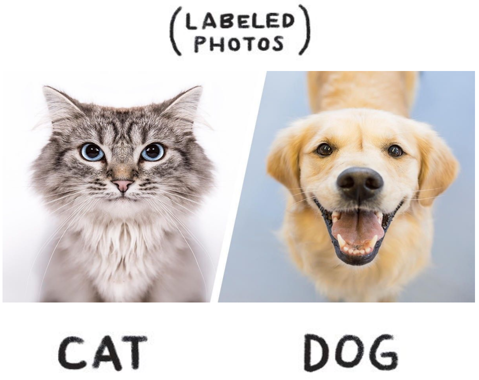
Learning to Generalize with Self-Supervision

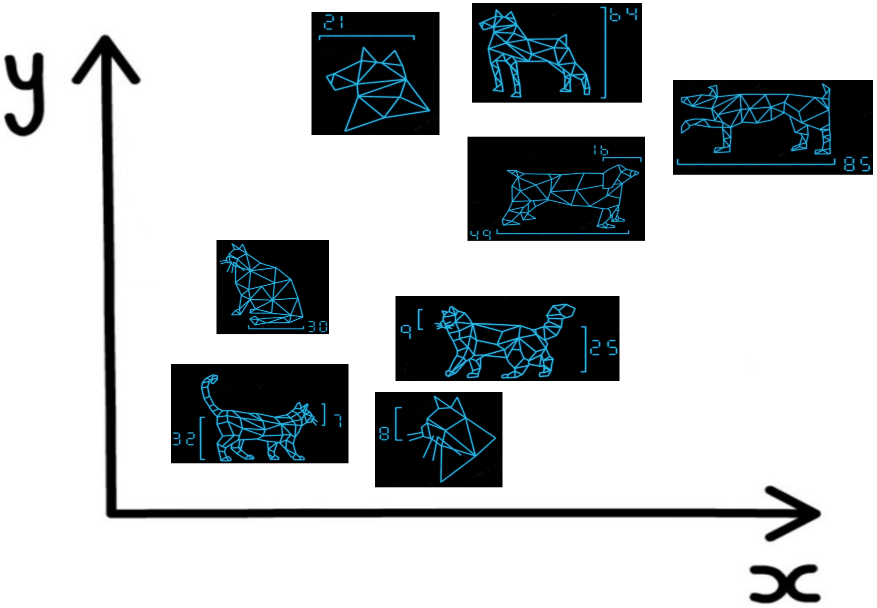
Tatiana Tommasi
Assistant Professor, Polytechnic University of Turin, Italy
Affiliated Researcher, Italian Institute of Technology

Computer Vision Winter Workshop
February 3rd, 2020

Machine Learning

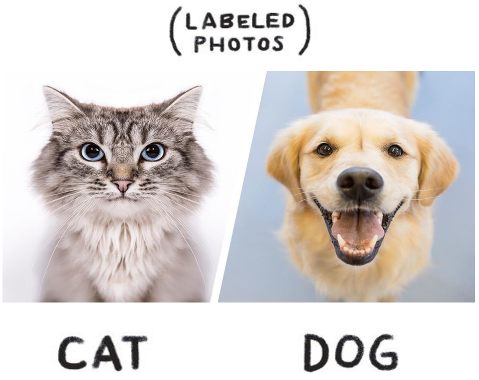
the study of algorithms that computer systems use to improve their **performance** a specific **task** relying on patterns and inference from **data (experience)**.

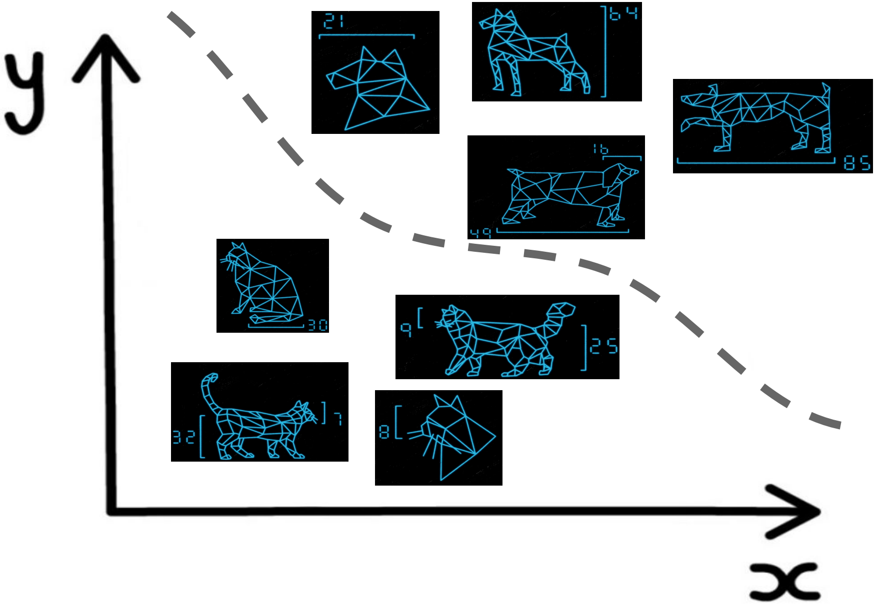




Machine Learning

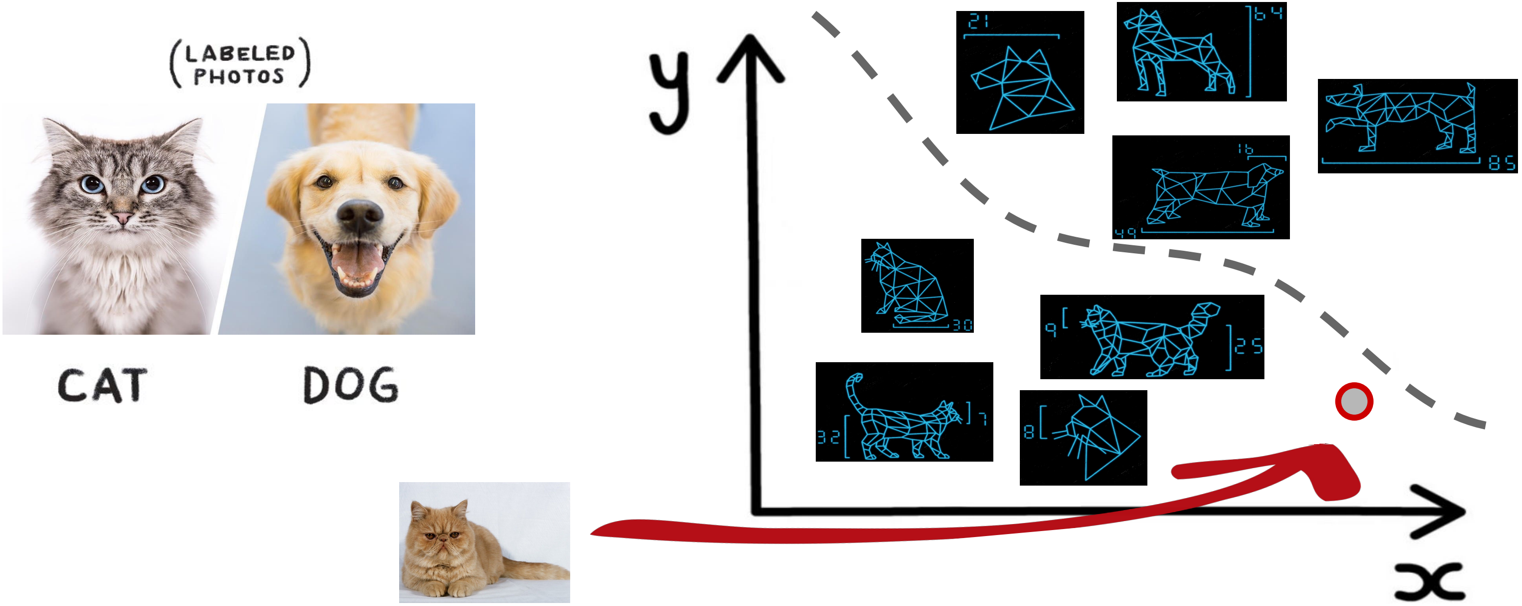
the study of algorithms that computer systems use to improve their **performance** a specific **task** relying on patterns and inference from **data (experience)**.





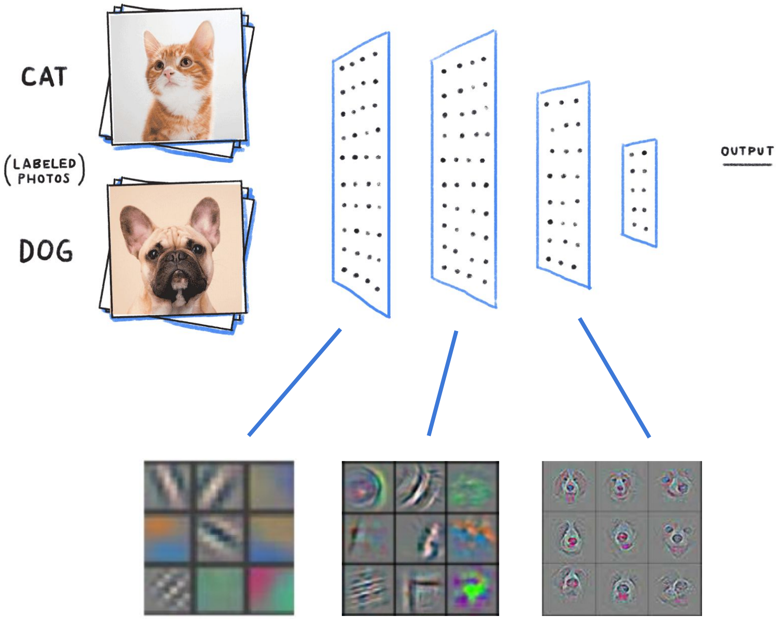
Machine Learning

the study of algorithms that computer systems use to improve their **performance** a specific **task** relying on patterns and inference from **data (experience)**.



Deep Machine Learning

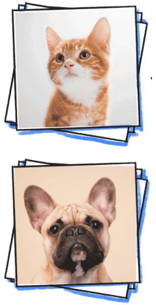
the study of algorithms that computer systems use to improve their **performance** a specific **task** relying on patterns and inference from *a lot of...* **data (experience)**.



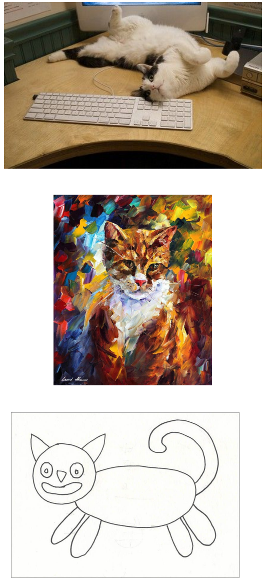
Filter Images from Zeiler & Fergus, 2013

Deep Machine Learning

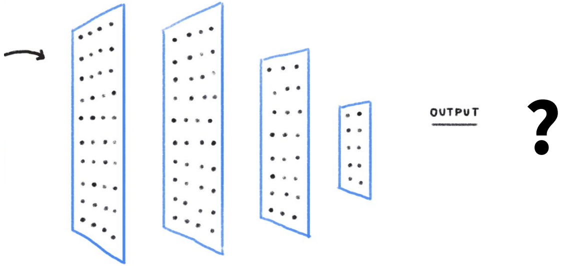
the study of algorithms that computer systems use to improve their **performance** a specific **task** relying on patterns and inference from *a lot of...* **data (experience)**.



