NEURAL INFORMATION PROCESSING SYSTEMS

Domain Feature Enhanced Unisolver for Weak Lensing Inference

THUML, School of Software, Tsinghua University

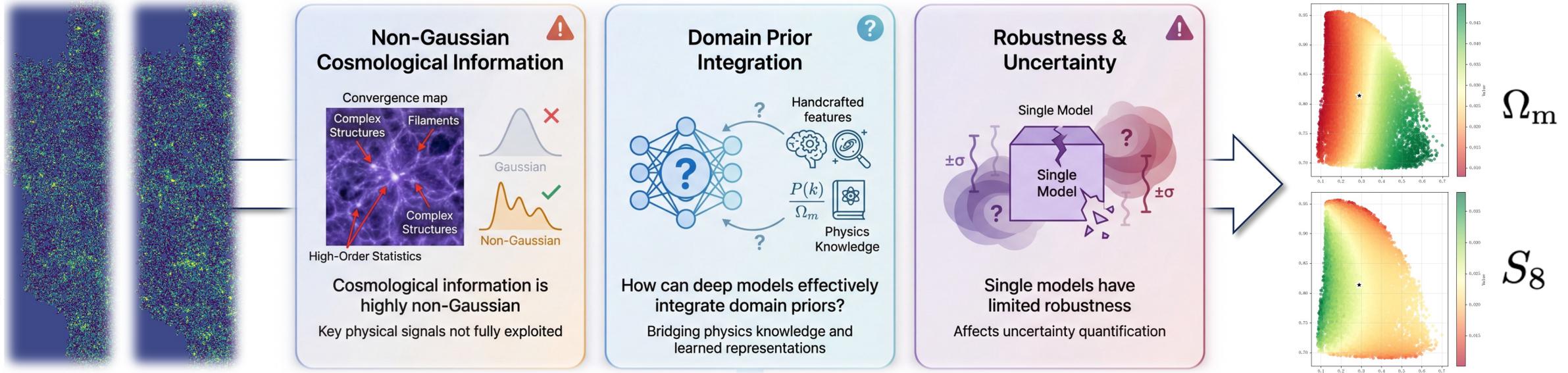Huikun Weng   Haonan ShangGuan   Haixu Wu   Yuezhou Ma   Hang Zhou   Yuanxu Sun   Mingsheng Long
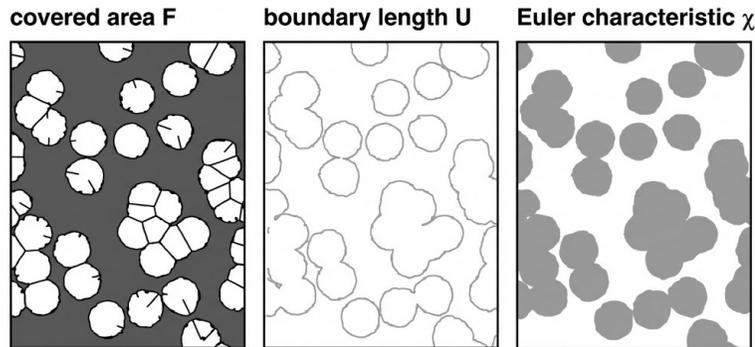
Challenges in Weak Lensing Inference

1. **Scientific problem:** requires capturing physical information behind observation
2. **Learning difficulty:** effectively integrate both deep representations and domain priors
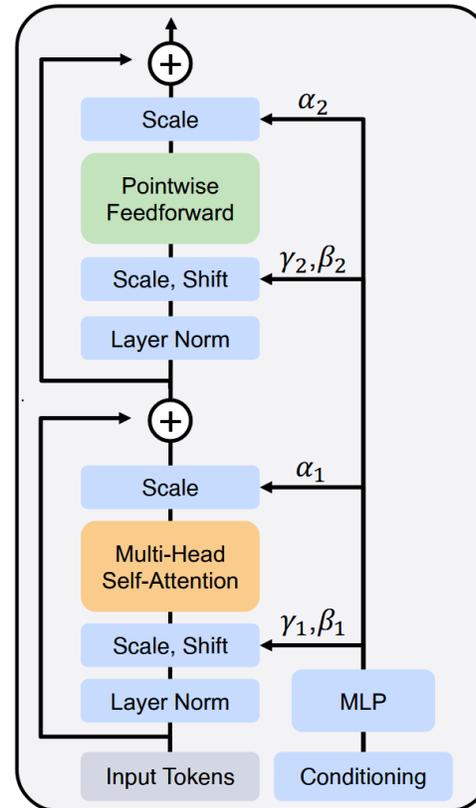3. **Ill-posed problem:** inferring underlying coefficients from partial observations

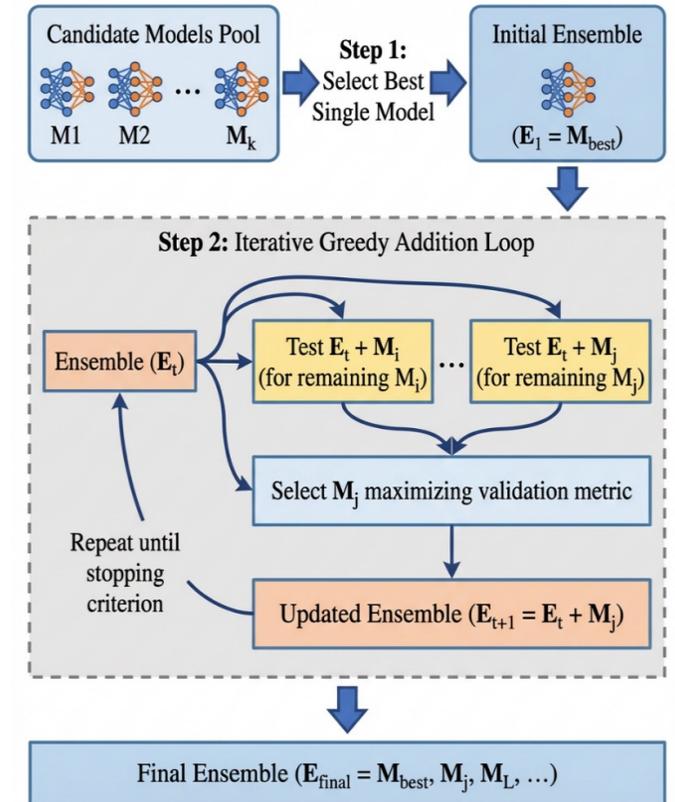# Key Designs in Our Method

(1) Elaborated Domain Priors　　**(2) Specialized Conditional Transformer**　　**(3) Robust Ensemble Algorithm**

