Learning to Detect Important People in Unlabelled Images for Semi-supervised Important People Detection

Fa–Ting Hong[1] , Wei–Hong Li[2] and Wei–Shi Zheng[1]

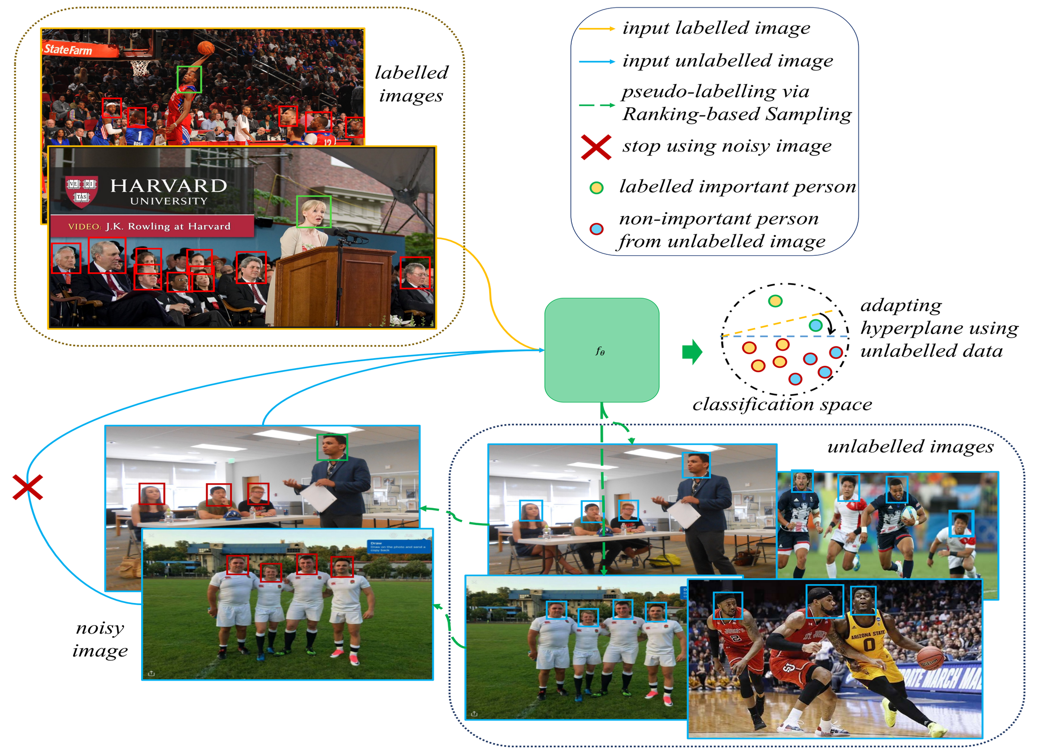
[1]Sun Yat-sen University [2]University of Edinburgh

CVPR SEATTLE WASHINGTON JUNE 16-18 2020

Introduction

★ Introduction & Motivation

* Existing methods of important people detection require massive quantities of labelled data and detecting important people in unlabeled images has not yet been developed.
* The imbalance between the number of important people and non-important people in the picture will cause pseudo-labelling imbalance problem.
* Not all unlabelled images contain important people; images without such people represent noisy unlabelled samples during learning.



★ Contributions

* The proposed approach is the first to study on learning important people detection from partially labelled data.
* we contribute two large datasets called Extended-MS (EMS) and Extended- NCAA (ENCAA) for evaluation of semi-supervised important people detection by augmenting existing datasets with a large number of unlabelled images collected from the internet
* Extensive experiments verify the efficacy of our proposed method on important people detection of semi-supervised phase.

