Imbalanced Label Distribution Learning

Appendix

A.1 Details of Datasets

A.1.1 Description of Datasets

- *Movie* (movie rating): The Movie dataset [Geng, 2016] is about the user ratings on movies. The dataset includes 7,755 movies and 54,242,292 ratings from 478,656 different users. The ratings come from Netflix, ranging from 1 to 5 stars (5 labels). The distribution of rating labels for each movie is calculated as the percentage of each rating level. The features are extracted from metadata, including genre, director, actor, country, budget and etc. Categorical attributes are converted into binary vectors. The final feature vector extracted from each movie is 1,869-dimensional.
- *SCUT-FBP* (facial beauty perception): SCUT-FBP [Xie *et al.*, 2015] contains 500 frontal, unoccluded faces aged from 15 to 60 with neutral expression in the size of 350×350 pixels. All the images are labeled with beauty scores by 75 volunteers. These ratings were collected by randomly displaying one image at a time and asking the raters to rate the attractiveness within 5 degrees. According to the score given by the rater, the label distribution of each instance is generated. For the imbalanced *SCUT-FBP* dataset, we extract a 2048-dimensional feature vector by the pre-trained ResNet [He *et al.*, 2016] provided by Dlib [King, 2009] for each image, and use PCA to reduce its dimension to 128.
- *Emotion6* (visual sentiment distribution perception): Emotion6 [Peng *et al.*, 2015] contains 1,980 images collected from Flickr. The images are annotated with the votes for seven emotional categories (i.e. *anger*, *disgust*, *joy*, *fear*, *sadness*, *surprise* in Ekman's basic emotion set [Ekman, 1992] and *neutral*). For the imbalanced *Emotion6* dataset, we extract a 2048-dimensional feature vector by ResNet-50 [He *et al.*, 2016] pre-trained on ImageNet [Russakovsky *et al.*, 2015] for each image, and use PCA to reduce its dimension to 128.
- *Flickr_LDL* (visual sentiment distribution perception): Flickr_LDL [Yang *et al.*, 2017] contains 11,150 images collected from Flickr. The images are annotated with eight emotional categories (i.e., *awe*, *disgust*, *fear*, *amusement*, *sad*, *contentment*, *excitement*, *anger* in Mikel's Wheel [Mikels *et al.*, 2005]). For the imbalanced *Flickr_LDL* dataset, we extract a 2048-dimensional feature vector by ResNet-50 [He *et al.*, 2016] pre-trained on ImageNet [Russakovsky *et al.*, 2015] for each image, and use PCA to reduce its dimension to 128.
- *Natural Scene* (visual sentiment distribution perception): Natural Scene [Geng, 2016] is collected from 2,000 natural scene images with the results of the inconsistent multilabel rankings. Nine possible labels are associated with these images, including sun, cloud, sky, building, water, mountain, snow, desert and plant. The images are labeled by ten human rankers. For each image, they first select the labels they think are relevant to the image from the nine candidate labels, and then arrange the relevant labels in descending order related to the image. Each human ranker makes decisions independently, so the resulting multi-label rankings are expected to be highly inconsistent. Then, the inconsistent rankings for each image are transformed into label distribution through the nonlinear programming process [Geng and Luo, 2014] and the most compatible common label distribution with all individual rankings is found. Finally, for each image, the method proposed in [Boutell *et al.*, 2004] is used to extract 294-dimensional feature vectors.
- *RAF-ML* (visual sentiment distribution perception): RAF-ML [Shang and Deng, 2019] contains 4,908 images. It is a facial expression dataset, where each image is described by 2048-dimensional DBM-CNN feature and 6-dimensional expression distribution. For the imbalanced *RAF-ML* dataset, we use PCA to reduce the dimension of each feature vector to 128.

A.1.2 Dataset Sampling Process

When sampling the datasets, we adopt the following settings to make ILDL tasks in accord with real-world imbalanced problems: we first rank all the classes by $\hat{N}_i$ in training data and obtain the label j which has the maximum sum description degree. Then, we persist any training point i that satisfies $\boldsymbol{d}_{\boldsymbol{x}_i}^{y_j} < \eta_1$, where η_1 is a threshold. In addition, for the label j' which has the minimum sum description degree, we disregard any training point i' that satisfies $\boldsymbol{d}_{\boldsymbol{x}_{i'}}^{y_{j'}} < \eta_2$, where η_2 is another threshold. We also leverage random sampling to prevent homogenization of training set samples. In this way, the DLD of the training set will be significantly different from the test set.

