FNet II for Spectral Classification: A Deep Learning Approach Beyond Line Identification

Fatemeh Rastegar Nia

International Center for Relativistic Astrophysics Network (ICRANet)

In collaboration with:

R. Moradi, M. T. Mirtorabi, Y. Wang, A. Vafaei Sadr

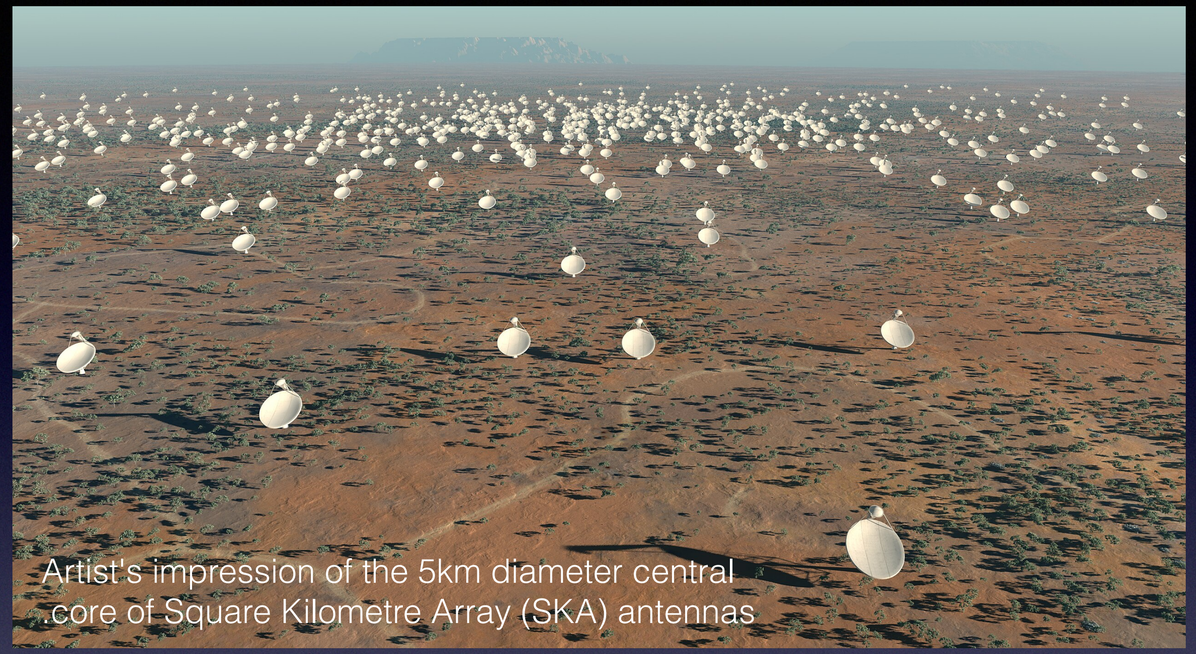
7th Galileo Xu-Guangqui meeting

July 7-11 2025 (Pescara, Italy)

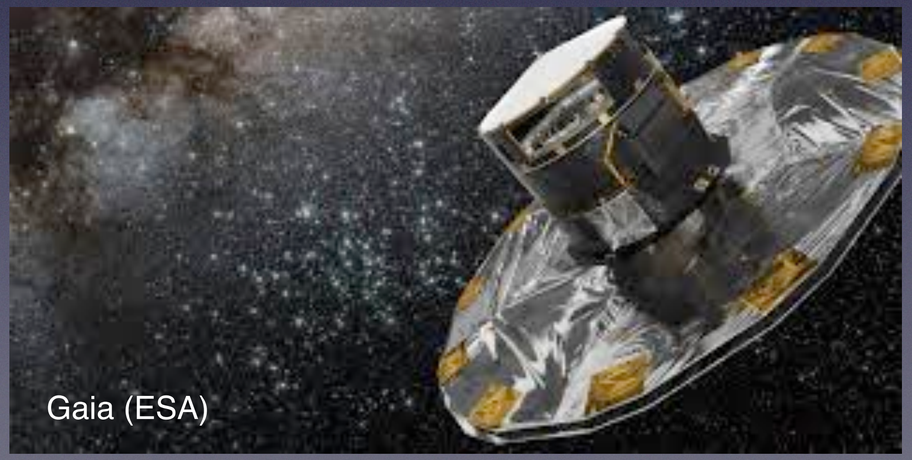
Outlines

- Motivation and Introduction
- Data and its challenges
- Convolutional Neural Networks (CNN)
- What is FNet?
- Results and Perspectives

The Data Revolution in Astronomy


Artist's impression of the 5km diameter central
.core of Square Kilometre Array (SKA) antennas


Vera C. Rubin Observatory


Gaia (ESA)


Euclid (ESA)


Pan-STARRS

Project Name	Type	Total Data Volume	Daily Data Production	Key Features & Notes
Square Kilometre Array (SKA)	Radio Telescope	~4.6 Exabytes	~150 TB/day	World's largest radio telescope; aims to address fundamental questions about the universe.
Vera C. Rubin Observatory (LSST)	Optical Survey Telescope	~60 Petabytes	~20 TB/night	Equipped with a 3.2-gigapixel camera; will conduct a 10-year survey capturing 10 million celestial events nightly.
Gaia (ESA)	Space Observatory	~200 TB	~5 GB/day	Mapping over a billion stars to create a precise 3D map of the Milky Way.
Euclid (ESA)	Space Telescope	~220 TB (est.)	~100 GB/day	Investigating dark energy and dark matter by mapping the geometry of the dark universe.
Pan-STARRS	Optical Survey Telescope	~40 Petabytes	~5 TB/day	Conducts continuous sky surveys to detect moving or variable objects.
LAMOST (China)	Spectroscopic Telescope	~10 TB (est.)	~20 GB/day (est.)	Has gathered data on over 7 million stars, creating the world's largest database of stellar spectra.
FAST (China)	Radio Telescope	~20 Petabytes	~55 TB/day	The world's largest filled-aperture radio telescope; focuses on pulsar observations and neutral hydrogen studies.
HERA (South Africa)	Radio Telescope Array	~5 Petabytes	~50 TB/night	Designed to detect the 21 cm hydrogen line from the early universe, aiding in understanding cosmic reionization.
Desert Fireball Network (DFN)	Meteor Observation Network	~0.5 Petabytes	~1.4 TB/day	Monitors meteorite falls using a network of cameras across Australia.

How Do We Classify the Universe?

