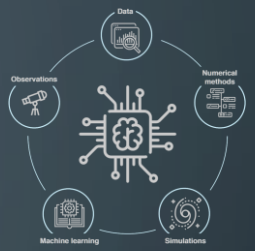


RECOGNITION OF TOTAL ECLIPSES IN BINARIES WITH COMPUTER VISION

OLIVERA LATKOVIĆ & ATTILA CSÉKI

ASTRONOMICAL OBSERVATORY OF BELGRADE, SERBIA

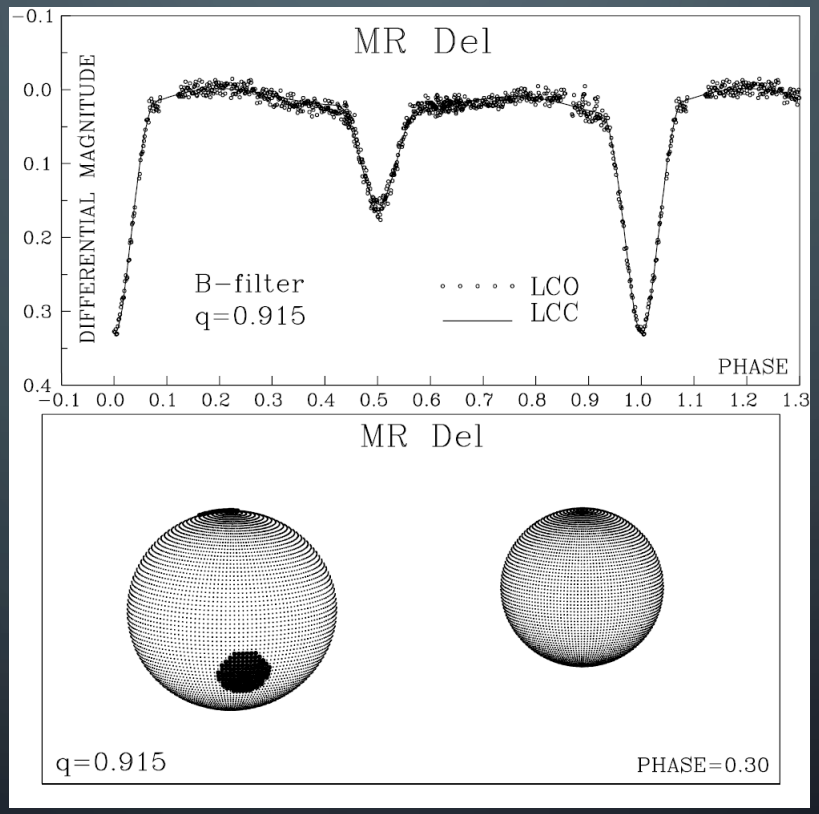
PROGRAM

- Eclipsing binaries
  - Modeling, automation, applications of ML
  - Contact binaries, importance of total eclipses
- Adventures with computer vision
  - Commercial image recognition model (CustomVision)
  - Custom-made image recognition model
  - Application on Kepler and ASAS data

ECLIPSING BINARIES

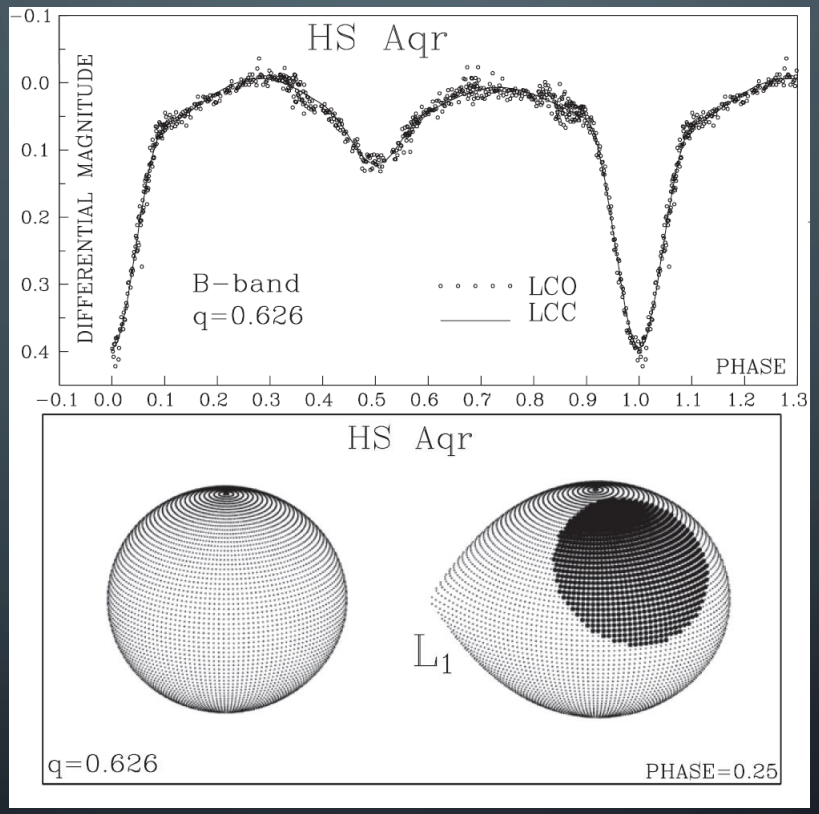
- Binary systems (mostly close) where component stars eclipse each other during each orbit
- At the forefront of astrophysics
- Measurement of stellar masses, radii and temperatures from first principles and simple geometric arguments
- Parameters are traditionally determined by modeling

MODELING ECLIPSING BINARIES
DETACHED



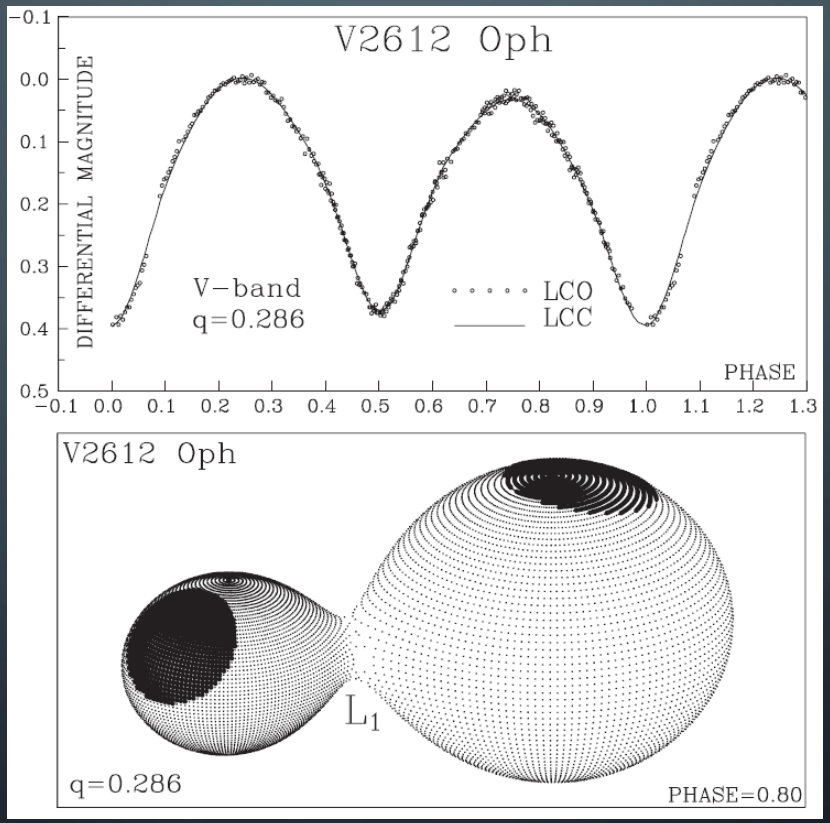
Light curve and model of MR Del
From Djurasevic et al., 2011, A&A, 525