A.2 Evaluation Criteria

Six standard LDL measures (Chebyshev Distance, Clark Distance, Canberra Metric, Kullback-Leibler Divergence, Cosine Coefficient, and Intersection Similarity between ground-truth description degree vectors and predicted description degree vectors) are selected to evaluate different methods for the prediction of the whole distribution. Besides, Euclidean Distance is also adopted to evaluate the performance of different methods on *tail*, *head* and *all* labels. Below we use D to denote the true description degree distribution, $\widehat{D}$ to denote the predicted description degree distribution and c to denote the number of labels.

The definition of Chebyshev Distance is:

$$Dis_1\left(D,\widehat{D}\right) = \max_j \left|d_j - \hat{d}_j\right|.$$

The definition of Clark Distance is:

$$Dis_2\left(D,\widehat{D}\right) = \sqrt{\sum_{j=1}^{c} \frac{\left(d_j - \hat{d}_j\right)^2}{\left(d_j + \hat{d}_j\right)^2}}.$$

The definition of Canberra Metric is:

$$Dis_3\left(D,\widehat{D}\right) = \sum_{j=1}^{c} \frac{\left|d_j - \hat{d}_j\right|}{d_j + \hat{d}_j}.$$

The definition of Kullback-Leibler Divergence is:

$$Dis_4\left(D,\widehat{D}\right) = \sum_{j=1}^{c} d_j \ln \frac{d_j}{\hat{d}_j}.$$

The definition of Cosine Coefficient is:

$$Sim_1\left(D,\widehat{D}\right) = \frac{\sum\limits_{j=1}^{c} d_j \hat{d}_j}{\sqrt{\sum\limits_{j=1}^{c} {d_j}^2}\sqrt{\sum\limits_{j=1}^{c} \hat{d}_j^2}}.$$

The definition of Intersection Similarity is:

$$Sim_2\left(D,\widehat{D}\right) = \sum_{j=1}^{c} \min\left(d_j, \hat{d}_j\right).$$

The definition of Euclidean Distance is:

$$Dis_5\left(D,\widehat{D}\right) = \sqrt{\sum_{j=1}^{c} \left(d_j - \hat{d}_j\right)^2}.$$

A.3 Algorithms and Hyperparameters

A.3.1 Existing SOTA Algorithms

Several existing state-of-the-art LDL algorithms, i.e., SA-BFGS [Geng, 2016], EDL-LRL [Jia *et al.*, 2019], LDLSF [Ren *et al.*, 2019a], LDL-LCLR [Ren *et al.*, 2019b], Adam-LDL-SCL [Jia *et al.*, 2021] and LDL-LDM [Wang and Geng, 2021] are set as baselines.

- SA-BFGS [Geng, 2016] optimizes the Kullback-Leibler divergence between the ground truth and the predicted distribution through the quasi-Newton method BFGS. In the experiments, the minimum log-likelihood difference for convergence is set to 10^{-5}.
- EDL-LRL [Jia *et al.*, 2019] exploits local label correlation by capturing low-rank structure on clusters of samples with trace-norm regularization. In the experiments, the number of clusters is set to 5 and the maximum number of iterations is set to 50 as suggested in the code provided by the authors.

- LDLSF [Ren *et al.*, 2019a] exploits the label correlations and learns the common features for all labels and specific features for each label simultaneously. In the experiments, λ_1 is set to 10^{-4}, λ_2 is set to 10^{-2}, λ_3 is set to 10^{-3}, and ρ is set to 10^{-3} as suggested in the code provided by the authors.
- LDL-LCLR [Ren *et al.*, 2019b] captures the local and global correlations simultaneously via clustering and the low-rank approximation on all instances. In the experiments, λ_1, λ_2, λ_3, λ_4 and ρ are set to 10^{-4}, 10^{-3}, 10^{-3}, 10^{-3} and 1 respectively as suggested in the code provided by the authors.
- Adam-LDL-SCL [Jia *et al.*, 2021] exploits label correlations on local samples and uses the Adam algorithm [Kingma and Ba, 2015] to optimize the LDL model. In the experiments, the number of clusters is set to 14, the learning rate is set to 0.001, the maximum number of iterations is set to 300, the batch size is set to 50, the parameters λ_1, λ_2 and λ_3 are set to 10^{-3} uniformly.
- LDL-LDM [Wang and Geng, 2021] exploits both global and local label correlations in a data-driven way. Its basic idea is that label distribution lies in a probability simplex whose underlying structure may encode label correlation. In the experiments, λ_1, λ_2, λ_3 amd g are set to 10^{-3}, 1, 1 and 10 respectively as suggested in the official code.

