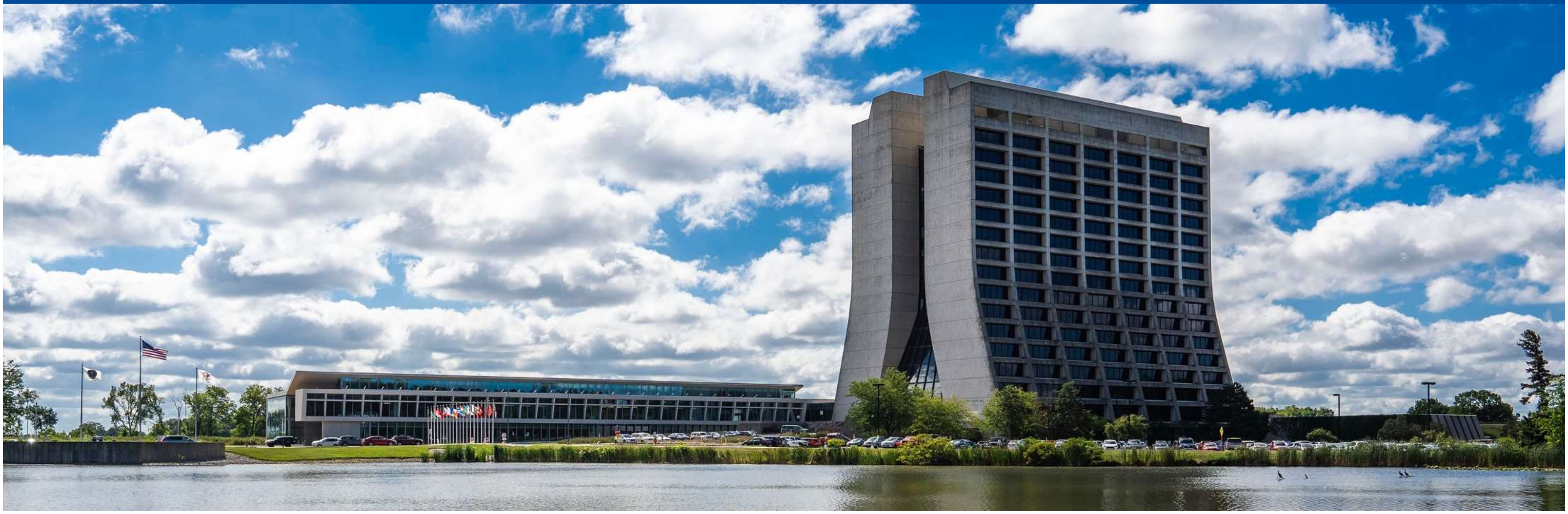
FERMILAB-SLIDES-23-392-STUDENT



Simulation Based Inference with Domain Adaptation for Strong Gravitational Lensing

Marcos E. Tamargo-Arizmendi

Supervisors: Aleksandra Ćiprijanović and Brian Nord

SULI Final Presentation

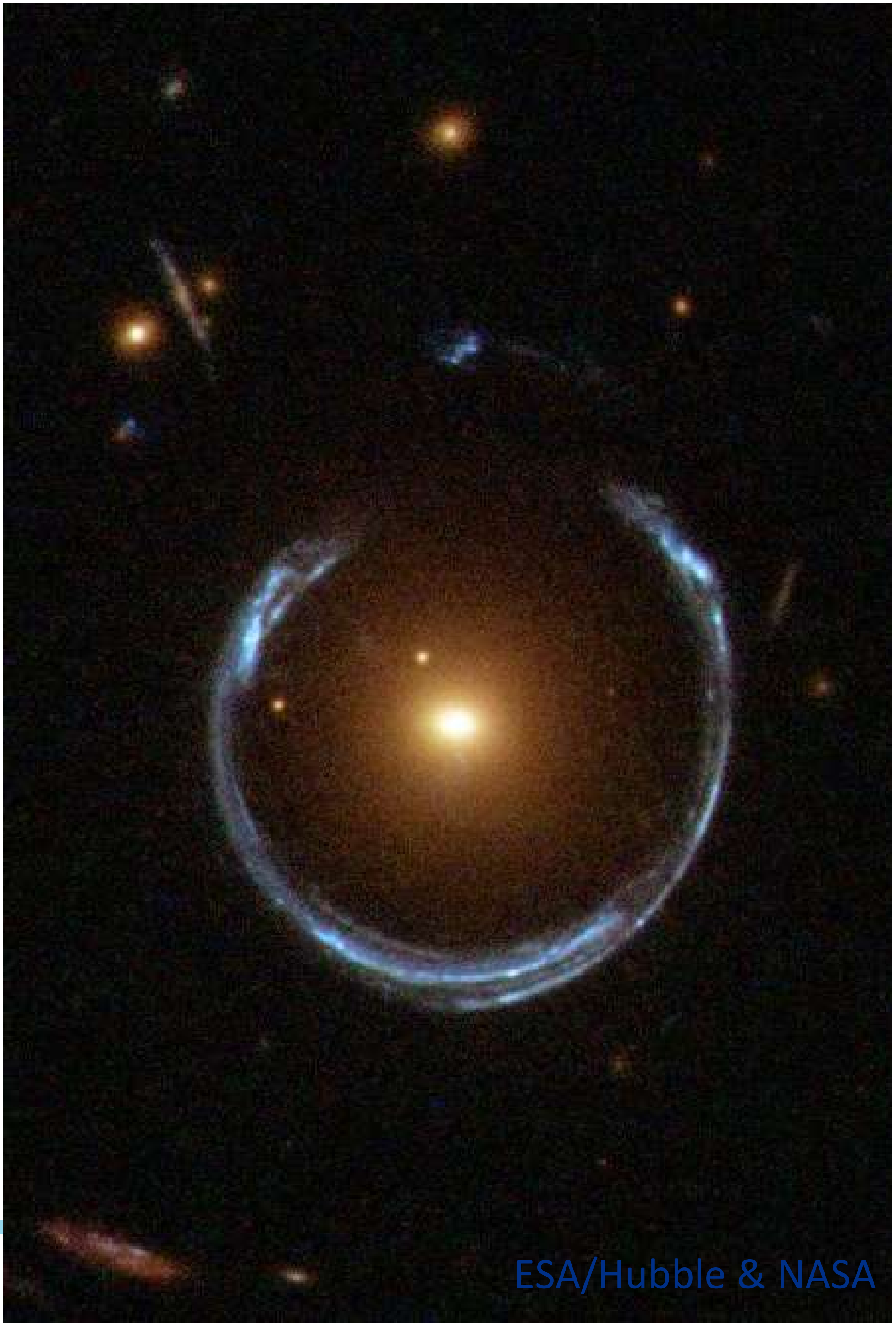
6 December 2023

This manuscript has been authored by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics.

This work was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI).

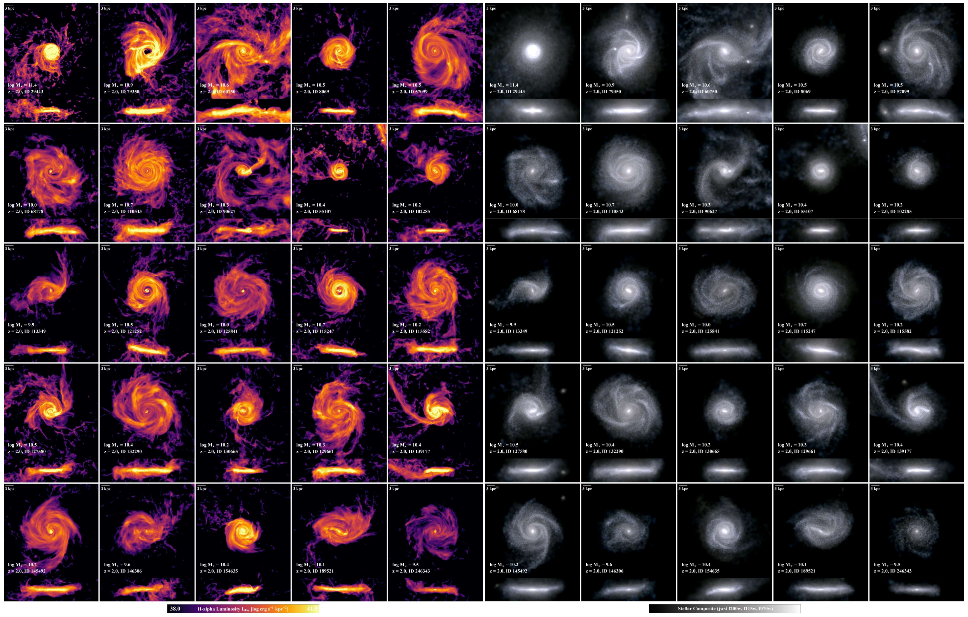
Gravitational Lensing

- Bending of light by gravity
- Cosmological probes
- Powerful telescopes
- Constrain mass distributions
- We keep discovering more and more – How do we keep up?



