RankMatch: A Novel Approach to Semi-Supervised Label Distribution Learning Leveraging Rank Correlation between Labels

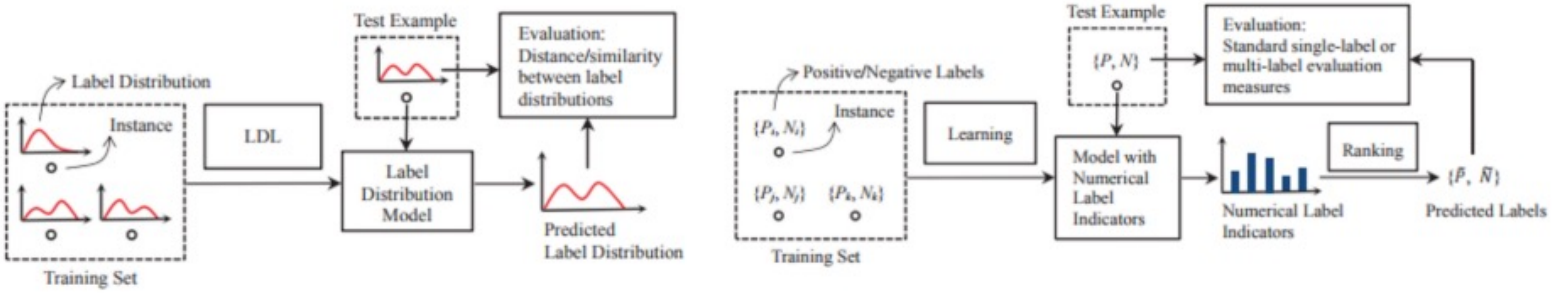
Zhiqiang Kou[1,2], Yucheng Xie[1,2], Hailin Wang[5], Junyang Chen[1,2], Jing Wang[1,2], Ming-kun Xie[3], Shuo Chen[3], Yuheng Jia[1,2]*, Tongliang Liu[4], Xin Geng[1,2]*

[1]School of Computer Science and Engineering, Southeast University, China
[2]Key Laboratory of New Generation Artificial Intelligence Technology and Its Interdisciplinary Applications (Southeast University), Ministry of Education, China
[3]RIKEN Center for Advanced Intelligence Project (AIP), Tokyo, Japan
[4]Sydney AI Centre, The University of Sydney, Australia
[5]School of Mathematics and Statistics, Xian Jiaotong University, China

{zhiqiang_kou, xieyc, chenjunyang, wangjing91, yhjia, xgeng}@seu.edu.cn,
{ming-kun.xie, shuo.chen.ya}@riken.jp, wanghailin97@163.com,
tongliang.liu@sydney.edu.au

When an instance is associated with multiple labels at the same time, those labels are seldom equally important for that instance; instead, they typically differ in their relative importance or priority.

(a)



(b) LDL

(c) other supervised learning Paradigm

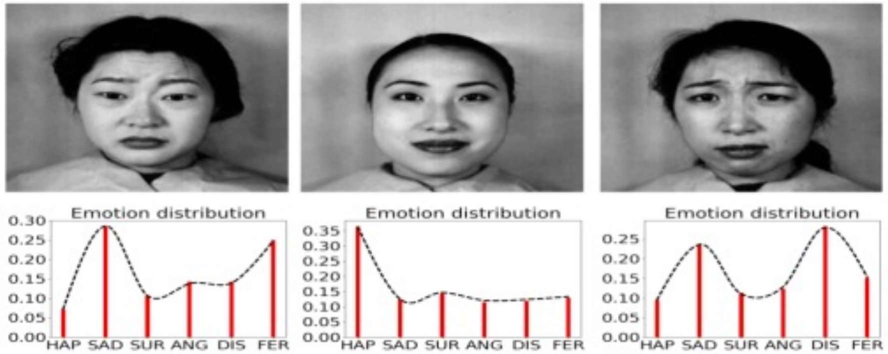
Figure 1 introduces single-label learning, multi-label learning, and label distribution learning.

[1] Geng, X., 2016. Label Distribution Learning. IEEE Transactions on Knowledge and Data Engineering, 28(7): 1734–1748

