"This project has received funding from the European Union's Horizon 2020 research and innovation programme for the period January 2019 until December 2022 under Grant Agreement No 824135 (SOLARNET)"

Project (Grant) Number:	824135
Project Acronym:	SOLARNET
Project title:	Integrating High Resolution Solar Physics

Document Details	
Document Title	Assessment report of suitable CNN architectures
Prepared by (Institution's Name):	Leibniz-Institut für Astrophysik Potsdam (AIP)
Work Package (WP) number & Title	WP5.2 – Data archiving and development of advanced data search techniques
Deliverable number & Title:	D5.6 - Assessment report of suitable CNN architectures
Serial number of Deliverable:	D48
Document code (inserted by project office)	D5.6_Version 1.0
File name (inserted by project office)	SOLARNET_D5.6_V1.0_CNN_Public_20191219
Date uploaded (inserted by project office)	Dec 19th, 2019

Leibniz-Institut für Sonnenphysik (KIS)
Schöneckstr. 6, 79104 Freiburg Germany
Tel.: +49-761-3198-224; Email: SOLARNET-office@leibniz-kis.de

"This project has received funding from the European Union's Horizon 2020 research and innovation programme for the period January 2019 until December 2022 under Grant Agreement No 824135 (SOLARNET)"

AUTHORS/ CONTRIBUTORS LIST

Name	Function	Organization
Christoph Kuckein	Postdoctoral Researcher (Task Leader)	Leibniz-Institut für Astrophysik Potsdam (AIP)

APPROVAL CONTROL FROM SUB-WP & WP LEAD

Control	Name	Organization	Function	Date
Prepared	Christoph Kuckein	AIP	Postdoc (Task Leader)	Dec 17th, 2019
Revised	Andrea Diercke	AIP	PhD Student (Task contributor)	Dec 18th, 2019
Revised	Marco Ziener	Latentine GmbH	Data Engineering Lead (Task contributor)	Dec 18th, 2019
Revised	Meetu Verma	AIP	Postdoc (Task contributor)	Dec 18th, 2019
Revised	Carsten Denker	AIP	Head Solar Group (sub-WP Leader)	Dec 18th, 2019
Approved	Nazaret Bello González	Leibniz-Institute for Solar Physics (KIS)	WP5 Leader	Dec 19th, 2019

APPROVAL CONTROL FROM PROJECT OFFICE

Control	Name	Organization	Function	Date
Approved	Tirtha Som	KIS	Project Manager	Dec 19th, 2019
Approved	Rolf Schlichenmaier	KIS	Project coordinator	Dec 19th, 2019
Authorized	Markus Roth	KIS	Project Scientist	Dec 19th, 2019

HISTORY OF DOCUMENT CHANGES

Issue	Date	Change Description
Version 1.0		Initial Issue

Leibniz-Institut für Sonnenphysik (KIS)
Schöneckstr. 6, 79104 Freiburg Germany
Tel.: +49-761-3198-224; Email: SOLARNET-office@leibniz-kis.de

"This project has received funding from the European Union's Horizon 2020 research and innovation programme for the period January 2019 until December 2022 under Grant Agreement No 824135 (SOLARNET)"

Table of Contents

List of Abbreviations

CNN	Convolutional neural network (also called ConvNet)
FPS	Frames per second
mAP	Mean average precision
NMS	Non-maximum suppression
R-CNN	Region-based convolutional neural networks
ROI	Region of interest
RPN	Region proposal network
YOLO	You only look once

Leibniz-Institut für Sonnenphysik (KIS)
Schöneckstr. 6, 79104 Freiburg Germany
Tel.: +49-761-3198-224; Email: SOLARNET-office@leibniz-kis.de

"This project has received funding from the European Union's Horizon 2020 research and innovation programme for the period January 2019 until December 2022 under Grant Agreement No 824135 (SOLARNET)"

1. Introduction

The main goal of this work is to automatically identify, classify, and provide the pixel-level segmentation of the features on the Sun, from the smallest to the largest spatial scales, in existing and new data. Many advantages arise from access to a fully characterized data base. For instance, unprecedented long-term studies of individual solar events can be accomplished, which further provides input for new models of solar features and predictions of eruptive events on the Sun. Deep Convolutional Neural Networks (CNNs) are ideally suited for this task. Mask R-CNN[1] is an example of this type of network, which facilitates recognizing a variety of object categories in images very quickly and reliably. Same types of networks can be adapted to work with solar data.

Machine learning techniques are being used in many fields of science. Instead of developing complex codes, a computer learns how to solve otherwise time-consuming problems that involve much manual labor. Typically, a huge amount of data is a prerequisite to teach the computer how to identify patterns in the training set and enabling it without any human interaction to recognize similar features in new data. The European Solar Data Centre will archive major data holdings providing an ideal environment for a machine to learn specific patterns present in the data.

In computer vision, object detection consists of two main tasks. On the one hand, objects need to be localized and distinguished from the background of an image. On the other hand, the object needs to be classified, that is, identified with a previously labelled tag. The outcome of an object detection program is one optimal bounding box around each localized object and the corresponding name of the predicted class. Each bounding box is represented by four values: the coordinates of the origin in x and y, the width w and the height of the box h. The classification part consists of a probability value, usually a percentage from 0 to 100%, which provides information about the accuracy of the predicted class.

