# OPBNet: Oscillatory Phase-Binding Networks for Interpretable Zero-Shot Medical Visual Reasoning Without Pretraining

Supplementary Material

## 9. Full Theoretical Proofs

This appendix provides complete mathematical proofs supporting the theoretical claims made in the main paper. The analysis relies on tools from nonlinear dynamical systems, spectral graph theory, and statistical analysis of circular random variables. All symbols and notations follow the definitions established in the main manuscript.

### 9.1. Proof of Theorem 1 (Convergence of Discrete Synchronization Dynamics)

We analyze the stability of the discrete phase evolution process in the oscillator system. Consider the phase vector

$$\boldsymbol{\phi}^{(t)} \in \mathbb{R}^N$$

evolving under the iterative map

$$\boldsymbol{\phi}^{(t+1)} = \boldsymbol{\phi}^{(t)} + \Delta t\, F(\boldsymbol{\phi}^{(t)}),$$

where $F(\boldsymbol{\phi})$ denotes the interaction field induced by oscillator coupling.

**Lyapunov Energy Construction**
Define the scalar function

$$V(\boldsymbol{\phi}) = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} K_{ij}\cos(\phi_i - \phi_j).$$

The gradient of this function with respect to phase variable $\phi_i$ is

$$\frac{\partial V}{\partial \phi_i} = \sum_{j=1}^{N} K_{ij}\sin(\phi_i - \phi_j).$$

Hence the interaction field satisfies

$$F_i(\boldsymbol{\phi}) = -\frac{\partial V}{\partial \phi_i}.$$

Therefore the dynamics correspond to a gradient descent system

$$\boldsymbol{\phi}^{(t+1)} = \boldsymbol{\phi}^{(t)} - \Delta t \nabla V(\boldsymbol{\phi}^{(t)}).$$

**Energy Decrease Property**
Using the Taylor expansion of $V$ around $\boldsymbol{\phi}^{(t)}$:

$$V(\boldsymbol{\phi}^{(t+1)}) = V(\boldsymbol{\phi}^{(t)}) + \nabla V(\boldsymbol{\phi}^{(t)})^\top(\boldsymbol{\phi}^{(t+1)} - \boldsymbol{\phi}^{(t)}) + \mathcal{O}(|\boldsymbol{\phi}^{(t+1)} - \boldsymbol{\phi}^{(t)}|^2).$$

Substituting the update rule:

$$\boldsymbol{\phi}^{(t+1)} - \boldsymbol{\phi}^{(t)} = -\Delta t \nabla V(\boldsymbol{\phi}^{(t)}),$$

we obtain

$$V(\boldsymbol{\phi}^{(t+1)}) = V(\boldsymbol{\phi}^{(t)}) - \Delta t|\nabla V(\boldsymbol{\phi}^{(t)})|^2 + \mathcal{O}(\Delta t^2).$$

If

$$\Delta t < \frac{1}{N\max_{ij}|K_{ij}|},$$

the second-order term remains bounded and the energy strictly decreases:

$$V(\boldsymbol{\phi}^{(t+1)}) < V(\boldsymbol{\phi}^{(t)}).$$

**Convergence Result**
Since the energy $V$ is bounded below and monotonically decreasing, the sequence

$$\{V(\boldsymbol{\phi}^{(t)})\}$$

converges. Consequently the gradient norm approaches zero:

$$|\nabla V(\boldsymbol{\phi}^{(t)})| \to 0.$$

Thus the phase configuration converges to a stationary point of the energy landscape, proving stability of the synchronization dynamics.

### 9.2. Proof of Proposition 1 (Synchronization Cluster Formation)

Consider a subset of oscillators $S \subset \{1, \ldots, N\}$. Define the intra-cluster coupling strength

$$K_{\text{intra}} = \min_{i\in S}\sum_{j\in S} K_{ij}.$$

Similarly define the external coupling strength

$$K_{\text{ext}} = \max_{i\in S}\sum_{j\notin S}|K_{ij}|.$$

Assume

$$K_{\text{intra}} > K_{\text{ext}}.$$

**Phase Difference Dynamics**
For oscillators $i, j \in S$, define the phase difference

$$\delta_{ij}(t) = \phi_i^{(t)} - \phi_j^{(t)}.$$

Subtracting their update equations yields

$$\delta_{ij}(t+1) = \delta_{ij}(t) + \Delta t\Big[\sum_k K_{ik}\sin(\phi_k - \phi_i) - \sum_k K_{jk}\sin(\phi_k - \phi_j)\Big].$$

Using the mean value theorem and Lipschitz continuity of the sine function, we obtain the bound

$$|\delta_{ij}(t+1)| \le (1 - \Delta t(K_{\text{intra}} - K_{\text{ext}}))|\delta_{ij}(t)|.$$

Because $K_{\text{intra}} > K_{\text{ext}}$, the coefficient is strictly smaller than one.

**Exponential Convergence**
Iterating the inequality yields

$$|\delta_{ij}(t)| \le \rho^t |\delta_{ij}(0)|, \quad \rho = 1 - \Delta t(K_{\text{intra}} - K_{\text{ext}}) < 1.$$

Therefore

$$\delta_{ij}(t) \to 0,$$

implying that oscillators within $S$ synchronize asymptotically. Thus clusters of mutually coupled oscillators emerge naturally from the dynamics.

## 9.3. Proof of Proposition 2 (Order Parameter as Confidence Indicator)

Consider the complex representation of oscillator phases

$$z_i = e^{i\phi_i}.$$

The order parameter is

$$r = \left|\frac{1}{N}\sum_{i=1}^{N} z_i\right|.$$

Assume phases follow a Von Mises distribution

$$p(\phi|\mu,\kappa) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa\cos(\phi-\mu)}.$$

**Expected Synchronization**
The expectation of $z_i$ under this distribution is

$$\mathbb{E}[z_i] = \frac{I_1(\kappa)}{I_0(\kappa)} e^{i\mu}.$$

Therefore

$$\mathbb{E}[r] = \frac{I_1(\kappa)}{I_0(\kappa)}.$$

**Monotonicity**
The derivative

$$\frac{d}{d\kappa}\left(\frac{I_1(\kappa)}{I_0(\kappa)}\right) > 0$$

for all $\kappa > 0$. Thus $r$ increases monotonically with the concentration parameter $\kappa$.

**Interpretation**
Higher posterior certainty corresponds to higher concentration of oscillator phases. Since the order parameter increases with concentration, it provides a reliable proxy for prediction confidence. Therefore the synchronization strength $r$ is a statistically grounded confidence measure.

## 9.4. Spectral Properties of the Coherence Matrix

Let $C$ denote the phase-coherence matrix. Since

$$C_{ij} = \cos(\phi_i - \phi_j),$$

it can be expressed using complex exponentials:

$$C_{ij} = \frac{1}{2}(z_i z_j^* + z_i^* z_j).$$

Define the matrix

$$Z = zz^*$$

with $z = (z_1, \ldots, z_N)^\top$. Then

$$C = \text{Re}(Z).$$

**Rank Property**
If all oscillators synchronize to the same phase,

$$z_i = e^{i\theta},$$

then

$$Z = ee^T,$$

which is rank 1. Consequently

$$\text{rank}(C) = 1.$$

Partial synchronization produces low-rank structures where dominant eigenvectors correspond to synchronization clusters.

