

Article

Evaluation of Deep Learning-Based Neural Network Methods for Cloud Detection and Segmentation

Stefan Hensel ¹, Marin B. Marinov ^{2,*} , Michael Koch ¹ and Dimitar Arnaudov ^{3,*} 

- ¹ Department for Electrical Engineering, University of Applied Sciences Offenburg, Badstraße 24, D-77652 Offenburg, Germany; stefan.hensel@hs-offenburg.de (S.H.); michael.koch@hs-offenburg.de (M.K.)
- ² Department of Electronics, Technical University of Sofia, 8, Kliment Ohridski Blvd., BG-1756 Sofia, Bulgaria
- ³ Department of Power Electronics, Technical University of Sofia, 8, Kliment Ohridski Blvd., BG-1756 Sofia, Bulgaria
- * Correspondence: mbm@tu-sofia.bg (M.B.M.); dda@tu-sofia.bg (D.A.)

Abstract: This paper presents a systematic approach for accurate short-time cloud coverage prediction based on a machine learning (ML) approach. Based on a newly built omnidirectional ground-based sky camera system, local training and evaluation data sets were created. These were used to train several state-of-the-art deep neural networks for object detection and segmentation. For this purpose, the camera-generated a full hemispherical image every 30 min over two months in daylight conditions with a fish-eye lens. From this data set, a subset of images was selected for training and evaluation according to various criteria. Deep neural networks, based on the two-stage R-CNN architecture, were trained and compared with a U-net segmentation approach implemented by CloudSegNet. All chosen deep networks were then evaluated and compared according to the local situation.

Keywords: machine learning; generation; ground-based sky image; irradiation; load scheduling; photovoltaic power; short-term forecasting; solar irradiance; solar photovoltaics; total cloud cover



Citation: Hensel, S.; Marinov, M.B.; Koch, M.; Arnaudov, D. Evaluation of Deep Learning-Based Neural Network Methods for Cloud Detection and Segmentation. *Energies* **2021**, *14*, 6156. <https://doi.org/10.3390/en14196156>

Academic Editor: Boštjan Blažič

Received: 31 July 2021

Accepted: 22 September 2021

Published: 27 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction and Motivation

Electric power load forecasting has been an integral part of managing electrical energy markets and infrastructure for many decades. Consequently, experiences, regulations, and planning by utilities and independent system operators are the dominant considerations for research and commercial development in this field. The cost of generating power from non-traditional energy sources can be reduced through the integration of solar energy into classical energy supply structures. However, such an integration has its challenges and costs [1,2]. These are mainly caused by the unstable conditions of renewable energy sources such as the dynamic change of sky conditions. Clouds are considered one of the key elements causing fluctuation in solar energy availability [3]. Thus, cloud coverage determines direct and non-direct solar irradiance. Accurate, short-term forecasting of cloud cover is required for a variety of applications, particularly for power generation from photovoltaic solar power plants, as their power output is heavily dependent on sky cloud coverage. The generated power decreases by up to 30 % with a light cloud cover of the sun as compared to cloudless conditions. The yield could decrease by 75 % in the case of sunshine dimmed by dense clouds [4].

The choice of a solar radiation forecast method depends significantly on the periods, which may vary from a few days ahead (intraweek), to a few hours (intraday), or a few minutes (intrahour). Depending on the forecasting application, different time horizons are relevant. The forecasting of the distributed photovoltaic (PV) power generation, which is the focus of this study, requires both intrahour and day-ahead forecasting of solar irradiance [5].

The parameter which is of interest for this study depends on the technology used for power generation. For non-concentrating systems (such as most PV systems), global irradiation (GI) on the inclined surface is required above all.

For different time horizons, however, different approaches are required:

- For relatively long time horizons, of the order of 6 h or more, physics-based models are typically used [6,7].
- Two- to six-hour time horizons use a combination of methods based on observations or predictions of clouds through Numerical Weather Prediction Models (NWPM) and satellite images with information about the optical depth of the cloud and the motion vector of the cloud [6,8].
- For a very short time (<30 min), a range of ground-based imaging techniques were developed for GI using the information on cloud positioning and deterministic models [9,10].

The different solar forecasting techniques and their inputs are summarized in Table 1.

Table 1. Main characteristics and inputs for different solar forecasting approaches [11].

Approach	Sampling Rate	Spatial Resolution	Forecast Horizon	Application
Total-sky imagery	30 s	10–100 m	minutes	Short-term ramps, regulation
Satellite imagery	15 min	1 km	5 h	Load following
NAM ¹ weather model	1 h	12 km	10 days	Unit commitment

¹ North American Mesoscale Model (NAM) is a numerical weather prediction model for short-term weather forecasting.

Numerical weather prediction and up-to-date geostationary satellite-based forecast approaches are restricted in terms of their spatial and temporal resolution and are too imprecise for very short-term forecasts. So, the use of a ground-based sky imager in forecasting is a promising approach as it provides high temporal and spatial cloud cover resolution [12].

Short-term cloud coverage prediction involves two main stages. The first stage includes the detection and segmentation of clouds using available images. The results obtained in the first stage are of great importance, as the quality of the actual prediction (the second stage) depends on the most elaborate representation possible of the clouds. This work presents a camera-based short-term cloud coverage prediction based on machine learning methods. The main contribution is the comparison and evaluation of deep neural network architectures, for instance, segmentation for clouds.

2. Materials and Methods

2.1. Camera-based Cloud Coverage Prediction

Over the last two decades, many studies have proposed various statistical methods for image processing [13,14]. These include various parametric approaches such as Bayesian model averaging [15], or non-homogeneous regression [16], or combined methods such as quantile mapping [17,18].

In recent times, machine learning methods have become increasingly popular in image processing [19]. The work of Taillardat et al. uses quantitative regression forests (QRF) to improve the accuracy of temperature and wind speed forecasts [20]. In [21], an approach based on neural networks to process ECMWF near-surface temperature predictions using QRF as a reference model is presented. Bakker et al. [22] propose several machine learning approaches for the post-processing of Numerical weather prediction (NWP) predictions for solar radiation based on quantum regression, including random forests, gradient amplification, and neural networks.

