Impact of filtering on CYGNO's Experiment

Igor Fonseca Pains

Advisors: Rafael Antunes Nóbrega
Igor Abritta Costa

September 27, 2022

Summary

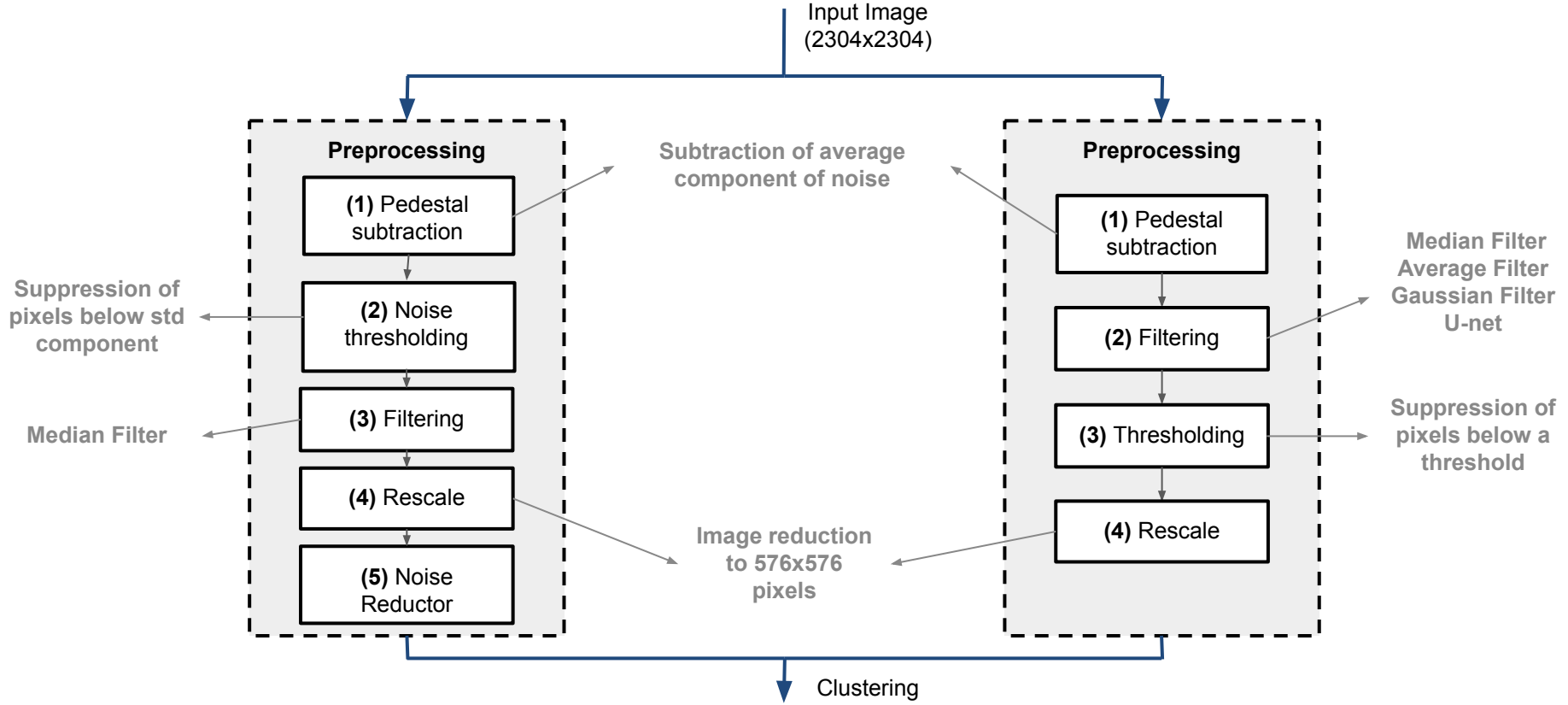
- Image filtering on CYGNO experiment
  - Motivation and goals
  - Proposal
  - Results
  - Conclusions
  - Next steps

Motivation and goals

- Motivations
  - Is it possible to improve SNR of CYGNO using filtering techniques?
  - What is their impact on detection efficiency?
  - What is their impact on energy estimation?

- Goals
  - Propose different pre-processing techniques.
  - Define a methodology to assess their performance.
  - Evaluate their impact on simulation and real data for the LIME detector.

