

Star Detection with RefineDet in Crowded Fields of CSST

Yan-Zhen Hao^{1,2}, Xi-Yan Peng^{1*}, Zhao-Xiang Qi^{1,2*}, Fen Li³, Shi-Long Liao^{1,2}, Yong Yu^{1,2}, Zheng-Hong Tang^{1,2}, Zheng-Yi Shao^{4,5}, Qi-Qi Wu^{1,2}, Zhen-Sen Fu^{1,2} and Wen-Feng Fang^{1,2}

¹ Shanghai Astronomical Observatory, Chinese Academy of Sciences, Shanghai, China,

² University of Chinese Academy of Sciences, Beijing, China,

³ Beijing Institute of Technology, Beijing, China,

⁴ Key Laboratory for Research in Galaxies and Cosmology, Shanghai Astronomical Observatory, Chinese Academy of Sciences, Shanghai, China,

⁵ Key Lab for Astrophysics, Shanghai 200234, People's Republic of China,

Correspondence*:

Zhao-Xiang Qi

zxqi@shao.ac.cn

Xi-Yan Peng

xypeng@shao.ac.cn

2 ABSTRACT

This research paper evaluates the source detection performance using a neural network model based on RefineDet, for the forthcoming Chinese Space Station Telescope (CSST) in crowded fields, where objects are often blended challenging the requirement of accurate position and flux measurement. A comprehensive dataset of simulated images and catalogs is utilized, along with a dataset of cropped images for training and evaluating an object detection model based on RefineDet. RefineDet is a high-precision and fast model in object detection, and work well with small objects which containing most of stars. Adaptations are made to the RefineDet algorithm for star detection in crowded regions, including a novel method for designing prior boxes. The results demonstrate the challenges faced in star detection as star density increases to 5.8 million stars in a 9232×9216 image. However, in images containing 0.16 million stars, the algorithm achieves a typical precision of 85% and recall of 80%, effectively identifying and localizing a significant number of stars with reasonable accuracy and position precision.

Keywords: CSST Survey, object detection, neural network, crowded field, RefineDet, deblending

1 INTRODUCTION

CSST (Chinese Space Station Telescope) is a forthcoming 2-meter aperture telescope designed to operate in low Earth orbit (Zhan, 2021). It is strategically positioned at a certain distance from the Chinese Space Station (CSS) to facilitate repairs and updates while minimizing disturbances from CSS-related vibrations and stray light. CSST Survey covers a substantial field of view of 1.1 square degrees and boasts exceptional imaging quality, characterized by nearly round star imaging and a dark magnitude limit of no brighter than

21 26.3 mag in g band, which is mainly tested in this paper (Zhan, 2021) . With a pixel scale of 74 mas and a
22 typical exposure time of 150 s, CSST Survey exhibits significant single-image astrometry capabilities. The
23 survey region encompasses the entire sky, except for regions with low galactic latitude and low ecliptic
24 latitude. Additionally, certain crowded regions contain high potential value, such as the galactic center
25 (GC) (Gallego-Cano et al., 2022), galactic anticenter, globular clusters, galaxy clusters (Connor et al.,
26 2017), and neighboring galaxies, etc.

27 In astrometry and photometry, a crowded field refers to a sky region with a higher stellar density compared
28 to normal fields. However, it is important to note that the definition of a crowded field may vary across
29 different processes (e.x. (Neichel et al., 2014)). In the context of image processing, a crowded field
30 specifically refers to a field where the high density of sources results in observations that are severely
31 blended, making it challenging to discern individual objects. To assess and describe the degree of crowding
32 in such fields, several methods have been devised. One approach involves estimating the Poissonian
33 probability of a pixel receiving photons from more than two sources based on the assumption of a spatially
34 uniform random distribution, deriving a formula as $P = 1 - e^{-\rho}(1 + \rho)$, where ρ represents the star density
35 (Feinstein et al., 2019). Another study suggests that an approximate natural confusion limit of crowded
36 field ranges from 1/25 to 1/50 beams (Kramer et al., 2022), where the beams is defined as πr_h^2 , with r_h
37 denoting the full width at half maximum (FWHM). For CSST Main Survey, which has an FWHM of
38 approximately 2 pixels, obtained from simulation data by SExtractor (Bertin and Arnouts, 1996), this limit
39 can be converted to a star density of approximately $\rho = \frac{1}{25 \sim 50 \times \pi r_h^2} \approx 0.0064 \sim 0.0127$ object/pixel or
40 $1.16 \sim 2.33$ object/ arcsec^2 . In a 9232×9216 picture, this corresponds to approximately 0.54 million \sim
41 1.08 million stars. This paper involves the generation of a series of mock catalogs and simulation images
42 exhibiting varying stellar densities across a range of 0.16 million to 6.5 million stars within a single picture.

43 To effectively handle crowded field images and extract valuable information, astrometry processing
44 plays a fundamental role. The CSST astrometry processing pipeline consists of following steps. Image
45 preprocessing encompasses tasks such as flat correction, bias correction, and cosmic ray elimination.
46 Object detection involves background deduction, deblending, and measuring position and flux of objects.
47 Single-frame astrometry utilizes objects detected in a single chip to correct distortion and derive a position
48 solution for all objects in that frame. Catalog merging aims to match objects across frame catalogs to
49 create a comprehensive catalog. Finally, the astrometry five-parameter solution utilizes the merged catalog
50 to calculate the position, proper motion, and parallax of each object. Since object detection forms the
51 foundation for subsequent processes, any errors or omissions in this step can lead to misalignments and
52 confusion in single-frame positioning and catalog merging, ultimately propagating errors throughout the
53 entire astrometry process. Therefore, achieving high performance in object detection is of great importance
54 and requirement.

55 Astrometry in crowded fields presents unique challenges. During preprocessing, determining the
56 background map becomes difficult because the number of object pixels can be comparable to the number
57 of background pixels, making it challenging to accurately identify background pixels. Additionally, cosmic
58 rays are pervasive in the images, often passing through one or multiple objects. Determining the point
59 spread function (PSF) is also problematic as it is typically derived from bright, isolated single stars
60 (Anderson and King, 2000), which are rare in crowded fields. In object detection, overlapped objects can
61 create large contiguous regions of light with random shapes, posing challenges for deblending. Moreover,
62 bright stars can obscure faint stars in their vicinity, and regions may contain flux extensions from multiple
63 objects, leading to similarities between small, faint objects and noise pixels. Further description can be
64 found in Melchior et al. (2021). Catalog merging in crowded fields is also challenging, as achieving ideal

65 one-to-one matches between objects is rare, and multi-to-multi matches are common instead, requiring
66 extra caution to avoid misleading associations (Budavári and Basu, 2016; Portillo et al., 2017; Shi et al.,
67 2019; Feder et al., 2020).

