Investigation of Algorithms for Detecting Cosmic Rays

Report

Karlo Mrakovčić[1]
Željko Ivezić[2]

April 28, 2022

[1] *University of Rijeka, Faculty of Physics; mrakovcickarlo@gmail.com*
[2] *Department of Astronomy, University of Washington, Seattle, WA 98195; ivezic@uw.edu*

Abstract

Cosmic rays (CRs) are energetic particles that form a trail when they impact CCD. Here we are investigating the possibility to recognise the CR trails morphologically, using only one snap instead of the currently baselined two snaps designed to find CRs by image differencing. If successful, morphological detection would allow the LSST to recover up to 8% of total observing time (saving on the closing/opening shutter and readout time). Evaluating different approaches, we created an overview of the performance of currently available CR detection methods in LSST Stack and external to the Rubin Construction Project. Two major paradigms are explored: the conventional approach using Laplacian edge detection (Rubin default CR detection algorithm, and AstroScrappy) as well as the machine learning approach (Cosmic-CoNN and Unsupervised learning, called LAICA). All the algorithms except LAICA are tested on both science and difference images on the Hyper Supreme-Cam data. Our analysis showed that all of the tested algorithms performed similarly. We produced a visually verified and labeled high-purity dataset of stamps labeled as "CR" and "NOT CR" that can be used in future investigations and algorithm optimization. A good next step in this investigation would be to use dark frames to quantify the sample completeness. The results of this CR study bode well for the suggestion to abolish the two-snaps-per-visit scenario.

Contents

1 Introduction

Cosmic rays (CRs) are caused by energetic particles (electrons, protons, muons) impacting CCD array. The counts of CRs (about 1 CR cm^{-2} sec^{-1} above the atmosphere and several orders of magnitude smaller at the sea level) can be non-negligible compared to anticipated alert counts, and thus their recognition and removal is an important step in alert production[1]. Current default plan for LSST is to difference the two snaps per visit and find CRs in resulting difference image. However, when considering the shutter open / close time and readout time, the two-snaps strategy results in using 8% more observational time than if there was no intermediate readout. Therefore, this report investigates the possibility of finding CRs without relying on the availability of two snaps, with the ultimate goal of informing decision whether LSST can be executed with visits without two snaps[2]

2 Conventional Approach

We used data provided by the Hyper Supreme-Cam (HSC) mounted on the Subaru telescope. The science images in "HSC/raw/all" and templates from "HSC/runs/RC2/w 2022 04/DM-33402" were used to create the dataset. For initial investigations, Visit 11710 was chosen, for no particular reason. This visit was obtained on December 11, 2014. The visit includes 103 detectors that collected data at the same time. Each detector has 4176×2948 pixels CCD; the size of each pixel is $15\mu m \times 15\mu m$. The exposure time was $t = 300s$ and the field center is at R.A=150.18 deg. and Dec =+1.73). The entire visit contains $8.8 \cdot 10^8$ pixels, with $2.3 \cdot 10^6$ pixels marked as bad. Given the size of the pixels, the number of pixels per CCD and the exposure time, approximately 10^4 CRs per CCD would be expected above the atmosphere. Giving Subaru's altitude of 4,139 m above sea level, we expect several orders of magnitude lower CR flux.

We have not considered yet dark frame images obtained with LSSTCam at SLAC, or with ComCam at the summit.

2.1 Default CR algorithm on science images

The current CR detection algorithm is detailed in the publication Lupton *et. al.*: "SDSS Image Processing II: The Photo Pipelines". This algorithm finds CRs using their morphology: in fully sampled images, CRs look sharper than point spread function. The same algorithm is implemented in Rubin software stack and used as a part of the current Alert Production (AP) workflow.

To assess the algorithm's performance, the interpolation over recognized CRs is disabled, allowing the CRs to stay visible on the difference images. This has no effect on the ability of the CR detector to locate CRs, but it forces the DIASource detector to find CRs as DIASources. This

[1]There are two types of Rubin image processing where CRs may affect the quality of the final data products: co-added image production and analysis, and alert production. It is expected that CRs will be treated similarly to all other transient sources during co-added image production and recognized as statistical outliers. Alert production presents a harder challenge regarding CRs than co-added image production because CRs may be misinterpreted as alerts. The baseline plan assumes that two back-to-back exposures (a.k.a. snaps) will be obtained for each visit, and that all sources detected in their difference will be easily identified as CRs and subsequently masked before adding snaps into a single-visit image.

[2]Traditional morphological methods (e.g. SDSS algorithm and L.A. Cosmic) relied on the fact that CRs are typically sharper than the point-spread function (PSF). Those algorithms were quite successful when applied to thin CCDs, where CRs would typically span only a few pixels. However, with thick CCDs such as used in Rubin's LSSTCam, CRs can be quite extended worm-like structures and it is not obvious that existing algorithms can be readily applied.

is done to test for the possibility that the DIASource detector mistakes CRs for DIA sources, and to calculate the severity of the DIASource contamination by CRs. Following the creation of a database with the coordinates of all the detected DIASources, 30×30 pixel "postage stamps" (hereafter "stamps") around DIASource's center are extracted from the difference images.

2.1.1 Exposure and Mask

To evaluate the algorithm's performance, each detector's mask and exposure must be used. Mask containing information about position of the pixels are available for Template, Science, as well as Difference objects. Template CR mask marks all pixels that were part of the CR from all the visits that were used to create the template (this is a documented bug that eventually will be fixed). Science CR mask marks only pixels determined as CRs in that visit. The difference mask inherits CR mask from both the Science and the Template, and thus it is of limited use because of the above mentioned bug. Inherited Template CRs are spread because of the process that widens Template objects PSF to match Science objects. Because of that the CRs on the Difference mask that are inherited from the Template mask look like rectangles. Since Difference mask is of limited use, only Science CR mask was used to get the information about CRs on each detector. Example of the mask is shown in image 1.

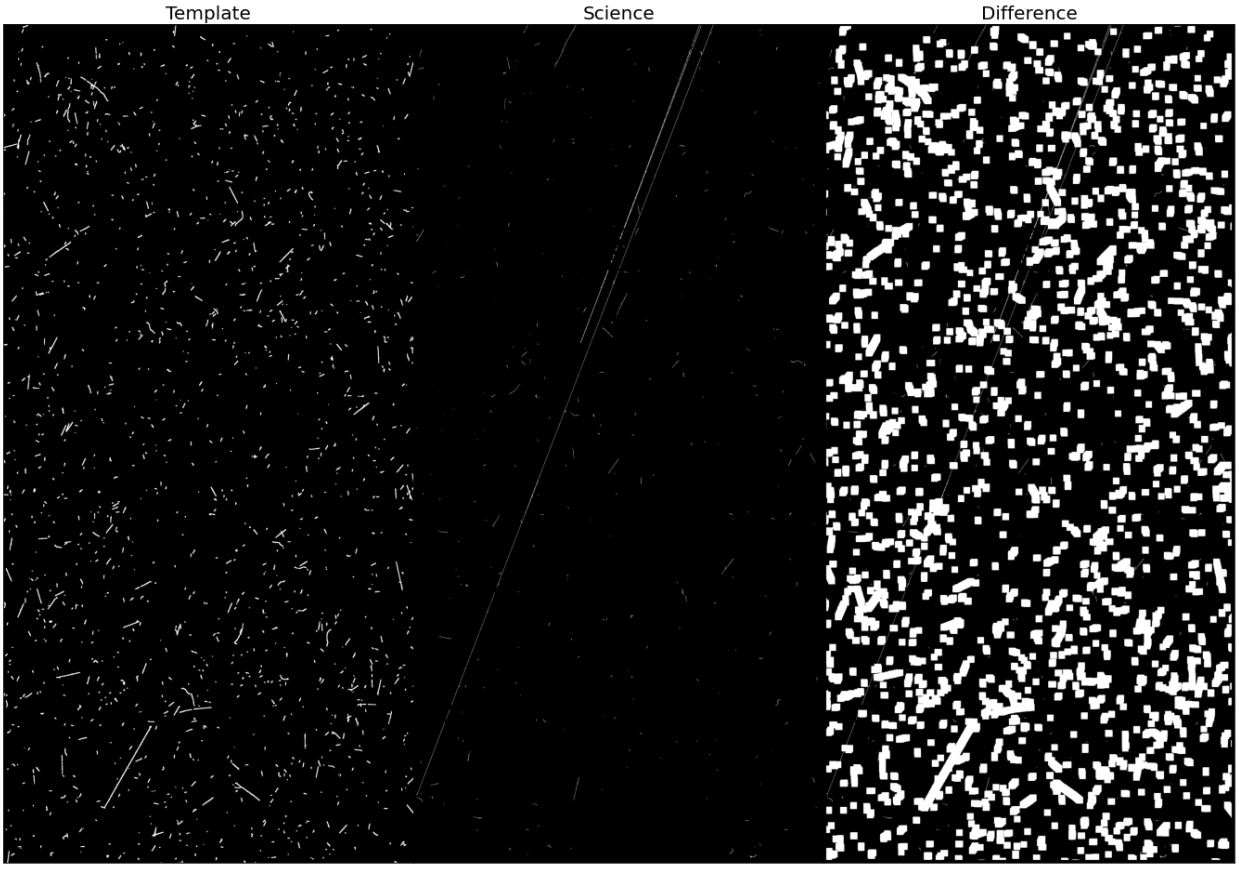


Figure 1: CR masks of the detector 42 from the visit 11710. The template mask contains all the CRs from the visits that were used to create the template. The Difference mask then inherits those CRs as well as the CRs from the science mask. Therefore, to receive valid information about the CRs in current visit, the science mask CRs must be used instead of difference mask.

There is a noticeable diagonal line detected as CR in the science image of the figure 1 and 2. Closer inspection revealed that the line was formed by a fast-moving satellite. For the current method, most satellite lines are identified as CRs. Out of the 107 detectors, 7 of them (6.8%) are contaminated with something that looks like a satellite trail.

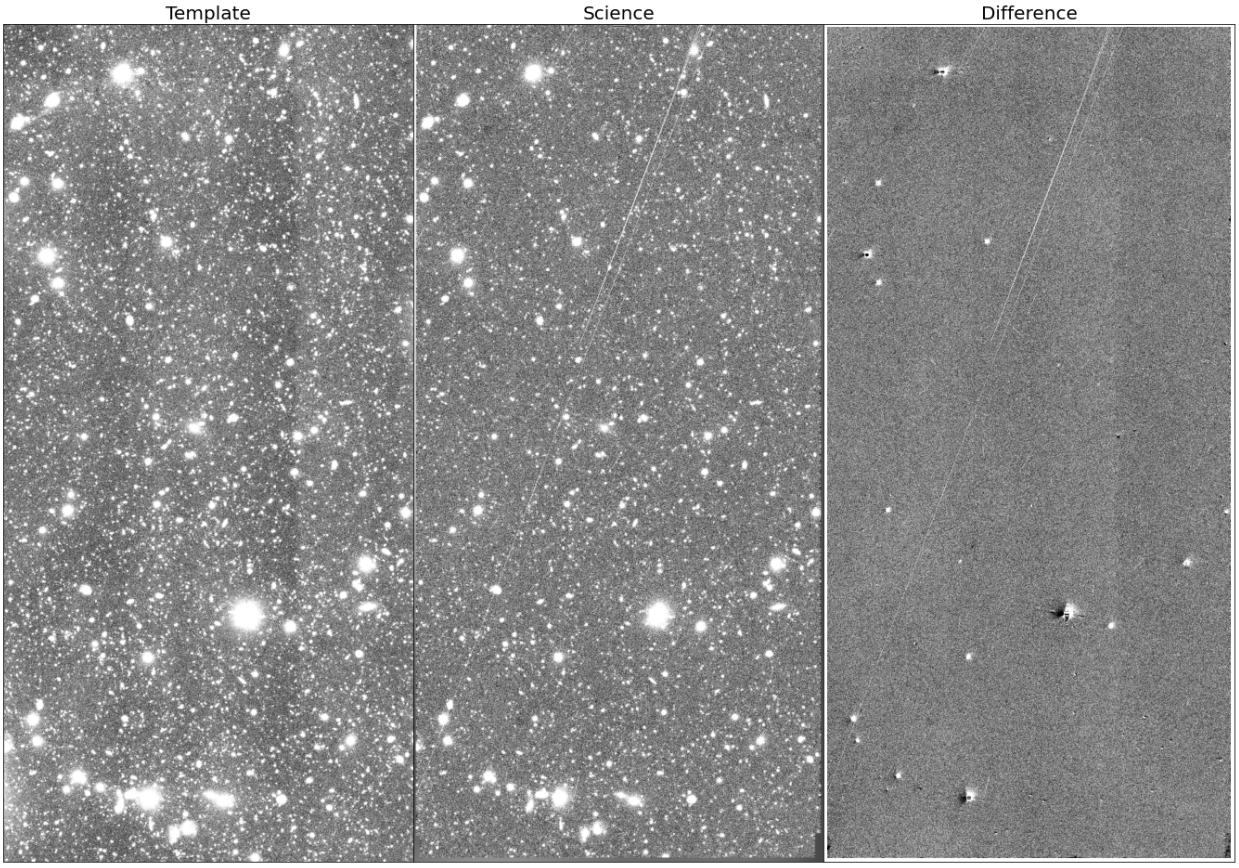


Figure 2: Representation of exposures for the detector 42 from the visit 11710. Value of each pixel is displayed using $tanh(exposure)$ stretch to enhance contrast. There are notable diagonal lines that are created by a fast moving satellite. Those lines are detected as CRs by the algorithm, as seen in figure 1.

This algorithm detected 878,807 pixels as CR pixels inside science mask. It is informative to calculate:

$$CR = \frac{N_{CRpix}}{N_{totpix}} = 0.0998\% = 997.6\, CR\, pix/M\, pix$$

where N_{CRpix} is the number of CR pixels in visit 11710 and N_{totpix} is the total number of pixels. The result is 997.6 cosmic ray pixels per megapixel, or around 0.1% of all pixels are associated with CRs.

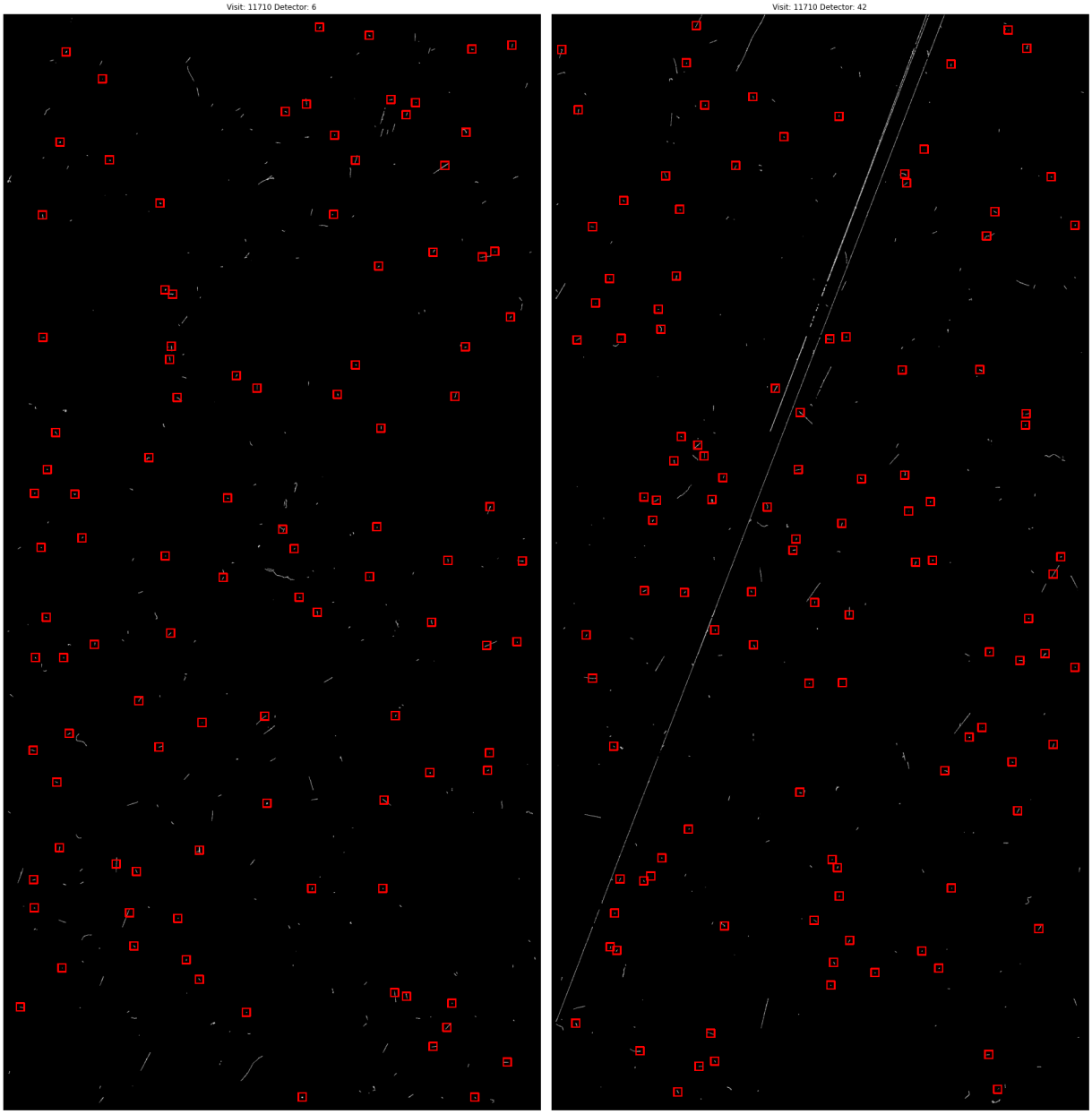
2.1.2 DIASources

LSST alerts will be produced for DIASources detected in difference images. In this dataset, some DIASources are CRs because interpolation of the CRs in Science images was turned off. This step was skipped in order to evaluate DIASource finder's ability to create DIASource from the CR signal.

DIASource detector detected 18,343 sources from the visit 11710. After each object is detected, a stamp is cut around each object so it is centered inside $30x30$ pixels image. Each stamp's CR mask is examined and 232,247 CR pixels are found on the cutouts. Rest of the CR pixels from difference images are not detected as DIASources. This number represents only 26.4% of all CR pixels. The number of CR pixels that are associated with DIASources is thus

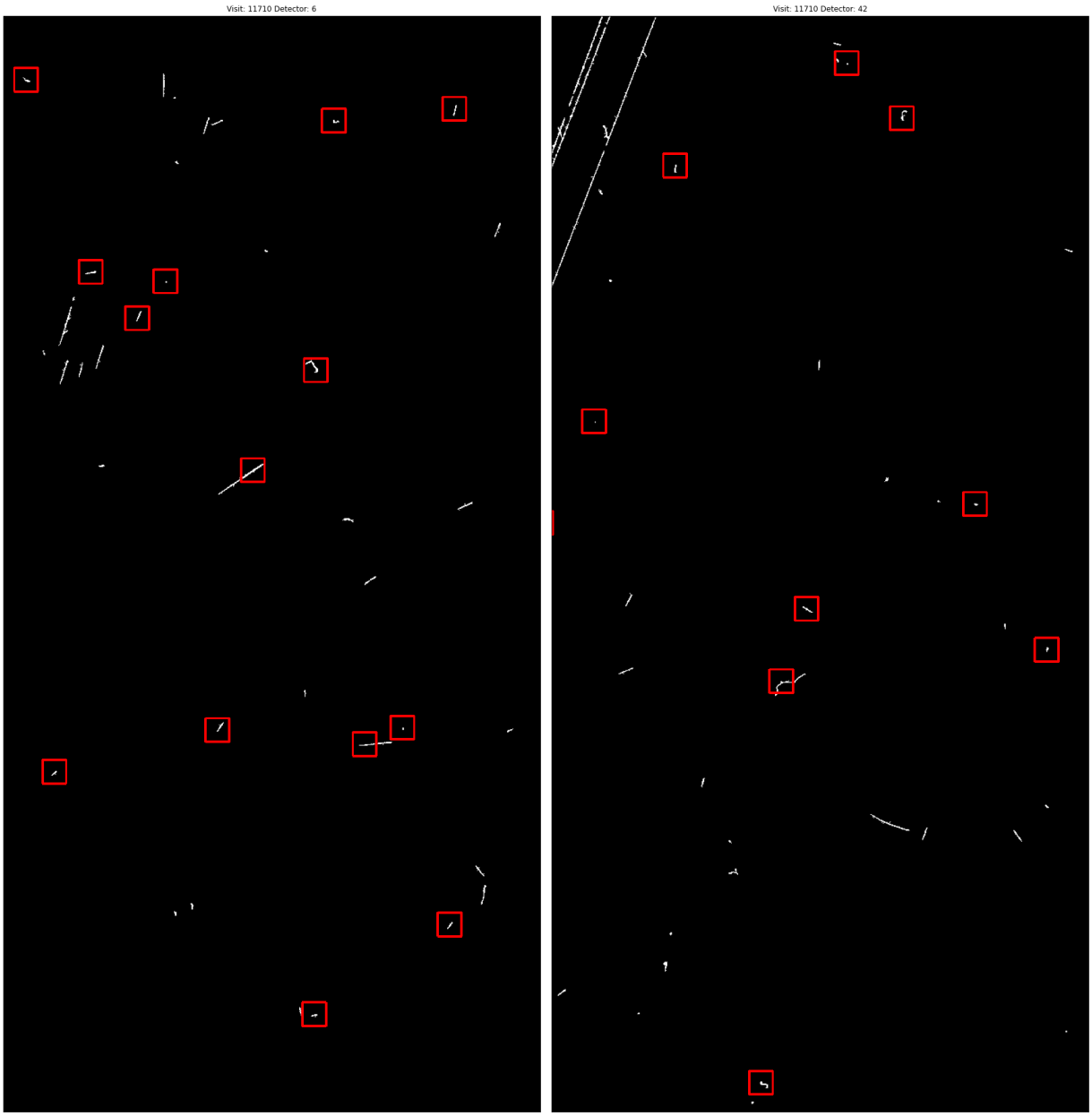
$$CR = 0.0264\% = 263.6\, CR\, pix/M\, pix$$

This is the number of potential CR pixels that could become DIASource and thus could potentially contaminate alerts. The rest of CR pixels are not associated with DIASources, and are of no consequence for alert production.



(a) Detector 6 (b) Detector 42

Figure 3: Masks showing in white color which pixels are detected as CRs by the Default CR detector on Science images. Red rectangles show stamps that are cutout around the DIASources and have one or more pixels detected as CR by the same algorithm. Out of 878,807 CR pixels, 232,247 are associated with an DIASource and are found inside red rectangles. Due to the process of PSF convolution step during the DIASource detection, great amount of the CR pixels are not detected as DIASource. This is by the design and is not a concerning issue.



(a) Detector 6

(b) Detector 42

Figure 4: Zoomed in upper right corner of the figure 3 showing in white color which pixels are detected as CRs by the Default CR detector on Science images. Red rectangles show stamps that are cutout around found DIASources and have one or more pixels detected as CR by the same algorithm. Mostly smaller CRs are detected as DIASources as well as only parts of longer ones. Please note this image contains no information about the brightness of the CR-s only the position because it's showing the CR mask.

2.1.3 Postage Stamps

Alert "postage stamp" is a cutout from the difference exposure with the shape of $30x30$ pixels around detected DIASource.

In order to investigate number of CR objects posing as DIASources, each stamp is labeled as either "CR" or "not CR". If a stamp has more CR pixels in a CR mask than the threshold L, the entire stamp is labeled "CR," otherwise "not CR."

