

Automatic Visual Reading of Meters Using Deep Learning

Karlo Koščević

Faculty of Electrical Engineering and Computing
University of Zagreb
Zagreb, Croatia Email: karlo.koscevic@gmail.com

Marko Subašić

Faculty of Electrical Engineering and Computing
University of Zagreb
Zagreb, Croatia Email: marko.subasic@fer.hr

Abstract—In this paper, we present a novel approach to the problem of reading residential meters using deep learning algorithms. As a starting point we use Faster R-CNN method and, to acquire more precise readings, we modify its functionality. As there were no databases for this kind of task, one had to be collected and properly annotated. This paper also provides a brief introduction to methods for image augmentation and a technique to augment annotated image dataset. For each part of the presented method as well as the whole method as one unit experiments were conducted to show the overall successfulness.

I. INTRODUCTION

Electricity, gas and water meters are common devices in all residential areas. Every day, many experts inspect these devices and many owners have a need to track their consumption either for personal statistics or because of their obligations to service providers. There are already some remote reading solutions for the problem of residential meter reading, but they are often very expensive and require some additional hardware on top of the meter itself or a new meter. Therefore, our goal was to develop a solution that is easy to use, accessible to everyone and inexpensive. Smartphones, deep neural networks and new frameworks that can connect these two concepts together are the main motivation for our solution. A piece of software that is accessible to all smartphone users drastically reduces costs of meter reading as it does not require custom hardware and simplifies interactions with systems for gas, water or electricity consumption. Our goal was to develop a deep learning method that would fulfill such requirements. The starting point of our solution is a deep neural network for object detection Faster R-CNN [10]. To achieve better readings we modified that network. As there were no databases for this sort of problem we manually annotated our own database and, additionally, implemented augmentation methods to bulk up our database.

II. RELATED WORK

The technology of telemetry has successfully been applied for the problem of reading of residential meters. Some PLC [1], RF [2], GSM [3], and etc. technologies are still widely used for automatic meter reading.

In [4] a method for residential meter reading that uses low-level image features is proposed. The process of determining meter type is done by searching nearest neighbors in a k-d tree which is made of features extracted from all sample images

of devices and image of interest. To extract features SURF [5] method is used. Once the type of meter is known input image is transformed to match the perspective of the corresponding sample. For each region of interest, digits are extracted using some image processing methods and recognized using the HOG [6] descriptor. For more in-depth insight into this method please refer to [4].

Examples of existing smartphone solutions include [7], [8], [9] and etc.

III. OBJECT DETECTION USING DEEP NETWORKS

Faster R-CNN [10] is an object detection method split up into two parts: Region Proposal Network (RPN) and Fast R-CNN detector [11]. Faster R-CNN is a deep neural network with *attention* [12] mechanism as RPN module tells Fast R-CNN detector where to look for objects [13]. Figure 1 shows Faster R-CNN architecture. To make this paper more self-contained we briefly describe this method in following paragraphs but for a more in-depth description of Faster R-CNN method please take a look at [10]. The goal of this method was to share RPN and Fast R-CNN detector computations and because of that, both networks share a common set of convolutional layers [10].

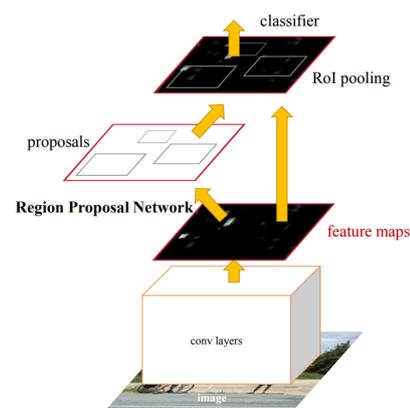


Fig. 1: Faster R-CNN architecture.

Region proposal network (RPN) [10] takes an image as input and outputs objects proposals. Region proposals are generated using the small convolutional neural network [14] on top of the last shared convolutional layer. This network works

in a sliding-window fashion and at the final stage it splits into two sibling fully-connected layers: a box-regression layer and box-classification layer. This small network is illustrated at a single location in figure 2. At each sliding window location, multiple region proposals are predicted. These proposals are parametrized relative to k reference boxes, that are called *anchors*. An anchor is centered at the sliding window and is associated with a scale and aspect ratio. Because of the fixed number of anchors, the input image is preprocessed to some fixed size [10].