68 Despite the difficulties inherent in astrometry processing for crowded fields, the potential value of such
69 observations encourage further exploration. Crowded fields offer unique opportunities to address important
70 and complex problems, such as understanding the dynamics and kinematics of globular clusters, unraveling
71 the structure and properties of galactic nuclei, identifying and characterizing substructures within nearby
72 galaxies, studying stellar populations in dense regions, and investigating statistical properties of the universe
73 (Kramer et al., 2022). Answer to these problems will deepen our understanding of astrophysical processes
74 and phenomena occurring in crowded field environments.

75 From the 1980s to now, various source separation methods were developed, including multi-threshold
76 (Bertin and Arnouts, 1996), image enhancement by convolution (Stetson, 1987), deconvolution (Magain
77 et al., 2007), iterative subtracting (Diolaiti et al., 2000), and synthesis methods (Zheng et al., 2015). Each
78 method performs well in different scenarios, but can hardly meet the demands of extremely crowded fields
79 with high confusion and overlap, a wide range of density variations, and complex noise conditions.

80 Therefore, new approaches such as variational inference (Liu et al., 2021) and machine learning (Paillassa
81 and Bertin, 2019; Hausen and Robertson, 2020) are expected for addressing these challenges. Machine
82 learning has emerged as a significant methodology for automatic learning by machines. One notable
83 machine learning technique is the Convolutional Neural Network (CNN), which specializes in image
84 processing tasks. Through multiple layers, the resulting feature maps capture contextual information in a
85 center-based feeling field and extract shape characteristics of the targets. Furthermore, CNN algorithms
86 for object detection have been extensively developed, refined, and tested, which can be categorized into
87 one-stage approaches (Liu et al., 2016) and two-stage approaches (Ren et al., 2015; Redmon and Farhadi,
88 2018). Two-stage approaches consist of two parts: the first generates candidate object proposals, and the
89 second assigns scores to each proposal and performs regression to refine their precise positions. In contrast,
90 one-stage approaches do not generate candidate proposals but instead use a sparse list of prior boxes that
91 match the sizes of outputs from different layers. Two-stage approaches offer higher precision and better
92 detection of small objects, while one-stage approaches provide faster training speeds. RefineDet (Zhang
93 et al., 2018), an example of a one-stage approach, offers both fast training speed and significantly improved
94 precision, reaching the performance of the best two-stage approach. Therefore, RefineDet is chosen as the
95 base network for conducting star detection experiments in crowded fields.

96 In this paper, detection is focused on, aiming at resolution of overlapped stars. A n-PSF(Point Spread
97 Function) fitting could be a follow-up technique for precise calculation. The rest part is structured as
98 follows: Section 2 presents the conduction of mock catalogs and simulation images, serving as the dataset
99 for RefineDet. Section 3 introduces the experiments on star extraction using RefineDet. Sections 4-5
100 provide a summary and comprehensive discussions of the entire research process.

2 DATA PREPARATION

101 To facilitate object detection training for the forthcoming CSST mission, we employ simulation techniques
102 as CSST has not yet been launched. Our data preparation involves two major steps: dataset generation
103 and data preprocessing. Dataset generation mainly contains mock catalog generation and observation
104 simulation, for generating data in the experiment. And data preprocessing process the raw data into the
105 format required by the model.

106 **2.1 Dataset Generation**

107 **2.1.1 Mock Catalog Generation**

108 Mock catalogs serve as input for observation simulation and require position and magnitude information
109 to construct star observations. Stellar atmospheric parameters are required for determining spectral type,
110 and astrometry parameters like proper motion and parallax, can be included for simulating star motion
111 across different epochs. Two methods are employed to construct mock catalogs for crowded field stars.

112 The first method assumes that all parameters follow independent statistical laws. Sampling is then
113 performed according to their respective distribution laws. Details of the assumed laws can be found in
114 Table 1.

Table 1. Distribution assumption settings for star samples in the first method. In table, gmag, grav, teff refers to magnitude in g band, surface gravity, efficient temperature,respectively. In the table, feh01 refers to metal abundance in thin disk of galaxy, and feh02 refers to metal abundance in thick disk of galaxy, with the amount ratio setted 4:1. Detail column mainly contain distribution formulas in which x means the parameter, besides ranges for uniform distribution and numbers for constant. A detailed description could be found at <https://docs.scipy.org/doc/scipy/reference/stats.html#module-scipy.stats>

parameter	distribution	detail
ra	uniform	270.3 to 271.7
dec	uniform	-41.7 to -40.3
gmag	exponential	$0 - \exp\left(\frac{3(x+27)}{8}\right)$
grav	uniform	3.6 to 4.6 (in log)
teff	exponential	$\exp\left(\frac{(x-2500)}{1800}\right)$
feh01	chi2	$\frac{1}{2^{k/2}\Gamma(k/2)}\left(\frac{x+0.6}{0.1}\right)^{k/2-1} \exp\left(-\frac{x+0.6}{0.1}/2\right)$
feh02	normal	$\frac{\exp\left(-\left(\frac{x+0.7}{0.25}\right)^2/2\right)}{\sqrt{2\pi}}$
pmra,pmdec,rv	constant	0
parallax	constant	1e-9

115 The second method involves generating a magnitude-complete base catalog and performing statistical
116 reconstruction to create a mock catalog based on this base catalog. In the base catalog generation step, we
117 employ several field star catalogs from Gaia DR3 (Prusti et al., 2016; Vallenari et al., 2023) and displace
118 stars from the neighboring region to the N6397 field. The displacement distance is extended with ensuring
119 that the magnitudes of the stars fall within the rational distribution, surpassing the magnitude limitation
120 of the Gaia catalog. This magnitude-complete catalog serves as the base for further reconstruction. In
121 the statistical reconstruction step, we begin by calculating the covariance matrix to assess the cross-
122 correlations among the parameters. Parameters with strong correlations are grouped together. Next, we
123 derive a multi-parameter probability distribution for each group and interpolate the discrete distributions
124 to obtain smooth and continuous distributions. We then employ Gibbs sampling on these interpolated

125 multi-parameter distributions (as shown in Fig 1). Spatial locations are assumed to be uniformly distributed,
 126 and the resulting catalog is generated by randomly combining all groups of parameters. Compared to the
 127 first method, which is an ideal assumption leading to a notable difference with fact, the second method
 128 containing fitting and sampling is believed to be closer to actual situation.