Figure 5 shows the distribution of stamps with different number of CR pixels in CR science mask. Median value of CR pixels is 18 CR pixels per stamp. It's also clear that a significant number of CRs are made up of merely a few pixels. This is mostly expected because size of the CR in CCD is mostly due to the CR angle of incidence and energy. There is a possibility that the CR's angle of incidence will be perpendicular to the CCD, and have such low energy that just a few pixels will be affected. If we use median value of 18 CR pixels per CR as a means to evaluate CR flux we get:

$$j = \frac{N_{CR}}{t \cdot A} = 0.082 \frac{CR}{cm^2 \cdot s}$$

where N_{CR} is rough number of CRs, A is the area of the telescope and t is the time of the exposure

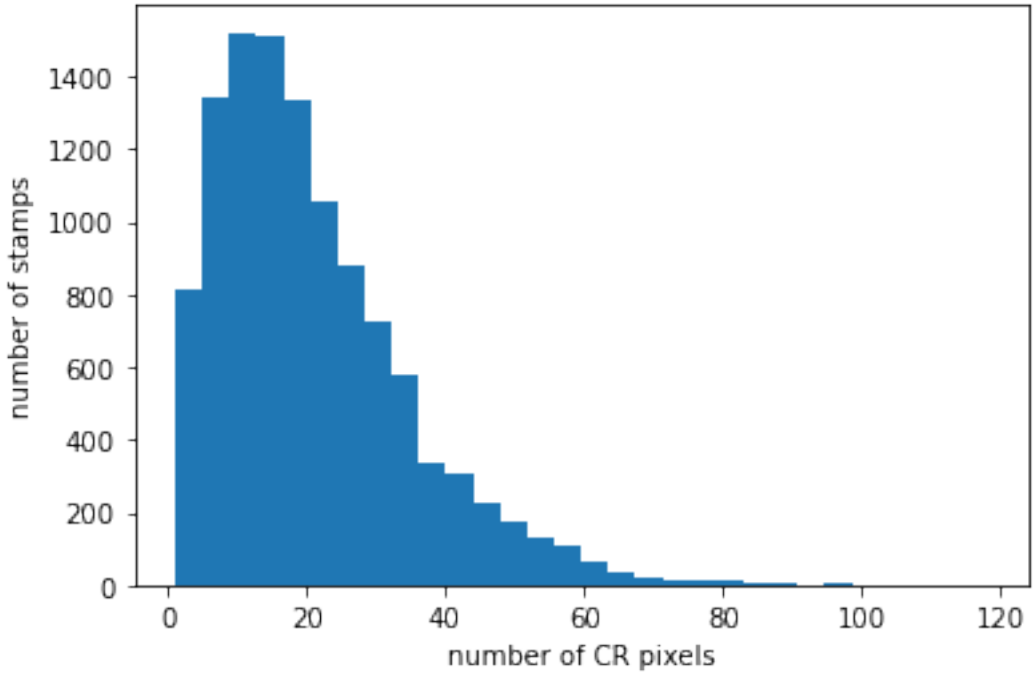
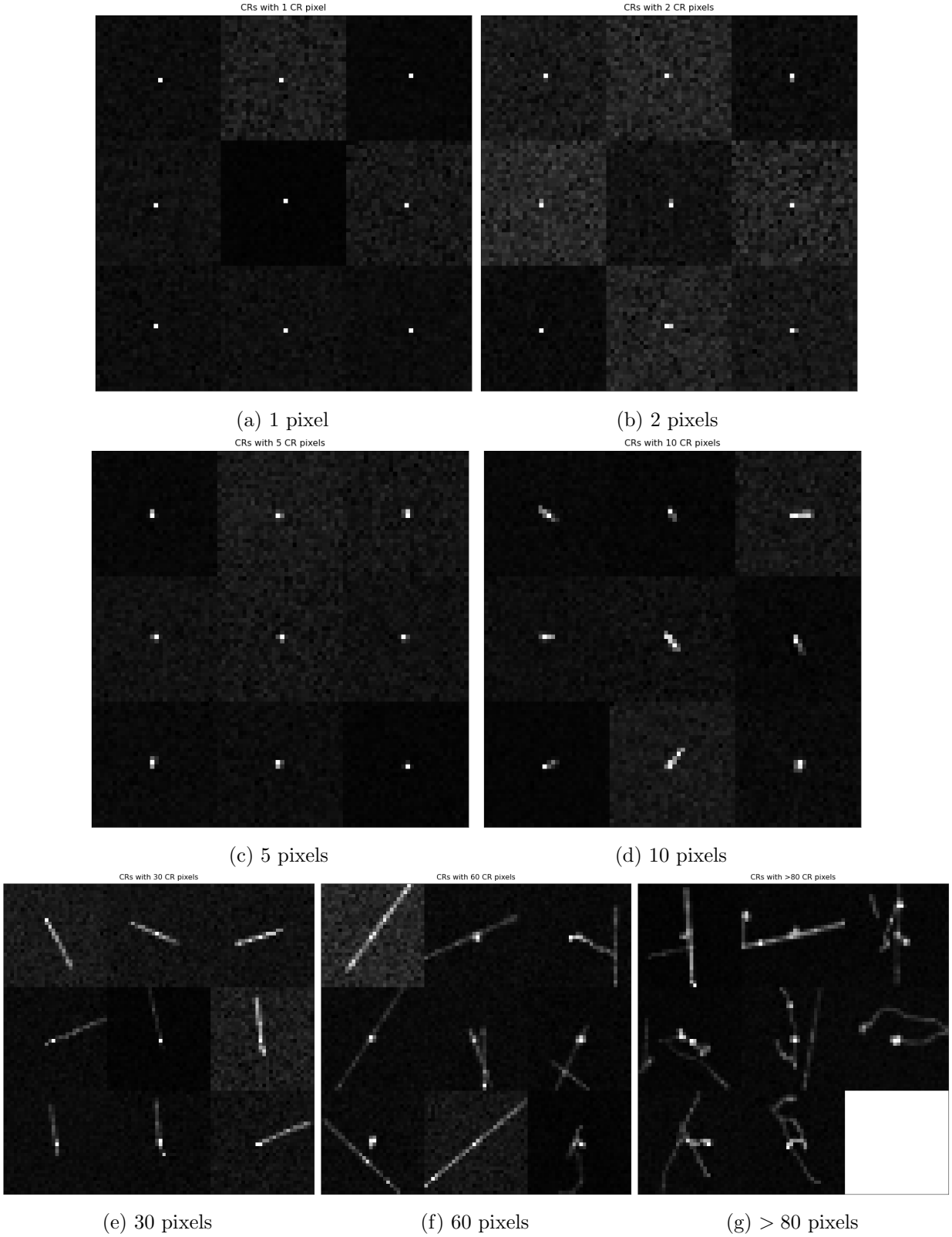


Figure 5: Histogram of the distribution of stamps with different number of CR pixels detected in the science mask created by the Default CR algorithm. Median number is 18 CR pixels per stamp. Stamps with 0 CR pixels are not taken into consideration.

Figure 6 shows stamps containing different number of CR pixels inside the science CR pixel mask. The CRs with 1, 2 and 5 CR pixels are present, and must be given the "CR" label. CRs with the count number 60 and > 80 are usual due to multiple cosmic rays are present in the same stamp. For stamps of dimension 30×30 the longest straight line is 43 pixels long. That is the the highest number of the CR pixels one stamp can have from one CR that has one pixel wide trail. Number significantly higher than that is due to either multiple CRs present in the same stamp, or the CR particle drifting inside CCD. A discussed below, $L = 1$ threshold is used for subsequent analysis.



(a) 1 pixel (b) 2 pixels

(c) 5 pixels (d) 10 pixels

(e) 30 pixels (f) 60 pixels (g) > 80 pixels

Figure 6: Examples of DIASource stamps with different numbers of CR flags in its mask created by the Default CR algorithm on the science images. Stamps are $30x30$ pixels. Observe various shapes and sizes of the CRs that algorithm detected.

To assess the optimal threshold L for the minimal number of CR pixels in the science mask for the stamp to become labeled as "CR", one needs to observe figure 7. The figure shows all stamps that contain only 1 CR pixel in their CR mask. It is obvious that most of those stamps really show cosmic rays that have such a combination of energy and incidence angle that only one pixel on the CCD is affected. Out of 82 stamps with 1 CR, 76 of them clearly show a cosmic ray in the middle of the stamp. Therefore, a threshold value $L = 1$ is adopted.

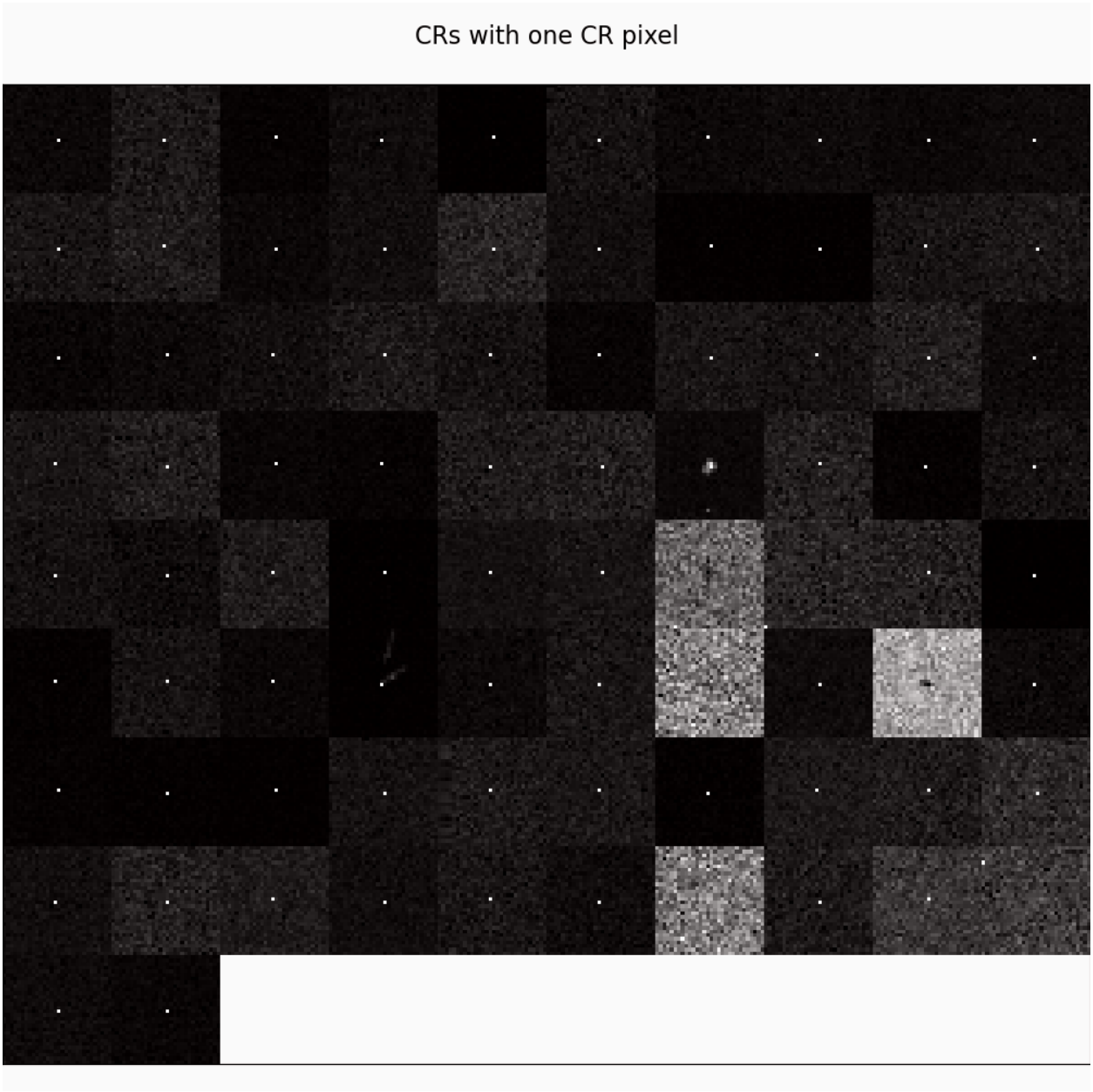


Figure 7: All of the 82 examples of stamps with 1 pixel flagged as CR for visit 11710. It is clear that most of the stamps are in fact CRs so the cutoff $L = 1$ is chosen. If $L > 1$ is chosen all of those examples would be classified as objects.

2.1.4 Labels

The following steps are done to classify postage stamps into "CR" and "not CR" subsets. To obtain a mask of CRs and the exposure of the Science image, the Default CR algorithm was first run on a science image without interpolation ("calexp"). The second step is to build a difference image by combining the exposures of the science image and the template (produced beforehand). Following the creation of the difference image, the DIASource detector searches for DIASources (including CRs) and returns a table of DIASources and their positions. Procedure to this point is the same as the standard procedure of the Alert Production, only difference is that in this case CR pixels are not interpolated. Template exposure is cut into 30×30 pixel stamps with DIASources in the middle. Then the Science CR mask is analyzed and if any CR pixels are found inside the stamp, whole stamp is labeled "CR". Otherwise, it's labeled "not CR".

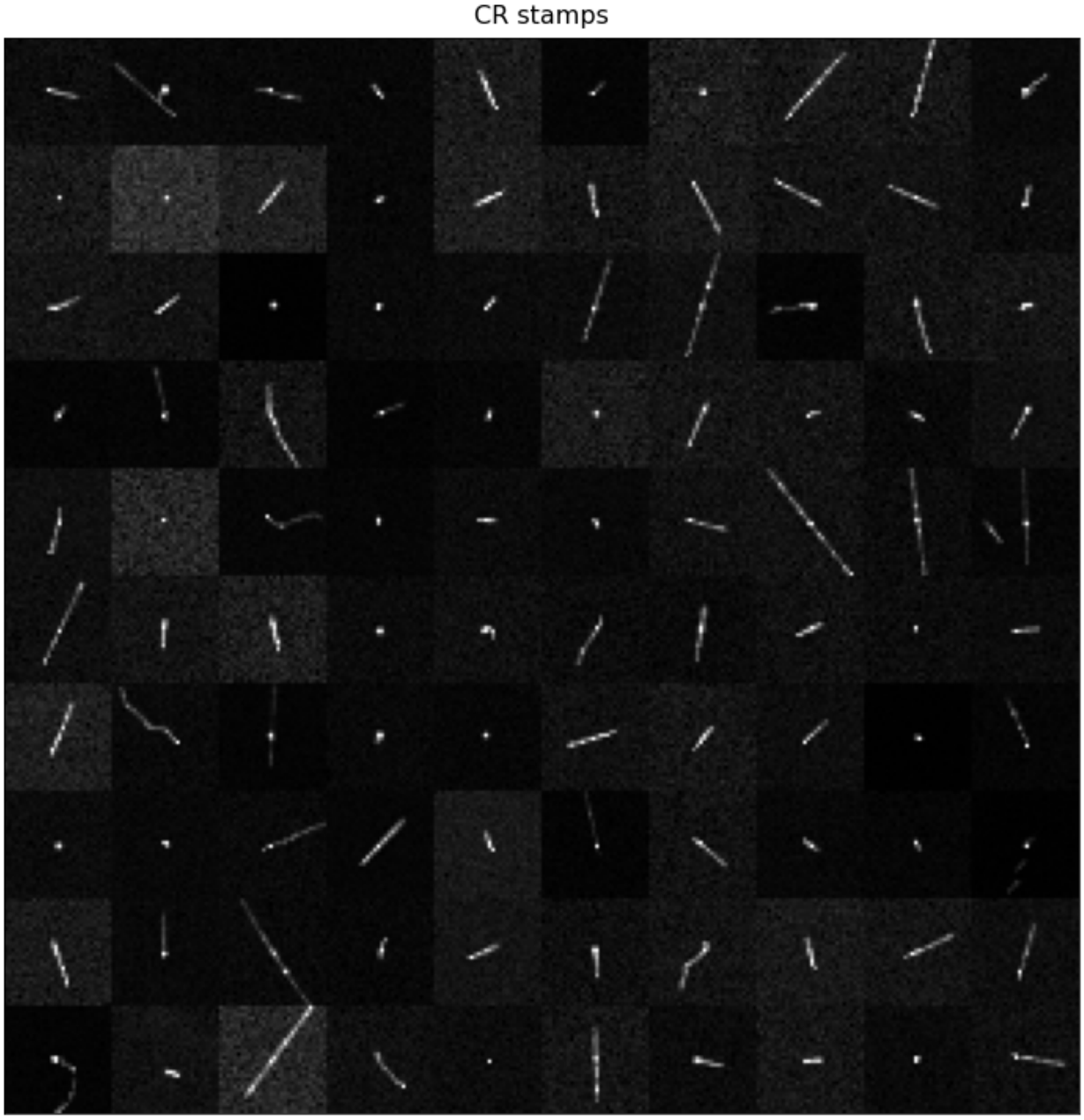


Figure 8: Random sample of 100 $30x30$ pixels stamps cut around middle of the DIASources that are labeled by the Default CR algorithm on science images as "CR".

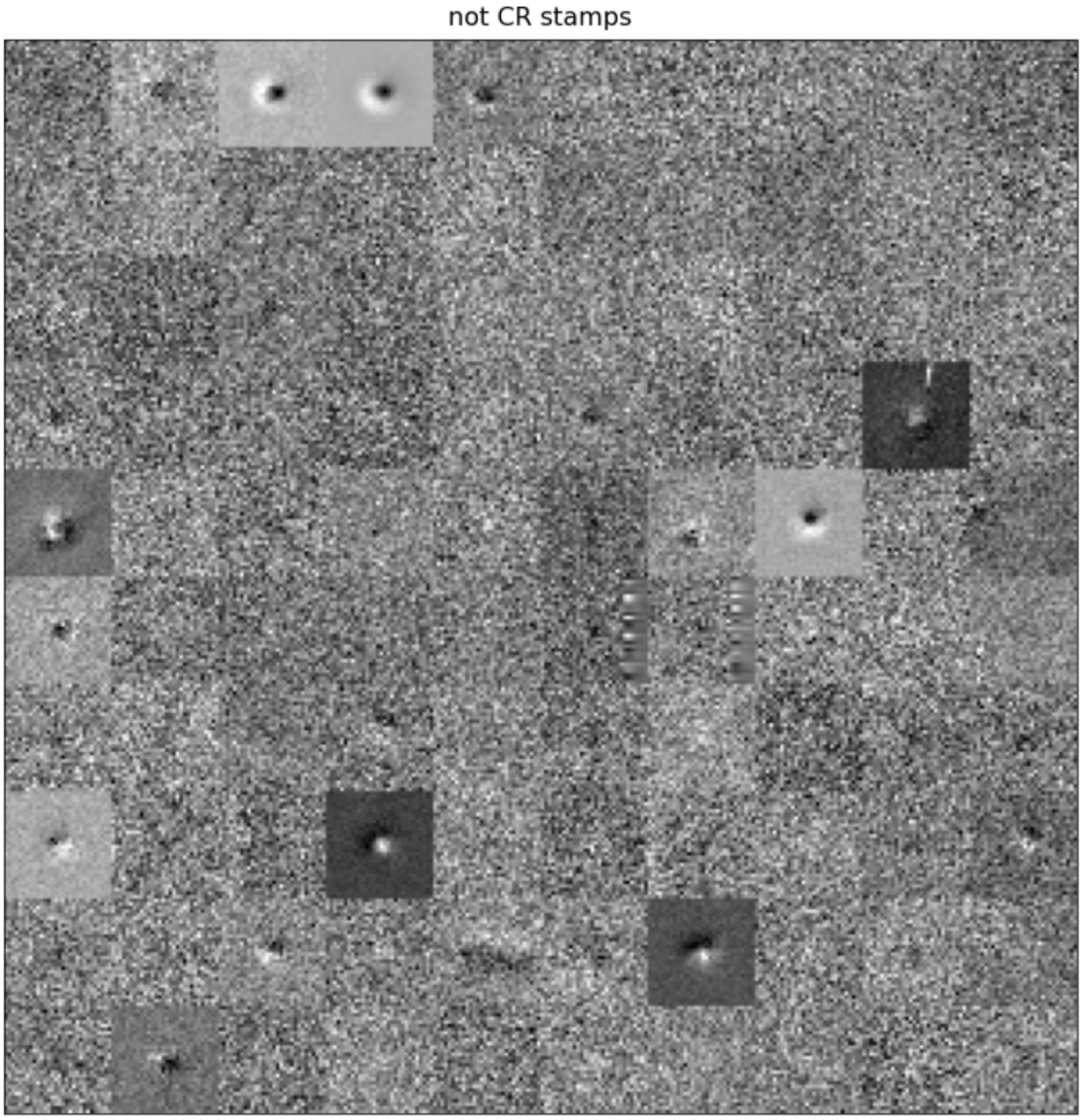


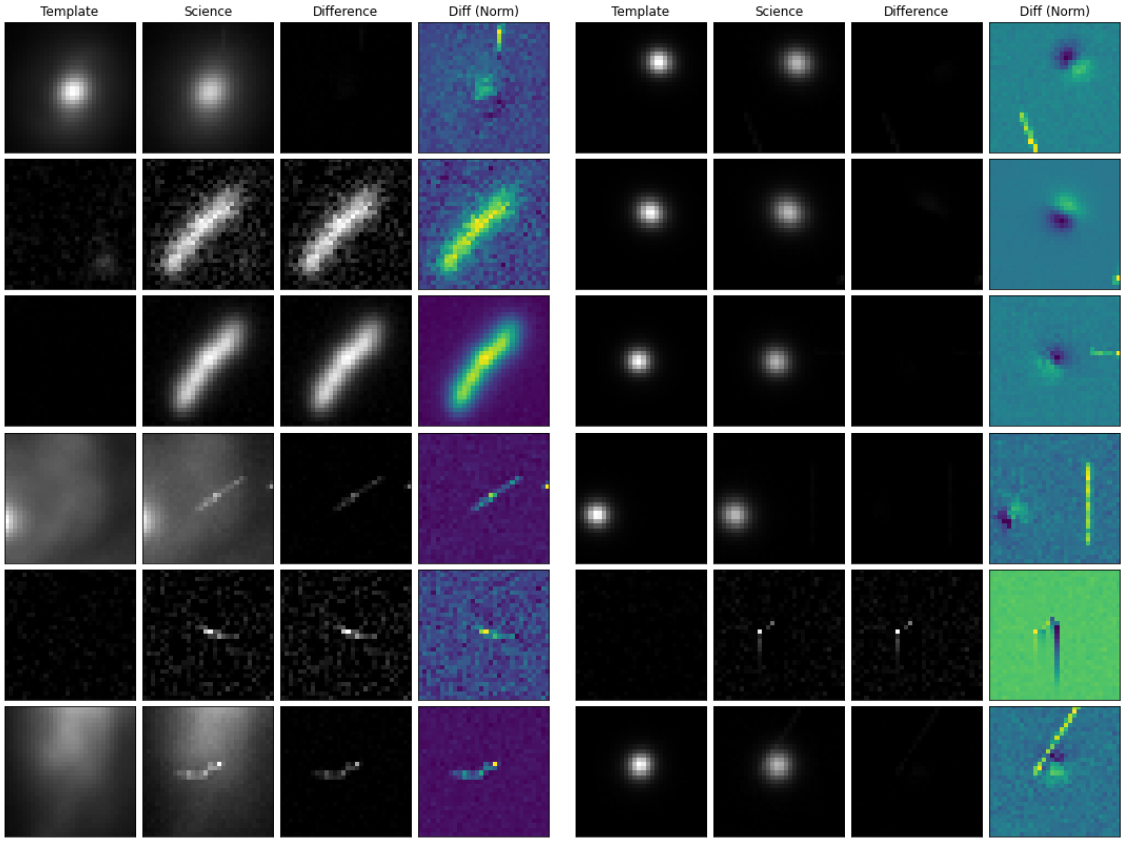
Figure 9: Random sample of 100 $30x30$ pixels stamps cut around middle of the DIASources that are labeled by the Default CR algorithm on science images as "not CR".

The stamp classification summary is listed in Table 1. Figure 8 shows example of "CR" stamps, while "non CR" examples are shown in figure 9.

Total	CR	not CR
18343	11211	7132
	61%	39%

Table 1: Distribution of the labels of the $30x30$ pixels stamps cut around middle of the DIASources that are labeled by the Default CR algorithm on science images inside the visit 11710

The adopted classification algorithm is not perfect and although most stamps are correctly labeled, there are also some mislabeled stamps. Figure 10 shows that some stamps labeled as "not CR" clearly contain CRs, while some examples of false negatives are shown in figure 10a. Therefore, the current implementation of Alert Production pipeline can produce alerts that are purely a result of misidentifying a CR. On the other hand, all visually examined "CR" stamps really contain a CR so false positive rate is very low.

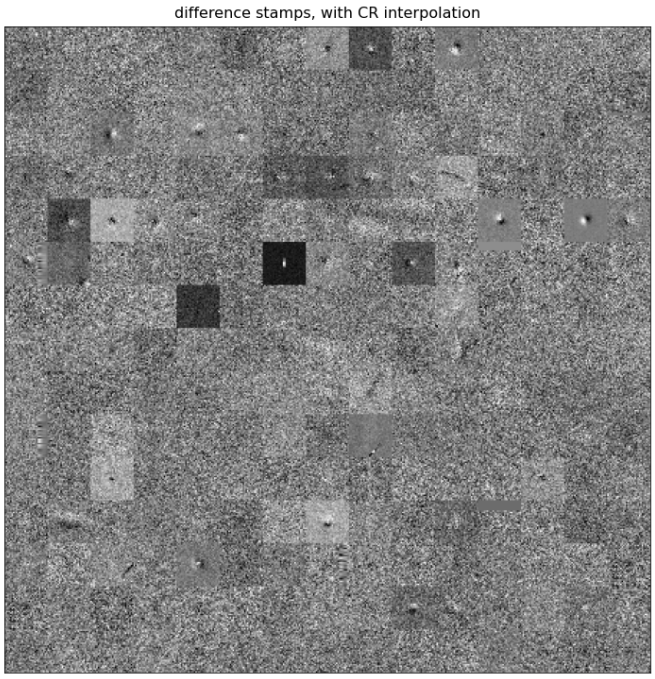


(a) Stamps labeled as "not CR". Rows 1, 4, 5 and 6 clearly contain CR

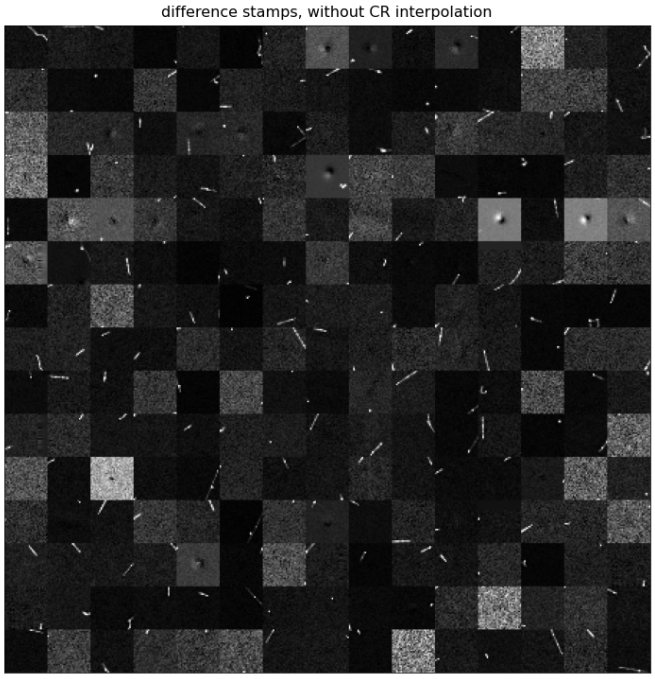
(b) Stamps labeled as "CR"

Figure 10: Examples of stamps that have been examined in greater detail. Template, Science and Difference images use the same normalization factor, meaning that the brightest pixel of all three is white and the darkest is black, with the color linearly proportional to values. This way the ratio of pixel values between Template, Science and Difference is uniform. Diff (norm), the fourth column, is a recolored version of the Difference that has been normalized by itself, making the image's greatest value light green and the lowest dark purple.

2.1.5 Interpolation



(a) CR interpolation.



(b) No CR interpolation.

