


## **Retraction Notice**

The Editor-in-Chief and the publisher have retracted this article, which was submitted as part of a guest-edited special section. An investigation uncovered evidence of systematic manipulation of the publication process, including compromised peer review. The Editor and publisher no longer have confidence in the results and conclusions of the article.

XY either did not respond directly or could not be reached.

# Multidimensional graphic art design method based on visual analysis technology in intelligent environment

Xiaoyu Yan \*

Shaanxi Xueqian Normal University, Academy of Fine Arts, Xi'an, China

**Abstract.** In the last few years, the artificial intelligence technology has provided unique methods for design analysis in the art field. With the development of China's economy and cultural prosperity, graphic design has changed people's lives greatly. Under the concept of modern scientific and technological creation, the variety, scale, and plasticity of graphic design art are constantly improving, and the diversification of design elements is becoming the development trend. We designed a graphic art element recognition model based on single shot multibox detector (SSD) method through deep learning of visual processing technology. This method can automatically identify various elements in multidimensional graphic art works, thus helping artists and learners analyze an art work better. In this method, we take the SSD structure as the main model backbone and use the improved attention mechanism module feature pyramid transformer to replace the original feature fusion module, inject long-distance dependency into the model, and improve the accuracy of object detection. In addition, we use the public dataset to make the relevant image target detection dataset. Different object detection evaluation metrics are used to evaluate the proposed methods, and several existing methods are selected for comparative experiments. Compared with YOLO V5 object detection model, our method improves 0.53%, 0.67%, 1.33%, and 1.28% on pixel accuracy, mean pixel accuracy, average recall, and mean intersection over union, respectively. The proposed algorithm has a great contribution to the performance improvement of object detection and the auxiliary analysis of multidimensional works of art. © 2023 SPIE and IS&T [DOI: [10.1117/1.JEL.32.6.062507](https://doi.org/10.1117/1.JEL.32.6.062507)]

**Keywords:** graphic art; multidimensional; single shot multibox detector; visual analysis; deep learning.

Paper 221317SS received Nov. 22, 2022; accepted for publication Dec. 28, 2022; published online Jan. 15, 2023; retracted Jul. 8, 2023.

## 1 Introduction

As a new technology, artificial intelligence is widely used in various fields due to its characteristics of automation and intelligence. Visual analysis technology has been applied in the field of art analysis due to its ability to comprehensively analyze images. In graphic art, we will see a large number of words, graphics, symbols, and so on. The works of art composed of these elements convey a variety of information to people. For graphic art design, it is necessary to create works with different life forms through multiangle, multilevel, and multiperspective creation,<sup>1</sup> which is extremely important for the overall development of art. By building multiple perception dimensions, enhancing the visual perception of design works, and improving the diversity of elements according to the preferences of target user groups, people can meet their increasing artistic pursuit. Therefore, the multidimensional elements of graphic art design can not only enhance the texture and visual impact of works of art<sup>2</sup> but also improve the form of expression of works. For the analysis of works of art, it is helpful for artists to appreciate and learn works by accurately identifying various elements of works through automatic methods.

In recent years, visual analysis technology has been tending to automation and intelligence, and artificial intelligence is increasingly used in the field of computer vision. Machine learning<sup>3,4</sup> enables it to learn rules from a large amount of historical data through algorithms so as to

---

\*Address all correspondence to Xiaoyu Yan, [xiaoyuyan6111@163.com](mailto:xiaoyuyan6111@163.com)

intelligently identify new samples or predict the future. Principal component analysis (PCA) is a common method of machine learning analysis. Jiang et al.<sup>5</sup> took advantage of the characteristics of PCA technology and used it as an efficient preprocessing method in hyperspectral image classification and analysis. In addition, the author improved the original algorithm to the method of super pixel PCA by some methods to solve the problem of different image spectra caused by different homogeneous regions. As an automatic classifier, support vector machine (SVM) is also widely used in the field of visual analysis. Jang et al.<sup>6</sup> predicted cloud motion vectors using SVM and atmospheric motion vector satellite images. The SVM is used as a classifier to classify a large number of historical satellite image data so as to predict the solar power of photovoltaic power station. Decision tree is a decision analysis algorithm. It analyzes and predicts possible events through the tree structure. Srabanti and Srishti et al.<sup>7</sup> used the method of decision tree combined with artificial neural network to predict heart disease. This is a typical application of decision trees in the field of data mining. Using decision trees to learn and predict a large number of patient information greatly reduces the number of medical examinations required by patients. Although these machine learning algorithms can solve problems automatically to a certain extent, in the context of big data, when facing large amounts of data, feature engineering in machine learning will consume a lot of time.

