The DEEPDIP ANR

– **Deep Learning techniques for SN's and Photometric redshifts**

DEEPDIP.net

LAM :	Treyer, Kraljic, de la Torre, Ilbert, Vibert, Gray, Moutard, Picouet, Arnouts
CPPM:	Fouchez, Bautista, Lin, Racine, …
IAP:	Bertin, McCracken, Codis, Laigle, Pichon, Dubois
Montpellier:	Pasquet (TETIS), Chaumont (LIRRM) …

—> 2 Postdocs: Katarina Kraljic (LAM) + Julian Bautista (CPPM)

—> 1 thèse ANR-IA: Reda Ait-Ouahmed (LAM +TETIS)

– **Science Goals:** cosmology with SNs & Cosmic Web analysis

prepare next generation surveys: LSST + Euclid

Why exploring Deep Learning techniques ?

—> A large family of Photometric redshift techniques:

— **SED fitting** needs a small training set for calibration but **computationally intensive**

— **Machine learning** (Aritificial Neural Network, kNearest NeigborsN, Random forest, SOM, …)
very good accuracy when using a large training set

—> One main Limiting factor: **input informations based on extracted features**
relies on flux extraction which can be sensitive to PSF, neighbors, profile models …

—> Deep Learning approach:

— **no feature extraction.** **Works at the pixel level !**
exploits all the informations (SB, sizes, inclinations, color gradients, neighbors)

— Now under reach thanks to large spec-z samples & 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)

—> ML + DL :

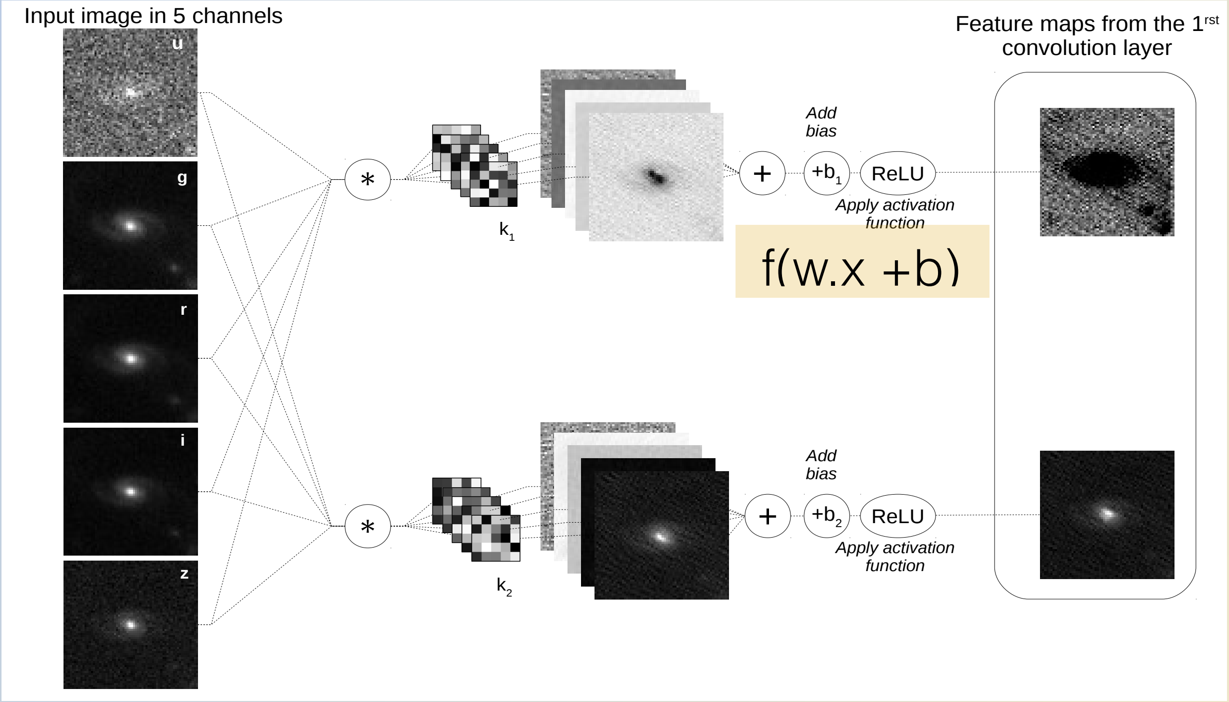
— **limited to the representativity of the training set**

