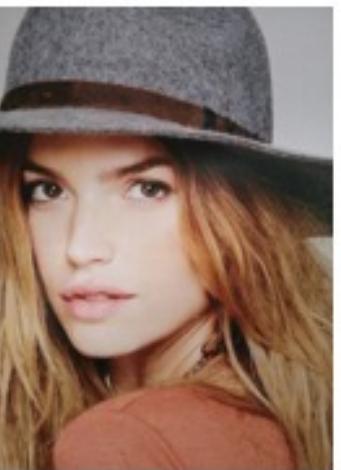


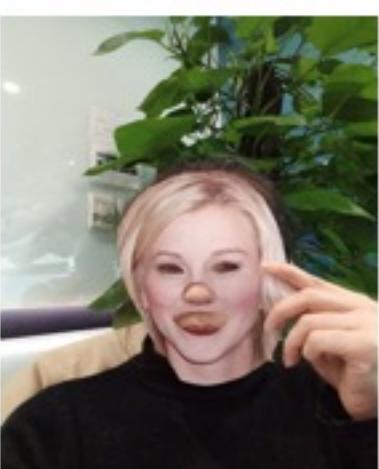
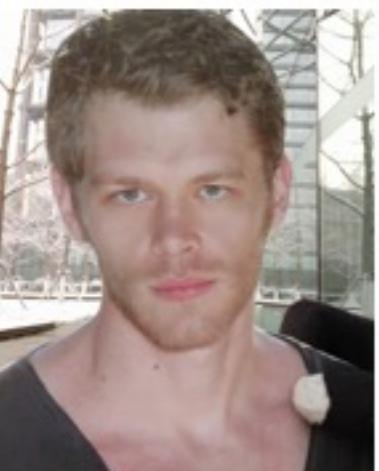
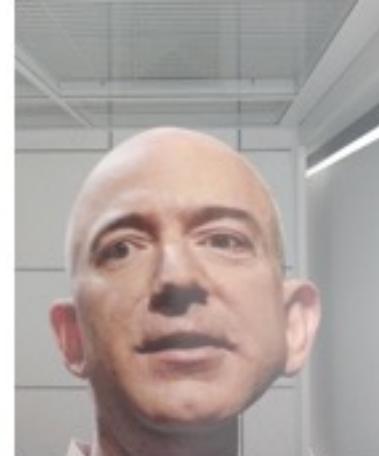


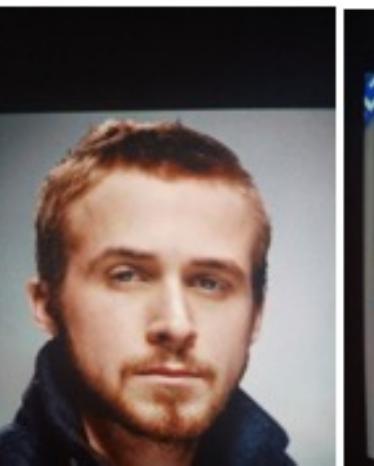
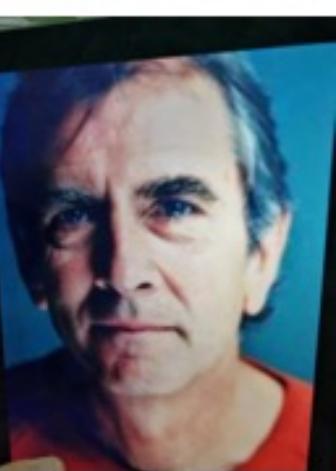
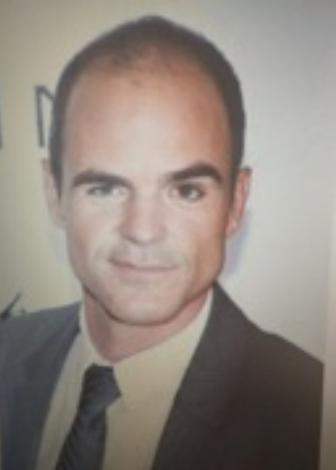
# Rich Annotations

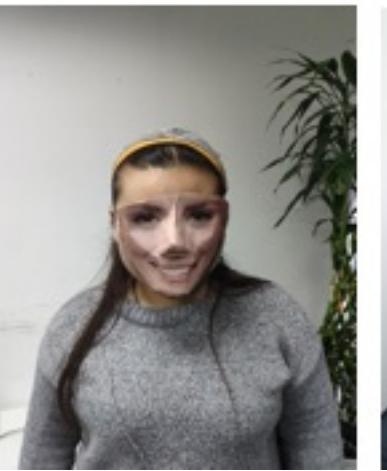
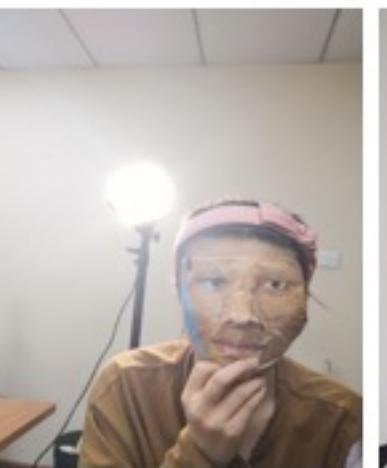
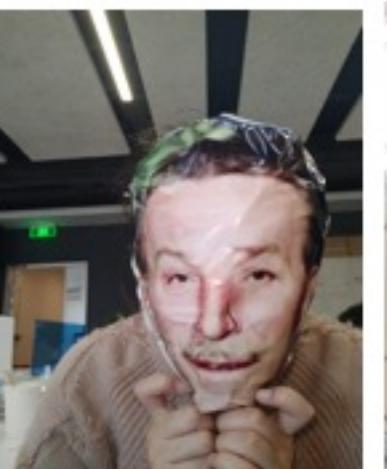
# Rich Annotations

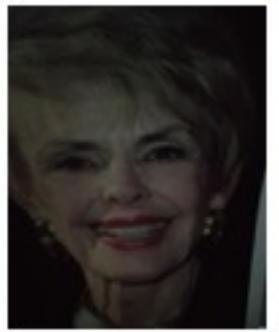
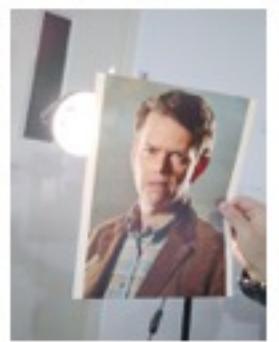
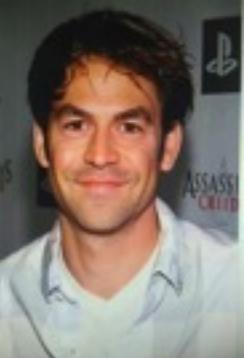
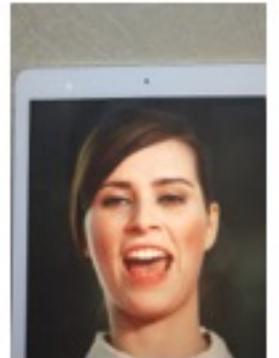
Illumination Condition	Spoof Type	Environment
Normal Strong Back Dark	Photo Poster A4 Face Mask Upper Body Mask Region Mask PC Tablet Phone 3D Mask	Indoor Outdoor













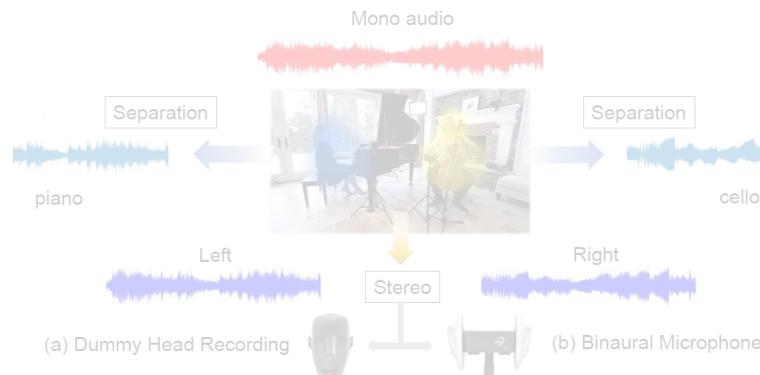


**Github Page:** <https://github.com/Davidzhangyuanhan/CelebA-Spoof>

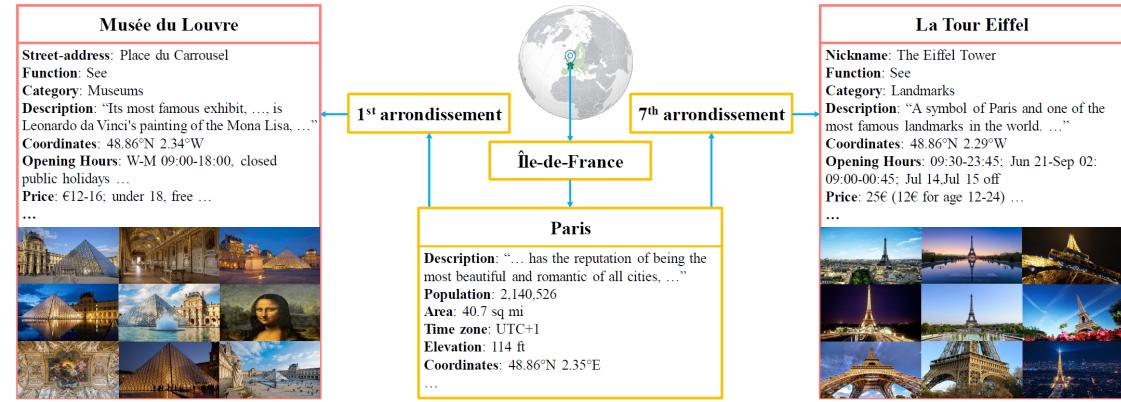




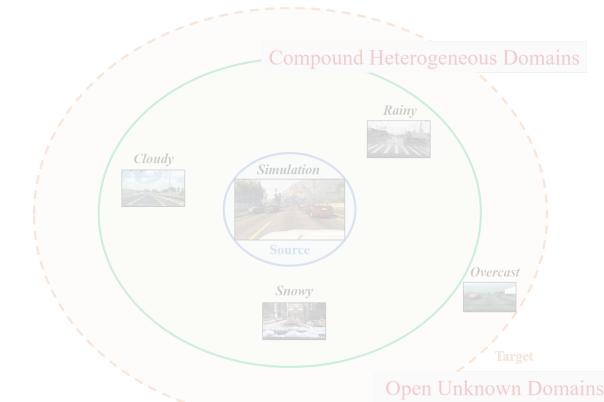
## Robust Sensing CelebA-Spoof



## Synthesizing across Modalities Sep-Stereo



## Understanding beyond Recognition Placepedia



## Open World Learning Open Compound Domain Adaptation



# Placepedia: Comprehensive Place Understanding with Multi-Faceted Annotations

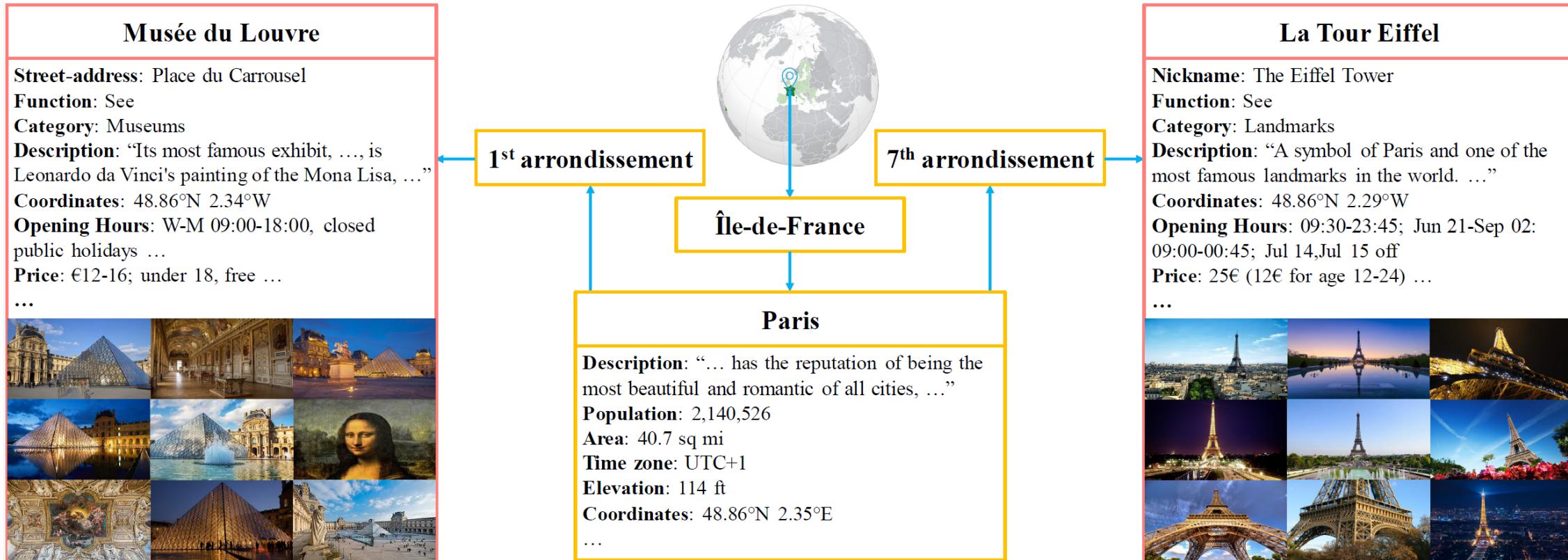
Huaiyi Huang , Yuqi Zhang, Qingqiu Huang, Zhengkui Guo, Ziwei Liu, Dahua Lin

The Chinese University of Hong Kong

# Dataset Overview

- Place is an important element in visual understanding.
  - Functionality
  - Cultural style
  - Economic type
- Comprehensive place understanding
  - Far beyond categorizing a place with an image
  - Requires information of multiple aspects

# Dataset Overview



# Get Very Rich Data of Place

- Wikivoyage!
- 360K places of 25K cities
- Lots of meta data

North America > United States of America > California > Bay Area > San Francisco > San Francisco/Chinatown–North Beach



Understand · Get in · Get around · See · Do · Buy · Eat · Drink · Sleep · Connect · Stay safe · Go next

See [edit] [add listing]

The listings in this article are geographically organized in roughly a south-to-north direction; meaning that they start with Chinatown firs

- Stockton Street Produce Markets. Stockton St (runs parallel to Grant Ave, one block west — between Sacramento St and Vallejo St). concentration of Chinese shops and Chinese shoppers can be found in the three blocks from Washington to Broadway. They are note Tangerines are important during Chinese New Years. You may need a gut check as well in the live produce markets — there are all ki on weekdays; weekends are even more crowded, when Chinese families that have moved up to the suburbs return for shopping on St 6PM, but the eateries will remain open into the evening hours. [edit](#)
- Chinatown Alleys. Though Grant Avenue has a lot to offer, it is quite touristy; thus, it is essential that you examine the more authentic Lane, and Ross Alley, between Grant and Stockton. Ross Alley is the oldest alley in the city and many movies have had scenes shot here got a real old-world feel and you will hear Cantonese conversations and the clicking sound of mahjong tiles being shuffled. [edit](#)
- 1 Golden Gate Fortune Cookie Factory, 56 Ross Alley (*between Jackson St, Washington St, Stockton St and Grant Ave*), ☎ +1-415-462-1415— produces more than 20,000 fortune cookies a day. The factory is in a small alley and it is tiny with only 3 people making fortune cookies. (bended) fortune cookie sample but photos cost 50 cents and the moment you walk in they are asking you in their broken English what a huge bag of cookies. [edit](#)
- 2 City Lights Bookstore, 261 Columbus Ave (*at Broadway St*), ☎ +1-415-362-8193, fax: +1-415-362-4921. 10AM–midnight daily. of the centers of the Beat community in the 1950s. It's iconic and has become synonymous with the literary Beat movement. Oh, don't prose and poetry. Why not buy a copy of *On the Road* while you're there — you won't find a better place to get it! [edit](#)
- 3 Jack Kerouac Alley, Jack Kerouac Alley (*at Columbus Ave and Broadway St*). This tiny paved pedestrian alley was named after the author of *On the Road*. The alley is a favorite of tourists and locals alike. It was intended to form a literary (and actual) connection between the communities of Chinatown and North Beach. The Chinese and Western poems from Kerouac, Confucius and John Steinbeck among others. [edit](#)
- 4 Telegraph Hill. Telegraph Hill earned its name in the days of the Gold Rush when it was used as a signaling post to relay messages in 1853. It rewards a weary traveler with some wonderful views over the city. Over time a quiet residential neighborhood built up along the hill. There is something to admire on your way up or down. Other neighbors include a colony of colorful feral parrots, predominantly red-masked parrots. One can drive to the top, but it's better to take one of the narrow steps leading up and down the sides of the hill (including the Filbert Steps). [edit](#)
- 5 Filbert Steps. The Filbert Steps are the part of Filbert Street that runs between Battery Street and Telegraph Hill Boulevard in North Beach. They are though somewhat strenuous — route for visitors of the tower. In fact, following the steps is at times faster than driving to Coit Tower. Visitors of the steps will see public gardens, stylish homes and views of North Beach and the bay; if a path is not gated or sparsely populated, it's easy to explore. Also, it pays to be adventurous, as some of the best gardens and views are off the stairs. Finally, there is more than one way up and down the hill, so consider taking a different route on the return leg. Just avoid private property. [edit](#)

Museums and galleries [edit] [add listing]

- 6 Chinese Culture Center, 750 Kearny St, 3rd floor (*From Portsmouth Sq: just walk across the footbridge that crosses Kearny St*). Info@c-c-c.org. Tu-Sa 10AM–4PM. The center was established in 1973 in order to promote understanding of Chinese and Chinese American culture through changing Chinese art exhibitions. Free. [edit](#)