Parameter Inference with Intractable Likelihoods

- Simulations are powerful research tools
  - Predictive power
  - Inform Observation
  - Test theory
- Poorly suited for parameter inference
  - Intractable likelihoods
  - Most methods computationally expensive



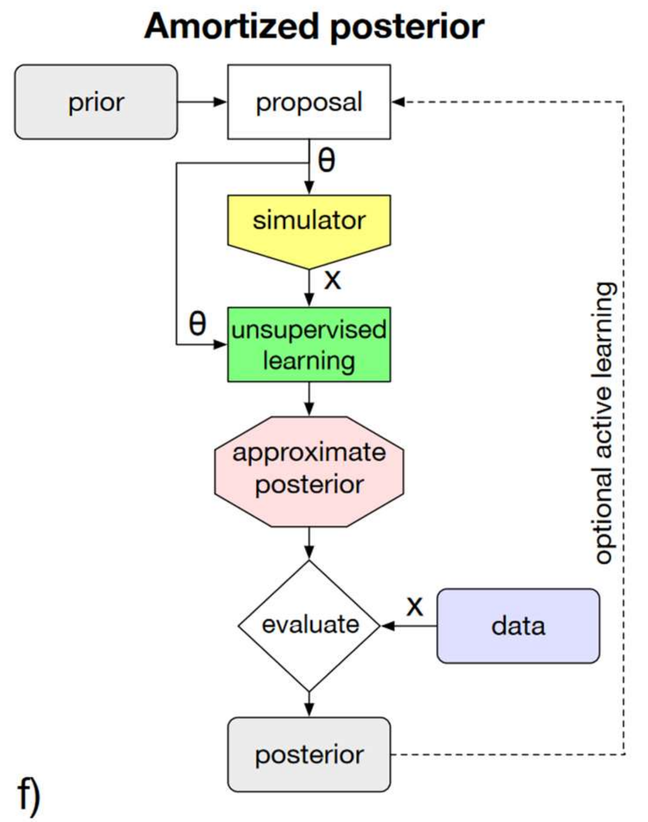
(Phillepich+, 2019)

Fermilab

Simulation Based Inference

Learn the posterior from sims
→ infer from real data

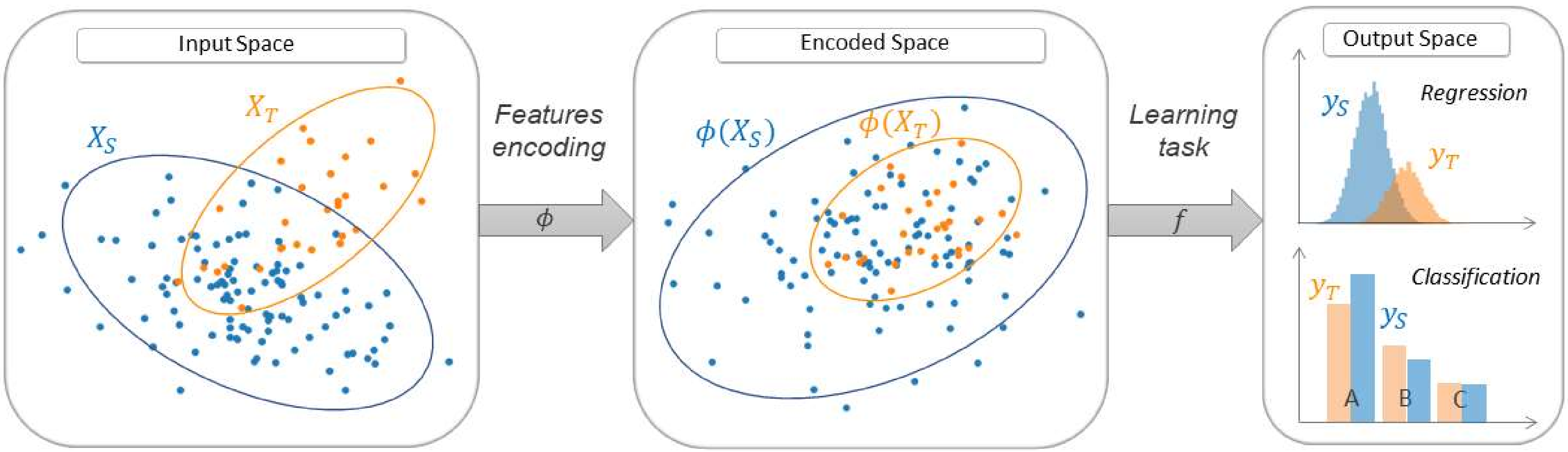
- Leverage ML
  - Prior → simulate data → train network →perform inference on real data
- Directly target posterior
- Problem! → how do you go from simulated data to real data?



(Cranmer+, 2020)

Domain Adaptation

- Family of techniques to deal with domain shift
- Source vs Target Domains
- Feature based domain adaptation
  - Align feature representations of the data



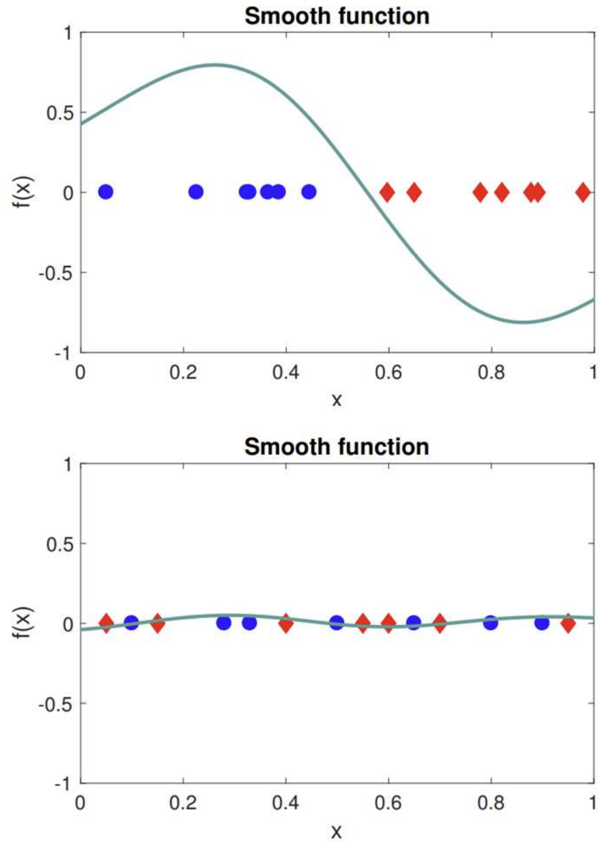
(de Mathelin, 2023)

Fermilab

Maximum Mean Discrepancy

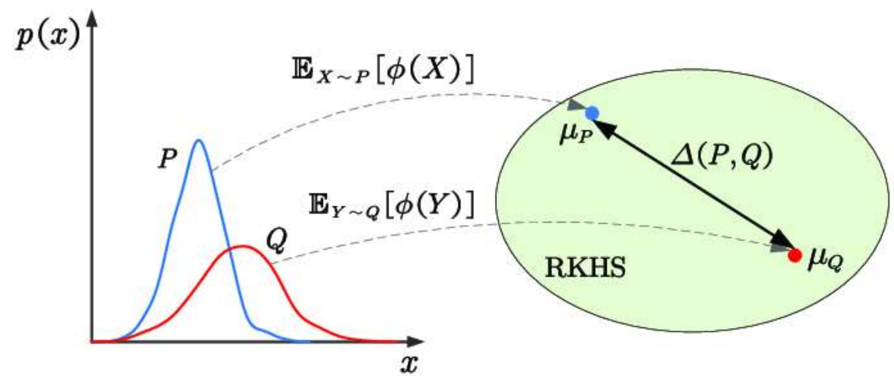
Find a "well behaved function" $f(x)$ to maximize

$$\mathbf{E}_P f(X) - \mathbf{E}_Q f(Y)$$



(Gretton, 2018)

