2026 JULY 19–22 VANCOUVER, BC
JOINT AAPM | COMP MEETING

# A Controlled Evaluation of Architectural Enhancements to 3D U-Net for Automated Radiotherapy Dose Prediction


T Kunkyab, PhD[1]; Y Lei, PhD[1]; H Guo, PhD[1]; J Yang, PhD[1]; J Xia, PhD[1]; J Zhang, PhD[1]; T Liu, PhD[1]; M Chao, PhD[1];

[1]Department of Radiation Oncology, The Mount Sinai Hospital



## ABSTRACT

Accurate 3D dose prediction depends heavily on a model's capacity to internalize complex patient anatomy, a trait often engineered through deep, multi-stage architectures. This work introduces a unified, single-network paradigm: a dual-head 3D U-Net that jointly optimizes clinical dose prediction and anatomical reconstruction under a single, synergistic learning objective. To force the shared encoder to extract highly generalized structural representations, we introduce a structure-aware biased masking strategy targeting critical tissue interfaces and an asymmetric skip-connection topology that selectively eliminates skip pathways for the reconstruction decoder. At a budget-matched 80k-iteration threshold, embedding this anatomical self-supervision task outperforms traditional training from scratch, yielding a 0.14 Gy reduction in combined error driven by substantial gains in both spatial dose (-0.10 Gy) and dose-volume histogram (DVH) accuracy (-0.18 Gy). To demonstrate its top-tier clinical competitiveness, the framework was benchmarked on the OpenKBP dataset against the challenge-winning Cascaded 3D U-Net (C3D). With test-time augmentation, the proposed streamlined model achieves a combined score of 2.187 Gy, establishing statistical parity with C3D's 2.134 Gy while eliminating its complex two-network cascade and knowledge distillation pipeline. These findings demonstrate that joint structural co-learning provides a highly efficient, high-performance alternative to complex, multi-stage architectures.



## CONTACT

Tenzin Kunkyab
The Mount Sinai Hospital
Department of Radiation Oncology
Tenzin.Kunkyab@mountsinai.org


## INTRODUCTION

Automated 3D dose prediction accelerates radiotherapy planning, but standard architectures trained from scratch often struggle to inherently internalize complex, patient-specific anatomy. To address this, we propose a novel single-network paradigm: a dual-head 3D U-Net that jointly optimizes clinical dose regression and self-supervised anatomical reconstruction via a unified, synergistic objective. To force the shared encoder to extract robust structural representations, we introduce a structure-aware biased masking strategy targeting critical tissue boundaries where dose gradients are steepest, paired with an asymmetric skip-connection topology that removes skip pathways exclusively from the reconstruction decoder to prevent feature leaking.

Evaluated under a strict, budget-matched 80k-iteration threshold, our synergistic framework outperforms the standard 3D U-Net model trained from scratch, yielding a 0.14 Gy reduction in combined error driven by clear gains across both spatial dose (-0.10 Gy) and DVH conformity (-0.18 Gy). Furthermore, to benchmark its high-tier clinical competitiveness, the model was evaluated against the OpenKBP challenge-winning Cascaded 3D U-Net (C3D). With test-time augmentation, our streamlined single network achieves statistical parity with the complex C3D baseline (2.187 Gy vs. 2.134 Gy) while eliminating the need for multi-network cascading or knowledge distillation.

## METHODS AND MATERIALS

**Synergistic Model Innovations**

- Dual-Head Joint-learning: A single 3D U-Net backbone optimizing dose prediction and anatomical reconstruction simultaneously under one joint objective.
- Structure-Aware Masking: The MAE branch selectively masks critical organ-at-risk (OAR) boundaries, forcing the network to resolve features where dose gradients are steepest.
- Asymmetric Skip Connections: Pathways are completely removed for the reconstruction decoder to prevent feature leaking and force geometric encoding at the latent bottleneck.

**Experimental Setup**

- Data Split: Evaluated on the OpenKBP dataset using the standard 200 train / 40 validation / 100 test patient cohort.
- Protocol & Metrics: Optimization held to a strict 80k-iteration budget with test-time augmentation (TTA). Framework accuracy is evaluated via voxel-by-voxel Dose, clinical DVH, and Combined scores against both a baseline trained from Scratch and the C3D cascade winner.

