



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

AI & ML applications in Biophysical problems

Daniel Remondini

DIFA – daniel.remondini@unibo.it

AI in biophysics – our group

We are involved since 20y in biomedical data analysis & modelling, with many national and EU projects and collaborations (Systems Medicine, Oncology, Ageing, Agrofood, Public Health, Biophysical modelling):

VEO - Versatile Emerging infectious disease Observatory

E-MUSE – Complex microbial ecosystems multiscale modelling

GenoMed4all – Genomics & Personalized Medicine through AI

INC-COST – International Nucleosome Consortium COST action

AIM (INFN) - Artificial Intelligence in Medicine



Background: biophysics

Heterogeneous data types: imaging, omics data (mRNA, DNA, metabolites, proteins, 3d structure of proteins and DNA, ...)

Many possible tasks: data processing, classification/clustering, regression, image segmentation & enhancement, optimal embedding

Many biological data are badly conditioned (many variables few samples), with nontrivial noise (pdf) and relations (hard to model with simulations)



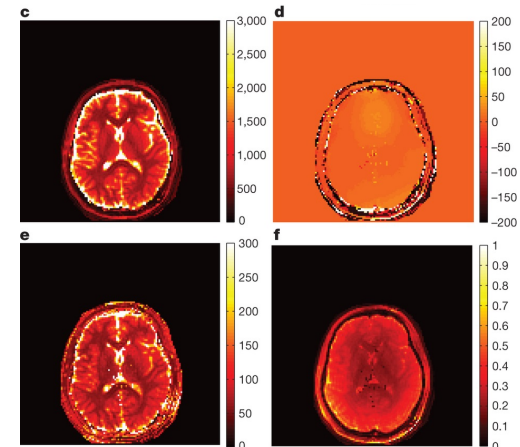
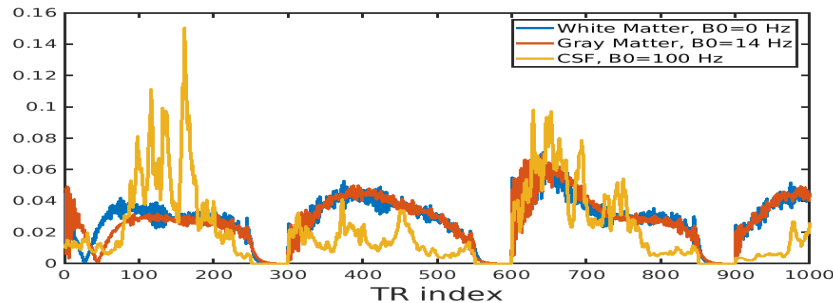
MR fingerprinting with DL: extracting physical parameters

FINGERPRINT approach to MR physical parameter estimation:

- generate response **patterns** to specific MR sequences
- estimate parameters with a known pattern-parameter **dictionary**

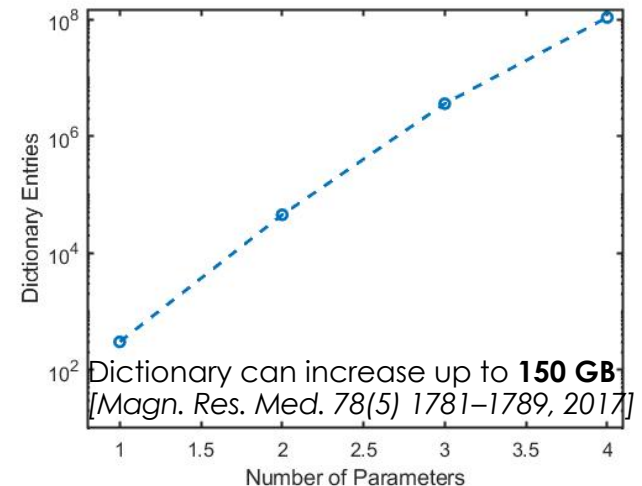
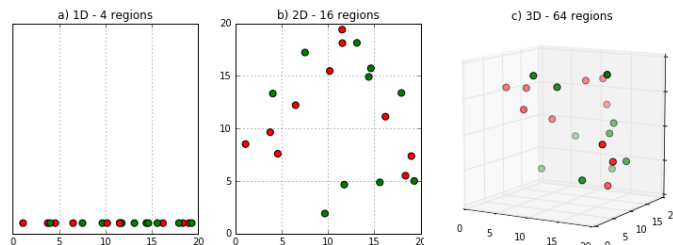
Magnetic resonance fingerprinting

Dan Ma¹, Vikas Gulani^{1,2}, Nicole Seiberlich¹, Kecheng Liu³, Jeffrey L. Sunshine², Jeffrey L. Duerk^{1,2} & Mark A. Griswold^{1,2}
14 MARCH 2013 | VOL 495 | NATURE | 187

