***ASTROCLIP: CROSS-MODAL PRE-TRAINING FOR ASTRONOMICAL FOUNDATION MODELS***

<https://arxiv.org/abs/2310.03024>

**Data**

* DESI Legacy Survey
  + MetaData
    - 41M images
    - 152 x 152 → cropped to 96 x 96
    - 197,979 pairs
    - grayscale/rgb

**Discussion/Analysis**

* Contrastive learning tasks => allows embedding of each modality to discover physical patterns about data
* Transformer-based model for modeling of galaxy spectra

***Self-supervised similarity search for large scientific datasets***

<https://www.semanticscholar.org/reader/7ebe772d4e6c92fe3cbb64a219d3e03ddfa40ca7>

**Data**

* DESI Legacy Imaging Surveys DR9

**Technique/Methodology**

* Self-supervised model based on MoCOV2
* CNN encoder to produce a lower representation of image x
* Pre-Training
  + Randomly size images between 90% and 110% of its original size by rescaling with bilinear interpolation
  + PSF blur
  + Jitter and crop
  + Gaussian noise
* Cosine similarity

**Discussion/Analysis**

* Self-supervised learning model can be used to extract information-rich representations from unlabelled data
* Can be used to discover any object that exist in a dataset
* Self-supervised reduces barriers of computational complexity or resource allocation

**Similar Papers**

***A brief review of contrastive learning applied to astrophysics***

<https://arxiv.org/pdf/2306.05528.pdf>

**Data**

n/a

**Technique**

* Contrastive learning: self-supervised framework to learn meaningful representations from a dataset X
  + Trainable function and parametrized by a neural network
  + Different views of same object are positive pairs while others are negative pairs
  + Uses contrastive loss
    - Approaches
      * Spring loss
      * Triplet loss
      * Normalized cross entropy
  + Primarily designed for datasets without labels
  + Possible implementations
    - CMC (Contrastive Multiview Coding)
    - MoCo (Momentum Contrastive)
    - SimCLR (Simple Contrastive Learning)
    - BYOL (Bootstrap Your Own Latent)
    - CLIP (Contrastive Language-Image Pre-training)

**Discussion/Analysis**

* Usefulness
  + Reduce number of required labels
  + No predefined encoded metric
  + Opens door for representations of different types
* Not Useful
  + Not as efficient as dimensionality reduction algorithms
  + More challenging interpretability of the extracted features

***Astroinformatics of galaxies and quasars: a new general method for photometric redshifts estimation***

<https://arxiv.org/abs/1107.3160>

**Data**

* SDSS-DR7: <https://classic.sdss.org/dr7/>
  + 2048 x 1489 (from DR9)
  + Channels: 1 channel
  + Spectrum: infrared (DR18)

**Technique**

* CNN with attention mechanism; WGE (Weak-Gated Experts )
  + Partition of feature spaces
  + Create model for each predictor (expert); maps feature space to target space
  + New gate maps patterns to targets

**Discussion/Analysis**

* WGE can combine various experts (unite predictors with different strengths)

***Attention-gating for improved radio galaxy classification***

<https://arxiv.org/abs/2012.01248>

**Data**

* VLA Sky Survey
  + FR-Deep and MiraBest Datasets
  + MetaData MiraBest:
    - 1256 images (batches with 157 images each)
    - 150 x 150 pixels
    - Seems like 1 channel
  + MetaData FR-Deep:
    - MetaData
      * 1360 images
      * 150 x 150 pixels
      * 1 channel

