Astronomy and Computing 42 (2023) 100687

Contents lists available at ScienceDirect

Astronomy and Computing

journal homepage: www.elsevier.com/locate/ascom

Full length article

A deep learning based astronomical target detection framework for multi-colour photometry sky survey projects[☆]

P. Jia[a,b,c,*], Y. Zheng[a], M. Wang[a], Z. Yang[a]

[a] College of Physics and Optoelectronics, Taiyuan University of Technology, Taiyuan, 030012, China
[b] Peng Cheng Lab, Shenzhen, 518066, China
[c] Department of Physics, Durham University, South Road, Durham, 695571, DH1 3LE, UK

ARTICLE INFO

Article history:
Received 21 November 2022
Accepted 11 January 2023
Available online 18 January 2023

Dataset link: https://nadc.china-vo.org/res/r101190/

Keywords:
Data analysis
Image processing
Surveys
Target detection
Deep neural network

ABSTRACT

Multi-colour photometry sky survey projects would obtain celestial object images of different colours with a wide field telescope and several different filters. Images of different colours could reveal different components of celestial objects. We would be able to investigate details of celestial objects and even discover many new celestial objects with these images. Nevertheless, celestial objects are randomly and sparsely distributed in observational images and we need to detect them from these images before we could carry out further study. According to properties of multi-colour images, we propose a novel deep learning based astronomical target detection framework. Our framework could adapt to images with any number of colours and directly detect celestial objects from these images. For extended targets, such as galaxies, our framework could detect and segment their images; for point like targets, such as stars or quasars, our framework could detect and locate their images. From practical application aspects, we have tested the performance of our framework by building a pipeline to process SDSS images with one, three and five colours. Results show that our framework could provide reliable detection results. Our framework could be used as a basic block to build pipelines for multi-colour photometry sky survey projects.

1. Introduction

Multi-colour photometry sky survey projects have attracted a lot of attentions, which could obtain celestial object images of several different colours. Since images of different colours could reveal different components of celestial objects, multi-colour images could be used to better study properties of celestial objects and possibly lead to some new discoveries, such as dark nebulae, green bean galaxies or blue pea galaxies. In recent years, many multi-colour photometry sky survey projects have been proposed, such as the SDSS (Accetta et al., 2022), the SkyMapper (Keller et al., 2007), the J-PAS (Benitez et al., 2014), the DES (Flaugher et al., 2015), the LSST (Ivezić et al., 2019), the Mephisto Yuan et al. (2020), the WFST (Lin et al., 2022), the Sitian (Liu et al., 2021b) and the CSST (Gong et al., 2019). These projects have already obtained or will obtain huge amount of observational images. Because celestial objects are randomly and sparsely distributed in observational images, detection of celestial objects from observational images is the first and the most important step. However, contemporary and future multi-colour sky survey projects would obtain huge amount of observational images and celestial objects have extremely different size and shapes, it would be hard to detect different celestial objects from observational images in a straight-forward way.

Traditionally, astronomers use source extraction algorithms, such as the SExtractor (Bertin and Arnouts, 1996) to obtain candidates of celestial objects from observational images. Then astronomers would classify these targets to different types either through manual inspections (Bom et al., 2017; Ignat and Bodeic, 2019; He et al., 2020; Xiang et al., 2022; Kim et al., 2020) or machine learning algorithms (De La Calleja and Fuentes, 2004; Tachibana and Miller, 2018; Duev et al., 2019; Xi et al., 2020; Sun et al., 2022; Jacquemont et al., 2021). However, scientists normally require to select images in a particular band for the source extraction algorithm to obtain candidates, which would bring difficulties in detection of some specific celestial objects, because some celestial objects have features that could only be obtained through comparing images of different bands, such as

[☆] This document is the results of the research funded by National Natural Science Foundation of China (NSFC) with funding number of 12173027 and 12173062, the science research grants from the China Manned Space Project with NO. CMS-CSST-2021-A01, the science research grants from the Square Kilometre Array (SKA) Project with NO. 2020SKA0110102.

[*] Corresponding author at: College of Physics and Optoelectronics, Taiyuan University of Technology, Taiyuan, 030012, China.
E-mail address: robinmartin20@gmail.com (P. Jia).
URL: http://aojp.lamost.org (P. Jia).

https://doi.org/10.1016/j.ascom.2023.100687
2213-1337/© 2023 Elsevier B.V. All rights reserved.

the emission nebulae. Therefore, end to end target detection algorithms, which could detect celestial objects directly from multi-colour photometry sky survey images are required for data processing pipelines.

In recent years, deep neural networks have been proposed for general purpose target detection tasks (Ren et al., 2015; Redmon and Farhadi, 2018; Carion et al., 2020; Zhu et al., 2020). Thanks to the development of the GPU technology and the huge amount of labelled data, deep neural network based target detection algorithms have achieved remarkable performance. Based on the structure of these deep neural networks, many detection algorithms have been proposed to process astronomical images. For general purpose sky survey projects, Astro R-CNN (Burke et al., 2019) uses R-CNN based neural network (Girshick et al., 2014) to detect, classify and segment stars and galaxies from astronomical images. Jia et al. (2020) proposed a detection and classification algorithm based on Faster R-CNN (Ren et al., 2015) for detection of stars and moving celestial objects from images obtained by wide-field small-aperture telescopes. Mask galaxy (Farias et al., 2020), an algorithm based on the Mask R-CNN (He et al., 2017) has been proposed for detection, segmentation and morphological classification of spiral or elliptical galaxies. Based on the YOLO (Redmon et al., 2016), APSCnet (He et al., 2021) is proposed to detect quasars, stars and galaxies from observational images.

These algorithms are directly used to process astronomical images. However, it should be noted that images obtained by multi-colour photometry sky survey projects have some unique properties, which make it inadequate to be directly processed by algorithms adopted from general purpose target detection neural networks.

1. The size of different celestial object images could be extremely different. For example, galaxies could be as large as several arcmin, while stars or quasars with moderate brightness for ground-based multi-colour sky survey projects would be only several arcsec.
2. Multi-colour sky survey projects would obtain celestial object images with many different bands, which would be much more than ordinary natural images with only three channels (R, G and B).
3. Ordinary source detection algorithms use functions that are designed to process images with gray scale of 2^8 (256) levels. However, astronomical images have gray scale which is much larger than that of ordinary images.

Therefore, it would be necessary for scientists to design a framework, instead of simply adapting general purpose target detection neural networks to process images obtained by multi-colour photometry sky surveys. In this paper, we have built a framework to process multi-colour photometry images. Firstly, we propose a multi-step detection strategy, which contains two neural network: a neural network based on attention-mechanism, which is better in detection of extended targets, to detect and segment images of extended targets; a neural network based on Faster R-CNN, which is better in detection of small targets, to detect and classify point like targets. Secondly, we have modified structures and data augmentation algorithms in these two neural networks to process images with any number of colours. With these designs and modifications, our framework could directly detect and segment celestial objects from observational images. This paper contains the following parts. In Section 2, we would discuss the data and the structure of neural networks used in our framework. We would test the performance of our algorithm in Section 3, make conclusions and anticipate our future works in Section 4.

2. Data properties and the target detection framework

2.1. Properties of multi-colour photometry sky survey images

In multi-colour photometry sky survey projects, filters are specially selected to trace different components of celestial objects. For example, H-alpha filters are used to observe distribution of hydrogen atoms in nearby galaxies and infrared filters (centred in 91521 Å with around 1001 Å width) could be used to trace hydrogen Lyman alpha emission from star formation region in distant galaxies. Nowadays, many multi-colour photometry sky survey projects are proposed with much more filters to discover and study celestial objects in detail. For example, the J-PAS uses 56 narrow band filters to observe celestial objects in optical band Benitez et al. (2014). The Chinese Space Station Telescope would use at least 7 filters from ultra-violet to optical band to observe celestial objects (Gong et al., 2019).

