Finding Strong Gravitational Lenses with Self-Attention

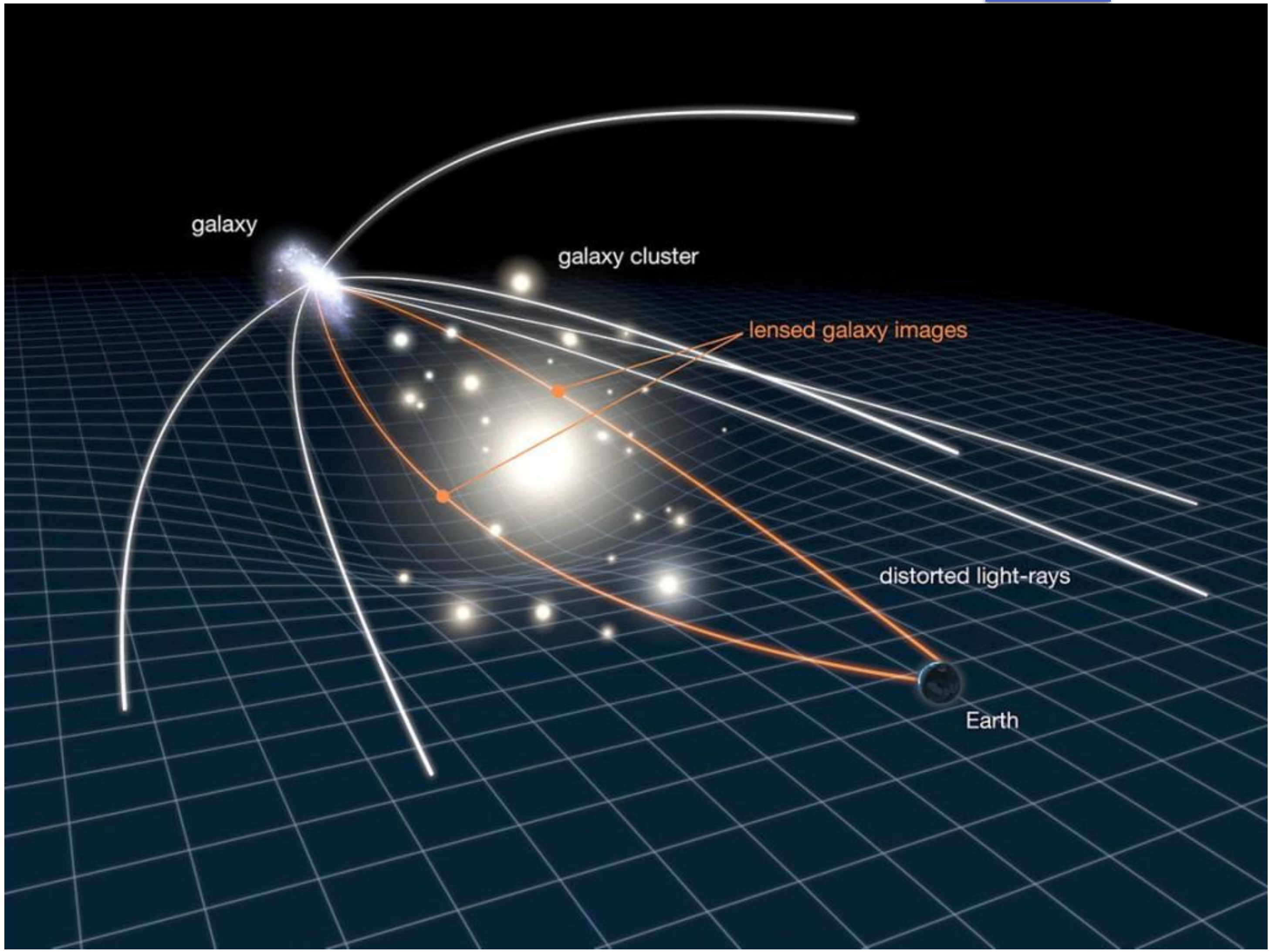
HAREESH THURUTHIPILLY & MARGHERITA GRESPAN,

NATIONAL CENTER FOR NUCLEAR RESEARCH (NCBJ), POLAND

NCBJ ŚWIERK

Strong Lensing

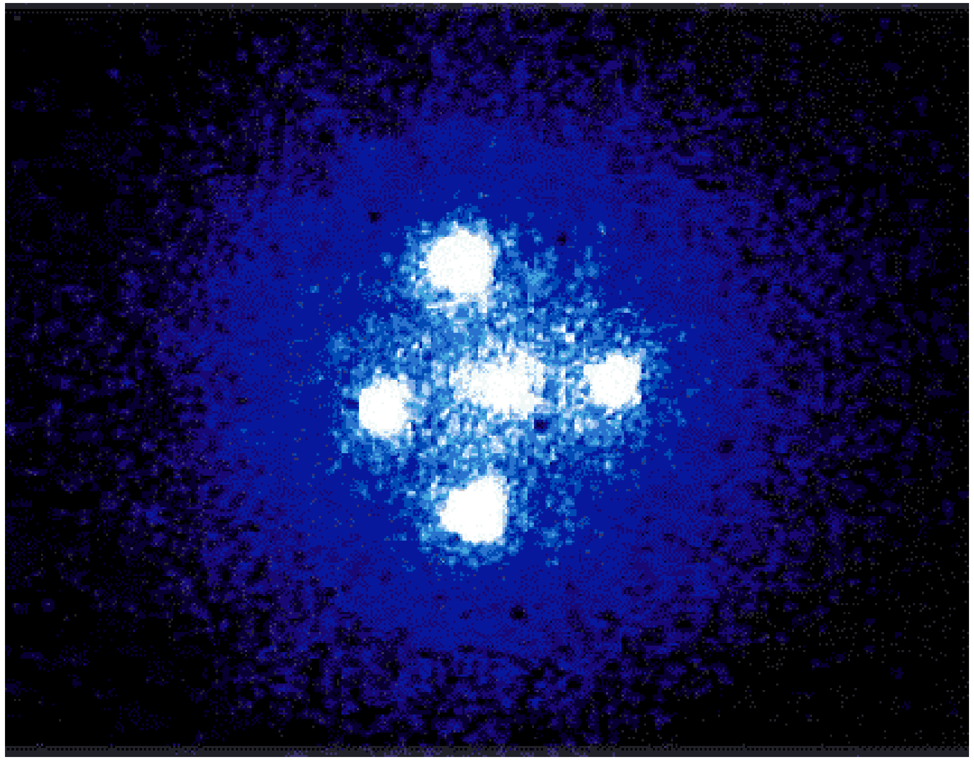
A distant galaxy or quasar produces multiple, highly distorted images because of the gravitational field of the foreground galaxy or a nearby massive astronomical body.



NCBJ ŚWIERK

A closer look at Strong Gravitational Lenses

Image, taken with the NASA/ESA Hubble Space Telescope. The two eyes are the galaxies SDSSCGB 8842.3 and SDSSCGB 8842.4 and the misleading smile lines are actually arcs caused by strong gravitational lensing.