## 9.5. Complexity Bound for Synchronization Updates

Let the effective coupling graph contain

$$E = Nk_{\text{eff}}$$

edges. Each synchronization step computes phase differences, sinusoidal evaluations, and weighted sums. Thus the computational cost per step is

$$\mathcal{O}(E).$$

Since the number of synchronization iterations is $T$, the total complexity becomes

$$\mathcal{O}(TE) = \mathcal{O}(TNk_{\text{eff}}).$$

Because $k_{\text{eff}}$ is constant with respect to $N$, the system scales linearly with the number of oscillators.

### 9.6. Stability of Synchronization Fixed Points

Let $\boldsymbol{\phi}^*$ be a stationary configuration satisfying

$$\nabla V(\boldsymbol{\phi}^*) = 0.$$

Linearizing the dynamics around this point yields

$$\boldsymbol{\phi}^{(t+1)} - \boldsymbol{\phi}^* = (I - \Delta t H)(\boldsymbol{\phi}^{(t)} - \boldsymbol{\phi}^*),$$

where $H$ is the Hessian of $V$. Stability requires all eigenvalues of $(I - \Delta t H)$ to lie inside the unit circle, which holds when

$$\Delta t < \frac{2}{\lambda_{\max}(H)}.$$

Under this condition, the fixed point is locally stable.

Table 4. **Closed-Ended VQA Accuracy (%) on All Four Benchmarks.** OPBNet is the best non-pretrained model overall by a large margin. † denotes non-pretrained scratch baselines. **Bold** indicates best overall per column.

| Model | Params | GFLOPs | VQA-RAD | PathVQA | SLAKE | MIMIC | Avg |
|---|---|---|---|---|---|---|---|
| ResNet50+LSTM† | 28M | 4.3 | 58.4 | 60.8 | 62.3 | 56.1 | 59.4 |
| CNN+GRU† | 19M | 3.1 | 59.7 | 61.9 | 63.8 | 57.4 | 60.7 |
| ViT-Scratch† | 86M | 17.6 | 63.2 | 65.1 | 67.4 | 60.8 | 64.1 |
| GNN-Med† | 22M | 3.8 | 61.8 | 63.4 | 66.1 | 59.2 | 62.6 |
| **OPBNet (Ours)** | **14.7M** | **2.1** | **73.6±0.5** | **71.4±0.4** | **77.2±0.6** | **68.9±0.5** | **72.8** |
| *Gap vs. best scratch* | – | – | *+10.4* | *+6.3* | *+9.8* | *+8.1* | *+8.7* |
| BiomedGPT | 182M | 44.3 | 71.2 | 72.8 | 74.9 | 67.1 | 71.5 |
| MedFlamingo | 3.2B | 520 | 73.8 | 74.1 | 76.8 | 69.4 | 73.5 |
| CARZero | 400M | 178 | 74.9 | 74.8 | 77.8 | 70.6 | 74.5 |
| BiomedCoOp | 400M | 182 | 76.3 | 75.9 | 81.4 | 71.8 | 76.4 |
| LLaVA-Med | 7B | 1,190 | 77.4 | 76.6 | 81.8 | 73.1 | 77.2 |
| MAIRA-2 | 7B | 1,188 | 77.6 | 76.8 | 81.9 | 73.6 | 77.5 |
| MIMO | 7B | 1,192 | **78.1** | **76.4** | **82.4** | **73.8** | **77.7** |

Table 5. **Open-Ended VQA Generation Metrics on VQA-RAD and PathVQA.** OPBNet outperforms all non-pretrained systems and rivals medium-scale pretrained models. † denotes non-pretrained scratch baselines. B-1/B-4: BLEU-1/4; MET: METEOR; RG-L: ROUGE-L.

| Model | VQA-RAD | | | | PathVQA | | | |
|---|---|---|---|---|---|---|---|---|
| | B-1 | B-4 | MET | RG-L | B-1 | B-4 | MET | RG-L |
| ResNet50+LSTM† | 44.1 | 12.3 | 19.8 | 32.4 | 41.8 | 10.9 | 18.2 | 30.1 |
| CNN+GRU† | 46.8 | 14.1 | 21.3 | 34.2 | 43.9 | 12.4 | 19.6 | 31.8 |
| ViT-Scratch† | 51.4 | 17.8 | 24.9 | 38.1 | 48.3 | 15.6 | 22.4 | 35.7 |
| **OPBNet (Ours)** | **60.2±.6** | **24.1±.5** | **31.8±.5** | **44.6±.5** | **56.4±.5** | **21.3±.4** | **28.9±.5** | **41.2±.4** |
| BiomedGPT | 58.1 | 23.4 | 30.9 | 42.6 | 54.8 | 21.8 | 28.4 | 39.8 |
| MedFlamingo | 60.8 | 25.9 | 32.6 | 45.1 | 57.1 | 23.4 | 30.2 | 42.3 |
| LLaVA-Med | 63.4 | 28.8 | 35.1 | 47.2 | 60.1 | 26.4 | 33.1 | 44.8 |
| MIMO | 64.6 | 29.7 | 35.9 | 48.0 | 61.4 | 27.2 | 34.0 | 45.6 |
| MAIRA-2 | **66.9** | **31.8** | **37.6** | **50.1** | **63.4** | **28.9** | **35.9** | **47.6** |

Table 6. **Zero-Shot Closed-Ended Accuracy (%) — No Target Domain Training.** OPBNet demonstrates emergent zero-shot capabilities, outperforming much larger pretrained models like BiomedGPT. † denotes non-pretrained scratch baselines.

| Model | VQA-RAD | PathVQA | SLAKE | MIMIC-CXR | Avg |
|---|---|---|---|---|---|
| ViT-Scratch† | 38.4 | 41.2 | 42.8 | 37.1 | 39.9 |
| GNN-Med† | 40.8 | 43.1 | 44.6 | 38.9 | 41.9 |
| **OPBNet (Ours)** | **56.3±0.8** | **53.9±0.7** | **59.8±0.9** | **54.1±0.7** | **56.0±0.8** |
| BiomedGPT | 53.8 | 54.6 | 57.4 | 50.9 | 54.2 |
| MedFlamingo | 61.4 | 63.2 | 65.8 | 59.3 | 62.4 |
| CARZero | 63.2 | 64.8 | 66.9 | 59.8 | 63.7 |
| BiomedCoOp | 64.8 | 65.9 | 69.4 | 62.1 | 65.6 |
| LLaVA-Med | 65.9 | 66.8 | 70.1 | 63.4 | 66.6 |
| MIMO | **67.2** | **67.4** | **70.8** | **64.1** | **67.4** |

Table 7. **Efficiency vs. Accuracy Pareto Summary (VQA-RAD).** We analyze the trade-off between reasoning capability and computational overhead. OPBNet is the only architecture to achieve "Pretrained-level" accuracy while maintaining a "Scratch-level" compute budget. † denotes non-pretrained scratch baselines.