In this paper, we use the SDSS DR17 data as examples of multi-colour photometry sky survey images (Accetta et al., 2022). The SDSS is one of the most successful multi-colour sky survey projects, partly due to its high quality data processing pipeline (Lupton et al., 2002, 2005). There are five filters used in the SDSS project, which include g, i, r, u and z filters. These filters cover from optical band to near infrared band. The SDSS data processing pipeline would normally detect celestial objects only from images obtained in the r band, because generally images in r band would have higher signal to noise ratio. However, as we have discussed above, there are some other celestial objects which have higher signal to noise ratio in other bands or could be better detected, if we could consider multiple bands. Although we would detect extended targets, point targets from the SDSS data to test the performance of the framework, advantages of our framework could further be released, if we detect celestial objects that require multi-colour information, such as dark nebulae or emission nebulae.

2.2. Evaluation criterion

The framework proposed in this paper is majorly used to detect celestial objects from observational images and we could obtain the following results with the framework: bounding boxes (rectangular boxes that outlines targets) for point targets and polygons for extended targets. Therefore, the position accuracy of detection results would be rough, which could reduce complexity of the neural network and help us to better train the neural network. Accurate astrometry and photometry results with uncertainty for point targets could be obtained by adding regression Bayes neural networks (Jia et al., 2021) and we could also develop separate algorithms to deblend stars in clusters or galaxies.

According to detection requirements, we could define evaluation criterion for the detection framework. Since we use bounding boxes and polygons to define detection results, we would use the IOU (intersection and merging ratio) to define whether neural networks have detected celestial objects. The IOU is used to describe the degree of overlap between bounding boxes or polygons predicted by neural networks and ground truth bounding boxes or polygons. When IOU is 0, bounding boxes or polygons do not overlap and there is no intersection; when IOU is 1, bounding boxes or polygons completely overlap; when IOU is between 0 and 1, the higher the value, the larger the area of the overlapping part between bounding boxes and polygons as defined in Eq. (1),

$$IOU = \frac{InterSectionArea}{UnionArea}. \tag{1}$$

In real applications, we would define a detection as True Positive detection (TP), if a positive detection has an IOU larger

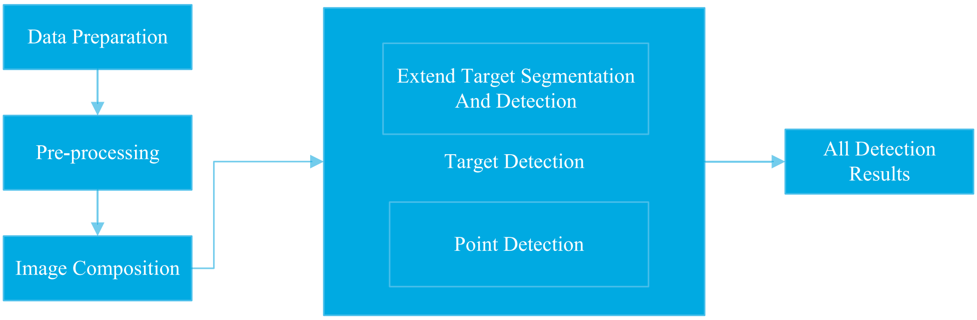


Fig. 1. The structure of the framework to detect celestial objects from multi-colour photometry sky survey images.

than a predefined value and a True Negative detection (TN), if a negative detection result has an IOU lower than a predefined value. Otherwise, we would have a False Positive detection (FP), if a positive detection result has an IOU larger than a predefined value and a False Negative detection (FN), if a negative detection result has an IOU smaller than a predefined value. In this paper, we set IOU as 0.5 for detection of extended targets and 0.8 for detection of point targets. With these definitions, we could calculate precision rate and recall rate as the evaluation criterion for the detection framework. The precision rate is defined in Eq. (2),

$$precision = \frac{TP}{TP + FP}, \tag{2}$$

which could show percentage of TPs in all detection results. The recall rate is defined in Eq. (3),

$$recall = \frac{TP}{TP + FN}, \tag{3}$$

which could show the probability of correctly predicting a positive sample out of all positive samples. With a given IOU as detection threshold, the precision and recall rate reflects the performance of a detection algorithm. Normally, we could plot the PR curve under a given IOU to represent the performance of a detection algorithm. The horizontal axis of the PR curve represents precision rate and the vertical axis represents recall rate. The area enclosed by the PR curve, known as the Mean Average Precision (MAP), reflects the performance of a detection algorithm, as defined in Eq. (4),

$$Map = \frac{\sum_1^n precision}{n}, \tag{4}$$

where n represents the number of all categories. In this paper, we would use the MAP and the PR curve as the evaluation criterion to evaluate the performance of an algorithm. Meanwhile, we will also evaluate precision and recall rate for our framework in detection of celestial objects from real observation data.

2.3. Data preparation for the framework

Our framework could be divided into four different parts as shown in Fig. 1, which includes data preparation, data pre-processing, image composition and target detection. In the data preparation part, we would mask bad pixels, use dark frame images and flat field images to calibrate observation images. In the data pre-processing part, we would use zscale algorithm defined in the DS9 to separately transform gray scale of observation images in different bands to increase signal to noise ratio of celestial objects with low magnitudes (Joye et al., 2003). In the image composition part, we would use the image registration algorithm to align multi-colour images according to the plate model of the camera. After the image composition part, we could obtain multi-colour images and celestial targets in these multi-colour images have been aligned. Original images in each wavelength band have a size of 1 × H × W and composited images have a size of N × H × W, where N stands for number of colours, H and W stand for height and width of images. It should be noted that the pixel scale difference is small for SDSS images. However, if we use observation images from ultra-violet to infrared, the difference of pixel scale would be very large and we need to rescale all observation images to a predefined pixel scale, normally the pixel scale for images with larger signal to noise ratio.

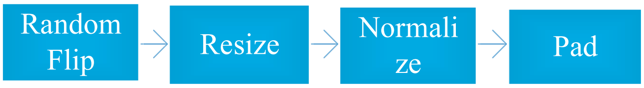


Fig. 2. The flow chart of data augmentation procedure.

Since we use supervised learning algorithms (neural networks) in the framework, we need to train these neural networks before we deploy them in real applications. We need images as well as their labels as the training data. Our detection framework would output bounding boxes for point targets and polygons for extended targets. Therefore, we would use Gaia DR2 (Gaia Collaboration et al., 2018) to obtain catalogue of point and extended targets. For point targets, we would directly define bounding boxes with 10 × 10 pixels. For extended targets, we would transform these images to JPG files and share these images with the labelme through the internet for citizen scientists (Russell et al., 2008). Then citizen scientists would click pixels as annotated pixels and images surrounded by annotated pixels will be defined as labels for extended targets. Since manual annotation would introduce error to polygons, we would discuss this problem in Section 3.4. We have obtained a total of 5000 images as the training data. Since it is quite expensive to obtain labels for training data, we need to obtain more training data with data augmentation technology. Therefore, after we obtain images and their corresponding labels, we would resize, random flip, random rotate and normalize aligned images to increase the number of images in the training data-set as shown in Fig. 2.

