HIGGS PRODUCTION CLASSIFIER USING WEAK SUPERVISION

Kai-Feng Chen, Yi-An Chen, Cheng-Wei Chiang, Feng-Yang Hsieh

Presented by Feng-Yang Hsieh 謝豐仰 2026/01/20

arXiv:2511.18726 [hep-ph]

OUTLINE

- Introduction: Physics motivation
- Method: Weak supervision, Neural Network
- Proposed strategies: Data augmentation, Transfer learning
- Results
- Summary

INTRODUCTION

PHYSICS MOTIVATION: VBF vs. GGF

- Since the discovery of the Higgs boson in 2012, much effort has been devoted to precision measurements of Higgs couplings and properties
- Production modes:
  - GGF (Gluon–gluon fusion): Dominant cross-section in Standard Model (SM)
  - VBF (Vector boson fusion): Subdominant but crucial. It probes Electroweak Symmetry Breaking and New Physics (BSM) effects
- Accurately separating VBF from the GGF is essential but difficult

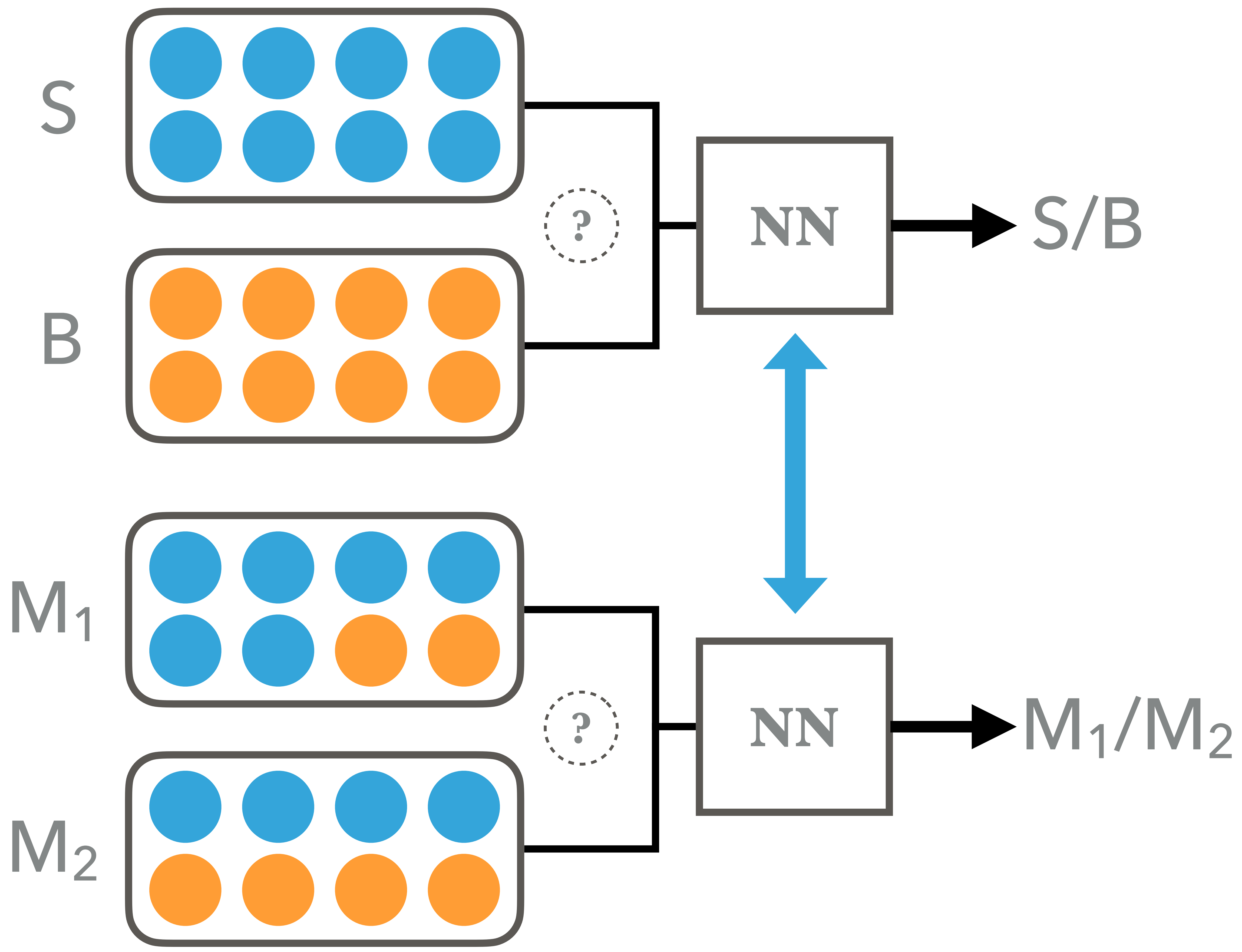
VBF/GGF CLASSIFICATION

- Traditional methods rely on sequential kinematic cuts or BDTs
- Deep neural networks (DNNs) demonstrate superior performance in classification and extracting features from low-level inputs
- In NN training, fully supervised learning requires event-level truth labels. However, they are unavailable in real collider data → Simulations
- Relying solely on Monte Carlo (MC) simulations introduces model dependencies and systematic uncertainties
    - Can we train a classifier on the dataset without truth labels?

