Primena veštačkih neuronskih mreža na LSST krive sjaja aktivnih galaksija

Nikola Andrić Mitrović (Student Katedre za računarstvo i informatiku MATF)

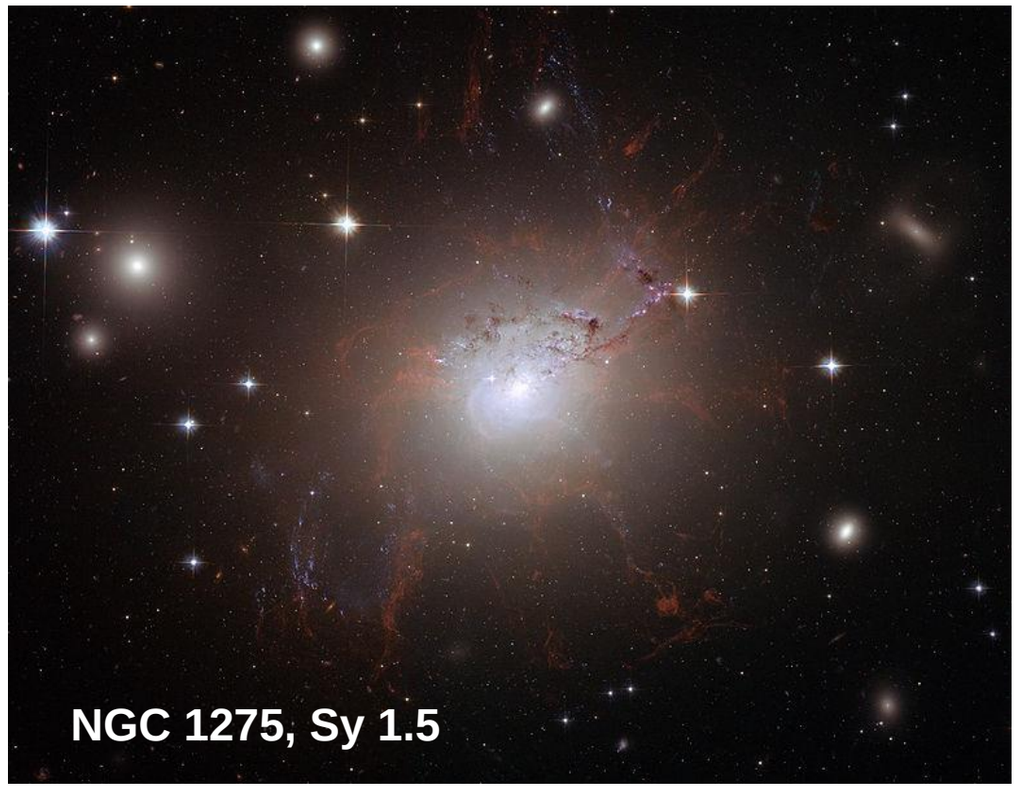
LSST Corporation Enabling Science: Building Deep Learning Engine (DLE) for AGN light-curves

CoPI: Andjelka Kovačević, Dragana Ilić (Katedra za astronomiju MATF)
Peer-reviewer: Iva Čvorović Hajdinjak (Katedra za astronomiju MATF)

