Learning to Zoom with Anatomical Relations for Medical Structure Detection

Bin Pu[1], Liwen Wang[2], Xingbo Dong[2], Xingguo Lv[1], Zhe Jin[2*]

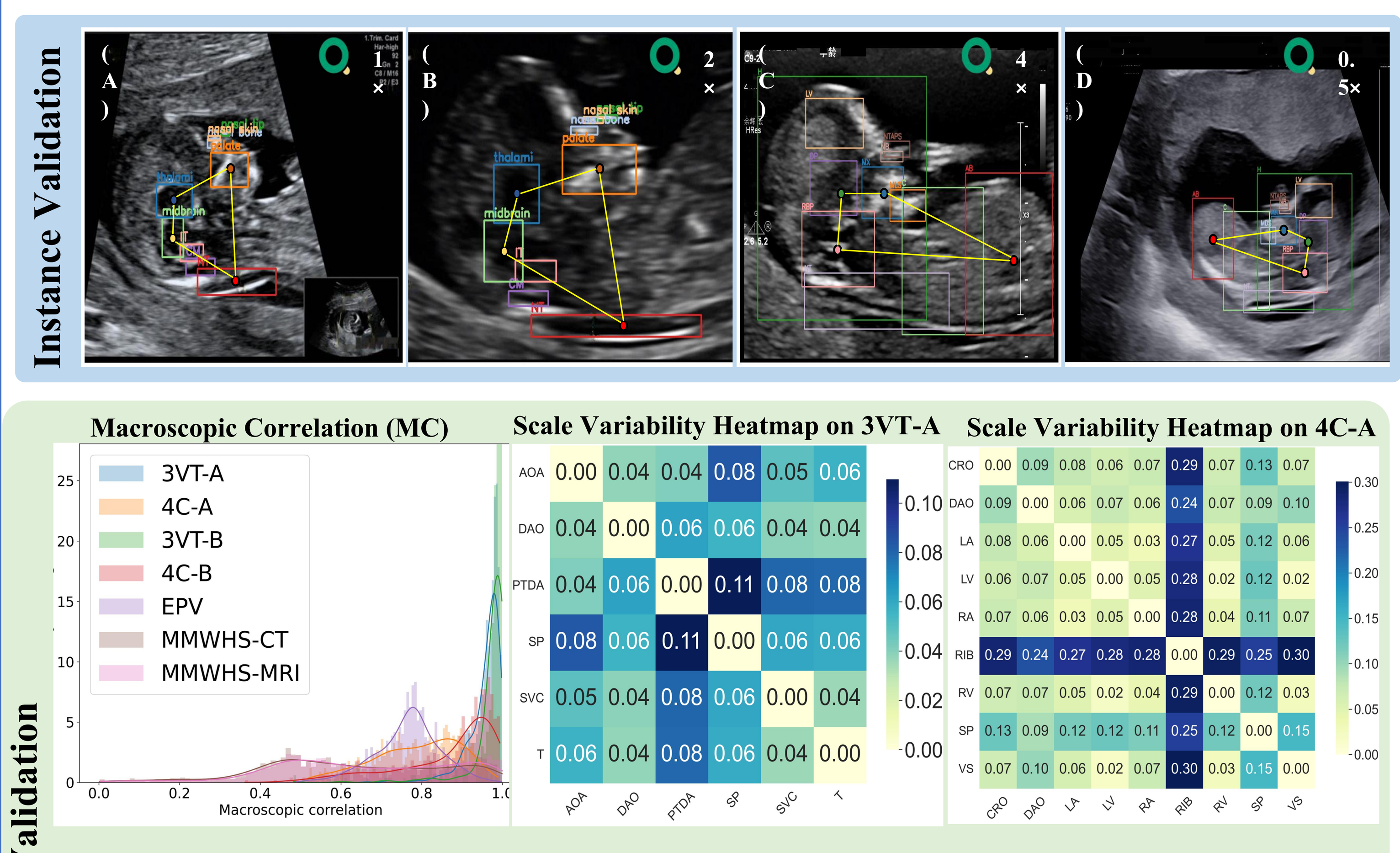
[1]Hunan University [2]Anhui University

NEURAL INFORMATION PROCESSING SYSTEMS

Problem & Motivation

Problem & Motivation

1. Anatomical structures in medical images exhibit **significant scale variability**.
2. Current algorithms **neglect zoom information** and **structural topological priors**.
3. Models struggle with **morphological variability**, **structural inconsistency**, and **detection uncertainty**.



