Bridging the gap between astronomical datasets with AI

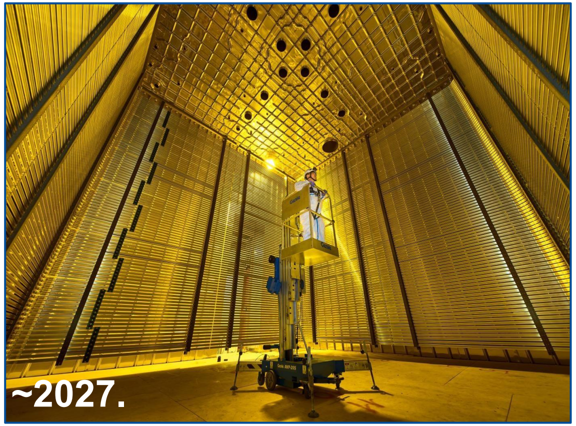
- Domain Shift, Model Robustness and Failure Modes

Aleksandra Ćiprijanović
(she/her/hers)

Fermilab, DSSL
aleksand@fnal.gov

KITP Review Talk
February, 2023

Vision of the Future



Rubin LSST

~ 20 TB / day
~ 100 PB total by DR11

DUNE

~ 30-60 PB / year (raw)
~ 114x4 TB / month (raw)
for Supernovae detection
(speed need for followups)

HL-LHC

~ order of magnitude more data
~ 650 PB / year

Vision of the Future

- **Real-time:**
  - data handling,
  - decision making
  - detection of interesting events
  - inference
- **Automated experiments**
- **Working with big data** later in the process

Rubin LSST

~ 20 TB / day
~ 100 PB total by DR11

DUNE

~ 30-60 PB / year (raw)
~ 114x4 TB / month (raw) for Supernovae detection (speed need for followups)

HL-LHC

~ order of magnitude more data
~ 650 PB / year

Fermilab

Vision of the Future

Rubin LSST

~ 20 TB / day
~ 100 PB total by DR11

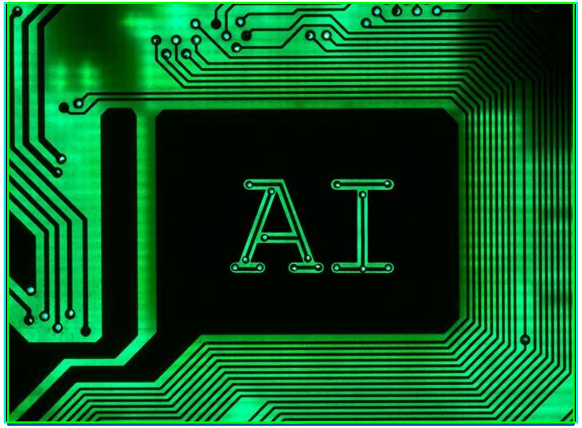
DUNE

~ 30-60 PB / year (raw)
~ 114x4 TB / month (raw)
for Supernovae detection
(speed need for followups)

HL-LHC

~ order of magnitude more data
~ 650 PB / year

Vision of the Future



Rubin LSST

~ 20 TB / day
~ 100 PB total by DR11

DUNE

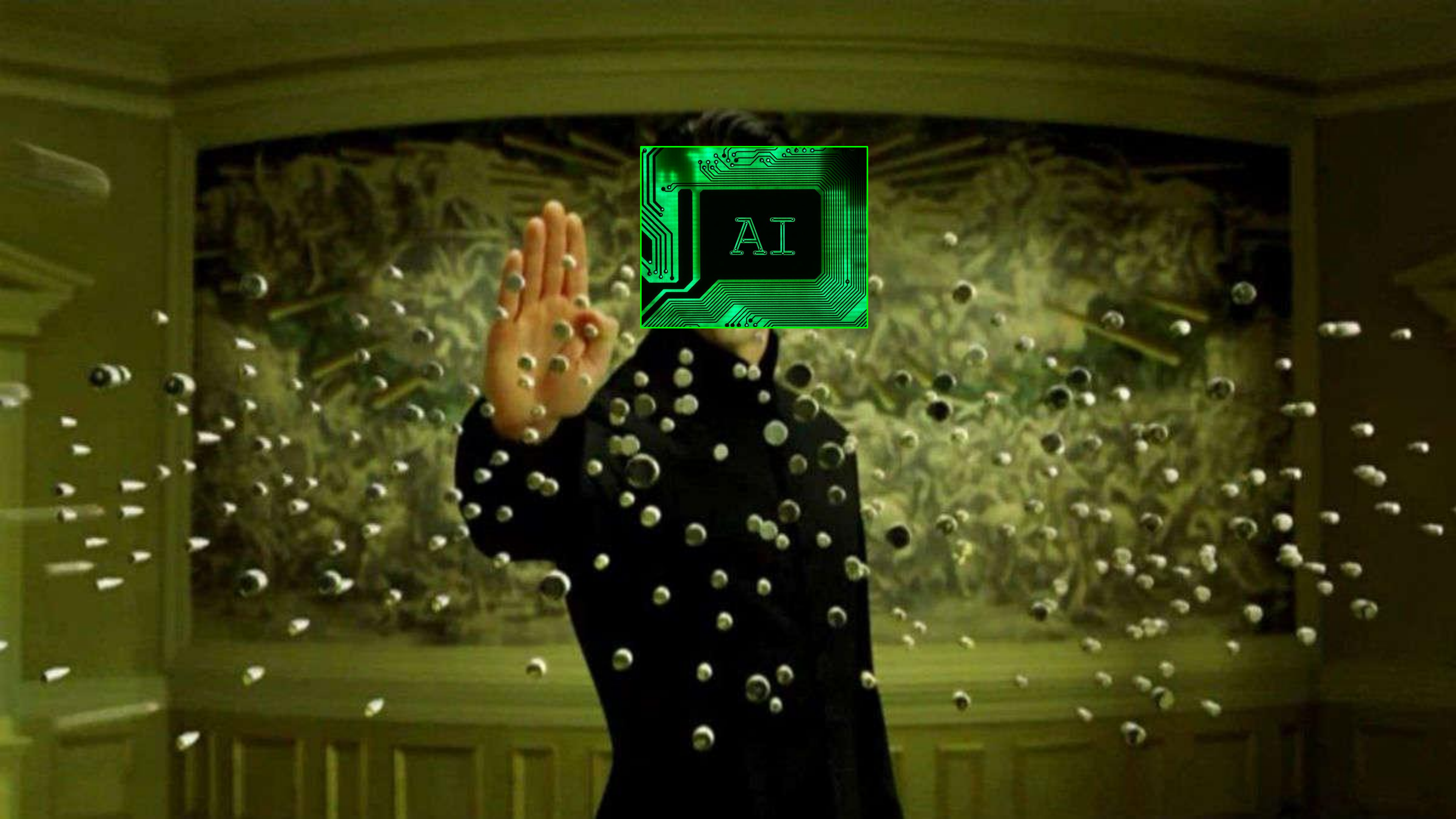
~ 30-60 PB / year (raw)
~ 114x4 TB / month (raw)
for Supernovae detection
(speed need for followups)

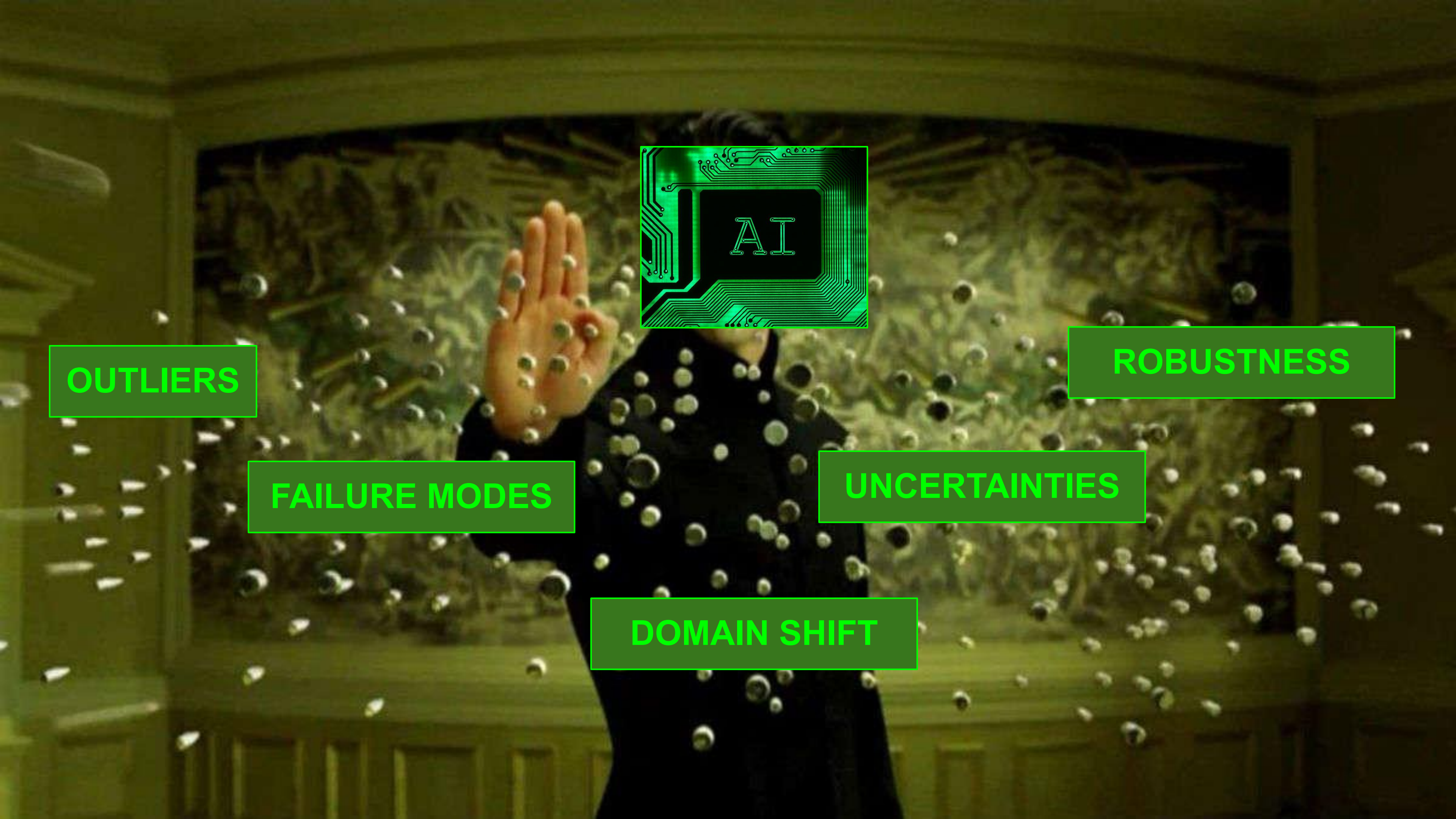
HL-LHC

~ order of magnitude more data
~ 650 PB / year

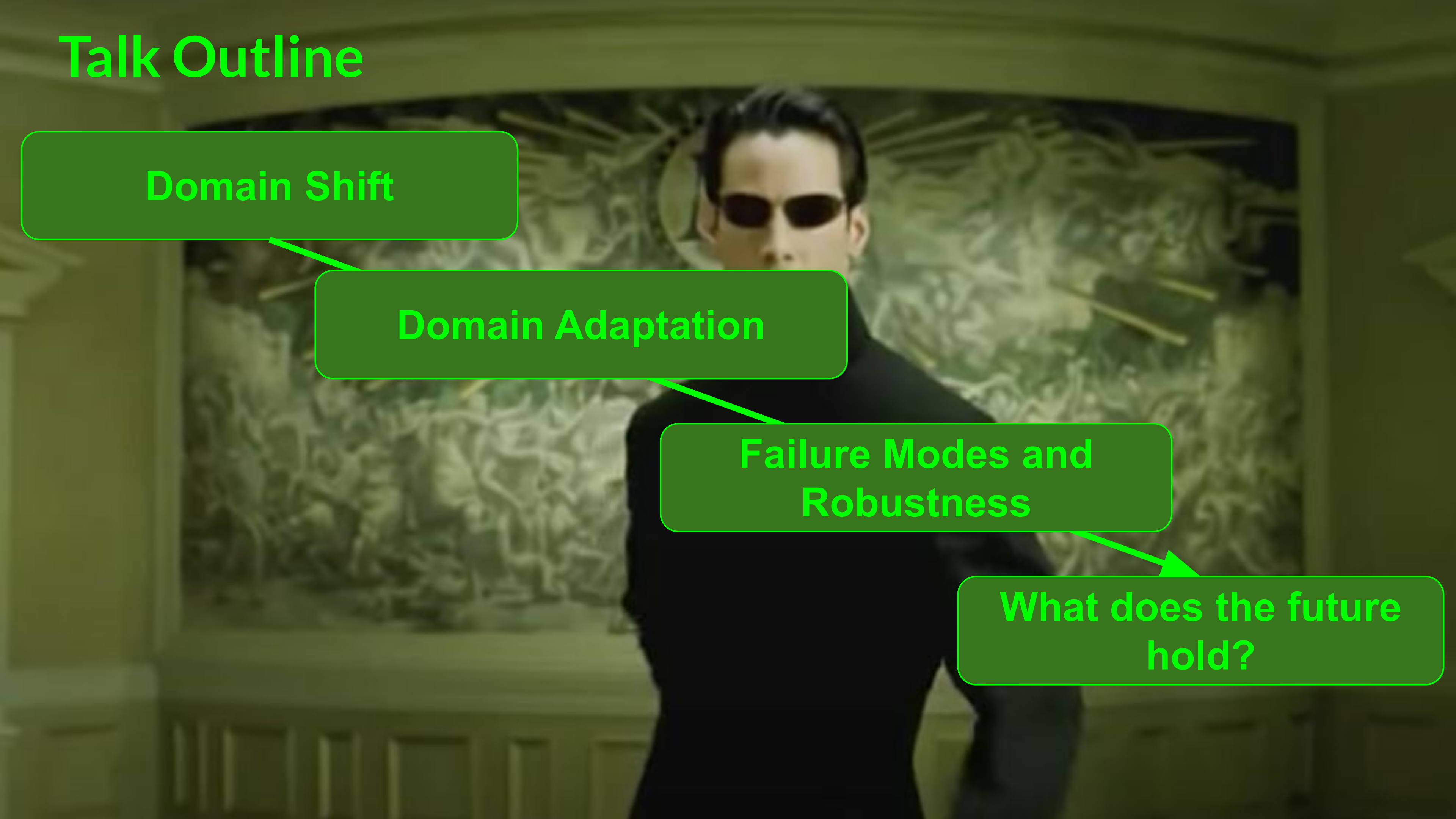
1999. reference….


AI


AI


AI
OUTLIERS
ROBUSTNESS
FAILURE MODES
UNCERTAINTIES
DOMAIN SHIFT

Talk Outline



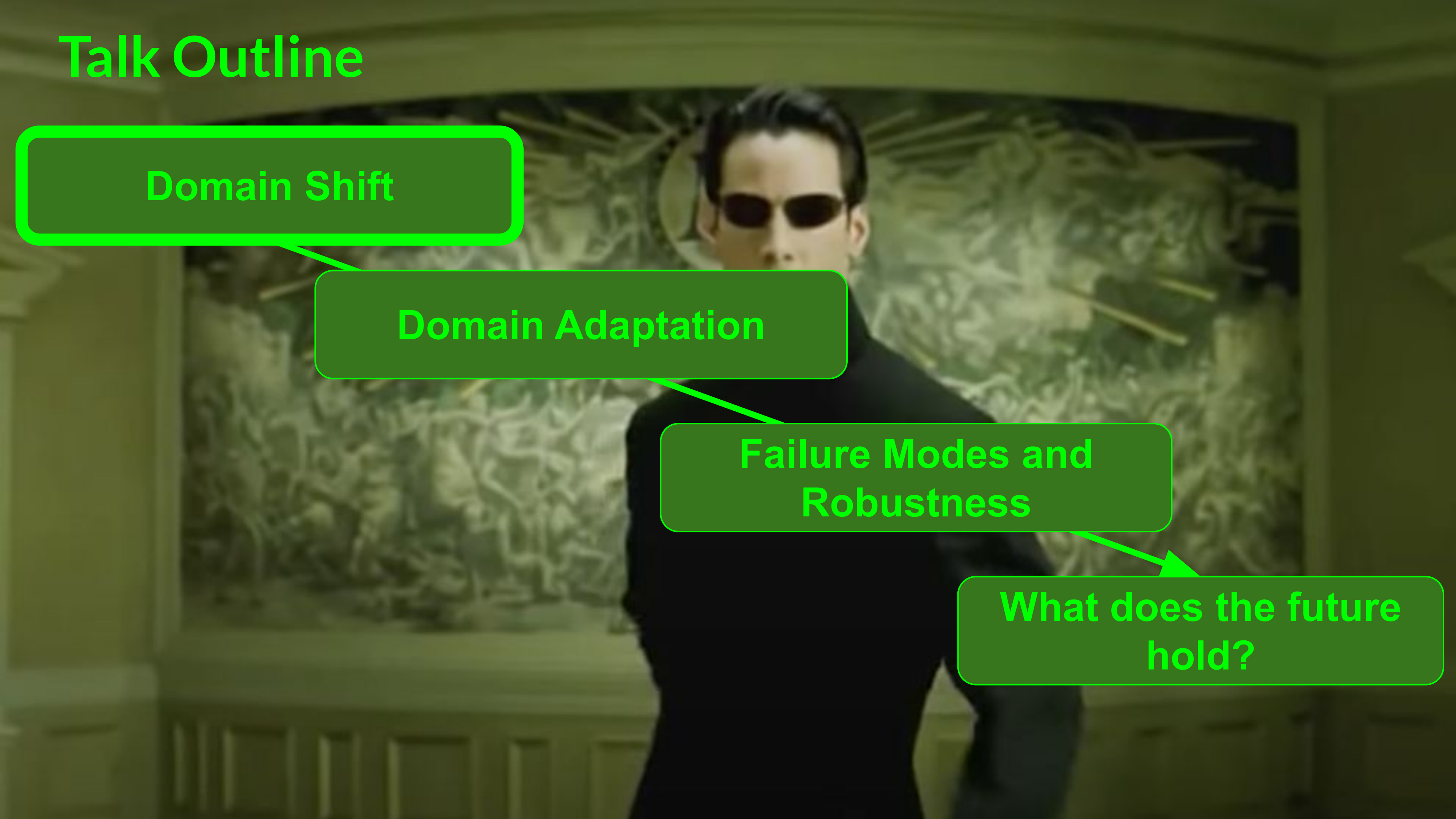
Domain Shift

Domain Adaptation

Failure Modes and Robustness

What does the future hold?

Talk Outline



- Domain Shift
- Domain Adaptation
- Failure Modes and Robustness
- What does the future hold?

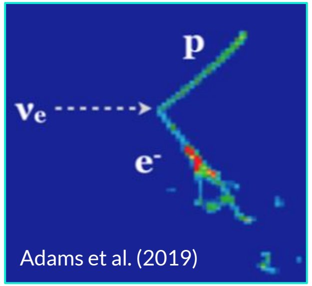
Combining Datasets

All areas of science often need to create **model trained on simulated data, that also work on real detector data**!

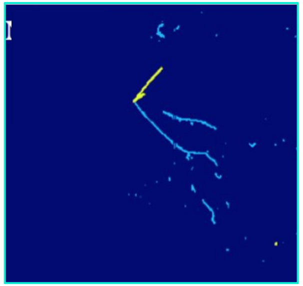
DATASETS ARE DIFFERENT!