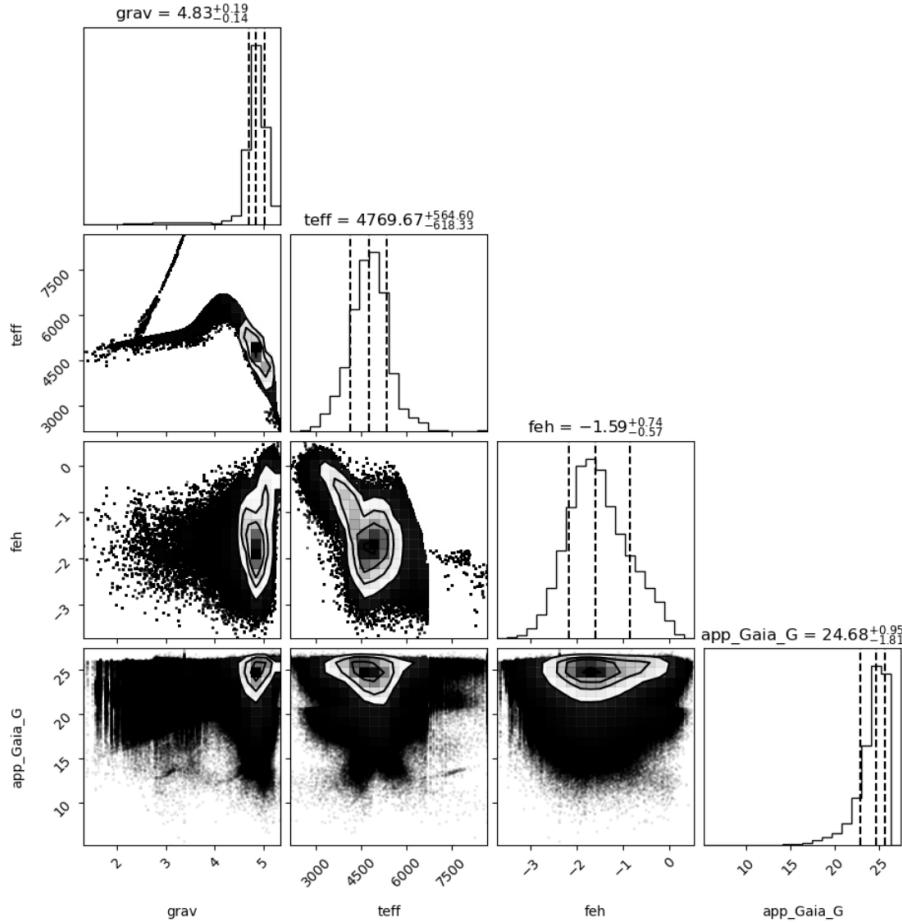


Figure 1. Gibbs sampling of a correlated parameter group, including magnitude in g band, effective temperature, surface gravity and metal abundance. Six subgraphs with triangles in the lower left corner show bivariate distributions, and the other four subgraphs show marginal probability distributions.

129 2.1.2 Observation Simulation

130 We employ CSST-simulation [cite], a software designed for simulating CSST observations, to generate
 131 simulated images and corresponding catalogs. CSST-simulation incorporates measurement effects and
 132 instrument characteristics observed during CSST ground tests. In our work, CSST-simulation takes mock
 133 catalogs and observation information as input, simulates observations for each source by generating
 134 observation stamps, and combines them to create a complete image. Background and effects such as field
 135 distortions, background, dark noise, and flat-fielding are considered, excluding cosmic rays, are then
 136 added to the image. The output consists of a scientific observation image and a corresponding catalog
 137 (sim-catalog) that contains the true positions in the image. In this simulation, star observation is focused
 138 solely on, omitting galaxies. Kinematic and dynamic effects are of minor concern. A complete list of mock
 139 catalogs and simulated images is presented in Table 2, and image examples of simulations with varying

140 density are illustrated in Fig 2. It is important to note that due to the high costs associated with simulating
 141 crowded fields, the number of simulation images in our study is limited.

Table 2. A complete list of mock catalogs and image simulations. N_{sky} means average star number in a sky of $1.4 \times 1.4\text{deg}$. $\bar{N}_{picture}$ means star number in a 9232×9216 picture. $N_{picture}^*$ means star number in practical simulation picture.

N_{sky}	$\bar{N}_{picture}$	method	simulation band	$N_{picture}^*$
1M	18k	01	i	0.16M
8M	146k	01	-	-
300M	5.5M	01	g	5.9M
500M	9.2M	01	-	-
300M	5.5M	02	g	6.5M

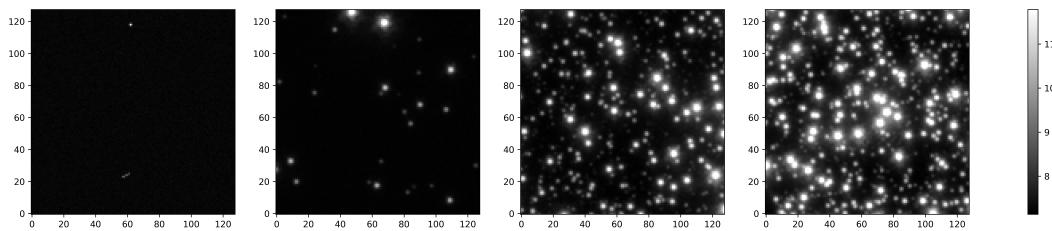


Figure 2. Simulation image examples with different density, cropped in 128×128 size. Subgraphs from left to right represent the whole picture with density of 26k, 0.16M, 5.9M and 6.5M, respectively.