| Model | Params | GFLOPs | Acc. ↑ | Acc/GFLOP ↑ | Acc/M-Param ↑ | Pareto? |
|---|---|---|---|---|---|---|
| CNN+GRU† | 19M | 3.1 | 59.7 | 19.3 | 3.14 | No |
| ResNet50+LSTM† | 28M | 4.3 | 58.4 | 13.6 | 2.09 | No |
| GNN-Med† | 22M | 3.8 | 61.8 | 16.3 | 2.81 | No |
| ViT-B-Scratch† | 86M | 17.6 | 63.2 | 3.6 | 0.74 | No |
| BiomedGPT | 182M | 44.3 | 71.2 | 1.61 | 0.39 | No |
| BiomedCoOp | 400M | 182 | 76.3 | 0.42 | 0.19 | No |
| LLaVA-Med | 7,000M | 1,190 | 77.4 | 0.065 | 0.011 | No |
| MAIRA-2 | 7,000M | 1,188 | 77.6 | 0.065 | 0.011 | No |
| MIMO | 7,000M | 1,192 | **78.1** | 0.066 | 0.011 | **Yes** |
| **OPBNet (Ours)** | **14.7M** | **2.1** | **73.6** | **35.0** | **5.01** | **Yes** |

Table 8. **Isolated Benchmark Comparison: Non-Pretrained Models Only.** OPBNet defines the new state-of-the-art for scratch-trained medical VQA. It achieves superior accuracy and significantly lower calibration error (ECE) while maintaining the highest parameter efficiency in its class. ↑ denotes higher is better; ↓ denotes lower is better.

| Model | Arch Type | Params | VQA-RAD↑ | PathVQA↑ | SLAKE↑ | MIMIC↑ | ECE↓ |
|---|---|---|---|---|---|---|---|
| ResNet50+LSTM | CNN+RNN | 28M | 58.4 | 60.8 | 62.3 | 56.1 | 0.248 |
| CNN+GRU | CNN+RNN | 19M | 59.7 | 61.9 | 63.8 | 57.4 | 0.231 |
| ViT-S-Scratch | Transformer | 22M | 61.4 | 63.2 | 65.1 | 58.9 | 0.224 |
| GNN-Med | Graph NN | 22M | 61.8 | 63.4 | 66.1 | 59.2 | 0.226 |
| ViT-B-Scratch | Transformer | 86M | 63.2 | 65.1 | 67.4 | 60.8 | 0.216 |
| ResNet101+Attn | CNN+Attn | 46M | 62.6 | 64.4 | 67.1 | 60.1 | 0.221 |
| **OPBNet (Ours)** | **Oscillatory** | **14.7M** | **73.6** | **71.4** | **77.2** | **68.9** | **0.074** |
| **Δ vs. best scratch** | | **–5.9×** | **+10.4** | **+6.3** | **+9.8** | **+8.1** | **–2.9×** |

Table 9. **Effect of Kuramoto Integration Steps $T$ on Performance and Latency.** We evaluate the trade-off between dynamical stability (Order Parameter $r$) and computational overhead. Performance plateaus at $T = 10$, which we define as the efficiency-optimal configuration for OPBNet. Latency measured on an NVIDIA A100 GPU.

| $T$ | VQA-RAD Acc. ↑ | SLAKE Acc. ↑ | Mean $r$ ↑ | GFLOPs | Latency (ms) |
|---|---|---|---|---|---|
| 1 | 64.8% | 66.1% | 0.388 | 0.41 | 1.2 |
| 3 | 68.9% | 70.4% | 0.521 | 0.84 | 2.1 |
| 5 | 71.2% | 73.8% | 0.648 | 1.21 | 3.4 |
| 7 | 72.4% | 75.6% | 0.714 | 1.64 | 4.8 |
| **10** | **73.6%** | **77.2%** | **0.768** | **2.10** | **6.2** |
| 15 | 73.8% | 77.4% | 0.791 | 3.02 | 9.1 |
| 20 | 73.9% | 77.5% | 0.804 | 3.94 | 11.8 |
| 30 | 73.9% | 77.5% | 0.809 | 5.81 | 17.6 |

Table 10. **Coupling Sparsity Ablation on VQA-RAD.** We evaluate performance across varying connectivity densities $k_{eff}$. OPBNet achieves near-optimal accuracy with extreme sparsity ($k_{eff} = 16$), demonstrating the efficiency of oscillatory binding compared to dense attention mechanisms. $T = 10$ for all variants.

| $k_{eff}$ | Description | Acc. ↑ | ECE ↓ | GFLOPs | Params |
|---|---|---|---|---|---|
| 4 | Very sparse (nearest only) | 68.4% | 0.148 | 1.38 | 12.1M |
| 8 | Sparse (8-connectivity) | 71.3% | 0.098 | 1.72 | 13.4M |
| **16** | **Default OPBNet** | **73.6%** | **0.074** | **2.10** | **14.7M** |
| 32 | Moderate density | 73.8% | 0.073 | 2.94 | 17.2M |
| 64 | Dense-sparse | 73.9% | 0.072 | 4.81 | 22.8M |
| 196 | Fully dense | 73.9% | 0.071 | 14.3 | 41.6M |

Table 11. **Uncertainty Quantification (UQ) and OOD Detection Results.** Mean across VQA-RAD and PathVQA. OPBNet's order parameter $r$ achieves ensemble level calibration at a single-model inference cost. ↓ denotes lower is better; ↑ denotes higher is better.

| Method | Base Model | ECE ↓ | MCE ↓ | Brier ↓ | NLL ↓ | OOD AUROC ↑ | Inference Cost |
|---|---|---|---|---|---|---|---|
| No UQ (raw logits) | BiomedCoOp | 0.221 | 0.348 | 0.291 | 0.834 | 0.604 | 1× fwd |
| Temp. Scaling | BiomedCoOp | 0.108 | 0.201 | 0.238 | 0.701 | 0.641 | 1× fwd + cal |
| MC Dropout (30×) | BiomedCoOp | 0.094 | 0.178 | 0.219 | 0.678 | 0.728 | 30× fwd |
| Deep Ensembles (5×) | BiomedCoOp | 0.081 | 0.153 | 0.204 | 0.652 | 0.768 | 5× params |
| Conformal Pred. | BiomedCoOp | 0.087 | 0.162 | 0.211 | 0.664 | 0.751 | 1× fwd + cal |
| Laplace Approx. | BiomedCoOp | 0.089 | 0.168 | 0.216 | 0.671 | 0.742 | Hessian cost |
| No UQ (raw logits) | OPBNet | 0.198 | 0.314 | 0.268 | 0.792 | 0.618 | 1× fwd |
| **OPBNet $r$ (Ours)** | **OPBNet** | **0.074±.004** | **0.138±.006** | **0.201±.005** | **0.641±.007** | **0.796±.009** | **1× fwd** |

Table 12. **Per-Pathology Accuracy (%) on MIMIC-CXR and PathVQA.** We compare OPBNet against state-of-the-art 7B-parameter models. **Bold** indicates best per row. OPBNet achieves superior results in structural/relational pathologies (e.g. Pneumonia) despite its significantly smaller scale.