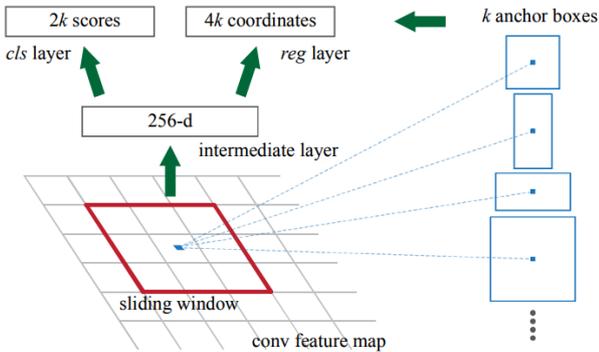


Fig. 2: Region proposal network.

Fast R-CNN detector [11] is also built on top of the last shared convolutional layer. It takes object proposals predicted by RPN and classifies each proposal in one of N classes and refines localization. This small network, for each proposal, computes a fix-sized feature vector and after few fully-connected layers, it splits into two sibling layers same as RPN. Layer named RoI pooling layer [15] computes this fix-sized vectors. RoI pooling divides the region of interest into a $H \times W$ grid of sub-windows and performs max-pooling in each sub-window to get the corresponding output grid cell [11].

RPN and Fast R-CNN detector have similar loss functions. Both are split into regression and classification losses. The only difference is how the classification loss is computed. RPN uses log loss over two classes while Fast R-CNN detector uses log loss over N classes [10].

IV. IMAGE DATABASE

There are two main classes of residential meters: gas and electricity meters. Additionally, as there are subtle differences within printed data in the window of a meter, each of these categories can be visually divided into many subclasses. Figure 3 shows different subclasses of gas meters. All the information that we need to read the residential meter is contained inside of the meter window. Most important are counter and serial number regions. By visual insight into images of various counters, it can be concluded that areas outside of the window do not necessarily have useful information for our task. These regions are usually uniform surfaces that provide much less information about the counter or serial number location. This can be observed in figure 3, where three counters in bottom row all classify as same class and subclass.



Fig. 3: Samples of gas meters

Our database consists of images of mechanical meters of two main classes: *gas* and *electricity*. Both are additionally divided into 9 subclasses. There are 312 images of the meters of class *gas* and 463 images of meters of class *electricity*. Images were captured with various types of smartphone cameras. The number of acquired images is not sufficient to train a high-quality object detector and therefore we extend our database using augmentation. It is common for meters for electrical power consumption measuring to have more than one counter. All of the meters of class *electricity* in this database have two counters.

Faster R-CNN is classified as a supervised learning method [16] which means that we have to provide the expected result for each input that we train our network with. Regarding object detection, we have to annotate all objects of interest in the image with the class and bounding box. Each image in this database has manually annotated corners of meter display, corners of the counter(s) and serial number and corners around each digit in those two regions. Figure 4 shows one sample for each of those annotations. Usually, object location is annotated in a form of a rectangular bounding box, but we propose a new approach for describing the object location. In our dataset bounding boxes are not rectangles, but irregular quadrangle. This means that we describe each object location with four points. The difference can be observed in figure 5. Using our approach borders of each object are localized more precisely which completely removes all surrounding areas from further processing and localization is more robust to different positions and orientations of captured objects. It is important to point out that all objects of interest for the problem of residential meter reading are quadrangles and new annotations system is therefore perfectly suitable.

A. Image augmentation

As the acquired dataset is rather small with the respect of the number of images, to enhance generalization, we use image augmentation techniques. There are many open-source libraries for image augmentation, but we have implemented our own augmentation system. We use standard image transformations, i.e. translation, rotation, warping, cropping, exposure and intensity adjustment and blurring. What makes our library different from others is that, as we create new images, we also output locations of all objects of interest corresponding