# Get Very Rich Data of Place

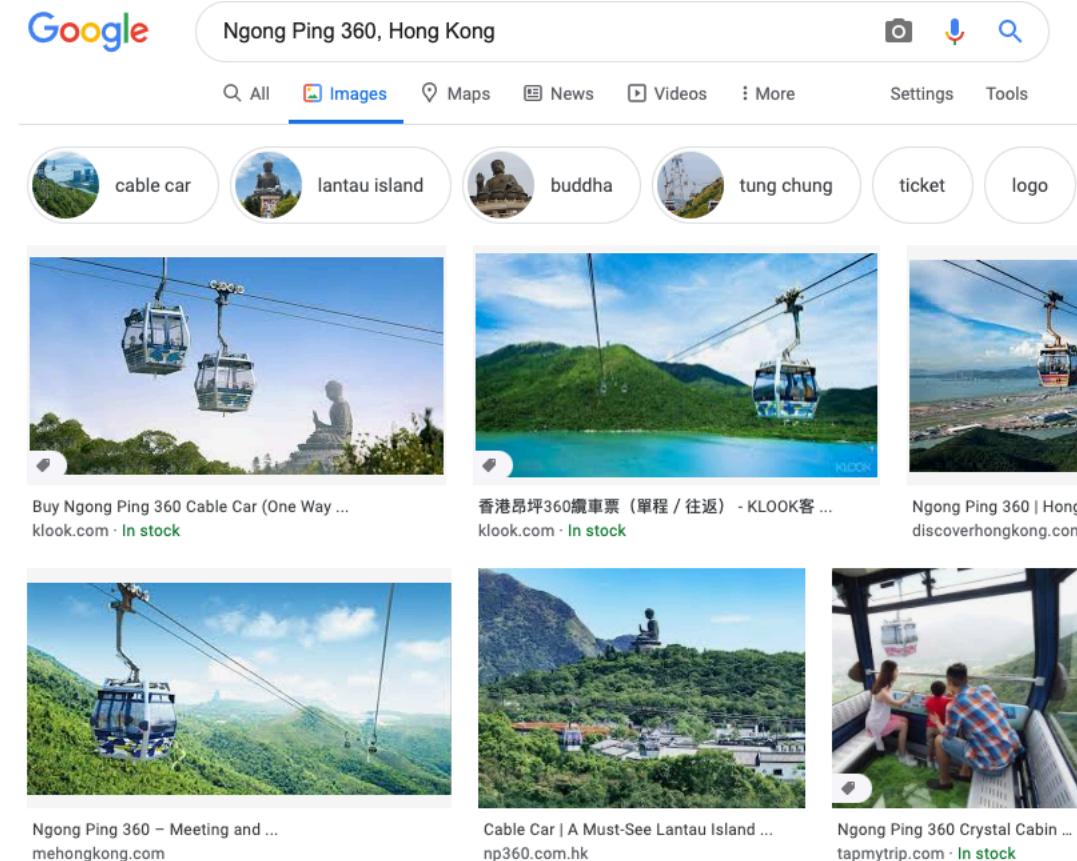
- Some are not place. Refine the place list!
  - It has the attribute GPS coordinates or address
  - It is identified as a location by Google Entity Recognition [1] or Stanford Entity Recognition [2]
    - 1 Grant Plaza Hotel<sup>1</sup>, 465 Grant Ave (*at Pine St*), ☎ +1-415-434-3883, fax: +1-412 night stay. Rates are reasonable. Make sure you ask for one of the outside rooms → Is a place.
    - Chinese New Year Festivities<sup>2</sup>. Jan or Feb. Celebrated for over 5,000 years, firecrackers, "lucky-money" envelopes, colorful banners, over 100 ornately the → Is not a place.
- Remains 320K places

[1] <https://cloud.google.com/natural-language>

[2] <https://nlp.stanford.edu/software/CRF-NER.html>

# Get Very Rich Data of Place

- Google Image open source
- 240K places
- 35M images



# Challenges



1) Daytime to nighttime

2) Different angles

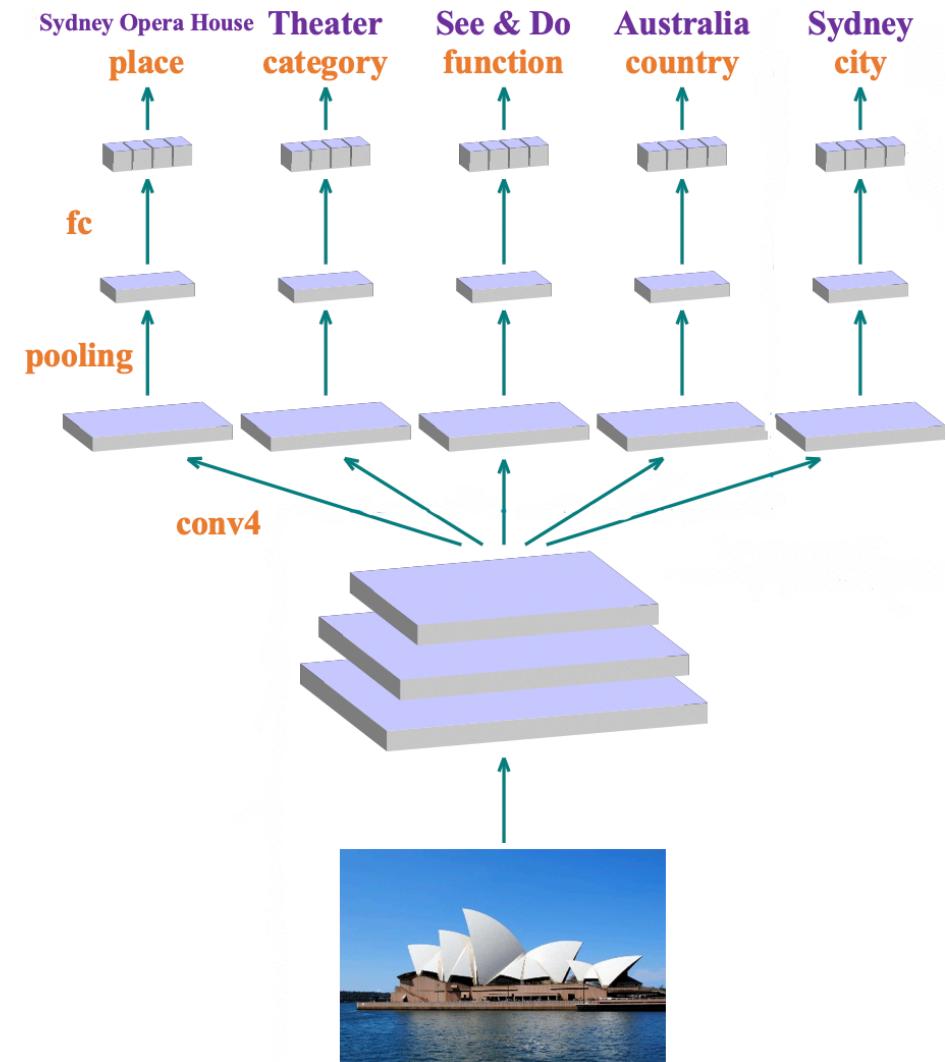
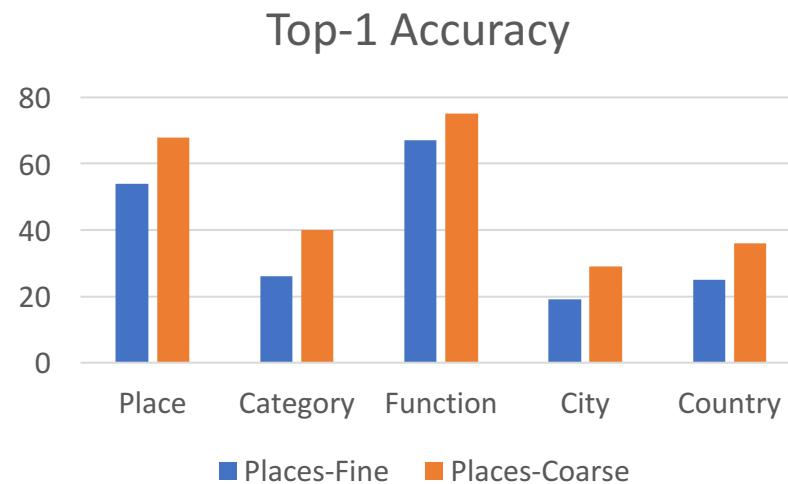
3) Inside and Outside

# Comprehensive Place Understanding

- Benchmarks
  - Datasets:
    - Places-Coarse: 26K places
    - Places-Fine: 1K places
  - Tasks:
    - Place Retrieval (determine if two images belong to the same place)
    - Place Categorization (classify places into categories like museums, parks, churches, and temples)
    - Function Categorization (classify places by their functionality such as eat, sleep, see, buy, and so on)
    - City/Country Recognition (classify places into their cities or countries)

# A unified Framework to Predict All Tasks

- Duplicate the last convolution/pooling/fc layers of ResNet50 to five branches



# Findings

Ground Truth

Parks  
Franklin Park



Tombs



Eat



Buy



Parks  
Parque México



Museums



Buy



Parks  
Parque Ecológico do Tietê



Pubs



Drink

Category Recognition  
Place Retrieval

Category Recognition

Function Categorization

Ground  
Truth

Florence



Florence



Florence

Beijing



Beijing



Beijing

Oslo



Singapore



Taipei



Toronto



Tokyo



Milan



Shanghai



Paris



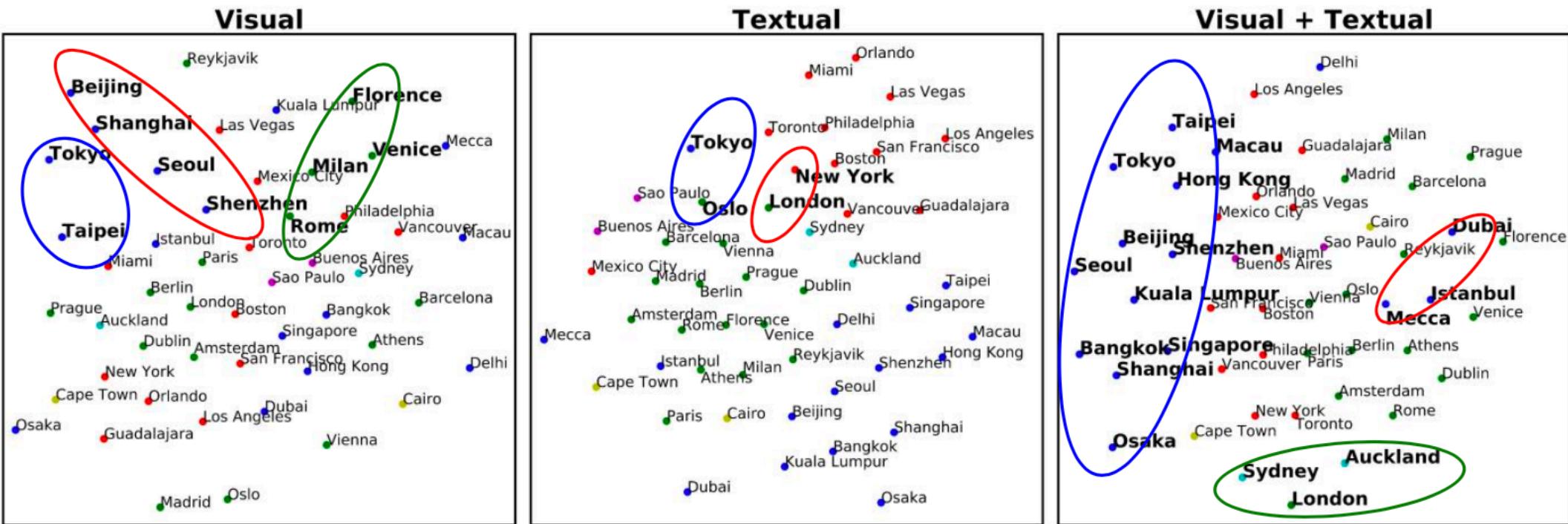
Seoul

City  
Recognition

# City Embedding

- For vision:
  - Using place images
  - Extract the feature from the city recognition model
- For text:
  - Using city descriptions
  - Embed the content of texts into numeric space based on Bert pre-defined model
- For both vision & text: concatenate two vectors above

# City Embedding Visualization



# City Embedding

- City description
- Calculate the weights
  - Economic
  - Cultural
  - Political
- as in [1]
- Pearson correlation
- Compare with neuron

**Economy**



**Culture**



**Politics**



[1] Son, J.S., Thill, J.C.: Is your city economic, cultural, or political? recognition of city image based on multidimensional scaling of quantified web pages. In: Spatial Analysis and Location Modeling in Urban and Regional Systems, pp. 63–95. Springer (2018)

# Conclusion

- A large-scale place dataset
  - comprehensively annotated with multiple aspects
- Explore place understanding
  - Build several benchmarks and study a unified model to recognize places
  - Remains lots of challenges
- Learn city embedding representations
  - Learning from both visual & textual domains can better characterize a city
  - Economic/cultural/political elements could be expressed in different types of images

# Placepedia: Comprehensive Place Understanding with Multi-Faceted Annotations

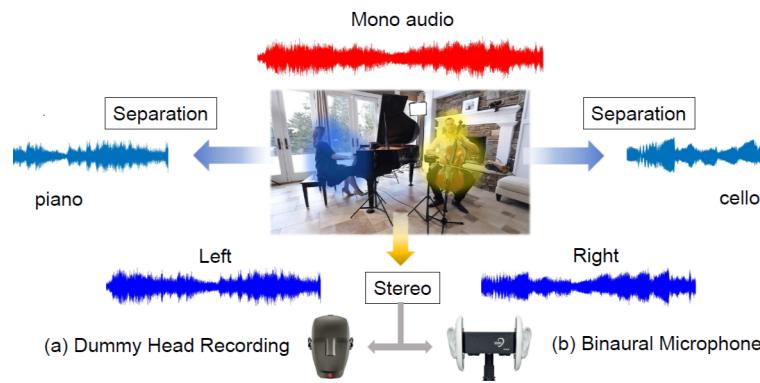
**Project page:** <https://hahehi.github.io/placepedia.html>

**Code and models:** <https://github.com/hahehi/placepedia>



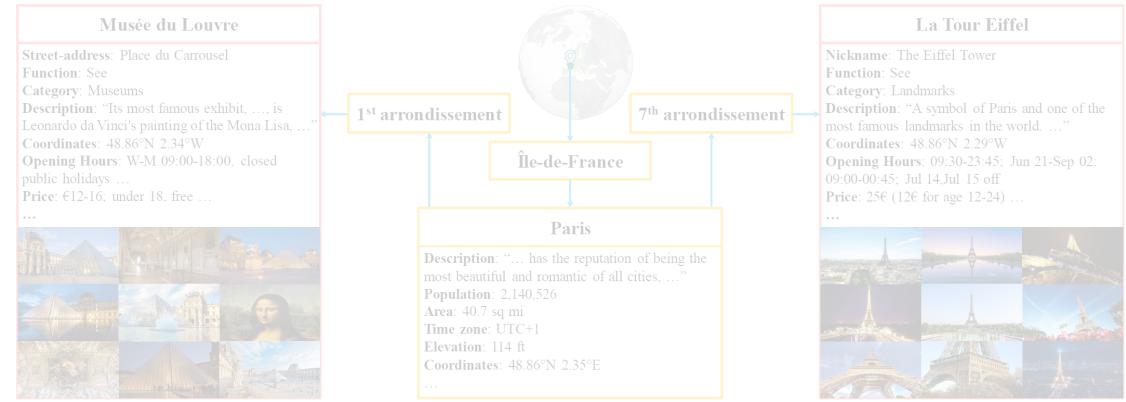
## Robust Sensing

### CelebA-Spoof



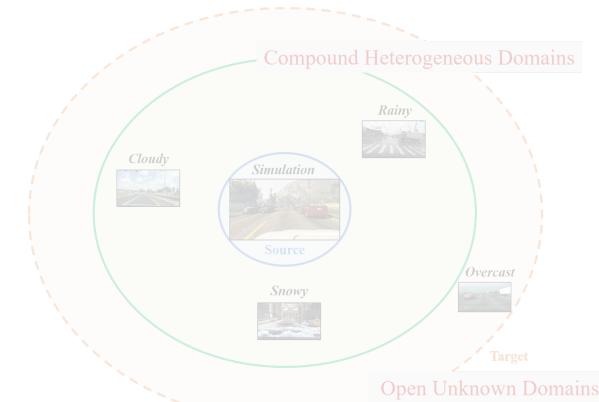
## Synthesizing across Modalities

### Sep-Stereo



## Understanding beyond Recognition

### Placepedia



## Open World Learning

### Open Compound Domain Adaptation



# Sep-Stereo: Visually Guided Stereophonic Audio Generation by Associating Source Separation

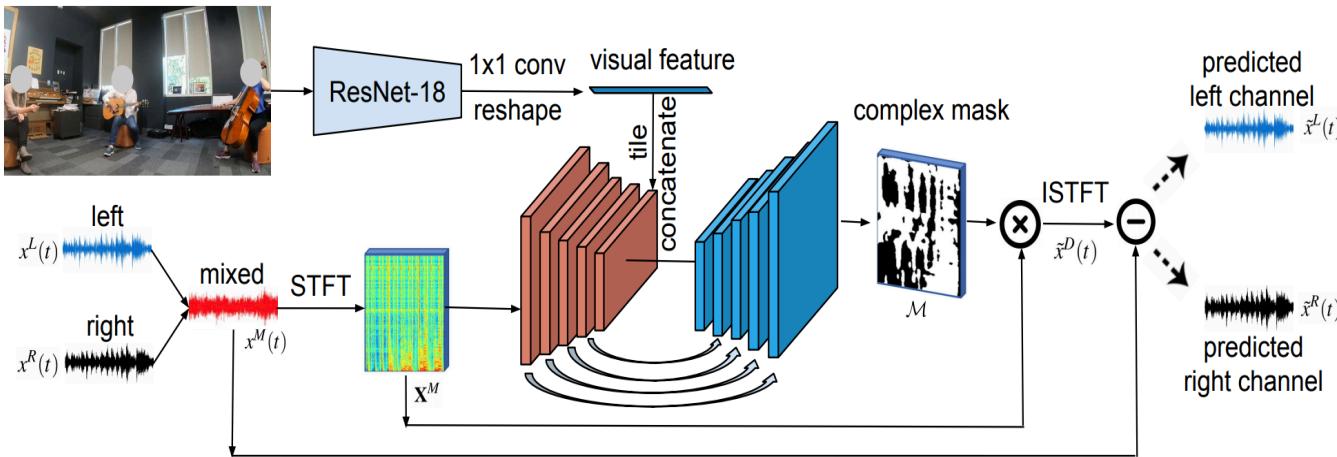


Hang Zhou\*, Xudong Xu\*, Dahua Lin, Xiaogang Wang, Ziwei Liu

CUHK – SenseTime Joint Lab, The Chinese University of Hong Kong

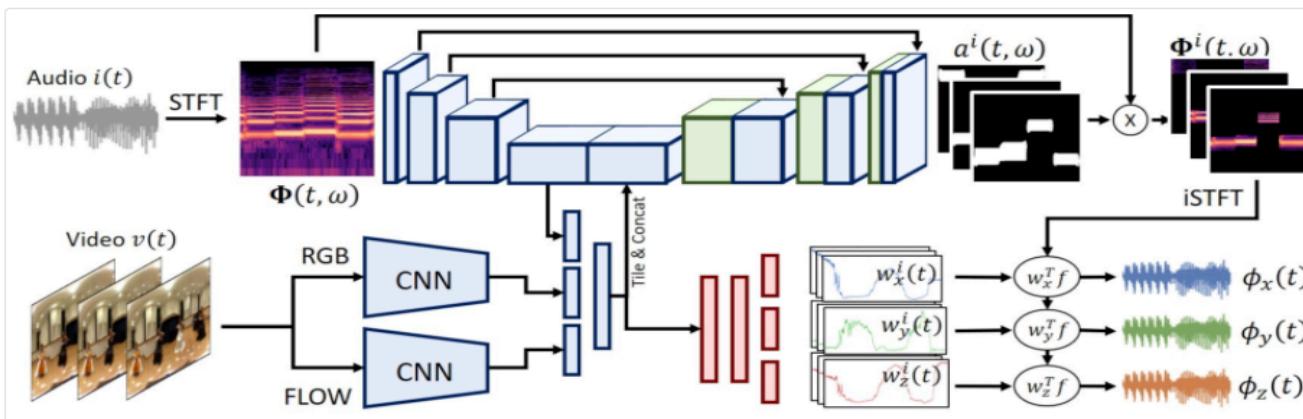
# Prior Works: Fully Supervised Training

2.5d Visual Sound (Gao et al, CVPR 2019)



Mono2Binaural relies on self-collected dataset, FAIR-Play, to train the network.