Train

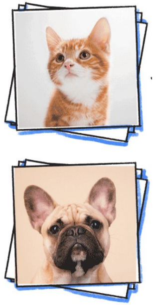


Test

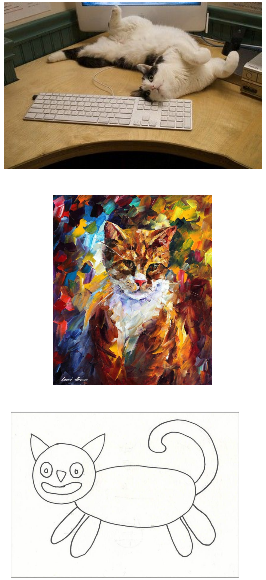


Deep Machine Learning

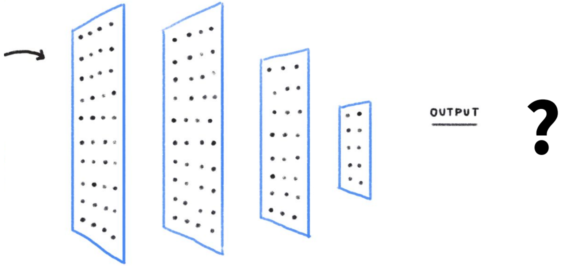
the study of algorithms that computer systems use to improve their **performance** a specific **task** relying on patterns and inference from *a lot of...* **data (experience)**.



Train



Test



Intelligent?

"Intelligence measures an agent's ability to adapt and achieve goals in a wide range of environments."

A Collection of Definitions of Intelligence, 2007 Conference on Advances in Artificial General Intelligence (Wikipedia)

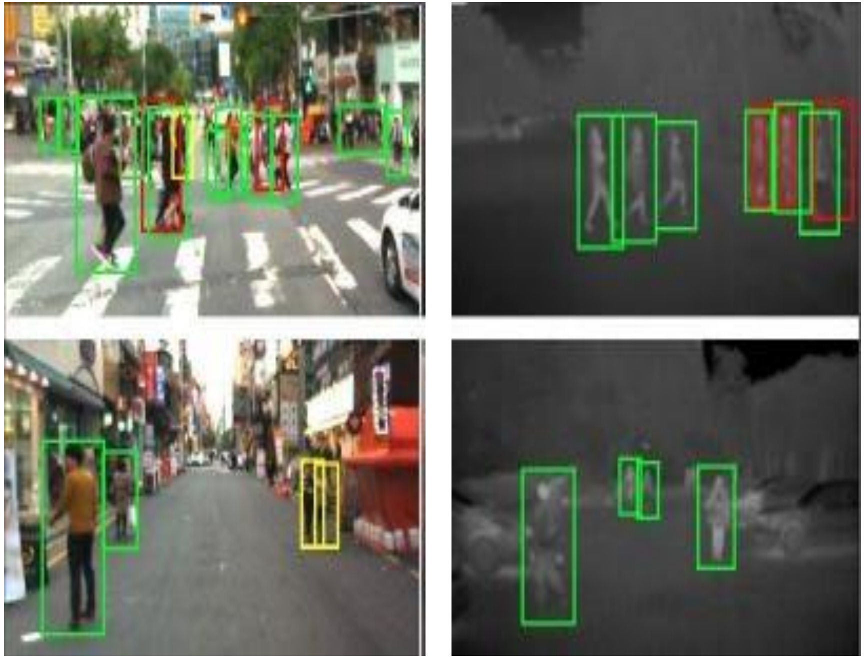
Domain Adaptive Learning: why?

Appearance changes due to seasonal and time changes.

Slide Credit: M. Mancini, Domain Adaptation Tutorial, ICIAP 2019

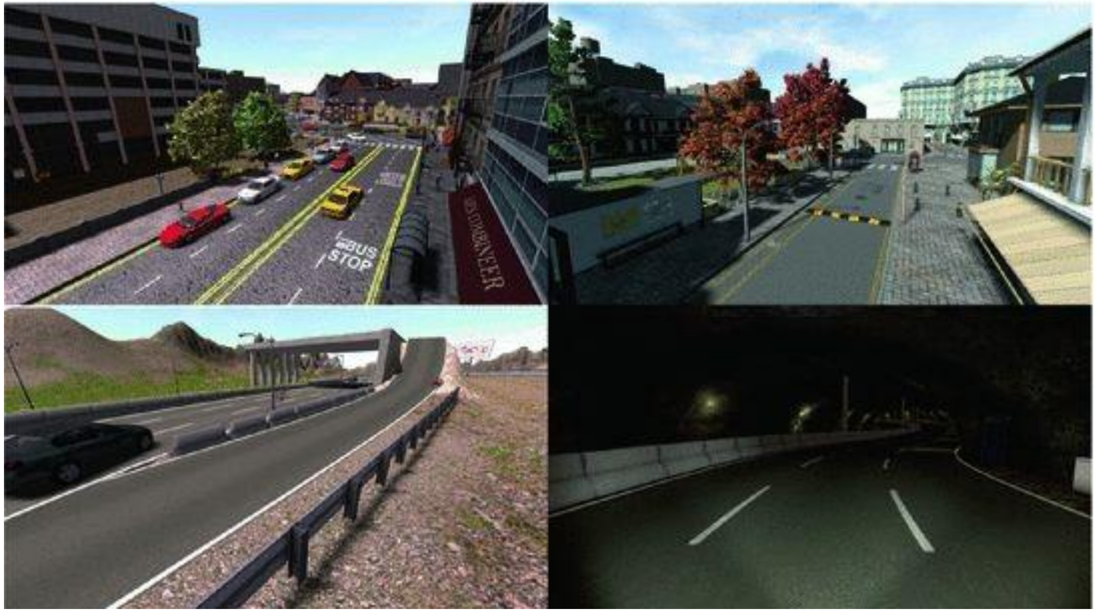
Domain Adaptive Learning: why?

Appearance changes due to different sensors.



Domain Adaptive Learning: why?

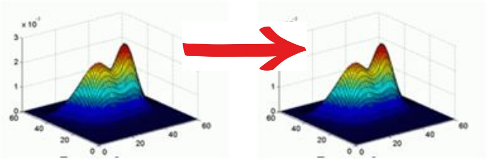
Use of synthetic data

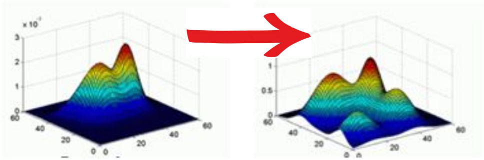


Domain Adaptive Learning: why?

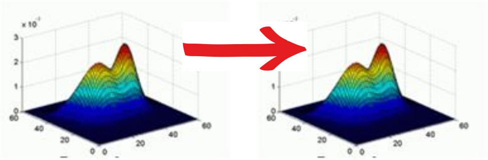
Overcoming costly/unfeasible data collection

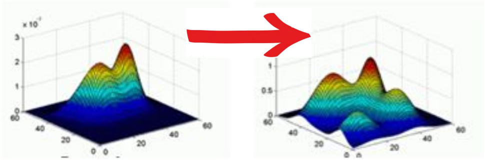
Domain Adaptation Settings

Train  Test

Source  Target

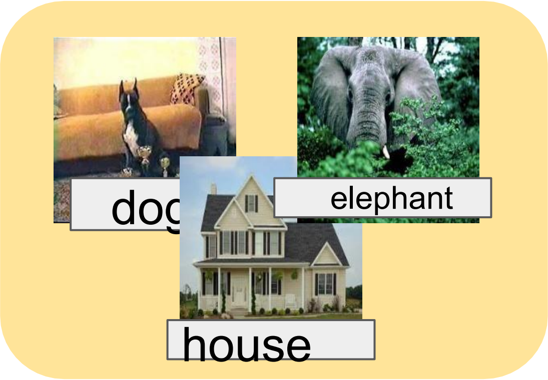
Domain Adaptation Settings

Train  Test

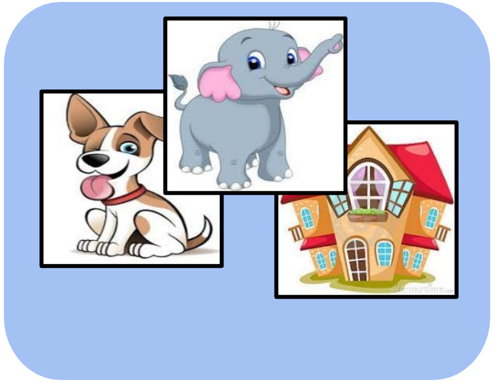
Source  Target

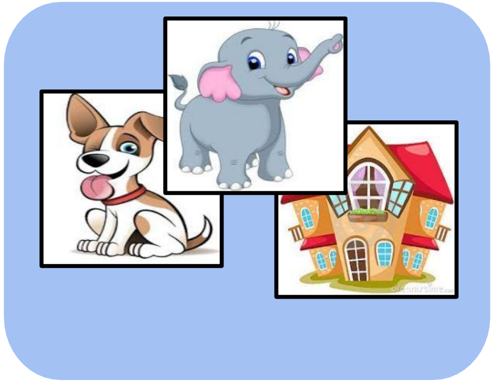
Unsupervised DA: transductive setting. T is available during training but is unlabeled