★ Detecting Noisy Unlabelled Images

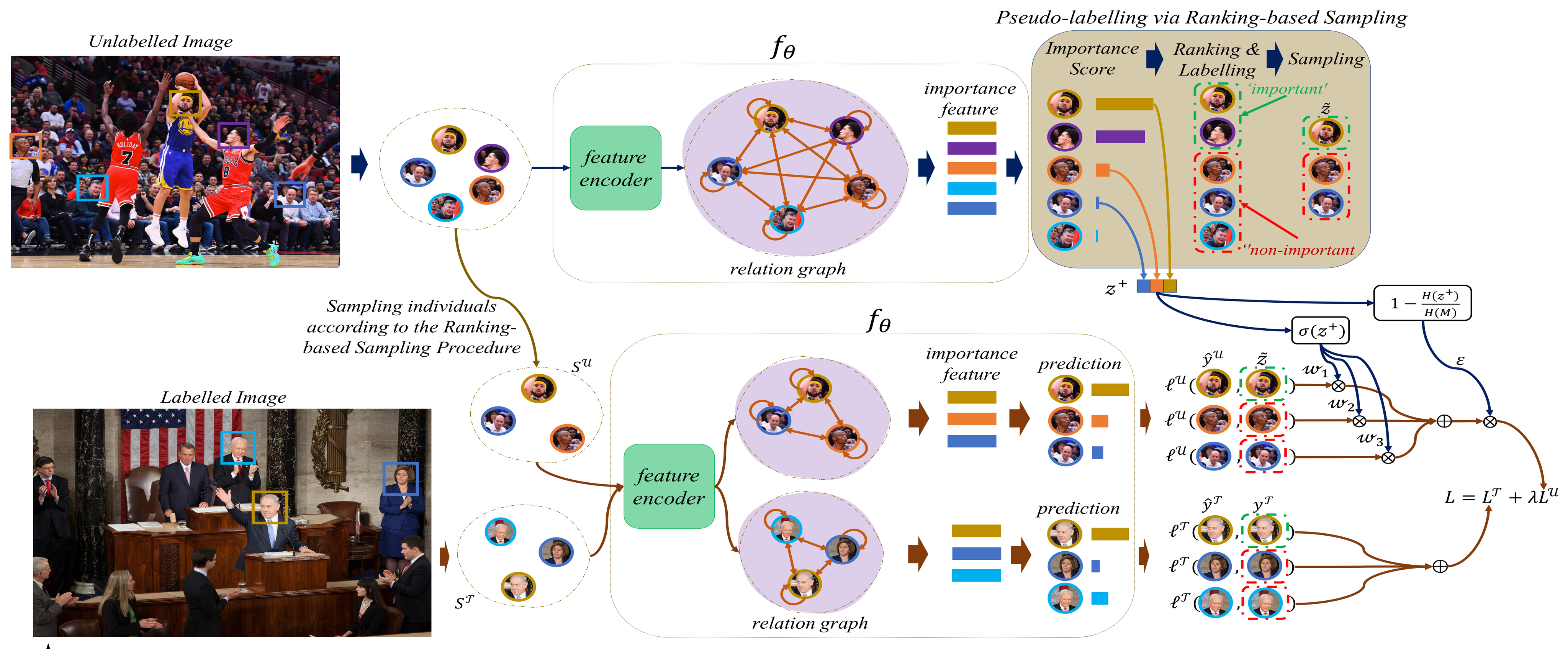
* Image-specific effectiveness weight : ε

$$\varepsilon = 1 - \frac{H(z^+)}{H(M)}$$

* Effectiveness weight reflect the confidence that an unlabelled image features important people：

$$L^{\mathcal{U}} = \frac{\lambda}{|\mathcal{U}|}\sum_{i=1}^{|\mathcal{U}|}\varepsilon_i \sum_{x_j^{\mathcal{U}} \in \mathcal{S}_i^{\mathcal{U}}} w_j \ell^{\mathcal{U}}(\hat{y}_j^{\mathcal{U}}, \tilde{z}_j)$$

