

# Panoptic Segmentation of Galactic Structures in LSB Images

anonymous

CONFIDENTIAL. For review only.

## Abstract

*We explore the use of deep learning to localise galactic structures in low surface brightness (LSB) images. LSB imaging reveals many interesting structures, though these are frequently confused with galactic dust contamination, due to a strong local visual similarity. We propose a novel unified approach to multi-class segmentation of galactic structures and of extended amorphous image contaminants. Our panoptic segmentation model combines Mask R-CNN with a contaminant specialised network and utilises an adaptive preprocessing layer to better capture the subtle features of LSB images. Further, a human-in-the-loop training scheme is employed to augment ground truth labels. These different approaches are evaluated in turn, and together greatly improve the detection of both galactic structures and contaminants in LSB images.*

## 1 Introduction

Recent advancements in astrophysics imaging techniques have made possible the high resolution capture of very faint, or low surface brightness (LSB), galactic structures. Of particular importance are collisional debris (e.g. tidal tails, streams, shells...) which are essential to better understand the history of galaxy evolution. The automatic determination of the physical properties of these objects requires their accurate localisation and segmentation. This task is complicated by the large number of objects (galaxies, foreground stars...) and image contaminants (inc. dust clouds and diffraction figures around bright foreground stars) that are visible in these new LSB images. Future surveys seek to produce LSB datasets far larger than can be feasibly manually classified, as in [3, 1, 17]. Development of a method to automatically classify LSB structures is therefore crucial for the sphere of LSB galaxy research.

There have been several applications of object detection and segmentation in astronomy works. Some works focused on detecting galaxies, e.g. González et al. [6] with YOLO [14] and Burke et al. [2], Farias et al. [4], Levy et al. [10] with Mask R-CNN [7]. Galaxy morphologies were analysed with various levels of detail, from simple classification of morphology types (elliptical, spiral, irregular) [6, 4], to classifying the presence of collisional debris [21, 16].

Few works study galaxies in LSB images with deep

learning, though their precision tends to suffer from image contamination. In [16] cirrus dust clouds impact the identification of tidal structures, while in [10] identification of LSB galaxies is confused by tidal structures and cirrus. In [19] LSB artefacts such as ghosted halos and scattered light are detected with poor precision. In Tanoglidis et al. [18], Walmsley et al. [21] galaxies and tidal structures (respectively) are reliably classified, but cirrus contamination appears weak in both studies. Segmentation of cirrus clouds has been attempted with an ensemble of attention networks [15], though foreground objects are not a focus in the study.

Building on these previous experiences, we account for image contaminants when analysing galaxy morphology in LSB images, through combining the two detection tasks into a unified approach where information is shared. To the best of our knowledge, unification of segmentation of galactic structures and contaminants has not yet been performed.

Typically, multi-class segmentation is divided into two distinct tasks requiring different methods: instance segmentation deals with foreground objects, suitable for galaxy and tidal structures, whereas semantic segmentation handles background amorphous regions suitable for cirrus clouds. Kirillov et al. [9] propose a novel task, panoptic segmentation, that unifies both segmentation tasks. A baseline for panoptic segmentation is proposed in [8], where a semantic model, FCN [12], is added into Mask R-CNN so that instance and semantic networks share a common feature generating backbone. We propose a panoptic model to identify foreground objects and background contamination in LSB images. Following the work of [9] we extend Mask R-CNN with a purpose-designed cirrus segmentation network (Section 3.1), and further adapt the model to the dynamic range and contrast of LSB images through an adaptive intensity scaling layer (Section 3.2). A training scheme is proposed to mitigate the small availability of LSB data and annotations (Sections 3.3 and 3.4).

## 2 Data

Our dataset contains 186 MATLAS [3] LSB images of average spatial size  $6000\text{px}^2$  with two spectral channels. Respectively, 80% and 20% of samples are used for training and testing. Each MATLAS image targets a galaxy of interest: we take a  $3000\text{px}^2$  crop around the target galaxy of each image and then downsize to

1024px<sup>2</sup>. Data augmentation is applied as a combination of random flips and 90° rotations, followed by element-wise Gaussian noise with  $\sigma = 0.1$ .

