Final Warm-Up Report

Hari Varshan Dharmendra Mohan Prabu
Arizona State University
hdharmen@asu.edu

Abstract

This single report consolidates seven warm-up exercises on miniJSRT teaching datasets: (T1) robust image I/O for PNG/JPG/DICOM with percentile windowing, (T2) CXR orientation classification using a frozen ResNet-18 head, (T3) gender classification, (T4) age regression, (T5) binary organ segmentation, (T6) multi-class segmentation (lungs + heart) with DeepLabV3-ResNet101 and mixed losses, and (T7) three-organ localization using boxes derived from cleaned masks with geometric failsafes. I also include the two unsupervised exercises: (T8) anomaly detection via a denoising convolutional autoencoder on `autoencoder_img`, and (T9) image clustering on `Directions01` using ImageNet features + PCA + k-means, with quantitative evaluation (AUROC/AP for T8, NMI/ARI/Purity/Silhouette for T9). Across tasks, I detail data handling, training recipes, evaluation, and practical heuristics that produce compact, reproducible baselines suitable for coursework and future extensions.

1. Introduction

The miniJSRT materials provide small, well-scoped tasks for practicing radiograph processing and basic deep learning workflows. This report unifies Tasks 1–7 into a coherent pipeline: from clean image I/O (Task 1) and a first supervised classifier (Task 2), to attribute modeling (Tasks 3–4), and finally to pixel- and box-level structure prediction (Tasks 5–7). I conclude with unsupervised tasks: anomaly detection (Task 8) and clustering (Task 9) that reuse the same clean I/O and feature backbones.

2. Data and Preprocessing

Datasets. For I/O (T1) I use miniJSRT `Practice_PNGandJPG` and `Practice_DICOM`. For orientation (T2) I use `Directions01` with its train/test split. For gender (T3) and age (T4), I use the provided educational subsets with stratified splits. For segmentation/localization (T5–T7) I use labeled masks compatible with lungs/heart classes. For T8 (anomaly detection) I use `autoencoder_img` where "flips" act as anomalies. For T9 (clustering) I use `Directions01`(up/down/left/right/flip).

Intensity windowing (T1). PNG/JPG are read via `PIL`. DICOM uses `pydicom` (`pixel_array`, optional rescale slope/intercept). Medical images are typically 12–16 bit, I convert to 8 bit with a percentile window (0.5–99.5%) and linear scaling:

$$I_{\text{8-bit}} = \text{round}\left(255 \cdot \frac{\text{clip}(I, p_{0.5}, p_{99.5}) - p_{0.5}}{p_{99.5} - p_{0.5} + \epsilon}\right).$$

Model input pipeline. Single-channel CXRs are replicated to three channels to match ImageNet backbones (ResNet/DeepLab). Images follow the default ImageNet resize/crop and mean/std normalization. For orientation (T2), no label-violating augmentations (e.g., flips) are used.

3. Methods

3.1. Task 1: PNG/JPG/DICOM I/O

Utilities:

1. `png_to_jpg` and `jpg_to_png`: `PIL`-based load, percentile windowing, save.
2. `dcm_to_png`: `pydicom` load, optional slope/intercept, same windowing, save 8-bit PNG.

These are non-destructive and facilitate quick visual QA for downstream tasks.

3.2. Task 2: Orientation Classification (ResNet-18)

I fine-tune only the final fully connected layer of a ResNet-18 (ImageNet init), freezing the backbone. Cross-entropy is optimized with AdamW:

$$\mathcal{L} = -\sum_{c \in \{\text{Up,Down,Left,Right}\}} y_c \log p_\theta(c \mid x).$$

Hyperparameters: epochs=5, batch=32, LR=1e-3, backbone frozen.

3.3. Task 3: Gender Classification

Backbone: ResNet-18 (ImageNet). I compare unweighted vs. inverse-frequency weighted cross-entropy on a near-balanced split. Training is head-only (frozen trunk) with AdamW, a brief low-LR fine-tune can be added if desired.

3.4. Task 4: Age Regression

Single-output regression head on ResNet-18. Targets are *z-scored* using the train split, training minimizes Smooth L1 (Huber). Metrics (MAE, RMSE, R^2, Pearson r) are reported in years. A two-stage schedule is used: head-only, then short fine-tune (unfreeze `layer4`+head) at low LR. I also report horizontal-flip test-time augmentation (TTA).

