Journal of Computer and Communications, 2018, 6, 54-63
http://www.scirp.org/journal/jcc
ISSN Online: 2327-5227
ISSN Print: 2327-5219

Research on Pedestrian Detection Technology Based on MSR and Faster R-CNN

Xueyun Zhao, Chaoju Hu

College of Control and Computer Engineering, North China Electric Power University, Baoding, China
Email: 983417471@qq.com

How to cite this paper: Zhao, X.Y. and Hu, C.J. (2018) Research on Pedestrian Detection Technology Based on MSR and Faster R-CNN. *Journal of Computer and Communications*, **6**, 54-63.
https://doi.org/10.4236/jcc.2018.67006

Received: July 1, 2018
Accepted: July 27, 2018
Published: July 30, 2018

Abstract

In order to avoid the problem of poor illumination characteristics and inaccurate positioning accuracy, this paper proposed a pedestrian detection algorithm suitable for low-light environments. The algorithm first applied the multi-scale Retinex image enhancement algorithm to the sample pre-processing of deep learning to improve the image resolution. Then the paper used the faster regional convolutional neural network to train the pedestrian detection model, extracted the pedestrian characteristics, and obtained the bounding boxes through classification and position regression. Finally, the pedestrian detection process was carried out by introducing the Soft-NMS algorithm, and the redundant bounding box was eliminated to obtain the best pedestrian detection position. The experimental results showed that the proposed detection algorithm achieves an average accuracy of 89.74% on the low-light dataset, and the pedestrian detection effect was more significant.

Keywords

Deep Learning, Pedestrian Detection, Region-Based Convolutional Neural Network, Image Enhancement, Non-Maximum Suppression

1. Introduction

Pedestrian detection refers to a research problem in which a pedestrian is judged in a specific scene and a specific position of the pedestrian is given. In recent years, it has been widely used in scenes such as video surveillance, vehicle assisted driving, and intelligent robots. Because pedestrians are easily affected by occlusion, background, lighting and other factors, pedestrian detection becomes a challenging and hot issue. Therefore, optimizing pedestrian detection has important implications.

Pedestrian detection can be roughly divided into three parts: feature extrac-

tion, classification and non-maximum suppression [1]. Currently, there are many technologies that are used in pedestrian detection. Dala *et al.* [2] proposed the use of the Histogram of Oriented Gradient (HOG) feature, combined with a linear SVM classifier to achieve pedestrian detection. One of the more important features in recent years is the Deformable Part Mode (DPM) proposed by Felenszwalb *et al.* [3]. Because DPM takes into account the internal structure of the target, it can well detect pedestrians with different postures and can distinguish between targets and backgrounds. Although the above detection method improves the object detection method to different extents, the hand-designed features are not very robust to target diversity changes in complex scenes. The biggest feature of the Convolutional Neural Network (CNN) is that it can automatically learn object features through a large amount of data, and send this feature into the classifier to obtain excellent classification performance. Sermanet *et al.* [4] proposed the application of convolutional neural networks to pedestrian detection. The image features extracted by deep learning are far superior to those extracted by traditional methods. The Faster R-CNN model proposed by Ren *et al.* [5] in 2015 achieved the highest accuracy of current target detection. Relative to R-CNN and Fast R-CNN, Faster R-CNN truly implements an end-to-end target detection framework, which further reduces the generation of bounding box time. The depth convolution feature proposed by the Cascaded Boosted Forest direct training area suggestion network is used in [6], which has achieved good results in pedestrian detection. In the paper [7], the regional proposal network is used to generate proposals, and the Faster R-CNN network framework is used to implement pedestrian detection in nighttime infrared images. On the basis of the Faster R-CNN model, the paper [8] incorporates the dark channel dehazing algorithm to effectively improve the pedestrian detection effect in harsh environments.

The above method shows that deep learning has been greatly improved in the field of pedestrian detection compared with the traditional method, but for the target shot in low light and small in the image, the Region-of-Interest (RoI) pooling layer has no distinguishing ability for features extracted at low resolution. In addition, the Faster R-CNN model uses the traditional NMS algorithm (Greedy-NMS) to eliminate redundant detection frames. This algorithm is based on a greedy strategy. If an object is within the preset overlap threshold, it may not be detected.