Figure 11: First image showing stamps that had one or more CR pixels even after interpolation of the CRs was turned on. Second image is showing same images but with the interpolation of the CRs turned off. The interpolation in most cases is good, although there are some cases interpolation caused artifacts in the image.

To evaluate interpolation, two runs were created. One run as described in subsection 2.1.4 that will be called the "non-interpolation" run, and the other one with the interpolation of the CRs that will be called the "interpolation" run. There are 227 stamps with more than one CR pixels in the CR mask. All of those stamps are shown in figure 11. Most of the objects are cases where both DIASource and CRs are present on the same stamp, but there are at least two cases where DIASource finder picked up bad interpolation. This would create an alert only due to poor interpolation.

2.1.6 Real DIASources contained in CR stamps

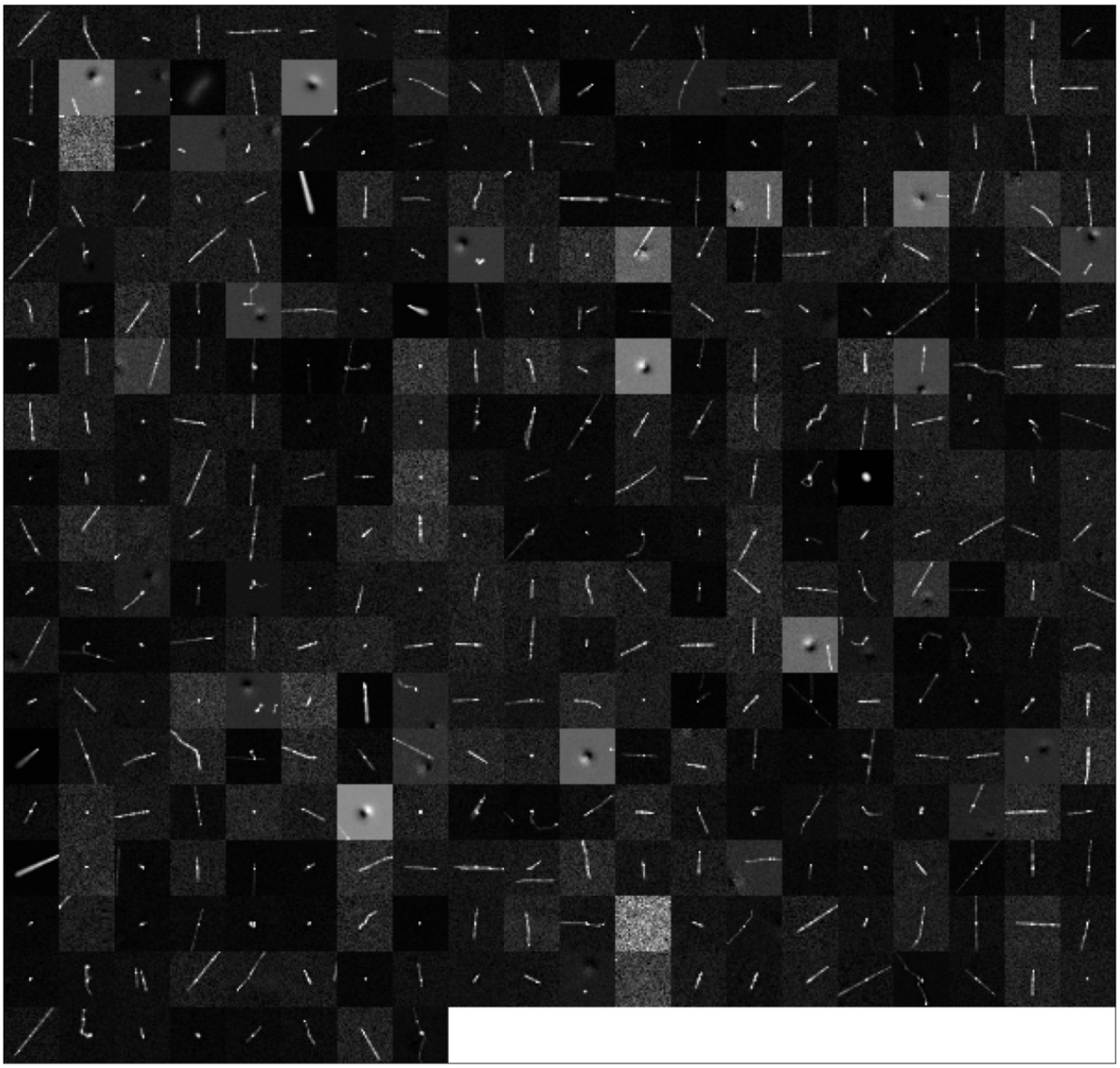


Figure 12: All of the stamps that are labeled as both containing OBJECT and CR. Out of 368, only 36 can be clearly confirmed to contain both. This means that the stamps labeled as CR are not heavily contaminated by the real astronomical objects.

It is useful to understand how many stamps contain both CRs and real DIASources. Another run was produced, this time with the interpolation turned on. This way the algorithm creates mask in the difference image that contains information which pixels are part of DIASources (similar to CR mask). This mask should contain only objects that algorithm thinks are true DIASources (because interpolation was turned on). Afterwards, every stamp is labeled as OBJECT or NO OBJECT similarly to the process described in 2.1.4. This way the number of stamps containing real DIASources can be found. 11,211 stamps were labeled as CR and 3,540 as OBJECT. These two labels are represented in the confusion matrix in table 2. After manual examination of the stamps labeled both OBJECT and CR, only 36 out of 368 are found to be true objects containing both cosmic rays and astronomical objects (figure 12). This means that after manual examination only 36 stamps contain both object and CR, out of the 3,540 that contain objects. In other words 1% of detected DIASources contain CR pixels.

	CR	NOT CR
OBJECT	368 (2.01%)	3172 (17.29%)
NO OBJECT	10843 (59.11%)	3960 (21.59%)

Table 2: Confusion matrix of the stamps labeled as CR by the Default CR algorithm with CR interpolation turned off, and stamps labeled as objects with the interpolation turned on. Most of the OBJECTS do not contain CR pixels.

2.2 Default CR algorithm on difference images

The same algorithm described in section 2.1 is now used again but instead of Science images, it is applied to difference images. The reason behind that is that difference images usually contain fewer objects and have flatter background, so differences between the "CR" and "not CR" pixels should be more obvious to the algorithm. Difference images are produced the same way as in the section 2.1, by running Default CR algorithm with interpolation turned off. Output of this approach is the CR mask which conveys the information which pixels were recognised as CRs. This algorithm detected 878,807 CR pixels inside science mask. Comparing that masks with the masks produced in section 2.1 confusion matrix can be calculated, showing how many pixels are detected by which algorithm. The confusion matrix is shown in table 3.
This algorithm detected 506,112 pixels as CR pixels. It is valuable to calculate:

$$CR = 0.061\% = 610.1\, CR\, pix/M\, pix$$

where N_{CRpix} is the number of CR pixels in difference mask and N_{totpix} is the total number of pixels. The result is 610.1 cosmic ray pixels per megapixel of detector or around 0.061% of all pixels are CRs. If we denote pixels which are detected as following (performance of Difference, relative to Science)

- Science image CR and Difference image CR = TP
- Science image CR and Difference image NOT CR = FN
- Science image NOT CR and Difference image CR = FP
- Science image NOT CR and Difference image NOT CR = TN

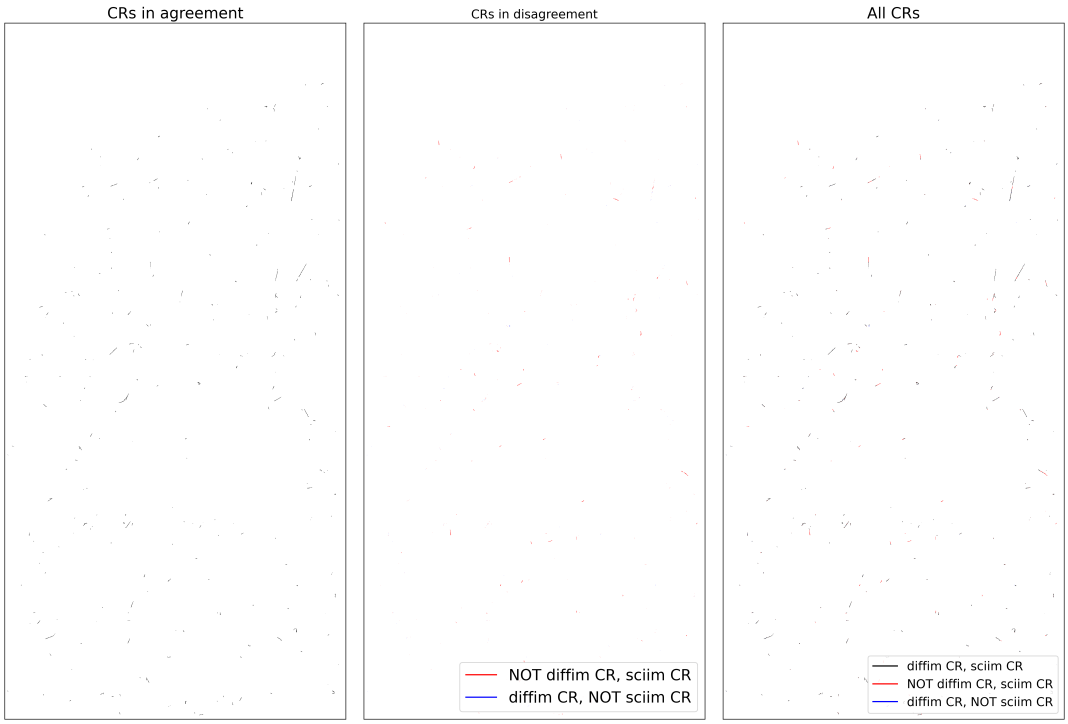
We can calculate:

$$\begin{aligned} purity &= \frac{TP}{TP+FP} \\ completeness &= \frac{TP}{FN+TP} \\ F1 &= 2 \cdot \frac{purity \cdot completeness}{purity + completeness} \end{aligned} \tag{1}$$

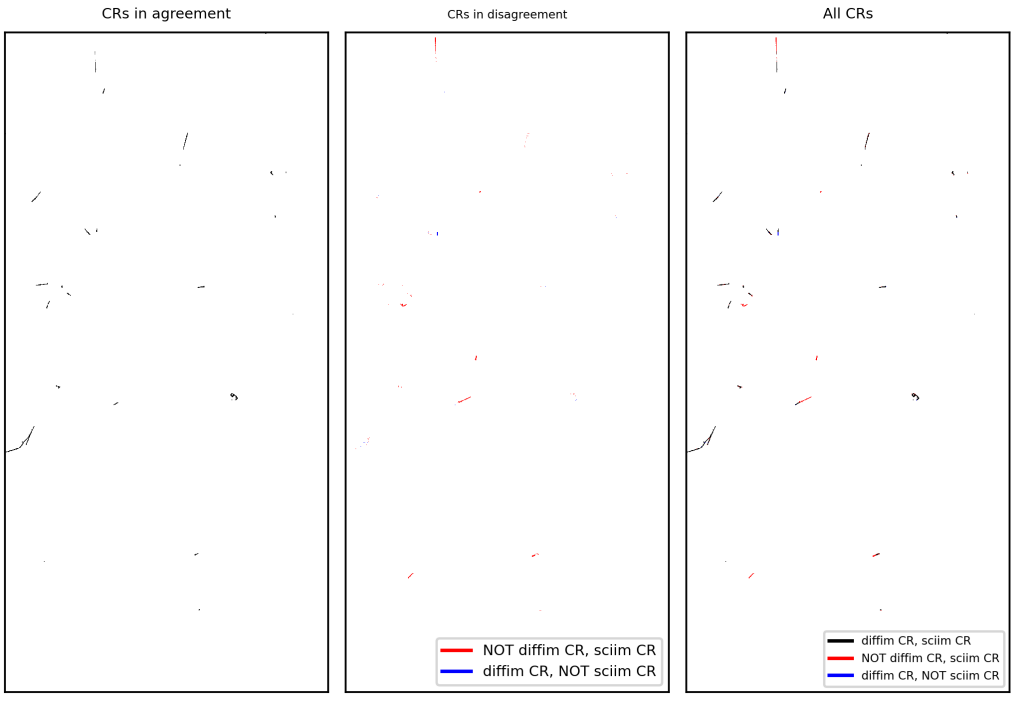
	Difference image CR	Difference image NOT CR
Science image CR	490941 (0.06%)	285297 (0.03%)
Science image NOT CR	15171 (0.002%)	828796047 (99.90%)
Purity: 0.970	Completeness: 0.632	F1 score: 0.766

Table 3: Pixel-wise comparison of the CR mask produced by the Default CR algorithm on science images and difference images. Purity, completeness and F1 score, for the performance of Difference relative to Science, are calculated using equations 1

The science image CR detection masks and difference image CR detection masks are analyzed. Pixels that are detected by both methods as well as those that are detected by only one method are closely examined. Out of the total 791,409 pixels that are detected by both methods, 1.9% are detected only on difference images, 36.1% are detected only on science images, and 62% are detected on both. An example for a single detector is shown in figure 13.



(a) Whole detector 62.



(b) Zoomed in center part of the detector 62.

Figure 13: Comparison of the CR masks obtained using different methods for detector 62 visit 11710. "diffim CR" are pixels that are detected as CRs using the Default CR algorithm on difference images, while "sciim CR" are the ones detected using Default CR algorithm on science images. Most of the "diffim CR" are also detected in "sciimg mask", but the "sciimg" mask contains 7,483 pixels labeled as CRs while "diffimg" mask contains only 5,803 for this detector.

Out of the 103 detectors, this algorithm evaluated only 97 detectors, and it failed for 6 with the error: 'Too many CR pixels (max 10000)'. Detectors that failed are: 16, 42, 43, 51, 52, 60. Coincidentally, those detectors are the ones that have some number of satellite trails. Probably detection of those pushed the CR pixel count over 10000 per detector causing the algorithm to fail. Execution of the CR detection part of the Default CR algorithm took around 3.4 sec of runtime per detector.
Difference images are identical as in section 2.1. Therefore, the DIASource detector is detecting exactly the same DIASources. The only difference is that, when applying the procedure described in section 2.1.4, some stamps have different labels. We use the same convention from 1, but instead of pixels, labels of the stamps are used. Note that the $F1 = 0.997$ match between two algorithms is very high; more quantitative details are listed in table 4

	Difference image CR	Difference image NOT CR
Science image CR	10480 (57.13%)	2 (0.01%)
Science image NOT CR	66 (0.36%)	7795 (42.5%)
Purity: 0.994	Completeness: 1.000	F1 score: 0.997

Table 4: Stamp-wise comparison of the CR masks produced by the Default CR algorithm on science images and difference images. Purity, completeness and F1 score for the performance of Difference relative to Science are calculated using equations 1, but instead of pixels, labels for the whole stamps are used.

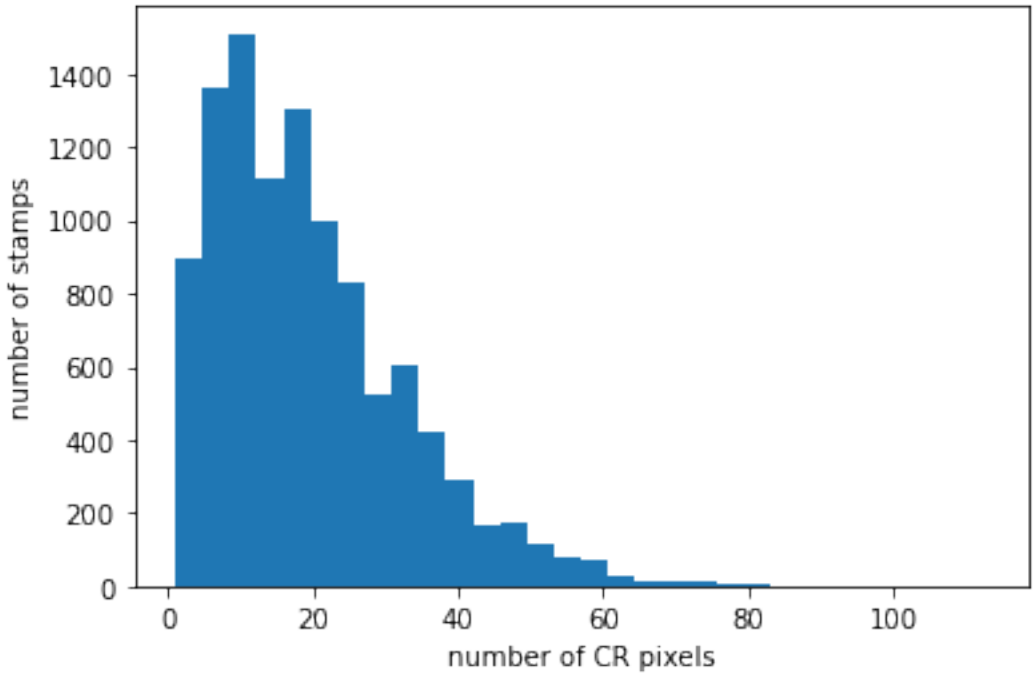


Figure 14: Histogram of the distribution of stamps with different number of CR pixels detected in the difference image mask. Median number is 17 CR pixels per stamp. Stamps with 0 CR pixels are not taken into consideration.

Image 14 shows distribution of the stamps with different number of CR pixels detected by the Default CR algorithm on difference images. Distribution is pretty similar to the distribution in figure 5, which is an indication that we are detecting most of the objects similarly to the Default CR algorithm on science images. If we use median value of 17 CR pixels per CR as a means to evaluate CR flux, we get:

$$j = 0.053 \frac{CR}{cm^2 \cdot s}$$

where N_{CR} is rough number of CRs, A is the area of the telescope and t is the time of the exposure.

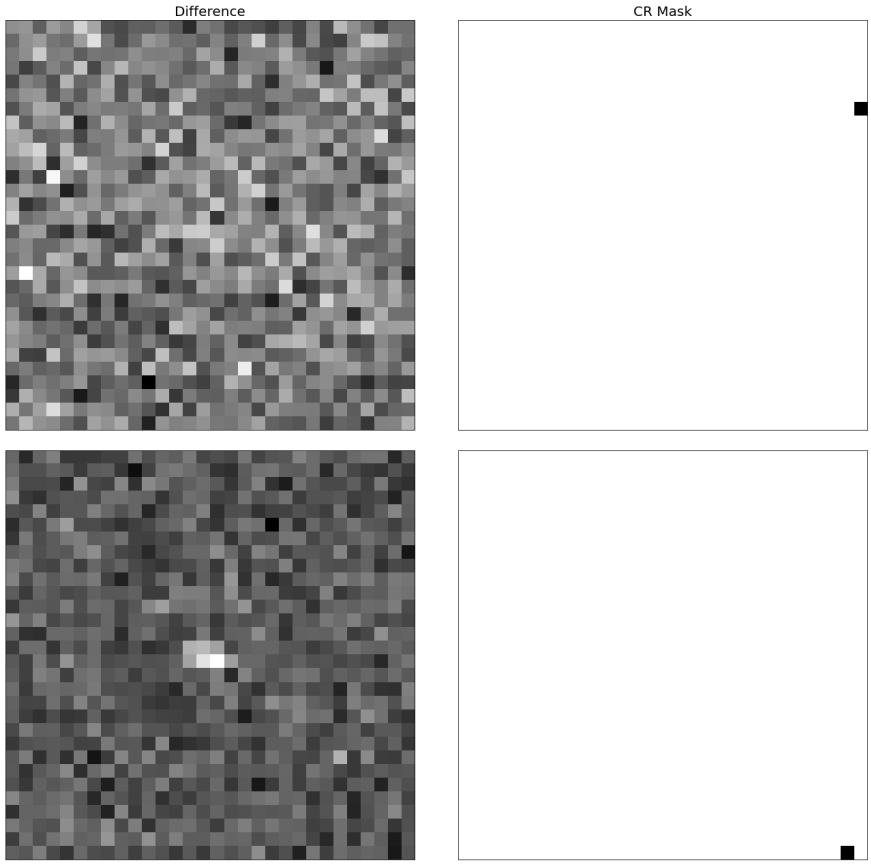
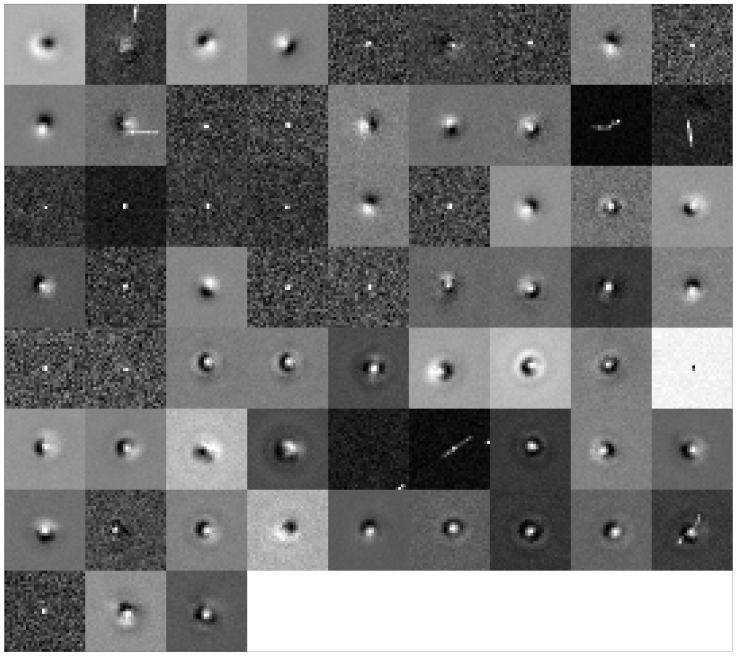
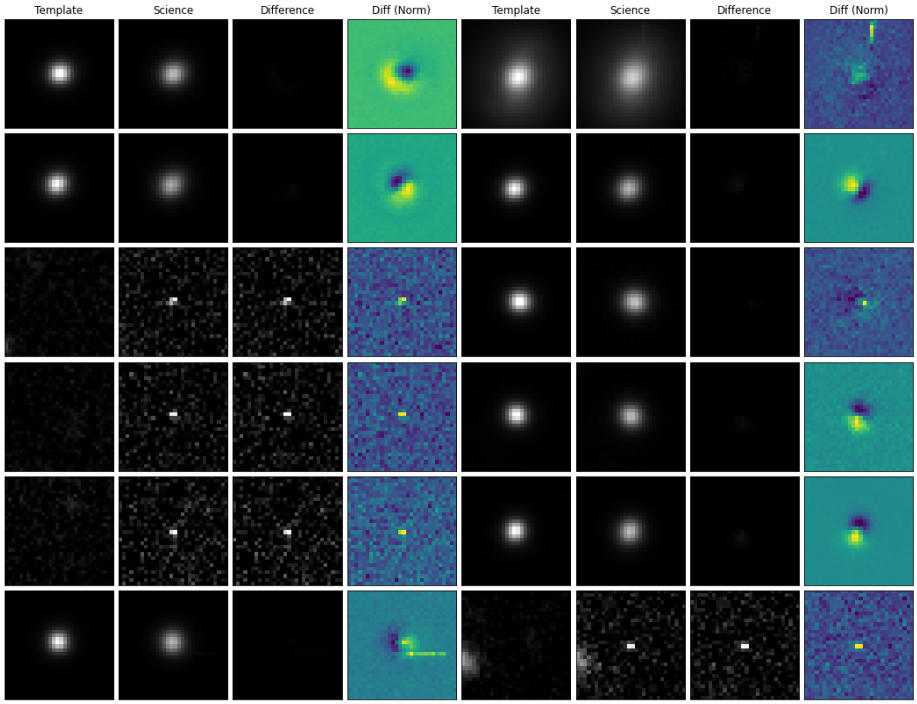


Figure 15: All of the stamps that the Default CR algorithm on science images detected as CR, and the Default CR algorithm on difference images detected as NOT CR. Left column shows difference stamps, and the right shows where does the Default CR algorithm on the science images think the CR position is.

Examples of the stamps that had different labels between the algorithms are shown in figures 15 and 16. A closer examination indicates that the predictions from Science images are more likely to be false, and there are no CR present.
Examples in figure 16 show examples of the opposite, where science image algorithm labeled stamps NOT CR and difference image as CR. From those examples it is clear that 24 (36.4%) of stamps show what appears to be a CR, 41 (62.1%) show what appears to be a dipole and 1 that is neither. This behaviour can be explained by the fact that the algorithm is made to work on images that do not contain negative values, and when the transition between positive and negative values in dipoles is too abrupt, algorithm interprets that object as CR.



(a) All of the stamps that the Default CR algorithm on science images detected as NOT CR, and the Default CR algorithm on difference images detected as CR. It appears that when algorithm is run on Difference images, it can find CRs missed when it's run on Science images. Also the algorithm detects dipoles as CRs, so further tuning is needed.



(b) First 12 images from the figure 16a showing template, science and difference cutouts, all together normalized, meaning that the brightest pixel of all three is white and the darkest is black and the color is linearly proportional to values. Diff (norm) is a recolored version of the difference that has been normalized by itself, making the image's greatest value light green and the lowest dark purple.