andricmitrovicnikola@yahoo.com

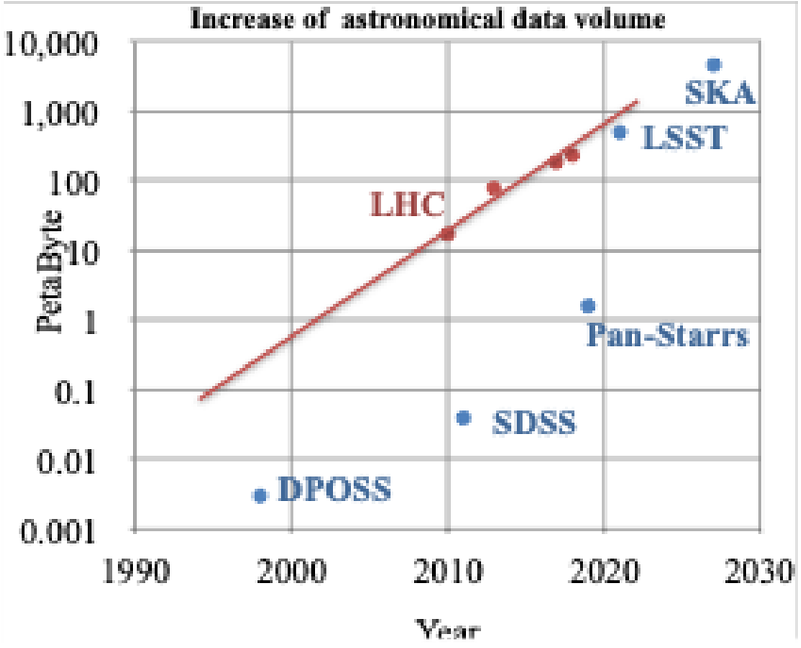
Overview

- Awarded with LSST Corp Enabling Science 2021 grant
- Studying Active galactic nuclei that are so far away that we can't capture the details of what's going on in the very core of active galaxies, so we map the central regions through the analysis of variability, LCs, which we model
- Developing deep learning engines (DLEs) for non-parametric modeling of active galactic nuclei (AGN) light-curves (LCs)
- Project aim:
  - Implementation of a new version of the Conditional Neural Process in PyTorch from TensorFlow v1
  - Code upgrade for allowing modeling of a large number of LCs and generalization instead of overfitting to a single LC
  - Batched training and stacking of LCs
  - Extraction, preprocessing and creating of dataset of LCs ready for modeling from big astronomical databases


NGC 1275, Sy 1.5

andricmitrovicnikola@yahoo.com

Legacy Survey of Space and Time (LSST)



The telescope will produce the deepest, widest, image of the Universe:

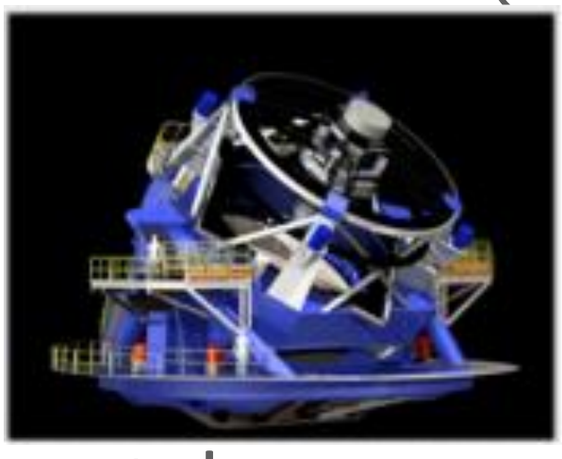
- 27-ft (8.4-m) mirror, the width of a singles tennis court
- 3200 megapixel camera
- Each image the size of 40 full moons
- 37 billion stars and galaxies
- 10 year survey of the sky
- Up to 10 million alerts,
- **20 Terabytes of data .. every night!**

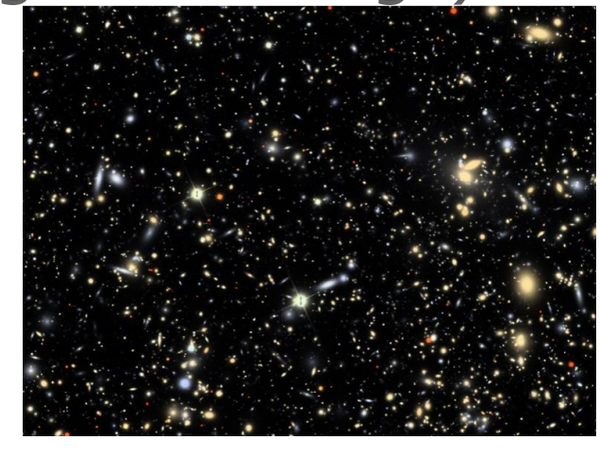
Vera C. Rubin observatory

-Under construction in Chile.
-Main task will be an astronomical survey:
the Legacy Survey of Space and Time (LSST)

$$LSST = \int \text{Observatory} + \text{Telescope} + \text{Camera} + \text{Data Management System}$$
$$= \text{Fully Reduced Data}$$

