NEW YORK UNIVERSITY

NEW YORK UNIVERSITY

Machine Learning for Astronomy

Rob Fergus

Dept. of Computer Science,
Courant Institute,
New York University

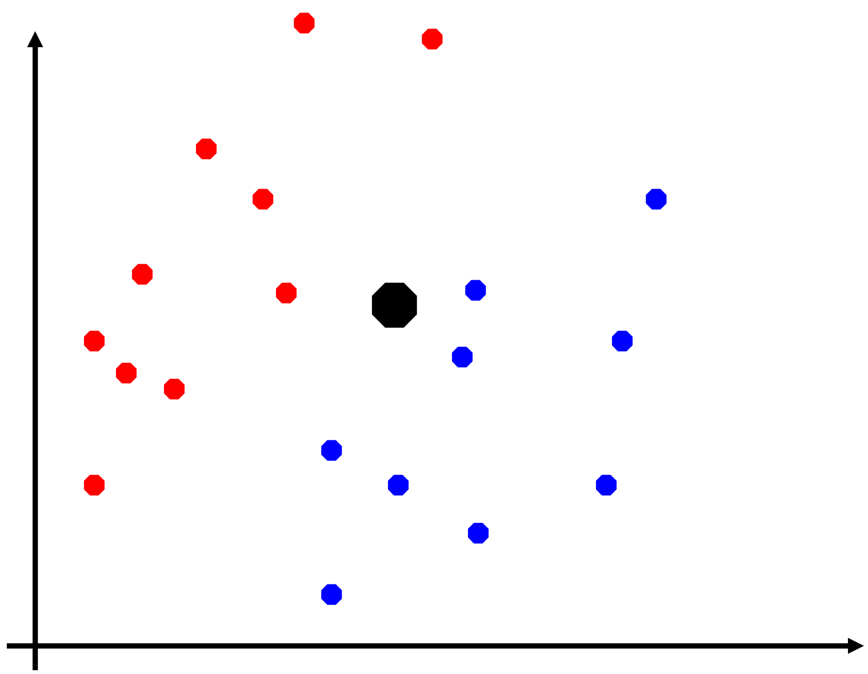
Overview

- High-level view of machine learning
  - Discuss generative & discriminative modeling of data
  - Not exhaustive survey
  - Try to illustrate important ML concepts
- Give examples of these models applied to problems in astronomy
- In particular, exoplanet detection algorithms

Generative vs Discriminative Modeling

- Key distinction in machine learning
- E.g toy classification dataset with labels

  (red=class 1, blue=class 2)

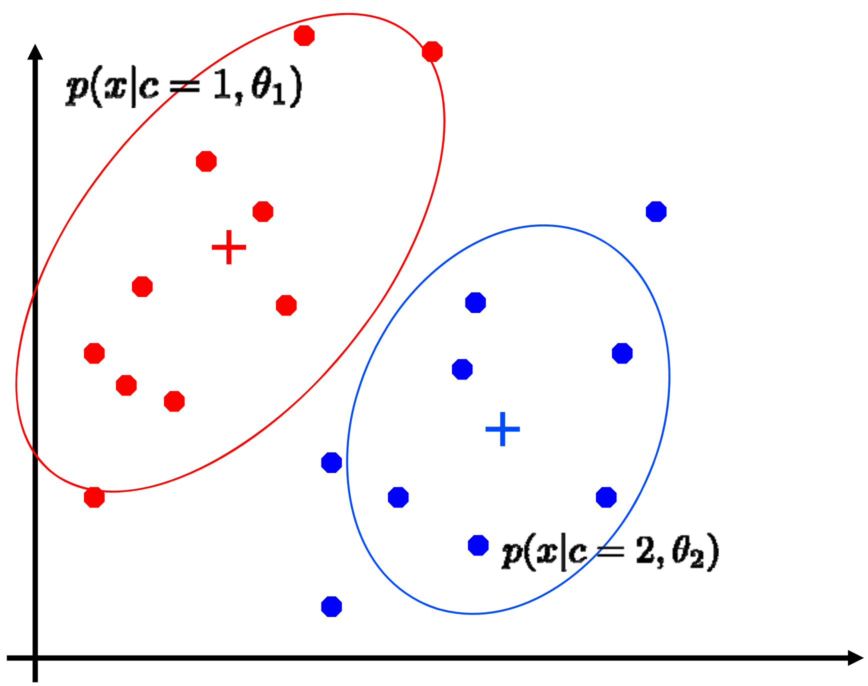


Generative vs Discriminative Modeling

- Given new point x
- Want to compute $\underbrace{p(C|x)}_{\text{Posterior}}$

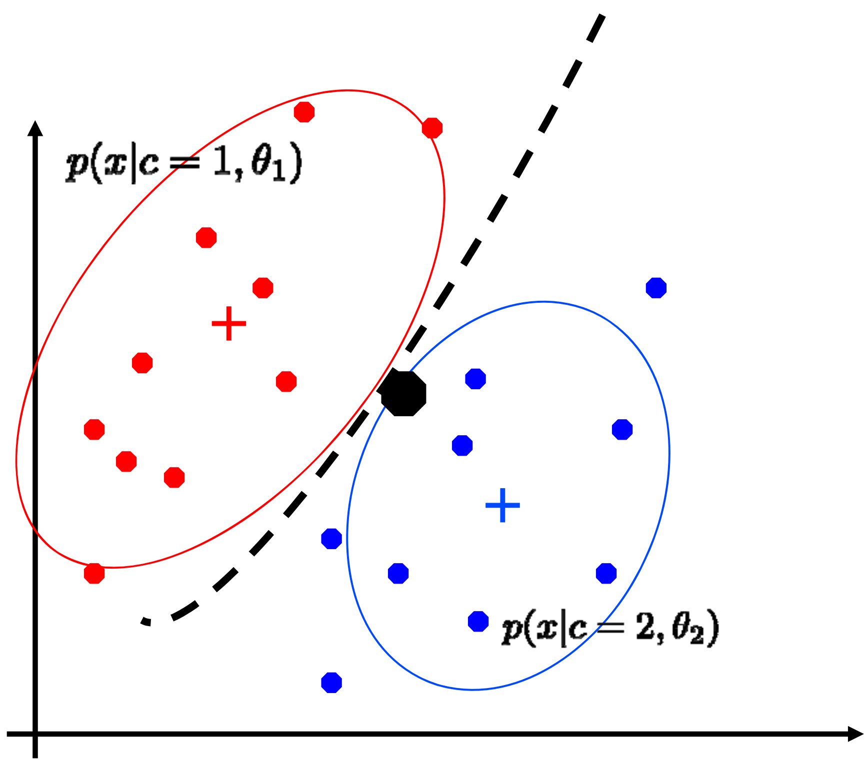
Generative Modeling

- Top-down interpretation of data
  - i.e. adjust model parameters to fit observed data
- E.g. Gaussian model, estimate $\theta = \{\mu_c, \Sigma_c\}$ that maximizes likelihood of data: $p(x|c,\theta)$



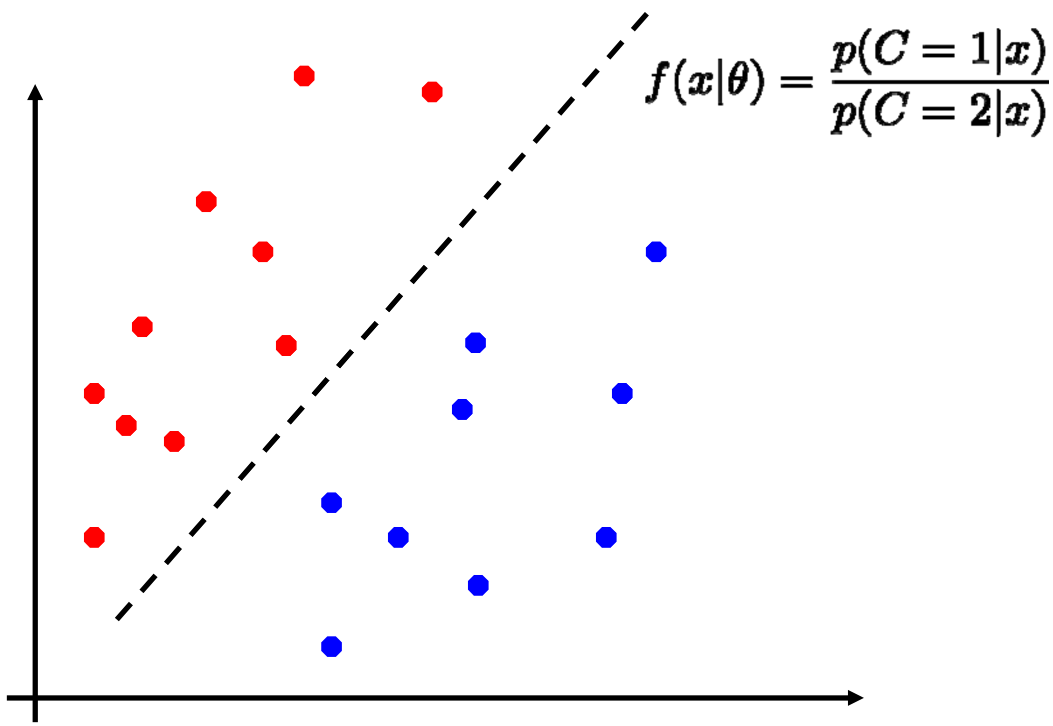
Generative Modeling

- Given new point, we can compute $\frac{p(x|c=1,\theta_1)}{p(x|c=2,\theta_2)}$
- Combine with prior to give posterior
- Likelihood ratio defines decision surface



Discriminative Modeling

- Model posterior directly (no model of data density)
- Fit decision surface directly $\frac{p(C=1|x)}{p(C=2|x)}$
- Bottom-up model: input=x, output=class prediction



Principal Components Analysis (PCA)

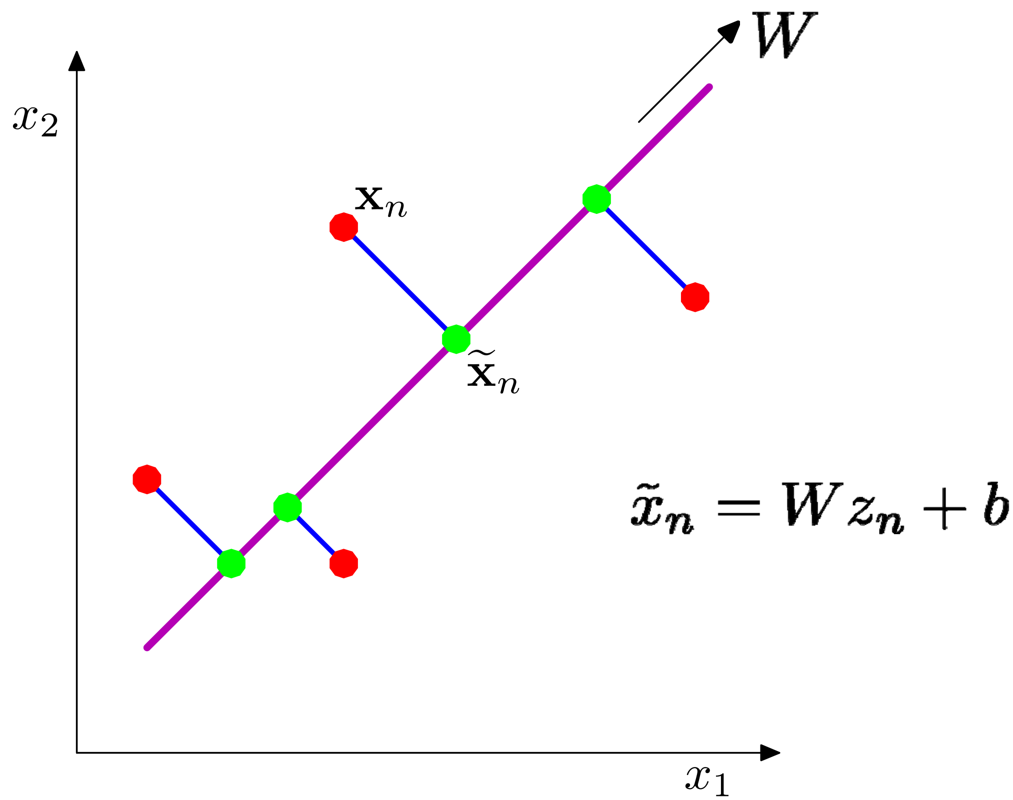
- Example of generative model (objective: compression)
- Observed data points: $x_n \in R^D, \quad n = 1, 2, \ldots, N$
- Hidden manifold coords.: $z_n \in R^M, \quad n = 1, 2, \ldots, N$
- Hidden linear mapping: $\tilde{x}_n = W z_n + b \quad W \in R^{D \times M}$

  $b \in R^{D \times 1}$

$$J(z, W, b | x, M) = \sum_{n=1}^{N} \|x_n - \tilde{x}_n\|^2 = \sum_{n=1}^{N} \|x_n - W z_n - b\|^2$$

- Find global optimum via eigendecomposition of sample covariance matrix

Principal Components Analysis (PCA)



C. Bishop, Pattern Recognition & Machine Learning

$$J(z, W, b | x, M) = \sum_{n=1}^{N} \|x_n - \tilde{x}_n\|^2 = \sum_{n=1}^{N} \|x_n - W z_n - b\|^2$$

Probabilistic Principal Components Analysis (PPCA)

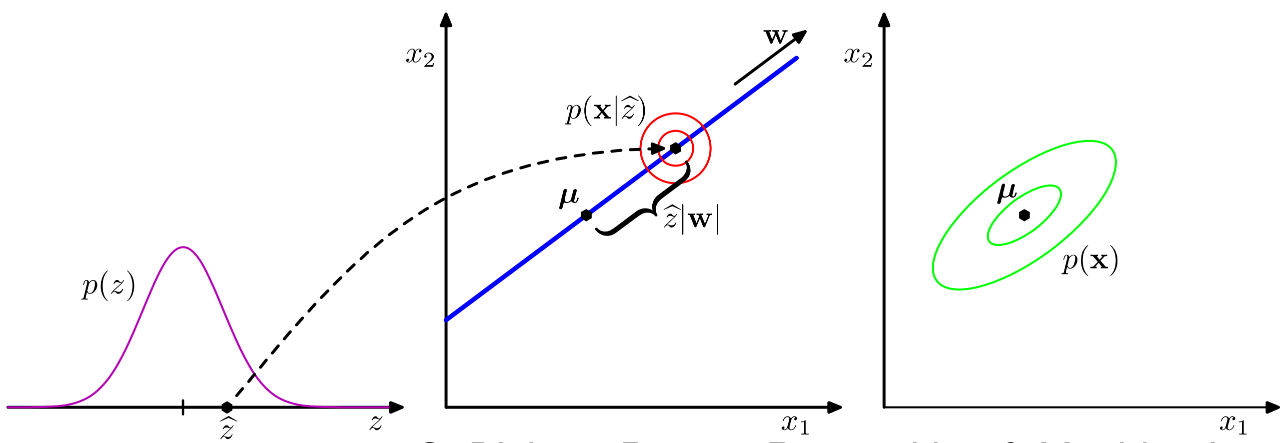
- Data is linear function of low-dimensional latent coordinates, plus Gaussian noise.

$$p(x_i \mid z_i, \theta) = \mathcal{N}(x_i \mid W z_i + \mu, \Psi) \qquad p(z_i \mid \theta) = \mathcal{N}(z_i \mid 0, I)$$

$$p(x_i \mid \theta) = \mathcal{N}(x_i \mid \mu, WW^T + \Psi)$$

low rank covariance parameterization

$$\Psi = \sigma^2 I$$

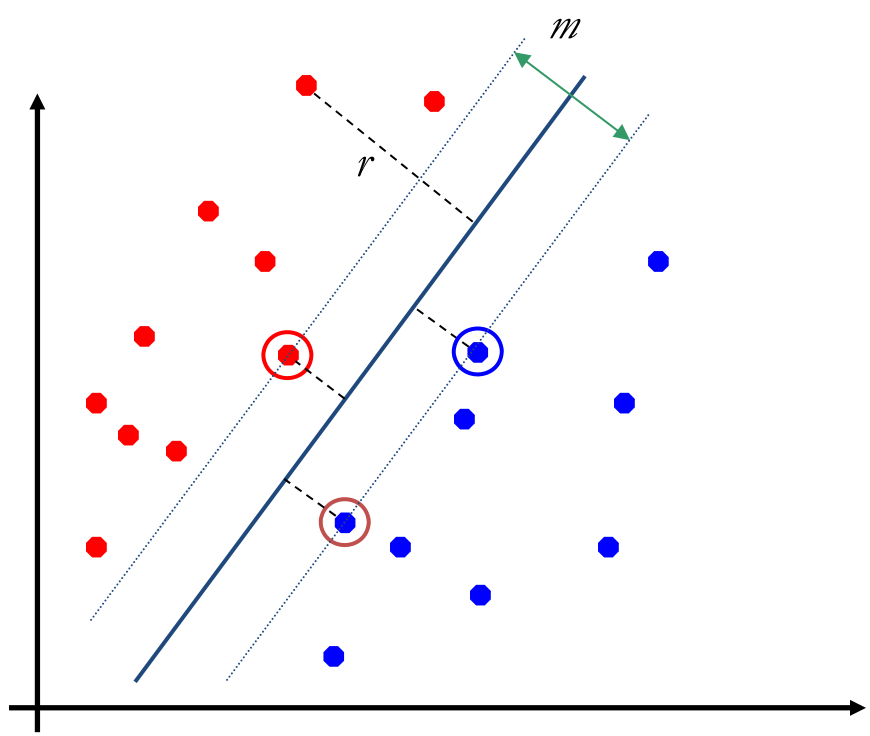


C. Bishop, Pattern Recognition & Machine Learning

Support Vector Machines (SVMs)

[Cortes; Vapnik; Schölkopf; others]

- Classic discriminative approach
- Formal notion of margin m, to aid generalization
- "Kernel trick" to give non-linear decision surfaces



Comparison

<u>Generative Models</u>

+ Labels not essential
+ Unsupervised or supervised
- • Models whole density
+ Interpretable result
- Can be hard to specify model structure

<u>Discriminative Models</u>

- **Need labels**
- Supervised only
- • Model only fits decision surface
+ Fast to evaluate
+ Can be very powerful

