# Incomplete Cigarette Code Recognition via Unified SPA Features and Graph Space Constraints

Huiming Ding[1], Zhifeng Xie[1(✉)], Jundong Lai[2], Yanmin Xu[3], and Lizhuang Ma[4]

[1] Shanghai University, Shanghai, China
zhifeng_xie@shu.ed.cn
[2] Shanghai Tobacco Group Co., Ltd., Shanghai, China
[3] Shanghai Tobacco Monopoly Administration, Shanghai, China
[4] Shanghai Jiao Tong University, Shanghai, China

**Abstract.** Cigarette code is a 32-character string printed on a cigarette package, which can be used by tobacco administrations to determine the legality of distribution. Unfortunately, the recognition task for incomplete cigarette code often suffers from lowered recognition accuracy and the destruction of semantic context due to complex backgrounds and damaged characters. This paper proposes an end-to-end recognition network for incomplete cigarette code to improve recognition accuracy and estimate character landmarks. The proposed network first extracts multi-scale features using feature pyramid networks (FPN), then utilizes a spatial attention (SPA) mechanism to yield unified SPA features and integrates them into instance segmentation. This strengthens spatial representation ability and improves the recognition accuracy. A graph convolutional network (GCN) is introduced to construct graph space constraints and calculate character spatial correlations and accurately estimates missing character landmarks. Finally, we employ the Hungarian algorithm to align recognition characters with estimated landmarks and fill missing characters with '*' to preserve the complete semantic context, and produce the final regularized cigarette code. The experimental results demonstrate that our proposed network reduces time consumption and improves recognition accuracy, surpassing the state-of-the-art methods.

**Keywords:** Cigarette code recognition · Spatial attention mechanism · Graph convolutional network · Deep learning

## 1 Introduction

Cigarette code is a 32-character string printed onto cigarette packages, which contain information including the production date, the manufacturer, and the cigarette retail store, often used by tobacco administration to determine the


L. Fang et al. (Eds.): CICAI 2022, LNAI 13605, pp. 59–70, 2022.
https://doi.org/10.1007/978-3-031-20500-2_5

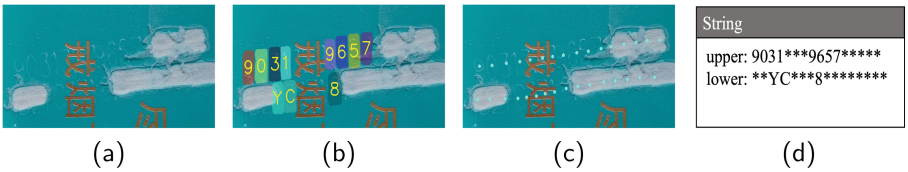


**Fig. 1.** Recognition task for cigarette code. (a) example incomplete cigarette code with complex background and damaged characters; (b) recognized cigarette code result without damaged characters; (c) estimated cigarette code landmarks including damaged characters; (d) final recognized result with 32 characters, where missing characters are filled by '*'.

legality of distribution [28,29]. Unfortunately, the recognition task for cigarette code suffers from many difficulties such as complex backgrounds, blurred printing, damaged characters, and even large broken areas, which greatly reduce the accuracy of character recognition. Missing character landmarks need to be estimated to produce a 32-character code with complete semantic correlation. As shown in Fig. 1, when given a source image of incomplete cigarette code, the recognition task first recognizes undamaged characters, then estimates full character landmarks, including those of damaged characters. Finally, the recognized characters and estimated landmarks are regularized to yield an 32-character cigarette code where any missing characters are filled by '*'.

The recognition of incomplete cigarette code can be divided into two subtasks: character recognition and landmark estimation. Current character recognition methods [8,15,17] have difficulty dealing with complex backgrounds and blurred printing to construct the spatial semantic context of full characters for incomplete cigarette code. Meanwhile, methods [6,22,27–31] were unable to establish the spatial correlation between characters for incomplete cigarette code in the process of landmark estimation.

