# CNN-Based Erratic Cigarette Code Recognition

Zhi-Feng Xie[1,2(✉)], Shu-Han Zhang[1(✉)], and Peng Wu[1(✉)]

[1] Department of Film and Television Engineering, Shanghai University, Shanghai 200072, China
zhifeng_xie@shu.edu.cn, 1993jerryzhang@163.com, 1097829202@qq.com
[2] Shanghai Engineering Research Center of Motion Picture Special Effects, Shanghai 200072, China

**Abstract.** Cigarette code is a string printed on the wrapper of cigarette packet as a basis of distinguishing illegal sales for tobacco administration. In general, the code is excerpted and entered to administration system manually during on-site inspection, which is quite time-consuming and laborious. In this paper, we propose a new solution based on convolutional neural network for intelligent transcription. Our recognition method is composed of four components: detection, identification, alignment, and regularization. First of all, the detection component fine-tunes an end-to-end detection network to obtain the bounding box region of cigarette code. Then the identification component constructs an optimized CNN architecture to recognize each character in the region of cigarette code. Meanwhile the alignment component trains a CPM-based network to estimate the positions of all characters including some missing characters. Finally, the regularization component develops a matching algorithm to produce a regularized result with all characters. The experimental results demonstrate that our proposed method can perform a better, faster and more labor-saving cigarette code transcription process.

**Keywords:** Cigarette code · Optical Character Recognition · Convolutional neural network

## 1 Introduction

Cigarette code is a string with 32 characters printed on the wrapper of cigarette packet, which can be used to distinguish illegal cigarette sales in China, and example is shown in Fig. 1. Presently, the code is excerpted and entered to administration system manually during on-site inspection. This manual recording method is quite time-consuming and laborious. Thus it is urgent that an intelligent method can be proposed to simplify manual operations and improve inspection efficiency.

OCR (Optical Character Recognition) is an ordinary method to recognize text from an image. However, our task to recognize cigarette code faces several


Y. Zhao et al. (Eds.): ICIG 2019, LNCS 11901, pp. 245–255, 2019.
https://doi.org/10.1007/978-3-030-34120-6_20

difficulties with classical OCR methods [9,19,29] or modern CNN-based OCR methods [1,2,5,7,8,13,14,22,24] : (a) Erratic layouts. The cigarette codes are printed by different administrations, their layouts, font types, and font sizes are miscellaneous. (b) Complicated backgrounds. Normally, cigarette code is printed on the its wrapper randomly, which may cause cigarette code contaminated by its background. (c) Geometric deformation. Due to imperfect printing technique, characters are often printed with distortion. (d) Man-made sabotage. In order to evade punishment, some retail stores with illegal sale would make sabotage to cigarette code that results in some indiscernible characters. (e) Semantic demands. Even if some characters are unable to recognize, the tobacco monopoly administration demands these characters to be regularized by $'*'$ character.

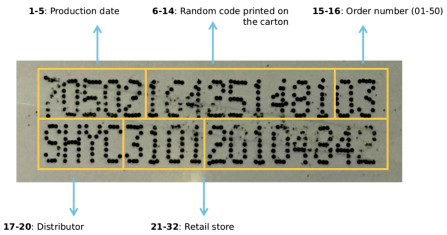


**Fig. 1.** Information of cigarette code. This carton of cigarettes is manufactured in May 2, 2017, it is the third one in a large box, and distributed from Shanghai Tobacco Monopoly Administration to a retail store with $'310120104842'$ identifier.