SIMULATED | REAL

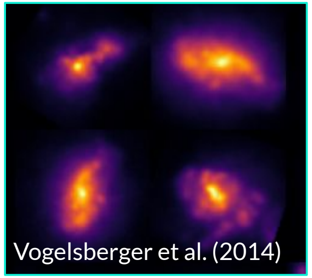
MicroBooNE (neutrinos)



Adams et al. (2019)



Illustris / Hubble (merging galaxies)


Vogelsberger et al. (2014)


Hubble

Fermilab

Combining Datasets

All areas of science often need to create **model trained on simulated data, that also work on real detector data**!

DATASETS ARE DIFFERENT!

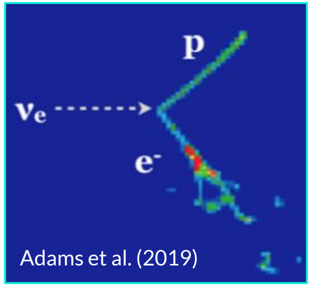
Missing and unknown physics, wrong geometry, background levels

Computational constraints for simulations

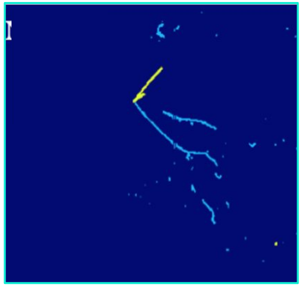
SIMULATED

REAL

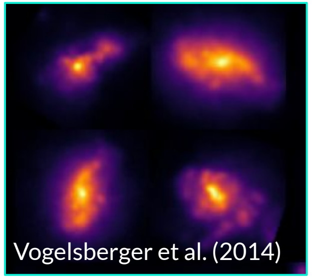
MicroBooNE (neutrinos)



Adams et al. (2019)



Illustris / Hubble (merging galaxies)


Vogelsberger et al. (2014)

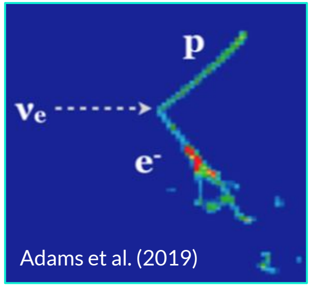
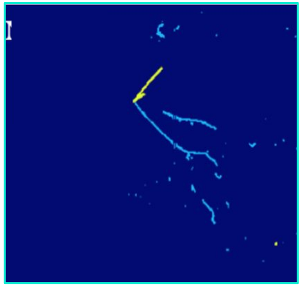
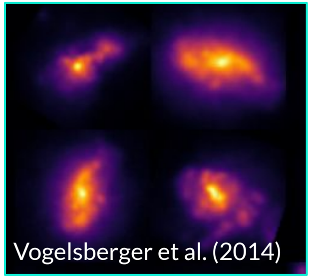

Hubble

Combining Datasets

All areas of science often need to create **model trained on simulated data, that also work on real detector data**!

DATASETS ARE DIFFERENT!

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

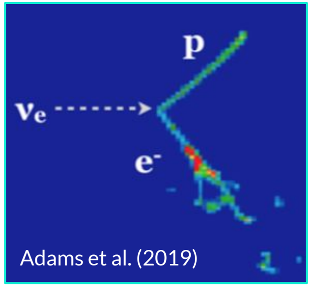
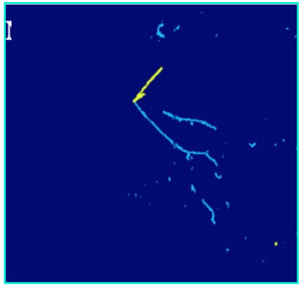
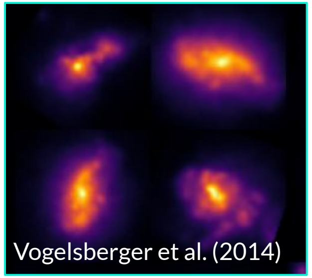
	SIMULATED	REAL
MicroBooNE (neutrinos)		
Illustris / Hubble (merging galaxies)		

Combining Datasets

All areas of science often need to create **model trained on simulated data, that also work on real detector data**!

DATASETS ARE DIFFERENT!

- 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)		

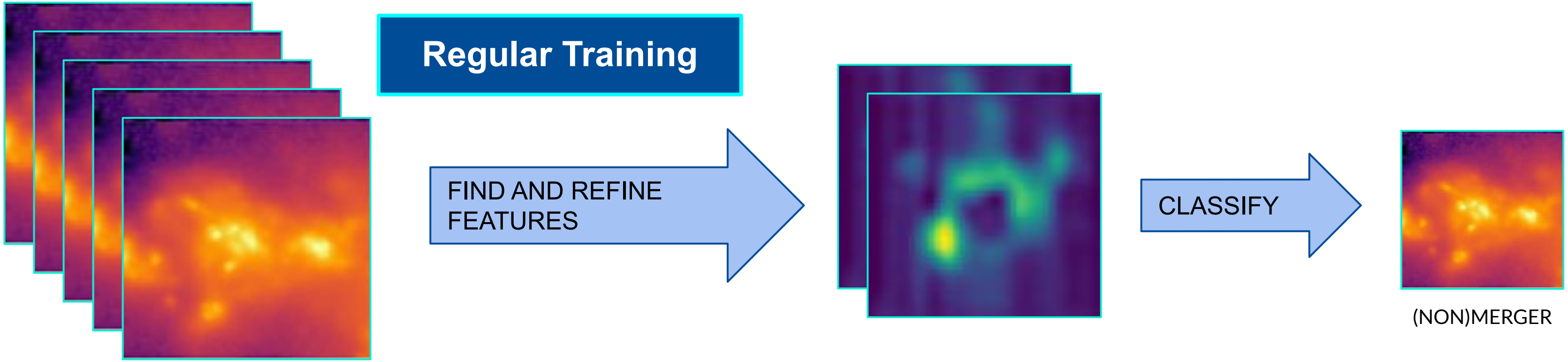
Adams et al. (2019)

Vogelsberger et al. (2014)

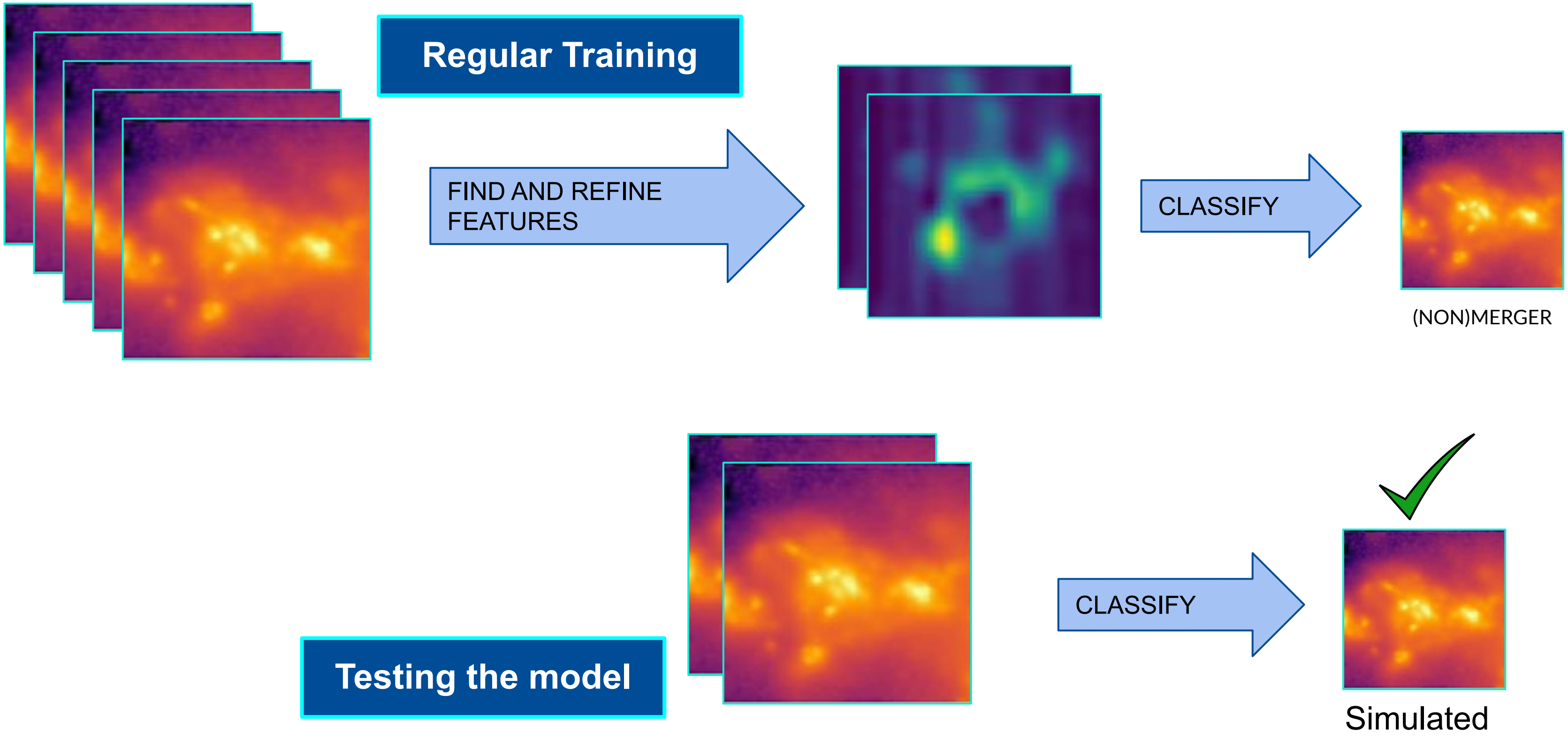
Hubble

Fermilab

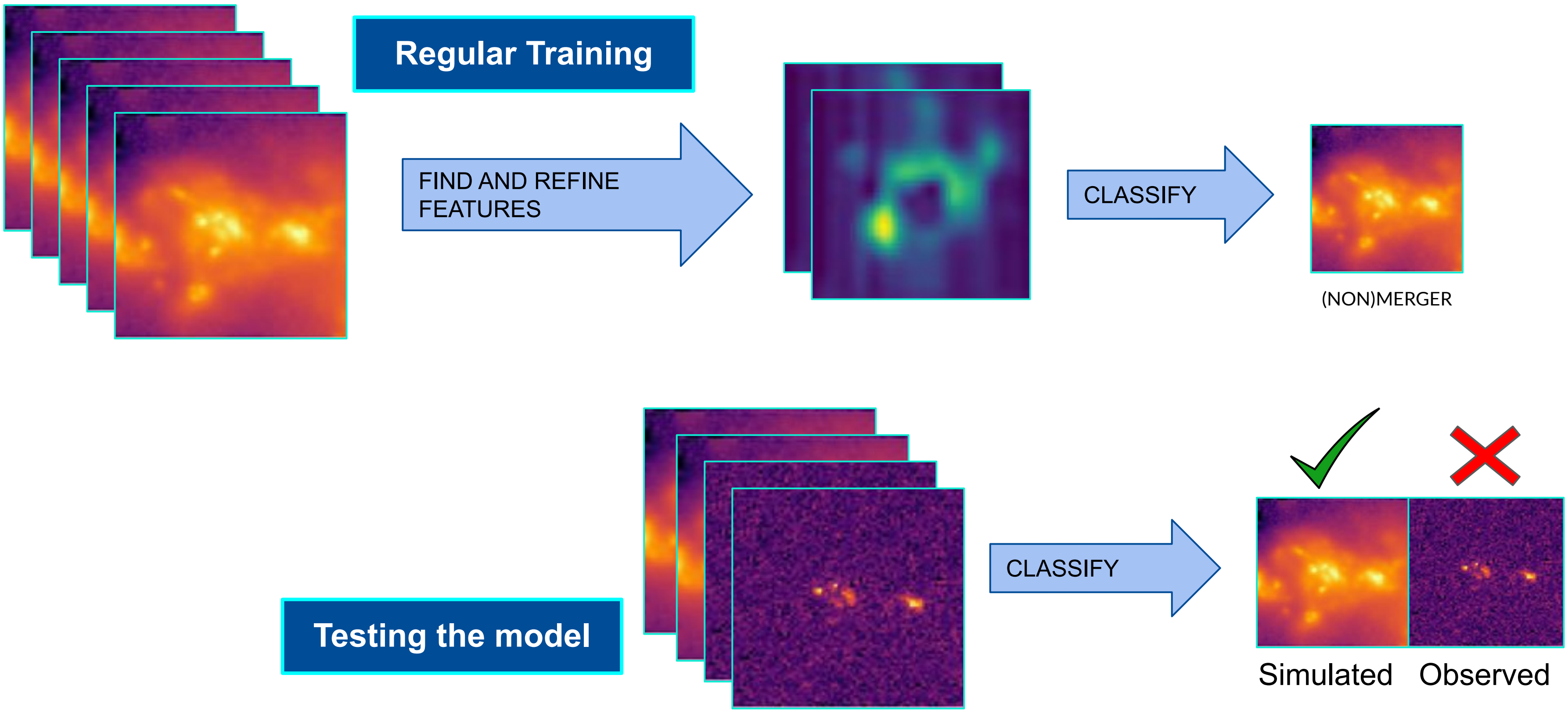
Combining Datasets



Combining Datasets



Combining Datasets



Combining Datasets

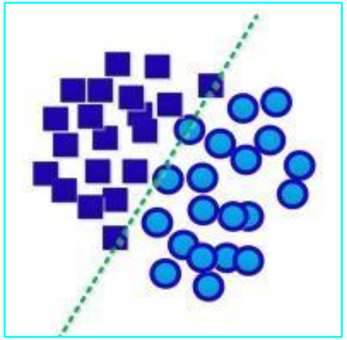
Why does this happen?

Combining Datasets

Why does this happen?

Source Domain

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



Fermilab

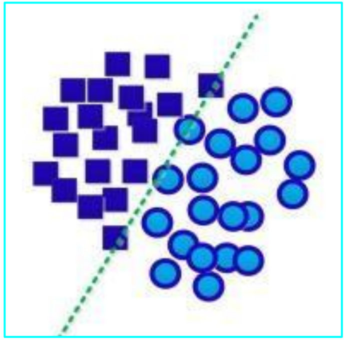
Combining Datasets

Why does this happen?

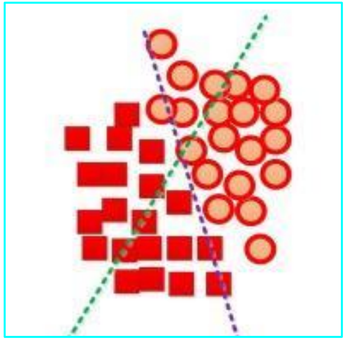
DOMAIN SHIFT !

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

Source Domain



Target Domain



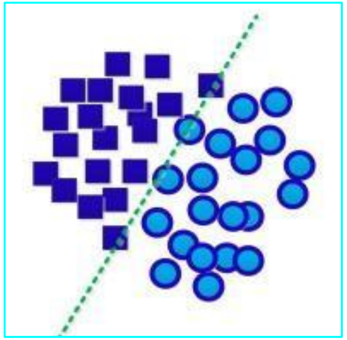
Combining Datasets

Why does this happen?

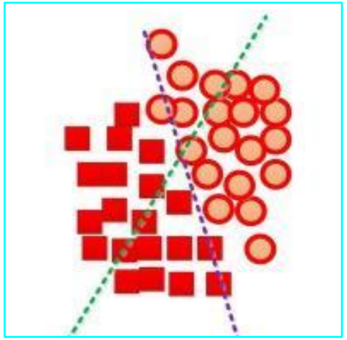
DOMAIN SHIFT !

We need to align the data during training!

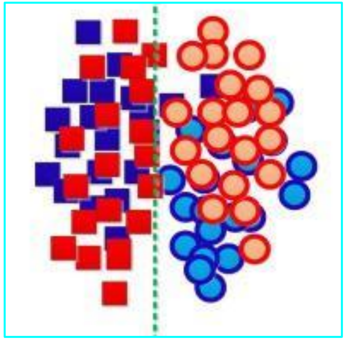
Source Domain



Target Domain



Domain Alignment



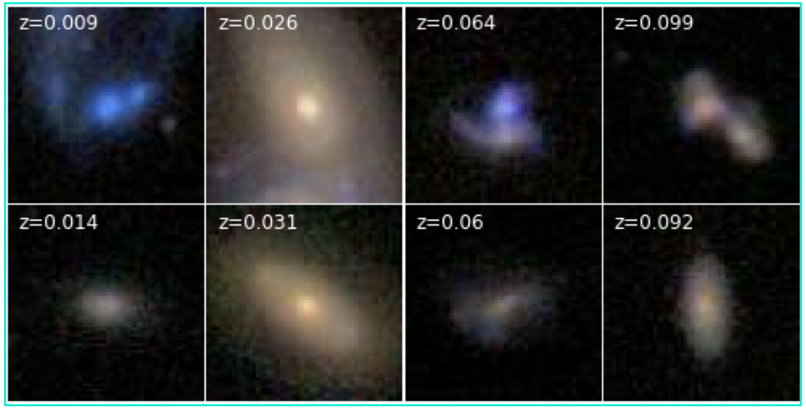
Domain Shift - Example 1

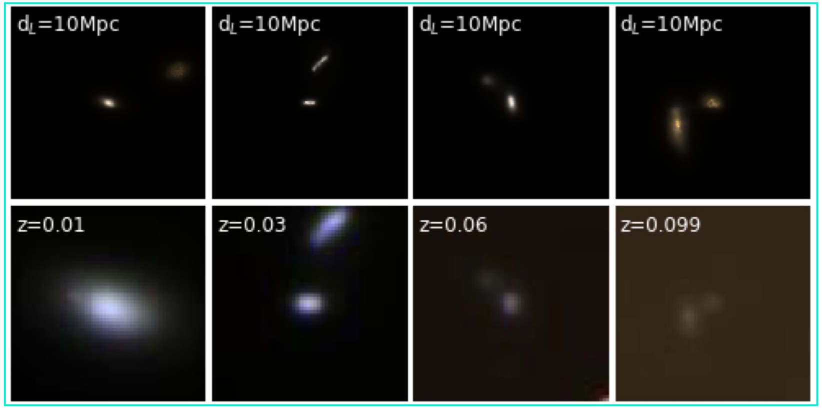
Mergers vs. non-mergers

SDSS

Pearson et al. 2019.