Detour

Deep Neural Networks for
Natural Image Classification

Deep Learning

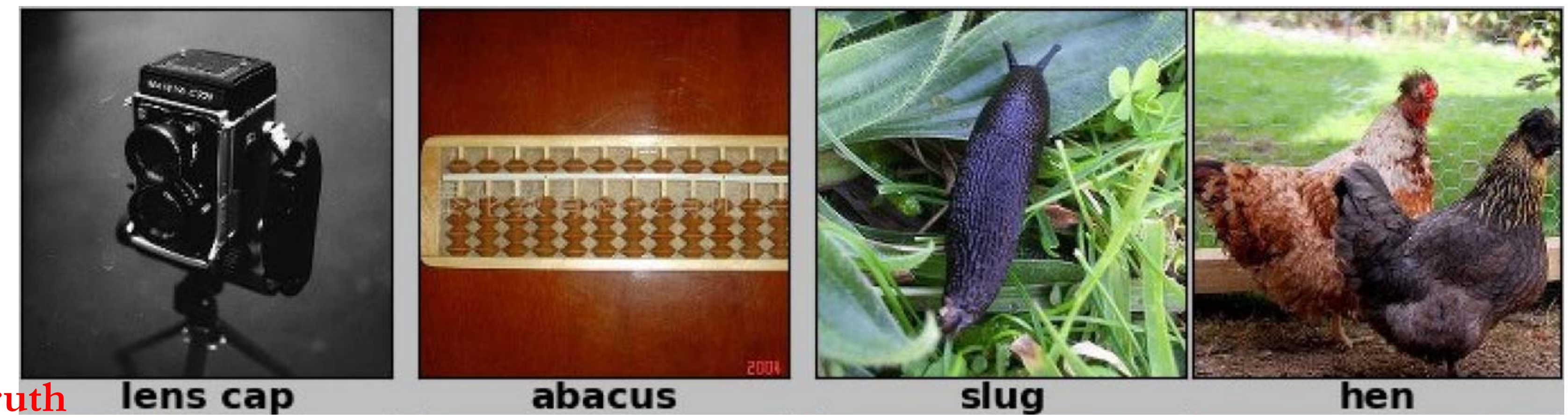
- Big gains in performance in last few years on:
  - Vision
  - Audition
  - Natural language processing

- Three ingredients:
  1. **Discriminative** neural network models (supervised training)
  2. Big labeled datasets
  3. Lots of computation

Computer Vision

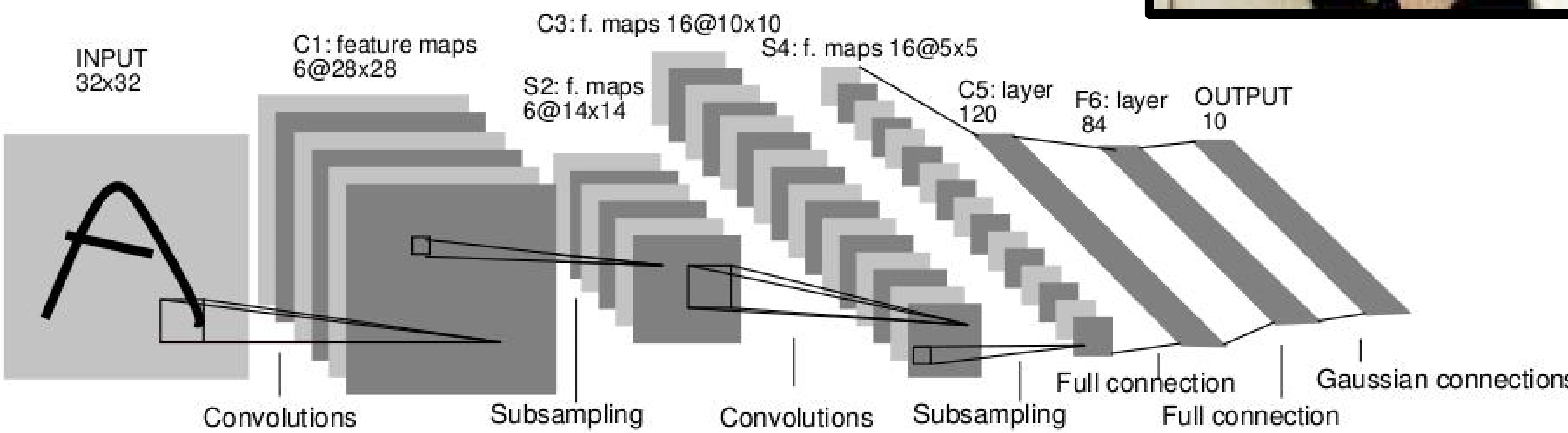
- Image Recognition
  - Input: Pixels
  - Output: Class Label

Ground Truth



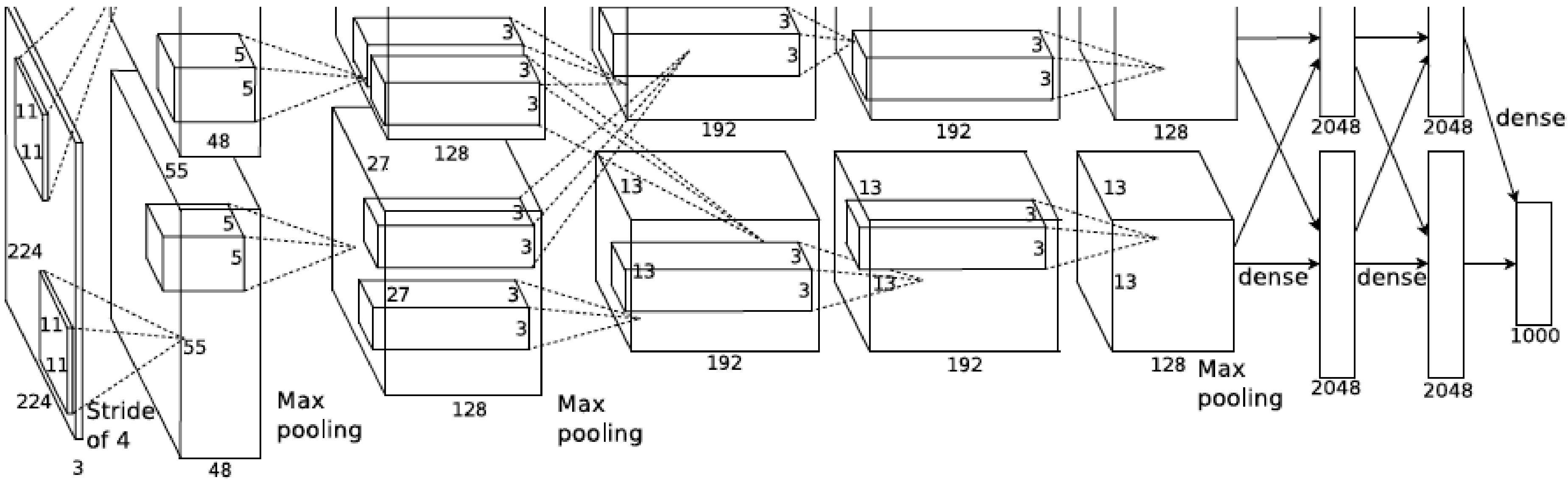
Convolutional Neural Networks

- LeCun et al. 1989
- Neural network with specialized connectivity structure



Convolutional Neural Network

- Krizhevsky et al. [NIPS2012]
  - 8 layer Convolutional network model [LeCun et al. '89]
  - Trained on 1.2 million ImageNet images (with labels)
  - GPU implementation (50x speedup over CPU)



- 7 hidden layers, 650,000 neurons, 60,000,000 parameters
- Trained on 2 GPUs for a week

Big Image Datasets

IMAGENET

- Stanford Vision group [Deng et al. 2009]
- ~14 million labeled images, 20k classes
- Images gathered from Internet
- Human labels via Amazon Turk

COCO
Common Objects in Context

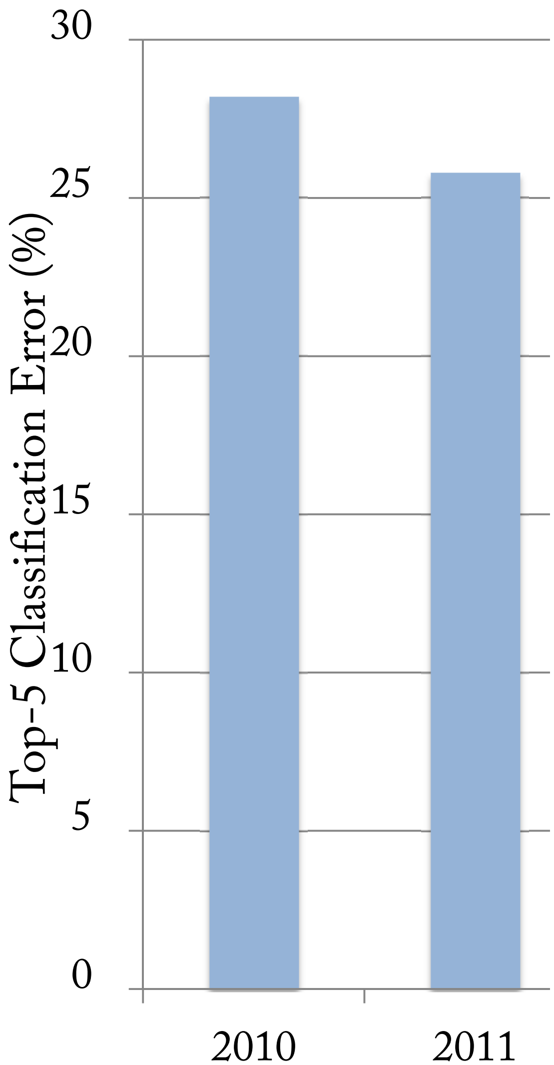
- Microsoft + academic collaboration
- 2 million objects in natural settings
- Human labels via Amazon Turk

Powerful Hardware

- Deep neural nets highly amenable to implementation on Graphics Processing Units (GPUs)
  - Mainly matrix multiply, 2D convolution operations

- Latest generation nVidia GPUs (Pascal) deliver 10 TFlops / card
  - Faster than fastest super-computer in world in 2000

ImageNet Performance over time



[Russakovsky et al. IJCV 2015]

Examples

- From Clarifai.com

Predicted Tags:

food	(16.00%)
dinner	(3.10%)
bbq	(2.90%)
market	(2.50%)
meal	(1.40%)
turkey	(1.40%)
grill	(1.30%)
pizza	(1.30%)
eat	(1.10%)
holiday	(1.00%)

Stats:

Size: 247.24 KB

Time: 110 ms

Examples

- From Clarifai.com

Predicted Tags:

ship	(2.30%)
helsinki	(1.80%)
fish	(1.40%)
port	(1.10%)
istanbul	(1.10%)
beach	(1.00%)
denmark	(1.00%)
copenhagen	(0.90%)
sea	(0.80%)
boat	(0.80%)

Examples

- From Clarifai.com

Predicted Tags:

barcelona	(6.50%)
street	(3.00%)
cave	(2.20%)
sagrada	(1.90%)
old	(1.80%)
night	(1.40%)
familia	(1.40%)
jerusalem	(1.40%)
guanajuato	(1.10%)
alley	(1.00%)

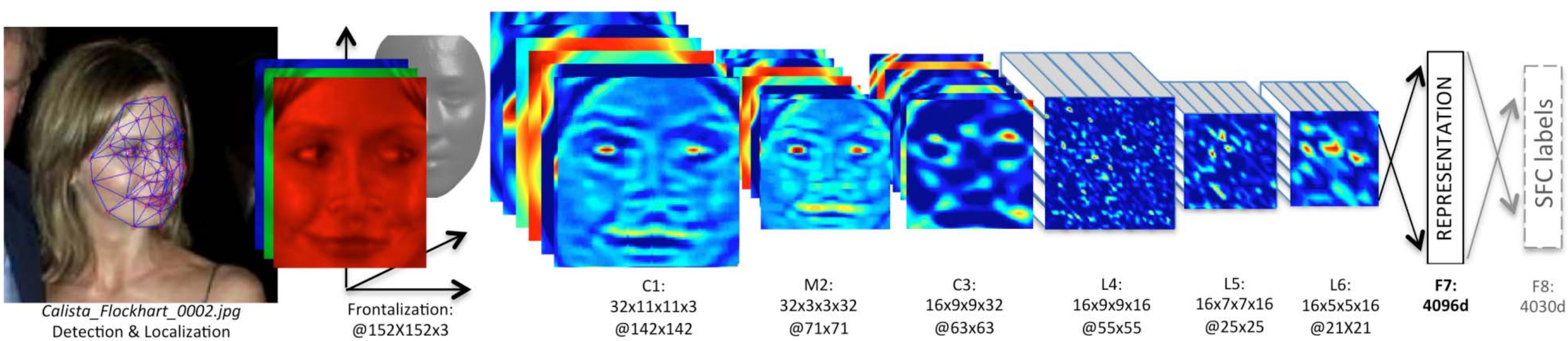
Stats:

Size: 278.96 KB

Time: 113 ms

Industry Deployment

- Widely used in Facebook, Google, Microsoft
- Face recognition, image search, photo organization….
- Very fast at test time (~100 images/sec/GPU)



[Taigman et al. DeepFace: Closing the Gap to Human-Level Performance in Face Verification, CVPR'14]

Success of DeepNets

- ConvNets work great for other types of data:
  - Medical imaging
  - Speech spectrograms
  - Particle physics traces

- Other types of deep neural nets (Recurrent Nets) work well for natural language

- **But need lots and lots of labeled data!!**

End of Detour

Galaxy Morphology Classification

- https://www.galaxyzoo.org/

- Crowd-sourced labels for different galaxy shapes

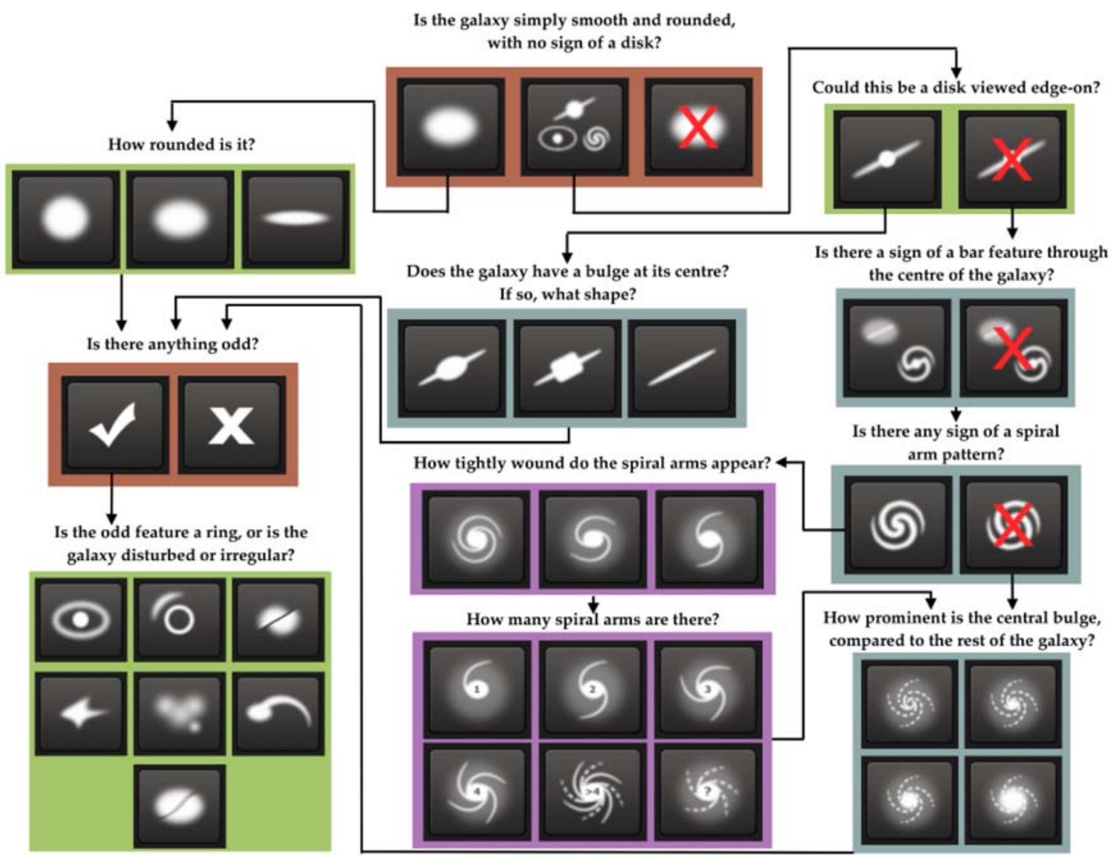
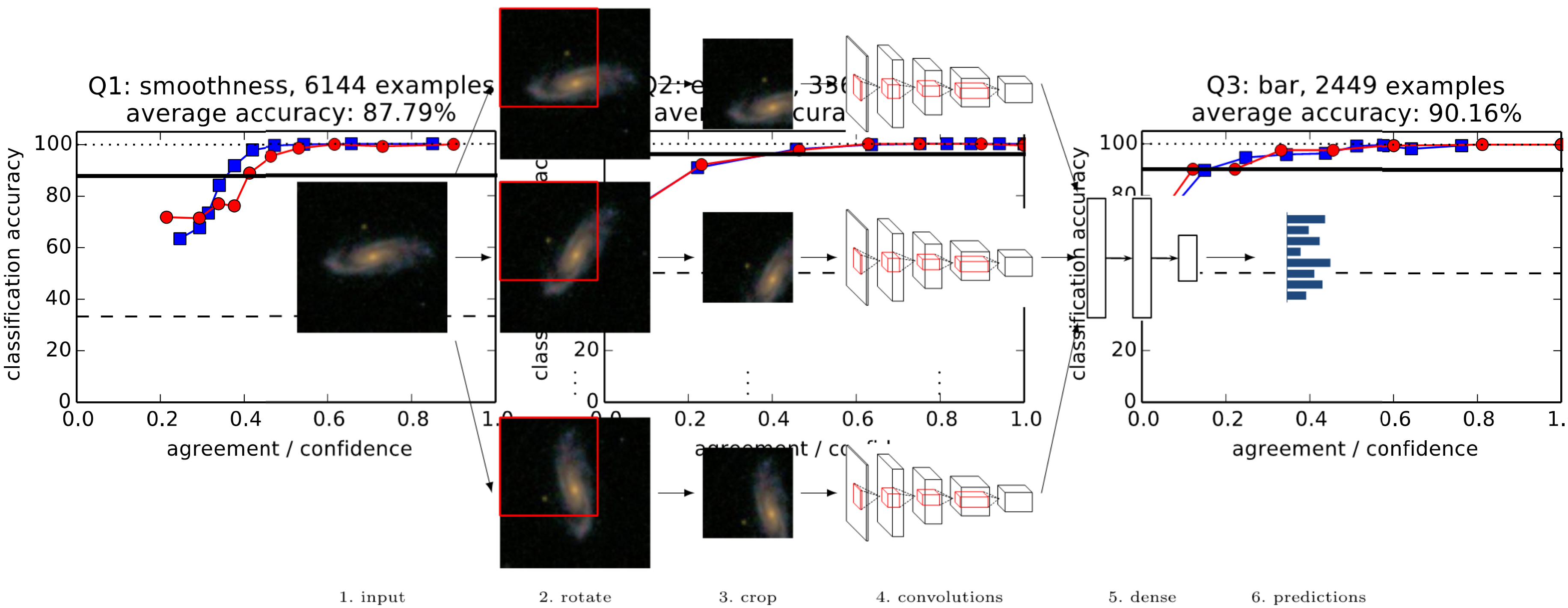


Figure 1. The Galaxy Zoo 2 decision tree. Reproduced from Figure 1 in Willett et al. (2013).

