Lights and pitfalls of convolutional neural networks for diatom identification

Anibal Pedraza, Gloria Bueno, Oscar Déniz, Jesus Ruiz-Santaquiteria, Carlos Sanchez, et al.

Anibal Pedraza, Gloria Bueno, Oscar Déniz, Jesus Ruiz-Santaquiteria, Carlos Sanchez, Saul Blanco, Maria Borrego-Ramos, Adriana Olenici, Gabriel Cristobal, "Lights and pitfalls of convolutional neural networks for diatom identification," Proc. SPIE 10679, Optics, Photonics, and Digital Technologies for Imaging Applications V, 106790G (24 May 2018); doi: 10.1117/12.2309488

Event: SPIE Photonics Europe, 2018, Strasbourg, France

Lights and Pitfalls of Convolutional Neural Networks for Diatom Identification

Anibal Pedraza[a], Gloria Bueno[a], Oscar Deniz[a], Jesus Ruiz-Santaquiteria[b], Carlos Sanchez[b], Saul Blanco[c], Maria Borrego-Ramos[c], Adriana Olenici[c], and Gabriel Cristobal[b]

[a]VISILAB-University of Castilla-La Mancha, ETSII, 13005 Ciudad Real, Spain
[b]Institute of Optics, CSIC, Serrano 121, 28006 Madrid, Spain
[c]The Institute of the Environment, University of Leon, E-24071 Leon, Spain

ABSTRACT

Diatom detection has been a challenging task for computer scientist and biologist during past years. In this work, the new state of art techniques based on the deep learning framework have been tested, in order to check whether they are suitable for this purpose. On the one hand, RCNNs (Region based Convolutional Neural Networks), which select candidate regions and applies a convolutional neural network and, on the other hand, YOLO (You Only Look Once), which applies a single neural network over the whole image, have been tested. The first one is able to reach poor results in out experimentation, with an average of 0.68 recall and some tricky aspects, as for example it is needed to apply a bounding box merging algorithm to get stable detections; but the second one gets remarkable results, with an average of 0.84 recall in the evaluation that have been carried out, and less aspects to take into account after the detection has been performed. Future work related to parameter tuning and processing are needed to increase the performance of deep learning in the detection task. However, as for classification it has been probed to provide succesfully performance.

Keywords: Deep learning, CNN, RCNN, YOLO, Diatoms detection, Water quality

1. INTRODUCTION

Diatoms are a kind of microscopic algae that have been extensively used for water quality assessment. For such reason, during recent years the identification of these organisms has been approached with several machine learning techniques. Nowadays, with the irruption of the deep learning paradigm, a new horizon is open to explore the power of this framework and its suitability for this problem. In this work, both detection and classification of 100 taxa have been tackled using this kind of algorithms. This work analyses different deep learning approaches including regional convolutional neural networks (RCNN) and their performance for taxon detection and classification.

One of the main points that need to be addressed with such techniques is the availability of enough data. For this purpose, three expert diatomists have digitized and labelled a significant amount of specimens per taxon. Once this information is processed, an initial dataset is obtained, which is further enhanced through data augmentation.

The pearls or lights that deep learning show in this problem are related to classification, were outstanding results are obtained when the number of available data is enough to take advantage from this technique, as shown in 1. In this work, fine-tuned versions of state-of-art models such as AlexNet have been tested providing outstanding results above 99% accuracy. This can be considered as a remarkable result considering the high amount of species and their variability. On the other hand, as stated in 2, with a reduced number of samples traditional methods are able to obtain better results than with CNN.

However, the use of CNN for detection by means of RCNN has shown some pitfalls. The problems comes in terms of detection, when several aspects such as the proposal of candidate regions has to be considered. These will be discussed in this work, as well as the methods proposed to overcome them.

Corresponding author: E-mail: gloria.bueno@uclm.es; http://visilab.etsii.uclm.es

2. MATERIALS AND METHODS

In order to approach an object detection problem, a dataset is needed with enough samples to train using a technique such as deep learning. For this reason, an extensive process of data collecting, labeling and processing has been performed.

2.1 Data acquisition

The first step is to capture a large number of images with real samples of diatoms as they are observed under the microscope. To achieve this, the collaboration of an expert diatomist is crucial. As a result, Dr S. Blanco and his team were responsible for taking close to 11,000 images from the 10 species that are considered in this work. The specific relation of the number of images per species is stated in table 1.