In recent years, deep learning<sup>8</sup> has been applied in various fields, especially in the context of big data. Deep learning for large-scale data has better performance than machine learning. Computer vision analysis has always been a major field of deep learning applications. Target detection<sup>9</sup> technology combines segmentation and recognition to track the target in the image. Regions with CNN features (R-CNN)<sup>10</sup> first applied deep learning to the field of target detection. R-CNN applied CNN (ConvNet) to calculate feature vectors for region proposals. From experience-driven features to data driven-features, R-CNN enhanced the representation of features. Fast R-CNN<sup>11</sup> combined with spatial pyramid pooling net (SPPNet)<sup>12</sup> improved R-CNN by extracting image features only once, and then mapping the feature map of candidate regions to the feature map of the whole image according to the algorithm, greatly improving the running speed of the model. You only look once (YOLO)<sup>13</sup> creatively treated the object detection task as a regression problem, and combined the candidate area and detection phases into one. People can know which objects and their positions are in each image at a glance. The single shot multibox detector (SSD)<sup>14</sup> model detected the feature map obtained from each convolution based on the feature pyramid. This detection method used the idea of multiscale feature fusion to detect feature maps of different scales, greatly improved the accuracy of small target detection. Although these target detection models can complete detection tasks on some datasets, some models do not consider multiscale features. Although SSD have feature fusion modules, long-range information is missing from features. How to further improve the accuracy of the object detection model and apply it to the element recognition of multidimensional graphic design is the problem we need to solve.

To better apply the deep learning object detection algorithm to the recognition of elements in graphic art design, in this paper, we combine the attention mechanism and the improved SSD model to build a recognition method. In addition, we annotated large-scale graphic art works in the object detection way to evaluate our proposed methods. The main contributions are as follows.

- (1) Use the method of deep learning to accurately identify the multidimensional elements in graphic art design, and help artists and art learners analyze art works.
- (2) The attention mechanism model of feature pyramid transformer (FPT) is improved so that its output feature map can meet the needs of target detection.
- (3) The feature fusion structure of SSD target detection model is improved, and the long-distance attention mechanism information is added to increase the accuracy of target detection. Compared with the existing models, the improved SSD model has a certain performance improvement.

The organization of this article is as follows. Section 2 introduces the related works; Sec. 3 describes the proposed methods, which includes the overall structure of the model, the attention mechanism module, and the dataset; and Sec. 4 describes the experiment and result analysis of the performance of proposed method. The conclusion is presented in Sec. 5.

## 2 Related Work

With the improvement of people's living standards, the multidimensional graphic art design has become the development trend and the goal pursued by people. The object detection algorithm based on deep learning can automatically locate and classify various elements from graphic design works, which is helpful for artists and art learners to analyze and learn works. To realize the automatic recognition of elements of multidimensional graphic art works with high accuracy, we have carried out relevant research based on SSD algorithm.

As a high-performance target detection model, SSD is widely used in various fields. Liao et al.<sup>15</sup> used MobileNet as the feature extractor of SSDs and extracted feature maps from the convolution results of different layers of MobileNet as the input of the feature fusion module in the original SSD to improve the performance of the original SSD. The improved SSD is used to realize automatic recognition of occlusion gesture. Yang et al.<sup>16</sup> introduced the divided evolution and attention residuals module into the original SSD model, fused the sparse pixel feature map with the dense pixel feature map, and improved the resolution. The purpose of the improvement is to improve the accuracy of small object detection. Liang et al.<sup>17</sup> improved the default box of the single shot multibox detector by remaking the shape of the default box on several feature maps and realized real-time detection of mangoes on trees. These models have improved the performance of the original SSD to a certain extent and also tried to generate the accuracy of target detection at different scales through new feature fusion methods, but they did not consider the long-range dependency and the need for global information.

