2023 IEEE Symposium Series on Computational Intelligence (SSCI)
Mexico City, Mexico. December 5-8, 2023

Return of Small-Scale Crowd Counting via Fast and Accurate Semi-Supervised Least Squares Model

Hao Luo
School of Software Engineering
Xi'an Jiaotong University
Xi'an, China
luohao4122@stu.xjtu.edu.cn

Shaoyi Du
Institute of Artificial Intelligence and Robotics
National Key Laboratory of Human-Machine Hybrid Augmented Intelligence
National Engineering Research Center for Visual Information and Applications
Xi'an Jiaotong University
Xi'an, China
dushaoyi@xjtu.edu.cn

Zhiqiang Tian*
School of Software Engineering
Xi'an Jiaotong University
Xi'an, China
zhiqiangtian@xjtu.edu.cn

***Abstract*—Existing crowd counting techniques have achieved significant progress with the emergence of deep learning. During development, emerging crowd counting methods have generally become more and more complex and enormous, enabling them to understand and process more prior knowledge from input data. However, they suffer from two major drawbacks: 1) they generally require a significant amount of labeled training samples, which is labor-intensive, and 2) they require increasing computational hardware resources, making it luxurious and impractical to apply directly in small-scale scenes. To address these issues, we formulate crowd counting as a classification problem and leverage least squares model with a novel semi-supervised strategy. Technically, we construct the least squares model based on only two regularization terms: a regression term and a discriminative relaxation term. Moreover, we propose a semi-supervised soft label correcting strategy incorporated in the model. As a result, a fast and accurate crowd counting method is achieved. Experimental results on five small-scale benchmarks demonstrate the proposed method outperforms the other competitors in terms of both regression metrics and consumed time.**

***Index Terms*—crowd counting, least squares model, classification, semi-supervised**

I. Introduction

Crowd counting aims to estimate the number of pedestrians in a static image or a video frame. Towards growing demands of social security, it has been a crucial video surveillance technique [1]. Nevertheless, the development of crowd counting techniques benefits to bring meaningful insights to other counting tasks since they share similar task structure, e.g., step counting [2], finger counting [3], and tree counting [4].

Zhiqiang Tian* is corresponding author. This work was supported by NSFC under Grant Nos. 62173269 and 61971343, the Natural Science Basic Research Plan in Shaanxi Province of China under Grant No. 2022JM-324, the Social Science Foundation of Shaanxi Province of China under Grant No. 2021K014, and the Huawei Intelligent Base Project under Grant No. 22ZNJZ10, Shaanxi Joint Key Laboratory for Artifact Intelligence, China.

Existing crowd counting methods essentially learn a mapping model from a set of labeled images to estimate the crowd count [5], [6] or density map [7], [8]. Despite the continuous improvement in the accuracy of crowd counting, most existing crowd counting models heavily rely on a large number of labeled samples during the learning phase, especially those based on deep learning techniques [7], [8]. In practice, the task of annotating a large number of crowd videos is expensive and time-consuming, which hinders the deployment of supervised learning-based methods in real-world scenarios. Therefore, it will safe a lot of annotating labors to derive a semi-supervised crowd counting method.

Until now, crowd counting has achieved significant development particularly in mid-to-large-scale scenarios due to the emergence of deep learning models. These models require abundant computational resources for training in order to achieve more powerful capabilities. However, such models are too elaborate to directly apply in small-scale scenarios. Additionally, small-scale scenes suffer from a lack of training samples, which may result in serious performance degradation of deep learning methods. Therefore, deriving a small-scale scenario-oriented crowd counting method would be beneficial, as it could save a significant amount of computational resources.