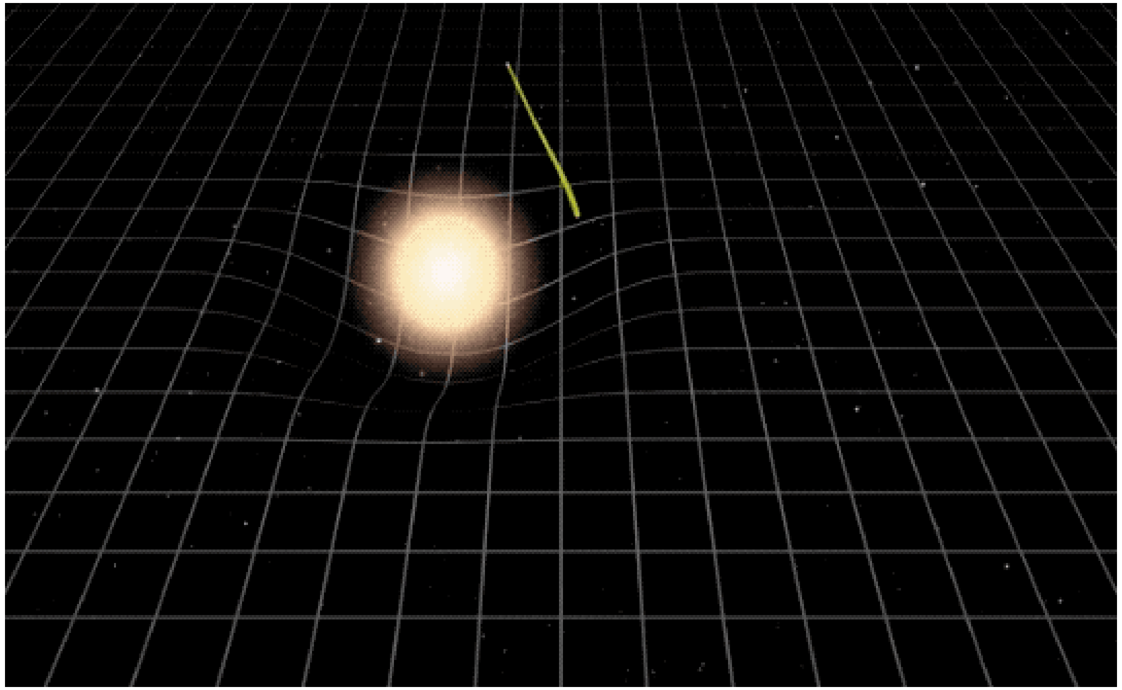
The image of an Einstein cross 2237+0305 as an example of a gravitational lens. The explanation for this pattern claims that it is produced by a galaxy which deflects the light from a quasar into four distinct images

Why SLs are Important

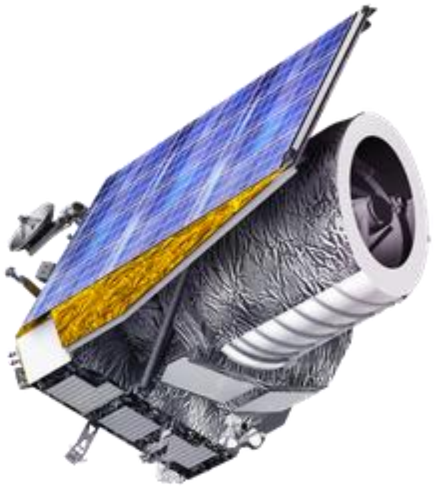
To estimate the universe's dark matter distribution.

To constrain the cosmological models.

To measure the Hubble constant independently of cosmic distance ladder



Upcoming Large Scale Surveys

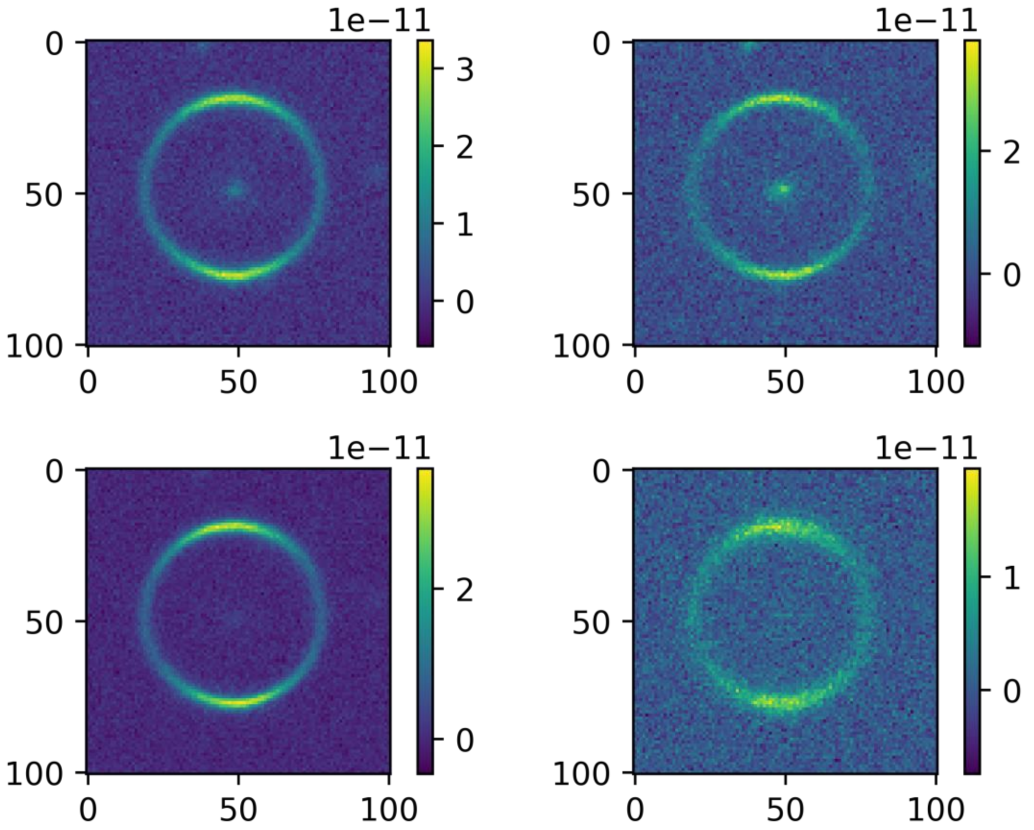


- The advanced missions such as the Euclid, and LSST is expected to find around 10^5 from around 10^9 astronomical objects.
- Non-automated techniques will be highly challenging and time-consuming.
- We propose a new automated architecture based on the principle of self-attention to find SLs.

NCBJ ŚWIERK

Bologna Lens Challenge

- This dataset mocks a ground based, multi-band survey.
- The training set 20'000 x 4 images.
- The challenge data set consists of 100'000, (101 x 101) px images in each of four bands (u,g,r,i).



Metcalf et al. 2019

Metrics for Evaluation

- Area under the receiver operating characteristic curve (AUROC) assesses the overall ability of a classifier to distinguish between classes.

$$TPR = \frac{N_{True\ positives}}{N_{True\ positives} + N_{False\ Negatives}}, \quad FPR = \frac{N_{False\ positives}}{N_{False\ Positives} + N_{True\ Negaitves}}$$

- **TPR_0** is defined as the highest TPR reached, as a function of the p threshold, before a single false positive occurs in the test set of 100,000 cases.
- **TPR_{10}** is defined as the TPR at the point where less than ten false positives are made.

NCBJ ŚWIERK

Self-Attention to Find SLs

- All the current models in astronomy uses CNNs even though the current state-of-art techniques in computer vision uses transformers and EfficientNets.

- Can self-attention based models or transformers can replace CNNs in astronomy ?

