
Masking S-PLUS

 **Natalí Soler Matubaro de Santi***
Department of Mathematical Physics
Institute of Physics
University of São Paulo
natali.santi@usp.br

August 26, 2021

ABSTRACT

The new generation of surveys has been and will be able to reach much deeper into the Universe, with better image quality, photometric accuracy, as well as much larger numbers of detected objects and that is the case of the Southern Photometric Local Universe Survey (S-PLUS). One key ingredient to use its data for many purposes (since Astronomy to Astrophysics and Cosmology) is to pre-process that data. In this way, masking some observed artifacts is a main task, to guarantee that the objects could be truthfully measured, without the interference of any bright and/or saturated star. Therefore, the objective of this term project is to provide a mask for bright stars (at least for some fields) to S-PLUS.

Keywords survey mask, computer vision, image segmentation, k-means

1 Introduction

The large scale structure of the universe could be measured studying the galaxy distribution in the recent observations [1, 2, 3, 4, 6, 7]. To do this we need multi-wavelength imaging data to measure the galaxy photometric redshifts, requiring high-precision data pre-processing, whose development and validation are driven by the increasing statistical power of the surveys and the increasing relative importance of the various systematic effects [8].

One of the systematic effects accounts for the bright stars, that saturate the CCD images, messing up measurements of nearby objects and losing the information beyond the saturation limit. The first approach to circumvent this problem usually is the use of external star catalogs to build photometric masks around the saturated stars. However, the loss in area using this method ranges between 10% and 20% [9, 10]. Another important thing to mention is that those masks need to cover those stars without masking any extended source of interest [8].

Considering all those challenges, in this work I provide a code that is able to mask almost only the saturated stars (not completely only them because I consider extremely nearby objects of the parts of those stars), that covers less than 5% of the observed area. In this way, this work is organized as follow: first I show a little about the structure of the bright stars in the section 2; in the section 3 I explain what was done, showing some results; the analysis of the results and discussions are presented in the section 4; and I close the manuscript with the conclusions, in the section 5.

2 Going deeper into the bright stars

The bright stars causes many optical (diffraction pattern, optical reflections and ghosts) and electronic effects (electrons leaking to nearby pixels, brighter-fatter effects and so on). Those effects affect directly the image generating point-spread function-shaped luminous patterns, extended luminous halos and ghosts, the vertical linear “bleed trails” and spikes. Almost all those effects could be seen in the Figure 1. Both images are *GRI* composite images, but the

*You can find more about my work in my personal page <https://natalidesanti.github.io/>.

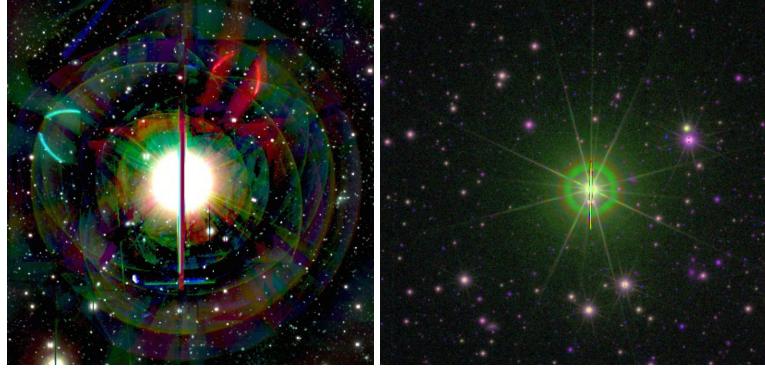


Figure 1: Example of saturated stars: on the left, $G_{Gaia} = 3.9$ in the HSC-SSP survey (from Coupon, J. et al 2017 [8]) and, on the right, one saturated star from the field STRIPE82, from S-PLUS. In the figures you can see the point-spread function, extended luminous halos and ghosts, the vertical linear “bleed trails” and the spikes.

$G_{Gaia} = 3.9$ image is processed in the way that the effects are more evident. There is a higher number of objects nearby those stars, that are behind or extremely nearby the effects. Then, the idea of the mask is to remove those stars, to saving the largest possible area of the total image.

3 Methodology and Partial Results

All the present work was done for two specific S-PLUS tiles STRIPE82-0113 and HYDRA-0167. I have used `fits` images from the bands R , G and I . All of them are images with 11000×11000 pixels.

3.1 Finding the bright stars

The first task of this project is to find the bright stars. To do that, I have used the existent masks for S-PLUS, that were made exactly with the first commented approach to build the masks: it uses the external star catalog Guide Star Catalog (GSC) and is a photometric mask around bright stars. The information for the original mask comes in tabular data for RA and Dec , indicating if the object is saturated or unreliable (i.e., nearby objects, inside the defined circle for each bright star). The visualization of the two original masks, for the two different tiles could be seen in the Figure 2.

Table 1: Original mask: area and covered area in pixels

	Area (pixel ²)	Covered percentage
STRIPE82-0113	23.504.215	26.9 %
HYDRA-0167	18.845.029	21.6 %

I have converted the object positions into an image matrix format. The process followed the steps: first I took the objects positions, converted them into pixel coordinates and made a 2D grid density; the result of this step was a matrix that I have enlarged up to the size of the image (without the borders); I filled the gaps between the pixel points, to cover the entire area of the circles; finally I have converted the pixels inside the circles with ones and the other with zeros, having the inverted mask, in order to consider the objects inside the mask. The covered area by them is presented in the table 1.

3.2 Building the new masks

The objective of this project is, using the selected region of bright stars and its surroundings of unreliable objects, provide masks that could encapsulate “just” the bright stars, i.e., masking only regions as close as possible to them, saving a considerable area of the observations. Moreover, I wanted to select only the bright stars and not all the largest objects. That is the reason because I use the bright stars positions from the previous mask as well.