142 2.2 Data Preprocessing

143 For training RefineDet, the entire 9232×9216 pixel image is cropped into small patches with 128×128
 144 pixels. Adjacent patches have a 10-pixel overlap to account for side-truncated stars. The training data
 145 for object detection performance as a matrix of dimensions $n_{object} \times 5$ for each patch. The first four
 146 columns represent location and size information in the form of $[x_{min}, y_{min}, x_{max}, y_{max}]$, while the fifth
 147 column represents the object class index. The location information is extracted from the sim-catalog using
 148 CSST-simulation, and the object class is set to 1 to indicate stars (background is represented by 0). As size
 149 information is not provided in any catalog, we employ a process involving sparse field extraction, matching,
 150 and artificial neural network (ANN) fitting. Using SExtractor (Bertin and Arnouts, 1996) on a low-density
 151 crowded field image of 9232×9216 pixels containing 150,000 stars, we obtain the fitted x, y coordinates,
 152 and iso-aperture areas of each object. To establish matches between truth information from sim-catalog and
 153 the results of SExtractor detection, we consider a true positive detection when the nearest neighbor distance
 154 (containing spatial distance and magnitude distance) is below a certain threshold. Subsequently, we employ
 155 an ANN to fit the sizes, where the ANN takes the magnitude and stellar atmospheric parameters from the
 156 sim-catalog as inputs and the sizes, calculated as $size = (\frac{isoarea}{\pi})^{0.5}$, from the true positive detections
 157 as outputs. The ANN model comprises 6 fully connected layers with sizes of (10-10-20-30-10-1), and
 158 employs the sigmoid activation function. The fitting results are shown in Fig 3.

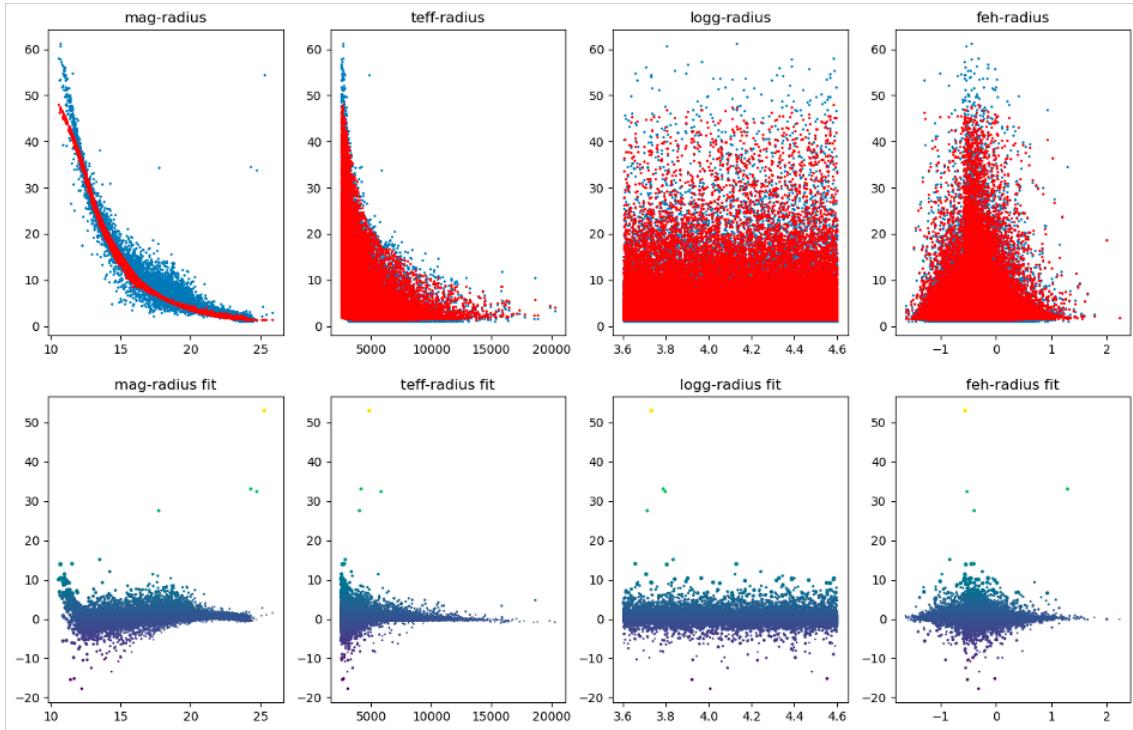


Figure 3. ANN fitting radius with four parameters in catalog. Scatters of single parameter and radius margin distribution and fitting are shown at top (blue means data, red means fitting result). Corresponding fitting residual scatters are shown at bottom.

159 Each picture is divided as one separate dataset, with 6084 cropped patches and corresponding boxes.
 160 Each dataset is split into training and testing sets, a train-test ratio of 4:1 is employed, where 80% of the
 161 patches are used for training and 20% for testing.

3 EXPERIMENT OF STAR EXTRACTION BASED ON REFINEDET

162 3.1 Network Structure

163 RefineDet is a highly effective CNN model for object detection, and performs well for small objects. It
 164 receives 3 channeled input images of size 320×320 or 512×512 and employs VGG-16 (Simonyan and
 165 Zisserman, 2014) as the backbone, incorporating layers 4-3, 5-3, and 6-2 as sources. Anchor refinement
 166 modules (ARMs) is utilized to predict the location and type of bounding boxes. Transfer connection blocks
 167 (TCBs) are employed to combine features from different source layers, while object detection modules
 168 (ODMs) are used for final predictions. Due to the high density of stars in our dataset, the input images
 169 are resized to 128×128 , reducing the number of stars to around 1,000 per image. We experiment with
 170 different combinations of source layers (1-4 or 2-5) and add additional convolutional layers before the
 171 ARM module to investigate the impact on learning efficiency.

172 3.2 Prior Boxes

173 The prediction modules in RefineDet, namely the ARM and ODM, generate offsets with respect to the
 174 prior boxes, where each pixel in one feature map corresponds to an offset holding a one-to-one relationship
 175 with a specific prior box. These prediction boxes need to be decoded using the prior boxes and the final
 176 feature map pixels. Thus channels of the feature map correspond to magnification power of prior boxes

amount. Consequently, it is essential to ensure that the prior boxes are covered by the feeling field of their corresponding pixels. Furthermore, deeper layers can aid in establishing the mapping between the feature map and the offset values for each prior box. The number of prior boxes, their positions, sizes, and spatial coverage collectively form the crucial elements of prior box design, which fundamentally supports our one-stage object detection algorithm.