METHOD

CWOLA FRAMEWORK

- Truth signal/background event labels are inaccessible in real collider data
- Classification Without Labels (CWoLa) operates on two mixed datasets with different signal fractions
- Optimal classifier for mixed samples is proven to be optimal for distinguishing pure Signal and Background

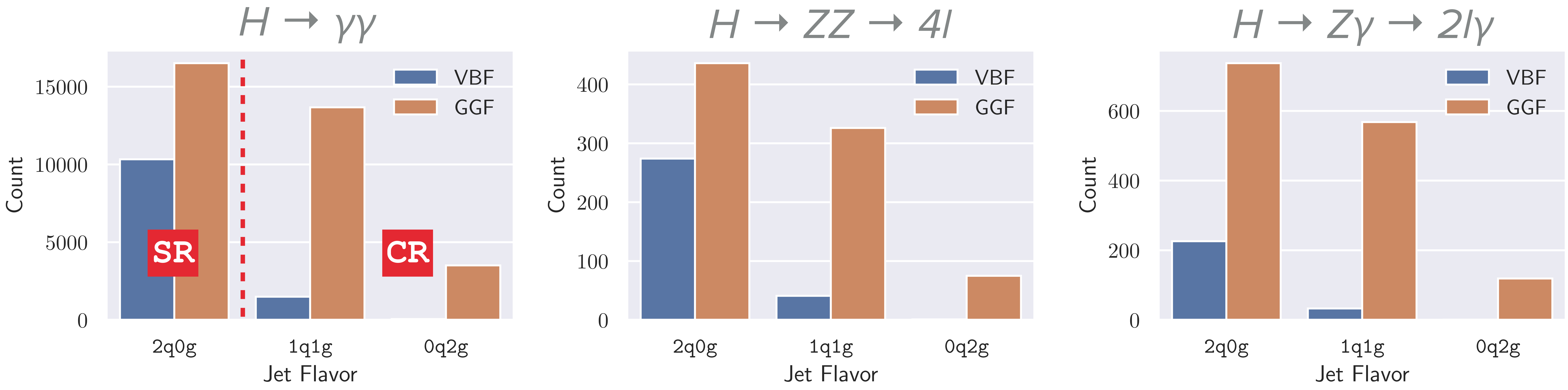


PHYSICS IMPLEMENTATION: SR & CR

- Target: Distinguish VBF (Signal) from GGF (Background)
- We define regions based on the parton flavor of the two leading jets. Then, use these regions to construct two mixed datasets
  - Signal Region (SR): Both jets originate from quarks (2q0g)
  - Control Region (CR): Contains at least one gluon jet (1q1g/0q2g)

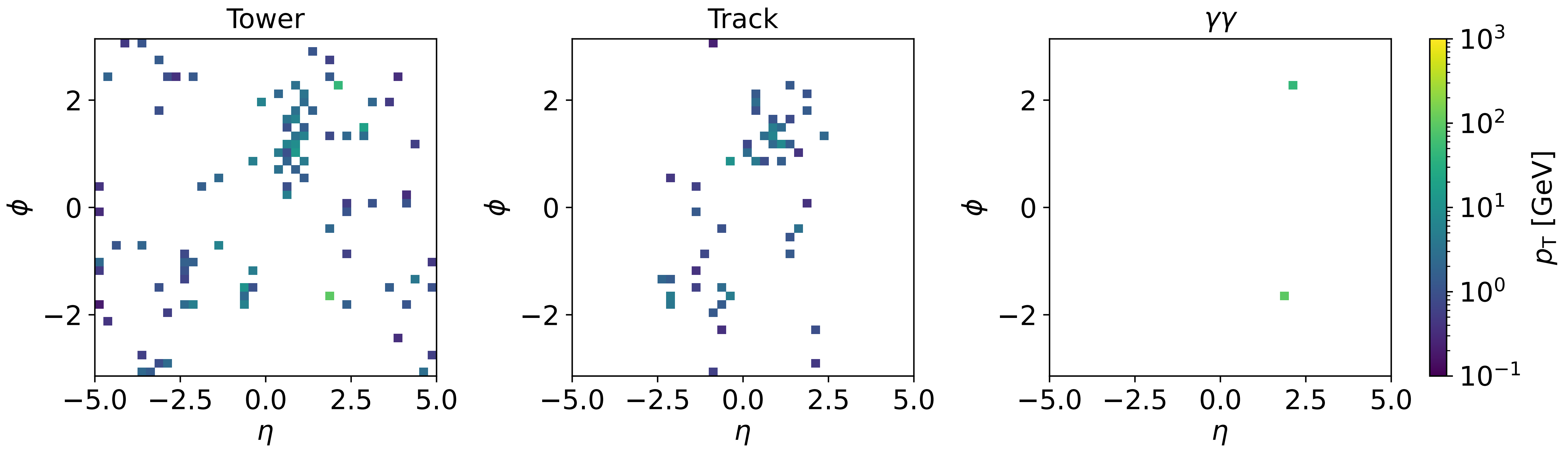
MONTE CARLO SAMPLES

- VBF/GGF Higgs boson production are simulated for the LHC at a center-of-mass energy of $\sqrt{s}$ = 14 TeV with different decay channels



NETWORK ARCHITECTURES: CNN vs. TRANSFORMER

- Convolutional Neural Network (CNN)
  - Based on the Event-CNN
  - Input: Event image
  - Captures local spatial correlations via convolutional layers



- Transformer
  - Based on the Particle Transformer
  - Input: (p_T, η, φ, type)
  - Capture global dependencies via Attention blocks