LSST (images+catalogs) → Sci. discoveries





```
$ head -n 3 output/galaxy_catalog.dat
galtileid, objectId, raJ2000, decJ2000, redshift, u_ab, g_ab, r_ab, i_ab, z_ab,
222500350435, 222500350435, 199.56648010, -9.28911042, 0.87100780, 24.72078514,
222501392641, 222501392641, 199.57937323, -9.29996667, 0.70250392, 26.08153725,
```

telecope → images → catalogues

andricmitrovicnikola@yahoo.com

Dataset

- Sánchez-Sáez, P., et al. "Alert classification for the alerce broker system: The light curve classifier." *The Astronomical Journal* 161.3 (2021): 141.
- Real data from ZTF telescope (The Zwicky Transient Facility https://www.ztf.caltech.edu/about-ztf.html)
- Parquet file with the light curves contains 50,468,778 observations [ID, name, oid_alerce, mjd, mag, magerr, catflags, ra, dec]
- Csv file with some variability features and information from the original catalogs (~50 features)

objID	oid_alerce	meanra	meandec	n_good_det	timespan_good	T2020_sigma2	mhps_ratio	mhps_low	mhps_high	mhps_non_zero	mhps_PN_flag	Amplitude	Anderson	Autocor_length

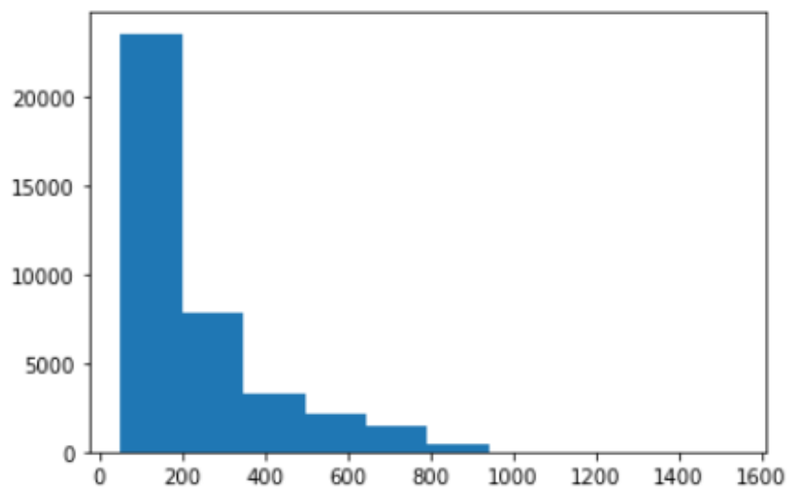
- 38,673 quasars

	oid_alerce	mjd	mag	magerr
270	600115400000062	58277.446759	20.499042	0.150769
271	600115400000062	58280.447870	20.510712	0.151645
272	600115400000062	58283.447500	20.546181	0.154301
273	600115400000062	58286.446076	20.422102	0.143933
274	600115400000062	58292.443738	20.381016	0.140310
...	...	...	...	...
106537366	498106400001568	59211.114086	19.524521	0.111328
106537367	498106400001568	59216.134167	19.357594	0.100389
106537368	498106400001568	59218.102951	19.321796	0.098138
106537369	498106400001568	59220.145822	19.447880	0.106221
106537370	498106400001568	59224.150370	19.445705	0.106078

50470539 rows × 4 columns

	oid_alerce	type
0	5.671073e+14	Q
1	5.771083e+14	AX
2	7.221141e+14	Q
3	8.201032e+14	Q
4	5.701162e+14	Seyfert_1
...	...	...
230459	6.331161e+14	QX
230460	6.801084e+14	Q
230461	4.731143e+14	Q
230462	6.241132e+14	Q
230463	6.241151e+14	QX

