COST-SENSITIVE SPARSE LINEAR REGRESSION FOR CROWD COUNTING WITH IMBALANCED TRAINING DATA

Xiaolin Huang, Yuexian Zou , Yi Wang*

ADSPLAB/ELIP, School of ECE, Peking University, Shenzhen, 518055, China
{xiaolinhuang, wygamle}@pku.edu.cn,*zouyx@pkusz.edu.cn

ABSTRACT

Video-based crowd counting (VCC) is a high demanded technique in many video applications. Existing supervised VCC methods essentially learn an intrinsic mapping function between image features and corresponding crowd counts. However, imbalanced training dataset degrades the performance of VCC significantly. Encouraged by recent success in cost-sensitive learning for image classification with imbalance dataset, we propose a novel cost-sensitive sparse linear regression VCC method (CS-SLR-VCC). Specifically, a sparse linear regression (SLR) model is firstly learned and the modelling errors associated with each training data are calculated accordingly. Then, aiming to eliminate the adverse effect of the high modelling errors of SLR model due to imbalanced data, all modelling errors are taken as prior knowledge to design sample-related weighting factors. Thus, a cost-sensitive SLR model is reformulated and its optimal solution is derived. Extensive experiments conducted on public UCSD and Mall benchmarks demonstrate the superior performance of our proposed CS-SLR-VCC method.

***Index Terms*—**Video-based crowd counting, sparse linear regression, cost-sensitive learning, imbalanced data, model fitting

1. INTRODUCTION

In recent years, visual-based crowd counting (VCC) has attracted much attention in the communities of multimedia and computer vision due to its potential applications in public security [3], retail sectors profiling [5], resource management, urban planning and emergency detection [7].

The task of VCC is to label a scene image automatically with the exact people count. However, the severe occlusion, low resolution and perspective distortion make it difficult to detect and track people. Most state-of-art VCC methods are regression based [6, 8-13]. Essentially, the regression based VCC methods learn an intrinsic mapping between image features and crowd counts with labeled training dataset, which are suitable for the real-word applications.

This work is supported by the Shenzhen Science & Technology Fundamental Research Program (No: JCYJ20150430162332418)

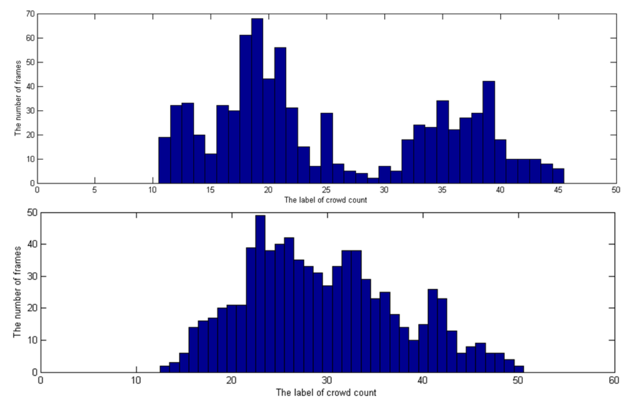


Fig.1. The data distribution of two benchmarking crowd counting datasets. Top: UCSD crowd dataset. Bottom: Mall crowd dataset.

In principle, existing regression based VCC methods consist of two key modules: feature extraction (FE) module and mapping function learning (MFL) module. For the FE module, the recent mainstream regression-based methods [8-11] make use of the combined features by concatenating texture, segment and edge features together to get better performance. However, combining features directly may generate correlated features to some extent [8], which may degrade the discriminative capability of the features. For the MFL module, many different regression models have been introduced for solving the VCC problem [6, 8, 12, 14], among which the Ridge Regression (RR) model (one linear regression model) has shown good performance in crowd counting [8]. Research outcomes also reveal that the VCC performance of the existing regression methods including the RR model degrades significantly when the distribution of training data is imbalanced [11]. Evaluating the existing benchmarking datasets for crowd counting such as Mall and UCSD that are imbalanced, we plot their training data distribution in Figure 1. It is clear to see that the distribution of the training data is imbalanced. Some of them have much higher quantity (Such as the crowd count label 19 in UCSD dataset), but some have much less samples (such as the crowd count label 29 in UCSD dataset). Inspired by recent success of cost-sensitive learning for solving imbalanced classification problem, in this study, we make an effort to provide a new solution to the VCC problem with imbalanced training data.