Images were annotated for various galactic structures and image contaminants [17]. We retain five classes: "galaxy", combining main galaxies and companions; "elongated tidal structures"; "diffuse halo"; "ghosted halo", a contaminating diffraction effect from foreground stars; and "cirrus", obstructing dust clouds. Elongated tidal structures include tidal tails, plumes and streams, which appear visually similar and are defined as a propulsion of stellar material from a galaxy. We discard image contaminants other than ghosted halos and cirrus, such as satellite trails and instrument artefacts, as they are either very rare or very difficult to correctly predict without a larger field of view.

Postprocessing of diffuse halo annotations is required to separate overlapping labels. A Euclidean distance transform is applied to the mask to obtain local maximum peaks, representing centres of overlapping shapes. Based on the number of galaxies present, the most separated peaks are chosen. These peaks are used as markers for the watershed algorithm [20] to identify the boundary separating each overlapping label. Finally, an ellipse is fit optimally to each separated part.

During the annotation of MATLAS images [17], astronomers labelled target subsections containing objects of interest - galaxies with tidal structures. Often, in MATLAS images, more galaxies lie outside these subsections which are not annotated. On average, annotators processed 1.7 (std 0.9) galaxies per image.

## 3 Method

### 3.1 Panoptic segmentation model

In this work, we wish to segment cirrus contamination along with localised objects, allowing these two tasks to support each other. Mask R-CNN [7] is designed to handle the latter category; segmentation of extended amorphous regions such as cirrus contamination fits poorly into this instance segmentation framework. Handling categories of objects that cannot be divided into discrete entities is typically handled by networks of different design, such as FCN [12] or attention networks [5, 15]. Thus, there is a strong motivation to extend Mask R-CNN to segment cirrus in parallel to instance segmentation, in a panoptic model.

We combine the gridded Gabor attention (GGA) network proposed in [15], designed for cirrus segmentation, with Mask R-CNN, as shown in Fig. 1. GGA performs a computationally efficient multi-scale analysis thanks to a gridded approach. By design, it is sensitive to the large scale orientation of textures, which provides a key advantage for recognising cirrus clouds. This is achieved through including a Gabor modulation of features in an attention mechanism.

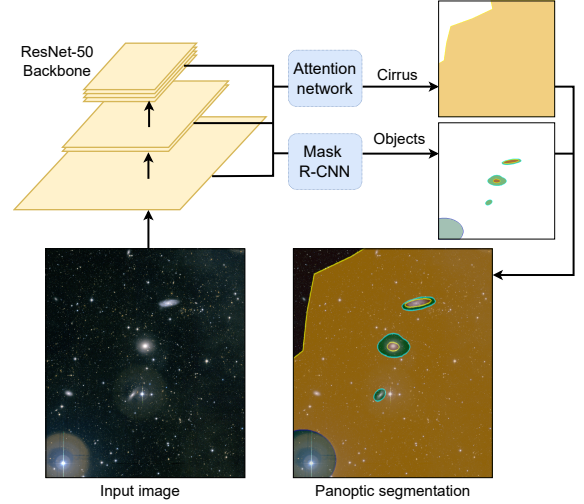


Figure 1. Proposed panoptic segmentation model.

The two networks share the same backbone ResNet-50 features, unifying segmentation of discrete objects and segmentation of galactic cirrus. Computation along each branch is performed in parallel yielding two segmentations which are overlapped in a multi-label style to achieve a segmentation of all structures in an LSB image.

**Selection of anchor sizes** - We implement Mask R-CNN with mostly default off-the-shelf parameters (see Torchvision [13]). While hyperparameters tuning could obtain minor performance improvements, this setup gave consistently good results. Thus, this computationally intensive optimisation is left for future work.

