Neighbor Matching for Semi-supervised Learning

Renzhen Wang , Yichen Wu , Huai Chen , Lisheng Wang , and Deyu Meng

Macau University of Science and Technology
Shanghai Jiao Tong University
Xi’an Jiaotong University

MICCAI2021

Introduction

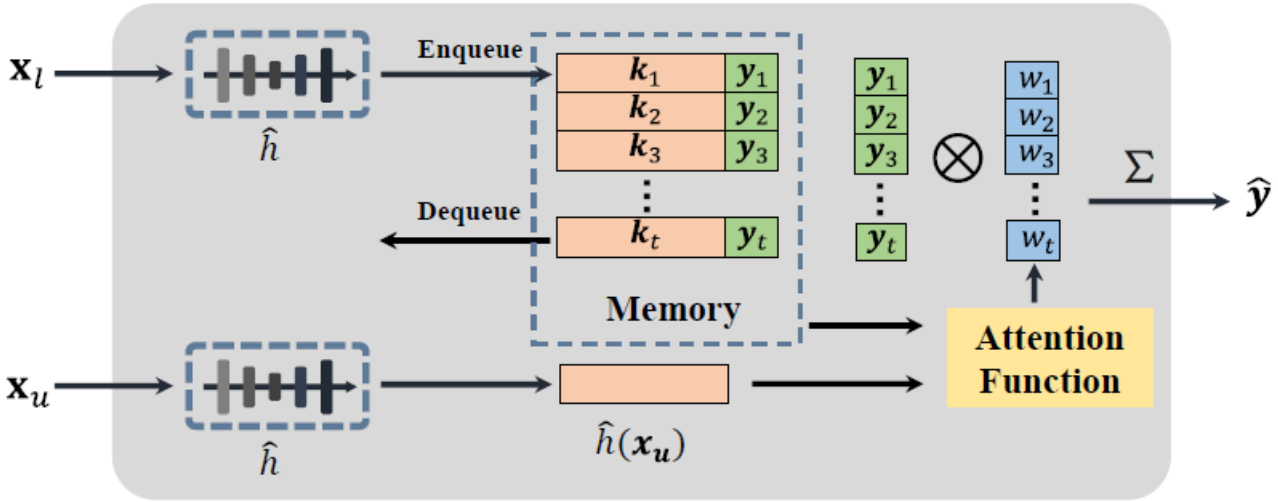
- Semi-supervised learning:

we are given a labeled training set D_l and an unlabeled set D_u

- Pseudo labeling:

Assigning a pseudo label to unlabeled data D_u for training

Method



- Neighbor Matching

$$P(\hat{\mathbf{y}}|\hat{h}(\mathbf{x}_u), \mathcal{M}) = \sum_{i=1}^{t} w(\hat{h}(\mathbf{x}_u), \mathbf{k}_i)\mathbf{y}_i,$$

- Attention mechanism

$$w(\hat{h}(\mathbf{x}_u), \mathbf{k}_i) = \frac{\exp(d(\hat{h}(\mathbf{x}_u), \mathbf{k}_i)/T)}{\sum_{j=1}^{t} \exp(d(\hat{h}(\mathbf{x}_u), \mathbf{k}_j)/T)}$$

Method

Memory Padding

- Batch:

Misclassify if the class unlabeled sample belong to was not memorized in *M*

- All sample:

extremely time consuming

dynamic updating and **unbiased updating**

- the memory bank M is updated through a queue
- the number of each class in *M* should be always kept equal, and only correctly classified samples whose embedding-label pairs can enqueue into *M*

$$(\hat{h}(\mathbf{x}), \mathbf{y}) \rightarrow \mathcal{M} \iff \delta[f(\mathbf{x})] = c$$

Method

Algorithm

Algorithm 1. Mini-batch training of SSL with Neighbor Matching

Input: Labeled training data $\mathcal{D}_l$, Unlabeled training data $\mathcal{D}_u$, batch size s, the size t of memory bank $\mathcal{M}$, max iteration T

Output: Parameters θ of the recognition network f

Initialize the memory bank $\mathcal{M}$.
for $t = 1$ **to** T **do**
  $\hat{\mathcal{D}}_l = \{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^{s} \leftarrow \text{SampleMiniBatch}(\mathcal{D}_l, s)$.
  $\hat{\mathcal{D}}_u = \{\mathbf{x}_j\}_{j=1}^{s} \leftarrow \text{SampleMiniBatch}(\mathcal{D}_u, s)$.
  Forward to compute the supervised loss: $\mathcal{L}_l = \frac{1}{|\hat{\mathcal{D}}_l|} \sum_{\mathbf{x}_i, \mathbf{y}_i \in \hat{\mathcal{D}}_l} \text{H}(\mathbf{y}_i, f(\mathbf{x}_i))$.
  Update the memory bank $\mathcal{M}$ according to Eq. (3).
  Estimate pseudo-labels for $\hat{\mathcal{D}}_u$ by Eq. (1).
  Forward to compute the unsupervised loss: $\mathcal{L}_u = \frac{1}{|\hat{\mathcal{D}}_u|} \sum_{\mathbf{x}_j \in \hat{\mathcal{D}}_u} \text{H}(\hat{\mathbf{y}}_j, f(\mathbf{x}_j))$.
  Backward according to Eq. (4) to update the parameters θ.
end for