covered area F　　boundary length U　　Euler characteristic χ

- Power Spectrum
- Peak Counts
- Minkowski Functionals
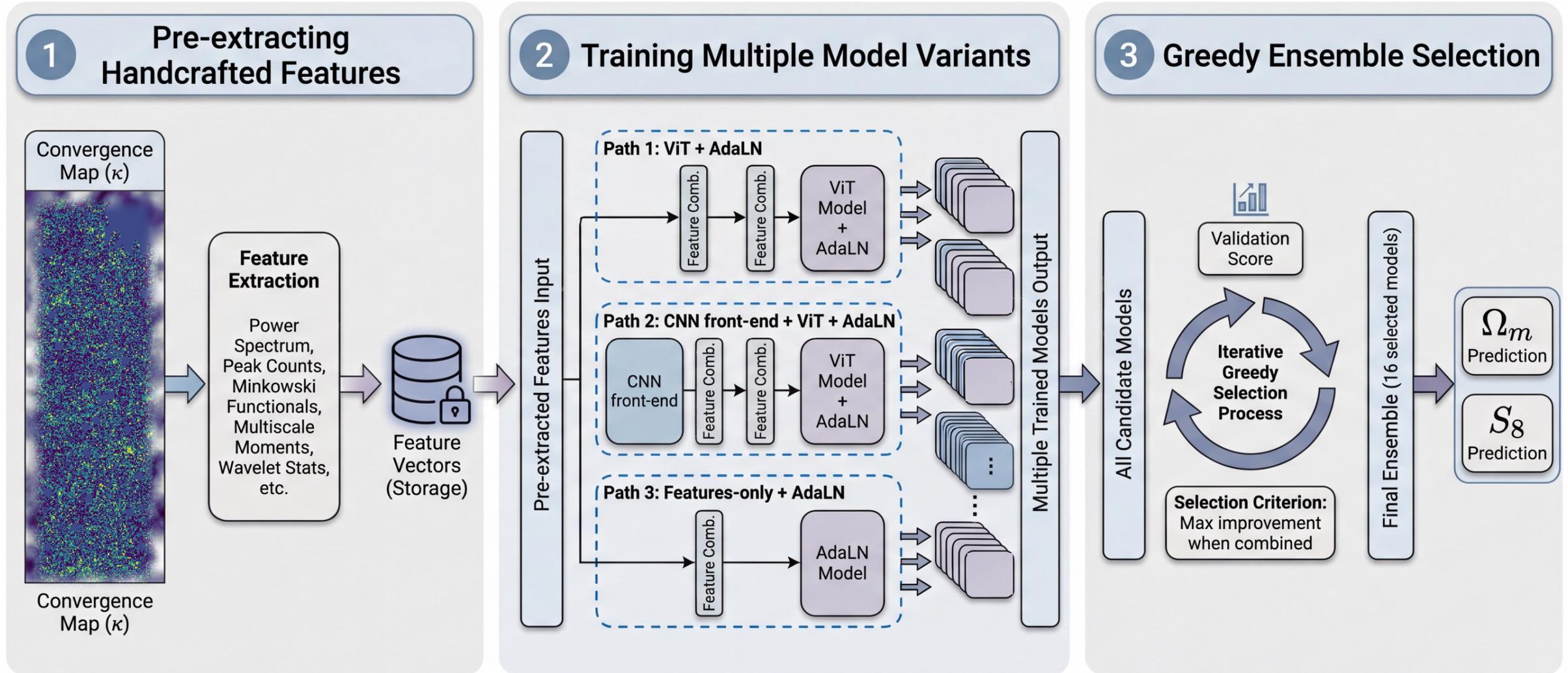- Multiscale Wavelet Statistics
- 1-point PDF
- ... ...

Overall Training and Inference Pipeline
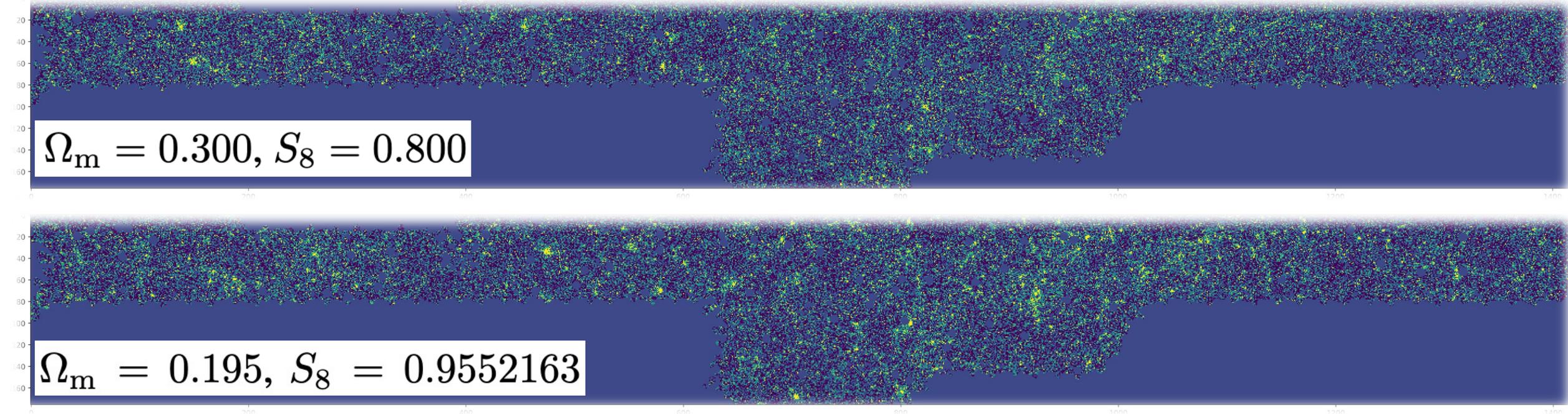
Stage 1: Starting From Domain Priors

$\Omega_m = 0.300, S_8 = 0.800$

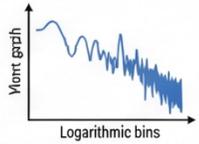$\Omega_m = 0.195, S_8 = 0.9552163$

- Weak lensing is a **highly structured physics signal**, not arbitrary images.

