Local Anomaly Detection in Rocket Fuel Combustion Data

WAW Machine Learning 6
October 27th – 29th 2020

Alexander Rüttgers (SC-HPC)
Institute for Software Technology

Joint work with Anna Petrarolo (RA-TRS)
and with Lars Steffens (AS-HYP)

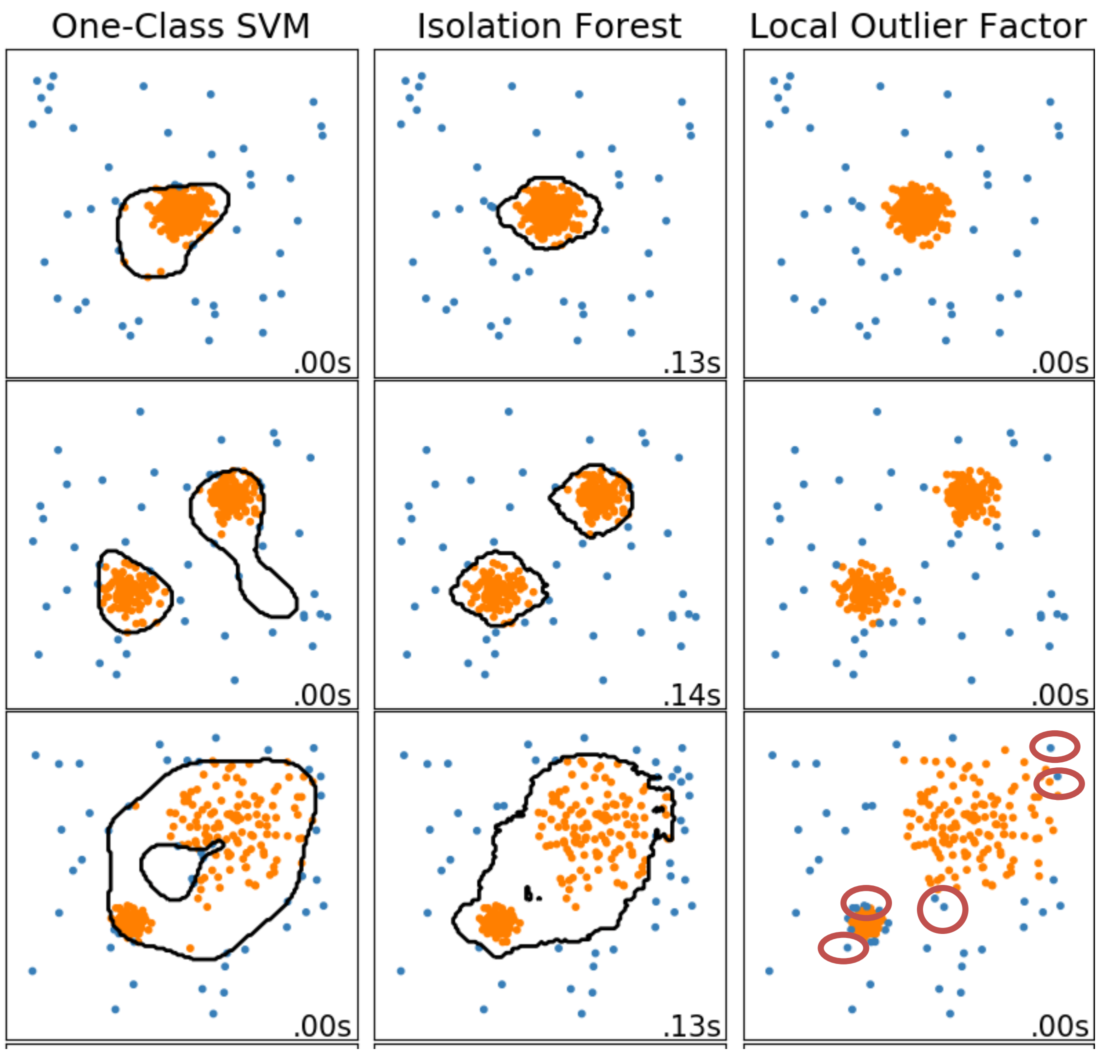
DLR

Overview of outlier detection methods

Some popular techniques are:

- Density-based techniques (LOF, Isolation Forests, …)
- Cluster analysis-based outlier detection (DBSCAN, OPTICS, …)
- One-class support vector machines (SVM)
- Adapted neural networks (autoencoders, variational autoencoders, …)
- Covariance estimation in Gaussian distributed dataset

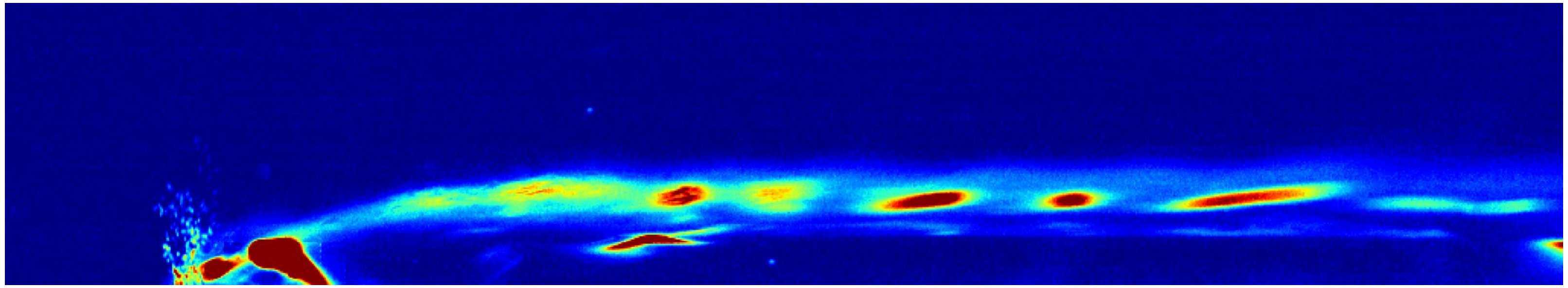
*https://scikit-learn.org/stable/modules/outlier_detection.html



Anomaly detection on 2d toy datset*

- orange: inlier
- blue: outlier
- black: decision boundary

Applications: Outlier Detection on Rocket Fuel Combustion Image Data



Application 1
Hybrid rocket fuel combustion
with RA-TRS
Test 284 (3 seconds, 30 000 images)

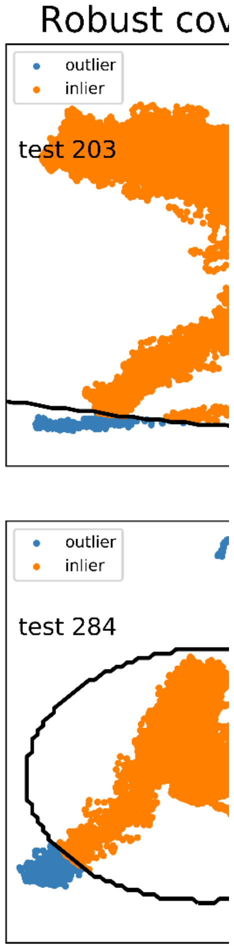
Application 2
Projects ATEK / STORT
with AS-HYP, MORABA, BT-KVS
Static Firing Tests

