Research on garment flat multi-component recognition based on Mask R-CNN

DOI: 10.35530/IT.074.01.202199

TAO LI
YE-XIN LYU
LING MA

YONG XIE
FENG-YUAN ZOU

ABSTRACT – REZUMAT

Research on garment flat multi-component recognition based on Mask R-CNN

The automatic recognition of garment flat information has been widely researched through computer vision. However, the unapparent visual feature and low recognition accuracy pose serious challenges to the application. Herein, inspired by multi-object instance segmentation, the method of mask region convolutional neural network (Mask R-CNN) for garment flat multi-component is proposed in this paper. The steps include feature enhancement, attribute annotation, feature extraction, and bounding box regression and recognition. First, the Laplacian was employed to enhance the image feature, and the Polygon annotated component attributes to reduce the interaction interference. Next, the ResNet was applied to realize identity mapping to characterize redundant information of components. Finally, the feature map was entered into two branches to achieve bounding box regression and recognition. The results demonstrated that the proposed method could realize multi-component recognition effectively. Compared with the unenhanced feature, the mAP increased by 2.27%, reaching 97.87%, and the average F_1 was 0.958. Compared to VGGNet and MobileNet, the ResNet backbone used for Mask R-CNN could improve the mAP by 11.55%. Mask R-CNN was more robust than the state-of-the-art methods and more suitable for garment flat multi-component recognition.

Keywords: *Mask R-CNN, garment flat, feature enhancement, multi-component network, component localization and recognition*

Cercetări privind recunoașterea multi-componentelor liniare ale articolelor de îmbrăcăminte bazate pe metoda Mask R-CNN

Recunoașterea automată a informațiilor despre îmbrăcăminte a fost cercetată pe scară largă prin tehnologia computerizată. Cu toate acestea, caracteristica vizuală neaparentă și acuratețea scăzută a recunoașterii reprezintă provocări serioase pentru aplicație. Aici, în această lucrare, inspirată de segmentarea instanțelor multi-obiect, este propusă metoda rețelei neuronale convoluționale regionale (Mask R-CNN) pentru multi-componentele liniare ale articolelor de îmbrăcăminte. Pașii includ îmbunătățirea caracteristicilor, adnotarea atributelor, extragerea caracteristicilor și regresia și recunoașterea spațiului de delimitare. În primul rând, operatorul Laplacian a fost utilizat pentru a îmbunătăți caracteristica imaginii, iar atributele componentelor adnotate Polygon au fost utilizate pentru a reduce interferența interacțiunii. Apoi, rețeaua ResNet a fost aplicată pentru a realiza maparea identității și pentru a caracteriza informații redundante ale componentelor. În cele din urmă, harta caracteristicilor a fost introdusă în două ramuri, pentru a obține regresia și recunoașterea spațiului de delimitare. Rezultatele au demonstrat că metoda propusă ar putea realiza în mod eficient recunoașterea multicomponentelor. Față de caracteristica neîmbunătățită, mAP a crescut cu 2,27%, ajungând la 97,87%, iar media F_1 a fost de 0,958. În comparație cu VGGNet și MobileNet, rețeaua backbone ResNet a fost utilizată pentru Mask R-CNN, care ar putea îmbunătăți mAP cu 11,55%. Mask R-CNN a fost mai robustă decât metodele de ultimă generație și mai potrivită pentru recunoașterea multicomponentelor liniare pentru articolele de îmbrăcăminte.

Cuvinte-cheie*: Mask R-CNN, îmbunătățirea caracteristicilor liniare ale articolelor de îmbrăcăminte, rețea multicomponentă, localizarea și recunoașterea componentelor*

INTRODUCTION

In the textile and garment industry, garment CAD technology intends to become increasingly versatile in garment flat, pattern making, grading, and layout modules [1]. However, the obtained garment flat information and pattern-related dimension acquisition still rely on visual inspection of trained patternmakers' experience [2]. It is highly subjective, inefficient and prone to misjudgment [3]. Currently, image recognition technology based on computer vision has sprung up to meet the needs of information exchange between garment flat and pattern-related dimensions [4].

Nowadays, the garment flat recognition framework is driven by two scenarios based on the attribute difference: single or multiple object recognition. Compared to single recognition, multiple object recognition could assign multiple attributes for each instance simultaneously [5]. Moreover, the multi-component learning framework is more effective by jointly inter-class and intra-class recognition tasks. They could boost each other to prevent overfitting and the mutual promotion to ensure the feature representations robust and discriminative [6]. Thus, a multi-component learning network was established to automatically recognize garment flat.

Nowadays, garment flat recognition methods mainly are divided into two types: mathematical model and machine learning approach [7]. Compared with other methods, deep learning could extract high-level semantic features, without requiring artificially designed features, and has advantages in image recognition, classification, and detection [8]. In terms of deep learning in multi-object recognition, the recognition methods could be divided into attribute combination [9], classifier combination [10], and deep learning architecture modification [11]. In related research, Guan et al. established a series of databases for different recognition attributes, and different learning models were selected [9]. However, this method requires prior knowledge and many different models are prone to random errors. Later, Donati et al. converted multi-object into multiple single-object combinations by combining deep learning, template matching, and other classifiers. Nonetheless, the recognition accuracy was only 73.8% [10]. The main reason is that the garment flat is and composed of curves without texture and colour features, resulting in inconspicuous features [12]. Compared to the classifier combination, the modified deep learning Hypotheses-CNN-Pooling was proposed. However, this method was based on the entire image, which leads to redundant calculations [11]. Since then, the recognition framework based on the garment component was proposed [13]. The component-based recognition studies will more conform to patternmakers' minds which can select pattern prototypes according to the component category. More importantly, the component has more discriminating features because of removing too many irrelevant factors in contrast to the integral garment. Zhou et al. applied a part-based deep neural network cascade mode to integrate different component-recognition sub-network into a cascade for human parsing. The parsing results verify the superiority of this framework [14]. Furthermore, Zhou et al. applied VGG16-CAM to component-based garment recognition. The components could be effectively recognized, but the recognition accuracy only reached 82.23% [15]. This is because simultaneous multi-component recognition is prone to interfere with each other, resulting in low recognition accuracy. Therefore, it is necessary to find a method to address the problem of the unapparent visual feature and low recognition accuracy in the multi-component recognition of garment flat.

In this work, inspired by multi-object instance segmentation, the method of Mask R-CNN was proposed. First, the Laplacian was used to enhance the image feature. The Polygon annotated component attributes reduce interaction interference. Second, the ResNet backbone was applied to extract the feature. Finally, it combined the region proposal network (RPN) and full connection layer (FCL) for joint multi-component detection and recognition. We demonstrated that the proposed method has better performance than an unenhanced feature, other backbone architectures, and recognition methods. It is of great practical significance to automatically obtain garment flat information and reduce the subjectivity of patternmakers.

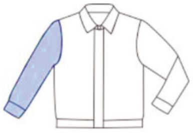
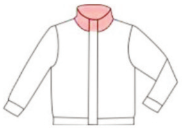
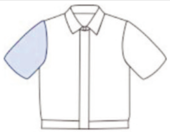
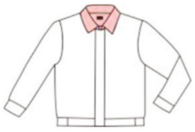
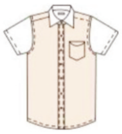
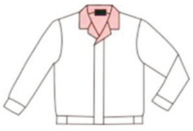
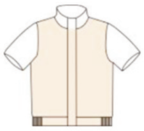
EXPERIMENTAL SECTION

Multi-attribute dataset annotation

Garment flats with collar, sleeve, and body components were selected for multi-component recognition. The dataset was established according to pattern prototype and component category, as shown in table 1.

Table 1

COMPONENT CATEGORY AND SCHEMATIC DIAGRAM

Component	Category	Schematic Diagram	Component	Category	Schematic Diagram
Sleeve	Long sleeve		Collar	Stand Collar	
	Short sleeve			Lapel	
Body	Shirt			Peak Lapel	
	Jacket		-	-	-

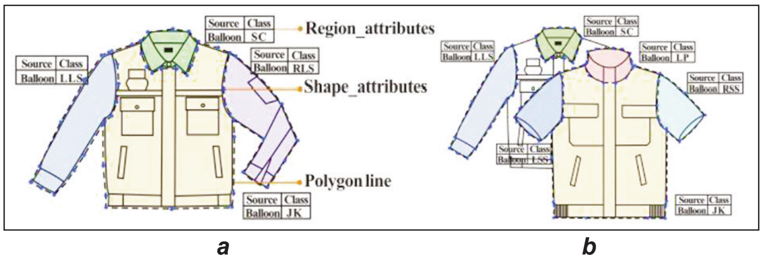


Fig. 1. Garment flat annotation: *a* – single annotation; *b* – multiple annotation

The 360 garment flats were kindly gifted from Zhejiang Lanting Garment Co., Ltd., China. The image resolution was 120 dpi, without other pre-processing. To meet the requirements of network training, VGG Image Annotation was applied to mask selection and attribute annotation. The attributes were annotated respectively (if the occlusion area exceeds 60%, it will not be annotated). Then the Polygon, instead of the Rectangle, annotated component attributes to reduce interaction interference (figure 1).

Multi-component recognition network establishment

Mask R-CNN, which integrates a full convolution network (FCN) and feature pyramid network (FPN) based on the Faster R-CNN was proposed [16]. It has low requirements on image quality and is more suitable for garment flat multi-component recognition. The results of feature extraction directly affect the recognition accuracy. Nowadays, the feature extraction networks, such as VGGNet, MobileNet, AlexNet, ResNet, etc. all have good recognition effects. Considering the particularity of the garment flat with less feature information, the ResNet-50 was applied as the backbone architecture to extract the feature. Compared to others, the problems of gradient explosion or dispersion caused by a deep network could be avoided by performing identity mapping on the redundant information of the garment flat. The residual block of ResNet is illustrated in figure 2 [17]. It transforms the fitting objective function into a residual function, which is more sensitive to small fluctuations. After feature extraction by ResNet, five feature layers (C2-C5) with different sizes and dimensions were fused in combination with FPN to generate the feature maps. Then the original garment flat had already obtained the highly abstract feature. Among them, C2 and C3 layers were used to extract shallow features to recognize the shape feature with obvious structures such as sleeves, collar, and body components. C4 and C5 layers were employed to recognize component subdivisions by extracting high-level semantic features.

Later, several regions of interest (RoIs) for each pixel position on the feature maps were set. The RoIs were sent into RPN for binary classification (foreground/ background) and bounding box regression to generate the refined RoIs. Then the refined RoIs were pooled into a fixed-size feature map through RoI Align to solve the misalignment problem caused by twice quantization processes. The goal is to ensure the pixels in the original image are completely aligned with the pixels in the feature maps. The backpropagation formula of RoI Align is illustrated in equation 1. The main purpose is to calculate the gradient between the output and the target value, and backpropagation to update the weight:

$$\text{the } \frac{\partial L}{\partial x_i} = \\ = \sum_r \sum_j \left[d(i, i(r,j)) < 1 \right] (1-\Delta h)(1-\Delta w) \frac{\partial L}{\partial y_{rj}} \quad (1)$$

where: *L* represents the function of RoI Align, x_i – the point on the feature map before pooling, y_{rj} – the *j*-th point in the *r*-th bin after pooling, *r* – the number of bins, *j* – the number of points in *r*-th bin, *i* – the point coordinate on the feature map, $i(r, j)$ – the floating-point coordinate after pooling, *d*(.) – the distance between pixel points, Δh, Δw – the horizontal and the vertical coordinate difference between *i* and $i(r, j)$.

Then the RoIs after RoI Align processed entered into two branches to achieve multi-component recognition. Among them, one branch realized component subdivision through RoI classification and bounding box regression. The other was the mask generation network composed of FCN, which generated masks consistent with the size and shape of garment components. Then the category classification, bounding box regression, and mask generation were realized (figure 3) [18].

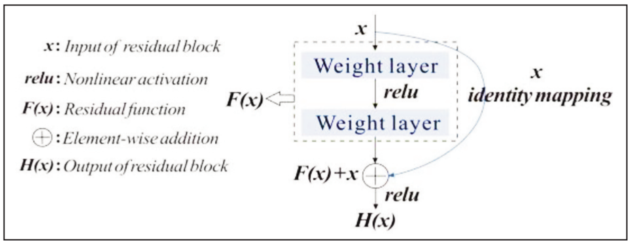


Fig. 2. Residual block of ResNet

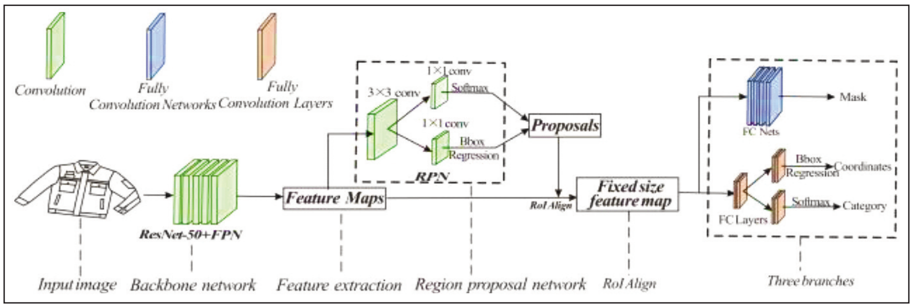


Fig. 3. Multi-component recognition framework

The network losses mainly contain class, bbox, and mask loss. The formula of the loss function is shown in equation 2. The classification and bounding box losses are identical to those defined in the work of Dai et al. [19]. The mask loss as the average binary cross-entropy loss is defined in the work of He et al. [16]. And the multi-target loss function was employed.

$$L = L_{cls} + L_{bbox} + L_{mask} \tag{2}$$

where: L_{cls} represents the class loss, L_{bbox} – the bbox loss and L_{mask} – the mask loss.

Multi-component recognition experiments

The experiments were implemented on a PC with AMD Ryzen 9 3900X CPU and AMD Radeon RX 6700XT GPU. The operation system was on Windows 10 with Tensorflow. The proposed method was performed on Python software.

According to the optimal standard of machine learning, the dataset was divided into training and testing set according to 8:2. Due to the small magnitude of the self-dataset, the experiments were pre-trained on the MS COCO 2014 to improve the generalization ability. The weights were transferred to the pre-training model by model similarity and transfer learning. The ResNet-50 was applied as the feature extraction backbone. The experiments were performed with a gradient of 25 epochs to determine the approximate training epochs. The trend was shown in figure 4. It was observed that mean average precision (mAP) tended to be stable after 100 epochs. In addition, the correlation between training time and epoch was analyzed. It could be observed that the epoch was significantly and positively correlated with the training time (R^2 = 0.97). Thus, considering the mAP and training time, the epoch was determined to be 100 and the number of images per iteration was 50. Since the garment components were relatively small relative to the entire garment and their size is uncertain, the scale of the RPN anchor was adjusted to better detect small targets. The main parameters of Mask R-CNN were shown in table 2.

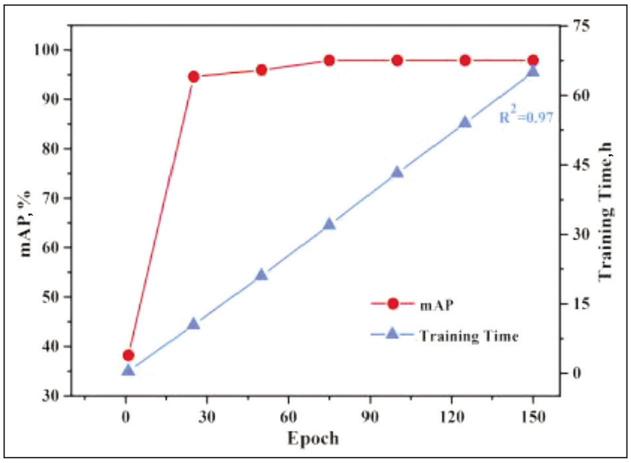


Fig. 4. mAP trend with epoch

RESULTS AND DISCUSSION

Evaluation metric selection

In target detection and recognition, average precision (AP) and mAP are often selected as evaluation metrics of the recognition network. Generally, the higher the value is, the better the network will be. According to the COCO dataset, the evaluation metrics are AP@50:5:95, $AP^{IoU=0.5}$, and $AP^{IoU=0.75}$. The IoU is the intersection over the union, as illustrated in equation 3:

$$IoU(i) = \frac{n_{ii}}{\sum_j (n_{ij} + n_{ji}) - n_{ii}} \tag{3}$$

where: n_{ii} represents the pixel numbers of the category i, which is predicted to be i, n_{ij} – the category j, which is predicted to be i. Otherwise, n_{ji} stands for the opposite.

For the evaluation of component recognition, accuracy and precision are used. Among them, the accuracy represents an evaluation of the overall correct recognition. And precision is the evaluation of a certain category. F_1 score is taken as the harmonic average, which takes into account the precision and recall, as shown in equation 4:

$$Accuracy = TP + TN / TP + FP + TN + FN$$
$$Precision = TP / (TP + FP),\ Recall = TP / (TP + FN)$$
$$F_1 = 2Precision * Recall / (Precision + Recall) \tag{4}$$

where: TP is the correct recognition, FP – the false, FN – the number of components.

Multi-component recognition results

The multi-component recognition network was trained on the self-constructed dataset. Figure 5, *a* is the network loss of the training set. All losses decreased rapidly in the iteration [0, 10] stage and tended to converge stably in the later stage. The total loss steadily converged to 0.25 after 100 iterations and the class, bbox, and mask losses all converged to 0.1. Figure 5, *b* is the losses of the testing set, which is similar to the training set. Although there had

Table 2

MAIN PARAMETERS OF MASK R-CNN

Configuration	Parameter	Configuration	Parameter
Backbone	ResNet-50	Detection_Min_Confidence	0.7
Backbone_Strides	[4, 8, 16, 32, 64]	Learning_Rate	0.001
RPN_Anchor_Scales	(32, 64, 128, 256, 512)	Learning_Momentum	0.9
RPN_NMS_Threshold	0.7	Weight_Decay	0.0001

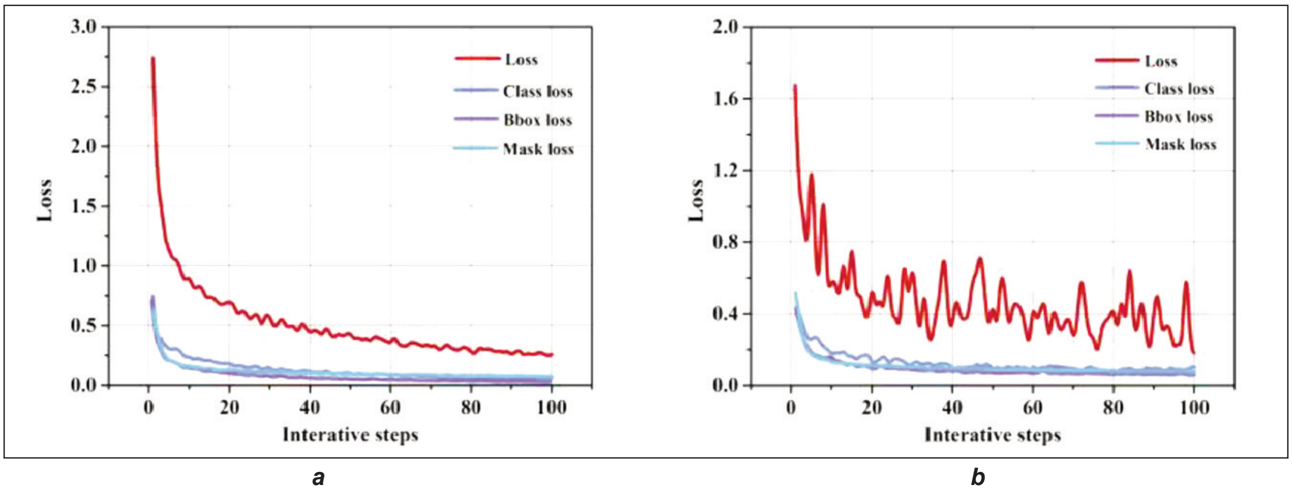


Fig. 5. Loss curve: *a* – training set; *b* – testing set

been a certain degree of small-range fluctuations in a total loss, the overall trend tended to converge. Thus, the network constructed in this paper was well-trained and had no over-fitting problem.

The trained multi-component recognition network was tested when the IoU = 0.5. The results demonstrated that the garment components could be recognized and localized. Among them, the rectangle represented the component position and the data was the mask quality. Different colour masks could accurately cover the component areas, and the bounding boxes could be positioned. It is applied to a garment flat containing one or more styles. The average mask quality score reached 99.2 % in garment flat containing one style, and in two or more styles was 98.1 % (figure 6). However, due to the insufficient features caused by mutual occlusion, some components appeared recognition omissions.

To further characterize the recognition effect, evaluation metrics were used. The mAP reached 97.87 %, and the average F_1 was 0.958 (table 3). The results showed that the Mask R-CNN based on ResNet-50 is effective. The garment components could be recognized at high precision. Among them, the F_1 of the sleeve and collar was higher than that of the body component. The reason was the shape of the sleeve and collar was more obvious than that of the body, and ResNet-50 could fully learn the component subdivision difference. The AP of the stand collar and

Table 3

EVALUATION METRICS OF MULTI-COMPONENT GARMENT FLAT

Category		Precision	Recall	F_1	Average F_1	AP/ %	mAP (%)
Sleeve	Long sleeve	0.944	1.000	0.971	0.964	98.4	97.87
	Short sleeve	0.957	0.957	0.957		98.83	
Body	Shirt	0.889	0.800	0.842	0.91	95.83	
	Jacket	0.955	1.000	0.977		98.14	
Collar	Stand collar	1.000	0.917	0.956	0.986	96.43	
	Lapel	1.000	1.000	1.000		98.74	
	Peak collar	1.000	1.000	1.000		98.73	

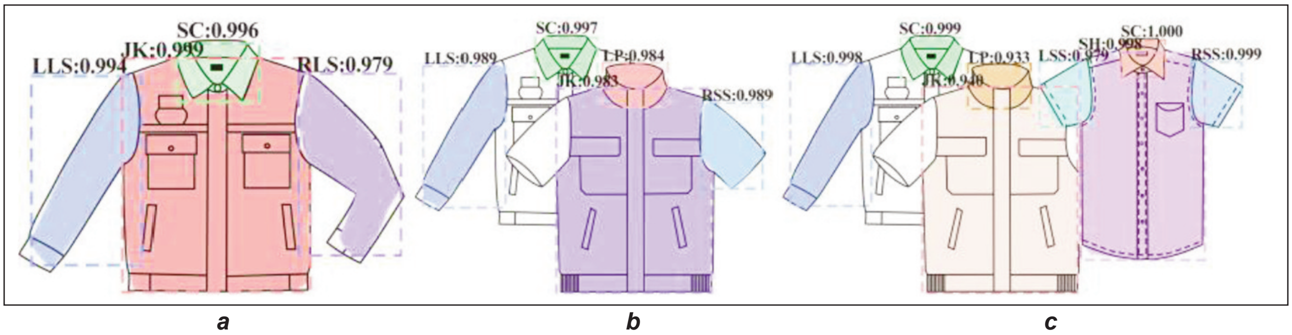


Fig. 6. Garment flat recognition results: *a* – single; *b* – double; *c* – multiple

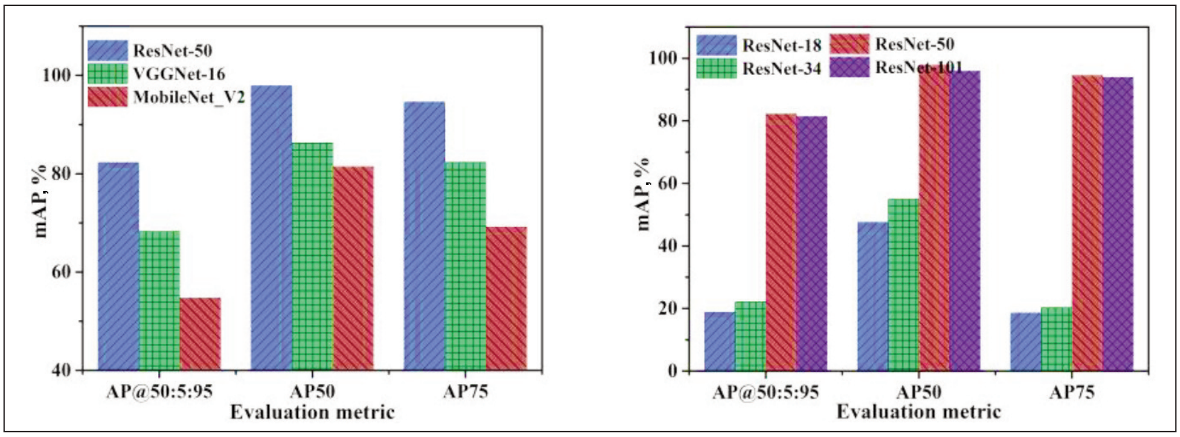
shirt was relatively small because the box localization and mask annotation could not completely coincide with the test set, resulting in a decrease.

Experimental comparative analysis

The specifications and proportions of the garment flat were not uniform. It was easily blurred, which directly affected the clarity of visual features. Edge sharpening was applied to enhance the shape feature to solve this problem. Currently, the main sharpening methods are first-order-based Sobel, Prewitt, and second-order-based Laplacian operators. Table 4 is the sharpening results.

It could be seen that Sobel and Prewitt operators formed pseudo edges. Among them, Prewitt expanded the edge silhouette and was almost distorted. And the internal details were still not obvious after Sobel sharpened. Compared with the aforementioned, the Laplacian operator not only enhanced the internal details but also heightened the edge feature. Therefore, the shape edge of the garment flat was enhanced by the Laplacian operator, and the definition of Laplacian is defined in [18]. Moreover, the multi-component experiments were carried out with enhanced features under the same network parameters. As we can see that the enhanced feature has a more distinguishable shape feature. It could effectively recognize the components, and the mAP improved by 2.27 % after feature enhancement (table 5).

To verify the ResNet-50 is more suitable in garment flat multi-component recognition, the ResNet-18, ResNet-34, ResNet-101 and other backbones such as VGGNet-16 and MobileNet_V2 were selected (figure 7). It could be seen that the mAP of ResNet-50



a *b*

Fig. 7. Comparison of backbones: *a* – other backbones; *b* – ResNet backbones

Table 4

RESULTS OF FEATURE ENHANCEMENT

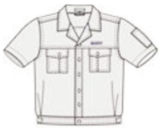
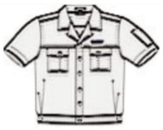
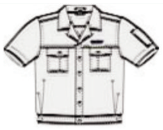
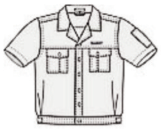
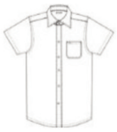
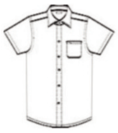
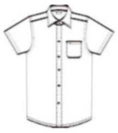
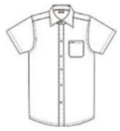
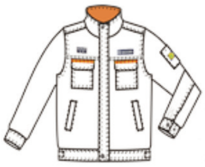
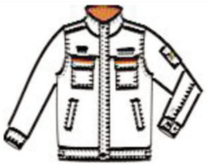
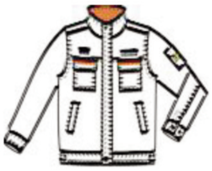
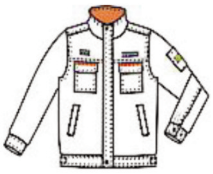
Sharpen operator	Origin	Sobel	Prewitt	Laplacian
Garment flat 1				
Garment flat 2				
Garment flat 3				

Table 5

COMPARATIVE EXPERIMENTS

Recognition network	Mask R-CNN	Mask R-CNN with enhanced feature
mAP (%)	95.6	97.87

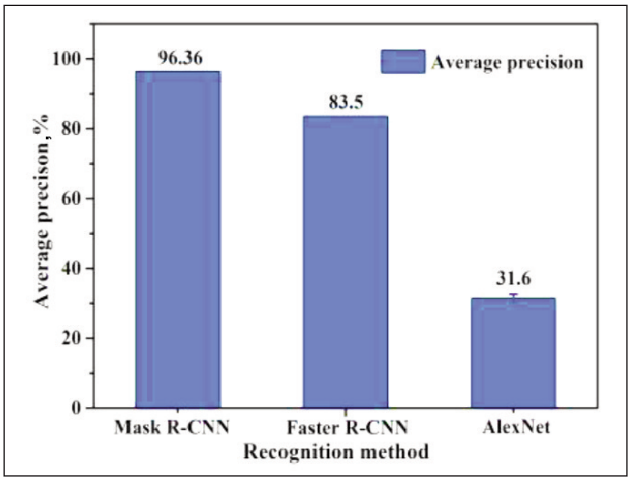


Fig. 8. Comparison of recognition methods

was improved by more than 11.55 % than other backbones. It is because the ResNet replaces the original target by fitting residual mapping, which makes the network more sensitive to small fluctuations [17]. On the other hand, ResNet-50 has a deeper network layer than others. The high-level convolutional kernel could cover larger-scale features and combine more low-level features. Compared to ResNet-101, the mAP of ResNet-50 only increased by 1.97%, but training time was reduced by more than double. The main reason is that the ResNet-50 already has a good feature extraction ability to extract the local gradient and edge shape features. However, ResNet-101 weakens the low-level feature while extracting more high-semantic information due to the depth network. And the weight needs to be updated layer by layer, resulting in a significant increase in training time. The core convolution layers 3 × 3 were adopted in ResNet-18 and ResNet-34 as a basic block, and network depth is relatively shallow that some components such as body components in garment flat could not be recognized effectively. Therefore, the ResNet-50 is more suitable as a feature extractor.

To verify the advantages of Mask R-CNN in garment flat recognition, the Faster R-CNN and AlexNet were selected (figure 8). The recognition precision of Mask R-CNN was significantly the highest. This is because of the pixel deviation of Faster R-CNN during the pooling process. The RoI Align is used in Mask R-CNN which is combined with the FCN's accurate pixel Mask to achieve higher precision. Moreover, the lowest precision of AlexNet is mainly due to feature interference. Other components weakened the feature, resulting in the precision being only 31.6%.

CONCLUSIONS

In this paper, the garment flat multi-component recognition method based on Mask R-CNN was proposed. The main conclusions were drawn as follows:

1. The proposed method could realize the recognition and localization of garment flat components effectively. The network was robust and had low requirements for image quality.
2. The feature enhanced by Laplacian could effectively improve the mAP by 2.27%, reaching 97.87%, and the average F_1 was 0.958.
3. According to the experimental comparison, Mask R-CNN based on ResNet-50 had higher recognition accuracy than that of other backbones and recognition methods.

Acknowledgements

This study is financially supported by the National Natural Science Foundation of China (NO. 11671009) and the Key Laboratory of Silk Culture Heritage and Products Design Digital Technology (NO. 2020WLB09).

REFERENCES

[1] Mok, T., Xu, J., Wang, X.X., Fan, J.T., Kwok, Y.L., Xin, J.H., *An IGA-based design support system for realistic and practical fashion designs*, In: Computer-Aided Design, 2013, 45, 1442–1458

[2] Hong, Y., Bruniaux, P., Zheng, J.J., Liu, K.X., Dong, M., Chen, Y., *Application of 3D-to-2D garment design for atypical morphology: a design case for physically disabled people with scoliosis,* In: Industria Textila, 2018, 69, 1, 59–64, http://doi.org/10.35530/IT.069.01.1377

[3] Yu, L.J., Wang, R.W., Zhou, I.F., *A novel approach for identification of pills based on the method of Depth from Focus*. In: Industria Textila, 2018, 69, 6, 466–471, http://doi.org/10.35530/IT.069.06.1271

[4] Hong, Y., Zeng, X.Y., Bruniaux, P., Chen, Y., Zhang, X.J., *Development of a new knowledge-based fabric recommendation system by integrating the collaborative design process and multi-criteria decision support*, In: Textile Research Journal, 2018, 88, 23, 1–17

[5] Liu, W.W., Wang, H.B., Shen, X.B., Tsang, I.W., *The emerging trends of multi-label learning*, In: IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 14, 8, 1–22

[6] Zhao, J.J., Peng, Y.X., He, X.T., *Attribute hierarchy based multi-task learning for fine-grained image classification*, In: Neurocomputing, 2020, 395, 150–159

[7] Li, T., Du, L., Huang, Z.H., Jiang, Y.P., Zou, F.Y., *Review on pattern conversion technology based on garment flat recognition*, In: Journal of Textile Research, 2020, 41, 8, 145–151

[8] Lecun, Y., Bengio, Y., Hinton, G., *Deep learning*, In: Science, 2015, 521, 7553, 436–444

[9] Guan, C.Y., Qin, S.F., Long, Y., *Apparel-based deep learning system design for apparel style recommendation*, In: International Journal of Clothing Science and Technology, 2019, 31, 3, 376–389

[10] Donati, L., Iotti, E., Mordonini, G.L., Prati, A., *Fashion product classification through deep learning and computer vision*, In: Applied Sciences, 2019, 9, 7, 1385

[11] Wei, Y.C., Xia, W., Lin, M., Huang, J.S., Ni, B.B., Dong, J., Zhao, Y., *HCP: A flexible CNN framework for multi-label image classification,* In: IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38, 9, 1901–1907

[12] Liu, K.X., Zeng, X.Y., Wang, J.P., Tao, X.Y., Xu, J., Jiang, X.W., Ren, J., Kamalha, E., Agrawal, T.K., Bruniaux, P., *Parametric design of garment flat based on body dimension*, In: International Journal of Industrial Ergonomics, 2018, 65, 46–59

[13] Hidayati, S.C., You, C.W., Cheng, W.H., Hua, K.L., *Learning and recognition of clothing genres from full-body images*, In: IEEE Transactions on Cybernetics, 2018, 48, 5, 1647–1659

[14] Zhou, Y.H., Mok, T., Zhou, S., *A part-based deep neural network cascade model for human parsing*, In: IEEE Access, 2019, 7, 160101–160111

[15] Zhou, L.P., Zhou, Z.Z., Zhang, L.Q., *Deep part-based image feature for clothing retrieval*, In: International Conference on Neural Information Processing, 2017, 10636, 340–347

[16] He, K.M., Gkioxari, G., Dollar, P., Girshick, R., *Mask R-CNN*, In: Proceedings of the IEEE International Conference on Computer Vision, 2017, 2961–2969

[17] He, K.M., Zhang, X.Y., Ren, S.Q., Sun, J., *Deep residual learning for image recognition*, In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, 770–778

[18] Yu, Y., Zhang, K.L., Yang, L., Zhang, D.X., *Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN*, In: Computers and Electronics in Agriculture, 2019, 163, 104846

[19] Dai, J.F., Li, Y., He, K.M., Sun, J., *R-FCN: Object detection via region-based fully convolutional networks*, In: NIPS, 2016, 1–11

[20] Satish, S.B., *Efficient medical image enhancement technique using transform HSV space and adaptive histogram equalization*, In: Soft Computing Based Medical Image Analysis, 2018, 51–60

Authors:

TAO LI[1], YE-XIN LYU[1], LING MA[1], YOUNG XIE[2], FENG-YUAN ZOU[1,3,4]

[1]School of Fashion Design and Engineering, Zhejiang Sci-Tech University, Hangzhou 310018, China

[2]School of Art and Design, Wuyi University, Jiangmen, 529020, China

[3]Key Laboratory of Silk Culture Heritage and Products Design Digital Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China

[4]Zhejiang Provincial Research Center of Clothing Engineering Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China

Corresponding author:

FENG-YUAN ZOU
e-mail: zfy166@zstu.edu.cn