Recently, some researchers have found that formulating crowd counting as a classification problem benefits to achieve very accurate results than those based on conventional philosophy towards small-scale scenarios [5], [6], [9], [10]. Specifically, sparse representation and random projection (SRRP) [9] is the pioneer work being the first to formulating crowd counting as a classification problem and obtain very accurate counting results than those of conventional types of methods. Inspired by SRRP, Zhang *et al*. [5] proposed that continual frames of small-scale crowd seances lie in a low-dimensional manifold, and exploring such information helps to improve

the counting performance. Towards the low-efficient l_0 and l_1 constraints used in SRRP, Liu *et al.* [10] proposed to replace such time-consuming constraints by a linear transformation, and proposed a linear dictionary learning model, achieving faster classification-based crowd counting than SRRP. Wang *et al.* adopted a class-specific structured to formulate a linear dictionary model to capture salient features in a pedestrian frames, resulting in higher results.

The above-discussed methods all formulate crowd counting as a classification problem and obtain more accurate counting results than those of conventional methods. However, they either include complex regularization terms or stacking class-specific structures, which are time-consuming and there is still room for improvement in speed. In addition, excessive pursuit of formulating a complex model may lead to over-fitting. To address the above issues, we propose to formulate a very simple yet effective classification-based least squares model with a proposed semi-supervised soft label correcting strategy. As a result, the proposed method is qualified to achieve more accurate and faster results than conventional classification-based methods. The major contributions of this paper are summarized as follows.

1) We propose a very simple yet effective least squares model and propose to formulate crowd counting as a classification problem.
2) We propose a semi-supervised strategy to progressively correct soft labels during iterations towards imbalanced crowd samples.

The remainder of this paper is arranged as follows. The overall structure and the idea of the proposed method is detailed in section II. The experimental results and relevant analysis are presented in III. Conclusions, remaining problems, and future works are described in section IV.

II. Proposed Method

A. Basic Least Squares Terms

Denote $X \in \mathbb{R}^{d\times n}$ as training data matrix, where d is the size of dimension and n is the number of training samples. Dentoe $H = [h_1, h_2, \cdots, h_n] \in \mathbb{R}^{C\times n}$ as the one-hot binary label matrix with its column $h = [0, \cdots, 1, \cdots, 0]^T \in \mathbb{R}^C$, where C is the number of classes. The basic least squares model is formulated as Eq. (1).

$$\min_{Q} \|QX - H\|_F^2 + \lambda \|Q\|_F^2, \tag{1}$$

where $Q \in \mathbb{R}^{C\times d}$ is the objective weight matrix to be learned, the first term enables the training samples to be projected onto the class space, and the second term ensures the computing is stable and avoids over-fitting. Typically, the weight Q can be computed through a closed-form least square method as Eq. (2).

$$Q = HX^T \left(XX^T + \lambda I\right)^{-1}. \tag{2}$$

With the weight Q learned, the class of a new test sample y in the same distribution with X can be predicted by Eq. (3).

$$\text{class}\,(y) = \arg\max_k \left([Qy]_k\right), \tag{3}$$

where $[Qy]_k$ denotes the k-th element of the vector Qy, and the computed index k denotes the predicted class of y.

However, the above model is not qualified to tackle crowd counting problem regarding the following two reasons.

1) Crowd counting especially in small-scale scenarios suffers from an extremely imbalanced number of samples of different pedestrians, which is unfavorable to a classification model.
2) The one-hot label matrix H is too hard to be a regression target, since a pedestrian frame with the same number of people might have large spacial variations due to the space shift of a pedestrian.

Towards the above issues, a least squares model with a new semi-supervised strategy is proposed in the next subsection.

B. Least Squares Model with Semi-Supervised Soft Label Correcting

To alleviate the negative influence of the hard regression constant imposed by H, we propose to introduce a relaxation term to learn a more flexible regression target, and incorporate this term into the objective function as Eq. (4).

$$\min_{Q,T,W} \|QX - T\|_F^2 + \eta \|WT - H\|_F^2 + \lambda \left(\|Q\|_F^2 + \|W\|_F^2\right), \tag{4}$$

where $T \in \mathbb{R}^{C\times n}$ and $W \in \mathbb{R}^{C\times C}$ are two newly-introduced matrices to be learned and η is a user-defined hyper-parameter. Specifically, T is the relaxed regression target and W is the weight matrix to project T onto the class space. Since the regression target is relaxed, it benefits to tackle the negative influence from the variations of the pedestrian frames, which results in a complete difference with the basic least squares model as Eq. (1). Next, the class of a given test sample y can be predicted by Eq. (5).

