



Low-cost neural summaries for cosmological inference

AN ENSEMBLE OF CONVOLUTIONAL NEURAL NETWORKS TRAINED ON APPROXIMATE SIMULATIONS

Iñigo Sáez-Casares

November 3rd, 2026

with: Davide Bianchi, Matteo Calabrese, Luigi Guzzo

Summary statistics for cosmological fields

$\longrightarrow$Extracting cosmological information from the large-scale structure:

- marginalise over initial conditions
- we compress cosmological fields with summary statistics
- e.g.: power spectrum, two-point correlation function

Summary statistics for cosmological fields

$\longrightarrow$Extracting cosmological information from the large-scale structure:

- marginalise over initial conditions
- we compress cosmological fields with summary statistics
- e.g.: power spectrum, two-point correlation function

$\longrightarrow$Non-Gaussian information at small scales

- accessed with high-order statistics
- traditional N-point statistics: bispectrum, trispectrum, ...
- other statistics: wavelet scattering transform, density splits, ...

Summary statistics for cosmological fields

$\longrightarrow$Extracting cosmological information from the large-scale structure:

- marginalise over initial conditions
- we compress cosmological fields with summary statistics
- e.g.: power spectrum, two-point correlation function

$\longrightarrow$Non-Gaussian information at small scales

- accessed with high-order statistics
- traditional N-point statistics: bispectrum, trispectrum, ...
- other statistics: wavelet scattering transform, density splits, ...

$\longrightarrow$Alternative: use neural networks to find a near-optimum compression, learnt from simulations.

Building a neural summary

Goal: train a neural network to compress density fields into a few coefficients

$\longrightarrow$These coefficients form a summary statistic

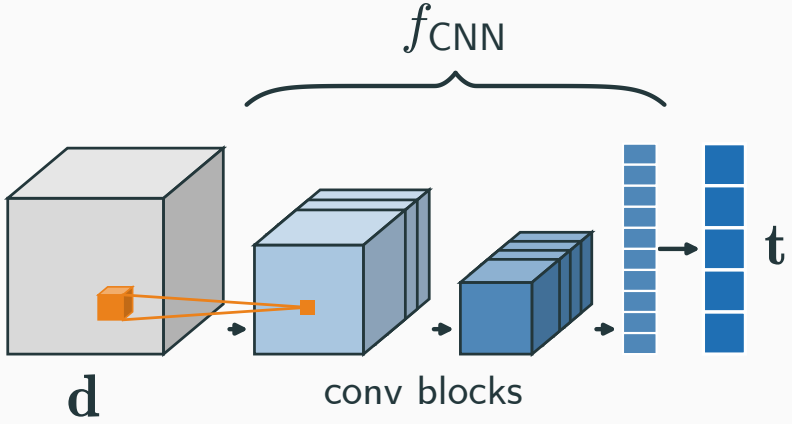
$\longrightarrow$They need to carry information about the cosmological parameters

Building a neural summary

Goal: train a neural network to compress density fields into a few coefficients

$\longrightarrow$These coefficients form a summary statistic

$\longrightarrow$They need to carry information about the cosmological parameters



Convolutional neural network (CNN):

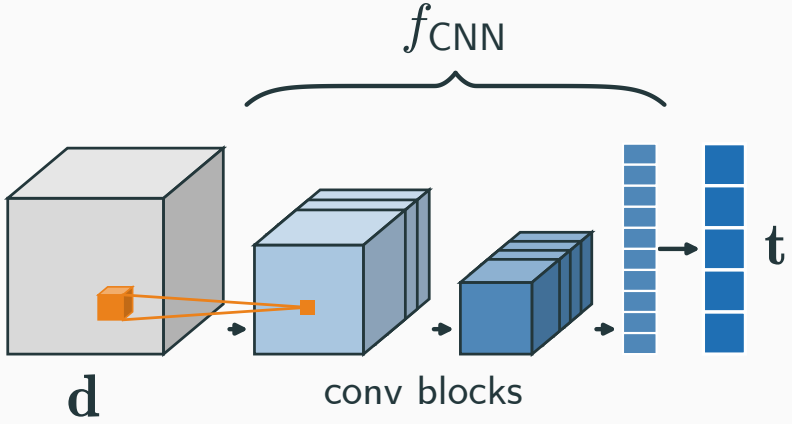
- Input: density on a 3D grid $\mathbf{d}$
- Output: 1D vector $\mathbf{t} = f_{\mathrm{CNN}}(\mathbf{d})$
- f_{CNN}: convolution kernels with free parameters that need to be optimised (trained)

Building a neural summary

Goal: train a neural network to compress density fields into a few coefficients

$\longrightarrow$These coefficients form a summary statistic

$\longrightarrow$They need to carry information about the cosmological parameters



Convolutional neural network (CNN):

- Input: density on a 3D grid $\mathbf{d}$
- Output: 1D vector $\mathbf{t} = f_{\mathrm{CNN}}(\mathbf{d})$
- f_{CNN}: convolution kernels with free parameters that need to be optimised (trained)

Training: minimise a loss function $\mathcal{L}$

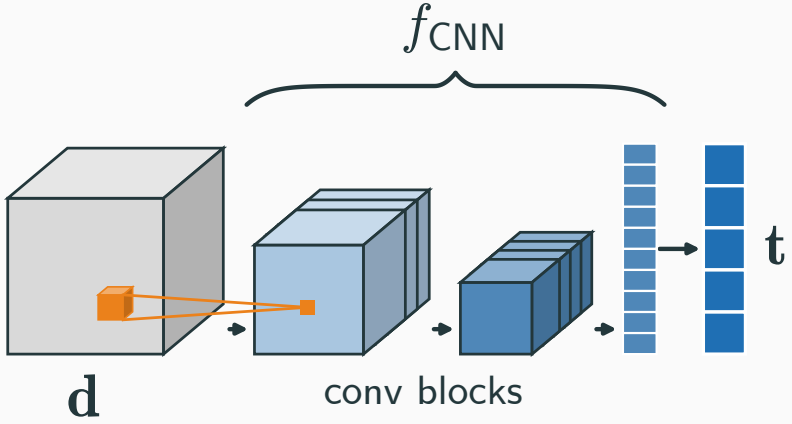
$\longrightarrow$Simplest choice: $\mathcal{L} = (\mathbf{t} - \boldsymbol{\theta})^2$

