YL8C4Net: A Novel Algorithm for Target Source Detection and Classification in Astronomical Photometric Images (Postprint)

Authors: Chen-Ying Zhao, Liang-Ping Tu, Jian-Xi Li, Jia-Wei Miao, Geng-Qi Lin, Fang-Yuan Chen and Yang-Yang Liu

Date: 2025-09-28T12:32:54+00:00

Abstract

In the task of classifying large-scale astronomical data, accurate classification of galaxies, stars, and quasars typically relies on spectral labels. However, spectral data constitute only a small fraction of all astronomical observation data, and the classification information of target sources within vast photometric data has not been accurately determined. To address this issue, we propose a novel deep learning-based algorithm, YL8C4Net, for the automatic detection and classification of target sources in photometric images. This algorithm integrates the YOLOv8 detection network with the Conv4Net classification network. Additionally, we propose a novel magnitude-based labeling method for target source annotation. In performance evaluation, YOLOv8 achieves remarkable performance with average precision scores of 0.824 for AP@0.5 and 0.795 for AP@0.5:0.95. Meanwhile, the constructed Conv4Net attains an accuracy of 0.8895. Overall, YL8C4Net offers advantages including fewer parameters, faster processing speed, and higher classification accuracy, rendering it particularly suitable for large-scale data processing tasks. Furthermore, we employed the YL8C4Net model to conduct target source detection and classification on photometric images from 20 sky regions in SDSS-DR17. Consequently, a catalog containing approximately 9.39 million target source classification results has been preliminarily constructed, thereby providing valuable reference data for astronomical research.

Full Text

Preamble

Research in Astronomy and Astrophysics, 25:085019 (15pp), 2025 August

 https://doi.org/10.1088/1674-4527/ade492
CSTR: 32081.14.RAA.ade492

YL8C4Net: A Novel Algorithm for Target Source Detection and Classification in Astronomical Photometric Images

Chen-Ying Zhao[1], Liang-Ping Tu[1,2], Jian-Xi Li[2], Jia-Wei Miao[1], Geng-Qi Lin[1], Fang-Yuan Chen[1], and Yang-Yang Liu[1]

[1] School of Science, University of Science and Technology Liaoning, Anshan 114051, China; tuliangping@ustl.edu.cn

[2] School of Mathematics and Statistics, Minnan Normal University, Zhangzhou 363000, China

Received 2025 April 15; revised 2025 May 29; accepted 2025 June 10; published 2025 July 11

Abstract

In the task of classifying massive celestial data, the accurate classification of galaxies, stars, and quasars usually relies on spectral labels. However, spectral data account for only a small fraction of all astronomical observation data, and the target source classification information in vast photometric data has not been accurately measured.

To address this, we propose a novel deep learning-based algorithm, YL8C4Net, for the automatic detection and classification of target sources in photometric images. This algorithm combines the YOLOv8 detection network with the Conv4Net classification network. Additionally, we propose a novel magnitude-based labeling method for target source annotation. In the performance evaluation, the YOLOv8 achieves impressive performance with average precision scores of 0.824 for AP@0.5 and 0.795 for AP@0.5:0.95. Meanwhile, the constructed Conv4Net attains an accuracy of 0.8895. Overall, YL8C4Net offers the advantages of fewer parameters, faster processing speed, and higher classification accuracy, making it particularly suitable for large-scale data processing tasks.

Furthermore, we employed the YL8C4Net model to conduct target source detection and classification on photometric images from 20 sky regions in SDSS-DR17. As a result, a catalog containing about 9.39 million target source classification results has been preliminarily constructed, thereby providing valuable reference data for astronomical research.

Key words: techniques: image processing –methods: data analysis –techniques: photometric –catalogs

1. Introduction

With the continuous advancement of astronomical observation technology and the ongoing progress of large-scale survey projects, astronomical data are currently growing at an exponential rate. Effectively processing these vast amounts of data has become a critical challenge in the field of astronomy. In the vast Universe we observe, the number of celestial objects such as galaxies, stars, and quasars can reach billions. Accurate identification and classification of these celestial objects are crucial for astronomical research and serve as a fundamental prerequisite for subsequent morphological classification of galaxies.

Sloan Digital Sky Survey (SDSS; York et al. 2000) uses modern digital detectors for survey data acquisition, which record the sky with CCD cameras, and acquire spectra with two spectrographs guided by optical fibers. The survey provides hundreds of catalogs and is widely used for various research tasks. However, the classification labels of target sources in the PhotoObjAll catalog are not accurate, which divides sources into unknown sources (type = 0), galaxies (type = 3), and stars (type = 6), and does not classify quasars separately. It is well known that stars are point sources, galaxies are extended sources, and quasars are also point sources. Therefore, the separation method (Enard 1982), which relies solely on morphological features, cannot accurately distinguish between stars and quasars. More accurate source classification information is available in the SpecPhoto catalog, which classifies sources into galaxies, stars, and quasars. However, this classification requires that the sources have been spectrally calibrated. In comparison to photometric imaging, spectral observations are not only more time consuming, but also more expensive. Consequently, most sources lack spectral observations, preventing the accurate determination of their classifications.

