Illinois Physics
The Grainger College of Engineering

ILLINOIS
NCSA | National Center for Supercomputing Applications

Deep Learning at Scale for Morphological Classification of Galaxies in DES

Asad Khan[a,b]

E. A. Huerta[a,c], Sibo Wang[a], Robert Gruendl[a,c], Elise Jennings[d], Huihuo Zheng[d]

NCSA Gravity Group
gravity.ncsa.illinois.edu
Numerical Relativity, Einstein Toolkit, Gravitational Wave Astrophysics, Astrodynamics, MMA
Deep Learning / Machine Learning, Data Analysis, HPC

[a] National Center for Supercomputing Applications
[b] Dept. of Physics, University of Illinois at Urbana-Champaign
[c] Dept. of Astronomy, University of Illinois at Urbana-Chmapign
[d] Argonne National Laboratory, Leadership Computing Facility

Motivation

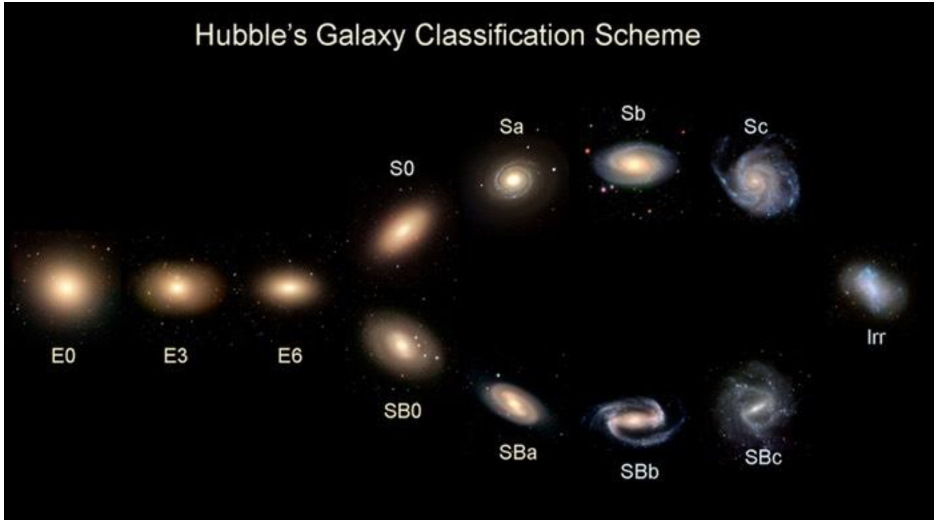
EM Surveys: key insights into Large Scale Structure, GW follow up, etc

As scale and depth continue to increase: need for low latency data analysis pipelines

Starting point of any large survey analysis: Object Classification.

For galaxies, broadly:

1. Elliptical
2. Spiral

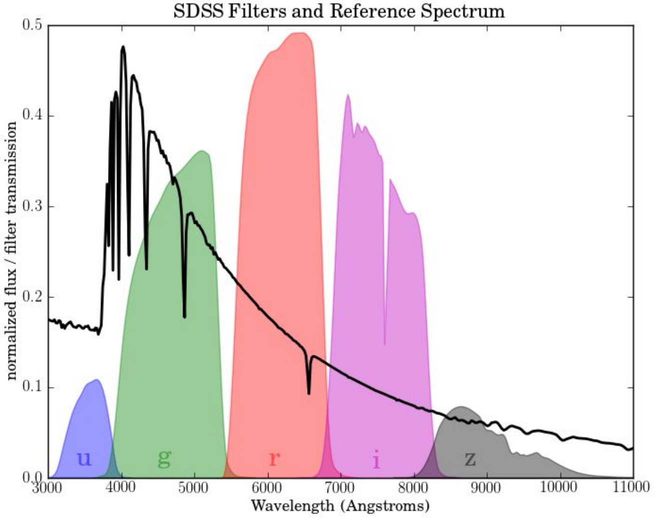


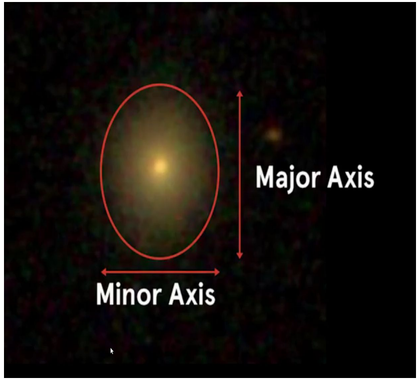
Traditional Methods

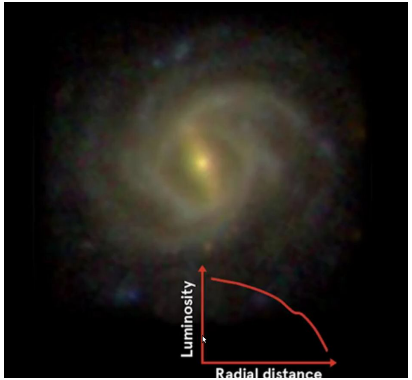
<u>Machine Learning</u>:

- Domain Knowledge, Slow Feature Engineering:
  - Color indices, Eccentricity, Adaptive Moment, Concentration, etc

- Classification accuracies: ~85% (Significantly below Human Level Performance)





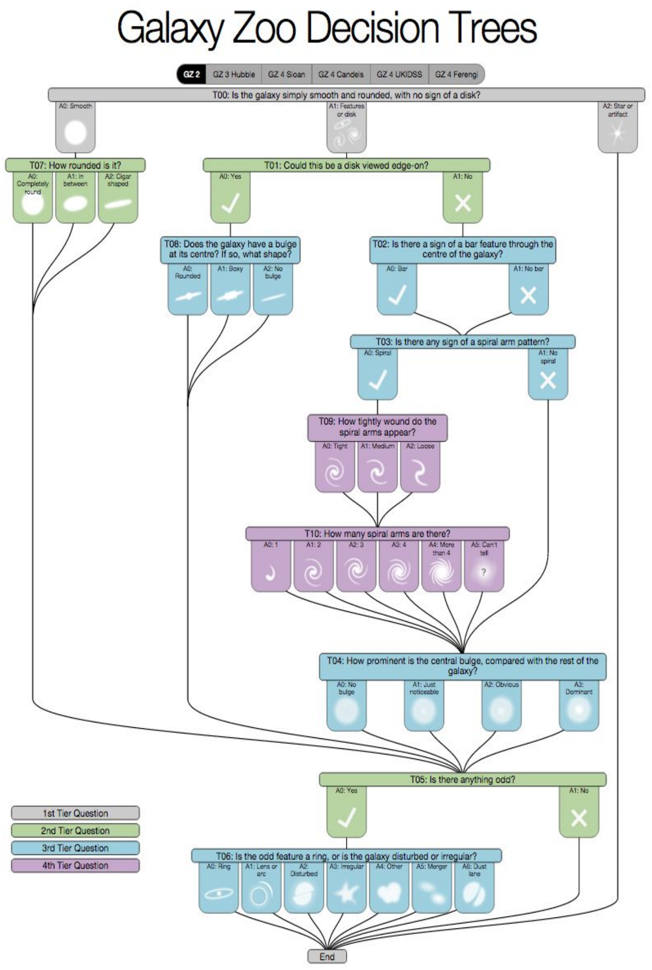


Traditional Methods

Citizen Science Approach:

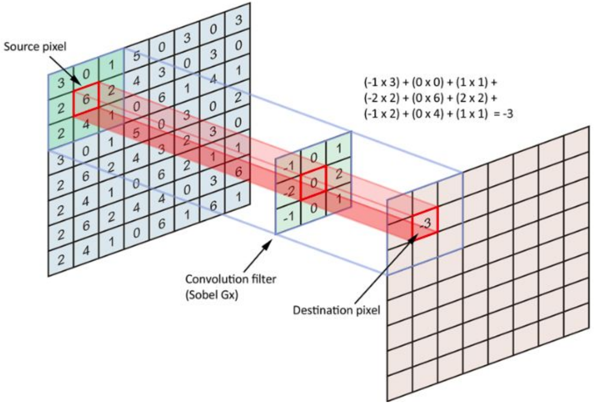
- Galaxy Zoo/Sloan Digital Sky Survey:
  - Crowd sourced Astronomy project, running since July, 2007
  - 50 million classifications received in the first year, contributed by 150,000 volunteers