Cost-sensitive learning is designed to solve imbalanced learning problem by using different cost matrices that describe the cost for misclassifying any data example [15]. There are some successful researches about the cost-sensitive learning based image classification with imbalanced training data. Nguyen [16] proposed a cost-sensitive extension of Regularized Least Square algorithm that penalizes errors of different samples with different weights. Sun [17] proposed the cost-sensitive boosting algorithms, which are developed by introducing cost items into the learning framework of AdaBoost. Both of Nguyen and Sun's works achieved significantly performance improvement.

Inspired by the works discussed above, this paper proposes a novel cost-sensitive sparse linear regression method to solve imbalanced VCC problem. To achieve this purpose, a two-stage VCC framework is developed. In the first stage, considering the nature of sparse constraint that selects discriminative features and the encouraging performance of linear regression, a sparse linear regression (SLR) model is learned based on the training dataset. Accordingly, the modelling error for each training data can be calculated. However, experimental results showed that some training data may not satisfy the assumed distribution due to complicated scenes which will cause much higher modelling errors. This motivates us to adopt the idea of cost-sensitive learning to eliminate the adverse effect of high modelling errors in the second stage. Specifically, we take all modelling errors as prior knowledge and design a sample-related weighting factor to formulate a novel cost-sensitive SLR model for VCC (named as CS-SLR-VCC method). Extensive experiments have been conducted to evaluate the performance of our proposed CS-SLR-VCC method. Experimental results on public UCSD and Mall databases demonstrate the effectiveness of our proposed CS-SLR-VCC method.

The rest of the paper is organized as following. Section 2 describes the framework and detailed methodology of our proposed CS-SLR-VCC method. Section 3 presents experimental results. Finally, Section 4 concludes our paper.

2. METHODOLOGY

2.1 Overview of the proposed method

The flowchart of our proposed two-stage VCC framework is illustrated in Figure 2. Firstly, in the training stage 1, feature extraction module will extract the segment feature, edge feature and texture features from input training images. According to the extracted image features, a sparse linear regression model is trained. Accordingly, the modelling errors associated all training data can be calculated and are used to design the weighting factor for each training image. In the training stage 2, a cost-sensitive sparse linear regression model (CS-SLR) will be jointly learned using input training data and their associated weighting factors. In the testing stage, an unseen image is given and its feature can be extracted as that in training stage, and then the extracted feature vector and trained $\boldsymbol{\beta}$ are used to predict the labels of testing images. The details will be discussed separately in the following subsections.

2.2. Feature extraction

Feature extraction is an important module for VCC as shown in Figure 2. In our study, a mainstream feature extraction method

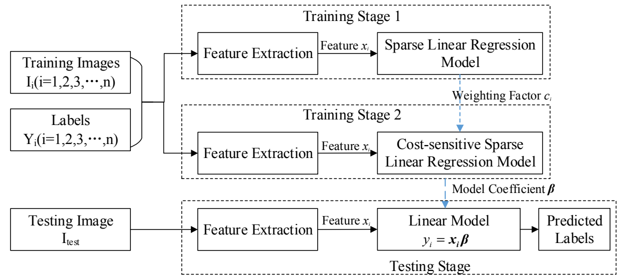


Fig.2. Diagram of our proposed method

is adopted as in [6] which combines the following three types of features to form a bank of feature vectors.

Segment features: The segment features are the most popular representation for VCC, which are obtained by background subtraction and can capture the shape and size of crowd. Here, the *Area*, *Perimeter*, *Perimeter edge orientation* and *Blob count* features are adopted for the representation of segment features.

Edge features: The edges features describe the strong information about the number of people in the segment. Normally, a canny edge detector is applied to extract the edge from the segment and then the *Edge length*, *Edge orientation* and *Minkowski dimension* features are obtained.

