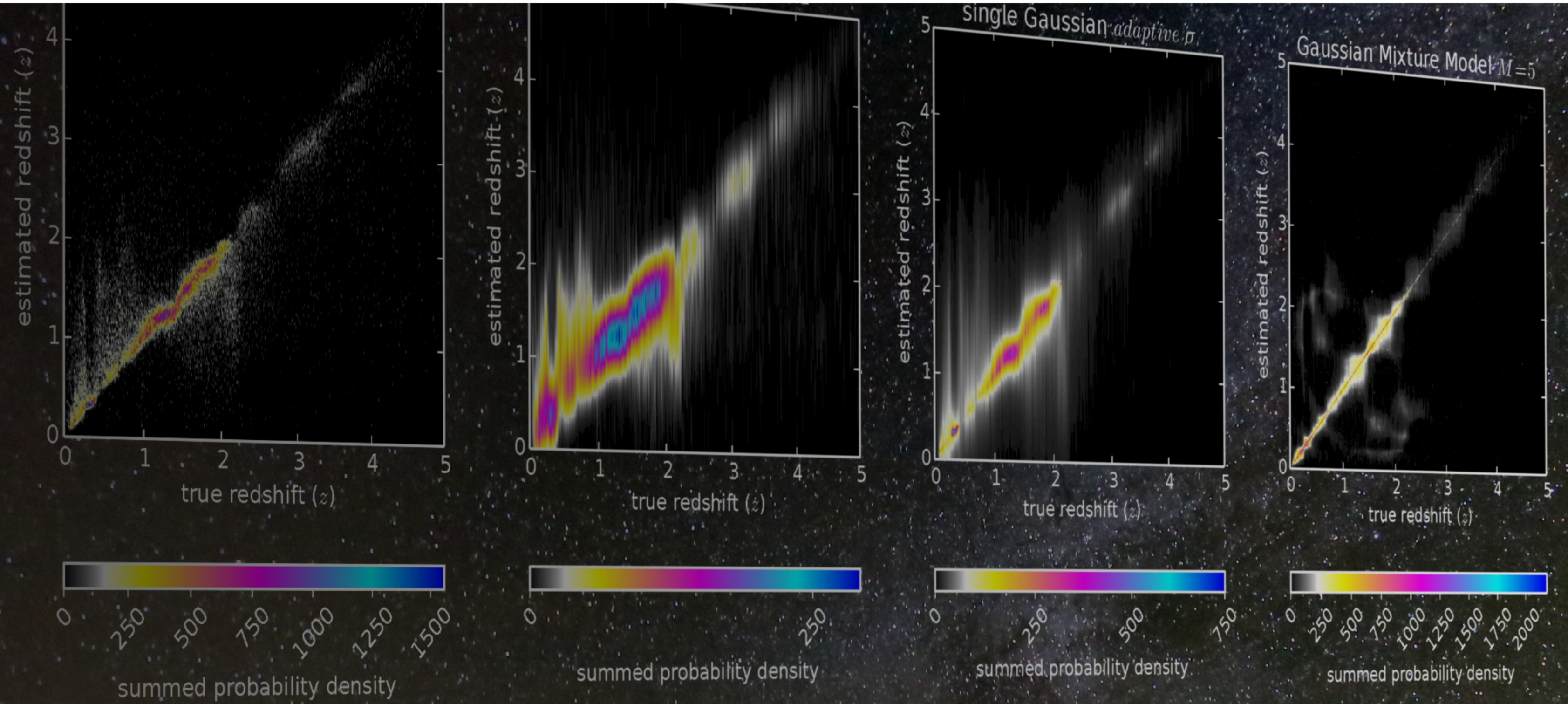
Heidelberg Institute for
Theoretical Studies