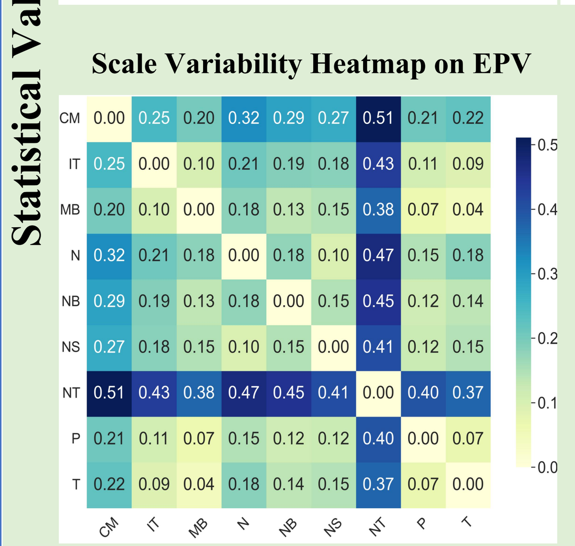


Figure 1: Motivation of our porposed ZR-DETR. The upper row illustrates the observed structural relation consistency and zoom patterns in our proposed ZR-DETR framework. The second row provides empirical validation through visualizations of Macroscopic Correlation (MC) across the employed datasets and category-wise scale variability heatmaps.

Our Contribution

1. Introduced **Zoom Relation Encoder** and **Anatomical Relation Constraints** for scale-aware structural priors.
2. Proposed a **Gaussian Process probabilistic detection framework** for uncertainty quantification.
3. Significantly outperformed baselines in **single-domain and domain adaptation scenarios**.