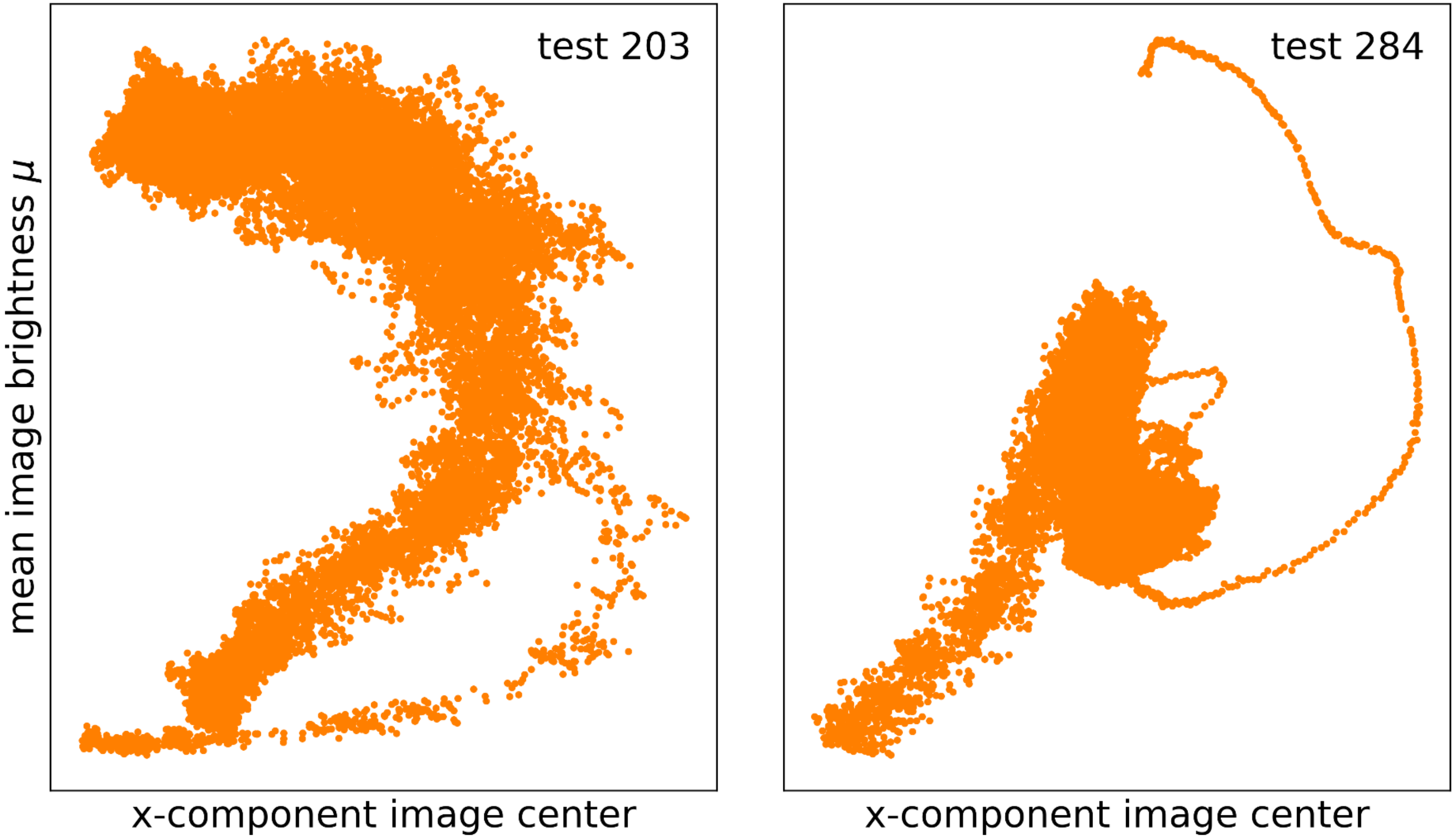
How to find an adequate algorithm for our applications?

- **Start:** Comparison of two features $(\mu, \bar{x})_j$ for all $j = 1, \ldots, 30000$ images of test 203 and 284 (Application 1).

test 203

test 284



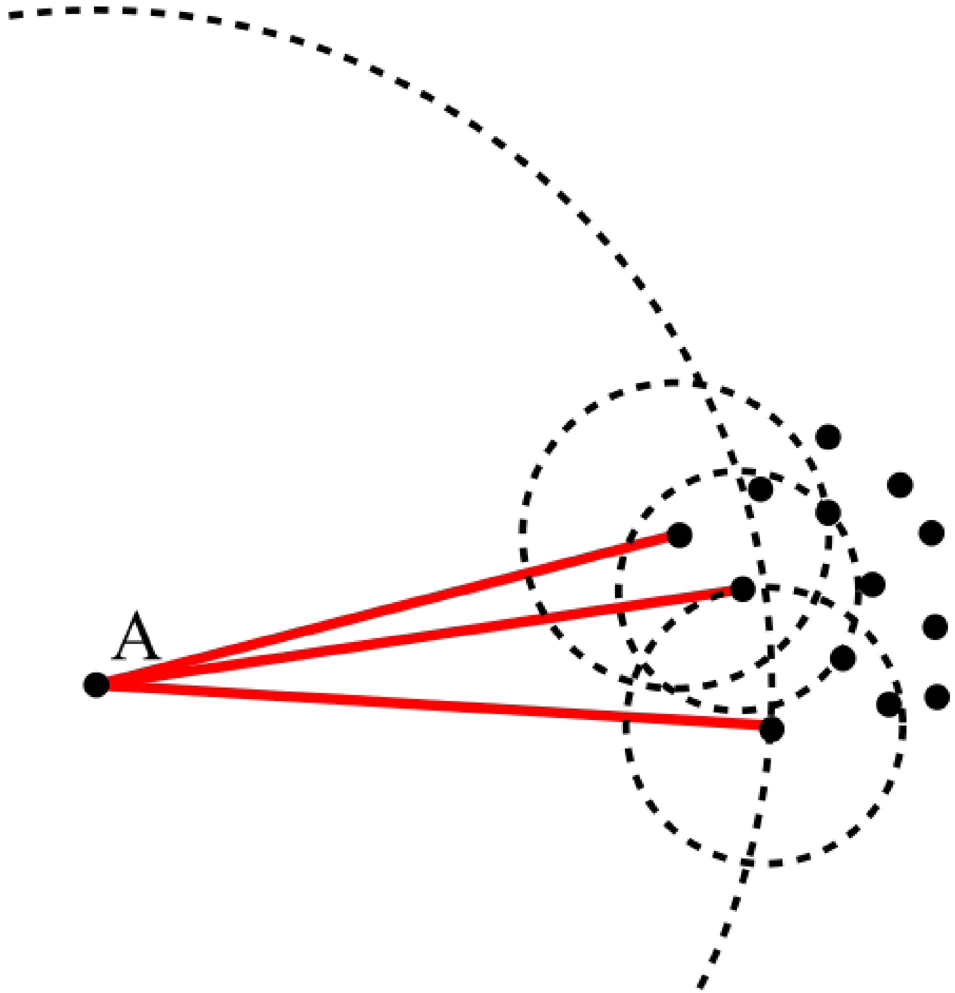


rizontal axis:
rizontal position of
e image center

rtical axis:
ean image
ightness

Local Outlier Factor (LOF)

- Algorithm that bases on local density of data points.
- Shares some concepts with clustering algorithms such as DBSCAN and OPTICS.
- Does not show a decision boundary, i.e. cannot directly be used on new data (not necessary here)
- **Core idea:** Compare local density of an object to the local densities of its neighbors.
- Ratio „Density of neighbors / local density of an objects"
    - ≈ 1.0 means similar density as neighbors
    - > 1.0 means lower density than neighbors (outlier candidate)

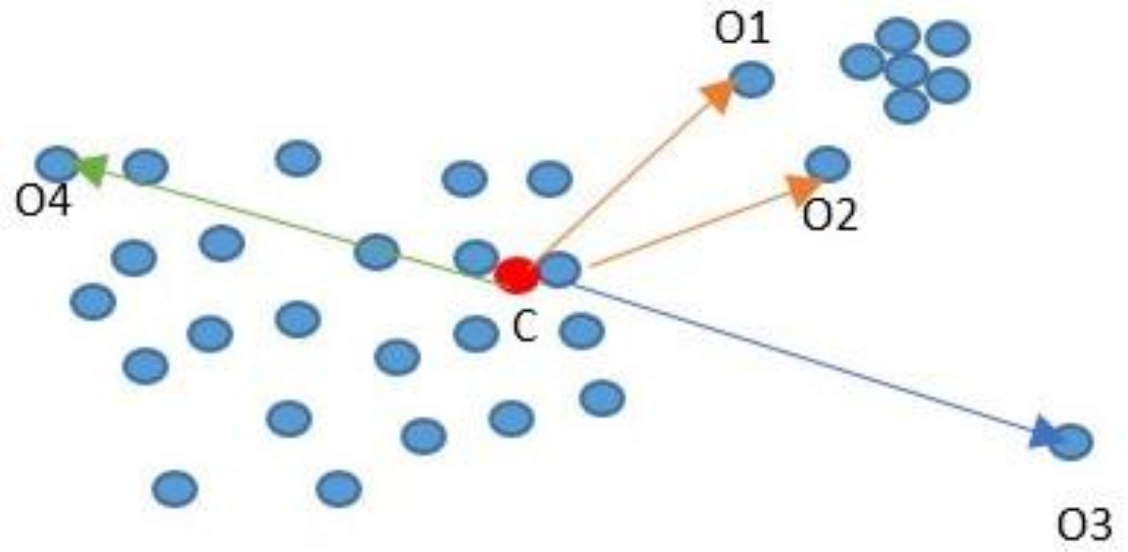


Point density with respect to *k*=3 closest neighbors

Pros and Cons of Local Outlier Factor (LOF)

Advantages

- Algorithm recognizes local outliers
  ⇨ Applications: detect anomalies in different combustion flow phases and not in transition regime.
- Deals with regions of varying densities.
- Only requires a dissimilarity function not a distance function (i.e. triangle inequality is not required).

