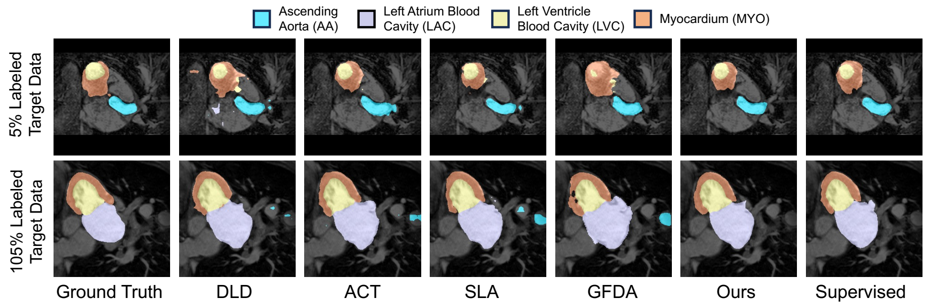

Fig. S1. Visualization results on MM-WHS with 5% and 10% labeled target data. Our method consistently generates more accurate predictions, with the completest region.

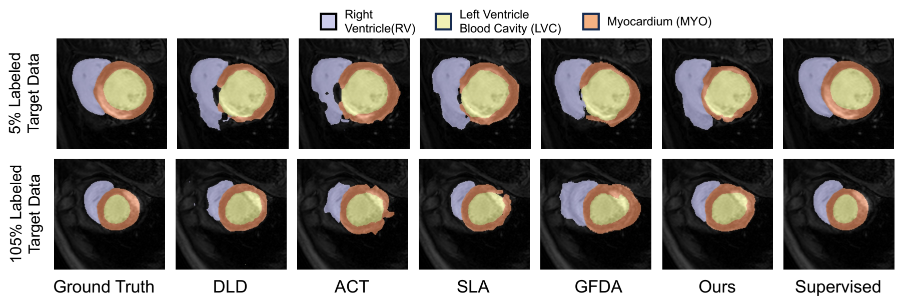

Fig. S2. Visualization results on MS-CMRSeg with 5% and 10% labeled target data. CMMTA outperforms other methods, generating better prediction on the challenging RV region.

Table S1. Ablation study for different data mixing methods on MM-WHS dataset.

Method	Cardiac CT → Cardiac MRI									
	Dice↑					ASSD↓				
	AA	LAC	LVC	MYO	Avg.	AA	LAC	LVC	MYO	Avg.
CutMix [24]	76.7	81.6	90.0	70.2	79.6±5.1	5.4	5.4	3.3	2.4	4.1±2.4
CMMTA	**80.8**	**83.7**	**91.6**	**76.1**	**83.1±4.6**	**3.9**	**2.7**	**2.2**	**1.6**	**2.6±0.9**

Table S2. Ablation study for single pseudo supervision provided by different teacher models on MM-WHS dataset.

Providing	Cardiac CT → Cardiac MRI									
	Dice↑					ASSD↓				
	AA	LAC	LVC	MYO	Avg.	AA	LAC	LVC	MYO	Avg.
Model θ_1^t	80.7	82.5	90.3	75.4	82.2±4.6	3.5	3.6	3.2	3.5	3.1±1.7
Model θ_2^t	80.8	81.6	90.9	74.2	81.9±4.5	4.7	4.3	2.7	1.7	3.3±1.4
CMMTA	**80.8**	**83.7**	**91.6**	**76.1**	**83.1±4.6**	**3.9**	**2.7**	**2.2**	**1.6**	**2.6±0.9**

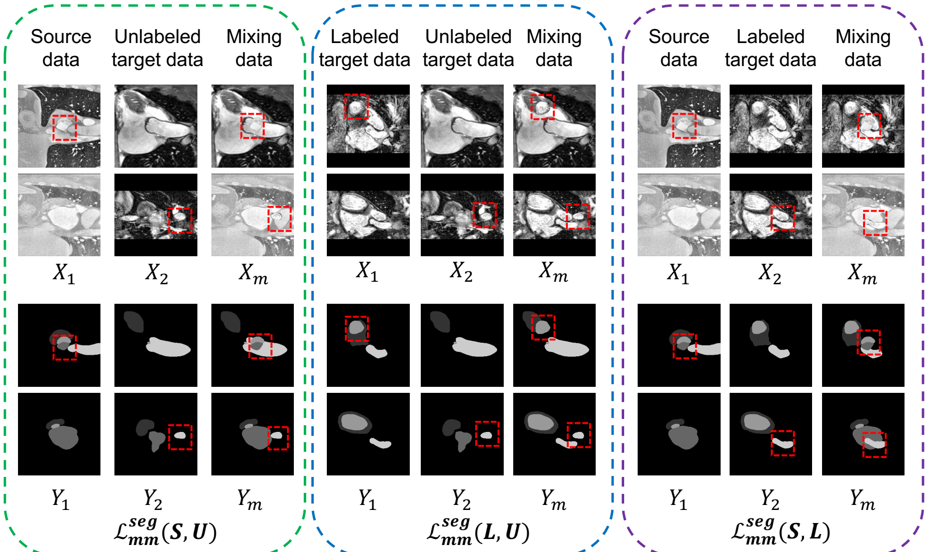

Fig. S3. Visualization of the class-aware mutual mixup. Here we show our detailed implementation for triple pathways.

Algorithm 1: The proposed training algorithm

Input: Labeled source data with label $S = \{x_s, y_s\}$, labeled target data with label $L = \{x_l, y_l\}$, unlabeled target data $U = \{x_u, y_u\}$

Initialize four network $\mathbf{F}$ parameter θ_1^t, θ_1^s, θ_2^t, θ_2^s. ;
for *step= 1* **to** n **do**
    $X_1, Y_1 \leftarrow$ Sample S, $X_2, Y_2 \leftarrow$ Sample L, $X_3 \leftarrow$ Sample U;
    Obtain $X_m(X_1, X_2), Y_m(Y_1, Y_2)$ by Eq.(1) and Eq.(2);
    Calculate the $S \leftrightarrows L$ consistency loss $\mathcal{L}_{mm}^{seg}(X_m(X_1, X_2), Y_m(Y_1, Y_2))$;
    Generate pseudo label $\hat{Y}_{u1} = \arg\max(\mathbf{F}(X_3, \theta_1^t))$ by network θ_1^t;
    Generate pseudo label $\hat{Y}_{u2} = \arg\max(\mathbf{F}(X_3, \theta_2^t))$ by network θ_2^t;
    Obtain $X_m(X_1, X_3), Y_m(Y_1, \hat{Y}_{u2})$, $X_m(X_2, X_3), Y_m(Y_2, \hat{Y}_{u1})$ by Eq.(1),(2);
    Calculate the $S \leftrightarrows U$ consistency loss $\mathcal{L}_{mm}^{seg}(X_m(X_1, X_3), Y_m(Y_1, \hat{Y}_{u2}))$;
    Calculate the $L \leftrightarrows U$ consistency loss $\mathcal{L}_{mm}^{seg}(X_m(X_2, X_3), Y_m(Y_2, \hat{Y}_{u1})$;
    Updated the θ_1^s and θ_2^s by Eq.(5);
    Updated the θ_1^t and θ_2^t by EMA with θ_1^s and θ_2^s;
end