## RESULTS

At the 80k-iteration threshold, the Synergistic MAE framework significantly outperforms the standard scratch baseline, reducing the spatial Dose score from 2.814 Gy to 2.719 Gy and the clinical DVH score from 1.836 Gy to 1.654 Gy. This isolated breakdown reveals a net MAE task contribution of -0.10 Gy for spatial dose and a more pronounced -0.18 Gy drop for DVH metrics, demonstrating that anatomical co-learning primarily drives superior dose-volume histogram accuracy. Crucially, these gains bring the streamlined single-network model into strict clinical competitiveness with the heavy C3D cascade winner (2.642 Gy Dose, 1.625 Gy DVH), achieving statistical parity in the cumulative combined score (2.187 Gy vs. 2.134 Gy) without any multi-stage overhead.

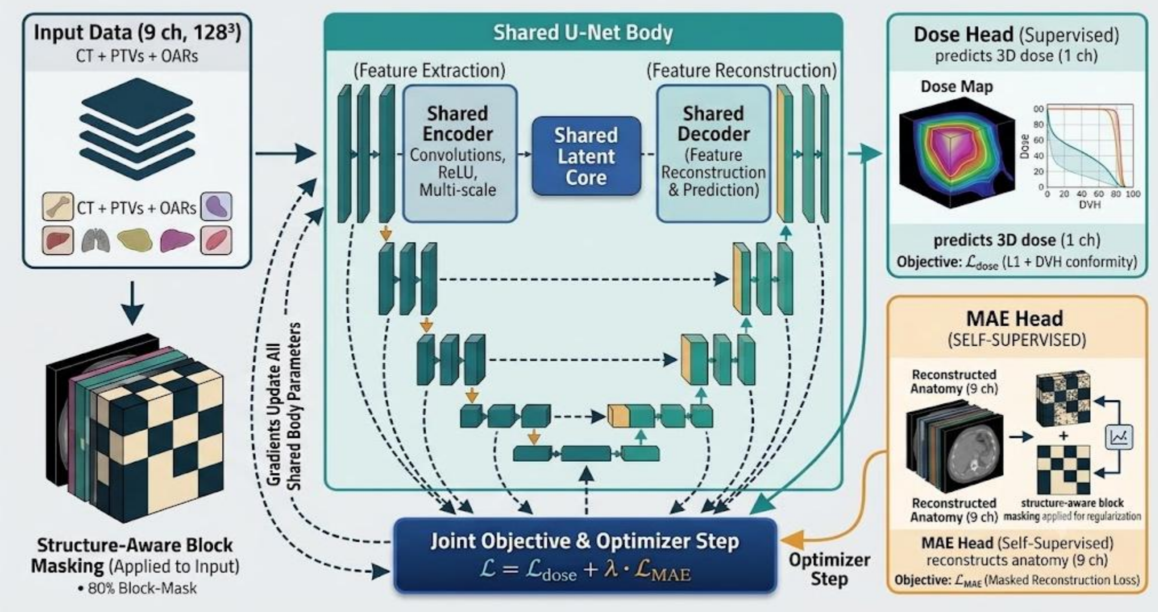


**Figure 1.** Synergistic Joint Training Workflow

**Table 1.** The performance of budget-matched OpenKBP test results.

| Model | Dose Score | DVH Score | Combined |
|---|---|---|---|
| Scratch, no MAE | 2.814 | 1.836 | 2.325 |
| **Synergistic MAE** | 2.719 | 1.654 | 2.187 |
| C3D Cascade Model | 2.642 | 1.625 | 2.134 |
| Δ MAE Contribution | -0.1 | -0.18 | -0.14 |