In crowded field, where the number of objects increases and object overlaps become more prevalent, the design considerations for prior boxes has to evolve accordingly. The primary objective of our design is to guarantee that each real star, regardless of its size and position, is associated with at least one prior box exclusively. To match boxes, the Intersection over Union (IoU) metric is employed, which quantifies the similarity between two bounding boxes. In crowded fields, where stars are densely scattered across a wide range of sizes, the IoU upper limit can be significantly lower when there is a substantial difference in magnification between the star sizes and the prior box sizes. And the IoU can be lower for the random positioning of stars. Therefore the design of prior boxes closely attaches to final performance, specifically for size and magnification power of amount.

To determine the sizes of the prior boxes, radius inverse cumulative distribution of the stars is calculated (see Fig 4). This analysis reveals that over 90% of stars have radius smaller than 3.5 pixels, and approximately 99% of stars have radius smaller than 11 pixels. Based on these findings, we finalize the sizes of the prior boxes as [3, 10, 26, 34] pixels.

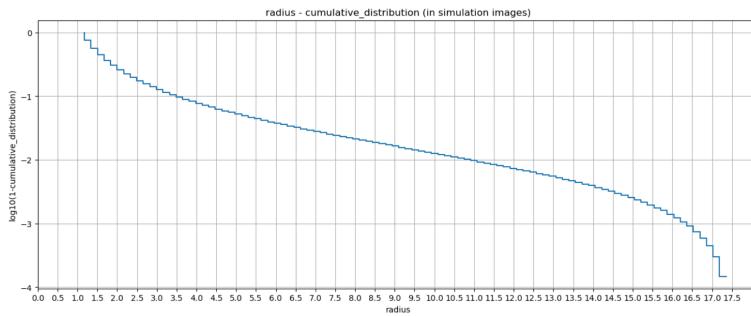


Figure 4. The inverse cumulative probability of stellar size distribution with radius in simulation images, using the 5.9M stars picture as an example. Cumulative distribution probability are in the form of log.

To determine the magnifying power of prior box numbers for each source layer, we introduce a concept called the "ambiguous unit." An ambiguous unit represents the smallest unit of the receptive field, denoting the minimal gap between two adjacent pixels on the feature map. In our experiments, each pooling layer increases the size of the ambiguous unit by a factor of 2. Both convolutional layers and pooling layers result in varying coverage of the ambiguous unit. Additionally, as the side length of a square block comprising ambiguous units increases, the number of feature pixels covering the block will differ. It is worth noting that the coverage of the square block of ambiguous units is, on average, one time when $\text{block side} \cdot \text{ambiguous unit side} = \text{image side}$. Naturally, the object density in the data should be lower than the coverage of both the ambiguous unit and the square block, which can be seen as the minimum magnifying power for determining the number of prior boxes. The relationship between the coverage of a block and the block's side length of ambiguous unit can be expressed as follows:

$$c_{\text{block}} = \left(\frac{s_{\text{ff}}}{s_{\text{au}}} - 1 + \frac{s_{\text{block}}}{s_{\text{au}}} \right)^2 \quad (1)$$

206 Where c_{block} represents the coverage of the block by the feature pixels, s_{ff} is the size of the feeling field,
 207 s_{au} is the size of the ambiguous unit, and s_{block} is the size of the block. The allowed maximum average
 208 density in pixels, d_{pixel} , can be calculated as:

$$d_{pixel} = \frac{c_{block}}{A_{block}} = \frac{c_{block}}{s_{block}^2} \quad (2)$$

209 In this equation, A_{block} represents the area of the block.

210 Fig 5 provides an example of the coverage ratio of a square block and object density, which illustrates
 211 that the magnifying power has a significant lower limit of [1, 2, 5, 10] for layers 1-4. However, due to
 212 factors such as matching overlap, uneven density fluctuations, and other effects, this calculation is not
 213 precise and should be used as a lower limit reference only.

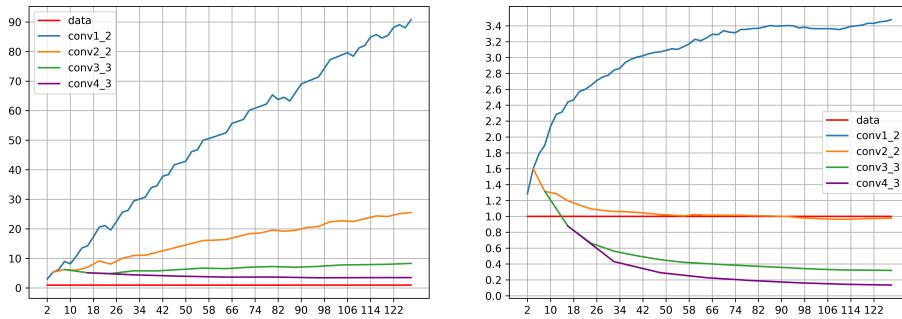


Figure 5. Ratio of the allowed maximum density from theoretical square block coverage and practical object density from statistic, containing data and sources of 1-4 layers respectively. Data of left and right subgraph is from picture with density of 0.16M and 5.9M respectively.

214 3.3 Training Parameters

215 The training process of the model is mainly influenced by four parameters.

216 The variance parameter controls the magnification ratio of the locations (x, y) and sizes (width, height),
 217 which in turn affects the proportion of loss attributed to each component. Since star images are typically
 218 represented as point-like objects, the location is of primary importance, while the size plays a supporting
 219 role. Therefore, this parameter is modified to increase the learning weight of the location.

220 The overlap threshold parameter determines the minimum IoU required for a prior box to be considered a
 221 positive match. This threshold influences the number and spatial coverage of matched prior boxes for each
 222 object. Conducting intermediate examinations can assist for this parameter (example shown in Fig 6).

223 The negpos ratio parameter is responsible for adjusting the process of hard negative mining during
 224 training. It aims to strike a balance between training speed and the learning environment. In a crowded
 225 field with a large number of prior boxes, the negpos ratio is increased to 10 to ensure a more robust training
 226 process.