Classical & Supervised Learning

Template Matching (TM)
Match to known patterns

KN N **K-Nearest Neighbors** (KNN)
Classify by proximity

SVM **Support Vector Machines** (SVM)
Learn boundaries between classes

DTE **Decision Trees & Ensembles**
Tree-based logic & combined models

Random Forests / Gradient Boosting
High-performance esembles

Deep Learning & Neural Approaches



Artificial Neural Networks (ANN)
Brain-inspired learning



Convolutional Neural Networks (CNN)
Pattern extraction from spectra/Images



Transformers
Discriminates In sequential & spectral data

Emerging & Hybrid Methods



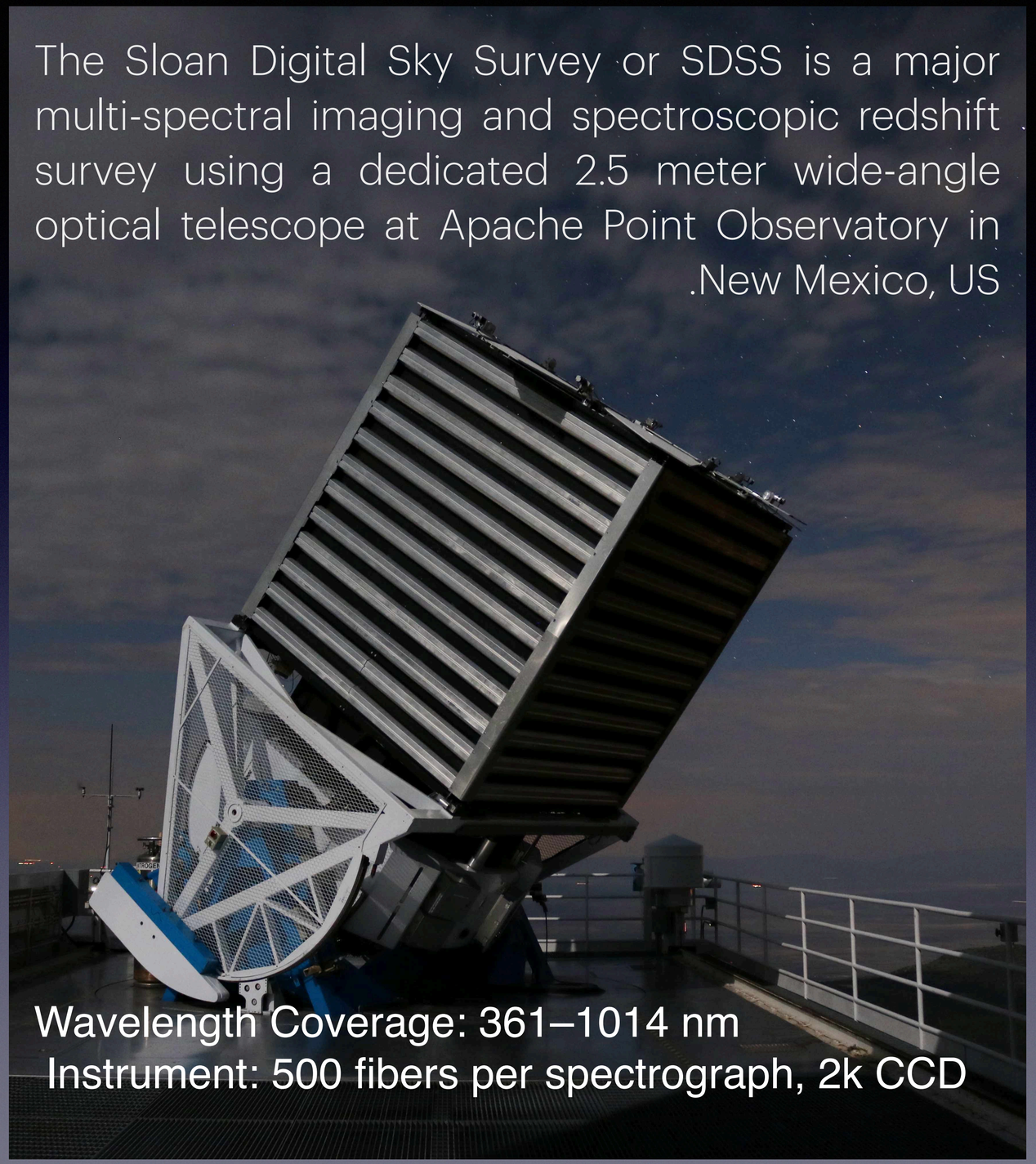
Ranking-Based Models
Graph and manifold learning



Unsupervised & Semi-Supervised Learning
For unlabeled, high-volume data

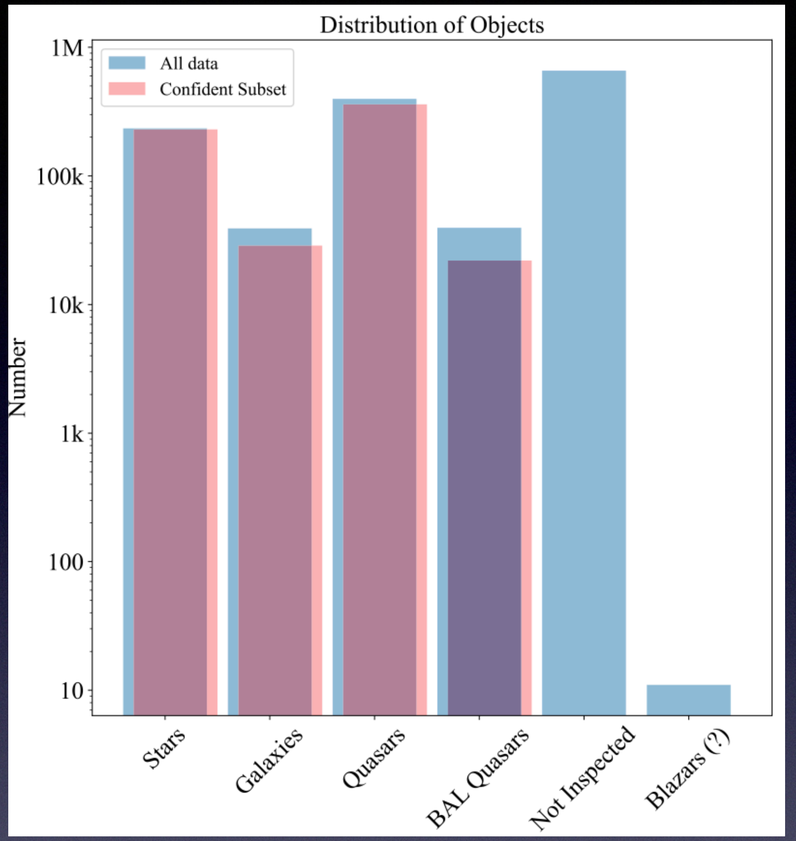
Dataset and Sample:

- For our work, we have used the seventeenth data release (DR17) of the Sloan Digital Sky Survey (SDSS).
- DR17 includes over 623 terabytes of data on the Science Archive Server (SAS).
- It features around 1.5 million unique spectra and objects.
- Our study specifically focuses on the eBOSS (extended Baryon Oscillation Spectroscopic Survey) catalogue.
- The eBOSS DR17 dataset alone contains 1.4 million high-quality spectra, forming the foundation of our current research.



Data Cleaning for High-Confidence Training

1. **ZCONF = 3** → High-confidence redshift
2. **BAL_PROB ≥ 0.9** → Reliable BAL Quasar detection
3. **ZVI > 0** → Only visually inspected spectra
4. **IS_QSO_FINAL** → **Quasars: = 1**
   → Galaxies/Stars: ≠ 1

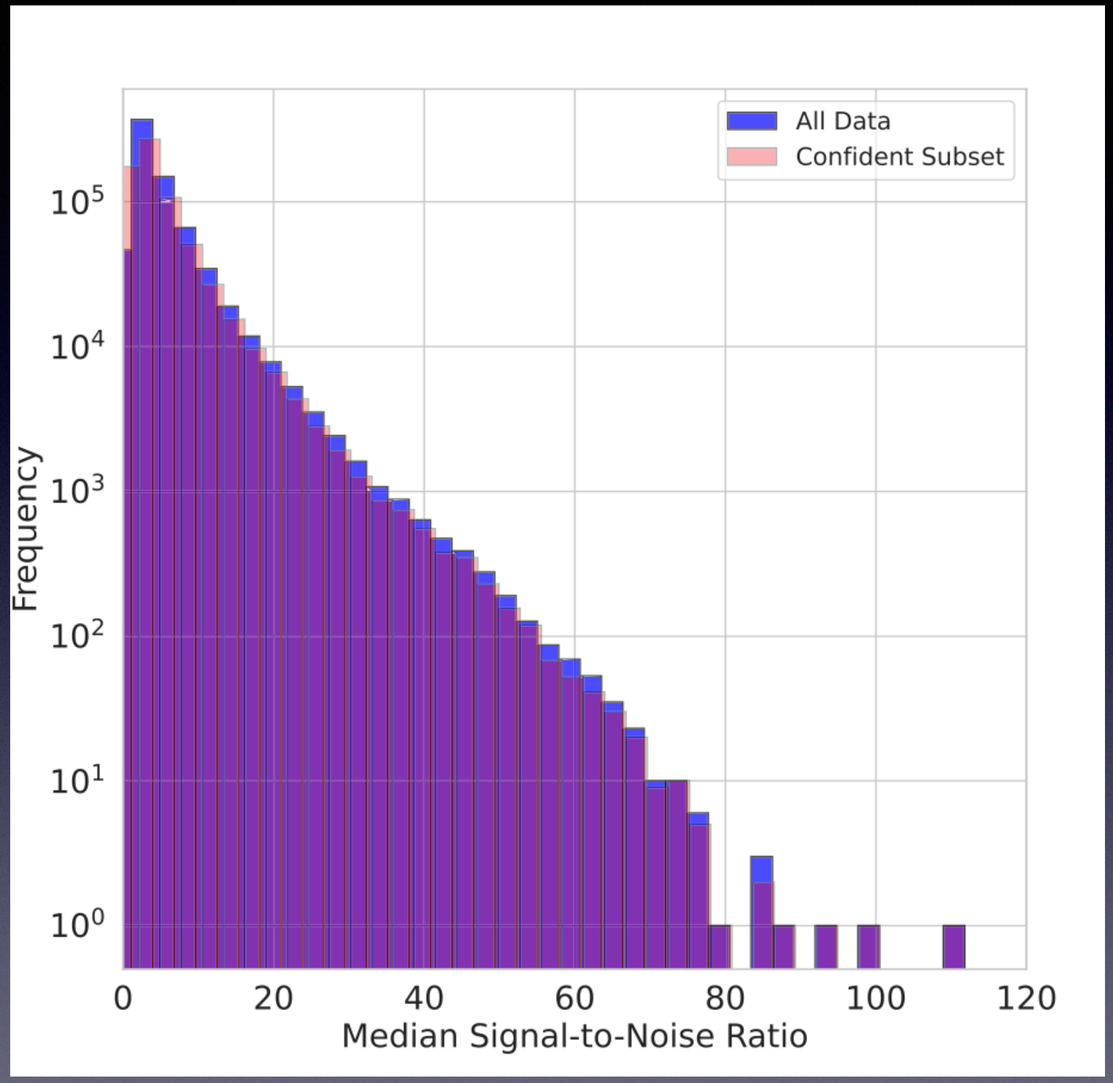


5. **Excluded: `AUTOCLASS_DR14Q = VI` (35,268 spectra with low SNR / ambiguous cl assification)**