Here I present two different methods:

- Only finding the objects;

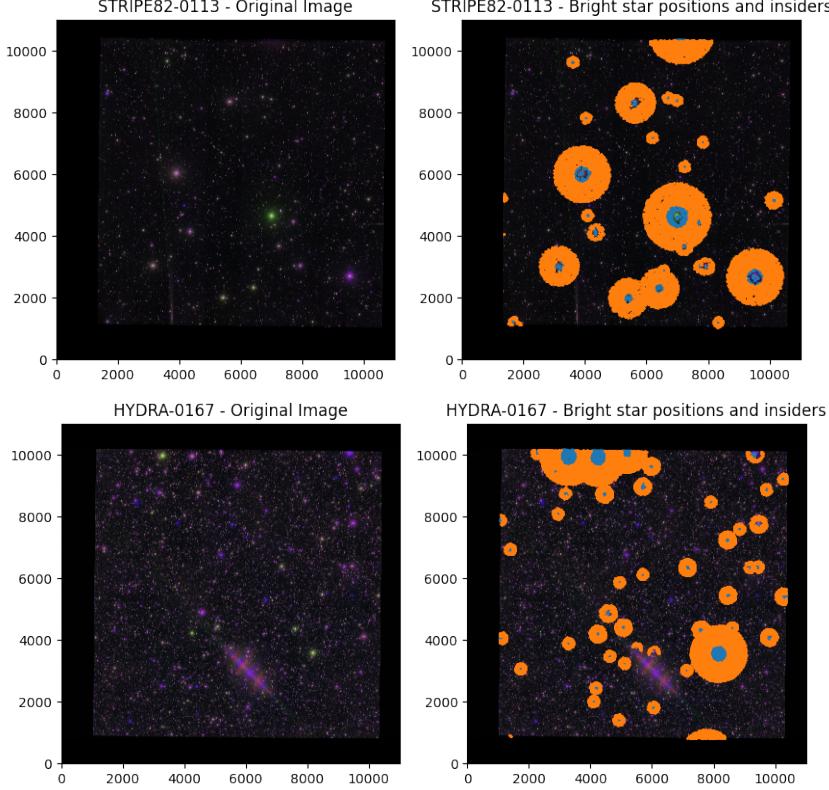


Figure 2: Original field *RGB* image (left) and the original masked positions evidenced (right). The first row is for STRIPE82-0113 and the second for HYDRA-0167. The blue points represents the bright objects and the orange points, the unreliable ones (insiders).

- Applying image segmentation and, then, finding the objects.

3.2.1 Finding the objects

The main idea of this method is to apply image processing techniques, from computer vision, for each image channel at time (R , G or B)². The first task was to label all the objects in the image. Any non-zero values in figure are counted as features and zero values are considered the background. For this purpose, it works like a clustering algorithm. Then, I found the labeled objects in rectangles containing each one of them. At this stage, the Figure 3 shows the distribution of them, for each channel and band, according to their square root of their size.

Then, I have defined a limit for the size of the objects, selecting the small and the larger ones. I removed the small objects, keeping the largest. For the largest I have selected only the ones which corresponded to the bright stars in the original mask. Thus, I have applied a binary closing image technique, in order to fill the small vacancies of the pixels in the mask and to remove the almost the whole vertical bleed trails. The final task was to do the union of all the three different bands mask, to be the final mask for the whole tile. In the Figure 4, I show a zoom around two bright stars, showing the mask for each band and the final union.

Table 2: Finding mask: area and covered area in pixels

	Area (pixel ²)	Covered percentage
STRIPE82-0113	3.383.033	3.9 %
HYDRA-0167	3.147.190	3.6 %

²I have used the Lupton algorithm [11] to convert the `fits` images to *RGB* images, using R channel as R band, G channel as G band and B channel as I band.

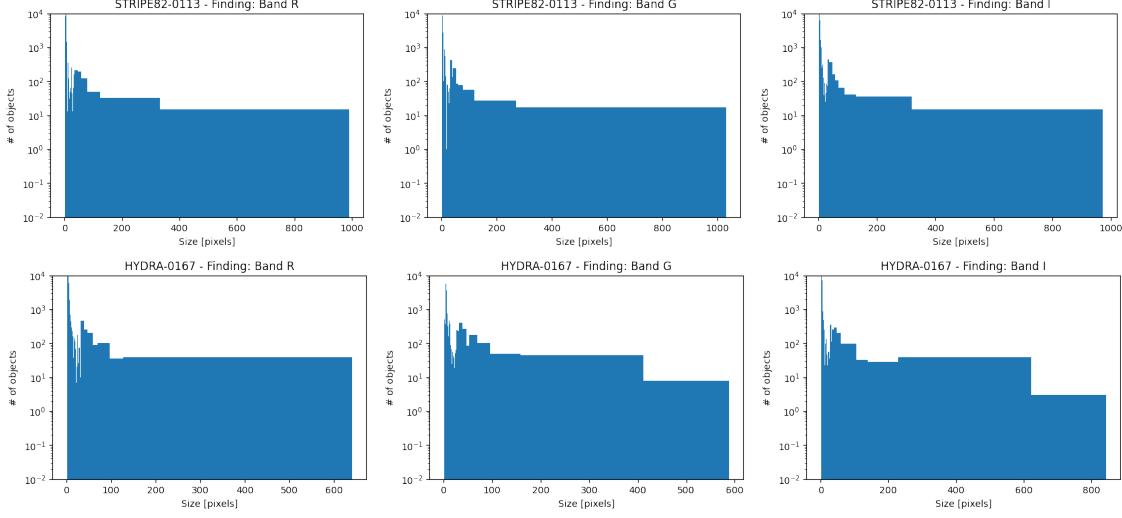


Figure 3: Finding the objects: distribution of the found objects for the three different bands (R band in the first column, G band in the second column and I band in the first column) for the two fields STRIPE82-0113 (first row) and HYDRA-0167 (second row).