**Technique**

* CNN with attention mechanism

**Discussion/Analysis**

* Possible model for highlighting additional areas of research

***CLASSIFYING RADIO GALAXIES WITH CONVOLUTIONAL NEURAL NETWORK***

<https://arxiv.org/pdf/1705.03413.pdf>

**Data**

* VLA
* FIRST

**Technique**

* Convolutional neural network

**Analysis/Discussion**

* Transfer learning is possible
* CNN works but there is more work needed

***Convolutional Deep Denoising Autoencoders for Radio Astronomical Images***

<https://arxiv.org/abs/2110.08618>

**Data**

* Sky Surveys from LOFAR instrument
* Generate clean and dirty images

**Technique**

* Data denoised
  + Training program reads parameters and hyperparameters
  + Read from FITS files
  + Set minimum of emissivity
  + Normalize results
  + Divide into tiles
  + Serialized to feed network
  + Mini-batches offloaded to GPU
  + Save network once it reaches convergence
  + Load network from file
* Added random noise to sky images
* Use CNN with denoising autoencoder

**Analysis/Discussion**

* Autoencoders effective in removing random noise, identifying the faint, extended outskirts of diffused structures

***Deep Learning Approach to Photometric Redshift Estimation***

<https://arxiv.org/pdf/2310.16304.pdf>

**Data**

* Sloan Digital Sky Survey

**Technique**

* Fully Connected Neural Network
* Decision Tree

**Analysis**

* Fully connected neural network did very well

***Deep learning based detection of cosmological diffuse radio sources***

<https://academic.oup.com/mnras/article-pdf/480/3/3749/25519514/sty2102.pdf>

**Data**

* Cosmodeep Dataset: <https://cosmosimfrazza.myfreesites.net/cosmodeep-training-datasets>
  + Sample dataset of sky models for training CNN Cosmodeep
  + Metadata
    - 61 images (.fits) data
    - 2000 x 2000 images

images seem to have 1 channel

***Deep-Learnt Classification of Light Curves***

<https://arxiv.org/pdf/1709.06257.pdf>

**Data**

* Catalina Real Time Transient Survey

**Technique**

* Transform time series to two-dimensional light curve representations in order
* Pushed processed data via CNN

**Analysis**

* Does not require techniques such as dimensionality reduction
* Multiple ways to improve results
* Possibility of using Generative Networks to create simulated examples

***DEEPLY UNCERTAIN: COMPARING METHODS OF UNCERTAINTY QUANTIFICATION IN DEEP LEARNING ALGORITHMS***

<https://arxiv.org/pdf/2004.10710.pdf>

**Data**

**Technique**

* Single pendulum experiment
* Deep Ensembles (DE)
* Bayesian Neural Networks (BNN)
* Concrete Dropout

**Analysis**

* Aleatoric uncertainty: data’s inherent randomness that cannot be explained
  + Wide-enough variation in noise present in training set => model not stuck predicting same relative uncertainty
    - BNN struggled with this problem
* Epistemic uncertainty: fidelity of the model
  + Very difficult to find
  + DE and BNN could detect out-of-distribution examples
* Statistical uncertainty: errors quantified by statistical analysis of a series of experimental measurements
  + Model can only infer the typical uncertainty in each region of inputs from the training set
* Systematic uncertainty: anything else
* Recommend DE
  + Performed best
  + Smallest conceptual load
  + Train network several times

***Deep Probabilistic Imaging: Uncertainty Quantification and Multi-modal Solution Characterization for Computational Imaging***

<https://arxiv.org/abs/2010.14462>

**Data**

* Radio interferometric astronomical imaging for blackholes using the event Horizon Telescope
* Lots of noise
  + Time-dependent telescope-based error
  + Telescope-based phase error
  + Baseline-based Gaussian thermal noise

**Technique**

* Deep generative models to learn a reconstructed image’s posterior distribution
  + Invertible flow-based generative model
  + Trained with MAP
* Variational bayesian method for predicting a distribution
  + Flow-based generative model for parametrizing latent sample distribution
  + Used KL divergence between generative model distribution and image posterior distribution

**Analysis/Discussion**

***Density of States Prediction of Crystalline Materials via Prompt-guided Multi-Modal Transformer***

<https://arxiv.org/pdf/2311.12856.pdf>

**Data**

* Phonon DOS
* Electron DOS

**Technique**

* DOSTransformer
  + Encode crystalline material with GNNs to learn representation of atom
  + Use cross-attention layers of multi-modal transformer to capture relationship between crystalline material and energy level
  + Energy decoder
  + Loss function is root mean squared error