Stars = 233 859
Galaxies = 39 054
Quasars = 396 843
BAL Quasars = 39 501
Not inspected objects = 658 020
Objects indicate potential blazars = 11

After cleaning

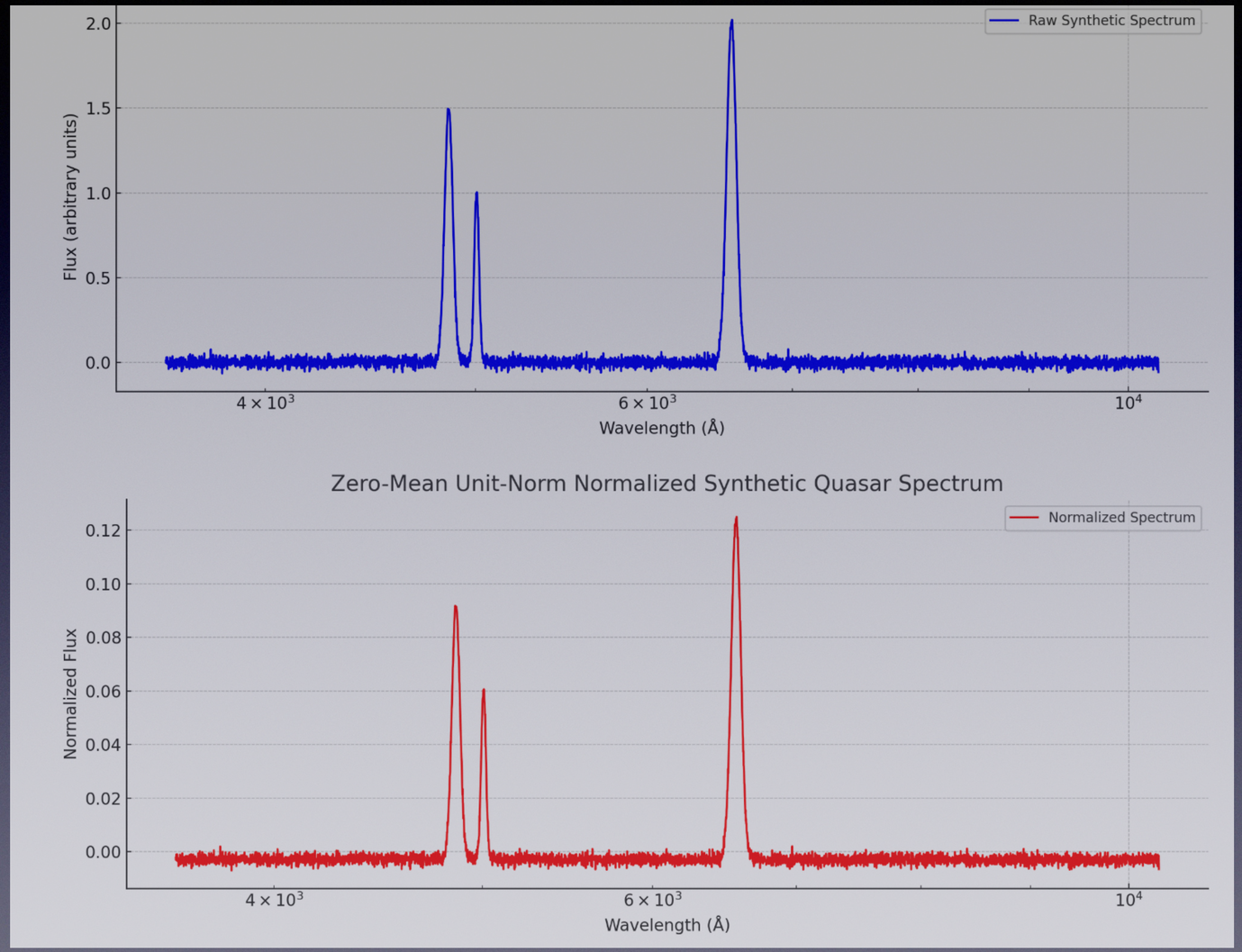
Stars = 229,647
Galaxies = 28,688
Quasars = 359,609
BAL Quasars = 21,962



*"The criteria we used to build our training and test sets did **not** impose any constraints on signal-to-noise ratio, or SNR."*

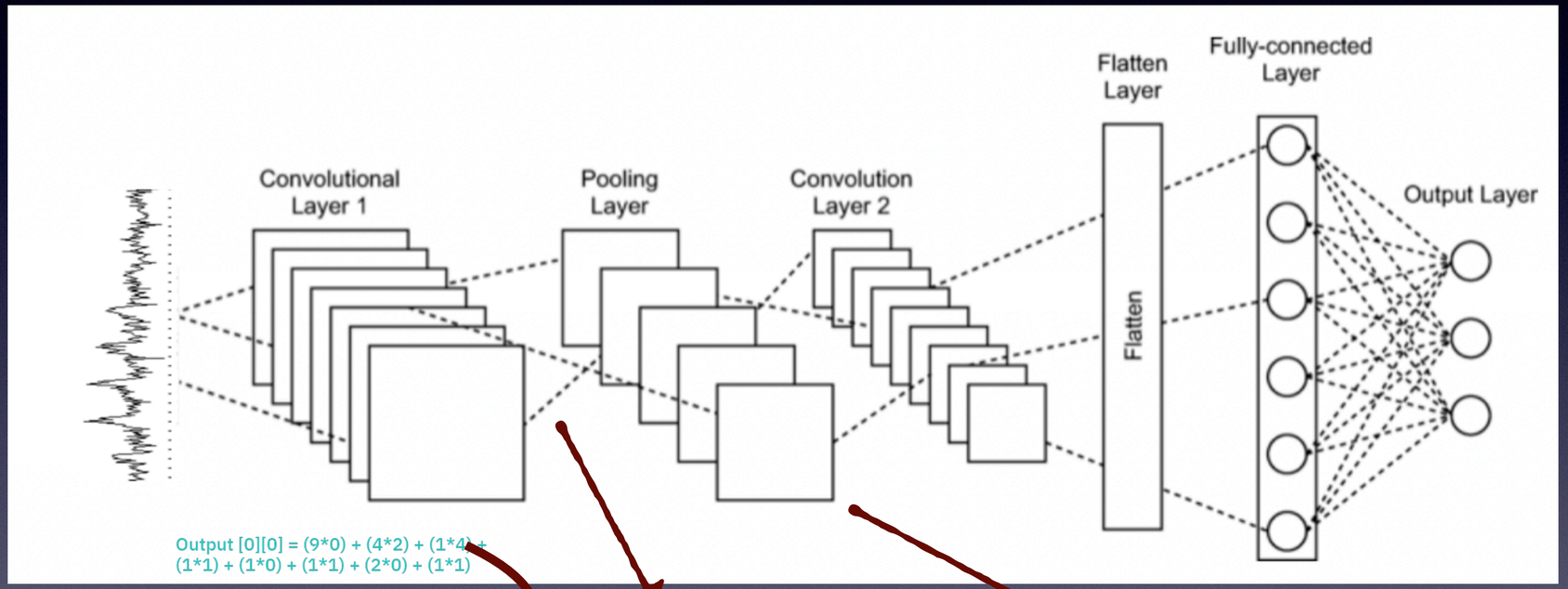
Data Preprocessing

Spectral Representation	Each spectrum is treated as a 1D time series, representing flux versus wavelength.
Standardization	Spectra are interpolated to a uniform grid of 4,618 data points (pixels) across the logarithmic wavelength range of 360–1032.5 nm.
Normalization	Applied zero-mean and unit-norm normalization to each spectrum to stabilize and accelerate DL training.
Data Structuring	Processed spectra are stored in an n × 4,618 matrix (n = number of quasars), with corresponding normalized redshifts in an n × 1 label matrix.



How a Convolutional Layer Works?