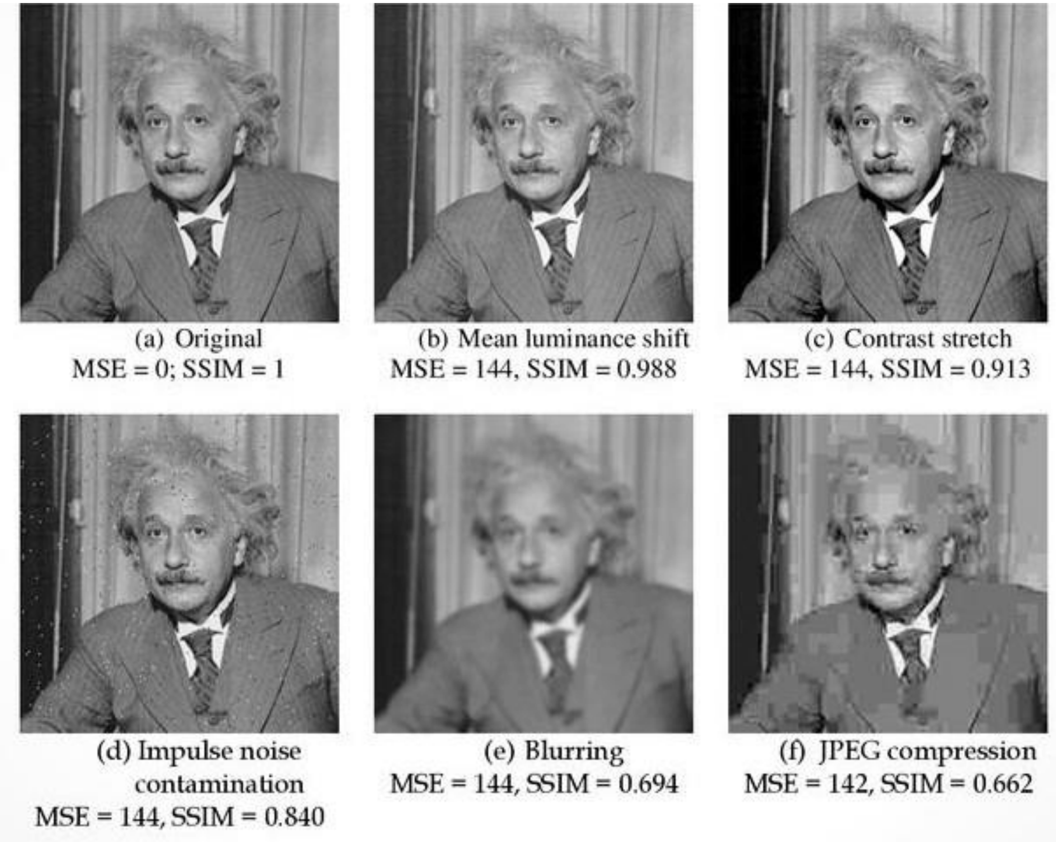


Disadvantages

- Outlier score > 1.0 is hard to interpret (threshold value is problem dependent).
- No decision boundary (important for additional data).
- How to determine hyperparameter *k* (number of neighbors that is considered)?

Dissimilarity measure for image data

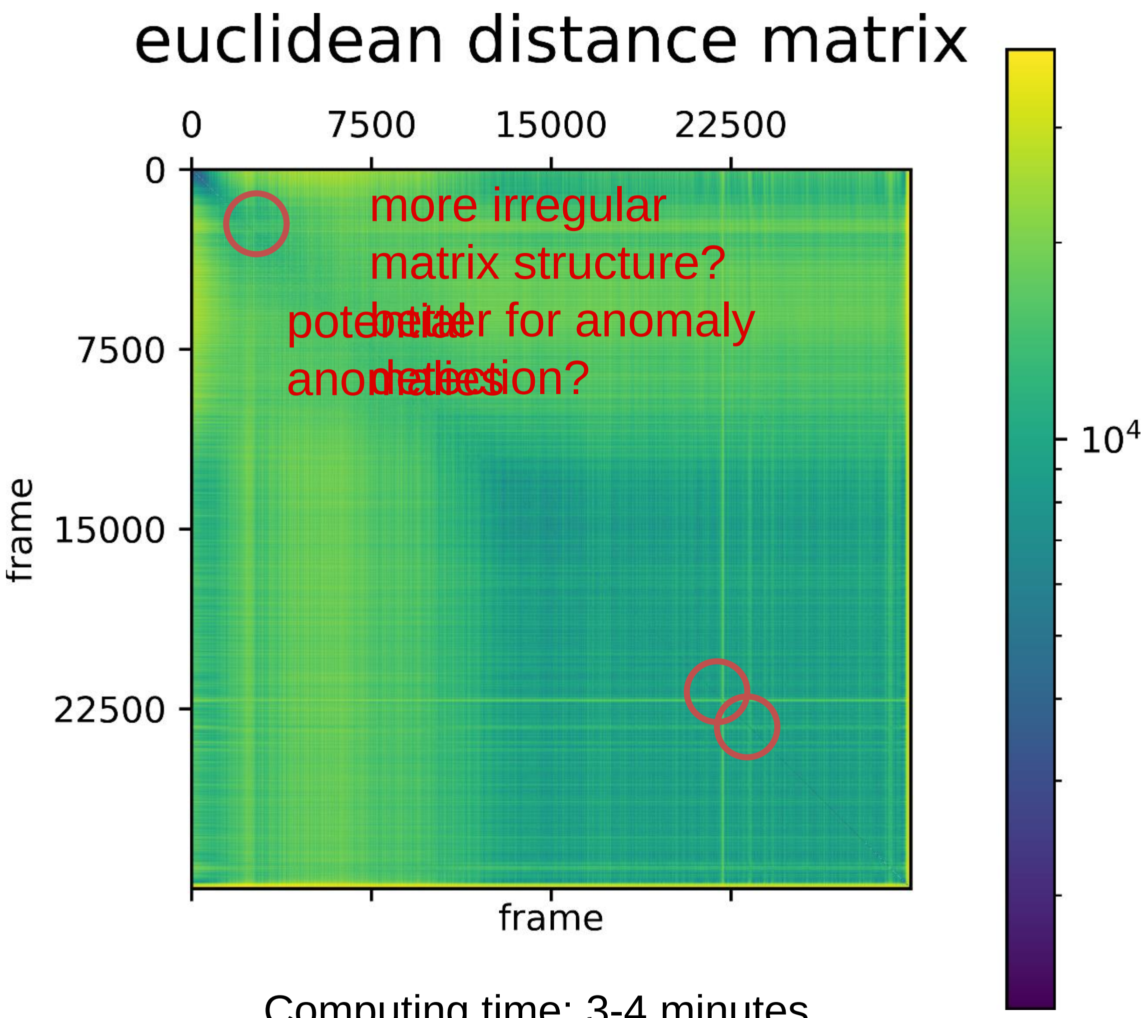
- LOF requires pairwise dissimilarity of images (matrix of size nr_of_images x nr_of_images).
- Standard approaches such as mean squared error (MSE) / discrete L^2-norm often differ from human recognition.
- Advanced dissimilarity measures such as structural similarity (SSIM) often perform better (considers luminance, contrast and structure) but are much more expensive.
- Structural similarity (SSIM)/ structural dissimilarity (DSSIM) is not a distance metric (but not required for LOF).