Collaboration vs Proposal



Evaluation proposal

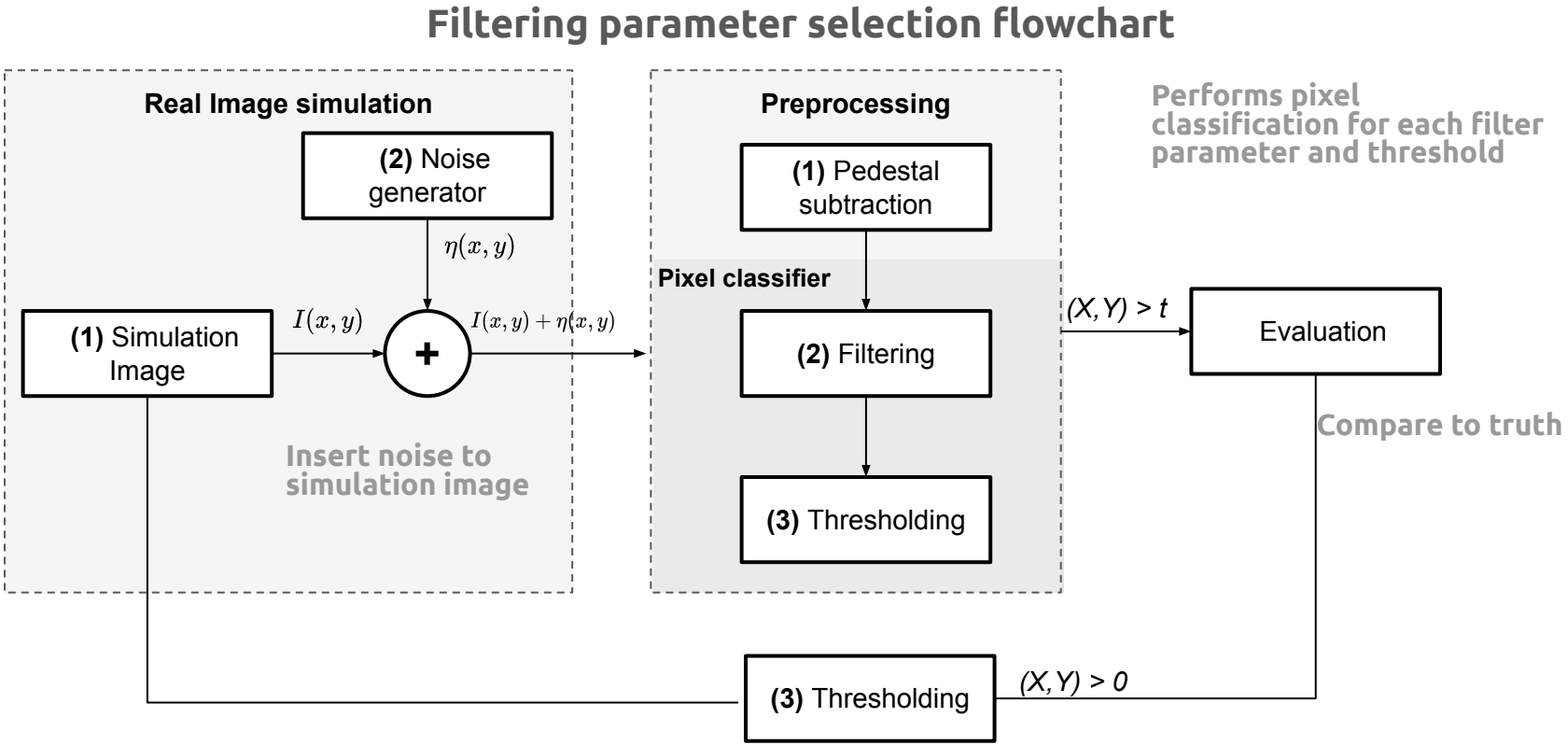
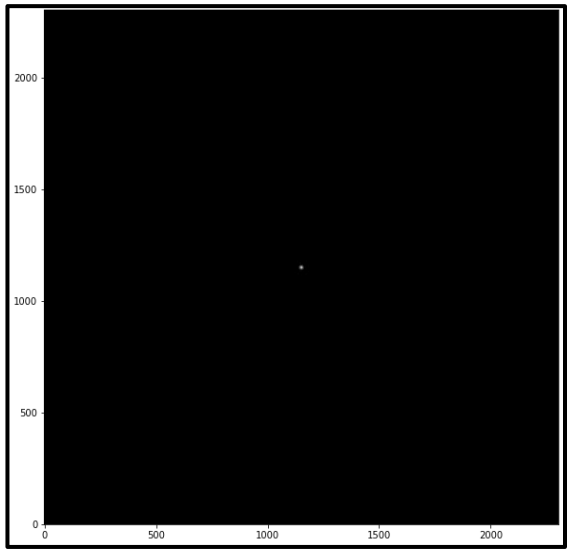
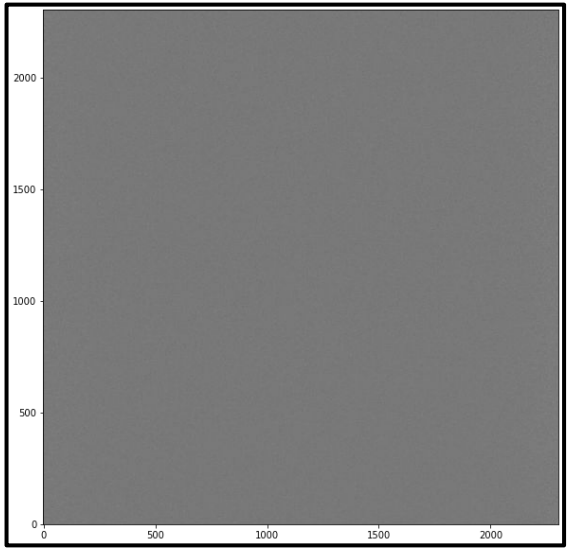


