HPNet: Detecting human parts in the wild

HPNet: Detectando partes humanas na natureza

DOI:10.34117/bjdv7n2-221

Recebimento dos originais: 10/01/2021
Aceitação para publicação: 10/02/2021

Claudemir Casa
Universidade Federal do Paraná - UFPR, Curitiba, Paraná
E-mail: claudemir.casa@ufpr.br

Jhonatan Souza
Universidade Federal do Paraná - UFPR, Curitiba, Paraná
E-mail: email@ufpr.br

Tiago Mota de Oliveira
Universidade Federal do Paraná - UFPR, Curitiba, Paraná
E-mail: tiufsc@gmail.com

ABSTRACT
The objective of this work is to present a model of neural network for the detection and segmentation of human's body parts (HPNet). We offer a multiplatform real-time solution that can be run on ordinary computers, as well as on mobile devices and embedded systems. Our proposal is characterized by presenting a compact solution, and by investigating a part of object detection still little explored. One of the striking features presented is the ability to recognize parts of the human body even in uncontrolled environments, due to the use of a random subset of Google's public database (Open Images Dataset) that contains images with objects in the most varied sizes, positions, lighting and occlusion conditions. At first, we offer a solution only for the detection and segmentation of the common parts of the human body, but we intend to expand its capabilities to detect other more specific parts and regions. The main purpose of our model is its use to solve specific problems that require the detection and segmentation of human's body parts, for example, in user authentication.

Keywords: Human part detection, compact model, real time model, neural networks.

RESUMO
O objetivo deste trabalho é apresentar um modelo de rede neural para detecção e segmentação de partes do corpo humano (HPNet). Oferecemos uma solução em tempo real multiplataforma que pode ser executada em computadores comuns, bem como em dispositivos móveis e sistemas embarcados. Nossa proposta se caracteriza por apresentar uma solução compacta, e por investigar uma parte da detecção de objetos ainda pouco explorada. Uma das características marcantes apresentadas é a capacidade de reconhecer partes do corpo humano mesmo em ambientes não controlados, devido ao uso de um subconjunto aleatório do banco de dados público do Google (Open Images Dataset) que contém imagens com objetos dos mais variados tamanhos e posições , condições de iluminação e oclusão. A princípio, oferecemos uma solução apenas para a detecção e

segmentação das partes comuns do corpo humano, mas pretendemos expandir suas capacidades para detectar outras partes e regiões mais específicas. O principal objetivo do nosso modelo é a sua utilização para resolver problemas específicos que requerem a detecção e segmentação de partes do corpo humano, por exemplo, na autenticação do usuário.

Palavras-chave: Detecção de partes humanas, modelo compacto, modelo em tempo real, redes neurais.

1 INTRODUCTION

Neural networks have been constantly cited when looking a solution for specific object detection problems, intelligent interpretation of images, audios and textual content. There is practically an implicit association between the term *"Artificial Intelligence"* and *"Neural Networks"*, this is due to the wide use of this type of approach to solve problems related to AI. One of the most common problems neural networks are used in today is object detection. And generally for this type of problem, a specific architecture called "*Convolutional Neural Networks"* is used [1].

In a quick search for the term "Object Detection" through indexing tools of scientific publications we find many quotes from different approaches related to object detection. It is easy to find citations for object detection as well as R-CNN [2], R- FCNN [3], FAST-RCNN [4], FASTER-RCNN [5], MASK-RCNN [6], MS-CNN [7], YOLO [8], among others. Each method is constituted by its small peculiarities, however, they all use CNNs in their pipeline.

One of the advantages of the great interest in CNN's is that we have available several tools and frameworks that allow us to carry out prototyping and implementation of neural networks. Many of these tools are made available by large institutions like the Tensorflow that was created by Google [9] or the Pytorch developed by Facebook [10]. This gives us a good margin to obtain good results when implementing new architectures. We can also mention those tools in the cloud, which further accelerate the training process for the new architectures.