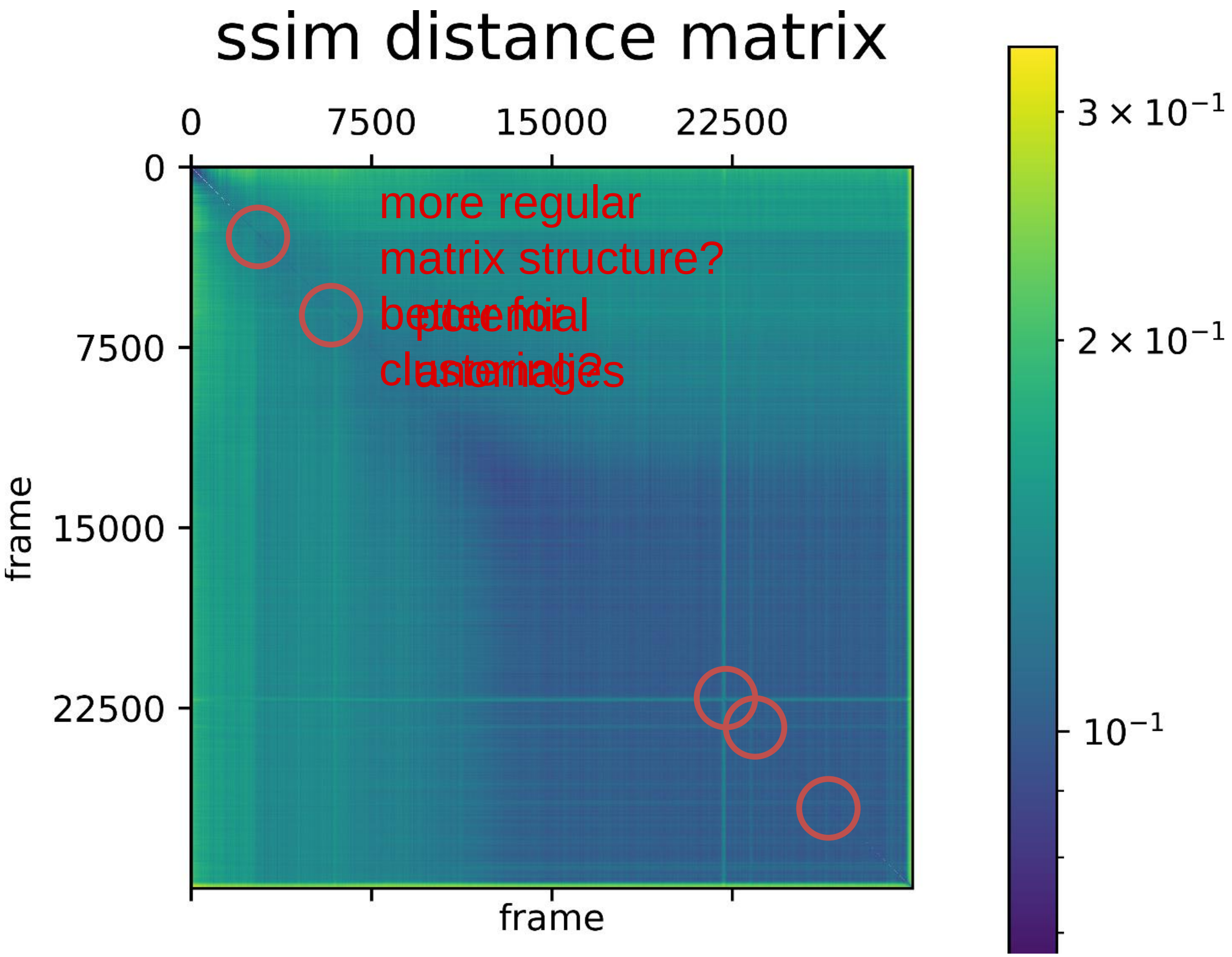
Example: (b)-(f) with same MSE, SSIM decreases*

*https://nsf.gov/news/mmg/mmg_disp.jsp?med_id=79419&from=

Pairwise distance matrices for test 284



Computing time: 3-4 minutes



Computing time: 5 days (OpenMP parallel, 56 cores)
one comparison ≈ 0.1 s (scikit-image)

Application 1: Experiments on hybrid rocket fuels (with RA-TRS)

- Combustion tests were performed with single-slab fuel with 20° forward facing ramp angle.
- Optically accessible combustion chamber is 450 mm long, 150 mm wide and 90 mm high.
- Combustion is captured with high-speed video camera with 10 000 frames / second
- Up to now, 18 tests have been investigated with LOF anomaly algorithm (≈ 500 000 images).

Fig. 1: Fuel slap configuration before (top) and after (bottom) combustion test.