Image Simulation



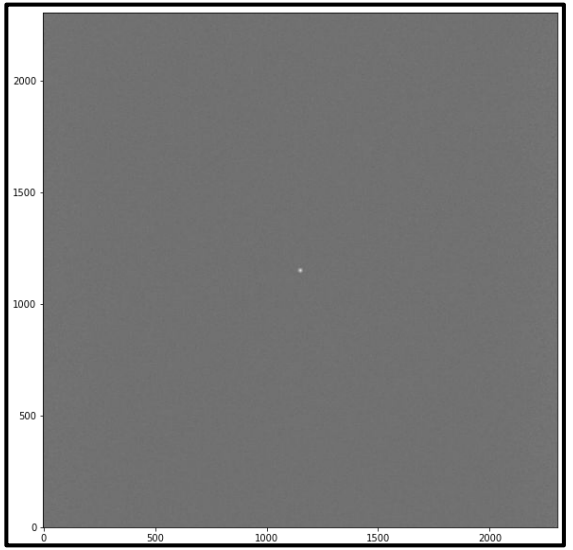
6 keV signal simulation

+



Noise simulation based on the ECDF of a noise run

=



Simulated image with signal and noise background

Filtering

Two approaches will be used:

- Spatial filtering:
  - Linear (Average and Gaussian filters);
  - Non linear (Median filter);
    - Parameter is the used mask (window size and filter type).

- Deep learning based:
  - U-Net;
    - Parameters are obtained after training process, by optimizing the layer's weights;

Evaluation

- Threshold is applied to define if a pixel is signal or background.
- The result is then compared to the truth image to access the precision, recall and f1 score through the confusion matrix.

Truth \ Predicted	Negativo	Positive
Negative	True Negative (TN)	False Positive (FN)
Positive	False Negative (FN)	True Positive (TP)

$$\textbf{Precision} = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

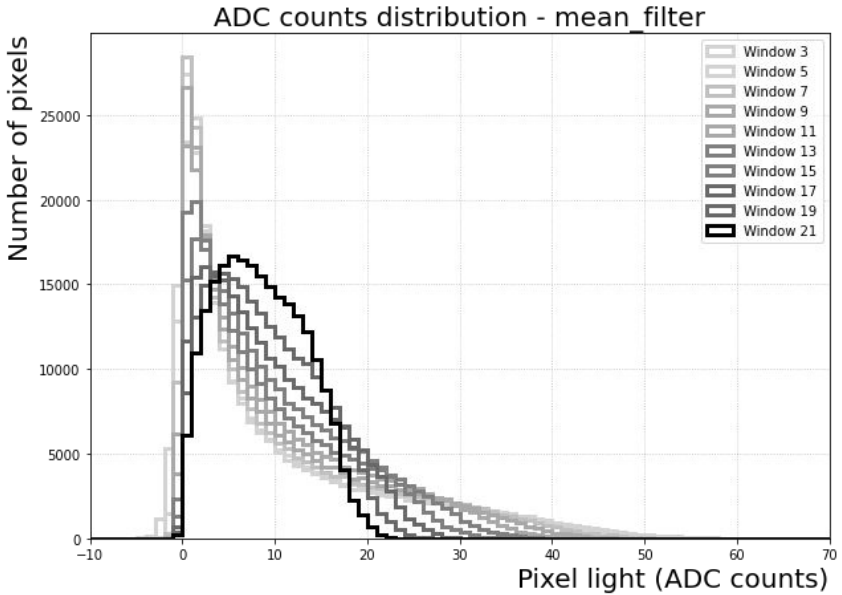
$$\textbf{Recall} = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

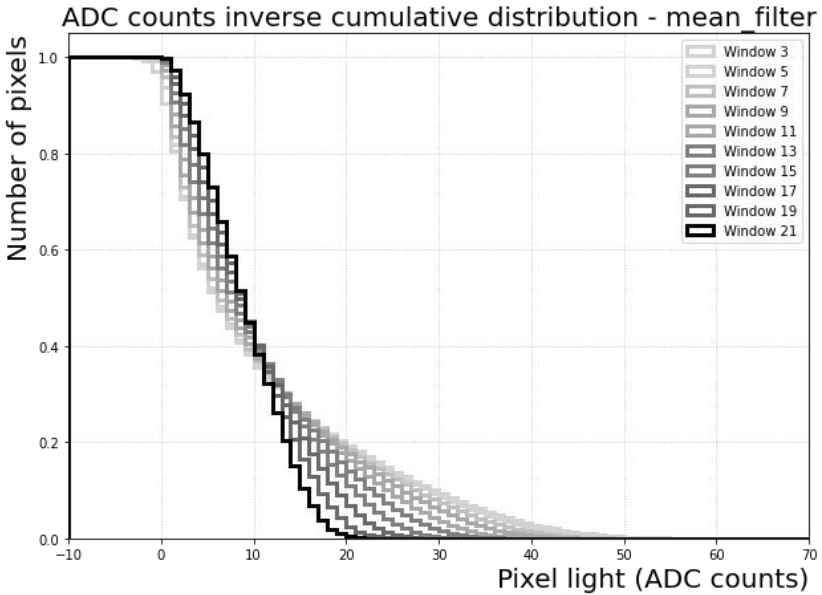
$$F1 = 2.\frac{Precision \times Recall}{Precision + Recall}$$

Optimization

- The optimization of the filter parameters *(window size and threshold)* was done through a scan aiming the maximum F1 score on the training dataset (200 images).
  - Average Filter: Completed
  - Gaussian Filter: Completed
  - Median Filter: Completed
  - U-net: Using the network trained with LEMON data. (For now)