Photometric redshifts with Deep CNN

Convolutional Neural Network

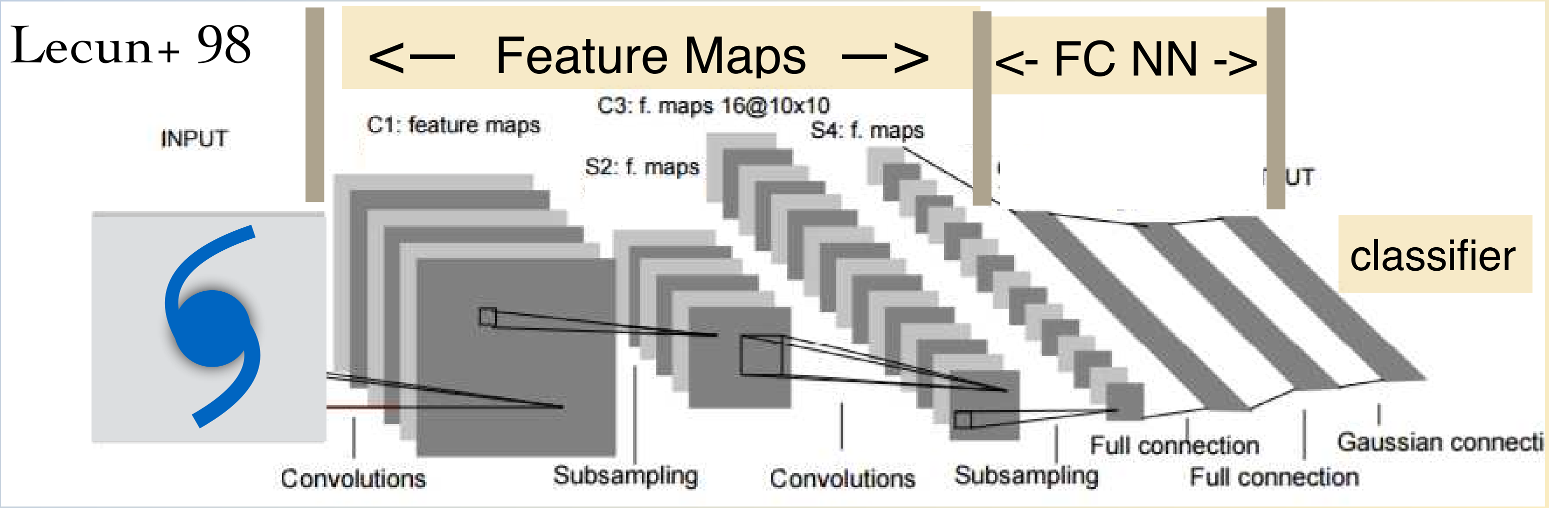
Pasquet+19

- **input** : 64x64 ugriz images



- **First steps :** convolution blocks
  apply convolution kernels to extract
  several feature maps
  (successive conv. blocks with pooling to
  reduce their sizes)

- **Second steps** : Fully Connected NN
  Final features maps + E(B-V)$_{gal}$
  are inputs for the FC NN



- **Last step** : classifier
  Output layer is a classifier with bins of δz
  providing a normalized PDF ($z = \sum_k z_k . P_k$)