230464 rows × 2 columns

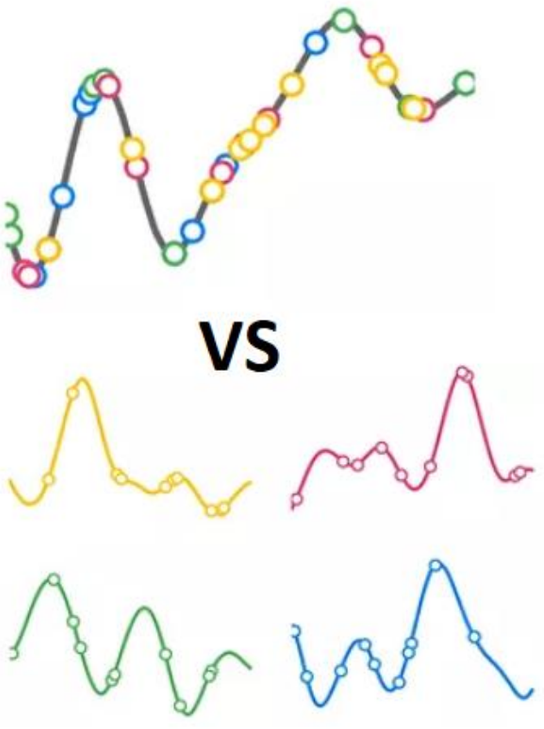


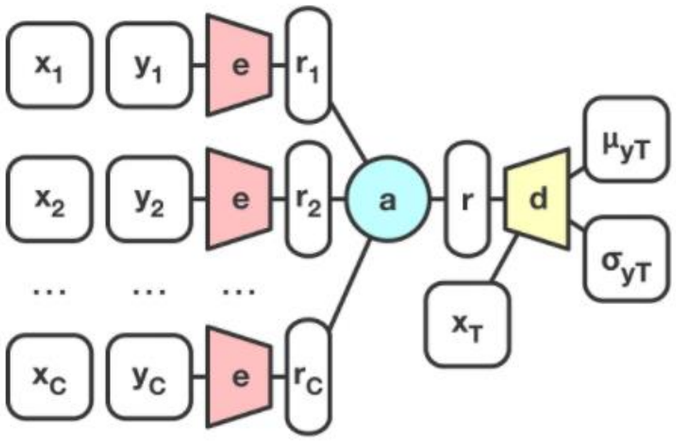
x: number of points
y: number of LCs

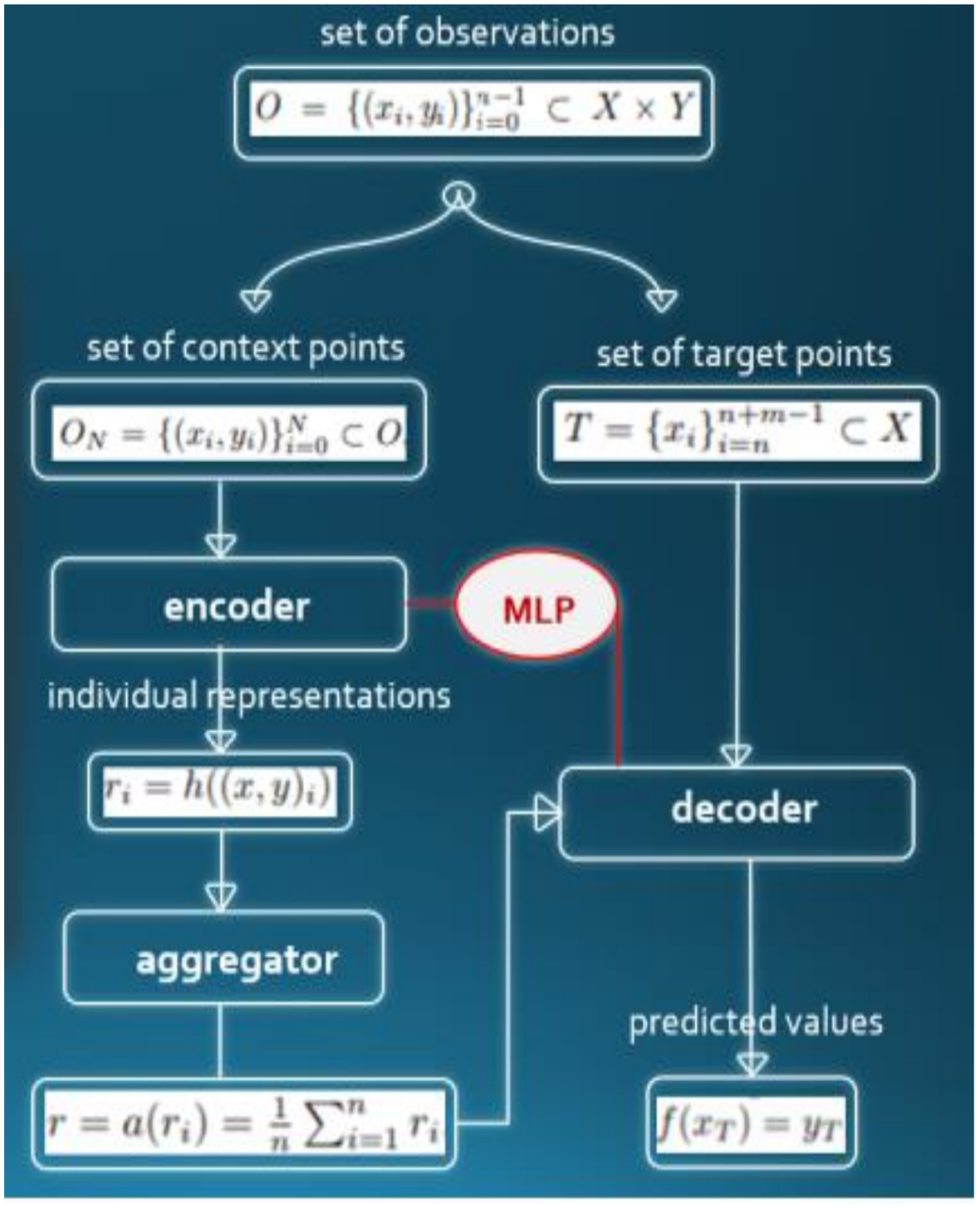
andricmitrovicnikola@yahoo.com

Method

- The task is to try to find the values of flux in the moments of time where we have gaps in our observations in order to obtain realistic and continuous representation of our light curve
- Conditional Neural Processes (Garnelo, Rosenbaum, et al., 2018) architecture
- Application of CNP to quasar LCs modeling by Cvorovic-Hajdinjak et. al 2021







Cvorovic-Hajdinjak et. al 2021

andricmitrovicnikola@yahoo.com