To address the above problems, this paper proposes a recognition network for incomplete cigarette code based on unified spatial attention (SPA) features and graph space constraints. The network architecture is divided into four modules, as shown in Fig. 2: (1) Feature Extraction. We use the residual network (ResNet) [5] and feature pyramid network (FPN) [11] to extract multi-scale features of incomplete cigarette code. (2) Instance Segmentation. Inspired by 'SOLOv2' [24], we introduce a spatial attention mechanism to yield unified SPA features, which enhance the spatial representation for characters and backgrounds. (3) Landmark Estimation. We first estimate initial landmarks using integral regression based on the unified SPA features. We then construct graph space constraints based on graph convolutional network (GCN) [6] to optimize the spatial correlation between characters, and therefore can accurately estimate the full character landmarks. (4) Text Regularization. The Hungarian algorithm [7] is introduced to align the instance segmentation results and estimated landmarks.

In addition, we construct and execute experiments on a dataset containing 15,000 images of incomplete cigarette code, including missing characters,

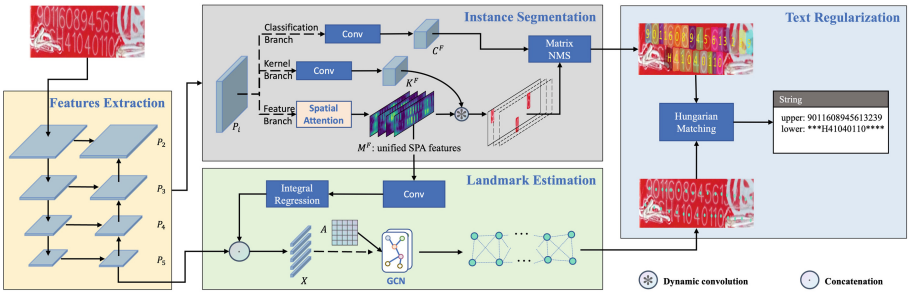


**Fig. 2.** The network architecture of our recognition network for incomplete cigarette code consists of four modules: (1) Features Extraction; (2) Instance Segmentation; (3) Landmark Estimation; (4) Text Regularization.

complex background, and blurred printing. The experimental results demonstrate that our proposed method can accurately recognize characters, effectively estimate the landmarks, and produce the recognized result of cigarette code with 32 characters.

## 2 Related Work

Some popular techniques relevant to the recognition task for incomplete cigarette code can be classified into three categories: text recognition, landmark estimation, and graph convolutional networks.

Recognition methods based on deep learning are widely used in complex scenes [1,8,15,17,28,29]. However, the performances of these methods were weak for incomplete cigarette code because of the complex backgrounds and blurred printing. Otherwise, attention mechanisms have been used in some studies [10,14,18] to draw attention to channel and spatial dimensions, which can strengthen feature representation and improve text recognition performance. To this end, inspired by SOLOv2 [25], in this paper we introduce a spatial attention mechanism to yield unified SPA features, and strengthen the spatial representation for characters and backgrounds.

Incomplete cigarette code recognition requires landmark estimation for both present and missing characters. State-of-the-art methods for landmark estimation [2,22,23,26] have been applied mainly to face localization and human pose estimation. They introduced heatmap-supervised [23], multi-stage supervised training [26], multi-scale features [2] to estimate landmarks. Notably, Sun et al. [22] reconsidered the task of human pose estimation using an integral regression perspective that can generate joint coordinates from heatmap. Inspired by this, in this study we apply integral regression to generate the initial character landmarks for further GCN-based optimization.

Character correlation can be modelled as a graph-based relationship, which can be learned using GCN [6,30]. GCN-based landmark estimation works [3,31]

usually built graph neural networks through a prior graph node relationship to achieve better detection performance. Therefore, we have employed GCN to construct graph space constraints for the 32 cigarette code characters to improve landmark estimation for incomplete cigarette code.

## 3 Method

In this paper, we propose an end-to-end recognition network for incomplete cigarette code which can: accurately recognize characters using multi-scale SPA features; estimate full-character landmarks by constraining the character spatial correlation; and produce recognition results with semantic context. As shown in Fig. 2, the network is divided into feature extraction, instance segmentation, landmark estimation, and text regularization modules.