Traditional target source detection methods primarily rely on techniques such as background estimation (Bertin & Arnouts 1996), threshold segmentation (Otsu 1979), and connected component analysis (Haralick & Shapiro 1985). However, the sensitivity of these methods to noise levels, parameter settings, and prior assumptions limits their applicability in modern astronomy. They exhibit significant limitations, particularly when dealing with complex backgrounds, blended targets, or low signal-to-noise ratio data. In the field of target source classification, although many researchers have applied traditional machine learning methods such as decision trees, random forests, and clustering to classify galaxies, stars, and quasars (Vasconcellos et al. 2011; McInnes et al. 2017), the performance of these methods is often unsatisfactory when coping with large-scale and complex astronomical classification tasks.

Convolutional Neural Networks (CNNs), as powerful deep learning models (LeCun et al. 2015), are capable of extracting implicit features from huge and high-dimensional image data, thereby providing robust support for the accurate detection and classification of target sources. Based on this, many researchers have adopted CNNs for target source localization, detection, and classification.

Stoppa et al. (2022) used the ASID-L algorithm to achieve fast localization of sources. Kim & Brunner (2016) successfully accomplished the classification of galaxies and stars using deep CNN. Li et al. (2023) employed a multi-scale convolutional capsule network for galaxy morphology classification. González et al. (2018) utilized the deep learning framework DARKNET and the real-time object detection system YOLO for galaxy detection and classification. Jia et al. (2020) successfully applied the Faster R-CNN object detection algorithm to detect and classify sources in astronomical images. Hausen & Robertson (2020) proposed the Morpheus framework based on deep learning, which performs pixel-wise source detection, segmentation, and morphological classification. Farias et al. (2020) successfully implemented galaxy detection, segmentation, and morphological classification using the Mask R-CNN framework. He et al. (2021) accomplished source detection and classification using photometric data. Shi et al. (2022) developed a photometric pipeline for SDSS image processing based on CNNs.

Although the above methods outperform traditional algorithms and yield promising results, their broader applicability is limited by the scarcity of annotated data. Researchers are often constrained to using publicly available annotated data sets or synthetic data as training sets, which restricts the widespread use of these methods. To overcome this limitation, we propose an innovative method that automates the annotation process by utilizing the magnitudes of target sources for data annotation. This method effectively alleviates the issue of insufficient annotated data in previous source detection tasks and reduces the reliance on synthetic data.

In this paper, YOLOv8 was used for target source detection. Due to the lack of accurate category labels in the PhotoObjAll catalog, YOLOv8 performed only the detection task without classification. Subsequently, we reconstructed a spectrally calibrated target source data set and used Conv4Net to complete the source classification task.

The paper is organized as follows: Section 2 details the data used in this study and the data processing methods. Section 3 describes the architectural design of the YOLOv8 detection network and the Conv4Net classification network, along with their training strategies. Section 4 shows the experimental results, comparing and analyzing the performance of the proposed models against other methods, and completes the construction of the catalog. Finally, Section 5 summarizes the research results of this work and looks forward to future research directions.

2. Data and Data Processing

2.1. Source Detection Data

2.1.1. Data Access The source detection data were accessed from the photometric images of the 3723 sky region in Sloan Digital Sky Survey Data Release 17 (SDSS-DR17) (Abdurro' uf et al. 2022). This region contains 1364 photometric images, of which 227 were randomly selected for the validation set and the remaining 1137 for the training set.

2.1.2. Photometric Image Partitioning The CCD photometric system (Schechter et al. 1993) of SDSS employs six sets of CCDs to simultaneously measure celestial objects across five bands (u, g, r, i, and z). These multi-band data have been extensively used in various astronomical studies. Due to the higher noise levels in the u and z bands and their limited contribution to this study, we selected the g, r, and i bands and used the officially synthesized RGB images as the data set. Since the RGB images are synthesized using the r-band as the baseline band, we used the FITS files of the r-band to correspond with the synthesized RGB images and proceeded with subsequent data processing.

The size of the photometric image is 2048 × 1489 pixels. It has been experimentally verified that directly inputting images of this size into the model leads to poor detection results. To optimize the training performance of the target detection model and prevent the loss of edge targets, we implemented an overlapping partitioning strategy. Each photometric image was divided into 12 smaller photometric images, each with a size of 640 × 640 pixels. Figure 1 [Figure 1: see original paper] shows the schematic of this partitioning process. Although this partitioning strategy will produce some duplicate detection results, we introduce the Non-Maximum Suppression algorithm as a post-processing step to merge the detection results of the overlapping areas (Van Etten 2018).

2.1.3. Label Creation Since a supervised deep learning model is employed, a fully annotated data set is required for model training. Given the large scale of the data set, manual annotation is obviously impractical. This highlights the urgent need for an efficient method capable of automatically annotating the target sources.