Texture features: The texture features are based on the Gray-Level Co-occurrence Matrix (GLCM) and contains significant information about the count of the crowd. *Homogeneity*, *Energy* and *Entropy* features based on GLCM are utilized as the measurement of the smooth, uniformity and complexity of texture, respectively.

Note that, all frames are perspective normalized and transformed to gray-scale to extract features.

2.3. The sparse linear regression model

As a kind of linear regression, RR has shown a promising performance for crowd counting regression [8]. Thus, in this paper, we choose RR as the basic linear regression model. Considering that the combined features may generate correlated features to some extent, in order to select better representative features, we adopt the linear regression model with sparse constraint. For making a complete presentation, the basic linear regression model RR will be introduced as follows.

Given n training samples (feature vectors), let $\boldsymbol{x}_i$ represents an m-Dim feature vector of the i-th sample, y_i represents its crowd count. The RR model can be obtained by the following optimization problem:

$$\arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left\| y_i - \boldsymbol{x}_i \boldsymbol{\beta} \right\|_2^2 + \lambda \left\| \boldsymbol{\beta} \right\|_2^2 \tag{1}$$

Where λ denotes the trade-off parameter between the regularized term and the loss function. $\boldsymbol{\beta}$ is the coefficient vector of the RR model. Research shows that the solution of RR is stable, however it does not give an easily interpretable model [18]. Moreover, m-dimension feature components are not equally important for the linear regression model. Therefore, it is necessary to select discriminative features. Obviously, a better interpretable linear regression model can be determined

Algorithm 1: *Algorithm for solving CS-SLR*

Input :

—parameter λ_1 and λ_2

—$\mathbf{X} \in \mathbb{R}^{m\times n}$ is the centred and standardized features

—$\mathbf{y} \in \mathbb{R}^{1\times n}$ is the labels of samples

Output:

—coefficient $\boldsymbol{\beta}_{CS\text{-}SLR}$

1. Execute SLR algorithm to obtain the coefficient $\boldsymbol{\beta}_{SLR}$
2. Obtain the absolute error between predictions and labels
3. Remove outliers
4. Calculate c_i and $\boldsymbol{V}$ according to Eqn. (6) and Eqn. (8)
5. Calculate $\boldsymbol{X}^*$, $\boldsymbol{\beta}^*$ and γ according to Eqn. (9)
6. Solve the Eqn. (10) using the LARS algorithm
7. Obtain the coefficient $\boldsymbol{\beta}_{CS\text{-}SLR}$ according to Eqn. (11)

by the following sparse constrained optimization problem, problem, termed as sparse linear regression (SLR) model:

$$\arg\min_{\boldsymbol{\beta}_1} \sum_{i=1}^{n} \left\| y_i - \boldsymbol{x}_i \boldsymbol{\beta}_1 \right\|_2^2 + \lambda_1 \left\| \boldsymbol{\beta}_1 \right\|_1 + \lambda_2 \left\| \boldsymbol{\beta}_1 \right\|_2^2 \tag{2}$$

where parameters λ_1 and λ_2 trade off the sparsity of the model coefficient versus the stability of model coefficient given by L_1 and L_2 norms respectively and $\boldsymbol{\beta}_1$ is a SLR coefficient vector. It is obvious that $\boldsymbol{\beta}_1$ is a sparse vector compared with $\boldsymbol{\beta}$, which indicates that some feature components are more important to represent training samples. Moreover, owing to the nature of L_1 norm, sparsity is able to reduce the disturbance of noisy variables and improves the interpretation of system, which is helpful for increasing the modeling accuracy and robustness.