- *The basic idea behind CNNs comes from how the human visual system works.*
- *Instead of telling the model where to look, we let it learn what matters.*



Output [0][0] = (9*0) + (4*2) + (1*4) + (1*1) + (1*0) + (1*1) + (2*0) + (1*1)

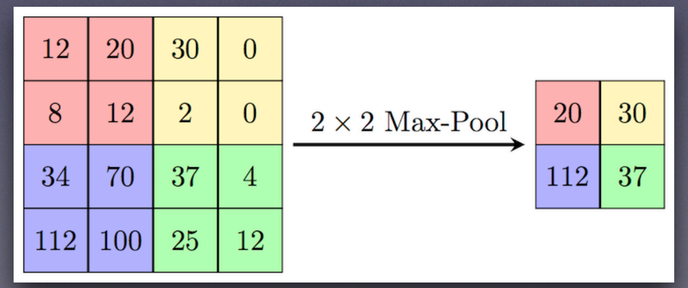
= 0 + 8 + 1 + 4 + 1 + 0 + 1 + 0 + 1

= 16

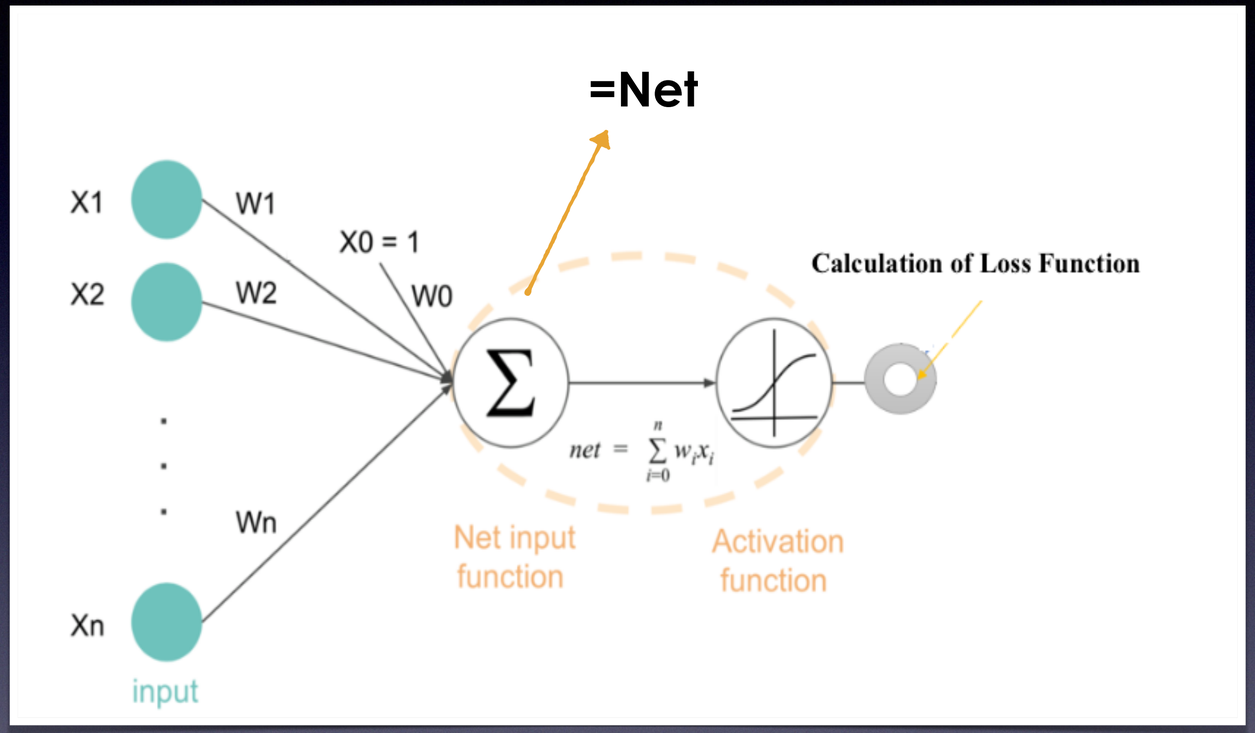


Activation functions decide whether a neuron should be .activated

$$\mathrm{ReLU}(x) = \max(0, x)$$



Fully Connected Layers (Dense)



Loss Function: Measuring the Model's Mistakes

- Quantifies the difference between predicted and true labels.
- Guides the learning process via **back-propagation**.
- Common loss for classification: **Cross-Entropy Loss**
- Lower loss → better model performance.
- Our model uses **Categorical Cross-Entropy** for multi-class output.

$$-\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)\right]$$

← Cross- entropy for classification

$$\frac{1}{N}\sum_{i=1}^{N}(y_i-\hat{y}_i)^2$$

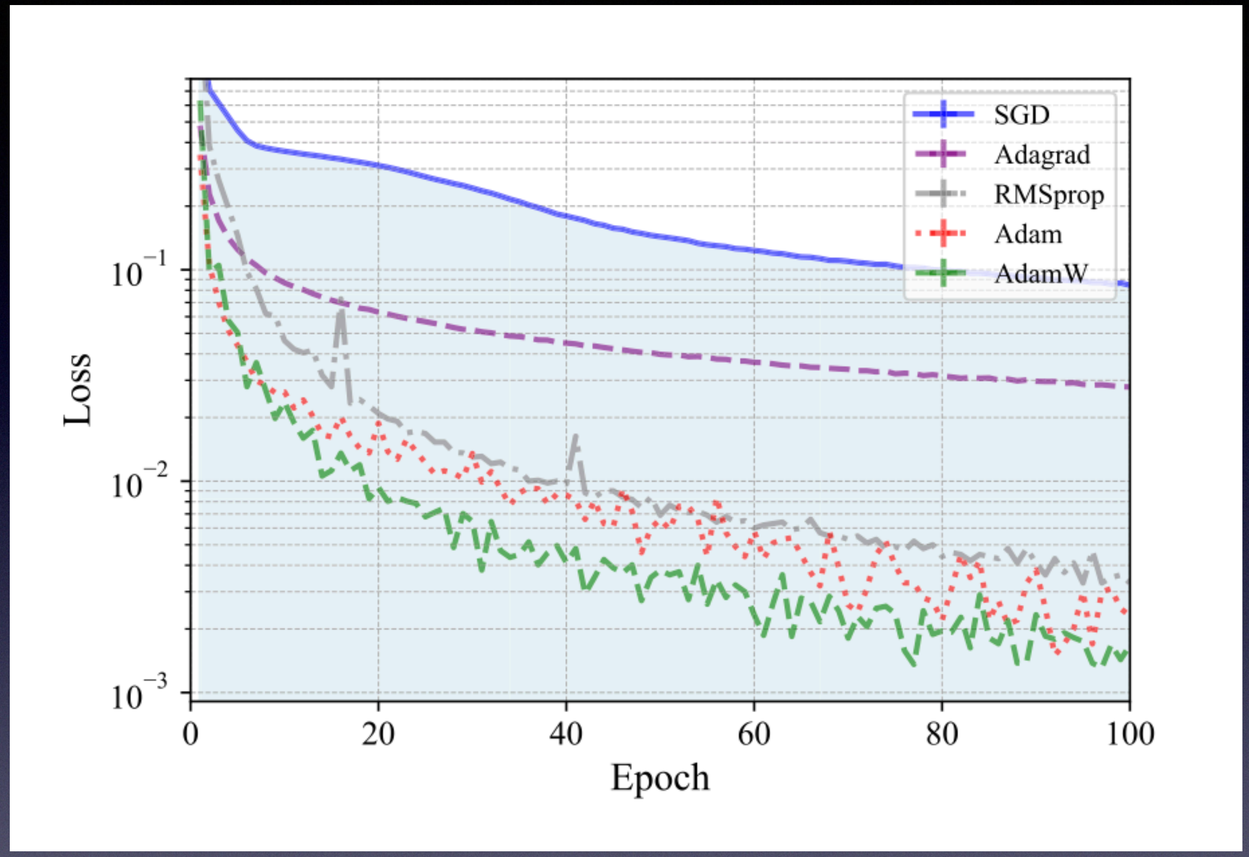
← Mean Square Error for regression

Optimization Method — AdamW

*The **optimizer** updates the model's weights to .**minimize the loss***

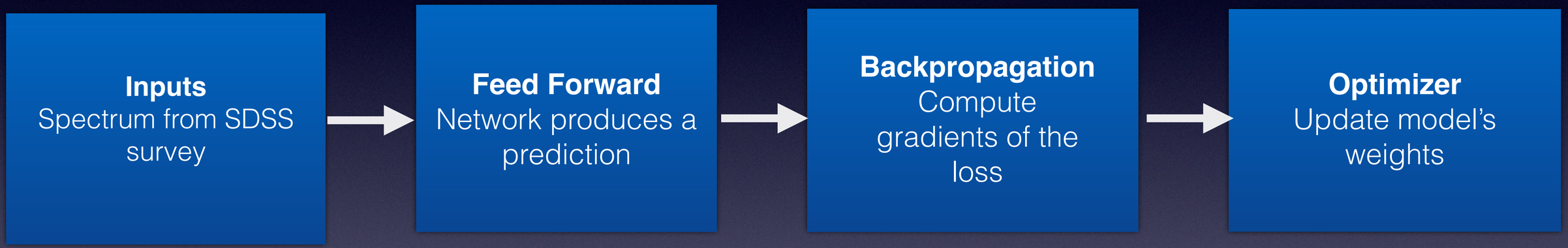
*:Our work uses **AdamW***

- *.A variant of **Adam** optimizer*
- *Handles **weight decay** separately from .gradients*
- *.Combines strengths of **AdaGrad** + **RMSProp***
- *.Faster convergence*
- *.Better generalization*
- *.Effective in high-dimensional, noisy datasets*