PROPOSED STRATEGIES

DATA AUGMENTATION: φ-SHIFTING

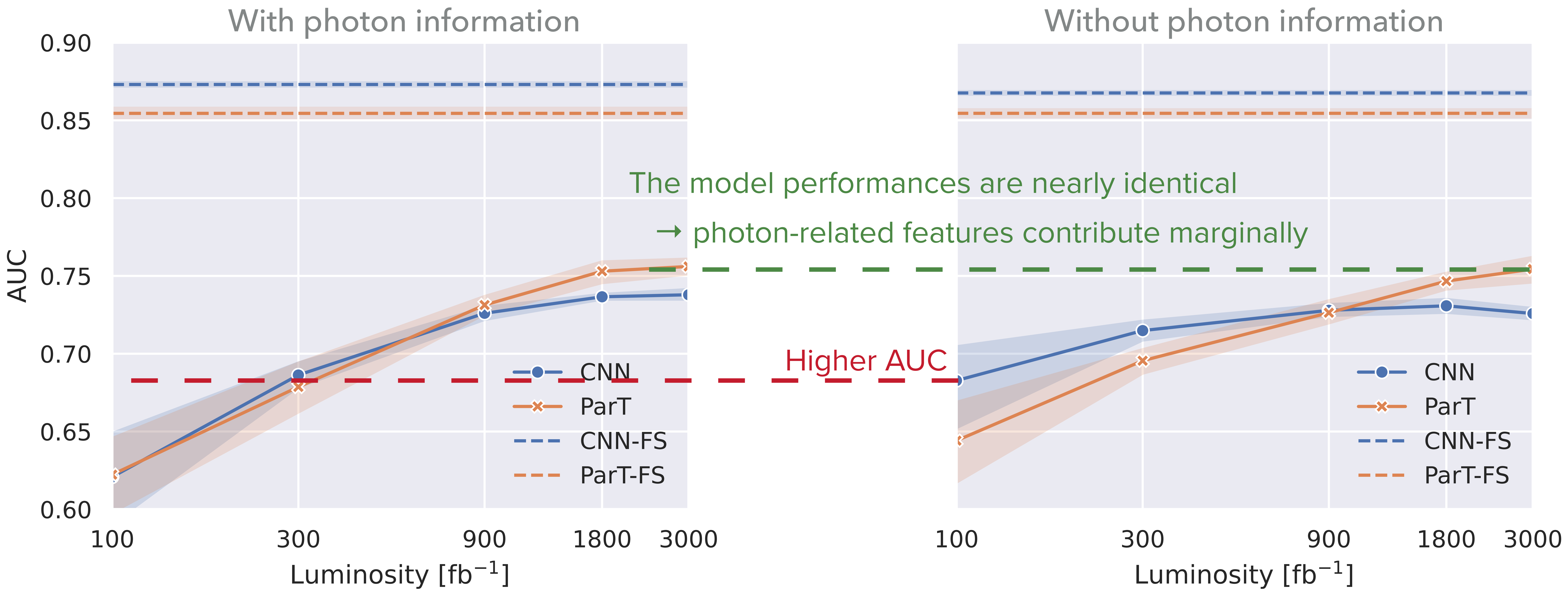
- LHC detectors exhibit cylindrical symmetry along the beam axis
- Rotate all event constituents by a random angle θ in $[-\pi, \pi]$
- φ-Shifting: $\varphi_i \rightarrow \varphi_i + \theta$
- Generates statistically independent samples without altering the underlying kinematics or topology

DECAY-MODE TRANSFERABILITY

- Proposed Workflow:
  - Train the model on the $H \to \gamma\gamma$ dataset (masking photon information)
  - Transfer the trained model directly to $H \to ZZ \to 4l$ and $H \to Z\gamma \to 2l\gamma$ events (masking decay products)
- The VBF/GGF production mechanisms (initial state radiation, forward jets) are largely factorized from the Higgs decay mode
- Hadronic activity alone provides sufficient discrimination power