### 3.1 Features Extraction Using FPN

The FPN structure can effectively extract the multi-scale features of an incomplete cigarette code to obtain the character location and semantic information [11]. We use ResNet [5] as the bottom-up part for forwarding features from the image to the top-down part. The top-down part applies layer-by-layer upsampling and lateral connection operations to the feature maps to obtain multi-scale features from $P_2$ to $P_5$. $P_2$ features are the stronger of these features for locational representation and the weaker for semantic representation in the image, while $P_5$ behaves oppositely. The feature extraction module can effectively utilize character location and semantic information from input images to provide multi-scale features for the instance segmentation and landmark estimation modules.

### 3.2 Instance Segmentation via Unified SPA Features

In instance segmentation module, the multi-scale features $P_i$, $i \in [2, 5]$ are first distributed to classification, kernel, and feature branches. A dynamic convolutional operation is then introduced to produce masks. Lastly, we employ a matrix non-maximum suppression (Matrix NMS) [25] to produce instances of the incomplete cigarette code.

In classification branch, each feature $P_i$ is aligned to the cell $s_i \times s_i$, where $s_i$ is the number of x-axis and y-axis for input image, and obtain $F_i^C \in \mathbb{R}^{s_i \times s_i \times d}$, where $d$ is the channel of the input image. After four convolutional operations, the features $C_i$ are aligned into the category feature shape $s_i \times s_i \times n$, where $n$ is the category number. Finally, the features $C_i$ are concatenated from different scales into feature $C^F \in \mathbb{R}^{s^2 \times n}$, where $s$ is the summation of $s_i$.

In kernel branch, we introduce the coordinate convolution [13] into feature $P_i$ to improve the mask spatial representation. After similar operations in the classification branch, the mask kernel $K^F \in \mathbb{R}^{s^2 \times e}$ is predicted, where $e$ is the channel number of the mask kernel.

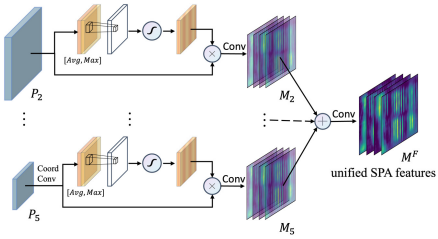


**Fig. 3.** Unified SPA features. Spatial attention mechanism is introduced to yield unified SPA features and strengthen the spatial representation for characters and backgrounds.

As shown in Fig. 3, we employ a spatial attention mechanism in the mask feature branch to strengthen the representation of cigarette code spatial features, and unify the multi-scale SPA features to generate the final mask features. The SPA features can be calculated as follows:

$$F_i^{spa} = \sigma(f(AvgPool(F_i^M) \odot MaxPool(F_i^M))), \tag{1}$$

where $\sigma$ denotes the sigmoid function, $f(\cdot)$ is the convolutional operation, and $\odot$ denotes concatenatal operation. The input feature $F_i^M$ is equal to $P_i$ if $i < 5$. Since the feature $P_5$ lacks the location information, the coordinate convolution is added to the input feature $F_5^M$. The unified SPA features $M^F$ is calculated as

$$M^F = Conv(\sum_{i=2}^{5} M_i) = Conv(\sum_{i=2}^{5}(f(F_i^{spa} \otimes F_i^M)), \tag{2}$$

where $\otimes$ denotes element-wise multiplication; $f(\cdot)$ is series upsample and convolutional operation to align the mask features shapes from multi-scales; $M_i$ denotes mask feature result of $i^{th}$ scale; and $Conv$ is a convolutional operation. Equation 1 uses spatial attention to obtain SPA features from $F_i^M$. Equation 2 first employs an element-wise multiplication operation between $F_i^M$ and $F_i^{spa}$ to enhance the representation ability of character location features. Then the $f(\cdot)$ upsamples the SPA features until its shape matches that of $P_2$. Additionally, an element-wise summation for each feature $M_i$ is performed to obtain the unified SPA features $M^+$. Finally, a $Conv$ operation is used to ascend dimension for $M^+$ to obtain the unified SPA mask features $M_F$ for the landmark estimation.

When the kernel feature $K^F$ and unified SPA mask feature $M^F$ are calculated, a dynamic convolutional operation is introduced to produce the instance masks $M$. For each kernel feature $K_z^F$ where $z \in [1, s^2]$, the instance mask $M_k$ can be calculated as $M_k = K_z^F \circledast M^F$, where $\circledast$ denotes dynamic convolution.