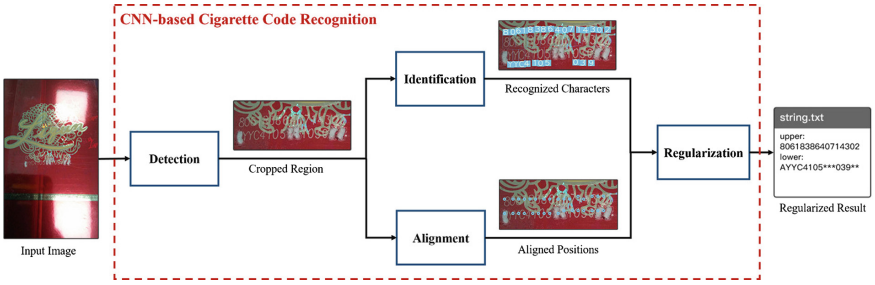


**Fig. 2.** Overview. Our CNN-based recognition method consists of four main components: detection, identification, alignment, and regularization.

To deal with these issues, we proposed a new CNN-based solution to achieve the efficient recording of erratic cigarette code in a single image. As shown in Fig. 2, our pipeline proceeds in four components: detection, identification, alignment, and regularization. Given an image with erratic cigarette code, the detection component first employs transfer training technique to fine-tune an

end-to-end detection network [23], which can classify the categories of cigarette code (black and white) and obtain the bounding box region of cigarette code. Then the identification component constructs an optimized CNN architecture by strengthening feature extraction and defining multi-parallel region proposal networks, which can recognize and locate each character in the cropped region of cigarette code. At the same time, the alignment component trains a CPM-based (Convolutional Pose Machine) network [31] to estimate the positions of all 32 characters in the cropped region, including some missing characters. Finally, the regularization component develops a matching algorithm to set up the mapping relationship between the identification and alignment results, and fill some $'*'$ characters to produce a regularized result with all 32 characters. The experimental results show that our proposed method can yield the detection accuracy of over 98%, and the recognition accuracy (all of 32 characters are correct in an image) of over 90% in the testing dataset.

## 2 Related Work

Text recognition is always a key task to recognize text content accurately in the process of OCR. Recently, a lot of state-of-the-art methods have been proposed to achieve high-accuracy text recognition. Wang et al. [29] proposed a system rooted in generic object recognition to achieve superior performance of text recognition. Neumann et al. [19] further proposed an ERs-based (Extremal Regions) real-time scene text recognition. Jaderberg et al. [9] present the text recognition problem as multi-class classification task with large number of labels. Since then, a lot of CNN-based methods are introduced into the term, which can be mainly divided into three categories: LSTM (Long Short-Term Memory) + CTC (Connectionist Temporal Classification) [2,7,13,16], LSTM + Attention [1,3–6,8,14,15,17,24] and object detection [21,22,30].

The perfect recognition of erratic cigarette code not only depends on identification accuracy for each character, but also need to fill some missing characters and produce a regularized recognition result with all 32 characters by estimating all positions of missing characters. Thus a excellent localization technique is very important for the regularization of erratic cigarette code. At present, many alignment algorithms have been applied successfully to locate key points, especially in human face and human pose. For face alignment, a number of CNN-based methods [25], such as TCDCN [35], TCNN [32], MTCNN [34], DAN [11], and so on, can produce the key points of face with partial occlusion efficiently and accurately. For pose estimation, many state-of-the-art methods [10,26,27,31] can also construct CNN-based models to yield a great performance even under body occlusion.

## 3 CNN-Based Recognition for Erratic Cigarette Code

Since erratic cigarette code is more complex and distinctive than traditional text, a number of state-of-the-art OCR techniques fail to produce satisfying

recognition results. Here, we propose a new CNN-based solution to recognize erratic cigarette code effectively. As shown in Fig. 2, its pipeline consists of four key components: detection, identification, alignment, and regularization.

### 3.1 Detection

In the detection component, we employ a end-to-end convolutional neural network to obtain a bounding box region of cigarette code and point out its category (black code or white code). Our end-to-end network concatenates three sub-networks: feature extraction, region proposal, regression and classification.