Methodology:ZR-DETR

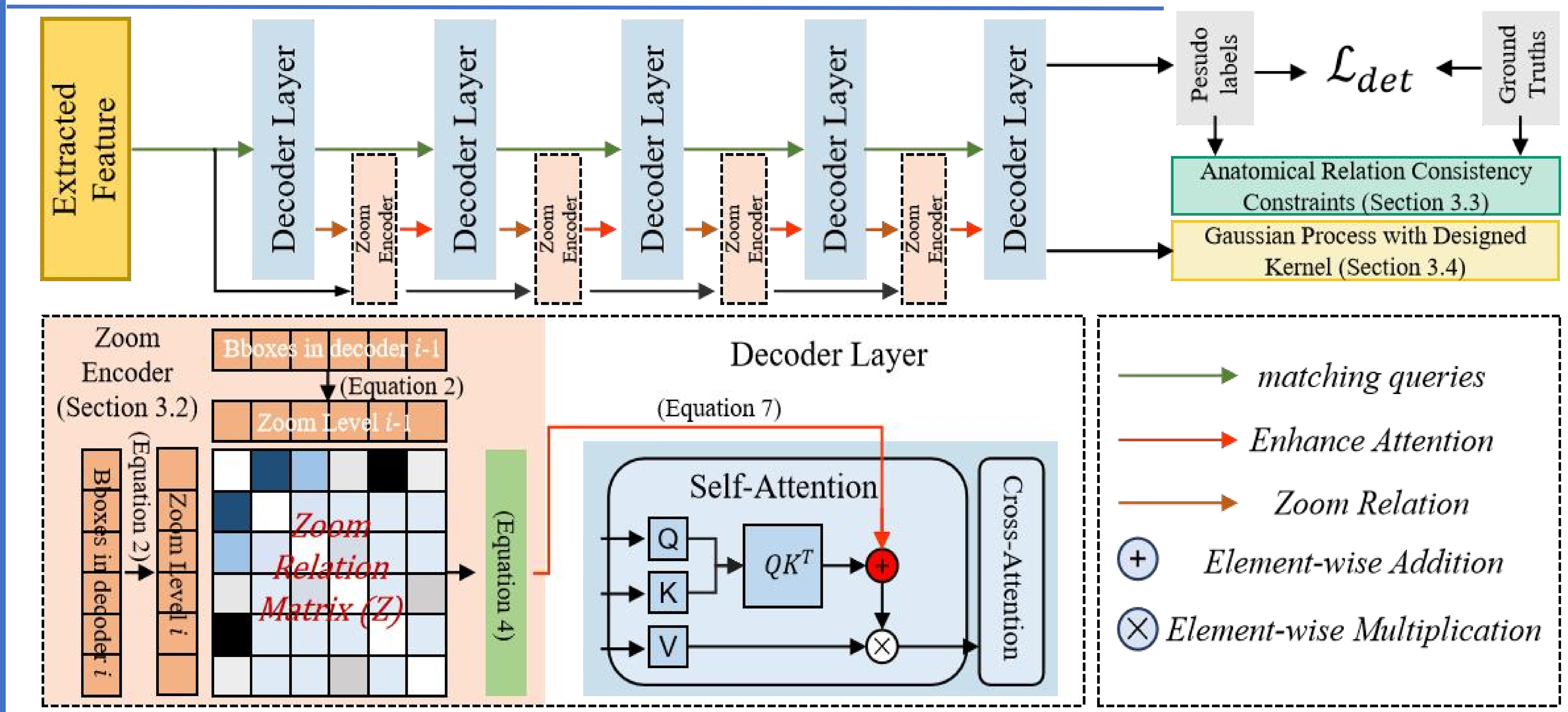


Figure 2: ZR-DETR integrates a Transformer-based encoder-decoder architecture, incorporating Zoom Embeddings to process multi-scale medical image features, Anatomical Relation Consistency Constraints to encode prior anatomical knowledge, and a Gaussian Process for uncertainty-aware detection.

Zoom Relation Encoder

- **Capture relative organ scale relationships**

Zoom Relation Matrix(Z): $\lambda_i = \log\left(\frac{A_i}{A_{\text{img}}}\right), \quad z_{ij} = \frac{|\lambda_i|}{|\lambda_j|}$

Scalar to high-dim vector:

$$\text{Embed}(z_{ij}, 2k) = \sin\left(\frac{s \cdot z_{ij}}{T^{2k/d_z}}\right)$$

$$\text{Embed}(z_{ij}, 2k+1) = \cos\left(\frac{s \cdot z_{ij}}{T^{2k/d_z}}\right)$$

- **Integrates zoom embeddings into attention**

Dual-branch decoder:

$$\text{Attn}(Q_m^l) = \text{Softmax}\left(\frac{Q_m Q_m^\top}{\sqrt{d_{\text{model}}}} + \text{Zoom}(b^{l-1}, b^l)\right) Q_m,$$

$$\text{Attn}(Q_h^l) = \text{Softmax}\left(\frac{Q_h Q_h^\top}{\sqrt{d_{\text{model}}}}\right) Q_h.$$

Anatomical Relation Consistency Constraints

Incorporates prior topological knowledge, penalizing implausible predictions by comparing the spatial proximity between predicted bounding boxes with the prior expected spatial proximity estimated from training data

$$\mathcal{L}_{\text{anatomy}} = \frac{1}{|\mathcal{P}|}\sum_{i=1}^{T}\sum_{j=1}^{T} \mathbf{M}_{a_i,a_j}\left\|k_{\text{spatial}}(\mathbf{b}_i,\mathbf{b}_j) - k_{\text{prior}}(a_i,a_j)\right\|_2$$

Probabilistic Detection with Gaussian Processes

Probabilistic Detection with Gaussian Processes quantifies predictive uncertainty by modeling each detection score as a Gaussian Process sample, using a composite kernel function that integrates visual features, zoom levels, and anatomical categories, thereby providing both predictive mean and variance.

Experimental Results

Results on FCS dataset

Table 1: The performance of different detection methods in FCS (4C) dataset [9].