- **Training : back propagation**
  to minimize 28 millions parameters

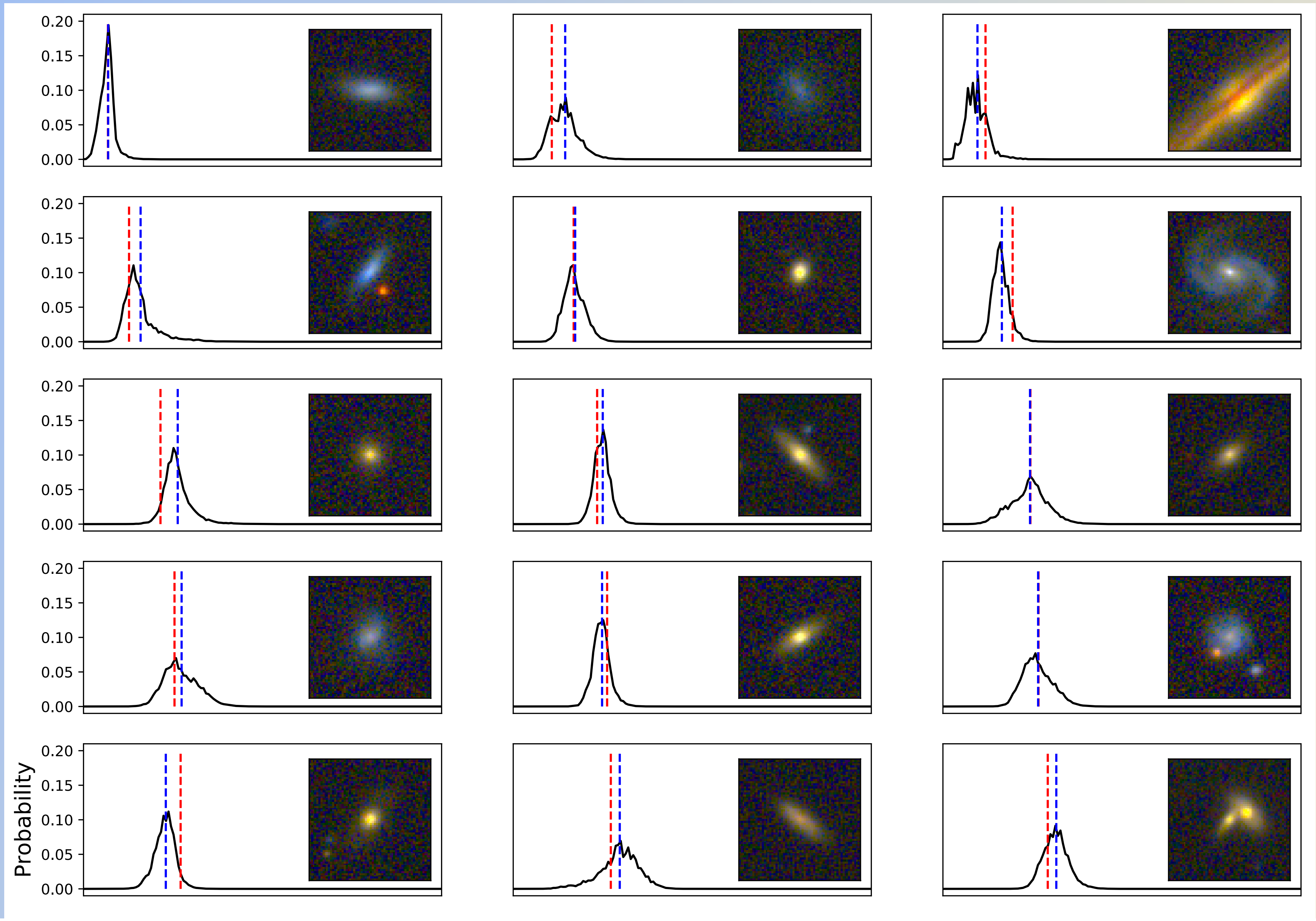
Photometric redshifts with Deep CNN

—> **Protocol with all the training set :** *Pasquet+ 19*