Our detection component refers to the concept of inductive transfer learning [20] for network training. We first collect tens of thousands of images with cigarette code, and their categories and bounding box coordinates are manually annotated. Then we construct the end-to-end detection network and achieve the fine-tuned training based on the VGG-16, RPN, and RCNN architectures. Finally, we apply the trained model and the NMS (Non-Maximum Suppression) algorithm to predict the bounding box and category of cigarette code accurately.

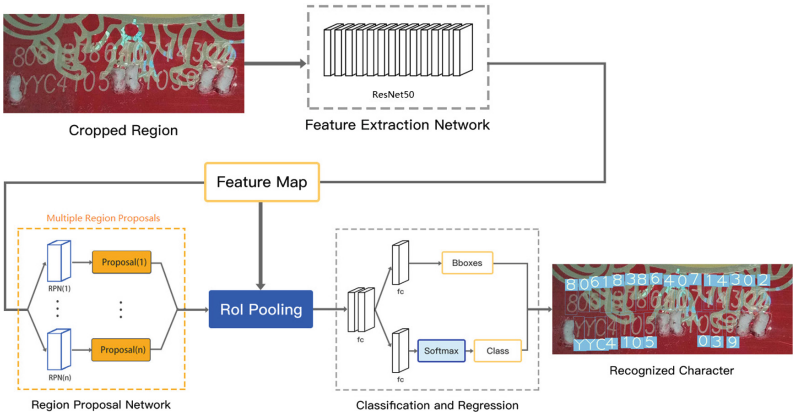


**Fig. 3.** Network of identification component. The identification network consists of three sub-networks: ResNet-based feature extraction, multi-parallel region proposal, classification and regression.

### 3.2 Identification

In the identification component, we construct a new convolutional neural network to identify and locate each character in the cropped region of cigarette code. As shown in Fig. 3, the identification network consists of three sub-networks: ResNet-based feature extraction, multi-parallel region proposal, classification and regression.

Inspired by the concept of CNN ensemble [33], we introduce extra anchors, with $0.25, 0.333, 0.5, 1, 2$ in ratio and $4, 6, 8, 12, 16$ in scale, yielding $5 \times 5 = 25$ anchors. The optimized architecture defines more small and tall anchors for the character shapes of cigarette code.

As shown in Fig. 4, the characters in two cropped cigarette codes and their bounding boxes are correctly marked by our identification component. However, since some characters in the white example are destructed deliberately, their positions cannot be specified and the identification result is also incomplete. Thus we must further estimate the positions of missing characters, introduce a special character $'*'$ to fill them and produce the recognized result with all 32 characters.

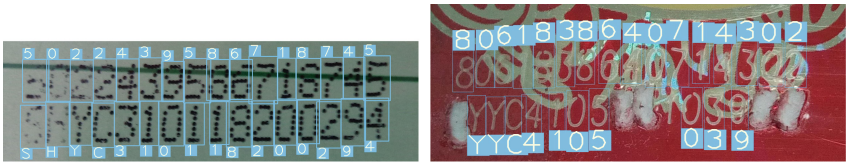


**Fig. 4.** Two examples of cigarette code identification. (a) Recognized characters of black cigarette code. (b) Recognized characters of white cigarette code.

### 3.3 Alignment

In the alignment component, we integrate the popular DeepPose's [28] concept to localize all characters of erratic cigarette code, especially including some missing characters. We can apply the fine-tuned training to optimize the network stage by stage, and then yield a predicted model to produce a final alignment result.