MODELING ECLIPSING BINARIES
SEMI-DETACHED



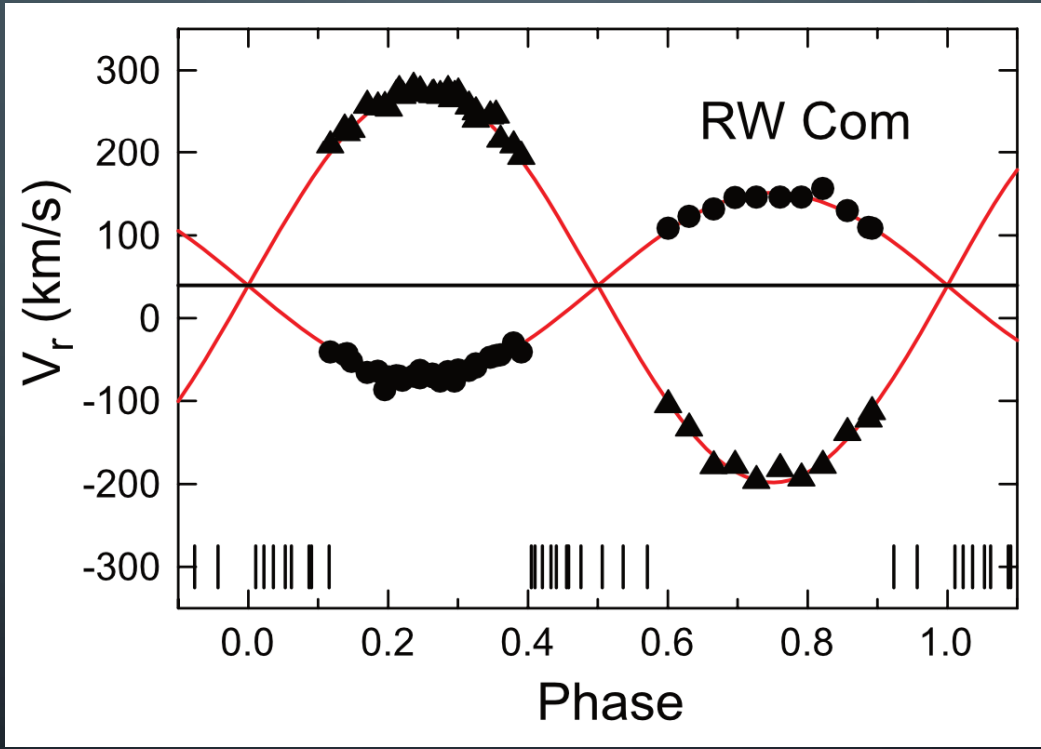
Light curve and model of HS Aqr
From Djurasevic et al., 2013, AJ, 145

