Boosting Vision Semantic Density with Anatomy Normality Modeling for Medical Vision-language Pre-training

Supplementary Material

6. More ablation studies

6.1. Variety in visual encoder selection

In the 3D CT VLP task, we discover that the CNN visual encoder outperforms the ViT. Consequently, we explore the impact of various CNN backbones on model performance. As illustrated in Table 6, both ResNet34 and ResNet50 demonstrate improved performance compared to ResNet18. However, considering the balance between computational cost and performance, we decide to utilize ResNet18 as the primary visual encoder in this study.

6.2. Different initializations for visual encoders

Aligned with Figure 3, Table 7 provides a numerical comparison of different initialization methods for visual encoder. The table clearly shows that the model initialized with weights derived from our proposed disease-level contrastive learning method achieves the highest AUC, outperforming the other two initialization approaches. These quantitative results further underscore the effectiveness of the proposed visual semantic enhancement.

6.3. Experiments on local and diffuse diseases

We assessed the improvement offered by the proposed model over the baseline model in diagnosing both local and diffuse diseases. A radiologist categorizes these abnormalities into local and diffuse diseases, as listed in Table 9. Detailed performances are presented in Table 8. As indicated in the table, there is a 4.0% increase in the AUC for local diseases, which surpasses the 2.8% improvement seen in diffuse diseases. This suggests that our approach significantly improves the model's ability to diagnose localized diseases.

7. More implementation details

For the MedVL-CT69K dataset, we utilize the pre-trained BERT-base [11] as the text encoder. Our ViSD-boost is trained with the Adam optimizer, where the learning rate increases linearly to 1e-4 in the first epoch and then decreases gradually to 1e-6 via a cosine decay scheduler. The model is trained over four phases for 60, 30, 60, and 30 epochs, utilizing 4 A100 GPUs and a batch size of 48. During training, we dynamically apply RandomCrop and RandomFlip augmentations. For the chest CT-RATE dataset, we employ the same image pre-processing methodology as CT-CLIP [14] to ensure a fair comparison with other methods. We also use the same CXR-Bert as the text encoder [14]. Furthermore, in line with the fVLM [35], we adopt the same anatomy and report parsing methods, facilitating anatomy-wise image-report alignment.

Methods	SE	SP	ACC	AUC
ResNet18	73.6	75.9	74.8	78.7
ResNet34	74.8	76.1	75.5	78.9
ResNet50	76.0	75.3	75.7	79.0

Table 6. Zero-shot performance comparison of different vision encoders on MedVL-CT69K validation set.

Methods	SE	SP	ACC	AUC
Random	75.9	73.6	74.8	78.7
Supervised	76.4	75.4	75.9	79.4
Ours (Disease-level CLP)	77.9	76.6	77.3	80.7

Table 7. Zero-shot performance comparison of different initialization solutions for vision encoder on MedVL-CT69K validation set. CLP: Contrastive Learning Pre-training.

Types	Methods	SE	SP	ACC	AUC	Δ
Local	Base	72.5	70.9	71.7	76.3	4.0
	Ours	75.4	74.5	75.0	80.3	
Diffuse	Base	78.7	81.7	80.2	85.2	2.8
	Ours	82.1	83.1	82.6	88.0	

Table 8. Comparison between the base model and our model regarding performance improvements in local and diffuse diseases.

8. More visualizations of semantic density

We present the distributions of visual tokens across additional anatomical structures, as illustrated in Figure 6. The figure clearly demonstrates that, for all organs, the visual tokens of the model exhibit increased sparsity after the implementation of VSDB, indicating that the model is prioritizing more important features.

9. Details about zero-shot performance

Table 10 displays the zero-shot performance of the proposed method across 54 abnormalities spanning 15 distinct anatomies.