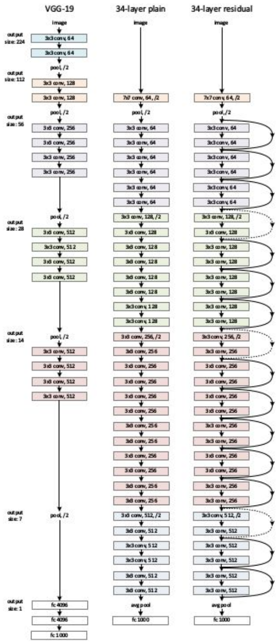
- As electromagnetic surveys continue to increase depth and coverage, campaigns of this nature may lack scalability

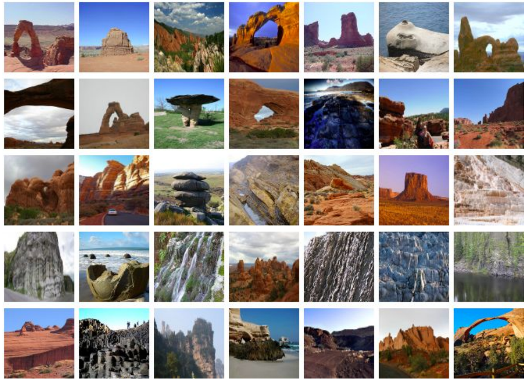


Deep Learning

- Convolutional Neural Network (CNN):





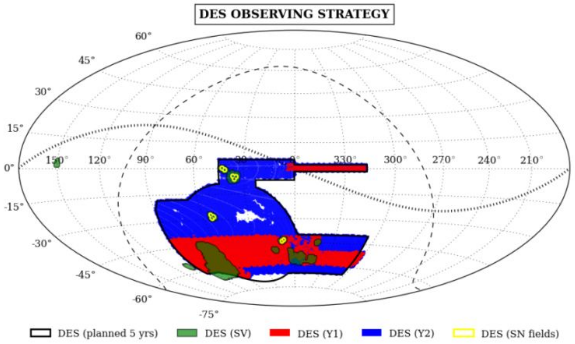


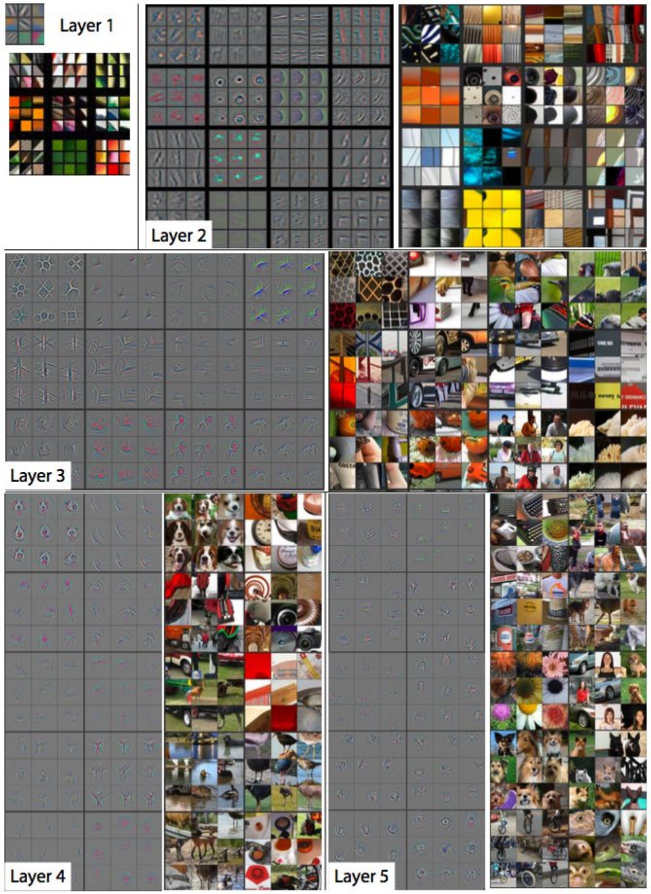
- ImageNet: 14 million images in 10,000 categories
    - Deep CNNs: ≥ human-level performance on object classification tasks
    - SOA Top 5 Accuracy: ~96%

Transfer Learning

- Deep Learning Algorithms: Data Hungry !
  - Transfer Learning: Domain adaptation with little re-learning/fine-tuning.

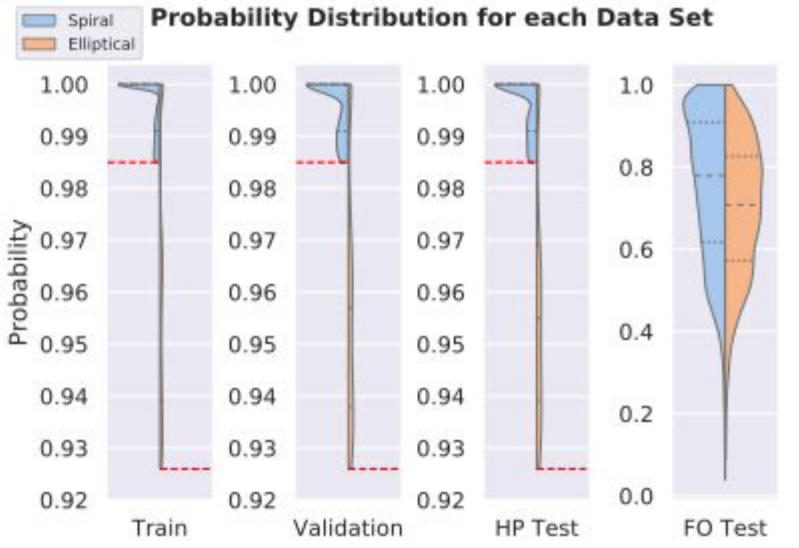
- Dark Energy Survey:
  - ~400 million catalogued objects
  - SDSS/Galaxy Zoo: seed dataset for fine-tuning
  - DES overlap with SDSS: Cross Validation

