Fermilab

IAIFI Summer School, August 2025

Shift Happens:
Building Robust AI Models with Domain Adaptation

Aleksandra Ciprijanvoic

Associate Scientist, Fermilab
aleksand@fnal.gov

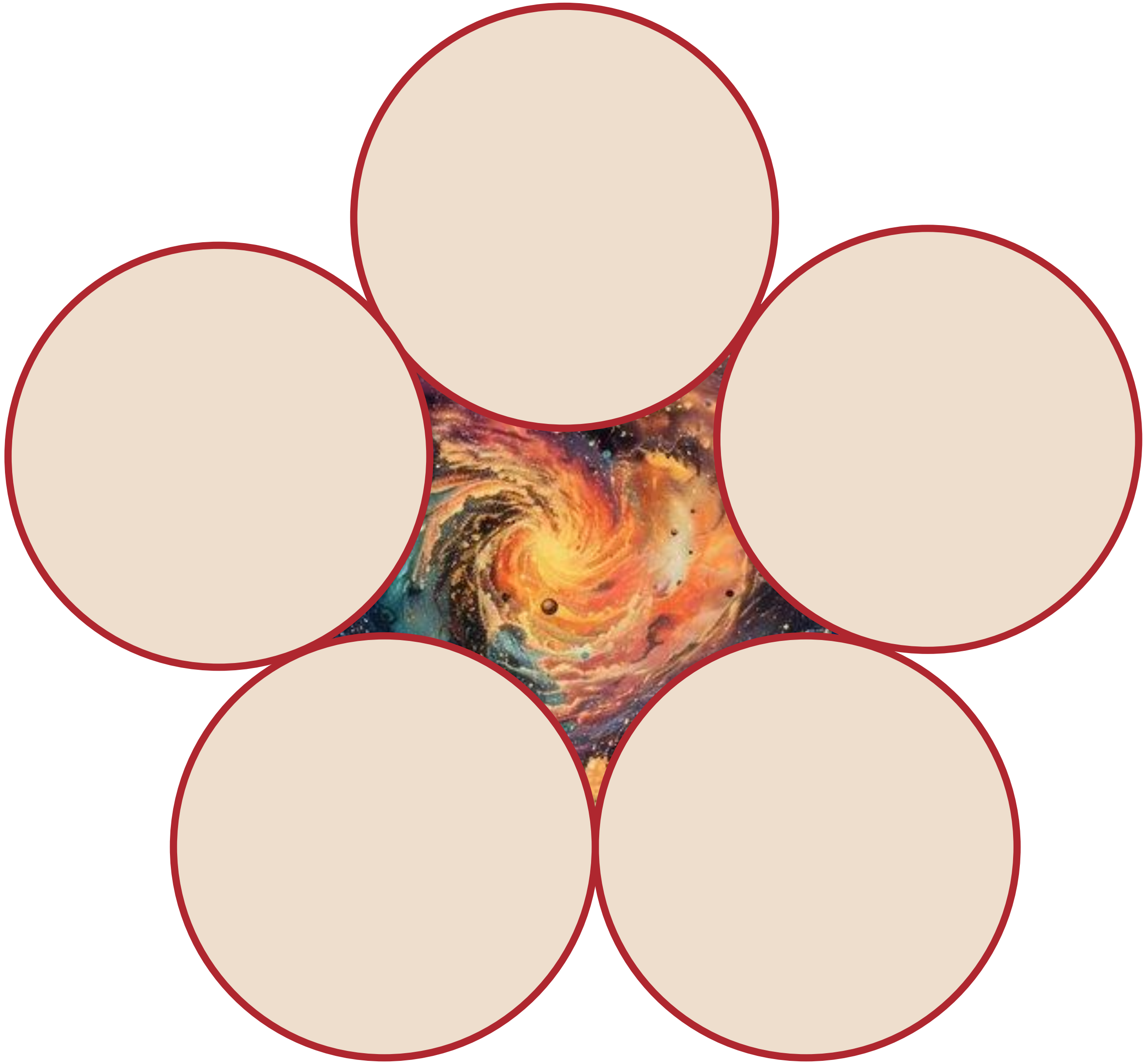
U.S. DEPARTMENT of ENERGY

Fermi National Accelerator Laboratory is managed by
FermiForward for the U.S. Department of Energy Office of Science

FERMI NATIONAL ACCELERATOR LABORATORY

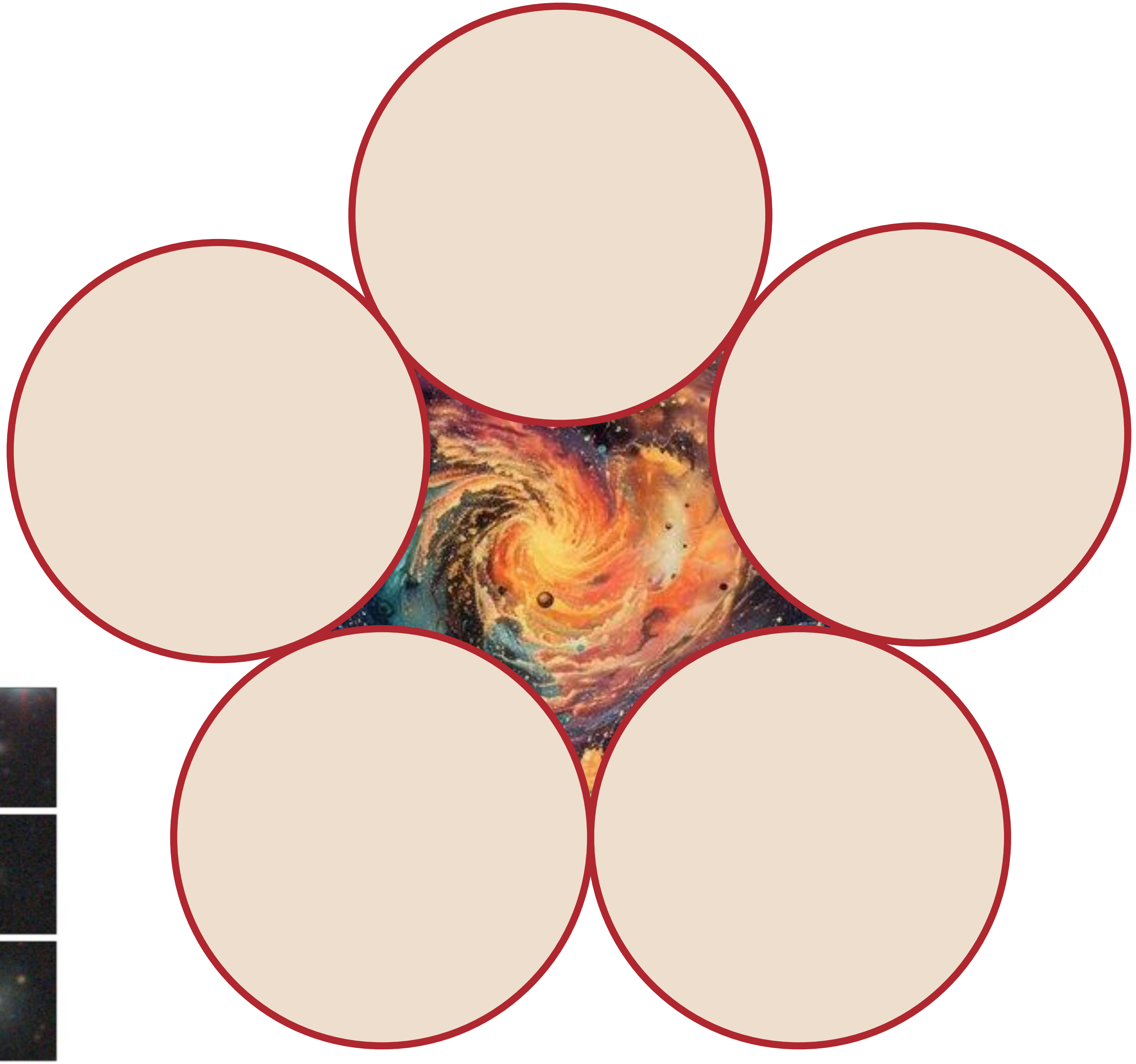
Major challenges for AI applications in Science

As listed in The Dawes Review 10: The impact of deep learning for the analysis of galaxy surveys
[Huertas-Company & Lanusse 2022]



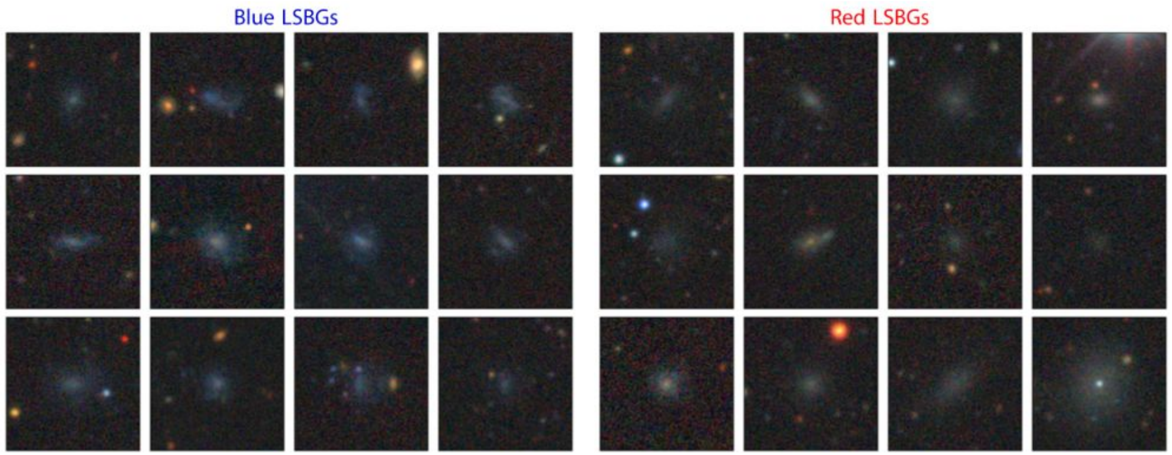
Major challenges for AI applications in Science

As listed in The Dawes Review 10: The impact of deep learning for the analysis of galaxy surveys
[Huertas-Company & Lanusse 2022]



Low Surface Brightness Galaxies (LSBGs)

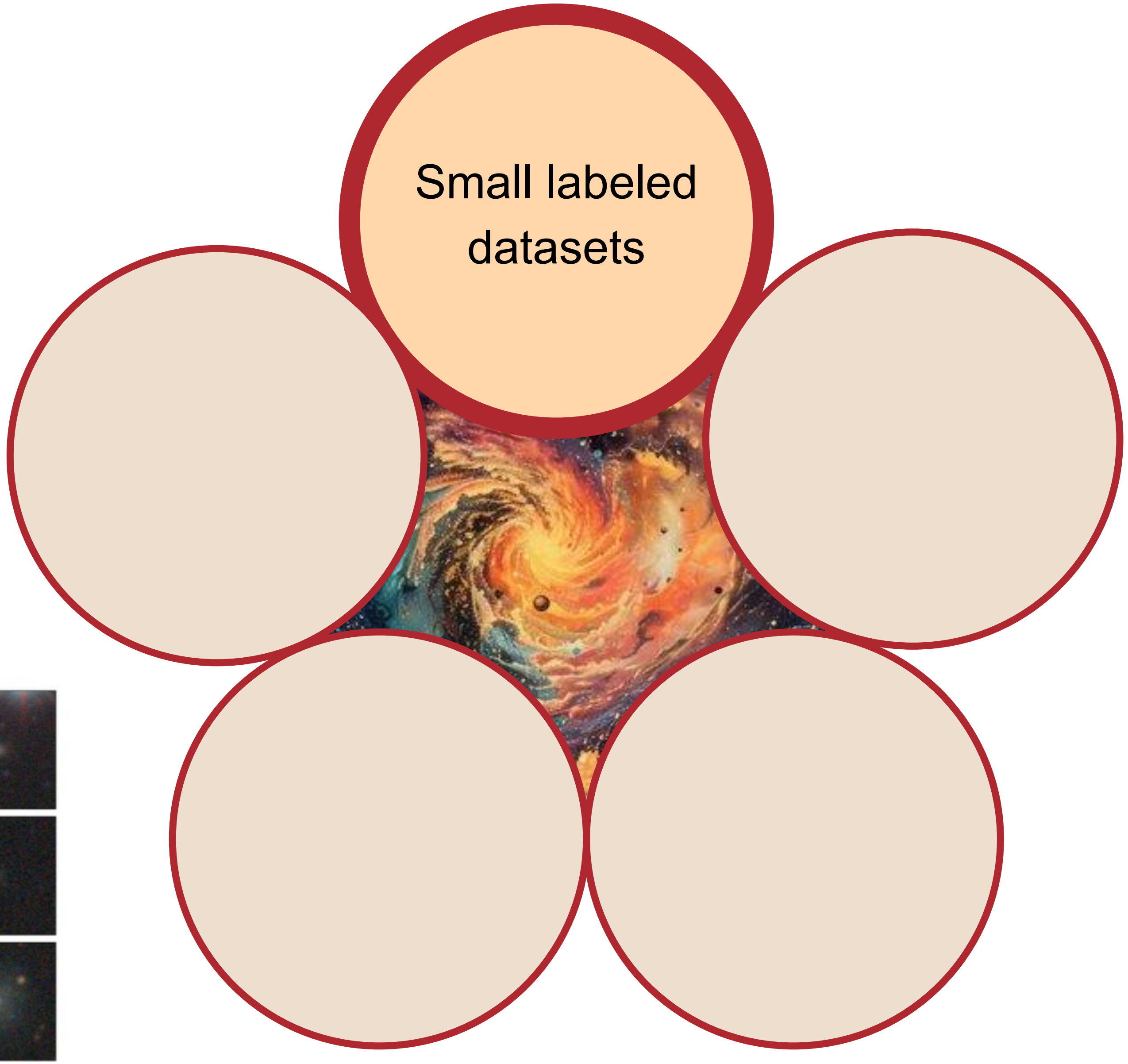
Tanogldis et al. 2021.



Major challenges for AI applications in Science

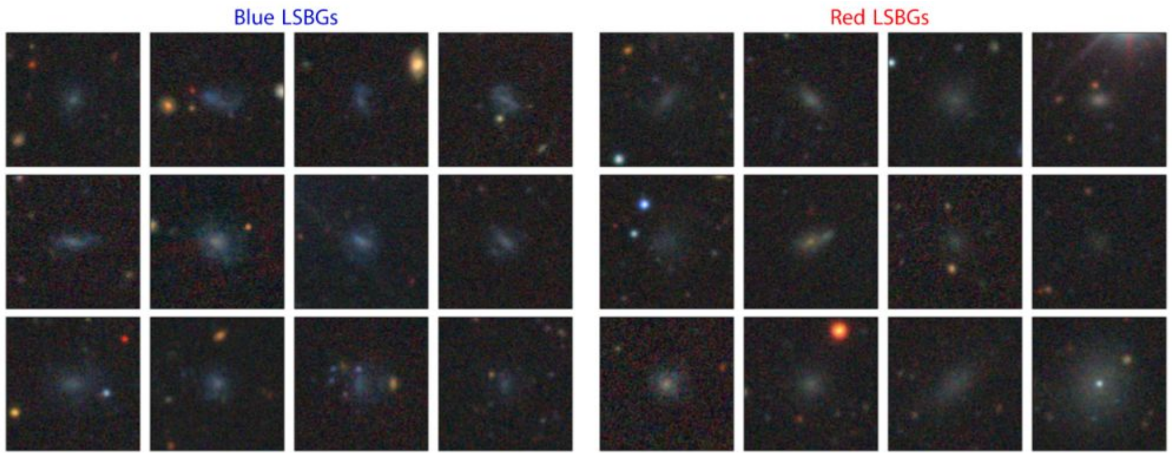
As listed in The Dawes Review 10: The impact of deep learning for the analysis of galaxy surveys
Huertas-Company & Lanusse 2022

I want to train my AI model by I don't have enough trustworthy labels?



Low Surface Brightness Galaxies (LSBGs)

Tanogldis et al. 2021.

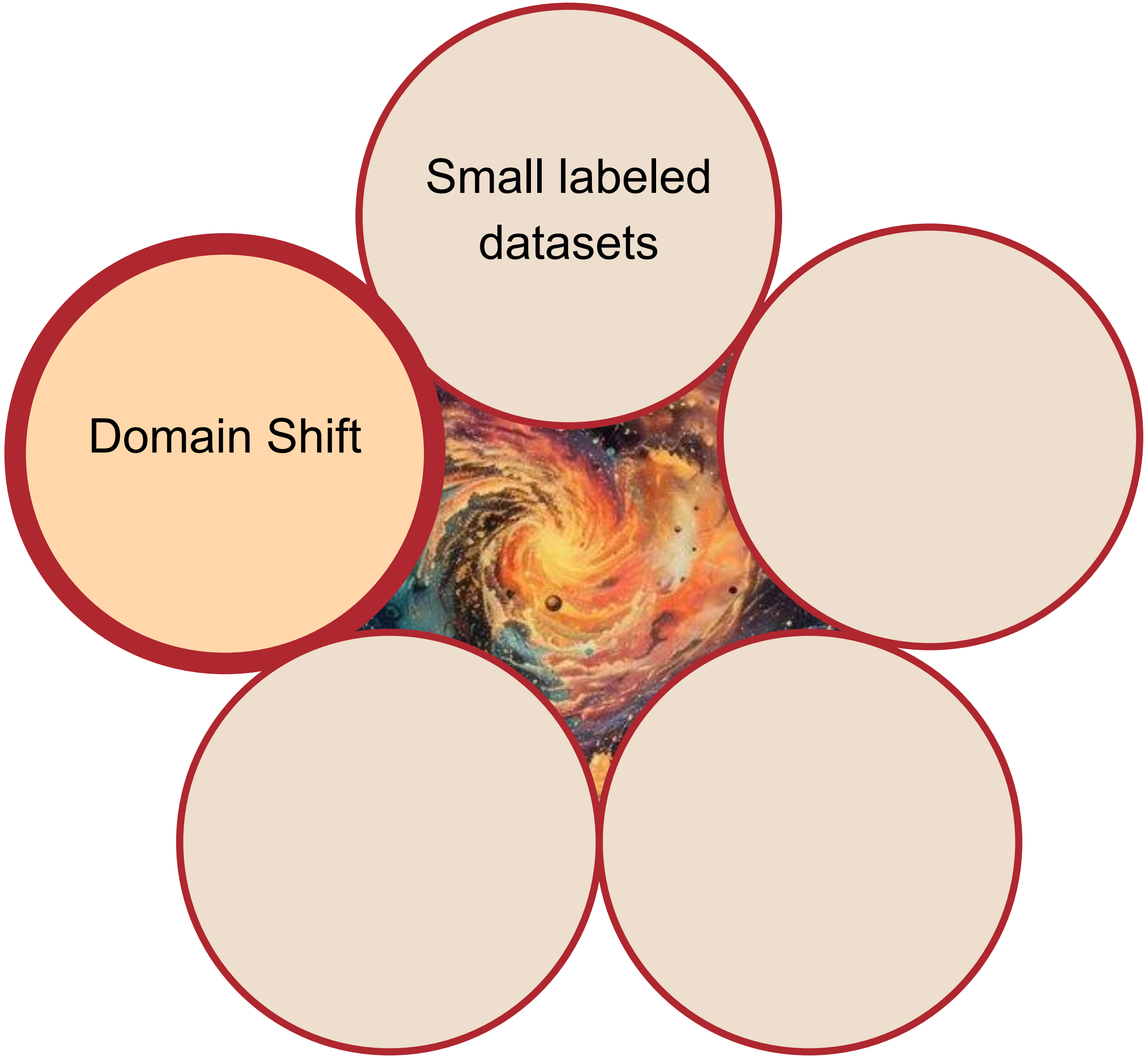


Major challenges for AI applications in Science

As listed in The Dawes Review 10: The impact of deep learning for the analysis of galaxy surveys
[Huertas-Company & Lanusse 2022]

I want to train my AI model by I don’t have enough trustworthy labels?

Can I train on simulations and apply the model to my real data?



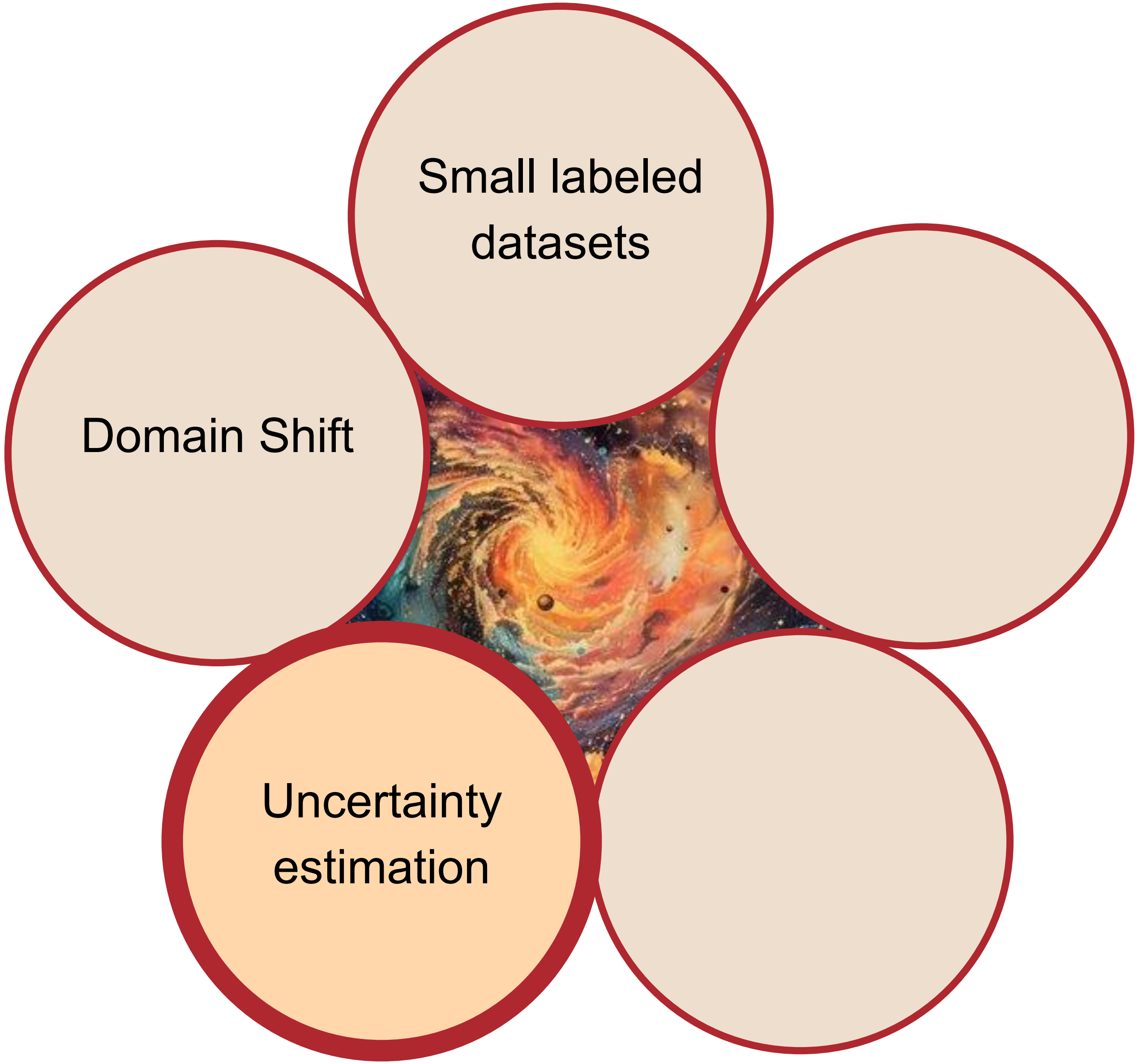
Major challenges for AI applications in Science

As listed in The Dawes Review 10: The impact of deep learning for the analysis of galaxy surveys
[Huertas-Company & Lanusse 2022]

I want to train my AI model by I don’t have enough trustworthy labels?

Can I train on simulations and apply the model to my real data?

How well is my model going to perform?



Major challenges for AI applications in Science

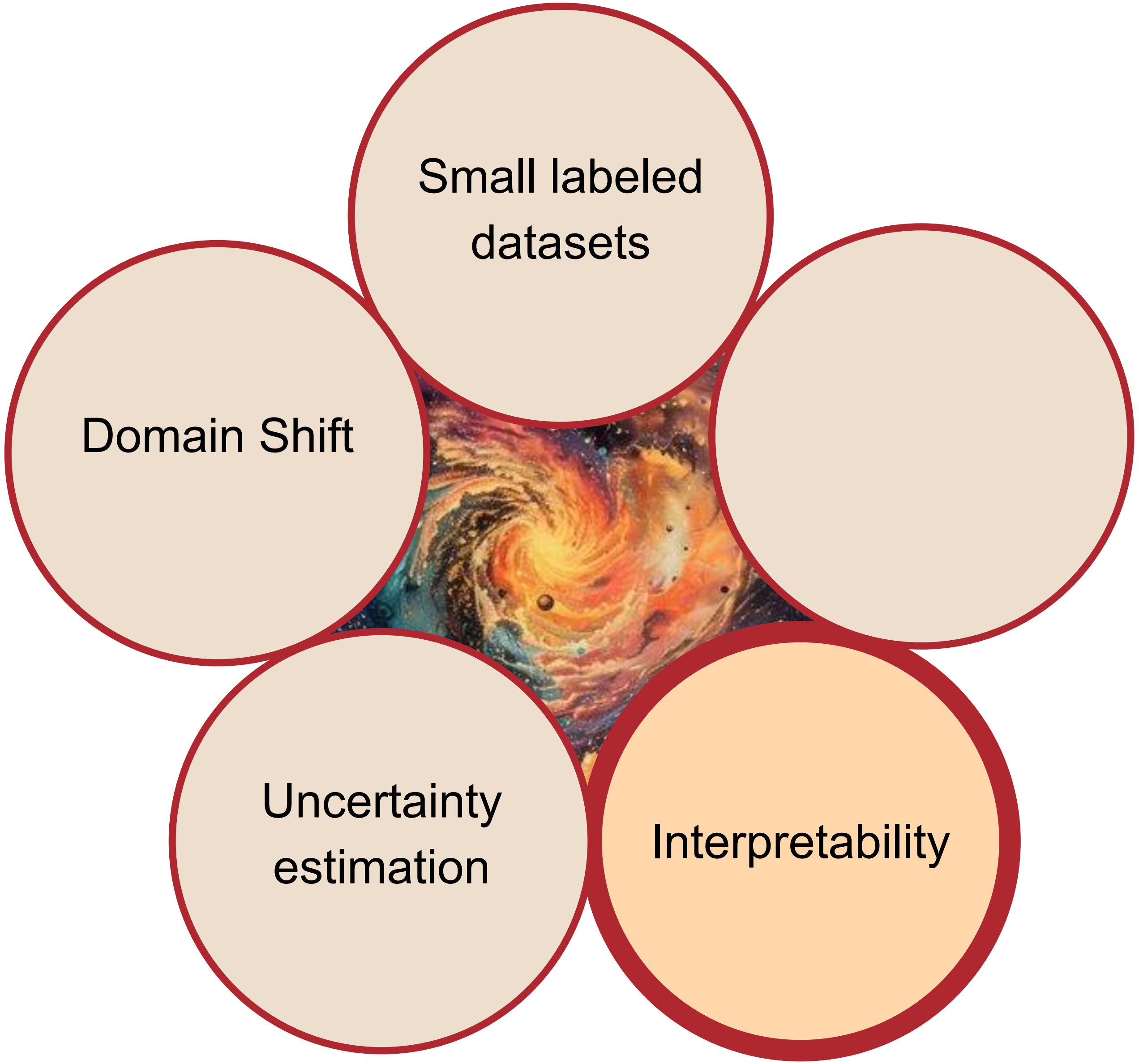
As listed in The Dawes Review 10: The impact of deep learning for the analysis of galaxy surveys
[Huertas-Company & Lanusse 2022]

I want to train my AI model by I don’t have enough trustworthy labels?

Can I train on simulations and apply the model to my real data?

How well is my model going to perform?

Can I understand its performance?



Major challenges for AI applications in Science

As listed in The Dawes Review 10: The impact of deep learning for the analysis of galaxy surveys
[Huertas-Company & Lanusse 2022]

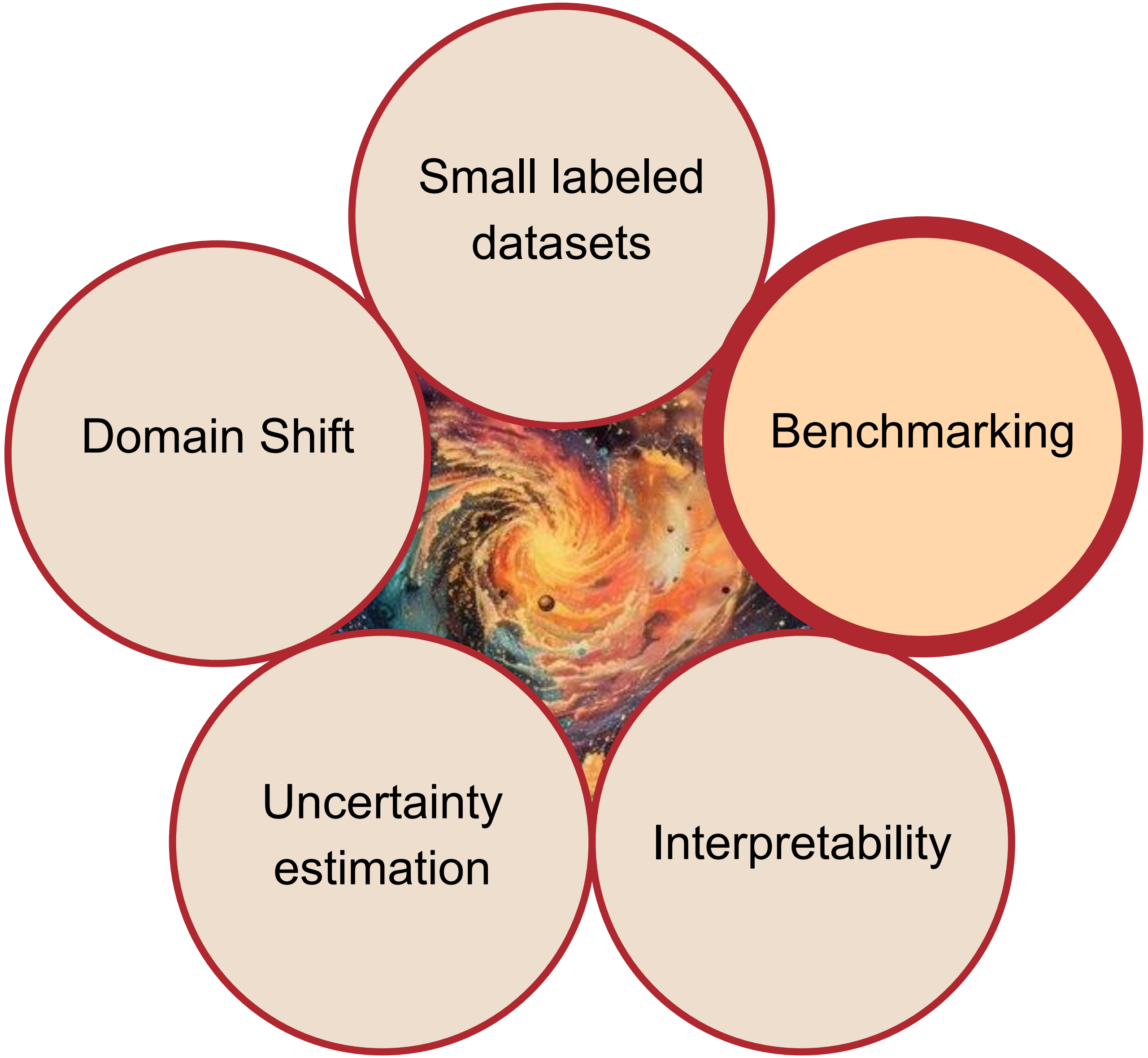
I want to train my AI model by I don’t have enough trustworthy labels?

Can I train on simulations and apply the model to my real data?

How well is my model going to perform?

Can I understand its performance?

Can I compare my model with other researchers?



Major challenges for AI applications in Science

As listed in The Dawes Review 10: The impact of deep learning for the analysis of galaxy surveys
[Huertas-Company & Lanusse 2022](#)

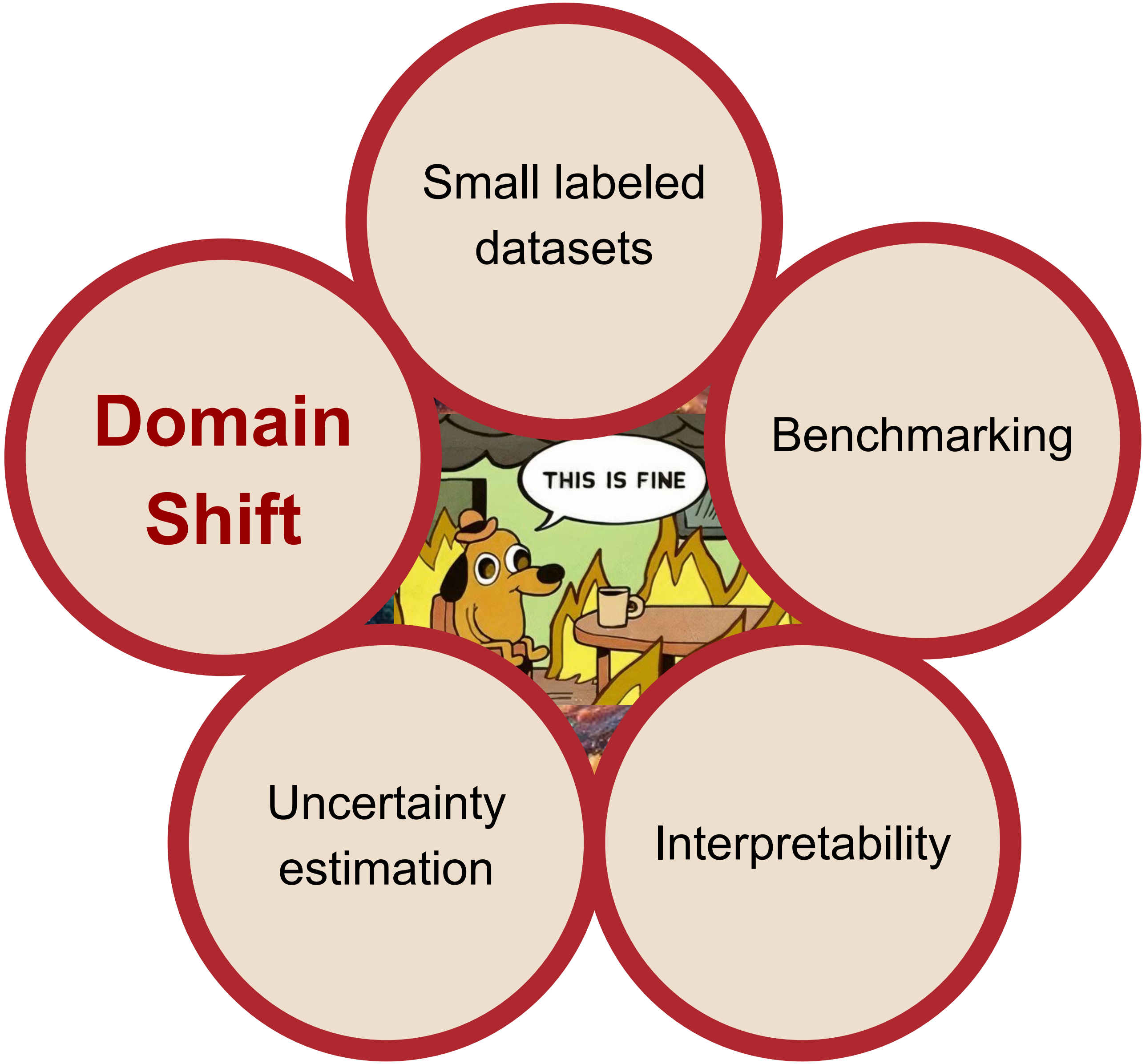
I want to train my AI model by I don’t have enough trustworthy labels?

Can I train on simulations and apply the model to my real data?

How well is my model going to perform?

Can I understand its performance?

Can I compare my model with other researchers?



Talk Outline

Part 1: Domain Shift Problem

Part 2: Domain Adaptation Techniques

Part 3: Examples

Part 4: Distance-based Methods

Part 5: Uncertainty Quantification *

Part 6: Open Questions & Research Trends

01

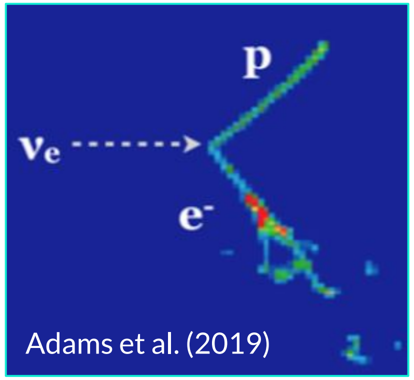
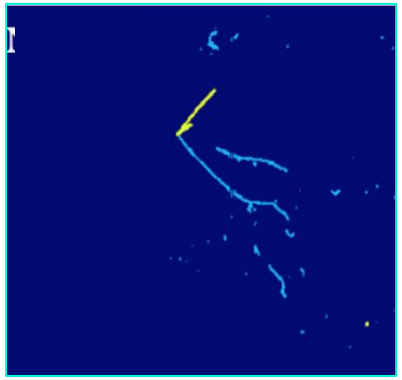
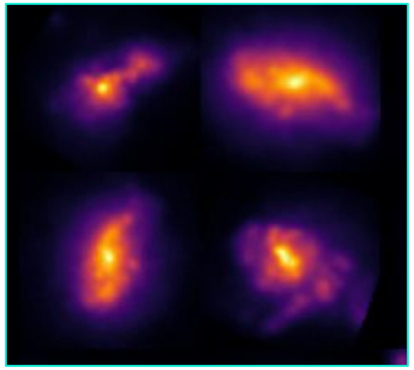
Domain Shift Problem

More Data - More problems

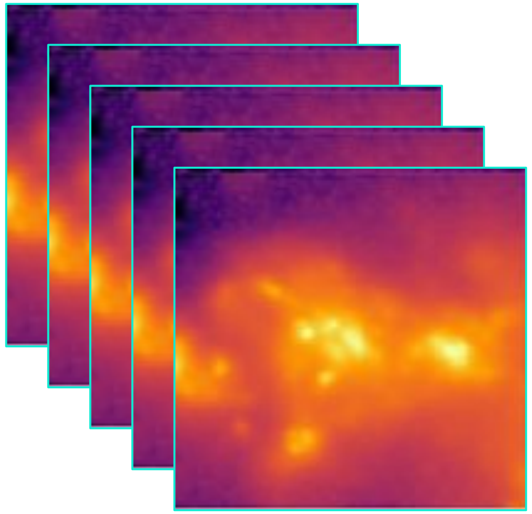
All areas of astro(physics) often need to create **model trained on simulated data, that also work on real data**!