3.5. Task 5: Binary Segmentation (Organs vs. Background)

Model: FCN32s-style head over ResNet-18 features (ImageNet init), with early layers frozen. Loss: weighted cross-entropy for class imbalance + Dice on the foreground. Optimizer: AdamW with cosine LR, BN in frozen stages set to eval.

3.6. Task 6: Multi-class Segmentation (Lungs + Heart)

Model: DeepLabV3-ResNet101 with class-matched classifier/aux heads. Training uses EMA weights and a short SWA sweep at the end. Loss blend:

- Focal cross-entropy (class weights from mask histograms),
- Multi-class soft Dice (ignore index 255),
- Lovász-Softmax (IoU surrogate).

Inference uses scales $\{0.75, 1.0, 1.25\}$, horizontal flips, softmax averaging, and semantic cleaning (remove speckles, fill holes, keep two largest lung components).

3.7. Task 7: Localization from Masks (3 Boxes)

Starting from cleaned T6 masks:

1. **Clean**: morphology (open/close), hole fill, keep two largest lung blobs.
2. **Left/right split**: if lungs fuse, 1-D k-means on x-coordinates (k=2), then enforce a midline rule.
3. **Heart clamp/synthesis**: clamp heart to central gap and plausible size (20–45% thorax width). If empty, synthesize a central box with height $\sim$70% of lung band.
4. **Boxes**: tight boxes around left lung, right lung, heart with small padding.

This guarantees three boxes per image, outputs include PNG/JPEG overlays and a CSV (image ID, organ boxes, image size).

3.8. Task 8: Unsupervised Anomaly Detection (Autoencoder)

I train a lightweight denoising convolutional autoencoder on *normal* images only (exclude paths containing "flip"). The encoder uses small convolutional blocks with dropout and a compact latent (32 channels). The decoder mirrors the stride pattern. Training minimizes $0.7 \times (1{-}\text{SSIM}) + 0.3 \times \|x{-}\hat{x}\|_1$ with AdamW and cosine LR, light Gaussian noise on inputs encourages robustness. At test time, per-image anomaly score is the mean of normalized L1 and $(1{-}\text{SSIM})$, higher is more anomalous.

3.9. Task 9: Unsupervised Clustering (Directions)

Each image is resized and fed to a frozen ImageNet backbone (ResNet-18) to extract a global average pooled feature. I reduce dimensionality with PCA (whitened) and cluster with k-means (k=5 for $\{$up, down, left, right, flip$\}$). Metrics include normalized mutual information (NMI), adjusted Rand index (ARI), cluster purity (via majority class mapping), and Silhouette score.

4. Experiments and Results

4.1. Task 1 (I/O Verification)

On a small subset (up to 5 examples/format), visual QA confirms preserved contrast after windowing and viewability in standard tools. Outputs are written to new folders.

4.2. Task 2 (Orientation)

Frozen-backbone ResNet-18 with a learned linear head achieves high accuracy on the provided split.

Model	Train Acc	Test Acc
ResNet-18 (frozen)	**0.993**	**1.000**

Table 1: Task 2 accuracy on the provided test split.

4.3. Task 3 (Gender)

Setting	Test Acc
Unweighted CE (frozen)	0.882
Weighted CE (frozen)	0.849

Table 2: Task 3 test accuracy on `Gender01`.

Per-class (unweighted). female: P=0.897, R=0.833, F1=0.864, male: P=0.870, R=0.922, F1=0.895.
Confusion (unweighted). $[35, 7;\ 4, 47]$.

Setting	MAE ↓	RMSE ↓	R^2	r
Head+FT (no TTA)	6.90	9.65	0.494	0.704
Head+FT (+TTA)	6.88	9.24	–	–

Table 3: Task 4 age regression metrics (years).

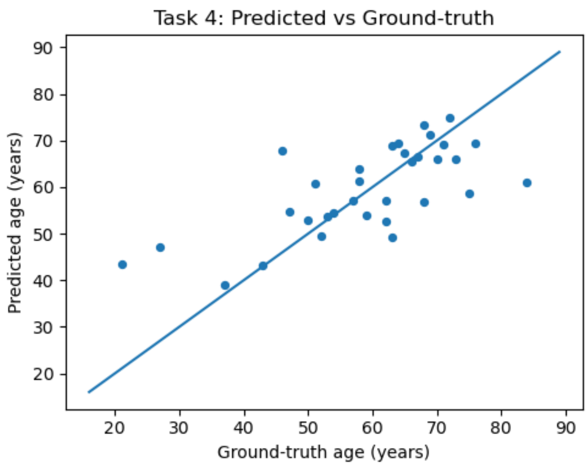

Figure 1: Task 4: Predicted vs. ground-truth ages.

