JD AI Research at Visual Domain Adaptation Challenge 2018 — Detection

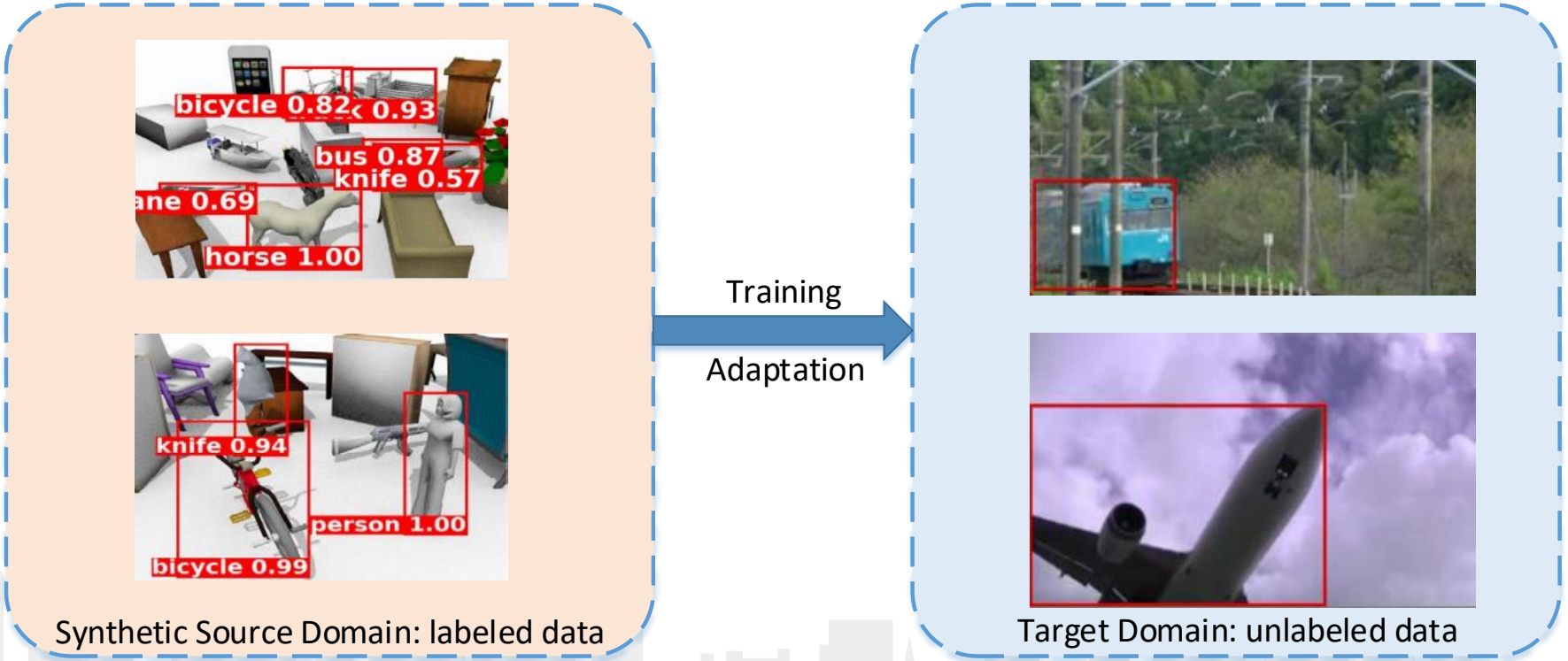
Yingwei Pan, Yehao Li, Qi Cai, Yiheng Zhang, Zhaofan Qiu, Ting Yao and Tao Mei