Method	LA ↑	RA ↑	LV ↑	RV ↑	VS ↑	CRO ↑	SP ↑	DAO ↑	RIB ↑	mAP ↑
Single Domain Structure Detection (Site A)										
FasterRCNN [37] CVPR16	92.0	96.6	94.1	91.5	96.9	97.0	94.2	93.1	73.0	92.1
Deformable-DETR [6] ICLR21	94.5	95.6	96.0	95.2	96.7	93.1	97.5	97.4	73.5	93.3
DAB-DETR [38] ICLR22	96.5	96.7	95.6	94.5	97.5	98.5	97.5	98.0	76.8	94.6
DN-DETR [39] CVPR22	97.2	98.0	98.9	97.7	99.8	97.7	99.2	98.5	85.4	96.9
DINO [41] ICLR23	97.3	97.8	99.4	97.5	99.9	97.9	98.1	98.5	86.4	97.0
Relation-DETR [30] ECCV24	97.6	98.4	97.8	97.2	99.6	97.9	99.9	99.6	87.8	97.3
ZR-DETR (Ours)	98.1	97.9	98.9	99.6	99.9	100	99.9	99.6	87.7	98.0
Single Domain Structure Detection (Site B)										
FasterRCNN [37] CVPR16	71.3	91.3	86.7	80.5	90.4	87.4	88.1	89.1	82.6	85.3
Deformable-DETR [6] ICLR21	75.7	87.4	90.7	85.2	91.0	85.0	90.0	91.0	84.3	86.7
DAB-DETR [38] ICLR22	75.5	92.2	92.9	88.4	94.3	86.7	92.1	92.1	86.6	89.0
DN-DETR [39] CVPR22	77.1	92.3	92.0	87.6	96.0	87.9	93.2	92.4	85.0	89.3
DINO [41] ICLR23	83.4	95.2	96.4	92.9	97.0	92.2	94.3	93.0	89.5	92.7
Relation-DETR [30] ECCV24	87.9	95.1	97.5	94.2	94.9	92.7	96.6	93.3	88.4	93.4
ZR-DETR (Ours)	89.0	95.0	97.4	95.5	96.1	94.0	96.4	94.4	90.0	94.2
Cross Domain Structure Detection (A→B)										
SIGMA [42] CVPR22	50.1	62.1	49.5	51.3	58.9	55.6	46.7	54.0	47.9	52.0
SIGMA++ [43] TPAMI 23	57.1	57.0	60.2	58.3	56.1	58.9	55.5	59.1	60.1	57.8
M^3-UDA [9] CVPR24	79.9	69.8	72.8	71.7	81.0	78.0	81.7	78.0	78.3	76.8
ToMo-UDA [10] ICML24	64.2	75.6	70.4	64.3	66.7	75.0	75.5	77.2	73.0	71.3
DATR [44] TIP25	81.4	62.4	68.9	74.1	81.8	73.9	82.6	83.7	62.3	74.6
ZR-DETR-UDA (Ours)	58.4	82.2	80.8	74.6	89.5	83.2	83.4	87.4	78.2	79.8

Table 2: The performance of different detection methods in FCS (3VT) dataset [9].

Method	DAO ↑	SP ↑	PTDA ↑	T ↑	SVC ↑	AOA ↑	mAP ↑
Single Domain Structure Detection (Site A)							
FasterRCNN [37] CVPR16	93.7	92.0	96.9	87.1	87.5	94.5	91.9
Deformable-DETR [6] ICLR21	97.1	90.2	96.9	91.4	91.0	98.8	94.2
DAB-DETR [38] ICLR22	96.5	96.3	96.8	92.0	89.0	97.9	94.7
DN-DETR [39] CVPR22	97.2	92.7	96.7	89.5	93.7	98.8	94.8
DINO [41] ICLR23	95.8	97.4	97.7	94.5	93.0	99.4	96.6
Relation-DETR [30] ECCV24	96.5	94.0	96.5	94.8	93.0	99.7	95.8
MI-DETR [30] CVPR25	97.8	97.4	97.7	94.5	93.0	99.4	96.6
ZR-DETR (Ours)	98.7	96.9	98.7	95.3	94.3	99.8	97.3
Single Domain Structure Detection (Site B)							
FasterRCNN [37] CVPR16	54.8	30.2	30.0	19.9	39.5	41.3	35.9
Deformable-DETR [6] ICLR21	51.3	22.4	31.9	20.5	36.9	56.0	36.8
DAB-DETR [38] ICLR22	44.6	35.7	29.6	27.2	37.8	62.4	39.6
DN-DETR [39] CVPR22	51.9	33.4	31.0	23.4	44.9	60.0	40.8
DINO [41] ICLR23	51.8	39.0	29.0	26.6	52.7	62.5	43.6
Relation-DETR [30] ECCV24	56.9	42.2	32.6	22.5	57.9	56.1	44.7
MI-DETR [30] CVPR25	53.2	43.2	32.7	22.3	62.4	54.8	44.8
ZR-DETR (Ours)	57.4	39.1	33.2	28.7	73.2	61.2	48.8
Cross Domain Structure Detection (A→B)							
SIGMA [42] CVPR22	42.9	42.8	59.4	39.6	41.7	60.0	47.7
SIGMA++ [43] TPAMI 23	42.3	37.4	45.4	29.0	32.0	42.9	38.1
M^3-UDA [9] CVPR24	59.5	59.7	70.1	51.9	52.4	68.9	60.4
ToMo-UDA [10] ICML24	45.4	60.1	81.5	27.6	45.6	63.7	54.0
DATR [44] TIP25	57.4	59.2	61.7	56.9	50.7	62.7	58.1
ZR-DETR-UDA (Ours)	61.1	67.1	67.8	60.8	49.4	67.9	62.4

Results on MM-WHS dataset

Table 4: The performance of different detection methods in MM-WHS dataset [34].