MODELING ECLIPSING BINARIES
CONTACT



Light curve and model of V2612 Oph
From Caliskan et al., 2014, AJ, 148

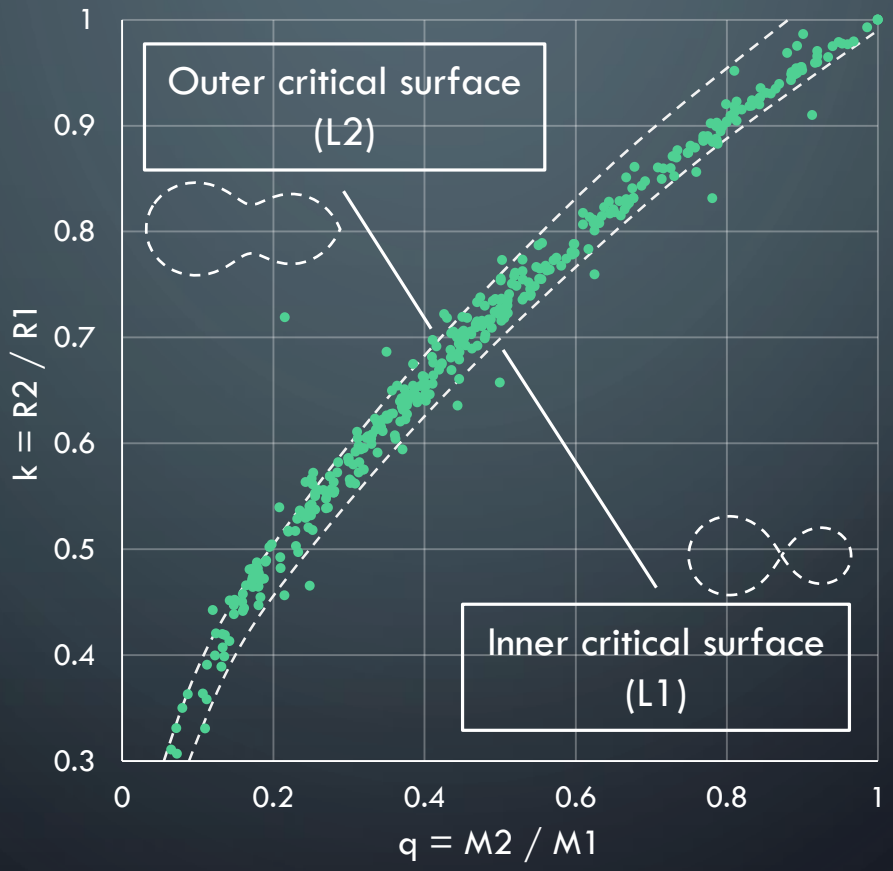
MODELING ECLIPSING BINARIES
MASS RATIO FROM SPECTROSCOPY



Radial velocity curves of RW Com
From Pribulla et al., 2009, AJ, 137

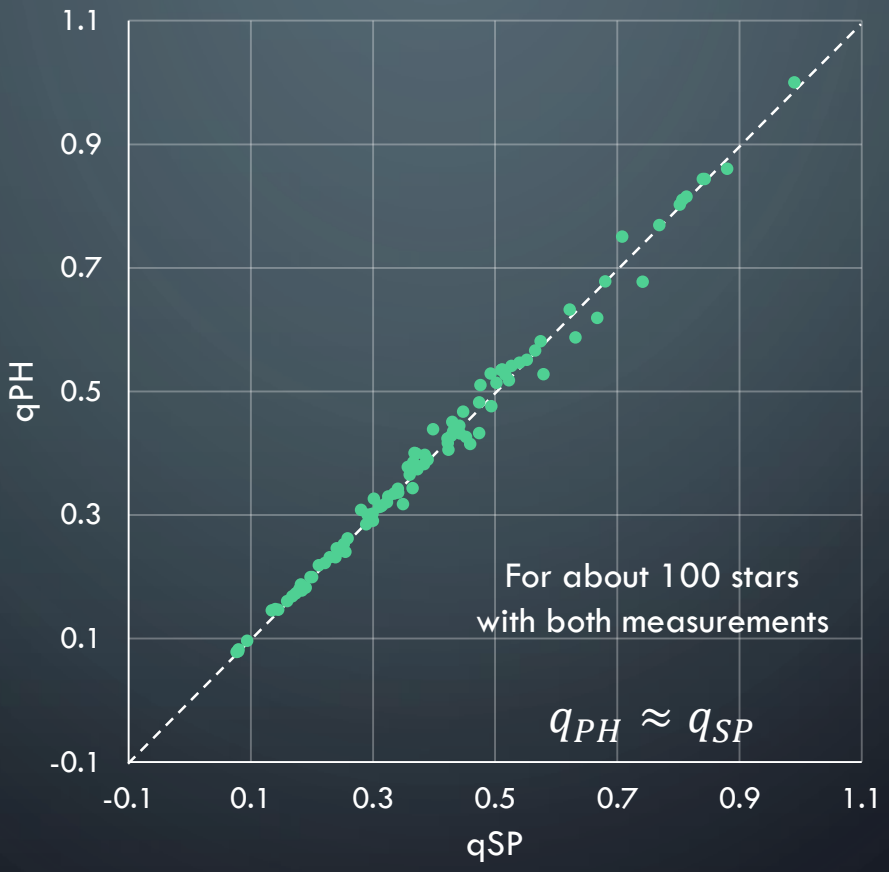
CONTACT BINARIES

MASS RATIO VS RADIUS RATIO



CONTACT BINARIES

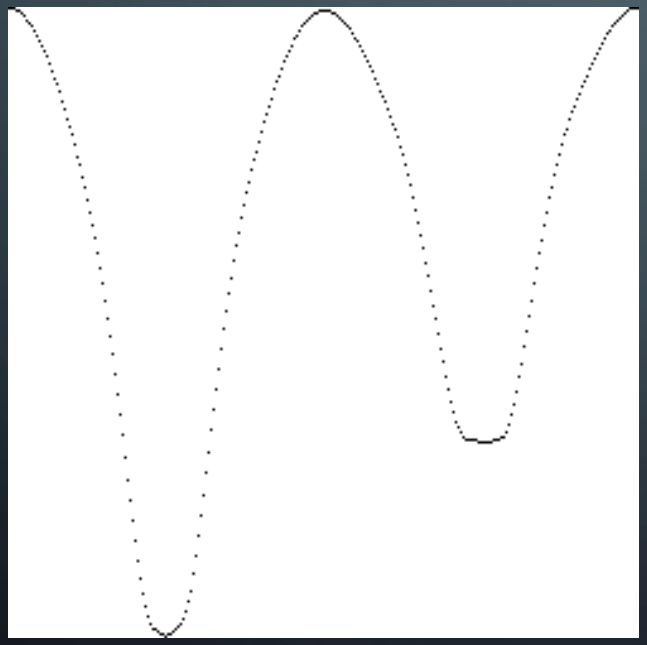
SPECTROSCOPIC VS PHOTOMETRIC MASS RATIO



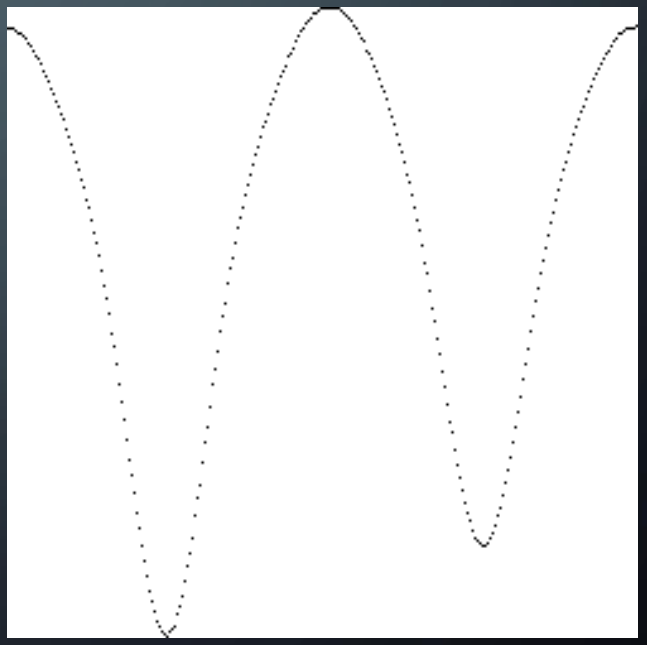
TOTAL VS PARTIAL ECLIPSES

EASY EXAMPLES (FROM KEPLER)

TOTAL



PARTIAL



TOTAL VS PARTIAL ECLIPSES
HARD EXAMPLES (FROM ASAS)