- Describe data using smooth function
- Find distance between distribution embeddings in RKHS

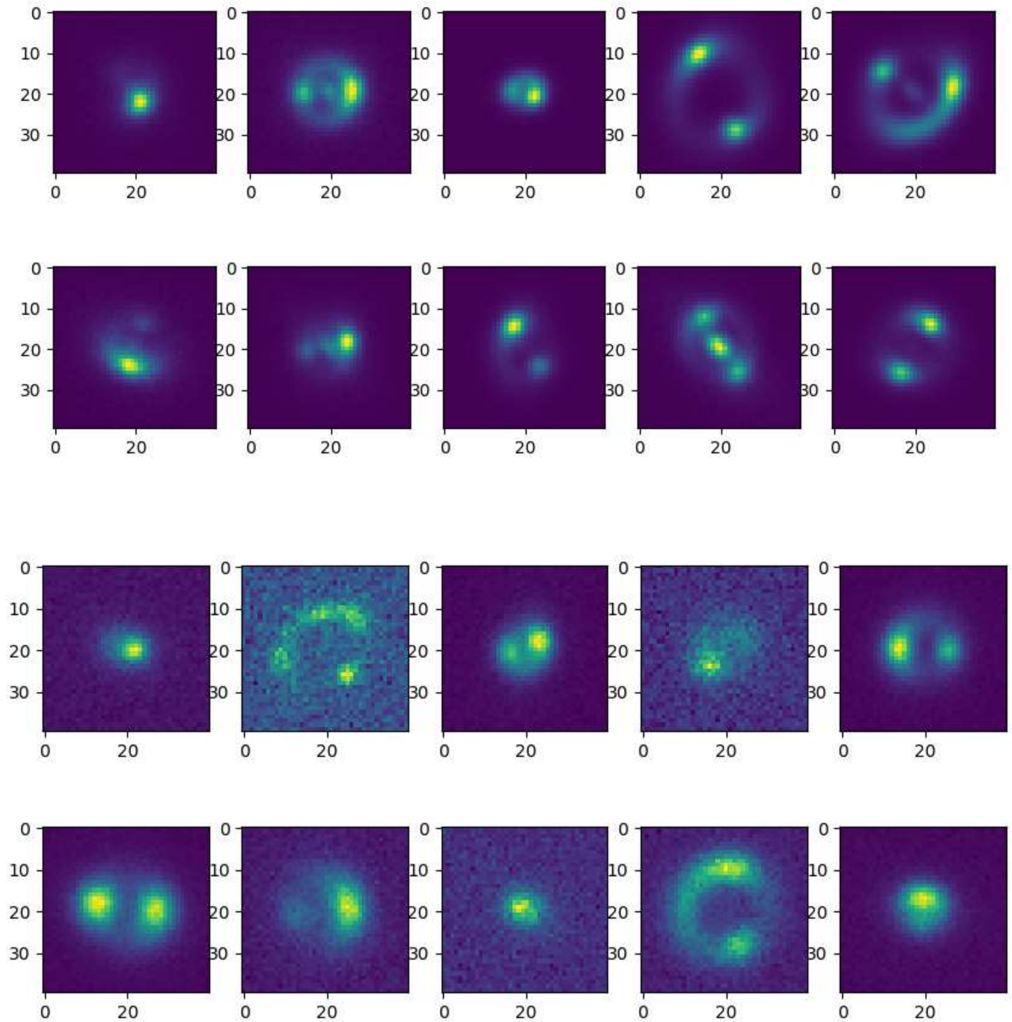


(Jun Shi+, 2011)

Strong Lensing Simulations

400k SIE lenses using Deeplenstronomy

- 200k source: low noise survey conditions
- 200k target: DES survey conditions
- 5 parameter fitting
  - θ_E
  - e_1, e_2
  - x, y



Results

- Two types of performance checks
  - Predictive performance
    - How well does the model recover the true parameter values
    - Metrics
      - χ^2 , R^2, MSE
  - Uncertainty calibration
    - How well does the model estimate prediction uncertainties
    - Metrics – Rank statistics
      - KS-test p-values, C2ST – DAP, C2ST – Ranks

Performance Metrics: No DA vs MMD

- Marginal improvement
- From χ^2 true values fall within uncertainty

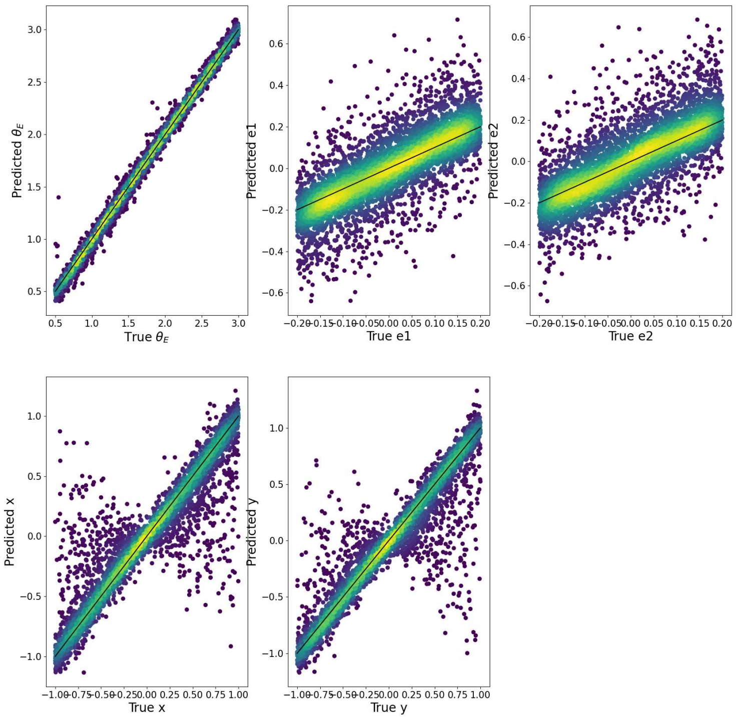
metric	No DA	MMD; $\lambda = 0.02$
χ^2	1.164	1.084
R^2	-0.1106	0.04701
MSE	0.04504	0.04160

- Ranks, on average not uniformly distributed
- From C2ST – deviations not large

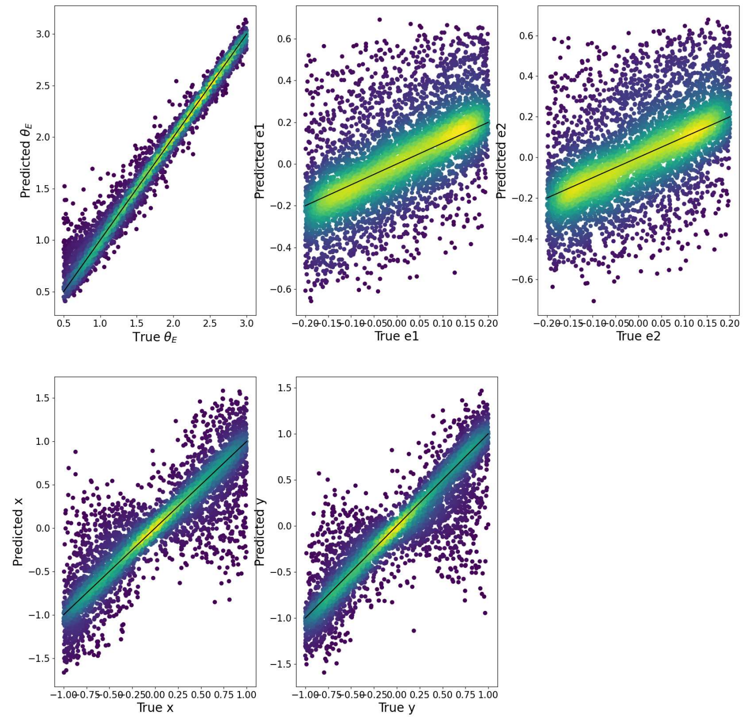
metric	No DA	MMD; $\lambda = 0.02$
KS p-val	1.099×10^{-4}	7.840×10^{-3}
C2ST - Ranks	0.5779	0.5786
C2ST - DAP	0.5668	0.5621