A.3.2 Objective Function Reshaping Methods

Three objective function reshaping methods presented in this paper, i.e., Focal loss (OFR-FL), Class-balanced focal loss (OFR-CB) and Distribution-balanced focal loss (OFR-DB), are performed in the experiments. In these methods, f_θ is set to a three-layer neural network with ReLU function [Glorot *et al.*, 2011]. For the focal loss in these methods, we use $\gamma = 2$ with a balance parameter of 2.

A.3.3 Representation Distribution Alignment (RDA)

In RDA, g_ϵ and h_ϕ are set as linear projections, $\mathcal{F}_\vartheta$ is set as ReLU function [Glorot *et al.*, 2011] with linear projection, $\mathcal{G}_\varepsilon$ and $\mathcal{H}_\varphi$ are set as single-layer neural network with two outputs including mean and variance of Gaussian. Hyperparameters λ_1, λ_2 and λ_3 are selected by grid search from the set $\{0.01, 0.05, 0.1, 0.2, 0.5\}$. We adopt distribution-balanced focal loss (DB) as the loss function V in RDA.

A.3.4 Training Details

All methods are implemented by PyTorch. All the computations are performed on a GPU server with NVIDIA Tesla V100, Intel Xeon Gold 6240 CPU 2.60 GHz processor and 32 GB GPU memory. We sample training sets, validation sets and test sets for 10 times. For each algorithm, we select the model that obtains the optimal *Kullback-Leibler Divergence* on the validation set for testing.

A.4 Additional Results: More Evaluation Criteria

Tables 4 to 12 tabulate the experimental results of different algorithms on the six ILDL datasets evaluated by nine distribution criteria, and the best performance on each dataset is highlighted by boldface. For each evaluation metric, "↓" indicates the smaller the better while "↑" indicates the larger the better. The ranks are given in the parentheses right after the performance values. The average rank of each algorithm over all the datasets is also calculated and given in the last row of each table.

From Tables 4 to 12, it can be observed that: 1) The existing LDL algorithms show poor performances in solving ILDL tasks. 2) The performances of the objective function reshaping approaches are improved compared to existing algorithms. 3) Compared with these baseline methods, Our RDA has achieved better results in all criteria. This observation indicates that the RDA can effectively calibrate the distributions of feature representations and label representations of instances to leverage the information hidden in different spaces and improve the performance.

When looking at the average ranks over all the six real-world datasets, RDA achieves rather competitive performance over other algorithms. When compared with these state-of-the-art algorithms, RDA achieves 1st in 85.2% cases and 2nd in 11.1% cases. Thus, RDA possesses rather superior performance over the state-of-the-art algorithms across all the evaluation criteria.

Algorithm	Movie	SCUT-FBP	Emotion6	Flickr_LDL	RAF-ML	Natural Scene	Avg. rank
SA-BFGS	0.342±0.007(5)	0.727±0.033(10)	0.829±0.018(10)	0.895±0.010(10)	0.757±0.015(10)	0.662±0.020(10)	9.17
EDL-LRL	0.364±0.012(8)	0.352±0.024(6)	0.417±0.007(5)	0.581±0.006(6)	0.478±0.014(5)	0.434±0.023(5)	5.83
LDLSF	0.362±0.011(7)	0.470±0.031(9)	0.435±0.011(6)	0.570±0.009(5)	0.418±0.017(3)	0.444±0.025(6)	6.00
LDL-LCLR	0.335±0.007(4)	0.333±0.025(3)	0.524±0.014(9)	0.703±0.013(9)	0.385±0.011(2)	0.568±0.022(9)	6.00
Adam-LDL-SCL	0.718±0.049(10)	0.446±0.022(8)	0.471±0.033(7)	0.671±0.055(8)	0.585±0.030(9)	0.477±0.034(8)	8.33
LDL-LDM	0.486±0.029(9)	0.403±0.044(7)	0.474±0.016(8)	0.582±0.009(7)	0.535±0.028(8)	0.477±0.023(7)	7.67
OFR-FL	0.342±0.015(6)	0.336±0.036(4)	0.391±0.010(3)	0.564±0.005(3)	0.508±0.024(7)	0.432±0.020(3)	4.33
OFR-CB	0.334±0.018(3)	0.345±0.029(5)	0.392±0.009(4)	0.566±0.006(4)	0.506±0.016(6)	0.433±0.021(4)	4.33
OFR-DB	0.255±0.008(2)	0.320±0.038(2)	0.377±0.007(2)	0.525±0.021(2)	0.464±0.020(4)	0.387±0.025(2)	2.33
RDA (Ours)	**0.196±0.007(1)**	**0.285±0.016(1)**	**0.360±0.008(1)**	**0.521±0.007(1)**	**0.376±0.007(1)**	**0.377±0.021(1)**	1.00