Draft version June 18, 2020
Preprint typeset using LaTeX style emulateapj v. 12/16/11

NEW HIGH-QUALITY STRONG LENS CANDIDATES WITH DEEP LEARI
IN THE KILO DEGREE SURVEY

R. Li[1], N. R. Napolitano[1], C. Tortora[2], C. Spiniello[3], L. V. E. Koopmans[4], Z. Huang[1], N. Ro
S. Chatterjee[4], B. Giblin[6], F. Getman[3], M. Radovich[7], G. Covone[3,8,9], and K. K
[1]School of Physics and Astronomy, Sun Yat-sen University, Zhuhai Campus, 2 Daxue Road, Xiangzhou Distri
napolitano@mail.sysu.edu.cn, lirui228@mail.sysu.edu.cn.
[2]Osservatorio Astrofisico di Arcetri, L.go E. Fermi 5, 50125 Firenze, Italy.
[3]INAF−Osservatorio Astronomico di Capodimonte, Salita Moiariello,16,80131 Napoli, It
[4]Kapteyn Astronomical Institute, University of Groningen, P.O.Box 800, 9700AV Groningen, the
[5]Institute of Astrophysics, Foundation for Research and Technology− Hellas (FORTH), GR-70013, He
[6]Institute for Astronomy,University of Edinburgh,Blackford Hill, Edinburgh,EH9 3HJ,U
[7]INAF−Osservatorio Astronomico di Padova, Vicolo Osservatorio 5, I-35122 Padova, Ita
[8]Dipartimento di Fisca "E. Pancini", University of Naples "Federico II", Naples,
[9]INFN, Sezione di Napoli, Naples, Italy and
[10]Leiden Observatory, Leiden University, P.O.Box 9513, 2300RA Leiden, The Neth
Draft version June 18, 2020

ABSTRACT

We report new high-quality galaxy scale strong lens candidates found in the Ki data release 4 using Machine Learning. We have developed a new Convolutional (CNN) classifier to search for gravitational arcs, following the prescription by Petr and using only $r-$band images. We have applied the CNN to two "predictive samp red galaxy (LRG) and a "bright galaxy" (BG) sample ($r < 21$). We have fou probability candidates, 133 from the LRG sample and 153 from the BG sample. We these candidates based on a value that combines the CNN likelihood to be a lens score resulting from visual inspection (P-value) and we present here the highest 82 r with P-values ≥ 0.5. All these high-quality candidates have obvious arc or point-lik the central red defector. Moreover, we define the best 26 objects, all with scores l a "golden sample" of candidates. This sample is expected to contain very few fa thus it is suitable for follow-up observations. The new lens candidates come parti more extended footprint adopted here with respect to the previous analyses, partia predictive sample (also including the BG sample). These results show that machi are very promising to find strong lenses in large surveys and more candidates that enlarging the predictive samples beyond the standard assumption of LRGs. In the f apply our CNN to the data from next-generation surveys such as the Large Synoptic Euclid, and the Chinese Space Station Optical Survey.
Subject headings: gravitational lensing: strong

MNRAS **000**, 1–12 (2017) Preprint 9 March 2017 Compiled using MNRAS LaTeX style file v3.0

CMU DeepLens: Deep Learning For Automatic Image-based Galaxy-Galaxy Strong Lens Finding

François Lanusse,[1★] Quanbin Ma,[2] Nan Li,[3,4] Thomas E. Collett,[5] Chun-Liang Li,[2] Siamak Ravanbakhsh,[2] Rachel Mandelbaum[1] and Barnabás Póczos[2]
[1]*McWilliams Center for Cosmology, Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213, USA*
[2]*School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA*
[3]*High Energy Physics Division, Argonne National Laboratory, Lemont, IL 60439, USA*
[4]*Department of Astronomy & Astrophysics, The University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637, USA*
[5]*Institute of Cosmology and Gravitation, University of Portsmouth, Burnaby Rd, Portsmouth, PO1 3FX, UK*

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT
Galaxy-scale strong gravitational lensing is not only a valuable probe of the dark

Astronomy & Astrophysics manuscript no. aanda
September 2, 2021

Strong lens systems search in the Dark Energy Survey using Convolutional Neural Networks

K. Rojas[1], E. Savary[1], B. Clément[1], M. Maus[1], F. Courbin[1], C. Lemon[1], J. H. H. Chan[1], G. Vernardos[1], R. Joseph[2], R. Cañameras[3], and A. Galan[1]

[1] Institute of Physics, Laboratory of Astrophysics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, 1290 Versoix, Switzerland
[2] Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544, USA
[3] Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, 85748 Garching, Germany

September 2, 2021

ABSTRACT

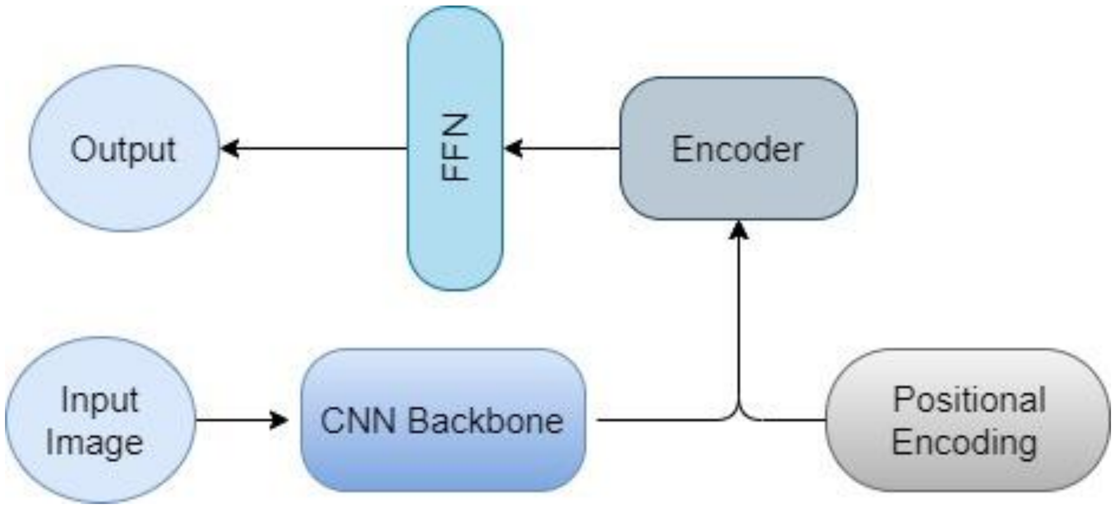
We performed a search for strong lens galaxy-scale systems in the first data release of the Dark Energy Survey (DES), from a color-selected parent sample of 18 745 029 Luminous Red Galaxies (LRGs). Our search was based on a Convolutional Neural Network (CNN) to grade our LRG selection with values between 0 (non-lens) and 1 (lens). Our training set was data-driven, i.e. using lensed sources taken from HST COSMOS images and where the light distribution of the lens plane was taken directly from DES images of our LRGs. A total of 76 582 cutouts obtained a score above 0.9. These were visually inspected and resulted in two catalogs. The first one contains 405 lens candidates, where 90 present clear lensing features and counterparts, while the others 315 require more evidence, such as higher resolution images or spectra to be conclusive. A total of 186 candidates were totally new identified in this search. The second catalog includes 539 ring galaxy candidates that will be useful to train CNNs against this type of false positives. For the 90 best lens candidates we carried out color-based deblending of the lens and source light without fitting any analytical profile to the data. The method turned out to be very efficient in the deblending, even for very compact objects and for objects with very complex morphology. Finally, from the 90 best lens candidates we selected 52 systems having one single deflector, to test an automated modeling pipeline which successfully modeled 79% of the sample within an acceptable amount of computing time.

Key words. Gravitational lensing: strong – Surveys – Techniques: image processing

Transformer model for Strong Lens Detection

We can assume that the Encoder model works in 3 phases.