Labelled **Source** Domain



Unlabelled **Target** Domain





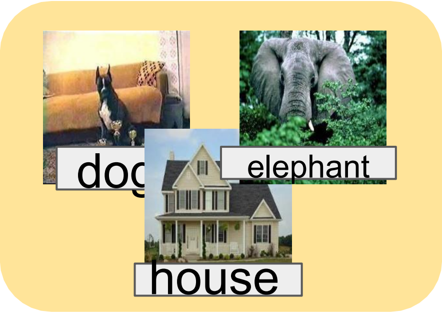
Train

Test

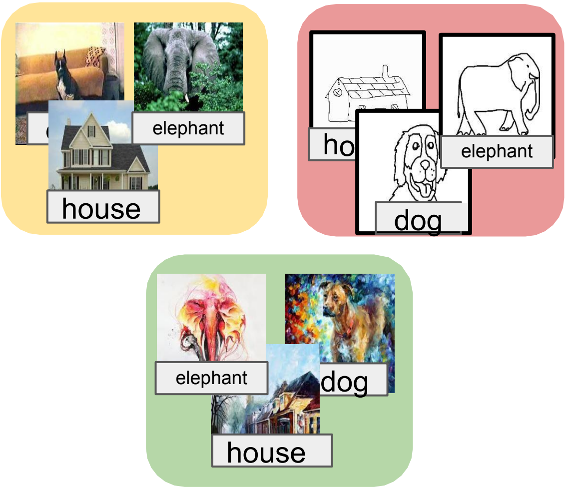
Domain Adaptation Settings

Number of Sources

Single



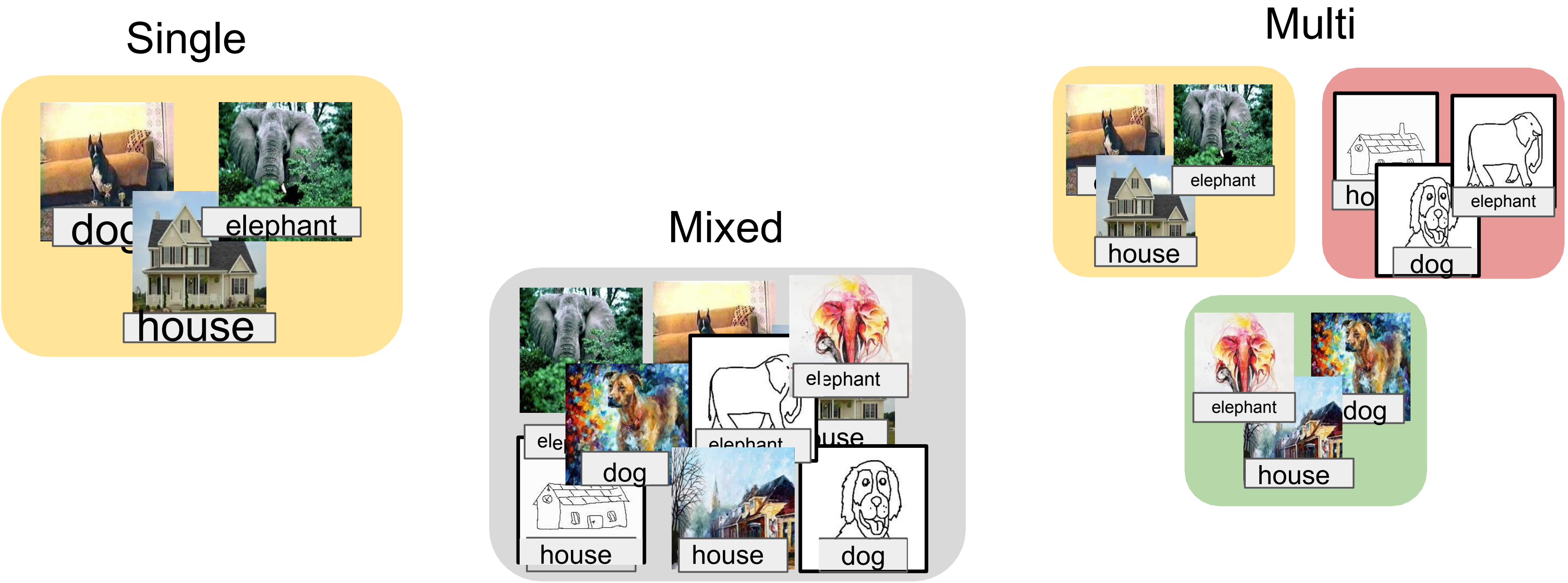
Multi



Slide Credit: M. Mancini, Domain Adaptation Tutorial, ICIAP 2019

Domain Adaptation Settings

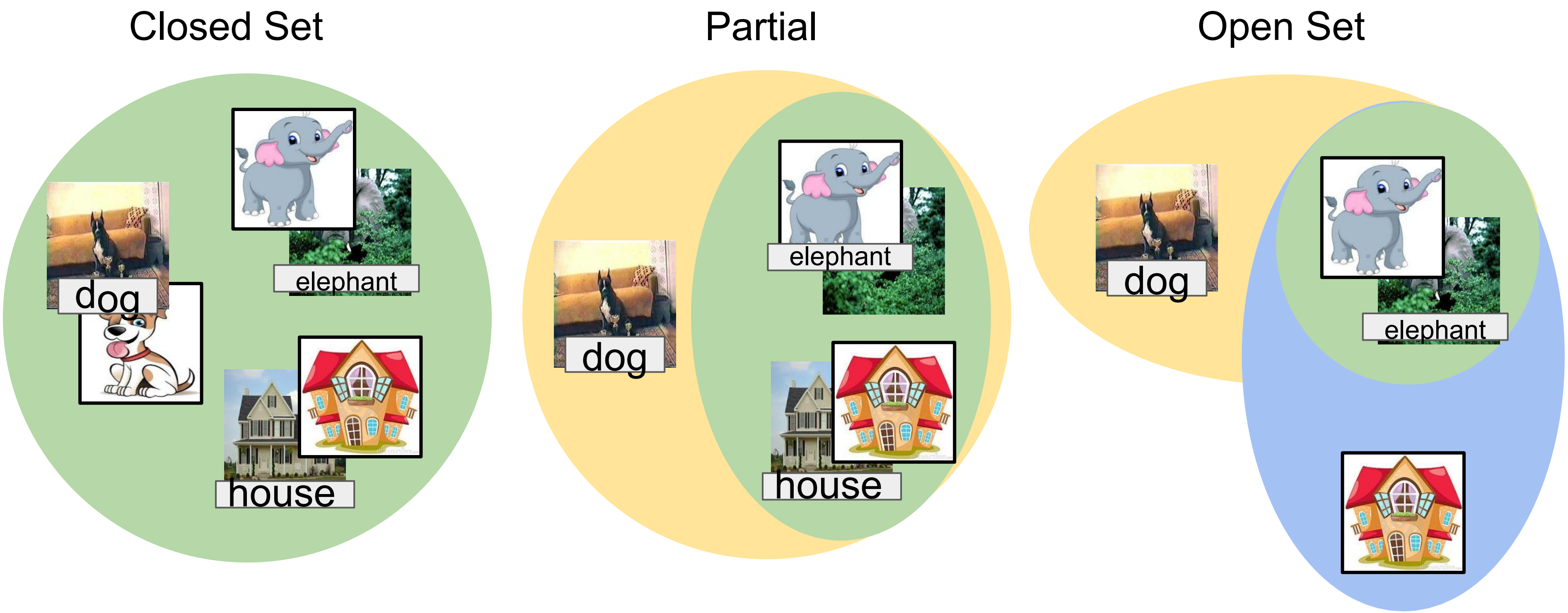
Number of Sources



Slide Credit: M. Mancini, Domain Adaptation Tutorial, ICIAP 2019

Domain Adaptation Settings

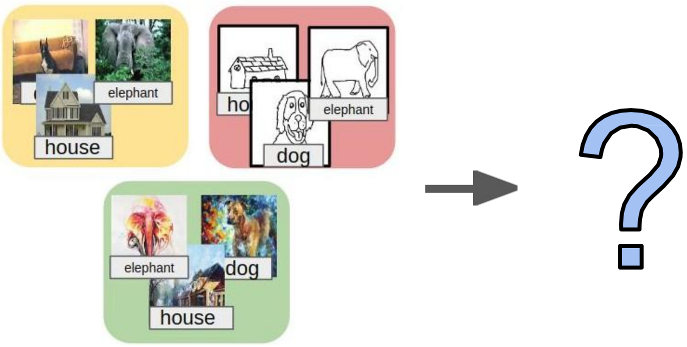
The Label Space



Slide Credit: M. Mancini, Domain Adaptation Tutorial, ICIAP 2019

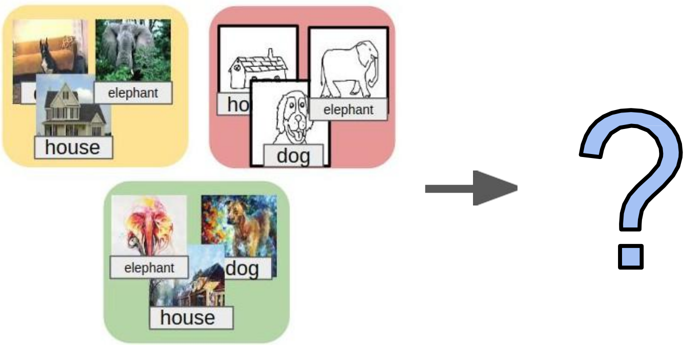
Without target data at training time

Domain Generalization

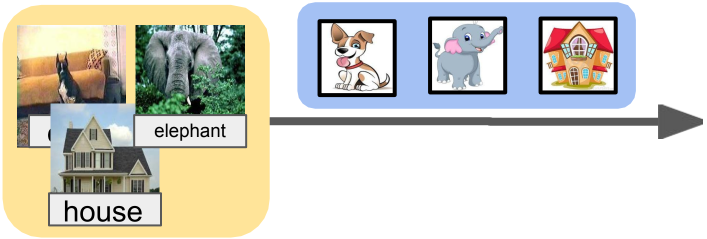


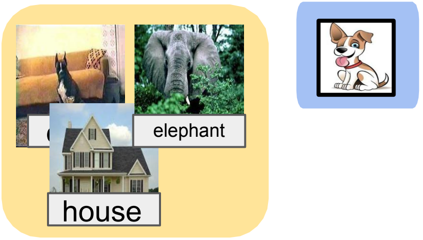
Without target data at training time

Domain Generalization



Continuous DA





Single Sample DA

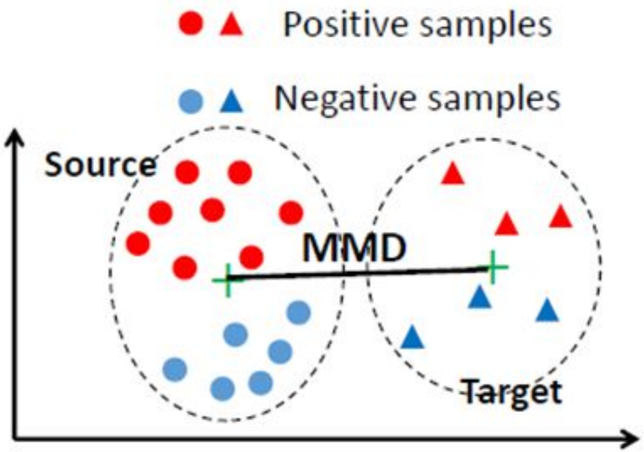
Measuring and Closing the Domain Shift

Maximum Mean Discrepancy (MMD): Distance between embeddings of the probability distributions in a reproducing kernel Hilbert space.

$\mathcal{H}$ Reproducing Kernel Hilbert space

$$h(x) = \langle h, \phi(x) \rangle_{\mathcal{H}}$$

$$\mathrm{MMD}^2 = \left\| \frac{1}{n^s} \sum_{i=1}^{n^s} \phi(x_i^s) - \frac{1}{n^t} \sum_{i=1}^{n^t} \phi(x_i^t) \right\|^2$$

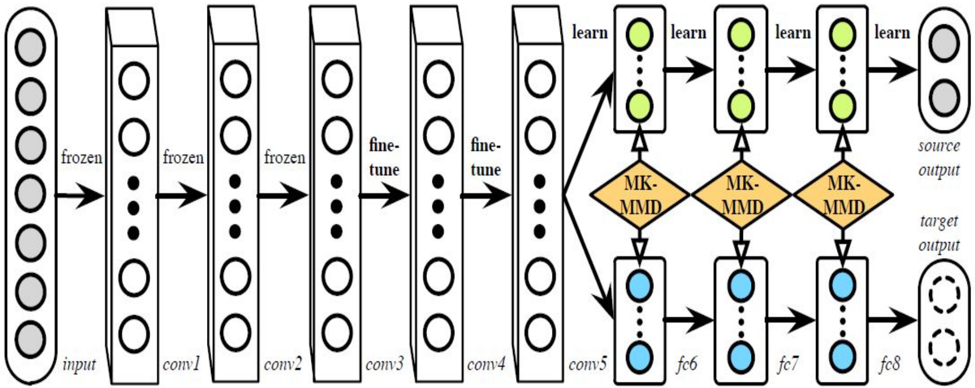


$$P^s(Y|X) = P^t(Y|X)$$ **Covariate Shift Assumption**

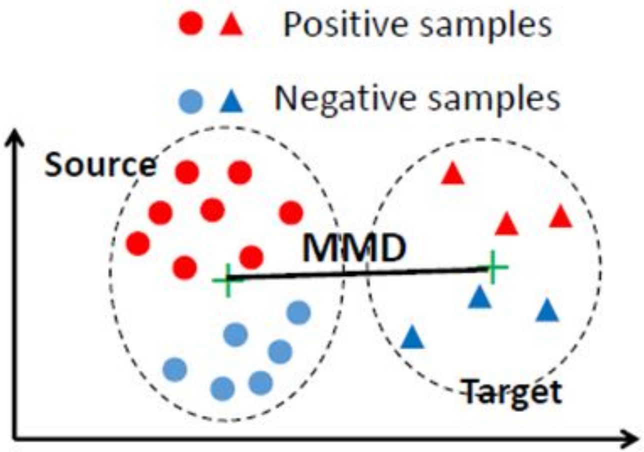
Gretton et al, NIPS 2007

Measuring and Closing the Domain Shift

Maximum Mean Discrepancy (MMD): Distance between embeddings of the probability distributions in a reproducing kernel Hilbert space.



