Automatic Scoring System for Handwritten Examination Papers Based on YOLO Algorithm

Mingliang Lu, Weili Zhou* and Ruijie Ji
School of Electronic and Information Engineering, Foshan University, Foshan, China
Email: willychow@163.com

Abstract. Written examination is an important way of measuring knowledge and ability, but manual scoring of papers is tedious and error-prone. Automated scoring is more fair, accurate and efficient. With the development of artificial intelligence, automatic scoring of papers through image recognition and object detection is becoming an achievable and promising technology today. The aim of this paper is to design an automatic scoring system for objective questions in the examination papers. The automatic scoring system uses YOLOv3 technique to detect and recognize handwritten numbers and characters on examination papers. It also addresses the problem of incorrect recognition due to scribbles. Compared to optical symbol recognition, it can recognize the handwritten answers without extra answer cards. In addition, there is no limit where the student can fill in the answer. The experiments show that the automatic scoring system performs satisfactorily and has good prospects for practical application in the future.

1. Introduction
Examinations become an important part of the educational process as a means of identifying and judging knowledge and skills. The scoring of the examination papers is an important part of the examination process, it is always necessary for teachers mark all examination papers one by one. However manual scoring of papers is tedious and error-prone. Some important exams, such as SSC [1], civil service and postgraduate exams where one question is misjudged may detriment the candidate’s future. In the early 1960s, IBM developed the IBM 1230, which first used the OMR (Optical Mark Reader) technology [2], and since then, automatic scoring systems entered the OMR era. The OMR marking technology, with its dedicated readers and answer cards, became the first choice for automated marking. Later, the Remark Office software became more sophisticated, with a rounded area for filling in answer cards, making it ideal for scoring objective questions.

In China, it was not until the 1980s that scholars began to study the scoring system based on OMR automatic identification technology, after years of research and development, the field of automatic marking has now a variety of mature and complete automatic marking system. The mainstream products include: Nanhao scanning online marking system, the paper and pencil king marking system, etc. [3].Most of these marking system use OMR technology with the related equipment to carry out the automatic scoring of the objective question answer, such as: scanner equipment, answer cards and other supporting additional products. The automatic scoring systems are expensive, and often used in the large-scale examinations.

This shows that there is a considerable demand for automated marking systems and that they are constantly being researched and continuously developed in, but most are OMR technologies. The current systems require the use of consumables such as answer cards. In this paper, we propose an
answer card-less automated scoring systems for examination paper based on YOLO algorithm. The examination paper is processed using YOLO with a convolutional neural network [4]. The handwritten characters such as student numbers and candidates' answers in the examination paper are automatic detected, recognized and scored. The system is suitable for most types of examinations and it is effective and accurate [5].

2. Automatic Scoring Algorithm for Objective Questions in Examination Papers
The overall process of handwritten character recognition based on YOLOv3 shown in Figure 1. First, images of the examination paper are collected and dividing them into a training set and a test; Then, the labellmg annotation tool is used to mark the position of the objective questions answers (such as A, B, C and D) on each image in the training set; then generate an XML file and convert the XML file into a TXT file. Modify the YOLOv3 algorithm by adding the categories of questions answers to the classification and train the dataset. After training, the weights of the YOLO are update. Finally, the Bounding Box of each multiple-choice question in each image is regressed to achieve the localization of multiple-choice questions.

![Figure 1. Overall flow chart.](image)

2.1. YOLOv3 Target Detection Algorithm
YOLOv3 [6-7] is the fastest and most accurate target detection network. YOLOv3 improves the accuracy and small target detection of the YOLO algorithms through a combination of advanced methods. The main improvements in YOLOv3 are the use of: 1) multi-scale predictions across scales (FPN-like); 2) the underlying network architecture is darknet-53 [8], which is darknet-19 compared to v2 classifier with better results; and 3) classifier-category prediction. YOLOv3 uses multiple independent binary cross entropy (also known as logistic regression) for binary classification instead of Softmax units. The advantage of this approach is that it can handle the problem of multiple overlapping labels and the use of multiple independent logistic classifiers instead of Softmax does not degrade the accuracy.