Spatial AudioGen (Morgado et al, NeurIPS 2018)



Spatial AudioGen exploits the stereophonic data on YouTube.

# Limitation: Stereophonic Data Collection



(a) Dummy Head Recording



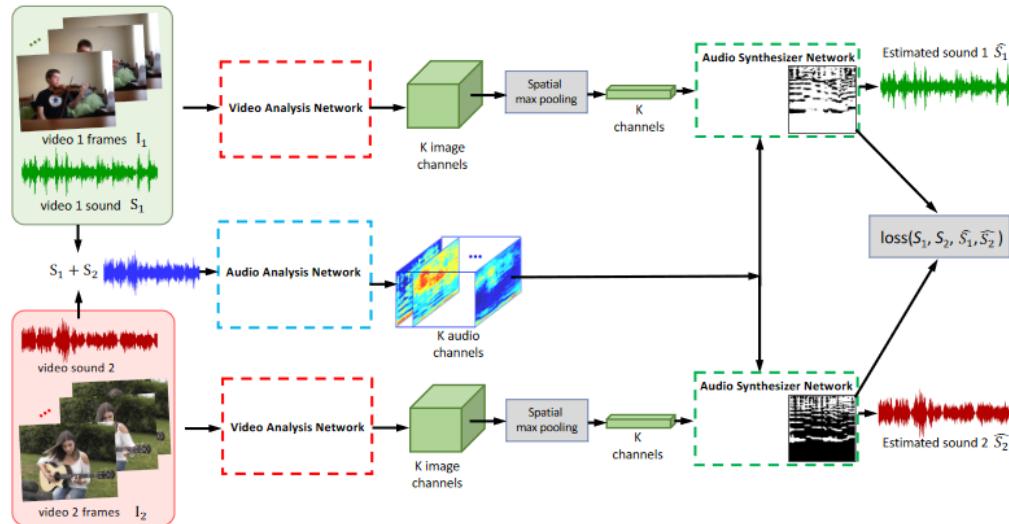
Data collection equipment in:  
2.5d Visual Sound



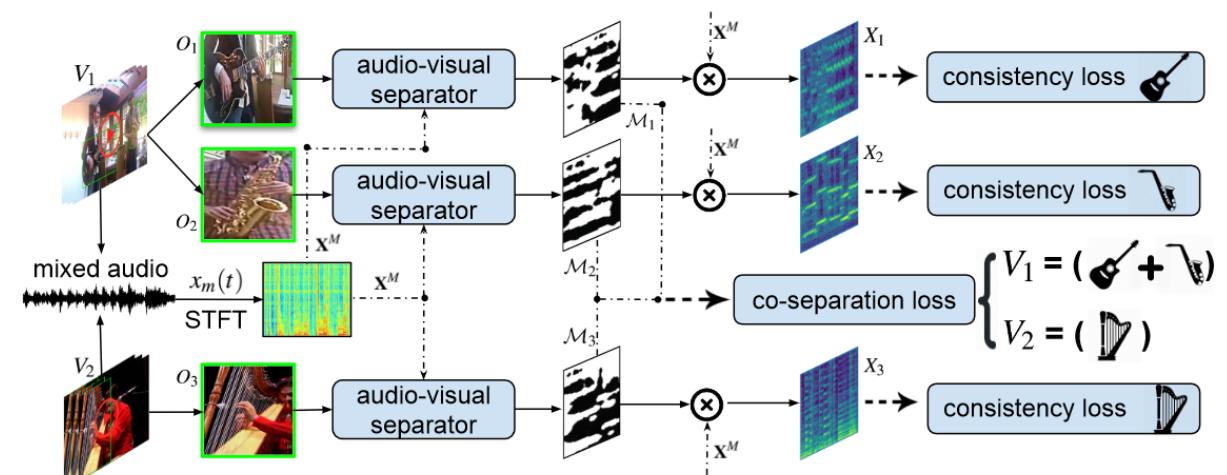
(b) Binaural Microphone

# Mono Audios in Source Separation

The Sound of Pixels  
(ECCV 2018)

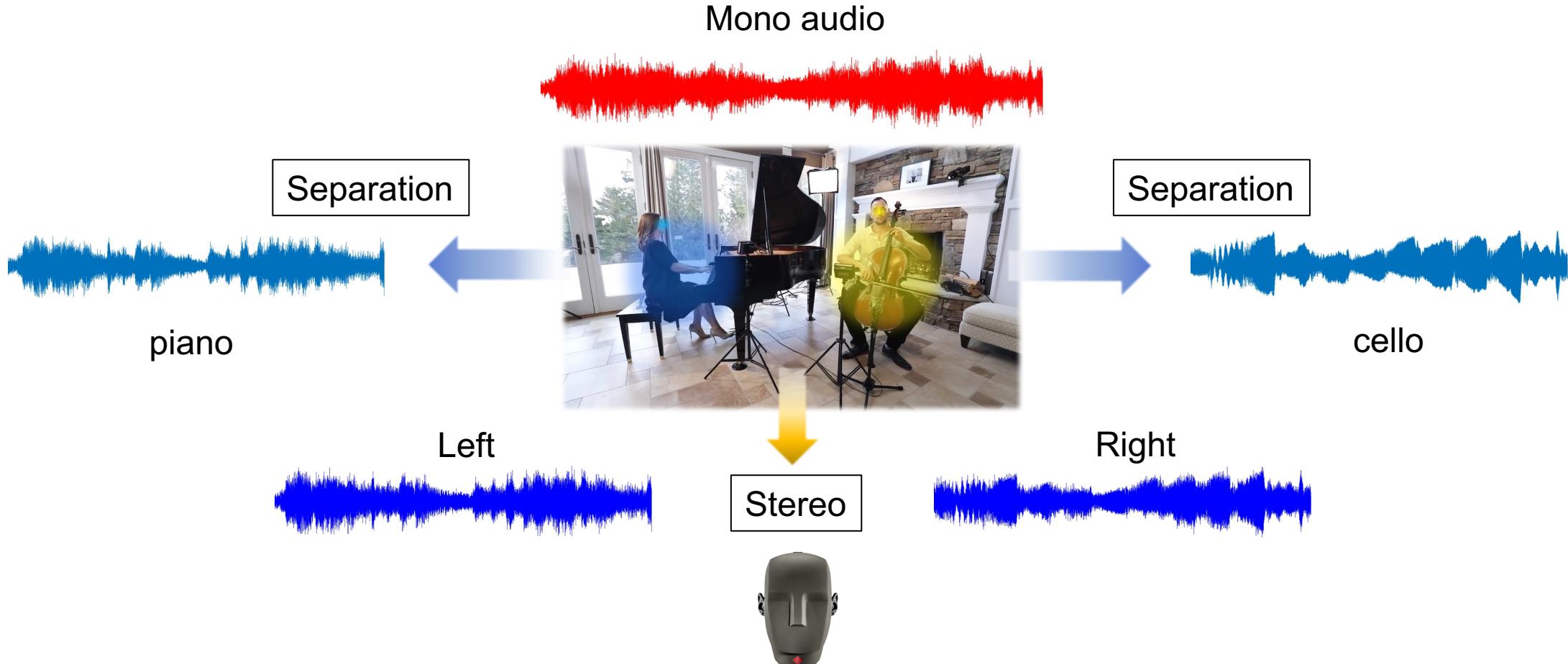


Co-Separating Sounds of Visual Objects  
(ICCV 2019)



- Massively availed mono audios have been successfully used in the field of source separation.

# Key Insight: Regard the problem of separating two audios as an extreme case of creating binaural audio

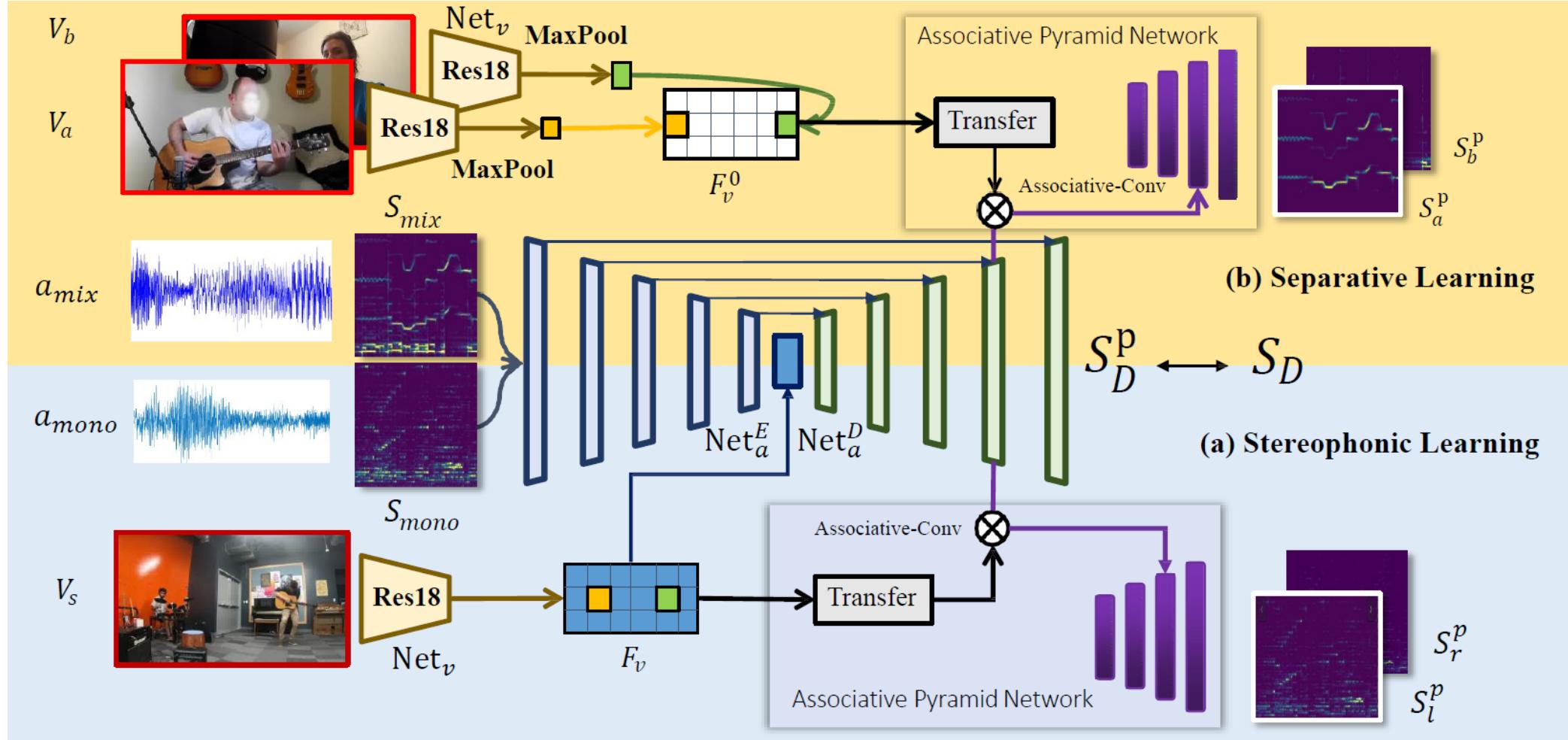


# Key Insight: Regard the problem of separating two audios as an extreme case of creating binaural audio

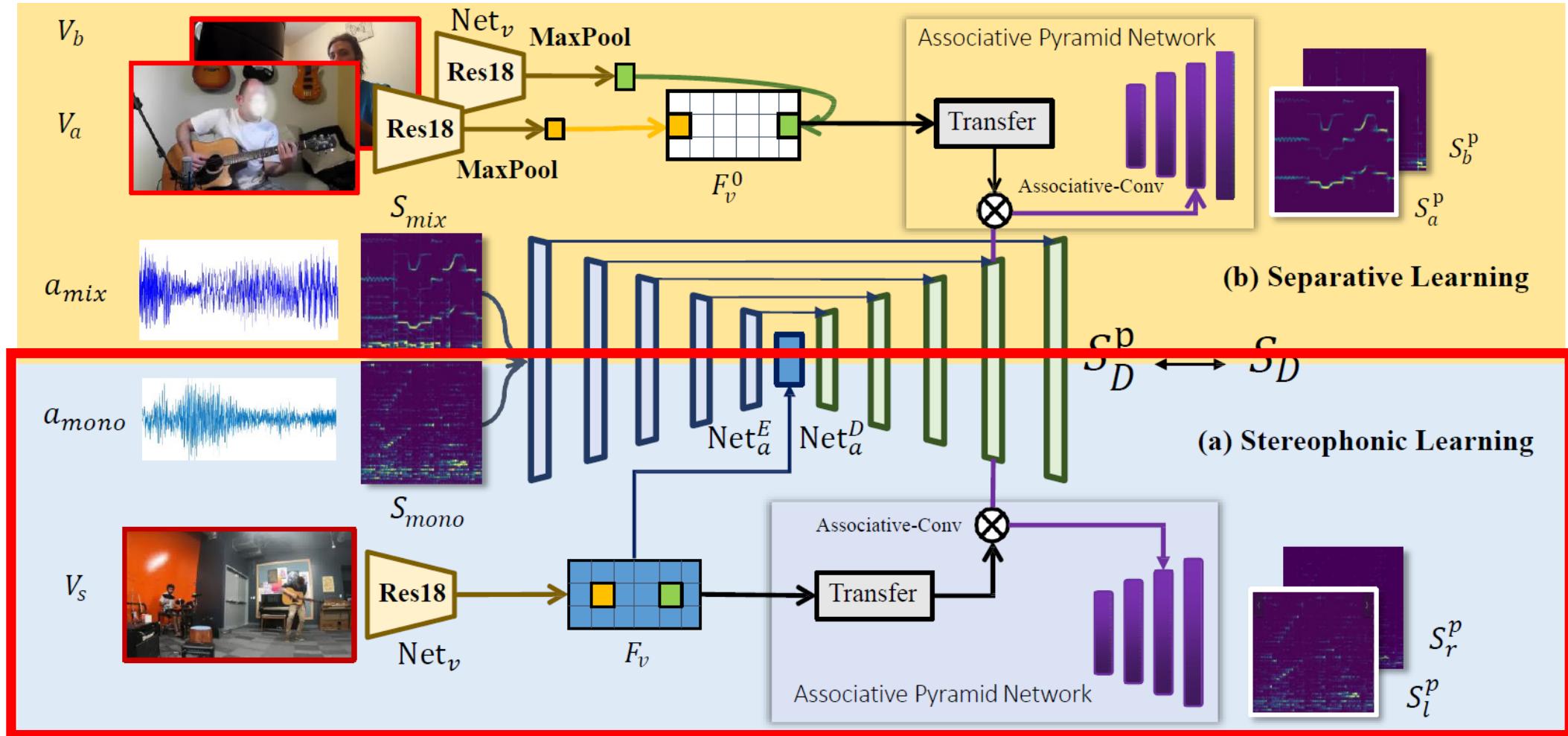


- Mono audios can be used to facilitate the generation of stereo audio.