For the problems of pedestrian detection under low light, the main work of this paper is as follows: 1) For the problem of difficult feature extraction in low-light environment, sample pre-processing is performed before the Faster R-CNN model training by multi-scale Retinex image enhancement method; 2) For the problem of inaccurate positioning of the bounding box, the Soft-NMS algorithm is more effective in improving the detection accuracy; 3) In order to verify the performance of the proposed algorithm, the low-light pedestrian image dataset with annotations is trained under the algorithm to evaluate the performance of the proposed algorithm.

2. Algorithm Principle

2.1. Multi-Scale Retinex Image Enhancement Algorithm (MSR)

Since the illumination condition is an important factor affecting the performance of pedestrian detection, it is the key to the feature extraction of pedestrian detection. Under different illumination conditions, the results of pedestrian detection will be different. Especially for dealing with relatively small pedestrians in low light, the resolution of the feature map extracted by the ROI pooling layer in the Faster R-CNN algorithm will be relatively low, so this paper takes a multi-scale Retinex image enhancement method on the data set to preprocess the sample to improve the resolution, thereby improving the accuracy of pedestrian detection.

The Retinex theory proposed is mainly to reduce the image or remove the influence of incident light, and then obtain the reflection characteristics of the object in the scene, and only retain the information in the original image that can reflect the basic features of the object, so as to achieve the purpose of image enhancement. The mathematical expression of Retinex theory can be expressed as:

$$I(x,y) = R(x,y) \cdot L(x,y) \tag{1}$$

where $I(x,y)$ represents the information of the initial image captured by the camera, $L(x,y)$ represents the illumination component of the incident light in the scene, and $R(x,y)$ reflects the reflected component of the essential information of the image.

According to the human eye, the perception of the brightness of the acquired image and the change of the brightness exhibit a logarithmic nonlinear relationship. For the formula (1), the illumination component in the scene is separated from the acquired image by taking the logarithm, that is, the relationship is [9]:

$$\log\left[R(x,y)\right] = \log\left[I(x,y)\right] - \log\left[L(x,y)\right] \tag{2}$$

$$r(x,y) = \log\left[I(x,y)\right] - \log\left[G(x,y) * I(x,y)\right] \tag{3}$$

where * represents a convolution operation.

$$G(x,y) = \lambda e^{-\frac{x^2+y^2}{c^2}} \tag{4}$$

In the formula (4), c represents a Gaussian scale, which represents a scale for satisfying $\iint G(x,y)dxdy = 1$.

In order to ensure rich image feature information and low color distortion, this paper uses the multi-scale Retinex algorithm (MSR) [10], which can be expressed as:

$$r_{\text{MSRV}} = \sum_{j=1}^{J} W_j \left\{ \log I_v(x,y) - \log\left[G_j(x,y) * I_v(x,y)\right] \right\} \tag{5}$$

where v represents the color channel, and this article $v = 3$, which represents the color image of the R, G, and B channels. r_{MSRV} represents the processing results of the v channels, j represents the number of scales, that is, the number of Gaus-

sian surround functions, W_j represents the weight corresponding to each scale. In general, the value of j is 3, because the execution time of the oversized algorithm will increase, the effect will not be significantly improved. Finally, the pixel value of the reflection information image of each channel obtained by the MSR algorithm is normalized to between 0 and 255 by using the formula (6).

$$R_i(x,y) = 255 \cdot \frac{r_{\text{MSRV}}(x,y) - r_{\min}}{r_{\max} - r_{\min}} \tag{6}$$

$r_{\max}$ and $r_{\min}$ represent the maximum values and minimum values of the reflection information for each channel.

The processing flow chart of the MSR algorithm can be used as shown in Figure 1.

2.2. Faster R-CNN

The main process of the Faster R-CNN algorithm for detecting pedestrians is: first, the network inputs a picture, generates a series of proposals through the Region Proposal Network (RPN), and then sends the picture and the proposals together to the FastR-CNN. The network outputs the final pedestrian test results. The detection process of Faster-RCNN is shown in Figure 2.