227 The loss weight parameter is used to assign weights to the four losses in the total loss function. These
 228 losses include the ARM location and size loss, ARM kind loss, ODM location and size loss, and ODM

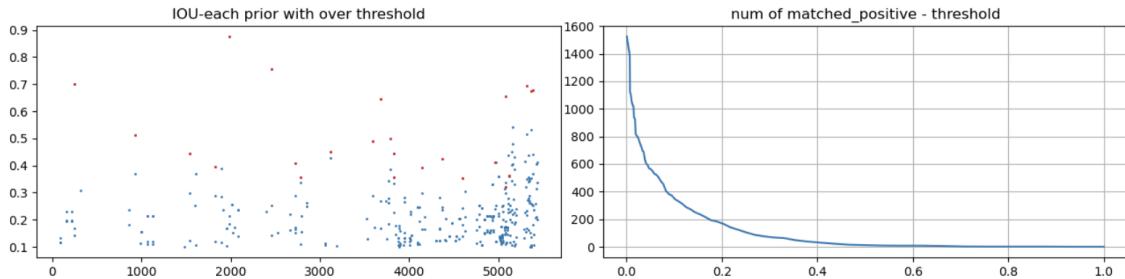


Figure 6. A set of graphs shows overlap between prior boxes and truth boxes in a 128×128 image. The left graph shows scatters of IoU overlap values for each prior, and the right graph shows the relationship between number of matched prior boxes and overlap threshold.

229 kind loss. To achieve a more precise result, a higher loss weight is set for the ODM loss component. This
230 emphasizes the importance of accurate predictions for object detection, leading to improved performance.

231 3.4 Detection and Result

232 The detection progress is governed by two parameters: thresholds and the number of remaining detections
233 based on sorted category scores. These parameters play a crucial role in restricting and selecting prior
234 boxes at different stages of the detection process. It is important to note that when using a Non-Maximum
235 Suppression (NMS) method for detection, prior boxes with an IoU above a certain threshold are eliminated,
236 which will introduce true detection eliminated caused by high star overlaps in crowded fields.

237 In the experiment, the results revealed that training on images with a density of 5.9M or 6.5M stars per
238 picture rendered the training process ineffective, as the loss curve failed to decline. However, training
239 the model on images with a density of 0.16 million stars per picture yielded successful performance. An
240 example of the detection results is presented in Fig 7.

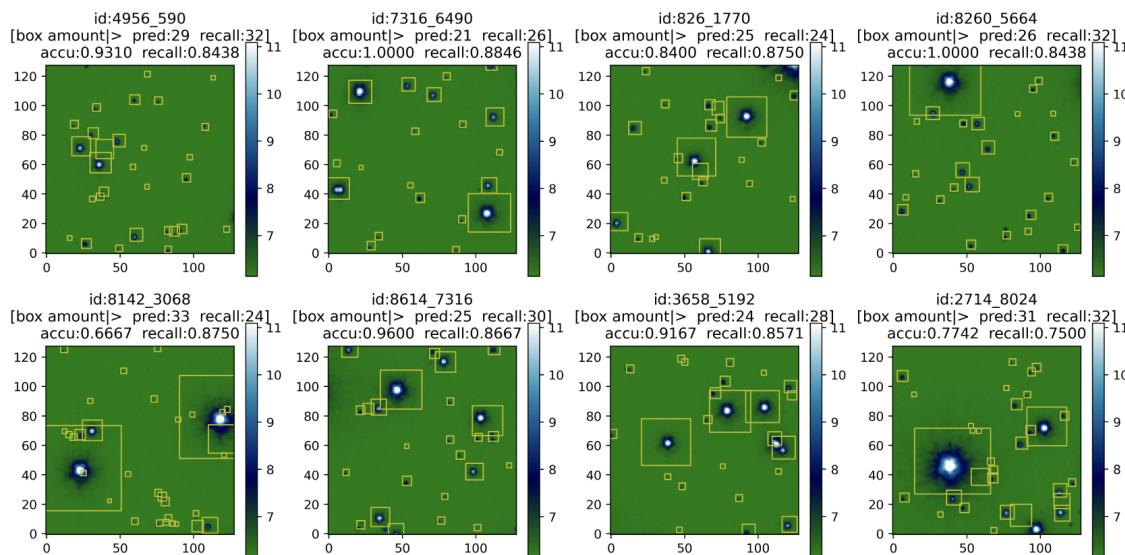


Figure 7. Samples of detection

241 One of the advantages of the approach described in this article is the ability to adjust the detection process
 242 through the use of detection parameters. The effects of different detection parameters on the detection
 243 results are examined in a detailed test, which is presented in Table 3. This table provides a description
 244 of each detection parameter and its potential influence on the detection process. By manipulating these
 245 parameters, the detection can be controlled and optimized for variance situations.

Table 3. Detection parameters description and possible influences while be higher on detection.

parameters	description	influences
objectness	object and background division	candidate boxes reduced
confidence	object believed degree	candidate boxes reduced
NMS	IoU threshold in NMS process	higher recall and lower precision
top_k	remained amount of priors before nms	higher recall
keep_top_k	remained amount of priors as result	limit the amount of result boxes

246 The output of the RefineDet model contains five parts: location of ARM, confidence point of ARM,
 247 location of ODM, confidence point of ODM and prior boxes. Location is constructed by a 4-element box
 248 coordinate array: center, width and height of a box. Confidence point shows the pointing situation predicted
 249 by network model for classification (In this paper, class of object is 0 for background and 1 for stellar).
 250 Prior boxes reserve the the location information of the prior boxes used in this training, and are the key of
 251 encoding and decoding process.

252 Furthermore, a comprehensive evaluation of the detection performance is presented in Fig 8. This figure
 253 provides a detailed distribution of the detection results, including the accuracy and recall distributions. It
 254 also includes a distribution analysis based on the sizes of the detected stars.

255 The results of the evaluation show that the detection algorithm achieves reasonable accuracy and recall
 256 for small, medium, and large objects. However, the performance is notably lower for tiny objects with a
 257 radius smaller than 1.5 pixels. One potential explanation is that the pooling layer preceding the conv1_2
 258 output introduces size confusion at the 2-pixel level. As the radius of the stars increases beyond 1.5 pixels,
 259 the detection performance improves significantly. Notably, large stars with a radius of 10 pixels or more
 260 consistently achieve perfect accuracy and recall of 100%. Overall, the detection algorithm demonstrates
 261 typical accuracy and recall rates of 85% and 80%, respectively.

262 In terms of position precision, the algorithm performs reasonably well with small objects, but it exhibits
 263 poorer precision with larger objects, possibly due to the adverse effects of multiple pooling processes that
 264 reduce spatial sensitivity. By adjusting the parameters of the detection algorithm, the accuracy and recall
 265 for different object sizes can be modified, indicating the adjustability of the detection process.

266 While decreasing density of picture, confusion and complexity will decline rapidly, so further experiment
 267 with lower density field is expected to have a better performance.