# A Unified Framework: Sep-Stereo



# Stereophonic Learning

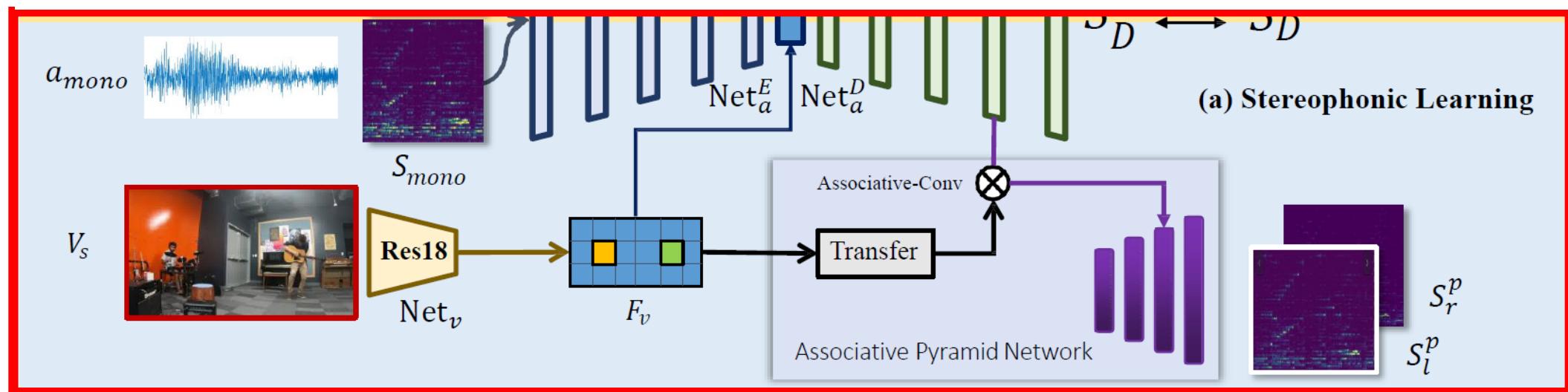


# Stereophonic Learning

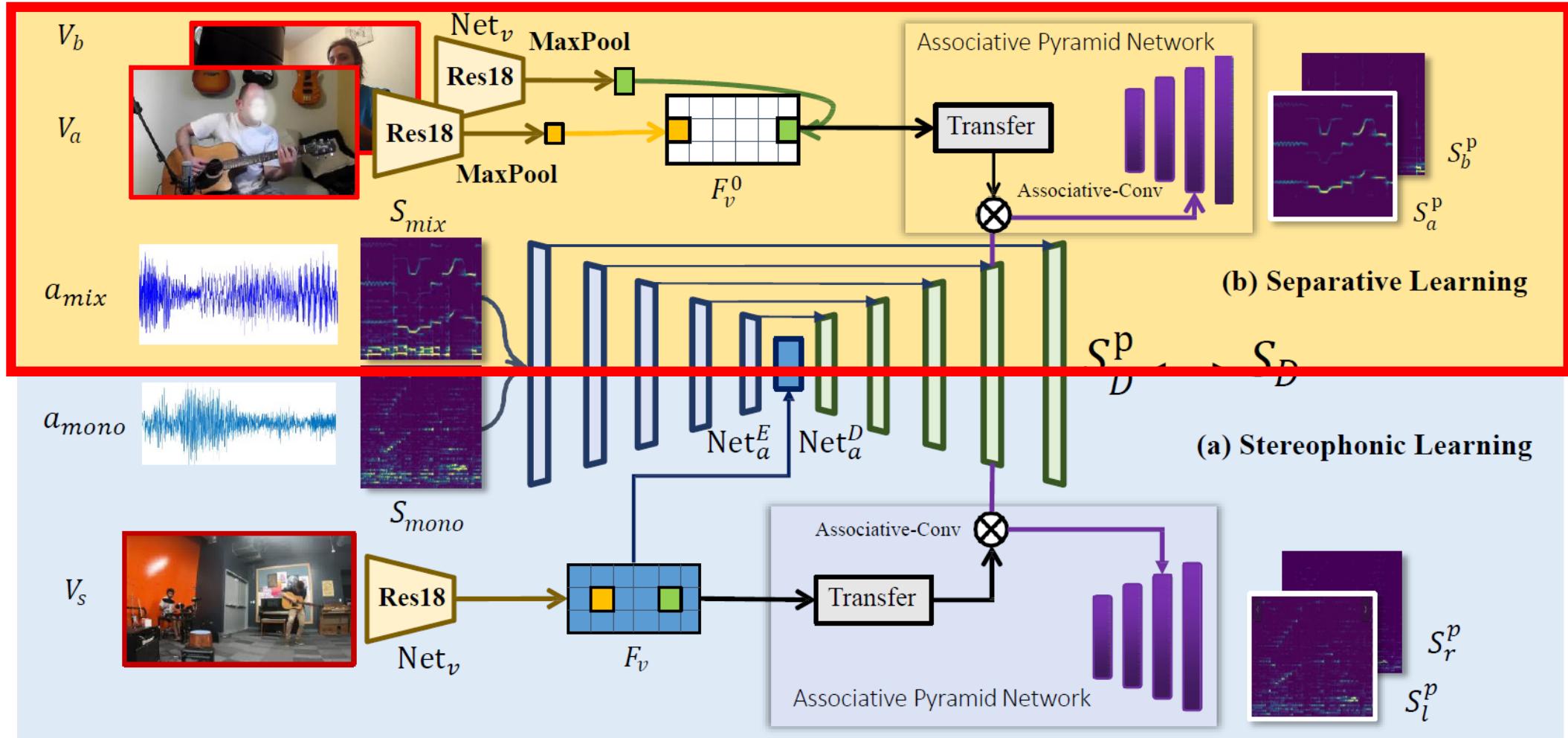
**Training Data Settings:** Same as 2.5D Visual Sound.

**Base Network:** Same as 2.5D Visual Sound.

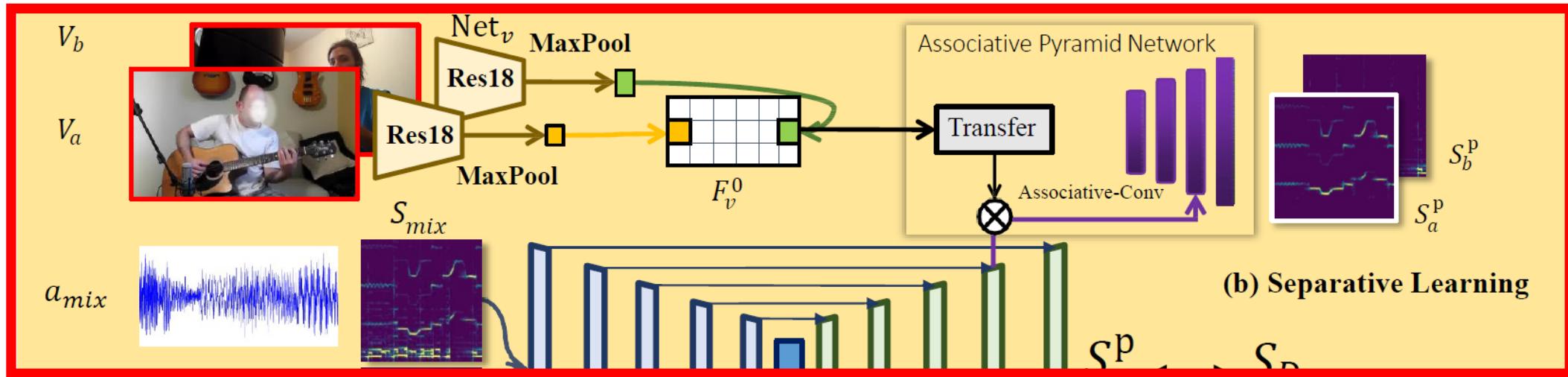
**Visual Feature:** Kept as 14x7 feature map.



# Separative Learning



# Separative Learning

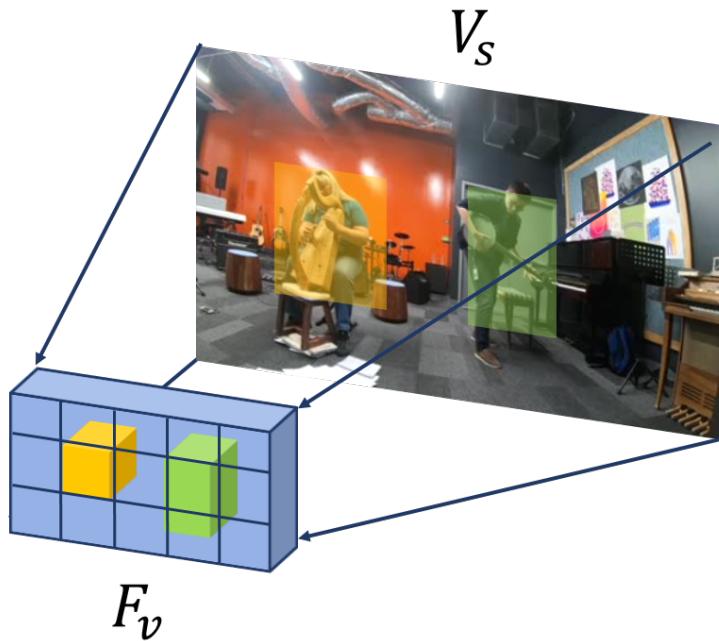


**Training Data Settings:** Same as Sound of Pixels.

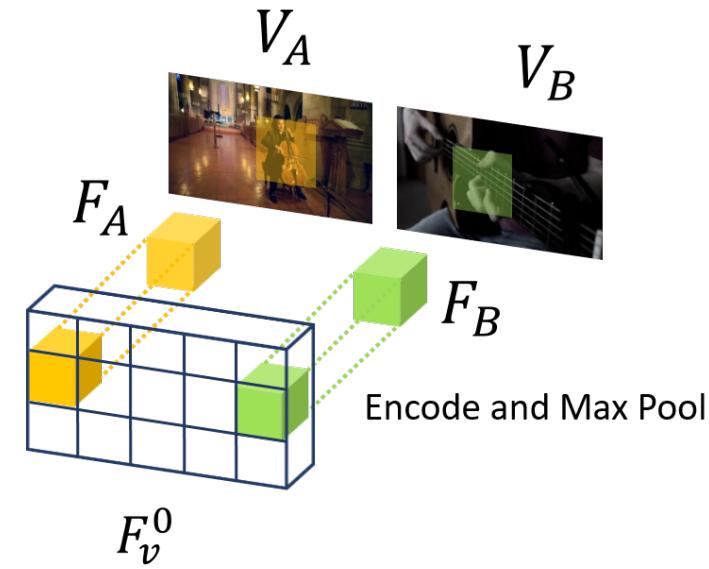
**Base Network:** Same as 2.5D Visual Sound.

**Visual Feature:** Max-pooled and rearranged.

# Difference: Visual Feature Rearrangement

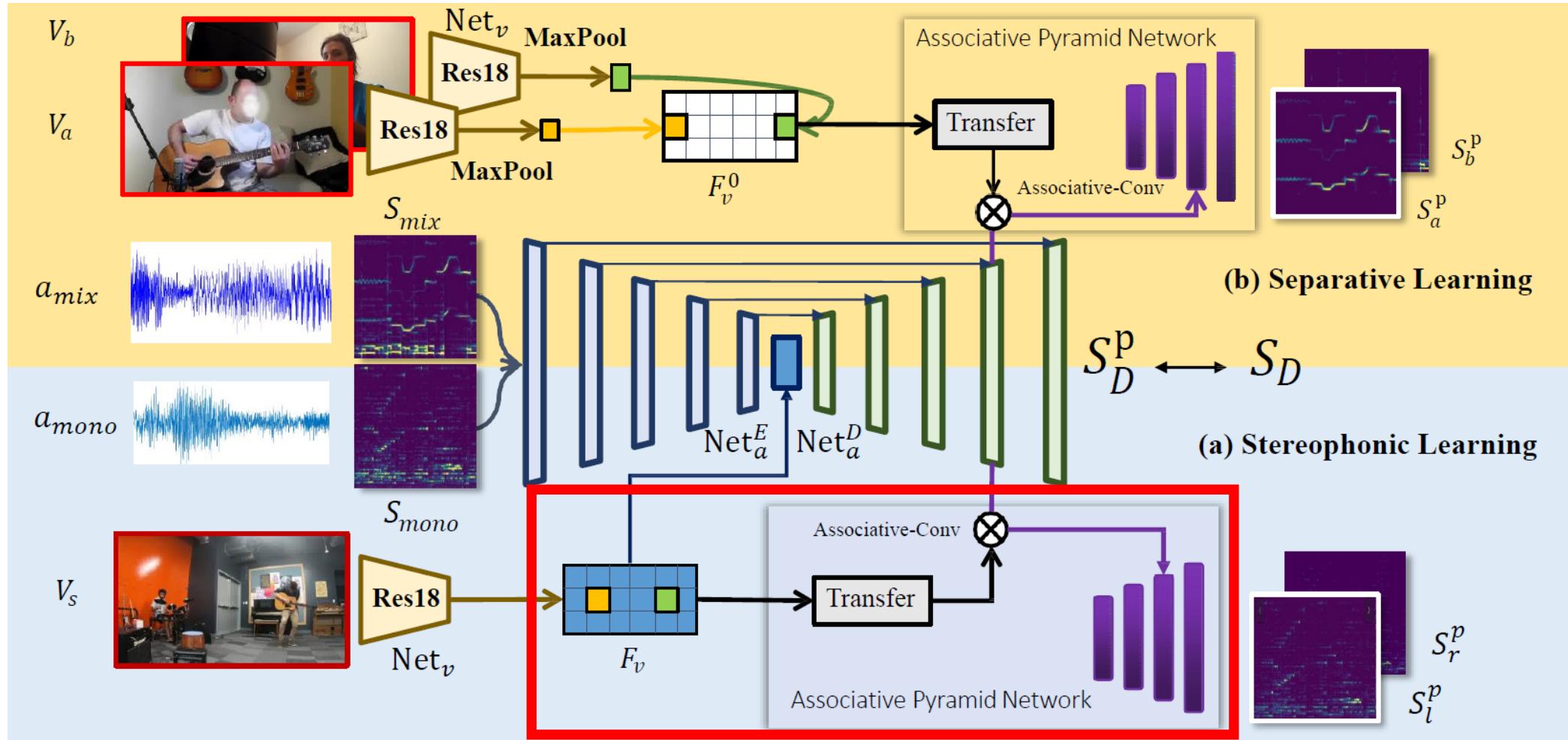


Stereophonic Learning



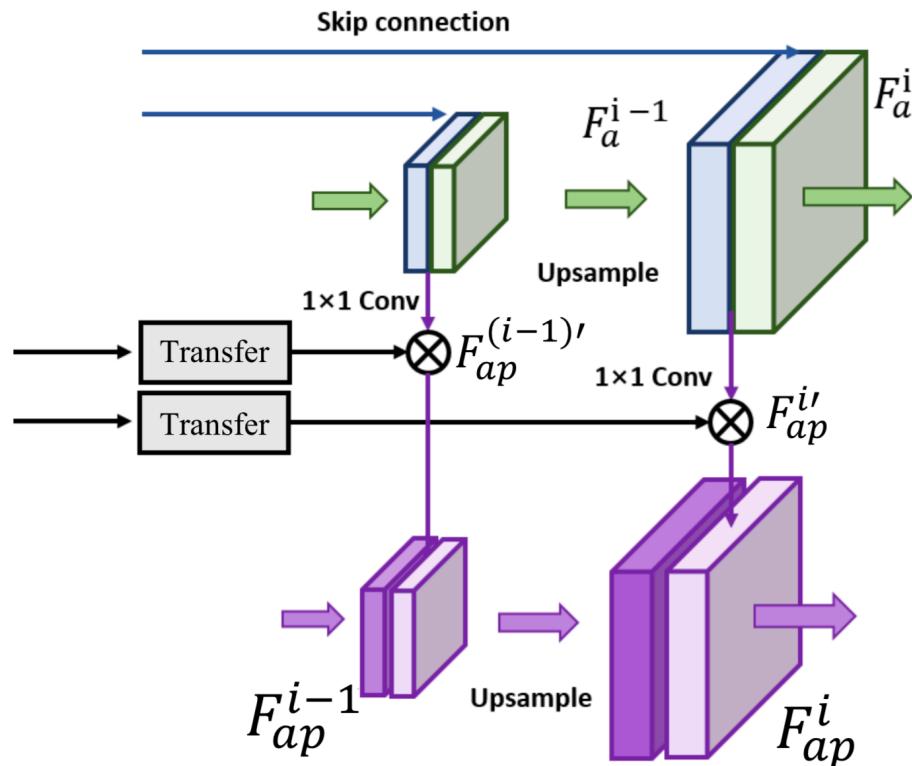
Separative Learning

# Associative Pyramid Network



# Associative Pyramid Network

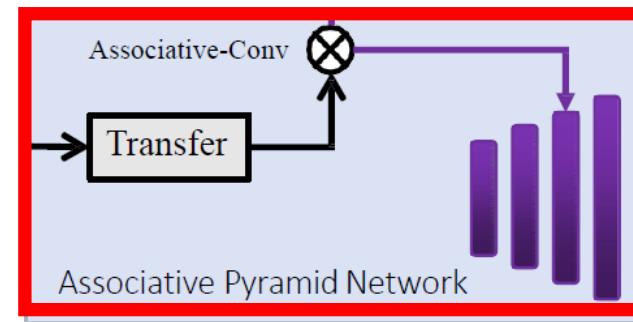
- Associative Pyramid Network better associates the visual features and the audio features with a learned Associative-Conv operation.



(a) APNet Structure

$$F_{ap}^{i'} = \text{Conv2d}_{K_v^i}(F_a^i)$$

$$F_{ap}^i = \text{Cat}([\text{DeConv}(F_{ap}^{i-1}), F_{ap}^{i'}])$$



# Mono audio helps stereo audio generation

- Extensive experiments demonstrate that Sep-Stereo can achieve better performance on the task of stereo audio generation with the help of mono audio data.

Method	Training Data		FAIR-Play		YT-Music	
	Stereo	Mono	$\text{STFT}_D$	$\text{ENV}_D$	$\text{STFT}_D$	$\text{ENV}_D$
Mono2Binaural	✓	✗	0.959	0.141	1.346	0.179
Baseline (MUSIC)	✓	✓	0.930	0.139	1.308	0.175
Associative-Conv	✓	✗	0.893	0.137	1.147	0.150
APNet	✓	✗	0.889	0.136	1.070	0.148
<b>Sep-Stereo (Ours)</b>	✓	✓	<b>0.879</b>	<b>0.135</b>	<b>1.051</b>	<b>0.145</b>

# Demo Results

For **stereo**, we compare our results with Mono2Binaural (model of 2.5D Visual Sound), we show the mono input, results from two models and the ground truth.

For **separation**, we compare with PixelPlayer (Sound of Pixels).

Results on MUSIC duets demonstrate the generalization of our method.

# Demo Results

Better watch with HIGH QUALITY  
earphones or headphones

# Future Work

- Separation:
  - Universal separation, tackling music and speech, even general sound with one model.
  - Adopt the ideas from state-of-the-art audio source separation for the pursuit of models with more capacity.
  - Exploring the task of visually guided audio generation and separation together.
- Stereo:
  - The problem of overfitting still remains unsolved.
  - How to incorporate the setting of the room into the generation of stereo.



