The challenges of Photometric redshifts with large imaging surveys

S. Arnouts
Laboratoire d'Astrophysique de Marseille

Deep Learning — Treyer, Ait-Ouahmed (LAM), Pasquet (Tetis), Bertin (IAP), Lin, Fouchez (CPPM)

DEEPDIP.AI

SED fitting — Picouet, Ilbert (LAM), Sawicki (Halifax) Desprez (Geneve)

Photometric redshifts

SED fitting

Priors on the SED of galaxies
physical knowledge

Supervised Machine Learning

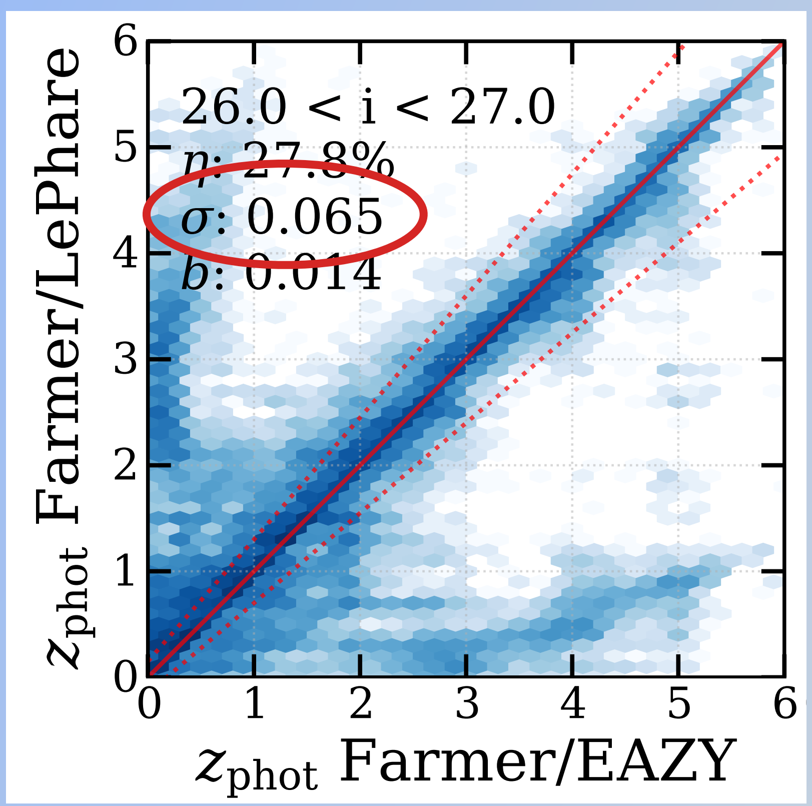
non linear mapping between inputs & z

Artificial Neural Network, k-Nearest NeigborsN,

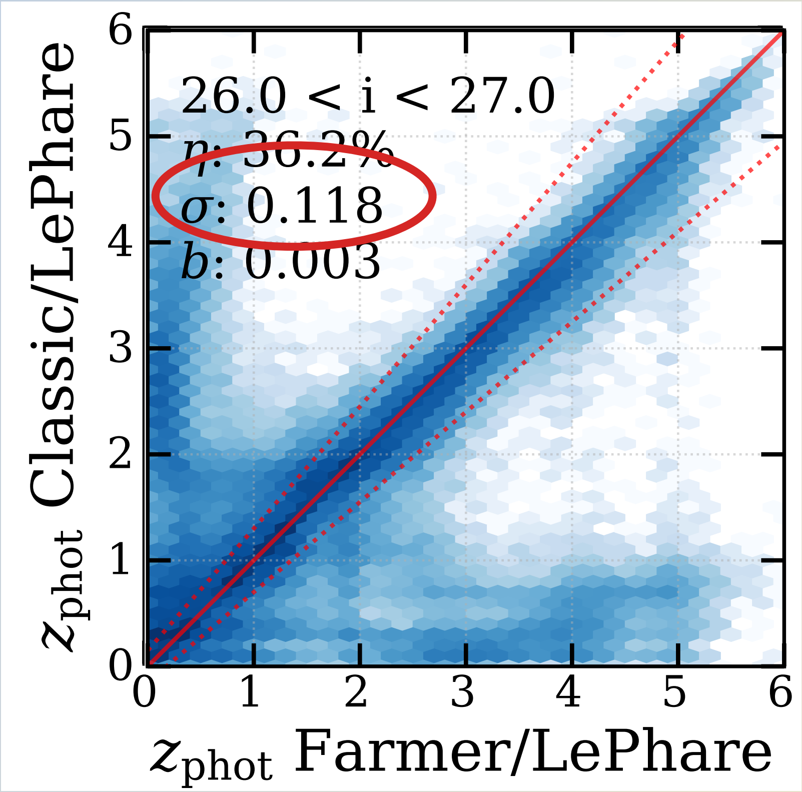
Random forest, SOM , …

input informations based on extracted features. Photometry is a critical step …

2 Ph-z codes / 1 photometry



2 photometry / 1 Ph-z code



COSMOS 2020
(Weaver+21)

—> Larger uncertainties due to photometric extractions than Photo-z codes

Photometric redshifts

SED fitting
Priors on the SED of galaxies
physical knowledge

Supervised Machine Learning
non linear mapping between inputs & z
Artificial Neural Network, k-Nearest NeigborsN,
Random forest, SOM , …

input informations based on extracted features.

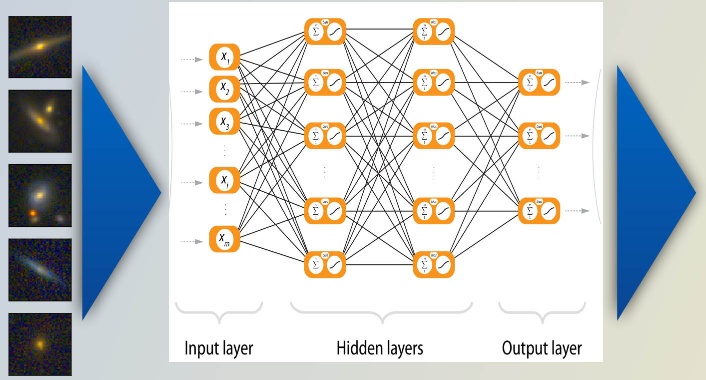
skip
photometric
step

use of Multi-band images with Convolutional Neural Networks

— **no feature extraction.**
Works at the pixel level !
exploits all the informations

(SB, sizes, inclinations, gradients, ….)

DNN



z
PDF(z)

— Now under reach thanks to large spec-z training set & GPU power

— **Hoyle+16** 60x60 jpeg RGBa images encoding (i-z,r-i,g-r, r mag), output: PDF

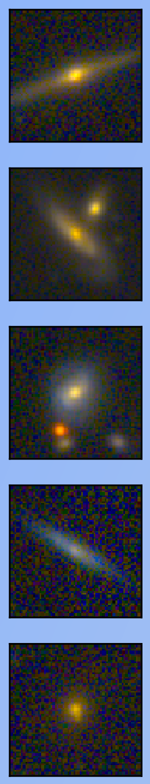
— **d'Isanto & Polester+18**: 28x28 ugriz fits images, output: PDF with Gaussian Mixture model

—> **Pasquet+19, Campagne19 , Menou19, Schuld+21 (netZ) …**

Convolutional Neural Network (Lecun+ 98)