Building a neural summary

Goal: train a neural network to compress density fields into a few coefficients

$\longrightarrow$These coefficients form a summary statistic

$\longrightarrow$They need to carry information about the cosmological parameters



Convolutional neural network (CNN):

- Input: density on a 3D grid $\mathbf{d}$
- Output: 1D vector $\mathbf{t} = f_{\mathrm{CNN}}(\mathbf{d})$
- f_{CNN}: convolution kernels with free parameters that need to be optimised (trained)

Training: minimise a loss function $\mathcal{L}$

$\longrightarrow$Simplest choice: $\mathcal{L} = (\mathbf{t} - \boldsymbol{\theta})^2$

Challenge: requires large number of simulated samples ($> 10^4$)

$\longrightarrow$Using realistic simulations is expensive

$\longrightarrow$*Can we rely on approximate simulations instead?*

Cosmological inference

Likelihood:

- assume Gaussian likelihood
- $\log \mathcal{L} \sim (\mathbf{t}_{\rm theo}(\boldsymbol{\theta}) - \mathbf{t}_{\rm data})^\top \mathbf{C}^{-1} (\mathbf{t}_{\rm theo}(\boldsymbol{\theta}) - \mathbf{t}_{\rm data})$
- no predictions for $\mathbf{t}_{\rm theo}$: need emulator trained on simulations

Cosmological inference

Likelihood:

- assume Gaussian likelihood
- $\log \mathcal{L} \sim (\mathbf{t}_{\rm theo}(\boldsymbol{\theta}) - \mathbf{t}_{\rm data})^\top \mathbf{C}^{-1} (\mathbf{t}_{\rm theo}(\boldsymbol{\theta}) - \mathbf{t}_{\rm data})$
- no predictions for $\mathbf{t}_{\rm theo}$: need emulator trained on simulations

Emulator:

- needs (expensive) realistic training simulations ($> 10^2$)
- public suites: Abacus, Bacco, ...

Cosmological inference

Likelihood:

- assume Gaussian likelihood
- $\log \mathcal{L} \sim (\mathbf{t}_{\rm theo}(\boldsymbol{\theta}) - \mathbf{t}_{\rm data})^\top \mathbf{C}^{-1} (\mathbf{t}_{\rm theo}(\boldsymbol{\theta}) - \mathbf{t}_{\rm data})$
- no predictions for $\mathbf{t}_{\rm theo}$: need emulator trained on simulations

Emulator:

- needs (expensive) realistic training simulations ($> 10^2$)
- public suites: Abacus, Bacco, ...

Covariance:

- $\mathbf{C} = \mathbf{C}_{\rm data} + \mathbf{C}_{\rm emu}$
- $\mathbf{C}_{\rm data}$: computed from mocks, cosmology independent
- $\mathbf{C}_{\rm emu}$: computed from cross-validation, cosmology independent
- Ideally $\mathbf{C}_{\rm emu} << \mathbf{C}_{\rm data}$

$\longrightarrow$Standard inference method for simulation-based models

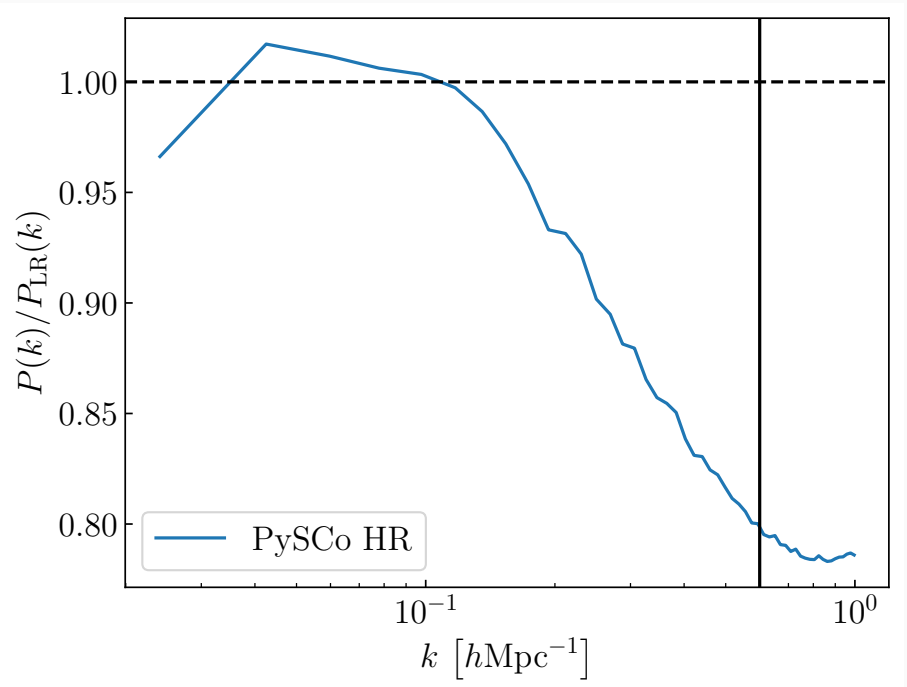
Simulation sets

Simulations:

- particle-mesh (PySCo, Breton+25)
- $L_{\rm box} = 328\ h^{-1}{\rm Mpc}$
- LR: $N_{\rm part} = 128^3$
- HR: $N_{\rm part} = 512^3$
- CNN training: 8192 LR sims.
- Emulator training: 256 HR sims.
- Covariance: 1000 HR sims.