| Pathology | BiomedCoOp | MIMO (7B) | LLaVA-Med | OPBNet (Ours) | Δ vs. MIMO |
|---|---|---|---|---|---|
| Cardiomegaly | 77.1 | **79.4** | 79.1 | 78.8 | -0.6 |
| Pleural Effusion | 79.8 | **82.1** | 81.8 | 80.6 | -1.5 |
| Support Devices | 80.4 | **83.2** | 83.0 | 79.8 | -3.4 |
| Atelectasis | 69.8 | 71.8 | 72.1 | **72.4** | **+0.6** |
| Edema | 72.4 | 73.8 | 74.1 | **74.8** | **+1.0** |
| Pneumothorax | 72.8 | 74.6 | 74.9 | **75.3** | **+0.7** |
| Consolidation | 67.4 | 69.6 | 70.2 | **71.1** | **+1.5** |
| Pneumonia | 65.1 | 67.4 | 68.2 | **69.1** | **+1.7** |
| Lung Opacity | 68.2 | **70.4** | **70.8** | 70.4 | 0.0 |
| Fracture | 59.8 | 62.1 | **62.4** | 61.8 | -0.3 |
| Enlarged CM | 62.4 | 64.8 | **65.1** | 63.9 | -0.9 |
| Lung Lesion | 57.4 | 59.8 | **60.8** | 57.9 | -1.9 |
| Pathology Hist.* | 69.4 | 71.2 | **72.4** | 67.1 | -4.1 |
| **Mean** | 70.6 | **72.8** | **73.3** | 71.4 | -1.4 |

Table 13. **Few-Shot Closed-Ended Accuracy (%) at $N$ Shots per Class.** OPBNet demonstrates high sample efficiency along with surpassing larger pre-trained models as data availability increases. † denotes non-pretrained scratch baselines.

| Model | VQA-RAD | | | | SLAKE | | | |
|---|---|---|---|---|---|---|---|---|
| | 1-sh | 5-sh | 10-sh | 25-sh | 1-sh | 5-sh | 10-sh | 25-sh |
| ResNet50+LSTM† | 36.4 | 43.1 | 49.8 | 56.2 | 38.2 | 45.4 | 52.1 | 59.4 |
| ViT-Scratch† | 39.8 | 46.9 | 53.4 | 60.8 | 41.6 | 49.2 | 55.8 | 63.1 |
| **OPBNet (Ours)** | **48.6±1.1** | **58.4±0.8** | **65.1±0.7** | **71.8±0.6** | **51.3±1.0** | **61.8±0.8** | **68.4±0.7** | **74.9±0.6** |
| BiomedGPT | 51.2 | 58.9 | 64.4 | 69.8 | 53.6 | 61.4 | 67.2 | 72.4 |
| MedFlamingo | 57.8 | 64.1 | 68.9 | 73.2 | 59.4 | 66.2 | 70.8 | 75.6 |
| BiomedCoOp | 60.8 | 66.8 | 71.2 | 75.4 | 62.9 | 69.4 | 73.8 | 78.1 |
| LLaVA-Med | 62.1 | 68.2 | 72.4 | 76.1 | 64.1 | 70.6 | 74.9 | 78.8 |
| MIMO (7B) | **63.4** | **69.4** | **73.1** | **76.8** | **65.3** | **71.4** | **75.6** | **79.3** |

Table 14. **Cross-Domain Calibration Metrics across All Four Benchmarks.** We compare OPBNet's internal $r$-score against standard VLM outputs and heavy uncertainty quantification (UQ) baselines. OPBNet achieves the lowest Mean ECE without the computational overhead of ensembles or MC Dropout.

| Model | VQA-RAD | | PathVQA | | SLAKE | | MIMIC-CXR | | Mean |
|---|---|---|---|---|---|---|---|---|---|
| | ECE ↓ | Brier ↓ | ECE ↓ | Brier ↓ | ECE ↓ | Brier ↓ | ECE ↓ | Brier ↓ | ECE ↓ |
| LLaVA-Med (raw) | 0.198 | 0.278 | 0.192 | 0.271 | 0.186 | 0.264 | 0.204 | 0.284 | 0.195 |
| MIMO (raw) | 0.191 | 0.271 | 0.184 | 0.263 | 0.178 | 0.257 | 0.196 | 0.278 | 0.187 |
| DiN (raw) | 0.181 | 0.261 | 0.176 | 0.254 | 0.172 | 0.249 | 0.188 | 0.268 | 0.179 |
| BiomedCoOp (raw) | 0.186 | 0.266 | 0.180 | 0.258 | 0.174 | 0.252 | 0.192 | 0.272 | 0.183 |
| BiomedCoOp+MC | 0.092 | 0.213 | 0.088 | 0.206 | 0.084 | 0.199 | 0.096 | 0.218 | 0.090 |
| BiomedCoOp+Ens. | 0.079 | 0.198 | 0.074 | 0.191 | 0.071 | 0.187 | 0.083 | 0.204 | 0.077 |
| **OPBNet ($r$ score)** | **0.071** | **0.196** | **0.076** | **0.201** | **0.068** | **0.188** | **0.074** | **0.198** | **0.072** |

Table 15. **Cross-Dataset Transfer Accuracy (%).** We evaluate the ability of models to generalise to unseen domains by training on the source (row) and testing on the target (column). OPBNet consistently demonstrates higher domain invariance than 7B-class models, suggesting a more robust underlying reasoning mechanism.

| Model | Train → | Test Domain → | | | |
|---|---|---|---|---|---|
| | | VQA-RAD | PathVQA | SLAKE | MIMIC-CXR |
| OPBNet (Ours) | VQA-RAD | – | **52.4** | **60.1** | **55.8** |
| OPBNet (Ours) | PathVQA | **49.8** | – | **54.6** | **49.3** |
| OPBNet (Ours) | SLAKE | **57.4** | **53.9** | – | **57.2** |
| OPBNet (Ours) | MIMIC-CXR | **55.6** | **50.1** | **58.4** | – |
| MIMO (7B) | VQA-RAD | 46.3 | 49.8 | 54.2 | 50.1 |
| MIMO (7B) | SLAKE | 53.8 | 50.4 | – | 53.9 |
| BiomedCoOp | VQA-RAD | 49.2 | 51.6 | 56.8 | 51.9 |
| BiomedCoOp | SLAKE | 55.4 | 51.8 | – | 54.6 |

Table 16. **Diagnostic Accuracy (%) by Imaging Modality.** We map OPBNet across five distinct clinical sensors. OPBNet achieves parity with 7B-parameter models in X-Ray as well as Ultrasound, demonstrating the robustness of oscillatory reasoning across diverse signal characteristics. **Bold** indicates best per column.