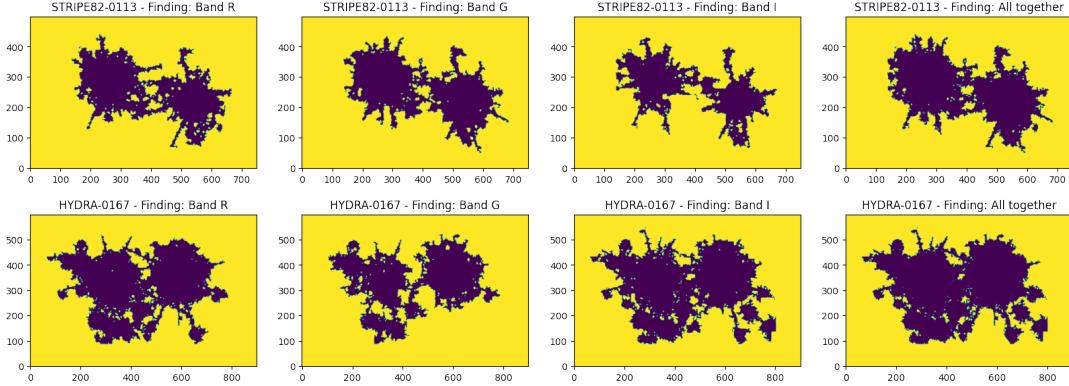


Figure 4: Finding the objects: a zoom of two different bright stars for the two different fields STRIPE82-0113 (first row) and HYDRA-0167 (second row). The first three columns represents the mask for each different band (R , G and I) and the last one is the union of all the three bands masks.

The covered area by the mask is presented in the table 2. This represents an improvement of $6.9\times$ for STRIPE82-0113 and $6.0\times$ for HYDRA-0167. And, the visual representation for the whole field could be seen in the Figure 5. Notice that all the bright stars were caught in both masks.

3.2.2 Image segmentation and finding the objects

The general idea from this second method was to use image segmentation techniques to improve the object finding. Image segmentation is the process of partitioning an image into multiple distinct regions, where each pixel has similar attributes.

The technique used here detect similarities in the regions of an image using the K-Means clustering algorithm. Here, the K-Means is used to find certain groups based on the color similarity of the image, with the number of groups represented by K . In this work, the best number of cluster was found using the elbow method. This method has the same general idea from the statistical tools like BIC or AIC, to find the best parameter as well. More specifically it measures how much the increase of K in the K-Means vary according to the intra-cluster. This quantity to be minimized is called as within-cluster sum of square (WCSS) too. I have computed the best K for just one small piece of the fields, because it is made for the original sklearn KMeans, which is not made to work with images, but

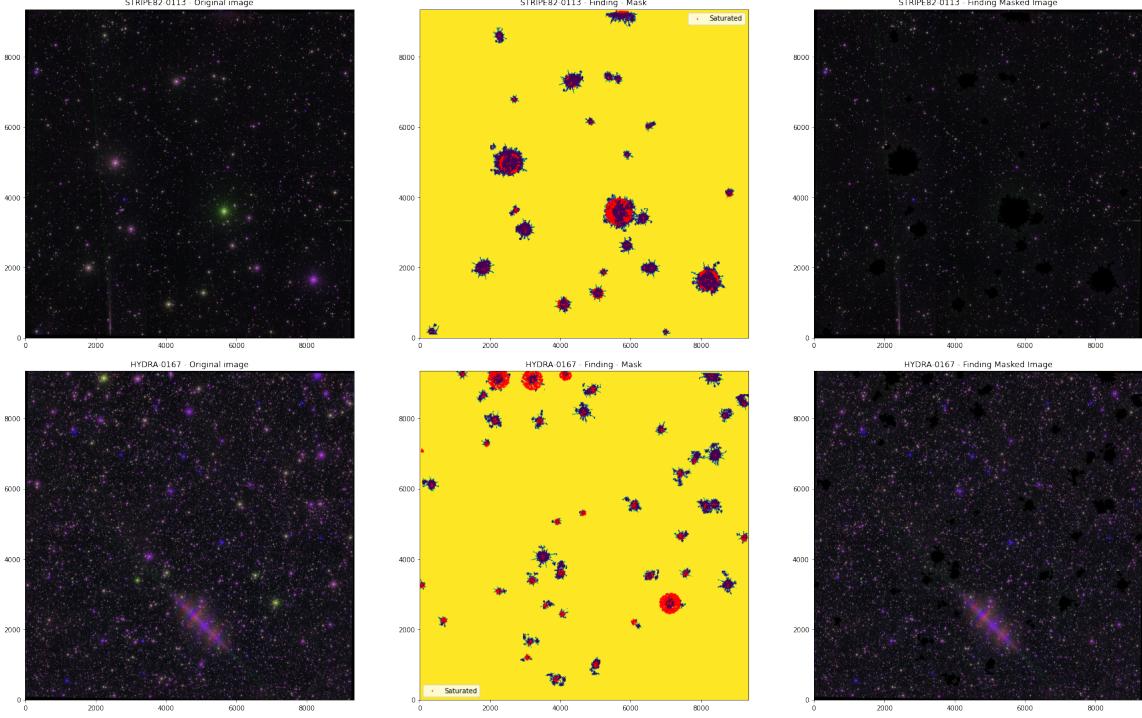


Figure 5: Finding the objects: a complete comparison of the original images, their masks (with the saturated stars in evidence) and the masked image. The two different fields STRIPE82-0113 (first row) and HYDRA-0167 (second row) are presented.