2.2. Marking of Objective Questions on Examination Papers Using the YOLOv3 Target Detection Algorithm

2.2.1. Construction of the Examination Paper Datasets. The development of deep learning in various industries is inseparable from the development of datasets [9], so establishing good datasets provides good prerequisites for deep learning. In order to better train the deep learning network, an examination
papers dataset was created. The specific creation process was as follows.

1) Take a shot at the native paper multiple choice questions and divide them into a training set and a test set.

2) The labelImg [10] annotation tool was used to annotate each image in the training set using Bounding Box to mark the specific location of each objective question, and to classify each handwritten answer in the image as A, B, C and D.

3) After each multiple-choice question in the image has been accurately annotated, an xml file is generated for each image. For example, if the original image is "88.jpg", the file will be generated as "88.xml".

4) Transform the xml format file into a txt format tag file.

2.2.2. Training Step

1) Prediction Bounding Box

The YOLOv3 algorithm obtains anchor boxes [11] by clustering, the bounding box can be predicted based on the offset of the top left corner of the picture.

2) Classification of bounding boxes

For each bounding box prediction classification (0 and 1 classification), the bounding box representing 0 is removed from the graph and the bounding box representing 1 is kept, and the bounding box is classified using multi-label classification [12]. the YOLOv3 algorithm uses simple logistic regression for classification, using a binary cross- entropy loss) function.

3) Cross-scale prediction

The network can predict boxes at three different scales. yolov3 uses a feature pyramid network to extract the special features. In the basic feature extractor, several convolutional layers are added. The last of these layers predicts Bounding Box (bounding box), objectness (object) and category prediction for 3D tensor encoding [13]. The bounding box priors are determined using k-means clustering, nine clusters and three scales are selected, and the clusters are then split evenly across the scales.

4) Performing feature extraction

The YOLOv3 feature extraction model is a hybrid of several models [14], using YOLOv2, Darknet-19 and a residual network, which uses the better performing 3×3 and 1×1 convolutional layers and also adds a shortcut connection structure. The final model has 53 convolutional layers, hence the name Darknet-53 [15].

5) Put the training set into the Darknet-53 network for training, output the positioning of each multiple choice question, train the weights file is automatically generated after completion. Save every 100 iterations when there are less than 1000 iterations, or every 10000 iterations when there are more than 1000 iterations [16].

2.2.3. Testing of Images Using the Completed Training Weight File. Test the images in the test set and output the test images with each multiple choice question precisely located on the Bounding Box, thus identifying multiple choice questions on a single image. The Bounding Box does not appear, but the corrected answers are also accurately identified [17-18].

3. Experiments

3.1. Experimental Settings

The environment chosen for this experiment is the tensorflow framework [19] and the model used is the YOLOV3 target detection algorithm. The training batch size was set to 8, the image size was set to 416 × 416 by default, the number of input channels was 3, the momentum was set to 0.9, and the learning rate was set to 0.001. The data set was divided into a training set and a validation set in the ratio of 9:1. YOLOv3 predicted the score (overlap rate) [14] of an object for each bounding box by logistic regression, with the overlap rate referring to the weighting between the predicted box and the true box. If the predicted bounding box overlaps the true bounding box by a large margin and is better
than the predicted value of all other bounding boxes, then the value is 1. If the overlap does not reach a threshold (set here to 0.5), then the predicted bounding box is ignored, representing a no-loss value [20].

3.2. Experimental Results

Figure 2 shows the performance of a Chinese test paper tested for multiple-choice positioning recognition. It seems that the system is not so sensitive to the position of the answers, and it can also be located and identified outside the brackets. Figure 3 shows the effect of an English test paper, the answers of the questions are close to each other. However, it does not affect the detection performance of the system. Figure 4 shows the performance of scribbled answers. It can be seen that no matter the degree of ambiguity, the detection still behaves well. Figure 5 is an example of the scoring of the answers. When scoring papers, the accurate answers will be shown on the left, and then students’ answers will be scanned for comparison. If the student’s answer is wrong, it will be displayed in red in the accurate answer.
Figure 5. Scoring of the answers.