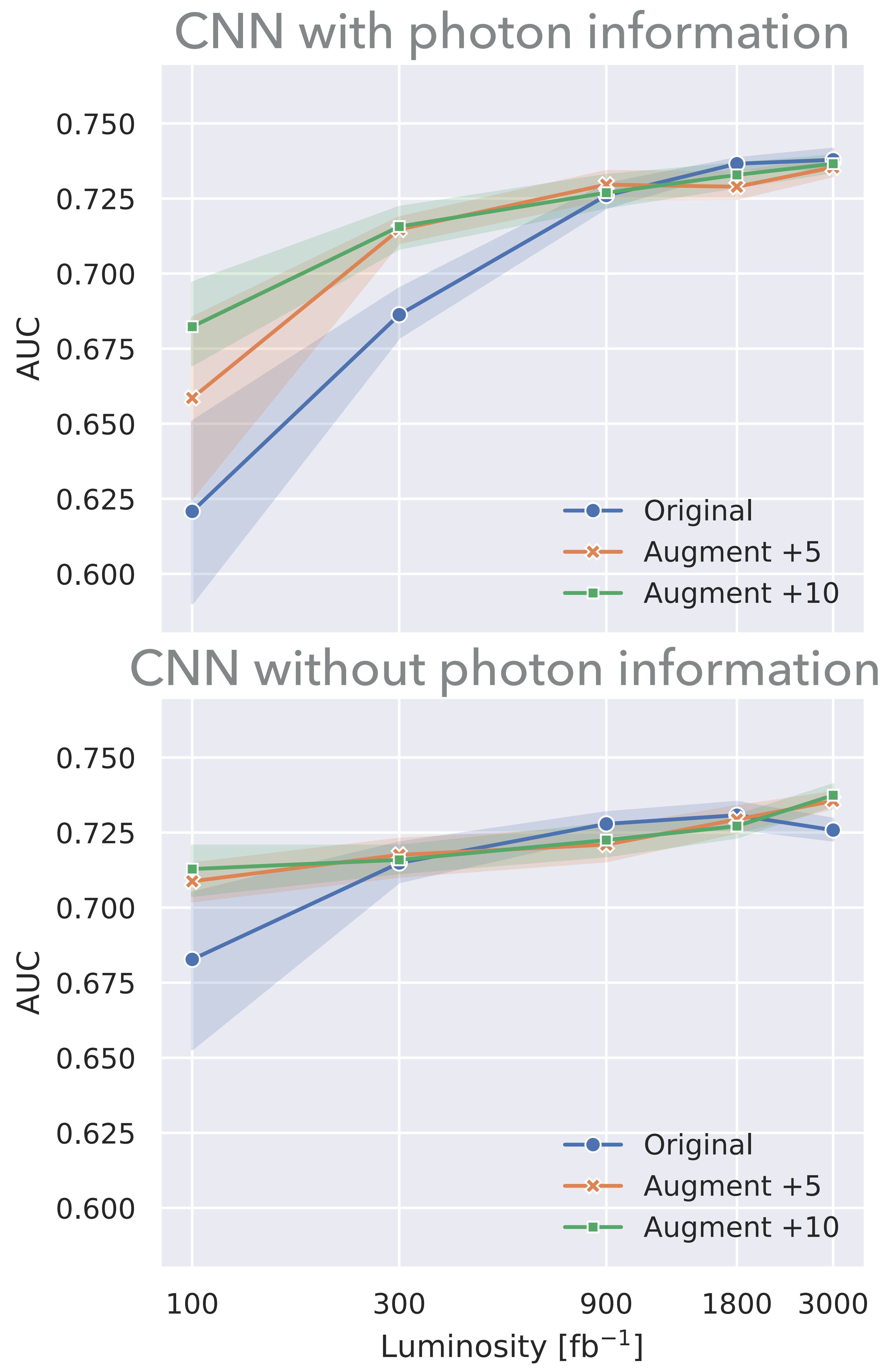
RESULTS

$H \to \gamma\gamma$ PERFORMANCE



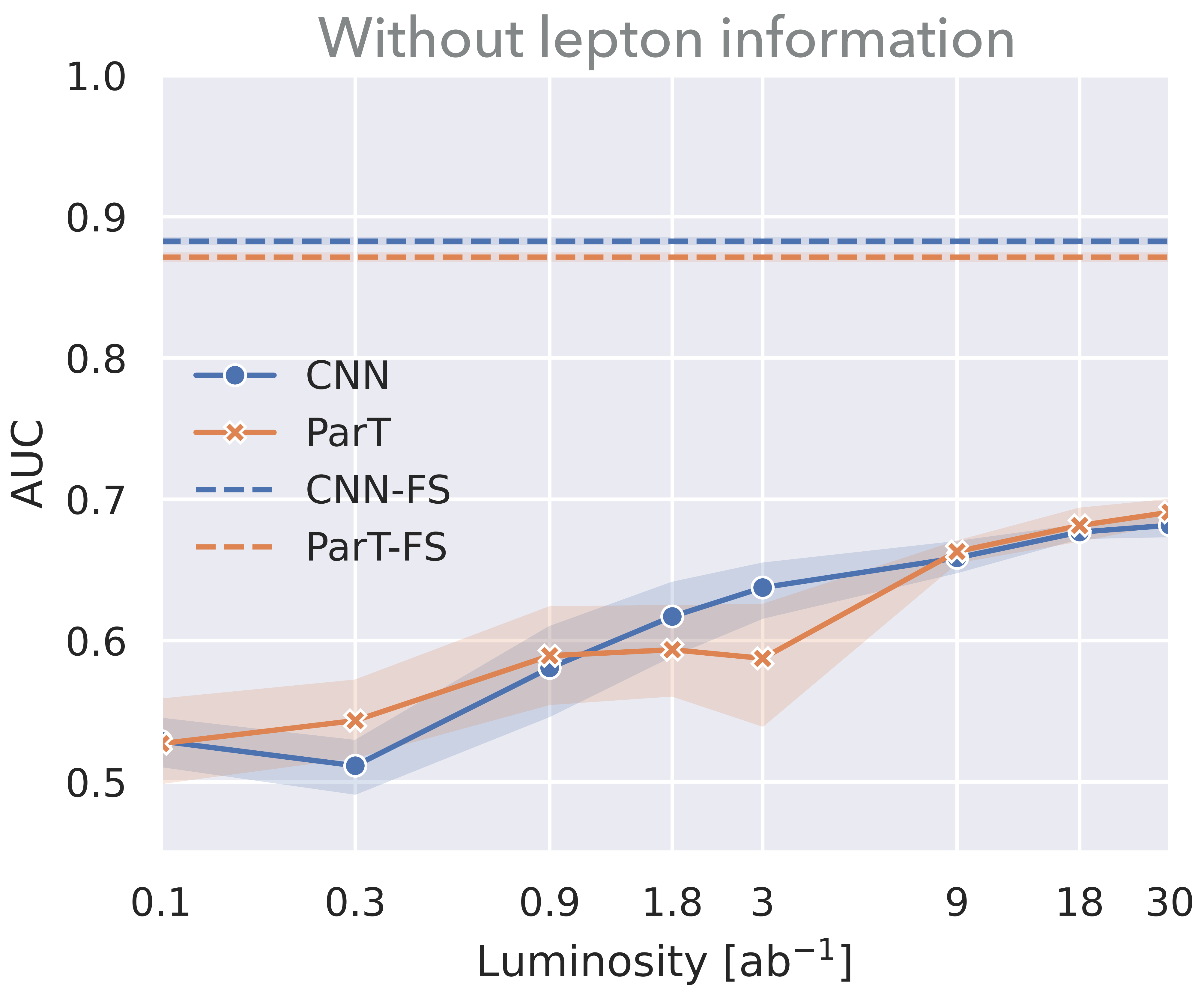
EFFECT OF DATA AUGMENTATION

- φ-shifting augmentation improves both the mean AUC and training stability in the low-luminosity regime
- Beyond $L \approx 900$ fb^{-1}, the improvement saturates
- φ-shifting augmentation provides a statistically efficient and physically consistent means of improving weakly supervised training

