Label-Free SOM–CNN Framework for Scalable Computational Imaging

Pixel-Level Self-Organizing Maps as Zero-Cost Teachers for Lightweight U-Net Students

Xinxin Sun, Ph.D.
Postdoctoral Candidate – Computational Imaging and AI
University of Maryland, College Park

Designed to build a repeatable, label-free SOM–CNN and registration pipeline that scales from CT-style slices to large 3D/4D biomedical imaging datasets with minimal GPU cost.

Motivation & Problem Setting

Target imaging scenario

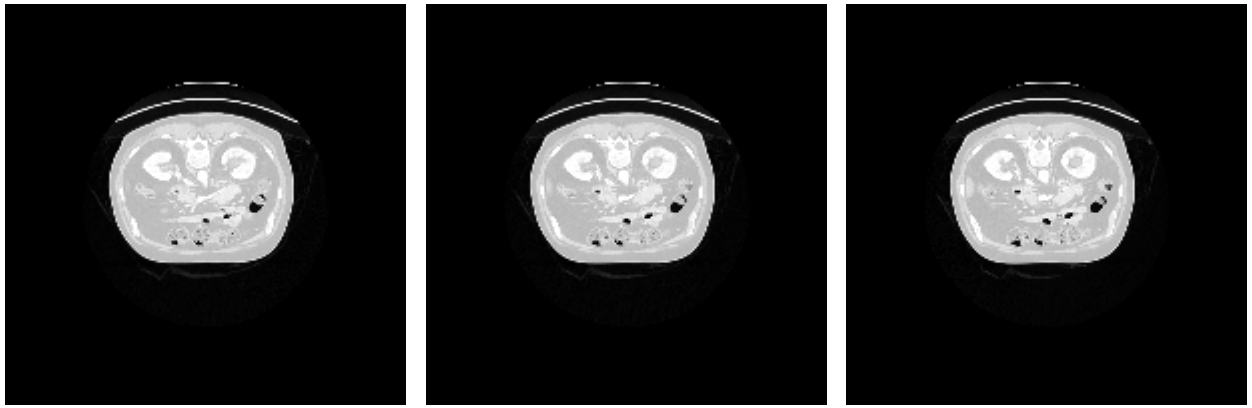
- Volumetric / time-resolved imaging (CT-style slices, MRI, XCT, etc.)
- Need to localize evolving “defect / lesion” patterns across slices and time

Key pain points

- Pixel-level annotations are expensive and often unavailable
- Classical unsupervised methods (thresholding, basic clustering) are not stable or transferable across slices
- Heavy supervised CNNs / Transformers require many labels and large GPU budgets

Goal of this demo

- Start from a **single slice** with no manual labels
- Use **pixel-level SOM** to obtain a repeatable, interpretable segmentation and transferable centroids
- Train a **lightweight U-Net** to learn from this SOM “teacher” and propagate labels to neighboring slices
- Align slices via feature-based registration to enable consistent 3D/4D analysis



z223 z224 z225

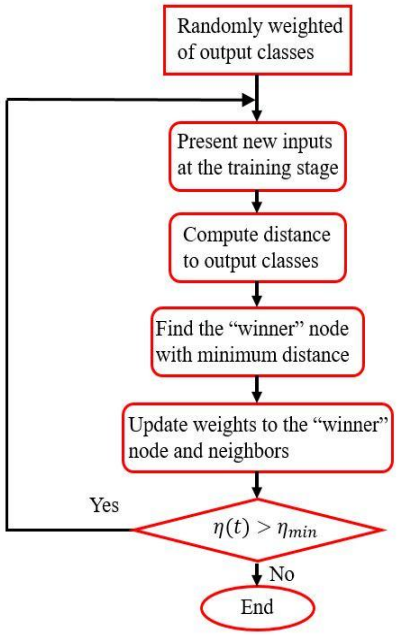
Example CT slice (cropped) from the AutoPET III whole-body FDG-PET/CT dataset (Gatidis & Kuestner, FDG-PET-CT-Lesions, The Cancer Imaging Archive, 2022); downloaded via the ClinicalDataScience/autoPETIII GitHub repository.