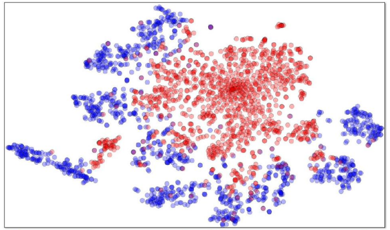
Long et al, ICML 2015



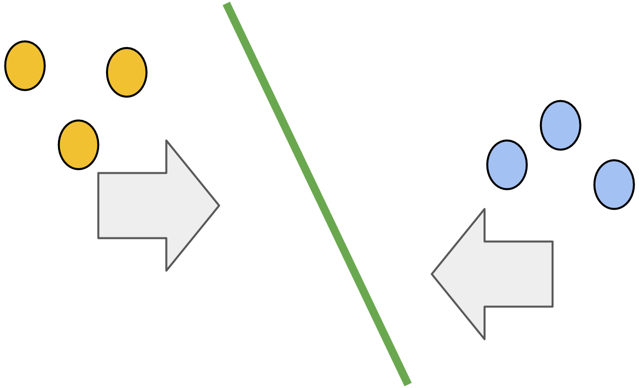
Measuring and Closing the Domain Shift

Adversarial Domain Classification

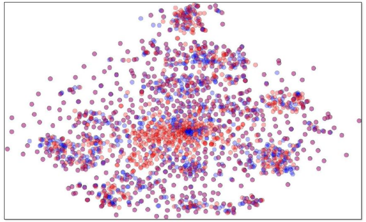
When trained on source only, feature distributions do not match:

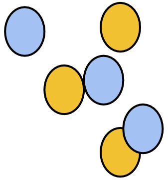


$$S(\mathbf{f}) = \{G_f(\mathbf{x};\theta_f)\,|\,\mathbf{x}{\sim}S(\mathbf{x})\}$$
$$T(\mathbf{f}) = \{G_f(\mathbf{x};\theta_f)\,|\,\mathbf{x}{\sim}T(\mathbf{x})\}$$

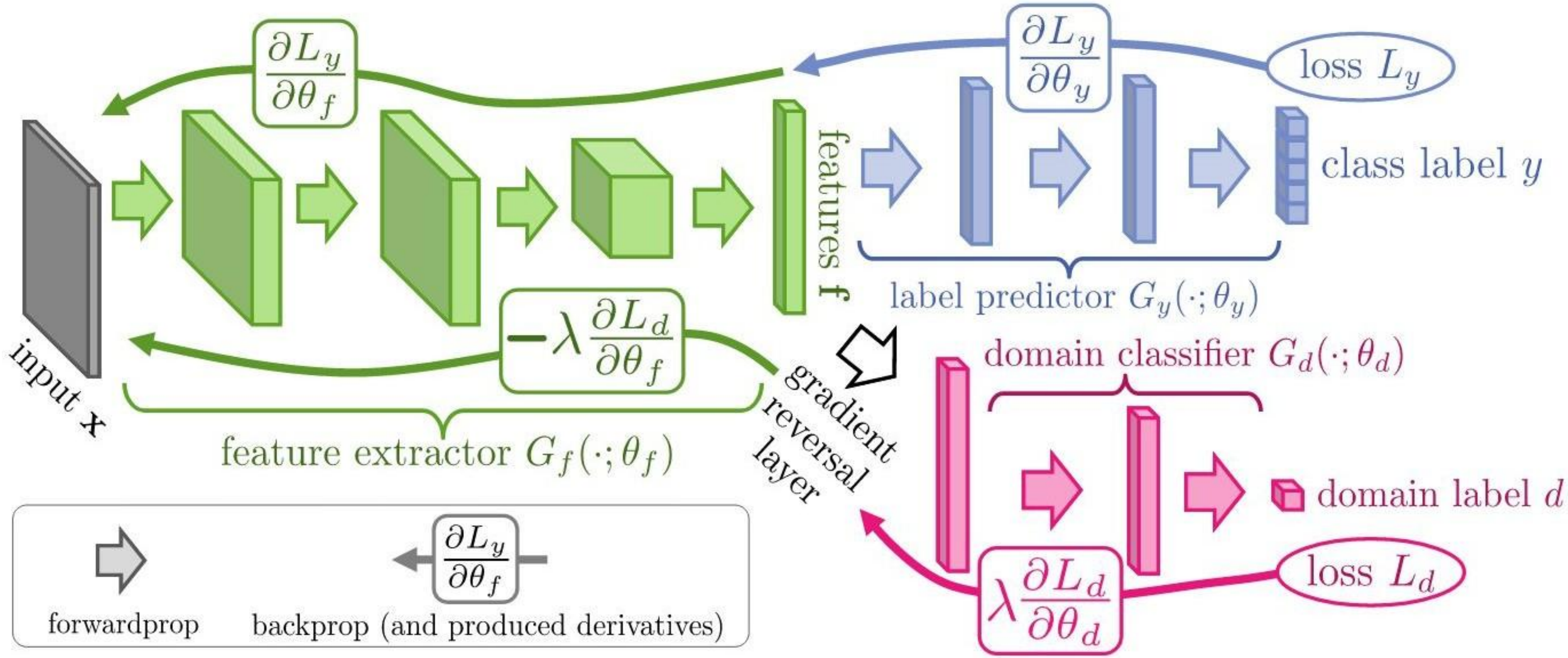


Our goal (after adaptation):





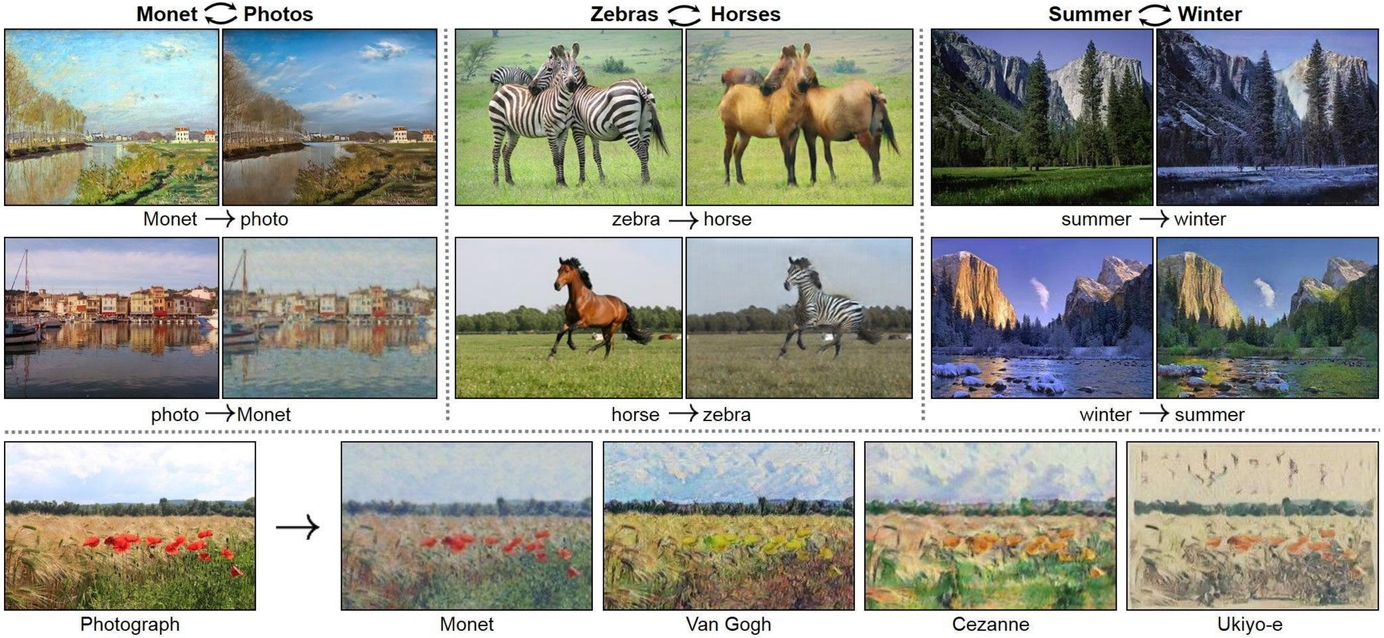
Measuring and Closing the Domain Shift



Ganin, Y., et al. “Domain-adversarial training of neural networks”. The Journal of Machine Learning Research, 17(1), 2096-2030, 2016..

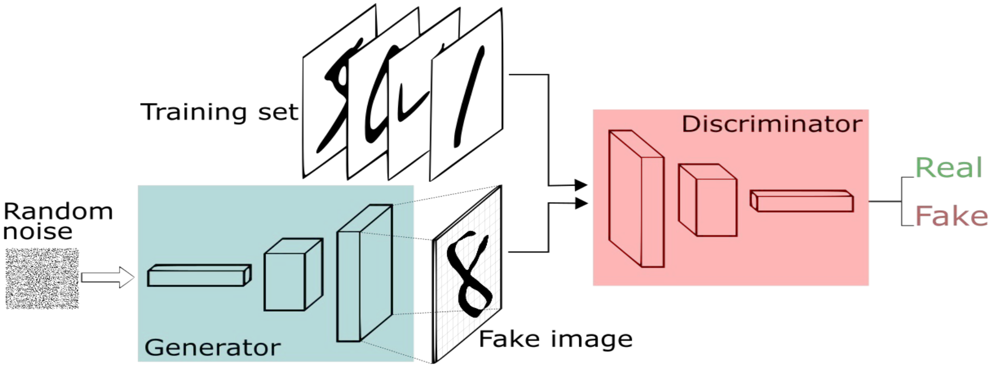
Measuring and Closing the Domain Shift

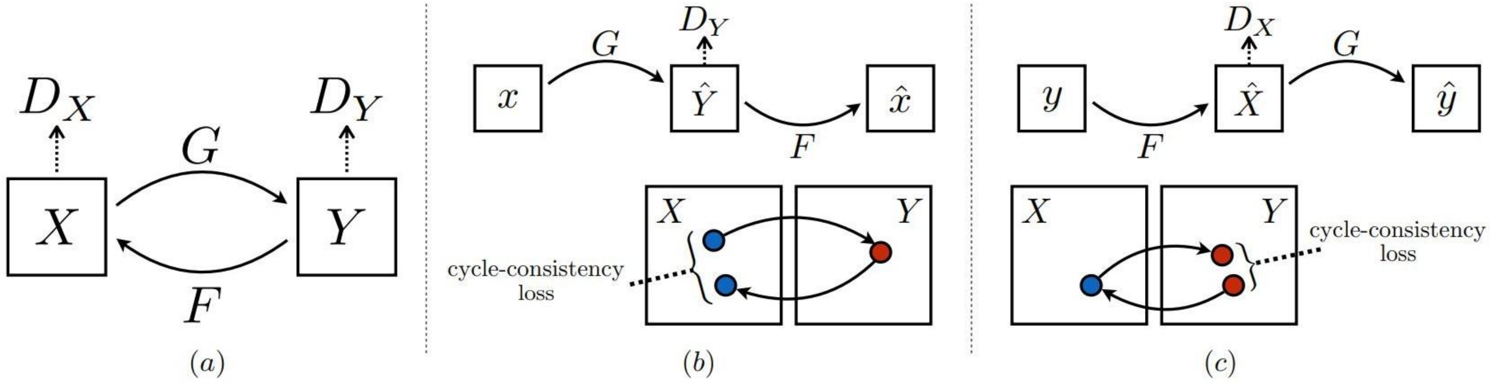
Visual (Pixel-Level) Adaptation



Zhu, J. Y., Park, T., Isola, P., & Efros, A.. “Unpaired image-to-image translation using cycle-consistent adversarial networks”. In ICCV 2017.