For category feature $C^F$ in the classification branch, we first filter out low confidence results using a 0.1 threshold and a 0.5 threshold to obtain $C'$ and $M$ respectively. Finally, Matrix NMS [25] is employed to obtain the final instance segmentation result from $C'$ and $M$.

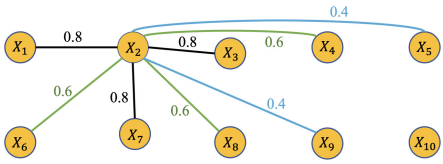


**Fig. 4.** Cigarette characters correlation. The black edges indicate the node's closest edge, whose weights are defined as 0.8; analogously, the weights of the green and blue edges are set to 0.6 and 0.4 respectively. If the distance between two nodes is $d \leq 3$, the value of correlation between two nodes is set to $0.2 \times (5 - d)$, otherwise is set to 0.(Color figure online)

### 3.3 Landmark Estimation via Graph Space Constraints

In landmark estimation module we first employ integral regression to estimate the initial landmarks of characters based on the unified SPA features, then we introduce GCN to construct graph-based space constraints that indicate the spatial correlations between individual characters.

In the integral regression process, firstly a convolutional operation is used to calculate the heatmap features from the unified SPA features $M^+$. Then, integral regression is performed using a discretization algorithm to generate the initial landmarks as follows:

$$X_i^{init} = \int_{p \in \Omega} (p \cdot \widetilde{F}_i^H(p)) = \sum_{x=1}^{h} \sum_{y=1}^{w} (p \cdot \widetilde{F}_i^H(p)), \tag{3}$$

where $\widetilde{F}_i^H(p)$ is the normalized heatmap feature; $p$ is the heatmap pixel point; $\Omega$ is the heatmap pixel region of width $w$ and height $h$; and $X_i^{init}$ denotes the landmark coordinates at channel $i$ after the discretization algorithm. As a result, the initial landmarks $X^{init} \in \mathbb{R}^{2\times 32}$ have been calculated.

The next step is to optimize the landmarks for the graph space constraints, by introducing GCN to constrain spatial semantic correlations between the 32 characters. Firstly, we align the shape of the FPN high-level feature $P_5$ as $F^{sem} \in \mathbb{R}^{h'w'\times 32}$ to 32 dimensions, corresponding to 32 characters. Secondly, the feature sequence $X$ of characters is calculated as $X = X^{init} \odot F^{sem}$, where $\odot$ denotes a concatenation operation. Finally, GCN is introduced to optimize landmarks, with a GCN operation shown as follows:

$$X' = f(X, A) = ReLU(AXW), \tag{4}$$

where $X'$ denotes the optimized landmarks through GCN operation $f(\cdot)$; $X$ denotes input feature sequences; $A$ is the adjacency matrix; $W$ denotes the GCN weights; and $ReLU$ is the activation function. The GCN weights $W$ of GCN are updated after network training is completed. The adjacency matrix $A$ is defined as shown in Fig. 4.

### 3.4 Text Regularization

The text regularization section uses the Hungarian matching algorithm to match the incomplete cigarette code recognition results with the estimated landmarks, and use '*' to fill the broken characters.

We define the mask centroids of character instances as a set $X^{mask}$ and the estimated landmarks as a set $X^{gcn}$. Then our matching task is transformed into an assignment problem [20] for each $i \in X^{mask}$, $j \in X^{gcn}$. The minimization assignment between $X^{mask}$ and $X^{gcn}$ is expressed as:

$$R = \arg \min_{R} (\sum_{i=1}^{n} \sum_{j=1}^{n} (d_{ij} \cdot r_{ij})), i, j \in (1, 2, \cdots, n), \tag{5}$$

where $R$ denotes minimization matching matrix; $d_{ij}$ is the Euclidean distance between $X_i^{mask}$ and $X_j^{gcn}$; $r_{ij} \in 0, 1$ indicates the matching relationship between $X_i^{mask}$ and $X_j^{gcn}$.