Galaxy Morphology Classification

[Rotation-invariant convolutional neural networks for galaxy morphology prediction, Dieleman, Willett, Dambre, R. Astron. Soc. March 2015]

- Train ConvNet on Galaxy Zoo data/labels
  - Won Kaggle competition
- Closely matches human performance



Direct Detection of Exoplanets using the S4 Algorithm

[Spatio-Spectral Speckle Suppression]

Rob Fergus [1], David W. Hogg [2],
Rebecca Oppenheimer [3], Doug Brenner [3], Laurent Pueyo [4]

1 Dept. of Computer Science, Courant Institute, New York University

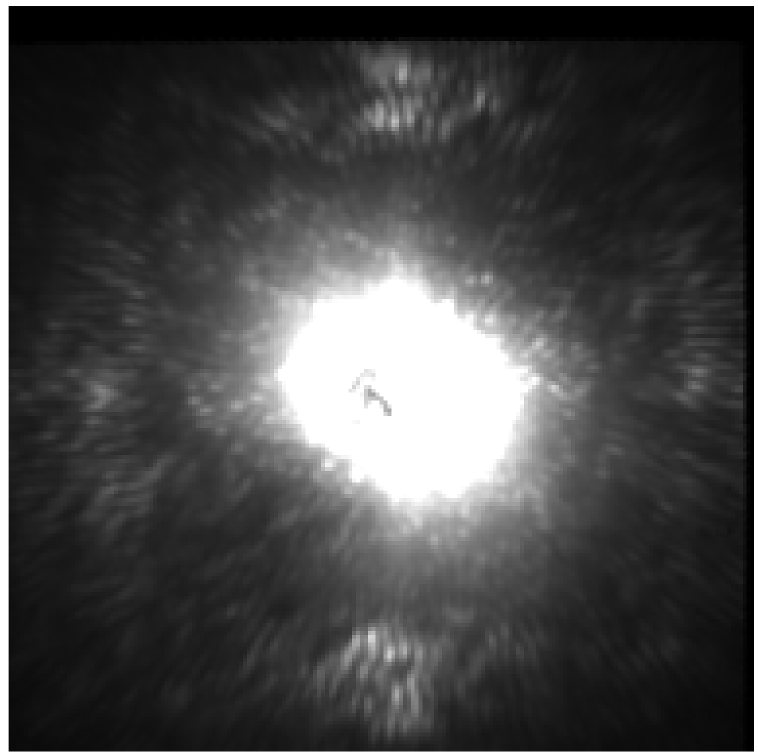
2 Center for Cosmology & Particle Physics, Dept. of Physics, New York University

3 Dept. of Astrophysics American Museum of Natural History

4 Space Telescope Science Institute

P1640 Data Cubes

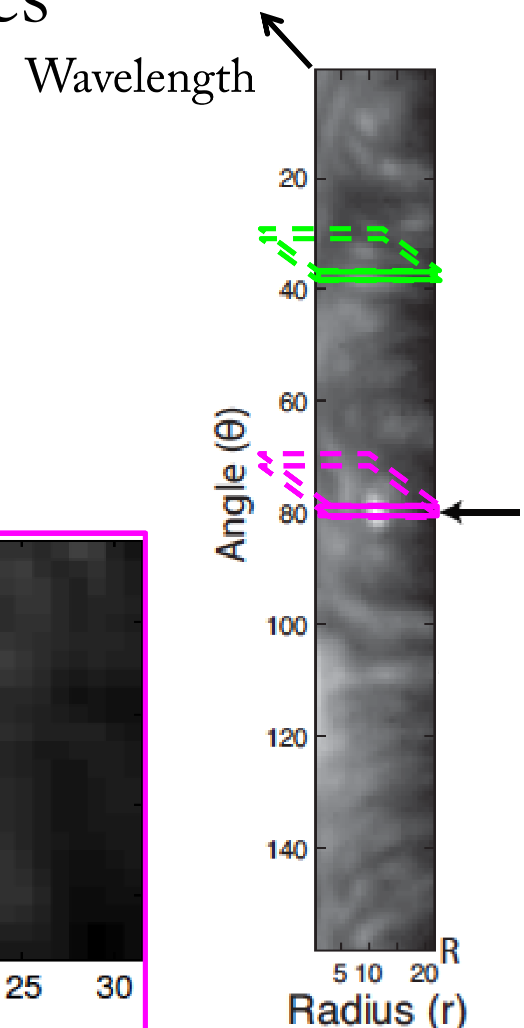
- Each exposure gives 32 wavelength bands
  (near IR 950-1770nm)
- Speckles are diffraction artifacts

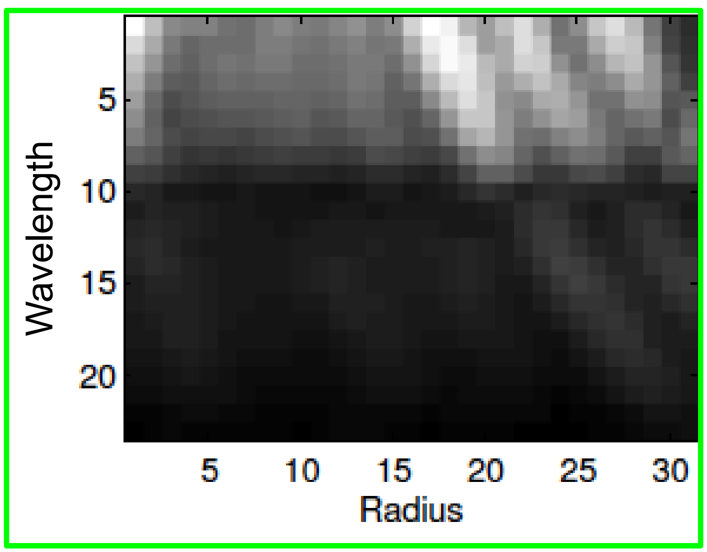


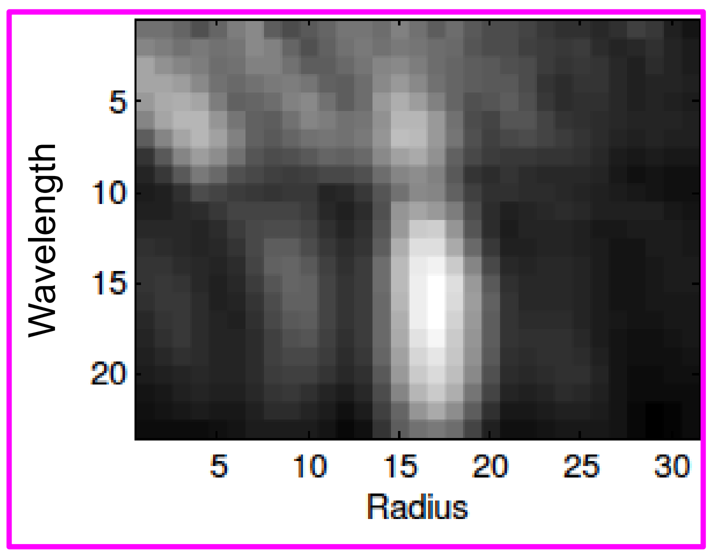
- Move radially with wavelength
- Planet stationary

Use Polar Representation

- Speckles become diagonal structures
- Planet is vertical
  - Key to separating the two
- Assume: independence to angle and exposure







Three versions of S4

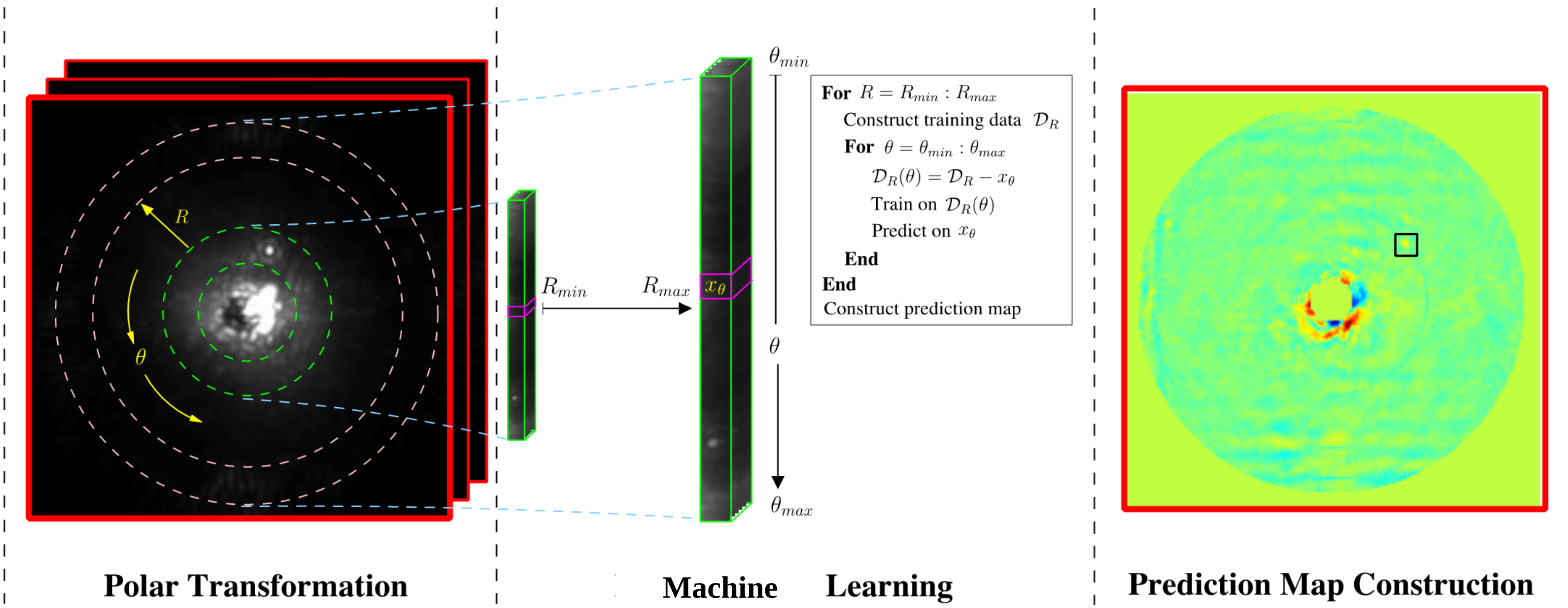
1. S4 Detect [Generative, PCA-based detection model]
2. DS4 Detect [Discriminative, SVM-based detection model]
   - [Munandet, Schölkopf, Oppenhiemer, Nilsson, Veicht]
3. S4 Spectra [Generative, spectra estimation model]

- All use same representation
- Just different ML approach

- Lots of related algorithms (KLIP, LOCI etc.)

Leave-Out Strategy for Detection (S4 Detect & DS4)

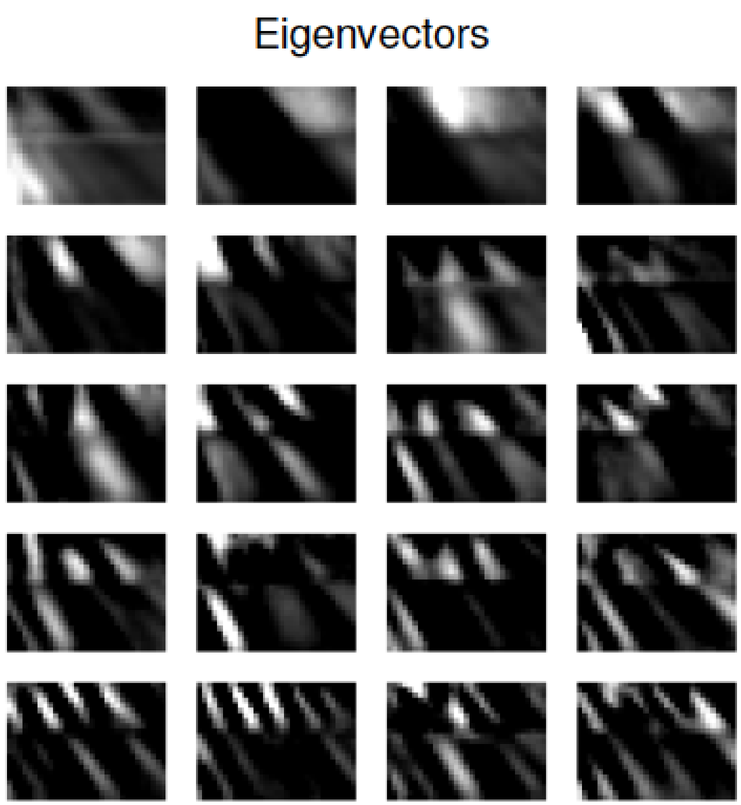
- Separate slices within annulus into train/test
- Train new model for each location

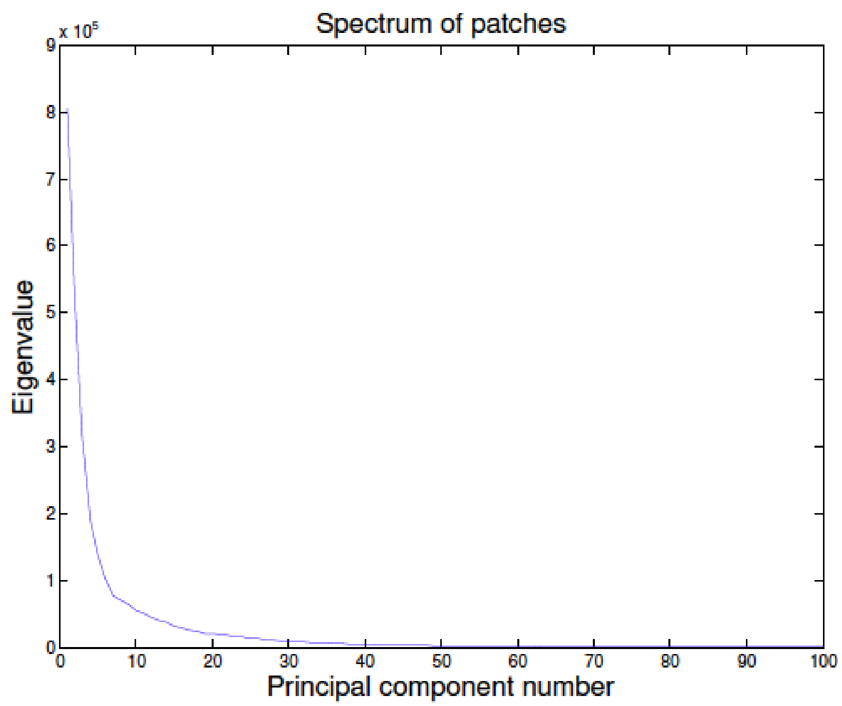


1. S4 Detect

S4 Detect PCA Model

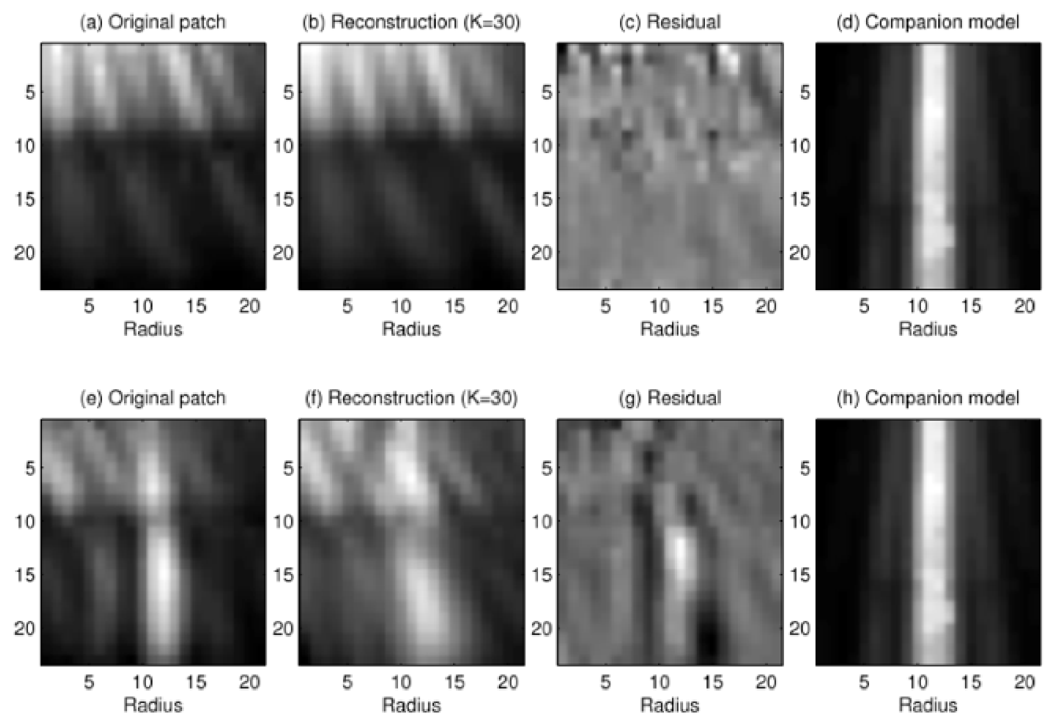
- Trained for each location





S4 Detect Summary

- Build PCA basis on training set
- Fit PCA model to test patches
- Companion should appear in residual
- Correlate residual with (fixed) companion model