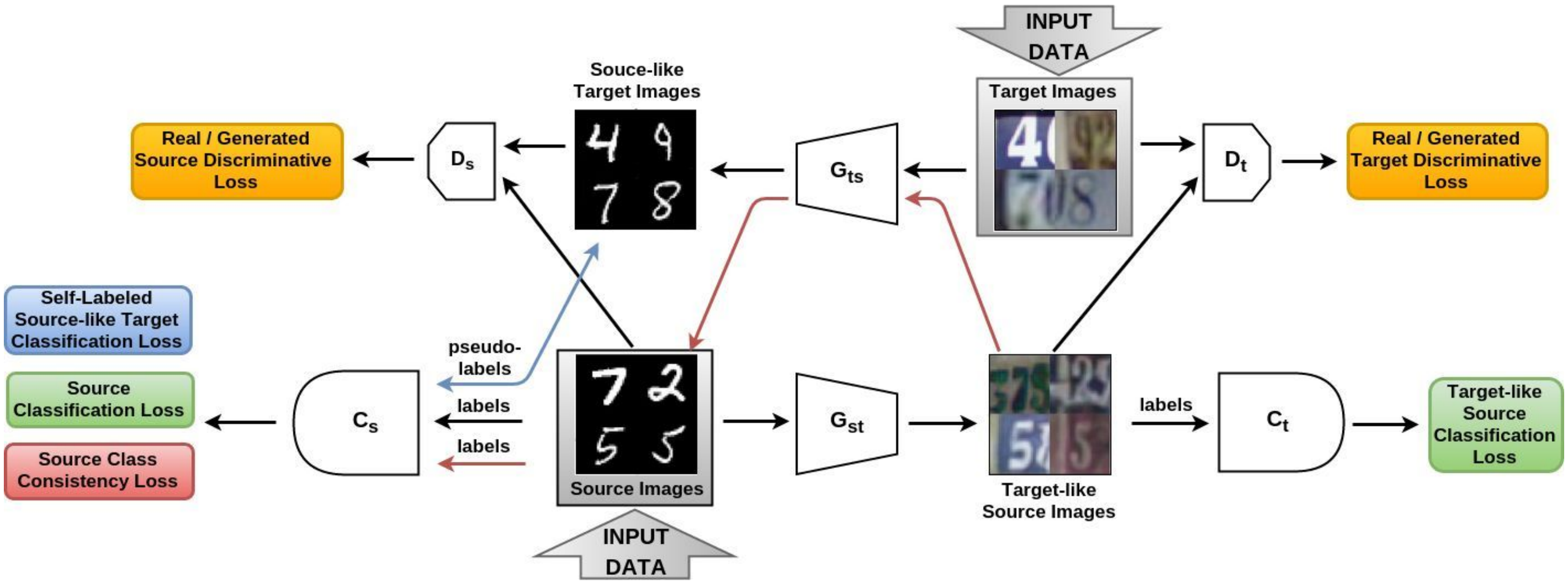
Measuring and Closing the Domain Shift





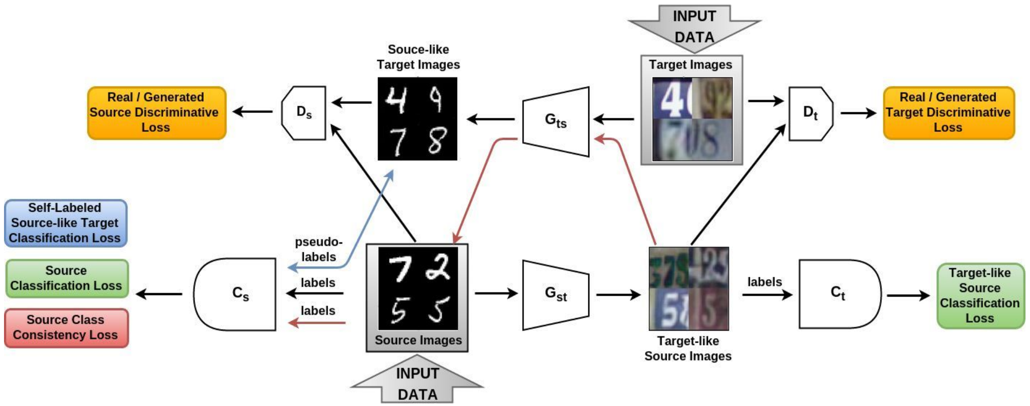
Zhu, J. Y., Park, T., Isola, P., & Efros, A.. “Unpaired image-to-image translation using cycle-consistent adversarial networks”. In ICCV 2017.

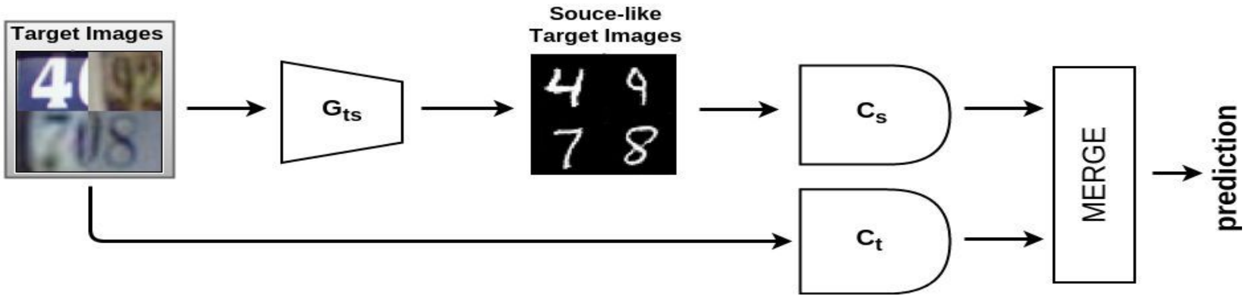
Measuring and Closing the Domain Shift



Russo, P., Carlucci, F. M., Tommasi, T., & Caputo, B. “From source to target and back: symmetric bi-directional adaptive gan”. In CVPR 2018.

Measuring and Closing the Domain Shift





Russo, P., Carlucci, F. M., Tommasi, T., & Caputo, B. “From source to target and back: symmetric bi-directional adaptive gan”. In CVPR 2018.

Self-Supervision

- A form of unsupervised learning where **the data provides its own supervision**
- In general, **withhold some part of the data**, and ask the network to predict it

- The task defines a proxy loss, and the network is forced to learn what we really care about, e.g. a semantic representation, in order to solve it

Zhang et al, Colorful Image Colorization. In ECCV 2016

Pathak et al, Image Inpainting, CVPR 2016

Self-Supervision: Why?

- A form of unsupervised learning where **the data provides its own supervision**
- In general, **withhold some part of the data**, and ask the network to predict it

- some areas are supervision starved
- availability of vast numbers of unlabelled images / videos
- how infants may learn

The Scientist in the Crib: What early learning tells us about the mind. Alison Gopnik et al

The Development of Embodied Cognition: Six Lessons from Babies. Linda Smith and Michael Gasser

Slide Credit: A. Zisserman

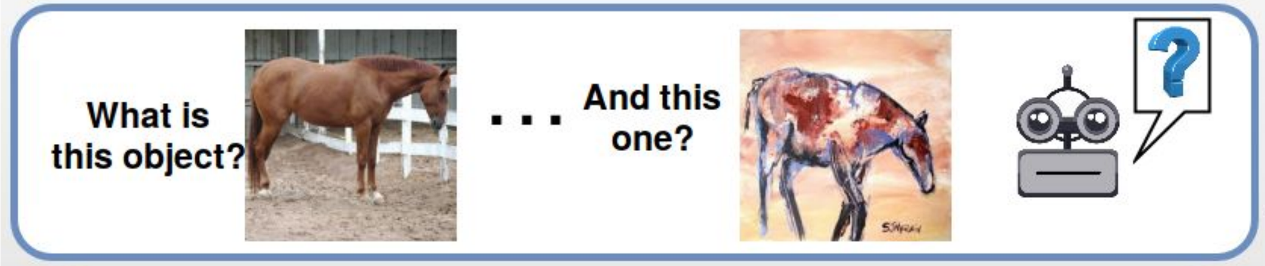
Visual Generalization by Solving Jigsaw Puzzles

- Decompose an image in patches
- Shuffle them = remove their spatial co-location

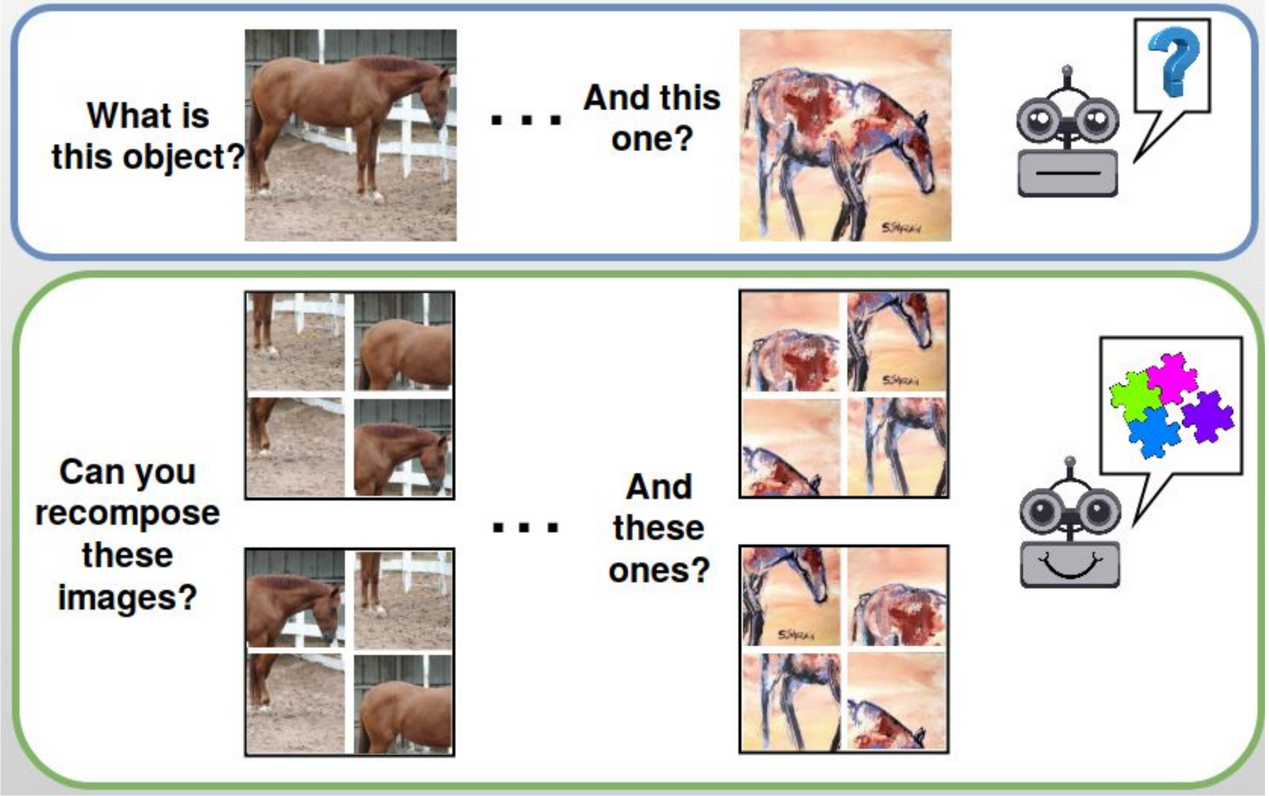
Visual Generalization by Solving Jigsaw Puzzles

