



Visage dans étoile,
Pablo Picasso (1947)

TOBIAS BUCK & STEFFEN WOLF



PAINTING INTRINSIC
ATTRIBUTES ONTO SDSS OBJECTS

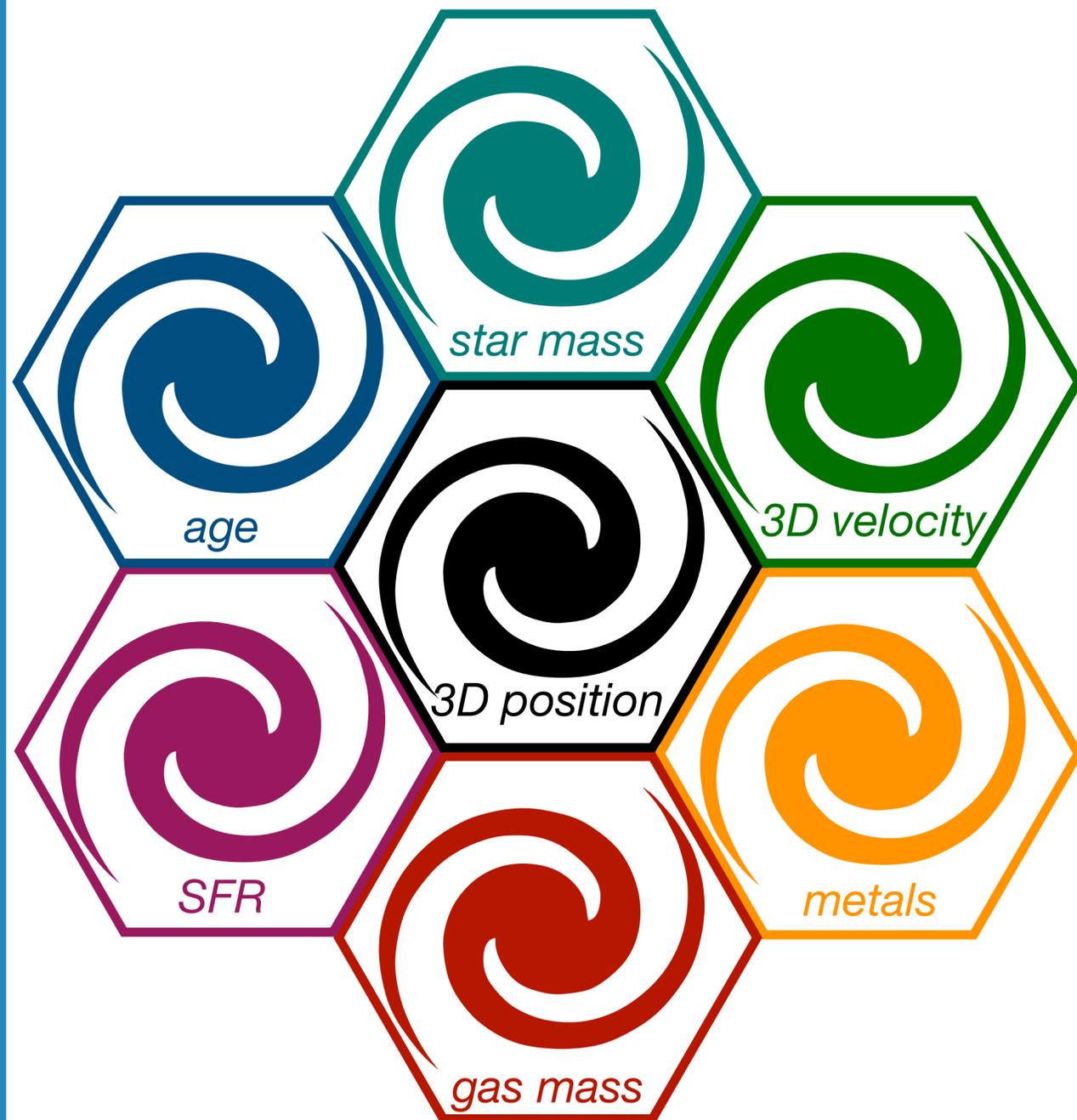
VIRTUAL ANNUAL MEETING OF THE GERMAN ASTRONOMICAL SOCIETY, 23. SEPTEMBER 2020

LEIBNIZ-INSTITUT FÜR ASTROPHYSIK POTSDAM (AIP),

tbuck@aip.de



derive physical parameters



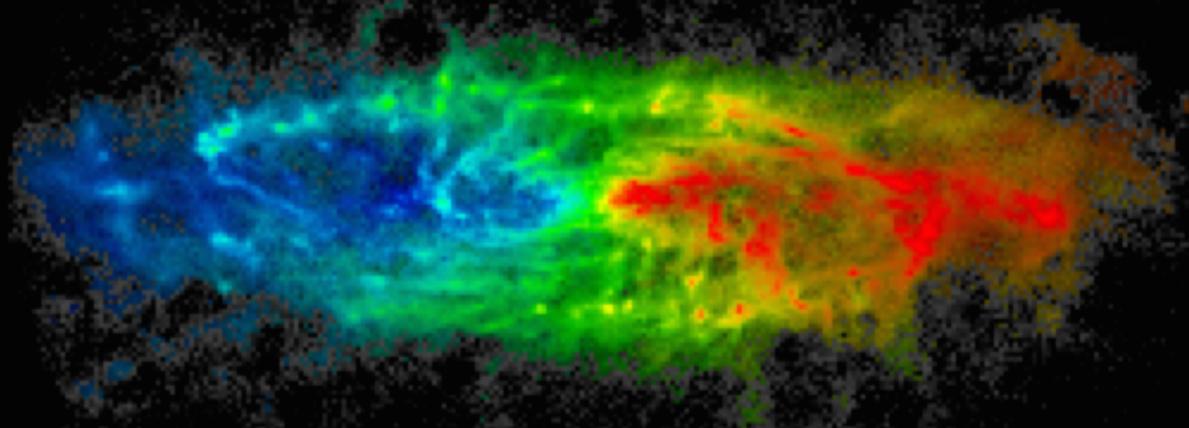
FULL FORWARD MODELLING: HYDRO SIMS

simulation

Gas + Spitzer + GALEX



Gas rotation

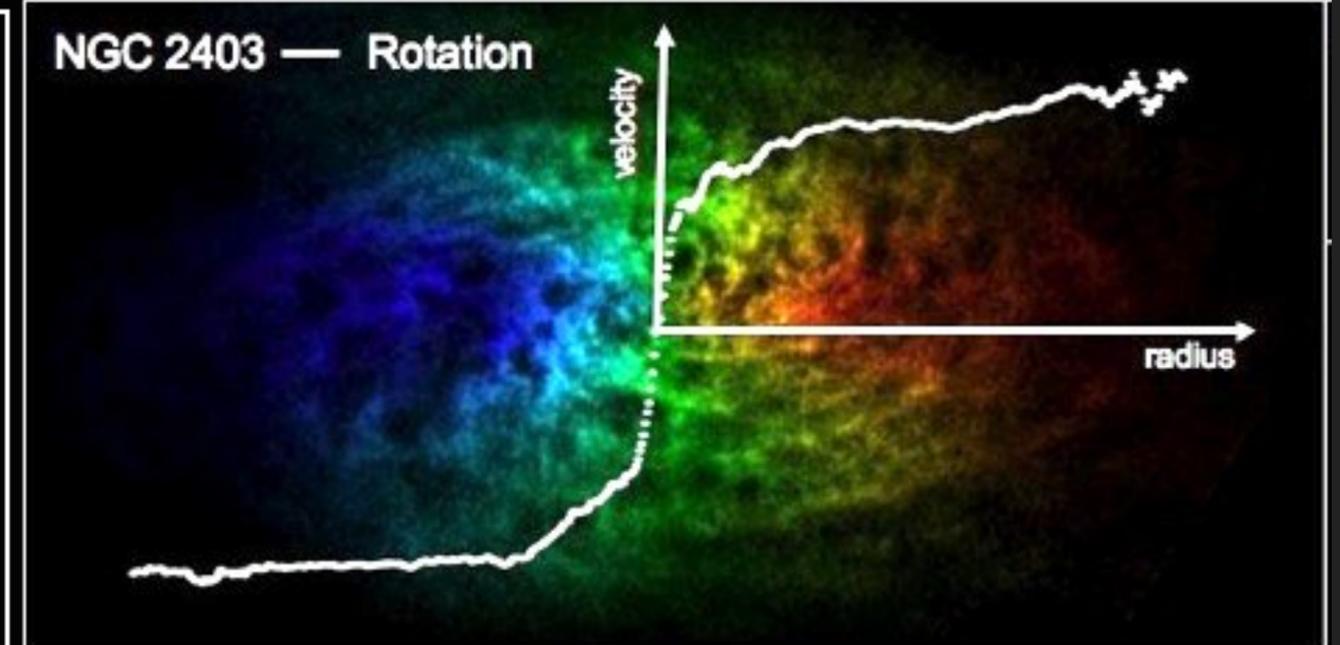
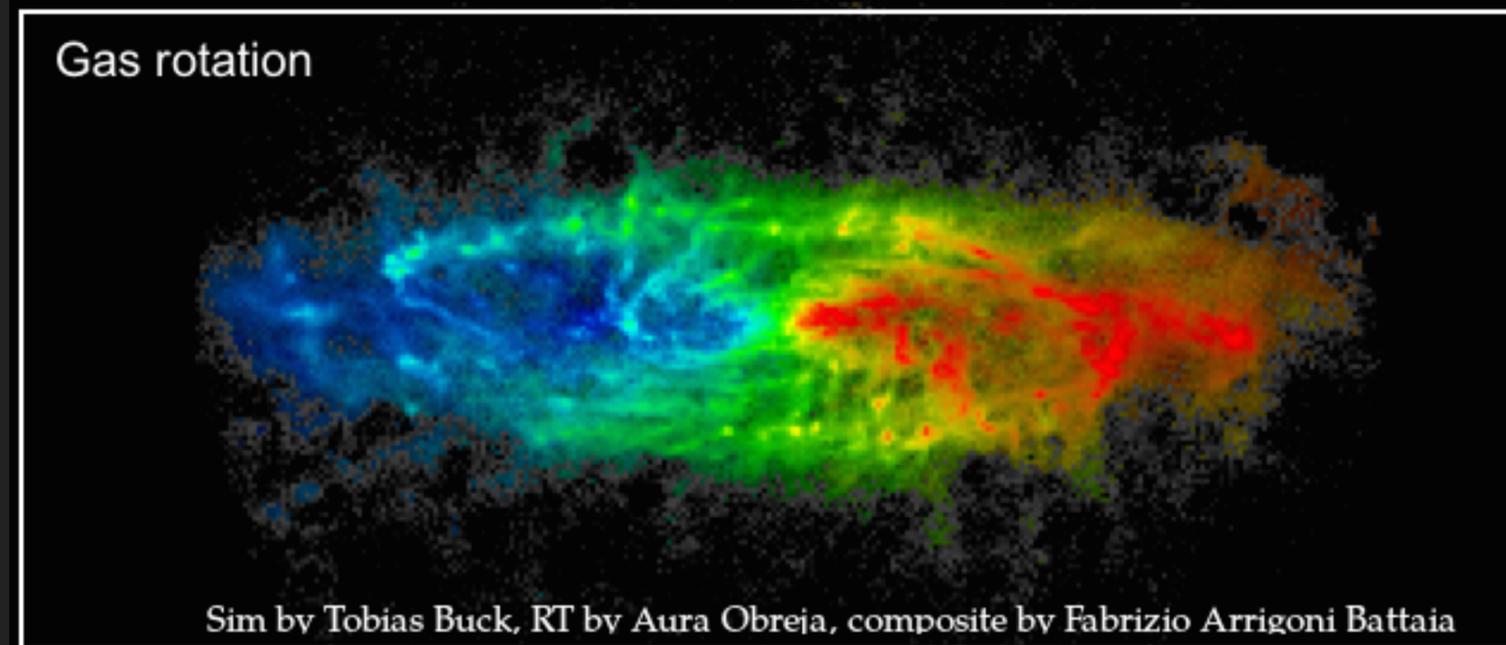
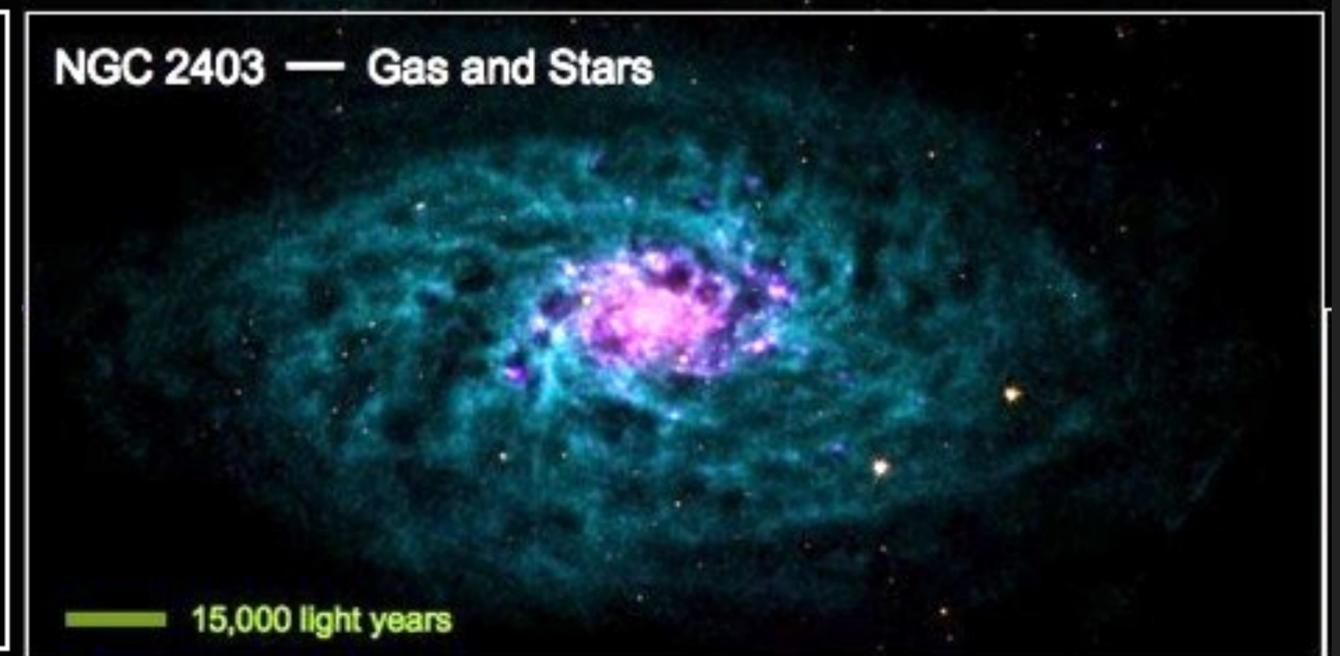
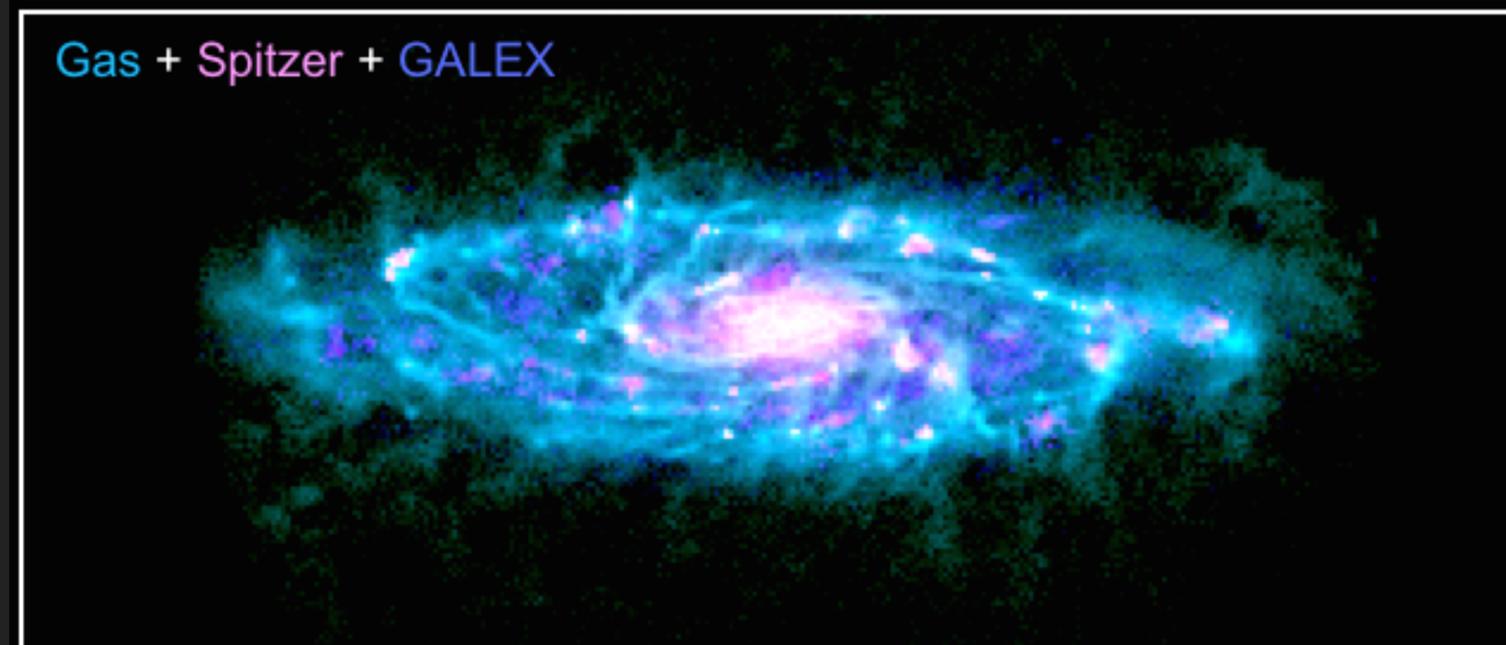


Sim by Tobias Buck, RT by Aura Obreja, composite by Fabrizio Arrigoni Battaia

FULL FORWARD MODELLING: HYDRO SIMS

simulation

observation



FULL FORWARD MODELLING: HYDRO SIMS

simulation

observation

Gas + Spitzer + GALEX

NGC 2403 — Gas and Stars

NOT FEASIBLE!