Figure 16

2.2.1 Conclusion

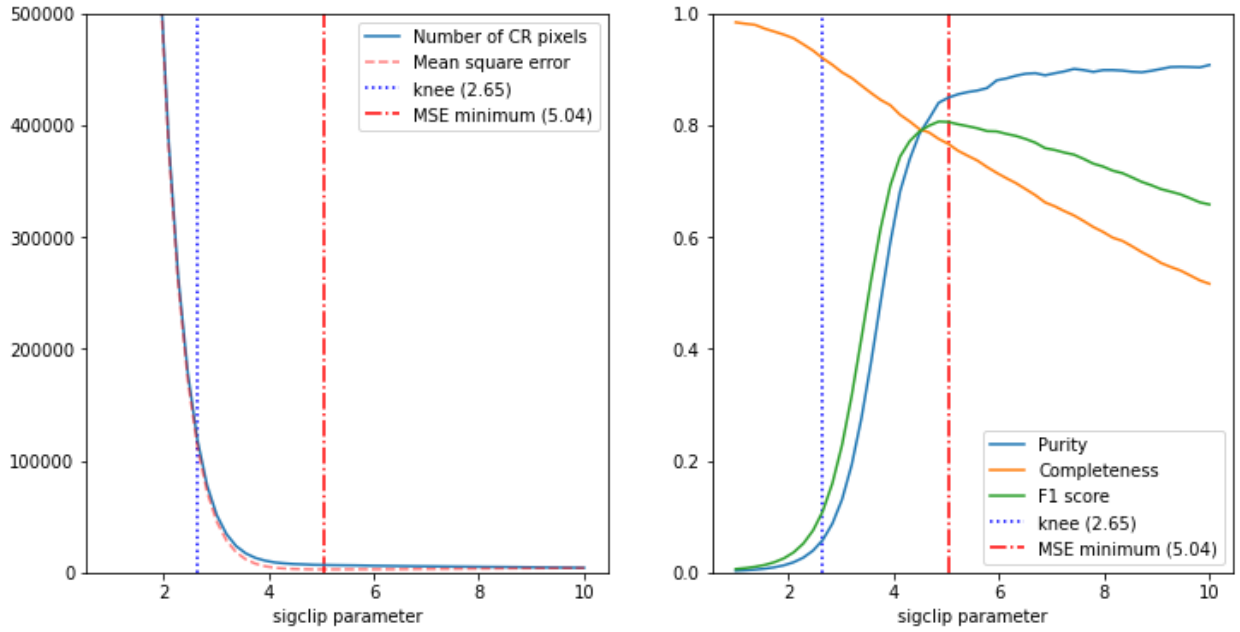
CR detection performed on difference images shows no significant improvement compared to science images. When compared to the Default CR algorithm on science images, the algorithm detected fewer pixels as the CR, but the performance after labeling the stamps does not differ significantly. Aside from the possibility that there are certain types of CRs that algorithm didn't find in either case, a high level of agreement between the two approaches is encouraging.

2.3 AstroScrappy

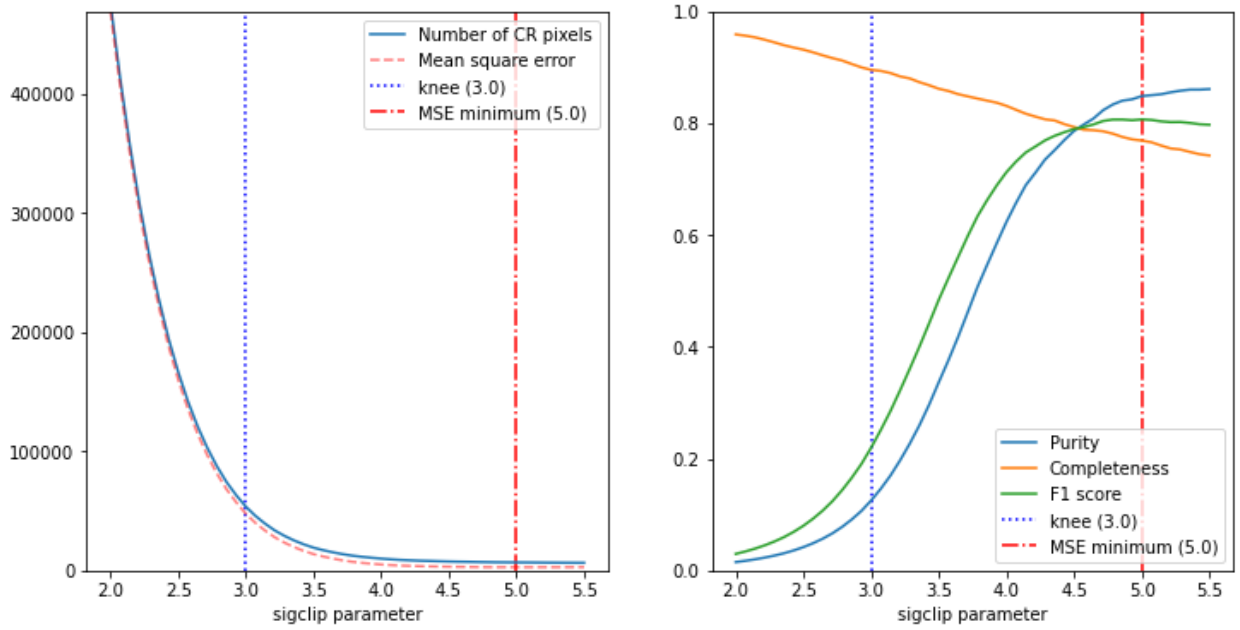
AstroScrappy is the openly available implementation of the Laplacian Edge Detection called LA Cosmic. The code can be found on github.com/astropy/astroscrappy, while the LA Cosmic is described in paper: van Dokkum, "Cosmic-Ray Rejection by Laplacian Edge Detection". This algorithm is similar to the one implemented in the Default CR algorithm, but its main advantage is that it has detailed documentation for usage. For that reason, it is easier to change parameters and fine-tune for a specific instrument. Algorithm was optimized by random search for the parameter called *sigclip*. This parameter controls Laplacian-to-noise limit for cosmic ray detection, and lower values will flag more pixels as cosmic rays. After testing, optimizing only this parameter was deemed necessary because changing other parameters did not produce any significant improvement. Optimization was done on only one detector, to save time. Two significant points in the parameter space were evaluated, one using the knee method and one in the mean square error compared to the Default CR algorithm in the difference images. A predetermined estimate of the data variance (ie., noise squared) in each pixel was provided by the Default CR algorithm inside the variance mask. Algorithm was tested on both difference and science image. For both types of images algorithm was first evaluated with such value of parameter sigclip for which minimum of the mean squared error compared to Default CR algorithm on science images, and after that at the value at which "knee" occurs in the graph of the sigclip vs. number of CR pixels graph. Second method refers to the "knee method" of optimisation and consists of choosing a point where diminishing returns are no longer worth the additional cost.
Runtime of AstroScrappy is around 90 s per detector, but it can be probably optimised to run in parallel mode. Further investigation is needed for more details of the shrinking AstroScrappy runtime.

2.3.1 Science images

AstroScrappy was created for use on the astronomical images that are most similar to science images. Parameter *sigclip* was identified to be mostly responsible for the amount of the pixels detected as the CR. Before evaluating the performance, the best value of the *sigclip* was found using two methods. First method is minimizing the mean square error (MSE) compared to Default CR algorithm on science images, and the other one is the "knee" method. For both methods two searches were performed: wide search and narrow search. In the wide search the value of *sigclip* was shifted for 0.18 each step in the range between 1 and 10. Both knee and minimum MSE points were found in the interval between 2 and 5.5, so the narrow search was performed, with the *sigclip* step of 0.07. For each step number of found CRs, MSE, purity and completeness compared to Default CR algorithm on science images were evaluated. Figure 17 contains detailed plot.



(a) Wide search for parameter sigclip.



(b) Narrow search for parameter sigclip.

Figure 17: Optimization of the AstroScrappy with respect to Default CR algorithm on science images. Wide search was performed with steps of 0.18, while narrow search had 0.07 step size.

2.3.2 Science images at MSE minimum

AstroScrappy algorithm detected 812489 CR pixels on the science images, when $sigclip = 5.0$, what represents minimum of the mean squared error compared to Default CR algorithm on science images.

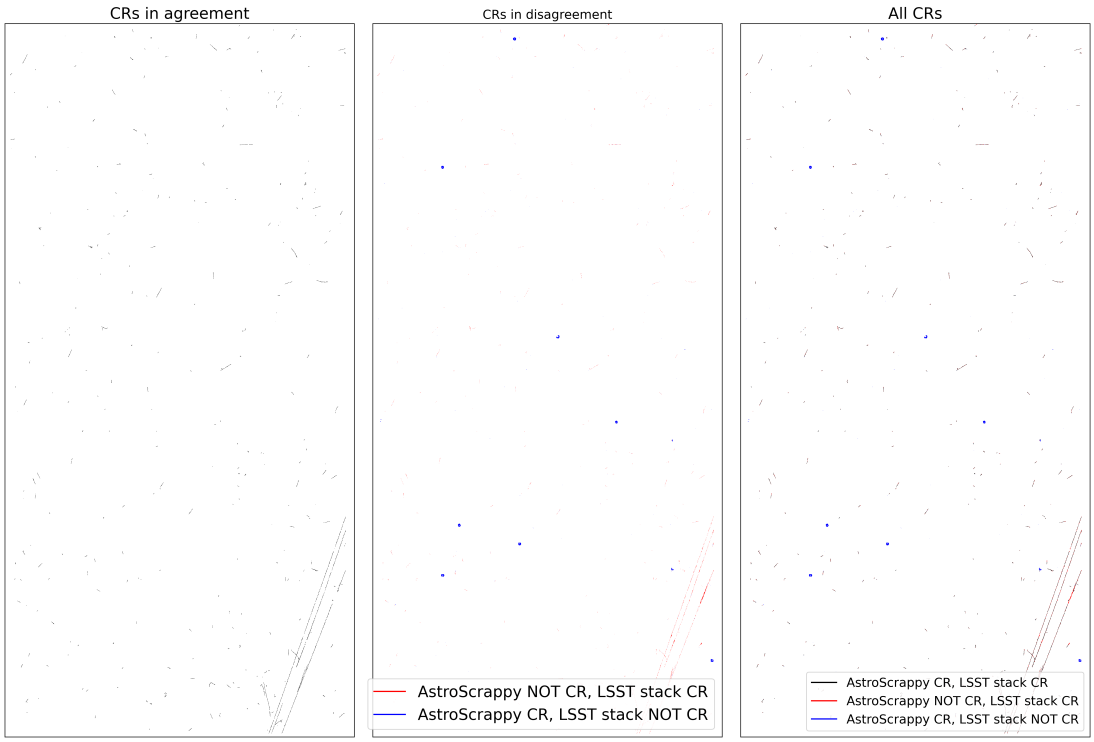
$$CR = 0.092\% = 922.3\, CR\, pix/M\, pix$$

Comparing that to Default CR algorithm detector on the science images of 997.6 $CR\,pix/M\,pix$, number of CR pixel is similar. Table 5 shows the confusion matrix between those two algorithms. After closer examination it's clear that agreement between the two algorithms was only $F1 = 0.792$. This value of *sigvalue* is chosen because the agreement between the algorithms is the highest.

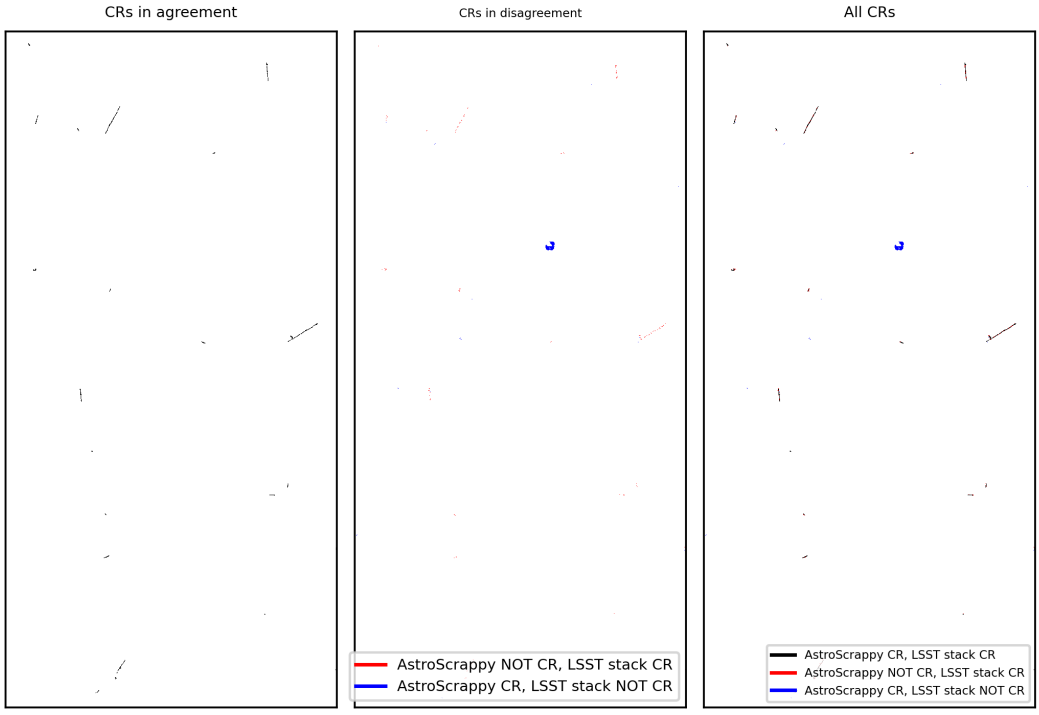
	AstroScrappy CR	AstroScrappy NOT CR
Default CR algorithm CR	669342 (0.076%)	209465 (0.024%)
Default CR algorithm NOT CR	143147 (0.016%)	879880190 (99.884%)
Purity: 0.824	Completeness: 0.762	F1 score: 0.792

Table 5: Pixel-wise comparison of the CR mask produced by the Default CR algorithm on science images and AstroScrappy on science images.

The CR detection masks produced by the algorithm was compared to the CR detection masks produced by the Default CR algorithm on science images. Pixels that are detected by both methods as well as those that are detected by only one method are closely examined. Out of the total 1021954 pixels that are detected by both methods, 14% are detected only with the AstroScrappy, 20.5% are detected only with Default CR algorithm on science images, and 65.5% are detected on both. An example for a single detector is shown in figure 18.
Peculiar behaviour of the AstroScrappy can be manifested by having circular, blob shaped structures of CR detection. It is not clear the reason for their existence. Closer look on those shapes are shown on figure 19. Those structures are not around any particular object, and closer look revealed no recognisable, so the reason of their occurrence is unknown.

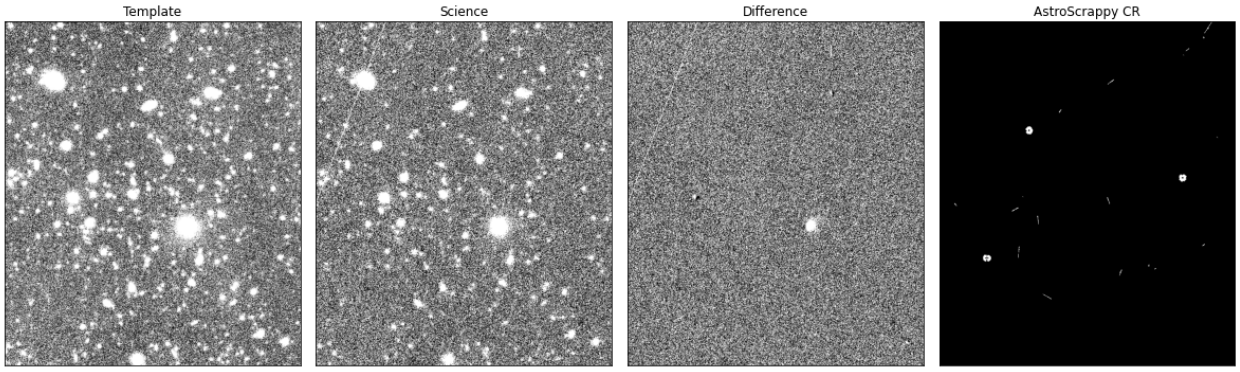


(a) Whole detector 42.

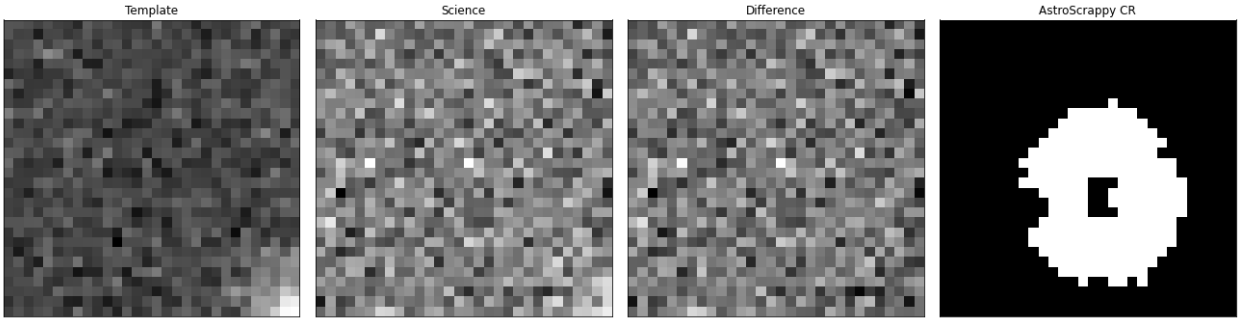


(b) Zoomed in center part of the detector 42.

Figure 18: Comparison of the CR masks of different methods on detector 42 visit 11710. LSST stack stands for the Default CR algorithm. There are some visible "blobs" of AstroScrappy misdetection of the CRs. Otherwise algorithms are mostly in agreement.



(a) Cluster of nearby mislabeled blobs.



(b) Zoomed in to one region of mislabels.

Figure 19: Exploring blobs of defective detection of CRs by AstroScrappy on science images. It is unclear what is the reason of the misdetection.

Mask produced by the AstroScrappy was used to determine CR and NOT CR labels for the DIASources described by the process in section 2.1.4, and compared to the same process using the Default CR algorithm on science images. Match between the two algorithms is high with $F1 = 0.998$. AstroScrappy determined that 61.35% of the DIASources are CR-s while 38.65% are NON CR. Details are listed in table 6

	AstroScrappy CR	AstroScrappy NOT CR
Default CR algorithm CR	11191 (61.01%)	20 (0.11%)
Default CR algorithm NOT CR	63 (0.34%)	7069 (38.54%)
Purity: 0.994	Completeness: 0.998	F1 score: 0.996

Table 6: Stamp-wise comparison of the CR mask produced by the Default CR algorithm on science images and AstroScrappy on science images.

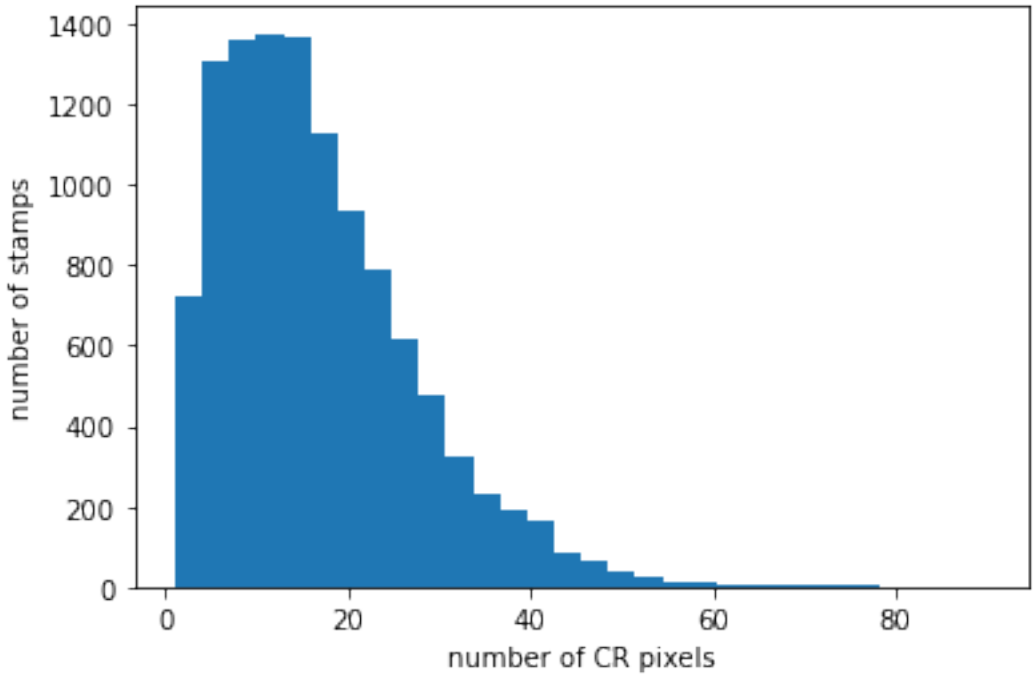


Figure 20: Histogram of the distribution of stamps with different number of CR pixels detected with the AstroScrappy algorithm on science images. Median number is 14 CR pixels per stamp. Stamps with 0 CR pixels are not taken into consideration.

With the median of 14 CR pixel per one cosmic ray particle, as determined on the figure 20, CR flux is:

$$j = 0.098 \frac{CR}{cm^2 \cdot s}$$

Images that are recognised by the Default CR algorithm on science images as CR, and by the AstroScrappy on science images as NOT CR are shown in figure 21. Those images show something that does not look like a CR, so most of this detection by Default CR algorithm is false positives. This shows that AstroScrappy has lower affinity for false positive detection.
Images that are recognised by the Default CR algorithm on science images as NOT CR, and by the AstroScrappy on science images as CR are shown in figure 22. Most of those images contain CR-s, so this proves that Astroscrappy can detect CR-s that Default CR algorithm can't. Some questionable detection is due to errors and artifacts in the images.
This shows that while the F1 score was lower on the table 5, on the CRs that could be mixed up with DIA sources, AstroScrappy has superb performance.

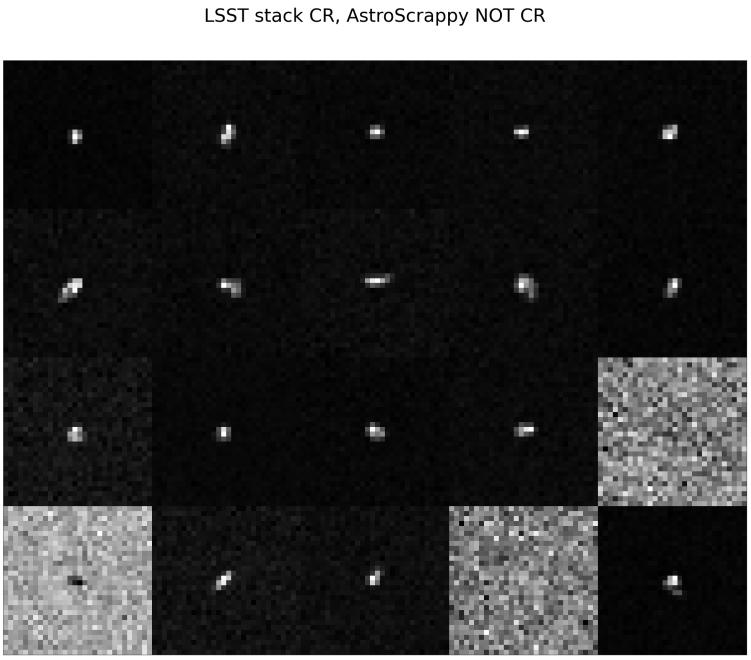


Figure 21: All of the stamps that the Default CR algorithm on science images detected as CR, and the AstroScrappy on science images detected as NOT CR. AstroScrappy mostly missed only CRs that don't look like a straight line.

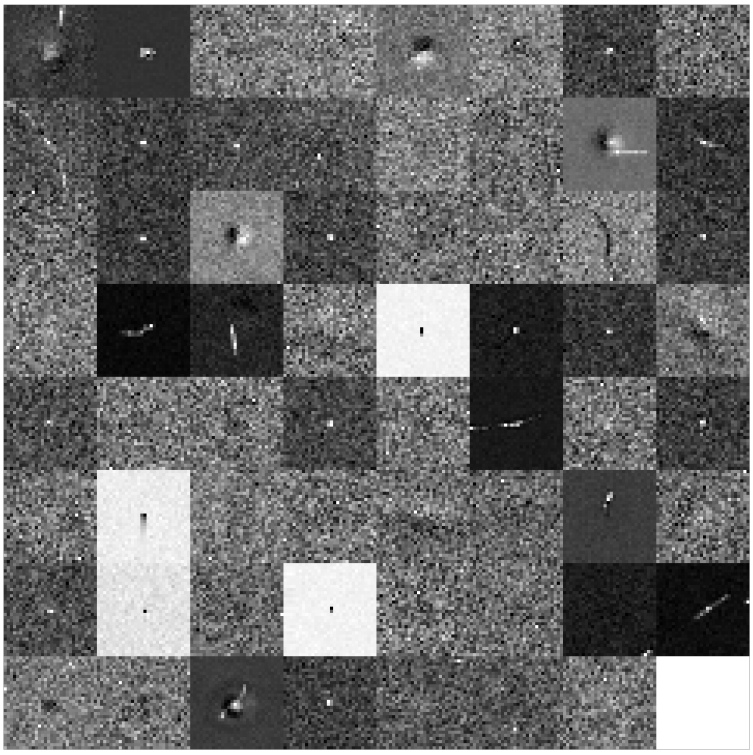


Figure 22: All of the stamps that the Default CR algorithm on science images detected as NOT CR, and the AstroScrappy on science images detected as CR. AstroScrappy detected CRs that Default CR algorithm missed.

2.3.3 Science images at knee

The point for which AstroScrappy has parameter $siglicp = 3.0$ represents the "knne" point. At that point diminishing increase in CR pixels does no longer follow the increase in the parameter. Since this point is as the significantly lower value of *sigclip*, higher number of detected CR pixels are expected. AstroScrappy algorithm detected 6,268,128 CR pixels on the science images, which is around seven times higher than Default CR algorithm on the science images.