Self-attention is often used in the field of computer vision to give global information and long-range dependency to feature maps. Nonlocal introduced the attention mechanism into the field of computer vision by using an ingenious feature matrix operation.<sup>18</sup> This module can be inserted into the CNN and directly act on the characteristic graph after the convolution operation. However, nonlocal has the problem of too much computation, and the processing process contains a lot of redundant information. Lin et al.<sup>19</sup> proposed a cross attention mechanism to reduce the number of parameters and remove redundant information by sampling the matrix operated in nonlocal. In the cross attention mechanism, two modules in series are used to collect complete global information. Zhang et al.<sup>20</sup> proposed the FPT module, which can also be inserted into the CNN for direct use. This module also learns from the nonlocal computing mode but also takes into account the multiscale information fusion. These attention mechanism modules have improved the performance of computer vision tasks to varying degrees. How to apply the attention mechanism to target detection technology and improve the automatic recognition of elements of multidimensional graphic art works is a problem we need to study.

## 3 Methods of Identifying Elements in Multidimensional Graphic

To accurately identify various elements in graphic art design works and assist in the learning and analysis of multidimensional graphic design, we propose a multidimensional graphic art design element recognition algorithm based on SSD. The overall method flow is shown in Fig. 1. In the overall approach, the first part is data collection. The dataset we use is an open source DesignNet graphic design dataset, which contains a large number of graphic art design works. Some of them were selected as the original data. The next step is the annotation of the dataset. We labeled the dataset with the object detection annotation way and then use the data enhancement method to enrich the dataset. At the same time, we build the SSD-based object detection overall model framework and FPT attention mechanism module. After that, we use the established dataset to train the object detection model and use the test set to evaluate the model. The trained model can automatically identify multiple elements in multidimensional graphic art design. Subsequent experiments have proved that our method is superior to some existing target detection models in detection accuracy and has a good practical performance in plane design dataset.

### 3.1 Object Detection Model Based on SSD

The overall structure of the model we designed is based on the SSD model, as shown in Fig. 2. First, the model has seven convolution modules and its internal structure is in the form of Conv

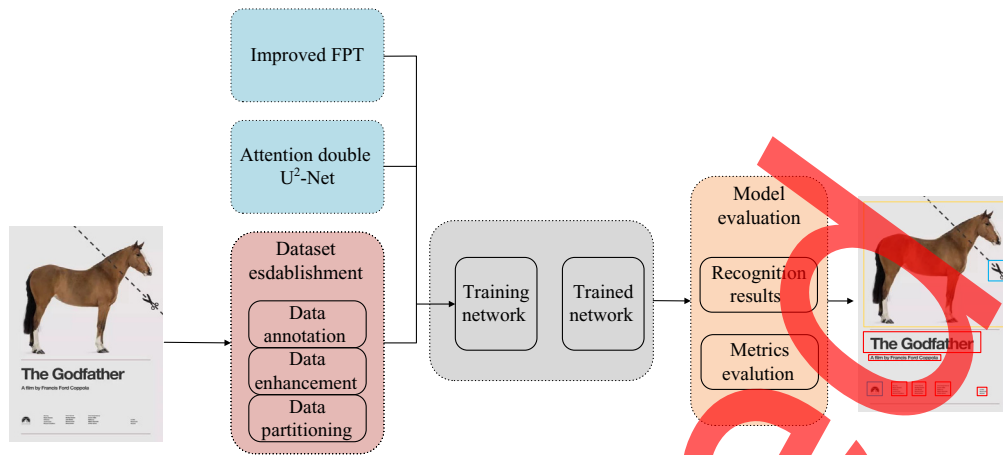


Fig. 1 Graphic art design element identification overall flow chart.

