galmask: A Python package for unsupervised galaxy masking

Yash Gondhalekar 1, Rafael S. de Souza 2, and Ana L. Chies-Santos 3, 2

1 Birla Institute of Technology and Science, Pilani, Goa, India
2 Shanghai Astronomical Observatory, Chinese Academy of Sciences, 80 Nandan Rd., Shanghai 200030, China
3 Instituto de Física, Universidade Federal do Rio Grande do Sul (UFRGS), Av. Bento Gonçalves, 9500, Porto Alegre, RS, Brazil

ABSTRACT

Galaxy morphological classification is a fundamental aspect of galaxy formation and evolution studies. Various machine learning tools have been developed for automated pipeline analysis of large-scale surveys, enabling a fast search for objects of interest. However, crowded regions in the image may pose a challenge as they can lead to bias in the learning algorithm. In this Research Note, we present galmask, an open-source package for unsupervised galaxy masking to isolate the central object of interest in the image. galmask is written in Python and can be installed from PyPI via the pip command.

Keywords: Astroinformatics – Astronomy software

INTRODUCTION

A galaxy’s morphology encodes valuable information about the underlying physical processes driving their formation and evolution (van der Wel 2008; Lee et al. 2013; Conselice 2014; Fang et al. 2017). It correlates with physical properties, such as environmental density, merger history, and star formation rate. The first step in deriving the morphology of a galaxy in large-scale surveys and crowded regions is to isolate it from sources in the same field. Thus, there is considerable interest in generating automated astronomical source detection and segmentation tools (Bertin & Arnouts 1996; Akhlaghi & Ichikawa 2015) scalable to large-scale surveys. Examples include Morpheus (Hausen & Robertson 2020), a deep learning approach for pixel-level classification using semantic segmentation that requires training the model with user-specified segmentation maps, and galcon (de Albernaz Ferreira & Ferrari 2018), which was designed to remove bright sources around a central galaxy by generating a non-target segmentation map and replacing the non-galaxy regions with a median background estimate. Also, Farias et al. (2020) developed an automated machine learning pipeline to perform detection, segmentation, and morphological classification of galaxies. A traditional machine learning classification scheme is usually designed to provide a single class per image. The learning method might be affected by neighboring objects such as stars and other galaxies in the same field. galmask is a general-purpose package to isolate the central object and remove unwanted detections. The only assumption is that the object of interest is placed near the center of the image.

DATA

To showcase the capabilities of our method, we apply it to the crowded field around the galaxy SDSS J095734.63+033901.7 taken with the Hyper Suprime-Cam (HSC) on the 8.2-m Subaru Telescope. We downloaded stamps in the $g$, $r$, and $i$-bands in fits format from hscMap 1. The galaxy is located in the COSMOS field and the region of the CAMIRA cluster HSCJ095728+033956 (Oguri et al. 2018) at $z \sim 0.16$. The equatorial coordinates of the galaxy are RA=09:57:34.607 and DEC=+03:39:01.9612.

Corresponding author: Rafael S. de Souza
drsouza@shao.ac.cn

Corresponding author: Yash Gondhalekar
yashgondhalekar567@gmail.com

1 https://hsc-release.mtk.nao.ac.jp/hscMap-pdr3/
RESULTS AND CONCLUSIONS

The default configuration of \texttt{galmask} employs three main steps for the galaxy masking process. Firstly, it receives the original galaxy image, and convolves it with either a user-specified kernel or a normalized 2D Gaussian kernel with full width at half maximum, FWHM = 3. Then it estimates an initial segmentation map using the \texttt{photutils} (Bradley et al. 2016) library by selecting sources above a user-specified sigma threshold level.

The second step is deblending on the segmentation map to remove outliers and delineate the region around the central galaxy. For this, we use the \texttt{deblend} method from \texttt{photutils} that uses multi-thresholding and watershed segmentation to deblend nearby or overlapping sources. We empirically found deblending unnecessary in simple cases, and hence we keep this step optional. Local peaks, i.e., local maxima, are searched in each distinct labeled region of the partially-cleaned segmentation map from previous steps using the \texttt{peak} method from the \texttt{scikit-image} library (van der Walt et al. 2014).

The third and final step employs connected-component labeling (CCL; Samet & Tamminen 1988; Dillencourt & Samet 1992) to isolate the central object. CCL is an algorithm to label each connected component in an image. Let \( S \) be a subset region of the original image. \( S \) is a connected component if, for each pixel in \( S \), there is a path to any other pixel in \( S \) consisting of pixels that only belong to \( S \). Our CCL implementation makes use of the \texttt{opencv-python} (Bradski 2000) library. We also ensure that the central galaxy is not masked during this step if the background sources are area-wise larger than the central galaxy or if they dominate the image region. Finally, we replace all non-source pixels with zero values in the final output image.