https://gricad-gitlab.univ-grenoble-alpes.fr/talks/fidle

Input



64x64 ugriz

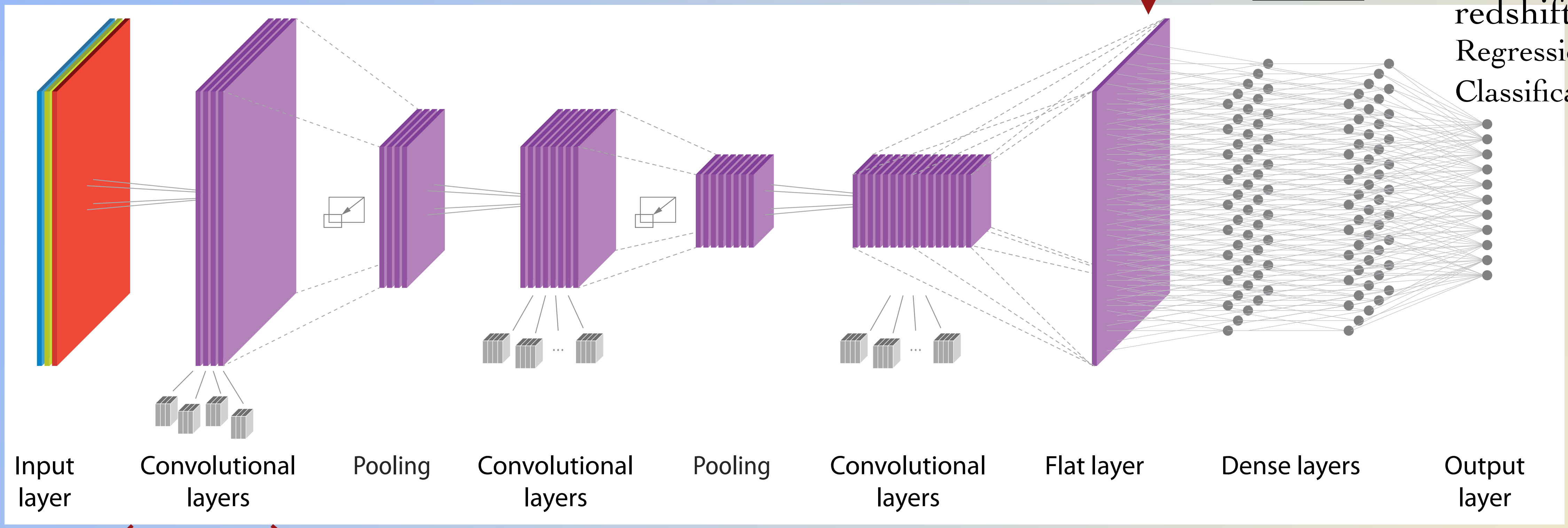
convolution blocks

Input DNN Feature Maps

DNN

Output

redshift
Regression/
Classification



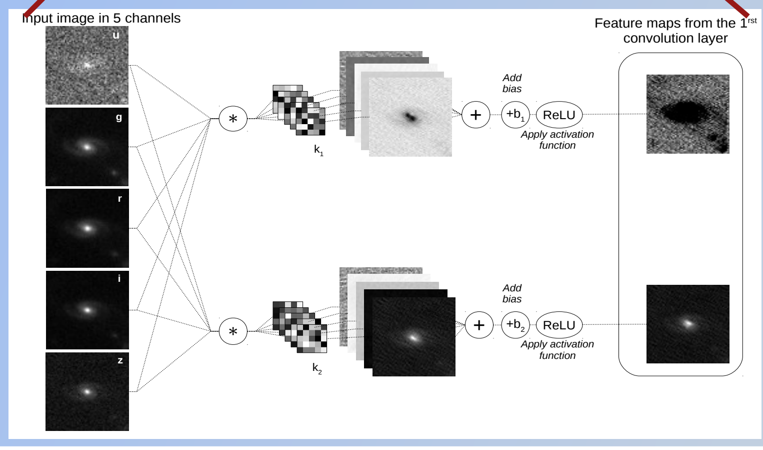
5 Conv Blocks

348 Feature Maps + E(B-V)

2x1024

classifier with 180 z

Pasquet+19

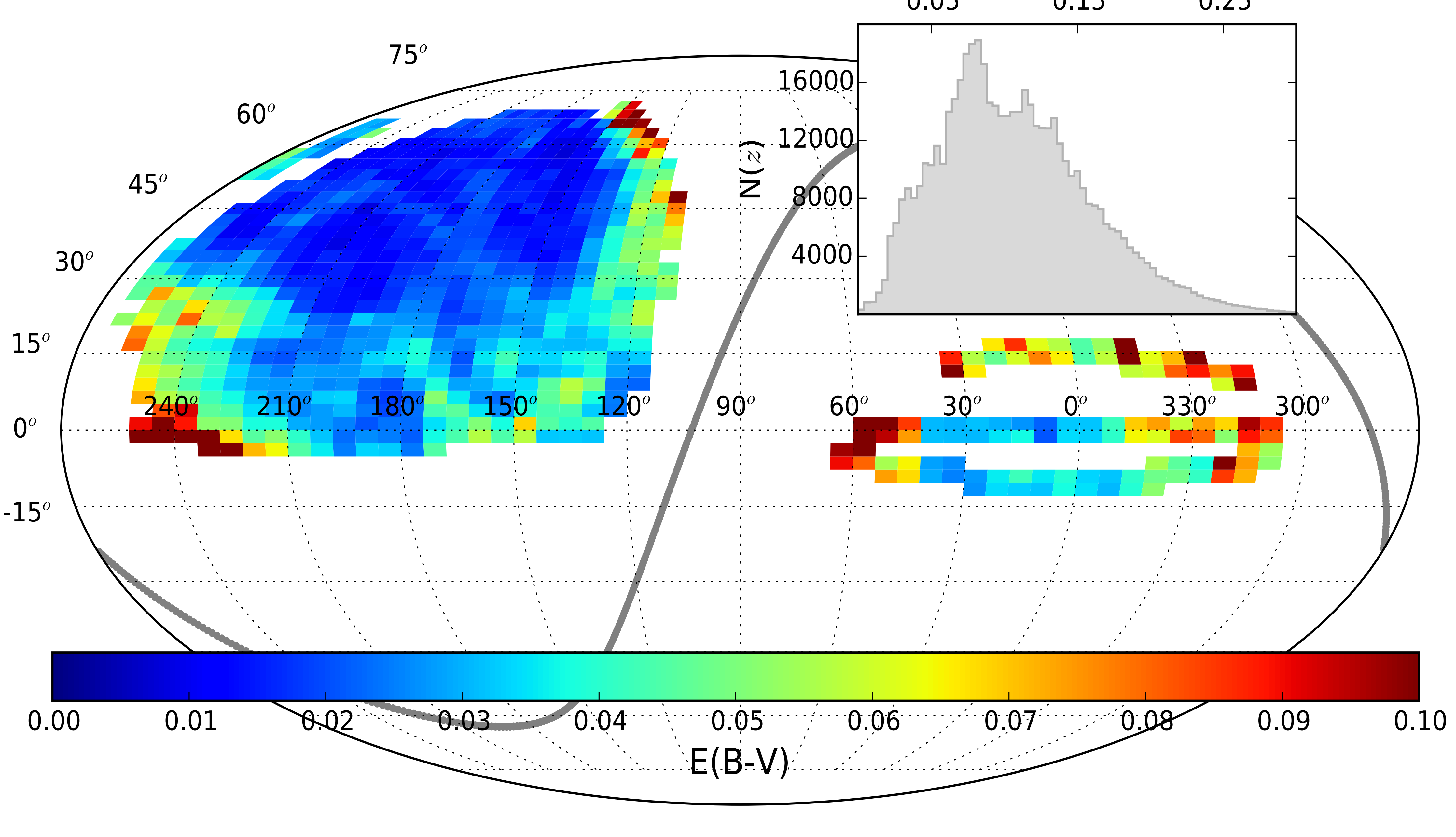


- CNN : several millions of parameters
  - -> training all weights/bias/kernels with back-propagation
  - -> cost function : cross-entropy
  - -> 5 CV with 80% train / 20% validation

Photometric redshifts with Deep CNN : SDSS

Pasquet+ 19

SDSS (DR12) : 516,000 galaxies with r<17.8



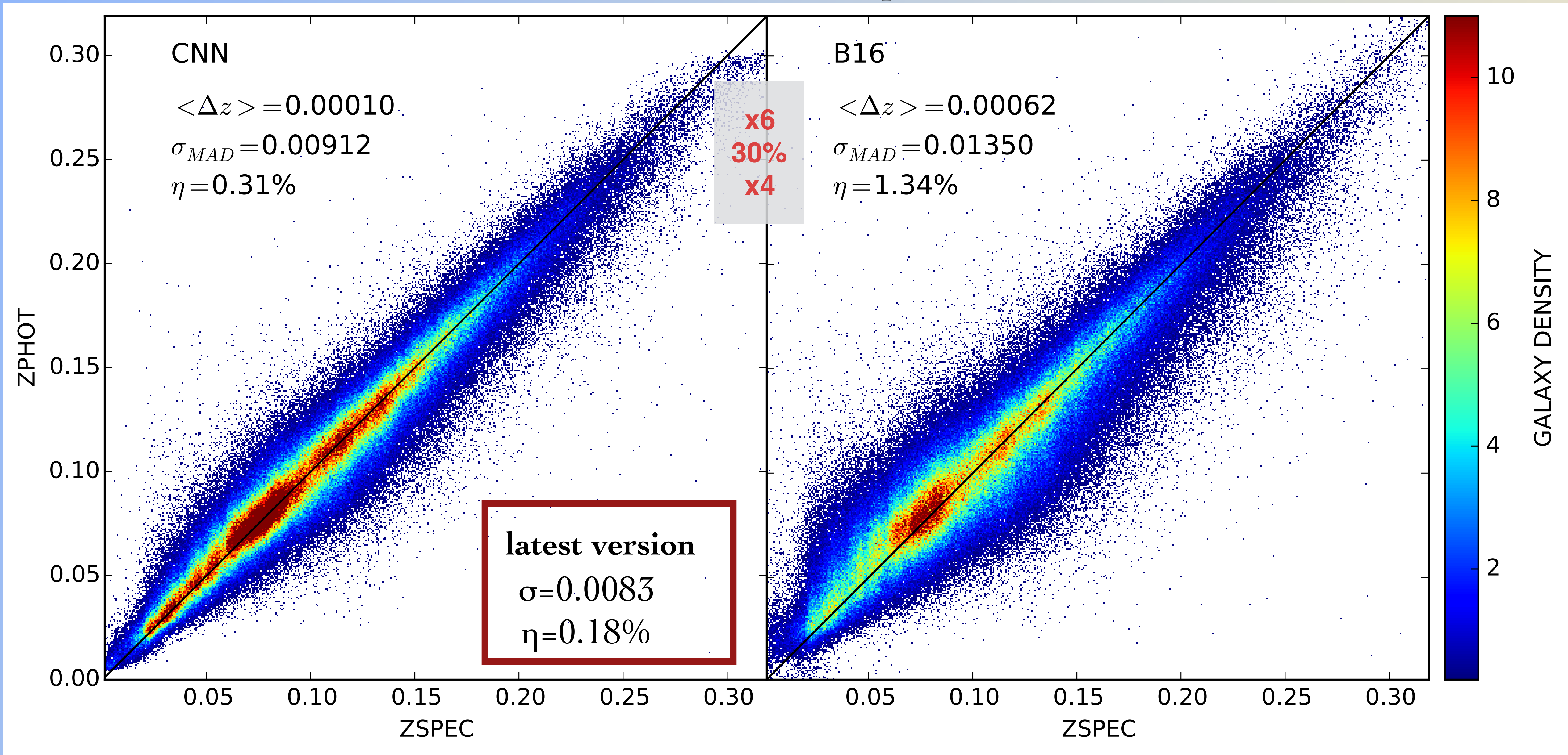
—> what photo-z accuracy can we get compared to other ML techniques ?

—> can we extract reliable PDF estimates ?