Result

Dataset:

- **Chexpert:**

A large chest X-ray dataset for thoracic disease recognition, in which 224,316 chest radiographs were collected from 65,240 patients, and 14 pathology categories

- **ISIC 2018 skin dataset:**

7 disease categories, in which 10015 dermoscopic images and associated labels are contained

Network:

AlexNet

Result

Table 1. Ablation analysis on ISIC 2018 skin dataset [5] with 800 labeled training samples. Note that t is the memory bank size.

Methods	AUC	MCA
Baseline	0.7835	0.4168
EntMin [7]	0.7916	0.4127
Ours (w/o Unbiased Updating, w/t = 126)	0.8147	0.4184
Ours (w/Unbiased Updating, w/t = 7)	0.8184	0.4524
Ours (w/Unbiased Updating, w/t = 126)	0.8271	0.4846
Ours (w/Unbiased Updating, w/t = 252)	**0.8303**	**0.4933**
Ours (w/Unbiased Updating, w/t = 504)	0.8178	0.4701

Result

Table 2. The performance on ISIC 2018 skin dataset [10].

Method	350		800		1200	
	AUC	MCA	AUC	MCA	AUC	MCA
Baseline	0.7692	0.3951	0.7835	0.4168	0.8340	0.4935
GLM [8]	0.7870	0.3856	0.8479	0.4515	0.8754	0.5034
Ours (w/o mixup)	0.7743	0.4145	0.8303	0.4933	0.8547	0.5539
Ours (w/mixup)	**0.8176**	**0.4729**	**0.8603**	**0.5186**	**0.8864**	**0.5887**

Result

Table 3. The performance (mean AUC) on Chexpert dataset [10].

Method	Size of labeled data set (k)				
	100	200	300	400	500
Baseline	0.5576	0.6166	0.6208	0.6343	0.6353
LSSE [9]	0.6200	0.6386	0.6484	0.6637	0.6697
GLM [8]	0.6512	0.6641	0.6739	0.6796	0.6847
Ours	**0.6515**	**0.6711**	**0.6830**	**0.6889**	**0.6934**

Unsupervised Representation Learning Meets Pseudo-Label Supervised Self-Distillation: A New Approach to Rare Disease Classification

Jinghan Sun, Dong Wei, Kai Ma, Liansheng Wang, and Yefeng Zheng
Xiamen University, Xiamen,
ChinaJarvis Lab, Shenzhen, China

MICCAI2021

Introduction

- few-shot learning (FSL) problem
- we model a specific task as $T = \{S,Q\}$ consisting of a support set $S = \{(x, y)\}$ and a query set $Q = \{(x, y)\}$, where x is an image and y is its label
- It's an N-way K-shot task includes N rare diseases, each with K instances in S, where K is small
- Only S is available for training and Q is solely for testing
- The target is optimal classification performance on Q given S and D_{base} we consider D_{base} to be unlabeled for a more generally applicable approach in practice
- There is a large base dataset D_{base} consisting of common diseases, S and Q randomly sampled from a dataset D_{rare} consisting of rare diseases