JD AI Research

Outline

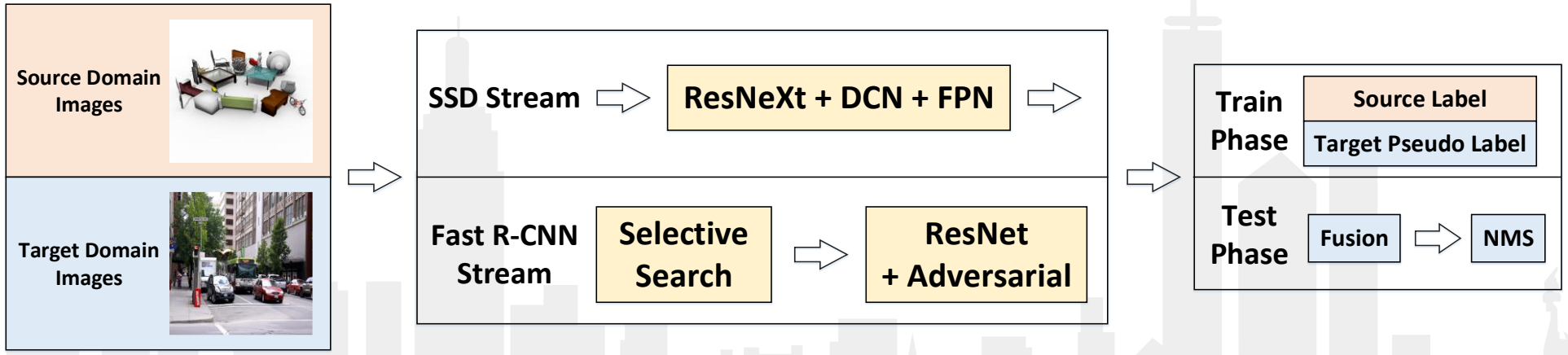
- Task
- Framework
- Results

Task — Detection



Framework

- Fusion of SSD and Fast R-CNN with pseudo labels
  - SSD detection module
  - Fast R-CNN detection module
  - Pseudo label assignment
  - Fusion strategy



SSD detection module

- One-stage detector.
- Backbone: A modified ResNeXt-101-64×4d network (ImageNet pre-trained) with DCN [1] and FPN [2].

[1] Dai, Jifeng, et al. Deformable Convolutional Networks. ICCV, 2017.
[2] Lin, Tsung Yi, et al. Feature Pyramid Networks for Object Detection. CVPR, 2017.

Fast R-CNN detection module

- Two-stage detector (bounding box proposal stage + classification stage)
- Backbone: ResNet101 (ImageNet pre-trained)

Bounding box proposal

- Employ selective search and randomly select 1,000 bounding boxes for each image.

Classification

- Classification loss on labeled source data
- Classification loss with the supervision of pseudo labels for target samples
- Adversarial loss to learn domain-invariant representations via a domain discriminator

Pseudo Label Assignment

- Assign pseudo labels with SSD detection module in **a three-stage strategy**

- For each target sample, select all the bounding boxes whose confidence scores are larger than 0.95 and perform clustering over them with K-means.
- **Bounding boxes (BB)** within each cluster -> one **candidate bounding box (CBB)** with pseudo label

- For each category, select CBB with top-2000 highest confidence scores and perform clustering over them with K-means
- Set the center of cluster with highest confidence score as the **prototype** of this category
- Assign pseudo labels to clusters with both top 50% highest confidence scores and top 50% minimum distances to the prototype

- For the clusters of each category filtered out in previous stage, if the distances between their centers and the prototypes are close enough, we assign the CBB in each cluster with the label of its nearest prototype

Fusion strategy

- Late fusion scheme (SSD + Fast R-CNN)
  - Linearly fuse the predicted score distributions of each bounding box from SSD and Fast R-CNN
  - Perform soft NMS to produce the final detection results

Results

#	Setting	Per Category Average Precision (AP)												mAP
		plane	bcycl	bus	car	horse	knife	mcycl	person	plant	sktbd	train	truck	
1	SSD source only	0.199	0.157	0.244	0.365	0.331	0.028	0.346	0.066	0.089	0.184	0.095	0.053	0.180
2	Fast R-CNN + Adversarial	0.561	0.167	0.406	0.611	0.419	0.166	0.399	0.094	0.162	0.111	0.170	0.095	0.280
3	SSD iteration 1	0.656	0.272	0.557	0.650	0.539	0.052	0.437	0.091	0.107	0.233	0.230	0.181	0.334
4	SSD + Fast R-CNN iteration 1	0.631	0.249	0.550	0.742	0.510	0.146	0.483	0.091	0.151	0.228	0.250	0.123	0.346
5	SSD iteration 2	0.737	0.301	0.696	0.742	0.577	0.101	0.486	0.139	0.223	0.188	0.363	0.559	0.426
6	SSD + Fast R-CNN iteration 2	0.758	0.328	0.708	0.808	0.607	0.140	0.525	0.063	0.226	0.209	0.375	0.569	0.443
7	SSD iteration 3	0.752	0.287	0.703	0.729	0.567	0.312	0.459	**0.148**	**0.246**	0.177	0.369	0.634	0.449
8	SSD + Fast R-CNN iteration 3	0.770	0.343	**0.720**	**0.819**	0.611	0.332	**0.542**	0.085	0.242	0.230	0.387	0.654	0.478
9	Ensemble	**0.784**	**0.367**	0.719	0.787	**0.614**	**0.342**	0.525	0.075	0.239	**0.275**	**0.399**	**0.702**	**0.486**

- Involving more iterations with pseudo labels tends to achieve better performance.

JDAI Research

THANKS!