Photometric redshifts with Deep CNN : SDSS

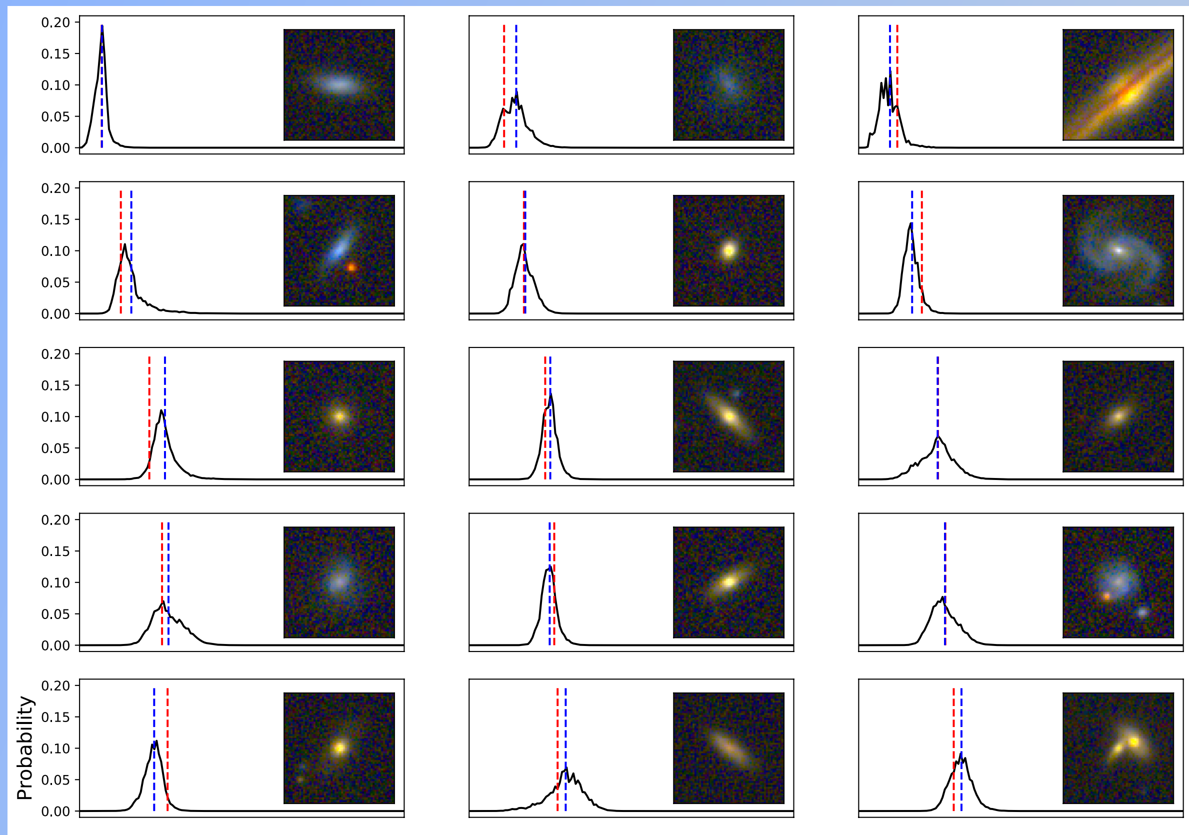
Pasquet+ 19

Comparison with Beck+16 based on k-NN



—> Better performance than the latest SDSS photo-zs

Photometric redshifts with Deep CNN : SDSS



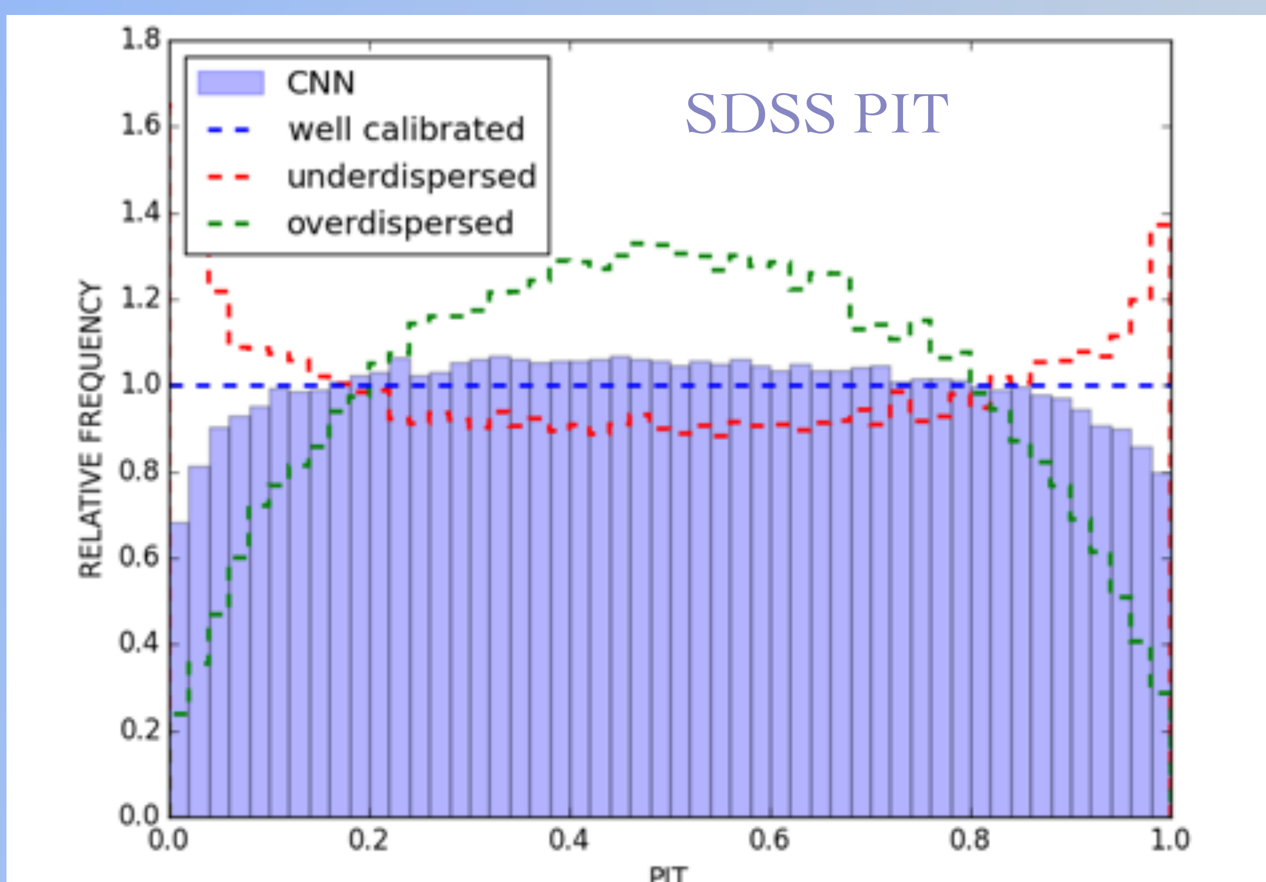
Pasquet+ 19

—> **PDF evaluations :**

Probability Integral Transform **PIT**

$$\mathrm{PIT}_i = \int_{-\infty}^{z_i} PDF_i(z)\,dz$$

Polsterer+ 16



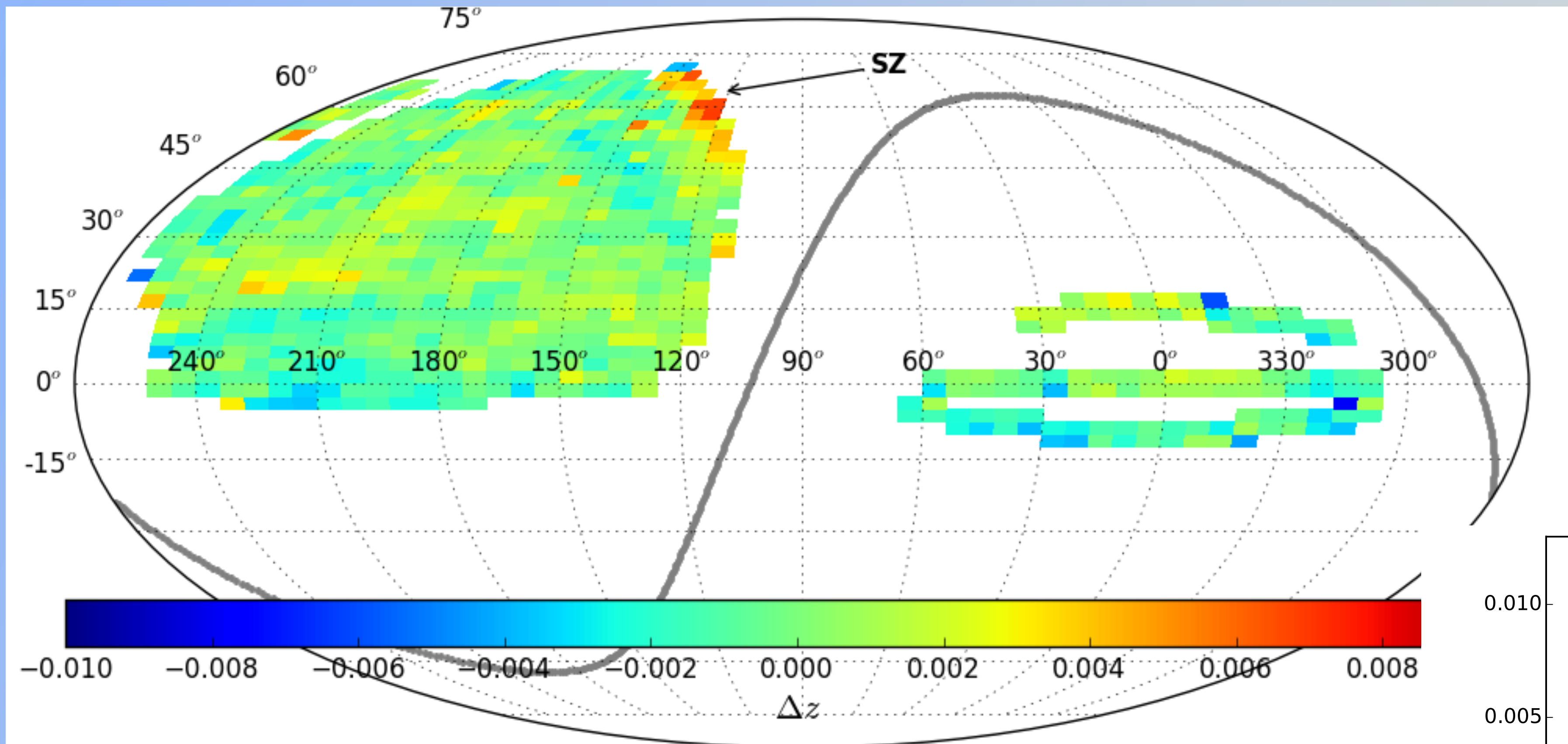
PDFs estimates well calibrated

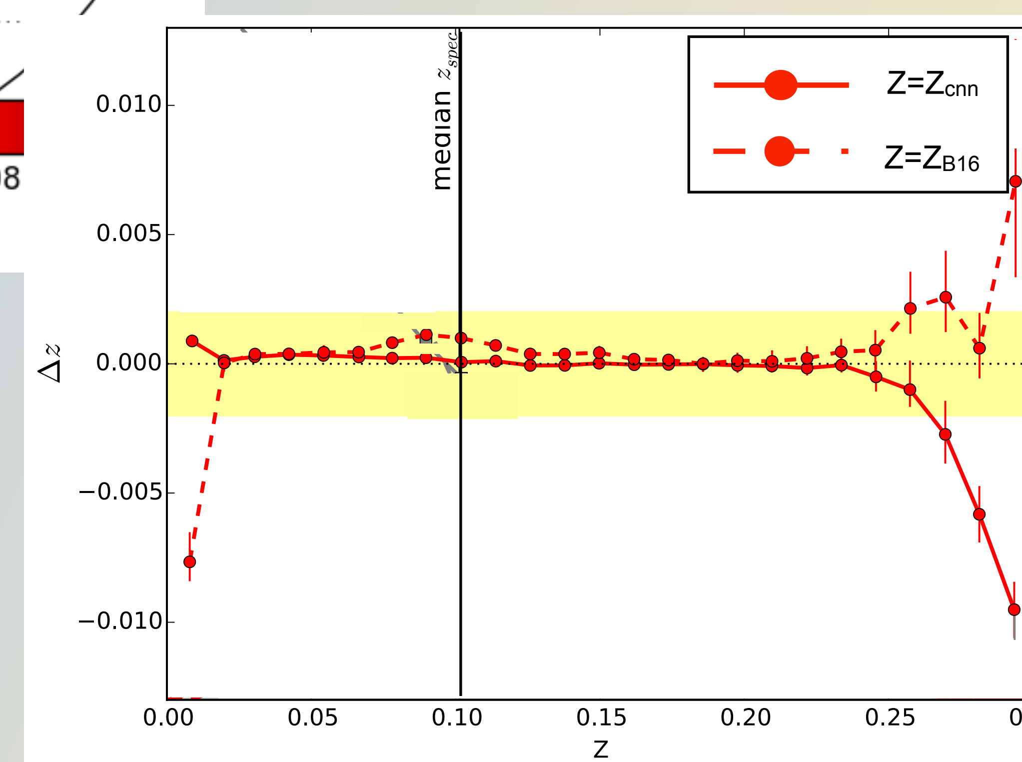
(PDF neither too broad nor too narrow)