Visual Generalization by Solving Jigsaw Puzzles

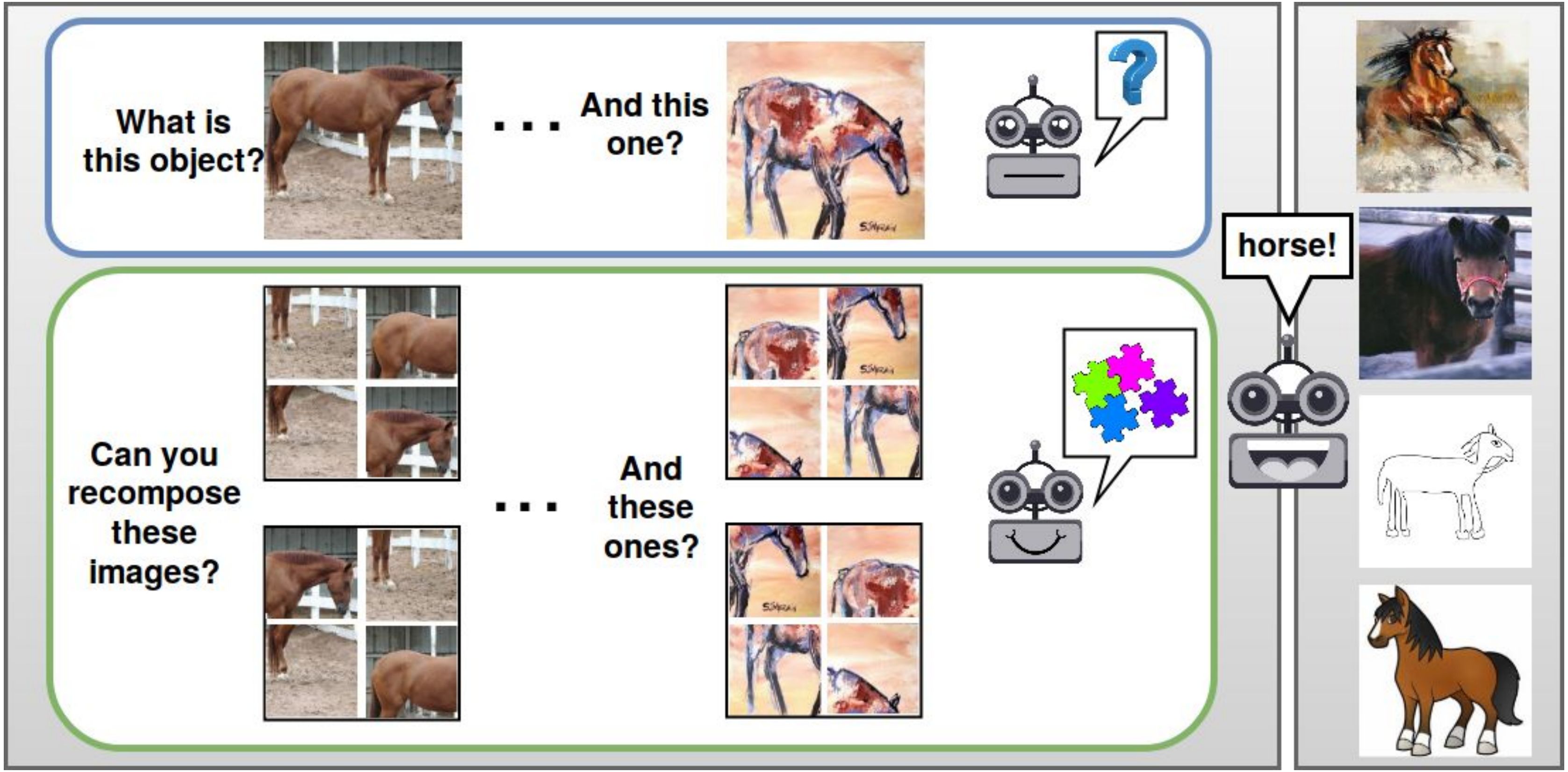
Visual Generalization by Solving Jigsaw Puzzles



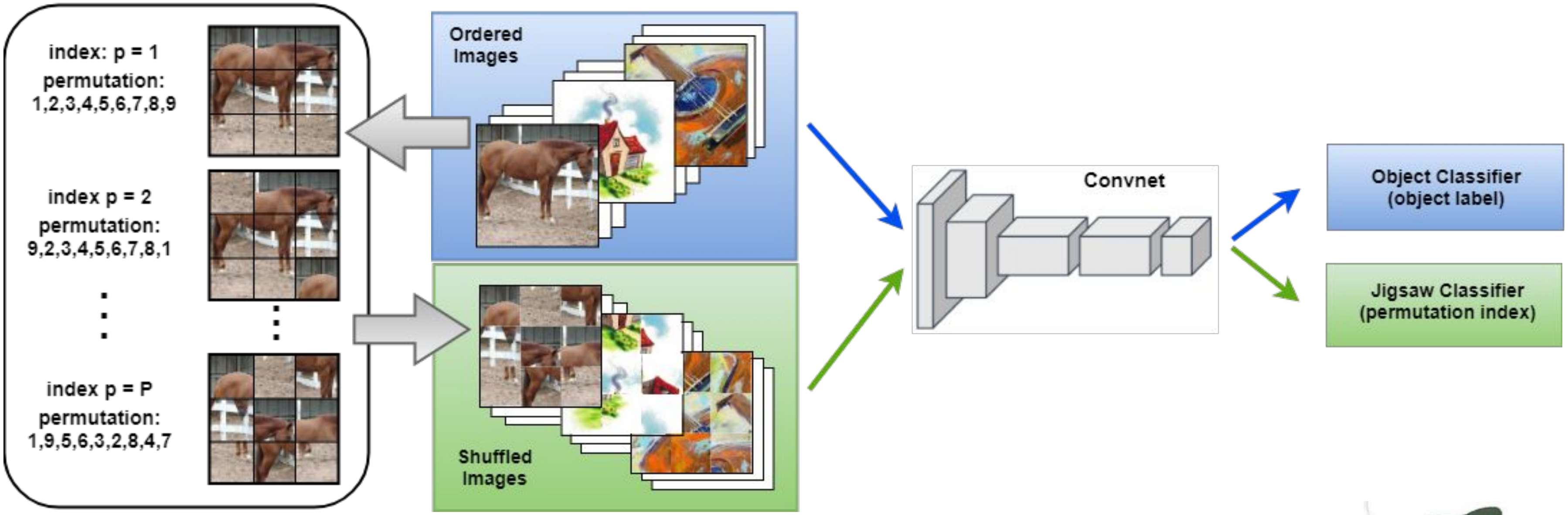
Visual Generalization by Solving Jigsaw Puzzles



Visual Generalization by Solving Jigsaw Puzzles



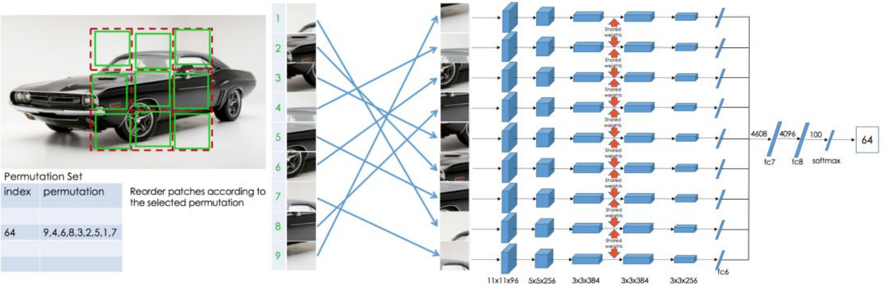
Visual Generalization by Solving Jigsaw Puzzles

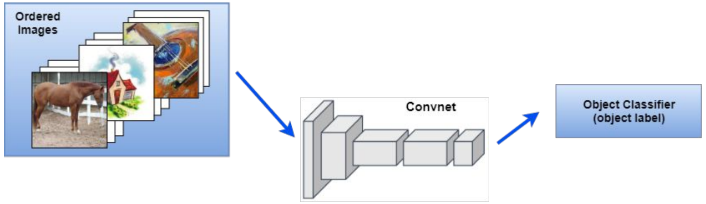


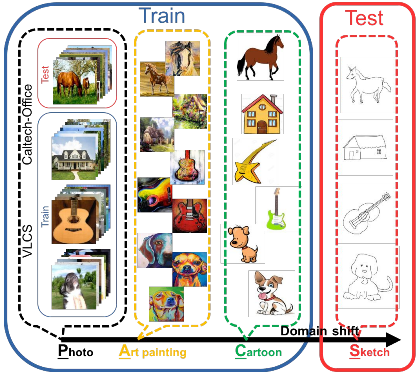
Jigsaw puzzle **Gen**eralization (JiGen)

"Domain Generalization by Solving Jigsaw Puzzle", IEEE CVPR 2019 (Oral)

From Features to Images & Multi-Task

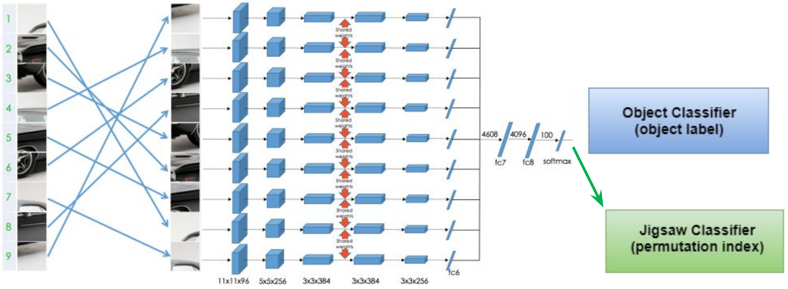


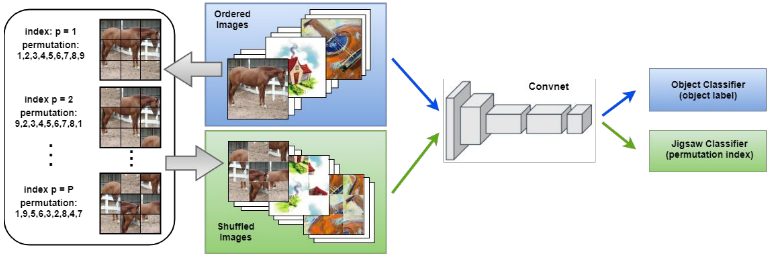


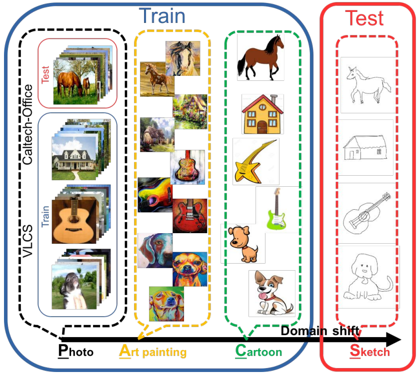


PACS	art_paint	cartoon	sketch	photo	Avg.
C-CFN-Deep All	59.69	59.88	45.66	85.42	62.66
Deep All	63.30	63.13	54.07	87.70	67.05

From Features to Images & Multi-Task

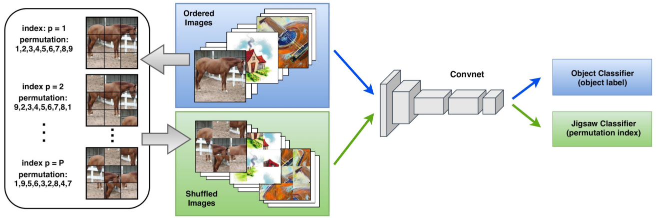






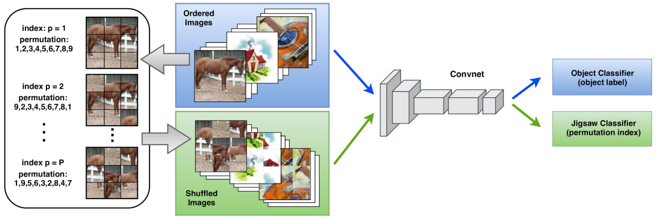
PACS	art_paint	cartoon	sketch	photo	Avg.
C-CFN-Deep All	59.69	59.88	45.66	85.42	62.66
C-CFN-**JiGen**	60.68	60.55	55.66	82.68	**64.89**
Deep All	63.30	63.13	54.07	87.70	67.05
JiGen	67.63	71.71	65.18	89.00	**73.38**