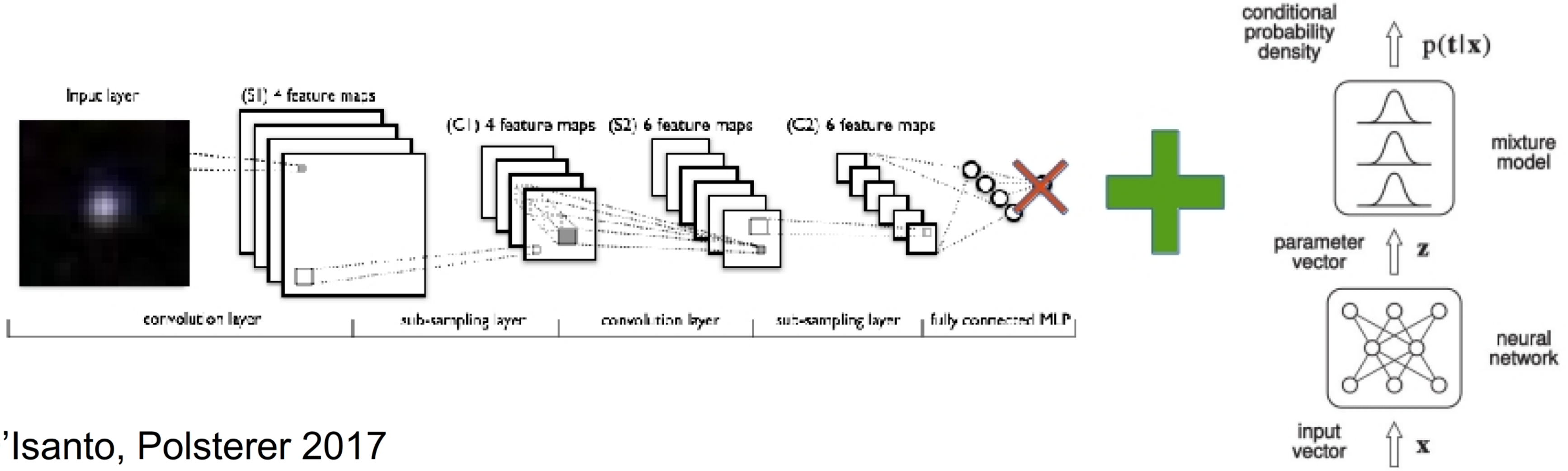
HITS



Code to the Data

a hard deep learning example

Deep Convolutional Network meets Mixture Density Network

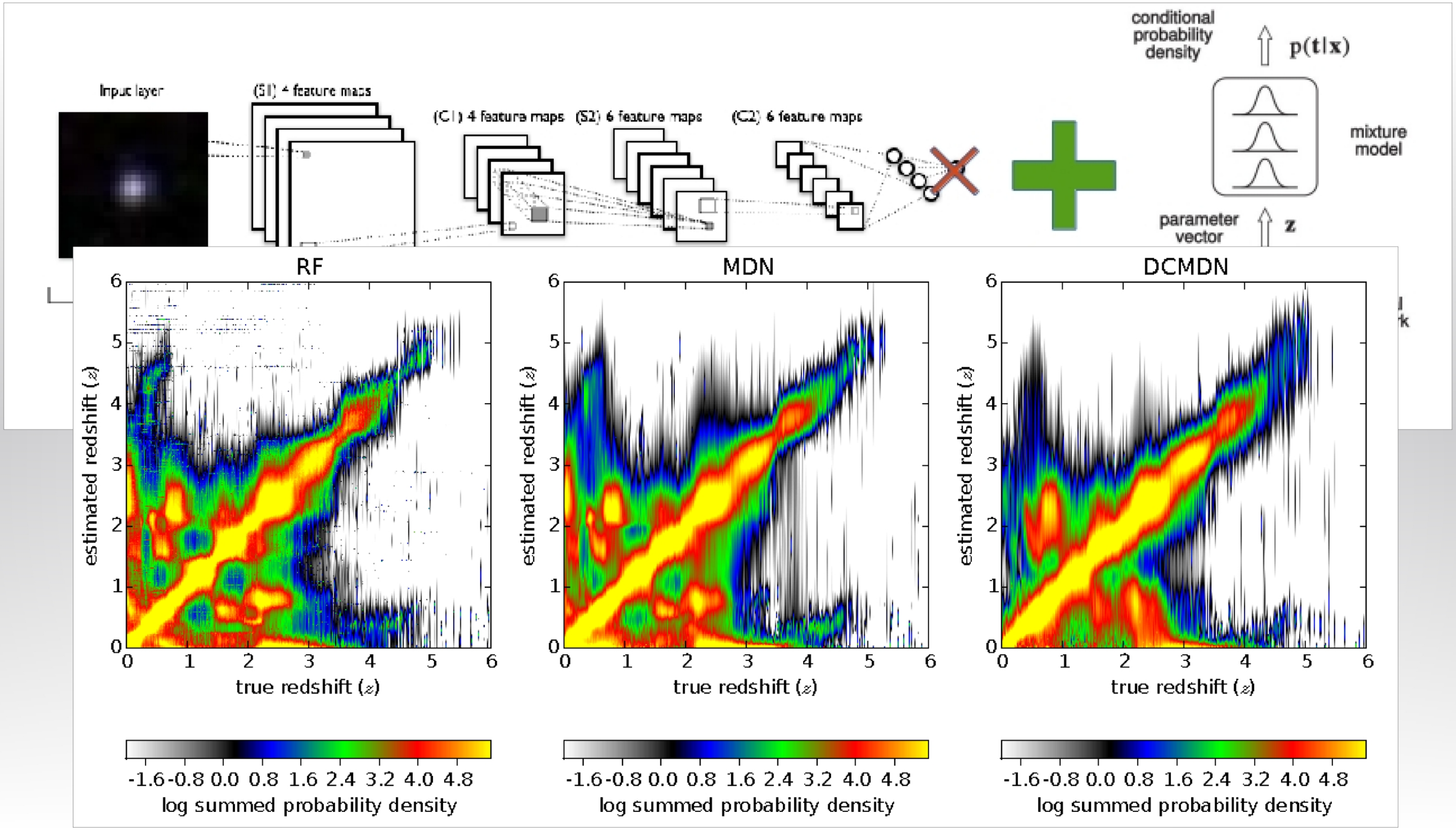


D'Isanto, Polsterer 2017

predict redshifts photometrically,
purely based on images

→ find new high redshift objects

Deep Convolutional Network meets Mixture Density Network