Label-free SOM defines structure; a compact CNN and simple registration pipeline propagate it across slices with minimal compute and zero human annotation.

Method Overview

Self-Organizing Map (SOM) for Label-Free, Physics-Informed Segmentation in CT / PET-CT Imaging

Feature	Physical / Statistical Meaning in CT / PET-CT
Grayscale intensity	Local attenuation / tissue density (CT) or tracer uptake (PET); basic structural and metabolic contrast at each voxel.
Edge magnitude	Sharp organ and lesion boundaries; separates parenchyma, background, and reconstruction edges.
Local contrast	Heterogeneity between lesion and surrounding tissue; highlights hyper- and hypo-intense regions.
Thinness / filament metric	Captures vessel-like or cortical-rim structures and small circular defects; sensitive to thin rims and small holes.
Local mean (3×3 / 5×5)	Neighborhood background estimator that smooths noise while preserving local context around lesions.
Anisotropy (p, q)	Sensitivity to directional streaking, motion artifacts, and elongated structures; encodes orientation information.
Intensity ratio surrogates	Simple normalized intensity ratios (row / column or local/global) that stabilize contrast across slices / patients and enrich feature diversity.



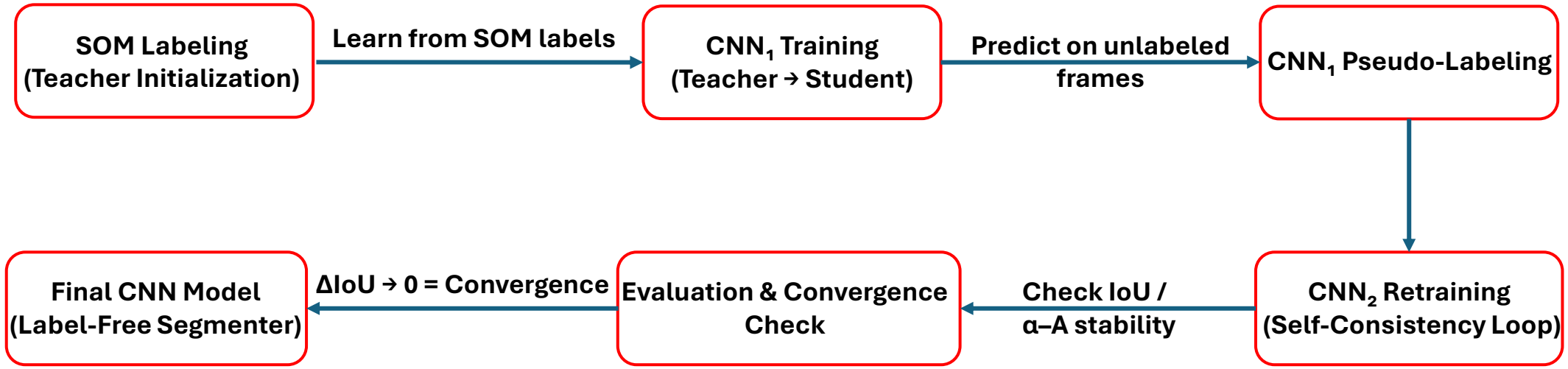
Workflow Summary

1. Extract a compact pixel-wise morphology feature vector (≈8–10 dimensions) from each 2D CT(-like) slice.
2. Apply SOM to cluster pixels into a small number of statistically distinct tissue / lesion regimes.
3. Interpret clusters as background, normal tissue, defect / lesion, and boundary / edge classes.
4. Obtain a class-label map and per-class centroids that can be reused on new slices.
5. Use simple visualizations (cluster map + radar plot) to quantify physical / clinical differences between clusters.