Many approaches have been proposed to solve this sparse linear regression model, Zou provided a solution to handle with this kind of model and named it as Elastic net [19]. Zou transformed the Eqn. (2) into the flowing formulation:

$$\arg\min_{\boldsymbol{\beta}^*} \left\| \boldsymbol{y}^* - \boldsymbol{X}^* \boldsymbol{\beta}^* \right\|_2^2 + \gamma \left\| \boldsymbol{\beta}^* \right\|_1 \tag{3}$$

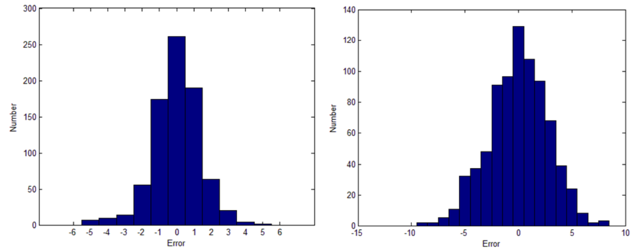
and the SLR model coefficient is obtained by

$$\begin{aligned} \boldsymbol{\beta}_{SLR} &= (1+\lambda_2)\boldsymbol{\beta} = (1+\lambda_2)\frac{1}{\sqrt{1+\lambda_2}}\boldsymbol{\beta}^* \\ &= \sqrt{1+\lambda_2}\,\boldsymbol{\beta}^* \end{aligned} \tag{4}$$

It is noted that the input features $\boldsymbol{x}_i$ are centered and standardized. The detailed information about the solution of this model can refer to literature [19].

2.4. The proposed VCC method based on Cost-sensitive sparse linear regression model

It is noted that, for imbalanced data classification problem, the standard classifiers generally perform poorly because they just pursue the minimization of the overall error and ignore the errors of minor classes, which lead the model to have a tendency to misclassify data to be major classes. The similar observation is found in linear regression problem. Due to the imbalance of data, the errors of minor classes contribute few to the overall errors, therefore the estimated model will have a



(a) UCSD (b) Mall

Fig. 3. Error distribution on two benchmark datasets

bias toward major classes. Consequently, test samples belonging to the minor classes will have high errors. Fortunately, cost-sensitive learning can handle the imbalanced classification problem by assuming higher misclassification costs with samples in minor class [17]. In order to solve the similar problem in regression, here we introduce cost items into the regression learning framework and propose a new cost-sensitive SLR (CS-SLR) model as follows:

$$\arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} c_i \left\| y_i - \boldsymbol{x}_i \boldsymbol{\beta} \right\|_2^2 + \lambda_1 \left\| \boldsymbol{\beta} \right\|_1 + \lambda_2 \left\| \boldsymbol{\beta} \right\|_2^2 \tag{5}$$

where $\boldsymbol{\beta}$ is the coefficient vector of the CS-SLR model, parameters λ_1 constrains the sparsity of the model coefficients and parameters λ_2 constrains the stability of model coefficients given by L_1 and L_2 norms respectively. In our study, we followed the common practice to select λ_1 and λ_2 where the curve of the counting accuracy versus λ_1 and λ_2 are obtained. The λ_1 and λ_2 associated with the peaks are chosen. Then the selected λ_1 and λ_2 are fixed for all experiments. And c_i is the cost weighting factor associated with the i-th modeling error.

Since the ignorance of minor classes leads to large modelling error, to eliminate the negative effect of the modelling error due to imbalanced data, more attention should be paid to the minor classes, which means that we should give a high cost weighting factor c_i for them. In order to achieve this purpose, an exponential function is used to determine the weighting factor as [17]:

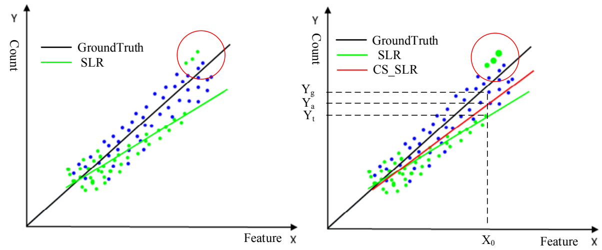
$$c_i = \exp(|\, y_i - \boldsymbol{x}_i \hat{\boldsymbol{\beta}} \,|) / Z \tag{6}$$

where $\hat{\boldsymbol{\beta}}$ is the model coefficient vector estimated by Eqn. (2) and Z is a normalized factor.