There are several practical applications using neural net- works to detect individual human parts. Facial recognition that applications can apply for security and authentication [11]– [13]. We can mention more specific applications, such as nose detection and tracking [14], Eye detection to see if the driver is drowsy [15], smile recognition [16] and pose estimation [17]. We can also find methodologies that use CNN's to au- thenticate users in order to provide more secure solutions. In some examples of applications that use this technique to solve authentication problems, we can highlight those that use the detection of parts of the human face based on their attributes [18], [19]. This type of approach uses segmentation of parts of the face to increase accuracy when authenticating.

Based on searches in the literature, we can see that the detection of parts of the human body using methods of object detection is something little explored in research.

Based on the information presented above, we propose a model of neural network for the detection of human's body parts in uncontrolled environments. Our proposal is aimed at delivering a model that can be used in applications that need to detect and segment such parts, for any problem within this scope. The intention is not to offer the most accurate model, but the one that can be used to solve most problems of this scope, as well as the one with the highest response speed and that can be used on most current platforms.

2 RELATED WORKS

In our searches we found some related works that we would like to quote. A. Jalal [17] presents a proposal for estimating pose in physical sports using human body parts. Hai-Wen Chen [20] proposed a method based on the intensity of colors to estimate human activities, walking patterns and facial recognition. Yang [21], [22] presents, respectively, an approach to facial detection based on face parts to increase accuracy, and an approach to estimate human poses based on the use of human body parts and the relationship between them. Another example we can cite is the work of Tian Y. [23] who also uses body parts to increase accuracy in detecting pedestrians.

3 MOTIVATIONS AND LIMITATIONS

Among the reasons for the elaboration of this work, those of greater prominence are related to the optimization and delivery of a specific model for the detection of human's body parts. We can find several examples of different implementations of models and approaches to detect common objects [24] - [27], however, based on our research, none specific like the one presented. We would also like to use a model like this to detect adverse situations in drivers (such as, for example, identifying when the driver is asleep at the wheel or distracted) and to authenticate users using multiple parts of the face simultaneously. We needed a model that was fast and compact, but also capable of adapting to different variations in the environment.

It is important to note that when we refer to optimization, we are talking about making the model fastest and compact as possible, in addition to making it executable on most current platforms. There are many approaches that are really optimized, but due to the complexity of the model's architecture, sometimes they cannot be executed in real time without the aid of specific hardware, sometimes they cannot be executed on more than one platform.

Our model contributes in several aspects, among them: A specific model for detecting human's body parts in uncontrolled environments; A compact model; It can be executed in real time without the aid of graphics hardware; It can be run without changes on desktop computers, embedded devices and browsers. Even with the contributions presented, the model has limitations, one of the most impacting limitations is accuracy. Due to the compact property, the model loses accuracy to gain speed and decrease its physical size, these gains have a direct impact on the loading and model execution.

4 PROPOSED MODEL

We propose a neural network model for the detection of the human body parts in uncontrolled environments. Our model was designed to detect 18 different classes related to humans, and can be executed on any platform (with or without the aid of GPU's) in real time. The model classes are divided as follows:

- People detection (whole body).
- Detection of 4 genders (adults and children).
- Detection of human body parts.

Our proposal is due to the small interest of community to develop a neural network for this specific scope. We see that there is a great interest in the use of models for the object detection and aimed at body parts like pose estimation works [28]. There are some small quotes using simpler methods like *Histograms of Oriented Gradients* [29] is a feature descriptor that uses shape and texture to select the main parts of an image and we have some works that use it to recognize human parts [30]. Another method is Haar Cascades [31] for some face parts, but no complete neural network model for detecting all visible human body parts.

At first, we would like to provide a single stage network. We also focus our efforts on providing an optimized and universal model to run on multiple platforms, as most neural network architectures are designed to run on computers containing a great deal of computational power (which is not the reality for many users).

The model was also configured and trained in order to achieve similar results even for those entries with many variations, for example, coming from a webcam. This was possible due to the large variation in the image dataset and the combination of some network parameters.

4.1 IMPLEMENTATION DETAILS

Our implementation is based primarily on the YOLO [8] object detection model, but with some modifications. The model written by Redmon proved to be a great candidate for generic object detection, it still has a good inference time and good accuracy. However, for detection in uncontrolled environments for complex objects, its original version is not a good option.

