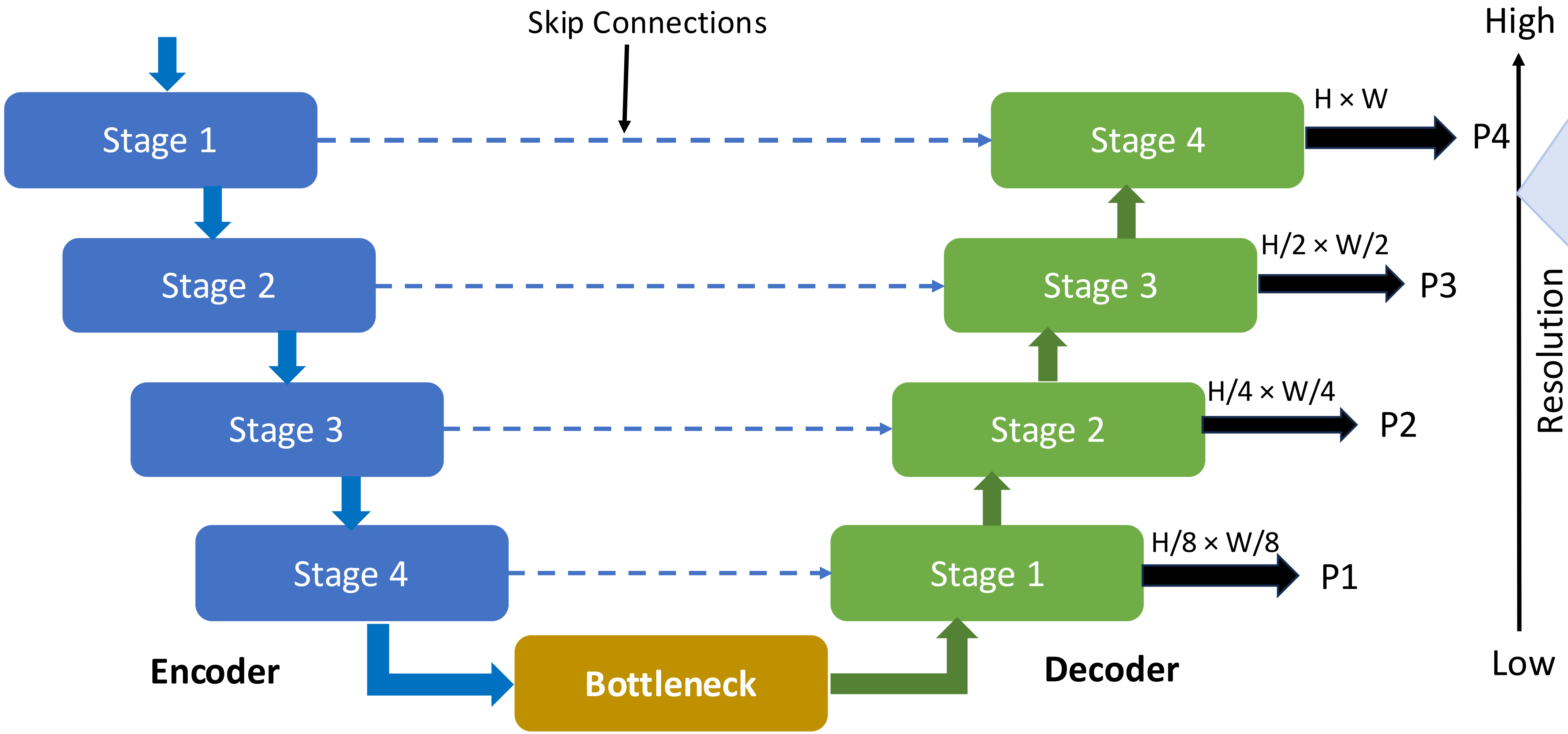
LoMix: Learnable Weighted Multi-Scale Logits Mixing for Medical Image Segmentation