The detection of clouds in sky imager scenarios is also developing rapidly from classical approaches based on support vector machines and Bayes classifiers, as in [23], to systems employing deep learning techniques. After starting with simple neural structures for remote sensing images, as in [24], current systems are built upon segmentation-based

approaches. These rely on encoder-decoder structures, first proposed in [25] and adapted recently for cloud coverage prediction in [26,27]. The importance and influence of image quality for object detection has been incorporated into deep learning approaches only recently, e.g., in [28].

In contrast to basic segmentation level tasks, the prediction of coverage improves when considering individual cloud objects for tracking and prediction. For this application, segmentation methods are the algorithms of choice. Most prominent and, in fact, ubiquitous in computer vision tasks such as pedestrian recognition is the two-stage approach of Mask R-CNN [29], which allows instance segmentation and bounding box prediction for a given set of classes. A third class of deep learning architectures is the so-called transformer networks, originally invented in the context of speech and natural language recognition. Current research focusses on applying transformers to object detection [30] and segmentation tasks [31].

2.2. Hardware and Imaging

Sky Camera

The present study used a ground-based sky camera to monitor the sky. It is situated at Offenburg University, where it was built based on the optical systems described in [32–34]. It comprises a high-sensitivity CCD-based camera chip combined with a 180° fish-eye lens for full hemispherical imaging. The camera system is combined with additional sensors to measure the actual ground solar irradiance and temperature. The resulting measurement station is shown in Figure 1.



Figure 1. The ground-based camera sensor system positioned at Offenburg University.

Data acquisition was carried out based on a LabVIEW application that stores the captured sky images as an exposure series at a given time interval. The whole hardware setup for image capturing and data storage is described in [35]. The sky imager system was calibrated beforehand based on non-linear distortion models of spherical lenses [36,37].

With classic image processing steps, attempts were made to detect and segment clouds on these images to subsequently be able to make a short-term prediction. It turned out that good detection and segmentation of the clouds is essential for later solar irradiance prediction. With the classical approach, based on a sky illumination prediction and adaptive thresholding as presented in [38], an accuracy of 76.7% could be achieved. In this subsequent work, the aim is to evaluate whether neural networks-based approaches with deep learning are more suitable for detection and segmentation, in the sense of computational speed and accuracy.

The ground-based camera system continuously generates a full hemispherical image. Images are selected from this data stream. Present clouds are marked in the images using pixelwise annotation. The classical system is able to work without a sun disc to block solar rays by using HDR images and a solar position prediction. It is therefore not necessary to

mark the sun or other objects to compare the neural network approach on equal terms. The labeled images are treated as a small database, separated into training and validation sets, only holding back a small sub-set for testing.

2.3. Neural Network-Based Instance Segmentation

Instance segmentation in computer vision has been dominated by deep neural networks since their advent, culminating in the publishing of Mask R-CNN. In this work, we compare and evaluate the power of two prominent neural network architectures, namely Mask R-CNN, which was adapted and trained for the given data set, and Cloud SegNet, an actual state-of-the-art segmentation network already trained on generic cloud data.

2.3.1. Mask R-CNN

Mask R-CNN, although published in a canonical form, allows for variation and adaptation, not only in hyperparameters, but also in more profound ways, such as feature generator architecture, loss functions, or mask sizes.

Framework

Our implementation is based on PyTorch and the Detectron2 archetypes as described in [39]. The structure is highly modular, allowing networks to be adapted and trained for detection and segmentation, the latter as a classic instance, or for panoptic variation.

In this contribution, we use transfer learning and fine-tuning of a pre-trained version. As the clouds do vary in scale and shape, we employed pyramid networks as a backbone to ensure scale invariance, and the data augmentation stack of PyTorch to substantially increase our image database and emulate variations in brightness and color. The following sections briefly explain the structure and adaptation of the chosen network architecture.

Base RCNN-FPN as Backbone

We employ Feature Pyramid Networks (FPN) [40], trained with a focal loss on the MS Coco data set. The FPN backbone is important in detecting clouds on several scales. The network is an object detector with a multi-task loss to allow for class prediction and bounding box estimation. The whole network is basically divided into three components:

The backbone network is a basic convolutional neural network to extract features on different scale levels. The feature maps of several layers are used to ensure scale invariance; the underlying ResNet architecture is reasonably fast for computation.

The classical two-stage approach makes reuse of these features in the Region Proposal Network, which is the second main component of the architecture. The feature maps are used as input and the ROI-align method is used to interpolate regions as possible object proposals for the last main component of the network.

The third stage, the so-called Box Head, consists of fully connected layers that predict the object class and perform a bounding box regression with a multi-task loss, in the case of the R-CNN-FPN base, on the proposed focal loss.

After the post-processing of the detector, non-maxima suppression ensures the efficient pruning of overlapping and wrong object detections. All in all, the RCNN-FPN network produces the typical output of an object detector, namely the most probable class and bounding box, which is exemplarily shown in Figure 2 for the detection of different clouds for a typical output of our system.