Type	Diseases
Local	kidney cyst, kidney stone, adrenal gland nodule, stomach cancer, gallstone, pancreatic cancer, small intestine diverticulum, small intestine intussusception, colon cancer, rectal cancer, colon diverticulum, colon appendicolith, liver cyst, liver cancer, liver abscess, liver metastase, spleen infarction, spleen hemangioma, bladder diverticulum, bladder stone, esophageal varicose veins, sacrum osteitis
Diffuse	colon obstruction, colonic gas, colon effusion, colon appendicitis, small intestine obstruction, small intestine gas, small intestine fluid accumulation, cardiomegaly, pericardial effusion, liver glisson's capsule effusion, liver cirrhosis, intrahepatic bile duct dilatation, fatty liver, bronchiectasis, emphysema, pneumonia, pleural effusion, atelectasis, kidney atrophy, hydronephrosis, adrenal hypertrophy, gastric wall thickening, cholecystitis, pancreatitis, pancreatic duct dilatation, pancreas steatosis, pancreas atrophy, splenomegaly, portal vein hypertension, portal vein thrombosis, esophageal hiatal hernia, gallbladder adenomyomatosis

Table 9. Classification of local and diffuse diseases.

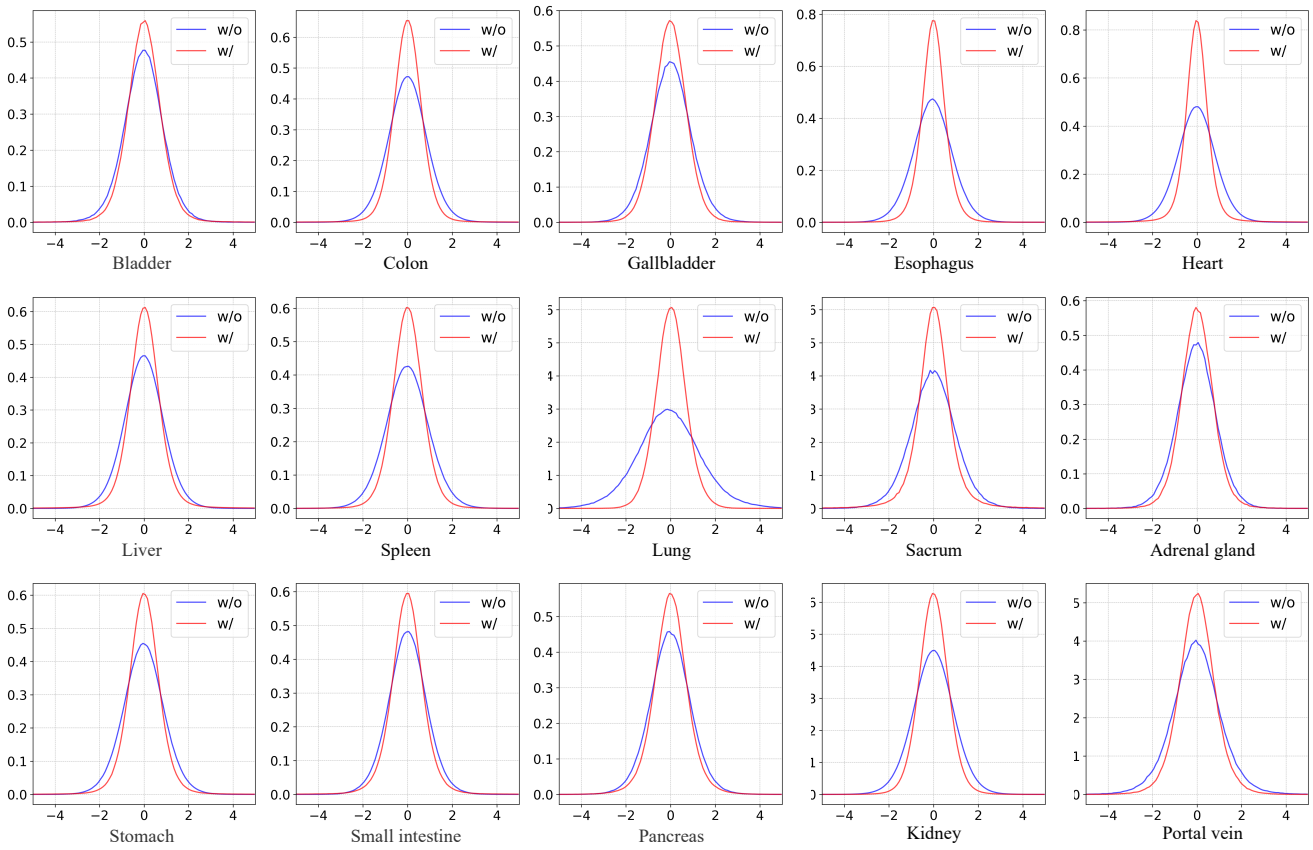

Figure 6. Vision semantic density comparison between models w/ and w/o VSDB.