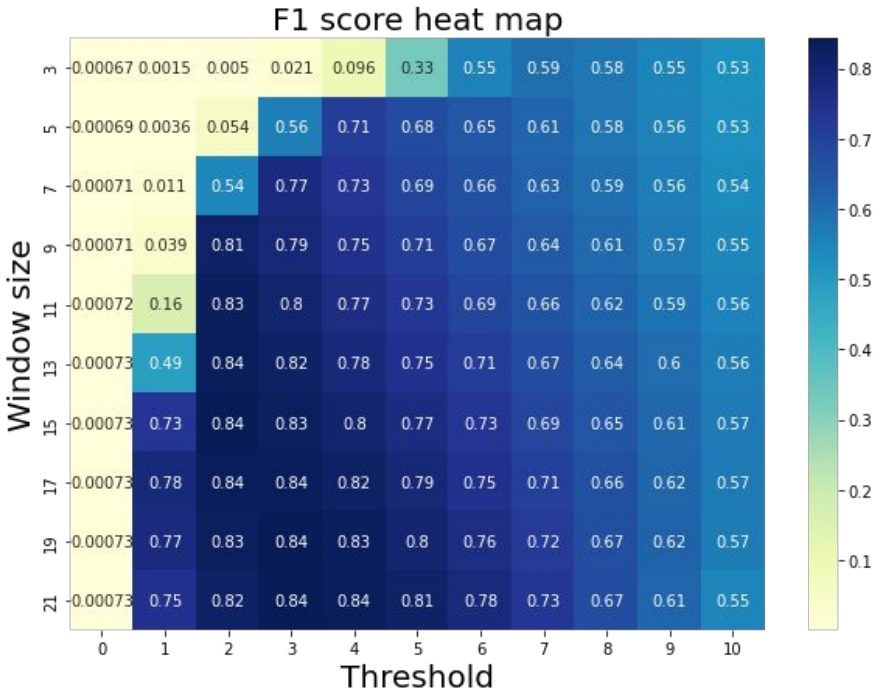
Threshold selection - Average Filter

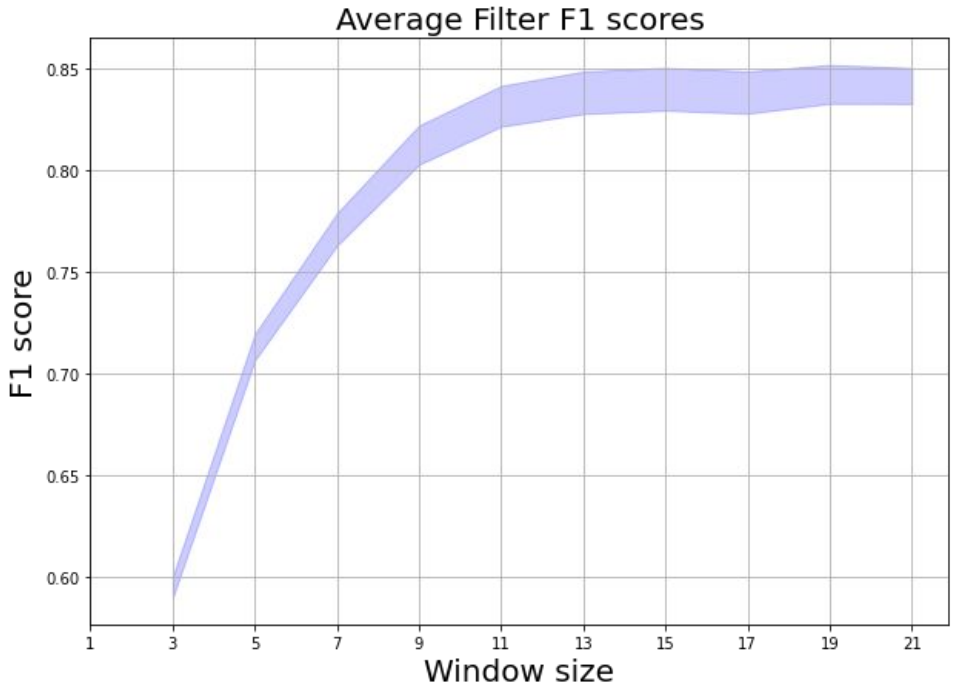




The range of possible thresholds was chosen by looking at the ADC counts distribution of signal pixels after filtering.

F1 score evaluation - Average Filter





The ideal window size seems to be around 13 whereas the threshold 2.