First, our proposal offers a compact model, with a lower number of layers in the network and smaller filters, and the most important point is that we use a specific type of layer that is not present in the original YOLO version using a technique called "Assisted Excitation of Activations" [32]. We also do not use pre-trained weights, as there is no model trained with the characteristics of ours, so we trained our model from the beginning.

We use the same tool used to train YOLO, called Darknet. We later used an external tool to convert our model into a universal format called ONNX. For this conversion, we opted to use an external tool, which transforms operators written for Darknet into operators that can be interpreted by ONNX. Finally, our model initially written for Darknet is structured just like in table I, but the final model already converted cannot be presented through a figure or table due to the number of operators. However, it is possible to visualize the ONNX model using external tools, such as Netron.

We decided to convert the model to ONNX because this intermediate representation can be run on any device without changes, while with Darknet this is not possible. In Darknet the code needs to be adapted and recompiled for each device that you want to support. At the time of conversion, we perform some optimizations to eliminate redundant operations and other automatic optimizations.

Table 1: HPNet base architecture

Layer	Filters	Size	Input	Output
0 conv	16	3x3/1	608x608x3	608x608x16
1 max		2x2/2	608x608x16	304x304x16
2 conv	32	3x3/1	304x304x16	304x304x32
3 max		2x2/2	304x304x32	152x152x32
4 conv	64	3x3/1	152x152x32	152x152x64
5 max		2x2/2	152x152x64	76x76x64
6 ae_conv	128	3x3/1	76x76x64	76x76x128
7 max		2x2/2	76x76x128	38x38x128
8 conv	256	3x3/1	38x38x128	38x38x256
9 max		2x2/2	38x38x256	19x19x256
10 conv	512	3x3/1	19x19x256	19x19x512
11 max		2x2/2	19x19x512	19x19x512
12 conv	1024	3x3/1	19x19x512	19x19x1024
13 conv	256	1x1/1	19x19x1024	19x19x256
14 conv	512	3x3/1	19x19x256	19x19x512
15 conv	69	1x1/1	19x19x512	19x19x69
16 yolo				
17 route	13			19x19x256
18 conv	128	1x1/1	19x19x256	19x19x128
19 upsample		2x	19x19x128	38x38x128
20 route	19, 8			38x38x384
21 conv	256	3x3/1	38x38x384	38x38x256
22 conv	69	1x1/1	38x38x256	38x38x69
23 yolo				
24 route	21			38x38x256
25 conv	128	1x1/1	38x38x256	38x38x128
26 upsample		2x	38x38x128	76x76x128
27 route	26, 6			76x76x256
28 ae_conv	128	3x3/1	76x76x256	76x76x128
29 ae_conv	69	1x1/1	76x76x128	76x76x69
30 yolo				

4.2 DATASET

In our model, we choose to use a randomly extract dataset from Google Open Images Dataset [33], [34]. We created a dataset with more than 120 thousand random and Human-Verified images, and manually annotated the images for the classes ”Nose” and ”Eyebrow”. We decided to use this approach because the images available in OID have many variations in lighting, occlusion, size, among others presented in figure 1. These variations are perfect for our proposal, as we present a model that operates in uncontrolled environments. With these configurations we obtain a physical model of 35 Megabytes.

Figure 1: Examples of image variations

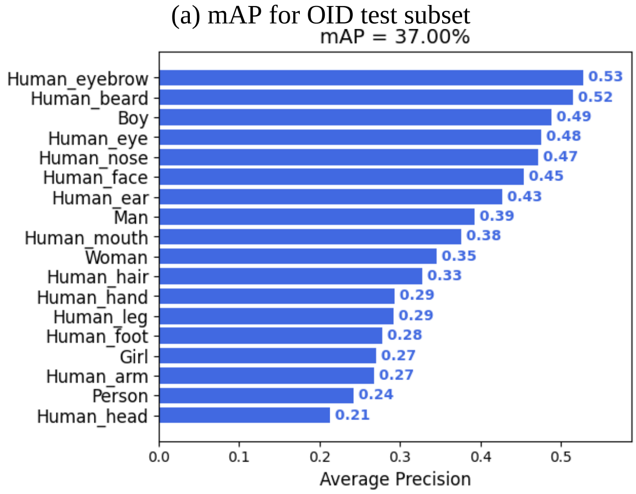
(a) Lighting (b) Oclusion (c) Rotation and Size