Table 1. List of diatom taxa and number of samples considered

Species	Number
Achnanthes subhudsonis	984
Eolimna rhombelliptica	1056
Nitzschia capitellata	984
Nitzschia inconspicua	2040
Skeletonema potamos	1240
Eolimna minima	1392
Gomphonema rhombicum	512
Nitzschia frustulum var frustulum	1808
Nitzschia palea var palea	100
Staurosira venter	696

2.2 Data labeling

Using these images, the expert was provided with a labeling tool so that he was capable of manually extract a significant amount of ROIs (Regions of Interest), thus fulfilling the experiment requirements. The labeling information was stored in a suitable format so that it was easier to process it.

2.3 Validation

To perform the experiments, 10-fold cross-validation has been applied, with 9 folds for training and the remaining one in each iteration for testing. Hence, joining the results from test folds, the whole dataset is validated.

In order to measure performance, the F-score (also known as F-measure) has been chosen. This is one of the most commonly used metrics to obtain the performance of a binary decision, such as diatom detection. T. Fawcett performs a comprehensive review of this topic in 3. Here, the most important concepts are summarized.

The first equation states the general expression of the F-score, in terms of True positive (TP), True negative (TN), Type 1 error (False positive, FP) and Type 2 error (False negative, FN), of a given confusion matrix:

$$F_\beta = \frac{(1+\beta^2)\cdot \mathrm{TP}}{(1+\beta^2)\cdot \mathrm{TP} + \beta^2 \cdot \mathrm{FN} + \mathrm{FP}} \tag{1}$$

This expression can be stated in terms of Precision (P) and Recall (R), as shown in the following equations.

$$P = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \tag{2}$$

$$R = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

The F-score can be understood as a weighted measure of false positives and negatives, varying the balance between them with the b value, so that recall is given b times more importance than precision. Some common values to b are 0.5, 1 and 2. For this problem the selected value is 1, so the equation used to calculate the measure in our experiments is:

$$F_1 = 2 \cdot \frac{P \cdot R}{P + R} \tag{4}$$

3. CNN FOR REGION DETECTION

3.1 RCNN

Region-based Convolutional Neural Networks is a technique for deep learning applied to object detection that has been extensively used in recent years, along with some variants that improve the performance of certain details. This deep learning-based object detection approach is based in four fundamental steps:

1. Edge boxes for region proposal
2. Region proposal rejection
3. Classification of regions
4. Box merging

These steps are described as follows.

3.1.1 Edge boxes for region proposal

Using the method described in (4), several candidate regions are obtained. These are proposed according to the concentration of edges over different areas of the input image. Moreover, a score is provided so that the likelihood that a specific region contains an object can be measured, and it can be used as a threshold too.

3.1.2 Region proposal rejection

Since the previous algorithm provides several candidate regions, and most of them are overlapped, it is necessary to process them. For this reason, this step takes as input the raw regions that are highlighted previously and merges them using these criteria and a threshold:

- Union area: given two bounding boxes, the overlap index is calculated using the intersection of both areas and also its union, as follows:

$$\text{Union} = \frac{\text{area}(A \cap B)}{\text{area}(A \cup B)} \tag{5}$$

- Minimum area: criterion is similar to the previous one but taking into account the area of the minimum box, as stated in equation (6).

$$\text{Min} = \frac{\text{area}(A \cap B)}{\min(\text{area}(A), \text{area}(B))} \tag{6}$$

Using these criteria, the obtained overlap intersection metric states how much both boxes overlap, so depending on the threshold it is decided whether the value is enough to merge them. To perform this operation, the minimum area that includes both boxes is calculated.

3.1.3 Classification of regions

Once the region proposals are processed, they are classified using the neural network that RCNN has trained.

3.1.4 Box merging (implementation and problems)

One of the facts that has to be considered when using this kind of region-based detectors is that the overall performance highly relies on how well proposed these regions are. Whether artifacts or inaccurate boxes are provided to the network or, on the other hand, are discarded, may have a significant impact over final results. For this reason, the parameters that rule these candidate box proposals have to be wisely chosen, depending on the intrinsic features of the input images and the kind of objects that are to be detected. By tuning these parameters, every raw box can be obtained or some of them may be thresholded, depending on their distance and similarity to other boxes.