268 Additionally, the time costs of the detection process are presented in Table 4. The tests were conducted
 269 on a system equipped with an Intel(R) Xeon(R) Gold 5222 CPU operating at 3.80GHz and an NVIDIA
 270 RTX A4000 GPU. The evaluation was performed on a large picture with dimensions of 9232 × 9216.

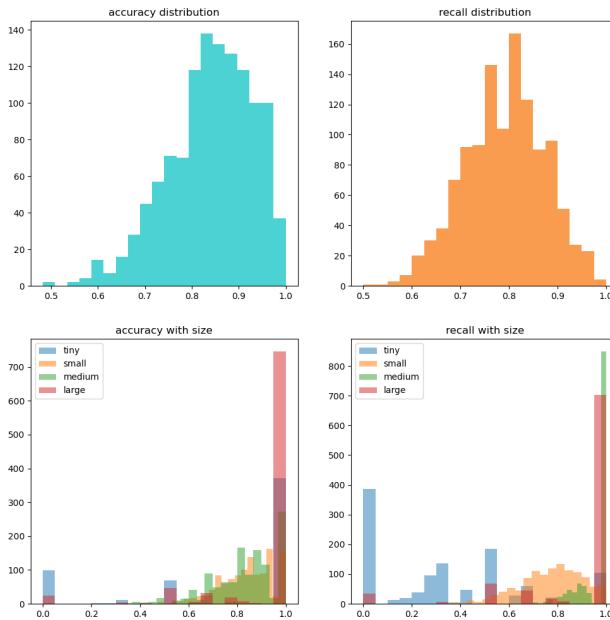


Figure 8. An example of a detection performance distribution on a set of 1216 images. On the top shows the histograms of accuracy and recall. On the bottom shows the different and detailed distribution of detection for different sizes. Sizes of tiny, small, medium, large refer to radius in $(0, 1.5)$, $(1.5, 3)$, $(3, 10)$, $(10, \text{inf})$ pixels, respectively

Table 4. Time costs in a single picture with 9232×9216 pixels.

mod	device	time(s)
train	GPU	29
test	CPU	420
test	GPU	857

4 SUMMARY

271 In this research paper, we have undertaken the generation of a comprehensive dataset that encompasses
 272 simulated images and corresponding catalogs featuring varying densities. The primary objective of this
 273 dataset is to facilitate the evaluation of the performance of the CSST in crowded fields. Additionally, we
 274 have curated a dataset comprising cropped images accompanied by truth boxes, which serve as the ground
 275 truth for both training and evaluating the object detection algorithm of CNN.

276 To overcome the challenges associated with crowded fields, we have made specific adaptations to the
 277 RefineDet algorithm, tailoring it specifically for star detection in densely populated regions. Furthermore,
 278 a novel method for designing prior boxes is developed, taking into account the unique characteristics of
 279 crowded fields. These prior boxes play a pivotal role in guiding the star detection process by providing
 280 anchor points for potential star locations.

281 The results obtained from this article reveal a noticeable decline in the performance of star detection
 282 as the density of stars within the field increases. For instance, in an image containing 0.16 million stars,

283 our algorithm achieves typical precision and recall of 85% and 80%, respectively. This signifies that
284 our algorithm successfully identifies and localizes a substantial portion of the stars in the image while
285 maintaining a reasonable level of accuracy, recall and position precision.

5 DISCUSSION

286 During the course of our experiments, we encountered several challenges related to prior box design, net
287 structure adjustment, fine control of parameters, and the associated time costs. One of the key difficulties was
288 striking the right balance in the detection algorithm. This involves making tradeoffs between objectiveness
289 and the elimination of noisy points, distinguishing between false candidate prior boxes and true overlapped
290 prior boxes, and dealing with both superfluous boxes containing obvious objects and the sole boxes
291 encompassing faint objects. The complexity inherent in crowded fields trusted to be higher than what our
292 CNN model was designed to handle, and we found that the guidance provided to the CNN model was
293 insufficient. Essentially, the model lacked a clear understanding of what it needed to learn and how it
294 should learn in order to accurately match all the truth boxes with their corresponding prior boxes. This was
295 particularly challenging due to the fact that the feeling field would not move with the predicted boxes, and
296 the truth boxes themselves often exhibited characteristics such as faintness, brightness, small size, large
297 size, and overlapping.

298 Further more, the inadaptation between one-stage methods in CNN object detection and crowded fields
299 can be attributed to the following factors:

- 300 1. High Complexity: Crowded fields pose a high level of complexity due to size variance and differences
301 in intensity and overlap among objects.
- 302 2. Fixed Feeling Field: The feeling field, representing the CNN model's receptive field, does not adjust
303 with predicted boxes, lacking a fixed positional relationship between objects and the model.
- 304 3. Prior Box Regression Precision: The precision of prior box regression, which determines the matching
305 with ground truth objects, is not always high, leading to confusion.
- 306 4. Pooling and Spatial Reliance: The pooling operation in CNNs reduces spatial reliance, resulting in
307 decreased precision for detecting larger objects and loss for tiny objects in crowded fields.

308 One possible avenue for future work is to explore a more elaborate design or alternative approaches. For
309 example, the introduction of a multi-parameter setting group could potentially enhance the accuracy of the
310 detection algorithm in different situations. Additionally, a spatially optimized design of the prior boxes
311 could lead to more effective matching of stars.

312 Another area of investigation is the implementation of an autoregulatory threshold mechanism based on
313 the surrounding density. Such a mechanism could help mitigate confusion in the detection process, and a
314 more detailed threshold design would likely prove beneficial in this regard.

315 Alternatively, a different approach could involve devising a system-designed method that divides the
316 detection process into multiple steps. This approach would rely on both pre-processing and post-processing
317 techniques to reduce complexity and better guide the decision-making of the neural network.

318 In conclusion, future research should focus on refining the design of the algorithm and exploring
319 alternative approaches to tackle the challenges posed by crowded fields. By incorporating innovative
320 strategies such as improved parameter settings, optimized spatial design, adaptive threshold mechanisms,

321 and system-designed methodologies, the accuracy and efficiency of star detection in crowded field are
322 expected to be enhanced.

AUTHOR CONTRIBUTIONS

323 YZ H is responsible for the implementation of research. XY P organized the entire project and topic, and
324 provided key guidance and full support throughout the process. ZX Q firstly proposed this mission and
325 provided precious theoretical references, critical material, and solid supports. F L firstly advised RefineDet
326 and supported guidance and assistant in the field of neural network. SL L, Y Y and ZH T kindly assist and
327 support this work, providing valuable theoretical support and direction guidance. CSST Simulation Group
328 support simulation software support. ZY S provided the base catalog for the mock catalog generation. QQ
329 W, ZS F and WF F participated in method discussion during this research.