Code to the Data: a story of failures

Creating training data

- find all objects in CFHTLS wide with SDSS spectra + redshift
  - merge Schneider et al. 2010 & Paris et al. 2017

    crossmatch with CFHTLS wide

    no MOC at CADC → CDS

    Aladin Beta → 1.000 objects, too few

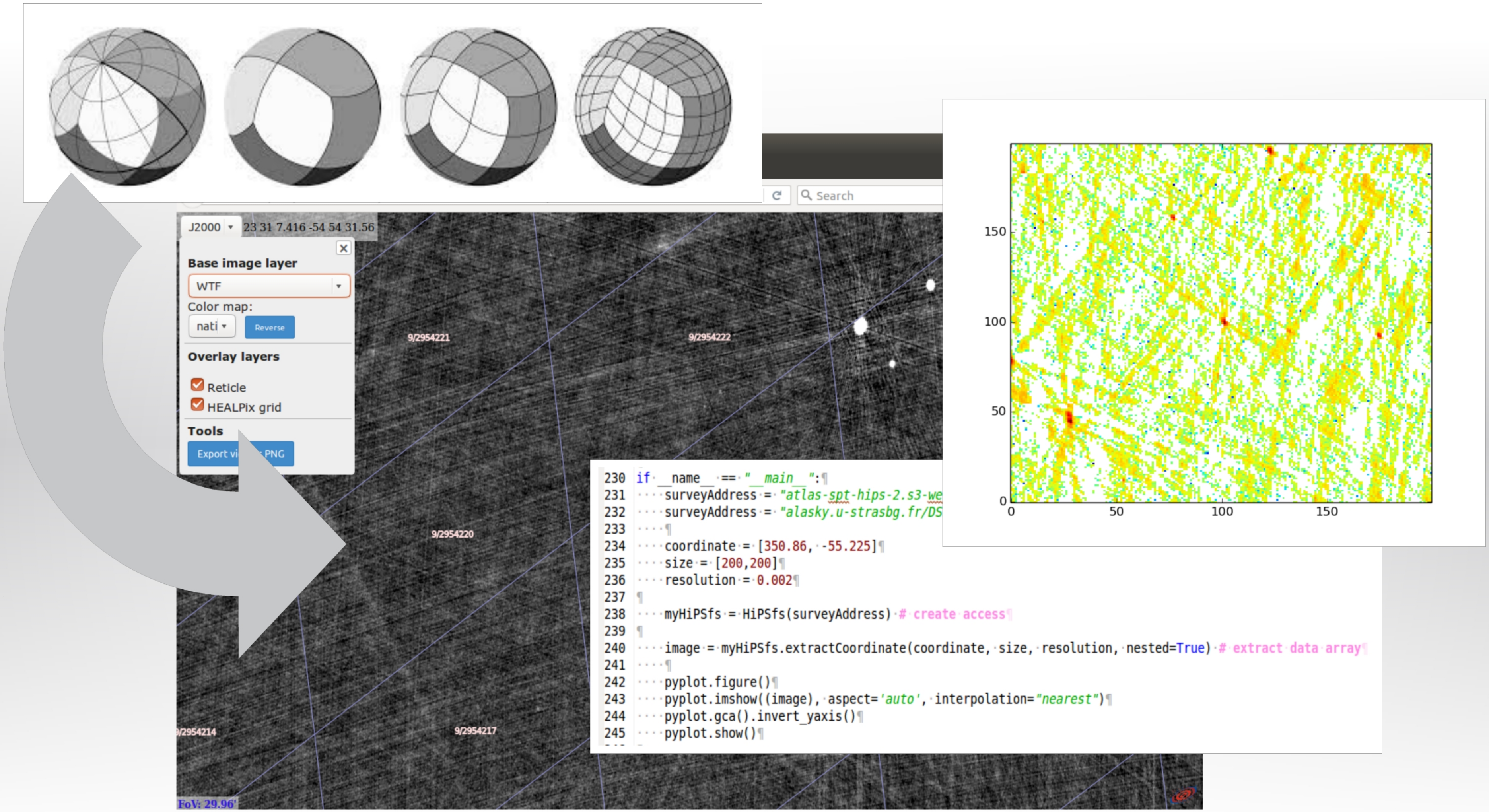
    problem with coordinate columns fixed → 3.000 objects, too few