— 5 cross validation samples with : **80% training + 20% testing**

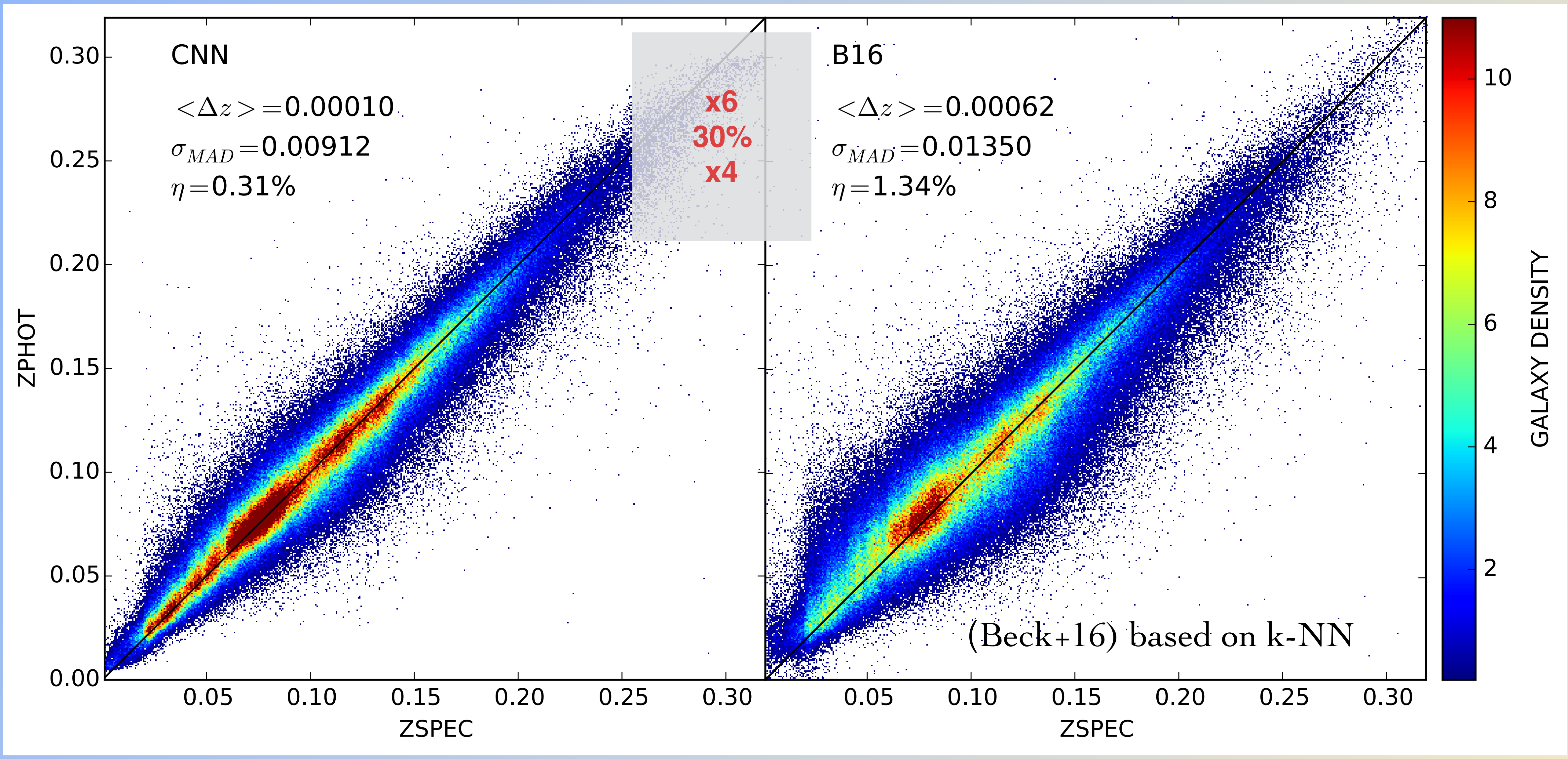
+ 6 ensembles (Training set augmented by rotation + Flip of the images)