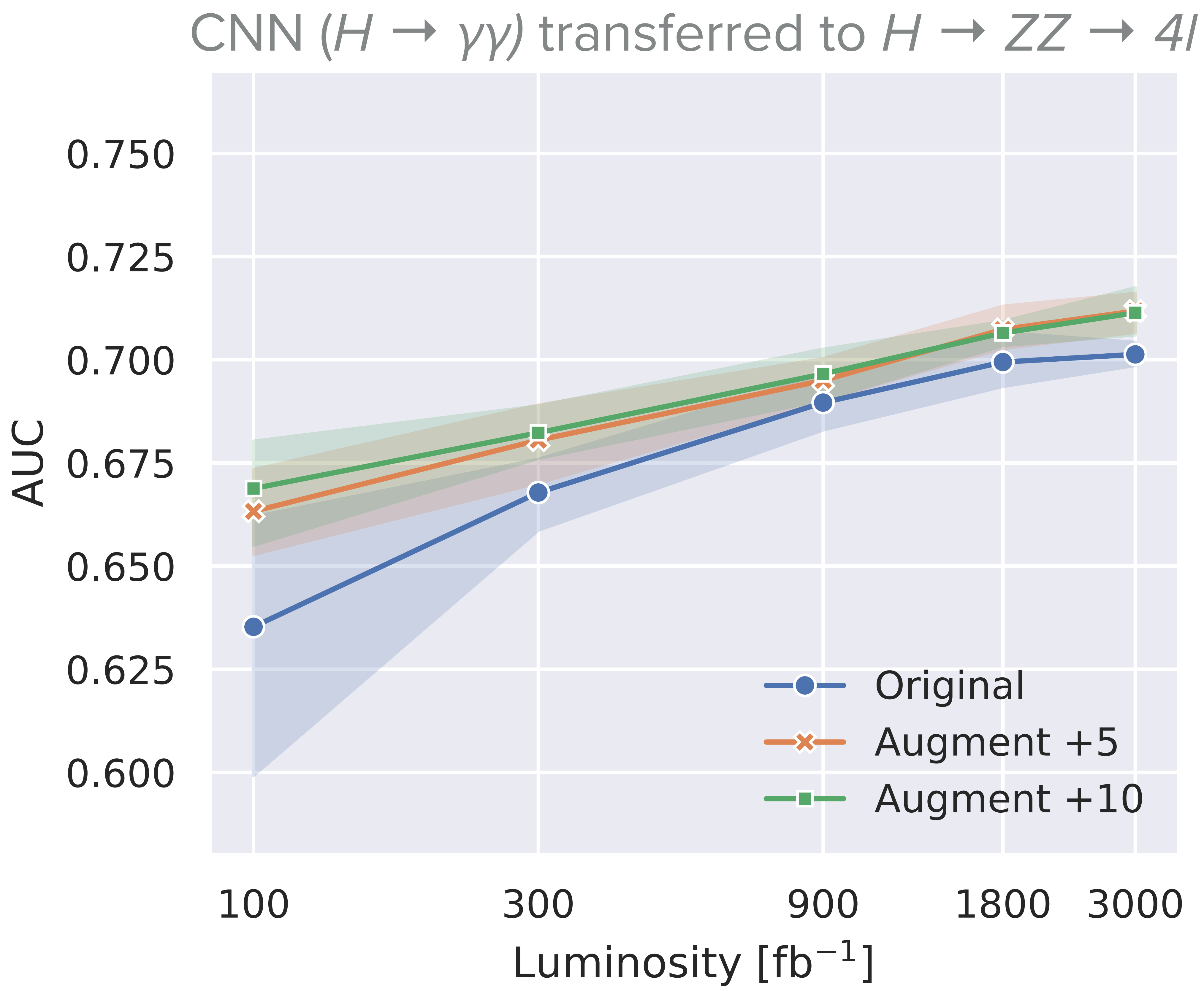


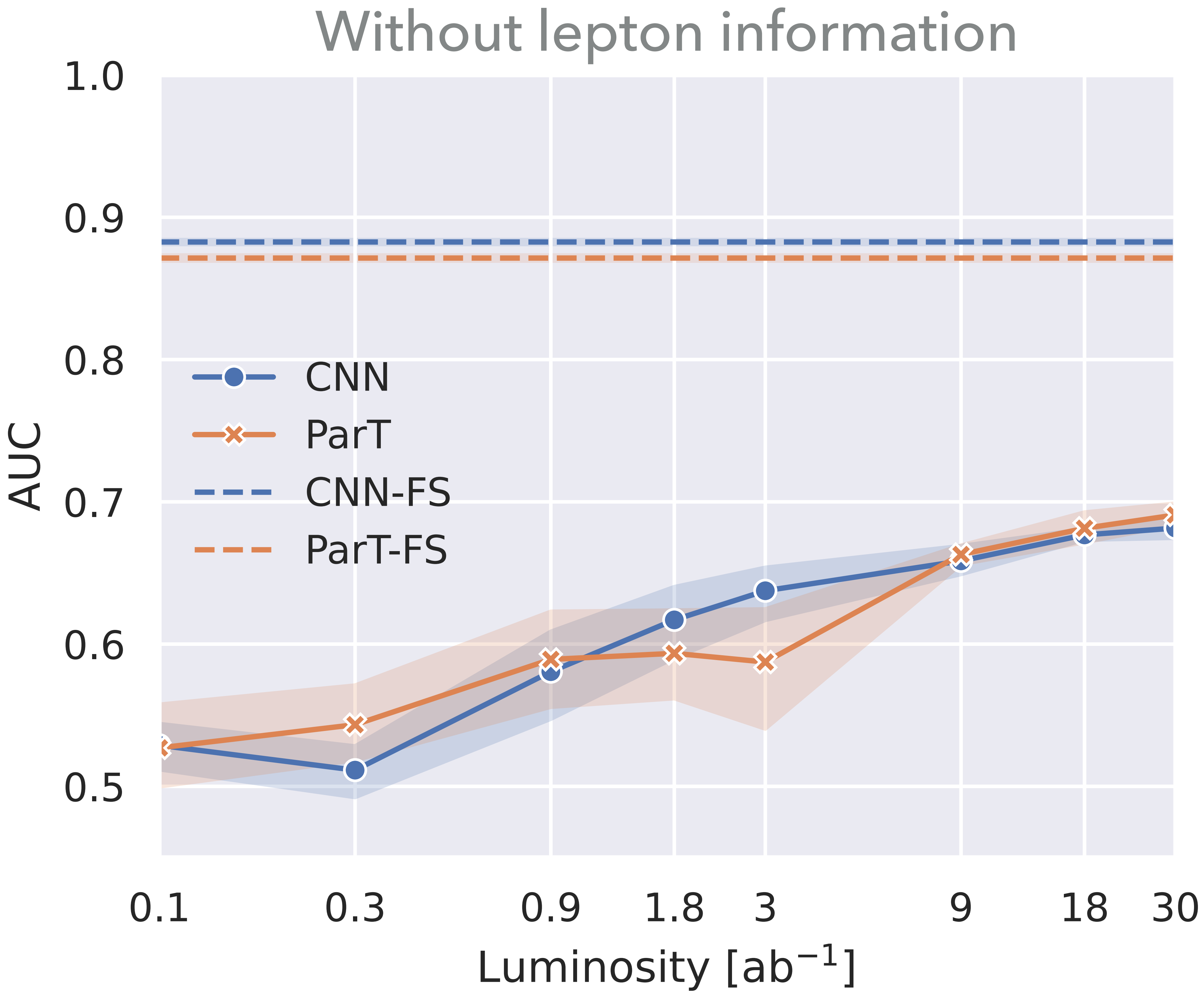
$H \to ZZ \to 4l$: A DIFFICULT CHANNEL

- The model performance improves gradually with increasing training luminosity
- Even at L = 3000 fb^{-1}, the achieved AUC values remain modest, reflecting the severe data scarcity
- The luminosity levels required to reach stable training are well beyond those achievable in realistic experimental conditions

