

Transformer-based source detection and morphological classification in LOFAR Deep-Field continuum images

Guangwen Chen^{1,2,3*}, Kristian Z. Adami^{1,4*}, John Abela⁵, Caijuan Yue⁶, Weibin Sun^{7,8}, Fujia Li^{7,8}, Zhaoting Chen⁹, Daniel Magro⁴, Yogesh Wadadekar¹⁰ and Leah K. Morabito^{11,12}

¹Sub-department of Astrophysics, Department of Physics, University of Oxford, Keble Road, Oxford OX1 3RH, UK

²Instituto de Astrofísica de Canarias, calle Vía Láctea s/n, E-38205 La Laguna, Tenerife, Spain

³Departamento de Astrofísica, Universidad de La Laguna, E-38206 La Laguna, Tenerife, Spain

⁴Institute of Space Sciences and Astronomy, University of Malta, Msida MSD2080, Malta

⁵Department of Information Systems, Faculty of ICT, University of Malta, Msida MSD2080, Malta

⁶School of Physics and Astronomy, Anqing Normal University, Anqing 246133, China

⁷Department of Astronomy, University of Science and Technology of China, Hefei 230026, China

⁸School of Astronomy and Space Sciences, University of Science and Technology of China, Hefei 230026, China

⁹Institute for Astronomy, School of Physics and Astronomy, The University of Edinburgh, Royal Observatory, Edinburgh EH9 3HJ, UK

¹⁰National Centre for Radio Astrophysics, Tata Institute of Fundamental Research, Post Bag 3, Ganeshkhind, Pune 411007, India

¹¹Centre for Extragalactic Astronomy, Department of Physics, Durham University, Durham DH1 3LE, UK

¹²Institute for Computational Cosmology, Department of Physics, Durham University, Durham DH1 3LE, UK

Accepted 2026 May 27. Received 2026 May 27; in original form 2026 February 10

ABSTRACT

Radio source detection and morphological classification are fundamental for exploiting the scientific potential of modern radio continuum surveys. However, the rapidly increasing data volumes and the wide diversity of radio morphologies make traditional visual inspection infeasible and pose significant challenges for automated source finding. We apply a transformer-based set-prediction detector (RF-DETR) to 150 MHz continuum images from the LOFAR Deep Fields for instance-level source detection and morphological classification. The method is adapted to multi-frequency-synthesis images of interferometric data and trained with a morphology-driven scheme using five mutually exclusive classes. The model is trained on the ELAIS-N1 Deep Field, where it achieves high detection and classification performance ($F1 \simeq 91$ per cent), and is then applied without retraining to the other three LOFAR Deep Fields. Across all four fields, the model yields consistent catalogues with modest field-to-field differences arising from survey depth and calibration. Compared with widely used PyBDSF catalogues, RF-DETR recovers the majority of PyBDSF sources while representing classical multi-component radio galaxies as single source-level detections rather than fragmented Gaussian components. Artefact-affected and spurious detections are identified as explicit classes, allowing these detections to be distinguished from general astrophysical sources in the resulting catalogues. As external validation, RF-DETR recovers the majority of visually identified extended and giant radio galaxies in the LOFAR Deep Fields and assigns them predominantly to extended morphological classes. These results indicate that transformer-based detectors provide a practical, scalable, morphology-aware approach to source finding in deep radio surveys, with clear relevance for forthcoming facilities such as SKA-Low.

Key words: methods: data analysis – methods: statistical – radio continuum: galaxies.

1 INTRODUCTION

Radio observations provide dust-unbiased views of star formation and nuclear activity in galaxies (e.g. P. N. Best & T. M. Heckman 2012; G. J. Bendo et al. 2015, 2016; M. J. Hardcastle & J. H. Croston 2020; G. Chen et al. 2023; R. K. Cochrane et al. 2023; G. Chen et al. 2024). Large-area, deep surveys with Square Kilometre Array (SKA) pathfinders, including the Low Frequency ARray

(LOFAR; M. P. Haarlem et al. 2013), the Murchison Widefield Array (S. J. Tingay et al. 2013), and the Australian SKA Pathfinder (ASKAP; A. W. Hotan et al. 2021), are transforming how we discover and characterize the extragalactic radio sky.

Accurate detection and classification are prerequisites for two complementary outcomes. On the one hand, complete, well-characterized catalogues enable number counts, angular clustering, and cross-correlations that link galaxy populations to their dark-matter environments and test structure growth on large scales (e.g. M. Jarvis et al. 2015; D. Alonso et al. 2021; C. L. Hale et al. 2024, 2025; K. Tanidis et al. 2025). On the other hand,

* E-mail: guangwen.chen@physics.ox.ac.uk (GC); kristian.zarbadami@physics.ox.ac.uk (KA)

morphological labels, in combination with multi-wavelength observations, provide indicators of the dominant emission mechanisms and facilitate the separation of star-formation-dominated systems from jetted active galactic nuclei (AGN), allowing interpretable inferences about AGN feedback, galaxy evolution, and environments (e.g. B. Mingo et al. 2019; G. Chen et al. 2020, 2022; B. Mingo et al. 2022; R. K. Cochrane et al. 2023; M. A. Horton et al. 2025; L. K. Morabito et al. 2025). Within this landscape, the LOFAR Two-metre Sky Survey (LoTSS) delivers arcsecond-scale imaging with sub-0.1 mJy beam⁻¹ sensitivity at 120–168 MHz, enabling the detection of complex, low-surface-brightness structures over tens of square degrees (T. W. Shimwell et al. 2017, 2019; C. Tasse et al. 2021; T. W. Shimwell et al. 2022, 2025). At the same time, the growing heterogeneity of radio-galaxy morphologies, spanning from compact cores to extended multi-component systems, complicates catalogue production at survey scale (L. Rudnick 2021; S. Ndung’u et al. 2023). Looking ahead, SKA Phase 1 will increase sensitivities and survey speed and confront the community with exascale data volumes, making automation essential (e.g. R. Braun et al. 2019; A. Bonaldi et al. 2021, 2025).

Classical radio source finders – such as the Python Blob Detection and Source Finder (PyBDSF; N. Mohan & D. Rafferty 2015) and AEGEAN (P. J. Hancock et al. 2012; P. J. Hancock, C. M. Trott & N. Hurley-Walker 2018) – have been the workhorses of continuum imaging for more than a decade. They model emission via thresholding and Gaussian decomposition (PyBDSF) or Laplacian-of-Gaussian clustering (AEGEAN), returning catalogues of compact/extended components with fitted shapes and fluxes. Despite their maturity, challenging regimes persist: artefacts around bright sources, diffuse or filamentary emission (e.g. N. Mohan & D. Rafferty 2015; S. Riggi et al. 2021), and physically connected multi-component systems that are split into separate islands and require a subsequent association step before cross-matching (e.g. M. Magliocchetti et al. 1998; W. L. Williams et al. 2019). These difficulties motivate methods that learn morphological priors directly from the data and provide richer, instance-level outputs beyond simple Gaussian components.

Motivated by these limitations, deep learning source finders – particularly those based on convolutional architectures – have evolved from early point-source detectors into multi-class, instance-level systems that jointly localize and label sources. DeepSource benchmarked point-source detection and reported completeness–purity gains over PyBDSF on simulations at fixed thresholds (A. Vafaei Sadr et al. 2019). Region-based models were also explored for radio maps, including CLARAN, a Faster R-CNN (S. Ren et al. 2017) variant that detects and assigns morphology labels using radio and infrared inputs (C. Wu et al. 2019). In addition, semantic-segmentation approaches such as ConvoSource were developed for source finding on SKA Science Data Challenge 1–style skies (V. Lukic, F. de Gasperin & M. Brüggén 2019a). More recently, instance-segmentation frameworks based on Mask R-CNN (K. He et al. 2017) have been adapted to radio images to jointly detect and classify compact and extended sources as well as spurious detections, enabling pixel-level masks and downstream morphological analyses; examples include the caesar-mrcnn line and related applications to ASKAP and the Meer Karoo Array Telescope (MeerKAT) data (S. Riggi et al. 2023, 2024). In parallel, one-stage detectors in the YOLO family have also been adapted to radio data; for example, YOLO-CIANNA introduces astronomy-specific adaptations and reports strong performance on the SKAO SDC1 simulations (D. Cornu et al. 2024).

Hybrid designs that blend convolutional backbones with Transformer blocks report improved handling of extended and diffuse emission via stronger context modelling (e.g. HeTu-v2; B. Lao et al. 2023). Recent work has also begun to explore the applicability of Vision Transformers (ViTs) and foundation models to astrophysical data, demonstrating strong performance in optical classification and competitive results in detection settings, while revealing significant limitations for radio galaxy morphology due to domain shift and task-specific challenges (E. Lastufka et al. 2025).

Complementary to detectors, supervised image-level classifiers operating on pre-centred cutouts – together with increasingly adopted self- and semi-supervised representation learning – aid morphology studies and improve label efficiency (e.g. W. Alhasan, A. R. Taylor & M. Vaccari 2018; V. Lukic et al. 2019b; A. Samudre et al. 2022; K. Brand et al. 2023; S. Ndung’u et al. 2024; N. Baron Perez et al. 2025), yet they presuppose existing candidates and provide no localization, and therefore cannot substitute for instance-level detection in catalogue production. In addition, practical constraints of current models persist at survey scale: anchor design and non-maximum suppression may yield duplicate or fragmented detections near bright, multi-component systems; dense-mask architectures increase annotation burden and memory footprint; and performance can degrade under domain shift across fields/instruments and when running tiled inference on large images (e.g. S. Ndung’u et al. 2023; R. Sortino et al. 2023; D. Cornu et al. 2024).

Beyond convolutional pipelines, transformer-based detectors offer a unified route to localization and labelling. The DETECTION TRansformer (DETR; N. Carion et al. 2020) casts detection as set prediction with bipartite (Hungarian) matching, which removes anchors and, by construction, duplicate assignments, while self-attention provides long-range context that is well suited to extended, multi-component radio morphology. For survey-grade cataloguing, such properties are attractive because they reduce reliance on hand-tuned anchors and non-maximum suppression and allow reasoning over diffuse structures within large tiles. A recent radio benchmark that evaluates DETR alongside Mask R-CNN and YOLO across cutouts from ASKAP, the Australia Telescope Compact Array (ATCA), the Karl G. Jansky Very Large Array (VLA), and the Radio Galaxy Zoo (RGZ; J. K. Banfield et al. 2015) project finds transformer detectors competitive yet still comparatively underused in this domain, indicating both potential and room for further development (R. Sortino et al. 2023).

Building on the original DETR framework, successors such as Deformable DETR, Conditional DETR and Lightweight DETR further improve convergence and small-object sensitivity (D. Meng et al. 2021; X. Zhu et al. 2021; Q. Chen et al. 2024). Among these, RF-DETR attains a favourable accuracy–throughput trade-off on the Common Objects in Context (COCO) benchmark by coupling a ViT encoder with a shallow DETR decoder and training-oriented refinements (I. Robinson et al. 2026). RF-DETR has already been applied beyond natural-image benchmarks (e.g. N. Dahiya et al. 2025; R. Sapkota et al. 2025), yet it has not been systematically evaluated on radio astronomical data.

In this work, we investigate whether a DETR-family, set-prediction transformer with strong self-supervised pretraining can be effectively repurposed for LOFAR deep-field continuum imaging. Specifically, we adapt RF-DETR to deliver unified, instance-level source detection with five mutually exclusive morphological classes, and assess its performance not only in terms of detection accuracy but also in terms of catalogue-level scientific

usability. Our aim is to determine whether such a framework can produce scalable and morphology-aware source catalogues across deep fields with differing depths and calibration characteristics, providing a practical reference for forthcoming SKA-Low surveys.

This paper is organized as follows. In Section 2, we describe the LoTSS Deep Fields data and the source classification scheme, together with the construction of the training and validation data sets. Section 3 introduces the RF-DETR transformer-based detector and outlines the methods adopted in this work. The model evaluation, the source catalogues, and the comparison with PyBDSF are presented in Section 4. A brief summary of the main results is given in Section 5.