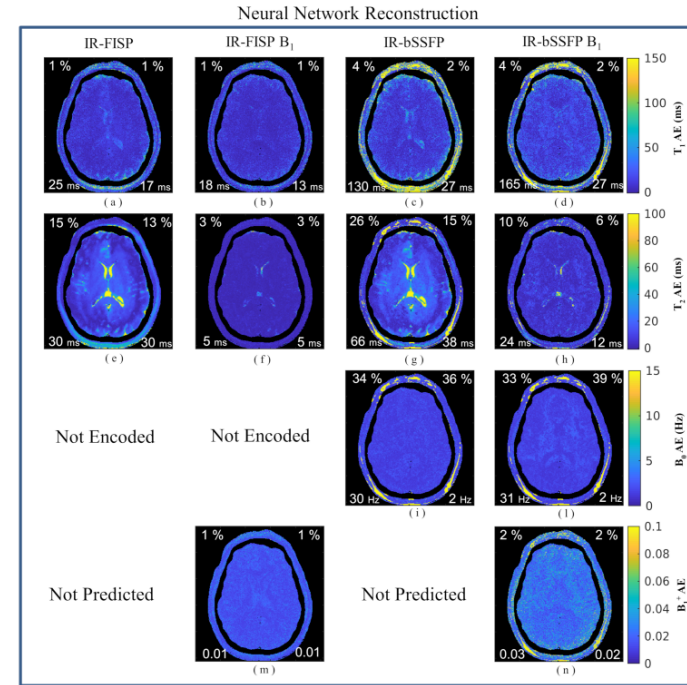
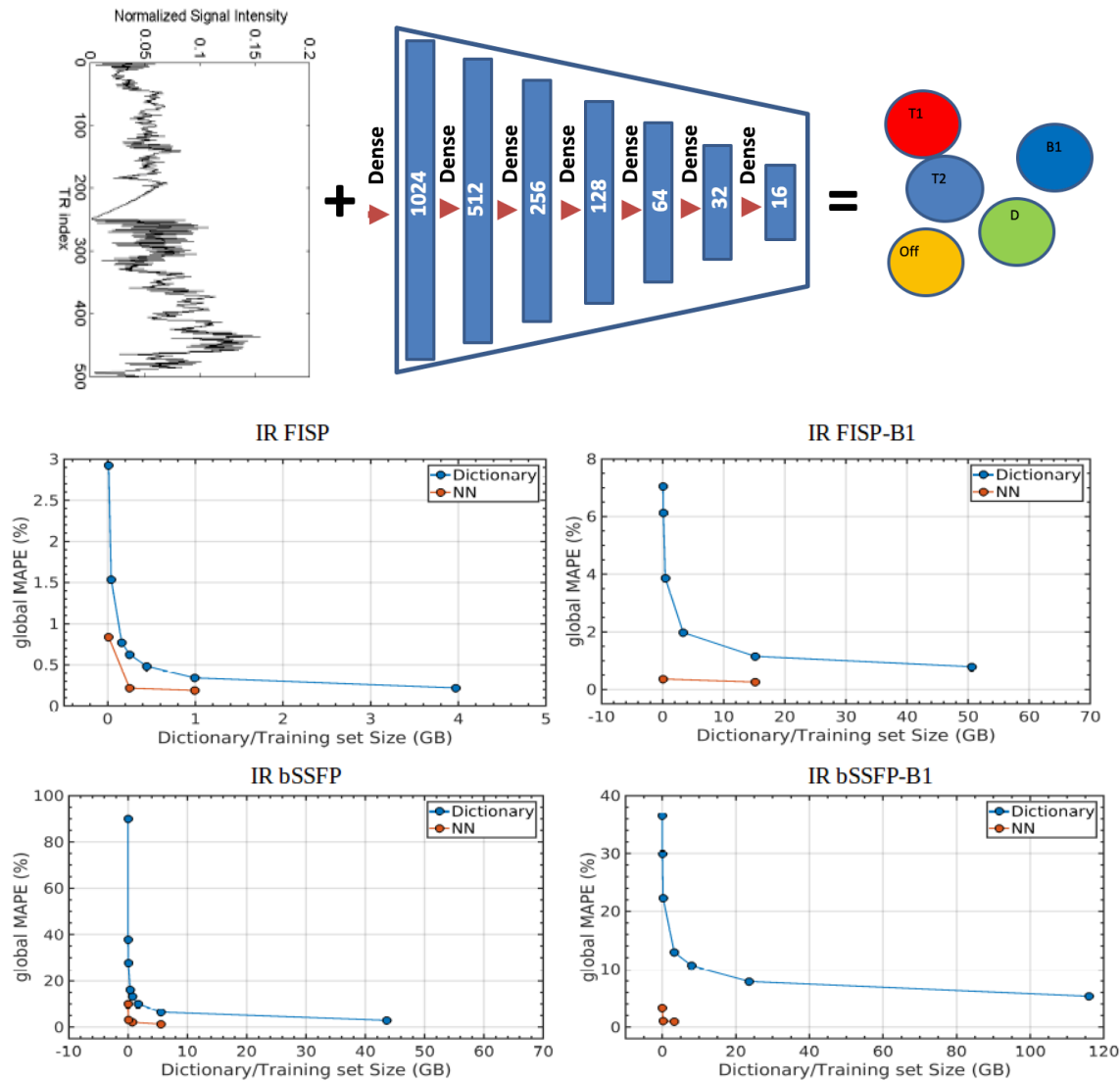


Drawbacks:

exponential growth of dictionary size
Time consuming mapping



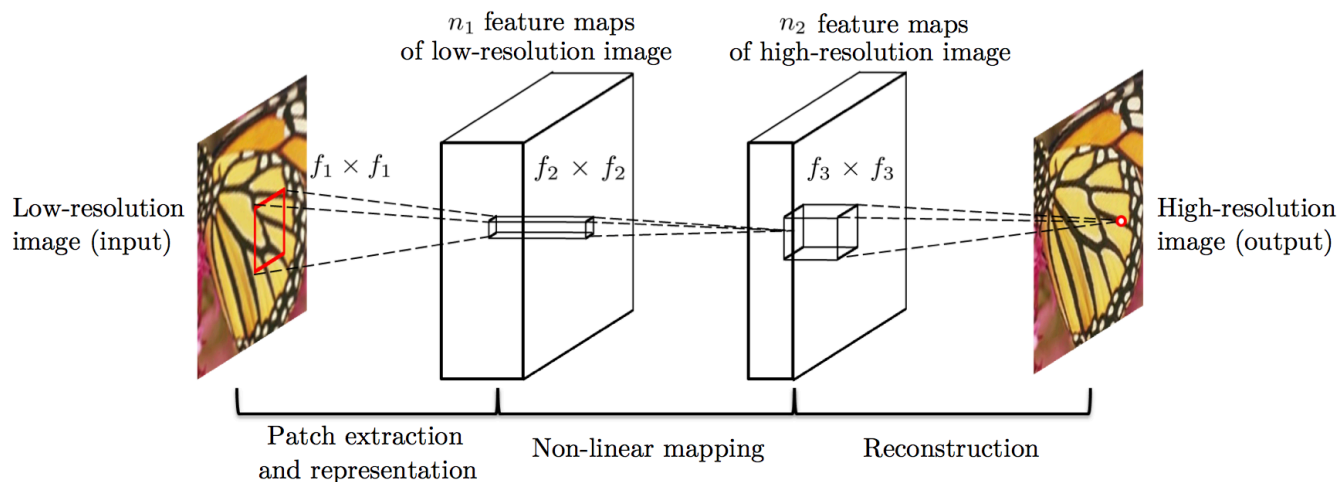
DL learns the transfer function: estimate parameters from feature vectors