- **CNN** – To extract the Features
- **Encoder** – To filter the relevant Features of the image
- **FFN** – To learn the relevant Features



Thuruthipilly et al. 2022

Results From the Models

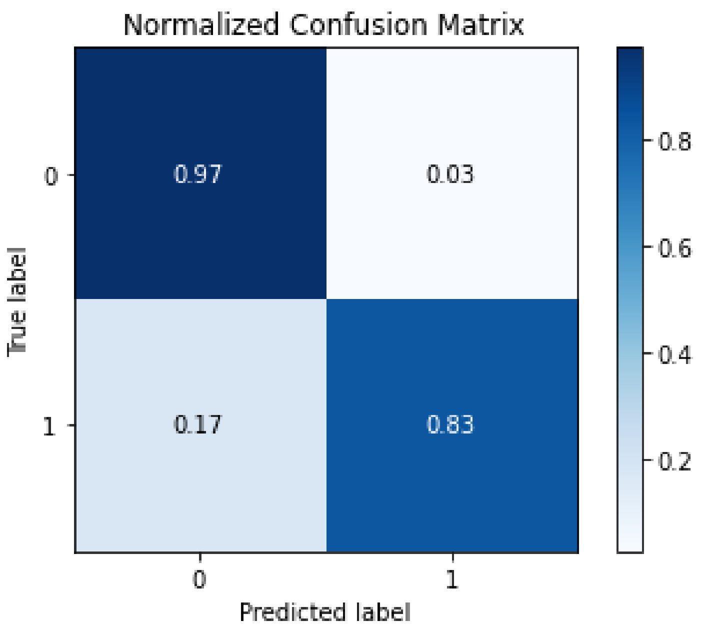
Model Name	Model Structure	Accuracy on Challenge Set
CNN 1	5 CNN Layers	88.21
CNN 2	4 CNN Layers	86.74
CNN 3	8 CNN Layers	89.91
CNN 4	3 CNN Layers	88.49
Lens Detector 1	CNN 1+1 H_{16}+1(E)	89.57
Lens Detector 2	CNN 2 + 1 H_{16} + 1(E)	88.13
Lens Detector 3	CNN 2 + 2 H_{16} + 1(E)	88.00
Lens Detector 4	CNN 2 + 2 H_{32} + 1(E)	88.12
Lens Detector 5	CNN 2 + 4 H_{64} + 2 (E)	88.46
Lens Detector 6	CNN 2 + 4 H_{128} + 4(E)	89.51
Lens Detector 7	CNN 3 + 8 H_{128} + 2(E)	91.45
Lens Detector 8	CNN 4 + 2 H_{384} + 2 (E)	89.43
Lens Detector 9	3 CNN Layers + 2 H_{384} + 2 (E)	89.61
Lens Detector 10	5 CNN Layers + 8 H_{128} + 2 (E)	90.58
Lens Detector 11	5 CNN Layers + 8 H_{128} + 4 (E)	90.45
Lens Detector 12	8 CNN Layers + 8 H_{128} + 4 (E)	89.82
Lens Detector 13	8 CNN Layers + 8 H_{128} + 4 (E)	91.94
Lens Detector 14	8 CNN Layers + 8 H_{128} + 4 (E)	91.95
Lens Detector 15	8 CNN Layers + 8 H_{128} + 4 (E)	**92.99**
Lens Detector 16	16 CNN Layers + 8 H_{128} + 8 (E)	90.97
Lens Detector 17	16 CNN Layers + 8 H_{128} + 8 (E)	92.19
Lens Detector 18	16 CNN Layers + 8 H_{128} + 8 (E)	92.21
Lens Detector 19	16 CNN Layers + 16 H_{128} + 8 (E)	90.03
Lens Detector 20	25 CNN Layers + 8 H_{128} + 4 (E)	91.26
Lens Detector 21	8 CNN Layers + 8 H_{128} + 4 (E)	92.79

NCBJ ŚWIERK

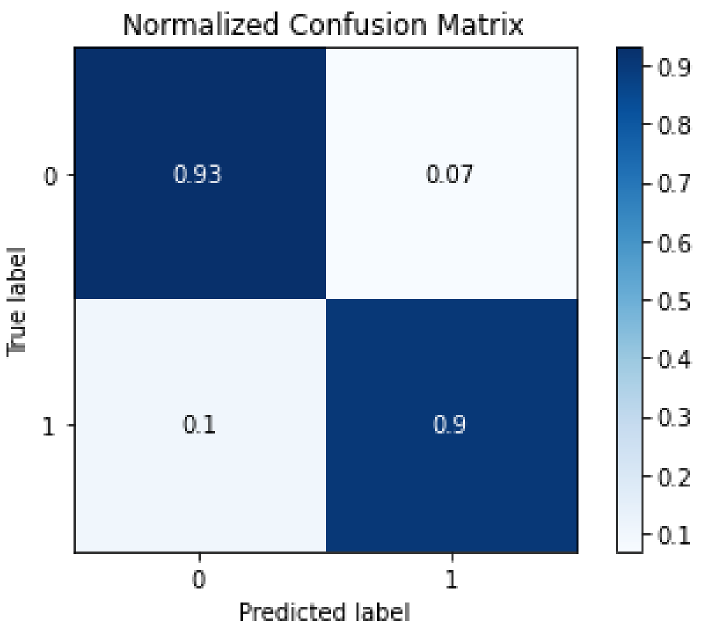
Thuruthipilly et al. 2022

Transformer Encoder vs CNN

- The Encoder Networks performs better than the CNN.
- The Encoder depends on the CNN to extract the features, and the model is only as good as the CNN, but always better.



The Confusion matrix of the **Convolution** Model 3 (Accuracy = 89.91) on the Challenge data.



The Confusion matrix of the **Transformer** model 7 (Accuracy = 91.45) with Convolution Model 3 as the backbone on the Challenge data.

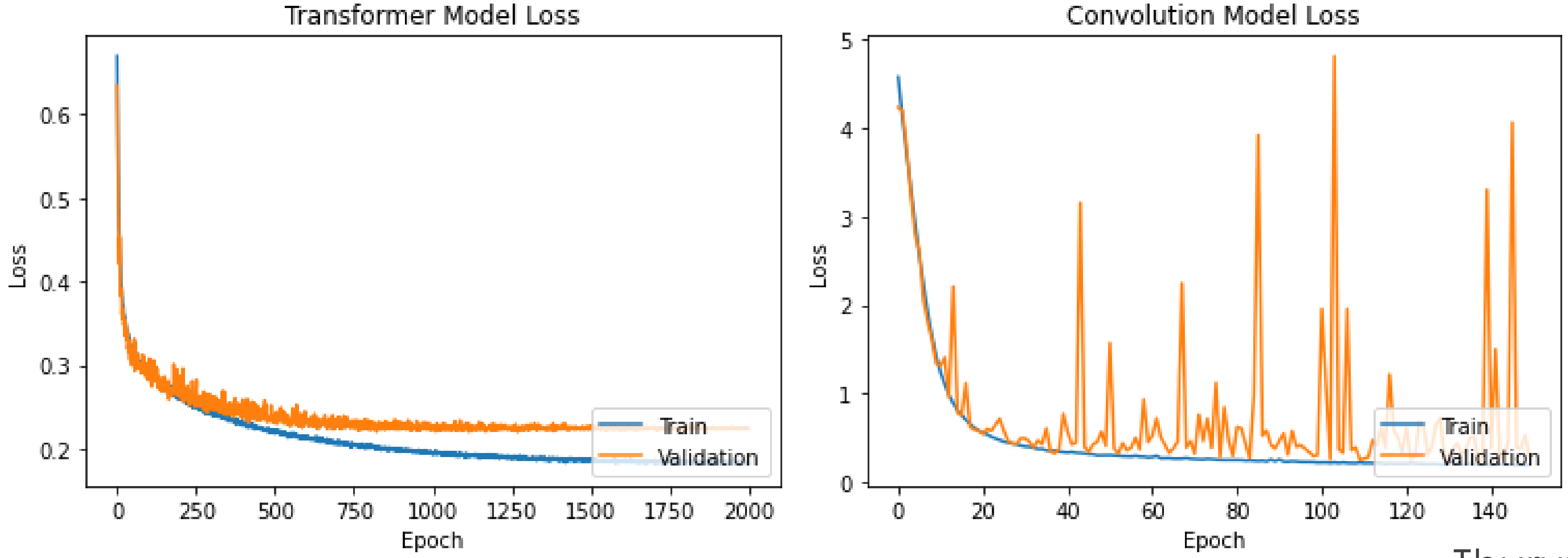
NCBJ ŚWIERK

Thuruthipilly et al. 2022

Stability

Since the Encoder is based on self-attention, the encoder layers **prevent the model from overfitting** by only learning the useful features extracted by the CNN and provides more stability to the Model.