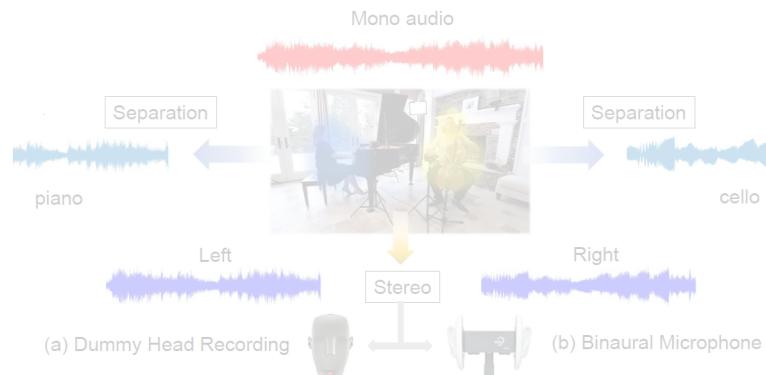
**Project page:** <https://hangz-nju-cuhk.github.io/projects/Sep-Stereo>

**Code and models:** [https://github.com/SheldonTsui/SepStereo\\_ECCV2020](https://github.com/SheldonTsui/SepStereo_ECCV2020)



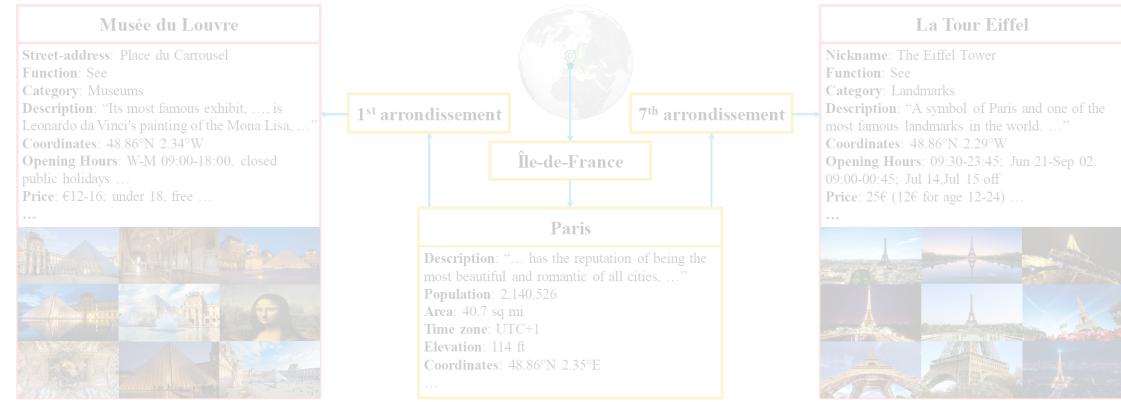
## Robust Sensing

CelebA-Spoof



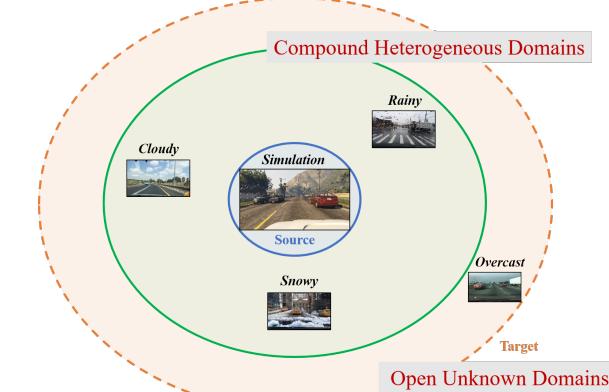
## Synthesizing across Modalities

Sep-Stereo



## Understanding beyond Recognition

Placepedia



## Open World Learning

Open Compound Domain Adaptation

# Open Compound Domain Adaptation

Ziwei Liu\* Zhongqi Miao\* Xingang Pan Xiaohang Zhan Dahua Lin Stella X. Yu Boqing Gong



The Chinese University of Hong Kong



UC Berkeley / ICSI



Google Inc.