It is noted that due to the fact that the higher error is assigned a larger c_i which makes our method sensitive to outliers. In order to reduce the negative effects caused by outliers, we use the preprocessing to remove the impact of outliers. Figure 3 shows the error distribution of UCSD and Mall dataset which accords with normal distribution. Miller [20] points out that data is taken at the desired confidence intervals (usually 95%) of the standard derivative. Here, we adopt the same confidence intervals as Miller. This means when the modelling error is beyond 95% of its standard derivative, its associated training sample will be considered as an outlier and will not be taken into account.

2.5. Solving the CS-SLR model

To solve the CS-SLR problem described in Eqn. (5), some operations are required to transform Eqn. (5) into following



(a) The SLR model (b) The CS-SLR model

Fig. 4. Conceptual illustration of the effect of CS-SLR (green dots represent the training data and blue dots represents the testing data)

expression:

$$\arg\min_{\beta}\sum_{i=1}^{n}\left\|\sqrt{c_i}\,y_i-\sqrt{c_i}\,\boldsymbol{x}_i\boldsymbol{\beta}\right\|_2^2+\lambda_1\|\boldsymbol{\beta}\|_1+\lambda_2\|\boldsymbol{\beta}\|_2^2 \qquad (7)$$

Let's denote

$$\begin{aligned}\boldsymbol{V}&=diag(c_1,c_2,\ldots,c_n)\\ \boldsymbol{V}^{1/2}&=(\boldsymbol{V}^{1/2})^T=diag(\sqrt{c_1},\sqrt{c_2},\ldots,\sqrt{c_n})\end{aligned} \qquad (8)$$

and

$$\begin{aligned}\boldsymbol{X}^*&=\frac{1}{\sqrt{1+\lambda_2}}\begin{pmatrix}\boldsymbol{V}^{1/2}\boldsymbol{X}\\ \sqrt{\lambda_2}\boldsymbol{I}\end{pmatrix},\ \boldsymbol{y}^*=\begin{pmatrix}\boldsymbol{V}^{1/2}\boldsymbol{y}\\ \boldsymbol{0}\end{pmatrix}\\ \gamma&=\frac{\lambda_1}{\sqrt{1+\lambda_2}},\ \boldsymbol{\beta}^*=\sqrt{1+\lambda_2}\,\boldsymbol{\beta}\end{aligned} \qquad (9)$$

By substituting Eqn. (9) to Eqn. (7), we get the following formulation which is similar with Eqn. (3) whereas the $\boldsymbol{X}^*$ and $\boldsymbol{y}^*$ here are different from Eqn. (3):

$$\arg\min_{\beta}\left\|\boldsymbol{y}^*-\boldsymbol{X}^*\boldsymbol{\beta}^*\right\|_2^2+\gamma\left\|\boldsymbol{\beta}^*\right\|_1 \qquad (10)$$

Obviously, Eqn. (10) is a standard lasso problem, which can be easily solved by the LARS algorithm [21]. Similar to the Eqn. (3), we obtain the $\boldsymbol{\beta}_{CS\text{-}SLR}$ using the following equation:

$$\boldsymbol{\beta}_{CS-SLR}=\sqrt{1+\lambda_2}\,\boldsymbol{\beta}^* \qquad (11)$$

The details of CS-SLR algorithm is summarized in **Algorithm 1.**

2.6. Some discussions about the CS-SLR

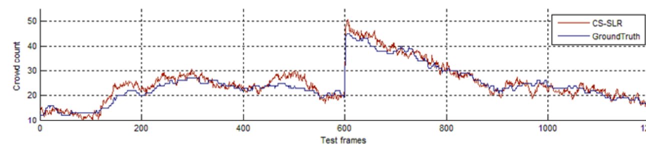
For the VCC problem, crowd may be thick in one period time, and thin in another period time, which usually leads to the imbalanced training data (the green dots circled in Figure 4 (a)). However for testing, due to unknown pedestrian flows, the ground truth crowd count may randomly distributed (blue dots in Figure 4). For conceptual visualization purpose, we take the 1-dimension data points for example and all these data points satisfy the linear distribution. In Figure 4, the black line represents the ground truth model. SLR model gives the result in green line which bias from the ground truth model due to imbalanced training data. Our proposed CS-SLR model shown in red line is more close to the ground truth model. As shown in Figure 4, it is clear that, for a given feature X_0 , crowd count y_g, y_t and y_a represent the ground truth crowd count, the count predicted by SLR and CS-SLR, respectively. We can see that the bias $|y_a-y_g|$ is smaller than the bias $|y_t-y_g|$. This illustrates the capability of CS-SLR to provide better performance than that of SLR. It also can be seen from Figure 4 that, when the training data is balanced, the modelling error of each sample is similar which means the weighting factors are close and CS-SLR degenerates to a SLR model.