Md Mostafijur Rahman and Radu Marculescu
The University of Texas at Austin

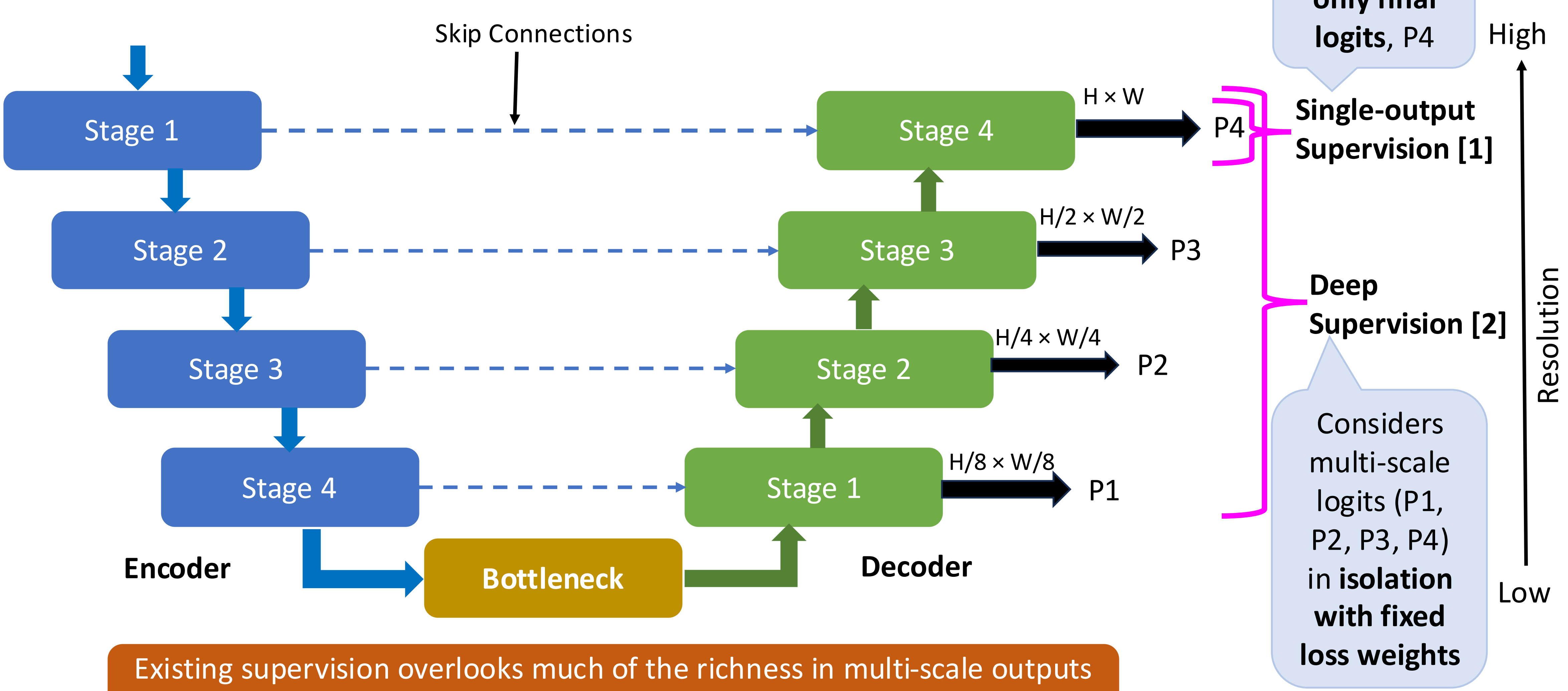
Motivation: Multi-scale Outputs Matter



- Small, thin objects rely on high resolution fine details (P3, P4)
- Large objects benefit from low resolution coarse context (P1, P2)

U-shaped architectures can generate **multi-scale decoder logits** containing **coarse-to-fine** information

Limitation of Existing Supervision



Existing supervision overlooks much of the richness in multi-scale outputs because it fails to use complementary information across scales