To determine reliable anchor sizes and shapes for our galactic structures, we compute histograms of the heights, widths, and aspect ratios of all target objects. The majority of objects have heights and widths between 32px and 512px, and aspect ratios between 0.25 and 2. We therefore set the RPN to consider anchor boxes of widths 32, 64, 128, 256 and 512, and for each box width three aspect ratios are considered: 0.5, 1, and 2. While there are a significant number of object bounding boxes with aspect ratios between 0.25 and 0.5, using unbalanced ratios has the possibility of introducing a bias into the model where tall but narrow objects are favoured by the RPN. This is unwanted as pose variation in localised astronomical objects is naturally balanced (ghosted halos which are artefacts also have unit aspect ratio in most cases). This setup results in 15 total anchor box sizes considered.

### 3.2 Adaptive intensity scaling layer

A challenge of LSB images comes from their large dynamic range and low contrast of LSB structures. An adaptive intensity scaling layer [15] is placed before

Table 1. Comparison of panoptic and instance detection of objects across different classes from models trained without HITL. Relative improvements of the panoptic model over the instance model are indicated as (+x%).

		Galaxy	Diffuse halo	Tidal structures	Ghosted halo	All
Panoptic	AP <sub>50</sub>	0.782 (+0.5%)	0.788 (+4.5%)	0.000 (+0%)	0.658 (+26.1%)	0.543 (+5.6%)
	AP <sub>75</sub>	0.217 (+3.3%)	0.411 (+7.0%)	0.000 (+0%)	0.658 (+34.3%)	0.330 (+21.8%)
Instance	AP <sub>50</sub>	0.778	0.754	0.000	0.522	0.514
	AP <sub>75</sub>	0.210	0.384	0.000	0.490	0.271

the backbone network to exaggerate fainter structures. It implements arcsinh scaling, popular in astronomy:  $X_s = \text{arcsinh}(aX + b)$  with  $a$  and  $b$  being learnable. We implement 2 layers in parallel, one per image channel, and concatenate their results with the original image to produce a 4-channel input to the ResNet-50 model.

### 3.3 Transfer learning

To further improve generalisation of the trained model and reduce overfitting due to the limited sample size, transfer learning is used. Mask R-CNN is pre-trained on MS-COCO [11]. While MS-COCO is made of natural images that appear different from astronomical images, learned features can still be closely applicable on both sets, possibly due to the adaptive intensity scaling layer. Further, this practice is standard in works combining astronomy and segmentation [2, 4, 19, 15]. To account for the extra scaling channels, the third convolutional input channel is duplicated.

For the cirrus (semantic) segmentation subnetwork, we follow the same pretraining on a synthesised cirrus dataset as in [15]. Fine-tuning is performed end-to-end on the complete network.

### 3.4 Human-in-the-loop training

A consequence of the non-exhaustive annotation of MATLAS images is that, during training, the model may be penalised for false positives which actually are correct predictions of unannotated objects. To combat this, we implement a human-in-the-loop (HITL) training protocol to iteratively construct a more densely annotated dataset. After an initial training period (30 epochs), we review predictions of all images. Mask predictions of good quality are retained and combined with previous annotations. The network is then trained on the new dataset for a shorter period (5 epochs for the following four rounds, then 10 epochs for three rounds to further adapt to a more greatly modified dataset) and new predictions are reviewed, with this process being repeated 7 times. After this reviewing stage, the model is trained until a total of 200 epochs is reached.

Elongated tidal structures were annotated exhaustively in all MATLAS images, therefore they do not require a HITL protocol. Only the galaxy, diffuse halo, and ghosted halo classes are considered.

## 4 Results

We assess our panoptic model by comparing against separate Mask R-CNN and GGA baselines for the tasks of localised objects and cirrus segmentation respectively on MATLAS LSB images. To ensure a fair comparison, we use the same training protocol without HITL for both the proposed model and separate baselines. We also quantify the effect of our HITL training.

We calculate the average precision (AP) score for various IoU (intersection over union) detection thresholds. AP at an IoU threshold  $x$  is denoted as AP<sub>100 $x$</sub> , e.g. AP<sub>50</sub> (threshold 0.5). AP is calculated for each class individually and then combined by averaging.

