



## Full length article

## Radio sources segmentation and classification with deep learning

B. Lao<sup>a,\*</sup>, S. Jaiswal<sup>b</sup>, Z. Zhao<sup>c</sup>, L. Lin<sup>d</sup>, J. Wang<sup>d</sup>, X. Sun<sup>a</sup>, S.-L. Qin<sup>a</sup><sup>a</sup> School of Physics and Astronomy, Yunnan University, Kunming, 650091, China<sup>b</sup> Shanghai Astronomical Observatory, Chinese Academy of Sciences, Shanghai, 20030, China<sup>c</sup> School of Electrical and Information Engineering, University of Sydney, Sydney, 2006, Australia<sup>d</sup> School of Information and Communication, Guilin University Of Electronic Technology, Guilin, 541004, China

## ARTICLE INFO

## Article history:

Received 14 February 2023

Accepted 30 May 2023

Available online 10 June 2023

Dataset link: <https://third.ucllnl.org>, <https://github.com/lao19881213/RGC-Mask-Trans-finer>

## Keywords:

Radio continuum survey

Radio sources

Image segmentation

Morphological classification

Deep learning

## ABSTRACT

Modern large radio continuum surveys have high sensitivity and resolution, and can resolve previously undetected extended and diffuse emissions, which brings great challenges for the detection and morphological classification of extended sources. We present HeTu-v2, a deep learning-based source detector that uses the combined networks of Mask Region-based Convolutional Neural Networks (Mask R-CNN) and a Transformer block to achieve high-quality radio sources segmentation and classification. The sources are classified into 5 categories: Compact or point-like sources (CS), Fanaroff–Riley Type I (FRI), Fanaroff–Riley Type II (FRII), Head–Tail (HT), and Core–Jet (CJ) sources. HeTu-v2 has been trained and validated with the data from the Faint Images of the Radio Sky at Twenty-one centimeters (FIRST). We found that HeTu-v2 has a high accuracy with a mean average precision ( $AP_{50:5:95}$ ) of 77.8%, which is 15.6 points and 11.3 points higher than that of HeTu-v1 and the original Mask R-CNN respectively. We produced a FIRST morphological catalog (FIRST-HeTu) using HeTu-v2, which contains 835,435 sources and achieves 98.6% of completeness and up to 98.5% of accuracy compared to the latest 2014 data release of the FIRST survey. HeTu-v2 could also be employed for other astronomical tasks like building sky models, associating radio components, and classifying radio galaxies.

© 2023 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The modern large radio telescopes, e.g. the Square Kilometre Array (SKA) pathfinders like Low Frequency Array (LOFAR; [van Haarlem et al., 2013](#)), Murchison Widefield Array (MWA; [Tingay et al., 2013](#)), and Australian SKA Pathfinder (ASKAP; [Hotan et al., 2021](#)), are designed with the novel technologies, such as phased array, digital beamforming, or Phased Array Feed (PAF), and thus have wide field of view, high resolution, high sensitivity, and broad frequency coverage. The radio continuum surveys carried out by these telescopes have an effective combination of depth and area, and can detect a lot of previously unseen or unresolved emissions ([Dabhade et al., 2020](#); [Delhaize et al., 2021](#); [Andernach et al., 2021](#)). The new detection presents various complicated morphologies beyond simple point-like sources. Examples are the LOFAR Two-metre Sky Survey (LoTSS; [Shimwell et al., 2017](#)), GaLactic and Extragalactic All-sky MWA eXtended (GLEAM-X) survey ([Hurley-Walker et al., 2022](#)), MeerKAT International GHz Tiered Extragalactic Exploration (MIGHTEE) survey ([Jarvis et al., 2016](#)), ASKAP Evolutionary Map of the Universe (EMU) survey ([Norris et al., 2011](#)), Apertif imaging survey ([Adams et al., 2022](#)), and Karl G. Jansky Very Large Array Sky

Survey (VLASS; [Lacy et al., 2020](#)). The EMU survey, as an instance, is expected to produce a catalog containing about 70 million radio sources ([Norris et al., 2011](#)). With this, considerable efforts in recent years have focused on developing algorithms for imaging and source-finding in a rapid, reliable, and automated way ([Lao et al., 2019](#); [Wieringa et al., 2020](#); [Hancock et al., 2018](#); [Riggi et al., 2016](#)).

In radio astronomy, a source-finding software, also called source finder, is typically used to identify and extract radio sources from images. The source finders widely used in existing surveys are Aegean ([Hancock et al., 2012, 2018](#)), Python Blob Detection and Source Finder (PyBDSF; [Mohan and Rafferty, 2015](#)), and SeLavy ([Whiting and Humphreys, 2012](#)); and they all form catalogs by fitting Gaussian components to sources. In the fitting procedure, compact or point-like sources are fitted with a single Gaussian component, and extended sources are fitted with a combination of Gaussian components. This approach works well for compact or point-like sources but is not necessarily true for extended sources because there can be over-fitting of components where Gaussians are not as appropriate for the emission shape ([Hale et al., 2019](#)). To address this problem, several new source finders have been proposed, such as Compact And Extended Source Automated Recognition (CAESAR; [Riggi et al., 2016, 2019](#)) and ProFound ([Robotham et al., 2018](#); [Hale et al., 2019](#)).

\* Corresponding author.

E-mail address: [lbq19881213@gmail.com](mailto:lbq19881213@gmail.com) (B. Lao).

Both source finders are developed based on the pixel segmentation technique and specifically designed for the extraction of extended sources, but their performances are still not as good as that achieved for compact sources. In particular, they cannot work for the extended source that has two-lobed jets with a large separation (Hale et al., 2019).

With the artificial intelligence awakening and developments of Convolutional Neural Networks (CNN) in recent times, deep learning algorithms have been widely used as an automated approach in many specific areas of radio astronomy (Mesarcik et al., 2020), including astronomical source detection (Lukic et al., 2019; Vafaei Sadr et al., 2019; Wu et al., 2019; Burke et al., 2019; Arslan, 2020; Farias et al., 2020; Lao et al., 2021; Riggi et al., 2023). The source-finding methods based on deep learning, called source detectors, allow automatic localization and classification of sources, and their detection speeds are significantly faster than the standard source finders by utilizing stronger computing power, viz. Graphics Processing Units (GPUs) or Neural Processor Units (NPUs). Moreover, a source detector performs classification of sources based on their morphologies, and the separate components of larger extended sources are automatically linked and classified. The existing source detectors, however, have some limitations on the grounds of accuracy and performance for extended sources. For example, the accuracy and performance of CLARAN (Wu et al., 2019) are limited by the lack of layers of a backbone network it uses.

Lao et al. (2021) developed a source detector HeTu (to be termed as HeTu-v1) using Faster Region-based CNN (Faster R-CNN; Ren et al., 2017). The HeTu-v1 is able to locate and classify four classes of radio sources, but its accuracy is limited because of the approximation method used in the network to get the bounding boxes of the sources, and fewer samples of the extended sources used in the model training. Recently, it has been found that the methods (e.g. Burke et al., 2019; Arslan, 2020; Farias et al., 2020; Riggi et al., 2023) using Mask R-CNN (He et al., 2017) have excellent performance in localization of the bounding boxes, as well as provide segmentation masks for each source that is implemented similar to the pixel segmentation-based source finders. However, the segmentation masks obtained by the original Mask R-CNN based software are still coarse and cannot satisfy scientific research. The quality of the segmentation can profoundly impact the accuracy of the flux density estimation of radio sources, especially for extended sources. Thus, high-quality segmentation is critical for source-finding in modern radio surveys.

This paper focuses on the upgrade of HeTu-v1 into HeTu-v2 using the combination network of Mask R-CNN and state-of-the-art object detection algorithm *i.e.* Transformer block (Carion et al., 2020) to achieve the high-quality segmentation and morphological classification of radio sources. HeTu-v2 can perform multiple astronomical tasks and has distinct advantages over other source detectors:

**Building morphological catalog.** The main function of HeTu-v2 is to build a catalog from radio image. Each source in the catalog is described by the detected information (class name, mask, bounding box, and score) and standard properties (flux density and location) of the source. The morphological classification covers five categories (see Section 3.1.1). The cataloged properties include Gaussian-fit information for the compact sources and deep-learning estimated information for the extended sources. Such a catalog allows its direct use in statistical analysis and astronomical scientific research of specific classes.

**Building accurate sky models.** HeTu-v2 can produce high-quality segmentation, which can be used to build a sky model that is typically used in the radio data processing pipeline like ProFound and CAESAR, and avoids the Gaussian over-fit when modeling the extended emission. More importantly, HeTu-v2 can automatically associate the separated radio lobes to build a complete model, which is more accurate than ProFound and CAESAR.

**Automatic associate radio components.** HeTu-v2 can be extended to find the radio components associated with a radio source in an automated way from the components-based catalog. The association method is based on the predicted segmentation mask of the radio source that is generated by HeTu-v2, and therefore the result is more accurate than the methods based on the bounding box. This greatly benefits in finding the corresponding host galaxies of radio sources and inferring their physical parameters.

**Classifying radio galaxies.** If there are a number of input radio images and each input image is considered to have exactly one interesting source, it is troublesome to morphologically classify the source in each image manually through visual inspection when using traditional source finders. HeTu-v2 can, however, be used as a classifier to automatically classify compact and extended radio galaxies into the five classes mentioned above. Different classes of radio galaxies can be used as tracers of the cosmic environment to address key cosmological questions. It is an important step towards achieving the scientific goals of modern radio surveys.

The rest of this paper is organized as follows: Section 2 provides a brief overview of the deep learning based source detectors. Section 3 introduces the framework and network of HeTu-v2, the dataset and its related methods including data generation, the data augmentation, the new morphological classification scheme, and the HeTu-v2 model training. The results of the evaluation and inference of the HeTu-v2 model are given in Section 4. Section 4 also describes the application prospects of HeTu-v2. Finally, Section 5 concludes the paper based on the current state of HeTu-v2.

## 2. Source detectors based on deep learning methods

Due to the advantages of deep learning methods in terms of their running speed and accuracy, many automated source detectors based on deep learning models have been developed in recent times mainly along two directions: either using a particular CNN architecture (Gheller et al., 2018; Lukic et al., 2019; Vafaei Sadr et al., 2019) or an object detection algorithm (Wu et al., 2019; Lao et al., 2021; Arslan, 2020; Burke et al., 2019; Farias et al., 2020; Zhang et al., 2022; Riggi et al., 2023).

In CNN architecture based methods, COSMODEEP (Gheller et al., 2018) is a classifier that first divides the input simulated radio image into a number of tiles, performs the source detection by classifying each tile as containing or not a radio source, and then mosaicing all tiles with the classification results to form the final detection. Such a method only detects the parts of radio sources, and lacks classification capability as it only performs binary classification. ConvoSource (Lukic et al., 2019) first reconstructs a solution map of the input real map (e.g. simulated SKA image), and then determines the locations of predicted sources by post-processing using reconstruction threshold. DeepSource (Vafaei Sadr et al., 2019) first generates a high signal-to-noise ratio image of the input simulated radio image, then finds the best threshold, detects blobs, and assigns one point source to each blob using

Thresholded Blob Detection (TBD) technique. Both ConvoSource and DeepSource are only used in image processing, not to directly perform source detection. Furthermore, ConvoSource and DeepSource are not able to extract the high-level features which can be a problem in the case of more complicated or faint sources as they are implemented in the basics of CNNs. In summary, the described CNN architecture based methods only utilize ideal simulated images and lack of practical application to real complex observational data.

To improve the CNN architectures, people have adopted the object detection algorithms that make use of Region Proposal Network (RPN; Ren et al., 2017). CLARAN (Wu et al., 2019) is based on the Faster R-CNN (Ren et al., 2017) framework and uses 16 layers of the Visual Geometry Group Network (VGGNet-16) as the backbone, which can automatically locate and classify individually separate and extended components of radio sources on the input radio image and its counterpart infrared image. Generally, the deeper the network, the more accurate the predictions, but the VGGNet only supports up to 19 layers on account of vanishing gradient and exploding gradient during the increase in the number of layers. Moreover, CLARAN only uses the highest feature map, and therefore it loses local details of objects from the lower feature maps that causes lower accuracy for multi-sources detection.

HeTu-v1 (Lao et al., 2021) is based on the improved Faster R-CNN algorithm using the backbone network architecture of the Residual Network (ResNet; He et al., 2016) with Feature Pyramid Network (FPN; Lin et al., 2017). ResNet adopts the back-propagation algorithm, which avoids the problems of vanishing gradient and exploding gradient, and the network can be deeper than VGGNet; 101 layers of ResNet (ResNet101) have been used in HeTu-v1. The combination of ResNet and FPN through the bottom-up pathway, top-down pathway, and lateral connection design allows extraction of the multi-scale feature maps that not only ensures the advantages of small object detection but also improves the detection accuracy of complex objects. However, the methods based on Faster R-CNN have limited positioning accuracy due to the use of Region of Interest (RoI) Pooling for double rounding. For this, high positioning accuracy methods have been developed by using Mask R-CNN (He et al., 2017), which improves the positioning accuracy by using the RoI Align method instead of using the RoI Pooling method and adding a branch to generate pixel masks for objects (He et al., 2017).

Arslan (2020) used Mask R-CNN to implement a source detector on the same dataset defined in CLARAN, and showed that it can classify radio source morphologies and generate masks at the same time. Astro R-CNN (Burke et al., 2019) applies Mask R-CNN to perform star and galaxy detection, classification, and deblending on simulated images. Mask Galaxy (Farias et al., 2020) also uses Mask R-CNN to perform detection, segmentation, and morphological classification for only one class of (Spiral or Elliptical) galaxy. However, Astro R-CNN and Mask Galaxy methods focus on processing optical images.

Caesar-mrcnn (Riggi et al., 2023) is a Mask R-CNN based source detector which is able to detect and classify compact, extended, spurious, and poorly imaged sources in radio continuum images. However, Caesar-mrcnn performs poorly on the extended sources due to the ambiguity of the segmentation masks obtained by the Mask R-CNN, and the morphological classification scheme of Caesar-mrcnn lacks astrophysical relevance for most extended sources. Furthermore, the current version of Caesar-mrcnn does not support building catalogs like the standard source finders, such as compact source Gaussian fitting and extended source properties estimation which are the essential features of a source detector.

As shown above, the implementations of Mask R-CNN have been found to provide powerful detectors that are well-suited for

executing source-finding and classification in radio surveys. However, it should be noted that they provide less accurate predicted masks in the boundary regions of objects (Tang et al., 2021), which is a crucial factor for the segmentation of high-resolution extended sources. Hence, the main contribution of this work is to explore a high-quality segmentation approach based on the improved Mask R-CNN that provides an appropriate baseline for future works.

### 3. Data and method

#### 3.1. Classification scheme and data

##### 3.1.1. Morphological classification scheme of HeTu-v2

The morphological classification of radio sources is a method of grouping sources based on their visual appearance. Studying the morphological classification of radio sources is able to reveal the physical properties of the sources, allowing us to achieve many critical scientific goals, such as the cosmological evolution of radio galaxy counts and luminosities (van Velzen et al., 2012), formation and evolution of radio galaxies and peculiar morphological radio sources (e.g. Super Massive Black Holes (SMBH) and Active Galactic Nuclei (AGN) feedback (Fabian, 2012; Padovani et al., 2017)), and identifying low brightness diffuse radio galaxies (e.g. nearby Star-Forming Galaxies (SFG; Hopkins, 2004)).

The radio sky usually includes a variety of compact point-like and extended sources. A compact source (CS) is an unresolved source with a single non-diffuse emission component or is a point-like source, and an extended source is a resolved source that has at least one diffuse component. Traditionally, the extended sources have been classified as Fanaroff–Riley type I (FRI) and Fanaroff–Riley type II (FR II) sources using the Fanaroff–Riley scheme (Fanaroff and Riley, 1974). In this scheme, the ratio of the distance between the regions of the highest brightness on opposite sides (radio jets) of the central core, to the total extent of the source up to the lowest brightness, is used to distinguish these two classes. Sources with ratios less than 0.5 are defined as FRI, and those with ratios greater than 0.5 are defined as FR II.

