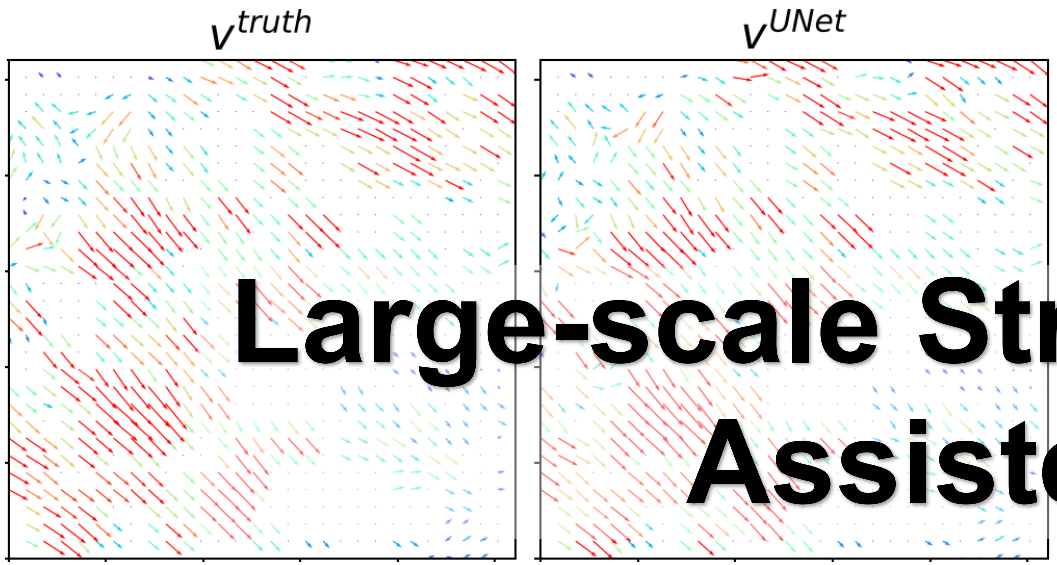
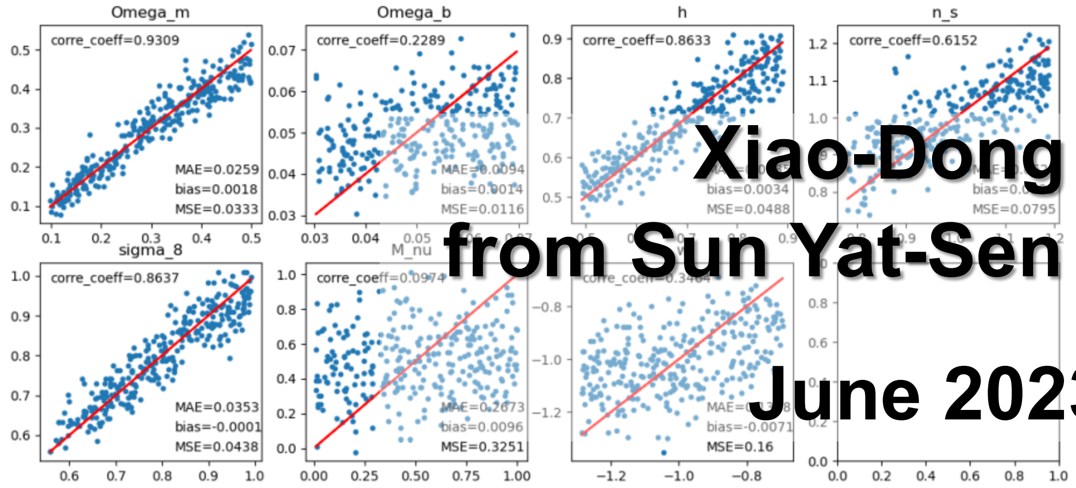
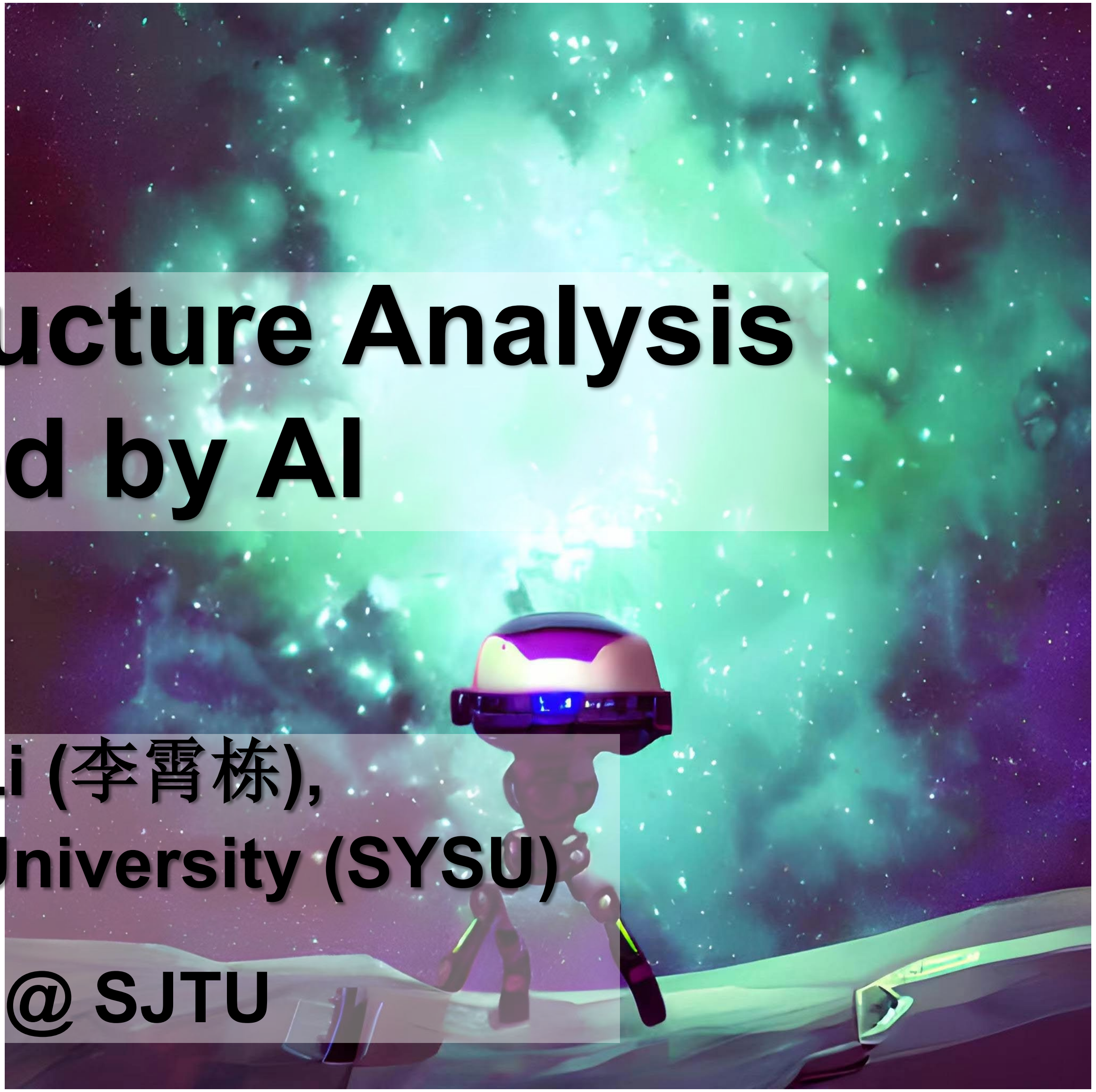




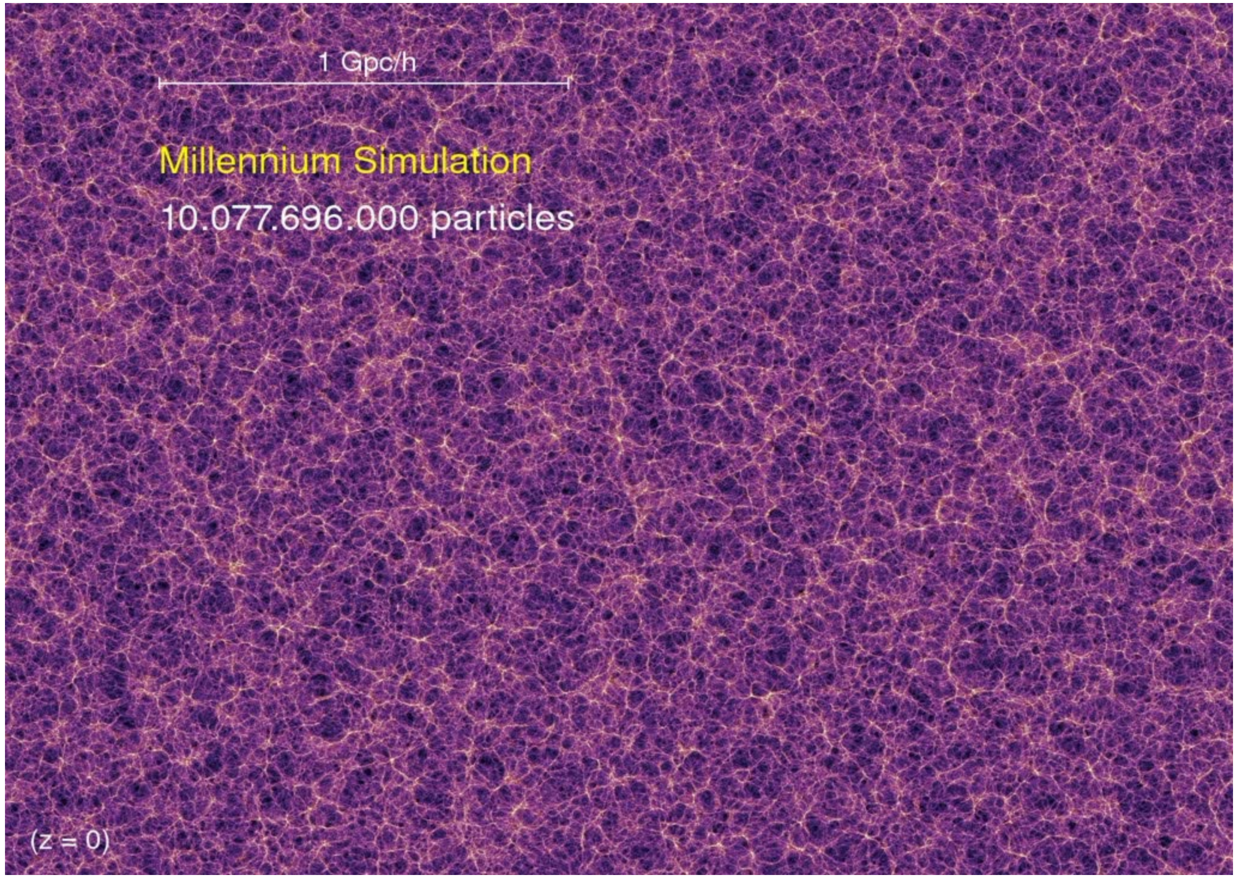


Large-scale Structure Analysis Assisted by AI

Xiao-Dong Li (李霄栋),
from Sun Yat-Sen University (SYSU)

June 2023 @ SJTU

Motivation of AI



The Extremely Complicated LSS

Team Members (Machine Learning for Cosmology)

Nicola R. Napolitano
SYSU

Xiao-Dong Li
SYSU

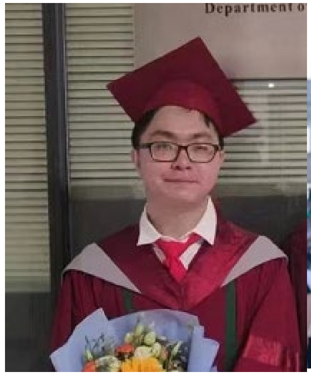
Yang Wang
PengCheng Lab

Le Zhang
SYSU

Xin Wang
SYSU

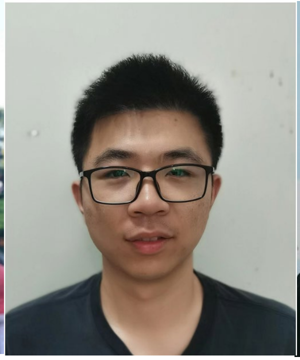
Xiaoru Li
South China Normal Univ.

Jinqu Zhang
South China Normal Univ.



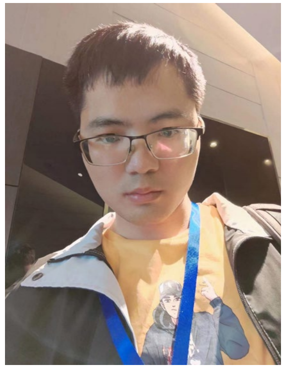
Xiaolin Luo
SJTU

Ziyong Wu
USTC



Jiacheng Ding
SYSU

Liang Xiao
SYSU



Zhiwei Min
SYSU

Qian Li
SYSU

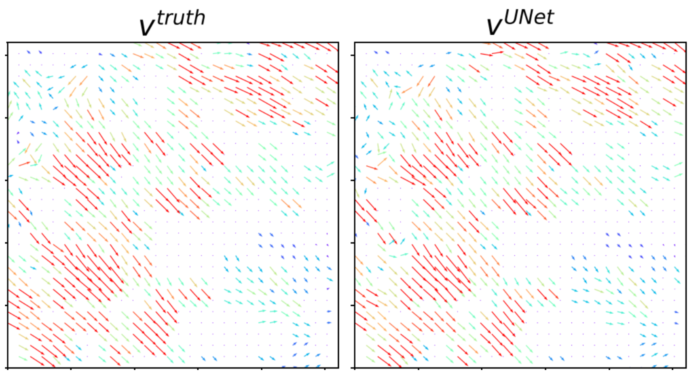
Zhujun Jiang
SYSU

Wenying Du
SYSU

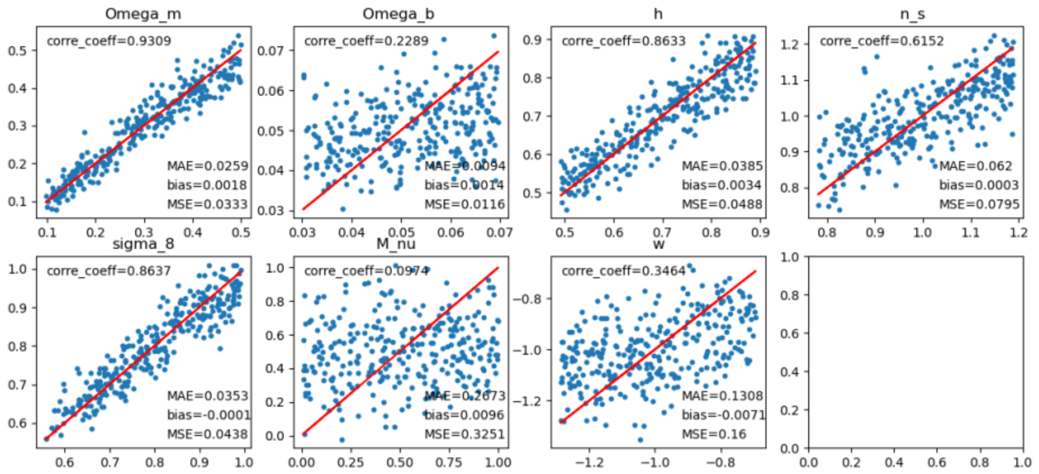
Xu Xiao
SYSU

Outline

- Velocity Reconstruction

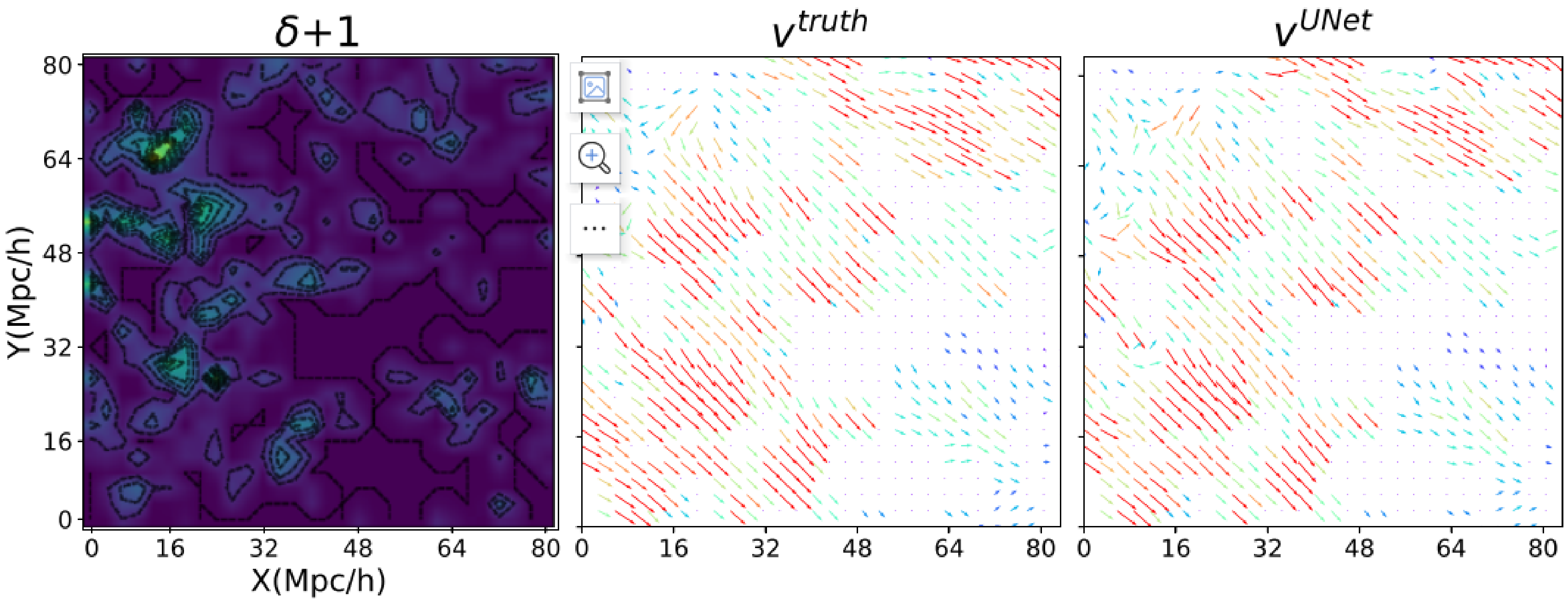


- Parameter Estimation

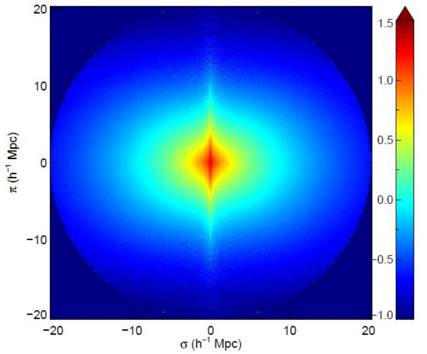


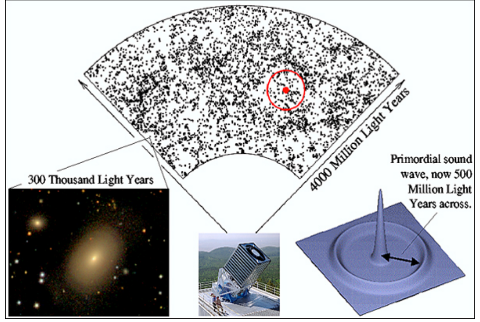
- Summary & Future

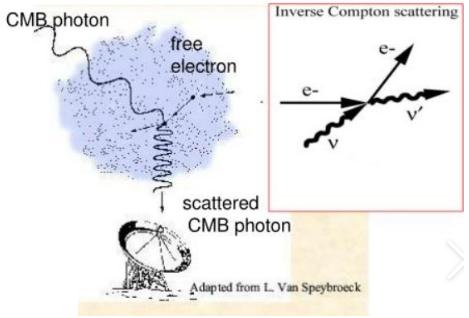
I. Velocity Reconstruction

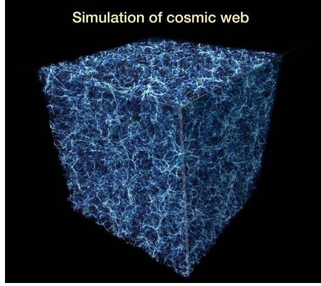


Motivation

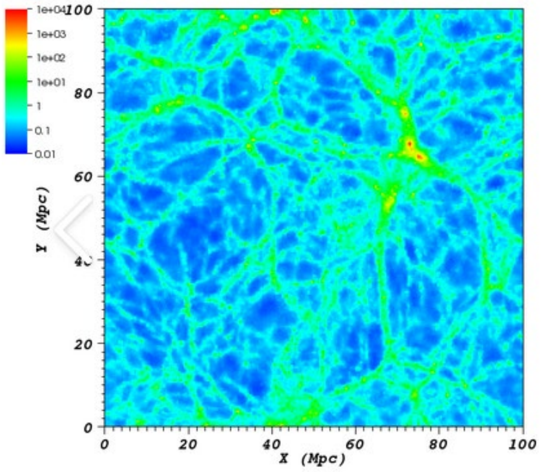








- Dynamics of objects in the Universe
- Redshift space distortion, AP test
- BAO reconstruction
- kSZ
- Cosmic Web



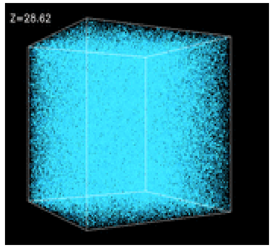
$$\dot{\delta} + \frac{1}{a}\nabla\cdot\mathbf{v} = -\frac{1}{a}\nabla\cdot(\mathbf{v}\delta)\ , \qquad \nabla\cdot\dot{\mathbf{v}} + H\nabla\cdot\mathbf{v} + \frac{3H^2}{2}a\Omega_m\delta = -\frac{1}{a}\nabla\cdot[(\mathbf{v}\cdot\nabla)\,\mathbf{v}]\ , \qquad \nabla^2\phi = 4\pi G\rho\ .$$

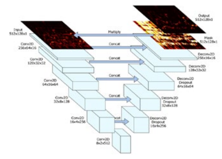
But velocities are difficult to observe !