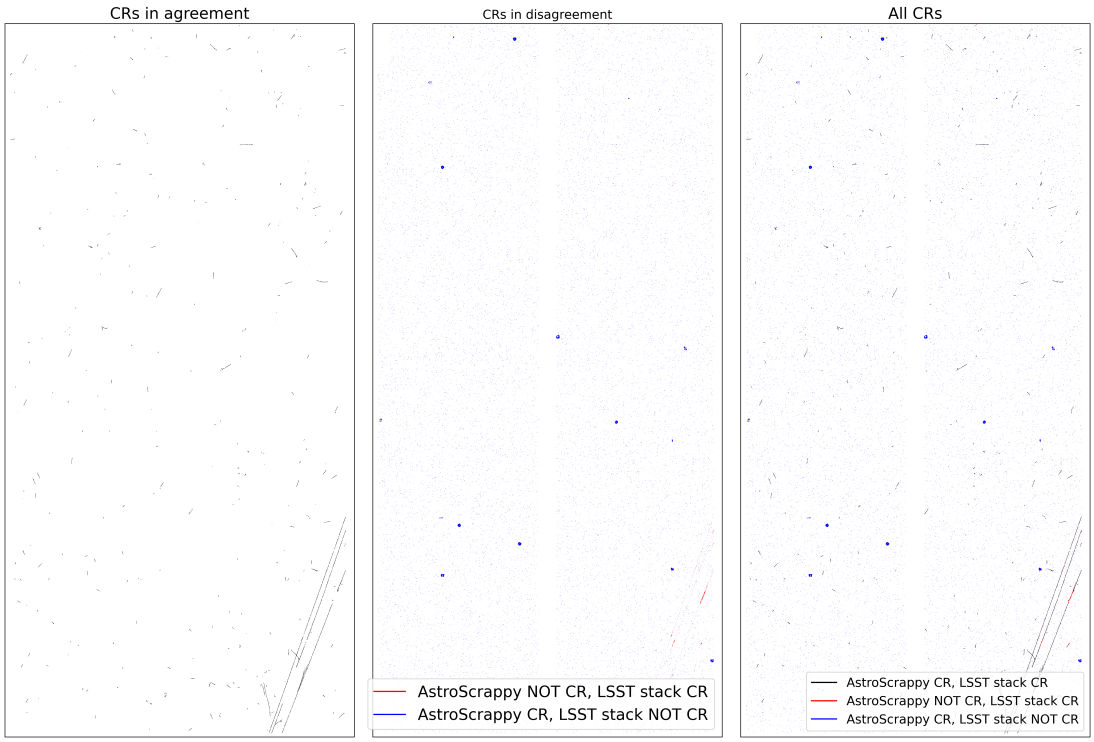
$$CR = 0.712\% = 7115.6\, CR\, pix/M\, pix$$

Agreement between the Default CR algorithm on science images and AstroScrappy on science images in the "knee" point is pretty low, with $F1 = 0.22$. This low score is mostly due to low purity, because AstroScrappy detected seven times more CR pixels.

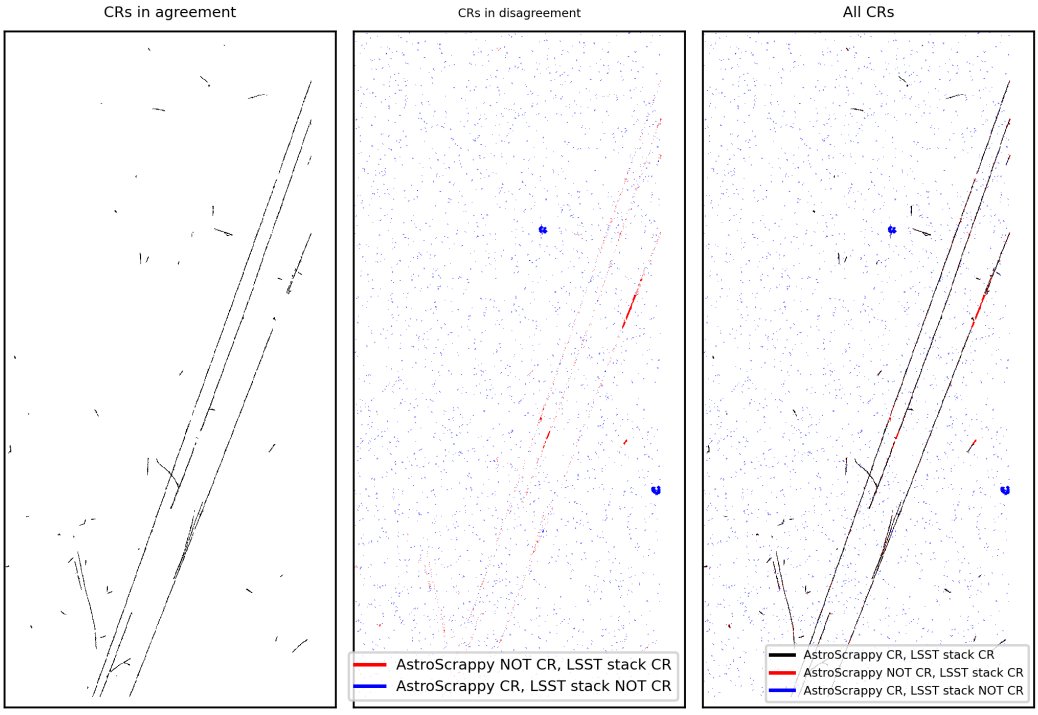
	AstroScrappy CR	AstroScrappy NOT CR
Default CR algorithm CR	785957 (0.089%)	92850 (0.011%)
Default CR algorithm NOT CR	5482171 (0.622%)	874541166 (99.278%)
Purity: 0.125	Completeness: 0.894	F1 score: 0.220

Table 7: Pixel-wise comparison of the CR mask produced by the Default CR algorithm on science images and AstroScrappy on science images.

Out of the total 6,360,978 pixels labeled as CR by both algorithms, only 12.4% are detected by both algorithms. 86.2% of CR pixels are detected only by AstroScrappy and 1.5% by only Default CR algorithm. This means that probably AstroScrappy is detecting a lot of false positives, because chance that Default CR algorithm mislabels pixels of that magnitude is unlikely. An example for a single detector is shown in figure 23. Peculiar structures of CR detection that is detected at the MSE minimum point, was also detected at the knee point. Its nature is unknown. Those structures are on exactly the same places as on the figures 18 and 19.



(a) Whole detector 42.



(b) Zoomed in bottom right part of the detector 42.

Figure 23: Comparison of the CR masks of different methods on the science images on detector 42 visit 11710. There is a high noise detection by the AstroScrappy in the "knee" point what means that probably the algorithm is not sufficiently tuned. Minimum MSE error point is the superior tuning method, therefore meaning the Default CR algorithm on the science images is also fine tuned to an acceptable degree.

Mask produced by the AstroScrappy was used to determine CR and NOT CR labels for the DIASources described by the process in section 2.1.4, and compared to the same process using the Default CR algorithm on science images. The match between the two algorithms is inadequate with $F1 = 0.77$. AstroScrappy determined that 97.4% of the DIASources are CR-s while 2.6% are NON CR. While completeness is 1, purity is very low; therefore, the algorithm optimized this way is not in agreement with the Default CR algorithm. Details are listed in table 8.

	AstroScrappy CR	AstroScrappy NOT CR
Default CR algorithm CR	11211 (61.12%)	0 (0.00%)
Default CR algorithm NOT CR	6651 (36.26%)	481 (2.62%)
Purity: 0.628	Completeness: 1.000	F1 score: 0.771

Table 8: Stamp-wise comparison of the CR mask produced by the Default CR algorithm on science images and AstroScrappy on science images.

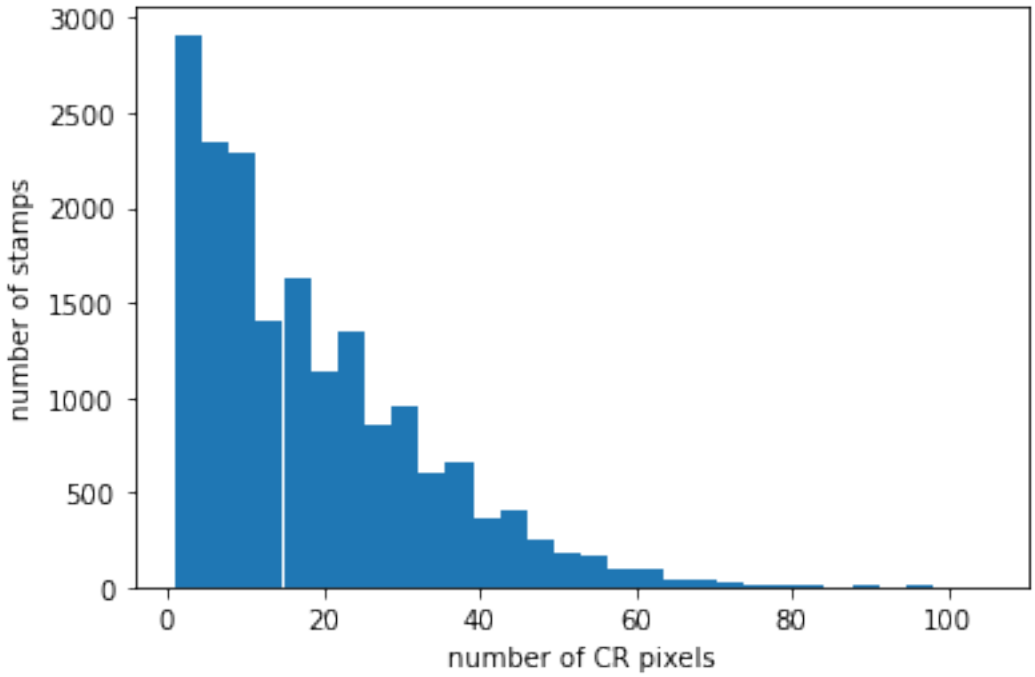


Figure 24: Histogram of the distribution of stamps with different number of CR pixels detected with the AstroScrappy algorithm on science images. Median number is 23 CR pixels per stamp. Stamps with 0 CR pixels are not taken into consideration.

With the median of 14 CR pixel per one cosmic ray particle, as determined on the figure 24, CR flux is:

$$j = 0.753 \frac{CR}{cm^2 \cdot s}$$

There are 0 stamps that are classified by Default CR algorithm on science images as CR and by AstroScrappy on science images in "knee" point as not CR. In addition, most of the stamps are classified as CR by AstroScrappy. This is very unusual and the primary suspect is poor choice of

sigclip. Figure 25 shows some examples of the different detection by the algorithms. It is obvious that most of the stamps are not showing CR-s.

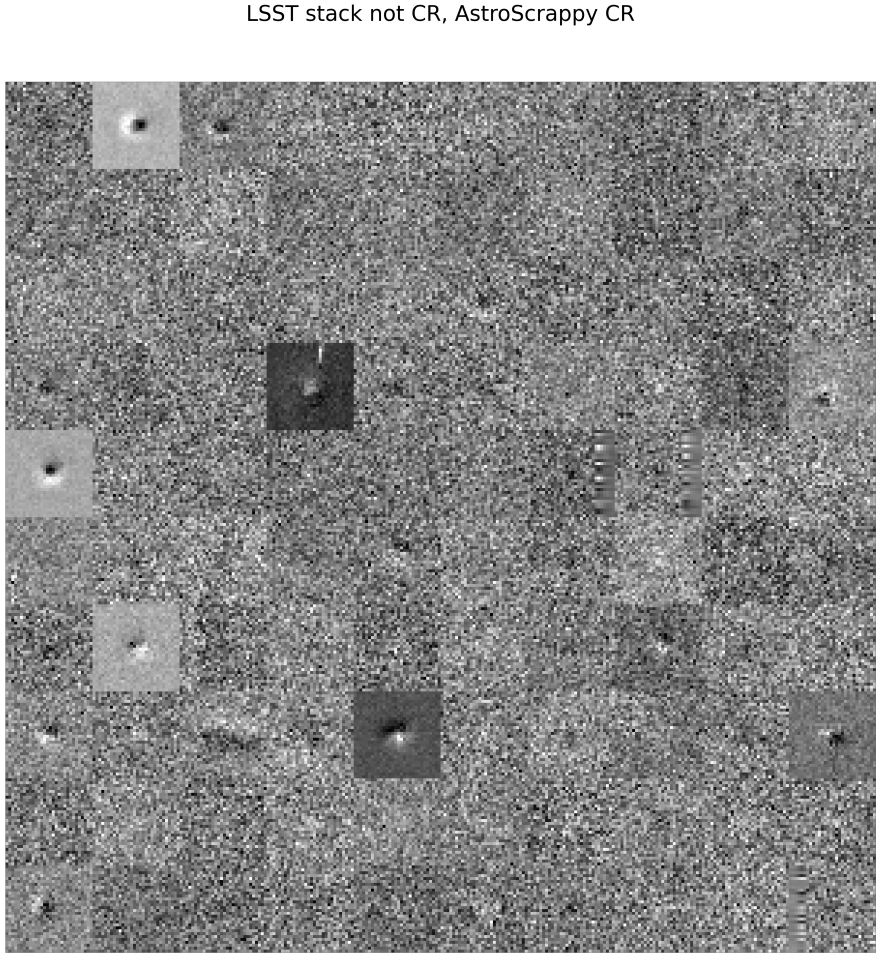
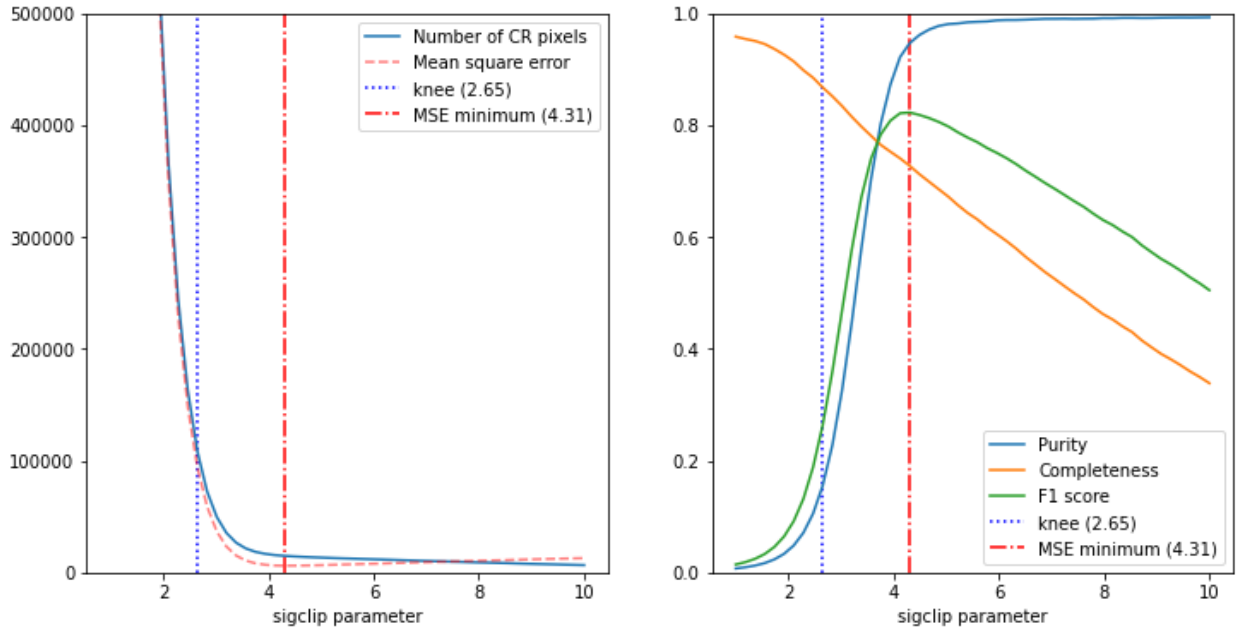


Figure 25: Some stamps that the Default CR algorithm on science images detected as NOT CR, and the AstroScrappy on science images detected as CR. Rate of false positives is high with this type of tuning.

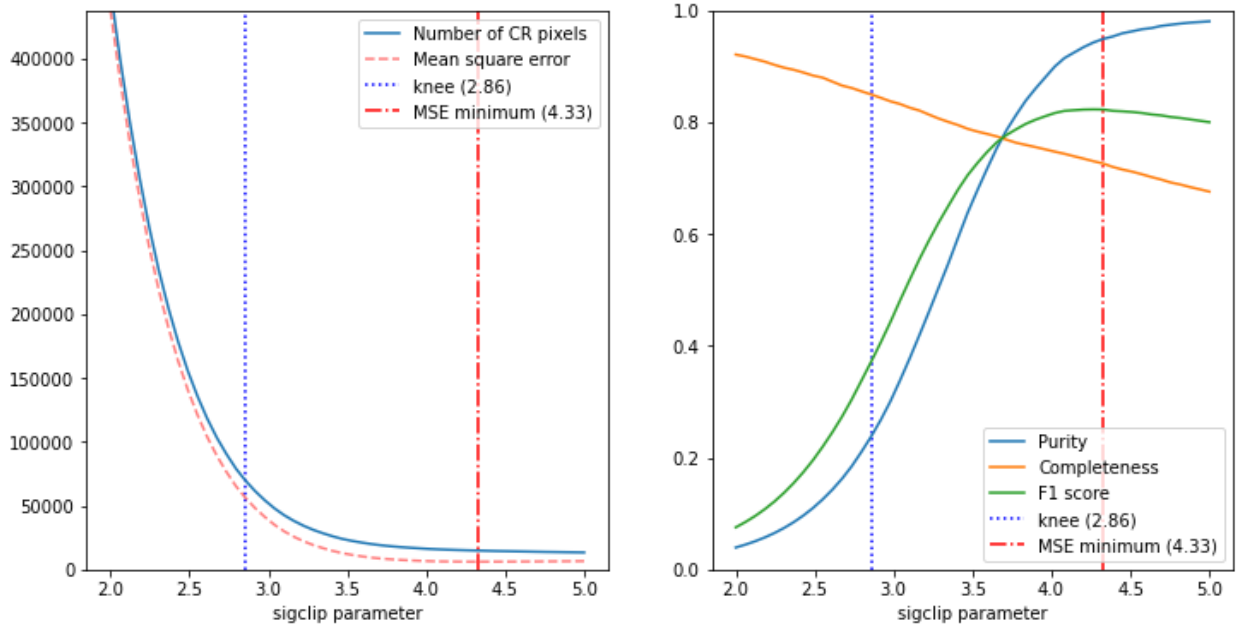
Optimizing AstroScrappy with the "knee" method yielded a poor performance, and it is not recommended.

2.3.4 Difference images

For the evaluation of the AstroScrappy on the difference images, the same optimization procedure of the parameter *sigclip* was necessary. In the wide search value of *sigclip* was increased for 0.18 in the interval between 1 and 10. Both minimum of MSE and "knee" point was located in the interval between 2 and 5, so the narrow search in that interval was performed with the step of 0.06. "Knee" point was located at 2.86 and MSE minimum at 4.33. For each step number of found CRs, MSE, purity and completeness compared to Default CR algorithm on science images were evaluated. Figure 26 contains detailed plot.



(a) Wide search for parameter sigclip.



(b) Narrow search for parameter sigclip.

Figure 26: Optimization of the AstroScrappy with respect to Default CR algorithm on science images. Wide search was performed with steps of 0.18, while narrow search had 0.06 step size.

2.3.5 Difference images at MSE minimum

AstroScrappy identified 665,506 pixels on the difference images as CRs, with the parameter *sigclip* = 4.33, what is recognised as the minimum of the mean squared error between the AstroScrappy al-

gorithm and the Default CR algorithm on science images.

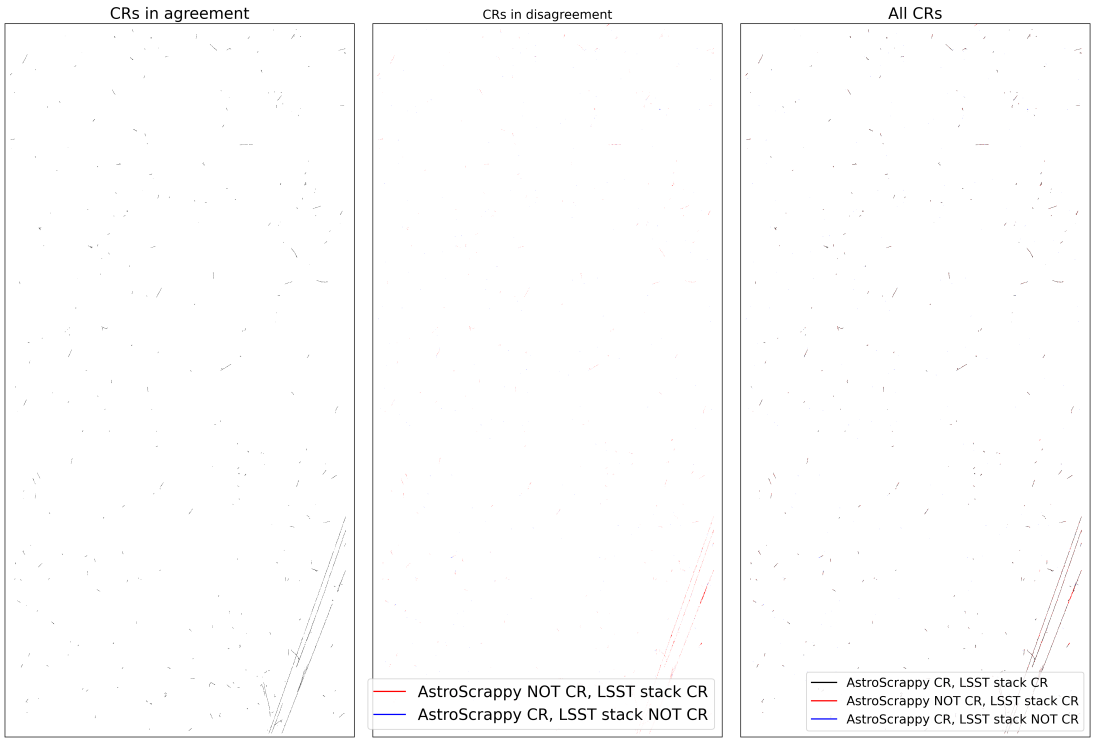
$$CR = 0.076\% = 755.5\, CR\, pix/M\, pix$$

Comparing that to Default CR algorithm detector on the science images of $997.6\, CR\, pix/M\, pix$, number of CR pixel detected is for around a quarter lower. Agreement between the algorithms is also low with the $F1 = 0.798$. This agreement is similar to the AstroScrappy on the science images at the MSE minimum.

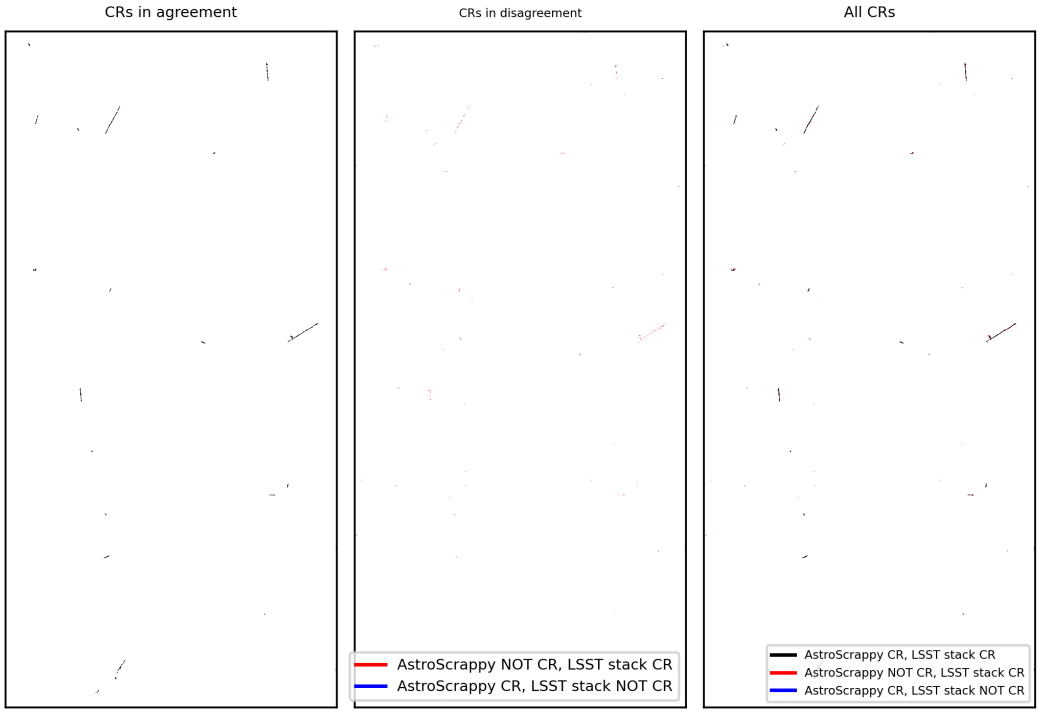
	AstroScrappy CR	AstroScrappy NOT CR
Default CR algorithm CR	616038 (0.070%)	262769 (0.030%)
Default CR algorithm NOT CR	49468 (0.006%)	879973869 (99.894%)
Purity: 0.926	Completeness: 0.701	F1 score: 0.798

Table 9: Pixel-wise comparison of the CR mask produced by the Default CR algorithm on science images and AstroScrappy on difference images.

The CR detection masks produced by the AstroScrappy algorithm was compared to the CR detection masks produced by the Default CR algorithm on science images. Pixels that are detected by both methods as well as those that are detected by only one method are closely examined. Out of the total 928275 pixels that are detected by both methods, 5.3% are detected only with the AstroScrappy, 28.3% are detected only with Default CR algorithm on science images, and 66.4% are detected on both. An example for a single detector is shown in figure 27.



(a) Whole detector 42.



(b) Zoomed in center part of the detector 42.

Figure 27: Comparison of the CR masks of different methods on detector 42 visit 11710. Agreement between the Default CR algorithm on the science images and AstroScrappy on the difference images tuned in the MSE minimum is high. This shows AstroScrappy can be tuned to reproduce Default CR algorithms results. performance.

	AstroScrappy CR	AstroScrappy NOT CR
Default CR algorithm CR	11193 (61.02%)	18 (0.10%)
Default CR algorithm NOT CR	180 (0.98%)	6952 (37.90%)
Purity: 0.984	Completeness: 0.998	F1 score: 0.991

Table 10: Stamp-wise comparison of the CR mask produced by the Default CR algorithm on science images and AstroScrappy on difference images.

Stamp images of the DIASources were labeled as CR or NOT CR with the process described in section 2.1.4. Match between the Default CR algorithm on science images is high with $F1 = 0.998$ agreement.