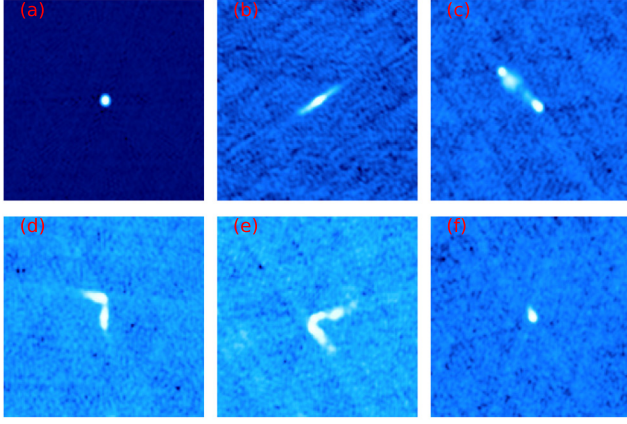
Besides the typical FR sources, there is a morphological subclass called bent-tailed sources or Head-Tail (HT) sources that present more complex bent morphologies of the two radio jets caused by environmental factors or intrinsic properties (Proctor, 2011). The bending of the two jets forms an opening angle of less than  $180^\circ$  at the center of the source, giving the overall shape of the radio source like a “L”, “C” or “V” shape. Depending on the size of the opening angle, HT sources can be further classified into two sub-classes. If the opening angle of the sources is greater than or equal to  $90^\circ$ , the source is classified as Wide-Angle Tailed (WAT); if the opening angle is less than  $90^\circ$ , the source is classified as Narrow-Angle Tailed (NAT) source. HT sources are generally found in galaxy clusters in the local Universe and can be used in observations as possible tracers of galaxy clusters at high redshifts ( $z \geq 1$ ) (Blanton et al., 2003; Dehghan et al., 2014).

In addition to the above four classes, there is another category of sources with a prominent compact component and one-sided weak extended feature. We called it as Core-Jet (CJ) sources. CJ sources form a considerable group and their radio emission characteristics are obviously different from FR sources, so this category also needs to be taken into account. In summary, we refine the morphological classification scheme of HeTu-v1 (i.e. CS, FRI, FR II, and CJ) and consider five morphological classes in this work: CS, FRI, FR II, HT, and CJ. Example images of the five classes are shown in Fig. 1.

##### 3.1.2. Data

We used the images from the VLA FIRST radio survey (Becker et al., 1995), and the catalog from the latest data release version 14Dec17 (FIRST-14dec17) that contains 946,432 sources at





**Fig. 1.** Examples of morphological classes. From panels (a) to (f) are CS, FRI, FRII, WAT, NAT, and CJ sources. Panels (d) and (e) are combined as HT sources in this work.

a detection limit of 1 mJy (Helfand et al., 2015). FIRST-14dec17 catalog was produced by the Astronomical Image Processing System (AIPS; Greisen, 1990) based source-finding program HAPPY (White et al., 1997) whose key algorithm is similar to Aegean and PyBDSF. The FIRST images were downloaded and pre-processed, and then input to the trained model of HeTu-v1 for preliminary source finding and classification. Next, the training and validation datasets of HeTu-v2 were generated based on preliminary results using the new morphological classification scheme (see Section 3.1.1). Finally, all processed images were used for HeTu-v2 inference experiments (see Section 4.2). This section describes data download, data pre-processing, conducting preliminary experiments, and building training and validation datasets.

To speed up the image downloads, we have implemented a Message Passing Interface (MPI)-based program to batch download the FITS images in parallel using multiple processes. In this program, the FIRST-14dec17 catalog has been equally divided into multiple sub-catalogs by the total number of MPI processes used. Then, FITS images of these sub-catalogs are downloaded in parallel. In each download operation, the coordinates of each source in the sub-catalog have been extracted as the image center of the FITS image and downloaded this image from the FIRST cutout webpage (Haridas et al., 2005) using the Python `request.post` method. The size of each FITS image is  $3.96 \times 3.96$  arcmins (*i.e.*  $132 \times 132$  pixels) in equinox of J2000. There are a total of 946,366 FITS images available for download, all are selected to flow preliminary experiments.

The image pre-processing is performed to generate the PNG images from the original FITS images as the PNG image is one of the essential formats for the input of the HeTu model. In the pre-processing, all PNG images are generated using the log-max-min color scale and ‘cool’ colormap. Following Joye and Mandel (2003) the log-max-min color scale is given in Eq. (1):

$$I_{\text{out}} = \frac{1}{3} \log_{10} \left( \frac{I_{\text{in}} - I_{\text{min}}}{I_{\text{width}}} 1000 + 1 \right), \quad (1)$$

where  $I_{\text{width}} = I_{\text{max}} - I_{\text{min}}$  represents the width pixel values in an image,  $I_{\text{max}}$  and  $I_{\text{min}}$  refer to the maximum and minimum pixel value in the image, respectively.  $I_{\text{out}}$  and  $I_{\text{in}}$  represent the final output and original input pixel values in the image, respectively.

The pre-processed PNG images were fed into the radio source detector HeTu-v1 for source finding and classification. The trained model of HeTu-v1 is the same as that used in the final testing experiments from our previous work (Lao et al., 2021). This trained model can automatically identify and locate sources with the

**Table 1**

Number of sources in training and validation datasets.

Class	Training	Rotation-augmented training	Validation
CS	1673	82,105	894
FRI	326	16,313	217
FRII	738	36,577	479
HT	406	20,466	270
CJ	458	23,132	304
Total	3601	178,593	2164

bounding box, label name, probability score, and other physical information from input images. These parameters are presented in the final catalog in CSV format. Finally, a total of 955,208 sources are detected by HeTu-v1 model, including 852,271 CS sources, 11,134 FRI sources, 35,904 FRII sources, and 55,719 CJ sources. Based on this result, new training and validation datasets can be quickly built through visual inspection and automated ways.

Based on the new classification scheme in Section 3.1.1, we utilized visual inspection to select images for HeTu-v2’s training and validation datasets from the above preliminary results generated by HeTu-v1. This involved carefully selecting images and fixing any annotation errors, such as reclassifying false positives and adjusting predicted bounding box positions. Finally, a total of 5118 corrected and annotated images have been selected through this process. Among these, 3172 images contained 3601 sources for training, while 1946 images contained 2164 sources for validation. Details of the two datasets are shown in Table 1.

### 3.2. Method

Radio sources segmentation in deep learning is called instance segmentation which can be roughly divided into two-stage (*e.g.* Mask R-CNN; He et al., 2017), one-stage (*e.g.* You Only Look At CoefficientTs, YOLCAT; Bolya et al., 2019), and query-based (*e.g.* Segmenting Objects by Learning Queries, SQLQ; Dong et al., 2021) methods. However, the quality of segmentation generated by using any of these methods alone is still unsatisfactory.

Recently, inspired by the powerful capability of Transformers (Carion et al., 2020) in object detections, a hybrid method, Mask Transfne (Ke et al., 2022), was proposed for high-quality instance segmentation. On top of the Mask R-CNN, the image regions have been decomposed to build a hierarchical quadtree, which is subsequently fed into a Transformer block to predict the final mask labels. Benefiting from the Transformer-based refining, significant performance improvements have been achieved on large-scale natural image datasets such as the Common Objects in Context (COCO; Lin et al., 2014), Cityscapes,<sup>1</sup> and BDD100K.<sup>2</sup>

In the images obtained from radio observations, the boundary of the sources is more smooth and not as sharp as in the natural images. Thus, correctly segmenting radio sources is more challenging, especially in the presence of noise. Inspired by Mask Transfner, we embark on the ambiguity areas and exploit the long-distance dependencies of the Transformer block to refine the segmentation boundary obtained from the Mask R-CNN. As can be seen from below, significant improvements are achieved.

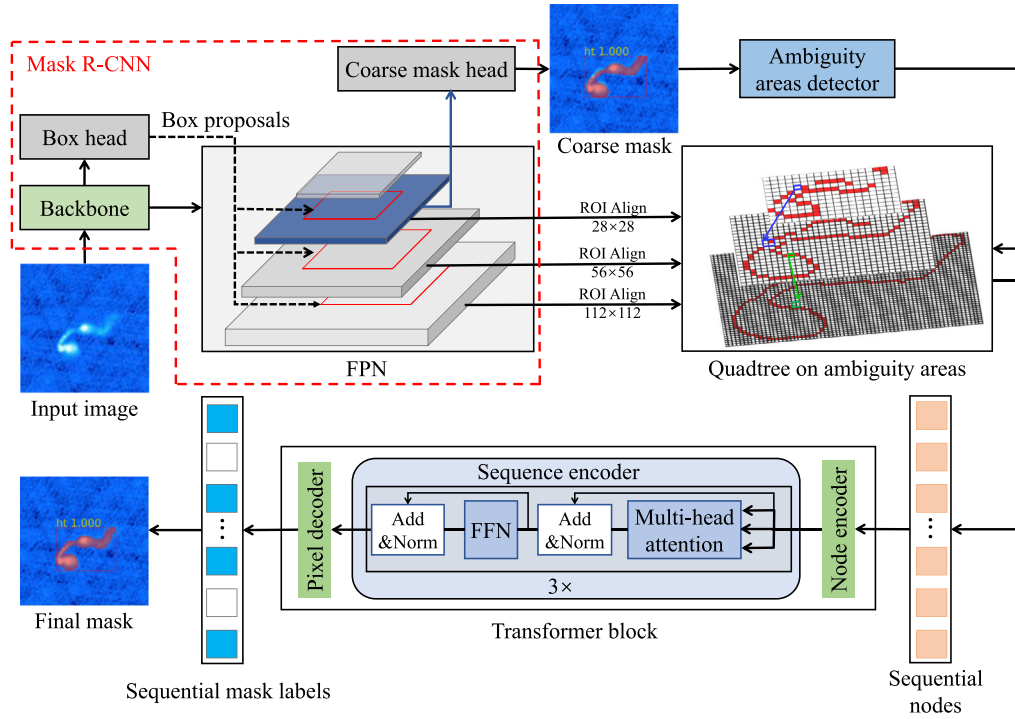
#### 3.2.1. Framework of HeTu-v2

HeTu-v2 implements Mask Transfne (Ke et al., 2022) which is built on the Pytorch (Paszke et al., 2019) based object detection library DETECTRON2.<sup>3</sup> In the training phase HeTu-v2 processes

<sup>1</sup> <https://www.cityscapes-dataset.com/>.

<sup>2</sup> <https://www.bdd100k.com/>.

<sup>3</sup> <https://github.com/facebookresearch/detectron2>.

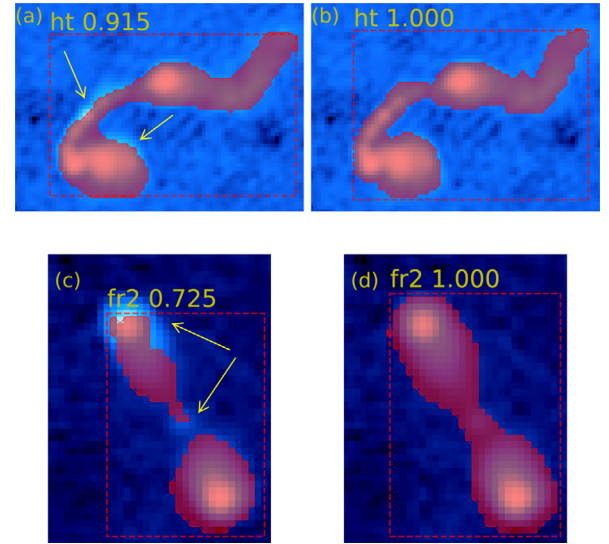


**Fig. 2.** The framework of HeTu-v2 for radio sources segmentation and classification.  
Source: Modified from Ke et al. (2022).

images in the PNG format only. In order to better adapt to astronomical tasks, in the inference phase HeTu-v2 accommodates multiple formats of astronomical data, including the PNG, MIRIAD files, and FITS image format (see Section 4.2 for details). HeTu-v2 is well suited for solving the source finding, classification, and segmentation problems in radio continuum surveys, and the framework is able to form a single software package to perform all tasks.

As shown in Fig. 2, the framework of HeTu-v2 mainly includes a two-stage detector Mask R-CNN, an ambiguity areas detector, and a Transformer block. Firstly, Mask R-CNN is used to generate an initial coarse mask and three levels RoI feature pyramid for a given input radio image. Examples of coarse mask predictions are seen in panels (a) and (c) in Fig. 3, where pixel misclassification occurs at the edge of the bottom left lobe of the HT source and at the edge of the top left lobe of the FRII source respectively. Secondly, the coarse mask and RoI feature pyramid are fed to the ambiguity areas detector to detect the ambiguity areas and output a quadtree (Finkel and Bentley, 1974) responding to the feature level. Finally, the sequence of quadtree nodes is fed into the Transformer block to reclassify the ambiguity pixels and output a fine mask, seen in panels (b) and (d) in Fig. 3, showing that the ambiguity regions of the HT and FRII sources are successfully detected, and the final mask results generated by the Transformer block are corrected and refined compared to panels (a) and (c) respectively. Details of the designed framework's network are described below.

Mask R-CNN uses 101 layers deep Residual Network (ResNet101) (He et al., 2016) combined with Feature Pyramid Network (FPN; Lin et al., 2017) as backbone network (ResNet101-FPN) to extract multi-scale level feature maps from a given input image. These feature maps are represented as  $[P_2, P_3, P_4, P_5, P_6]$  with pixels of size from large to small. They are fed into the RPN (Ren et al., 2017) with the input image to generate region proposals, and the RoI is extracted from the feature maps based on the region proposals using a pooling operation RoI Align (He et al., 2017), where  $[P_2, P_3, P_4, P_5]$  of the feature maps have



**Fig. 3.** Examples of coarse and refined mask prediction. Panels (a) and (c) are the coarse mask predictions by Mask R-CNN, and panels (b) and (d) are the final refined mask predictions by HeTu-v2. Coarse masks appear at the boundaries of the source where the yellow arrows point in panels (a) and (c). The rectangular boxes and masks represent the identified sources. Each source is labeled with a class name (ht: HT, fr2: FRII) and a score between 0 and 1 on the top left of the bounding box.

been utilized. High-level (low-resolution) RoI features contain more contextual and semantic information, and low-level (high-resolution) RoI features resolve more local details. RoI has been used to generate the bounding box, segmentation, or mask and label by the object detection head (i.e. box head) (Lin et al., 2017) and coarse mask head network, respectively. The coarse mask head network consists of a Fully Convolutional Networks (FCN) with four  $3 \times 3$  convolutional layers. Due to the adoption of

an advanced backbone network, Mask R-CNN achieves excellent performance. RoI Align is used in the later steps of the network, which greatly reduces the problem of feature mapping and averaging errors, so the quality of the bounding box produced by Mask R-CNN is higher. However, the segmentation masks are still very coarse due to the downsampling operation used in the feature pyramid and RoI Align steps (Zhang et al., 2021). For example, in Fig. 2, the high-level RoI feature (i.e.  $28 \times 28$ ) has been selected to perform mask prediction, which reduces spatial resolution, so the generated segmented mask is called a coarse mask.

High-quality instance segmentation needs to consider the high-level semantics of the image as well as high-resolution deep features, but the large size of high-resolution features requires much more computational and memory resources (Wang et al., 2021). To refine the segmented mask with low resource usage, the loss information of multi-levels caused by downsampling operation are defined as ambiguity regions which are detected by an ambiguity areas detector. As shown in Fig. 2, the ambiguity areas detector follows the cascaded design to detect information loss nodes at three levels on the RoI feature pyramid:  $28 \times 28$ ,  $56 \times 56$ , and  $128 \times 128$ . Taking the information loss point detected from the lowest level ( $28 \times 28$ ) as the root node, the four sub-quadrant points are recursively expanded from top to bottom, and a multi-level quadtree is constructed. In the ambiguity areas detector, the lowest-level RoI feature and the predicted coarse mask are first fed to a lightweight FCN (i.e. four  $3 \times 3$  convolutions) to detect the root node of the quadtree. Each root node is decomposed into 4 sub-nodes in the previous layer, combined with the corresponding RoI feature, uses a single  $1 \times 1$  convolutional layer to predict finer information loss nodes to keep the detection module lightweight until the mask of the highest layer is generated.

The Transformer block is composed of a node encoder, a sequence encoder, and a pixel decoder. The node encoder first enriches the feature representation of each point, such as the position encoding of the supplementary point and the local detail information of the neighboring points. Then, the node sequence as a Query into the sequence encoder to perform feature fusion and update between points on the input sequence. Finally, the pixel decoder predicts the instance label corresponding to each point. The node encoder encodes each quadtree node using four different pieces of information: fine-grained deep features, coarse mask, relative position encoding, and surrounding point information. Sequence encoder is a Transformer based module, each layer mainly consists of a multi-head self-attention module and a Fully Connected Feed-forward Network (FFN). Pixel decoder is a simple two-layer Multi-Layer Perceptron (MLP; Kirillov et al., 2020) without a multi-head attention module.