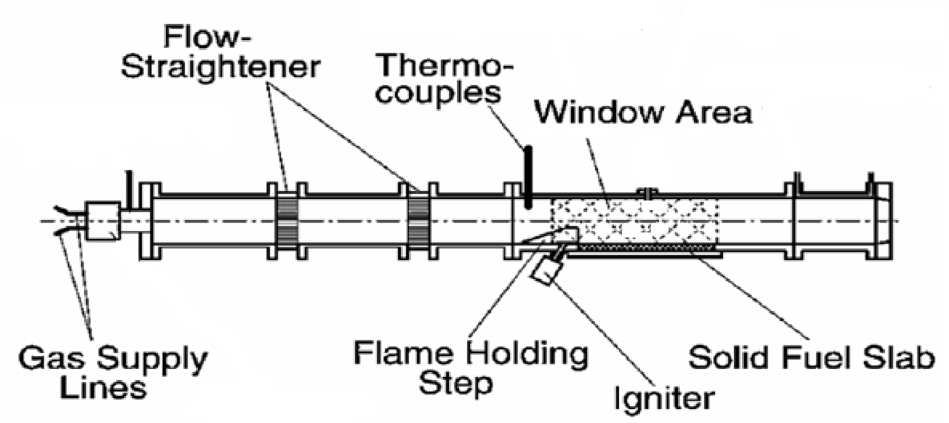


Fig. 2: Side view of combustion chamber

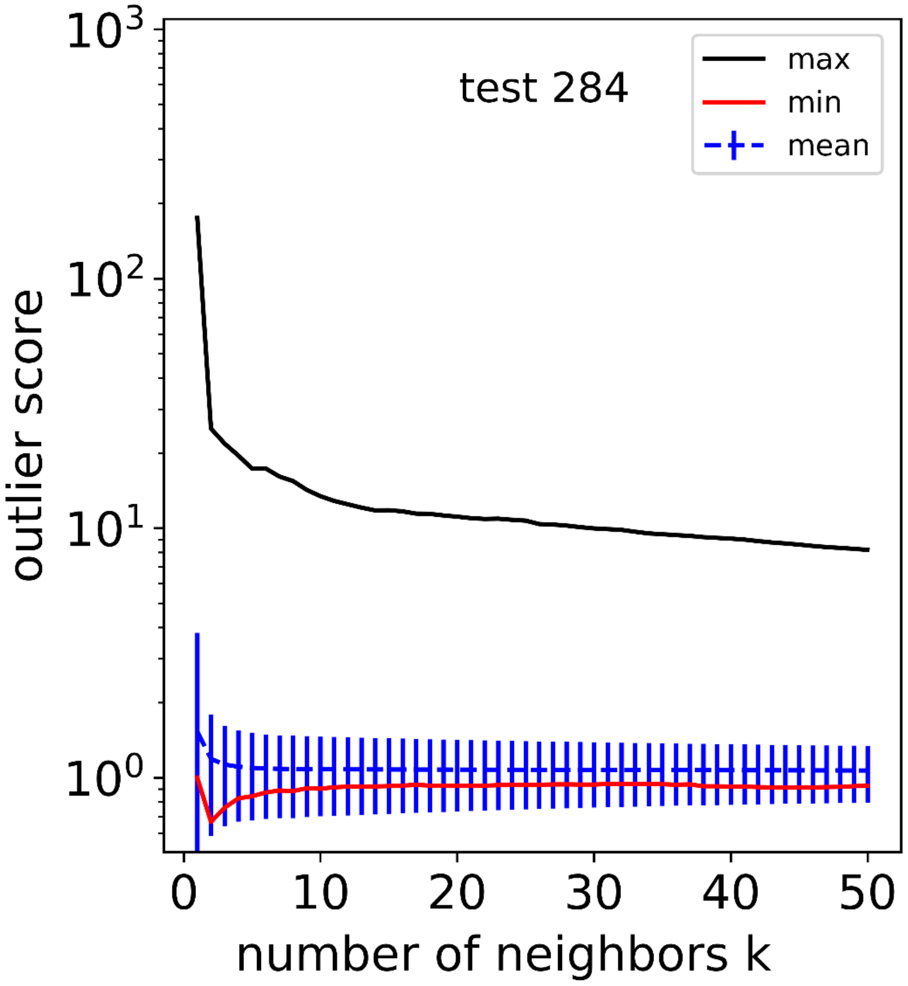
Test 284

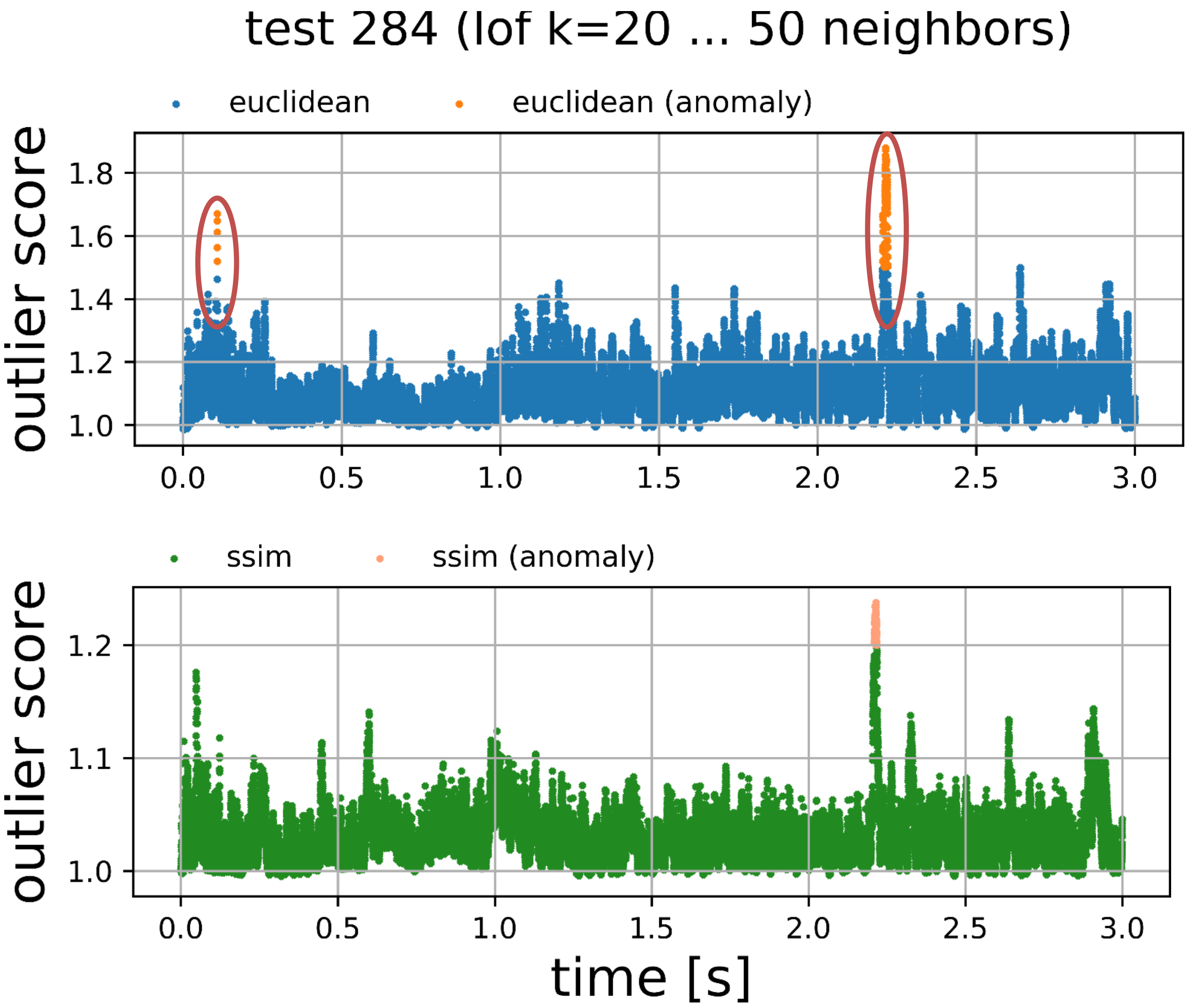
Schlieren video:
(test extract)

fuel = pure paraffin 6805
oxidizer mass flow = 50 *g*/*s*,
CH*-filter (i.e. wavelengths emitted from CH* are filmed)
test 3s = 30 000 frames / 8GB data per test

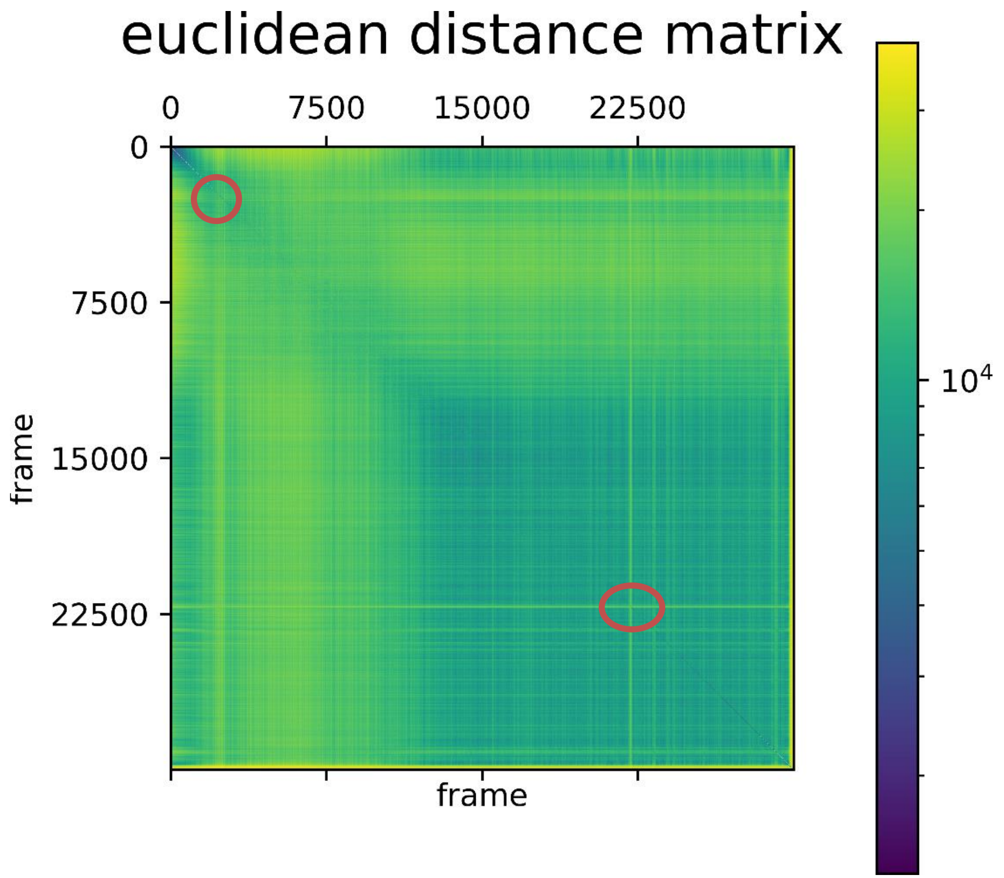
How does the hyperparameter *k* affect the LOF result?

- If *k* is chosen too small, the result is affected by stochastic oscillations.
- If *k* is chosen too large, LOF becomes a global algorithm.
- In the literature, a lot of authors recommend *k*=20.
- Here: We compute LOF values for a range of different hyperparameter values, i.e. $\max_k(\mathrm{LOF\ value\ (image}\ j\ ))$ pointwise for image $j = 1, \ldots, 30000$ and $k \in \{k_{\min}, k_{\max}\}$.