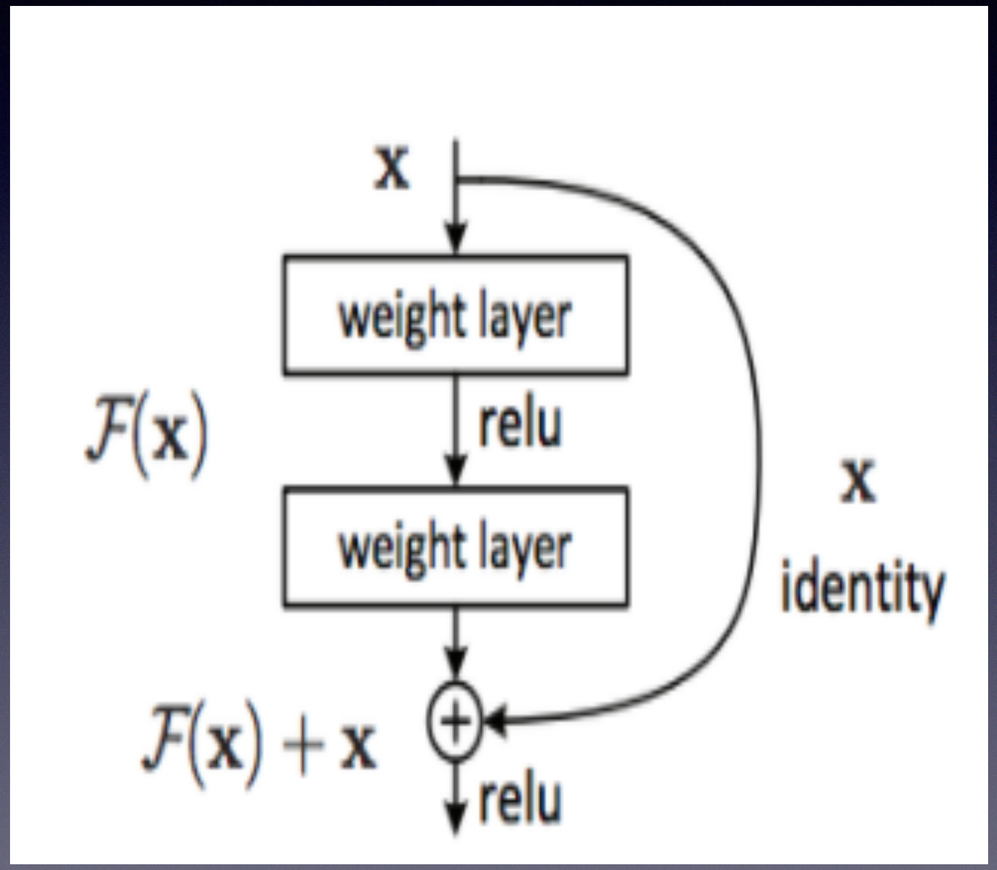
The loss decay (on a logarithmic scale) of `FNet` for 40 000 spectra of DR17Q catalogue over 100 epochs for the `AdamW`, `Adam`, `SGD`, `AdaGrad`, and `RMSProp` optimizers is presented. The `AdamW` optimizer leads to a more rapid decline in the loss function, resulting in more accurate prediction.

The Learning Process



ResNet

Vanishing/Exploding gradient



It is easier to optimize the residual mapping than to optimize the original, unreferenced .mapping

The formulation of F(x) +x can be realized by feedforward neural networks with “shortcut connections”. Shortcut connections are those skipping one or .more layers

Identity shortcut connections add neither extra parameter nor computational complexity. The entire network can still be trained end-to-end by Adam with .backpropagation

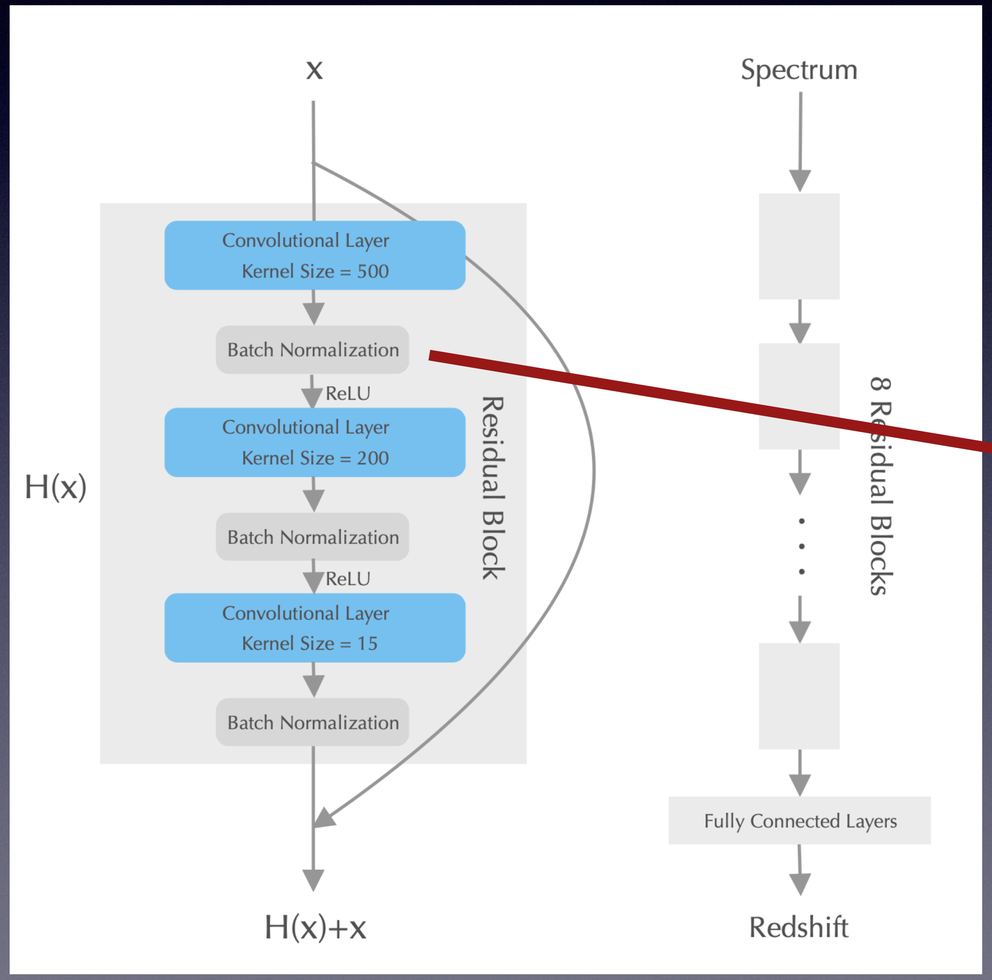
He et.al. (2015)

?What is FNet? And how does it work



?What is FNet

- **FNet I (2022)**: Redshift estimation for .quasars, $|\Delta v| < 6000$ km/s accuracy: 97%



FNet is a deep, one-dimensional convolutional neural network, designed specifically to estimate the .**redshifts of quasars** from their spectral flux

The architecture consists of **24 convolutional layers** grouped into residual blocks. Each block combines **large, medium, and small kernels** to capture both broad and fine-grained features across the .spectrum

To maintain the gradients of all layers in an appropriate range and

prevent of gradient vanishing

Hyperparameter Tuning

- **The number of layers** → 24 Convolutional layer, 8 Residual Block
- **The size of filters** → 500,200,15
- **The number of convolutional filters** → 32-64
- **Network Weight Initialization** → He Normal initializer (He et al. 2015)
- **Number of Epoches** → 100
- **Activation Function** → Rectified Linear Unit (ReLU)

Hyperparameter Tuning

We use Adam optimizer :

First moment estimate, which tracks the mean of the gradients → $m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t$

Second moment estimate, which tracks the uncentered variance of the gradients → $v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2$

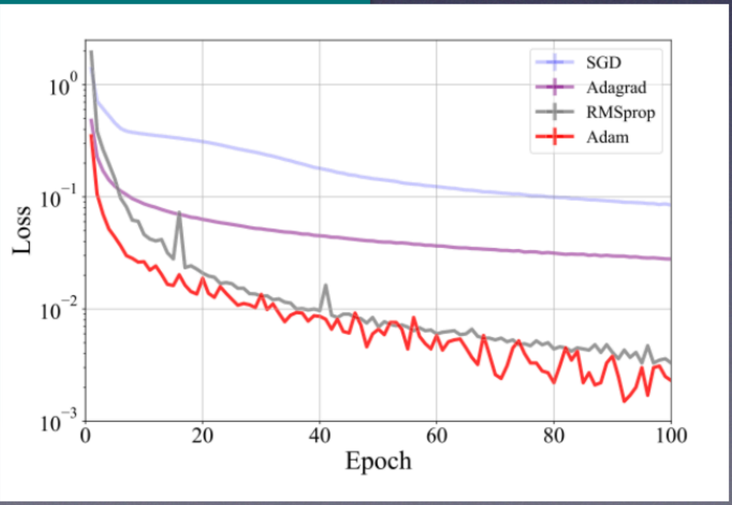
β_1=0.9

β_2=0.999

ϵ= 10^{-5} → weight decay

the step learning rate is used, namely after some epochs lr is reduced by an order of 0.1.