There are two neural networks in the target detection part: extended target detection and segmentation neural network and point target detection neural network. Extended targets are targets that could not be resolved to point sources, such as galaxies, nebulae or globular clusters. Point targets are resolved targets, which include stars, moving targets and quasars. During the training stage, we would send labelled images separately to these two neural networks and we would train these neural networks with these images. During the deployment stage, we would firstly detect and segment extended targets with the extended target detection and segmentation neural network and then detect point sources with the point target detection neural network. We will discuss details of these two neural networks in following sub-sections.

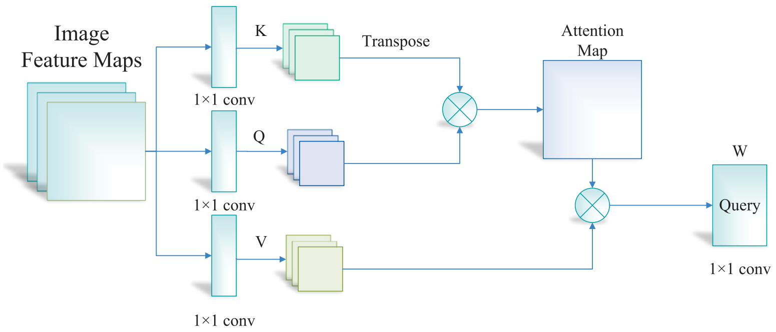


Fig. 3. The schematic draw of the vision transformer.

2.4. Detection and segmentation of extended targets with the swin transformer

There are extended targets, such as galaxies or nebulae, in the multi-colour photometry sky survey images. Since extended targets have complex structures and would be obstructed by foreground targets, it would be quite hard to detect these targets with high efficiency. For extended targets, a type of attention-based neural network, the vision transformer has achieved remarkable performance for general purpose target detection tasks. The vision transformer is a type of transformer, which is firstly used for natural language processing and later in vision tasks (Vaswani et al., 2017; Gal et al., 2022). The vision transformer has the same mechanism as the human vision system, which uses attention mechanism to increase its ability in building long range relations between pixels in images as well as reduce the cost of computation resources during the training stage. For target detection tasks, the vision transformer would pay attention to important parts of a target and ignores gaps or areas that are not dense with information. Therefore, we use the vision transformer as the basic block to build the detection and segmentation algorithm for extended targets. The structure of the vision transformer is shown in Fig. 3.

There are three different components in the vision transformer: Key, Query, Value. The vision transformer would randomly send N Queries and Keys to obtain scores and calculate correlations between scores and Values to obtain important information for target detection (Carion et al., 2020). However, since each pixel has to capture global contextual information in ordinary vision transformer, the self-attention mechanism module has a high computational complexity and would cost a large amount of GPU memory, when it is used to process images. Meanwhile, the ordinary vision transformer directly divides the image evenly into disjoint patches and calculates the self-attention between two patches, which is less suitable to extract features with multiple different scales, particularly for astronomical targets, who have features of different size.

As a new type of visual transformer, the Swin Transformer proposes the idea of the sliding window method to reduce the amount of computation by limiting the computation of attention to each window and uses a pooling-like operation called patch merging to capture features with different scales (Liu et al., 2021a). In the Swin Transformer, the window is used as basic element for each query and each window will move by half of the window size from the top left to the bottom right of the input image. Each time, we would calculate a window-based multi-headed self-attention between each windows to extract features. Besides, the Swin Transformer merges patches that are close to each other, so that the receptive field would become larger and features of multi-scale targets can be captured, which is adequate in detection of extended celestial objects.

The schematic diagram of the Swin Transformer used in this paper is shown in Fig. 4. Since we would adapt the Swin Transformer to detect targets from multi-colour images, we design the Swin Transformer in a hierarchical manner, which contains four stages. The Patch Partition layer would generate 4 images with different resolution according to the original image. Each generated images are multi-colour images with lower resolution, which could help the Swin Transformer to obtain larger perception field. If we set the patch size as M and images with N colours, the patch partition module would have as size of $M \times M$. Then the input image would be transformed to a matrix with size of $\frac{H}{M} \times \frac{W}{M} \times (M \times M \times N)$.

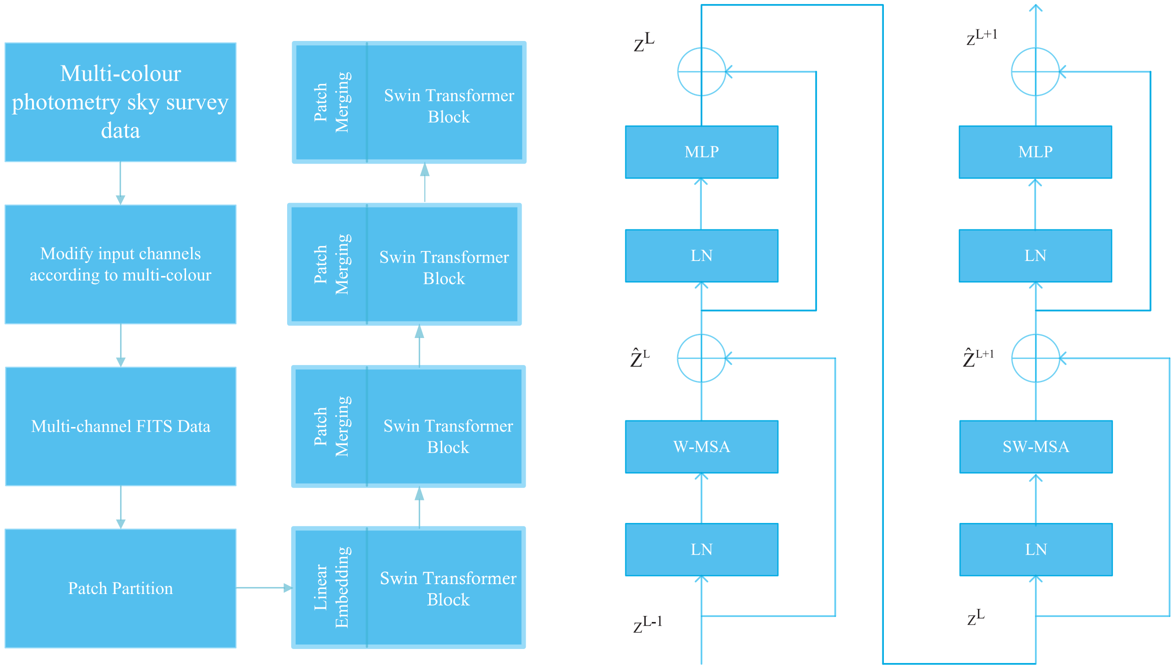
Then the transformed matrix would be sent to the linear embedding layer and Swin Transformer Blocks. There are four different layers in the Swin Transformer block: LN (Layer Normalization), W-MSA (Window Masked Attention Layer), SW-MSA (Shifted Window Masked Attention Layer) and MLP (Multilayer Perceptron). The LN is used to normalize vectors and the MLP is used to extract output. The W-MSA and the WS-MSA are normally used by pairs, therefore the number of Swin Transformer Blocks is always even number as shown in Fig. 4.

The W-MSA would calculate attention between windows and the WS-MSA would calculate attention between shifted windows. With these Swin Transformer Blocks, we could extract feature maps of different levels for further detection and segmentation. The Swin transformer block would carry out the following calculations, as defined in Eq. (5),

$$\begin{aligned}
\hat{z}^{l} &= W - MSA\left(LN\left(z^{l-1}\right)\right) + z^{l-1},\\
\hat{z}^{l} &= MLP\left(LN\left(\hat{z}^{l}\right)\right) + \hat{z}^{l-1},\\
\hat{z}^{l+1} &= SW - MSA\left(LN\left(\hat{z}^{l}\right)\right) + z^{l},\\
\hat{z}^{l+1} &= MLP\left(LN\left(\hat{z}^{l+1}\right)\right) + \hat{z}^{l+1},
\end{aligned} \tag{5}$$

where $\hat{z}^{l}$ and z^{l} represent the output features of the $(S)W - MSA$ module and the MLP module for block l respectively. $W - MSA$ and $SW - MSA$ represent window based multi-head self-attention using regular and shifted window partitioning configurations respectively (Liu et al., 2021a). In order to achieve equivalent calculations for W-MSA and SW-MSA with the same number of windows, the Swin Transformer has set a reasonable mask to isolate the information in different areas. In this paper, the Swin Transformer contains 8 Swin Transformer Blocks (4 pairs) to process observational images.