Use LDL for Mars composition prediction

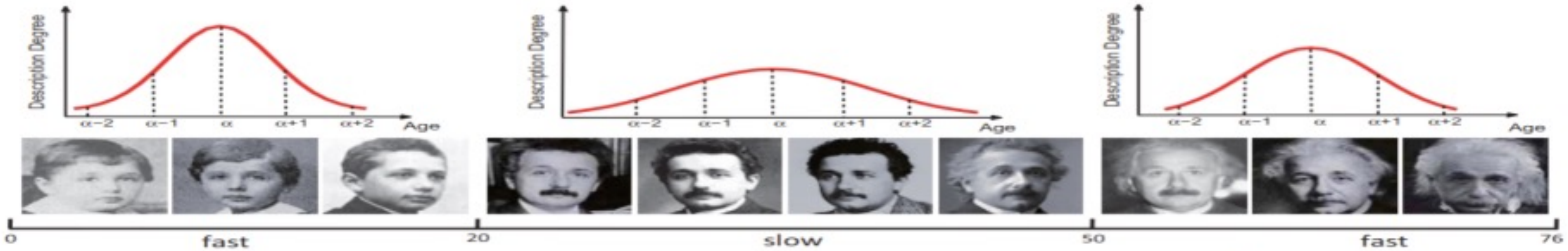
NASA [1] uses LDL to predict trace elements in Martian craters.

Using LDL for Facial Expression Recognition



Using LDL for more accurate facial emotion representation [2].

Using LDL for Facial Age Estimation



Using LDL to better capture the gradual aging process [3].

[1] Chow B J. Forming infrastructural materials by mechanical compaction of lunar and Martian regolith simulants[M]. University of California, San Diego, 2016.
[2] Geng X, Zhou Z H, Smith-Miles K. Automatic age estimation based on facial aging patterns[J]. IEEE Transactions on pattern analysis and machine intelligence, 2007, 29(12): 2234-2240.
[3] Geng X, Wang Q, Xia Y. Facial age estimation by adaptive label distribution learning[C]. 2014 22nd International Conference on Pattern Recognition. IEEE, 2014: 4465-4470.

Challenge： Obtaining label distribution is very expensive and time-consuming.

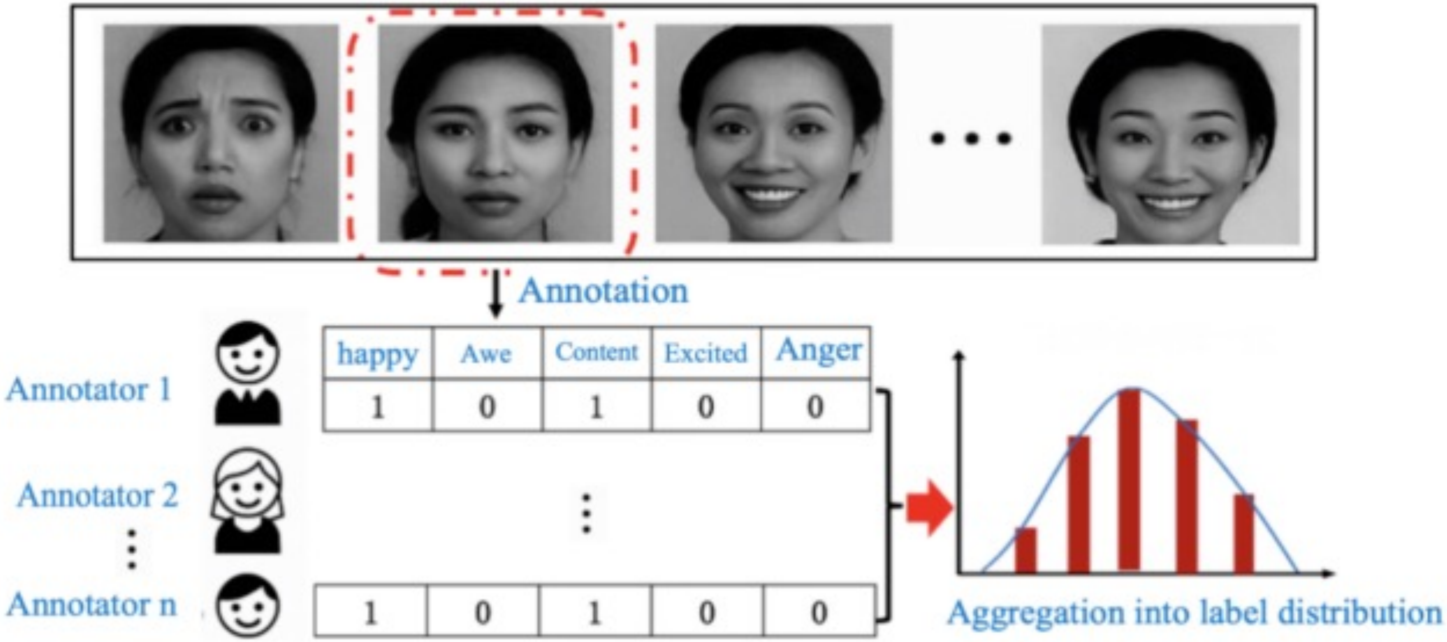
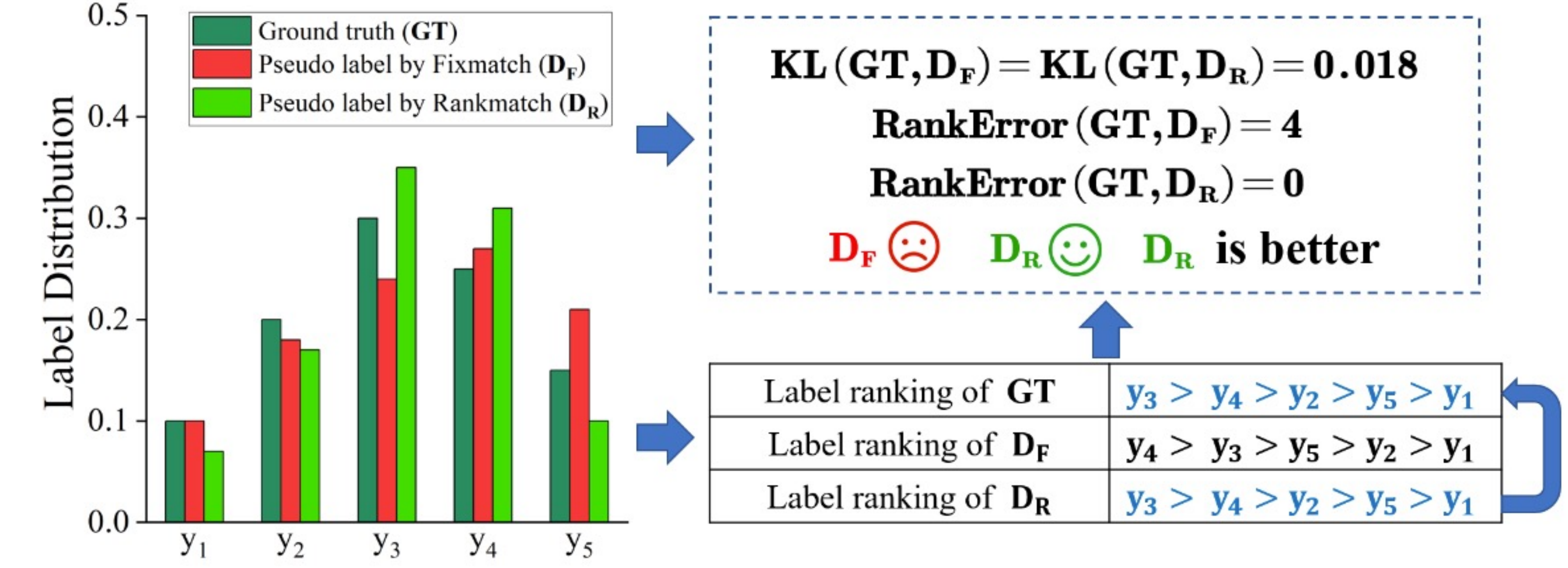