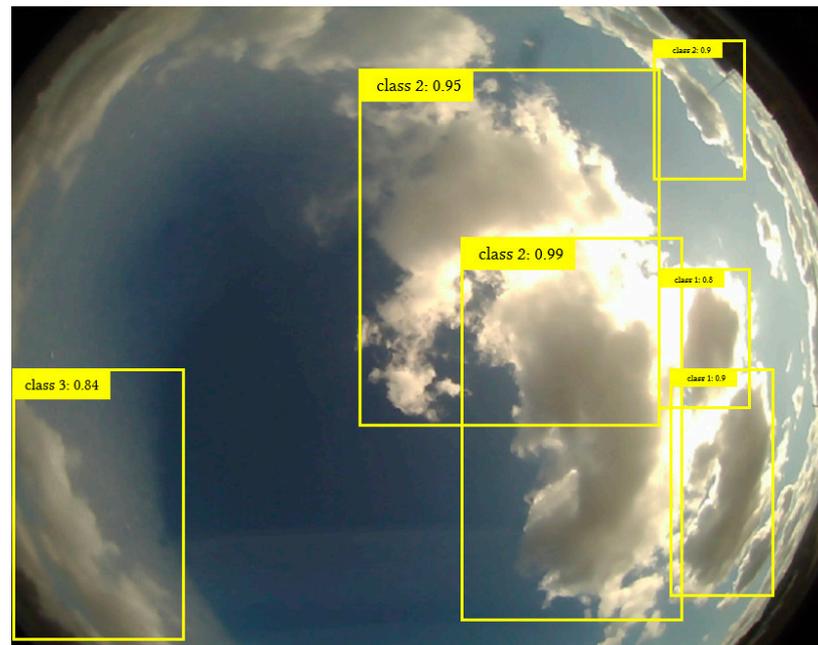


Figure 2. Object detection with retina net-like two-stage detector.

Mask Head

Mask R-CNN is the next step in augmenting the base network described above. An additional third head is added to the object detection-based network. This last head is called a mask head, and estimates a binary mask, based on two subsequent convolutional layers. Training can be performed in one seamless stage, adapting the weights and parameters of all the networks (region proposal, bounding box, class, and mask) simultaneously. This instance of segmentation is shown in Figure 3, adapting the seminal picture in [29] slightly for our case.

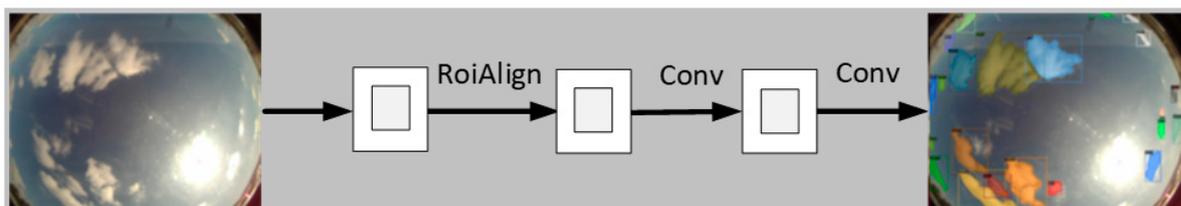


Figure 3. Instance segmentation using Mask R-CNN, based upon the R-CNN detection network (adapted from [29]).

Figure 3 closely summarizes the two preceding paragraphs, depicting the base R-CNN backbone and the subsequent box head, called *class box* within the figure. The two additional convolutional layers for the segmentation step with Mask R-CNN are depicted symbolically to show the upsampling of the detected masks in the final image.

2.3.2. CloudSegNet

The second architecture this contribution evaluates is CloudSegNet. This is a classical encoder–decoder neural network. CloudSegNet focuses on its initial training set on the segmentation of day and night images within a single framework and achieved state-of-the-art results [26]. The network architecture and the associated training data are also open-source [41].

CloudSegNet Architecture

CloudSegNet is a semantic segmentation network specifically designed to segment clouds from the background. In comparison to large image databases and classes, the cloud

segmentation has significantly less texture, structure, and classes, as a plain architecture is chosen. The CloudSegNet architecture has the classical encoder–decoder structure used before U-Net. It is therefore comparable to the fully convolutional nets as described in [42]. This allows for few layers and thus few parameters to be trained. An overview of the architecture is shown in Figure 4, showing the encoder and decoder layers.

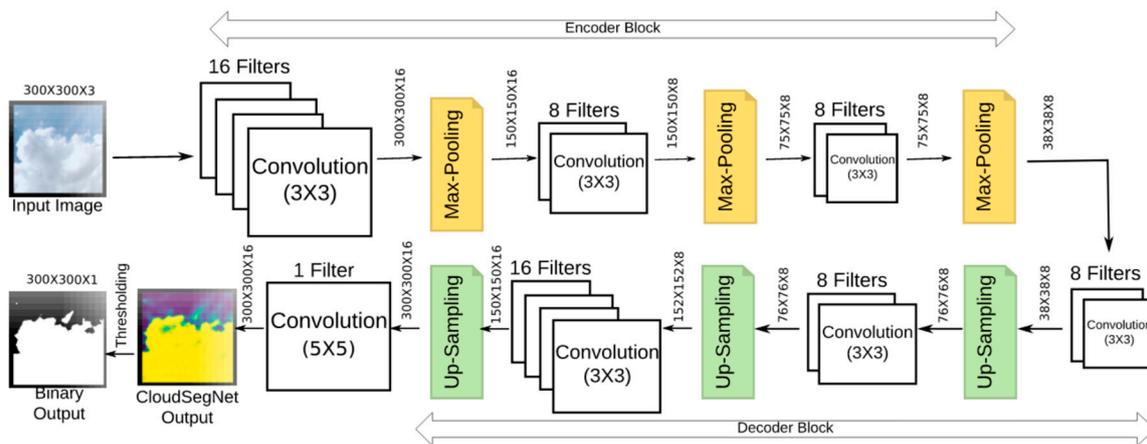


Figure 4. Architectural overview of CloudSegNet (taken from [41]).

Encoder

The network's encoder block is built upon only three layers; the input size of the image is assumed to be 300×300 pixels, limiting the possible resolution. As described in its origins in [43,44], the lower convolution layers encode basic image features, e.g., lines. Later layers set together more and more complex features and can detect clouds in larger receptive fields. The input is condensed into a representation of $38 \times 38 \times 8$ pixels.

Decoder

The subsequent decoder upsamples the image based on the deconvolution operation. The output is upsampled by three layers back to its original size, but only one channel with the probabilities for the classes of each pixel. This output is finally converted to a binary mask by a simple threshold.