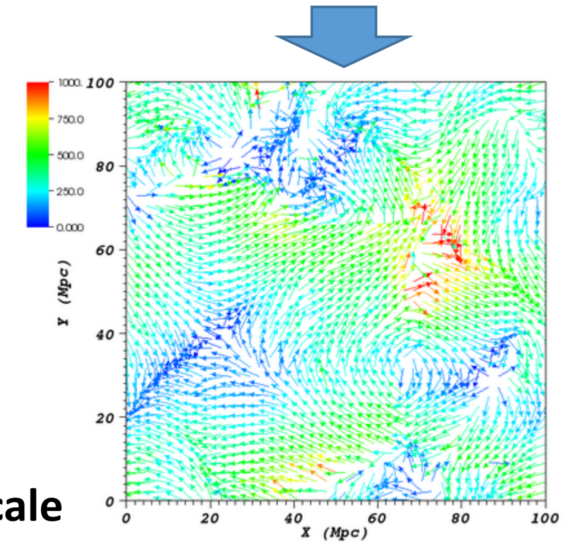
(z-independent determination of distance,

e.g. SNIa; fundmental plane, Tully-Fisher, ...)

-> Reconstruct it (from density)







AI advantage: Non-linear, small scale

Velocity Reconstrunction (of DM particles)

Ziyong Wu et al., 2021, ApJ (eprint: 2105.09450)

THE ASTROPHYSICAL JOURNAL, 913:2 (10pp), 2021 May 20 https://doi.org/10.3847/1538-4357/abf3bb

CrossMark

Cosmic Velocity Field Reconstruction Using AI

Ziyong Wu[1], Zhenyu Zhang[1], Shuyang Pan[1], Haitao Miao[1], Xiaolin Luo[1], Xin Wang[1], Cristiano G. Sabiu[2,3], Jaime Forero-Romero[4], Yang Wang[1], and Xiao-Dong Li[1]

[1] School of Physics and Astronomy, Sun Yat-Sen University, Guangzhou 510297, People's Republic of China; lixiaod25@mail.sysu.edu.cn, wangyang23@mail.sysu.edu.cn, wangxin35@mail.sysu.edu.cn

[2] Department of Astronomy, Yonsei University, 50 Yonsei-ro, Seoul 03722, Republic of Korea

[3] Natural Science Research Institute, University of Seoul, 163 Seoulsiripdaero, Dongdaemun-gu, Seoul 02504, Republic of Korea

[4] Departamento de Física, Universidad de los Andes, Cra. 1 No. 18A-10 Edificio Ip, CP 111711, Bogotá, Colombia

Received 2020 December 1; revised 2021 March 19; accepted 2021 March 29; published 2021 May 18

Abstract

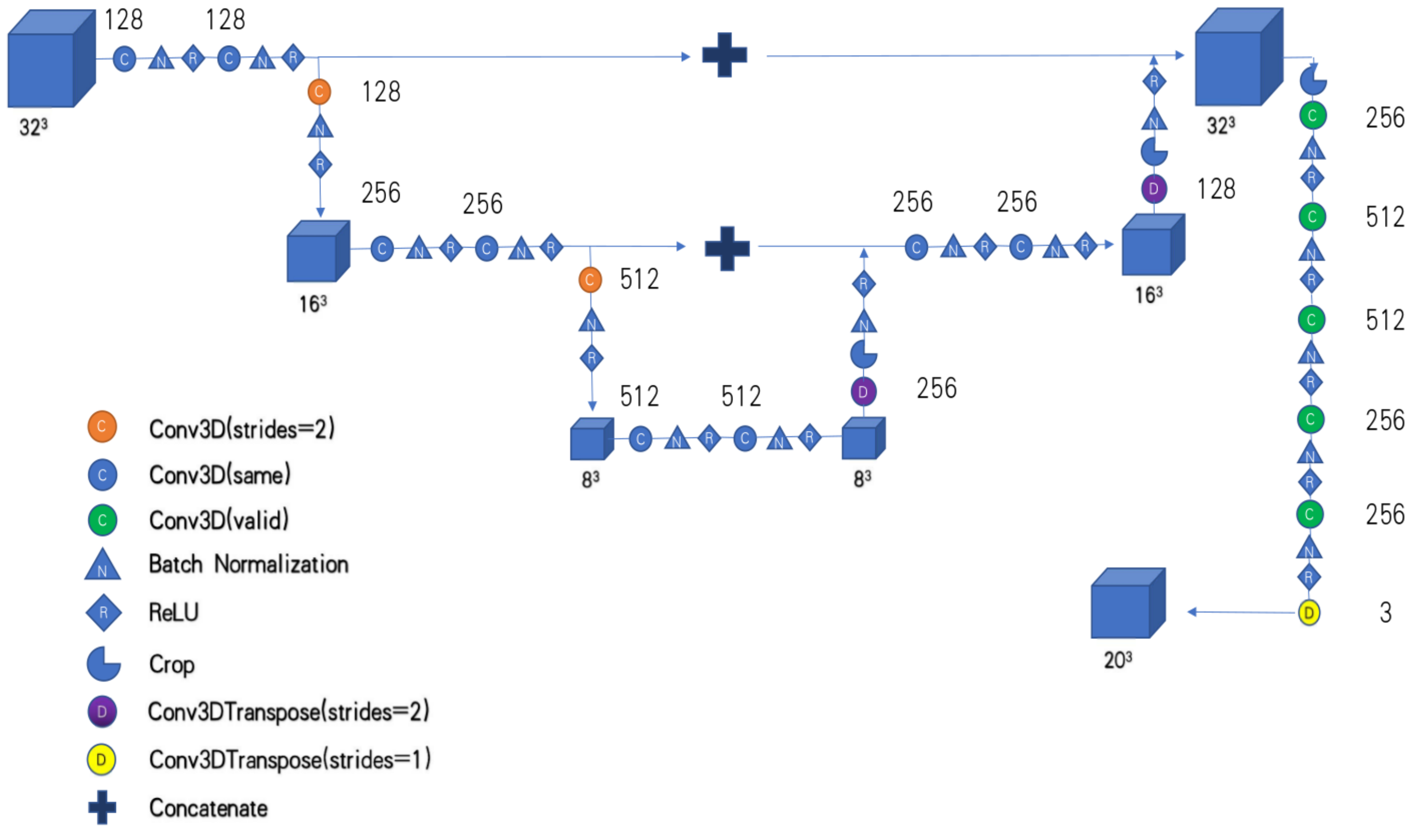
We develop a deep-learning technique to infer the nonlinear velocity field from the dark matter density field. The deep-learning architecture we use is a "U-net" style convolutional neural network, which consists of 15 convolution layers and 2 deconvolution layers. This setup maps the three-dimensional density field of 32^3 voxels to the three-dimensional velocity or momentum fields of 20^3 voxels. Through the analysis of the dark matter simulation with a resolution of $2h^{-1}$ Mpc, we find that the network can predict the the nonlinearity, complexity, and vorticity of the velocity and momentum fields, as well as the power spectra of their value, divergence, and vorticity and its prediction accuracy reaches the range of $k \simeq 1.4\ h\ \mathrm{Mpc}^{-1}$ with a relative error ranging from 1% to $\lesssim$10%. A simple comparison shows that neural networks may have an overwhelming advantage over perturbation theory in the reconstruction of velocity or momentum fields.

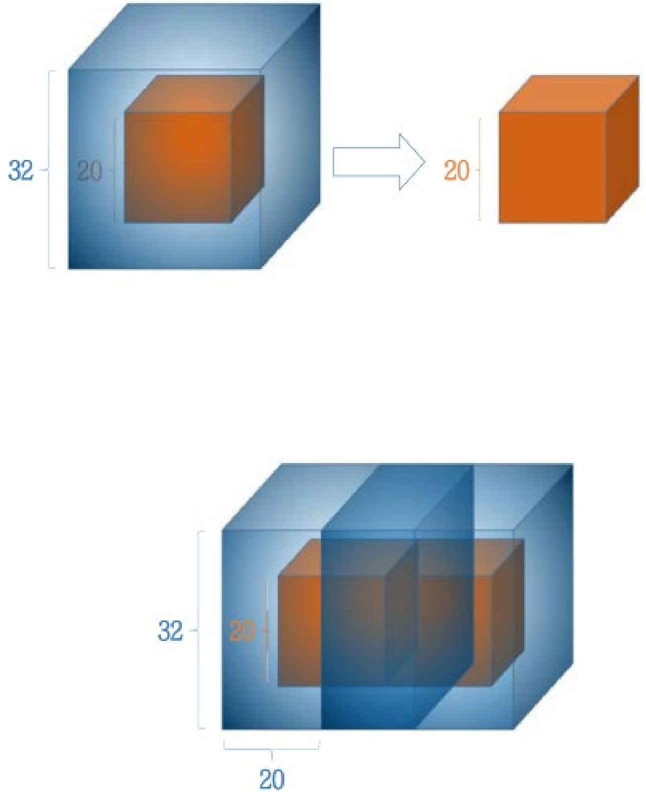
Unified Astronomy Thesaurus concepts: Large-scale structure of the universe (902); Cosmic web (330); Dark matter distribution (356); Cosmology (343); Astrostatistics (1882)

Ziyng Wu
Zhejiang Univ.

Zhenyu Zhang
Peking Univ.

Shuyang Pan
Sun Yat-Sen Univ.

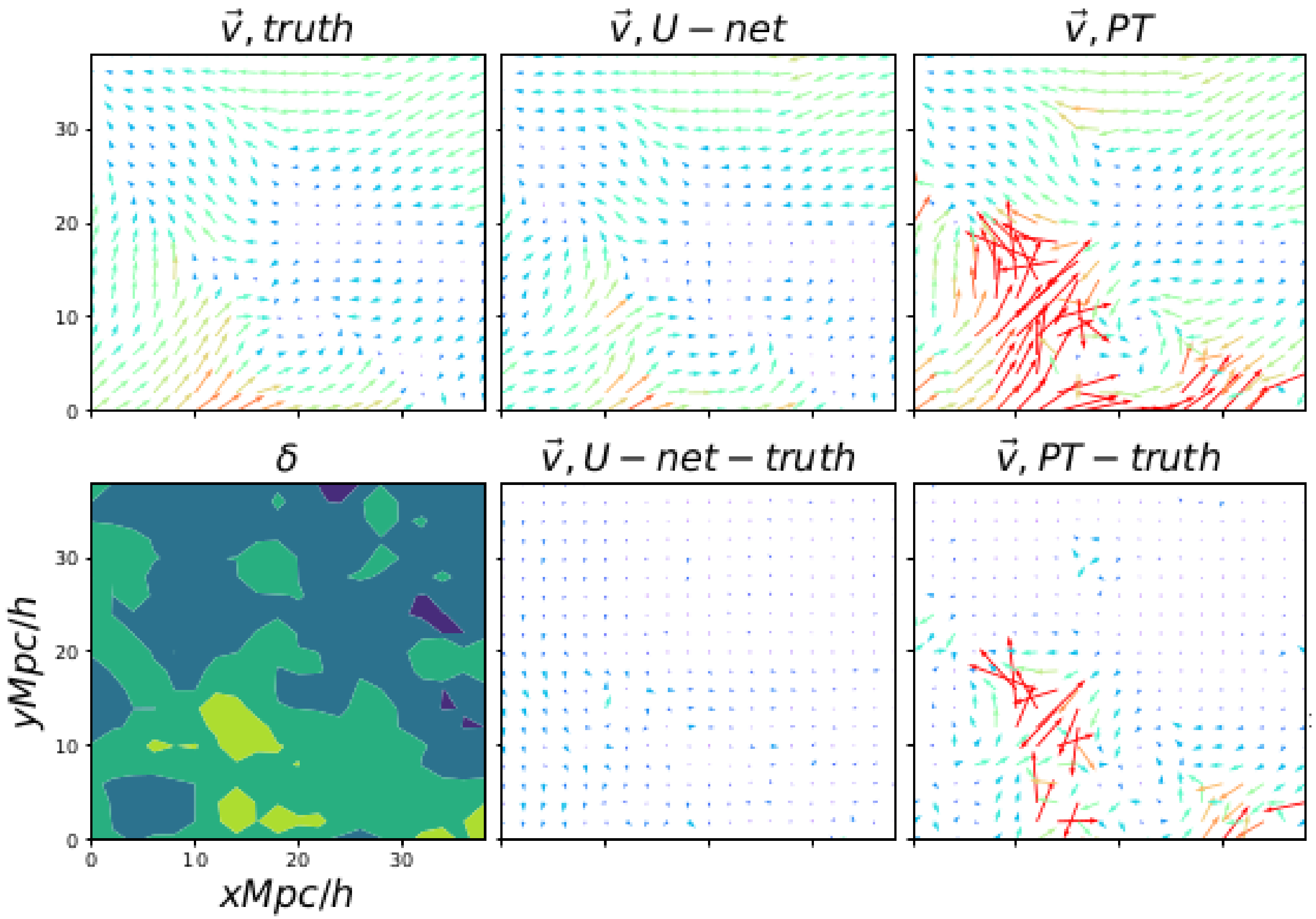




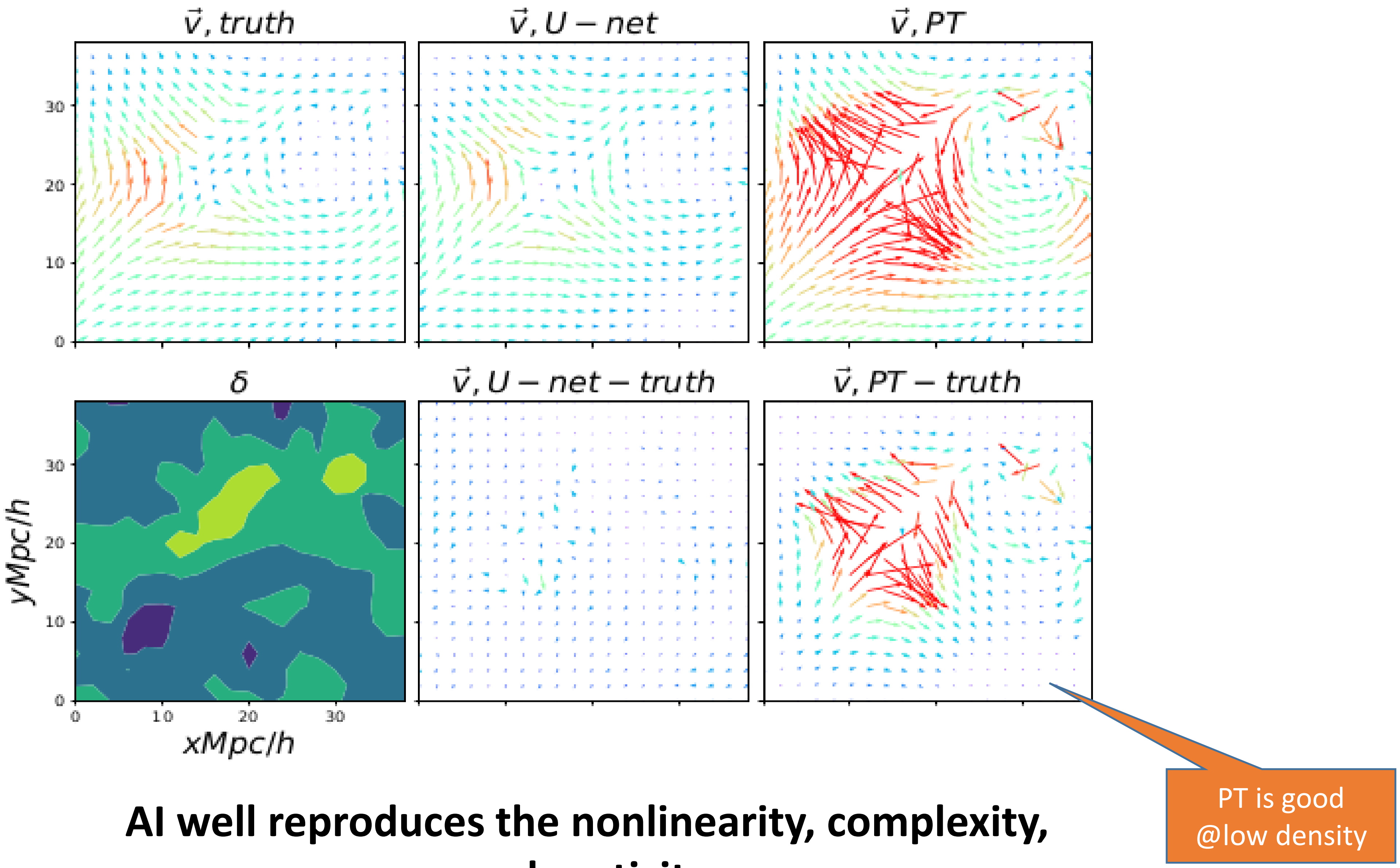
U-net model, transforming density to velocity

Input/output: CIC fields , **2Mpc/h grid**

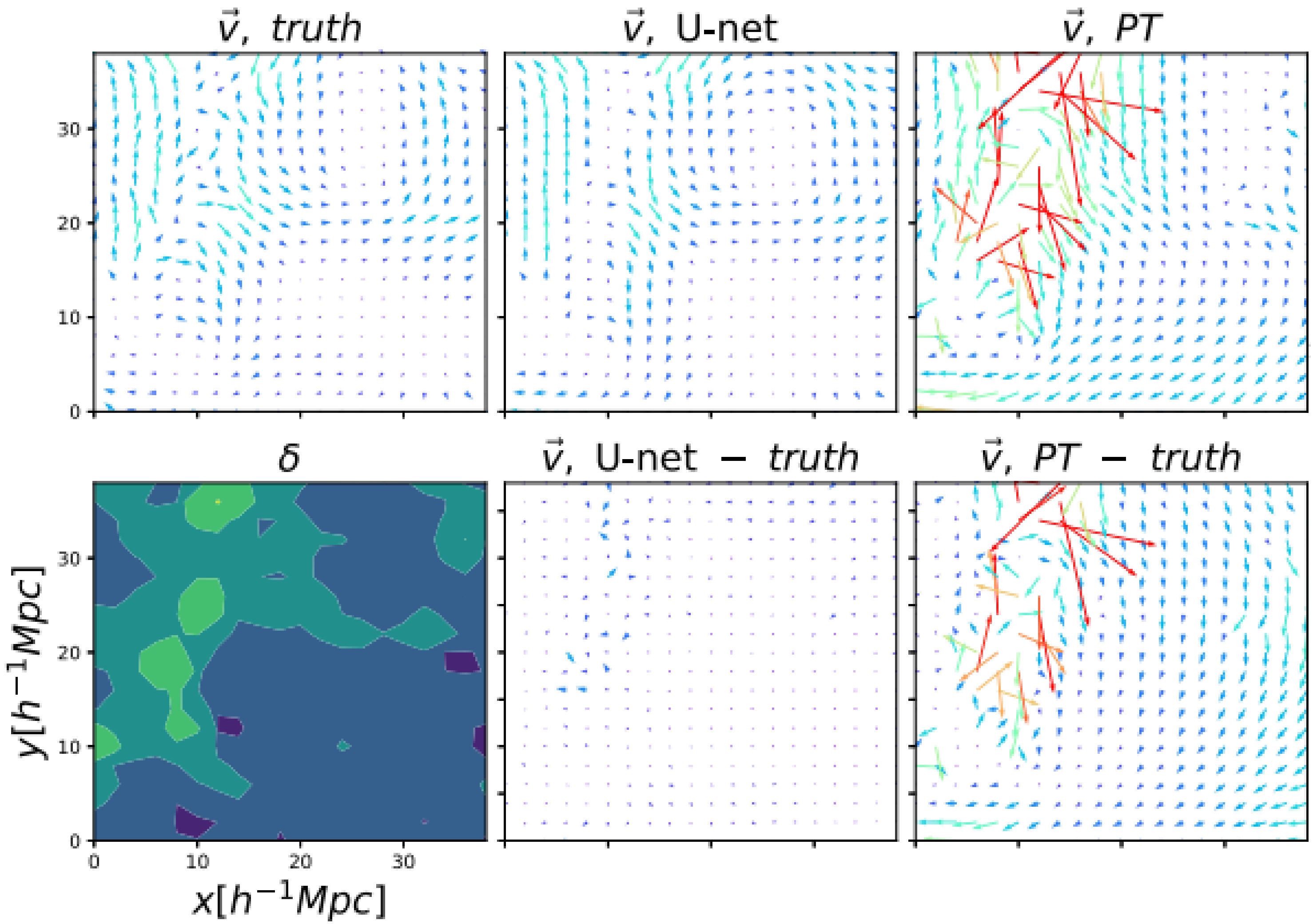
Trained on Tianhe2 GPUs



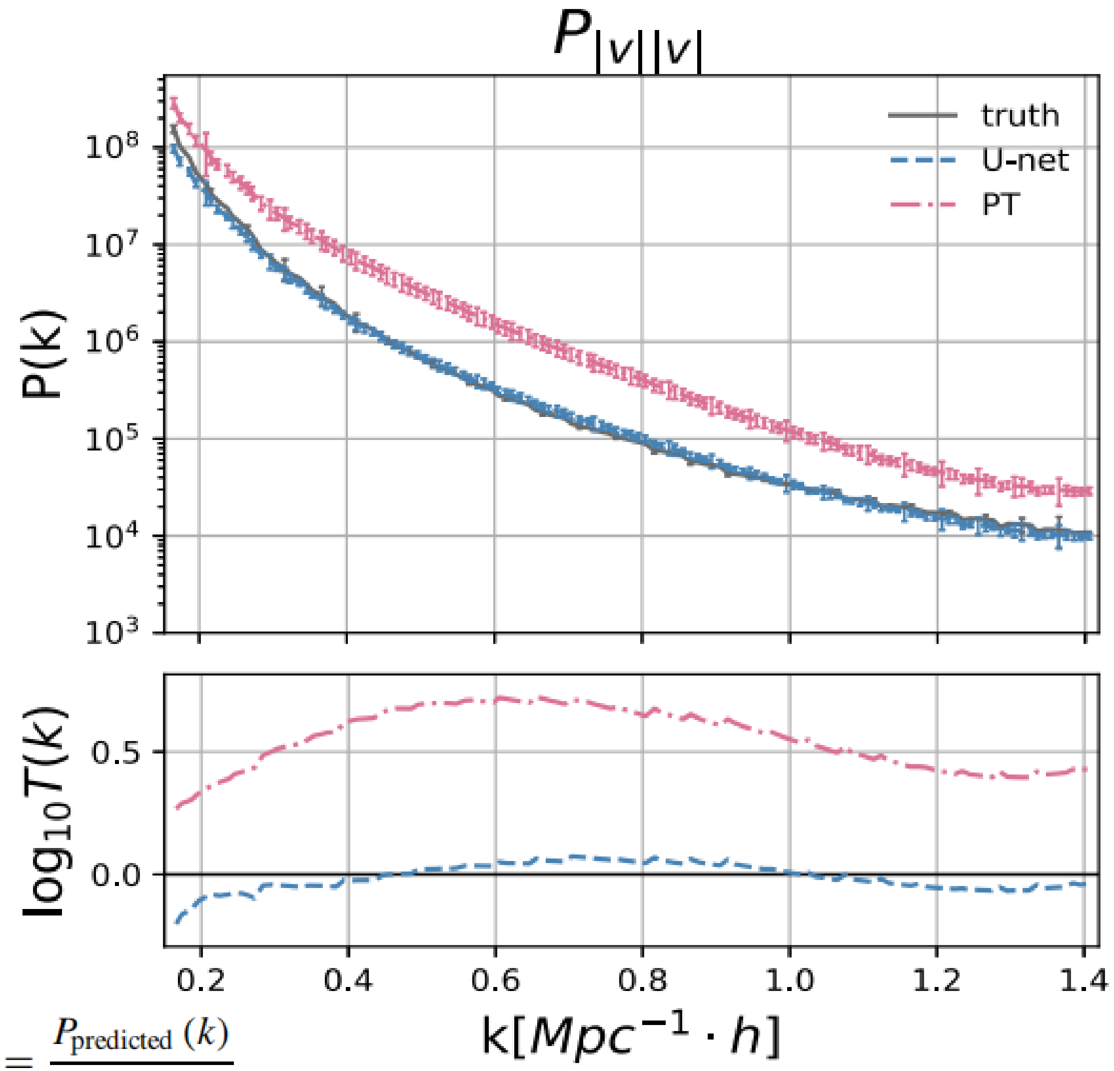
AI well reproduces the nonlinearity, complexity, and vorticity



AI well reproduces the nonlinearity, complexity, and vorticity



AI well reproduces the nonlinearity, complexity, and vorticity



$$T(k) = \frac{P_{\text{predicted}}(k)}{P_{\text{true}}(k)}$$

U-net accurately predicts the power spectra of |v|, |p|, as well as their div and curl

up to k~1.4 h/Mpc

error: 1%-10%

AAS | AAS Journals | BAAS | AAS-IOP eBooks | AAS Nova | AASTeX | Astronomy Image Explorer

AAS NOVA

Research highlights from the journals of the American Astronomical Society

HOME HIGHLIGHTS JOURNALS DIGEST

A Peculiar Use of AI: Predicting Cosmic Velocities with Neural Networks

By Astrobites on 2 June 2021 ASTROBITES

Share:

While the neural network used in this paper can definitely be improved — perhaps by further optimizing its architecture or by using more training data — the authors have shown that neural nets can be valuable tools for predicting peculiar velocities. With such programs as DESI, EUCLID, the Rubin Observatory, and the Nancy Grace Roman Space Telescope promising to map out an unprecedented volume of the cosmos within the next decade, it is of utmost importance that we possess fast and accurate methods for parsing the new data — and neural networks are surely at the forefront of these methods. Maybe the rise of machines isn' t such a bad thing after all!