**Analysis/Discussion**

* More precise prediction
* Better captures peak points
* Encoding may provide conflicting signals

***Detecting Tidal Features using Self-Supervised Representation Learning***

<https://arxiv.org/pdf/2307.04967.pdf>

**Data**

* Ultradeep layer of the Hyper Suprime-Cam Subaru Strategic Program optical imaging survey

**Technique**

* Pre-processing
  + Orientation
  + Gaussian Noise
  + Jitter and Crop
* Nearest Neighbor Contrastive Learning self-supervised learning algorithm
  + Adam optimizer pre-trained self-supervised encoder converts image to representations
* Model from Pearson et al with a couple of modifications
  + Adam optimizer

**Analysis/Discussion**

* Can isolate majority of galaxies with tidal features from a large sampling of galaxies
* SSL easy to re-train on data from different surveys
* SSL outperforms supervised model

***Fanaroff-Riley classification of radio galaxies using group-equivariant convolutional neural networks***

<https://arxiv.org/pdf/2102.08252.pdf>

**Data**

* MiraBest Dataset: <https://zenodo.org/record/4288837>
  + MetaData:
    - 1256 images (batches with 157 images each)
    - 150 x 150 pixels
    - Seems like 1 channel

**Technique**

* LeNet CNN architecture

**Analysis/Discussion**

* Distribution of data based on rotation of radio galaxies
* Equivariant CNN’s converge more rapidly

***Gravity Spy: Integrating Advanced LIGO Detector Characterization, Machine Learning, and Citizen Science***

<https://arxiv.org/pdf/1611.04596.pdf>

**Data**

* LIGO

**Technique**

* Combines crowdsourcing and machine learning in order to categorize glitches into morphological classes
  + Discover new classes as detectors evolve

**Analysis/Discussion**

* Helpful in aiding LIGO detector characterization
* “Scalable”; able to add new classes
* Large quantity of data may render crowdsourcing unavailable

***IMAGE AS FIRST-ORDER NORM+LINEAR AUTOREGRESSION: UNVEILING MATHEMATICAL INVARIANCE***

<https://arxiv.org/pdf/2305.16319.pdf>

**Data**

* ImageNet

**Technique**

* FINOLA: First-Order Norm + Linear Autoregression
  + Generates W x H feature map autoregressively
  + First-order process => predicting position using only immediately previous neighbor
    - Two separate linear models for predicting independently x and y

**Analysis/Discussion**

* Presence of partial differential equations governing latent feature space
* Successful image reconstruction
* Does not require extensive fine-tuning

***New Approaches to Object Classification in Synoptic Sky Surveys***

<https://arxiv.org/pdf/0810.4945.pdf>

**Data**

* Palomar-Quest (PQ) Survey

**Technique**

* Multilayer Perceptron
  + Softmax Activation Function
  + Cross-Entropy Error
* Performance of classifiers criteria
  + Completeness
  + contamination
  + Overall classification rate

**Analysis**

* Improved performance obtained by combining models together

***Machine Learning for the Zwicky Transient Facility***

<https://arxiv.org/ftp/arxiv/papers/1902/1902.01936.pdf>

**Data**

* Zwicky Transient Facility dataset

**Technique**

* Separate bogus from real candidates
  + Used Real/Bogus classifier
  + Used verification
* Classification
  + Random forest
  + CNN

**Analysis**

* ZTF is good

***Model-Aware Contrastive learning: Towards Escaping Uniformity-Tolerance Dilemma in Training***

<https://arxiv.org/pdf/2207.07874.pdf>

**Problem**

* Uniformity-tolerance dilemma
  + Smaller temperature can cause larger penalties on high similarity regions
  + Larger temperature tends to give all negative samples equal magnitude of penalties