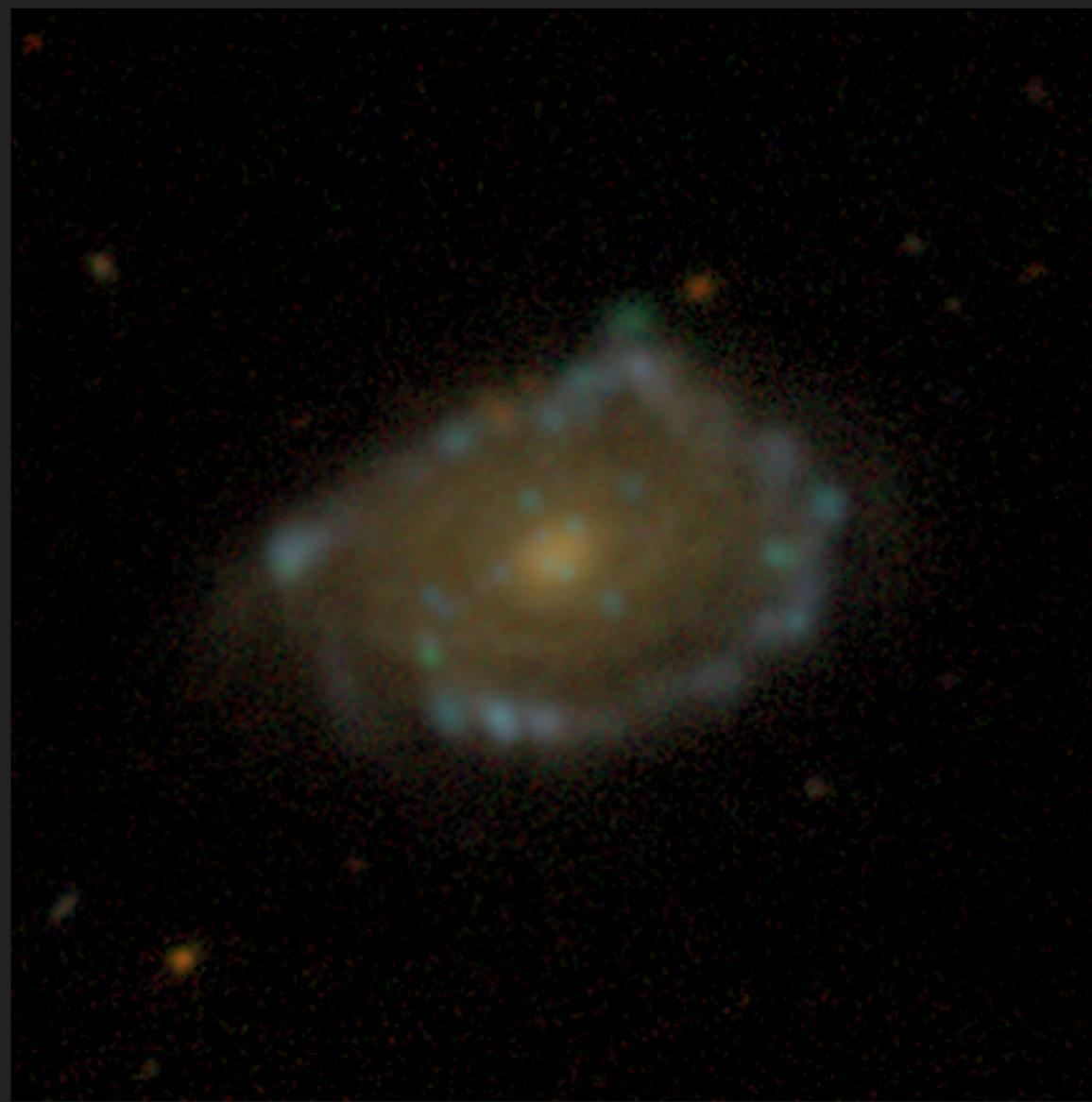
**MODEL UNCERTAINTIES,
TOO LOW SAMPLE SIZES, YOU'LL NEVER FIND
A CLOSE MATCH TO AN OBSERVED GALAXY**

Gas rotation

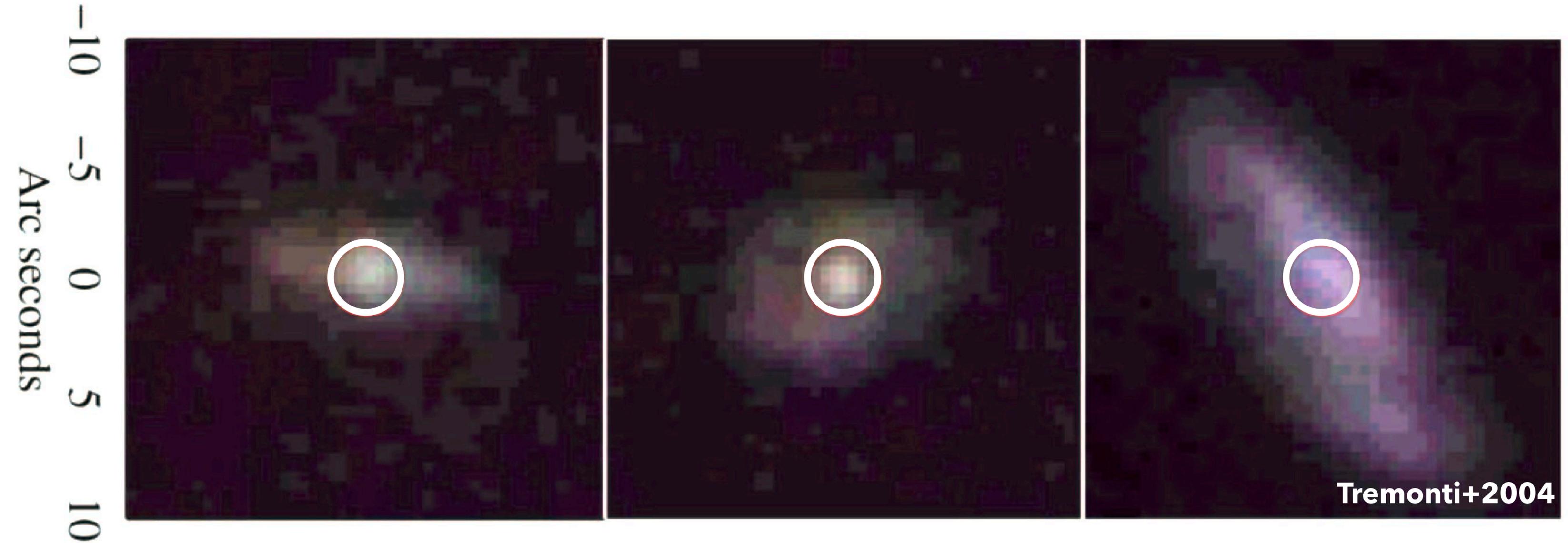
radius

Sim by Tobias Buck, RT by Aura Obreja, composite by Fabrizio Arrigoni Battaia

OBSERVATIONS

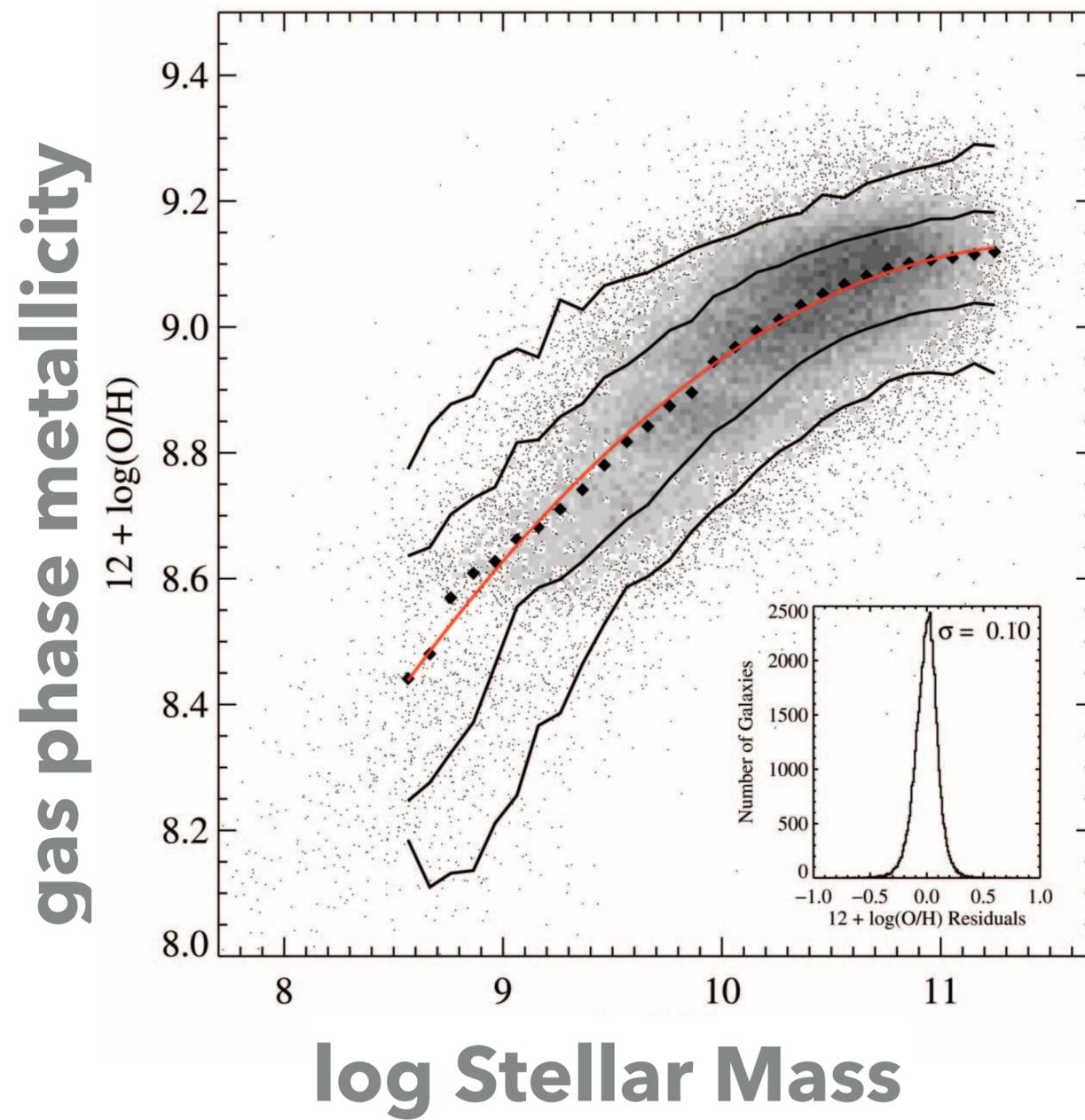


CLASSIC SDSS $\sim 100,000$ GALAXIES



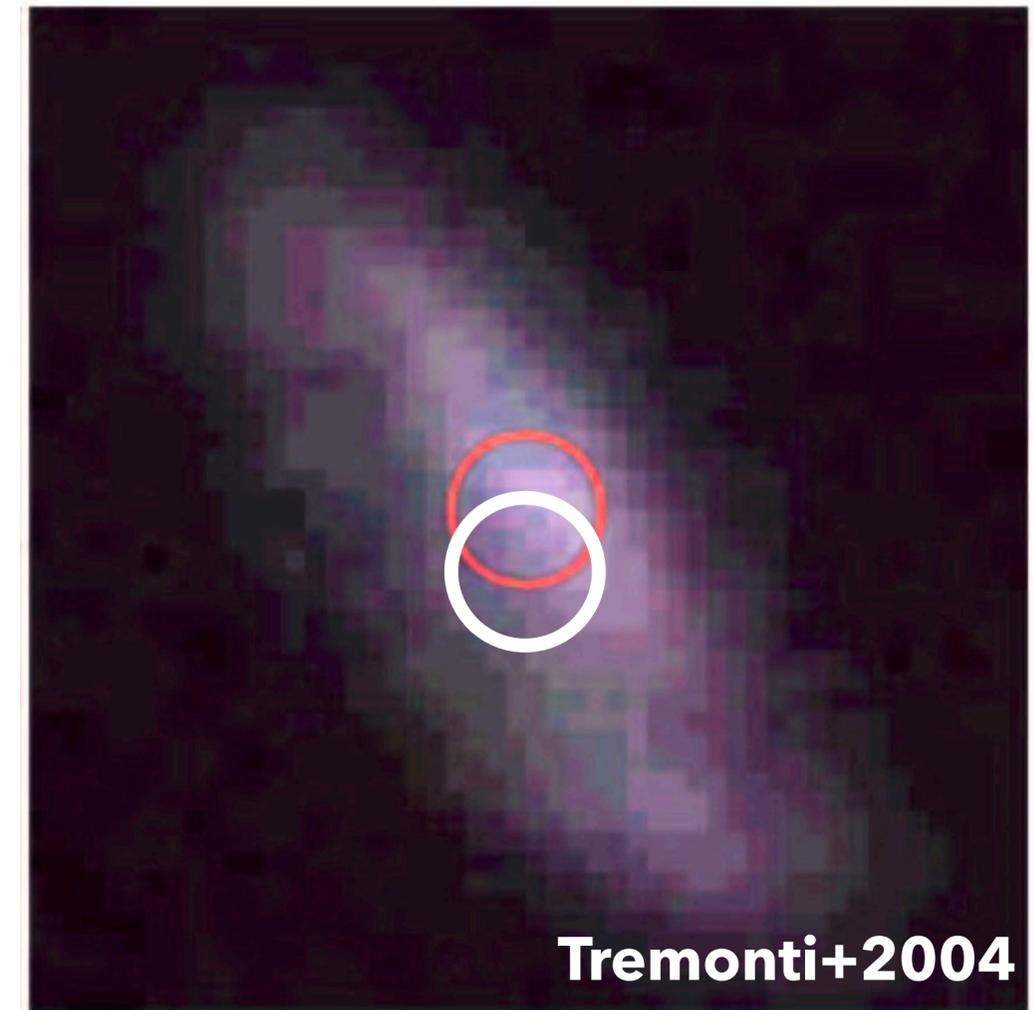
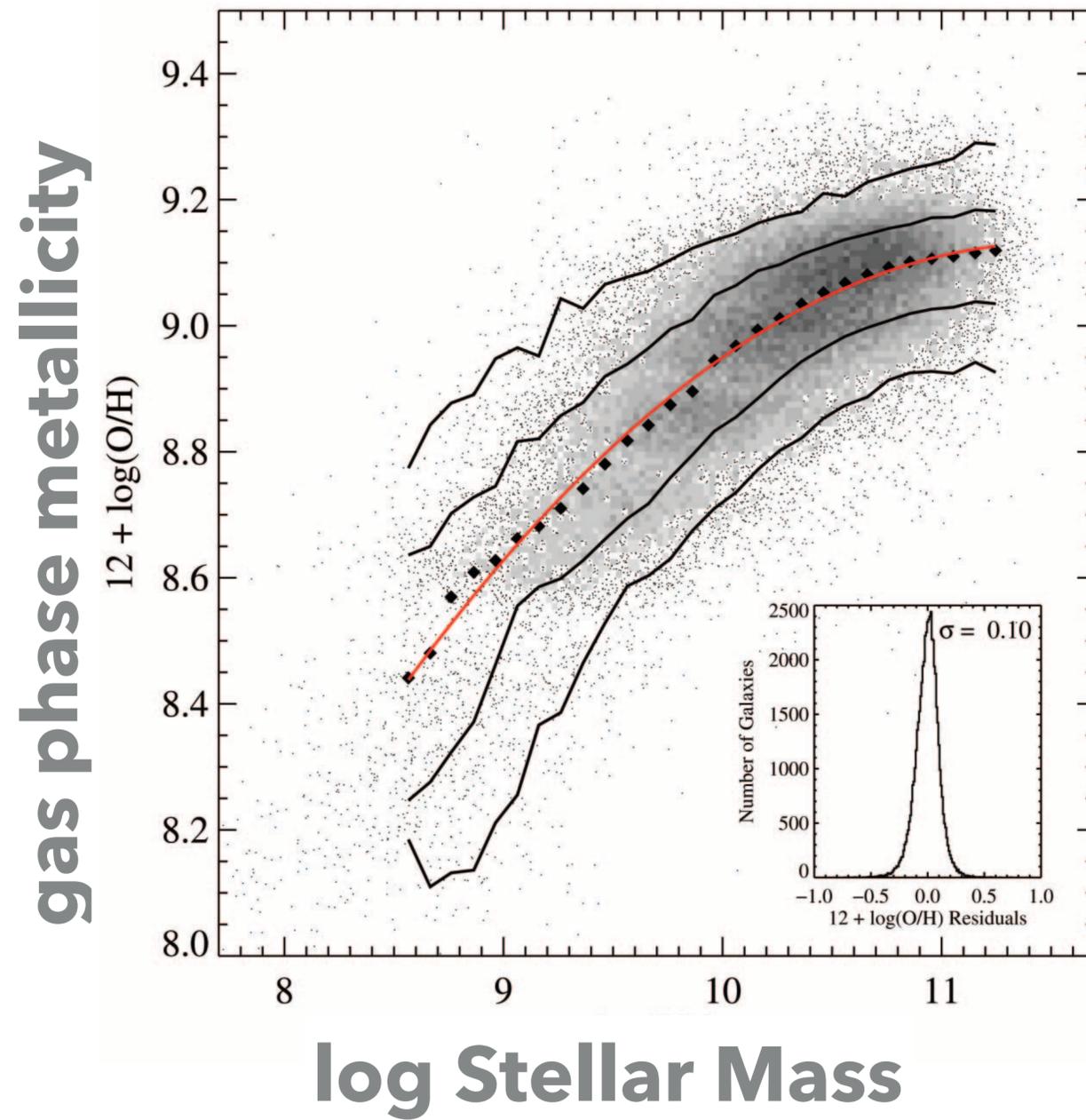
CLASSIC SDSS