### 3.2.2. Auto-generate polygonal annotation

The supervised learning method has been adopted in the training procedure of HeTu-v2, in which machines are trained and validated using well-labeled datasets. In these datasets, the ground truth of each object is annotated. In HeTu-v1, the ground truth annotation information includes bounding boxes and labels, while in HeTu-v2 the polygonal annotation of object segmentation is added. Currently, the widely used annotation tools are LABELME (Russell et al., 2008) and VGG Image Annotator (VIA) tools (Dutta et al., 2016). Both LABELME and VIA support image polygonal annotation through manual mode. The polygonal annotation generated in this mode is greatly affected by the visual. Due to the complex morphology of the radio source, the annotation task requires a lot of human effort and time. We have, therefore, developed a script to automatically generate polygonal annotation using the results from HeTu-v1 model. The input data of this script includes the above FIRST FITS images in Table 1, and

the bounding boxes and the labels name of the detected results of these images. The flow chart of this automated method is shown in Fig. 4, where 2 FRII sources and 2 CS sources are detected by HeTu-v1 and indicated with the yellow and red bounding boxes in the input image respectively. The flow chart is described as follows.

Firstly, image segmentation is performed to the input image using Python module **photutils.segmentation**.<sup>4</sup> Image segmentation is the process of assigning a label to each pixel in an image, and pixels with the same label belong to the same source. Detected sources must have a minimum number of connected pixels, each greater than a specified threshold for the image. The threshold level is usually defined as a multiple of the background noise ( $\sigma$  level) above the background. Here, we use threshold level of  $4\sigma$  and find sources that have at least 5 connected pixels. In the segmented image in Fig. 4, for example, there are 6 radio sources that have been detected and labeled with numbers 1–6, and those segments are filled-in with 6 different colors. However, some extended sources, detected by **photutils.segmentation** and presented as different segments in the segmented image, could belong to one true extended source. For example, there are two FRII sources given by (2,3) and (5,6) labels in Fig. 4; each FRII source in the segmented image is detected as 2 segments or sources in two different colors.

Secondly, the segments within the bounding box of the true source are automatically reassigned with a new and same label number. Then, the coordinates of the boundary points of the newly labeled segments are calculated and used to generate the polygonal annotation. For example, in Fig. 4 the segments within the yellow color bounding box in the input image are labeled to the same number, and the polygonal points of the newly labeled sources are plotted with the same edge colors, see the annotated image. After steps of reassigning segments and generating polygonal annotation, the separate segments of each of the two FRII sources have been combined into one segment, namely 2 and 4 with their polygon points plotted using edge colors of green and orange, respectively.

Thirdly, the polygonal annotation of each image and its related information, including image name, image width and height, image data, label name, and polygonal point coordinates of each radio source, are written to a LABELME format JSON file. These files can be opened with LABELME with visualization to double-check the correctness of the labeled information.

Finally, as HeTu-v2 uses the COCO format Lin et al. (2014) dataset in training and validation experiments, all of LABELME format JSON files for the training and validation sets are converted to COCO format JSON files respectively.

### 3.2.3. Image data augmentation

As the motive power of deep learning, data are crucial for model training. Sufficient training data can not only alleviate the over-fitting problem of the model during training but also further expand the parameter search space and help the model to further optimize towards the global optimal solution. For training of large models, the original amount and style of samples in Table 1 are inadequate. Data augmentation is an effective technique to increase training samples and makes samples as diverse as possible, which has become a necessary part of the successful application of deep learning models on image data (Yang et al., 2022).

To reduce the over-fitting of HeTu-v2 network and to improve the generalization ability and robustness of the network, an augmentation method combining offline and online is adopted in training. For online mode, we directly use the image resize and flip methods in DETETrON2. The input images are first resized to

<sup>4</sup> <https://photutils.readthedocs.io/en/stable/segmentation.html>.



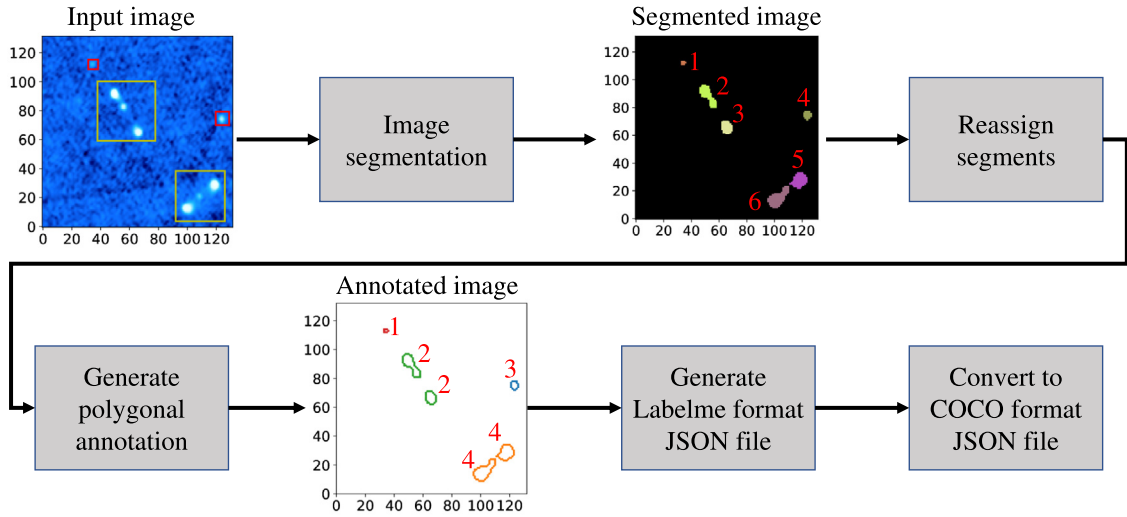


Fig. 4. The flow chart of the auto-generate polygonal annotation.

the shorter edge range from 640 to 800 pixels with step size of 32 pixels, and then each image performed a horizontal flip.

The number of training sets after online augmentation is still unable to reach the requirements of HeTu-v2 model training, so we consider adding image random-rotation methods before online augmentation. When the image pixel size is kept unchanged, the radio source instance in the image after rotation may overflow beyond the image pixel size, see Fig. 5. It should be noted that the blank regions in the rotated images, such as those in panels (b) and (d) of Fig. 5, do not impact the model training. These overflowing images need to be removed manually, so the rotation method is more suitable for offline mode. Our offline augmentation method is implemented using the Python package `imgaug`<sup>5</sup>, in which each image is randomly selected for rotation from  $-180$  to  $180$  degrees, and 50 selections are conducted. The rotation-augmented dataset has 159,303 images and 178,593 sources. More details can be seen in Table 1.

### 3.3. Model training

All experiments including training, evaluation, and inference are conducted in an Intel Xeon Silver 4314 CPU server running at 2.4 GHz (64 cores, 256 GB RAM memory) equipped with three NVIDIA GPUs. Two of the GPUs are NVIDIA RTX A4000, and each has 3072 CUDA cores and 16 GB global memory; the remaining one is NVIDIA A40, and has 5376 CUDA cores and 48 GB global memory. We used all the three GPUs for training.

The goal of model training is to find the model parameters through the sample data to minimize the loss function. The loss function of the HeTu-v2 model consists of multiple losses, and the total loss is calculated by (Ke et al., 2022):

$$L_{\text{total}} = \lambda_1 L_{\text{detection}} + \lambda_2 L_{\text{coarse}} + \lambda_3 L_{\text{refine}} + \lambda_4 L_{\text{ambiguity}}, \quad (2)$$

where  $L_{\text{detect}}$  is the detection loss including the bounding box and label losses from the Mask R-CNN,  $L_{\text{coarse}}$  denotes the loss of the initial coarse mask predicted by the Mask R-CNN,  $L_{\text{refine}}$  is the refinement with mean absolute error (L1) loss of labels for ambiguity nodes from the Transformer block,  $L_{\text{ambiguity}}$  is a Binary Cross Entropy loss of ambiguity regions detection by an ambiguity areas detector, and  $[\lambda_1, \lambda_2, \lambda_3, \lambda_4]$  are hyperparameter weights [1.0, 1.0, 1.0, 0.5].

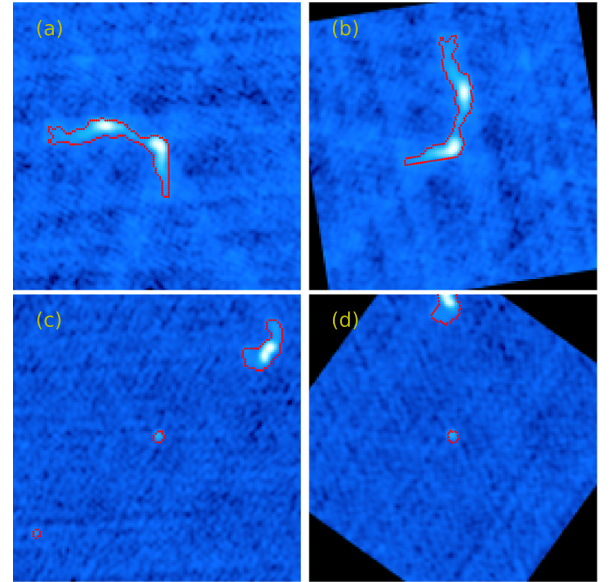
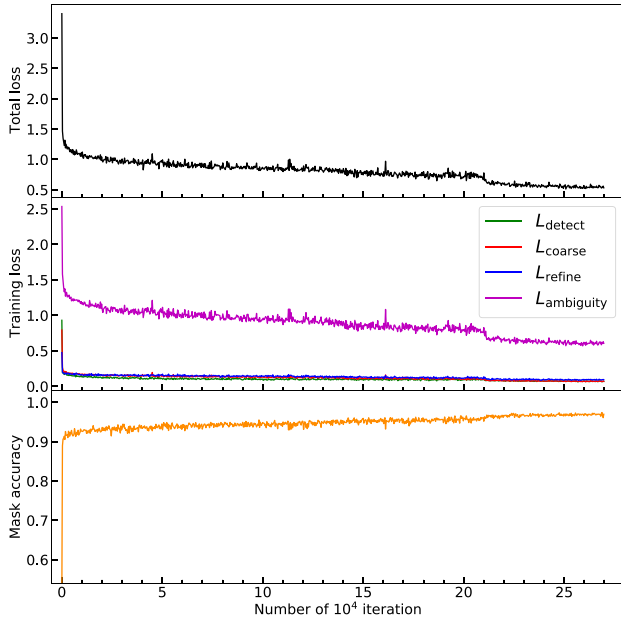


Fig. 5. Examples of random augmentation in the training set, the red polygon points in these examples are the polygon annotation of radio source instances. Panels (a) and (c) are the original images, and panels (b) and (d) are their augmented images, respectively. In panel (b), after the image and instance are rotated, the entire instance is still within the pixel size of the image, and such augmented images can be used for model training. In panel (d), two instances are missing or incomplete after rotation, and such augmented images must be removed.

HeTu-v2 model is trained using the  $3\times$  training schedule (He et al., 2019). With 9 images per batch on 159,303 radio images, the learning rate decreases by a factor of 0.1 at 210,000 and 250,000 iterations, and finally terminates at 270,000 iterations, where an iteration refers to the process of training on a batch of images. To optimize training and speed up model convergence, Stochastic Gradient Descent (SGD) with a momentum of 0.9 and 1000 iteration linear warm-up with an initial learning rate of 0.01 and weight decay of 0.0001 is used. In addition, distributed data-parallel strategy has been employed to reduce training time. The model is replicated on each GPU device, and the total batch data are equally divided into 3 different GPUs for calculation, each GPU worker loads its own data from disk. During training, each GPU process communicates with other processes through

<sup>5</sup> <https://github.com/aleju/imgaug>.



**Fig. 6.** Training loss and mask accuracy curves with the number of iterations. The top panel shows the total loss, the middle panel shows the loss of the tasks, and the bottom panel shows the mask accuracy.

the “All-Reduce” method and exchanges their gradients to obtain the average gradient of all processes. Each GPU process uses the averaged gradient to update its own parameters, making each GPU parameter synchronize.

During training, the training and validation metrics (e.g. time, multi-task losses, mask, and class accuracy) have been recorded into a JSON file every 20 iterations, and the trained model file is saved per 10,000 iterations. The training has taken  $\sim 67.5$  h to finish. The trained HeTu-v2 model can be performed on radio images with size of  $132 \times 132$  pixels in less than  $\sim 200$  ms per sample using one GPU.

Fig. 6 shows the training loss and mask accuracy curves. As the number of iterations increases, the loss of the model on the training dataset continues to decrease. The first 10,000 iterations show a significant drop in the total loss due to a large difference between the parameters of the model and the optimal parameters. By continuously optimizing the model parameters on the validation set, the model will eventually approach the optimal value, resulting in improved performance in locating and classifying radio sources. The total loss steadily decreases from 10,000 to 209,999 iterations as the model undergoes local optimization. At the 210,000th iteration, there appears to be a sudden drop in the total loss, indicating that the model had deviated from its previous locally optimized state as the learning rate decreased tenfold from 0.01 to 0.001 at this iteration. Finally, the total loss function begins to converge at the 229,999th iteration, and the mask accuracy is the highest for the first time at this iteration. After this iteration, the total loss flattens out and stabilizes at around  $L_{\text{total}} = 0.53$ , with  $L_{\text{detection}} = 0.07$ ,  $L_{\text{coarse}} = 0.07$ ,  $L_{\text{refine}} = 0.09$ , and  $L_{\text{ambiguity}} = 0.6$ . Therefore, the output model of the 229,999th iteration is used as the final trained model of HeTu-v2.

## 4. Results and discussion

### 4.1. Model performance and evaluation

To quantitatively evaluate the HeTu-v2 network performance, we have applied the trained HeTu-v2 model to the validation

dataset (see Table 1) and calculated the evaluation metrics. Since the validation dataset is in COCO format, the Microsoft COCO Evaluation Metrics (Lin et al., 2014) are used to quantitatively evaluate the performance of HeTu-v2 network, including Average Precisions (APs) and Average Recalls (ARs), but ARs are not commonly used. Compared with the Pascal VOC metrics (Everingham et al., 2010), Microsoft COCO Evaluation Metrics are more comprehensive.

To understand AP, we first introduce Precision ( $P$ ) and Recall ( $R$ ), which represent the reliability (accuracy) and completeness of the detection results generated by the HeTu-v2 model on the validation set.  $P$  is the True Positive ( $TP$ ) prediction rate of all positive predictions for a class:

$$P = \frac{TP}{TP + FP} = \frac{TP}{\# \text{predictions}}, \quad (3)$$

where  $TP$  denotes the number of positive class sources that are correctly predicted to be in the positive class and whose Intersection over Union ( $IoU$ ) value exceeds the specified threshold ( $IoU_{\text{threshold}} \in 0.0 \dots 1.0$ ), and  $FP$  represents the number of negative class sources that are incorrectly predicted to be in the positive class or whose  $IoU$  value is below the  $IoU_{\text{threshold}}$ . The  $IoU$  is given by the ratio of the area of intersection ( $\cap$ ) between the predicted mask ( $M_p$ ) and ground truth mask ( $M_{gt}$ ) to the area of their union ( $\cup$ ):

$$IoU = \frac{\text{area}(M_p \cap M_{gt})}{\text{area}(M_p \cup M_{gt})}. \quad (4)$$

$R$  is the  $TP$  prediction rate of all ground truth sources in a class:

$$R = \frac{TP}{TP + FN} = \frac{TP}{\# \text{ground truths}}, \quad (5)$$

where False Negative ( $FN$ ) is the number of positive class sources but are incorrectly predicted to be in the negative class.

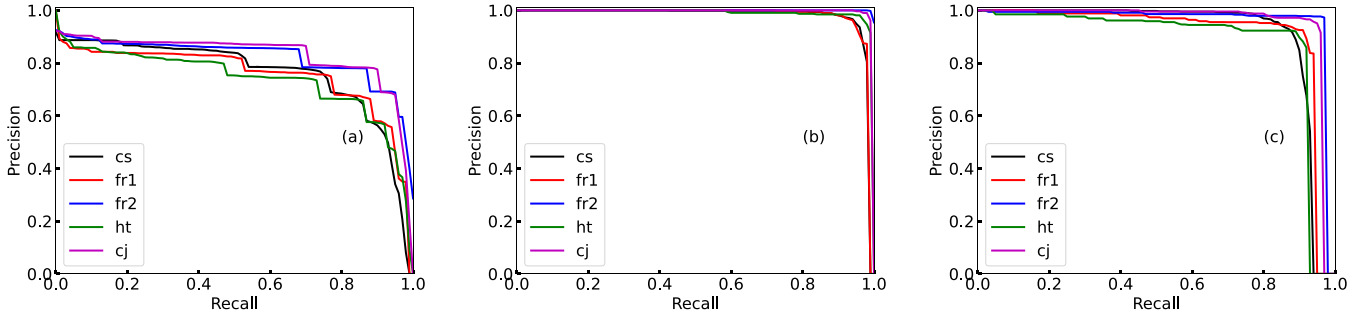
The area under the  $P - R$  curve of a given class is called Average Precision ( $AP$ ). A higher  $AP$  indicates better detection performance for the network. In practice,  $AP$  represents the precision averaged across all recall values that are equally spaced between 0.0 and 1.0. These equally spaced recall values are typically achieved by adjusting the  $IoU$  threshold level to obtain the desired recall. Here, the  $AP$  is calculated as the mean of precision values at a set of 101 equally spaced recall levels  $R \in [0, 0.01, \dots, 1.0]$ :

$$AP = \frac{1}{101} \sum_{R \in [0, 0.01, \dots, 1.0]} P(R), \quad (6)$$

where  $P(R)$  is the maximum precision in interval  $\Delta R$ .

