Deep Learning for Galaxies
(a progress report)

Joel Primack
UCSC

Project MAC (the Project on Mathematics and Computation was launched at MIT with a $2 million grant from the Defense Advanced Research Projects Agency (DARPA) in 1963. The "AI Group" including Marvin Minsky (the director), John McCarthy (who invented Lisp) and others. The MIT Artificial Intelligence Lab was started in 1970. Early leaders included Minsky and Seymour Papert. They were initially quite optimistic about how quickly AI would become practical.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Group
Vision Memo. No. 100.

July 7, 1966

THE SUMMER VISION PROJECT

Seymour Papert

The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

‘Simple’ problems proved most difficult.

- For decades we tried to write down every possible rule for everyday tasks —> impossible

- Every day tasks we consider blindingly obvious have been exceedingly difficult for computers.

cat?

from “Deep Learning for Vision” lecture at CERN by Jon Shlens, Google Research - author of TensorFlow

Machine learning applied everywhere.

- The last decade has shown that if we teach computers to perform a task, they can perform exceedingly better.

machine translation	speech recognition
face recognition	time series analysis
molecular activity prediction	image recognition
road hazard detection	object detection
optical character recognition	motor planning
motor activity planning	syntax parsing
language understanding	…

face recognition for galaxies?

Examples of artificial vision in action

- fine-grain classification

hibiscus

dahila

- generalization

meal

meal

- sensible errors

snake

dog

*** Trained a model for whole image recognition using Inception-v3 architecture.*

A toy model of a neuron: “perceptron”

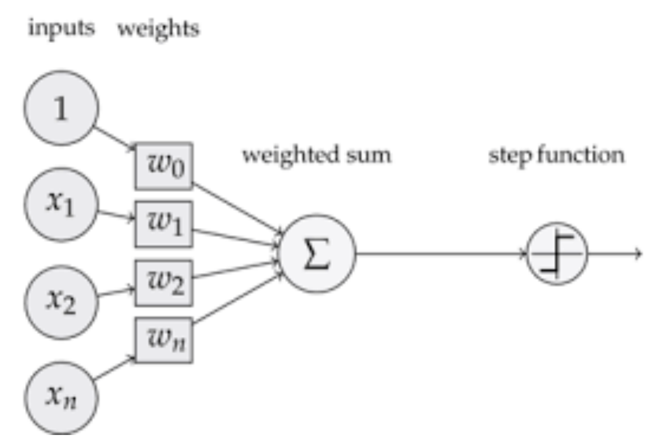
Frank Rosenblatt
1928-1971
vs.

Marvin Minsky
1927-2016

Simplify the neuron to a sum over weighted inputs and a nonlinear activation function.

- no spikes
- no recurrence or feedback *
- no dynamics or state *
- no biophysics

$$y = f(\sum_i w_i x_i + b)$$



$$f(z) = max(0, z)$$

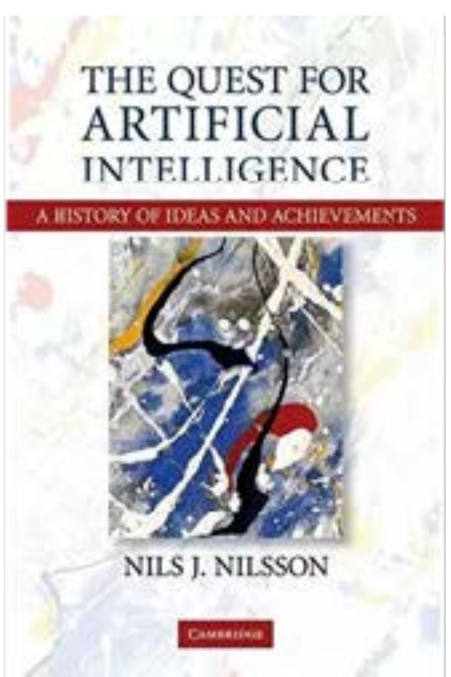
The perceptron: a probabilistic model for information storage and organization in the brain.
F Rosenblatt (1958)

“During the late 1950s and early 1960s … Rosenblatt and Minsky debated on the floors of scientific conferences the value of biologically inspired computation, Rosenblatt arguing that his neural networks could do almost anything and Minsky countering that they could do little.”

Web version of *The Quest for Artificial Intelligence* by Nils Nilsson, nicely covers Minsky and Rosenblatt (as well as a lot of other relevant AI material).

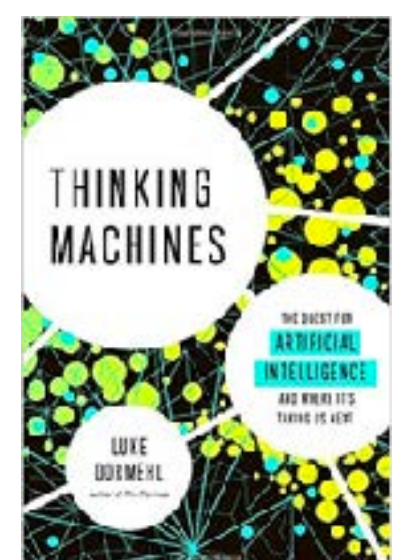
During the 1960s, neural net researchers employed various methods for changing a network's adjustable weights so that the entire network made appropriate output responses to a set of "training" inputs. For example, **Frank Rosenblatt** at Cornell adjusted weight values in the final layer of what he called the three-layer alpha-perceptron. But what stymied us all was how to change weights in more than one layer of multilayer networks…. Inventive schemes were tried for making weight changes; none seemed to work out.

That problem was solved in the mid-1980s by the invention of a technique called "back propagation" (**backprop** for short) introduced by David Rumelhart, **Geoffrey E. Hinton**, and Ronald J. Williams. The basic idea behind backprop is simple…. In response to an error in the network's output, backprop makes small adjustments in all of the weights so as to reduce that error. It can be regarded as a hill-descending method – searching for low values of error over the landscape of weights. But rather than actually trying out all possible small weight changes and deciding on that set of them that corresponds to the steepest descent downhill, backprop uses calculus to precompute the best set of weight changes.



From *The Quest for Artificial Intelligence* by Nils Nilsson, Chapter 29.

See also

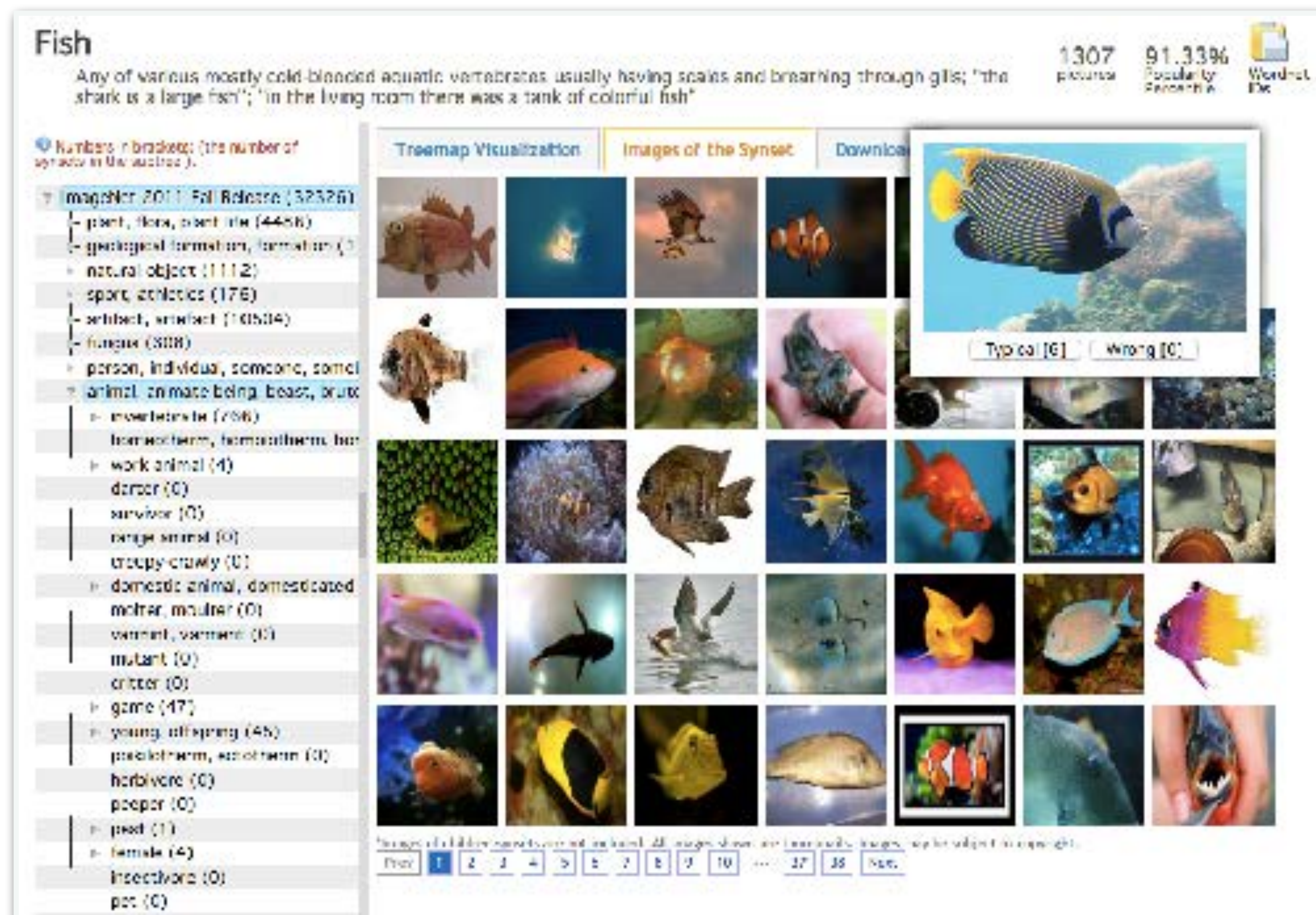


The computer vision competition: IMAGENET

Large scale academic competition focused on predicting 1000 object classes (~1.2M images).

classes

- `electric ray`
- `barracuda`
- `coho salmon`
- `tench`
- `goldfish`
- `sawfish`
- `smalltooth sawfish`
- `guitarfish`
- `stingray`
- `roughtail stingray`
- `...`