We use features extracted by the Swin Transformer to segment targets from observational images, as shown in Fig. 5. The structure of the detection and segmentation part is adapted from convolutional neural network based target detection and segmentation neural networks. Multiple roi proposals are obtained by



(a) Swin Transformer

(b) Swin Transformer Block

Fig. 4. The schematic draw of the swin transformer and swin transformer block.

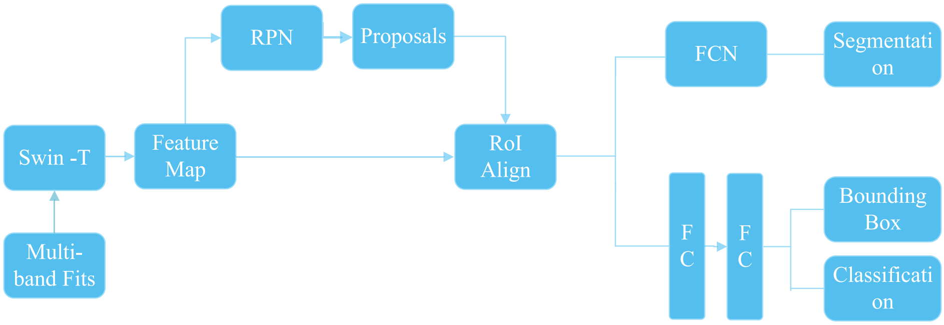


Fig. 5. The overall architecture of the Swin Transformer for target detection and segmentation.

setting roi of each point in the feature map. These roi proposals are sent to the RPN network for binary classification (foreground or background) and bounding box regression. We would obtain bounding boxes that contain foreground (targets) and use the Roi Alignment to modify size of these bounding boxes. At last we would send these bounding boxes to the FCN (fully convolutional neural network) for target segmentation and the FC (Fully connected neural network) for target detection.

2.5. Detection of point sources with the Faster R-CNN

We use the Faster R-CNN to detect resolved targets, such as: stars or quasars. The Faster R-CNN is a two-stage algorithm that integrates feature extraction, proposal extraction, bounding box regression and classification into a single neural network, as shown in Fig. 7. The Faster R-CNN is used in Jia et al. (2020) to detect point-like and streak-like targets from single colour images obtained by wide field small aperture telescopes. In this paper, we have modified the structure of the Faster R-CNN neural network to make it possible to process images with any number of channels (colours). The modified Faster R-CNN neural network includes the following four different parts.

1. Feature extraction. This part is to used to extract features from multi-colour images and these features would be sent to the RPN network, which will generate candidate regions for classification and position regression. We adapt our original design, which uses ResNet-50 as the backbone neural network for feature extraction and use convolutional kernel with size of 3 × 3 pixels. However, our neural network would adjust its structure according to the number of colours in the input image, which would automatically set the number of channels in the convolution kernel for feature extraction based on the number of colours in the input image. We set the number of convolutional kernels as twice of the number of the colour of the input image c and perform feature extraction on images with convolutional kernels with size of 3 × 3 pixels. Then, the number of convolution kernels in the first BottleNeck of ResNet is set to 2^c, and we could obtain feature maps with size of $2^c \times 4$. The next three BottleNecks would perform similar operations to extract features. The modified ResNet is shown in Fig. 6.

2. Region Proposal. In this part, the RPN network generates region proposals with celestial target information for subsequent networks. We would carry out region proposal regressions to obtain proposals with more accurate positions.

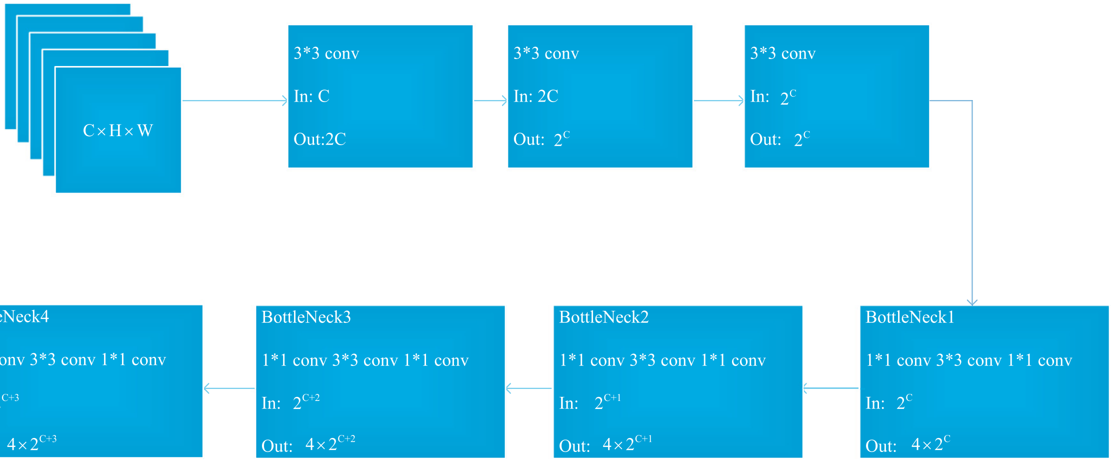


Fig. 6. The structure of the modified ResNet50. *C* stands for number of colour and *H* and *W* stand for the size of input images.

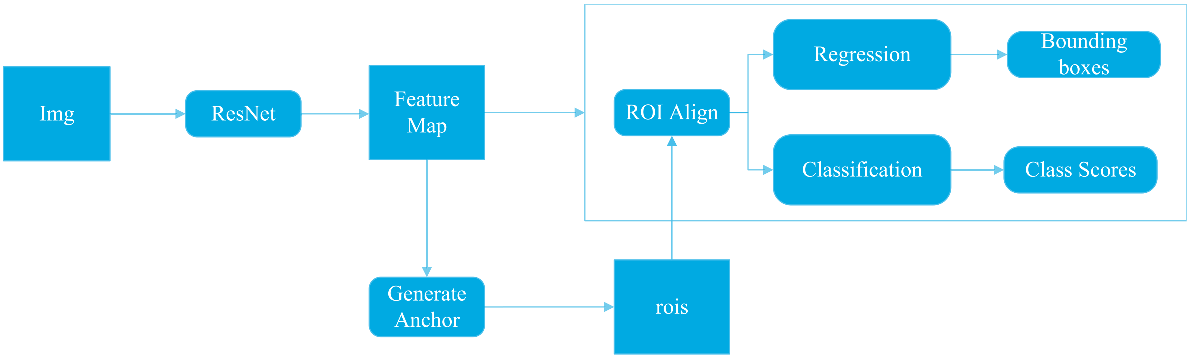


Fig. 7. The overall architecture of the Faster R-CNN.

3. Roi alignment. Roi alignment will use a bi-linear interpolation method to obtain positions of proposals for two variables (x and *y* coordinates). For multi-colour images, we assume the position of a target in images of different colours would be the same. Then we send feature maps extracted from multi-colour images to the subsequent fully connected layer to classify them into backgrounds or targets.