Method	CT				mAP ↑	MRI				mAP ↑
Single Domain Structure Detection	LV-MYO ↑	LA-blood ↑	LV-blood ↑	AA ↑		LV-MYO ↑	LA-blood ↑	LV-blood ↑	AA ↑	
FasterRCNN [37]	92.7	80.7	88.1	85.9	86.9	80.9	78.2	74.2	60.1	73.4
Deformable-DETR [6]	90.6	82.9	90.4	88.0	88.0	72.8	76.7	75.5	66.5	72.9
DAB-DETR [38]	91.3	83.2	86.7	93.8	88.8	87.1	80.5	74.9	57.1	74.9
DN-DETR [39]	96.3	86.5	89.9	89.0	89.4	84.7	85.8	75.4	66.1	78.0
DINO [41]	83.7	81.9	81.6	82.2	82.4	86.7	82.8	82.9	71.9	81.1
Relation-DETR [30]	89.2	88.7	98.1	95.7	92.9	90.9	81.3	87.9	74.6	83.7
MI-DETR [40]	90.4	87.6	96.5	95.8	92.6	89.8	84.6	86.0	76.9	84.3
ZR-DETR (Ours)	93.5	89.9	99.8	96.0	94.8	87.3	85.7	87.2	78.0	84.5
Cross Domain Structure Detection	CT → MRI				mAP ↑	MRI → CT				mAP ↑
Method	LV-MYO ↑	LA-blood ↑	LV-blood ↑	AA ↑		LV-MYO ↑	LA-blood ↑	LV-blood ↑	AA ↑	
SIGMA [42]	84.4	38.9	77.5	46.3	61.8	61.4	75.8	69.5	77.0	70.9
SIGMA++ [43]	81.2	51.5	78.7	47.1	65.4	67.2	76.4	74.4	77.1	73.8
M^3-UDA [9]	84.6	55.8	80.8	46.9	67.0	70.9	77.5	78.6	76.6	75.9
ToMo-UDA [10]	84.2	60.8	80.9	47.1	68.2	68.4	76.8	77.3	77.0	74.9
DATR [44]	84.7	62.1	81.1	47.4	68.8	73.4	77.1	79.3	76.3	76.5
ZR-DETR-UDA (Ours)	84.5	64.8	81.3	46.9	69.4	78.5	77.1	80.9	75.2	77.9

Results on EPV dataset

Table 3: The performance of different detection methods in EPV dataset [33].

Method	T ↑	NB ↑	P ↑	NS ↑	NT ↑	MB ↑	NT ↑	IT ↑	CM ↑	mAP ↑
Single Domain Structure Detection										
FasterRCNN [37] CVPR16	98.5	80.2	97.0	62.1	86.0	98.3	66.4	94.7	73.5	84.1
Deformable-DETR [6] ICLR21	98.6	87.3	95.5	81.5	92.3	99.6	73.0	95.2	78.2	89.0
DAB-DETR [38] ICLR22	97.7	93.0	96.4	85.5	92.2	98.4	82.8	94.4	77.5	90.9
DN-DETR [39] CVPR22	98.7	93.7	97.3	85.2	94.6	97.4	83.1	96.7	79.1	91.7
DINO [41] ICLR23	98.3	92.3	97.1	86.8	93.8	99.8	82.4	94.6	81.5	91.8
Relation-DETR [30] ECCV24	98.7	94.9	96.7	85.9	95.2	98.3	84.8	96.0	87.8	93.1
MI-DETR [40] CVPR25	98.9	94.6	98.6	89.2	94.5	100	87.2	94.2	84.2	93.5
ZR-DETR(Ours)	99.0	95.1	98.4	91.6	95.4	100	90.1	94.6	85.2	94.4

Computational Efficiency

Table 6: Comparison of computational efficiency across different detection methods.

Metric	FasterRCNN [37]	Deformable-DETR [6]	DAB-DETR[38]	DN-DETR [39]	DINO [41]	Relation-DETR [30]	MI-DETR [40]	ZR-DETR (Ours)
Times (ms) ↓	41.405	40.101	41.558	46.133	42.102	46.2	55.6	45.172
Params (M) ↓	41.1	65.2	48.5	88.5	90.6	111.2	138.1	115.4
TFlops ↓	0.192	0.18	0.0874	0.244	0.256	0.258	0.292	0.268

Code & data: https://github.com/LiwenWang919/ZR-DETR

Please contact *pubin@hnu.edu.cn* for further information.