Table 4: Experimental results on ILDL datasets measured by Chebyshev Distance ↓.

Algorithm	Movie	SCUT-FBP	Emotion6	Flickr_LDL	RAF-ML	Natural Scene	Avg. rank
SA-BFGS	1.164±0.024(6)	1.935±0.028(10)	2.227±0.027(10)	**1.914±0.064(1)**	2.030±0.014(9)	2.506±0.017(4)	6.67
EDL-LRL	1.186±0.030(7)	1.516±0.028(6)	1.900±0.024(5)	2.498±0.065(6)	1.746±0.034(4)	2.531±0.049(7)	5.83
LDLSF	1.271±0.027(8)	1.630±0.065(8)	1.935±0.034(6)	2.104±0.018(2)	1.707±0.035(3)	2.653±0.020(9)	6.00
LDL-LCLR	1.127±0.020(5)	1.373±0.024(3)	2.121±0.020(8)	2.658±0.005(10)	1.767±0.023(7)	**2.449±0.015(1)**	5.67
Adam-LDL-SCL	2.030±0.026(10)	1.688±0.104(9)	2.132±0.110(9)	2.353±0.258(3)	2.160±0.104(10)	2.667±0.111(10)	8.50
LDL-LDM	1.469±0.065(9)	1.518±0.056(7)	2.004±0.020(7)	2.537±0.013(9)	1.882±0.051(8)	2.557±0.017(8)	8.00
OFR-FL	1.113±0.044(4)	1.394±0.038(4)	1.826±0.033(3)	2.517±0.016(8)	1.758±0.036(6)	2.527±0.029(6)	5.17
OFR-CB	1.102±0.048(3)	1.398±0.032(5)	1.827±0.028(4)	2.516±0.019(7)	1.749±0.030(5)	2.524±0.026(5)	4.83
OFR-DB	0.915±0.014(2)	1.350±0.039(2)	1.743±0.021(2)	2.409±0.028(5)	1.695±0.025(2)	2.493±0.030(3)	2.67
RDA (Ours)	**0.796±0.016(1)**	**1.320±0.024(1)**	**1.699±0.023(1)**	2.384±0.013(4)	**1.602±0.015(1)**	2.486±0.026(2)	1.67

Table 5: Experimental results on ILDL datasets measured by Clark Distance ↓.

Algorithm	Movie	SCUT-FBP	Emotion6	Flickr_LDL	RAF-ML	Natural Scene	Avg. rank
SA-BFGS	2.357±0.052(6)	4.240±0.083(10)	5.618±0.178(10)	**4.100±0.613(1)**	4.733±0.038(9)	7.107±0.064(6)	7.00
EDL-LRL	2.412±0.074(7)	2.988±0.080(6)	4.526±0.072(6)	6.955±0.044(9)	3.872±0.090(5)	7.111±0.187(7)	6.67
LDLSF	2.572±0.062(8)	3.244±0.180(8)	4.422±0.120(5)	4.851±0.065(2)	3.555±0.100(2)	7.494±0.092(9)	5.67
LDL-LCLR	2.298±0.048(5)	2.645±0.070(3)	5.241±0.064(8)	7.388±0.026(10)	3.865±0.065(4)	**6.882±0.055(1)**	5.17
Adam-LDL-SCL	4.438±0.080(10)	3.433±0.242(9)	5.459±0.384(9)	6.293±1.178(4)	5.186±0.235(10)	7.545±0.418(10)	8.67
LDL-LDM	3.010±0.151(9)	3.019±0.139(7)	4.873±0.061(7)	6.892±0.045(8)	4.267±0.150(8)	7.273±0.071(8)	7.83
OFR-FL	2.296±0.099(4)	2.718±0.113(4)	4.321±0.111(3)	6.803±0.067(6)	3.905±0.098(7)	7.103±0.115(5)	4.83
OFR-CB	2.270±0.115(3)	2.732±0.096(5)	4.321±0.085(3)	6.804±0.062(7)	3.884±0.081(6)	7.097±0.106(4)	4.67
OFR-DB	1.837±0.034(2)	2.576±0.113(2)	4.056±0.051(2)	6.354±0.099(5)	3.760±0.066(3)	6.962±0.124(3)	2.83
RDA (Ours)	**1.549±0.036(1)**	**2.479±0.057(1)**	**3.893±0.063(1)**	6.232±0.048(3)	**3.496±0.041(1)**	6.932±0.111(2)	1.50

Table 6: Experimental results on ILDL datasets measured by Canberra Metric ↓.