Eagle





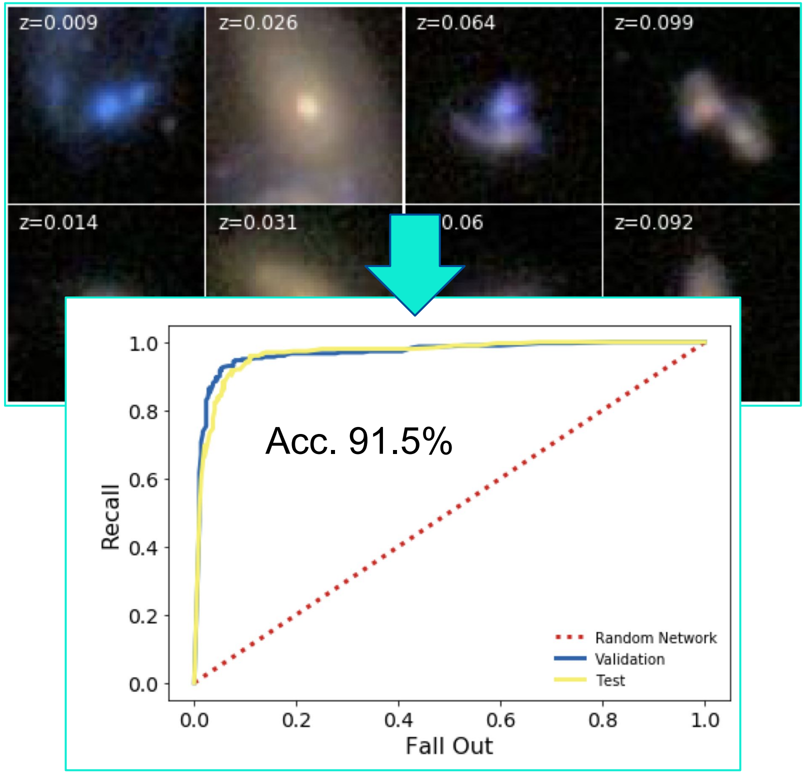
Domain Shift - Example 1

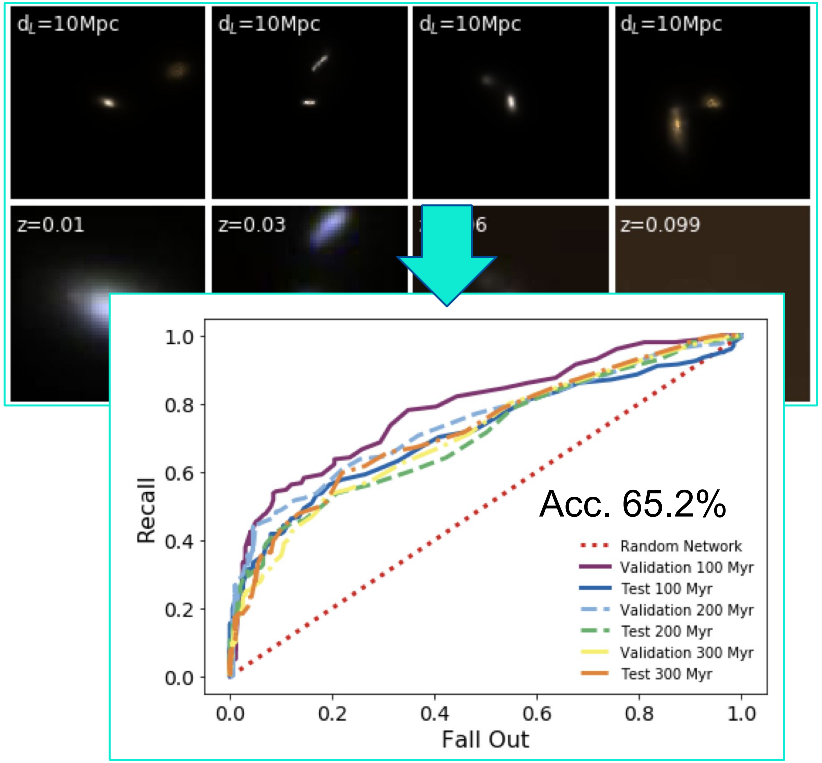
Mergers vs. non-mergers

SDSS

Pearson et al. 2019.

Eagle





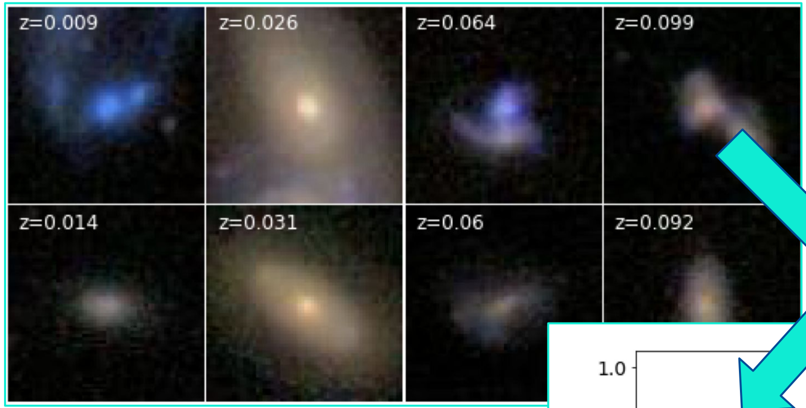
Fermilab

Domain Shift - Example 1

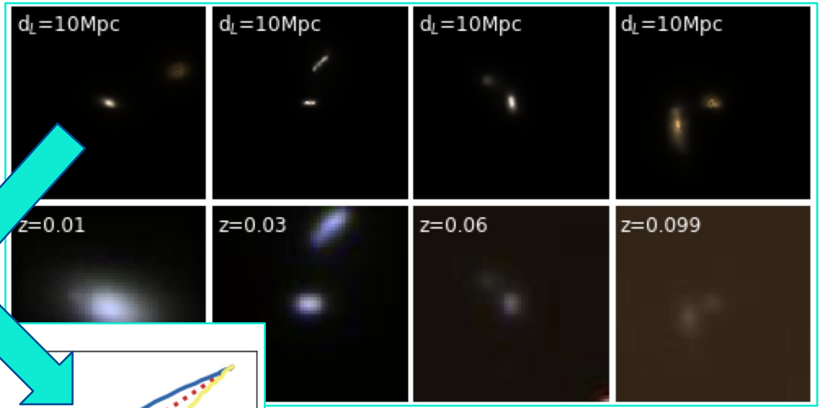
Mergers vs. non-mergers

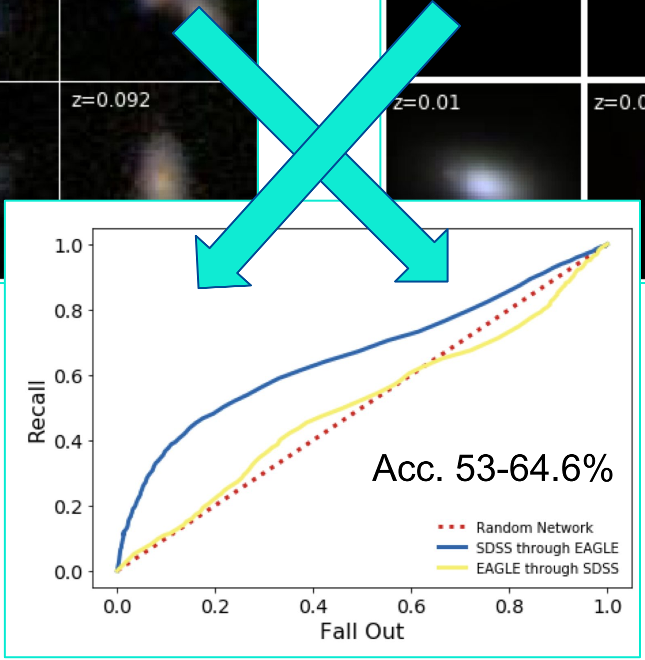
Pearson et al. 2019.

SDSS



Eagle



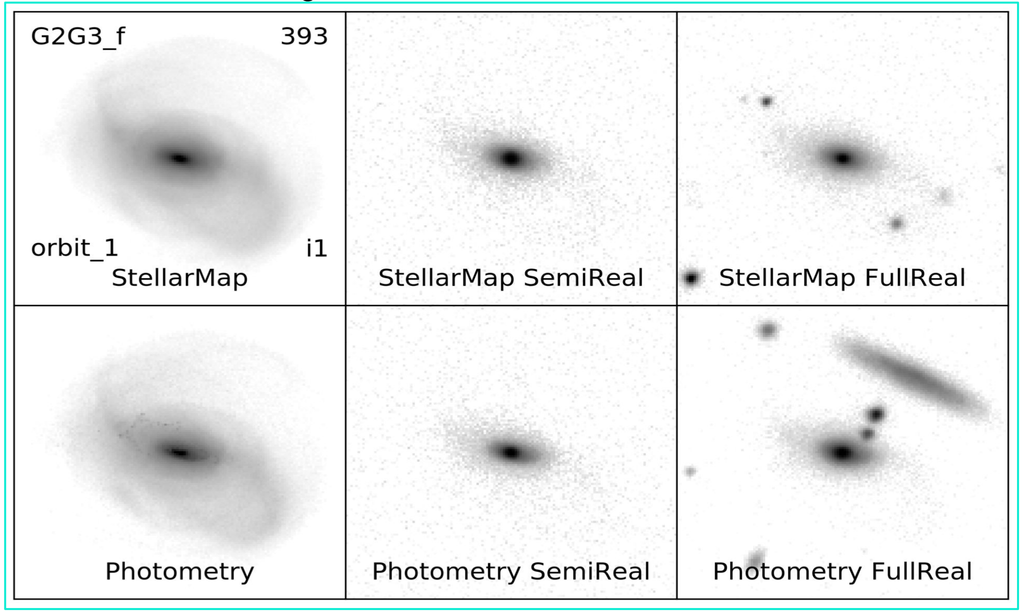


Domain Shift - Example 2

Classifying merger stage

Bottrell et al. 2019.

Moreno et al. 2019 gal. interaction simulation



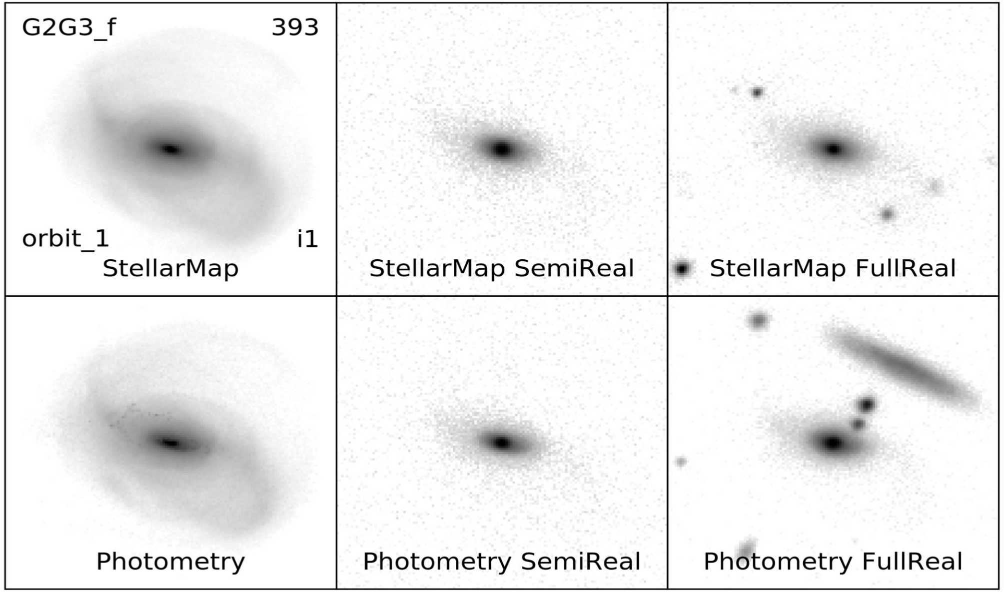
No obs. effects | PSF+noise | real sky background

Domain Shift - Example 2

Classifying merger stage

Bottrell et al. 2019.

Moreno et al. 2019 gal. interaction simulation

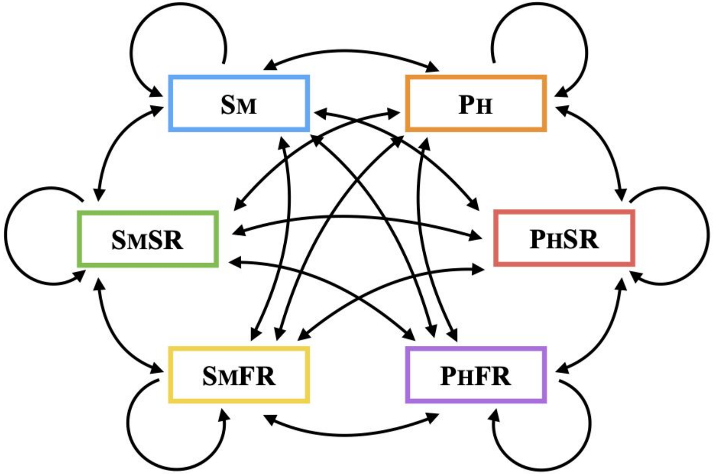


No obs. effects

PSF+noise

real sky background

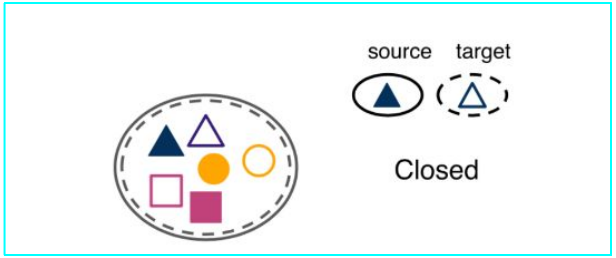
fully realistic training images are needed to achieve high acc. on fully-realistic test set



Otherwise you often get as low as ~50% acc.

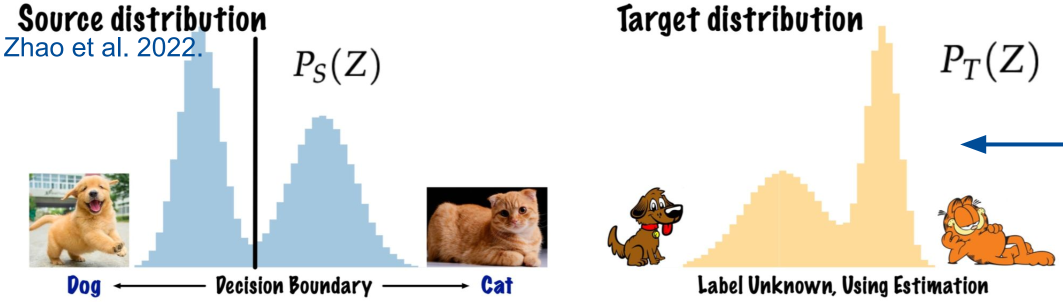
Fermilab

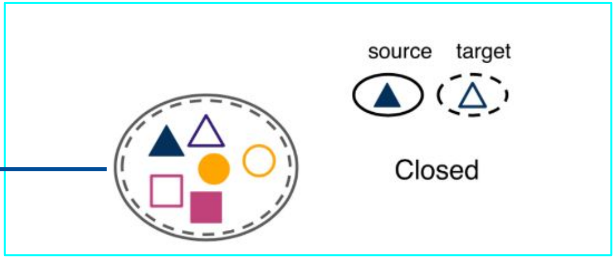
Types of Domain Shift Problems



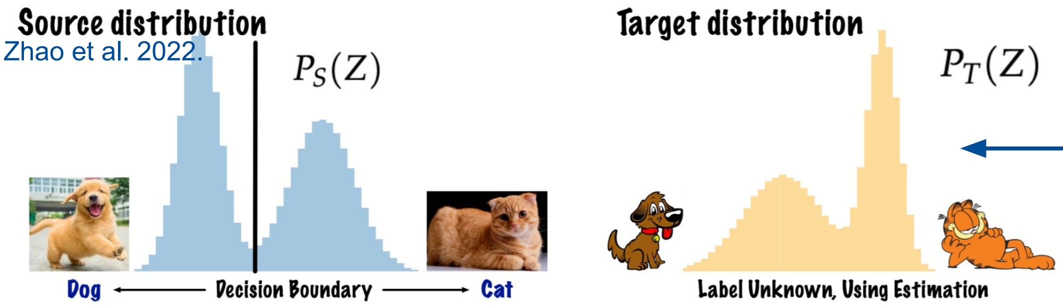
Types of Domain Shift Problems

Zhao et al. 2022.



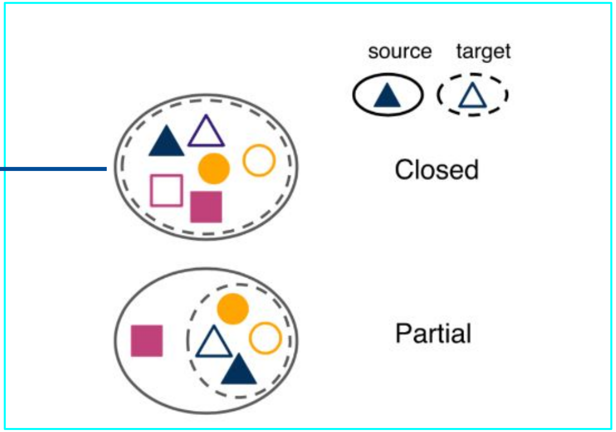


Types of Domain Shift Problems



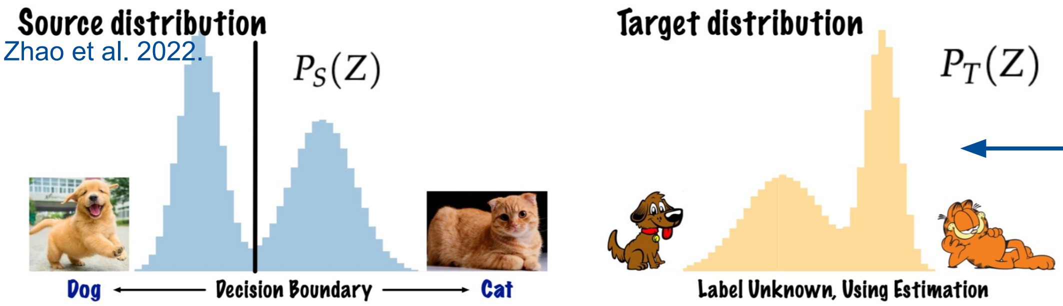
Zhao et al. 2022.

"Bird"

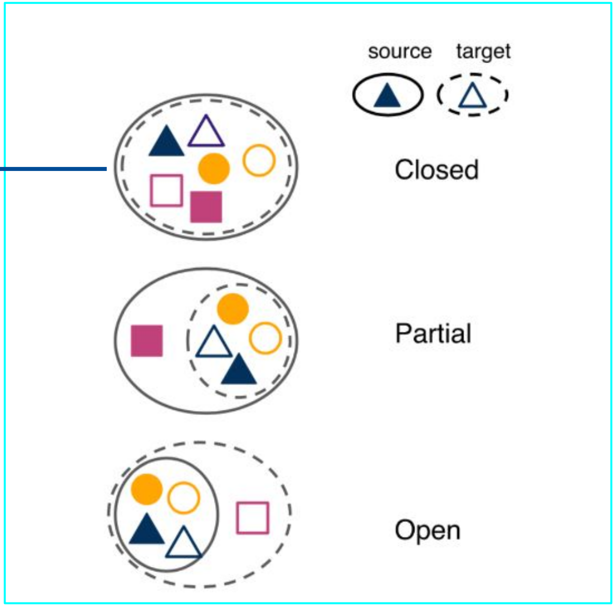


Types of Domain Shift Problems

Zhao et al. 2022.