- Domain-specific priors are based on **decades of accumulated knowledge** in the field.

Effectively capture information from low signal-to-noise ratio cosmological observations
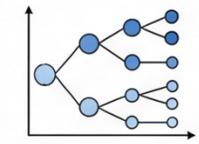
Stage 1: Domain Features (11 Categories)
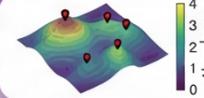
Spectral & Frequency Features

Power Spectrum
Second-order statistics, Gaussian modes, logarithmic multipole bins

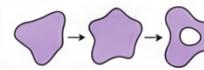**Multiscale Wavelet Statistics**
Amplitude, variation, higher-order stats at different scales
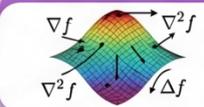
Morphological & Topological Features

Peak Counts
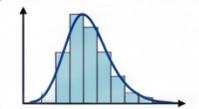Local maxima abundance, non-Gaussian information

Minkowski Functionals
V0, V1, V2 - area, perimeter, genus over thresholds

Gradient and Curvature Statistics
1st & 2nd derivatives, Laplacian, Hessian

Distribution & Density Features

1-point PDF
Density fluctuation distribution
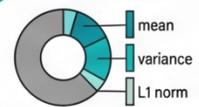
Aperture Mass Statistics
Mean, variance, L1 norm, skewness
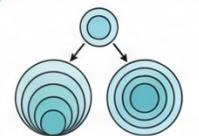
Multiscale Moments
Smoothed moments, edge statistics, multi-resolution

Statistical & Information Features

Lightweight Statistical Descriptors
Moments, autocorrelation, spectral slope

Entropy and Information-Theoretic Features
Fisher info, KL divergence, mutual info

Spatial Autocorrelation Features
Moran's I, Geary's C, spatial entropy

Domain Features

6

Stage 1: Performance of Each Domain Feature

Validation Performance Across Feature Categories
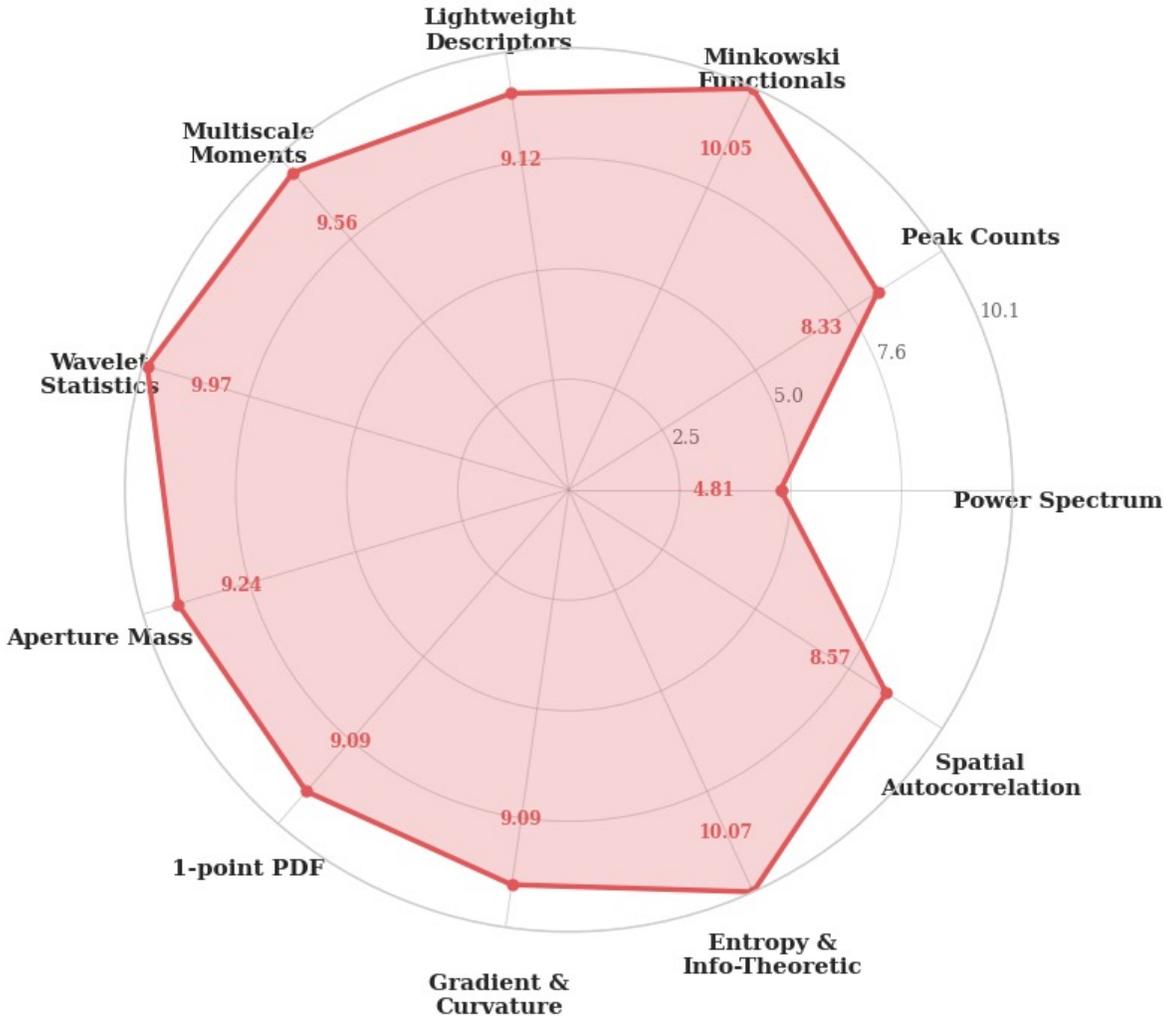
- Splits the public data for train and val as **8:2**.