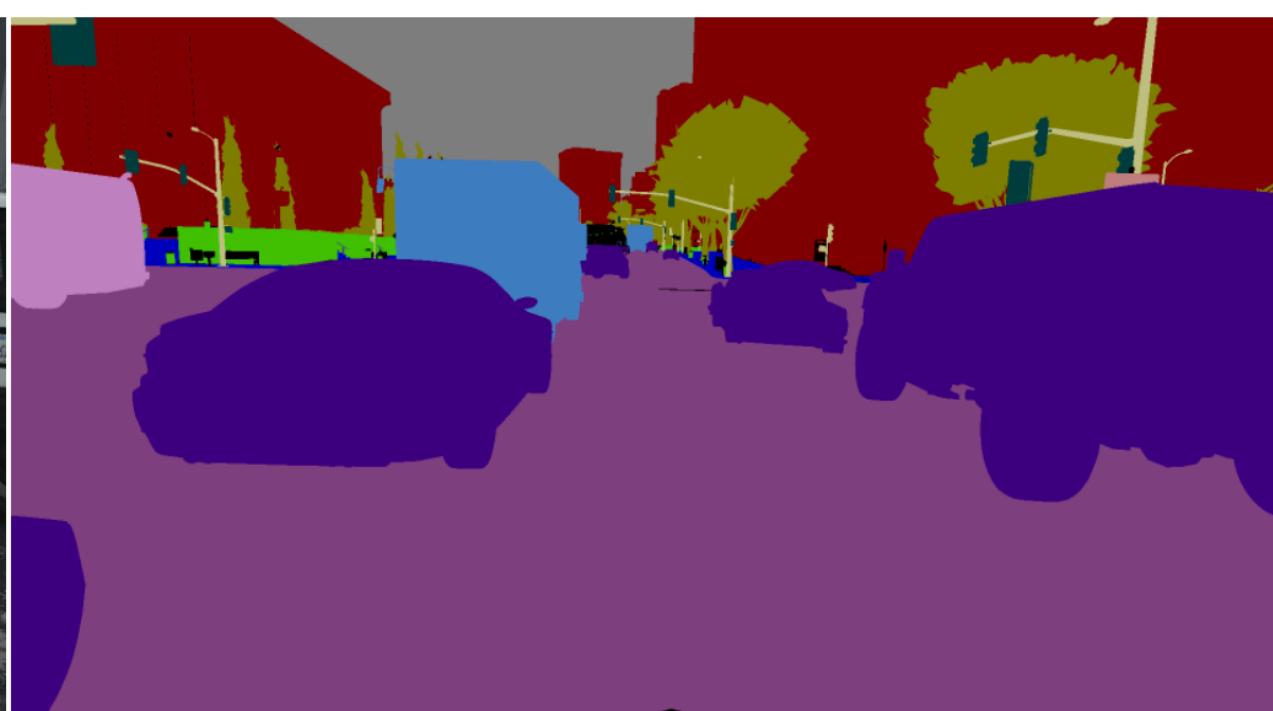
# Perception for Autonomous Driving

SELF DRIVING  
MODE

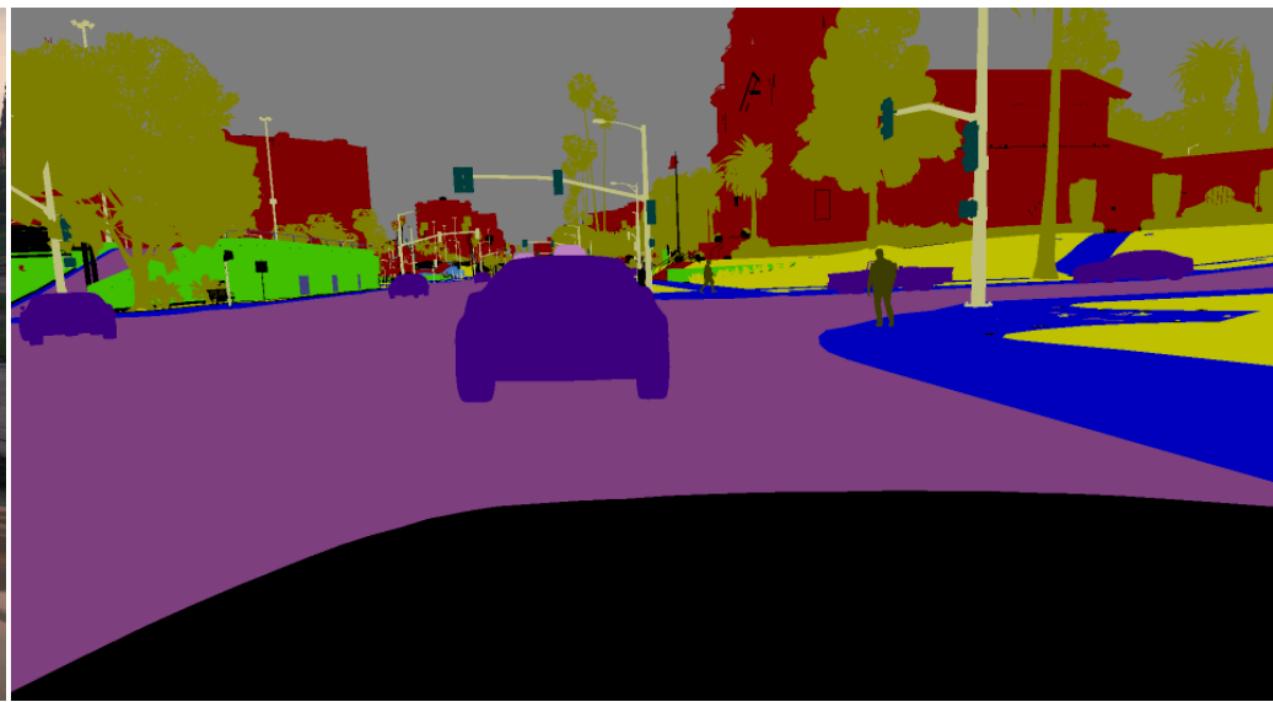
40



# Simulation



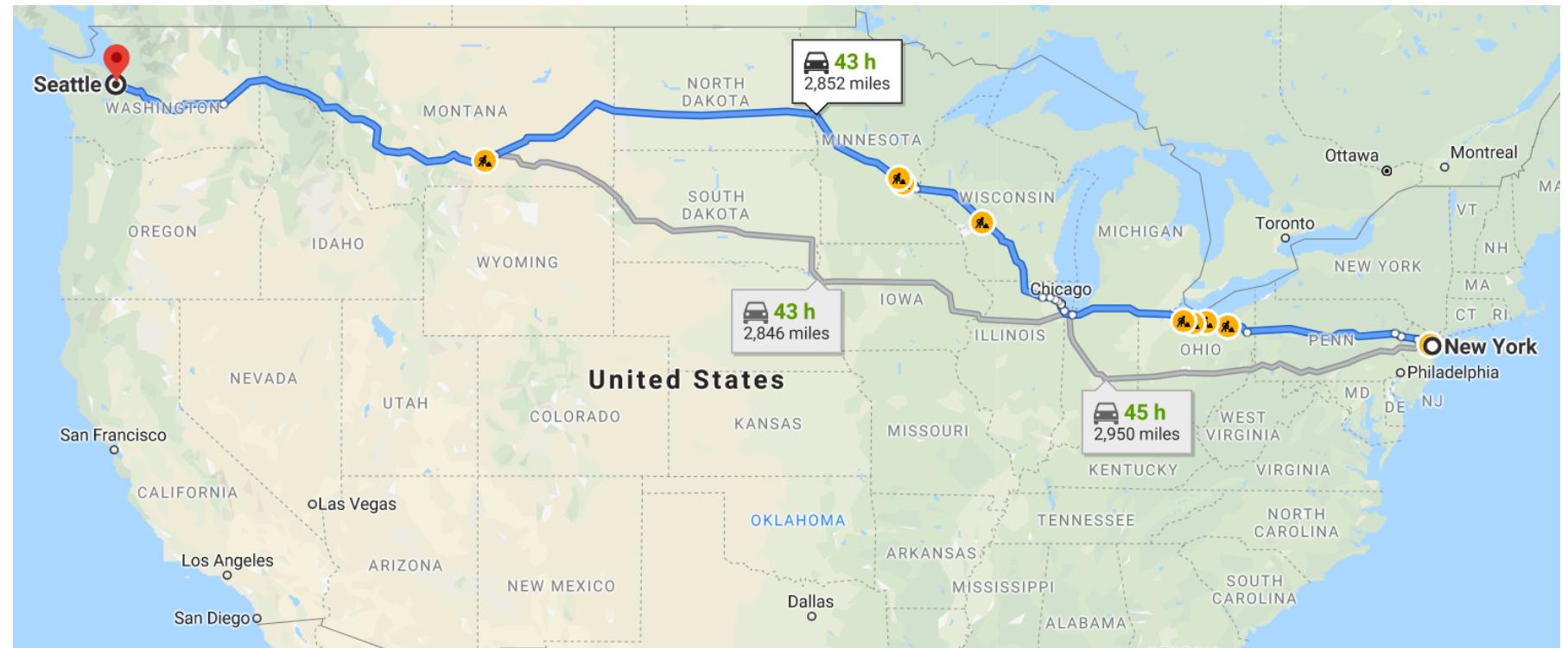
# Simulation

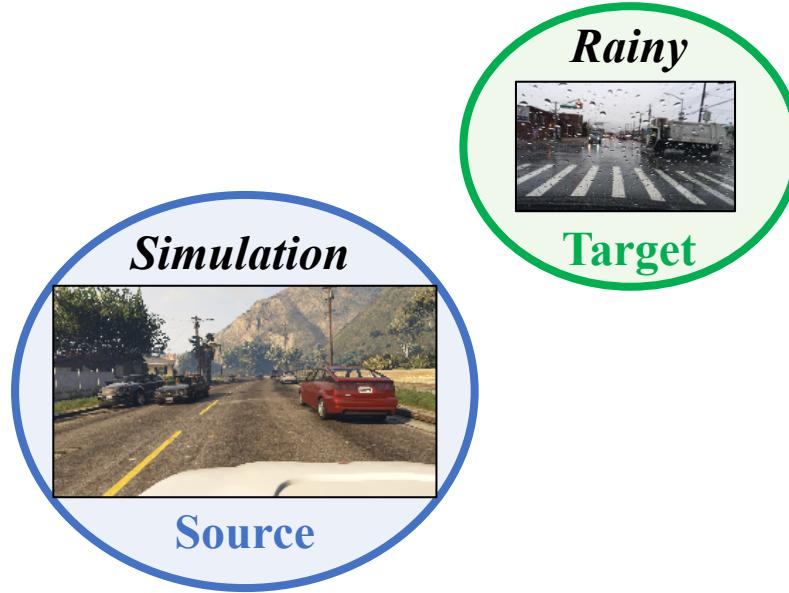


## Simulation

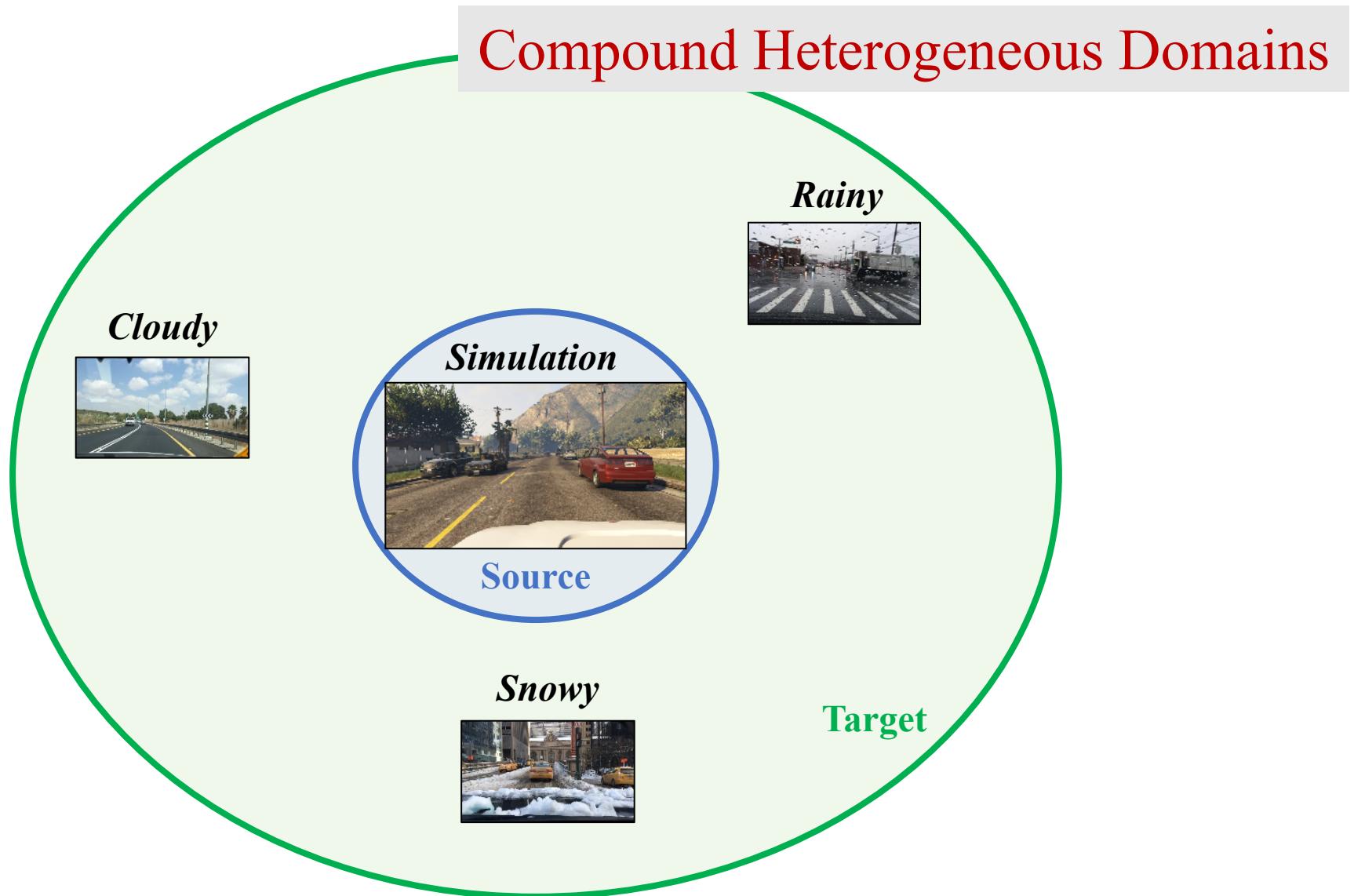


## Open World Driving Conditions

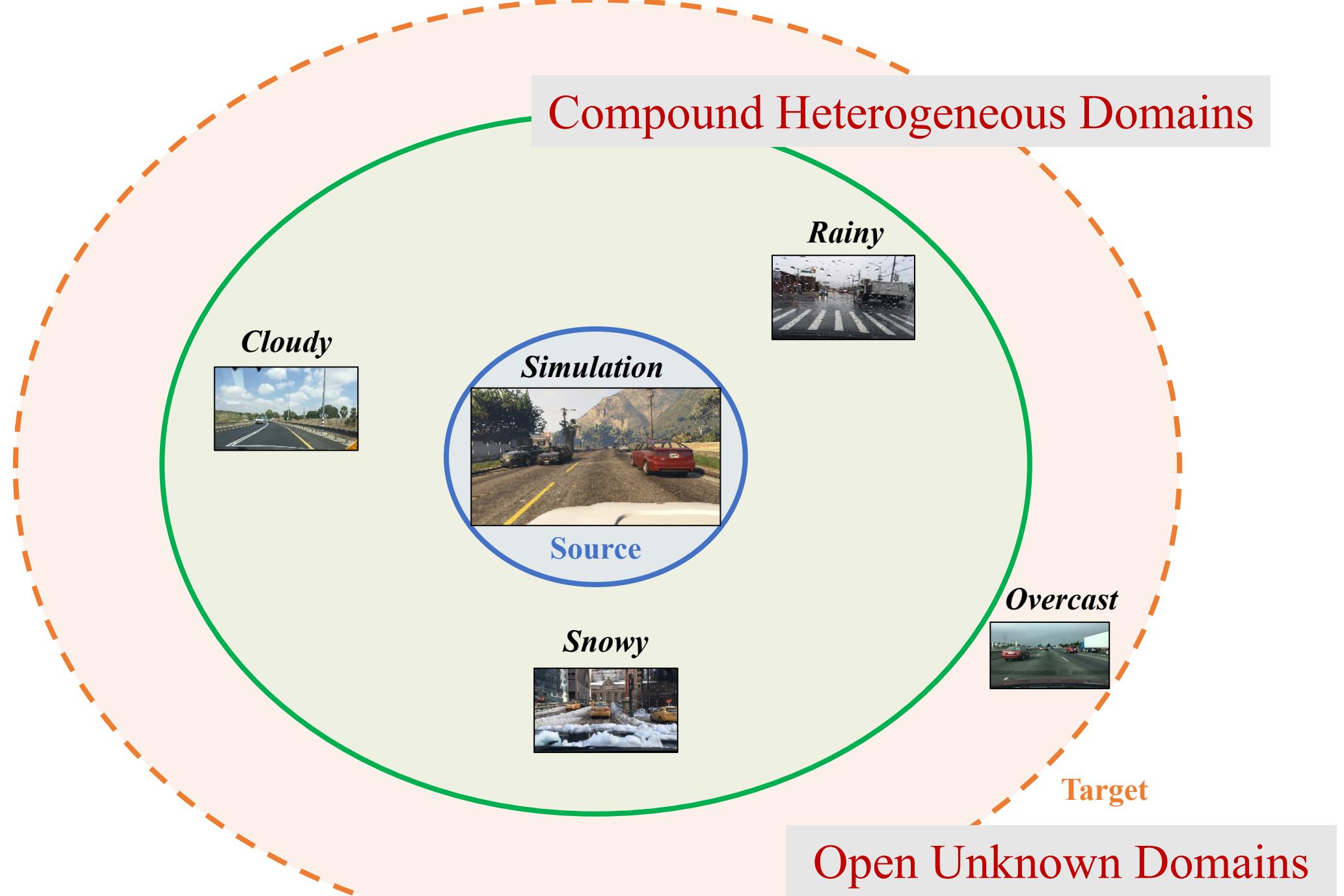


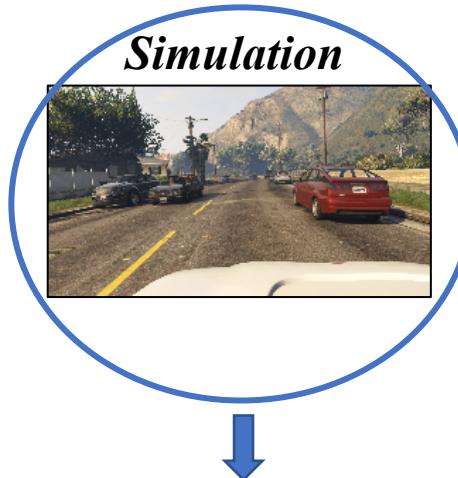


# Compound Heterogeneous Domains



# Compound Heterogeneous Domains





Source domain



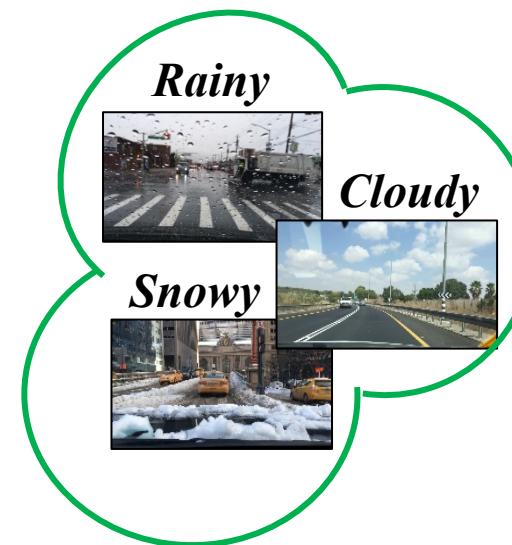
Single target domain

(a) Unsupervised  
Domain Adaptation



Multiple target domains

(b) Multi-Target  
Domain Adaptation



A compound target domain



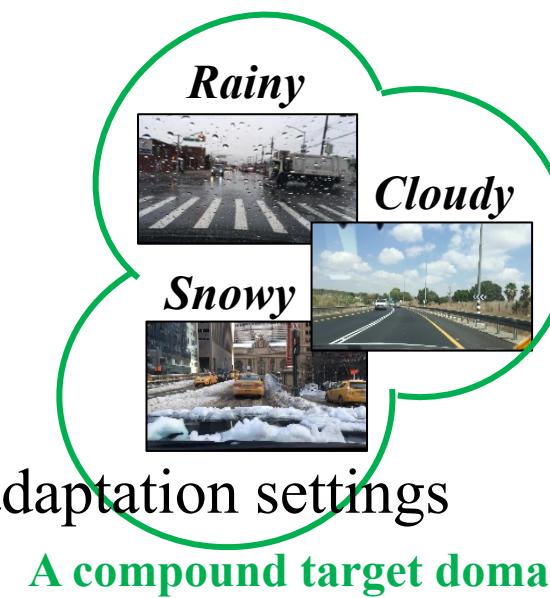
Unseen weather  
and more

Open Compound Domain Adaptation

## Challenges:

### 1) Compound Heterogeneous Domains

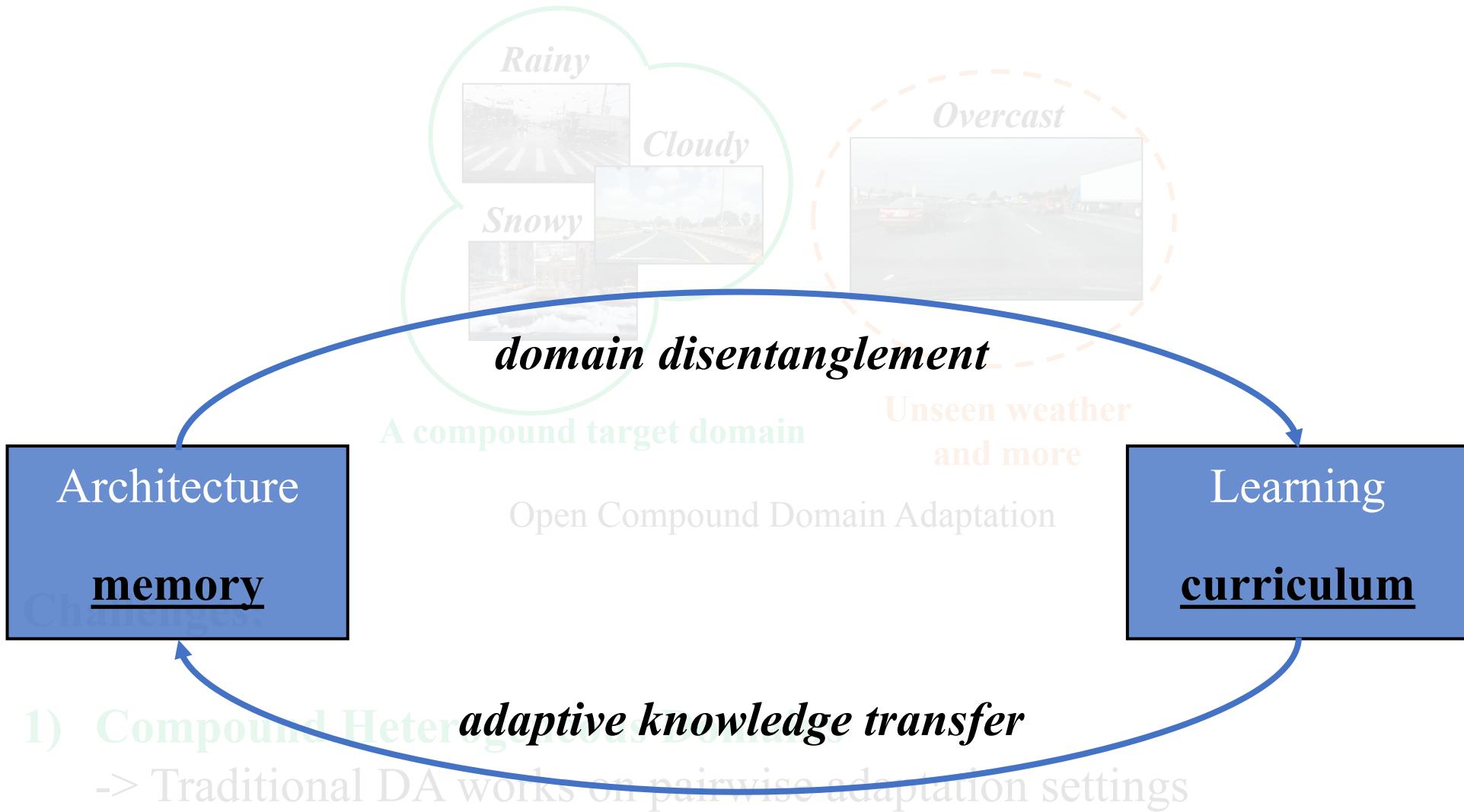
-> Traditional DA works on pairwise adaptation settings



### 1) Open Unknown Domains

-> Traditional DA assumes prior access to domain data during training

Open Compound Domain Adaptation



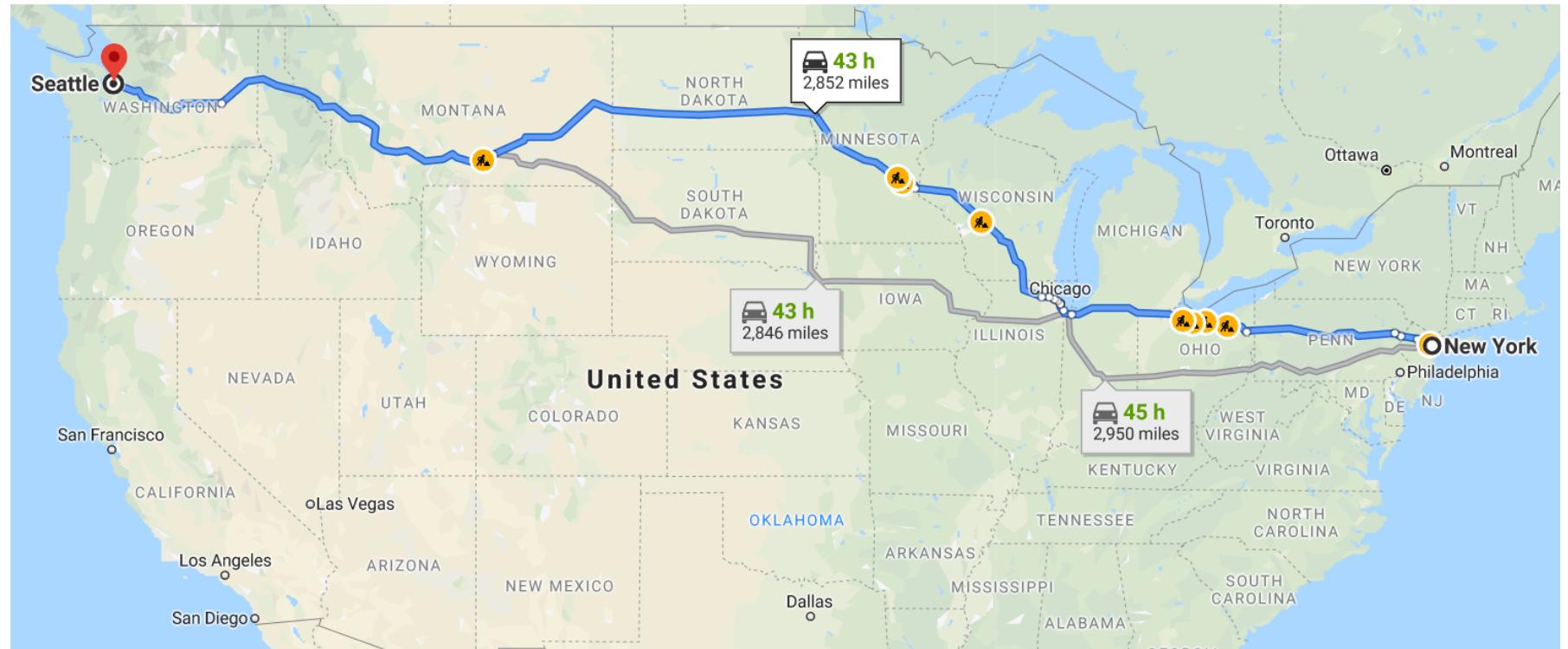
## 1) Open Unknown Domains

-> Traditional DA assumes prior access to domain data during training

## Simulation



## Open World Driving Conditions



## Source

Simulation



## Compound Targets

Open World Driving Conditions



Cloudy

Rainy

Overcast

Continuous Adaptation

## Source

Simulation



...

## Compound Targets

Open World Driving Conditions



...

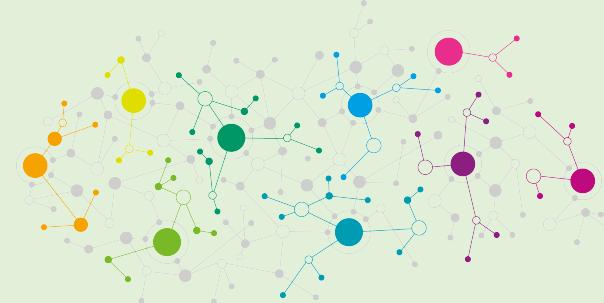


Cloudy

Rainy

Overcast

instance-wise curriculum



domain memory



Continuous Adaptation

## Source

Simulation



## Compound Targets

Open World Driving Conditions



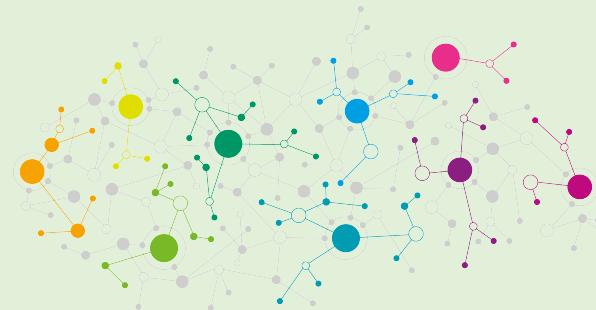
Domain  
Disentanglement

Cloudy

Rainy

Overcast

instance-wise curriculum



Adaptive  
Knowledge Transfer



Continuous Adaptation

## Source



SVHN

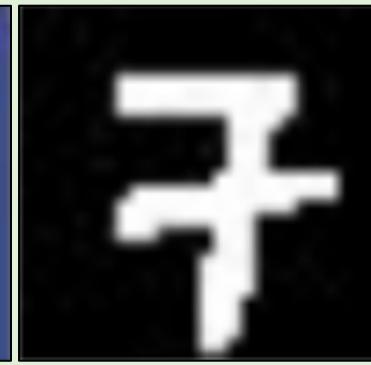
## Compound Targets

Open Compound Domain Digits Classification

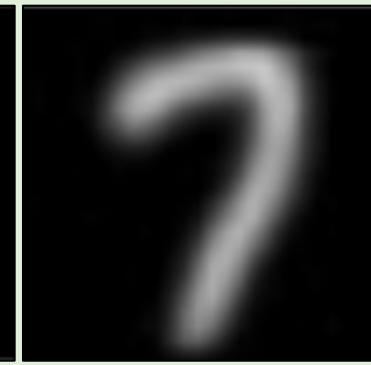


...