Super resolution imaging

improve image quality (i.e. pixel density):

We have (re)implemented pre-trained WDSR-CNN that allow **x2, x4, x8** super resolution, for application to biomedical images



Original image 0861 from DIV2K



HR
(PSNR/SSIM)



Bicubic
(21.23 dB/0.648)



WDSR



EDSR

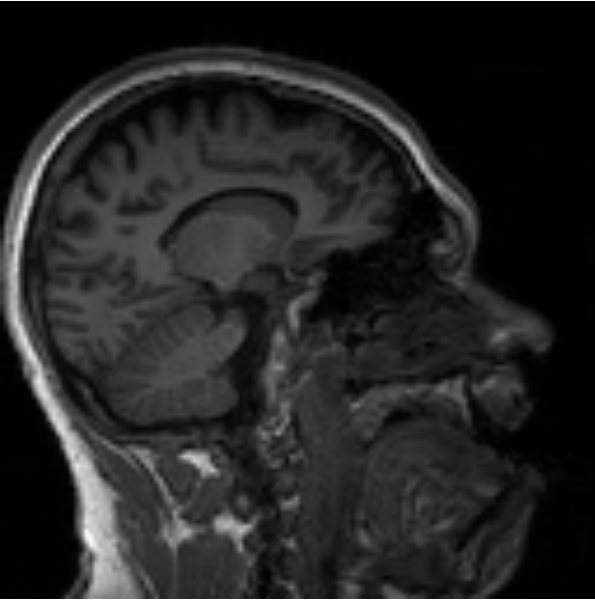
(23.17 dB/0.776)(23.36 dB/0.783)

- DIV2k training set
- 10^5 parameters
- 28x28 patches

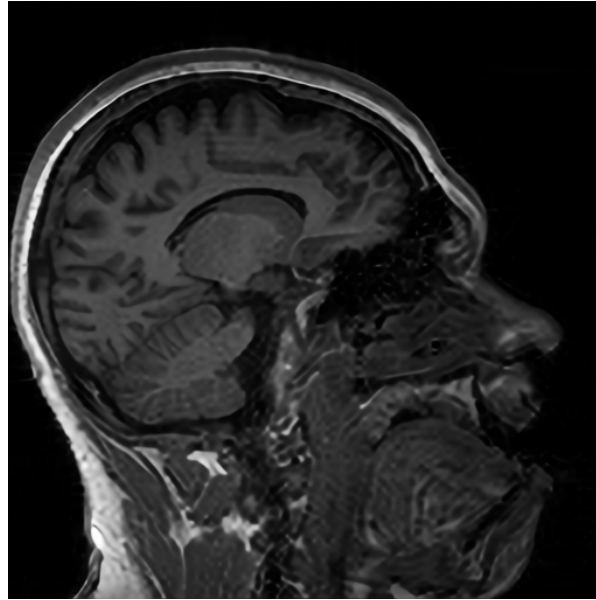


ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Raw (128x128)

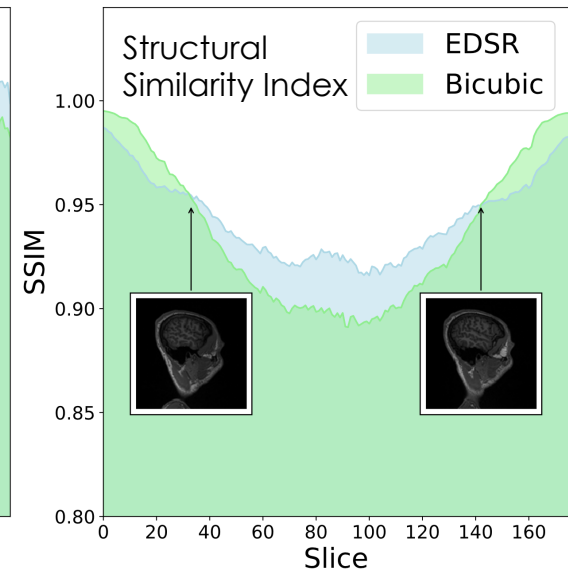
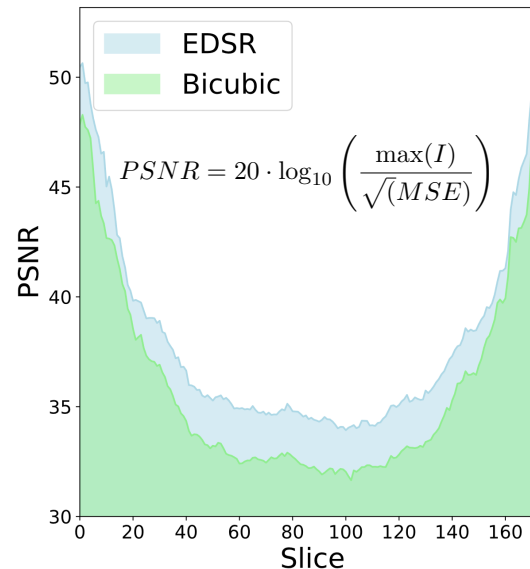