- Linearly regress the coefficients with MLP.
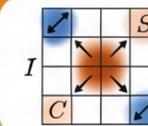
Feature Name	Val Score
Power Spectrum	4.81
Peak Counts	8.33
🏆 Minkowski Functionals	10.05
Lightweight Statistical Descriptors	9.12
Multiscale Moments	9.56
🏆 Multiscale Wavelet Statistics	9.97
Aperture Mass Statistics	9.24
1-point PDF	9.09
Gradient and Curvature Statistics	9.09
🏆 Entropy and Information-Theoretic Features	10.07
Spatial Autocorrelation Features	8.57

Stage 1: Complementarity Among Domain Features

Method	Test Score
CNN	9.30
+Power Spectrum	10.00
+Power Spectrum +Peak Counts	10.31
+Power Spectrum +Peak Counts +Minkowski Functionals	10.95
+Power Spectrum +Peak Counts +Minkowski Functionals +Lightweight Statistical Descriptors	11.08
+Power Spectrum +Peak Counts +Minkowski Functionals +Lightweight Statistical Descriptors +Multiscale Moments	11.24
+All Domain Features	11.22
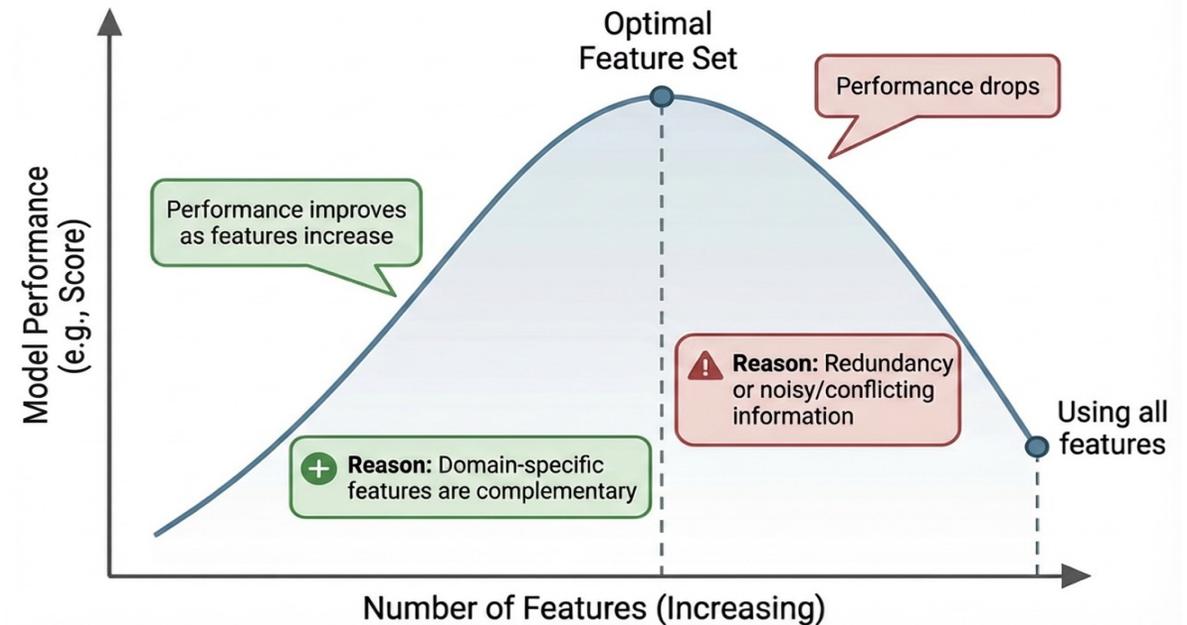
- There exists a **hidden complementarity** among the 11 extracted domain features.
- **But, Too Many Cooks Spoil the Broth.**

Stage 2: Integrating Deep and Domain features

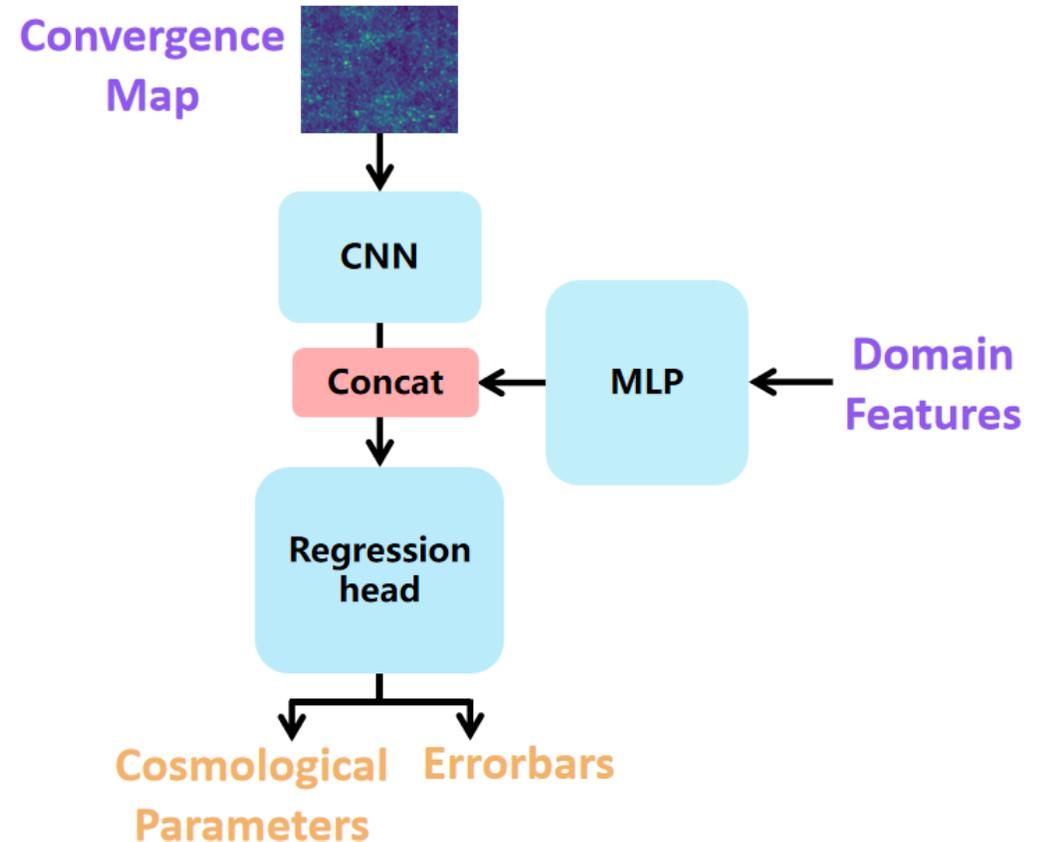Architecture 1: Only domain features
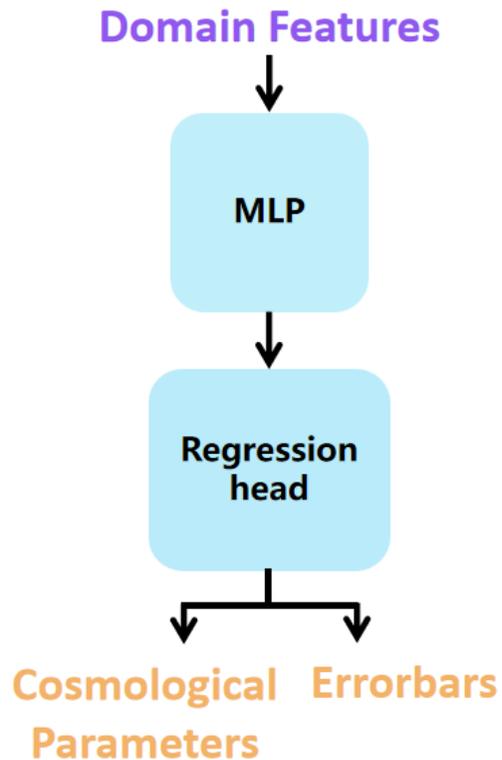
Architecture 2: Simple integration in the last layer