Imagenet: A large-scale hierarchical image database
J Deng et al (2009)

History of techniques in ImageNet Challenge

ImageNet 2010

Locality constrained linear coding + SVM	NEC & UIUC
Fisher kernel + SVM	Xerox Research Center Europe
SIFT features + LI2C	Nanyang Technological Institute
SIFT features + k-Nearest Neighbors	Laboratoire d'Informatique de Grenoble
Color features + canonical correlation analysis	National Institute of Informatics, Tokyo

ImageNet 2011

Compressed Fisher kernel + SVM	Xerox Research Center Europe
SIFT bag-of-words + VQ + SVM	University of Amsterdam & University of
SIFT + ?	ISI Lab, Tokyo University

ImageNet 2012

Deep convolutional neural network	University of Toronto
Discriminatively trained DPMs	University of Oxford
Fisher-based SIFT features + SVM	ISI Lab, Tokyo University

ImageNet Classification with Deep Convolutional Neural Networks

Alex Krizhevsky
University of Toronto
kriz@cs.utoronto.ca

Ilya Sutskever
University of Toronto
ilya@cs.utoronto.ca

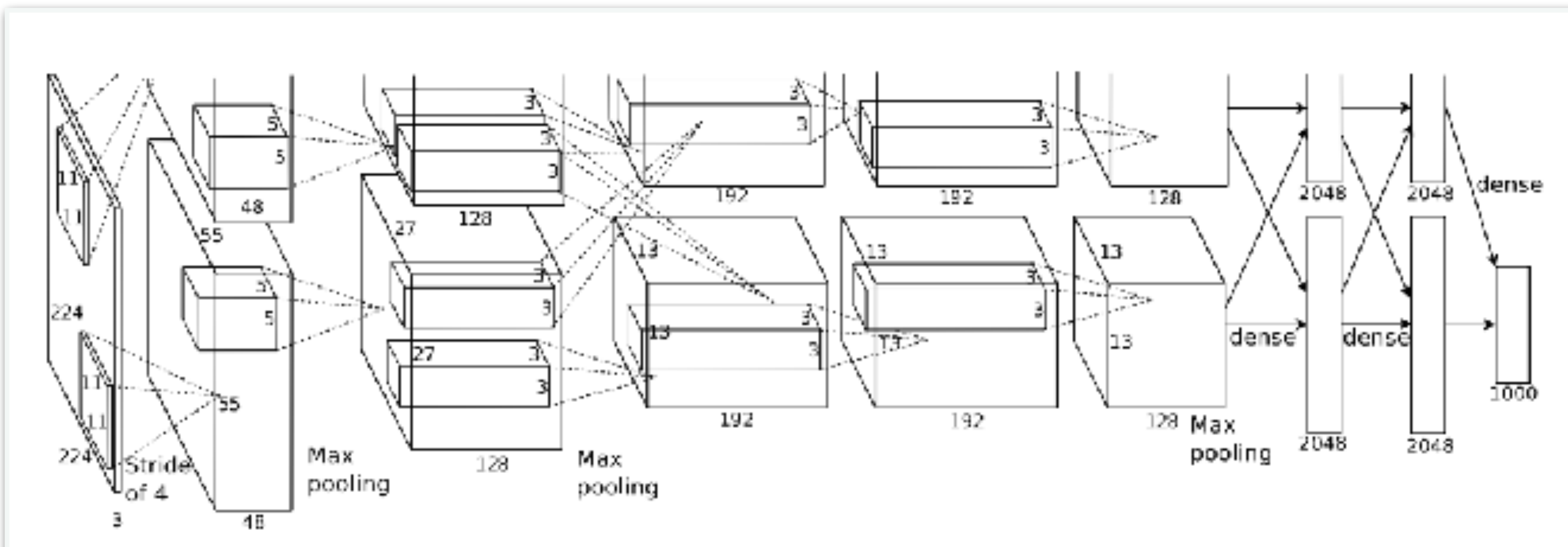
Geoffrey E. Hinton
University of Toronto
hinton@cs.utoronto.ca

Advances in Neural Information Processing Systems 25 (NIPS 2012) [PDF]

Abstract

We trained a large, deep convolutional neural network to classify the 1.2 million high-resolution images in the ImageNet LSVRC-2010 contest into the 1000 different classes. On the test data, we achieved top-1 and top-5 error rates of 37.5% and 17.0% which is considerably better than the previous state-of-the-art. The neural network, which has 60 million parameters and 650,000 neurons, consists of five convolutional layers, some of which are followed by max-pooling layers, and three fully-connected layers with a final 1000-way softmax. To make training faster, we used non-saturating neurons and a very efficient GPU implementation of the convolution operation. To reduce overfitting in the fully-connected layers we employed a recently-developed regularization method called "dropout" that proved to be very effective. We also entered a variant of this model in the ILSVRC-2012 competition and achieved a winning top-5 test error rate of 15.3%, compared to 26.2% achieved by the second-best entry.

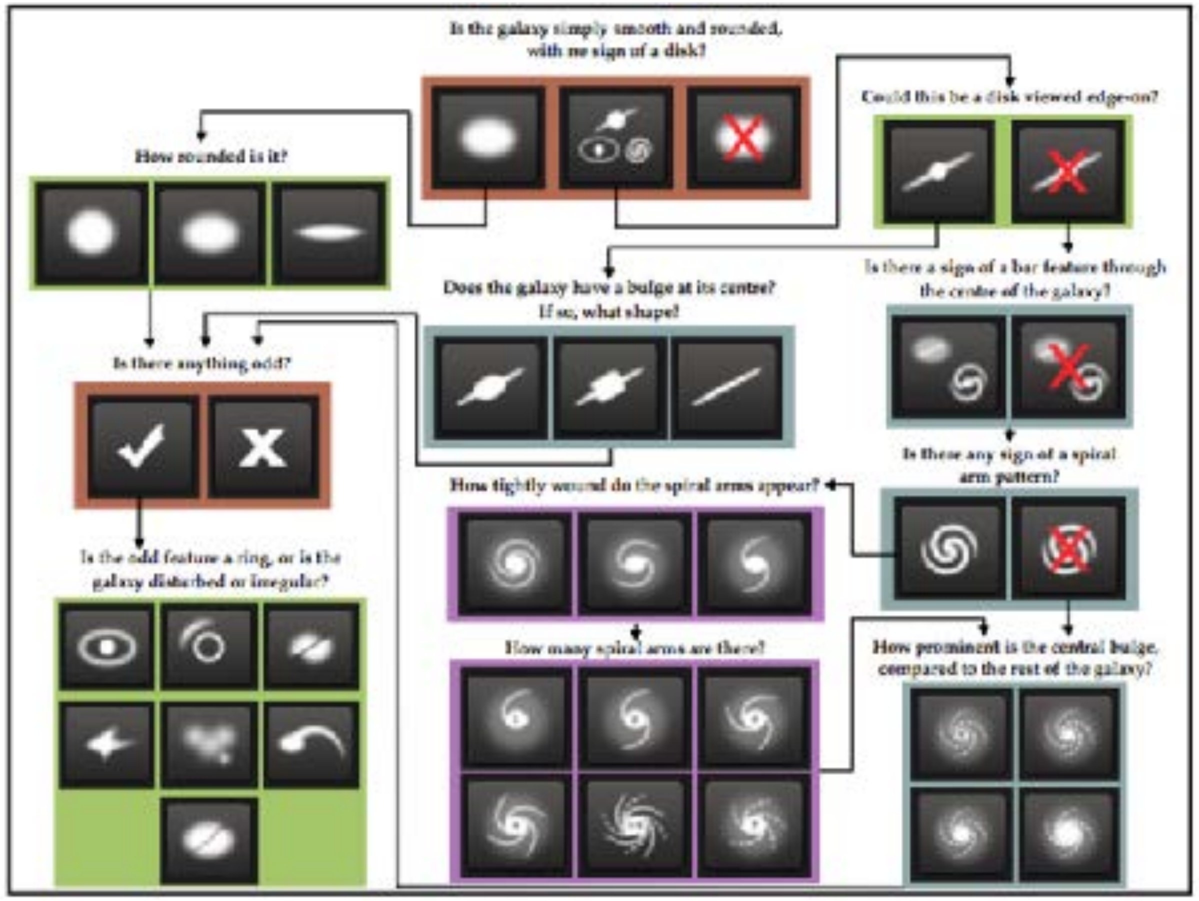
Deep convolutional neural networks



ImageNet Classification with Deep Convolutional Neural Networks
A Krizhevsky I Sutskever, G Hinton (2012)

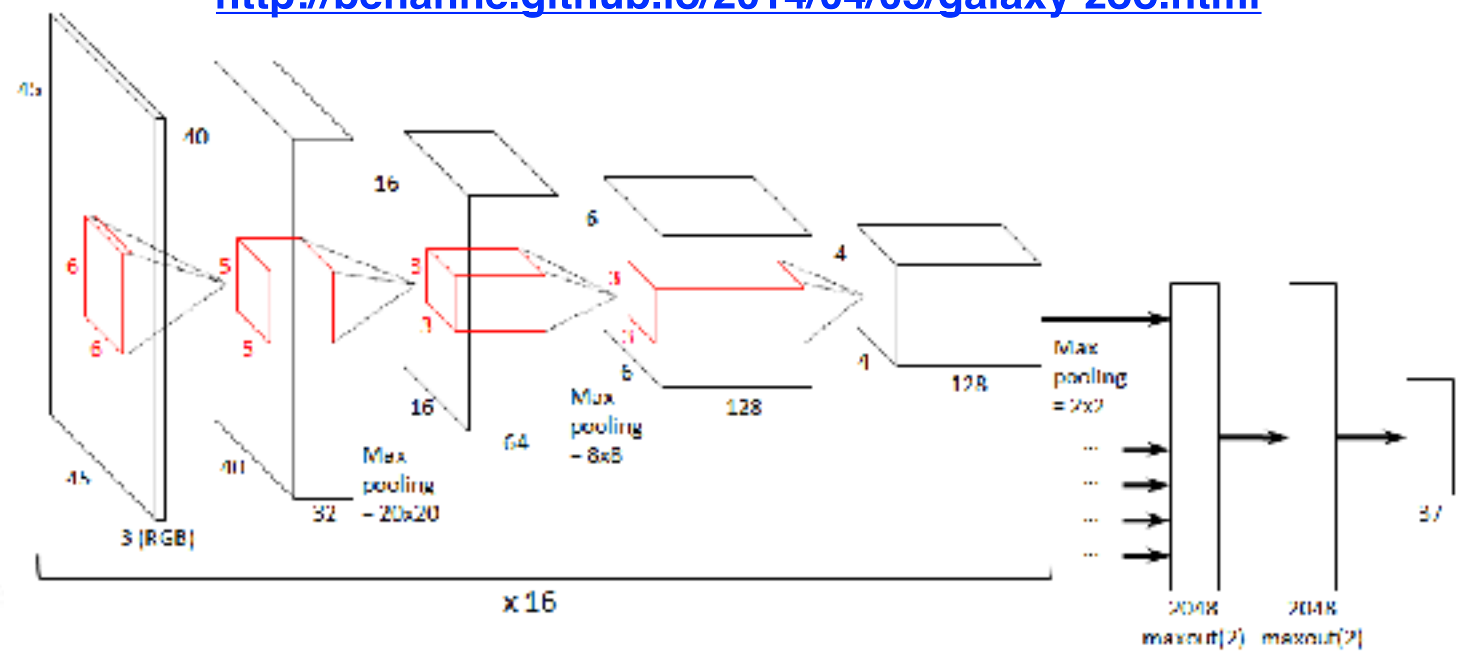
- Multi-layer perceptron trained with back-propagation are ideas known since the 1980's.
- The success of deep learning in the past 5 years is due to more powerful computers (GPUs) and better code.