DOMAIN SHIFT

- Missing and unknown physics, wrong geometry, background levels
- Computational constraints for simulations
- Detector problems, transients, errors, data compression
- Imperfect addition of observational effects
- Different detectors or telescopes

	SIMULATED	REAL
MicroBooNE (neutrinos)		
Illustris / Hubble (merging galaxies)		

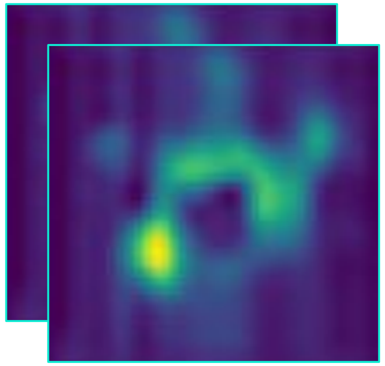
Traditional ML

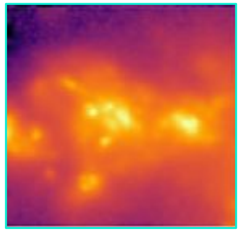


Regular Training

FIND AND REFINE FEATURES

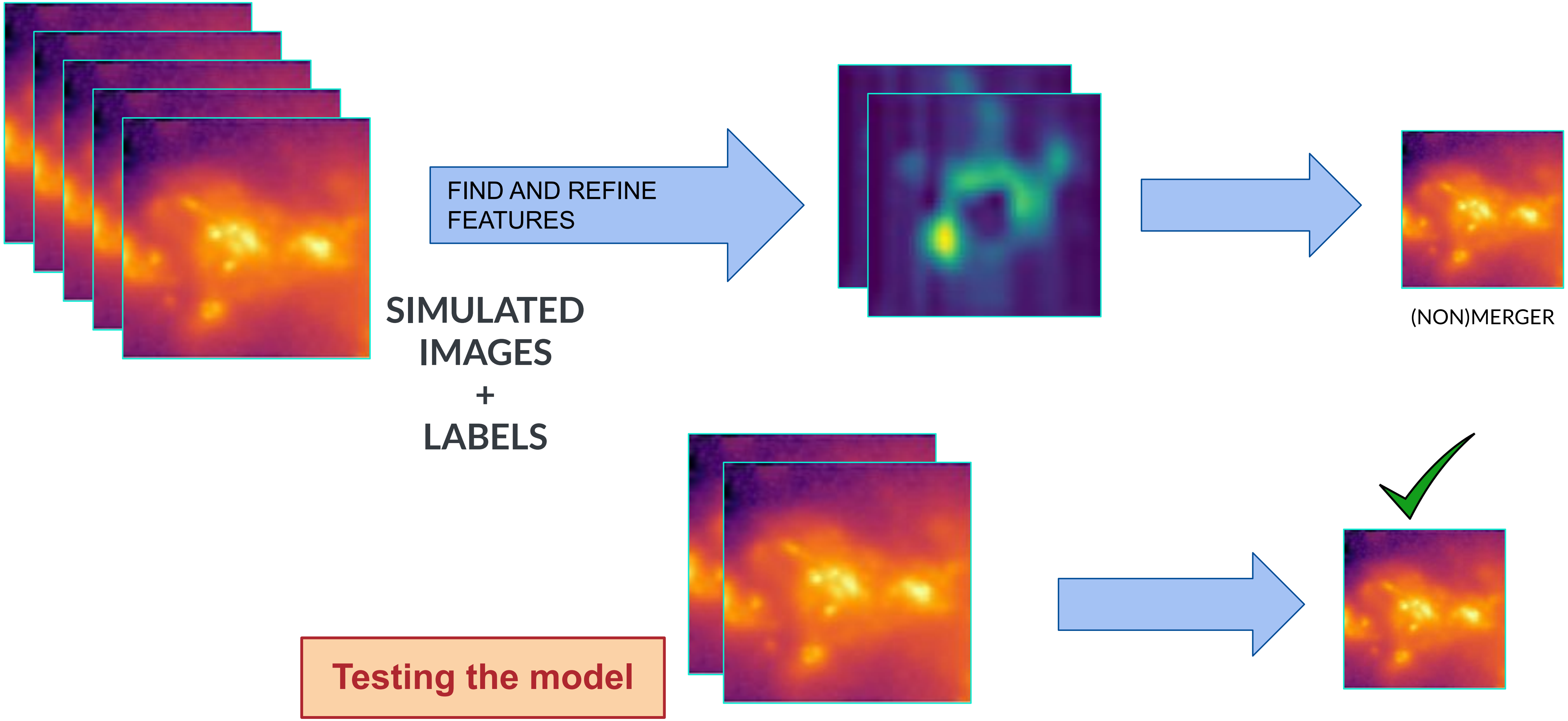
SIMULATED IMAGES + LABELS



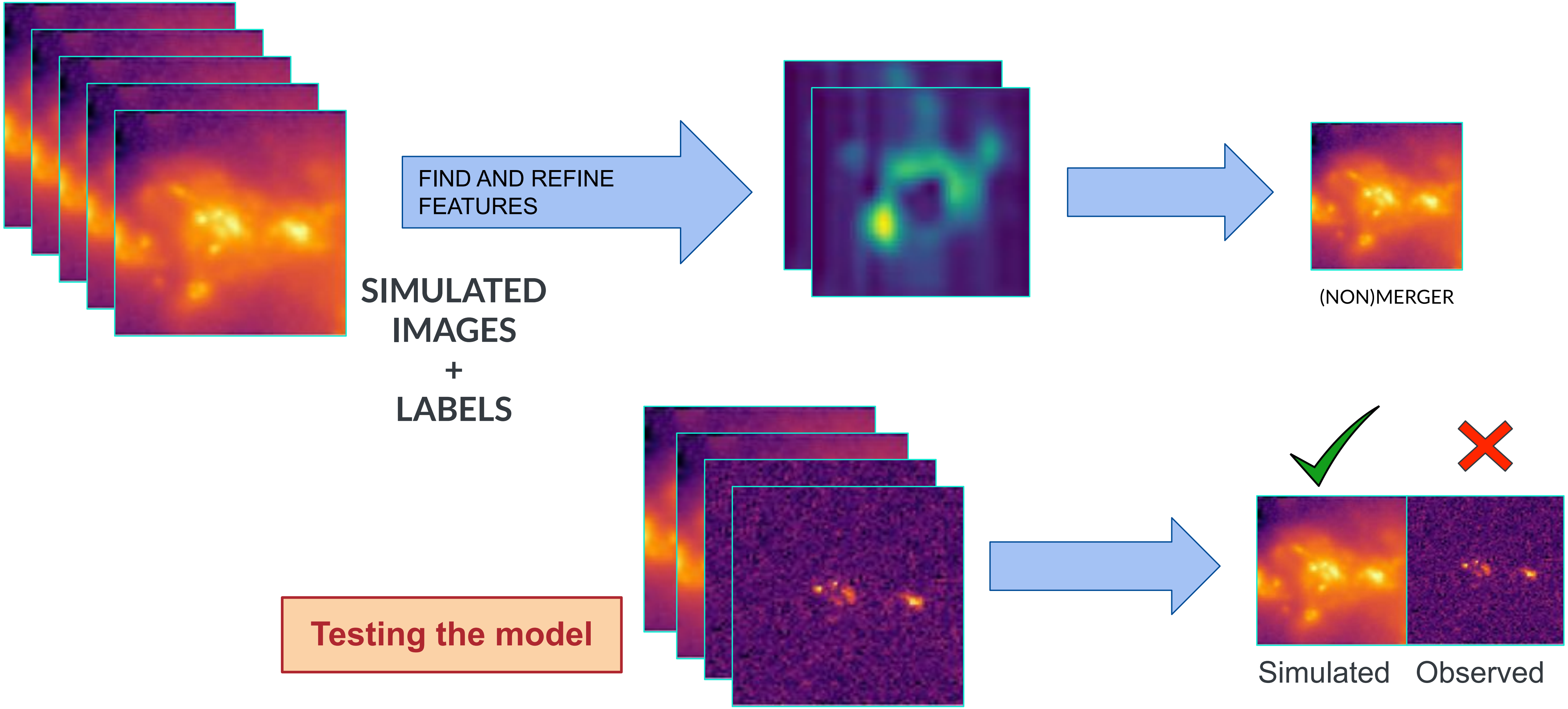


(NON)MERGER

Traditional ML



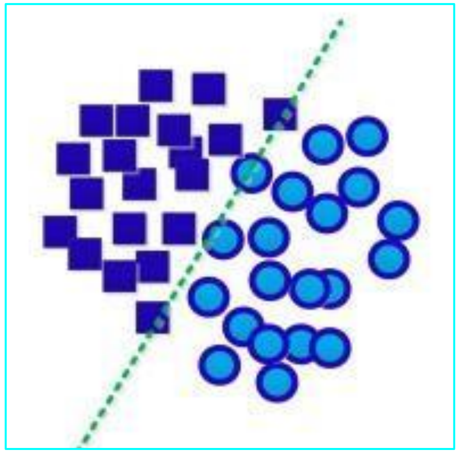
Traditional ML



Why do models fail?

Source Domain

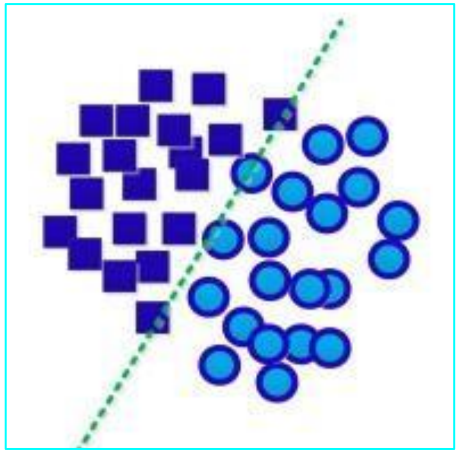
Train the model on source dataset and find the decision boundary.



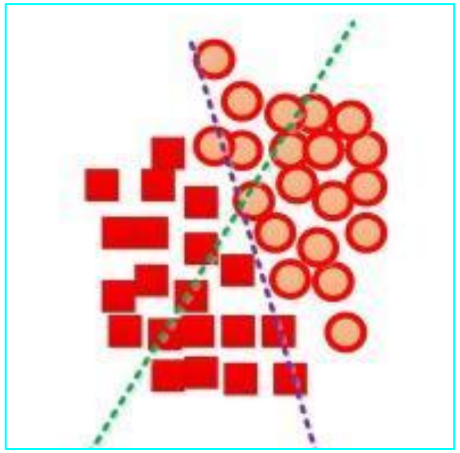
Why do models fail?

New domain is shifted,
learned decision boundary doesn't work.

Source Domain



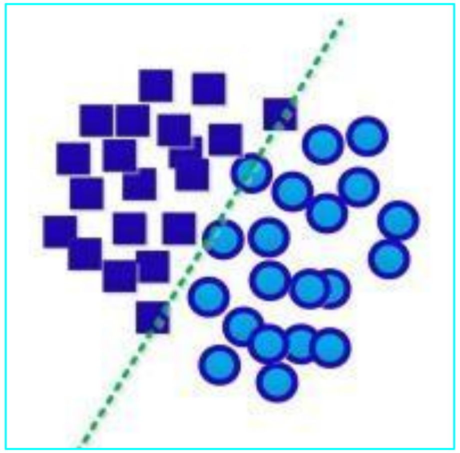
Target Domain



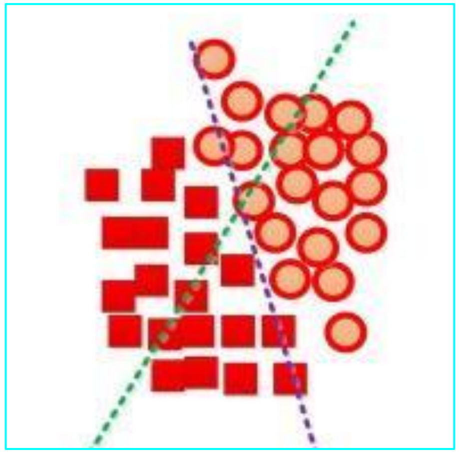
Why do models fail?

We need to align the data during training!

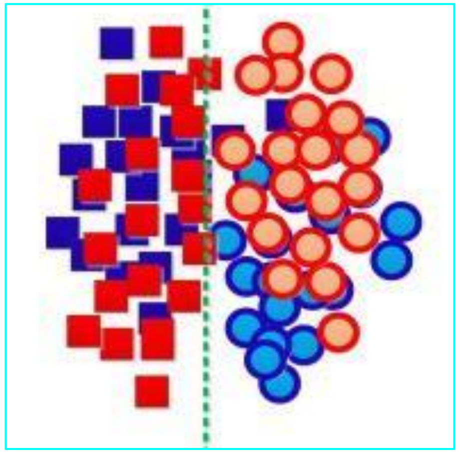
Source Domain



Target Domain

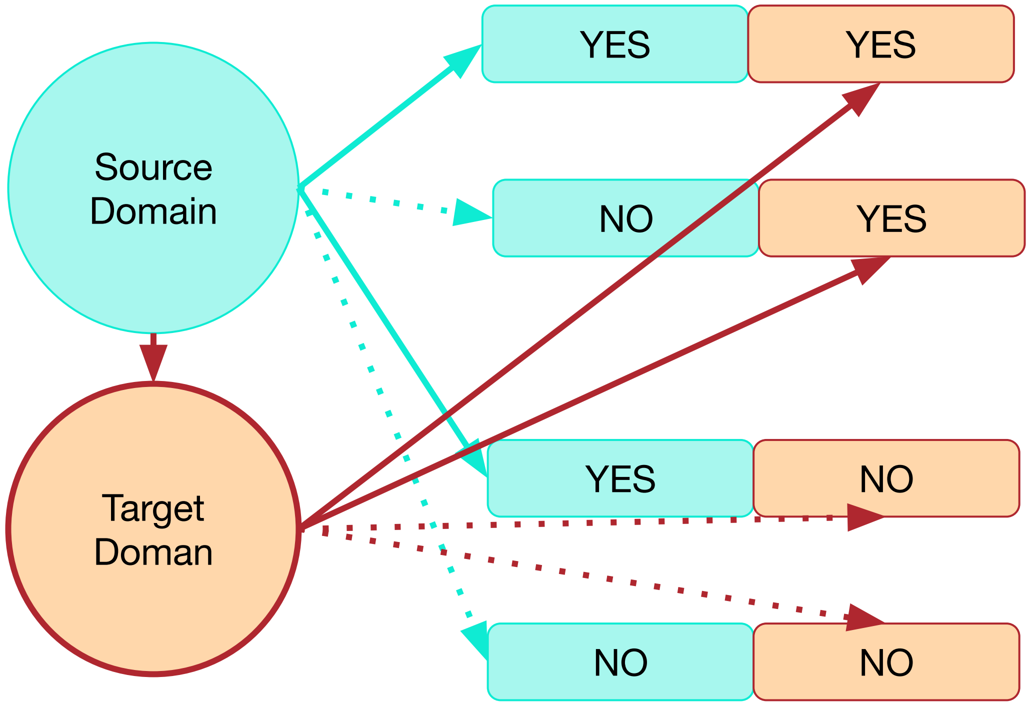


Domain Alignment



The Landscape of Multi-Dataset Learning

Do we have labels?



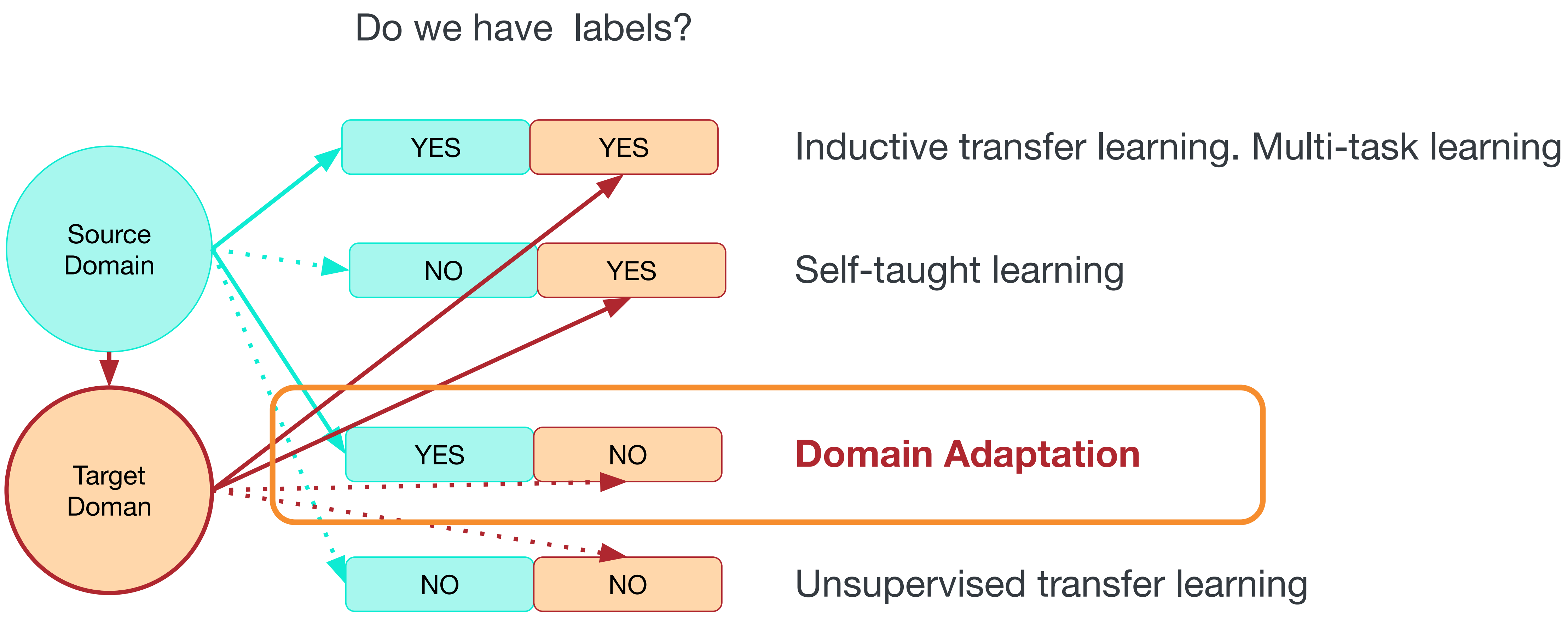
Inductive transfer learning. Multi-task learning

Self-taught learning

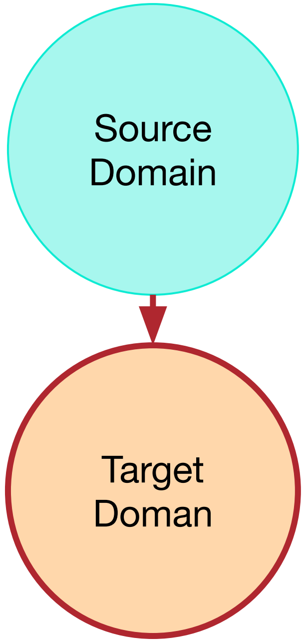
Domain Adaptation

Unsupervised transfer learning

The Landscape of Multi-Dataset Learning



Datasets and Types of Shifts



$$\mathcal{D}_S = \{(x_i^S, y_i^S)\}_{i=1}^{n_S}$$

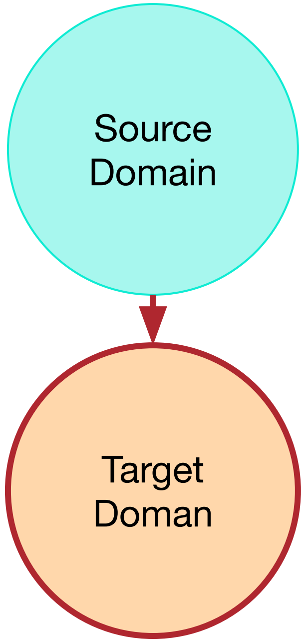
with n_S labelled samples, from joint distribution $P_S(x, y)$

$$\mathcal{D}_T = \{x_i^T\}_{i=1}^{n_T}$$

with n_T unlabelled samples, from joint distribution $P_T(x)$

$$P_S(x, y) \neq P_T(x, y)$$

Datasets and Types of Shifts



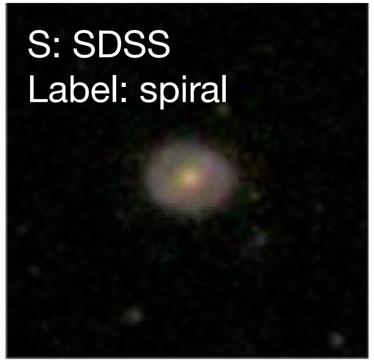
$$\mathcal{D}_S = \{(x_i^S, y_i^S)\}_{i=1}^{n_S}$$

with n_S labelled samples, from joint distribution $P_S(x, y)$

$$\mathcal{D}_T = \{x_i^T\}_{i=1}^{n_T}$$

with n_T unlabelled samples, from joint distribution $P_T(x)$

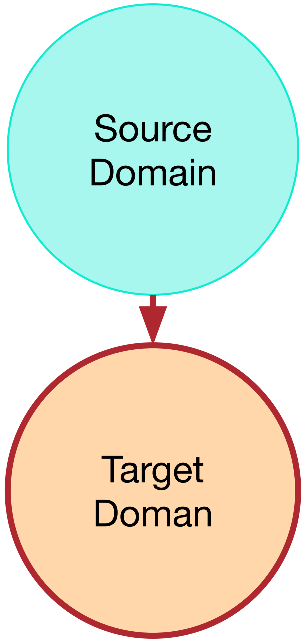
$$P_S(x, y) \neq P_T(x, y)$$



Label Shift: $P_s(Y) \neq P_t(Y)$ while $P_s(X \mid Y) = P_t(X \mid Y)$

Conditional distribution of data given a label stays the same, but the label distribution changes i.e., the frequency of classes differs across domains (e.g., class imbalance), but **each class "looks" the same in both domains**.

Datasets and Types of Shifts



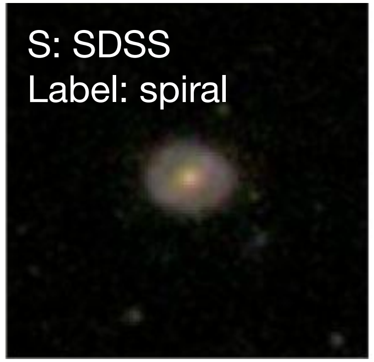
$$\mathcal{D}_S = \{(x_i^S, y_i^S)\}_{i=1}^{n_S}$$

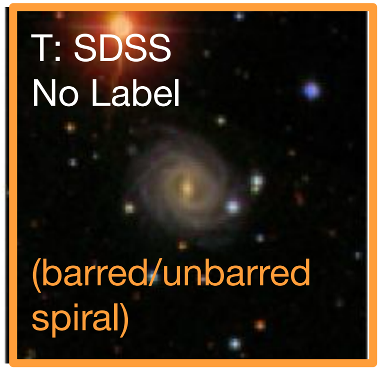
with n_S labelled samples, from joint distribution $P_S(x, y)$

$$\mathcal{D}_T = \{x_i^T\}_{i=1}^{n_T}$$

with n_T unlabelled samples, from joint distribution $P_T(x)$

$$P_S(x, y) \neq P_T(x, y)$$

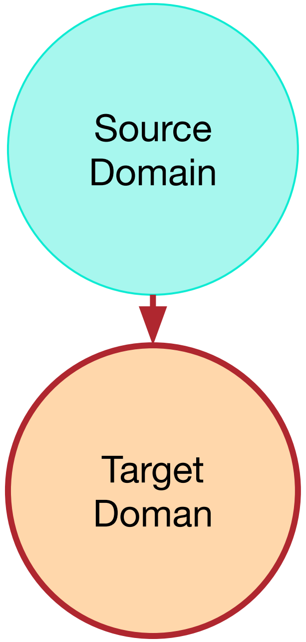




Concept Shift: $P_s(X) = P_t(X)$ but $P_s(Y \mid X) \neq P_t(Y \mid X)$

The input distribution of data stays the same, but the mapping from input to label changes i.e., the same input might be labeled differently across domains (e.g., due to different labeling policies or evolving concepts.

Datasets and Types of Shifts



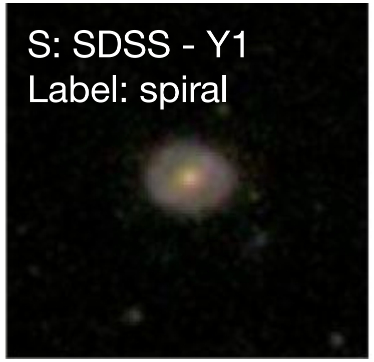
$$\mathcal{D}_S = \{(x_i^S, y_i^S)\}_{i=1}^{n_S}$$

with n_S labelled samples, from joint distribution $P_S(x, y)$

$$\mathcal{D}_T = \{x_i^T\}_{i=1}^{n_T}$$

with n_T unlabelled samples, from joint distribution $P_T(x)$

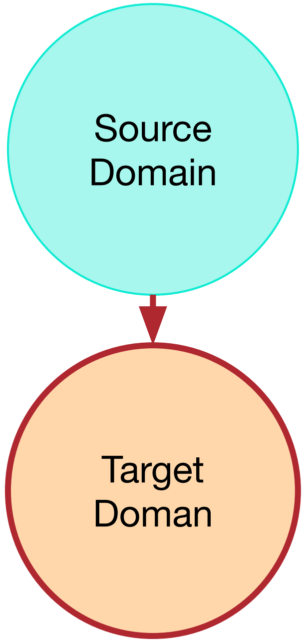
$$P_S(x, y) \neq P_T(x, y)$$



Conditional Shift: $P_S(Y) = P_T(Y)$ but $P_S(X \mid Y) \neq P_T(X \mid Y)$

The label distribution the same, but the mapping from input to label changes i.e., the same labeled corresponds to different data due to measurement differences.

Datasets and Types of Shifts



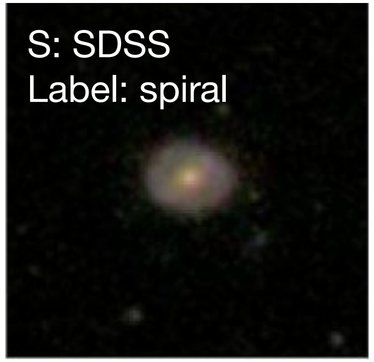
$$\mathcal{D}_S = \{(x_i^S, y_i^S)\}_{i=1}^{n_S}$$

with n_S labelled samples, from joint distribution $P_S(x, y)$

$$\mathcal{D}_T = \{x_i^T\}_{i=1}^{n_T}$$

with n_T unlabelled samples, from joint distribution $P_T(x)$

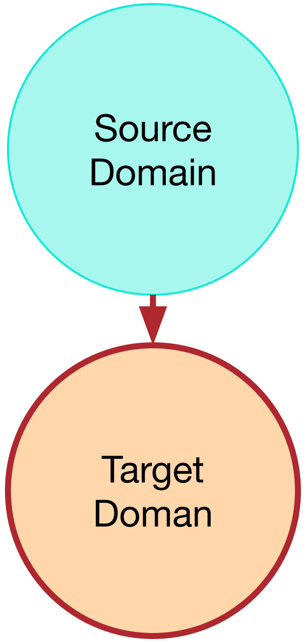
$$P_S(x, y) \neq P_T(x, y)$$



Covariate Shift: $P_s(X) \neq P_t(X)$ but $P_s(Y \mid X) = P_t(Y \mid X)$

Data distribution changes (often not just image properties but class distribution as well for example e.g., different sensors, sampling bias), but the conditional distribution of labels given the data remains the same.

Datasets and Types of Shifts

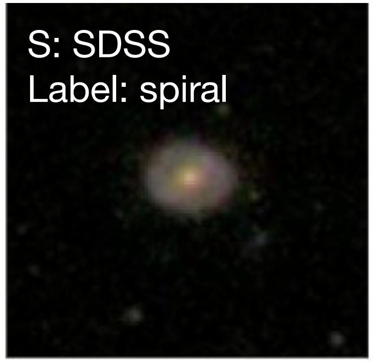


$$\mathcal{D}_S = \{(x_i^S, y_i^S)\}_{i=1}^{n_S}$$

with n_S labelled samples, from joint distribution $P_S(x, y)$

$$\mathcal{D}_T = \{x_i^T\}_{i=1}^{n_T}$$

with n_T unlabelled samples, from joint distribution $P_T(x)$



$$P_S(x, y) \neq P_T(x, y)$$

Compound Shift:

Any potential combination of two or more shift!

Dataset Shift Quiz !

Scenario 1:

You trained a model to detect a rare disease in a hospital with a low prevalence rate (5%). It's now deployed in a different region where the prevalence is 30%. Patient features (symptoms, test values) behave similarly for each diagnosis.

Covariate	Concept / Conditional	Label

Answer:

Dataset Shift Quiz !

Scenario 1:

You trained a model to detect a rare disease in a hospital with a low prevalence rate (5%). It's now deployed in a different region where the prevalence is 30%. Patient features (symptoms, test values) behave similarly for each diagnosis.

Covariate	Concept / Conditional	Label

Answer:

Label shift — $P(y)$ has changed (disease rate), but $P(x \mid y)$ is stable.

Dataset Shift Quiz !

Scenario 2:

Your spam classifier was trained on English-language emails. Now it's deployed to classify emails in both English and French. The spam-to-nonspam ratio is about the same, and spam emails still tend to have similar structures (e.g., lots of links, certain keywords).

Covariate	Concept / Conditional	Label

Answer:

Dataset Shift Quiz !

Scenario 2:

Your spam classifier was trained on English-language emails. Now it's deployed to classify emails in both English and French. The spam-to-nonspam ratio is about the same, and spam emails still tend to have similar structures (e.g., lots of links, certain keywords).

Covariate	Concept / Conditional	Label

Answer:

Covariate shift — the input distribution P(x) changed (language), but the conditional P(y | x) stayed relatively consistent.

Dataset Shift Quiz !

Scenario 3:
Your credit risk model was trained 5 years ago. Now, due to economic changes and inflation, people with low income are more likely to default than they used to. Income distribution remains unchanged.

Covariate	Concept / Conditional	Label

Answer:

Dataset Shift Quiz !

Scenario 3:

Your credit risk model was trained 5 years ago. Now, due to economic changes and inflation, people with low income are more likely to default than they used to. Income distribution remains unchanged.

Covariate	**Concept** / Conditional	Label

Answer:

Concept shift — the relationship P(y | x) changed due to evolving economic factors even though P(x) remained the same.

FERMI NATIONAL ACCELERATOR LABORATORY

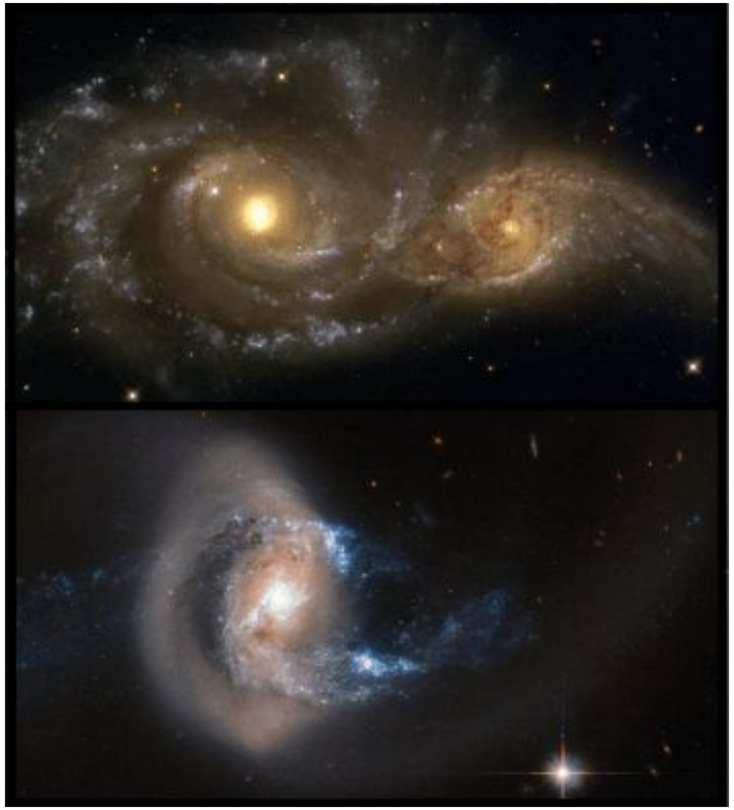
Dataset Shift Quiz !

Scenario 4:

You trained a model on a balanced dataset of merging and non-merging galaxies. At test time, your realistic catalog contains 90% of the non-merging galaxy images, and the remaining 10% are merging. The images for each galaxy still look the same.

Covariate	Concept / Conditional	Label

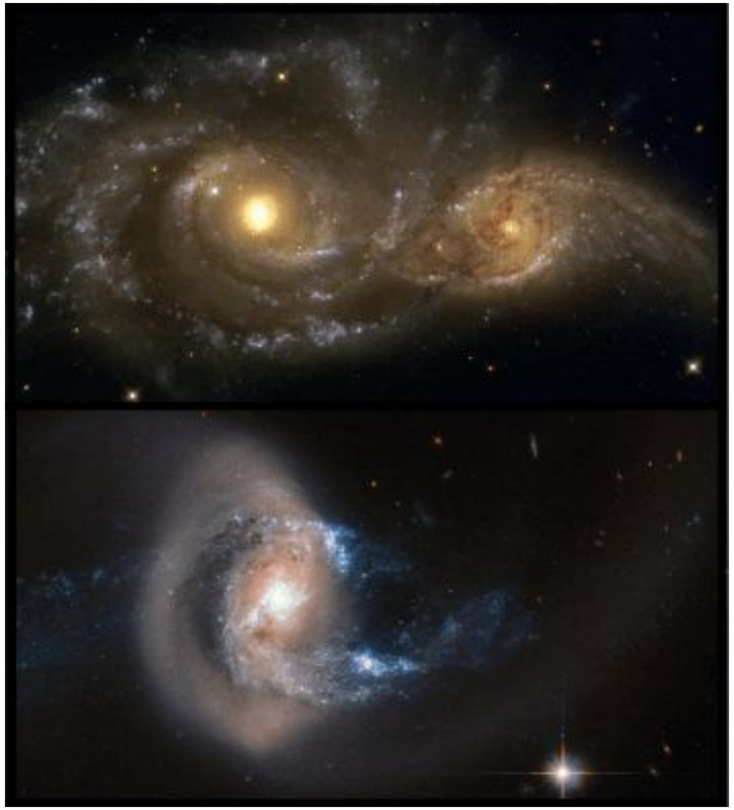
Answer:



Dataset Shift Quiz !

Scenario 4:

You trained a model on a balanced dataset of merging and non-merging galaxies. At test time, your realistic catalog contains 90% of the non-merging galaxy images, and the remaining 10% are merging. The images for each galaxy still look the same.



Covariate	Concept / Conditional	Label

Answer:

Label shift — the label distribution P(y) changed (mergers are rare), but image features (P(x | y)) stayed the same.

Dataset Shift Quiz !

Scenario 5:

A remote sensing model is trained to classify crop types (e.g., wheat, corn, rice) using satellite imagery from **Region A** during a specific season. Later, the model is applied to **Region B**, where crop types and labels are the same, but **soil composition, weather patterns, and sunlight levels differ**, affecting how the crops look in satellite images

Covariate	Concept / Conditional	Label

Answer:

Dataset Shift Quiz !

Scenario 5:

A remote sensing model is trained to classify crop types (e.g., wheat, corn, rice) using satellite imagery from **Region A** during a specific season. Later, the model is applied to **Region B**, where crop types and labels are the same, but **soil composition, weather patterns, and sunlight levels differ**, affecting how the crops look in satellite images .

Covariate	Concept / Conditional	Label

Answer:

Covariate shift — the input distribution **P(x)** (satellite pixel features) has changed due to environmental factors, but **P(y | x)** (how features map to crop types) is assumed to remain the same.

Useful Resources:

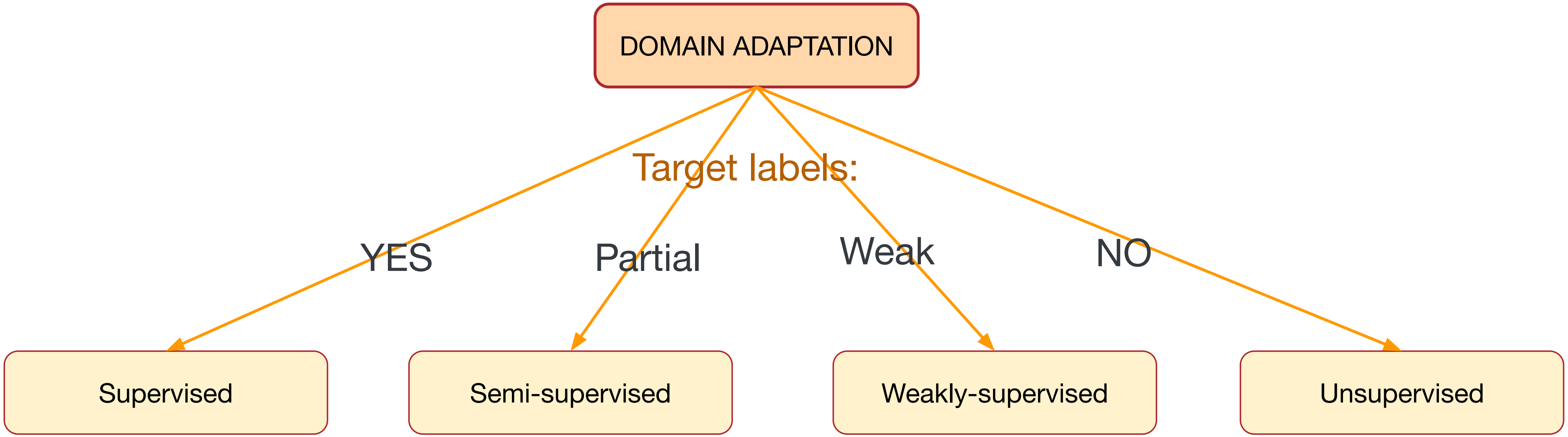
- “A Brief Review of Domain Adaptation” Farahini et al. 2020.
- “Deep Visual Domain Adaptation: A Survey” Wand & Deng 2018.
- “Domain Adaptation for Visual Applications: A Comprehensive Survey” Csurka 2017.

02

Domain Adaptation Techniques

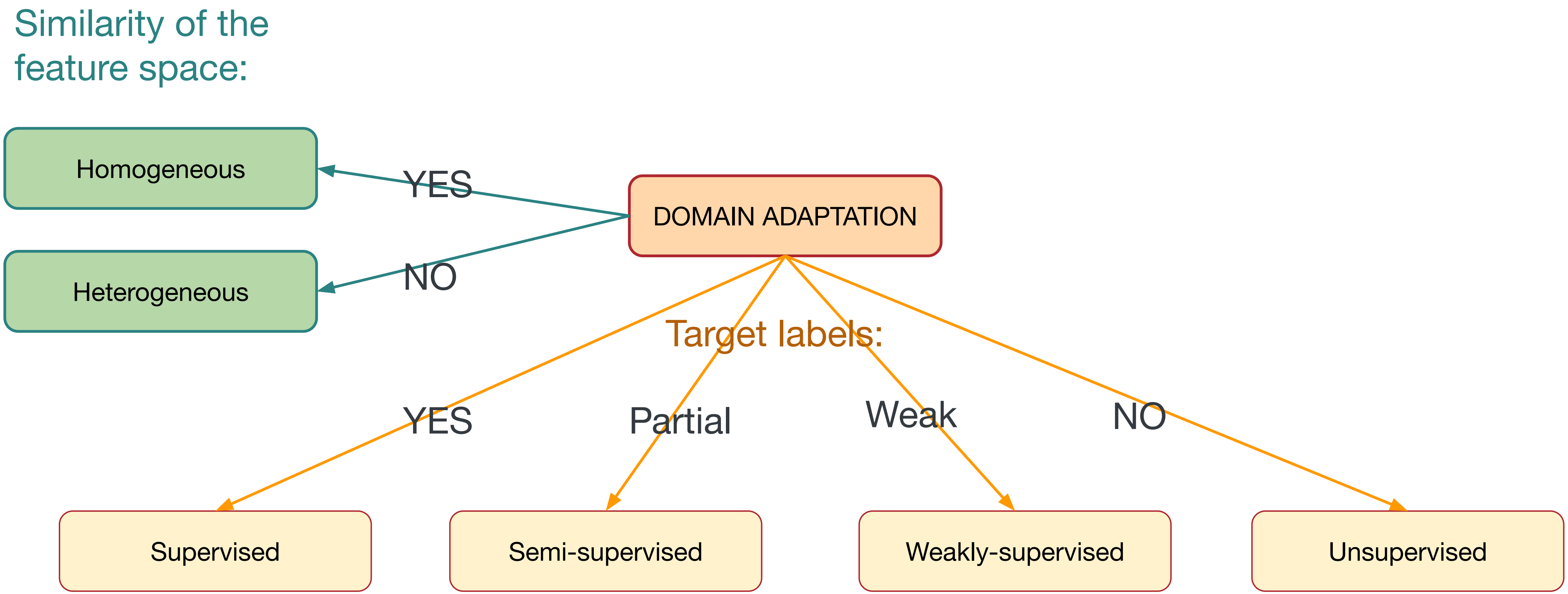
Types of Domain Adaptation Approaches

There is a huge number of domain adaptation methods to suit different needs.



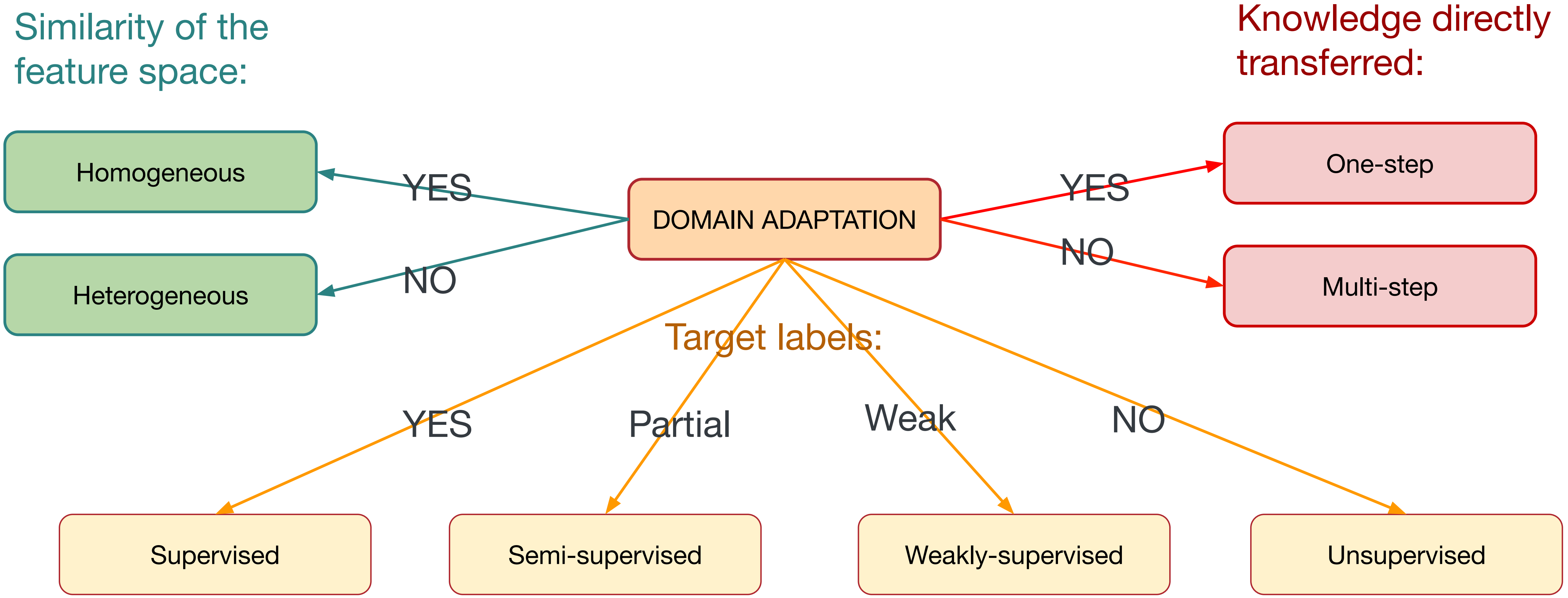
Types of Domain Adaptation Approaches

There is a huge number of domain adaptation methods to suit different needs.