The above two simple architectures were explored. **Performance gains were limited.**

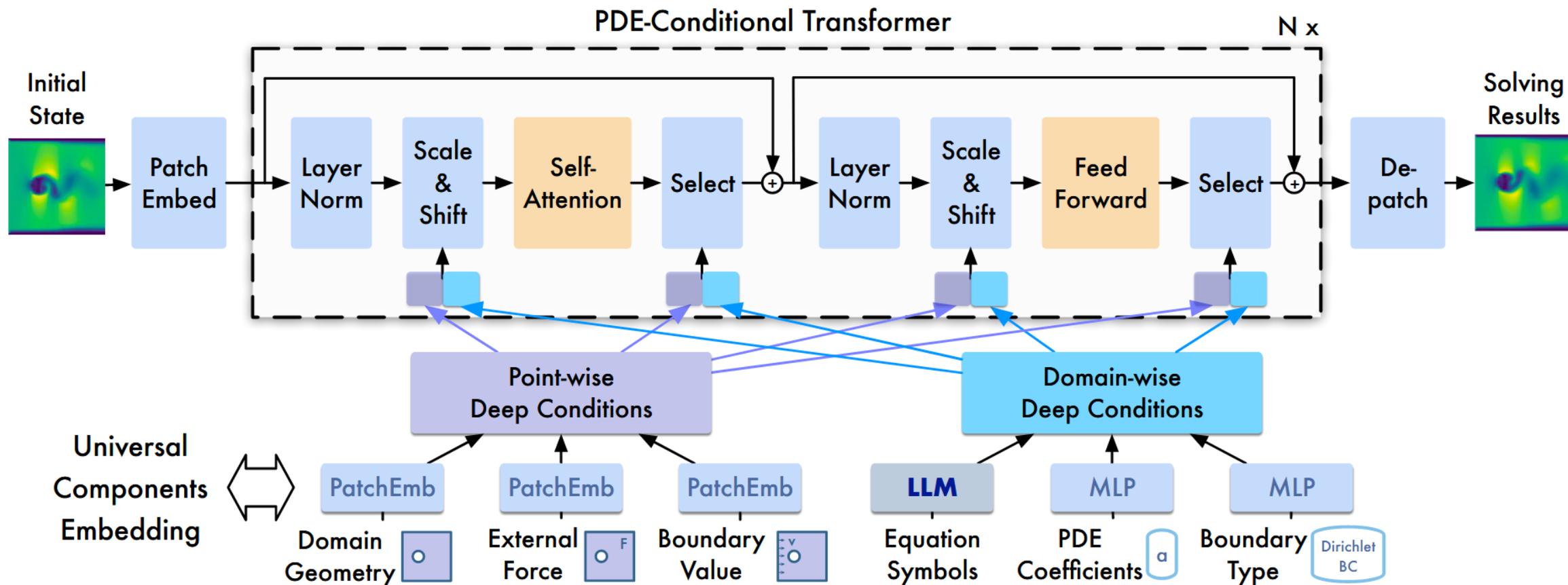# Stage 2: PDE-Conditional Transformer (Unisolver)

Figure 3. Overview of Unisolver. We universally embed all PDE components, such as domain geometries, equation symbols and boundary types into deep conditions and employ a conditional Transformer to aggregate deep conditions in the decoupled subspace.