Sander Dieleman used a deep learning code to predict Galaxy Zoo nearby galaxy image classifications with high accuracy, winning the 2014 Kaggle competition



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

http://benanne.github.io/2014/04/05/galaxy-zoo.html



Krizhevsky-style diagram of the architecture of the best performing network.

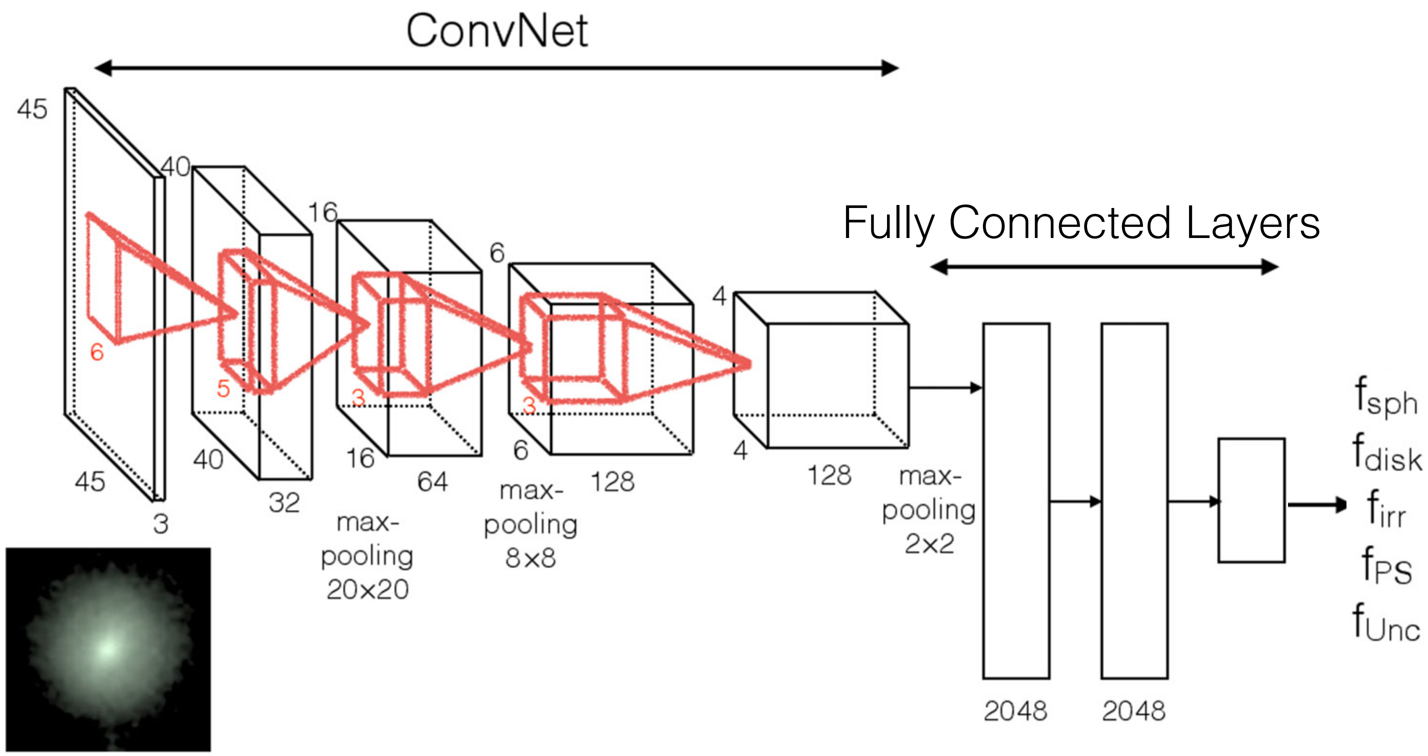
Dieleman, Willett, Dambre 2015, Rotation-invariant convolutional neural networks for galaxy morphology prediction, MNRAS

We present a deep neural network model for galaxy morphology classification which exploits translational and rotational symmetry. For images with high agreement among the Galaxy Zoo participants, our model is able to reproduce their consensus with near-perfect accuracy (>99 per cent) for most questions.

Marc Huertas-Company used Dieleman's code to classify CANDELS galaxy images

H-C et al. 2015, Catalog of Visual-like Morphologies in 5 CANDELS Fields Using Deep Learning

In this work, we mimic human perception with deep learning using convolutional neural networks (ConvNets). The ConvNet is trained to reproduce the CANDELS visual morphological classification based on the efforts of 65 individual classifiers who contributed to the visual inspection of all of the galaxies in the GOODS-S field. It was then applied to the other four CANDELS fields. The galaxy classification data was then released to the astronomical community.



Configuration of the Convolutional Neural Network used in this paper, based on the one used by Dieleman et al. (2015) on SDSS galaxies. It is made of 5 convolutional layers followed by 2 fully connected perceptron layers.

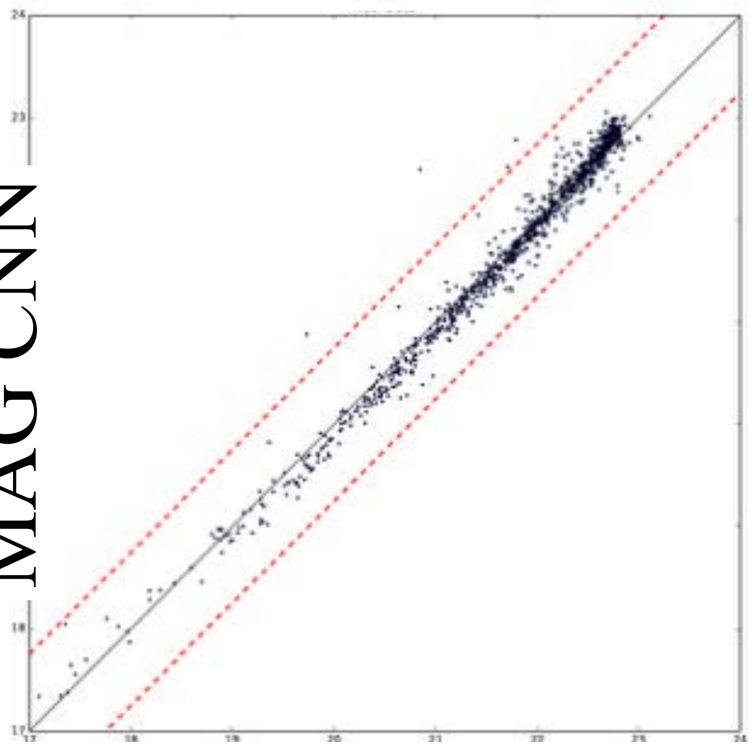
Following the approach in CANDELS, we associate five real numbers with each galaxy corresponding to the frequency at which expert classifiers flagged a galaxy as having a bulge, having a disk, presenting an irregularity, being compact or point-source, and being unclassifiable. Galaxy images are interpolated to a fixed size, rotated, and randomly perturbed before feeding the network to (i) avoid over-fitting and (ii) reach a comparable ratio of background versus galaxy pixels in all images. ConvNets are able to predict the votes of expert classifiers with a <10% bias and a ~10% scatter. This makes the classification almost equivalent to a visual-based classification. The training took 10 days on a GPU and the classification is performed at a rate of 1000 galaxies/hour.

H-C et al. 2016, Mass assembly and morphological transformations since z ~ 3 from CANDELS

We quantify the evolution of star-forming and quiescent galaxies as a function of morphology from z ~ 3 to the present. Our main results are: 1) At z ~ 2, 80% of the stellar mass density of star-forming galaxies is in irregular systems. However, by z ~ 0.5, irregular objects only dominate at stellar masses below $10^9 M_\odot$. 2) Quenching: We confirm that galaxies reaching a stellar mass $M_* \sim 10^{10.8} M_\odot$ tend to quench. Also, quenching implies the presence of a bulge: the abundance of massive red disks is negligible at all redshifts

Marc Huertas-Company and his group have used deep learning to emulate GALFIT. The deep learning (convolutional neural net CNN) emulator measurements agree with with GALFIT about as well as GALFIT run again on the images.