The algorithm consists of two major modules:

1) Region Proposal Network

RPN is a full convolutional network. The network consists of a convolutional layer, an intermediate layer, a classification layer, and a regression layer. The convolutional layer is consistent with Fast R-CNN, and the intermediate layer input is fully connected to the $n \times n$ region on the last layer of the convolutional layer feature map. The network traverses the convolution extraction feature using a $n \times n$-sized sliding window, encoding each convolution map location as a low-dimensional feature vector. This paper uses the 512-dimensional VGG-16 network structure, as shown in Table 1. The position in each window corresponds to k anchors of different scales and aspect ratios simultaneously sampled. In this paper, the value of k is 9. The output of the network is a classification layer and a regression layer, indicating the category score of the image area and the position correction of the bounding box.

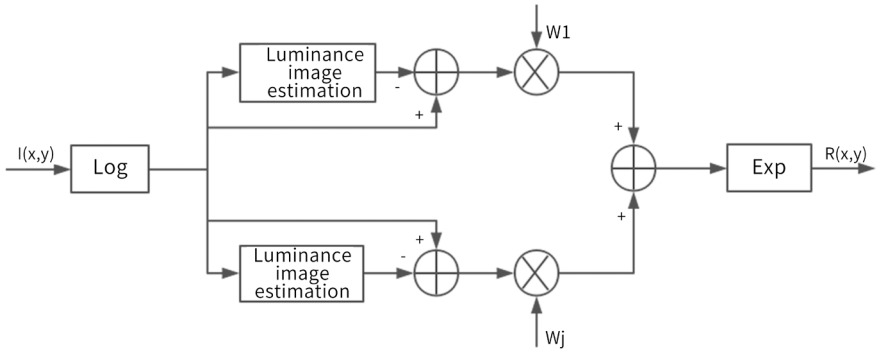

Figure 1. MSR algorithm flow chart.

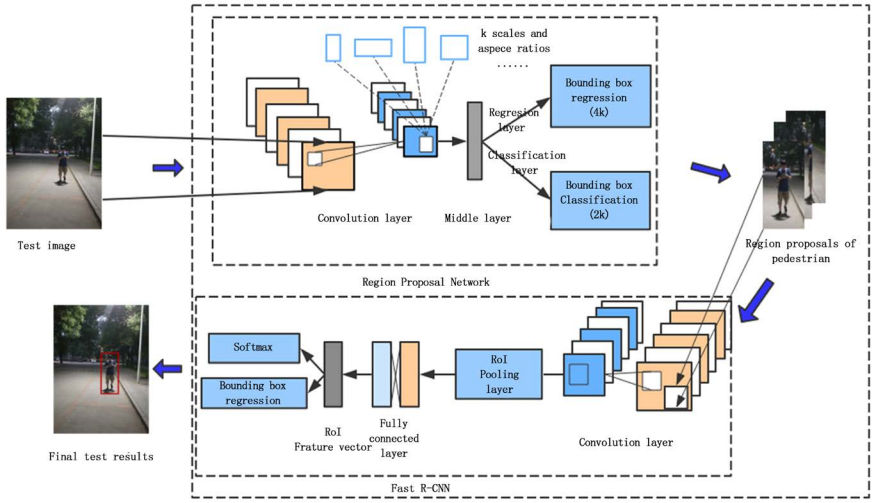

Figure 2. Faster R-CNN detection flow chart.

Table 1. VGG-16 network structure table.

Type/Layer	Number of kernel	Kernel size/Stride
Conv1_x/2	64	3 × 3/1
Maxpool		2 × 2/2
Conv2_x/2	128	3 × 3/1
Maxpool		2 × 2/2
Conv3_x/3	256	3 × 3/1
Maxpool		2 × 2/2
Con4_x/3	512	3 × 3/1
Maxpool		2 × 2/2
Conv5_x/3	512	3 × 3/1

2) Fast R-CNN

After the proposals are obtained by the RPN output, it is regarded as the input of another Fast R-CNN. The RoI pooling layer uses the proposal window to extract the proposal feature from the feature map and send it to the subsequent full connection and softmax network for classification. Through the accurate image classification and positioning correction again, the final target detection result is obtained.