2 DATA SET

2.1 LoTSS Deep Fields DR1

The LoTSS Deep Fields data release 1 (DR1)¹ provides ultra-deep 120–168 MHz imaging in four extragalactic regions: European Large-Area ISO Survey-North 1 (ELAIS-N1), Lockman Hole, Boötes, and the Euclid Deep Field North (EDFN), reaching root mean square (rms) depths in the central regions down to $\sim 20, 22, 32,$ and $32 \mu\text{Jy beam}^{-1}$ at central frequencies ~ 150 MHz, respectively (J. Sabater et al. 2021; C. Tasse et al. 2021; M. Bondi et al. 2024). These depths are comparable to those anticipated for SKA1-Low ($\sim 26 \mu\text{Jy beam}^{-1}$ at ~ 110 MHz for a 1-h continuum observation; R. Braun et al. 2019). Each pointing covers approximately 20 deg^2 , with an angular resolution of $\simeq 6$ arcsec. The Deep Fields also benefit from extensive multi-wavelength coverage from X-ray to far-infrared over a combined cross-matched area of $\sim 26 \text{ deg}^2$ (R. Kondapally et al. 2021), which we use for verification and contextual interpretation of radio sources.

It is worth noting that a DR2 release for ELAIS-N1 is now available, but DR2 products for the other Deep Fields have not yet been released. To maintain uniformity across training and evaluation, we adopt DR1 throughout and will use DR2 for future analysis. Specifically, we train and validate on the ELAIS-N1 DR1 continuum images, and assess generalization by applying the trained model to the Lockman Hole, Boötes, and EDFN DR1 continuum images.

2.2 Source classification scheme

We adopt a morphology-driven taxonomy of five *mutually exclusive* classes, designed for automated source finding in next-generation radio surveys and conceptually aligned with S. Riggi et al. (2023). Throughout, an island denotes a contiguous region of emission in the radio image as identified by standard source finders.

(i) *Compact*: single-island system with point-like or only slightly resolved structure, broadly consistent with the synthesized beam. Blending of one or more unresolved components within the same island is allowed.

(ii) *Single-Extended*: single-island system exhibiting clear, intrinsic resolved structure beyond that expected from ionospheric smearing or beam broadening (e.g. lobes, jets, or extended haloes), possibly comprising several blended components within one connected island.

(iii) *Multi-Extended*: physically associated system comprising two or more disconnected islands judged to belong to the same astrophysical source (e.g. classical double-lobed radio galaxies, a core with lobes, or multiknot jets). For such systems, a single source-level annotation is defined as a rectangular bounding box drawn to enclose the full spatial extent of the system, covering the combined area spanned by all constituent islands.

(iv) *Flagged*: single-island system with compact or extended morphology whose measurements are unreliable because of severe imaging artefacts or dynamic-range limitations near bright objects. Such sources are kept for completeness in statistics but are explicitly flagged in science catalogues since derived quantities (flux density, size, morphology) are not trustworthy.

(v) *Spurious*: false detection produced by different types of artefacts (e.g. sidelobes, ripples, striping) introduced by calibration and/or imaging. We focus on the sidelobes, which can mimic compact or elongated sources and frequently exceed standard 5σ thresholds, thereby compromising catalogue reliability.

This scheme is intentionally conservative and oriented towards source detection rather than detailed astrophysical taxonomy. It nonetheless covers nearly all kinds of sources in radio observations: *Compact* captures most ordinary radio sources; *Single-Extended* together with *Multi-Extended* encompass the majority of classical Fanaroff–Riley I/II galaxies; and *Flagged* plus *Spurious* capture artefact-related detections, supporting quality control and reliability assessment.

Definitions are based on the LOFAR 150 MHz continuum image from the ELAIS-N1 field. Ancillary images, including $4.5 \mu\text{m}$ data from the *Spitzer* Wide-area Infrared Extragalactic survey (SWIRE; C. J. Lonsdale et al. 2003), *K*-band imaging from the UK Infrared Deep Sky Survey (UKIDSS) Deep Extragalactic Survey (DXS) DR10 (A. Lawrence et al. 2007), and *i*-band data from the Panoramic Survey Telescope and Rapid Response System (Pan-STARRS; N. Kaiser et al. 2010), are used exclusively during the manual annotation process. They provide complementary morphological information to assist in judging source association and in identifying artefact-affected and spurious detections, particularly for *Multi-Extended*, *Flagged*, and *Spurious* cases. These ancillary images are not used as inputs to the model during training or inference and serve no role beyond human-guided annotation and visual verification (Fig. 1).

2.3 Data set preparation

Annotations were created by a team of trained annotators within a private Zooniverse project. Annotators performed direct visual inspection of the LOFAR 150 MHz deep-field continuum images and were not provided with any pre-existing source catalogues or automated source-finding outputs. Radio sources were identified solely on the basis of visual assessment of the continuum images, following the morphological definitions described in Section 2.2. For each identified source, annotators manually drew an axis-aligned rectangular bounding box enclosing the perceived radio emission and assigned one of the five mutually exclusive classes. Each bounding box together with its assigned class constitutes one annotation. Sources near the edges of the cutouts were treated in the same manner as interior sources and were annotated whenever clear radio emission was visible within the cutout.

Illustrative examples of all classes are shown in Fig. 2. For *Multi-Extended* systems, these bounding boxes are intended

¹<https://lofar-surveys.org/deepfields.html>

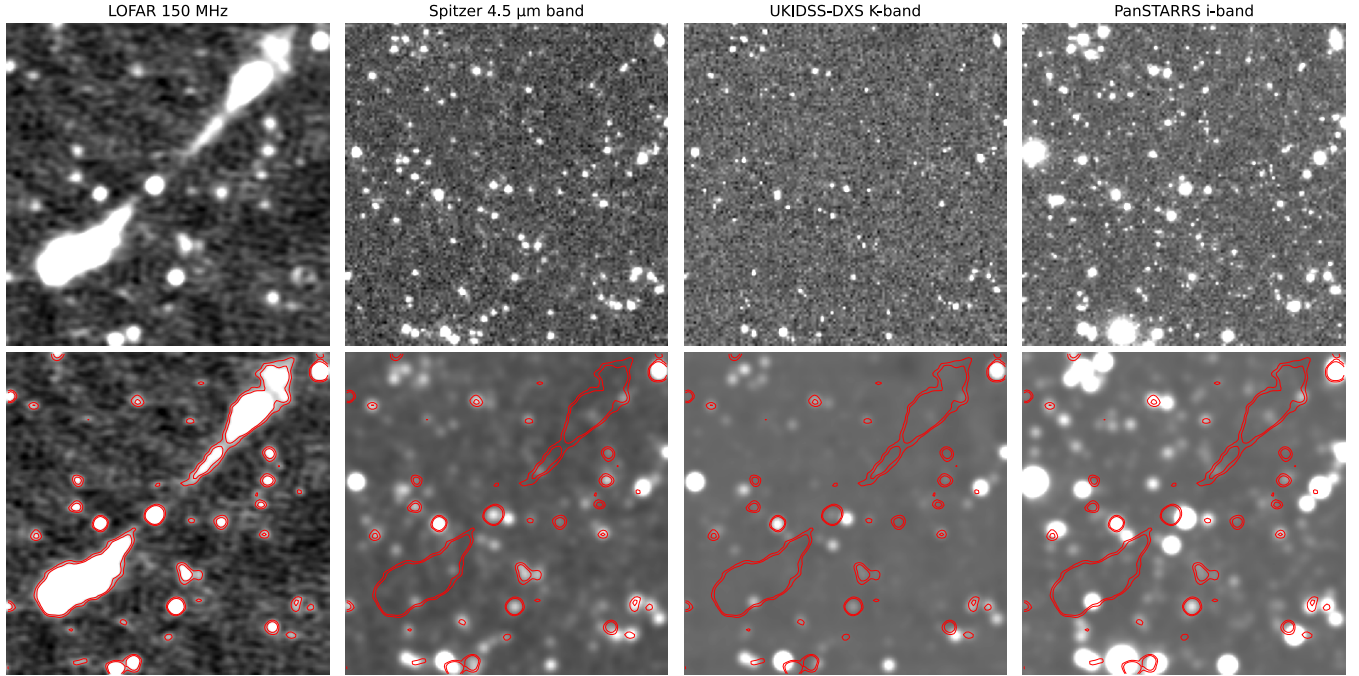


Figure 1. Examples of LOFAR 150 MHz and ancillary imaging for a 132×132 pixel cutout (3.3×3.3 arcmin at 1.5 arcsec pixel $^{-1}$), displayed with a ZScale stretch (contrast = 0.2). Left to right: LOFAR 150 MHz, *Spitzer* 4.5 μ m, UKIDSS-DXS *K*, and Pan-STARRS *i*. Upper panels: native ancillary images (no PSF convolution). Lower panels: ancillary images PSF-matched to the LOFAR synthesized beam (FWHM $\simeq 6$ arcsec) to enable direct morphological comparison. Contours trace LOFAR surface brightness at 3σ and 5σ (with σ estimated from the local rms noise). Ancillary images are used solely for visual verification and are not included in the training set.

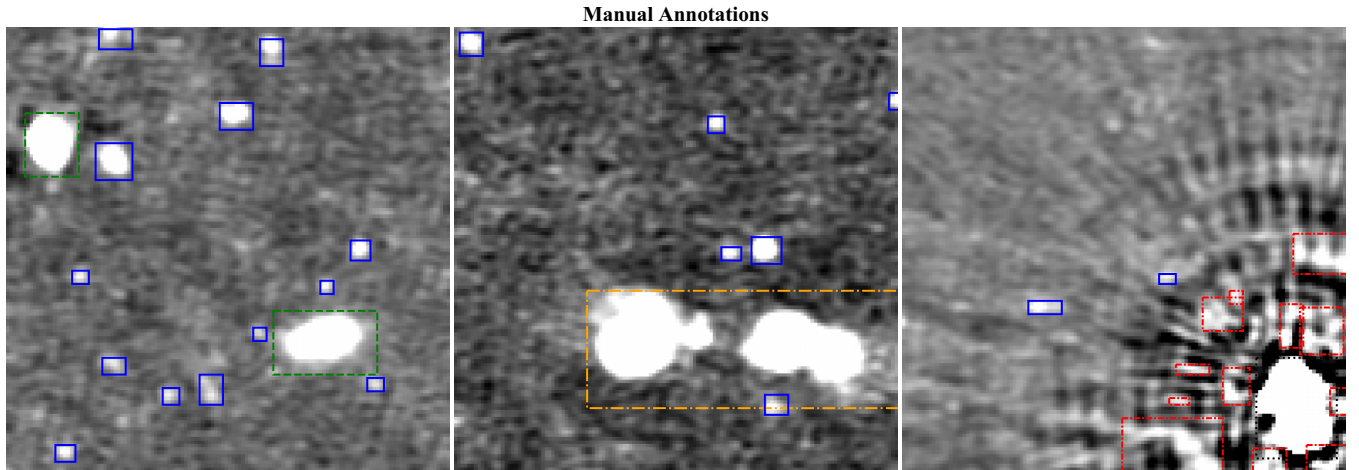


Figure 2. Examples of the annotation classes shown on 132×132 pixel cutouts from the LOFAR 150 MHz continuum images, with a ZScale stretch (contrast = 0.2). Bounding boxes are distinguished by line style: Compact (solid), Single-Extended (dashed), Multi-Extended (dash-dotted), Flagged (dotted), and Spurious (dash-dot-dotted). Left: predominantly Compact sources and two Single-Extended sources. Middle: a Multi-Extended system together with nearby Compact sources. Right: a bright source affected by severe imaging artefacts is labelled Flagged, and surrounding artefact-induced false detections are marked as Spurious. All annotations are axis-aligned rectangular bounding boxes drawn by annotators.

to represent source-level envelopes enclosing all physically associated islands, rather than precise morphological outlines. While more general inclined rectangles could in principle provide a tighter geometric description, we adopt axis-aligned bounding boxes throughout to ensure consistency with the representation required by DETR-style detectors used in this work. The set-prediction framework of RF-DETR allows multiple, potentially overlapping bounding boxes to be predicted, enabling compact or single-extended sources embedded within larger Multi-

Extended envelopes to be detected and classified without ambiguity.

To minimize confusion between genuine radio sources and artefact-induced responses around bright objects, annotators considered not only the morphology in the LOFAR 150 MHz images but also the spatial context and symmetry of emission, as well as information from ancillary infrared and optical images, which were used to identify likely sidelobes or calibration artefacts that lack plausible multi-wavelength counterparts. Ambigu-

Table 1. Class counts in our original training and validation data sets.

Class	Total	Training	Validation
Compact	12 535	10 199	2336
Single-Extended	520	409	111
Multi-Extended	112	94	18
Flagged	90	62	28
Spurious	564	444	120

ous cases in the vicinity of bright sources were conservatively labelled as *Flagged* or *Spurious*, rather than as astrophysical sources.

As a further quality-control measure, a subset of identical or overlapping cutouts was independently inspected by more than one annotator, and the resulting labels were cross-checked prior to model training to ensure consistent application of the morphological definitions. Following this quality-control step, all annotated cutouts were retained for training and validation, including those with partial spatial overlap.