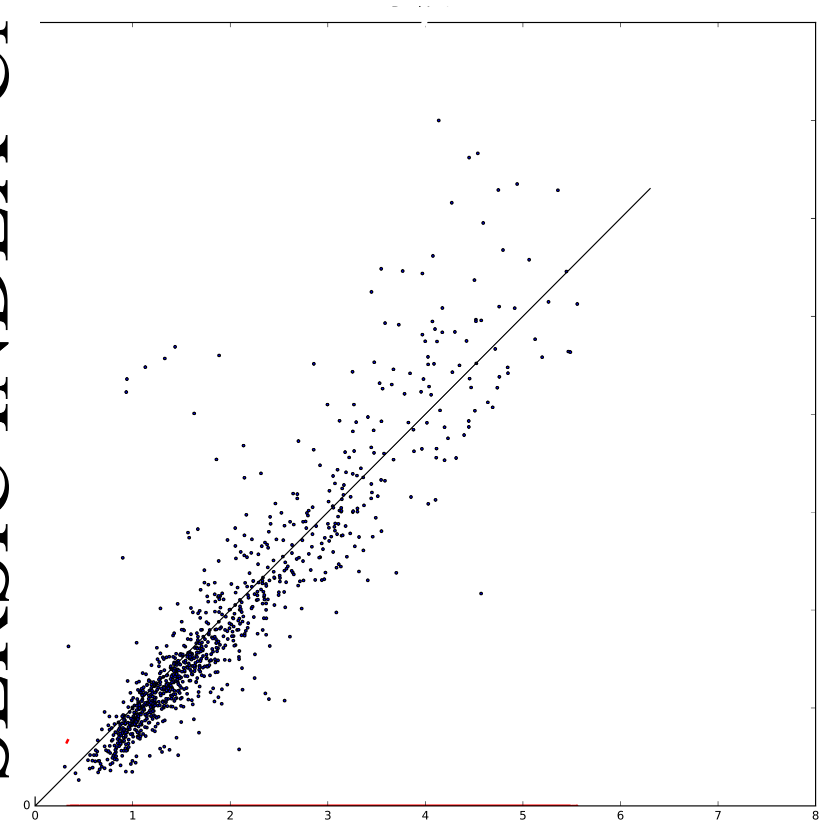
MAG CNN



MAGNITUDE GALFIT

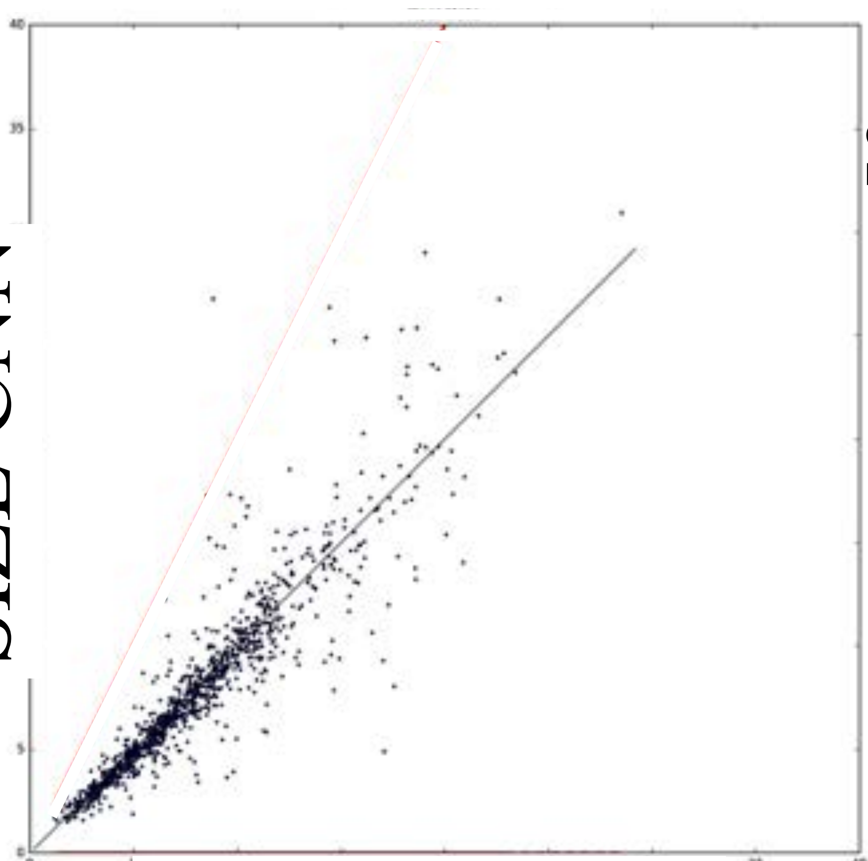
CNN analyzes ~1000 images per second while GALFIT takes hours and sometimes is problematic.

SERSIC INDEX CNN



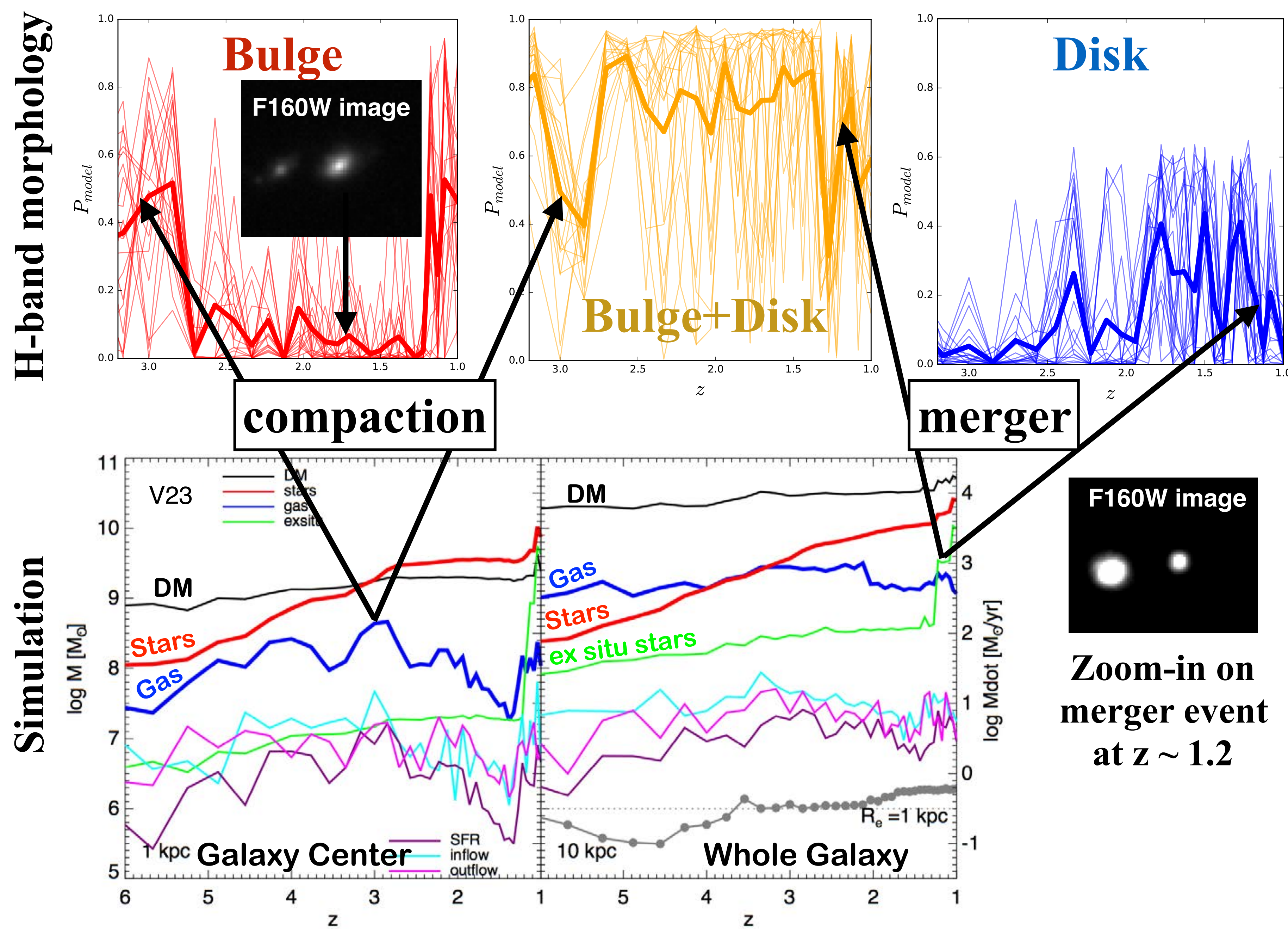
SERSIC INDEX GALFIT

SIZE CNN



SIZE GALFIT

Google has supported Marc H-C's visits to UCSC in summers 2016 and 2017, and his grad student Fernando Caro's visit March-August 2017 using deep learning, CANDELS images, and our galaxy simulations to understand galaxy formation



Evolution of zoom-in galaxy simulation VELA23-RP. The upper three panels show the probabilities that the galaxy is best fit by GALFIT as a single-Sersic Bulge or Disk, or instead as a double Sersic Bulge+Disk, based on classifications by a deep learning code trained using synthetic images. (Note that these probabilities do not need to sum to unity, since they are independent.) Classifications are plotted for 19 different orientations, with the medians plotted as heavy lines.