Algorithm	Movie	SCUT-FBP	Emotion6	Flickr_LDL	RAF-ML	Natural Scene	Avg. rank
SA-BFGS	0.801±0.054(7)	13.042±4.101(10)	21.851±1.052(10)	27.126±1.551(10)	18.205±1.202(10)	4.798±0.373(8)	9.17
EDL-LRL	0.780±0.047(6)	0.811±0.108(6)	1.435±0.116(5)	9.914±4.576(7)	1.284±0.099(4)	2.586±1.583(6)	5.67
LDLSF	3.134±0.379(9)	8.414±1.657(9)	9.437±0.506(9)	12.851±1.051(8)	7.068±1.141(9)	8.845±0.559(9)	8.83
LDL-LCLR	0.680±0.031(5)	0.603±0.079(3)	2.282±0.158(7)	6.217±0.290(6)	1.011±0.070(2)	2.945±0.253(7)	5.00
Adam-LDL-SCL	19.172±1.630(10)	2.377±1.173(8)	8.112±4.890(8)	17.194±8.519(9)	6.117±4.256(8)	9.621±4.899(10)	8.83
LDL-LDM	1.812±0.279(8)	1.025±0.219(7)	1.789±0.137(6)	2.742±0.210(5)	1.916±0.225(7)	1.775±0.206(5)	6.33
OFR-FL	0.646±0.057(4)	0.641±0.144(4)	1.183±0.096(3)	2.599±0.165(3)	1.367±0.168(6)	1.336±0.098(4)	4.00
OFR-CB	0.629±0.060(3)	0.658±0.117(5)	1.190±0.078(4)	2.628±0.377(4)	1.326±0.111(5)	1.328±0.093(3)	4.00
OFR-DB	0.388±0.016(2)	0.558±0.132(2)	0.924±0.024(2)	1.775±0.286(2)	1.148±0.082(3)	1.175±0.090(2)	2.17
RDA (Ours)	**0.249±0.015(1)**	**0.431±0.033(1)**	**0.768±0.022(1)**	**1.607±0.111(1)**	**0.706±0.020(1)**	**1.119±0.059(1)**	1.00

Table 7: Experimental results on ILDL datasets measured by Kullback-Leibler Divergence ↓.

Algorithm	Movie	SCUT-FBP	Emotion6	Flickr_LDL	RAF-ML	Natural Scene	Avg. rank
SA-BFGS	0.612±0.011(4)	0.337±0.035(10)	0.253±0.024(10)	0.149±0.006(10)	0.336±0.022(9)	0.334±0.026(10)	8.83
EDL-LRL	0.589±0.018(6)	0.679±0.041(4)	0.467±0.017(5)	0.219±0.008(8)	0.429±0.019(4)	0.505±0.034(4)	5.17
LDLSF	0.573±0.015(8)	0.471±0.050(8)	0.459±0.020(6)	0.290±0.017(3)	0.570±0.032(3)	0.457±0.029(6)	5.67
LDL-LCLR	0.613±0.012(3)	0.705±0.042(3)	0.408±0.018(7)	0.224±0.016(7)	**0.678±0.016(1)**	0.384±0.026(9)	5.00
Adam-LDL-SCL	0.398±0.092(10)	0.438±0.038(9)	0.331±0.031(9)	0.156±0.016(9)	0.302±0.016(10)	0.420±0.042(8)	9.17
LDL-LDM	0.521±0.040(9)	0.616±0.067(7)	0.399±0.026(8)	0.238±0.018(4)	0.373±0.033(8)	0.446±0.029(7)	7.17
OFR-FL	0.589±0.021(6)	0.674±0.072(5)	0.482±0.030(3)	0.229±0.012(5)	0.386±0.037(7)	0.506±0.028(3)	4.83
OFR-CB	0.598±0.027(5)	0.659±0.061(6)	0.480±0.024(4)	0.226±0.008(6)	0.390±0.024(6)	0.504±0.026(5)	5.33
OFR-DB	0.709±0.011(2)	0.715±0.082(2)	0.543±0.015(2)	0.350±0.060(2)	0.422±0.030(5)	0.575±0.026(2)	2.50
RDA (Ours)	**0.816±0.012(1)**	**0.795±0.022(1)**	**0.622±0.009(1)**	**0.388±0.024(1)**	0.659±0.012(2)	**0.597±0.017(1)**	1.17

Table 8: Experimental results on ILDL datasets measured by Cosine Coefficient ↑.