Using this taxonomy, we curated 1057 132×132 pixel cutouts from the ELAIS-N1 150 MHz mosaic ($1.5 \text{ arcsec pixel}^{-1}$), yielding 13 821 labelled instances distributed as: *Compact*=12 535, *Single-Extended*=520, *Multi-Extended*=112, *Flagged*=90, *Spurious*=564. This corresponds to an average surface density of ≈ 13.1 sources per cutout, i.e. $\approx 1.21 \text{ arcmin}^{-2}$. For comparison, S. Riggi et al. (2023) trained Mask R-CNN source detectors on data with a surface density of $\approx 0.22 \text{ arcmin}^{-2}$ (with 0.15 arcmin^{-2} from VLA, 0.42 arcmin^{-2} from ASKAP, and 0.33 arcmin^{-2} from ATCA, respectively). In addition, B. Lao et al. (2023) trained the HeTu-v2 source detector on the VLA FIRST survey at $\approx 0.10 \text{ arcmin}^{-2}$. By contrast, based on deeper observations, our data set operates at $\sim 5.5\times$ and $\sim 12\times$ higher source densities, respectively, than these previous studies. While the increased depth of the LoTSS Deep Fields implies higher SNRs for many detected sources, which can aid the identification of individual radio components, the corresponding rise in source density also leads to substantially more crowded fields. This introduces additional challenges for source-level morphology classification, including increased blending, ambiguous association of neighbouring emission, and the coexistence of compact and extended structures within small angular scales. Consequently, classification in deep LOFAR images constitutes a particularly stringent test of a model’s ability to reason over complex spatial context at survey depth.

We split the data set $\approx 80:20$ by cutout into training (850) and validation (207) subsets to avoid leakage. Because the cutouts are partially overlapping, a small fraction of validation cutouts also overlap spatially with those in the training set. In practice, such overlap is mostly confined to sources located near cutout boundaries and involves only a small number of instances (≈ 1.8 per cent of the validation annotations). This limited overlap does not materially affect the statistical independence of the validation set. The details of the subsets are summarized in Table 1. No explicit rebalancing or resampling was applied to the class composition. Training and evaluation under such survey-like conditions allows us to assess the performance of RF-DETR in a realistic catalogue production setting, rather than optimizing detection or classification metrics under artificially balanced class proportions. In this context, our primary aim is to evaluate whether RF-DETR can reliably detect and classify the full range

of radio source morphologies encountered in LOFAR deep-field data, including rarer extended and multi-component systems.

To improve robustness to source orientation and image contrast variations, we apply a uniform data augmentation scheme to the training subset. Each training cutout is randomly rotated by multiples of 90° with equal probability and horizontally flipped with a probability of 0.6, so that extended and elongated sources are seen under different orientations without altering their intrinsic morphology. In addition, a ZScale-style contrast remapping is applied with probability 0.6, with the contrast parameter independently sampled per cutout in the range 0.1–0.6, to account for variations in background level and dynamic range across the images. The original training instances are retained, and multiple augmented views are generated per cutout. Taken together with the probabilities above, this results in approximately 15 per cent of training instances remaining unchanged in both geometry and contrast. All geometric transformations are applied consistently to the corresponding bounding boxes and class labels, and the augmented data are converted into the model-specific input formats described in Section 3.

3 METHODS

3.1 Transformer detector: RF-DETR

We adopt RF-DETR v1.3.0² (I. Robinson et al. 2026) as our deep-learning detector for LOFAR 150 MHz continuum images. The public implementation follows the DETR set-prediction paradigm (N. Carion et al. 2020), in which a fixed set of object queries is matched one-to-one to ground-truth instances via the Hungarian algorithm. A vision backbone together with a Transformer encoder–decoder maps the queries to class logits and normalized bounding-box parameters. A schematic overview of the RF-DETR architecture as used in this work is shown in Fig. 3.

RF-DETR builds on LW-DETR by coupling a ViT backbone with a real-time transformer decoder and introducing a number of training-effective refinements (Q. Chen et al. 2024). The backbone is initialized with DINOv2 self-supervised features to provide strong general-purpose representations across domains (M. Oquab et al. 2024), while decoder conditioning with learned reference points follows the Conditional-DETR formulation (D. Meng et al. 2021). A key component of RF-DETR is multi-scale deformable attention (X. Zhu et al. 2021). Rather than attending densely over all spatial positions, each object query samples a small number of locations around learned reference points across multiple feature scales. This sparse attention mechanism reduces the computational cost associated with high-resolution feature maps while focusing attention on informative regions of the image. In radio continuum data, where faint compact sources occupy only a few pixels and extended systems span a wide range of angular scales, this combination of multi-scale representation and spatially focused attention provides an effective framework for capturing both compact and complex morphologies.

Given the optimal Hungarian matching, the loss is computed on the matched prediction–target pairs. For each decoder output, the objective is

$$\mathcal{L} = \lambda_{\text{cls}} \mathcal{L}_{\text{cls}} + \lambda_{\ell_1} \mathcal{L}_{\ell_1} + \lambda_{\text{giou}} \mathcal{L}_{\text{giou}}, \quad (1)$$

²<https://github.com/robflow/rf-detr>

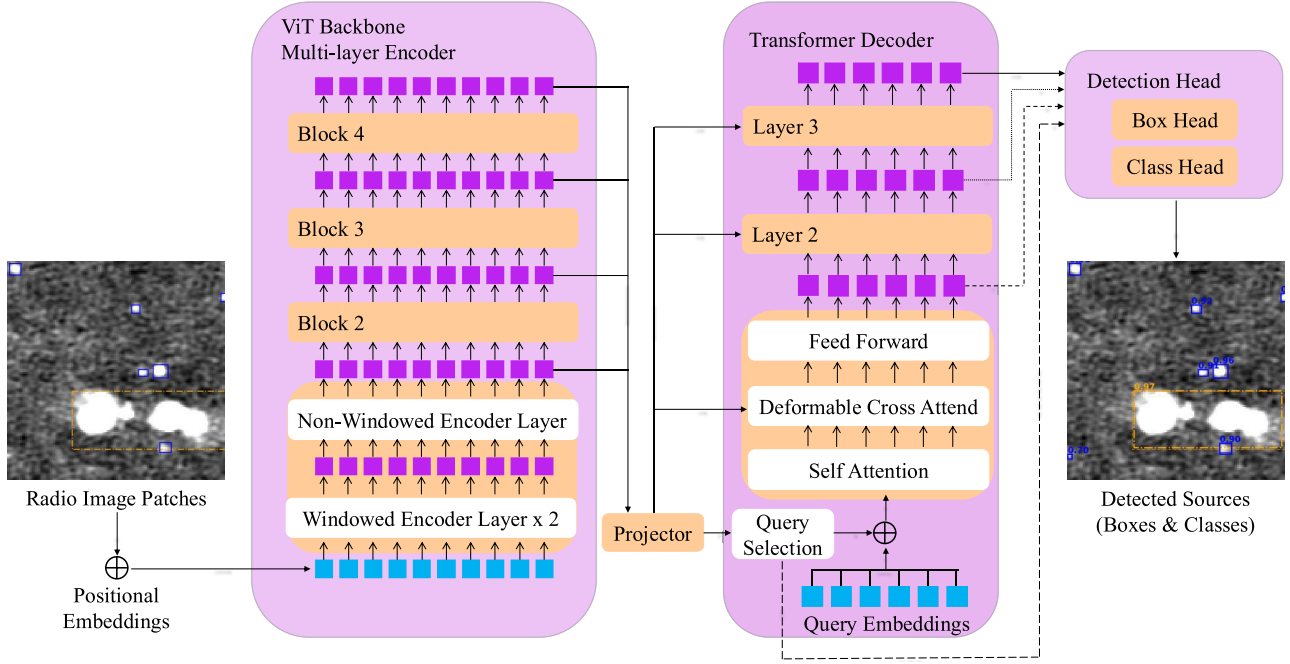


Figure 3. Schematic overview of the RF-DETR architecture as used in this work. A LOFAR 150 MHz radio image cutout is processed by a ViT-based backbone with interleaved windowed and non-windowed encoder layers to extract image features, which are then passed to a Transformer encoder-decoder. A fixed set of object queries interacts with the encoded features and produces a set of predictions through the detection head, yielding source-level bounding boxes and corresponding class labels. This formulation enables one-to-one matching between predictions and radio sources, providing a unified framework for instance-level source detection and morphological classification.

where \mathcal{L}_{cls} is a sigmoid focal classification loss, and \mathcal{L}_{ℓ_1} and $\mathcal{L}_{\text{giou}}$ are the standard L_1 and generalized-intersection-over-union (GIoU) losses for box regression. The focal formulation down-weights easy negative predictions and mitigates the strong class imbalance arising from the large number of unmatched queries. The same weighted loss is applied to intermediate decoder layers as auxiliary supervision, while the final layer provides the main predictions. A cardinality term is evaluated for monitoring purposes but is not included in the optimization.

For training, the annotations described in Section 2.3 were converted to the COCO format using axis-aligned bounding boxes parametrized as $[x, y, w, h]$, where (x, y) denotes the coordinates of the upper-left corner of the box and w and h its width and height, respectively. All geometric augmentations were applied consistently to the bounding boxes. Single-channel LOFAR cutouts were replicated to three channels to match the RGB input expected by the pre-trained encoder. We adopted the RFDETR_{Medium} variant with a DINOv2 encoder (specified as `dinov2_windowed_small`), initialized from a COCO-pretrained checkpoint. The network was trained for 200 epochs using mixed-precision training and synchronized batch normalization. Optimization employed AdamW with a base learning rate of 5×10^{-5} (encoder learning rate 2×10^{-5}), a weight decay of 0.02, and a cosine learning-rate schedule with a 10-epoch linear warm-up and a final minimum learning-rate factor of 0.05. Gradient accumulation was used (effective batch size 4×4) to stabilize optimization, and exponential moving average (EMA) tracking was enabled with a decay factor of 0.9998. Prediction-target assignment followed DETR’s bipartite (Hungarian) matching. We retained the default matching costs and loss coefficients for classification, L_1 , and GIoU, giving the usual stronger weight to the L_1 box term relative to GIoU, following standard DETR formulation

(N. Carion et al. 2020). Transformer drop-path regularization was left at the model default for this run. Checkpoints were selected by the highest validation mean Average Precision from per-epoch COCO evaluation. The main parameters are summarized in Table 2.

3.2 Evaluation protocol and metrics

We evaluate RF-DETR on the validation set under a strict one-to-one matching between ground-truth (GT) sources and model detections (DTs). For each image, we retain detections with score ≥ 0.5 and deem a GT-DT pair valid when their axis-aligned bounding boxes have intersection-over-union (IoU) ≥ 0.5 . All valid pairs are sorted by IoU and greedily matched to enforce that each GT and each DT appears in at most one pair; ties are broken by higher IoU and then by higher detection score. This protocol prevents double counting and therefore yields more conservative estimates than more permissive matching schemes.

We adopt metric definitions consistent with prior machine learning studies on radio data (e.g. S. Riggi et al. 2023, 2024). For the detection metrics, we compute

$$C = \frac{|\mathcal{A}_{1-1}|}{|G|}, \quad R = \frac{|\mathcal{A}_{1-1}|}{|D|}, \quad \text{F1}_{\text{det}} = \frac{2CR}{C+R}, \quad (2)$$

where G is the set of GT objects, D is the set of filtered detections, and $\mathcal{A}_{1-1} \subseteq G \times D$ is the one-to-one matched-pair set defined above. Here, *completeness* (C) is the fraction of true sources recovered by a matched detection, *reliability* (R) is the fraction of detections corresponding to true sources, and F1_{det} is their harmonic mean.

Table 2. Main parameters of RF-DETR used in this work.