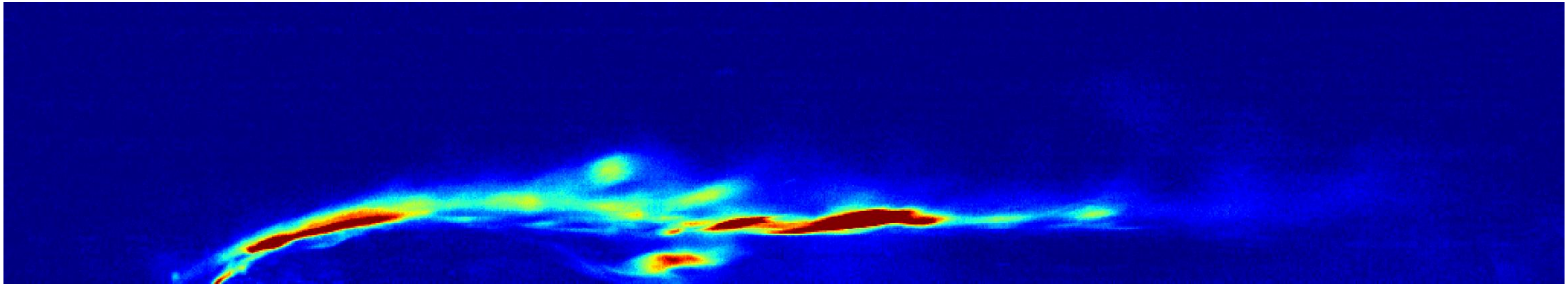




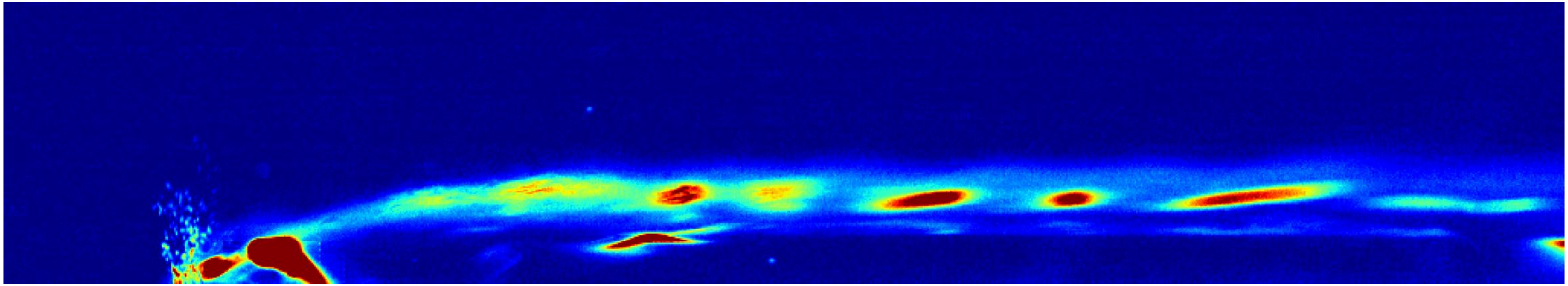
- Euclidean distance norm returns larger outlier score values (due to irregular matrix?).
- SSIM and Euclidean distance share some anomalies but there are differences.



Peak outliers of Euclidean metric (test 284)

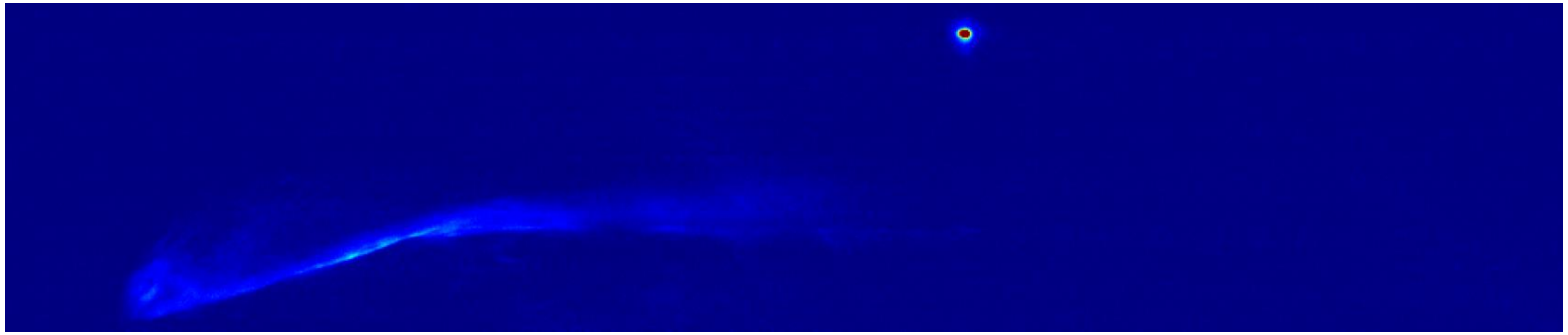


Flame fluctuations in ignition phase at t = 0.1078 s



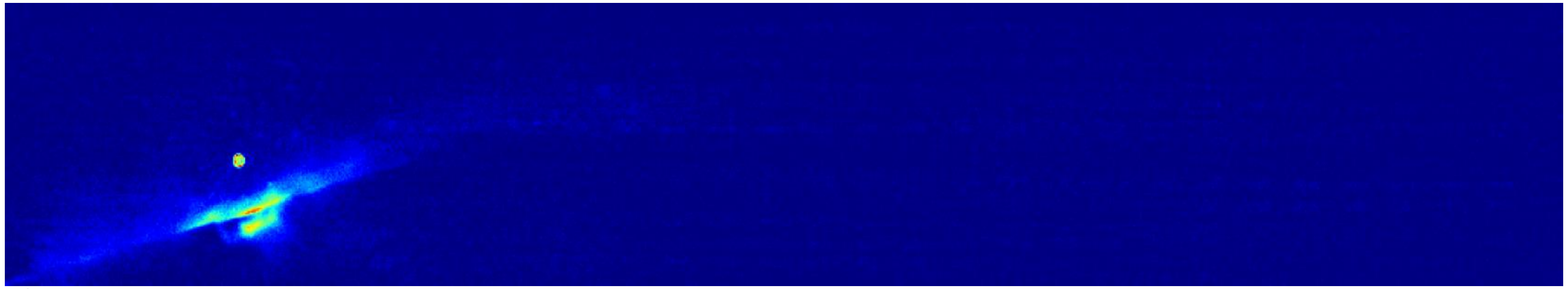
Droplet detection towards end of combustion at t = 2.2055 s