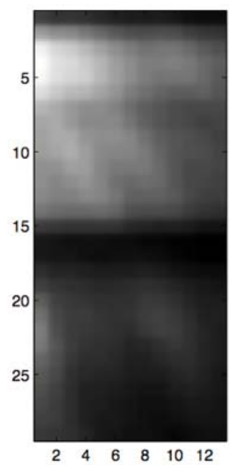


2. DS4 Detect

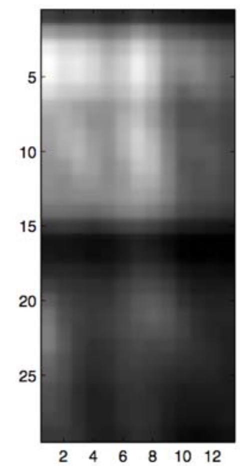
DS4 Detect Summary

- Generate training set
  - Discriminative models need labeled examples
  - Negative examples: take directly from data
  - Positive examples: add artificial companion (different spectra)

Negative Example
(image patch with no planet)



Positive Example
(fake companion realistic brightness and spectra)

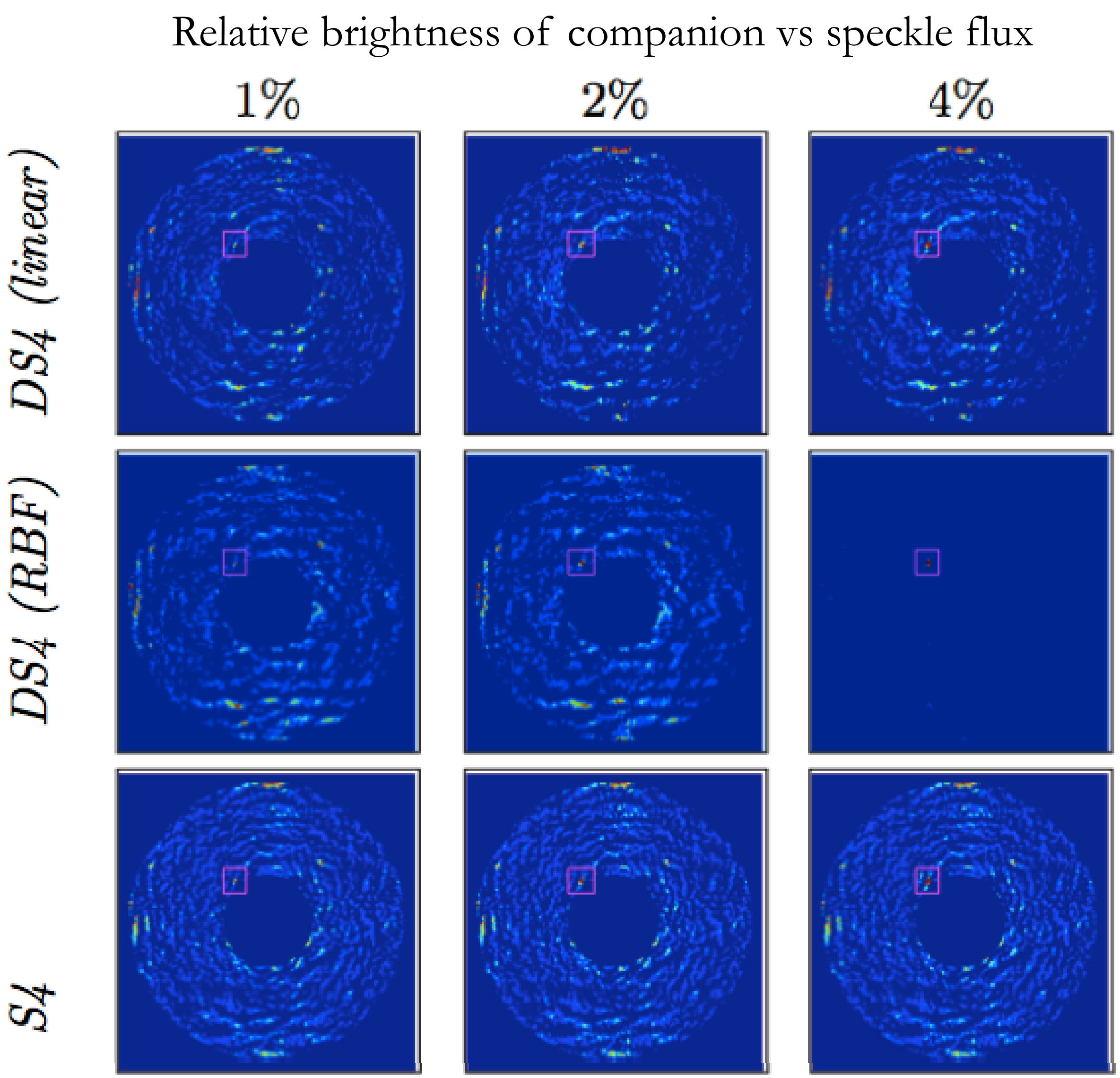


- Train Support Vector Machine (SVM)
- Use SVM on test patches to estimate p(companion|patch)

S4 Detect vs DS4 Comparison

Method	Data	Algorithm	Detection
S4	Background data (speckle)	Principle Component Analysis (Unsupervised learning)	Correlation between residual and template
DS4	Background data + artificially generated data	Support Vector Machine (Supervised learning)	Prediction value of the model

S4 Detect vs DS4 Detect

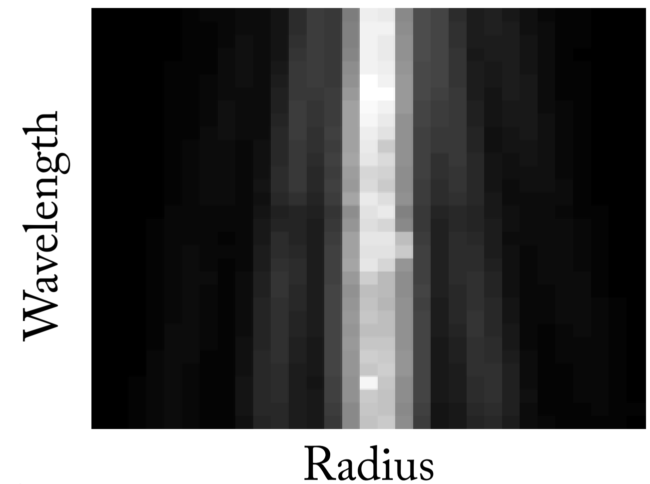


3. S4 Spectra

True Generative Model for Spectra

- S4 Detect: spectrum of planet fixed (white)

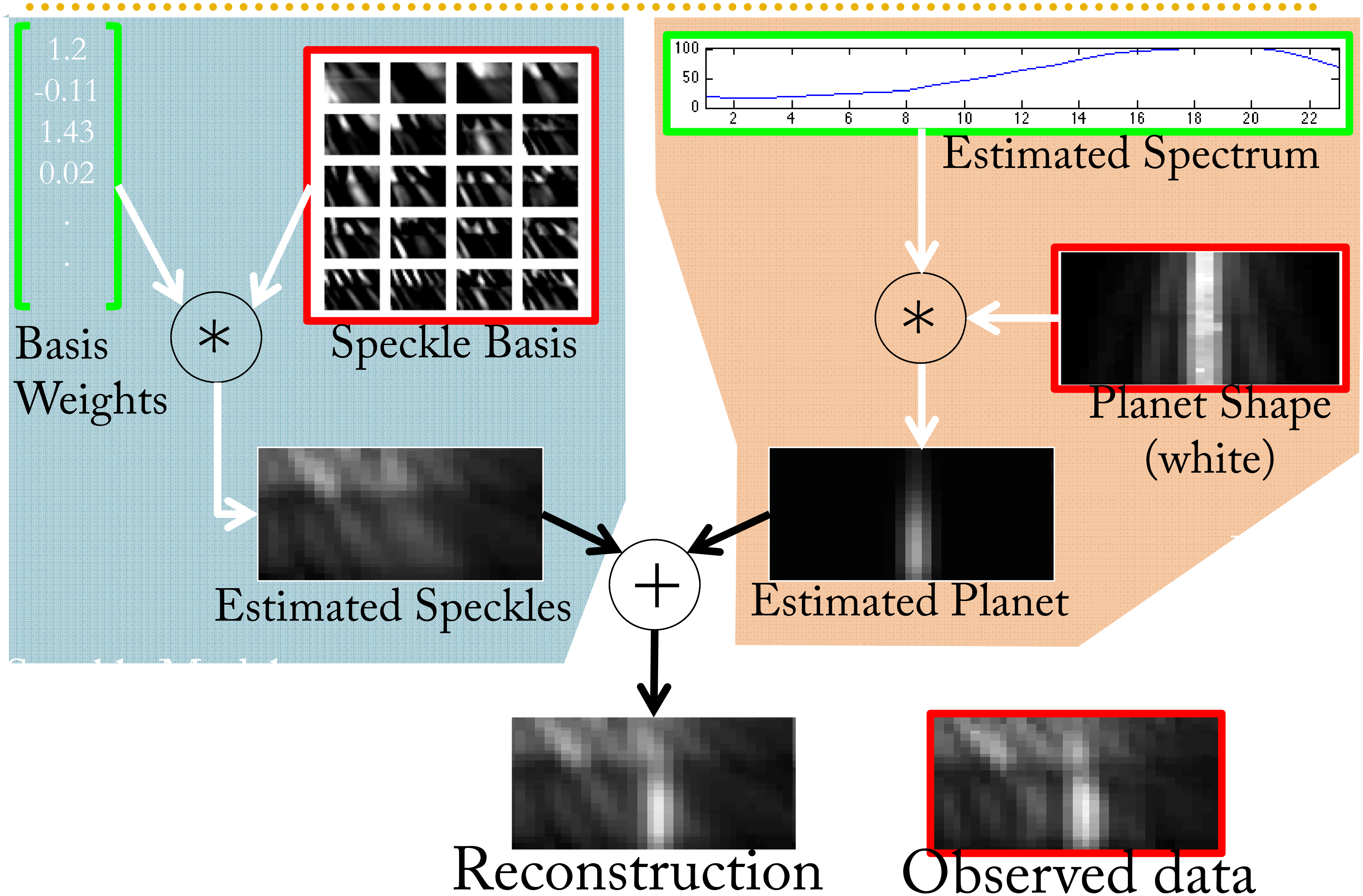
- Now spectra is unknown
  -- Treat as latent variable



- Observed data = PCA speckle model
  +
  Fixed (spatial) planet model with latent spectra

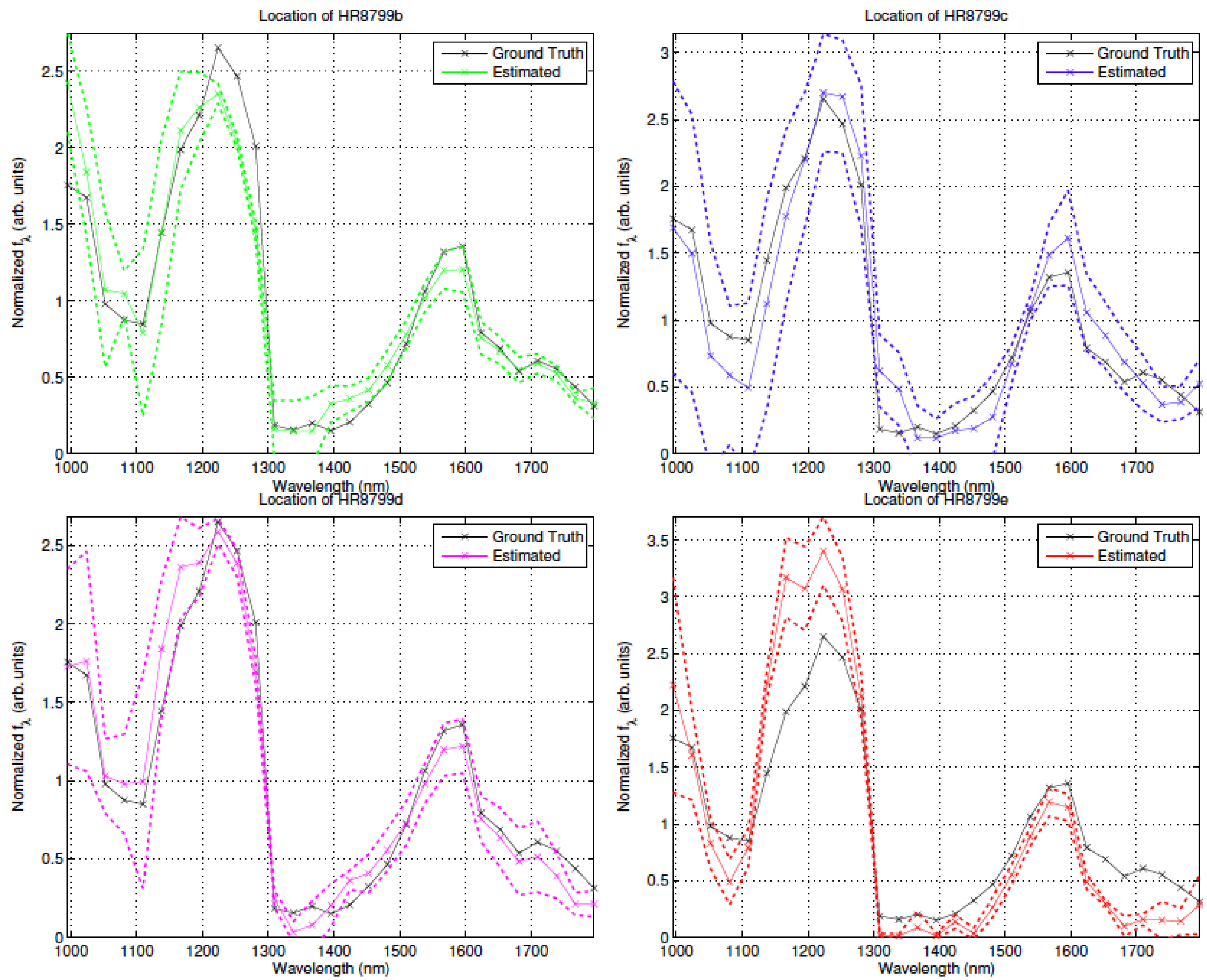
- Gaussian noise assumption

S4 Spectra Algorithm



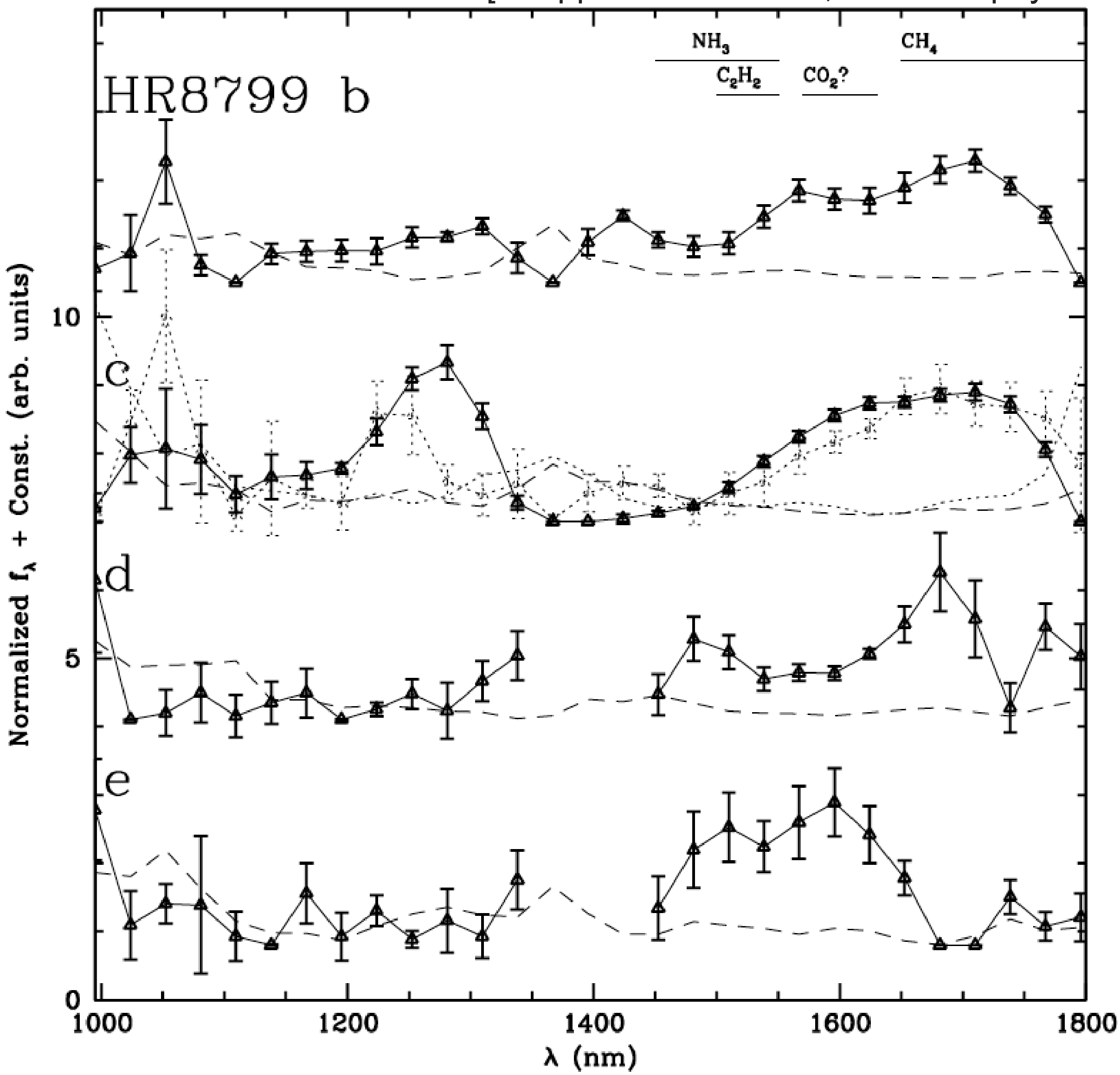
Spectra of Fake Insertions

- Insert T4.5 standard 2MASS J0559-1404 at same strength as real companions into HR8799 data



Spectra of HR8799 system

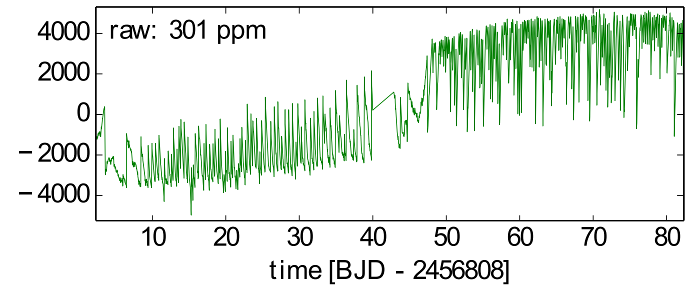
[R Oppenheimer et al., The Astrophysical Journal, April 2013.]