More Results

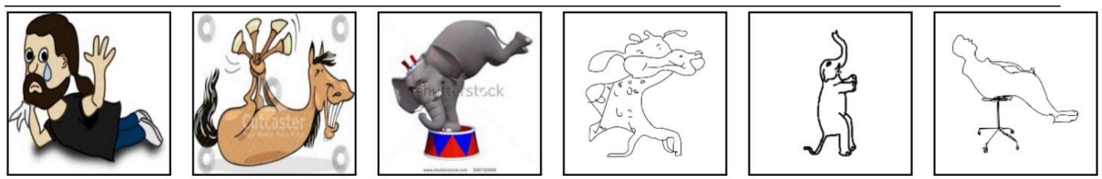


PACS	art_paint	cartoon	sketch	photo	Avg.
C-CFN-Deep All	59.69	59.88	45.66	85.42	62.66
C-CFN-**JiGen**	60.68	60.55	55.66	82.68	**64.89**
Deep All	63.30	63.13	54.07	87.70	67.05
JiGen	67.63	71.71	65.18	89.00	**73.38**
AAAI 2018	**66.23**	**66.88**	**58.96**	**88.00**	70.02
GCPR 2018	63.87	70.70	64.66	85.55	71.20

More Results

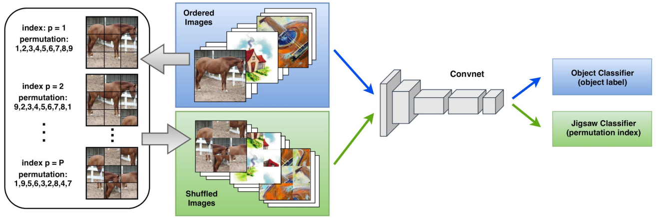


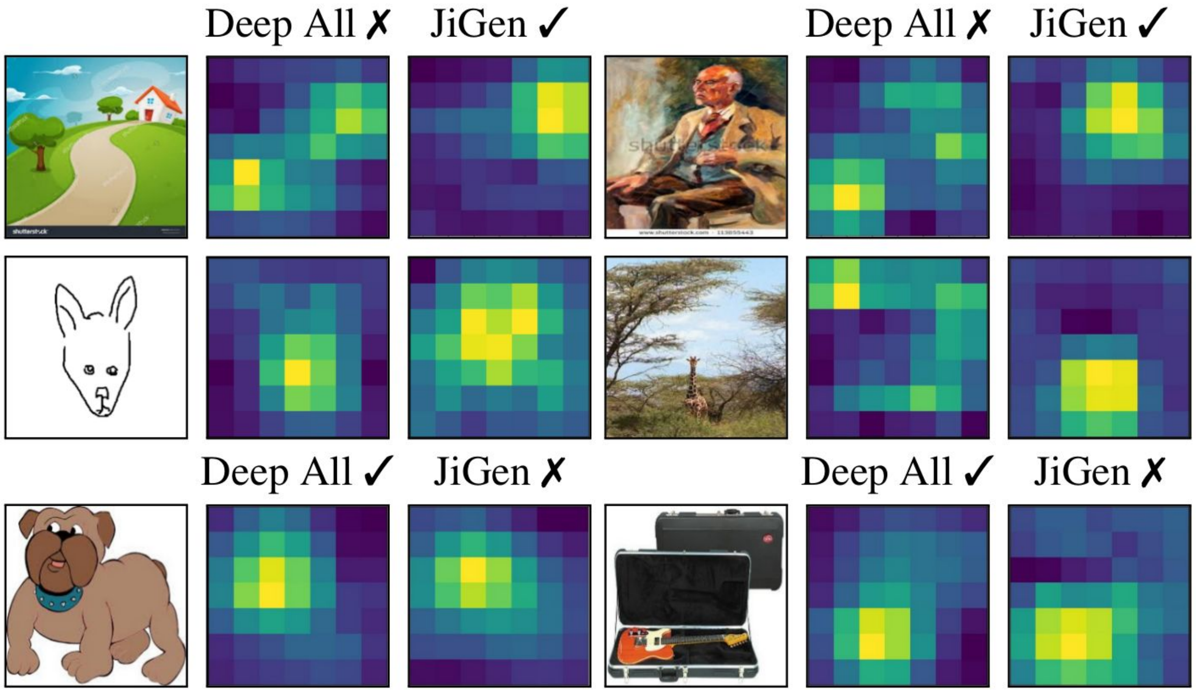
PACS	art_paint	cartoon	sketch	photo	Avg.
C-CFN-Deep All	59.69	59.88	45.66	85.42	62.66
C-CFN-**JiGen**	60.68	60.55	55.66	82.68	**64.89**
Deep All	63.30	63.13	54.07	87.70	67.05
JiGen	67.63	71.71	65.18	89.00	**73.38**
AAAI 2018	**66.23**	**66.88**	**58.96**	**88.00**	70.02
GCPR 2018	63.87	70.70	64.66	85.55	71.20
Rotation	**67.67**	**69.83**	**61.04**	**89.98**	71.52



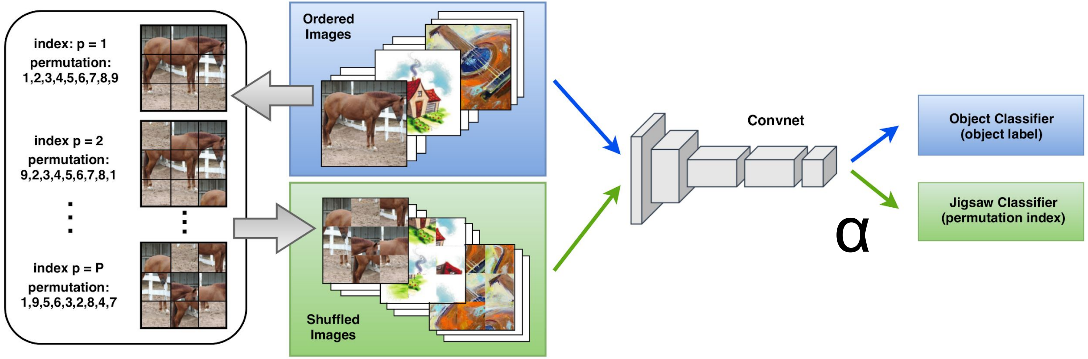
Gidaris et al, “Unsupervised Representation Learning by predicting Image Rotations”. In CVPR 2018.

More Results



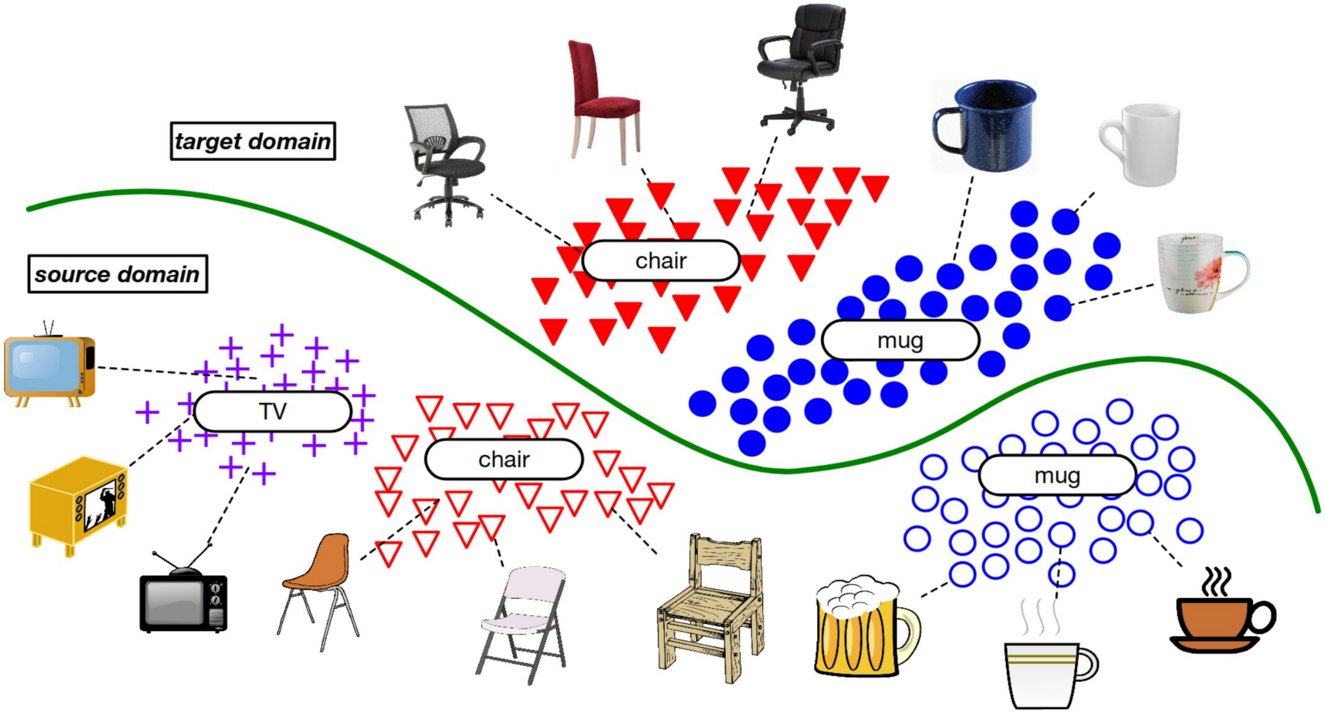


Domain Adaptation



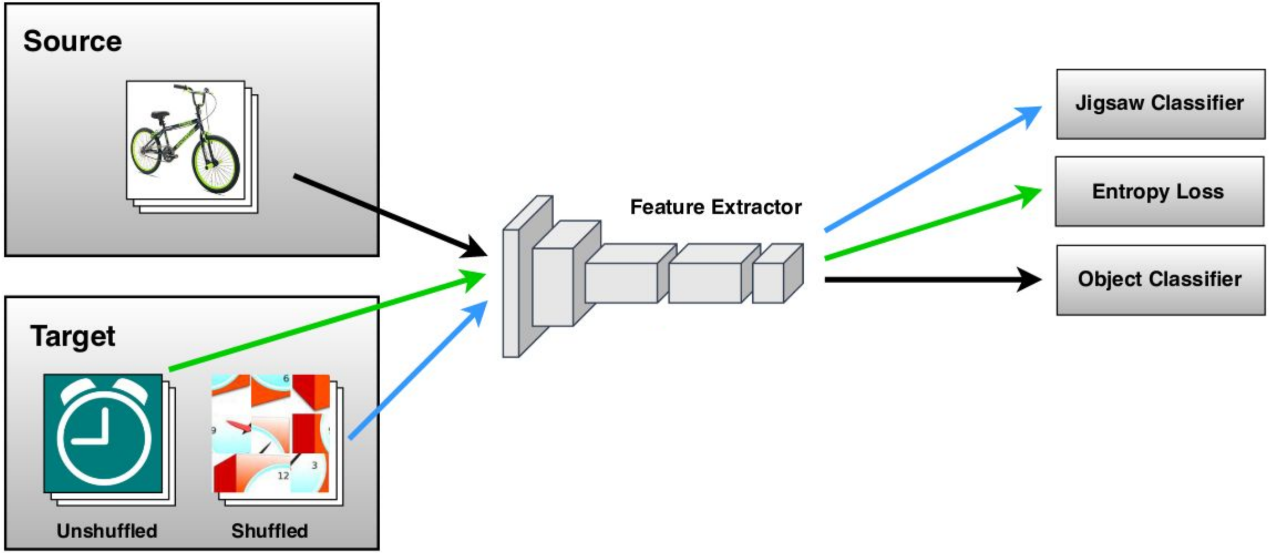
	PACS-DA	art_paint.	cartoon	sketches	photo	Avg.
	Resnet-18					
	Deep All	74.70	72.40	60.10	92.90	75.03
[31]	Dial	87.30	85.50	66.80	97.00	84.15
	DDiscovery	**87.70**	**86.90**	69.60	97.00	85.30
	Deep All	77.85	74.86	67.74	95.73	79.05
	JiGen $_{\alpha^s=\alpha^t=0.7}$	84.88	81.07	**79.05**	**97.96**	**85.74**
	JiGen $_{\alpha^t=0.1}$	85.58	82.18	78.61	98.26	86.15
	JiGen $_{\alpha^t=0.3}$	85.08	81.28	81.50	97.96	86.46
	JiGen $_{\alpha^t=0.5}$	85.73	82.58	78.34	98.10	86.19
	JiGen $_{\alpha^t=0.9}$	85.32	80.56	79.93	97.63	85.86