Optimization

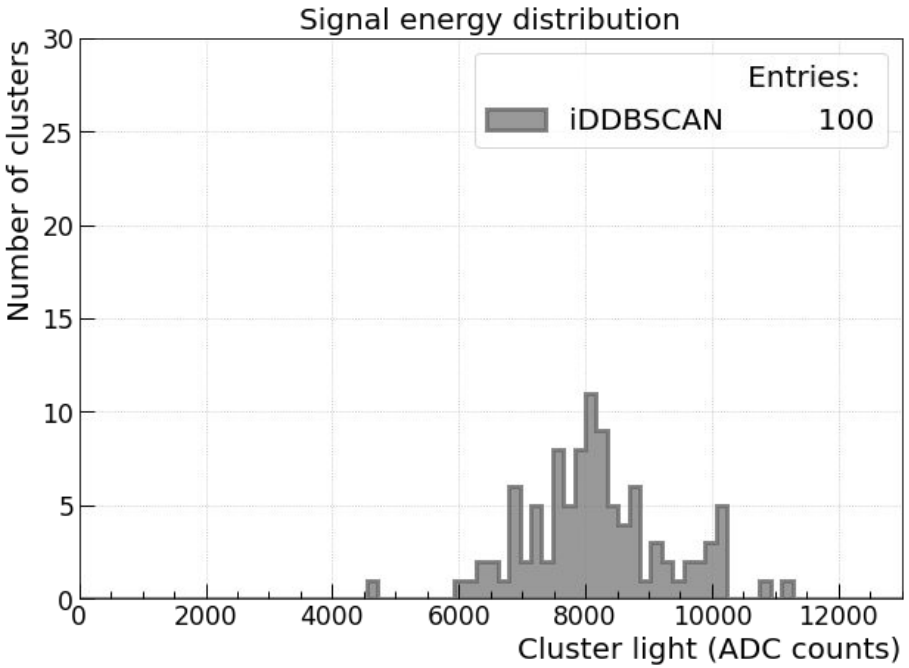
- The following parameters were chosen for each filter ([Window size, threshold]):
  - Average Filter: [13, 2]
  - Gaussian Filter: [13, 2]
  - Median Filter: [15, 2]
  - U-net: Using the network trained with LEMON data. (For now)

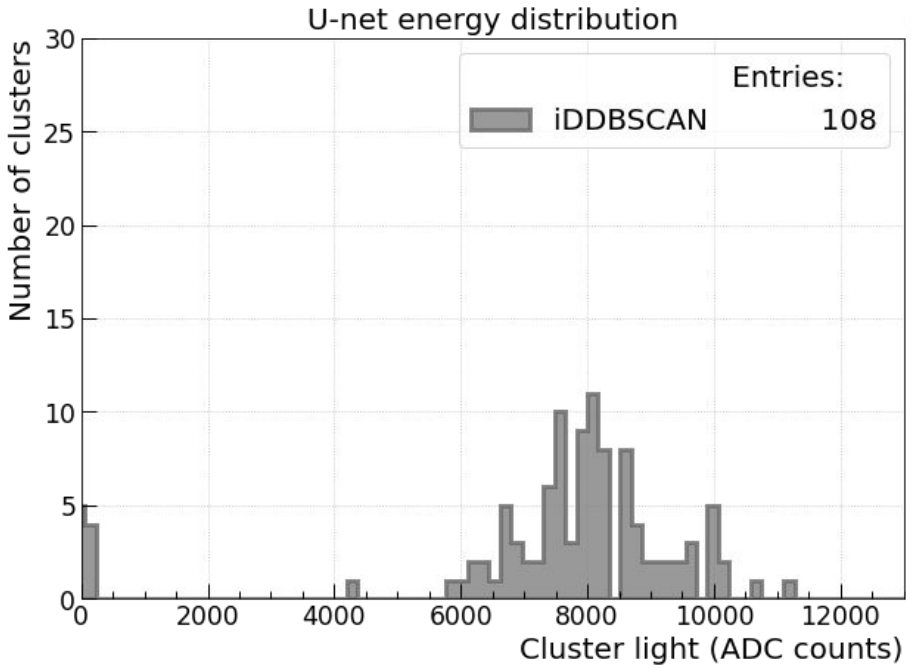
- These parameters were used on a modified version of the reconstruction algorithm in order to create reco files from the test dataset (100 images) for comparison.

Results

	Number of clusters	Precision	Recall	F1 score	Energy estimation accuracy (%)
CYGNO	295	0.8021 ± 0.0051	0.8172 ± 0.0053	0.8088 ± 0.0046	-1.26 ± 0.17
CYGNO (1.4)	106	0.8344 ± 0.0050	0.7848 ± 0.0052	0.8081 ± 0.0044	-1.92 ± 0.17
Average Filter	100	0.7991 ± 0.0045	**0.8511 ± 0.0055**	0.8237 ± 0.0046	**-0.97 ± 0.16**
Gaussian Filter	100	0.7991 ± 0.0045	**0.8509 ± 0.0056**	0.8238 ± 0.0045	**-0.92 ± 0.16**
Median Filter	100	**0.8619 ± 0.0051**	0.7726 ± 0.0048	0.8143 ± 0.0045	-2.42 ± 0.16
U-net	108	0.8063 ± 0.0045	0.8454 ± 0.0051	**0.8250 ± 0.0044**	-1.11 ± 0.16