Fig. The process of labeling the label distribution.

Due to the high expertise required for annotation and the substantial labeling cost, real-world datasets often contain few labeled samples.

RankMatch: A Novel Approach to Semi-Supervised Label Distribution Learning Leveraging Rank Correlation between Labels



Existing semi-supervised techniques often fail in Semi-Supervised Label Distribution Learning (SSLDL) because, during training on labeled data, the model focuses solely on minimizing the overall similarity between the predicted and true label distributions (e.g., using KL divergence as the loss function), without learning the rank correlation among labels. This neglect of label ordering leads to biased pseudo-labels.

Problem Statement and Notation

In SSLDL, the training data consists of a labeled dataset $\mathcal{D}_L = \{(\mathbf{x}_i, \mathbf{d}_i)|i = 1, 2, ..., n\}$ and an unlabeled dataset $\mathcal{D}_U = \{\mathbf{x}_g|g = 1, 2, ..., m\}$. Here, n and m represent the number of labeled and unlabeled samples, respectively. In the labeled dataset $\mathcal{D}_L$, $\mathbf{x}_i$ is a labeled sample, and $\mathbf{d}_i = \{d_{\mathbf{x}_i}^{y_1}, d_{\mathbf{x}_i}^{y_2}, ..., d_{\mathbf{x}_i}^{y_c}\}$ is the corresponding label distribution, where $d_{\mathbf{x}_i}^{y_j}$ represents the importance or relevance of label y_j to sample $\mathbf{x}_i$. The label distribution satisfies the normalization constraint $\sum_{j=1}^{c} d_{\mathbf{x}_i}^{y_j} = 1$. c denotes the number of labels in the label space $\mathcal{Y} = \{y_1, y_2, ..., y_c\}$.

