

Supplementary Material: Ranking Distance Calibration for Cross-Domain Few-Shot Learning

Pan Li, Shaogang Gong
Queen Mary University of London
{pan.li, s.gong}@qmul.ac.uk

Chengjie Wang
Tencent Youtu Lab
jasoncjwang@tencent.com

Yanwei Fu
Fudan University
yanweif@fudan.edu.cn

In this supplementary material, we present:

- To validate the robustness of the proposed RDC and RDC-FT methods, we analyse the sensitivities of different hyper-parameters, *i.e.* λ , p and α (in Sec. A);
- To qualitatively show the effectiveness of the proposed RDC and RDC-FT methods, we show the ranking lists of a case study with and *w/o* RDC (Figure b in Sec. B), and further use t-SNE [2] to visualise the embeddings of three target domains with and *w/o* RDC-FT (Figure c in Sec. B);
- For better understanding the computing process of the Jaccard distance, we illustrate the algorithm of the Jaccard distance in Sec. C.
- The notations for all symbols and hyper-parameters used in the main paper are defined (in Sec. D);

A. Sensitivity analysis of the hyper-parameters

In all experiments of the main paper, we reported the results on 8 target domains with the same hyper-parameters. In practice, our method is robust to the hyper-parameters selection as shown in Fig. a. Further, we analyse in depth three key hyper-parameters, λ , p and α .

A.1. Effect of the trade-off scalar λ

The trade-off scalar λ is used to balance the original distance and the Jaccard distance for the proposed RDC method, thus it is a critical hyper-parameter for RDC. We conducted experiments to test RDC with $\lambda = \{0.1, 0.3, 0.5, 0.7, 0.9\}$ on the pre-trained space.

The results are shown in Tab. a and Fig. a(a), from which we can see that assigning smaller weights to the original distance (smaller λ) is a better choice for RDC. In particular, the best λ for 1-shot is 0.3 while that for 5-shot is 0.5. This indicates that the original distance becomes more robust when the shot increases, thus the original space should occupy larger weights in the calibrated distance. Besides, when λ is between 0.1 to 0.5, the average accuracies of

λ	5-way 1-shot								Ave.
	CUB	Car	Places	Plantae	Crop	Euro.	ISIC	Chestx	
0.1	47.10	37.01	57.52	41.06	80.15	65.66	31.34	22.40	47.78
0.3	47.46	37.47	57.67	41.29	79.82	65.67	31.72	22.52	47.95
0.5	47.51	37.79	57.28	41.27	78.87	65.21	31.92	22.61	47.81
0.7	47.12	37.85	55.74	40.92	76.71	64.13	31.98	22.70	47.14
0.9	45.94	37.37	52.51	40.12	71.72	62.19	31.98	22.78	45.58

λ	5-way 5-shot								Ave.
	CUB	Car	Places	Plantae	Crop	Euro.	ISIC	Chestx	
0.1	60.94	49.42	70.66	55.79	88.67	76.95	40.11	25.03	58.45
0.3	61.64	50.25	71.00	56.12	88.52	76.96	40.72	25.32	58.82
0.5	62.04	50.90	71.01	56.10	88.00	76.54	41.00	25.66	58.91
0.7	61.81	51.07	70.30	55.41	86.74	75.44	40.86	25.75	58.42
0.9	60.60	50.63	67.97	53.85	83.70	73.30	40.45	25.70	57.03

Table a. **Analysis of the trade-off scalar λ .** Results of RDC with $\lambda = \{0.1, 0.3, 0.5, 0.7, 0.9\}$ on the pre-trained space.

p	5-way 1-shot								Ave.
	CUB	Car	Places	Plantae	Crop	Euro.	ISIC	Chestx	
16	48.56	37.12	53.27	41.71	71.98	61.57	31.63	22.81	46.08
32	47.43	37.48	51.38	40.06	69.93	60.54	31.45	22.87	45.14
64	47.65	37.98	52.07	40.51	69.95	60.42	31.45	22.87	45.36
128	47.78	38.27	53.10	40.91	70.37	61.22	31.60	22.90	45.77
256	44.88	36.39	48.54	40.08	67.56	61.70	31.63	22.89	44.21
512	44.33	36.14	48.75	39.24	66.76	61.00	31.34	22.79	43.79

p	5-way 5-shot								Ave.
	CUB	Car	Places	Plantae	Crop	Euro.	ISIC	Chestx	
16	65.00	49.58	70.46	56.29	88.54	76.50	40.15	26.01	59.07
32	65.38	51.80	70.29	56.48	88.41	76.89	41.12	26.38	59.59
64	65.19	52.43	70.55	56.46	88.27	77.20	41.74	26.46	59.77
128	64.93	52.77	71.21	56.59	88.27	77.20	41.81	26.49	59.91
256	63.94	52.34	69.99	56.58	88.41	77.81	41.81	26.49	59.67
512	63.10	52.13	69.95	55.39	87.78	76.88	41.29	26.35	59.11