2.3. Non-Maximum Suppression (NMS)

A to-be-detected image has initially detected multiple quasi-targets in the image, but due to the influence of scale and traversal, there will be multiple Bounding Boxes (BB) in the same target, so it is necessary to suppress the extra bounding boxes and find the best detection location. Therefore, non-maximum suppression is a post-processing process for pedestrian detection. The purpose is to remove redundant bounding boxes and retain the best one. However, the biggest

problem with the traditional NMS algorithm is that it forces the scores of adjacent bounding boxes to zero. In this case, if a pedestrian appears in the overlapping area, it will cause the detection of the pedestrian to fail and reduce the average precision (AP).

For this problem with NMS, we introduce the Soft-NMS [11] algorithm to suppress redundant information in the bounding box.

The formula for the NMS algorithm and Soft-NMS is as follows:

$$s_i = \begin{cases} s_i & iou(M, b_i) < N_t \\ 0 & iou(M, b_i) \geq N_t \end{cases} \tag{7}$$

$$s_i = \begin{cases} s_i & iou(M, b_i) < N_t \\ s_i\left(1 - iou(M, b_i)\right) & iou(M, b_i) \geq N_t \end{cases} \tag{8}$$

when IoU is less than the threshold N_t, the detection score is s_i; when IoU is greater than the threshold N_t, the score is 0. This process is applied recursively to the remaining bounding boxes. The Soft-NMS attenuates the detection score of the non-maximum bounding box instead of completely removing it. After IoU is greater than the threshold N_t, the score value s_i is $s_i\left(1 - iou(M, b_i)\right)$. Simple changes are made in the traditional NMS algorithm, and without additional parameters, the detection accuracy can be improved by about 1.2% and the detection speed.

3. Experimental

3.1. Data Set

In the experiment, this paper selects the pictures collected under low light as the data set. This data set has various scenes, including low-light pictures in various pedestrian situations such as subway stations, roads and shopping malls. The data set uses 8000 images as the training model, in which 6400 pictures are randomly selected as the training set, and the remaining 1600 pictures are used as the test set. Name the image according to the format, write the program in Python, mark the real bounding box in each image, and save the bounding coordinate information of the label to the .xml file. An example of an experimental data set is shown in **Figure 3**.

3.2. Experimental Setup and Evaluation Criteria

This paper uses the most advanced pedestrian detection method Faster R-CNN model. This pedestrian detection model is implemented on the popular deep learning framework Tensor Flow. For the VGG-16 network pre-trained by Image Net, initialize both RPN and Fast R-CNN, and then use the data set to train the system. In the first training stage, since the previous layers usually extract very similar pixel-level features, no adjustments are needed. We only need to train conv3_1 and higher of the VGG-16 network. In the second stage, only the conv5_3 layer and higher layers in the RPN and the fully-connected layers in the Fast R-CNN are tuned. The system is trained using Stochastic Gradient Descent

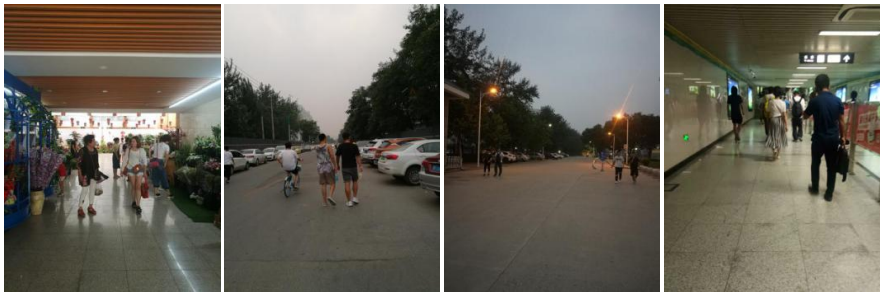

Figure 3. Example diagram of the experimental data set.

(SGD) with a momentum of 0.9 and a weight decay of 0.0005. The layer parameters are updated at an initial learning rate of 0.001. After 50,000 iterations, the learning rate lowered to 1/10 of the current rate, and the total number of iterations is 100,000.

We refer to the model obtained by training as model 1, and then continue to train the sample set processed by the MSR image enhancement algorithm to obtain model 2, and compare the models 1 and 2 through the pedestrian test set. In addition, we introduce Soft-NMS into the model and compare it with the traditional NMS algorithm. The experimental flow chart is shown in Figure 4.

3.3. Experimental Results and Analysis

The comparison of pre-processed and unprocessed images using the multi-scale Retinex image enhancement algorithm is shown in Figure 5.