Method

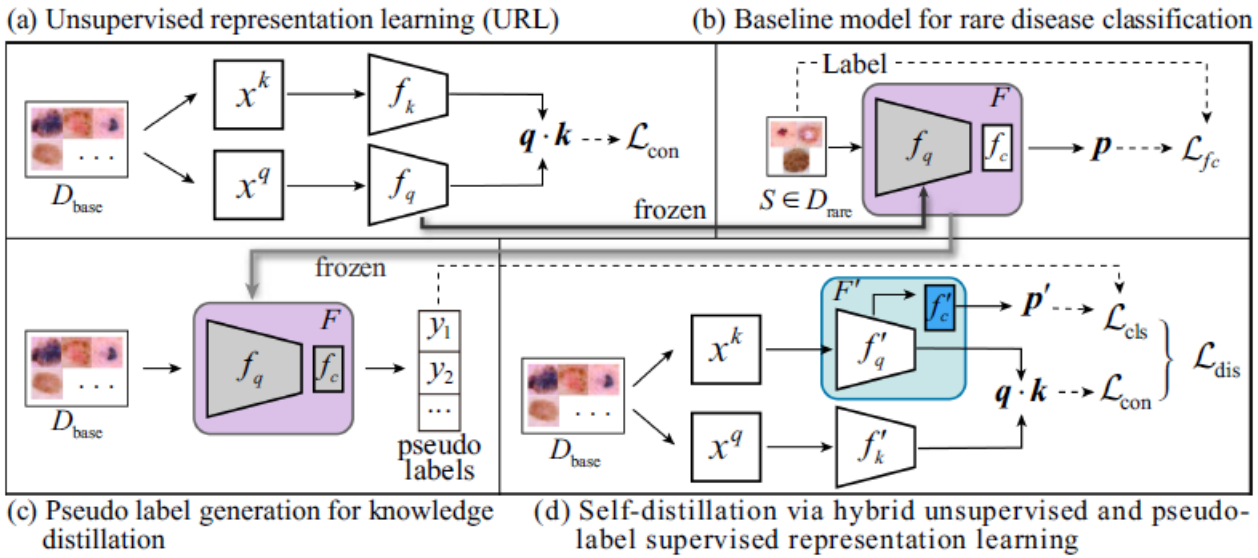


Fig. 1. Overview of the proposed approach. Solid line: information flow; dashed line: loss computation. Note that $\mathcal{L}_{f_c}$ in (b) can be any loss suitable for the classifier f_c.

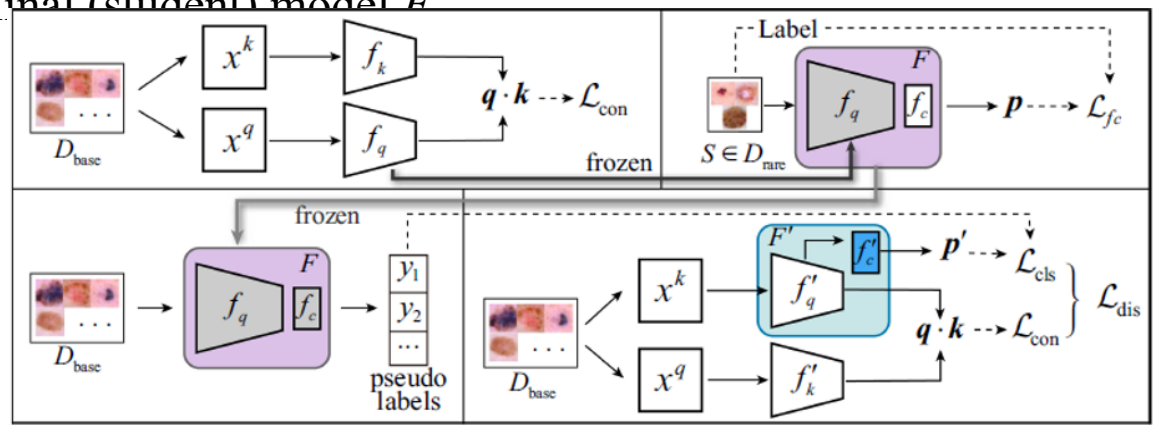
Method

Given D_{base}, we first perform unsupervised representation learning (URL) to train the embedding function f_q

Next, a simple classifier f_c is appended to the learned f_q(with frozen parameters) to compose a baseline model F wheref_c is optimized on S.

Then, F is employed to assign each CDNC **(**common diseases and normal controls**)** instance in D_{base} a pseudo label of the rare diseases

Lastly, a self-distillation via hybrid unsupervised and pseudo-label supervised representation learning is performed on D_{base} to produce the final (student) model F'



Method

URL on CDNC(common diseases and normal controls**) for Rare Disease Classification**

$$\mathcal{L}_{\text{con}}(x_i) = -\log\left[\exp\left(\boldsymbol{q}_i \cdot \boldsymbol{k}_i/\tau\right)/\left(\exp\left(\boldsymbol{q}_i \cdot \boldsymbol{k}_i/\tau\right) + \sum_{j=1}^{L} \exp\left(\boldsymbol{q}_i \cdot \boldsymbol{k}_j/\tau\right)\right)\right]$$

$$\boldsymbol{q}_i = f_q(x_i^q;\theta_q) \qquad \boldsymbol{k}_i = f_k(x_i^k;\theta_k)$$

Self-Distillation of Rare Disease Knowledge

$$\mathcal{L}_{\text{dis}} = \mathcal{L}_{\text{con}}(x;\theta_q',\theta_k') + \mathcal{L}_{\text{cls}}\left(\boldsymbol{y}, F'(x;\theta_q',\theta_c')\right)$$

Method

Adaptive Pseudo Labels.