Table b. **Analysis of the number of the reduced dimensions p .** Results of NPC on the proposed non-linear subspaces with $p = \{16, 32, 64, 128, 256, 512\}$.

RDC are stable, verifying the robustness of λ . Therefore, we set $\lambda = 0.5$ in all experiments of the main paper.

A.2. Influence of the reduced dimensions p

The dimensions p in the subspace is a key parameter to build our non-linear space. Typically, we choose $p = \{16, 32, 64, 128, 256, 512\}$ ($p = 512$ represents the original space) to test the effects of different dimension p .

Table b and Fig. a(b) show that the performance on different subspaces are stable when p is smaller than 128. This observation shows that the subspaces constructed by the hyperbolic tangent transformation are not sensitive to

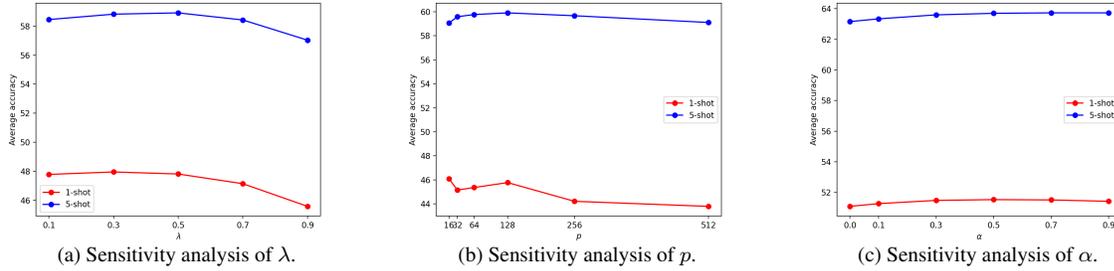


Figure a. **Analysis of the hyper-parameters, i.e. λ , p and α .** The evaluation results are the average accuracies of 8 target domains.

α	5-way 1-shot								
	CUB	Car	Places	Plantae	Crop	Euro.	ISIC	Chestx	Ave.
0	51.13	39.15	60.18	43.93	86.01	70.71	35.34	22.29	51.09
0.1	51.14	39.21	60.63	44.07	86.16	71.11	35.56	22.29	51.27
0.3	51.22	39.17	61.25	44.33	86.31	71.49	35.79	22.28	51.48
0.5	51.20	39.20	61.50	44.33	86.33	71.57	35.84	22.27	51.53
0.7	51.17	39.17	61.50	44.23	86.24	71.52	35.98	22.23	51.51
0.9	50.97	39.13	61.41	44.05	86.05	71.46	36.04	22.23	51.42

α	5-way 5-shot								
	CUB	Car	Places	Plantae	Crop	Euro.	ISIC	Chestx	Ave.
0	67.35	53.80	73.62	60.19	93.25	83.85	47.63	25.58	63.16
0.1	67.46	53.81	73.96	60.48	93.34	84.14	48.02	25.54	63.34
0.3	67.65	53.80	74.42	60.72	93.48	84.51	48.72	25.51	63.60
0.5	67.77	53.75	74.65	60.63	93.55	84.65	49.06	25.48	63.70
0.7	67.87	53.69	74.74	60.56	93.57	84.70	49.26	25.47	63.73
0.9	67.94	53.65	74.80	60.42	93.54	84.62	49.50	25.47	63.73

Table c. **Analysis of the attention scalar α .** Results of RDC-FT with the attention scalar $\alpha = \{0, 0.1, 0.3, 0.5, 0.7, 0.9\}$. $\alpha = 0$ represents the RDC-FT results without attention strategy.

the reduced dimensions. In particular, the subspace with $p = 16$ is the best dimension for 1-shot learning and that with $p = 128$ is the best dimension for 5-shot learning. To make a balance among different shot learning, we set $p = 64$ in all experiments of the main paper.