It can be observed from the figure that the pedestrian features in the image processed by the MSR algorithm are more prominent, and the picture quality is significantly improved in both brightness and contrast.

The performance improvement of soft-NMS relative to NMS under different overlapping thresholds was found through experiments. With the increase of the overlap threshold and the retrieval rate, the soft-NMS has a greater improvement in accuracy. As shown in Figure 6.

Finally, our proposed improved model achieved an average accuracy of 89.74% in pedestrian detection under low light as shown in Table 2. Observations show that the proposed method achieves the desired test results on the test set. Compared with the original Faster R-CNN, our improved algorithm improves performance by 1.5%, which is better than the original algorithm. This is because the pedestrian feature is easier to extract after using the MSR image enhancement algorithm in front of the Faster R-CNN network. In addition, the introduction of the Soft-NMS algorithm enables the detection rate of overlapping pedestrians to be effectively improved and the pedestrian position to be more accurate. Some examples of pedestrian detection using the proposed method are shown in Figure 7.

4. Conclusion

The improved model presented in this paper has a higher accuracy than the

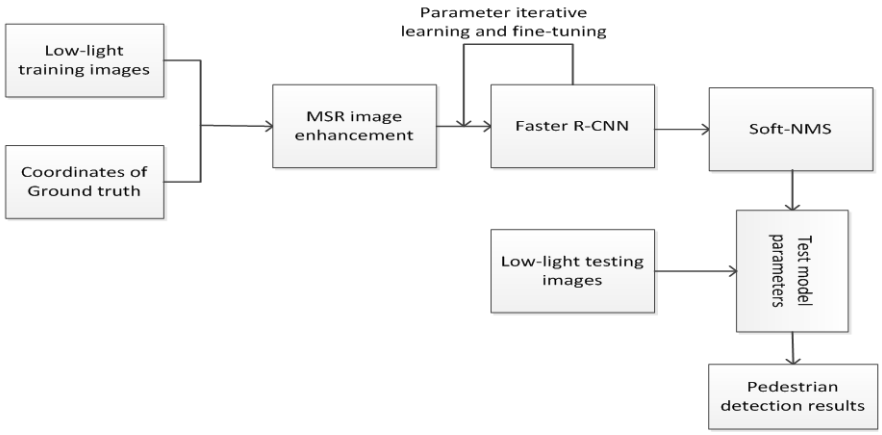

Figure 4. Flow char of experiment.

Figure 5. Comparison of MSR results.