Algorithm	Movie	SCUT-FBP	Emotion6	Flickr_LDL	RAF-ML	Natural Scene	Avg. rank
SA-BFGS	0.546±0.010(4)	0.244±0.026(10)	0.154±0.015(10)	0.099±0.005(10)	0.228±0.016(10)	0.247±0.020(10)	9.00
EDL-LRL	0.529±0.016(7)	0.573±0.032(4)	0.417±0.012(5)	0.209±0.005(7)	0.395±0.013(4)	0.394±0.025(3)	5.00
LDLSF	0.514±0.012(8)	0.416±0.038(8)	0.397±0.017(6)	0.253±0.013(3)	0.476±0.026(3)	0.359±0.021(6)	5.67
LDL-LCLR	0.550±0.010(3)	0.593±0.031(2)	0.347±0.014(8)	0.186±0.012(8)	**0.559±0.012(1)**	0.301±0.017(9)	5.17
Adam-LDL-SCL	0.240±0.049(10)	0.395±0.031(9)	0.295±0.031(9)	0.141±0.015(9)	0.259±0.016(9)	0.345±0.034(7)	8.83
LDL-LDM	0.437±0.031(9)	0.521±0.050(7)	0.361±0.015(7)	0.217±0.012(6)	0.339±0.027(8)	0.343±0.022(8)	7.50
OFR-FL	0.534±0.018(6)	0.569±0.049(5)	0.430±0.022(3)	0.219±0.010(4)	0.368±0.027(7)	0.390±0.022(4)	4.83
OFR-CB	0.540±0.023(5)	0.559±0.043(6)	0.430±0.017(4)	0.218±0.006(5)	0.373±0.018(6)	0.389±0.021(5)	5.17
OFR-DB	0.631±0.008(2)	0.577±0.052(3)	0.469±0.010(2)	0.288±0.031(2)	0.395±0.018(5)	0.396±0.020(2)	2.67
RDA (Ours)	**0.709±0.009(1)**	**0.630±0.014(1)**	**0.512±0.008(1)**	**0.306±0.011(1)**	0.509±0.008(2)	**0.403±0.018(1)**	1.17

Table 9: Experimental results on ILDL datasets measured by Intersection Similarity ↑.

Algorithm	Movie	SCUT-FBP	Emotion6	Flickr_LDL	RAF-ML	Natural Scene	Avg. rank
SA-BFGS	0.473±0.010(5)	0.914±0.034(10)	0.992±0.018(10)	1.108±0.009(10)	0.936±0.017(10)	0.834±0.028(10)	9.17
EDL-LRL	0.495±0.015(7)	0.473±0.034(5)	0.572±0.010(5)	0.779±0.009(6)	0.652±0.018(5)	0.582±0.026(5)	5.50
LDLSF	0.502±0.013(8)	0.643±0.043(9)	0.602±0.015(6)	0.771±0.013(5)	0.565±0.026(3)	0.611±0.026(6)	6.17
LDL-LCLR	0.466±0.010(3)	0.449±0.035(3)	0.692±0.017(9)	0.919±0.016(9)	0.502±0.014(2)	0.736±0.025(9)	5.83
Adam-LDL-SCL	0.856±0.054(10)	0.639±0.030(8)	0.672±0.039(8)	0.914±0.054(8)	0.806±0.026(9)	0.649±0.044(8)	8.50
LDL-LDM	0.618±0.033(9)	0.539±0.058(7)	0.643±0.019(7)	0.779±0.012(6)	0.721±0.034(8)	0.634±0.026(7)	7.33
OFR-FL	0.479±0.017(6)	0.469±0.050(4)	0.540±0.018(3)	0.755±0.011(3)	0.691±0.033(7)	0.574±0.021(3)	4.33
OFR-CB	0.472±0.022(4)	0.481±0.045(6)	0.541±0.014(4)	0.760±0.009(4)	0.687±0.022(6)	0.575±0.023(4)	4.67
OFR-DB	0.377±0.008(2)	0.443±0.058(2)	0.503±0.009(2)	0.666±0.027(2)	0.636±0.026(4)	0.526±0.024(2)	2.33
RDA (Ours)	**0.295±0.010(1)**	**0.386±0.017(1)**	**0.464±0.008(1)**	**0.645±0.010(1)**	**0.484±0.007(1)**	**0.515±0.020(1)**	1.00