~100.000 GALAXIES



CLASSIC SDSS

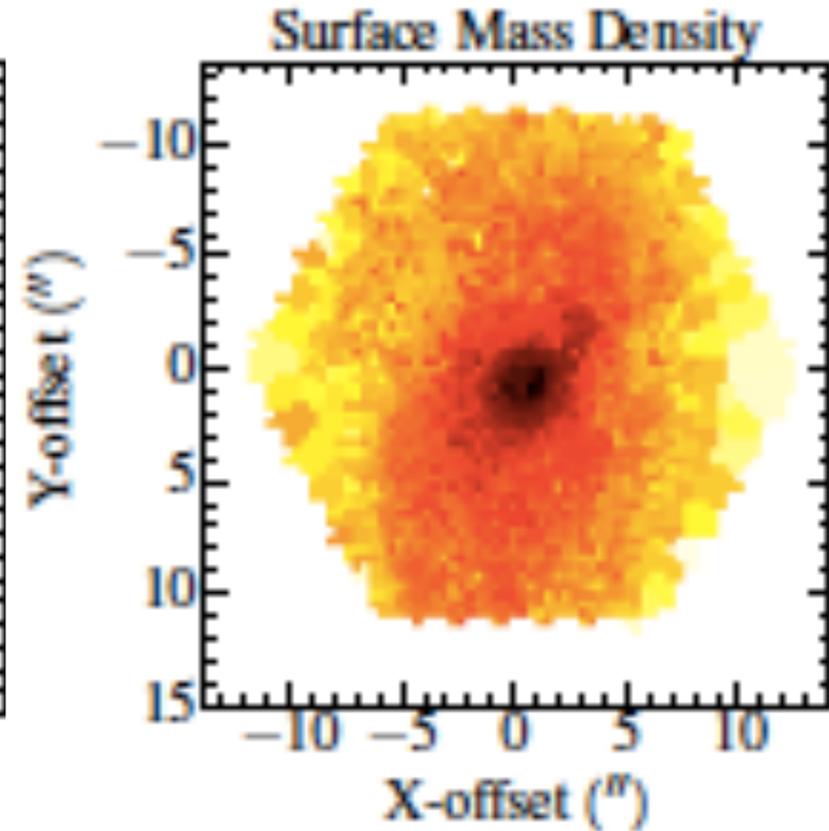
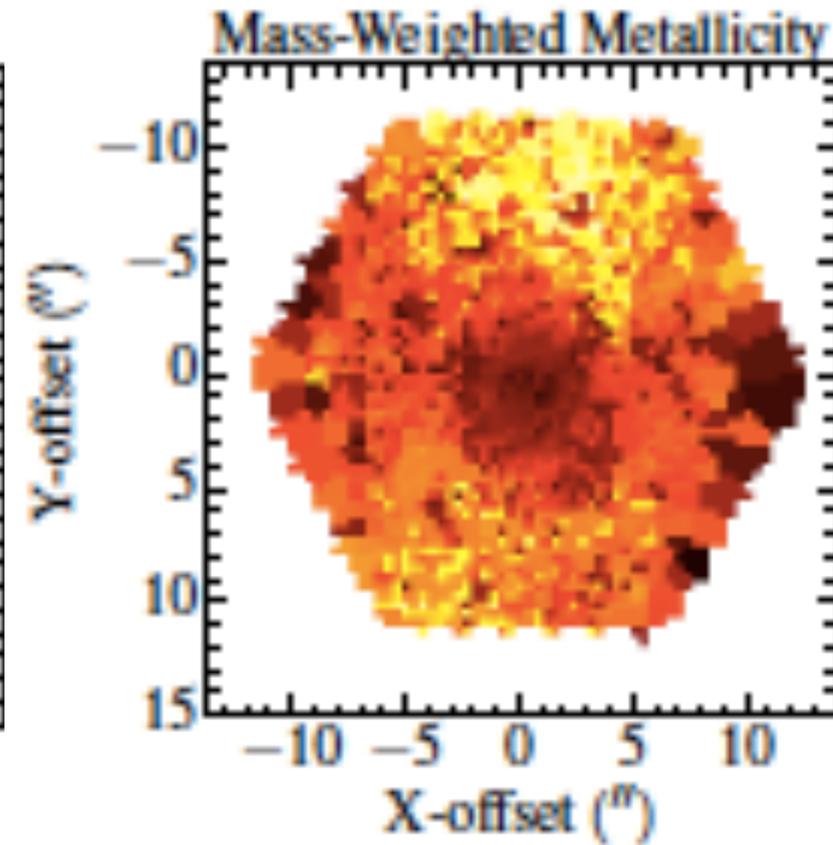
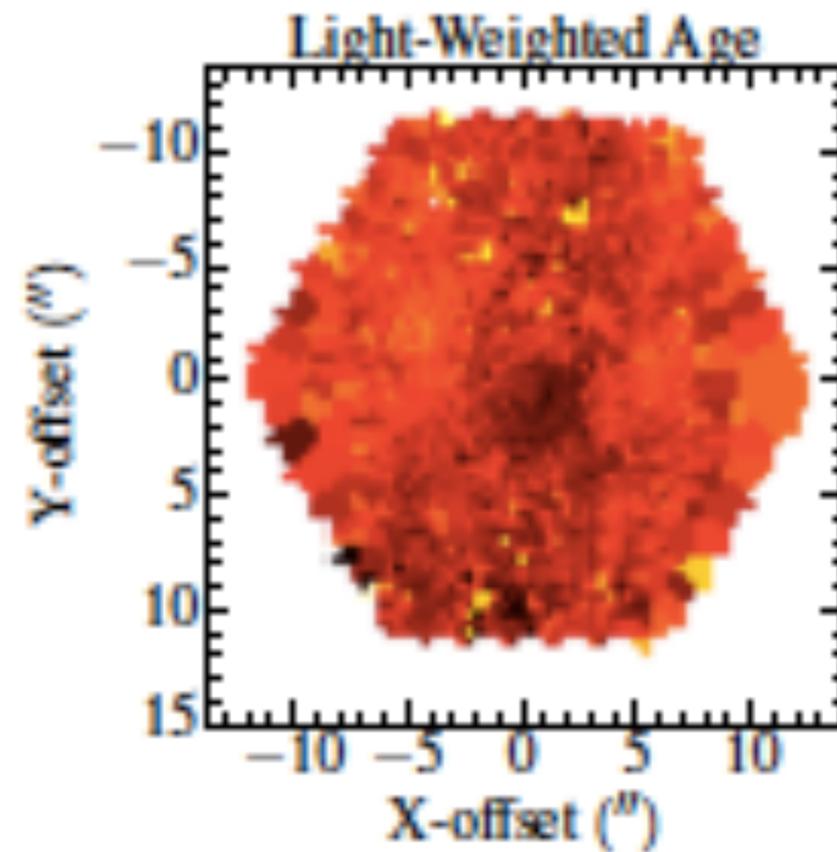
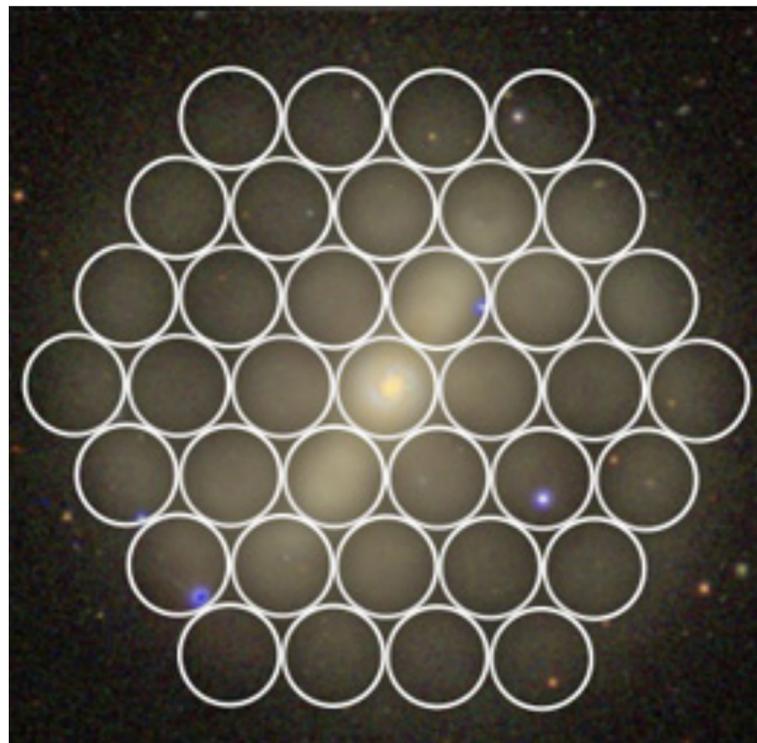
~100,000 GALAXIES



SDSS MANGA $\sim 10,000$ GALAXIES

EXPENSIVE INTEGRAL FIELD SPECTROSCOPY

OTHER EXAMPLES: SAMI, CALIFA, FORNAX3D, ETC.



LARGE SCALE SURVEYS: CHALLENGE FOR CONVENTIONAL ANALYSIS / MODELLING

PHOTOMETRIC DATA FOR MILLIONS OF GALAXIES,
(EUCLID, LSST, DES, COSMOS, DEEP2, BUT ALSO LEGACY DATA LIKE SDSS)

CLASSICAL ANALYSIS/CLASSIFICATION (VISUAL OR GALAXY ZOO LIKE) NOT FEASIBLE

DATA EXPLORATION BEYOND (SIMPLE) MORPHOLOGICAL CLASSIFICATION

- ▶ RESOLVED GALAXY PROPERTIES

**HOW MUCH INFORMATION
IS ENCODED IN BROAD
BAND GALAXY
IMAGES?**



CAN WE BUILD AN ANALYSIS TOOL WHICH:

▶ WORKS ON LARGE PHOTOMETRIC DATA SETS

A. FAST

▶ IS EASY TO HANDLE

C. AUTOMATION

D. GENERALIZATION

➔ FAST, OFF-THE-SHELF TOOL, READY TO USE

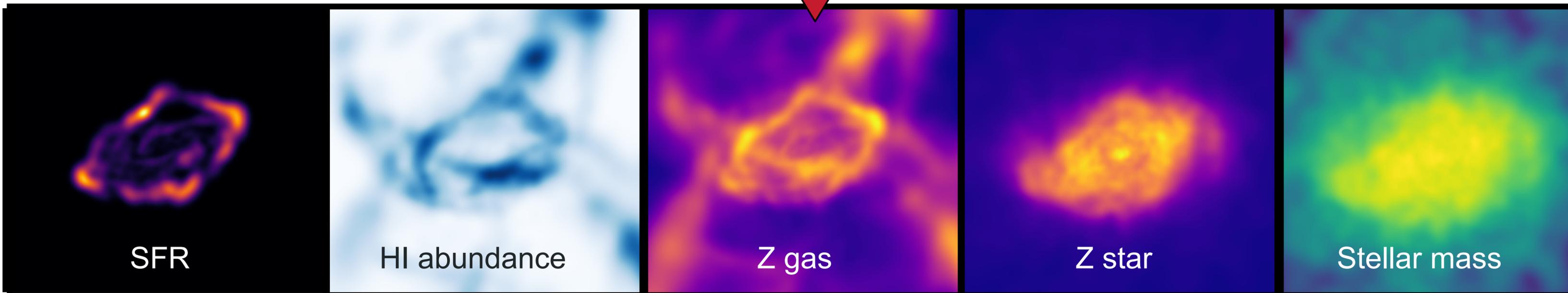
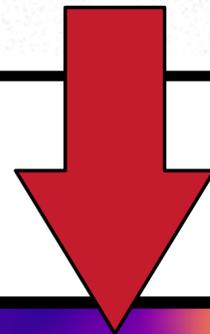
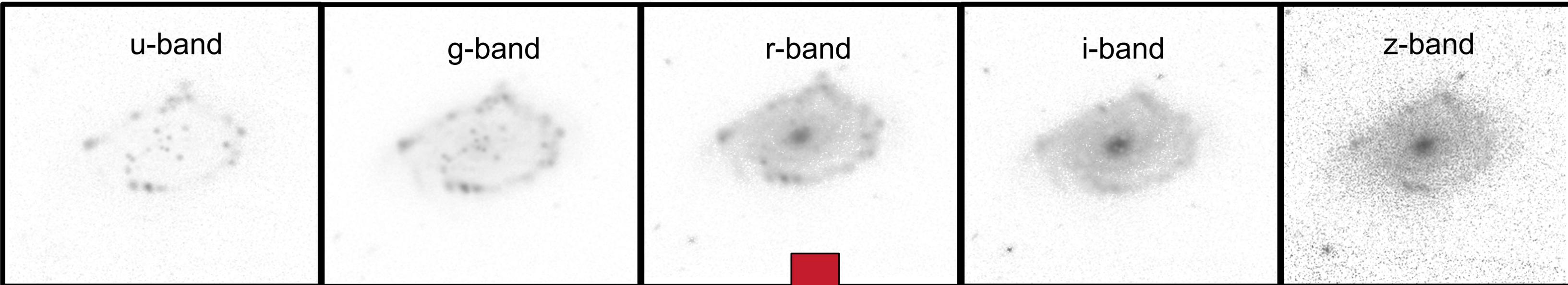
MOTIVATION/ROAD MAP

- ▶ Proof-of-concept: Does multi-band photometry contain enough information to recover resolved maps of intrinsic properties —> Knowledge transfer from IFU surveys
- ▶ Which properties can we recover? Can we do kinematics?
- ▶ What do we learn about galaxies? —> Inspect the latent space. How does the machine reconstructs galaxies?
- ▶ Can we make the model physically interpretable?
- ▶ How can we incorporate such models in future pipelines?
—> Sampling from latent space to create close analogues to observed galaxies

MOTIVATION/ROAD MAP

- ▶ Proof-of-concept: Does multi-band photometry contain enough information to recover resolved maps of intrinsic properties —> Knowledge transfer from IFU surveys
- ▶ Which properties can we recover? Can we do kinematics?
- ▶ What do we learn about galaxies? —> Inspect the latent space. How does the machine reconstructs galaxies?
- ▶ Can we make the model physically interpretable?
- ▶ How can we incorporate such models in future pipelines?
—> Sampling from latent space to create close analogues to observed galaxies

PHOTOMETRY TO PHYSICAL PROPERTIES



METHOD: DEEP LEARNING

THIS IS YOUR MACHINE LEARNING SYSTEM?

YUP! YOU POUR THE DATA INTO THIS BIG PILE OF LINEAR ALGEBRA, THEN COLLECT THE ANSWERS ON THE OTHER SIDE.