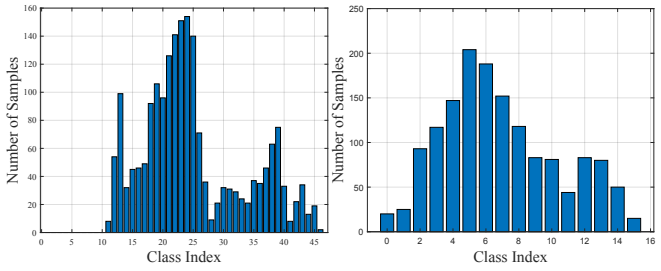
$$\text{class}\,(y) = NN\left(QWX, QWy\right), \tag{5}$$

where $NN(\cdot)$ denotes the nearest-neighbor classifier.

Since a crowd counting samples are extremely imbalanced in terms of the number of people as shown in Fig. 1, it will lead to a performance degradation if we train our model using the conventional supervised manner. Based on this considerations, we propose a novel semi-supervised strategy to correct the soft label progressively. Specifically, suppose the training process contains $T+1$ stages, where the first stage corresponds to Eq. (4) that trains with the labeled data matrix X, and the remained T stages jointly train with $\mathcal{X} = [X, X_u]$, where $X_u \in \mathbb{R}^{d\times m}$ is the unlabeled data matrix with m samples. In the first stage, we start training with Eq. (4), then the soft label vector of a given test sample y can be computed by Eq. (6).

$$Soft\,(y) = QWy. \tag{6}$$

Please note that, here we define the soft label vector QWy with its each element being a probability to each class varying from 0 to 1, instead of a one-hot label vector. This benefits to generate a stable augmented results, and can be corrected progressively with iterations going on. Then, given the unlabeled



(a) UCSD (b) Fudan

Fig. 1. Distribution of each class (number of pedestrians). The number of samples of each class is extremely imblanced, which might lead to negtive influence to a classification model. It will be beneficial if we adopt all samples for training with semi-supervised learning.

test data matrix $X_u \in \mathbb{R}^{d\times m}$, its pseudo label matrix Y_u in the t-th stage can be computed using Eq. (7).

$$Y_u^{(t)} = NN\left(Q^{(t)}W^{(t)}X, Q^{(t)}W^{(t)}\frac{1}{t}\sum_{t=1}^{T}X_u\right). \quad (7)$$

Next, we update the one-hot label matrix of the unlabeled test data $H_u^{(t)}$ using the last iteration pseudo label $Y_u^{(t)}$, and we augment it with the one-hot label matrix of the labeled samples H being $\mathcal{H}^{(t)} = \left[H, H_u^{(t)}\right]$. Then, the t-stage optimization variables can be computed with Eq. (8).

$$\min_{Q^{(t)},T^{(t)},W^{(t)}} \left\|Q^{(t)}\mathcal{X}^{(t)} - T^{(t)}\right\|_F^2 + \eta\left\|W^{(t)}T^{(t)} - H^{(t)}\right\|_F^2 + \lambda\left(\left\|Q^{(t)}\right\|_F^2 + \left\|W^{(t)}\right\|_F^2\right). \quad (8)$$