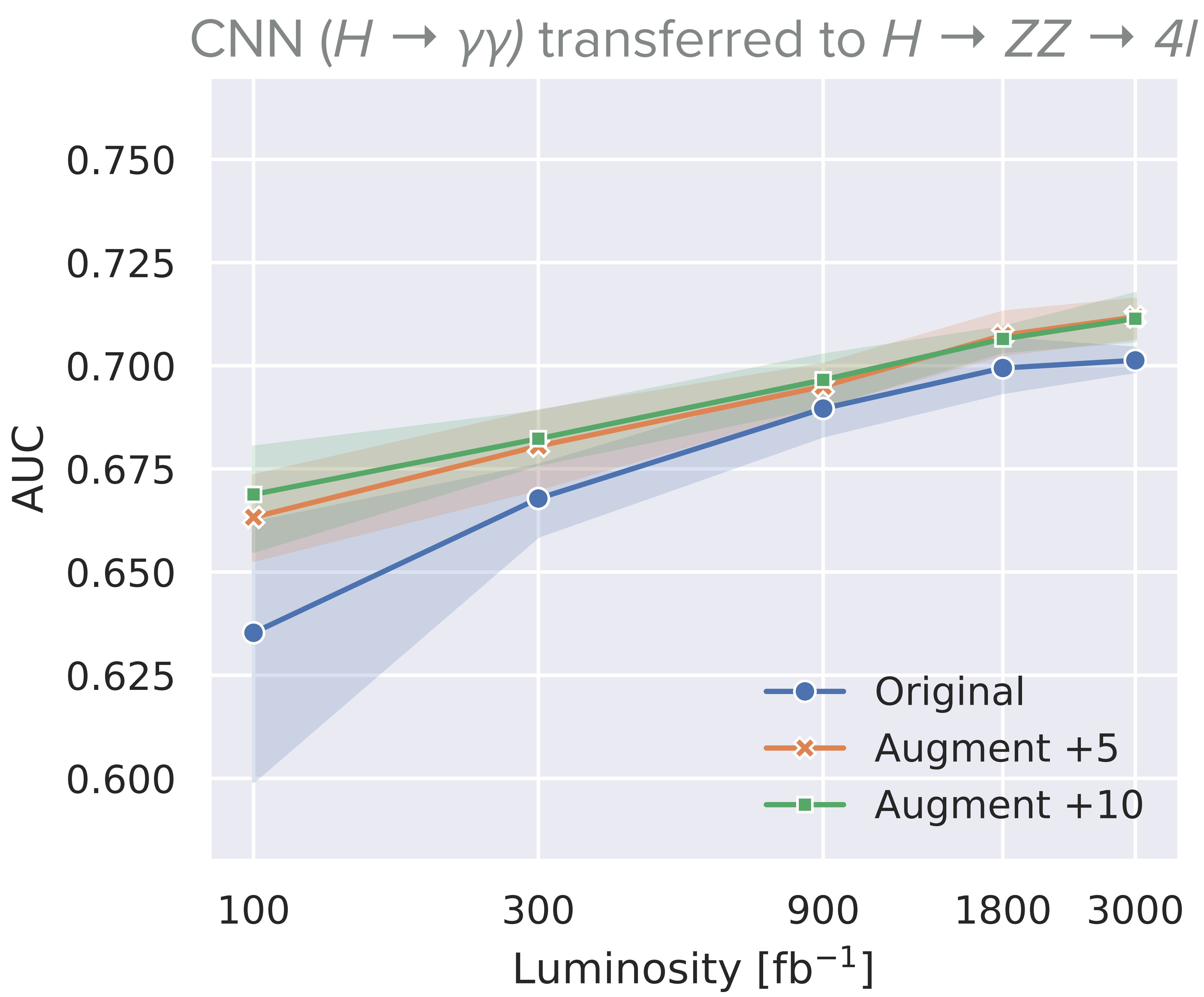


TRANSFERABILITY TEST



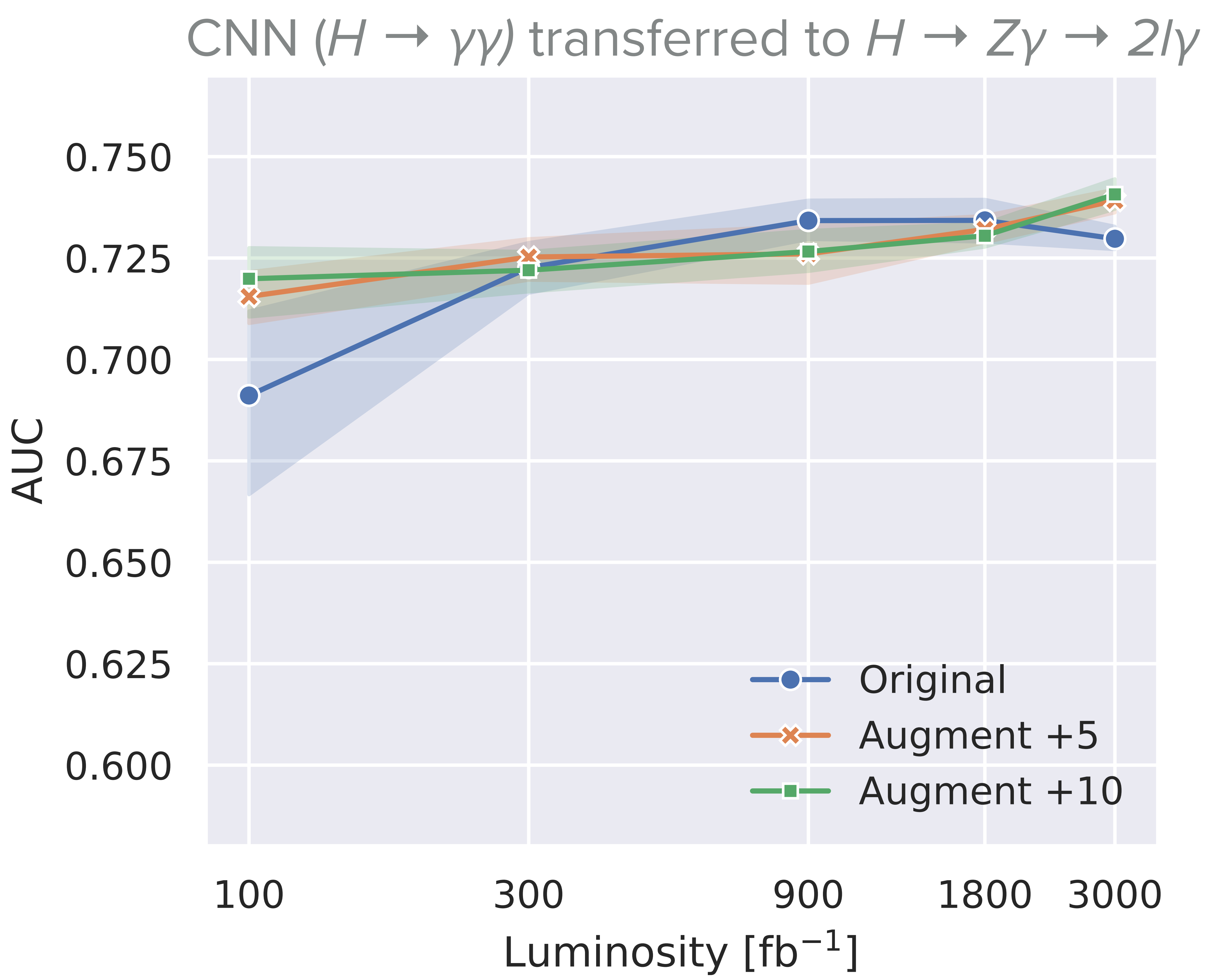


TRANSFERABILITY TEST



- Transferred performance at L = 100 fb^{-1} corresponds to L = 9000 — 18000 fb^{-1} for training directly on the $H \to ZZ \to 4l$ data
- The transfer model retains strong discrimination power and stability, highlighting a promising level of generalization