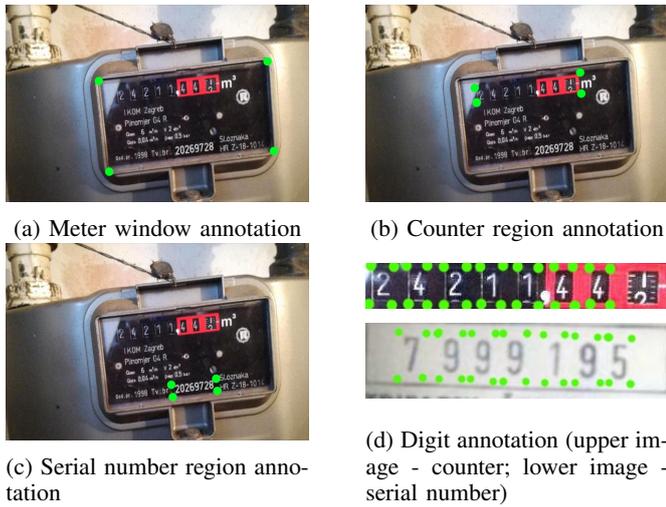


Fig. 4: Annotation samples



Fig. 5: Difference between standard bounding box notation and our four-point notation

to a reference image. Using this approach we do not have to manually annotate each augmented image. Our system is fully parameterized. For each operation range of acceptable values can be specified. We filter all images based on the position of a region of interest in the augmented image, i.e. if the region is out of the image we discard that image. Additionally, we expand the space of possible augmented images by providing the option to specify the probability that each operation will be applied.

To bulk up our database, we split the original data into train and test sets and for each set separately, using reference images, we produce a set of artificial images.

V. FASTER R-CNN MODIFICATION

In this paper, we present a modified version of Faster R-CNN method. In the original implementation (described in section III) object location is annotated using rectangular bounding box where each bounding box is described with x and y coordinates of its upper left corner, width and height. In section IV we have shown that this representation of bounding box may often add surrounding areas which may have a bad impact on further processing to the region of interest and we have proposed a new notation where bounding box is described with x and y coordinates of its four corners. To store this information we use following notation: $(x_1, y_1, x_2, y_2, x_3, y_3, x_4, y_4)$ where pairs of x_i and y_i numbers

represent coordinates of bounding box corners. (x_1, y_1) is the top-left corner and the following points are stored clockwise.

With new bounding box notation, it was necessary to modify Faster R-CNN method.

A. Modification of Fast R-CNN detector

Fast R-CNN detector does fine-tuning of regions proposed by the RPN module. It is aware of all classes of objects and performs multiclass classification. Also, it refines proposed object locations [10].

Fast R-CNN detector output is the output of the Faster R-CNN method and localization output for our implementation must be in form of a four-point bounding box. Regression head outputs $8k$ values, instead of original $4k$ values, where k is the number of classes. Classification head remains the same. New object location notation requires modification of loss function. For each corner we have to calculate the error. Unlike the equation 2 defined in [10] we do not use box's center coordinates and its width and height. Instead, we have to apply the offset calculation that is used for the box's center coordinates for each of our quadrangle's corner coordinates and disregard the offset calculations for the width and height. For the w_a and h_a we simply use the length of the longer horizontal and vertical side of quadrangle. Our loss function does not require bounding box width and height values as we are computing the error for each corner separately.

B. RPN adjustment

We can describe the RPN module as a coarse object detector as it only does binary classification [10]. In our solution, RPN retains that same functionality and uses standard bounding box notation. We did not have to modify this module, but in order to make it work properly with the rest of our method, some modifications were made regarding the RPN inputs and outputs.

During the process of meter reading, object locations are annotated with all four points of a bounding box. These notations must be converted in standard bounding box form, i.e. each bounding box must be annotated with its upper left corner position, width and height. We conduct this conversion according to the equation 1 where x_1, y_1 is top-left corner and following corners are observed in the clockwise manner. A bounding box for the RPN module is the smallest rectangle that can fit the whole region of interest annotated with four points. This conversion is shown in figure 6.

$$\begin{aligned} x^* &= \min(x_1, x_4), & y^* &= \min(y_1, y_2) \\ w^* &= \max(x_2, x_3) - x^*, & h^* &= \max(y_4, y_3) - y^* \end{aligned} \quad (1)$$

As Faster R-CNN loss is calculated with respect to the bounding boxes calculated by RPN module we have to convert these rectangular bounding boxes from standard bounding box notation to our four-point notation. We use the equation 2 to achieve that. This conversion does not affect object localization, but only changes the form that bounding box is written in.