C. Optimization

There are 3 optimization variables needed to update, i.e., $Q^{(t)}, T^{(t)}$, and $W^{(t)}$. Since it is non-convex to jointly optimize these variables, we adopt the alternative convex search (ACS) [11] algorithm to alternatively update each variable. By dividing the optimization process into 3 sub-problems respectively regarding to each of 3 optimization variables, it is easy to solve their closed-form solutions using the least square method. Specifically, we obtain the corresponding closed-form solutions of $Q^{(t)}, T^{(t)}$ and $W^{(t)}$ with Eq. (9), Eq. (10), and Eq. (11), respectively.

$$Q^{(t)} = \mathcal{X}^{(t-1)}\mathcal{X}^{(t-1)^T}\left(Q^{(t-1)}Q^{(t-1)^T} + \lambda I\right)^{-1}. \quad (9)$$

$$T^{(t)} = \left(\eta W^{(t-1)^T}W^{(t-1)} + \eta I\right)^{-1}\left(\eta W^{(t-1)^T}\mathcal{H}^{(t-1)} + Q^{(t-1)}\mathcal{X}^{(t-1)}\right). \quad (10)$$

$$W^{(t)} = \eta T^{(t-1)}T^{(t-1)^T}\left(\eta W^{(t-1)}W^{(t-1)^T} + \lambda I\right)^{-1}. \quad (11)$$

The overall optimization procedure of the proposed method is summarized in Algorithm 1.

Algorithm 1: Least squares model with semi-supervised strategy for crowd counting

input : Training samples X, one-hot label matrix H, hyper-parameters η and λ, number of semi-supervised training stages T.
output: Predicted Label of X_u.

```
Initialize W^(0) using the unit Frombenius norm, t = 0,
  compute the first-stage pseudo label H^(0) using the
  weight Q^(0) solved by Eq. (2), augment
  X = [X, X_u];
while t < T do
    t = t + 1;
    while not converged do
        Update Q^(t) by Eq. (9);
        Update T^(t) by Eq. (10);
        Update W^(t) by Eq. (11);
    end
    Update the pseudo label of X_u by Eq. (7).
end
Predict the label of test data by Eq. (5).
```

(a) UCSD (b) Fudan (c) Bus (d) Canteen (e) Classroom

Fig. 2. Some representative samples on UCSD [12], Fudan [13], and three real-scene datasets [14]

III. Experimental Results and Analysis

A. Benchmarks and experimental configuration

To evaluate the effectiveness of the proposed method, five small-scale scenarios benchmarks are selected, i.e., UCSD [12], Fudan [13], and three real-scene datasets [14]. Some representative samples of these datasets are presented in Fig. 2. Specifically, the UCSD dataset includes 2000 video frames, with crowd counts ranging from 11 to 46. In our experiments, we adopt 800 frames for training and the remaining 1200 for testing. The Fudan dataset contains 1500 frames, with 500 frames used for training and the remaining 1000 frames used for testing. For the three real-scene datasets, three different subsets called Bus, Canteen, and Classroom datasets are included, which consist of 3000, 6000, and 4000 video frames, respectively. In the following experiments, we strictly follow the protocol proposed by [14], i.e., half of them is used for training and the rest for testing. Detailed information for these datasets are summarized in Table I. By following [5], we adopt the Gist [15] feature as the input feature for all datasets used in our experiments. All experiments are implemented by MATLAB R2021a on a PC with AMD R9 5900HX CPU and 16-GB RAM.

TABLE I
DETAIL INFORMATION ON UCSD [12], FUDAN [13], AND THREE REAL-SCENE DATASETS [14]

Dataset	#Frames	#Training	#Test	Min	Max	#Instance
UCSD	2000	800	1200	11	46	49885
Fudan	1500	500	1000	3	18	10308
Bus	3000	1500	1500	0	39	46273
Canteen	6000	3000	3000	0	19	66864
Classroom	4000	2000	2000	24	37	125637

B. Evaluation metrics

Following other crowd counting methods, two standard metrics are adopted, namely Mean Absolute Error (MAE) and Mean Square Error (MSE) fomulated as Eq. (12) and Eq. (13), respectively.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|Pre_i - GT_i|, \tag{12}$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(Pre_i - GT_i)^2, \tag{13}$$

where N is the number of test samples, and Pre_i and GT_i represent the prediction pedestrian number and ground truth of the i-th sample, respectively. While MAE relects the accuracy of the evaluated method, MSE evaluates the robustness of the evaluated algorithm.