Methodology



★ Overview of Proposed Method

* To alleviate the pseudo-labelling imbalance problem, we introduce a ranking strategy for pseudo-label estimation, and also introduce two weighting strategies applied to unlabelled data loss.
* The final objective function can be expressed as:

$$L = L^{\mathcal{T}} + \lambda L^{\mathcal{U}}$$

$$= \frac{1}{|\mathcal{T}|K}\sum_{i=1}^{|\mathcal{T}|}\sum_{x_j^{\mathcal{T}} \in \mathcal{S}_i^{\mathcal{T}}} \ell(\hat{y}_j^{\mathcal{T}}, y_j^{\mathcal{T}}) + \frac{\lambda}{|\mathcal{U}|}\sum_{i=1}^{|\mathcal{U}|}\varepsilon_i \sum_{x_j^{\mathcal{U}} \in \mathcal{S}_i^{\mathcal{U}}} w_j \ell^{\mathcal{U}}(\hat{y}_j^{\mathcal{U}}, \tilde{z}_j)$$

$$s.t. \quad w_j \in \boldsymbol{w} = \sigma(z^+), \varepsilon_i = 1 - \frac{H(z^+)}{H(M)}$$

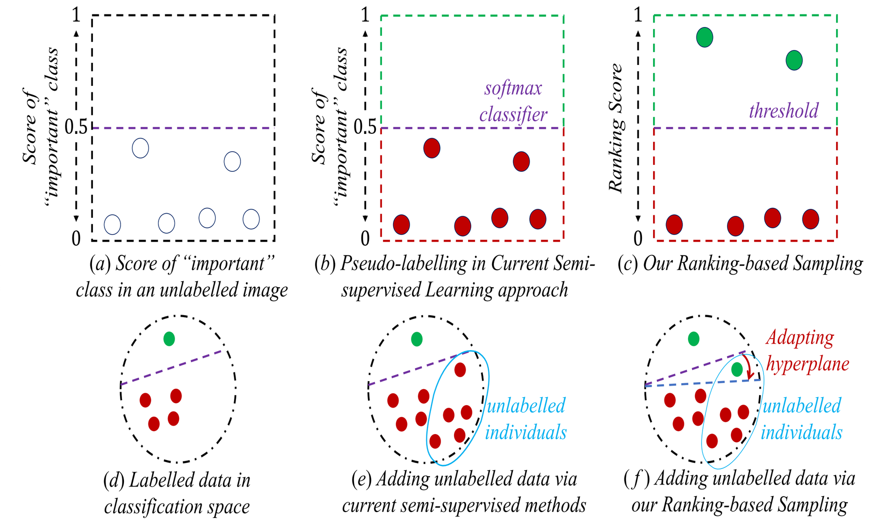
★ Pseudo-labelling by Ranking-based Sampling:

* Ranking-based sampling procedure:

$$\mathcal{S}_i^{\mathcal{U}}, \tilde{z} = RankS(f_\theta, \{x_j^{\mathcal{U}}\}_{x_j^{\mathcal{U}} \in I_i^{\mathcal{U}}}, \alpha, K)$$

* Replacing the unlabelled people and corresponding pseudo-labels with those sampled by RankS:

$$L^{\mathcal{U}} = \frac{\lambda}{|\mathcal{U}|}\sum_{i=1}^{|\mathcal{U}|}\sum_{x_j^{\mathcal{U}} \in \mathcal{S}_i^{\mathcal{U}}} \ell^{\mathcal{U}}(\hat{y}_j^{\mathcal{U}}, \tilde{z}_j)$$



(a) Score of "important" class in an unlabelled image (b) Pseudo-labelling in Current Semi-supervised Learning approach (c) Our Ranking-based Sampling (d) Labelled data in classification space (e) Adding unlabelled data via current semi-supervised methods (f) Adding unlabelled data via our Ranking-based Sampling

★ Balancing Loss via Importance Score Weighting:

* Person-specific importance score weight: w_j

$$L^{\mathcal{U}} = \frac{\lambda}{|\mathcal{U}|}\sum_{i=1}^{|\mathcal{U}|}\sum_{x_j^{\mathcal{U}} \in \mathcal{S}_i^{\mathcal{U}}} w_j \ell^{\mathcal{U}}(\hat{y}_j^{\mathcal{U}}, \tilde{z}_j), \text{s.t.} \sum_{j=1}^{K} w_j = 1, w_j > 0.$$