3. Experimental Results with Selected Neural Networks

3.1. Creation of the Data Sets

3.1.1. Selection of Images

The given camera systems provide sky images for several months, taken with a frequency of one image every 10 min. Since its installation two years ago, a large amount of data is available that needs to be pre-sorted for the given task. To obtain sensible comparisons, the images were screened and several situations and weather scenarios have been pruned in advance. These include insects on the lens, too many raindrops upon the lens, dirt on the lens, a closed cloud cover, and heavy fog.

Examples of the removed images are shown in Figure 5.

From the remaining data, 76 images were randomly selected for the training data set and 14 for the test set. The training was performed using k-fold cross-validation, with the aim of minimizing the necessary amount of training data. To achieve a greater variation of the displayed clouds, the time interval between selected recordings was set to at least one hour and limited to between 8:00 a.m and 5:00 p.m.

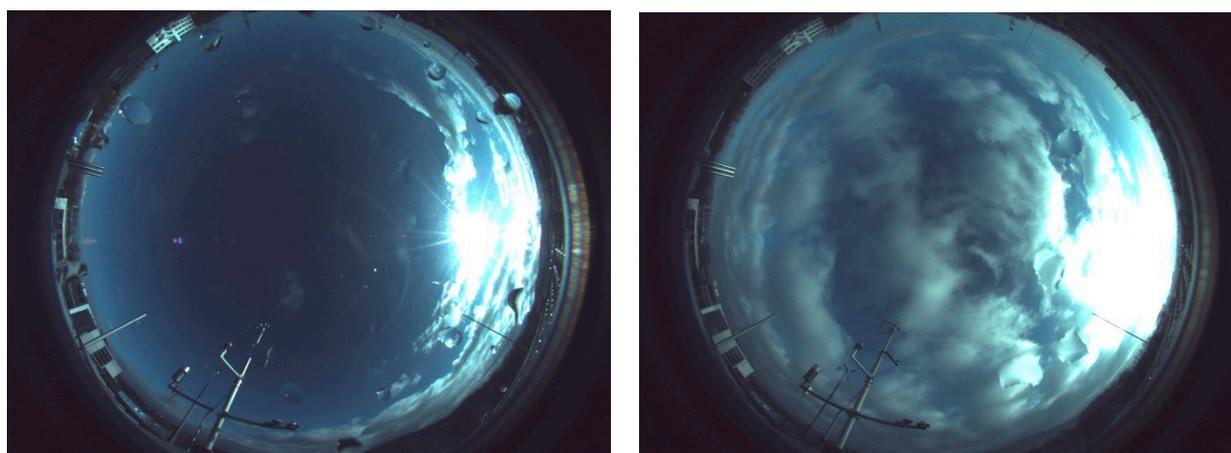


Figure 5. Examples of the removed images with many raindrops upon the lens, dirt on the lens, a closed cloud cover, and heavy fog.

The overall numbers and characteristics of the image database used for training are summarized in Table 2.

Table 2. Overview of the collected image data set.

Number of Training Images	Number of Validation Images	Number of Cloud Objects	The Time Interval between Recording	Time Interval of Recording
67	18	956	60 min	8:00 a.m–5:00 p.m.

If a later contribution uses the segmentation as input, the interval can be easily scaled up. An exemplary image sample is shown in Figure 6.



Figure 6. Exemplary training image with clouds, taken with the ground-based all-sky imager.

3.1.2. Marking the Clouds

To complete instanced segmentation, the time-consuming part is the pixel-wise labeling of the training data. Open-source tools were used and a representative segmentation was completed at the pixel level. Examples are again shown, this time in Figure 7.

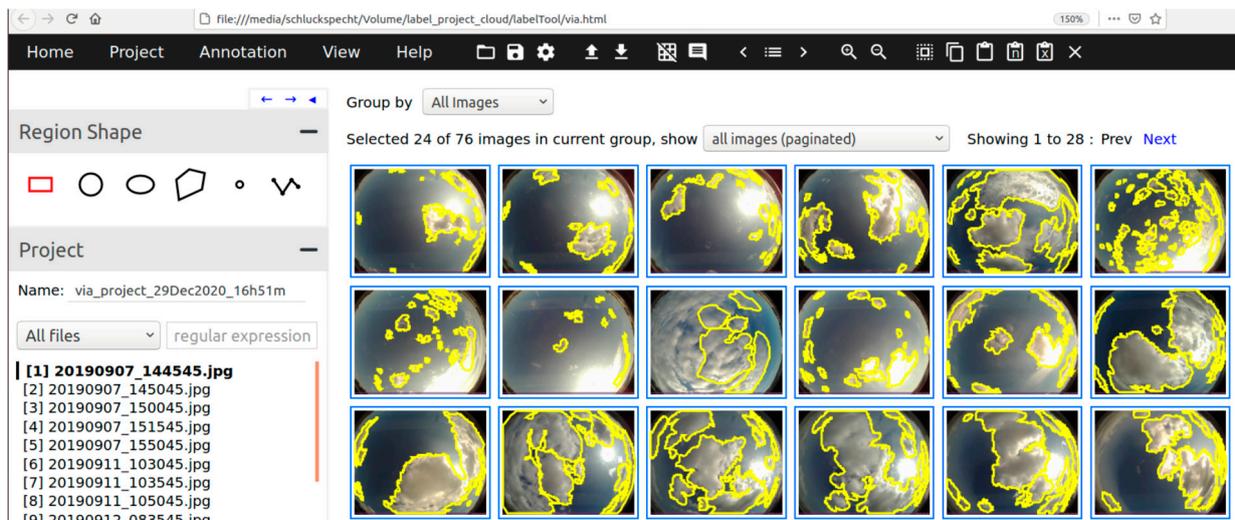


Figure 7. Annotated clouds for training.

For the input, we chose images that were non-rectified and not preprocessed to allow on the one hand for a comparison with CloudSegNet, and on the other hand for a test of the capability of cloud detection under severe optical distortions. The problem arose in the peripheral areas, where clouds are labeled with large difficulties, as shown in Figure 8.