  - all SDSS objects with spectra + other → CSV 5 mio. objects

    Aladin Beta → out of memory

    Topcat → 60.000 objects + augmentation (480k = enough)

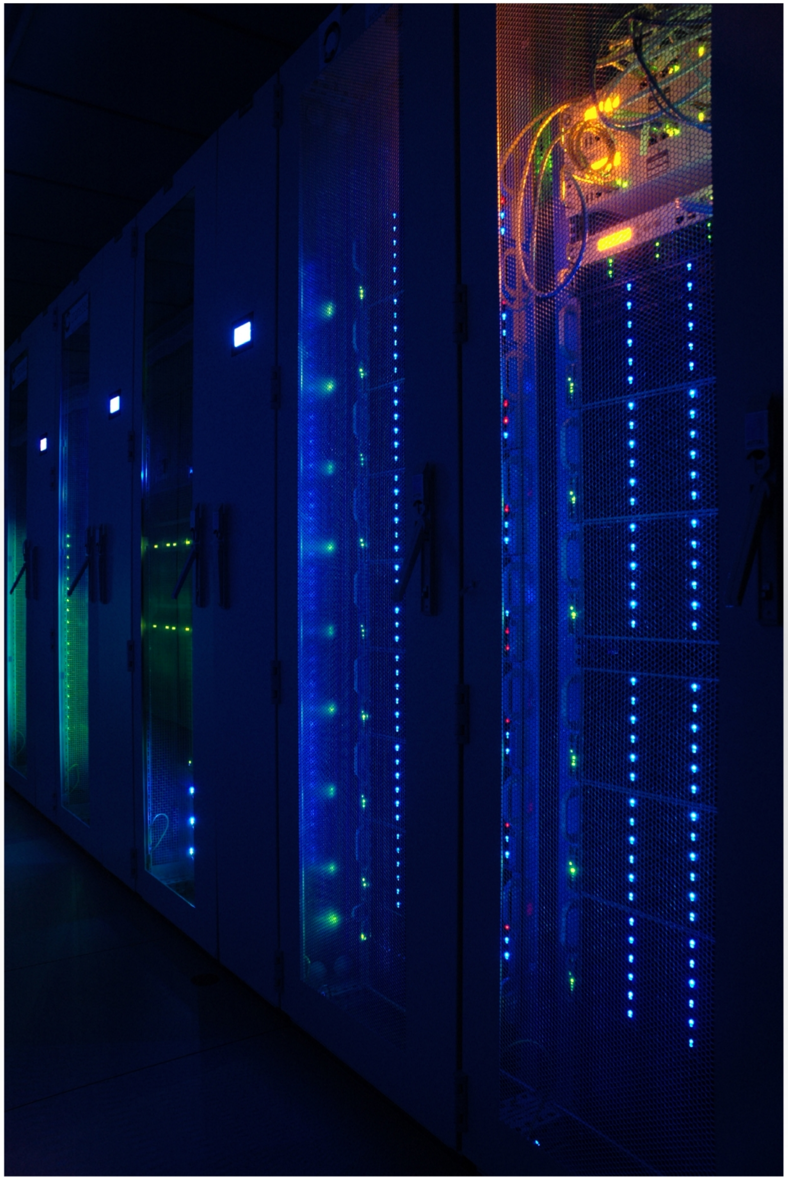
Download images via Healpix / HiPS



Training Model

on local GPU – Cluster

- several days
- exclusive hardware



KD Schema

provide the necessary tools to access the data and transfer only the required parts

consider the usage of optimized hardware

provide the data for processing & schedule, optimize and distribute the workload

transfer / post-process data / collect, generate provenance

provide analysis capabilities

select training objects → extract training data → train model → use model → save results → analyze results

- select training objects: find/use/merge data, annotate objects, statistical analysis
- extract training data: SIA / SSA / HIPS ...
- train model: CPU / GPU, IO / storage, special purpose HW
- use model: data-access, compute, GWS
- save results: storage, VO-space, provenance, DM
- analyze results: remote visualization, statistical tools