Supplementary Material for M^3-UDA: A New Benchmark for Unsupervised Domain Adaptive Fetal Cardiac Structure Detection

Bin Pu[1*] Liwen Wang[2*] Jiewen Yang[1*] Guannan He[3] Xingbo Dong[2] Shengli Li[4]
Ying Tan[4] Ming Chen[5] Zhe Jin[2] Kenli Li[6†] Xiaomeng Li[1†]
[1] The Hong Kong University of Science and Technology.
[2] Anhui Provincial International Joint Research Center for Advanced Technology in Medical Imaging, Anhui University, China. [3] Sichuan Provincial Maternity and Child Health Care Hospital.
[4] Shenzhen Maternity and Child Healthcare Hospital. [5] Harbin Red Cross Central Hospital.[6] Hunan University.
{eebinpu, eexmli}@ust.hk, jyangcu@connect.ust.hk, {jinzhe, xingbo.dong}@ahu.edu.cn, lkl@hnu.edu.cn

Effects of backbone and detector on our approach

Table 1 shows the effect of different backbones as the feature extractor and various detection heads on the adaptation task.

Algorithm Pipeline

Based on our description of the proposed M^3-UDA in Section 3 method, the following algorithm description is presented for the HM, SM, and GM components.

Table 1. Compare different detection heads and feature extractors on 4CC.

Hospital 1→2											
Detection head	Feature extractor	RA	RV	LV	VS	SP	LA	CR	DAO	RIB	mAP (%)
Faster-RCNN	VGG-16	76.40	85.31	78.20	77.32	74.73	85.43	75.52	79.71	72.93	78.39
	ResNet-50	65.31	62.66	63.16	52.93	63.36	63.12	58.57	59.31	65.73	61.57
	ResNet-101	70.64	61.52	63.16	66.58	66.72	62.34	75.31	68.64	68.92	67.09
	ResNet-152	79.18	65.87	66.13	66.32	74.84	67.26	72.21	64.42	67.48	69.31
FCOS	VGG-16	58.56	46.50	52.43	48.99	56.56	42.98	49.52	54.23	35.76	49.52
	ResNet-50	81.09	65.29	70.54	67.12	77.56	69.33	79.28	77.86	77.06	73.90
	ResNet-101	79.94	69.77	72.84	71.73	81.02	77.99	81.70	77.99	78.28	76.81
	ResNet-152	83.79	77.28	80.98	76.68	83.30	80.07	84.34	79.96	82.24	80.96

Algorithm 1 Histogram Matching (HM)

Output: $X^{t'}$: The X^t after the Histogram Matching with X^s.
Input: $X^{s,t}$: Input images from source and target domain.
1: $P^{s/t} = \{p_i^{s/t}\}_{i=0}^{k} \leftarrow$: Get the normalized histogram of X_s and X_t;
2: $\sigma := \arg\min \sum_{i=1}^{k} \mathbf{d}(M(P^s), P^t) \leftarrow$: Conduct the optimal mapping $\sigma(\cdot)$;
3: $X^{t'} = \sigma(X^t) \leftarrow$: conduct the histogram matching via optimal mapping $\sigma(\cdot)$; Equation (1) in paper
4: **return** $X^{t'}$;

*Equal Contribution.
†Corresponding author.

Algorithm 2 Sub-structure Matching (SM)

Output: $\mathcal{L}_{SM}$: The graph matching loss of SM;
$\mathcal{L}_{sup}$: The loss of detection result.
Input: $X^{s,t}$: Input images from source and target domain;
Y^s : The ground truth of the source domain;
$\mathcal{L}_{bce}(\cdot)$:Binary cross-entropy Loss;
$\mathcal{L}_{iou}(\cdot)$: Intersection over Union Loss;
Encoder$(\cdot)$: The parameter shared feature extractor;
Decoder$(\cdot)$: The parameter shared detection head for generating the detection result;