4x Super-Res (512x512)

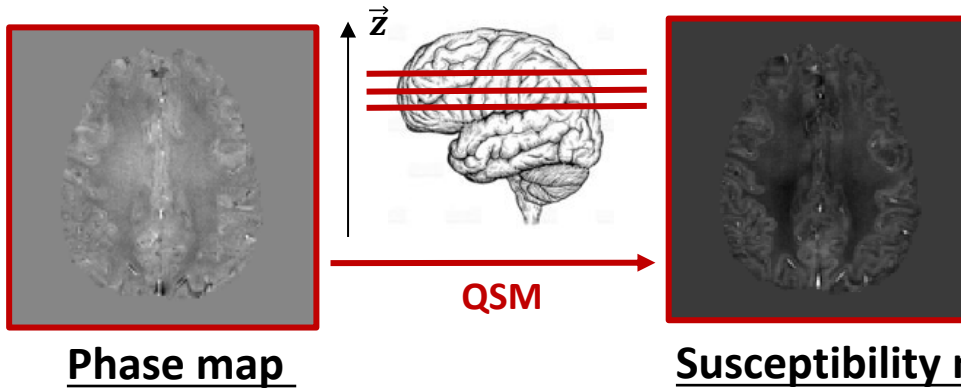


Biomedical MR Imaging

NN is better in the central slices of scan (brain) and worse in the side slices (skull)
SR Training on natural images learned complex shapes/contours/structures



Quantitative Susceptibility Mapping



- Susceptibility $\chi(r)$: response to B_0
- Biomarkers: water, myelin, iron, calcium
- Useful for Neuroimaging

Mathematical problem: given phase map in k-space convert to χ

$$\Delta B(k) = B_0 \cdot \chi(k) \cdot \left(\frac{1}{3} - (\hat{k} \cdot \hat{z})^2 \right) \quad \longrightarrow \quad \chi(k) = \frac{\Delta B(k)}{B_0 \cdot D(k)} \quad \begin{matrix} D(k) = 0 \\ (\hat{k} \cdot \hat{z}) = \pm \frac{1}{\sqrt{3}} \\ \vartheta_m = 54.7^\circ \end{matrix}$$

Ill-posed problem: inversion has **singularities** in k-space (magic angle)

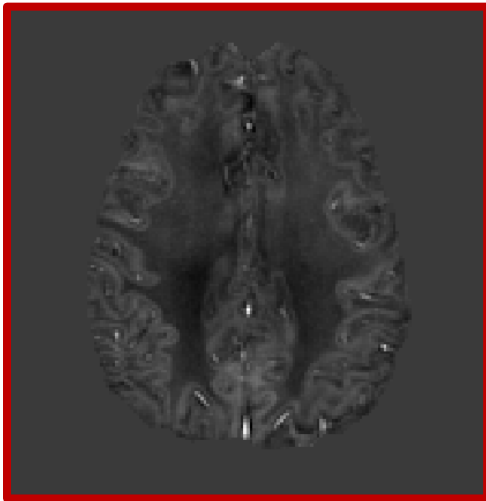
Golden standard solution: combine multiple acquisitions (COSMOS)

Limits: time consuming

TKD and COSMOS

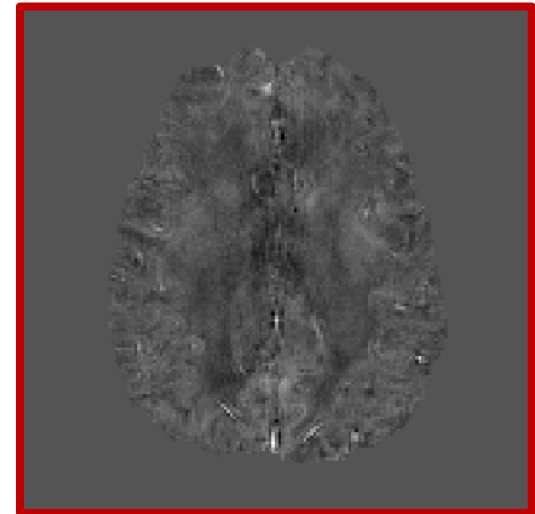
COSMOS - Calculation Of Susceptibility through Multiple Orientation Sampling

- **Multiple head-orientation acquisition**
- Long acquisition time
- Uncomfortable for the patient
- Accurate and precise reconstruction