Notebook: preprocessing

- Splitting raw data into LCs
- Extracting quasar type LCs
- Applying transformations
- Saving LCs ready for modeling as train, test, and validation sets

andricmitrovicnikola@yahoo.com

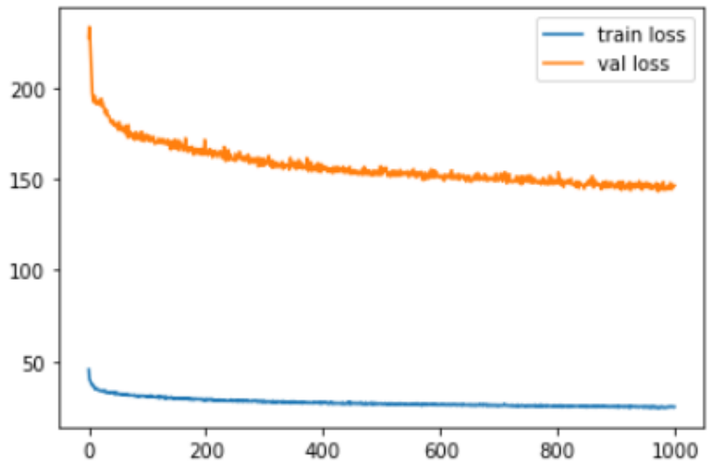
Notebook: modeling

- Data retrieval class
- CNP model – Encoder-Decoder Architecture
- Custom collate function for stacking variable-size inputs to enable batched processing of light curves
- Selection of context and target points
- Train and validation loop with early stopping criterion to prevent the network from overfitting on the train set
- Displaying of modeled LCs from train and test set