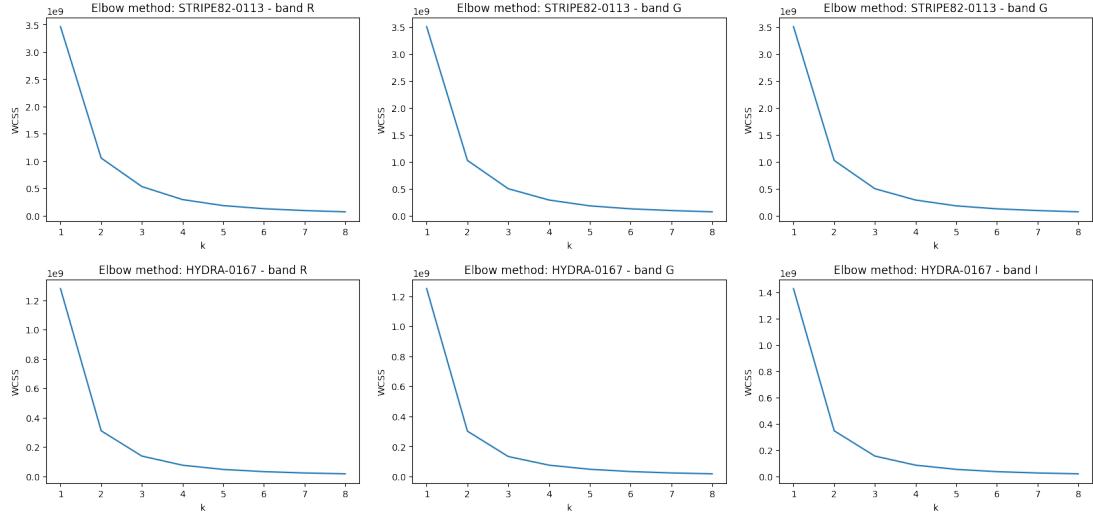


Figure 6: Elbow method: results for the three different bands (R band in the first column, G band in the second column and I band in the third column) for the two fields STRIPE82-0113 (first row) and HYDRA-0167 (second row). The elbow happens, in all the cases, for $K = 3$, but, to ensure the results, I went a little further, choosing $K = 5$.

tabular data. The results are in the Figure 6. As explained in the caption, the elbow happens, in all the cases, for $K = 3$, but, to ensure the results, I choose $K = 5$.

After choosing the best K value, I have applied the image segmentation, using kmeans, with the specific library to work the images, from cv2. The visual results are found in the Figure 7. The segmentation evidences the brightest and the darkest regions of the images and this improves the object identification from the finding objects method.

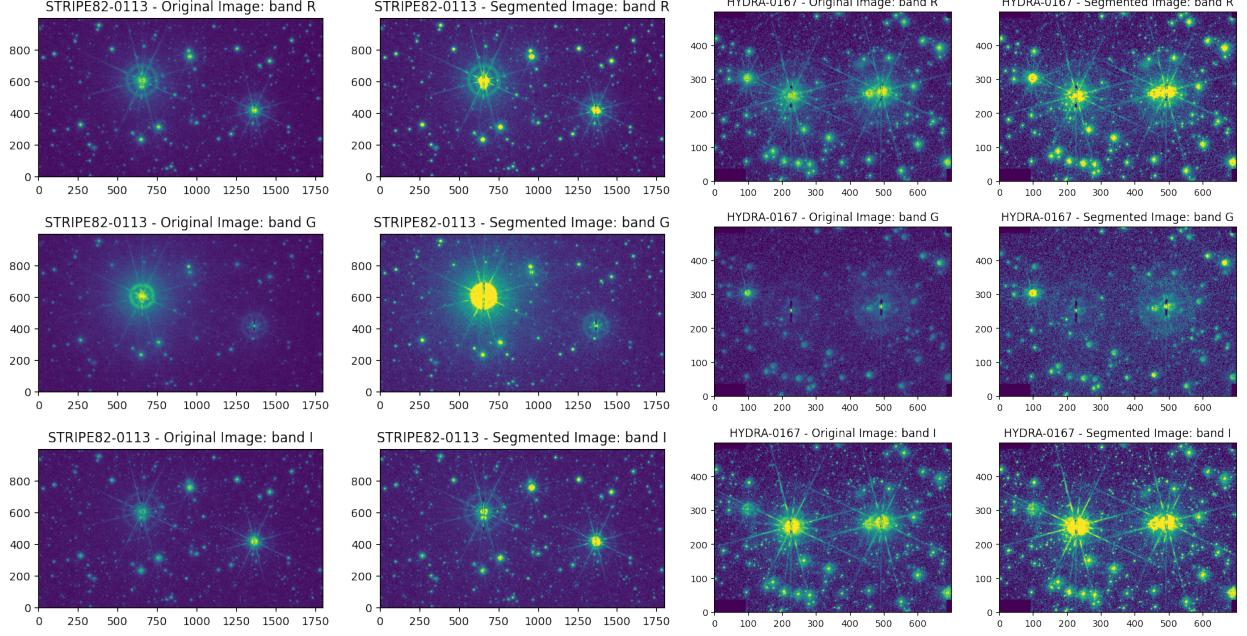


Figure 7: Comparison for a zoom in the the original image and the segmented images (using $K = 5$) for the three different bands (R band as first row, G band as second row and I band as third row) for the two fields STRIPE82-0113 (first and second columns) and HYDRA-0167 (third and fourth columns).