Zhou et al., Unisolver: PDE-Conditional Transformers Towards Universal Neural PDE Solvers, ICML 2025

Stage 2: From PDE to Cosmology

Domain-Feature-Conditioned Unisolver

$$\text{AdaLN}(h \mid f) = \gamma(f) \cdot \frac{h - \mu(h)}{\sigma(h)} + \beta(f)$$
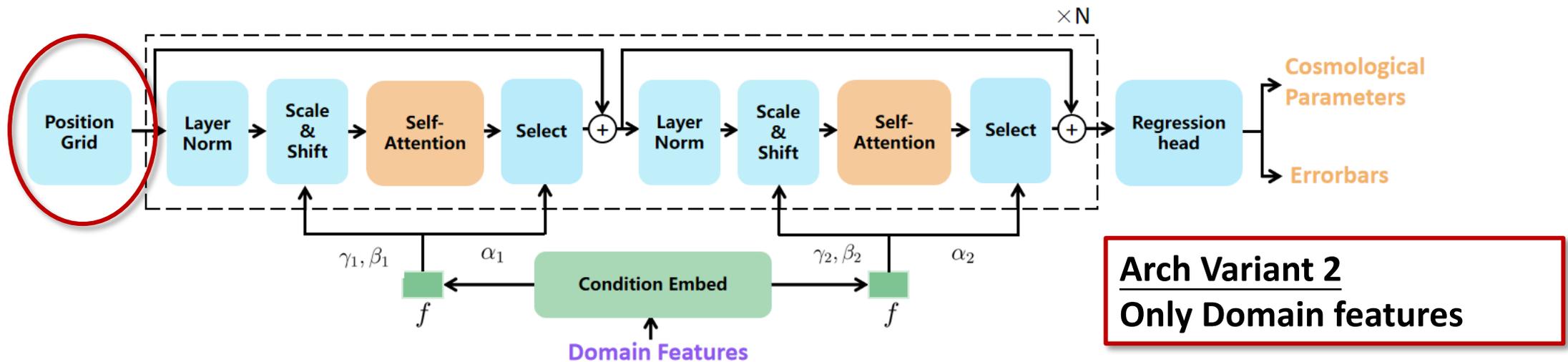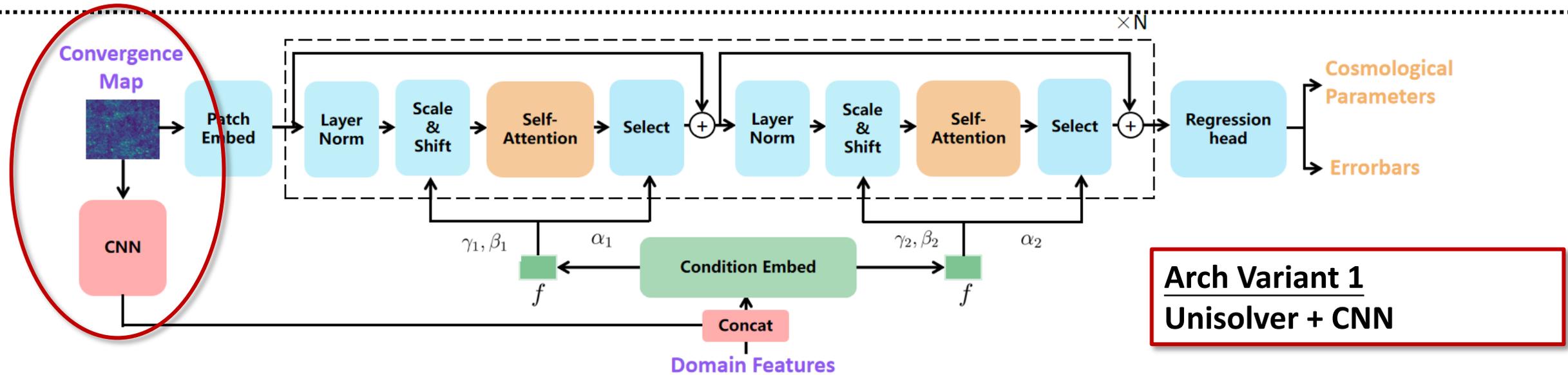
global physical context
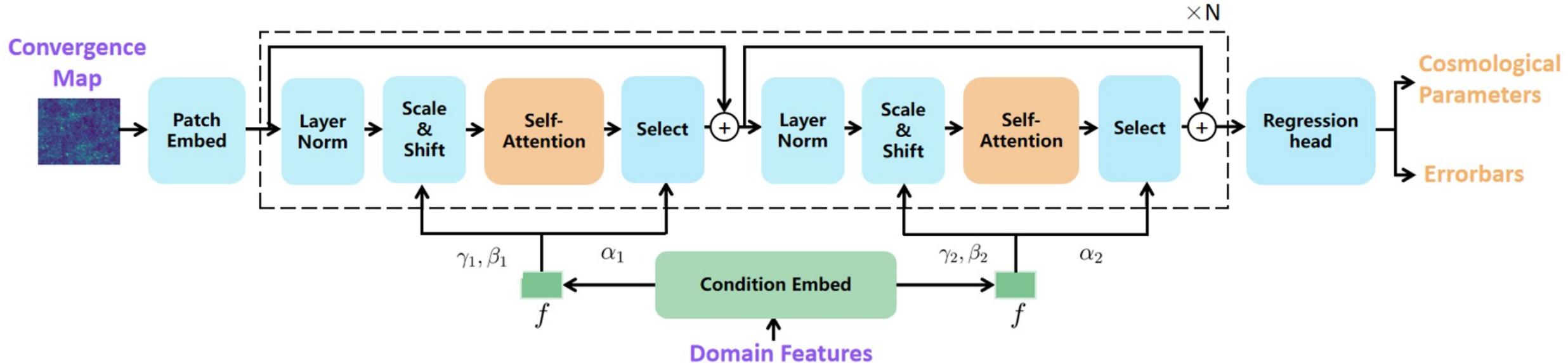
✓ From CNN-based to Transformer-based architecture; **Enhanced model capacity!**

✓ Adaptively rescale and shift internal representations based on **cosmological conditions**.

Stage 3: Ensemble Algorithm

Heterogeneity: (1) Domain Feature Combinations; (2) Architecture Variants.