Table 10: Experimental results on ILDL datasets measured by Euclidean Distance on *all* labels ↓.

Algorithm	Movie	SCUT-FBP	Emotion6	Flickr_LDL	RAF-ML	Natural Scene	Avg. rank
SA-BFGS	0.346±0.010(6)	0.760±0.044(10)	0.888±0.018(10)	0.617±0.013(10)	0.849±0.014(10)	0.736±0.026(10)	9.33
EDL-LRL	0.360±0.014(8)	0.323±0.018(3)	0.508±0.012(5)	0.567±0.006(7)	0.490±0.008(3)	0.511±0.023(5)	5.17
LDLSF	0.359±0.009(7)	0.442±0.023(8)	0.540±0.016(6)	0.552±0.018(3)	0.490±0.028(3)	0.565±0.025(7)	5.67
LDL-LCLR	0.337±0.008(5)	0.318±0.019(2)	0.625±0.021(9)	0.599±0.012(9)	**0.438±0.013(1)**	0.672±0.023(9)	5.83
Adam-LDL-SCL	0.676±0.079(10)	0.461±0.025(9)	0.618±0.034(8)	0.579±0.006(8)	0.587±0.027(9)	0.563±0.037(6)	8.33
LDL-LDM	0.484±0.019(9)	0.420±0.058(7)	0.583±0.020(7)	0.563±0.010(6)	0.555±0.016(8)	0.578±0.027(8)	7.50
OFR-FL	0.336±0.010(4)	0.337±0.034(5)	0.495±0.012(3)	0.556±0.007(5)	0.496±0.016(6)	0.499±0.023(3)	4.33
OFR-CB	0.334±0.014(3)	0.342±0.032(6)	0.497±0.011(4)	0.555±0.011(4)	0.491±0.010(5)	0.499±0.023(3)	4.17
OFR-DB	0.285±0.007(2)	0.332±0.036(4)	0.465±0.009(2)	**0.524±0.030(1)**	0.498±0.009(7)	0.492±0.024(2)	3.00
RDA (Ours)	**0.245±0.006(1)**	**0.299±0.016(1)**	**0.429±0.008(1)**	0.526±0.011(2)	0.454±0.007(2)	**0.486±0.020(1)**	1.33

Table 11: Experimental results on ILDL datasets measured by Euclidean Distance on *tail* labels ↓.

Algorithm	Movie	SCUT-FBP	Emotion6	Flickr_LDL	RAF-ML	Natural Scene	Avg. rank
SA-BFGS	0.303±0.011(4)	0.355±0.073(8)	0.269±0.037(10)	0.864±0.018(10)	0.221±0.037(4)	0.257±0.022(9)	7.50
EDL-LRL	0.315±0.017(6)	0.326±0.038(6)	0.214±0.016(8)	0.501±0.012(7)	0.416±0.022(6)	0.242±0.011(6)	6.50
LDLSF	0.337±0.014(8)	0.447±0.051(10)	0.198±0.034(5)	0.492±0.018(5)	0.218±0.021(3)	0.181±0.013(3)	5.67
LDL-LCLR	0.309±0.010(5)	0.298±0.036(4)	0.201±0.033(6)	0.632±0.023(8)	0.177±0.018(2)	0.211±0.012(5)	5.00
Adam-LDL-SCL	0.378±0.186(10)	0.435±0.019(9)	0.225±0.017(9)	0.679±0.069(9)	0.527±0.049(10)	0.297±0.027(10)	9.50
LDL-LDM	0.297±0.053(3)	0.286±0.030(3)	0.203±0.024(7)	0.498±0.014(6)	0.416±0.051(6)	0.197±0.017(4)	4.83
OFR-FL	0.338±0.017(9)	0.315±0.040(5)	0.190±0.027(3)	0.480±0.014(3)	0.473±0.034(8)	0.253±0.011(7)	5.83
OFR-CB	0.329±0.020(7)	0.330±0.033(7)	0.190±0.014(3)	0.488±0.015(4)	0.473±0.024(8)	0.256±0.016(8)	6.17
OFR-DB	0.243±0.009(2)	0.284±0.049(2)	0.163±0.012(2)	0.374±0.022(2)	0.384±0.036(5)	0.162±0.008(2)	2.50
RDA (Ours)	**0.157±0.010(1)**	**0.234±0.013(1)**	**0.133±0.004(1)**	**0.333±0.005(1)**	**0.132±0.006(1)**	**0.151±0.003(1)**	1.00