First of all, we annotate the bounding boxes of 32 characters in the cigarette code region, $(x_i^1, y_i^1, x_i^2, y_i^2), i \in [1, 32]$, where $(x^1, y^1)$ and $(x^2, y^2)$ are the top left and bottom right points of each bounding box. Then, we compute their center points as the ground-truth positions $Z = (z_1, \ldots, z_{32})$, where $z_i = \{(x_i^1 + x_i^2)/2, (y_i^1 + y_i^2)/2\}, i \in [1, 32]$, and create the ideal belief map $b_*^p(z)$ for each position $z$ by putting Gaussian peaks at ground truth locations of the $p$-th position. Next we construct the CPM-based alignment network and define the cost function of each stage $t \in [1, 6]$:

$$f_t = \sum_{p=1}^{32} \sum_{z \in Z} \| b_t^p(z) - b_*^p(z) \|_2^2 \tag{1}$$

where $b_t^p$ denotes the belief map of the $p$-th estimated position by stage $t$, $b_*^p$ denotes the ideal belief map of the $p$-th ground-truth position. Finally, we add the losses at each stage $F = \sum f_t$, and use standard stochastic gradient descend to jointly train all stages in the network. As shown in Fig. 5, the alignment network can effectively estimate character positions even with characters indiscernible.

### 3.4 Regularization

In the regularization component, we further propose a matching algorithm to set up a corresponding relationship between the identification and alignment results, and then employ a special character $'*'$ to fill some missing characters and produce a regularized result with all 32 characters.

First of all, we obtain the locations of identification by computing the central points of bounding boxes $Y^{rb}$ in the identification result. Then we denote the locations of identification as $Y^r$ and the positions of alignment as $Y^a$. The mathematical model for our matching task can be defined as a typical assignment problem [18]. Based on this, we introduce Hungarian algorithm [12] to minimize $\Phi$ and calculate the mapping matrix $\mathbf{X}$ in order to match the identification locations $Y^r$ whose elements are $\leq$32 with the estimated 32 alignment positions $Y^a$.

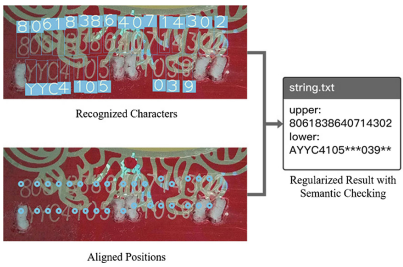


**Fig. 5.** Regularization results.

With the output of mapping matrix $\mathbf{X}$, we can assign the identification result $Y_i^r$ into the $j$-th position of the output string for each $(x_{ij} \neq 0) \in \mathbf{X}$. Since the length of output string is fixed to 32, we need to fill the rest positions in output string with $'*'$ if $Y^r$ has less than 32 characters. If there exists a unique matching between the output string and the dictionary element, we can replace these corresponding $'*'$s with certain characters from the semantic dictionary. As shown in Fig. 5, by our matching algorithm, we can yield the output string with 32 characters "8061838640714302" and "*YYC4105***039**".

## 4 Data Preparation and Network Training

To train our networks, we need to prepare a lot of images with annotations. We first collect 21,861 images including the black and white cigarette codes for the detection network. The annotation of each input image contains the bounding box of cigarette code and its category $C_r = \{black, white\}$. Then we train the detection network to predict the bounding box and category of cigarette code, which can be used to produce the cropped region image of cigarette code. We collect these cropped images to construct the identification dataset, including

21,740 images of the black cigarette code and 17,044 images of the white code. The annotation of each cropped region image image contains the bounding box of each character and its corresponding category $C_c = \{0\text{–}9, \text{A–Z}, *\}$, where $'*'$ denotes the indiscernible character. We randomly pick 20,000, 20,000, 16,000 samples for training and make the rest 1,861, 1,740, 1,044 samples for testing.

With all the data preparation done, we can start our network training process. For the detection network, we set some necessary training parameters, including the total iterations of 40,000 with step learning rate at $\lambda = 0.001$, gamma $= 0.1$, stepsize $= 50,000$, momentum set at 0.9, weight decay at $5e-4$. The training parameters of the identification network is similar with the detection network except a total iterations of 100,000. For the alignment network, we integrate all identification datasets as the alignment dataset, and compute the central points of their bounding boxes as the ground-truth positions. We set its necessary training parameters, including the total iterations of 62,500 with step learning rate at $\lambda = 4e-6$, gamma $= 0.333$, stepsize $= 13,275$, momentum set at 0.9, weight decay at $5e-4$.