Object detection algorithms within the framework of machine learning and convolutional neural networks have lately become very popular. This is mainly due to the huge amount of available digital images and due to the development of faster computers and graphic cards. The applications of fast object-detection tools grow rapidly, for instance, autonomous vehicles, medical image diagnosis or road monitoring, among many others. In this report, we will evaluate state-of-the-art object detection architectures based on convolutional neural networks which are suitable for object identification and classification of high-resolution images of the Sun.

2. Object detection architectures

There are four main ingredients for object detection[2]: (1) Regions of interest (ROIs), which can be either obtained by a deep learning model or with segmentation algorithms. (2) Visual feature extraction from the ROIs and prediction of its content. (3) Non-maximum suppression (NMS), combines overlapping bounding boxes of the same object into a single bounding box. (4) Metrics for evaluating the accuracy of the detection performance. The convolutional neural networks architecture is especially useful in object detection because successively complex features can be extracted in each layer.

The most popular object-detection algorithms are: R-CNN family of networks, SSD and the YOLO family of networks. In addition, U-NET can also be used for object detection although it is based on image segmentation. Below we will shortly review their main properties.

2a. Region-based convolutional neural networks (R-CNNs)

The R-CNN is considered as one of the first large and successful integration of a convolutional neural network for object detection. The architecture was developed by Girshick et al.[3] in 2014 and was expanded to Fast-RCNN and Faster-RCNN in 2015 and 2016, respectively. The original R-CNN included a very slow, computational expensive object detection algorithm called *Selective Search*. This was improved in the Fast-RCNN, but remained the bottleneck of the architecture. Finally, the Faster-RCNN[4] architecture accelerated the object detection by a factor of ten, and it is entirely a deep learning object detector. We will only concentrate on the fastest and latest version of the R-CNN. The architecture of the Faster-RCNN is as follows:

Leibniz-Institut für Sonnenphysik (KIS)
Schöneckstr. 6, 79104 Freiburg Germany
Tel.: +49-761-3198-224; Email: SOLARNET-office@leibniz-kis.de

"This project has received funding from the European Union's Horizon 2020 research and innovation programme for the period January 2019 until December 2022 under Grant Agreement No 824135 (SOLARNET)"

Network 1 →	Network 2
Region proposal network (RPN) based on a CNN. Outputs a probability of being an object or not and the bounding box location.	Fast R-CNN: (1) pre-trained CNN model to extract features from the image; (2) ROI pooling layer; (3) Output layer with (a) softmax classifier to output the class probability and (b) a bounding box prediction.

Limitations of the R-CNN architecture compared to other modern architectures:

- Slow network: long training and slow inference time of objects. Real-time object detection not possible.
- Multiple phases are necessary for the training (region proposal and classification are treated separately).

Frames per second (FPS) is a common metric to evaluate the speed of object detection algorithms. Quantitatively, according to Elgendy[2] the Faster-RCNN algorithm can detect objects in images of 512×512 pixels at a rate of 7 FPS, with a mean average precision (mAP) of 73%.

2b. Single shot detection (SSD)

In 2015, Liu et al.[5] presented an one-stage object-detection architecture called *Single Shot Detection* (SSD). In contrast to R-CNNs, there are not two separate stages, the ROI proposal and the classification, for object detection in images. The algorithm is based on a CNN, which has groups of boxes with a fixed size. The main characteristics of the SSD architecture are:

Network →	Convolutional layers →	NMS
Base network used for object detection: a standard pre-trained network without including the classification layers.	After the base network a series of convolutional filters are added which decrease gradually the size of the images (lower the resolution) to allow detections at multiple scales.	Non-maximum suppression: to unify overlapping bounding boxes and use only one prediction box for each object.

The predicted output from the SSD are four variables describing the bounding box (*x, y, w, h*), one value which scores whether there is an object or not within the bounding box, and the probability of each object class inside the box. Here "class" refers to the type of object, which needs to be pre-trained and labelled beforehand. The bounding boxes are always kept with the same size, contrary to the image dimensions, which are successively resized to assure object detection of multiple scales.

As presented by Elgendy[2], SSD can detect objects in images of 512×512 pixels at a rate of 22 FPS, with a mAP of ~77%.

2c. You only look once (YOLO)

Similar to R-CNNs, there is the state-of-the-art family of three *You only look once* (YOLO) networks: YOLOv1[6], YOLOv2[7] (also called YOLO9000), and YOLOv3[8]. They are all end-to-end deep learning models and were conceived for very fast object detection. However, the accuracy of the object detection is lower than the previous algorithms, in expense of having a near real-time performance. As SSD, YOLO is also a single-stage detector and splits the image into a grid. The architecture of YOLOv3 is described as:

Single network
The network unifies object detection and classification. Feature extraction is carried out by a pre-trained CNN called Darknet-53 with 53 convolutional layers. Additional 53 convolutional layers (that is a total of 106 layers) are added for the object detection task.

