Title

CT (Non-Contrast) Hemorrhage Detection with CT-Native Self-Supervised Pretraining (JEPA) vs ImageNet

Problem

Determine whether CT-specific self-supervised pretraining (JEPA) on unlabeled CT slices outperform standard ImageNet pretraining for intracranial hemorrhage detection on RSNA brain CT.

Data

- 1,200 labeled slices (400 hemorrhage, 800 non-hemorrhage) from RSNA Brain CT (non-contrast).
- 10,000 additional unlabeled CT slices from the same cohort.
- Slices brain-windowed (L=40, W=80), normalized, resized to 224×224.

Methods

- ResNet-18 (≈11M params) Backbone, same supervised pipeline across all tests.
- Self-supervised JEPA pretraining
    - Student/teacher ResNet-18 encoders with EMA teacher update.
    - Two augmented CT views per slice including affine, flips, HU jitter, and noise.
    - Cosine prediction loss + variance regularization on 10k unlabeled slices (5 epochs).
    - Save student encoder weights as CT-native initialization.

- Supervised fine-tuning
    - Binary hemorrhage classifier with Dropout + Linear head.
    - Trained on 1,200 labeled slices with class-balanced sampling and BCE+logits loss.
    - Compared three init modes:
        1. ImageNet-pretrained ResNet-18
        2. JEPA-pretrained CT encoder
        3. Random initialization ("scratch")
- Evaluation: ROC AUC, sensitivity, specificity on a held-out 20% validation split + Grad-CAM heatmaps.

Results (Val set, n=240)

- **ImageNet**: AUC ≈ 0.878, Sens ≈ 0.70, Spec ≈ 0.89
- **Scratch**: AUC ≈ 0.829, Sens ≈ 0.79, Spec ≈ 0.74
- **JEPA (CT-SSL)**: AUC ≈ 0.752, Sens ≈ 0.85, Spec ≈ 0.53

From a qualitative perspective, Grad-CAM shows that all three models attend to lesion-like regions. JEPA produces more diffuse attention, whereas ImageNet and scratch produce sharper, more focal hotspots.

Takeaways

- On this small RSNA subset and with a lightweight JEPA implementation, ImageNet pretraining maintains the best overall AUC and specificity.
- **JEPA pretraining improves hemorrhage sensitivity** but lacks specificity, consistent with the idea that CT-native SSL may be more label-efficient but needs more careful scaling and tuning to surpass ImageNet.
- This small study empirically supports the observation that "vanilla" self-supervised methods do not automatically outperform ImageNet for CT, and provides a baseline for more advanced CT-specific SSL (e.g., a closer reproduction of the Self-Directed Learning architecture, longer pretraining, or label-efficiency experiments).