We introduce the Hungarian algorithm to calculate the minimization matching matrix $R$. Then, we use the matrix $R$ to match outcomes from the instance segmentation and landmark estimation modules to produce the recognition result for the incomplete cigarette code. Lastly, we fill missing character locations with '*' to regularize the 32-character cigarette code recognition result.

### 3.5 Loss Function

In the classification branch, we introduce the Focal loss [12] to calculate the classification loss $L_{cate}$. In the mask kernel and mask feature branches, we employ Dice loss [19] to calculate the loss $L_{mask}$ for the predicted mask area.

In the landmark initialization and optimization tasks, we introduce the loss function $\ell_1$ loss [21] to calculate the initial landmarks loss $L_{init}$ and optimized landmarks loss $L_{opt}$.

Ultimately, the overall loss function for our end-to-end recognition network is defined as:

$$L = L_{cate} + \lambda_0 L_{mask} + \lambda_1 L_{init} + \lambda_2 L_{opt}, \tag{6}$$

where $\lambda_0$, $\lambda_1$, $\lambda_2$ denote weights of each loss function. The optimal network performance is achieved for $(\lambda_0, \lambda_1, \lambda_2)$ equal to $(3, 0.1, 1)$.

## 4 Experiments

In this section, we extensively evaluate the proposed recognition network for incomplete cigarette code on a dataset, containing 15,000 images of incomplete cigarette code, including missing characters, complex background, and blurred printing. Firstly, we evaluate the instance segmentation module; secondly, we evaluate the landmark estimation module; we then discuss the performance of the end-to-end network in recognizing cigarette code; in addition, we demonstrate the effectiveness of the core components through ablation experiments.

**Table 1.** Instance segmentation mask AP (%) on test set of incomplete cigarette code.

| Method | AP | $AP_{50}$ | $AP_{75}$ | $AP_M$ | $AP_L$ |
|---|---|---|---|---|---|
| Mask R-CNN [4] | 52.7 | 84.6 | 59.6 | 48.5 | 54.0 |
| SOLOv2 [25] | 53.7 | 83.9 | 59.2 | 48.6 | 54.6 |
| **Ours** | **54.5** | **87.1** | **62.9** | **49.0** | **55.7** |

### 4.1 Evaluation of Instance Segmentation

For evaluating the instance segmentation performance, we compared the Mask R-CNN [4], the SOLOv2 [25] and our method. In the instance segmentation module, we evaluated the performance metrics AP, $AP_{50}$, $AP_{75}$, $AP_M$, and $AP_L$ for mask segmentation of Mask R-CNN, SOLOv2, and ours.

Table 1 shows that our instance segmentation module achieved an AP of 54.5%, which is a 1.8% and 0.8% improvement over Mask R-CNN and SOLOv2, respectively. These results demonstrate that our instance segmentation module has better performance for incomplete cigarette code.

### 4.2 Performance of Landmark Estimation

Our landmark estimation module introduces GCN to constrain characters in graph space and establish a spatial semantic correlation, which can effectively estimate the landmarks of incomplete cigarette code.

As shown in Fig. 5, we compared the landmark estimation performance between Wu et al. [27] and ours for multiple situations including: complex background on the first row; damaged characters on the second row; and skewed alignment on all rows. It is shown that our method obtained centered and smooth character locations.

### 4.3 End-to-End Performance

To evaluate the performance of our method for incomplete cigarette code, we evaluate the accuracy based on the test set. The accuracy was determined by the percentage of predicted results that correctly classified all 32 aligned characters. Our method achieves an accuracy rate of **90.2%**. Ten results of the recognition process for incomplete cigarette code are shown in Fig. 6. The results show that our network has good recognition performance for incomplete cigarette code.

As other text detection methods were incapable of recognition task for incomplete cigarette code, we used 500 images with complete cigarette code to fairly compare our method to other text recognition methods. We then evaluated each text recognition method in terms of time consumption and accuracy. The results are shown in Table 2, where time consumption is indicated by time taken to recognize one image in seconds, and accuracy by the percentage of predicted results that correctly classify 32 aligned characters. In summary, our method yields the optimal performance in both time consumption and recognition accuracy.