Component	Value
<i>Architecture & initialization</i>	
Model variant	RFDETRMedium
Backbone	dinov2 windowed
Pretrained checkpoint	rf-detr-medium
Number of classes	5
Number of decoding layers	3
Number of query slots	300
<i>Optimization & schedule</i>	
Optimizer	AdamW
Base learning rate	5×10^{-5}
Backbone learning rate	2×10^{-5}
LR schedule	cosine
Warm-up epochs	10
Weight decay	0.02
<i>Batching & train length</i>	
Multi-scale training	on
Batch size	4
Gradient accumulation	4
Max epochs	200
<i>Stability</i>	
EMA	on
EMA decay	0.9998
<i>Matching & loss</i>	
Matching cost (class)	2
Matching cost (L_1)	5
Matching cost (GIoU)	2
Loss coefficient (cls)	2
Loss coefficient (bbox L_1)	5
Loss coefficient (GIoU)	2

Classification is assessed only on the matched pairs $\mathcal{A}_{1,1}$. For each class, a GT–DT pair counts as a true positive (TP) if the GT label and the predicted label are the same; otherwise it contributes a false negative (FN) for the GT class and a false positive (FP) for the predicted class. We then compute

$$\mathcal{R} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad \mathcal{P} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad \text{F1}_{\text{cls}} = \frac{2\mathcal{P}\mathcal{R}}{\mathcal{P} + \mathcal{R}}. \quad (3)$$

Here, *recall* (\mathcal{R}) is the fraction of matched GT objects of a class that are correctly classified; *precision* (\mathcal{P}) is the fraction of matched detections predicted as that class that are correct; and F1_{cls} is their harmonic mean. Both the detection metrics (F1_{det}) and the classification metrics (F1_{cls}) are computed separately for each class. The overall classification metrics referred to as ‘All’ are obtained by pooling matched pairs across classes (micro average). By construction, classification metrics are conditional on successful detection and matching; end-to-end performance is bounded by detection completeness.

All detection and classification metrics are computed in the same manner for all five classes. However, the interpretation of these metrics differs between genuine source classes and the *Spurious* category. The *Compact*, *Single-Extended*, *Multi-Extended*, and *Flagged* classes all correspond to real sources, with *Flagged* identifying systems whose measurements are unreliable owing to severe imaging artefacts or dynamic-range limitations. By contrast, the *Spurious* class represents false detections induced by imaging artefacts rather than genuine sources. For this reason, the detection completeness or reliability of the *Spurious* class is not a primary determinant of catalogue quality. The purpose of introducing an explicit *Spurious* class is to distinguish artefact-induced detections from real

sources. Accordingly, the primary concern is whether genuine astrophysical sources are incorrectly classified as *Spurious*, which would lead to their exclusion from subsequent analyses.

3.3 Inference on LoTSS Deep Field continuum maps

We run RF-DETR on each LoTSS Deep Field mosaic using a padded sliding window and a two-stage consolidation of per tile detections. Windows of 132×132 pixels with a stride of 66 pixels in both directions are laid out with end compensation so that the right and bottom edges are fully covered. Each window is extracted with a symmetric halo of $4 \times \text{FWHM}$ (in 16 pixels) using reflection padding outside the image. RF-DETR is applied to the haloed tile; we then perform per tile non-maximum suppression (NMS) and retain only boxes whose centroids fall inside the inner 132×132 region to avoid halo artefacts. Surviving detections are mapped back to mosaic coordinates and concatenated.

A class-aware consolidation step encodes radio-specific priors consistent with the annotation scheme used for training and evaluation. First, we retain detections with score ≥ 0.5 and remove boxes with width or height < 3 pixels that are consistent with noise fragments. Next, a global per-class suppression step resolves duplicate detections while applying class-dependent overlap rules that preserve *Multi-Extended* envelopes, which are intended to represent physically associated multi-component radio systems. Finally, containment and same class overlap rules prune nested or near duplicate boxes that arise from tile boundaries or closely spaced detections, keeping the highest confidence instance so that each source is represented once. These consolidation rules are designed to reflect the same source-level assumptions encoded in the training annotations. The resulting catalogue provides world coordinate system (WCS) positions, pixel space boxes, and corresponding classes as well as their fluxes. Details on the computational resources and runtimes associated with training and inference are provided in Appendix B.

4 RESULTS

4.1 Model performance on the validation set

We trained RF-DETR on the LOFAR images with the main settings listed in Table 2. Fig. 4 illustrates representative RF-DETR evaluation examples on the validation set for the same image cutouts shown in Fig. 2. In fields dominated by *Compact* sources with occasional *Single-Extended* objects (*left*), detections are dense and labels are correct. In the presence of large, multi-component systems (*middle*), the model assigns the *Multi-Extended* label to the aggregate structure while preserving nearby *Compact* sources; this behaviour is desirable for radio surveys, as accurate identification of extended and especially *Multi-Extended* morphologies facilitates the discovery of nearby Fanaroff–Riley I/II galaxies and/or giant radio galaxies. Around bright *Flagged* regions (*right*), the detector also returns *Flagged* and *Spurious* predictions; however, predicted bounding boxes can differ from the human annotations, and the model occasionally produces multiple detections associated with ambiguous artefacts, which locally reduces completeness and reliability. This is expected because artefact morphology depends on bright-source sidelobes, u coverage, and calibration residuals, and thus lacks a stable pattern. Nevertheless, the detector seldom confuses real sources with *Spurious* detections, so the

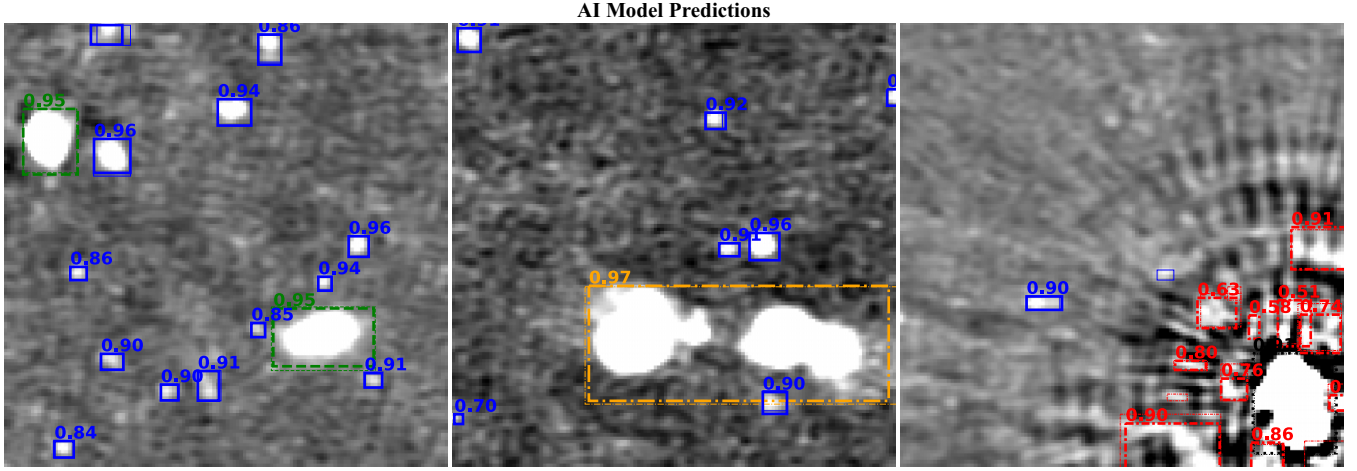


Figure 4. RF-DETR predictions on the validation set for the same image cutouts shown in Fig. 2, using the same bounding-box convention. Predicted source instances are shown with thicker bounding boxes and annotated with detector confidence scores, while ground-truth annotations are shown with thinner boxes. Only predictions with score ≥ 0.5 are displayed. The left panel shows Compact and Single-Extended predictions, the middle panel shows a Multi-Extended system together with nearby Compact sources, and the right panel contains Spurious detections around a Flagged source.

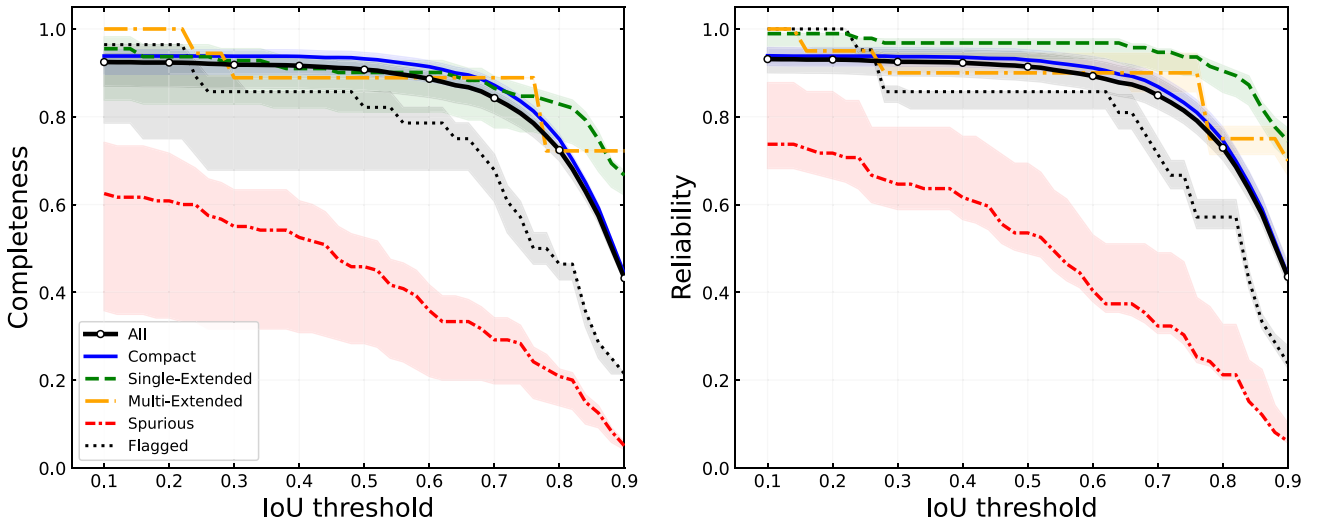


Figure 5. Detection completeness (left) and reliability (right) as a function of the IoU threshold for the full sample (All) and the five morphological classes. The curves correspond to the reference score threshold of 0.5, while shaded regions indicate the variation of score thresholds in the range 0.25–0.75.

uncertainty in these classes – being imaging artefacts rather than astrophysical objects – does not bias detections of real galaxies.

To assess the robustness of the detection metrics with respect to the IoU threshold, we examine completeness–reliability curves (as defined in Section 3.2) as a function of the IoU threshold, shown in Fig. 5. For the three main morphological classes (Compact, Single-Extended, and Multi-Extended), both completeness and reliability remain nearly unchanged over the range $0.4 \lesssim \text{IoU} \lesssim 0.6$, indicating that the detection performance is stable within this interval. At higher IoU thresholds ($\gtrsim 0.7$), both metrics decrease, as expected from the increasingly stringent matching criterion, which amplifies small spatial mismatches between predicted and reference bounding boxes. This effect is more pronounced for diffuse or extended radio emission, where axis-aligned bounding boxes act as source-level envelopes rather

than precise geometric descriptions. The shaded regions indicate the variation across detection score thresholds, demonstrating robustness to both IoU and score thresholds. As a sensitivity check, adopting more permissive IoU thresholds (e.g. $\text{IoU} \geq 0.3$) leads to a modest increase in completeness for the Single-Extended and Multi-Extended classes, without altering the relative ordering of classes or the overall conclusions. We therefore adopt IoU and score thresholds both set to 0.5 throughout this work as a conservative and uniform operating point. Table 3 reports the corresponding overall and per-class detection and classification metrics evaluated under this configuration.

Under this configuration, RF-DETR achieves strong performance on the ELAIS-N1 validation set, with an overall detection F1 score of $\simeq 91\%$ (completeness $C \simeq 91\%$, reliability $R \simeq 91\%$). Classification performance after one-to-one source match-

Table 3. Validation results for RF-DETR on the validation set using one-to-one matching at $\text{IoU} \geq 0.5$ and $\text{score} \geq 0.5$.

Class	Detection			Classification		
	Completeness (%)	Reliability (%)	F1 score (%)	Recall (%)	Precision (%)	F1 score (%)
All	90.7	91.4	91.1	98.7	98.7	98.7
Compact	93.2	92.9	93.1	99.9	99.1	99.5
Single-Extended	90.1	96.8	93.3	85.0	95.5	89.9
Multi-Extended	88.9	90.0	89.4	100	88.9	94.1
Flagged	82.1	85.7	83.9	73.9	94.4	82.9
Spurious	45.8	53.5	49.4	89.1	92.5	90.7

Table 4. RF-DETR detections in the LoTSS Deep Fields.