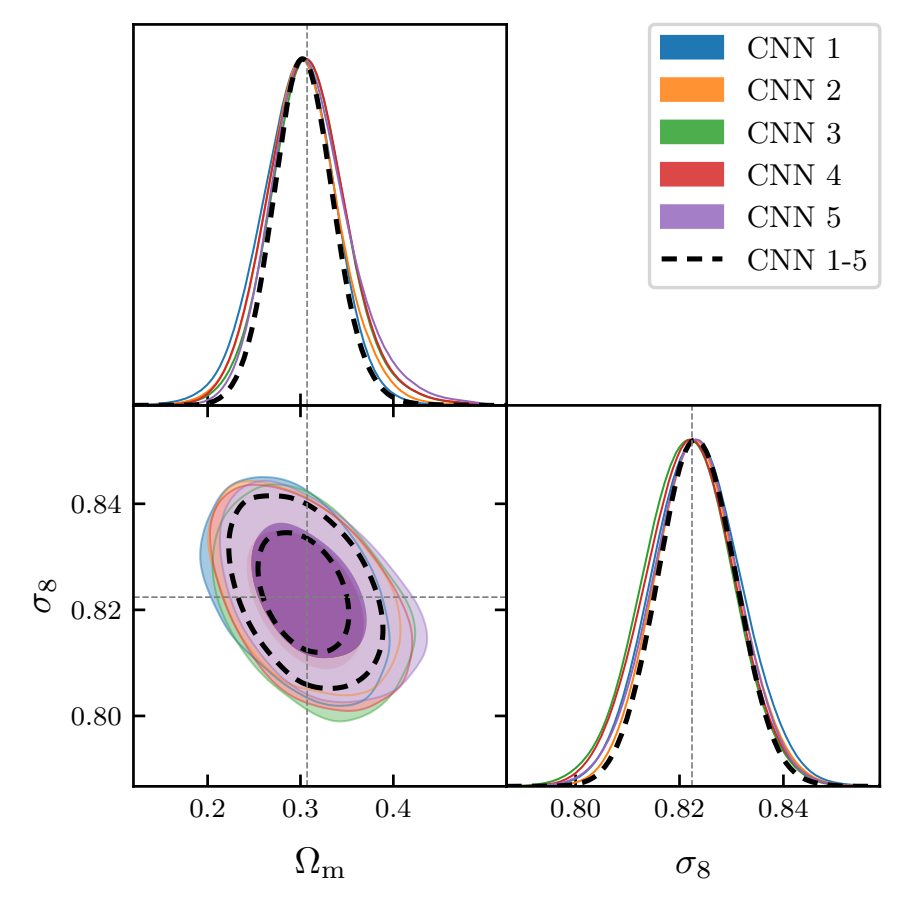
Datasets:

- matter fields (64^3 grid)
- power spectrum
- $k_{\rm Nyq}^{\rm CNN} \sim k_{\rm max}^{\rm PS} \sim 0.6\ h{\rm Mpc}^{-1}$
- $z = 0$



Cosmological constraints from the neural summary

Low resolution set



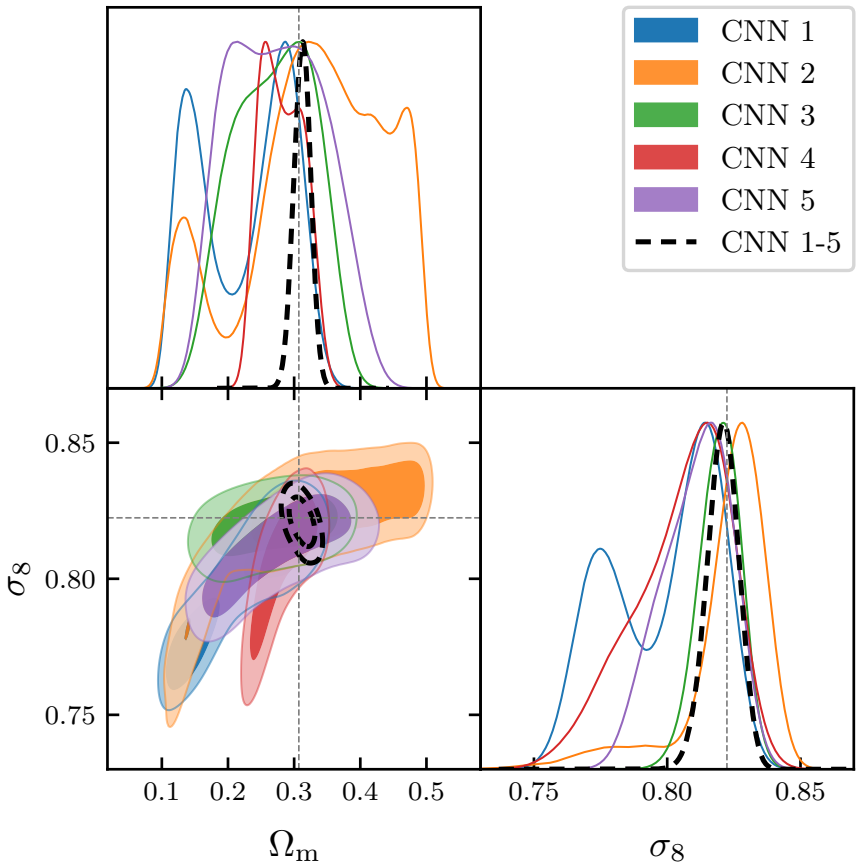
MCMC: vary Ω_m and σ_8 (others fixed)

Inference setup: synthetic data, covariance matrix, emulator built from low resolution simulations

- 5 CNNs with the best final loss
- similar performance for all of them
- combination: slight improvement