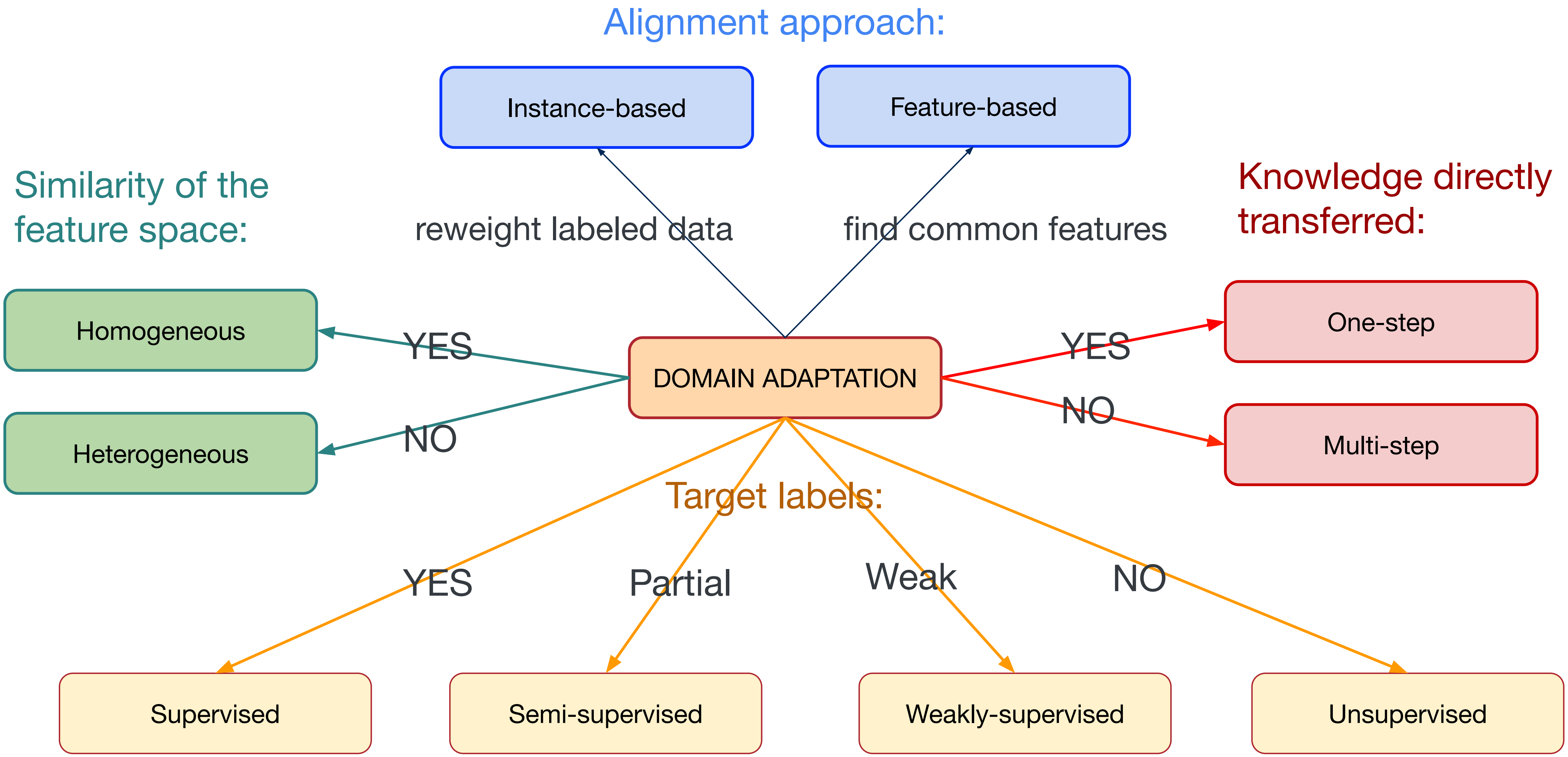
Types of Domain Adaptation Approaches

There is a huge number of domain adaptation methods to suit different needs.



Types of Domain Adaptation Approaches

There is a huge number of domain adaptation methods to suit different needs.



Domain Adaptation Quiz !

Scenario 1:

You trained a cosmic ray classifier using labeled simulated data. You also have a small set of real cosmic ray events with noisy labels from an old detector. What is this setup best described as?

Supervised	Weakly supervised	Unsupervised

Answer:

Domain Adaptation Quiz !

Scenario 1:

You trained a cosmic ray classifier using labeled simulated data. You also have a small set of real cosmic ray events with noisy labels from an old detector. What is this setup best described as?

Supervised	Weakly supervised	Unsupervised

Answer:

You have real data labels, but they are **noisy or imprecise** — so you need to use **weak supervision**. Supervised DA assumes accurate labels, while unsupervised DA assumes no labels. Weak supervision methods handle uncertainty, label correction, or probabilistic label modeling.

Domain Adaptation Quiz !

Scenario 2:

You align galaxy image features across domains by first performing PCA, then minimizing some domain adaptation loss. You then fine-tune a regressor on the aligned target features. What type of method is this?

One-step, instance based	Multi-step, feature based	Multi-step, instance based

Answer:

Domain Adaptation Quiz !

Scenario 2:

You align galaxy image features across domains by first performing PCA, then minimizing some domain adaptation loss. You then fine-tune a regressor on the aligned target features. What type of method is this?

One-step, instance based	Multi-step, feature based	Multi-step, instance based

Answer:

This approach first **transforms features** (PCA), then **aligns distributions** using a DA method, and finally adapts the model — making it **multi-step and feature-based**. Instance-based methods directly reweight or select examples data points.

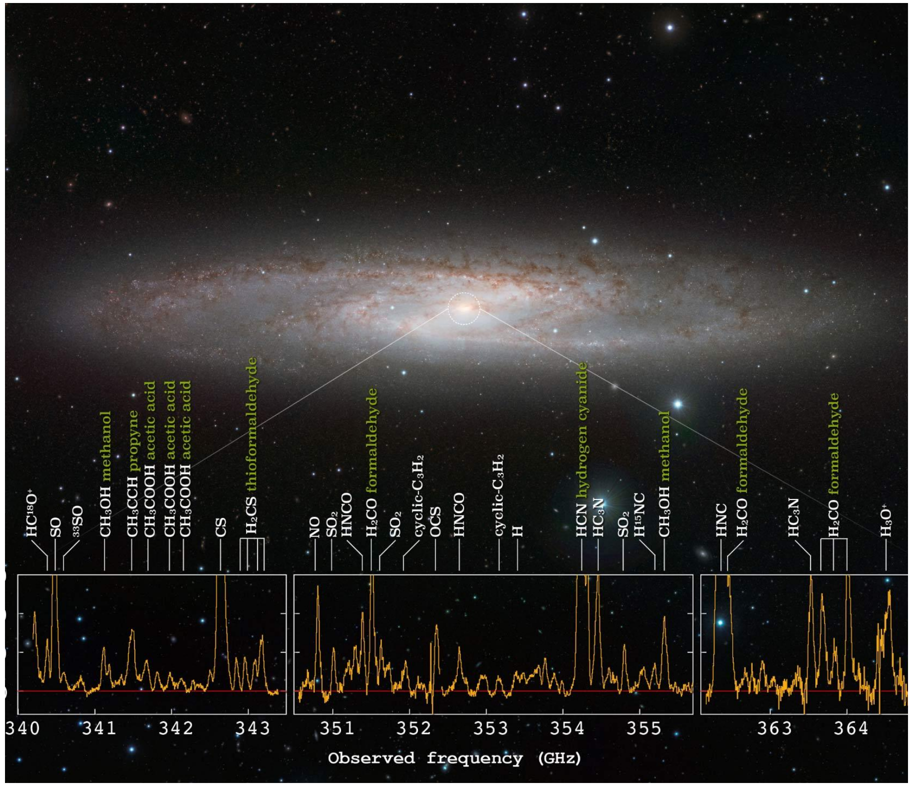
Domain Adaptation Quiz !

Scenario 3:

You trained a redshift estimation model using spectroscopic data (rich, high-dimensional spectra). You now want to apply it to photometric data (broadband magnitudes). What kind of domain adaptation challenge is this?

Unsupervised	Homogeneous	Heterogeneous

Answer:

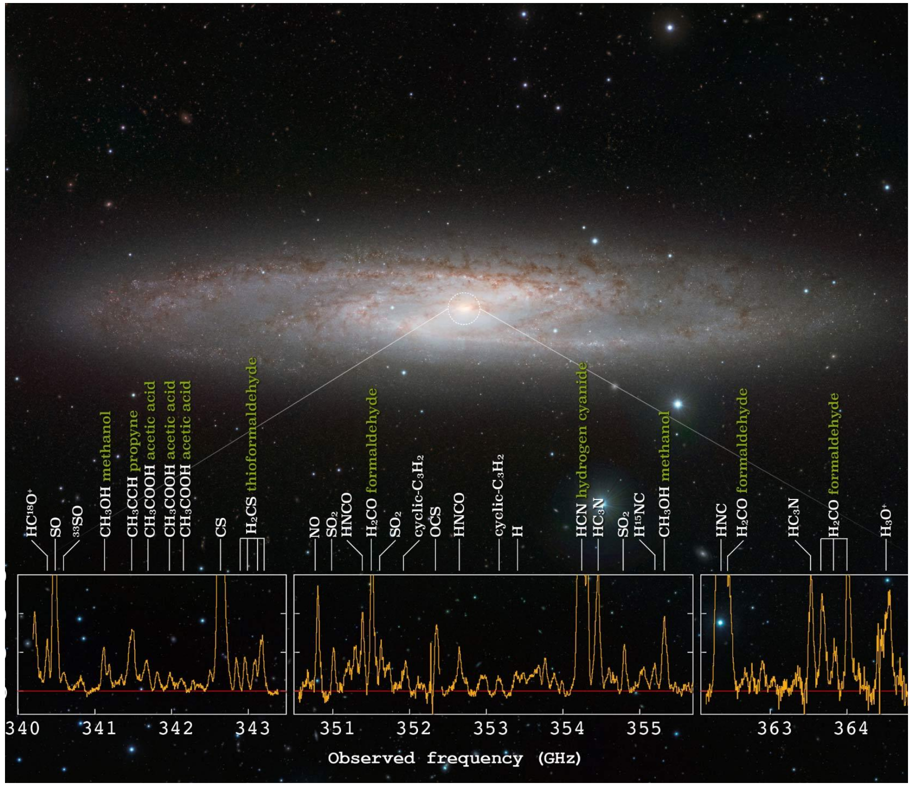


Domain Adaptation Quiz !

Scenario 3:

You trained a redshift estimation model using spectroscopic data (rich, high-dimensional spectra). You now want to apply it to photometric data (broadband magnitudes). What kind of domain adaptation challenge is this?

Unsupervised	Homogeneous	Heterogeneous

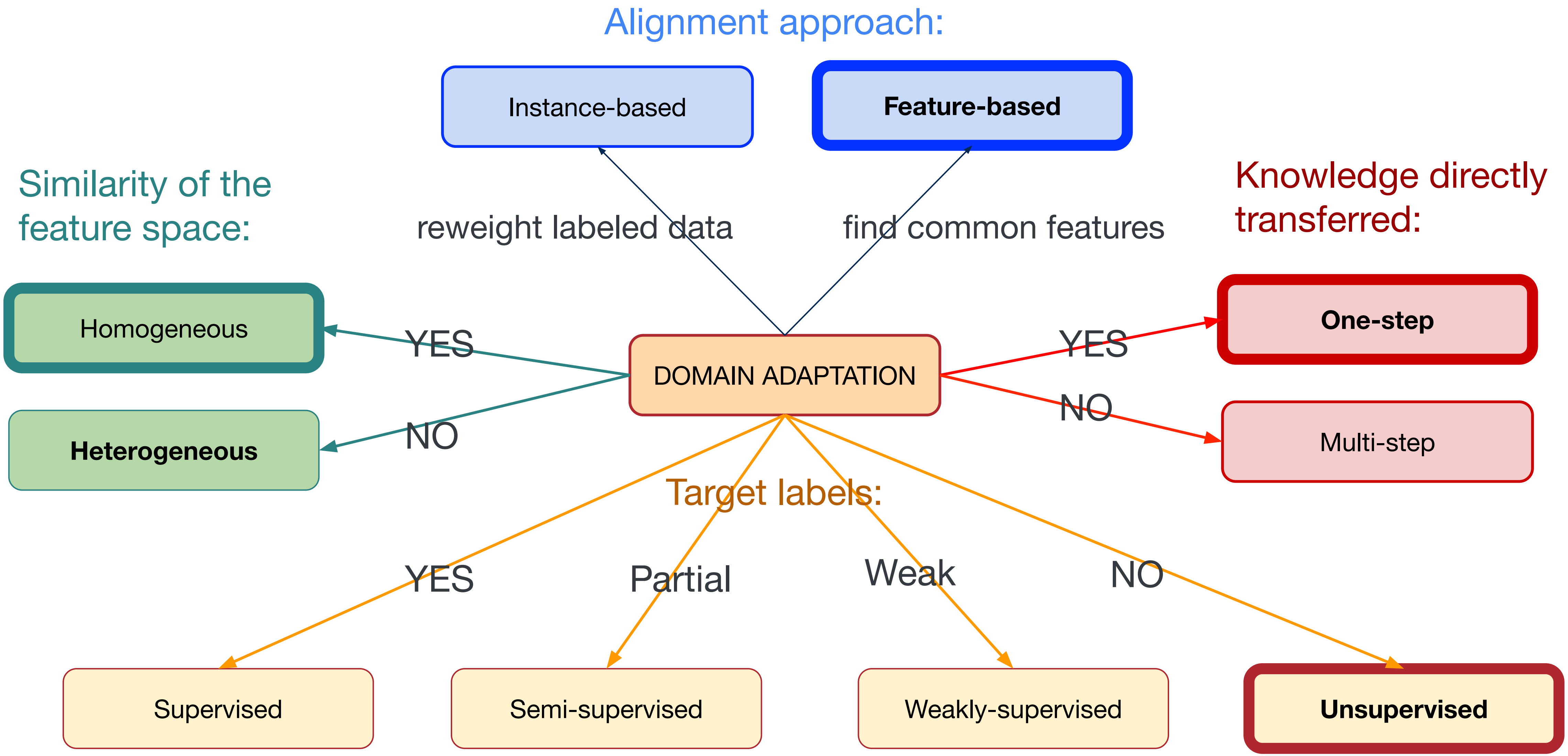


Answer:

The **input features differ in type and dimension** — spectra vs. photometry — so this is a **heterogeneous DA** problem. Homogeneous DA assumes the same input space. The key challenge here lies in mapping between incompatible feature spaces (e.g., using shared latent representations or cross-modal alignment).

Types of Domain Adaptation Approaches

There is a huge number of domain adaptation methods to suit different needs.



Combining Datasets

DOMAIN ADAPTATION

Align data distributions in the latent space of the network by forcing the network to **find more robust domain-invariant features**.

Combining Datasets

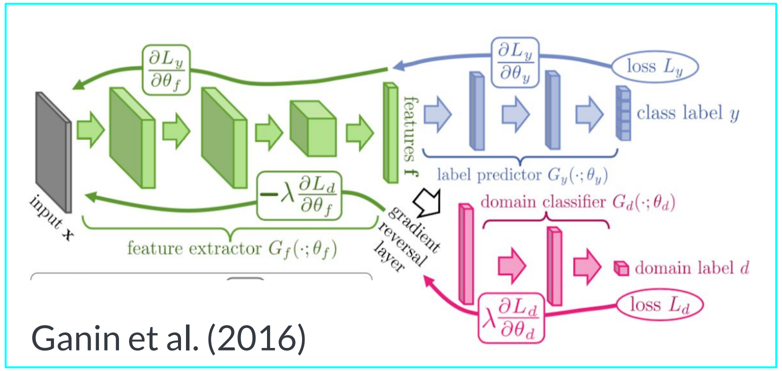
DOMAIN ADAPTATION

Align data distributions in the latent space of the network by forcing the network to **find more robust domain-invariant features.**

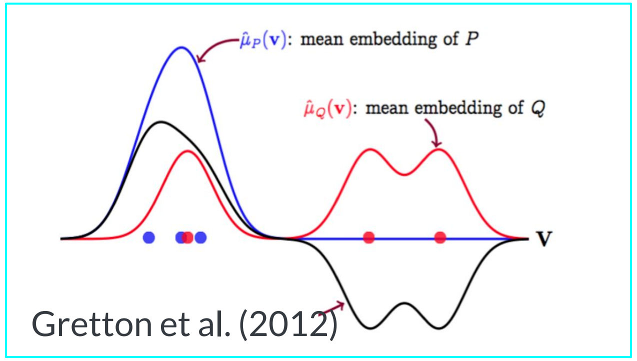
Adversarial

Distance-based methods

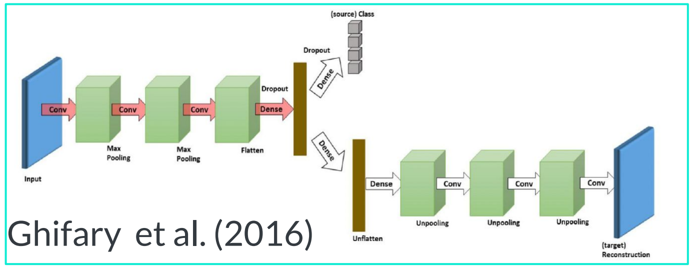
Reconstruction-based



Using domain discriminators to encourage domain confusion through an adversarial objective.



Minimize a distance metric between source and target data.



Data reconstruction as an auxiliary task to ensure feature invariance.

FERMI NATIONAL ACCELERATOR LABORATORY

Combining Datasets

DOMAIN ADAPTATION

Align data distributions in the latent space of the network by forcing the network to **find more robust domain-invariant features**.

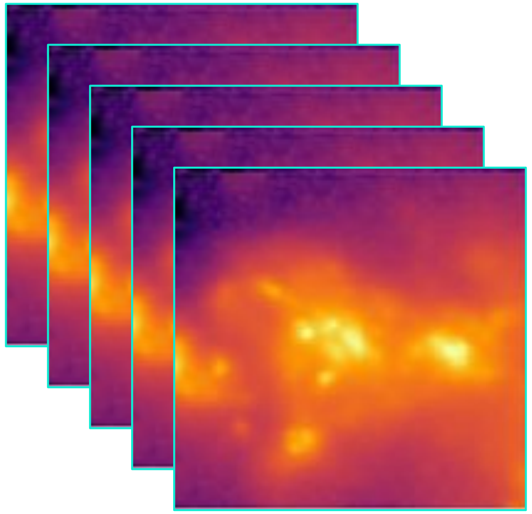
Adversarial

Distance-based methods

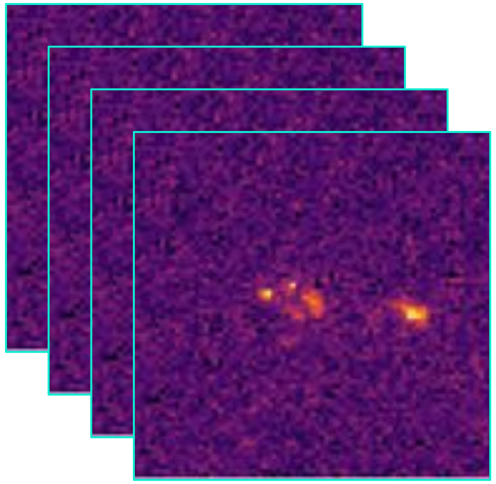
Reconstruction-based

Works on **unlabeled target domain**!
Can be applied to **new data**, no need for scientists to label anything.

Training with Domain Adaptation



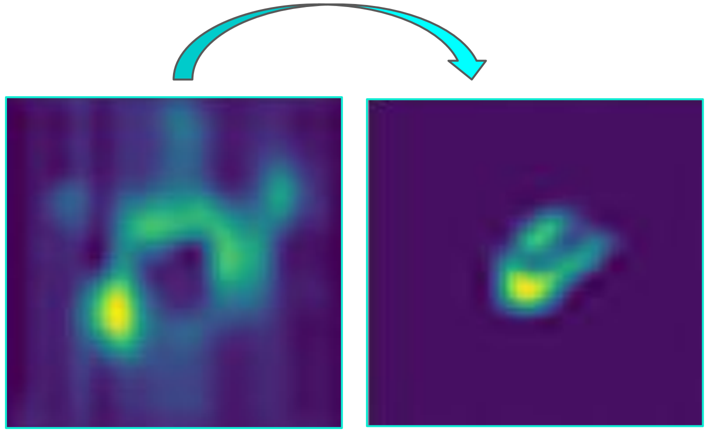
SIMULATED IMAGES + LABELS

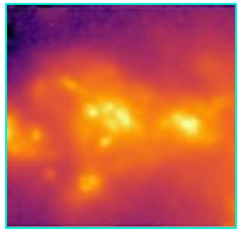


OBSERVED IMAGES

Domain Adaptation

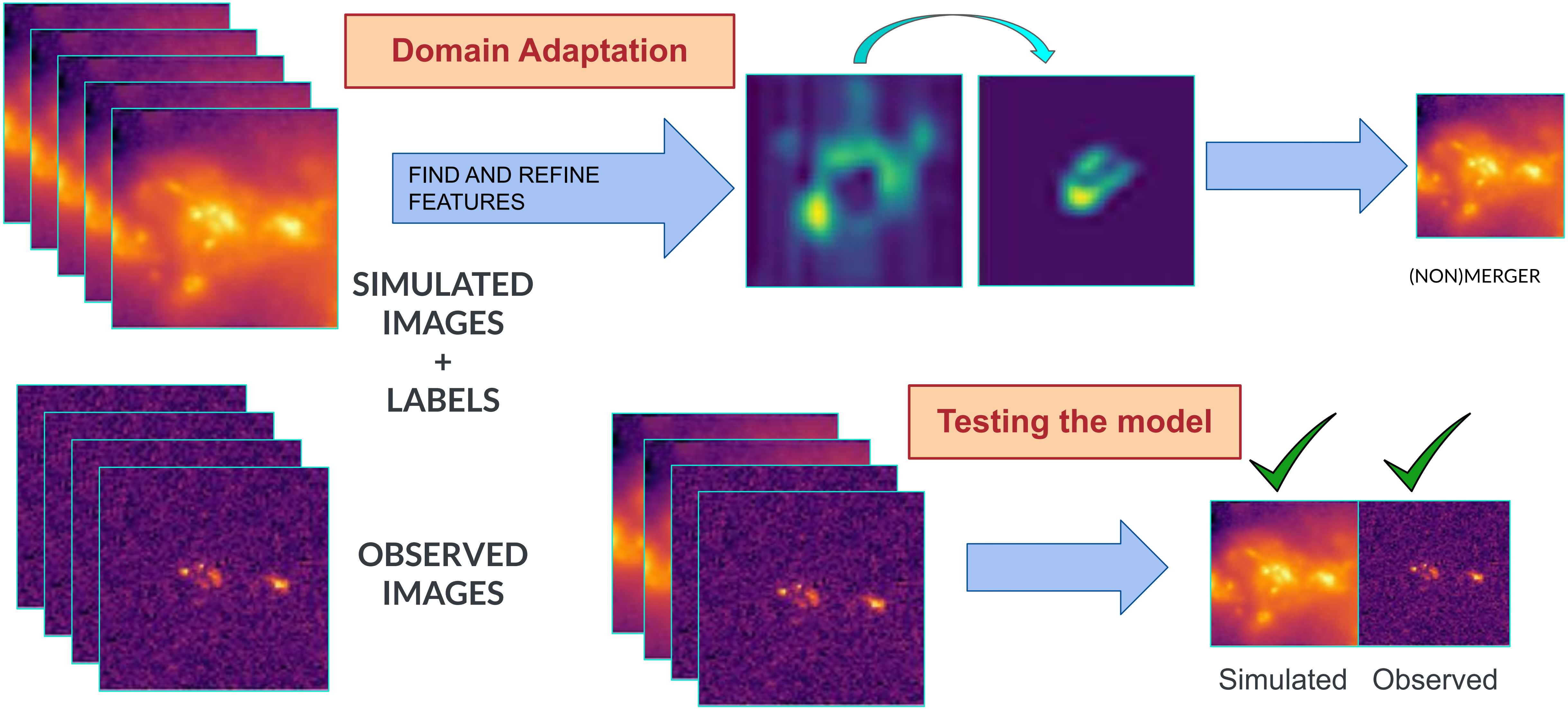
FIND AND REFINE FEATURES





(NON)MERGER

Training with Domain Adaptation

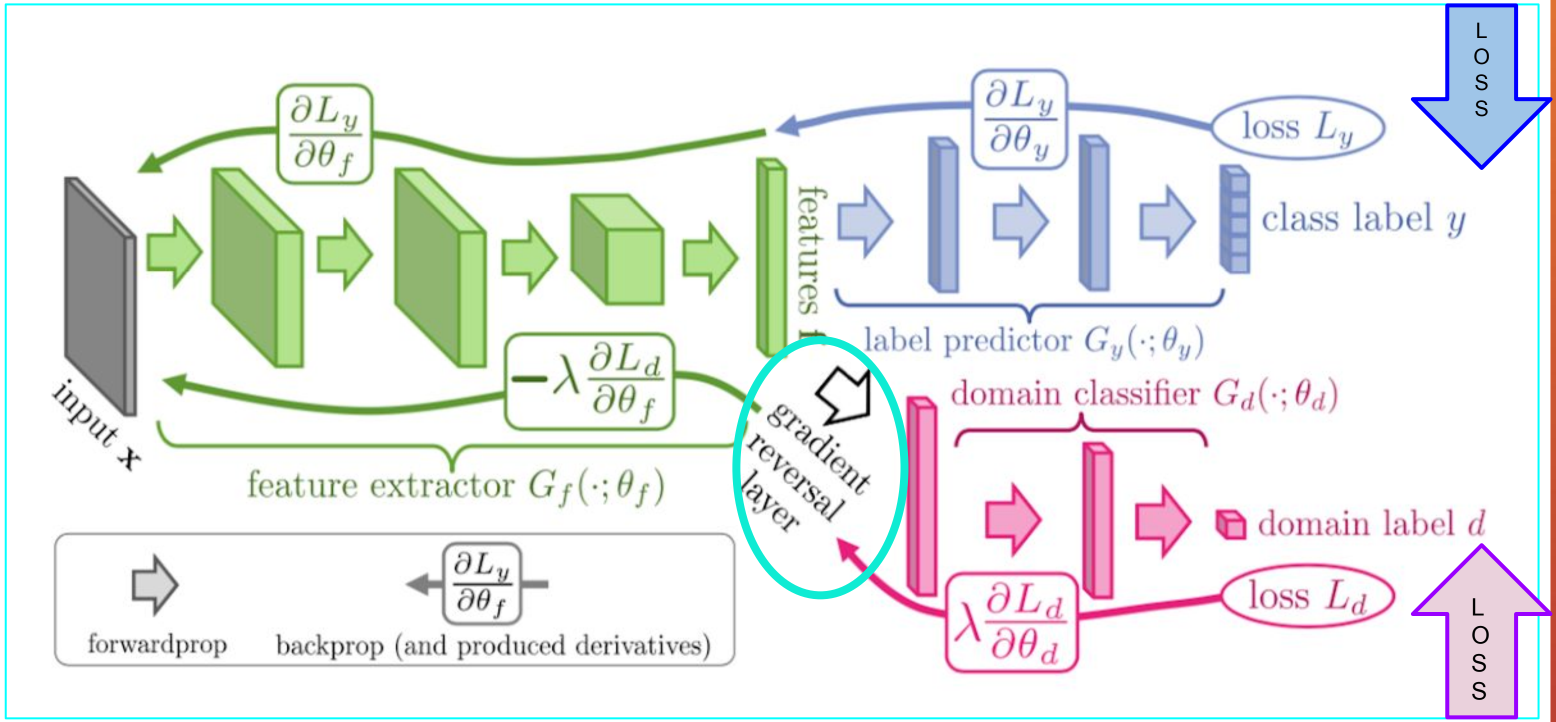


Adversarial Training

Domain Adversarial Neural Networks - DANNs

DANN - feature extractor + label predictor + domain classifier

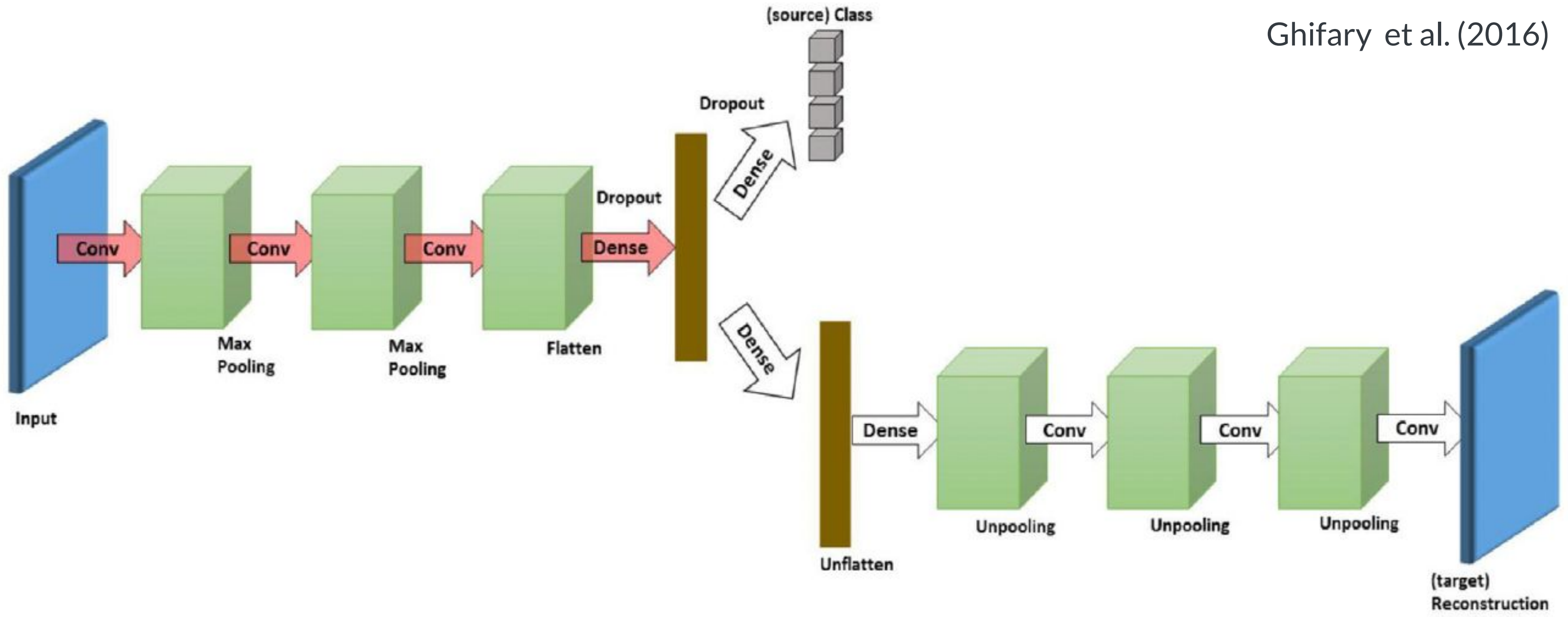
- **Gradient reversal layer** - multiplies the gradient by a negative constant during the backpropagation.
- Results in the extraction of **domain-invariant features**.
- Only source domain images are labeled during training.



Ganin et al. (2016)

Reconstruction-Based Training

Deep Reconstruction-Classification Networks - DRCN



Ghifary et al. (2016)

Model jointly learns a shared encoding representation for two tasks:

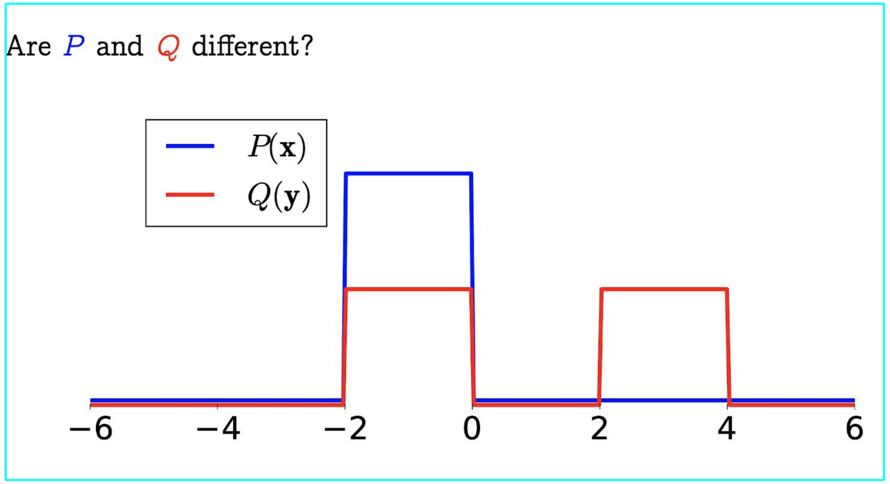
- **supervised classification** of labeled source data
- **unsupervised reconstruction** of unlabeled target data.

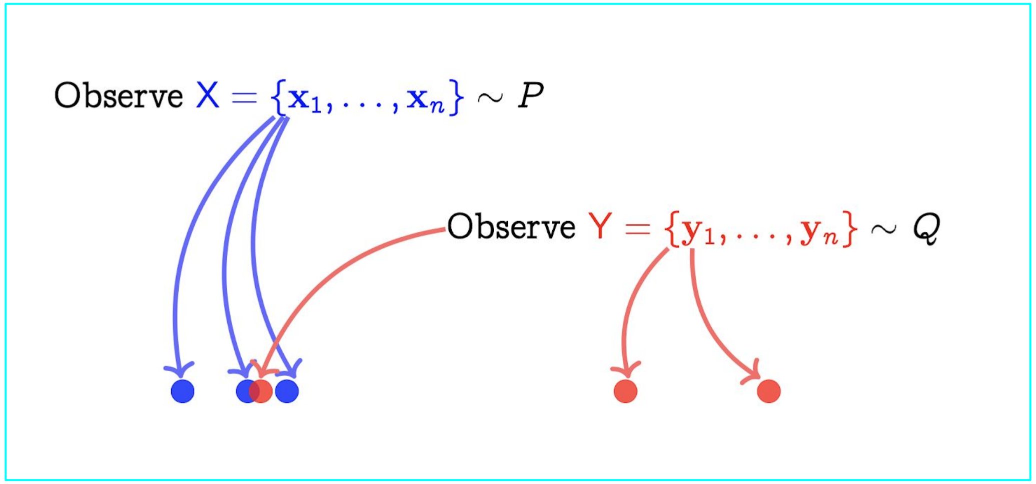
The learnt representation not only preserves discriminability, but also encodes useful information from the target domain.

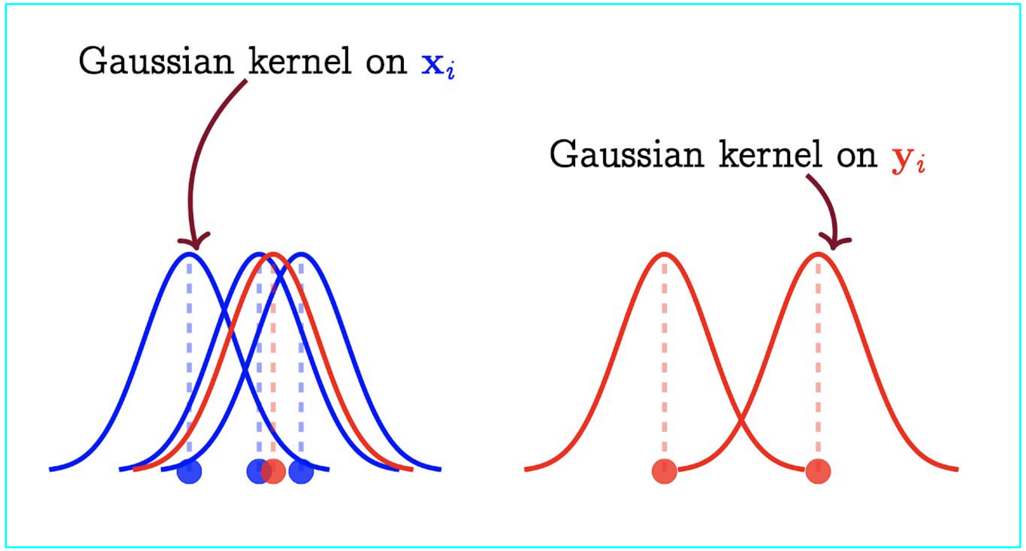
Distance-Based Training

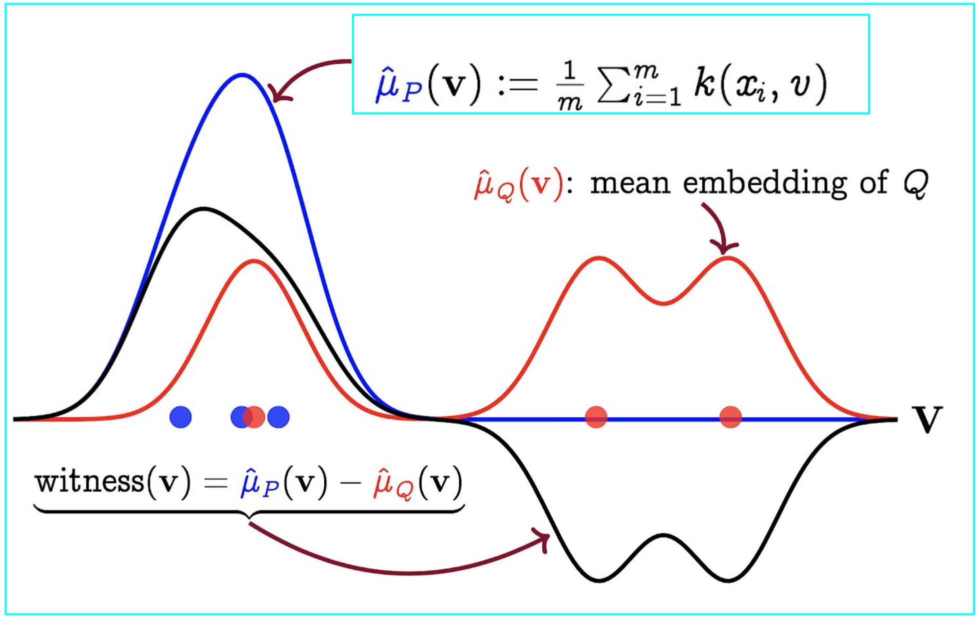
Maximum Mean Discrepancy - MMD

Smola et al. (2007)
Gretton et al. (2012)





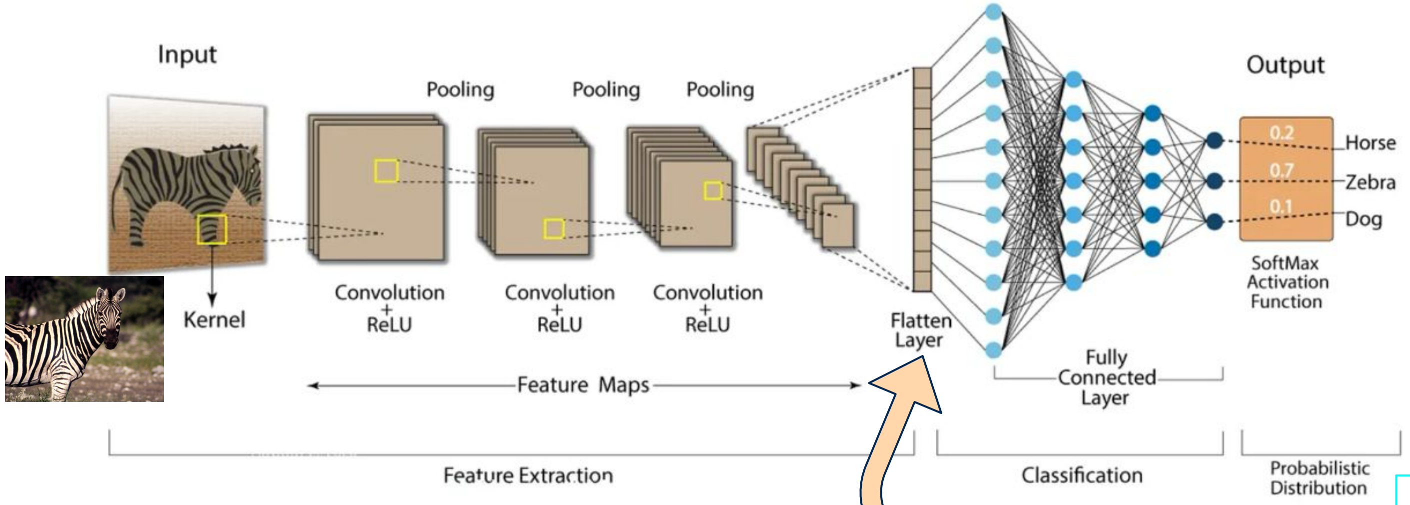


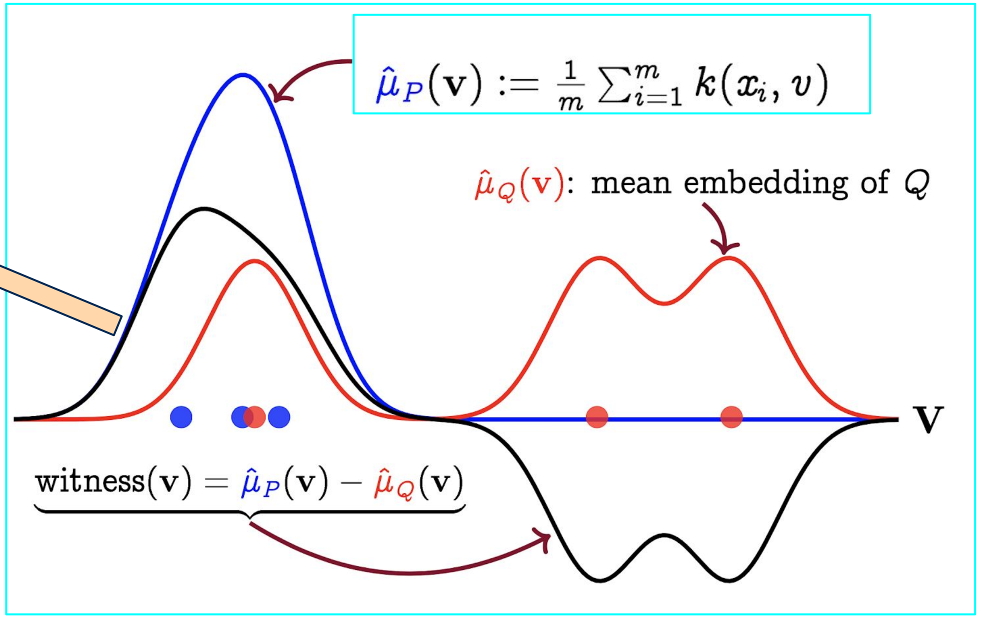


Distance-Based Training

Maximum Mean Discrepancy - MMD

Smola et al. (2007)
Gretton et al. (2012)





FERMI NATIONAL ACCELERATOR LABORATORY

BREAK

How do visualise what DA is doing?

Dimensionality reduction techniques

We need to be able to examine the multi-dimensional laten (embedding) space i.e. go from ND to 2D or 3D.

- tSNE
- Isomap
- Umap

How do visualise what DA is doing?