Concretely, given the prediction $\boldsymbol{p}$ by the student model and pseudo labels $\boldsymbol{y}$ defined above, we combine them as our new training target:

$$y^{\text{adpt}} = (1-\alpha) \times \boldsymbol{y} + \alpha \times \boldsymbol{p},$$

where α is a confidence parameter controlling how much we trust the teacher's knowledge

Therefore, we adopt a linear growth rate for α at the t-th epoch: $\alpha_t = \alpha_T \times (t/T)$, where α_T is the last-epoch value and set to 0.7

Result

DATASET:
ISIC 2018 skin lesion classification dataset
melanocytic nevus (6,705), benign keratosis (1,099), melanoma (1,113), basal cell carcinoma (514), actinic keratosis (327), dermatofibroma (115), and vascular lesion (142).

Network:
ResNet-12

Dataset division：
we use the four classes with the most cases as the CDNC dataset D_{base}, and the other three as the rare disease dataset D_{rare}

Result

Method	$(N, K) = (3, 1)$		$(N, K) = (3, 3)$		$(N, K) = (3, 5)$	
	Accuracy (%)	F1 score (%)	Accuracy (%)	F1 score (%)	Accuracy (%)	F1 score (%)
Training from scratch	37.74 (1.07)	29.90 (3.65)	39.76 (0.88)	35.60 (1.72)	45.36 (3.76)	38.41 (4.06)
▷ SML MAML [5]	47.49 (5.38)	42.33 (6.16)	55.55 (3.12)	49.19 (4.20)	58.94 (2.59)	53.51 (2.46)
RelationNet [28]	46.10 (4.80)	39.98 (6.73)	47.29 (2.77)	43.37 (3.65)	55.71 (3.30)	49.34 (3.57)
ProtoNets [25]	35.18 (3.12)	30.81 (3.09)	38.59 (1.91)	33.11 (2.08)	42.45 (2.45)	34.92 (3.70)
DAML [18]	50.05 (5.18)	41.65 (3.98)	55.57 (3.55)	49.01 (6.62)	59.44 (3.17)	54.66 (2.43)
▷ UML UMTRA [14]	45.88 (3.63)	41.44 (4.37)	51.29 (3.54)	45.91 (3.96)	57.33 (1.76)	53.06 (0.89)
CACTUs-MAML [9]	42.98 (2.91)	35.38 (3.08)	44.44 (3.35)	39.94 (3.65)	48.11 (4.20)	44.32 (3.65)
CACTUs-ProtoNets [9]	42.67 (2.43)	39.24 (2.72)	45.00 (3.26)	39.69 (2.66)	47.95 (3.52)	44.08 (2.63)
▷ SRL SRL-simple [29]	54.45 (5.82)	51.02 (6.93)	61.31 (6.31)	57.65 (3.46)	70.53 (2.17)	65.58 (3.72)
SRL-distil [29]	55.43 (7.36)	51.18 (5.50)	64.92 (6.00)	59.88 (4.87)	72.78 (1.67)	65.89 (2.54)
▷ URL SimCLR [1]	52.43 (5.01)	44.70 (8.24)	63.82 (3.70)	57.55 (3.67)	70.18 (1.76)	63.73 (1.78)
MoCo_v2 [3]	59.95 (4.73)	55.98 (3.81)	70.84 (2.91)	64.77 (3.69)	75.80 (1.85)	70.69 (2.13)
MoCo_v1 (baseline) [7]	61.90 (2.92)	56.30 (1.48)	74.92 (2.96)	69.50 (5.72)	79.01 (2.00)	74.47 (3.03)
Hybrid distil (ours)	**64.15** (2.86)	**61.01** (1.30)	**75.82** (2.47)	**73.34** (2.30)	**81.16** (2.60)	**77.35** (4.21)

Result

$\mathcal{L}_{\text{dis}}$	$(N, K) = (3, 1)$		$(N, K) = (3, 3)$		$(N, K) = (3, 5)$	
	Accuracy (%)	F1 score (%)	Accuracy (%)	F1 score (%)	Accuracy (%)	F1 score (%)
N.A.	61.90 (2.92)	56.30 (1.48)	74.92 (2.96)	69.50 (5.72)	79.01 (2.00)	74.47 (3.03)
$\mathcal{L}_{\text{cls}}$	63.70 (3.39)	57.31 (7.73)	74.92 (2.10)	70.28 (3.97)	80.24 (1.61)	77.29 (2.91)
$\mathcal{L}_{\text{con}} + \mathcal{L}_{\text{cls}}$	**64.15** (2.86)	**61.01** (1.30)	**75.82** (2.47)	**73.34** (2.30)	**81.16** (2.60)	**77.35** (4.21)
$\mathcal{L}_{\text{con}} + \mathcal{L}_{\text{reg}}$	62.20 (5.18)	56.19 (4.28)	74.43 (2.88)	69.74 (4.13)	79.14 (2.09)	74.41 (2.65)