4. Summary
The whole system is based on YOLO's target detection to accurately locate and recognize handwritten characters on the original test paper, from taking pictures to create the dataset, to recognizing handwritten characters anywhere on the entire test paper, and making some optimisations to solve the interference caused by students' corrections to the original answers, whether it's obvious scribbles or those that only make one or two horizontal, vertical and diagonal changes to the original answers. Both the more obvious scribbles and those with only one or two horizontal, vertical and diagonal strokes on the original answer are accurately identified and corrected. However, there are still areas of improvement in the application of the system that require further research and analysis in the future, including: (1) more convenient input methods, more accurate character and number segmentation areas, and a more efficient and accurate identification method. (2) A more intelligent and user-friendly system, so that users can get started more easily.

References
[1] Ciresan D C, Meier U, Gambardella L M, et al. 2011 Convolutional neural network committees for handwritten character classification International Conference on Document Analysis and Recognition (IEEE Computer Society) pp 1135-1139.
[2] Le C Y, Bengio Y and Hinton G 2015 Deep learning Nature 521 (7553) 436.
[3] Li X Y and Yuan H 2012 Research on several image scaling algorithms Modern Electronic Technology 35 (5) 48-51.
[4] Le C Y, Bottou L, Bengio Y, et al. 1998 Gradient-based learning applied to document recognition Proceedings of the IEEE 86 (11) 2278-2324.
[5] Xu K, Ba J, Kiros R, et al. 2015 Neural image caption generation with visual attention computer Show, Attend and Tell 53 (20) 2048-2057.
[6] Li Q 2008 Overview of binarization algorithms for document images Journal of Communication University of China (Natural Science Edition) 15 (4) 66-70.
[7] Krizhevsky A, et al. 2017 image net classification with deep convolutional neural networks Communications of the ACM 60 (6) 84-90.
[8] Bidyut B 2015 A survey of Hough Transform Pattern Recognition 48 (3) 993-1010.
[9] Goodfellow I, Bengio Y and Courville A 2016 Introduction to deep learning Deep Learning (MIT Press) p 192.
[10] Marieke L, Yevhen H and Riitta H 2012 What differs in visual recognition of handwritten vs printed letters? An fMRI study Human Brain Mapping 32 (8) 1250-1259.
[11] He K, Zhang X, Ren S, et al. 2016 Deep residual learning for image recognition Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition pp 770-778.
[12] Zeiler D and Fergus R 2014 Visualizing and understanding convolutional networks European Conference on Computer Vision pp 818-833.
[13] Srivastava N, Hinton G, Krizhevsky A, et al. 2014 Dropout: A simple way to prevent neural networks from over fitting Journal of Machine Learning Research 15 (1) 1929-1958.
[14] Wang P S, Song Y and Dai L R 2019 Fine grained image classification based on multi-channel visual attention Data Acquisition and Processing 34 (01) 157-166.

[15] Li Y, Ren L L 2019 Optimal design of intelligent marking system Optimal Design of Intelligent Marking System Based on Image Recognition Technology (China Computer and Communication) pp 2-3.

[16] Wang M and Li D X 2020 Based on improvement VGG-16 handwritten numeral recognition based on and naive Bayes Modern Electronic Technology 43 (12) 176-181.

[17] Chen D Y, Yin F and Wu R 2011 Perspective distortion correction method of scene text based on vanishing point Journal of Nanjing University of Technology/Journal of Nanjing University of Science and Technology 35 (4) 436-441.

[18] Guo L J, Sun S Y and Duan X S 2008 Research on support vector machine and kernel function Science, Technology and Engineering 8 (2) 487-490.

[19] Rajni R and Anutam A 2013 Image denoising techniques-An overview International Journal of Computer Applications 86 (16) 13-17.

[20] Jaddi N S, Abdullah S and Hamdan A R 2016 A solution representation of genetic algorithm for neural network weights and structure Information Processing Letters 116 (1) 22-25.