-> good estimator of photo-z accuracy for single object

Photometric redshifts with Deep CNN : SDSS

—> Keep photo-z bias <Δz> under control over large area and Z range

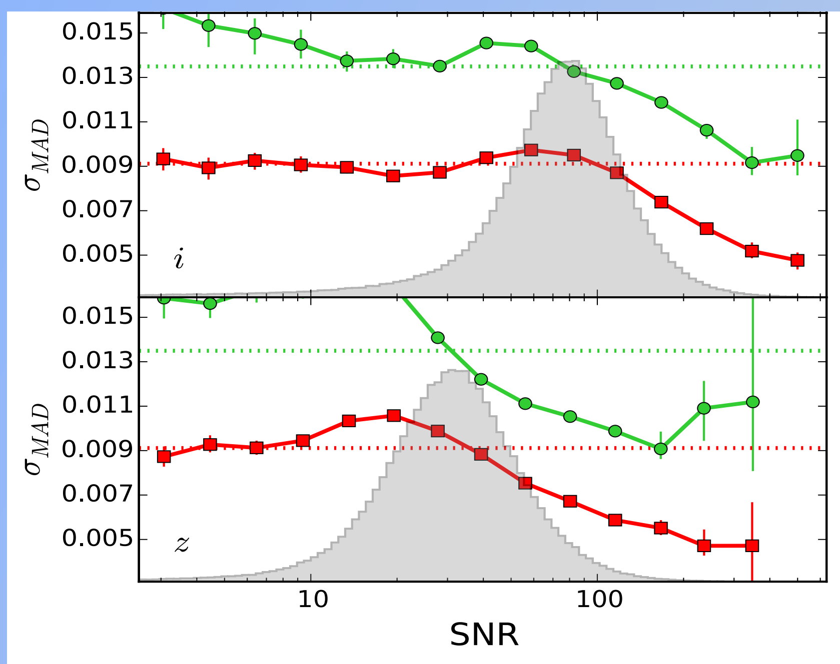




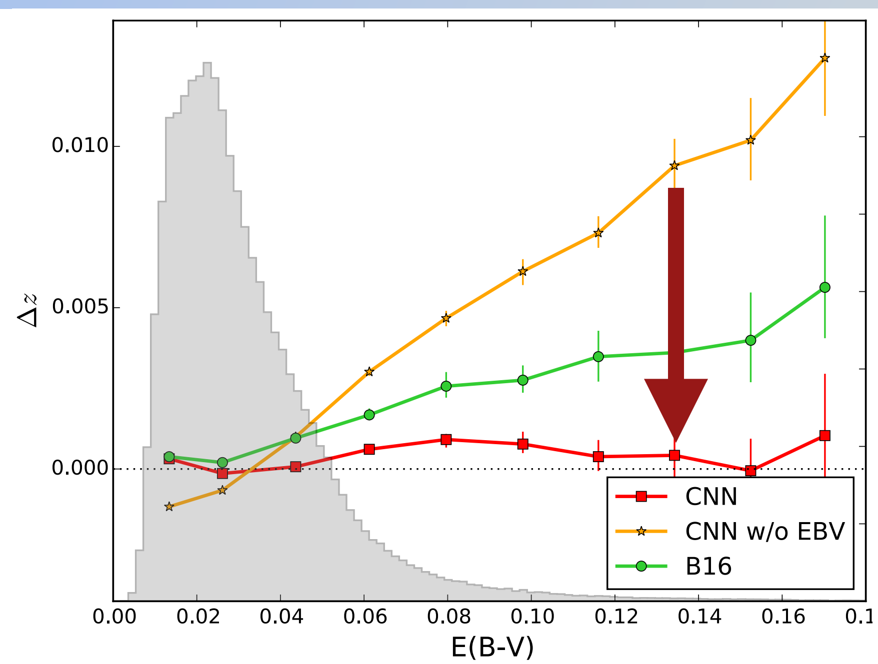
Pasquet+ 19

—> Main features

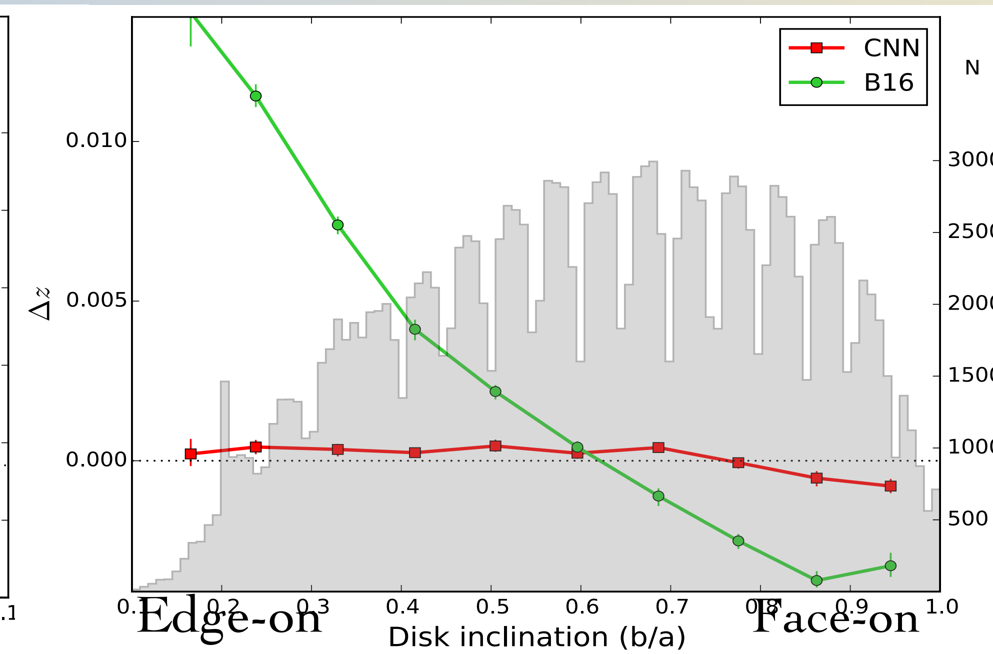
Accuracy vs SNR



bias vs E(B-V)



bias vs Gal. inclination



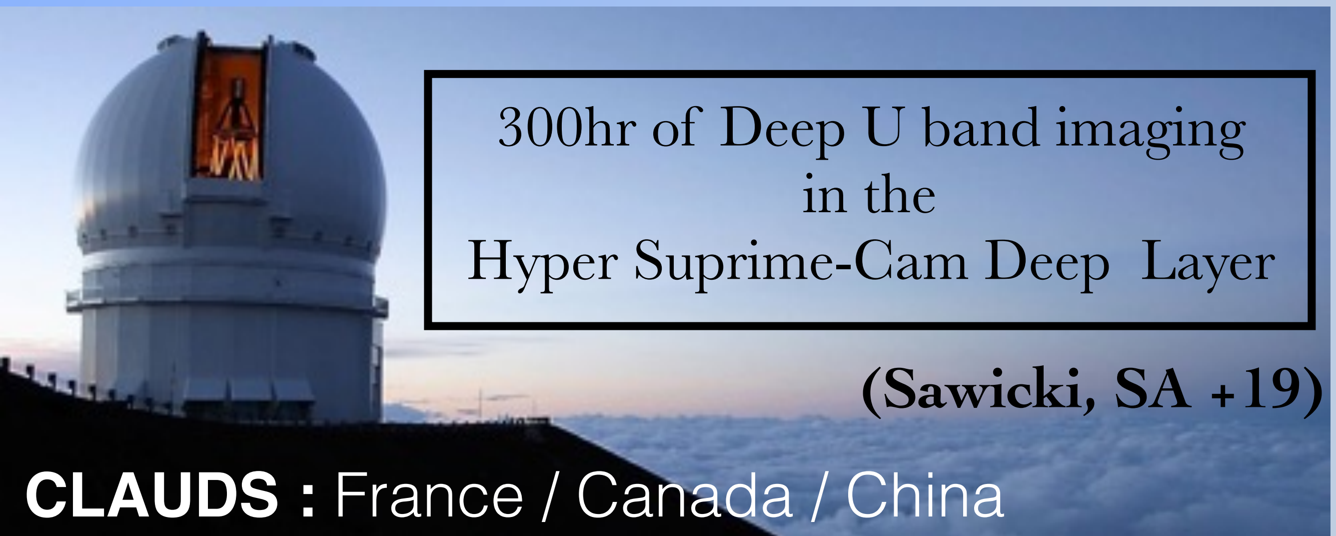
—> Training sizes

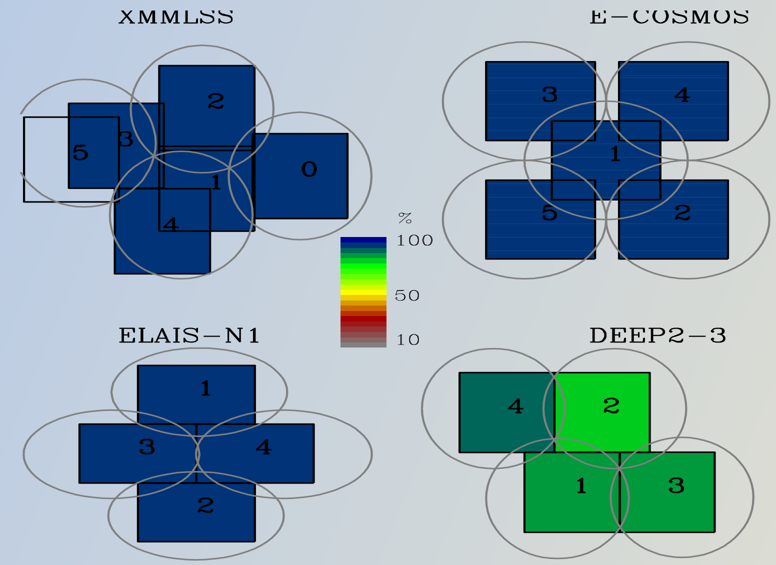
Training with 50% of the dataset★	250,000	252,500	0.00007	0.00910	0.29	0.00672
Training with 20% of the dataset	99,001	385,970	-0.00001	0.00914	0.30	0.00677
Training with 2% of the dataset	10,100	434,228	-0.00017	0.01433	1.26	0.01009

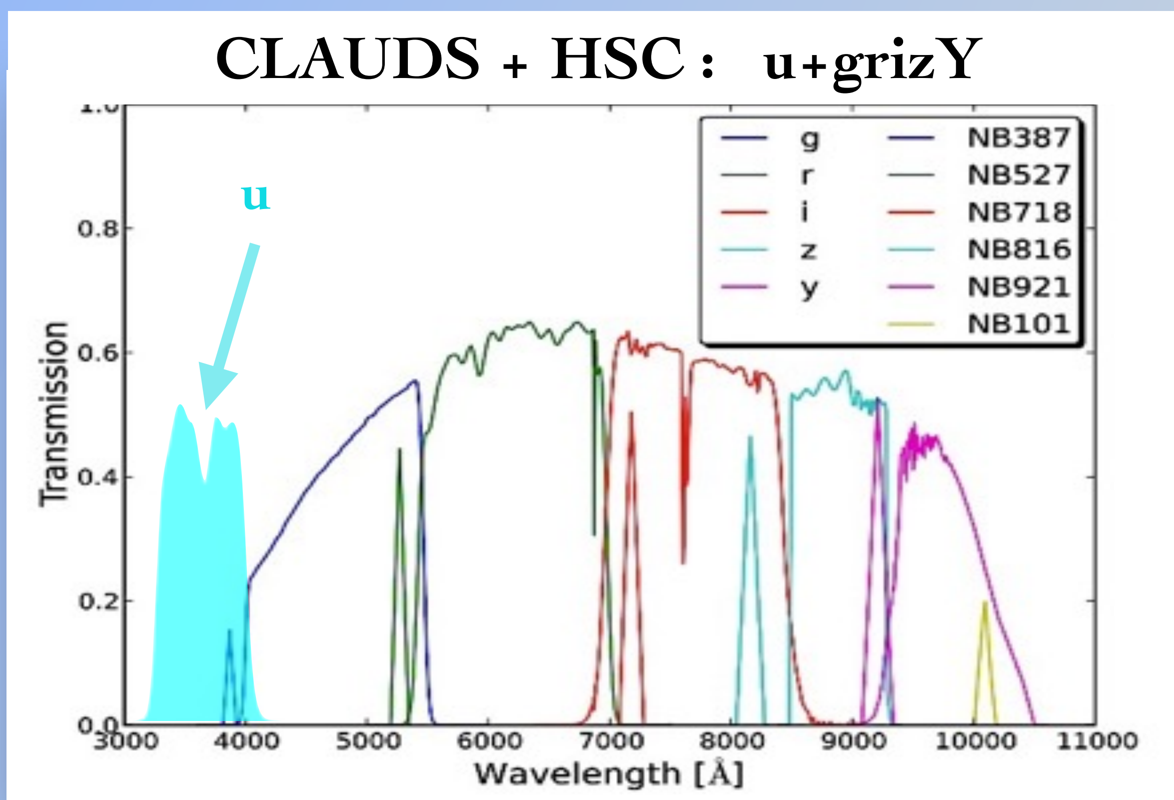
—> training with only 2% sample as well as Beck+16 k-NN results