TKD - Truncated K-space Division

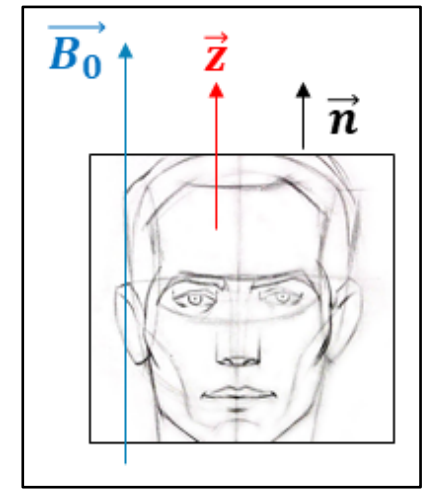
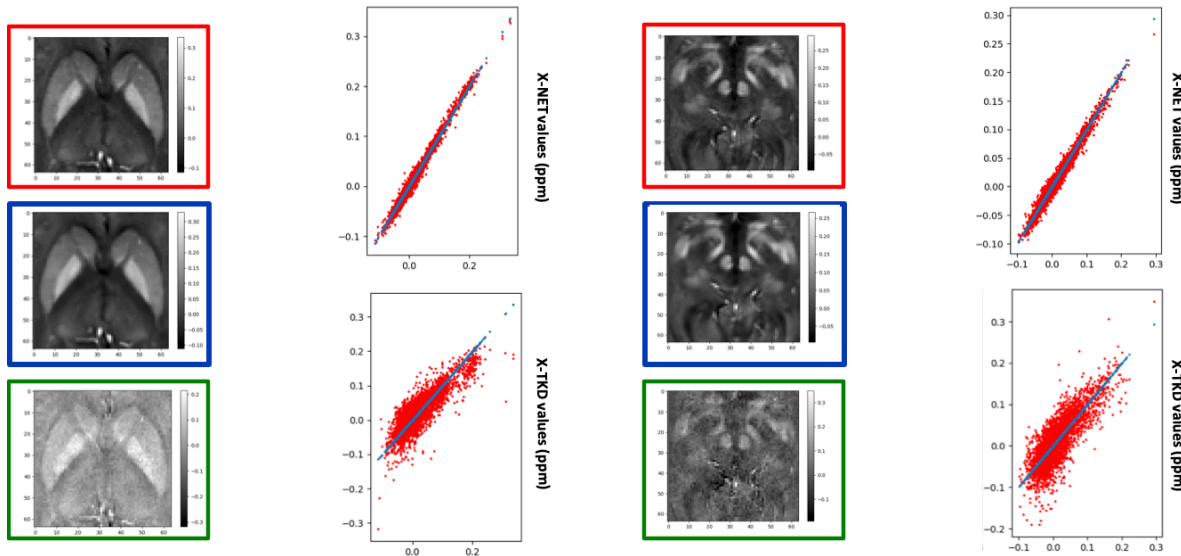
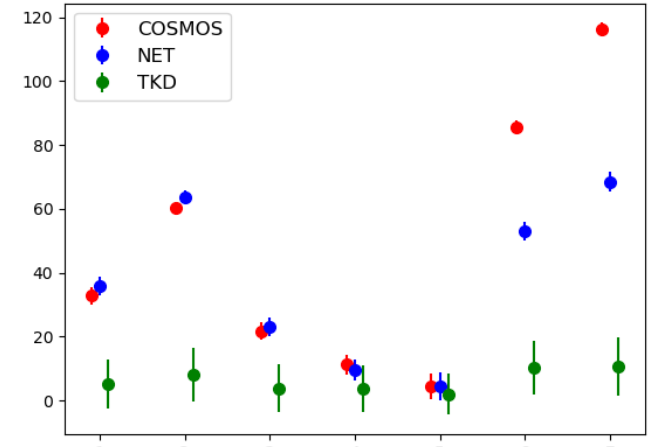
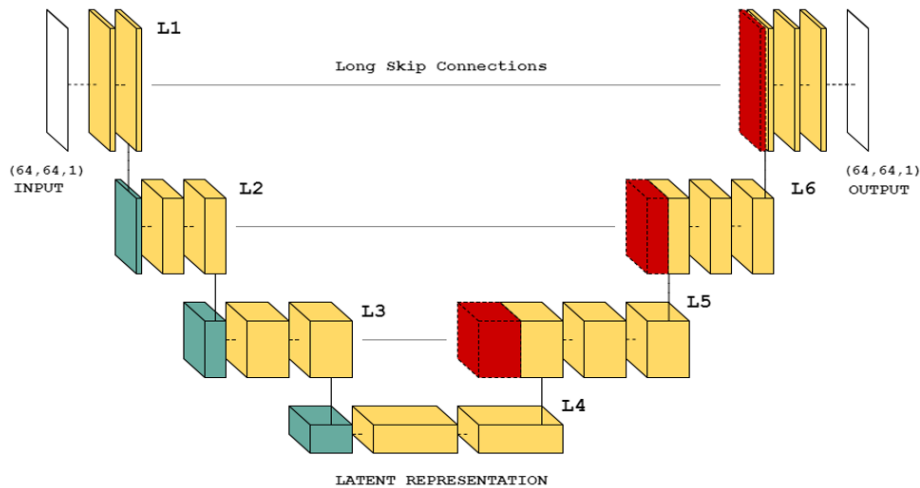
- **Single head-orientation acquisition**
- Short acquisition time
- Numerical strategy: k-space cutoff
- Noisy reconstruction