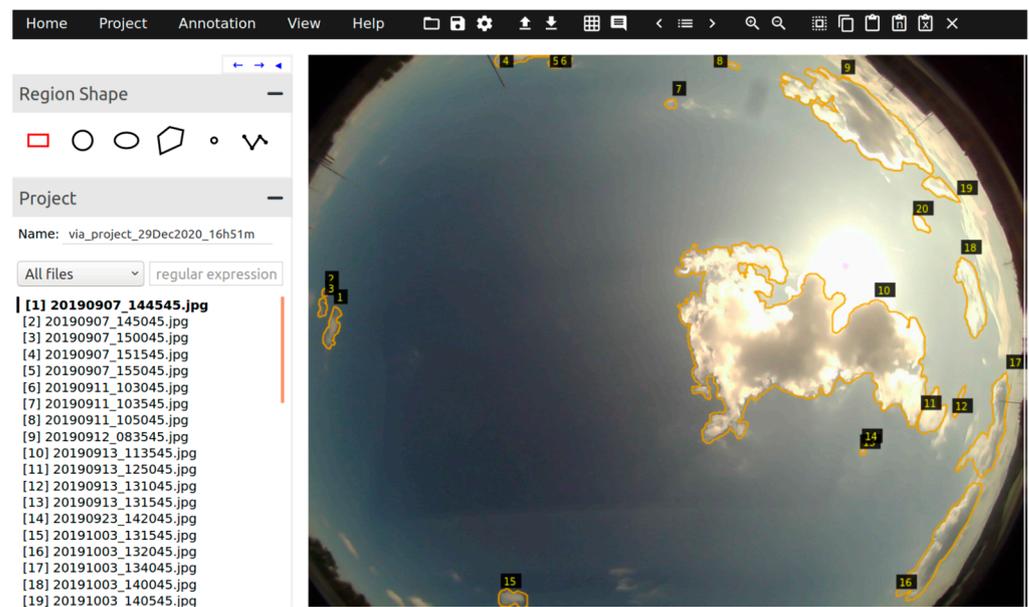


Figure 8. Screenshot of an annotated training data image with clouds in the peripheral area.

The masks are binary in both cases, but Mask R-CNN also uses additional bounding box information generated from the positive areas.

3.2. Mask R-CNN

Given the training and test data, the hyperparameters and overall pipeline for Mask R-CNN had to be set up.

3.2.1. Training

For the training, the hyperparameters were adapted to our problem and data set. Using ADAM optimization [45], the learning rate was scheduled, starting with $\alpha = 0.00025$. Validation and training data were separated with k-fold cross-validation. Convergence of

the training loss could be observed after roughly 10,000 epochs. No further improvement could be achieved by varying the hyperparameters.

3.2.2. Visualization and Qualitative Assessment

After completing the training as described above, the results for the test data set were visually inspected. Results of the network forward pass are shown in Figure 9. On the left side, the input image is shown; the right side depicts results with the object mask and its detection bounding box.

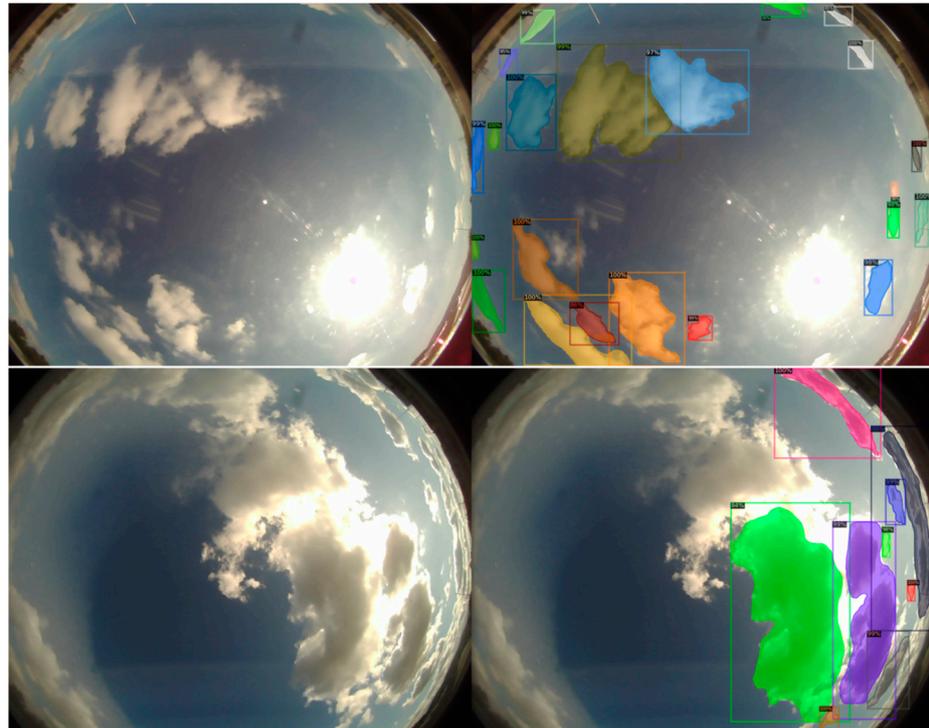


Figure 9. Visualization of the trained Mask R-CNN network using the test set data (left without a mask, right with mask, successful segmentation above, insufficient segmentation below).

Two possible outcomes are shown in the figure. In the upper half, a successful detection and segmentation of the clouds can be seen. It should be noted that the network is somewhat robust concerning disturbances, as the sun was not falsely detected as a cloud. The lower half of the figure shows a very large cloud that was only detected partially. Another problem is that a large portion of the remaining cloud was not detected at all. Our best solution so far is to massively extend the training data set. The quantitative evaluation follows in subsequent sections.