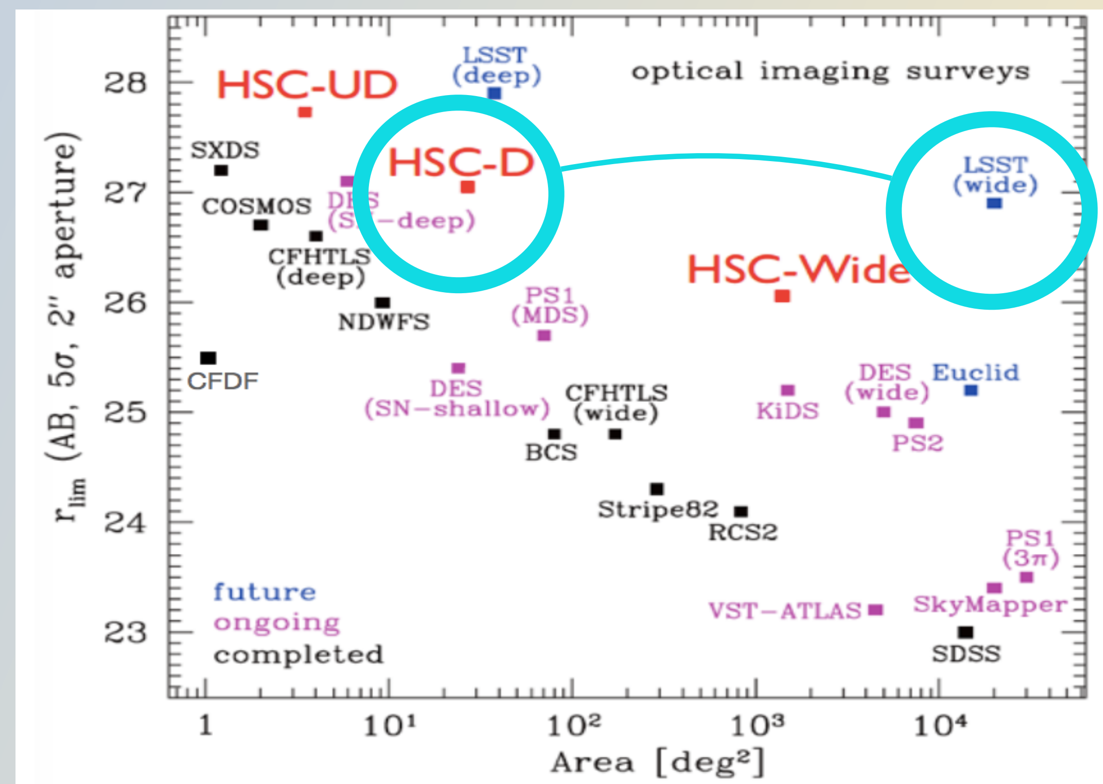
CNN photo-z are highly competitive and do not required large training set

Prepare LSST with CLAUDS - HSC Deep Survey









HSC Deep : **28 deg² at r~27** **+NB** (in progress)
CLAUDS : **25 deg² at u~27** (done)

CLAUDS+HSC deep
a unique dataset until LSST

CLAUDS - HSC Deep

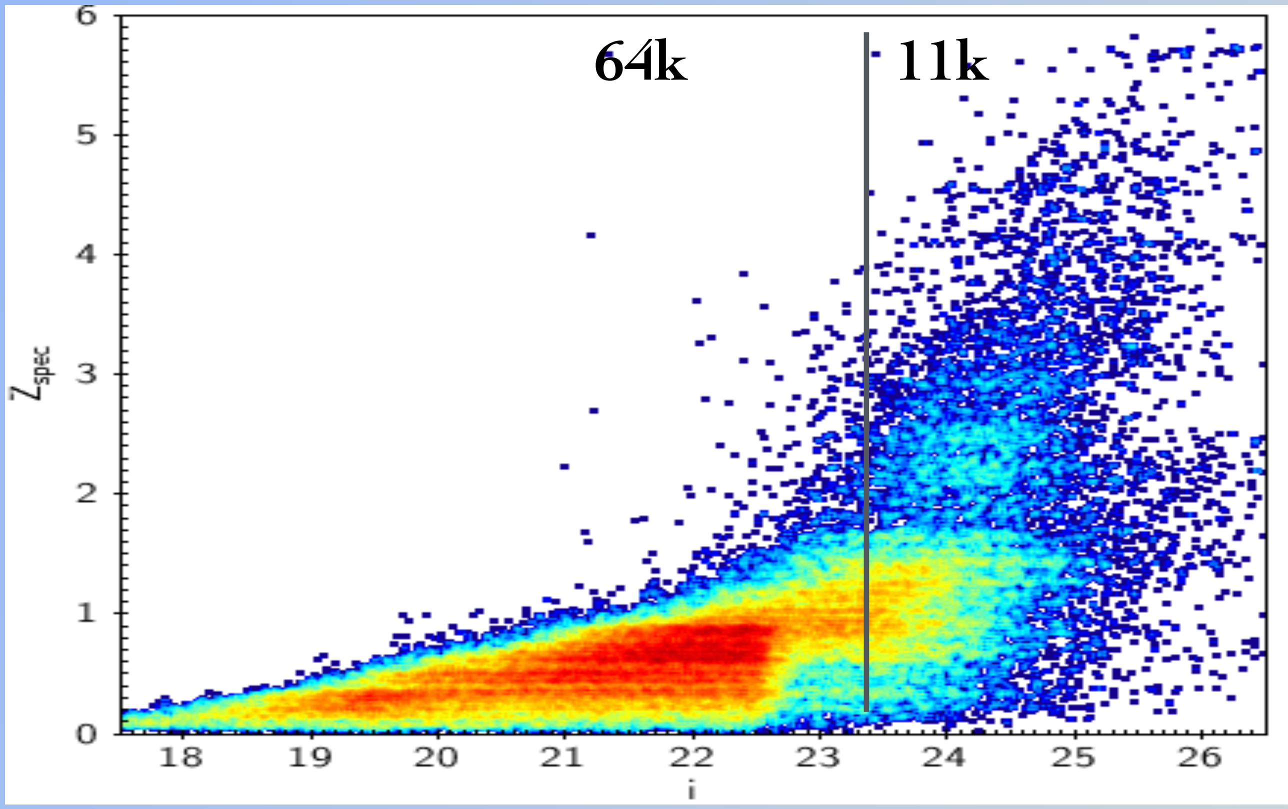
—> **HSC-CLAUDS catalogues** (Desprez, Picouet+, in prep)

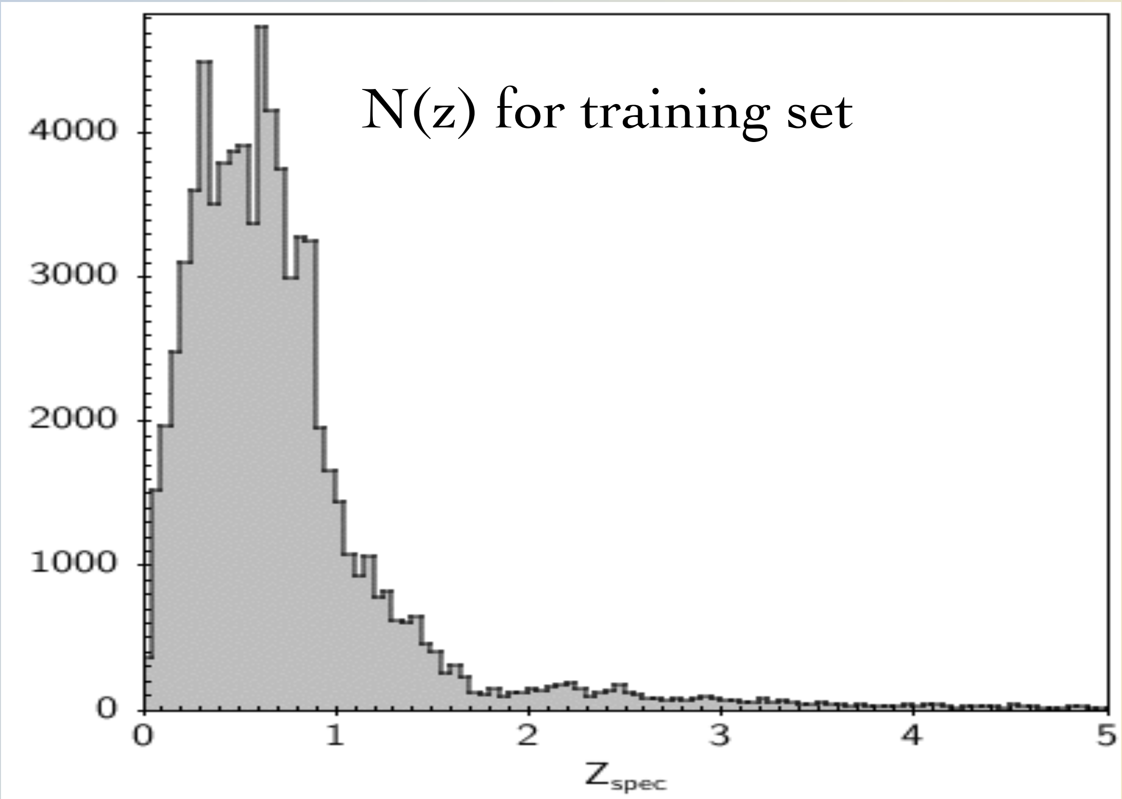
SExtractor + HSC-pipe photometry

Le Phare + Phosphoros photometric redshifts

- First release : This summer

—> **HSC Spectroscopic dataset with 75,000 spectroscopic redshifts**





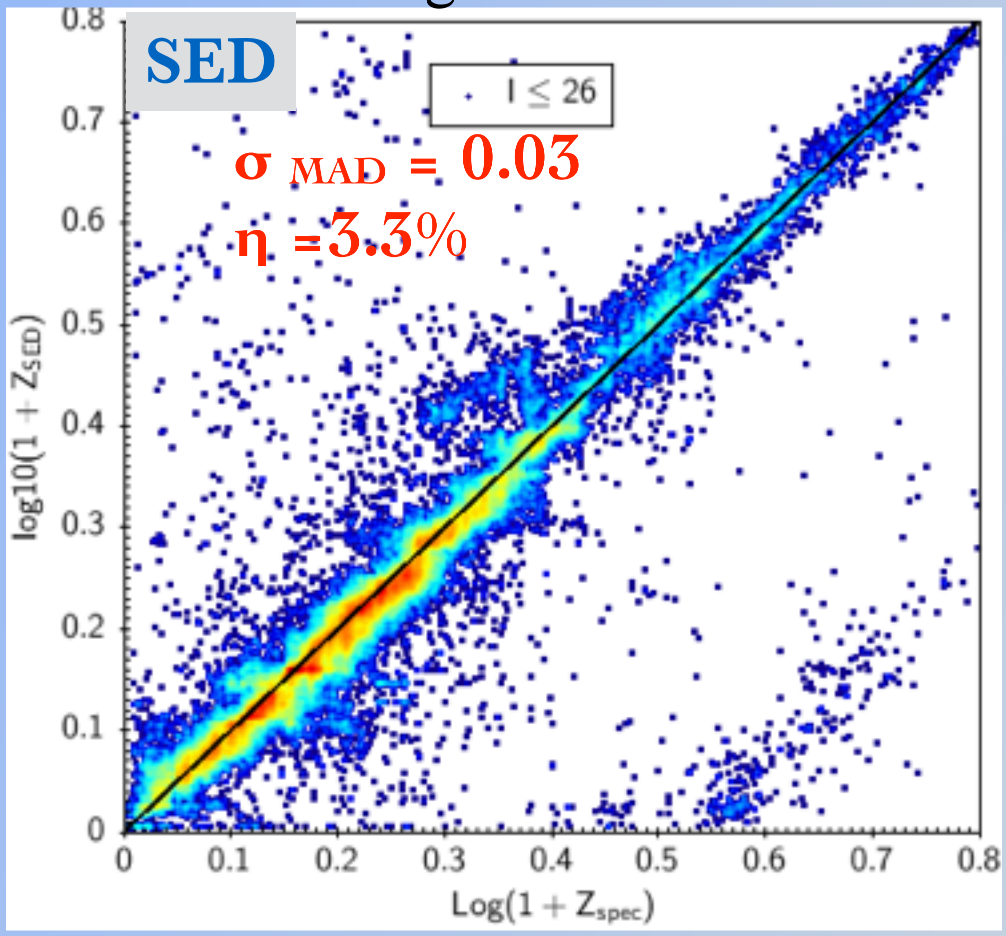
unbalanced low/high z

poor training set & poor representativity at i > 25 !