Experimental Results

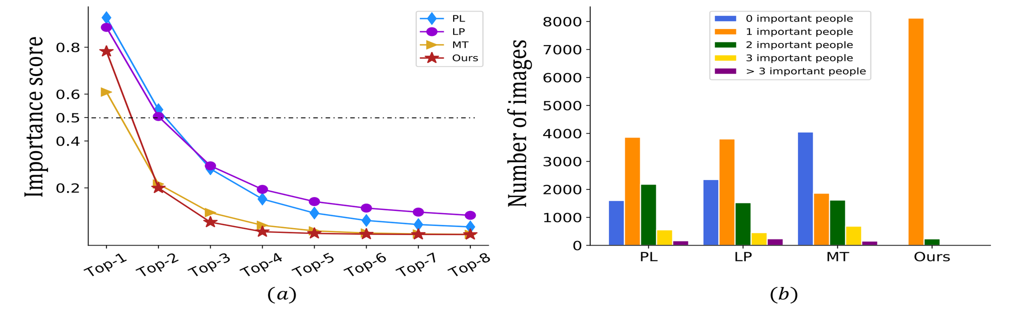
★ Datasets

* EMS Dataset: 10, 687 images from multiple scenes.
* ENCAA Dataset: 28, 798 frames of basketball game video.

★ Comparisons with related methods on both datasets

Dataset	EMS			ENCAA		
#labelled images	33 %	66 %	100 %	33 %	66 %	100 %
POINT (fully supervised)	83.36	85.97	88.48	84.60	88.21	89.75
Pseudo Label (PL)	83.37	85.35	88.57	85.70	88.43	90.56
Label Propagation (LP)	82.34	86.33	86.66	85.36	88.61	90.18
Mean Teacher (MT)	84.50	86.29	87.55	83.33	84.66	87.55
Ours	**87.81**	**88.44**	**89.79**	**88.75**	**90.86**	**92.03**

Table 1. Comparison with related methods on both datasets.



Fgiure (a) is the distribution of top 8 importance score in testing set in EMS datasets and Figure (b) is the statistics of unlabelled data's pseudo-labels on EMS dataset. Better view in color.

★ Evaluations of different components and techniques used to estimate importance score in proposed method

Dataset	EMS			ENCAA		
#labelled images	33 %	66 %	100 %	33 %	66 %	100 %
$\text{Ours}_{\text{w/o Ranks, ISW and EW}}$	83.70	86.81	87.67	84.35	87.66	89.93
$\text{Ours}_{\text{w/o ISW and EW}}$	85.55	87.25	88.53	87.13	90.53	91.49
$\text{Ours}_{\text{w/o EW}}$	86.34	87.45	89.67	87.68	90.60	92.00
Ours	87.81	88.44	89.79	88.75	90.86	92.03

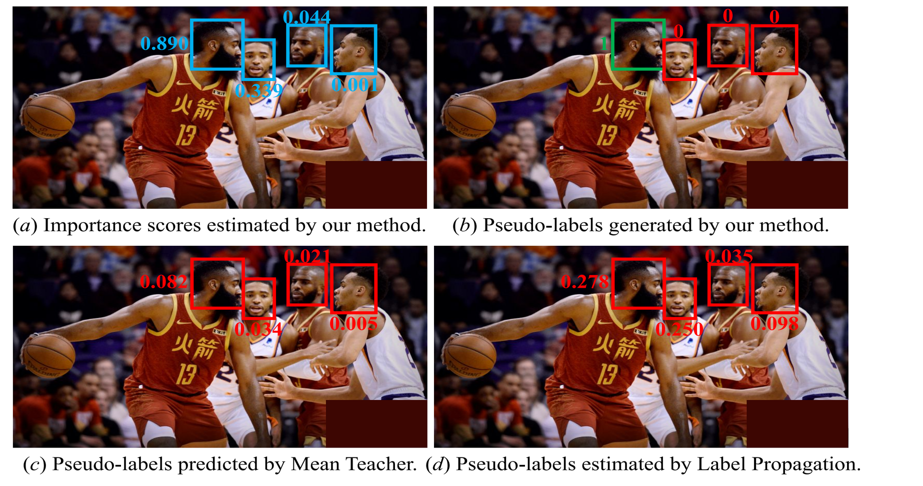
Table 2. Ablation study on both datasets. RankS represents ranking-based sampling while ISW and EW indicate importance score weighting and effectiveness weighting, respectively. $\text{Ours}_{\text{w/o ISW and EW}}$ means our model without using ISW and EW.