TRANSFERABILITY TEST



- NNs trained on $H \to \gamma\gamma$ datasets are applied to $H \to Z\gamma \to 2l\gamma$
- NNs achieve slightly higher AUCs than those obtained on the *ZZ* transfer tests, and their fluctuations across are smaller
- Decay-agnostic classifiers can be trained in one Higgs decay channel and successfully applied to others

SUMMARY

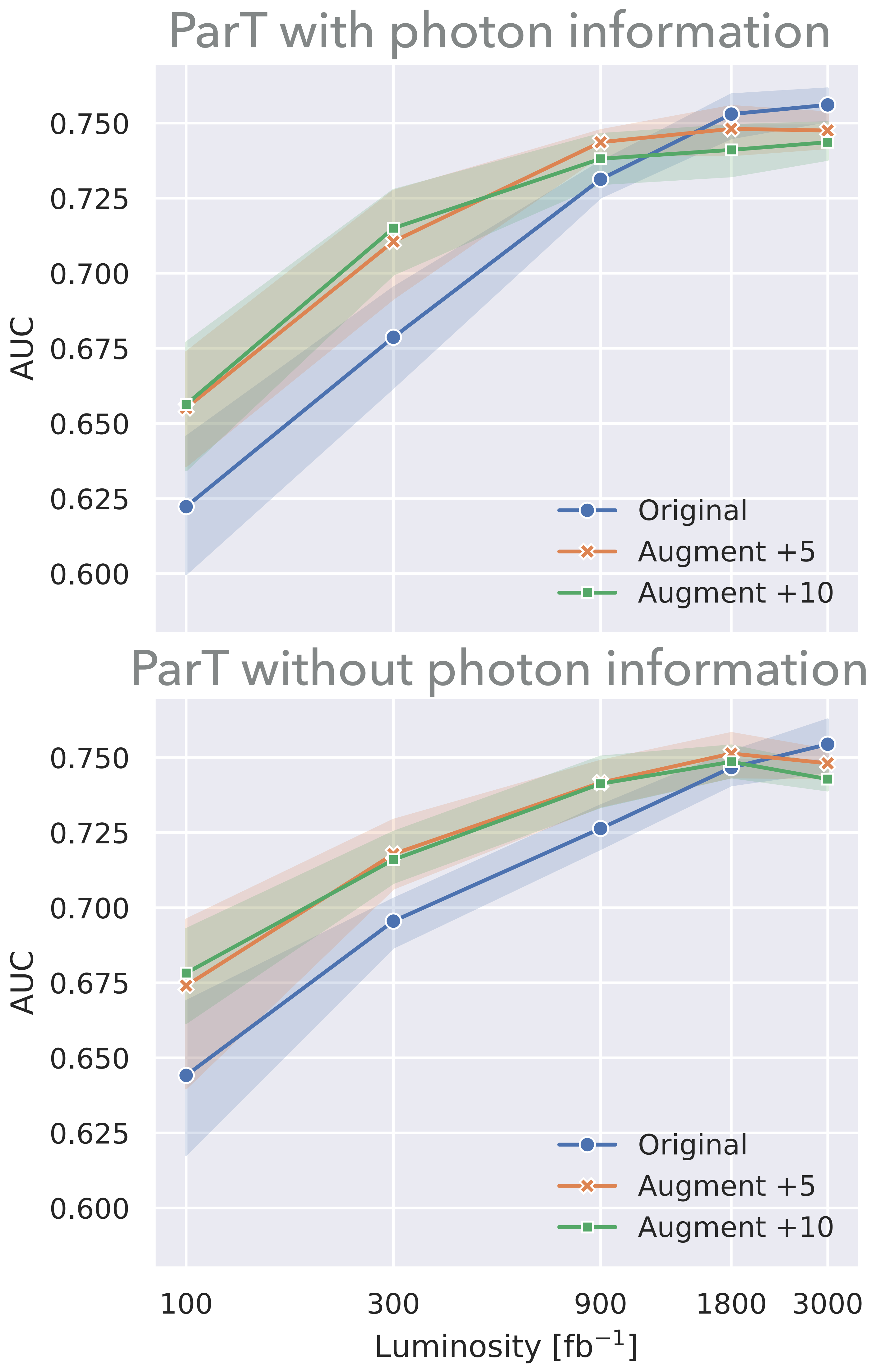
SUMMARY

- Successfully applied the CWoLa framework to distinguish VBF/GGF production mechanisms using only mixed data without explicit labels
- For the $H \to \gamma\gamma$ channel, the CWoLa captures the differences between VBF and GGF processes and demonstrates that hadronic activity is sufficient for discrimination
- Augmentation: φ-shifting improves AUC and stability in low-luminosity regimes
- Transfer Learning: Models pre-trained on the abundant $H \to \gamma\gamma$ channel generalize successfully to rare $H \to ZZ \to 4l$ and $H \to Z\gamma \to 2l\gamma$ channels
- Establishes a practical, data-efficient strategy for real LHC analyses

BACKUP

EFFECT OF DATA AUGMENTATION

- For ParT models, similar trends are observed: enhances performance up to $L \approx 900\ \mathrm{fb}^{-1}$
- Fivefold replication is already sufficient. Tenfold offers no further gain
- The φ-shifting augmentation provides a statistically efficient and physically consistent means of improving weakly supervised training



$H \to ZZ \to 4l$: A DIFFICULT CHANNEL

- Even with the application of data augmentation, the performance gain remains marginal
- These results indicate that direct weakly supervised training on $H \to ZZ \to 4l$ is constrained by statistical limitations