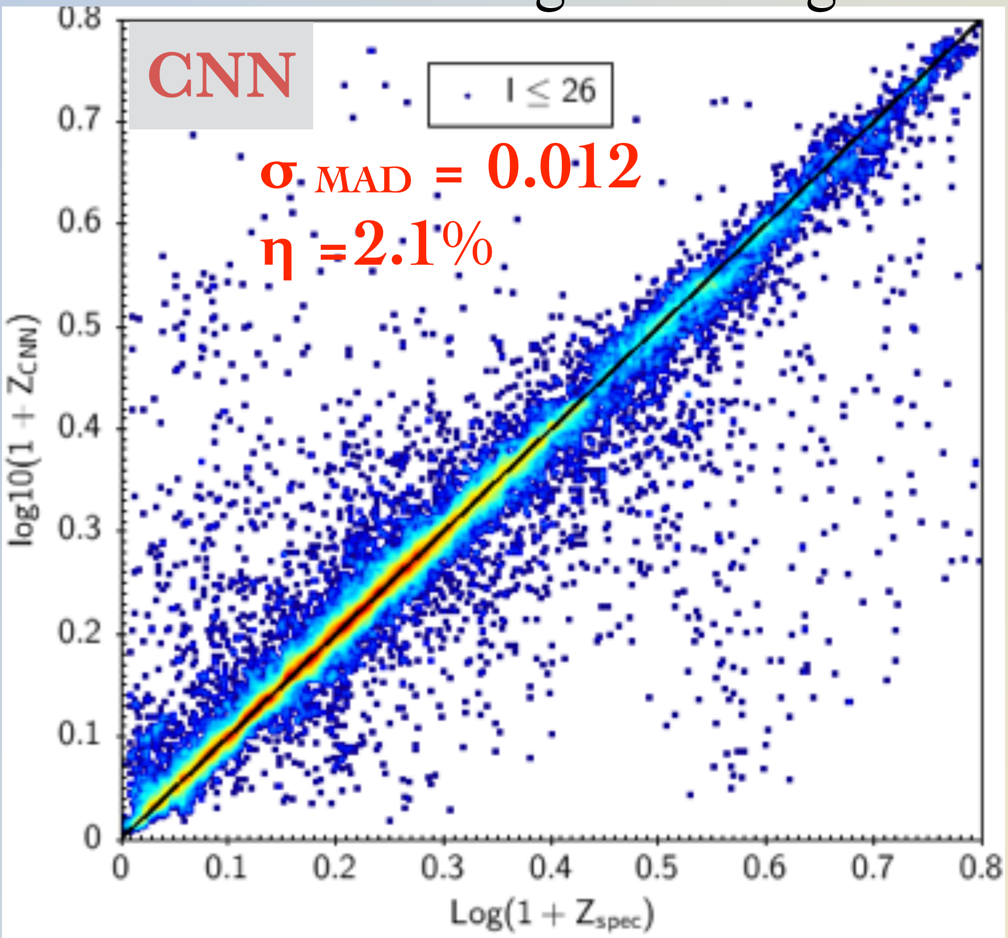
I : Comparison with spectroscopic redshifts

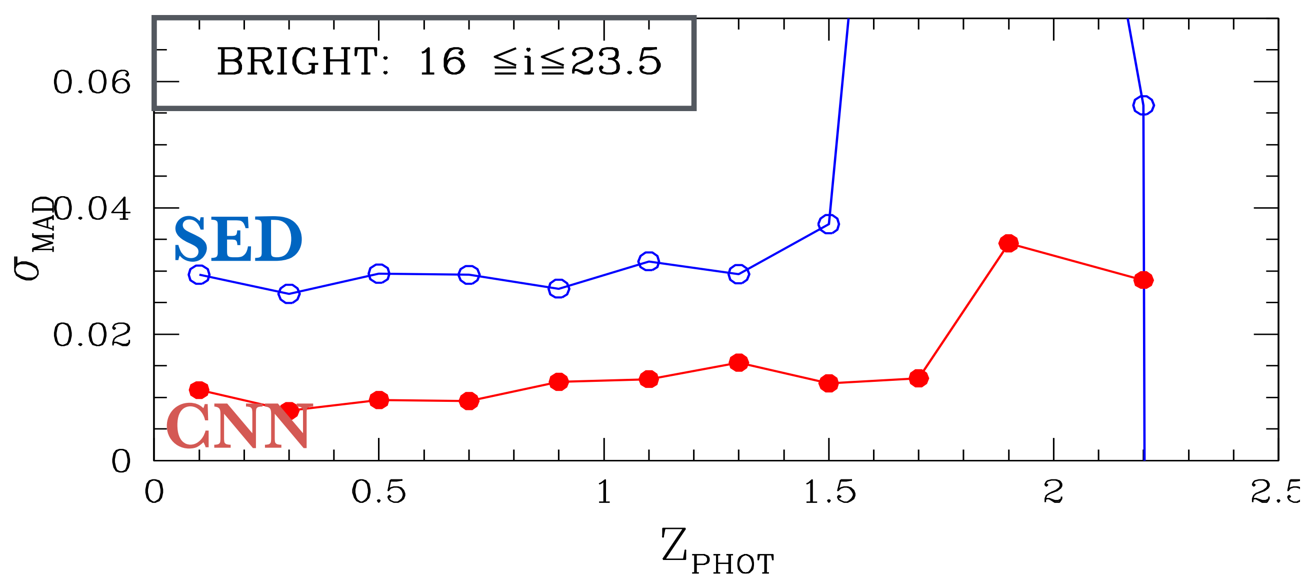
Treyer, in prep

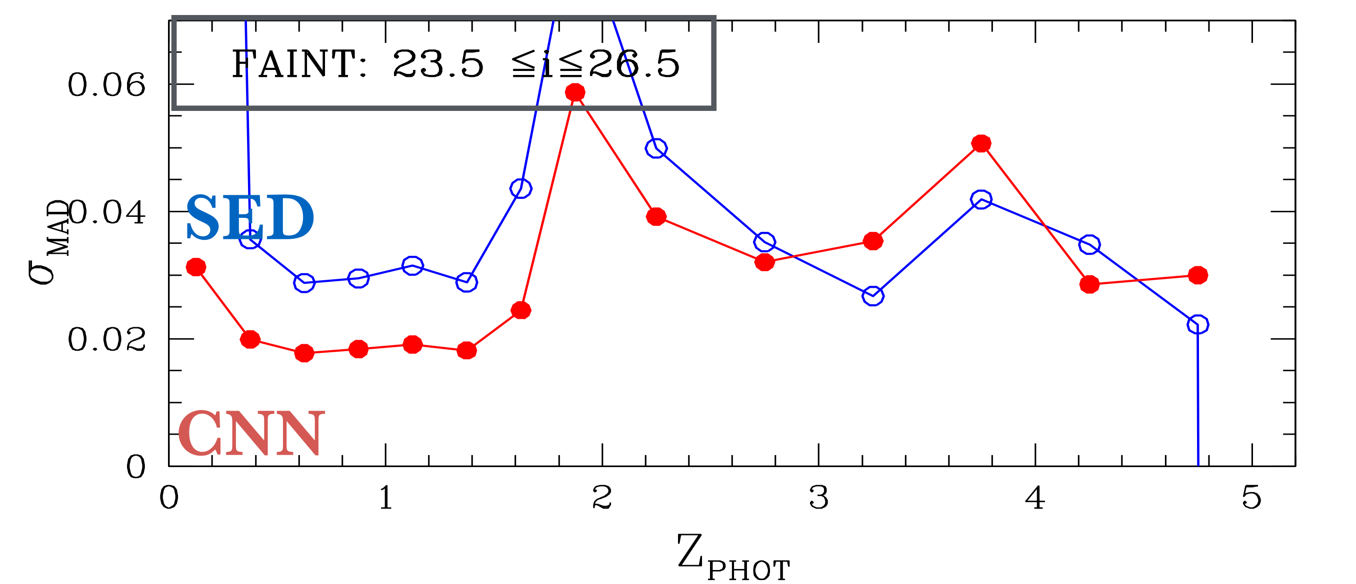
LePhare —> UgrizY+GALEX+JHK



CNN —> UgrizY images







II : Comparison with external dataset : COSMOS2020 (Weaver+21)

Treyer, in prep

—> 30 bands « **ultimate** » photo-z

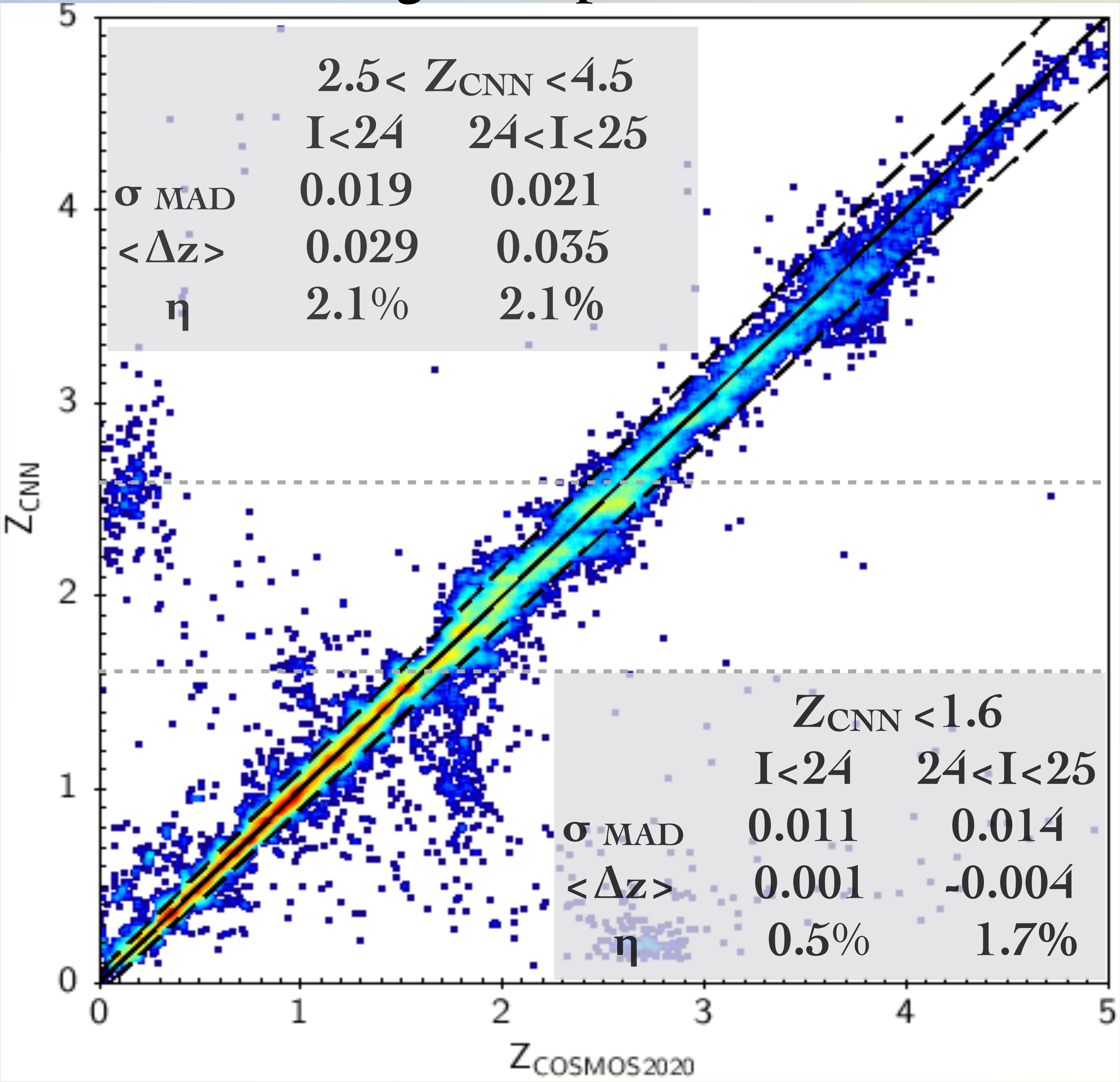
(U -> IRAC)