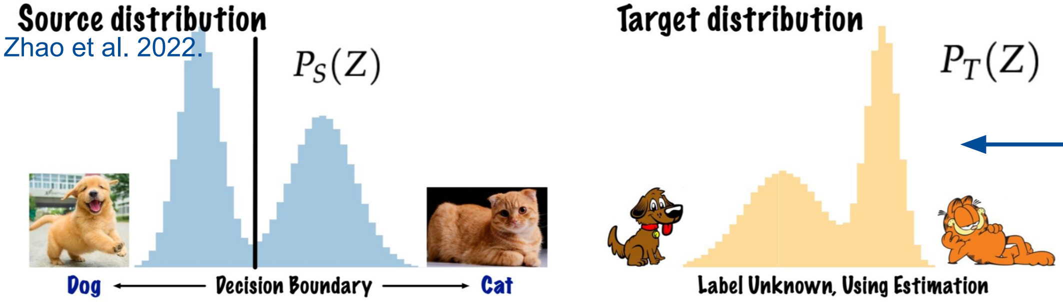


"Bird"



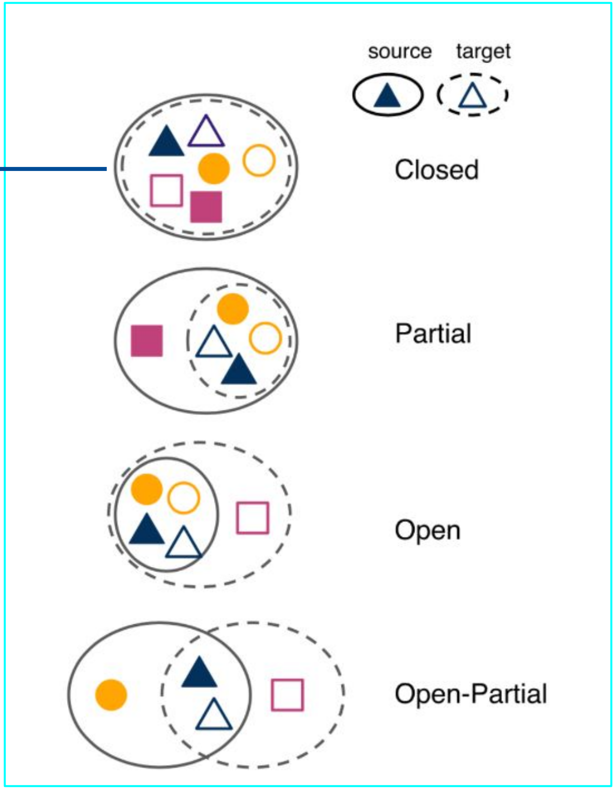
Types of Domain Shift Problems

Zhao et al. 2022.



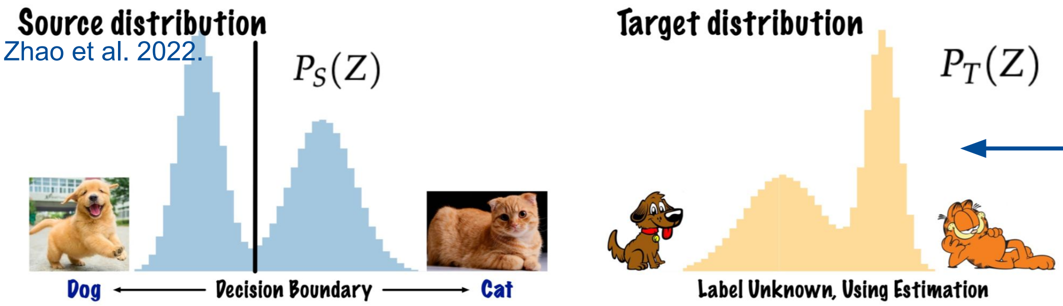
"Bear"

"Bird"



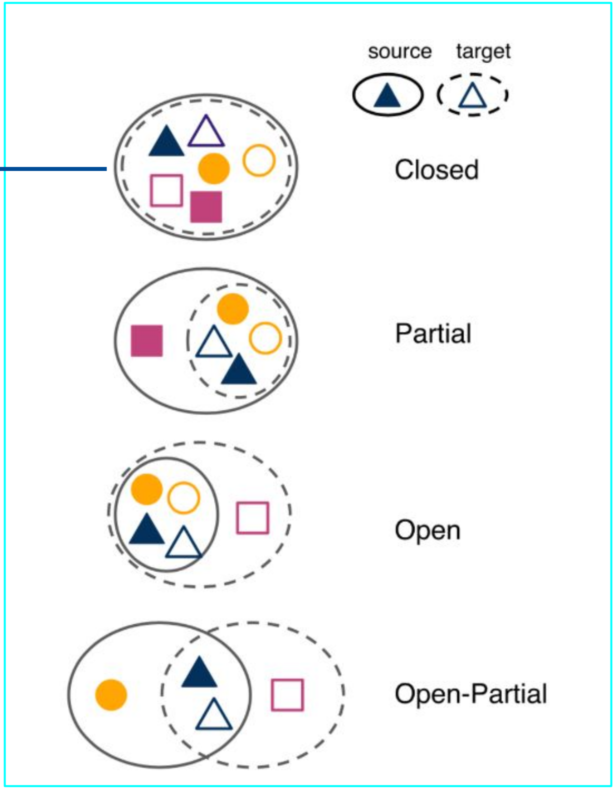
Types of Domain Shift Problems

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.

Combining Datasets

Solution?

Talk Outline

- Domain Shift
- **Domain Adaptation**
- Failure Modes and Robustness
- What does the future hold?

Combining Datasets

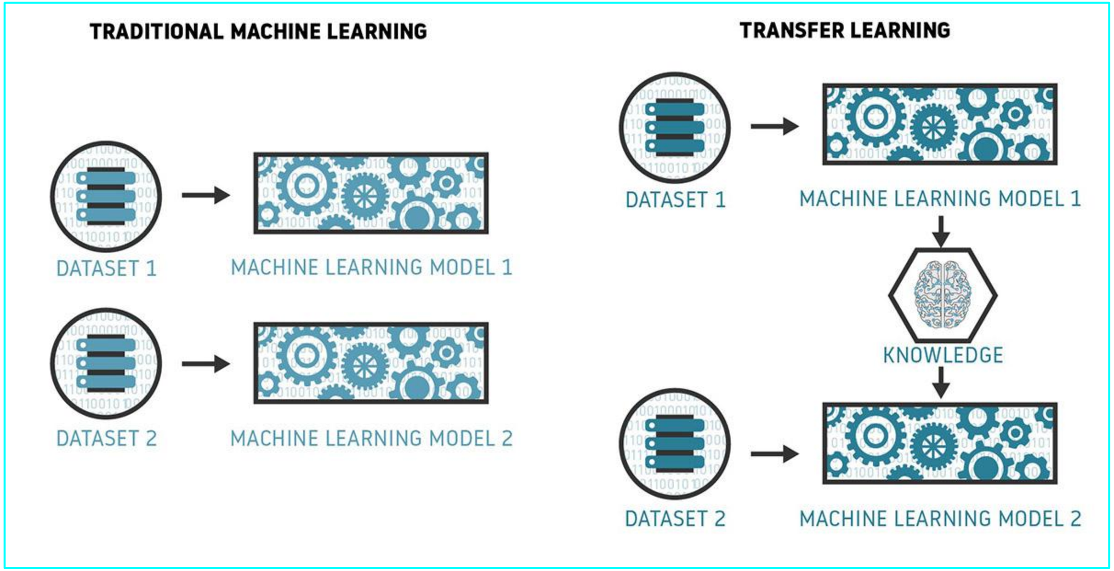
Solution?

OPTION 1: My new datasets is fully or partially labeled

Combining Datasets

TRANSFER LEARNING

OPTION 1: My new datasets is fully or partially labeled



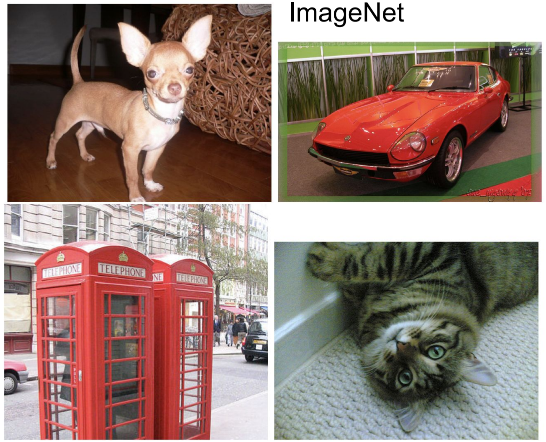
Fermilab

Combining Datasets

TRANSFER LEARNING

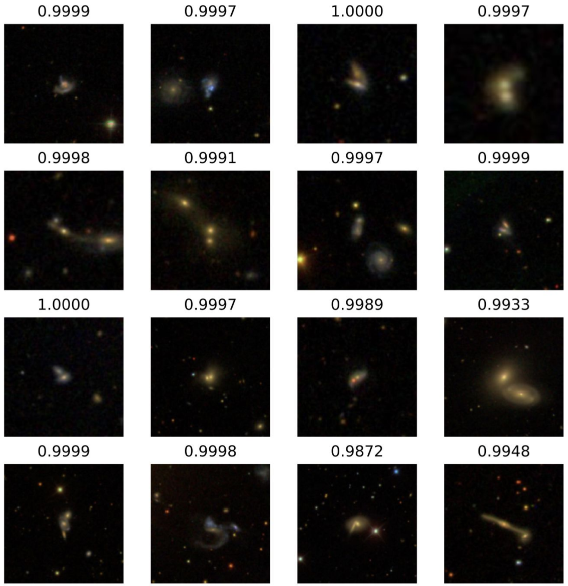
OPTION 1: My new datasets is fully or partially labeled

ImageNet



Ackermann et al. 2018.

SDSS DR7



Fermilab

Combining Datasets

Solution?

OPTION 2: My new datasets is sadly unlabeled

Combining Datasets

DOMAIN ADAPTATION

OPTION 2: My new datasets is sadly unlabeled

Combining Datasets

DOMAIN ADAPTATION

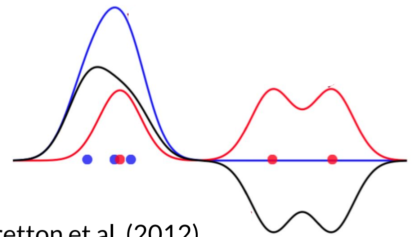
Align data distributions in the latent space of the network by forcing the network to **find more robust domain-invariant features.**

Combining Datasets

DOMAIN ADAPTATION

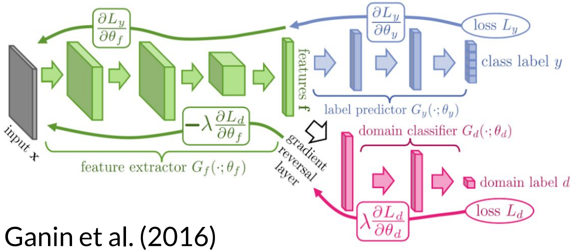
Align data distributions in the latent space of the network by forcing the network to **find more robust domain-invariant features.**

Distance-based methods



Gretton et al. (2012)

Adversarial methods



Ganin et al. (2016)

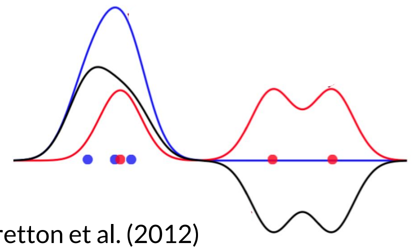
Combining Datasets

DOMAIN ADAPTATION

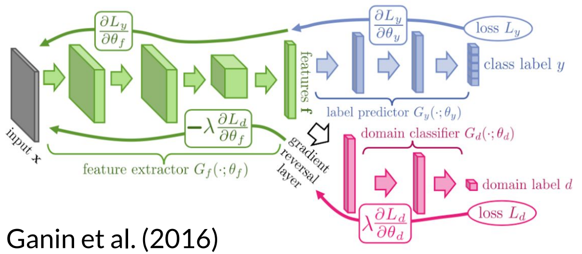
Align data distributions in the latent space of the network by forcing the network to **find more robust domain-invariant features.**

Distance-based methods

Adversarial methods


Gretton et al. (2012)


Closed



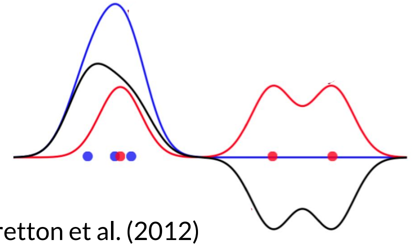
Ganin et al. (2016)

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.**

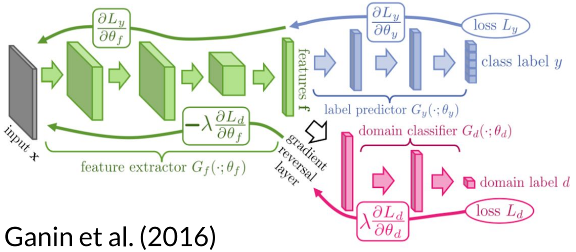
Distance-based methods



Gretton et al. (2012)

Training = Task Loss + DA Loss

Adversarial methods



Ganin et al. (2016)

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.**

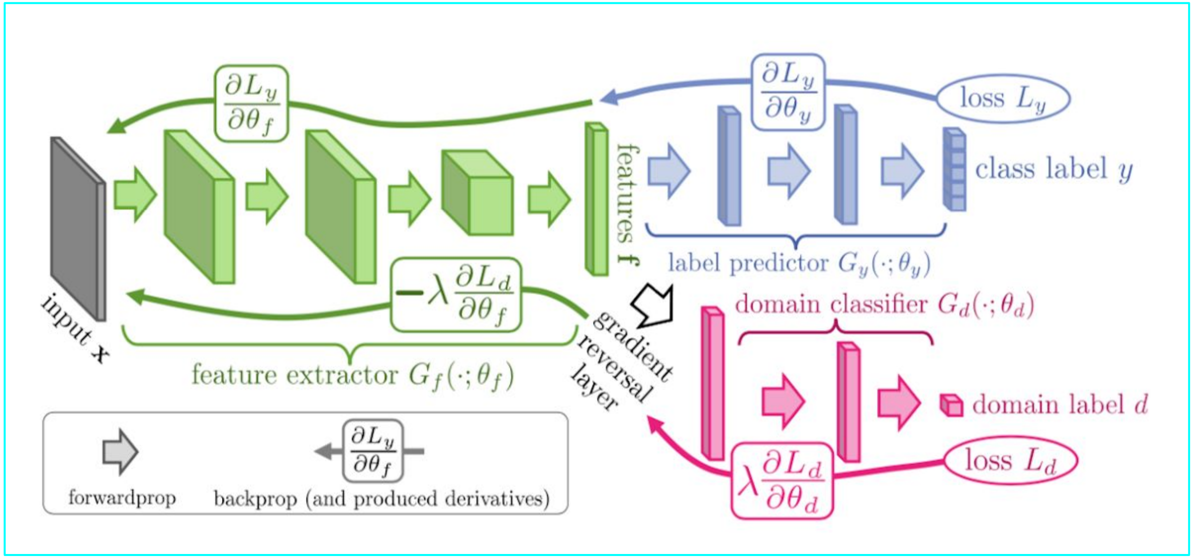
- Distance-based methods
- Adversarial methods

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

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.

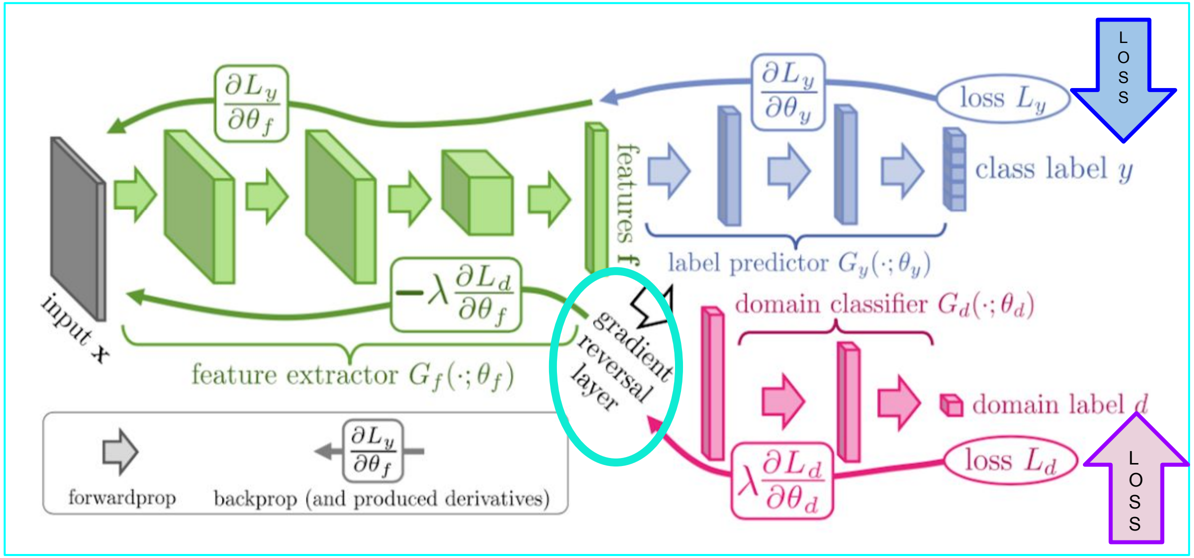


Ganin et al. (2016)

Domain Adversarial Neural Networks - DANNs

DANN - feature extractor + label predictor + domain classifier

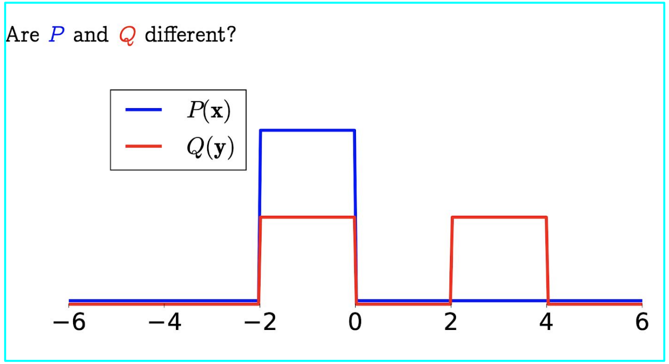
- **Gradient reversal layer** - multiplies the gradient by a negative constant during the backpropagation.
- Results in the extraction of **domain-invariant features**.
- Only source domain images are labeled during training.