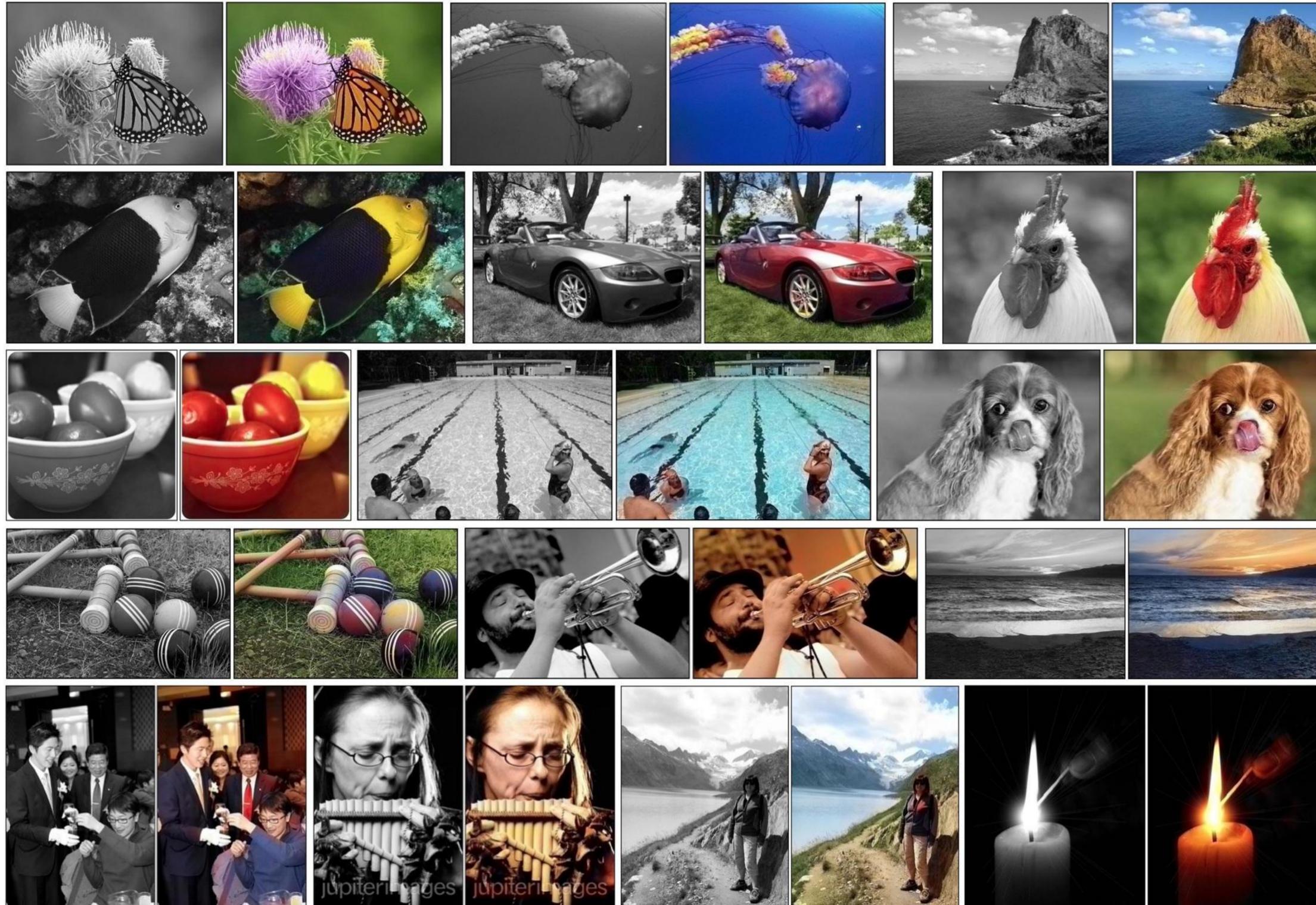
WHAT IF THE ANSWERS ARE WRONG?

JUST STIR THE PILE UNTIL THEY START LOOKING RIGHT.



SIMILAR APPLICATIONS

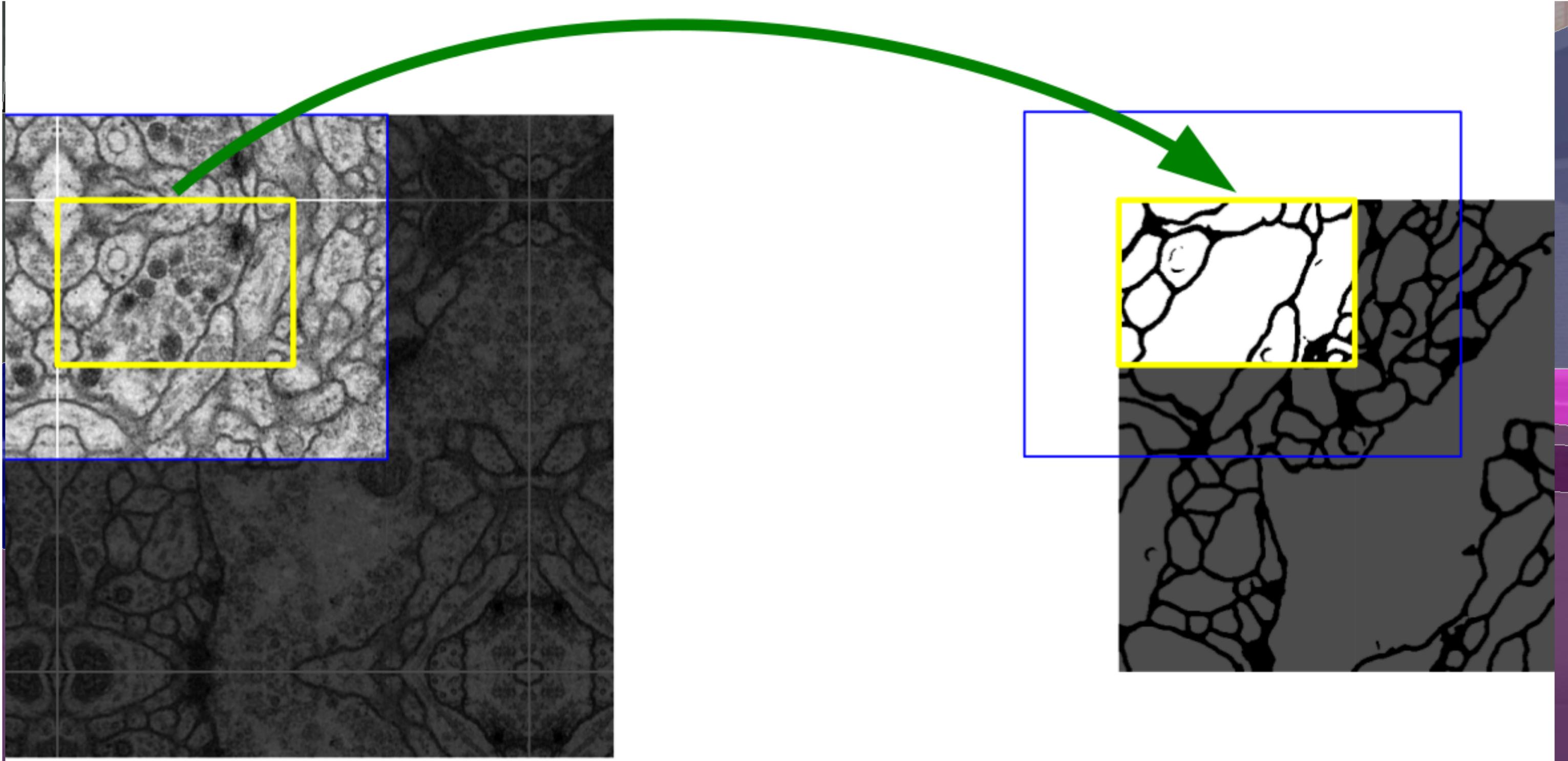
SIMILAR APPLICATIONS



SIMILAR APPLICATIONS



SIMILAR APPLICATIONS



SIMILAR APPLICATIONS

The image displays a grid of image transformation examples, organized into four rows. Each row shows an 'Input' image and an 'Output' image, with a caption below indicating the transformation. The examples are:

- Row 1:** Input: A brown horse standing in a field. Output: A zebra standing in the same field. Caption: horse → zebra.
- Row 2:** Input: Two brown horses standing on a beach. Output: Two zebras standing on the same beach. Caption: horse → zebra.
- Row 3:** Input: A brown horse lying down in a field. Output: A zebra lying down in the same field. Caption: horse → zebra.
- Row 4:** Input: A zebra standing in a savanna. Output: A brown horse standing in the same savanna. Caption: zebra → horse.
- Row 5:** Input: A group of zebras in a savanna. Output: A group of brown horses in the same savanna. Caption: zebra → horse.
- Row 6:** Input: Two zebras rearing up in a savanna. Output: Two brown horses rearing up in the same savanna. Caption: zebra → horse.
- Row 7:** Input: A basket of various fruits including apples and oranges. Output: A basket of various fruits including oranges and apples. Caption: apple → orange.
- Row 8:** Input: A tree with red apples. Output: A tree with orange fruit. Caption: apple → orange.
- Row 9:** Input: A basket of oranges. Output: A basket of red apples. Caption: orange → apple.
- Row 10:** Input: A tree with orange fruit. Output: A tree with red apples. Caption: orange → apple.

On the left side, there is a large grayscale texture image with a yellow bounding box and a green arrow pointing to it. On the right side, there is a large grayscale texture image with a blue bounding box and a yellow arrow pointing to it.

SIMILAR APPLICATIONS

Input



Output



Input



Output



Input



Output



Share a common Architecture:

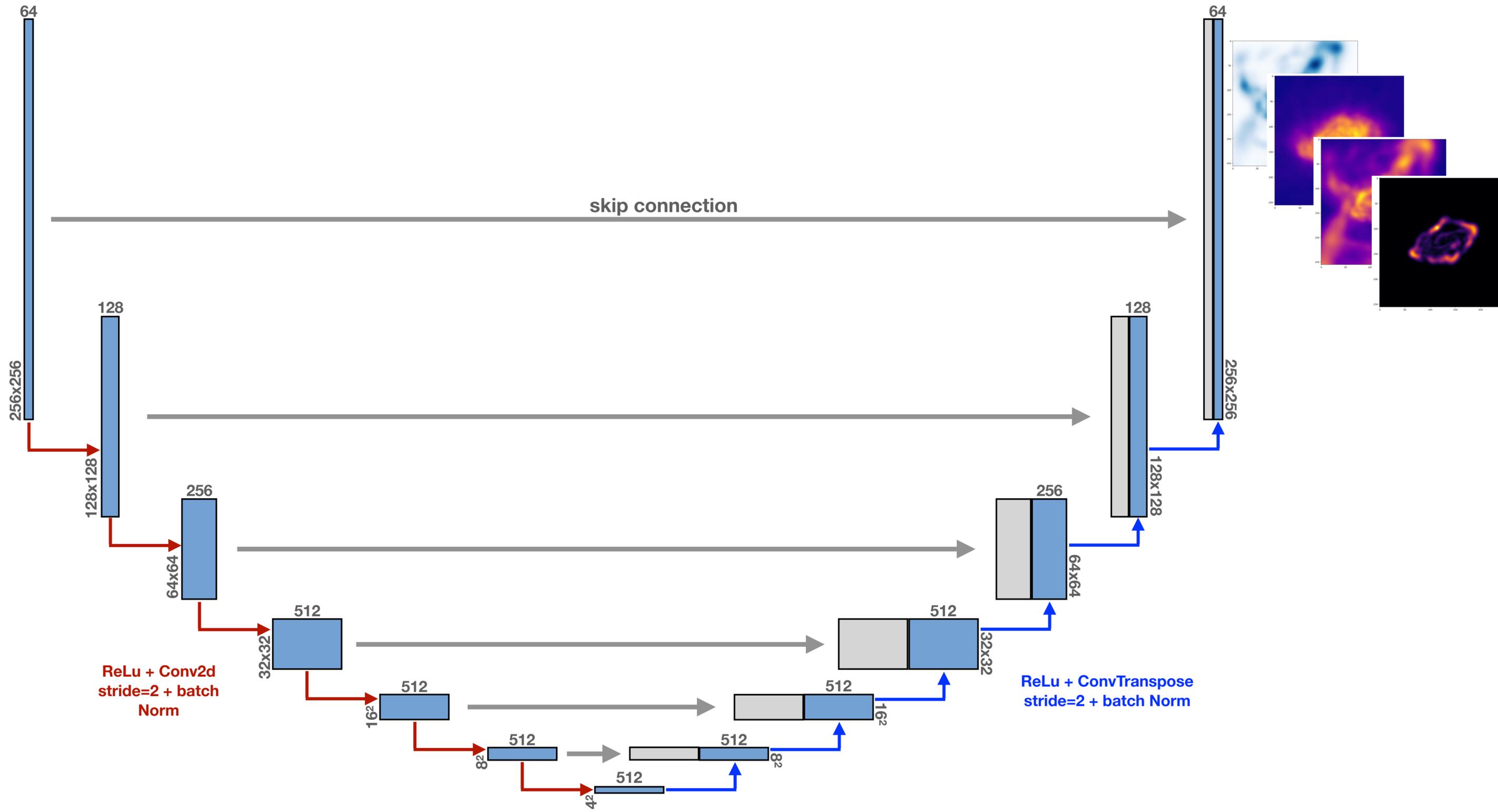
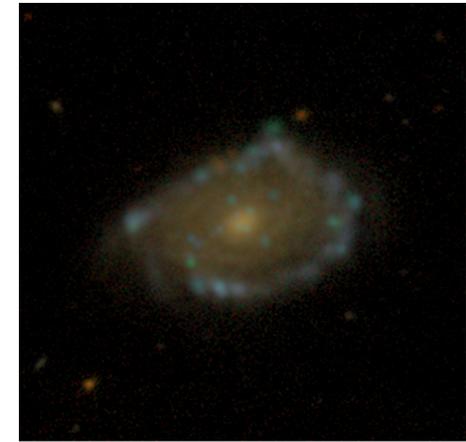
UNet (Ronneberger+2015)

apple → orange



orange → apple

U-NET ARCHITECTURE: PIX2PIX Isola+ CVPR 2017 (CONDITIONAL GAN)



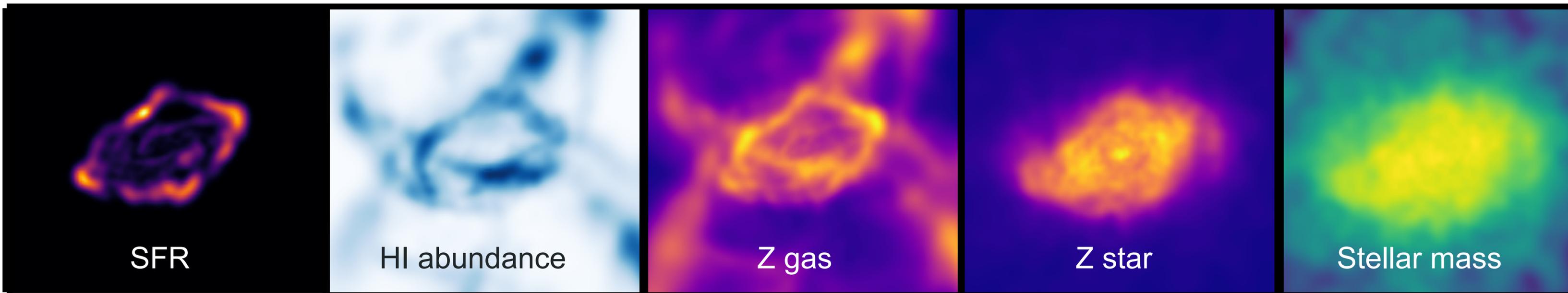
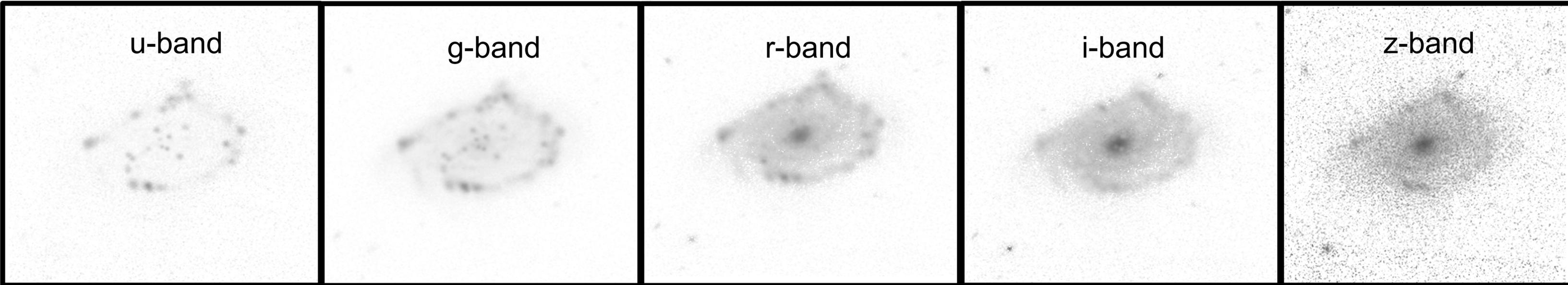
WHAT IS DIFFERENT WHEN PREDICTING PHYSICAL PROPERTIES