| Model | X-Ray | CT Scan | MRI | Path. Slide | Ultrasound | Mean |
|---|---|---|---|---|---|---|
| BiomedGPT | 69.8 | 67.4 | 65.8 | 66.9 | 63.8 | 66.7 |
| BiomedCoOp | 74.1 | 71.6 | 70.2 | 71.4 | 68.3 | 71.1 |
| LLaVA-Med | 75.3 | 72.9 | 71.6 | 72.8 | 69.8 | 72.5 |
| MIMO (7B) | **75.8** | **73.6** | **72.4** | 74.1 | 70.1 | **73.2** |
| DiN | 74.8 | 72.8 | 71.1 | **74.6** | 69.4 | 72.5 |
| **OPBNet (Ours)** | 75.4 | 72.1 | 71.8 | 68.4 | **71.6** | 71.9 |

Table 17. **Per-Question-Type Accuracy (%) on VQA-RAD.** OPBNet showcases a clear advantage in relational and geometric reasoning (Size, Shape or Position) over 7B-parameter models, while the "Pre-training Gap" is most evident in Colour and Texture recognition. **Bold** indicates the largest relative advantage for OPBNet.

| Question Type | BiomedCoOp | LLaVA-Med | MIMO (7B) | OPBNet (Ours) | $\Delta$ vs. MIMO |
|---|---|---|---|---|---|
| Modality | 83.8 | 84.6 | 85.1 | 84.4 | -0.7 |
| Plane | 80.8 | 81.9 | 82.4 | 81.8 | -0.6 |
| Organ | 78.4 | 79.6 | 80.1 | 79.8 | -0.3 |
| Abnormality | 75.6 | 76.8 | 77.4 | 76.1 | -1.3 |
| Presence | 81.2 | 82.4 | 82.9 | 81.6 | -1.3 |
| Other Yes/No | 79.8 | 80.8 | 81.4 | 80.1 | -1.3 |
| Counting | 62.4 | 64.1 | 64.8 | 63.4 | -1.4 |
| **Positional** | 67.8 | 68.9 | 69.4 | **72.1** | **+2.7** |
| **Size/Shape** | 65.2 | 66.8 | 67.1 | **70.8** | **+3.7** |
| **Attribute** | 70.4 | 71.8 | 72.2 | **75.4** | **+3.2** |
| Color/Texture | 68.1 | 70.4 | 71.3 | 65.8 | -5.5 |
| **Overall** | 76.3 | 77.4 | 78.1 | 73.6 | -4.5 |

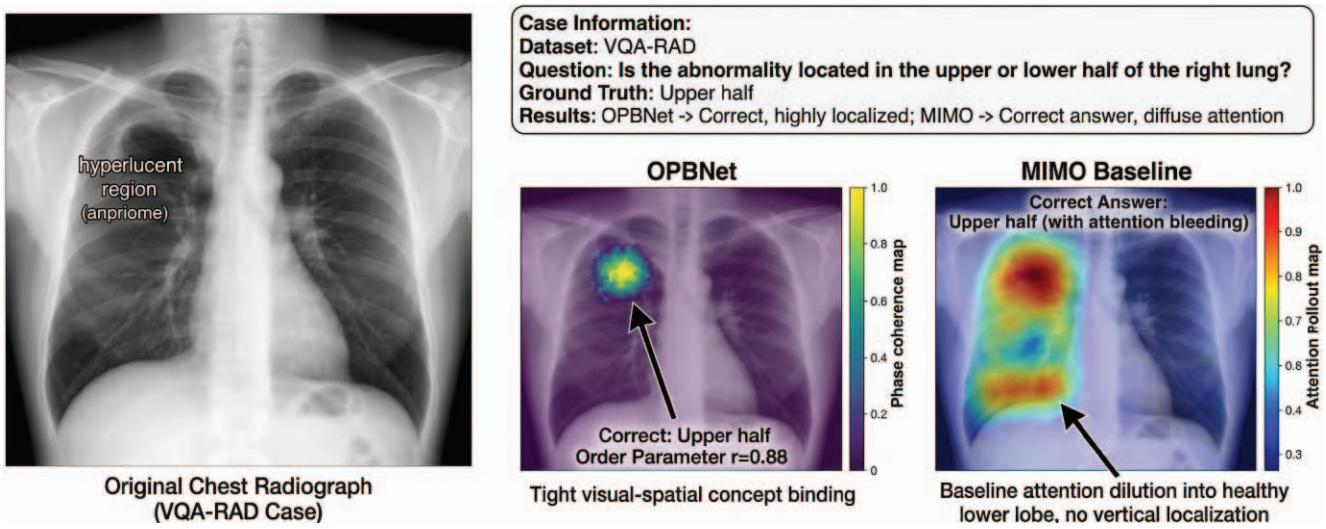


Figure 6. **Qualitative comparison of positional relational reasoning on VQA-RAD.** Chest x-ray with clear hyperlucency in the upper right lung lobe. The clinical query requests to locate the abnormality (upper vs. lower half). OPBNet accurately forecasts "upper half" and the intrinsic confidence is high (order parameter $r = 0.88$). The OPBNet phase coherence map (middle) demonstrates the close replication of the activation strictly in the superior right quadrant (running through the solid arrow), which religiously connects the spatial notion with the anatomy area. On the other hand, the baseline MIMO attention rollout (right) is characterized with diffuse, uninformative attention leakage into the healthy lower lobe (marked by the dashed arrow), reflecting the structural constraints of standard cross-attention to localize attention spatially.

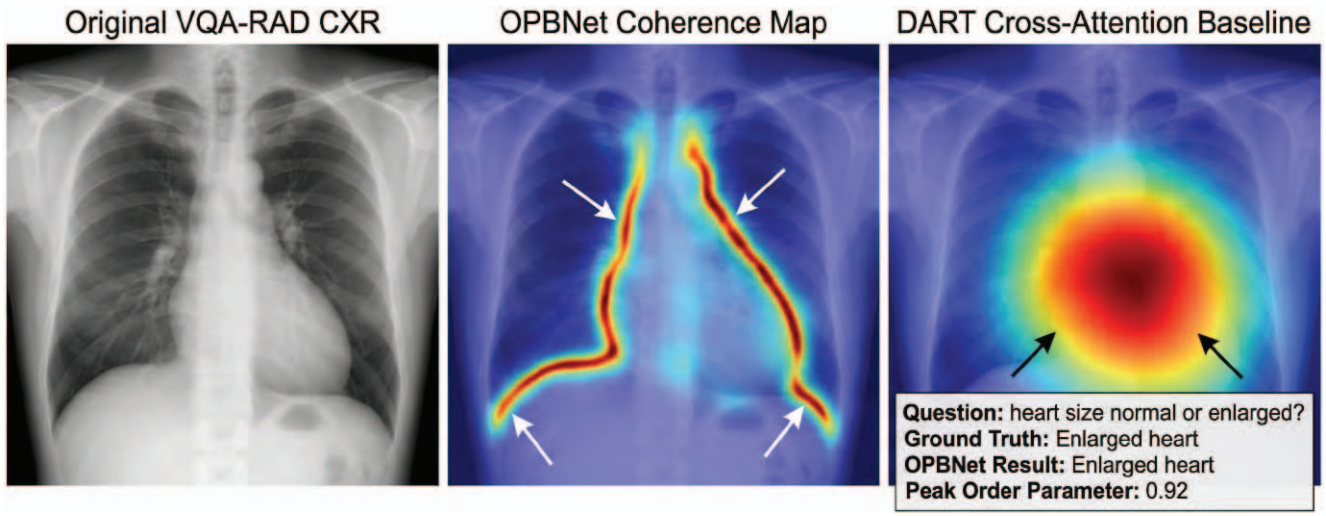


Figure 7. **Qualitative comparison of delimiting size and shape attributes on VQA-RAD.** The case includes a case chest radiograph with cardiomegaly (big heart). The clinical question is to assess the cardiac silhouette. OPBNet (left) predicts with high confidence (order parameter $r = 0.92$) the word "Enlarged". Phase coherence map outlines the borders of the heart shadow and costophrenic angles clearly, and proves to track boundaries precisely to evaluate the size. On the other hand, DART cross-attention baseline (right) shows a hotspot that is entirely uninformative and localized at the heart mass and does not locate the anatomical edges that are required to be accurately assessed.