Supervised training phase . First, this work adopts the KL divergence as the loss function, where $\mathrm{Aug}_{w}(\boldsymbol{X}_i)$ denotes the data augmentation applied to the input.

$$\mathcal{L}_s = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{c} d_{\mathbf{x}_i}^{y_j} \ln \left(\frac{d_{\mathbf{x}_i}^{y_j}}{h(y_j \mid \mathrm{Aug}_w(\mathbf{x}_i); \theta)} \right)$$

$$h(y_j | \mathbf{x}_i; \theta) = \frac{\exp(f_j(\mathbf{x}_i; \theta))}{\sum_{q=1}^{c} \exp(f_q(\mathbf{x}_i; \theta))}$$

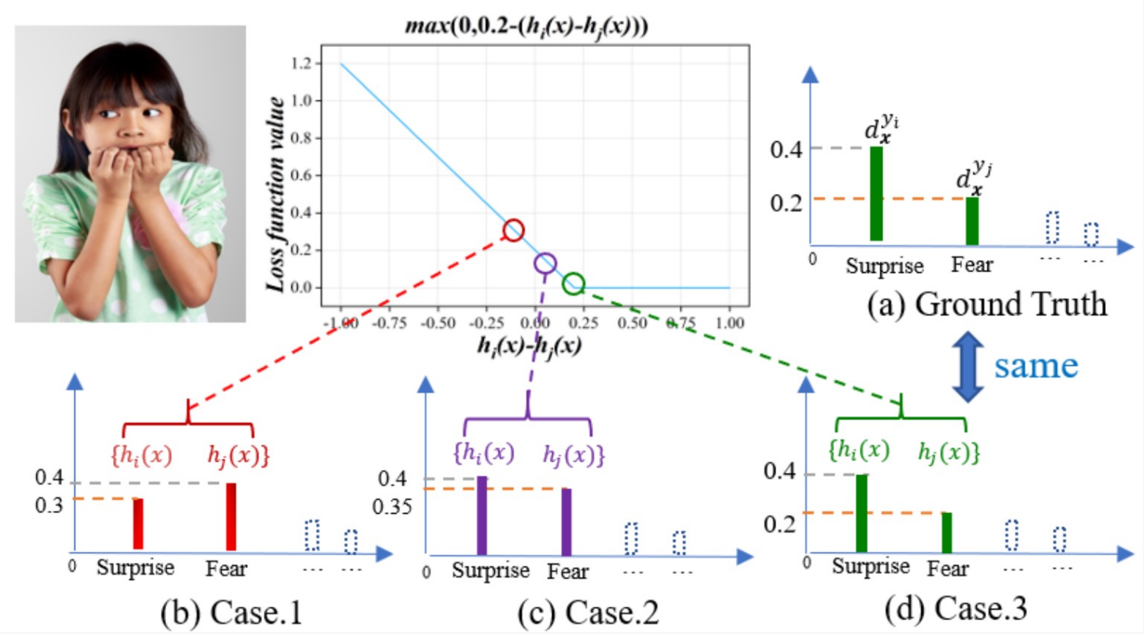
Supervised training phase: To generate more reliable pseudo-labels, we propose a pairwise relevance ranking loss:

$$\mathcal{L}_{PRR_L} = \sum_{1<j<k<c} \Big(s(j,k) \cdot g_\delta(j,k) + s(k,j) \cdot g_\delta(k,j) \Big),$$

Here, $\delta = d_{\mathbf{x}_i}^{y_j} - d_{\mathbf{x}_i}^{y_k}$, and $f(j,k)$ and $g_\delta(j,k)$ are defined as follows:

$$s(j,k) = \begin{cases} 1, & \text{if } d_{\mathbf{x}_i}^{y_j} > d_{\mathbf{x}_i}^{y_k} \text{ and } d_{\mathbf{x}_i}^{y_j} - d_{\mathbf{x}_i}^{y_k} > t, \\ 0, & \text{otherwise.} \end{cases}$$

$$g_\delta(j,k) = \begin{cases} 0, & \text{if } h_j(\mathbf{x}_i) - h_k(\mathbf{x}_i) \geq \delta, \\ \delta - (h_j(\mathbf{x}_i) - h_k(\mathbf{x}_i)), & \text{otherwise.} \end{cases}$$



Unsupervised training phase: An ensemble learning approach is used to assign pseudo-labels to unlabeled samples.

$$\mathbf{p}_i(y_j) = \frac{\exp\left(\frac{1}{H}\sum_{k=1}^{H} f_j(\mathrm{Aug}_w(\mathbf{x})_k;\theta)\right)}{\sum_{q=1}^{c}\exp\left(\frac{1}{H}\sum_{k=1}^{H} f_q(\mathrm{Aug}_w(\mathbf{x})_k;\theta)\right)},$$

The average of the model outputs from multiple weak augmentations of unlabeled images is used as the pseudo-label distribution. Then, KL divergence is used as the loss function.