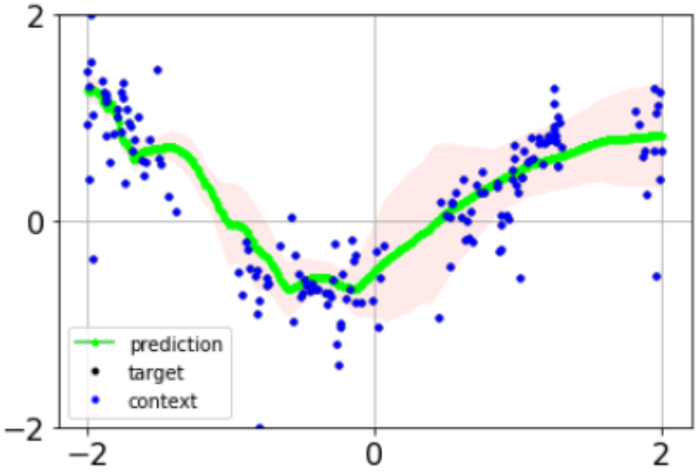
andricmitrovicnikola@yahoo.com

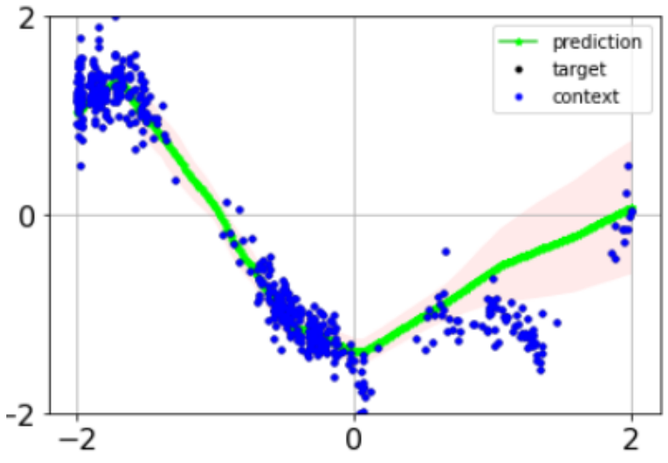
Results

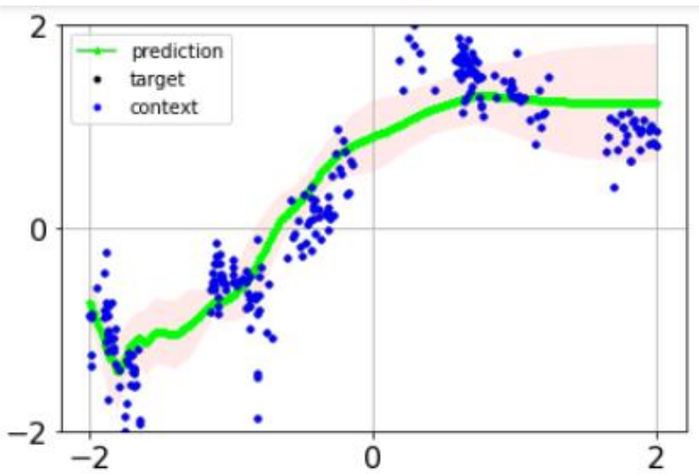
- Training took ~30h on 2,560 CUDA cores