Dimensionality reduction techniques

We need to be able to examine the multi-dimensional laten (embedding) space i.e. go from ND to 2D or 3D.

tSNE | Isomap | Umap

Figure from: [Distil]

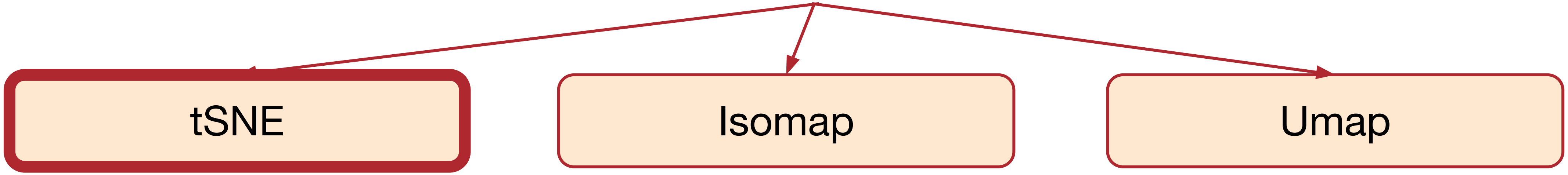
t-distributed Stochastic Neighbor Embedding

A nonlinear dimensionality reduction technique that preserves local similarities by modeling high-dimensional data into low-dimensional space using probability distributions. It excels at visualizing clusters but can distort global structure and is sensitive to parameters like perplexity. van der Maaten et al. 2008.

How do visualise what DA is doing?

Dimensionality reduction techniques

We need to be able to examine the multi-dimensional laten (embedding) space i.e. go from ND to 2D or 3D.



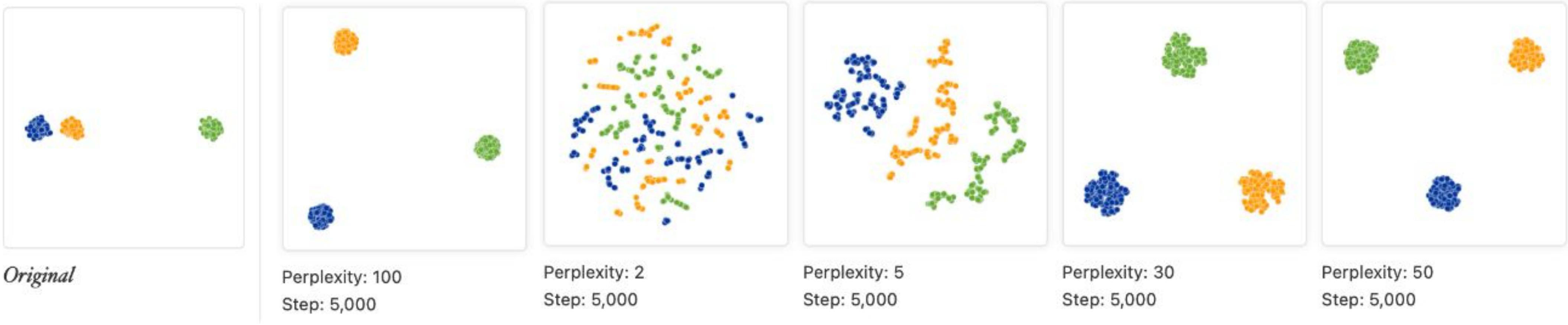


Figure from: [Distil]

How do visualise what DA is doing?

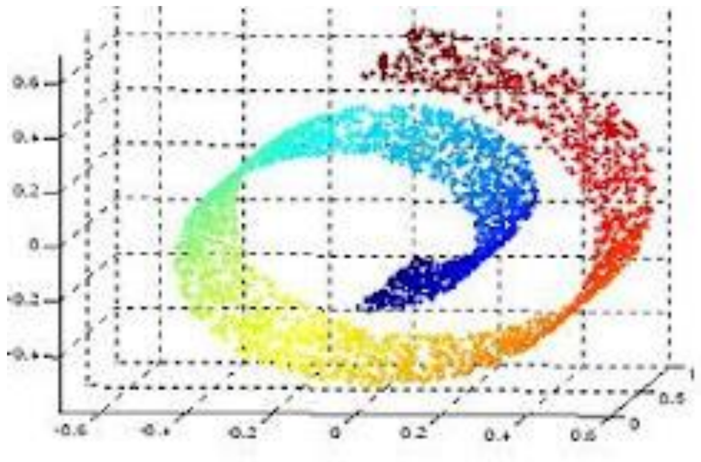
Dimensionality reduction techniques

We need to be able to examine the multi-dimensional laten (embedding) space i.e. go from ND to 2D or 3D.

tSNE

Isomap

Umap



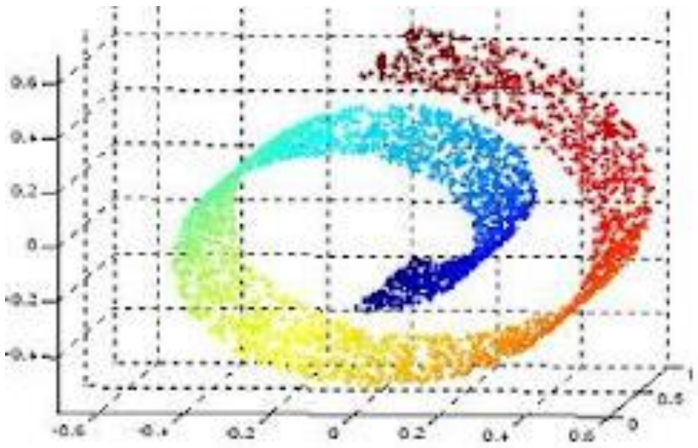
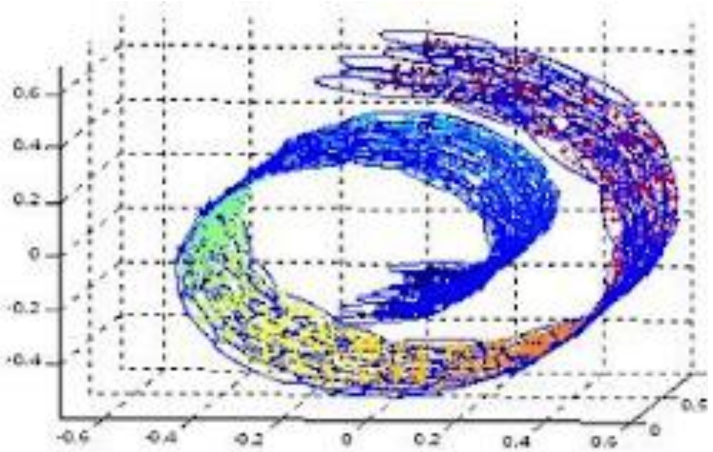
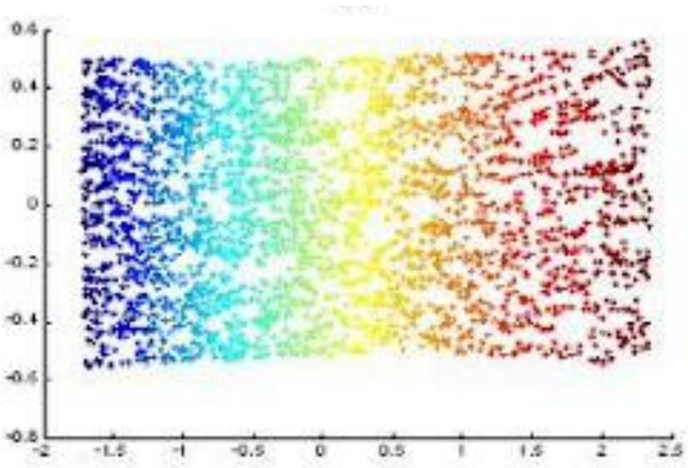
Isometric Mapping

Isomap is a manifold learning method that preserves global geometry by computing geodesic distances along a graph connecting nearest neighbors. It is useful for capturing the underlying structure of data that lies on a nonlinear manifold. Tenenbaum et al. 2000.

How do visualise what DA is doing?

Dimensionality reduction techniques

We need to be able to examine the multi-dimensional laten (embedding) space i.e. go from ND to 2D or 3D.

tSNE	**Isomap**	Umap
		

How do visualise what DA is doing?

Dimensionality reduction techniques

We need to be able to examine the multi-dimensional laten (embedding) space i.e. go from ND to 2D or 3D.

tSNE

Isomap

Umap

~4 million single-cell transcriptomes from adult mouse brain labeled by source brain region represented by a UMAP (Yao Z. et al. 2023, bioRxiv).

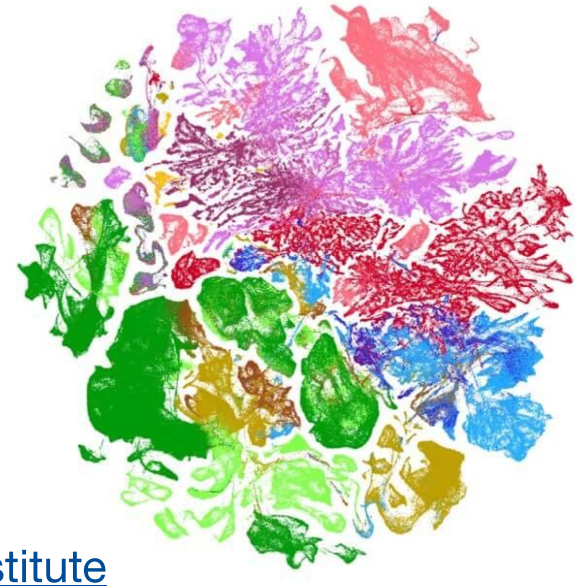


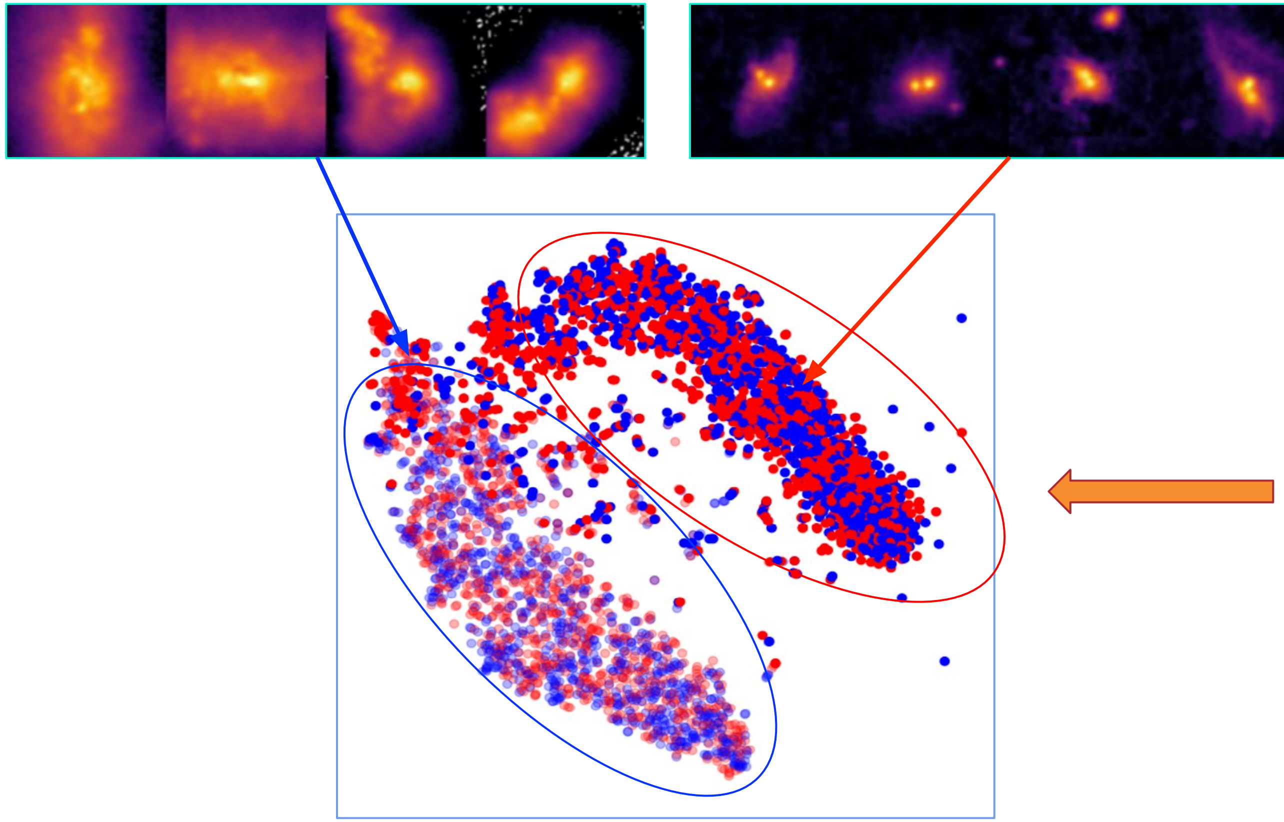
Figure from: [Allen Institute]

Uniform Manifold Approximation and Projection

UMAP is a fast, scalable nonlinear dimensionality reduction technique that preserves both local and some global structure using fuzzy topological representations. It often outperforms t-SNE in preserving global relationships and is more suitable for large datasets. McInnes et al. 2020.

Latent Space Alignment

Source - Illustris **Target - SDSS observations**

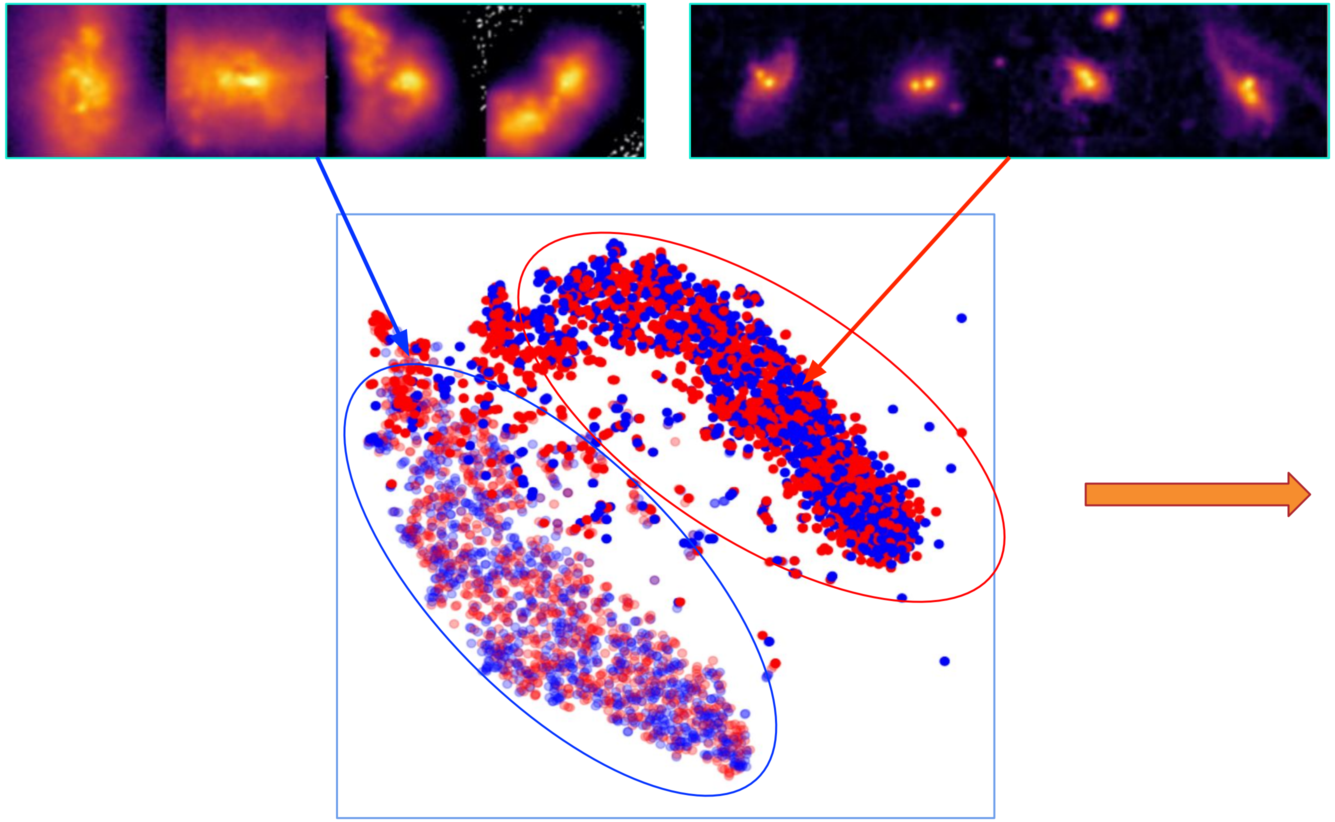


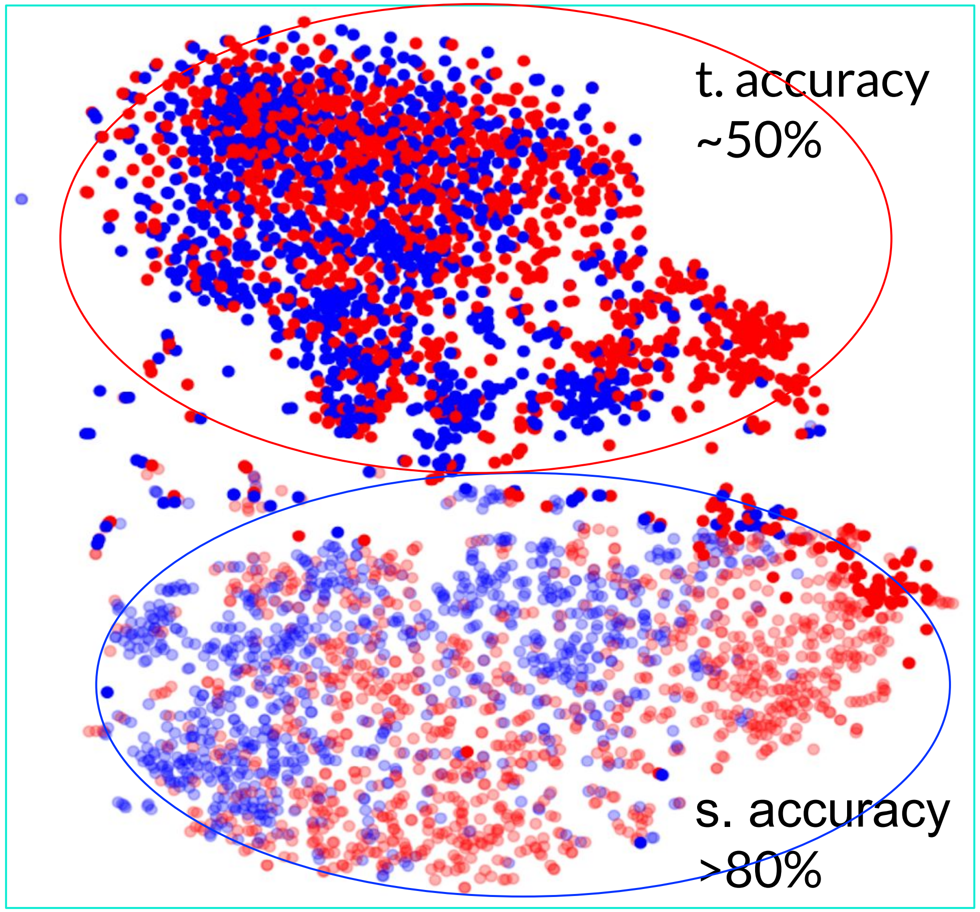
This is how the network sees the data.
2D representation of network's latent space.

Ćiprijanović et al. 2020.
Ćiprijanović et al. 2021.

Latent Space Alignment

Source - Illustris **Target - SDSS observations**



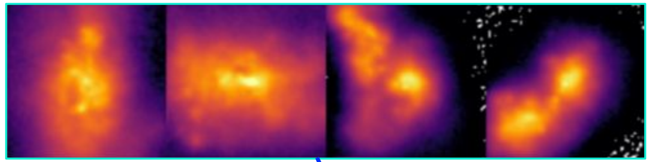


Regular Training

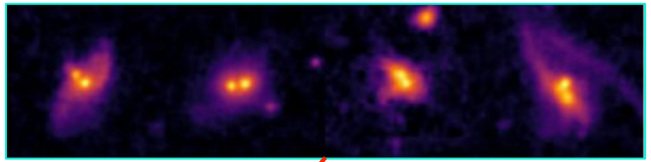
Ćiprijanović et al. 2020.
Ćiprijanović et al. 2021.

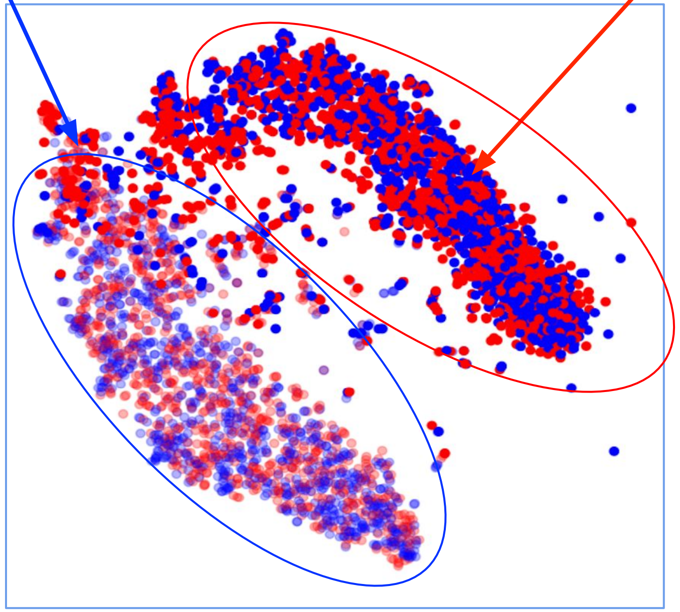
Latent Space Alignment

Source - Illustris



Target - SDSS observations

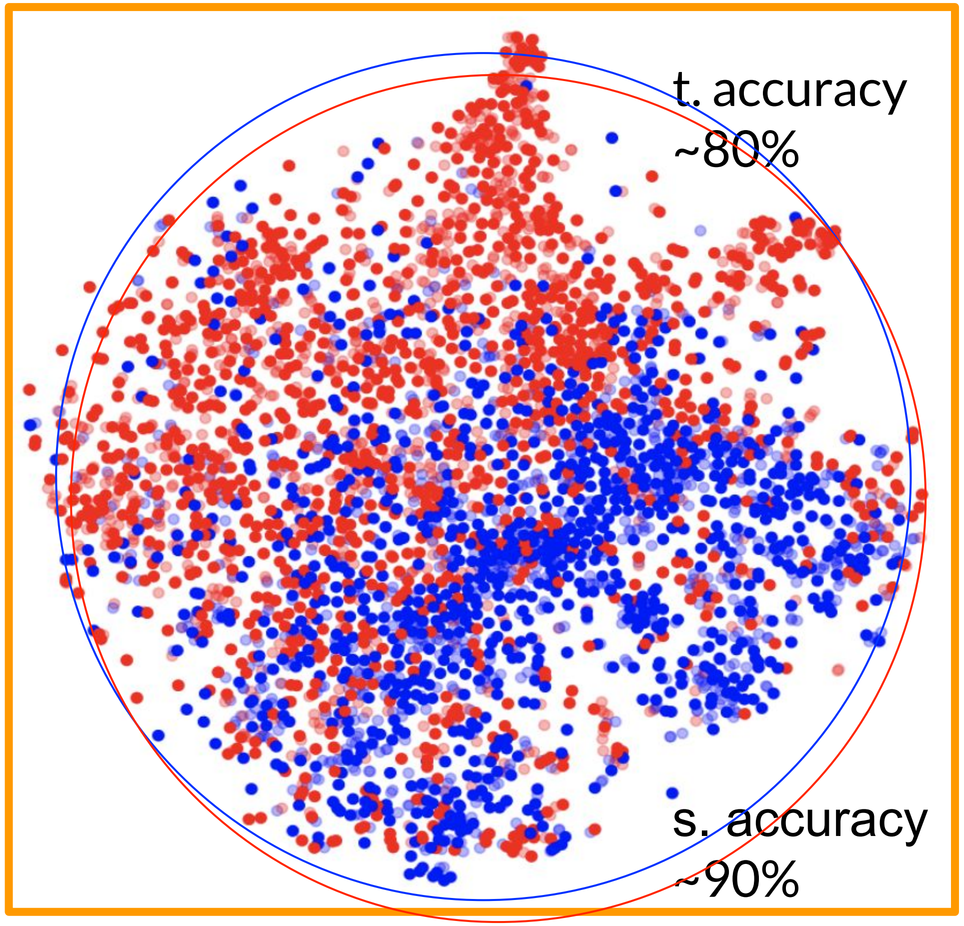




Ćiprijanović et al. 2020.
Ćiprijanović et al. 2021.

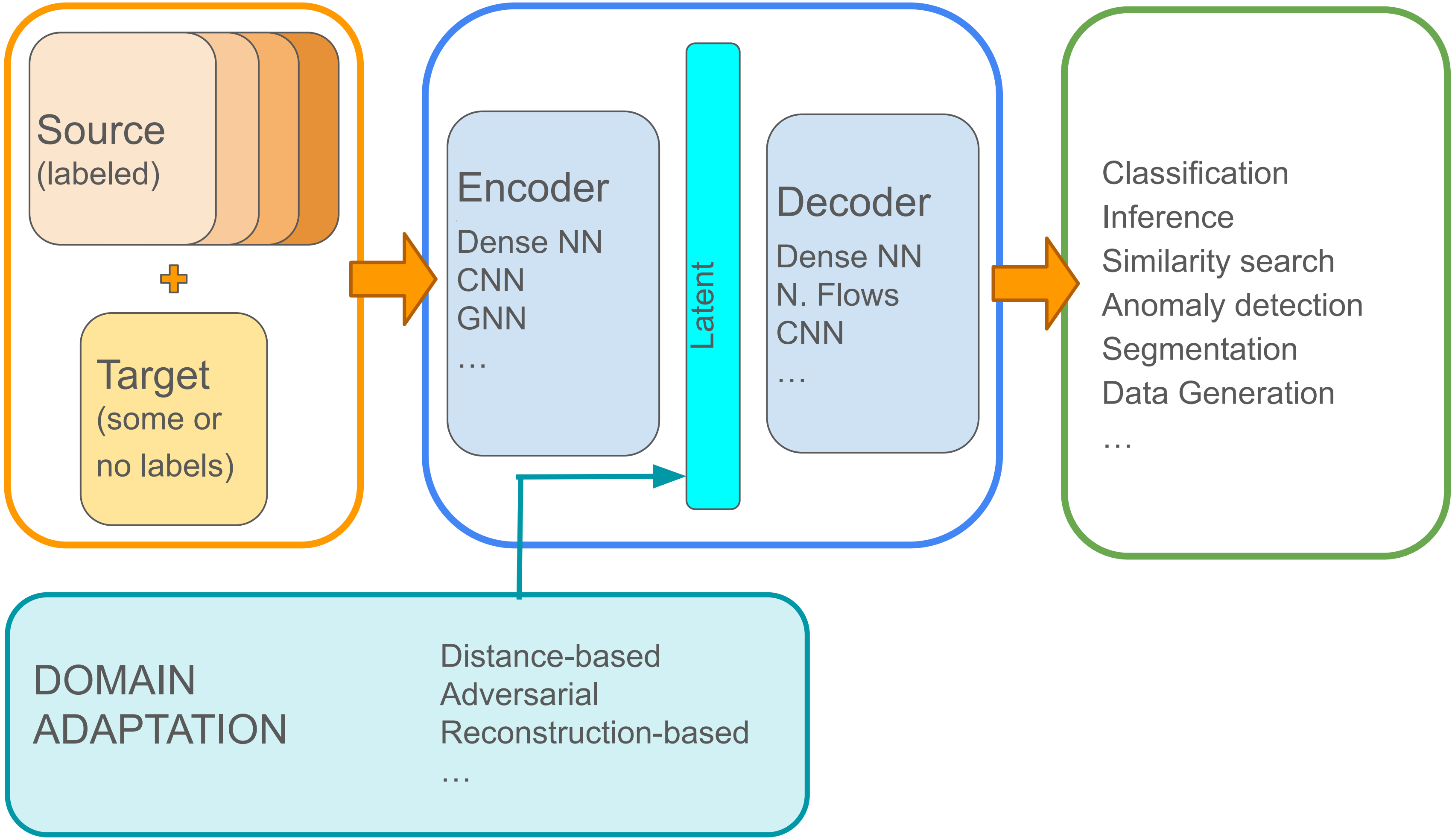
Domain Adaptation

Up to 30% increase!



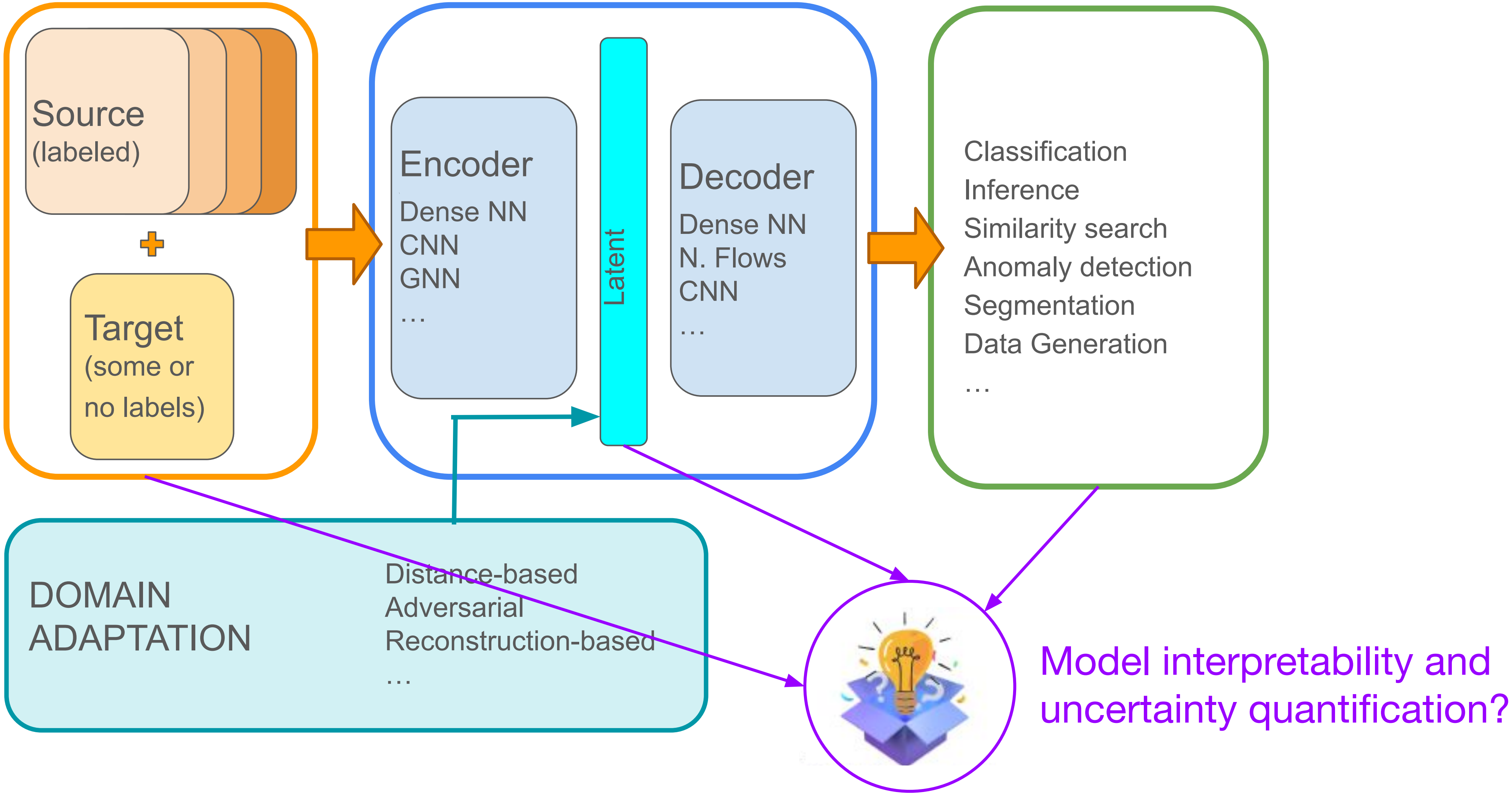
Where can we use it?

Any data, any task, any problem (within reason!)



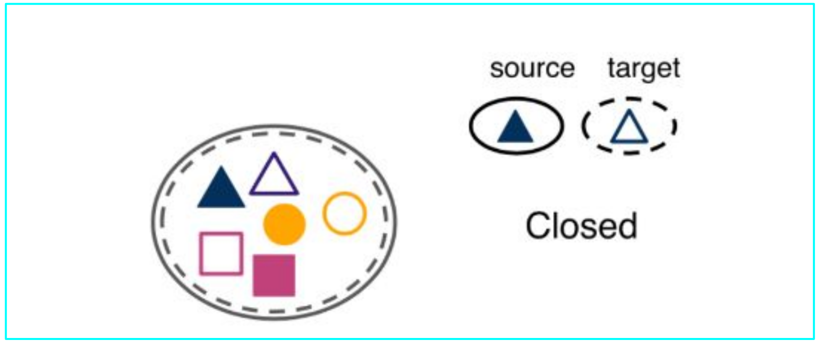
Where can we use it?

Any data, any task, any problem (within reason!)



Limitations and things to be careful about

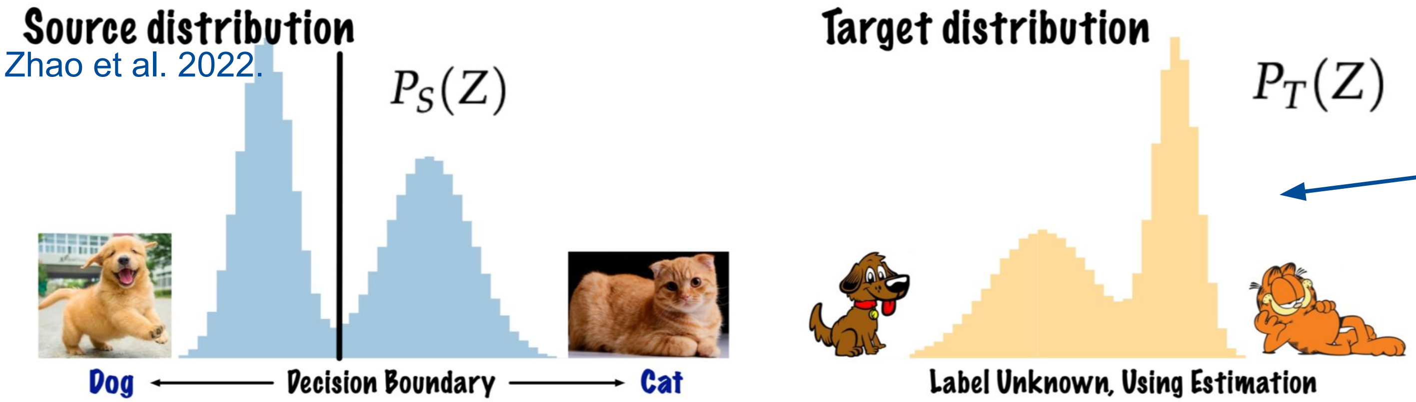
Types of domain overlap

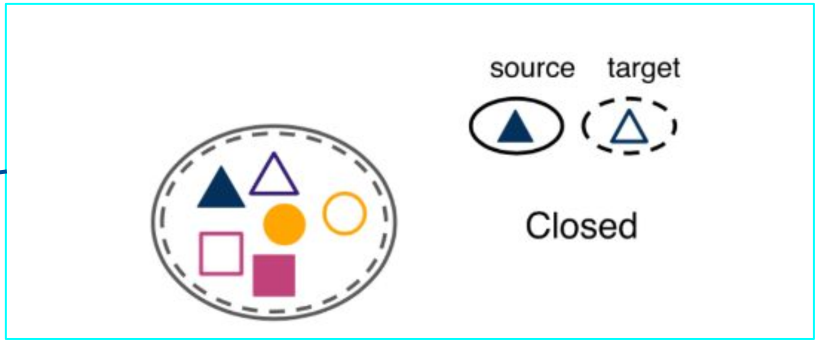


Limitations and things to be careful about

Types of domain overlap

Zhao et al. 2022.

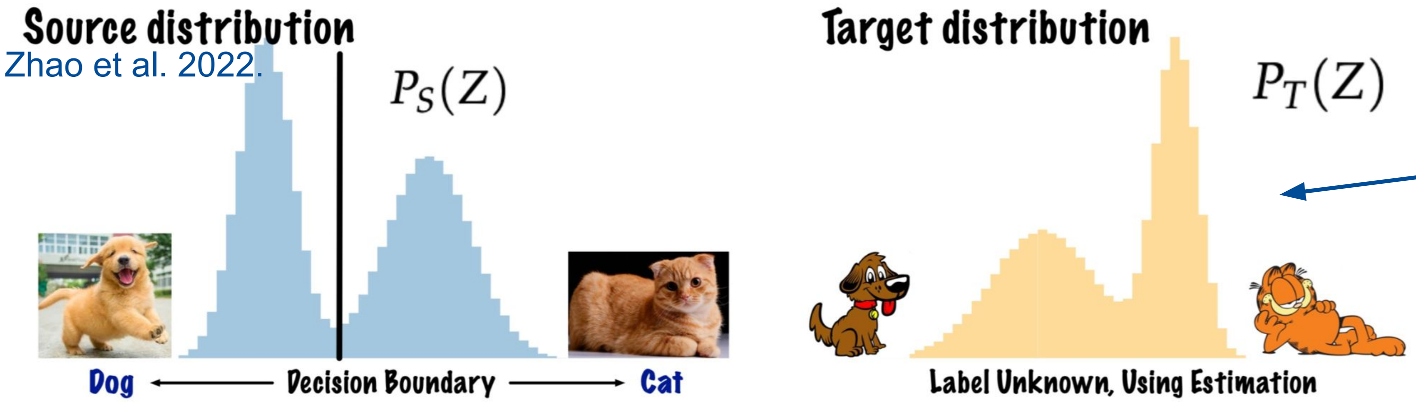




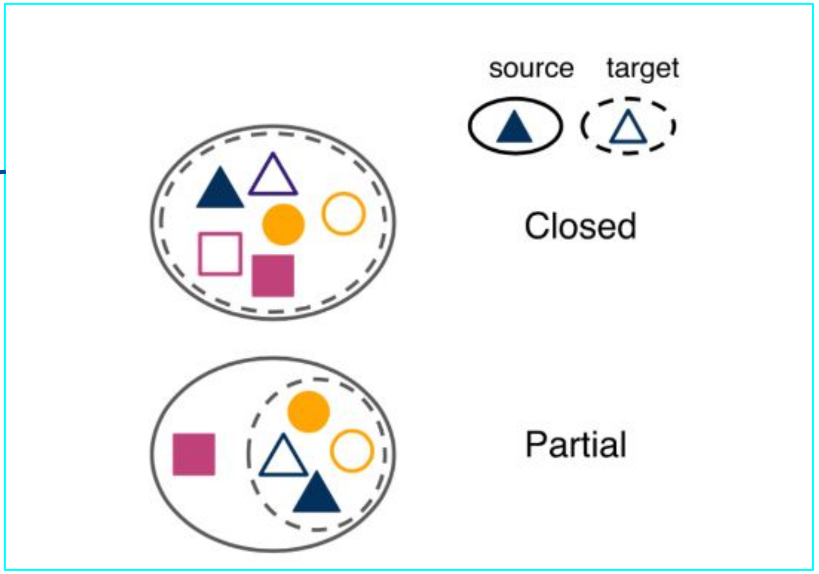
Limitations and things to be careful about

Types of domain overlap

Zhao et al. 2022.