COSMOS: “golden standard” for QSM
- requires multiple acquisitions



CNN learn COSMOS output using **single orientation** data



Better agreement between COSMOS and CNN



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

NLP & protein sequences

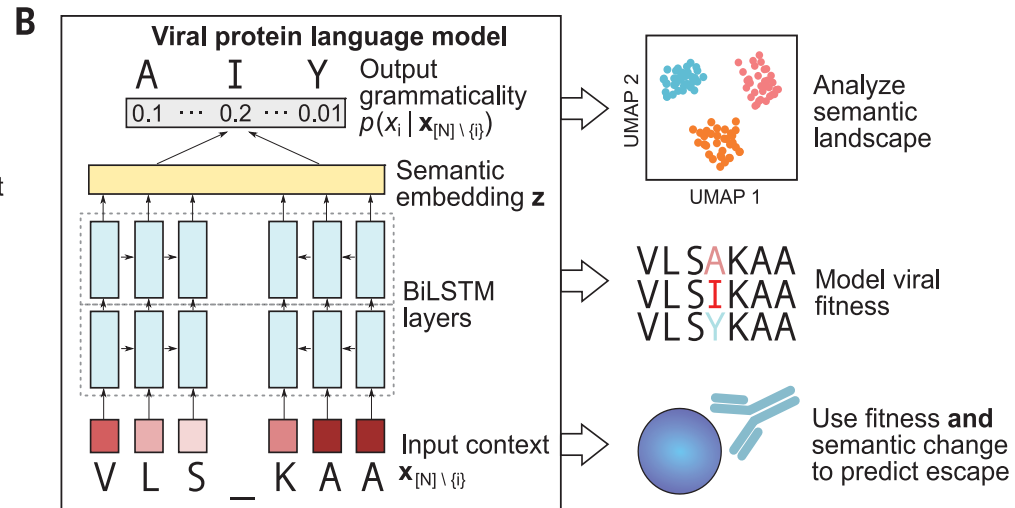
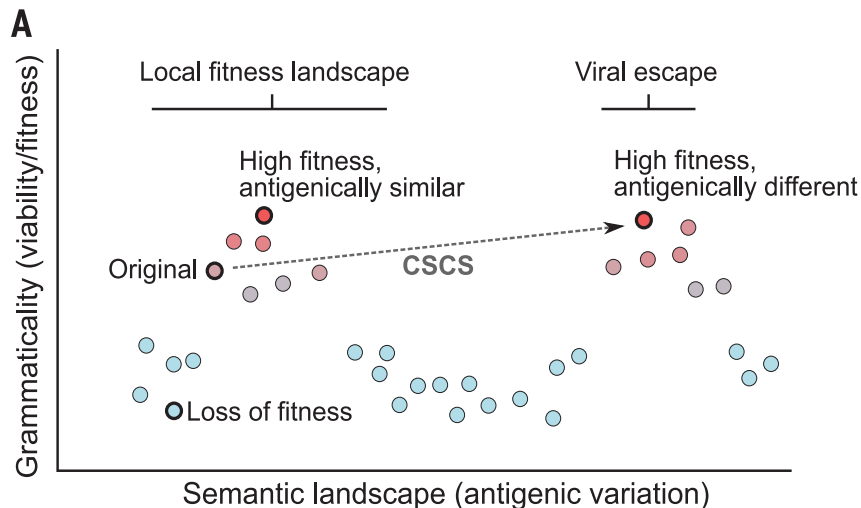
H2020 VEO Versatile Emerging infectious disease Observatory

WP2 Data-Mining tools: UNIBO co-leader

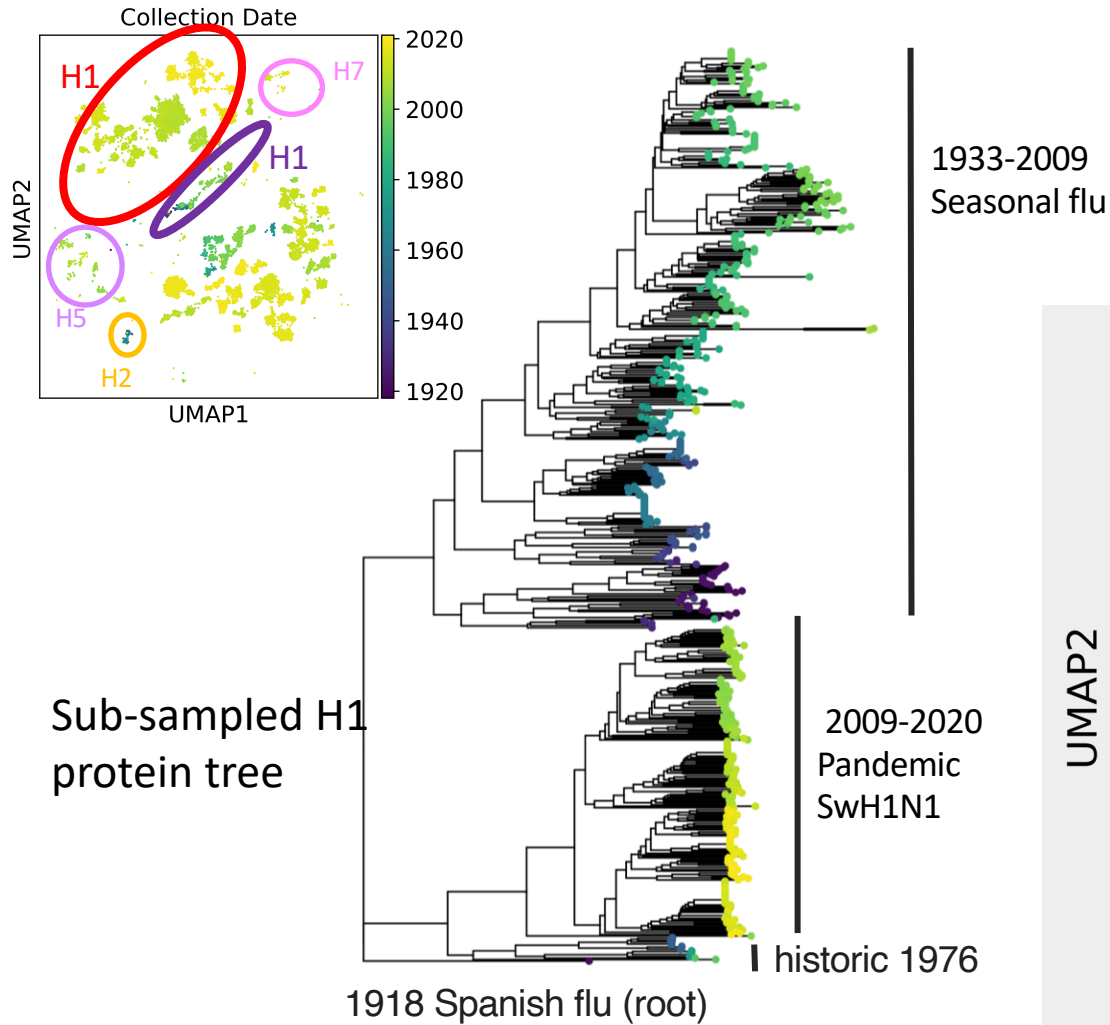
(Prof. S. Lycett, Univ. Edinburgh UK)