(a) The UCSD dataset (b) The Mall dataset

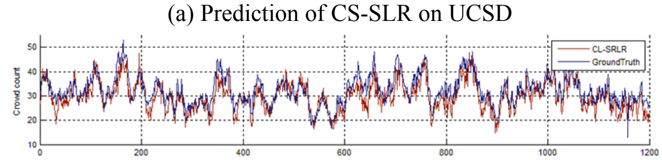
Fig. 5. Example samples in two benchmark datasets

Table 1. Dataset properties: N_f is the number of frames, R is the resolution, FPS is the frame rate, D is the density (minimum and maximum number of people in the ROI), and T_p is the number of crowd in total.

Dataset	N_f	R	FPS	D	T_p
UCSD	2000	238×158	10	11-46	49885
Mall	2000	320×240	<2	13-53	62325



(a) Prediction of CS-SLR on UCSD



(b) Prediction of CS-SLR on Mall

Fig. 6. Comparison between ground truth and prediction

3. EXPERIMENTS

3.1. Datasets and settings

Datasets – To demonstrate the effectiveness of our proposed method, two benchmarking datasets are introduced here: UCSD [6] and Mall [8] which represent the indoor and outdoor scene respectively. The detailed information of the two datasets are given in Table 1. As shown in the Figure 5, both datasets have object occlusion and object shadow, and Mall dataset have more sever lighting challenging and glass surface reflection influence.

Evaluation protocol – Following the work [8], we employ the Frame 601-1400 for training and the rest for testing in UCSD dataset. And for the Mall dataset, we utilize Frame 1-800 as the training dataset and keep the remaining 1200 frames as the testing dataset.

Table 2. Crowd counting performance comparison

Method	UCSD			Mall		
	MAE	MSE	MDE	MAE	MSE	MDE
LSSVR[1]	2.20	7.29	0.107	3.51	1.82	0.108
KRR[2]	2.16	7.45	0.107	3.51	18.1	0.108
RFR[4]	2.42	8.47	0.116	3.91	21.5	0.121
GPR[6]	2.24	7.91	0.112	3.72	20.1	0.115
RR[8]	2.25	7.82	0.110	3.59	19.0	0.110
CA-RR[11]	2.07	6.86	0.102	3.43	17.7	0.105
WRR[10]	2.05	6.75	0.102	3.44	18.0	0.105
SLR	2.03	5.96	0.089	3.35	16.62	0.105
CS-SLR	**1.83**	**5.04**	**0.079**	**3.23**	**15.77**	**0.104**

Table 3. Cumulative score on UCSD at different levels (%)

	1	**2**	**3**	**4**	**5**	**6**	**7**
SLR	27.67	53.17	75.83	90.83	97.33	99.58	100.00
CS-SLR	30.92	62.75	80.83	93.17	97.66	99.33	100.00

Table 4. Cumulative score on Mall at different levels (%)

	1	**2**	**3**	**4**	**5**	**6**	**7**
SLR	17. 17	33. 58	48.67	65.08	77.08	86.67	92.83
CS-SLR	18. 17	36.50	52.17	67.17	78.75	88.42	93.17

Table 5. Performance on imbalanced UCSD dataset

	0	**11-15**	**16-20**	**21-25**	**26-30**	**31-35**	**36-40**
SLR	2.03	2.41	2.64	2.75	4.22	5.81	13.78
CS-SLR	1.83	1.99	2.27	2.29	3.27	7.54	12.13

Evaluation metrics – We employ three evaluation metrics as used in [8]. They are mean absolute error (MAE) ε_{abs}, mean squared error (MSE) ε_{sqr} and mean deviation error (MDE) ε_{dev} respectively.

$$\varepsilon_{abs} = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - \hat{y}_i\right|, \varepsilon_{abs} = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - \hat{y}_i\right|, \varepsilon_{abs} = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - \hat{y}_i\right|,$$

where N is the total number of test frames, y_i is the actual number of crowd and $\hat{y}_i$ is the predicted number of crowd in the i-th frame.