—> **2 photometric catalogues**
(SExtractor + Tractor)

—> **2 photometric redshift codes**
(EASY, Le Phare)

—> independent test with
~180,000 z = mean value of 4
photo-z (σ<0.1)

Bright sample : I<25

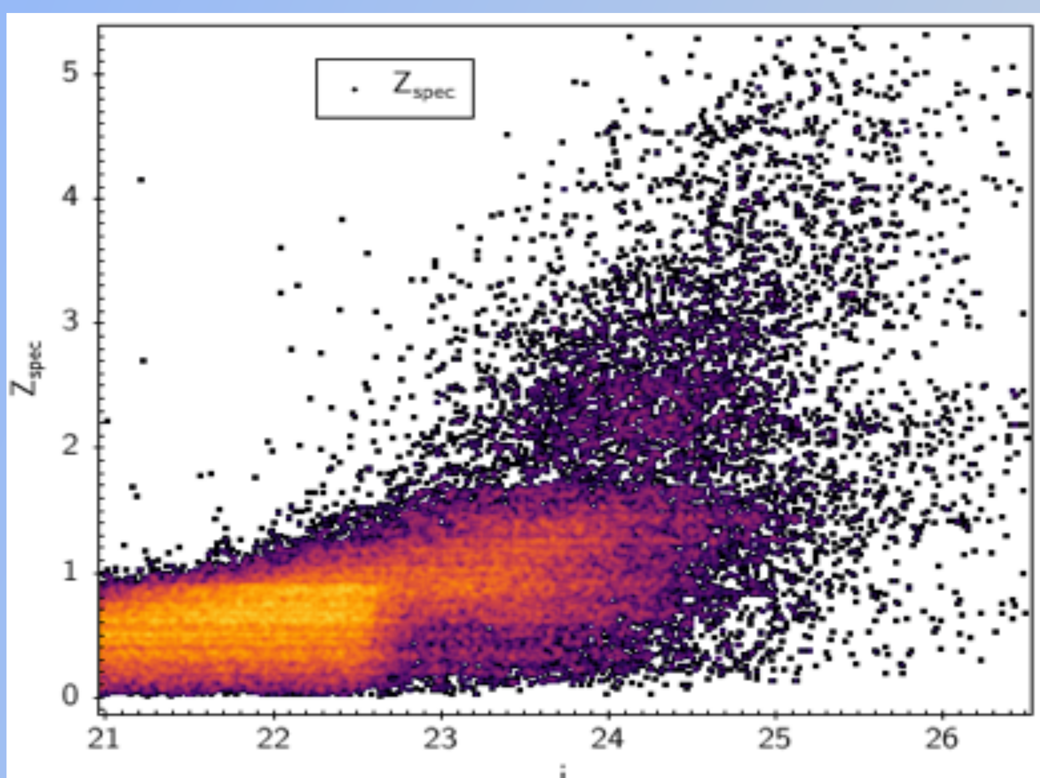


II : Comparison with external dataset : COSMOS2020 (Weaver+21)

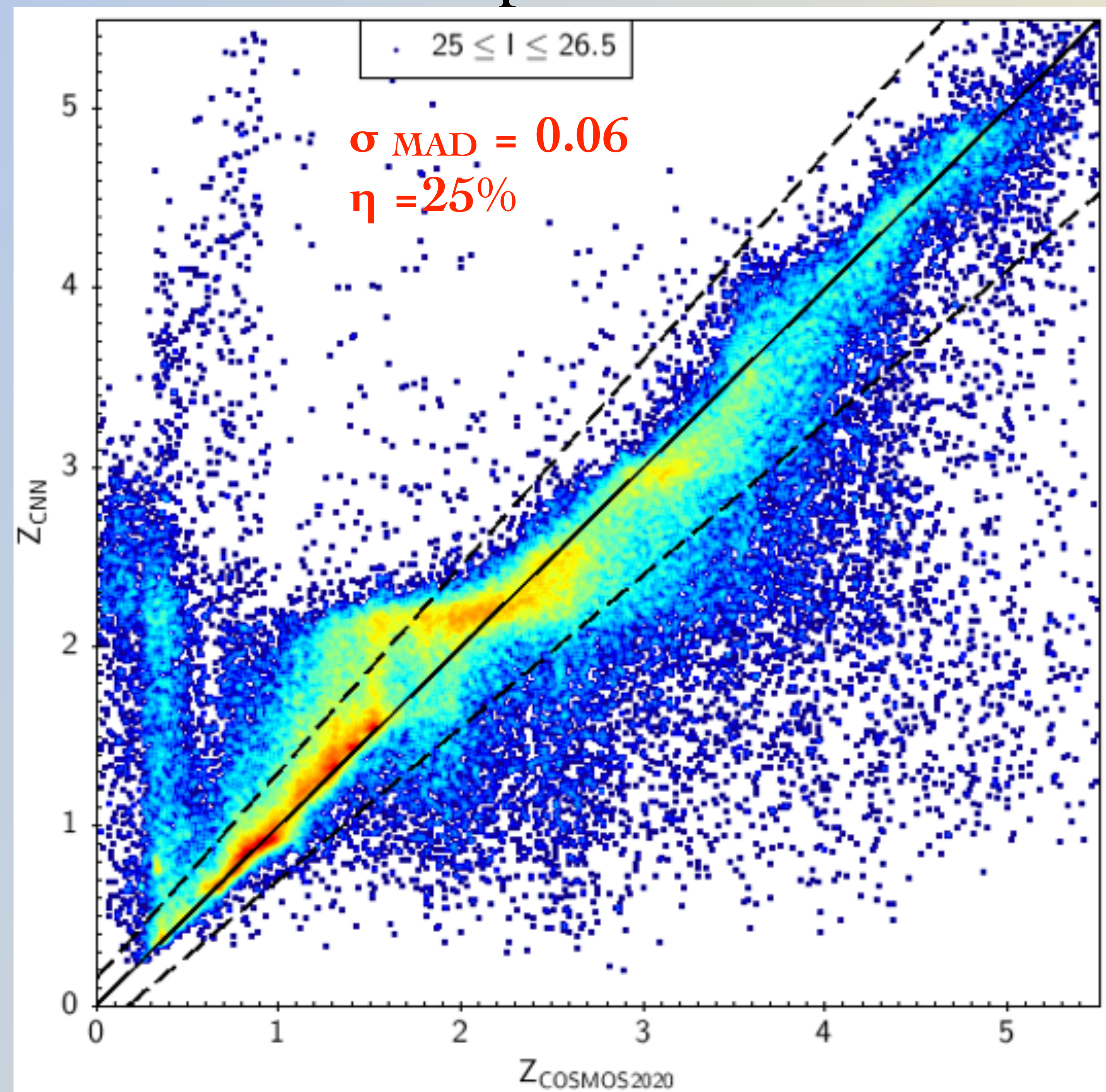
Treyer, in prep

high discrepancy at faint magnitudes

—> CNN suffers from lack of z training



Faint sample : 25<I<26.5



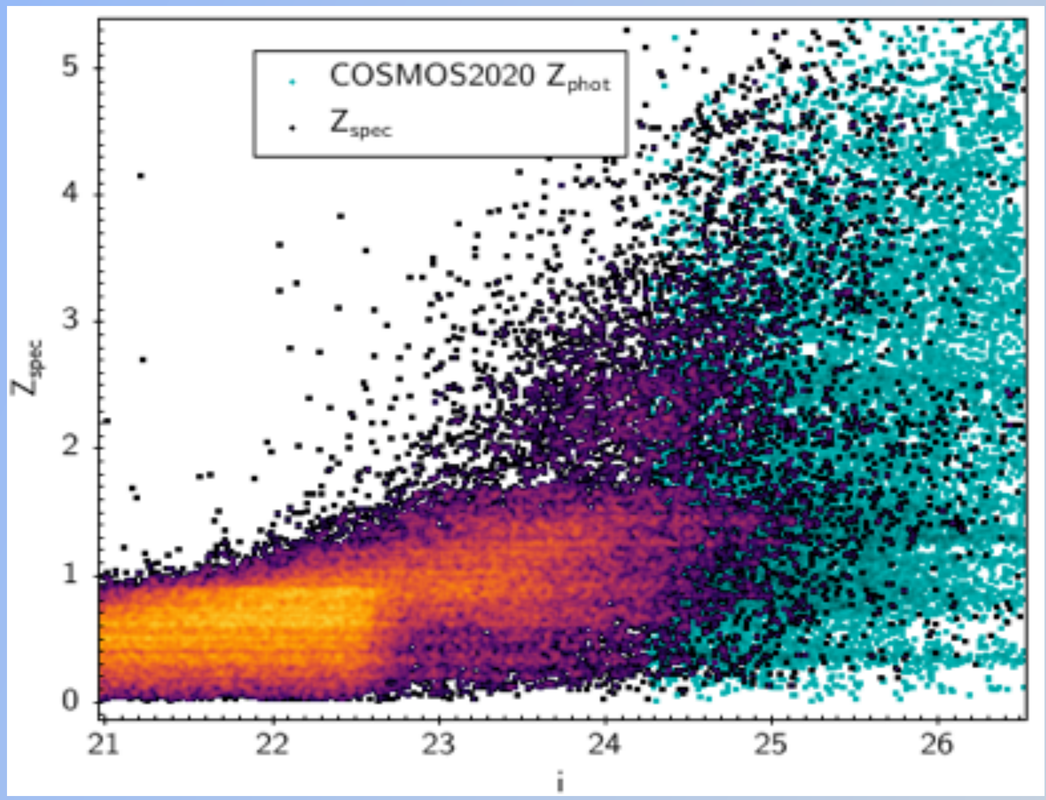
II : Comparison with external dataset : COSMOS2020 (Weaver+21)