Fig. 6: RPN input transformation.

$$\begin{aligned}
 x'_1 &= x^*, & y'_1 &= y^* \\
 x'_2 &= x^* + w^*, & y'_2 &= y^* \\
 x'_3 &= x^* + w^*, & y'_3 &= y^* + h^* \\
 x'_4 &= x^*, & y'_4 &= y^* + h^*
 \end{aligned} \tag{2}$$

VI. PROPOSED METHOD FOR METER READING

The method that we propose is functionally divided into two major parts, but in order to establish proper communication between them, we created an additional layer, called the bridge. Both major parts perform the same functionality, but differ in types of objects that they are trained to detect. First part is trained to detect counter and serial number regions in the image of a residential meter. Second part of our method performs OCR (Optical Character Reading) of counter and serial number regions.

A. Preprocessing

Both major parts start by preprocessing the input image. The image is normalized by subtracting the predefined mean pixel values and then scaled. We uniformly scale the input image so that shorter side is not longer than s_1 and longer side is not longer than s_2 . Section III provides more detailed description of the importance of scaling of input images.

B. Post-processing

Faster R-CNN method outputs many detections for each input image and not all of them are valid. In order to extract only detections that we are interested in, some post-processing is also done. Each detection is described with classification score and bounding box coordinates. For each class only detections with classification score higher than some threshold are taken into account, i.e. all detections with a score lower than a threshold value are discarded from the next steps. It is common that multiple overlapping detections of the same object occur. This problem is solved using the Non-Maxima Suppression (NMS)[17] method. Since the class is determined based on maximum response NMS does not have effect on detection, but notably reduces the number of region proposals. When post-processing is finished only one and the most accurate detection remains for each class of objects. The proper choice of post-processing thresholds is very important. If the classification score threshold is too high all detections may be discarded and if it is too low, many false positives will remain.

C. Counter and serial number detection

The main part of this module is modified Faster R-CNN network that we proposed which is trained to detect counters and serial number in the image of a residential meter. It takes an image as input and gives one or more detections of objects of a class counter and exactly one detection for objects of a class serial number. Considering the number of detected objects, all images where none or more than three detections occurred can be considered as faulty. This part of the method is in total aware of three object classes. As it is already stated there are counter and serial number class, but additional class named *background* also exist. This class is a catch-all class where all detections that are neither counter or serial number are classified.

D. The bridge

The bridge converts the result of the counter and serial number detection part into a format that is required for digit detection which must be an image containing only a counter or serial number region. To achieve this we crop required regions from the original image of meter.

Our goal is to preserve as much data from the region of interest as possible and at the same time minimize the amount of surrounding areas that may badly affect digit detection. It is important to notice that we represent our region of interest with four points which leads to different shapes of these regions. To achieve our goal we have to transform the region of interest into a rectangle. Planar homography is used to map the image of a meter into en-face view which removes unwanted perspective distortion. Using this approach region of interest can be cropped from original image without any loss of data and, in a perfect situation, without any excess regions. This process is shown in figure 7.



Fig. 7: The bridge functionality.

E. Digit reading

Here the OCR (Optical Character Recognition) of counter and serial number regions is performed. As input, it takes strictly image of one of those two regions. Its main part is, as in section VI-C, modified Faster R-CNN network which is now trained to detect digits. The number of classes for this part is 11 as there are ten possible digits (numbers 0-9) and additional catch-all class *background*. Considering meter and region type, this part can detect five to eight digits. After the detection and standard post-processing are done, digits must be formed into a meaningful number. To achieve this we order them according to the x coordinate of their bounding box, starting from the leftmost one.

The number of executions of this part of our proposed method for residential meter reading is in accordance with the number of counter and serial regions that were previously detected.

VII. RESULTS

As described in section VI our method consists of two main parts. In this section, we present the results achieved for each part as individual component and the results of our method in as one unit. Unlike standard Intersection over Union measure [18] for object localization, we calculate the error for each corner of bounding box separately. The error is written with respect to the average diagonal length of the bounding box of the observed object class. Experiments were performed using Nvidia Titan Xp GPU and Tensorflow framework.