Cosmological constraints from the neural summary

High resolution set

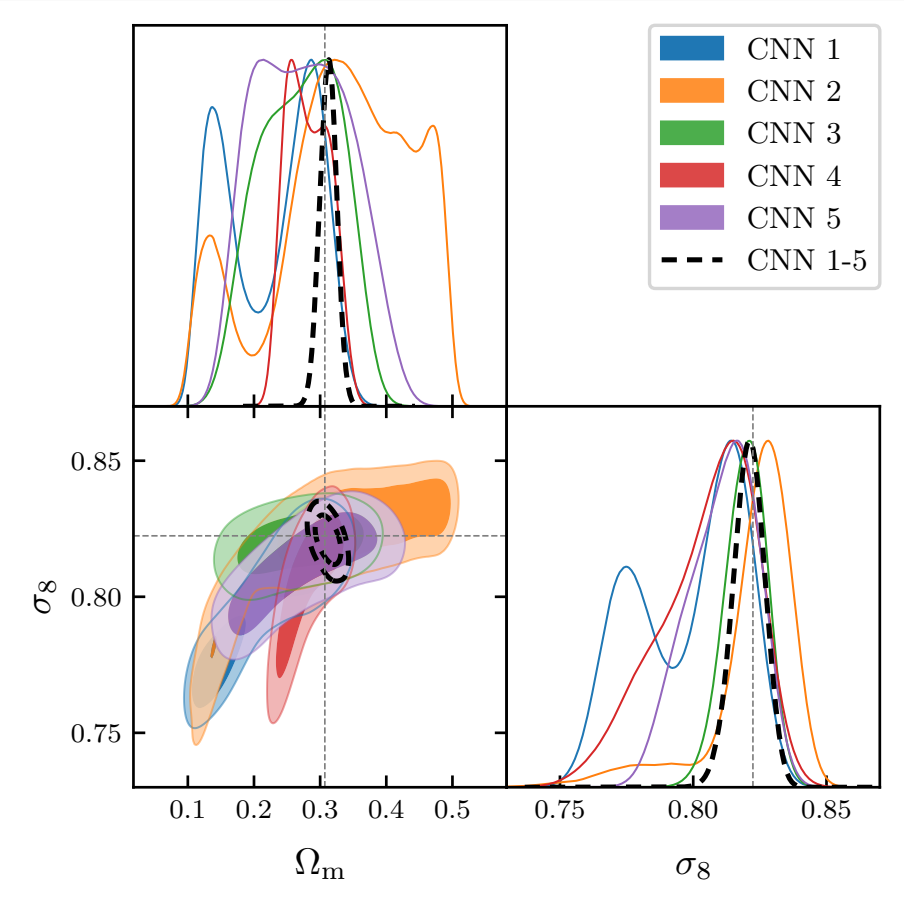


Inference setup: synthetic data, covariance matrix, emulator built from high resolution simulations

- significantly worse constraints
- large variations between CNNs
- combination: large improvement!

Cosmological constraints from the neural summary

High resolution set



Inference setup: synthetic data, covariance matrix, emulator built from high resolution simulations

- significantly worse constraints
- large variations between CNNs
- combination: large improvement!

$\longrightarrow$Neural summary trained on low resolution samples does not give an informative compression for high resolution data

Strategy: build a neural summary from an ensemble of networks

A neural ensemble summary statistic

Ensemble generation: multiple CNNs with enough variety

- CNNs with different architectures (e.g.: number of convolution layers)
- train each CNN independently with the same dataset

A neural ensemble summary statistic

Ensemble generation: multiple CNNs with enough variety

- CNNs with different architectures (e.g.: number of convolution layers)
- train each CNN independently with the same dataset

Ensemble pruning: select sub-ensemble to build the summary statistic

$\longrightarrow$quantify performance with posterior volume: $\mathcal{V}_{\rm p} = \sqrt{\det \mathbf{C}_{\boldsymbol{\theta}}}$

$\longrightarrow$find the smallest sub-ensemble that minimises $\mathcal{V}_{\rm p}$

$\longrightarrow$brute force approach: try all possible combinations, too expensive!

A neural ensemble summary statistic

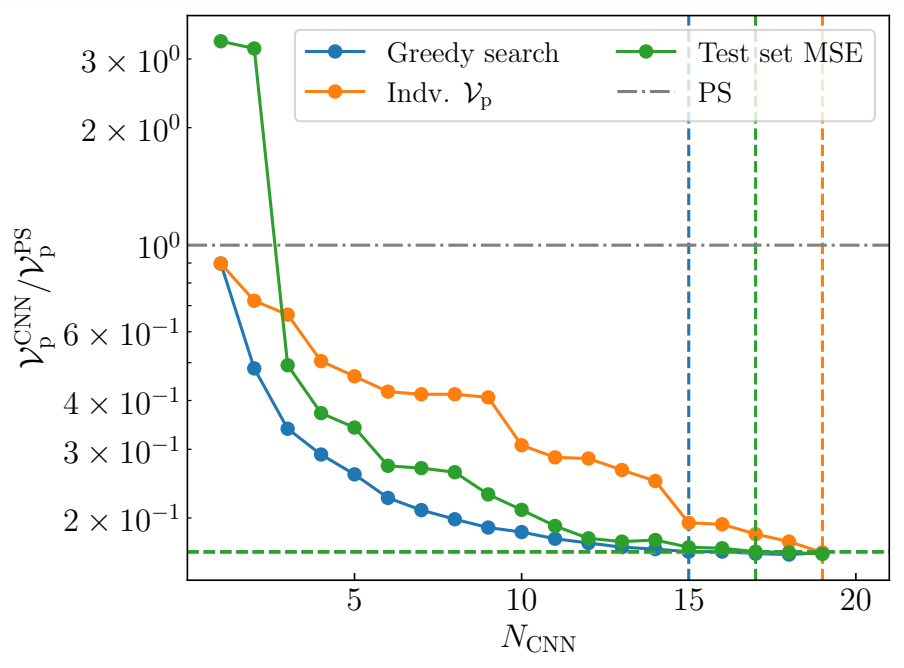
Ensemble generation: multiple CNNs with enough variety

- CNNs with different architectures (e.g.: number of convolution layers)
- train each CNN independently with the same dataset

Ensemble pruning: select sub-ensemble to build the summary statistic
$\longrightarrow$quantify performance with posterior volume: $\mathcal{V}_{\rm p} = \sqrt{\det \mathbf{C}_{\boldsymbol{\theta}}}$
$\longrightarrow$find the smallest sub-ensemble that minimises $\mathcal{V}_{\rm p}$
$\longrightarrow$brute force approach: try all possible combinations, too expensive!

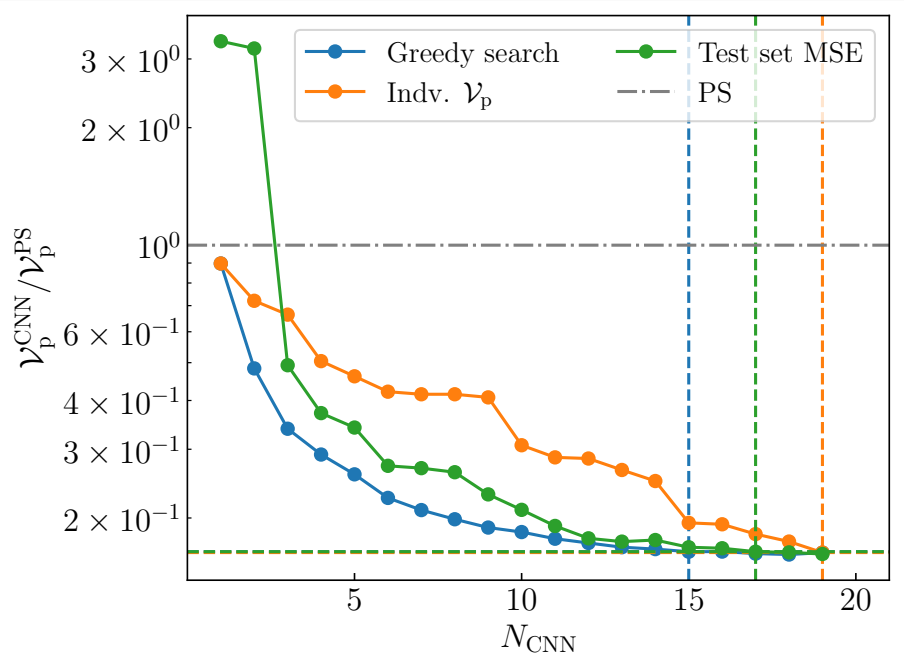
- Option 1: rank CNNs by loss function, pick $N_{\rm CNN}$ best ones
  $\longrightarrow$no information from HR sims.
- Option 2: rank CNNs by $\mathcal{V}_{\rm p}$, pick $N_{\rm CNN}$ best ones
  $\longrightarrow$no complementarity between CNNs
- Option 3: greedy search to minimise $\mathcal{V}_{\rm p}$