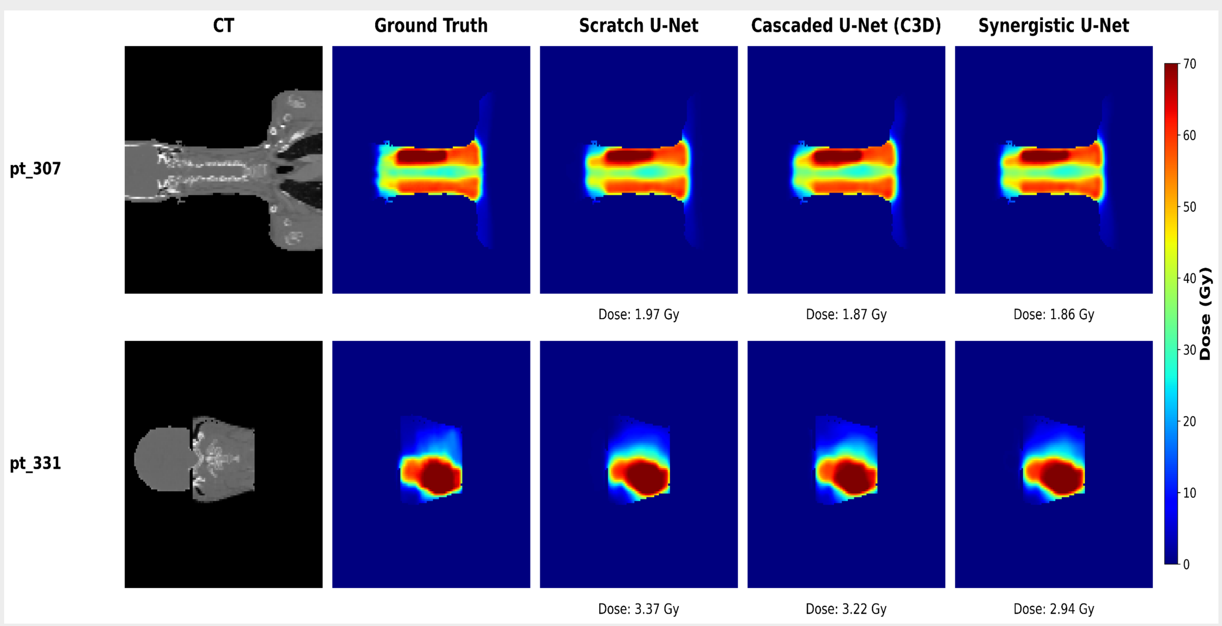


**Figure 2.** Comparison of the predicted dose vs the ground truth in the three models.

## DISCUSSION

A single, structure-aware U-Net matches complex multi-network cascades by sharing a self-supervised anatomical reconstruction task. This joint-learning forces the shared encoder to internalize robust structural representations, driving a -0.14 Gy reduction in combined error over a standard scratch baseline. The benefit is heavily concentrated in the clinical DVH score (-0.18 Gy drop vs. -0.10 Gy spatial dose), proving that targeted structural regularization provides critical geometric precision precisely where steep dose gradients intersect tight organ boundaries.

By achieving statistical parity with the OpenKBP challenge-winning C3D cascade (2.187 Gy vs. 2.134 Gy), this framework demonstrates that sufficient training runway paired with strong structural priors can successfully substitute for heavy multi-stage pipelines. Clinically, a single-network paradigm is vastly simpler to train, hyperparameter-tune, and deploy than a traditional cascade-plus-distillation pipeline. This eliminates substantial operational and computational overhead with zero penalty to final dosimetric accuracy.

To evaluate true clinical generalizability beyond synthetic OpenKBP plans, external validation on an institutional Emory head-and-neck cohort is underway. This work characterizes model degradation under real-world clinical planning variations and low-sample data scaling constraints. Preliminary scaling results indicate that synergistic initialization preserves its performance edge, mitigating model drift when subjected to the challenging cross-modality domain shift from fixed-field IMRT to Volumetric Modulated Arc Therapy (VMAT).

## CONCLUSIONS

This work demonstrates that a single 3D U-Net with structure-aware masked autoencoding can entirely replace complex multi-stage cascade and distillation pipelines. By driving a -0.14 Gy reduction in combined error over a scratch baseline, the self-supervised MAE task inherently captures the steep-gradient geometric precision required for superior DVH accuracy. Under an equal training budget, this streamlined architecture achieves statistical parity with the OpenKBP challenge winner (2.187 Gy vs. 2.134 Gy). Ultimately, prioritizing targeted structural regularization and adequate training runway over architectural complexity yields a vastly simpler, clinically deployable alternative with zero compromise on final dosimetric accuracy.

## REFERENCES


1. Babier, A., Zhang, B., Mahmood, R., Moore, K. L., Purdie, T. G., McNiven, A. L., & Chan, T. C. (2021). OpenKBP: the open-access knowledge-based planning grand challenge and dataset. Medical Physics, 48(9), 5549-5561.