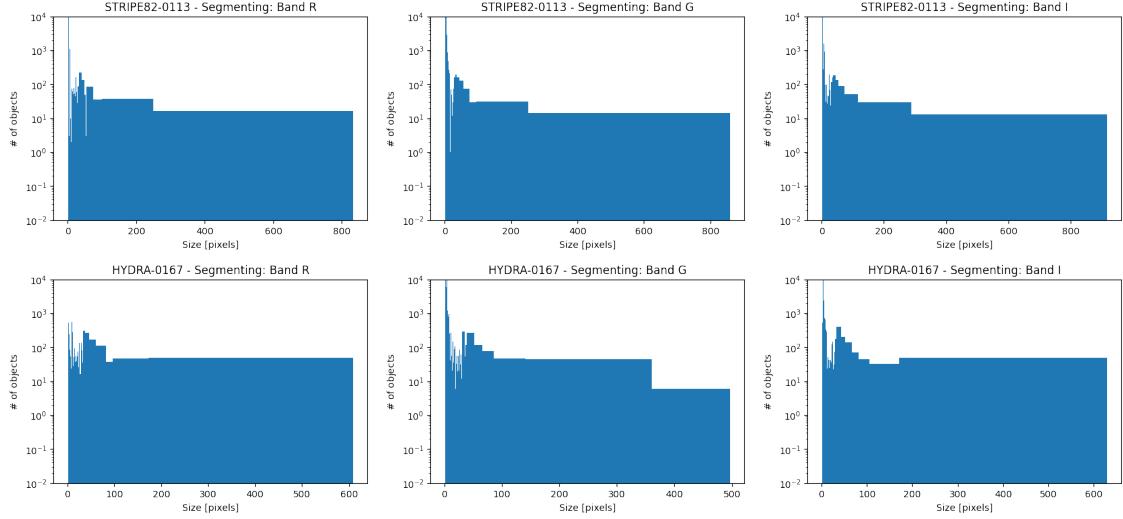


Figure 8: Segmenting the objects: distribution of the found objects for the three different bands (R band in the first column, G band in the second column and I band in the third column) for the two fields STRIPE82-0113 (first row) and HYDRA-0167 (second row).

The rest of the work done here followed the same steps from the previous method, but now finds the objects in the segmented images. Then, the size of the rectangular objects are found in the Figure 8. Comparing those results with the previous ones (see Figure 3) we can barely see a big difference, but it is evidenced in the next results.

After defining a new limit value for the size of the objects, to select the small and largest ones for this new method and performing exact the same steps described in the previous sub-subsection, I obtained the mask for each band and field. At the end I have obtained the final mask as the union of all the band masks, for each field as well. And, in the Figure 9, I show the zoom (for the same bright stars showed previously), showing the mask for each band and the final one.

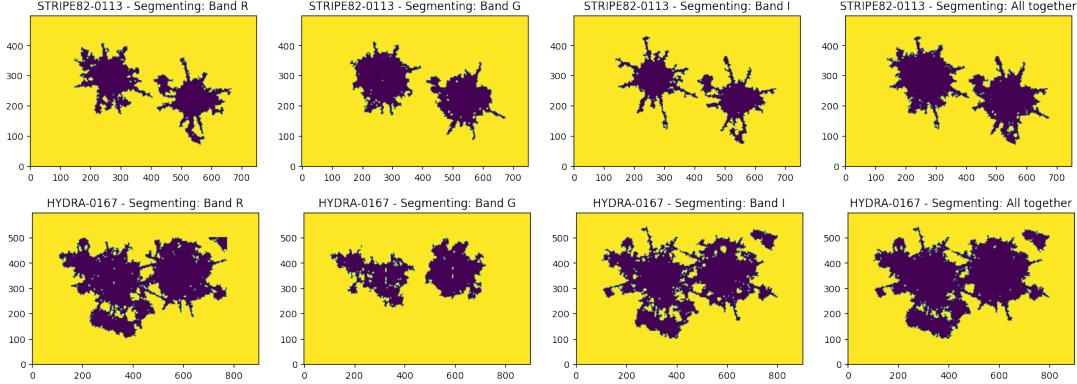


Figure 9: Segmenting the objects: a zoom of two bright stars for the two different fields STRIPE82-0113 (first row) and HYDRA-0167 (second row). The first three columns represents the mask for each different band (R , G and I) and the last one is the union of all the three bands masks.

Table 3: Segmenting mask: area and covered area in pixels

	Area (pixel 2)	Covered percentage
STRIPE82-0113	2.269.400	2.6 %
HYDRA-0167	2.427.164	2.8 %

The final result comes with the covered area by the mask, that is presented in the table 3. This represents an improvement of $10.3\times$ for STRIPE82-0113 and $7.7\times$ for HYDRA-0167. In other words, the image segmentation improves a lot the gain in area for both fields. And, the visual representation for the whole field could be seen in the Figure 10. One more time, you can see that all the bright stars where caught in both masks, covering a small area, compared to the previous method.

4 Results analysis and discussions

The main results, beyond the ones presented in the previous section are the image comparison with the original RGB image, the masked image (using the two methods) and the masked resulted image. Those results are in the Figures 11-12. Looking at those figures, you can clearly see that the segmenting method improves the masked area, decreasing and defining more the spikes and getting a low number of small objects to consider.

The last and final results were obtained running SExtractor in the original and masked images. A compilation of the number of detected objects and covered percentage (in amount of objects), by the mask, can be found in the Table 4. Notice that the number of percentage of the measured objects is quite different (is higher), compared with the covered area percentage, but all the times the segmented image is better than the finding ones. If you take a look at the magnitudes of the measured objects (see Figure 13) you can see that they are quite similar, when compared to the original image in all cases.

5 Conclusions

I was able to provide masks of zeros and ones that cover less than 5 % of the observed area and less than 15 % of the measured objects. There is a mask for each band, but the final mask is an union of all of the three band masks for the two different methods. Besides, the mask covers almost all the main image effects, like the point-spread function shaped luminous patterns, extended luminous halos and ghosts, the linear bleed trails and the spikes. Finally, the two methods presented here could be easily extended to all the fields of S-PLUS.

References

- [1] The Dark Energy Survey Collaboration. The Dark Energy Survey. In *arXiv: astro-ph/0510346*, 2005.
- [2] Ivezić, Z., Tyson, J. A., Acosta, E., et al. LSST: from Science Drivers to Reference Design and Anticipated Data Products. In *arXiv:0805.2366*, 2008.