Treyer, in prep

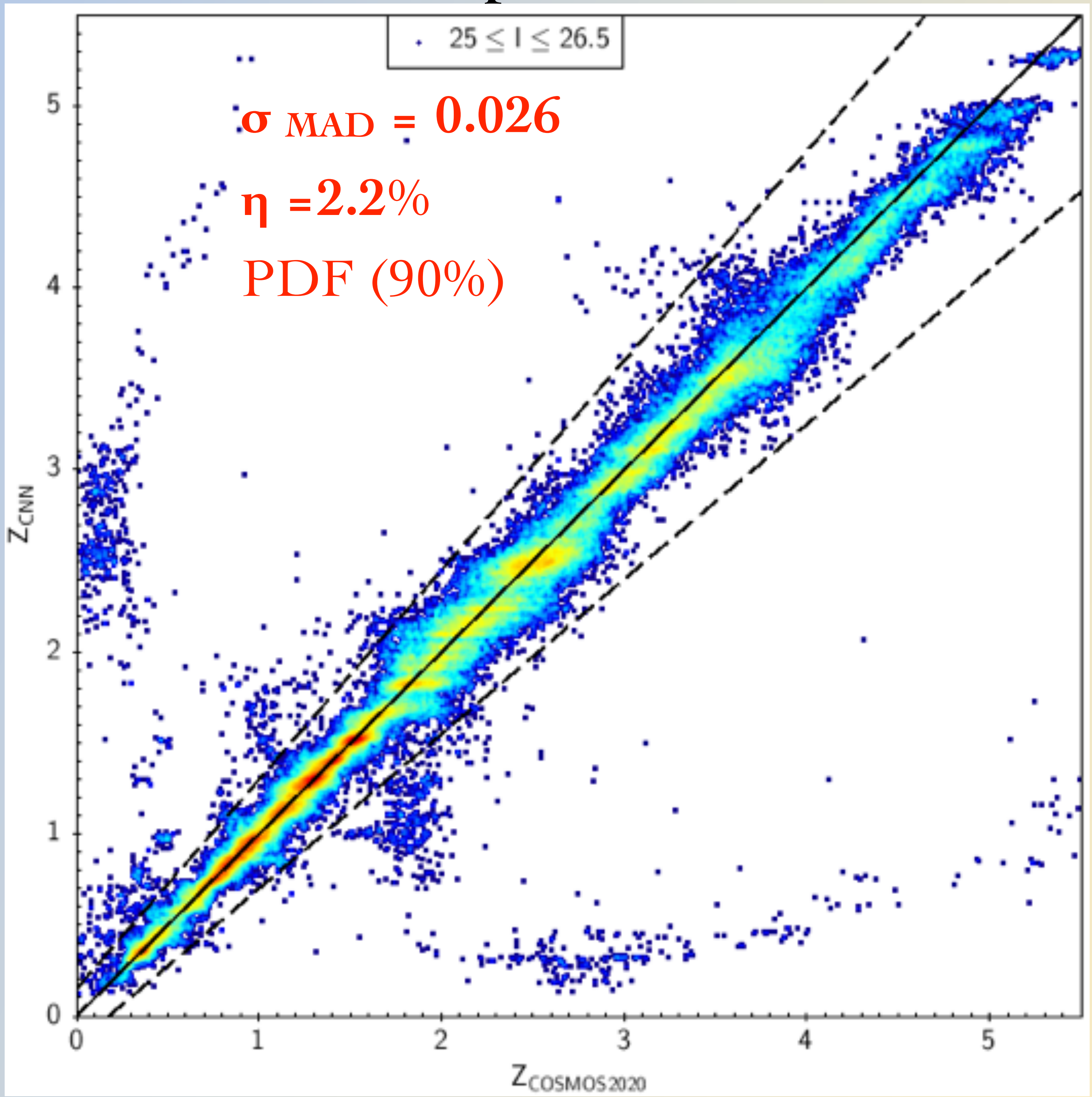
high discrepancy at faint magnitudes

—> CNN suffers from lack of z training

—> add 15k COSMOS photo-z at i>24.5

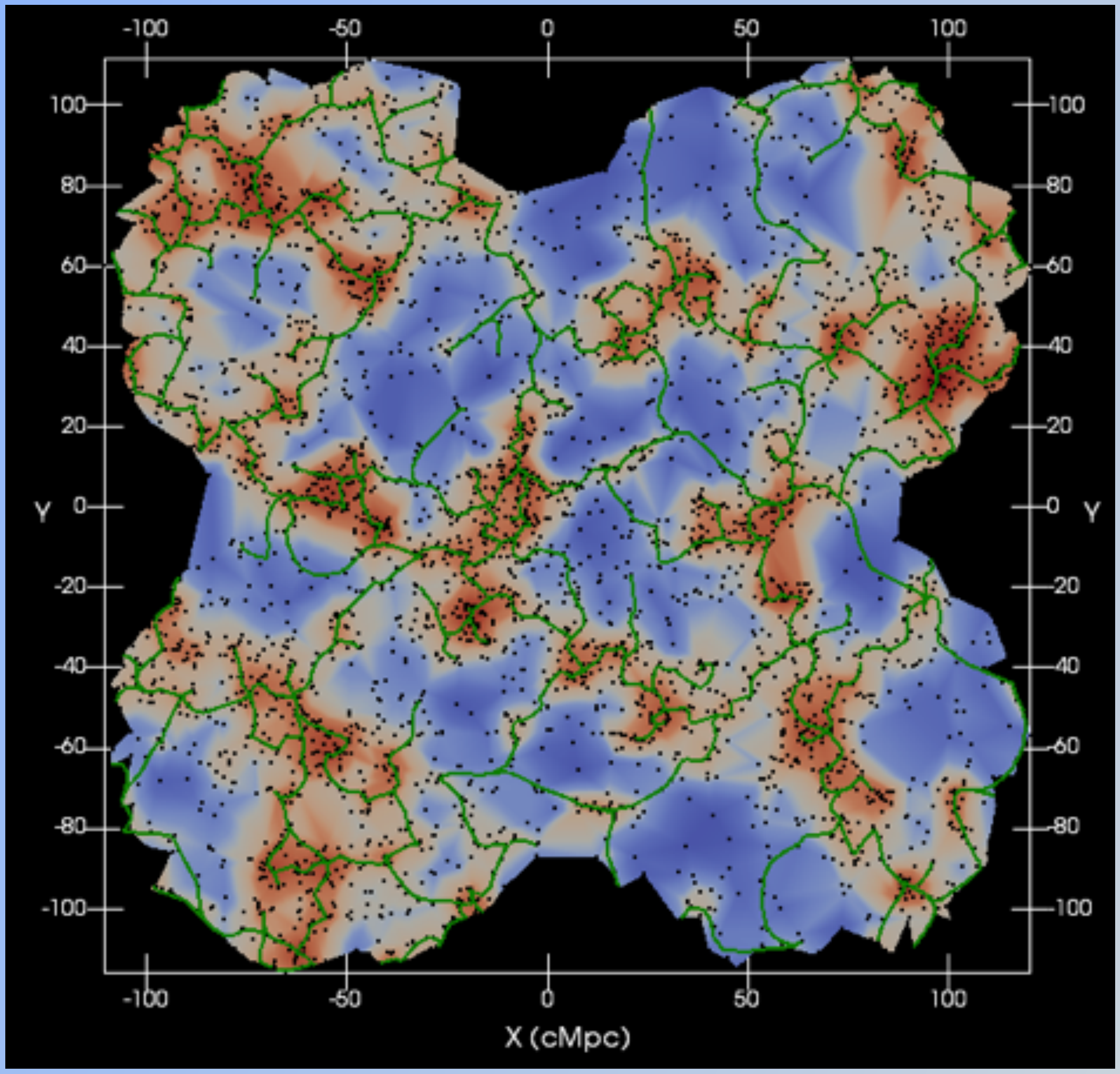


Faint sample : 25<I<26.5

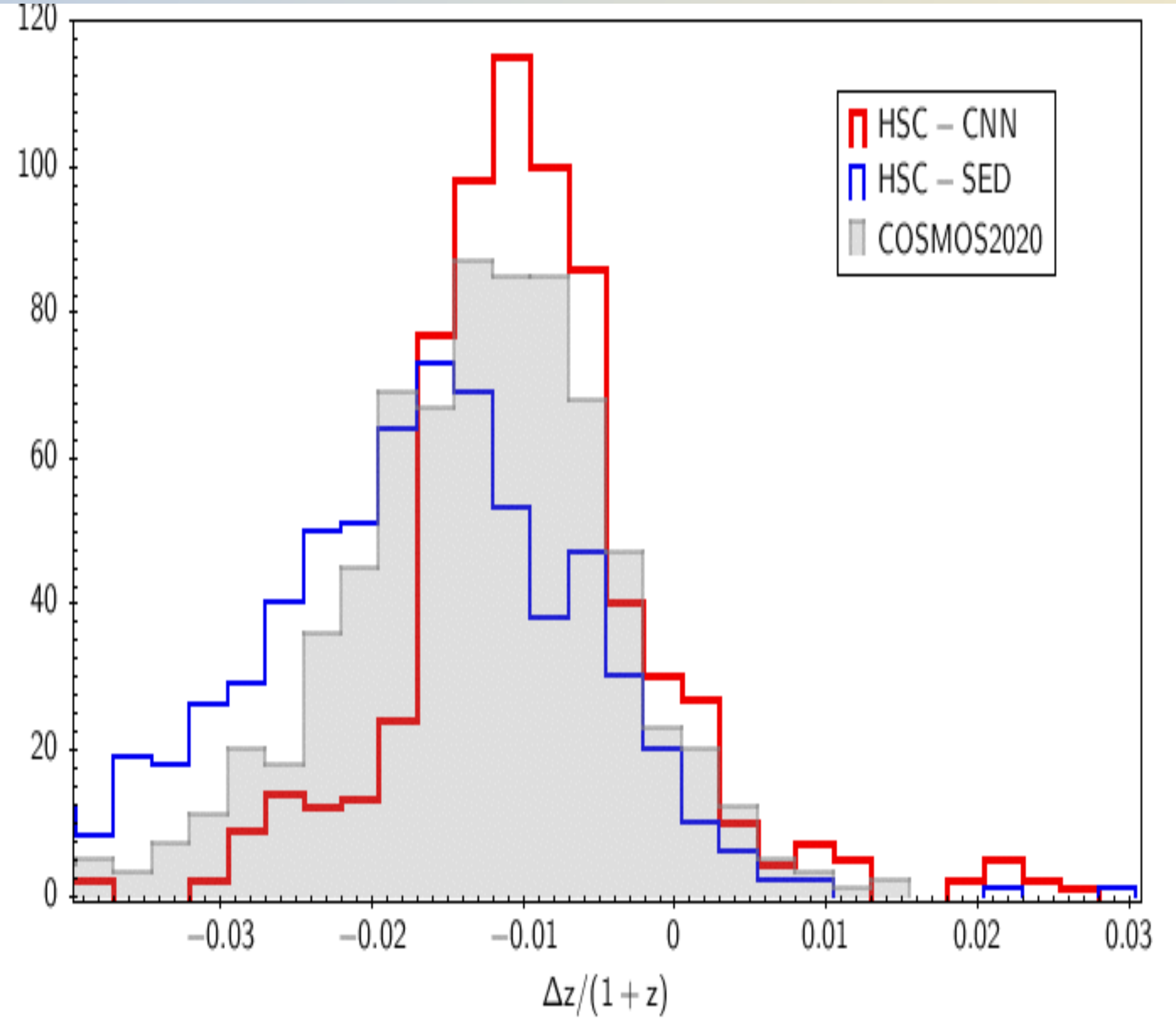


— Mixed CNN training in critical regimes

Comparison with COSMOS2020 (Weaver+21)



~750 OII Emission Line galaxies in NB912
at <z>=1.47 with i<26



CNN photo-z Summary

– **CNN photo-z is a powerful and promising alternative for large imaging surveys with limited photometric passbands**

—> at high-z : include NIR images (or fluxes) will help (WFIRST, Euclid)

– **Need to control the representativity of the training set**

—> to be improved at high-z / faint mag regimes : PFS, MOONS, KMOS, MSE…

– **Explore un/supervised methods with un/poor labeled data**

—> pretrain on large unlabeled data + fine-tuning with some labels (contrastive learning)

help to reduce the required training set (Hayat+21)

—> domain shift with un/label data