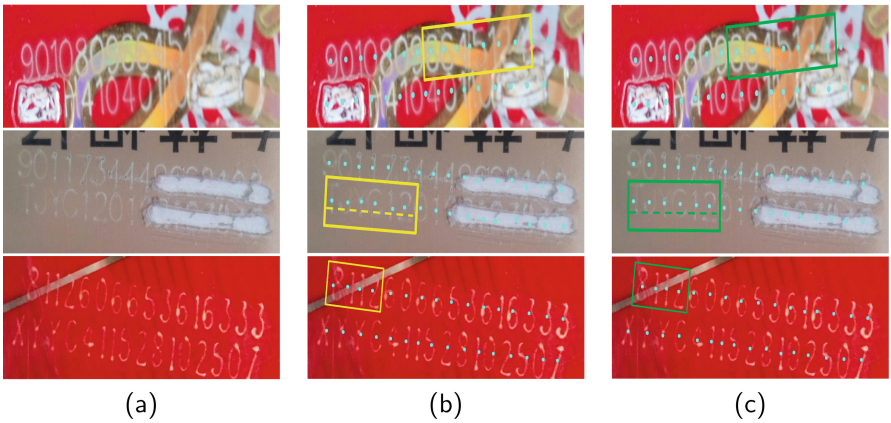


**Fig. 5.** Comparison of landmark estimation performance. (a) cigarette code image; (b) Wu et al. [27]; (c) ours. Locations are shown by yellow boxes for Wu et al. and green boxes for our method. Compared to Wu et al.'s method in the case of complex background, damaged characters, and skewed alignment, our method obtains centered and smooth character locations

**Table 2.** Comparison of recognition performance for complete cigarette code between our method and several state-of-the-art methods.

| Method | Time/s | Accuracy/% |
|---|---|---|
| Fots [15] | 0.564 | 69.6 |
| Mask textspotter [9] | 0.588 | 73.4 |
| ABCNet [16] | 0.216 | 71.3 |
| Xie et al. [29] | 0.636 | 87.4 |
| Wu et al. [27] | 0.324 | 90.6 |
| **Ours** | **0.192** | **92.6** |

### 4.4 Ablation Experiments

We present ablation experiments on the two key components, spatial attention mechanism and graph space constraints, to investigate the effectiveness of them.

To investigate the effectiveness of the SPA, we divided the experimental results into two cases by controlling whether to add SPA to the network, and calculated the network accuracy separately. The model achieves an accuracy of 90.2% with SPA component, while it only achieves 88.5% without SPA. The SPA component improved the character recognition accuracy by 1.7%, which proves that SPA can effectively strengthen the character representation ability.

To investigate the effectiveness of the GCN, we employed mean absolute error (MAE) and root mean square error (RMSE) to evaluate landmark estimation, and the results are shown in Table 3. The MAE and RMSE from GCN-based landmark estimation decreased by 3.978% and 2.546% respectively over