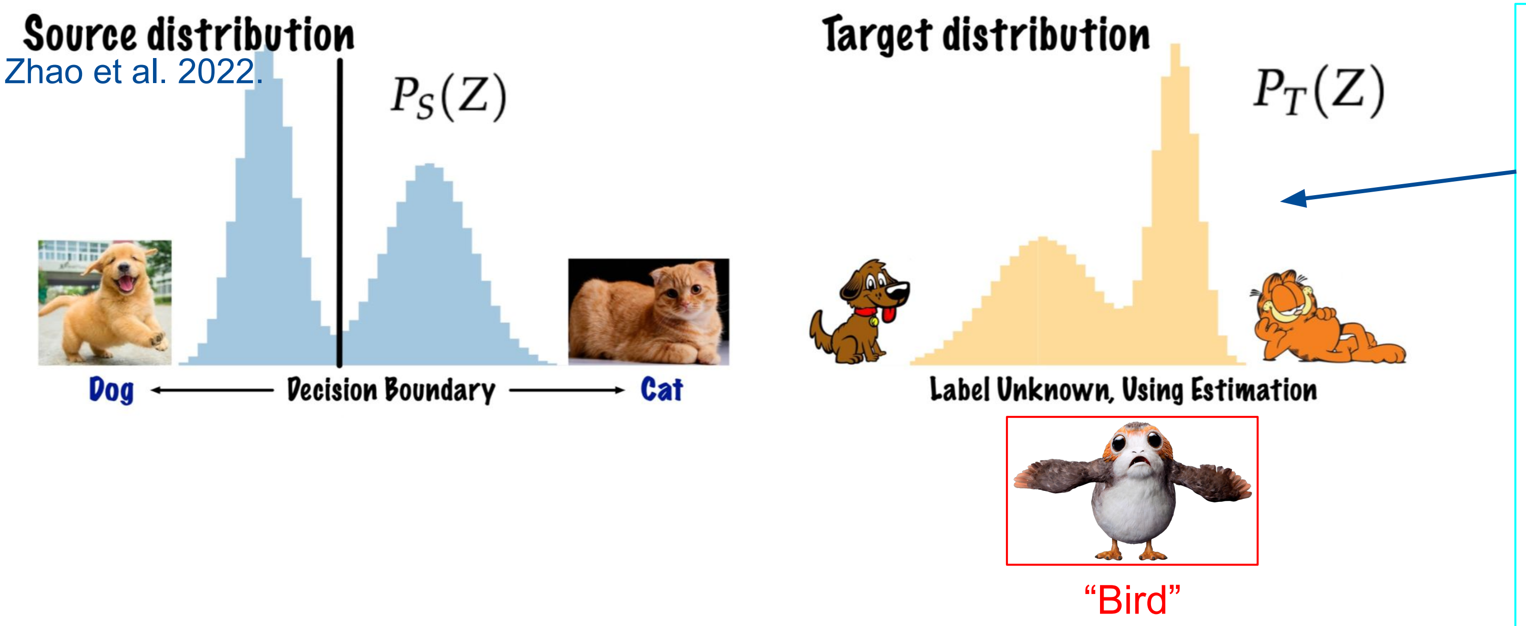


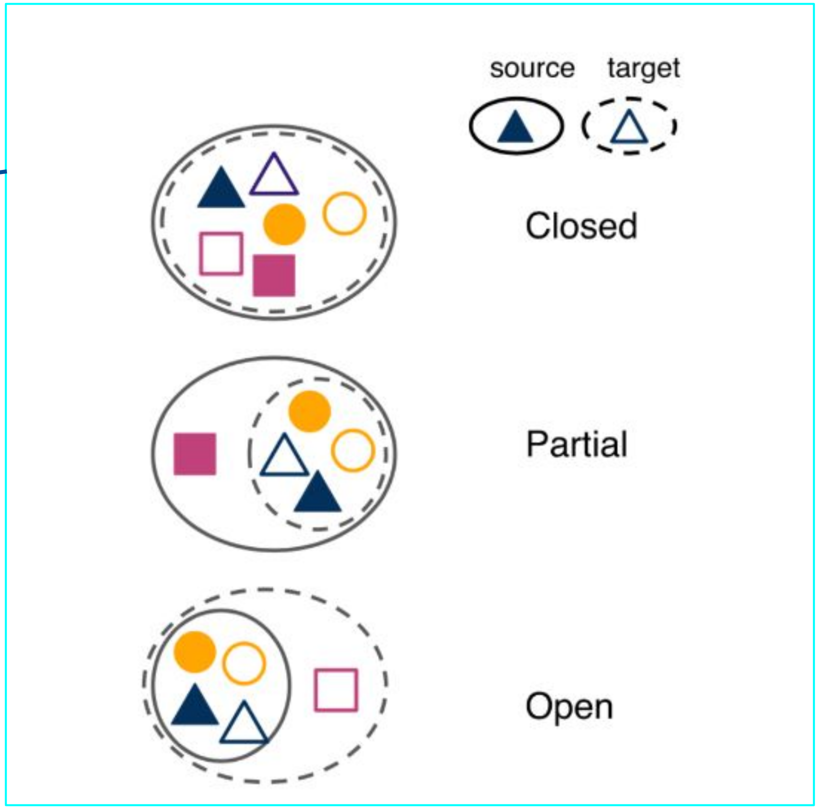
"Bird"



Limitations and things to be careful about

Types of domain overlap

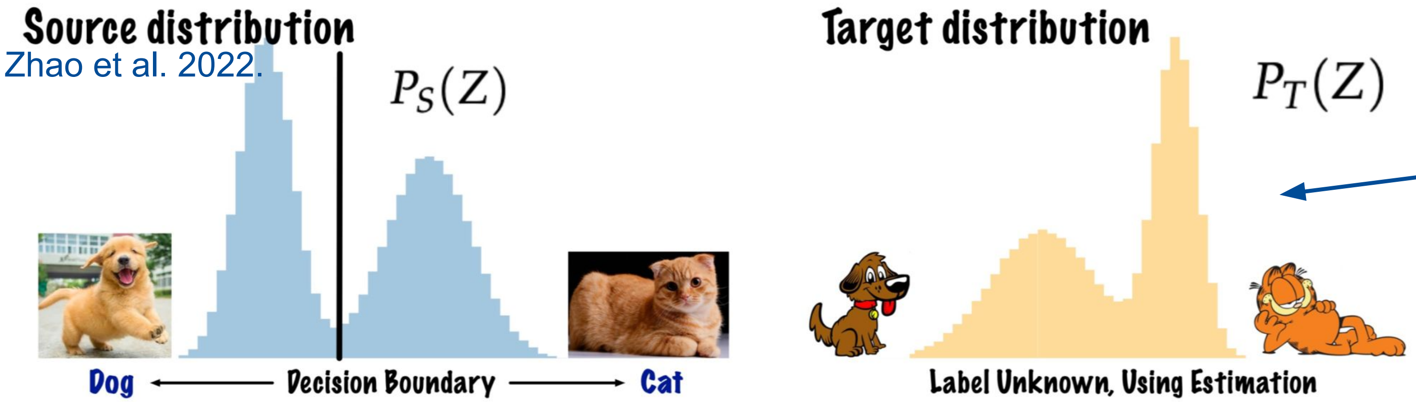




Limitations and things to be careful about

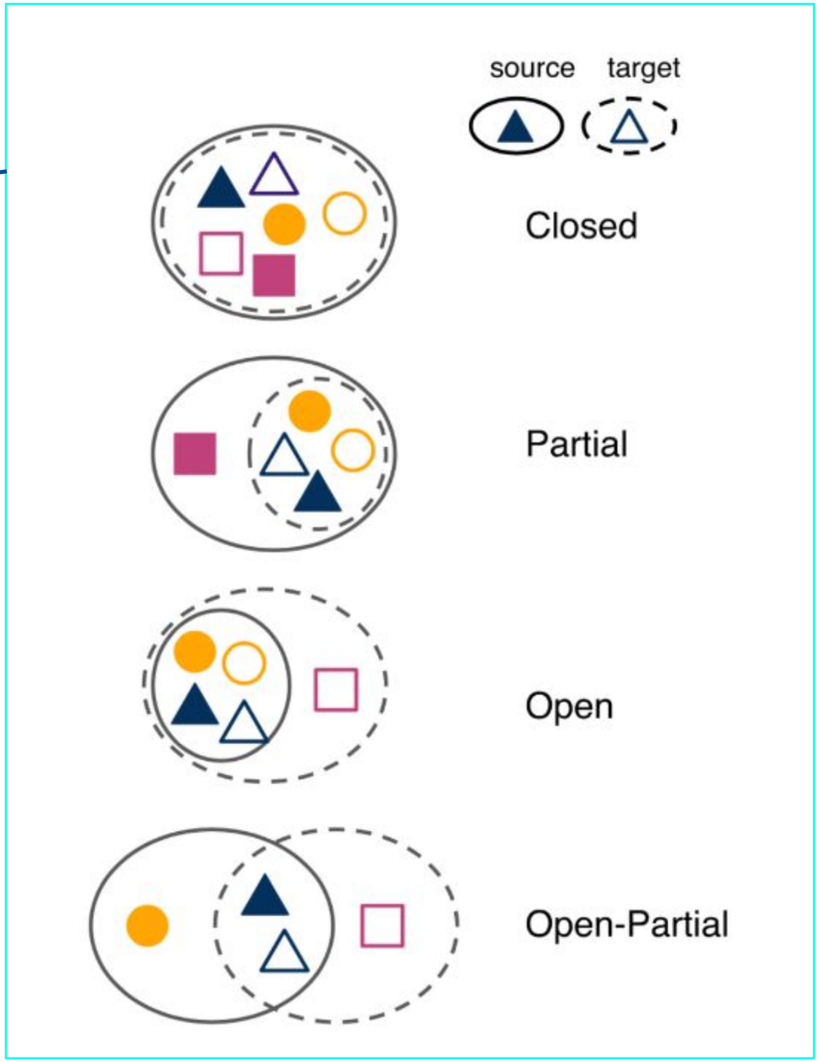
Types of domain overlap

Zhao et al. 2022.



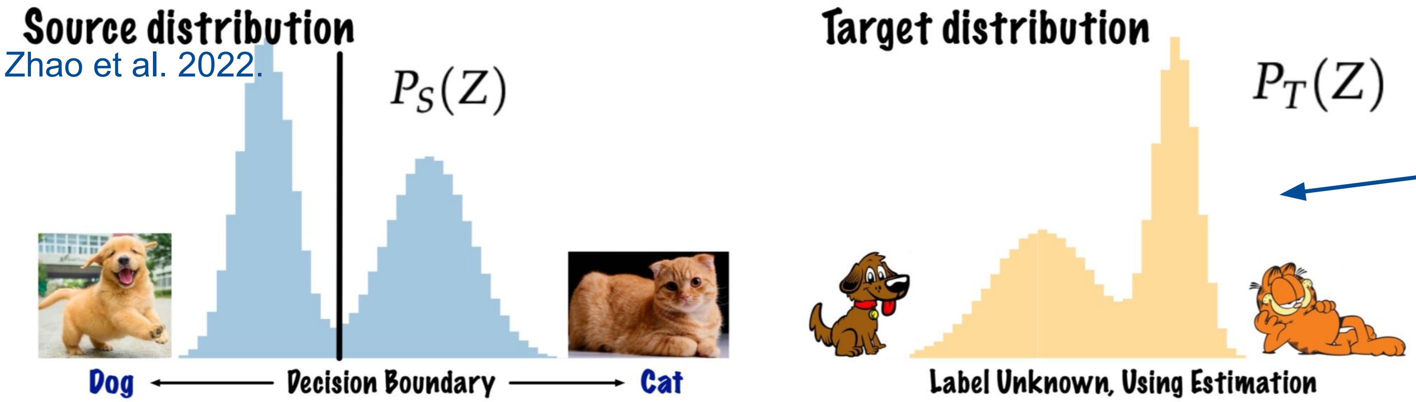
"Bear"

"Bird"



Limitations and things to be careful about

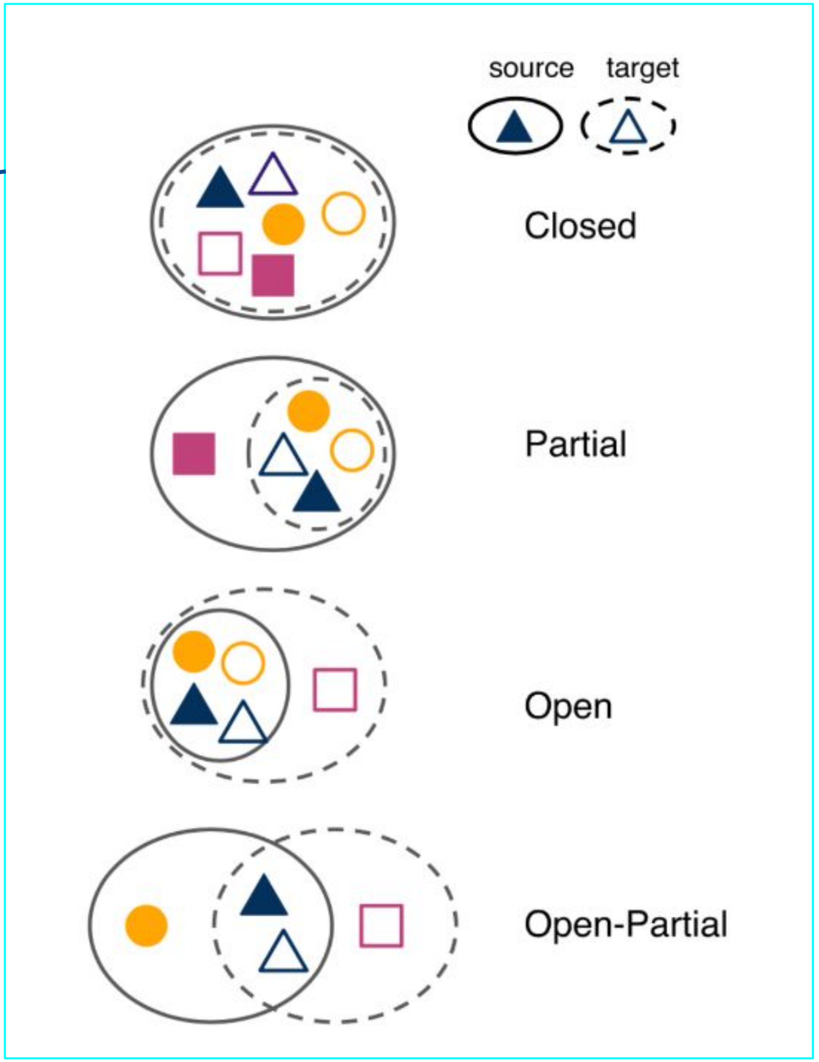
Types of domain overlap



Zhao et al. 2022.

“Bear”

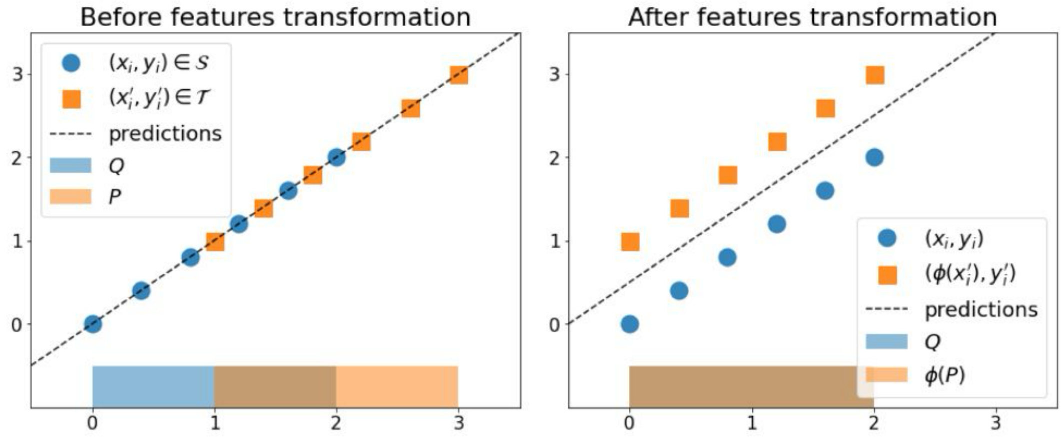
“Bird”



- We can have extra classes present in only one or both of the datasets!
- We might not even know about it!
    - Mutual classes should overlap.
    - Non-overlapping classes should not be aligned with anything - not all DA methods can do this!

Limitations for regression problems

Be careful if prior ranges of parameters being inferred do not match across domains!



de Mathelin et al. 2021.

Fig. 1. Negative transfer caused by feature-based methods. Matching Q and P is detrimental here, as the shift between the two distributions is informative with respect to the task.

Feature-based algorithms (distance or adversarial based) cannot be used if prior ranges don't match.

Only is differences between the domains is related to parameters not being inferred (noise, pixel scale etc).

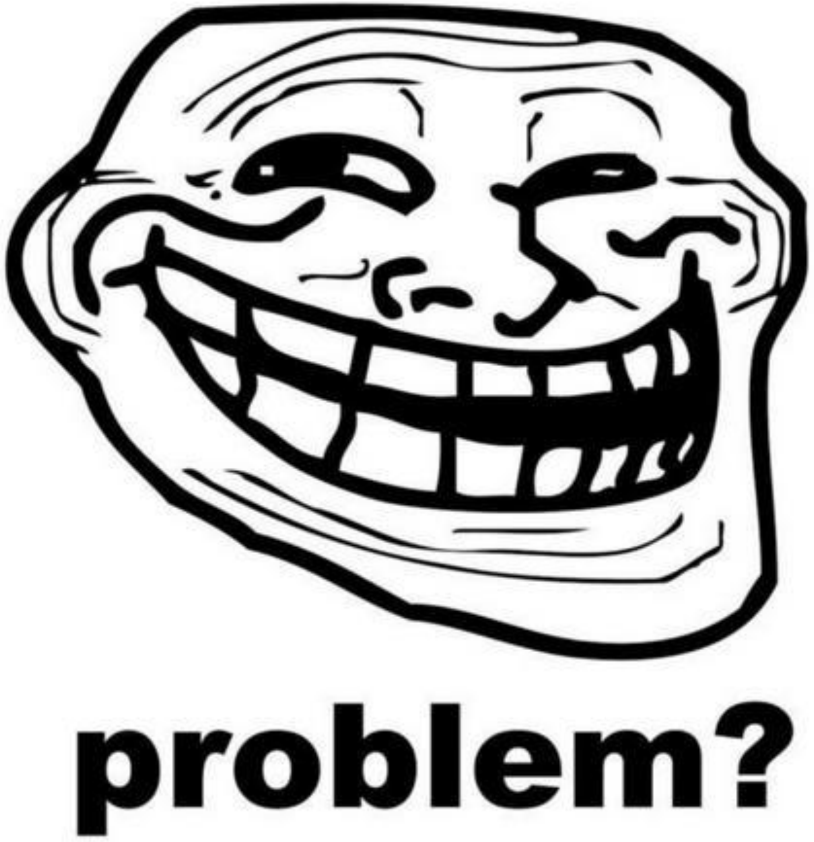
Instance-based algorithms are necessary - reweighting importance of labeled instances in the overlapping region.

Domain Adaptation Problems Quiz !

Scenario 1:
You apply a domain adaptation technique to align galaxy images from two telescopes. After adaptation, your model performs ***worse*** on the source data and loses its ability to distinguish spiral and elliptical galaxies. What likely went wrong?

Learning rate too small	Overfitting to source data	Too aggressive domain alignment

Answer:

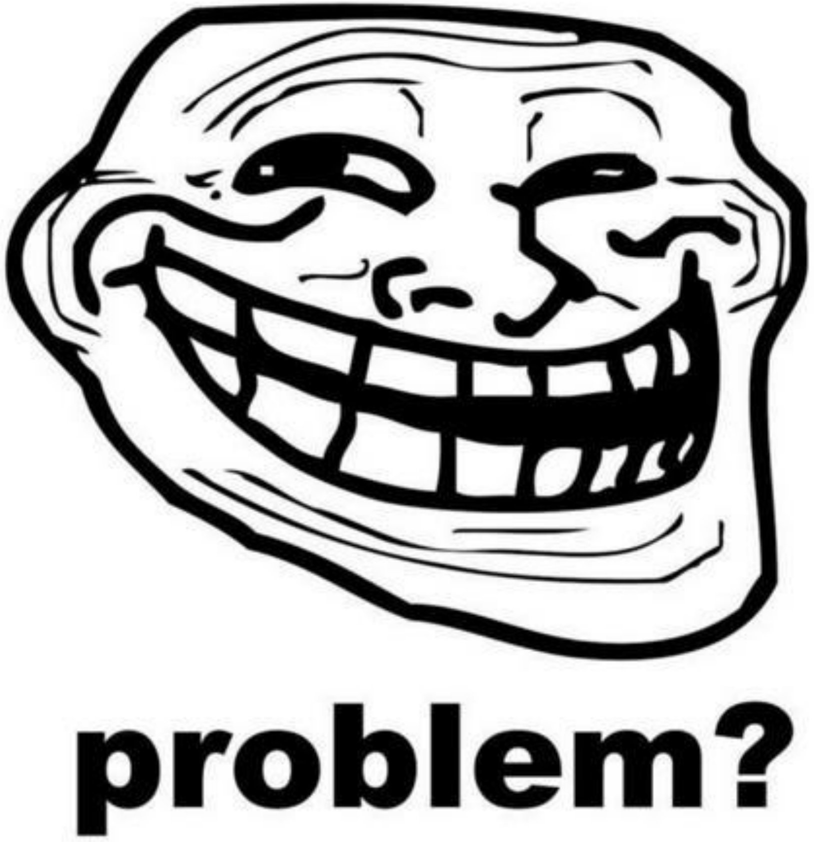


Domain Adaptation Problems Quiz !

Scenario 1:

You apply a domain adaptation technique to align galaxy images from two telescopes. After adaptation, your model performs ***worse*** on the source data and loses its ability to distinguish spiral and elliptical galaxies. What likely went wrong?

Learning rate too small	Overfitting to source data	Too aggressive domain alignment



Answer:

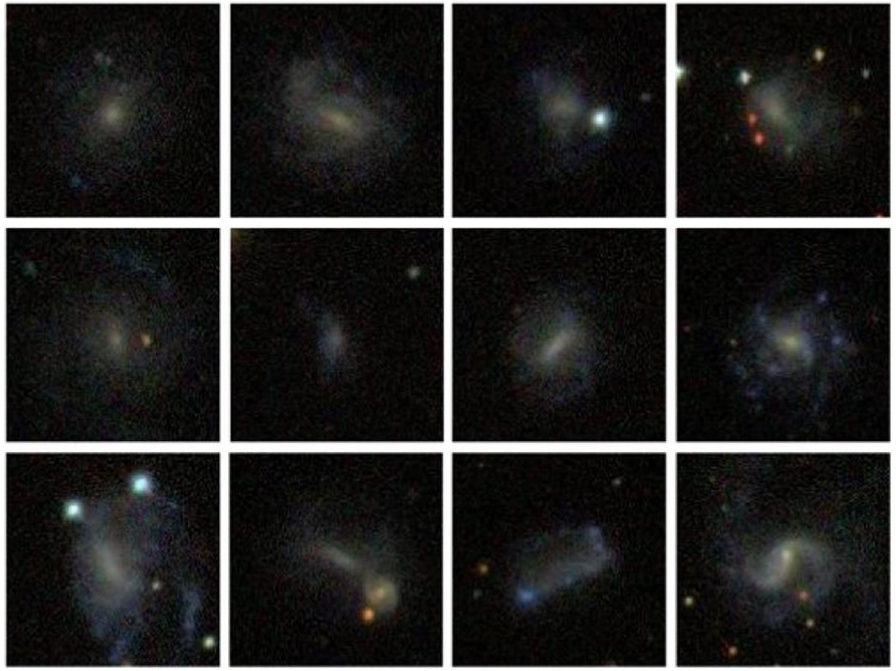
This is a classic **over-alignment** issue: when features are overly forced to match across domains, the model can discard **discriminative features** that are important for the task. This is especially risky in scientific domains where small differences matter, or domain overlap is small. Careful balance between alignment and task supervision is crucial.

Domain Adaptation Problems Quiz !

Scenario 2:

You trained a classifier on old survey data where low-surface brightness galaxies are rare, but in the new LSST survey, they dominate. Your adapted model performs poorly on the target domain, despite good feature alignment. What is the most likely cause?

We didn't account for label distribution shift	Feature extractor wasn't pre-trained	Training didn't converge



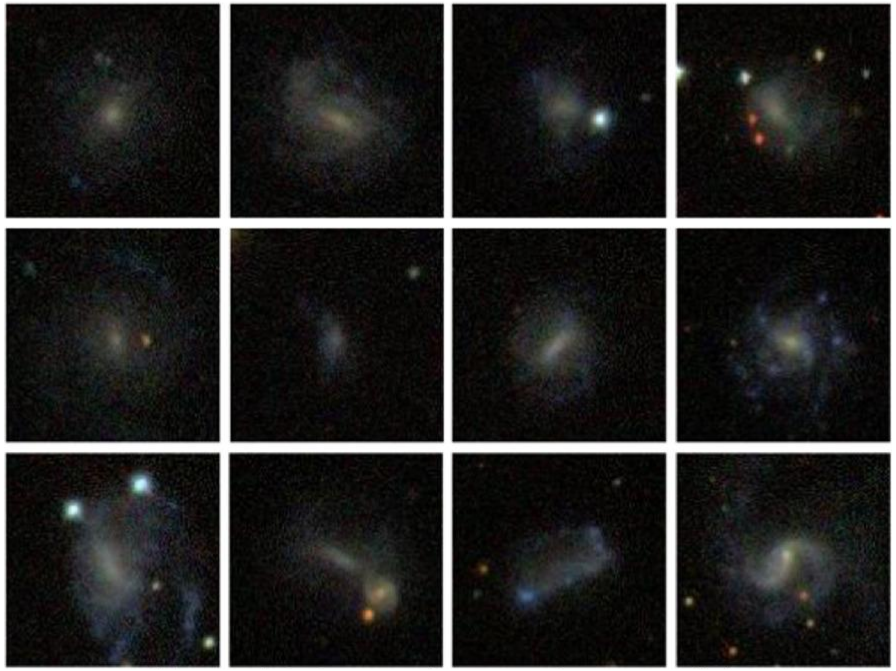
Answer:

Domain Adaptation Problems Quiz !

Scenario 2:

You trained a classifier on old survey data where low-surface brightness galaxies are rare, but in the new LSST survey, they dominate. Your adapted model performs poorly on the target domain, despite good feature alignment. What is the most likely cause?

We didn't account for label distribution shift	Feature extractor wasn't pre-trained	Training didn't converge



Answer:

Most domain adaptation methods assume that P**(Y|X)** stays the same (covariate shift). If instead P**(Y)** or P**(Y|X)** changes, models trained on one distribution may misclassify dominant classes in the new domain. Solutions like target-prior correction, self-supervision or instance reweighting are needed.

FERMI NATIONAL ACCELERATOR LABORATORY

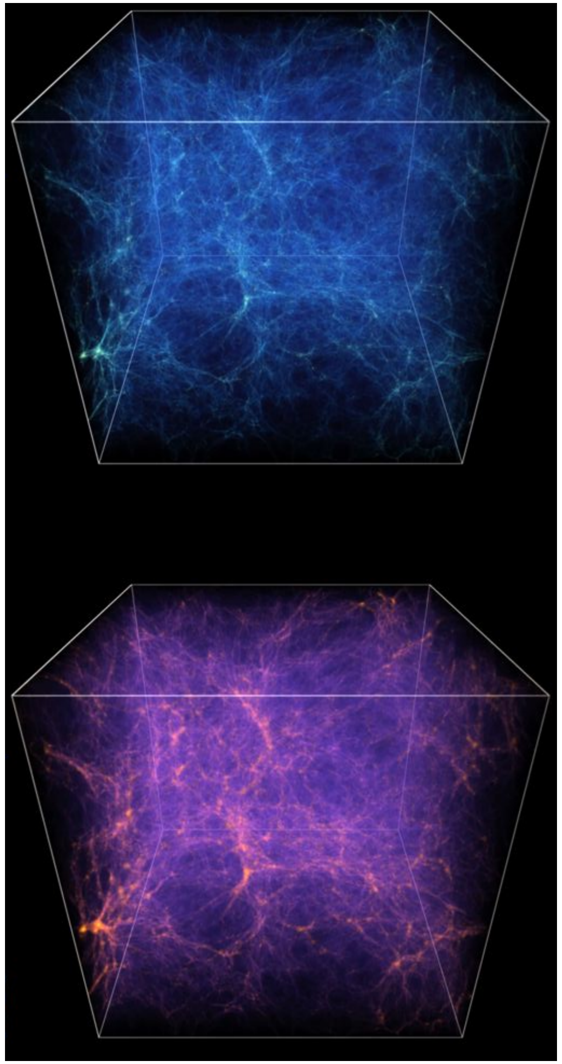
Domain Adaptation Problems Quiz !

Scenario 3:

You apply a domain adaptation method assuming two datasets come from distinct domains. However, both were collected under nearly identical conditions, just with different random seeds. Model performance drops after adaptation. What likely went wrong?

You overestimated domain shift	Domain alignment was overdone	Overfitting

Answer:

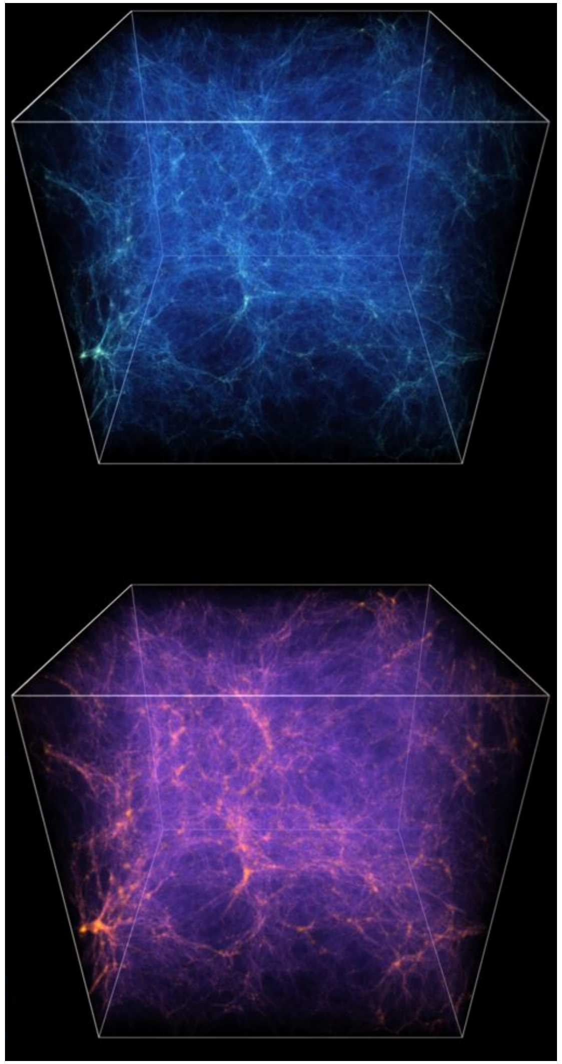


Domain Adaptation Problems Quiz !

Scenario 3:

You apply a domain adaptation method assuming two datasets come from distinct domains. However, both were collected under nearly identical conditions, just with different random seeds. Model performance drops after adaptation. What likely went wrong?

You overestimated domain shift	Domain alignment was overdone	Overfitting



Answer:

Applying domain adaptation when **no meaningful domain gap exists** can **inject noise** or unnecessary transformations into the model. **False domain shift** can degrade performance because of **spurious alignment pressure**. You end up learning representations to solve a non-existent problem, which may reduce task performance and hurt generalization.

Useful Resources:

- [“SIDDA: SInkhorn Dynamic Domain Adaptation for Image Classification with Equivariant Neural Networks” Pandya et al. 2025.]
- [“Domain-Adaptive Neural Posterior Estimation for Strong Gravitational Lens Analysis” Swierc et al. 2024.]
- [“Domain Adaptive Graph Neural Networks for Constraining Cosmological Parameters Across Multiple Data Sets” Roncoli et al. 2023.]

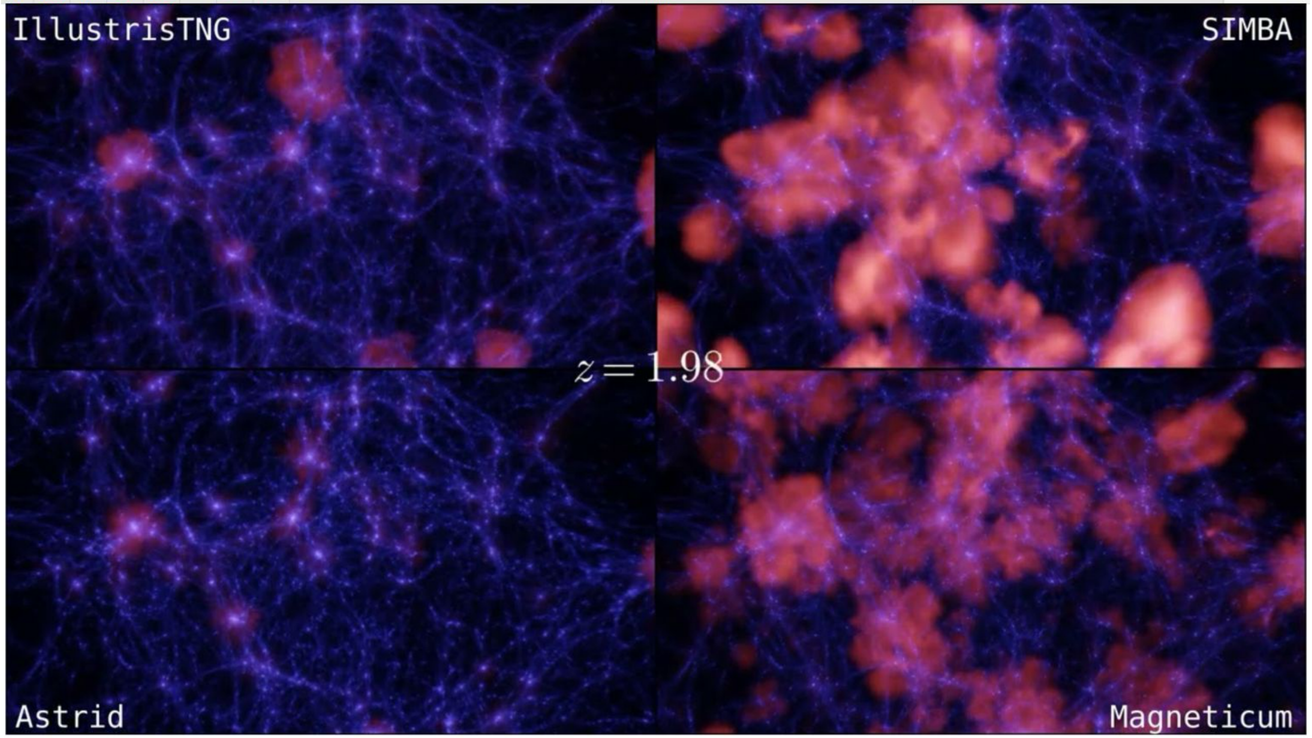
03

Examples

Cosmology with Graphs

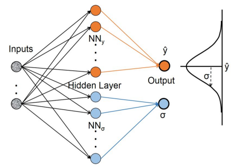
NeurIPS 2023.
Roncoli et al. 2023.

Can we correctly predict cosmology across different cosmological simulations?



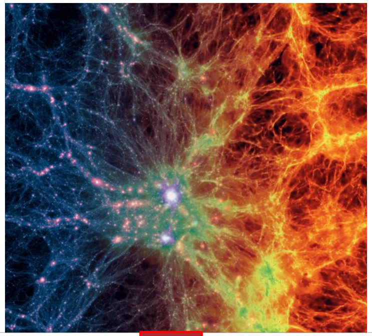
Cosmology with Graphs

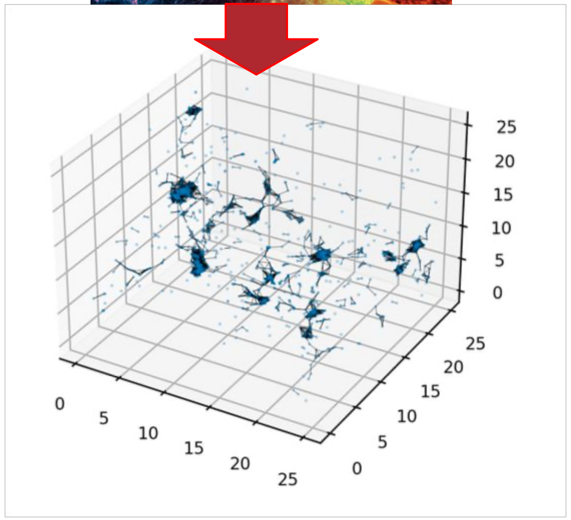
MVE: Infer matter density cosmological parameter



NeurIPS 2023.
Roncoli et al. 2023.

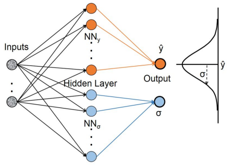
Graph Neural Networks:
ideal for sparse
galaxy catalogs!





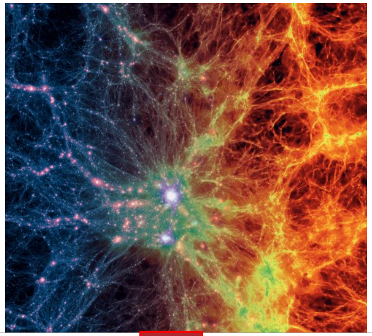
Cosmology with Graphs

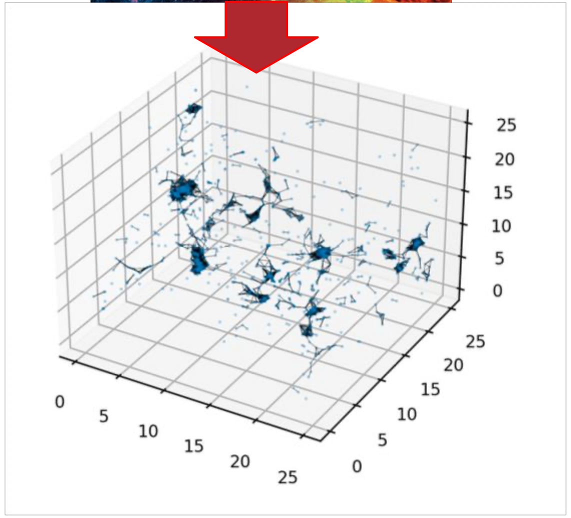
MVE: Infer matter density cosmological parameter

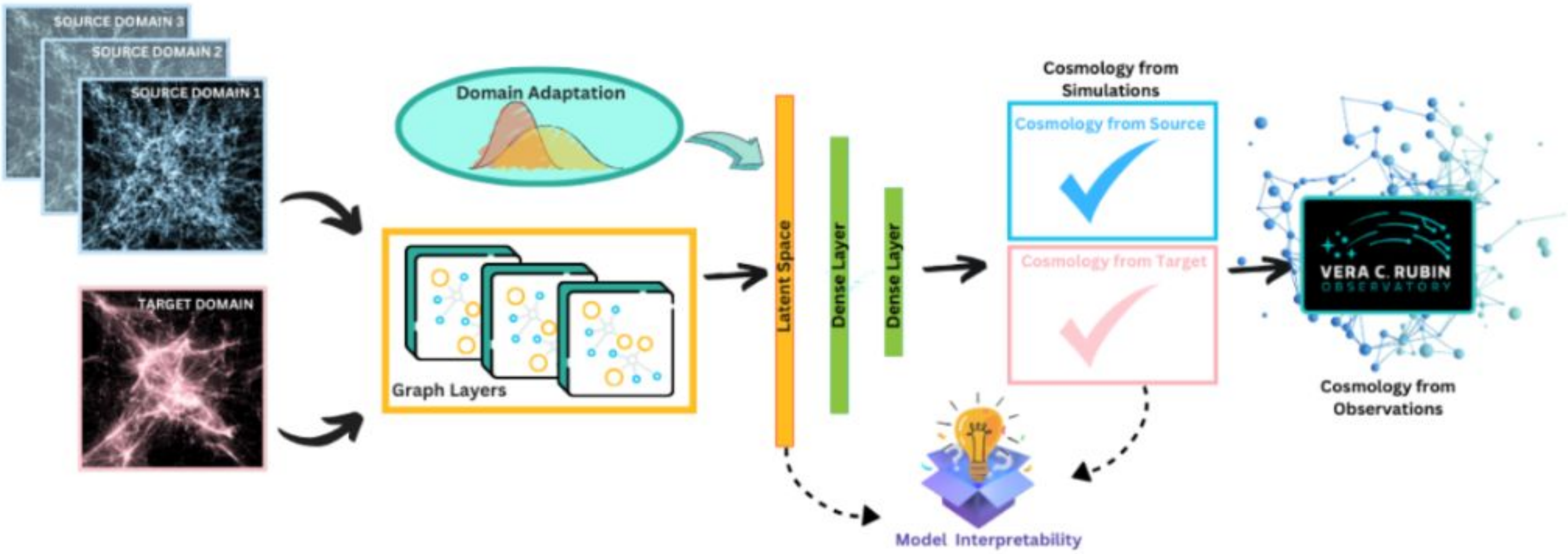


NeurIPS 2023.
Roncoli et al. 2023.

Graph Neural Networks:
ideal for sparse
galaxy catalogs!



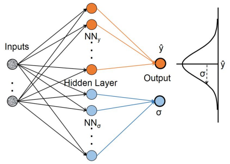




FERMI NATIONAL ACCELERATOR LABORATORY

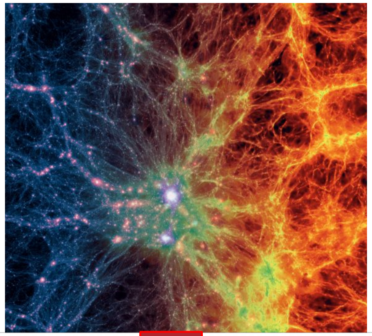
Cosmology with Graphs

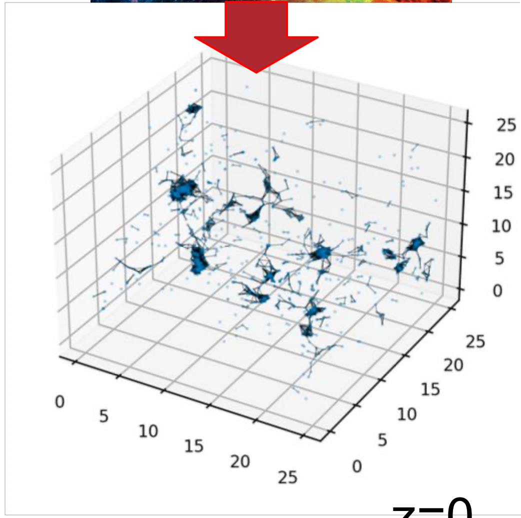
MVE: Infer matter density cosmological parameter

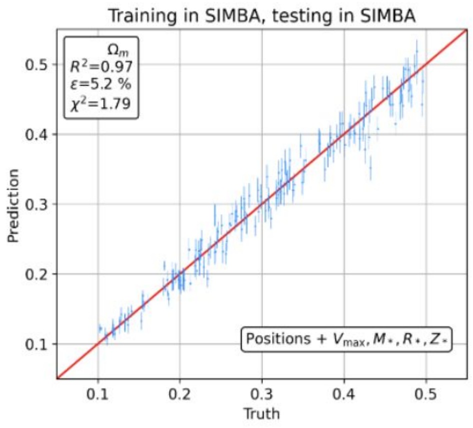


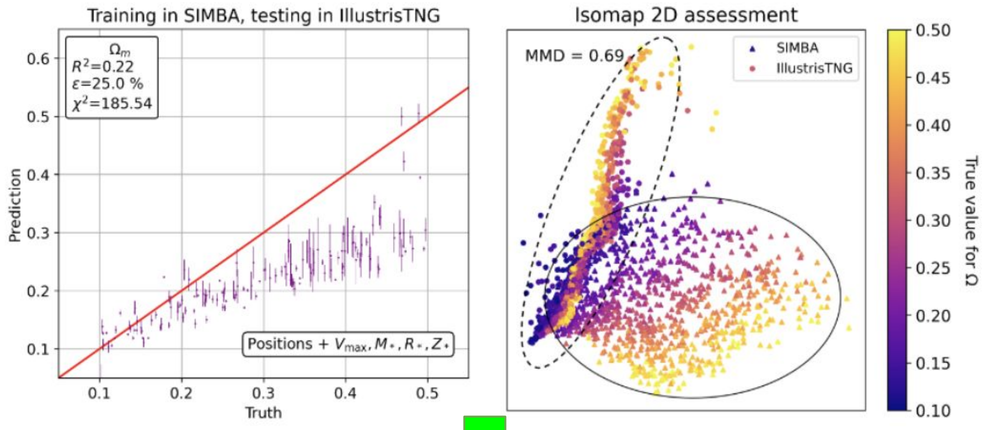
NeurIPS 2023.
Roncoli et al. 2023.