## 5 Experimental Results

In this section, we perform a lot of experiments to evaluate the detection, identification, and alignment components one by one. On the other hand, we also demonstrate the excellent end-to-end performance of our proposed method.

### 5.1 Evaluation of Detection Component

We evaluate the detection component with accuracy and region correctness. The accuracy of category classification is near 99.6%, and 8 images with cigarette code aren't classified correctly.

For region detection, we expand the bounding box of detected region by 6.25% horizontally and 3.125% vertically, and define its correctness if the annotated bounding box $B_a$ is fully included by the expanded bounding box region $B_r$, that is $B_a \subset B_r$, it is different with traditional definition. The detection component achieves the 98.6% accuracy of region detection in total testing dataset, including 98.8% and 98.3% in two testing datasets of black and white codes respectively.

### 5.2 Evaluation of Identification Component

We employ the detection component to produce the cropped region of cigarette code as identification dataset.

To perform a fair comparison, all the state-of-the-art methods are trained and tested in our same identification dataset, and the definition of identification correctness is that all characters excluding $'*'$ in cigarette code region must be correctly recognized. As shown in Table 1, we observe that most of state-of-the-art methods can achieve the good identification of simple black code but

**Table 1.** Accuracy of identification component.

| | Correct results (black) | Accuracy (black) | Correct result (white) | Accuracy (white) |
|---|---|---|---|---|
| Deep text spotter [2] | 1,489/1,740 | 85.6% | 792/1,044 | 75.9% |
| Attention OCR [5] | 1,522/1,740 | 87.5% | 851/1,044 | 81.5% |
| SEE [1] | 1,445/1,740 | 83.0% | 718/1,044 | 68.8% |
| TextSpotter [8] | 1,501/1,740 | 86.3% | 830/1,044 | 79.5% |
| Aster [24] | 1,529/1,740 | 87.9% | 857/1,044 | 82.1% |
| Our identification component | 1,578/1,740 | **90.7%** | 901/1,044 | **86.3%** |

be difficult to handle the complex white code. In contrast, our identification component can reach the higher identification accuracy on both two testing datasets.

### 5.3 Evaluation of Alignment Component

The definition of alignment correctness is that the estimated character position locates inside our beforehand artificially annotated bounding box with a 10% expansion both horizontally and vertically. Our alignment component achieves 92.7% accuracy in the testing dataset, including 94.7% accuracy on black code and 89.4% accuracy on white code respectively.

### 5.4 End-to-End Performance

To evaluate the overall performance of our proposed method, we execute an end-to-end verification in the detection testing dataset. We define the principle of correctness transcription as follow: all of the characters are recognized correctly, with $'*'$ character labeling unrecognized characters, and all the characters must be in right order. Our proposed method achieves 92.2% accuracy in total, 95.8% on black code and 87.2% on white one respectively.

During on-site cigarette inspection, we randomly pick 500 cartons of cigarettes and make comparison with artificial transcription of cigarette code. Our transcription system achieves 90.8% accuracy of recognition, which is slightly lower than 95.2% accuracy of artificial transcription. But our system only takes 43 min to finish the whole process, which is higher-efficiency and more labour-saving than 382 min of artificial transcription. The comparison result also demonstrates that our transcription system can further simplify manual operations and improve inspection efficiency.