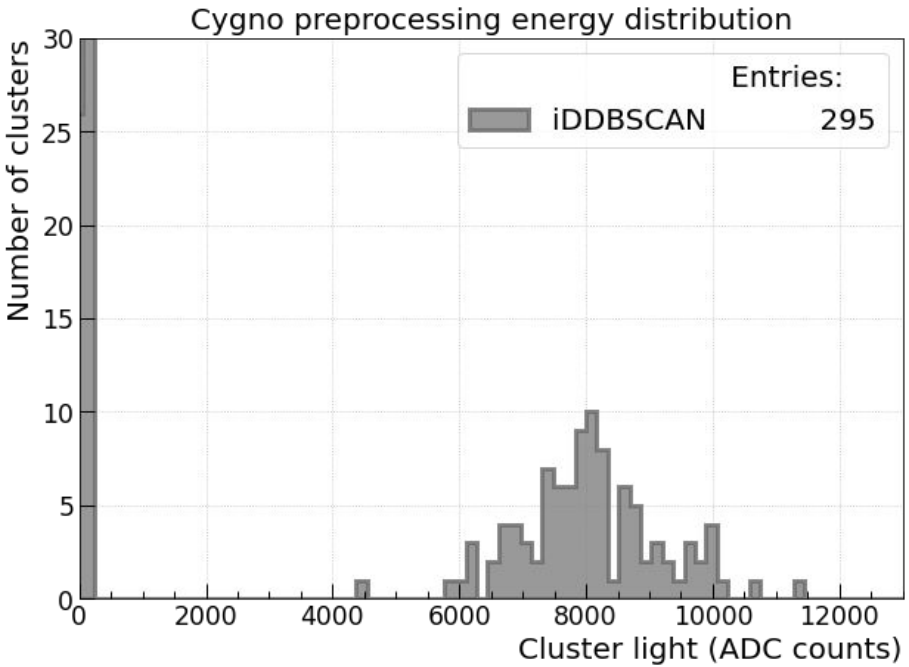
Energy distribution

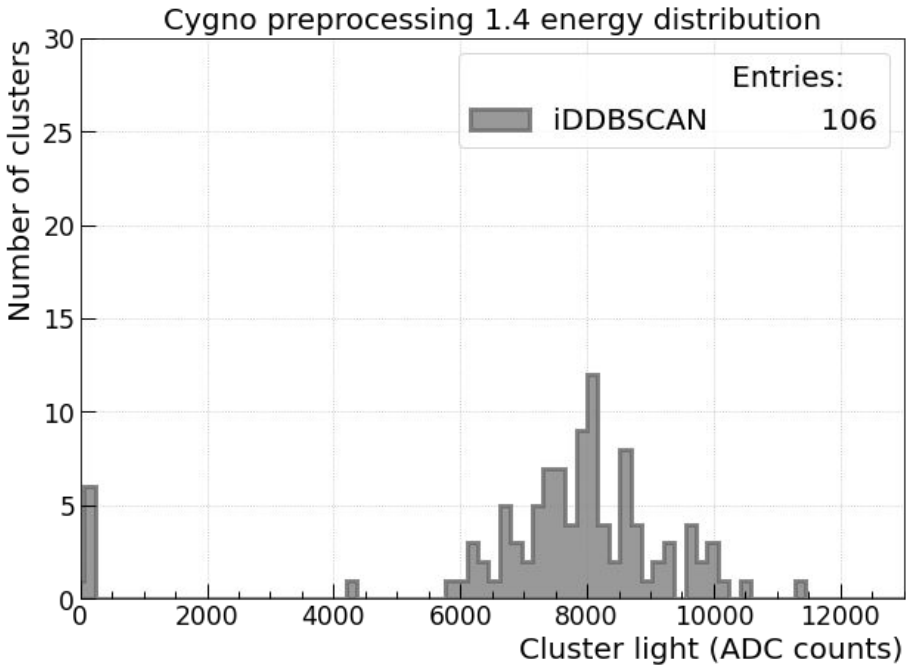




All the fake clusters are located in the very low-energy region, thus a simple energy cut would be enough to remove them.