C. Comparison experiments

In this section, we report the experimental results on five widely-used datasets in terms of MAE and MSE. The following baselines are chosen: four regression-based competitors termed GPR [12], SRRP [9], MOG-LDL [14], and CSRNet [16],three density map based algorithms termed E3D [17], BSAD [18],and PaDNet [19], and other six classification-based methods termed LC-PDL [20], RA-DPL [21], RBD-DPL [22], PG-DPL [10], SLatDPL [23], and SDR-DPL [6]. Beside MAE and MSE, we also evaluate the consumed time of our methods and its competitors ranked top three: SlatDPL, PG-DPL, and SDR-DPL. We report the average results on 10 random trials.

The results of the comparison experiments on the UCSD datasets are reported in Table II, and that on the Fudan data and the three real-scene datasets are reported in Table III, respectively. As reported, the proposed method outperforms the other competitors in terms of both MAE and MSE on all datasets except for the Canteen dataset. At the meanwhile, the proposed method exceeds its top three competitors in terms of consumed time, which is shown by Fig. 3. These experimental results demonstrate that the proposed method achieve fast and accurate crowd counting by formulating it as a classification problem.

To further investigate our method, we also present the fitting curves on the used five datasets in comparison of SLatDPL and SDR-DPL as shown in Fig. 4. Specifically, we randomly select a specific frame on each corresponding dataset, then we present the fitting curves and their corresponding MAE and MSE placed in each subtitle. As shown, SLatDPL as a ordinary classification method, does not exhibit enough capability for crowd counting because it generates bad value on the bus and the canteen datasets. SDR-DPL is a classification method specially developed for crowd counting and exhibits competitive results. At last, the proposed method still overwhelmingly outperforms its competitors in terms of the shape of fitting curves and the quantitative results, which further verifies the effective of the proposed method.

TABLE II
COMPARISON WITH OTHER REPRESENTATIVE METHODS ON THE UCSD DATASET, AND THE BOLD REPRESENTS THE BEST PERFORMANCE

Dataset	UCSD	
Evaluation Metric	MAE	MSE
GPR [12]	2.24	7.97
SRRP [9]	0.43	3.34
LC-PDL [20]	1.10	3.34
RA-DPL [21]	0.32	0.76
SLatDPL [23]	0.32	0.45
RBD-DPL [22]	0.52	1.28
PG-DPL [10]	0.30	0.37
BSAD [18]	1.00	1.40
E3D [17]	0.93	1.17
PaDNet [17]	0.85	1.06
SDR-DPL [6]	0.27	0.43
Ours	**0.24**	**0.28**

TABLE III
COMPARISON WITH OTHER REPRESENTATIVE METHODS ON THE FUDAN DATASET AND THE THREE REAL-SCENE DATASETS, AND THE BOLD REPRESENTS THE BEST PERFORMANCE

Dataset	Fudan		Bus		Classroom		Canteen	
Evaluation metric	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE
GPR [12]	1.01	2.42	3.98	25.70	0.98	1.64	2.17	6.91
SRRP [9]	0.76	3.09	1.65	28.32	0.38	1.10	**0.91**	3.76
LC-PDL [20]	1.38	4.55	0.70	8.51	0.32	0.90	1.30	4.71
RA-DPL [21]	0.82	2.13	N/A	N/A	0.43	1.92	1.42	6.25
SLatDPL [23]	0.76	2.13	1.26	18.64	0.46	1.96	1.17	4.86
RBD-DPL [22]	0.84	3.49	1.01	9.67	0.45	1.60	1.45	5.42
PG-DPL [10]	0.68	1.34	0.80	9.42	0.32	1.03	0.98	3.19
CSRNet [16]	N/A	N/A	3.84	21.16	0.88	1.19	1.67	4.12
MOG-LPL [14]	N/A	N/A	3.14	18.49	0.67	0.99	1.78	5.24
SDR-DPL [6]	0.75	2.09	0.65	8.31	**0.17**	0.62	1.05	4.57
Ours	**0.47**	**0.82**	**0.56**	**6.20**	**0.17**	**0.47**	0.93	**3.42**