Arch Variant 1
Unisolver + CNN

Arch Variant 2
Only Domain features

Stage 3: Greedy Model Selection Beyond All-in Ensemble
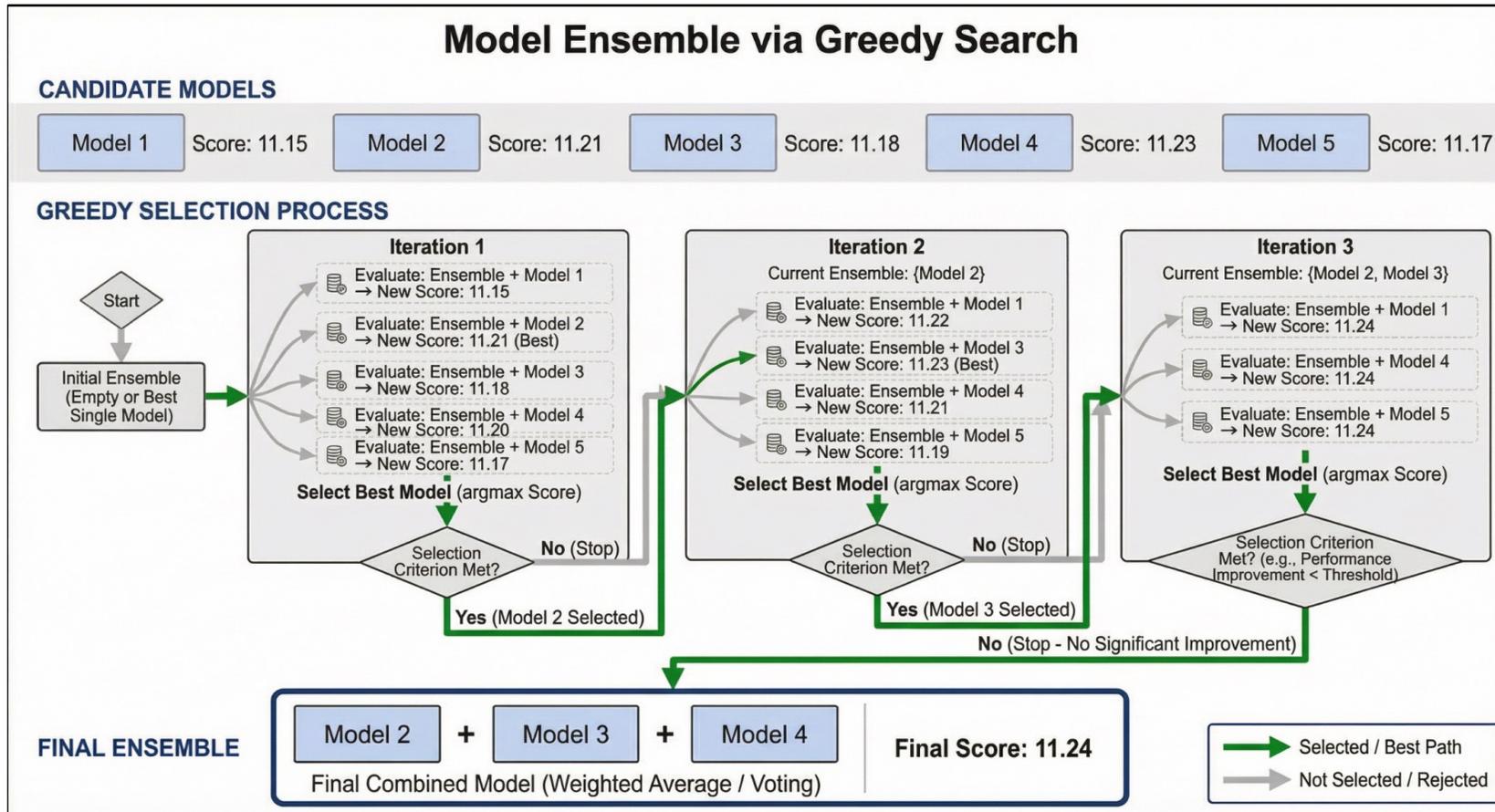
Huge ensemble Space

Combination of 11 Features

X

3 Architecture Variants

We need to reduce the search complexity

- (Search Order) Adds single models one by one **from high score to low score;**
- (Greedy Ensemble) Keep the newly added model only **if it can improve the performance.**

Stage 3: Results of the Greedy Search Ensemble

IDs listed under **Handcrafted Feature Combinations** correspond to the following feature groups:

(1) Power Spectrum,

(2) Peak Counts,

(3) Minkowski Functionals,

(4) Lightweight Statistical Descriptors,

(5) Multiscale Moments,

(6) Multiscale Wavelet Statistics,

(7) Aperture Mass Statistics,

(8) 1-point PDF,

(9) Gradient and Curvature Statistics,

(10) Entropy and Information-Theoretic Features,

(11) Spatial Autocorrelation Features.

IDs listed under **Model Architecture** correspond to the following model variants:

(1) Standard architecture - ViT with AdaLN,

(2) Standard architecture with a CNN front-end,

(3) Standard architecture without convergence map input.

| Model ID	Handcrafted Feature Combinations	Model Architecture
1	1,3,4,6,8	1
2	1,4,5,7,8	2
3	1,2,3,6,7	2
4	2,4,5,7,8	2
5	1,2,4,5,7	2
6	1,2,3,4,7	2
7	1,2,3,6,8	2
8	2,3,4,5,7	3
9	2,5,6,7,10	3
10	2,4,5,6,7,10	3
11	2,4,5,7,8	3
12	4,5,7,9,11	3
13	2,4,5,9,10	3
14	2,4,5,9,11	3
15	2,4,8,9,10	3
16	2,5,7,8,9	3