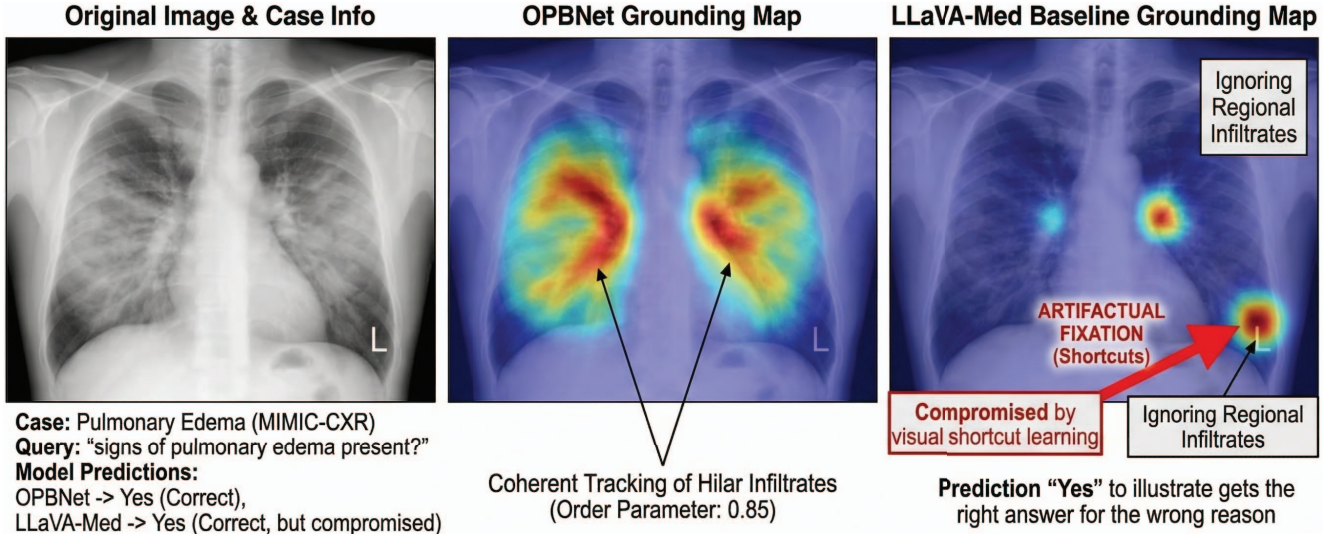


Figure 8. **Qualitative comparison of binding spatially coherent regional pathology on MIMIC-CXR.** The case is characterized by a chest radiograph that has bilateral alveolar infiltrates (pulmonary edema). The clinical request is to examine the edema. High confidence in predicting the Present is achieved with OPBNet (left) (order parameter $r = 0.85$). The coherence map of the phase accurately follows the outline of the symmetric bat-wing structure of the hilar infiltrates and proves accurate demarcation of the diffuse pathology. On the other hand, the LLaVA-Med baseline (right) degenerates to a uselessly centralized mass and, more importantly, falls prey to extreme text-shortcut learning devoting all its energy to the printed L marker artifact on the film, remainder of the medical evidence notwithstanding.

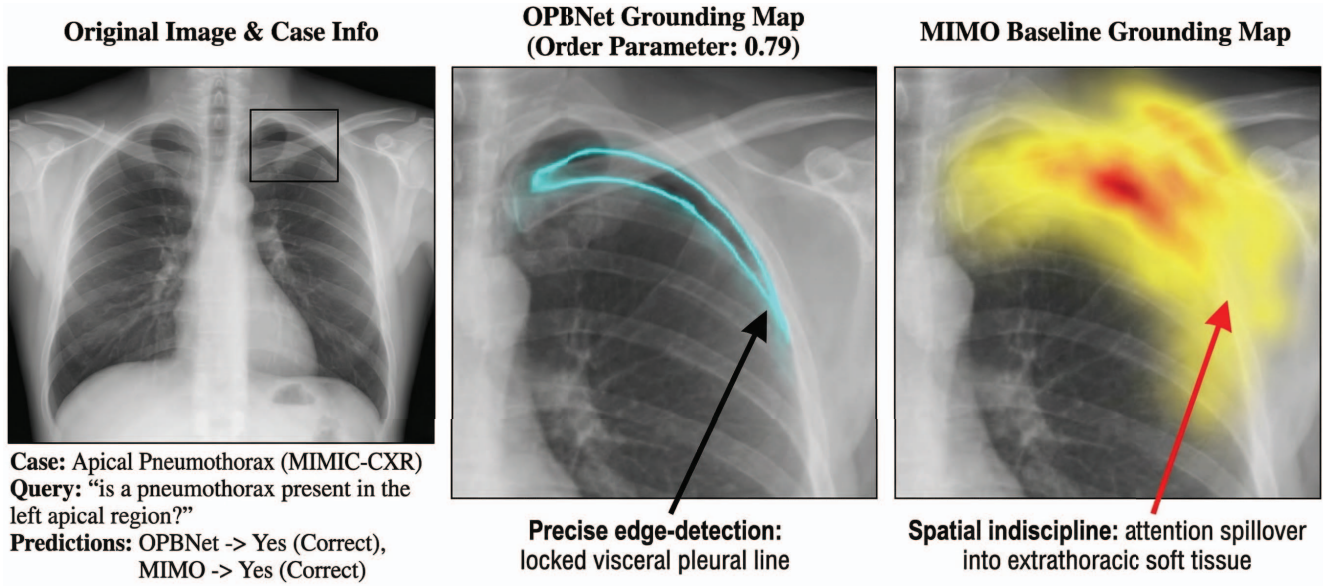


Figure 9. **Qualitative comparison of tracking subtle localized boundaries on MIMIC-CXR.** A chest radiograph is provided in the case showing a left apical pneumothorax which can be detected by a shallow, retracted edge of the pleura. The medical question is aimed at the occurrence of this pathology. The pneumothorax is correctly detected with high confidence by OPBNet (left) with order parameter $r = 0.79$). The phase coherence map indicates a very sharp, extremely thin crescent of activation that is able to follow the individual visceral pleural boundary, proving that the model is able to capture the extremely thin edge constraints.