In the Microsoft COCO Evaluation Metrics, the performance of the network in distinguishing objects is refined by varying the  $IoU_{\text{threshold}}$  from 0.5 to 0.95 with a step size of 0.05, resulting in 10 sets of  $P - R$  curves. The 10  $AP$ s calculated from the 10 sets of  $P - R$  curves are averaged and expressed as  $AP_{@50:5:95}$ , which is the primary challenge metric. When the  $IoU_{\text{threshold}}$  is fixed, such as  $IoU_{\text{threshold}} = 0.5$  and  $IoU_{\text{threshold}} = 0.75$ , the metric is written as  $AP_{50}$  (i.e. Pascal VOC metric) and  $AP_{75}$ , respectively. In addition, the  $AP$  of sources with different pixel areas is also calculated to evaluate the network's detection performance on objects of varying sizes. Regarding radio source detection, sources with an area less than  $16^2$  pixels are represented as  $AP_{\text{small}}$ , those with an area larger than  $16^2$  pixels but less than  $32^2$  pixels are represented as  $AP_{\text{medium}}$ , and those with an area larger than  $32^2$  pixels are represented as  $AP_{\text{large}}$ . These metrics are computed using the same  $IoU$  threshold range as that of  $AP_{@50:5:95}$ . The definitions of  $AP_{\text{small}}$ ,  $AP_{\text{medium}}$ , and  $AP_{\text{large}}$  used for radio source detection differ from those defined in the original COCO Evaluation Metrics





**Fig. 7.** Precision–recall ( $P - R$ ) curves of HeTu-v2 calculated at different  $IoU$  thresholds averaged over the validation data samples for five classes. Panel (a) is calculated at  $IoU$  threshold ranging from 0.5 to 0.95. Panel (b) is calculated at an  $IoU$  threshold of 0.5, and Panel (c) is calculated at an  $IoU$  threshold of 0.75.

because the pixel area of radio sources is typically much smaller. In object detection, the mean  $AP$  ( $mAP$ ) is typically used as the final evaluation metric. This metric is defined as the average of the  $AP$  metric across all classes in a model, and is given by:

$$mAP = \frac{\sum_{i=1}^{N_c} AP_i}{N_c}, \quad (7)$$

where  $N_c$  is the number of classes.

In Table 2, various  $AP$  metrics are to evaluate the performance of HeTu-v2, HeTu-v1, and Mask R-CNN networks. These metrics show the model's accuracy across different  $IoU$  thresholds and object scales. The three methods use the same training and validation datasets. HeTu-v2 obtains the highest mean  $AP_{@50:5:95}$  of 77.8% and outperforms HeTu-v1 and Mask R-CNN by 15.6 points and 11.3 points, respectively. Other  $mAP$ s of HeTu-v2 also have varying degrees of improvement compared to HeTu-v1 and Mask R-CNN. The significant advancements reveal the effectiveness of the framework upgrade from HeTu-v1 to HeTu-v2. This also indicates that HeTu-v2 network can effectively solve the loss of spatial information problem in Mask R-CNN, especially for large-scale extended sources that the mean  $AP_{large}$  increased by 23.8 points. Furthermore, the  $AP_{50}$ s of all classes are no less than 95%, particularly the  $AP_{50}$  of FRII is 99.9% i.e. close to 100%. HeTu-v2 model has the highest accuracy on a medium scale of sources with mean  $AP_{medium}$  of 80.5% and improved by 10.6 points and 11.8 points than HeTu-v1 and Mask R-CNN, respectively. This suggests that HeTu-v2 is the most appropriate method for modern radio surveys as they can resolve more complex sources.

The results shown in Fig. 7 summarize the performance of HeTu-v2 network against ground truth images in the validation dataset. In the validation data, each sample is assigned a prediction score by the HeTu-v2 model, which represents the probability that the sample is positive. The samples for each class are then sorted based on their prediction scores. Samples with prediction scores larger than the threshold are classified as  $TP$ , while those smaller than the threshold are  $FP$ . As the number of samples for evaluation increases, the HeTu-v2 model identifies more  $TP$  or  $FP$  detections according to the division rule, resulting in an increase in  $R$  but a decrease in  $P$ . This pattern may oscillate depending on the characteristics of the data being evaluated. Taking panel (a) of Fig. 7 as an example, all  $P - R$  curves show a decrease in  $P$  as  $R$  increases, suggesting the relationship of a trade-off between  $P$  and  $R$ . In general, a higher  $P$  at a certain  $R$  indicates better model detection ability. This is reflected in the larger area under the  $P - R$  curve, i.e., the higher the  $AP$ , the better the detection ability of the model. Panels (a)–(c) in Fig. 7 are used to calculate the  $AP_{@50:5:95}$ ,  $AP_{50}$ , and  $AP_{75}$  through Eq. (6), indicating that the area under  $P - R$  curves consistent with the  $AP$ s in Table 2. The  $P - R$  curves of the object scales are similar to Fig. 7 and therefore are not shown in this paper.

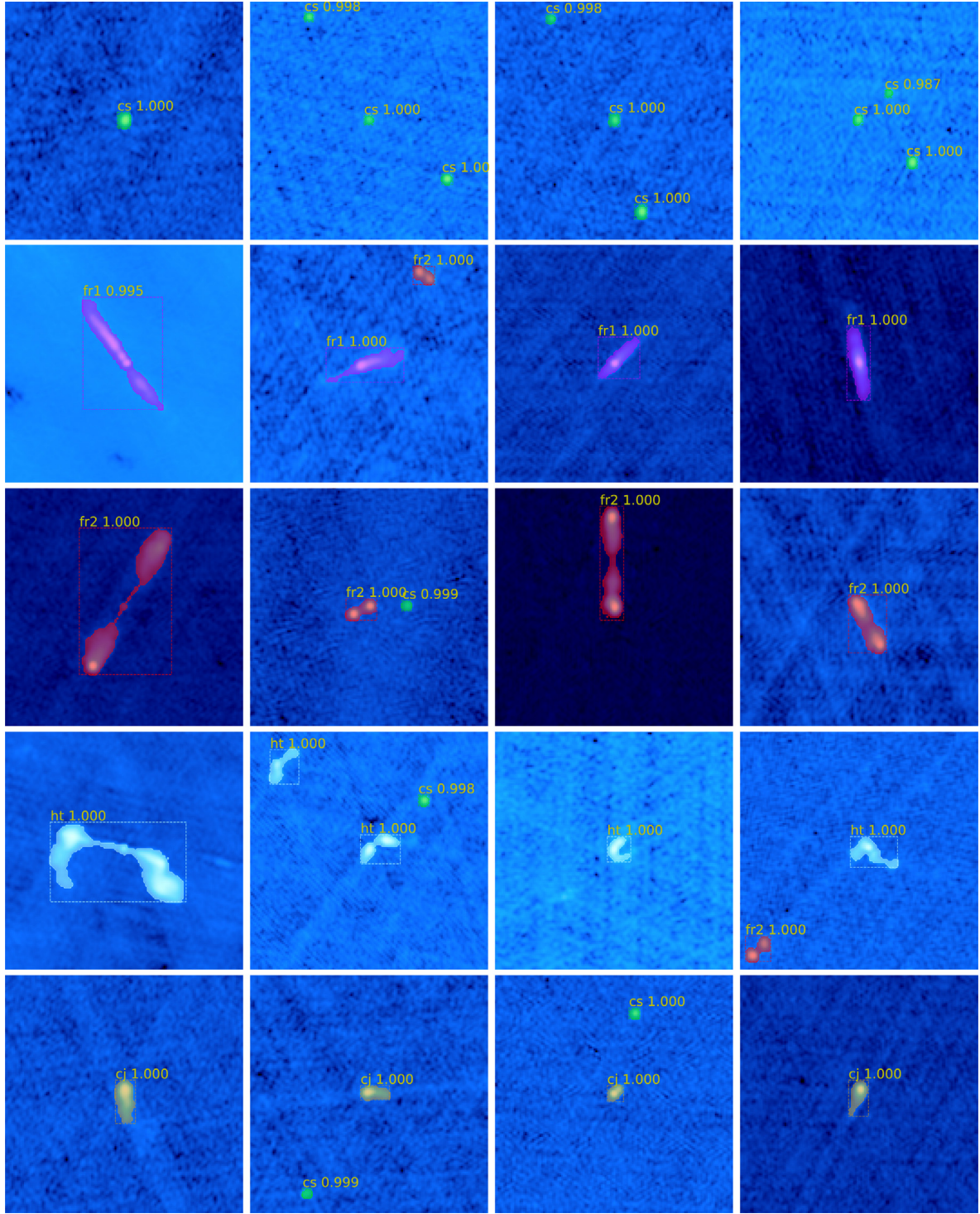
**Table 2**

Summary of  $AP$  results on validation dataset using different methods.  $AP_{@50:5:95}$  represents the  $AP$  at  $IoU$  threshold values ranging from 0.5 to 0.95. Meanwhile,  $AP_{50}$  and  $AP_{75}$  refer to the  $AP$  at  $IoU$  thresholds of 0.5 and 0.75, respectively. Additionally,  $AP_{small}$ ,  $AP_{medium}$ , and  $AP_{large}$  denote the  $AP$  values for sources with small (area <  $16^2$  pixels), medium ( $16^2$  pixels < area <  $32^2$  pixels), and large (area >  $32^2$  pixels) sizes, respectively, at  $IoU$  threshold values ranging from 0.5 to 0.95.  $mAP$  is the mean  $AP$  across all classes.

Methods	Class	$AP_{@50:5:95}$	$AP_{50}$	$AP_{75}$	$AP_{small}$	$AP_{medium}$	$AP_{large}$
HeTu-v2	CS	75.8%	97.3%	91.1%	75.8%	–	–
	FRI	75.4%	97.4%	91.3%	75.6%	79.0%	79.2%
	FRII	82.0%	99.9%	95.9%	81.9%	84.1%	78.2%
	HT	73.4%	98.4%	88.1%	68.2%	77.5%	72.7%
	CJ	82.5%	98.9%	95.3%	82.7%	81.4%	–
	<b>mAP</b>	<b>77.8%</b>	<b>98.4%</b>	<b>92.3%</b>	<b>76.8%</b>	<b>80.5%</b>	<b>76.7%</b>
HeTu-v1	CS	42.2%	89.1%	33.3%	42.2%	–	–
	FRI	58.7%	94.7%	72.4%	60.4%	59.8%	45.4%
	FRII	79.4%	99.9%	96.6%	73.4%	81.9%	85.4%
	HT	68.6%	99.2%	85.4%	62.3%	70.2%	70.2%
	CJ	62.1%	98.9%	73.7%	62.0%	67.9%	–
	<b>mAP</b>	<b>62.2%</b>	<b>96.4%</b>	<b>72.3%</b>	<b>60.1%</b>	<b>69.9%</b>	<b>67.0%</b>
Mask R-CNN	CS	57.6%	93.7%	65.1%	57.6%	–	–
	FRI	67.1%	98.2%	82.0%	69.8%	65.0%	45.8%
	FRII	68.2%	99.7%	86.0%	74.4%	68.3%	54.9%
	HT	65.6%	98.5%	82.8%	67.2%	68.0%	58.0%
	CJ	73.1%	98.8%	93.1%	73.3%	73.5%	–
	<b>mAP</b>	<b>66.3%</b>	<b>97.8%</b>	<b>81.8%</b>	<b>68.5%</b>	<b>68.7%</b>	<b>52.9%</b>

Moreover, the final precision and recall of HeTu-v2 on all validation data were calculated at thresholds of 0.5 and 0.75, and the results are shown in Table 3. Table 3 shows that the final precision for the five classes is 93.4% and 91.2% at thresholds of 0.5 and 0.7, respectively. Additionally, the final recall for the five classes is 95.1% and 87.1% at thresholds of 0.5 and 0.7, respectively. These findings highlight the high accuracy and completeness of the HeTu-v2 model in locating and classifying radio sources within validation images. As the threshold increases, the number of true or false positives decreases. This leads to a decrease in the overall recall and a fluctuation in the precision. The change in final recall and precision for FRII sources is minimal, with only a 0.2% decrease in recall and a 1.3% fluctuation in precision. This indicates that the HeTu-v2 model has a high level of confidence in detecting FRII sources.

Fig. 8 shows examples of segmentation and classification on the validation dataset. The localization and size of the detected sources are labeled with five colored rectangular boxes and masks: green for CS, magenta for FRI, red for FRII, light-cyan for HT, and orange for CJ. Each source is also labeled with the associated class or label name with a probability score between 0 and 1 on the top-left of a rectangular box. In Fig. 8, all detected radio sources are segmented and classified with a higher score and refined masks, respectively. This suggests the high accuracy of the predicted masks by HeTu-v2.



**Fig. 8.** Examples of segmentation and classification on the validation dataset. The different colored rectangular boxes and masks represent the identified sources: green for CS (cs), magenta for FRI (fr1), red for FRII (fr2), light-cyan for HT (ht), and orange for CJ (cj). Each source is labeled with a class name and a score between 0 and 1 on the top left of the bounding box.

## 4.2. Inference experiments

### 4.2.1. Building catalog

To investigate the inference ability of trained HeTu-v2 model and the completeness, reliability, and accuracy of the catalog produced by this model, we used 946,366 processed images from the FIRST survey described in Section 3.1.2. Removing the repeated positions of components, the images contain a total of 946,366

components in the FIRST-14dec17 catalog. To speed up the prediction procedure, the images have been split into three parts (*i.e.* 315,455, 315,455, and 315,456), and then run them simultaneously on three different GPU devices. The inferring experiments took about 17.5 h to finish.

A batch prediction is performed on all images, and at the same time each predicted source is calculated for their properties information to form a catalog. Thus, the HeTu-v2 is said to function as a source detector. The workflow of the HeTu-v2 based



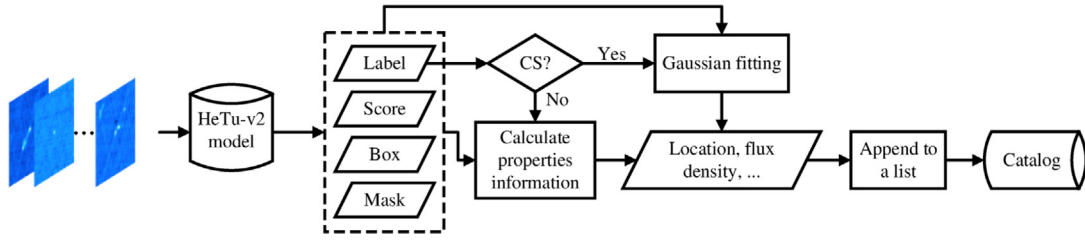


Fig. 9. Workflow of the HeTu-v2 based source detector.

Table 3

The final precision and recall of HeTu-v2 on all validation data using thresholds of 0.5 and 0.75 respectively.