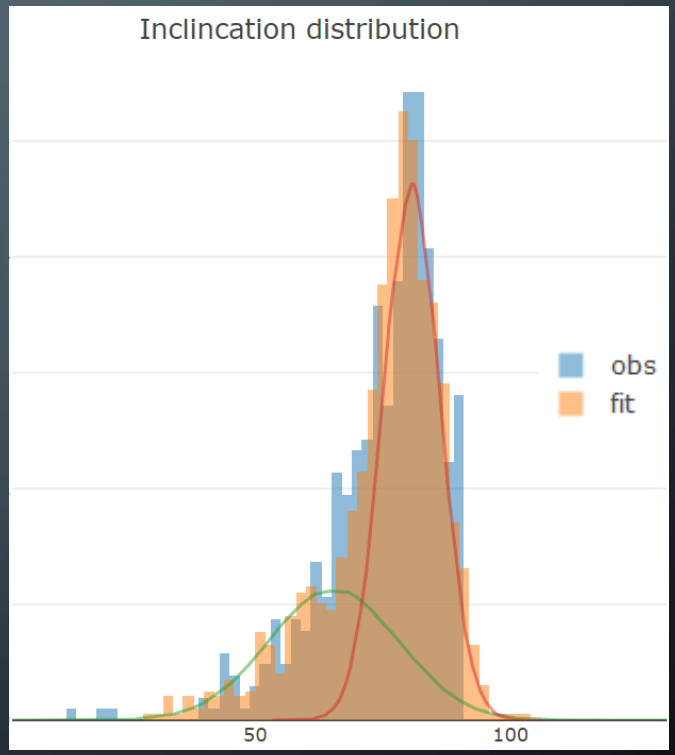
TOTAL

PARTIAL

ADVENTURES WITH COMPUTER VISION

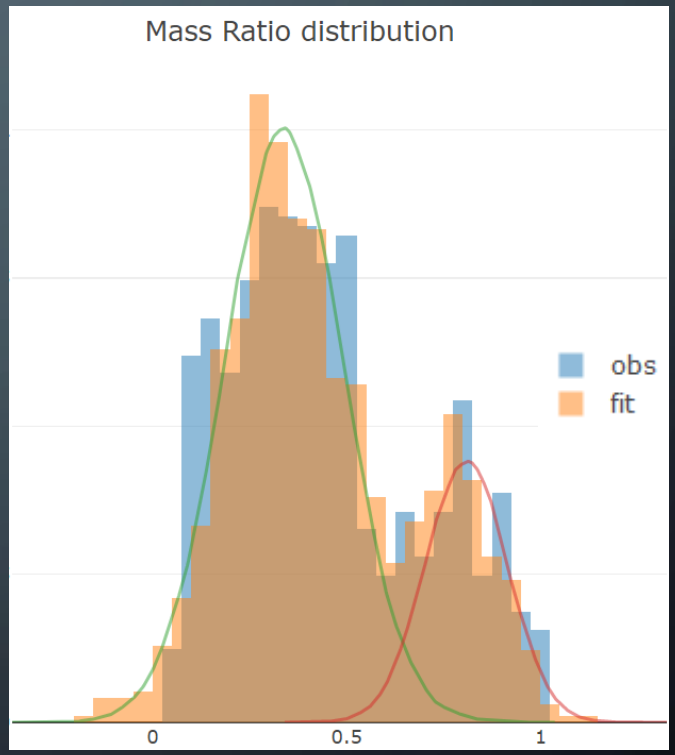
TRAINING DATA

- Synthetic LCs
- Labels based on light contribution of eclipsed star
- Parameters sampled from observed distributions
  - Inclination
  - Mass ratio
  - Temperatures
  - Filling factor (radii)
  - Spot size & location



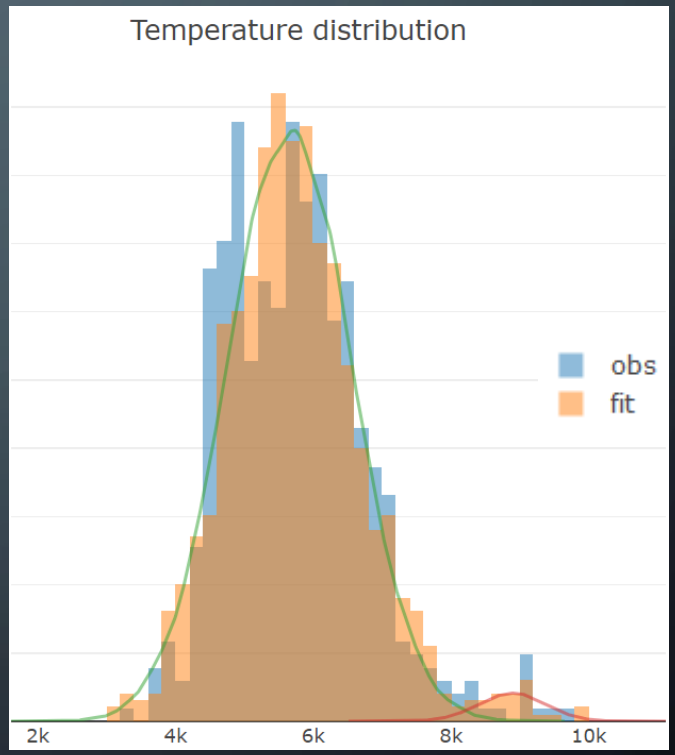
TRAINING DATA

- Synthetic LCs
- Labels based on light contribution of eclipsed star
- Parameters sampled from observed distributions
  - Inclination
  - Mass ratio
  - Temperatures
  - Filling factor (radii)
  - Spot size & location



TRAINING DATA

- Synthetic LCs
- Labels based on light contribution of eclipsed star
- Parameters sampled from observed distributions
  - Inclination
  - Mass ratio
  - Temperatures
  - Filling factor (radii)
  - Spot size & location



TRAINING DATA

- No noise (0.00)

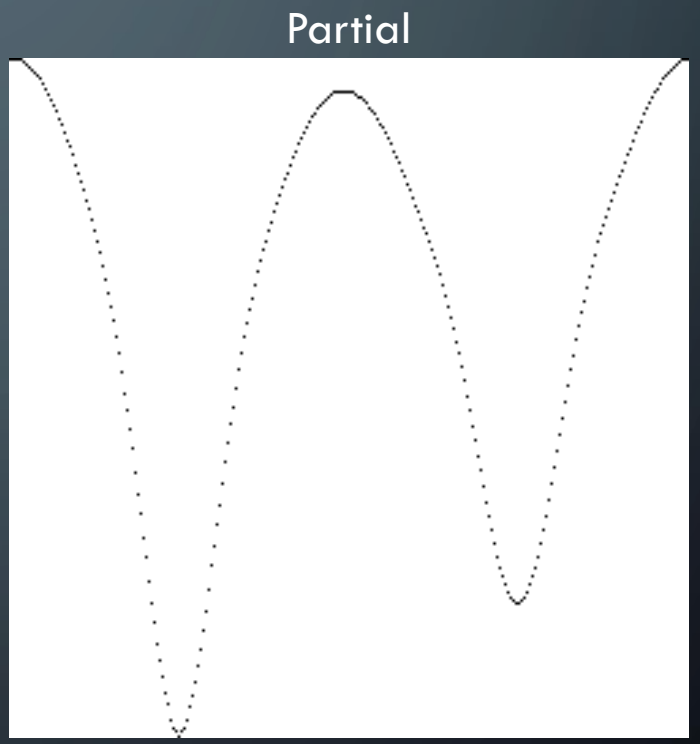
  Like space-based observations

- Moderate noise (0.02)

  Like ground-based observations with 1m class telescopes

- High noise (0.05)

  Like survey data from ASAS, CSS, OGLE etc.



TRAINING DATA

- No noise (0.00)

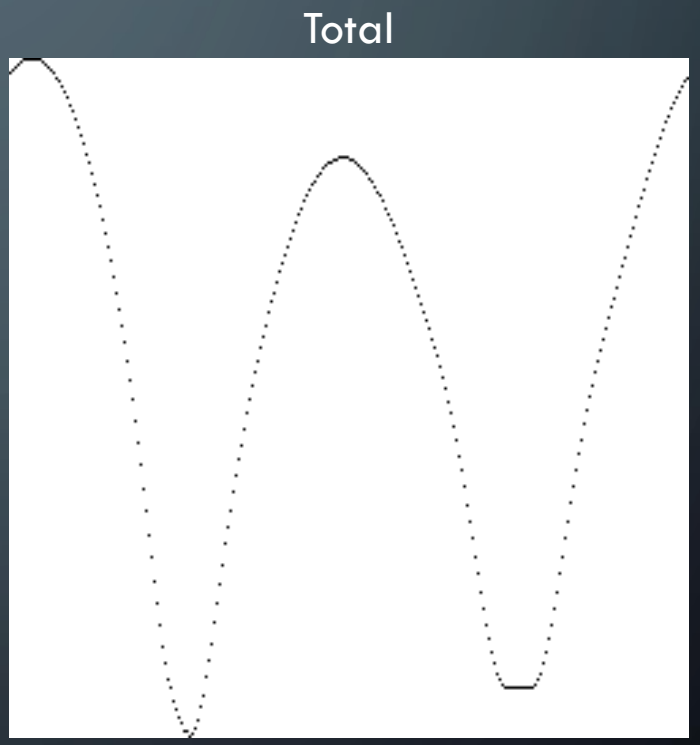
  Like space-based observations

- Moderate noise (0.02)