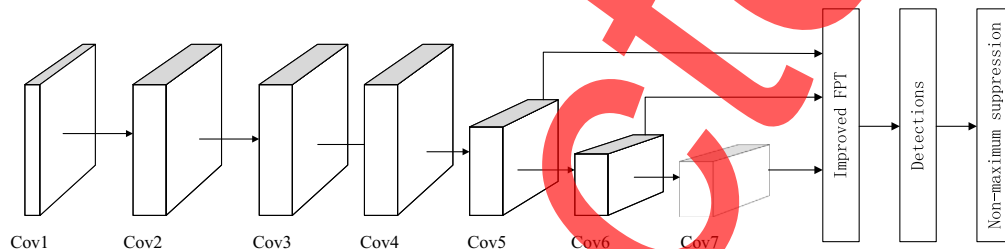


Fig. 2 Object detection model structure based on SSD.

+ReLU. Seven convolutional blocks constitute the feature extractor of the model. Three of the seven convolution blocks contain down sampling, which is different from the original SSD model. Because there are very few small objects that cannot be distinguished visually in the multidimensional plane art dataset we make, too much down sampling will only increase the pixel loss of the image but will not help to detect small objects. After the seven convolution modules, we propose the improved FPT module. The input of this module is the characteristic diagram of the output of three convolution blocks: Cov5, Cov6, and Cov7. This module uses the idea of attention fusion to not only add long-range dependency to the feature graph but also perform feature fusion. After the improved FPT is the detections module to complete the classification task in object detection. The final non-maximum suppression module is used to locate the detection frame in object detection.

### 3.2 Improved FPT Module

In image tasks, such as object detection, the convolution operation receptive field of convolutional neural network is limited and can only provide short-range dependence of different scales. The detection of large objects in images requires a long-range dependence and global information. In addition, feature fusion has been proved by many methods to effectively improve the performance of computer vision models. To solve the above problems and give consideration to the global information and feature fusion, we replace the feature fusion structure in the original SSD with the improved FPT structure. The overall structure of FPT is shown in Fig. 3. In this structure, the three input feature maps are from the previous convolution module and have different scales. After self-attention or mutual attention with other scales, these feature maps can generate three new feature maps of different scales by concatenation operation. These new feature maps are full of global information from different scales and local information from different scales. What is different from the original FPT is that in the end, we did the up sampling operation on the two small feature maps. The up sampling method is bilinear interpolation. After

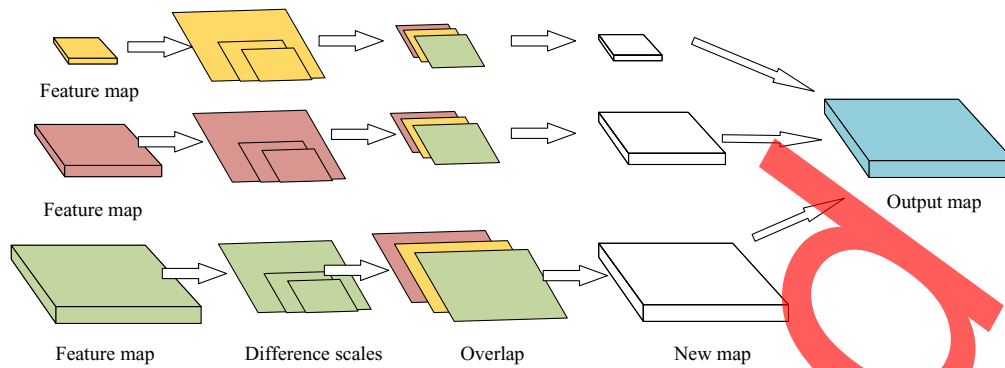


Fig. 3 Structure of the improved FPT.

upsampling, the three feature maps have the same scale. Concatenation is performed to generate a single, multichannel feature map. The reason for this is first to meet the needs of target detection, and second to better integrate information of different scales. Experiments show that the addition of this module can give the model better performance.