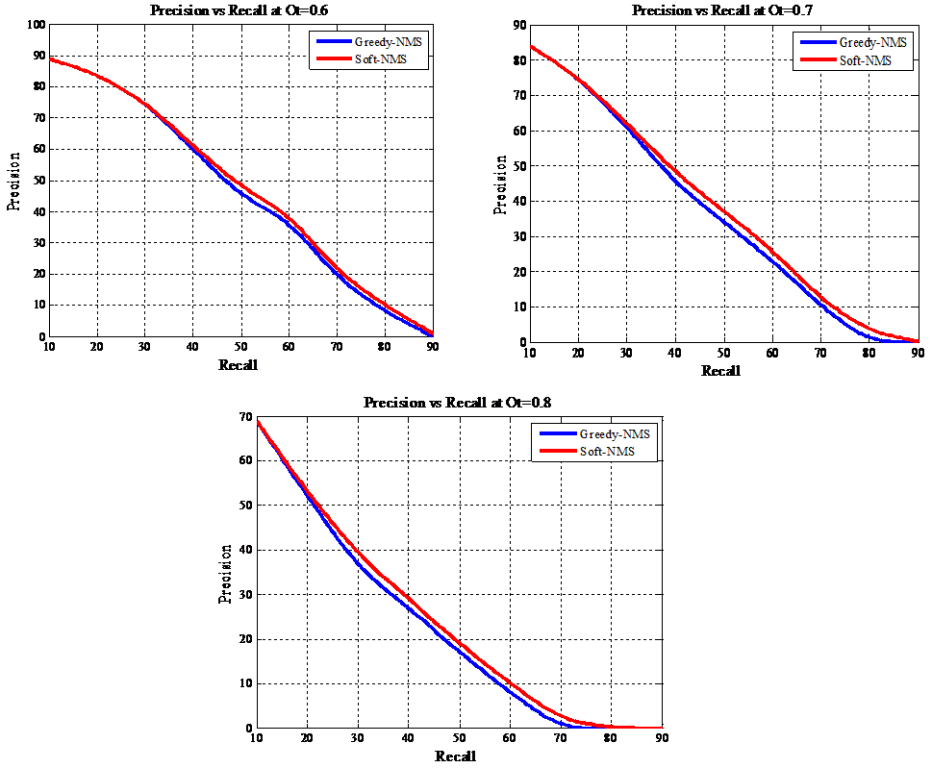

Figure 6. Precisionvs Recall at multiple overlap thresholds (Ot).

Table 2. Accuracy of different methods.

Method	AP (%)
Faster R-CNN	88.23
Faster R-CNN + MSR	88.62
MSR + Faster R-CNN + Soft-NMS	89.74

Figure 7. Examples of pedestrian detection.

original Faster R-CNN model. This is because after the image is enhanced by the multi-scale Retinex algorithm, the contrast of the image is improved, the difference between the target and the background area is more obvious, and the target outline is also clearer. Therefore, the network can better extract the characteristics of pedestrians. In addition, based on the Faster RCNN, the Soft-NMS algorithm makes it possible to obtain higher accuracy for pedestrian detection with higher overlap and small scale. The results show that the detection effect of the model is more significant, but the detection speed needs to be improved. How to improve the detection speed is the main direction of our next research.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Chen, J.H. and Ye, X.N. (2015) Improvement of Non-Maximum Suppression in Pedestrian Detection. *Journal of East China University of Science and Technology*, **41**, 371-378.

[2] Dalal, N. and Triggs, B. (2005) Histograms of Oriented Gradients for Human Detection. 2005 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (*CVPR*'05) San Diego, 20-25 June 2005, 886-893. https://doi.org/10.1109/CVPR.2005.177

[3] Felzenszwalb, P.F., Girshick, R.B., Mc Allester, D. and Ramanan, D. (2010) Object Detection with Discriminatively Trained Part-Based Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **32**, 1627-1645. https://doi.org/10.1109/TPAMI.2009.167

[4] Sermanet, P., Kavukcuoglu, K., Chintala, S. and Lecun, Y. (2013) Pedestrian Detection with Unsupervised Multi-Stage Feature Learning. 2013 *IEEE Conference on Computer Vision and Pattern Recognition*, Portland, 23-28 June 2013, 3626-3633. https://doi.org/10.1109/CVPR.2013.465

[5] Ren, S., He, K., Girshick, R. and Sun, J. (2015) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 91-99.

[6] Zhang, L., Lin, L., Liang, X., et al. (2016) Is Faster R-CNN Doing Well for Pedestrian Detection? *European Conference on Computer Vision*, 443-457.

[7] Ye, G.L., Sun, S.Y., Gao, K.J., *et al*. (2017) Nighttime Pedestrian Detection Based on Faster Region Convolution Neural Network. *Laser* & *Optoelectronics Progress*, **54**, 117-123.

[8] Tian, Q., Yuan, T.Y., Yang, D. and Wei, Y. (2018) A Pedestrian Detection Method Based on Dark Channel Defogging and Deep Learning. *Laser* & *Optoelectronics Progress*, **55**, 111007.

[9] Jobson, D.J., Rahman, Z. and Woodell, G.A. (1997) Properties and Performance of a Center/Surround Retinex. *IEEE Transactions on Image Processing a Publication of the IEEE Signal Processing Society*, **6**, 451-462. https://doi.org/10.1109/83.557356

[10] Jobson, D.J., Rahman, Z.U. and Woodell, G.A. (2002) A Multiscale Retinex for Bridging the Gap between Color Images and the Human Observation of Scenes. *IEEE Trans Image Process*, **6**, 965-976. https://doi.org/10.1109/83.597272

[11] Bodla, N., Singh, B., Chellappa, R., *et al*. (2017) Soft-NMS—Improving Object Detection with One Line of Code. 2017 *IEEE International Conference on Computer Vision* (*ICCV*), Venice, 22-29 October 2017, 5562-5570. https://doi.org/10.1109/ICCV.2017.593