# Cosmological Parameter Estimation via Parameter-Efficient DenseNet and Tunable Loss Function

## Winning Solution: Phase 1 of FAIR Universe Weak Lensing ML Uncertainty Challenge

Shubhojit Naskar

Yardi School of Artificial Intelligence
Indian Institute of Technology (IIT) Delhi

aib242285@iitd.ac.in

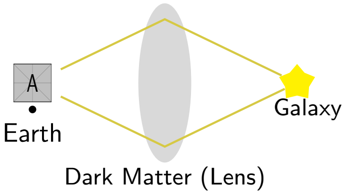
NeurIPS 2025 Workshop

Public Leaderboard Score: **11.61** (3rd Place)

# Cosmology 101: The Invisible Universe

**What is Weak Gravitational Lensing?**

- Most matter in the universe is **Dark Matter** (invisible).
- According to Einstein (General Relativity), gravity bends light.
- As light from distant galaxies travels to us, it passes through Dark Matter structures.
- **Result:** The galaxies appear slightly distorted (sheared).




We observe distorted shapes,

not the true galaxy position.

### The Goal

By analyzing these distortions (Convergence Maps), we can "weigh" the universe and map the invisible dark matter.

# The Machine Learning Task: Input to Output

We treat Cosmology as an Image Regression problem with Uncertainty.

**Input Map**
$1424 \times 176$ **px**

**(Simulated HSC Survey)**

**Represents Matter Density**

Pixel Data $\rightarrow$

**Deep Learning Model**
(DenseNet)

Predictions $\rightarrow$
$(\mu, \sigma)$

**Cosmological Parameters**

- $\boldsymbol{\Omega_m}$: Total Matter Density
  *(How heavy is the universe?)*
- $\boldsymbol{S_8}$: Fluctuation Amplitude
  *(How "clumpy" is it?)*

*Model must predict $\sigma$ (uncertainty) to handle simulation errors.

# The Core Challenge: Simulation-to-Real Gap

## The Problem

We train on **Simulations**, but test on data that mimics **Reality**.

**Sources of Mismatch (Systematics):**

1. **Baryonic Physics:** Gas interactions, supernovae feedback.
2. **Redshift Uncertainty:** Errors in measuring galaxy distances.
3. **Intrinsic Alignment:** Galaxies physically aligning with each other.

## The Data

- **Training Set:** $\sim 26,000$ maps (101 cosmologies).
- **Requirement:** The model must implicitly marginalize over these nuisance parameters.

# The Journey I: From Statistics to Deep Learning

My approach evolved through rigorous experimentation, moving from statistical baselines to advanced computer vision architectures.

- **Phase 1: Statistical Baseline (Score: 4.54)**
  Used Power Spectrum (1D summary statistic) + MCMC Emulator. *Limitation: Discarded non-Gaussian information in the maps.*

**Simple CNN** (Score: 8.46) 5-Fold Ensemble → **Adaptive Pooling** (Score: 9.02) Reduced Parameters → **ResNet Backbone** (Score: 9.82) Residual Connections

- **Key Insight:** Residual connections allowed the model to learn deeper features from the complex cosmic web structure.

# The Journey II: Optimization & Final Solution

Once the architecture was stable, the focus shifted to **Loss Engineering** and maximizing **Parameter Efficiency**.

## 1. Loss Engineering (Score: 11.48)

**Problem:** Standard Likelihood loss was unstable.
**Solution:** Decomposed loss into three tunable terms (Uncertainty Penalty + Weighted MSE + Pure MSE).

**Impact:** Gave precise control over the model's "laziness" (tendency to predict infinite uncertainty).

## 2. Final: DenseNet (Score: 11.61)

**Move:** ResNet → DenseNet.

**Why?** Feature concatenation preserves low-level lensing signals better than summation.
**Efficiency:** 766k → **472k parameters**.
**Result:** Superior generalization.

# Data Pipeline: Bridging Simulation and Reality

## 1. Efficient Preprocessing

- **Memory Mapping:** Converted 25,856 maps to `.mmap` files with `float16` precision.
- **Impact:** Allowed rapid, out-of-core data loading without crashing RAM.
- **Normalization:** Statistics (mean/std) computed on a random 5k subset to prevent overflow.