D. t-SNE Feature Visualization

Since we have formulated crowd counting as a classification problem, it is meaningful to visualize the feature space, so that it can be observed whether the proposed method exhibits enough linear separability. Based on such considerations, the t-SNE algorithm [24] is chosen to project the features extracted by the proposed method to two-dimensional subspaces, as shown in Fig. 5. As shown, the original raw features of each crowd dataset lie on different complex manifolds, which is not suitable for classification. However, after learning by the proposed method, the learned feature spaces exhibit satisfactory linear separations, which is the key to generate high

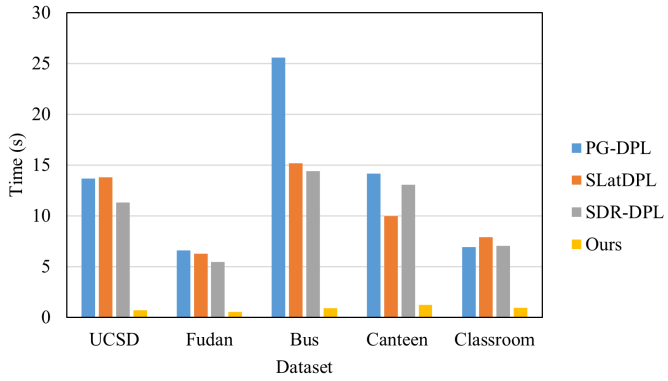


Fig. 3. Consumed time of each method

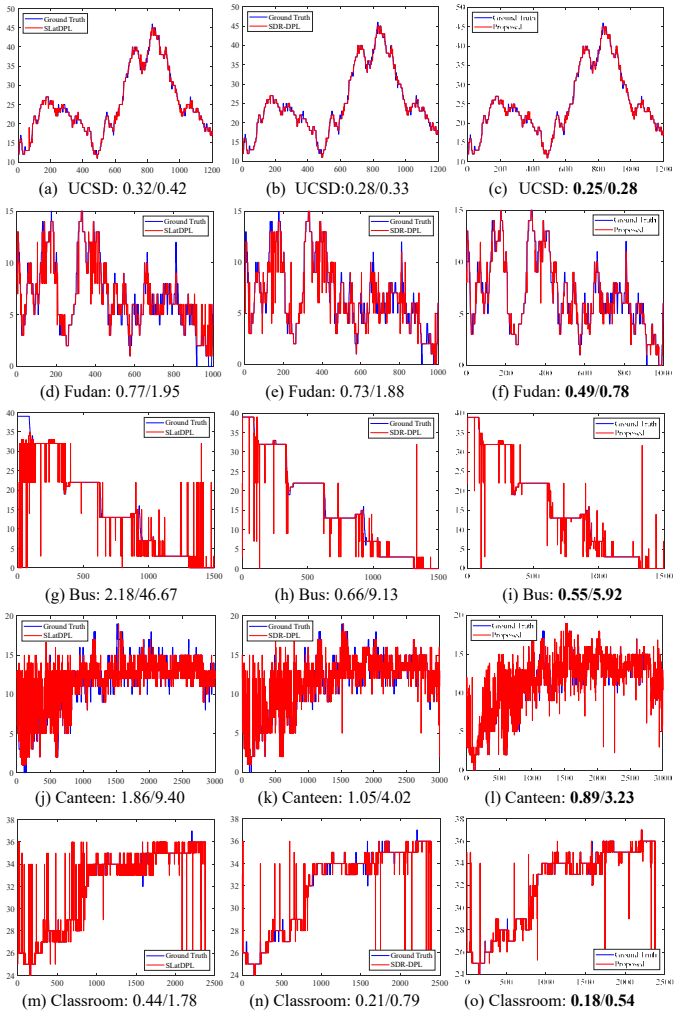


Fig. 4. Fitting curves of SlatDPL, SDR-DPL, and ours on different datasets. The first number in every subtitle indicates MAE while the second denotes MSE