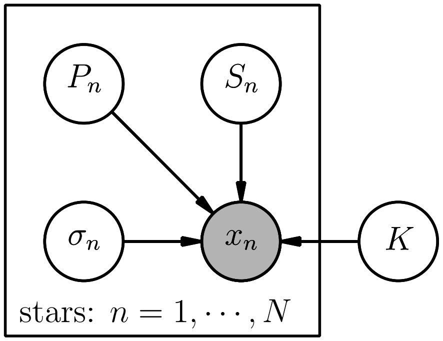


Finding Planets in Kepler 2.0 data

Foreman-Mackey, Montet, Hogg, *et al.* (arXiv:1502.04715)

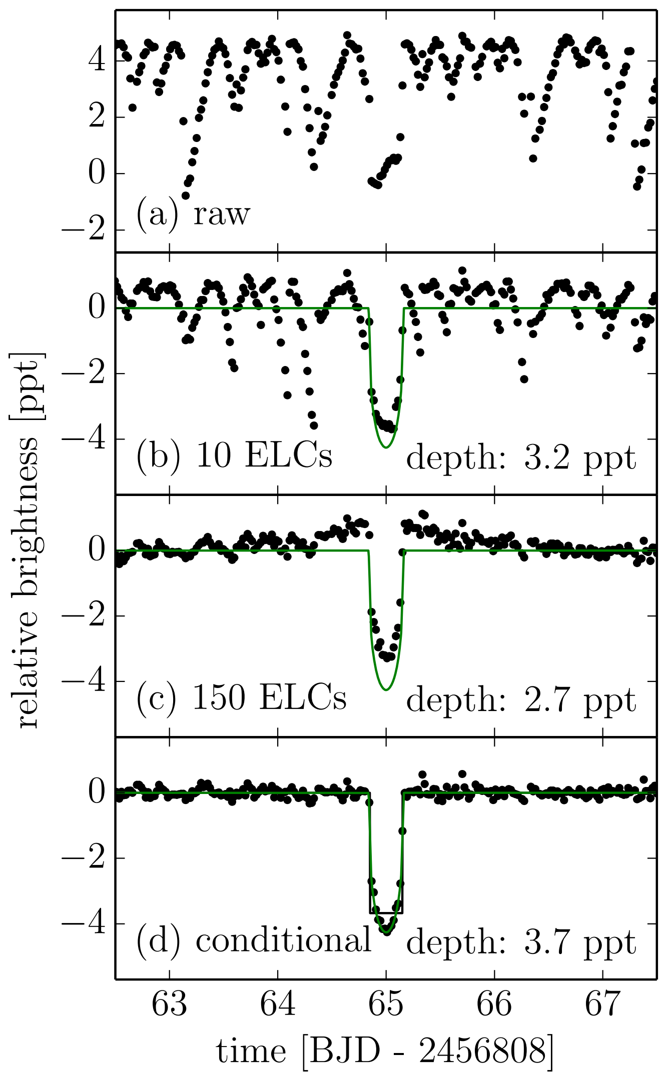
- Generative model of K2 data
- Simultaneous fit of:
  - Planet: physics & geometry
  - Star: Gaussian Process
  - CCD Noise: Poisson distribution
  - Space-craft: Data-driven linear model
- 36 plant candidates, 18 confirmed planets





Finding Planets in Kepler 2.0 data

Foreman-Mackey, Montet, Hogg, *et al.* (arXiv:1502.04715)



Raw data
\+ synthetic planet transit

PCA fit with
few components → systematics remain

PCA fit with
lots of components → systematics removed,
but transit signal attenuated

Simultaneous fit → systematics removed,
transit signal preserved.

36 plant candidates, 18 confirmed planets

Comparison

<u>Generative Models</u>

+ Labels not essential
+ Unsupervised or supervised
- • Models whole density
+ Interpretable result
- Can be hard to specify model structure

<u>Discriminative Models</u>

- **Need labels**
- Supervised only
- • Model only fits decision surface
+ Fast to evaluate
+ Can be very powerful

Final Thoughts

- Generative models feasible for many astronomy problems
  - Well understood signal formation process

- Discriminative models very powerful for other tasks where input features must be learned too

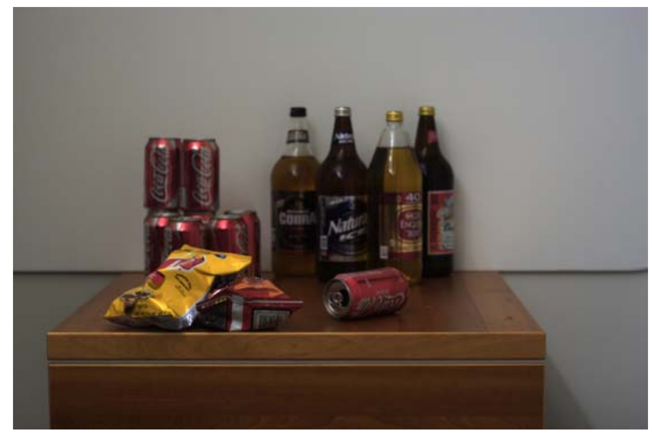
- Use machine learning to help design the coronograph itself
  - To maximize discriminability of planet vs speckles

Depth from Defocus using a Coded Aperture

[Levin, Fergus, Durand, Freeman, SIGGRAPH 2007]

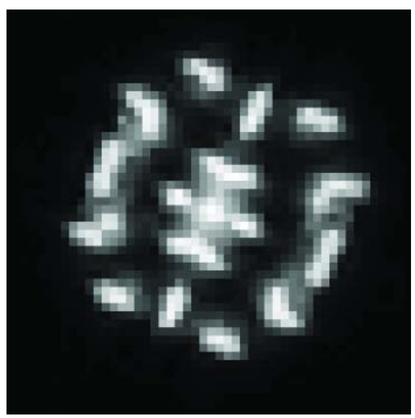
- Using generative model of natural images to design shape of aperture mask
  - Maximize discriminability between different defocus blur

Single input image (shallow D.o.F)



Modified Canon lens

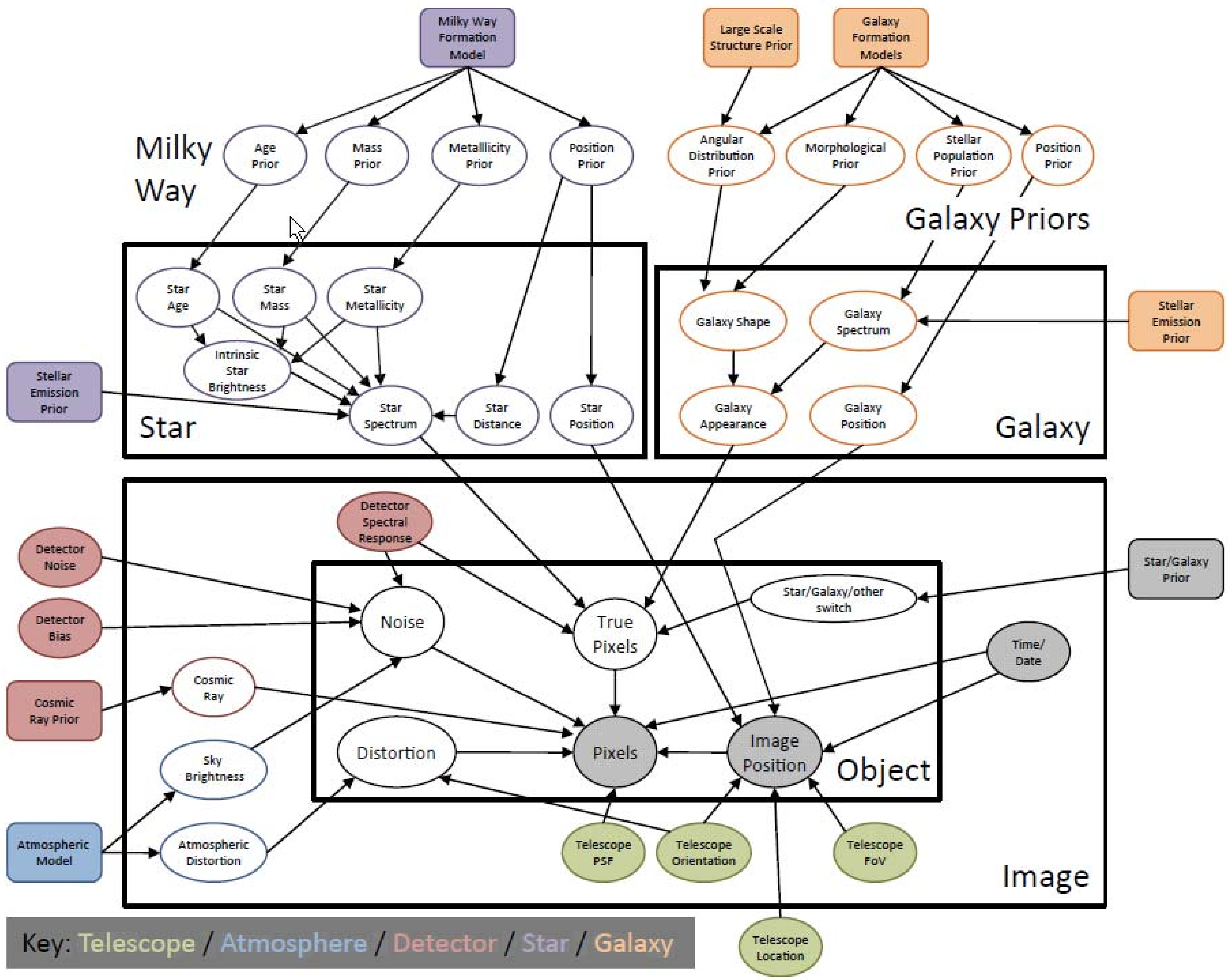
PSF



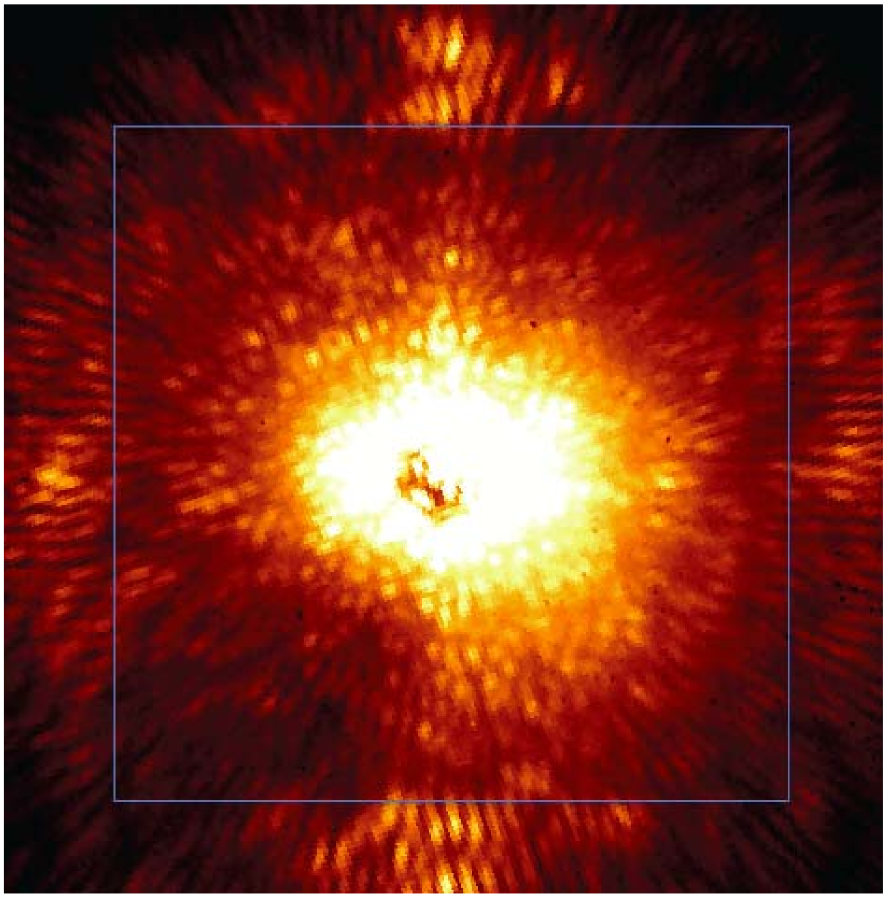
Inferred depth map

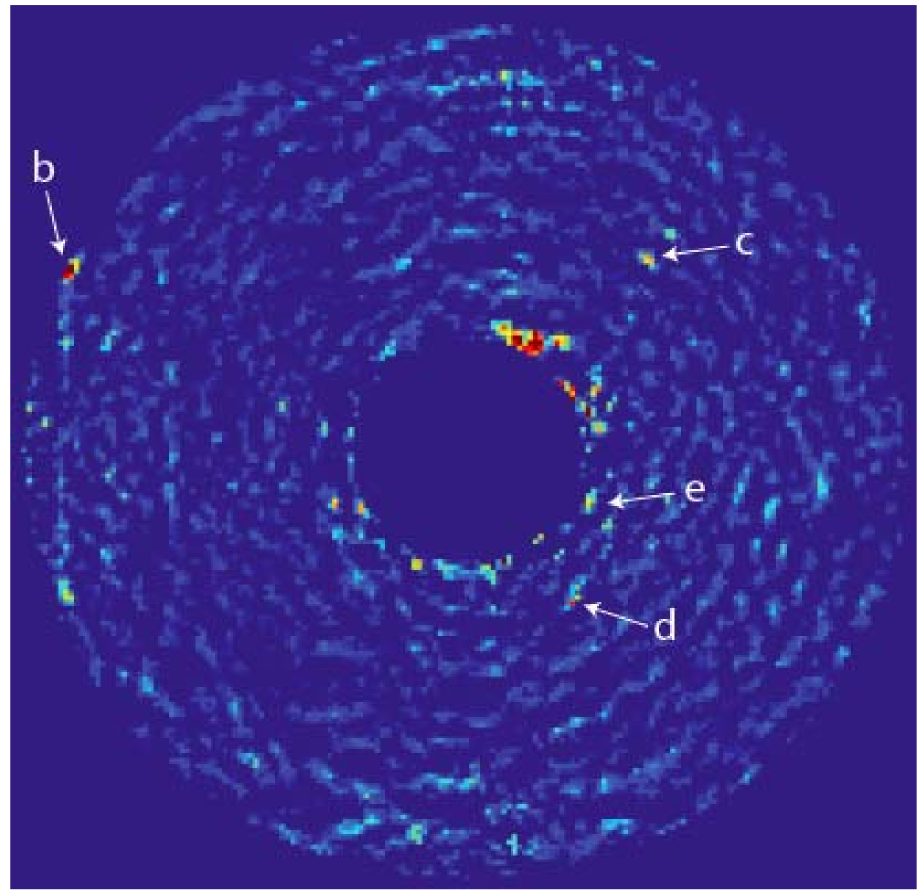
"Unified" Generative Model of Astronomical Images

- Unified Bayesian model
- Propagate uncertainty from pixels
- Physics-informed priors