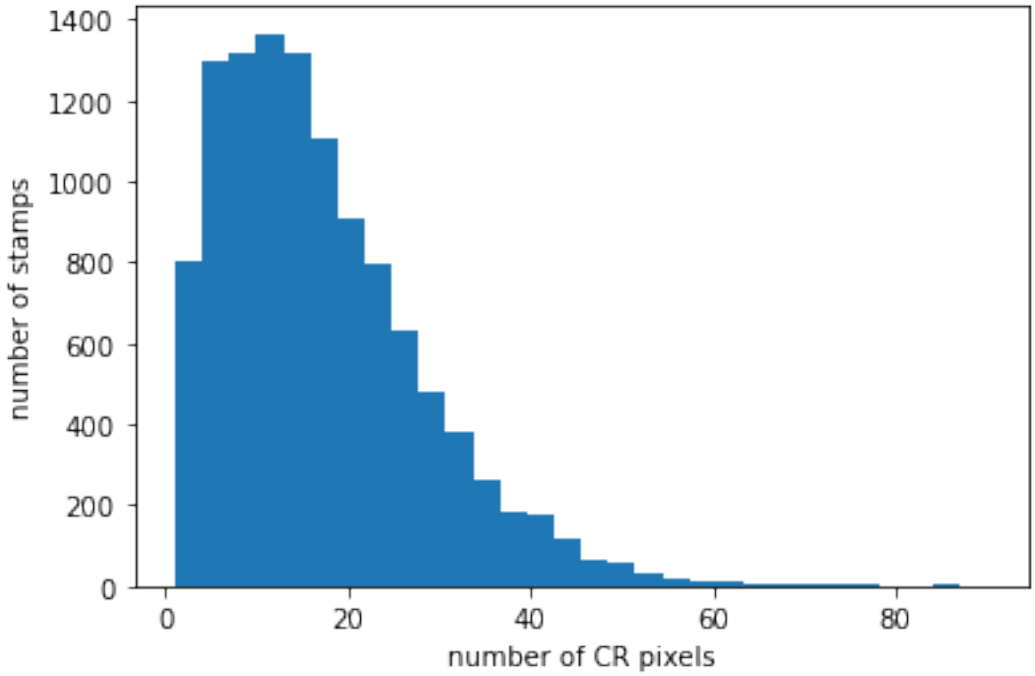


Figure 28: Histogram of the distribution of stamps with different number of CR pixels detected with the AstroScrappy algorithm on difference images. Median number is 15 CR pixels per stamp. Stamps with 0 CR pixels are not taken into consideration.

Median of 15 CR pixels per stamp is used to determine CR flux:

$$j = 0.075 \frac{CR}{cm^2 \cdot s}$$

Images that are recognised by the Default CR algorithm on science images as CR, and by the AstroScrappy on difference images as NOT CR are shown in figure 29. Those show that AstroScrappy properly labeled objects that look like CRs but are not CR.
Images that are recognised by the Default CR algorithm on science images as NOT CR, and by the AstroScrappy on difference images as CR are shown in figure 30. Most of those images contain CR-s, so this proves that Astroscrappy can detect CR-s that Default CR algorithm can't. However,

there are some examples where AstroScrappy detected dipoles as CR

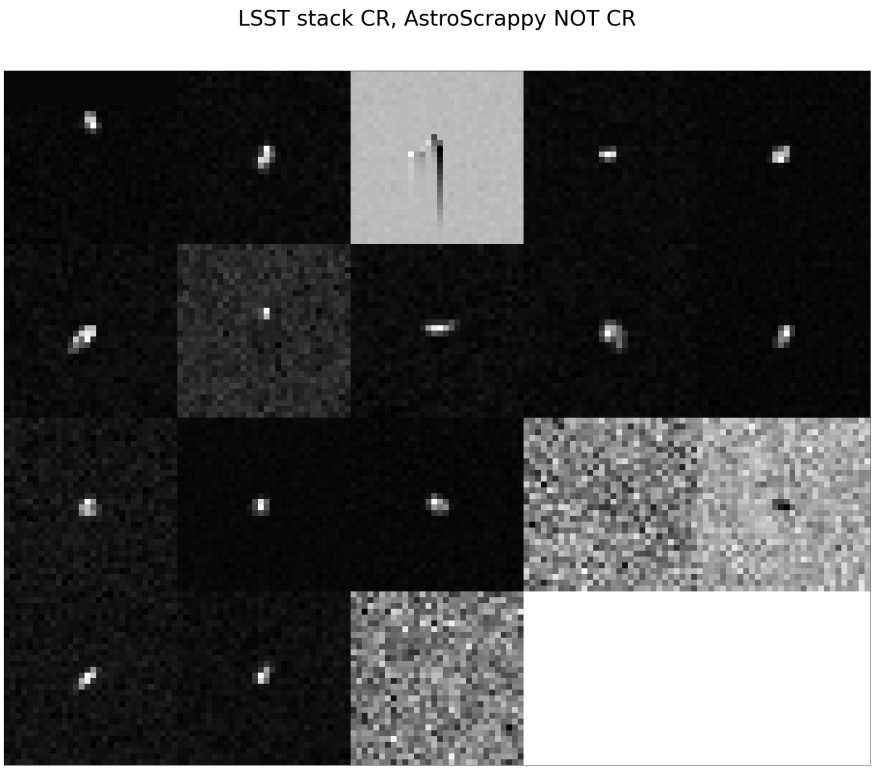


Figure 29: All of the stamps that the Default CR algorithm on science images detected as CR, and the AstroScrappy on difference images detected as NOT CR. AstroScrappy missed multiple CRs that don't look like a line. This is not completely obvious is it a misdetection by the AstroScrappy or the Default CR algorithm.

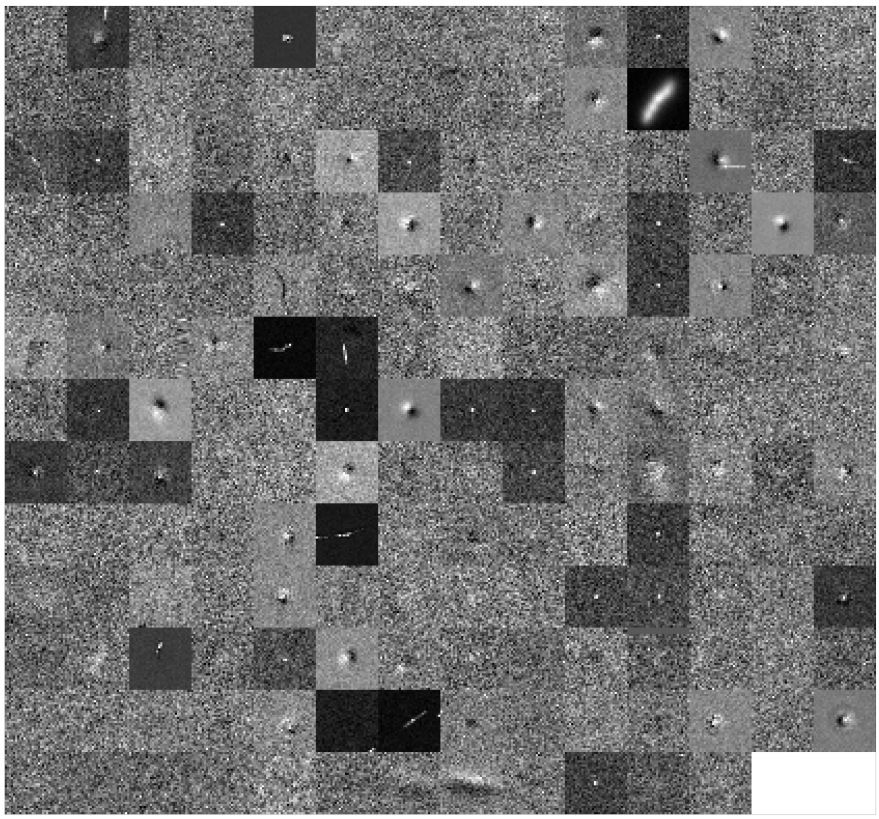


Figure 30: All of the stamps that the Default CR algorithm on science images detected as NOT CR, and the AstroScrappy on difference images detected as CR. Astroscrappy detects multiple dipoles as CRs.

AstroScrappy on the difference images shows great accuracy. Examining different predictions between the Default CR algorithm and the AstroScrappy, it is discovered that in most cases AstroScrappy was correct.

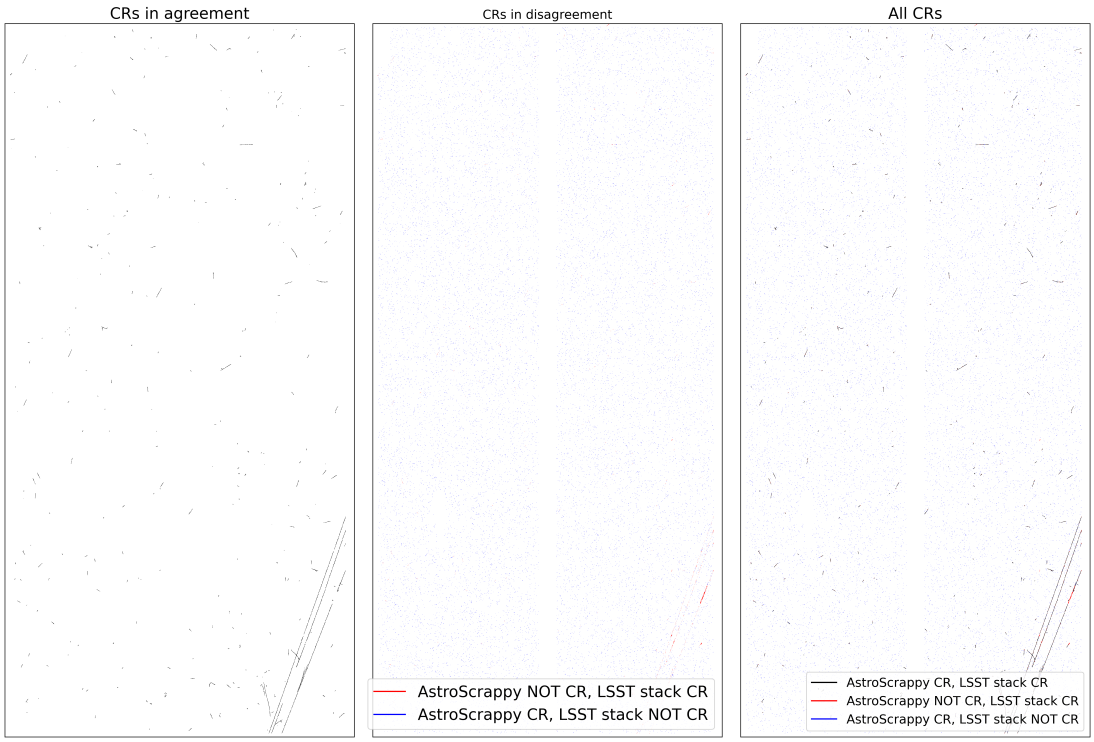
2.3.6 Difference images at "knee"

For completeness, AstroScrappy was evaluated in the "knee" point on difference images. Taught by experience, "knee" point has a significant purity problems. Following are the results:

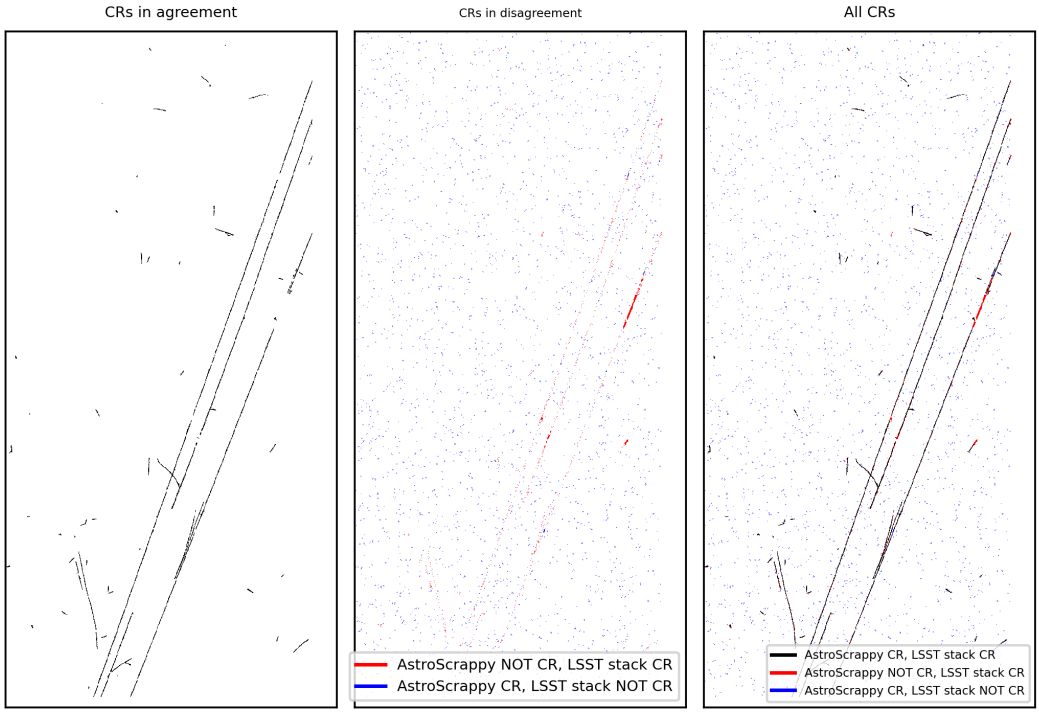
$$CR = 0.587\% = 5870.8\, CR\, pix/M\, pix$$

	AstroScrappy CR	AstroScrappy NOT CR
Default CR algorithm CR	698001 (0.079%)	180806 (0.021%)
Default CR algorithm NOT CR	4473643 (0.508%)	875549694 (99.392%)
Purity: 0.135	Completeness: 0.794	F1 score: 0.231

Table 11: Pixel-wise comparison of the CR mask produced by the Default CR algorithm on science images and AstroScrappy on difference images.



(a) Whole detector 42.



(b) Zoomed in bottom right part of the detector 42.

Figure 31: Comparison of the CR masks of different methods on detector 42 visit 11710. Inflated number of CR detection means this is not the optimal parameter tuning.

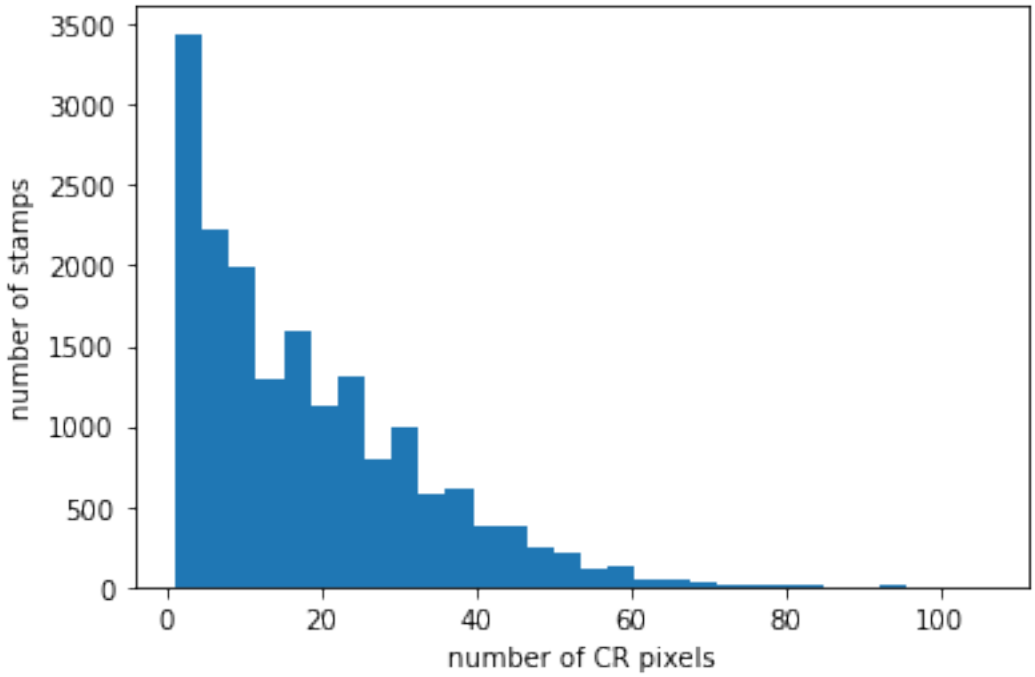


Figure 32: Histogram of the distribution of stamps with different number of CR pixels detected with the AstroScrappy algorithm on difference images. Median number is 15 CR pixels per stamp. Stamps with 0 CR pixels are not taken into consideration.

$$j = 0.621 \frac{CR}{cm^2 \cdot s}$$

	AstroScrappy CR	AstroScrappy NOT CR
Default CR algorithm CR	11209 (61.11%)	2 (0.01%)
Default CR algorithm NOT CR	6386 (34.81%)	746 (4.07%)
Purity: 0.637	Completeness: 1.000	F1 score: 0.778

Table 12: Stamp-wise comparison of the CR mask produced by the Default CR algorithm on science images and AstroScrappy on difference images.

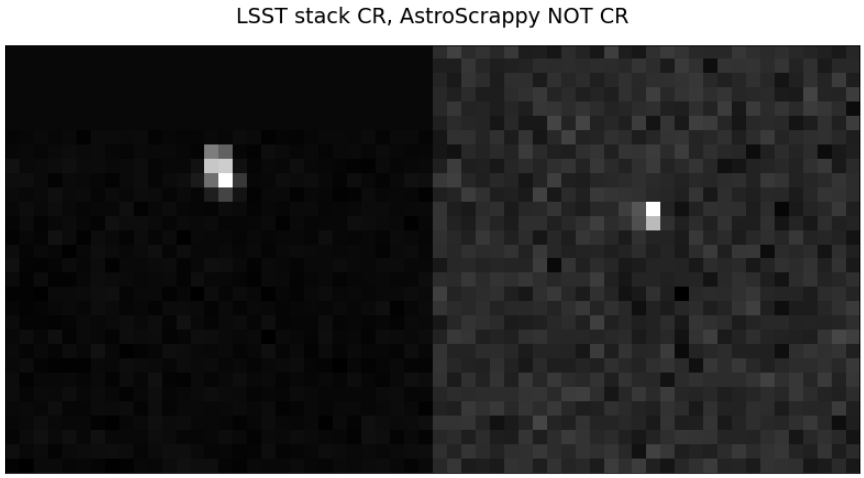


Figure 33: All of the stamps that the Default CR algorithm on science images detected as CR, and the AstroScrappy on difference images detected as NOT CR.

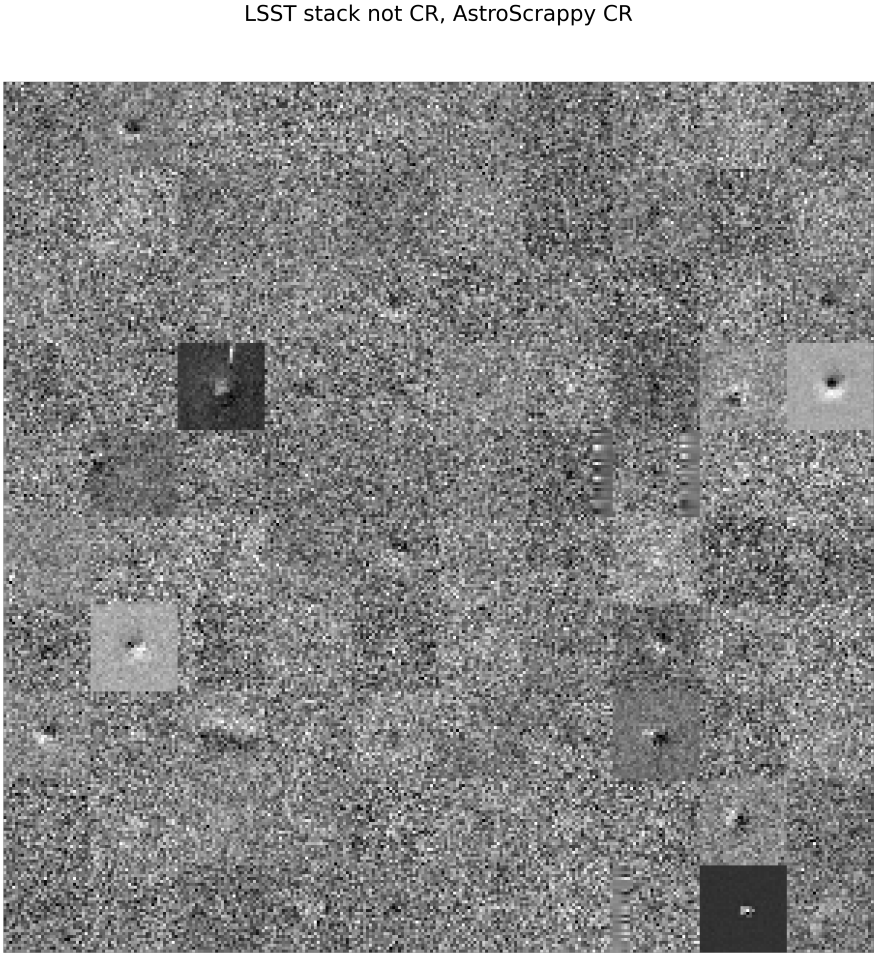


Figure 34: Some of the stamps that the Default CR algorithm on science images detected as NOT CR, and the AstroScrappy on difference images detected as CR.

3 Machine learning approach

Conventional CR-detection algorithms require tuning multiple parameters experimentally making it hard to fine tune it for different instruments. For that reason, the machine learning approach was investigated. In this approach, a different paradigm of programming is taken. Instead of programming explicitly rules that pixels have to satisfy in order to be detected as CR, data is fed to the machine learning (ML) algorithm. ML then learned to extract that rule from the data, and to generalize it to data yet unseen. Learning in which the program is given the data (images) as well as the correct answers, in this case the correct CR mask, is called supervised learning. Another type of learning is unsupervised, where program is asked to establish connections in data and come to conclusions without knowledge about the "correct answers". This type of learning does not require manual examination of the data, but the ML program is usually less successful in performing the task.

3.1 Cosmic-CoNN

Machine learning algorithm Cosmic-CoNN is the U-net type of deep convolution that uses deep layers interconnected with convolution, maxpooling and upsampling operations. This type of deep models are great in the segmentation type tasks such as this one and it was developed originally for biomedical image segmentation. Cosmic-CoNN was trained on the dataset produced from the thousands of images from the Las Cumbres Observatory global telescope network. Details about the models can be found in the publication: Xu *et. al.*: "Cosmic-CoNN: A Cosmic Ray Detection Deep-Learning Framework, Dataset, and Toolkit".

We tested Cosmic-CoNN on the HSC data and compared the performance to the Default CR algorithm on the science images. Runs with Cosmic-CoNN CR detection on the both science images and difference images are carried out, and compared using the Default CR algorithm run on science images as "meta truth". Cosmic-CoNN was not trained on any detector that has similar thickness of the HSC CCD detector, and consequently was presented during training with the CRs that don't necessary have the same shape.

3.1.1 Science images

This algorithm was trained on images that are most similar to our science images, and when tested on science images it detected 758082 pixels as CR.

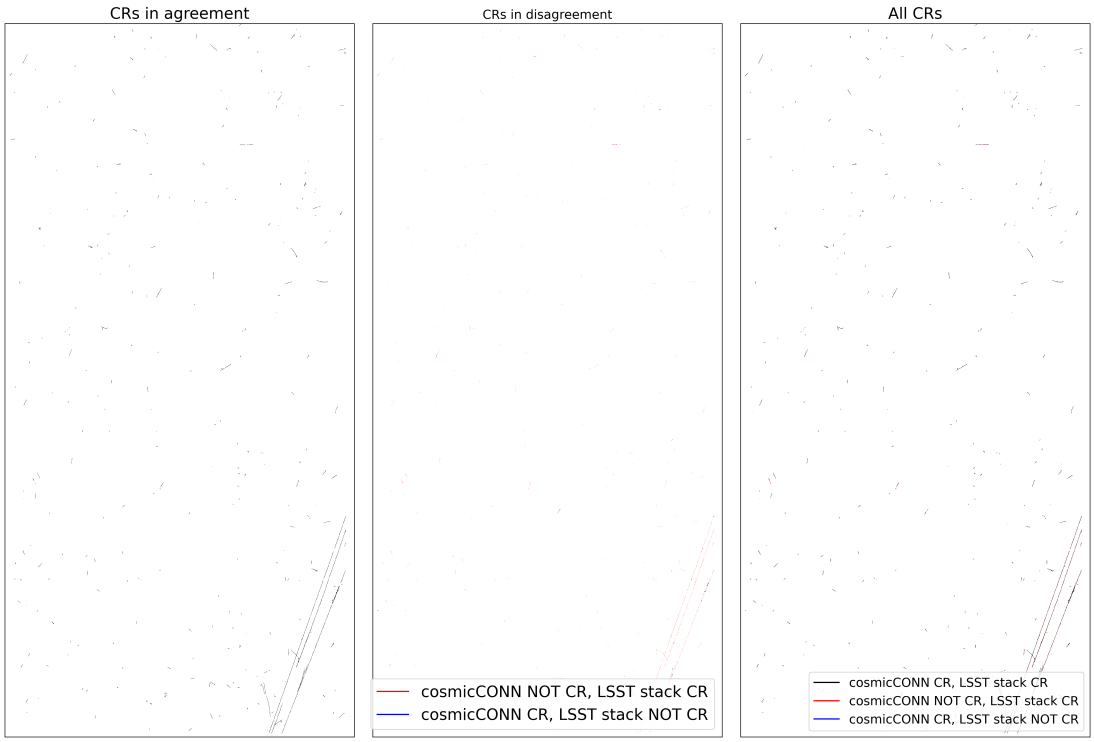
$$CR = 0.086\% = 860.6\, CR\, pix/M\, pix$$

Comparing that to Default CR algorithm detector on the science images of $997.6\, CR\, pix/M\, pix$, results are pretty similar.