### 3.3 Dataset Establishment

Although there are open source graphic art design datasets, these datasets are original images without annotation. To solve this problem, we use manual annotation and data enhancement to establish a multidimensional graphic art design element identification dataset. The process of dataset creation is shown in Fig. 4. The first is the annotation of datasets. The tool we use is the open-source data annotation tool Labelme. All labeling work is done manually. We have marked 3000 pieces of graphic art works with multidimensional characteristics. In these works, we have marked five different types of elements: text elements, graphic elements, symbol elements, natural elements, and abstract elements. Each different element is distinguished by different color detection boxes.

Data enhancement can enable limited data to generate more data, increase the number and diversity of training samples, and thus improve the robustness of the model. We have used different data enhancement methods to enrich our datasets, including random angle rotation, random flip, random cropping, random contrast enhancement, and color change. The final size of the dataset after several rounds of enhancement is 30,000 pictures, of which 25,000 are used as training sets, 2500 are used as verification sets, and 2500 are used as test sets.

## 4 Experiment and Analysis

To verify the recognition performance of proposed SSD based target detection algorithm on the graphic art design dataset, we have done a series of experiments. First, we selected the evaluation metrics commonly used in the internal test to evaluate the model. These metrics include average precision (AP), mean AP (mAP), average recall (AR), and mean intersection over union (MIoU).

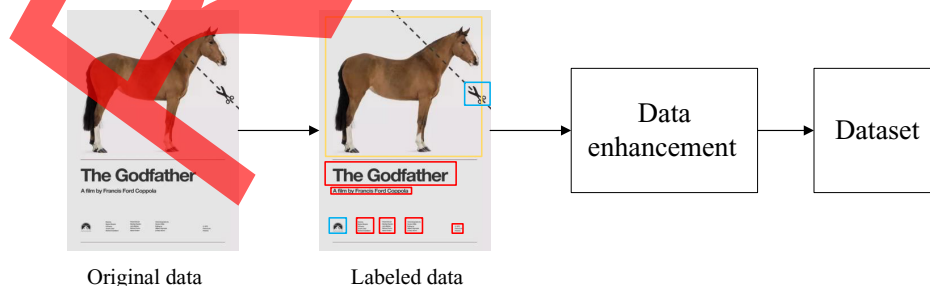


Fig. 4 Dataset production process.

In addition, we selected several high-performance objected detection models for comparative experiments, including R-CNN, raw SSD, Fast R-CNN, and YOLO V5. All experiments were completed in the following environments: CPU Intel Conroe i9-13900K, GPU 3080 with memory of 10g, 32g RAM, and Win10 operating system.

#### 4.1 Evaluation Metrics

Common evaluation indicators for target detection are as follows.

- (1) AP describes the number of correct classifiers describing the population

$$AP = \frac{TP + TN}{\text{all detections}}. \quad (1)$$

- (2) Mean AP describes the average number of correct classifiers per category year

$$mAP = \frac{TP}{\text{all detections} \cdot \text{ClassNum}}. \quad (2)$$

- (3) AR describes the number of correct predictions in the real label precision

$$AR = \frac{TP}{\text{all groundtruth}}. \quad (3)$$

- (4) Mean intersection over union (MIoU) describes the ratio between the intersection and the union of the rectangular box of the detection result and the rectangular box labeled by the sample under the category average. This evaluation metric is used to evaluate the accuracy of object detection model detection frame positioning

$$MIoU = \frac{1}{k+1} \sum_{i=0}^k \frac{TP}{FN + FP + TP}, \quad (4)$$

where  $k$  represents the total number of categories, and  $k+1$  represents including background classes.

#### 4.2 Contrast Experiment

To verify the effectiveness of our proposed object detection model in the element classification of graphic art works, we designed a contrast experiment. The existing object detection models involved in the comparison are: (1) R-CNN model, which is the first model to apply deep learning to object detection. (2) Original SSD, model without the attention mechanism module. (3) The Fast R-CNN model, improved from the R-CNN model, has strong performance. (3) YOLO V5 object detection model, the latest generation of YOLO series, with adaptive anchor algorithm and focus. The results of the comparison experiment are shown in Table 1.