$$\mathcal{L}_{uc} = \frac{1}{m}\sum_{u=1}^{m}\sum_{j=1}^{c} p_{\mathbf{x}_u}^{y_j} \ln\left(\frac{p_{\mathbf{x}_u}^{y_j}}{h\left(y_j \mid \mathrm{Aug}_s(\mathbf{x}_u);\theta\right)}\right),$$

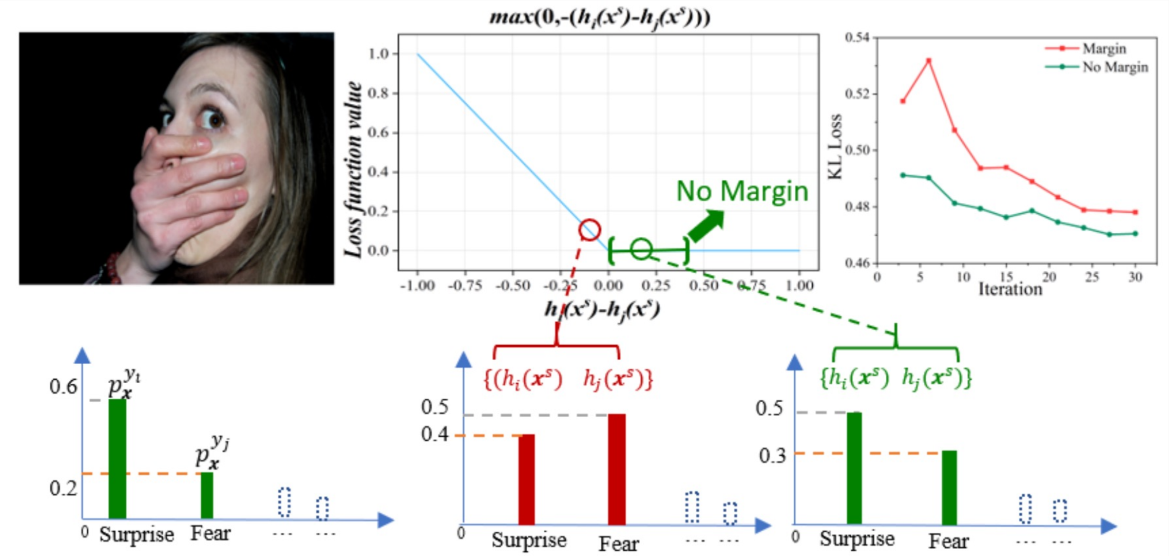
Unsupervised training phase: flexibly utilizing the ranking relationships between pseudo-labels

$$\mathcal{L}_{PRR_u} = \sum_{1<j<k<c} \Big(s(j,k) \cdot g_0(j,k) + s(k,j) \cdot g_0(k,j) \Big),$$

where:

$$s(j,k) = \begin{cases} 1, & \text{if } p_{\mathbf{x}_i}^{y_j} >_{\mathbf{x}_i}^{y_k} \text{ and } p_{\mathbf{x}_i}^{y_j} - p_{\mathbf{x}_i}^{y_k} > t, \\ 0, & \text{otherwise.} \end{cases}$$

$$g_0(j,k) = \begin{cases} 0, & \text{if } h_j(\mathbf{x}_i^s) - h_k(\mathbf{x}_i^s) \geq 0, \\ h_k(\mathbf{x}_i^s) - h_j(\mathbf{x}_i^s), & \text{otherwise.} \end{cases}$$



(a) The pseudo-label distributions (b) Case.1 (c) Case.2

A collection of numbers: A collection of four LDL images and emotions.

Twitter-LDL: A large-scale visual sentiment distribution dataset built from Twitter, containing eight distinct emotions: pleasure, anger, awe, satisfaction, disgust, excitement, fear, and sadness. Approximately 30,000 images were collected by searching various sentiment keywords such as "sadness," "heartbreak," and "grief." Eight annotators were then hired to label the dataset. The final Twitter LDL dataset contains 10,045 images.

Flickr-LDL: A subset of the Flickr dataset. Unlike other datasets that use sentiment words to search for images, the Flickr dataset collects 1200 adjective-noun pairs, totaling 500,000 images. We hired 11 annotators to label this subset with eight common emotions. The final Flickr LDL dataset contains 10,700 images, with roughly equal numbers of images in each category.

Emotion6: We used six category keywords and their synonyms as search terms, collecting 1,980 images from Flickr for Emotion6. 330 images were collected for each category, and each image was assigned to only one category (dominant sentiment). Emotion6 represents the sentiment associated with each image as a probability distribution, containing 7 intervals, including Ekman's 6 basic sentiments and neutrality.