Dataset	EMS			ENCAA		
#labelled images	33 %	66 %	100 %	33 %	66 %	100 %
Ours_{LP}	87.51	88.10	89.65	88.95	91.06	91.98
Ours_{MT}	87.23	88.56	90.72	88.97	90.93	91.62
Ours	87.81	88.44	89.79	88.75	90.86	92.03

Table 3. Evaluation of different techniques (*i.e.*, LP and MT) when used for instantiating pseudo-label estimation function (*i.e.*, $g(\cdot)$) instead of using Softmax function.

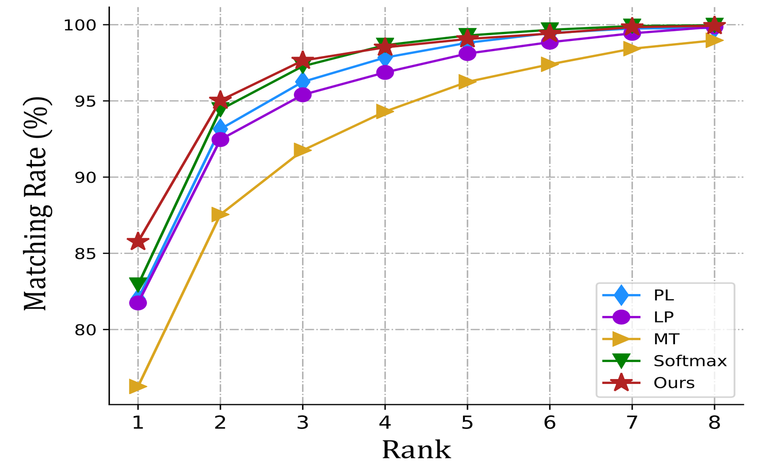
★ Visualize results

Examples of unlabelled images and their effectiveness weights estimated automatically by our method.

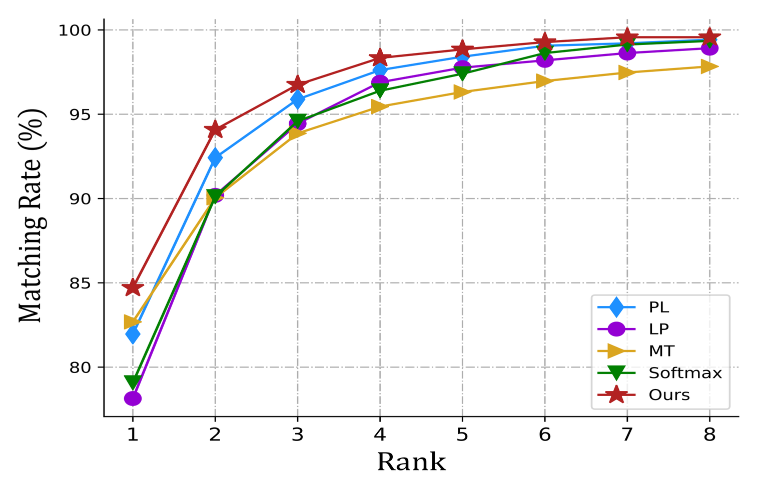


(a) Importance scores estimated by our method. (b) Pseudo-labels generated by our method. (c) Pseudo-labels predicted by Mean Teacher. (d) Pseudo-labels estimated by Label Propagation.

Examples of our method's results and pseudo-labels estimated by different methods during training



(a) CMC curve on EMS dataset



(b) CMC curve on ENCAA dataset