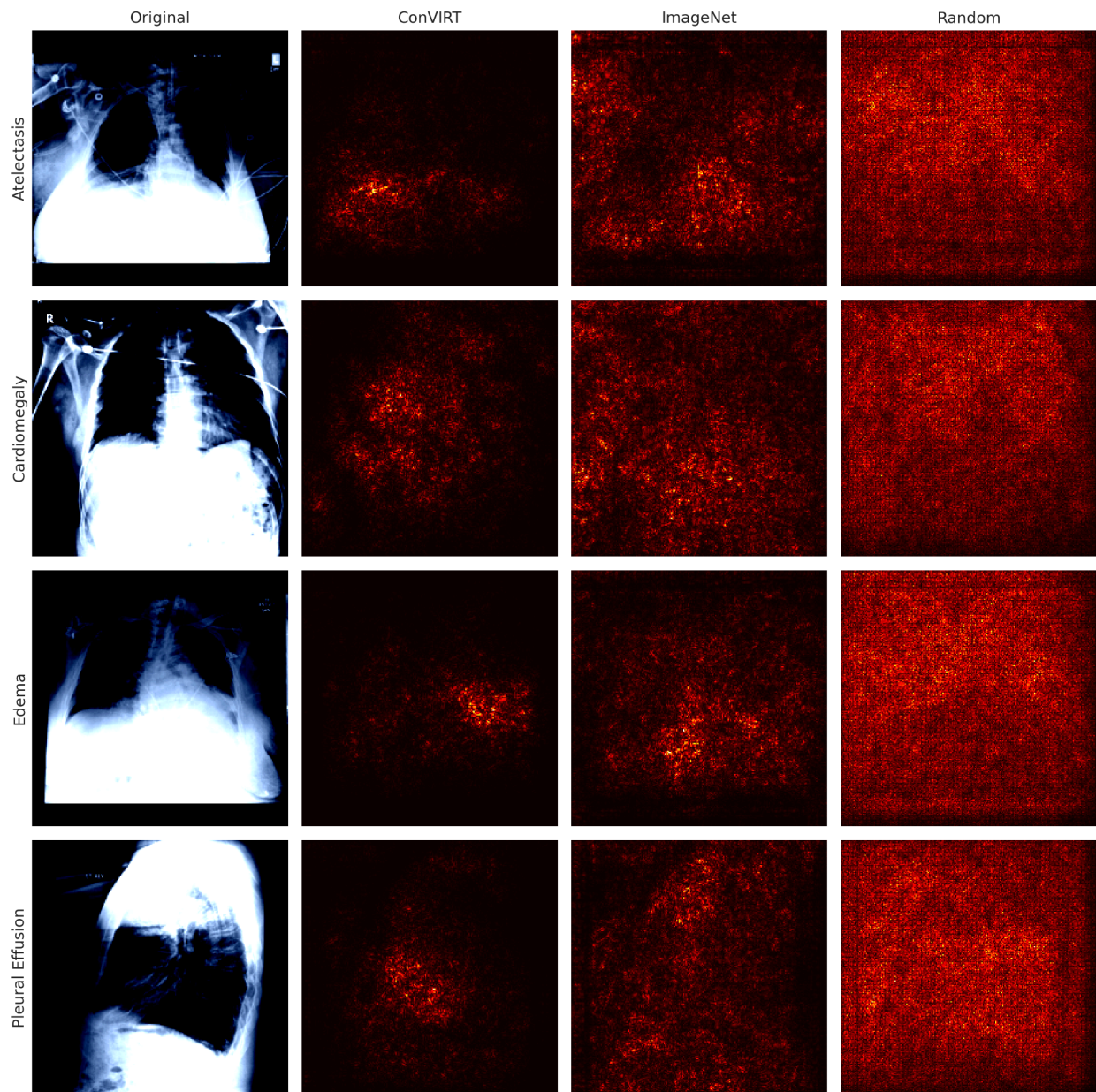


CheXpert Saliency map Visualization

Comparison: 10% Fine-Tuning vs Baselines

This document contains the results of gpt-4o-mini's analysis of the saliency map images.



ConVIRT's 10% fine-tune saliency maps exhibit substantially more focused, pathology-relevant activations compared to both ImageNet-pretrained and randomly-initialized baselines. Below is a detailed breakdown:

1. Atelectasis

- **ConVIRT:** Strong activation in the upper-lateral region of the affected lung, matching the typical collapse zone.
- **ImageNet:** Diffuse highlights across the lung fields, with no clear anatomical focus.
- **Random:** Uniform noise—no discernible pattern.

2. Cardiomegaly

- **ConVIRT:** Concentrated saliency over the cardiac silhouette, especially at the medial heart border where enlargement is most evident.
- **ImageNet:** Sprinkled attention across the mid-thorax without clear correspondence to the heart shape.
- **Random:** Scattered, meaningless activations.

3. Edema

- **ConVIRT:** Localizes around the perihilar regions—consistent with interstitial fluid patterns (“bat-wing” distribution).
- **ImageNet:** Broad, patchy saliency across both lungs, diluting clinical relevance.
- **Random:** Random speckles, no clinical alignment.

4. Pleural Effusion

- **ConVIRT:** Focuses on dependent lung areas—costophrenic angles and diaphragmatic interface—where fluid accumulates.
 - **ImageNet:** Signals spread across the lung fields; fails to isolate pleural spaces.
 - **Random:** Evenly distributed noise.
-

Multiple Perspectives

1. **Clinical Interpretability**

ConVIRT's maps align with established radiographic signs (e.g., perihilar fluid for edema). This increases clinician trust by showing the model "looks" at the right regions.

2. **Data-Efficiency**

Achieving precise localization with just 10% of labels demonstrates that contrastive pretraining yields robust feature priors, reducing dependence on large annotated sets .

3. **Model Reliability**

Compared to ImageNet-fine-tuned models, ConVIRT's tighter activations suggest less reliance on spurious background cues, potentially improving generalization on unseen datasets.

Recommendations & Next Steps

- **Quantify Localization:** Compute overlap (e.g., IoU) between saliency peaks and expert-annotated bounding boxes to move beyond qualitative assessment.
- **Alternative Explainability:** Apply Grad-CAM++ or Integrated Gradients to compare sensitivity profiles.
- **Fine-tuning Strategies:** Investigate layer-wise learning rates to further sharpen focus under low-data regimes.
- **Cross-modal Validation:** Use corresponding report text to verify whether highlighted image regions match described findings.