Partial Domain Adaptation



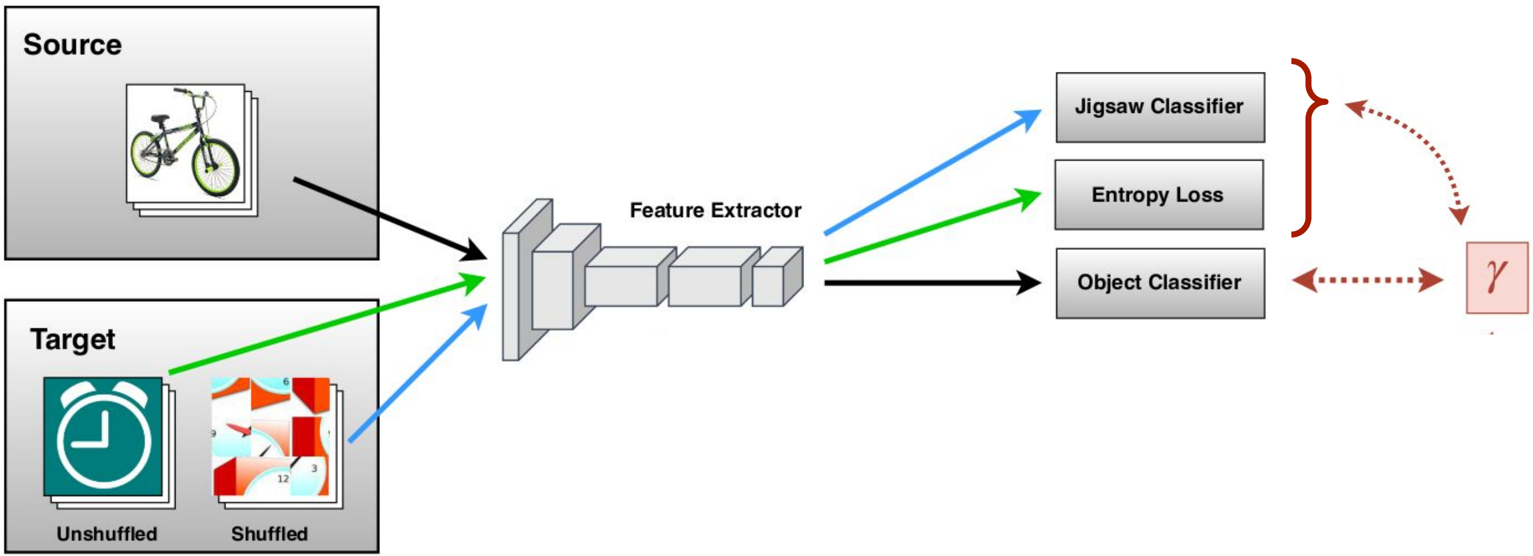
Cao et al, ECCV 2018

Partial Domain Adaptation

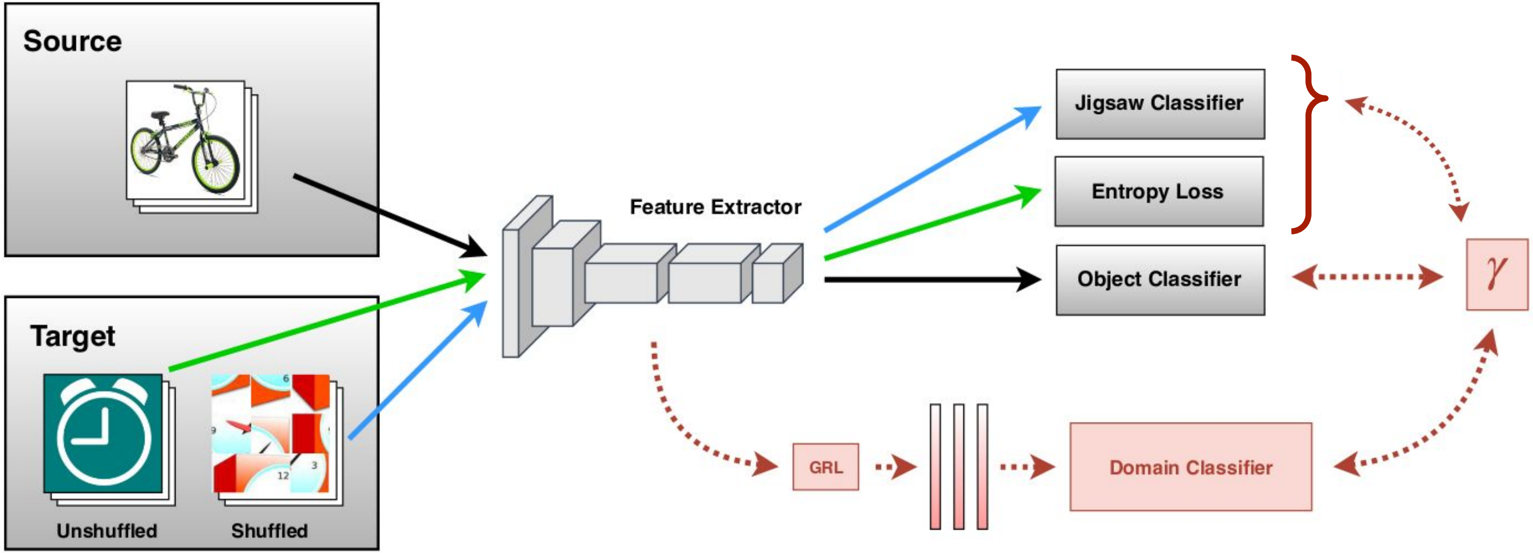


S. Bucci, A. D’Innocente, T. Tommasi “Tackling Partial Domain Adaptation with Self Supervision”. In ICIAP 2019

Partial Domain Adaptation



Partial Domain Adaptation

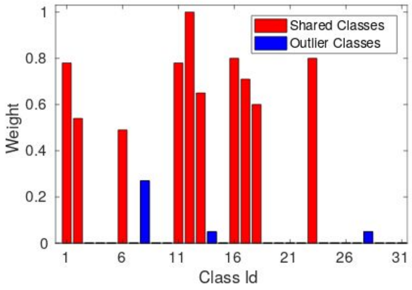


Partial Domain Adaptation

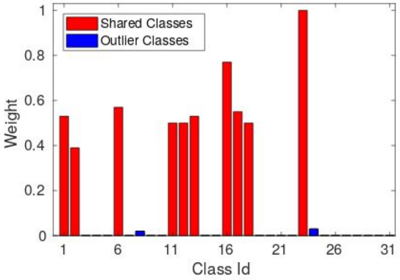
	Office-31						
	A→W	D→W	W→D	A →D	D→A	W→A	Avg.
Resnet-50	75.37	94.13	98.84	79.19	81.28	85.49	85.73
DAN[15]	59.32	73.90	90.45	61.78	74.95	67.64	71.34
DANN[10]	75.56	96.27	98.73	81.53	82.78	86.12	86.50
ADDA[27]	75.67	95.38	99.85	83.41	83.62	84.25	87.03
RTN[16]	78.98	93.22	85.35	77.07	89.25	89.46	85.56
IWAN [32]	89.15	99.32	99.36	90.45	**95.62**	94.26	94.69
SAN [3]	93.90	99.32	99.36	94.27	94.15	88.73	94.96
PADA[4]	86.54	**99.32**	**100**	82.17	92.69	**95.41**	92.69
TWIN [20]	86.00	99.30	**100**	86.80	94.70	94.50	93.60
JiGen [5]	92.88	92.43	98.94	89.6	84.06	92.94	91.81
SSPDA	91.52	92.88	98.94	90.87	90.61	94.36	93.20
SSPDA-γ	99.32	94.69	99.36	96.39	86.36	94.22	95.06
SSPDA-PADA	**99.66**	94.46	99.57	**97.67**	87.33	94.26	**95.49**

Partial Domain Adaptation

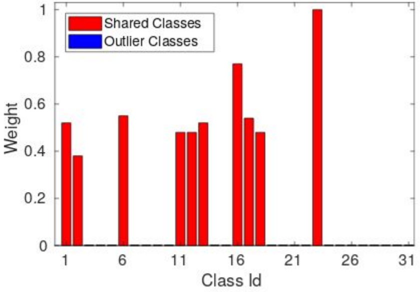
	Office-31						
	A$\rightarrow$W	**D$\rightarrow$W**	**W$\rightarrow$D**	**A $\rightarrow$D**	**D$\rightarrow$A**	**W$\rightarrow$A**	**Avg.**
Resnet-50	75.37	94.13	98.84	79.19	81.28	85.49	85.73
DAN[15]	59.32	73.90	90.45	61.78	74.95	67.64	71.34
DANN[10]	75.56	96.27	98.73	81.53	82.78	86.12	86.50
ADDA[27]	75.67	95.38	99.85	83.41	83.62	84.25	87.03
RTN[16]	78.98	93.22	85.35	77.07	89.25	89.46	85.56
IWAN [32]	89.15	99.32	99.36	90.45	**95.62**	94.26	94.69
SAN [3]	93.90	99.32	99.36	94.27	94.15	88.73	94.96
PADA[4]	86.54	**99.32**	**100**	82.17	92.69	**95.41**	92.69
TWIN [20]	86.00	99.30	**100**	86.80	94.70	94.50	93.60
JiGen [5]	92.88	92.43	98.94	89.6	84.06	92.94	91.81
SSPDA	91.52	92.88	98.94	90.87	90.61	94.36	93.20
SSPDA-γ	99.32	94.69	99.36	96.39	86.36	94.22	95.06
SSPDA-PADA	**99.66**	94.46	99.57	**97.67**	87.33	94.26	**95.49**



PADA



SSPDA-γ



SSPDA-PADA

Take Home Message

- Deep Learning
  - ❖ powerful but data-hungry
  - ❖ not robust across domains, lacks in generalization

- Domain Adaptation Techniques
  - ❖ allow recognition across domains
  - ❖ need target unlabeled data at training time

- Domain Generalization
  - ❖ no need of target data at training time
  - ❖ can deal with one or multiple sources, no need of domain label (mixed sources)

- Self-Supervision
  - ❖ powerful tool to exploit unlabeled data and reduce deep-learning data hunger
  - ❖ powerful tool to support learning across domains, in adaptation and generalization
    ...
  - ❖ improve robustness to label noise, novelty detection...

Take Home Message

- Deep Learning
  - ❖ powerful but data-hungry
  - ❖ not robust across domains, lacks in generalization

- Domain Adaptation Techniques
  - ❖ allow recognition across domains
  - ❖ need target unlabeled data at training time

- Domain Generalization
  - ❖ no need of target data at training time
  - ❖ can deal with one or multiple sources, no need of domain label (mixed sources)

- Self-Supervision
  - ❖ powerful tool to exploit unlabeled data and reduce deep-learning data hunger
  - ❖ powerful tool to support learning across domains, in adaptation and generalization
  - ...
  - ❖ improve robustness to label noise, novelty detection...

Thanks! Questions?