### A1: MIMIC-CXR classification results split by disease label

| Disease Label                  | Pooch et al. [32] | Seyyed et al. [38] | Base model | +REFLACX Bounding boxes | +REFLACX eye gaze | +EGD-CXR eye gaze |
|-------------------------------|-------------------|-------------------|------------|-------------------------|------------------|------------------|
| No Finding                    | -                 | -                 | 0.815      | 0.819                   | 0.830            | 0.836            |
| Enlarged Cardiomegastinum     | -                 | -                 | 0.812      | **0.872**               | 0.759            | 0.867            |
| Cardiomegaly                  | -                 | -                 | 0.727      | 0.705                   | 0.729            | 0.791            |
| Lung Opacity                  | -                 | -                 | 0.735      | **0.746**               | 0.688            | 0.700            |
| Lung Lesion                   | -                 | -                 | 0.678      | 0.614                   | 0.712            | **0.732**        |
| Edema                         | -                 | -                 | 0.772      | 0.777                   | **0.822**        | 0.819            |
| Consolidation                 | -                 | -                 | 0.734      | 0.771                   | 0.785            | **0.799**        |
| Pneumonia                     | -                 | -                 | 0.594      | 0.572                   | 0.621            | **0.654**        |
| Atelectasis                   | -                 | -                 | 0.663      | **0.776**               | 0.761            | 0.704            |
| Pneumothorax                  | -                 | -                 | 0.681      | **0.699**               | 0.686            | 0.684            |
| Pleural effusion              | -                 | -                 | 0.888      | 0.851                   | 0.832            | **0.917**        |
| Pleural other                 | -                 | -                 | 0.704      | 0.797                   | 0.754            | **0.841**        |
| Fracture                      | -                 | -                 | 0.635      | 0.688                   | **0.756**        | 0.685            |
| Support device                | -                 | -                 | 0.842      | **0.906**               | 0.852            | 0.867            |
| **Average**                   | 0.828             | 0.834             | 0.807      | 0.821                   | 0.827            | **0.836**        |

Table 1: AUC scores of our proposed method on MIMIC-CXR with a DenseNet121 CNN backbone, split by disease label. Base model scores indicate the performance after the training stage I with the large MIMIC-CXR base dataset. We also show results on the same test set of the base dataset after integration of object level annotations subsets (REFLACX bounding boxes & eye gaze and EGD-CXR eye gaze) in training stage II.
A2: Chest X-ray14 classification results split by disease label

| Disease                  | Wang et al. [44] | Yao et al. [48] | Gi zendel et al. [7] | Kim et al. [20] | ViT [41] | Tsakl imi et al. [41] | Li et al. [25] Base model | Li et al. [25] +Bounding boxes | Base model +Bounding boxes |
|--------------------------|------------------|-----------------|----------------------|-----------------|---------|-----------------------|---------------------------|---------------------------|---------------------------|
| Cardiomegaly             | 0.810            | 0.856           | 0.883                | 0.891           | 0.875   | 0.881                 | 0.81                      | 0.87                      | 0.825                     | 0.874                    |
| Edema                    | 0.805            | 0.806           | 0.835                | 0.842           | 0.848   | 0.854                 | 0.81                      | 0.88                      | 0.752                     | 0.792                    |
| Consolidation            | 0.703            | 0.711           | 0.745                | 0.754           | 0.748   | 0.747                 | 0.70                      | 0.88                      | 0.638                     | 0.737                    |
| Pneumonia                | 0.658            | 0.684           | 0.731                | 0.665           | 0.713   | 0.730                 | 0.66                      | 0.67                      | 0.662                     | 0.724                    |
| Atelectasis              | 0.708            | 0.733           | 0.767                | 0.741           | 0.781   | 0.782                 | 0.70                      | 0.88                      | 0.691                     | 0.740                    |
| Pneumothorax             | 0.799            | 0.805           | 0.846                | 0.838           | 0.871   | 0.874                 | 0.80                      | 0.87                      | 0.827                     | 0.861                    |
| Infiltration             | 0.661            | 0.673           | 0.709                | 0.687           | 0.701   | 0.715                 | 0.66                      | 0.70                      | 0.617                     | 0.718                    |
| Emphysema                | 0.833            | 0.842           | 0.895                | 0.832           | 0.914   | 0.956                 | 0.83                      | 0.91                      | 0.836                     | 0.938                    |
| Fibrosis                 | 0.786            | 0.743           | 0.818                | 0.787           | 0.826   | 0.815                 | 0.78                      | 0.79                      | 0.807                     | 0.812                    |
| Pleural thickening       | 0.684            | 0.724           | 0.761                | 0.755           | 0.778   | 0.798                 | 0.68                      | 0.79                      | 0.751                     | 0.764                    |
| Nodule                   | 0.669            | 0.724           | 0.758                | 0.753           | 0.799   | 0.879                 | 0.67                      | 0.75                      | 0.661                     | 0.771                    |
| Mass                     | 0.693            | 0.777           | 0.821                | 0.788           | 0.822   | 0.834                 | 0.69                      | 0.83                      | 0.700                     | 0.767                    |
| Hernia                   | 0.672            | 0.775           | 0.596                | 0.667           | 0.655   | 0.896                 | 0.71                      | 0.87                      | 0.781                     | 0.882                    |
| Effusion                 | 0.759            | 0.806           | 0.828                | 0.813           | 0.824   | 0.836                 | 0.76                      | 0.87                      | 0.767                     | 0.834                    |
| Average                  | 0.745            | 0.761           | 0.807                | 0.779           | 0.810   | 0.820                 | 0.746                     | 0.797                     | 0.772                     | 0.809                    |

Table 2: AUC scores of our proposed method on Chest X-ray14 and comparison to prior works, split by disease label. Our method has a DenseNet121 CNN backbone. Base model scores indicate the performance after the training stage I with the large Chest X-ray 14 base dataset. We also show results on the same test set of the base dataset after integration of an object level annotation subset (bounding boxes) in training stage II.