Velocity Reconstrunction (of DM halos/subhalos)

Ziyong Wu et al., 2023, MNRAS (eprint: 2301.04586)

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

Royal Astronomical Society

MNRAS **522,** 4748–4765 (2023) https://doi.org/10.1093/mnras/stad1290
Advance Access publication 2023 May 1

AI-assisted reconstruction of cosmic velocity field from redshift-space spatial distribution of haloes

Ziyong Wu,[1,2] Liang Xiao,[3,4★] Xu Xiao,[3] Jie Wang,[5,6] Xi Kang,[1,7] Yang Wang,[8] Xin Wang,[3,4] Le Zhang[3,4★] and Xiao-Dong Li[3,4★]

[1] *Purple Mountain Observatory, Chinese Academy of Sciences, 10 Yuanhua Road, Nanjing 210033, China*
[2] *School of Astronomy and Space Sciences, University of Science and Technology of China, Hefei 230026, China*
[3] *School of Physics and Astronomy, Sun Yat-Sen University, Guangzhou 510297, P. R. China*
[4] *CSST Science Center for the Guangdong-Hong Kong–Macau Greater Bay Area, SYSU, Zhuhai 519082, P. R. Chi*
[5] *National Astronomical Observatories, Chinese Academy of Sciences, 20A Datun Road, Beijing 100101, China*
[6] *University of Chinese Academy of Sciences, Beijing 100049, China*
[7] *Institute for Astronomy, The School of Physics, Zhejiang University, Hangzhou 310037, China*
[8] *Peng Cheng Laboratory, No. 2, Xingke 1st Street, Shenzhen 518000, P. R. China*

Ziyong Wu
USTC

Liang Xiao
SYSU

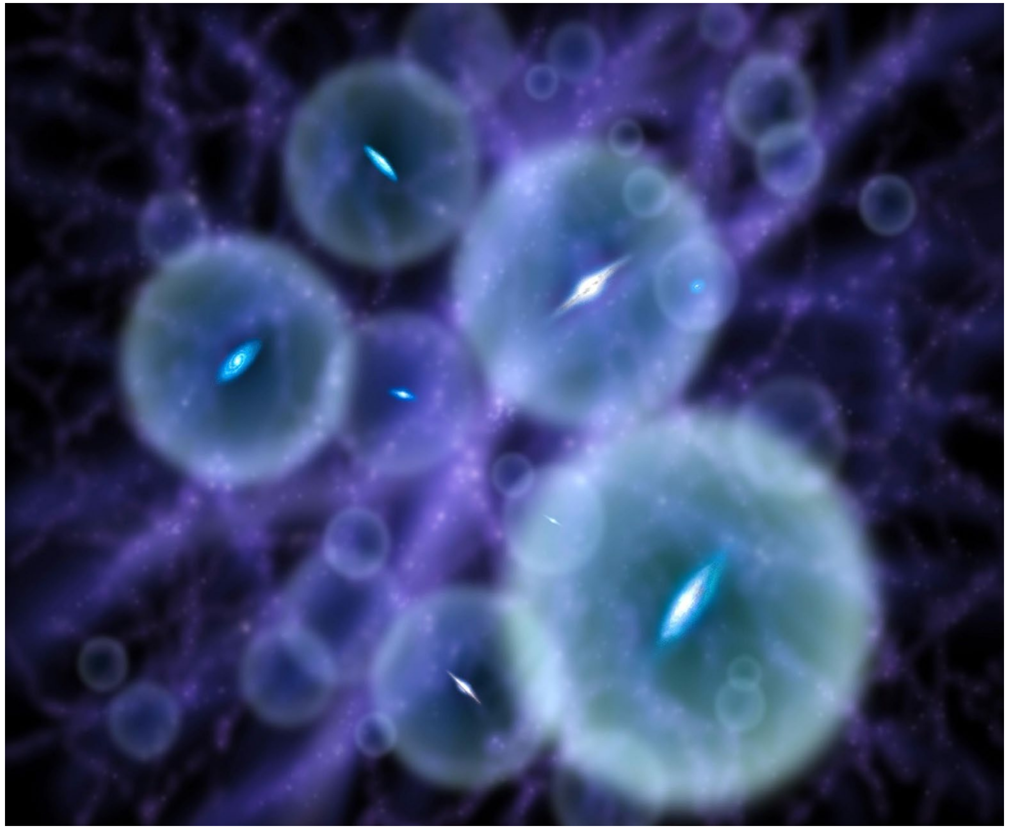
Xu Xiao
SYSU

Velocity Reconstrunction (of DM halos/subhalos)

Ziyong Wu et al., 2023, MNRAS (eprint: 2301.04586)

Why halos/subhalos?

- Close to observed objects (galaxies)
- Sparser -> more difficult



Credit: YU Jingchuan, Beijing Planetarium

Velocity Reconstrunction (of DM halos/subhalos)

Ziyong Wu et al., 2023, MNRAS (eprint: 2301.04586)

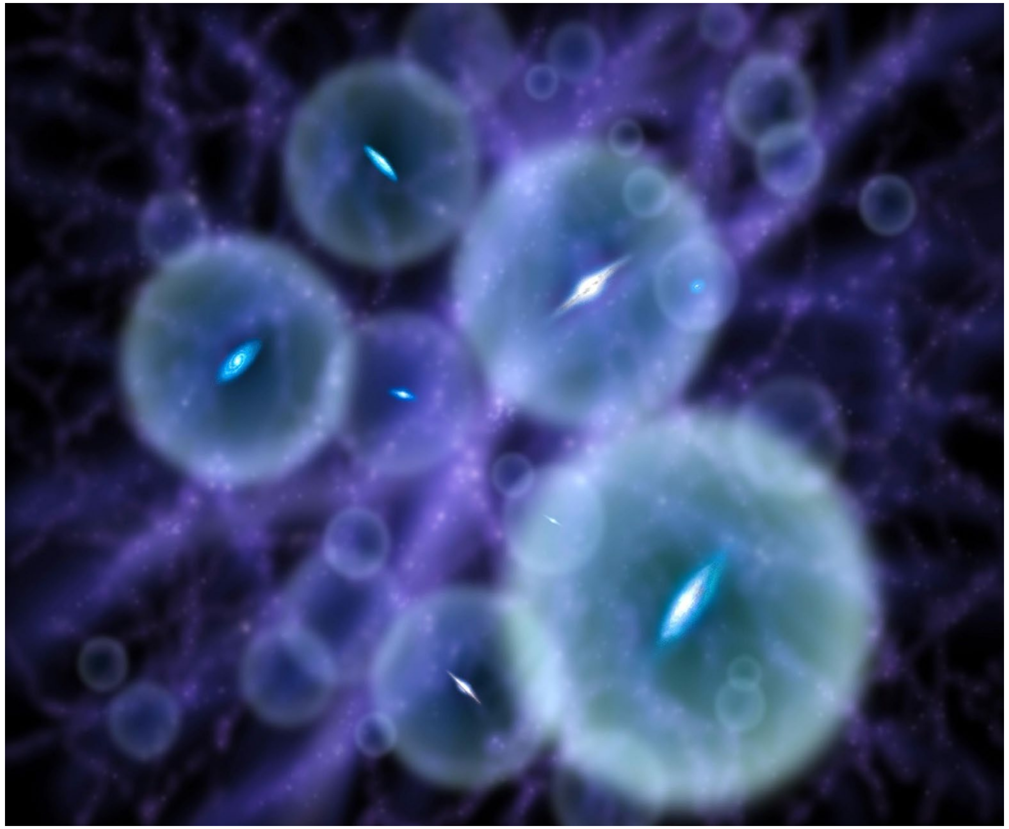
Why halos/subhalos?

- Close to observed objects (galaxies)
- Sparser -> more difficult

We did a work based on the BigMDPL simulation

- 2.5 Gpc/h, 3840^3 particles
- $M>10^{12}$ MSun, Rockstar halos/subhalos
- 8000 CIC sub-fields (500 for training, others for test), **resolution=2.78**

https://www.cosmosim.org/

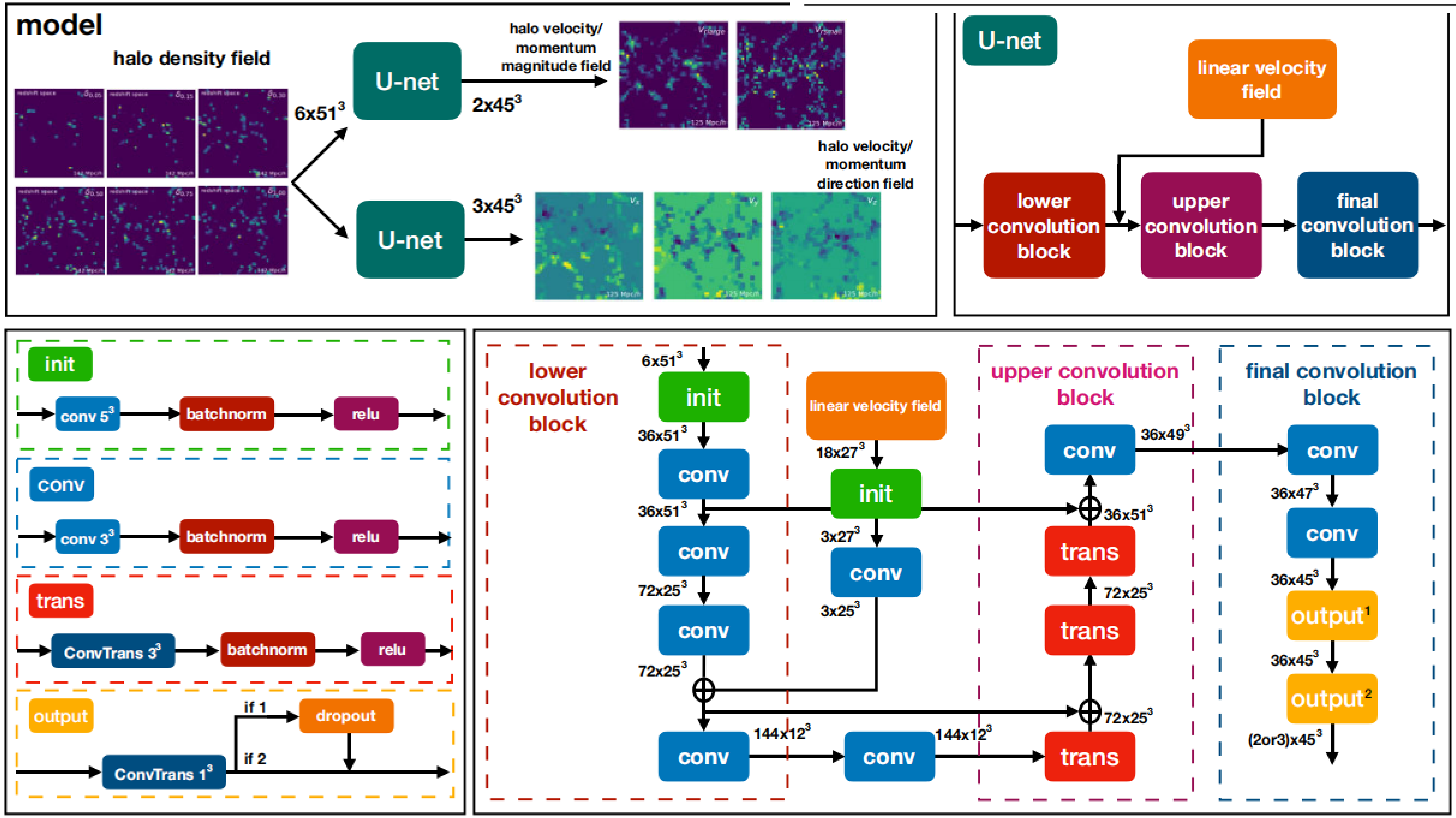


Credit: YU Jingchuan, Beijing Planetarium

Field	Grid size	Box size $(\mathrm{Mpc}\,h^{-1})^3$	Channel
Density	51^3	141.67^3	6
Velocity	45^3	125^3	5
Linear velocity	27^3	375^3	18

Input 1:
Density fields in 6 mass bins

$\log_{10}(M/M_\odot) \in$
$[15.09,\ 13.52,\ 13.12,\ 12.84,\ 12.62,\ 12.43,\ 12.30]$

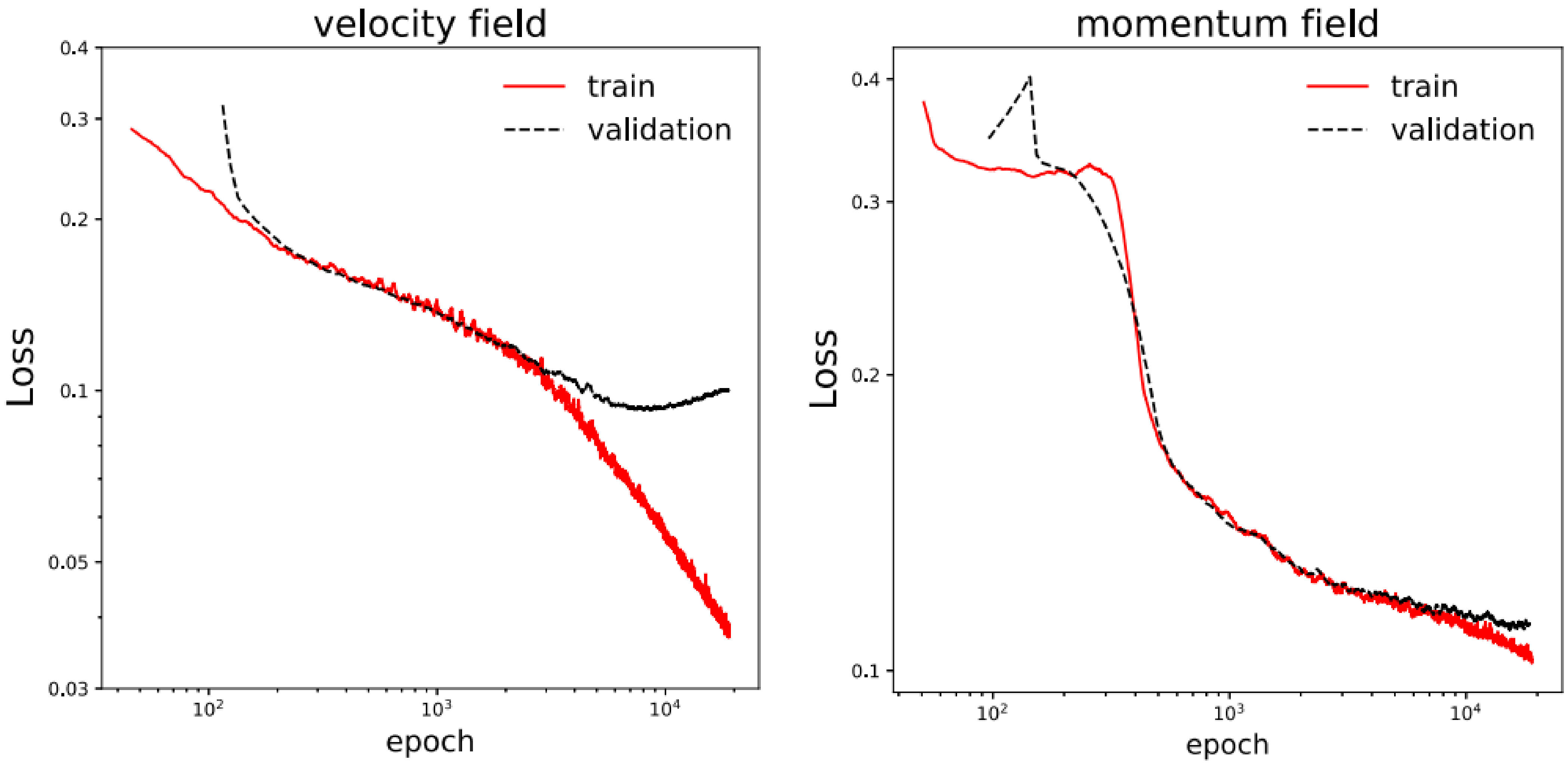


Input 2:
Linear PT results as input
-> adding large scale mode

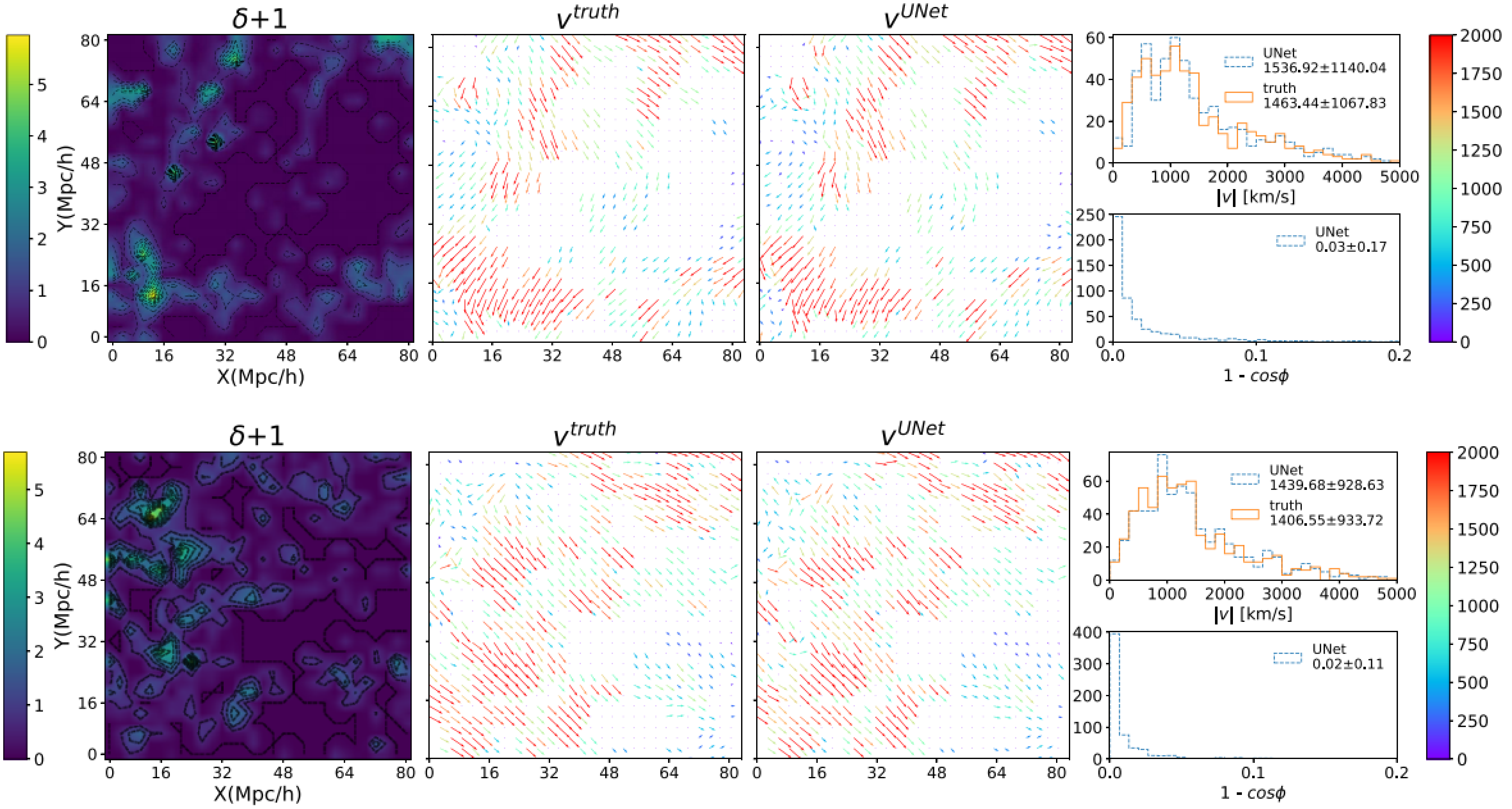
Output:
magnitude (split to 2) + direction

to improve the accuracy and the convergence speed, the velocity magnitude in the output is normalized, where the normalization factor c is chosen by $c = 1/200$ for $v > 60\ \mathrm{km\,s^{-1}}$ and $c = 1/12$ otherwise.