RAF-LDL: RAF-LDL is a multi-labeled facial expression dataset containing approximately 5,000 diverse facial images downloaded from the internet. These images vary in sentiment, subject identity, head pose, lighting conditions, and occlusion. During the annotation process, 315 trained annotators were employed to ensure each image was annotated a sufficient number of times independently. Images with a multi-peaked label distribution were selected to constitute RAF-LDL.

Experimental results

		Emotion6			Flickr-LDL			Twitter-LDL			RAF-LDL		
	Method	10%	20%	40%	10%	20%	40%	10%	20%	40%	10%	20%	40%
Can.↓	Rankmatch	**3.3902**	**3.3176**	3.2504	**4.4060**	**3.9964**	**3.9013**	**3.7370**	**3.6962**	**3.2913**	**3.0178**	**2.9358**	**2.8341**
	SSMLL-CAP	3.7951	3.7613	3.7248	5.3827	5.3235	5.2676	5.8983	5.7659	5.6366	3.4385	3.2808	3.1966
	PCLP	3.7011	3.6017	3.6030	5.2781	5.2292	5.1966	5.4909	5.3738	5.4133	3.3696	3.3383	3.3310
	Fixmatch-LDL	3.5080	3.5680	3.6050	5.5570	5.5310	5.4350	6.1750	6.0060	5.8340	3.1220	3.0920	3.0770
	Mixmatch-LDL	3.6080	3.4860	3.4880	5.6450	5.5026	5.5750	6.3530	6.2489	6.2960	3.1580	3.1111	3.0630
	GCT-LDL	3.5980	3.5490	3.6410	5.5860	5.5872	5.5260	6.3010	6.3078	6.2380	3.1920	3.1260	3.1470
	SALDL	3.4836	3.3737	**3.1931**	5.4612	4.7789	4.8199	5.0380	4.0868	4.0742	3.1947	3.1415	3.0527
	sLDLF	4.4164	4.3398	4.1322	6.2280	6.1238	6.2589	5.3084	6.0008	6.1910	4.0586	4.1705	4.1189
	DF-LDL	4.2427	4.0717	3.7221	5.5348	5.5549	5.5207	6.4184	6.3120	6.2588	3.3281	3.3865	3.3582
	LDL-LRR	4.6528	4.0496	3.7719	5.6325	5.4988	5.4319	6.4215	6.3295	6.2905	3.8677	4.0116	4.1890
	Adam-LDL-SCL	4.0815	4.1128	4.1204	6.1634	5.9889	5.6508	6.5220	6.4081	6.3575	3.0891	3.0242	2.9912

Experimental results

		Emotion6			Flickr-LDL			Twitter-LDL			RAF-LDL		
	Method	10%	20%	40%	10%	20%	40%	10%	20%	40%	10%	20%	40%
	Rankmatch	**1.5298**	**1.5050**	**1.4834**	**1.8189**	**1.7051**	**1.6737**	**1.6480**	**1.6190**	**1.5138**	**1.4506**	**1.4190**	**1.3843**
	SSMLL-CAP	1.6705	1.6611	1.6502	2.1222	2.0988	2.0820	2.2590	2.2155	2.1733	1.5918	1.5332	1.5082
	PCLP	1.6397	1.6059	1.6083	2.0601	2.0478	2.0328	2.1002	2.0623	2.0728	1.5689	1.5636	1.5593
	Fixmatch-LDL	1.5950	1.6230	1.6390	2.2220	2.2110	2.1910	2.3830	2.3310	2.2820	1.5130	1.5060	1.5050
Cla. ↓	Mixmatch-LDL	1.6240	1.5810	1.5840	2.2330	2.1996	2.2160	2.4280	2.4034	2.4150	1.5150	1.5020	1.4870
	GCT-LDL	1.6090	1.6050	1.6390	2.2200	2.2238	2.2080	2.4170	2.4216	2.4060	1.5350	1.5170	1.5290
	SALDL	1.6019	1.5751	1.5100	2.1967	2.0369	2.0446	2.1288	1.8938	1.8964	1.5445	1.5288	1.5035
	sLDLF	1.8922	1.8566	1.8049	2.3722	2.3436	2.3761	2.1480	2.3384	2.3746	1.9300	1.9645	1.9750
	DF-LDL	1.8217	1.7746	1.6781	2.2253	2.2072	2.1992	2.4313	2.4108	2.4033	1.6071	1.6229	1.6138
	LDL-LRR	1.9899	1.7745	1.6953	2.2285	2.2026	2.1919	2.4429	2.4223	2.4121	1.7907	1.8298	1.8919
	Adam-LDL-SCL	1.7851	1.7976	1.8014	2.3534	2.3093	2.2312	2.4639	2.4324	2.4160	1.5134	1.4980	1.4905