Parameter Inference- No DA

Source



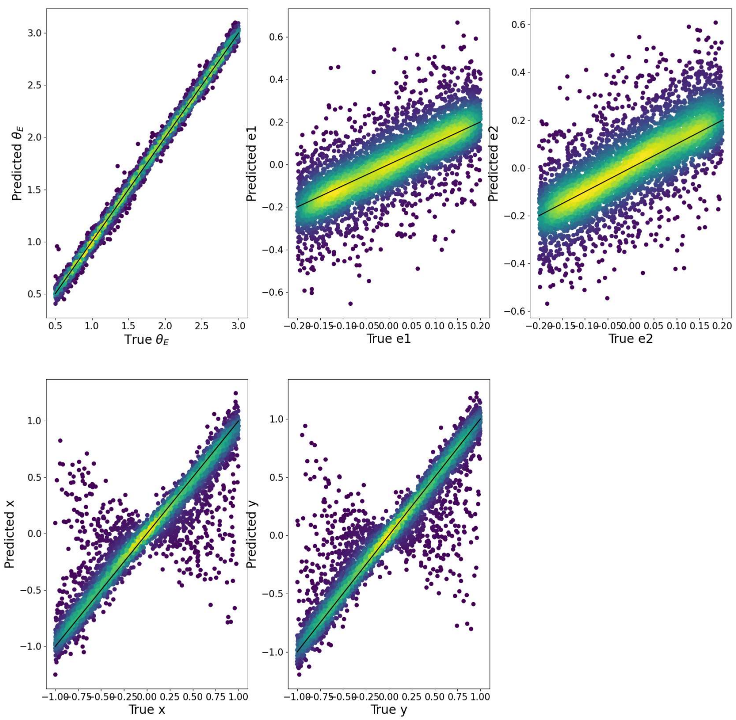
Target



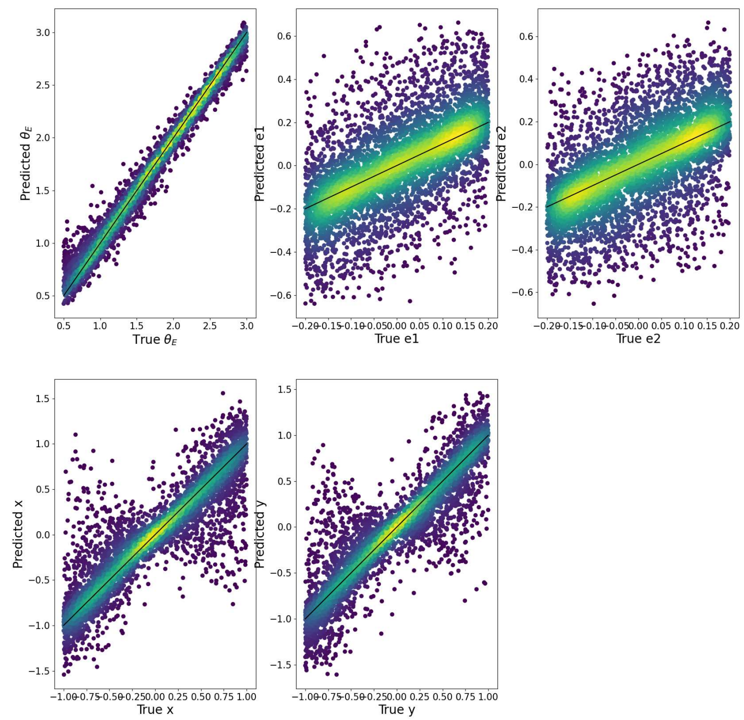
- Similar performance on source and target

Parameter Inference- MMD, λ = 0.02

Source



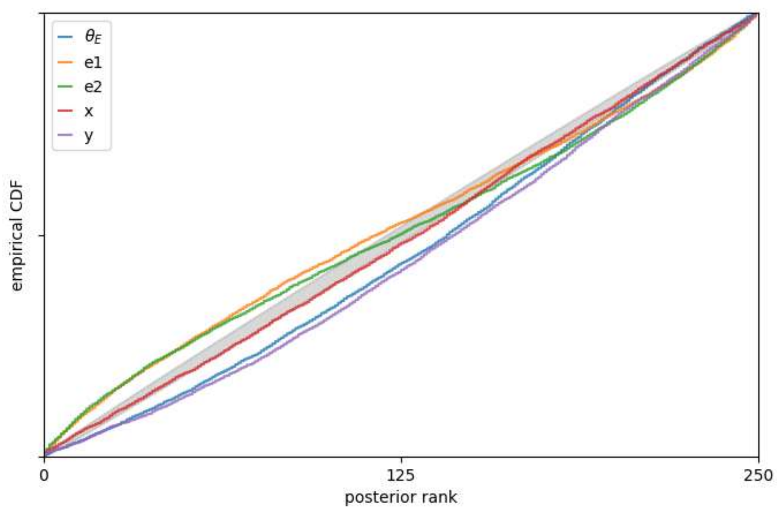
Target



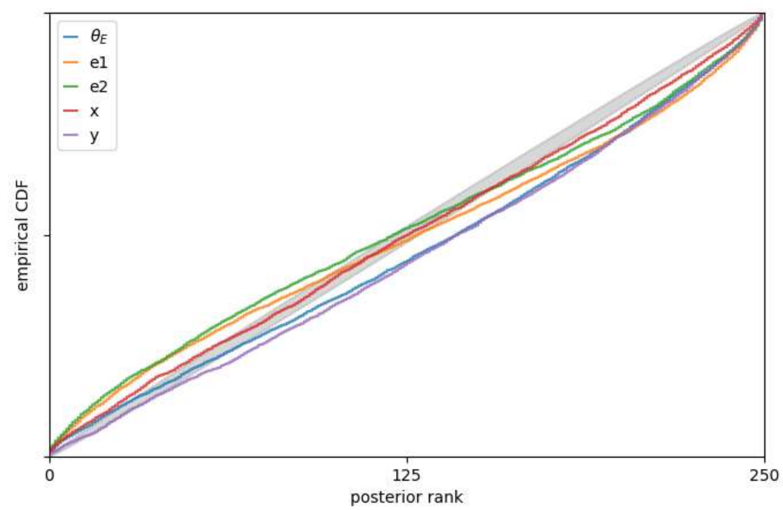
- Slight improvement but not equally across all parameters

Ranks CDF - No DA

Source



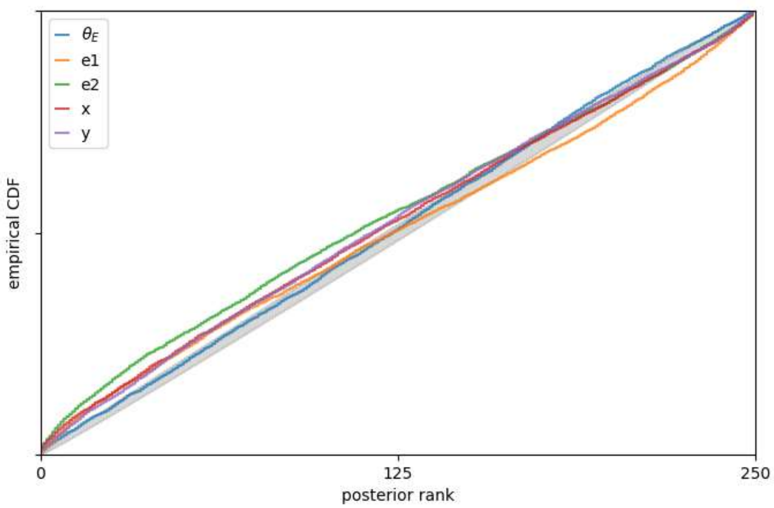
Target



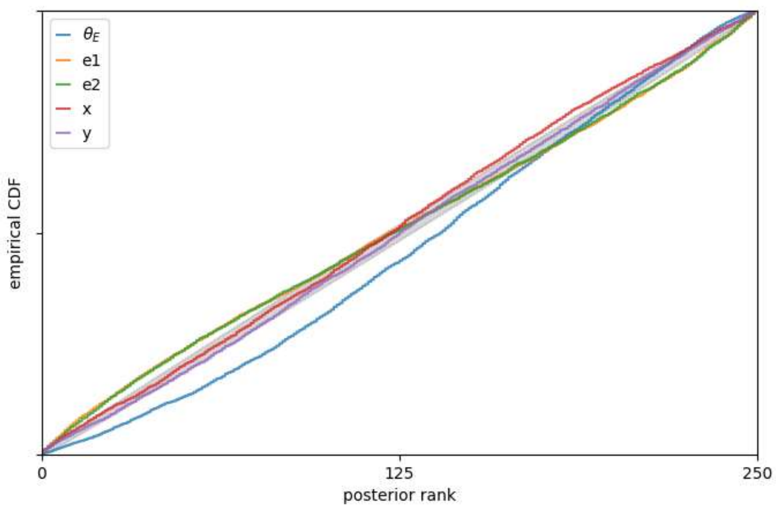
- Deviations from uniformity across all parameters