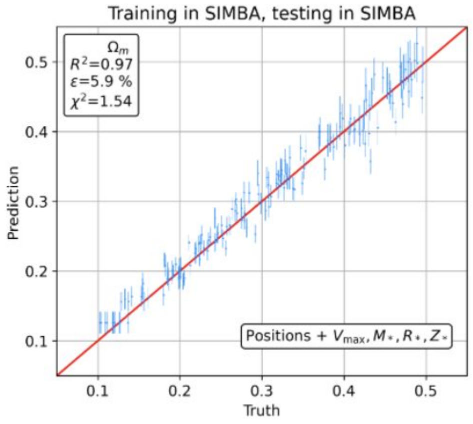
Graph Neural Networks:
ideal for sparse
galaxy catalogs!

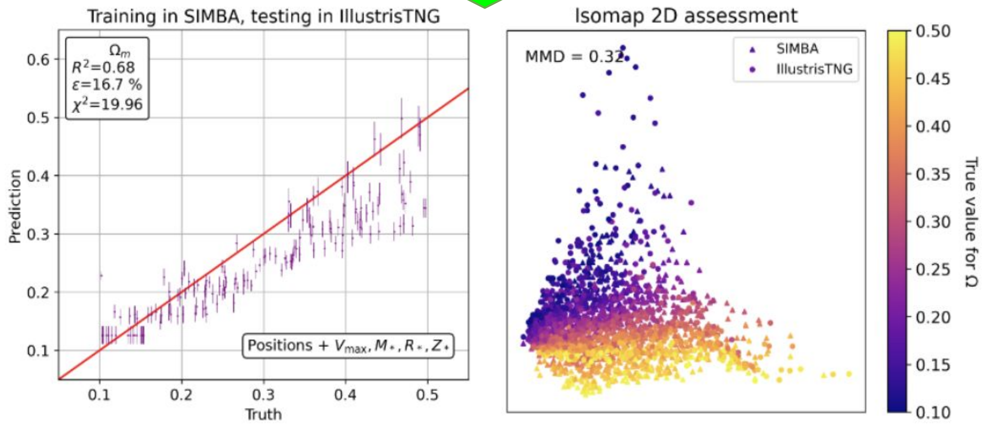












z=0 1000 simulations each

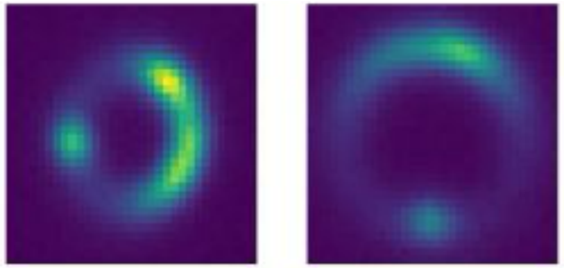
28% better relative error

Order of mag. better χ^2

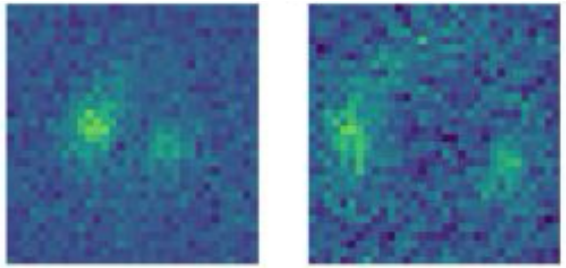
FERMI NATIONAL ACCELERATOR LABORATORY

Strong gravitational lensing

Can we infer strong lensing parameters (Einstein radius, ellipticity, position/offset) robustly in both simulated and real data?



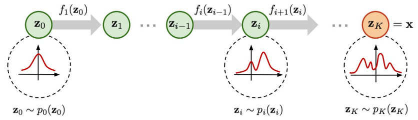
Source - Clean

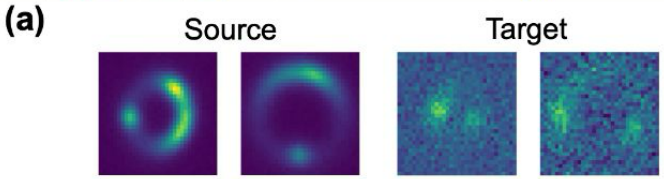


Target - DES

Strong gravitational lensing

SBI - infer Einstein radius, ellipticity, position/offset



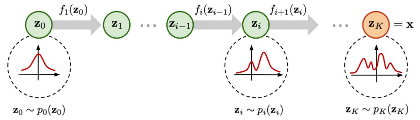


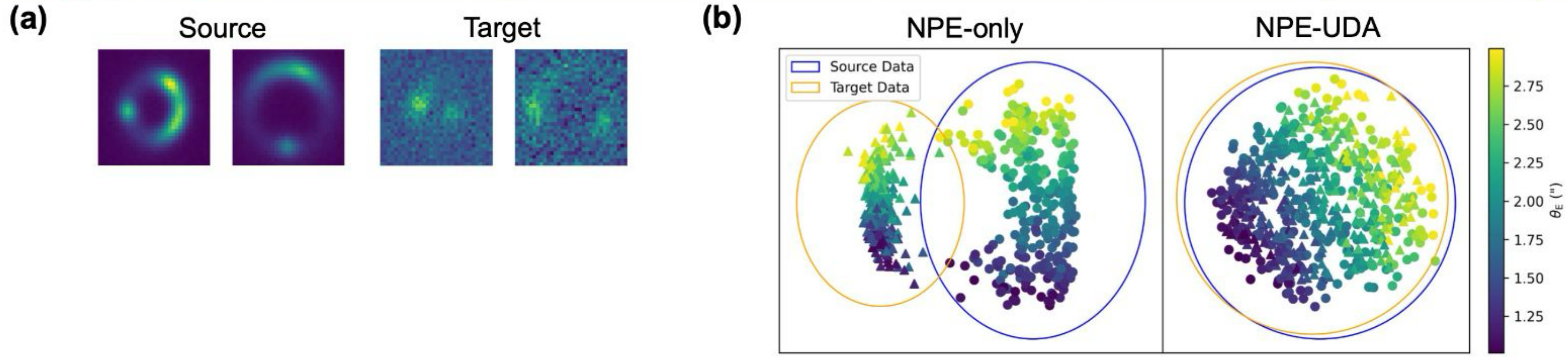
a. 100k images (1-filter) with and without noise
b. **Latent distributions are correctly aligned**.
c. Differences between true values and predicted posteriors are small. **Accuracy improves up to 2 orders of magnitude**.
d. Posterior are well calibrated and no longer overconfident.

Swierc et al. 2024.

Strong gravitational lensing

SBI - infer Einstein radius, ellipticity, position/offset



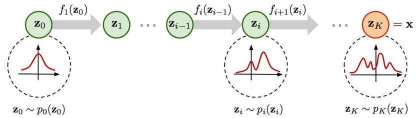


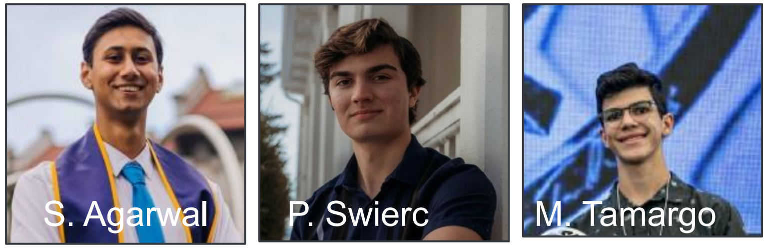
a. 100k images (1-filter) with and without noise
b. **Latent distributions are correctly aligned**.
c. Differences between true values and predicted posteriors are small. **Accuracy improves up to 2 orders of magnitude**.
d. Posterior are well calibrated and no longer overconfident.

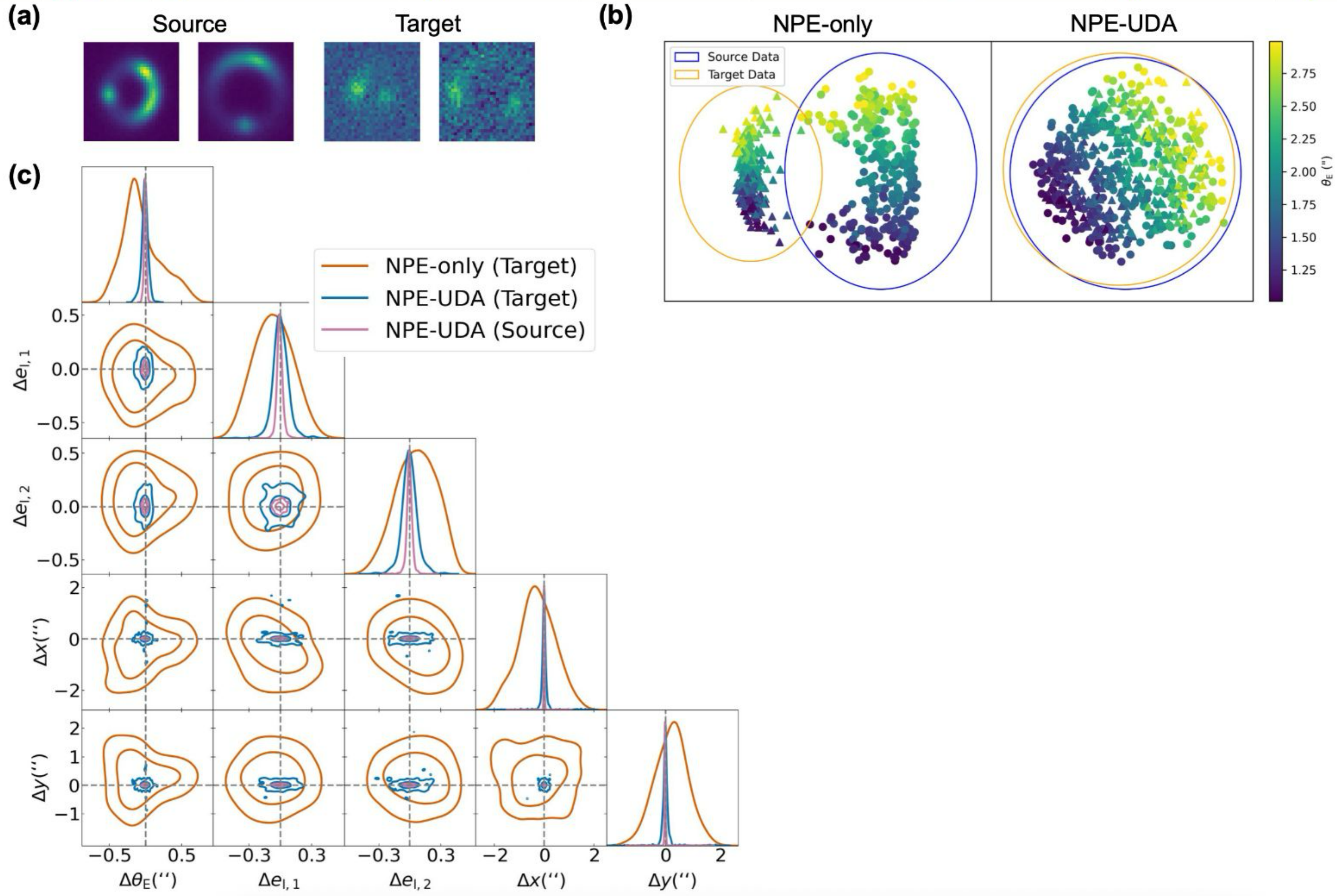
Swierc et al. 2024.

Strong gravitational lensing

SBI - infer Einstein radius, ellipticity, position/offset





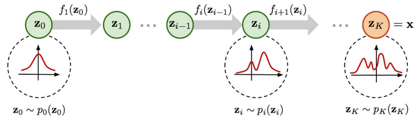


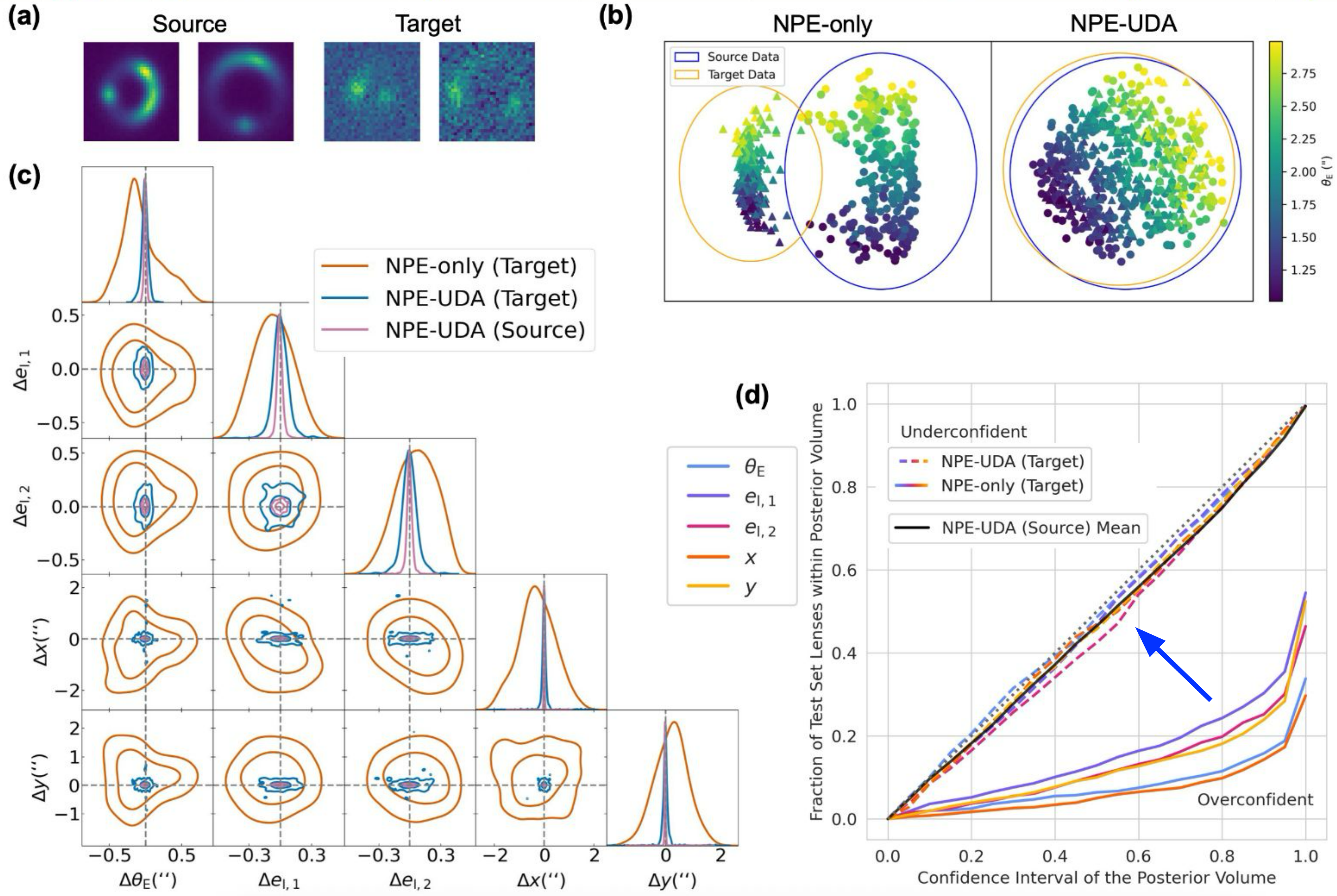
a. 100k images (1-filter) with and without noise
b. **Latent distributions are correctly aligned**.
c. Differences between true values and predicted posteriors are small. **Accuracy improves up to 2 orders of magnitude**.
d. Posterior are well calibrated and no longer overconfident.

Swierc et al. 2024.

Strong gravitational lensing

SBI - infer Einstein radius, ellipticity, position/offset





a. 100k images (1-filter) with and without noise
b. **Latent distributions are correctly aligned**.
c. Differences between true values and predicted posteriors are small. **Accuracy improves up to 2 orders of magnitude**.
d. Posterior are well calibrated and no longer overconfident.

Swierc et al. 2024.

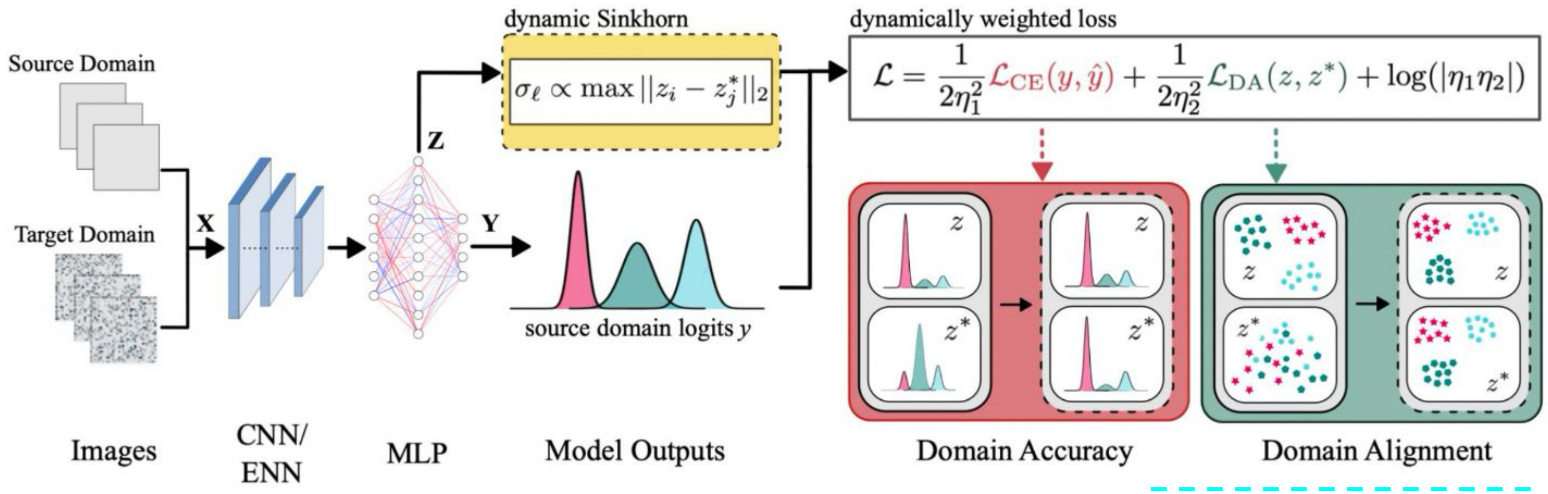
SIDDA: SInkhorn Dynamic Domain Adaptation

Can we perform automated domain alignment and avoid time-consuming hyperparameter tuning?

SIDDA: SInkhorn Dynamic Domain Adaptation

Automated domain alignment for “any” application

- **Trainable scaling** of the main task loos and DA loss.
- **Trainable Sinkhorn plan** i.e., how detailed should the distance measure be.
- **No tuning needed**!
- Up to **40% better accuracy** on unlabeled data.

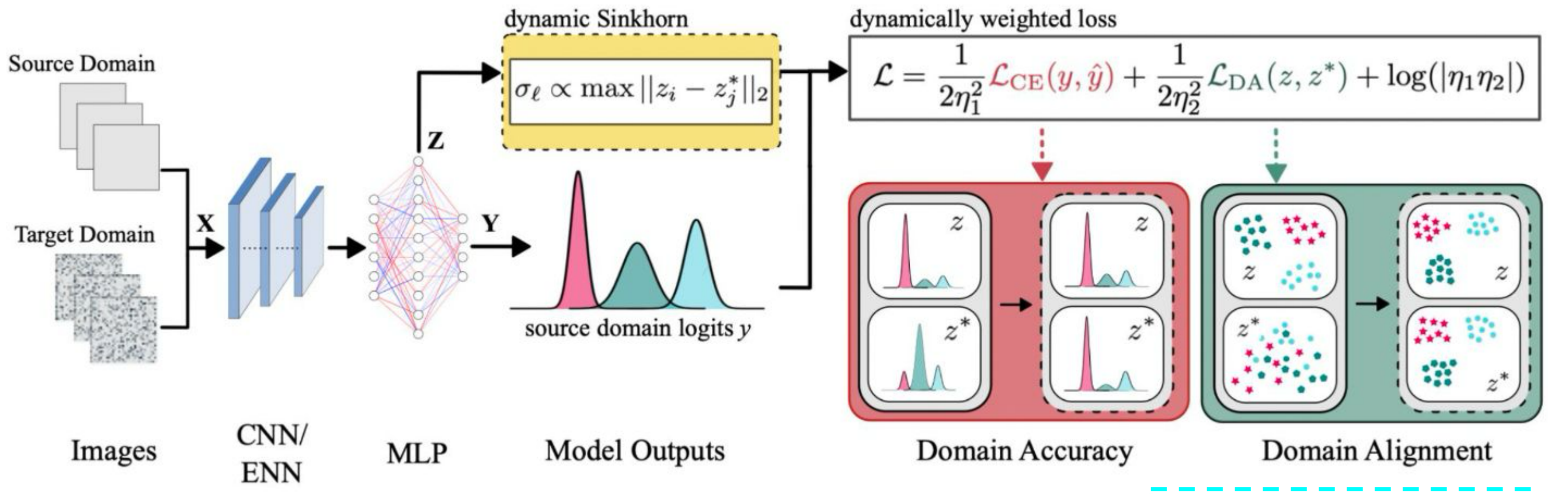


Pandya et al. 2025.

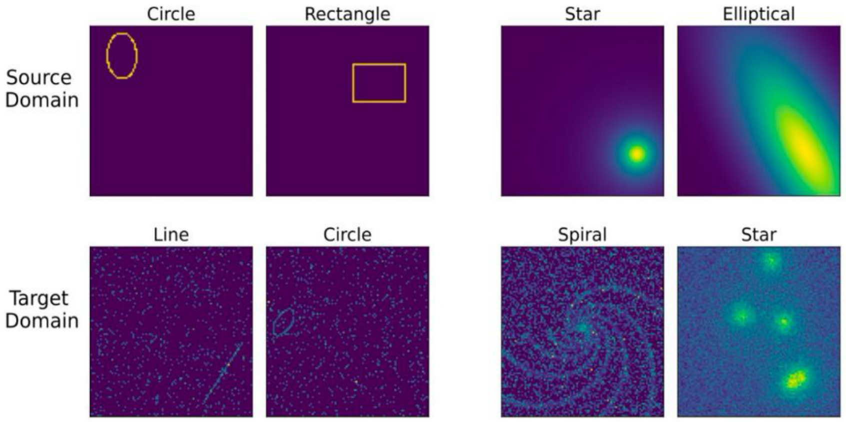
SIDDA: SInkhorn Dynamic Domain Adaptation

Automated domain alignment for “any” application

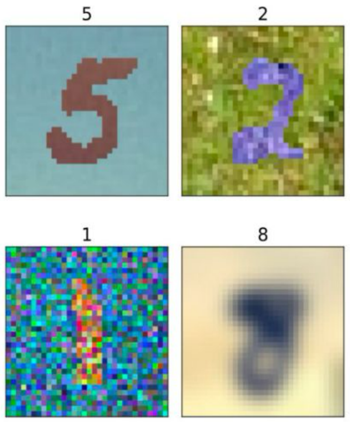
- **Trainable scaling** of the main task loos and DA loss.
- **Trainable Sinkhorn plan** i.e., how detailed should the distance measure be.
- **No tuning needed**!
- Up to **40% better accuracy** on unlabeled data.

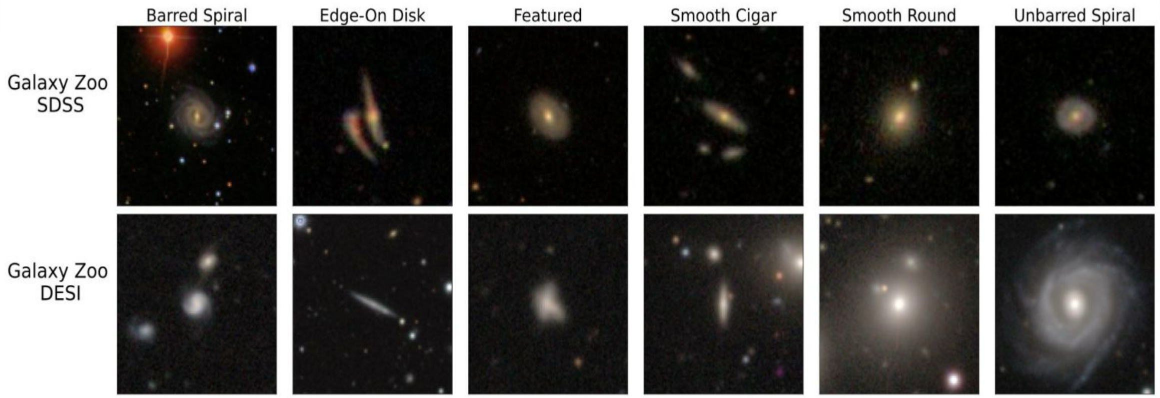


Pandya et al. 2025.



Voetberg et al. 2025



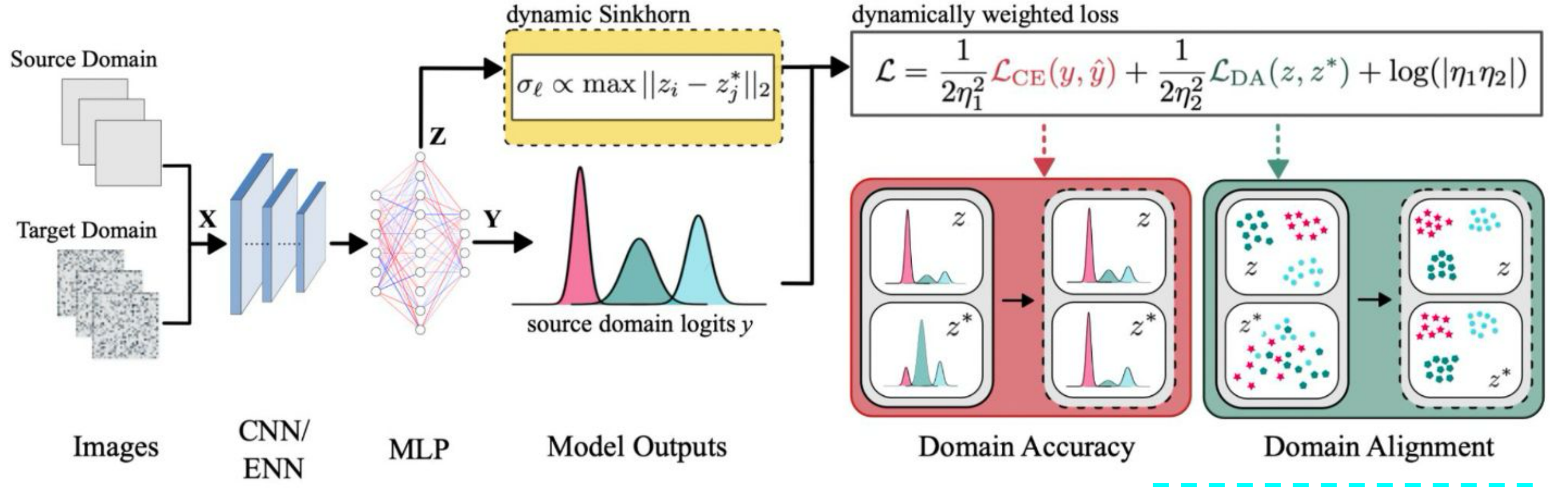


FERMI NATIONAL ACCELERATOR LABORATORY

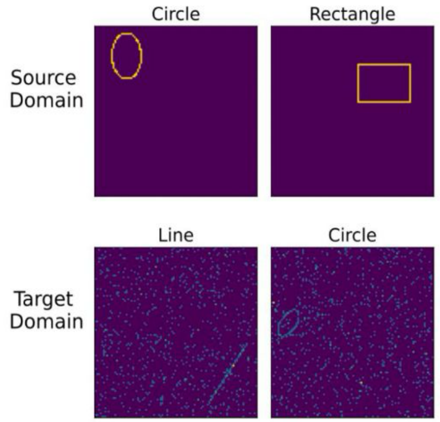
SIDDA: SInkhorn Dynamic Domain Adaptation

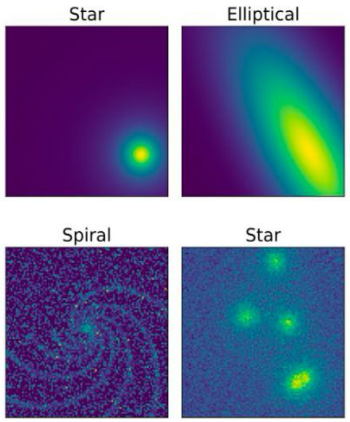
Automated domain alignment for “any” application

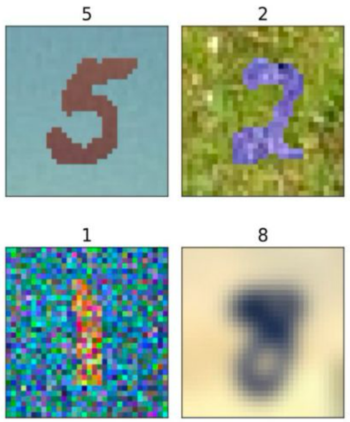
- **Trainable scaling** of the main task loos and DA loss.
- **Trainable Sinkhorn plan** i.e., how detailed should the distance measure be.
- **No tuning needed**!
- Up to **40% better accuracy** on unlabeled data.



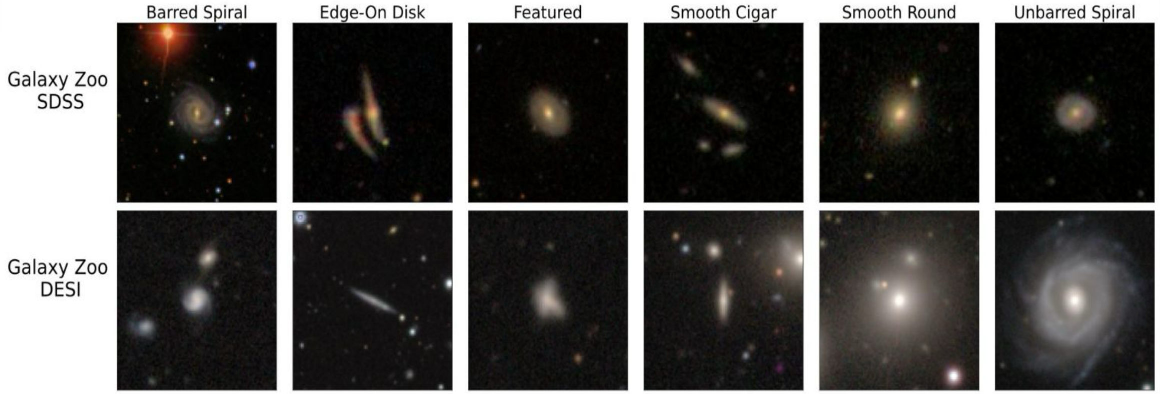
Pandya et al. 2025.







Voetberg et al. 2025



Stay tuned for the hands-on session to talk to Sneh!

Useful Resources:

- “A Kernel Two-Sample Test” Gretton et al. 2012., Video 1, Video 2
- “Domain Adversarial Training of Neural Networks” Ganin et al. 2016.
- “Deep Reconstruction-Classification Networks for Unsupervised Domain Adaptation” Ghifary et al. 2016.
- “Interpolating between Optimal Transport and MMD using Sinkhorn Divergences” Feydy et al. 2018

04

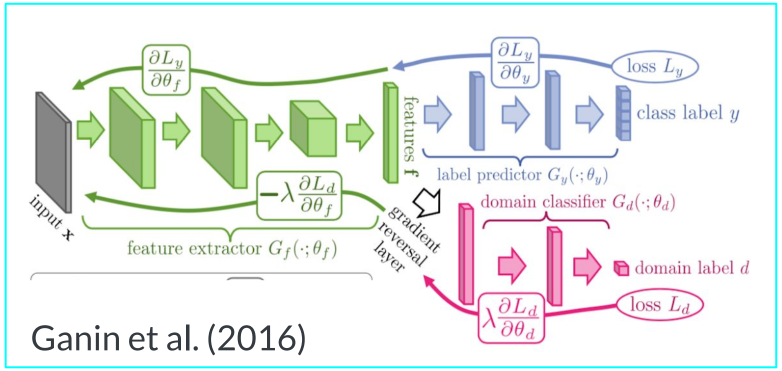
Diving Deeper Into Distance-Based Methods

Combining Datasets

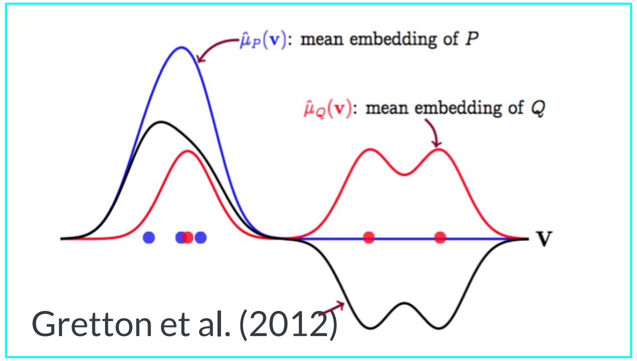
DOMAIN ADAPTATION

Align data distributions in the latent space of the network by forcing the network to **find more robust domain-invariant features.**

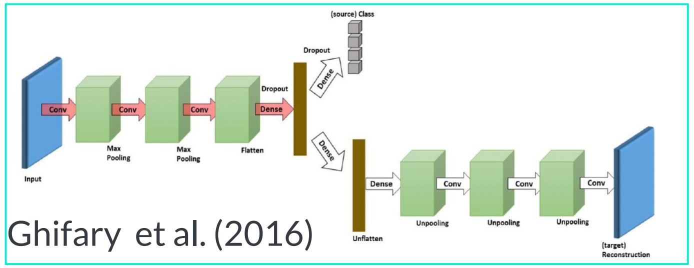
Adversarial	Distance-based methods	Reconstruction-based



Ganin et al. (2016)



Gretton et al. (2012)



Ghifary et al. (2016)

Using domain discriminators to encourage domain confusion through an adversarial objective.

Minimize a distance metric between source and target data.

Data reconstruction as an auxiliary task to ensure feature invariance.

Common types of distance measures

φ-divergences

Czisar 1963.

- Use convex function φ to measure the ratio of the two distributions.
- Likelihood based models and information theory.

$$D_\varphi(\alpha\|\beta) = \int \varphi\left(\frac{d\alpha}{d\beta}\right) d\beta$$

- Examples:

Kullback-Liebler (KL) divergence ($\varphi = x \log x$)
Jensen-Shannon (JS) distance
Total Variation (TV) distance
□^{2} distance

Kernel distances

Gretton 2006. ("king of kernels")

- Kernel-based metric between distributions.
- Measures distance in Reproducing Kernel Hilbert Space (RKHS).
- Used in generative modeling and domain adaptation.

$$\mathrm{MMD}^2(\alpha,\beta) = \mathbb{E}_{x,x'\sim\alpha}[k(x,x')] + \mathbb{E}_{y,y'\sim\beta}[k(y,y')] - 2\mathbb{E}_{x\sim\alpha,y\sim\beta}[k(x,y)]$$

Optimal transport

Monge 1781, Kantorovitch 1942

- Finding the transport plan π that minimized the total cost of moving all the mass (c - the cost of moving a unit mass) from α to □?

$$\mathrm{OT}(\alpha,\beta) = \inf_{\gamma\in\Pi(\alpha,\beta)} \int c(x,y)\, d\gamma(x,y)$$

$$c(x,y) = \|x-y\|^p$$

- Examples metrics:

Wasserstein distance ($p=2$)

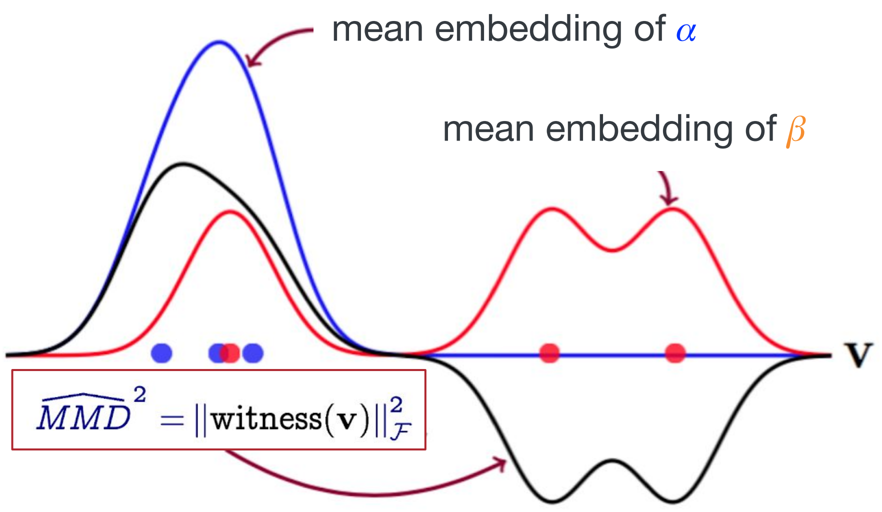
Earthmovers distance ($p=1$)

Sinkhorn divergence (regularized OT)

Maximum Mean Discrepancy (MMD)

Gaussian (RBF) kernel

$$k(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right)$$



$$\mathrm{MMD}^2(\alpha, \beta) = \mathbb{E}_{x,x' \sim \alpha}[k(x, x')] + \mathbb{E}_{y,y' \sim \beta}[k(y, y')] - 2\mathbb{E}_{x \sim \alpha, y \sim \beta}[k(x, y)]$$

Maximum Mean Discrepancy (MMD)

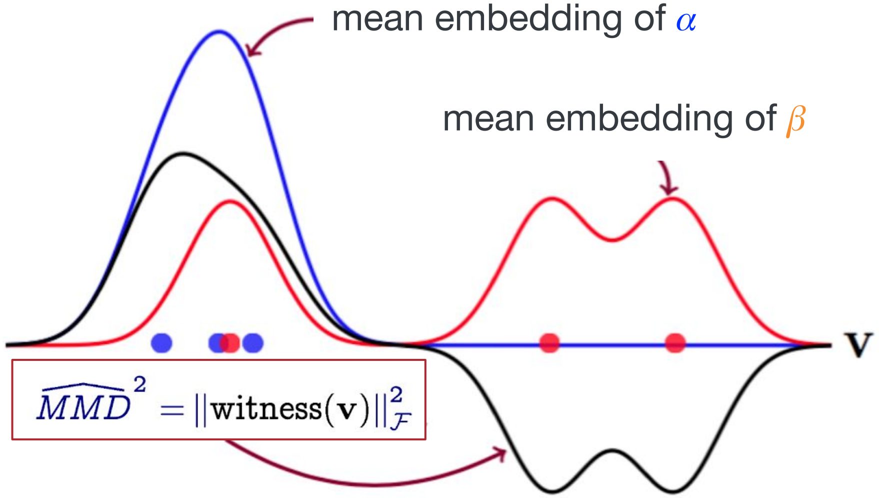
Smola et al. (2007)
Gretton et al. (2012)

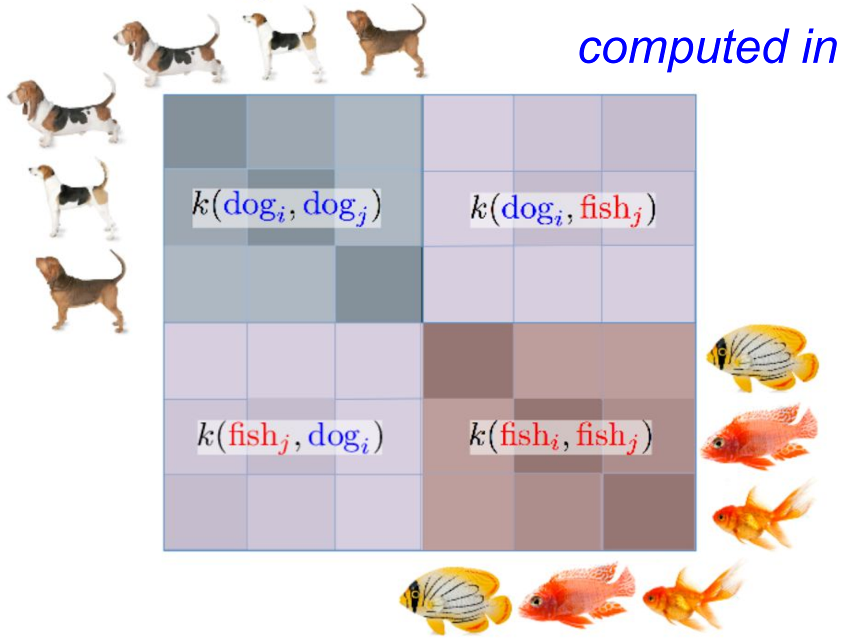
Gaussian (RBF) kernel

$$k(x,y) = \exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right)$$

$$\widehat{MMD}^2 = \frac{1}{n(n-1)}\sum_{i\neq j} k(\mathrm{dog}_i,\mathrm{dog}_j) + \frac{1}{n(n-1)}\sum_{i\neq j} k(\mathrm{fish}_i,\mathrm{fish}_j) - \frac{2}{n^2}\sum_{i,j} k(\mathrm{dog}_i,\mathrm{fish}_j)$$

computed in $O(n^2)$





$$\mathrm{MMD}^2(\alpha,\beta) = \mathbb{E}_{x,x'\sim\alpha}[k(x,x')] + \mathbb{E}_{y,y'\sim\beta}[k(y,y')] - 2\mathbb{E}_{x\sim\alpha,y\sim\beta}[k(x,y)]$$

FERMI NATIONAL ACCELERATOR LABORATORY

Maximum Mean Discrepancy (MMD)

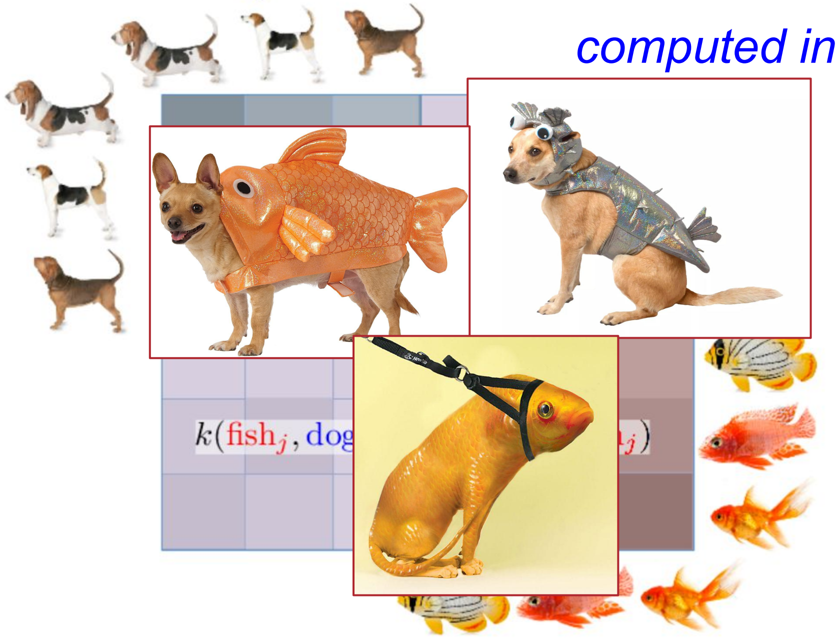
Smola et al. (2007)
Gretton et al. (2012)

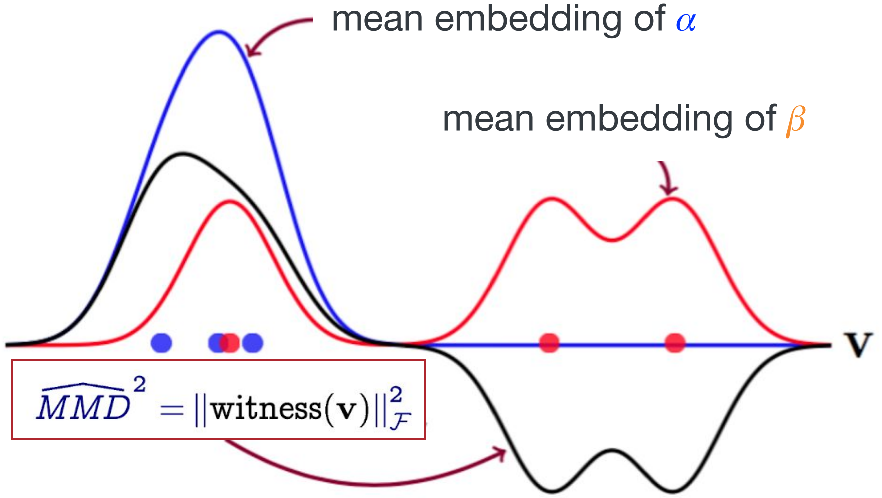
Gaussian (RBF) kernel

$$k(x,y) = \exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right)$$

$$\widehat{MMD}^2 = \frac{1}{n(n-1)}\sum_{i\neq j} k(\mathrm{dog}_i, \mathrm{dog}_j) + \frac{1}{n(n-1)}\sum_{i\neq j} k(\mathrm{fish}_i, \mathrm{fish}_j) - \frac{2}{n^2}\sum_{i,j} k(\mathrm{dog}_i, \mathrm{fish}_j)$$

computed in $O(n^2)$

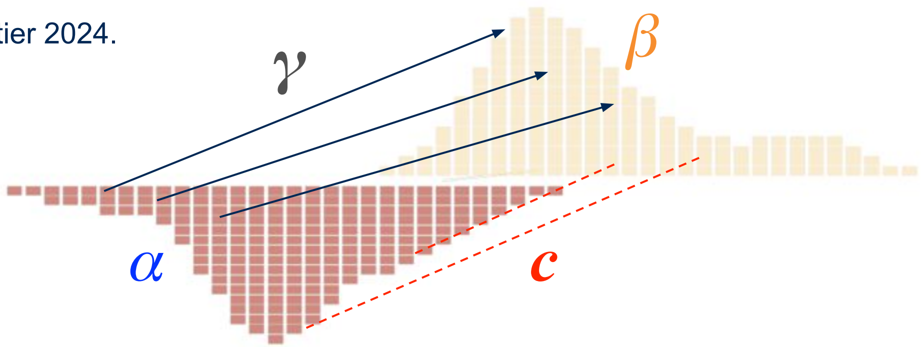




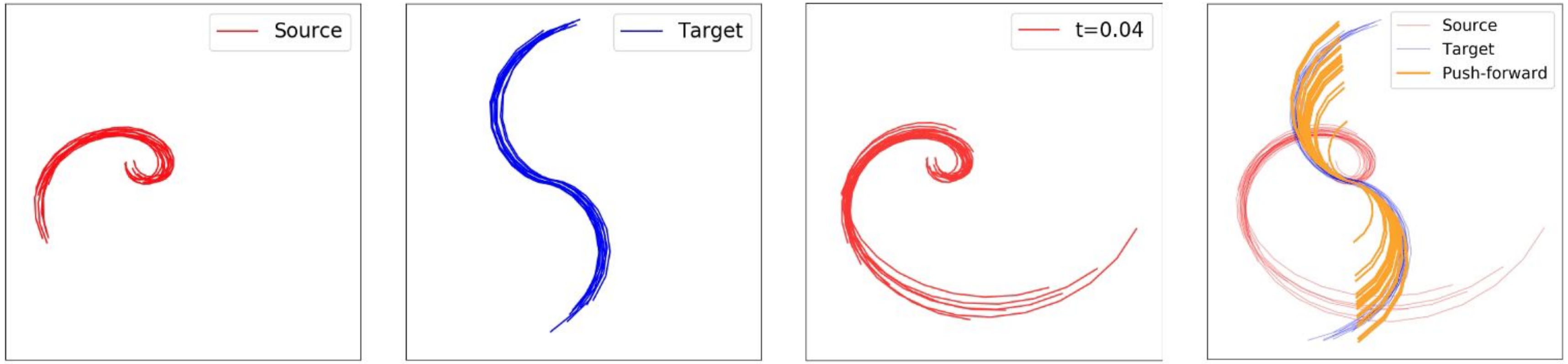
$$\mathrm{MMD}^2(\alpha,\beta) = \mathbb{E}_{x,x'\sim\alpha}[k(x,x')] + \mathbb{E}_{y,y'\sim\beta}[k(y,y')] - 2\mathbb{E}_{x\sim\alpha,y\sim\beta}[k(x,y)]$$

Optimal Transport

Source: Charpentier 2024.