classification results. Such visualization results further verify the effectiveness of the proposed method.

IV. CONCLUSION

In this paper, towards crowd counting, we formulated it as a classification problem and leveraged the philosophy of the least squares model. Aiming at the issue of extremely imbalanced crowd datasets, we proposed a novel semi-supervised strategy to progressively correct the pseudo label of the least squares model during each iteration. Experimental results demonstrate that the proposed method outperformed its competitors in terms of both accuracy and speed. Moreover, the visualization of feature spaces of the proposed method exhibited excellent linear separation, which further proved the feasibility of building a crowd counting-oriented classification method. As a result, fast and accurate crowd counting was achieved.

REFERENCES

[1] H. Li, F. Wang, F. Song, and L. Wang, "Crowd counting method on sparse scene," in *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 2016, pp. 1–6.

[2] S. S. Khan and A. Abedi, "Step counting with attention-based lstm," in *2022 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 2022, pp. 559–566.

[3] L. Pecyna, A. Cangelosi, and A. Di Nuovo, "A deep neural network for finger counting and numerosity estimation," in *2019 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 2019, pp. 1422–1429.

[4] A. Dokania, N. Varia, and J. Senthilnath, "Eobjcount: An evolving spectral and spatial approach for tree count using multispectral satellite images," in *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 2018, pp. 1896–1901.

[5] K. Zhang, H. Wang, W. Liu, M. Li, J. Lu, and Z. Liu, "An efficient semi-supervised manifold embedding for crowd counting," *Applied Soft Computing*, vol. 96, p. 106634, 2020.

[6] T. Wang, H. Luo, K. Zhang, H. Wang, M. Li, and J. Lu, "Salient double reconstruction-based discriminative projective dictionary pair learning for crowd counting," *Applied Intelligence*, vol. 53, no. 2, pp. 1981–1996, 2023.

[7] T. Wang, T. Zhang, K. Zhang, H. Wang, M. Li, and J. Lu, "Context attention fusion network for crowd counting," *Knowledge-Based Systems*, vol. 271, p. 110541, 2023.

[8] J. Wan and A. Chan, "Adaptive density map generation for crowd counting," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 1130–1139.

[9] H. Foroughi, N. Ray, and H. Zhang, "Robust people counting using sparse representation and random projection," *Pattern Recognition*, vol. 48, no. 10, pp. 3038–3052, 2015.

[10] W. Liu, H. Wang, H. Luo, K. Zhang, J. Lu, and Z. Xiong, "Pseudo-label growth dictionary pair learning for crowd counting," *Applied Intelligence*, vol. 51, p. 8913–8927, 2021.

[11] J. Gorski, F. Pfeuffer, and K. Klamroth, "Biconvex sets and optimization with biconvex functions: a survey and extensions," *Mathematical methods of operations research*, vol. 66, pp. 373–407, 2007.

[12] A. B. Chan, Z.-S. J. Liang, and N. Vasconcelos, "Privacy preserving crowd monitoring: Counting people without people models or tracking," in *2008 IEEE conference on computer vision and pattern recognition*. IEEE, 2008, pp. 1–7.

[13] K. Chen, C. C. Loy, S. Gong, and T. Xiang, "Feature mining for localised crowd counting." in *Bmvc*, vol. 1, no. 2, 2012, p. 3.