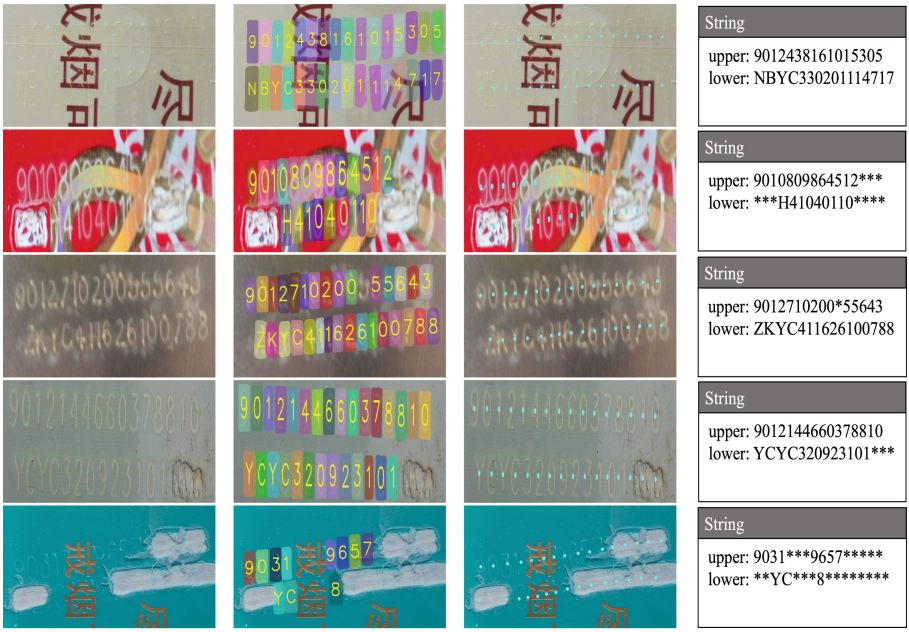


**Fig. 6.** Recognized results of our network. (a) cigarette code image; (b) character instances yielded from instance segmentation module; (c) character landmarks produced from landmark estimation module; (d) recognition result. Results are divided into 5 categories from top to bottom: similar color, complex background, blurred printing, damaged characters, and large broken area.

**Table 3.** Effectiveness of GCN.

| Component | MAE | RMSE |
|---|---|---|
| Without GCN | 0.2212 | 0.1807 |
| **With GCN** | **0.2124** | **0.1761** |

the method without GCN, indicating that GCN contributes to constructing character correlation.

## 5 Conclusion

In this work, we propose an end-to-end recognition network for incomplete cigarette code. The network reduces time consumption and improves recognition accuracy. Specifically, we utilize a spatial attention mechanism to yield unified SPA features, which can strengthen character and background representations, and improve the recognition accuracy; we construct graph space constraints through GCN to achieve high-accuracy landmark estimation, which can

establish spatial semantic correlation between characters and estimate the landmarks effectively for incomplete cigarette code.

**Acknowledgments.** This work was supported by the Shanghai Natural Science Foundation of China No. 19ZR1419100.

## References

1. Chen, X., Jin, L., Zhu, Y., Luo, C., Wang, T.: Text recognition in the wild: a survey. ACM Comput. Surv. **54**(2), 42:1–42:35 (2021)
2. Chen, Y., Wang, Z., Peng, Y., Zhang, Z., Yu, G., Sun, J.: Cascaded pyramid network for multi-person pose estimation. In: The IEEE Conference on Computer Vision and Pattern Recognition, pp. 7103–7112 (2018)
3. Doosti, B., Naha, S., Mirbagheri, M., Crandall, D.J.: Hope-net: a graph-based model for hand-object pose estimation. In: The IEEE Conference on Computer Vision and Pattern Recognition, pp. 6607–6616 (2020)
4. He, K., Gkioxari, G., Dollár, P., Girshick, R.B.: Mask R-CNN. In: IEEE International Conference on Computer Vision, pp. 2980–2988 (2017)
5. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: The IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
6. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. In: International Conference on Learning Representations, pp. 1–14 (2017)
7. Kuhn, H.W.: The Hungarian method for the assignment problem. In: Jünger, M., et al. (eds.) 50 Years of Integer Programming 1958-2008, pp. 29–47. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-540-68279-0_2
8. Li, H., Wang, P., Shen, C., Zhang, G.: Show, attend and read: a simple and strong baseline for irregular text recognition. In: The AAAI Conference on Artificial Intelligence, pp. 8610–8617 (2019)
9. Liao, M., Lyu, P., He, M., Yao, C., Wu, W., Bai, X.: Mask TextSpotter: an end-to-end trainable neural network for spotting text with arbitrary shapes. IEEE Trans. Pattern Anal. Mach. Intell. **43**(2), 532–548 (2021)
10. Lin, Q., Luo, C., Jin, L., Lai, S.: STAN: a sequential transformation attention-based network for scene text recognition. Pattern Recogn. **111**, 107692 (2021)
11. Lin, T., Dollár, P., Girshick, R.B., He, K., Hariharan, B., Belongie, S.J.: Feature pyramid networks for object detection. In: The IEEE Conference on Computer Vision and Pattern Recognition, pp. 936–944 (2017)
12. Lin, T., Goyal, P., Girshick, R.B., He, K., Dollár, P.: Focal loss for dense object detection. In: IEEE International Conference on Computer Vision, pp. 2999–3007 (2017)
13. Liu, R., et al.: An intriguing failing of convolutional neural networks and the CoordConv solution. In: Conference on Neural Information Processing Systems, pp. 9628–9639 (2018)
14. Liu, W., Chen, C., Wong, K.Y.K.: Char-net: a character-aware neural network for distorted scene text recognition. In: The AAAI Conference on Artificial Intelligence, pp. 7154–7161 (2018)
15. Liu, X., Liang, D., Yan, S., Chen, D., Qiao, Y., Yan, J.: FOTS: fast oriented text spotting with a unified network. In: The IEEE Conference on Computer Vision and Pattern Recognition, pp. 5676–5685 (2018)