Some outliers found in other combustion tests



Test 291:
satellite droplet at t = 0.0253 s

Test 296:
satellite droplet at t = 0.0017 s

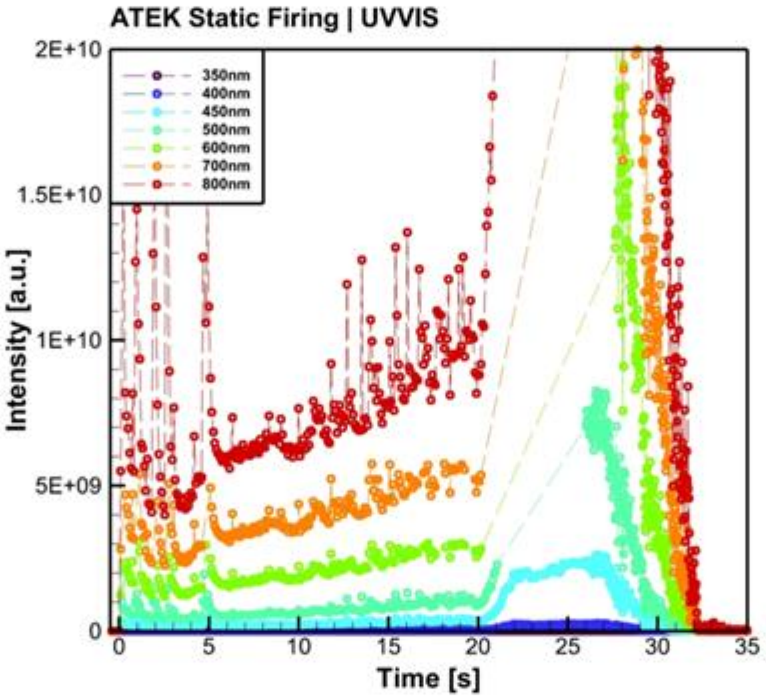


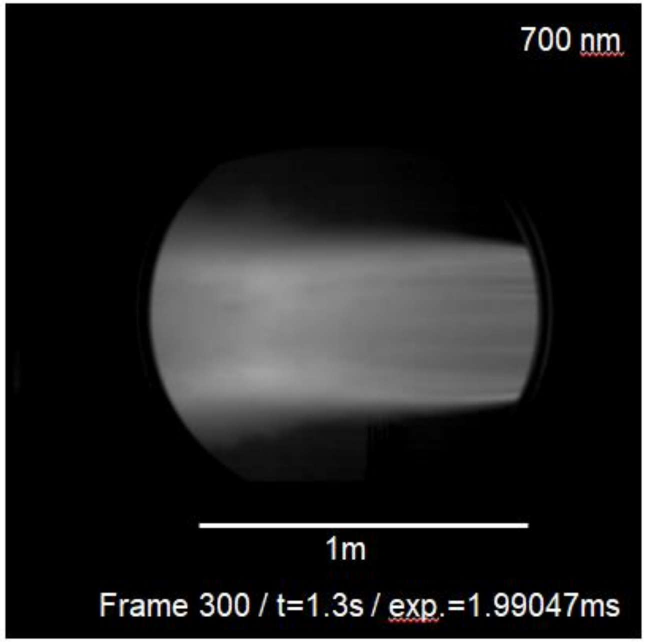
Test 296:
satellite droplet at t = 0.0223 s

Application 2: ATEK Static Firing Test (with AS-HYP, MORABA, BT-KVS)

- ATEK Static Firing Test was performed on April 27th 2018 at Esrange rocket range (Sweden). ATEK rocket flight was on June 13th 2019 (⇨ YouTube video: “Mission ATEK: Vom hohen Norden ins All”).

- Three different datasets were obtained (video data, UVVIS spectral data, AEM video data).





Optical video data

UVVIS spectral data

Alumina Emission Measurement (AEM) data

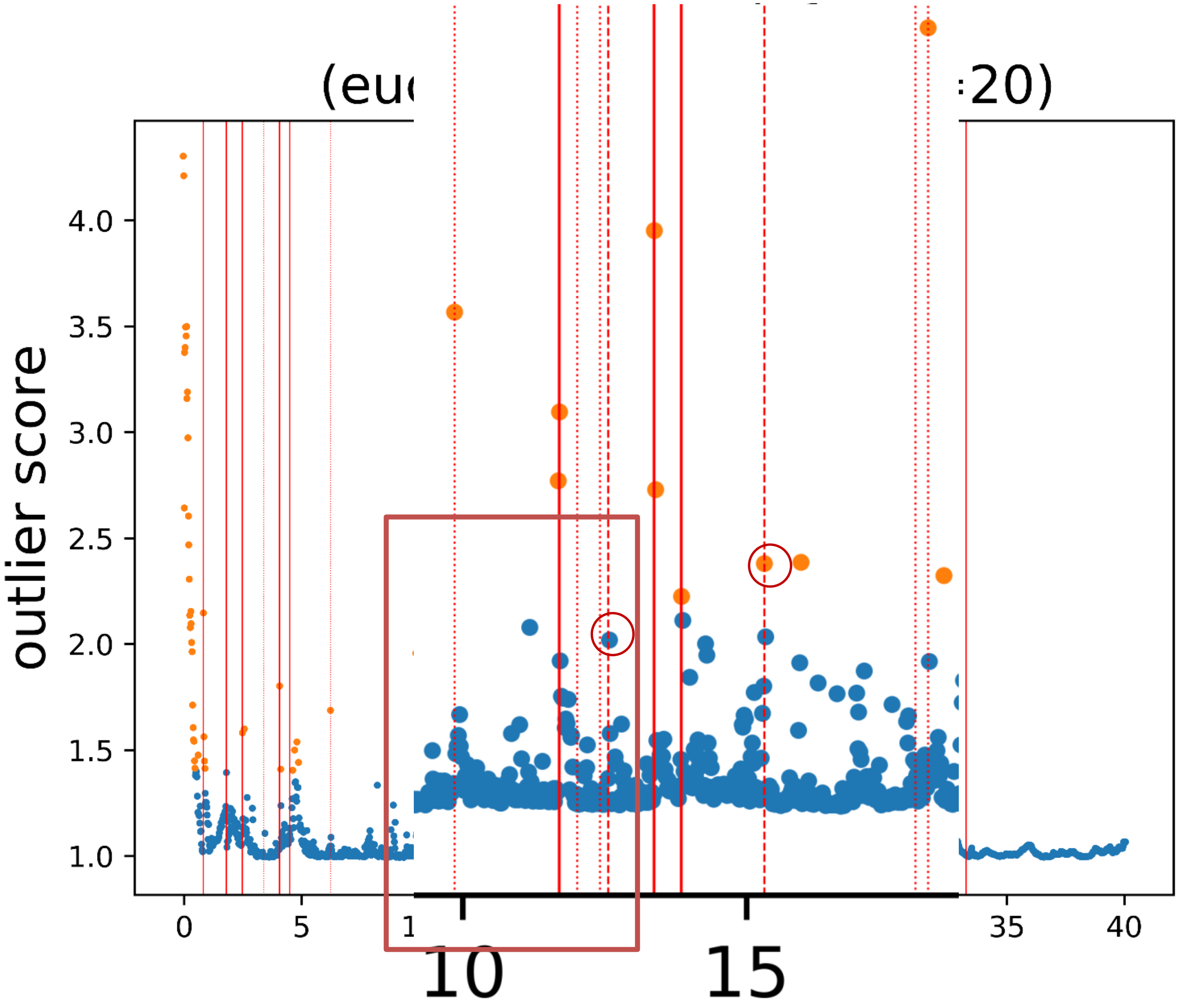
LOF algorithm manual video a

LOF algorithm on spectrum vs. manual video analysis (not shown here)

LOF anomaly detection on ATEK static firing video data (only left camera)

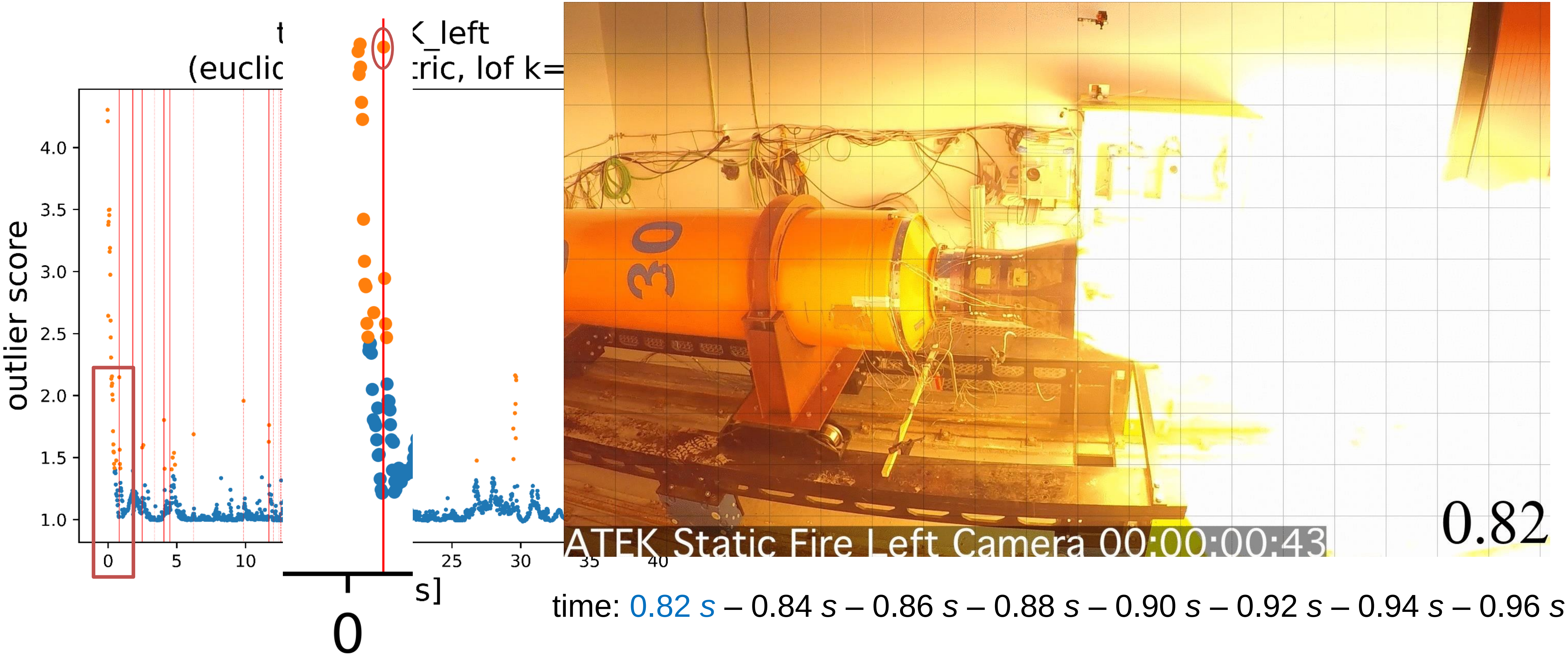
manually detected anomalies (RB-MRB)