In order to overcome this problem, a custom algorithm was developed that, given a list of candidate bounding boxes, merges them by taking into account the classes they predict and their overlap ratio. The main steps of this algorithm are:

1. Check the overlap of each pair of boxes, using the criteria exposed before about union or minimum overlap
2. When a positive merge is detected a new box is built that includes both of them, and they are appended to the list for the next iteration. If not, they are appended as is (if they were not included before)
3. Repeat the process until the list does not change in two consecutive iterations.

Fig. 1 and Fig. 2 show how this problem is overcome with the algorithm described above.

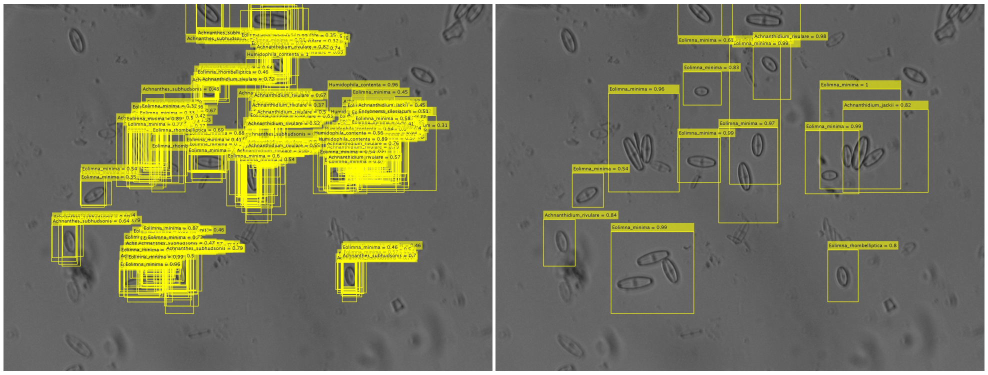

Figure 1. Bounding box merging applying overlap threshold. Left column: a) raw bounding boxes; Right column: b) merged bounding boxes

Moreover, some of the problems that these methods exhibit in our experiments are related to the inability to correctly identify diatoms that are very close to each other, giving a single bounding box and, as a result, failing to quantify the number of diatoms in the slide. This is clearly shown in Fig. 1.

3.2 Object Detection with YOLO

The same problem but with a different approach has been faced using Darknet (see 5). This framework is based on a specific network that follows a different philosophy than traditional R-CNN or even Fast/Faster-RCNN, that increases the running time significantly, which is known as YOLO 6. The traditional approaches are based on applying the model to the input image at multiple locations and scales. However, this recent approach applies a single neural network to the input image, once. This pass is responsible of dividing the image into candidate regions (instead of trusting on additional algorithms that add extra cost). Additionally, as the whole image is fed to the network (instead of several patches), the model has global information about the context of the object, which is more advantageous for accurate decisions.

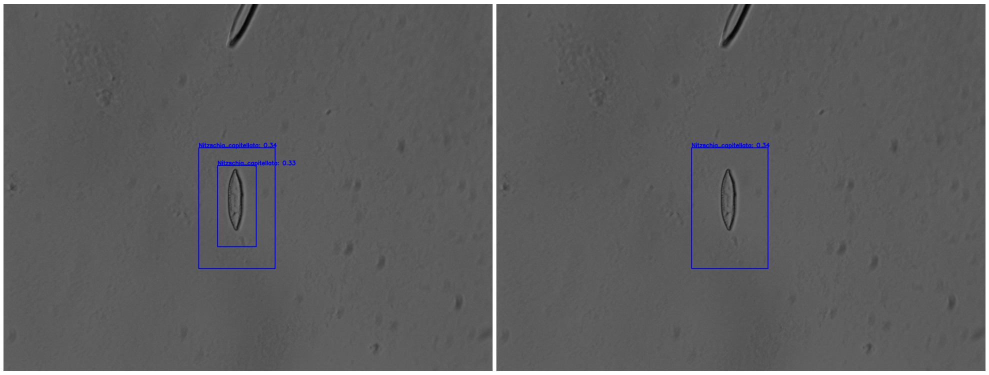
Nitzschia_capitellata: 0.34
Nitzschia_capitellata: 0.33
Nitzschia_capitellata: 0.34

Figure 2. Bounding box merging for the case of multiple detections over the same object. Left column: a) multiple boxes in raw detections; Right column: b) single box after applying the algorithm

In detail, this framework divides the images in a cell matrix, so each cell is responsible for proposing a fixed number of candidate regions. Then, each box is moved according to a predicted offset, so that it fits in size and location to a candidate object. Also, the prediction provides a confidence for both the object class and the bounding box, with the likelihood that they are object. The result is that myriads of boxes are proposed, but most of them have very low confidence, so they can be easily rejected using a threshold.