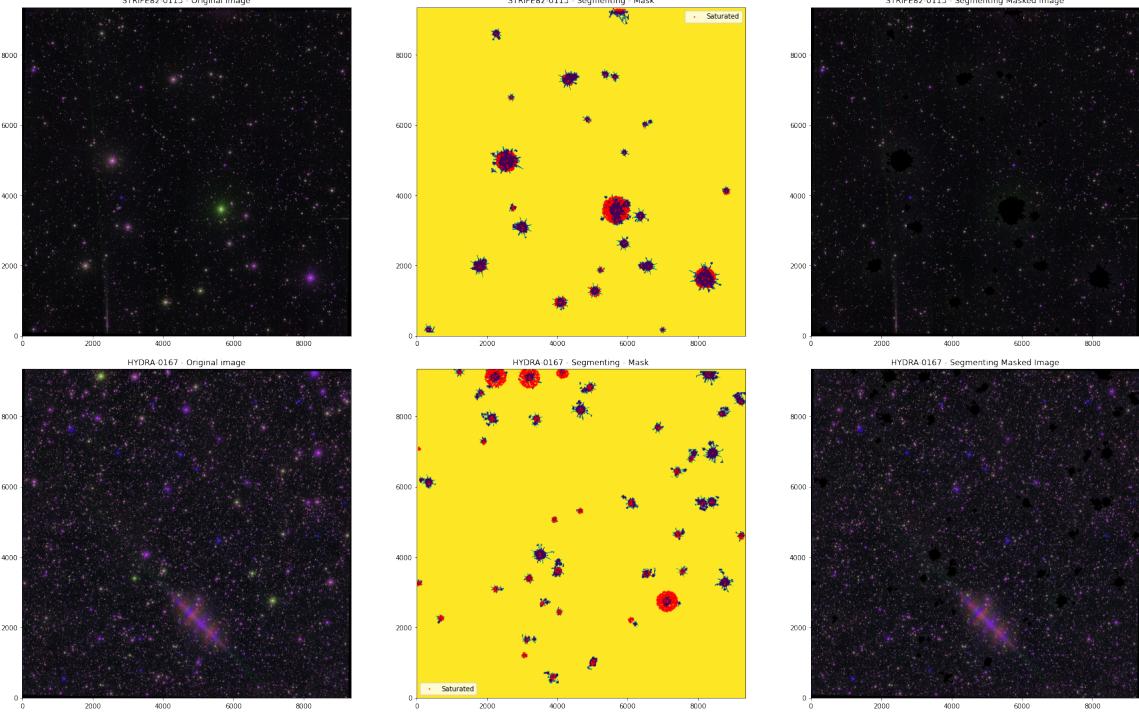


Figure 10: Segmenting the objects: a complete comparison of the original fields, their masks (with the saturated stars in evidence) and the masked image. The two different fields STRIPE82-0113 (first row) and HYDRA-0167 (second row) are presented.

Table 4: SExtractor results: number of detected objects and covered percentage number in quantity

Field	Band	Image	Number of detected objects	Covered percentage by the mask
STRIPE82-0113	R	Original	38.601	—
		Finding Masked	34.153	11.5 %
		Segmenting Masked	35.100	9.1 %
	G	Original	25.758	—
		Finding Masked	22.870	11.2 %
		Segmenting Masked	23.471	8.9 %
	I	Original	48.453	—
		Finding Masked	42.928	11.4 %
		Segmenting Masked	44.151	8.9 %
HYDRA-0167	R	Original	119.214	—
		Finding Masked	107.818	9.6 %
		Segmenting Masked	110.087	7.7 %
	G	Original	83.805	—
		Finding Masked	75.870	9.5 %
		Segmenting Masked	77325	7.7 %
	I	Original	144.522	—
		Finding Masked	130.575	9.7 %
		Segmenting Masked	133.299	7.8 %

- [3] Amendola, L., Appleby, S., Bacon, D., et al. Cosmology and fundamental physics with the Euclid satellite. In *arXiv:1206.1225*, 2012.
- [4] Ellis, R., Takada, M., Aihara, H., et al. Extragalactic Science, Cosmology and Galactic Archaeology with the Subaru Prime Focus Spectrograph (PFS). In *arXiv:1206.0737*, 2012.
- [5] Benitez, N., et al. J-PAS: The Javalambre-Physics of the Accelerated Universe Astrophysical Survey. In *arXiv*:

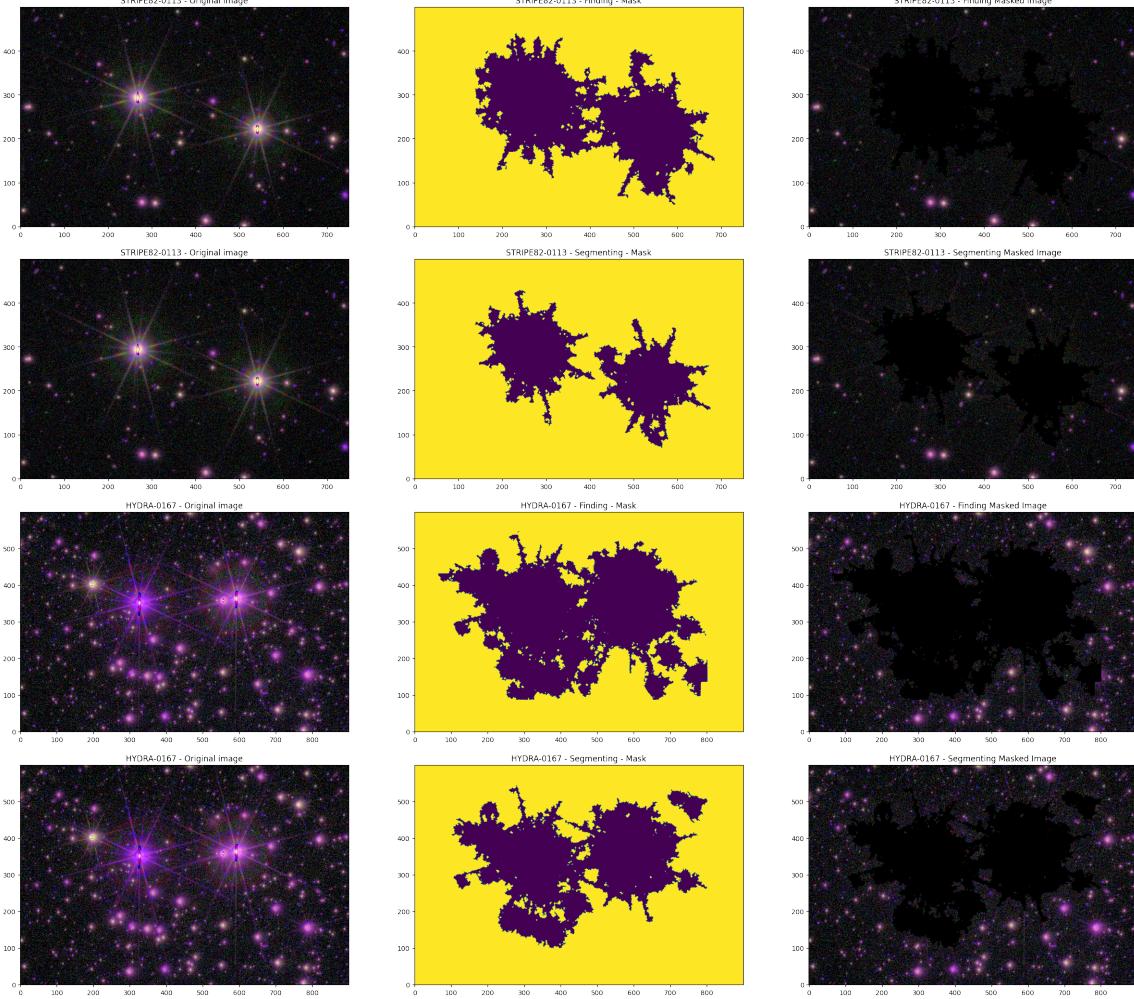


Figure 11: Final mask comparison (ZOOM 1): the original *RGB* image, the masked image (using the two methods: finding and segmenting) and the masked resulted image. The two different fields STRIPE82-0113 (first two rows) and HYDRA-0167 (last two rows) are presented.

1403.5237, 2014.

- [6] The DESI Collaboration. The DESI Experiment Part I: Science, Targeting, and Survey Design. In *arXiv:1611.00036*, 2016.
- [7] Mendes de Oliveira, C. and Ribeiro, T. and Schoenell, W. and Kanaan, A. and Overzier, R. A. and Molino, A. and Sampedro, L. and Coelho, P. and Barbosa, C. E. and Cortesi, A. and et al. The Southern Photometric Local Universe Survey (S-PLUS): improved SEDs, morphologies, and redshifts with 12 optical filters. In *arXiv:1907.01567*, 2019.
- [8] Coupon, Jean and Czakon, Nicole and Bosch, James and Komiyama, Yutaka and Medezinski, Elinor and Miyazaki, Satoshi and Oguri, Masamune. The bright-star masks for the HSC-SSP survey. In *arXiv:1705.00622*, 2017.
- [9] Coupon, J. and Ilbert, O. and Kilbinger, M. and McCracken, H. J. and Mellier, Y. and Arnouts, S. and Bertin, E. and Hudelot, P. and Schultheis, M. and Le Fèvre, O. and et al. Photometric redshifts for the CFHTLS T0004 deep and wide fields. In *arXiv:0811.3326*, 2008.
- [10] Heymans, Catherine and Van Waerbeke, Ludovic and Miller, Lance and Erben, Thomas and Hildebrandt, Hendrik and Hoekstra, Henk and Kitching, Thomas D. and Mellier, Yannick and Simon, Patrick and Bonnett, Christopher and et al. CFHTLenS: the Canada–France–Hawaii Telescope Lensing Survey. In *arXiv:1210.0032*, 2012.
- [11] Lupton, Robert and Blanton, Michael R. and Fekete, George and Hogg, David W. and O’Mullane, Wil and Szalay, Alex and Wherry, Nicholas. Preparing Red-Green-Blue Images from CCD Data. In *arXiv:0312483*, 2003.