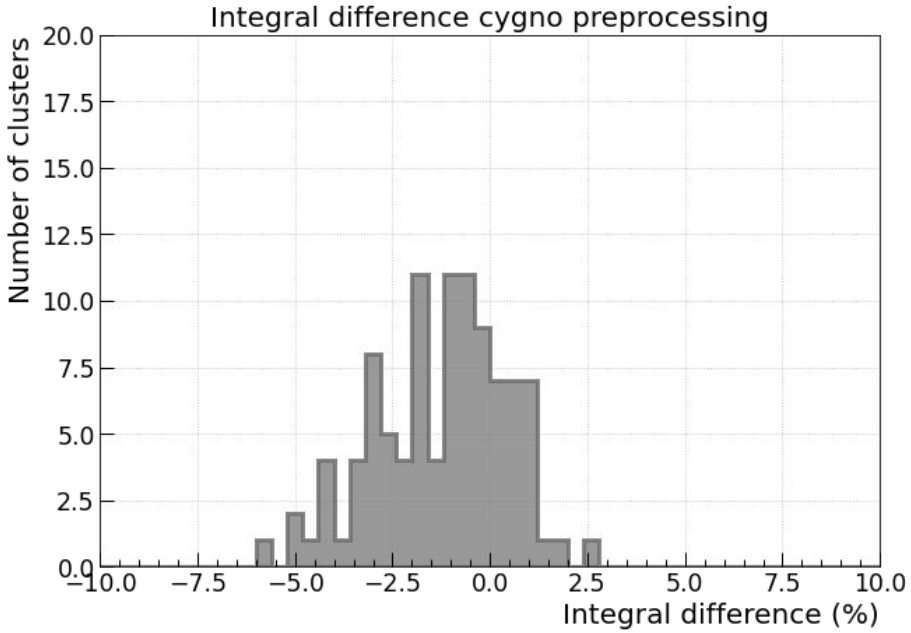
Energy distribution

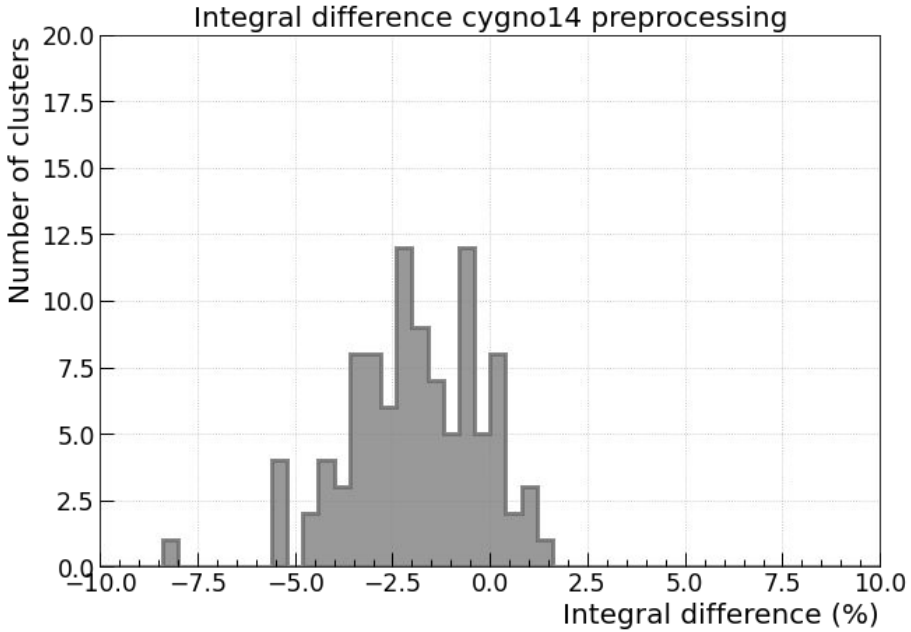




All the fake clusters are located in the very low-energy region, thus a simple energy cut would be enough to remove them.