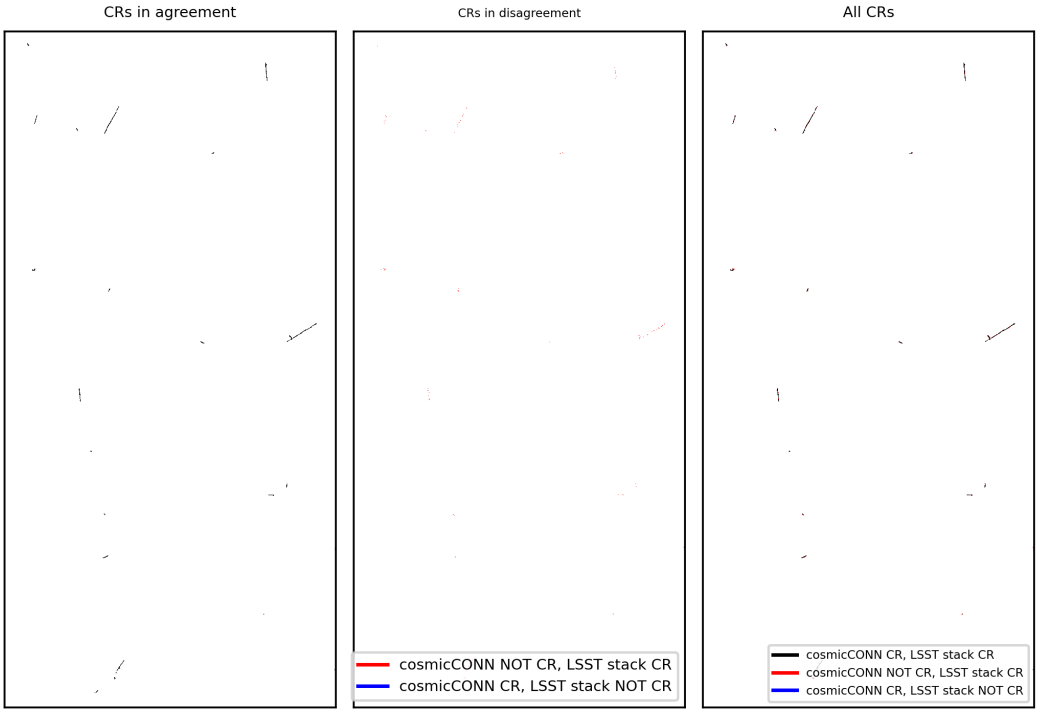
	Cosmic-CoNN CR	Cosmic-CoNN NOT CR
Default CR algorithm CR	691085 (0.078%)	187722 (0.021%)
Default CR algorithm NOT CR	66997 (0.008%)	879956340 (99.893%)
Purity: 0.912	Completeness: 0.786	F1 score: 0.844

Table 13: Pixel-wise comparison of the CR mask produced by the Default CR algorithm on science images and Cosmic-CoNN on science images.

The CR detection masks produced by the algorithm was compared to the CR detection masks produced by the Default CR algorithm on science images. Pixels that are detected by both methods as well as those that are detected by only one method are closely examined. Out of the total 945804 pixels that are detected by both methods, 7% are detected only with the Cosmic-CoNN, 19.9% are detected only with Default CR algorithm on science images, and 73.1% are detected on both. An example for a single detector is shown in figure 35.



(a) Whole detector 42.



(b) Zoomed in center part of the detector 42.

Figure 35: Comparison of the CR masks of different methods on detector 42 visit 11710. Some pixels that are obviously part of the CR that Cosmic-CoNN detected, were not masked as CR. This shows that this approach can yield inconsistent results which depends upon training data. Further training is required.

Runtime of Cosmic-CoNN took about 150 sec per detector, run on CPU. Deep models are slow while running on the CPU, and GPU run should improve runtime performance. Also deep model should, in theory, be able to run multiple detector images in parallel, but we were not able to produce this behaviour.
Mask produced by the Cosmic-CoNN was used to determine CR and NOT CR labels for the DIASources described by the process in section 2.1.4, and compared to the same process using the Default CR algorithm on science images. Match between the two algorithms is high with $F1 = 0.988$. Cosmic-CoNN determined that 59.83% of the DIASources are CR-s while 40.17% are NON CR. Details are listed in table 14

	Cosmic-CoNN CR	Cosmic-CoNN NOT CR
Default CR algorithm CR	10963 (59.77%)	248 (1.35%)
Default CR algorithm NOT CR	11 (0.06%)	7121 (38.82%)
Purity: 0.999	Completeness: 0.978	F1 score: 0.988

Table 14: Stamp-wise comparison of the CR mask produced by the Default CR algorithm on science images and Cosmic-CoNN on science images.

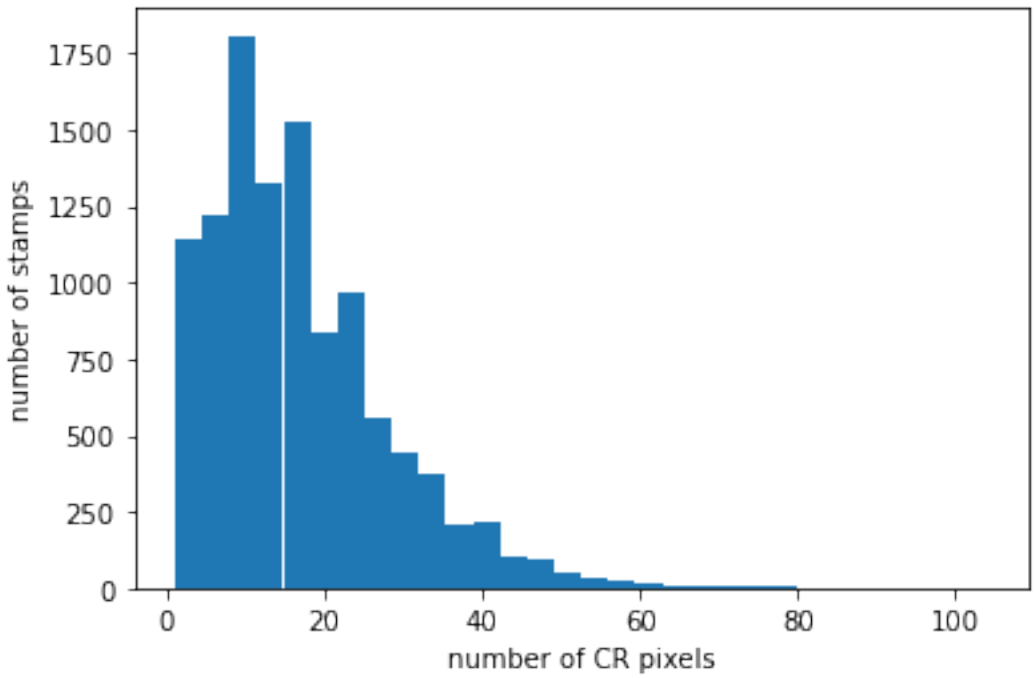


Figure 36: Histogram of the distribution of stamps with different number of CR pixels detected with the Cosmic-CoNN algorithm on science images. Median number is 14 CR pixels per stamp. Stamps with 0 CR pixels are not taken into consideration.

With the median of 14 CR pixel per one cosmic ray particle, as determined on the figure 36, CR flux is:

$$j = 0.091 \frac{CR}{cm^2 \cdot s}$$

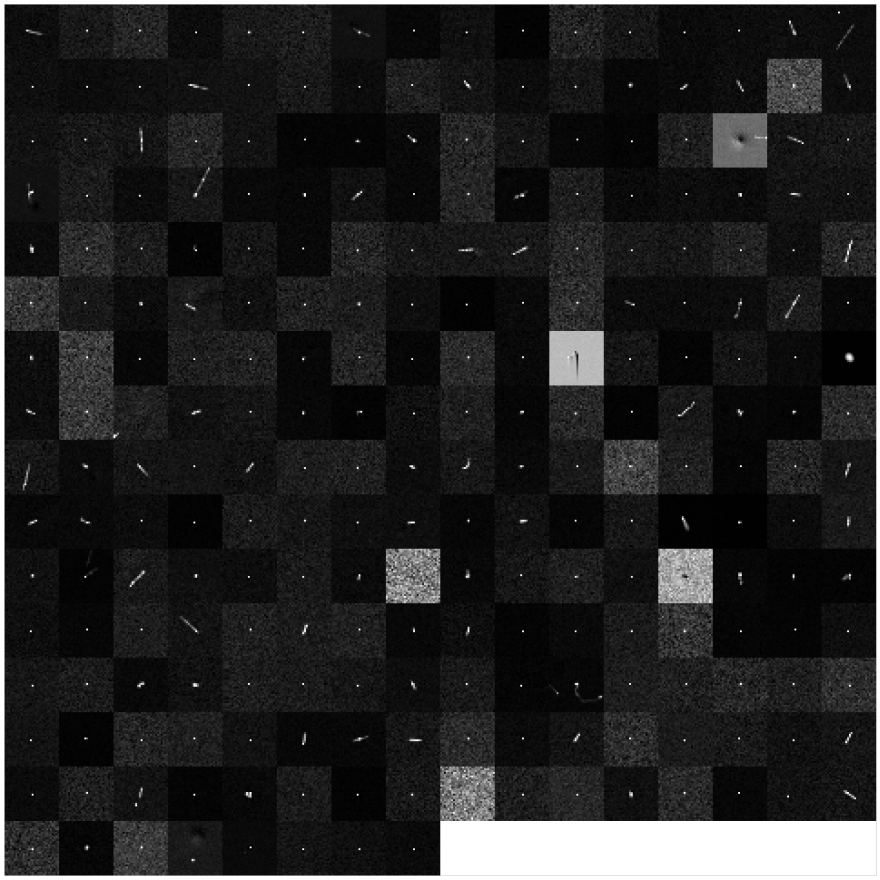


Figure 37: All of the stamps that the Default CR algorithm on science images detected as CR, and the Cosmic-CoNN on science images detected as NOT CR. Cosmic-CoNN has troubles with the detection of single pixel CR, and misses multiple cases of CRs. The machine learning is highly dependent on the training set, so further fine-tuning is required.

Images that are recognised by the Default CR algorithm on science images as CR, and by the Cosmic-CoNN on science images as NOT CR are shown in figure 37. It is clear that on most of those conflicting stamps are CR-s, but a few examples of the real and valuable DIASources can be found. Those would usually be interpolated in Default CR algorithm as CR.
Images that are recognised by the Default CR algorithm on science images as NOT CR, and by the Cosmic-CoNN on science images as CR are shown in figure 38. Only 11 of those examples are found, and while some of them is a CR, others are not entirely clear what they contain.

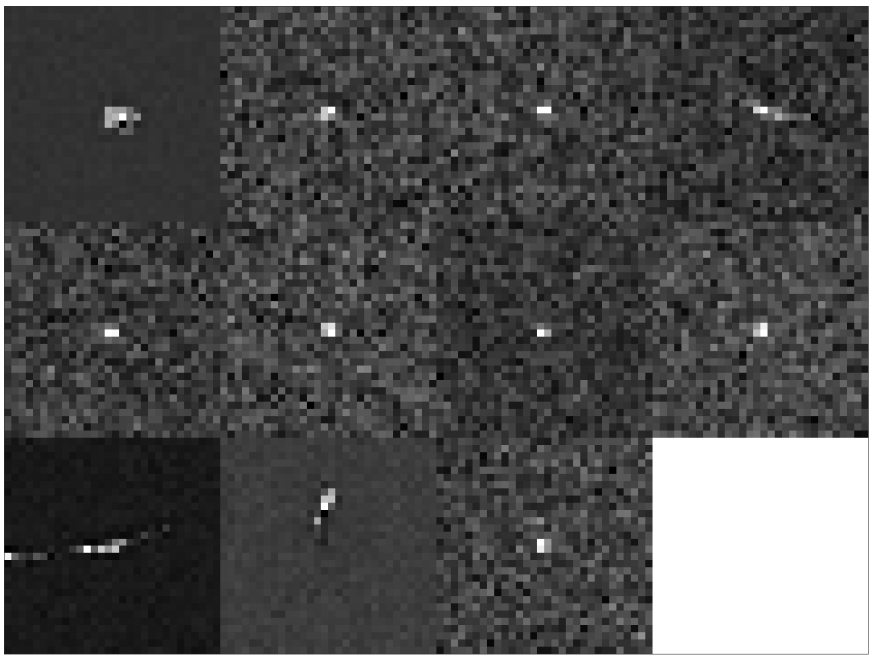


Figure 38: All of the stamps that the Default CR algorithm on science images detected as NOT CR, and the Cosmic-CoNN on science images detected as CR. Cosmic-CoNN detects only a few CR-s that have not been detected by the Default CR algorithm. Good news there are no false positive defections which shows great potential for this algorithm.

3.1.2 Difference images

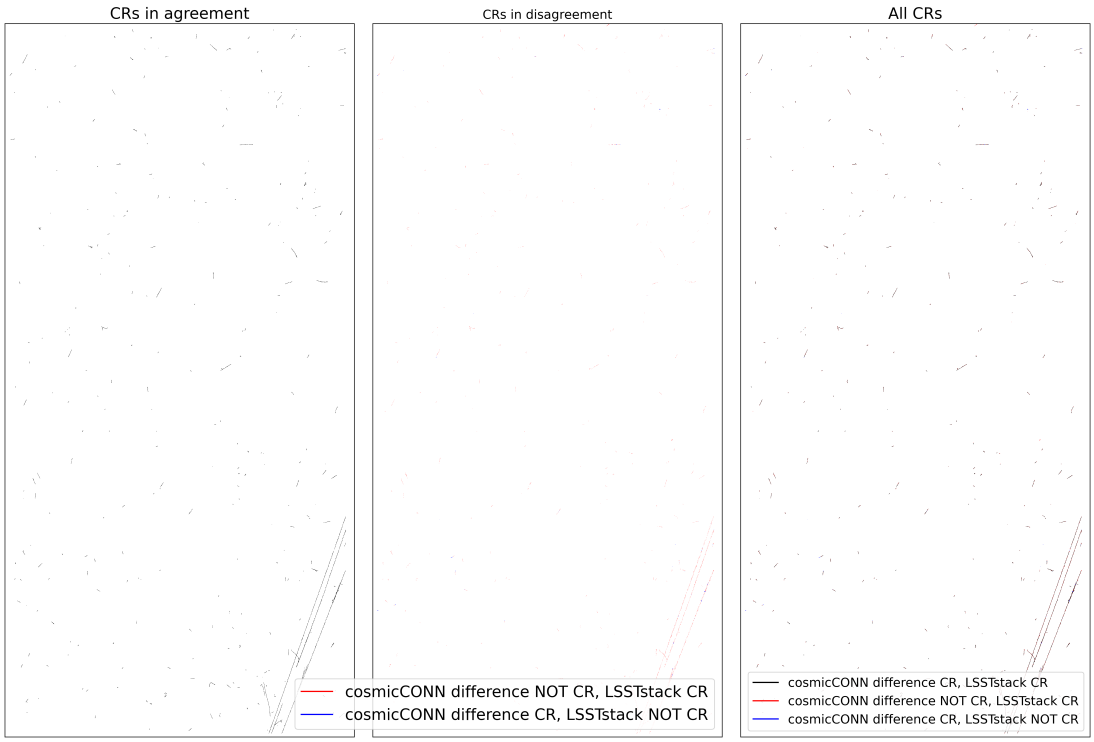
Although Cosmic-CoNN was trained on the images that are most similar to science images, using the algorithm on the difference images was evaluated. The CR-s should be able to get detected easier on the difference images, and for that reason the algorithm was tested. It detected 689245 pixels as CR, compared to Default CR algorithm on science images 878807 pixels and 506112 pixels on difference images.

$$CR = 0.078\% = 782.4\, CR\, pix/M\, pix$$

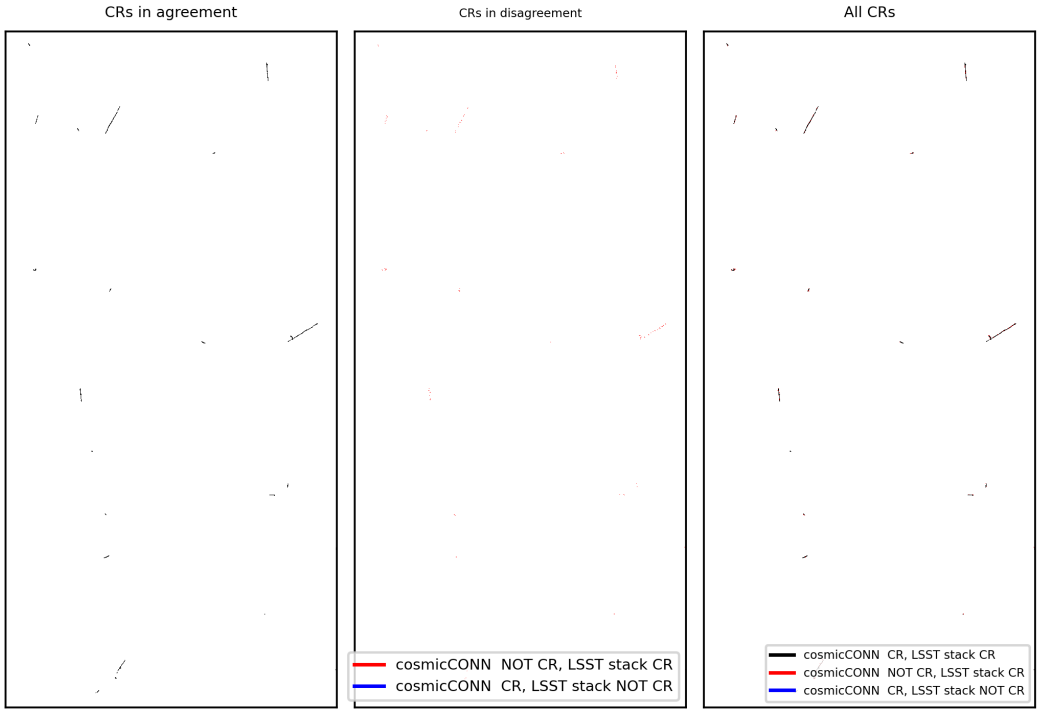
The number of the CR detected is somewhat lower compared to Cosmic-CoNN on science images ($860.6\, CR\, pix/M\, pix$) and Default CR algorithm on the science images ($997.6\, CR\, pix/M\, pix$), but little higher than Default CR algorithm on the difference images ($755.5\, CR\, pix/M\, pix$). This could be attributed to both algorithms not being fully optimized to be run on the difference images. Another explanation could be that most of the CR detection on the science images with both algorithms are false positives, but after examining the stamp images this was deemed unlikely.

	Cosmic-CoNN CR	Cosmic-CoNN NOT CR
Default CR algorithm CR	637950 (0.072%)	240857 (0.027%)
Default CR algorithm NOT CR	51295 (0.006%)	879972042 (99.894%)
Purity: 0.926	Completeness: 0.726	F1 score: 0.814

Table 15: Pixel-wise comparison of the CR mask produced by the Default CR algorithm on science images and Cosmic-CoNN on difference images.



(a) Whole detector 42.



(b) Zoomed in center part of the detector 42.

Figure 39: Comparison of the CR masks of different methods on detector 42 visit 11710. No significant improvement over the Cosmic-CoNN on the science images (figure 35) is made.

Comparison of the Cosmic-CoNN algorithm on the difference images with the Default CR algorithm algorithm on the science images is shown in table 15. The agreement value between the algorithm CR pixel mask is $F1 = 0.844$. From the 945804 CR pixels detected by either one of the algorithms, 70.1% were detected by both. 19.8% were detected only using the Default CR algorithm on science images, and 7.1% using only Cosmic-CoNN on difference images. An example for a single detector is shown in figure 39.

	Cosmic-CoNN CR	Cosmic-CoNN NOT CR
Default CR algorithm CR	11022 (60.09%)	189 (1.03%)
Default CR algorithm NOT CR	126 (0.68%)	7006 (38.19%)
Purity: 0.989	Completeness: 0.983	F1 score: 0.986

Table 16: Stamp-wise comparison of the CR mask produced by the Default CR algorithm on science images and Cosmic-CoNN on difference images.

Mask produced by the Cosmic-CoNN was used to determine CR and NOT CR labels for the DIASources described by the process in section 2.1.4, and compared to the same process using the Default CR algorithm on science images. Cosmic-CoNN determined that 60.78% of the DIASources are CR-s while 39.22% are NON CR. The match between the Cosmic-CoNN on difference images and the Default CR algorithm on science images is high with $F1 = 0.986$. Comparing that to the Cosmic-CoNN on difference images match of $F1 = 0.988$, we conclude that results are very similar. Cosmic-CoNN on difference images (Purity: 0.989 Completeness: 0.983) has lower purity and higher completeness than on science images (Purity: 0.999 Completeness: 0.978). Details are listed in table 16

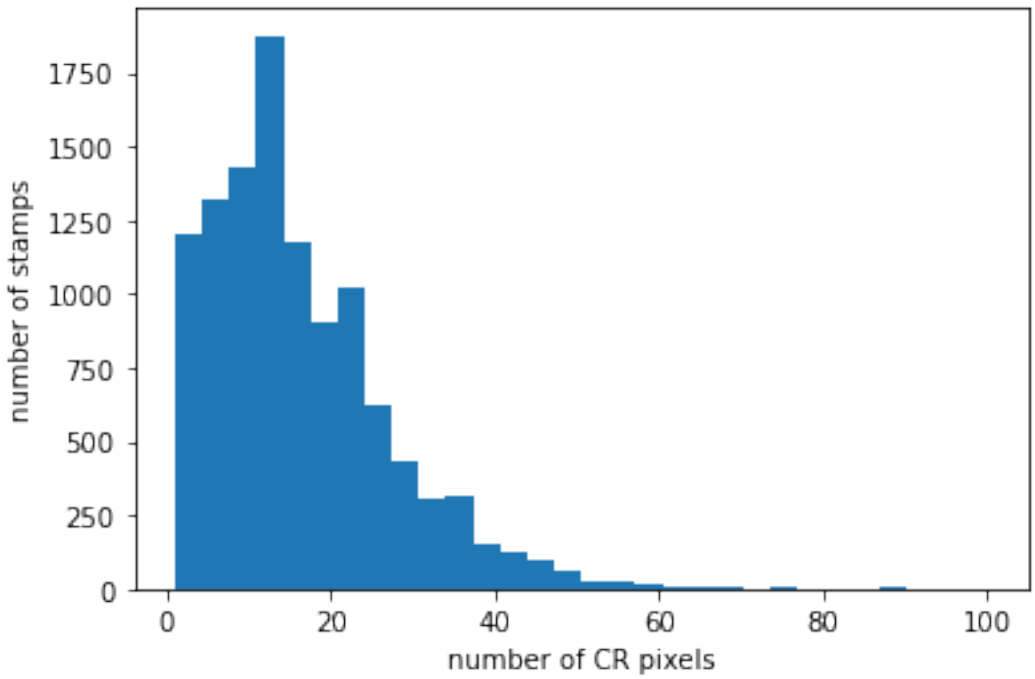


Figure 40: Histogram of the distribution of stamps with different number of CR pixels detected with the Cosmic-CoNN algorithm on difference images. Median number is 14 CR pixels per stamp. Stamps with 0 CR pixels are not taken into consideration.

With the median of 14 CR pixel per one cosmic ray particle, as determined on the figure 40, CR flux is:

$$j = 0.083 \frac{CR}{cm^2 \cdot s}$$

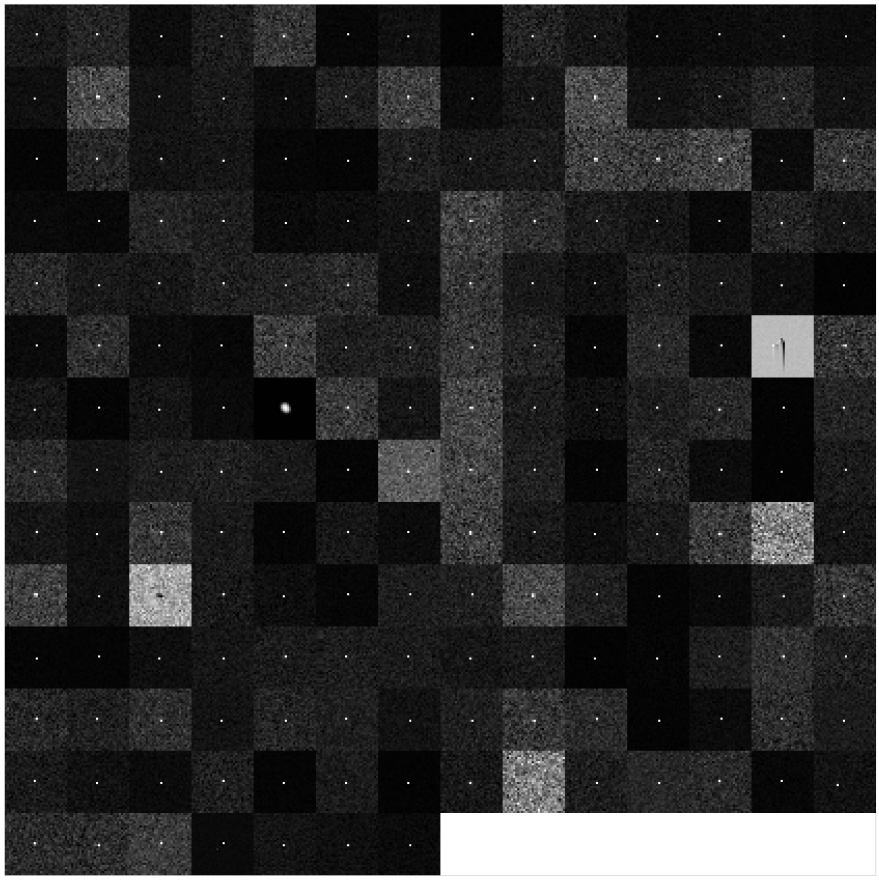


Figure 41: All of the stamps that the Default CR algorithm on science images detected as CR, and the Cosmic-CoNN on difference images detected as NOT CR. Cosmic-CoNN doesn't detect multiple cases of single pixel CRs. This type of stamps have to be included in the further training set of the Cosmic-CoNN

Images that are recognized by the Default CR algorithm on science images as CR, and by the Cosmic-CoNN on difference images as NOT CR are shown in figure 41. Some stamps contain objects that are clearly not a consequence of a CR particle, but the Default CR algorithm detected it as CR. On the other hand, there are a lot of stamps that contain a few pixels CR that were not detected by the Cosmic-CoNN. If there could be found a way to improve detection for the few pixel CRs, this would yield great improvement to CR detection.
Images that are recognised by the Default CR algorithm on science images as NOT CR, and by the Cosmic-CoNN on difference images as CR are shown in figure 42. It is obvious that Cosmic-CoNN is struggling with the dipole objects because these kind of objects were not present in the training dataset. Besides that, there are some examples of CRs that were detected correctly.