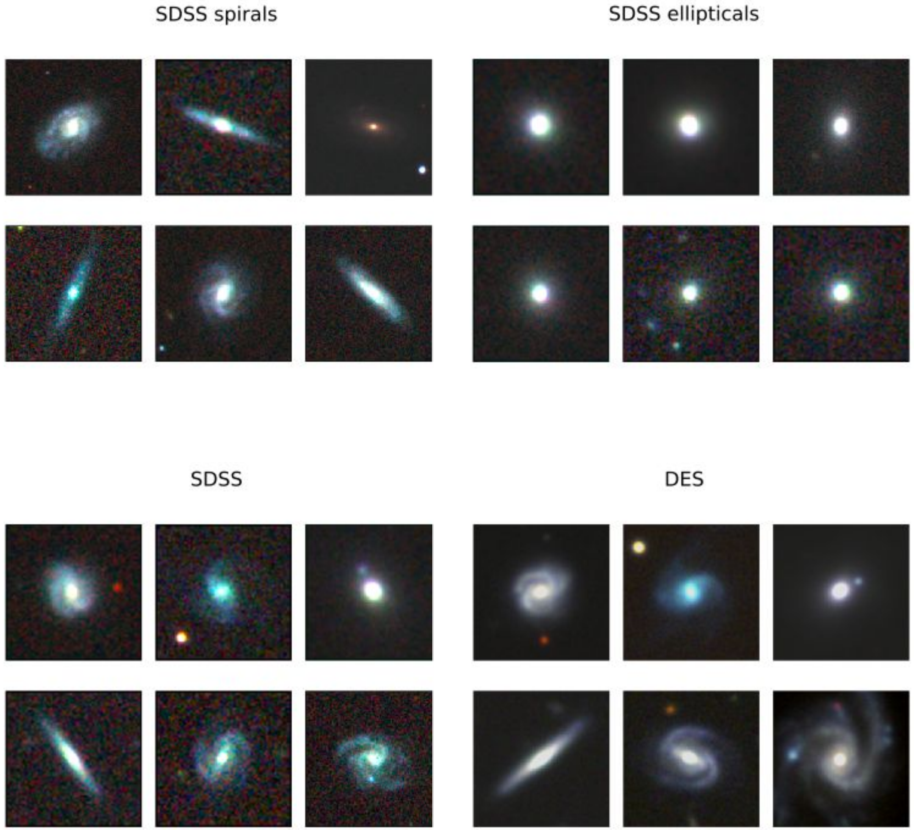




Data Curation

- **Training/Validation Sets:** SDSS
- **Test Sets:** SDSS and DES crossmatched

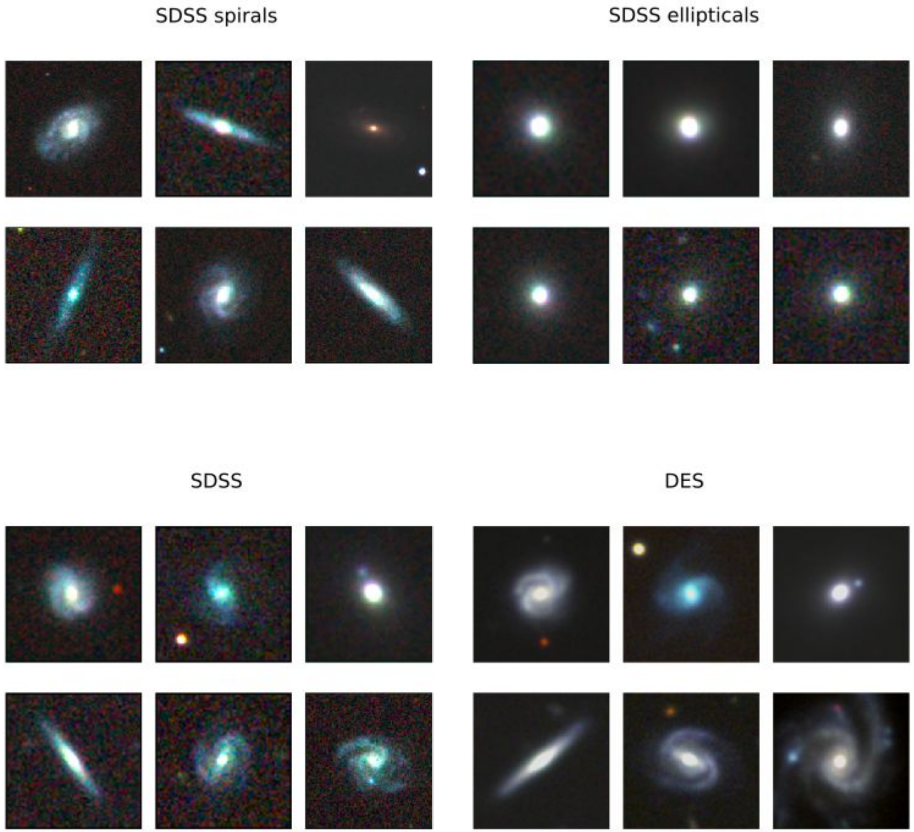




Data Curation

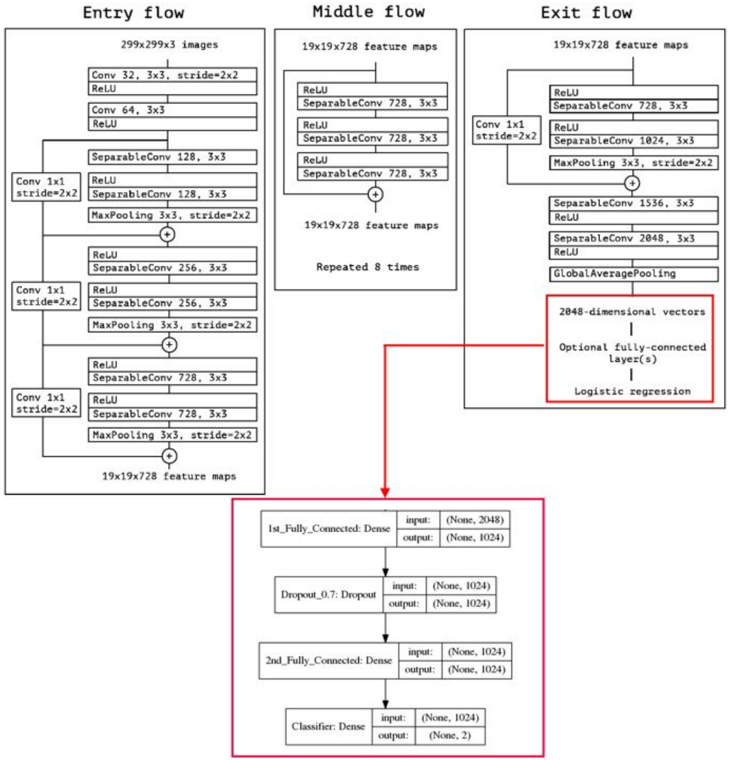
- **Training/Validation Sets:** SDSS
- **Test Sets:** SDSS and DES crossmatched

Dataset	Spirals	Ellipticals
Training set	18,352	18,268
HP SDSS Test Set	516	550
HP DES Test Set	516	550
FO SDSS Test Set	6,677	5,904
FO DES Test Set	6,677	5,904



Model Selection

- Architectures with better ImageNet performance ➔ Better transferrable representations.[1]
- Xception Model:
  - Best ImageNet performance at the time

