Automatic Extraction of Solar Filaments Using Machine Learning Techniques

A. Diercke[1,2,3], R. Jarolim[4], C. Kuckein[2,5,6], S.J. González Manrique[5,6,7,8], M. Ziener[9], A.M. Veronig[4,10], C. Denker[2], W. Pötzi[4,10], T. Podladchikova[11]

1 National Solar Observatory (NSO), USA, 2 Leibniz-Institut für Astrophysik (AIP), Germany, 3 University of Potsdam, Germany, 4 University of Graz, Austria, 5 Instituto de Astrofísica de Canarias (IAC), Spain, 6 Universidad de La Laguna, Spain, 7 Leibniz-Institut für Sonnenphysik (KIS), Germany, 8 Astronomical Institute, Slovak Academy of Sciences, Slovak Republic, 9 Tidy GmbH, Germany, 10 Kanzelhöhe Observatory, Austria, 11 Skolkovo Institute of Science and Technology, Russia

Abstract. Filaments are omnipresent features in the solar chromosphere. Regular full-disk Hα observations allow us to analyze statistical properties of filaments. Therefore, filaments have to be extracted from the images. Manual extraction is tedious and takes too much time; extraction with morphological image processing tools produces a large number of false-positive detections. Automatic object detection and extraction allows us to process more data in a shorter time. We use a semi-supervised approach to reliably extract filaments from Hα data. The trained neural network is used to create a catalog of filaments starting 2010, which will accelerate research on filaments.

Data. In order to train a neural network to perform object detection of filaments, we had to label these filaments in Hα full-disk filtergrams. For this purpose, we decided to use the dataset of Chromospheric Telescope (ChroTel, Kentischer et al. 2008, Bethge et al. 2011). The regular ChroTel observations started in 2012 and are ongoing. The dataset contains 1056 days of observations until September 2020, whereby on each day the qualitatively best filtergram was selected. We apply a data pre-processing where we use a dark- and flat-field correction, and adjust the center-to-limb variation. Furthermore, the images are normalized to the median intensity of the solar disk and are limb darkening corrected. The Lyot filter introduced a non-uniform intensity variation, which is corrected by approximation with Zernike polynomials (Shen et al. 2018). Chrotel is our primary dataset, which we we use for manual labeling of solar filaments. As secondary datasets we consider Hα filtergrams of the Global Oscillation Network Group (GONG, Harvey et al. 1996) and Kanzelhöhe Solar Observatory (KSO, Pötzi et al. 2021). We use this data for automatically extending our training dataset and additional verification. The data is automatically read into algorithm, where basic image reduction steps are performed, such as, rotation to the solar north, normalization to the median intensity, as well as a correction of the intensity with Zernike polynomials, to standardize the images in comparison to the ChroTel images.