A.3. Effect of the attention scalar α

The attention scalar α is used to increase the weights of the calibrated distance occurred in \hat{R} , here we investigate the effectiveness of different α .

The results in Tab. c and Fig. a(c) show that this attention strategy can benefit the representation adaptation for FSL task in the target domain. In specific, moderately increasing the attention scalar (α from 0.1 to 0.5) can improve the effectiveness of the attention strategy. To the contrary, overly increasing the attention scalar (α from 0.5 to 0.9) will introduce less even negative effect, resulting the decrease (slight increase) of the performance on 1(5)-shot learning. Therefore, the choice of $\alpha = 0.5$ in the main paper is a moderate and robust parameter for the attention strategy.

B. Visualisation

To qualitatively show the effectiveness of our RDC and RDC-FT methods. We first show a case study of a FSL task from *CUB* by comparing the original ranking list and the ranking list with RDC. As in Fig. b, for a given query

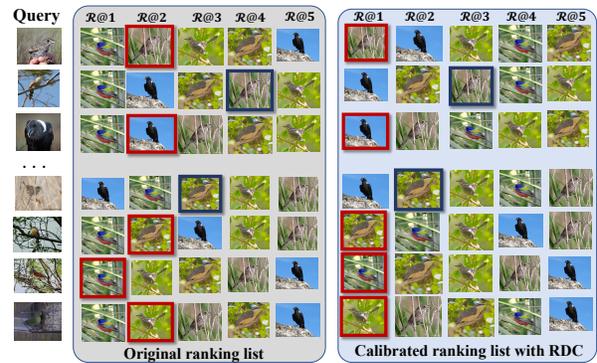


Figure b. **Ranking lists of a 5-way 1-shot task from *CUB*.** The images with red/blue rectangle are the ground-truth support data for a given query. The RDC method calibrates the original ranking list to yield correct recognition results (the images with red rectangle) or closer pairwise distances (the images with blue rectangle).

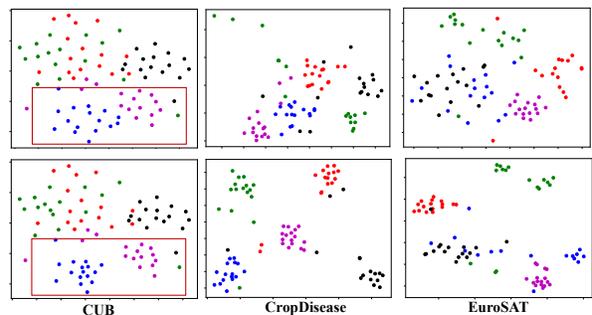


Figure c. **T-SNE visualisation of 5-way 1-shot tasks.** Different colours refer to different classes. We visualise the task features before (the 1st row) and after (the 2nd row) the RDC-FT method.

data, our RDC method pulls the ground-truth support data closer to the query data, arriving at a more accurate position. This process is achieved by the calibration process of our RDC method. For the RDC-FT method, we use t-SNE [3] to visualise the feature embeddings of FSL tasks randomly selected from target domains, i.e. *CUB*, *CropDisease* and *EuroSAT*. As in Fig. c, the feature representations with RDC-FT (in the 2nd row plots) have less within-class variations and large class margins compared to these without RDC-FT process (in the 1st row plots), showing that the RDC-FT method can guild a task-specific embedding

Algorithm C1: Jaccard distance computing

Data: pre-trained feature extractor f_Φ ; FSL task \mathcal{T} ; k -nearest neighbors set: k_1, k_2 .
Result: Jaccard distance $D_J = \{d_J(i, g_q), \text{ where } i, g_q \in [1, n]\}$.