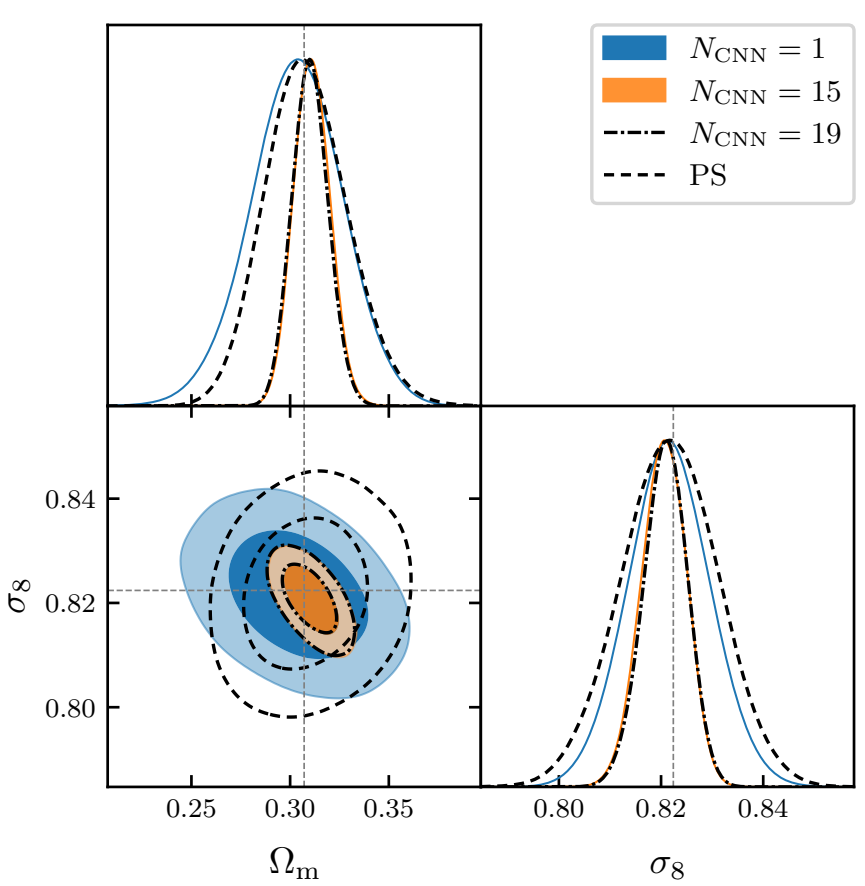
Comparison of pruning methods



- Best CNN from LR metric does not give best constraints on HR
- Greedy search gives best constraints at each N_CNN

