Target	AGG	Ensemble	Epi-FCR(S2B)	First Blk	Third Blk	Fixed DSNN	Epi-FCR
A.	77.6	79.3	80.4	78.8	81.0	79.1	82.1
C.	73.9	75.9	75.3	75.5	75.6	76.6	77.0
P.	94.4	95.4	94.4	93.7	92.2	93.5	93.9
S.	70.3	71.2	73.4	74.5	74.1	74.7	73.0
Ave.	79.1	80.4	80.9	80.6	80.7	80.9	81.5

Table 7: Further evaluation on PACS using ResNet-18.

A. Additional analysis

We conduct some analysis to better understand our episodic training method, and its contributions.

Comparison model ensemble In our current implementation of episodic training, besides the AGG model we regularize, n domain-specific branches are used for generating DG episodes. In this way, it increases the total parameters during training to $n+1$ times that of AGG, although in the end only a single AGG branch is used for testing. To verify that the benefit is not solely due to additional parameters, we compare our episodic-training method with the ensemble of $n+1$ AGG models on PACS using ResNet-18. The result in Table 7 shows that our episodic training is more effective than the ensemble model. Crucially this is despite the fact that Epi-FCR is $1/(n+1)$th the size of the full ensemble during testing.

Flexibility of episodic training a) *Global sharing parameters:* In our current demonstration, we use $n+1$ branches to conduct the episodic training. To reduce the total trainable parameters, we can also globally share the bottom feature layers and episodic-train the rest feature layers and classifier. For example, globally sharing the first two blocks and episodic-training the remaining parameters still leads to a 1.8% improvement over the baseline (see Table 7, S2B). b) *Intermediate feature layers:* Applying episodic training around a typical feature vs classifier module split is intuitive, but other options are possible. For example, we evaluate splitting the modules at intermediate feature layers: first block, and third block of ResNet-18 on PACS. The results, in Table 7 (First Blk and Third Blk), show that episodic training can also work with intermediate layer splits, but the original design is better. c) *Fixed domain-specific branches:* Currently, we train the domain-specific classification losses and the episodic training losses jointly. We also evaluate our episodic training using the pre-trained domain-specific branches. From the result in Table 7 (Fixed DSNN), we can see that episodic training using fixed domain-specific branches is still effective with 1.8% performance improvement over AGG.