—> $z_{\mathbf{cnn}} = \sum_k z_k \,.\, \mathrm{PDF}_k$



Pasquet+ 19

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



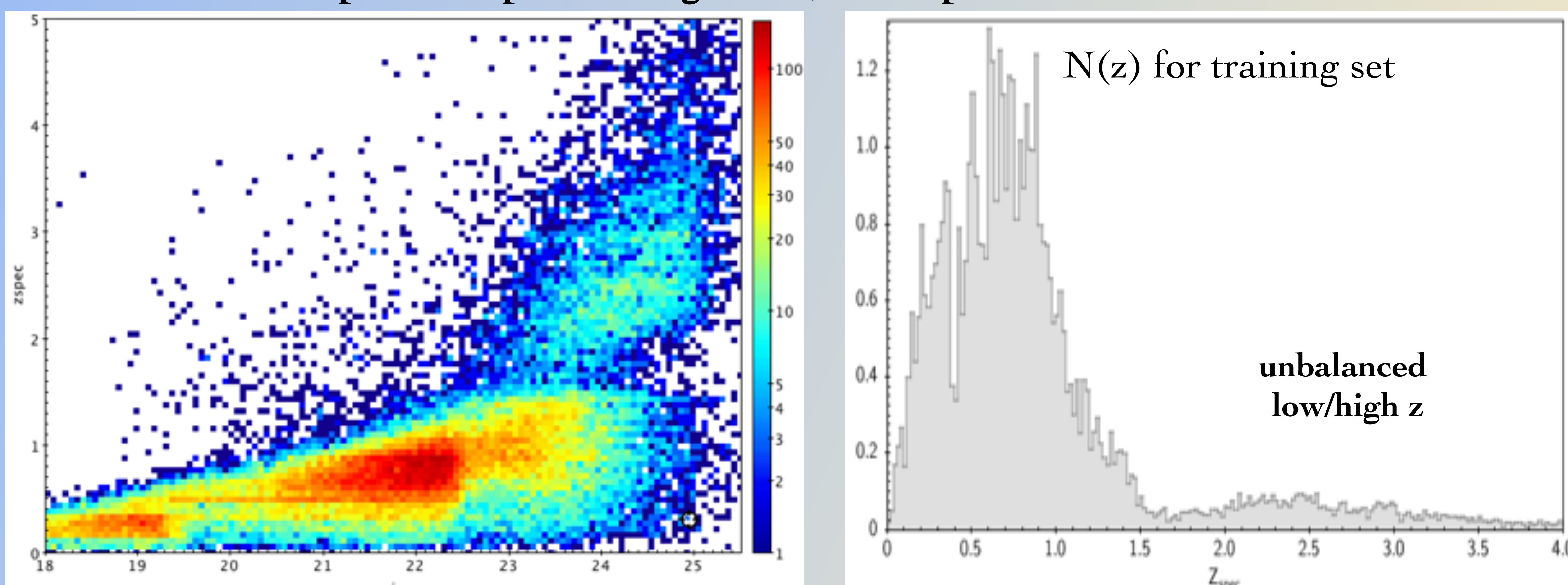
—> Better performance than the latest SDSS photo-zs