Detection of Planets

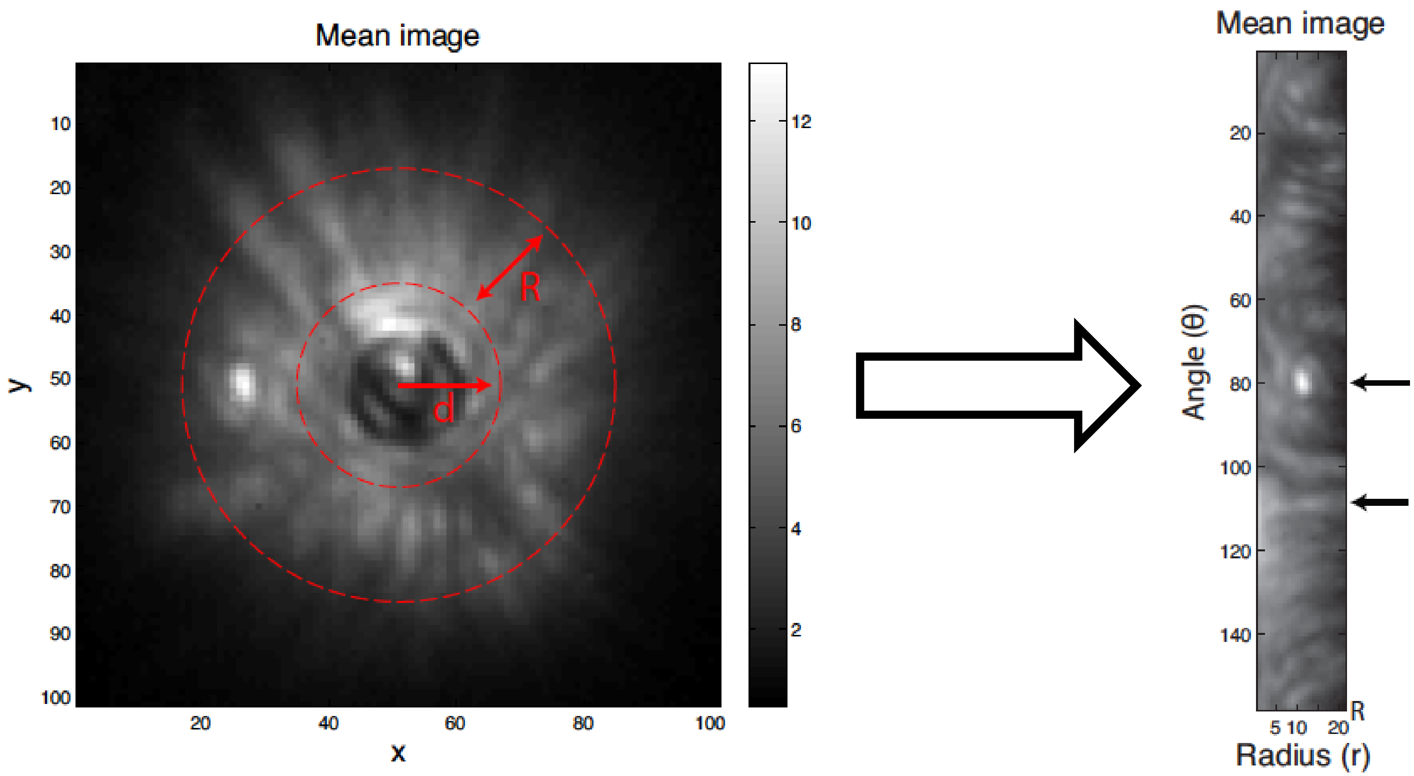




HR 8799 Input | S4 Output map

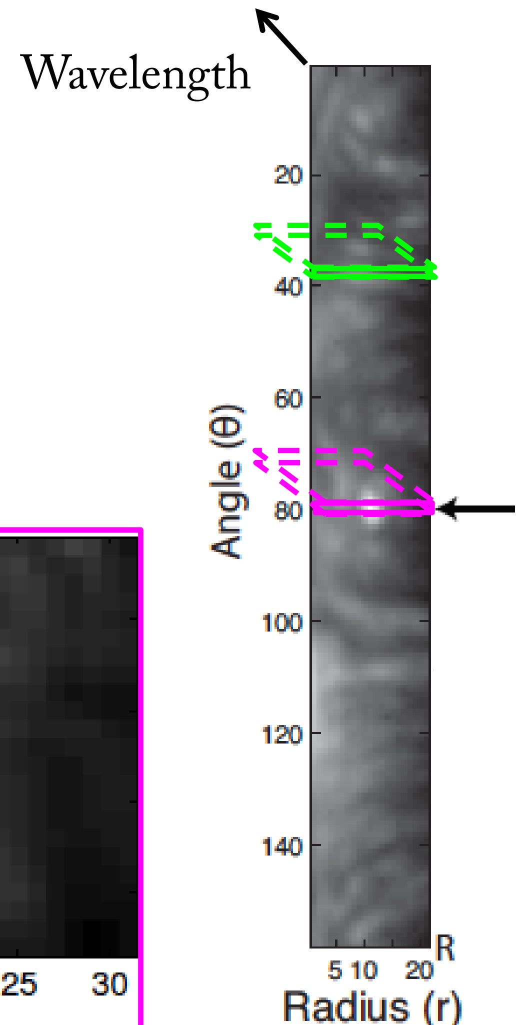
Algorithm Overview

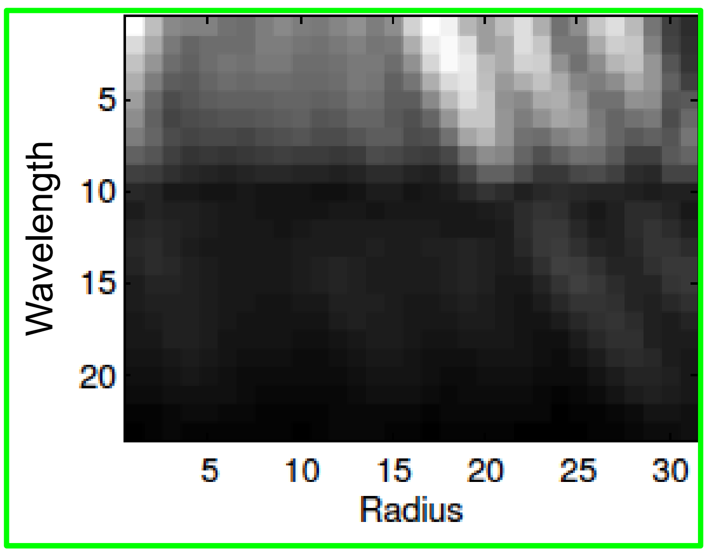
- Exploit radial motion of speckles (vs wavelength)
  - Build model in polar domain
  - Speckle motion is now 1D

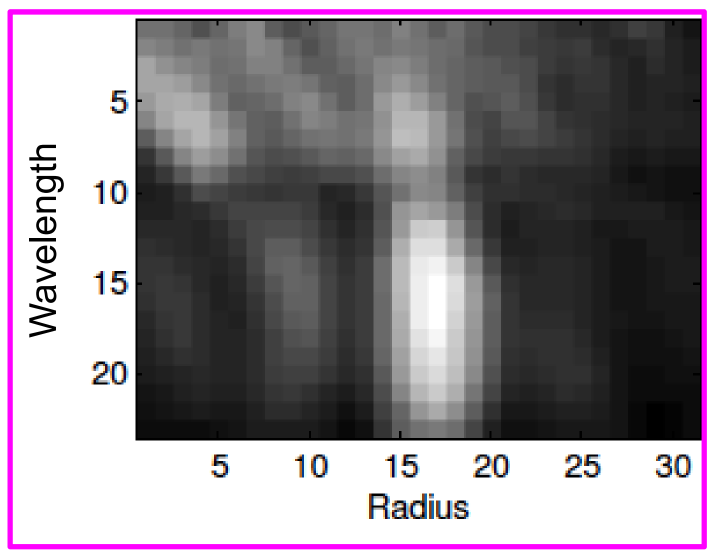


Joint Radius-Wavelength Model

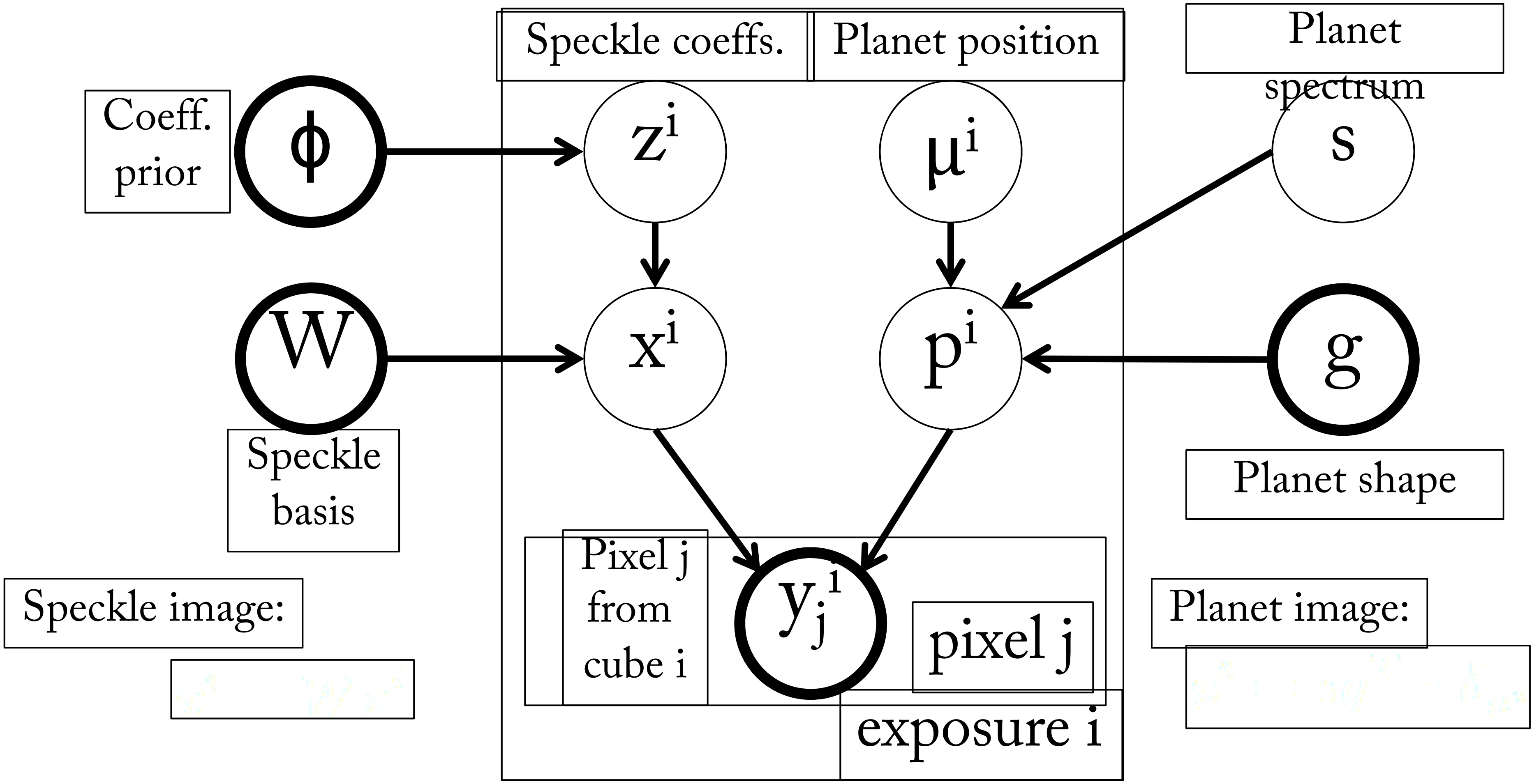
- Speckles are diagonal structures
- Planet is vertical
  - Key to separating the two
- Assume: independence to angle and exposure







S4 Graphical Model



Assume Gaussian distributions, yields overall cost:

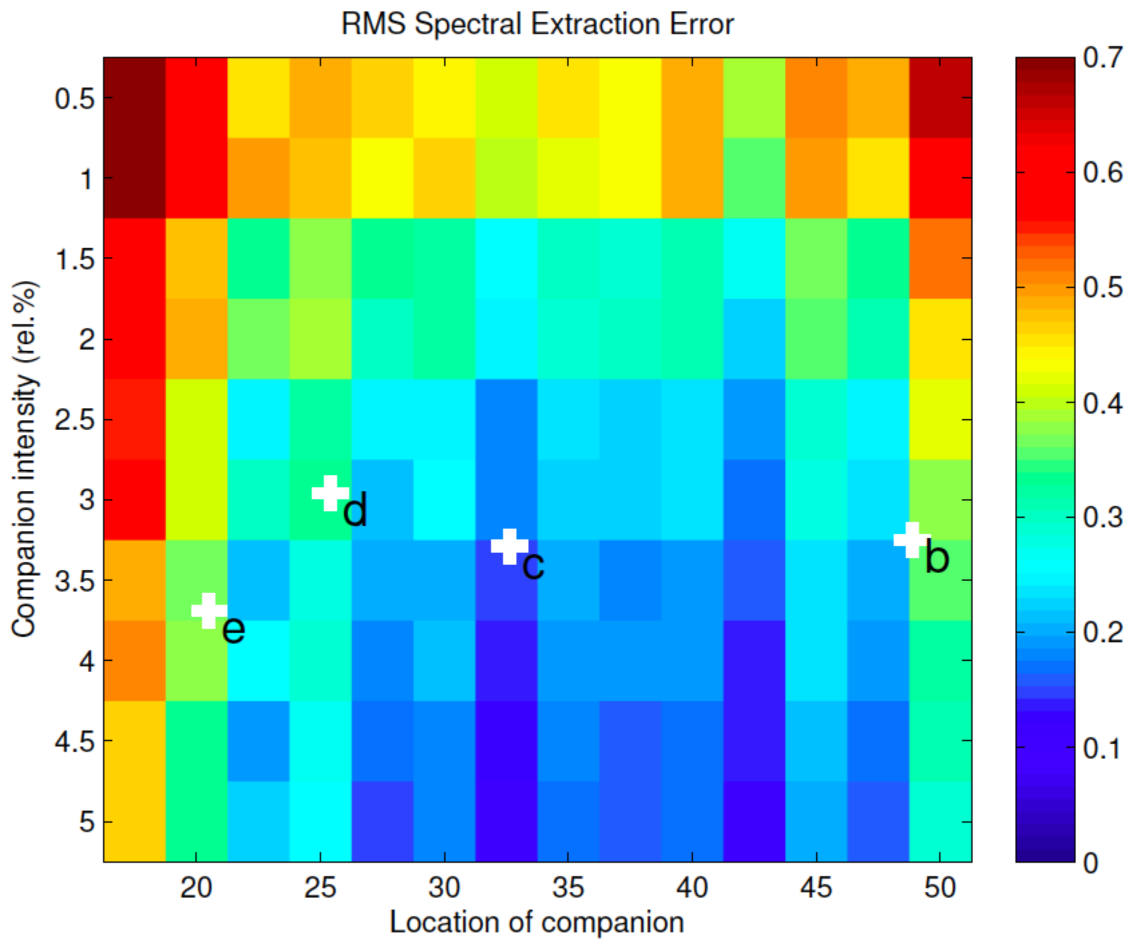
$$\sum_{ij} \|y^{ji} - (x^{ji} + p^{ji})\|^2 + \lambda \|z^{ji}\|^2$$

Approach

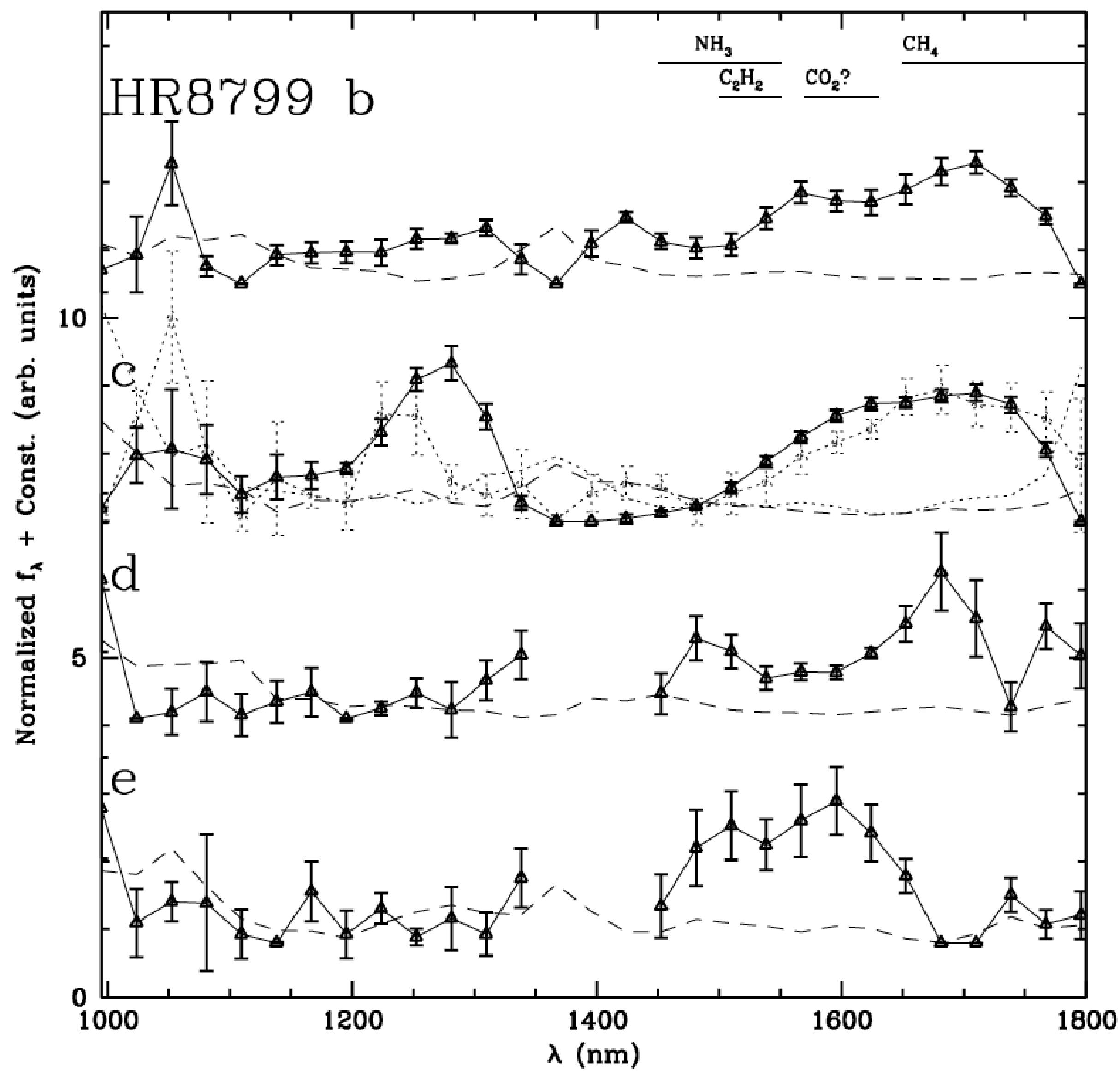
- Build statistical model of speckles
  - Physical model of optics too complex
- Few exposures of a given star (5-10)
  - Little data from which build model
- Need to exploit problem structure to yield more samples of speckles