## 2. Geometric Augmentation

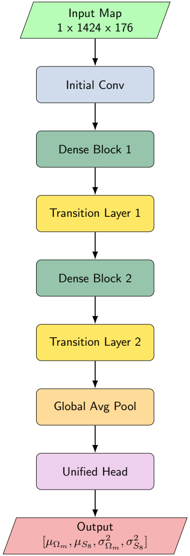
- Random Horizontal/Vertical Flips applied to every sample.
- **Why?** The Universe is isotropic; orientation should not affect cosmological parameters.

## 3. Physics-Based Noise Injection (Crucial Step)

**The Challenge:** Training images are *noiseless*, but Test images emulate a real telescope (noisy).
**The Solution:** We inject Gaussian noise derived from the survey's instrument parameters:

$$n_g = 30.0 \text{ gal/arcmin}^2 \quad | \quad \text{Pixel Scale} = 2.0'$$

# Architecture: Parameter-Efficient DenseNet




**Architecture Details:**

- **Input:** Single channel $(1424 \times 176)$.
- **Dense Blocks:** Layers connect to all subsequent layers.
- **Bottleneck:** Adaptive Global Average Pooling reduces spatial dims to $1 \times 1$.
- **Output Head:** A single MLP predicts both means and log-variances.

**Initialization Trick:** The output bias for $\log(\sigma^2)$ was initialized to $-5.0$. This prevents the loss from exploding at the start of training.

# Core Innovation: Tunable Three-Term Loss

Standard NLL is often unstable. I used a decomposed loss to balance Accuracy vs. Uncertainty:

$$\mathcal{L} = \underbrace{\lambda_{unc} \sum \log(\sigma^2)}_{\textbf{1. Uncertainty Penalty}} + \underbrace{\lambda_{wMSE} \sum \frac{(y-\mu)^2}{\sigma^2}}_{\textbf{2. Weighted MSE}} + \underbrace{\lambda_{MSE} \sum (y-\mu)^2}_{\textbf{3. Pure MSE (Anchor)}}$$

**The Logic:**

1. **Uncertainty Penalty:** Explicitly penalizes high uncertainty. Prevents the model from "cheating" by saying "I don't know" ($\sigma \to \infty$) to everything.
2. **Weighted MSE:** The standard probabilistic error term.
3. **Pure MSE Anchor:** Ensures the mean prediction $\mu$ remains accurate even if the uncertainty estimation is fluctuating.

**Optimal Tuning:** $\lambda_{unc} = 2.0, \lambda_{wMSE} = 1.0, \lambda_{MSE} = 0.0$

# Results & Ablation Study

| Method | Parameters | Public Score |
|---|---|---|
| Power Spectrum (Baseline) | N/A | 4.54 |
| CNN Ensemble | 1.2M | 8.46 |
| ResNet Backbone | 766k | 11.48 |
| **DenseNet (Ours)** | **472k** | **11.61** |

## Key Findings

1. **Efficiency > Complexity:** The smaller DenseNet generalized better.
2. **Loss Decoupling:** Controlling the $\log(\sigma^2)$ weight was more effective than architectural changes for calibration.
3. **Physics Wins:** Models trained with physics-based augmentation outperformed deeper models without it.

# Scientific Impact: Towards LSST & Euclid

Why is this important for Physics?

**1. Implicit Marginalization**

- Physicists usually have to mathematically integrate over nuisance parameters.
- Our DenseNet *learns* to be invariant to systematics (baryons, IA) automatically.

**2. Trustworthy AI**

- Future surveys (Rubin/LSST) will have massive data volume.
- We need models that provide **calibrated uncertainty**.

**The Takeaway**

This approach demonstrates that we can use imperfect simulations to extract precise cosmological parameters from real observations, provided the uncertainty is modeled correctly.

# Conclusion

**Summary of Contribution:**

1. **Compact Architecture:** DenseNet (472k params) prevents overfitting to simulation quirks.
2. **Physics-Informed Augmentation:** Injected known Instrument Noise into clean training data to match Test conditions.
3. **Tunable Loss:** Allows manual calibration of the Precision-Uncertainty trade-off.

**Thank You! Questions?**