- Almost all CNNs are classifiers: $Y \in \{0, 1\}^N$
 - Here $Y \in \mathbb{R}^N$ with multiple orders of magnitude
1. Predict $\log(Y)$
 2. Quantized Regression [Güler et al. *CVPR* 2017]

Bins $B = \{-14, -12, \dots, 0, 2\}$
quantiles $q \in [0, 1]^{|\mathcal{B}|-1}$
residuals $r \in [0, 1]^{|\mathcal{B}|-1}$

$$f_{\theta}(x) = \sum_{i=0}^{|\mathcal{B}|-2} q_i (B_i + r_i (B_{i+1} - B_i))$$

PROOF OF CONCEPT: ILLUSTRIS DATA



PROOF OF CONCEPT: ILLUSTRIS DATA

u-band

g-band

r-band

i-band

z-band

- ▶ SDSS MOCK IMAGES 256X256 PIXELS TORREY+2014, SNYDER+2015
- ▶ RADIATIVE TRANSFER, BACKGROUND STARS, PSF, NOISE, SURFACE BRIGHTNESS CUT
- ▶ PHYSICAL PROPERTIES ON SAME SCALE

SFR

HI abundance

Z gas

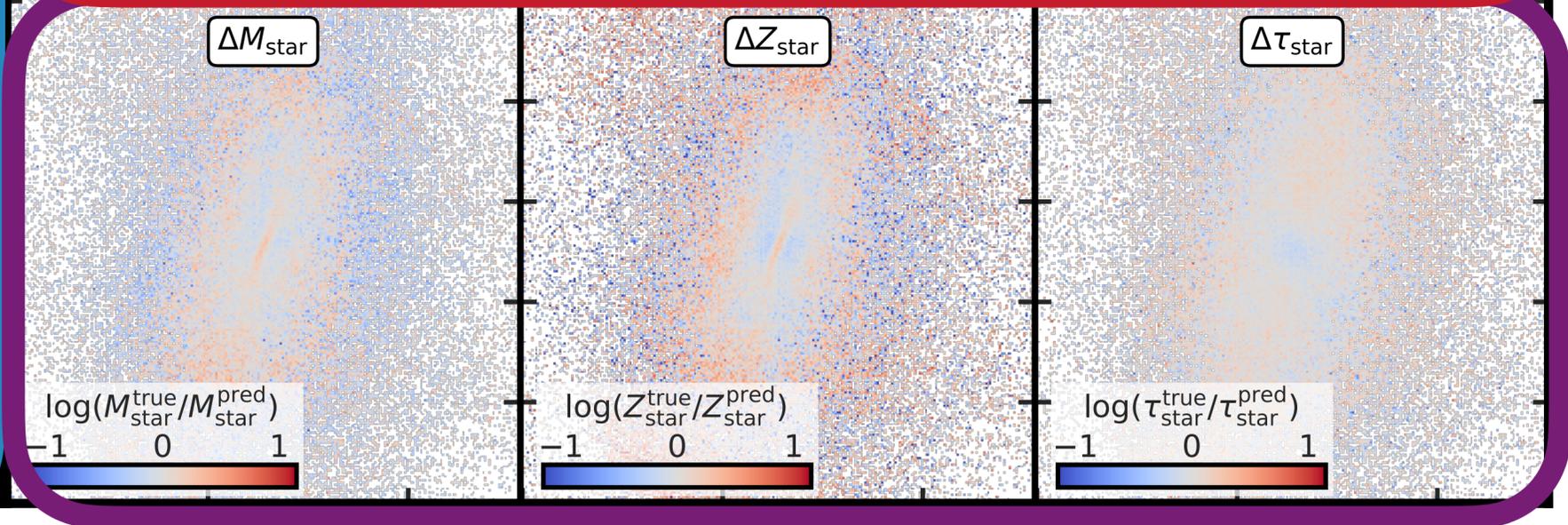
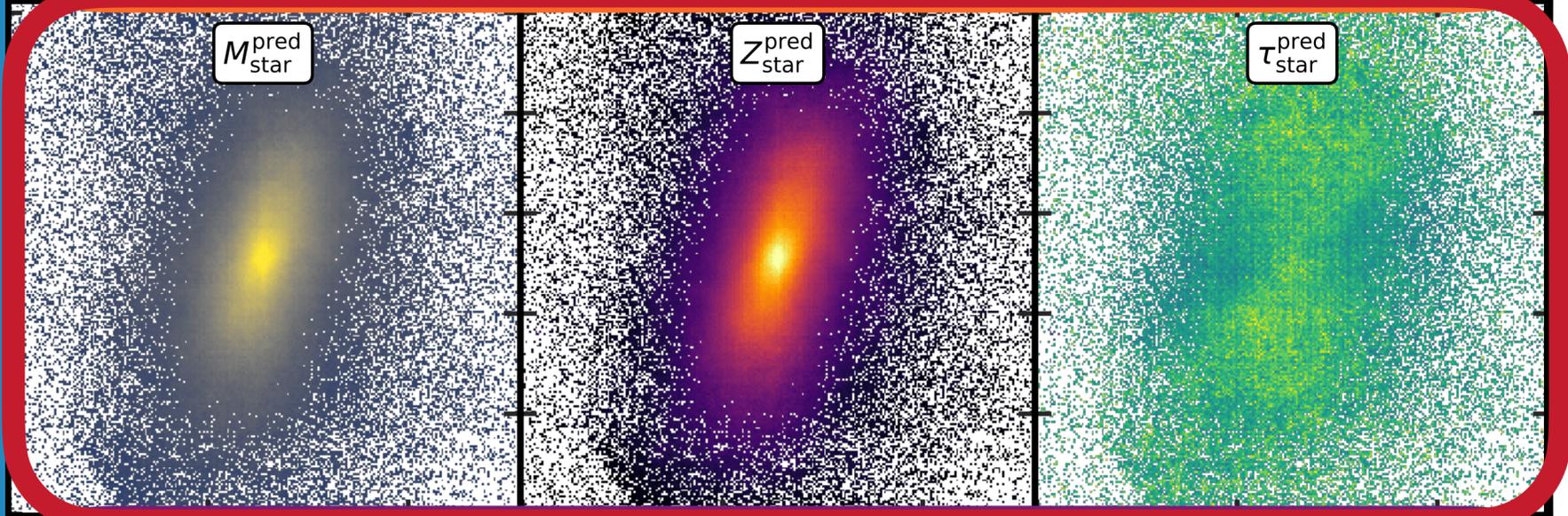
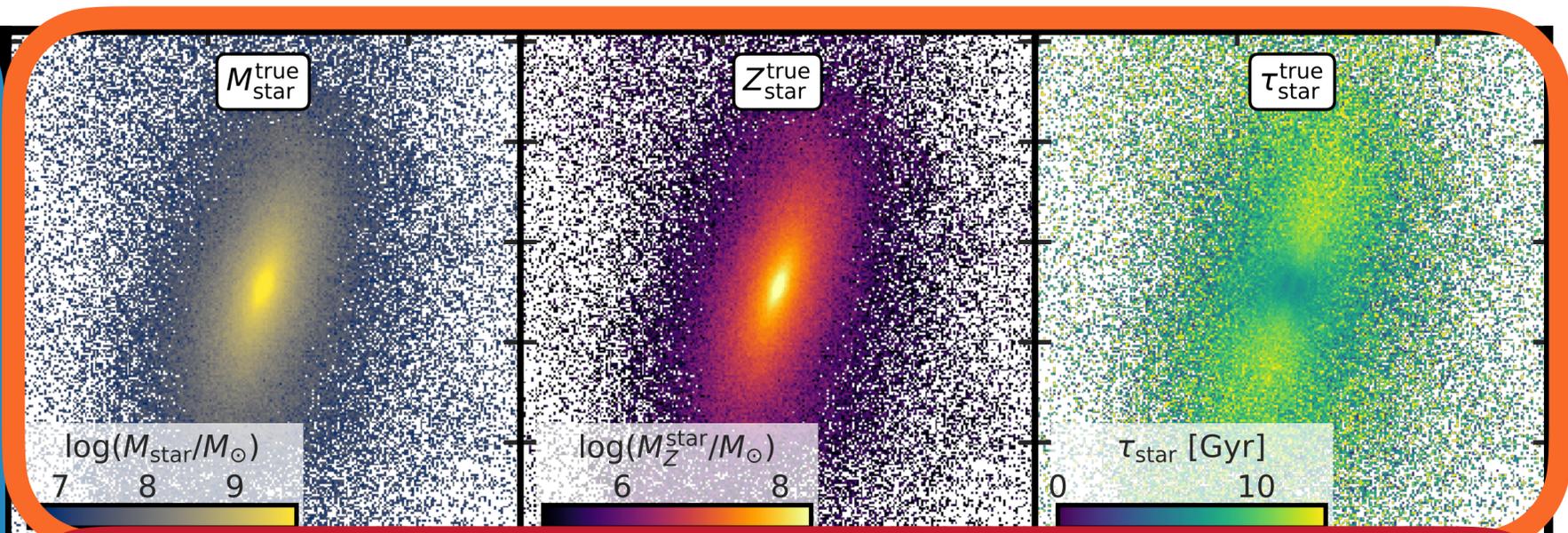
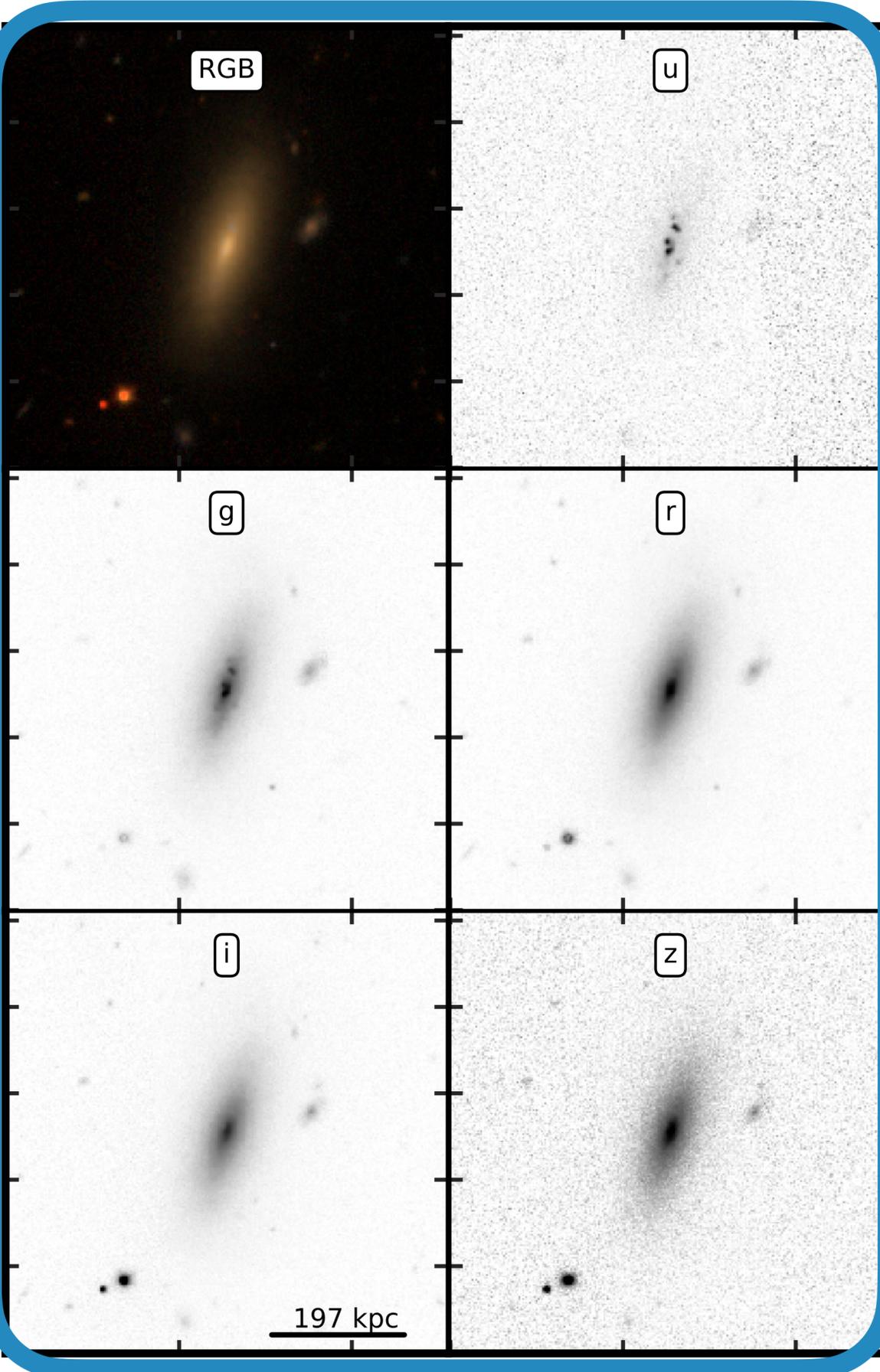
Z star