Field	All	Compact	Single-Extended	Multi-Extended	Flagged	Spurious
ELAIS-N1	73 539	69 441 (94.4 %)	1434 (2.0 %)	130 (0.18 %)	246 (0.3 %)	2288 (3.1 %)
Lockman Hole	75 609	69 865 (92.4 %)	1855 (2.4 %)	196 (0.26 %)	390 (0.5 %)	3303 (4.4 %)
Boötes	43 805	41 187 (94.0 %)	870 (2.0 %)	91 (0.21 %)	184 (0.4 %)	1473 (3.4 %)
EDFN	56 969	53 029 (93.1 %)	1306 (2.3 %)	142 (0.25 %)	266 (0.5 %)	2226 (3.9 %)

ing is excellent, with both recall and precision exceeding 98 per cent, yielding a classification F1 of $\simeq 99\%$. We emphasize that these classification metrics are conditional on successful detection and matching; the overall catalogue-level performance is accordingly bounded by the detection completeness. For the dominant *Compact* population, the model performs exceptionally well in both detection and classification ($F1_{\text{det}} \simeq 93\%$; $F1_{\text{cls}} \simeq 99\%$), reflecting the relative morphological simplicity and abundance of these sources. *Single-Extended* systems are detected reliably ($F1_{\text{det}} \simeq 93\%$), while classification is slightly more challenging due to their structural diversity, with recall $\simeq 85\%$, precision $\simeq 96\%$, and $F1_{\text{cls}} \simeq 90\%$. *Multi-Extended* sources exhibit moderately lower completeness ($\simeq 89\%$) but high reliability ($\simeq 90\%$), consistent with occasional partial overlaps between predicted and reference bounding regions for complex, multi-component systems. Importantly, their classification is robust, achieving perfect recall and a classification F1 of $\simeq 94\%$, demonstrating that physically associated multi-island systems are rarely misidentified once detected. The *Flagged* class shows moderate detection performance ($F1_{\text{det}} \simeq 84\%$) and lower but still reasonable classification accuracy ($F1_{\text{cls}} \simeq 83\%$), reflecting the heterogeneous morphologies and limited training support for this category. For the *Spurious* category, the reduced detection completeness is primarily driven by ground-truth *Spurious* instances that are not detected at all, rather than by systematic confusion with astrophysical source classes. These unmatched cases correspond to artefact-like structures in the radio images that do not give rise to sufficiently confident detections. This behaviour is consistent with the high classification precision achieved for the *Spurious* class (precision $\simeq 93\%$), indicating that genuine radio sources are rarely assigned a *Spurious* label once detected. The spatial association of *Spurious* detections with regions affected by strong imaging artefacts, as reflected by the *Flagged* class, further supports this interpretation. Other morphological classes do not exhibit comparable clustering around such regions. Consequently, the *Spurious* class predominantly captures artefact-related features, and does not significantly impact the identification of genuine radio sources.

Recently, S. Riggi et al. (2023) developed *caesar-mrcnn* from Mask R-CNN and successfully applied it to VLA/ASKAP/ATCA

data sets. With IoU and score thresholds of 0.6 and 0.5, they report overall $C \simeq 88\%$ and $R \simeq 64\%$, with a typical class ordering: *Compact* showing the highest in completeness, *Extended* classes trading completeness for reliability, and *Spurious* difficult to detect ($C \simeq 46\%$, $R \simeq 37\%$). Our results follow the same qualitative trend while operating at, on average, $\simeq 5.5\times$ higher source density driven by deeper LOFAR imaging (Section 2.3). To provide a controlled in-domain benchmark, we also trained the latest *caesar-mrcnn-tf2* on our training set and evaluated it on our LOFAR validation set at the same thresholds (see Table A2 in Appendix A). It remains competitive on *Compact* sources ($F1_{\text{det}} \simeq 88\%$), but shows larger gaps on the complex classes relative to RF-DETR, indicating that RF-DETR handles complex morphology better in our denser and class-imbalanced data set.

4.2 RF-DETR source catalogues in the LoTSS Deep Fields

We apply the trained RF-DETR model to the full continuum maps of all four LoTSS Deep Fields (ELAIS-N1, Lockman Hole, Boötes, and EDFN) using the tiled inference and consolidation procedure described in Section 3.3. Although the model is trained exclusively on ELAIS-N1, the same inference configuration and operating point are adopted for all fields. The four Deep Fields cover comparable sky areas of order $\sim 20 \text{ deg}^2$, but differ in achieved depth and calibration history, providing a realistic test of cross-field transfer under heterogeneous imaging conditions.

Table 4 summarizes the RF-DETR detections in each field, reporting the total number of sources and their breakdown by morphological class. As expected, fields reaching lower central rms levels, such as ELAIS-N1 and Lockman Hole, yield systematically larger detection counts than the shallower Boötes and EDFN fields. In all cases, however, the resulting catalogues are strongly dominated by *Compact* sources, which account for $\simeq 92\text{--}95\%$ per cent of all detections, in line with expectations for deep low-frequency radio surveys. The relative contributions of *Single-Extended* and *Multi-Extended* sources are small and remarkably stable across the four Deep Fields, together comprising only $\simeq 2\text{--}3\%$ per cent of the catalogues. This consistency indicates that the RF-DETR detector generalizes well beyond the training field, producing comparable morphological populations

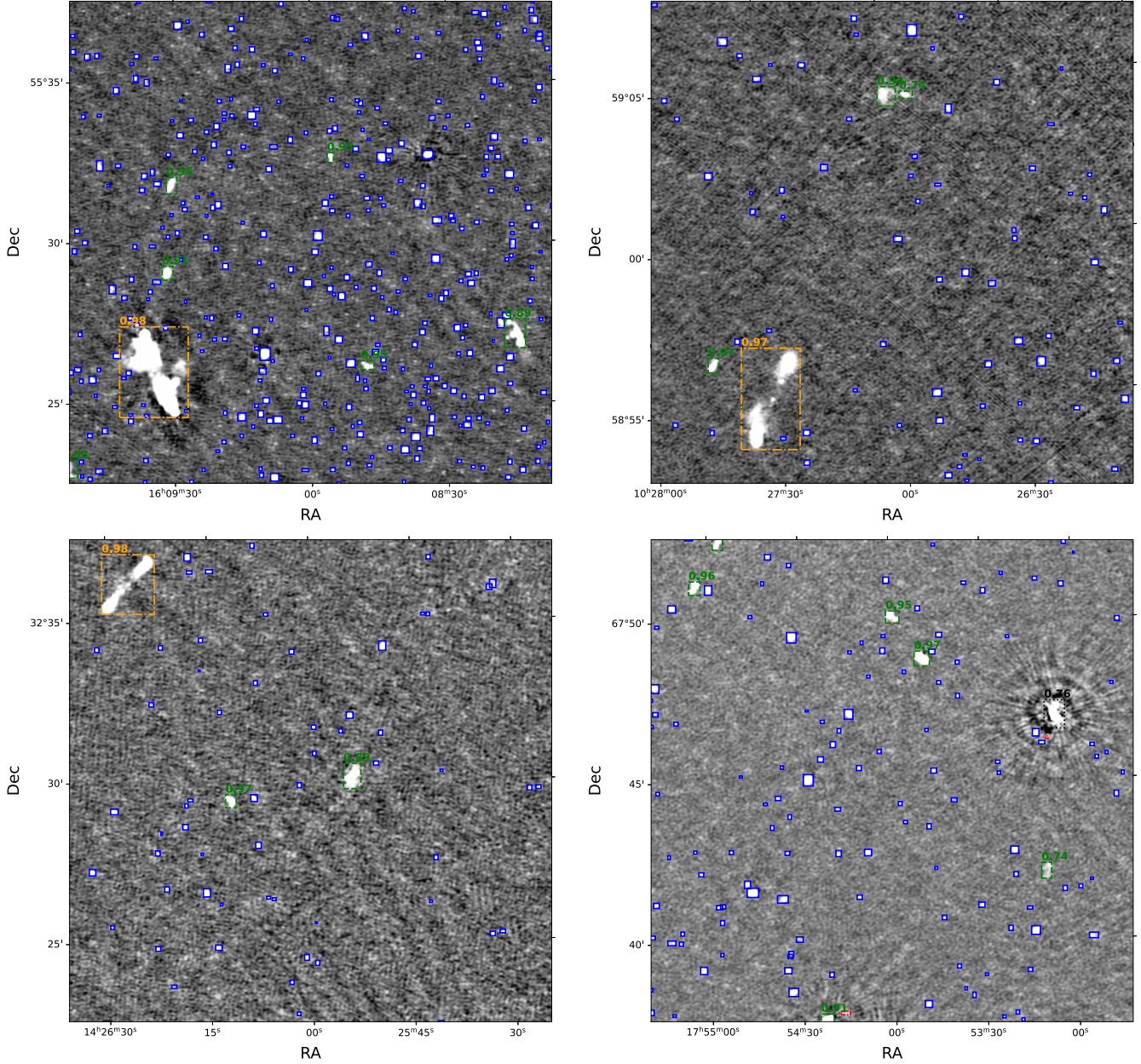


Figure 6. RF-DETR detections in randomly selected 15 arcmin \times 15 arcmin regions from the four LoTSS Deep Fields (ELAIS-N1, Lockman Hole, Boötes, and EDNF). Predicted bounding boxes follow the same conventions as in Fig. 2. Detector confidence scores are shown only for Single-Extended, Multi-Extended, and Flagged for clarity.

despite differences in depth, sky coverage, and calibration strategy. Representative cutouts from each field, selected to illustrate the range of detected morphologies, are shown in Fig. 6. Minor field-to-field variations are primarily confined to the Flagged and Spurious categories. Such behaviour is consistent with differences in imaging characteristics, including stronger sidelobes, elevated or spatially varying noise, and direction-dependent calibration residuals, rather than with intrinsic changes in the underlying source population (J. Sabater et al. 2021; C. Tasse et al. 2021; M. Bondi et al. 2024). Overall, these variations do not affect the stability of the primary source populations and have no impact on the scientific interpretation of the resulting catalogues.

4.3 Comparison with PyBDSF catalogues

Having characterized the RF-DETR catalogues in the LoTSS Deep Fields, we now compare them with the widely used component-based catalogues produced by PyBDSF on the same continuum maps so as to assess catalogue-level completeness, overlap, and structural differences between the two approaches. The PyBDSF catalogues used in this comparison are the publicly released LoTSS Deep Fields DR1 products (e.g. J. Sabater et al. 2021; C. Tasse et al. 2021). To establish correspondences between the two catalogues, we adopt a PyBDSF-centric containment-based matching scheme. Specifically, a PyBDSF source is considered matched if its catalogue position (RA, Dec.) lies within the area of a predicted RF-DETR bounding box. In cases where a PyBDSF

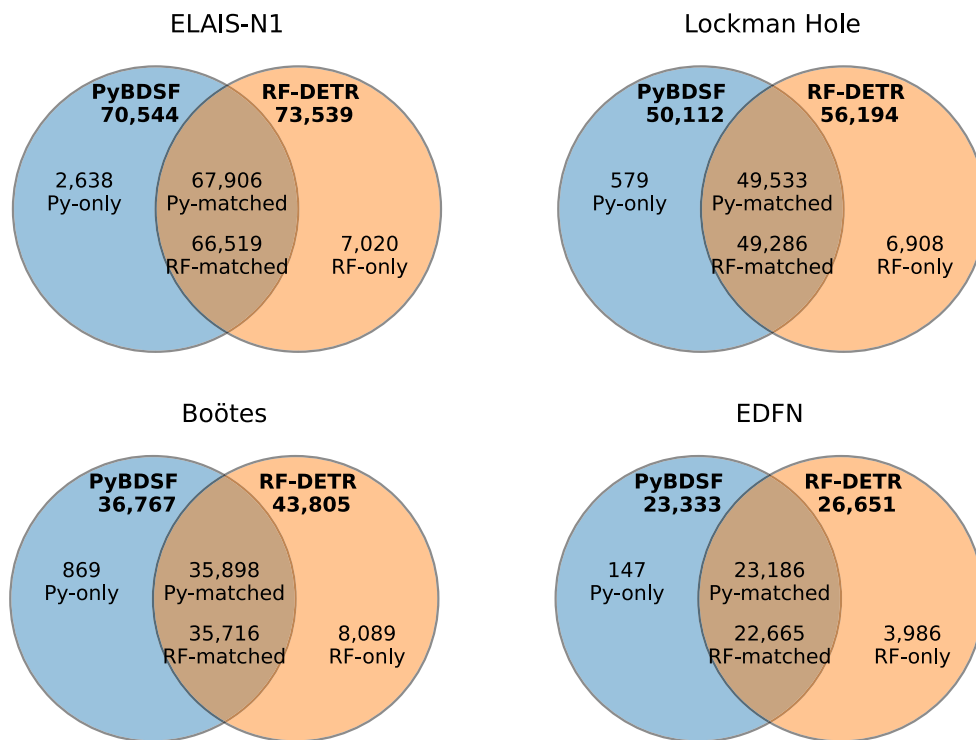


Figure 7. Venn diagrams illustrating the catalogue overlap between PyBDSF and RF-DETR in the four LoTSS Deep Fields. Numbers inside each circle indicate the number of detections in the corresponding catalogue. The overlap region shows the number of Py-matched sources, defined as PyBDSF sources that are associated with at least one RF-DETR bounding box, as well as the corresponding number of RF-matched detections. Py-only sources correspond to PyBDSF detections without an RF-DETR counterpart, while RF-only detections denote RF-DETR bounding boxes that are not matched to any PyBDSF source. Specifically, for the Lockman Hole and EDFN fields, both RF-DETR and PyBDSF counts are restricted to the central regions covered by the publicly available PyBDSF catalogues, matching the sky coverage used for the comparison. See more details in the text.

source falls within more than one RF-DETR bounding box, the match is assigned to the RF-DETR detection whose bounding-box centre is closest to the PyBDSF position. Because multiple PyBDSF sources can be associated with a single RF-DETR bounding box, the number of Py-matched sources is systematically larger than the number of RF-DETR detections that contain at least one PyBDSF source.