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{2}{5}(v_i - v_i^{\mathrm{true}})^2 + \frac{3}{5}(1 - \cos\phi_i)\right]$$

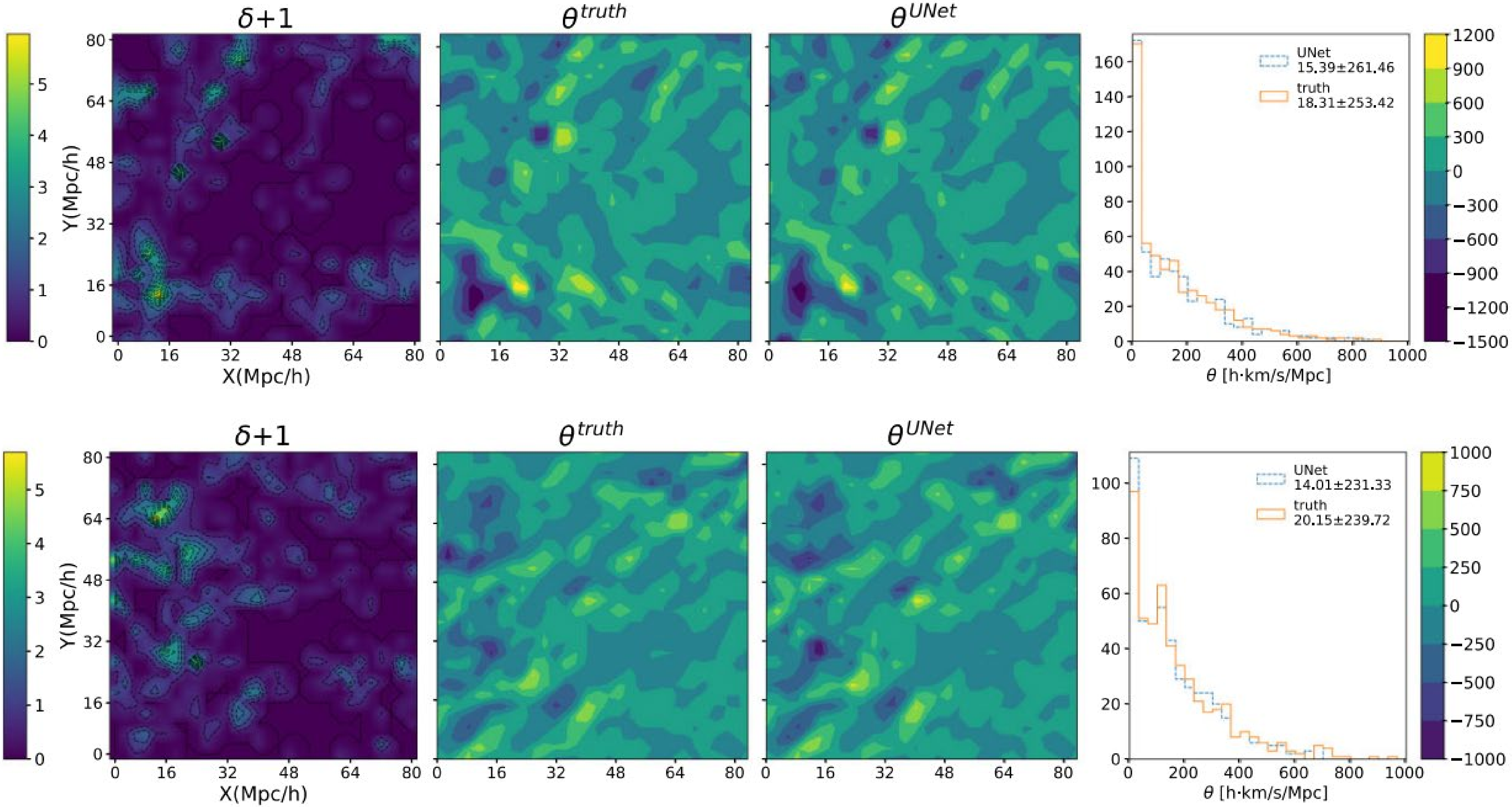


Converge after 10000-20000 epochs (very slow)

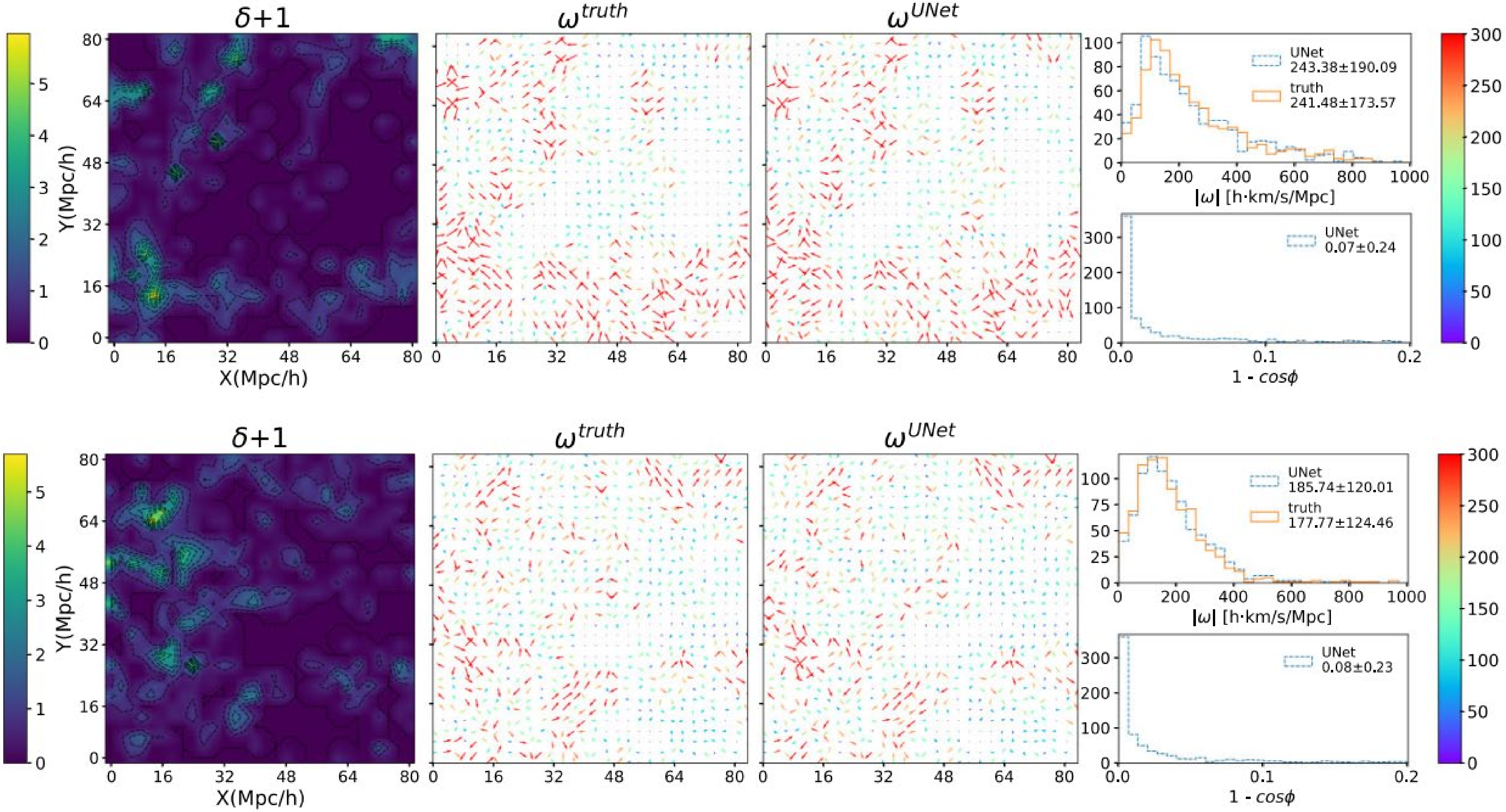


δ+1
v^{truth}
v^{UNet}
Y(Mpc/h)
X(Mpc/h)
UNet
1536.92±1140.04
truth
1463.44±1067.83
|v| [km/s]
UNet
0.03±0.17
1 - cosϕ
δ+1
v^{truth}
v^{UNet}
Y(Mpc/h)
X(Mpc/h)
UNet
1439.68±928.63
truth
1406.55±933.72
|v| [km/s]
UNet
0.02±0.11
1 - cosϕ

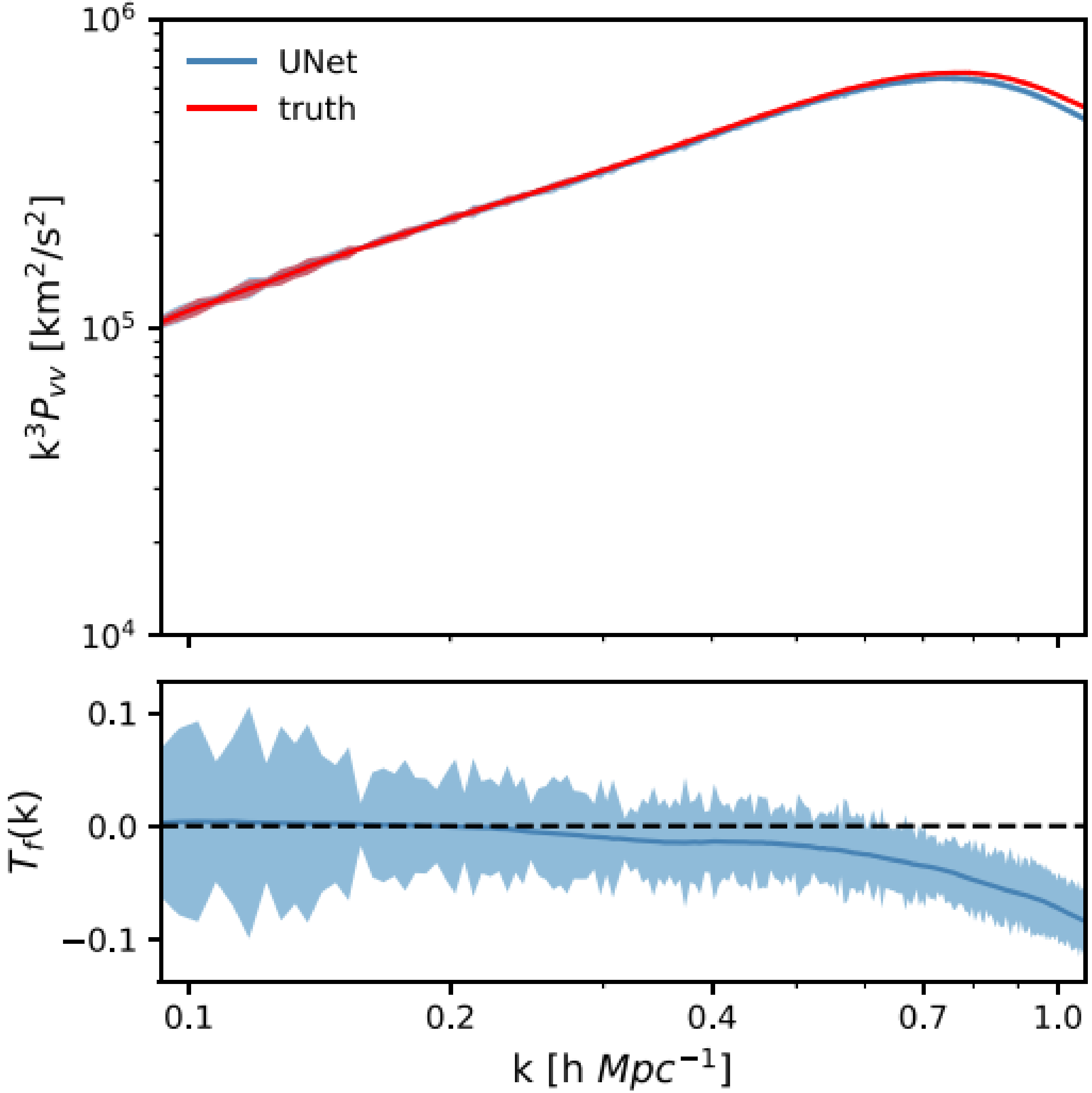
$$\theta \equiv \nabla \cdot \boldsymbol{v}$$



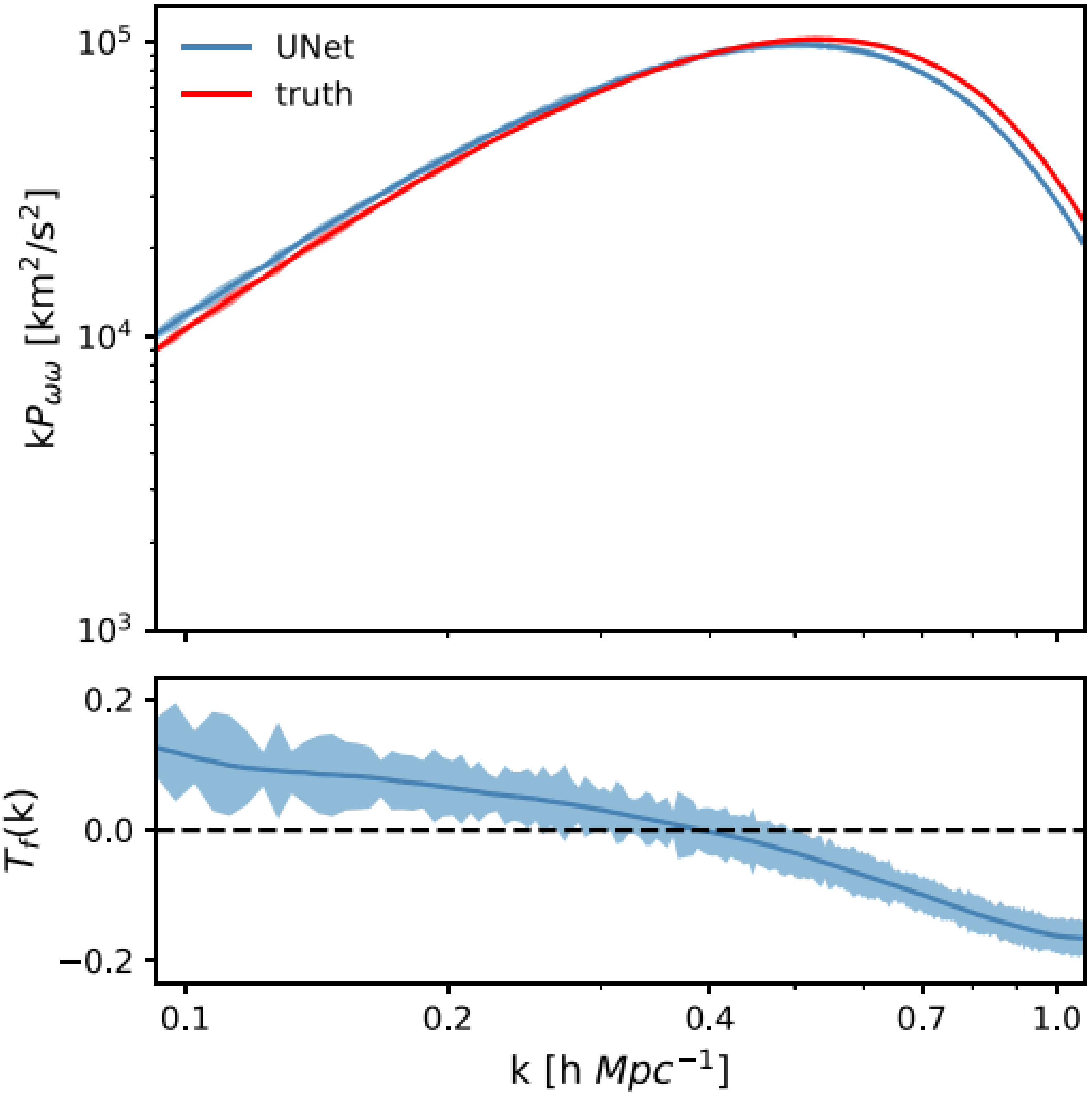
$\boldsymbol{\omega} = \nabla \times \boldsymbol{v}$.

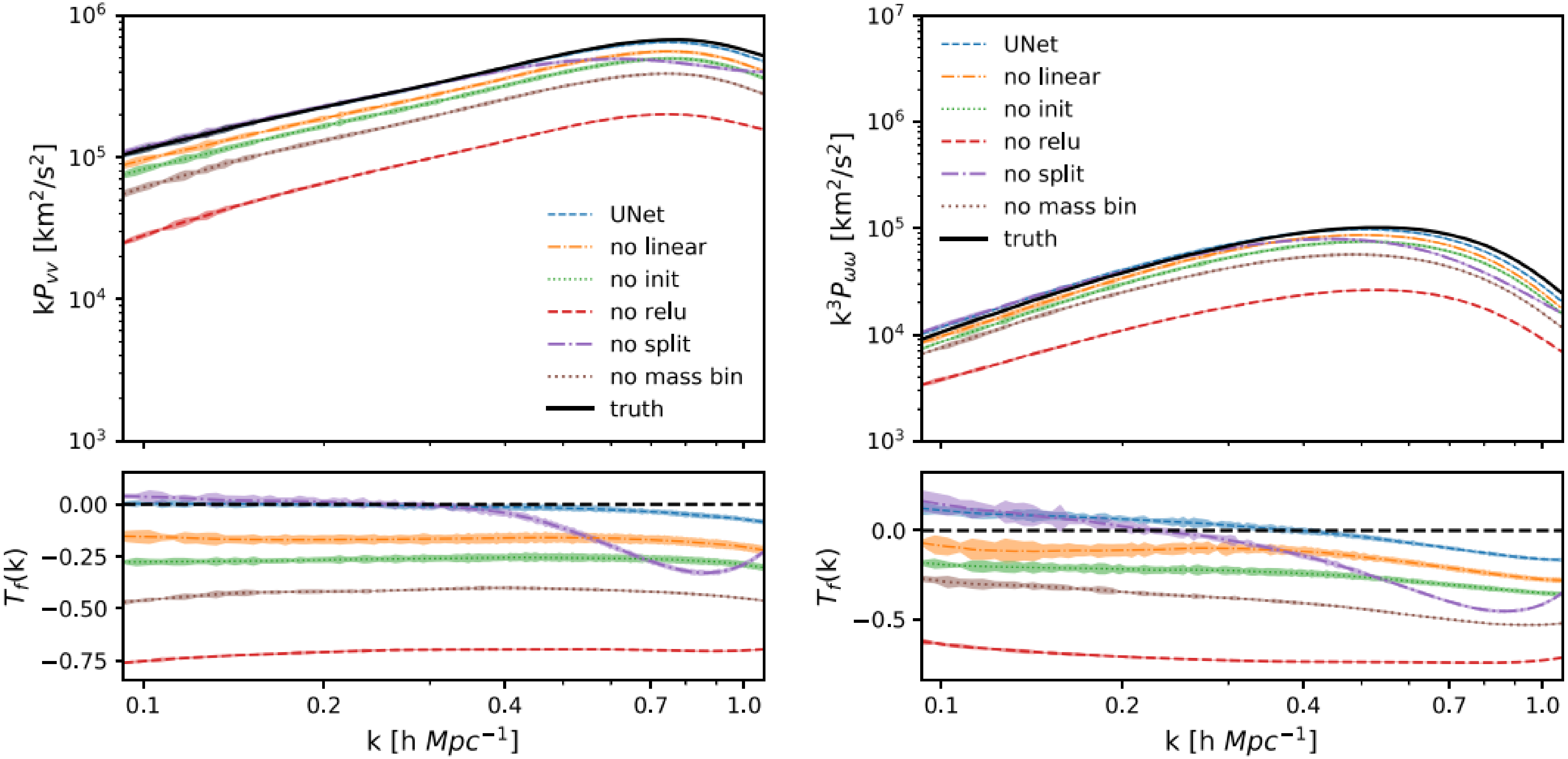


power spectra of
|v |



power spectra of
|∇×v |





UNet
no linear
no init
no relu
no split
no mass bin
truth
kP_{vv} [km^2/s^2]
$k^3P_{\omega\omega}$ [km^2/s^2]
T_f(k)
k [h Mpc^{-1}]