**Table 1** Comparison between proposed method and others.

Methods	AP	mAP	AR	MIoU
R-CNN	96.70%	95.21%	88.71%	85.42%
SSD	97.05%	95.65%	89.78%	87.32%
Fast R-CNN	97.43%	96.34%	90.31%	87.56%
YOLO V5	98.04%	96.87%	91.54%	88.61%
Ours	98.57%	97.54%	92.87%	89.87%

The experimental data show that our method has a comprehensive lead in the recognition of graphic art elements. Compared with the unimproved SSD algorithm, our method has improved 1.52%, 1.89%, 3.09%, and 2.57% on the four evaluation metrics respectively. The last two evaluation metrics can be said to be greatly improved, which means that the proposed method has not only greatly improved the classification accuracy but also more accurate in the element positioning. Compared with YOLO V5 model with the second best performance, our method leads by 0.53%, 0.67%, 1.33%, and 1.28% in four evaluation metrics, respectively. These data prove that our method has better performance, which means that the classification and positioning of various elements in multidimensional graphic art works will be more accurate. All kinds of small elements that could not be recognized would also be recognized because of our attention mechanism and multiscale fusion modules. The experiment proves that our method is practical and valuable.

### 4.3 Results Display

In this section, we randomly select some samples from the test set, identify them through the multidimensional graphic art design element identification method, and show the results. The results are shown in Fig. 5. It can be seen that different multi-dimensional art elements are marked by different color detection boxes. In the second picture, the small text marked by the red detection box at the lower right corner indicates that our method can accurately locate and classify small targets. The second picture has a variety of text elements. These text elements have different fonts, sizes, and even some fonts are abstract. These abstract or normal fonts can be accurately recognized. In the first picture, the purple detection box detects abstract elements, which is also a difficulty in identifying graphic art elements. Only when the deep learning model has certain performance and a large number of relevant samples are fully trained, can it distinguish abstract graphics and general graphics. The results show that our method has a good practical value and has a certain contribution to multidimensional graphic art design.

To show the robustness of our method, we selected the results in complex situations, as shown in Fig. 6. In the situation of dense elements, a wide variety of works of art, and it is difficult to define the types of elements, the method we proposed still completes the task well. The whole work of art uses the style of cartoon, but it still recognizes the natural elements that can appear in real life. In addition, some abstract text elements with strange colors and shapes are also accurately recognized. Some small target text elements that are not clear enough are also accurately located. This shows that our method is not only accurate but also robust and universal.

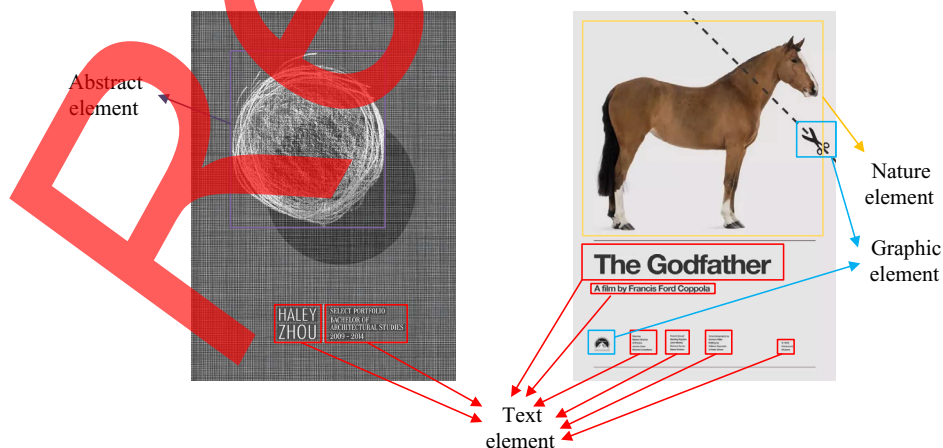


Fig. 5 Recognition results.





Fig. 6 Presentation of results in complex situations.