Fig. 7 illustrates the overlap between RF-DETR and PyBDSF catalogues across all four LoTSS Deep Fields, showing the total number of detections, as well as the Py-only, RF-only, and matched populations. The majority of PyBDSF sources are successfully recovered by RF-DETR. In the ELAIS-N1 field, approximately 96 per cent of PyBDSF sources (67 906 out of 70 544) are matched to at least one RF-DETR detection, with similarly high recovery fractions ($\gtrsim 97\%$) observed in the other fields. In addition to this high level of overlap, RF-DETR yields a slightly larger number of catalogue entries than PyBDSF. This difference reflects the distinct detection strategies adopted by the two methods. At the operating point used in Section 3, RF-DETR applies a single confidence threshold to the detector’s class probabilities, without a subsequent island-growth stage. In contrast, PyBDSF first identifies emission islands via pixel-thresholding and then fits one or more Gaussians; sources that do not seed an island are therefore not included in the catalogue. As a result, spatially coherent but low-surface-brightness sources that fall below per-pixel thresholds may be recovered by RF-DETR while remaining absent from PyBDSF. For the Lockman Hole and EDFN fields,

the publicly available PyBDSF catalogues are restricted to the central regions of the continuum maps and therefore contain fewer sources overall (50 112 and 23 333, respectively). Accordingly, the RF-DETR counts reported in Fig. 7 for these two fields are restricted to the same sky areas covered by the PyBDSF catalogues. In view of the reduced sky coverage of the available catalogues for the Lockman Hole and EDFN fields, and to avoid introducing additional field-dependent systematics associated with variations in noise properties and survey depth, we therefore focus the detailed comparison between RF-DETR and PyBDSF on the full ELAIS-N1 map, which serves as the deepest and most uniformly sensitive LoTSS Deep Field and thus provides a natural reference for catalogue-level comparisons (e.g. J. Sabater et al. 2021).

Fig. 8 compares total (left) and peak (right) flux densities for matched RF-DETR and PyBDSF sources in the ELAIS-N1 field. Total flux densities are obtained by summation within each RF-DETR bounding box and within the associated PyBDSF island, while peak flux densities correspond to the maximum pixel intensity inside the RF-DETR bounding box and the fitted Gaussian peak for PyBDSF. For bright sources, including those classified as *Flagged*, the flux densities measured by RF-DETR and PyBDSF closely follow the one-to-one relation in both total and peak flux, down to approximately 0.01 Jy and 0.01 Jy beam⁻¹, respectively. *Flagged* sources are identified as real radio sources rather than false detections; this class reflects limitations in flux measurements caused by dynamic-range and calibration artefacts around bright emission, rather than uncertainty in source identification

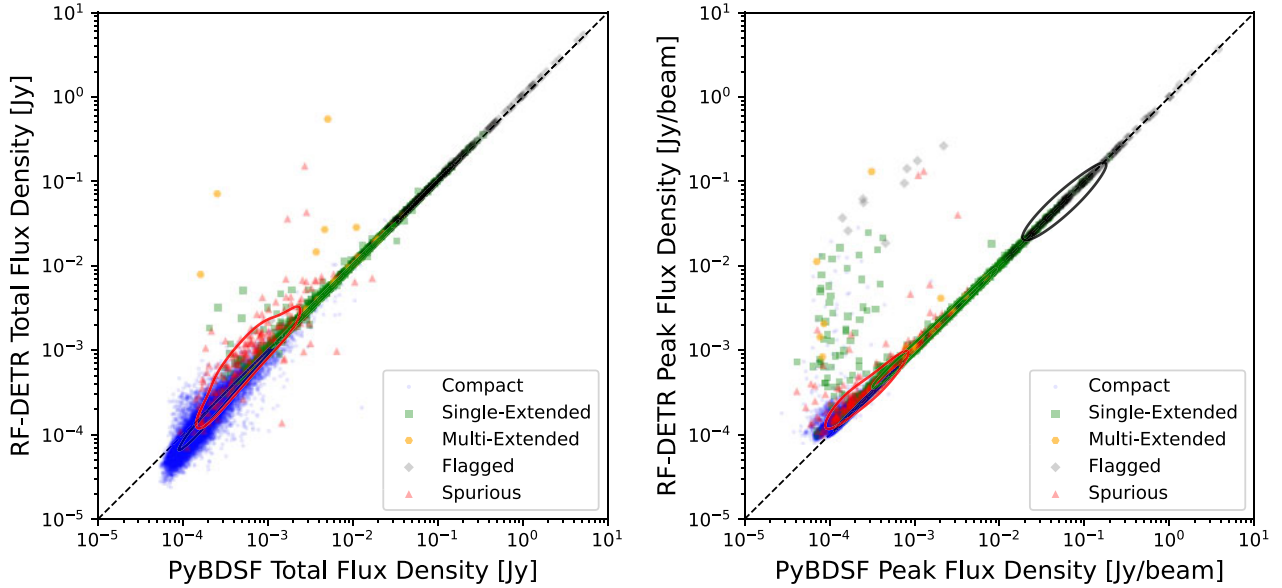


Figure 8. Comparison of RF-DETR and PyBDSF flux densities for matched sources in the ELAIS-N1 field. Left panel: comparison of total flux densities obtained via direct pixel summation within RF-DETR bounding boxes versus PyBDSF islands. Right panel: comparison of peak flux densities, where RF-DETR values correspond to the maximum pixel intensity within the bounding boxes, while PyBDSF measurements are derived from Gaussian component fitting within the associated islands. Data points are distinguished according to the RF-DETR morphological class. For classes with more than 30 matched sources, kernel-density contours are overlaid to enclose the densest 68.3 per cent of the points.

(Section 2.2). Accordingly, these sources are retained in the catalogue for completeness and source association, while their flux densities are best refined through dedicated re-imaging or improved direction-dependent calibration when used for quantitative analyses.

Deviations from the one-to-one relation reflect methodological differences: RF-DETR often reports higher total fluxes for *Multi-Extended* systems because a single RF-DETR bounding box can enclose spatially separated components that PyBDSF treats as multiple islands. A systematic excess is also seen for a subset of RF-DETR *Single-Extended* sources, particularly in the peak flux comparison. This may reflect differences in how PyBDSF peak fluxes are derived from Gaussian component fitting and can be suppressed for structured extended emission, while RF-DETR values are measured directly from pixel-level intensities and are more sensitive to localized brightness enhancements. At lower flux densities, *Spurious* detections exhibit substantial scatter, and since these detections are produced by different types of artefacts, their flux density measurements are not expected to be reliable and should not be used for quantitative analyses. *Compact* sources display increased scatter towards the faint end, and a subset of these sources show systematically lower RF-DETR total flux densities relative to PyBDSF. This behaviour is consistent with differences in how low-surface-brightness emission near the detection threshold is treated by the two approaches, particularly for faint *Compact* sources whose extended emission can be partially excluded by RF-DETR bounding boxes, and this effect is further quantified below.

Fig. 9 places these source-level differences into a population context by comparing total flux distributions decomposed by morphological class. The overall RF-DETR flux distribution closely tracks PyBDSF at intermediate and high fluxes, including *Single-Extended*, *Multi-Extended*, and *Flagged* classes. Notably, RF-DETR shows a modest excess of faint detec-

tions relative to PyBDSF that is dominated by *Compact* sources. To interpret this excess, we examine the detection behaviour as a function of signal-to-noise ratio (SNR), focusing on *Compact* sources where discrepancies are most pronounced. The peak SNR is defined as the ratio of the peak flux density to the local background rms. Using the matching method described above, approximately 7000 RF-DETR detections are not matched to any PyBDSF source, of which about 5000 are classified as *Compact* and about 2000 as *Spurious*. These RF-DETR-only *Compact* detections are predominantly low-SNR candidates; approximately half of these sources have peak SNR below 5, with a median peak SNR of $\simeq 5.1$, compared to a median peak SNR of 9.0 for *Compact* sources matched between RF-DETR and PyBDSF. In addition, these RF-DETR-only *Compact* detections are not spatially clustered around regions classified as *Flagged* or *Spurious*; none lie within 10 arcsec of such regions, and fewer than 5 per cent are located within 60 arcsec. This indicates that the majority of the additional low-SNR compact detections returned by RF-DETR are not trivially attributable to obvious imaging artefacts, but instead represent a broader population of low-SNR compact source candidates near the detection threshold. Conversely, approximately 2600 PyBDSF sources are not matched to any RF-DETR detection. These unmatched PyBDSF sources are systematically faint, with a median total flux density of 2.5×10^{-4} Jy, compared to 3.4×10^{-4} Jy for PyBDSF sources recovered by RF-DETR. A Kolmogorov–Smirnov test confirms that the detected and missing PyBDSF populations are statistically distinct ($p \ll 0.01$), indicating that incompleteness in both catalogues is concentrated towards the low-SNR, faint source regime. These results indicate that RF-DETR tends to return a larger population of low-SNR compact candidates, whereas PyBDSF is more conservative in this regime but may recover some faint Gaussian components that RF-DETR does not.

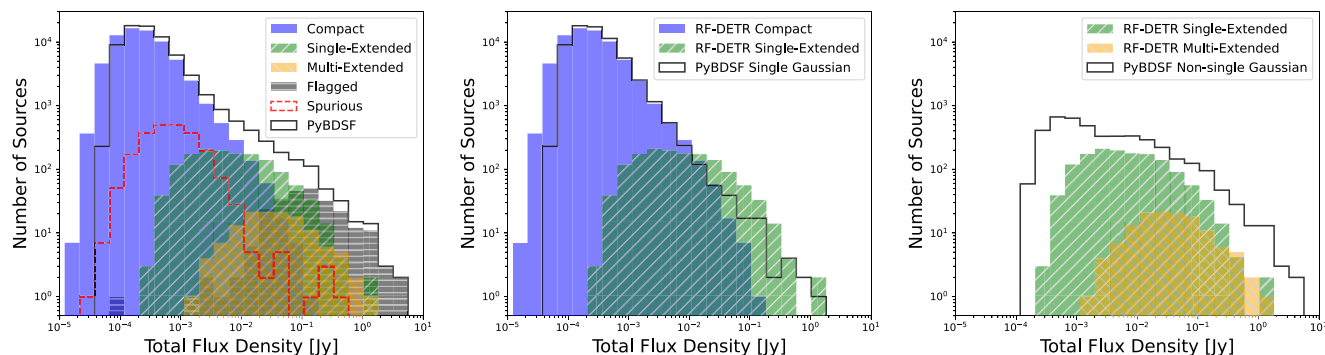


Figure 9. Total flux distributions for sources detected in the ELAIS-N1 field. Left: RF-DETR catalogue decomposed by morphological class (Compact, Single-Extended, Multi-Extended, Flagged and Spurious), overlaid with the PyBDSF total flux distribution. Middle: comparison between RF-DETR Compact, Single-Extended and PyBDSF single Gaussian sources. Right: comparison between RF-DETR extended classes and PyBDSF component groupings (M and C).

Nevertheless, the modest increase in RF-DETR faint detections does not lead to a higher catalogue-level integrated flux density; summing the total flux densities over all detected sources in the ELAIS-N1 field yields 150 Jy for RF-DETR, compared to 157 Jy for the PyBDSF catalogue, corresponding to a relative difference of -4.6 per cent. In particular, RF-DETR total fluxes are obtained by direct pixel summation within annotation-driven bounding boxes, which can be conservative around low-surface-brightness emission, while PyBDSF measurements integrate emission over the contiguous islands, including flux associated with fitted Gaussian wings (N. Mohan & D. Rafferty 2015). As a result, although RF-DETR returns a slightly larger number of faint detections, their individual flux contributions are typically small and do not offset the systematically larger per-source apertures adopted by PyBDSF, leading to a modestly lower total integrated flux density for RF-DETR at the catalogue level. In this sense, RF-DETR can be regarded as a generator of faint compact source candidates, for which SNR-based post-processing or additional validation should be applied to balance completeness and reliability according to specific scientific objectives.