In general, a target source with a smaller magnitude corresponds to a higher brightness and a relatively larger scale. Thus, we propose a new method for annotating target sources based on their magnitudes. This method can adaptively adjust the annotation range when handling targets with varying sizes and brightness levels. Subsequent experimental results demonstrate that this innovative annotation method is both highly efficient and effective, making it well-suited for annotating large-scale astronomical photometric data. Moreover, it addresses the challenges of automatic source cropping and scale adaptation in wide-field surveys to a certain extent.

First, we queried the photometric data of the entire 3723 sky region through the CasJobs server, and extracted the R.A., decl., and r-band magnitude information of all target sources. Then, based on the World Coordinate System (WCS) metadata in the r-band FITS file, the R.A. and decl. were converted into pixel coordinates in the image, and the bounding box was determined according to the r-band magnitude of the target source. The limiting magnitudes for the u, g, r, i, and z bands in the SDSS are 22.0, 22.2, 22.2, 21.3, and 20.5, respectively.

Sources with magnitudes below the limiting magnitudes are too faint to be detected effectively. Therefore, we used the limiting magnitude of the r-band as the threshold to exclude target sources exceeding this limit. After several experiments and analyses, the optimal bounding box sizes, as shown in Table 1 , were determined. If a bounding box extends beyond the image boundary, it is cropped to align with the image edges.

2.1.4. Source Pasting During the generation of the PhotoObjAll catalog, SDSS applies multiple layers of filtering and screening to exclude low-quality, noise-contaminated, or incomplete observational records. Furthermore, SDSS opts not to record certain target sources that have been extensively analyzed in other research. Therefore, the SDSS PhotoObjAll catalog does not cover all target sources in the photometric images. In the 3723 sky region data we used, certain regions contain a large number of target sources that are not recorded in the PhotoObjAll catalog. Figure 2 [Figure 2: see original paper] provides an example, showing that the number of target sources annotated based on the PhotoObjAll catalog is significantly lower than the actual number present. Consequently, training on data sets annotated using the aforementioned method may impact model performance, as some target sources remain unannotated.

To solve this problem, we first generate background images, then crop the annotated target sources from the images and paste them back into their original positions on the background images. For generating background images, we explore several methods. One method involves selecting regions with a "type" parameter of 8 (indicating sky regions) from the PhotoObjAll catalog to create the background images. However, this method has a limitation: the size of the cropped sky regions must be neither too large nor too small. If the size is too large, it may include target sources, while if it is too small, it requires cropping a substantial amount of sky background, increasing the workload. Therefore, although the method is theoretically feasible, it is not the optimal choice in practice.

Ultimately, we used the Photutils toolkit in Python to generate background images that are as consistent as possible with the background of the original image. Photutils is a widely used tool for astronomical image analysis, capable of detecting astronomical sources and performing semantic segmentation. Based on its background level assessment principles, we set a very low detection threshold (2σ) and a very small connected region (3 pixels) to ensure that all target sources can be accurately detected. Next, the detected sources were masked out from the photometric images to generate mask images containing only the sky background. The mean pixel value of the background region was then calculated and used to fill the masked areas, resulting in a complete background image. Finally, the cropped sources were pasted back into their original positions on the background image.

The entire processing workflow is shown in Figure 3 [Figure 3: see original paper]. Panel (a) of Figure 3 shows the original photometric image with a size of 640 × 640, while panel (b) presents the masked image generated after processing. Panel (c) illustrates the background image obtained by filling the masked regions with the mean pixel values, and panel (d) depicts the image used for model training, created by pasting the target sources back onto the background based on their original positions. This image excludes sources that exceed the limiting magnitude in the r-band as well as those unrecorded in the PhotoObjAll catalog.

Despite the above data processing process, there are still some inaccuracies in this labeling method for some large-scale extended galaxies. In this context, defining the bounding box becomes more complex: if the bounding box is too large, it may include other target sources, compromising accurate identification; if it is too small, it may fail to fully encompass the entire galaxy. To further refine the process, we used the LabelImg tool to manually adjust the inaccurate bounding box. Figure 4 [Figure 4: see original paper] shows the final annotation results.

2.2. Source Classification Data

2.2.1. Data Access The source classification data were accessed from the photometric images of 12 sky regions in SDSS-DR17, including regions such as 3697, 3698, and 3699. We used the CasJobs server to perform SQL queries for cross-matching the PhotoObjAll and SpecPhoto catalogs across these 12 sky regions, and extracted sources with anomalous redshifts (zWarning $\neq 0$) and those exceeding the r-band limiting magnitude were excluded.

Subsequently, following the method outlined in Section 2.1.3, the target sources were cropped from the photometric images based on their magnitude, resulting in a total of 220,599 spectrally calibrated target sources. Given the extreme imbalance in the data set, with galaxies substantially exceeding stars and quasars, a total of 35,280 galaxies, 34,828 quasars, and 34,183 stars were ultimately selected at random. The data set was then partitioned into training, validation, and test sets in an 8:1:1 ratio.