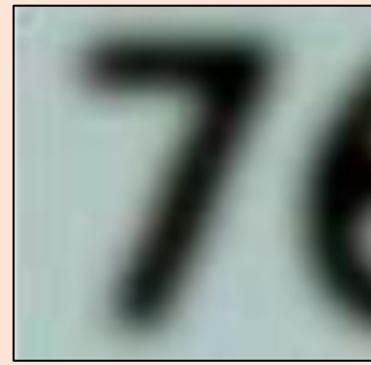
MNIST-M



MNIST

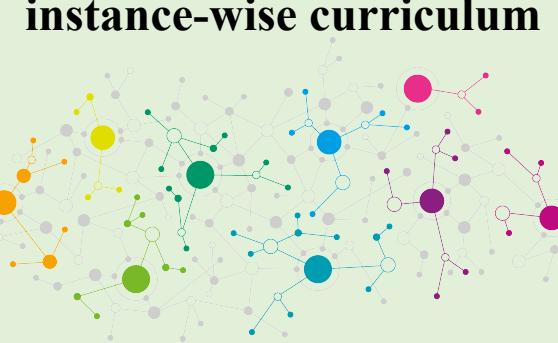


USPS



SymNum

Domain  
Disentanglement



Adaptive  
Knowledge Transfer



Continuous Adaptation

# Adversarial Domain Characteristics

## Disentanglement

$$\min_{E_{domain}} - \sum_i z_{random}^i \log D(E_{domain}(x^i))$$

$$\min_D - \sum_i y^i \log D(E_{domain}(x^i))$$

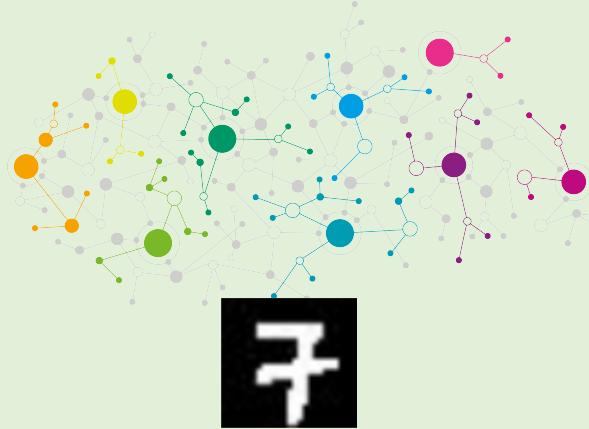


Source



Compound Targets

instance-wise curriculum



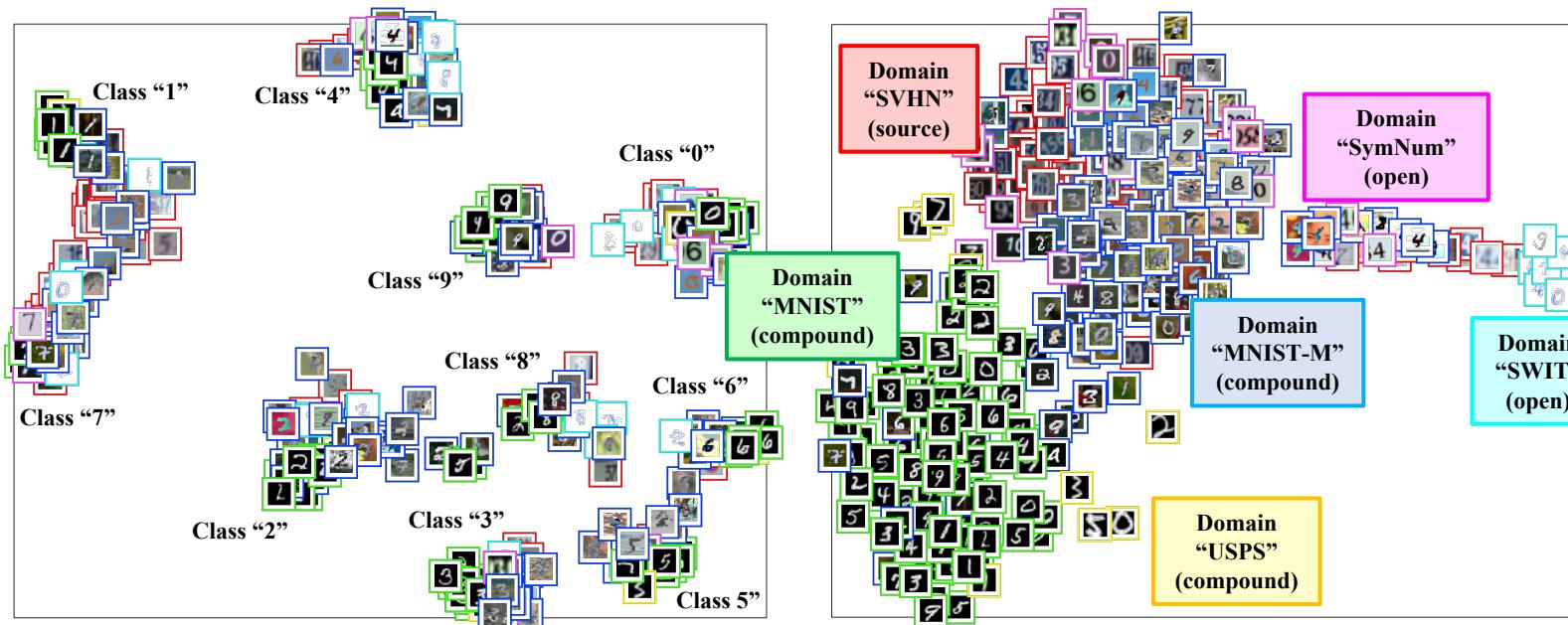
Open Targets

domain memory

Adaptive  
Knowledge Transfer



Continuous Adaptation



Source



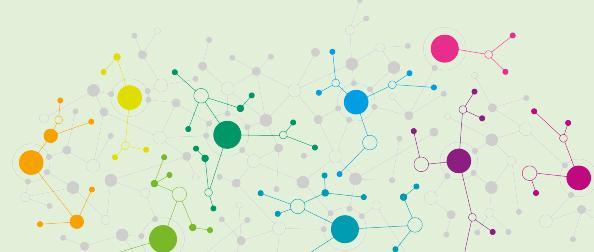
Domain  
Disentanglement

...



Compound Targets

instance-wise curriculum



Adaptive  
Knowledge Transfer

...



Open Targets

domain memory

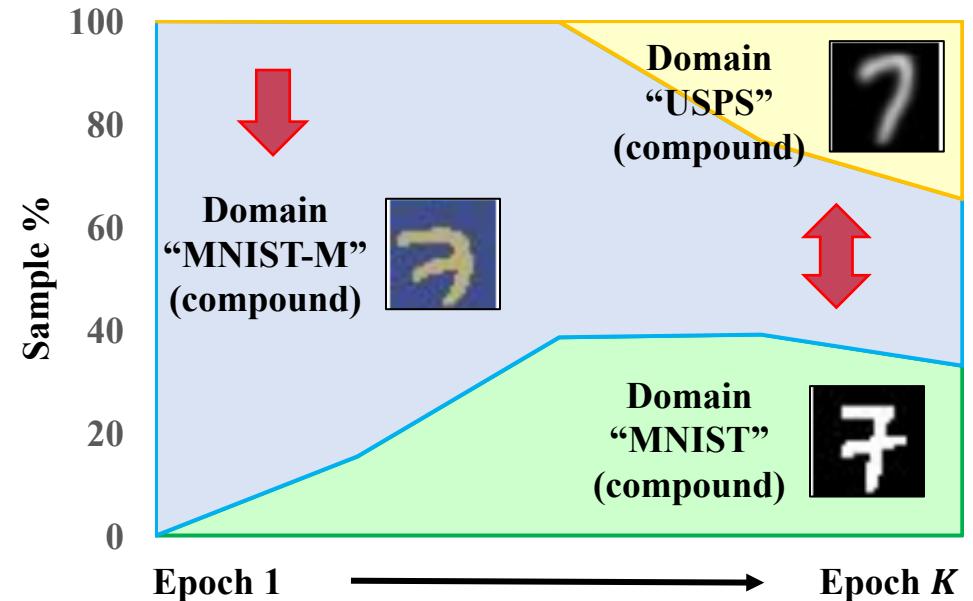


Continuous Adaptation

# Curriculum according to Domain Characteristics



Domain "SVHN" (source)

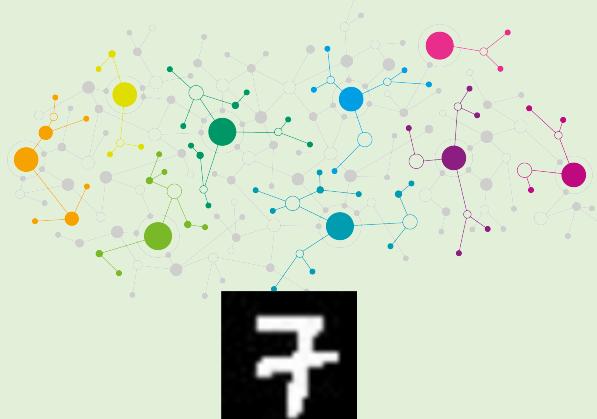


Source



Compound Targets

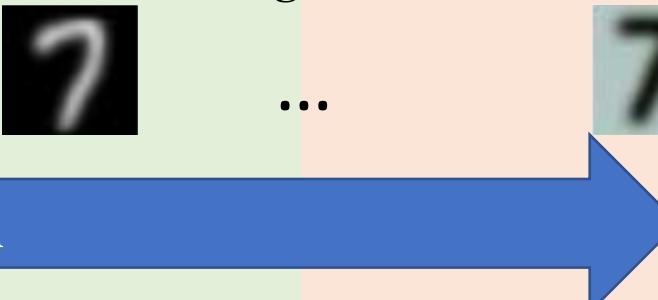
instance-wise curriculum



Open Targets

domain memory

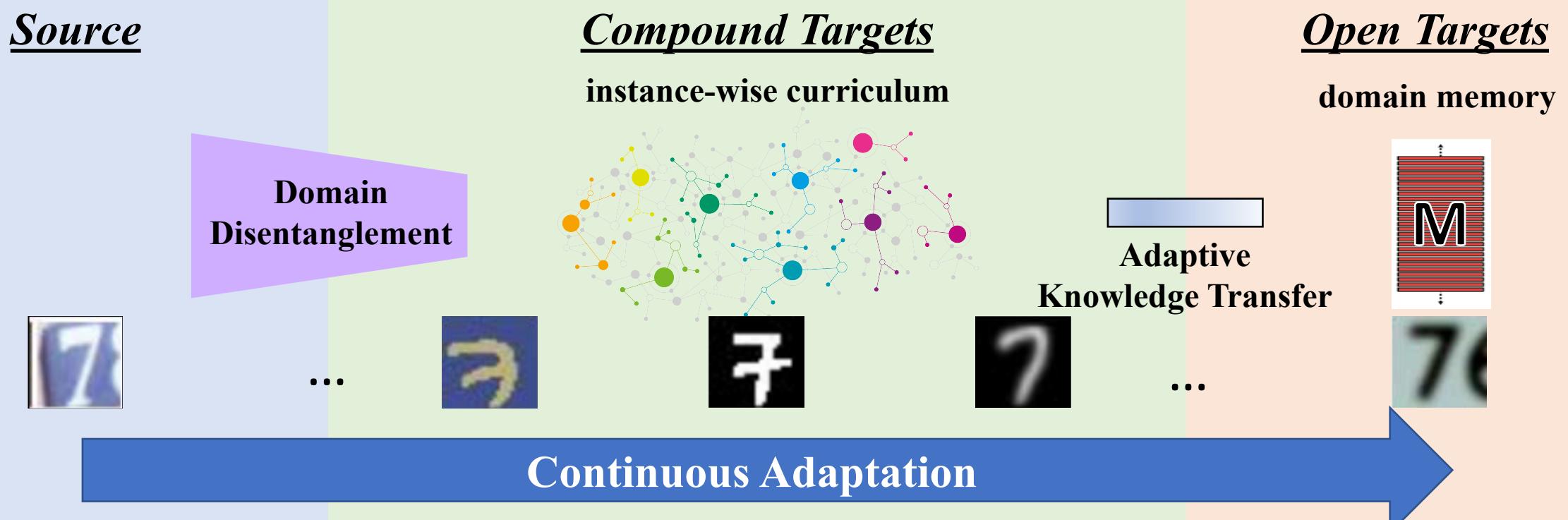
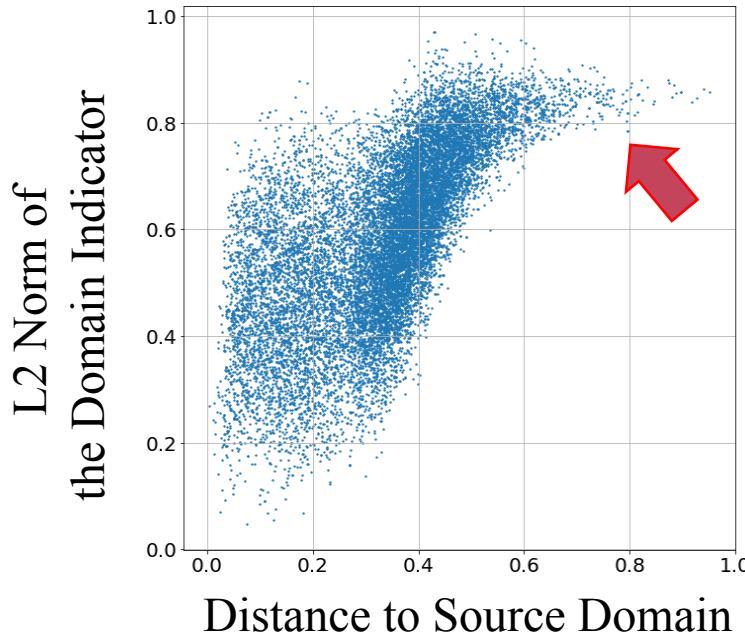
Adaptive Knowledge Transfer



Continuous Adaptation

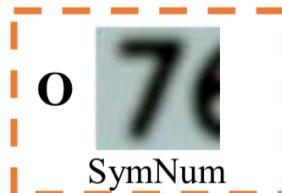
# Memory-Augmented Domain Indicator

$$v_{transfer} = v_{direct} + e_{domain} \otimes v_{enhance}$$



### C-Digits Benchmark

Absolute Performance Gain: ~5%



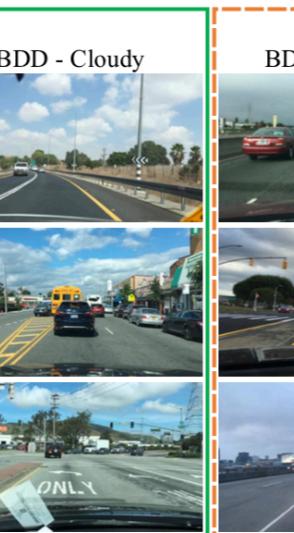
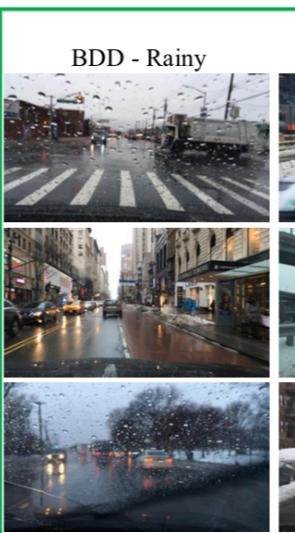
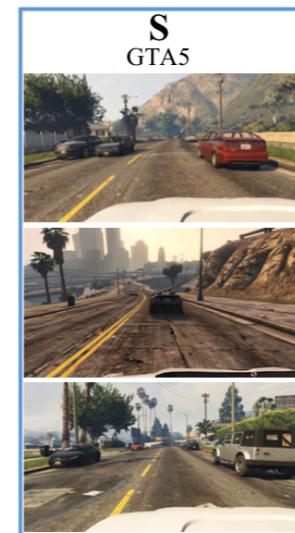
### C-Faces Benchmark

Absolute Performance Gain: ~10%



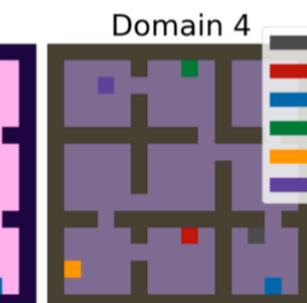
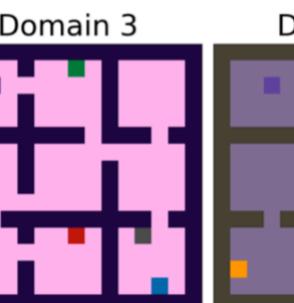
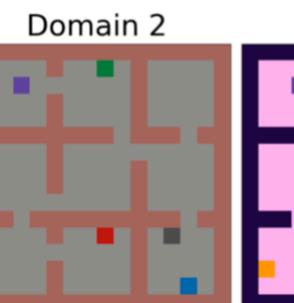
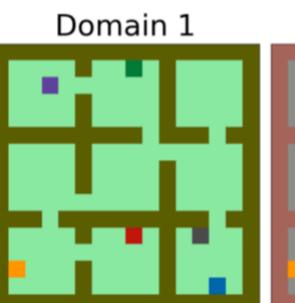
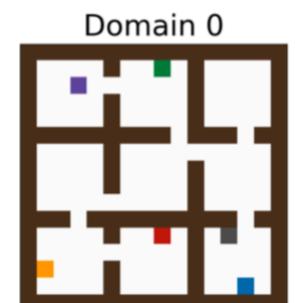
### C-Driving Benchmark

Absolute Performance Gain: ~2%



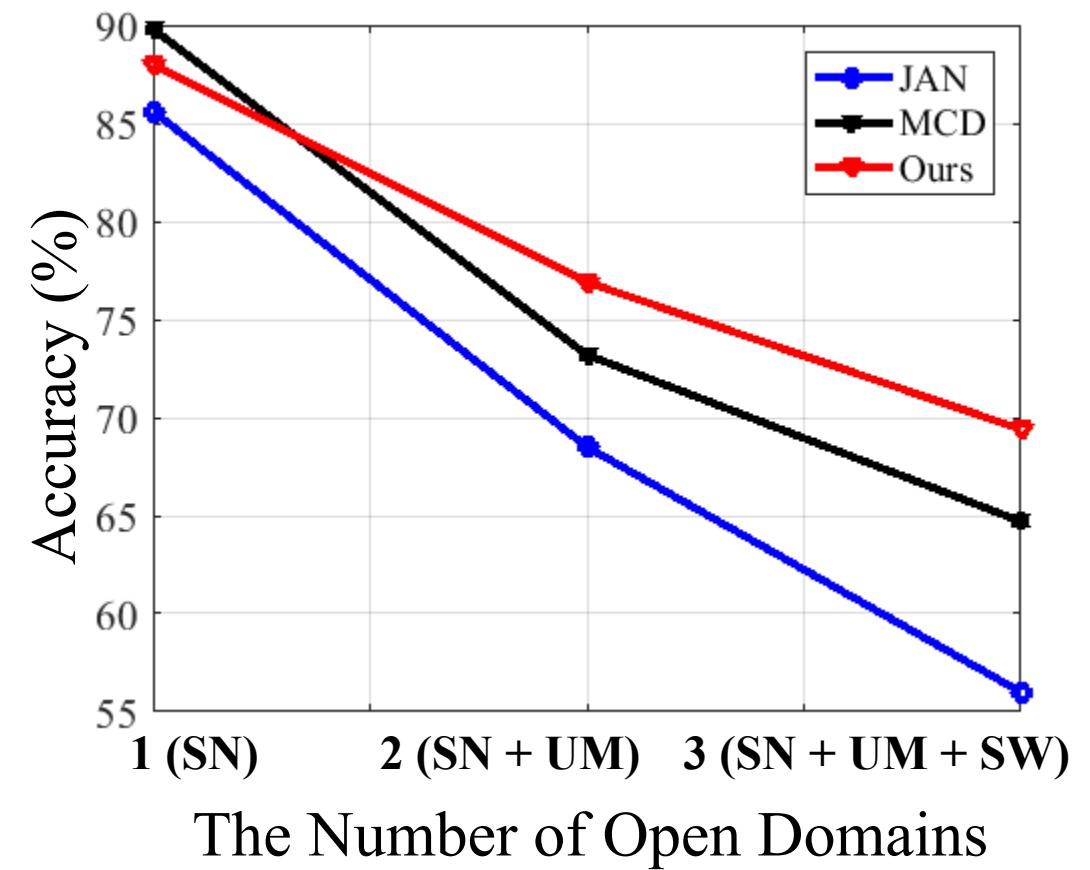
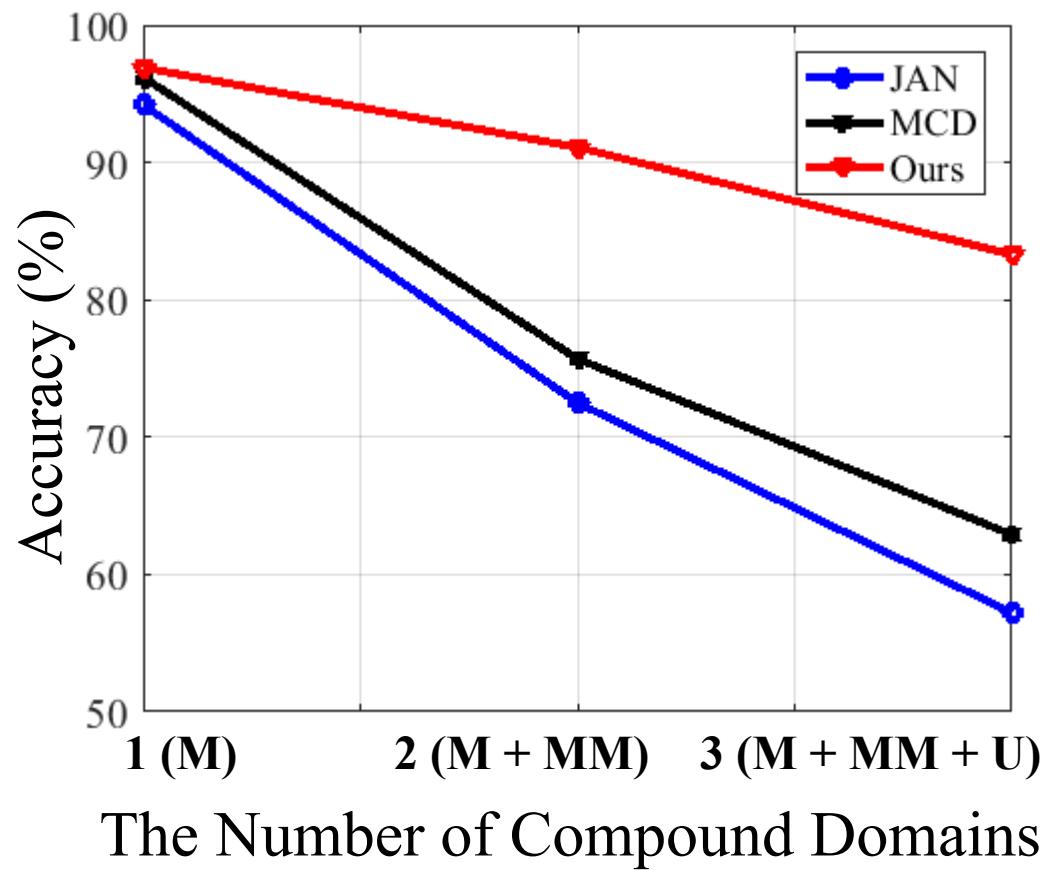
### C-Mazes Benchmark