FUNDING

330 We acknowledge the science research grants from the China Manned Space Project with NO.CMS-CSST-
331 2021-A12, NO. CMS-CSST-2021-B10, and NO.CMS-CSST-2021-B04.

ACKNOWLEDGMENTS

332 The observation simulation work is based on CSST-Sim in the scientific data processing
333 software system of the Chinese Space Station Telescope under the China Manned Space Project
334 (https://csst-tb.bao.ac.cn/code/csst_sim/csst_simulation/-/tree/release_v2.0/)

REFERENCES

- 335 Anderson, J. and King, I. R. (2000). Toward high-precision astrometry with wfpc2. i. deriving an accurate
336 point-spread function. *Publications of the Astronomical Society of the Pacific* 112, 1360
337 Bertin, E. and Arnouts, S. (1996). SExtractor: Software for source extraction. *Astronomy and astrophysics*
338 *supplement series* 117, 393–404
339 Budavári, T. and Basu, A. (2016). Probabilistic cross-identification in crowded fields as an assignment
340 problem. *The Astronomical Journal* 152, 86
341 Connor, T., Donahue, M., Kelson, D. D., Moustakas, J., Coe, D., Postman, M., et al. (2017). Crowded field
342 galaxy photometry: precision colors in the clash clusters. *The Astrophysical Journal* 848, 37
343 Diolaiti, E., Bendinelli, O., Bonaccini, D., Close, L. M., Currie, D. G., and Parmeggiani, G. (2000).
344 Starfinder: an idl gui-based code to analyze crowded fields with isoplanatic correcting psf fitting. In
345 *Adaptive Optical Systems Technology* (SPIE), vol. 4007, 879–888
346 Feder, R. M., Portillo, S. K., Daylan, T., and Finkbeiner, D. (2020). Multiband probabilistic cataloging: a
347 joint fitting approach to point-source detection and deblending. *The Astronomical Journal* 159, 163
348 Feinstein, C., Baume, G., Rodríguez, J., and Vergne, M. (2019). Extragalactic stellar photometry and the
349 blending problem. *arXiv preprint arXiv:1903.10989*
350 Gallego-Cano, E., Schödel, R., Gallego-Calvente, A., and Ghez, A. (2022). Accurate uncertainty estimation
351 in crowded fields: adaptive optics and speckle data. *arXiv preprint arXiv:2208.05483*
352 Hausen, R. and Robertson, B. E. (2020). Morpheus: A deep learning framework for the pixel-level analysis
353 of astronomical image data. *The Astrophysical Journal Supplement Series* 248, 20
354 Kramer, D. M., Carleton, T., Cohen, S. H., Jansen, R., Windhorst, R. A., Grogin, N., et al. (2022).
355 Skysurf-3: Testing crowded object catalogs in the hubble extreme deep field mosaics to study sample

- 356 incompleteness from an extragalactic background light perspective. *The Astrophysical Journal Letters*
357 940, L15
- 358 Liu, R., McAuliffe, J. D., and Regier, J. (2021). Variational inference for deblending crowded starfields.
359 *arXiv preprint arXiv:2102.02409*
- 360 Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., et al. (2016). Ssd: Single shot multibox
361 detector. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands,
362 October 11–14, 2016, Proceedings, Part I 14* (Springer), 21–37
- 363 Magain, P., Courbin, F., Gillon, M., Sohy, S., Letawe, G., Chantry, V., et al. (2007). A deconvolution-based
364 algorithm for crowded field photometry with unknown point spread function. *Astronomy & Astrophysics*
365 461, 373–379
- 366 Melchior, P., Joseph, R., Sanchez, J., MacCrann, N., and Gruen, D. (2021). The challenge of blending in
367 large sky surveys. *Nature Reviews Physics* 3, 712–718
- 368 Neichel, B., Lu, J. R., Rigaut, F., Ammons, S. M., Carrasco, E. R., and Lassalle, E. (2014). Astrometric
369 performance of the gemini multiconjugate adaptive optics system in crowded fields. *Monthly Notices of
370 the Royal Astronomical Society* 445, 500–514
- 371 Paillassa, M. and Bertin, E. (2019). Deblending in crowded star fields using convolutional neural networks.
372 *Astronomical Data Analysis Software and Systems XXVI* 521, 382
- 373 Portillo, S. K., Lee, B. C., Daylan, T., and Finkbeiner, D. P. (2017). Improved point-source detection in
374 crowded fields using probabilistic cataloging. *The Astronomical Journal* 154, 132
- 375 Prusti, T., De Bruijne, J., Brown, A. G., Vallenari, A., Babusiaux, C., Bailer-Jones, C., et al. (2016). The
376 gaia mission. *Astronomy & astrophysics* 595, A1
- 377 Redmon, J. and Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv preprint
378 arXiv:1804.02767*
- 379 Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster r-cnn: Towards real-time object detection with
380 region proposal networks. *Advances in neural information processing systems* 28
- 381 Shi, X., Budavári, T., and Basu, A. (2019). Probabilistic cross-identification of multiple catalogs in
382 crowded fields. *The Astrophysical Journal* 870, 51
- 383 Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image
384 recognition. *arXiv preprint arXiv:1409.1556*
- 385 Stetson, P. B. (1987). Daophot: A computer program for crowded-field stellar photometry. *Publications of
386 the Astronomical Society of the Pacific* 99, 191
- 387 Vallenari, A., Brown, A., Prusti, T., de Bruijne, J., Arenou, F., Babusiaux, C., et al. (2023). Gaia data
388 release 3-summary of the content and survey properties. *Astronomy & Astrophysics* 674, A1
- 389 Zhan, H. (2021). The wide-field multiband imaging and slitless spectroscopy survey to be carried out by
390 the survey space telescope of china manned space program. *Chinese Science Bulletin* 66, 1290–1298
- 391 Zhang, S., Wen, L., Bian, X., Lei, Z., and Li, S. Z. (2018). Single-shot refinement neural network for
392 object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*.
393 4203–4212
- 394 Zheng, C., Pulido, J., Thorman, P., and Hamann, B. (2015). An improved method for object detection in
395 astronomical images. *Monthly Notices of the Royal Astronomical Society* 451, 4445–4459