### 4.1 Panoptic vs. instance segmentation

We compare object segmentation performance through a unified panoptic approach, and an isolated instance approach. To obtain a performance baseline, Mask R-CNN is trained and evaluated on only classes that contain localised objects from the annotated MATLAS dataset. This is similar to the method of Levy et al. [10], although they focus on galaxies and do not attempt to detect tidal structures. Both models are trained for 200 epochs (no HITL is used for this experiment).

It can be seen in Table 1 that AP scores are increased in the panoptic model for all classes except elongated tidal structures, across all IoU thresholds.

Segmentation of elongated tidal structures proves to be a very difficult task, with neither network making any positive detections of such structures on the test set at any IoU threshold. The limited number of training samples and difficult class imbalance is likely the culprit here, and a larger study would be required to discern whether or not the panoptic approach improves on the task of segmenting elongated tidal structures.

For the galaxy class, the proposed panoptic approach offers a small improvement over only instance segmentation, with AP<sub>50</sub> increasing by 0.5% and AP<sub>75</sub> by 3.3%. This difference is likely minimal for the easier detection task at IoU threshold of 0.5 as the galaxy core is a strong structure and can be delineated relatively easily even in highly contaminated areas.

The synergistic benefits of tying the tasks of contamination and object prediction is stronger for diffuse

Table 2. AP<sub>50</sub> scores for standard and HITL models, evaluated on the HITL augmented test data. Relative differences of HITL training over standard training are indicated as (+x%).

Training	Galaxy	Diffuse halo	Tidal structures	Ghosted halo	All
Standard	0.533	0.503	0.000	0.450	0.371
HITL	0.797 (+49.5%)	0.856 (+70.2%)	0.000 (+0.0%)	0.814 (+80.9%)	0.617 (+66.3%)

halos and ghosted halos, which obtain a more significant increase of AP<sub>50</sub> scores by 4.5% and 26.1%, respectively, and 7.0% and 34.3% for AP<sub>75</sub>. This concurs with the previous insight, as boundaries of such structures are impacted by cirrus contamination, and thus the panoptic approach offers a larger synergistic benefit for these objects than for galaxy cores.

Overall, the benefit of the panoptic approach, quantified through improved AP scores, increases with the IoU threshold, i.e. with increased difficulty for the detection task and required precision of the object delineation. This indicates that predicted boundaries overlap better in the panoptic approach, i.e. correct detections/classifications are of better quality.

## 4.2 Panoptic vs. cirrus segmentation

In addition to improving on instance segmentation performance, the panoptic model also scores higher on the cirrus segmentation task. In [15], the GGA model scored an IoU of 74.5% as a standalone predictor on the same benchmark dataset, and 79.0% as an ensemble predictor where 5 versions of the model were trained and their results averaged to produce the final segmentation mask (see [15]). The same GGA network (without the ensemble prediction), as part of our panoptic model, scored an IoU of 85.5%, representing a relative increase of 14.9% and 8.2% over the standalone and ensemble contamination-only models. Based on this significant increase, it would seem that instance segmentation serves as a significantly beneficial auxiliary task for cirrus contamination segmentation. Given that the two tasks are combined through sharing a feature generating backbone, it follows that the addition of more semantic classes allows the model to generate features that better discriminate between cirrus and non-cirrus pixels.

## 4.3 Human-in-the-loop training

In this experiment, we study how the HITL training protocol further improves the results of our panoptic model. In total, 914 objects, out of 1928 apparent false positives, are added into the dataset. Over the first four review stages, 67% of false positive predictions are added into the dataset. Galaxies, in particular, are reliably predicted, with an acceptance rate of 89% over these review stages. In the second half of reviewing, the acceptance rate across all classes drops to 41%, likely as the annotation fields at this stage are saturated with no clear detections left to add.