We have first conducted the test for counter and serial number detection part. For this test, we have used a total of 9277 images that were split into the train (7915 images) and test (1362 images) sets. During image preprocessing side length thresholds s_1 and s_2 are set to 600 and 1000. As for the anchors, we use side ratios 1:1, 1:2, 2:1 and areas of 128^2 , 256^2 , 512^2 . We set the classification score threshold to 0.8. Table I displays results for both counter and serial number objects. We have achieved high classification accuracy for both classes. Localization error is greater for serial number objects due to the fact that counters are more salient regions of a residential meter.

We also tested digit reading part. For this experiment, we have used counter and serial number images with a total of 22262 digits. These images are much smaller than those in the previous experiment so the scaling dimensions are also smaller. We set side length thresholds s_1 and s_2 to 100 and 400. Consequently, we use smaller anchors. Side ratios remain the same as, but the areas are now 8^2 , 16^2 , 32^2 . Here we have also used the classification score threshold value of 0.8. Our model detected 95.00% of digits and correctly classified 99.96% of them. We have used the same measure for localization error as in the first experiment. Average localization error was 13.19% of the average diagonal length.

Finally, we have tested our proposed method as one unit. For this experiment, we have used images of residential meters as input for the counter and serial number detection part and as the output of our method one or more counter regions and only one serial number region is expected. In order to classify the reading of counter or serial number as successful, each digit must be properly detected. To conduct this experiment we used 7915 train and 1362 test images. In the test set there were 2009 counter regions and 1362 serial number regions. Our method successfully read 97.01% of counter regions and 63.51% serial number regions. The difference in accuracy is due to the greater localization error of serial numbers. On average, our model failed to detect 2.4 digits per region.

Average execution time we achieved was 0.36 seconds per image. It takes 0.11 seconds to detect counter and serial number regions and 0.08 seconds to read the digits from these detections. It took in average 0.17 seconds to convert

detected counter and serial number regions into valid inputs for OCR part where the majority of time was spent on planar homography.

We show some examples of detections in figures 8 and 9. Points marked with red color are ground-truth locations, while green points are predicted locations. We label ground-truth classes with gt index and predicted classes with p index.

Figure 8 shows some examples of unexpected, but proper detections. There is an image with too low exposure and an image taken from a challenging perspective. These are the best examples of how our method is robust and invariant to many challenging situations. An important feature of our method is that counter and serial number regions are detected regardless of their position and number of digits which vary depending on the meter type.

There are also some cases when our method failed. In figure 9 we can observe an image that was taken from the great distance and very oblique angle. Our method was not able to read the serial number of the captured meter. Additionally, in the same figure, we present some examples where we expect that all digits will be properly detected, but instead our method failed to detect them.

TABLE I: Test results for counter and serial number detection part

RoI	Accuracy (%)	Diagonal length (pixel)	Average error (%)			
			UL	UR	LL	LR
counter	100.00	292.19	0.68	0.42	0.37	0.68
serial No.	99.86	163.33	8.01	7.62	6.71	7.01

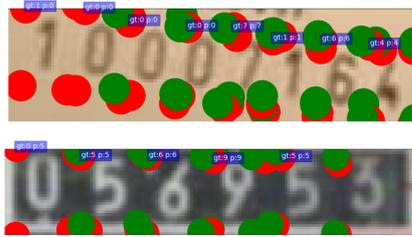


(a) Successful detection of image with low exposure (left) and contrast-enhanced version for visualization purposes(right)



(b) Successful detection of image taken from oblique angle.

Fig. 8: Examples of successful detection.



(a) Unexpected unsuccessful detections.



(b) Image taken from large distance and oblique angle.

Fig. 9: Examples of unsuccessful detection.

VIII. CONCLUSION

In this paper, we show that object detection algorithm Faster R-CNN can be successfully applied to the problem of reading residential meters. The method that we propose has shown to be robust to many different types of residential meters and with great accuracy detects several regions of interest, such as counter and serial number region. Each part of this method can also be used individually for some other tasks. The only thing that is needed is to prepare the annotated dataset for that specific task.

Faster R-CNN modification has proven to be very useful for the problem of reading residential meters. By enabling the network to represent the region of interest with all four corners of bounding box separately, counter and serial number region can be extracted so that surrounding noise is minimized and all relevant data inside the region preserved.