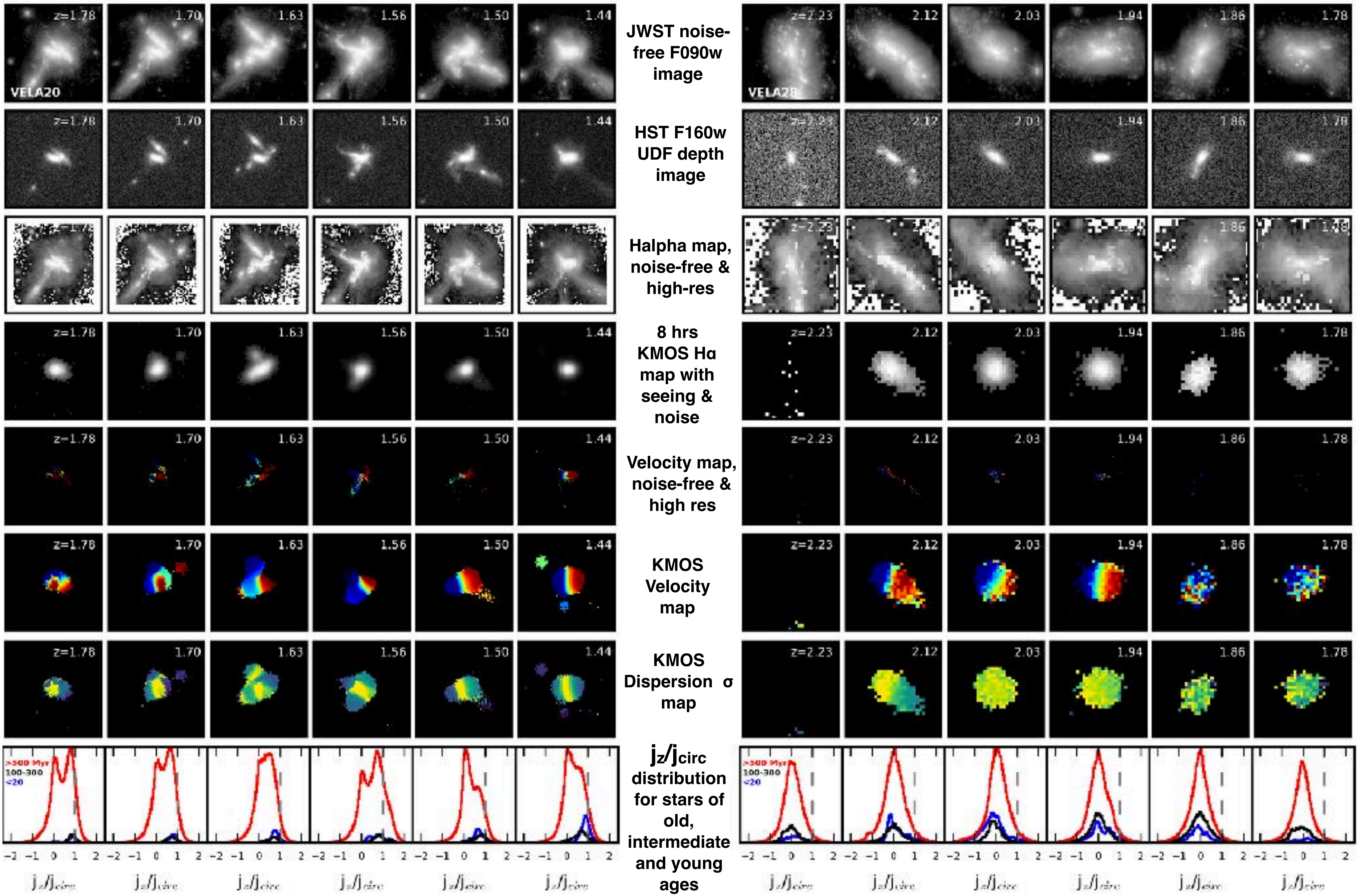
We want to give DL mock images and spectra + simulation metadata (recent major and minor mergers, counter-rotating gas flows, gas inflows, …) as a training set and see if DL can successfully predict the key phenomena from the images or the images + spectra.

For example, in the best cases of S/N and resolution, this might help discriminate between different causes of compaction. The images + spectra can also help discriminate between shear caused by mergers vs. rotation.

Greg Snyder and Raymond Simons have created a software pipeline to generate mock images and IFU data cubes from all the VELA simulations, with resolution appropriate for ground-based, HST, and JWST. It will work with essentially all current hydro simulations.

Avishai Dekel's Hebrew University group, including Santi Roca-Fabrege and Sharon Lapiner, Nir Mandelker at Yale, and others at UCSC are analyzing VELA and other simulations to create the simulation metadata set.

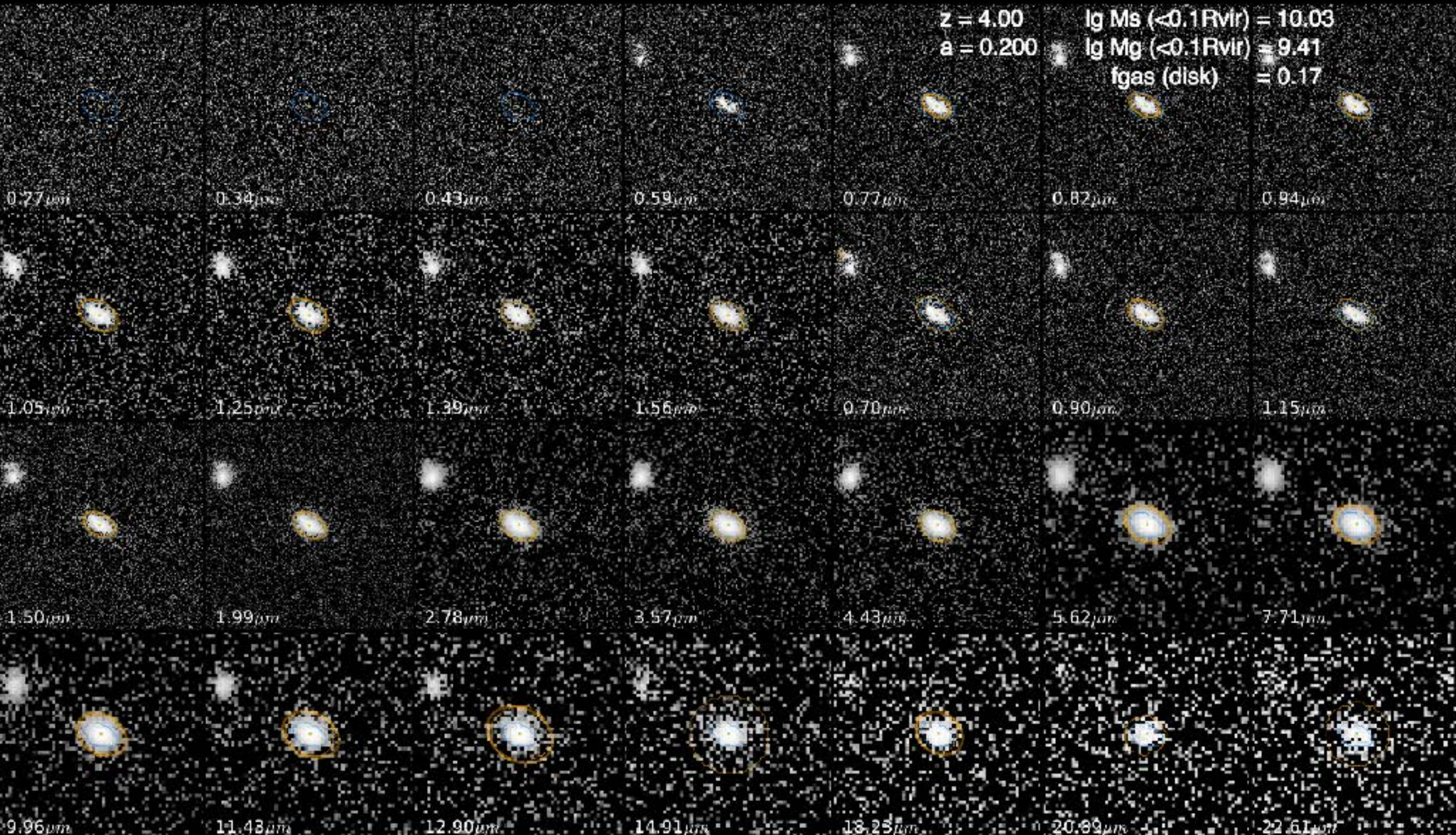
We want to give DL mock images and spectra + simulation metadata



Credit: Greg Snyder & Raymond Simons

HUDF S/N 27 mag/(arc sec)^2

VELA22-RP z = 4.00

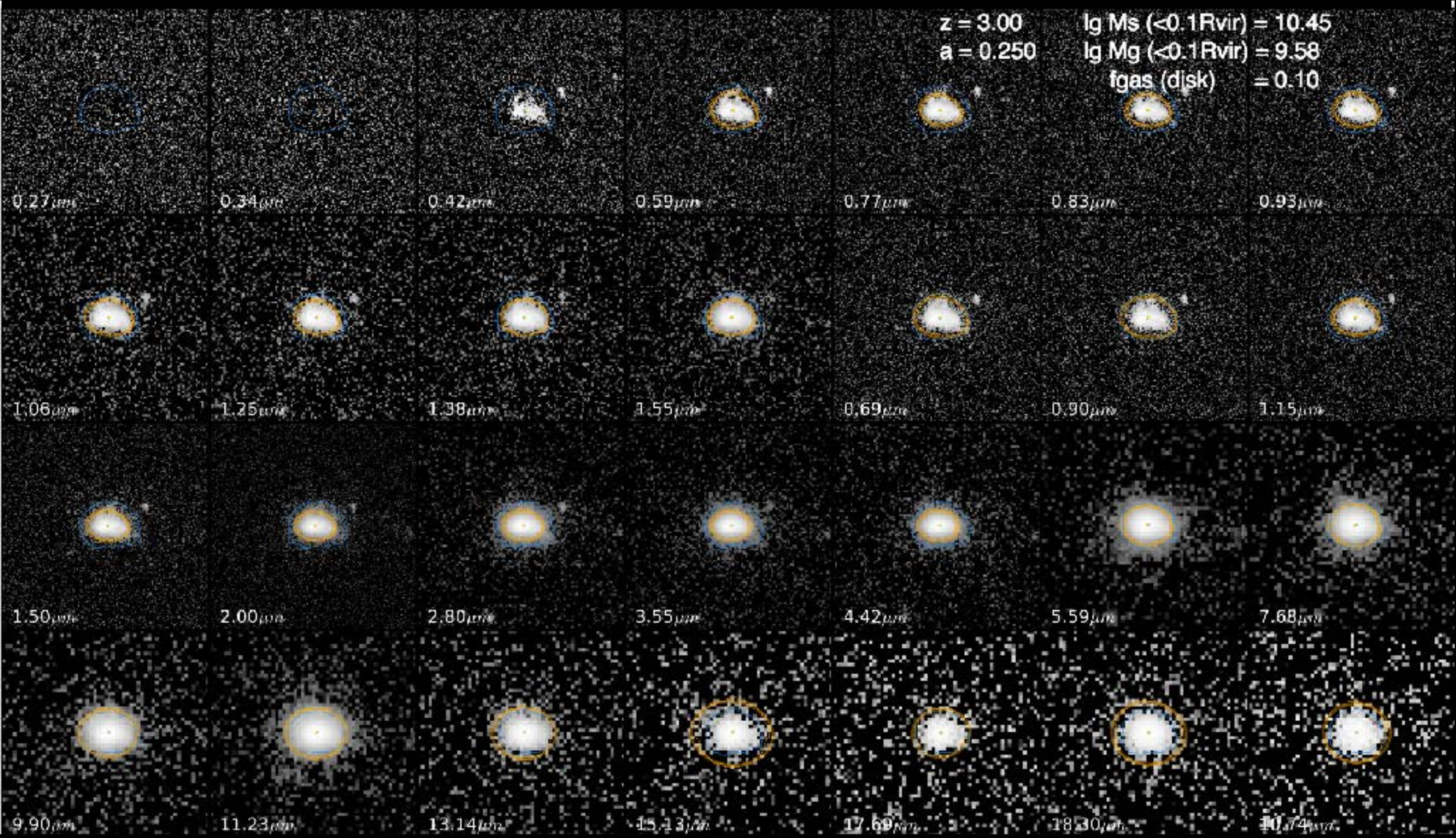


Camera 10 (fixed in simulation coordinates)