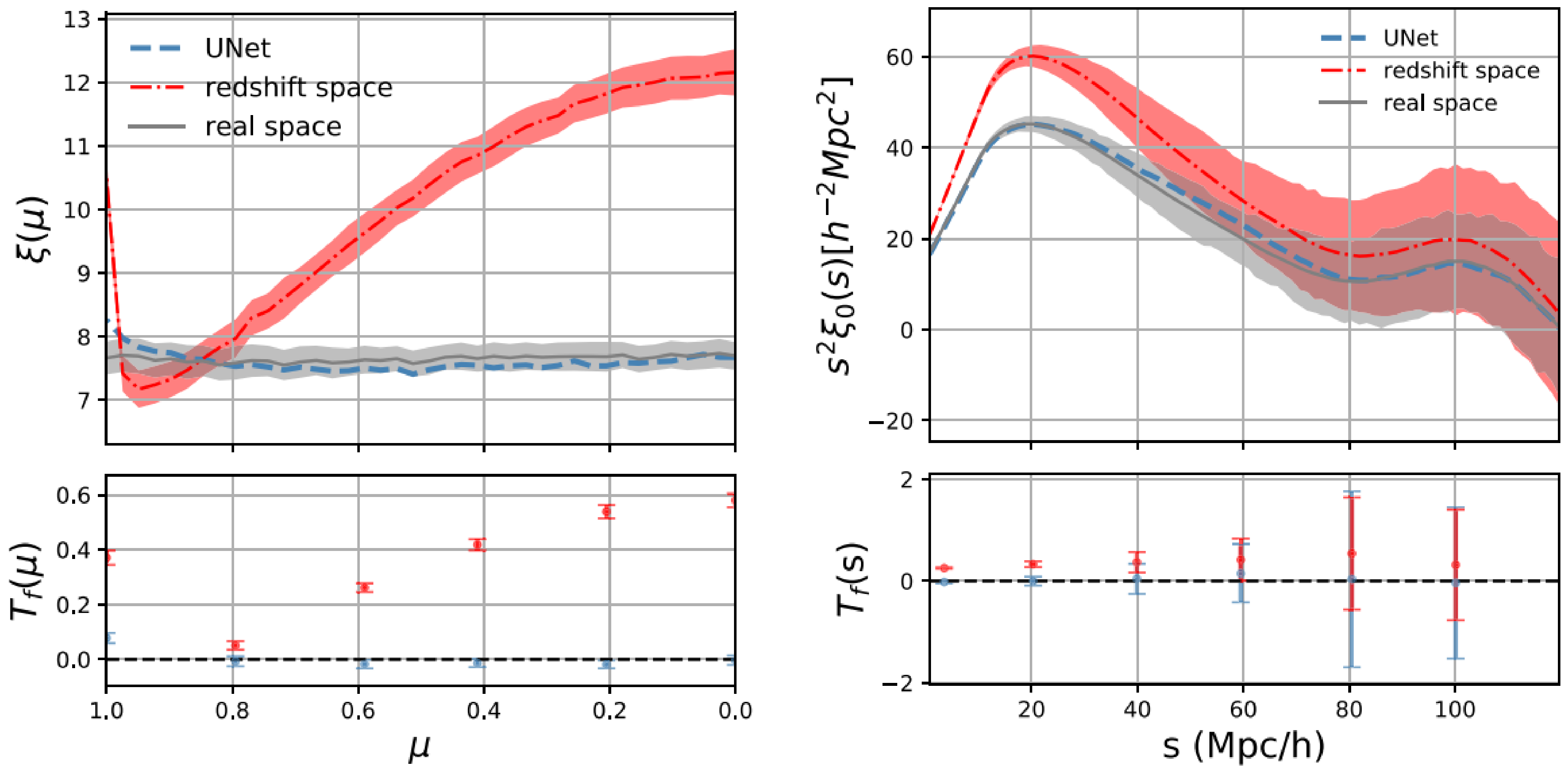
Real-space 2pcf

(interpolation **v** from grid to objects)

$$\xi(\boldsymbol{r}) = \sum_{\ell=0}^{\infty} \xi_\ell(r) L_\ell(\mu),$$

with

$$\xi_\ell(r) = \frac{2\ell+1}{2} \int_{-1}^{1} \xi(r,\mu) L_\ell(\mu) \mathrm{d}\mu$$



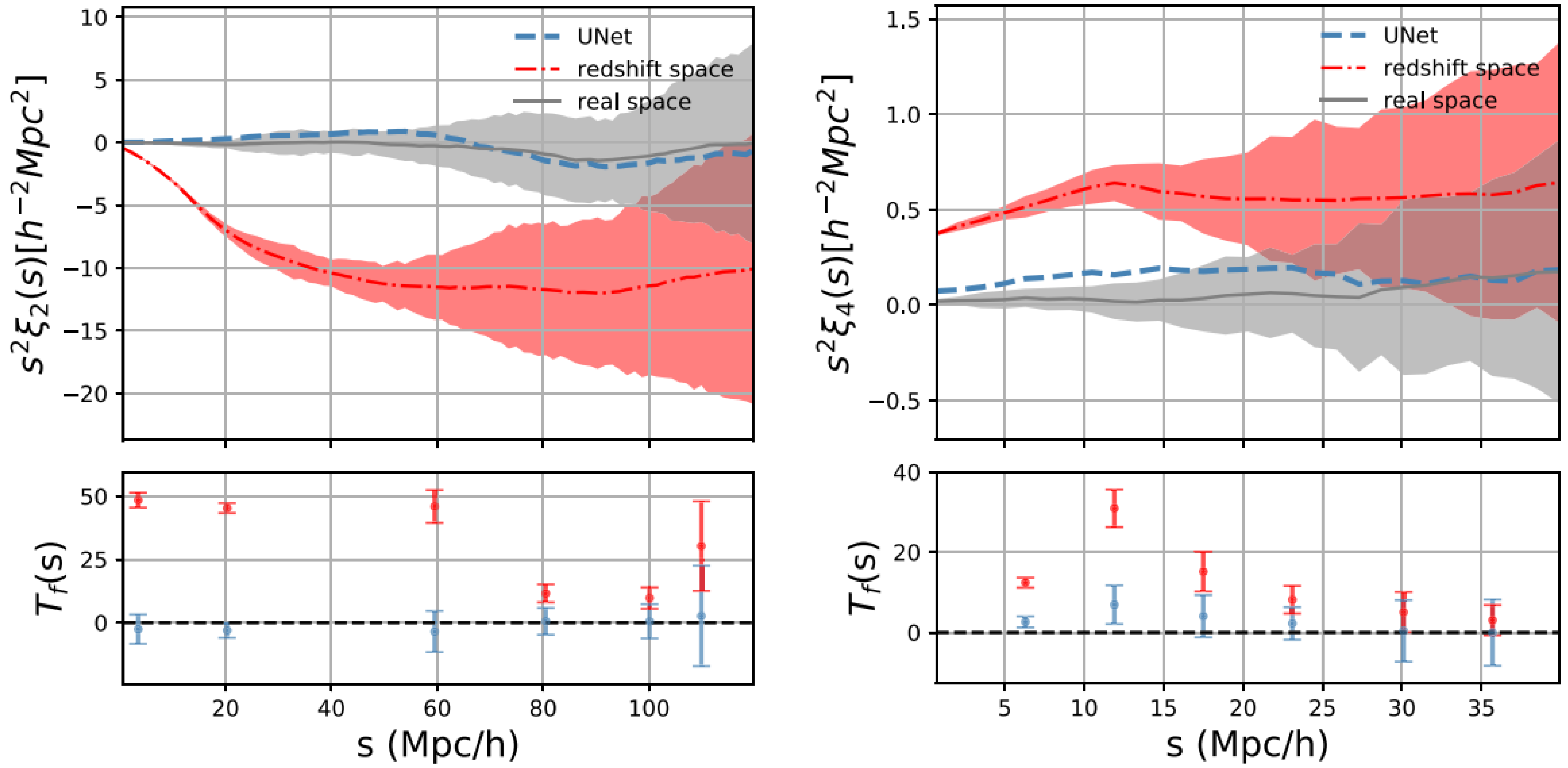
$$\xi_{1\mathrm{D}}(\mu) = \int_0^\infty \mathrm{d}r \xi(r,\mu) \mathrm{d}r \ .$$

Real-space 2pcf

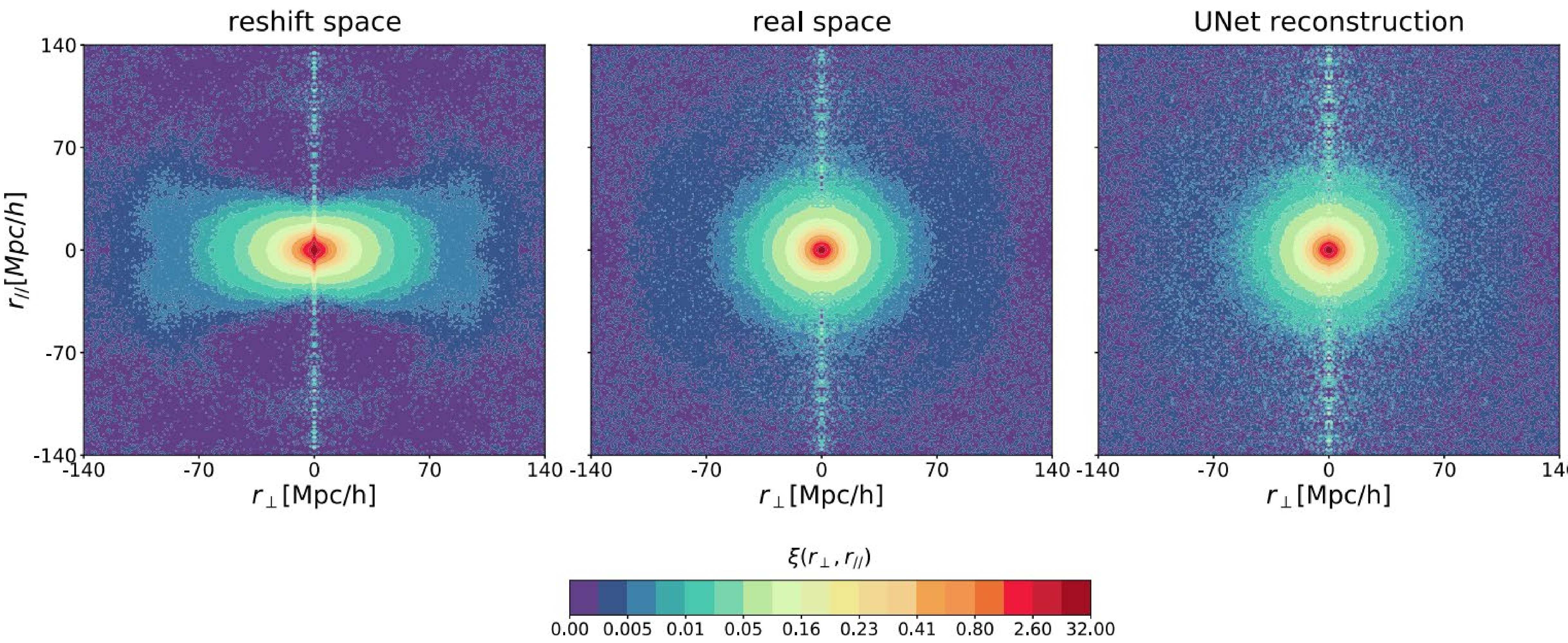
$$\xi(\boldsymbol{r}) = \sum_{\ell=0}^{\infty} \xi_\ell(r) L_\ell(\mu),$$

with

$$\xi_\ell(r) = \frac{2\ell+1}{2} \int_{-1}^{1} \xi(r,\mu) L_\ell(\mu) \mathrm{d}\mu$$



Real-space 2pcf

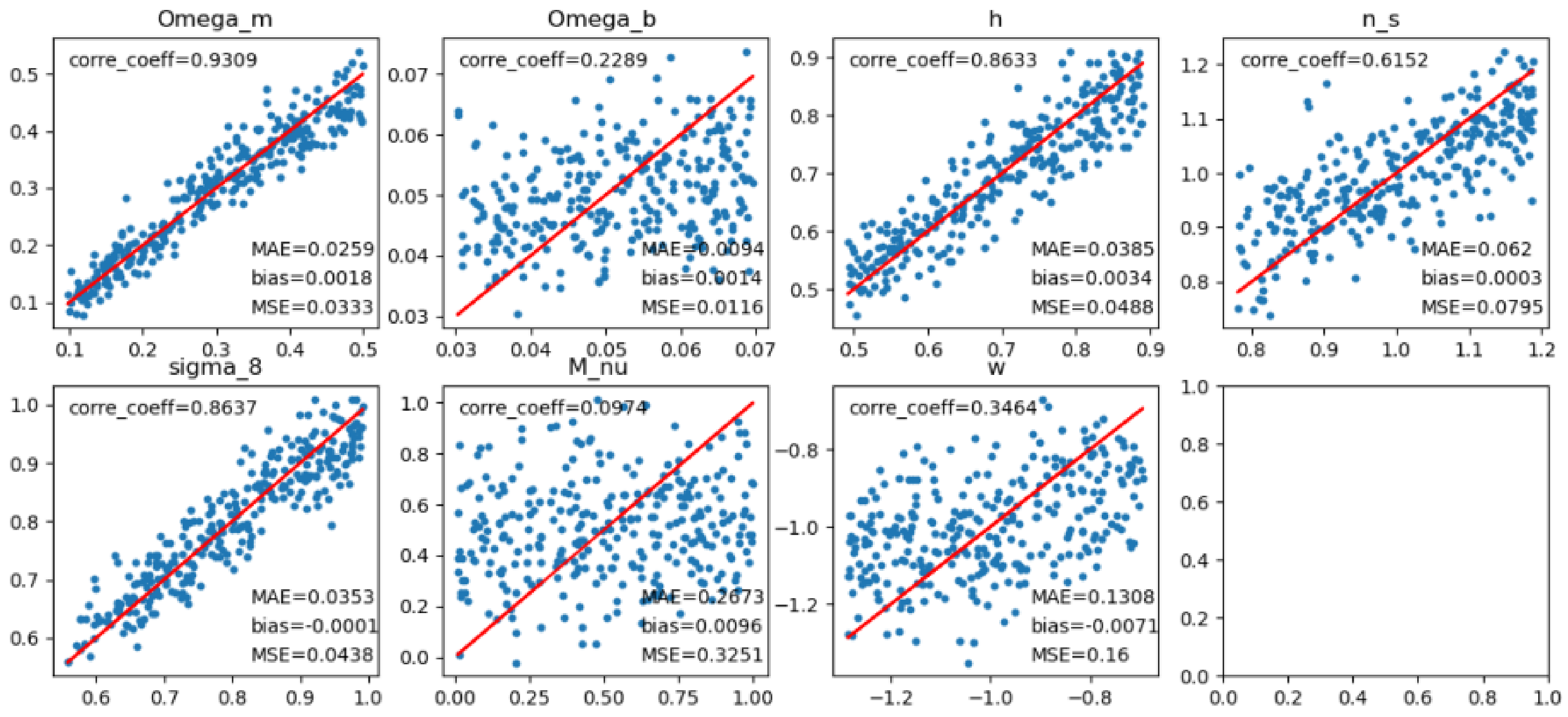


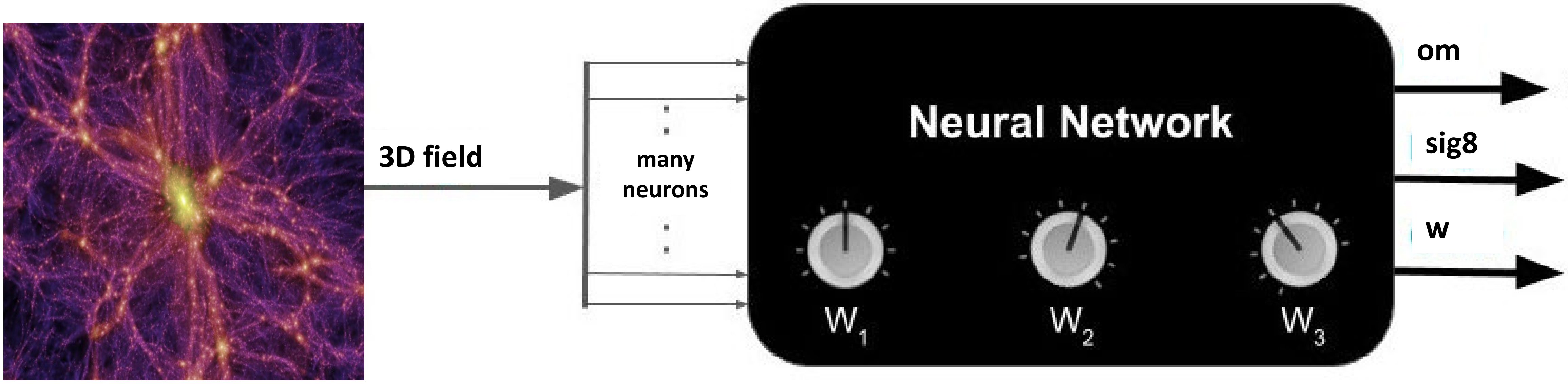
Summary

- **AI is very successful in reconstructing the velocities of DM particle/halo/subhalo @ 2-3 h^{-1} Mpc / k ~ 1-1.4**

- **Related works:**
  - Tian-Xiang Mao, Jie Wang, Baojiu Li, et al., 2020 (eprint: 2002.10218, **BAO Reconstruction**)
  - Fei Qin, David Parkinson, Sungwook E. Hong, Cristiano G. Sabiu (eprint: 2302.02087, **density and velocity reconstruction**)
  - Feng Shi et al. (see his talk, **density and velocity reconstruction**)

II. Parameter Estimation

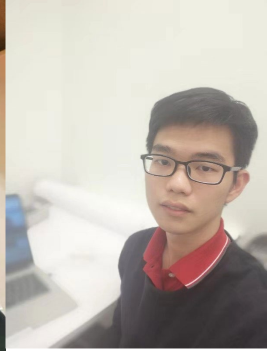



3D field
many
neurons
Neural Network
W_1
W_2
W_3
om
sig8
w

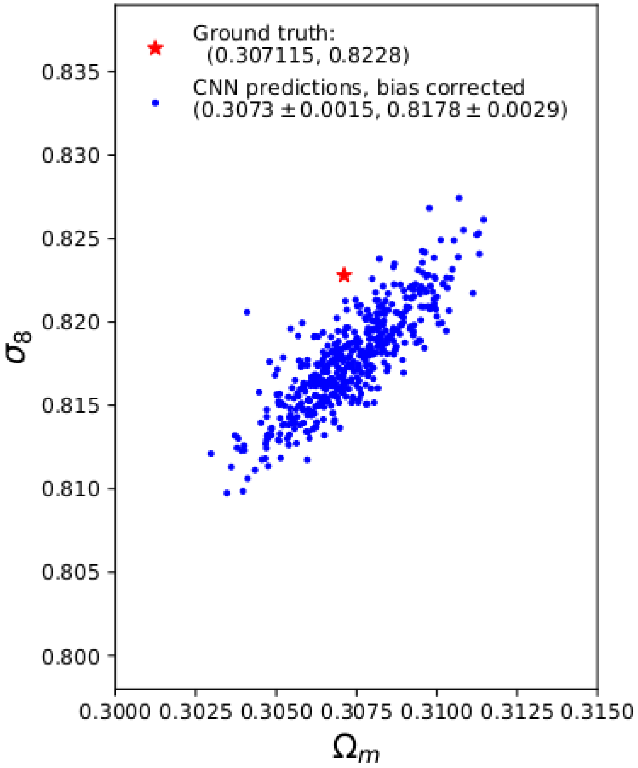
Parameter estimation (based on DM particles)

Pan et al., 2020, SCPMA

Shuyang Pan



Miaoxin Liu



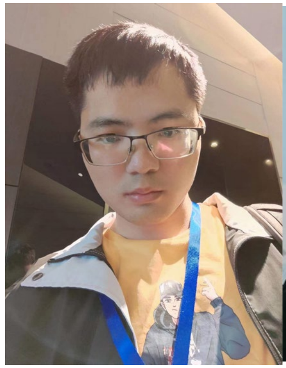
Method	Relative error of (Ω_m, σ_8)[a]	
CNN	(0.0048, 0.0053)	
2pcf, $s \in (0, 130)\ h^{-1}$ Mpc	(0.017, 0.012)	x3
2pcf, $s \in (10, 130)\ h^{-1}$ Mpc	(0.1, 0.06)	x10-20

Based on ~1000 simulations (each 128^3 particles, 256 h^{-1} Mpc box)

$0.16 \leq \Omega_m \leq 0.46$, $2.0 \leq 10^9 A_s \leq 2.3$

Parameter estimation (based on DM halos)

Zhiwei Min et al., in progress



Zhiwei Min
SYSU

Liang Xiao
SYSU

Xu Xiao
SYSU

Quijote LHvw simulations (1 Gpc/h box)

Name	Ω_m	Ω_b	h	n_s	σ_8	M_ν (e V)
LHvw	[0.1 , 0.5]	[0.03 , 0.07]	[0.5 , 0.9]	[0.8 , 1.2]	[0.6 , 1.0]	[0 , 1]
w	δ_b	realizations	simulations	ICs	$N_c^{1/3}$	$N_\nu^{1/3}$
[-1.3, -0.7]	0	5000	standard	Zeldovich	512	512

RSD added @ z-direction
convert to Planck cosmology background, use 744 Mpc/h volume
select ~1750 simulations

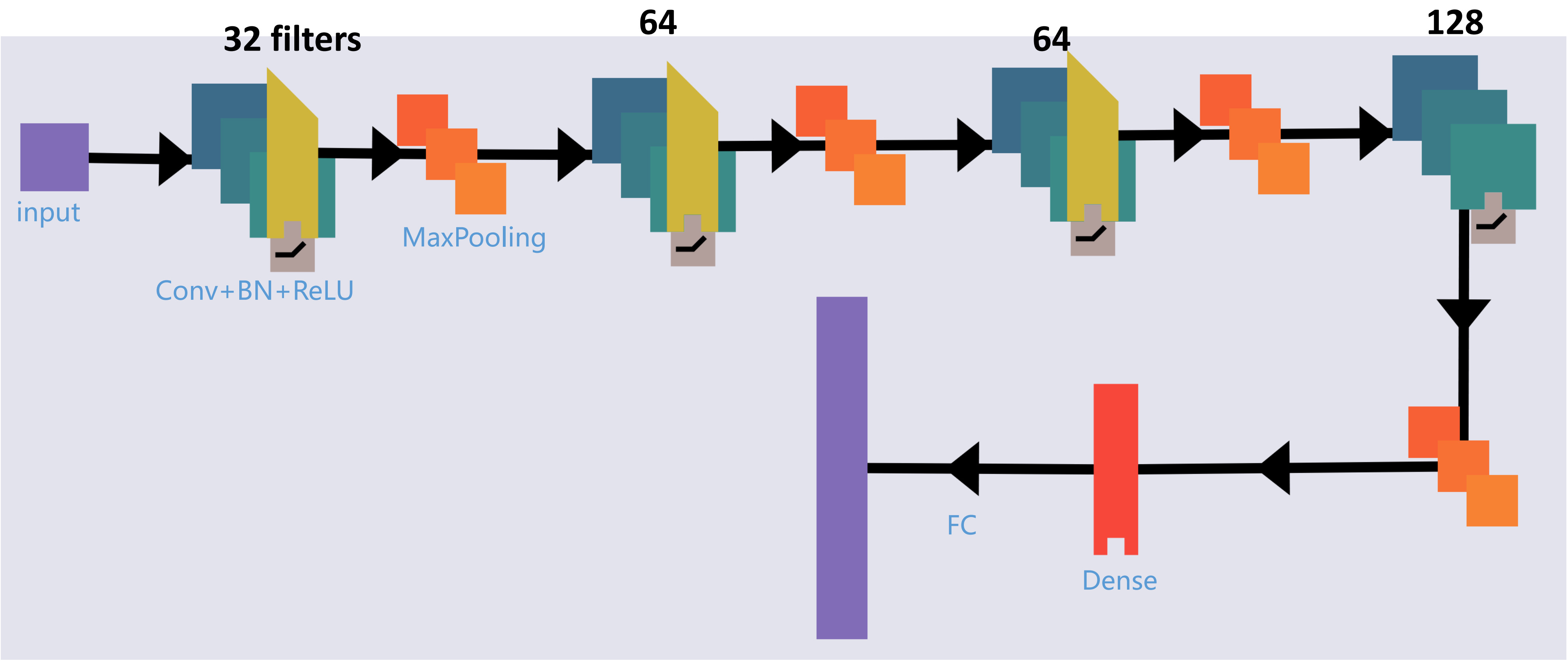
CNN architecture (simple version)