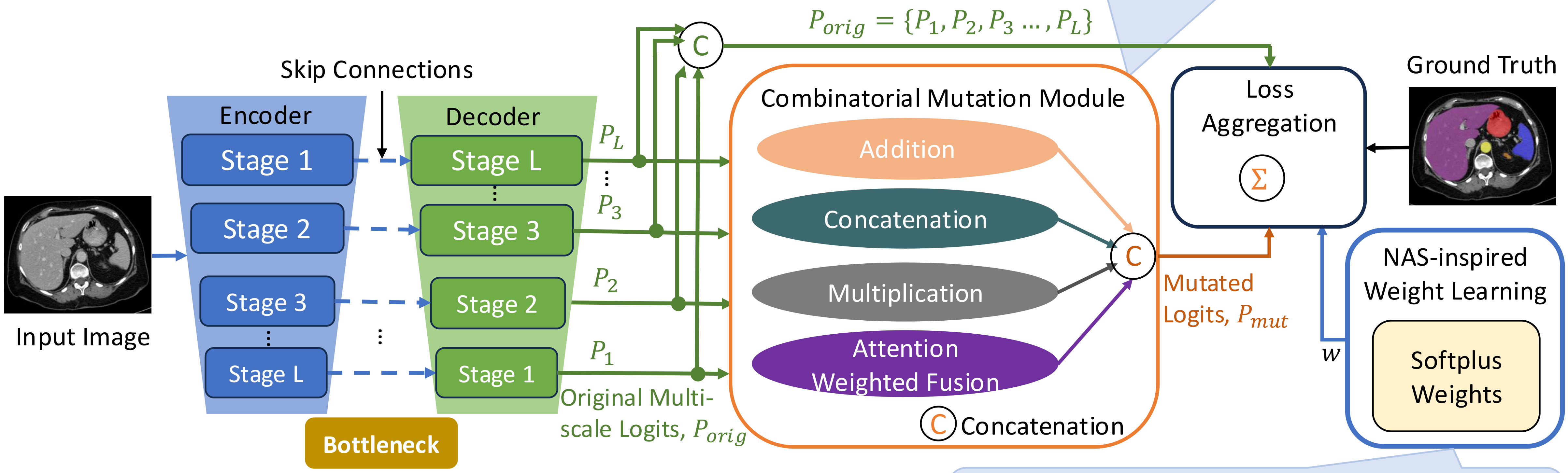
[1] Ronneberger, O., Fischer, P. and Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention,* 2015.
[2] Zhou, Z. et al. Unet++: A nested u-net architecture for medical image segmentation. In *International Workshop on Deep Learning in Medical Image Analysis,* 2018.

Our Solution: Learnable Logit Mixing (LoMix)

Key innovations

- Combinatorial Mutation Module (CMM)
- NAS-Inspired Softplus Weight Learning

- Mixes every subset of decoder logits using 4 differentiable operators
- Generates 4(2^L-1-L) synthetic logits
- Learns jointly with the U-shaped network