Furthermore, we accessed all photometric images from 20 sky regions, such as 3829, 3830, and 3835, for source detection and classification. Based on these processes, a catalog of the classification results was constructed.

2.2.2. Data Processing Due to the inconsistency in the sizes of target sources within the classification data set, we employed a bilinear interpolation function (Gribbon & Bailey 2004) to scale them to a uniform size. Following several experiments, we determined that the optimal classification performance was achieved when target sources were scaled to 48 × 48 pixels. To further enhance the model' s performance and robustness, we implemented data augmentation techniques to increase data set diversity. These included randomly flipping images horizontally or vertically with a 50% probability and rotating images randomly within a range of 0°–180°. Finally, Min–Max normalization and Z-score standardization were applied to accelerate model convergence.

3. Methods

In this paper, multiple target detection networks were employed for target source detection, and various classification networks were used to classify the detected sources. After comparative analysis, the YOLOv8 network was selected for source detection, and the Conv4Net network was chosen for source classification. This section provides a detailed description of the performance and application of these two networks.

3.1. Source Detection Network

The source detection task addressed in this study pertains to small target detection, as the target itself lacks prominent features, resulting in limited available information. Additionally, some features may be lost during the pooling stage of feature extraction, and the complex sky background creates further interference, culminating in a significant challenge for this detection task. YOLOv8 is among the most advanced networks in target detection (Terven et al. 2023), demonstrating exceptional performance, particularly in small target detection. Its network architecture is shown in Figure 5 [Figure 5: see original paper].

The YOLOv8 network consists of three main components: the Backbone, Neck, and Head. The Backbone layer is responsible for feature extraction from the input image, where the introduction of the C2f module enhances network efficiency, significantly accelerating detection speed. The Neck layer adopts a PANet structure for multi-scale feature fusion, integrating Feature Pyramid Networks and Path Aggregation Networks to achieve comprehensive fusion of both upper and lower feature information. The Head layer utilizes a decoupled detection head, separating the classification and regression tasks into two independent branches, thereby enhancing the model' s generalization capability. Additionally, YOLOv8 adopts an anchor-free design, reducing reliance on prior information while enabling more accurate predictions of target location and size.

Meanwhile, YOLOv8 integrates various data augmentation techniques, including Mosaic, Mixup, Random Perspective, and HSV augmentation. These augmentation methods increase the diversity of the image data, enrich the detection background, and enhance the model' s adaptability to complex scenarios. In deep learning, the configuration of hyperparameters significantly impacts the model' s training performance (Bengio 2012). We predefined a set of hyperparameters, as detailed in Table 2 , where lr0 represents the initial learning rate, lrf denotes the learning rate decay factor, and the final learning rate (lr) is calculated using the formula: $\mathrm{lr} = \mathrm{lr0} \times \mathrm{lrf}$.

For the YOLOv8 model, the training epoch was set to 200, and an early stopping strategy was implemented. Specifically, if the model does not demonstrate performance improvement for 50 consecutive epochs, training is halted. This early stopping strategy aims to mitigate the risk of overfitting while balancing model performance and training time. Consequently, the actual number of training epochs for the YOLOv8 model was 156. We recorded the loss variation of the YOLOv8 model on both the training and validation sets, as shown in Figure 6 [Figure 6: see original paper]. Figure 6 illustrates that the model' s loss decreased progressively during training, with a similar downward trend observed on the validation set. This indicates the model gradually converges during training and ultimately reaches a stable state.

3.2. Source Classification Network

In astronomical research, accurately classifying celestial targets such as galaxies, stars, and quasars based solely on photometric images presents a significant challenge. With the advancement of deep learning techniques, deep neural networks can automatically learn features and extract useful information from image data, facilitating precise classification of these targets.

In this paper, the Conv4Net network is employed for the classification of galaxies, stars, and quasars. Conv4Net is an efficient and streamlined CNN with excellent feature extraction and classification capabilities. Its architecture comprises four convolutional layers, four pooling layers, and three fully connected layers, with a CBAM module introduced after the second convolutional layer to enhance feature representation. A ReLU activation function is applied after each convolutional layer, followed by a max-pooling layer. A detailed description of the network architecture is provided in Figure 7 [Figure 7: see original paper] and Table 3 .

Although the Conv4Net network is comparable to traditional CNNs in terms of classification accuracy (see Section 4.2.2 for detailed experimental comparison results), it has fewer parameters and higher Frames Per Second (FPS). From the perspectives of computational efficiency and resource utilization, Conv4Net demonstrates superior performance. Its advantages lie not only in maintaining high classification accuracy but also in requiring fewer computational resources and faster execution for the same tasks. This characteristic is particularly crucial for large-scale data processing, real-time applications, and resource-constrained environments.