## 6 Conclusion

In this paper, we mainly propose a new solution to detect and recognize erratic cigarette code accurately and rapidly. Although some existing techniques are

applied into our solution, we still put forward some new ideas and make some important contributions. First of all, we collect more than 40 thousands images of cigarette code and annotate their bounding boxes and character information. Compared with a number of traditional OCR datasets, our cigarette code dataset is more complex and distinctive, such as erratic layouts, various fonts, complicated backgrounds, geometric deformation, man-made sabotage, and so on. It is a new challenge to solve these issues by existing state-of-the-art OCR techniques. In the future, we will share our cigarette code dataset with all researchers. Secondly, our new solution not only integrates the existing models but also further optimizes them to improve the recognition accuracy of erratic cigarette code. On one hand, we construct multi-parallel RPN units to strengthen the effect of region proposal and avoid missing some characters with different shapes. On the other hand, we propose a novel regularization method by training CPM-based network and developing an optimal string matching algorithm. The experimental results have demonstrated the effectiveness of our new improvement. Finally, with a view to the practical application, we employ our new solution to implement an intelligent transcription system of cigarette code.

Although our proposed method can achieve a higher-efficiency recognition for erratic cigarette code, its recognition accuracy is still slightly lower than artificial transcription. Therefore, we must first extend the training dataset of cigarette code, and further optimize the network architecture of our model to solve some bottleneck problems, such as rotation, various character shapes, alignment accuracy with many missing characters, and so on.

## References

1. Bartz, C., Yang, H., Meinel, C.: SEE: towards semi-supervised end-to-end scene text recognition. arXiv preprint arXiv:1712.05404 (2017)
2. Busta, M., Neumann, L., Matas, J.: Deep TextSpotter: an end-to-end trainable scene text localization and recognition framework. In: IEEE International Conference on Computer Vision, pp. 2223–2231 (2017)
3. Cheng, Z., Xu, Y., Bai, F., Niu, Y., Pu, S., Zhou, S.: AON: towards arbitrarily-oriented text recognition. In: 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, 18–22 June 2018, pp. 5571–5579 (2018). https://doi.org/10.1109/CVPR.2018.00584
4. Deng, Y., Kanervisto, A., Ling, J., Rush, A.M.: Image-to-markup generation with coarse-to-fine attention. arXiv preprint arXiv:1609.04938 (2016)
5. Deng, Y., Kanervisto, A., Rush, A.M.: What you get is what you see: a visual markup decompiler (2016)
6. Ghosh, S.K., Valveny, E., Bagdanov, A.D.: Visual attention models for scene text recognition. arXiv preprint arXiv:1706.01487 (2017)
7. He, P., Huang, W., Qiao, Y., Chen, C.L., Tang, X.: Reading scene text in deep convolutional sequences, vol. 116, no. 1, pp. 3501–3508 (2015)
8. He, T., Tian, Z., Huang, W., Shen, C., Qiao, Y., Sun, C.: An end-to-end textspotter with explicit alignment and attention. In: 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, 18–22 June 2018, pp. 5020–5029 (2018). https://doi.org/10.1109/CVPR.2018.00527