Spectral Estimation Error

- Function of radius & companion brightness



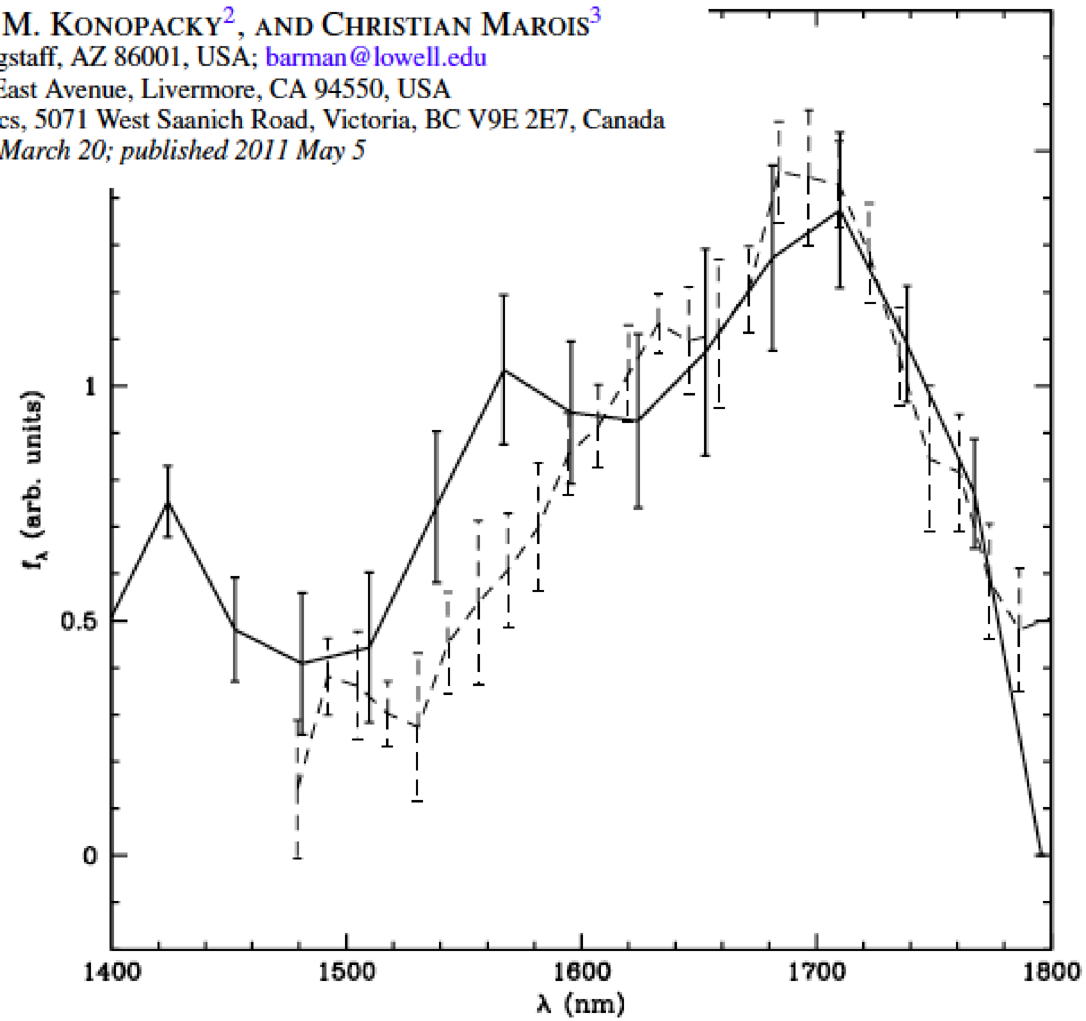
Spectra of HR8799 system



Comparison with Existing Spectrum of HR8799b

CLOUDS AND CHEMISTRY IN THE ATMOSPHERE OF EXTRASOLAR PLANET HR8799b

TRAVIS S. BARMAN[1], BRUCE MACINTOSH[2], QUINN M. KONOPACKY[2], AND CHRISTIAN MAROIS[3]
[1] Lowell Observatory, 1400 West Mars Hill Road, Flagstaff, AZ 86001, USA; barman@lowell.edu
[2] Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, CA 94550, USA
[3] National Research Council Canada, Herzberg Institute of Astrophysics, 5071 West Saanich Road, Victoria, BC V9E 2E7, Canada
Received 2011 January 26; accepted 2011 March 20; published 2011 May 5



Astronomy & Computer Vision

- Both fields concerned with images
  - Astronomy images simpler than natural scenes
  - Some hope that generative models could work

- Much work in vision on learning statistical models of natural scenes
  - Use as statistical priors for ill-posed or low S/N problems
  - Lots of ways to apply these to astronomy images

Single Image Blind Deconvolution

R. Fergus, B. Singh, A. Hertzmann, S.T. Roweis & W.T. Freeman, SIGGRAPH 2006

- Uses prior on image gradients to regularize problem

Original

Output

Close-up

Original

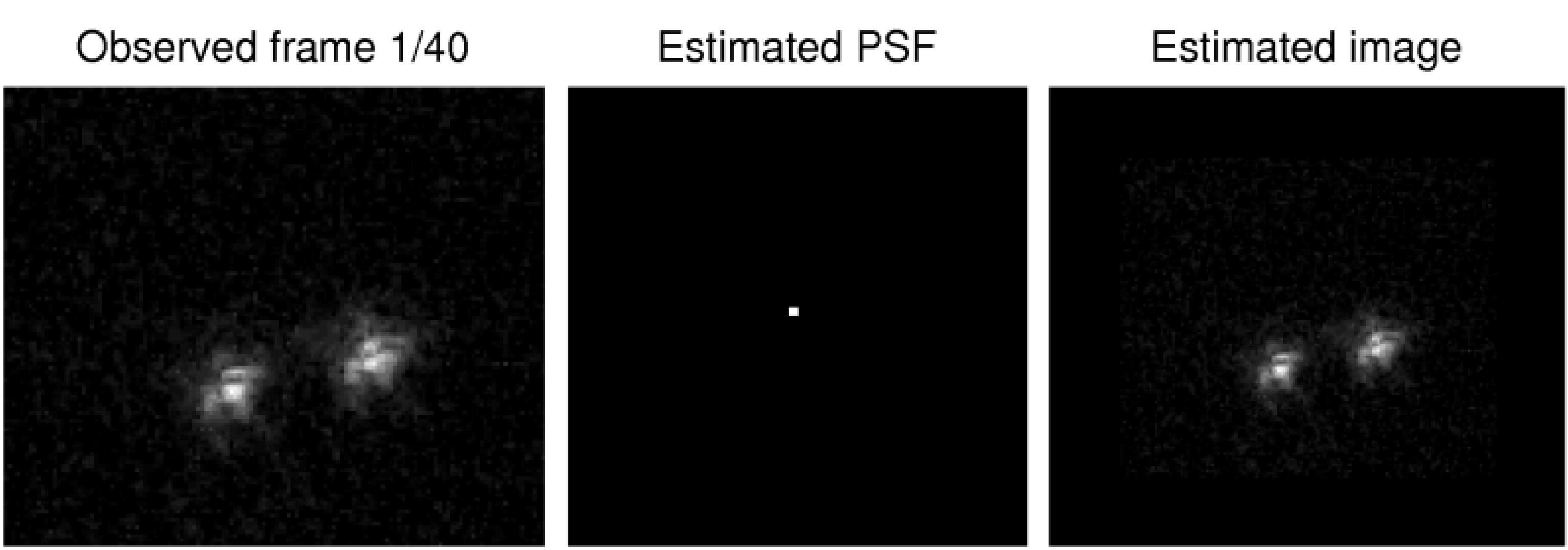
Naïve Sharpening

Our algorithm

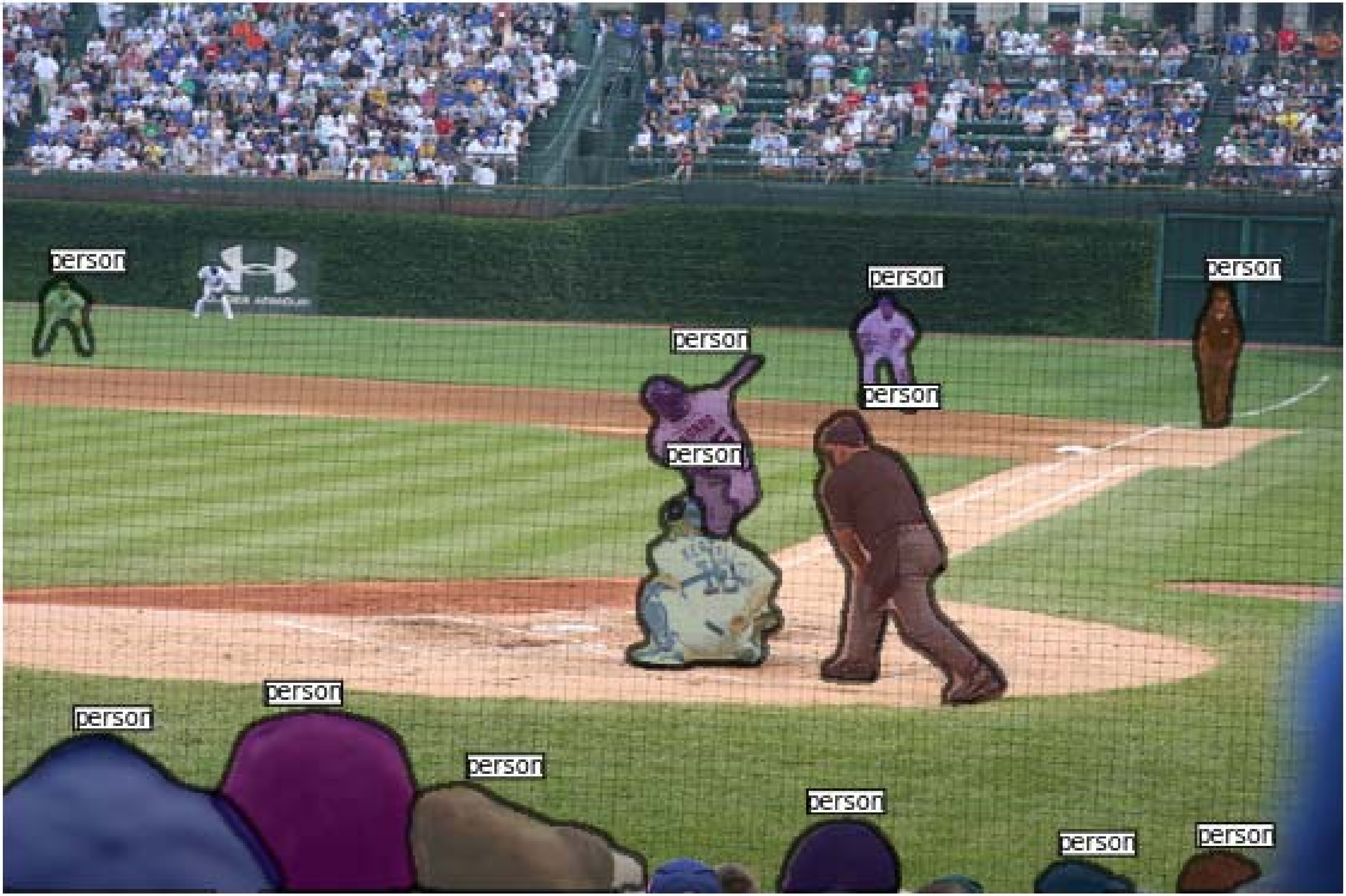
Online Blind Deconvolution

- Remove blur due to atmospheric turbulence
- Alterative to “lucky imaging” (keep best few %)

Hirsch, Harmeling, Sra & Schölkopf, Astronomy & Astrophysics 2011




person
cell phone
cell phone
laptop
laptop


person
person
person
person
person
person
person
person
person
person
person
person
person


traffic light
traffic light
car
car
person
car
car
car


sheep

Plan

- Generative vs Discriminative modeling [12 mins]
  - PCA & PPCA
  - SVMs
  - Deep Nets
- Examples of G & D modeling [10 mins]
  - Galaxy Zoo
  - Kepler DFM
- Examples of G &D modeling for direct imaging of exoplanets [20 mins]
  - S4 Detect
  - S4 Discriminative
  - S4 Spectra

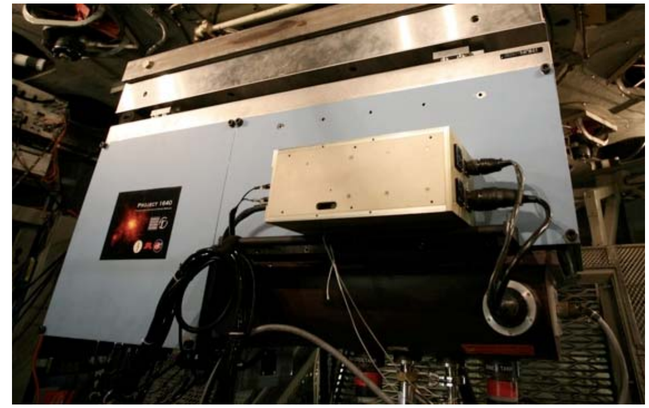
Project 1640

- Hale Telescope @ Palomar, CA
- Integral Field Spectrometer, Coronagraph, Adaptive Optics

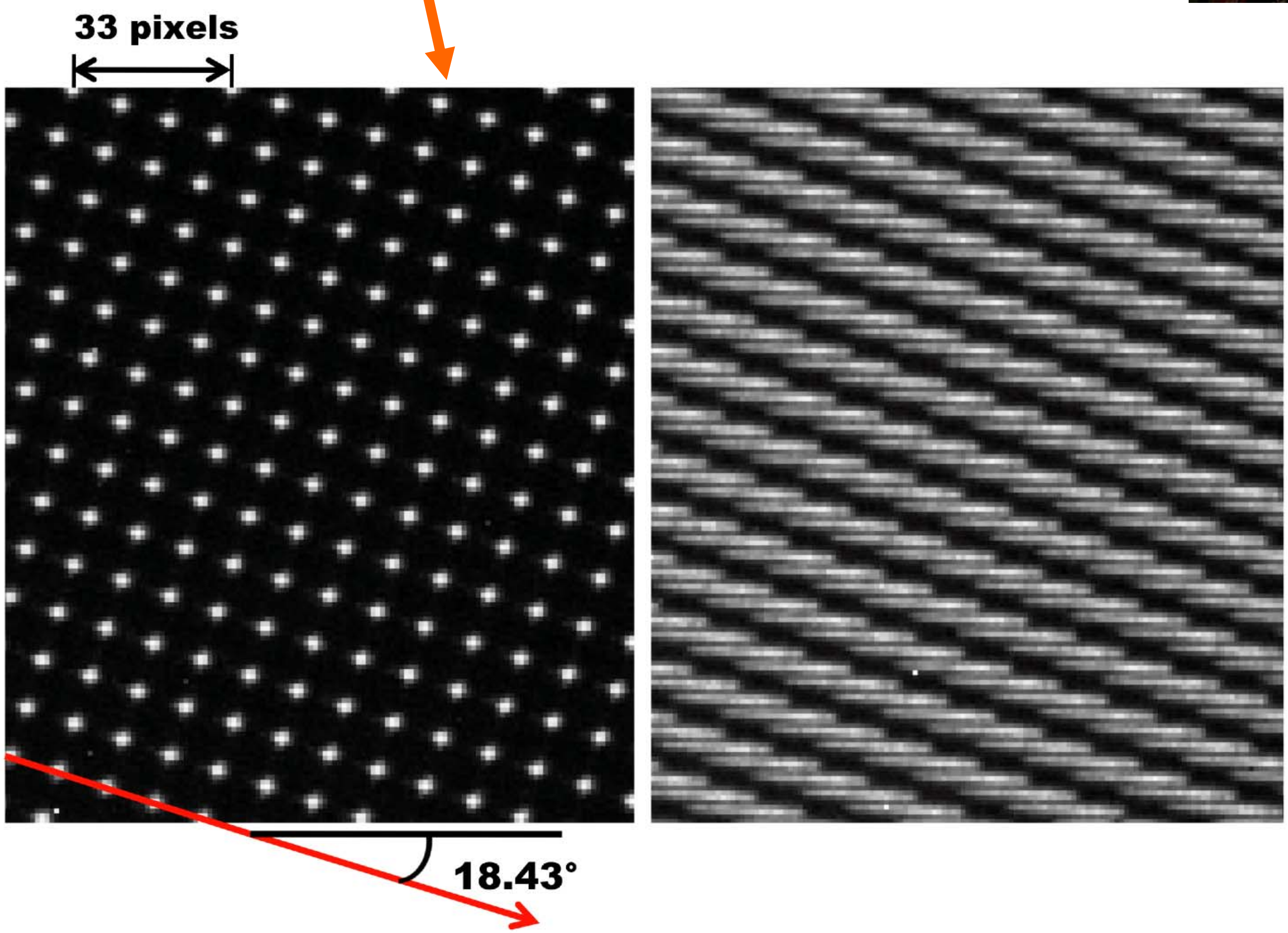
[redacted]

[Slide: R. Oppenhiemer]

Integrated Field Spectrometer



Monochromatic 1330nm light source



Broadband white light source

[Slide: R. Oppenhiemer]

Hinkley et al. 2011c (PASP, 123, 74)

Data Matrix

(#angles – held out zone) * #exposures

(~30-300) (~10)

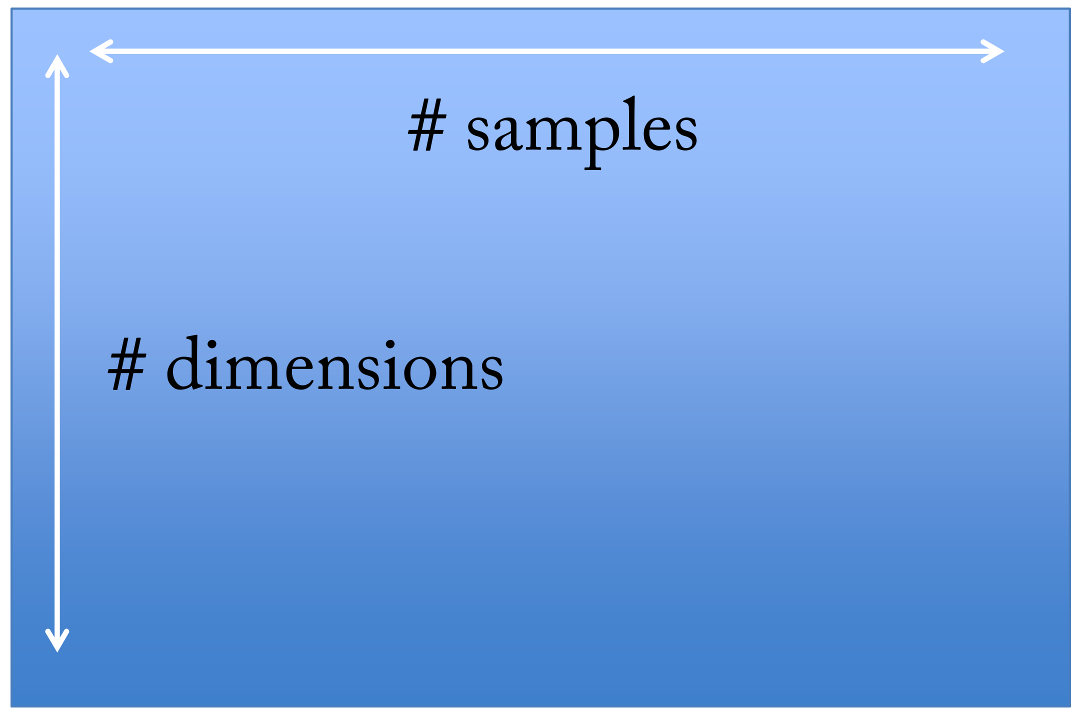
Annulus width (~20)