4. Classification and regression. In this part, we will use multi-colour feature maps of proposals to compute categories of proposals. Meanwhile we will re-scale these proposals and use bounding box regression to achieve higher regression and classification accuracy. The detection results would be bounding boxes which outline positions of celestial objects and types of these different celestial objects.

2.6. Integration of the detection algorithm and the segmentation algorithm

Two deep learning based algorithms discussed above would be integrated together in real applications as shown in Fig. 8. In the framework, we would firstly detect and segment extended targets with the Swin Transformer discussed in Section 2.4. Then pixels that belong to detected extended targets will be set as mean values of the whole image and we would send masked images to the Faster R-CNN discussed in Section 2.5 to detect point targets. At last, all detection results would be merged together and stored as final detection results. There is very small possibility that point targets would be detected as extended targets in the first step, which would lead to FPs to extended target detection results and FNs to point target detection results.

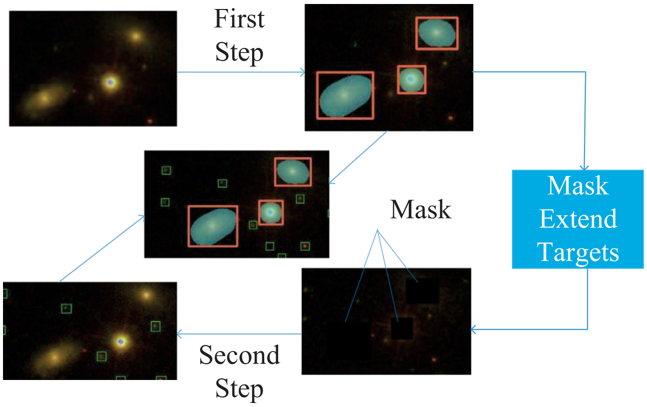
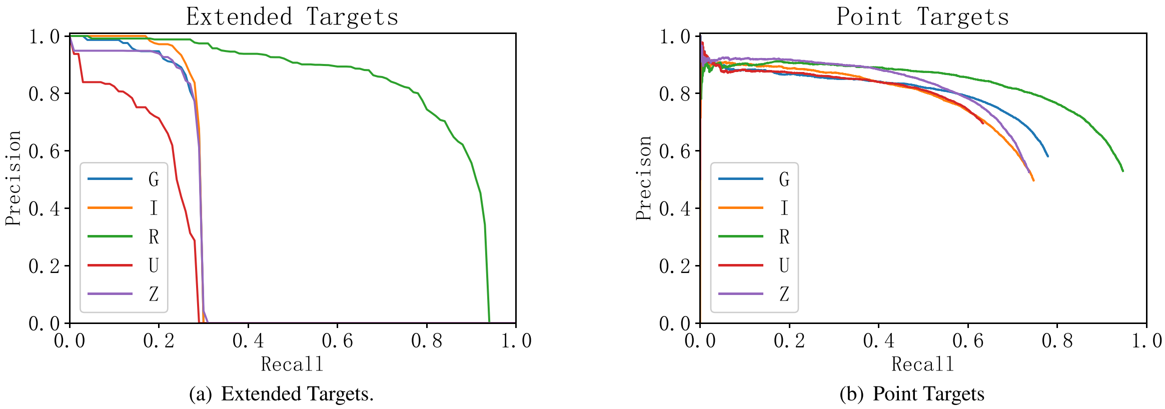


Fig. 8. The schematic draw of the integrated target segmentation and detection framework.

3. Training and performance test of the framework

In this section, we will train and test the performance of our framework with data from the SDSS DR17. We have processed these data with the method discussed in Section 2.3. Images from the SDSS DR17 data are obtained by 5 different filters. Therefore, we could test the performance of our algorithm in detection celestial objects from images with single colour, three colours and five colours. Besides, we would also test the robustness of our algorithm, if labels of extended celestial objects are not well annotated, which is a common situation for data labelled by human experts.



(a) Extended Targets.

(b) Point Targets

Fig. 9. Detection results for images with single colour. Our framework could obtain best detection results for image obtained in r band.

Table 1
The confusion matrix of detection results from images of single colour.

Class	Results		
	MAP	Precision	Recall
Extended targets	0.824	0.738	0.838
Point targets	0.828	0.808	0.913

Table 2
The confusion matrix of detection results from images of three colours.

Class	Results		
	MAP	Precision	Recall
Extended targets	0.864	0.803	0.852
Point targets	0.830	0.828	0.914

Table 3
The detection results of five Colours.

Class	Results		
	MAP	Precision	Recall
Extended targets	0.868	0.825	0.875
Point targets	0.838	0.886	0.924

3.1. Detection results for images of single colour

In order to show the advantageous of our algorithm in detection of celestial objects from multi-colour images, we would firstly show the performance of our algorithm in detection of celestial objects from images of single colour as baseline. We have separately trained five frameworks with images of ugriz bands and test the performance of these frameworks in detection of extended targets and point targets from these images. We use 2052 images in each band to test the performance of our framework. The best detection results could be obtained for images obtained in the r band, which is consistent with our experience (celestial objects have higher signal to noise ratio in images of r band). The p-r curves are shown in Fig. 9.

As shown in this figure, our framework has different performance in detection of point targets and extended targets. The precision in detection of point targets is relatively stable with different recall rates, while the precision and the recall are would change for extended targets, albeit the MAP is almost the same for both of these two targets. The difference indicates us that generally extended targets are harder to be detected than point targets and there are some point targets that would not be detected even if we set very small threshold. The confusion matrix of detection results is shown in Table 1. As shown in this table, we could confirm that point targets are easier to be detected than extended targets, presumably because extended targets have quite complex structures and low signal to noise ratio.

3.2. Detection results for images of three colours

In this subsection, we would test the performance of our framework in detection of celestial objects from images of three colours. We have chosen original images obtained from g, r and i bands and train our framework with these images. We also use 2052 images to test the performance of our framework. The p-r curve are shown in Fig. 10 and the confusion matrix is shown in Table 2.

As shown in the p-r curve, we could obtain better detection results with images of three colours, particularly for extended targets. However, there is a small decrease of precision rate for point targets, when the recall rate is high, which is probably caused by images of point targets with low signal to noise ratio. Besides, as shown in the confusion matrix, the detection accuracy of extended targets has been improved, since we could include more colour information with images of three colours. The results indicate the possibility that our framework could better extract information of multi-colour images to detect celestial objects.

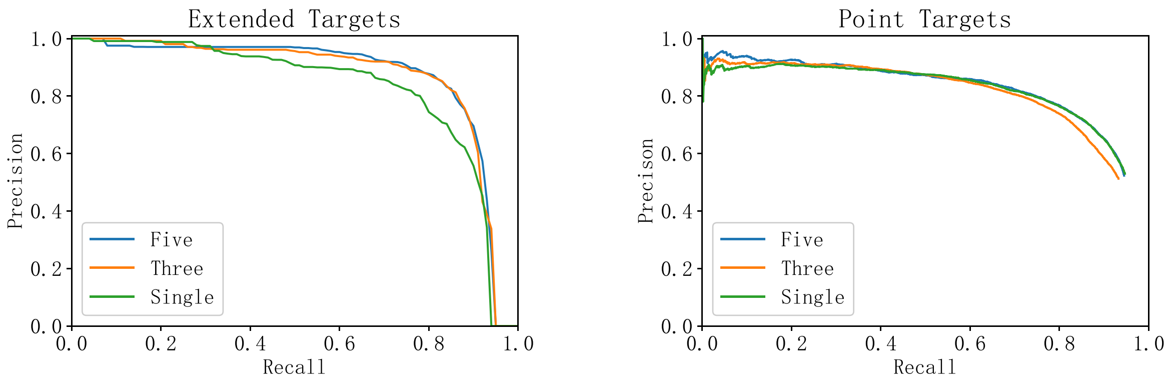
3.3. Detection results for images of five colours

In this subsection, we would further test the performance of our framework in detection of celestial objects from images of five colours. We use images of all five bands to train and test our framework. All these images are aligned with steps discussed above. We use 2052 images to test the performance of our framework and the results are shown in Fig. 10 and the confusion matrix is shown in Table 3.