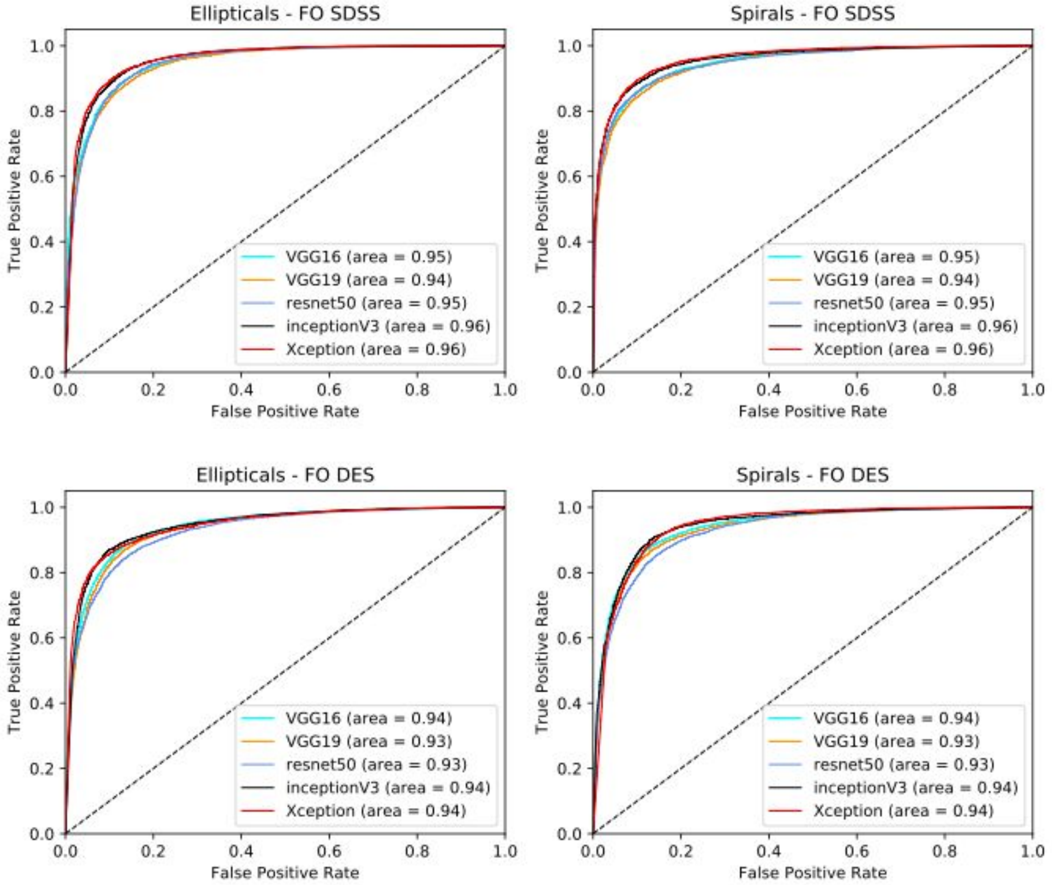


[1] S. Kornblith, J. Shlens, and Q. V. Le, “Do better imagenet models transfer better?” (2018), arXiv:1805.08974 .

Model Selection

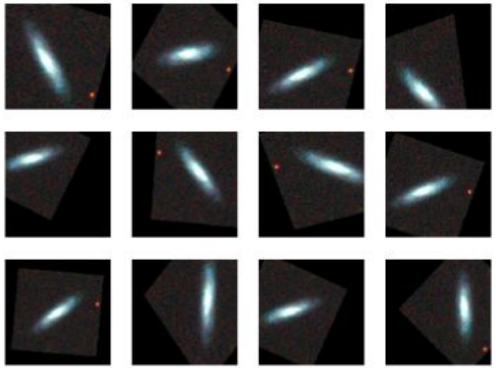
Case Studies:

Receiver Operating Characteristic (ROC) for several different fine-tuned state of the art architectures