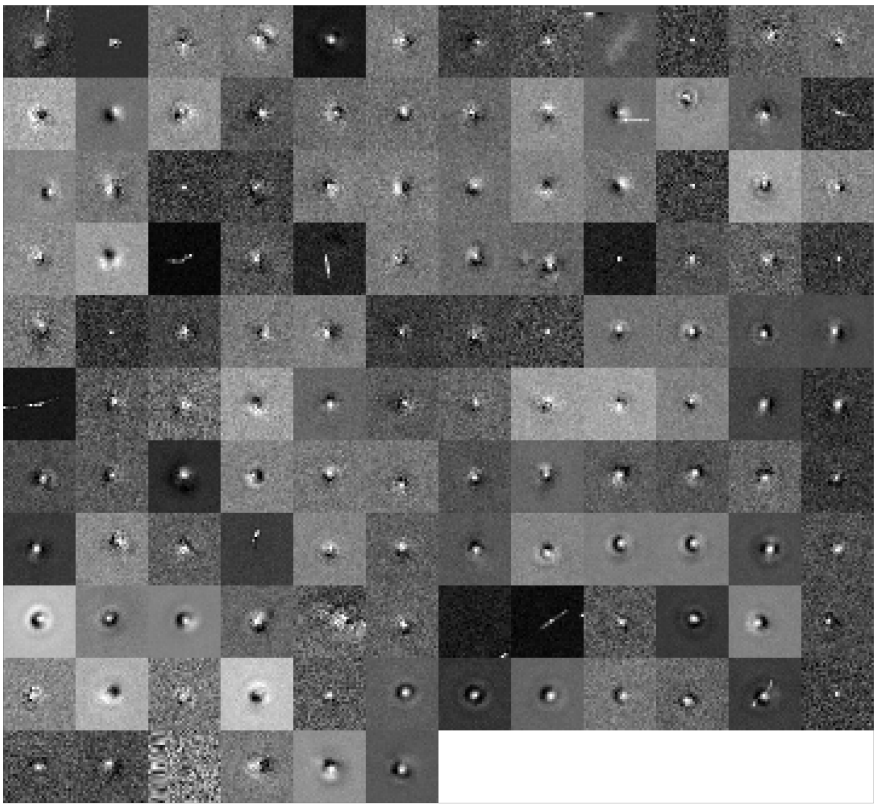


Figure 42: All of the stamps that the Default CR algorithm on science images detected as NOT CR, and the Cosmic-CoNN on difference images detected as CR. Cosmic-CoNN detects lots of dipoles as CR. Because Cosmic-CoNN wasn't created for the difference images, dipoles were not included in its training set.

Agreement between the Cosmic-CoNN run on the science images and the difference images is good, which is shown on the tables 17 and 18. This is showing that there is no major change between using the algorithm on science images or the difference images.

	science image CR	science image NOT CR
difference image CR	665545 (0.076%)	23700 (0.003%)
difference image NOT CR	92537 (0.011%)	880120362 (99.912%)
Purity: 0.878	Completeness: 0.966	F1 score: 0.920

Table 17: Pixel-wise comparison of the CR mask produced by the Cosmic-CoNN on science images and Cosmic-CoNN on difference images.

	science image CR	science image NOT CR
difference image CR	10939 (59.64%)	209 (1.14%)
difference image NOT CR	35 (0.19%)	7160 (39.03%)
Purity: 0.997	Completeness: 0.981	F1 score: 0.989

Table 18: Stamp-wise comparison of the CR mask produced by the Cosmic-CoNN on science images and Cosmic-CoNN on difference images.

3.1.3 Conclusion

Using the deep convolutional supervised learning proved to be successful. Without any major disagreement between Cosmic-CoNN and Default CR algorithm we conclude that most of the CRs are detected by Cosmic-CoNN. In order to improve that, some fine tuning and additional training is required to provide the model with the specific inside on the HSC (or in future LSST) data. For that, the creation of the training dataset is required, which can be tedious work.

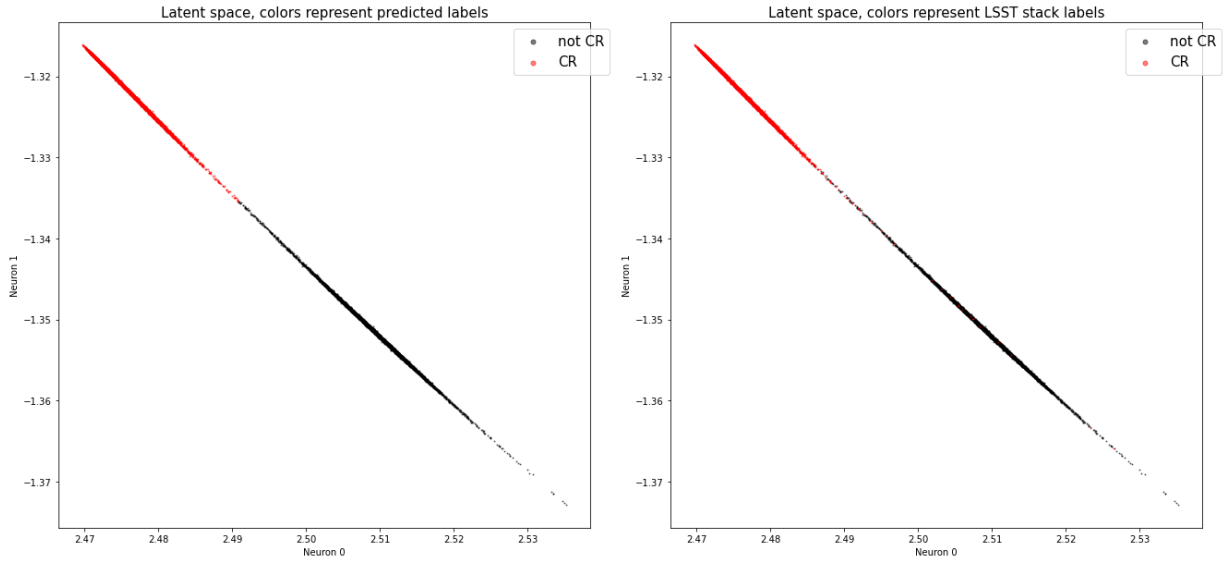
3.2 Unsupervised Learning

Unsupervised learning in artificial intelligence is the term used for the algorithms that do not use labeled data to learn. This allows the model to work on its own to discover patterns and information that was previously undetected. This independence on label data allows this types of algorithms to be used to be used on the problems in which the labels are hard or impossible to obtain. The Default CR algorithm created more that 18 000 stamps which, more than a half, contain cosmic rays. To manually classify that amount of stamps is laborious and unreliable, because in some occasions it's hard for humans to tell whether a stamp contain CR or not. For that reason unsupervised machine learning was used in a hope to find such model, which will, without any knowledge about the problem we are solving, successfully separate stamps into two groups: CR and NOT CR. Types of models that were used in this search were improved deep embedded clustering models, which belong to the broader category of convolutional autoencoder models. This types of models learn how to recreate the stamps by using dimensionality reduction followed by the dimensionality expansion. Between dimensionality reduction and expansion is a layer of neurons called "latent layer" which contains the representation of the stamps with far less bits than $30x30$ pixels stamps. Idea is that by performing this operation only relevant information about the stamp is contained in the latent layer. The values of neurons in latent layers represent points in latent space, which is then basis for the clustering algorithm. More information about this type of neural networks is in: "Guo, Gao *et. al*: Improved Deep Embedded Clustering with Local Structure Preservation"
In order to find models that can represent stamps in such way in the parameter space, so it can easily be clustered is done by hyperparameter search. Detailed procedure was taken from: "Mrakovčić: Automated Classification of LSST Images Using Convolutional Neural Networks". Changing the hiperparameters, networks were trained and evaluated by the ability to represent stamps in latent space in such way that it could be easily clustered into two clusters: one with majority CR labels and one with the majority NON CR. While the CR and NON CR labels were taken from the Default CR algorithm, please note that network was not trained on them. Labels were only used to evaluate which network represents stamps in the most convenient way for this problem. Out of the 1000 models, one with the highest F1 score on the validation set compared to the Default CR algorithm labels was chosen. Results of the labels produced by this model is in table 19.

	LAICA CR	LAICA NOT CR
Default CR algorithm CR	11185 (60.98%)	26 (0.14%)
Default CR algorithm NOT CR	57 (0.31%)	7075 (38.57%)
Purity: 0.995	Completeness: 0.998	F1 score: 0.996

Table 19: Stamp-wise comparison of the CR mask produced by the Default CR algorithm on science stamps and LAICA on difference stamps.

This model, which is called LAICA model, in the latent layer contains 2 neurons. In other words it is encoding every stamp in 2D space. Plotting those encodings of stamps is shown on figure 43. Curios observation is made: almost every point in latent space lie on the single line. This means that although 2 neurons spread 2D latent space, representation of the stamps could be represented with only one variable. Models that have only one neuron in the latent layer also performed very well, therefore, the stamps can be encoded with only one variable.



(a) Predicted labels

(b) Default CR algorithm labels

Figure 43: Parameter space of the LAICA model. The points are nicely clustered together and all points lie on the almost straight line. This means that the Autoencoder CNN can probably use single parameter to represent the stamp. By creating this way of representation it has decided to place CR and NOT CR stamps on the opposite side of the latent space. This fact was used by the unsupervised clustering to predict labels.

Figure 44 shows the same representation, only showing the stamps that were not used for training, but only for validation. Distribution of validation set on the latent space follows closely distribution of the training data. In other words no overfitting was detected.

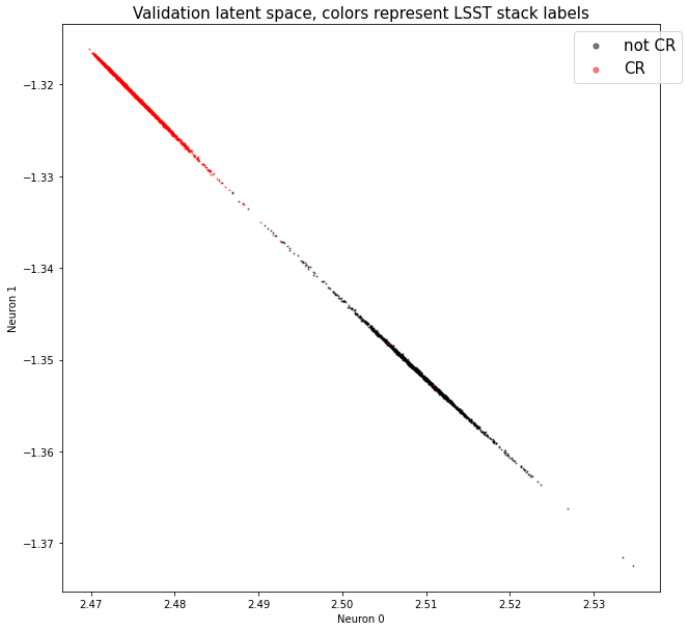
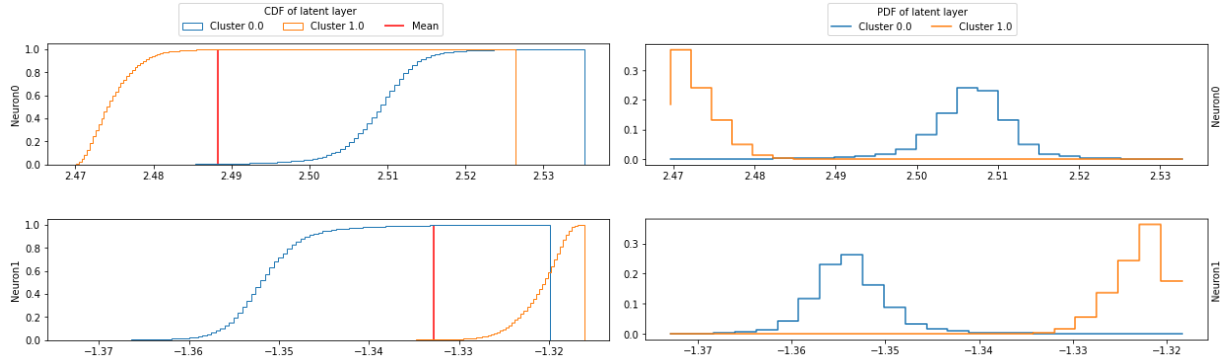


Figure 44: Parameter space of the LAICA model for only stamps that were used in the validation set. Distribution does not diverge from the figure 43. This means it can confidently predict yet unseen data, and is not restricted on using only the training set.

It is clear that model successfully created parameter space in which CR stamps and NON CR stamps are clustered together. The PDF and CDF for values of each neuron is shown on figure 45. This proves that the model successfully learned in unsupervised way to distinguish stamps that look like CR and the ones that look like NON CR.



(a) Predicted labels

(b) Default CR algorithm labels

Figure 45: Distribution of the values of the neurons in the latent layers. Cluster 0 is NOT CR and cluster 1 is CR. Clusters are well separated. Reason why this model creates representation of the stamps in a way that it could be easily separated between CR and NOT CR is by design. By trying 1000 different models we choose one that does this the best.

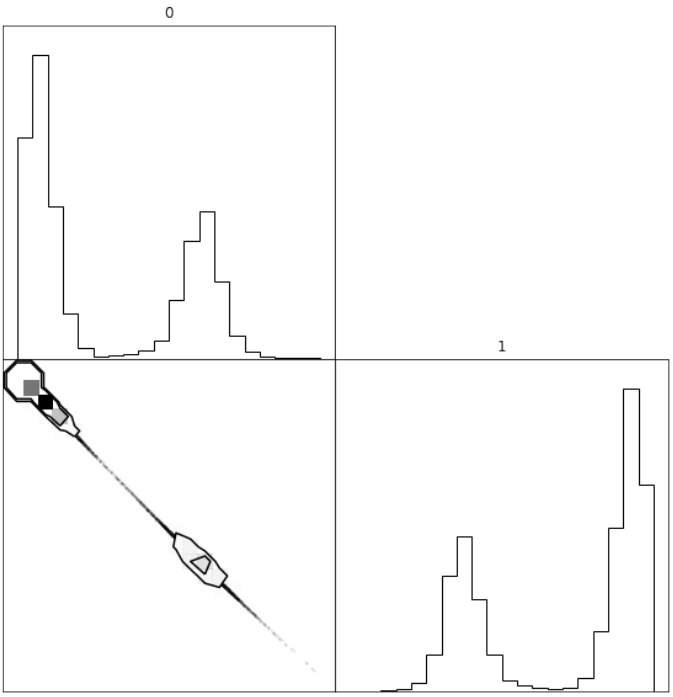


Figure 46: Corner plot of the distribution of the stamps in the latent space. 0 and 1 represent distributions for the neuron 0 and neuron 1 respectively. Separation of the latent layer creates a confident clustering of two classes.

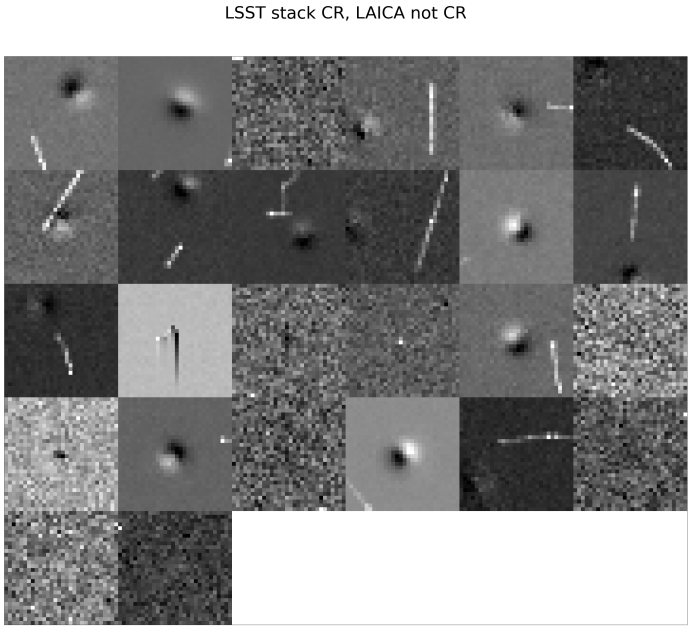


Figure 47: All of the stamps that the Default CR algorithm on science images detected as CR, and the LAICA detected as NOT CR. LAICA labels stamps with both objects and CRs as NOT CR. This can't be improved by further training, and should be investigated further.

Images that are recognized by the Default CR algorithm on science images as CR, and by the

LAICA model as NOT CR are shown in figure 47. Most of the stamps contain both objects and CR-s. LAICA prefers to label stamps that contain both object and CR as NOT CR. Additionally, stamps with high SNR are also labeled as NOT CR. Stamps which Default CR algorithm mislabeled as CR are scarce.
Images that are recognised by the Default CR algorithm on science images as NOT CR, and by the LAICA model as CR are shown in figure 48. LAICA detects stamps with "banana" shapes as CR, as well as some very bright dipoles. In addition, some true CR can be seen labeled as CR by LAICA.

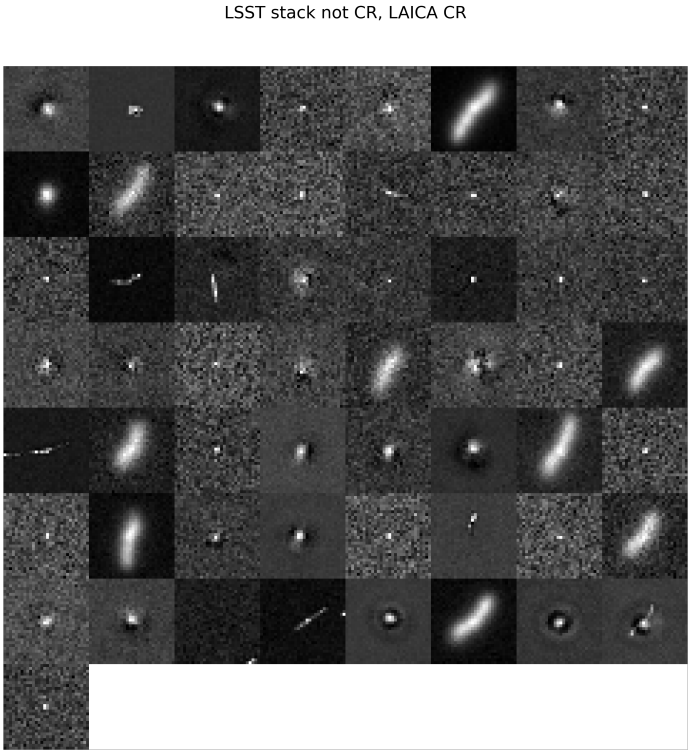


Figure 48: All of the stamps that the Default CR algorithm on science images detected as NOT CR, and the LAICA detected as CR. LAICA detected a few stamps as CR that look like artefact. Possiblity of using more than two labels while clustering is a possibility that could solve this type of mislabeling.

It is important to stress that this algorithm does not perform the same task as the others that were examined. Every other algorithm was tested with the whole detector panel and asked to create a mask where the CR pixels are. The mask was then used to determine label of each stamp. Contrariwise, this algorithm has no ability to produce CR mask and it can only predict labels for stamps. While the results are not perfect, and the unsupervised algorithm is mislabeling some stamps, it is a neat way to create somewhat trustworthy dataset of stamps and their labels. More research is needed to improve this approach.

4 Manual examination

To create a dataset of labeled stamps that contain both CR and NOT CR, manual examination effort was required. In order to avoid manually labeling the whole set, only the stamps that have

at least one algorithm's prediction differ from the others, are hand labeled. We claim that the stamps that are labeled by every algorithm consistently are correct, and the ones that aren't are examined. This way the labeled dataset with greater degree of confidence is produced. Once the "Truth" labels are produced they are used to evaluate the performance of every algorithm. 632 out of total 18343 stamps (3.45 %) were closely examined. The ratio between the CR and NON CR stamps in the hand labeled subset is shown in the table 20. Note that Default CR algorithm on difference images have smaller subset because not all of the detectors could be analysed.

	CR	NOT CR
Default CR algorithm sciim	320	312
Default CR algorithm diffim	337	254
AstroScrappy sciim	363	269
AstroScrappy diffim	482	150
Cosmic CoNN sciim	83	549
Cosmic CoNN diffim	257	375
LAICA	351	281
Truth	323	309

Table 20: Number of the labels in the subset that was manually labeled.

	LSST sciim CR	LSST sciim NOT CR
Truth CR	11182 (60.96%)	32 (0.17%)
Truth NOT CR	29 (0.16%)	7100 (38.71%)
Purity: 0.997	Completeness: 0.997	F1 score: 0.997

Table 21: Stamp-wise evaluation of the CR mask produced by the Default CR algorithm on science stamps.

	LSST diffim CR	LSST diffim NOT CR
Truth CR	10477 (60.81%)	7 (0.04%)
Truth NOT CR	69 (0.40%)	6677 (38.75%)
Purity: 0.993	Completeness: 0.999	F1 score: 0.996

Table 22: Stamp-wise evaluation of the CR mask produced by the Default CR algorithm on difference stamps.

	AstroScrappy sciim CR	AstroScrappy sciim NOT CR
Truth CR	11207 (62.00%)	7 (0.04%)
Truth NOT CR	47 (0.26%)	7082 (38.61%)
Purity: 0.996	Completeness: 0.999	F1 score: 0.998

Table 23: Stamp-wise evaluation of the CR mask produced by the AstroScrappy on science stamps. AstroScrappy parameters were taken in the mean square error minimum.

	AstroScrappy diffim CR	AstroScrappy diffim NOT CR
Truth CR	11207 (62.00%)	7 (0.04%)
Truth NOT CR	166 (0.90%)	6963 (37.96%)
Purity: 0.985	Completeness: 0.999	F1 score: 0.992

Table 24: Stamp-wise evaluation of the CR mask produced by the AstroScrappy on difference stamps. AstroScrappy parameters were taken in the mean square error minimum.

	Cosmic CoNN sciim CR	Cosmic CoNN sciim NOT CR
Truth CR	10951 (59.70%)	263 (1.43%)
Truth NOT CR	23 (0.13%)	7106 (38.74%)
Purity: 0.998	Completeness: 0.977	F1 score: 0.987

Table 25: Stamp-wise evaluation of the CR mask produced by the Cosmic CoNN on science stamps.

	Cosmic CoNN diffim CR	Cosmic CoNN diffim NOT CR
Truth CR	11017 (60.06%)	197 (1.07%)
Truth NOT CR	131 (0.71%)	6998 (38.15%)
Purity: 0.988	Completeness: 0.982	F1 score: 0.985

Table 26: Stamp-wise evaluation of the CR mask produced by the Cosmic CoNN on difference stamps.

	LAICA CR	LAICA NOT CR
Truth CR	11191 (61.01%)	23 (0.13%)
Truth NOT CR	51 (0.28%)	7078 (38.58%)
Purity: 0.995	Completeness: 0.998	F1 score: 0.997

Table 27: Stamp-wise evaluation of the CR mask produced by the LAICA on difference stamps.

Tables 21 - 27 show the evaluation of the different algorithms on compared to the manually examined and labeled dataset (called TRUTH here, which we know has high purity, and suspect also high completeness, though the latter needs to be verified using dark frame data). This gives us the information about the absolute performance of each algorithm relative to TRUTH. F1 score for every algorithm is very high as well as purity and completeness.

Every algorithm evaluated performed similarly, and no significant improvement of the Default CR algorithm was found. AstroScrappy on the science images with the parameter $sigclip = 5.0$ has the best overall performance of $F1 = 0.998$, but the long runtime means it needs to be further optimised to perform the task faster. Currently implemented CR algorithm in the LSST stack on the science images does the detection with high $F1 = 0.997$ score, meaning that with the current

implementation of AP pipe, CR contamination of alerts would likely not be significant.

These results also shows that unsupervised machine learning approach could be considered as a sufficient alternative as a CR detector with no appreciable loss of performance.

5 Conclusions

After experimenting with two conventional cosmic ray detectors and two machine learning ones on both science and difference images, the conclusion is that there are no major discrepancies between algorithms. For that reason, we believe that no major improvements are made in CR detection. Furthermore, after examining disagreements between the algorithms conclusion is made that every algorithm mislabels different types of objects. For that reason, in order to produce a sample of high purity and high completeness, we recommend using multiple algorithms. In order to avoid mislabeling specific type of object it is advised not to trust only single type of algorithm.

The questions that this investigation raised and need further research are:

- How does each of the tested algorithms perform with different seeing? Which algorithm is the most robust to change in seeing?
- How does each algorithm perform in dark frames, where every signal is a CR?
- Why does the ratio between Default CR algorithm on science image CR - algorithm NOT CR, and Default CR algorithm on science image NOT CR - algorithm CR changes drastically between pixel-wise and stamp-wise comparison for every algorithm evaluated. (recognized when comparing pairs of tables: 3-4, 5-6, 9-10, 13-14, 15-16)
- Can AstroScrappy be optimized further to be comparable to Default CR algorithm runtime of 3 seconds per detector?
- Can Cosmic-CoNN be further trained on HSC data to improve detection performance?
- What can be learned from the latent space created by unsupervised learning?