Figure 8 [Figure 8: see original paper] illustrates the changes in accuracy and loss values of the Conv4Net network on both the training and validation sets. Notably, as the number of training epochs increases, the classification accuracy progressively improves, while the loss continues to decrease. This indicates the network effectively fits the training data during the learning process, without exhibiting signs of overfitting or underfitting.

4. Experiments and Results

4.1. Experimental Setup

4.1.1. Experimental Environment The experiments in this study were conducted on a server equipped with NVIDIA A100 GPUs. All programs used in the experiments were written in Python and implemented using the PyTorch deep learning framework.

4.1.2. Performance Evaluation Metrics To comprehensively and objectively evaluate the model' s performance, we calculated a series of metrics, including accuracy, precision, recall, F1 score, mAP@0.5, and mAP@0.5:0.95. The formulas for these metrics are presented below.

The mean Average Precision (mAP) is a comprehensive evaluation metric for object detection models (Everingham et al. 2010), which integrates precision and recall by calculating the average precision (AP) across various categories to assess the overall performance of the model. In the object detection task, Intersection over Union (IoU) is a critical metric used to quantify the overlap between the predicted bounding box and the ground truth bounding box. It is calculated as the area of the intersection of the predicted and ground truth bounding boxes divided by the area of their union. mAP@0.5 indicates that at an IoU threshold of 0.5, the AP for each category is calculated, and the AP values of all categories are subsequently averaged. The mAP@0.5:0.95 indicates the average performance of the model evaluated at multiple IoU thresholds (from 0.5 to 0.95 in 0.05 step increments), providing a more granular performance evaluation. As the detection task in this study does not differentiate between target categories, the mAP is equivalent to the AP.

4.2. Model Results and Comparative Analysis

4.2.1. Source Detection The YOLOv8 model is trained on the source detection data set constructed in Section 2.1, achieving a precision of 0.993, a recall of 0.690, an F1 score of 0.814, an AP@0.5 of 0.824, and an AP@0.5:0.95 of 0.795 on the validation set (see Table 4 for details. In this paper, the bolded values in the tables represent the best performance among the models under the same evaluation metric.). YOLOv8 achieves a high score on the AP@0.5:0.95 metric, indicating that the model possesses excellent target detection capabilities and accurately predicts the bounding box positions of targets, thereby effectively aiding in the separation of blended sources. Figure 9 [Figure 9: see original paper] shows the training curves of AP@0.5 and AP@0.5:0.95 as epochs vary.

We conducted an in-depth analysis of the reasons behind the model' s low recall rate. We generated heatmaps of validation set images using the weights from the best epoch during training (see Figure 10 [Figure 10: see original paper]). The heatmap illustrates the model' s target detection process within the image. The colors indicate the model' s confidence or attention to specific regions, with warm tones representing areas deemed highly relevant or important, while cool tones correspond to regions of lower relevance or significance. When multiple target sources are in close proximity or undergo blending, their representations on the heatmap become superimposed, resulting in deeper colors (as shown in the yellow regions of Figure 10), which typically indicate a higher data density in that region.

As can be seen from Figure 10, although the majority of target sources can be detected successfully, some target sources with low brightness and small pixel size still miss detection. We plotted the relationship between different magnitude ranges and the missed detection rate (see Figure 11 [Figure 11: see original paper]). The results show that as the magnitude increases (i.e., brightness decreases), the missed detection rate rises significantly. Generally, targets with larger magnitudes and lower brightness tend to have smaller sizes, making them more prone to missed detections. Our further analysis found that another reason for the low recall rate is that the information of some target sources in the g, r and i bands is sparse or even completely missing, and the information of these target sources is mainly concentrated in the u and z bands. Because the information of u and z bands is not fully utilized, some target sources are not successfully detected.

For the detection task in this paper, the proportion of some target sources in the image is very small, occupying only a few pixels, and it is even difficult to distinguish them as noise points or actual target sources visually. Because the characteristic information of such a target source is very limited, even if the model can detect it, it is often accompanied by a high classification error rate in the subsequent classification process. These misclassification results not only have no practical significance for astronomical research, but may cause interference. In addition, too much attention to these difficult-to-distinguish small target sources may disperse the computational resources of the model, affecting the detection and classification performance of the key target sources. Therefore, the omission of such small target sources is acceptable and will not have a substantial impact on the overall results of the study.

We compared YOLOv8 with several representative target detection networks, including Faster R-CNN (Ren et al. 2017), YOLOv4, and YOLOv5. The results for precision, recall, F1 score, AP@0.5, and AP@0.5:0.95 were obtained and summarized in Table 4 .

As shown in Table 4, YOLOv8 demonstrates superior performance in the target detection task of this study. Compared with other models, YOLOv8 not only achieves higher detection accuracy but also exhibits faster computational speed and lower memory consumption. Despite our efforts to incorporate an additional layer specifically designed for small target detection to further optimize the model, the experimental results indicate that this modification does not yield the anticipated performance improvements. Instead, it significantly increases computational and memory consumption (see Section 4.3 for details).