Anatomy	Abnormality	SE	SP	ACC	AUC
Adrenal gland	Adrenal Hypertrophy	61.5	66.3	63.9	68.0
	Nodule	65.5	63.6	64.6	68.9
Bladder	Diverticulum	71.4	78.8	75.1	81.6
	Stones	78.6	69.9	74.2	80.9
Colon	Colonic Gas	74.4	81.5	78.0	85.3
	Effusion	80.0	81.8	80.9	85.3
	Obstruction	100	95.4	97.7	99.3
	Diverticulum	75.0	61.1	68.0	72.3
	Colon Cancer	77.1	68.8	72.9	80.8
	Rectal Cancer	80.8	88.7	84.8	92.9
	Appendicitis	68.4	75.3	71.9	75.8
	Appendicolith	64.9	57.9	61.4	63.7
Esophagus	Hiatal Hernia	100.0	88.3	94.2	96.9
	Varicose Veins	98.7	97.1	97.9	99.6
Gallbladder	Cholecystitis	67.1	69.7	68.4	74.4
	Gallstone	68.2	79.4	73.8	80.4
	Adenomyomatosis	61.7	60.0	60.9	63.0
Heart	Cardiomegaly	90.0	91.4	90.7	97.0
	Pericardial Effusion	79.2	74.1	76.6	84.1
Kidney	Atrophy	78.4	89.6	84.0	89.5
	Cyst	62.7	62.2	62.5	67.5
	Hydronephrosis	85.1	84.4	84.7	89.9
	Renal Calculus	63.5	61.9	62.7	67.1
Liver	Fatty Liver	84.0	78.4	81.2	90.4
	Glisson's Capsule Effusion	89.7	84.8	87.2	93.8
	Metastase	73.8	82.8	78.3	86.4
	Intrahepatic Bile Duct Dilatation	74.6	73.1	73.9	80.4
	Cancer	86.9	89.3	88.1	93.8
	Cyst	61.0	54.3	57.6	61.0
	Abscess	66.7	92.7	79.7	85.6
	Cirrhosis	90.4	87.2	88.8	96.0
Lung	Atelectasis	95.6	95.9	95.8	99.0
	Bronchiectasis	94.4	85.6	90.0	96.2
	Emphysema	80.0	79.7	79.8	79.0
	Pneumonia	81.1	82.0	81.6	88.9
	Pleural Effusion	95.7	95.8	95.7	98.2
Pancreas	Pancreatic Cancer	93.1	82.6	87.8	94.7
	Atrophy	83.8	86.1	84.9	91.1
	Pancreatitis	85.7	93.6	89.6	95.2
	Pancreatic Duct Dilatation	58.5	84.7	71.6	77.8
	Steatosis	82.2	78.8	80.5	85.7
Portal vein	Hypertension	94.4	90.8	92.6	97.9
	Thrombosis	90.9	94.5	92.7	96.7
Small Intestine	Gas Accumulation	81.9	83.2	82.6	89.3
	Fluid Accumulation	80.3	82.3	81.3	87.9
	Obstruction	85.2	86.9	86.1	92.9
	Diverticulum	84.1	77.1	80.6	88.2
	Intussusception	88.9	66.1	77.5	83.4
Spleen	Hemangioma	68.1	65.9	67.0	69.4
	Infarction	81.8	81.9	81.9	87.5
	Splenomegaly	84.7	83.6	84.1	91.8
Stomach	Gastric Wall Thickening	67.5	74.4	70.9	77.7
	Gastric Cancer	78.6	80.3	79.5	84.5
Sacrum	Osteiti	70.6	75.7	73.2	77.5
Average		79.4	79.6	79.5	84.9

Table 10. Detailed zero-shot performance of our method on each abnormality.