9. Jaderberg, M., Simonyan, K., Vedaldi, A., Zisserman, A.: Synthetic data and artificial neural networks for natural scene text recognition. Eprint Arxiv (2014)
10. Jain, A., Tompson, J., LeCun, Y., Bregler, C.: MoDeep: a deep learning framework using motion features for human pose estimation. In: Cremers, D., Reid, I., Saito, H., Yang, M.-H. (eds.) ACCV 2014. LNCS, vol. 9004, pp. 302–315. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-16808-1_21
11. Kowalski, M., Naruniec, J., Trzcinski, T.: Deep alignment network: a convolutional neural network for robust face alignment. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 2034–2043 (2017)
12. Kuhn, H.W.: The hungarian method for the assignment problem. Nav. Res. Logist. Q. **2**(1–2), 83–97 (1955)
13. Li, W., Cao, L., Zhao, D., Cui, X.: CRNN: integrating classification rules into neural network. In: International Joint Conference on Neural Networks, pp. 1–8 (2013)
14. Liu, W., Chen, C., Wong, K., Su, Z., Han, J.: STAR-NET: a spatial attention residue network for scene text recognition (2016)
15. Liu, W., Chen, C., Wong, K.K.: Char-Net: a character-aware neural network for distorted scene text recognition. In: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI 2018), The 30th Innovative Applications of Artificial Intelligence (IAAI 2018), and The 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI 2018), New Orleans, Louisiana, USA, 2–7 February 2018, pp. 7154–7161 (2018)
16. Liu, Y., Wang, Z., Jin, H., Wassell, I.: Synthetically supervised feature learning for scene text recognition. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11209, pp. 449–465. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01228-1_27
17. Liu, Z., Li, Y., Ren, F., Goh, W.L., Yu, H.: SqueezedText: a real-time scene text recognition by binary convolutional encoder-decoder network. In: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, AAAI 2018, The 30th Innovative Applications of Artificial Intelligence (IAAI 2018), and The 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI 2018), New Orleans, Louisiana, USA, 2–7 February 2018, pp. 7194–7201 (2018)
18. Mulmuley, K., Vazirani, U.V., Vazirani, V.V.: Matching is as easy as matrix inversion. Combinatorica **7**(1), 105–113 (1987). https://doi.org/10.1007/BF02579206
19. Neumann, L., Matas, J.: Real-time scene text localization and recognition. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3538–3545. IEEE (2012)
20. Pan, S.J., Yang, Q., et al.: A survey on transfer learning. IEEE Trans. Knowl. Data Eng. **22**(10), 1345–1359 (2010)
21. Prasad, S., Kong, A.W.K.: Using object information for spotting text. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11220, pp. 559–576. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01270-0_33
22. Redmon, J., Farhadi, A.: YOLOv3: an incremental improvement. CoRR arXiv:abs/1804.02767 (2018)
23. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems, pp. 91–99 (2015)
24. Shi, B., Yang, M., Wang, X., Lyu, P., Yao, C., Bai, X.: ASTER: an attentional scene text recognizer with flexible rectification. IEEE Trans. Pattern Anal. Mach. Intell. 1 (2018). https://doi.org/10.1109/TPAMI.2018.2848939

25. Sun, Y., Wang, X., Tang, X.: Deep convolutional network cascade for facial point detection. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 3476–3483 (2013)
26. Tompson, J., Goroshin, R., Jain, A., Lecun, Y., Bregler, C.: Efficient object localization using convolutional networks, pp. 648–656 (2014)
27. Tompson, J., Jain, A., Lecun, Y., Bregler, C.: Joint training of a convolutional network and a graphical model for human pose estimation. Eprint Arxiv, pp. 1799–1807 (2014)
28. Toshev, A., Szegedy, C.: DeepPose: human pose estimation via deep neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1653–1660 (2014)
29. Wang, K., Babenko, B., Belongie, S.: End-to-end scene text recognition. In: 2011 IEEE International Conference on Computer Vision (ICCV), pp. 1457–1464. IEEE (2011)
30. Wang, T., Wu, D.J., Coates, A., Ng, A.Y.: End-to-end text recognition with convolutional neural networks. In: 2012 21st International Conference on Pattern Recognition (ICPR), pp. 3304–3308. IEEE (2012)
31. Wei, S.E., Ramakrishna, V., Kanade, T., Sheikh, Y.: Convolutional pose machines. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016
32. Wu, Y., Hassner, T., Kim, K., Medioni, G., Natarajan, P.: Facial landmark detection with tweaked convolutional neural networks. IEEE Trans. Pattern Anal. Mach. Intell. **PP**(99), 1 (2015)
33. Xu, Y., et al.: End-to-end subtitle detection and recognition for videos in East Asian languages via CNN ensemble. Sig. Process. Image Commun. **60**, 131–143 (2018)
34. Zhang, K., Zhang, Z., Li, Z., Qiao, Y.: Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Sig. Process. Lett. **23**(10), 1499–1503 (2016)
35. Zhang, Z., Luo, P., Loy, C.C., Tang, X.: Facial landmark detection by deep multi-task learning. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8694, pp. 94–108. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10599-4_7