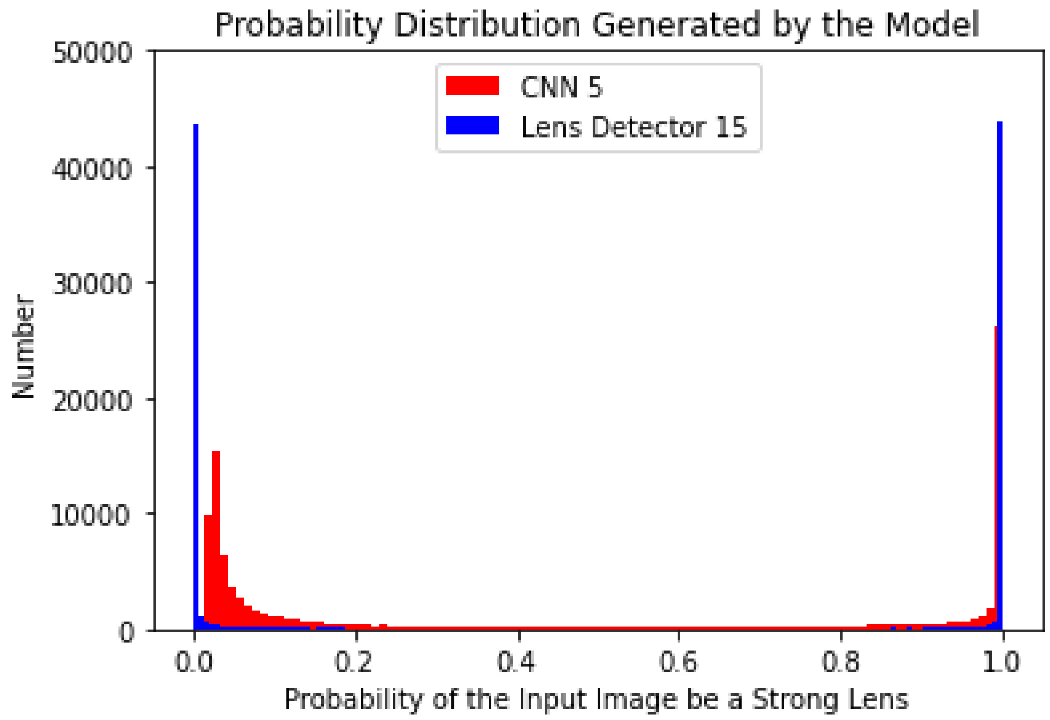
Variation of Loss Function with epochs for Lens Detector 13 and CNN 3, respectively.



NCBJ ŚWIERK

Thuruthipilly et al. 2022

Prediction Capacity



The encoder model can assign a probability for an input to be lens ($P \approx 1$) or non-lens ($P \approx 0$) with greater confidence than the CNN.

Thuruthipilly et al. 2022

Testing on KiDS survey images

Kilo Degree Survey (KiDS)

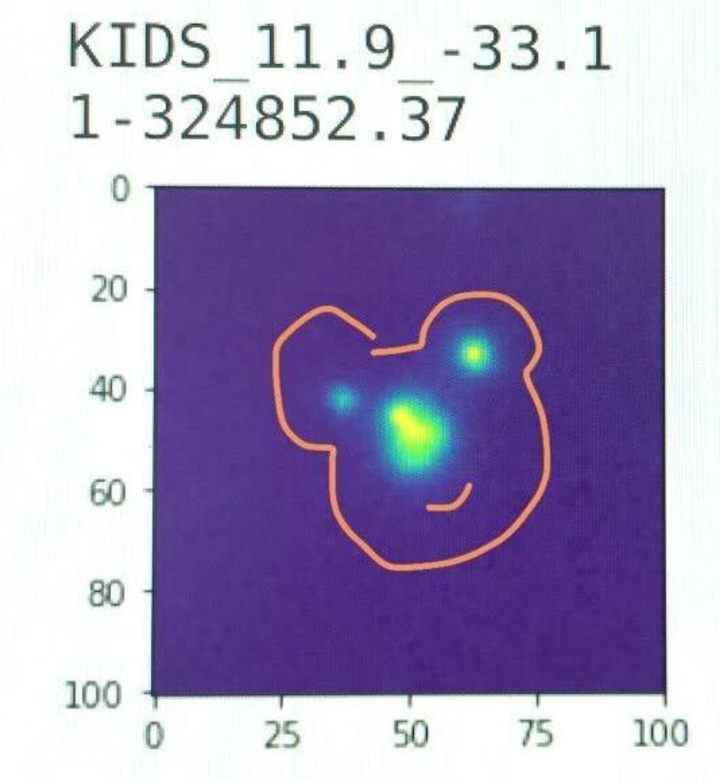
- KiDS, the Kilo-Degree Survey, is a large **optical imaging** survey in the **Southern sky**
- Using the **VLT Survey Telescope** (VST), located at the ESO Paranal Observatory, KiDS has mapped **1350** square degrees of the night sky in **four broad-band** filters (u, g, r, i).

Preliminary Tests on KiDS data

- We are searching for lensing galaxies **z_l < 0.4** in KiDS DR4
- We created cutouts of 200 tiles (~20 % of total)
- Cutouts of galaxies in 4 and 3 optical filters given to two different models
- Most probable candidates (prediction probability > 0.95) are human inspected

Human Inspection

- Candidates were checked by multiple humans (4 people).
- Scores: **5** (sure), **3** (maybe), **1** (interesting), **0** (nope).
- Candidates that were recognized by at **least 2 poeple** as potencial lens **or** got **5** (sure lense were further considered).



5 : For sure a bear

Some numbers

- KiDS collaboration found **268** HQ strong lens candidates in ~1300 tiles.
- In Li et al. 2021 (latest publication) they use two **ResNets** trained on Synthetic data and tested it on ~300 tiles.
- After the galaxy selection the model retrieves out of ~1'500'000 galaxies
  - 5 810 candidates
  - 97 HQ strong lenses
  - **1.7%** true positives

- We test our model on ~200 tiles
- With only a redshift selection we get ~1'500'000 galaxies
- Our model selects
  - 16 389 candidates
  - ~ 70 HQ strong lens
  - **0.4%** true positives

Too many false positives!