Ranks CDF - MMD, λ = 0.02

Source



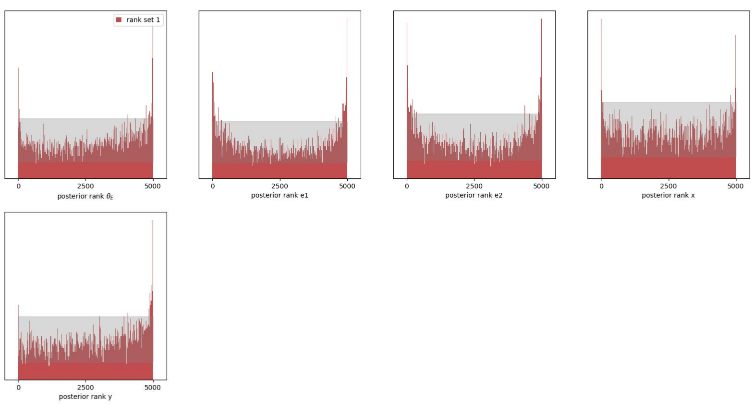
Target



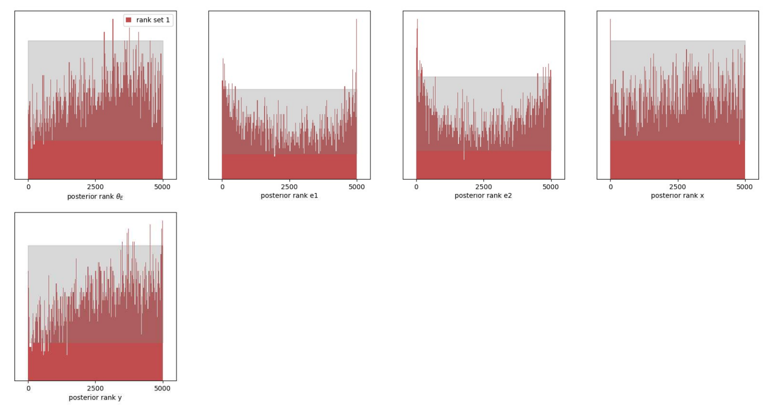
- Improvement on both source and target

Ranks Histogram - No DA

Source



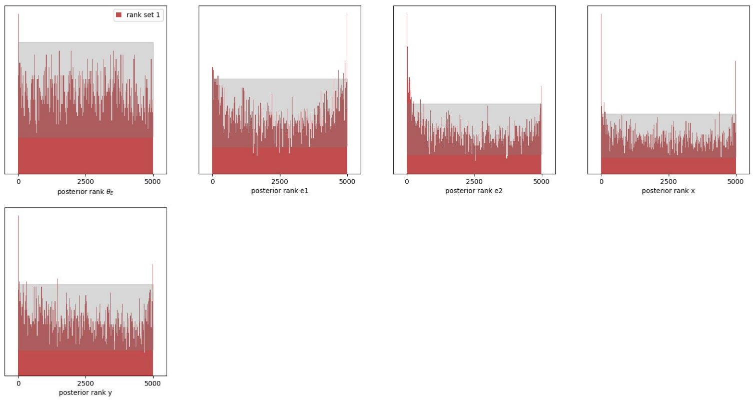
Target



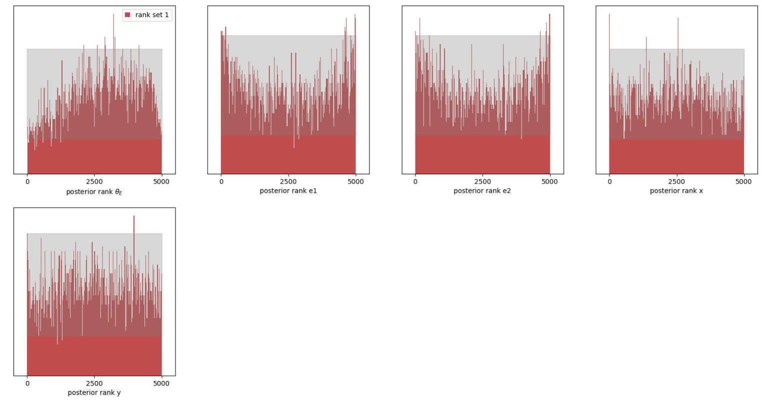
- Histograms help understand errors in uncertainty.
  - underdispersed e_1, e_2
  - Skewed θ_E, y
  - Large deviations in source domain

Ranks Histogram - MMD, λ = 0.02

Source



Target



- Marginal improvements in uncertainty calibration
  - Variance still underestimated for e_1, e_2 but discrepancy is mitigated
  - Skew in θ_E ranks still present but gone for x
  - Source domain still largely deviates from uniform

Future Work

- SBI robust against changes in data – test on source and target with larger discrepancy
- Different distance measures
  - CORAL
  - KL-Divergence
- Use real lensing data
- Test different networks
  - Larger CNN
  - Different network architectures

Works Cited

- K. Cranmer, J. Brehmer, and G. Louppe, Proceedings of the National Academy of Science 117, 30055–30062 (2020)
- S. Talts, M. Betancourt, D. Simpson, A. Vehtari, and A. Gelman, arXiv:1804.06788 (2018)
- A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Sch ̈olkopf, and A. Smola, Journal of Machine Learning Research 13, 723–773 (2012).

Extra Slides

Maximum Mean Discrepancy

$$\begin{aligned}\mathrm{MMD}^2[\mathcal{F},p,q] &= \left[\sup_{\|f\|_{\mathcal{H}}\leq 1}\left(\mathbf{E}_x[f(x)]-\mathbf{E}_y[f(y)]\right)\right]^2 \\ &= \left[\sup_{\|f\|_{\mathcal{H}}\leq 1}\langle \mu_p-\mu_q,f\rangle_{\mathcal{H}}\right]^2 \\ &= \|\mu_p-\mu_q\|^2_{\mathcal{H}}.\end{aligned}$$

$$\begin{aligned}&= \langle \mu_p,\mu_p\rangle_{\mathcal{H}}+\langle \mu_q,\mu_q\rangle_{\mathcal{H}}-2\langle \mu_p,\mu_q\rangle_{\mathcal{H}} \\ &= \mathbf{E}_{x,x'}\langle \varphi(x),\varphi(x')\rangle_{\mathcal{H}}+\mathbf{E}_{y,y'}\langle \varphi(y),\varphi(y')\rangle_{\mathcal{H}}-2\mathbf{E}_{x,y}\langle \varphi(x),\varphi(y)\rangle_{\mathcal{H}},\end{aligned}$$

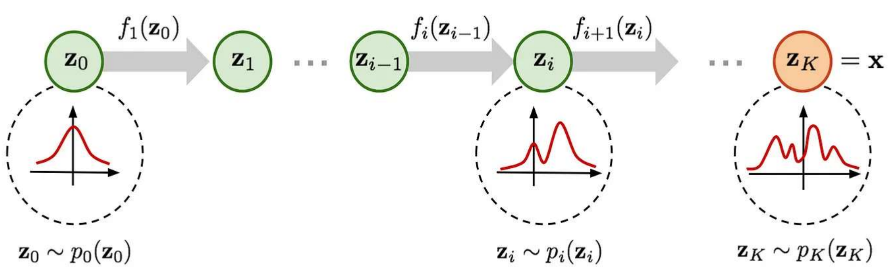
$$\langle \varphi(x),\varphi(y)\rangle_{\mathcal{H}} = k(x,y)$$

$$= \frac{1}{m^2}\sum_{i,j=1}^{M}k(x_i,x_j)+\frac{1}{n^2}\sum_{i,j=1}^{M}k(y_i,y_j)-\frac{2}{mn}\sum_{i,j=1}^{M}k(x_i,y_j)$$