—> Moving to Higher redshift with Deep imaging surveys

New challenges :

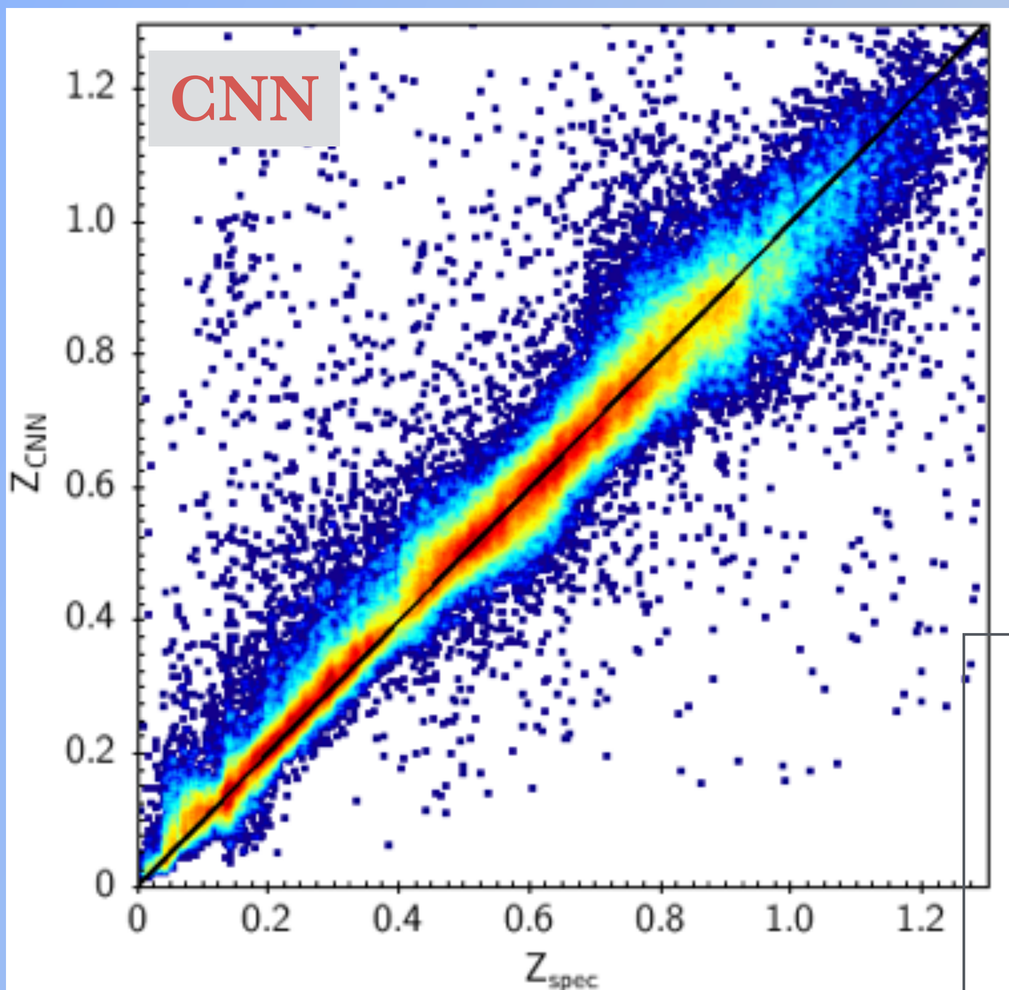
- Large redshift range —> large number of classes for training
- Smaller and inhomogeneous training set
  over/under-represented training set
  unbalanced in z