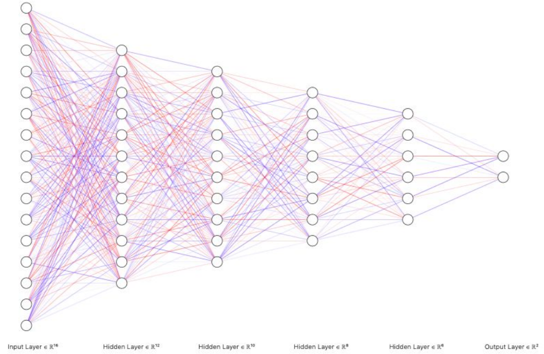


Training

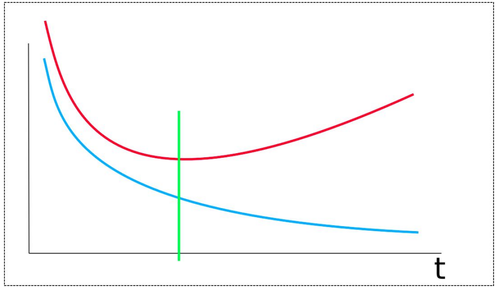
- Data Augmentation



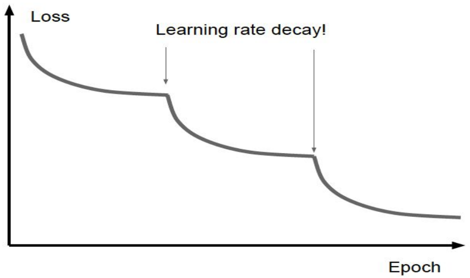
- Progressive unfreezing



- Early Stopping

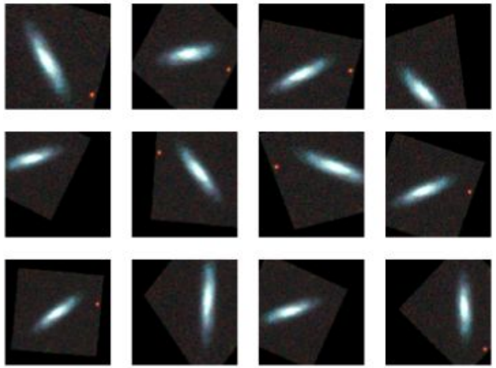


- Reduce Learning Rate on plateau

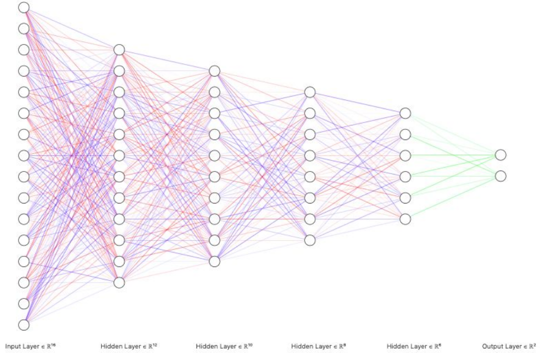


Training

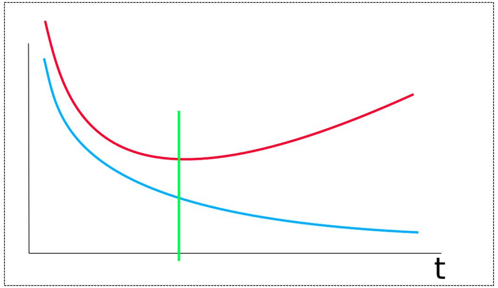
- Data Augmentation



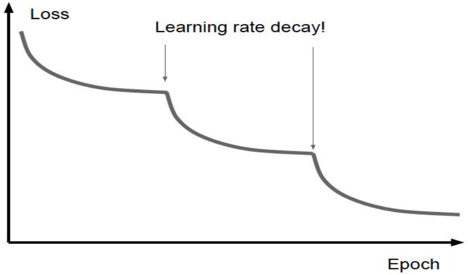
- Progressive unfreezing



- Early Stopping

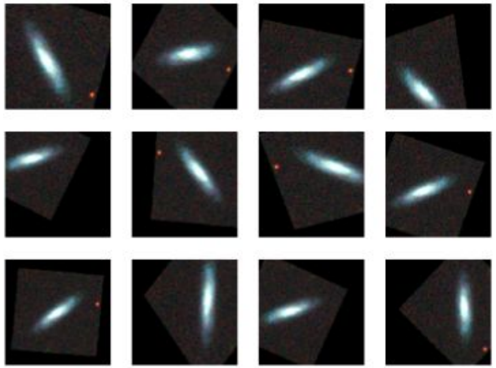


- Reduce Learning Rate on plateau

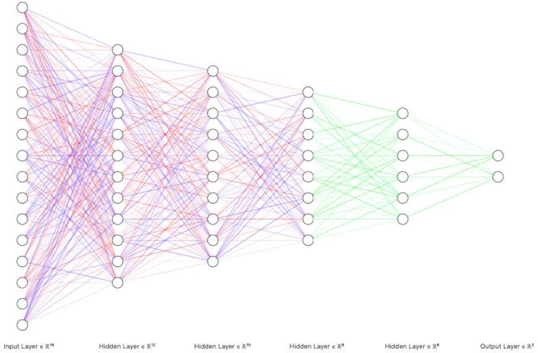


Training

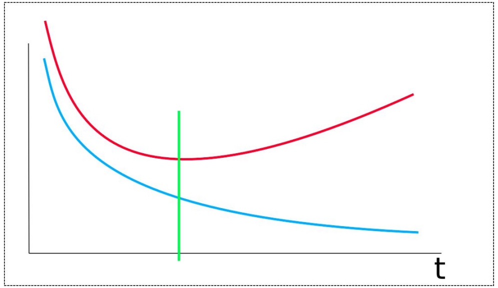
- Data Augmentation



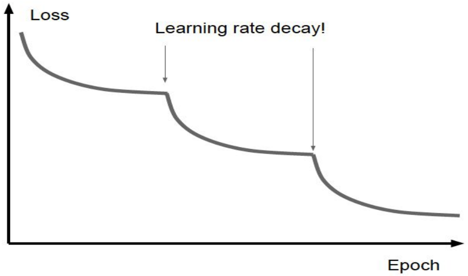
- Progressive unfreezing



- Early Stopping

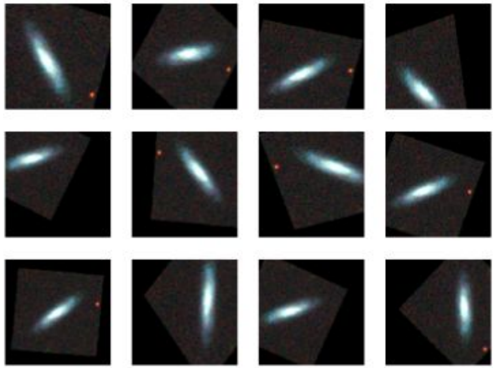


- Reduce Learning Rate on plateau

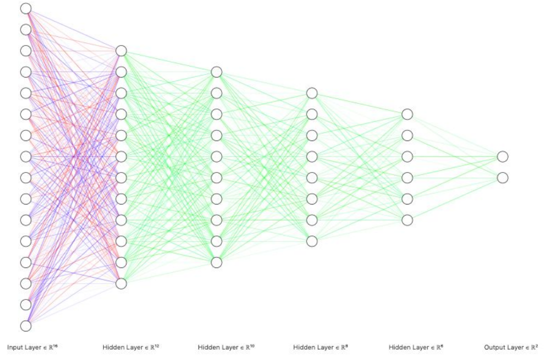


Training

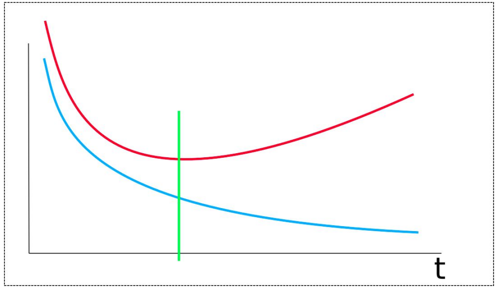
- Data Augmentation



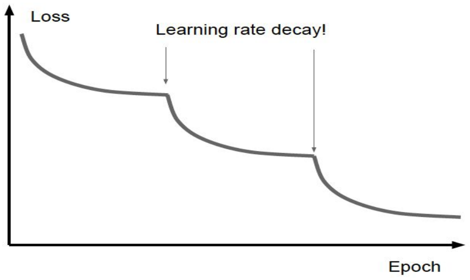
- Progressive unfreezing



- Early Stopping

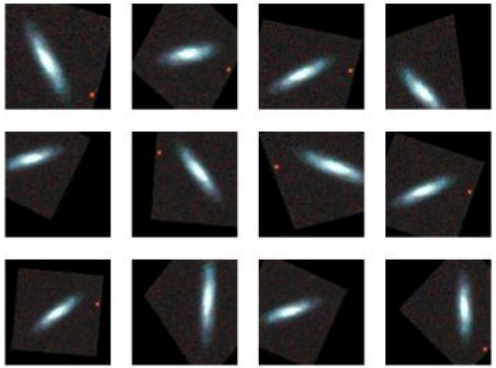


- Reduce Learning Rate on plateau

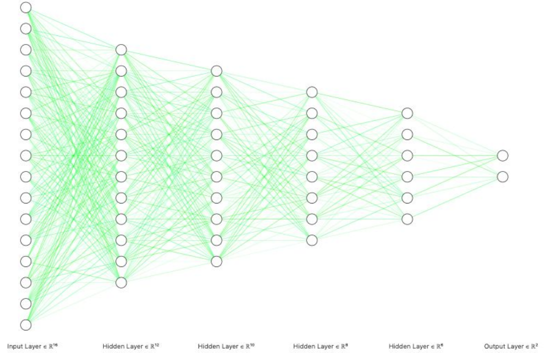


Training

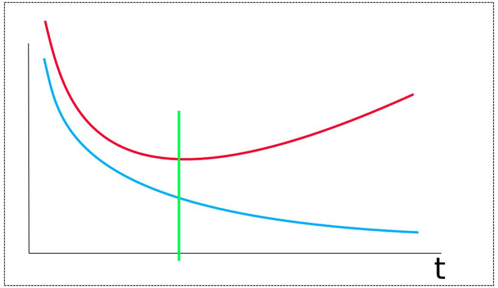
- Data Augmentation



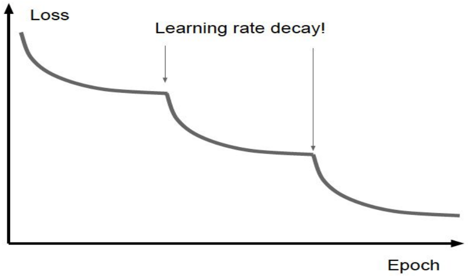
- Progressive unfreezing



- Early Stopping

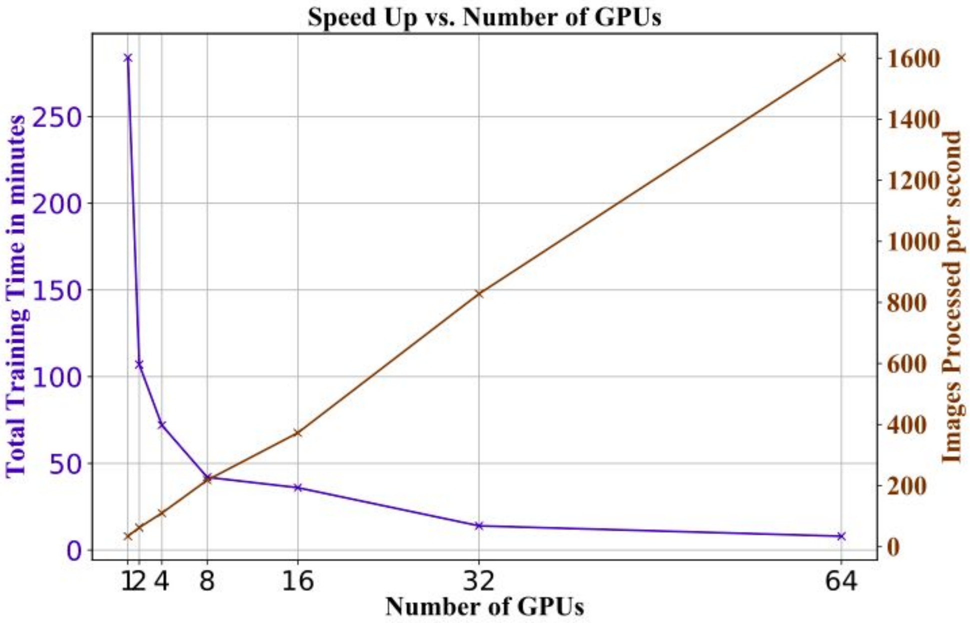


- Reduce Learning Rate on plateau



Training

- **Single GPU:** 5 hours on a Tesla P100 GPU on XSEDE for 36,500 images
- **Distributed Learning:** 8 minutes on 64 K80 GPUs on Cooley Supercomputer at Argonne

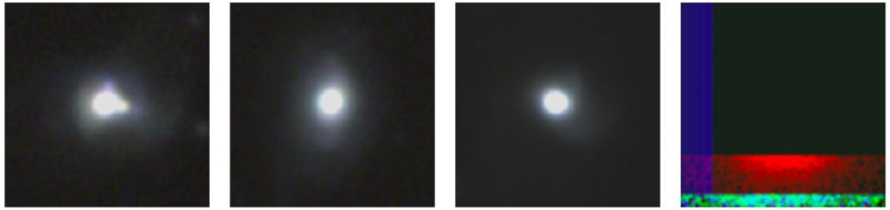


Results

Dataset	Precision	Recall	FPR	Accuracy	F1 score
Training set				99.81%	0.9998
HP SDSS Test Set	0.996	1	0.004	99.81%	0.9980
HP DES Test Set	0.998	0.995	0.002	99.62%	0.9961
FO SDSS Test Set	0.945	0.991	0.055	96.76%	0.9675
FO DES Test Set	0.965	0.946	0.025	96.32%	0.9685