Key Advantages

- **Label-free** – No manual voxel-wise annotation required.
- **Physics-informed** – Features encode attenuation / tracer uptake patterns and reconstruction artifacts.
- **Interpretable** – Each SOM cluster corresponds to a statistically distinct, clinically meaningful regime.
- **Noise-tolerant** – Stable under low-dose, streaking, and moderate motion artifacts.
- **Repeatable & transferable** – Learned centroids can be applied directly to new slices / subjects in the same protocol.
- **Efficient** – Runs in seconds on CPU for 2D slices; ideal as a zero-cost "teacher" for downstream CNNs.

SOM–CNN Integration Workflow for Physics-Consistent Label Expansion



- **Stage 1 – Initial SOM Labeling (Slice z223)**

Apply pixel-level SOM to a single CT-style slice (z223) using morphology-driven features (contrast, edges, thin structures, local statistics).
The SOM produces an interpretable multi-class segmentation, and a binary "lesion / defect" mask is selected as the teacher label.

- **Stage 2 – CNN Training on SOM Labels (Teacher → Student)**

A lightweight U-Net is trained in PyTorch using only this SOM-labeled slice as supervision.
The CNN learns to reproduce the lesion mask and boundary structure defined by the SOM teacher.

- **Stage 3 – Pseudo-Label Generation on Neighboring Slices (z224, z225)**

The trained CNN predicts lesion masks on two neighboring slices (z224, z225), which were never SOM-labeled.
These predictions become pseudo-labels that propagate the SOM knowledge through the volume at essentially zero additional annotation cost.

- **Stage 4 – Refined Training with Real + Pseudo Labels**

The CNN is retrained on the combination of the original SOM label (z223) and the pseudo-labels (z224, z225).
This self-consistency loop stabilizes the morphology, sharpens boundaries, and improves coverage of small lesions.

- **Stage 5 – Convergence & Evaluation**

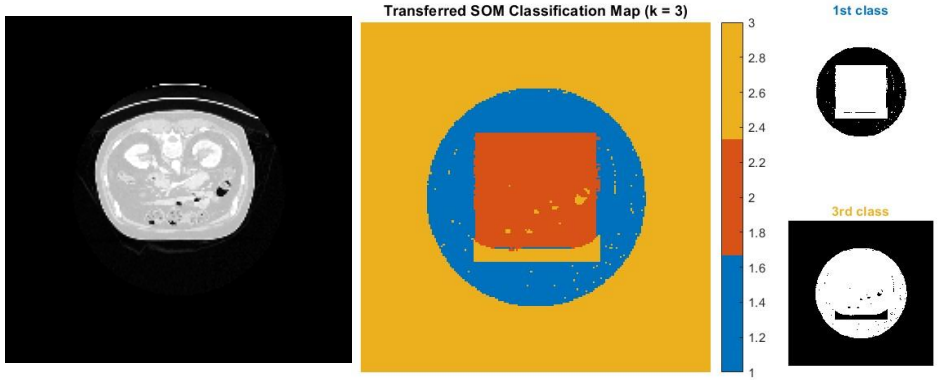
Final predictions are compared against SOM masks (used as teacher truth) using IoU, Dice, accuracy, precision, and recall.
In this CT-style demo, the CNN achieves Dice ≈ 0.99 and IoU ≈ 0.98 across three slices, indicating that a compact student network has successfully absorbed and extended the label-free SOM reasoning.

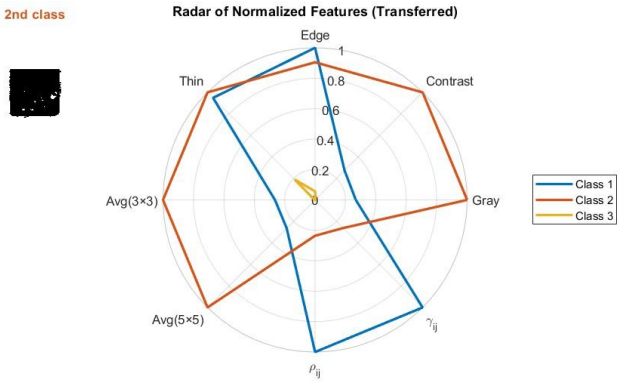
- **Stage 6 – Convergence Validation**

Mean IoU between the CNN and the SOM teacher increases steadily from **A → A + B → A + B + C** (A = slice z223, B = slice z224, C = slice z225).
When ΔIoU → 0, the CNN has effectively reached SOM-level reasoning and can replace SOM for fast slice-wise segmentation.