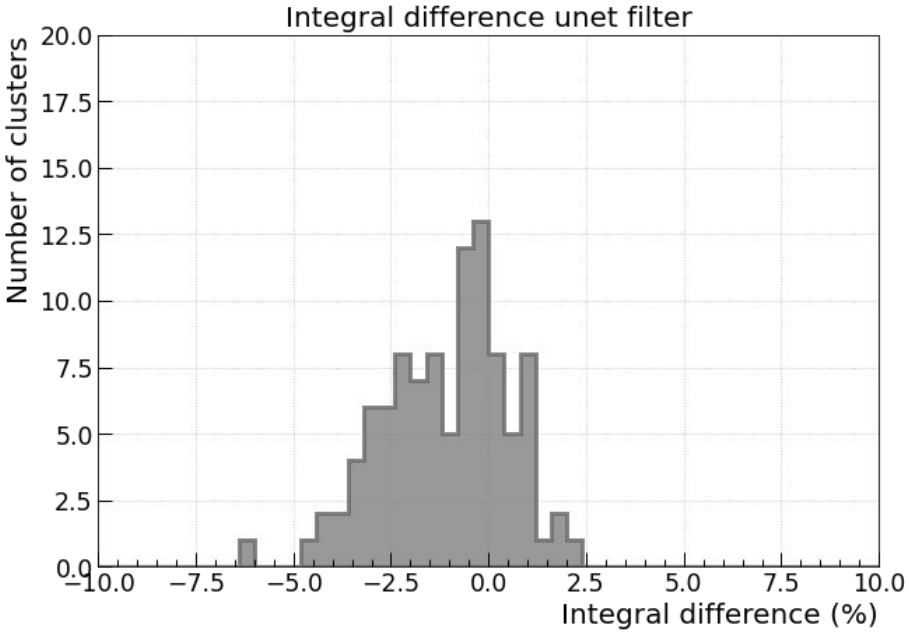
Energy estimation error

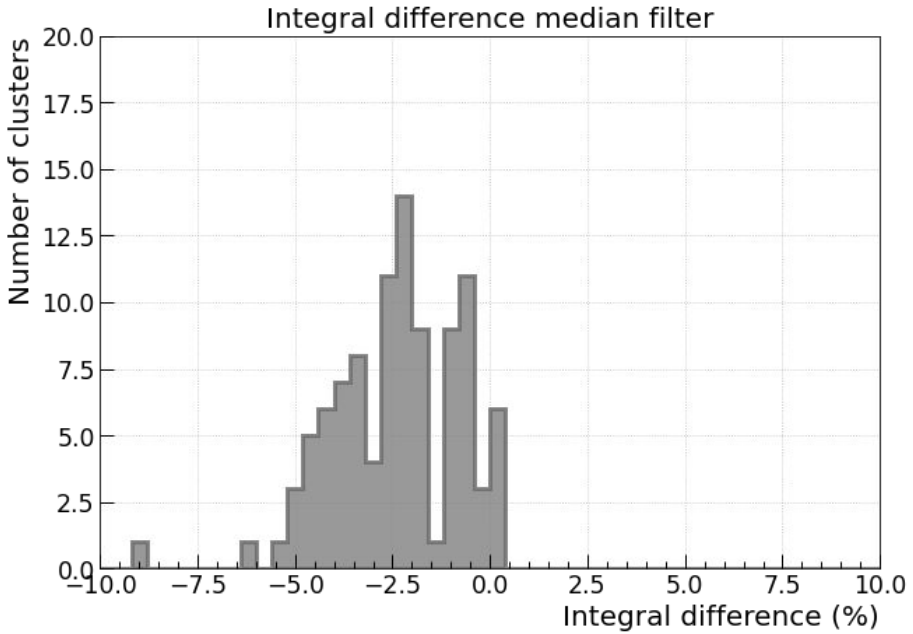




Both histograms seems pretty close, showing a trend to lose energy. Also, an increase in sigma means a bigger loss in energy reconstruction.

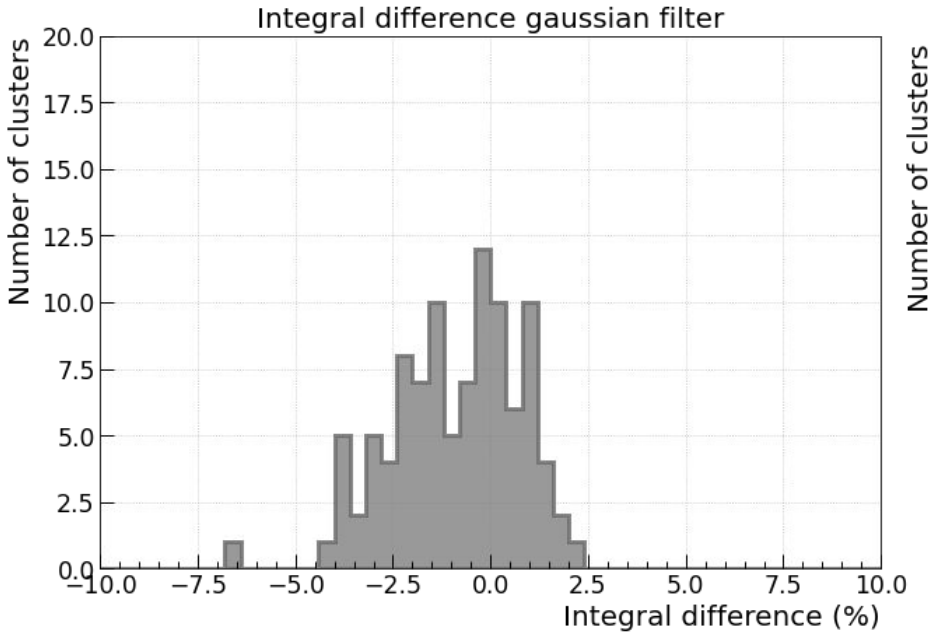
Energy estimation error

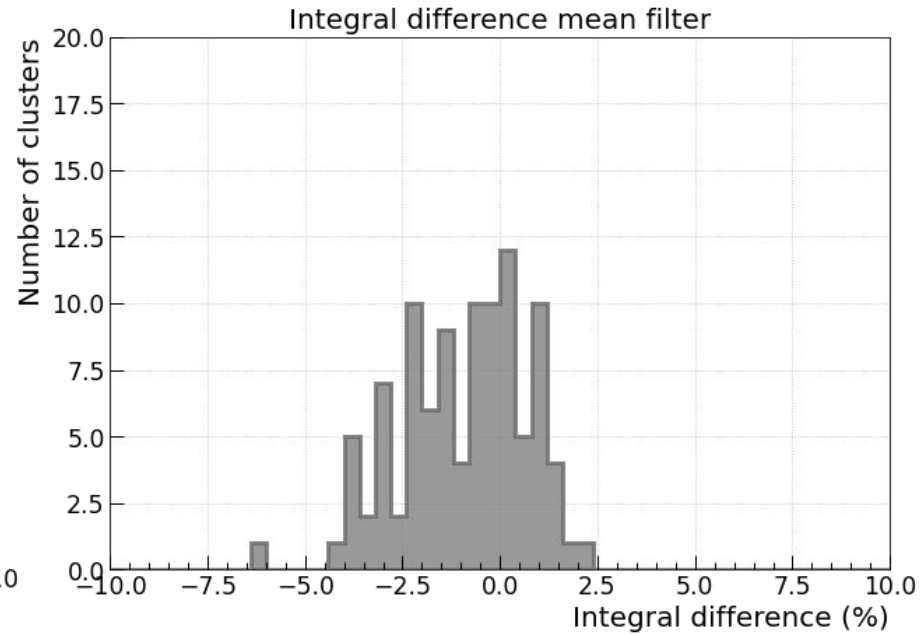




The same trend can be seen here. However, the median filter has a higher loss in energy reconstruction.

Estimation error





The two filters that resulted in the closest estimation of the signal energy.

Conclusions

- The gaussian and average filters alongside the U-net showed the best results concerning the F1 score.
- The median filter had the best precision, but its smaller recall reduced the F1 score.
- A larger dataset containing more diverse images should be used to justify this proposal.

Next steps

- Use a dataset containing more diverse images (more signals with different energies per image).

- Train the U-net using LIME data.