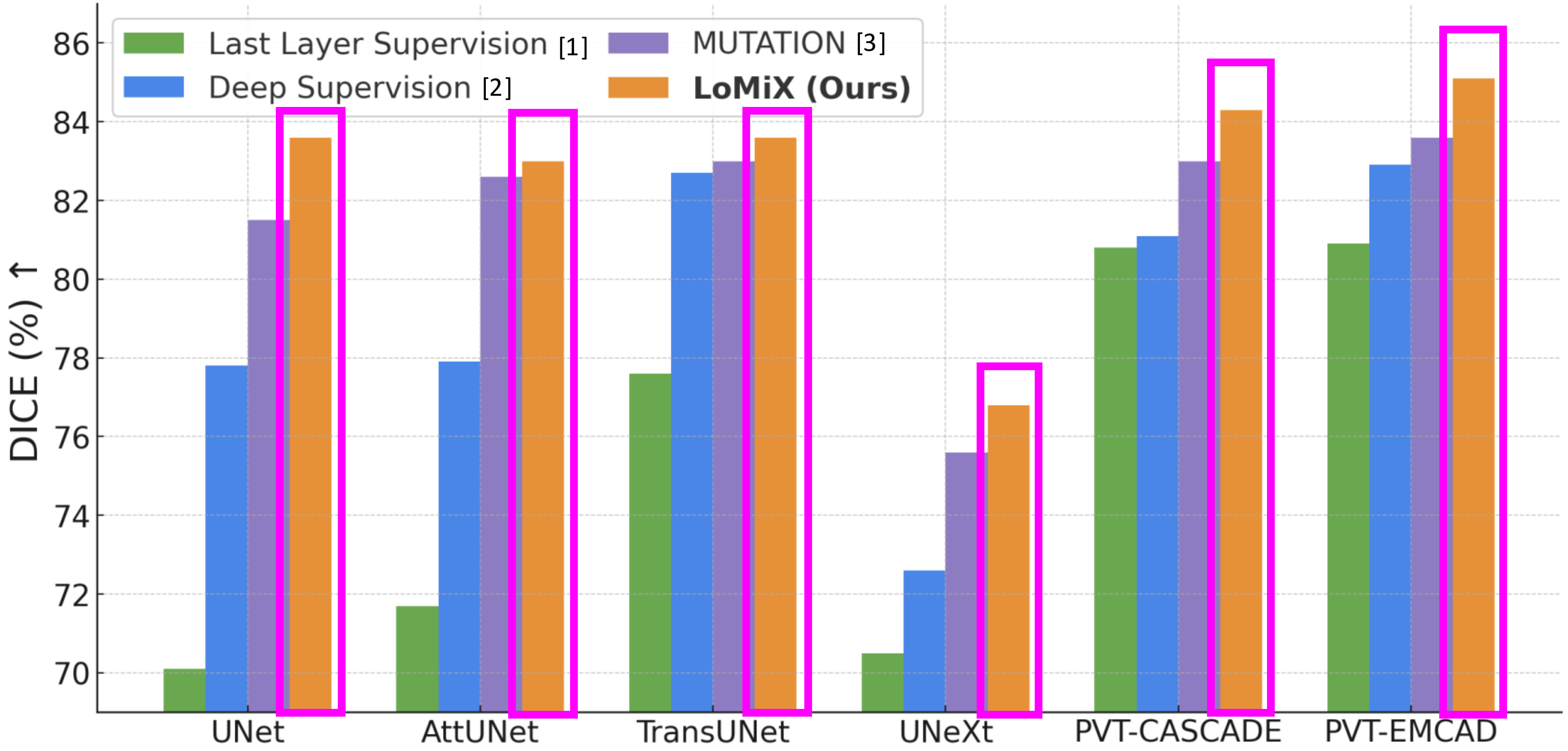


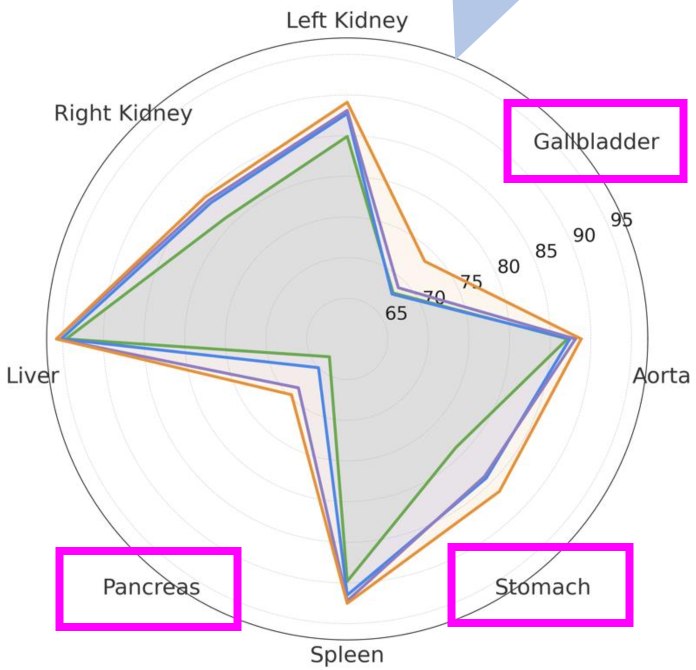
LoMix Turns fixed supervision into **learnable multi-scale generative logits mixing**