Pitfalls and False Positives - I

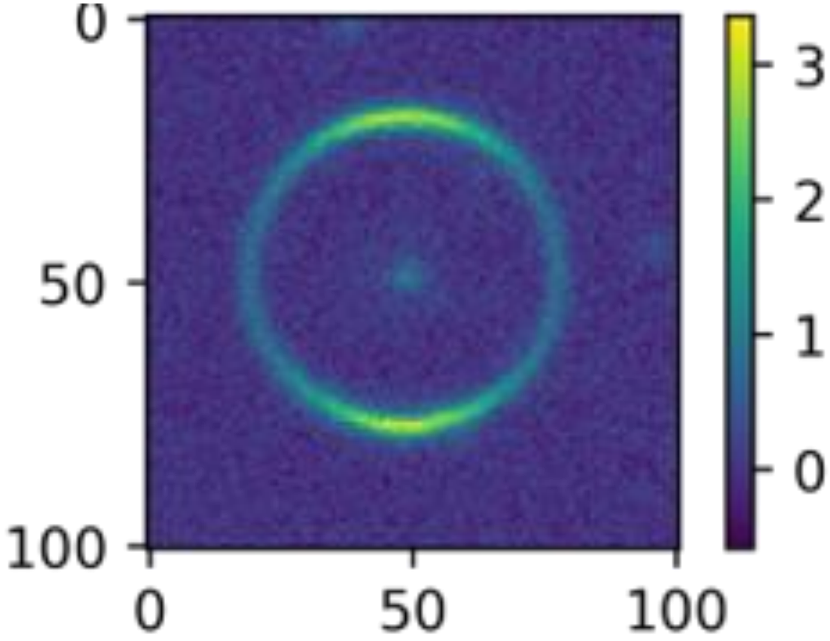
The training data is purely **simulated** and relatively simple compared to the data from the KiDS survey.

We are looking for SLs in the **entire data sample**, whereas the previous searches on KiDS only focused on the Luminous Red Galaxies (LRGs) and bright galaxies (BGs).

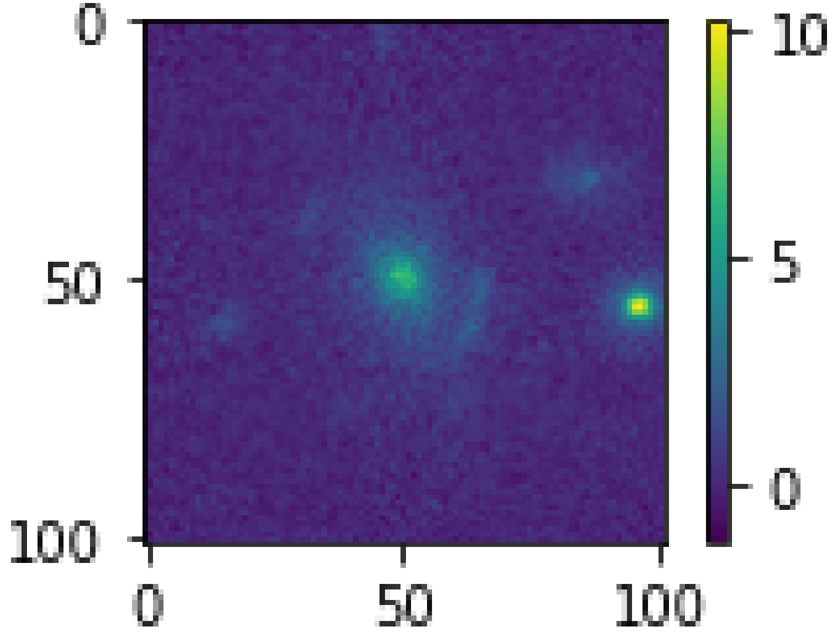
We have **not cleaned** the data to remove glitched images.

As a result, the number of false positives in the candidate sample can be high.

Pitfalls and False Positives - I

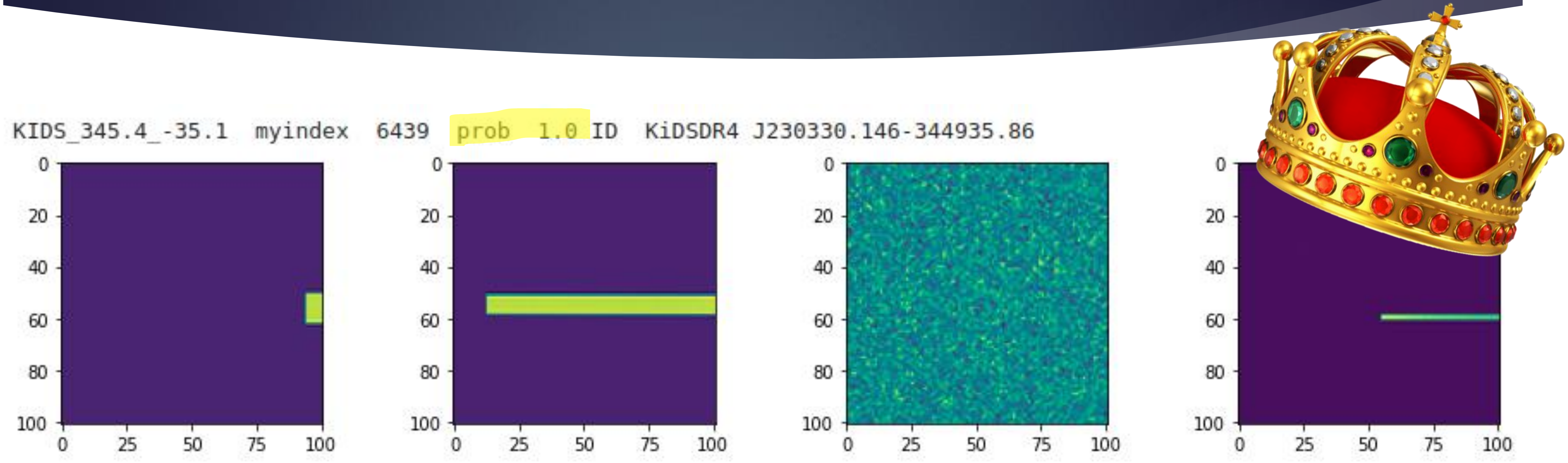


An example from simulated dataset



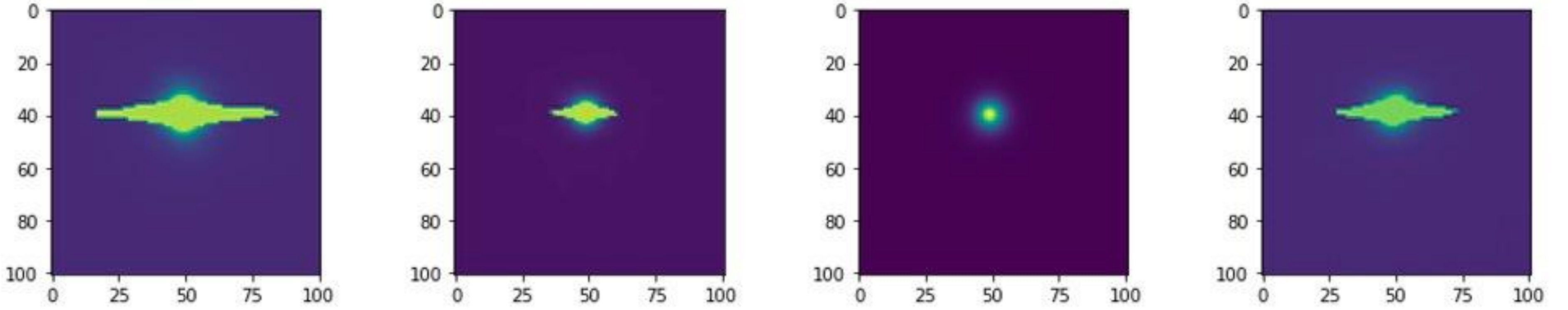
A real strong gravitational strong lens seen from the survey