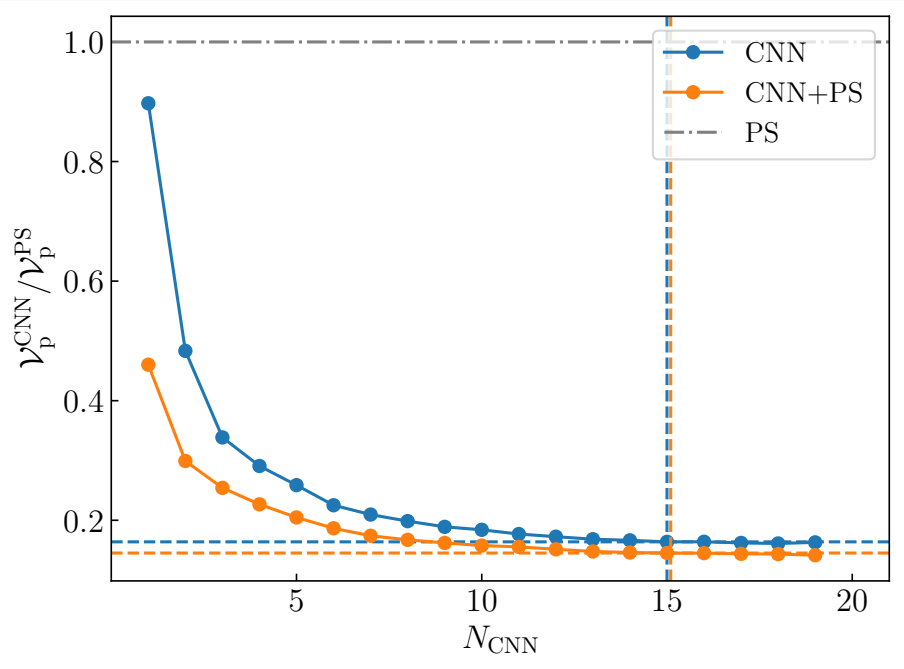
Comparison of pruning methods



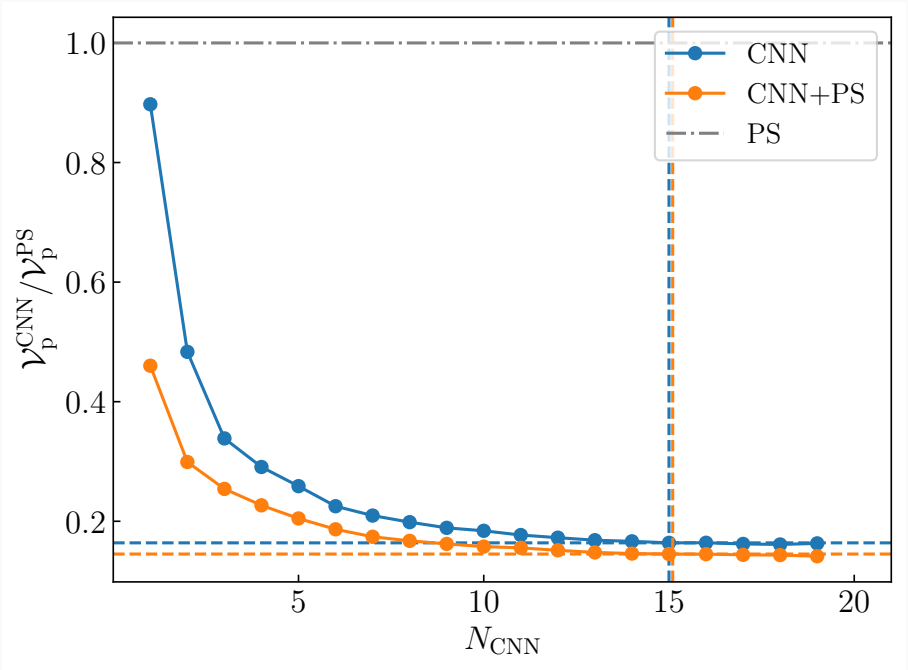


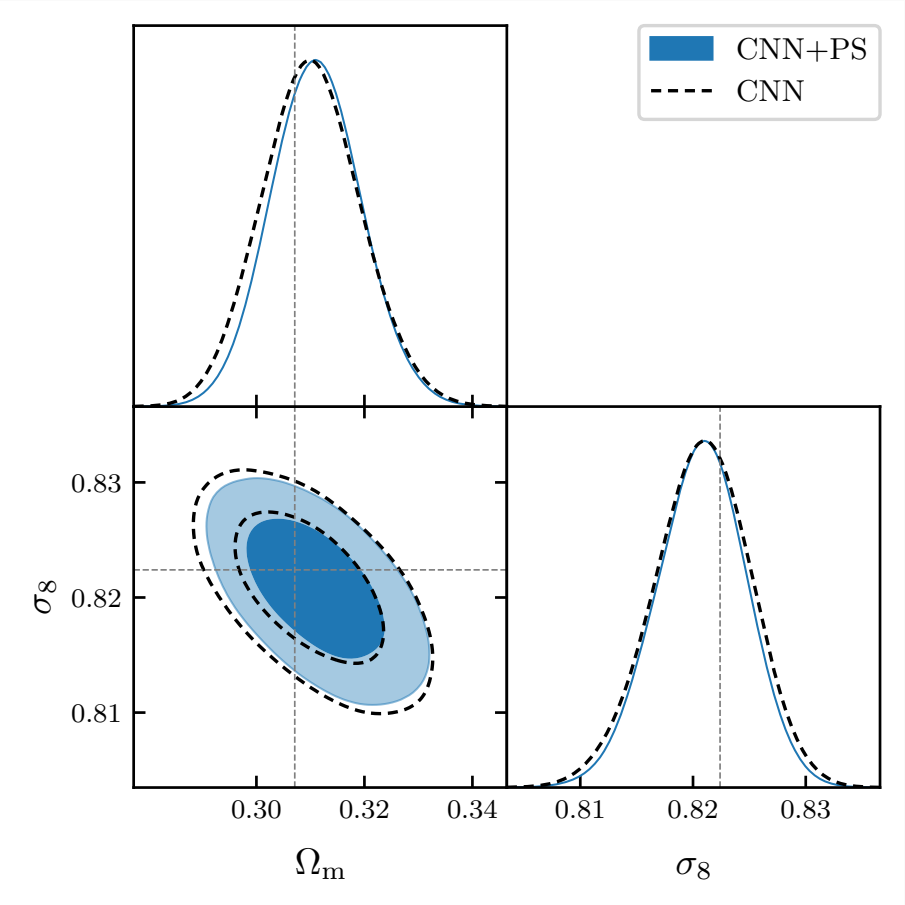
- Best CNN from LR metric does not give best constraints on HR
- Greedy search gives best constraints at each N_CNN

Combination with the power spectrum



Combination with the power spectrum





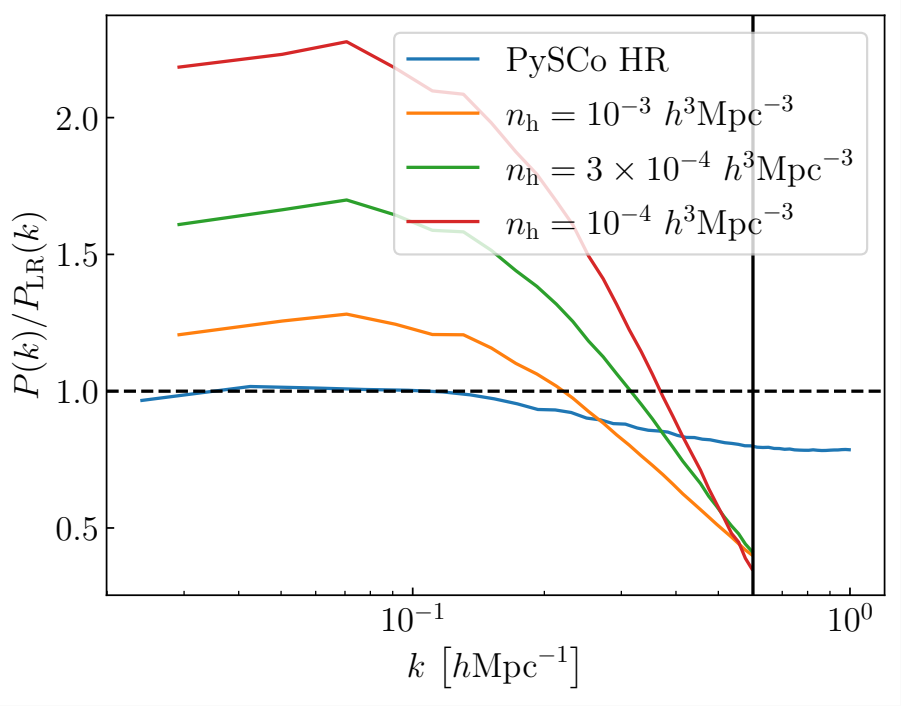
$\longrightarrow$greedy search does not converge sooner