**Technique**

* Model-Aware Temperature Strategy
  + Scaling factor alpha
  + Fine-grained adjusting approach
* Gradient Reduction Dilemma
  + Computationally costly
  + Two propositions
    - As K approaches positive infinity, W approaches 1
    - If tau approaches positive infinity, then Wi approaches its bound K/(K+1)

**Analysis/Discussion**

* New approaches were very helpful in making contrastive learning better

***ParSNIP: Generative Models of Transient Light Curves with Physics-Enabled Deep Learning***

<https://arxiv.org/pdf/2109.13999.pdf>

**Data**

* Pan-STARRS1 Data: <https://outerspace.stsci.edu/display/PANSTARRS/>
  + Dataset for Pan-STARRS
  + MetaData
    - 2885 light curves with host-galaxy redshifts; 557 have spectroscopically-confirmed types
    - 60 Orthogonal Transfer Arrays devices (OTA); each device has 8 x 8 array of “cells”; single OTA format is 4846 x 4868 pixel array; each device has 64 cells where each cell is 590 x 598 pixels
    - Seems grayscale (1 channel) when processed through a bandpass; full display uses 3 channels
* PLAsTiCC Dataset from Kaggle: <https://www.kaggle.com/code/michaelapers/the-plasticc-astronomy-starter-kit>
  + Dataset from Kaggle for astronomy purposes
  + Metadata
    - Tabular data or time series
    - First table: 12 features; Second Table: 6 features
    - Test set: ~3.5M objects; Training Data: 8000 sources, maybe ~3.5M objects

**Technique**

* Generative model
  + Process
    - Neural network to predict intrinsic spectra of a given transient as a function of three intrinsic latent variables
    - Passed through a physics layer
* Variational autoencoder

**Analysis**

* Can estimate distance to well-observed light curves of SNe Ia with some uncertainty
* Good results for photometric classification
* Stable
* Parsnip software

***Photometric redshifts for Quasars in multi band Surveys***

<https://arxiv.org/pdf/1305.5641.pdf>

**Data**

* SDSS - DR7
* GALEX - DR6/7
* UKDSS - DR9
* WISE

**Technique**

* Multi layer perceptron trained by a learning rule based on the Quasi Newton Algorithm

**Analysis**

* Multi layer perceptron with QNA seemed to perform well

***Pulsar Candidate Identification Using Semi-Supervised Generative Adversarial Networks***

<https://arxiv.org/pdf/2010.07457.pdf>

**Data**

* HTRS-Survey: <https://sites.google.com/site/htrupublications/htru-discoveries>
  + Has primarily medlat and hilate
  + Metadata
    - >100 pulsar discoveries
    - Tabular data

**Technique**

* CNN on the “Time-Phase” and “Freq-Phase” features and MLP for “DM-Curve” and “Pulse Profile” features
* Semi-supervised GAN model
  + Parts
    - Supervised discriminator
    - Unsupervised discriminator
    - Unsupervised generator
  + First output layer solves unsupervised task and outputs REAL/FAKE
  + Second output layer outputs if signal is from pulsar or not

**Analysis**

* SGAN has longer training time
* Static trained supervised model is not optimal approach

***Quantifying Uncertainty in Deep Learning Approaches to Radio Galaxy Classification***

<https://arxiv.org/pdf/2201.01203.pdf>

**Data**

* MiraBest Dataset: <https://zenodo.org/record/4288837>
  + MetaData:
    - 1256 images (batches with 157 images each)
    - 150 x 150 pixels
    - Seems like 1 channel

**Technique**

* Variational inference
  + Reduces Bayesian inference to optimization problem
  + BNN

**Analysis/Discussion**

* Model trained with Laplace prior performs best
* Bayes by BackProp di well labeled radio galaxies
* Method that combines Fisher information with weight magnitude allows more weights to be pruned without compromising performance