  Like ground-based observations with 1m class telescopes

- High noise (0.05)

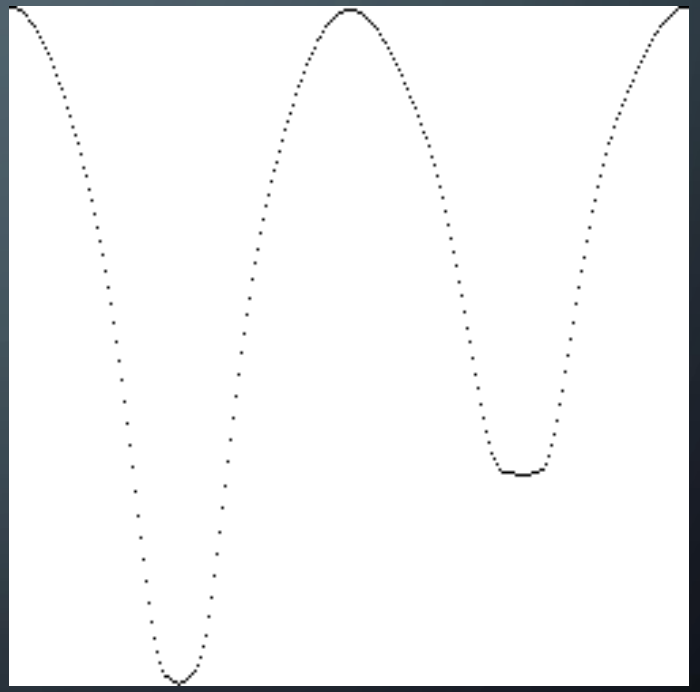
  Like survey data from ASAS, CSS, OGLE etc.



TRAINING DATA

- No noise (0.00)
  - Like space-based observations
- Moderate noise (0.02)
  - Like ground-based observations with 1m class telescopes
- High noise (0.05)
  - Like survey data from ASAS, CSS, OGLE etc.

Kepler



TRAINING DATA

- No noise (0.00)

  Like space-based observations

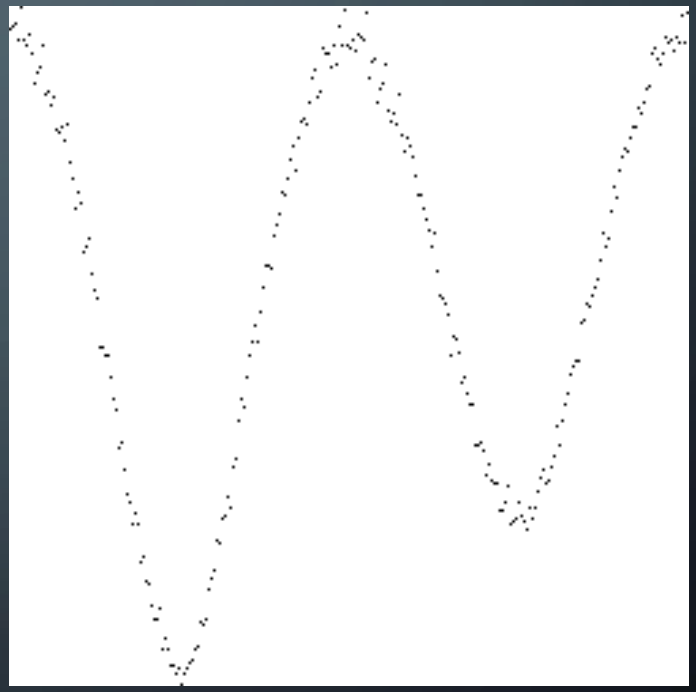
- Moderate noise (0.02)

  Like ground-based observations with 1m class telescopes

- High noise (0.05)

  Like survey data from ASAS, CSS, OGLE etc.

Partial



TRAINING DATA

- No noise (0.00)

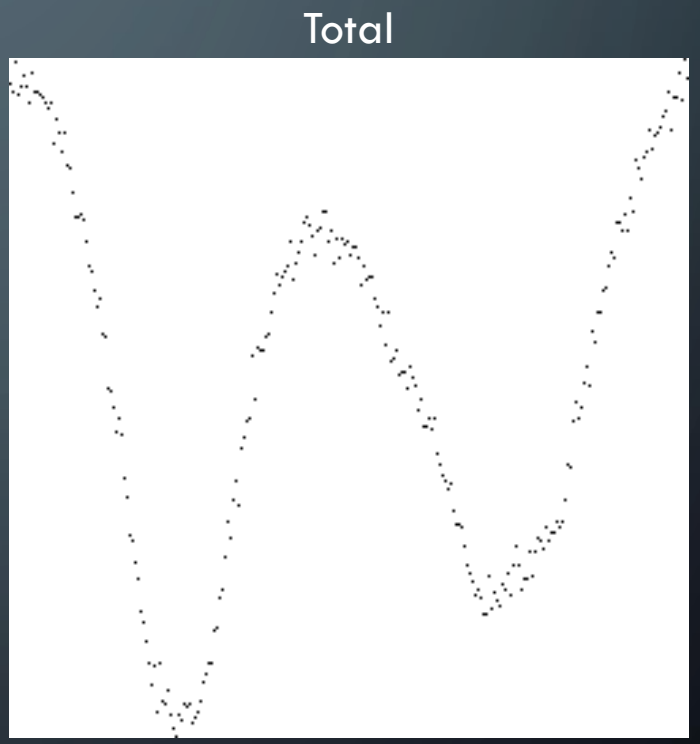
  Like space-based observations

- Moderate noise (0.02)

  Like ground-based observations with 1m class telescopes

- High noise (0.05)

  Like survey data from ASAS, CSS, OGLE etc.



TRAINING DATA

- No noise (0.00)

  Like space-based observations

- Moderate noise (0.02)

  Like ground-based observations with 1m class telescopes

- High noise (0.05)

  Like survey data from ASAS, CSS, OGLE etc.

Partial

TRAINING DATA

- No noise (0.00)

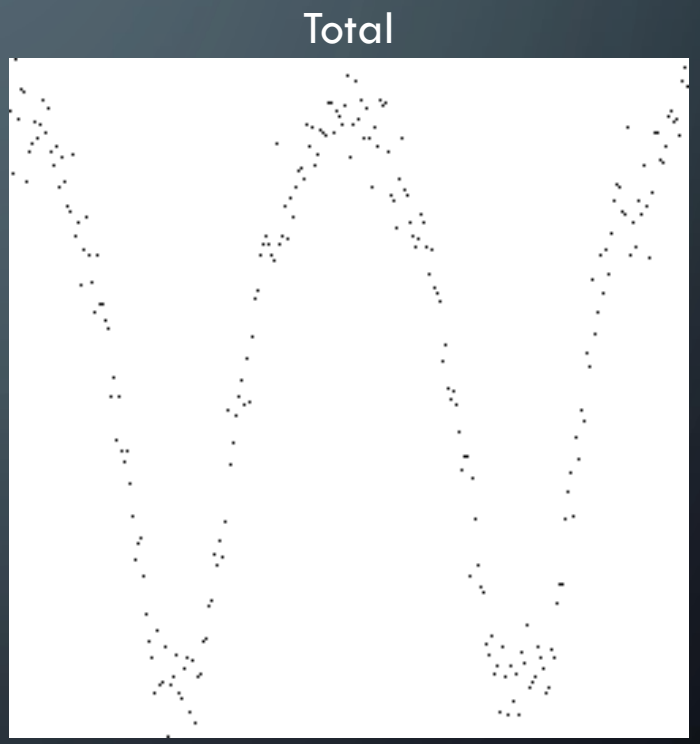
  Like space-based observations

- Moderate noise (0.02)