$f^{s,t} \leftarrow \text{Encoder}(X^{s,t})$
$\hat{Y}^{s,t} \leftarrow \text{Decoder}(f^{s,t})$
$\mathcal{L}_{sup} \leftarrow \mathcal{L}_{bce}(\hat{Y}^s, Y^s) + \mathcal{L}_{iou}(\hat{Y}^s, Y^s)$
(1). Sub-structure Relation Construction.
$\{c^{s,t}\} \leftarrow$ Get the location of different organs according to the corresponding position of each organ from Y^s and $\hat{Y}^t$.
$A^{s/t} = \begin{pmatrix} \theta(c_0, c_0) & \cdots & \theta(c_0, c_j) \\ \vdots & \vdots & \vdots \\ \theta(c_i, c_0) & \cdots & \theta(c_i, c_j) \end{pmatrix}^{s/t} \leftarrow$ Generate adjacency matrix according to $c^{s/t}$. Equation (2) in paper
(2). Sub-structure Matching.
$\mathcal{L}_{\text{SM}} = \sum_{i=0}^{n} \sum_{j=0}^{n} ||A^s_{i,j} - A^t_{i,j}||_2$. Equation (3) in paper
return $\mathcal{L}_{SM}, \mathcal{L}_{sup}$

Algorithm 3 Global-structure Matching (GM)

Output: $\mathcal{L}_{GM}$: The graph matching loss of GM; $\mathcal{L}_{\text{cls}}$: The classification loss on node;
Input: $X^{s,t}$: Input images from source and target domain;
Y^s : The ground truth of the source domain;
Encoder$(\cdot)$: The parameter shared feature extractor;
Decoder$(\cdot)$: The parameter shared detection head for prediction of the detection result;
GNN$(\cdot)$: The graph convolutional neural network to transform the visual graph;
W_q, W_k and W_v : The learnable weight of three independent projection layers that project the graph to different latent space as query, key and value vector;

$f^{s,t} \leftarrow \text{Encoder}(X^{s,t})$
$\hat{Y}^{s,t} \leftarrow \text{Decoder}(f^{s,t})$
(1).Organ Feature Extraction and Visual Node Sampling.
sample total M visual nodes from $F^{s/t}$ as $\mathcal{V}^{s/t} \in \mathbb{R}^{M \times d}$, according to the corresponding position of each organ from the Y^s and $\hat{Y}^t$,
compute the edge $\mathcal{E}^{s/t} \in \mathbb{R}^{M \times M}$ that represents the connectivity between each node of organs in $\mathcal{V}^{s/t}$ by using spectral clustering.
(2). Global Graphical Representation Construction.
$\mathcal{G}^{s/t} = \text{GNN}(\{\mathcal{V}, \mathcal{E}\}^{s/t}) \in \mathbb{R}^{M \times d}$. Equation (4) in paper
(3). Organ Morphology Knowledge Transfer.
$\dot{\mathcal{G}}^{s/t} = \lambda \cdot \left(W_q \mathcal{G}^{s/t} (W_k \mathcal{G})^{\mathsf{T}}\right) W_v(\mathcal{G}^s \oplus \mathcal{G}^t) \in \mathbb{R}^{M \times d}$ Equation (5) in paper
(4). Graph Matching and Node Classification.
$\mathcal{L}_{\text{GM}} = \sum_{i=0}^{d} \sum_{j=0}^{M} \left\| \dot{\mathcal{G}}^s_{i,j}, \dot{\mathcal{G}}^t_{i,\pi(j)} \right\|_2^2$. Equation (6) in paper
$\mathcal{L}_{\text{cls}} = \text{CE}(\dot{\mathcal{G}}^s, Y^s) + \text{CE}(\dot{\mathcal{G}}^t, \hat{Y}^t)$ Equation (7) in paper
return $\mathcal{L}_{GM}$,$\mathcal{L}_{\text{cls}}$