3.2. Experimental results

Table 2 compares the crowd counting performance of our proposed method with eight different methods, all of which are based on regression by using the UCSD and Mall benchmark dataset. The results show that the proposed CS-SLR method perform well on both benchmark datasets using all three metrics. By comparing the method SLR and CS-SLR, we can find that CS-SLR has a better performance on both benchmarking datasets, which demonstrates the superiority of our proposed method. Figure 6 shows the comparison between the ground truth and our CS-SLR method's prediction. To further evaluate the performance of our proposed method, we conduct another two experiments introduced in the following sections.

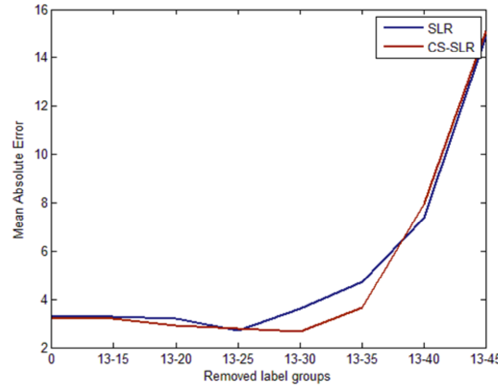


Fig.7. Performances on imbalanced Mall dataset

3.3. The effectiveness of CS-SLR

To evaluate the effectiveness of the proposed CS-SLR for the VCC problem, the cumulative score (CS) metric is adopted which was first applied in [22]. Specifically, we define $\mathrm{CumScore}(l) = \frac{M_{e\le l}}{M} \times 100\%$ where $M_{e\le l}$ is the number of test images on which the count estimation makes an absolute error no higher than integral l, which is termed as error level. Table 3 and Table 4 show the cumulative score information on UCSD and Mall dataset at different error levels. For cumulative score, the higher value means better performance.

It is clear to see from Table 3 and 4 that for most error levels, CS-SLR has a higher cumulative score, which indicates that crowd counts predicted by our proposed CS-SLR model are much closer to the ground truth. These results validate the conclusions presented in Section 2.6.

3.4. Against imbalanced data distribution

According to the discussion in Section 1, the imbalanced data is one major challenge for learning a good regression function. In this subsection, we verify the effect of our proposed CS-SLR method against the imbalanced data. In order to achieve this goal, data of certain crowd counts are removed to make the data more imbalanced. In our experiments, we divide crowd counts into seven groups according to their labels and the label-based group is removed from training data one by one. Table 5 and Figure 7 show the performance of SLR and CS-SLR with the remove of label-based groups. It is evident that when more label-based groups were removed, the performance of both SLR and CS-SLR methods degrades, however, the performance of our CS-SLR model degrades more slowly, which verify the effectiveness of the proposed CS-SLR method with imbalanced training data. It is also noticed that when only two label-based groups left, CS-SLR is not competitive. This is due to the fact that the training dataset is not large enough to represent the entire distribution when only two label-based groups are left. In real-word applications, we observe that the crowd counts normally have more than 40 counting levels. As a conclusion, our proposed CS-SLR method is suitable for crowd counting with imbalanced data distribution.

4. CONCLUSION

This paper introduces a novel cost-sensitive sparse linear regression method for solving the imbalanced crowd counting problem. More specifically, a two-stage VCC framework is developed. In the first stage, we adopt the SLR to select discriminative features and obtain the modelling error of each training sample accordingly. In the second stage, a novel cost-sensitive SLR is formulated by taking all modelling errors as prior knowledge and gives different sample with different weighting factor to combat the imbalanced data distribution. Extensive experiments conducted on public UCSD and Mall benchmarks have confirmed the effectiveness of the proposed method for crowd counting with the imbalanced data.