Initial *lr*	$lr_1 = 0.00024$
After 75 epoch	$lr_2 = 0.1\ lr_1$
After 85 epoch	$lr_3 = 0.01\ lr_1$
After 90 epoch	$lr_4 = 0.001\ lr_1$



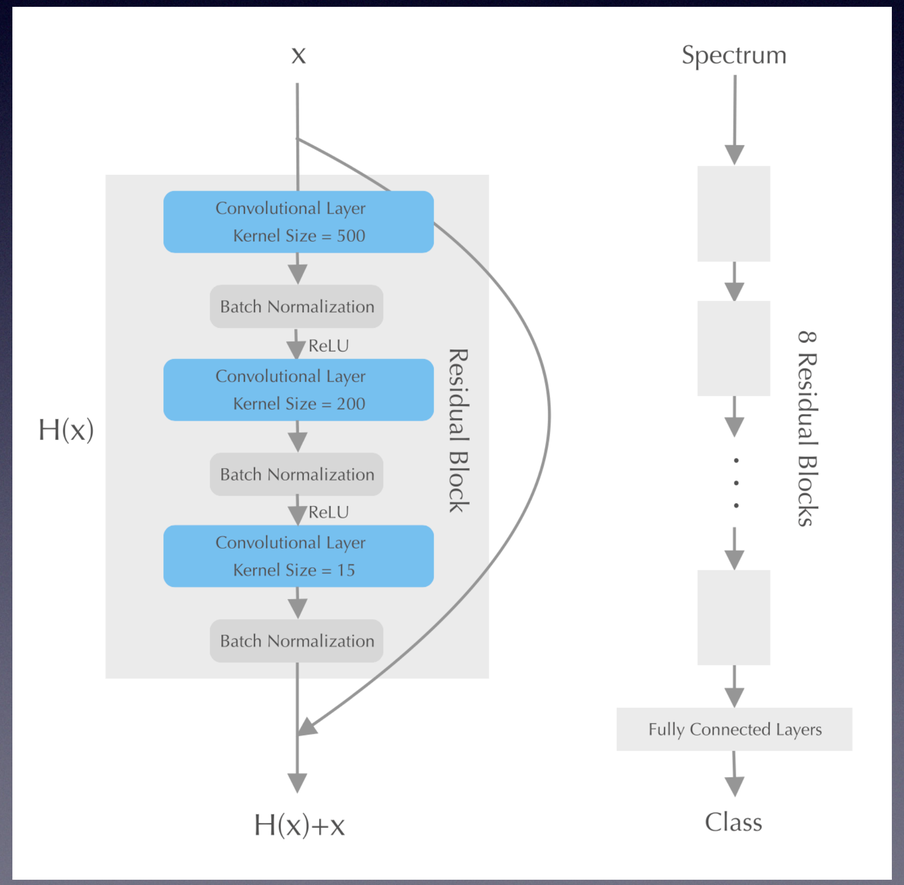
FNet II as an Effective Classifier for 1D Spectral Data

- **FNet II (2024)**: Expanded to classify stars, galaxies, BAL and non-BAL quasars.

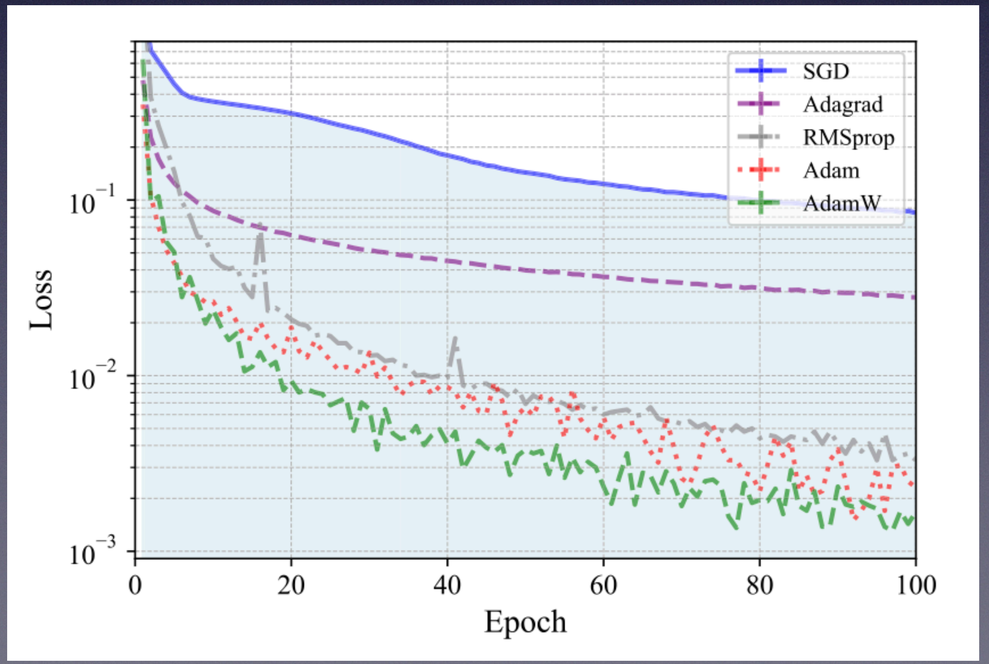
1. Output Layer Redesign
"First, we replaced the original single output neuron with a **softmax layer** that outputs class probabilities.
Now, instead of estimating redshift, the network assigns likelihoods for the spectrum being a **galaxy, quasar, BAL quasar, or star**."

2. Loss Function Update
"Next, we swapped the regression loss function with one suitable for classification: **categorical cross-entropy**.
This allows the network to be penalized more heavily for wrong classifications — especially when it's confident but wrong — and to learn better decision boundaries between classes."

3. Training Data Reconfiguration
"We also expanded the dataset.



Feature	Mean Squared Error (MSE)	Cross-Entropy (CE)
Definition	$\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$	$-\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)\right]$
Typical Use	Regression (continuous targets)	Classification (probabilistic outputs)
Error Sensitivity	Penalizes large errors quadratically → heavy on outliers	Penalizes confident wrong predictions strongly



Handling Imbalance in Data

Stars = 229,647

Galaxies = 28,688

Quasars = 359,609

BAL Quasars = 21,962

Imbalanced training led to:

- **High completeness** for stars (98.8%) and quasars (98.3%)
- **Low completeness** for galaxies (86.9%) and BAL quasars (78.8%)

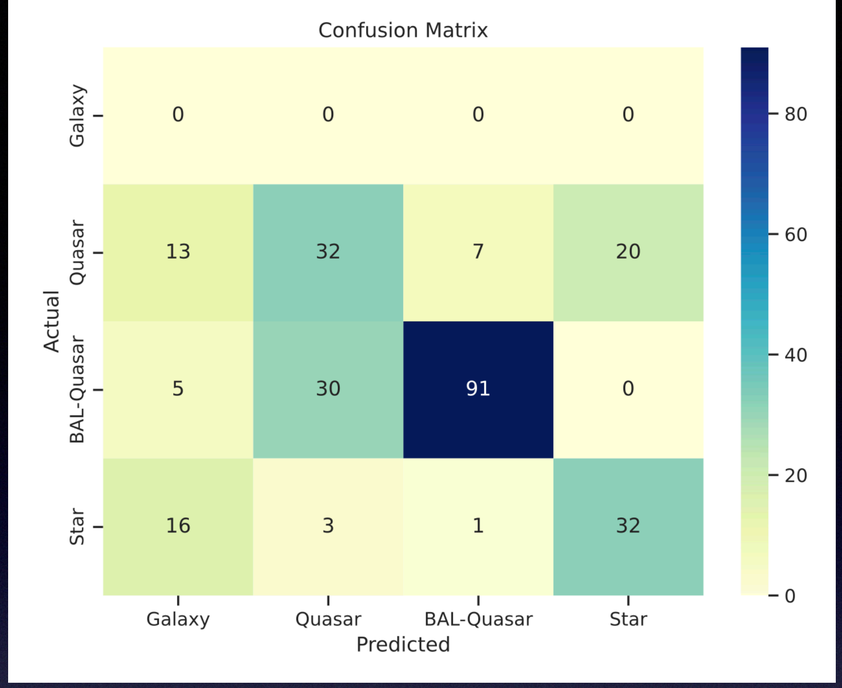
✅ **Our Strategy:**

- **Class balancing by equal sampling**
- Trained FNet using **equal number of spectra (N = 26,667)** for each class

Problem of BAL-quasar identification

Challenge: Absorption troughs conceal emission lines.

- **QuasarNET Limitation**: Relies on BAL_CIV tagging.
- **FNet Strategy**:
  - 500-wide kernels detect broad troughs.
  - Learns BAL patterns without manual tagging.