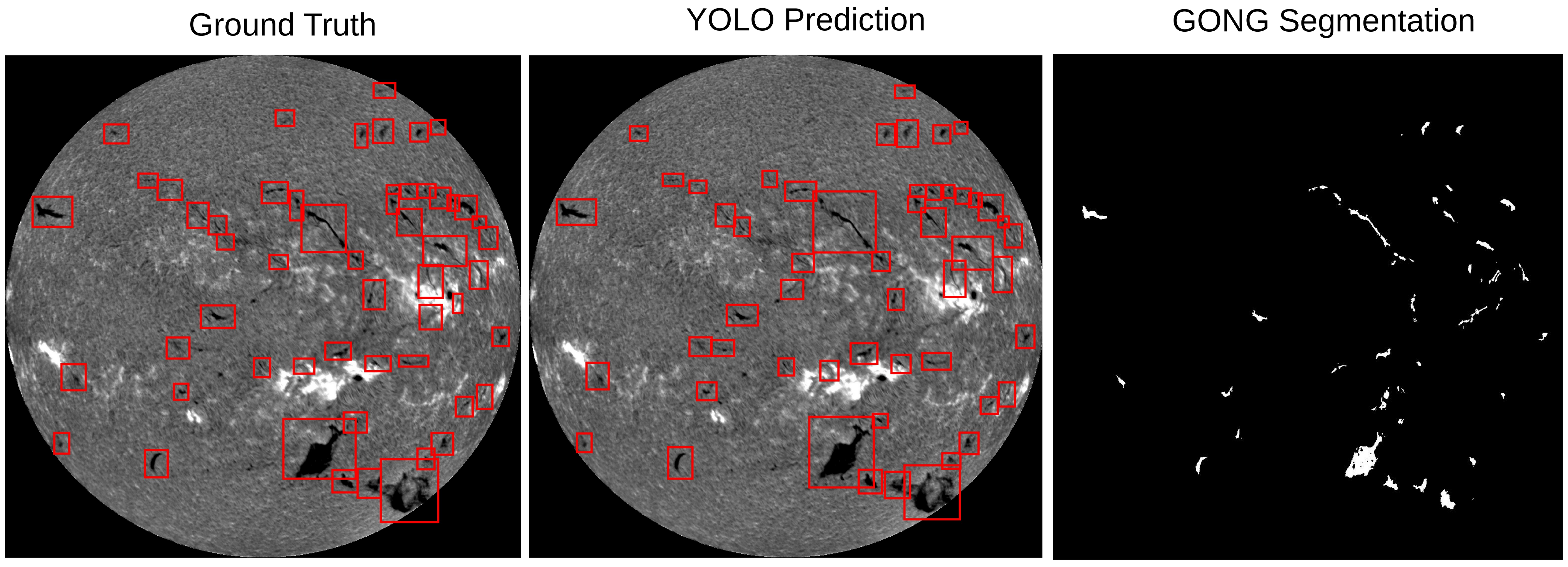

Fig 1: Sample Hα filtergrams from ChroTel for 2013 October 26. The manually label bounding boxes (red) are used as the input data (left panel) to train the YOLOv5 neural network. In the middle panel, the predicted bounding boxes by YOLOv5 are displayed. The extended dataset of GONG is used to train a general segmentation Network (U-Net) on Hα filtergrams resulting in a segmentation map (right panel).

Data Labeling. We manually labeled one image of each day from the entire ChroTel dataset between 2012 and 2018, which includes 955 observing days. The images are displayed at a lower resolution with an image size of 1000 x 1000 pixels and a rectangle is drawn around each filament of the images. In some cases the filament is split in several parts, whereby we labeled each part individual with a bounding box. The labeled data cover observations from the maximum and minimum of Solar Cycle 24. The labels of the bounding boxes and the images with a resolution of 1024 x 1024 pixels are used as ground-truth input data for the training of YOLOv5. The dataset is split in a test set, a training set, and a validation set. The training set is the largest with 680 images (70 %) and the validation set contains 85 images (10 %). The test set contains the remaining 190 images (20 %), whereby we select for each year one continuous block of 20 % of images for the test set.

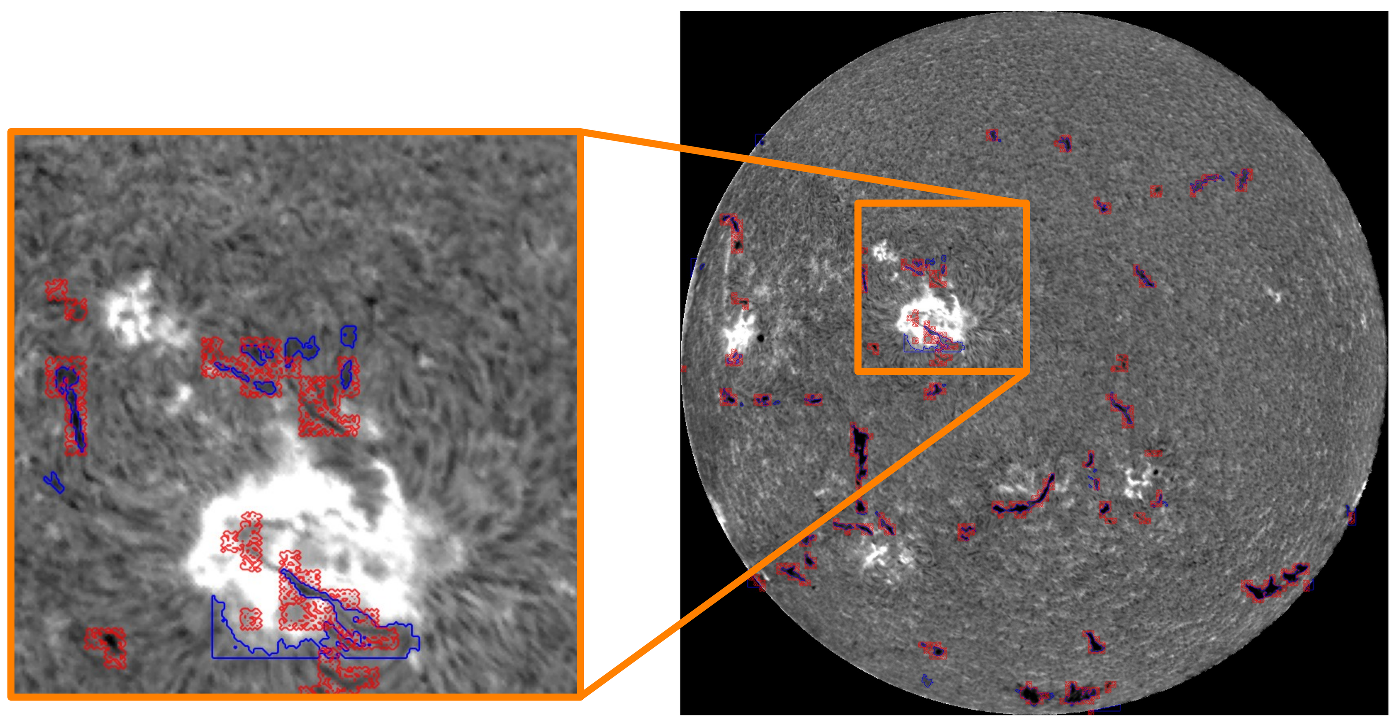

Fig 2: Hα filtergrams from ChroTel for 2012 October 14 with the segmentation from the YOLOv5 prediction (blue) and final segmentation from U-Net (red) for full-disk (right) and for a cut-out region (left).