Credit: Greg Snyder

HUDF S/N

VELA22-RP z = 3.00

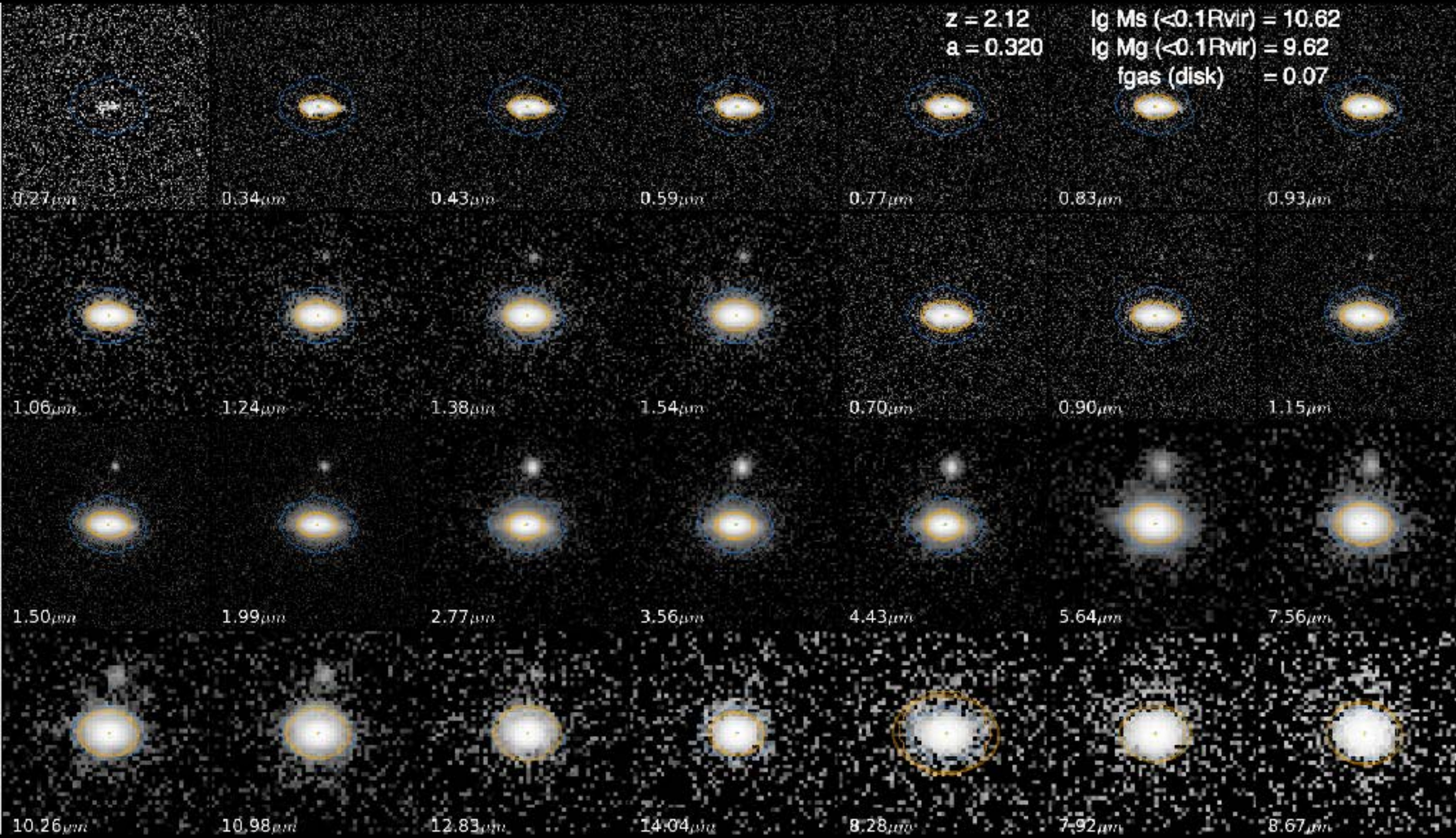


Camera 10 (fixed in simulation coordinates)

Credit: Greg Snyder

HUDF S/N

VELA22-RP z = 2.12

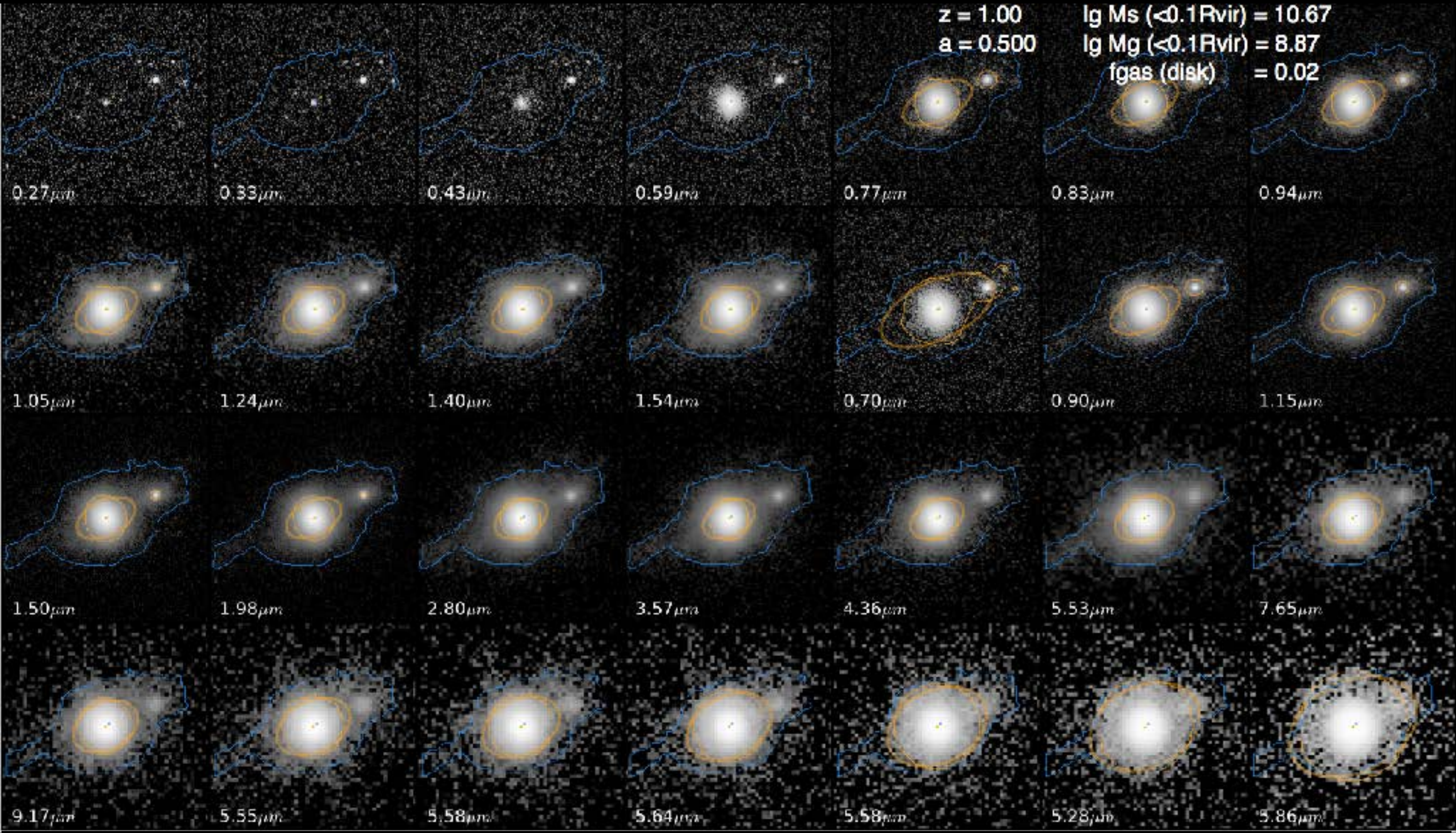


Camera 10 (fixed in simulation coordinates)