IoU threshold	Class	Precision	Recall
0.5	CS	92.5%	92.9%
	FRI	93.0%	91.2%
	FRII	96.7%	99.8%
	HT	88.1%	95.9%
	CJ	96.1%	96.4%
	ALL	93.4%	95.1%
0.75	CS	87.2%	80.0%
	FRI	93.8%	82.9%
	FRII	96.9%	98.5%
	HT	87.3%	91.9%
	CJ	95.4%	89.1%
	ALL	91.2%	87.1%

source detector is shown in Fig. 9. At first, the FIRST images are sequentially input to the HeTu-v2 model and then output the prediction results of the detected sources, including predicted label, predicted score, predicted bounding box, and predicted mask. Next, depending on the prediction information, each of predicted source properties such as location and flux density has been calculated. For each of CS sources, Gaussian fitting is performed using the IMFIT task of MIRIAD software (Sault et al., 1995), which provides the fitting information: location (*i.e.* right ascension and declination), local Root Mean Square (RMS) noise, peak flux density, integrated flux density, beam (major and minor axes length, and the position angle) in addition to the deconvolved beam information. For each of extended sources, the center position of the predicted box, local RMS noise, peak flux density at the maximum pixel value in the predicted box, and the total flux density of the detected source have been calculated. Here, the total flux density is given by Eq. (8)<sup>6</sup>:

$$S_{\text{total}} = \left( \sum I_{\text{in}}[\text{mask}] \right) / \left( (2\pi \cdot b_{\text{maj}} \cdot b_{\text{min}}) / G_{\text{factor}}^2 \right), \quad (8)$$

where *mask* is the predicted masks of the detected source in a type of Boolean. The function  $\sum I_{\text{in}}[\text{mask}]$  refers to the sum of the pixel values at the position where the *mask* is *True* in the input image. The *b<sub>maj</sub>* and *b<sub>min</sub>* represent the beam semi-major and semi-minor axis length of the input image in pixels scale. The *G<sub>factor</sub>* is a factor of beam area and is defined as  $G_{\text{factor}} = 2\sqrt{2(\ln(2))}$ . At last, the information of each detected source is appended to a list and thereby HeTu-v2 generates the initial catalog in a CSV file. The details of more catalog columns can be seen in Table 4.

Since the image samples have overlapping sky regions and the sources detected on one image are detected again on other images, the generated initial catalog contains duplicate sources which need to be removed. Consequently, a post-processing program for de-duplication has been developed. Due to the extended source large size and complex structure, the small size (diameter

$\leq 10$  pixels) and the simple structure (Gaussian or point) of the CS sources, different methods are adopted for these two groups of sources.

To begin with, the initial catalog is split into five sub-catalogs based on the label name. For CS sources, first, the initial CS sub-catalog performs intra-table cross-match on the location columns (centre\_ra and centre\_dec) using STILTS package (Taylor, 2006) within 5 arcsec (*i.e.* the full width at half-maximum (FWHM) beam size in the FIRST survey) of the max error to generate the matched CSV file. Then, duplicate rows are removed from the matching results and keep the row of the first occurrence. Now the unmatched rows are appended to the remaining results of the matched CSV file to create a new CSV file as the final results. For each morphological class sub-catalog of extended sources, the bounding box of the *i*th ( $1 \leq i \leq N_{\text{source}}$ ) row defining a location region, check whether the location from (*i* + 1)th to *N<sub>source</sub>*th is within the region, and the sources within this region are regarded as duplicated with the *i*th row and recorded in a TXT file. Here, *N<sub>source</sub>* is the total number of rows or sources for the sub-catalog of extended sources. Following this method we can get all duplicate sources by looping through all rows. Finally, the duplicate sources in the last generated TXT file are removed from the sub-catalog. In the end, all processed sub-catalogs are combined to obtain the final catalog.

#### 4.2.2. Completeness of the catalog

The final catalog (called FIRST-HeTu) detected by HeTu-v2 is shown in Table 5. In Table 5, HeTu-v2 has detected a total of 835,435 sources, including 700,609 CS sources, 12,501 FRI sources, 47,582 FRII sources, 22,187 HT sources, and 52,557 CJ sources. We have compared FIRST-HeTu with FIRST-14dec17 to investigate its completeness.

The HAPPY program that was used to generate the FIRST-14dec17 catalog is a component-based source finder, which identifies a CS source with a single component while identifying an extended source with multiple discrete components or sources. Diffuse emission between close components cannot be recognized as Gaussian components, so HAPPY cannot directly determine whether there is a connection between adjacent components, resulting in its lack of ability to automatically classify extended sources. Classification needs to be done by manual visual inspection in post-processing. Two methods have, therefore, been used to compare catalogs. For CS sources, we have used STILTS to perform cross-match on the fitted location columns between FIRST-HeTu (ra and dec) and FIRST-14dec17 (RA and DEC) within a searching radius of 5 arcsec and obtained 689,877 CS sources detected by both. For extended sources, we have developed a script to automatically associate FIRST-14dec17 components with FIRST-HeTu sources, which takes the FIRST-14dec17 components positioned within the predicted masks of the extended source as its associated components. There are 21,835 (FRI), 97,940 (FRII), 49,044 (HT), and 74,588 (CJ) FIRST-14dec17 components to be associated in different classes of extended sources. Thus, a total of 933,284 FIRST-14dec17 components are

<sup>6</sup> <https://github.com/mhardcastle/radioflux>.



**Table 4**

The columns information of output catalog by HeTu-v2 based source detector.

Column name	Description
source_name	Predicted source name in JHHMSSDDMMSS format.
image_filename	Input image file name in PNG format.
label	Predicted label name (cs: CS, fr1: FRI, fr2: FRII, ht: HT, cj: CJ).
score	Predicted score in the range of 0.0 to 1.0.
box	Predicted bounding box points in the format xmin-ymin-xmax-ymax.
mask	Predicted mask encode in run-length encoding (RLE) format.
local_rms	Local noise level of input image in Jy-beam <sup>-1</sup> .
peak_flux	Peak flux density of the fitted Gaussian model component in Jy-beam <sup>-1</sup> for CS sources, or flux density at the peak within the predicted box in Jy-beam <sup>-1</sup> for extended sources.
err_peak_flux	Fitting error of the peak flux density in Jy-beam <sup>-1</sup> , only for CS sources.
int_flux	Integrated flux density of fitted Gaussian model component in Jy for CS sources, or total flux density of extended sources.
err_int_flux	Fitting error of the integrated flux density in Jy, only for CS sources.
ra	Right Ascension of the emission peak of the fitted Gaussian model component in degrees for CS sources, or Right Ascension of the emission peak within the predicted box in degrees for extended sources.
dec	Declination of the emission peak of the fitted Gaussian model component in degrees for CS sources, or Declination of the emission peak within the predicted box in degrees for extended sources.
centre_ra	Right Ascension at the center position of the predicted box in degrees.
centre_dec	Declination at the center position of the predicted box in degrees.
major	Semi-major axis length of the fitted Gaussian model component in arcsec, only for CS sources.
err_major	Fitting error of the fitted semi-major axis length in arcsec, only for CS sources.
minor	Semi-minor axis length of the fitted Gaussian model component in arcsec, only for CS sources.
err_minor	Fitting error of the fitted semi-minor axis length in arcsec, only for CS sources.
pa	Position angle of the fitted Gaussian model component in degrees, only for CS sources.
err_pa	Fitting error of the Gaussian model component's position angle in degrees, only for CS sources.
deconv_major	Deconvolved major axis in arcsec, only for CS sources.
deconv_minor	Deconvolved minor axis in arcsec, only for CS sources.
deconv_pa	Deconvolved position angle in degrees, only for CS sources.

**Table 5**

Comparison the catalog results between FIRST-HeTu and FIRST-14dec17.

Class	FIRST-HeTu	Components of the FIRST-14dec17 cross-matched or associated with the FIRST-HeTu sources
CS	700,609	689,877
FRI	12,501	21,835
FRII	47,582	97,940
HT	22,186	49,044
CJ	52,557	74,588
Total	835,435	933,284

associated with the sources detected by HeTu-v2. The cross-matching rates achieve 98.6%, showing excellent completeness of HeTu-v2 detecting FIRST radio sources. The remaining 1.4% of components that were not detected by HeTu-v2 are limited by the current model's performance.

Since HAPPY detection is related to the threshold value and strict acceptance criteria while HeTu-v2 detection is based on the source morphology and unlimited on the flux density threshold and acceptance criteria (White et al., 1997), HeTu-v2 can detect more CS sources than HAPPY. HeTu-v2 has detected 10,732 CS sources that were not detected by HAPPY, with 3311 of them having optical counterparts within a search radius of 5 arcsec from Sloan Digital Sky Survey Data Release 16 (SDSS DR16; Ahumada et al., 2020). Among these 3311 sources, approximately 75% have a signal-to-noise ratio greater than or equal to 5, which is the threshold value used in HAPPY.

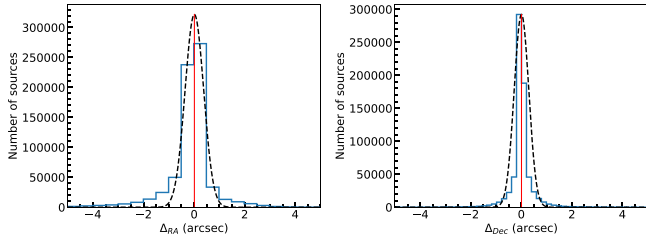
The number of sources in the latest released catalogs of FRI, FRII, and HT sources are 122 (Capetti et al., 2017), 219 (Capetti et al., 2017), and 717 (Sasmal et al., 2022), respectively. By the cross-matching script described in Section 4.2.1, we have found that the extended sources in the FIRST-HeTu catalog completely cover the latest published results, and we have identified more new sources. Therefore, FIRST-HeTu is currently the most complete catalog for the morphological classification and identification of extended sources in the FIRST survey.

On comparing the results obtained from HeTu-v1 in Section 3.1.2 with those from HeTu-v2, we observe that the total number of CS sources detected by HeTu-v1 (i.e., 852,271) is higher than that detected by HeTu-v2 (i.e., 700,609). However, most of the CS sources detected by HeTu-v1 but not by HeTu-v2 are found to be false positives by visual inspection. This suggests that HeTu-v2 detects fewer false positives in CS sources. Furthermore, the total number of detected extended sources by HeTu-v2 is approximately 1.3 times that of HeTu-v1, and all of these extended sources have associated components in the FIRST-14dec17 catalog, highlighting the significantly improved locating accuracy and detection performance of HeTu-v2 on extended sources. These findings demonstrate that HeTu-v2 has excellent accuracy on the FIRST survey and has established a high completeness catalog.

#### 4.2.3. Reliability and accuracy of the catalog

For each CS source that is cross-matched with a FIRST-14dec17 component, we have measured the offset in position to check the precision of the positions of the CS sources detected by HeTu-v2. The result is presented in Fig. 10, showing a mean offset of 0.0097 arcsec and 0.0072 arcsec on the Right Ascension (RA) axis and Declination (Dec) axis, respectively. The distributions in the left and right panels are concentrated around 0 arcsec. This indicates that the positions of CS sources fitted by HeTu-v2 are basically consistent with those of HAPPY.

The flux density accuracy of the detected CS sources has been estimated by comparing the peak and integrated flux densities of cross-matched CS sources in the FIRST-HeTu and FIRST-14dec17 catalogs. The results are shown in Fig. 11. Defining  $R = S_{\text{FIRST-HeTu}} / S_{\text{FIRST-14dec17}}$  as the ratio between the two peak flux densities, the median value of  $R$  is 0.93 with a median absolute deviation of 0.08. The median and absolute deviation values of the ratio between two integrated flux ( $S_{\text{FIRST-HeTu}}^{\text{INT}} / S_{\text{FIRST-14dec17}}^{\text{INT}}$ ) densities are equivalent to those of  $R$ . In addition, in Fig. 11, a linear fit to  $S_{\text{FIRST-HeTu}}$  vs.  $S_{\text{FIRST-14dec17}}$  gives the slope of 1.0 and intercept of 0.00003, the linear regression analysis of  $S_{\text{FIRST-HeTu}}^{\text{INT}}$



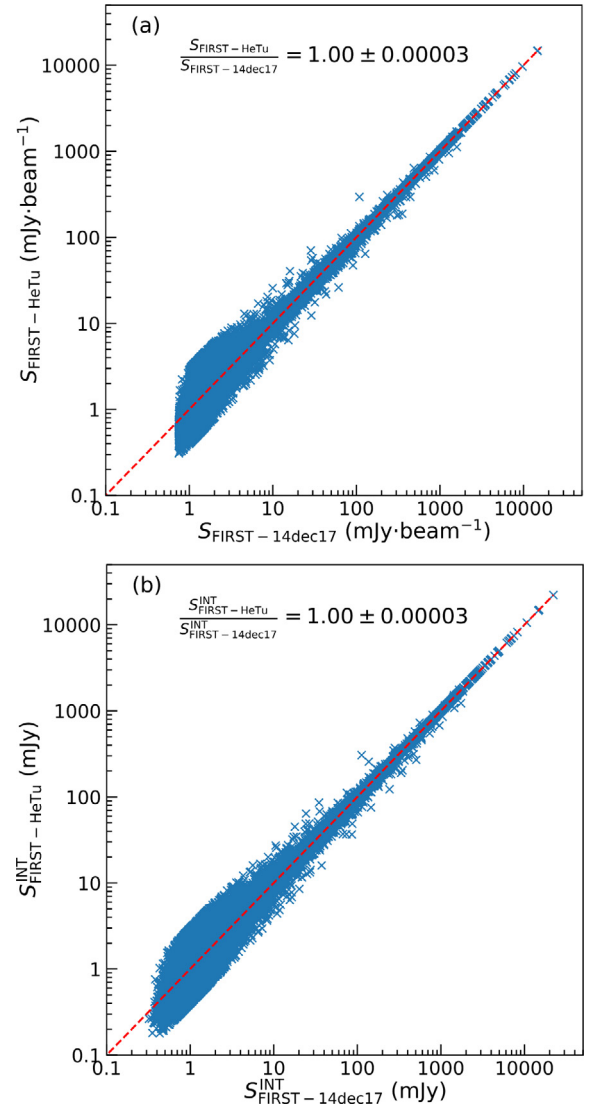
**Fig. 10.** Histograms of the difference in position coordinates (i.e. RA and Dec) of CS sources detected by HeTu-v2 compared to the matching FIRST-14dec17 sources, showing a mean offset of 0.0097 arcsec and 0.0073 arcsec on the RA axis and Dec axis, respectively. The black dotted curves show Gaussian fits to the histograms, and the red solid vertical lines indicate the mean offsets.

vs.  $S_{\text{FIRST-14dec17}}^{\text{INT}}$  yields the same slope and intercept as that of  $S_{\text{FIRST-HeTu}}$  vs.  $S_{\text{FIRST-14dec17}}$ . It indicates that the flux densities fitted by HeTu-v2 and HAPPY are in basic agreement. The flux density shows larger fluctuations below 10 mJy, indicating that the flux density fitted by HeTu-v2 is somewhat different from that by HAPPY in weak sources. Each fitted component by HAPPY lies in the contiguous set of pixels exceeding the threshold (i.e.  $5\sigma$ ), while the Gaussian fit by HeTu-v2 is performed only in the pixels region within the predicted box. For weak point-like sources, the regions of the sources used for fitting by HeTu-v2 are larger than those used in the HAPPY fitting. Furthermore, this also leads to differences in the shape and size of the Gaussian models that they fit. Fig. 12 compares the semi-major and semi-minor axis lengths of the fitted Gaussian models between the two catalogs for all cross-matched CS sources, showing that the fitted Gaussian models by HeTu-v2 are slightly larger than those by HAPPY.

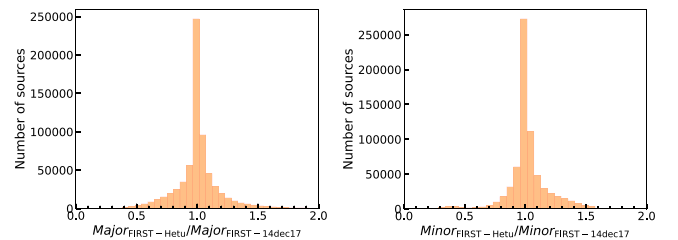
By defining the source that cross-matched with FIRST-14dec17 as true positive, the detection precision for the CS source is 98.5%. In addition, the analysis revealed that the ratio of integrated flux to peak flux for the cross-matched CS sources in FIRST-14dec17 is consistently less than 1.2, which suggests that these sources are indeed CS sources. For extended sources, we used the prediction score to estimate the detection accuracy; the higher the score, the higher the accuracy. Fig. 13 shows the prediction score histograms of the detected FRI, FRII, HT, and CJ sources. The histograms of four classes show protruded shoulders from 0.9 to 1.0 indicating that the majority of detected extended sources have high accuracy. The mean prediction scores of FRI, FRII, HT, and CJ sources have achieved 0.84, 0.93, 0.86, and 0.88, respectively. In summary, the FIRST-HeTu is a reliable and accurate FIRST morphological catalog.