The way this is performed is shown in Fig. 3. In this figure, it is observed how the image is divided in cells for selecting candidate regions, and how some of them are not taken into account due to a very low confidence.

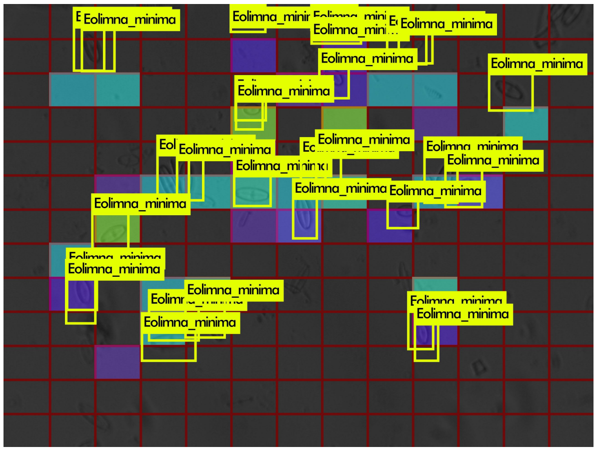

Figure 3. YOLO/Darknet visualization of cells and detections

The steps that have been taken to apply this framework are the following:

1. Process the labeling information to obtain the expected format
2. Modify some network parameters and hyperparameters of the Darknet architecture and the framework to produce the expected number of filters/layers suitable to our problem
3. Perform the training process using the compiled binaries and a pretrained network (based on a more general-purpose problem)
4. Test the model with new images, tuning the detection threshold depending on the response of the model to the different classes.

The same problem about overlapped regions appears with this method, but here it has minor impact. For this reason, the algorithm described before is applied here as well, with results as shown in Fig. 4. In this case, the framework is able to perform better in terms of the ability to split each diatom into a single bounding box, which is critical to perform the biological part of the task, related to water quality assessment.

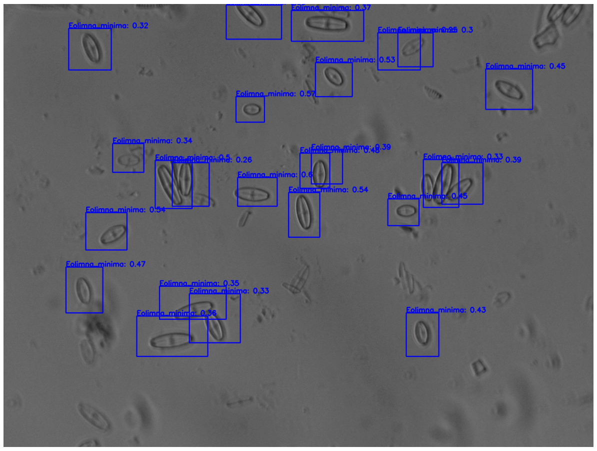

Figure 4. YOLO/Darknet detection

4. RESULTS

In this section, the results obtained with the methods described in previous sections are shown. The performance has been measured from two perspectives:

- Pixelwise: each pixel is considered as a single element in the detection. With this method the detector is checked in terms of its ability to fit the detected boxes to the groundtruth as accurately as possible.
- Objectwise: only single objects as a whole are considered to measure the performance, so in this way it is possible to handle missing objects, as well as false positives in the detection.

The obtained output masks state the performance of the detection in comparison with the groundtruth regarding the following colour code:

- Purple: regions that corresponds to false negatives, that is, regions that are positive in the groundtruth but not detected.
- Yellow: regions that corresponds to false positives, that is, regions that are detected but in fact they are not valid according to the groundtruth,
- Blue: regions that corresponds to true positives, that is, regions that match in both the groundtruth and the detection.

The results are stated in terms of the measures that have been explained before (precision, recall and F-measure) and the approaches based on pixels and objects.