- 1 Extract the embeddings for \mathcal{T} : $X = f_\Phi(\mathcal{T}), X \in \mathbb{R}^{n \times m}$;
- 2 Compute original distances D_o ;
/* k -reciprocal discovery and encoding */
- 3 **for** i **in** n **do**
- 4 Compute the k -nearest neighbors ranking list $R_i(k)$ for x_i ;
/* k -reciprocal neighbor discovery process */
- 5 **for** x_g **in** $R_i(k)$ **do**
- 6 Compute the k -reciprocal nearest neighbors set $\mathcal{R}_g(k)$ for x_g ;
- 7 Expand $\mathcal{R}_g(k)$ by mining hard-positive samples:
 $\hat{R}_i(k) \leftarrow R_i(k) \cup \mathcal{R}_g(\frac{1}{2}k)$ s.t. $|R_i(k) \cap \mathcal{R}_g(\frac{1}{2}k)| \geq \frac{2}{3} |R_g(\frac{1}{2}k)|$;
- 8 **end**
/* k -reciprocal encoding process */
- 9 Encode the expanded set \hat{R}_i as $\mathcal{V}_i = [\mathcal{V}_{i,g_1}, \mathcal{V}_{i,g_2}, \dots, \mathcal{V}_{i,g_n}]$ by $\mathcal{V}_{i,g_q} = \begin{cases} e^{-d_o(x_i, x_{g_q})} & \text{if } g_q \in \hat{R}_i(k) \\ 0 & \text{otherwise.} \end{cases}$;
- 10 **end**
/* query expansion and Jaccard distance computing */
- 11 **for** i **in** n **do**
- 12 Expand the feature of x_i as $\mathcal{V}_i = \frac{1}{|\hat{R}_i(k_2)|} \sum_{g_q \in \hat{R}_i(k_2)} \mathcal{V}_{g_q}$; // query expansion
- 13 Compute the Jaccard distance $d_J(i, g_q) = 1 - \frac{|\hat{R}_i(k) \cap \hat{R}_{g_q}(k)|}{|\hat{R}_i(k) \cup \hat{R}_{g_q}(k)|} = 1 - \frac{\sum_{j=1}^n \min(\mathcal{V}_{i,g_j}, \mathcal{V}_{g_q,g_j})}{\sum_{j=1}^n \max(\mathcal{V}_{i,g_j}, \mathcal{V}_{g_q,g_j})}$;
- 14 **end**

Symbol	Meaning
\mathcal{T}	FSL task in the target domain
x_i	Feature of i th sample in \mathcal{T}
D_o	Euclidean distance matrix in the original space
D_J	Jaccard distance matrix
\hat{D}_o	Calibrated distance matrix in the original space
\hat{D}_{sub}	Calibrated distance matrix in the subspace
\hat{D}_{com}	Complementary calibrated distance matrix
$d_o(i, j)$	Pairwise distance between x_i and x_j
$\mathbf{d}_o(i, :)$	Pairwise distances between x_i and $x_j \in \mathcal{T}$
$d_J(i, g_q)$	Jaccard distance between x_i and x_{g_q}
$R_i(k)$	k -nearest neighbors ranking list of x_i
$\hat{R}_i(k)$	Expanded k -nearest neighbors ranking list of x_i
\mathcal{V}_{i,g_q}	Gaussian kernel of pairwise distance between x_i and x_{g_q}

Table d. Explanation of the symbols.

Hyper-parameter	Meaning
k	Number of candidates in $R_i(k)$
k_2	Number of samples for updating \mathcal{V}_i
λ	Trade-off scalar to balance D_o and \hat{D}_{com}
p	Dimensions of feature in the subspace
T	Number of epochs in fine-tuning stage
τ	Temperature-scaling hyper-parameter
α	Attention scalar

Table e. Explanation of the hyper-parameters.

where the samples can easily be classified by a simple NPC classifier. Moreover, our RDC-FT method, as expected, is functioning as an implicit clustering process for FSL task. This can be qualitatively verified by the observation of the clustering effect as in the 2nd row plots of Fig. c.

C. Details of Jaccard distance

The Jaccard distance computing is an important part of RDC. In specific, the concept of Jaccard distance derives from [1] and the re-weighting strategy for Jaccard distance is also used in [4]. We briefly introduce the computing process of Jaccard distance in the main paper. Here, we further illustrate more details for clearer description as in Algorithm C1. In this pseudo-code, we illustrate the computing process of k -reciprocal discovery and encoding in line 3-9, and the discovery process and encoding process are presented in line 5-8 and line 9, respectively. Then, the query expansion and Jaccard distance computing process are illustrated in line 11-14 of Algorithm C1.

D. Symbols and hyper-parameters

To clearly and fast understand the equations in the main paper, we list the symbols and hyper-parameters in the Tab. d and Tab. e, respectively.

References

- [1] Song Bai and Xiang Bai. Sparse contextual activation for efficient visual re-ranking. *IEEE TIP*, 25(3):1056–1069, 2016. 3
- [2] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. 9(Nov):2579–2605, 2008. 1
- [3] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008. 2
- [4] Zhun Zhong, Liang Zheng, Donglin Cao, and Shaozi Li. Re-ranking person re-identification with k-reciprocal encoding. In *CVPR*, pages 1318–1327, 2017. 3