As shown in the p-r curve, our framework could achieve better performance in detection of celestial objects from images of five colours, particularly in detection of point targets that are hard to be detected (targets with low recall rate and high precision) and classification of extended targets that are easier to be merged with backgrounds (targets with low precision and high recall rate). Meanwhile the confusion matrix shows that the detection accuracy could further increase, if we have images with more colours.

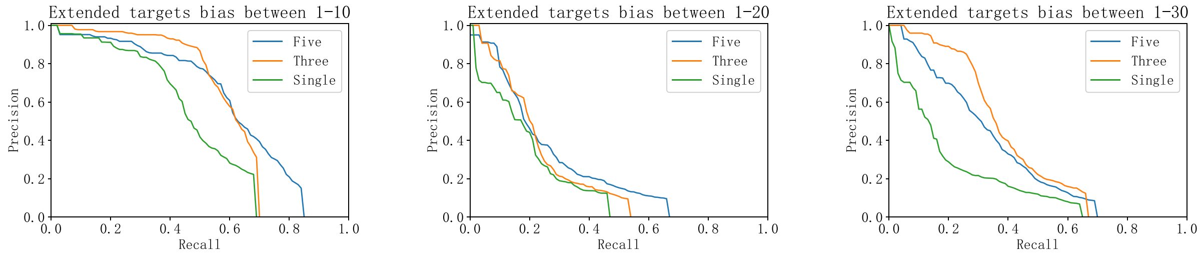
3.4. Robustness of the framework with labels of low quality

Since our framework is based on supervised learning algorithms, the training data is important and would directly affect the performance of our framework. In this paper, we use Gaia DR2 to locate central points of stars, quasars and galaxies in each



(a) Extended Targets (b) Point Targets

Fig. 10. Detection results for images with single colour, three colours and five colours.



(a) 1-10 (b) 1-20 (c) 1-30

Fig. 11. Detection results for images with different levels of annotation error.

Table 4
The recall rate and the precision rate (R/P) when IOU=0.5.

R/P Band Error	single	three	five
1-10	0.245/0.469	0.324/0.589	0.319/0.604
10-20	0.134/0.184	0.168/0.228	0.212/0.257
20-30	0.206/0.192	0.239/0.386	0.216/0.328

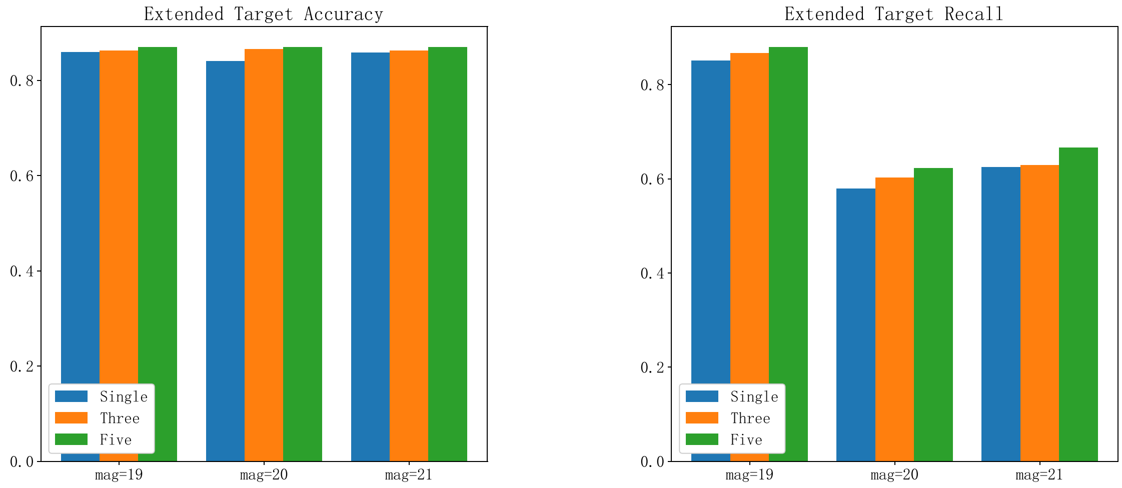
images. Then we use central points and bounding boxes with fixed size to label stars and quasars, which would be unlikely to introduce error. However, we use central points and polygons obtained through manual annotation to label extended targets. Therefore, there would be some positional errors induced by manual annotation, which would affect the performance of the extended targets detection algorithm in our framework.

To test the robustness of our framework, we add random shift to labels of galaxies and train the framework with these new images. We have generated three data-sets which have positioning error with 0 to 10 pixels, 1 to 20 pixels and 1 to 30 pixels. Detection results obtained by the extended targets detection algorithms are shown in Fig. 11. It can be seen that the area of the image enclosed by the p-r curve for each band decreases as the offset increases, in other words, the MAP decreases. Besides, these figures also show that the label offset would mainly affect the accuracy, and would slightly affect the recall. Besides, we could also find that multi-colour images would lead to better detection results and multi-colour images are relatively more robust to the label error. Although the P-R curve has changed significantly as we have label errors, we need to mention that our algorithm would set IOU of 0.5 as detection threshold and we could reduce IOU threshold, which would obtain better detection results and lower positioning accuracy (see Table 4).

3.5. Application of the pipeline in processing of observation images

At last, we would integrate our framework with astrometry (Lang et al., 2010) and astroquery (Ginsburg et al., 2019) to build a pipeline to process real observational images. The real observation images are also from SDSS DR 17, however in a different stripe. The code which contains our framework and a jupyter notebook file could be found in China-VO. In real applications, we would firstly set channels of these two neural networks as the number of colours. Then we would generate training data (normalized images with five colours) from original fits files with the data preparation steps mentioned above. For the training set, positions of point targets and extend targets are obtained from Gaia DR 2. Bounding boxes for point targets are obtained automatically with fixed size (defined manually according to the full width half magnitude of the point spread function). Polygons that are used to label extend targets are obtained by citizen scientists. We have defined a data loader to load these training data and generate new training data with data augmentation methods discussed above. Then we will train these two neural networks separately with data generated by the data loader.

After training, we would deploy these two trained neural network in target detection. Original data would also be processed from fits files to normalized images with five colours. Then these images would be firstly put into the Swin transformer to obtain detection results for galaxies. Then extended targets would be masked out and we would put masked images into the Faster R-CNN to obtain point targets. We would use positions of stars as reference and use the Astrometry to obtain plate model. Then we would transform positions of all detection results from camera coordinates to WCS coordinates. We use centre positions of bounding boxes as positions of point targets. Meanwhile, we use the gravity centre of polygons as positions of galaxies. We cross-match all our detection results with SDSS catalog (Alam et al., 2016) by Astroquery to obtain positions of different celestial



(a) Extended Target Accuracy

(b) Extended Target Recall

Fig. 12. The accuracy rate and the precision rate of our framework in detection of extended targets when cross matched with catalog of different limiting magnitude.

Table 5
The recall rate and the accuracy rate when limiting magnitude is 19.