  Like ground-based observations with 1m class telescopes

- High noise (0.05)

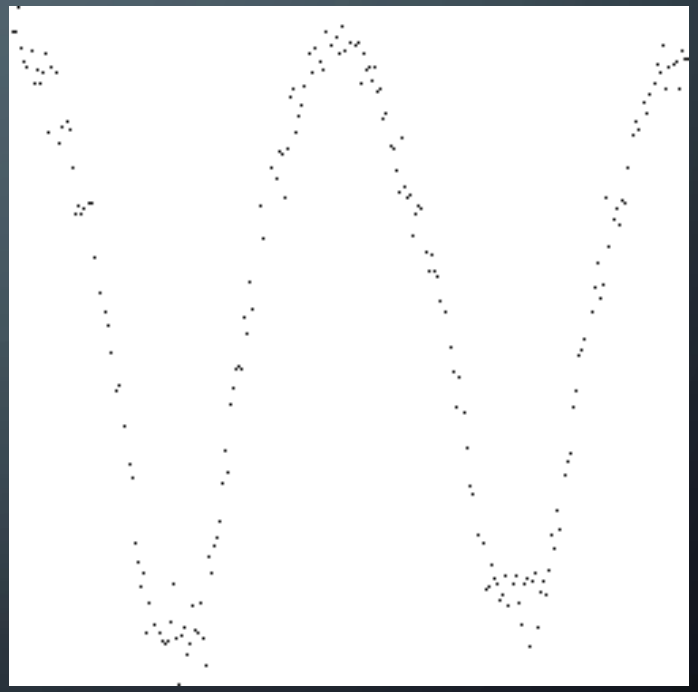
  Like survey data from ASAS, CSS, OGLE etc.



TRAINING DATA

- No noise (0.00)
  - Like space-based observations
- Moderate noise (0.02)
  - Like ground-based observations with 1m class telescopes
- High noise (0.05)
  - Like survey data from ASAS, CSS, OGLE etc.