Absolute Performance Gain: ~30%



Agent  
Treasure A  
Treasure B  
Treasure C  
Treasure D  
Treasure E

# Robustness to the complexity of compound domains and open domains



# Adaptation Results on C-Driving

(semantic segmentation)



Source Domain (Simulation)



Source Only



Ours



Compound Target Domain (Rainy)



Source Only



Ours



Open Target Domain (Overcast)



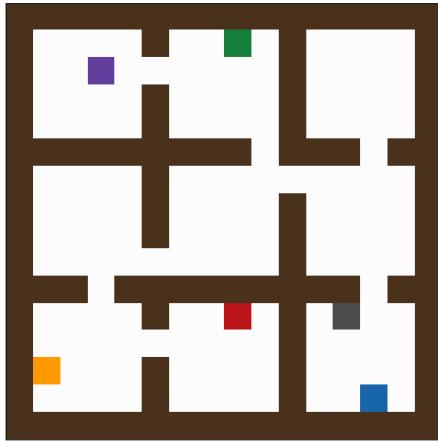
Source Only



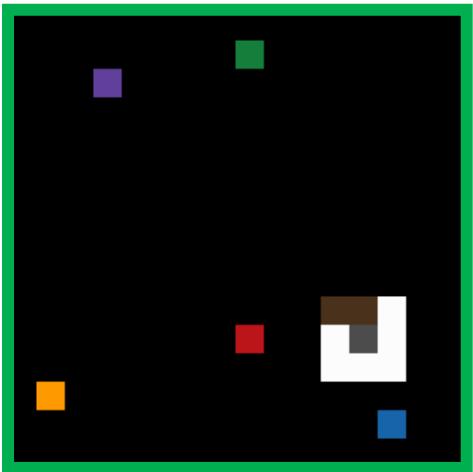
Ours

# Adaptation Results on C-Mazes

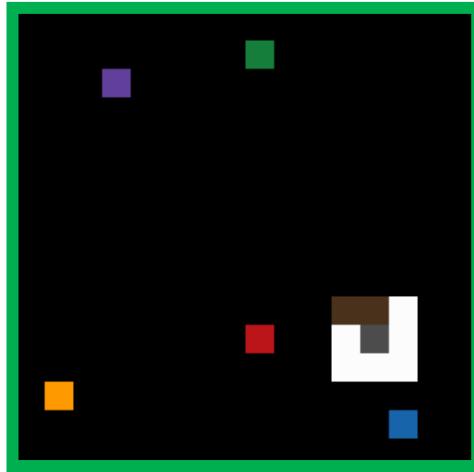
(reinforcement learning)



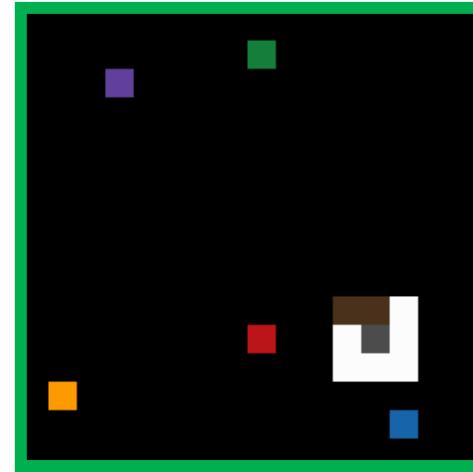
Source Domain



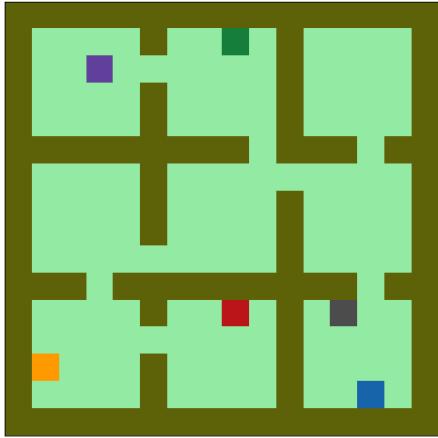
MTL  
(succeed)



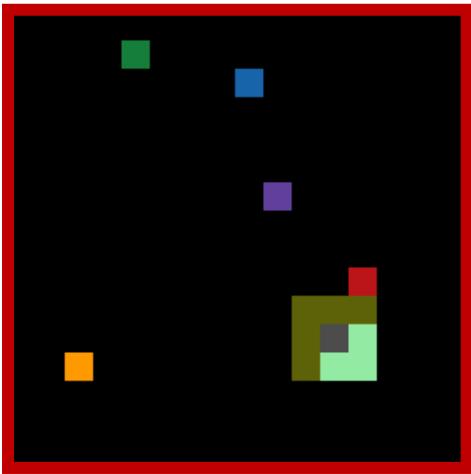
SynPo  
(succeed)



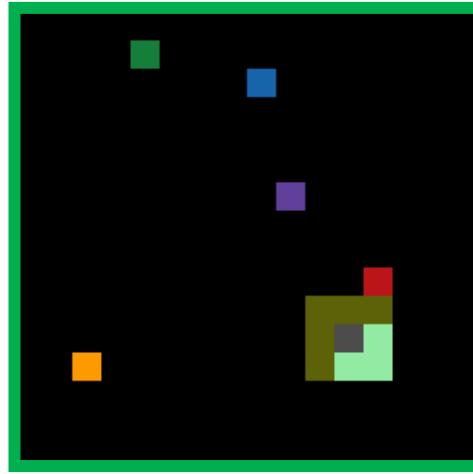
Ours  
(succeed)



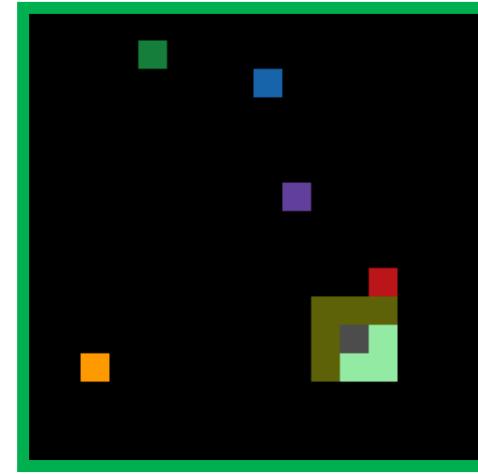
Open Target Domain 1



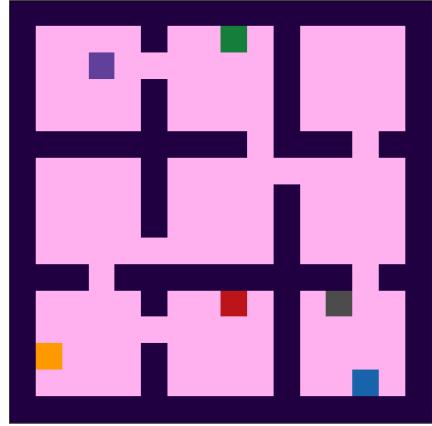
MTL  
(fail)



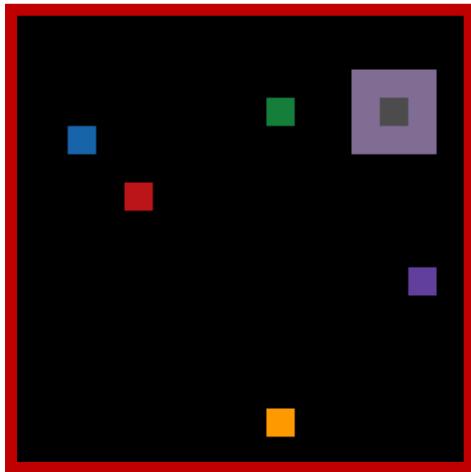
SynPo  
(succeed)



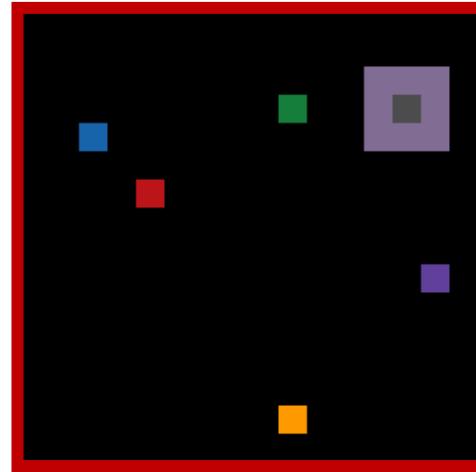
Ours  
(succeed)



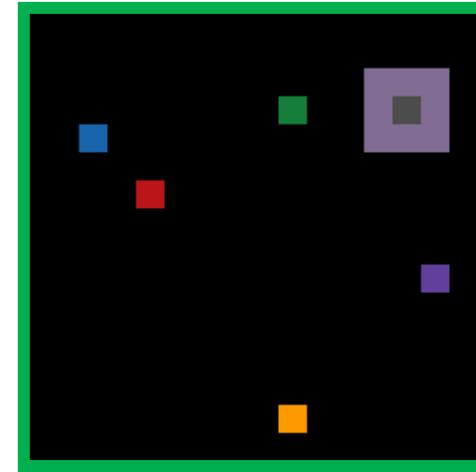
Open Target Domain 2



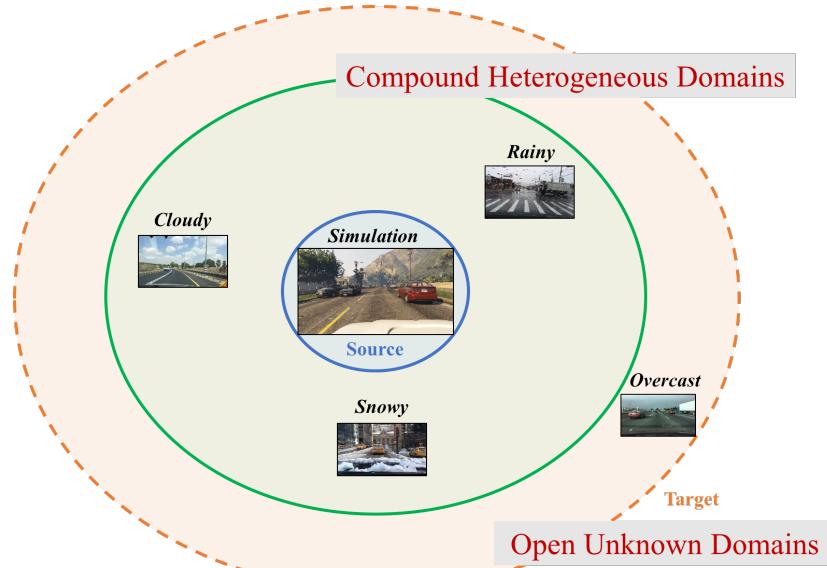
MTL  
(fail)



SynPo  
(fail)

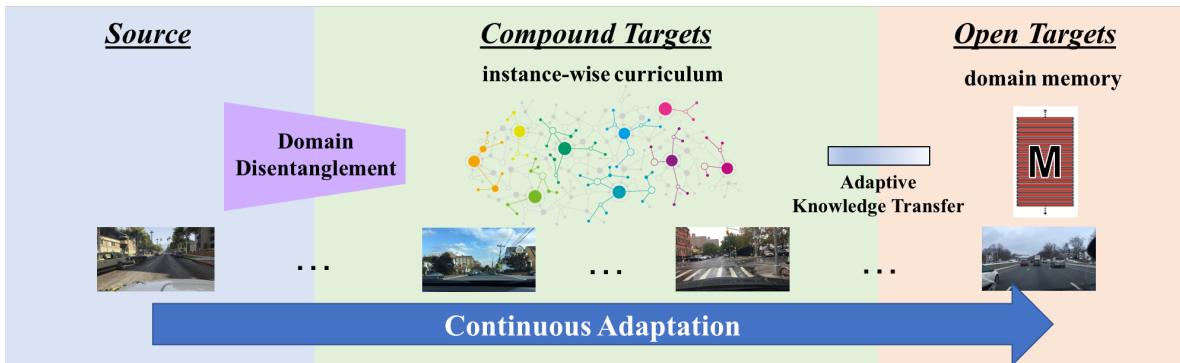


Ours  
(succeed)



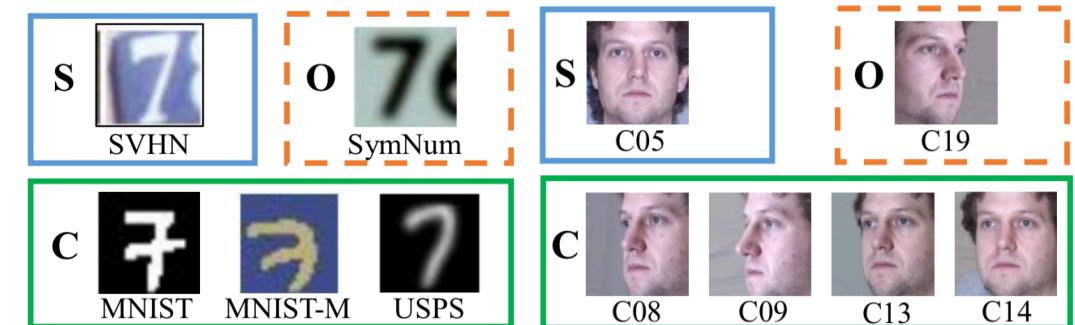
## New Task

Open Compound Domain Adaptation(OCDA)



## New Approach

Instance-wise Curriculum + Domain Memory



## New Benchmarks

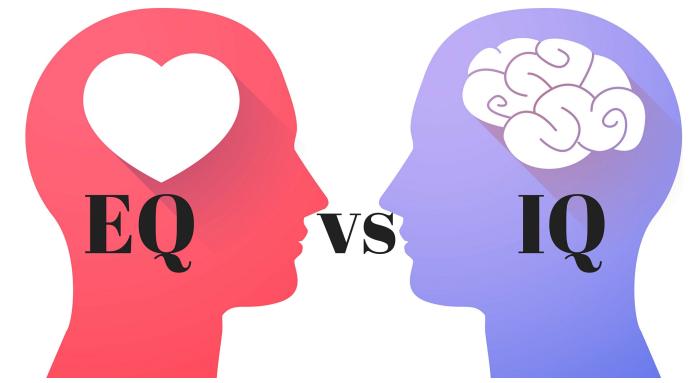
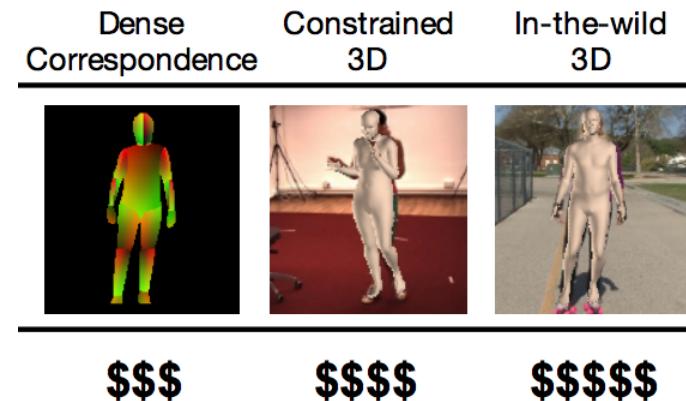
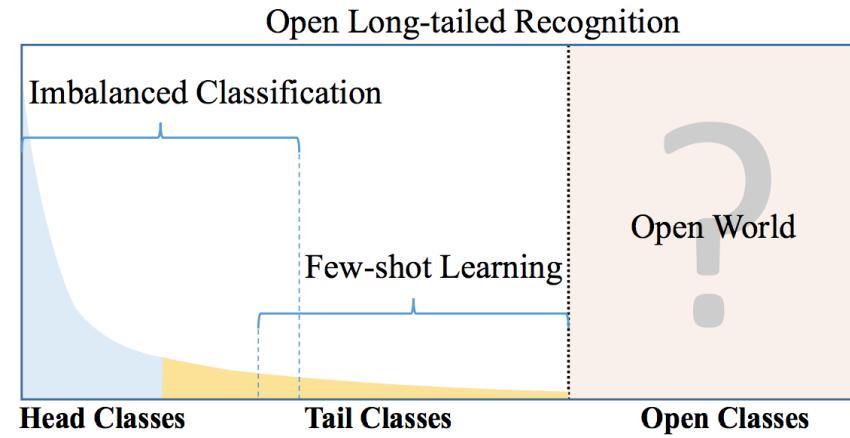
C-Digits, C-Faces, C-Driving, and C-Mazes



Code, models and benchmarks are available at

Project Page: <https://liuziwei7.github.io/projects/CompoundDomain.html>

# What's Next



“Devils are in the Tails”

“Blessing of Dimension”

“Ghost in the Shell”

# Thanks!

*Science is what we understand well enough to explain to a computer. Art is everything else we do.*

Homepage: <https://liuziwei7.github.io/>