16. Liu, Y., Chen, H., Shen, C., He, T., Jin, L., Wang, L.: ABCNet: real-time scene text spotting with adaptive Bezier-curve network. In: The IEEE Conference on Computer Vision and Pattern Recognition, pp. 9806–9815 (2020)
17. Long, S., He, X., Yao, C.: Scene text detection and recognition: the deep learning era. Int. J. Comput. Vis. **129**(1), 161–184 (2021)
18. Luo, C., Jin, L., Sun, Z.: MORAN: a multi-object rectified attention network for scene text recognition. Pattern Recogn. **90**, 109–118 (2019)
19. Milletari, F., Navab, N., Ahmadi, S.: V-net: fully convolutional neural networks for volumetric medical image segmentation. In: Fourth International Conference on 3D Vision, pp. 565–571 (2016)
20. Neumann, L., Matas, J.: A method for text localization and recognition in real-world images. In: Kimmel, R., Klette, R., Sugimoto, A. (eds.) ACCV 2010. LNCS, vol. 6494, pp. 770–783. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-19318-7_60
21. Shalev-Shwartz, S., Tewari, A.: Stochastic methods for $l_1$-regularized loss minimization. J. Mach. Learn. Res. **12**, 1865–1892 (2011)
22. Sun, X., Xiao, B., Wei, F., Liang, S., Wei, Y.: Integral human pose regression. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11210, pp. 536–553. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01231-1_33
23. Tompson, J., Goroshin, R., Jain, A., LeCun, Y., Bregler, C.: Efficient object localization using convolutional networks. In: The IEEE Conference on Computer Vision and Pattern Recognition, pp. 648–656 (2015)
24. Wang, X., Kong, T., Shen, C., Jiang, Y., Li, L.: SOLO: segmenting objects by locations. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12363, pp. 649–665. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58523-5_38
25. Wang, X., Zhang, R., Kong, T., Li, L., Shen, C.: SOLOv2: dynamic and fast instance segmentation. In: Conference on Neural Information Processing Systems, pp. 1–12 (2020)
26. Wei, S., Ramakrishna, V., Kanade, T., Sheikh, Y.: Convolutional pose machines. In: The IEEE Conference on Computer Vision and Pattern Recognition, pp. 4724–4732 (2016)
27. Wu, P., Zhou, Z., Huang, J., Xie, Z., Sheng, B.: Multi-scale feature fusion for incomplete cigarette code recognition. J. Comput.-Aided Des. Comput. Graph. **33**(5), 780–788 (2021)
28. Xie, Z., Wu, J., Zhang, S., Tang, Z., Fan, J., Ma, L.: Intelligent recognition method for cigarette code based on deep neural networks. J. Comput.-Aided Des. Comput. Graph. **31**(1), 111–117 (2019)
29. Xie, Z.-F., Zhang, S.-H., Wu, P.: CNN-based erratic cigarette code recognition. In: Zhao, Y., Barnes, N., Chen, B., Westermann, R., Kong, X., Lin, C. (eds.) ICIG 2019. LNCS, vol. 11901, pp. 245–255. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-34120-6_20
30. Xie, Z., Zhang, W., Sheng, B., Li, P., Chen, C.L.P.: BaGFN: broad attentive graph fusion network for high-order feature interactions. IEEE Trans. Neural Netw. Learn. Syst. Early Access 1–15 (2021)
31. Xin, M., Mo, S., Lin, Y.: EVA-GCN: head pose estimation based on graph convolutional networks. In: The IEEE Conference on Computer Vision and Pattern Recognition, pp. 1462–1471 (2021)