Pitfalls and False Positives - II

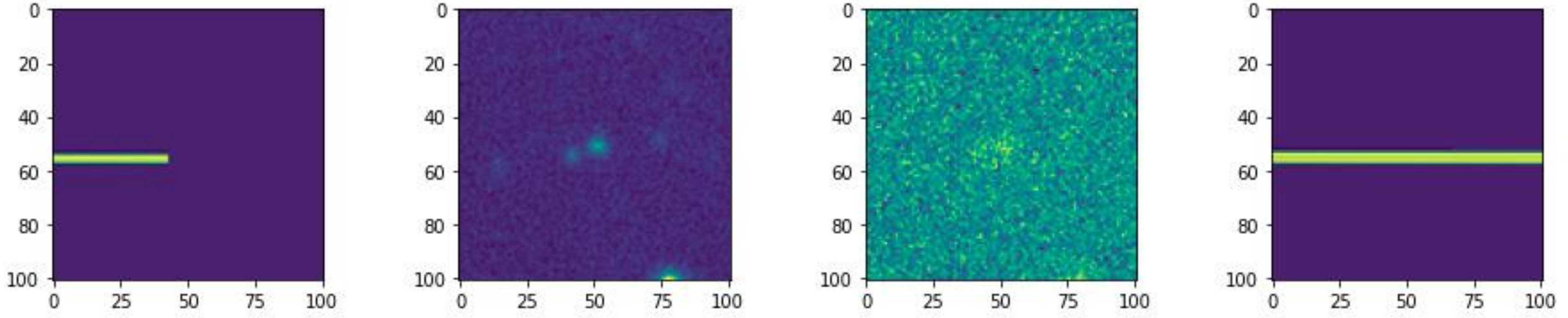


Our prodigal candidate !!!

Pitfalls and False Positives - III



Probability assigned by our model = 0.99998873

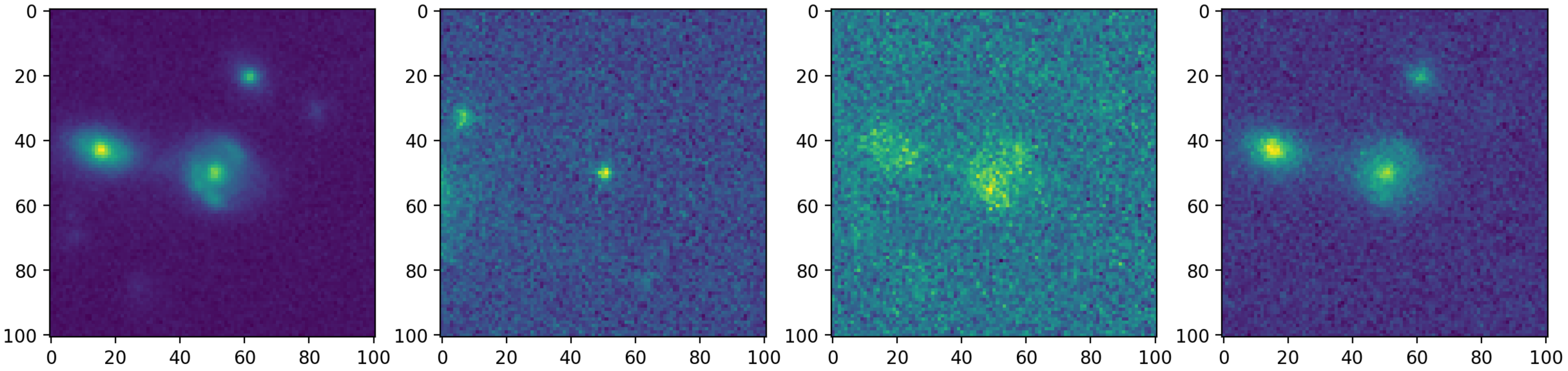


Probability assigned by our model = 0.973774

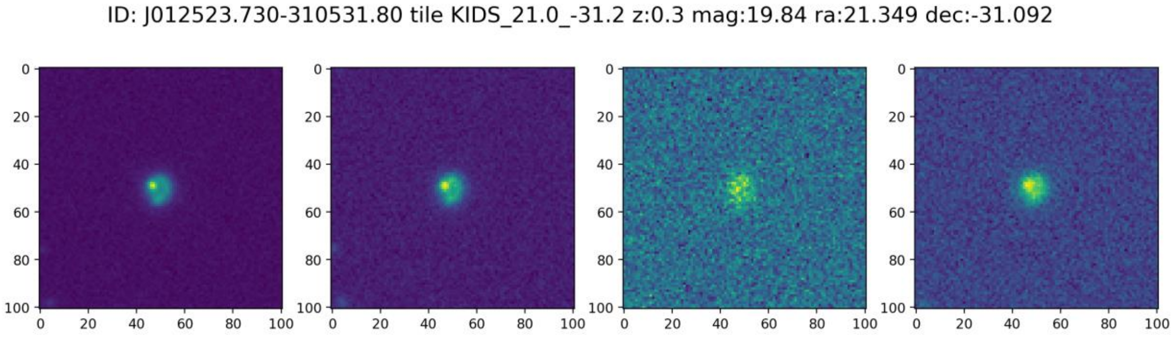
BUT…

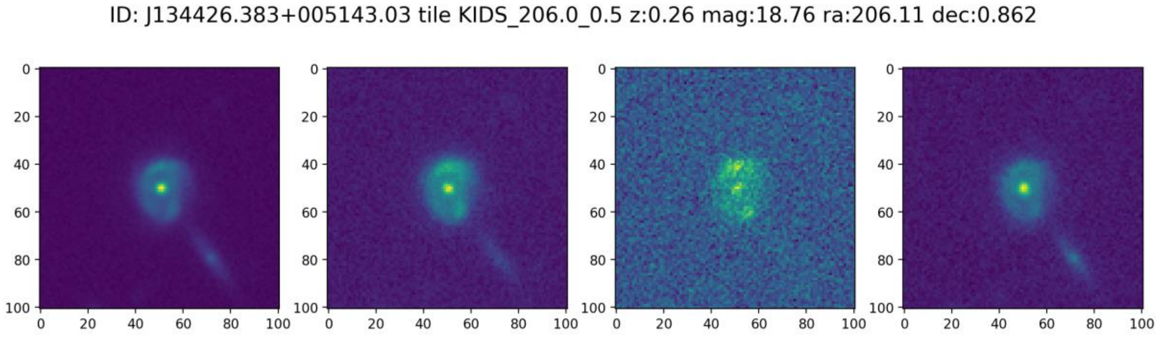
We found one lens :')

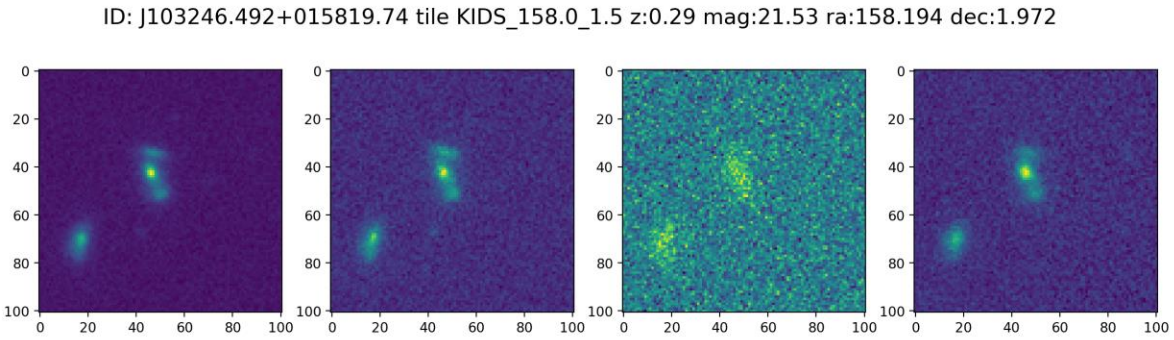
ID: J000014.372-281135.77 tile KIDS_0.0_-28.2 z:0.31 mag:18.89 ra:0.06 dec:-28.193