Spectroscopic training set 60,000 z-spec for HSC-CLAUDS

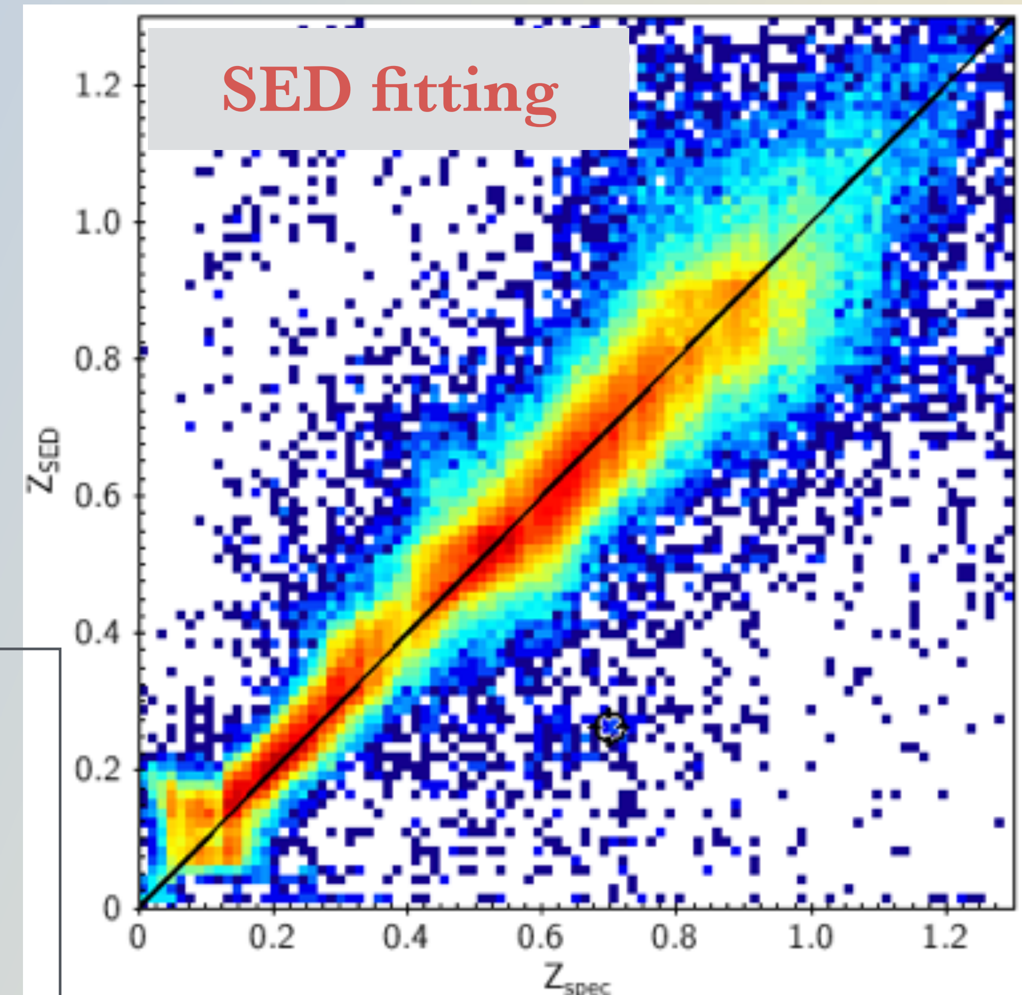


Photometric redshifts in Deep Surveys : CFHTLS

–>Performance at i<22.5 with CFHTLS - WIDE images



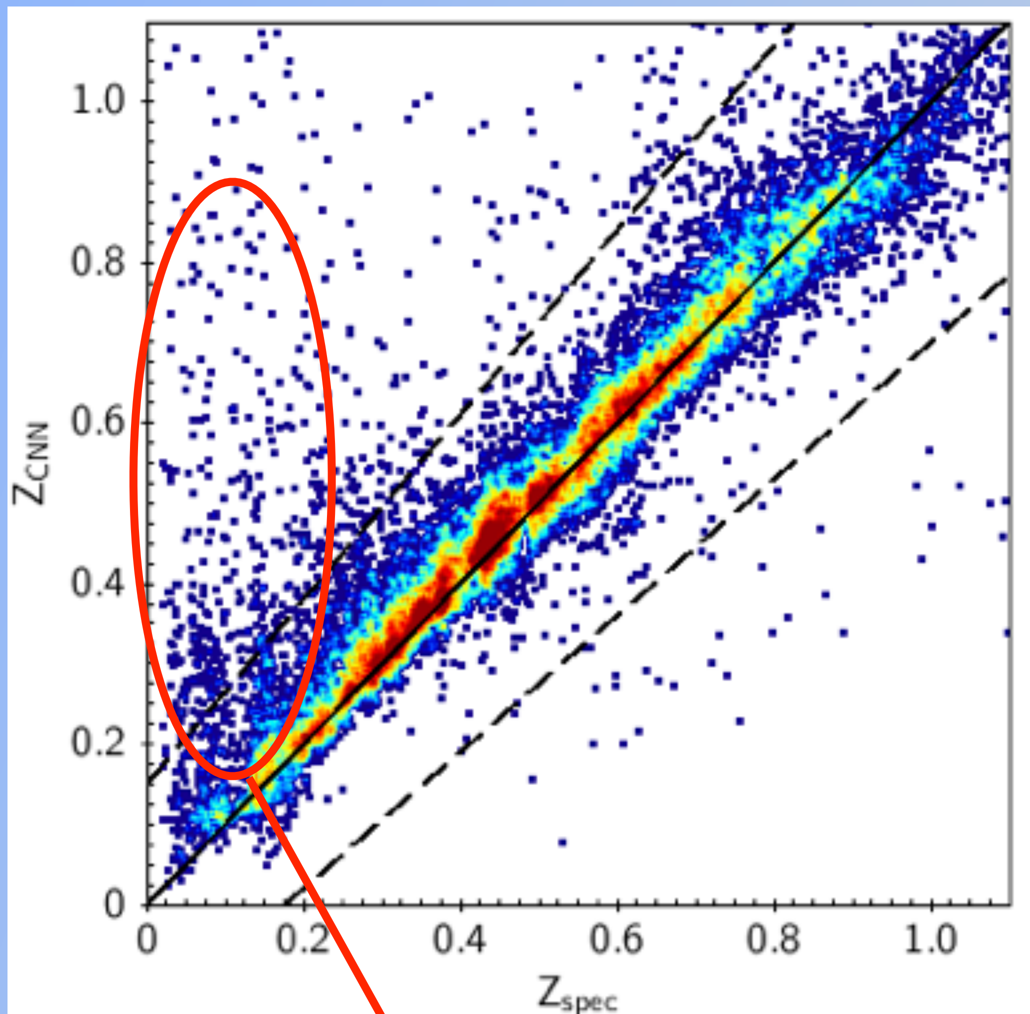
WIDE	125k z_s	
0.019	σ MAD	0.035
-0.002	**<bias>**	0.005
1.9%	η	2.9%



~ factor 2 improvement vs SED fitting

Photometric redshifts in Deep Surveys : CFHTLS

- External test with PRIMUS + 3D-HST (low resolution spec-z , i<22.5)