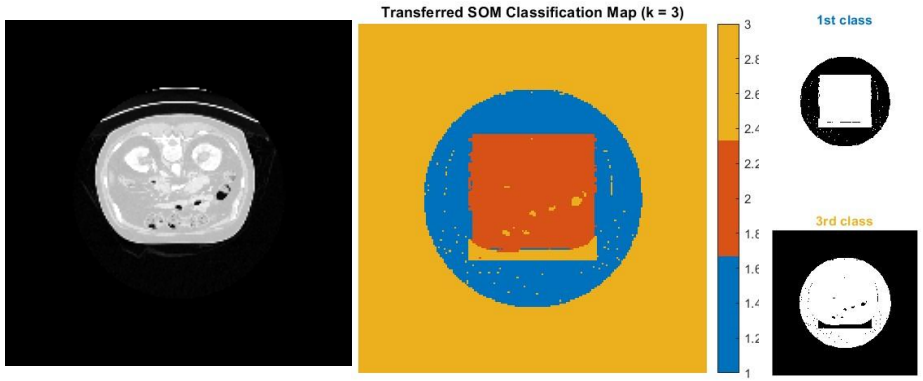
Unsupervised SOM Results on CT Slices

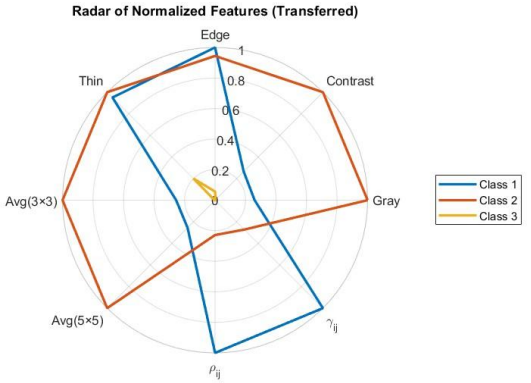
z223



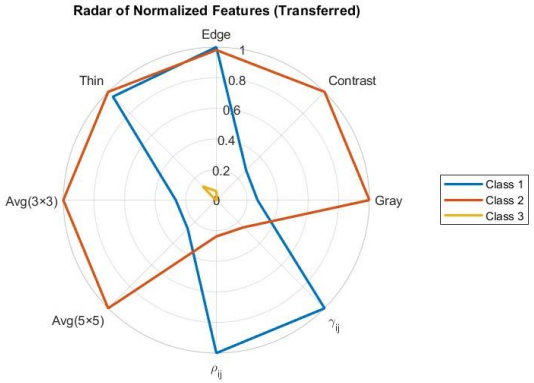


z224





z225



Pixel-Level SOM Segmentation and Slice-to-Slice Transfer

- **z223 (teacher slice)**

Unsupervised pixel-level SOM is trained on this single CT slice using morphology features (intensity, local contrast, edge, thinness, neighborhood means, etc.), discovering stable tissue / lesion–like clusters without any manual labels.

- **z224 & z225 (transferred slices)**

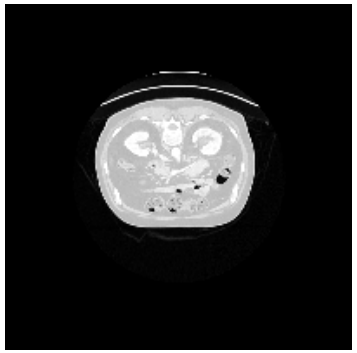
The learned SOM centroids from z223 are reused as fixed “tissue prototypes”: each pixel in z224 and z225 is assigned to the nearest centroid in feature space, giving consistent background–organ–lesion masks with zero re-training and zero annotation.