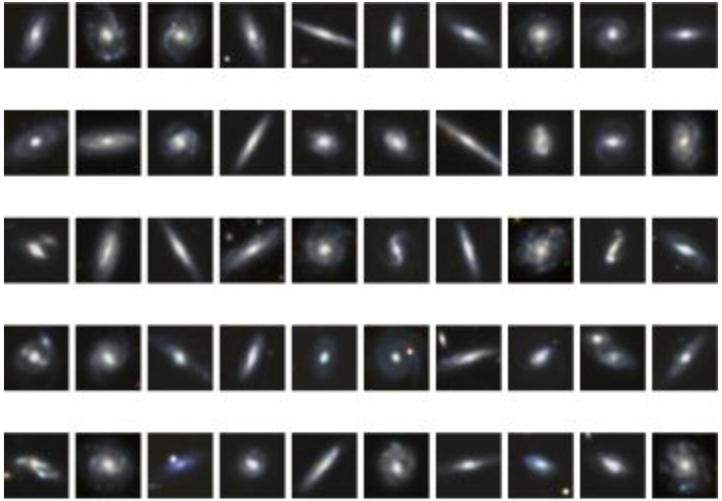
- Examples of misclassifications

HP DES. Predicted Class: Spiral

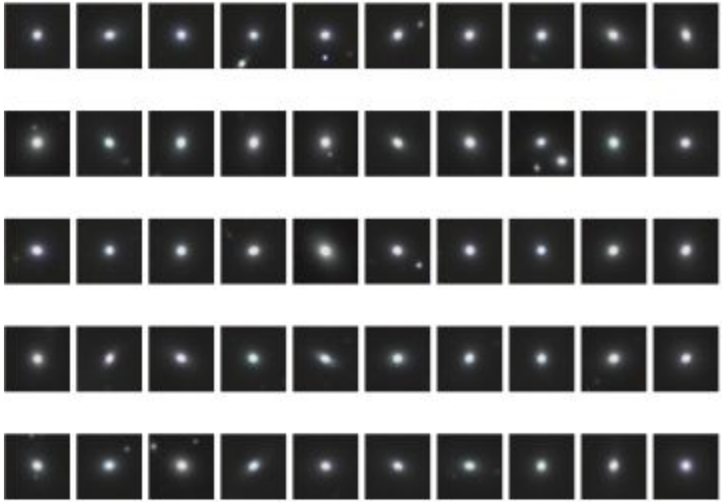


Results

- Unlabelled DES



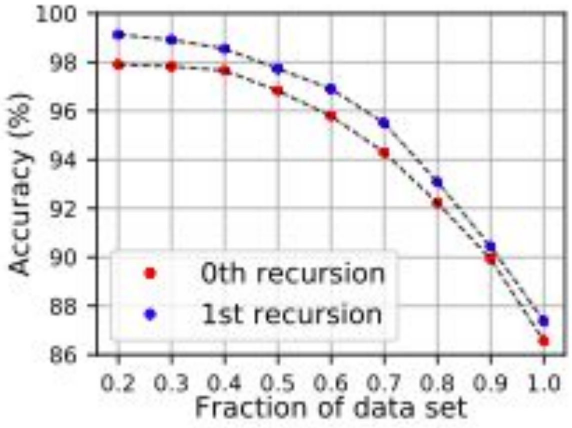
Predicted Spirals



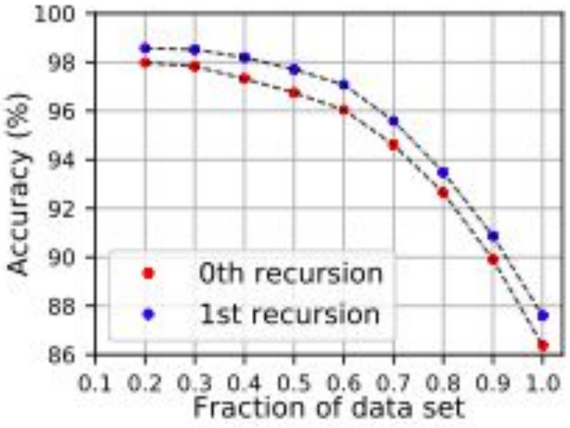
Predicted Ellipticals

Recursive Training:

- Accuracy vs. N high confidence predictions as a fraction of total FO test datasets