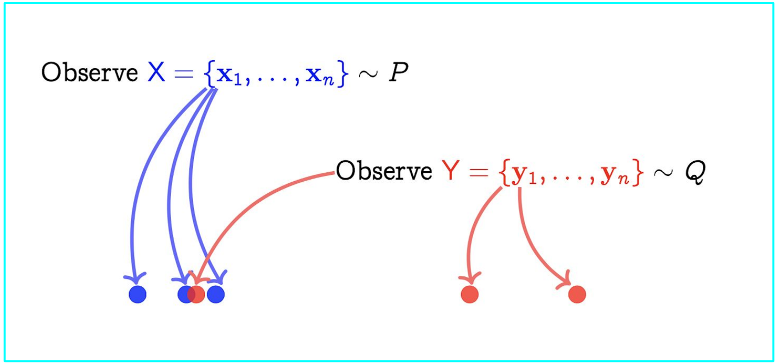


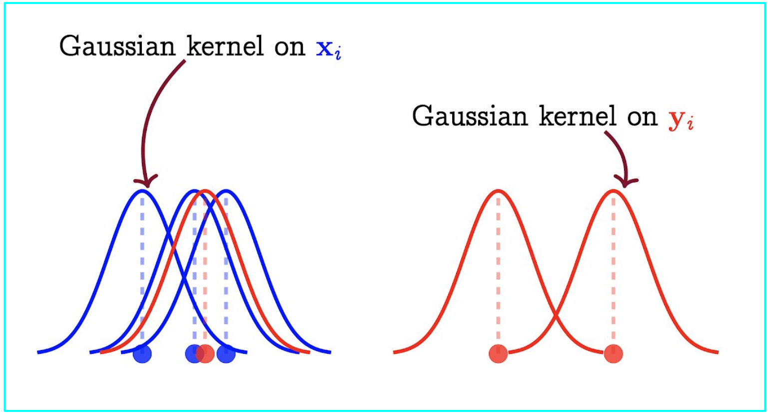
Ganin et al. (2016)

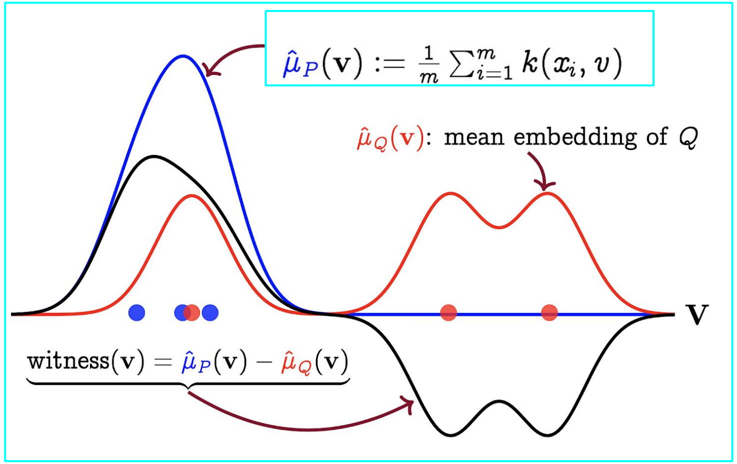
Maximum Mean Discrepancy - MMD

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

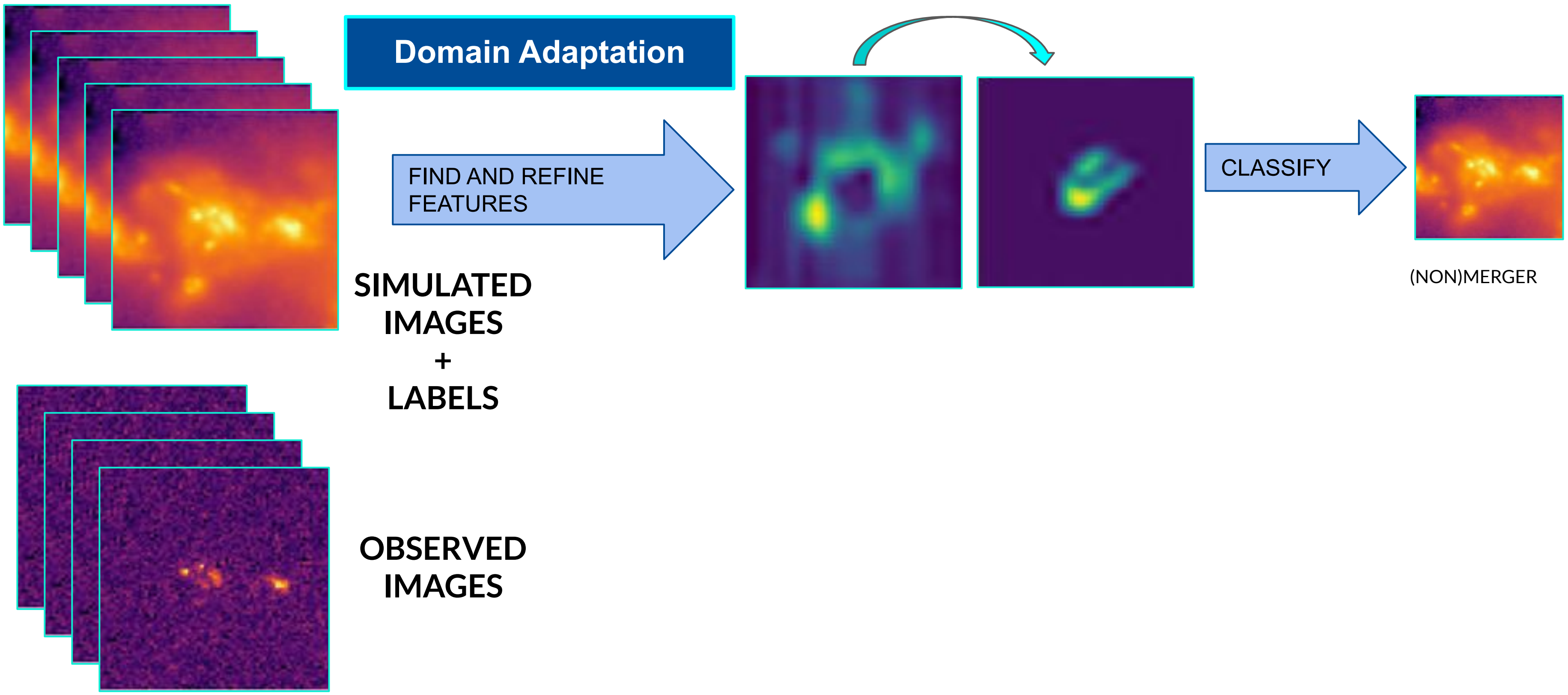






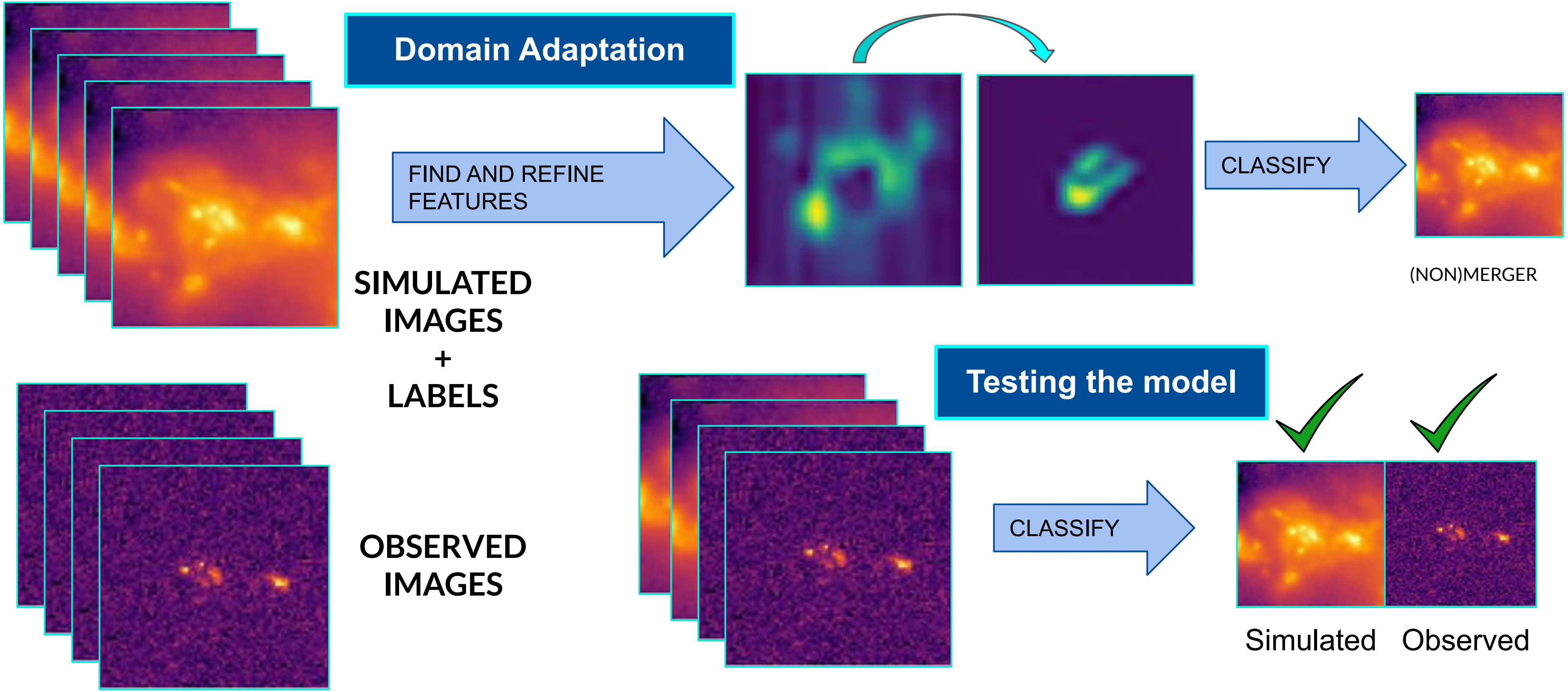


Combining Datasets

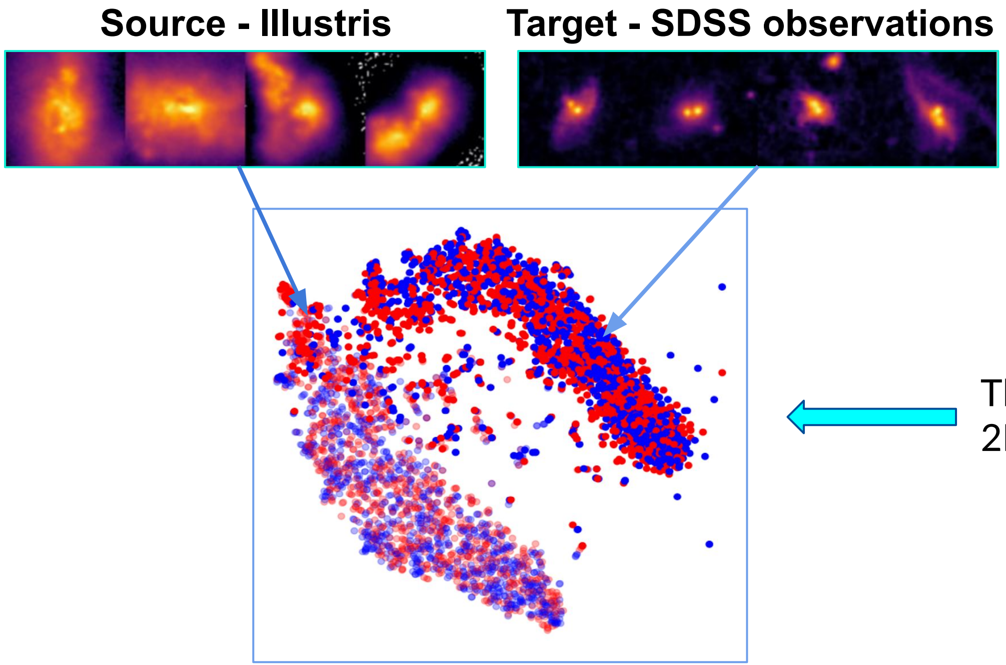


Fermilab

Combining Datasets



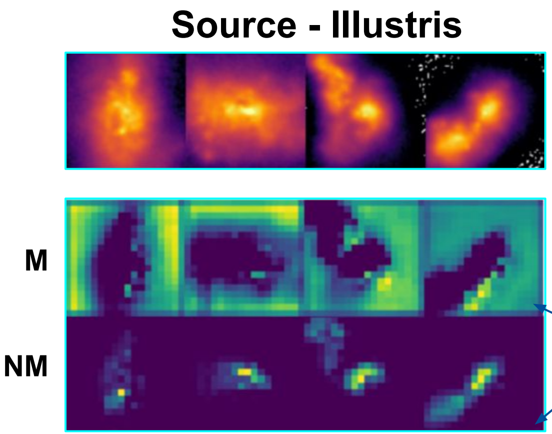
Combining Datasets



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

Ćiprijanović et al. 2020b.
Ćiprijanović et al. 2021a.

Combining Datasets



Important regions are highlighted!

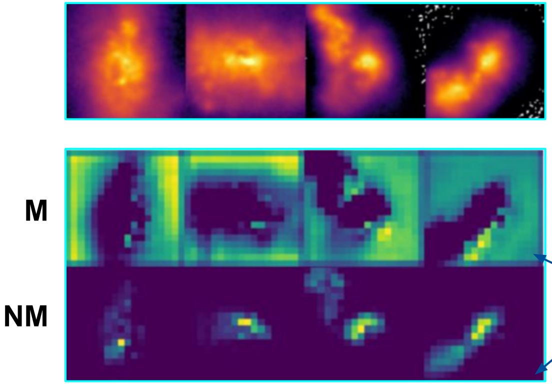
Regular Training

Ćiprijanović et al. 2020b.
Ćiprijanović et al. 202

Fermilab

Combining Datasets

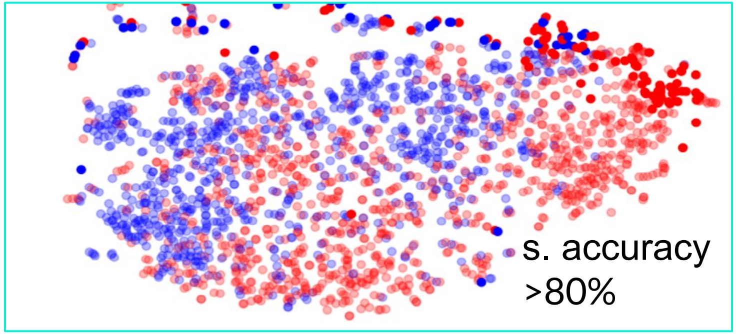
Source - Illustris



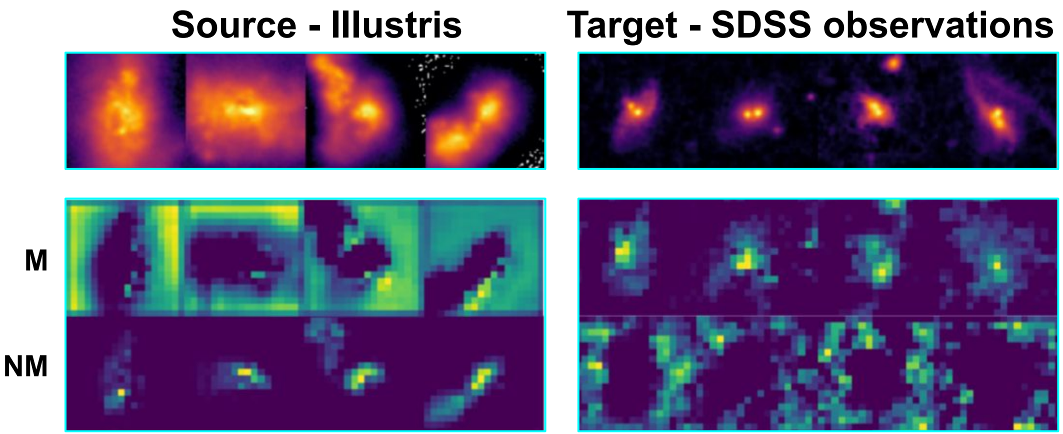
Important regions are highlighted!