Experimental results

		Emotion6			Flickr-LDL			Twitter-LDL			RAF-LDL		
	Method	10%	20%	40%	10%	20%	40%	10%	20%	40%	10%	20%	40%
Int. ↑	Rankmatch	**0.6735**	**0.6832**	**0.6940**	**0.6921**	**0.7073**	**0.7151**	**0.7036**	**0.7190**	**0.7316**	0.6551	**0.6813**	**0.7044**
	SSMLL-CAP	0.5479	0.5587	0.5666	0.5815	0.6125	0.6377	0.6034	0.6324	0.6577	0.5264	0.5876	0.6092
	PCLP	0.6059	0.6370	0.6363	0.6392	0.6469	0.6490	0.6707	0.6784	0.6780	0.5471	0.5588	0.5590
	fixmatch-LDL	0.6638	0.6797	0.6916	0.6857	0.7042	0.7119	0.7009	0.7147	0.7283	**0.6570**	0.6760	0.6987
	Mixmatch-LDL	0.6372	0.6418	0.6496	0.6639	0.6686	0.6831	0.6819	0.6806	0.6986	0.6133	0.6381	0.6534
	GCT-LDL	0.6116	0.6602	0.6770	0.6639	0.6879	0.6863	0.6787	0.7018	0.7102	0.6321	0.6669	0.6910
	SALDL	0.6457	0.6612	0.6723	0.5559	0.5108	0.5091	0.6632	0.5724	0.5687	0.6298	0.6504	0.6708
	sLDLF	0.5935	0.5861	0.6162	0.4813	0.4750	0.4616	0.6487	0.5652	0.5336	0.2433	0.2315	0.2199
	DF-LDL	0.5057	0.5461	0.6353	0.4173	0.4176	0.4169	0.3541	0.3536	0.3505	0.7022	0.7083	0.7085
	LDL-LRR	0.3721	0.6213	0.6626	0.5322	0.5519	0.5600	0.5746	0.5904	0.5979	0.5649	0.5389	0.4411
	Adam-LDL-SCL	0.3409	0.5627	0.6040	0.4724	0.3933	0.4628	0.5488	0.5828	0.5200	0.6177	0.5768	0.4843

Experimental results

		Emotion6			Flickr-LDL			Twitter-LDL			RAF-LDL		
Cos. ↑	Rankmatch	**0.8121**	**0.8257**	**0.8331**	**0.8489**	**0.8614**	**0.8679**	**0.8544**	**0.8698**	**0.8790**	**0.7901**	**0.8140**	**0.8375**
	SSMLL-CAP	0.6850	0.6994	0.7185	0.7634	0.7885	0.8144	0.8109	0.8270	0.8442	0.6456	0.7119	0.7329
	PCLP	0.7421	0.7737	0.7778	0.8057	0.8146	0.8151	0.8391	0.8436	0.8448	0.6815	0.6962	0.6969
	Fixmatch-LDL	0.8079	0.8200	0.8312	0.8487	0.8573	0.8673	0.8517	0.8647	0.8758	0.7881	0.8123	0.8311
	Mixmatch-LDL	0.7585	0.7863	0.7901	0.7888	0.8381	0.8468	0.8463	0.8552	0.8602	0.7536	0.7680	0.7820
	GCT-LDL	0.7530	0.8017	0.8134	0.8313	0.8508	0.8531	0.8499	0.8587	0.8716	0.7660	0.7977	0.8181
	SALDL	0.7784	0.7874	0.7981	0.7361	0.6643	0.6624	0.8479	0.7612	0.7615	0.7711	0.7938	0.8135
	sLDLF	0.7037	0.6980	0.7350	0.6276	0.6066	0.5897	0.8002	0.7454	0.6988	0.3262	0.3506	0.3459
	DF-LDL	0.6035	0.6470	0.7689	0.5436	0.5539	0.5569	0.5069	0.5233	0.5209	0.8427	0.8492	0.8470
	LDL-LRR	0.4604	0.7362	0.7905	0.7020	0.7316	0.7399	0.7767	0.8027	0.8125	0.7253	0.6938	0.5757
	Adam-LDL-SCL	0.4311	0.6670	0.7144	0.6104	0.4888	0.6166	0.7163	0.7661	0.7403	0.7717	0.7337	0.6191