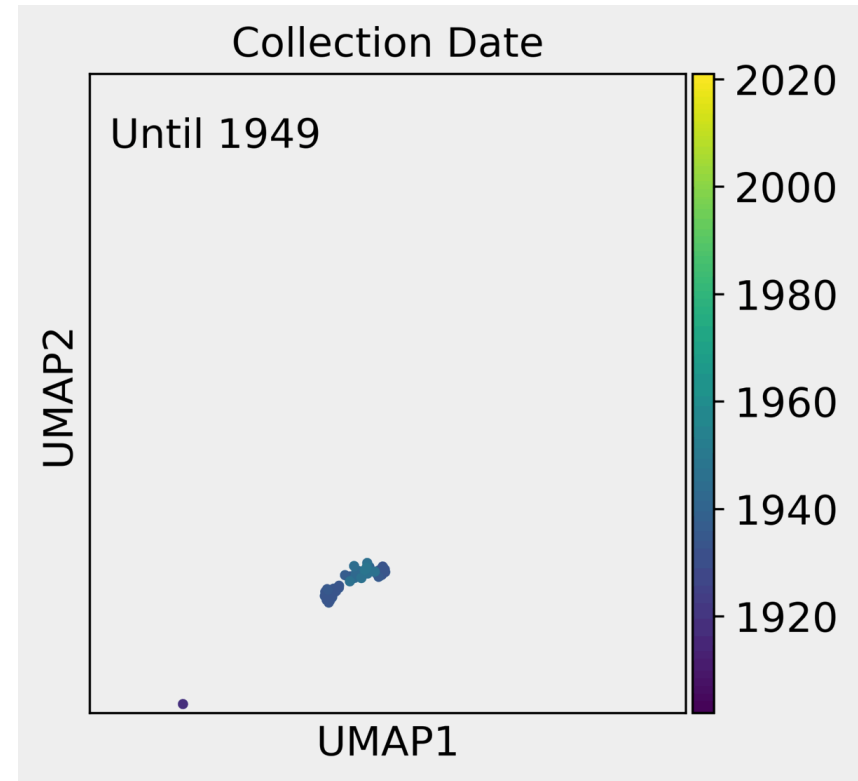
- Background: protein sequences as *strings* of symbols
- Aim: infer knowledge about virus from sequence alone
- We applied NLP AI algorithms (bi-LSTM and BERT-like) to encode protein sequences into numerical vectors ("prot2vec") [Hie et al., Science 2021]