5. REFERENCES

[1] T. Van Gestel, J. Suykens, B. De Moor, and J. Vandewalle, "Automatic relevance determination for least squares support vector machine regression," in International Joint Conference on Neural Networks, 2001, pp. 2416-2421.

[2] S. An, W. Liu, and S. Venkatesh, "Face recognition using kernel ridge regression," in IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp. 1-7.

[3] A. Albiol and J. Silla, "Statistical video analysis for crowds counting," in IEEE International Conference on Image Processing (ICIP), 2009, pp. 2569-2572.

[4] A. Liaw and M. Wiener, "Classification and regression by randomForest," R news, vol. 2, pp. 18-22, 2002.

[5] C. C. Loy, K. Chen, S. Gong, and T. Xiang, "Crowd counting and profiling: Methodology and evaluation," in Modeling, Simulation and Visual Analysis of Crowds, ed: Springer, 2013, pp. 347-382.

[6] A. B. Chan, Z.-S. J. Liang, and N. Vasconcelos, "Privacy preserving crowd monitoring: Counting people without people models or tracking," in IEEE Conference on Computer Vision and Pattern Recognition, 2008, pp. 1-7.

[7] G. L. Hamza-Lup, K. Hua, M. Le, and R. Peng, "Dynamic plan generation and real-time management techniques for traffic evacuation," IEEE Transactions on Intelligent Transportation Systems, vol. 9, pp. 615-624, 2008.

[8] K. Chen, C. C. Loy, S. Gong, and T. Xiang, "Feature Mining for Localised Crowd Counting," in BMVC, 2012, p. 3.

[9] Z. Zhang, M. Wang, and X. Geng, "Crowd counting in public video surveillance by label distribution learning," Neurocomputing, 2015.

[10] K. Chen and J.-K. Kamarainen, "Learning to Count with Back-propagated Information," in 22nd International Conference on Pattern Recognition (ICPR), 2014, pp. 4672-4677.

[11] K. Chen, S. Gong, T. Xiang, and C. C. Loy, "Cumulative attribute space for age and crowd density estimation," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2013, pp. 2467-2474.

[12] A. B. Chan and N. Vasconcelos, "Counting people with low-level features and Bayesian regression," IEEE Transactions on Image Processing, vol. 21, pp. 2160-2177, 2012.

[13] R. Ma, L. Li, W. Huang, and Q. Tian, "On pixel count based crowd density estimation for visual surveillance," in 2004 IEEE Conference on Cybernetics and Intelligent Systems, 2004, pp. 170-173.

[14] L. Fiaschi, R. Nair, U. Koethe, and F. Hamprecht, "Learning to count with regression forest and structured labels," in 21st International Conference on Pattern Recognition (ICPR), 2012, pp. 2685-2688.

[15] H. He and E. Garcia, "Learning from imbalanced data," IEEE Transactions on Knowledge and Data Engineering, vol. 21, pp. 1263-1284, 2009.

[16] N. H. Vo and Y. Won, "Classification of unbalanced medical data with weighted regularized least squares," in Frontiers in the Convergence of Bioscience and Information Technologies, 2007, pp. 347-352.

[17] Y. Sun, M. S. Kamel, A. K. Wong, and Y. Wang, "Cost-sensitive boosting for classification of imbalanced data," Pattern Recognition, vol. 40, pp. 3358-3378, 2007.

[18] R. Tibshirani, "Regression shrinkage and selection via the lasso," Journal of the Royal Statistical Society. Series B (Methodological), pp. 267-288, 1996.

[19] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," Journal of the Royal Statistical Society: Series B (Statistical Methodology), vol. 67, pp. 301-320, 2005.

[20] J. N. Miller, "Basic statistical methods for analytical chemistry. Part 2. Calibration and regression methods. A review," Analyst, vol. 116, pp. 3-14, 1991.

[21] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," The Annals of statistics, vol. 32, pp. 407-499, 2004.

[22] X. Geng, Z.-H. Zhou, and K. Smith-Miles, "Automatic age estimation based on facial aging patterns," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, pp. 2234-2240, 2007.