Ćiprijanović et al. 2020b.
Ćiprijanović et al. 202

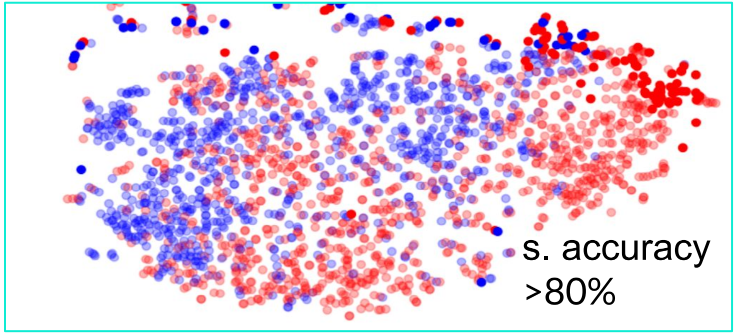
Regular Training



Combining Datasets

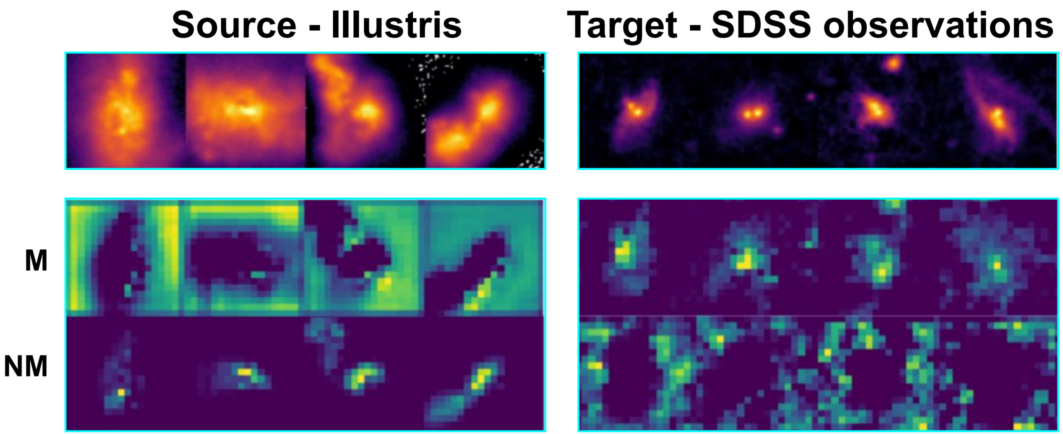


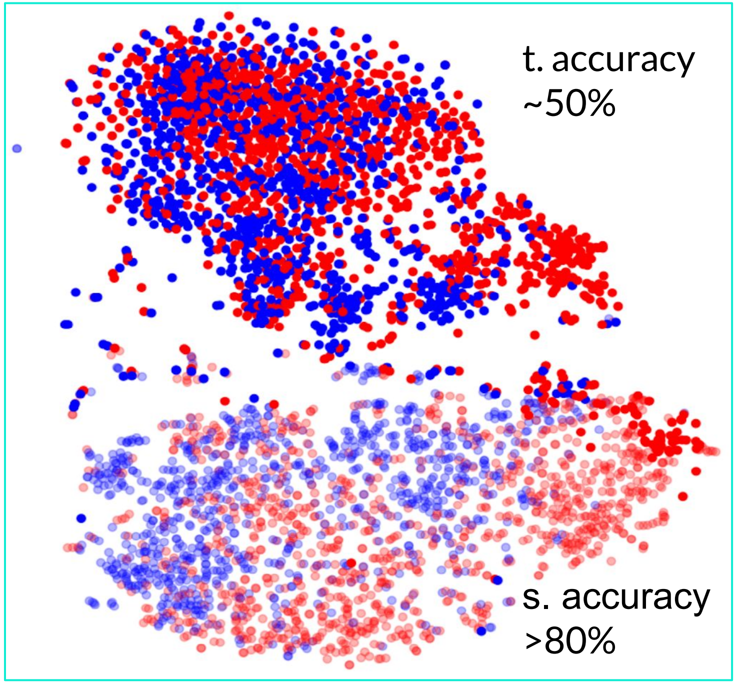
Regular Training



Ćiprijanović et al. 2020b.
Ćiprijanović et al. 202

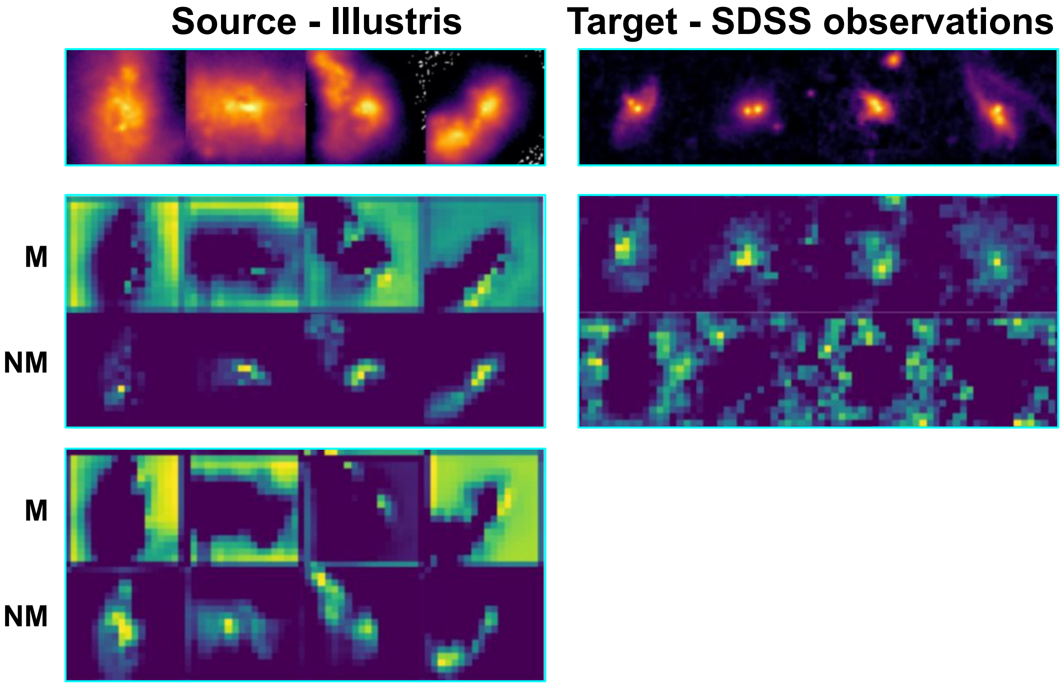
Combining Datasets





Ćiprijanović et al. 2020b.
Ćiprijanović et al. 202

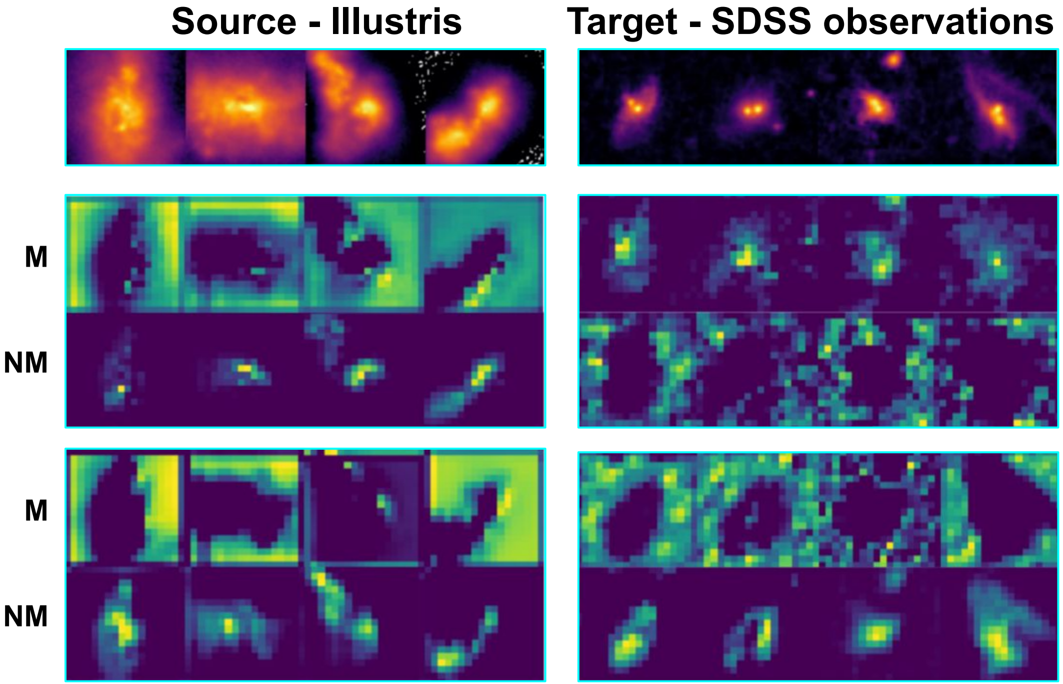
Combining Datasets



Domain Adaptation

Ćiprijanović et al. 2020b.
Ćiprijanović et al. 202

Combining Datasets



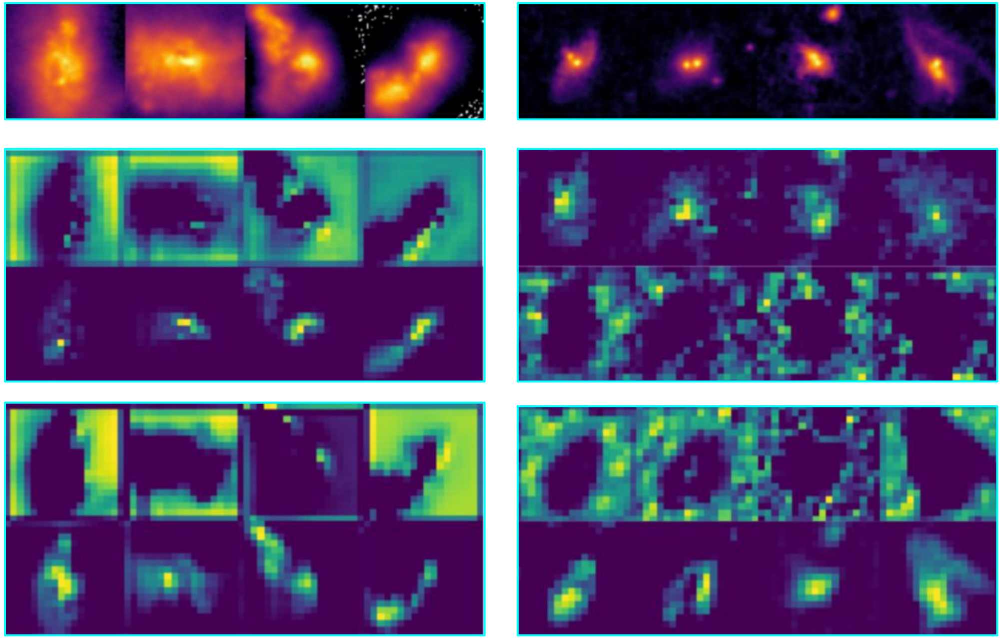
Domain Adaptation

Ćiprijanović et al. 2020b.
Ćiprijanović et al. 2021a.

Combining Datasets

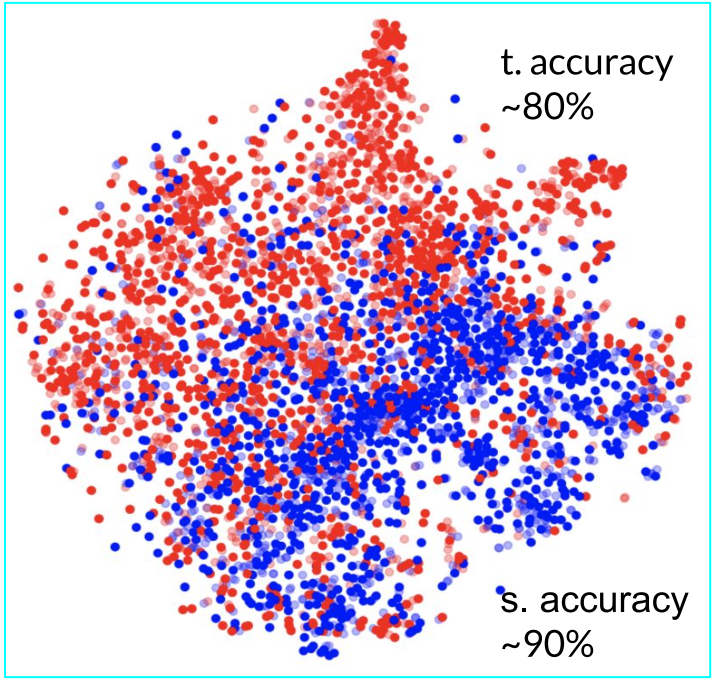
Source - Illustris **Target - SDSS observations**

M
NM
M
NM



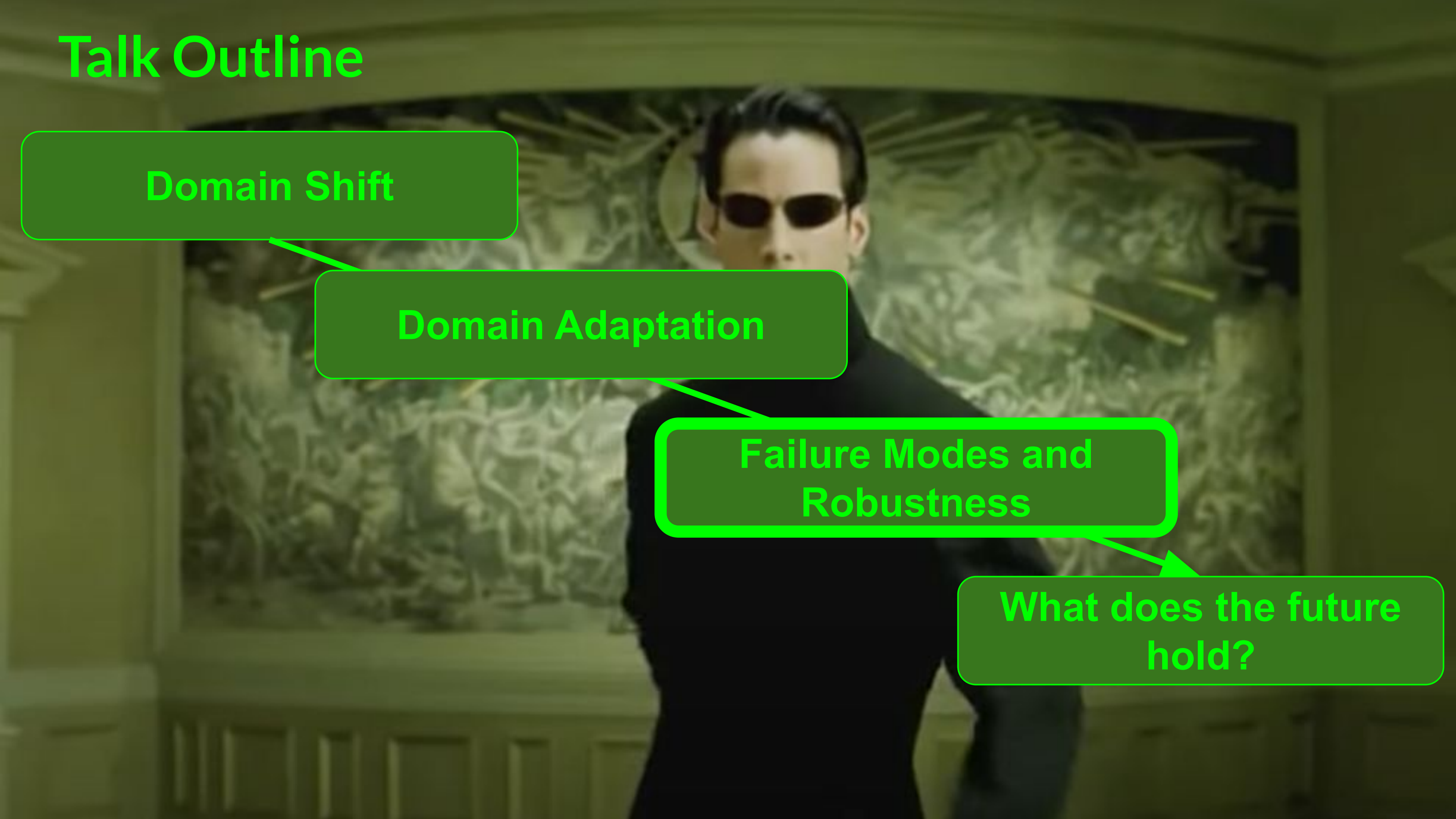
Ćiprijanović et al. 2020b.
Ćiprijanović et al. 2021a

Up to 30% increase!



Fermilab

Talk Outline



- Domain Shift
- Domain Adaptation
- Failure Modes and Robustness
- What does the future hold?

Failure modes and Model Robustness

Scientific data pipelines will introduce **inadvertent data perturbations:**

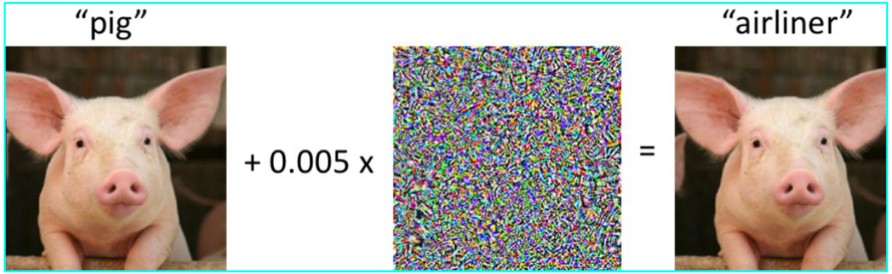
- image compression or blurring
- noise
- data pre-processing
- detector errors
- transient phenomena …

Failure modes and Model Robustness

Scientific data pipelines will introduce **inadvertent data perturbations:**

- image compression or blurring
- noise
- data pre-processing
- detector errors
- transient phenomena …

Model performance drops (sometimes catastrophically)



Targeted attack!

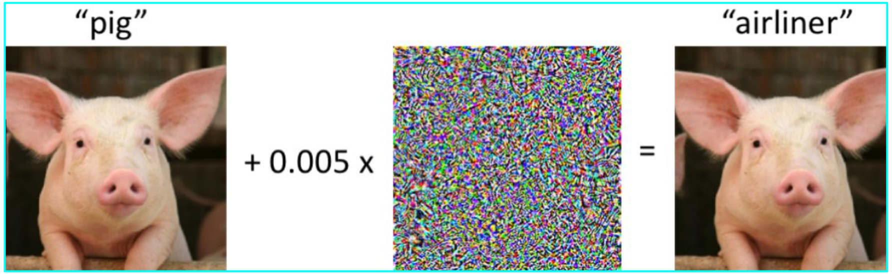
Failure modes and Model Robustness

Scientific data pipelines will introduce **inadvertent data perturbations:**

- image compression or blurring
- noise
- data pre-processing
- detector errors
- transient phenomena …

Model performance drops (sometimes catastrophically)

In science, **attacks won't be targeted**, so we also need a more general defense mechanism!



Targeted attack!

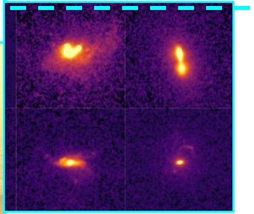
Failure modes and Model Robustness

Scientific data pipelines will introduce **inadvertent data perturbations:**

- image compression or blurring
- noise
- data pre-processing
- detector errors
- transient phenomena …

Model performance drops (sometimes catastrophically)

Ćiprijanović et al. 2021b.
Ćiprijanović et al. 2022a.



If we perturb **a single pixel,** model will classify the object **incorrectly**!

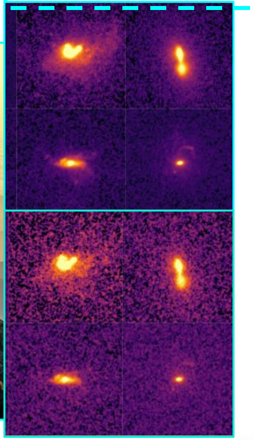
Failure modes and Model Robustness

Scientific data pipelines will introduce **inadvertent data perturbations:**

- image compression or blurring
- noise
- data pre-processing
- detector errors
- transient phenomena …

Model performance drops (sometimes catastrophically)

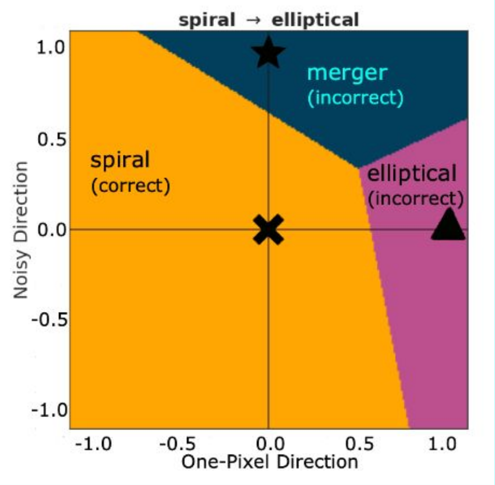
Ćiprijanović et al. 2021b.
Ćiprijanović et al. 2022a.



If we perturb **a single pixel,** model will classify the object **incorrectly**!

Old data can help!

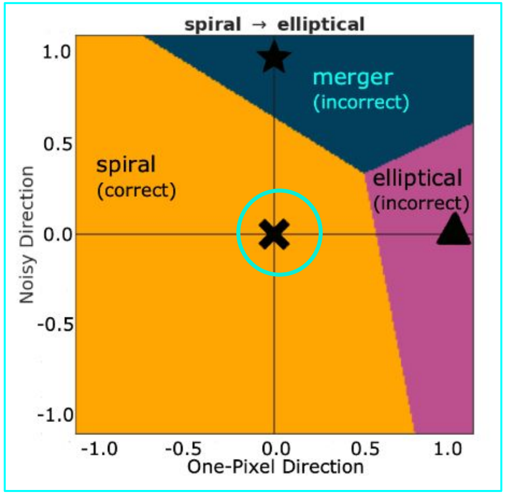
Model Robustness



Ćiprijanović et al. 2021b.
Ćiprijanović et al. 2022a.

Regular Training on Y10 data

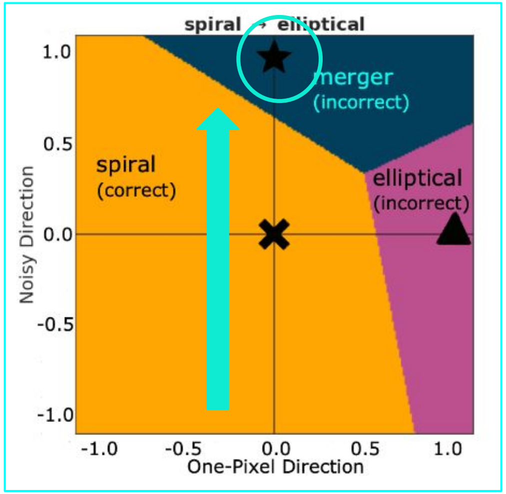
Model Robustness



Ćiprijanović et al. 2021b.
Ćiprijanović et al. 2022a.

Regular Training on Y10 data

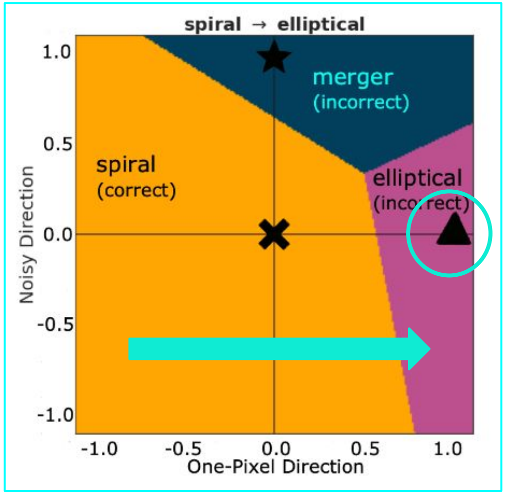
Model Robustness



Ćiprijanović et al. 2021b.
Ćiprijanović et al. 2022a.

Regular Training on Y10 data

Model Robustness

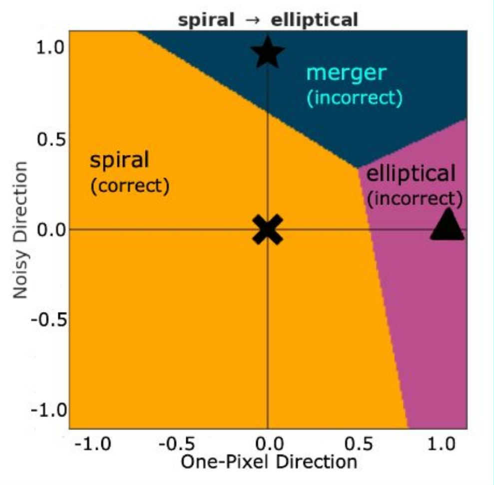


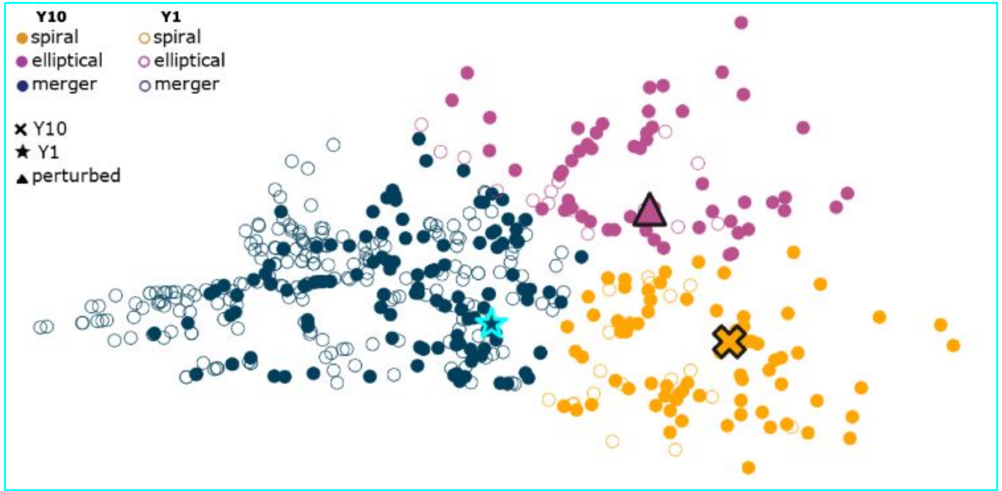
Ćiprijanović et al. 2021b.
Ćiprijanović et al. 2022a.