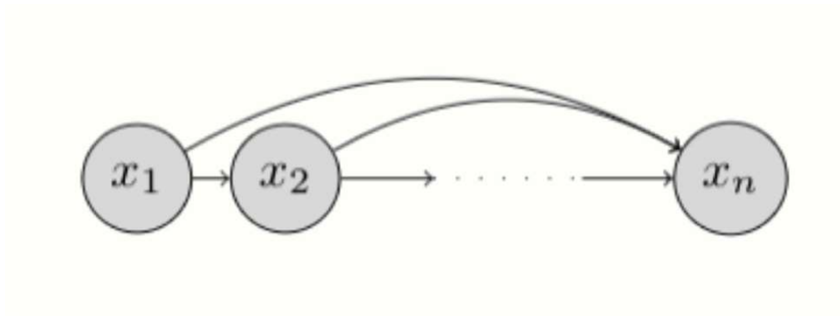
$$k(x,x') = e^{-\frac{\|x-x'\|^2}{2\sigma^2}}$$

Density Estimation With Masked Autoregressive Flow

- Represent the target density as an invertible transformation of a simpler density
- Decompose density into product of 1D conditionals



(Weng, 2018)



(Ermon, 2018)

Normalizing Flows

$$p(x) = \pi_u(f^{-1}(x)) \left| \det \left(\frac{\partial f^{-1}}{\partial x} \right) \right|$$

Autoregressive Models

$$p(x) = \prod_i p(x_i | x_{1:i-1})$$

SBC

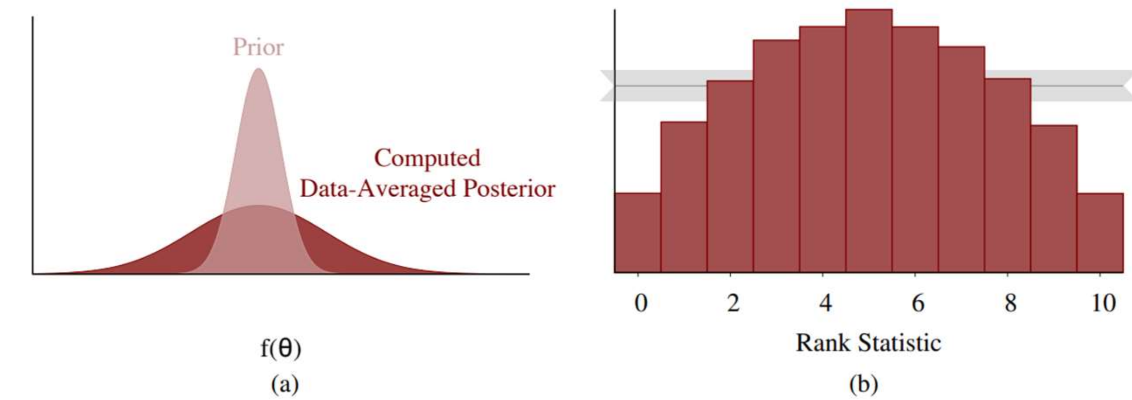


FIG 5. *A symmetric, ∩-shaped distribution indicates that the computed data-averaged posterior distribution (dark red) is overdispersed relative to the prior distribution (light red). This implies that on average the computed posterior will be wider than the true posterior.*

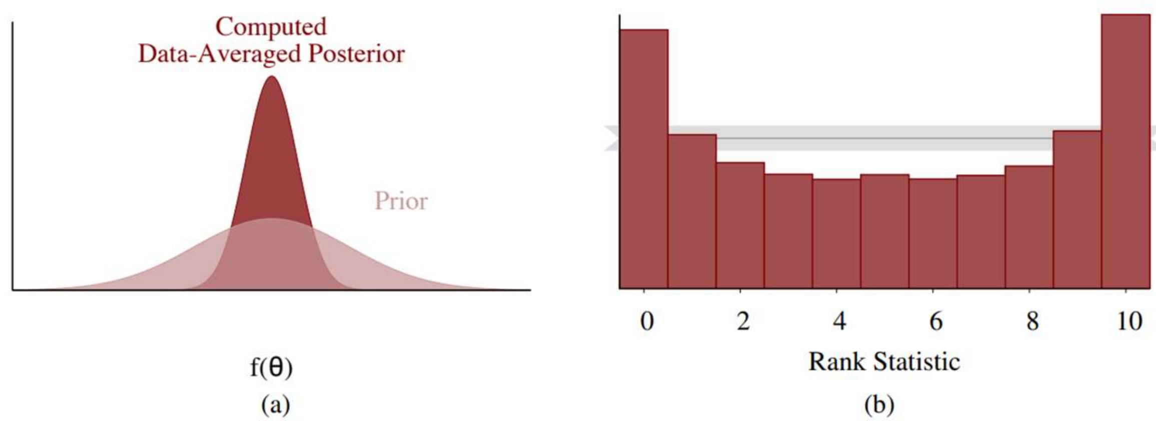


FIG 6. *A symmetric ∪ shape indicates that the computed data-averaged posterior distribution (dark red) is underdispersed relative to the prior distribution (light red). This implies that on average the computed posterior will be narrower than the true posterior.*

(Talts+, 2018)

SBC cont.

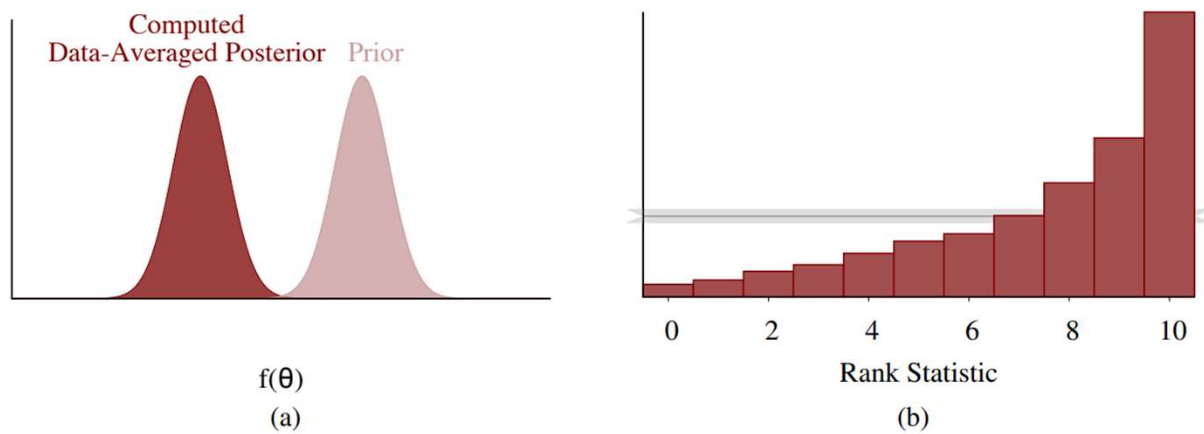


FIG 7. *Asymmetry in the rank histogram indicates that the computed data-averaged posterior distribution (dark red) will be biased in the opposite direction relative to the prior distribution (light red). This implies that on average the computed posterior will be biased in the same opposite direction.*

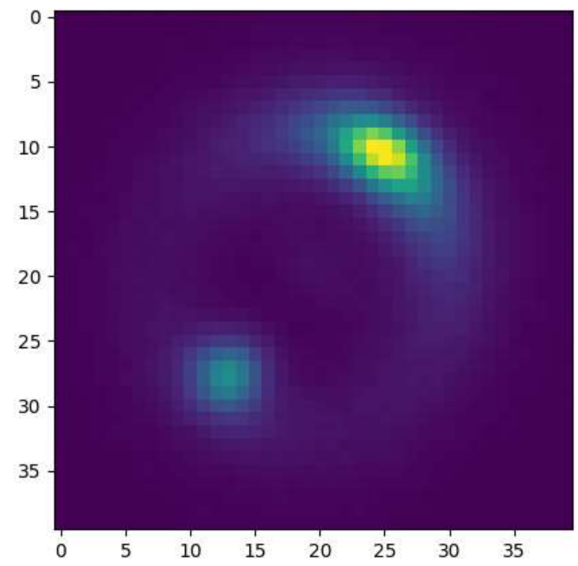
(Talts+, 2018)