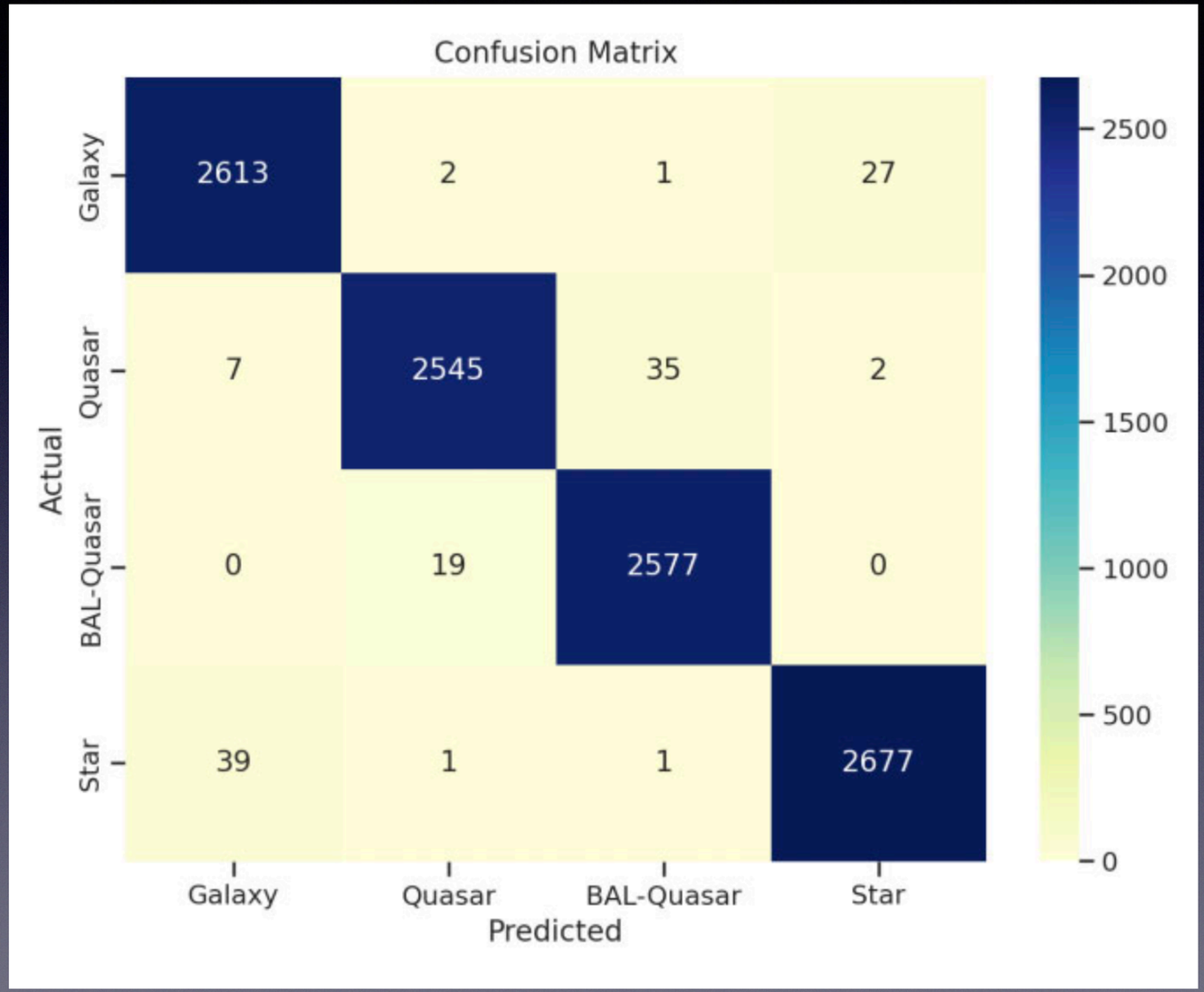
- Ambiguous sample: ~2% of SDSS spectra with low annotation confidence (Z_CONF <3)
- Galaxies in this sample are negligible and omitted

QuasarNet vs. FNet

- **QuasarNet**
  - Classifies only 20% of ambiguous spectra as quasars with reliable redshifts
- **FNet**
  - **Quasar completeness**: 45%
  - **BAL−quasar completeness**: 72%
  - **Star completeness**: 62%

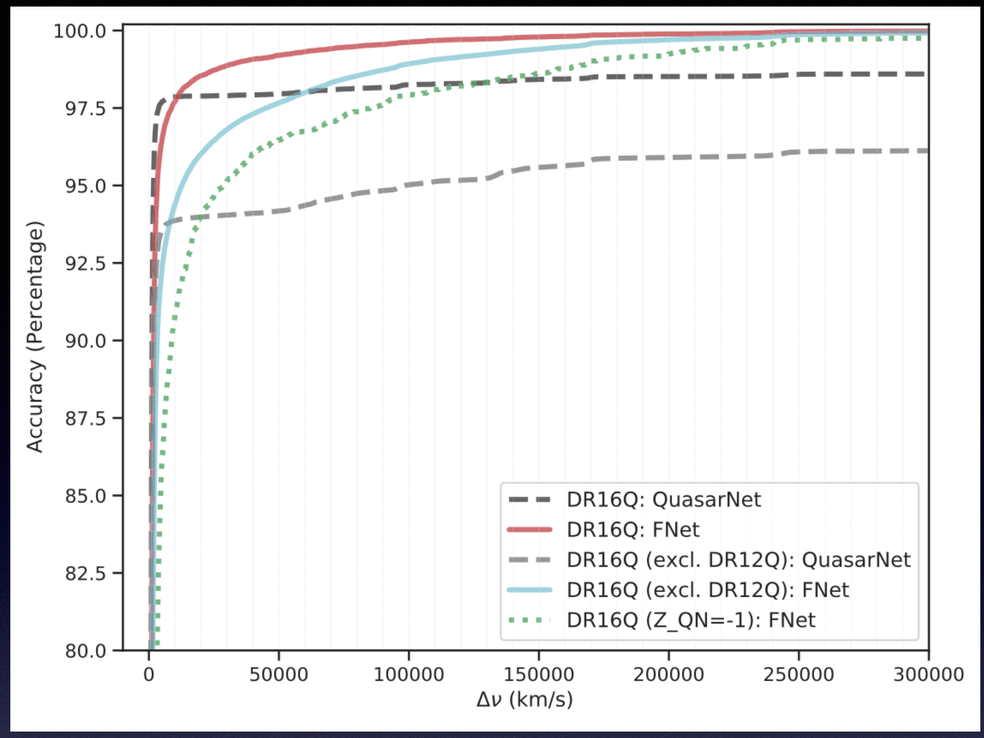
Classification Results (FNet II)

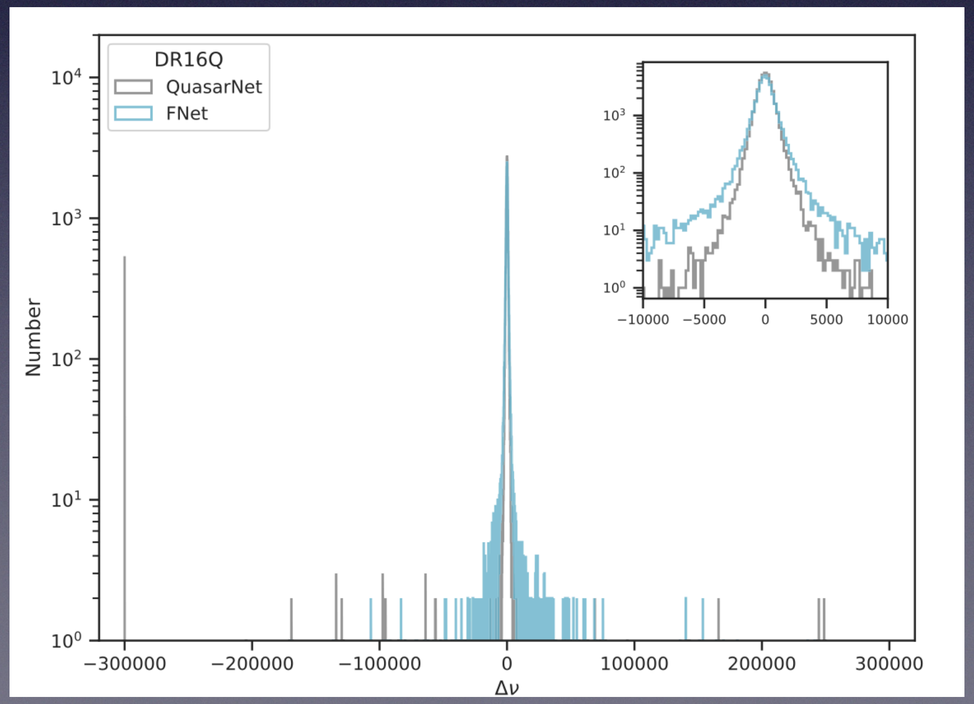
- **Completeness Metrics**:
  - Galaxies: 99.00%,
  - Quasars: 98.50%,
  - BALs: 99.00%,
  - Stars: 98.80%
- **Comparison**: Matches or outperforms QuasarNET across all classes.