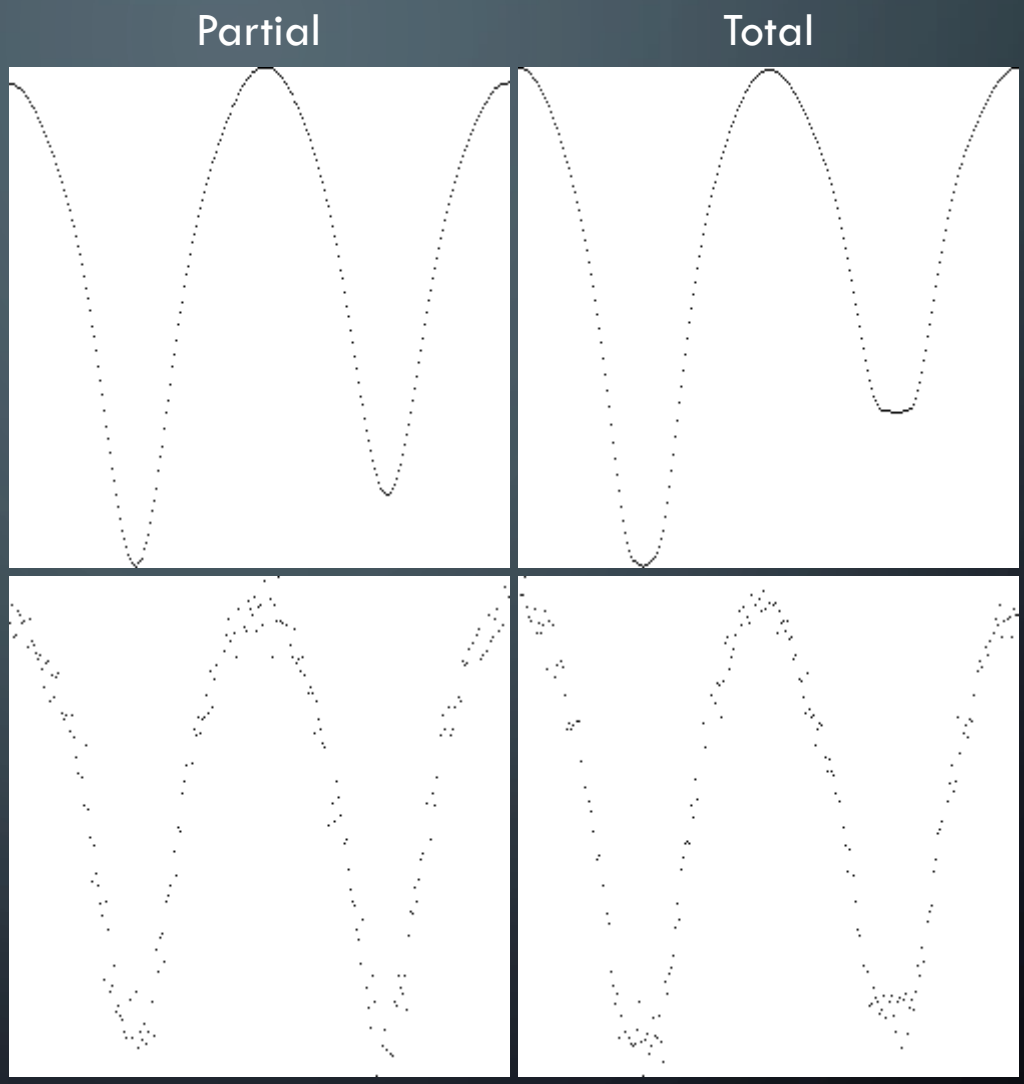
ASAS



TEST DATA

Kepler
300 LCs
45 total eclipses

ASAS
250 LCs
30 total eclipses



PROCESSING OF OBSERVATIONS

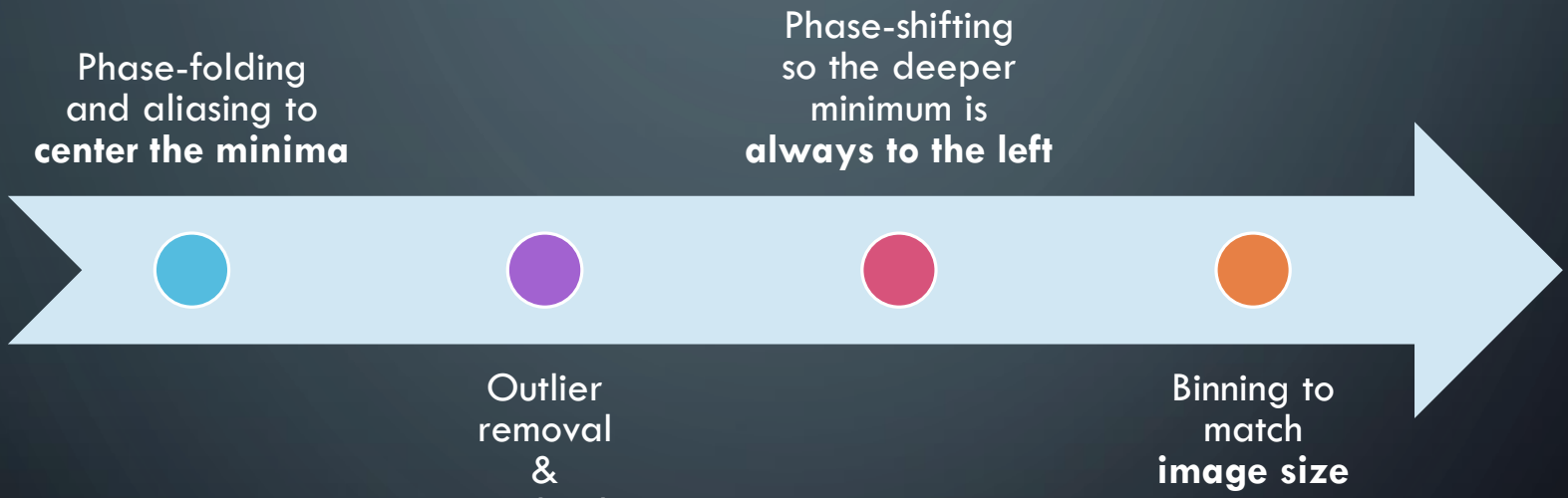
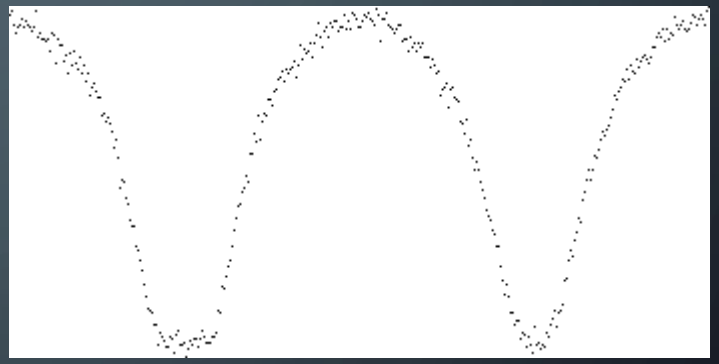


IMAGE SIZES
WITH MODERATE NOISE

100×100
(CNN10)

250×250
(CNN15)



350×175
(CNN15)

COMMERCIAL IMAGE RECOGNITION

http://CustomVision.ai

- Part of Microsoft Cognitive Services
- Free (with bandwidth restrictions)
- Pretrained model for transfer learning
- $F1 \approx 95\%$ on generated test data

CUSTOM-MADE IMAGE RECOGNITION

- Tools:
  - Keras on top of Tensorflow in Python
- Variants:
  - "CNN10"
  - "CNN15"

```
Layer         Output Shape              Param #
=================================================
Conv2D        (None, 100, 100, 16)      160
-------------------------------------------------
MaxPooling2D  (None, 50, 50, 16)        0
-------------------------------------------------
Conv2D        (None, 50, 50, 32)        4640
-------------------------------------------------
MaxPooling2D  (None, 25, 25, 32)        0
-------------------------------------------------
Conv2D        (None, 25, 25, 64)        18496
-------------------------------------------------
MaxPooling2D  (None, 12, 12, 64)        0
-------------------------------------------------
Conv2D        (None, 12, 12, 64)        36928
-------------------------------------------------
MaxPooling2D  (None, 6, 6, 64)          0
-------------------------------------------------
Conv2D        (None, 6, 6, 64)          36928
-------------------------------------------------
MaxPooling2D  (None, 3, 3, 64)          0
-------------------------------------------------
Flatten       (None, 576)               0
-------------------------------------------------
Dense         (None, 512)               295424
-------------------------------------------------
Dense         (None, 1)                 513
=================================================
Total params: 393,089
Trainable params: 393,089
Non-trainable params: 0
```

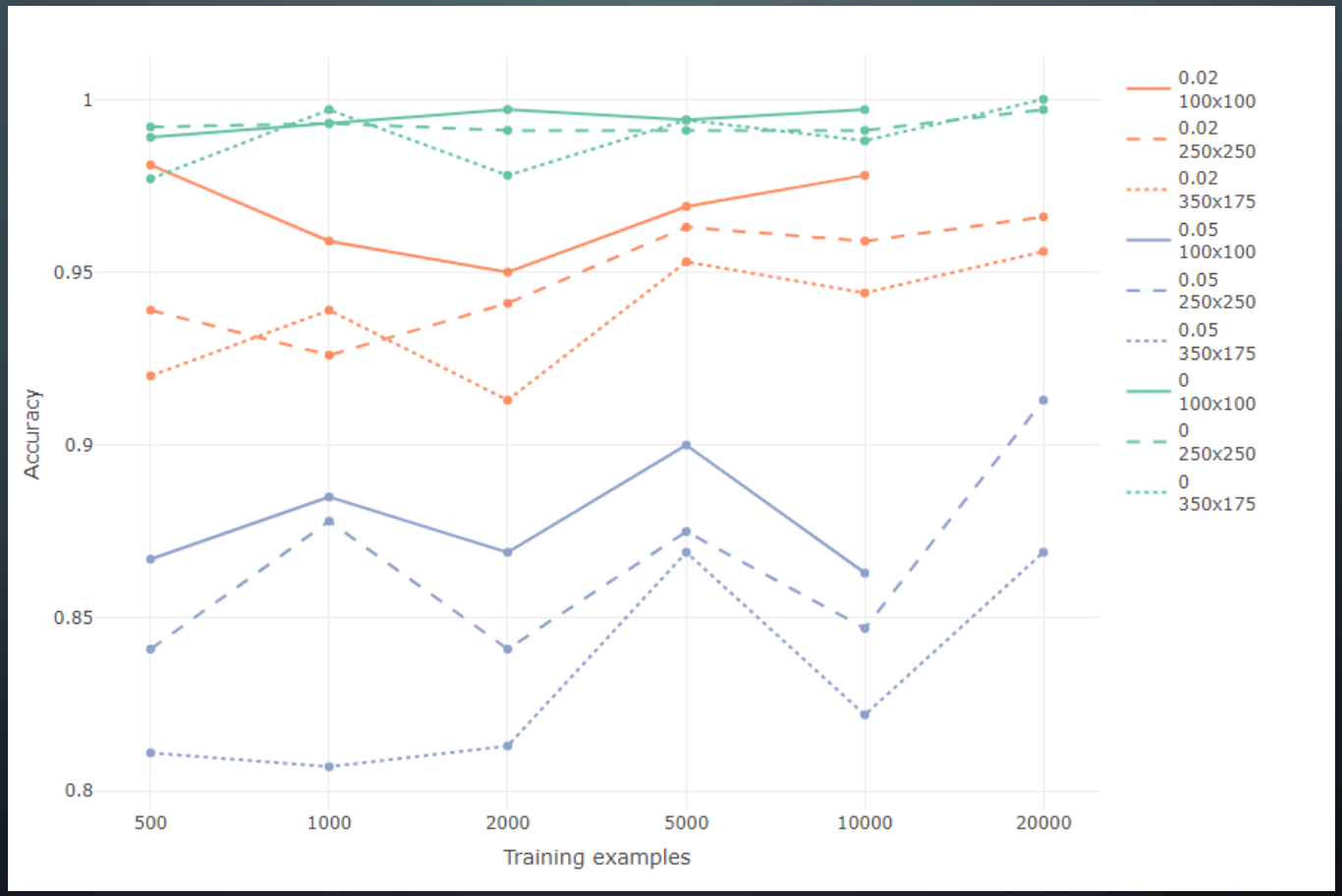
CUSTOM-MADE IMAGE RECOGNITION

- Tools:
  - Keras on top of Tensorflow in Python
- Variants:
  - "CNN10"
  - "CNN15"