Table 12: Experimental results on ILDL datasets measured by Euclidean Distance on *head* labels ↓.

References

[Boutell *et al.*, 2004] Matthew R. Boutell, Jiebo Luo, Xipeng Shen, and Christopher M. Brown. Learning multi-label scene classification. *Pattern Recognit.*, 37(9):1757–1771, 2004.

[Ekman, 1992] Paul Ekman. An argument for basic emotions. *Cognition & emotion*, 6(3-4):169–200, 1992.

[Geng and Luo, 2014] Xin Geng and Longrun Luo. Multilabel ranking with inconsistent rankers. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3742–3747, 2014.

[Geng, 2016] Xin Geng. Label distribution learning. *IEEE Trans. Knowl. Data Eng.*, 28(7):1734–1748, 2016.

[Glorot *et al.*, 2011] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics*, pages 315–323, 2011.

[He *et al.*, 2016] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.

[Jia *et al.*, 2019] Xiuyi Jia, Xiang Zheng, Weiwei Li, Changqing Zhang, and Zechao Li. Facial emotion distribution learning by exploiting low-rank label correlations locally. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9841–9850, 2019.

[Jia *et al.*, 2021] Xiuyi Jia, Zechao Li, Xiang Zheng, Weiwei Li, and Sheng-Jun Huang. Label distribution learning with label correlations on local samples. *IEEE Trans. Knowl. Data Eng.*, 33(4):1619–1631, 2021.

[King, 2009] Davis E. King. Dlib-ml: A machine learning toolkit. *J. Mach. Learn. Res.*, 10:1755–1758, 2009.

[Kingma and Ba, 2015] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *Proceedings of the 3rd International Conference on Learning Representations*, 2015.

[Mikels *et al.*, 2005] Joseph A Mikels, Barbara L Fredrickson, Gregory R Larkin, Casey M Lindberg, Sam J Maglio, and Patricia A Reuter-Lorenz. Emotional category data on images from the international affective picture system. *Behavior research methods*, 37(4):626–630, 2005.

[Peng *et al.*, 2015] Kuan-Chuan Peng, Tsuhan Chen, Amir Sadovnik, and Andrew C. Gallagher. A mixed bag of emotions: Model, predict, and transfer emotion distributions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 860–868, 2015.

[Ren *et al.*, 2019a] Tingting Ren, Xiuyi Jia, Weiwei Li, Lei Chen, and Zechao Li. Label distribution learning with label-specific features. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, pages 3318–3324, 2019.

[Ren *et al.*, 2019b] Tingting Ren, Xiuyi Jia, Weiwei Li, and Shu Zhao. Label distribution learning with label correlations via low-rank approximation. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, pages 3325–3331, 2019.

[Russakovsky *et al.*, 2015] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael S. Bernstein, Alexander C. Berg, and Li Fei-Fei. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.*, 115(3):211–252, 2015.

[Shang and Deng, 2019] Li Shang and Weihong Deng. Blended emotion in-the-wild: Multi-label facial expression recognition using crowdsourced annotations and deep locality feature learning. *Int. J. Comput. Vis.*, 127(6-7):884–906, 2019.

[Wang and Geng, 2021] Jing Wang and Xin Geng. Label distribution learning by exploiting label distribution manifold. *IEEE transactions on neural networks and learning systems*, 2021.

[Xie *et al.*, 2015] Duorui Xie, Lingyu Liang, Lianwen Jin, Jie Xu, and Mengru Li. SCUT-FBP: A benchmark dataset for facial beauty perception. In *Proceedings of the 2015 IEEE International Conference on Systems, Man, and Cybernetics*, pages 1821–1826, 2015.

[Yang *et al.*, 2017] Jufeng Yang, Ming Sun, and Xiaoxiao Sun. Learning visual sentiment distributions via augmented conditional probability neural network. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence*, pages 224–230, 2017.