Experimental results

Metric	Dataset	10% Labeled Data						20% Labeled Data					
		RankMatch	GCT	CAP	PCLP	FixMatch	MixMatch	RankMatch	GCT	CAP	PCLP	FixMatch	MixMatch
τ_K	RAF	**0.5696**	0.4258	0.2079	0.2750	0.4643	0.4066	**0.5463**	0.4811	0.3598	0.2949	0.4946	0.4524
	Emotion6	**0.5535**	0.3718	0.1237	0.3325	0.4899	0.4562	**0.5620**	0.4594	0.1536	0.4394	0.4985	0.4528
	Flickr	**0.5618**	0.5005	0.4030	0.4904	0.5215	0.5119	**0.5627**	0.5215	0.4475	0.5005	0.5416	0.5265
	Twitter	0.4927	0.4806	0.4407	0.4887	0.4744	**0.5016**	**0.5452**	0.5121	0.4828	0.5037	0.5039	0.5177
ρ_S	RAF	**0.6726**	0.5122	0.2474	0.3365	0.5564	0.4900	**0.6649**	0.5754	0.4297	0.3602	0.5917	0.5497
	Emotion6	**0.6545**	0.4495	0.1529	0.4037	0.5933	0.5528	**0.6512**	0.5624	0.1923	0.5332	0.5989	0.5559
	Flickr	**0.6537**	0.5904	0.4831	0.5793	0.6097	0.6019	**0.6551**	0.6112	0.5335	0.5903	0.6304	0.6171
	Twitter	0.5740	0.5637	0.5193	0.5735	0.5517	**0.5864**	**0.6315**	0.5966	0.5658	0.5898	0.5853	0.6024

Experimental results

		Che.↓	Cla.↓	Can.↓	KL↓	Cos.↑	Int.↑
Flickr	pretrain	0.2411	2.2594	5.6885	0.5371	0.8427	0.6873
	pretrain + consistency	0.2262(6.2%↑)	2.1131(6.5%↑)	5.1536(9.4%↑)	0.5293(1.5%↑)	0.8633(2.4%↑)	0.7188(4.6%↑)
	pretrain + consistency+PRR loss	**0.2184(3.4%↑)**	**2.0158(4.6%↑)**	**4.9008(4.9%↑)**	**0.5227(1.2%↑)**	**0.8714(0.9%↑)**	**0.7208(0.3%↑)**
		Che.↓	Cla.↓	Can.↓	KL↓	Cos.↑	Int.↑
RAF	pretrain	0.2938	1.5412	3.206	0.5146	0.7687	0.6411
	pretrain + consistency	0.255(13.2%↑)	1.5021(2.5%↑)	3.1345(2.2%↑)	0.3699(28.1%↑)	0.8189(28.1%↑)	0.7073(10.3%↑)
	pretrain + consistency+PRR loss	**0.2341(8.2%↑)**	**1.4914(0.7%↑)**	**3.0459(2.8%↑)**	**0.3464(6.4%↑)**	**0.8476(3.5%↑)**	**0.7194(1.7%↑)**

Experimental results

	anger	digust	fear	joy	sad	surprise	netural
Ground Truth	0.07	0.17	0.06	0.03	**0.36**	0.03	0.28
Rankmatch-LDL	0.02	0.04	0.05	0.08	**0.50**	0.08	0.23
Fixmatch-LDL	0.03	0.08	0.13	0.17	0.16	0.16	**0.27**
Mixmatch-LDL	0.04	0.06	**0.28**	0.12	0.23	0.09	0.18
GCT-LDL	0.04	0.06	0.12	0.24	0.15	0.10	**0.29**

	surprise	fear	disgust	happy	sad	anger
Ground Truth	0.45	**0.52**	0.00	0.00	0.00	0.03
Rankmatch-LDL	0.37	**0.52**	0.02	0.00	0.08	0.01
Fixmatch-LDL	**0.45**	0.32	0.04	0.03	0.14	0.02
Mixmatch-LDL	**0.39**	0.10	0.19	0.09	0.13	0.10
GCT-LDL	0.30	0.19	0.05	0.03	**0.42**	0.01

	anger	digust	fear	joy	sad	surprise	netural
Ground Truth	0.00	0.00	0.00	**0.63**	0.00	0.20	0.17
Rankmatch-LDL	0.01	0.04	0.11	**0.34**	0.07	0.20	0.23
Fixmatch-LDL	0.02	0.12	0.06	0.20	0.06	0.09	**0.45**
Mixmatch-LDL	0.04	0.17	0.07	0.18	0.10	0.09	**0.35**
GCT-LDL	0.04	0.14	0.10	0.23	0.13	0.11	**0.25**

	surprise	fear	disgust	happy	sad	anger
Ground Truth	0.27	0.00	0.03	**0.67**	0.03	0.00
Rankmatch-LDL	0.35	0.02	0.02	**0.59**	0.00	0.02
Fixmatch-LDL	**0.64**	0.10	0.02	0.22	0.01	0.01
Mixmatch-LDL	**0.55**	0.13	0.07	0.06	0.15	0.03
GCT-LDL	**0.44**	0.17	0.08	0.12	0.15	0.04