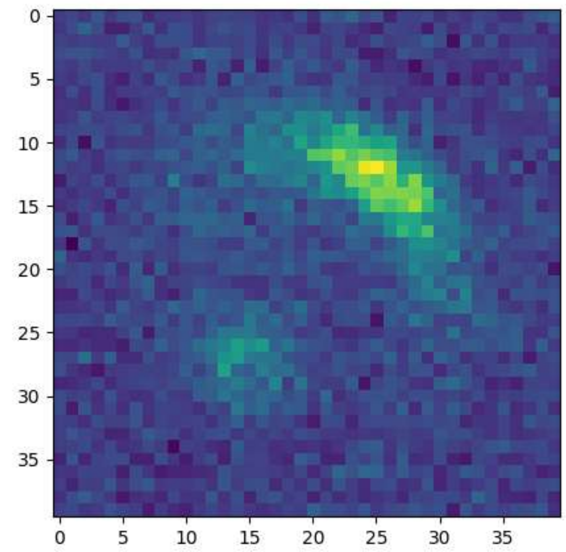
Fermilab

Domain Shift

Covariate Shift:

- Change in data space but not in label space
  - Change in conditional probability but not marginal probabilities





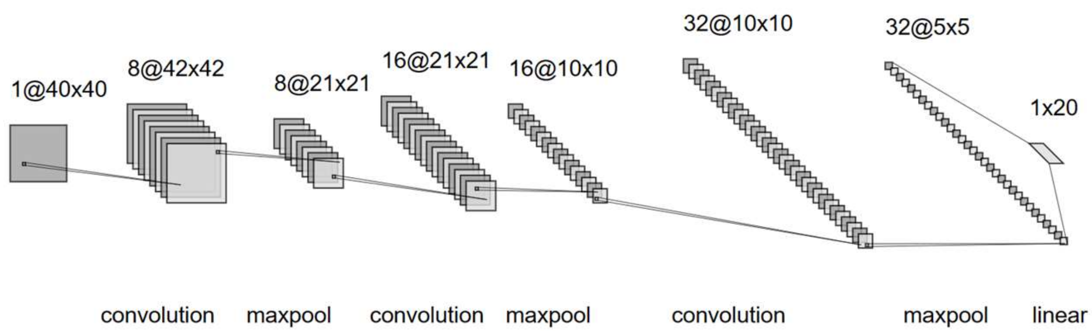
Embedding Network

3 convolutional layers

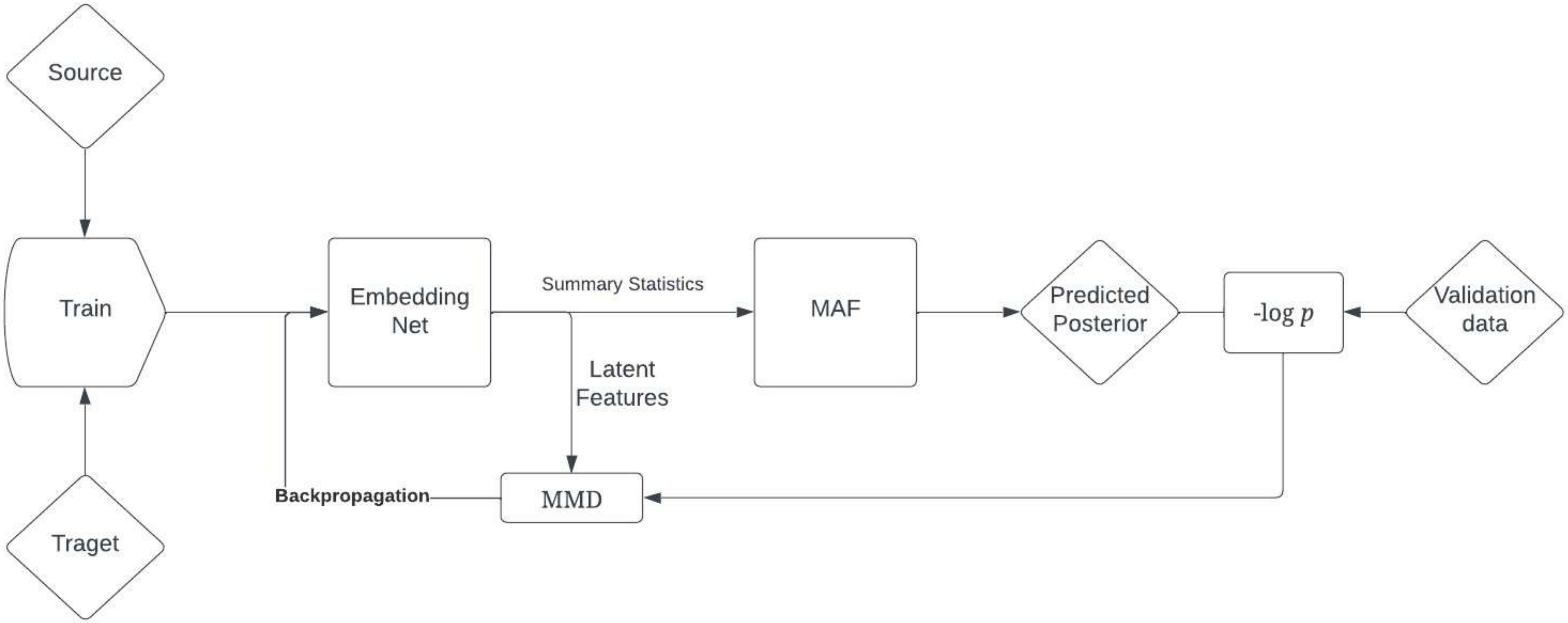
1 fully connected layer

20 summary outputs

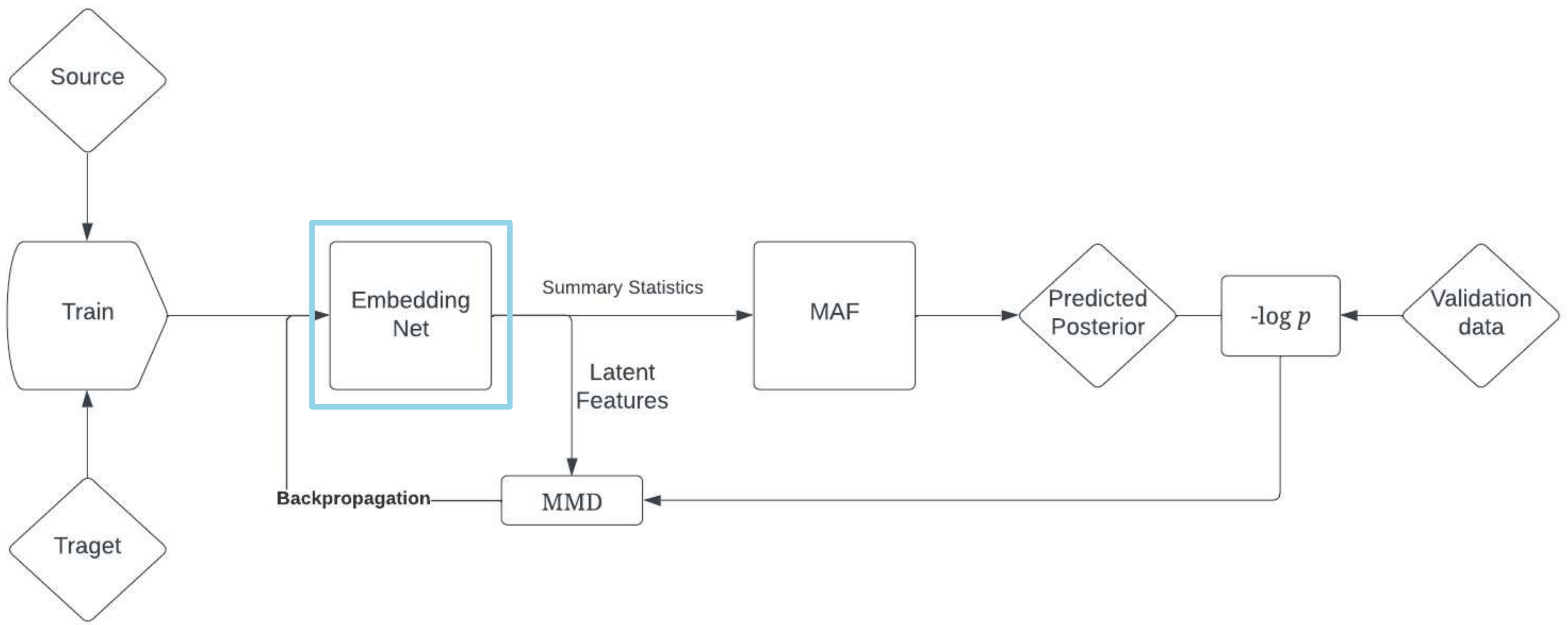
Layer	Output Shape	Parameters
Conv2d	[-1,8,42,42]	k=3, s=1, p=2
BatchNorm2d	[-1,8,42,42]	k=3, s=1
MaxPool2d	[-1,8,21,21]	k=2, s=2
Conv2d	[-1,16,21,21]	k=3, s=1, p=1
BatchNorm2d	[-1,16,21,21]	k=3, s=1
MaxPool2d	[-1,16,10,10]	k=2, s=2
Conv2d	[-1,32,10,10]	k=3, p=same
BatchNorm2d	[-1,32,10,10]	k=3, s=1
MaxPool2d	[-1,32,5,5]	k=2, s=2
Linear	[-1,20]	in_ ft=32 × 5 × 5, 4