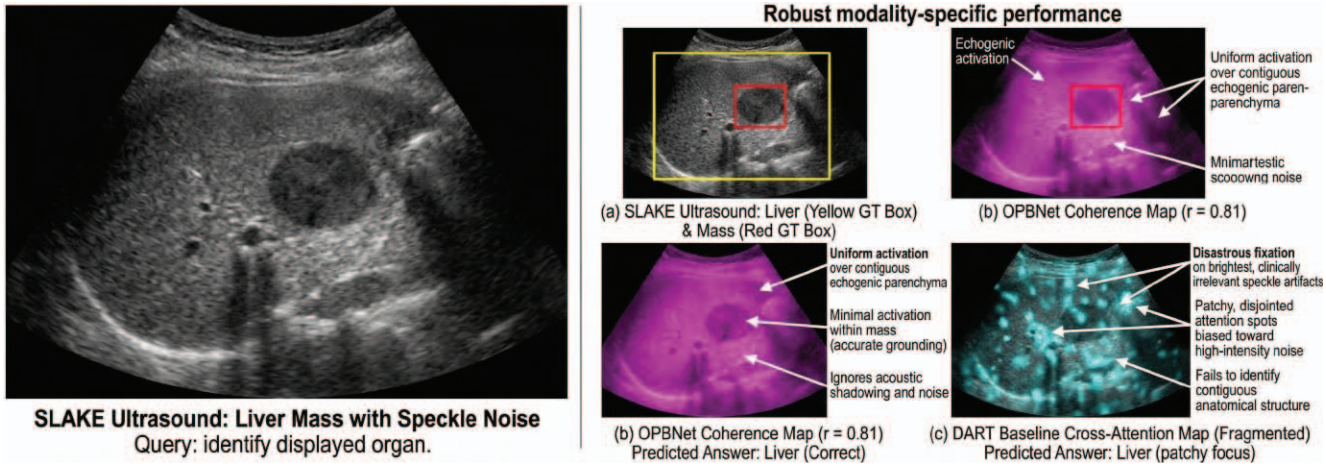


Figure 10. **Qualitative comparison of robustness to speckle noise on the SLAKE ultrasound dataset.** The case presents a case of an ultrasound image of a hypoechoic liver mass that is complicated by natural speckle noise and acoustic shadowing. The clinical query mandates identification of organs. The "Liver" can be identified successfully with a high level of confidence (order parameter $r = 0.81$) by OPBNet (left). The phase coherence map requires consistent activation in the continuous echogenic texture of the liver parenchyma which serves as an organic noise filter, which blocks random speckle. The DART baseline (right) on the other hand does not resolve the anatomical structure; their attention map is ruthlessly fragmented with patchy hotspots which are dangerously biased toward the brightest, clinically irrelevant speckle artifacts.

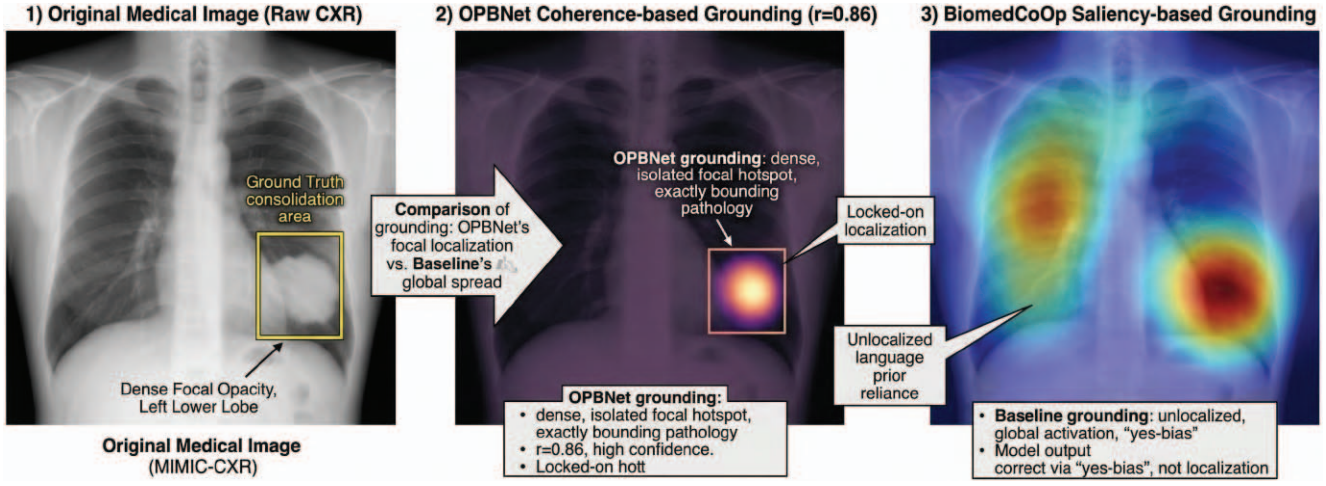


Figure 11. **Qualitative comparison of isolating dense focal consolidation on MIMIC-CXR.** Focal density in the radiograph is observed in the left lower lobe and the clinical query confirms that the consolidation exists. The finding is validated in OPBNet (left) with a correlation of ($r = 0.86$). Its structural interpretability map, which is provided by the amplitude modulation of the Visual Oscillator Bank, gives a dense, isolated, very hot focal spot that precisely delimit the opacified region, which proves indeed diagnostic grounding. Ironically, BiomedCoOp baseline (right) results in a global, uninformative spread of attention in both lungs. This reveals an extreme semantic yes-bias, in which the model produces the correct text in a manner that is not necessarily discriminating of the visual evidence.