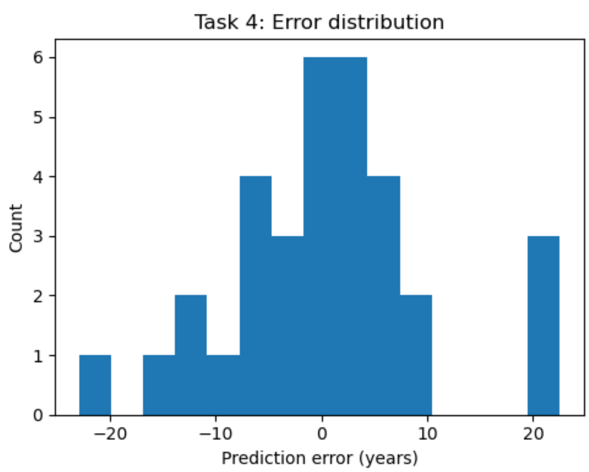

Figure 2: Task 4: Prediction error histogram.

4.4. Task 4 (Age): Table 3, Figure 1 and Figure 2

4.5. Task 5 (Binary Segmentation): Figure 3

Qualitatively, the lightweight FCN32s-style model separates thoracic foreground from background with stable boundaries after Dice weighting and morphological cleanup.

4.6. Task 6 (Multi-class Segmentation): Figure 4

DeepLabV3-ResNet101 with a mixed loss (Focal CE + soft Dice + Lovász) and simple TTA yields strong qualitative masks, an example hold-out printout indicates mean IoU ≈ **0.91** (no back-

Figure 3: **Task 5:** Example binary mask overlaid on input/prediction.

Figure 4: **Task 6:** Input, ground truth, prediction. (Training is 3-class: bg, lung, heart.)

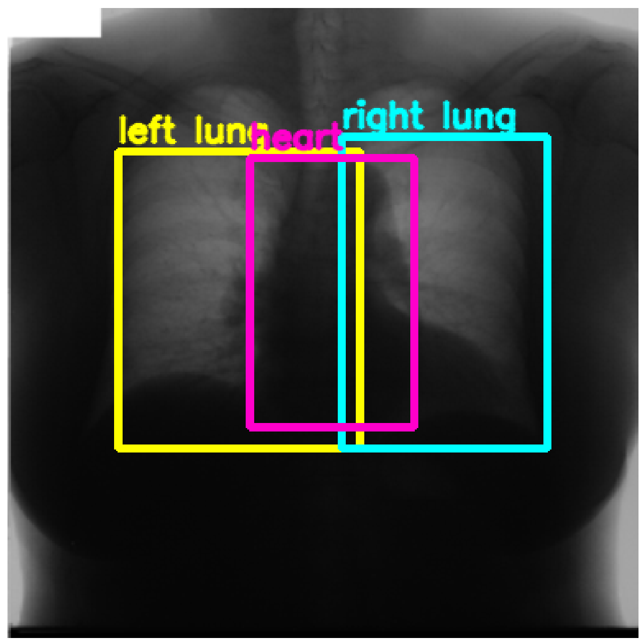

Figure 5: **Task 7:** Final boxes: left lung (yellow), right lung (cyan), heart (magenta).

ground).

4.7. Task 7 (Localization from Masks): Figure 5

The geometric post-process robustly returns three boxes per image (left/right lung, heart) even when the heart mask is weak or lungs are fused.

4.8. Task 8 (Anomaly Detection)

We treat flips as anomalies and normals as inliers. On the held set, the autoencoder achieves AUROC=**0.460** and AP=**0.173**. A threshold may be chosen by maximizing F1 on a tiny labeled slice, or by selecting the top $p\%$ of scores (I show p=5 in figures).

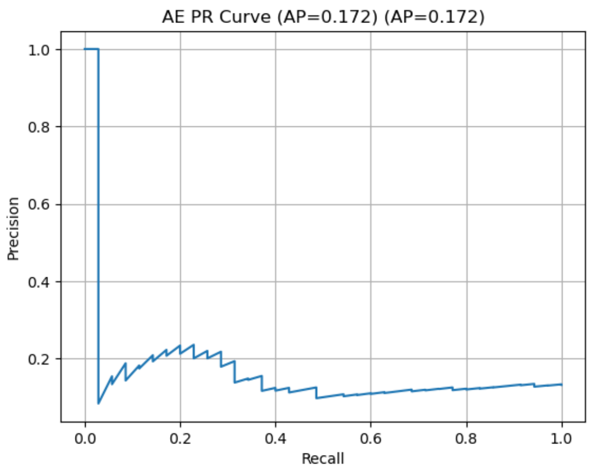

Figure 6: Task 8: PR curve for anomaly score (higher is better).