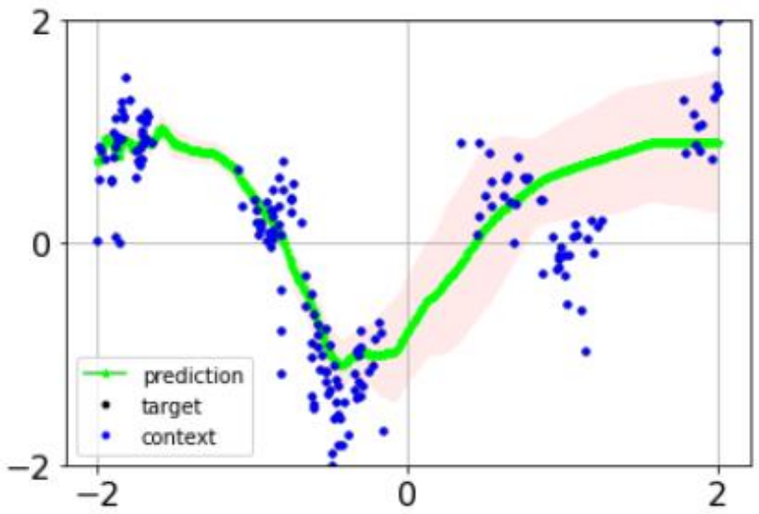
- Train

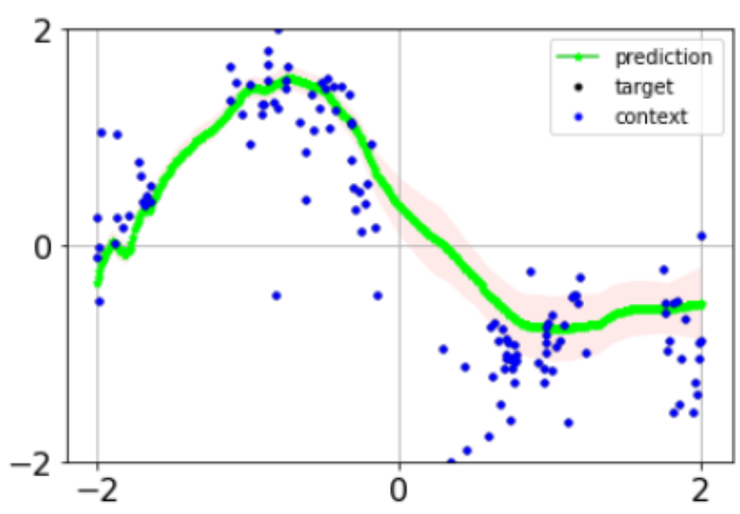


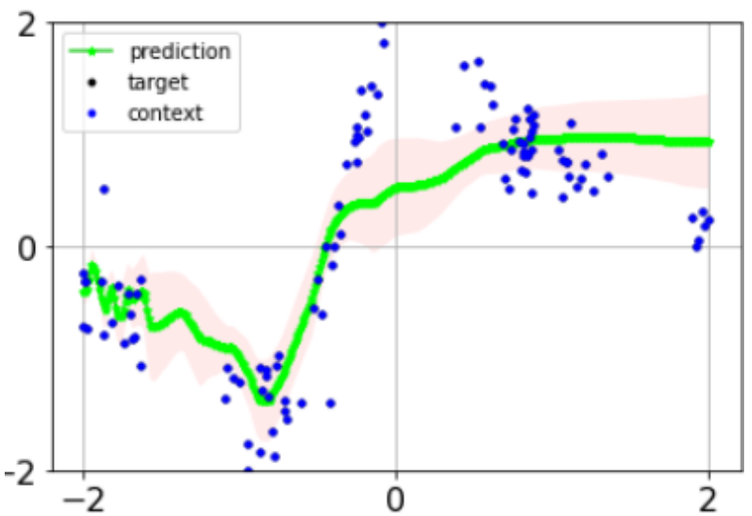




- Test







andricmitrovicnikola@yahoo.com

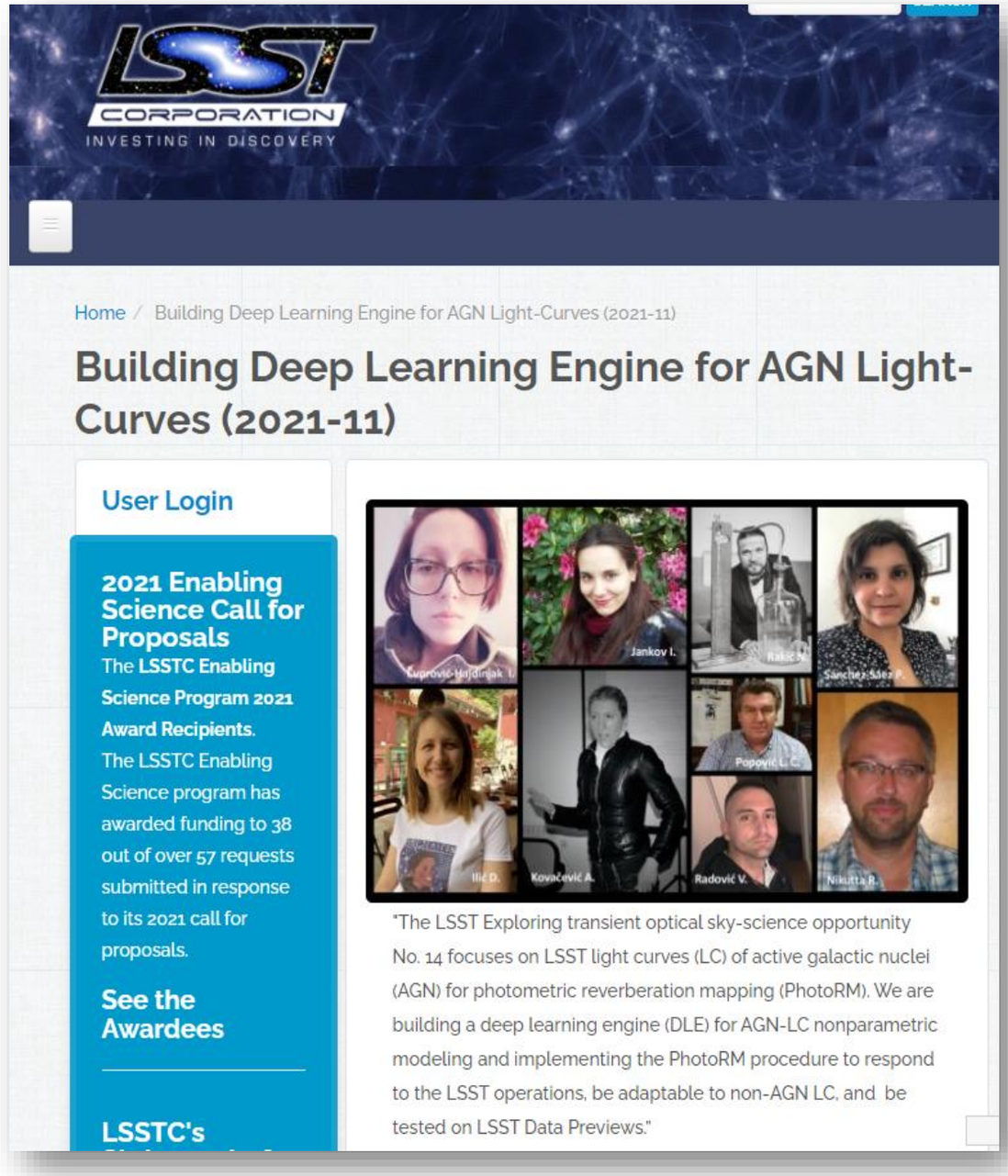
Future work

- Stratified train/test/validation split based on LCs parameters
- Splitting LCs into multiple LCs based on densities of observations (gaining ~ 100k LCs dataset)
- Hyperparameter optimization using Weights & Biases
- Notebook will be a part of periodicity detection pipeline in LSST
- Experiment with other architectures:
  - Convolutional Conditional Neural Process (Gordon et al., 2020)
  - Attentive Neural Process (Kim et al., 2019)

andricmitrovicnikola@yahoo.com

DLEs: Deep Learning Engines

- https://github.com/LSST-sersag/dle



andricmitrovicnikola@yahoo.com