Fig. 10 presents a qualitative comparison between RF-DETR detections (left panels) and PyBDSF catalogue entries (right panels) over representative $15 \text{ arcmin} \times 15 \text{ arcmin}$ regions of the ELAIS-N1 field. The upper panels illustrate a region containing a classical double-lobed radio galaxy embedded in a dense field of compact sources. RF-DETR assigns a single Multi-Extended detection whose bounding box encloses both lobes and the core, corresponding to the astrophysical interpretation of a single radio source. In contrast, the PyBDSF catalogue decomposes the same system into multiple Gaussian components, listed as separate catalogue entries. It is important to note that the PyBDSF structural flag encodes the composition of the Gaussian model (S: single Gaussian; M: multiple Gaussians; C: single Gaussian within an island containing other sources), rather than a source-level morphological classification (N. Mohan & D. Rafferty 2015). As a result, physically connected multi-component systems appear as multiple rows in the PyBDSF catalogue and require an additional association step before they can be treated as a single astrophysical source. By construction, RF-DETR instead provides a unified, source-level entry, which is advantageous for cross-identification and catalogue compactness.

The lower panels show a region dominated by three bright sources with pronounced sidelobes and deconvolution artefacts.

RF-DETR correctly identifies the bright artefact-affected systems as Flagged and labels a small number of nearby responses as Spurious, while simultaneously preserving numerous genuine faint Compact detections in the surrounding area. PyBDSF relies on fitting Gaussian components to emission islands identified above user-defined thresholds and does not provide explicit labels for artefacts. While PyBDSF supports adjustable parameters such as rms estimation boxes and island thresholds that can be tuned to mitigate the inclusion of sidelobe and other instrumental artefacts, these settings do not categorically exclude artefact emission. In practice, under conditions of elevated and spatially varying noise due to bright sources, artefact-dominated islands can still be fitted with Gaussian components that enter the PyBDSF catalogue. The explicit Flagged and Spurious labels produced by RF-DETR thus enable artefact-aware catalogue filtering without image-specific parameter tuning, while retaining genuine sources in the vicinity of bright objects.

4.4 Recovery of extended and giant radio galaxies

To further assess the practical performance of our RF-DETR-based source finder, we cross-matched our LOFAR deep-field catalogues with the sample of extended radio galaxies (ERGs) compiled by M. Simonte et al. (2024), which was constructed through careful visual inspection of the ELAIS-N1, Lockman Hole and Boötes deep fields. Their study provides a stringent external benchmark, as it targets large, morphologically complex radio galaxies that are known to be challenging for automated source-finding algorithms.

Restricting the comparison to ERGs whose host positions lie within the LOFAR image footprints, we find that RF-DETR successfully recovers the vast majority of these sources across all three fields. Overall, approximately 94 per cent (1512 out of 1609) of the ERGs are detected by RF-DETR, while the recovery fraction for giant radio galaxies (GRGs; defined as systems with projected linear sizes $LLS \geq 0.7 \text{ Mpc}$) remains high at ≈ 86 per cent. In this analysis, a source is considered recovered if the host position reported by M. Simonte et al. (2024) falls within at least one RF-DETR bounding box, which constitutes a conservative matching criterion, particularly for very large, lobe-dominated systems. More than 83 per cent of recovered ERGs and GRGs are assigned to extended morphological classes by RF-DETR, either Single-Extended or Multi-Extended. The latter class is frequently

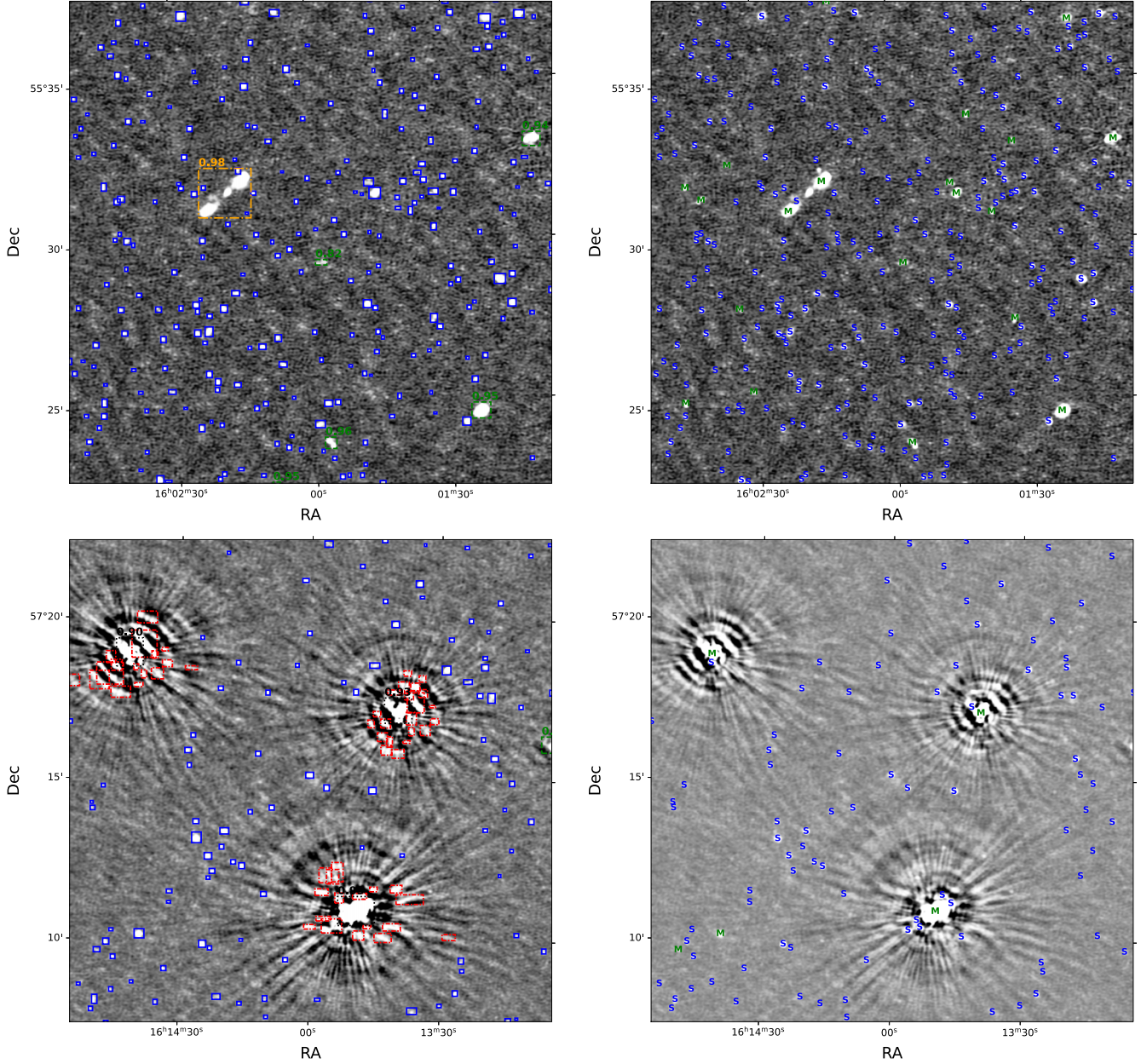


Figure 10. Left: RF-DETR detections overlaid on the 15 arcmin \times 15 arcmin cutouts of ELAIS-N1 mosaic. Right: PyBDSF catalogue entries on the same regions, with labels indicating single Gaussian (S) and multiple Gaussians (M) sources.

associated with classical double-lobed and multi-component systems, consistent with the physical nature of large radio galaxies. Only a small fraction of ERGs are classified as Flagged or Compact, indicating that RF-DETR does not confuse genuine extended emission with artefacts or noise features.

Beyond recovering the majority of visually identified ERGs, RF-DETR assigns a substantially larger number of sources in the LoTSS Deep Fields to extended morphological classes than those included in the M. Simonte et al. (2024) catalogue. While not all sources classified as Single-Extended or Multi-Extended by RF-DETR necessarily correspond to genuine ERGs as defined by projected linear size, this indicates the presence of additional extended radio galaxy candidates beyond those identified in visual catalogues. Given that compact sources constitute more than 90 per cent of detections in LoTSS Deep Fields, starting

from the RF-DETR extended classes provides a highly efficient and physically motivated pre-selection for subsequent ERG and GRG searches. These results demonstrate that RF-DETR not only achieves high completeness on a visually curated sample of extended radio galaxies but also produces morphologically consistent classifications in a fully automated and scalable manner, underscoring its potential for large-scale, systematic searches for extended and giant radio galaxies in forthcoming wide-area radio surveys.

5 CONCLUSIONS

We have demonstrated that a transformer-based, set-prediction detector, RF-DETR, can be successfully adapted for instance-level source detection and morphology classification in deep,

low-frequency radio continuum surveys. Our approach involves a balance between survey-scale source detection and capturing the complexity of radio morphologies, and adopts a conservative morphology scheme designed for automated source finding. Applied to the LoTSS Deep Fields at 150 MHz, the model delivers unified, source-level catalogues that simultaneously localize radio emission and assign physically motivated morphological labels.

Trained on the ELAIS-N1 Deep Field and applied without retraining to three additional fields, RF-DETR achieves robust and stable performance under realistic deep-field conditions. For the main source populations, including `Compact`, `Single-Extended`, and `Multi-Extended` systems, the model attains detection and classification performance at the $\gtrsim 90$ per cent level on a held-out validation set. Importantly, physically associated multi-lobed systems are preferentially detected as single instances, reflecting the astrophysical concept of a radio source rather than a collection of individual components.

At the catalogue level, RF-DETR recovers the vast majority of sources identified by the widely used PyBDSF source finder and maintains good overall agreement in flux measurements, while exhibiting distinct detection behaviour in specific regimes. In particular, RF-DETR provides source-level representations of classical multi-component radio galaxies as unified entries rather than fragmented Gaussian components, and returns a broader set of compact source candidates near the detection threshold. Beyond conventional morphology classes, the explicit identification of `Flagged` and `Spurious` detections enables artefact-aware filtering and facilitates reliable catalogue construction in the vicinity of bright sources.

As an external validation, we compared the RF-DETR catalogues with a visually curated sample of extended and giant radio galaxies. The detector successfully recovers the majority of these large, morphologically complex systems and assigns them predominantly to extended morphological classes. This demonstrates its effectiveness on physically associated, multi-component radio galaxies that are known to challenge traditional automated source-finding approaches.

These results demonstrate that modern transformer-based detectors can provide a practical and scalable framework for morphology-aware source finding in deep radio continuum surveys. The RF-DETR approach is well suited for source detection and population-level morphological studies, with clear applicability to forthcoming large-area surveys with facilities such as SKA-Low.

ACKNOWLEDGEMENTS

This work was funded by the International Science Partnerships Fund (ISPF), and this grant was awarded by the Science and Technology Facilities Council (STFC) on behalf of UK Research and Innovation (UKRI). GC also acknowledges financial support provided by the Spanish Ministerio de Ciencia, Innovación y Universidades (MICIU) through the project PID2023-153342NB-I00/10.13039/501100011033. LKM is grateful for support from ISPF via STFC (ST/Y004159/1). YW acknowledges the support of the Department of Atomic Energy, Government of India, under project no. 12-R&D-TFR5.02-0700. ZC was funded by a UKRI Future Leaders Fellowship (grant MR/X005399/1; PI: Alkistis Pourtsidou).

This publication also used data generated via the [Zooniverse.org](https://zooniverse.org) platform, development of which was funded by generous

support, including a Global Impact Award from Google, and by a grant from the Alfred P. Sloan Foundation.

DATA AVAILABILITY

The LoTSS Deep Fields DR1 continuum images are publicly available through the LoTSS data releases, and ancillary imaging data (SWIRE, UKIDSS-DXS, Pan-STARRS) are distributed by their respective surveys. The trained RF-DETR model weights and source catalogues generated in this work are publicly available at: <https://github.com/GuangwenChen/RF-DETR-LoTSS-DeepFields>. The annotation data underlying this work are not publicly released due to data-use constraints associated with the original annotation platform, but may be made available upon reasonable request to the corresponding authors.