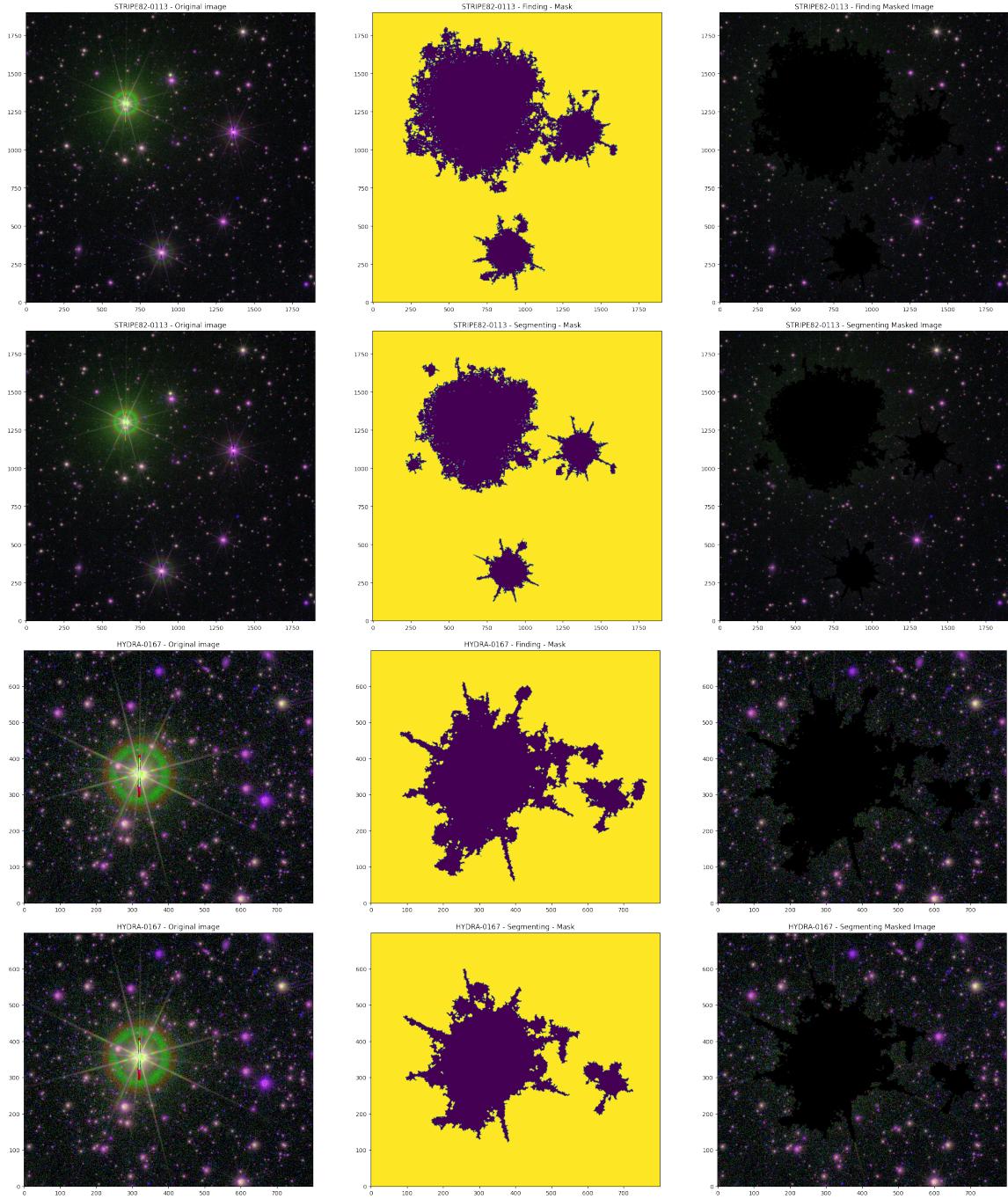


Figure 12: Final mask comparison (ZOOM 2): the original *RGB* image, the masked image (using the two methods: finding and segmenting) and the masked resulted image. The two different fields STRIPE82-0113 (first two rows) and HYDRA-0167 (last two rows) are presented.

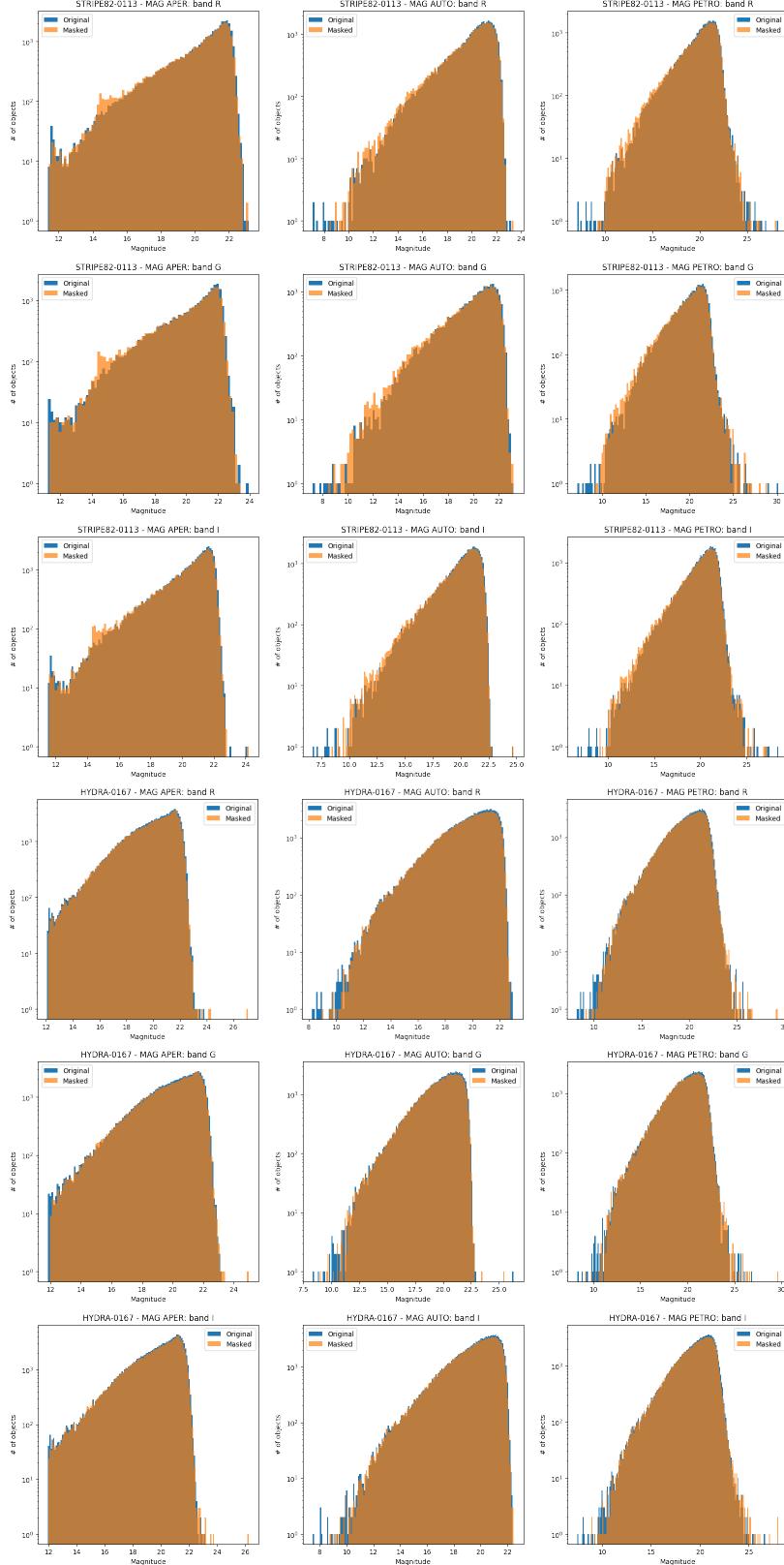


Figure 13: Final magnitudes comparison: comparison of the number of objects measured by SExtractor of the original images and the masked images (finding and segmenting). I present the results for the bands R , G and I and the magnitudes aper, auto and petro.