Method. While bounding boxes allow for a convenient labeling, our goal is a pixel-wise identification of solar filaments. As a baseline approach we use a global threshold of 1σ and select only pixels enclosed in bounding boxes. The problem with this approach is that extended filaments and bounding boxes that overlap with sunspots lead to invalid identifications. Furthermore, we have found that the prediction of bounding boxes is more robust than a pixel-wise segmentation, where the limited amount of training data leads to high dependence on the used instrument (overfitting).
In our study, we aim for a general method that can provide reliable solar filament detection for arbitrary Hα observations. We employ a semi-supervised approach where we take advantage of the more robust bounding boxes and the reliability of a pixel-wise segmentation. We start by training a YOLOv5 model (Redmon et al. 2016, Jocher et al. 2020) to predict bounding-boxes, based on our labeled ChroTel observation. The resulting model is used to label the full GONG archive of solar observations at a 4-hour cadence, where we use an equal splitting between the available observing sites.
From this we apply our thresholding approach to create a large data set of solar filament segmentations, which we use for training a final U-Net model (Ronneberger et al. 2015). We note that the baseline thresholding approach inevitably contains invalid pixel-wise classifications. The error cases are small as compared to the full data set size, and appear only as noise during model training. Thus, the final segmentation model distinguishes more reliable between sunspots and filament.

Catalog. To foster novel research with solar filaments, we provide a catalog based on the GONG data set. For this we apply our trained segmentation method to the full 4-hour cadence dataset. We use the Image-quality Assessment from Jarolim et al. (2020) to filter observations, which suffer from atmospheric degradation. The resulting pixel-wise segmentation has a high temporal cadence and high precision on the full-disk Hα filtergrams.

The catalog of solar filaments enables new studies of solar filaments. For example, statistical studies of filaments over several decades can be faster obtained. Furthermore, the pixel-wise segmentation allows to obtain parameters such as the size or orientation of the filaments. Ultimatly, the catalog paths the way towards space weather forecast, such as the detection of filament eruptions and to study the triggers of such events.

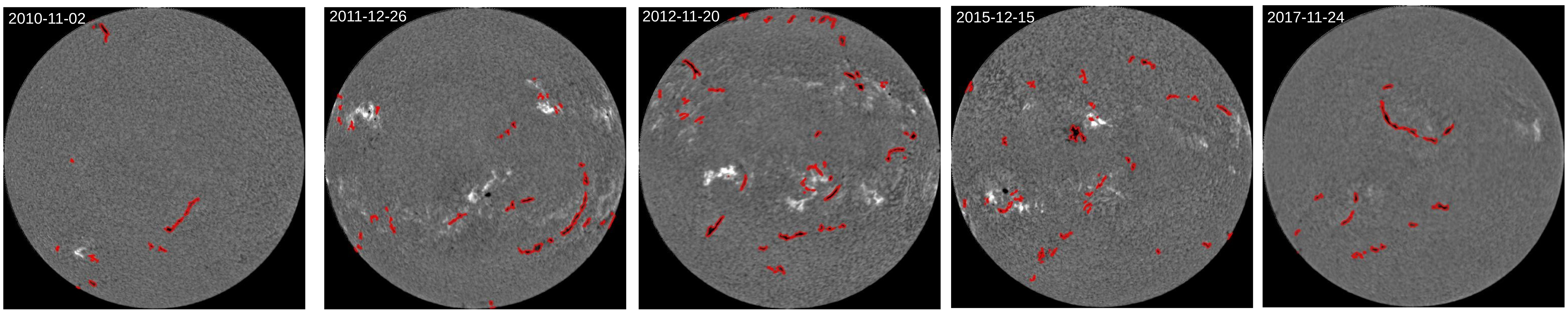

Fig 3: Sample Hα filtergrams from GONG and the corresponding predictions as segmentation maps for different conditions of the solar cycle and different seeing conditions.

References.
- Bethge, C., Peter, H., Kentischer, T. J., et al. 2011, A&A, 534, A105
- Jarolim, R., Veronig, A. M. ; Pötzi, W. ; Podladchikova, T., (2020) A&A 643, A72
- Jocher, G., Stoken, A., Borovec, J., et al. 2020, ultralytics/yolov5: v3.0 (https://github.com/ultralytics/yolov5)
- Harvey, J. W., Hill, F., Hubbard, R. P., et al. 1996, 272, 1284
- Kentischer, T. J., Bethge, C., Elmore, D. F., et al. 2008, in Proc. SPIE, Vol. 7014, 701413
- Pötzi, W., Veronig, A. M., Jarolim. R, et al. 2021, Sol. Phys., 296, 164
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. 2016, You Only Look Once: 620, Unified, Real-Time Object Detection
- Ronneberger, O., Fischer, P., & Brox, T. 2015, in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, 630, 234–241
- Shen, Z., Diercke, A., & Denker, C. 2018, AN, 339, 661