Stellar mass

RESULTS



Input

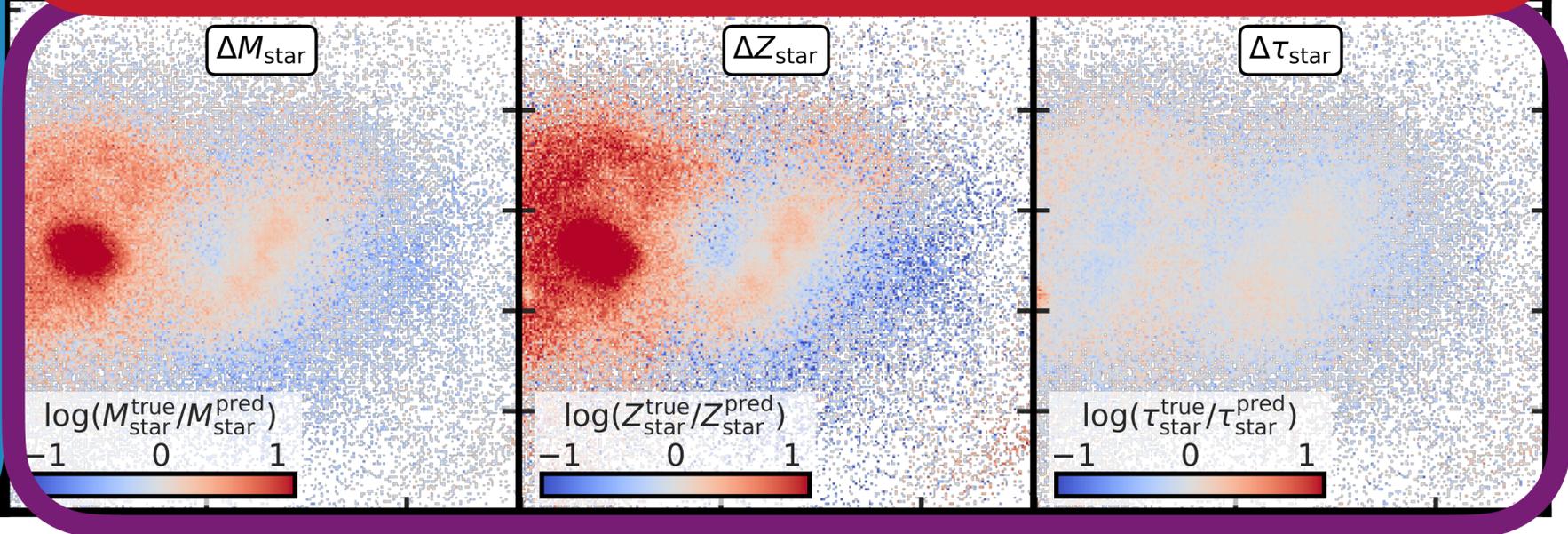
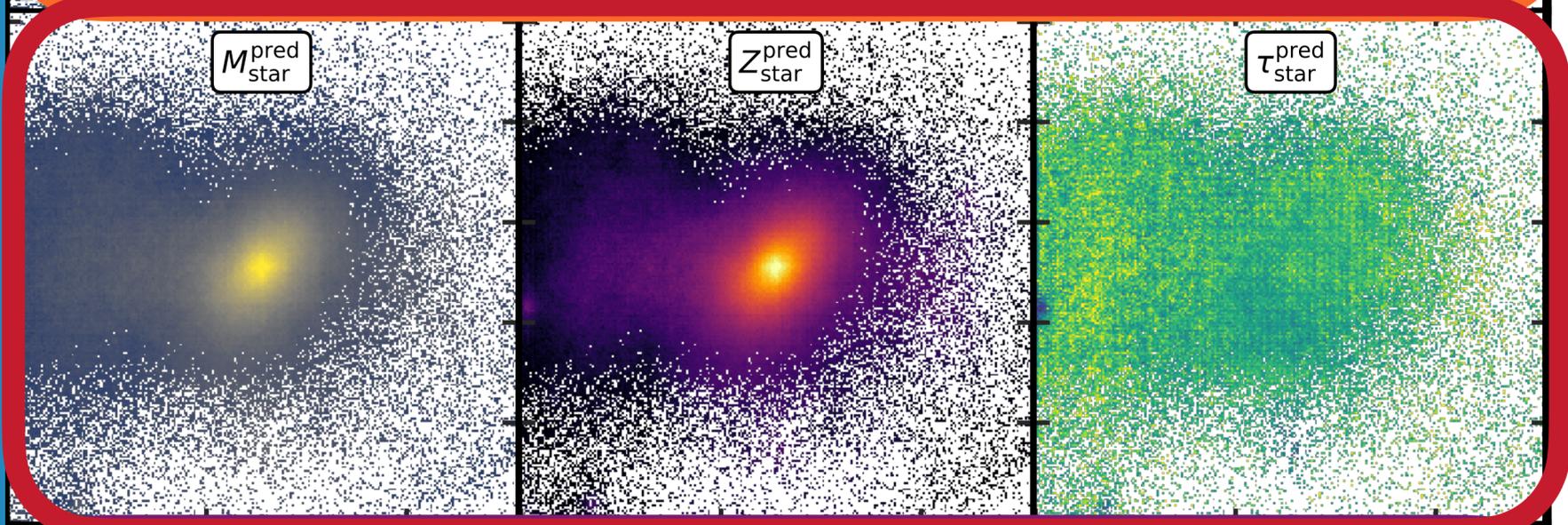
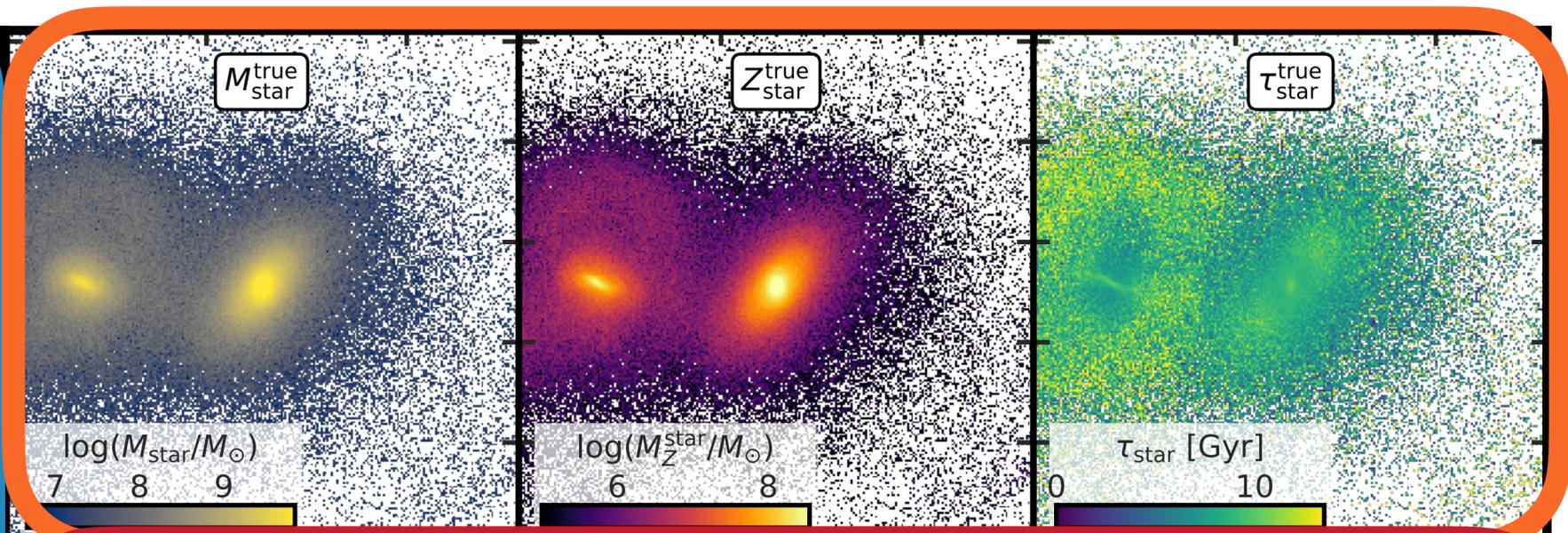
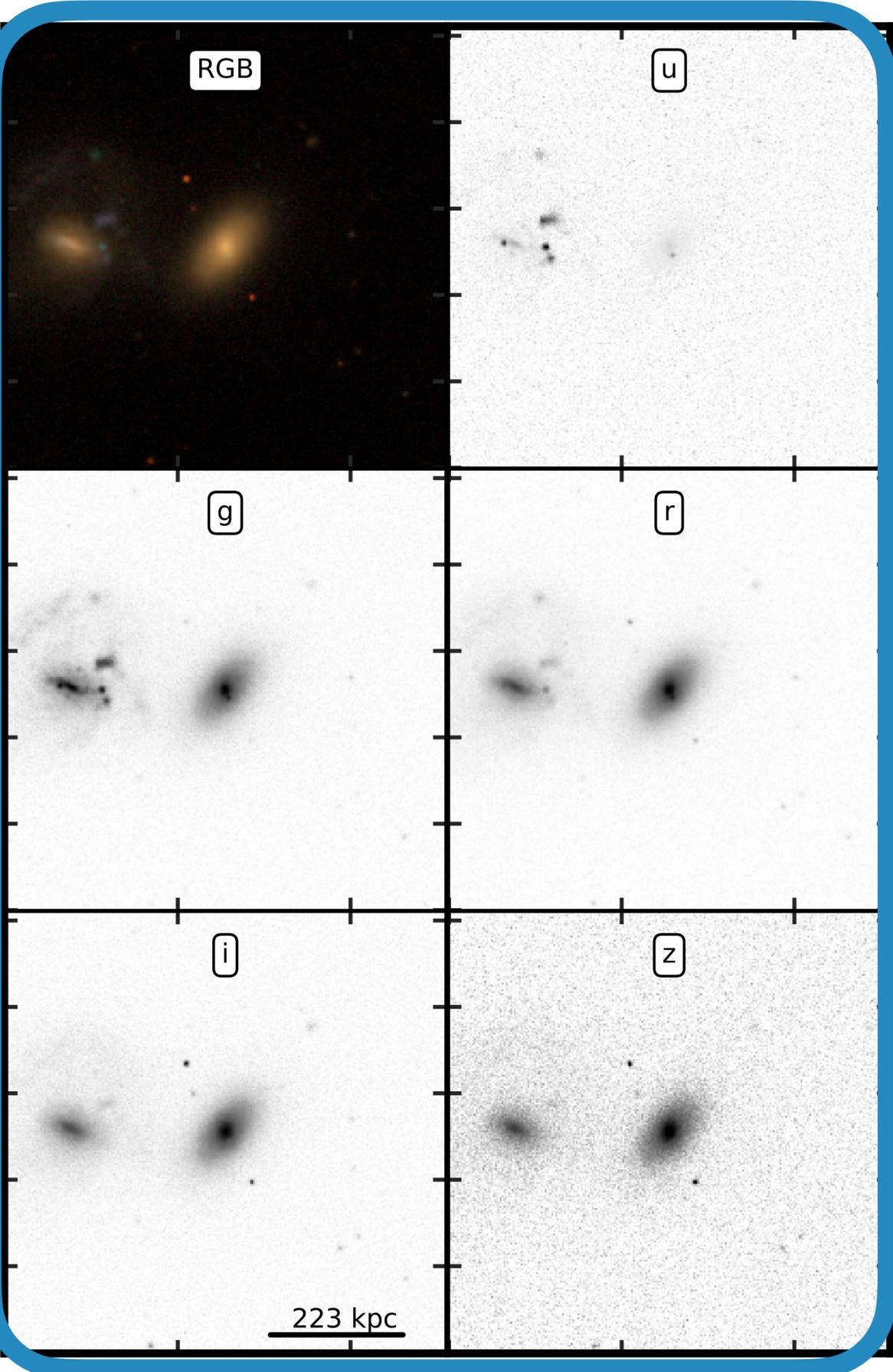


Truth

Prediction

Difference

Input

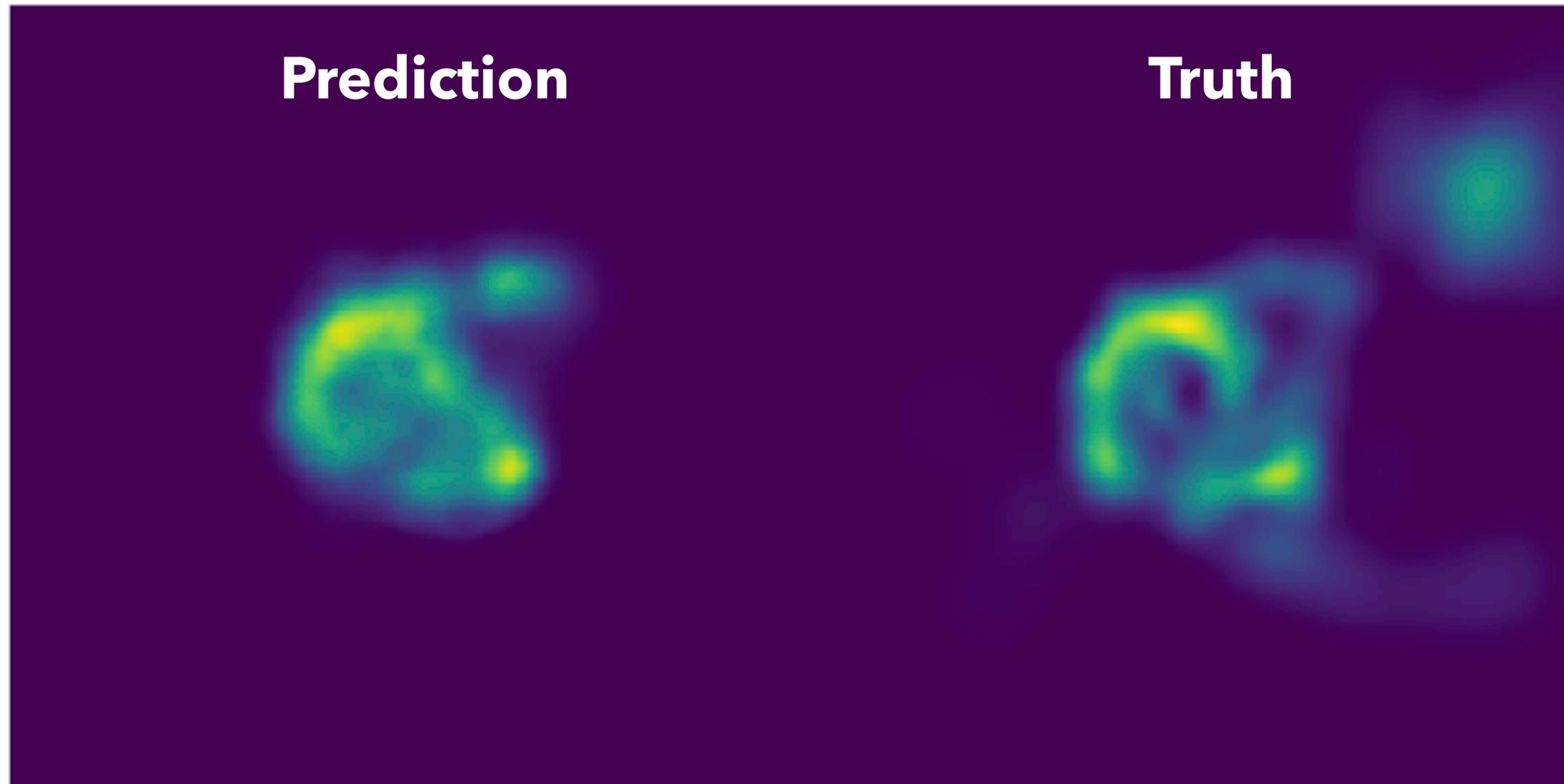


Truth

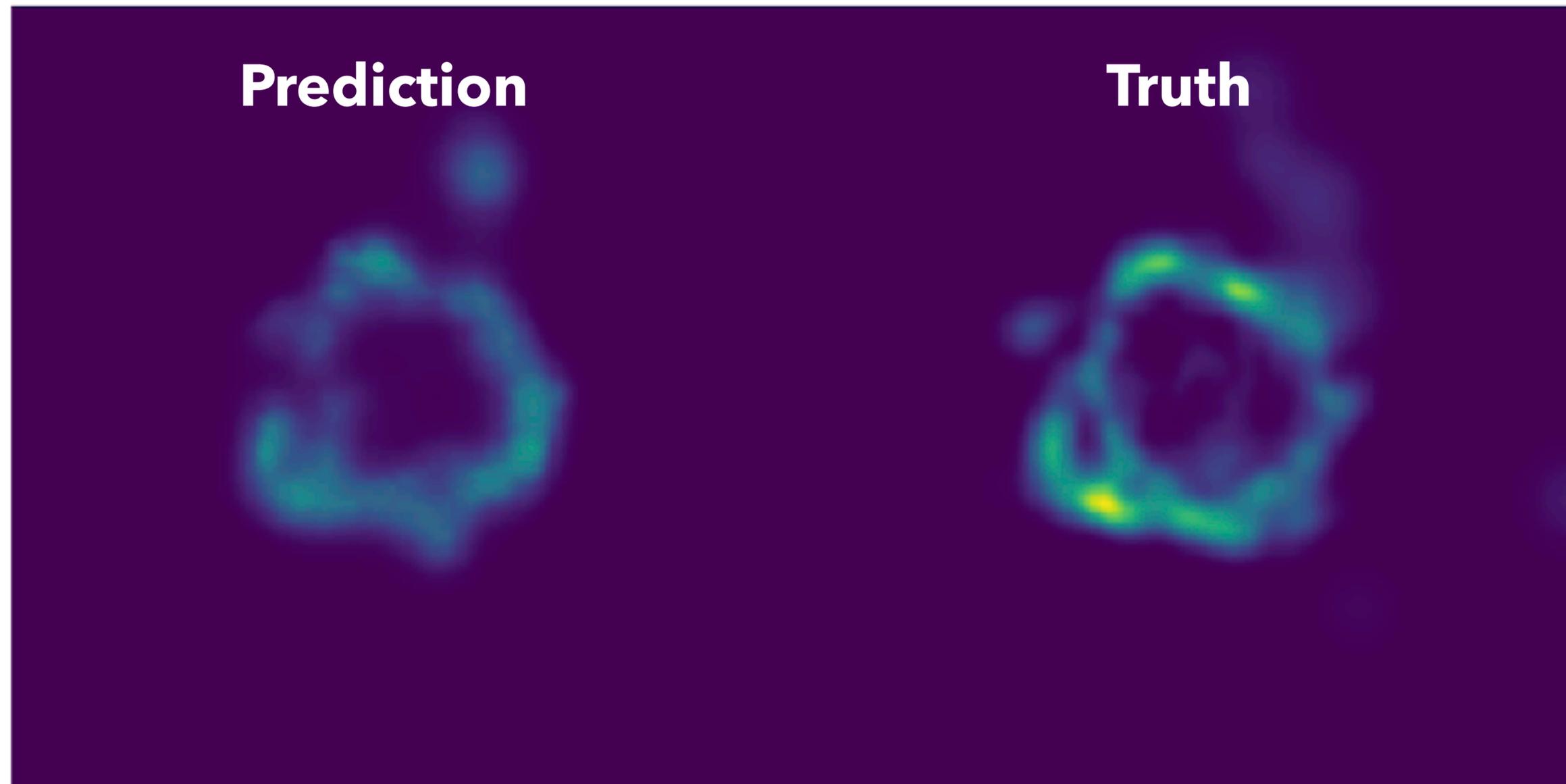
Prediction

Difference

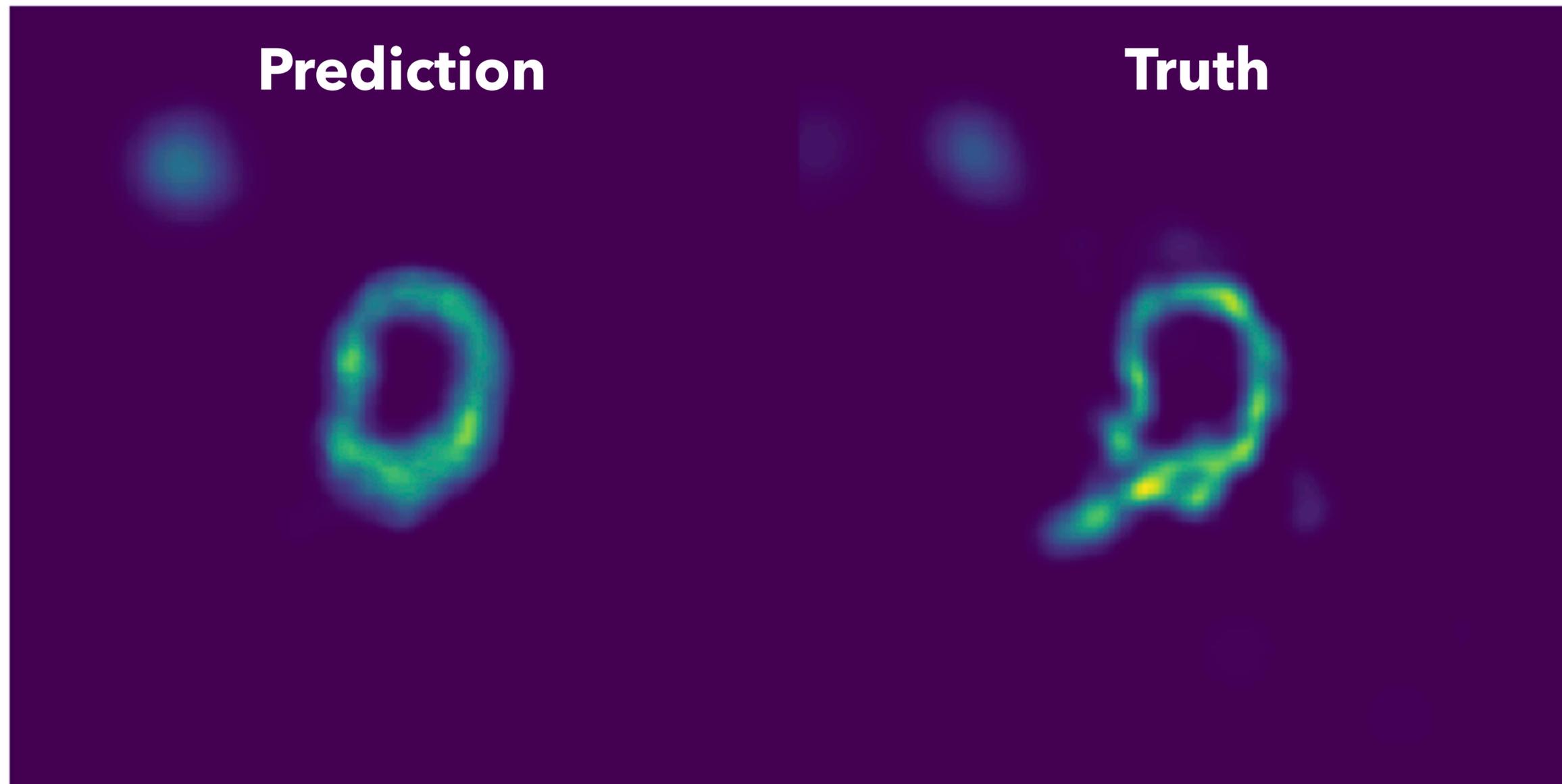
COMPARING TRUE AND PREDICTED SFR – 100TH QUANT.