Input subsample:

60^3 voxels, 148.8 Mpc/h

grid resolution = 2.48 Mpc/h

```
input:   torch.Size([25, 1, 60, 60, 60])
conv1:   torch.Size([25, 32, 30, 30, 30])
conv2:   torch.Size([25, 64, 15, 15, 15])
conv3:   torch.Size([25, 64, 6, 6, 6])
conv4:   torch.Size([25, 128, 2, 2, 2])
linear1:   torch.Size([25, 128])
linear2:   torch.Size([25, 64])
output:  torch.Size([25, 64])
```

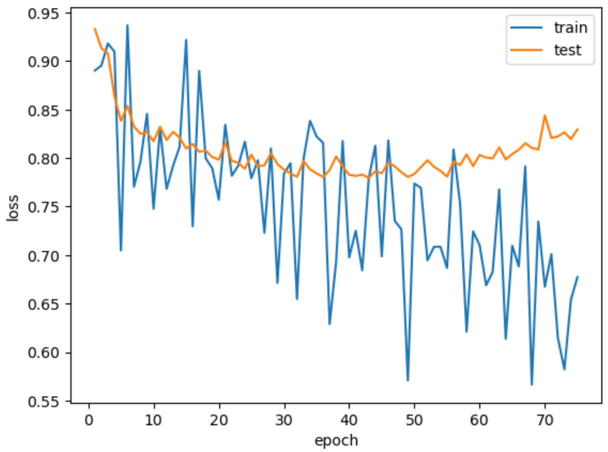


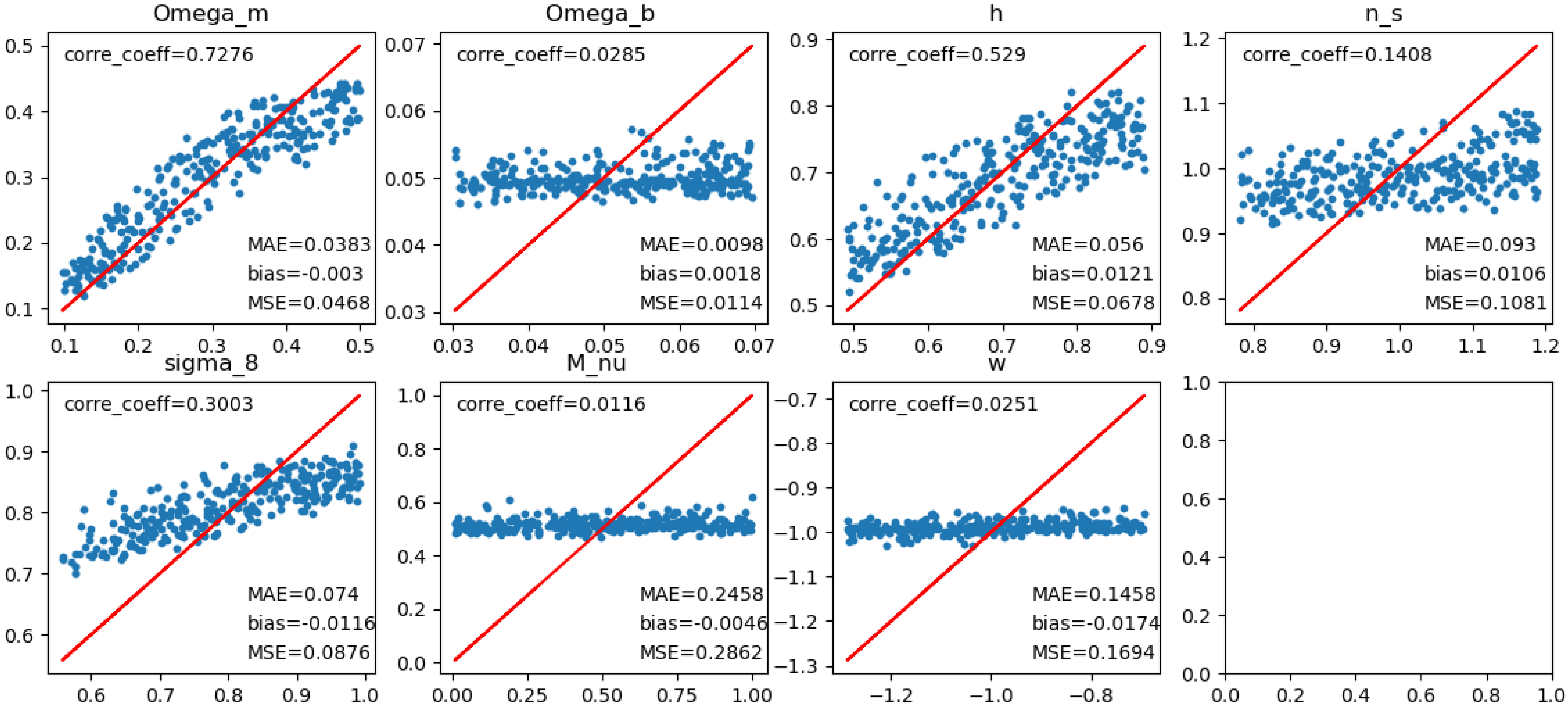
Configuration Space Result

Input subsample:

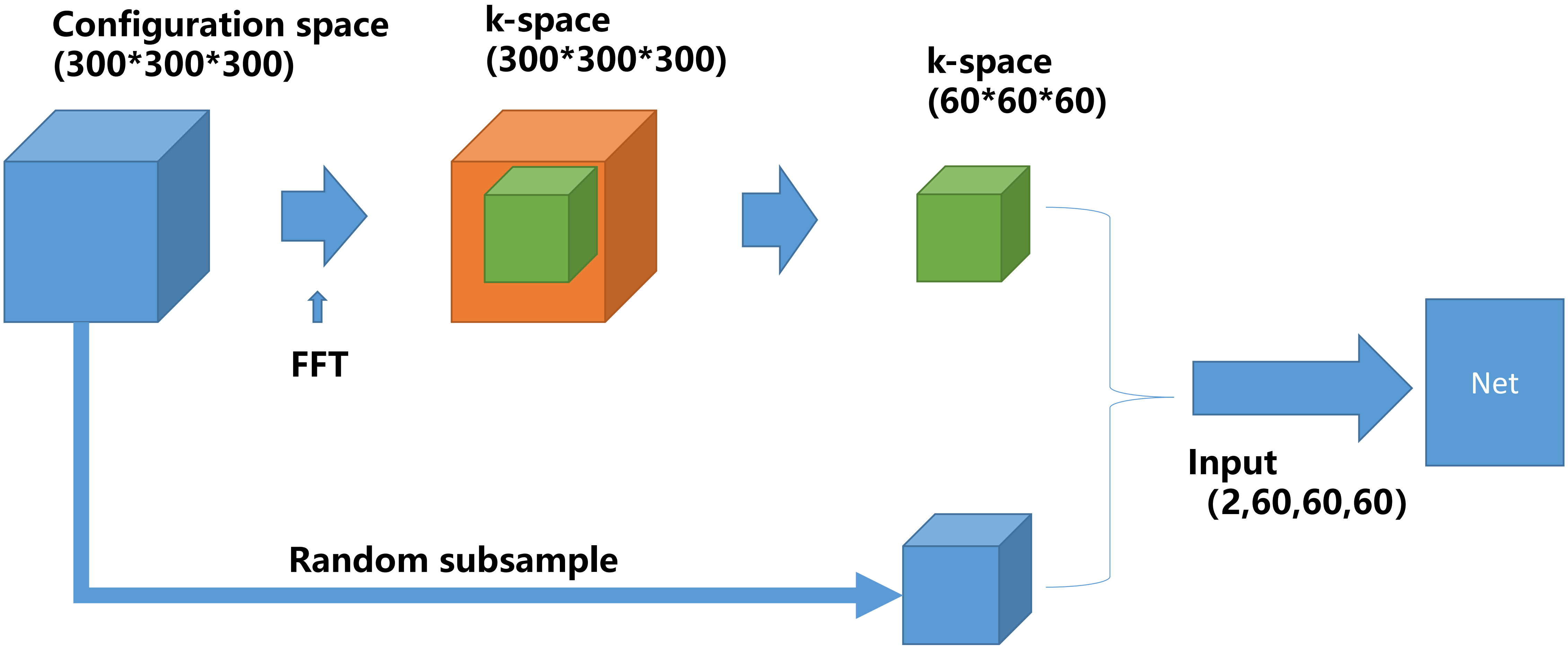
60^3 voxels, 148.8 Mpc/h

grid resolution = 2.48 Mpc/h





Configuration Space + Fourier space

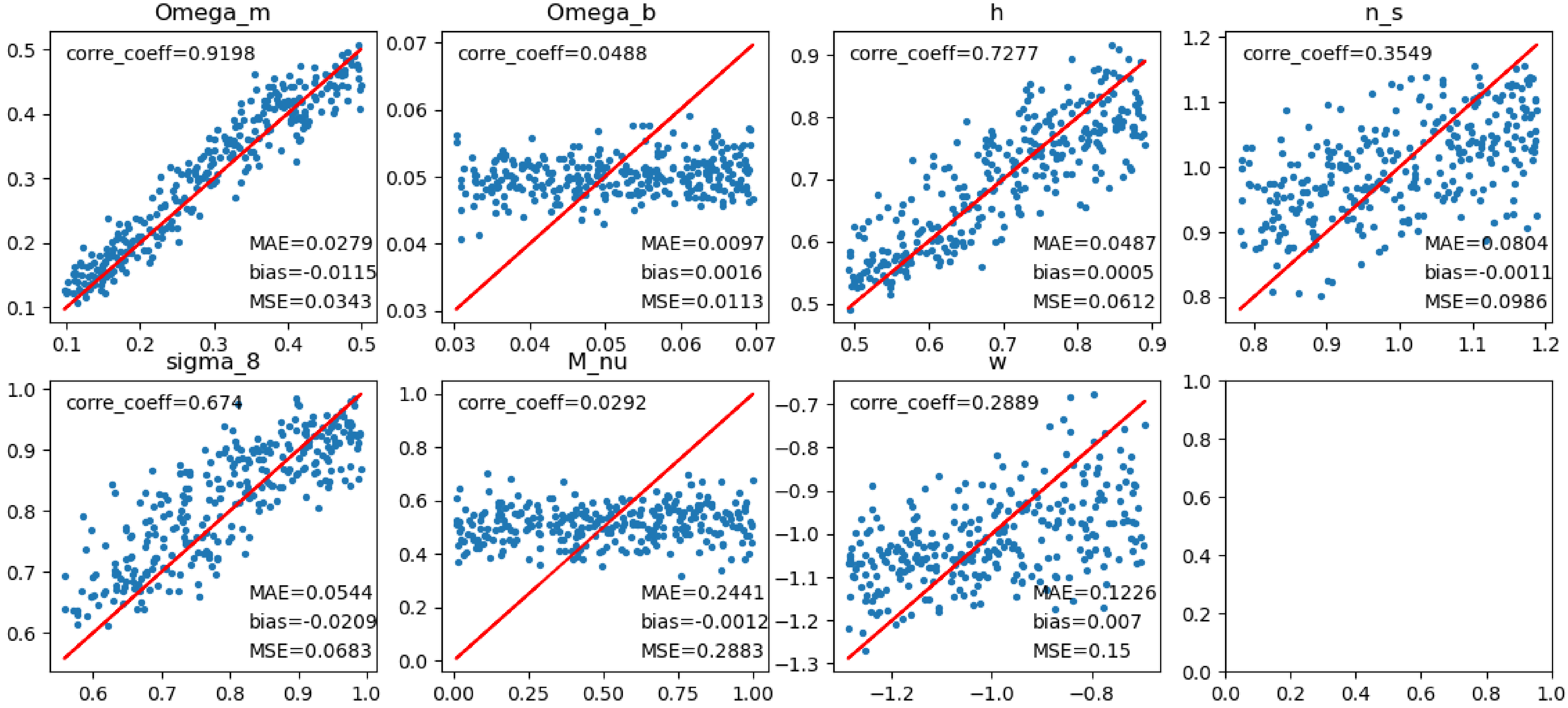


Configuration Space + Fourier space

Adding k-space is helpful

MSE reduced by 20-30%

n_s becomes reasonable, w shows some signal

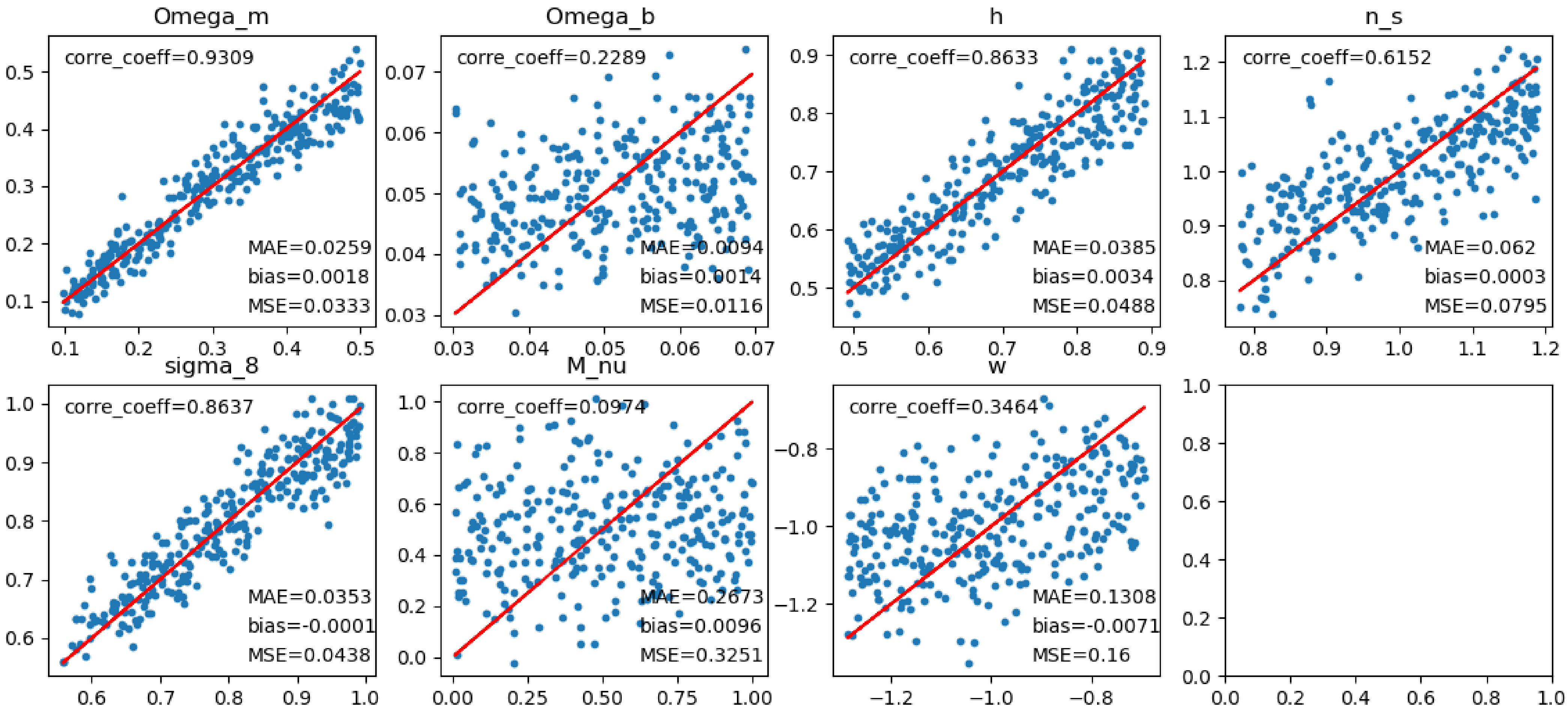


Configuration Space + PS + 2pcf + ScatteringTransform

Adding data summaries is helpful

Comparable to + k-space

better sigma_8 constraint

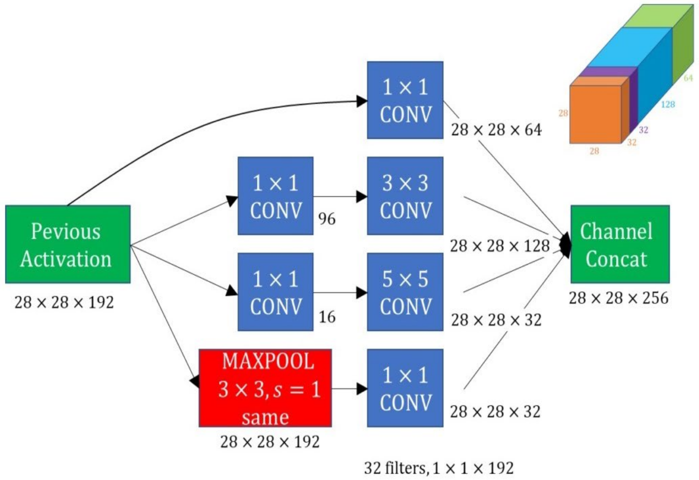


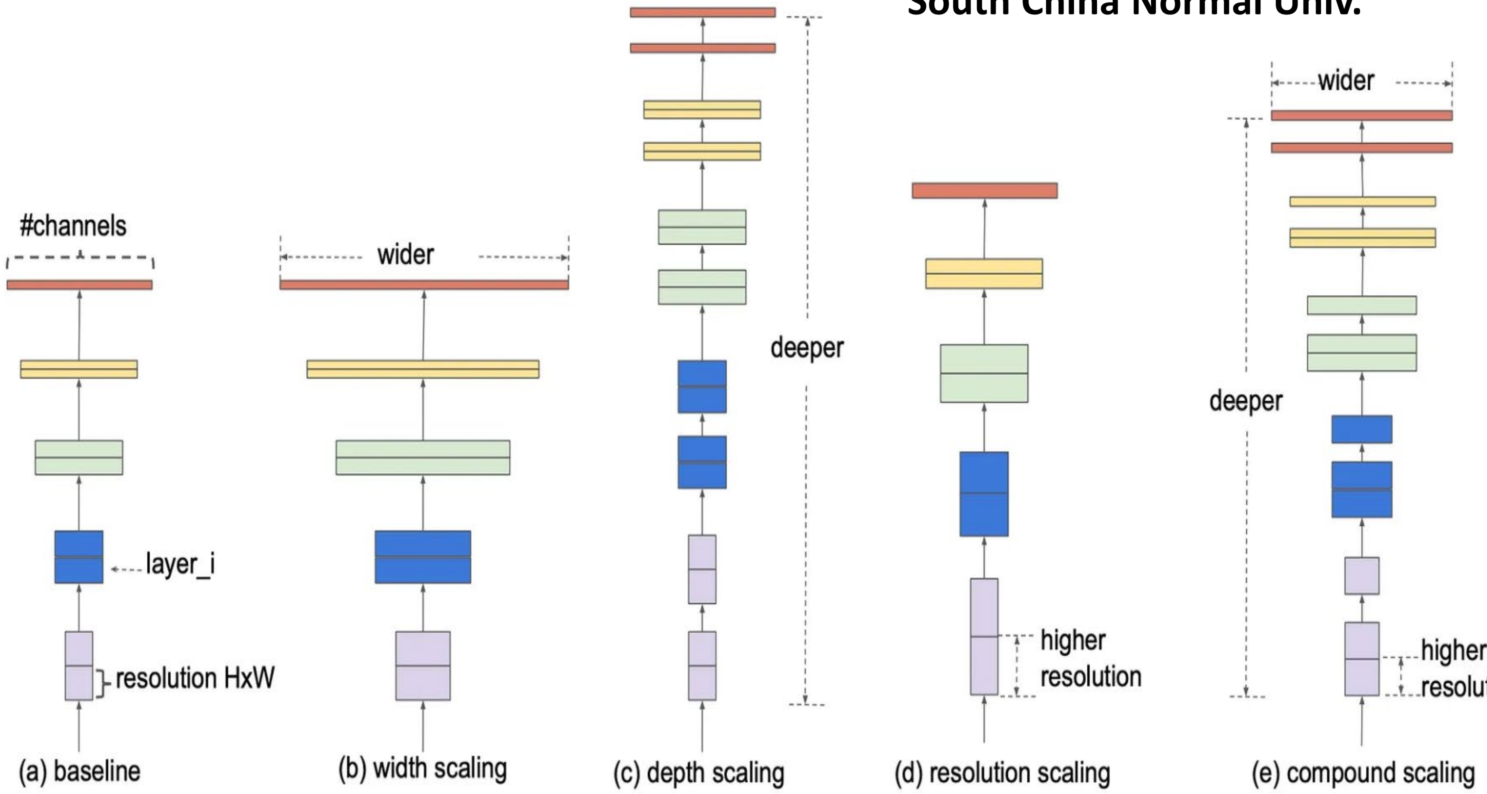
Advanced Architecture (Google's EfficientNet)

https://github.com/tensorflow/tpu/tree/master/models/official/efficientnet

Jinqu Zhang

South China Normal Univ.