4.1 RCNN

The overall performance of this method is stated in Table 2. To provide a better assessment of the detector performance, Table 3 states the performance for each species. The species with better performance are: *Gomphonema rhombicum*, *Achnanthes subhudsonis* and *Nitzschia palea var palea*. It is observed that the performance is very poor in general in terms of object counting, and more acceptable regarding pixel regions. Fig. 5 shows how the main drawbacks of this technique appear in our experiments, such as missing objects and oversized bounding boxes, which is translated into a seriously increasing number of false positives and false negatives, lowering the performance measures.

Table 2. Overall performance of RCNNs.

Pixelwise Precision	Pixelwise Recall	Pixelwise fMeasure	Objectwise Precision	Objectwise Recall	Objectwise fMeasure
0,29	**0,68**	0,37	0,10	0,13	0,10

Table 3. Performance of RCNNs by species.

	Pixelwise Precision	Pixelwise Recall	Pixelwise fMeasure	Objectwise Precision	Objectwise Recall	Objectwise fMeasure
Achnanthes subhudsonis	0,29	0,78	**0,42**	0,11	0,12	0,11
Eolina minima	0,28	0,75	0,40	0,08	0,09	0,09
Eolimna rhombelliptica	0,22	0,63	0,30	0,13	0,22	0,15
Gomphonema rhombicum	0,42	0,56	**0,45**	0,05	0,06	0,05
Nitzschia capitellata	0,24	0,87	0,35	0,06	0,06	0,06
Nitzschia frustulum var frustulum	0,43	0,39	0,38	0,19	0,12	0,15
Nitzschia inconspicua	0,25	0,70	0,36	0,15	0,21	0,17
Nitzschia palea var palea	0,29	0,75	**0,41**	0,01	0,01	0,01
Skeletonema potamos	0,07	0,45	0,13	0,03	0,07	0,04
Staurosira venter	0,20	0,83	0,30	0,11	0,28	0,16

4.2 YOLO

The overall performance of this method is stated in Table 4. Table 5 states the performance for each species. The species with better performance are: *Gomphonema rhombicum*, *Nitzschia frustulum var frustulum* and *Nitzschia inconspicu*a. With this method, most of the problems from the RCNN approach are solved. For example, most of the diatoms are contained in a single bounding box, and most of them are identified, usually with no missing samples, as shown in Fig. 6.

Table 4. Overall performance of YOLO.

Pixelwise Precision	Pixelwise Recall	Pixelwise fMeasure	Objectwise Precision	Objectwise Recall	Objectwise fMeasure
0,75	**0,84**	0,78	0,74	0,74	0,72

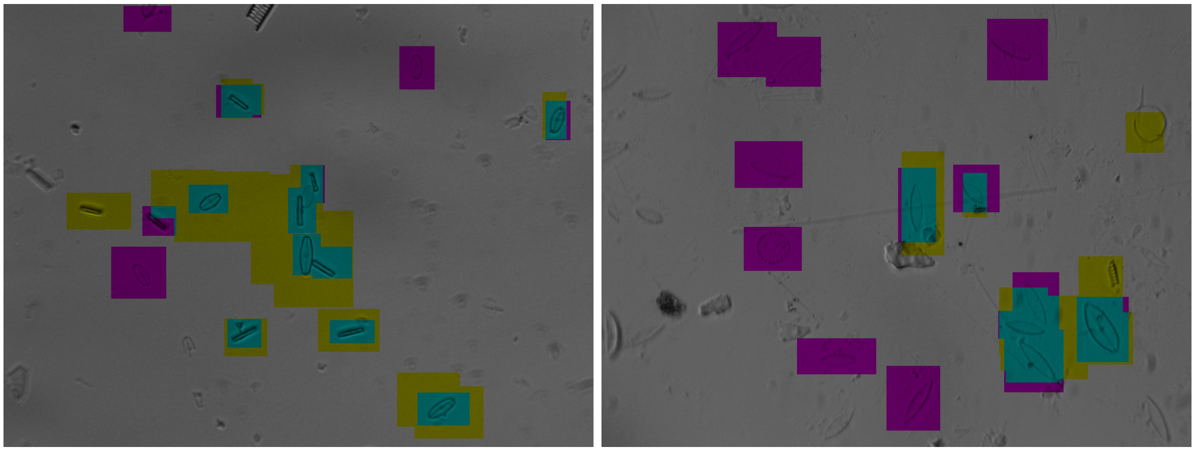

Figure 5. RCNN evaluation example. Left column: a) with false positive regions; Right column: b) with false negative regions