The method that we propose has well-defined input and output parameters. Image of residential meters is expected as the input and numerical values for each region of counter and serial number are expected outputs. Considering that, this method can also be integrated with some existing or new systems for tracking of gas, electricity or water consumption.

ACKNOWLEDGEMENT

Experiments presented in this paper were performed using Nvidia Titan Xp GPU card obtained through Nvidia GPU Grant program. The authors would like to thank Dr. Ross Girshick for permission to use figures 1 and 2 published in [10].

REFERENCES

- [1] P. Oksa, M. Soini, L. Sydanheimo, M. Kivikoski Considerations of using power line communication in the AMR system *Power Line Communications and Its Applications, 2006 IEEE International Symposium on.*, pp. 20 8-211, 2006.
- [2] C. Brasek Urban utilities warm up to the idea of wireless automatic meter reading *Computing & Control Engineering Journal*, vol. 15, no. 6, pp. 10-14, 2004.
- [3] A. Suneel Telemetry system for utility metering application through GSM network *Publications of problems and application in engineering research*, vol. 4, no. 6, pp. 1137-1149, 2017.
- [4] K. Koščević, M. Subašić. Automated computer vision-based reading of residential meters. CCVW 2016 Proceedings of the Croatian Computer Vision, Workshop, pp. 30-35
- [5] H. Bay, T. Tuytelaars, i L. Van Gool. Surf: Speeded-up robust features. pp. 404-417, 2006.
- [6] N. Dalal i B. Triggs. Histograms of oriented gradients for human detection. *Proc. IEEE Conf. Comput. Vis. Pattern Recognition*, 2005.
- [7] SPMR SPMR <https://www.smartphonemeterreading.com/>, Accessed: 09.09.2018.
- [8] Anyline Use The Best OCR Meter Reading Solution To Your Advantage <https://anyline.com/products/ocr-meter-reading>, Accessed: 09.09.2018.
- [9] H. Bouckley Now you can read your electricity meter using your mobile phone <http://home.bt.com/tech-gadgets/tech-news/now-you-can-read-your-electricity-meter-using-your-phone-11364013434362>, Accessed: 09.09.2018.
- [10] S. Ren, K. He, R. Girshick, i J. Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 39, no. 6, pp. 1137-1149, 2017.
- [11] R. Girshick. Fast R-CNN. *Proceedings of the IEEE international conference on computer vision* pp. 1440-1448, 2015.
- [12] J. Chorowski, D. Bahdanau, D. Serdyuk nad K. Cho, i Y. Bengio. Attention-based models for speech recognition. *Neural Information Processing Systems (NIPS)*, 2015.
- [13] P. Yang, Y. Zhong, i Y. Yuan. Fast R-CNN with region proposal refinement.
- [14] A. Karpathy. Stanford university cs231n: Convolutional neural networks for visual recognition. <http://cs231n.github.io/convolutional-networks/>, 23.05.2017.
- [15] B. Xue. Object detection review. <https://shadowthink.com/blog/research/2017/05/07/review-object-detection>, 07.05.2017.
- [16] S. B. Kotsiantis. Supervised machine learning: A review of classification techniques. *Proceedings of the 2007 Conference on Emerging Artificial Intelligence Applications in Computer Engineering: Real World AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies*, pp. 3-24, 2007.
- [17] R. Rothe, M. Guillaumin, i L. Van Gool. Non-maximum suppression for object detection by passing messages between windows. *U Computer Vision – ACCV 2014*, pp. 290-306, 2015.
- [18] A. Rosebrock. Intersection over union (IoU) for object detection. <https://www.pyimagesearch.com/2016/11/07/intersection-over-union-iou-for-object-detection/>, 07.11.2016.
- [19] R. Girshick, J. Donahue, T. Darrell, i J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition* pp. 580-587, 2014.
- [20] D. Rumelhart, G. Hinton, i R. Williams. Learning representations by back-propagating errors. *Nature*, pp. 533-536, 1986.
- [21] K. Simonyan i A. Zisserman. Very deep convolutional networks for large-scale image recognition. *International Conference on Learning Representations (ICLR)*,
- [22] J. R. Uijlings, K. E. Van De Sande, T. Gevers, i A. W. Smulders. Selective search for object recognition. 104(2).