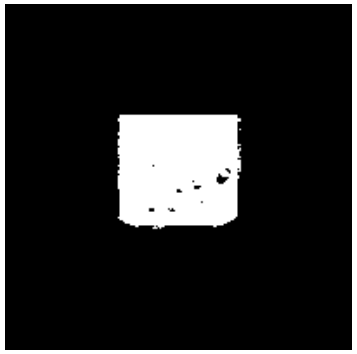
- **Radar plots**

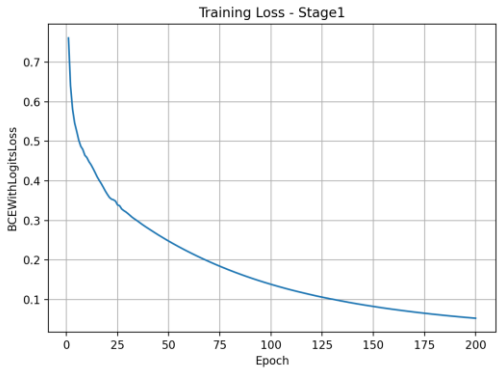
Cluster-wise feature profiles show that the “lesion” cluster maintains a similar signature across slices, indicating a repeatable, physiology-consistent phenotypic segmentation that can serve as a teacher for downstream CNNs.

SOM–CNN Results: Teacher–Student Agreement Across Slices

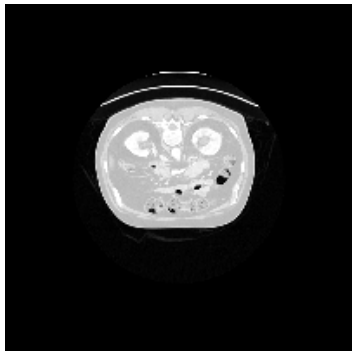
z223

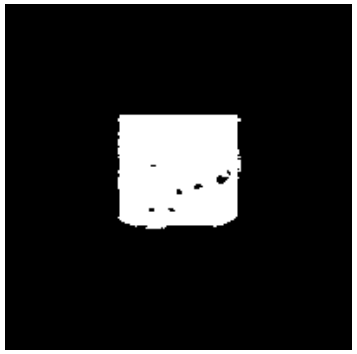


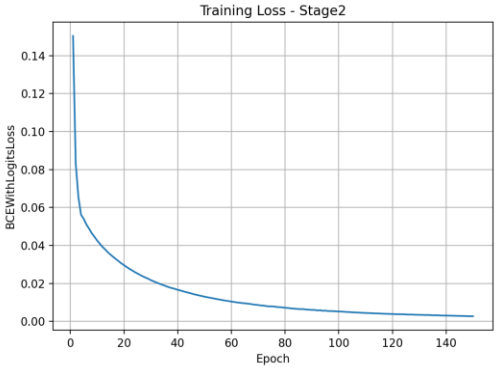




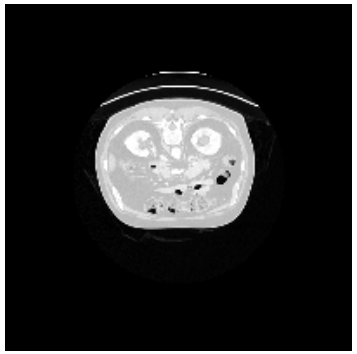
z224

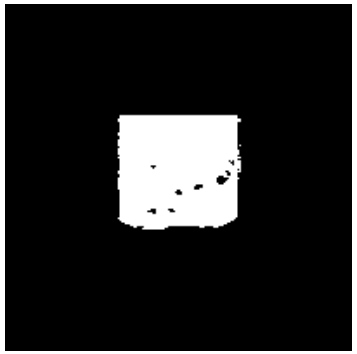


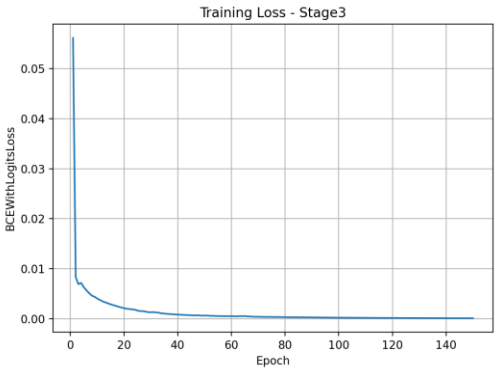




z225







Pixel-Level CNN Student Trained from SOM Teacher

- **z223 (teacher slice)**

CNN is trained on the SOM mask from z223 only. Loss curve shows fast convergence to a stable solution.

