

Supplementary material of MICCAI submission 1356

0.1 Supplementary Datasets details

We show a summary of datasets details in Table 1.

Table 1: Datasets details for our experiment

Datasets	Modailties	Train	Test
RSNA	CXR	3851	2000
VinDr-CXR	CXR	4000	2000
CXAD	CXR	2000	1000
Brain MRI	MRI	1000	1200
LAG	Retinal fundus	1500	1622

0.2 Supplementary Experiments

Hyperparameter λ . We investigate the impact of λ , and present findings in Fig. 1(a).

Layer to set the repulsion. We analyze the impact of the layer l to set the repulsion, and show results in Fig. 1(b).

AS histograms. We employ Anomaly Score (AS) histograms to graphically illustrate the distribution of anomaly scores throughout the dataset as illustrated in Fig. 2.

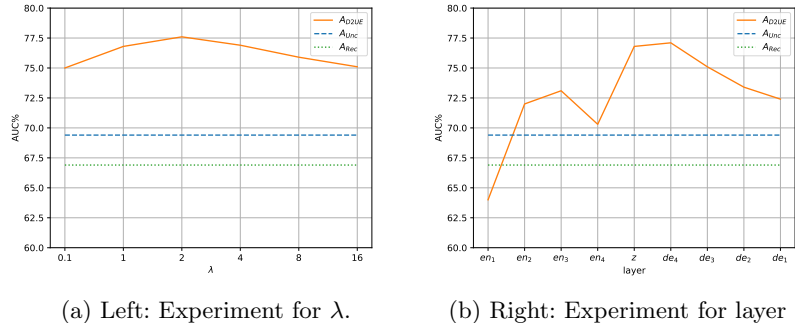


Fig. 1: Experiment results for λ and layer l in RAR. The vertical axis is AUC %. A_{D2UE} , A_{Unc} , and A_{Rec} denote D2UE, output space uncertainty estimation, and ensemble reconstruction. To ensure a fair comparison, we employ the vanilla Autoencoder on RSNA and the DSU is excluded. **Left:** The experiment reveals that the performance enhancement resulting from RAR is not susceptible to variations in λ . **Right:** Input X is encoded sequentially according to the horizontal axis direction. Except for encoder’s initial feature layer, repulsion across all other layers bolsters accuracy. This can likely be ascribed to the first layer of encoders predominantly capturing elementary visual characteristics of an image. Should these generalized features diverge during training, learners may struggle to reconstruct normal samples effectively.

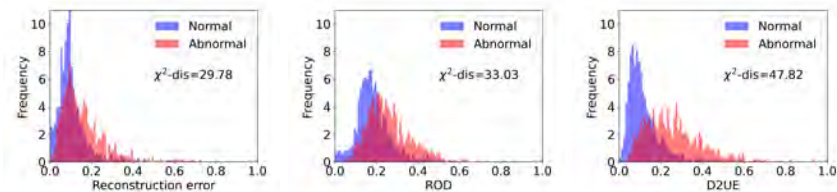


Fig. 2: Histograms of anomaly scores for normal (blue) and abnormal (red) samples in RSNA dataset. Scores are normalized to $[0,1]$. Our evaluation reveals a diminished overlap between normal and aberrant scores in D2UE. This observation implies a reduction in indiscernible samples within our methodology. Furthermore, we incorporate the χ^2 test to gauge the disparity between normal and abnormal scores. Our method exhibits a more considerable distance (47.82) compared to others, indicating an enhanced capacity to distinguish between normal and abnormal samples.