$\longrightarrow$CNN captures most of the information already

Application to dark matter haloes

Halo catalogues:

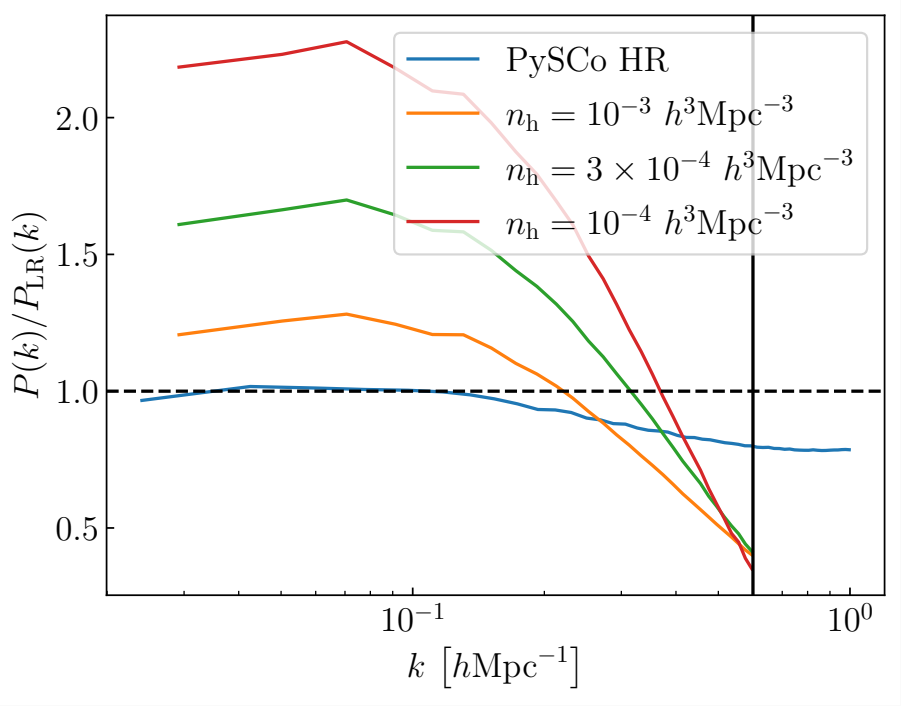
- e-MANTIS suite (Sáez-Casares+24)
- N-body sims. (RAMSES, AMR)
- $L_{\rm box} = 328\ h^{-1}{\rm Mpc}$
- $N_{\rm part} = 512^3$
- Emulator: 80 sims.
- Covariance: 384 sims.
- number density bins: 10^{-3}, $3 \cdot 10^{-4}$, $10^{-4}\ h^3{\rm Mpc}^{-3}$



Application to dark matter haloes

Halo catalogues:

- e-MANTIS suite (Sáez-Casares+24)
- N-body sims. (RAMSES, AMR)
- $L_{\rm box} = 328\ h^{-1}{\rm Mpc}$
- $N_{\rm part} = 512^3$
- Emulator: 80 sims.
- Covariance: 384 sims.
- number density bins: 10^{-3}, $3 \cdot 10^{-4}$, $10^{-4}\ h^3{\rm Mpc}^{-3}$

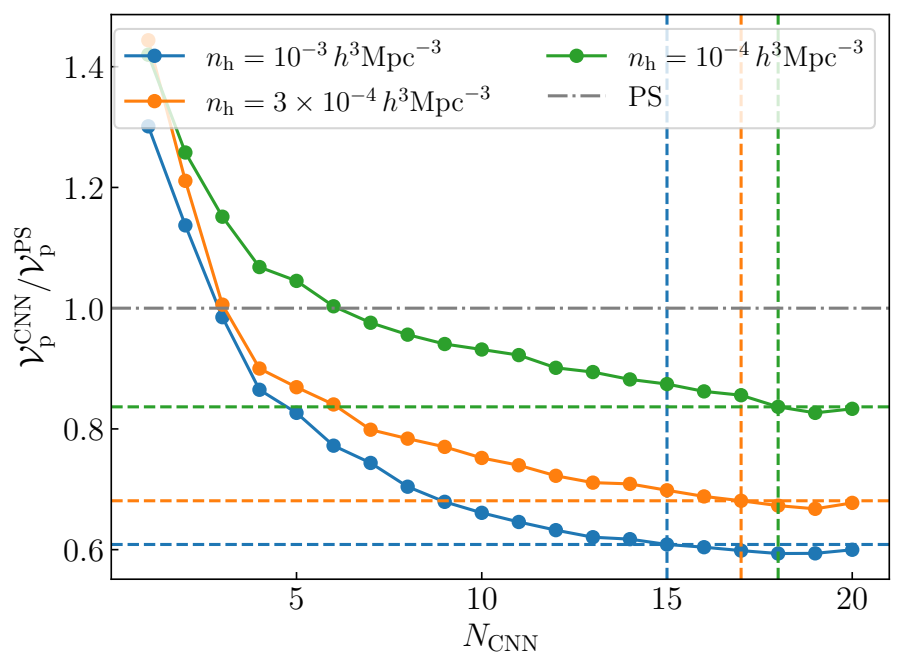


Inference setup:

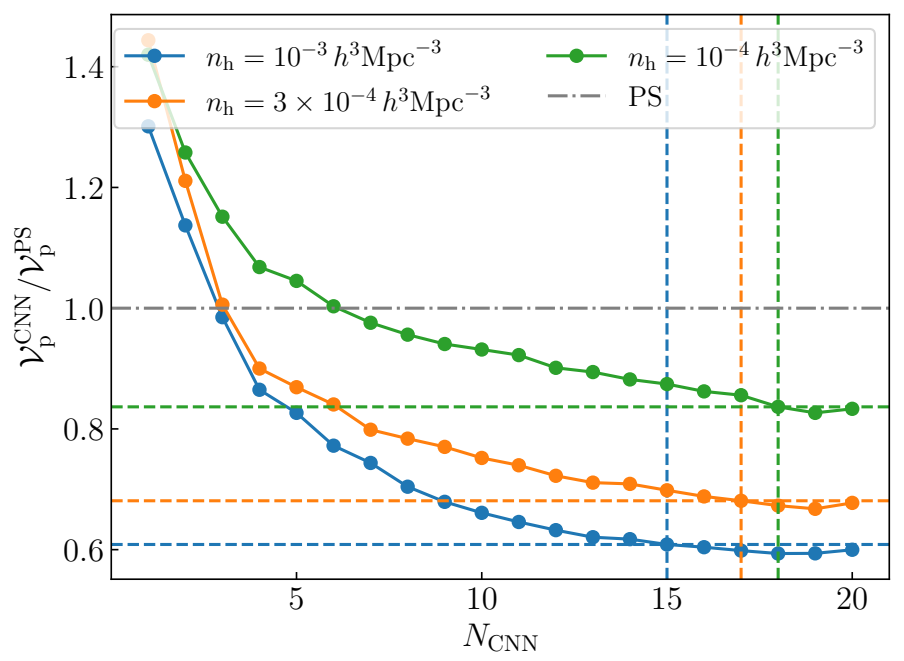
$\longrightarrow$Vary $\Omega_{\rm m}$ and σ_8 (others fixed)