OPTIMAL TRANSPORT:
Finding the "most efficient" transportation plan γ that minimizes the total transportation cost c of moving the entire mass from probability α distribution to β.



Source data: Swiss-roll curves

Target data: Wave curves

The **geodesic** of the **Mapping**

The resulting push-forward

Source: Zhu et al. 2023.

Optimal Transport

Finding the "most efficient" transportation plan γ that minimizes the total transportation cost c of moving the entire mass from probability α distribution to β.

Optimal Transport

$$\mathrm{OT}(\alpha,\beta) = \inf_{\gamma\in\Pi(\alpha,\beta)} \int c(x,y)\, d\gamma(x,y)$$

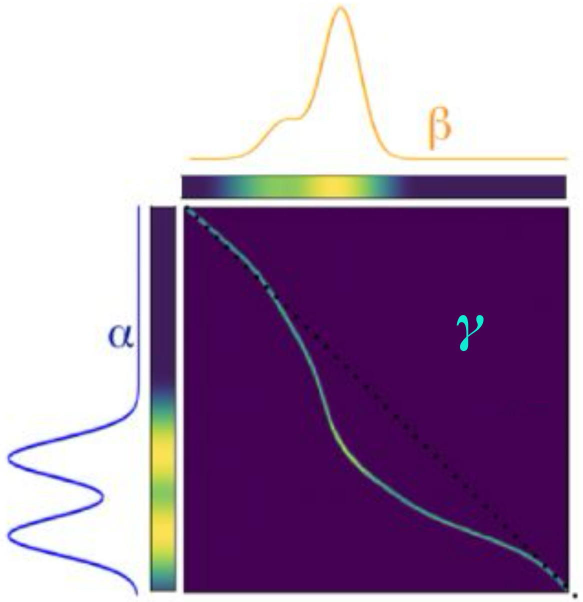
Solving OT exactly becomes impossible for large datasets as it scales as O(n^3logn)!

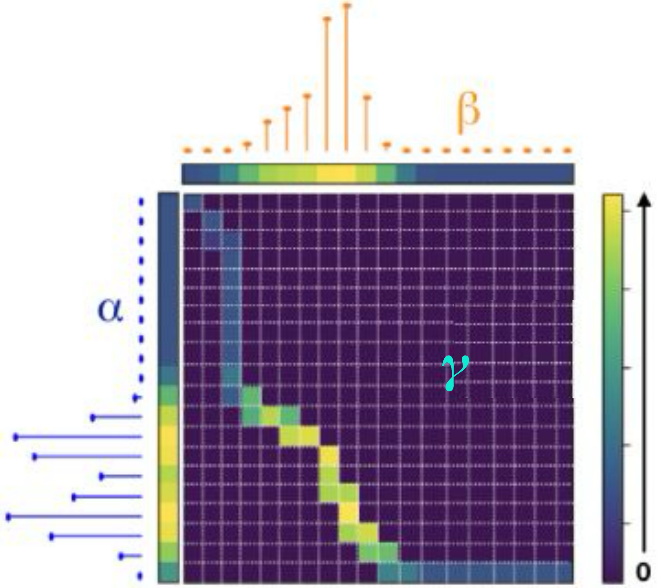
Wasserstein distance
p=2

$$W_p(\alpha,\beta) = \left(\inf_{\gamma\in\Pi(\alpha,\beta)} \int \boxed{\|x-y\|^p}\, d\gamma(x,y)\right)^{1/p}$$

$\boxed{c(x,y)}$ cost of moving a unit of mass from x to y

$\boxed{\gamma(x,y)}$ transport plan (coupling) - how much mass moves from x to y





Genevay 2019.

Optimal Transport

Finding the “most efficient” transportation plan γ that minimizes the total transportation cost c of moving the entire mass from probability α distribution to β.

Optimal Transport

$$\mathrm{OT}(\alpha, \beta) = \inf_{\gamma \in \Pi(\alpha,\beta)} \int c(x,y)\, d\gamma(x,y)$$

Solving OT exactly becomes impossible for large datasets as it scales as $O(n^3 \log n)$!

Wasserstein distance p=2

$$W_p(\alpha, \beta) = \left(\inf_{\gamma \in \Pi(\alpha,\beta)} \int \boxed{\|x-y\|^p}\, d\gamma(x,y) \right)^{1/p}$$

Entropy Regularization

$$OT_{\varepsilon,p}(\alpha, \beta) = \inf_{\gamma \in \Pi(\alpha,\beta)} \int_{\mathcal{X}\times\mathcal{Y}} \|x-y\|^p\, d\gamma(x,y) \boxed{- \varepsilon \int_{\mathcal{X}\times\mathcal{Y}} \log\left(\frac{d\gamma}{dxdy}\right) d\gamma(x,y)}$$

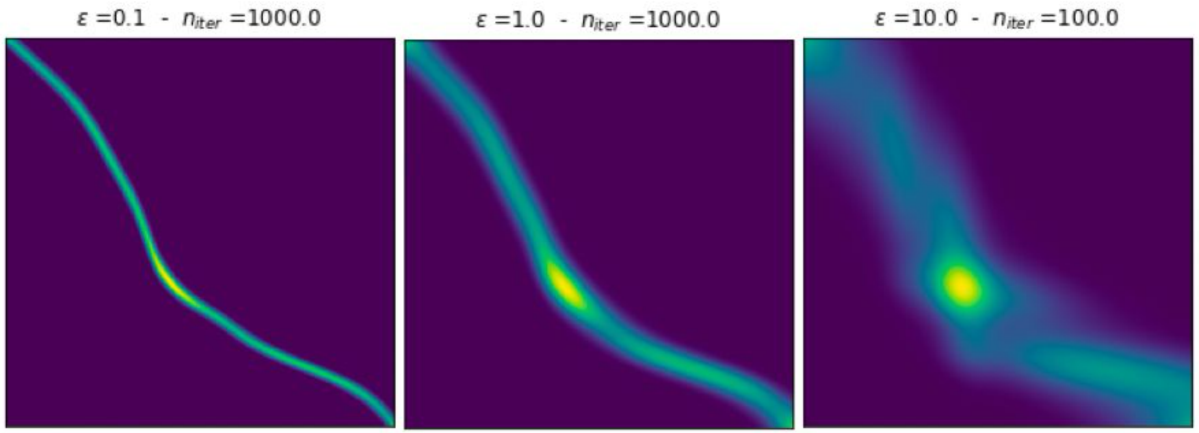
The **entropy term** smooths the optimization and ensures faster convergence with better numerical properties. ε controls the **strength of the entropy regularization**.

FERMI NATIONAL ACCELERATOR LABORATORY

Optimal Transport

Finding the “most efficient” transportation plan γ that minimizes the total transportation cost c of moving the entire mass from probability α distribution to β.

ε smooths the problem to avoid overfitting.



Genevay 2019.

Entropy Regularization

$$OT_{\varepsilon,p}(\alpha,\beta) = \inf_{\gamma\in\Pi(\alpha,\beta)} \int_{\mathcal{X}\times\mathcal{Y}} \|x-y\|^p \, d\gamma(x,y) \boxed{- \varepsilon \int_{\mathcal{X}\times\mathcal{Y}} \log\left(\frac{d\gamma}{dxdy}\right) d\gamma(x,y)}$$

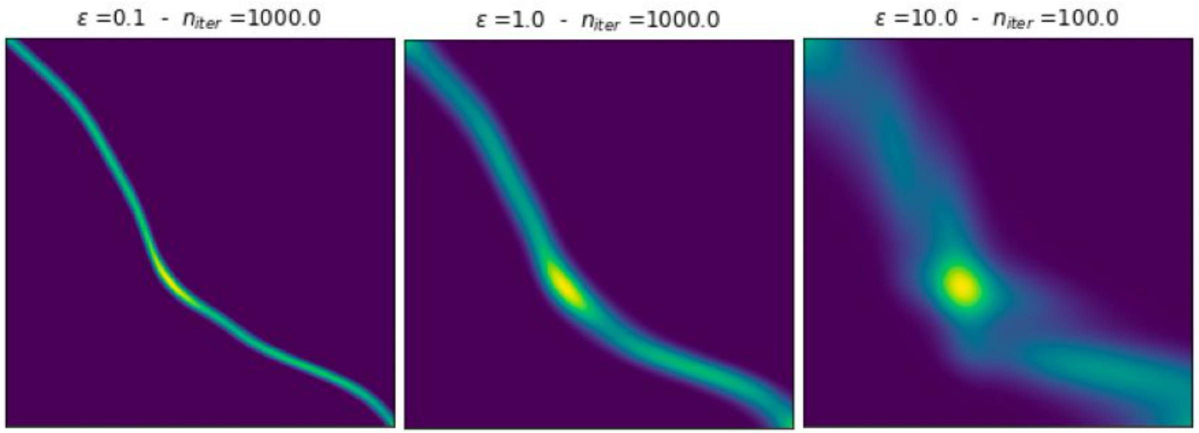
The **entropy term** smooths the optimization and ensures faster convergence with better numerical properties. ε controls the **strength of the entropy regularization**.

FERMI NATIONAL ACCELERATOR LABORATORY

Optimal Transport

Finding the "most efficient" transportation plan γ that minimizes the total transportation cost c of moving the entire mass from probability α distribution to β.

ε smooths the problem to avoid overfitting.
But **Entropy Regularization introduces a bias.** $OT_{\varepsilon,p}(\alpha,\alpha) \neq 0$



Genevay 2019.

Entropy Regularization

$$OT_{\varepsilon,p}(\alpha,\beta) = \inf_{\gamma\in\Pi(\alpha,\beta)} \int_{\mathcal{X}\times\mathcal{Y}} \|x-y\|^p \, d\gamma(x,y) \boxed{- \varepsilon \int_{\mathcal{X}\times\mathcal{Y}} \log\left(\frac{d\gamma}{dx dy}\right) d\gamma(x,y)}$$

The **entropy term** smooths the optimization and ensures faster convergence with better numerical properties. ε controls the **strength of the entropy regularization**.

Sinkhorn Algorithm
Debiased and efficient OT

$$\text{Sink}_\varepsilon(\alpha, \beta) = OT_\varepsilon(\alpha, \beta) - \tfrac{1}{2}\left[OT_\varepsilon(\alpha, \alpha) + OT_\varepsilon(\beta, \beta)\right]$$

Entropic OT between source and target distributions.

Self-distances of source and target (should be zero in classical OT, but not in entropic OT).

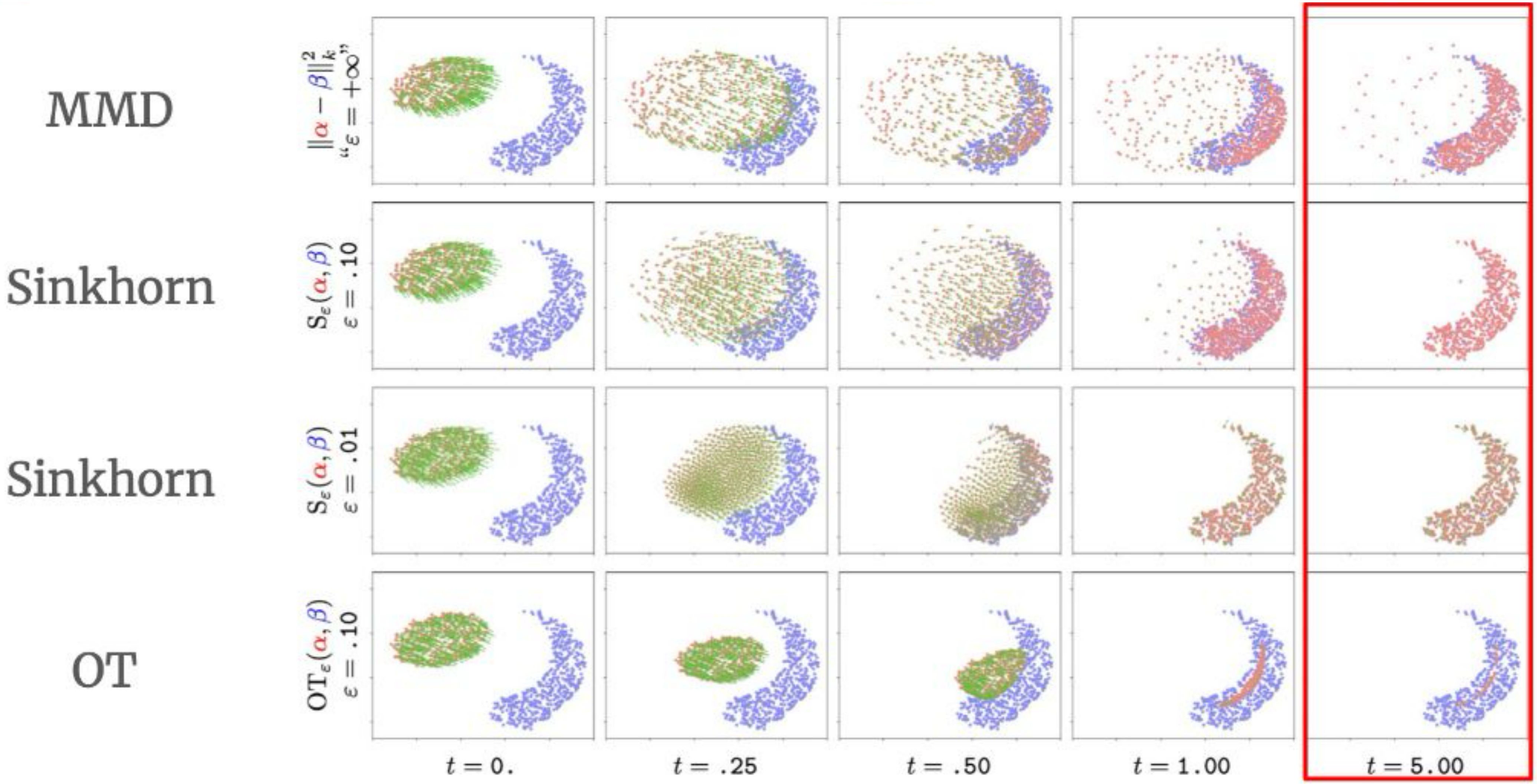
Asymptotic behaviours:

$$\text{Sink}_\varepsilon(\alpha, \beta) \rightarrow \begin{cases} \text{Wasserstein} & \text{as } \varepsilon \rightarrow 0 \\ \text{MMD-like mean-matching} & \text{as } \varepsilon \rightarrow \infty \end{cases}$$

Feydy et al. 2018.

Sinkhorn - best of both worlds

Interpolating between MMD and Optimal Transport (OT)



Feydy et al. 2018.

MMD and Optimal Transport (OT)

MMD PROS:

- **Simple and fast** to compute - $O(n^2)$.
- **Differentiable**
- Works well in **high dimensions** (if kernel is well-chosen).
- Captures subtle distributional differences.

CONS**:**

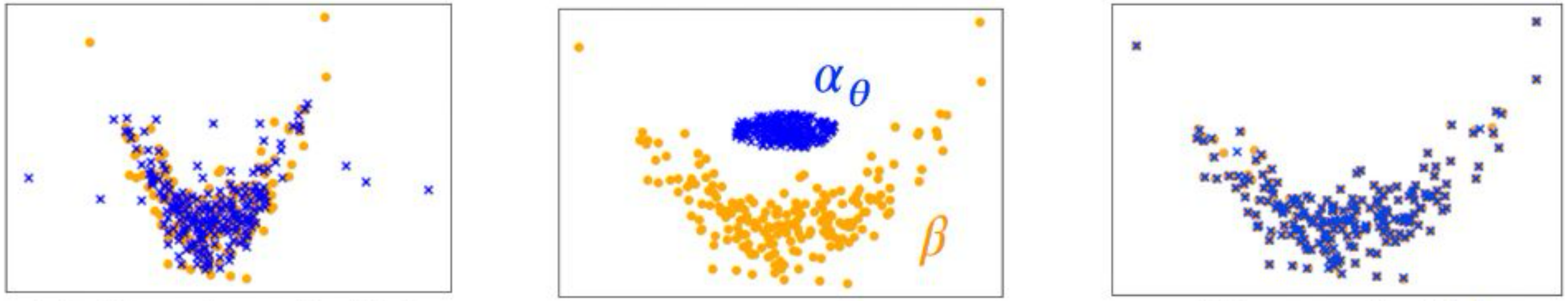
- **Sensitive to kernel choice**.
- Bad in low-density environments.
- Fails to capture **geometric structure**
- Less interpretable.

OT PROS:

- Captures **geometry of distributions**.
- More **interpretable**: measures "how much mass needs to move".
- Good for structured data (images, point clouds, time series).

CONS:

- **Computationally expensive** $O(n^3 \log(n))$.
- Curse of dimensionality.
- **Non-differentiable** (needs entropic regularization, but this requires tuning). **Sinkhorn to the rescue!**



MMD_k - $k = -\|\cdot\|_2^{1.5}$ | Initial Setting | W_c - $c = \|\cdot\|_2^{1.5}$

MMD and Optimal Transport (OT)

MMD PROS:

- **Simple and fast** to compute - O(n^2).
- **Differentiable**
- Works well in **high dimensions** (if kernel is well-chosen).
- Captures subtle distributional differences.

CONS:

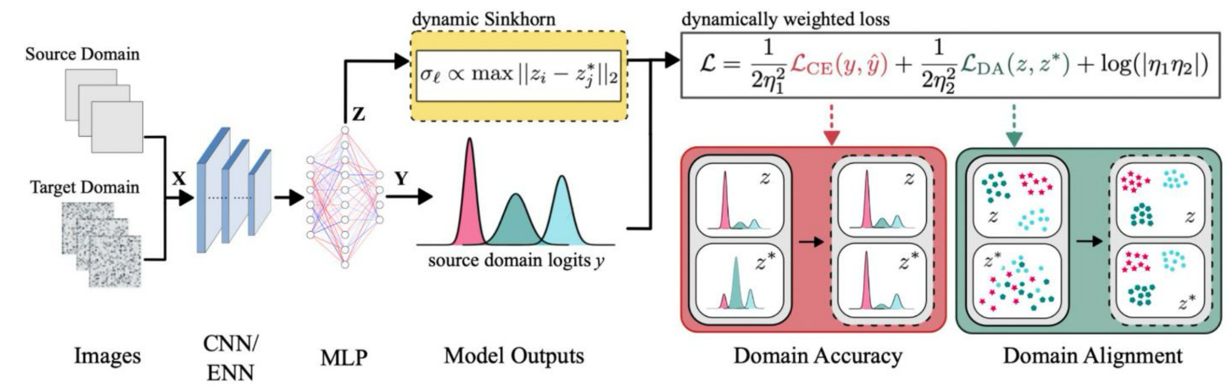
- **Sensitive to kernel choice**.
- Bad in low-density environments.
- Fails to capture **geometric structure**
- Less interpretable.

OT PROS:

- Captures **geometry of distributions**.
- More **interpretable**: measures "how much mass needs to move".
- Good for structured data (images, point clouds, time series).

CONS:

- **Computationally expensive** *O(n^3 log(n))*.
- Curse of dimensionality.
- **Non-differentiable** (needs entropic regularization, but this requires tuning). **Sinkhorn to the rescue!**



Useful Resources:

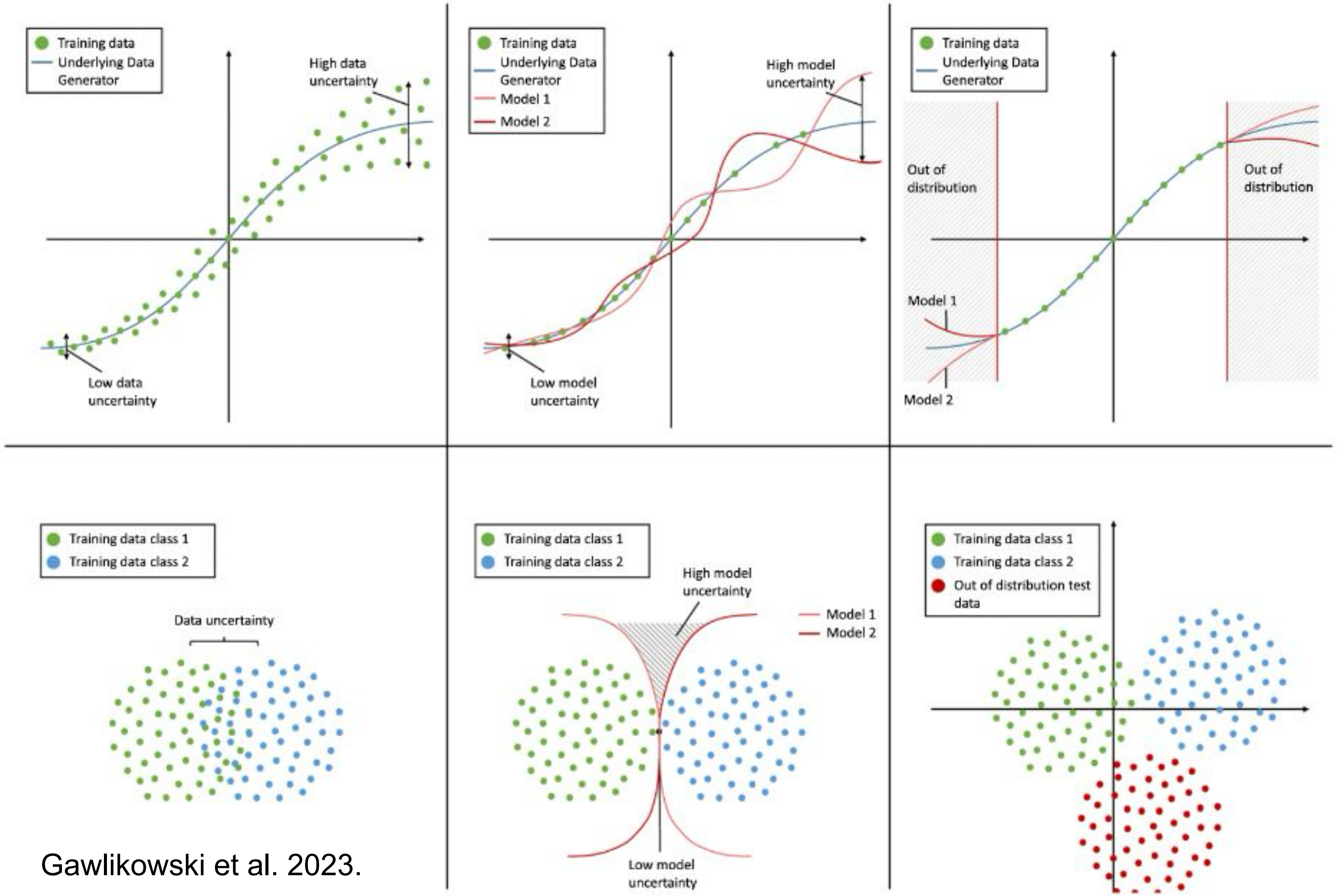
- [“A survey of uncertainty in deep neural networks” Gawlikowski et al. 2023.]
- [“On the Need to Align Intent and Implementation in Uncertainty Quantification for Machine Learning” Trivedi & Nord 2025.]

05

Uncertainty Quantification

Sources of uncertainties

Data | Model | Distributional



Gawlikowski et al. 2023.

Sources of uncertainties

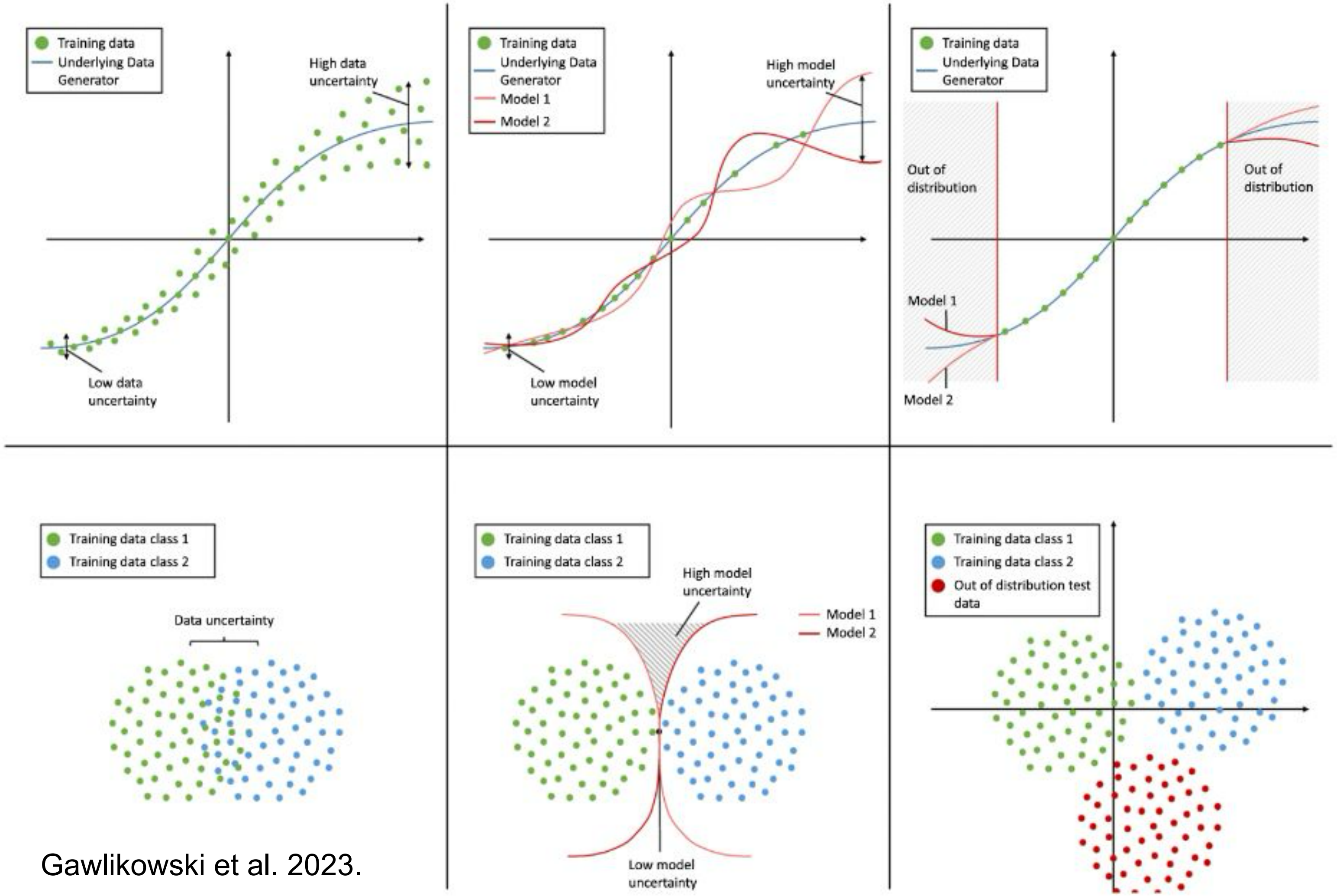
All of this will influence model parameters and model outputs!

Systematic
persistent bias
from instruments, calibration, or imperfect theory
Statistical
variation due to limited data

Epistemic
model or data-driven uncertainty
Aleatoric
irreducible noise in the data

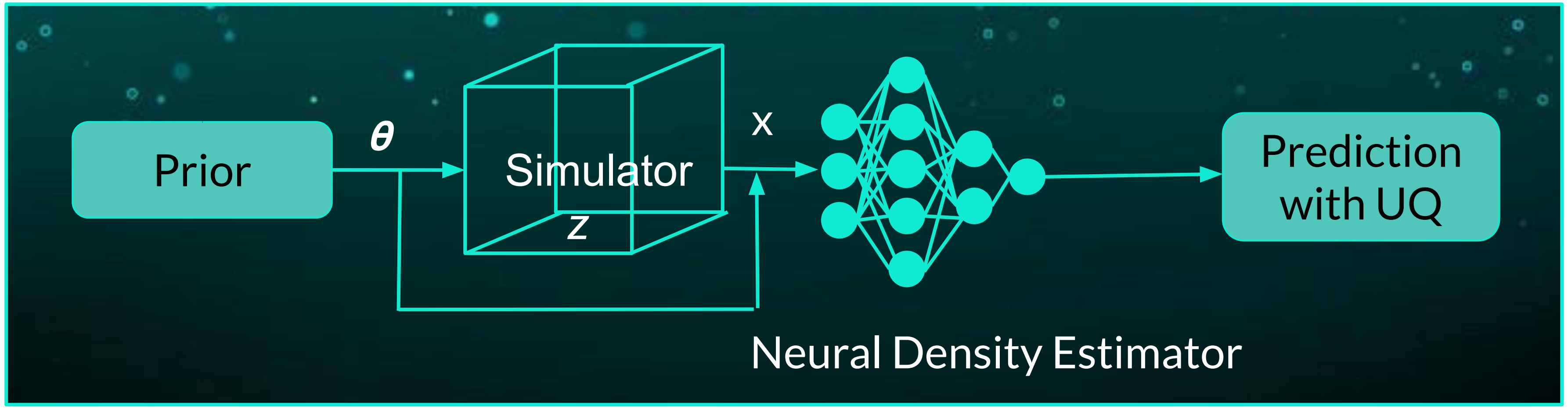
These vocabularies are not interchangeable!

Data **Model** **Distributional**



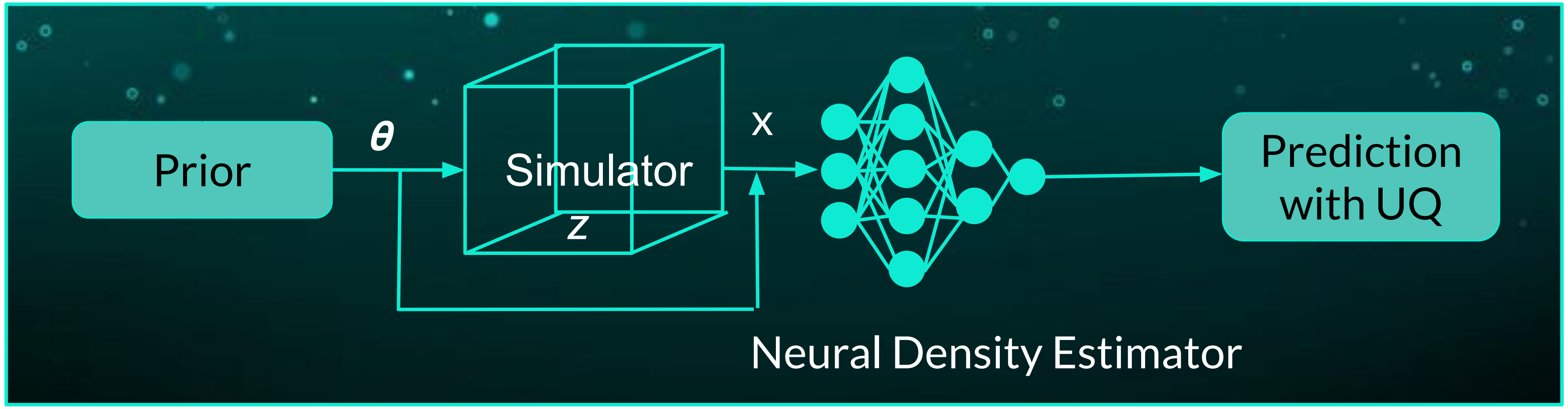
Gawlikowski et al. 2023.