## 5 Conclusion

In this paper, we design a multidimensional graphic art design element recognition method based on SSD object detection model and attention mechanism. In this method, a proposed attention mechanism module is inserted with SSD as the theme. The experiment proves that our method has good performance, compared with YOLO V5 target detection model, our method improves by 0.53%, 0.67%, 1.33%, and 1.28% on pixel accuracy, mean pixel accuracy, AR, and MIoU, respectively. This method can help artists and art learners to distinguish various elements in art works and has good practical value.

For the subsequent work, we consider using more samples for training to improve the robustness of the model. This also means building larger datasets and doing more data annotation work. In addition, the follow-up work will also broaden the element categories, and refine the element categories on the basis of this article. Finally, we hope to add a function of overall evaluation to automatically identify and evaluate the style and color style of a graphic art work. It is hoped that we can achieve more functions in the subsequent work and make the model have better performance.

## Acknowledgments

The work received no funding. The authors declare that there are no conflicts of interest.

## Code, Data, and Materials Availability

The dataset can be accessed upon request.

## References

1. J. Liu, "Research on multi-dimensional practical teaching system of art design major in Ming and Qing dynasty furniture design based on sample data analysis," *J. Phys. Conf. Ser.* **1852**, 042099 (2021)
2. S. Frank, *Understanding Visual Images in Picturebooks*, Talking beyond the page, pp. 10–25, Routledge (2020).
3. I. H. Sarker, "Machine learning: algorithms, real-world applications and research directions," *SN Comput. Sci.* **2**, 160 (2021).
4. Q. Bi et al., "What is machine learning? A primer for the epidemiologist," *Am. J. Epidemiol.* **188**, 2222–2239 (2019)

5. J. Jiang et al., "SuperPCA: a superpixelwise PCA approach for unsupervised feature extraction of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.* **56**, 4581–4593 (2018).
6. H. S. Jang et al., "Solar power prediction based on satellite images and support vector machine," *IEEE Trans. Sustain. Energy* **7**, 1255–1263 (2016).
7. M. Srabanti and A. Srishti, "Decision tree algorithms for prediction of heart disease," in *Information and Communication Technology for Competitive Strategies*, pp. 447–454, Springer, Singapore (2019).
8. Y. Guo et al., "Deep learning for visual understanding: a review," *Neurocomputing* **187**, 27–48 (2016).
9. Z. Zou et al., "Object detection in 20 years: a survey," arXiv:1905.05055 (2019).
10. R. Girshick et al., "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 580–587 (2013).
11. R. Girshick, "Fast R-cnn," in *Proc. IEEE Int. Conf. Comput. Vision*, pp. 1440–1448 (2015).
12. K. He et al., "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.* **37**, 1904–1916 (2015).
13. J. Redmon et al., "You only look once: unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 779–788 (2015).
14. W. Liu et al., "SSD: single shot multibox detector," *Lect. Notes Comput. Sci.* **9905**, 21–37 (2016).
15. S. Liao et al., "Occlusion gesture recognition based on improved SSD," *Concurr. Comput.: Pract. Exp.* **33**, e6063 (2021).
16. L. Yang et al., "Pipeline magnetic flux leakage image detection algorithm based on multi-scale SSD network," *IEEE Trans. Ind. Inf.* **16**, 501–509 (2019).
17. Q. Liang et al., "A real-time detection framework for on-tree mango based on SSD network," *Lect. Notes Comput. Sci.* **10985**, 423–436 (2018).
18. X. Wang et al., "Non-local neural networks," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 7794–7803 (2017).
19. H. Lin et al., "CAT: Cross attention in vision transformer," in *IEEE Int. Conf. Multimedia and Expo (ICME)*, IEEE, pp. 1–6 (2022).
20. D. Zhang et al., "Feature pyramid transformer," *Lect. Notes Comput. Sci.* **12373**, 323–339 (2020).

**Xiaoyu Yan** is a lecturer at Fine Arts Department of Shaanxi Xueqian Normal University; he is mainly engaged in design teaching and has successively served as an external teacher at the School of Continuing Education of Xi'an Academy of Fine Arts and the High tech College of Xi'an University of Science and Technology. He is now a member of Shaanxi Intangible Cultural Heritage Association.