Regular Training on Y10 data

Fermilab

Model Robustness



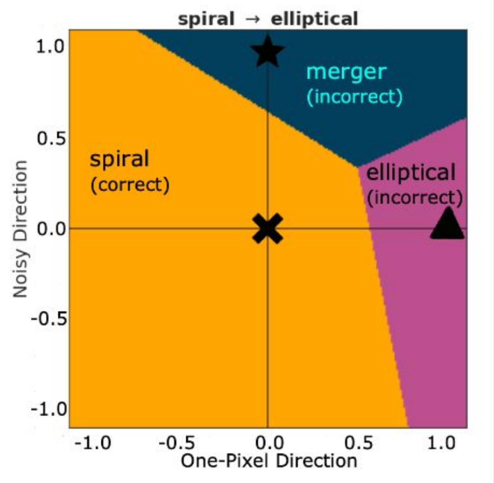


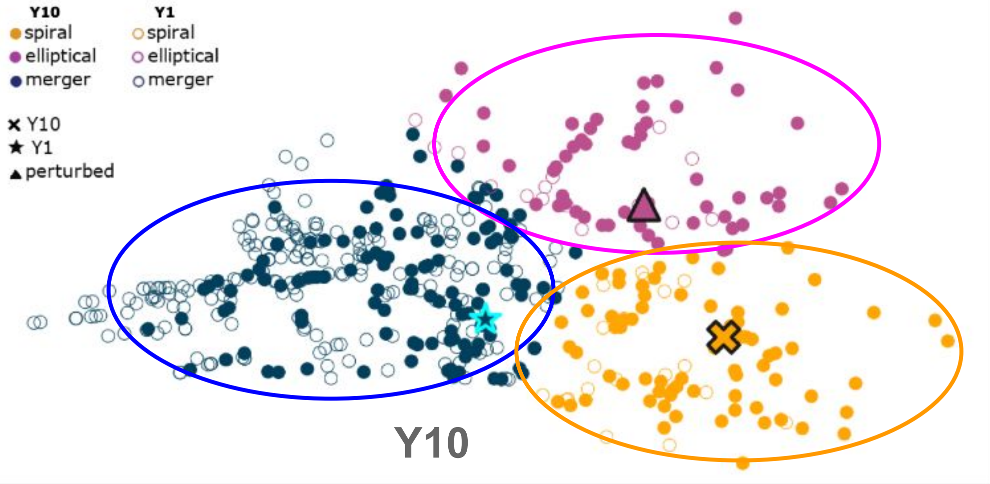
Ćiprijanović et al. 2021b.
Ćiprijanović et al. 2022a.

Regular Training on Y10 data

Fermilab

Model Robustness



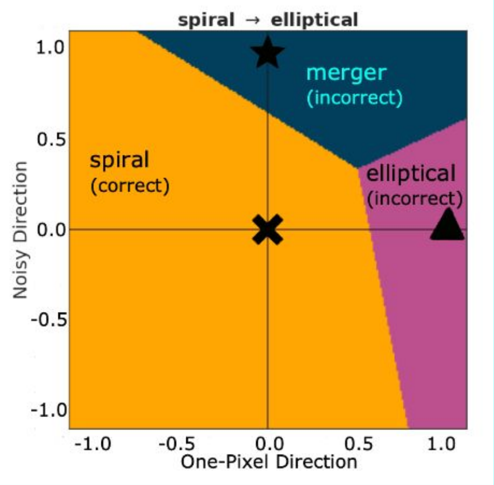


Ćiprijanović et al. 2021b.
Ćiprijanović et al. 2022a.

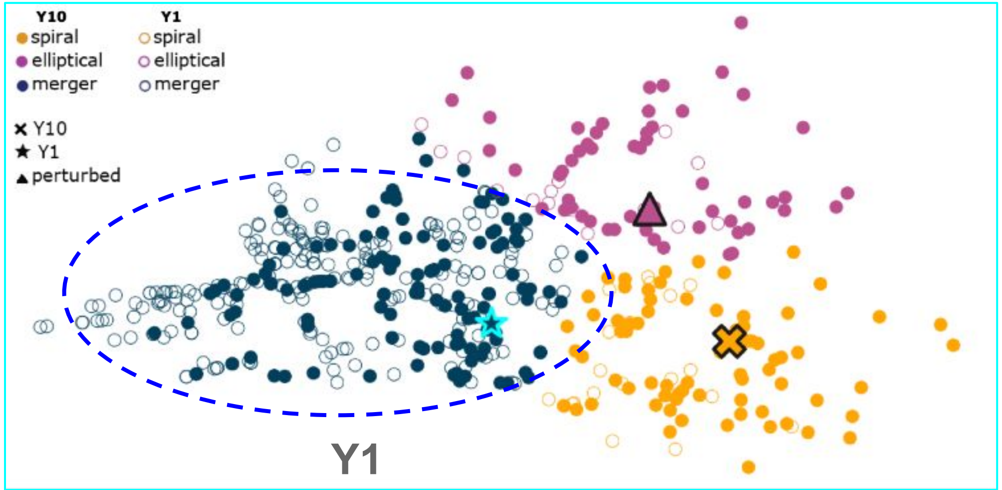
Regular Training on Y10 data

Fermilab

Model Robustness

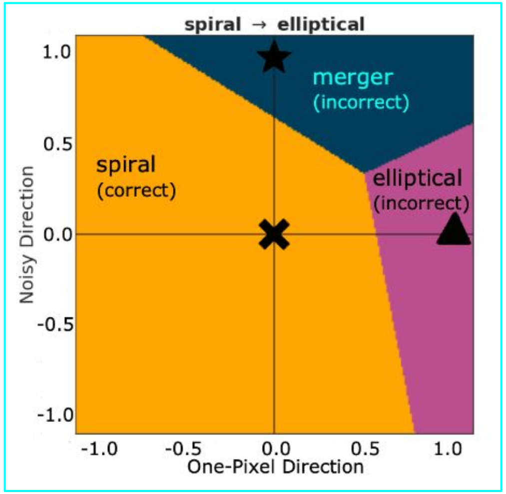


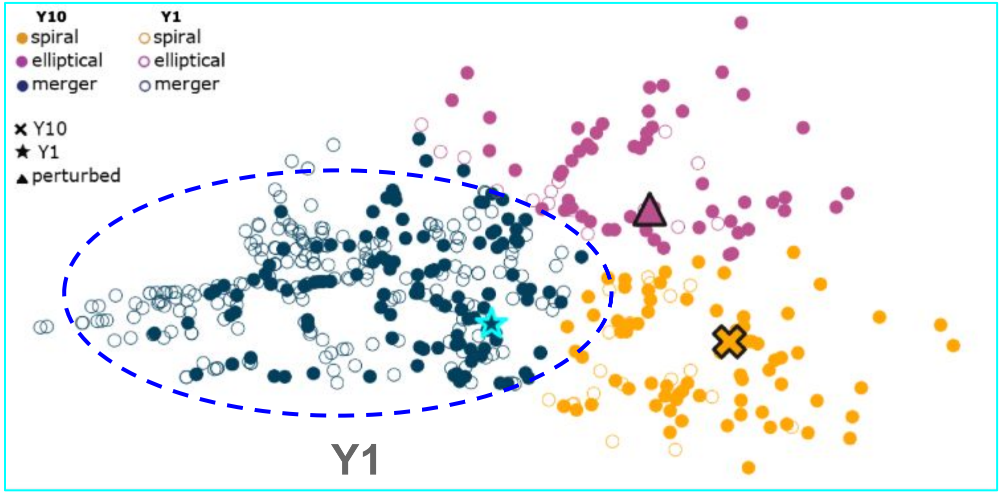
Ćiprijanović et al. 2021b.
Ćiprijanović et al. 2022a.



Regular Training on Y10 data

Model Robustness





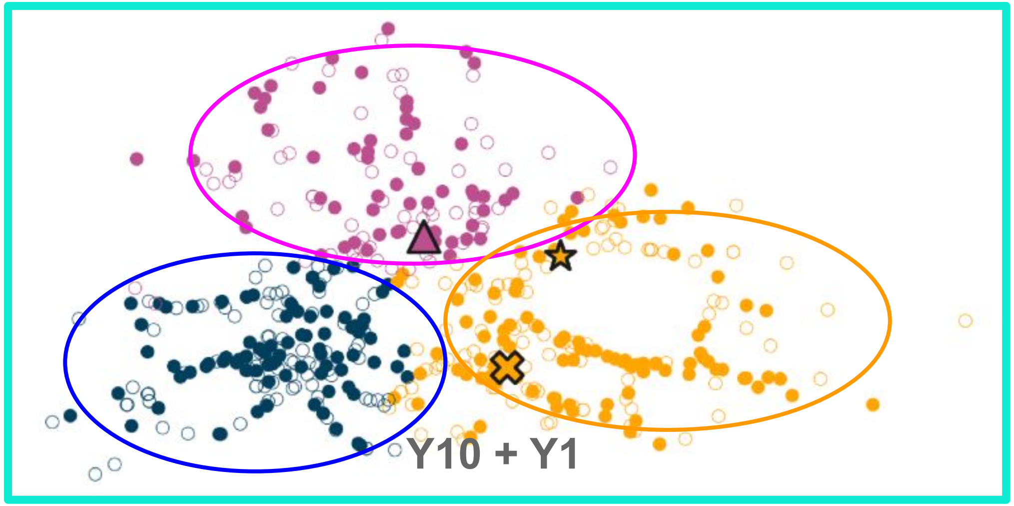
Ćiprijanović et al. 2021b.
Ćiprijanović et al. 2022a.

Regular Training on Y10 data

Domain Adaptation using Y1 data

Model Robustness

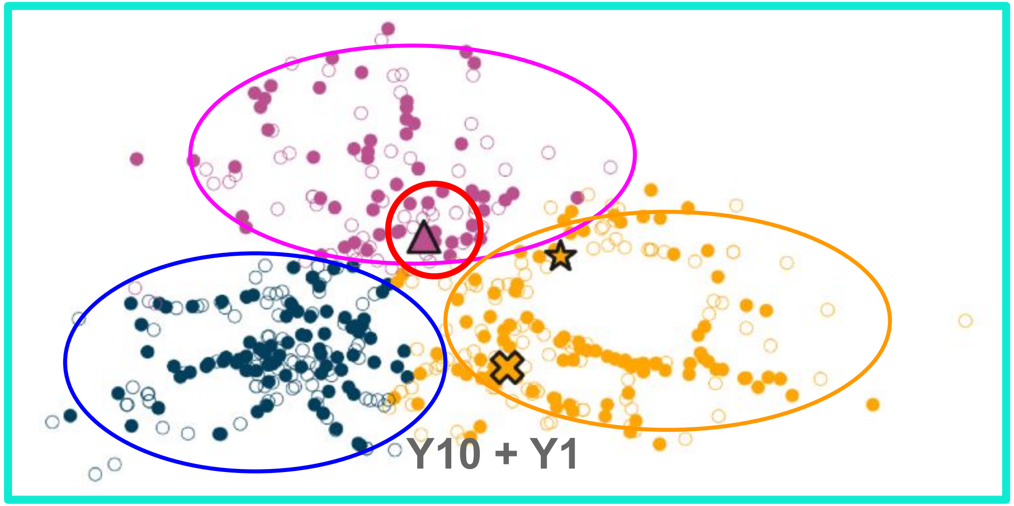
- Accuracy on both datasets increases (up to 23%)!



Regular Training on Y10 data → Domain Adaptation using Y1 data

Model Robustness

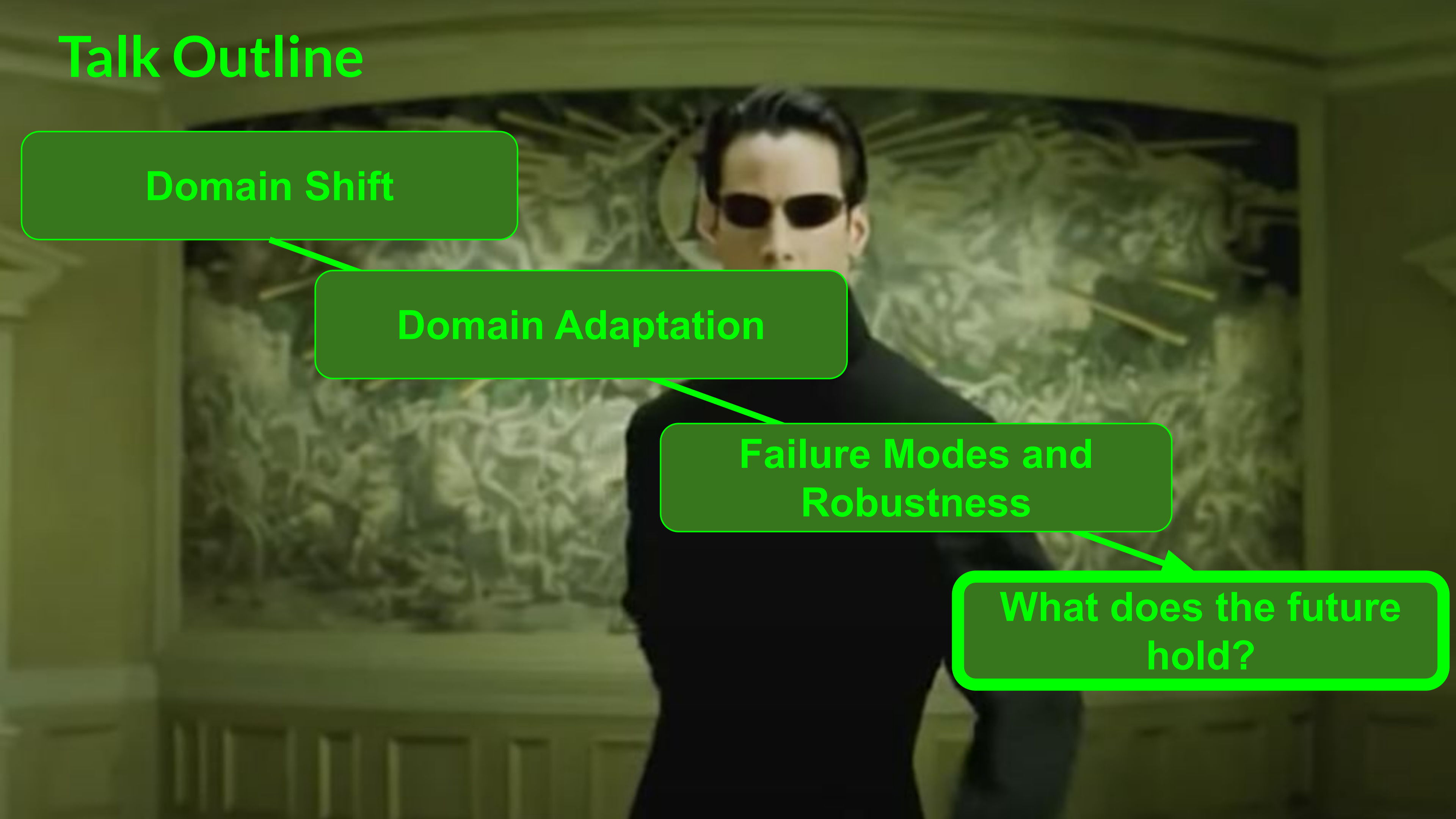
- Accuracy on both datasets increases (up to 23%)!
- Distance to the wrong class increases ~2.3!
- **Robustness to inadvertent perturbations increases!**



Regular Training on Y10 data

Domain Adaptation using Y1 data

Talk Outline



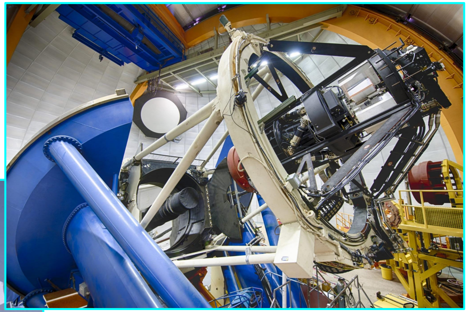
- Domain Shift
- Domain Adaptation
- Failure Modes and Robustness
- **What does the future hold?**

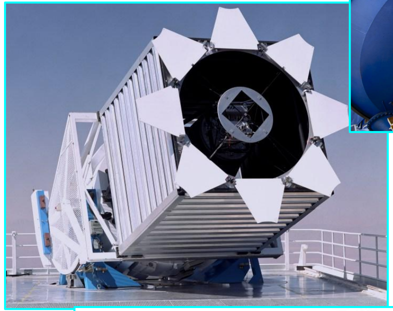
Multi-dataset inference

Bridging between observations is much harder!
We need general and flexible algorithms.

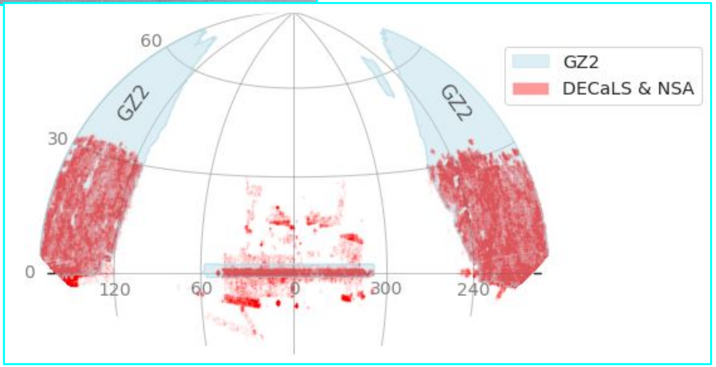
The gap between observational datasets is much larger:

- Noise, PSF
- Pixel scale
- Depth of the survey
- Magnitude limit
- Perhaps different filters
- Different data distributions….





SDSS to DECaLS?



Fermilab

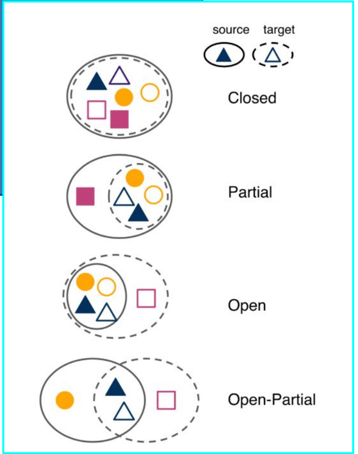
Multi-dataset inference

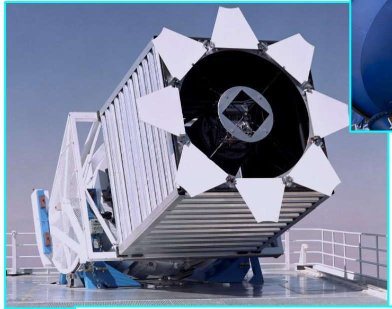
Bridging between observations is much harder!
We need general and flexible algorithms.