Channel&Bands	Accuracy		Recall	
	Extend	Point	Extend	Point
Single Band(r)	0.859	0.867	0.851	0.827
Three Band(i,r,u)	0.863	0.872	0.867	0.835
Five Band(g,i,r,u,z)	0.870	0.877	0.880	0.852

Table 6
The recall rate and the accuracy rate when limiting magnitude is 20.

Channel&Bands	Accuracy		Recall	
	Extend	Point	Extend	Point
Single Band(r)	0.841	0.800	0.579	0.855
Three Band(i,r,u)	0.866	0.846	0.603	0.859
Five Band(g,i,r,u,z)	0.870	0.872	0.623	0.860

objects. As we have discussed above, the position accuracy is not the core for our method. During the cross-match stage, we would set the position accuracy for point targets as 1 pixel and 5 pixels for extend targets. With steps mentioned above, we could obtain a catalog and cross-match the catalog with SDSS catalog. Results are shown in Tables 5–7 with different limiting magnitude. As shown in these tables, our method could achieve over 87% accuracy and 85% recall rate in detection of celestial objects from single frame images. Better results could be achieved, if we detect celestial objects from stacked images. Since this part is used to show the performance of our framework in detection of celestial objects with multiple colours, we would not further discuss the strategy for image stacking. Further comparisons of the detection results could be found in Figs. 12 and 13. As shown in these figures, our detection framework could achieve better detection results from images with more channels: the accuracy and the precision are both increasing, as we increase the number of colours of these images. These results indicate that our framework is effective.

At last, we have shown the performance of our framework and some other deep learning based target detection algorithms in Table 8. Results show that our framework could achieve more than 80% precision rate and more than 85% recall rate for all targets, which would be worse than other methods. We need to mention that our framework directly detects celestial objects from original SDSS images obtained by single epoch instead from some generated or simulated images. Therefore, noises or variations of backgrounds would limit the performance of our algorithm. Our algorithm could achieve better results, if we could use stacked images for detection.

Table 7
The recall rate and the accuracy rate when limiting magnitude is 21.

Channel&Bands	Accuracy		Recall	
	Extend	Point	Extend	Point
Single Band(r)	0.858	0.877	0.625	0.593
Three Band(i,r,u)	0.863	0.874	0.629	0.642
Five Band(g,i,r,u,z)	0.870	0.901	0.667	0.651

4. Conclusions and future works

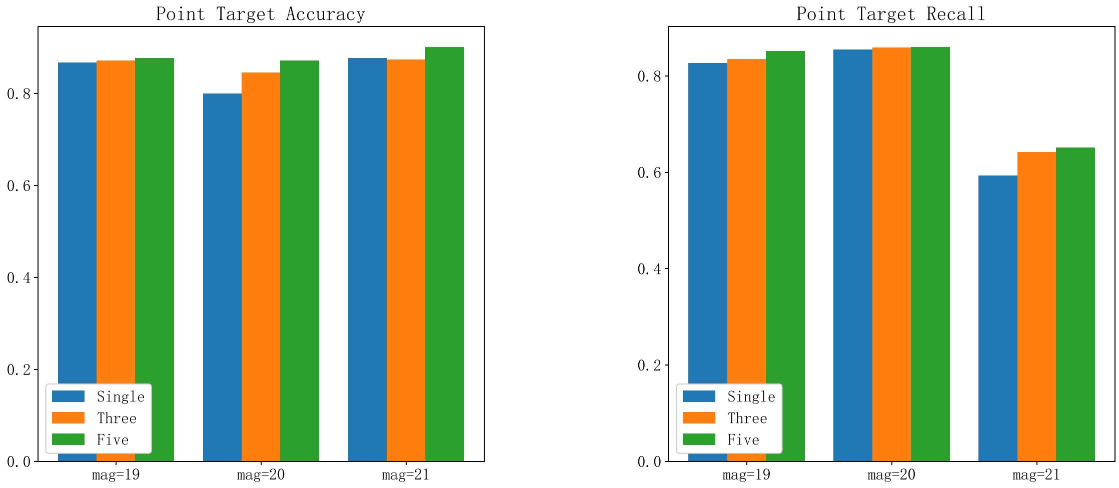
In this paper, we propose a deep learning based framework to process images obtained by multi-colour photometry sky survey projects. Our framework could detect extended targets and point targets directly from observation images, according to their multi-colour information. We have tested the performance of our framework with real observation data from SDSS and results show that our method could achieve relatively stable results after training. However, there are still some tricks left for us to further increase the performance of our framework. In the future, we would use neural network architecture search method to obtain better structure for different neural networks. Since our framework could process images with any number of colours, we would not only use our framework to process data obtained by optical telescopes, such as the Sitian and the CSST, but also data from radio telescopes, such as the SKA, which would obtain data with up to hundreds of channels.

CRediT authorship contribution statement

P. Jia: Conceptualization, Writing – original draft, Coding. **Y. Zheng:** Data processing, Pipeline Design. **M. Wang:** Data labelling. **Z. Yang:** Data labelling, Data Processing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.



(a) Point Target Accuracy

(b) Point Target Recall

Fig. 13. The accuracy rate and the precision rate of our framework in detection of point targets when cross matched with catalog of different limiting magnitude.

Table 8
Comparison with deep learning based methods.

	Channels&Bands	Quasars		Star/Point targets		Galaxy/ Extend targets	
		Precision	Recall	Precision	Recall	Precision	Recall
APSCNet	u,g,r,i	0.841	0.932	0.945	0.846	0.958	0.951
Astron R-CNN	g,r,z			0.882	0.938	0.950	0.990
Mask Galaxy	R,G,B					0.986	0.991
Our Framework Single	r			0.808	0.913	0.738	0.838
Our Framework Three	i,r,u			0.828	0.914	0.803	0.852
Our Framework Five	g,i,r,u,z			0.886	0.924	0.825	0.875

Data availability

The data and code for review are shared in: https://nadc.china-vo.org/res/r101190/.

References

Accetta, K., Aerts, C., Aguirre, V.S., Ahumada, R., Ajgaonkar, N., Ak, N.F., Alam, S., Prieto, C.A., Almeida, A., Anders, F., et al., 2022. The seventeenth data release of the sloan digital sky surveys: Complete release of MaNGA, MaStar, and APOGEE-2 data. Astrophys. J. Suppl. Ser. 259 (2), 35.

Alam, S., et al., 2016. Vizier online data catalog: The SDSS photometric catalogue, release 12 (Alam+, 2015). VizieR Online Data Catalog V–147.

Benitez, N., Dupke, R., Moles, M., Sodre, L., Cenarro, J., Marin-Franch, A., Taylor, K., Cristobal, D., Fernandez-Soto, A., de Oliveira, C.M., et al., 2014. J-PAS: the javalambre-physics of the accelerated universe astrophysical survey. arXiv preprint arXiv:1403.5237.

Bertin, E., Arnouts, S., 1996. Sextractor: Software for source extraction. Astron. Astrophys. Suppl. Ser. 117 (2), 393–404.

Bom, C., Makler, M., Albuquerque, M., Brandt, C., 2017. A neural network gravitational arc finder based on the mediatrix filamentation method. Astron. Astrophys. 597, A135.

Burke, C.J., Aleo, P.D., Chen, Y.-C., Liu, X., Peterson, J.R., Sembroski, G.H., Lin, J.Y.-Y., 2019. Deblending and classifying astronomical sources with mask R-CNN deep learning. Mon. Not. R. Astron. Soc. 490 (3), 3952–3965.

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S., 2020. End-to-end object detection with transformers. In: European Conference on Computer Vision. Springer, pp. 213–229.

De La Calleja, J., Fuentes, O., 2004. Machine learning and image analysis for morphological galaxy classification. Mon. Not. R. Astron. Soc. 349 (1), 87–93.