*

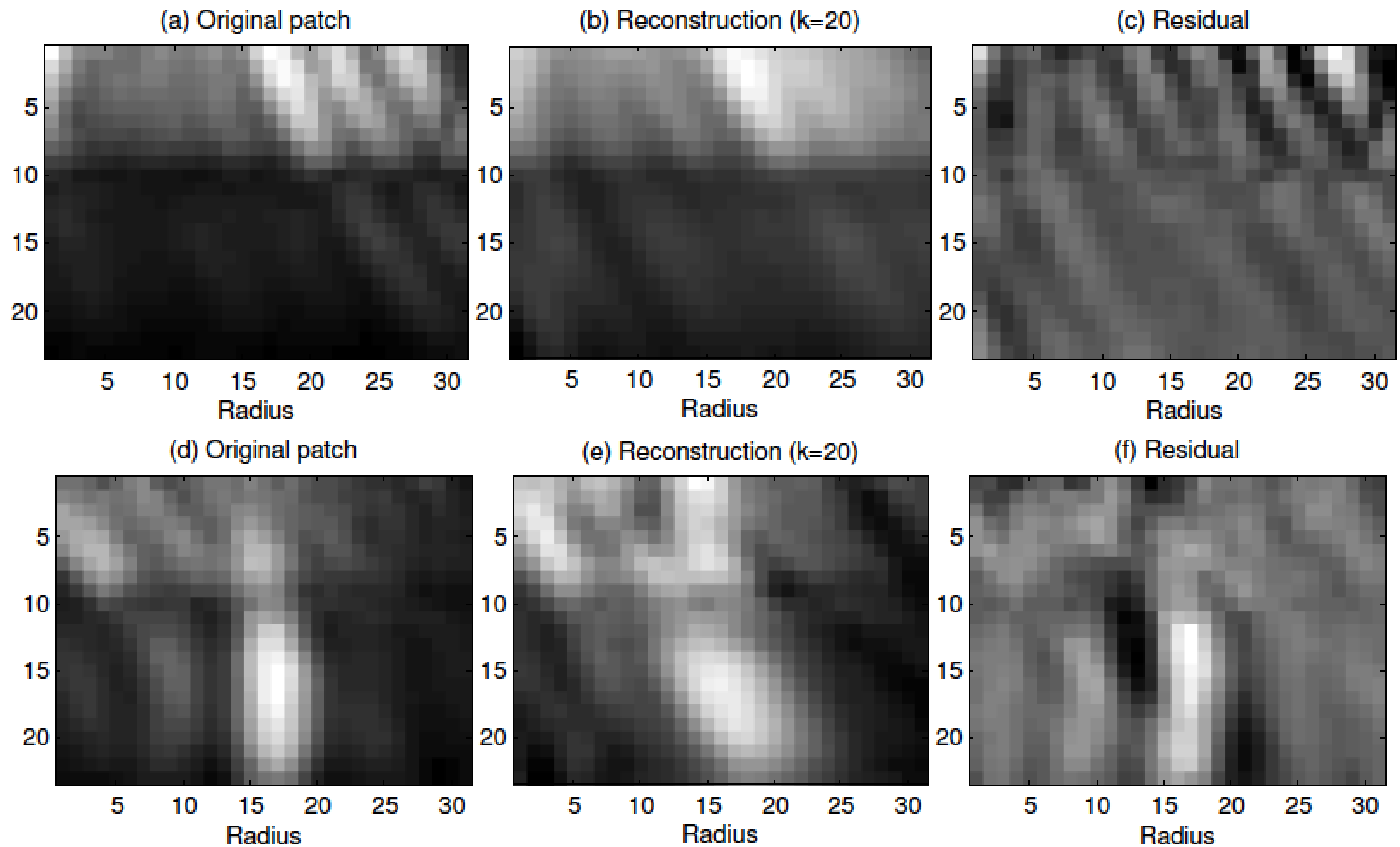
wavelengths (~30)

*

Patch width in angle (~3)

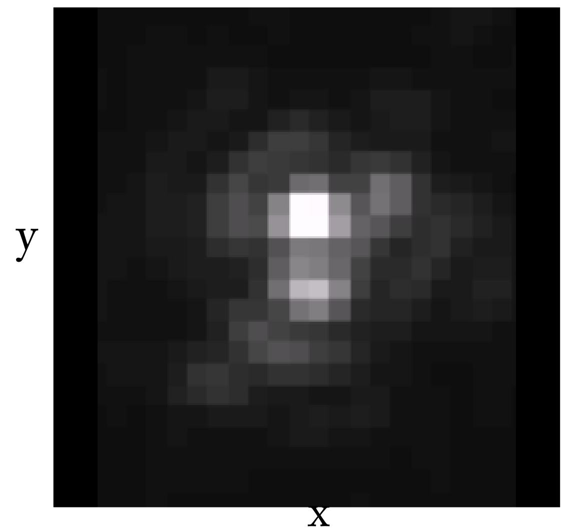


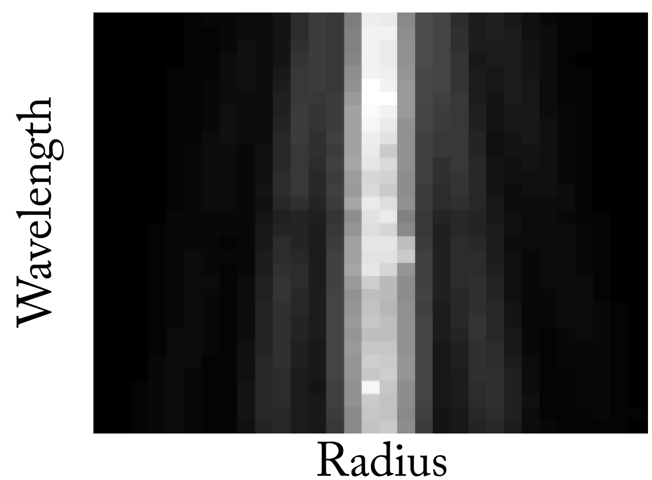
Residual Error of PCA Model



Planet Model

- Use model of planet
- Obtained from instrument calibration (spatially invariant)
- Spectra fixed: assume white



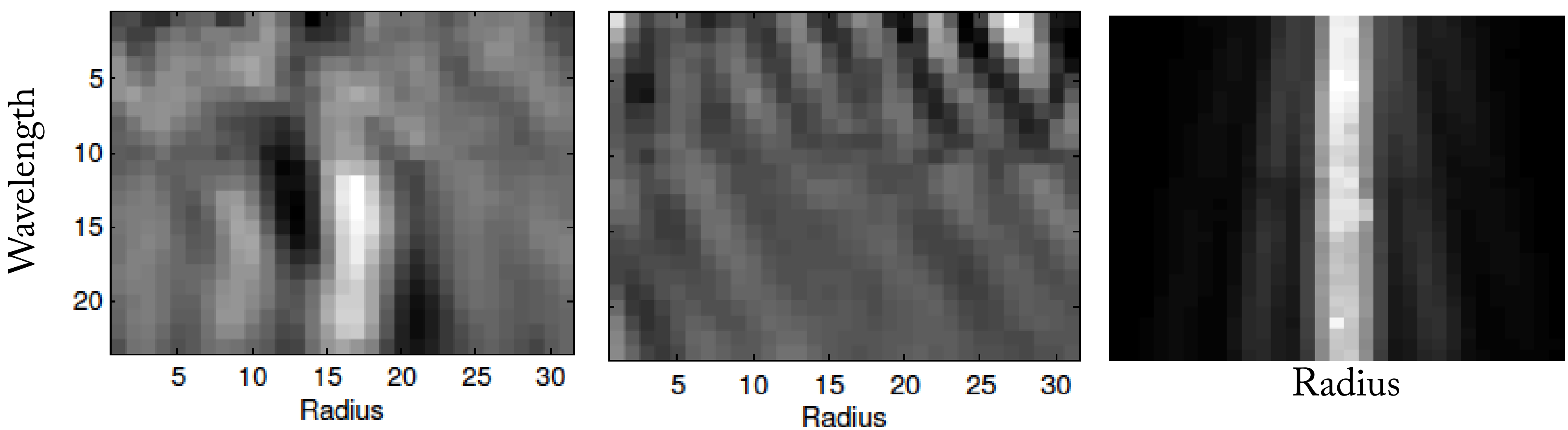


Correlation with Planet Model

- Correlation between planet model & residual error

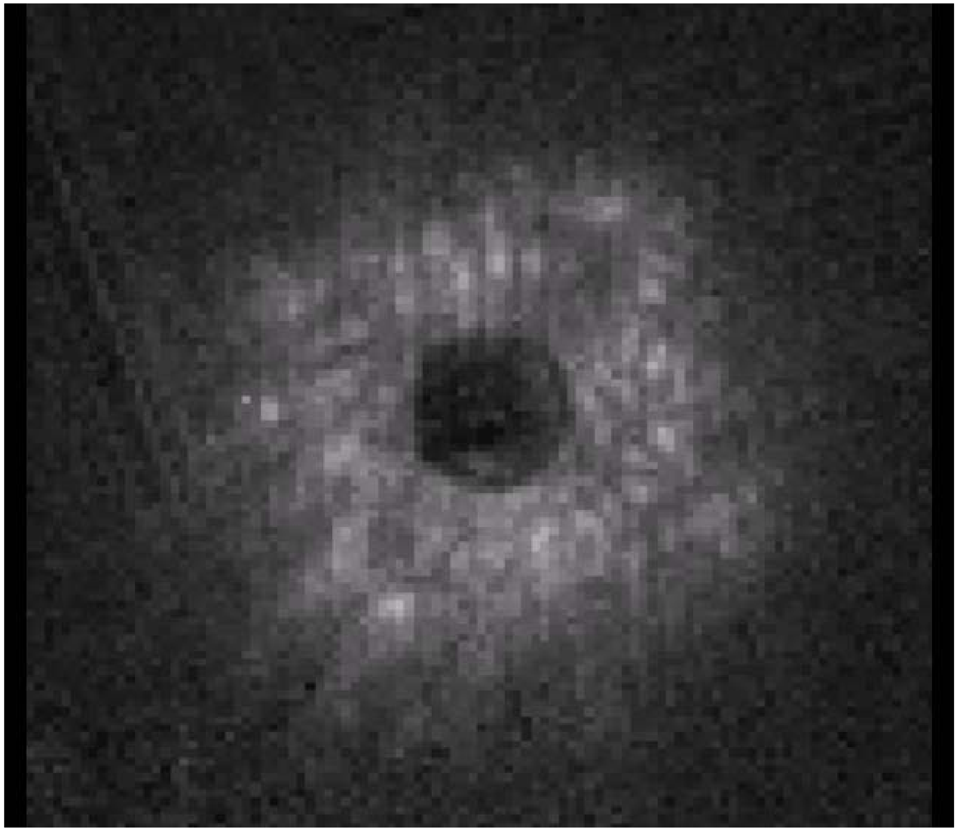
Residual (with planet)

Residual (no planet)



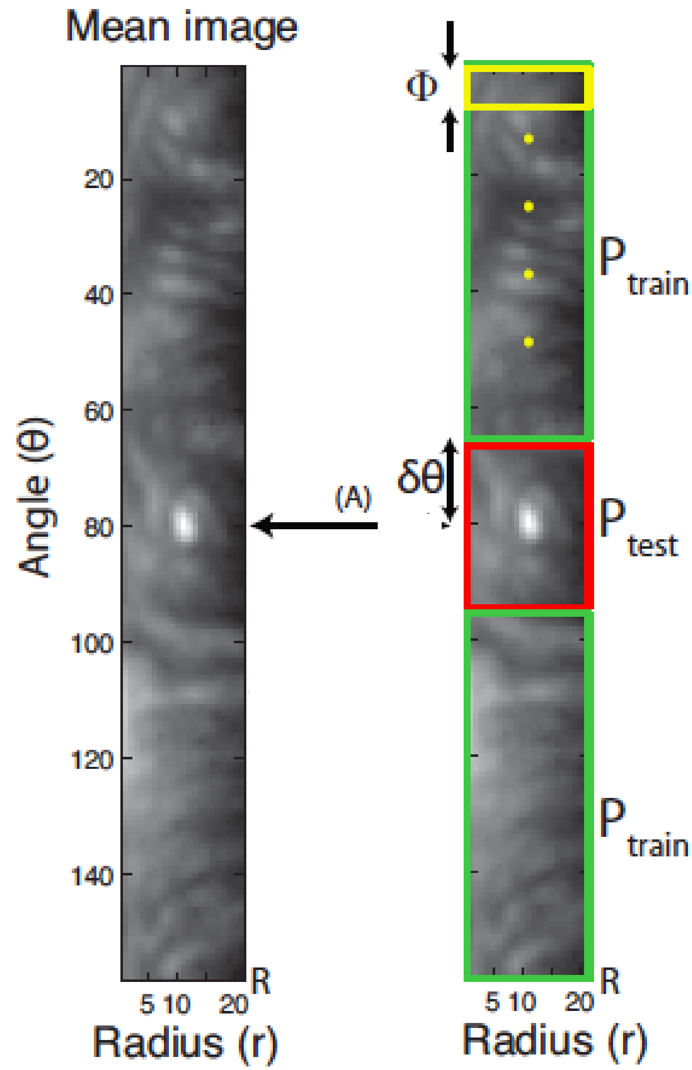
Data Cubes

- Each exposure gives 32 wavelength bands
  (near IR 950-1770nm)
- Speckles are diffraction artifacts
- Move radially with wavelength
- Planet stationary



Leave-Out Strategy

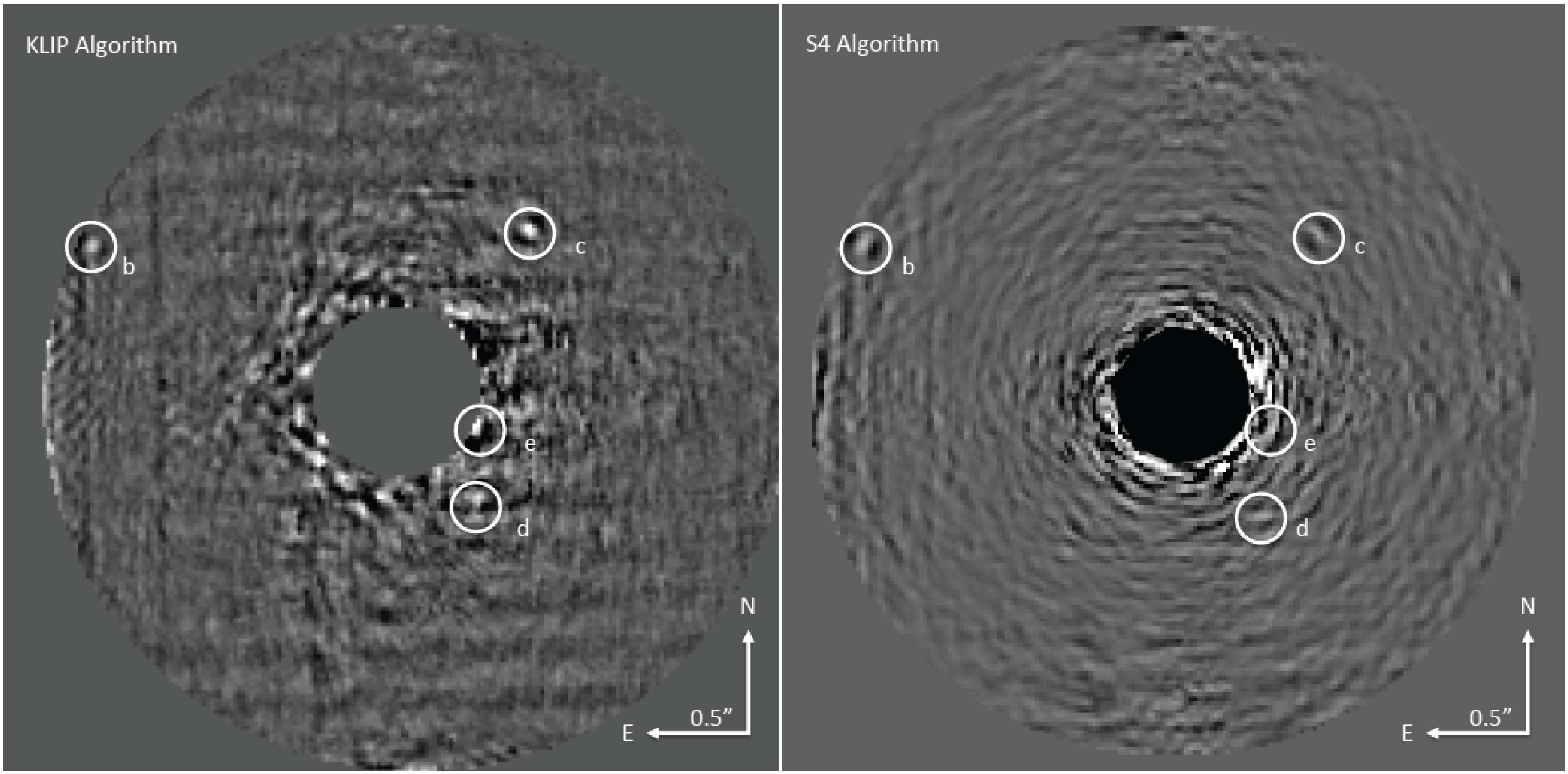
- Separate slices within annulus into train/test
- Build speckle model on train slices
  - Lots of them:
    ~ #exposures * #angle
  - Use patches with small extent in angle
- Use model to reconstruct test slices



Evaluation

- 10 exposures of star HR8799 from June 2012
- Compare to leading astronomy algorithms:
  - LOCI (Local Combination Of Images)
    Lafrenière et al. , The Astrophysical Journal, 660:770-780, May 2007
    - Models speckles as linear combination of speckles from other wavelengths/exposures
  - KLIP: Detection and Characterization of Exoplanets and Disks using Projections on Karhunen-Loeve Eigenimages, Remi Soummer et al., arXiv:1207.4197, July 2012
    - PCA-based but does not exploit radius-wavelength structure

PCA Residuals for HR8799



Spectra of HR8799 Planets