Credit: Greg Snyder

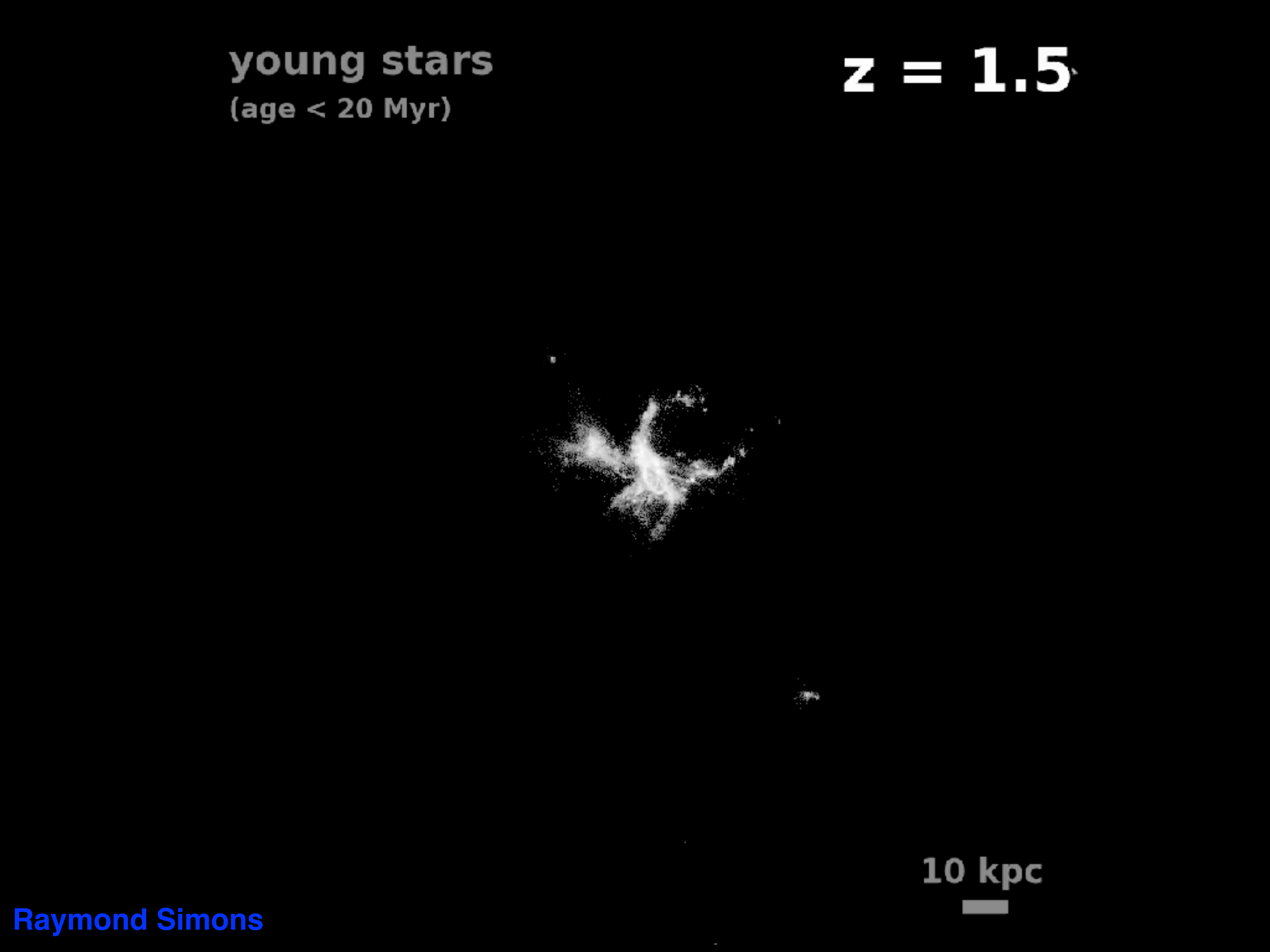
HUDF S/N

VELA22-RP z = 1.00



Camera 10 (fixed in simulation coordinates)

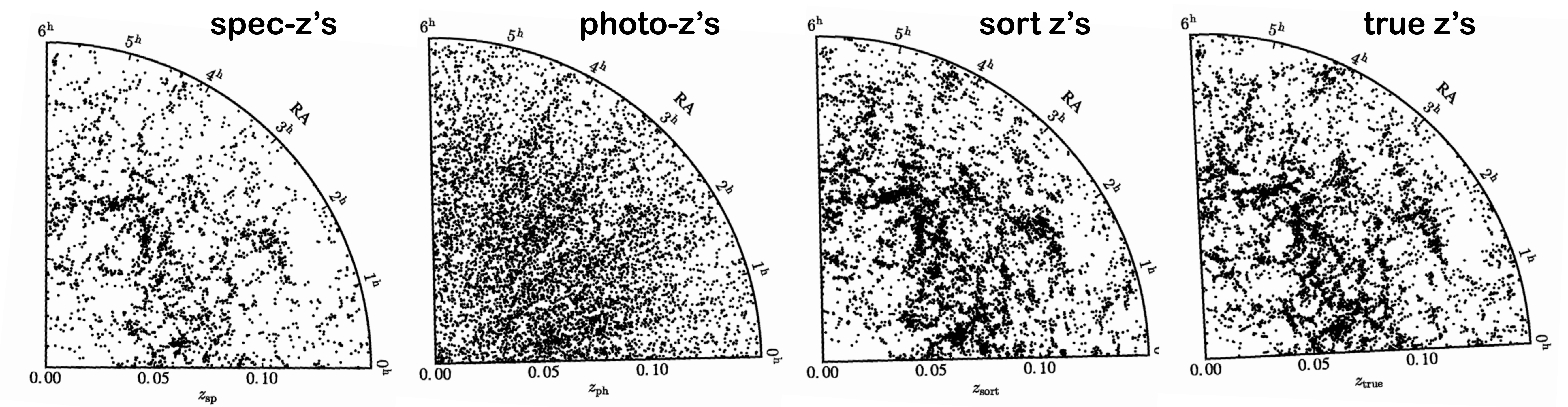
Credit: Greg Snyder

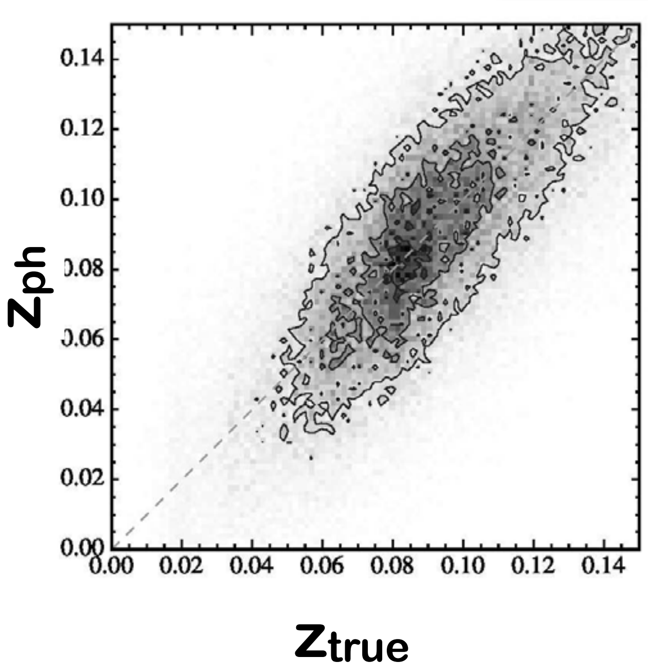

young stars
(age < 20 Myr)
z = 1.5
10 kpc
Raymond Simons

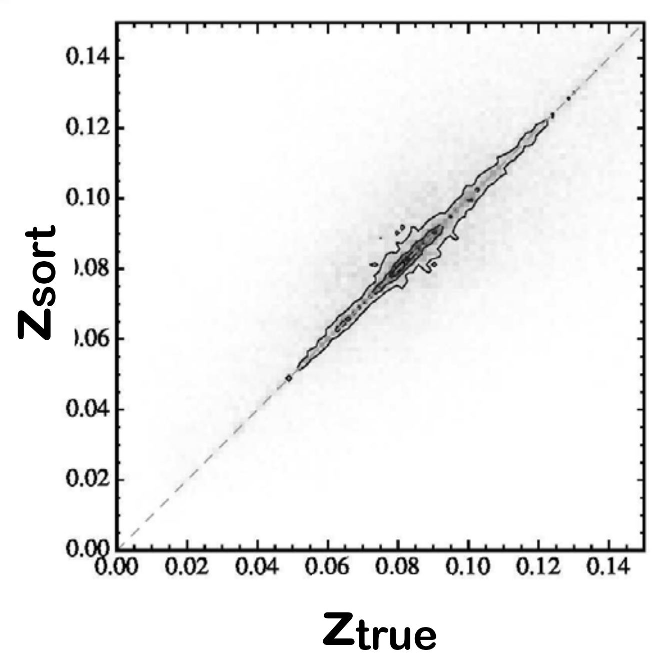
Another UCSC deep learning project: better galaxy environment estimates

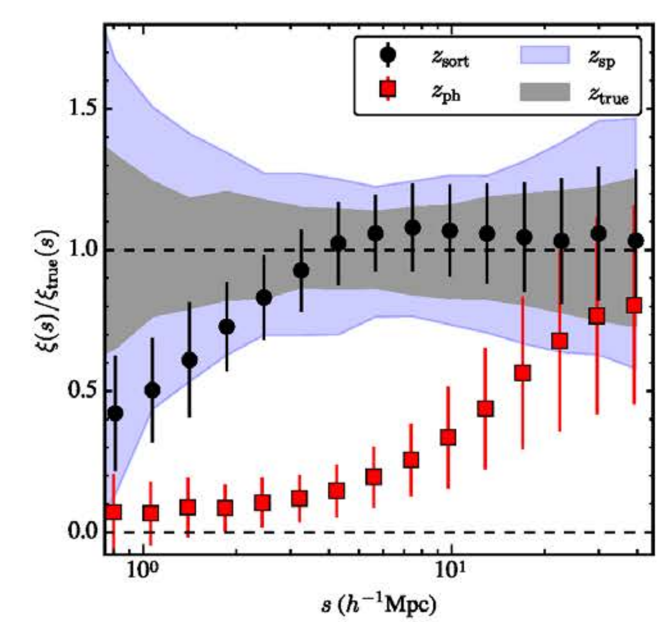
Joel Primack, Dave Koo, Doug Hellinger, UCSC grad students James Kakos, Dominic Pasquali

Images at various wavelengths (=>photometric redshifts, photo-z's) are much more plentiful than spectroscopic redshifts. How can we best combine a few spectroscopic z's with many photo-z's to estimate the environment of each galaxy? A preprint by Nicholas Tejos, Aldo Rodriguez-Puebla, and Joel Primack introduces a method ("SORT") to do this. Bryce Menard and collaborators have proposed a different approach. Can deep learning do better?







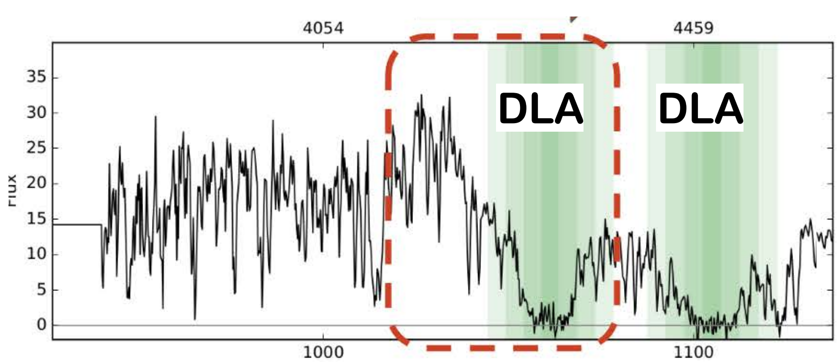


Ratio of measured and true 2-point correlation function as a function of redshift space distance s. Sort gets it right for s > 4 h^{-1} Mpc, while photo-z's fail even at s > 40 h^{-1} Mpc.