[14] M. Ling and X. Geng, "Indoor crowd counting by mixture of gaussians label distribution learning," *IEEE Transactions on Image Processing*, vol. 28, no. 11, pp. 5691–5701, 2019.

[15] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *International journal of computer vision*, vol. 42, pp. 145–175, 2001.

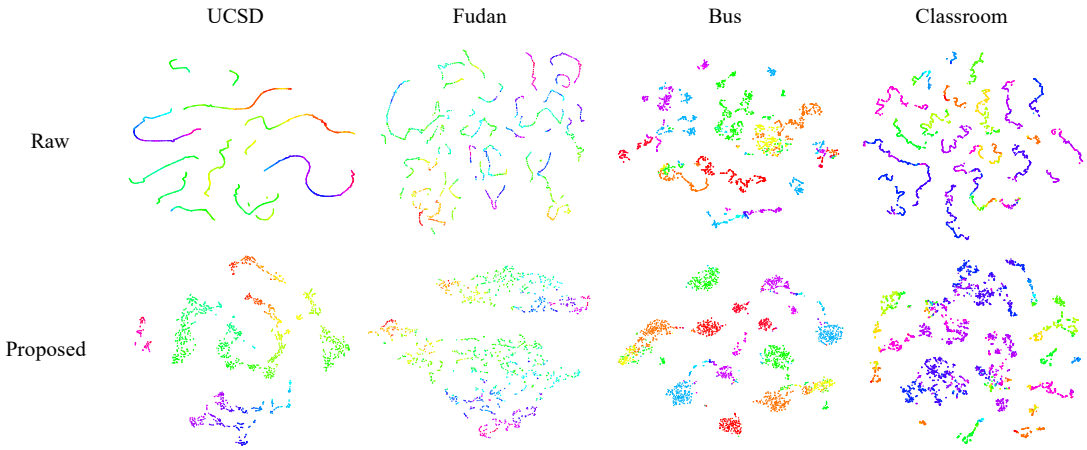


Fig. 5. The visualization of the raw features (the first row) and the features after the proposed method (the bottom row) on each dataset.

[16] Y. C. Li, "Dilated convolutional neural networks for understanding the highly congested scenes/y. li, x. zhang, d. chen," in *Proceedings of the IEEE conference on computer vision and pattern recognition.–IEEE*, 2018, pp. 1091–1100.

[17] Z. Zou, H. Shao, X. Qu, W. Wei, and P. Zhou, "Enhanced 3d convolutional networks for crowd counting," *arXiv preprint arXiv:1908.04121*, 2019.

[18] S. Huang, X. Li, Z. Zhang, F. Wu, S. Gao, R. Ji, and J. Han, "Body structure aware deep crowd counting," *IEEE Transactions on Image Processing*, vol. 27, no. 3, pp. 1049–1059, 2017.

[19] Y. Tian, Y. Lei, J. Zhang, and J. Z. Wang, "Padnet: Pan-density crowd counting," *IEEE Transactions on Image Processing*, vol. 29, pp. 2714–2727, 2019.

[20] Z. Zhang, W. Jiang, Z. Zhang, S. Li, G. Liu, and J. Qin, "Scalable block-diagonal locality-constrained projective dictionary learning," *arXiv preprint arXiv:1905.10568*, 2019.

[21] Y. Sun, Z. Zhang, W. Jiang, Z. Zhang, L. Zhang, S. Yan, and M. Wang, "Discriminative local sparse representation by robust adaptive dictionary pair learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 10, pp. 4303–4317, 2020.

[22] Z. Chen, X.-J. Wu, and J. Kittler, "Relaxed block-diagonal dictionary pair learning with locality constraint for image recognition," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 8, pp. 3645–3659, 2021.

[23] Z. Zhang, Y. Sun, Y. Wang, Z. Zhang, H. Zhang, G. Liu, and M. Wang, "Twin-incoherent self-expressive locality-adaptive latent dictionary pair learning for classification," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 3, pp. 947–961, 2020.