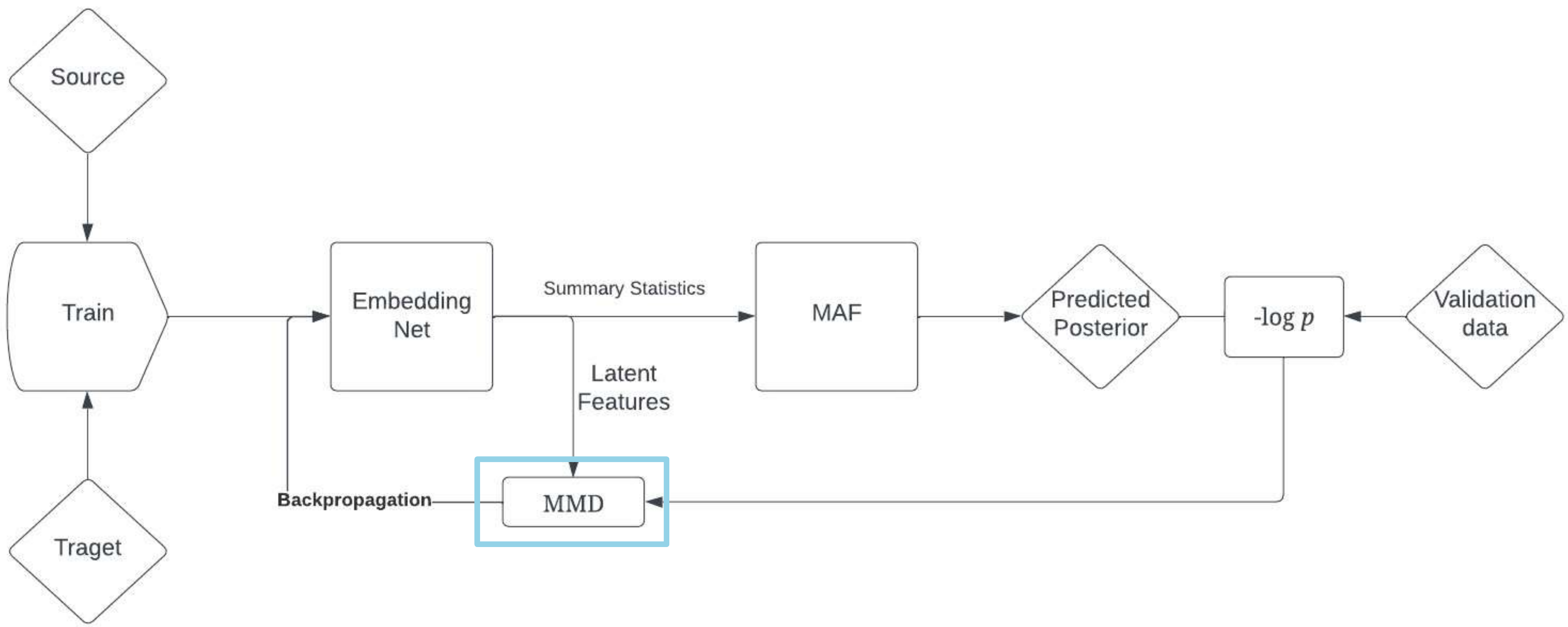
Density Estimator



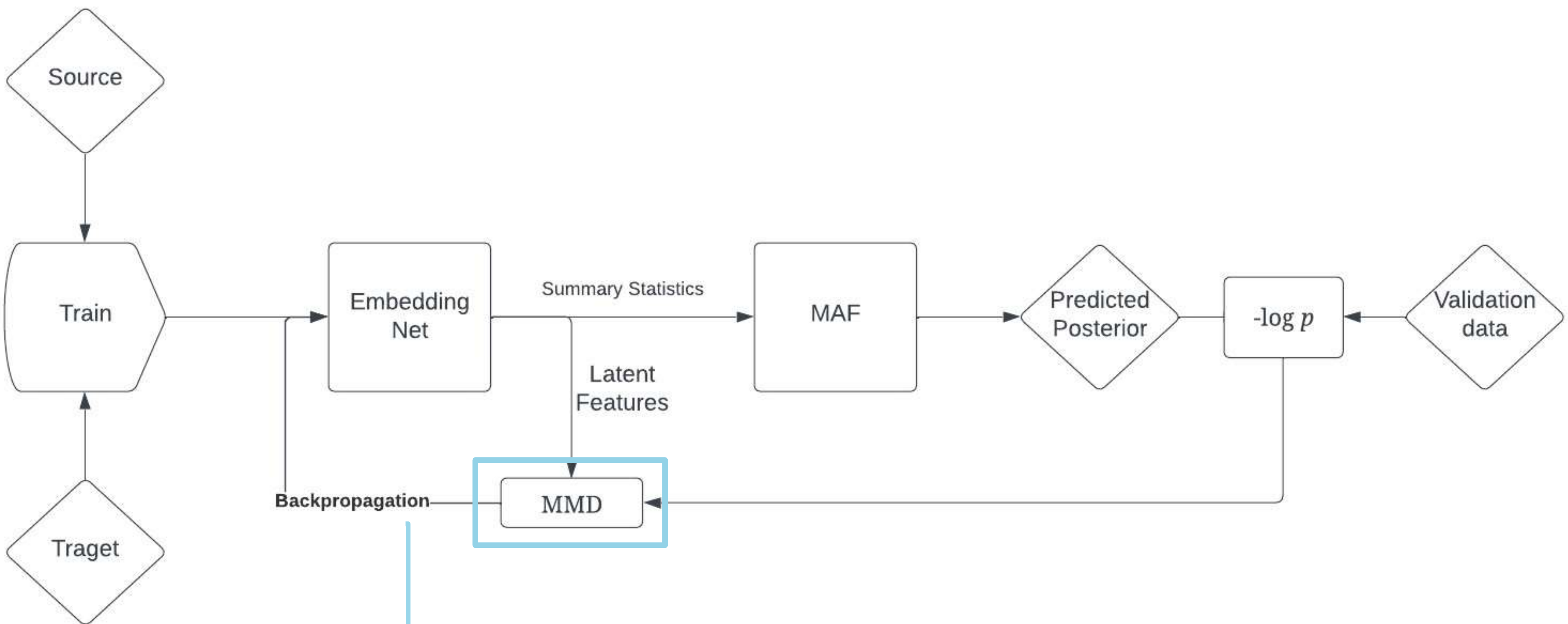
Density Estimator



Density Estimator



Density Estimator



$$\mathcal{L}_{tot} = -\sum_{i=1}^{N} \log p(\theta_i) + \lambda \mathcal{L}_{\mathrm{MMD}^2},$$

Density Estimator