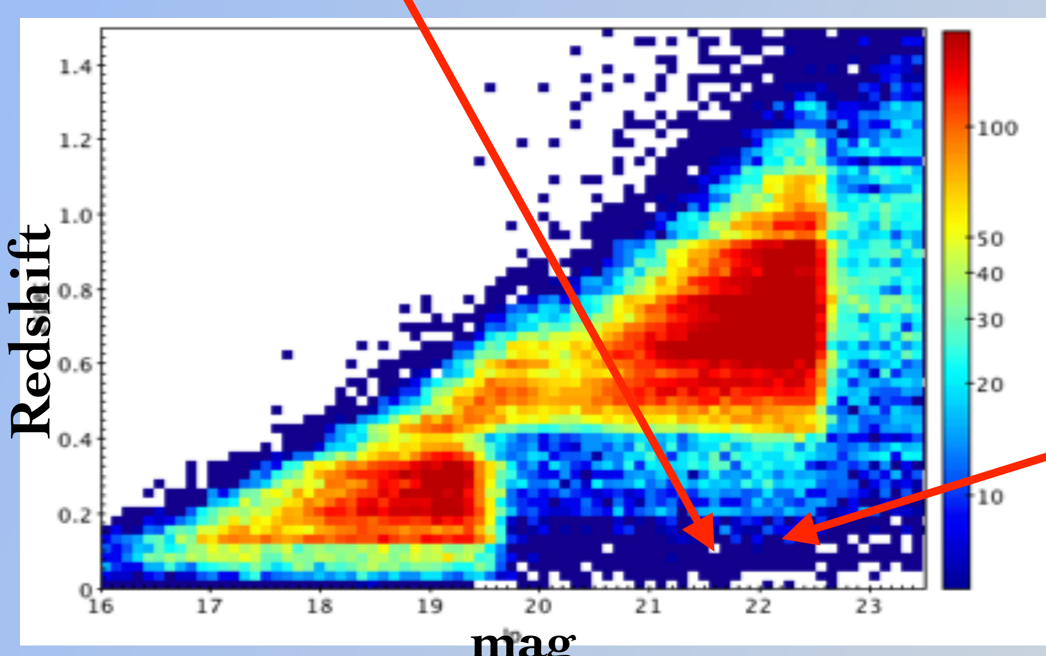


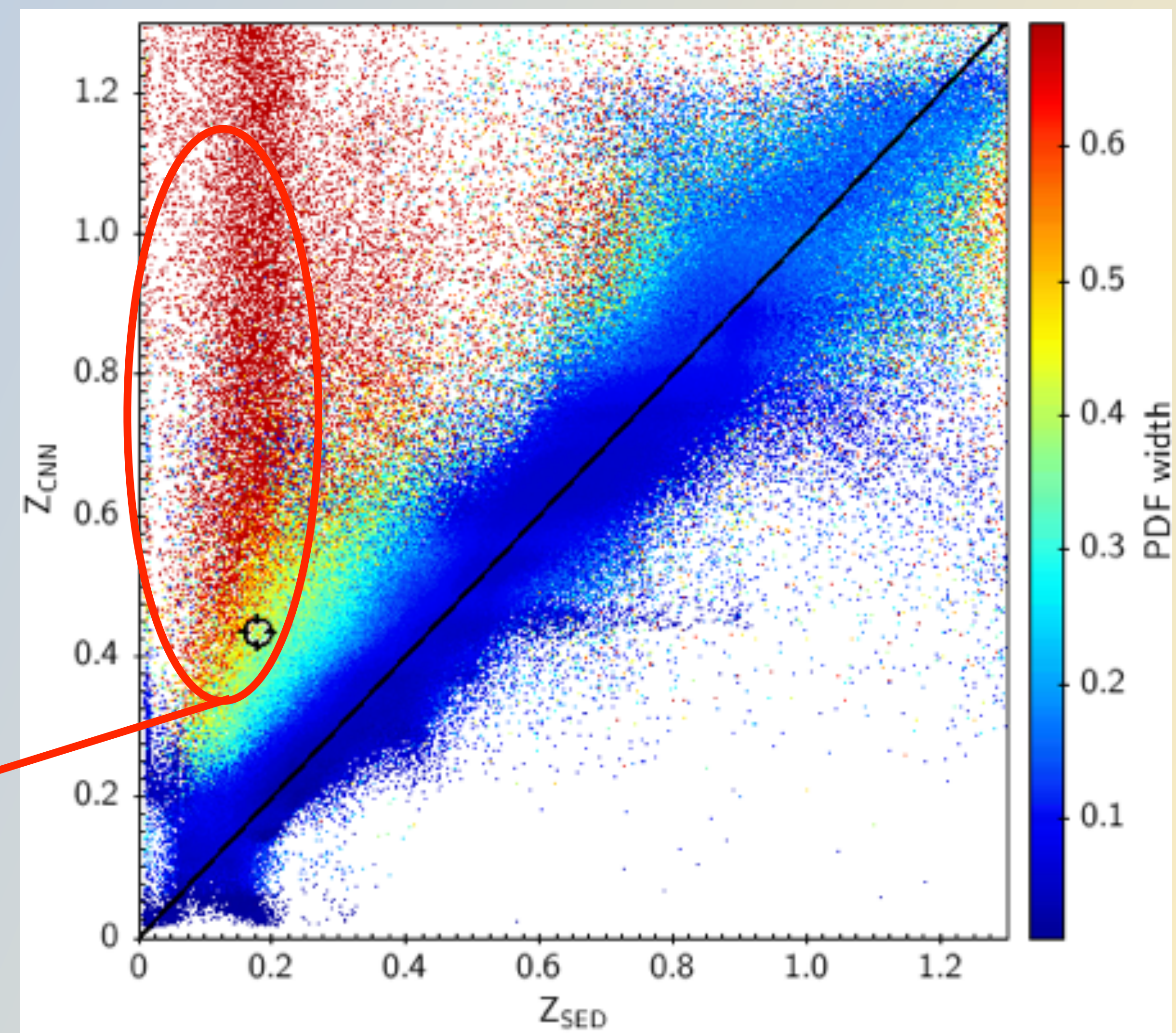
	CNN		SED
—>	0.021	σ MAD	0.036

—> **1st issue :** impact of the training set

bias at low-z due to lack of faint spec-z

—> translated in the individual PDF





conclusion on CNN photo-z

CNN photo-z is a promising approach but we need

—> to better handle under-represented & unbalanced training set

-1- upcoming spectroscopic surveys to improve the training set for DL

(PFS, MOONS, JPASS, WAVE, WFIRST, Euclid, …) but not for now …

-2- develop alternative DL approaches

- dealing with incomplete photometry (mising bands)
- dealing with under represented regions in training set : GAN ?
- develop transfer learning from one dataset to another
- find a way to exploit the large number of unlabelled galaxies
- extend analysis to other informations (physical parameters)

—> ANR DEEPDIP : to get ready for LSST+Euclid

Ressources : 5 GPUs installed on Cluster at LAM (need more)

—> collaboration with CESAM welcome