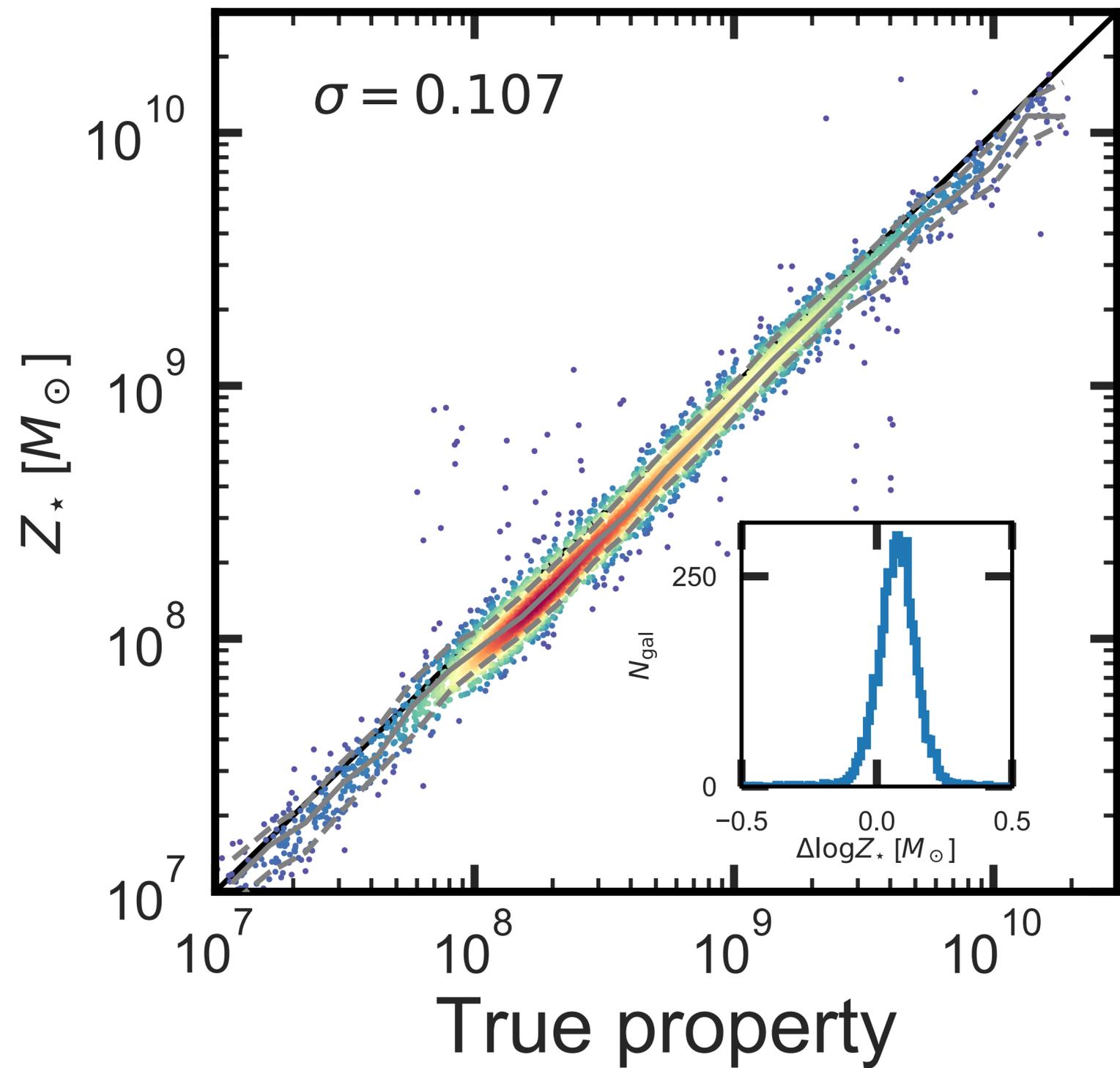
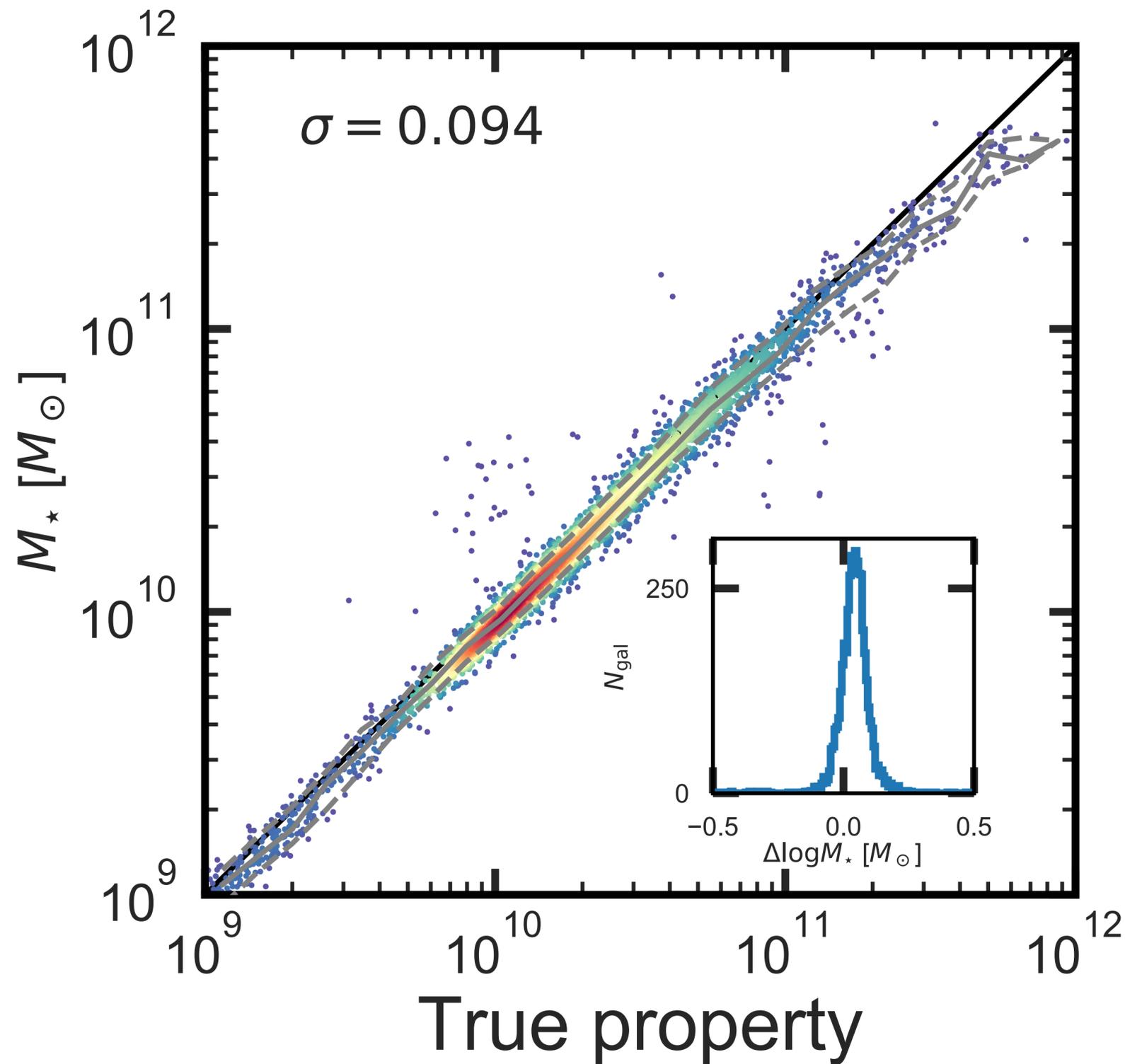
COMPARING TRUE AND PREDICTED SFR – 70TH QUANT.



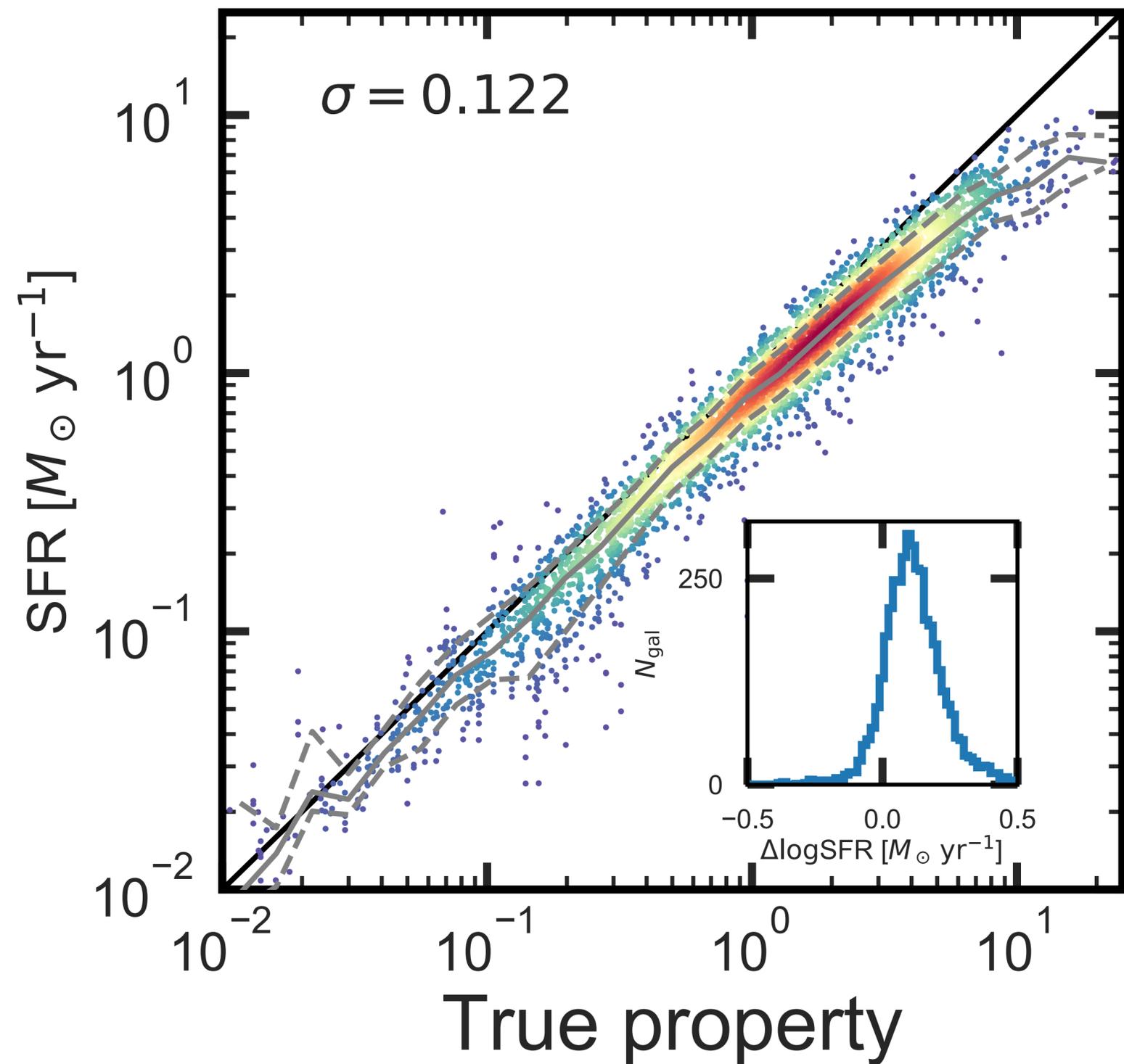
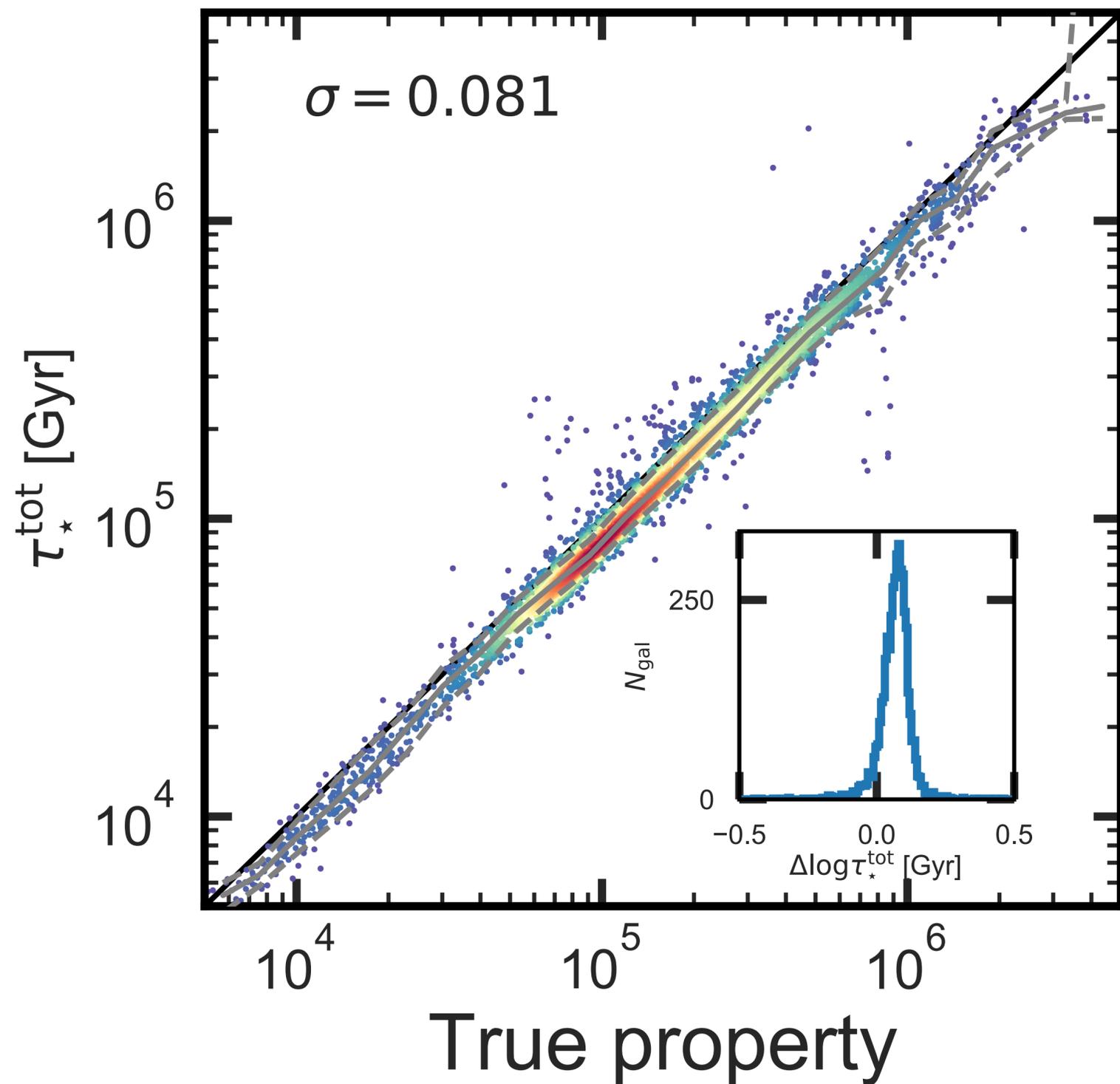
COMPARING TRUE AND PREDICTED SFR - 40TH QUANT.