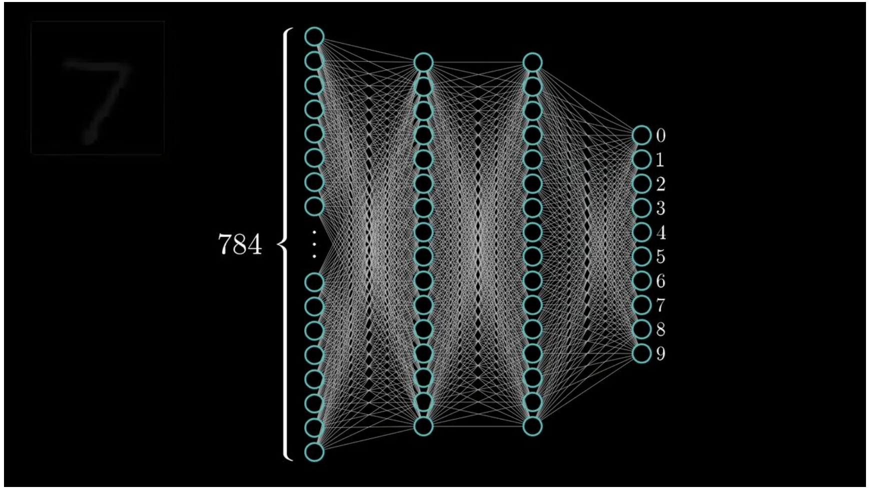


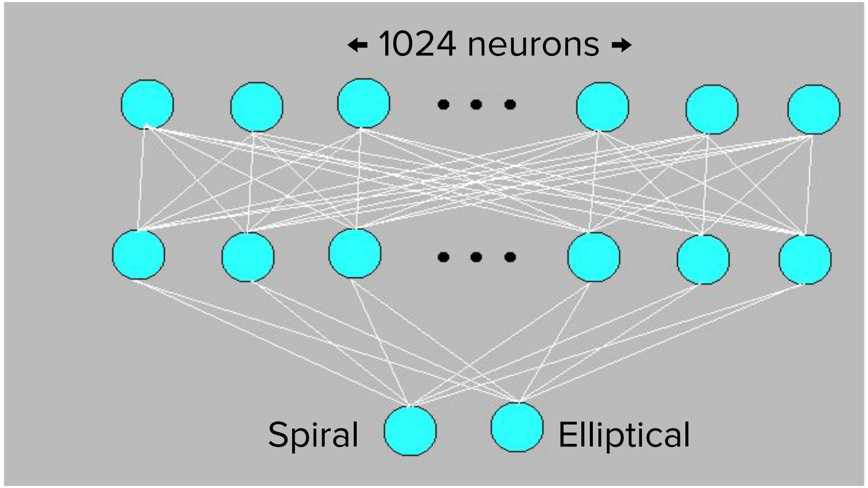
SDSS



DES

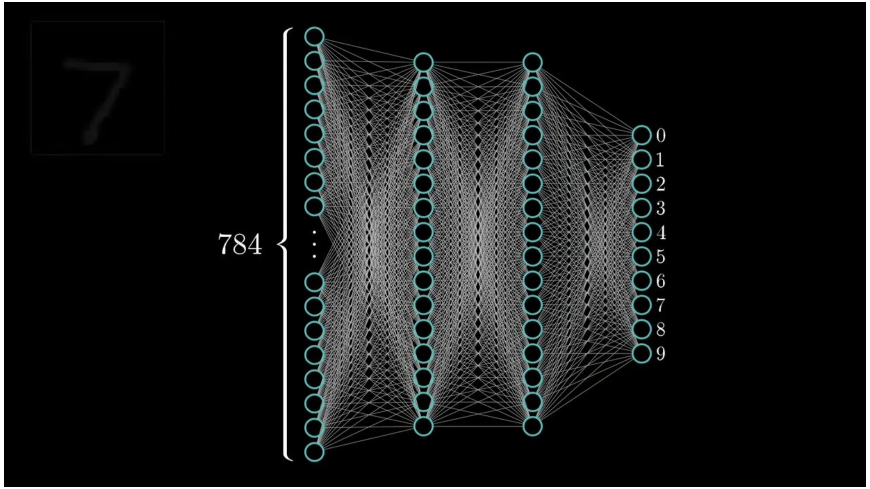
Clustering: A Heuristic Check

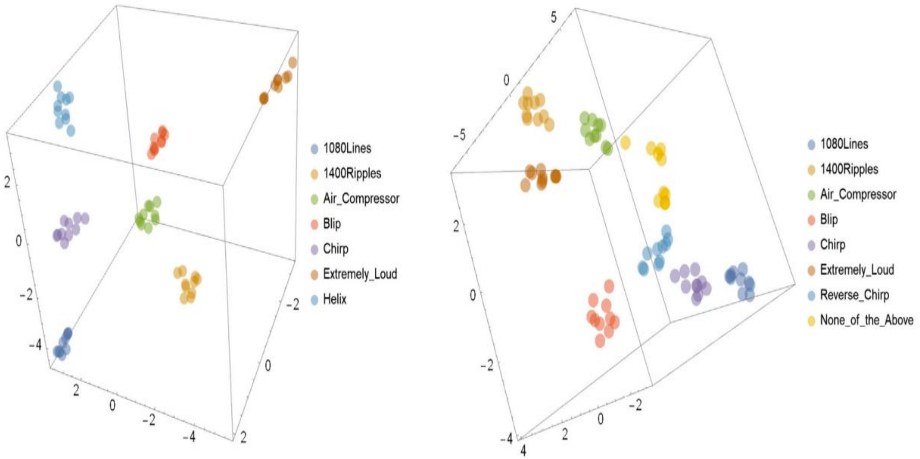




Source: 3Blue1Brown

Clustering: A Heuristic Check

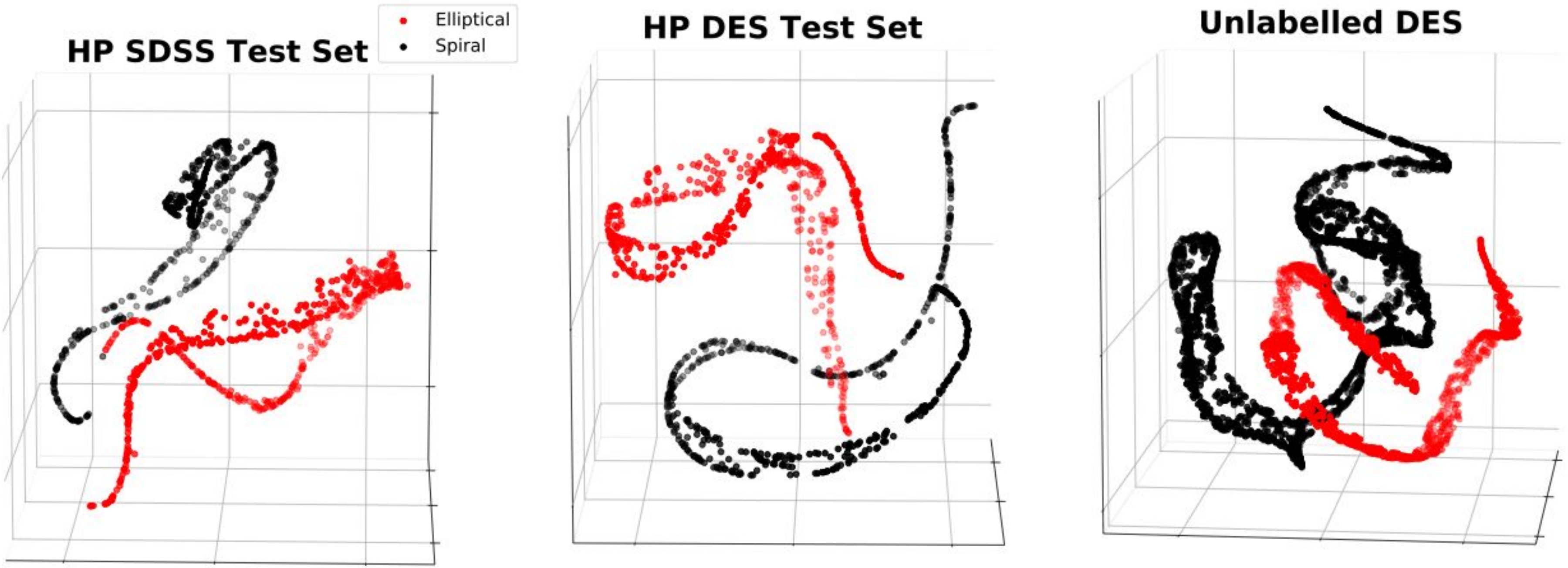




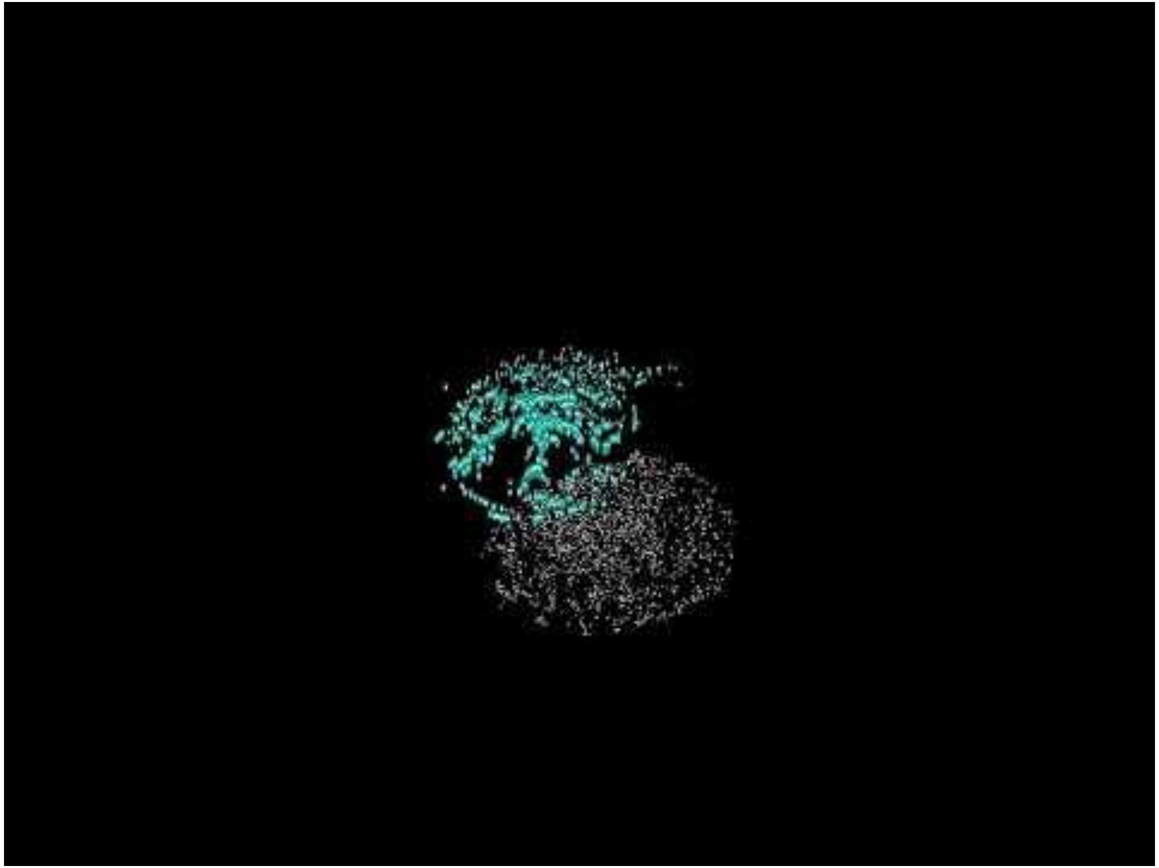
Source: 3Blue1Brown

Daniel George, Hongyu Shen, E. A. Huerta
“Classification and Unsupervised Clustering of LIGO Data with Deep Transfer Learning”

t-Distributed Stochastic Neighbor Embedding (tSNE)