After training on the extended HITL dataset for a further 120 epochs, we evaluate the model on the HITL-augmented test set. Results are significantly improved compared to the original non-HITL model, indicating that HITL training is beneficial. Indeed, the HITL model evaluated on HITL test data outperforms all panoptic scores by an average of 66.3%, demonstrating the effectiveness of the HITL training scheme to mitigate against limited data. In particular, performance on diffuse and ghosted halos sustains respectively 70.2% and 80.9% higher AP<sub>50</sub> providing a strong motivation utilising HITL to generate labels of diffuse and ghosted halos.

Cirrus segmentation was not significantly impacted by HITL training. Thus, while the addition of semantic classes improved the support of features for cirrus segmentation (Section 4.2), their refined definition from additional training did not have as large an effect.

## 5 Conclusion

In this paper, a method for automated cataloguing of galactic structures in LSB images was presented. Previous methods often suffer from a lack of unification of localised objects and large image contaminant presented as homogeneous textures. We proposed to address this issue by a panoptic segmentation model where Mask R-CNN was combined with a semantic segmentation network designed for detection of galactic cirrus. The proposed method was used to simultaneously segment galactic structures and cirrus contamination, where we showed that unification of the tasks improved their respective performance. A human-in-the-loop training protocol was utilised to create dense annotations, and significantly improved segmentation accuracy of all detected objects, especially for diffuse and ghosted halos that present more challenging boundaries and overlaps, motivating future studies on active or semi-supervised learning in astronomy.

We demonstrated that fine localisation of galactic structures through segmentation is feasible with deep learning, even in contaminated images. Galaxies and their surrounding diffuse halos, and ghosted halo contaminants were detected reliably, with reasonable delineation boundaries in areas of high cirrus. Predictions of elongated tidal structures remains a challenge. Given the good performance on other classes, further investigation is warranted. We hypothesise that with more training data, automated detection and segmentation of these subtle structures should be possible.