#### 4.3. Application prospects

Building a catalog from radio images is the primary purpose and function of HeTu-v2. The generated catalog by HeTu-v2 has three new useful features compared to the traditional catalog: the morphological class for each source, the total flux density for the extended source, and the central location for each source. As a novel tool, HeTu-v2 like traditional source-finding software, the created catalog results can be used to build sky models which can be used in the calibration steps of the standard data processing pipeline. For CS sources, the sky models are still created based on the Gaussian fitted information by HeTu-v2. For extended sources, however, the prediction result by HeTu-v2 allows building the sky model in pixel levels like the pixel segmentation-based source finder ProFound (Hale et al., 2019). Fig. 14 shows how HeTu-v2 is able to trace the shape of the source and so model its radio emission. In the case shown here, PyBDSF does not adequately model the emission seen in the image. The residual



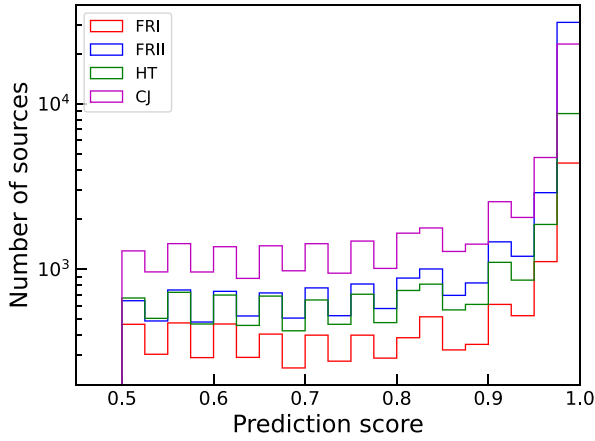
**Fig. 11.** Comparison of fitted source peak (a) and integrated (b) flux densities between FIRST-HeTu and FIRST-14dec17 for all cross-matched CS sources. The red solid line is the equal flux line.



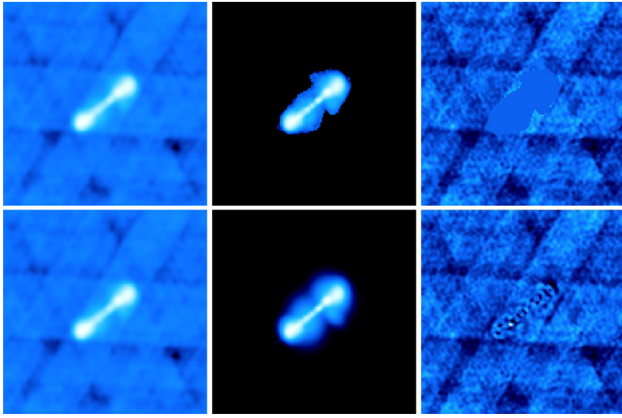
**Fig. 12.** Comparison of semi-major and semi-minor axis length of the fitted Gaussian model component between FIRST-HeTu and FIRST-14dec17 for all cross-matched CS sources.

images in Fig. 14 also show how HeTu-v2 is tracing the shape well in comparison to the PyBDSF whose residual image appears to over-fit in areas, which can leave negative and excess positive residuals around the source.

HeTu-v2 is also able to perform automatic association of radio components like the recent works by Mostert et al. (2022) through a small amount of extended development. Based on the results of traditional source-finding software, this method can



**Fig. 13.** Histograms of the prediction score of HeTu-v2 detected extended sources.

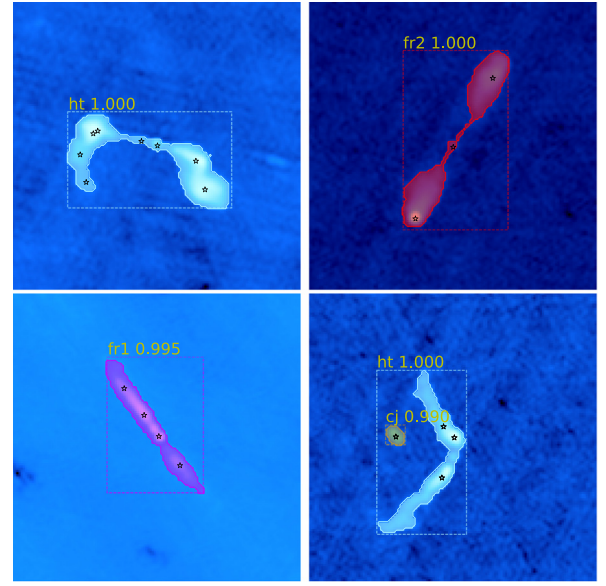


**Fig. 14.** Comparison of the model (middle column) and residual (right column) images from HeTu-v2 (top row) and PyBDSF (bottom row) of source J085556.090 + 491113.15 (left column) from the FIRST survey. Here, PyBDSF having version v1.10.1 generates the model and residual images using the task 'process\_image' with default parameters.

automatically and correctly combine the separated radio components that belong to a source, which can further help to find the corresponding optical host galaxy (Banfield et al., 2015) and infer the physical parameters such as the size and luminosity of the radio source. Traditionally, radio components association is used manually by visual inspection which is a time-consuming process even with extensive manpower for the huge amount of sources from the modern survey. In our method, the components of the source are associated by finding the components from the FIRST catalog within the predicted masks by HeTu-v2.

Fig. 15 shows examples of automatic radio components association using HeTu-v2. The components with black star markers that belong to a single radio source have been successfully associated. In the bottom-right panel, there is a CJ source inside the predicted bounding box of the HT source, and our method automatically ignores the CJ source as the associated components of the HT source. However, the method of Mostert et al. (2022) finds the components within the bounding box, which can treat the component of the CJ source as false positive for the HT source. This suggests that our method is more appropriate for automatically separating radio component associations.

HeTu-v2 can also be used as a multi-class classifier when only focusing on determining whether the input image contains an interested celestial object and no need to pay attention to the location of the object and the number of celestial objects contained. HeTu-v2 based classifier allows classifying radio images



**Fig. 15.** Examples of automatic radio components association using HeTu-v2. The different colored rectangular boxes and masks represent the identified sources: magenta for FRI (fr1), red for FRII (fr2), light-cyan for HT (ht), and orange for CJ (cj). Each source is labeled with a class name and a score between 0 and 1 on the top left of the bounding box. The black star indicates the position of the FIRST radio components.

into 5 classes of radio galaxies: CS, FRI, FRII, HT, and CJ. Examples of radio galaxy classification can be seen in the first column of panels in Fig. 8. Such a classifier can automatically provide a large number of morphologically classified samples to scientists, which greatly helps to investigate and address key cosmological questions, such as the formation and evolution of SMBH.

## 5. Summary and conclusions

Radio source detection is generally done by component-based Gaussian fitting or pixels segmentation in source-finding software. These methods do not work well on extended sources and the classification of components needs to be done manually at a later stage. Such a laborious approach is not applicable to modern large-scale radio surveys. The novel method based on the deep learning object detection algorithm using Mask R-CNN can automatically segment and classify radio sources. However, the obtained segmentation masks are still coarse at the boundaries of sources. In this paper, we developed a source detector called HeTu-v2 for automated detection and classification of radio sources in high-quality segmentation. Based on the combination of Mask R-CNN and Transformer block network, HeTu-v2 has achieved excellent performance on FIRST radio source detection with mean  $AP_{50:5:95}$  of 77.8% which improved by 15.6 points and 11.3 points compared to HeTu-v1 and Mask R-CNN alone. HeTu-v2 is able to distinguish the five most common different types of radio source morphologies: CS, FRI, FRII, HT, and CJ. HeTu-v2 has established a new reliable morphological catalog for the FIRST survey with 98.6% of completeness and up to 98.5% accuracy compared to the latest release catalog of FIRST. HeTu-v2 also works well on other astronomical tasks of building sky models, radio components association, and radio galaxies classification.

For the FIRST survey, the input images for HeTu-v2 were generated by image cutouts centered on the FIRST-14dec17 catalog. This method is highly suitable for the FIRST survey, where the original images are not available, and guarantees the completeness of detection results to a great extent. However, it may not be as efficient when applied to other radio survey data because



the total area of all cutouts is much larger than the original total area of the survey. To apply the HeTu-v2 model to other modern radio surveys, the following image cutout method can be adopted. Modern radio continuum surveys often generate image data with very large sizes. For instance, in the EMU survey, each image has a pixel size of approximately  $16,000 \times 14,000$ . Radio sources appear small and clustered in the image ( $\sim 25,000$  components per image), making it difficult to accurately detect and recognize them directly from the large image using HeTu-v2. To address this issue, a sliding cropping method is typically employed to extract multiple small images from the larger image, which are then subjected to detection. To start, the large image is cropped into smaller fixed-sized images ( $M \times M$  pixels) sequentially from the top-left corner to the bottom-right corner. To avoid some sources being segmented and cut off between two small images, an overlap ( $N_{\text{overlap}}$  pixels) region is set between them. The value of  $M$  can vary between 132 and 250, whereas  $N_{\text{overlap}}$  corresponds to the pixel size of the largest radio source within the survey area. Finally, the results obtained from each small image are merged, and any duplicate sources detected within overlapping regions are removed to obtain the final result for the entire large image area.

### CRedit authorship contribution statement

**B. Lao:** Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing. **S. Jaiswal:** Data Processing and analysis, Writing – review & editing. **Z. Zhao:** Network design, Writing – original draft. **L. Lin:** Data labels. **J. Wang:** Workflow design. **X. Sun:** Data analysis, Writing – review & editing. **S.-L. Qin:** Writing – review & editing.

### 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.

### Data availability

The data used in this study are available at the FIRST survey archive (<https://third.ucllnl.org>). The software code is publicly available at <https://github.com/lao19881213/RGC-Mask-Transfiner>.

### Acknowledgments

This work was supported by National SKA Program of China (2022SKA0120101). The preliminary experiments used resources of China SKA Regional Centre prototype (An et al., 2019, 2022) funded by National SKA Program of China (2022SKA0130103) and the National Key R&D Program of China (2018YFA0404603). We thank Tao An and Ailin Wang for the helpful discussions.

### References

Adams, E.A.K., Adebahr, B., de Blok, W.J.G., Dénes, H., Hess, K.M., van der Hulst, J.M., Kutkin, A., Lucero, D.M., Morganti, R., Moss, V.A., Oosterloo, T.A., Orrù, E., Schulz, R., van Amesfoort, A.S., Berger, A., Boersma, O.M., Bouwheuis, M., van den Brink, R., van Cappellen, W.A., Connor, L., Coolen, A.H.W.M., Damstra, S., van Diepen, G.N.J., Dijkema, T.J., Ebbendorf, N., Grange, Y.G., de Goei, R., Gunst, A.W., Holties, H.A., Hut, B., Ivashina, M.V., Józsa, G.I.G., van Leeuwen, J., Loose, G.M., Maan, Y., Mancini, M., Mika, A., Mulder, H., Norden, M.J., Offringa, A.R., Oostrum, L.C., Pastor-Marazuela, I., Pisano, D.J., Ponomareva, A.A., Romein, J.W., Ruiter, M., Schoenmakers, A.P., van der Schuur, D., Sluman, J.J., Smits, R., Stuurwold, K.J.C., Versteppen, J., Vilchez, N.P.E., Vohl, D., Wierenga, K.J., Wijnholds, S.J., Woestenburger, E.E.M., Zanting, A.W., Ziemke, J., 2022. First release of Apertif imaging survey data. *Astron. Astrophys.* 667, A38. doi:10.1051/0004-6361/202244007, arXiv:2208.05348.

Ahumada, R., Allende Prieto, C., Almeida, A., Anders, F., Anderson, S.F., Andrews, B.H., Anguiano, B., Arcodia, R., Armengaud, E., Aubert, M., Avila, S., Avila-Reese, V., Badenes, C., Balland, C., Barger, K., Barrera-Ballesteros, J.K., Basu, S., Bautista, J., Beaton, R.L., Beers, T.C., Benavides, B.I.T., Bender, C.F., Bernardi, M., Bershad, M., Beutler, F., Bidin, C.M., Bird, J., Bizyaev, D., Blanc, G.A., Blanton, M.R., Boquien, M., Borissova, J., Bovy, J., Brandt, W.N., Brinkmann, J., Brownstein, J.R., Bundy, K., Bureau, M., Burgasser, A., Burtin, E., Cano-Díaz, M., Capasso, R., Cappellari, M., Carrera, R., Chabaniar, S., Chaplin, W., Chapman, M., Cherinka, B., Chiappini, C., Doohyun Choi, P., Chojnowski, S.D., Chung, H., Clerc, N., Coffey, D., Comerford, J.M., Comparat, J., da Costa, L., Cousinou, M.-C., Covey, K., Crane, J.D., Cunha, K., Ilha, G.D.S., Dai, Y.S., Damsted, S.B., Darling, J., Davidson, J., Davies, R., Dawson, K., De, N., de la Macorra, A., De Lee, N., Queiroz, A.B.d.A., Deconto Machado, A., de la Torre, S., Dell'Agli, F., du Mas des Bourboux, H., Diamond-Stanic, A.M., Dillon, S., Donor, J., Drory, N., Duckworth, C., Dwelly, T., Ebelke, G., Eftekhazadeh, S., Davis Eigenbrot, A., Elsworth, Y.P., Eracleous, M., Erfanianfar, G., Escoffier, S., Fan, X., Farr, E., Fernández-Trincado, J.G., Feuillet, D., Finoguenov, A., Fofie, P., Fraser-McKelvie, A., Frinchaboy, P.M., Fromenteau, S., Fu, H., Galbany, L., Garcia, R.A., García-Hernández, D.A., Garma Oehmichen, L.A., Ge, J., Geimba Maia, M.A., Geisler, D., Gelfand, J., Goddy, J., Gonzalez-Perez, V., Grabowski, K., Green, P., Grier, C.J., Guo, H., Guy, J., Harding, P., Hasselquist, S., Hawken, A.J., Hayes, C.R., Hearty, F., Hekker, S., Hogg, D.W., Holtzman, J.A., Horta, D., Hou, J., Hsieh, B.-C., Huber, D., Hunt, J.A.S., Ider Chitham, J., Imig, J., Jaber, M., Jimenez Angel, C.E., Johnson, J.A., Jones, A.M., Jönsson, H., Jullo, E., Kim, Y., Kinemuchi, K., Kirkpatrick, I., Kite, G.W., Klaene, M., Kneib, J.-P., Kollmeier, J.A., Kong, H., Kounkel, M., Krishnarao, D., Lacerna, I., Lan, T.-W., Lane, R.R., Law, D.R., Le Goff, J.-M., Leung, H.W., Lewis, H., Li, C., Lian, J., Lin, L., Long, D., Longa-Peña, P., Lundgren, B., Lyke, B.W., Mackereth, J.T., MacLeod, C.L., Majewski, S.R., Machado, A., Maraston, C., Martini, P., Masseron, T., Masters, K.L., Mathur, S., McDermid, R.M., Merloni, A., Merrifield, M., Mészáros, S., Miglio, A., Minniti, D., Minsley, R., Miyaji, T., Mohammad, F.G., Mosser, B., Mueller, E.-M., Muna, D., Muñoz-Gutiérrez, A., Myers, A.D., Nadathur, S., Nair, P., Nandra, K., Correa do Nascimento, J., Nevin, R.J., Newman, J.A., Nidever, D.L., Nitschelm, C., Noterdaeme, P., O'Connell, J.E., Olmstead, M.D., Oravetz, D., Oravetz, A., Osorio, Y., Pace, Z.J., Padilla, N., Palanque-Delabrouille, N., Palicio, P.A., Pan, H.-A., Pan, K., Parker, J., Paviot, R., Peirani, S., Ramfiez, K.P., Penny, S., Percival, W.J., Perez-Fournon, I., Pérez-Ràfols, I., Petitjean, P., Pieri, M.M., Pinsonneault, M., Poovelil, V.J., Povich, J.T., Prakash, A., Price-Whelan, A.M., Raddick, M.J., Raichoor, A., Ray, A., Rembold, S.B., Rezaie, M., Riffel, R.A., Riffel, R., Rix, H.-W., Robin, A.C., Roman-Lopes, A., Román-Zúñiga, C., Rose, B., Ross, A.J., Rossi, G., Rowlands, K., Rubin, K.H.R., Salvato, M., Sánchez, A.G., Sánchez-Menguiano, L., Sánchez-Gallego, J.R., Sayres, C., Schaefer, A., Schiavon, R.P., Schimoia, J.S., Schlafly, E., Schlegel, D., Schneider, D.P., Schultheis, M., Schwöpe, A., Seo, H.-J., Serenelli, A., Shafieloo, A., Shamsi, S.J., Shao, Z., Shen, S., Shetrone, M., Shirley, R., Silva Aguirre, V., Simon, J.D., Skrutskie, M.F., Slosar, A., Smethurst, R., Sobek, J., Sodi, B.C., Souto, D., Stark, D.V., Stassun, K.G., Steinmetz, M., Stello, D., Stermer, J., Storch-Bergmann, T., Streblyanska, A., Stringfellow, G.S., Stutz, A., Suárez, G., Sun, J., Taghizadeh-Popp, M., Talbot, M.S., Tayar, J., Thakar, A.R., Theriault, R., Thomas, D., Thomas, Z.C., Tinker, J., Tojeiro, R., Toledo, H.H., Tremonti, C.A., Troup, N.W., Tuttle, S., Unda-Sanzana, E., Valentini, M., Vargas-González, J., Vargas-Magaña, M., Vázquez-Mata, J.A., Vivek, M., Wake, D., Wang, Y., Weaver, B.A., Weijmans, A.-M., Wild, V., Wilson, J.C., Wilson, R.F., Wolthuis, N., Wood-Vasey, W.M., Yan, R., Yang, M., Yèche, C., Zamora, O., Zarrouk, P., Zasowski, G., Zhang, K., Zhao, C., Zhao, G., Zheng, Z., Zheng, Z., Zhu, G., Zou, H., 2020. The 16th Data Release of the Sloan Digital Sky Surveys: First Release from the APOGEE-2 Southern Survey and Full Release of eBOSS Spectra. *Astrophys. J. Suppl. Ser.* 249 (1), 3. doi:10.3847/1538-4365/ab929e, arXiv:1912.02905.