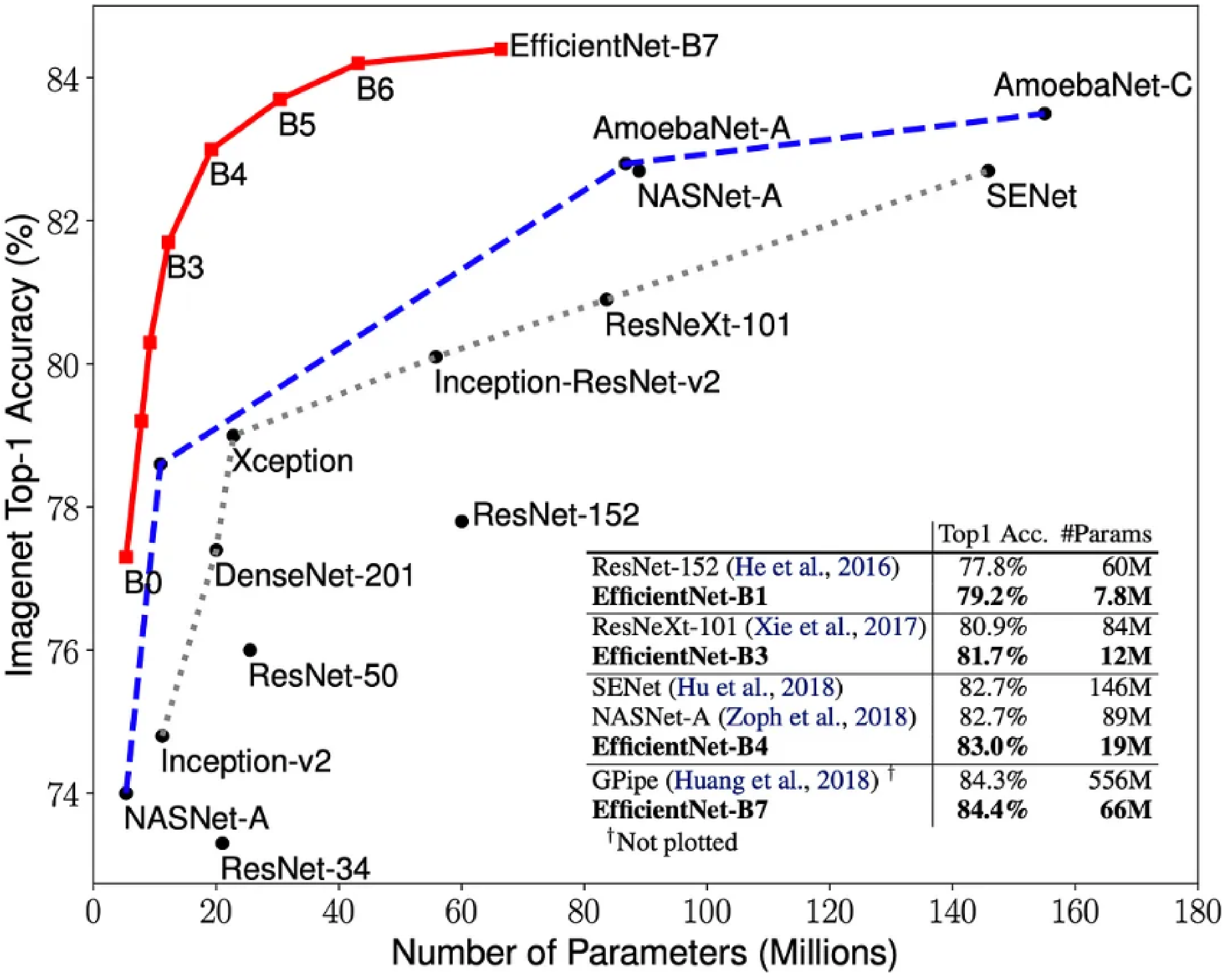




Inception **Scaling Conv Blocks**

Advanced Architecture (Google’s EfficientNet)

https://github.com/tensorflow/tpu/tree/master/models/official/efficientnet



	Top1 Acc.	#Params
ResNet-152 (He et al., 2016)	77.8%	60M
EfficientNet-B1	**79.2%**	**7.8M**
ResNeXt-101 (Xie et al., 2017)	80.9%	84M
EfficientNet-B3	**81.7%**	**12M**
SENet (Hu et al., 2018)	82.7%	146M
NASNet-A (Zoph et al., 2018)	82.7%	89M
EfficientNet-B4	**83.0%**	**19M**
GPipe (Huang et al., 2018) †	84.3%	556M
EfficientNet-B7	**84.4%**	**66M**

†Not plotted

Performance of the EfficientNet family compared to other classifiers. Taken from Tan & Le, 2019

Jinqu Zhang

South China Normal Univ.

EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks

Mingxing Tan, Quoc V. Le

Convolutional Neural Networks (ConvNets) are commonly developed at a fixed resource budget, and then scaled up for better accuracy if more resources are available. In this paper, we systematically study model scaling and identify that carefully balancing network depth, width, and resolution can lead to better performance. Based on this observation, we propose a new scaling method that uniformly scales all dimensions of depth/width/resolution using a simple yet highly effective compound coefficient. We demonstrate the effectiveness of this method on scaling up MobileNets and ResNet.

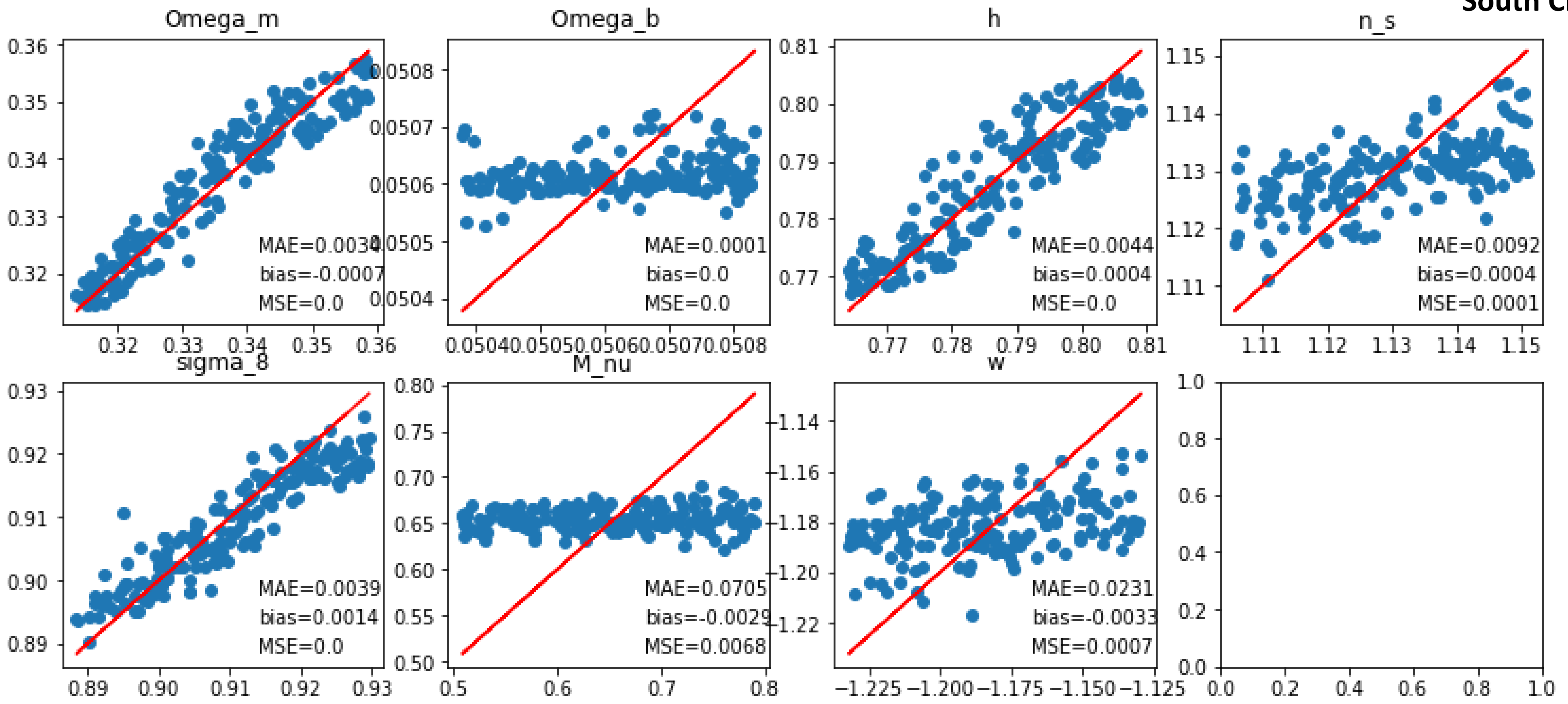
To go even further, we use neural architecture search to design a new baseline network and scale it up to obtain a family of models, called EfficientNets, which achieve much better accuracy and efficiency than previous ConvNets. In particular, our EfficientNet-B7 achieves state-of-the-art 84.3% top-1 accuracy on ImageNet, while being 8.4x smaller and 6.1x faster on inference than the best existing ConvNet. Our EfficientNets also transfer well and achieve state-of-the-art accuracy on CIFAR-100 (91.7%), Flowers (98.8%), and 3 other transfer learning datasets, with an order of magnitude fewer parameters. Source code is at this https URL.

Advanced Architecture (Google's EfficientNet)

https://github.com/tensorflow/tpu/tree/master/models/official/efficientnet

Jinqu Zhang

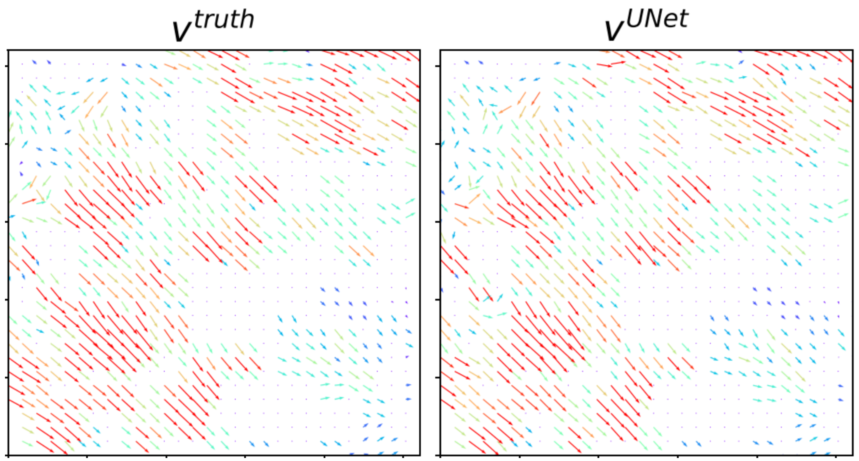
South China Normal Univ.

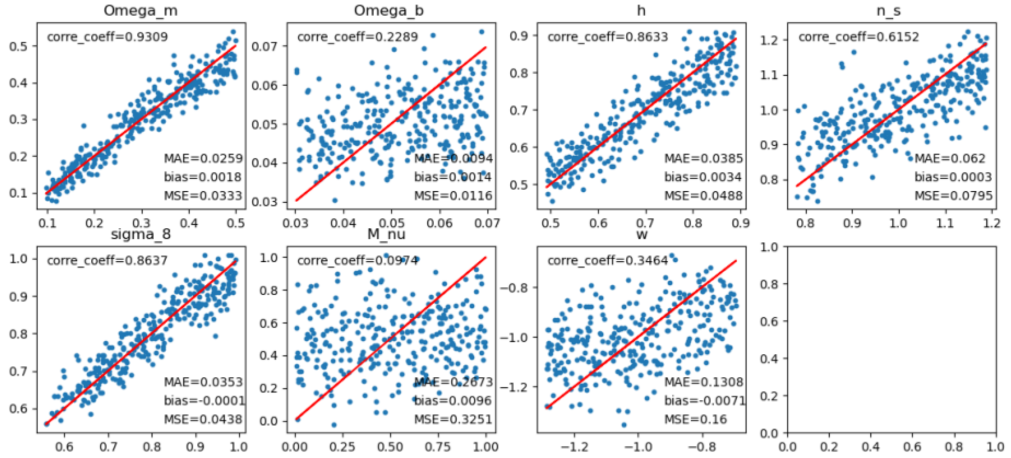


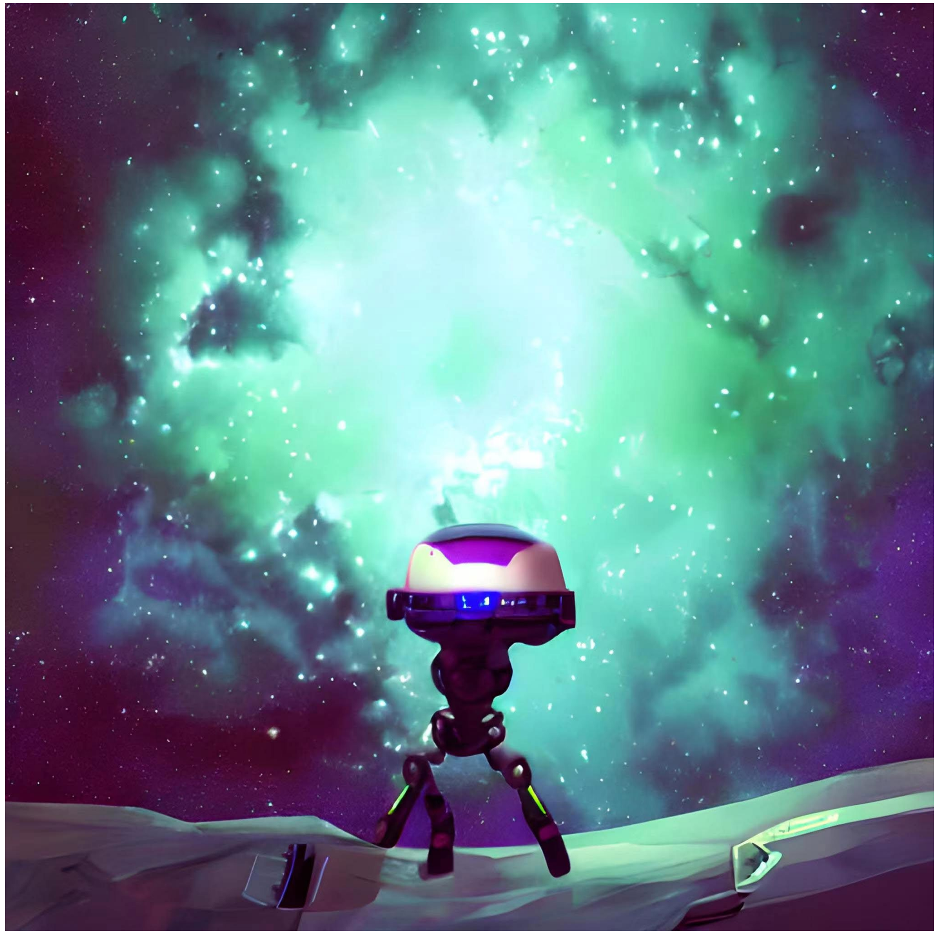
Summary

- + k-space / + data-summary is helpful
- advanced architecture is helpful
- Neutrino seems difficult...
    - Maybe need high resolution simulation?

III. Summar & Future







Many more things

Velocity Statistics

Parameter Estimation

Velocity Reconstruction

kSZ; RSD; Real-Space statistics

Information Reference

Traditional Method

Share technique; Complimentary

Photo-z (using GNN, in plan)

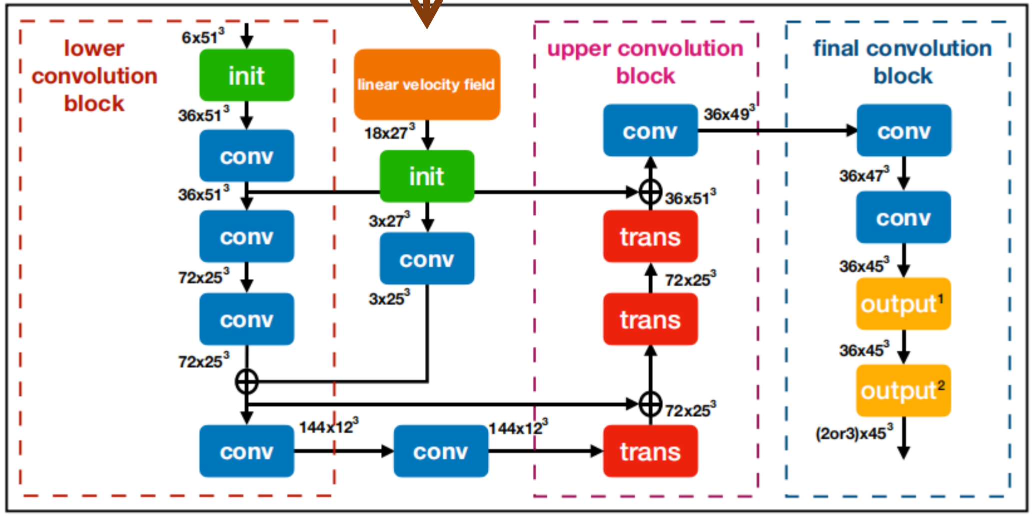
DM Particles (Pass) → Halo/Subhalo (almost pass) → Real Observation (start)

Numerous applications are NOT mentioned in this talk:

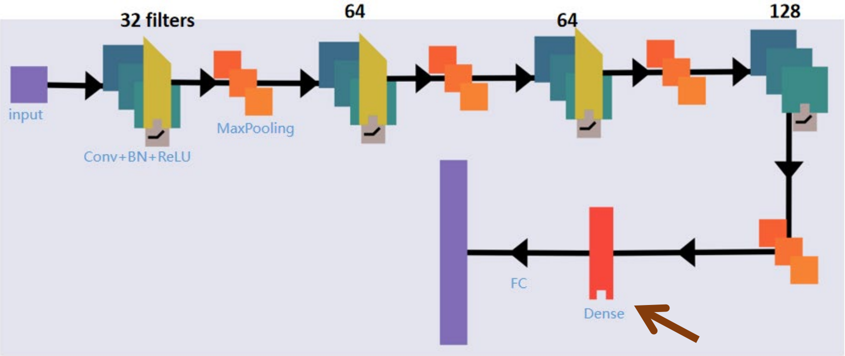
- WL, SL
- AIGC mock
- IC/DM Reconstruction
- CMB, 21cm ...
- ...

AI and Traditional Analysis are Complimentary

PT result added as input



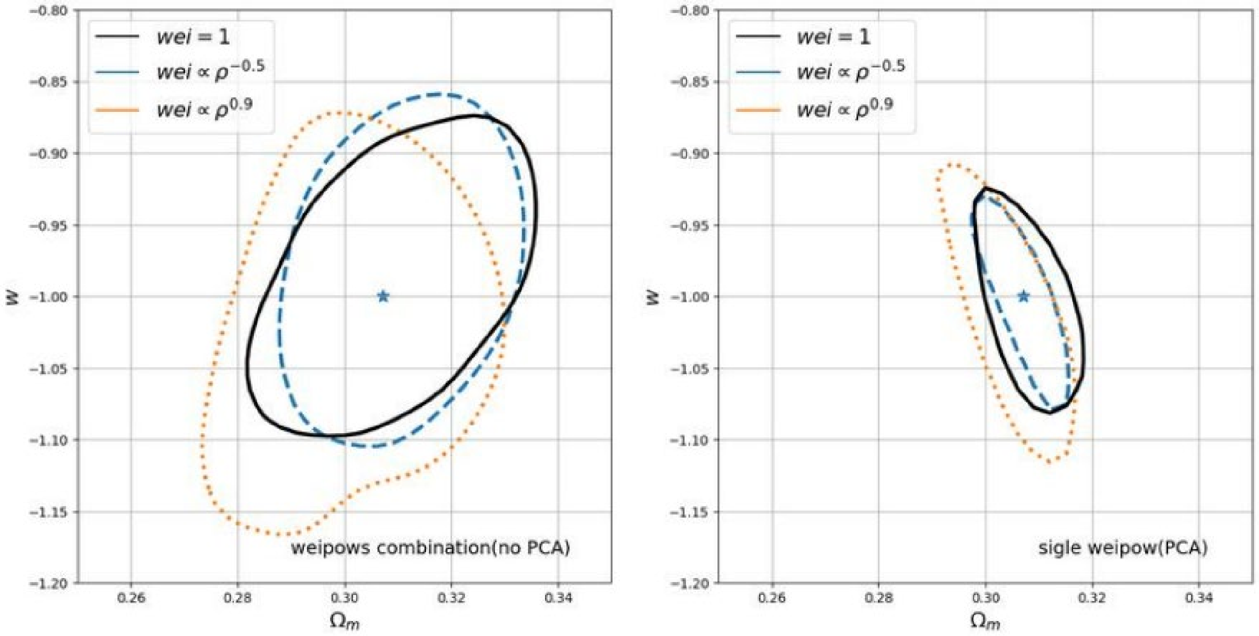
Velocity Reconstruction assisted by PT



Adding PS/2pcf to data vector

Parameter estimation assisted by PS/2pcf

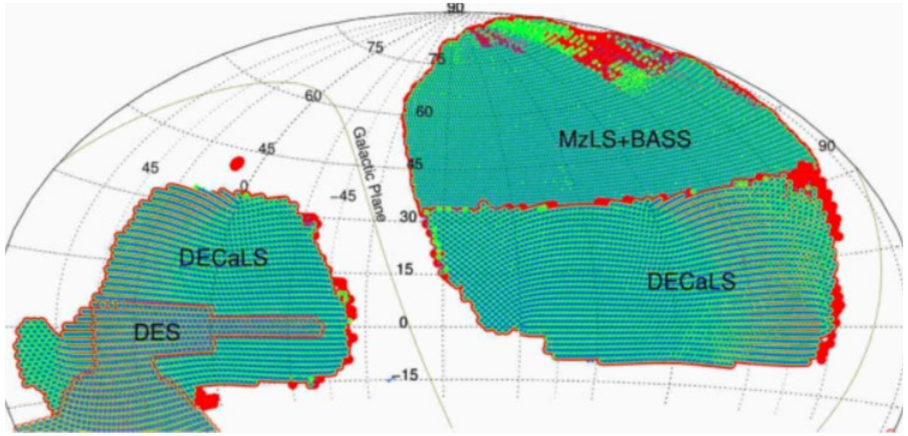
Using PCA result in Tomographic AP analysis