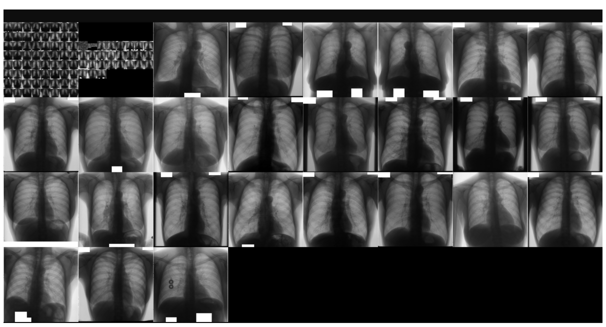

Figure 7: Task 8: Top-scoring anomalies visualized.

4.9. Task 9 (Clustering)

Clustering `Directions01` embeddings yields strong partitioning: NMI=**0.981**, ARI=**0.989**, Purity=**0.996**, Silhouette=**0.411**. Most confusion occurs between mirrored classes when pose is ambiguous.

Metric	NMI	ARI	Purity	Sil.
ResNet18+PCA+k-means	0.981	0.989	0.996	0.411

Table 4: Task 9 clustering metrics (1.0 is best, Silhouette $\in [-1, 1]$).

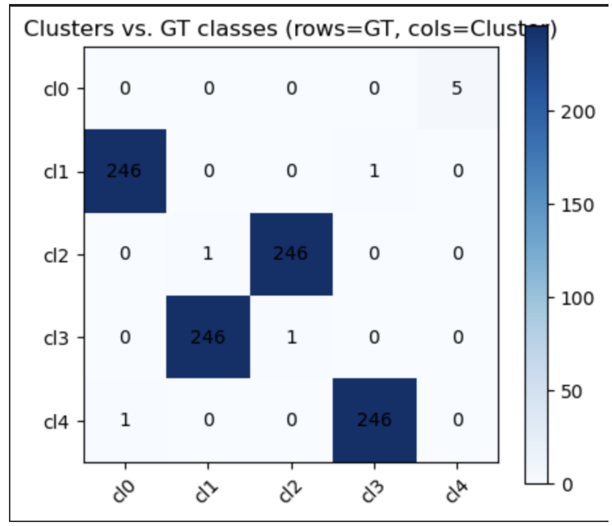

Figure 8: Task 9: GT classes vs. cluster assignments.

5. Discussion

Classification (T2–T3). For global attributes (orientation, gender), a frozen ResNet-18 with a trained head is often sufficient on small splits. Class weighting for near-balanced gender data modestly underperformed unweighted CE here.

Regression (T4). Z-scoring targets stabilizes optimization, brief low-LR fine-tuning of `layer4`+head and simple TTA provide small but consistent gains.

Segmentation & Boxes (T5–T7). A lightweight binary segmenter offers clean foreground masks, upgrading to DeepLabV3-ResNet101 with mixed losses improves organ delineation. The deterministic box-generator ensures usable localization even in failure cases by leveraging midline rules and plausible heart geometry.

Unsupervised (T8–T9). The denoising AE deliberately underfits anomalies, while simple, it already separates flips moderately well (AUROC 0.460). Performance can be raised by masking lungs to reduce background leakage, using perceptual losses, or scoring via latent-space likelihood. For clustering, ImageNet features with PCA+k-means nearly recover the ground-truth directions (NMI/ARI $\approx$ 0.98/0.99), small errors concentrate in borderline cases, suggesting either stronger augment-invariant features or a rotation-aware metric.

6. Conclusion

I present compact, reproducible baselines for Tasks 1–9 on miniJSRT: robust I/O, frozen-backbone classification/regression heads, and a practical segmentation→localization pipeline. These foundations are easy to extend (e.g., stronger augments, uncertainty, multi-task heads) while remaining coursework-friendly.

Code

The code for these warm-up exercises can be found here.

References

[1] L.-C. Chen, et al. Rethinking atrous convolution for semantic image segmentation. *arXiv:1706.05587*, 2017.

[2] M. Berman, et al. The Lovász-Softmax loss: A tractable surrogate for IoU optimization. *CVPR*, 2018.