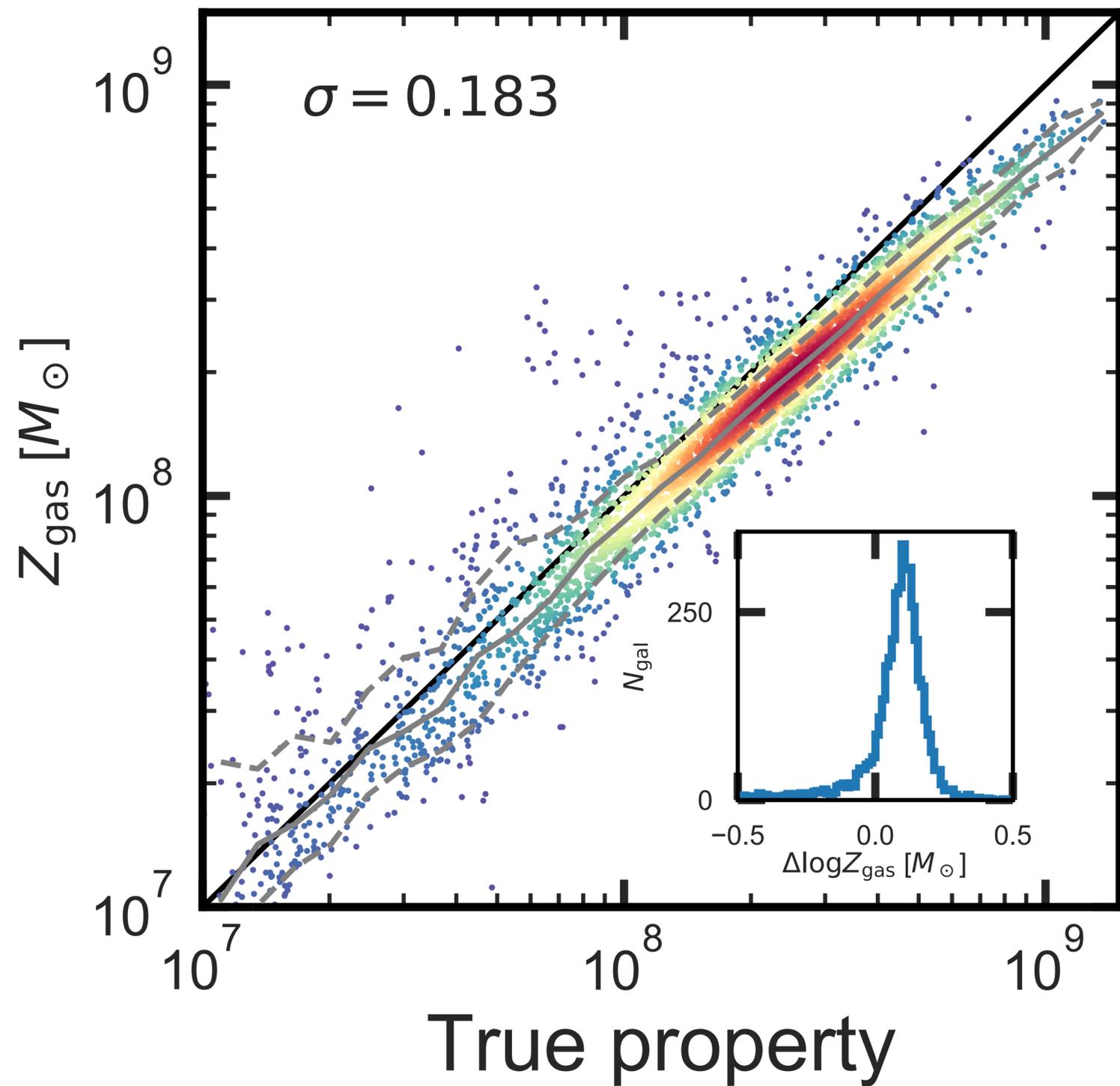
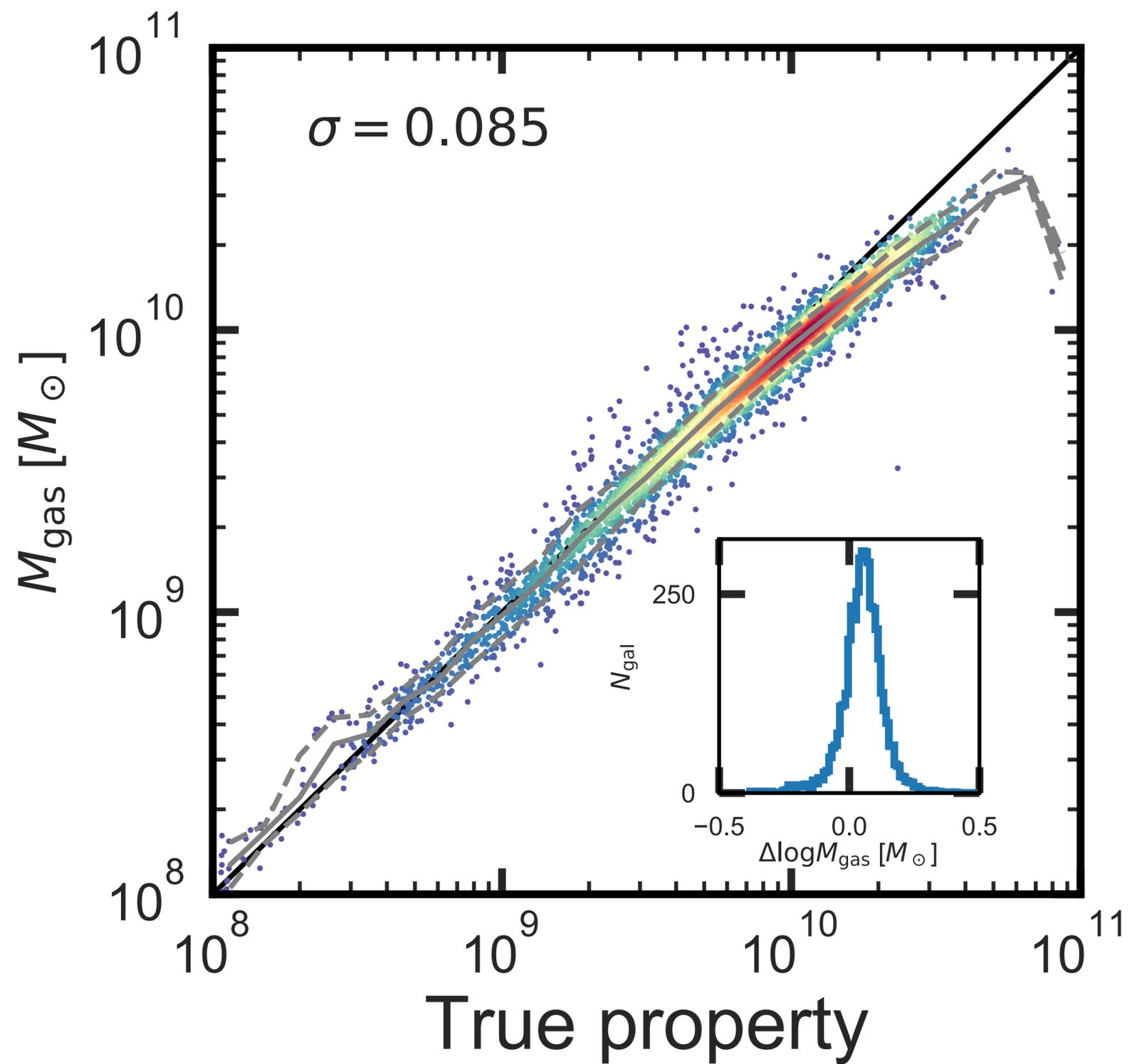
STELLAR PROPERTIES



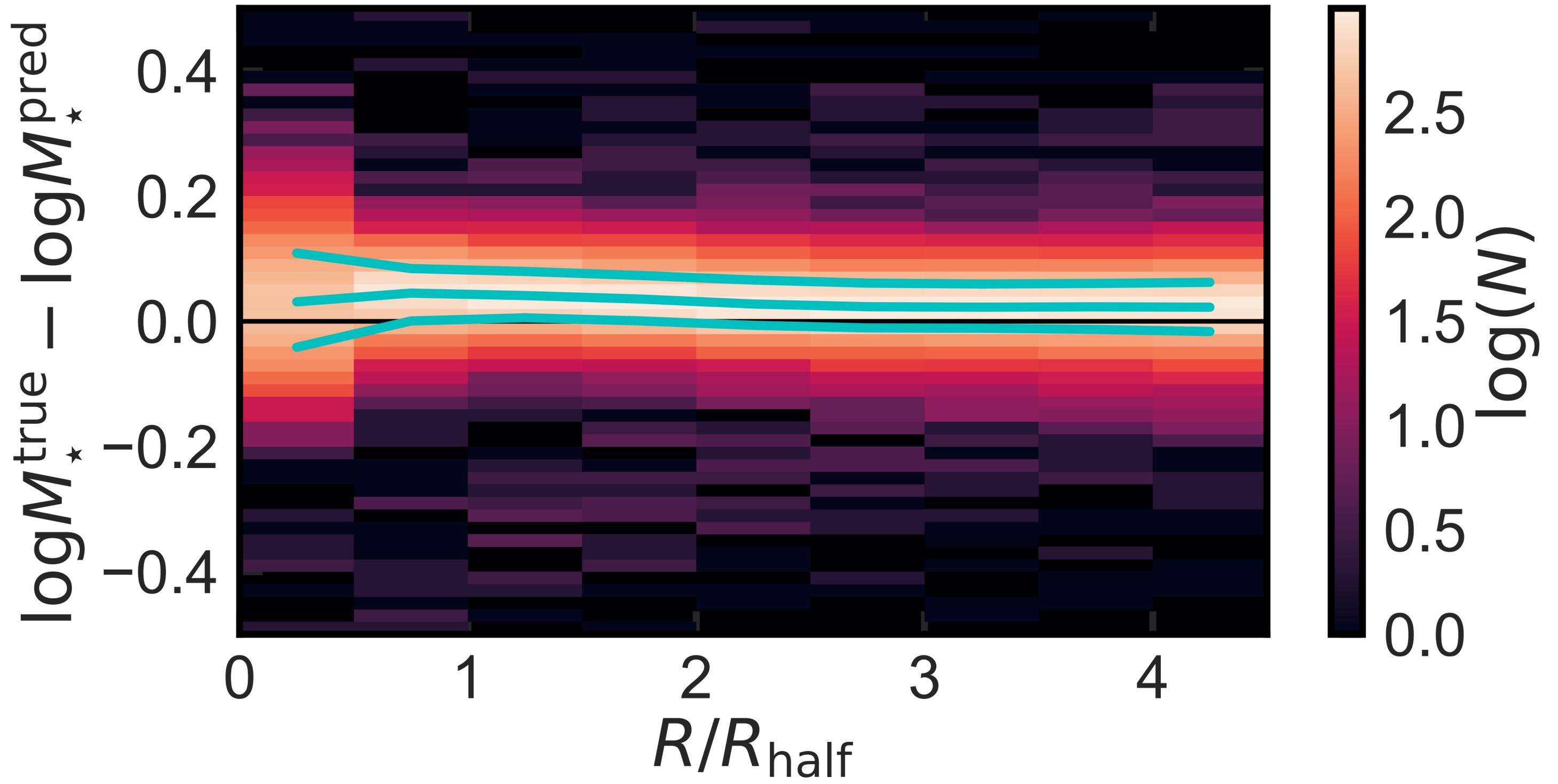
STELLAR PROPERTIES



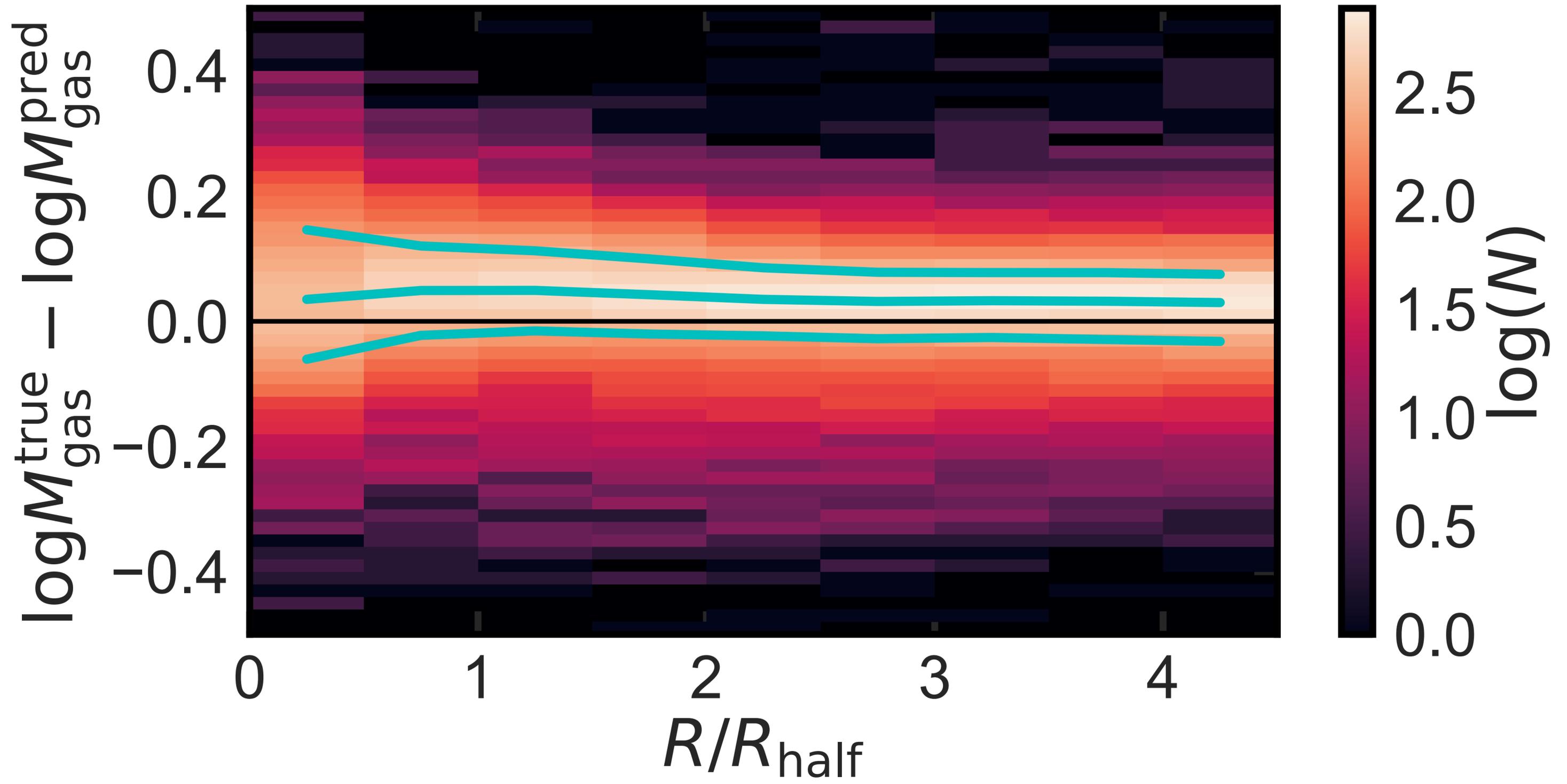
GASEOUS PROPERTIES



RADIAL SYSTEMATICS?



RADIAL SYSTEMATICS?



SUMMARY

SDSS (MOCK) *U,G,R,I,Z* IMAGES CONTAIN ENOUGH INFORMATION TO *PREDICT* PHYSICAL PROPERTIES OF GALAXIES ON A *PIXEL-BY-PIXEL* BASIS

NEXT STEPS: REAL LIFE APPLICATION

USE PICASSO ON REAL SDSS IMAGES WITH SDSS MANGA, SAMI, OR CALIFA AS TRAININGS SAMPLE

NEXT STEPS:

PROOF-OF-CONCEPT WORKS

QUANTIFY WHAT IS LEARNED: MORPHOLOGY OR COLOR?

QUANTIFY DEPENDENCE ON:

- ▶ **IMAGE RESOLUTION** (STABLE AGAINST FACTOR 2/4 LOWER RES)
- ▶ **TRAINING SET SIZE**
- ▶ **NUMBER OF INPUT BANDS**

REAL LIFE APPLICATION: IFU SURVEY DATA (E.G. MANGA)

RELEASE IT AS READY-TO-USE TOOL?