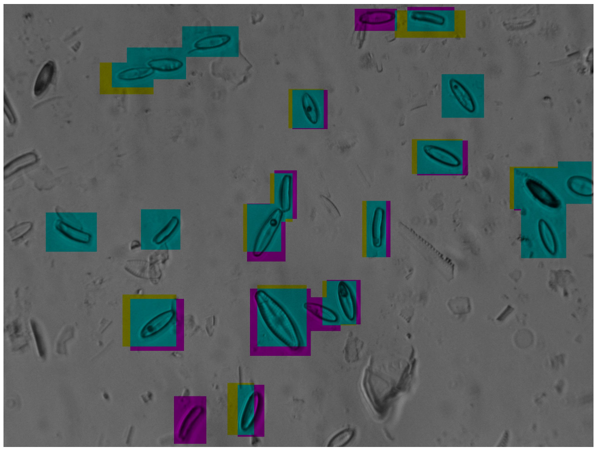

Figure 6. YOLO evaluation example

5. CONCLUSION

In this work, two state of art methods for object detection with deep learning have been compared. The first one, RCNN, is shown in our experimentation to be difficult to be used in terms of parameters and post-processing, since it has been shown that it is important to directly manage the candidate region proposal part and box merging in order to get acceptable results.

On the other hand, YOLO, which is a new approach, where the image is fed to the deep convolutional neural network only once, is shown to be more accurate, with a top performance of 0.84 in F-measure which is close to the standards in object detection for problems of this difficulty.

The evaluation of deep learning applied to this problem can be extended with more experimentation in terms of architectures, paradigms and pre/post processing of datasets. For example, a proposal of future work could be the application of better techniques of box post-processing, with algorithms more complex than overlap thresholding.

The experimental part of this in our work has shown to be useful for validating the suitability of deep learning approaches in such a difficult problem.

Table 5. Performance of YOLO species.

	Pixelwise Precision	Pixelwise Recall	Pixelwise fMeasure	Objectwise Precision	Objectwise Recall	Objectwise fMeasure
Achnanthes subhudsonis	0,84	0,73	0,77	0,85	0,66	0,74
Eolina minima	0,64	0,90	0,74	0,61	0,71	0,65
Eolimna rhombelliptica	0,72	0,87	0,78	0,69	0,81	0,73
Gomphonema rhombicum	0,86	0,84	**0,84**	0,77	0,70	0,72
Nitzschia capitellata	0,66	0,95	0,75	0,72	0,66	0,69
Nitzschia frustulum var frustulum	0,80	0,85	**0,82**	0,79	0,78	0,78
Nitzschia inconspicua	0,79	0,85	**0,81**	0,74	0,81	0,77
Nitzschia palea var palea	0,73	0,85	0,78	0,63	0,79	0,69
Skeletonema potamos	0,50	0,76	0,59	0,45	0,66	0,53
Staurosira venter	0,73	0,86	0,76	0,79	0,82	0,79

ACKNOWLEDGMENTS

The authors acknowledge financial support of the Spanish Government under the Aqualitas-retos project (Ref. CTM2014-51907-C2-2-R-MINECO). Project website: http://aqualitas-retos.es/en/.

REFERENCES

[1] Pedraza, A., Bueno, G., Deniz, O., Cristóbal, G., Blanco, S., and Borrego-Ramos, M., "Automated diatom classification (part b): a deep learning approach," *Applied Sciences* **7**(5), 460 (2017).

[2] Bueno, G., Deniz, O., Pedraza, A., Ruiz-Santaquiteria, J., Salido, J., Cristóbal, G., Borrego-Ramos, M., and Blanco, S., "Automated diatom classification (part a): handcrafted feature approaches," *Applied Sciences* **7**(8), 753 (2017).

[3] Fawcett, T., "An introduction to roc analysis," *Pattern Recognition Letters* **27**(8), 861–874 (2006).

[4] Zitnick, C. L. and Dollár, P., "Edge boxes: Locating object proposals from edges," *European Conference on Computer Vision* , 391–405 (2014).

[5] Redmon, J., "Darknet: Open source neural networks in c." http://pjreddie.com/darknet/ (2013–2016). (Accessed: 12 March 2018).

[6] Redmon, J. and Farhadi, A., "YOLO9000: better, faster, stronger," *CoRR* **abs/1612.08242** (2016).