3.2.3. Evaluation

The evaluation was performed with the fine-tuned network for the test set data. As the training loss function is not very helpful in determining the overall quality, we chose the common recall or hit-rate value and the precision or accuracy to assess the quality of the segmentation. As we have a large number of negatives in the image, we calculated the F-score, defined as $2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$, where the precision is the so-called positive prediction value, the quotient of all correctly identified objects (true positive value), and all positively classified objects (true positive cases and false-positive cases). The F-Score combines this value with the recall, or sensitivity, which is the quotient of the true positive values and the combination of true positive and false negative (missed objects) cases. We found that the F-Score is a superior quality measure compared to individual cases, clearly indicating the relevance of the results.

In addition, we detailed the evaluation in further categories: the cloud segmentation was assessed for bounding box accuracy and pixel-wise segmentation, and separated for different sizes of clouds for detection. Finally, large clouds covering roughly a third of the input image are called large, those half the size of large are medium, and the remaining ones are small. Total area means all results summed up. The detailed results are listed in Table 3.

Table 3. Evaluation of Mask R-CNN for quality measures.

Number of Images	Type	Area	Hit Rate	Accuracy	F-Score
67	Box	small	0.124	0.057	0.0781
46	Box	small	0.081	0.022	0.0346
67	Box	medium	0.469	0.354	0.4035
46	Box	medium	0.474	0.355	0.4060
67	Box	large	0.506	0.519	0.5124
46	Box	large	0.607	0.508	0.5531
67	Box	total	0.506	0.633	0.5624
46	Box	total	0.506	0.616	0.5556
67	Seg	small	0.100	0.019	0.0319
46	Seg	small	0.086	0.012	0.0211
67	Seg	medium	0.432	0.314	0.3637
46	Seg	medium	0.431	0.303	0.3558
67	Seg	large	0.541	0.479	0.5081
46	Seg	large	0.538	0.472	0.5028
67	Seg	total	0.457	0.629	0.5294
46	Seg	total	0.454	0.620	0.5242

3.3. CloudSegNet

The CloudSegNet network was used as described in the publication. The network was also fine-tuned with our data set. The framework is based on TensorFlow with Keras, the official repository that was used for the setup.

3.3.1. Preparation of the Data Sets and Training

The CloudSegNet network requires the image data in RGB format and the associated ground truth mask is stored as a binary image. We also used data augmentation with rotation, mirroring, and distorting to enlarge the training image data set.

3.3.2. Visualization

The trained CloudSegNet was visualized as Mask R-CNN, except for the bounding boxes. Exemplary results are shown in Figure 10. The segmentation works well, even for the small database. The upper half shows a near-perfect segmentation; the lower half depicts a problem for misdetecting a bright cloud as the sun.

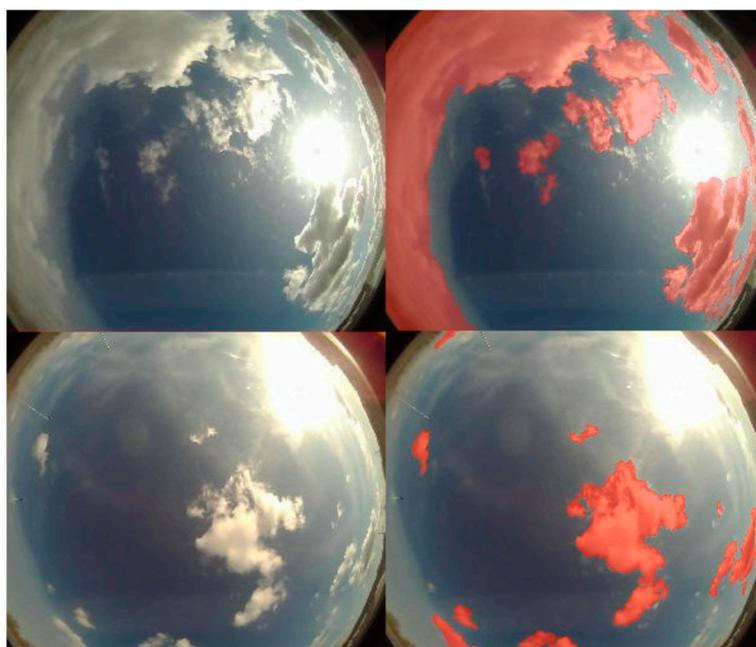


Figure 10. Visualization of the CloudSegNet network using the evaluation data set (left side—input image, right side—segmentation mask).

3.3.3. Evaluation

We used the same quality measures and images as for Mask R-CNN. The results with respect to accuracy and F-Score are far superior to Mask R-CNN. Therefore, we also list the results concerning training progress and complexity. The network could already be used after 500 epochs of fine-tuning, and after 3500 epochs the results are converged. The actual numbers are shown in Table 4.

Table 4. Evaluation of the CloudSegNet in different epochs.

Epoch	Hit rate	Accuracy	F-Score	Error
15	0.6843	0.6136	0.5915	0.1894
500	0.8047	0.7712	0.7726	0.0938
3500	0.8615	0.8428	0.8464	0.0753

4. Conclusions

The evaluation of two different deep neural network approaches showed promising results, albeit with Mask R-CNN lacking in efficiency. As we also have access to a wholly classical machine learning-based approach from [38], a comparison between the two deep learning methods and the pre-neural network method is shown in Table 5. It is worth mentioning that the semantic segmentation has the highest recall and precision, and therefore also the highest F-score. In terms of usage for cloud movement prediction and tracking, this could be used with an additional post-processing step as is needed for the classical approach. Interestingly, the most sophisticated model, Mask R-CNN, performs the worst. As this seems surprising, we conclude that this is due to the lack of training data. CloudSegNet has far fewer parameters to train and is explicitly suited to dealing with binary classes, whereas Mask R-CNN performs the best on large data sets and class numbers.

Table 5. Evaluation of the investigated cloud detection methods.

Method	Hit Rate	Accuracy	F-Score
Classic image processing	0.692	0.767	0.728
Mask R-CNN	0.583	0.500	0.538
CloudSegNet	0.862	0.843	0.846

Another advantage of Mask R-CNN is the bounding box prediction, which allows it to be used as direct input for the subsequent tracking and prediction of individual clouds. The pixel-wise segmentation offers usage for the coverage prediction. Both algorithms are reasonably fast in the evaluation (not training) and outclass the classical approach, which has to generate HDR images out of a small image sequence first.