An, T., Wu, X.-P., Hong, X., 2019. SKA data take centre stage in China. *Nat. Astron.* 3, 1030. doi:10.1038/s41550-019-0943-4.

An, T., Wu, X., Lao, B., Guo, S., Xu, Z., Lv, W., Zhang, Y., Zhang, Z., 2022. Status and progress of China SKA Regional Centre prototype. *Sci. China-Phys. Mech. Astron.* 65 (12), 129501. doi:10.1007/s11433-022-1981-8, arXiv:2206.13022.

Andernach, H., Jiménez-Andrade, E.F., Willis, A.G., 2021. Discovery of 178 Giant Radio Galaxies in 1059 deg<sup>2</sup> of the Rapid ASKAP Continuum Survey at 888 MHz. *Galaxies* 9 (4), 99. doi:10.3390/galaxies9040099, arXiv:2111.08807.

Arslan, E.A., 2020. Radio galaxy morphology classification with mask R-CNN. In: Proceedings of the 2020 4th International Conference on Vision, Image and Signal Processing, pp. 1–5. doi:10.1145/3448823.3448881.

Banfield, J.K., Wong, O.I., Willett, K.W., Norris, R.P., Rudnick, L., Shabala, S.S., Simons, B.D., Snyder, C., Garon, A., Seymour, N., Middleton, E., Andernach, H., Lintott, C.J., Jacob, K., Kapińska, A.D., Mao, M.Y., Masters, K.L., Jarvis, M.J., Schawinski, K., Paget, E., Simpson, R., Klöckner, H.R., Bamford, S., Burbell, T., Chow, K.E., Cotter, G., Fortson, L., Heywood, I., Jones, T.W., Kaviraj, S., López-Sánchez, Á.R., Maksym, W.P., Polsterer, K., Borden, K., Hollow, R.P., Whyte, L., 2015. Radio Galaxy Zoo: host galaxies and radio morphologies derived from visual inspection. *Mon. Not. R. Astron. Soc.* 453 (3), 2326–2340. doi:10.1093/mnras/stv1688, arXiv:1507.07272.

Becker, R.H., White, R.L., Helfand, D.J., 1995. The FIRST survey: faint images of the radio sky at twenty centimeters. *Astrophys. J.* 450, 559. doi:10.1086/176166.

- Blanton, E.L., Gregg, M.D., Helfand, D.J., Becker, R.H., White, R.L., 2003. Discovery of a high-redshift ( $z=0.96$ ) cluster of galaxies using a first survey wide-angle-tailed radio source. *Astron. J.* 125 (4), 1635. doi:10.1086/368140, arXiv:astro-ph/0212407.
- Bolya, D., Zhou, C., Xiao, F., Lee, Y.J., 2019. Yolact: Real-time instance segmentation. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. pp. 9157–9166.
- 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. doi:10.1093/mnras/stz2845, arXiv:1908.02748.
- Capetti, A., Massaro, F., Baldi, R.D., 2017. FRICAT: a FIRST catalog of FR I radio galaxies. *Astron. Astrophys.* 598, A49. doi:10.1051/0004-6361/201629287, arXiv:1610.09376.
- Capetti, A., Massaro, F., Baldi, R.D., 2017. FRIICAT: A FIRST catalog of FR II radio galaxies. *Astron. Astrophys.* 601, A81. doi:10.1051/0004-6361/201630247, arXiv:1703.03427.
- Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S., 2020. End-to-end object detection with transformers. In: *Proceedings of the European Conference on Computer Vision*. Springer, pp. 213–229.
- Dabhade, P., Röttgering, H.J.A., Bagchi, J., Shimwell, T.W., Hardcastle, M.J., Sankhyayan, S., Morganti, R., Jamroz, M., Shulevski, A., Duncan, K.J., 2020. Giant radio galaxies in the LOFAR Two-metre Sky Survey. I. Radio and environmental properties. *Astron. Astrophys.* 635, A5. doi:10.1051/0004-6361/201935589, arXiv:1904.00409.
- Dehghan, S., Johnston-Hollitt, M., Franzen, T.M., Norris, R.P., Miller, N.A., 2014. Bent-tailed radio sources in the Australia telescope large area survey of the chandra deep field south. *Astron. J.* 148 (5), 75. doi:10.1088/0004-6256/148/5/75, arXiv:1405.6484.
- Delhaize, J., Heywood, I., Prescott, M., Jarvis, M.J., Delvecchio, I., Whittam, I.H., White, S.V., Hardcastle, M.J., Hale, C.L., Afonso, J., Ao, Y., Brienza, M., Brüggén, M., Collier, J.D., Daddi, E., Glowacki, M., Maddox, N., Morabito, L.K., Prandoni, I., Randriamanakoto, Z., Sekhar, S., An, F., Adams, N.J., Blyth, S., Bowler, R.A.A., Leeuw, L., Marchetti, L., Randriamampandry, S.M., Thorat, K., Seymour, N., Smirnov, O., Taylor, A.R., Tasse, C., Vaccari, M., 2021. MIGHTEE: are giant radio galaxies more common than we thought? *Mon. Not. R. Astron. Soc.* 501 (3), 3833–3845. doi:10.1093/mnras/staa3837, arXiv:2012.05759.
- Dong, B., Zeng, F., Wang, T., Zhang, X., Wei, Y., 2021. Solq: segmenting objects by learning queries. In: *Advances in Neural Information Processing Systems*. 34, pp. 21898–21909.
- Dutta, A., Gupta, A., Zissermann, A., 2016. VGG image annotator (VIA). URL: <http://www.robots.ox.ac.uk/vgg/software/via> 2.
- Everingham, M., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A., 2010. The pascal visual object classes (voc) challenge. *Int. J. Comput. Vis.* 88 (2), 303–338.
- Fabian, A.C., 2012. Observational evidence of active galactic nuclei feedback. *Annu. Rev. Astron. Astrophys.* 50, 455–489. doi:10.1146/annurev-astro-081811-125521, arXiv:1204.4114.
- Fanaroff, B.L., Riley, J.M., 1974. The morphology of extragalactic radio sources of high and low luminosity. *Mon. Not. R. Astron. Soc.* 167 (1), 31P–36P. doi:10.1093/mnras/167.1.31P.
- Farias, H., Ortiz, D., Damke, G., Arancibia, M.J., Solar, M., 2020. Mask galaxy: Morphological segmentation of galaxies. *Astron. Comput.* 33, 100420. doi:10.1016/j.ascom.2020.100420.
- Finkel, R.A., Bentley, J.L., 1974. Quad trees a data structure for retrieval on composite keys. *Acta Inform.* 4, 1–9. doi:10.1007/BF00288933.
- Gheller, C., Vazza, F., Bonafede, A., 2018. Deep learning based detection of cosmological diffuse radio sources. *Mon. Not. R. Astron. Soc.* 480 (3), 3749–3761. doi:10.1093/mnras/sty2102, arXiv:1809.03315.
- Greisen, E.W., 1990. The Astronomical Image Processing System. In: *Acquisition, Processing and Archiving of Astronomical Images*. pp. 125–142.
- Hale, C., Robotham, A., Davies, L., Jarvis, M., Driver, S.P., Heywood, I., 2019. Radio source extraction with ProFound. *Mon. Not. R. Astron. Soc.* 487 (3), 3971–3989. doi:10.1093/mnras/stz1462, arXiv:1902.01440.
- Hancock, P.J., Murphy, T., Gaensler, B.M., Hopkins, A., Curran, J.R., 2012. Compact continuum source finding for next generation radio surveys. *Mon. Not. R. Astron. Soc.* 422 (2), 1812–1824. doi:10.1111/j.1365-2966.2012.20768.x, arXiv:1202.4500.
- Hancock, P.J., Trott, C.M., Hurley-Walker, N., 2018. Source Finding in the Era of the SKA (Precursors): Aegean 2.0. *Mon. Not. R. Astron. Soc.* 35, e011. doi:10.1017/pasa.2018.3, arXiv:1801.05548.
- Haridas, V., O'Mullane, W., Budavári, T., Nieto-Santisteban, M., Szalay, A., Cariles, S., Thakar, A., 2005. FITS image cutout web services. In: *Astronomical Data Analysis Software and Systems XIV*, Vol. 347. p. 665.
- He, K., Girshick, R., Dollár, P., 2019. Rethinking imagenet pre-training. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. pp. 4918–4927.
- 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, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 770–778.
- Helfand, D.J., White, R.L., Becker, R.H., 2015. The last of FIRST: The final catalog and source identifications. *Astrophys. J.* 801 (1), 26. doi:10.1088/0004-637X/801/1/26, arXiv:1501.01555.
- Hopkins, A.M., 2004. On the evolution of star-forming galaxies. *Astrophys. J.* 615 (1), 209. doi:10.1086/424032, arXiv:astro-ph/0407170.
- Hotan, A.W., Bunton, J.D., Chippendale, A.P., Whiting, M., Tuthill, J., Moss, V.A., McConnell, D., Amy, S.W., Huynh, M.T., Allison, J.R., Anderson, C.S., Bannister, K.W., Bastholm, E., Beresford, R., Bock, D.C.J., Bolton, R., Chapman, J.M., Chow, K., Collier, J.D., Cooray, F.R., Cornwell, T.J., Diamond, P.J., Edwards, P.G., Feain, I.J., Franzen, T.M.O., George, D., Gupta, N., Hampson, G.A., Harvey-Smith, L., Hayman, D.B., Heywood, I., Jacka, C., Jackson, C.A., Jackson, S., Jeganathan, K., Johnston, S., Kesteven, M., Kleiner, D., Koribalski, B.S., Lee-Waddell, K., Lenc, E., Lensson, E.S., Mackay, S., Mahony, E.K., McClure-Griffiths, N.M., McConigley, R., Mirtschin, P., Ng, A.K., Norris, R.P., Pearce, S.E., Phillips, C., Pilawa, M.A., Raja, W., Reynolds, J.E., Roberts, P., Roxby, D.N., Sadler, E.M., Shields, M., Schinckel, A.E.T., Serra, P., Shaw, R.D., Sweetnam, T., Troup, E.R., Tzioumis, A., Voronkov, M.A., Westmeier, T., 2021. Australian square kilometre array pathfinder: I. system description. *Publ. Astron. Soc. Aust.* 38, e009. doi:10.1017/pasa.2021.1, arXiv:2102.01870.
- Hurley-Walker, N., Galvin, T.J., Duchesne, S.W., Zhang, X., Morgan, J., Hancock, P.J., An, T., Franzen, T.M.O., Heald, G., Ross, K., Vernstrom, T., Anderson, G.E., Gaensler, B.M., Johnston-Hollitt, M., Kaplan, D.L., Riseley, C.J., Tingay, S.J., Walker, M., 2022. GaLactic and Extragalactic All-sky Murchison Widefield Array survey eXtended (GLEAM-X) I: Survey description and initial data release. *Publ. Astron. Soc. Aust.* 39, e035. doi:10.1017/pasa.2022.17, arXiv:2204.12762.
- Jarvis, M., Taylor, R., Agudo, I., Allison, J.R., Deane, R.P., Frank, B., Gupta, N., Heywood, I., Maddox, N., McAlpine, K., Santos, M., Scaife, A.M.M., Vaccari, M., Zwart, J.T.L., Adams, E., Bacon, D.J., Baker, A.J., Bassett, B.A., Best, P.N., Beswick, R., Blyth, S., Brown, M.L., Brüggén, M., Cluver, M., Colafrancesco, S., Cotter, G., Cress, C., Davé, R., Ferrari, C., Hardcastle, M.J., Hale, C.L., Harrison, I., Hatfield, P.W., Klockner, H.R., Kolwa, S., Malefahlo, E., Marubini, T., Mauch, T., Moodley, K., Morganti, R., Norris, R.P., Peters, J.A., Prandoni, I., Prescott, M., Oliver, S., Oozeer, N., Röttgering, H.J.A., Seymour, N., Simpson, C., Smirnov, O., Smith, D.J.B., 2016. The MeerKAT International GHz Tiered Extragalactic Exploration (MIGHTEE) Survey. In: *MeerKAT Science: On the Pathway To the SKA*. p. 6. doi:10.22323/1.277.0006, arXiv:1709.01901.
- Joye, W.A., Mandel, E., 2003. New Features of SAOImage DS9. In: *Payne, H.E., Jedrzejewski, R.L., Hook, R.N. (Eds.), Astronomical Data Analysis Software and Systems XII*. In: *Astronomical Society of the Pacific Conference Series*, vol. 295, p. 489.
- Ke, L., Danelljan, M., Li, X., Tai, Y.-W., Tang, C.-K., Yu, F., 2022. Mask transfiner for high-quality instance segmentation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 4412–4421.
- Kirillov, A., Wu, Y., He, K., Girshick, R., 2020. Pointrend: Image segmentation as rendering. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 9799–9808.
- Lacy, M., Baum, S.A., Chandler, C.J., Chatterjee, S., Clarke, T.E., Deustua, S., English, J., Farnes, J., Gaensler, B.M., Gugliucci, N., Hallinan, G., Kent, B.R., Kimball, A., Law, C.J., Lazio, T.J.W., Marvil, J., Mao, S.A., Medlin, D., Mooley, K., Murphy, E.J., Myers, S., Osten, R., Richards, G.T., Rosolowsky, E., Rudnick, L., Schinzel, F., Sivakoff, G.R., Sjouwerman, L.O., Taylor, R., White, R.L., Wrobel, J., Andernach, H., Beasley, A.J., Berger, E., Bhatnager, S., Birkinshaw, M., Bower, G.C., Brandt, W.N., Brown, S., Burke-Spolaor, S., Butler, B.J., Comerford, J., Demorest, P.B., Fu, H., Giacintucci, S., Golap, K., Güth, T., Hales, C.A., Hiriart, R., Hodge, J., Horesh, A., Ivezic, Z., Jarvis, M.J., Kamble, A., Kassim, N., Liu, X., Loinard, L., Lyons, D.K., Masters, J., Mezcua, M., Moellenbrock, G.A., Mroczkowski, T., Nyland, K., O'Dea, C.P., O'Sullivan, S.P., Peters, W.M., Radford, K., Rao, U., Robnett, J., Salcido, J., Shen, Y., Sobotka, A., Witz, S., Vaccari, M., van Weeren, R.J., Vargas, A., Williams, P.K.G., Yoon, I., 2020. The Karl G. Jansky Very Large Array Sky Survey (VLASS). *Science Case and Survey Design. Publ. Astron. Soc. Pac.* 132 (1009), 035001. doi:10.1088/1538-3873/ab63eb, arXiv:1907.01981.
- Lao, B., An, T., Wang, A., Xu, Z., Guo, S., Lv, W., Wu, X., Zhang, Y., 2021. Artificial intelligence for celestial object census: the latest technology meets the oldest science. *Sci. Bull.* 66 (21), 2145–2147. doi:10.1016/j.scib.2021.07.015, arXiv:2107.03082.
- Lao, B., An, T., Yu, A., Zhang, W., Wang, J., Guo, Q., Guo, S., Wu, X., 2019. Parallel implementation of w-projection wide-field imaging. *Sci. Bull.* 64 (9), 586–594. doi:10.1016/j.scib.2019.04.004, arXiv:1905.03213.
- Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S., 2017. Feature pyramid networks for object detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 2117–2125.
- Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L., 2014. Microsoft coco: Common objects in context. In: *Proceedings of the European Conference on Computer Vision*. Springer, pp. 740–755.
- Lukic, V., de Gasperin, F., Brüggén, M., 2019. ConvoSource: Radio-Astronomical Source-Finding with Convolutional Neural Networks. *Galaxies* 8 (1), 3. doi:10.3390/galaxies8010003, arXiv:1910.03631.
- Mesarcik, M., Boonstra, A.-J., Meijer, C., Jansen, W., Rangelova, E., van Nieuwpoort, R.V., 2020. Deep learning assisted data inspection for radio astronomy. *Mon. Not. R. Astron. Soc.* 496 (2), 1517–1529. doi:10.1093/mnras/staa1412, arXiv:2005.13373.