- Learns jointly with the U-shaped network
- Automatically weights **which scales** and which **fusion operators** matter most

LoMix Improves Segmentation Across Architectures

PVT-EMCAD architecture with different supervisions





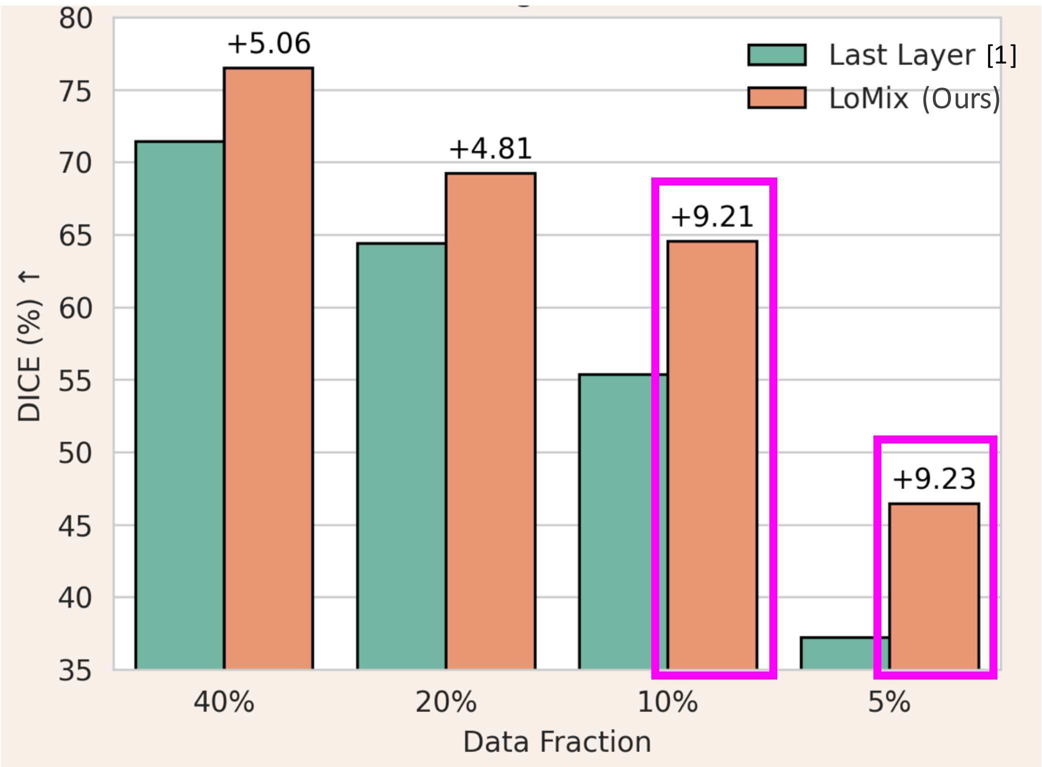
- 8 abdomen organ segmentation results on Synapse dataset
- 18 CT scans are used for training and 12 CT scans for testing