4.2.2. Source Classification The classification accuracy of the Conv4Net model for galaxies, stars, and quasars on the test set is 0.8895. The average precision, recall, F1 score, and AUC value are 0.8932, 0.8887, 0.8887, and 0.9689, respectively (see Table 5). Detailed classification performance for each category is described in Table 6 .

Figure 12 [Figure 12: see original paper] shows the confusion matrix and ROC curve for Conv4Net on the test set. The results indicate that the model excels in galaxy classification, achieving an AUC value of 0.9922, which suggests that the majority of galaxies can be accurately identified. However, the number of misclassifications for quasars and stars is significantly higher compared to galaxies, particularly with quasars being frequently misclassified as stars and vice versa. From a morphological perspective, galaxies generally appear as extended surface sources, whereas quasars and stars resemble point sources, making them more susceptible to confusion in image classification.

Table 5 provides a detailed comparison of the performance of the Conv4Net network with other prominent neural networks. These networks represent a variety of types and architectures in the field of deep learning, including traditional CNNs (such as AlexNet (Krizhevsky et al. 2012), VGG11, VGG19 (Simonyan & Zisserman 2014), GoogLeNet (Szegedy et al. 2015), RegNet (Radosavovic et al. 2020), and ConvNeXt (Liu et al. 2022)), residual networks (ResNet34 (He et al. 2016)), the EfficientNet-B0 model from the EfficientNet series (Tan & Le 2019), which is well-suited for small-scale image classification, and Transformer-based vision models (Vision Transformer (Dosovitskiy et al. 2021)). In addition, it also includes the new models TinyViT (Wu et al. 2022) and ConvNeXt V2 (Woo et al. 2023), which have been optimized for the network structure in recent years, as well as the TSCNet (Shi et al. 2022) network proposed by Shi et al. for the classification of small-scale target sources. By comparing these diverse types of networks, we aim to enhance the reliability of the comparison and thereby increase the persuasiveness of the results. When these networks are applied to classification tasks, the results indicate no significant differences in classification performance, with the networks showing relatively similar effects.

We further compare the total number of parameters and FPS for these models, as shown in Figure 13 [Figure 13: see original paper]. The Conv4Net has the fewest parameters and the highest FPS among all the models compared. The total number of parameters reflects the model's scale and complexity, which is directly related to its ability to represent features and fit complex data distributions, while also significantly affecting computational and storage costs. FPS serves as a metric for the model's inference efficiency, with higher FPS values indicating faster inference speed and greater efficiency.

The Conv4Net network contains 1,635,749 parameters and achieves approximately 4239 FPS. Compared to AlexNet, which has a very simple structure, the number of parameters in Conv4Net decreased by 46.26% and the FPS increased by 11.00%. Therefore, Conv4Net demonstrates competitive performance compared to other prominent networks and exhibits high inference efficiency, making it particularly well-suited for the rapid processing and classification of large-scale data in this study.

4.3. Ablation Experiment

The source detection task in this study mainly focuses on small target detection. Given the minimal pixel proportion occupied by small targets, they are challenging to detect. To enhance the model' s ability to detect small targets, we introduce a 160 × 160 small target detection head in the Head layer, enabling the model to more effectively capture the characteristic features of small targets. The network, following the addition of this detection head, is designated YOLOv8-ST (with ST standing for Small Targets), and its network architecture is illustrated in Figure 14 [Figure 14: see original paper].

We utilized the YOLOv8-ST model for source detection, with the batch size set to 32, consistent with the YOLOv8 configuration. The model' s performance on the validation set is summarized in Table 7 : precision is 0.994, recall is 0.691, F1 score is 0.815, AP@0.5 is 0.824, and AP@0.5:0.95 is 0.808.

As shown in Table 7, the detection performance of the YOLOv8-ST model does not exhibit a significant improvement, with the AP@0.5:0.95 score increasing by only 1.3% and other metrics remaining almost unchanged. However, with the addition of the detection head, the Giga Floating-point Operations Per Second (GFLOPs) of YOLOv8-ST increased by 25.60% compared to YOLOv8, and memory consumption increased by more than twice. This results in a substantial decrease in inference speed during practical applications, particularly on resource-constrained devices.

To balance the trade-off between performance improvement and computational resource consumption, we ultimately decide to use the original YOLOv8 model for the source detection task. Future work will focus on exploring more optimized model architectures, aiming to maintain high detection performance while minimizing computational resource usage.

4.4. Catalog Construction

In this study, the trained YL8C4Net model is employed for source detection and classification on photometric images from 20 sky regions. The detection results are cross-validated with the PhotoObjAll catalog, and a catalog containing about 9.39 million target source classification results is initially constructed.