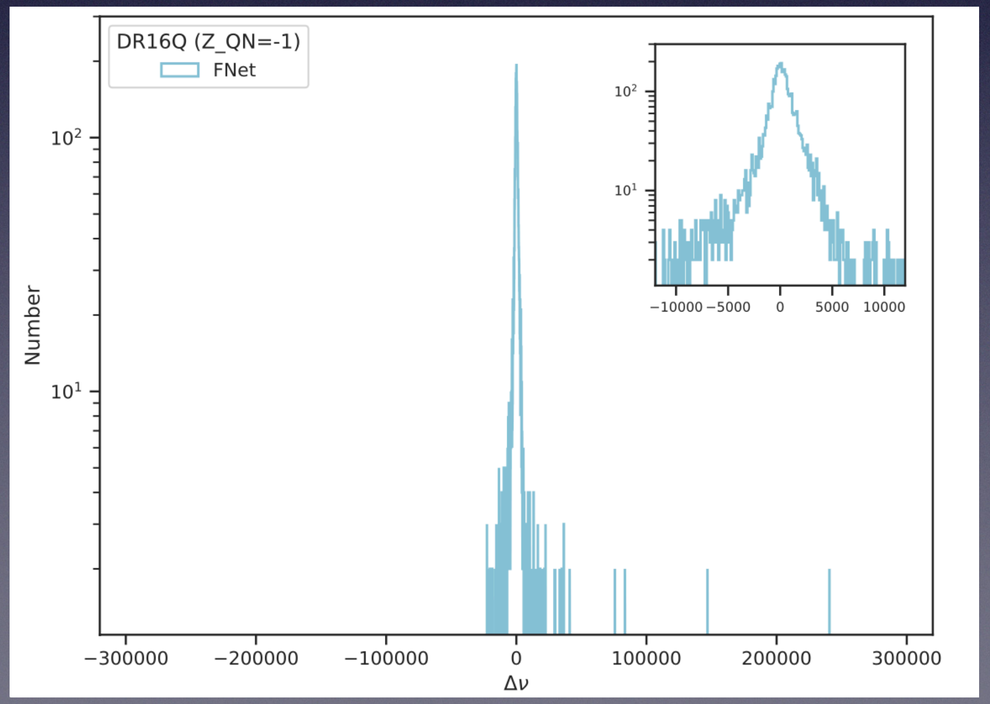


Redshift Estimation Results (FNet I)

- **Accuracy Metrics**:
  - |Δv| < 6000 km/s: 97%,
  - |Δv| <12,000 km/s: 98%,
  - |Δv| <30,000 km/s: 98.9%
- **Handling Edge Cases**: BAL quasars missed by QuasarNET recovered at 87.4%.







Novelty of FNet II – Spectral Classification

✓ **1. Self-Learning Architecture**

- Learns **local & global patterns** directly from data — **no templates or predefined features**.
- Uses:
  - **Multi-scale kernels** (500, 200, 15 pixels) → captures both broad BAL troughs & narrow emission lines.
  - **ResNet blocks** (24 layers) with skip connections → avoids vanishing gradients.

✓ **2. BAL-Quasar Detection Without Manual Rules**

- **No explicit BAL_CIV line detection** required.
- FNet achieves **99.0% BAL-quasar completeness**, even in imbalanced data.
- QuasarNET: ~78.8% for BALs (requires feature thresholds).

✓ 3. Classification of Ambiguous Spectra

- Handles **low-SNR / noisy** data with **Z_CONF ≠ 3**.
- Completeness rates:
  - **Quasars**: 45%
  - **BAL-quasars**: 72%
  - **Stars**: 62%
  - (vs. QuasarNET: 20% total)

✓ 4. Implicit Feature Discovery

- Detects subtle patterns missed by human labeling.
- Classifies spectra **mislabeled in SDSS** (e.g., stars as quasars).
- Identifies weak absorptions & continuum curvature without fitting lines.

✓ 5. High Computational Efficiency

- **14 ms/spectrum** on NVIDIA V100.
- Uses **AdamW** for faster, more stable training than Adam/SGD

- **Training Time**: 90 minutes/epoch on 639k spectra.
- **Inference Speed**: 14 ms/spectrum – suitable for LSST-scale input.
- **Accessibility**: Code and data openly available (GitHub/Kaggle).

✓ 6. Non-Conventional Insight

FNet avoids line-fitting entirely, learning from raw flux data to classify even ambiguous or rare spectra more accurately than traditional methods.

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **511,** 4490–4499 (2022)
Advance Access publication 2022 January 19

https://doi.org/10.1093/mnras/stac076

Deep learning in searching the spectroscopic redshift of quasars

F. Rastegarnia,[1,2,3] M. T. Mirtorabi,[1,2★] R. Moradi,[2,3,4★] A. Vafaei Sadr[5]

[1]*Department of Physics, Faculty of Physics and Chemistry, Alzahra University, Vanak, 1993891176, Tehra*
[2]*ICRANet, Piazza della Repubblica 10, I-65122 Pescara, Italy*
[3]*ICRA, Dipartimento di Fisica, Università di Roma 'La Sapienza', Piazzale Aldo Moro 5, I-00185 Roma, It*
[4]*INAF – Osservatorio Astronomico d'Abruzzo, Via M. Maggini snc, I-64100 Teramo, Italy*
[5]*Departement de Physique Theorique, Universite de Geneve, 1211 Geneva 4, Switzerland*

Accepted 2022 January 5. Received 2022 January 3; in original form 2021 June 23

ABSTRACT

Studying the cosmological sources at their cosmological rest frames is crucial to tr
compact objects. In view of the increasing data volume of existing and upcoming tele
convolutional neural network (CNN) with a residual neural network (ResNet) structur
Sloan Digital Sky Survey IV (SDSS-IV) catalogue from the Data Release 16 Quasa
Oscillation Spectroscopic Survey on a broad range of signal-to-noise ratios, named FN
and the ResNet structure with different kernel sizes of 500, 200, and 15, FNet is able
in the whole sample of spectra by a self-learning procedure. It reaches the accuracy of
for redshift, $|\Delta v| < 6000\,\mathrm{km\,s^{-1}}$, and 98.0 per cent for $|\Delta v| < 12\,000\,\mathrm{km\,s^{-1}}$, while
adopted in the SDSS routine and is constructed of four convolutional layers (no Re
to measure the redshift via identifying seven emission lines (*local* patterns), fails in

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **533,** 1976–1985 (2024)
Advance Access publication 2024 August 06

https://doi.org/10.1093/mnras/stae1878

FNet II: spectral classification of quasars, galaxies, stars, and broad absorption line (BAL) quasars

R. Moradi,[1★] F. Rastegarnia,[2,3,4★] Y. Wang[2,3,5★] and M. T. Mirtorabi[2,3,4]

[1]*Key Laboratory of Particle Astrophysics, Institute of High Energy Physics, Chinese Academy of Sciences, 100049 Beijing, People's Republic of China*
[2]*ICRANet, Piazza della Repubblica 10, I-65122 Pescara, Italy*
[3]*ICRA, Dipartimento di Fisica, Università di Roma 'La Sapienza', Piazzale Aldo Moro 5, I-00185 Roma, Italy*
[4]*Department of Fundamental Physics, Faculty of Physics, Alzahra University, Tehran 1993891176, Iran*
[5]*INAF – Osservatorio Astronomico d'Abruzzo,Via M. Maggini snc, I-64100 Teramo, Italy*

Accepted 2024 July 30. Received 2024 July 30; in original form 2024 May 20

ABSTRACT

In this work, we enhance the FNet, a 1D convolutional neural network (CNN) with a residual neural network (ResNet) architecture, to perform spectral classification of quasars, galaxies, stars, and broad absorption line (BAL)-quasars in the SDSS-IV catalogue from DR17 of eBOSS. Leveraging its convolutional layers and the ResNet structure with different kernel sizes, FNet autonomously identifies various patterns within the entire sample of spectra. Since FNet does not require the intermediate step of identifying specific lines, a simple modification enabled our current network to classify all SDSS spectra. This modification involves changing the final output layer from a single value (redshift) to multiple values (probabilities of all classes), and accordingly adjusting the loss function from mean squared error to cross-entropy. FNet achieves a completeness of 99.00 per cent ± 0.20 for galaxies, 98.50 per cent ± 0.30 for quasars, 99.00 per cent ± 0.18 for BAL-quasars, and 98.80 per cent ± 0.20 for stars. These results are comparable to those obtained using QuasarNET, a standard CNN employed in the SDSS routine, comprises convolutional layers without the ResNet structure with equal kernel sizes, and is utilized for

The code can be found in the public domain: https://github.com/AGNNet/FNet.git

The prepared data set can be found in https://www.kaggle.com/ywangscience/sdss-iii-iv

Thank you