Duev, D.A., Mahabal, A., Masci, F.J., Graham, M.J., Rusholme, B., Walters, R., Karmarkar, I., Frederick, S., Kasliwal, M.M., Rebbapragada, U., et al., 2019. Real-bogus classification for the zwicky transient facility using deep learning. Mon. Not. R. Astron. Soc. 489 (3), 3582–3590.

Farias, H., Ortiz, D., Damke, G., Arancibia, M.J., Solar, M., 2020. Mask galaxy: Morphological segmentation of galaxies. Astron. Comput. 33, 100420.

Flaugher, B., Diehl, H., Honscheid, K., Abbott, T., Alvarez, O., Angstadt, R., Annis, J., Antonik, M., Ballester, O., Beaufore, L., et al., 2015. The dark energy camera. Astron. J. 150 (5), 150.

Gaia Collaboration, et al., 2018. Vizier online data catalog: Gaia DR2 (gaia collaboration, 2018). VizieR Online Data Catalog I–345.

Gal, R., Alaluf, Y., Atzmon, Y., Patashnik, O., Bermano, A.H., Chechik, G., Cohen-Or, D., 2022. An image is worth one word: Personalizing text-to-image generation using textual inversion. arXiv preprint arXiv:2208.01618.

Ginsburg, A., Sipőcz, B.M., Brasseur, C., Cowperthwaite, P.S., Craig, M.W., Deil, C., Groener, A.M., Guillochon, J., Guzman, G., Liedtke, S., et al., 2019. Astroquery: an astronomical web-querying package in Python. Astron. J. 157 (3), 98.

Girshick, R., Donahue, J., Darrell, T., Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 580–587.

Gong, Y., Liu, X., Cao, Y., Chen, X., Fan, Z., Li, R., Li, X.-D., Li, Z., Zhang, X., Zhan, H., 2019. Cosmology from the Chinese space station optical survey (CSS-OS). Astrophys. J. 883 (2), 203.

He, Z., Er, X., Long, Q., Liu, D., Liu, X., Li, Z., Liu, Y., Deng, W., Fan, Z., 2020. Deep learning for strong lensing search: tests of the convolutional neural networks and new candidates from KiDS DR3. Mon. Not. R. Astron. Soc. 497 (1), 556–571.

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask r-cnn. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2961–2969.

He, Z., Qiu, B., Luo, A.-L., Shi, J., Kong, X., Jiang, X., 2021. Deep learning applications based on SDSS photometric data: detection and classification of sources. Mon. Not. R. Astron. Soc. 508 (2), 2039–2052.

Ignat, I.M.M.M.K., Bodeic, A.D.L., 2019. The use of deep learning and neural networks in imaging: Welcome to the new mathematical milieu of medicine.

Ivezić, Ž., Kahn, S.M., Tyson, J.A., Abel, B., Acosta, E., Allsman, R., Alonso, D., AlSayyad, Y., Anderson, S.F., Andrew, J., et al., 2019. LSST: from science drivers to reference design and anticipated data products. Astrophys. J. 873 (2), 111.

Jacquemont, M., Vuillaume, T., Benoit, A., Maurin, G., Lambert, P., Lamanna, G., 2021. First full-event reconstruction from imaging atmospheric cherenkov telescope real data with deep learning. In: 2021 International Conference on Content-Based Multimedia Indexing. CBMI, IEEE, pp. 1–6.

Jia, P., Liu, Q., Sun, Y., 2020. Detection and classification of astronomical targets with deep neural networks in wide-field small aperture telescopes. Astron. J. 159 (5), 212.

Jia, P., Sun, Y., Liu, Q., 2021. The deep neural network based photometry framework for wide field small aperture telescopes. arXiv preprint arXiv:2106.14349.

Joye, W., Mandel, E., Payne, H., Jedrzejewski, R., Hook, R., 2003. ASP conf. Ser. Vol. 295, astronomical data analysis software and systems XII.

Keller, S.C., Schmidt, B.P., Bessell, M.S., Conroy, P.G., Francis, P., Granlund, A., Kowald, E., Oates, A., Martin-Jones, T., Preston, T., et al., 2007. The SkyMapper telescope and the southern sky survey. Publ. Astron. Soc. Aust. 24 (1), 1–12.

Kim, B., Lee, S., Park, C., Kim, H., Song, W.J., 2020. The nebula benchmark suite: Implications of lightweight neural networks. IEEE Trans. Comput. 70 (11), 1887–1900.

Lang, D., Hogg, D.W., Mierle, K., Blanton, M., Roweis, S., 2010. Astrometry. net: Blind astrometric calibration of arbitrary astronomical images. Astron. J. 139 (5), 1782.

Lin, Z., Jiang, N., Kong, X., 2022. The prospects of finding tidal disruption events with 2.5-m wide-field survey telescope based on mock observations. Mon. Not. R. Astron. Soc. 513 (2), 2422–2436.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B., 2021a. Swin transformer: Hierarchical vision transformer using shifted windows. arXiv preprint arXiv:2103.14030.

Liu, J., Soria, R., Wu, X.-F., Wu, H., Shang, Z., 2021b. The SiTian project. Anais da Academia Brasileira de Ciências 93.

Lupton, R.H., Ivezic, Z., Gunn, J.E., Knapp, G., Strauss, M.A., Richmond, M., Ellman, N., Newburg, H., Fan, X., Yasuda, N., et al., 2005. SDSS image processing II: The photo pipelines.

Lupton, R.H., Ivezic, Z., Gunn, J.E., Knapp, G., Strauss, M.A., Yasuda, N., 2002. SDSS imaging pipelines. In: Survey and Other Telescope Technologies and Discoveries. Vol. 4836, SPIE, pp. 350–356.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: Unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 779–788.

Redmon, J., Farhadi, A., 2018. Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767.

Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. Adv. Neural Inf. Process. Syst. 28.

Russell, B.C., Torralba, A., Murphy, K.P., Freeman, W.T., 2008. Labelme: a database and web-based tool for image annotation. Int. J. Comput. Vis. 77 (1), 157–173.

Sun, T., Hu, L., Zhang, S., Li, X., Meng, K., Wu, X., Wang, L., Castro-Tirado, A.J., 2022. Pipeline for antarctic survey telescope 3-3 in yaoan, yunnan. arXiv preprint arXiv:2205.05063.

Tachibana, Y., Miller, A.A., 2018. A morphological classification model to identify unresolved panstarrs1 sources: Application in the ztf real-time pipeline. Publ. Astron. Soc. Pac. 130 (994), 128001.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. Adv. Neural Inf. Process. Syst. 30.

Xi, J., Xiang, Y., Ersoy, O.K., Cong, M., Wei, X., Gu, J., 2020. Space debris detection using feature learning of candidate regions in optical image sequences. IEEE Access 8, 150864–150877.

Xiang, Y., Wu, Z., Gong, W., Ding, S., Mo, X., Liu, Y., Wang, S., Liu, P., Hou, Y., Li, L., et al., 2022. Nebula-I: A general framework for collaboratively training deep learning models on low-bandwidth cloud clusters. arXiv preprint arXiv:2205.09470.

Yuan, X., Li, Z., Liu, X., Niu, D., Lu, Q., Jiang, F., Wang, Y., Li, X., Liang, Y., Wang, H., et al., 2020. Development of the multi-channel photometric survey telescope. In: Ground-Based and Airborne Telescopes VIII. Vol. 11445, SPIE, pp. 1372–1378.

Zhu, X., Su, W., Lu, L., Li, B., Wang, X., Dai, J., 2020. Deformable detr: Deformable transformers for end-to-end object detection. arXiv preprint arXiv:2010.04159.