- visible on left + right camera
- only visible on left camera
- only visible on right camera



- Red vertical lines: manually detected by RB-MRB
- orange: images with outlier score > 1.4
- blue: images with outlier score < 1.4

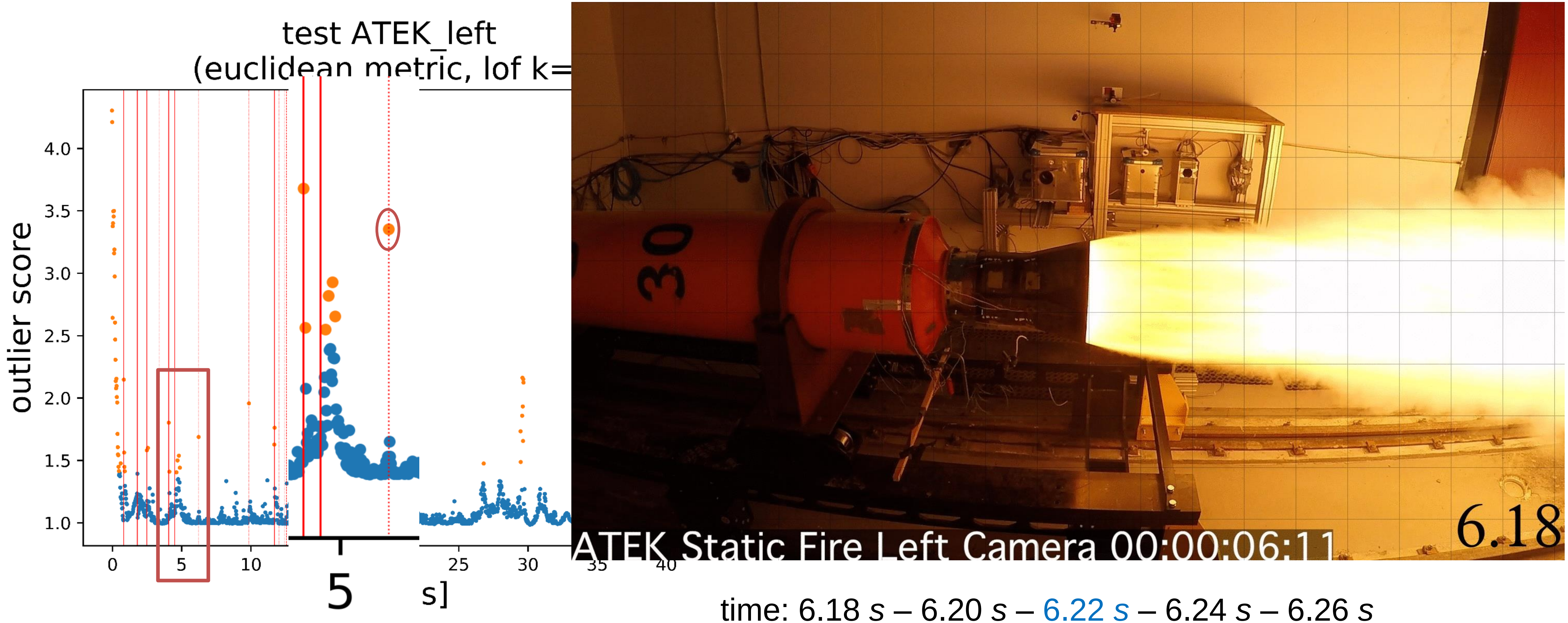
Anomaly 2
0.82 seconds



time: 0.82 *s* – 0.84 *s* – 0.86 *s* – 0.88 *s* – 0.90 *s* – 0.92 *s* – 0.94 *s* – 0.96 *s*

DLR

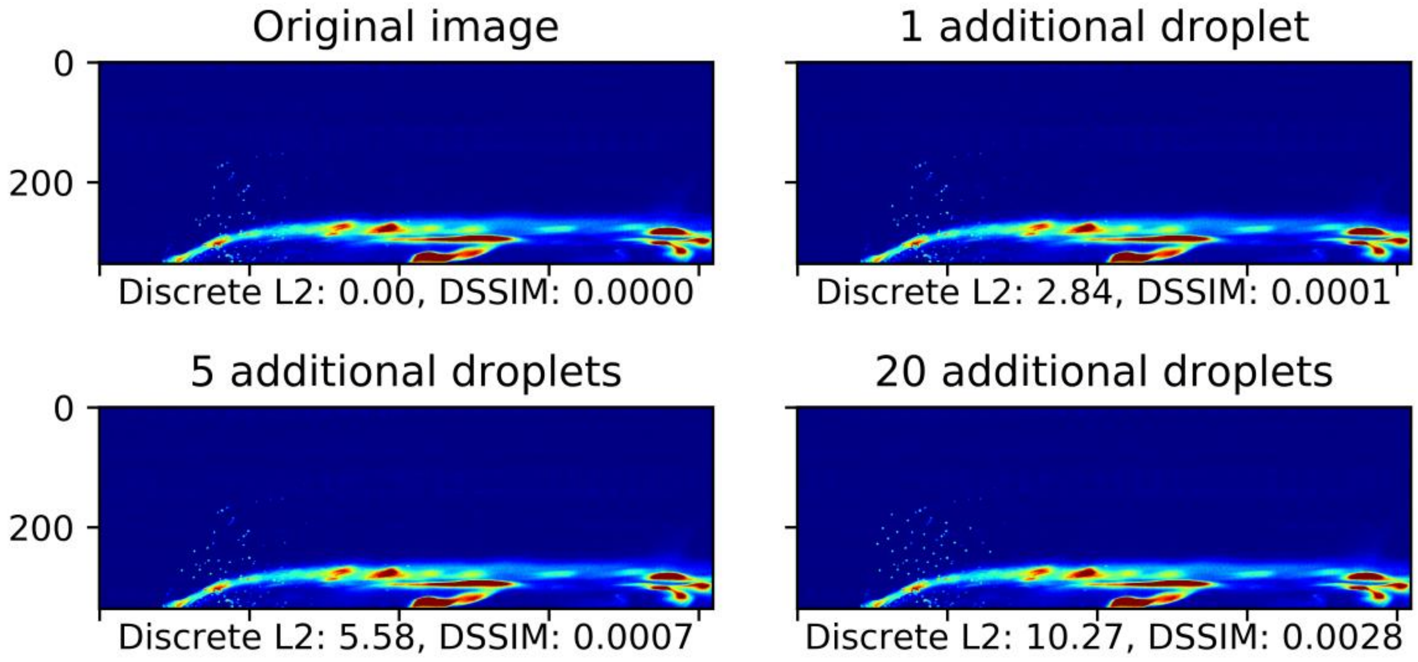
Anomaly 2
6.22 seconds



time: 6.18 *s* – 6.20 *s* – 6.22 *s* – 6.24 *s* – 6.26 *s*

Conclusion and outlook

- Local Outlier Factor is able to detect anomalies in image data provided that distance measure is adequate.
- Further insights are possible if datasets are combined (e.g. anomaly detection in spectral and image data).
- Future work is spent on distance measures that are more adapted to the „interesting anomalies“.



Thank you for your attention!