- Mohan, N., Rafferty, D., 2015. PyBDSF: Python Blob Detection and Source Finder. Astrophysics Source Code Library, record ascl:1502.007, ascl:1502.007.
- Mostert, R.I.J., Duncan, K.J., Alegre, L., Röttgering, H.J.A., Williams, W.L., Best, P.N., Hardcastle, M.J., Morganti, R., 2022. Radio source-component association for the LOFAR Two-metre Sky Survey with region-based convolutional neural networks. *Astron. Astrophys.* 668, A28. doi:10.1051/0004-6361/202243478, arXiv:2209.14226.
- Norris, R.P., Hopkins, A.M., Afonso, J., Brown, S., Condon, J.J., Dunne, L., Feain, I., Hollow, R., Jarvis, M., Johnston-Hollitt, M., Lenc, E., Middelberg, E., Padovani, P., Prandoni, I., Rudnick, L., Seymour, N., Umana, G., Ander-nach, H., Alexander, D.M., Appleton, P.N., Bacon, D., Banfield, J., Becker, W., Brown, M.J.I., Cilieg, P., Jackson, C., Eales, S., Edge, A.C., Gaensler, B.M., Giovannini, G., Hales, C.A., Hancock, P., Huynh, M.T., Ibar, E., Ivison, R.J., Ken-nicutt, R., Kimball, A.E., Koekemoer, A.M., Koribalski, B.S., López-Sánchez, Á.R., Mao, M.Y., Murphy, T., Messias, H., Pimblett, K.A., Raccanelli, A., Ran-dall, K.E., Reiprich, T.H., Roseboom, I.G., Röttgering, H., Saikia, D.J., Sharp, R.G., Slee, O.B., Smail, I., Thompson, M.A., Urquhart, J.S., Wall, J.V., Zhao, G.B., 2011. EMU: Evolutionary Map of the Universe. *Publ. Astron. Soc. Aust.* 28 (3), 215–248. doi:10.1071/AS11021, arXiv:1106.3219.
- Padovani, P., Alexander, D., Assef, R., De Marco, B., Giommi, P., Hickox, R., Richards, G., Smolčić, V., Hatziminaoglou, E., Mainieri, V., et al., 2017. Active galactic nuclei: what's in a name? *Astron. Astrophys. Rev.* 25 (1), 1–91. doi:10.1007/s00159-017-0102-9, arXiv:1707.07134.
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimeshe, N., Antiga, L., et al., 2019. Pytorch: an imperative style, high-performance deep learning library. In: *Advances in Neural Information Processing Systems*. 32, pp. 8024–8035.
- Proctor, D.D., 2011. Morphological Annotations for Groups in the First Database. *Astrophys. J. Suppl. Ser.* 194 (2), 31. doi:10.1088/0067-0049/194/2/31, arXiv:1104.3896.
- Ren, S., He, K., Girshick, R., Sun, J., 2017. Faster r-cnn: towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 39 (6), 1137–1149. doi:10.1109/TPAMI.2016.2577031, arXiv:1506.01497.
- Riggi, S., Ingallinera, A., Leto, P., Cavallaro, F., Bufano, F., Schillirò, F., Triglio, C., Umana, G., Buemi, C.S., Norris, R.P., 2016. Automated detection of extended sources in radio maps: progress from the SCORPIO survey. *Mon. Not. R. Astron. Soc.* 460 (2), 1486–1499. doi:10.1093/mnras/stw982, arXiv:1605.01852.
- Riggi, S., Magro, D., Sortino, R., De Marco, A., Bordi, C., Cecconello, T., Hopkins, A.M., Marvil, J., Umana, G., Sciacca, E., Vitello, F., Bufano, F., Ingallinera, A., Fiameni, G., Spampinato, C., Zarb Adami, K., 2023. Astronomical source detection in radio continuum maps with deep neural networks. *Astron. Comput.* 42, 100682. doi:10.1016/j.ascom.2022.100682, arXiv:2212.02538.
- Riggi, S., Vitello, F., Becciani, U., Buemi, C., Bufano, F., Calanducci, A., Cavallaro, F., Costa, A., Ingallinera, A., Leto, P., et al., 2019. Caesar source finder: recent developments and testing. *Publ. Astron. Soc. Aust.* 36, doi:10.1017/pasa.2019.29, arXiv:1909.06116.
- Robotham, A.S.G., Davies, L.J.M., Driver, S.P., Koushan, S., Taranu, D.S., Casura, S., Liske, J., 2018. ProFound: Source Extraction and Application to Modern Survey Data. *Mon. Not. R. Astron. Soc.* 476 (3), 3137–3159. doi:10.1093/mnras/sty440, arXiv:1802.00937.
- 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. doi:10.1007/s11263-007-0090-8.
- Sasmal, T.K., Bera, S., Pal, S., Mondal, S., 2022. A New Catalog of Head-Tail Radio Galaxies from the VLA FIRST Survey. *Astrophys. J. Suppl. Ser.* 259 (2), 31. doi:10.3847/1538-4365/ac4473.
- Sault, R.J., Teuben, P.J., Wright, M.C., 1995. A retrospective view of MIRIAD. In: *Astronomical Data Analysis Software and Systems IV*, Vol. 77, p. 433.
- Shimwell, T.W., Röttgering, H.J.A., Best, P.N., Williams, W.L., Dijkema, T.J., de Gasperin, F., Hardcastle, M.J., Heald, G.H., Hoang, D.N., Horneffer, A., Intema, H., Mahony, E.K., Mandal, S., Mechev, A.P., Morabito, L., Oonk, J.B.R., Rafferty, D., Retana-Montenegro, E., Sabater, J., Tasse, C., van Weeren, R.J., Brüggen, M., Brunetti, G., Chyży, K.T., Conway, J.E., Haverkorn, M., Jackson, N., Jarvis, M.J., McKean, J.P., Miley, G.K., Morganti, R., White, G.J., Wise, M.W., van Bemmell, I.M., Beck, R., Brienza, M., Bonafede, A., Calistro Rivera, G., Cas-sano, R., Clarke, A.O., Cseh, D., Deller, A., Drabent, A., van Driel, W., Engels, D., Falcke, H., Ferrari, C., Fröhlich, S., Garrett, M.A., Harwood, J.J., Heesen, V., Hoeft, M., Horellou, C., Israel, F.P., Kapińska, A.D., Kunert-Bajraszewska, M., McKay, D.J., Mohan, N.R., Orrú, E., Pizzo, R.F., Prandoni, I., Schwarz, D.J., Shulevski, A., Sipior, M., Smith, D.J.B., Sridhar, S.S., Steinmetz, M., Stroe, A., Varenus, E., van der Werf, P.P., Zensus, J.A., Zwart, J.T.L., 2017. The LOFAR Two-metre Sky Survey. I. Survey description and preliminary data release. *Astron. Astrophys.* 598, A104. doi:10.1051/0004-6361/201629313, arXiv:1611.02700.
- Tang, C., Chen, H., Li, X., Li, J., Zhang, Z., Hu, X., 2021. Look closer to segment better: Boundary patch refinement for instance segmentation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 13926–13935.
- Taylor, M.B., 2006. STILTS - A Package for Command-Line Processing of Tabular Data. In: Gabriel, C., Arviset, C., Ponz, D., Enrique, S. (Eds.), *Astronomical Data Analysis Software and Systems XV*. In: *Astronomical Society of the Pacific Conference Series*, vol. 351, p. 666.
- Tingay, S.J., Goetze, R., Bowman, J.D., Emrich, D., Ord, S.M., Mitchell, D.A., Morales, M.F., Booler, T., Crosse, B., Wayth, R.B., Lonsdale, C.J., Trem-blay, S., Pallot, D., Colegate, T., Wicenc, A., Kudryavtseva, N., Arcus, W., Barnes, D., Bernardi, G., Briggs, F., Burns, S., Buntun, J.D., Cappallo, R.J., Corey, B.E., Deshpande, A., Desouza, L., Gaensler, B.M., Greenhill, L.J., Hall, P.J., Hazelton, B.J., Herne, D., Hewitt, J.N., Johnston-Hollitt, M., Kaplan, D.L., Kasper, J.C., Kincaid, B.B., Koenig, R., Kratzenberg, E., Lynch, M.J., McKinley, B., Mcwhirter, S.R., Morgan, E., Oberoi, D., Pathikulangara, J., Prabu, T., Remil-lard, R.A., Rogers, A.E.E., Rosh, A., Salah, J.E., Sault, R.J., Udaya-Shankar, N., Schlagenhafer, F., Srivani, K.S., Stevens, J., Subrahmanyam, R., Waterson, M., Webster, R.L., Whitney, A.R., Williams, A., Williams, C.L., Wyithe, J.S.B., 2013. The Murchison Widefield Array: The Square Kilometre Array Precursor at Low Radio Frequencies. *Publ. Astron. Soc. Aust.* 30, e007. doi:10.1017/pasa.2012.007, arXiv:1206.6945.
- Vafaei Sadr, A., Vos, E.E., Bassett, B.A., Hosenie, Z., Oozeer, N., Lochner, M., 2019. DEEPSOURCE: point source detection using deep learning. *Mon. Not. R. Astron. Soc.* 484 (2), 2793–2806. doi:10.1093/mnras/stz131, arXiv:1807.02701.
- van Haarlem, M.P., Wise, M.W., Gunst, A.W., Heald, G., McKean, J.P., Hes-sels, J.W.T., de Bruyn, A.G., Nijboer, R., Swinbank, J., Fallows, R., Brentjens, M., Nelles, A., Beck, R., Falcke, H., Fender, R., Hörandel, J., Koopmans, L.V.E., Mann, G., Miley, G., Röttgering, H., Stappers, B.W., Wijers, R.A.M.J., Zaroubi, S., van den Akker, M., Alexov, A., Anderson, J., Anderson, K., van Ardenne, A., Arts, M., Asgekar, A., Avruch, I.M., Batejat, F., Bähren, L., Bell, M.E., Bell, M.R., van Bemmell, I., Bannema, P., Bentum, M.J., Bernardi, G., Best, P., Birzan, L., Bonafede, A., Boonstra, A.J., Braun, R., Bregman, J., Breiling, F., van de Brink, R.H., Broderick, J., Broekema, P.C., Brouw, W.N., Brüggem, M., Butcher, H.R., van Cappellen, W., Ciardi, B., Coenen, T., Conway, J., Coolen, A., Corstanje, A., Damstra, S., Davies, O., Deller, A.T., Dettmar, R.J., van Diepen, G., Dijkstra, K., Donker, P., Doorduyn, A., Dromer, J., Drost, M., van Duin, A., Eislöffel, J., van Enst, J., Ferrari, C., Frieswijk, W., Gankema, H., Garrett, M.A., de Gasperin, F., Gerbers, M., de Geus, E., Griebmeier, J.M., Grit, T., Grup-pen, P., Hamaker, J.P., Hassall, T., Hoeft, M., Holties, H.A., Horneffer, A., van der Horst, A., van Houwelingen, A., Huijgen, A., Iacobelli, M., Intema, H., Jackson, N., Jelic, V., de Jong, A., Juette, E., Kant, D., Karastergiou, A., Koers, A., Kollen, H., Kondratiev, V.I., Koostra, E., Koopman, Y., Koster, A., Kuniyoshi, M., Kramer, M., Kuper, G., Lambropoulos, P., Law, C., van Leeuwen, J., Lemaitre, J., Looze, M., Maat, P., Macario, G., Markoff, S., Masters, J., McFadden, R.A., McKay-Bukowski, D., Meijering, H., Meulman, H., Mevius, M., Middelberg, E., Millenaar, R., Miller-Jones, J.C.A., Mohan, R.N., Mol, J.D., Morawietz, J., Morganti, R., Mulcahy, D.D., Mulder, E., Munk, H., Nieuwenhuis, L., van Nieuwpoort, R., Noordam, J.E., Norden, M., Noutsos, A., Offringa, A.R., Olofsson, H., Omar, A., Orrú, E., Overeem, R., Paas, H., Pandey-Pommier, M., Pandey, V.N., Pizzo, R., Polatidis, A., Rafferty, D., Rawlings, S., Reich, W., de Reijer, J.P., Reitsma, J., Renting, G.A., Riemers, P., Rol, E., Romein, J.W., Roosjen, J., Ruiter, M., Scaife, A., van der Schaaf, K., Scheers, B., Schellart, P., Schoenmakers, A., Schoonderbeek, G., Serylak, M., Shulevski, A., Sluman, J., Smirnov, O., Sobey, C., Spreeuw, H., Steinmetz, M., Sterks, C.G.M., Stiepel, H.J., Stuurwold, K., Taggar, M., Tang, Y., Tasse, C., Thomas, I., Thoudam, S., Toribio, M.C., van der Tol, B., Usov, O., van Veen, M., van der Veen, A.J., ter Veen, S., Verbiest, J.P.W., Vermeulen, R., Vermaas, N., Vocks, C., Vogt, C., de Vos, M., van der Wal, E., van Weeren, R., Weggemans, H., Weltevred, P., White, S., Wijnholds, S.J., Wilhelmsson, T., Wucknitz, O., Yatawatta, S., Zarka, P., Zensus, A., van Zwielen, J., 2013. LOFAR: The LOW-Frequency ARray. *Astron. Astrophys.* 556, A2. doi:10.1051/0004-6361/201220873, arXiv:1305.3550.
- van Velzen, S., Falcke, H., Schellart, P., Nierstenhöfer, N., Kampert, K.-H., 2012. Radio galaxies of the local universe-all-sky catalog, luminosity functions, and clustering. *Astron. Astrophys.* 544, A18. doi:10.1051/0004-6361/201219389, arXiv:1206.0031.
- Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., Liu, D., Mu, Y., Tan, M., Wang, X., Liu, W., Xiao, B., 2021. Deep high-resolution representation learning for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 43 (10), 3349–3364. doi:10.1109/tpami.2020.2983686.
- White, R.L., Becker, R.H., Helfand, D.J., Gregg, M.D., 1997. A Catalog of 1.4 GHz Radio Sources from the FIRST Survey. *Astrophys. J.* 475 (2), 479–493. doi:10.1086/303564.
- Whiting, M., Humphreys, B., 2012. Source-Finding for the Australian Square Kilometre Array Pathfinder. *Publ. Astron. Soc. Aust.* 29 (3), 371–381. doi:10.1071/AS12028, arXiv:1208.2479.
- Wieringa, M., Raja, W., Ord, S., 2020. ASKAPsoft Pipeline Gets Ready for the Pilot Surveys. In: Pizzo, R., Deul, E.R., Mol, J.D., de Plaa, J., Verkoeter, H. (Eds.), *Astronomical Data Analysis Software and Systems XXIX*. In: *Astronomical Society of the Pacific Conference Series*, vol. 527, p. 591.
- Wu, C., Wong, O.I., Rudnick, L., Shabala, S.S., Alger, M.J., Banfield, J.K., Ong, C.S., White, S.V., Garon, A.F., Norris, R.P., Ander-nach, H., Tate, J., Lukic, V., Tang, H.,



- Schawinski, K., Diakogiannis, F.I., 2019. Radio Galaxy Zoo: CLARAN - a deep learning classifier for radio morphologies. *Mon. Not. R. Astron. Soc.* 482 (1), 1211–1230. doi:[10.1093/mnras/sty2646](https://doi.org/10.1093/mnras/sty2646), arXiv:[1805.12008](https://arxiv.org/abs/1805.12008).
- Yang, S., Xiao, W., Zhang, M., Guo, S., Zhao, J., Shen, F., 2022. Image Data Augmentation for Deep Learning: A Survey. doi:[10.48550/arXiv.2204.08610](https://doi.org/10.48550/arXiv.2204.08610), arXiv:[2204.08610](https://arxiv.org/abs/2204.08610).
- Zhang, Z., Jiang, B., Zhang, Y., 2022. Automatic Detection and Classification of Radio Galaxy Images by Deep Learning. *Publ. Astron. Soc. Pac.* 134 (1036), 064503. doi:[10.1088/1538-3873/ac67b1](https://doi.org/10.1088/1538-3873/ac67b1).
- Zhang, G., Lu, X., Tan, J., Li, J., Zhang, Z., Li, Q., Hu, X., 2021. Refinemask: Towards high-quality instance segmentation with fine-grained features. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 6861–6869.