In conclusion, we propose using CloudSegNet for cloud segmentation and detection but will try to facilitate Mask R-CNN with additional data augmentation techniques, improving the amount of training data.

Another important task to look at is the viability for several different classes of clouds, as there could be cirrostratus and misty layers in contrast to the rather well-defined cumulus, cumulonimbus, or altostratus clouds. This will be tackled with advanced matting techniques and deep learning, as presented in [46].

Author Contributions: Conceptualization, S.H. and M.B.M.; methodology, S.H. and M.K.; software, M.K.; validation, S.H.; investigation, S.H. and D.A.; resources, S.H.; writing—original draft preparation, S.H. and M.B.M.; writing—original draft preparation, S.H. and M.B.M.; writing—review and editing, S.H. and M.B.M.; visualization, D.A.; funding acquisition, D.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by the Bulgarian National Science Fund in the scope of the project “Exploration the application of statistics and machine learning in electronics” under contract number KII-06-H42/1.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Sauter, P.S.; Karg, P.; Kluwe, M.; Hohmann, S. Load Forecasting in Distribution Grids with High Renewable Energy Penetration for Predictive Energy Management Systems. In Proceedings of the 2018 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe), Sarajevo, Bosnia and Herzegovina, 21–25 October 2018.
- Maurer, J.; Sauter, P.S.; Kluwe, M.; Hohmann, S. Optimal energy management of low level multi-carrier distribution grids. In Proceedings of the 2016 IEEE International Conference on Power System Technology (POWERCON), Wollongong, NSW, Australia, 28 September–1 October 2016.
- Kim, M.; Kim, H.; Jung, J. A Study of Developing a Prediction Equation of Electricity Energy Output via Photovoltaic Modules. *Energies* **2021**, *14*, 1503. [[CrossRef](#)]
- Sun, S.; Ernst, J.; Sapkota, A.; Ritzhaupt-Kleissl, E.; Wiles, J.; Bamberger, J.; Chen, T. Short term cloud coverage prediction using ground based all sky imager. In Proceedings of the 2014 IEEE International Conference on Smart Grid Communications (SmartGridComm), Venice, Italy, 3–6 November 2014.
- Lorenz, E.; Hurka, J.; Heinemann, D.; Beyer, H.G. Irradiance Forecasting for the Power Prediction of Grid-Connected Photovoltaic Systems. *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* **2009**, *2*, 2–10. [[CrossRef](#)]
- Hammer, A.; Heinemann, D.; Hoyer, C.; Kuhlemann, R.; Lorenz, E.; Müller, R.; Beyer, H.G. Solar energy assessment using remote sensing technologies. *Remote. Sens. Environ.* **2003**, *86*, 423–432. [[CrossRef](#)]
- Perez, R.; Kivalov, S.; Schlemmer, J.; Hemker, K.; Renné, D.; Hoff, T.E. Validation of short and medium term operational solar radiation forecasts in the US. *Sol. Energy* **2010**, *84*, 2161–2172. [[CrossRef](#)]
- Takeyoshi, K. Chapter 4—Prediction of photovoltaic power generation output and network operation. In *Integration of Distributed Energy Resources in Power Systems*; Funabashi, T., Ed.; Academic Press: Cambridge, MA, USA, 2016; pp. 77–108.
- Marquez, R.; Gueorguiev, V.; Coimbra, C. Forecasting solar irradiance using sky cover indices. *ASME J. Sol. Energy Eng.* **2013**, *135*, 011017. [[CrossRef](#)]