- **z224 & z225 (unlabeled slices)**

The trained U-Net predicts masks on neighboring slices that were never manually or SOM-labeled. Predicted masks closely match the SOM teacher masks (IoU ≈ 0.98, Dice ≈ 0.99 on average).

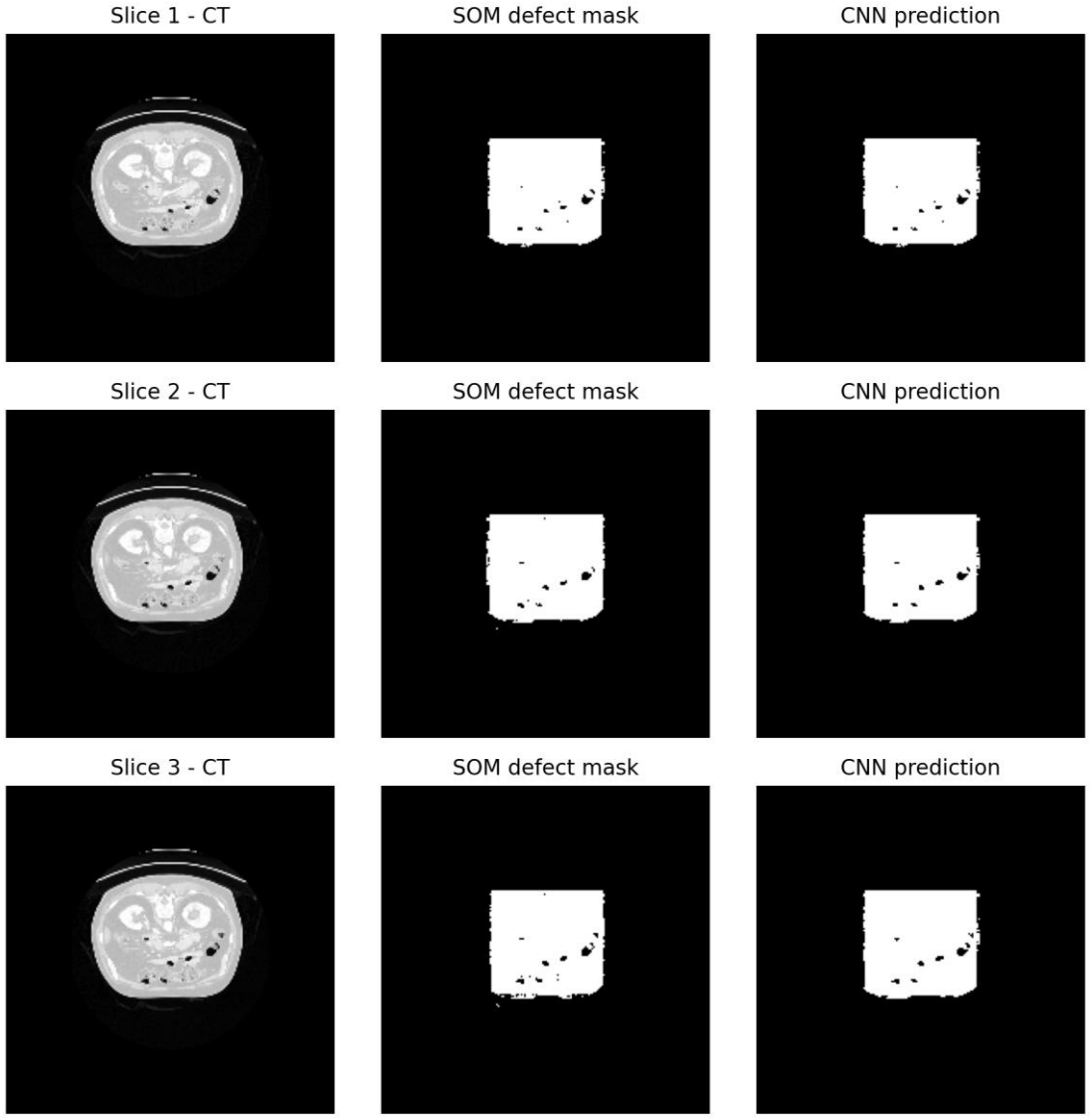
- **Self-consistency**

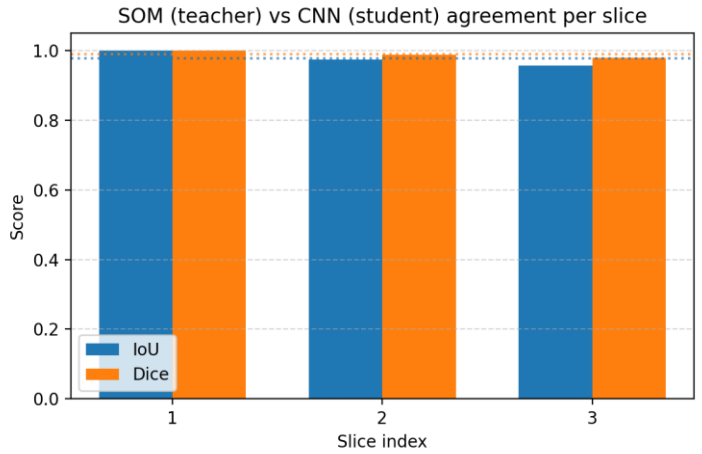
Loss curves remain low and smooth across stages, indicating that the CNN has learned the SOM's morphology-aware decision rule rather than overfitting a single slice.

- **Take-home message**

A single SOM-labeled slice is enough to train a lightweight CNN that generalizes to new CT slices with near-teacher performance and zero human annotation.

Direct SOM Transfer vs SOM-CNN Student: Two Modes of Label-Free Supervision





Direct SOM Slice-to-Slice Transfer (Top Row)

- Train pixel-level SOM once on one reference CT slice (z223).
- Reuse the learned centroids as fixed "tissue prototypes" and assign each pixel in new slices (z224, z225) to the nearest centroid.