Tomographic AP assisted by PCA

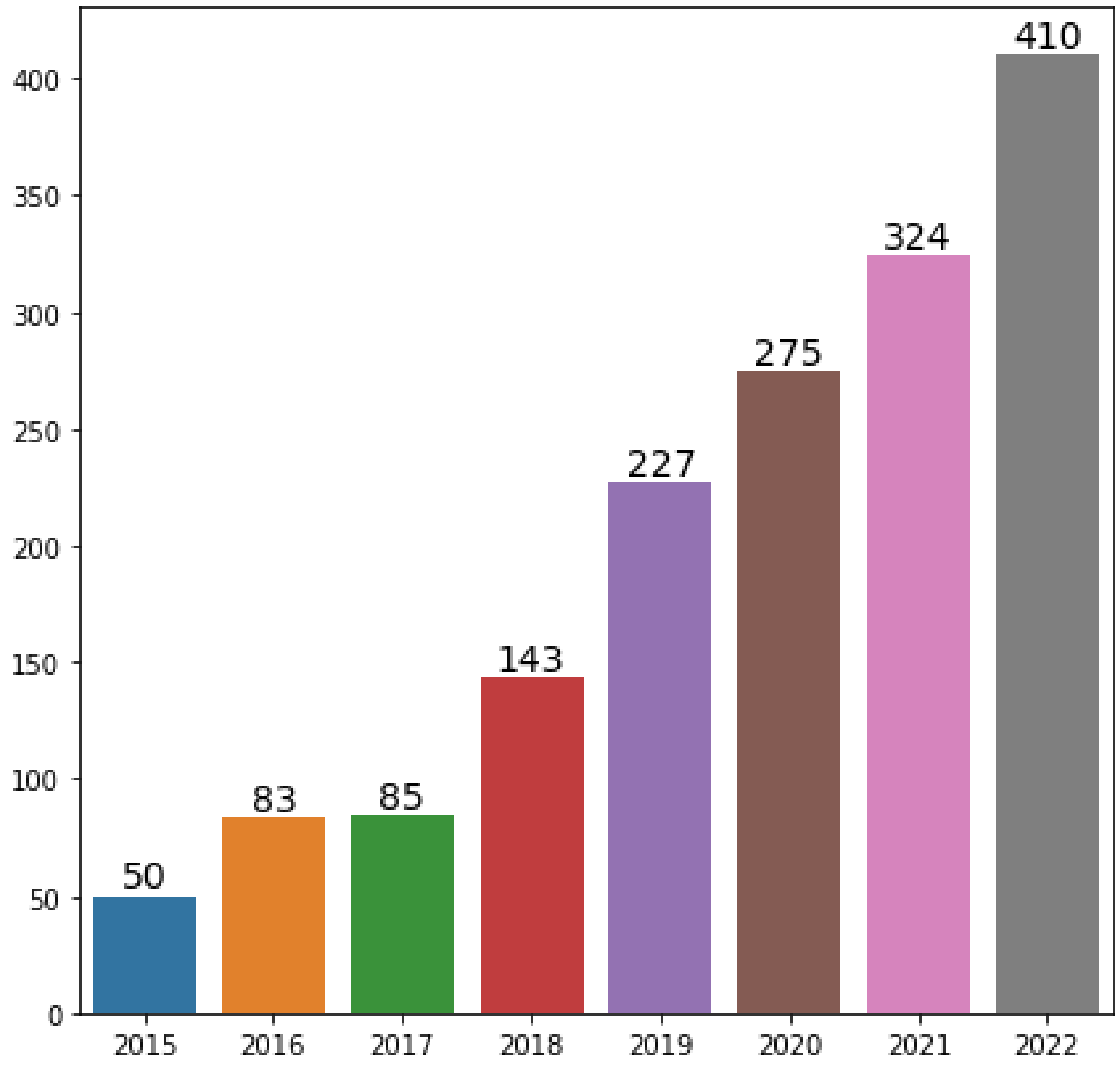
(test on BigMD simulation @ z=0.6 / z=1)

Challenges for AI analysis (from Stage-IV surveys)

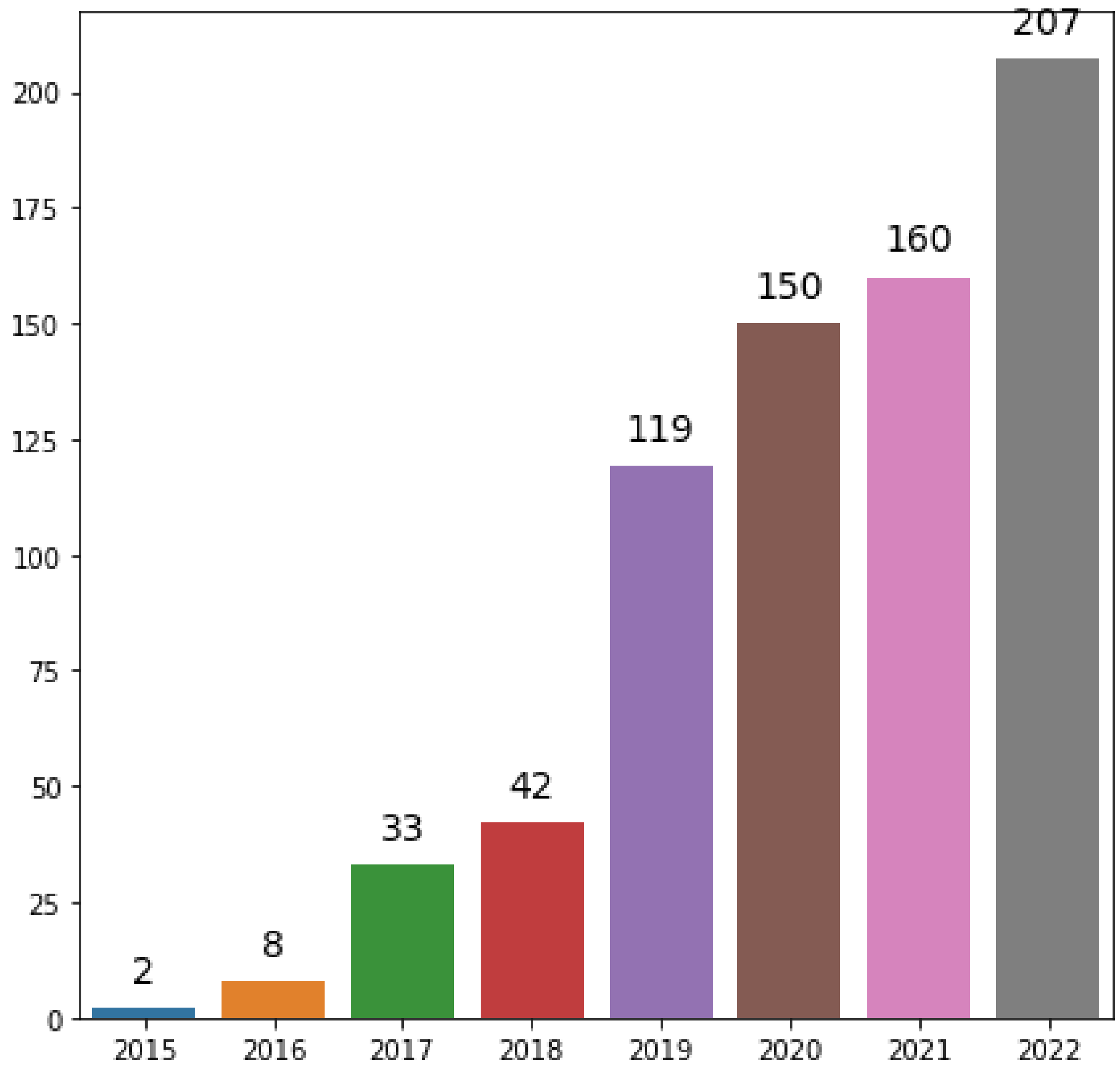


- Big data -> Much bigger training set?
- Lightcone
- Systematics
  - halo-galaxy connection
  - CSST slitless systematics (redshift error, overlapping)
  - ...

AI @ astro： arXiv paper burst



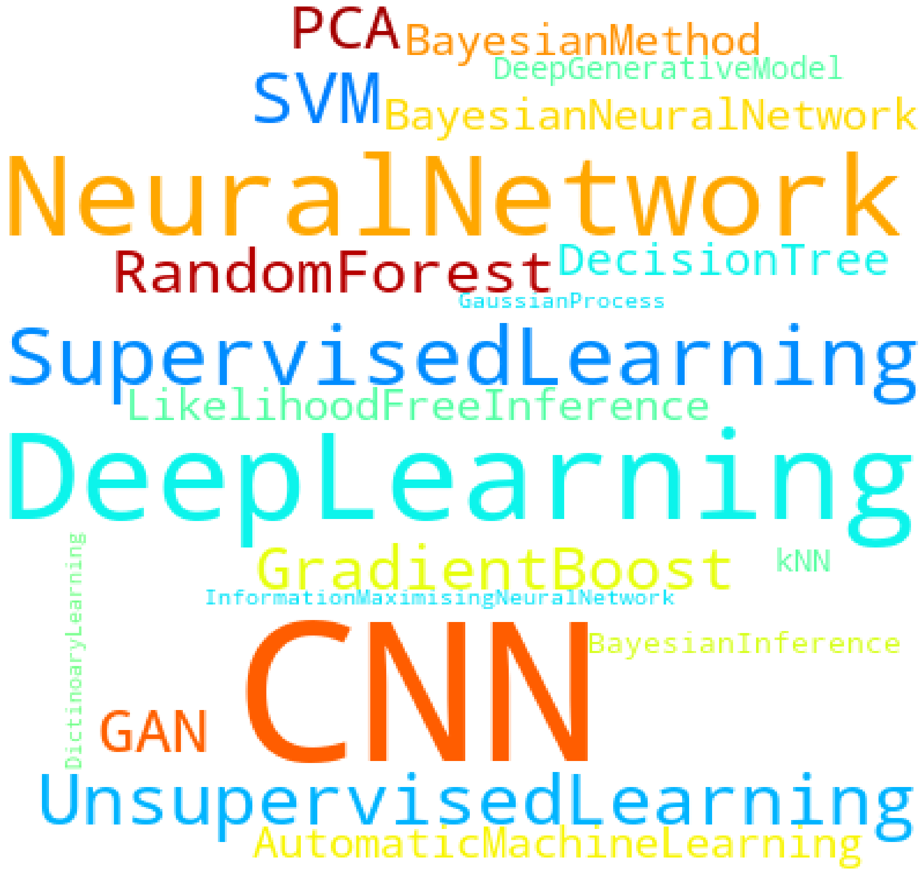
Abstract contains "Machine Learning"



Abstract contains "Deep Learning"

But, 科学分析 相对数据处理仍然较少

For Euclid survey, I found 54 arxiv Papers Containing “Euclid” and “Machine Learning” in abstract



目前 AI 趋势：大模型 + 预训练

	~2016: 上一轮人工智能	2023：以大模型为代表的 人工智能工业化时期
主流模型	以ResNet，VGG，AlexNet为代表的卷积神经网络模型	以GPT、BERT为代表的Transformer神经网络模型
参数量	几亿量级 ResNet-50 (4600万), VGG (1.4亿)	百亿-万亿 GPT-3 (1750亿) Switch-Transformer (1.6万亿)
主要计算	二维卷积，矩阵乘法	矩阵乘法为主
算力需求	单机单卡-单机多卡 +百GB内存空间 ResNet-50 用一块V100训练约3天	大规模并行训练 + PB内存空间 GPT-3 训练浮点运算次数：314ZFLOPTs 相当于100PFLOPs训练36.4天

（from 2023人工智能前沿交叉论坛， 陈左宁院士报告）

Transformer in one slide

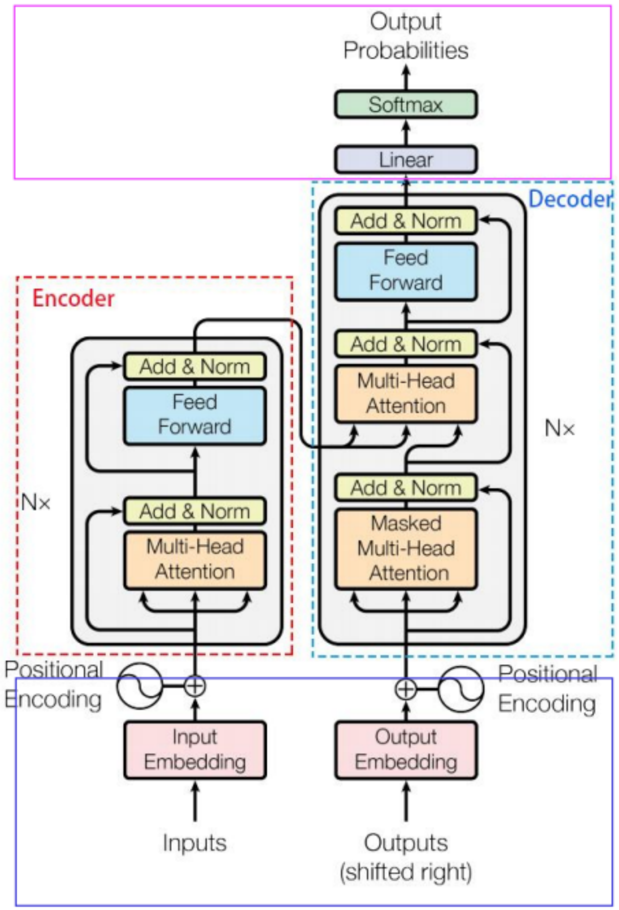
使用“自注意力（Self-attention）”机制捕捉序列内部关系的深度学习模型，广泛用于 NLP

(4) Output

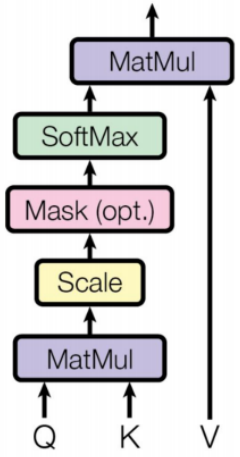
(3)Decoder

(2) Encoder

(1) Input

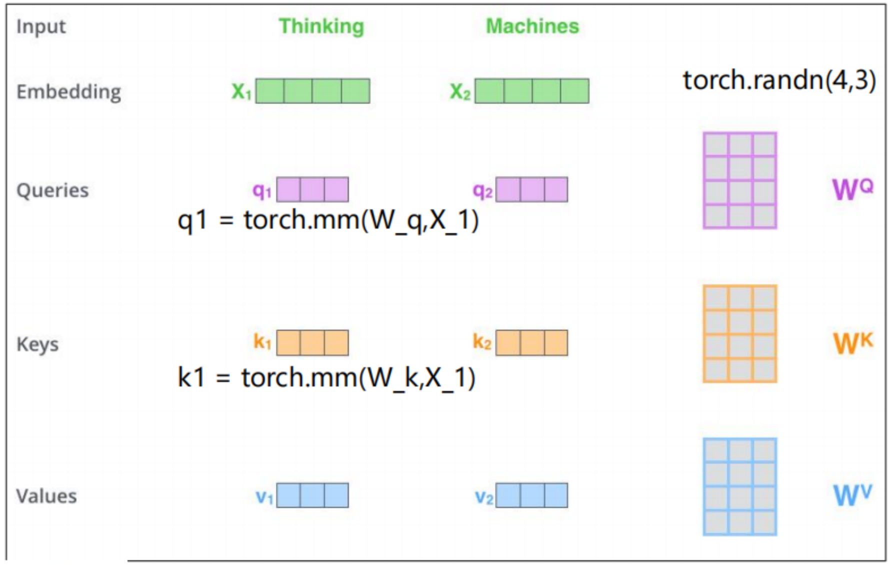


Scaled Dot-Product Attention



$$\mathrm{Attention}(Q,K,V)=\mathrm{softmax}(\frac{QK^T}{\sqrt{d_k}})V$$

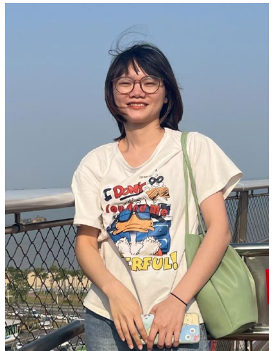
自注意力机制使模型能够在序列中的不同位置之间建立关联，捕捉到输入序列中各个元素之间的依赖关系。



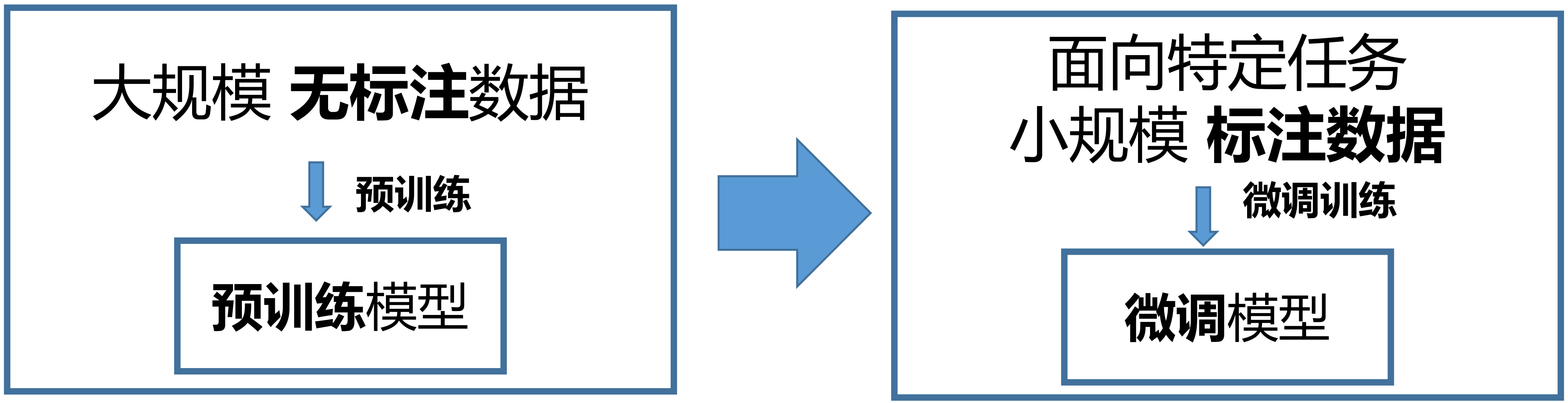
(batch_size, seq_len, d_model)

$dk = dv = d_{model}/h$

（from 李倩 @ our group）



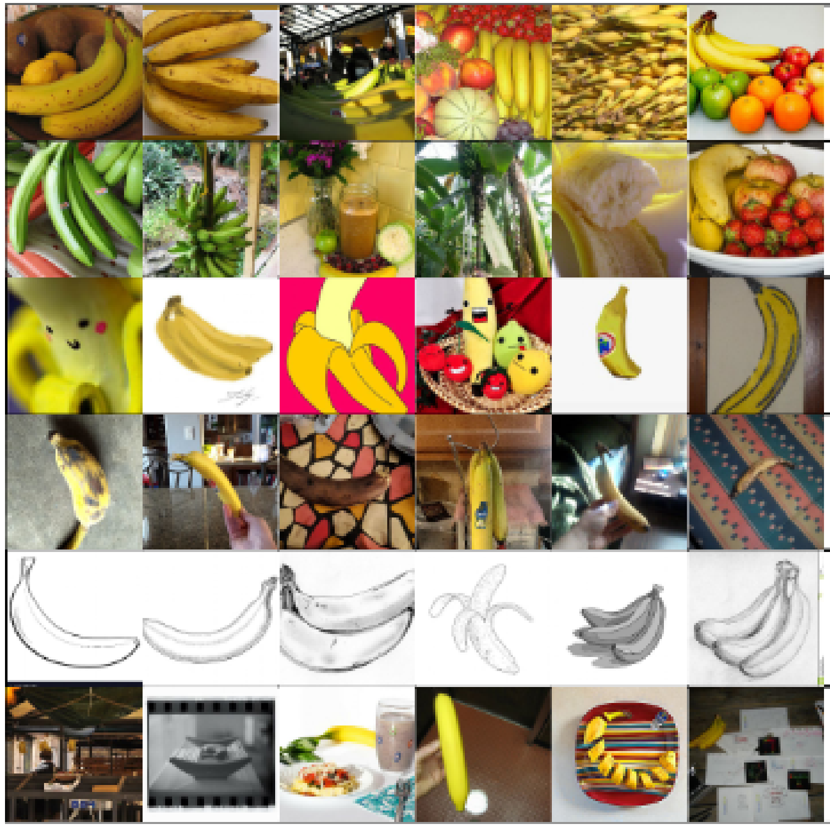
目前 AI 趋势：大模型 + 预训练



（from 2023人工智能前沿交叉论坛，陈左宁院士报告）

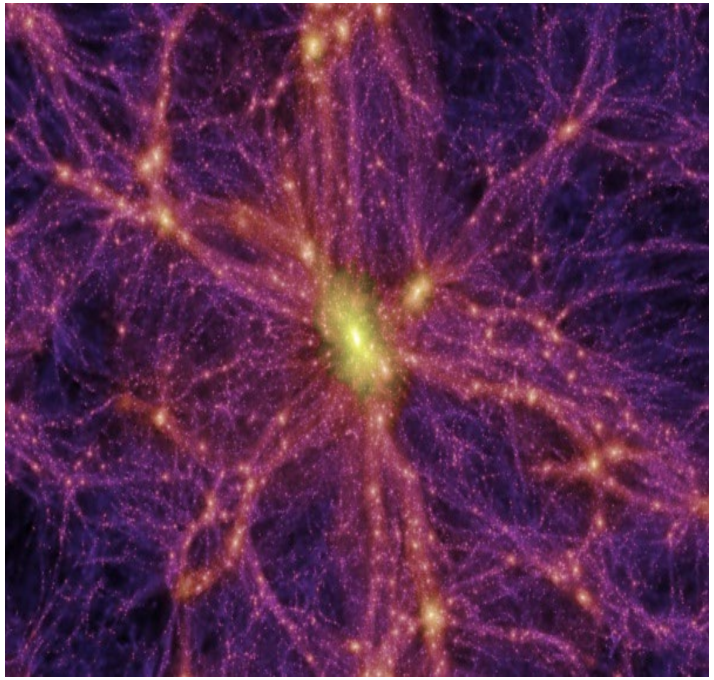
大模型 + 预训练 （即将）解决通用 CV 问题，
但 科学数据的通用分析 仍难以实现

Dataset Examples



	ImageNet ResNet101	Zero-Shot CLIP	Δ Score
ImageNet	76.2	76.2	0%
ImageNetV2	64.3	70.1	+5.8%
ImageNet-R	37.7	88.9	+51.2%
ObjectNet	32.6	72.3	+39.7%
ImageNet Sketch	25.2	60.2	+35.0%
ImageNet-A	2.7	77.1	+74.4%

CLIP (Contrastive Language-Image Pre-training)，
arXiv:2103.00020



我认为科学数据分析的大模型目前还难以实现。
特定科学问题仍需使用特定数据训练特定模型。
(answer from AI experts @ meeting)

专业科学问题；系统性效应多且敏感；数据规模巨大；专业特征有别于通用特征 （e.g. 宇宙学数据的特征、噪音与传统问题恰好相反）

Thank you!