Test Score in the final holdout dataset: **11.2847**

Rank: **3**

Interesting Finding I —— Errorbar Distribution of Parameters
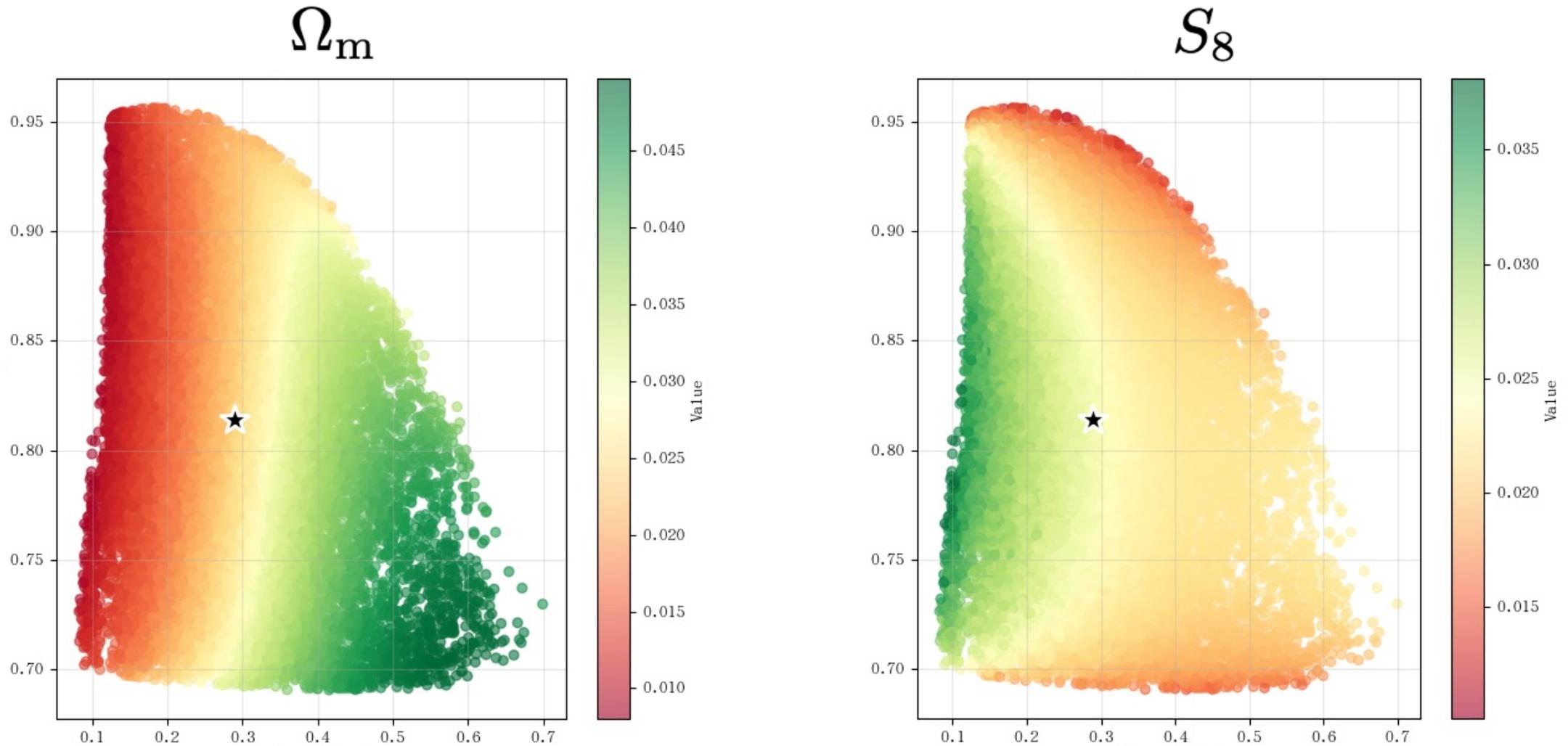
Errorbar of both parameters exhibits **strong dependence** on Ω_{m}, but **opposite trends**.

Interesting Finding II —— Score Distribution of Regions

Scatter: 2D label coordinates colored by score
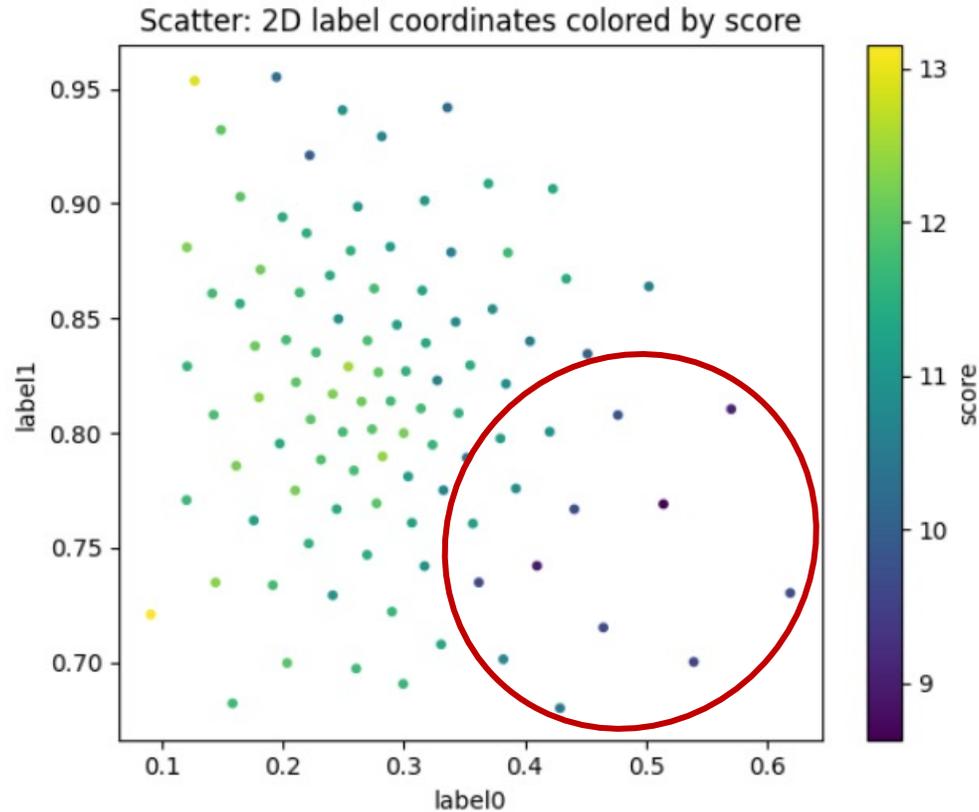
Samples located in the **bottom-right** region of the parameter space tend to exhibit lower scores.

Possible reasons: observation hardness / insufficient data collection.

Thank You!
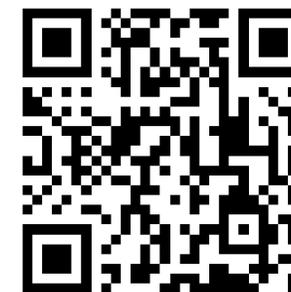
wenghk22@mails.Tsinghua.edu.cn / wuhaixu98@gmail.com
A new technical report is coming soon.

Code for Unisolver
https://github.com/thuml/Unisolver

Paper for Unisolver
https://openreview.net/pdf?id=r1ryQoI9iZ