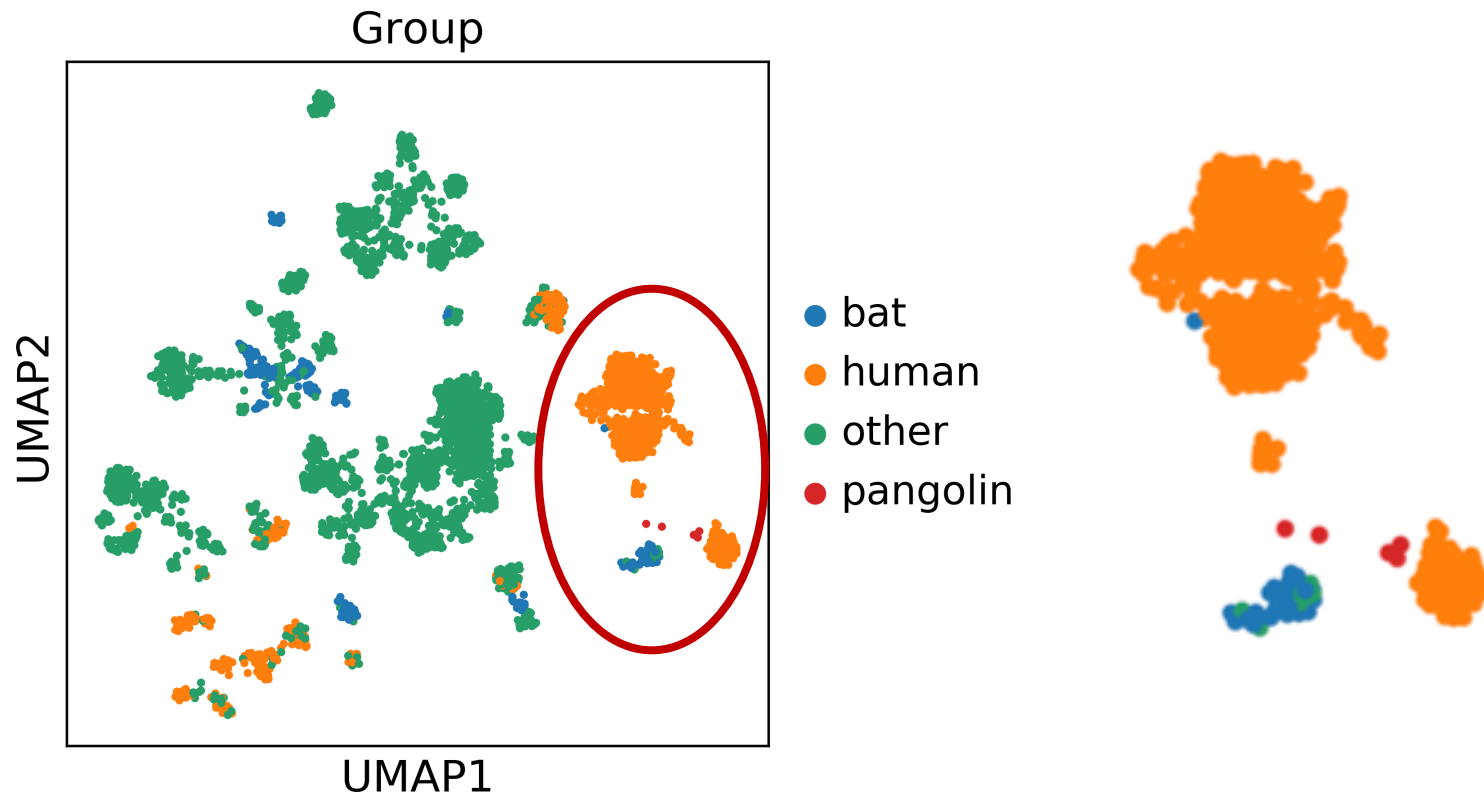
Human influenza strain evolution



- Evolution history of H1 Human flu in embedded space



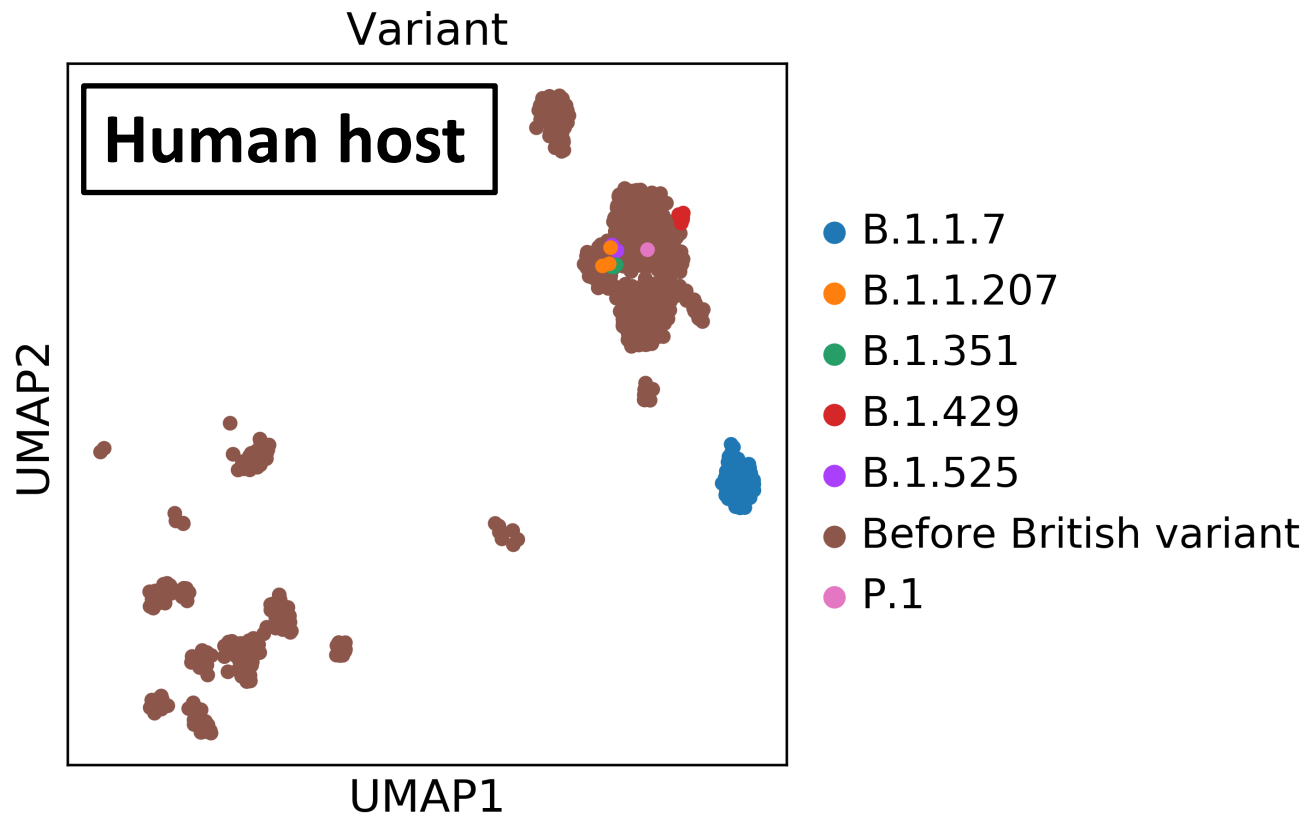
SARS-CoV-2 Spike protein (up to Sept '20 – GISAID website data)



Embedding of different hosts



Adding recent variants: UK B.1.1.7



- B.1.1.7 has N501Y and deletion in spike protein
- Shows as its own 'distant' cluster here
- B.1.1.7 **not** in training data at all

Physical issues

- DL encoders (CNN, Transformers , " x^2 vec) provide tools for dimensional embedding, just as recent spectral tools (ISOMAP, UMAP, based on discrete Laplace-Beltrami operator on networks) trying to reconstruct the underlying manifold in which data lie (link to Ricci flows & Heat kernels)
- Math/phys and AI research can complement on these topics, e.g. with applications to self- and semi-supervised learning





Department of Physics and Astronomy – DIFA

Credits to all the PhDs and Postdocs in our group, in particular:

Marco Barbieri for MRF (Prof. C. testa)

Nico Curti for Super-Res (Prof. G. Castellani)

Cristiana Fiscone for QSM (Prof. C. testa & Prof. R. Bowtell)

Francesco Durazzi and Lorenzo Dall'Olio for NLP (Prof. S. Lycett)



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA