CBL: A CLOTHING BRAND LOGO DATASET AND A NEW METHOD FOR CLOTHING BRAND RECOGNITION

Kuan-Hsien Liu[1], Tsung-Jung Liu[2], Fei Wang[2]
[1]*National Taichung University of Science and Technology, Taichung, Taiwan*
[2]*National Chung Hsing University, Taichung, Taiwan*
khliu@nutc.edu.tw, tjliu@dragon.nchu.edu.tw, z7785715@gmail.com

***Abstract*—In this work, we presented a novel clothing brand logo prediction method which is rooted on a dense-block based deep convolutional neural network for brand logo detection and recognition. To learn convolutional neural networks deeper and more accurately, we adopted dense blocks into deep convolutional networks to make connections between layers shorter. In our work, we propose several dense-block structure designs to improve detection and recognition accuracy on clothing brand logos. We also built a new large-scale clothing brand logo (CBL) dataset with the brand attribute and logo information to facilitate this task. To reduce complexity for the proposed framework, two pixel search steps for the bounding movement are implemented in the training procedure. In the experiment, we demonstrate our search reduced model can outperform some state-of-the-art methods and achieve very good results.**

***Index Terms*—clothing brand logo dataset, detection and recognition, prediction, dense block, convolutional neural network**

I. Introduction

Recently, clothing classification [1] [2], clothing attribute prediction [3] [4] and clothing retrieval [5] [6] have drawn much attention in computer vision and pattern recognition fields. However, there is no related work on the analysis of clothing brand, which motivates us to study this subject. There are many factors, such as brand, price, style, color, material, and pattern, could influence people to choose their clothes. Among these factors, brand is a very important one and clothing brand prediction is also a practical and very challenging task. Fig. 1 shows some clothing images for 4 brands: H&M, Adidas, SuperDry, and ASOS.

A lot of approaches based on deep convolutional neural networks (DCNNs) have shown significant breakthroughs in the identification of clothing types, clothing retrieval [7], object detection and recognition, age estimation [8]–[14], image inpainting [15], image super-resolution [16], quality of experience [17] and visual quality assessment [18]–[21]. Since YOLOv3 [22] has achieved promising results in object detection, we design a new DCNN model by modifying YOLOv3 with incorporation of dense blocks [23] for clothing brand prediction.

Currently, there is no existing clothing dataset containing brand information. To the best of our knowledge, we are the first one to construct a large-scale clothing dataset with brand information.

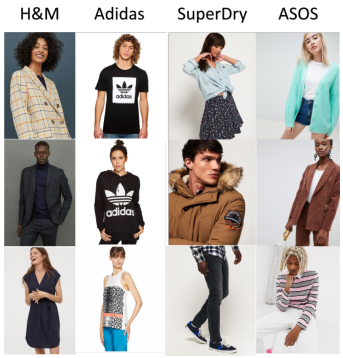

Fig. 1. Some sample images for 4 clothing brands.

Our experiments show that the dense-block YOLOv3 we proposed can achieve higher brand prediction accuracy than original YOLOv3 on our constructed clothing brand dataset. We also show that our proposed method outperforms several state-of-the-art methods for clothing brand prediction on this new clothing dataset.

II. Related Work

Here we first give a brief review on some existing clothing datasets. Then, some recent proposed DCNN models are also visited. Finally, we review some previous methods on clothing retrieval and classification.

A. Clothing datasets

In the past ten years, there are many methods proposed for the outfit recommendation and clothing retrieval. Due to the great need in this field, several clothing datasets were constructed. Some datasets are briefly introduced below.

Clothing Attribute (CA) dataset is proposed by Chen et al. [4]. CA includes 1,856 images and 26 attributes. They analyzed semantics to generate a list of attributes to achieve the purpose of clothing retrieval.

Apparel Classification with Style (ACS) [2] dataset introduced a complete pipeline for identifying and classifying

TABLE I
COMPARISON ON CLOTHING DATASETS

Dataset	No. of images	Type	Color	Material	Pair	Brand	Price	Lowest resolution	Highest resolution
ACS [2]	89,484	Y	N	N	N	N	N	132x116	244x192
CA [4]	1,856	Y	Y	N	N	N	N	750x742	864x1296
CF [24]	2,628	Y	Y	Y	N	N	N	400x600	400x600
DARN [25]	341,021	Y	Y	Y	Y	N	N	-	-
DDAN [26]	182,780	Y	Y	N	N	N	N	-	-
MVC [27]	161,260	Y	Y	Y	Y	N	N	1920x2240	1920x2240
DF [28]	800k up	Y	Y	Y	Y	N	N	349x512	750x1101
CB [Ours]	1,000k up	Y	Y	Y	Y	Y	Y	820x1000	1900x2375
CBL [Ours]	57,000	Y	Y	Y	Y	Y	Y	820x1000	1900x2375

people's clothes in natural scenes. ACS can be used as a benchmark dataset for the clothing classification task, which contains more than 80,000 images with 15 classes.

A new fashion image dataset called Colorful-Fashion (CF) was constructed by Liu et al. [24] in which all 2,682 images were labeled with pixel-level color category labels. They solved the problem of automatically analyzing the fashion images.

DARN [25] solves cross-domain image retrieval which is consisted of two subnets, one for each domain, whose retrieval feature representation is driven by semantic attribute learning. This attribute-led learning is a key factor in improving search accuracy.

Multi-View Clothing (MVC) [27] dataset contains 161,260 annotated images of 1920×2240 resolutions with 264 attribute labels. MVC is also the first built dataset to consider multiple viewing angle on clothing retrieval. It has four different views for each clothing item and can be used for view-invariant clothing retrieval application.

DeepFashion (DF) [28] dataset consists of three sub-datasets and has more than 800,000 images totally. In this dataset, each image contains more than one thousand labels, including pair information of clothing items and clothing landmarks. A powerful image retrieval method is reached by learning the locations of the landmarks and predicting a large number of attributes.

The comparison for clothing dataset is also summarized in Table I.

B. Some recent DCNN models

Classification is one of the most widely used approach for different applications in many fields, such as computer vision, image processing, and pattern recognition. Many DCNN based models have been proposed to deal with classification problems.

Alexnet [29], a very early and famous architecture consists of eight layers, including five convolutional layers and some of them are followed by max-pooling layers, and two fully connected layers. VGG16 [30] investigate the relationship between the convolutional network depth on its accuracy in the large-scale image recognition setting. He et al. [31] proposed ResNet50 which utilizes residual learning framework to ease network training. ResNet50 is substantially deeper than other network models used previously and gets outstanding performance in many applications.

DenseNet [23] was designed based on the idea that each layer is connected to each other layer in a feed-forward manner. For each layer, the feature maps from all previous layers are used as input, and its own feature map is also used as input for all subsequent layers.

The Residual-Dense Block (RDB) [32] combines the residual block and the dense block, and would contain both the advantage from residual block and the advantage from dense block. This RDB based CNN method would have a significant performance improvement over the previous DCNN methods.

C. Clothing retrieval/classification methods

Lin et al. [26] proposed a hierarchical deep search framework to achieve a rapid clothing retrieval. In order to make this model faster, they added a latent layer into the network and have neurons in this layer to learn hashes-like representations while fine-tuning it on the clothing dataset.

Lao et al. [33] applied deep neural networks for many tasks in the field of clothing related research, including the prediction of clothing types, clothing retrieval, object detection, attribute prediction, and etc.

FashionNet [28] was proposed with the use of clothing landmarks, which are more important in the description of clothing. The landmark points are marked and used to assist in the classification of clothing types.

Yu et al. [34] divided the landmarks into 32 important positions and built a dataset of more than 200,000 images. They used a three-stage learning method for fashion landmark detection. The first stage maps the features onto each landmark, and the second stage uses the clustering method to learn the correlation between the points. Finally, the feature is restored to the location of each landmark through a deconvolutional layer.

III. DATASET CONSTRUCTION

Firstly, we search clothing brands on the Internet and go to their official websites if they have one. In addition, we go

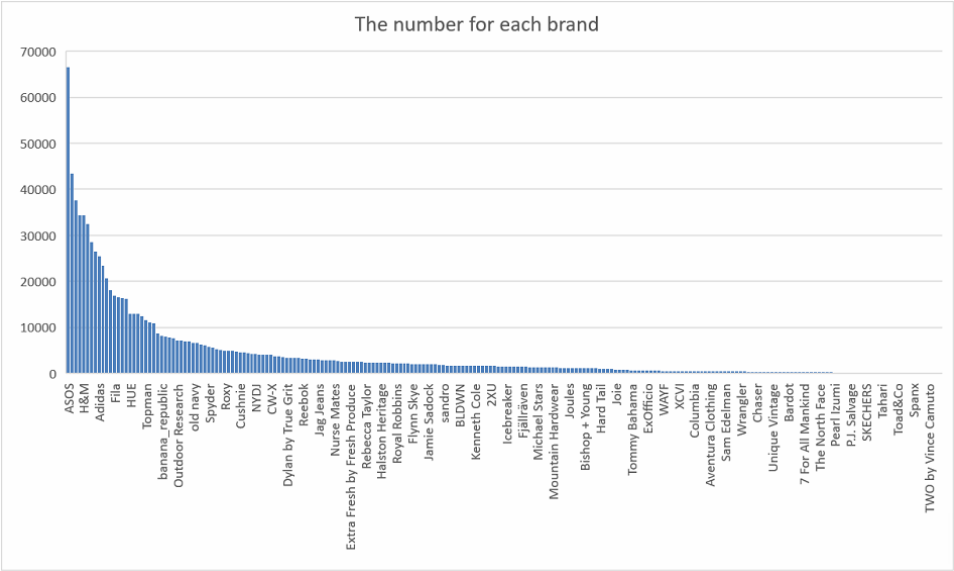

Fig. 2. The number of images for the 224 clothing brands.

to some clothing online shopping websites which have some other clothing brands. In this way, we could collect as many brands as possible.

Secondly, we start to crawl pictures from every website we found, including official websites: H&M, Superdry, Forever21, ROOTS, MANGO and so on, and online shopping websites: Amazon, Zappos, Yahoo, and etc.

In the process of image collection, each clothing item may contain 4 to 8 images corresponding to different viewing angles. All the images of the same item are considered as a pair. Different brands have different number of images and different resolutions. The number of images for all 224 brands are shown in Fig. 2. One difficulty we encountered in collecting clothing attributes is inconsistency, where different websites show attribute in different format. In the end of this stage, our benchmark dataset, clothing brand (CB) dataset, has collected more than one million images with several attributes, such as brand, type, color, material, price, pair, and so on. Fig. 1 shows some sample images for 4 clothing brands. It can be see that some brands have logo and some do not contain any logo.

Finally, we manually label the 250k images from the selected 25 clothing brands with brand logos to form a subset of the CB dataset: clothing brand logo (CBL) dataset. However, there will be no logos for many images. For example, some logos can be seen in front view and cannot be found in other viewing angles for the same clothing item. After the labelling procedure, 57,000 images with clear logos are kept and all of them have brand and bounding box information.

We also compare our dataset with some existing datasets: Clothing Attribute (CA) [4], Apparel Classification with Style (ACS) [2], Colorful-Fashion (CF) [24], Multi-view Clothing (MVC) [27], and DeepFashion (DF) [28]. The comparison is made in several aspects, such as resolution, number of images, types, colors, brands and price. The comparison results are summarized in Table 1. It can be seen that our dataset has the highest resolution and is the only one with brand and price information.

IV. Proposed Method

A. Structure design on dense block

Since YOLOv3 has demonstrated outstanding performance in object detection and recognition, our proposed method is based on YOLOv3 by replacing the residual block with dense block. The overall dense-block based YOLOv3 framework for clothing brand logo prediction is depicted in Fig. 3. The reason we use dense block is to train the DCNN more efficiently and deeper. It is obvious that the connections between all layers are shorter, where every layer is connected to all other layers in feed-forward manner.

In our work, we design 4 different dense-block structures: 2, 3, 4, and 5 convolutional layers, as shown in Fig. 4, 5, 6, and 7, respectively. The dense-block is indicated in the Fig. 3. Each convolutional layer has the size 3x3 and is followed by a ReLU activation layer. All other parameters are set the same as the original YOLOv3 model.

B. Bounding box movement

In general, the bounding box is moved at 1-pixel step to search an object and determine if an object is in the box during the training stage for object detection. However, this 1-pixel movement training process is time-consuming and could be modified to reduce system complexity.

To decide the step size in moving the bounding box, we examined the sizes of all labeled bounding boxes in the training data, where the minimum size is 50×50 and maximun size is $1,400 \times 1,400$.

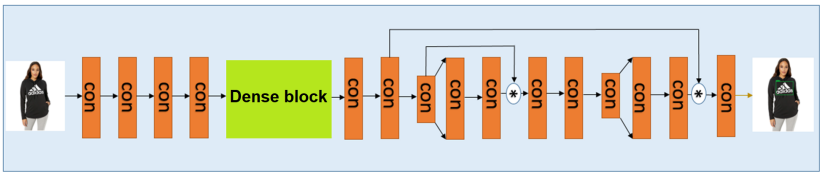

Fig. 3. Proposed dense-block based YOLOv3 framework for clothing brand logo prediction.

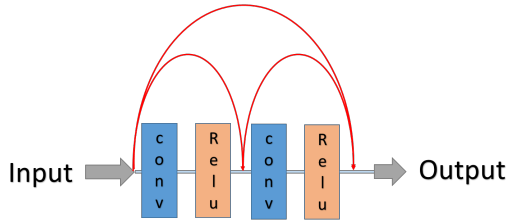

Fig. 4. 2 convolutional layers dense block.

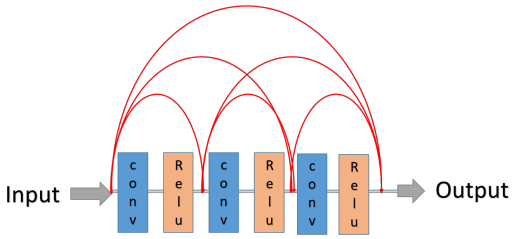

Fig. 5. 3 convolutional layers dense block.

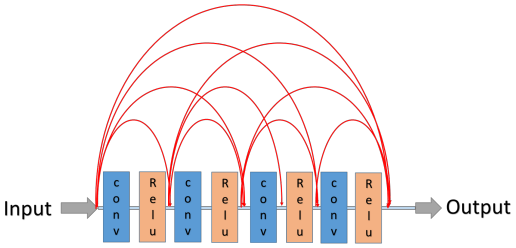

Fig. 6. 4 convolutional layers dense block.

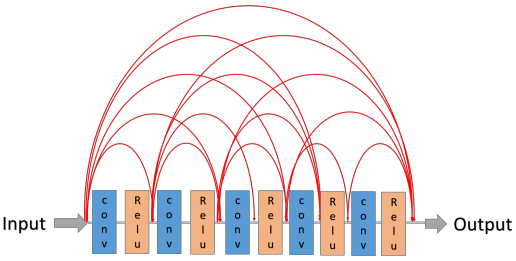

Fig. 7. 5 convolutional layers dense block.

In our proposed framework, we adopt 5-pixel and 10-pixel search steps to for the bounding box movement to reduce the training time and make little performance sacrifice. The experimental results will be demonstrated in the experiment section.

V. Experiments

In the experiment, we use 57,000 clothing images with brand and logo information to validated our proposed framework. For the experimental setting, 70%, 10%, and 20% of images are used in training, validation, and testing, respectively.

We tested our framework for 4 designs of dense block: 2-, 3-, 4-, and 5-layer Dense block. The results are shown in Table II. We also compare our framework with some state-of-the-art methods, where the results are also listed in Table II. The results show that our method can improve the performance greatly and has highest accuracy 62.59%. It proves that our dense-block based YOLOv3 is efficient.

To reduce the training time in our proposed framework, we implement another two bounding box movement strategies in our model. We experiment on 5-pixel and 10-pixel search steps for the bounding box movement in our 5-convolutional-layer dense-block based YOLOv3 model. The experiment results are shown in Table III. One can find that the accuracies are 62.55% and 62.53% for the 5-pixel and 10-pixel search steps, respectively. Compared with original 1-pixel search step, the performance is only decreased 0.04% and 0.06% for the 5-pixel and 10-pixel search steps, respectively. However, the time-complexity is greatly reduced as shown in Table III. As we observed, the time reduction is 12.1% and 24.9% for the 5-pixel and 10-pixel search steps. We believe this search time reduced (search step increased) consideration in our proposed framework is an efficient strategy.

VI. Conclusion

We constructed a new large-scale clothing brand dataset. It is the only clothing dataset containing brand (logo) and price information. A new dense-block based YOLOv3 framework is proposed to tackle with clothing brand logo prediction problem and achieve better performance than several state-of-the-art methods. To reduce model complexity, two pixel search steps are examined for model efficiency. In the future, we would consider two much more challenging tasks: one is clothing brand prediction in the whole CB dataset where no logo information is available for most of images, and another one is clothing price prediction in the whole CB dataset.

References

[1] Y. Kalantidis, L. Kennedy, and L.-J. Li, "Getting the look: clothing recognition and segmentation for automatic product suggestions in everyday photos," in *Proceedings of the 3rd ACM conference on*

TABLE II
COMPARISON OF CLOTHING BRAND LOGO PREDICTION ACCURACY FOR DIFFERENT METHODS

Method	Accuracy
YOLOv2 [35]	45.83%
YOLOv3 [22]	51.22%
RCNN [36]	45.59%
Fast-RCNN [37]	47.38%
Faster-RCNN [38]	48.97%
2-conv-layer dense-block YOLOv3 [ours]	**57.59%**
3-conv-layer dense-block YOLOv3 [ours]	**58.20%**
4-conv-layer dense-block YOLOv3 [ours]	**62.28%**
5-conv-layer dense-block YOLOv3 [ours]	**62.59%**

TABLE III
COMPARISON OF PIXEL SEARCH STEPS ON 5-CONVLUTIONAL-LAYER DENSE BLOCK FOR CLOTHING BRAND LOGO PREDICTION ACCURACY

pixel search step	1	5	10
Time reduction	0.0%	12.1%	24.9%
Accuracy	62.59%	62.55%	62.53%

International conference on multimedia retrieval, pp. 105–112, ACM, 2013.

[2] L. Bossard, M. Dantone, C. Leistner, C. Wengert, T. Quack, and L. Van Gool, "Apparel classification with style," in *Asian conference on computer vision*, pp. 321–335, Springer, 2012.

[3] M. H. Kiapour, K. Yamaguchi, A. C. Berg, and T. L. Berg, "Hipster wars: Discovering elements of fashion styles," in *European conference on computer vision*, pp. 472–488, Springer, 2014.

[4] H. Chen, A. Gallagher, and B. Girod, "Describing clothing by semantic attributes," in *European conference on computer vision*, pp. 609–623, Springer, 2012.

[5] W. Di, C. Wah, A. Bhardwaj, R. Piramuthu, and N. Sundaresan, "Style finder: Fine-grained clothing style detection and retrieval," in *Proceedings of the IEEE Conference on computer vision and pattern recognition workshops*, pp. 8–13, 2013.

[6] S. Liu, Z. Song, G. Liu, C. Xu, H. Lu, and S. Yan, "Street-to-shop: Cross-scenario clothing retrieval via parts alignment and auxiliary set," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pp. 3330–3337, IEEE, 2012.

[7] K.-H. Liu, F. Wang, and T.-J. Liu, "A clothing recommendation dataset for online shopping," in *2019 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW)*, pp. 1–2, IEEE, 2019.

[8] K.-H. Liu and T.-J. Liu, "A structure-based human facial age estimation framework under a constrained condition," *IEEE Transactions on Image Processing*, vol. 28, no. 10, pp. 5187–5200, 2019.

[9] K.-H. Liu, S. Yan, and C.-C. J. Kuo, "Age estimation via grouping and decision fusion," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 11, pp. 2408–2423, 2015.

[10] X. Zeng, C. Ding, Y. Wen, and D. Tao, "Soft-ranking label encoding for robust facial age estimation," *arXiv preprint arXiv:1906.03625*, 2019.

[11] K.-H. Liu, H.-H. Liu, P. K. Chan, T.-J. Liu, and S.-C. Pei, "Age estimation via fusion of depthwise separable convolutional neural networks," in *2018 IEEE International Workshop on Information Forensics and Security (WIFS)*, pp. 1–8, IEEE, 2018.

[12] K.-H. Liu, P. K. Chan, T.-J. Liu, and H.-A. Her, "Age estimation for low-quality facial images: from separate dcnns to a decision fuser," in *2019 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pp. 168–173, IEEE, 2019.

[13] K.-H. Liu, H.-H. Liu, S.-C. Pei, T.-J. Liu, and C.-T. Chang, "Age estimation on low quality face images," in *2019 IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS)*, pp. 295–296, IEEE, 2019.

[14] K.-H. Liu, C.-T. Chang, and T.-J. Liu, "Age estimation via modern convolutional neural networks," in *2019 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW)*, pp. 1–2, IEEE, 2019.

[15] Y.-Z. Su, T.-J. Liu, K.-H. Liu, H.-H. Liu, and S.-C. Pei, "Image inpainting for random areas using dense context features," in *2019 IEEE International Conference on Image Processing (ICIP)*, pp. 4679–4683, IEEE, 2019.

[16] B.-X. Chen, T.-J. Liu, K.-H. Liu, H.-H. Liu, and S.-C. Pei, "Image super-resolution using complex dense block on generative adversarial networks," in *2019 IEEE International Conference on Image Processing (ICIP)*, pp. 2866–2870, IEEE, 2019.

[17] T.-J. Liu, K.-H. Liu, and K.-H. Shen, "Learning based no-reference metric for assessing quality of experience of stereoscopic images," *Journal of Visual Communication and Image Representation*, vol. 61, pp. 272–283, 2019.

[18] T.-J. Liu and K.-H. Liu, "No-reference image quality assessment by wide-perceptual-domain scorer ensemble method," *IEEE Transactions on Image Processing*, vol. 27, no. 3, pp. 1138–1151, 2018.

[19] T.-J. Liu, W. Lin, and C.-C. J. Kuo, "Image quality assessment using multi-method fusion," *IEEE Transactions on image processing*, vol. 22, no. 5, pp. 1793–1807, 2013.

[20] T.-J. Liu, K.-H. Liu, J. Y. Lin, W. Lin, and C.-C. J. Kuo, "A paraboost method to image quality assessment," *IEEE transactions on neural networks and learning systems*, vol. 28, no. 1, pp. 107–121, 2017.

[21] T.-J. Liu, H.-H. Liu, S.-C. Pei, and K.-H. Liu, "A high-definition diversity-scene database for image quality assessment," *IEEE Access*, vol. 6, pp. 45427–45438, 2018.

[22] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[23] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708, 2017.

[24] S. Liu, J. Feng, C. Domokos, H. Xu, J. Huang, Z. Hu, and S. Yan, "Fashion parsing with weak color-category labels," *IEEE Transactions on Multimedia*, vol. 16, no. 1, pp. 253–265, 2014.

[25] J. Huang, R. S. Feris, Q. Chen, and S. Yan, "Cross-domain image retrieval with a dual attribute-aware ranking network," in *Proceedings of the IEEE international conference on computer vision*, pp. 1062–1070, 2015.

[26] K. Lin, H.-F. Yang, K.-H. Liu, J.-H. Hsiao, and C.-S. Chen, "Rapid clothing retrieval via deep learning of binary codes and hierarchical search," in *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*, pp. 499–502, ACM, 2015.

[27] K.-H. Liu, T.-Y. Chen, and C.-S. Chen, "Mvc: A dataset for view-invariant clothing retrieval and attribute prediction," in *Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval*, pp. 313–316, 2016.

[28] Z. Liu, P. Luo, S. Qiu, X. Wang, and X. Tang, "Deepfashion: Powering robust clothes recognition and retrieval with rich annotations," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1096–1104, 2016.

[29] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[30] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[32] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2472–2481, 2018.

[33] B. Lao and K. Jagadeesh, "Convolutional neural networks for fashion classification and object detection," *CCCV 2015: Computer Vision*, pp. 120–129, 2016.

[34] W. Yu, X. Liang, K. Gong, C. Jiang, N. Xiao, and L. Lin, "Layout-graph reasoning for fashion landmark detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2937–2945, 2019.

[35] J. Zhang, M. Huang, X. Jin, and X. Li, "A real-time chinese traffic sign detection algorithm based on modified yolov2," *Algorithms*, vol. 10, no. 4, p. 127, 2017.

[36] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969, 2017.

[37] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, pp. 1440–1448, 2015.

[38] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, pp. 91–99, 2015.