Due to an approximate 10% overlap between adjacent photometric images, this may lead to the repeated detection of the same target sources. To eliminate this redundancy, this study stipulates that two sources are considered to be the same target source when the distance between their centers is less than 2 pixels. We converted the center pixel coordinates of detected target sources to R.A. and decl. using the WCS metadata from the r-band FITS files. Subsequently, we calculated the angular distance between any two target sources based on their

R.A. and decl., and removed any sources with an angular distance of less than 0.5 arcseconds to eliminate duplicate detection.

After removing duplicate detections, we ultimately detected approximately 9.41 million target sources. Employing a nearest neighbor search algorithm with a query radius of 2 arcseconds, we cross-validated the detected sources against the PhotoObjAll catalog, successfully matching approximately 9.39 million sources. This discrepancy suggests that our model detected roughly 20,000 additional sources beyond those recorded in the PhotoObjAll catalog, potentially identifying previously unrecorded sources. However, it is important to note that we cannot definitively confirm the authenticity of all these additional detections, as some may be false positives resulting from high-noise backgrounds. To ensure the reliability of the final classification results, we retained only the 9.39 million sources matched to the PhotoObjAll catalog.

These matched sources were then integrated with positional data, and detailed predicted classifications and magnitude information were extracted from the PhotoObjAll catalog to construct the final catalog. The catalog is available at the following URL: https://nadc.china-vo.org/res/r101453/.

In current practical applications, the limitation of spectral labels poses an important challenge to the verification of classification results. It should be pointed out that this catalog is only a preliminary result. Due to the differences in observation conditions and the characteristics of each sky region itself, there may be significant differences in the background noise characteristics and the morphology of the target source in different photometric images. These differences will affect the generalization ability of the object detection and classification model, causing the performance of the model in practical applications to possibly deviate from the training stage. The reliability of the results of this study will be further evaluated in subsequent work. To enhance the scientific value and credibility of the research, we sincerely invite experts and scholars in related fields to participate in the data verification work. Due to time constraints, this study has only carried out target source detection and classification on photometric images of 20 sky regions. We encourage interested researchers to refer to the method proposed in this paper and carry out target source detection and classification studies on more sky regions, so as to further validate the universality of the method and promote the in-depth development of the related fields.

Figure 15 [Figure 15: see original paper] shows an example of detection and classification prediction, with red boxes for galaxies, blue boxes for stars, and green boxes for quasars. As shown in Figure 15, the YOLOv8 model successfully detects the vast majority of target sources and demonstrates strong detection capability, particularly when handling blended target sources. Figure 16 [Figure 16: see original paper] provides a zoomed-in view of the detection results for blended target sources.

5. Conclusions and Future Work

In this paper, the YL8C4Net algorithm was employed to detect and classify target sources, including galaxies, stars, and quasars, based on photometric images.

The YOLOv8 target detection framework was employed to achieve automatic detection of target sources. Given the necessity of manually generating labels for the detection data set, an innovative annotation method based on the target source' s magnitude is introduced. This method not only provides accurate and efficient labeling but also addresses the challenges of automatic cropping and scale adaptation in wide-field surveys. The YOLOv8 network performs well on the validation set, achieving AP@0.5 and AP@0.5:0.95 scores of 0.824 and 0.795, respectively. Additionally, a Conv4Net network, which excels in source classification tasks, was designed, achieving a classification accuracy of 0.8895 on the test set. The network comprises four convolutional layers, four pooling layers, three fully connected layers, and a CBAM module, offering a simple yet computationally efficient architecture. Compared with more than ten other network types, Conv4Net demonstrates the highest classification accuracy, with the fewest parameters and fastest computation speed, making it ideal for large-scale data processing.

We further performed source detection and classification on photometric images from 20 sky regions in the SDSS, and constructed a preliminary catalog containing classification results. The catalog provides classification information for target sources that have not yet been spectrally calibrated, to some extent compensating for the inaccurate "type" label in the PhotoObjAll catalog. In future work, we plan to further evaluate the accuracy of the catalog' s classification results, in order to provide a valuable data resource for astronomical research.

Future research will be dedicated to constructing a simulated photometric image data set with high authenticity, strong diversity, and precise labeling to support the development and application of a one-stage joint detection-classification model. Specifically, we plan to extract spectrally calibrated target sources of galaxies, stars, and quasars from different sky regions, and based on their magnitude information, we will use the annotation method described in Section 2.1.3 to annotate and crop the target sources. Subsequently, the cropped target sources are pasted onto the photometric background images generated in Section 2.1.4, and the Poisson blending algorithm is applied to reduce edge artifacts and enhance the naturalness of the resulting images.

During the pasting process, we will comprehensively consider the actual distribution characteristics of source in astronomical observations to achieve random position placement and density control of the target source in the image, thereby more realistically simulating the distribution state of the source in the photometric image. In addition, to enhance the diversity and representativeness of the data set, we will simulate background images from multiple sky regions, covering a wide range of observational conditions and noise environments. This approach allows for a more comprehensive representation of typical sky region characteristics. The large-scale synthetic images generated through the above process will be constructed into a custom data set with clear category labels. This data set is expected to significantly improve the model' s generalization ability under complex backgrounds and partially mitigate the issue of unreliable classification results caused by limited spectral labels.