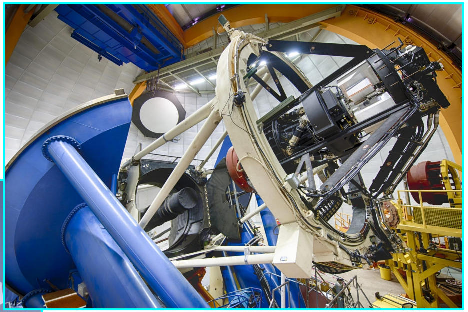
The gap between observational datasets is much larger:

- Noise, PSF
- Pixel scale
- Depth of the survey
- Magnitude limit
- Perhaps different filters
- Different data distributions….

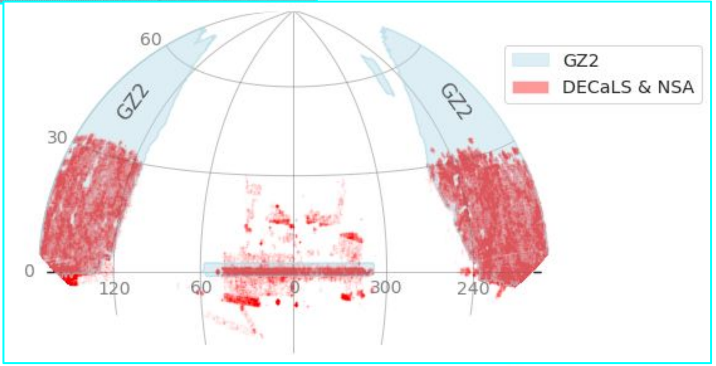
How do we build something flexible enough to handle any kind of data distribution shifts?







SDSS to DECaLS?



Universal Domain Adaptation (DeepAstroUDA)

preliminary

Ćiprijanović et al. 2022.
Ćiprijanović et al. 2023 (soon)

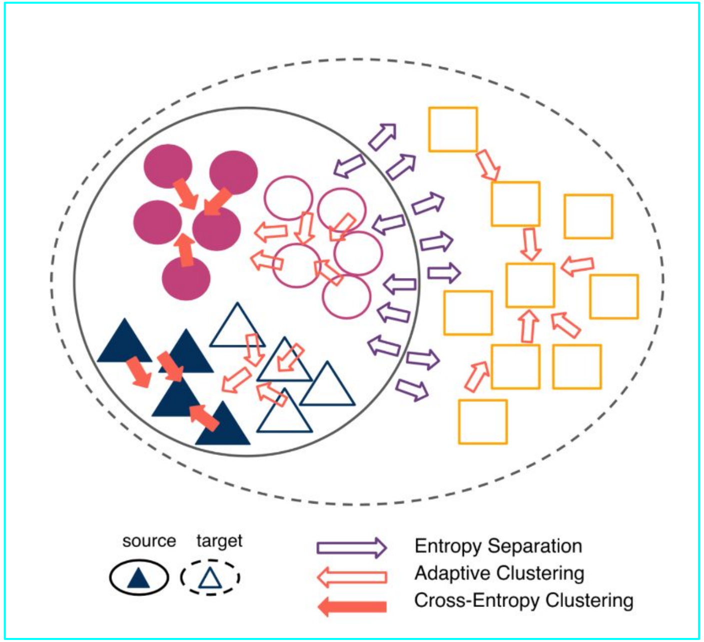
Classification of known classes

+

Clustering of similar known and unknown samples

+

Separation of different (anomalous) unknown samples



Universal Domain Adaptation (DeepAstroUDA)

preliminary

Ćiprijanović et al. 2022.
Ćiprijanović et al. 2023 (soon)

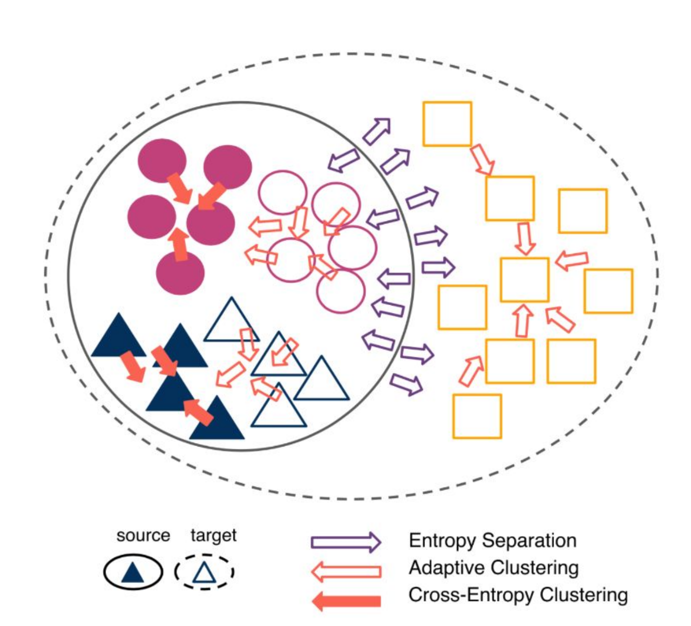
Classification of known classes

+

Clustering of similar known and unknown samples

+

Separation of different (anomalous) unknown samples

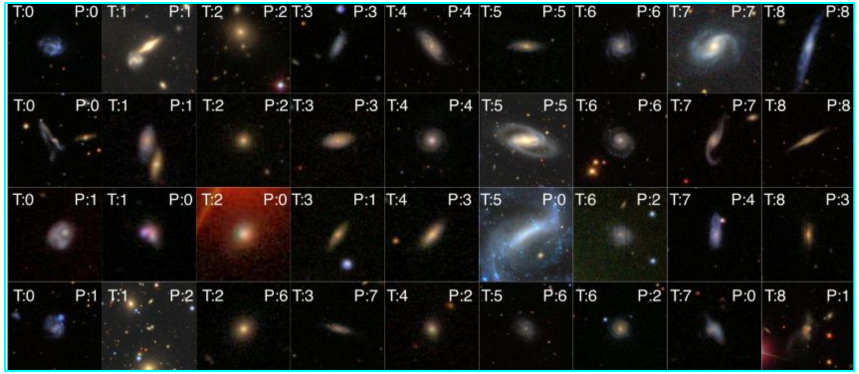


Via self-supervision for more flexibility!

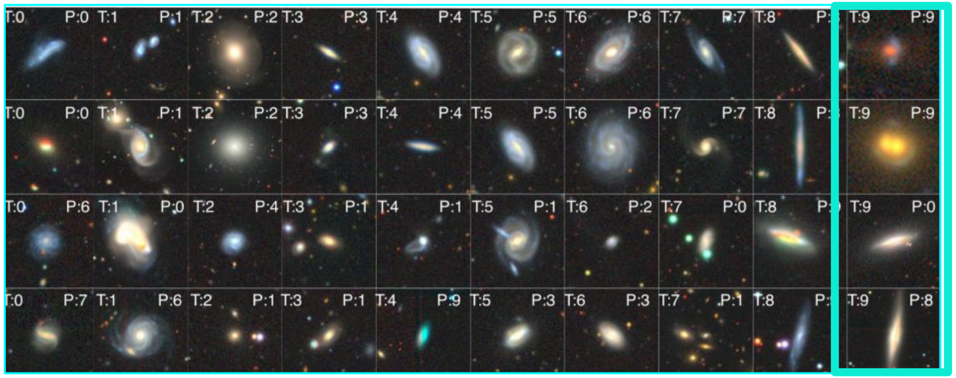
Fermilab

Universal Domain Adaptation (DeepAstroUDA)

SDSS



DECaLS



Class labels are from Galaxy Zoo 2 & 3 (crowdsourcing labels ~10^5 volunteers).

preliminary

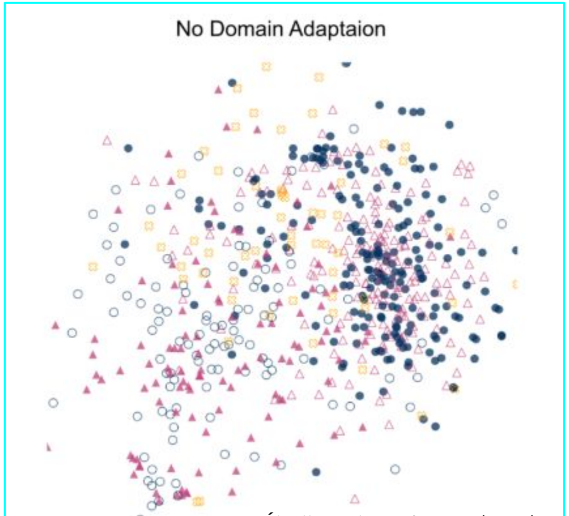
Known classes:

Disturbed (0)
Merging (1)
Round smooth (2)
Cigar shaped smooth (3)
Barred spiral (4)
Unbarred tight spiral (5),
Unbarred loose spiral (6)
Edge-on without bulge (7),
Edge-on with bulge (8),

Unknown anomaly class (only in DECaLS):

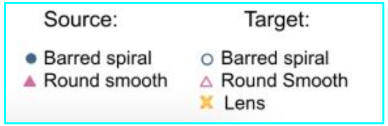
Strong gravitational lens (9)

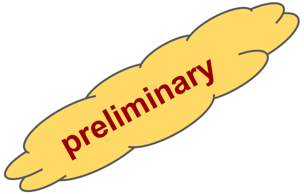
Universal Domain Adaptation (DeepAstroUDA)



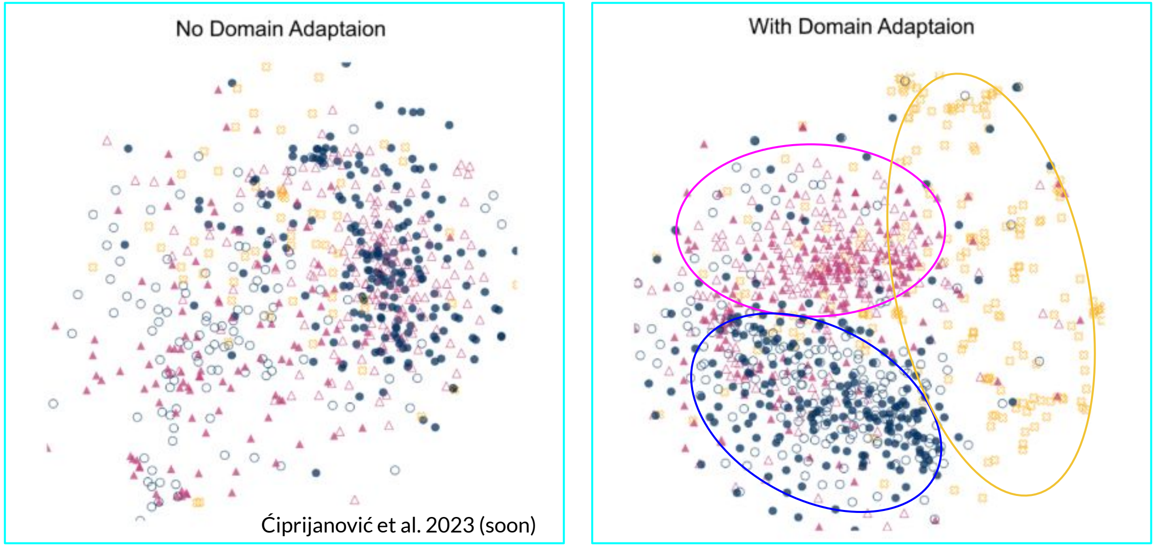
Ćiprijanović et al. 2023 (soon)

Classes are mixed!





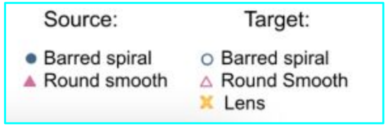
Universal Domain Adaptation (DeepAstroUDA)



Ćiprijanović et al. 2023 (soon)

Classes are mixed!

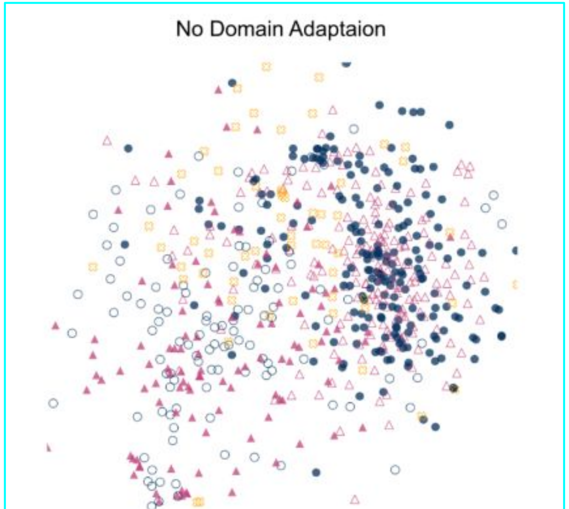
preliminary



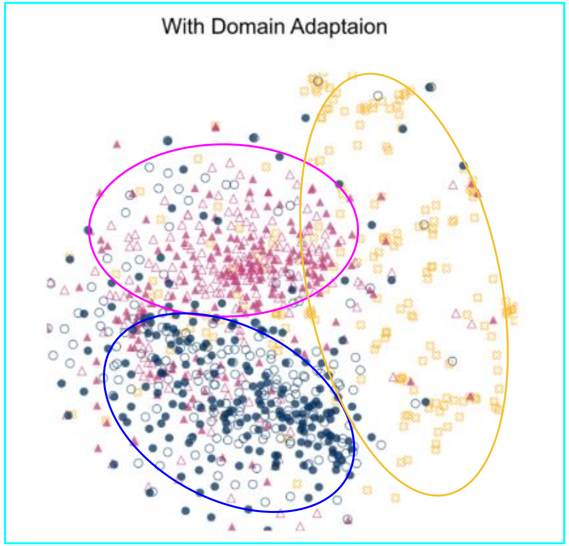
Known classes overlap,
unknown is pushed to the side.

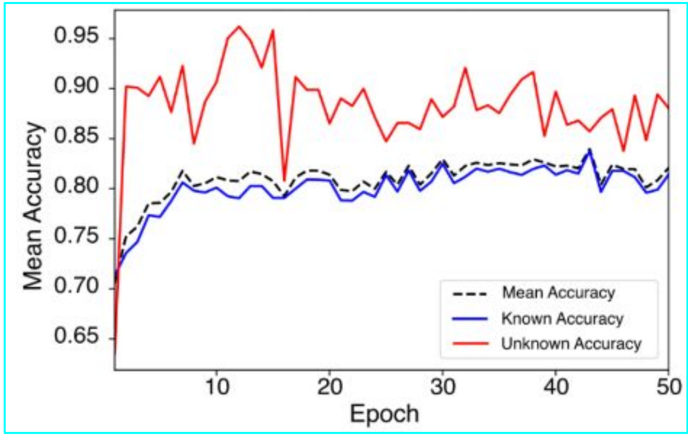
Fermilab

Universal Domain Adaptation (DeepAstroUDA)



Ćiprijanović et al. 2023 (soon)





Classes are mixed!

preliminary

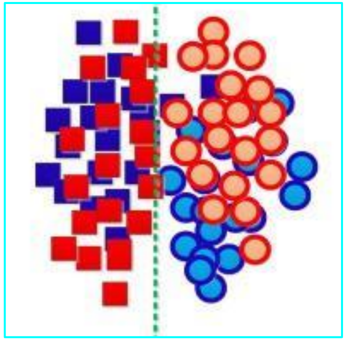
Source:	Target:
● Barred spiral	○ Barred spiral
▲ Round smooth	△ Round Smooth
	✕ Lens

Known classes overlap,
unknown is pushed to the side.

Fermilab

Questions to think about

- Is **the domain shift a big problem in astrophysics/cosmology and how do we solve it**?
    - transfer learning (labeled data), domain adaptation (unlabeled data), or something completely new?



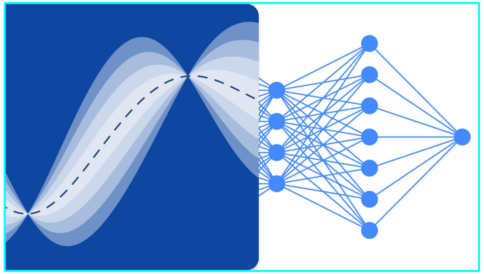
Questions to think about

- Is **the domain shift a big problem in astrophysics/cosmology and how do we solve it**?
    - transfer learning (labeled data), domain adaptation (unlabeled data), or something completely new?
- How do we **fight against unknown unknowns causing the domain shift**?

[KK] : Knowledge Known Knowns	[KU] : Awareness Known Unknowns
[UK] : Bias Unknown Knowns	[UU] : Ignorance Unknown Unknowns

Questions to think about

- Is **the domain shift a big problem in astrophysics/cosmology and how do we solve it**?
    - transfer learning (labeled data), domain adaptation (unlabeled data), or something completely new?
- How do we **fight against unknown unknowns causing the domain shift**?
- **Uncertainty quantification in AI and do we trust it**?
    - BNNs, Dropout…?

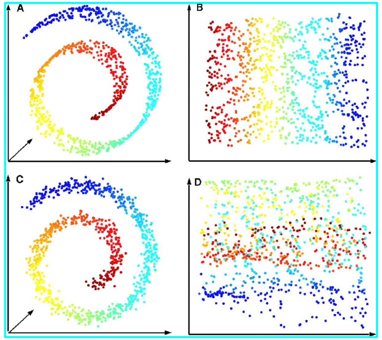


Questions to think about

- Is **the domain shift a big problem in astrophysics/cosmology and how do we solve it**?
  - transfer learning (labeled data), domain adaptation (unlabeled data), or something completely new?
- How do we **fight against unknown unknowns causing the domain shift**?
- **Uncertainty quantification in AI and do we trust it**?
  - BNNs, Dropout…?
- Do we need to standardize model **testing and reporting results**?
  - Does our model work on data different from our initial test dataset

Questions to think about

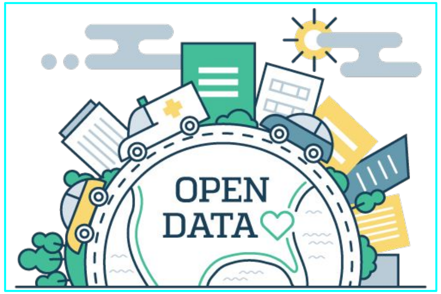
- Is **the domain shift a big problem in astrophysics/cosmology and how do we solve it**?
  - transfer learning (labeled data), domain adaptation (unlabeled data), or something completely new?
- How do we **fight against unknown unknowns causing the domain shift**?
- **Uncertainty quantification in AI and do we trust it**?
  - BNNs, Dropout…?
- Do we need to standardize model **testing and reporting results**?
  - Does our model work on data different from our initial test dataset?
- How to make our **AI models more transparent**?
  - Model interpretability, visualizations, ablation studies



Fermilab

Questions to think about

- Is **the domain shift a big problem in astrophysics/cosmology and how do we solve it**?
    - transfer learning (labeled data), domain adaptation (unlabeled data), or something completely new?
- How do we **fight against unknown unknowns causing the domain shift**?
- **Uncertainty quantification in AI and do we trust it**?
    - BNNs, Dropout…?
- Do we need to standardize model **testing and reporting results**?
    - Does our model work on data different from our initial test dataset?
- How to make our **AI models more transparent**?
    - Model interpretability, visualizations, ablation studies
- How to make sure our results are **reproducible**?
    - Open data and code, setting community standards, astro benchmarks



THANK YOU!

Aleksandra Ćiprijanović
(she/her/hers)

Fermilab, DSSL
aleksand@fnal.gov

KITP
February, 2023