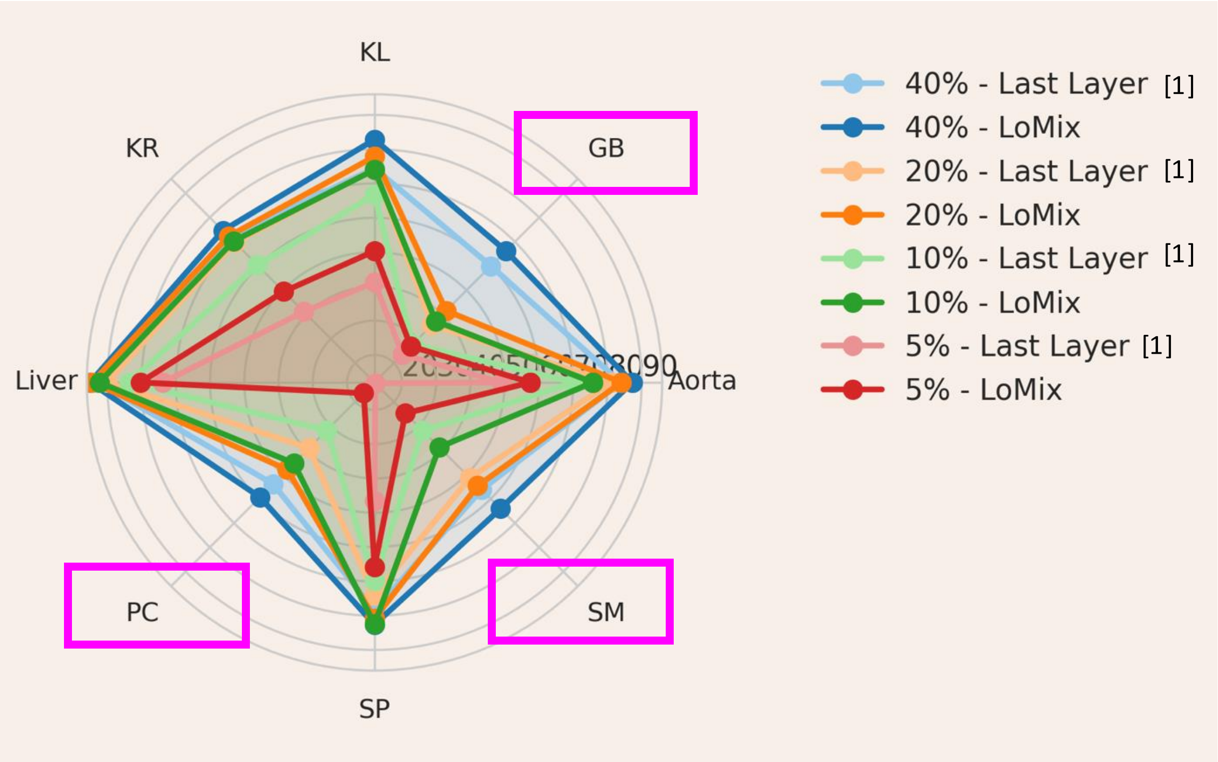
On average across all architectures, LoMiX improves DICE by +7.5% over Last Layer, +3.6% over Deep Supervision, and +1.2% over MUTATION

The largest DICE gains on smaller low-contrast organs: Gallbladder (+5.5), Pancreas (+6.6); and highly variable organ: Stomach (+7.6)

[1] Ronneberger, O., Fischer, P. and Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention,* 2015.
[2] Zhou, Z. et al. Unet++: A nested u-net architecture for medical image segmentation. In *International Workshop on Deep Learning in Medical Image Analysis,* 2018.
[3] Rahman, M.M. and Marculescu, R. Multi-scale hierarchical vision transformer with cascaded attention decoding for medical image segmentation. In *Medical Imaging with Deep Learning,* 2024.

LoMix is More Effective in Limited Data Scenarios





Synapse limited-data results with PVT-EMCAD: +5.1% with 40% scans, +4.8% with 20% scans, +9% with 10% or 5% scans

Strongest gains on small, difficult organs such as Gallbladder (GB), Pancreas (PC), and Stomach (SM)

[1] Ronneberger, O., Fischer, P. and Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention,* 2015.

Key Takeaways

Conceptual

- **First generative and learnable** multi-scale fusion at the logit level
- **Interpretable:** learned per-scale weights show where gains come from
- **Plug-and-play with U-shaped architectures and no inference overhead:** no architecture changes needed
- **Data-efficient:** synthetic logits enable high performance with minimal data

Results

- **Universal gains:** boosts 10 tested networks across six tasks
- **Outperforms standard supervision:** up to +13.5 vs Last Layer, +5.8 vs Deep Supervision, +2.1 vs MUTATION
- **Largest gains on small and hard organs:** +5.5 (Gallbladder), +6.6 (Pancreas), +7.6 (Stomach)

Code

SLD Group

Thank You

Contact <mostafijur.rahman@utexas.edu>