- Produces stable lesion / background / organ masks in **seconds per 256×256 slice on CPU**, with **no retraining and no manual labels**.
- Ideal when a clinician or physicist needs **fast labels for a few individual CT slices** (e.g., quality control, ROI initialization, or quick annotation for new protocols).

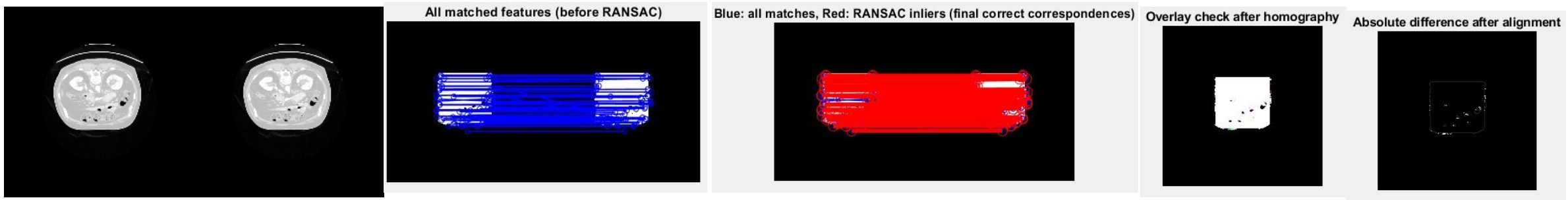
SOM-CNN Student Trained from SOM Labels (Bottom Row)

- A lightweight U-Net is trained on **one SOM-labeled slice**, then iteratively refined using pseudo-labels on neighboring slices.
- On z223–z225, the CNN predictions match the SOM teacher with **IoU ≈ 0.98 and Dice ≈ 0.99 on average**, while generalizing to slices never labeled by SOM directly.
- Inference runs on a **single GPU (or CPU) in milliseconds per slice**; training completes in **minutes**, not hours or days.
- Well-suited as a **scalable "label engine"** for large 3D/4D datasets or as a **low-cost supervision signal** for training larger vision models.