***Radio Galaxy Zoo: Leveraging laten space representations from variational autoencoder***

<https://arxiv.org/pdf/2311.08331.pdf>

**Data**

* Radio Galaxy Zoo Data Release 1
* MiraBest
* FR-Deep NVSS

**Technique**

* Very deep variational autoencoder (VDVAE)
  + Encoder
    - 6 stages
  + Decoder
    - 6 stages
  + Prior
    - Computed by residual block
  + Used elbo method
  + Residual block computes prior, posterior, and latent variable
  + RMSProp optimizer
* SimCLR method used as contrastive learning algorithm
  + Two stochastic transformations to an image
* BYOL considered
* SimSiam method considered

**Analysis/Discussion**

* Model able to recover inputs
* Classifiers trained on VDVAE latent codes achieved better accuracy
* Learned representation used for similarity search
* Decoder is capable of generating new images

***Radio Galaxy Zoo: Towards building the first multi-purpose foundation model for radio astronomy with self-supervised learning***

<https://arxiv.org/abs/2305.16127>

**Data**

* MiraBest

**Technique**

* Bootstrap Your Own Latent (BYOL) self-supervised learning algorithm
  + Instance differentiation
  + Trained to minimize the distance between different random augmentations of same image in representation space of the encoder
  + Momentum encoding to calculate positive pair losses
  + Loss is means squared error between representations of momentum cncoder and online encoder
  + ResNet architecture to train supervised baseline and self-supevised model
* Heavily used augmentations

**Analysis/Discussion**

* Outperforms baseline at all label volumes
* Very applicable
* Would like to wider range of data in the future

***Separation of pulsar signals from noise using supervised machine learning algorithms***

<https://arxiv.org/pdf/1704.04659.pdf>

**Data**

* HTRU Survey
* SMOTE Dataset

**Technique**

* Artificial Neural Networks
* Adaboost
* Gradient Boosting Classifier (GBC)
* eXtreme Gradient Boosting

**Analysis**

* Tree-based algorithms yield feature importance

***Transfer Learning for Radio Galaxy Classification***

<https://arxiv.org/pdf/1903.11921.pdf>

**Data**

* NVSS radio survey
* FIRST radio survey

**Technique**

* Network architecture
* Similar to AlexNet CNN but with 13 layers
  + Removed network component for parallel computing
  + Fully connected layer to reduce over-fitting
  + Adaptive mini-batch optimizer AdaGrad
* Transfer learning used

**Analysis**

***VoLTA: Vision-Language Transformer with Weakly-Supervised Local-Feature Alignment***

<https://arxiv.org/pdf/2210.04135.pdf>

**Data**

**Technique**

* VoLTA
  + Intra- and inter-modality redundancy reduction
    - Barlow Twins (BT) => foundational objective of VoLTA
  + Weakly-supervised cross-modal alignment of local features
    - OT-based learning methods
    - GOT => intricate images
      * WD for node matching
      * GWD for edge matching
  + Cross-modal attention fusion
    - BT and GOT computed in dual encoder settings

**Analysis/Discussion**

* Achieves fine-grained region-level image understanding
* Good low-level matching criterion
* Effective on wide range of course- and fine-grained tasks

***Weight Pruning and Uncertainty in Radio Galaxy Classification***

<https://arxiv.org/pdf/2111.11654.pdf>

**Data**

* MiraBest Dataset: <https://zenodo.org/record/4288837>
  + MetaData:
    - 1256 images (batches with 157 images each)
    - 150 x 150 pixels
    - Seems like 1 channel

**Technique**

* Expanded LeNet-5 architecture
* Used variational inference approach using Adam optimizer
  + Prior distributions
    - Gaussian Mixture Model
    - Laplace Prior
    - Simple Gaussian
    - Laplace Mixture Model

**Analysis/Discussion**

* BNN with variational inference useful for returning degree of model confidence in an individual classification
* Key elements include availability of radio astronomy data