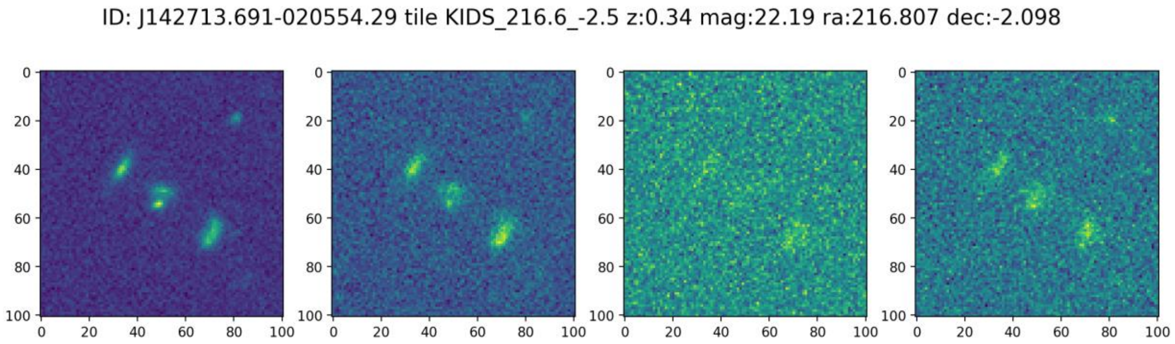


and then more (___/) (͡° ͜ʖ ͡°) \╭☞ \╭☞









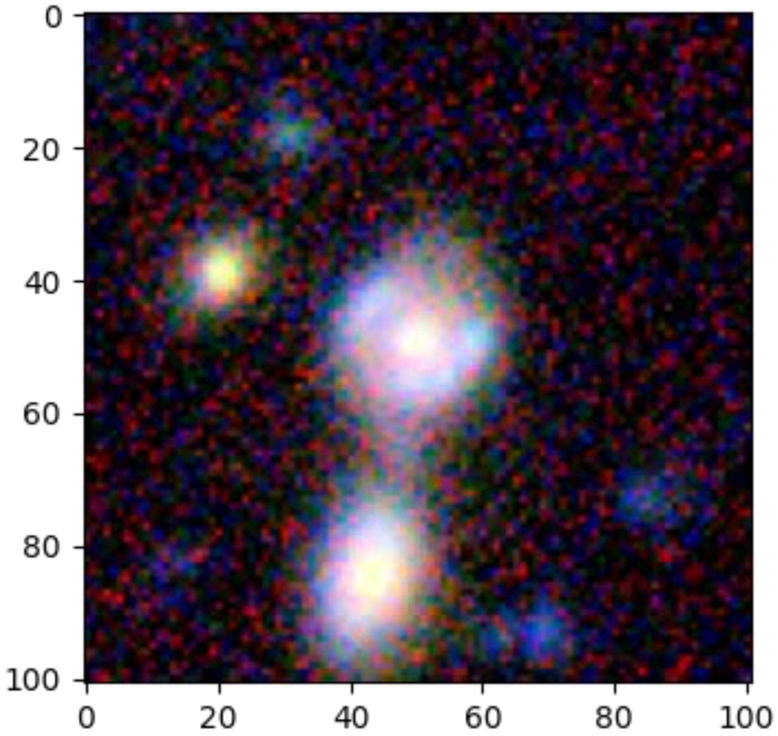
Work in Progress...

The simulated images are clearly too different from the real life lenses – We know it!

Future perspectives:

- Retrain on real life lenses
- Apply data augmentation
- Keep training with an active learning approach

We will keep you posted! Wish us luck



Summary

Encoder models have more stability than CNN's, which minimizes the need for monitoring.

The architecture proposed here is very simple and robust and has a high resistance to overfitting.

The proposed method might be **complementary** to the methods used previously on KiDS data – we find some lenses!

On going project!

Thank you

QUESTIONS ?

NCBJ ŚWIERK

Backup

No of Heads and Encoders

- Increasing the number of heads and the depth of the Encoder fastens the learning curve for the model and give better performance.

Model Name	Model Structure	Accuracy on Challenge Set
Lens Detector 2	CNN 2 + 1 H_{16} + 1(E)	88.13
Lens Detector 3	CNN 2 + 2 H_{16} + 1(E)	88.00
Lens Detector 4	CNN 2 + 2 H_{32} + 1(E)	88.12
Lens Detector 5	CNN 2 + 4 H_{64} + 2 (E)	88.46
Lens Detector 6	CNN 2 + 4 H_{128} + 4(E)	89.51

- The Encoder's performance depends upon the total no of parameters; the more the parameters, the better the learning curve and validation accuracy.

Transfer Learning

- Training time was considerably lowered by using transfer learning.

Model Name	Model Structure	Accuracy on Challenge Set
Lens Detector 5	CNN 2 + 4 H_{64} + 2 (E)	88.46
Lens Detector 6	CNN 2 + 4 H_{128} + 4(E)	89.51
Lens Detector 13	8 CNN Layers + 8 H_{128} + 4 (E)	91.94
Lens Detector 14	8 CNN Layers + 8 H_{128} + 4 (E)	91.95
Lens Detector 15	8 CNN Layers + 8 H_{128} + 4 (E)	**92.99**
Lens Detector 16	16 CNN Layers + 8 H_{128} + 8 (E)	90.97
Lens Detector 17	16 CNN Layers + 8 H_{128} + 8 (E)	92.19
Lens Detector 18	16 CNN Layers + 8 H_{128} + 8 (E)	92.21

- Models without Transfer learning performs slightly better than Models with pretrained CNNs.

Encoder Models vs Models participated in the Challenge

Name	AUROC	TPR_0	TPR_{10}	Model Type
Lens Detector 16	0.962	0.225	0.24	Transformer
Manchester SVM	0.93	0.220	0.35	SVM/Gabor
Lens Detector 11	0.966	0.219	0.34	Transformer
Lens Detector 15	0.978	0.14	0.48	Transformer
CMU-DeepLens Resnet-ground3	0.98	0.09	0.45	CNN
LASTRO EPFL	0.97	0.07	0.11	CNN

TRP_0

Lens Detector 16 surpassed all the models participated in the Challenge.

Name	AUROC	TPR_0	TPR_{10}	Model Type
Lens Detector 9	0.959	0.0	0.789	Transformer
Lens Detector 8	0.954	0.0	0.758	Transformer
Lens Detector 17	0.973	0.0	0.717	Transformer
CMU-DeepLens Resnet-ground3	0.98	0.09	0.45	CNN
Manchester SVM	0.93	0.220	0.35	SVM/Gabor
LASTRO EPFL	0.97	0.07	0.11	CNN

TRP_{10}

Lens Detector 9 surpassed all the models participated in the Challenge.

Name	AUROC	TPR_0	TPR_{10}	Model Type
CMU-DeepLens Resnet-ground3	0.98	0.09	0.45	CNN
Lens Detector 21	0.98	0.0	0.64	Transformer
CMU-DeepLens Resnet-Voting	0.98	0.02	0.10	CNN
Lens Detector 15	0.978	0.140	0.48	Transformer
Lens Detector 18	0.976	0.113	0.59	Transformer
LASTRO EPFL	0.97	0.07	0.11	CNN

AUROC

Lens Detector 21 scored equal to the highest reported AUROC in the Challenge.

Comparison to previous searches based on KiDS data 1/2

LinKS, Petrillo et al. 2019:

- **4 channels** model
  - Lenses associated with galaxies z < 0.4 - 11 out of 41 in the computed tiles
  - Our method found **4 out of 11**
- **3 channels** model
  - Lenses connected with galaxies z < 0.4 - 11 out of 41 in the computed tiles
  - Our method found **3 out of 11**

Compared against: https://www.astro.rug.nl/lensesinkids/1_bonus.html