REFERENCES

- Alhassan W., Taylor A. R., Vaccari M., 2018, *MNRAS*, 480, 2085
 Alonso D., Bellini E., Hale C., Jarvis M. J., Schwarz D. J., 2021, *MNRAS*, 502, 876
 Banfield J. K. et al., 2015, *MNRAS*, 453, 2326
 Baron Perez N., Brügger M., Kasieczka G., Lucie-Smith L., 2025, *A&A*, 699, A302
 Bendo G. J., Beswick R. J., D’Cruze M. J., Dickinson C., Fuller G. A., Muxlow T. W. B., 2015, *MNRAS*, 450, L80
 Bendo G. J., Henkel C., D’Cruze M. J., Dickinson C., Fuller G. A., Karim A., 2016, *MNRAS*, 463, 252
 Best P. N., Heckman T. M., 2012, *MNRAS*, 421, 1569
 Bonaldi A. et al., 2021, *MNRAS*, 500, 3821
 Bonaldi A. et al., 2025, *MNRAS*, 543, 1092
 Bondi M. et al., 2024, *A&A*, 683, A179
 Brand K., Grobler T. L., Kleynhans W., Vaccari M., Prescott M., Becker B., 2023, *MNRAS*, 522, 292
 Braun R., Bonaldi A., Bourke T., Keane E., Wagg J., 2019, preprint (arXiv:1912.12699)
 Carion N., Massa F., Synnaeve G., Usunier N., Kirillov A., Zagoruyko S., 2020, in Vedaldi A., Bischof H., Brox T., Frahm J.-M. eds, *Computer Vision—ECCV 2020*. Springer International Publishing, Cham, p. 213
 Chen G., Zhang H.-X., Kong X., Lin Z., Liang Z., Chen X., Chen Z., Song Z., 2020, *ApJ*, 895, 146
 Chen G., Zhang H.-X., Kong X., Lin Z., Liang Z., Chen Z., Tang Y., Chen X., 2022, *ApJ*, 934, L35
 Chen G., Bendo G. J., Fuller G. A., Henkel C., Kong X., 2023, *MNRAS*, 525, 3645
 Chen G., Bendo G. J., Fuller G. A., Zhang H.-X., Kong X., 2024, *MNRAS*, 530, 819
 Chen Q. et al., 2024, preprint (arXiv:2406.03459)
 Cochrane R. K. et al., 2023, *MNRAS*, 523, 6082
 Cornu D. et al., 2024, *A&A*, 690, A211
 Dahiya N., Prakash D., Kundu S., Kuttan S. R., Suwalka I., Ayadi M., Dubale M., Hashmi A., 2025, *Sci. Rep.*, 15, 25842
 Hale C. L. et al., 2024, *MNRAS*, 527, 6540
 Hale C. L. et al., 2025, *MNRAS*, 544, 1323
 Hancock P. J., Murphy T., Gaensler B. M., Hopkins A., Curran J. R., 2012, *MNRAS*, 422, 1812
 Hancock P. J., Trott C. M., Hurley-Walker N., 2018, *Publ. Astron. Soc. Aust.*, 35, e011
 Hardcastle M. J., Croston J. H., 2020, *New Astron. Rev.*, 88, 101539
 He K., Gkioxari G., Dollár P., Girshick R., 2017, in *Mask R-CNN, presented at the IEEE International Conference on Computer Vision (ICCV)*, p. 2980
 Horton M. A., Hardcastle M. J., Miley G. K., Tasse C., Shimwell T., 2025, *A&A*, 699, A338
 Hotan A. W. et al., 2021, *Publ. Astron. Soc. Aust.*, 38, e009

- Jarvis M., Bacon D., Blake C., Brown M., Lindsay S., Raccanelli A., Santos M., Schwarz D. J., 2015, in *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, p. 18
- Kaiser N. et al., 2010, in Stepp L. M., Gilmozzi R., Hall H. J., eds, SPIE Conf. Ser. Vol. 7733, Ground-based and Airborne Telescopes III. SPIE, Bellingham, p. 77330E
- Kondapally R. et al., 2021, *A&A*, 648, A3
- Lao B., Jaiswal S., Zhao Z., Lin L., Wang J., Sun X., Qin S. L., 2023, *Astron. Comput.*, 44, 100728
- Lastufka E. et al., 2025, *A&A*, 703, A217
- Lawrence A. et al., 2007, *MNRAS*, 379, 1599
- Lonsdale C. J. et al., 2003, *PASP*, 115, 897
- Lukic V., de Gasperin F., Brüggen M., 2019a, *Galaxies*, 8, 3
- Lukic V., Brüggen M., Mingo B., Croston J. H., Kasieczka G., Best P. N., 2019b, *MNRAS*, 487, 1729
- Magliocchetti M., Maddox S. J., Lahav O., Wall J. V., 1998, *MNRAS*, 300, 257
- Meng D., Chen X., Fan Z., Zeng G., Li H., Yuan Y., Sun L., Wang J., 2021, in Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), p. 3651
- Mingo B. et al., 2019, *MNRAS*, 488, 2701
- Mingo B. et al., 2022, *MNRAS*, 511, 3250
- Mohan N., Rafferty D., 2015, PyBDSF: Python Blob Detection and Source Finder, Astrophysics Source Code Library, record ascl:1502.007
- Morabito L. K. et al., 2025, *MNRAS*, 536, L32
- Ndung'u S., Grobler T., Wijnholds S. J., Karastoyanova D., Azzopardi G., 2023, *New Astron. Rev.*, 97, 101685
- Ndung'u S., Grobler T., Wijnholds S. J., Karastoyanova D., Azzopardi G., 2024, *MNRAS*, 530, 783
- Oquab M. et al., 2024, *Trans. Mach. Learn. Res.*
- Ren S., He K., Girshick R., Sun J., 2017, *IEEE Trans. Pattern Anal. Mach. Intell.*, 39, 1137
- Riggi S. et al., 2021, *MNRAS*, 502, 60
- Riggi S. et al., 2023, *Astron. Comput.*, 42, 100682
- Riggi S. et al., 2024, *Publ. Astron. Soc. Aust.*, 41, e085
- Robinson I., Robicheaux P., Popov M., Ramanan D., Peri N., 2026, in International Conference on Learning Representations (ICLR)
- Rudnick L., 2021, *Galaxies*, 9, 85
- Sabater J. et al., 2021, *A&A*, 648, A2
- Samudre A., George L. T., Bansal M., Wadadekar Y., 2022, *MNRAS*, 509, 2269
- Sapkota R., Harsha Cheppally R., Sharda A., Karkee M., 2025, preprint (arXiv:2504.13099)
- Shimwell T. W. et al., 2017, *A&A*, 598, A104
- Shimwell T. W. et al., 2019, *A&A*, 622, A1
- Shimwell T. W. et al., 2022, *A&A*, 659, A1
- Shimwell T. W. et al., 2025, *A&A*, 695, A80
- Simonte M., Andernach H., Brüggen M., Miley G. K., Barthel P., 2024, *A&A*, 686, A21
- Sortino R. et al., 2023, *Exp. Astron.*, 56, 293
- Tanidis K. et al., 2025, *Publ. Astron. Soc. Aust.*, 42, e062
- Tasse C. et al., 2021, *A&A*, 648, A1
- Tingay S. J. et al., 2013, *Publ. Astron. Soc. Aust.*, 30, e007
- Vafaei Sadr A., Vos E. E., Bassett B. A., Hosenie Z., Oozeer N., Lochner M., 2019, *MNRAS*, 484, 2793
- van Haarlem M. P. et al., 2013, *A&A*, 556, A2
- Williams W. L. et al., 2019, *A&A*, 622, A2
- Wu C. et al., 2019, *MNRAS*, 482, 1211
- Zhu X., Su W., Lu L., Li B., Wang X., Dai J., 2021, in International Conference on Learning Representations (ICLR)

APPENDIX A: MASK R-CNN BASELINE:

`caesar-mrcnn`

Mask R-CNN combines object detection, classification, and instance segmentation in a two-stage framework (K. He et al. 2017).

Table A1. Main parameters of `caesar-mrcnn` used in this work.

Component	Value
<i>Architecture & optimization</i>	
Backbone	ResNet-101
Backbone strides	[4, 8, 16, 32, 64]
Optimiser	Adam
Learning rate	5×10^{-4}
Epochs	250
Batch size	2
<i>RPN/ROI configuration</i>	
RPN anchor ratios	[0.2, 0.3, 0.5, 1, 2, 3, 4, 5]
RPN anchor scales	[8, 16, 32, 64, 128]
RPN NMS threshold	0.7
Train anchors per image	256
Train ROIs per image	256
Max GT instances	100
<i>Loss weights</i>	
RPN class	1.0
RPN bbox	0.1
MRCNN class	1.0
MRCNN bbox	0.1
MRCNN mask	0.1

We adopt `caesar-mrcnn`³ as a radio-oriented Mask R-CNN baseline against the transformer-based RF-DETR. `Caesar-mrcnn` is a refactored TensorFlow 2 implementation of the system presented by S. Riggi et al. (2023), featuring an improved preprocessing pipeline and additional backbone options. In this work, we use `caesar-mrcnn` as a detection-oriented baseline: for each radio cutout, it returns an instance mask, an associated bounding box, and a class label, while detailed source characterization is deferred to downstream analysis tools.

For comparison, `caesar-mrcnn` uses the same cutouts, data splits, and augmentation policy as RF-DETR (Section 2.3), namely 132×132 pixel single-channel LOFAR images at a scale of $1.5 \text{ arcsec pixel}^{-1}$. As Mask R-CNN requires per-instance binary masks, we derive these masks via threshold-based segmentation on the ZScale-remapped images using a 3σ threshold. The segmentation is recomputed after each geometric augmentation to avoid WCS misalignment and to preserve consistency between masks, bounding boxes, and class labels.

Algorithmically, `caesar-mrcnn` follows the standard two-stage Mask R-CNN pipeline: a ResNet-FPN backbone feeds a Region Proposal Network (RPN), whose proposals are refined via ROIAlign and passed to (i) a classification and bounding-box regression head and (ii) a fully convolutional mask head producing 28×28 per-instance masks. Training optimizes the standard multi-task loss $L = L_{\text{class}} + L_{\text{box}} + L_{\text{mask}}$. Relative to a vanilla Mask R-CNN configuration, the radio-oriented implementation extends the RPN anchor scales (to ~ 8 –128 pixels) and broadens the anchor aspect-ratio priors (0.2–5.0) to better capture elongated and multi-island radio emission. In addition, the bounding-box and mask loss terms are down-weighted relative to the classification loss to balance the segmentation-dominated objective, following the recommendations of S. Riggi et al. (2023).

Preprocessing is radio-specific and handled internally by the `caesar-mrcnn` package, including single-channel FITS ingestion, NaN sanitization, and ZScale intensity remapping. We there-

³<https://github.com/SKA-INAF/caesar-mrcnn-tf2>

Table A2. Validation results for `caesar-mrcnn` on the same ELAIS-N1 validation set at $\text{IoU} \geq 0.5$ and $\text{score} \geq 0.5$.

Class	Detection			Classification		
	Completeness (%)	Reliability (%)	F1 score (%)	Recall (%)	Precision (%)	F1 score (%)
All	85.6	82.1	83.8	86.4	86.4	86.4
Compact	90.0	85.7	87.8	89.9	98.4	94.0
Single-Extended	80.4	86.5	83.3	41.1	95.7	57.5
Multi-Extended	65.2	35.6	46.1	52.2	32.4	40.0
Flagged	48.4	87.5	62.2	22.6	100.0	36.8
Spurious	18.6	22.1	20.2	12.1	88.2	21.3

fore ingest LOFAR cutouts directly as single-channel FITS images, with conversion to network tensors performed internally. While the input representation differs from that of RF-DETR, the training/validation split and augmentation strategy are kept identical across both models for consistency. We largely follow the default configuration recommended by S. Riggi et al. (2023), adopting a ResNet-101 backbone with FPN, optimization with Adam at a learning rate of 5×10^{-4} , and wide RPN anchor ratios and scales together with reduced bounding-box loss weights to accommodate both compact and extended radio sources. The full model configuration is summarized in Table A1. We evaluate `caesar-mrcnn` on the same ELAIS-N1 validation set using the identical matching criteria and thresholds adopted for RF-DETR. The resulting detection and classification metrics are summarized in Table A2, providing a useful point of reference for comparison with RF-DETR.

APPENDIX B: COMPUTATIONAL RUNTIMES

All computations were conducted on a high-performance computing cluster using a single NVIDIA Tesla V100 GPU. Training RF-DETR on the ELAIS-N1 data set for 200 epochs, using the configuration listed in Table 2, required approximately 70 h of wall-clock time on a single GPU. Inference on the full LoTSS Deep Field continuum images was performed using the tiled sliding-window strategy described in Section 3.3. For RF-DETR, wall-clock runtimes for the detection and consolidation stage ranged from approximately 0.4–1.2 h per field on a single GPU, depending on mosaic size and source density. Additional photometry and post-processing steps may extend the total end-to-end pipeline runtime.

This paper has been typeset from a $\text{T}_{\text{E}}\text{X}/\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ file prepared by the author.