Layer	Output Shape	Param #
Conv2D	(None, 250, 250, 16)	160
MaxPooling2D	(None, 125, 125, 16)	0
Conv2D	(None, 125, 125, 32)	4640
MaxPooling2D	(None, 62, 62, 32)	0
Conv2D	(None, 62, 62, 64)	18496
MaxPooling2D	(None, 31, 31, 64)	0
Conv2D	(None, 31, 31, 64)	36928
MaxPooling2D	(None, 15, 15, 64)	0
Conv2D	(None, 15, 15, 64)	36928
MaxPooling2D	(None, 7, 7, 64)	0
Conv2D	(None, 7, 7, 64)	36928
MaxPooling2D	(None, 3, 3, 64)	0
Flatten	(None, 576)	0
Dense	(None, 512)	295424
Dropout	(None, 512)	0
Dense	(None, 1)	513

Total params: 430,017
Trainable params: 430,017
Non-trainable params: 0

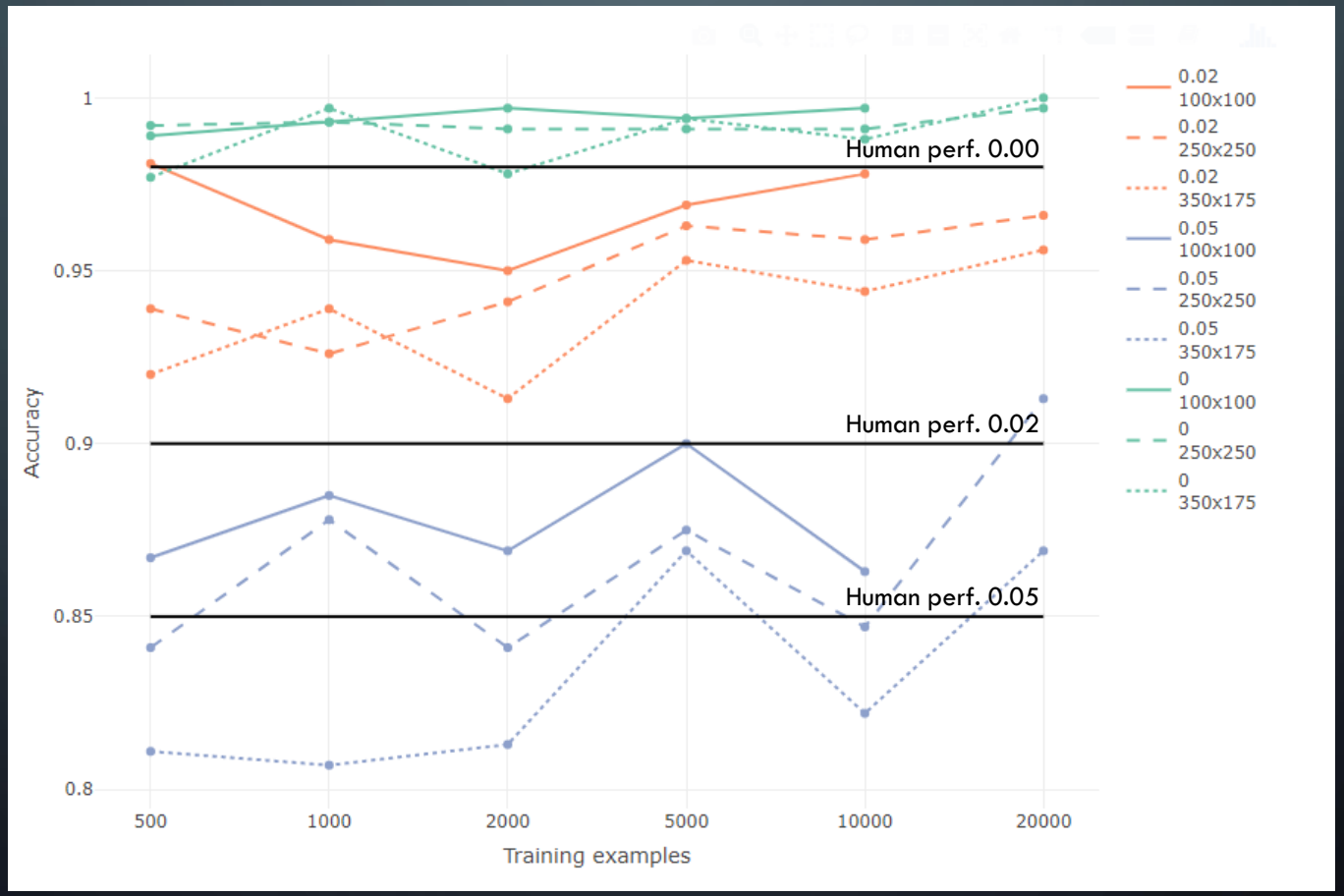
COMPARISON OF MODELS



TOTAL VS PARTIAL ECLIPSES
HUMAN PERFORMANCE

Noise/Size	Balanced Accuracy	F1 Score
0.00	**0.98**	**0.97**
250x250	0.96	0.96
350x175	0.97	0.97
0.02	**0.90**	**0.85**
250x250	0.88	0.81
350x175	0.92	0.90
0.05	**0.85**	**0.75**
250x250	0.84	0.69
350x175	0.85	0.81

COMPARISON OF MODELS



APPLICATION

Accuracy compared to human classification:

- Kepler – 10 misclassifications out of 300 ($\mathrm{Acc} \approx 97\%$)
- ASAS – 23 misclassifications out of 250 ($\mathrm{Acc} \approx 90\%$)
  - But very bad recall $\approx 10\%$

INSIGHTS

- Recognizing total eclipses in noisy data is a hard problem both for humans and for machines
- Models get in trouble with overfitting when training with noisy data
  - Dropout & regularization only have merit with large training sets
- Training on noisy data doesn't improve performance

RESOURCES

- Azure Machine Learning Studio
  studio.azureml.net
- Deep Learning Specialization online courses
  www.coursera.org/specializations/deep-learning
- TensorFlow in Practice Specialization online courses
  www.coursera.org/specializations/tensorflow-in-practice
- Machine Learning Yearning, a book by A. Ng