10. Ghonima, M.S.; Urquhart, B.; Chow, C.W.; Shields, J.E.; Cazorla, A.; Kleissl, J. A method for cloud detection and opacity classification based on ground based sky imagery. *Atmospheric Meas. Tech.* **2012**, *5*, 2881–2892. [CrossRef]
11. Kleissl, J. *Solar Energy Forecasting and Resource Assessment*; Academic Press: Cambridge, MA, USA, 2013.
12. Chow, C.W.; Urquhart, B.; Lave, M.; Dominguez, A.; Kleissl, J.; Shields, J.; Washom, B. Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed. *Sol. Energy* **2011**, *85*, 2881–2893. [CrossRef]
13. Williams, R.M.; Ferro, C.A.T.; Kwasniok, F. A comparison of ensemble post-processing methods for extreme events. *Q. J. R. Meteorol. Soc.* **2014**, *140*, 1112–1120. [CrossRef]
14. Su, X.; Li, T.; An, C.; Wang, G. Prediction of Short-Time Cloud Motion Using a Deep-Learning Model. *Atmosphere* **2020**, *11*, 1151. [CrossRef]
15. Raftery, A.E.; Balabdaoui, F.; Gneiting, T.; Polakowski, M. Using Bayesian Model Averaging to Calibrate Forecast Ensembles. *Mon. Weather. Rev.* **2003**, *133*, 1155–1174. [CrossRef]
16. Gneiting, T.; Raftery, A.E.; Westveld, A.H.; Goldman, T. Calibrated Probabilistic Forecasting Using Ensemble Model Output Statistics and Minimum CRPS Estimation. *Mon. Weather. Rev.* **2005**, *133*, 1098–1118. [CrossRef]
17. Hamill, T.M.; Scheuerer, M. Probabilistic Precipitation Forecast Postprocessing Using Quantile Mapping and Rank-Weighted Best-Member Dressing. *Mon. Weather. Rev.* **2018**, *146*, 4079–4098. [CrossRef]
18. Baran, Á.; Lerch, S.; El Ayari, M.; Baran, S. Machine learning for total cloud cover prediction. *Neural Comput. Appl.* **2021**, *33*, 2605–2620. [CrossRef]
19. Berthomier, L.; Pradel, B.; Perez, L. Cloud Cover Nowcasting with Deep Learning. In Proceedings of the 2020 Tenth International Conference on Image Processing Theory, Tools and Applications (IPTA), Paris, France, 9–12 November 2020.
20. Taillardat, M.; Mestre, O.; Zamo, M.; Naveau, P. Calibrated Ensemble Forecasts Using Quantile Regression Forests and Ensemble Model Output Statistics. *Mon. Weather. Rev.* **2016**, *144*, 2375–2393. [CrossRef]
21. Rasp, S.; Lerch, S. Neural Networks for Postprocessing Ensemble Weather Forecasts. *Mon. Weather. Rev.* **2018**, *146*, 3885–3900. [CrossRef]
22. Bakker, K.; Whan, K.; Knap, W.; Schmeits, M. Comparison of statistical post-processing methods for probabilistic NWP forecasts of solar radiation. *Sol. Energy* **2019**, *191*, 138–150. [CrossRef]
23. Cheng, H.-Y.; Lin, C.-L. Cloud detection in all-sky images via multi-scale neighborhood features and multiple supervised learning techniques. *Atmospheric Meas. Tech.* **2017**, *10*, 199–208. [CrossRef]
24. Shi, M.; Xie, F.; Zi, Y.; Yin, J. Cloud detection of remote sensing images by deep learning. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016.
25. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Cham, Switzerland, 2015; pp. 234–241.
26. Dev, S.; Nautiyal, A.; Lee, Y.H.; Winkler, S. CloudSegNet: A Deep Network for Nychthemeron Cloud Image Segmentation. *IEEE Geosci. Remote. Sens. Lett.* **2019**, *16*, 1814–1818. [CrossRef]
27. Li, Z.; Shen, H.; Wei, Y.; Cheng, Q.; Yuan, Q. Cloud detection by fusing multi-scale convolutional features. *ISPRS Ann. Photogramm. Remote. Sens. Spat. Inf. Sci.* **2018**, *IV-3*, 149–152. [CrossRef]
28. Varga, D. Multi-Pooled Inception Features for No-Reference Image Quality Assessment. *Appl. Sci.* **2020**, *10*, 2186. [CrossRef]
29. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. *arXiv* **2017**, arXiv:1703.06870.
30. Carion, N.; Massa, F.; Synnaeve, G.; Usunier, N.; Kirillov, A.; Zagoruyko, S. End-to-End Object Detection with Transformers. In *Computer Vision—ECCV 2020. ECCV 2020. Lecture Notes in Computer Science*; Vedaldi, A., Bischof, H., Brox, T., Frahm, J., Eds.; Springer: Cham, Switzerland, 2020; Volume 12346.
31. Zheng, S.; Lu, J.; Zhao, H.; Zhu, X.; Luo, Z.; Wang, Y.; Fu, Y.; Feng, J.; Xiang, T.; Torr, P.H.; et al. Rethinking Semantic Segmentation from a Sequence-to-Sequence Perspective with Transformers. *arXiv* **2020**, arXiv:2012.1584.
32. Kleissl, J.; Urquhart, B.; Ghonima, M.; Dahlin, E.; Nguyen, A.; Kurtz, B.; Chow, C.W.; Mejia, F.A. *Sky Imager Cloud Position Study Field Campaign Report*; University of California: San Diego, CA, USA, 2016.
33. Cazorla, A.; Olmo, F.J.; Alados-Arboledas, L. Development of a sky imager for cloud cover assessment. *J. Opt. Soc. Am. A* **2007**, *25*, 29–39. [CrossRef]
34. Gauchet, C.; Blanc, P.; Espinar, B.; Charbonnier, B.; Demengel, D. Surface solar irradiance estimation with low-cost fish-eye camera. In *Workshop on Remote Sensing Measurements for Renewable Energy*; HAL CCSD: Risoe, Denmark, 2012.
35. Kömm, T. *Development of a Cloud Camera for Short-Term Solar Energy Prediction*; University of Offenburg: Offenburg, Germany, 2016.
36. Hensel, S.; Marinov, M.B.; Schwarz, R. Fisheye Camera Calibration and Distortion Correction for Ground Based Sky Imagery. In Proceedings of the 2018 IEEE XXVII International Scientific Conference Electronics—ET, Sozopol, Bulgaria, 13–15 September 2018.
37. Hu, X.; Zheng, H.; Chen, Y.; Chen, L. Dense crowd counting based on perspective weight model using a fisheye camera. *Optik* **2015**, *126*, 123–130. [CrossRef]
38. Hensel, S.; Marinov, M.B.; Schwarz, R.; Topalov, I. Ground Sky Imager Based Short Term Cloud Coverage Prediction. In Proceedings of the FABULOUS 2019—4th EAI International Conference on Future Access Enablers of Ubiquitous and Intelligent Infrastructures, Sofia, Bulgaria, 28–29 March 2019.
39. Wu, Y. Detectron2. 2019. Available online: <https://github.com/facebookresearch/detectron2> (accessed on 25 June 2021).

40. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
41. Soumyabrata. 2019. Available online: <https://github.com/Soumyabrata/CloudSegNet> (accessed on 21 June 2020).
42. Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [[CrossRef](#)] [[PubMed](#)]
43. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
44. Ren, S.; He, K.; Zhang, X.; Sun, J. Deep residual learning for image recognition. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* **2016**, 770–778. Available online: https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html (accessed on 21 June 2020).
45. Diederik, P.; Ba, K.; Ba, J. ADAM: A Method for Stochastic Optimization. *arXiv* **2017**, arXiv:1412.6980.
46. Xu, N.; Price, B.; Cohen, S.; Huang, T. Deep Image Matting. *Proc. 2017 IEEE Conf. Comput. Vis. Pattern Recognit.* **2017**, 2970–2979. Available online: https://openaccess.thecvf.com/content_cvpr_2017/html/Xu_Deep_Image_Matting_CVPR_2017_paper.html (accessed on 21 June 2020).