Figure 1 shows the application of \texttt{galmask} to the region of the CAMIRA cluster HSCJ095728+033956. The panels depict different steps of the analysis. The original input image is displayed on the leftmost panel. The first segmentation, using sigma clipping, is depicted in the second panel, and the final mask of the central objects via CLL appears in the third panel. Furthermore, the rightmost panel provides a visualization of the masked galaxy image. \texttt{galmask} successfully removes the background sources surrounding the central galaxy and nearby sources.

![Image of application of galmask](https://example.com/image.png)

**Figure 1.** Application of \texttt{galmask} on the crowded field around the galaxy SDSS J095734.63+033901.7 observed by Subaru/HSC. The leftmost panel shows a \texttt{gri} color composite image of the original field. The second panel shows the initial segmentation map computed by \texttt{galmask}. The third panel shows the final mask, and the fourth panel shows a color composite image of the masked galaxy. The map between the \texttt{gri} bands and RGB colors is made using an asinh stretch (see, e.g. Lupton et al. 2004).

\texttt{galmask} is available as an open-source software package on GitHub\(^2\) and Zenodo.\(^3\) The package can be installed via the \texttt{pip} command.\(^4\)

**Software:** \texttt{Astropy} (Astropy Collaboration et al. 2013, 2018), \texttt{photutils} (Bradley et al. 2016), \texttt{OpenCV} (Bradski 2000), \texttt{scikit-image} (van der Walt et al. 2014).

\(^2\) https://github.com/Yash-10/galmask/
\(^3\) https://doi.org/10.5281/zenodo.6626668
\(^4\) https://pypi.org/project/galmask/
RSS thanks the National Natural Science Foundation of China, grant E045191001. ACS thanks the Chinese Academy of Sciences (CAS) President’s International Fellowship Initiative (PIFI) through grant E085201009 and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) through grant CNPq-314301/2021-6. We thank Carolina Queiroz de Abreu Silva for useful discussions. This research made use of the Photutils and Astropy packages for detection and photometry of astronomical sources.

REFERENCES

Akhlaghi, M., & Ichikawa, T. 2015, ApJS, 220, 1, doi: 10.1088/0067-0049/220/1/1
Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33, doi: 10.1051/0004-6361/201322068
Astropy Collaboration, Price-Whelan, A. M., Sipócz, B. M., et al. 2018, AJ, 156, 123, doi: 10.3847/1538-3881/aabc4f
Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393, doi: 10.1051/aas:1996164
Bradley, L., Sipocz, B., Robitaille, T., et al. 2016, Photutils: Photometry tools, Astrophysics Source Code Library, record ascl:1609.011. http://ascl.net/1609.011
Bradski, G. 2000, Dr. Dobb's Journal of Software Tools
Conselice, C. J. 2014, ARA&A, 52, 291, doi: 10.1146/annurev-astro-081913-040037
de Albernaz Ferreira, L., & Ferrari, F. 2018, MNRAS, 473, 2701, doi: 10.1093/mnras/stx2266
Dillencourt, M. B., & Samet, H. 1992, Journal of the ACM, 39, 253
Fang, G. W., Lin, Z. S., & Kong, X. 2017, Acta Astronomica Sinica, 58, 9
Farias, H., Ortiz, D., Damke, G., Jaque Arancibia, M., & Solar, M. 2020, Astronomy and Computing, 33, 100420, doi: 10.1016/j.ascom.2020.100420
Hausen, R., & Robertson, B. E. 2020, ApJS, 248, 20, doi: 10.3847/1538-4365/ab8868
Lee, B., Giavalisco, M., Williams, C. C., et al. 2013, ApJ, 774, 47, doi: 10.1088/0004-637X/774/1/47
Lupton, R., Blanton, M. R., Fekete, G., et al. 2004, PASP, 116, 133, doi: 10.1086/382245
Oguri, M., Lin, Y.-T., Lin, S.-C., et al. 2018, PASJ, 70, S20, doi: 10.1093/pasj/psx042
Samet, H., & Tamminen, M. 1988, IEEE Transactions on Pattern Analysis and Machine Intelligence, 10, 579, doi: 10.1109/34.3918
van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., et al. 2014, PeerJ, 2, e453, doi: 10.7717/peerj.453
van der Wel, A. 2008, ApJL, 675, L13, doi: 10.1086/529432