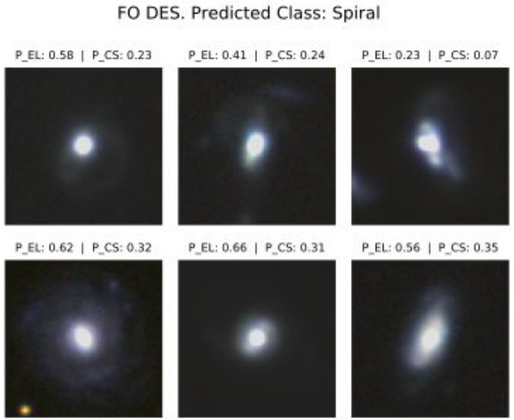
t-Distributed Stochastic Neighbor Embedding (tSNE)

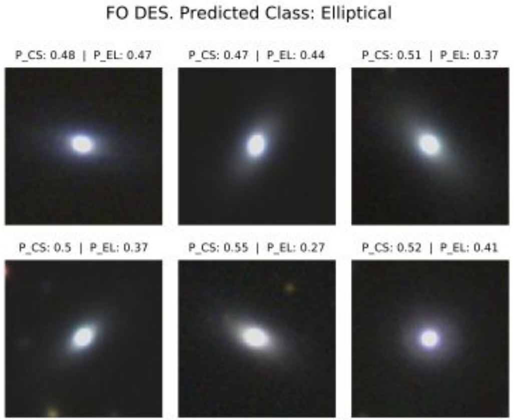


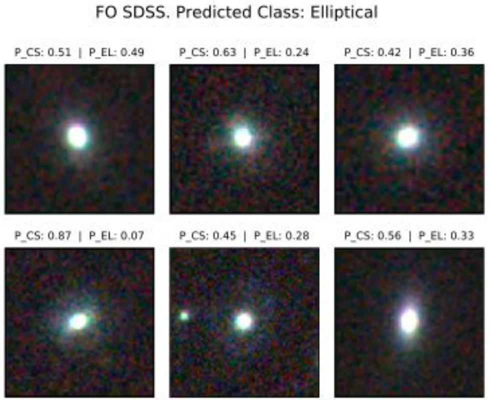
Conclusion:

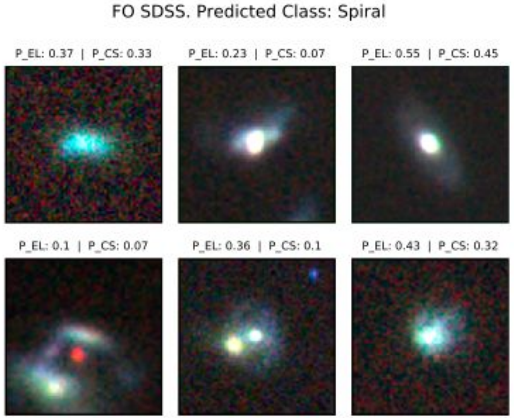
- First application of Deep Transfer Learning, using disparate datasets, combined with distributed training for galaxy classification.
- State-of-the-art classification accuracies for SDSS and DES galaxies.
- Label over 10,000 DES galaxies that had not been classified in previous surveys.
- Interpretability study to assess the robustness of the classification of unlabelled DES images.
- Recursive training on the most confident predictions from the newly labeled DES galaxies, boosting the classification accuracy for FO SDSS and DES test sets.

Appendix: Misclassifications







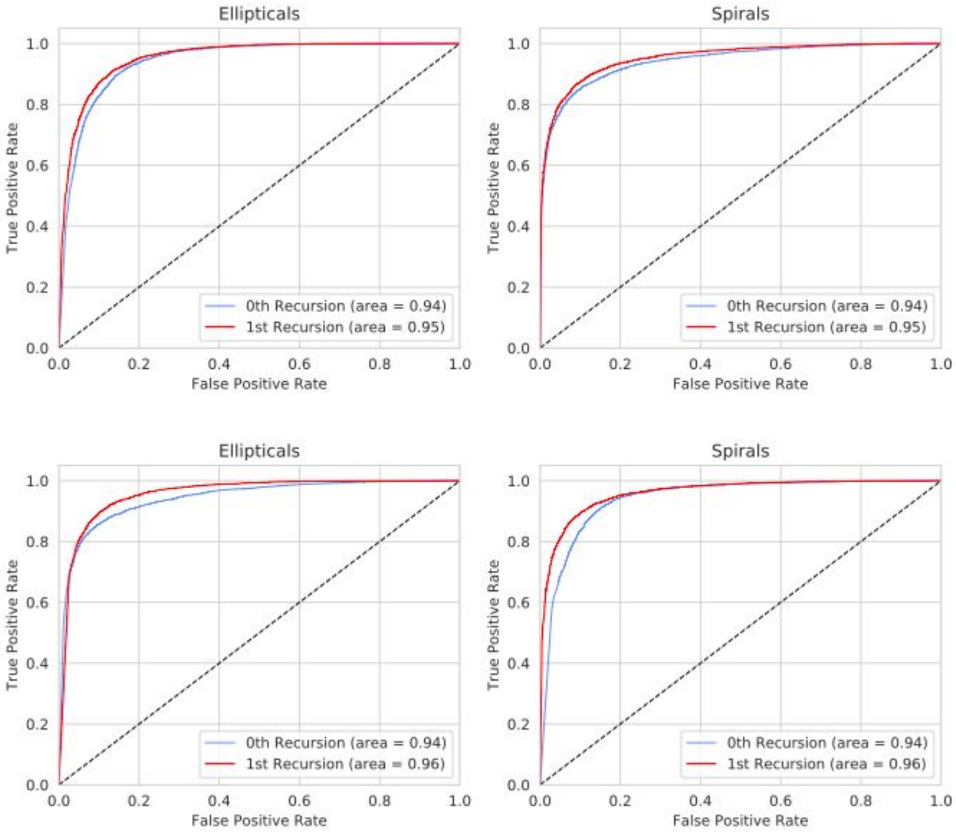


Appendix: Scaling Results

GPUs	Time per epoch (s)	# epochs	Total time	Accuracy	Val Accuracy
1	410 922 1626	1 11 4	4h 44 m	0.9992	0.9979
2	231 481 830	1 6 4	1h 47m	0.9993	0.9990
4	119 246 427	1 5 7	1h 12m	0.9995	0.9990
8	64 124 214	1 6 8	42m	0.9991	0.9979
16	35 63 109	1 4 17	36m	0.9993	0.9980
32	20 31 53	1 6 12	14m	0.9993	0.9990
64	13 15 27	1 5 15	8m	0.9993	0.9990

Appendix: Recursive Training

(Top: SDSS, Bottom: DES)



t-Distributed Stochastic Neighbor Embedding (tSNE)