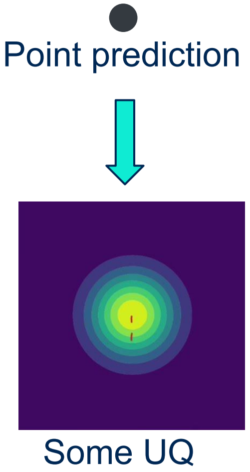
Inference with UQ

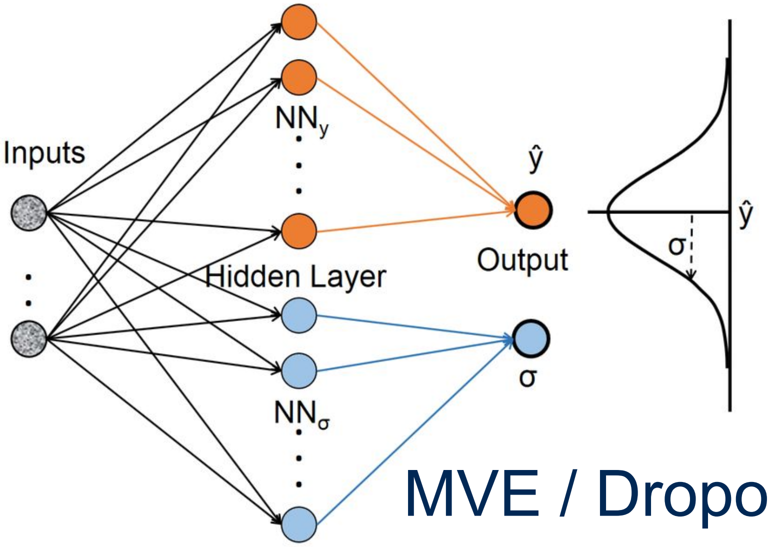


Point prediction

Inference with UQ

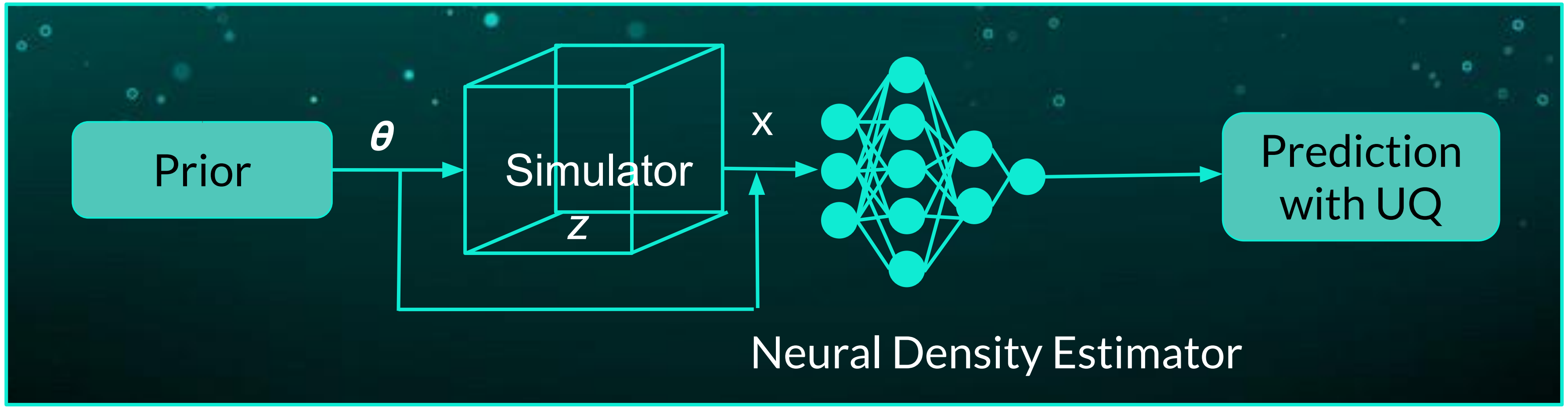


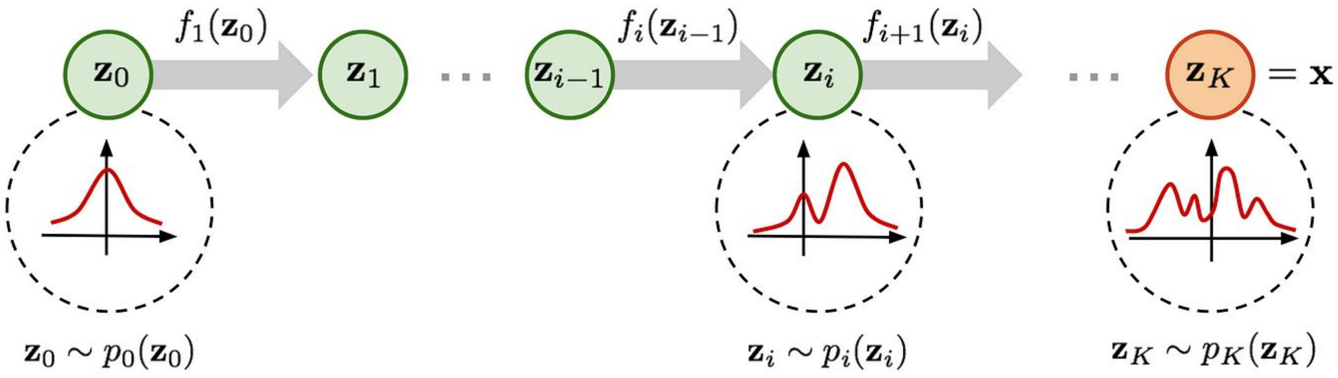




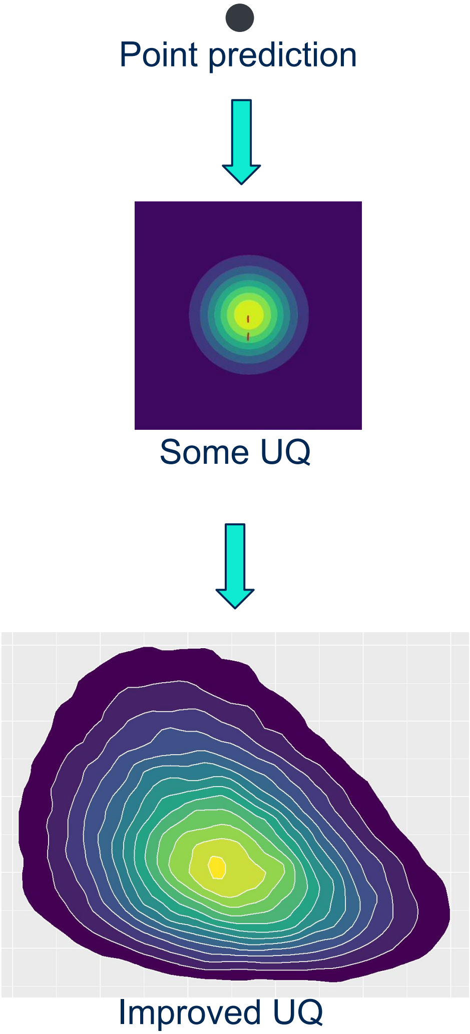
MVE / Dropout / BNN

Inference with UQ

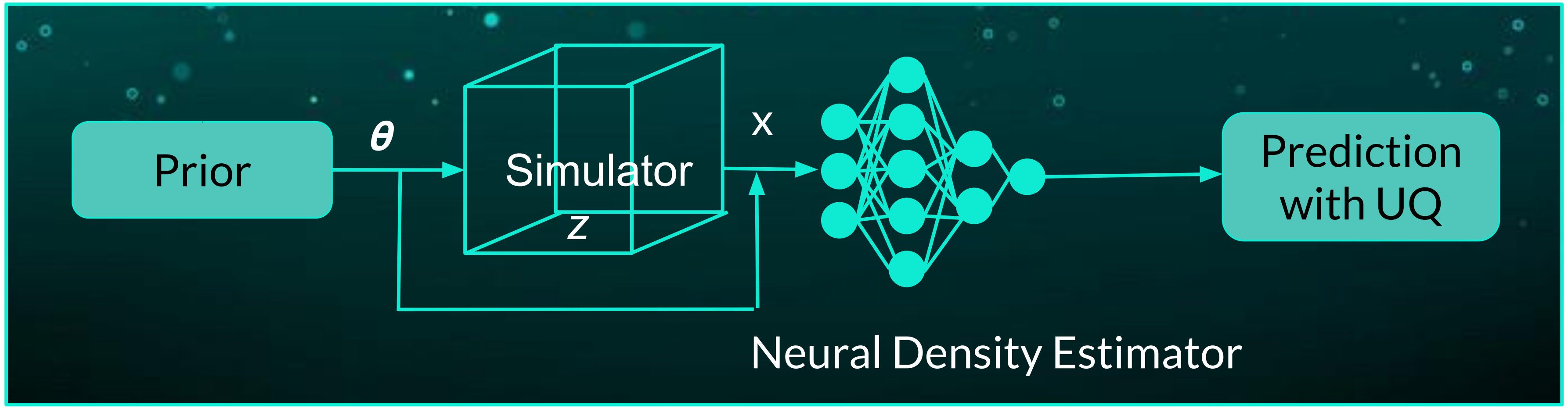


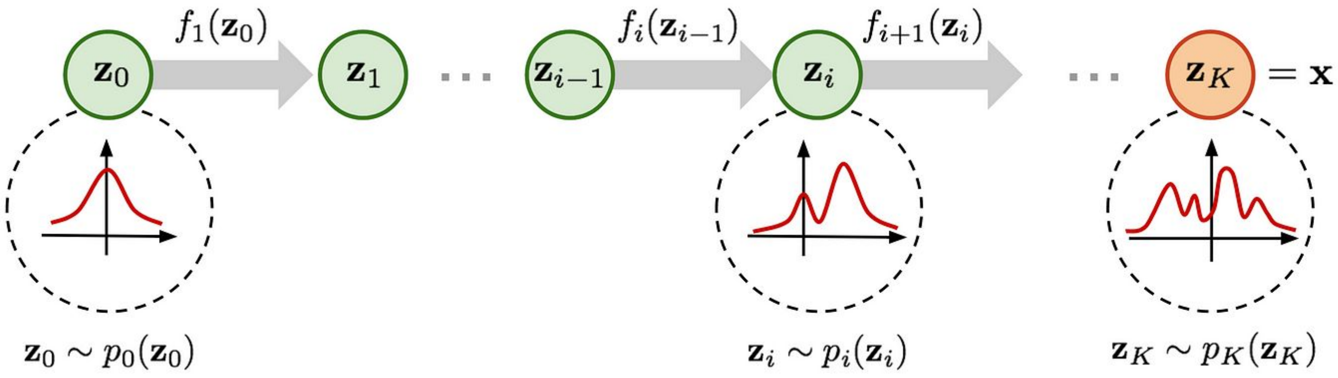


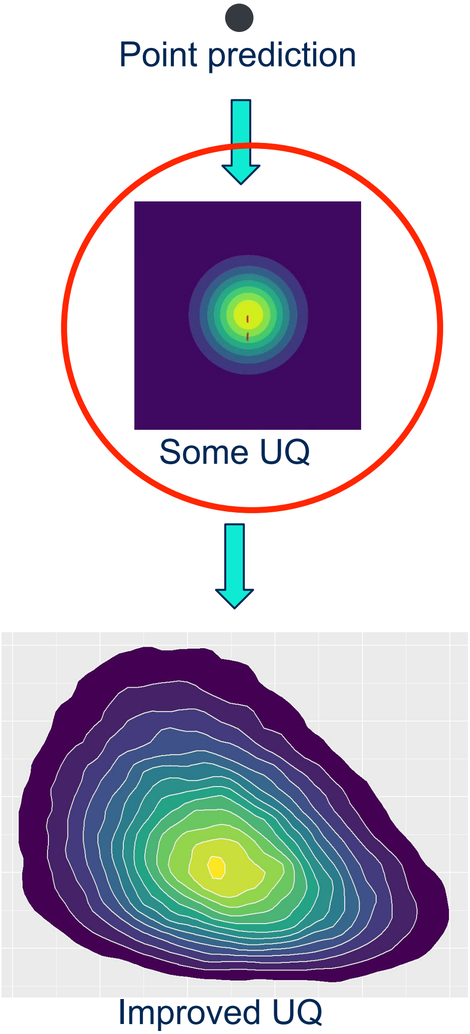
Estimate the full posterior (Normalizing Flows)



Inference with UQ





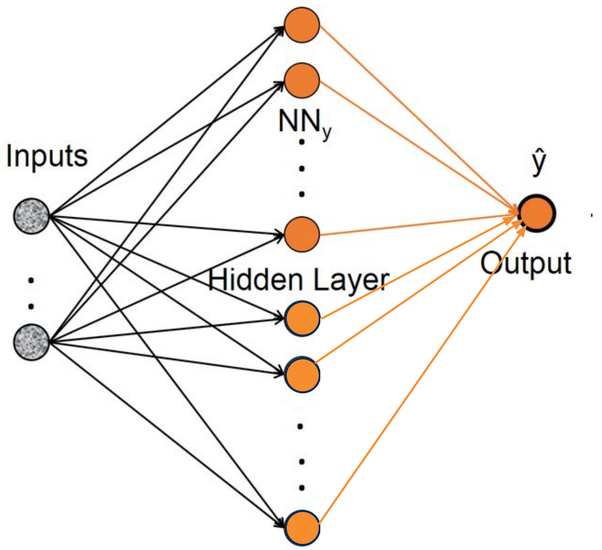


Estimate the full posterior (Normalizing Flows)

UQ for different types of models

Epistemic, aleatoric or both?

Deterministic Model

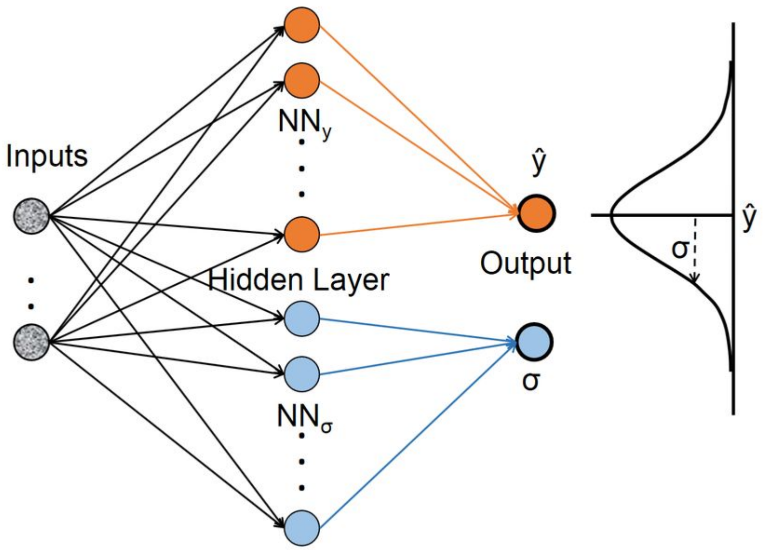


No Errors

UQ for different types of models

Epistemic, aleatoric or both?

Deterministic Model - MVE

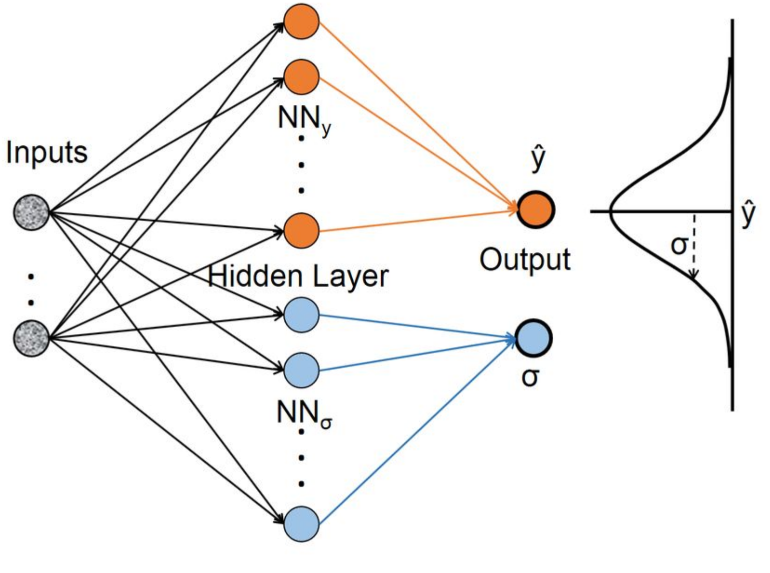


Data (Aleatoric)

UQ for different types of models

Epistemic, aleatoric or both?

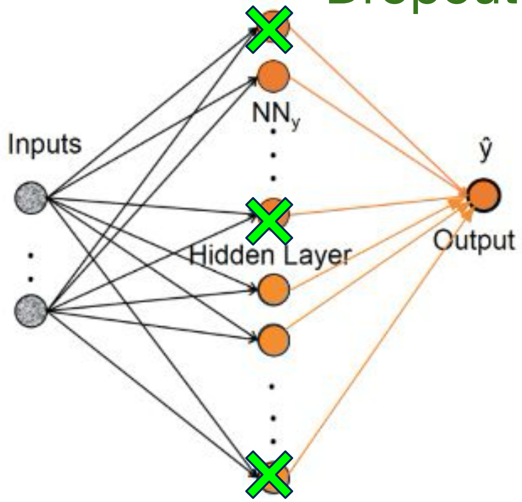
Deterministic Model - MVE



Data (Aleatoric)

Ensembles

Dropout

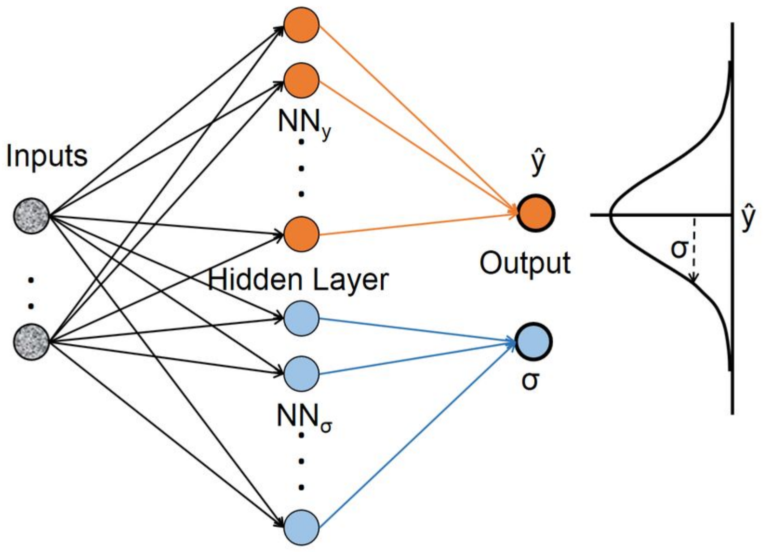


Mean, Var ŷ

UQ for different types of models

Epistemic, aleatoric or both?

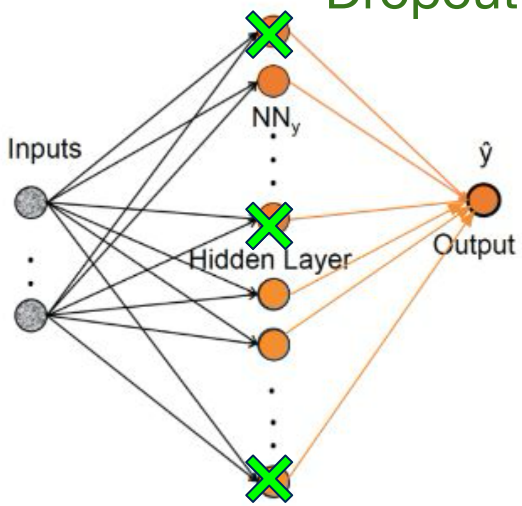
Deterministic Model - MVE



Data (Aleatoric)

Ensembles

Dropout



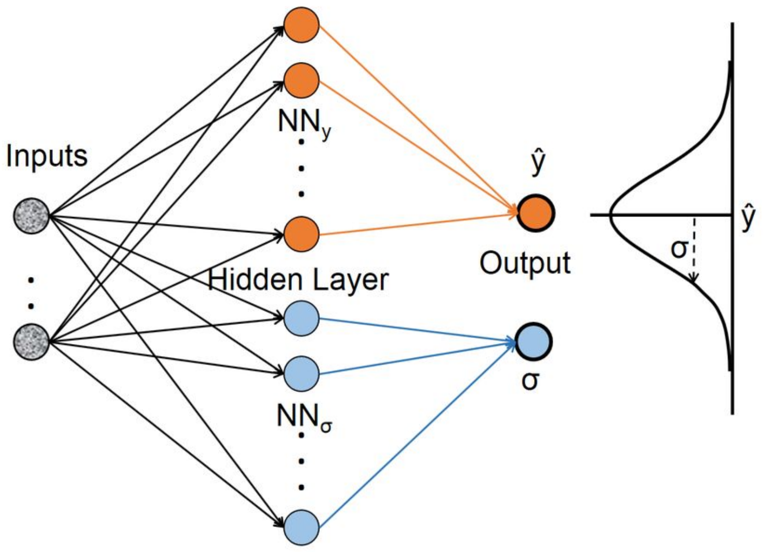
Mean, Var ŷ

Model (Epistemic)

UQ for different types of models

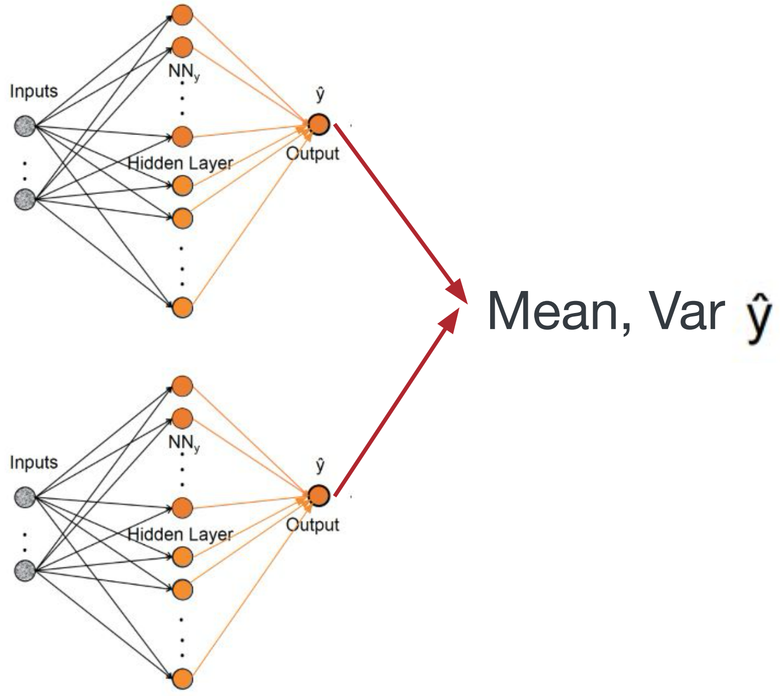
Epistemic, aleatoric or both?

Deterministic Model - MVE



Data (Aleatoric)

Ensembles

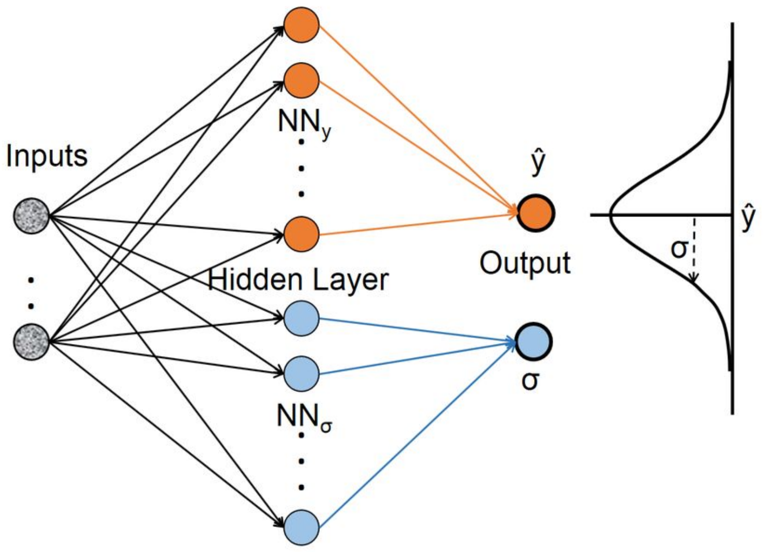


Model (Epistemic)

UQ for different types of models

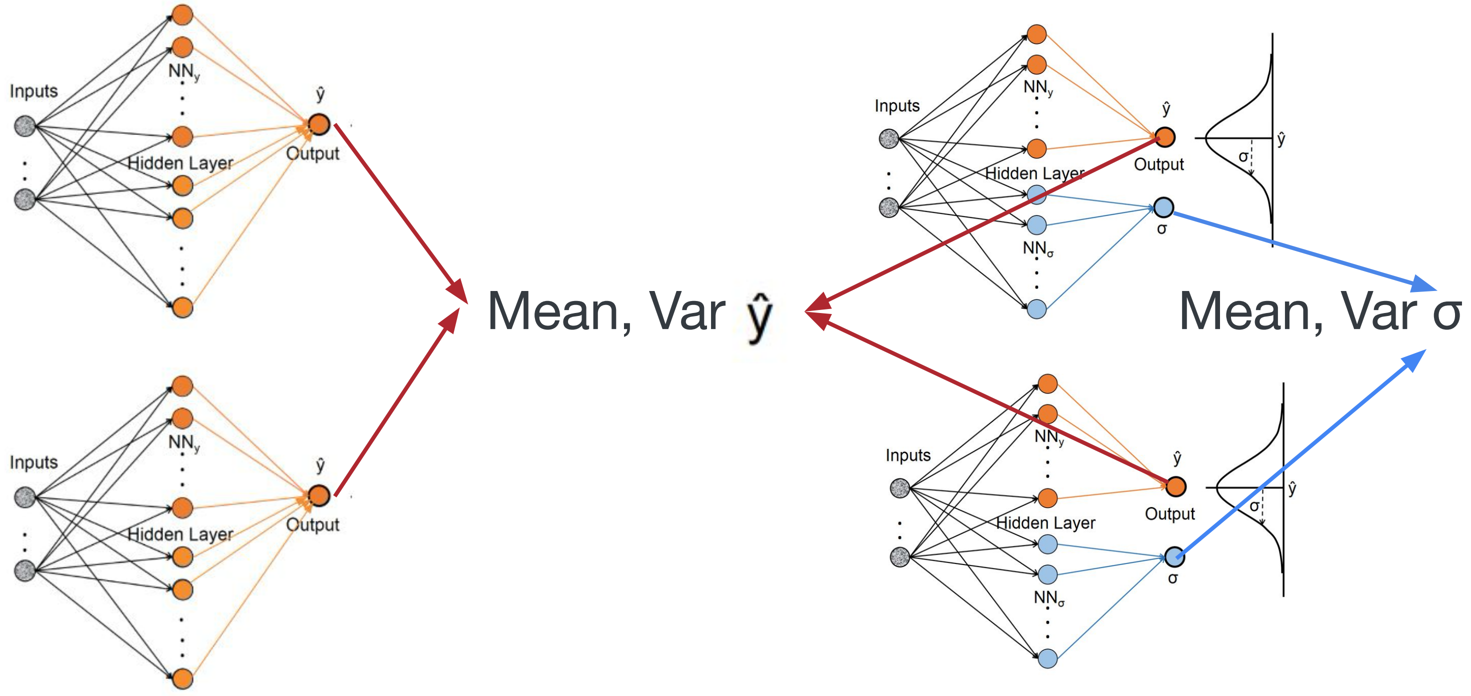
Epistemic, aleatoric or both?

Deterministic Model - MVE



Data (Aleatoric)

Ensembles



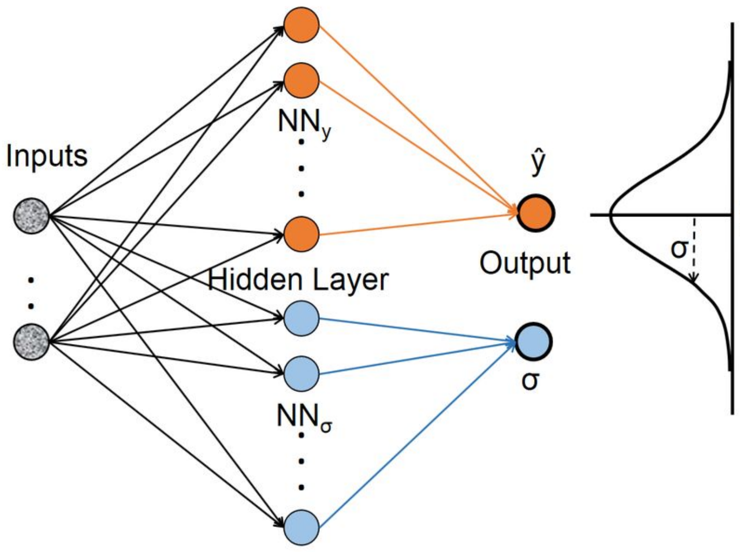
Model (Epistemic)

Data (Aleatoric)

UQ for different types of models

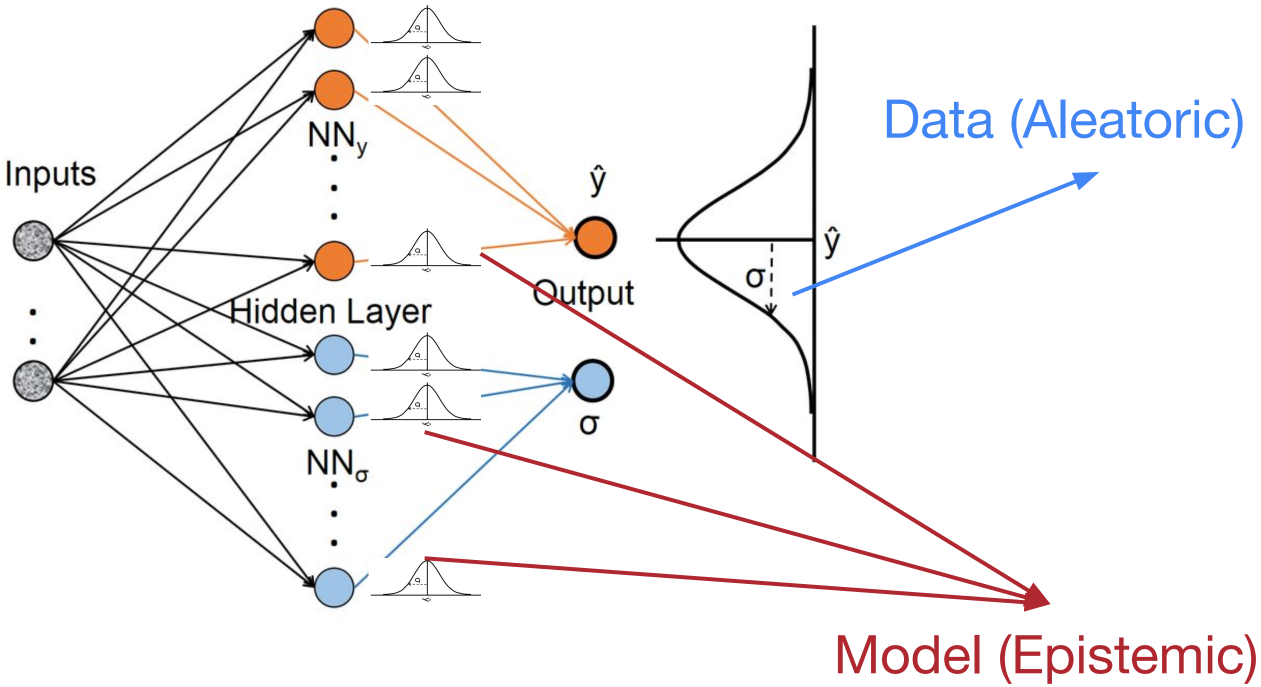
Epistemic, aleatoric or both?

Deterministic Model - MVE



Data (Aleatoric)

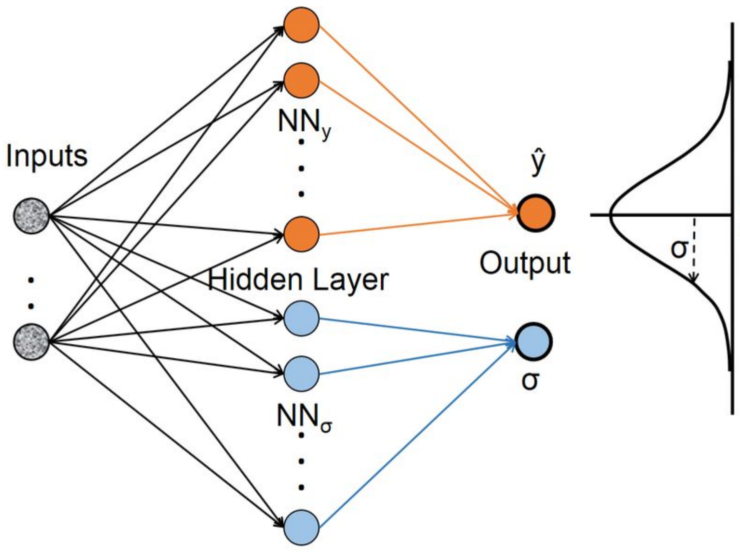
Bayesian Neural Networks



UQ for different types of models

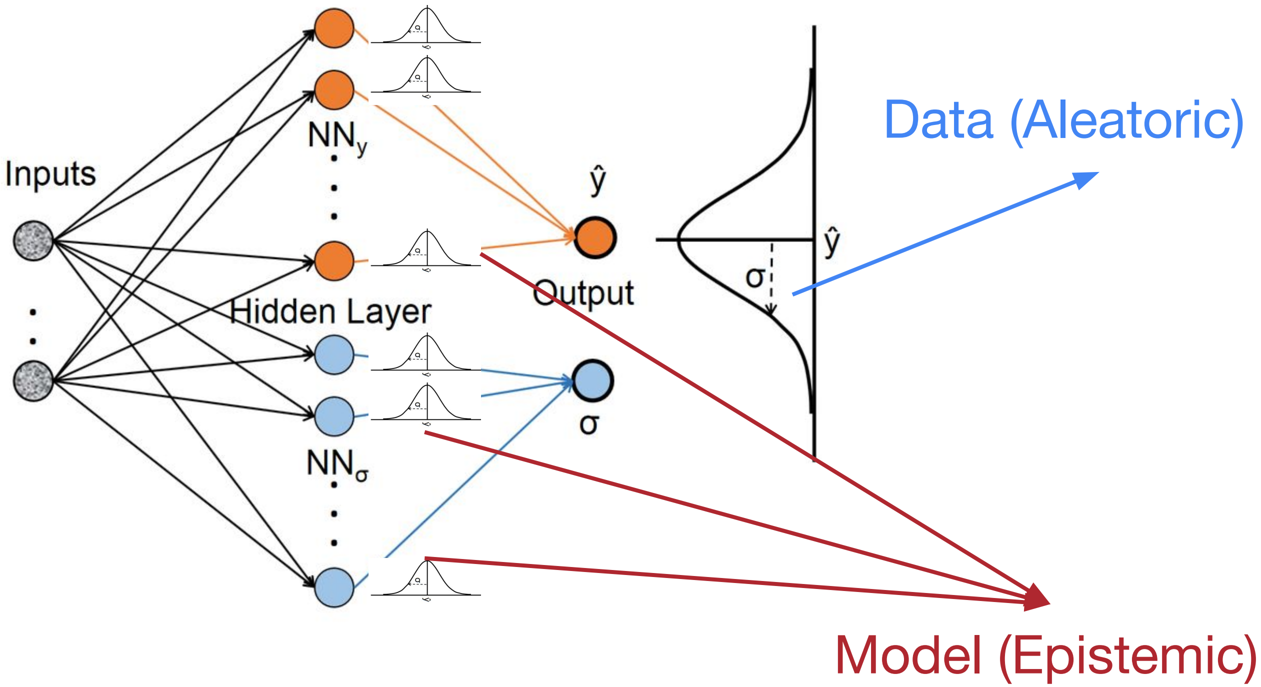
Epistemic, aleatoric or both?

Deterministic Model - MVE



Data (Aleatoric)

Bayesian Neural Networks



When domain shift is added to this picture interpretation of errors becomes even more difficult!

UQ and Domain Shift

No UQ method is robust to domain shift by default.

- Standard UQ methods often assume **i.i.d. data** → can lead to **misleading uncertainty estimates** when this is violated.

- For scientific or safety-critical applications, **overconfident but wrong predictions** under domain shift are a serious risk.

- **Key challenges:**
  - Overconfidence - highly certain (small errors) even on out of domain data.
  - Shifted epistemic/aleatoric balance - epistemic error dominates.
  - Miscalibrated models - confidence ≠ accuracy.
  - Lack of target labels - we cannot evaluate UQ.

FERMI NATIONAL ACCELERATOR LABORATORY

06

Open Questions & Recent Trends

Trending Topics

1. Unsupervised and Few-shot DA

- Moving beyond labeled source-heavy settings.
- Combining **few-shot learning** and **DA** to adapt with just a few target labels.

2. Test-time Adaptation (TTA) & Source-free Domain Adaptation

- Models adapt **on-the-fly** during inference using only target data (entropy minimization, pseudo-labeling).
- Adapting a pretrained model using **only a trained source model**, not the data.
- No access to source data required; useful in privacy-constrained or streaming settings.

3. Multi-source and Multi-target DA

- Leveraging multiple source domains or generalizing to multiple target domains.
- Key challenge: dealing with **conflicting domain shifts** and source reliability.

Trending Topics

4. Out-of-Distribution Generalization

- Blurring lines between DA and OOD robustness.
- Focus on **domain generalization**: training on multiple domains and generalizing to unseen ones without adaptation.

5. Domain Adaptation for Regression / Scientific Applications

- DA research has mostly focused on classification; regression problems (e.g., physics, medical data) are getting more attention.
- Requires new alignment methods and uncertainty-aware losses.

6. Vision-Language Models & Multimodal DA

- Adapting **multimodal representations** (e.g., text+image) to new domains.
- Requires alignment across **both modalities and domains**.

Unsolved Problems

1. Mismatch of Label Distributions

- Most methods assume P**(Y|X)** remains the same, but in real-world settings it may change → **label shift**, which is under-addressed.

2. Lack of Theoretical Guarantees

- Few practical methods come with **provable generalization bounds** under domain shift.
- Need better **theory for real-world, high-dimensional settings**.

3. Interpretability and Trust

- Black-box adaptations make it hard to understand **what features are being aligned**.
- Especially important in **scientific and safety-critical domains**.

Unsolved Problems

4. Measuring Adaptation Success

- Evaluating whether a model has "adapted well" remains vague.
- Need better **metrics and diagnostics** (e.g., generalization error decomposition, target risk estimation).

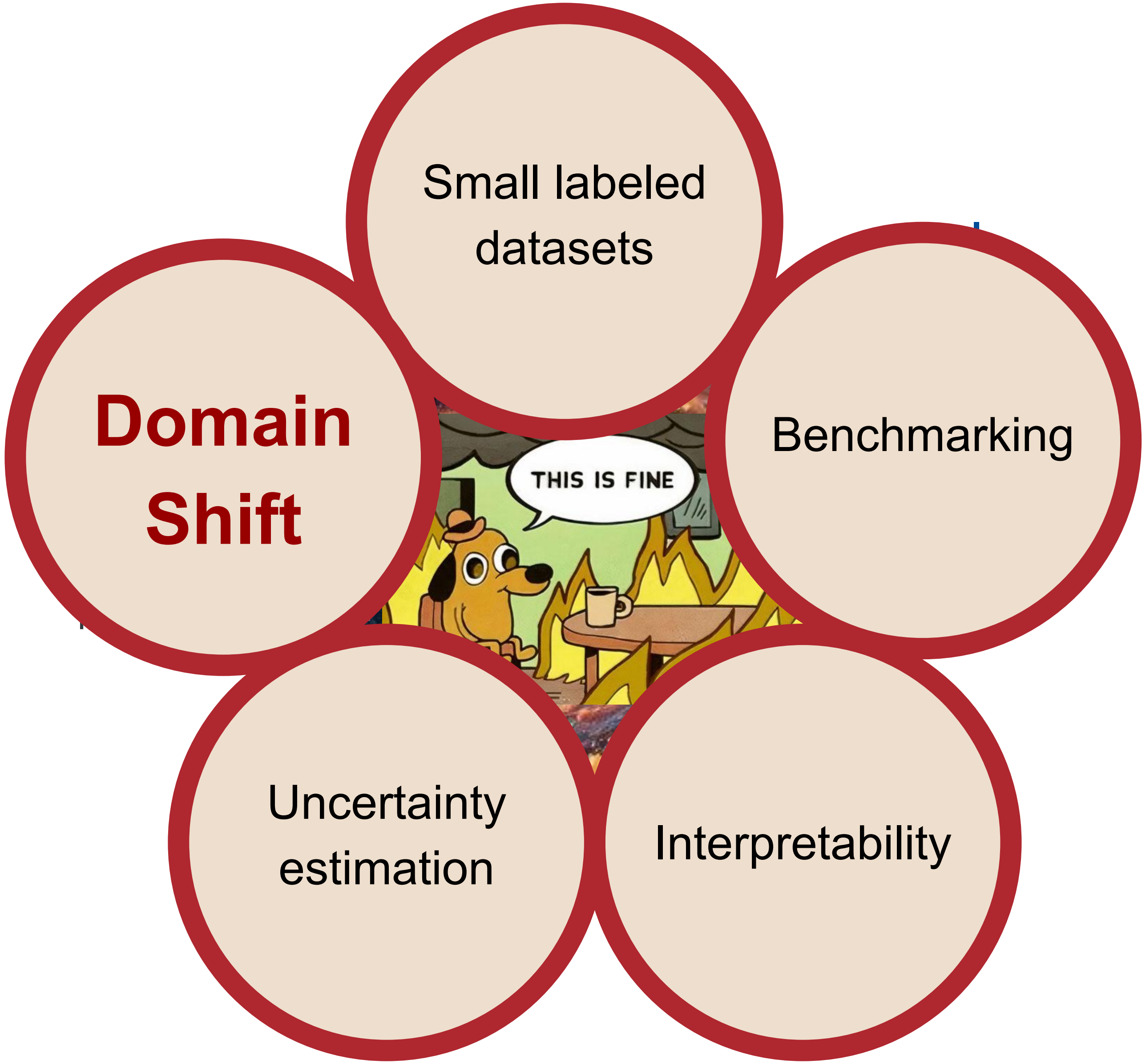
5. Continual and Lifelong Adaptation

- Adapting to a **sequence of domains** without forgetting past ones.
- Combines domain adaptation with **continual learning**.

6. Efficient and Scalable Adaptation

- Many DA algorithms are compute-heavy or memory-intensive.
- Need **efficient, lightweight** solutions for deployment in edge devices or real-time systems.

Summary

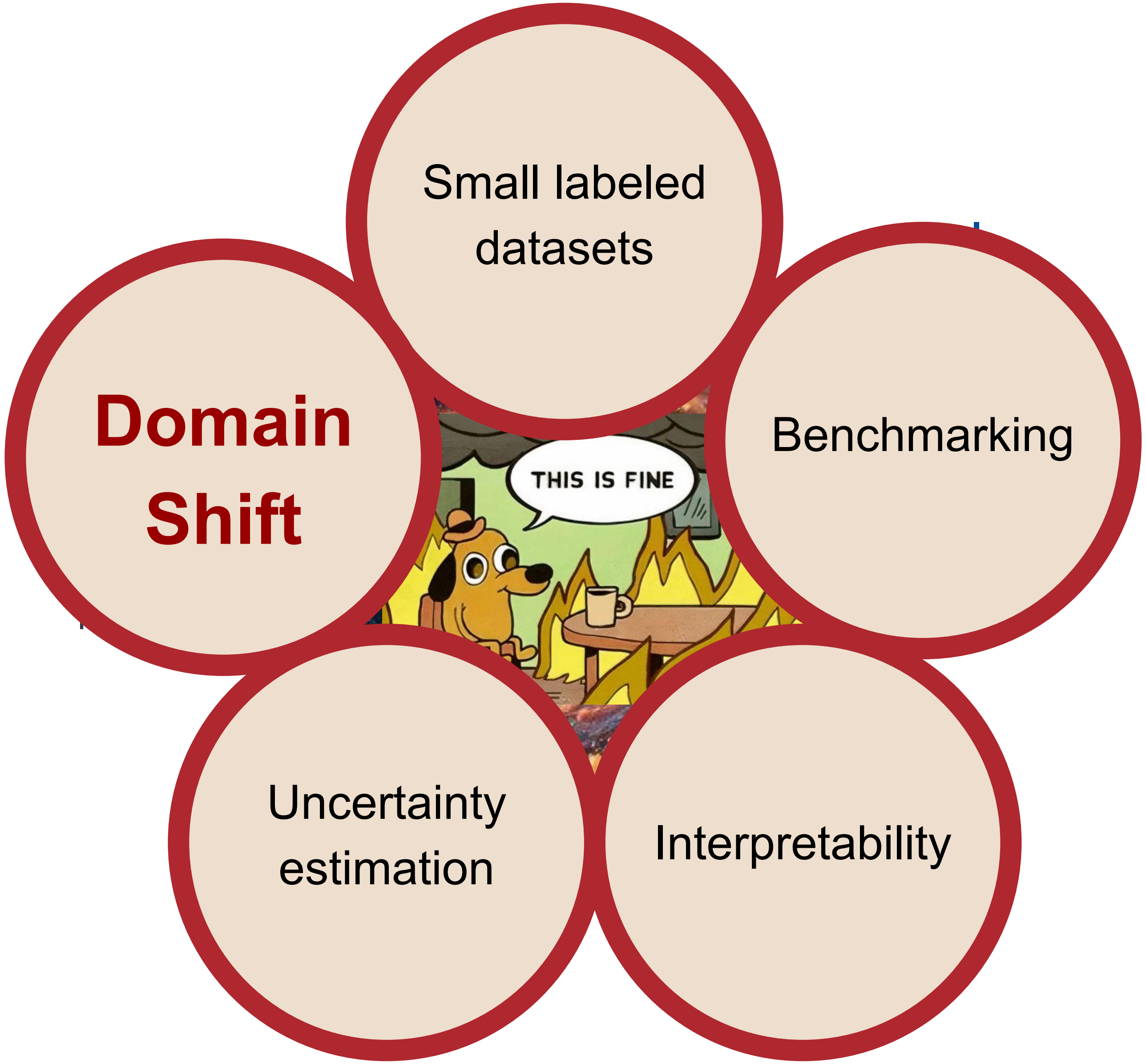


- Era of big astro surveys - **large amount of data but not all of it labeled** (i.e. suitable for easy AI applications).
- Simulations and old data are different - **domain shift problem**!
- **Domain Adaptation** can help but it cannot solve all problems.
- Hands on session:
  - Learn about different DA methods & how to tune them.

Summary

THANK YOU!

aleksand@fnal.gov



- Era of big astro surveys - **large amount of data but not all of it labeled** (i.e. suitable for easy AI applications).
- Simulations and old data are different - **domain shift problem**!
- **Domain Adaptation** can help but it cannot solve all problems.
- Hands on session:
  - Learn about different DA methods & how to tune them.

Fermilab

FermiFORWARD

U.S. DEPARTMENT of ENERGY

FERMI NATIONAL ACCELERATOR LABORATORY