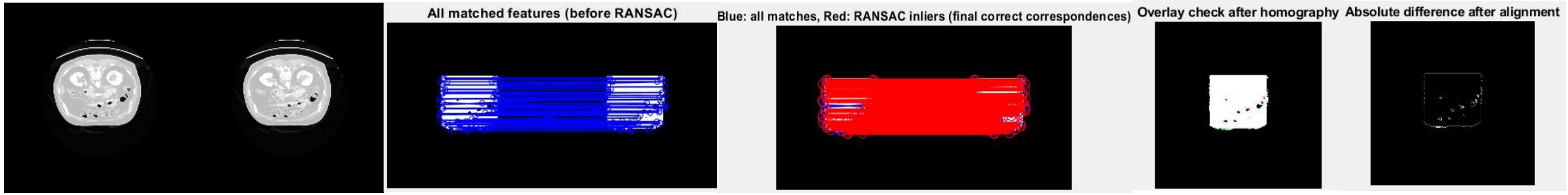
Key message:

- **Direct SOM transfer** is the fastest way to obtain pixel-wise labels for individual CT slices.
- **SOM-CNN** amortizes that cost, providing high-throughput, near-teacher segmentations for large cohorts using **single-GPU, label-free training.**

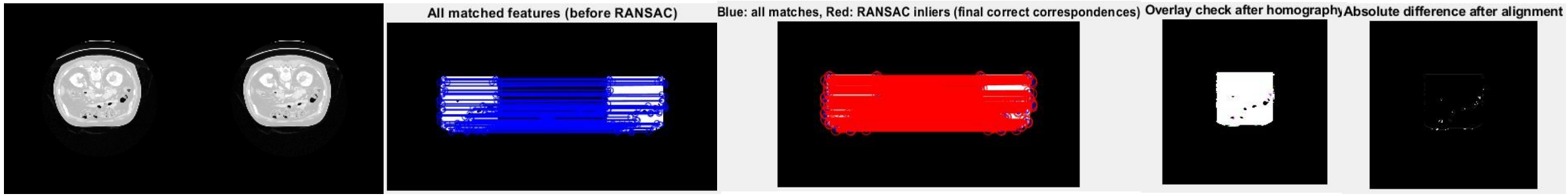
Slice-to-Slice Registration via KAZE + RANSAC



z224 z223



z225 z223



z225 z224

- KAZE detects stable anatomical keypoints on each pair of CT slices.
- RANSAC keeps only geometrically consistent matches (red inliers) and estimates the slice-to-slice homography.
- Overlays and absolute-difference maps confirm that most residual changes reflect lesion evolution rather than mis-alignment.

Conclusion & Next Steps

Key takeaways

- **Pixel-level SOM** provides repeatable, interpretable lesion / tissue segmentation and transferable centroids from a *single* labeled CT slice.
- **SOM–CNN teacher–student** pipeline achieves near-teacher IoU/Dice on new slices with **no manual annotations**, minutes of training, and single-GPU / CPU-friendly cost.
- **Slice-to-slice registration (KAZE + RANSAC)** aligns morphology across slices so that residual differences mainly reflect *true lesion evolution* rather than mis-positioning.
- Together, these components form a **label-free, scalable computational imaging workflow** that can serve as a fast supervision signal or initialization for larger vision models and 3D/4D analyses.

Future directions

- Extend from 2D slices to full **3D/4D PET/CT volumes** and longitudinal studies.
- Couple SOM-CNN supervision with **transformers / foundation models** as a physics-aware pretraining signal.
- Integrate with **HPC pipelines** for large-cohort training and real-time clinical deployment.