## References

- [1] Michal Bílek, Pierre-Alain Duc, Jean-Charles Cuillandre, Stephen Gwyn, Michele Cappellari, David V Bekaert, Paolo Bonfini, Theodoros Bitsakis, Sanjaya Paudel, Davor Krajnović, and others. Census and classification of low-surface-brightness structures in nearby early-type galaxies from the MATLAS survey. *Monthly Notices of the Royal Astronomical Society*, 498(2):2138–2166, 2020.
- [2] Colin J Burke, Patrick D Aleo, Yu-Ching Chen, Xin Liu, John R Peterson, Glenn H Sembroski, and Joshua Yao-Yu Lin. Deblending and classifying astronomical sources with Mask R-CNN deep learning. *Monthly Notices of the Royal Astronomical Society*, 490(3):3952–3965, 12 2019. doi: 10.1093/mnras/stz2845.
- [3] Pierre-Alain Duc, Jean-Charles Cuillandre, Emin Karabal, Michele Cappellari, Katherine Alatalo, Leo Blitz, Frédéric Bournaud, Martin Bureau, Alison F Crocker, Roger L Davies, and others. The ATLAS3D project–XXIX. The new look of early-type galaxies and surrounding fields disclosed by extremely deep optical images. *Monthly Notices of the Royal Astronomical Society*, 446(1):120–143, 2015.
- [4] Humberto Farias, Daniel Ortiz, Guillermo Damke, M Jaque Arancibia, and Mauricio Solar. Mask galaxy: Morphological segmentation of galaxies. *Astronomy and Computing*, page 100420, 2020.
- [5] Jun Fu, Jing Liu, Haijie Tian, Yong Li, Yongjun Bao, Zhiwei Fang, and Hanqing Lu. Dual attention network for scene segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3146–3154, 2019.
- [6] Roberto E González, Roberto P Munoz, and Cristian A Hernández. Galaxy detection and identification using deep learning and data augmentation. *Astronomy and computing*, 25:103–109, 2018.
- [7] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.
- [8] Alexander Kirillov, Ross Girshick, Kaiming He, and Piotr Dollár. Panoptic feature pyramid networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6399–6408, 2019.
- [9] Alexander Kirillov, Kaiming He, Ross Girshick, Carsten Rother, and Piotr Dollár. Panoptic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9404–9413, 2019.
- [10] Caleb Levy, Aleksandra Ciprijanovic, Alex Drlica-Wagner, Burçin Mutlu-Pakdil, Brian Nord, and Dimitrios Tanoglidis. Detecting Low Surface Brightness Galaxies with Mask R-CNN. Technical report, Fermi National Accelerator Lab.(FNAL), Batavia, IL (United States), 2021.
- [11] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755, 2014.
- [12] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.
- [13] Sébastien Marcel and Yann Rodriguez. Torchvision the machine-vision package of torch. In *Proceedings of the 18th ACM international conference on Multimedia*, pages 1485–1488, 2010.
- [14] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.
- [15] F Richards, X Xie, A Paiement, E Sola, and P-A Duc. Multi-scale gridded gabor attention for cirrus segmentation. In *2022 IEEE International Conference on Image Processing (ICIP)*, pages 3733–3737. IEEE, 2022.
- [16] H Domínguez Sánchez, G Martin, I Damjanov, F Buitrago, M Huertas-Company, C Bottrell, M Bernardi, JH Knapen, J Vega-Ferrero, R Hausen, et al. Identification of tidal features in deep optical galaxy images with convolutional neural networks. *Monthly Notices of the Royal Astronomical Society*, 521(3):3861–3872, 2023.
- [17] Elisabeth Sola, Pierre-Alain Duc, Felix Richards, Adeline Paiement, Mathias Urbano, Julie Klehammer, Michal Bílek, Jean-Charles Cuillandre, Stephen Gwyn, and Alan McConnachie. Characterization of low surface brightness structures in annotated deep images. *Astronomy & Astrophysics*, 662:A124, 2022.
- [18] Dimitrios Tanoglidis, Aleksandra Čiprijanović, and Alex Drlica-Wagner. DeepShadows: Separating low surface brightness galaxies from artifacts using deep learning. *Astronomy and Computing*, 35:100469, 2021.
- [19] Dimitrios Tanoglidis, Aleksandra Čiprijanović, Alex Drlica-Wagner, Brian Nord, Michael H L S Wang, Ariel Jacob Amsellem, Kathryn Downey, Sydney Jenkins, Diana Kafkes, and Zhuoqi Zhang. DeepGhostBusters: Using Mask R-CNN to Detect and Mask Ghosting and Scattered-Light Artifacts from Optical Survey Images. *arXiv preprint arXiv:2109.08246*, 2021.
- [20] Luc Vincent and Pierre Soille. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 13(06):583–598, 1991.
- [21] Mike Walmsley, Annette M N Ferguson, Robert G Mann, and Chris J Lintott. Identification of low surface brightness tidal features in galaxies using convolutional neural networks. *Monthly Notices of the Royal Astronomical Society*, 483(3):2968–2982, 2019.

## A Training parameters

The proposed model’s instance segmentation branch and the baseline Mask R-CNN model both use SGD with a learning rate of 0.01 which is halved every 25 epochs, and L2-regularisation penalty of  $5 \times 10^{-4}$ . The attention-based cirrus sub-network is trained with the Adam optimiser using a learning rate of  $10^{-3}$  which is exponentially decayed by a factor of 0.98 per epoch, and L2-regularisation penalty of  $5 \times 10^{-7}$ , similarly to [15].

## B Dataset statistics

After annotation by astronomers, the dataset contains the following number of labels:

1. Galaxy: 256 in total, per image mean=1.64, std=0.91
2. Elongated tidal structures: 66 in total, per image mean=0.42, std=0.71

3. Diffuse halo: 305 in total, per image mean=1.96, std=1.18
4. Ghosted halo: 262 in total, per image mean=1.68, std=1.49

In addition, 48 images out of 186 have cirrus annotation.

After HITL training, the numbers of labels are:

1. Galaxy: 439 in total, per image mean=2.81, std=1.64
2. Elongated tidal structures: 66 in total, per image mean=0.42, std=0.71
3. Diffuse halo: 529 in total, per image mean=3.39, std=2.06
4. Ghosted halo: 483 in total, per image mean=3.10, std=2.24