Building on this foundation, we will further explore a one-stage joint detection and classification approach based on object detection algorithms. This method enables simultaneous localization and identification of celestial objects within a single framework, significantly improving data processing efficiency. It is particularly well-suited for the rapid analysis of large-scale astronomical image data. We believe that this approach will provide strong support for astronomical research.

Acknowledgments

We sincerely acknowledge the Sloan Digital Sky Survey project for making their open-access data available, which has been instrumental in advancing our research. The data utilized in this paper are publicly available on the websites: https://data.sdss.org/sas/dr17/eboss/photoObj/frames/301 and http://skyserver.sdss.org/CasJobs.

This work was supported by the National Natural Science Foundation of China (NSFC, Grant No. U1731128).

References

Abdurro' uf, Accetta, K., Aerts, C., et al. 2022, ApJS, 259, 35 Bengio, Y. 2012, Practical Recommendations for Gradient-based Training of Deep Architectures, Neural Networks: Tricks of the Trade (2nd ed.; Berlin: Springer), 437 Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393 Dosovitskiy, A., Beyer, L., Kolesnikov, A., et al. 2021, in Int. Conf. on Learning Representations Enard, D. 1982, Proc. SPIE, 331, 232 Everingham, M., Gool, L. V., Williams, C. K. I., Winn, J. M., & Zisserman, A. 2010, International Journal of Computer Vision, 88, 303 Farias, H., Ortiz, D., Damke, G., Jaque Arancibia, M., & Solar, M. 2020, A&C, 33, 100420 González, R., Muñoz, R., & Hernández, C. 2018, A&C, 25, 103 Gribbon, K., & Bailey, D. 2004, in Second IEEE Int. Workshop on Electronic Design, Test and Applications (Piscataway, NJ: IEEE), 126 Haralick, R. M., & Shapiro, L. G. 1985, CVGIP, 29, 100 Hausen, R., & Robertson, B. E. 2020, ApJS, 248, 20 He, K., Zhang, X., Ren, S., & Sun, J. 2016, in 2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) (Piscataway, NJ: IEEE), 770 He, Z., Qiu, B., Luo, A.-L., et al. 2021, MNRAS, 508, 2039 Jia, P., Liu, Q., & Sun, Y. 2020, AJ, 159, 212 Kim, E. J., & Brunner, R. J. 2016, MNRAS, 464, 4463 Krizhevsky, A., Sutskever, I., & Hinton, G. 2012, in Neural Information Processing Systems (Red Hook, NY: Curran Associates), 25 LeCun, Y., Bengio, Y., & Hinton, G. 2015, Natur, 521, 436 Li, G., Xu, T., Li, L., et al. 2023, MNRAS, 523, 488 Liu, Z., Mao, H., Wu, C.-Y., et al. 2022, in 2022 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR) (Piscataway, NJ: IEEE), 11966 McInnes, L., Healy, J., & Astels, S. 2017, JOSS, 2, 205 Otsu, N. 1979, ITSMC, 9, 62 Radosavovic, I., Kosaraju, R. P., Girshick, R. B., He, K., & Dollár, P. 2020, in 2020 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR) (Piscataway, NJ: IEEE), 10425 Ren, S., He, K., Girshick, R., & Sun, J. 2017, ITPAM, 39, 1137 Schechter, P. L., Mateo, M., & Saha, A. 1993, PASP, 105, 1342 Shi, J.-H., Qiu, B., Luo, A.-L., et al. 2022, MNRAS, 516, 264 Simonyan, K., & Zisserman, A. 2015, in Int. Conf. on Learning Representations Stoppa, F., Vreeswijk, P., Bloemen, S., et al. 2022, A&A, 662, A109 Szegedy, C., Liu, W., Jia, Y., et al. 2015, in 2015 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) (Piscataway, NJ: IEEE), 1 Tan, M., & Le, Q. 2019, in Proc. Machine Learning Research, Vol. 97, Proc. Int. Conf. on Machine Learning, ed. K. Chaudhuri & R. Salakhutdinov (PMLR), 6105 Terven, J., Córdova-Esparza, D.-M., & Romero-González, J.-A. 2023, Machine Learning and Knowledge Extraction, 5, 1680 Van Etten, A. 2018, arXiv:1805.09512 Vasconcellos, E. C., de Carvalho, R. R., Gal, R. R., et al. 2011, AJ, 141, 189 Woo, S., Debnath, S., Hu, R., et al. 2023, in Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (Piscataway, NJ: IEEE), 16133 Wu, K., Zhang, J., Peng, H., et al. 2022, in European Conf. on Computer Vision (Berlin: Springer), 68 York, D. G., Adelman, J., John, E., Anderson, J., et al. 2000, AJ, 120, 1579

Note: Figure translations are in progress. See original paper for figures.