$\longrightarrow$Halo bias known (no cosmology-bias degeneracy)

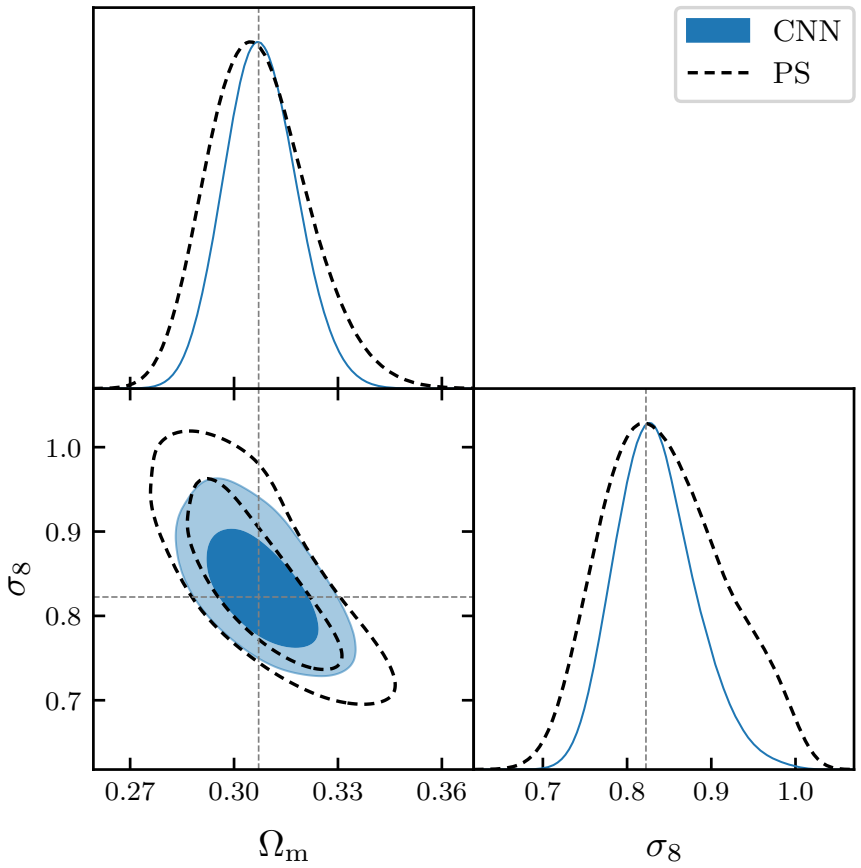
Application to dark matter haloes



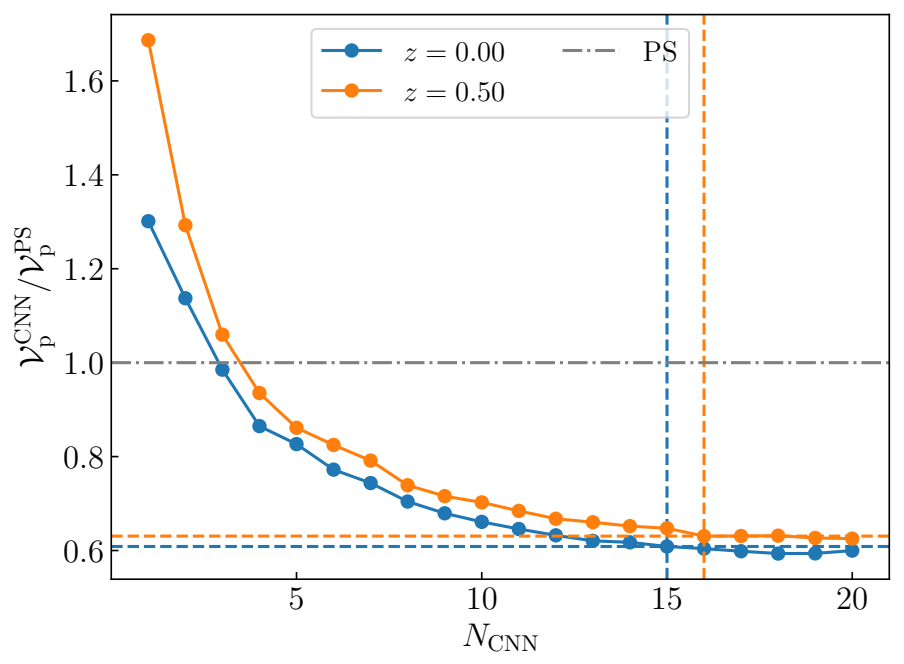
Application to dark matter haloes



$n_\mathrm{h} = 10^{-3}\ h^3\mathrm{Mpc}^3$



Application to dark matter haloes



- Application to different redshift
- Same smooth convergence

Conclusions and perspectives

Main results:

- Informative summary built from an ensemble of networks trained on approximate simulations
- Paves the way to cheaper neural summaries
- Paper: Sáez-Casares+26 (arXiv:2610.19891)

Conclusions and perspectives

Main results:

- Informative summary built from an ensemble of networks trained on approximate simulations
- Paves the way to cheaper neural summaries
- Paper: Sáez-Casares+26 (arXiv:2610.19891)

Open questions:

- Trade-off between resolution and number of samples
- Open parameter space: cosmology + galaxy modeling (bias, HOD, etc.)
- What is the best way to generate the ensemble?