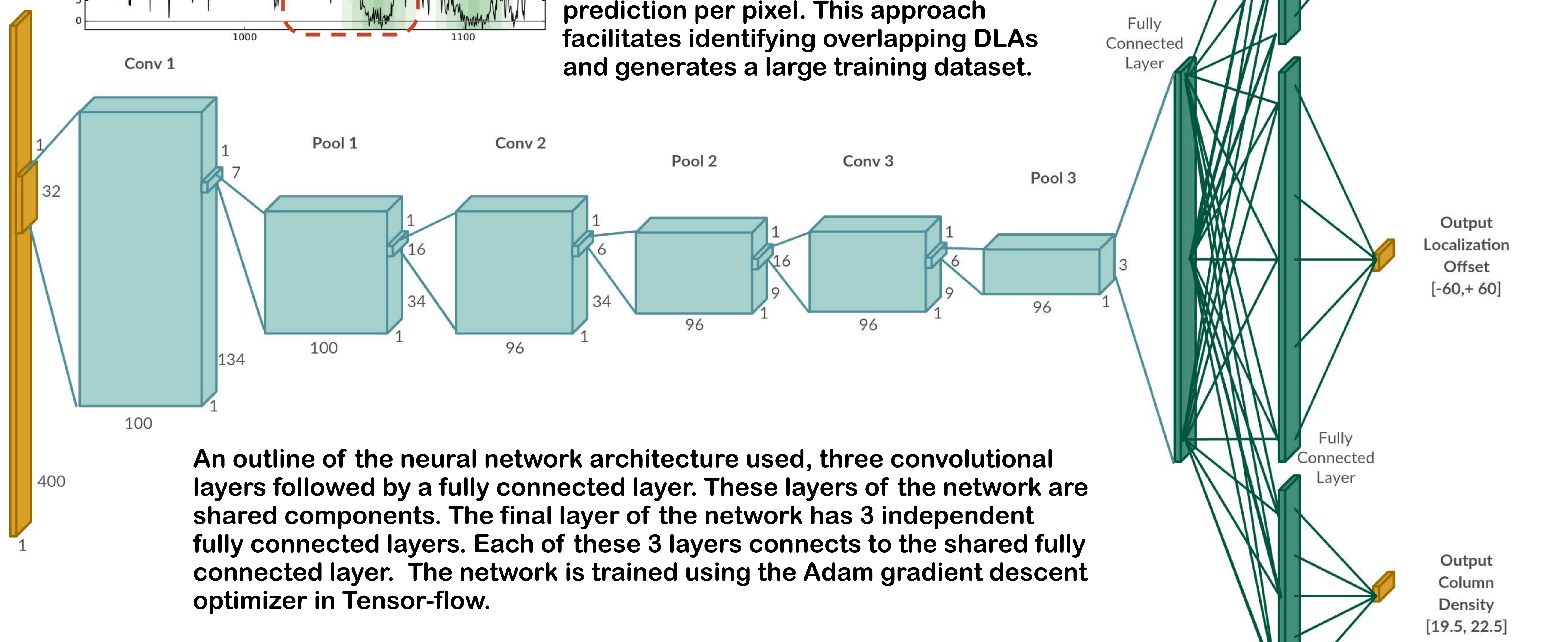
Another UCSC deep learning project: damped Lyα (DLA) systems in SDSS spectra

UCSC grad student David Park, Shawfeng Dong, J. Xavier Prochaska, Zheng Cai

DLA systems seen in quasar spectra, corresponding to at least 2×10^{20} hydrogen atoms/cm^2, represent most of the neutral hydrogen in the universe at redshifts z = 2 to 4. About 7000 DLAs were identified by astronomers in about 100,000 quasar spectra. The additional 270,000 sightlines that recently became available from the Sloan Digital Sky Survey were scanned for DLAs by a deep learning code, and the resulting DLA catalog will be made publicly available.



The sightline is broken into 400 pixel segments in a sliding window, so 1748 inference computations must be made for each sightline. Using each of the 1748 pixels in the sightline as the center point of a 400 pixel window generates a prediction per pixel. This approach facilitates identifying overlapping DLAs and generates a large training dataset.



An outline of the neural network architecture used, three convolutional layers followed by a fully connected layer. These layers of the network are shared components. The final layer of the network has 3 independent fully connected layers. Each of these 3 layers connects to the shared fully connected layer. The network is trained using the Adam gradient descent optimizer in Tensor-flow.

Astro data and computation are increasing exponentially
This will be challenging!

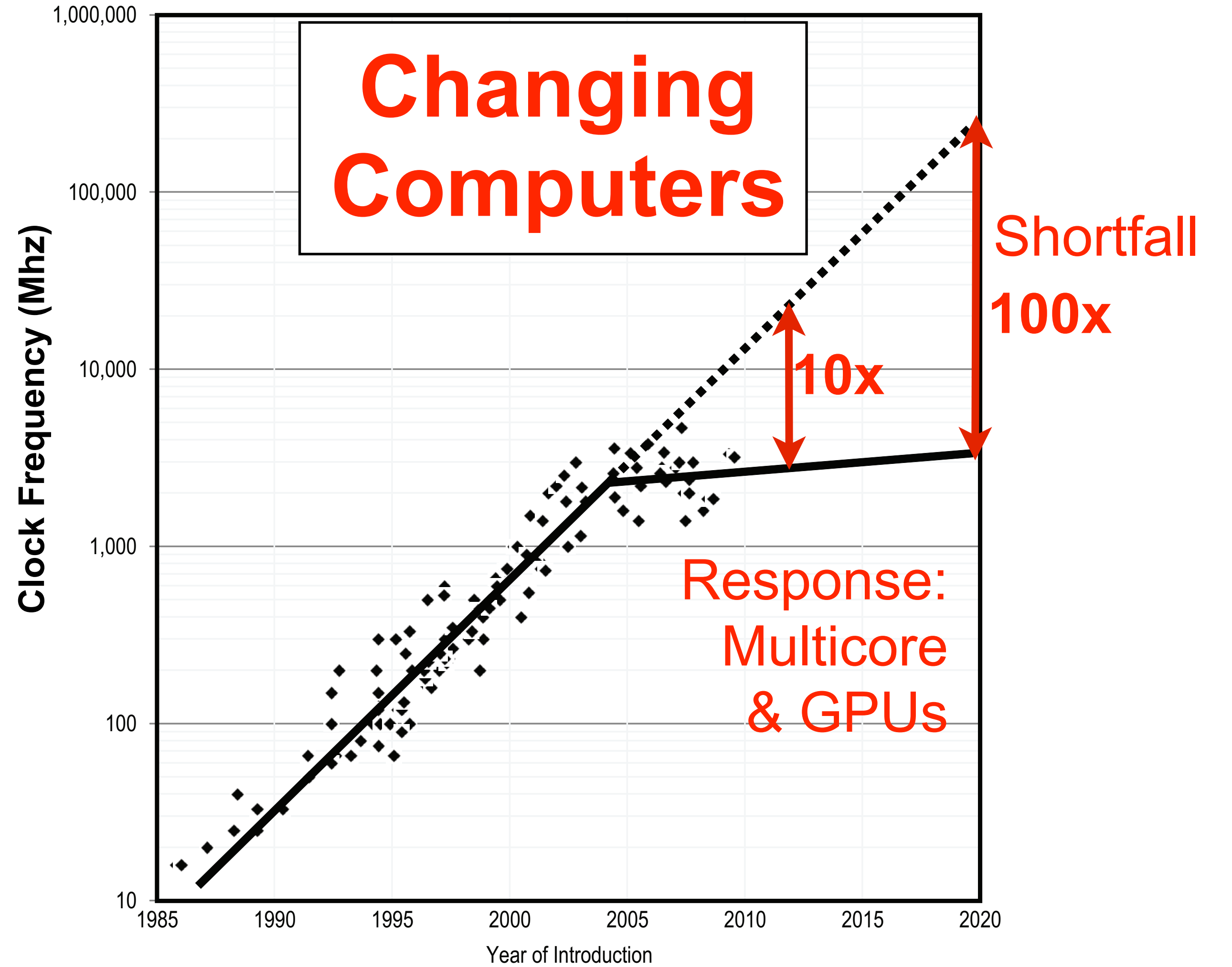
Big Data

Sloan Digital Sky Survey (SDSS) 2008
2.5 Terapixels of images
40 TB raw data ➠ 120 TB processed
35 TB catalogs

Mikulski Archive for Space Telescopes (MAST) 2013
185 TB of images
25 TB/year ingest rate
>100 TB/year retrieval rate

Large Synoptic Survey Telescope (LSST) ~2022
15 TB per night for 10 years
100 PB image archive
20 PB final database catalog

Square Kilometer Array (SKA) ~2024
1 EB per day (~ internet traffic today)
100 PFlop/s processing power
~1 EB processed data/year



Increasingly inhomogeneous computers are harder to program! We need computational scientists and engineers and new compilers that generate code for nodes with cores+accelerators with automatic load balancing and fault tolerance.

Deep Learning for Galaxies (a progress report)

Joel Primack

A deep learning code accurately predicted Galaxy Zoo galaxy image classifications, winning 2014 Kaggle competition

Marc Huertas-Company used deep learning to classify CANDELS galaxy images

H-C et al. 2015, Catalog of Visual-like Morphologies in 5 CANDELS Fields Using Deep Learning
H-C et al. 2016, Mass assembly and morphological transformations since z ~ 3 from CANDELS
Dimauro, H-C et al. 2017, Bulge and disk evolution in CANDELS — H-C's talk on Monday

Marc Huertas-Company and his group use DL to emulate GALFIT, etc.

Google supports Marc H-C's visits to UCSC Summer 2016 and 2017, and his grad student Fernando Caro's visit March-August 2017, using deep learning, HST and JWST images and spectra, and galaxy simulations to understand galaxy formation

Training set = mock images (or mock images plus spectra) plus simulation metadata, to see whether deep learning can successfully determine causes of morphological transformations

Better galaxy environment estimates with mostly photo-z's + some spec-z's

Another UCSC deep learning project: finding damped Lyα systems in SDSS spectra