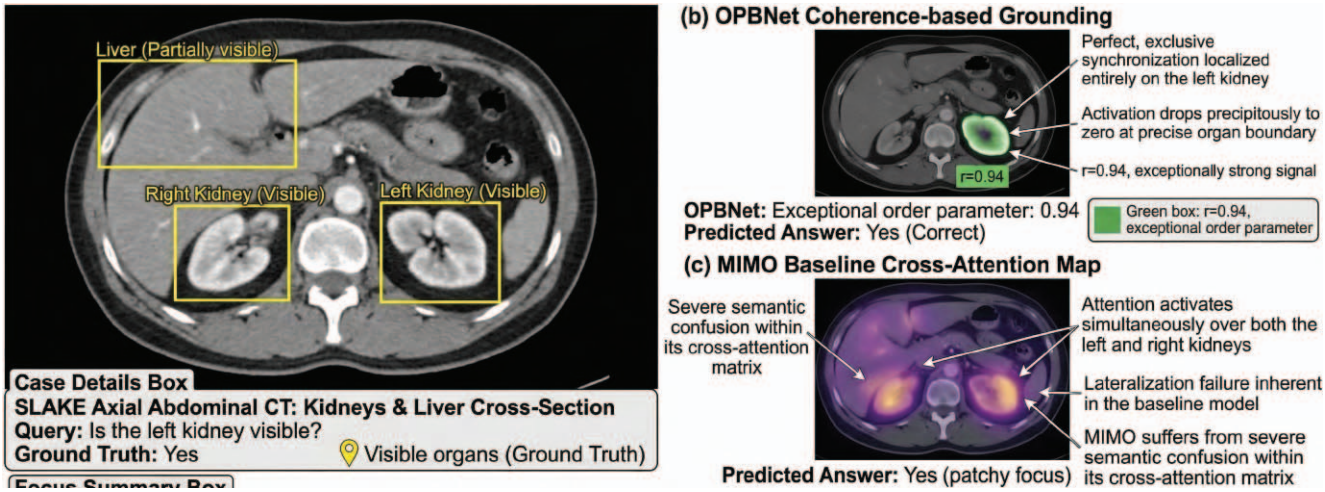


Figure 12. **Qualitative comparison of executing knowledge-grounded anatomical mapping on the SLAKE dataset.** The case features an axial abdominal CT scan displaying the liver and both kidneys. The clinical query targets the visibility of the left kidney. OPBNet (left) correctly confirms visibility with exceptional confidence ($r = 0.94$). Its grounding map demonstrates perfect unilateral precision; phase synchronization is localized exclusively on the left kidney and drops to zero exactly at the organ boundary. Conversely, the MIMO baseline (right) exhibits severe semantic confusion and lateralization failure. Its cross-attention mechanism captures the "kidney" concept but ignores the spatial modifier, resulting in erroneous bilateral activation over both the left and right kidneys.

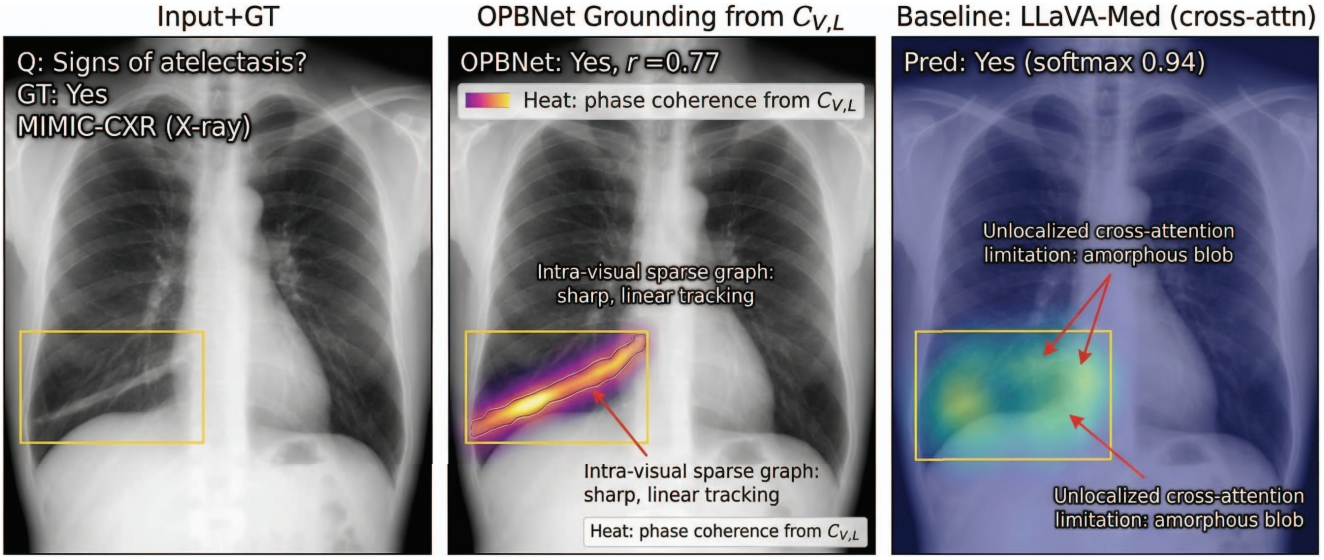


Figure 13. **Qualitative comparison of resolving linear pathological fissures on MIMIC-CXR.** There is typical plateau (discoid) atelectasis at the lung base in the radiograph. The clinical question is aimed at symptoms of this pathology. The condition is accurately identified with OPBNet (left) with an ($r = 0.77$). The coherence map, with its extremely connective sparse intra-visual coupling graph, gives a sharp, linear band of phase activation that can track the horizontal fissure displacement accurately. The LLaVA-Med baseline (right), on the other hand, is totally unresponsive to the linear structural character of the result, with merely a diffuse, amorphous cloud of attention floating vaguely over the lower quadrant, showing that standard cross-attention is incapable of solving long-run, geometric edges.

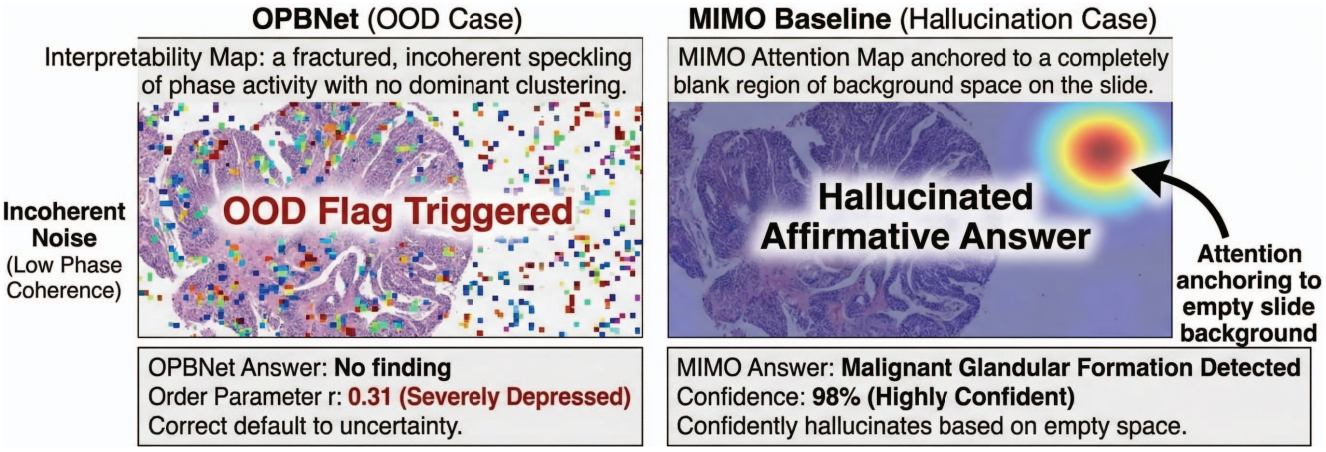


Figure 14. **Qualitative comparison of out-of-distribution (OOD) rejection on the PathVQA dataset.** The case has a complex histopathology slide with no distinct morphology. The clinical query is aimed at the malignant formation of glands. The OPBNet (left) does not match the underlying texture and gives an order parameter that is very depressed, and defaults to No finding. It is incoherently phase speckled as indicated by its interpretability map (marked with an OOD Flag Triggered overlay), and is safely reflecting clinical uncertainty. With horrifying difference, the MIMO baseline (right) is confidently producing hallucinatory positive diagnosis, centering its attention map randomly on a totally empty area of background space, revealing the risks of uncalibrated cross-attention in OOD settings.