Leibniz-Institut für Sonnenphysik (KIS)
Schöneckstr. 6, 79104 Freiburg Germany
Tel.: +49-761-3198-224; Email: SOLARNET-office@leibniz-kis.de

"This project has received funding from the European Union's Horizon 2020 research and innovation programme for the period January 2019 until December 2022 under Grant Agreement No 824135 (SOLARNET)"

YOLOv3 predicts bounding boxes at three different scales. The network is able to do so by downsampling the input image with strides of 32, 16 and 8 pixels. Larger objects will be detected by strongly downsampled images, while smaller objects are extracted from higher resolution images. The three predictions are made at layers 82, 94, and 106. Each prediction box will have the box coordinates (*x, y, w, h*), one value which scores whether there is an object or not within the bounding box, and the probability of each object class inside the box. Since the images are resized to three different scales, for each cell of a grid of *nx*×*ny* pixels we will have three predicted bounding boxes. Finally, boxes below a given threshold are ignored and NMS avoids multiple detections of the same object.
YOLOv3 detects objects in images of 608×608 pixels at a rate of 78 FPS, with a mAP of ~58%, as described by Elgendy[2].

2d. U-Net

Last, but not the least, there is the U-Net[9] architecture, a U-shaped CNN presented in 2015 for fast and precise image segmentation. The algorithm was specifically developed for biomedical images, with the aim of localising and classifying objects. The advantage of U-Net is that it is based on a modified fully convolutional network that works with few training images, as there are not many available in biomedicine. The network architecture can be described as follows:

Downsampling network →	Upsampling network
Typical convolutional network with repetitions of convolutions, rectified linear unit and max pooling operations with stride 2 for gradually downsampling the image.	The so-called expansive path consists of upsampling of the feature map, up-convolution, convolution, and rectified linear unit layers.

In total, 23 convolutional layers are used. The output is a binary segmentation mask for the whole image. Data augmentation was used to increase the number of training data. A study of Vuola et al.[10] in 2019 shows that U-Net has an mAP of about 51%. A different version of R-CNN called *Mask R-CNN*[1] was recently developed with the goal of image segmentation instead of object detection.

3. Conclusions

We presented the most popular state-of-the-art object-detection algorithms based on CNNs with their main properties and architecture. All algorithms can be used for the goal of preparing a toolkit for object identification in high-resolution solar images. The first algorithm which will be tested on our high resolution solar images is YOLOv3, because of its fast performance. This document will be expanded within the SOLARNET timeframe if new suitable algorithms become available.

References

[1]Kaiming He, Georgia Gkioxari, Piotr Dollár and Ross Girshick 2018: *Mask R-CNN*. ArXiv (https://arxiv.org/abs/1703.06870v3)

[2]Mohamed Elgendy 2019: *Deep Learning for Vision Systems*. Manning Early Access Program. Manning Publications. ISBN 9781617296192

[3]Ross Girshick, Jeff Donahue, Trevor Darrell and Jitendra Malik 2014: *Rich feature hierarchies for accurate object detection and semantic segmentation*. ArXiv (https://arxiv.org/abs/1311.2524)

[4]Shaoqing Ren, Kaiming He, Ross Girshick and Jian Sun 2016: *Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks*. ArXiv (https://arxiv.org/abs/1506.01497)

Leibniz-Institut für Sonnenphysik (KIS)
Schöneckstr. 6, 79104 Freiburg Germany
Tel.: +49-761-3198-224; Email: SOLARNET-office@leibniz-kis.de

[5]Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu and Alexander C. Berg 2016: *SSD: Single Shot MultiBox Detector*. In: Leibe B., Matas J., Sebe N., Welling M. (eds) Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science, vol 9905. Springer, Cham (https://arxiv.org/abs/1512.02325)

[6]Joseph Redmon, Santosh Divvala, Ross Girshick and Ali Farhadi 2015: *You Only Look Once: Unified, Real-Time Object Detection*. ArXiv (https://arxiv.org/abs/1506.02640v5)

[7]Joseph Redmon and Ali Farhadi 2017: *YOLO9000: Better, Faster, Stronger*. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). ArXiv (https://arxiv.org/abs/1612.08242v1)

[8]Joseph Redmon and Ali Farhadi 2018: *YOLOv3: An Incremental Improvement*. ArXiv. (https://arxiv.org/abs/1804.02767v1)

[9]Olaf Ronneberger, Philipp Fischer and Thomas Brox 2015: *U-Net: Convolutional Networks for Biomedical Image Segmentation*. In: Navab N., Hornegger J., Wells W., Frangi A. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science, vol 9351. Springer, Cham. (https://arxiv.org/abs/1505.04597v1)

[10]Aarno Oskar Vuola, Saad Ullah Akram and Juho Kannala 2019: *Mask-RCNN and U-net Ensembled for Nuclei Segmentation*. ArXiv (https://arxiv.org/abs/1901.10170v1)

Leibniz-Institut für Sonnenphysik (KIS)
Schöneckstr. 6, 79104 Freiburg Germany
Tel.: +49-761-3198-224; Email: SOLARNET-office@leibniz-kis.de