ISSN: 2525-8761

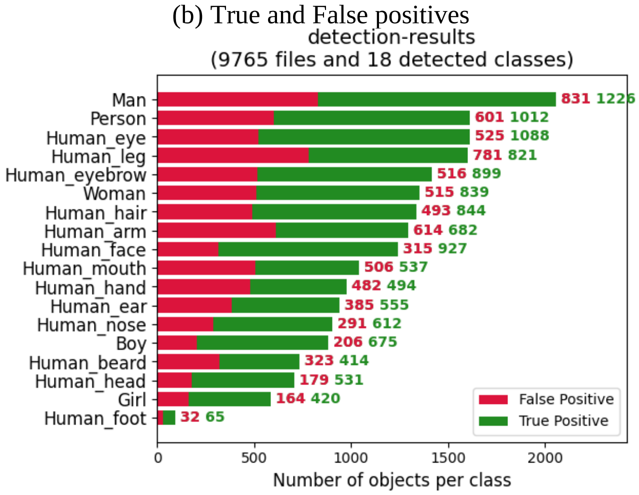
To extract a random subset of images from the OID, we use an external tool called the OIDv4 Toolkit [35], [36]. This tool extracts images from Google's servers according to the necessary parameters, which in our case are the images for all necessary classes and with the variations in the images.

Figure 2: Results obtained for the test dataset

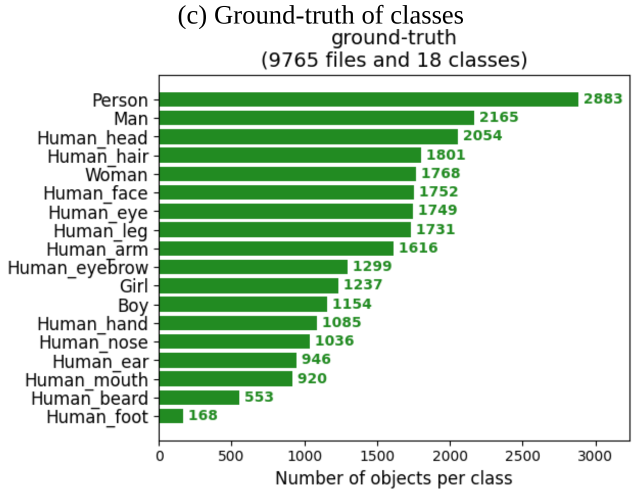
(a) mAP for OID test subset



(b) True and False positives



(c) Ground-truth of classes



4.3 EXPERIMENTS AND RESULTS

With the final model we achieved an average in all classes of 37% accuracy. It is possible to see in Fig. 2 the results obtained separated by class.

In the tests we selected 10% of the images at random, we believe that due to the fact that several images have a high level of variation, they influence more than expected in the results for accuracy. Running the model in real time on various devices visually we obtained great results. It is possible to observe in the figure 3 an example of the results obtained in the execution of the model.

Figure 3: Example of model prediction

(a) Example 1 (b) Example 2 (c) Example 3

We also carry out other experiments for use in real applications. An example was the use of videos of drivers directed to driving tests, in this type of video the variation in lighting and occlusion is much greater, but even so, we get interesting results that can be accessed at https://youtu.be/uyR9GFrIWRY. There are other methods that can be used to detect and extract body parts in addition to CNNs, such as using HOG, HAAR and SVM classifiers [37]–[41]. In some situations, the use of this type of method can have advantages, if well optimized, they can be faster and more accurate, however, when the number of objects is large and there are variations in the environment do not work correctly. Table II presents a simple comparison between our model and other commonly used approaches.

Table 2: Comparisons between approaches in lower computers without dedicated GPUs

Model	Train	Test	mAP	FPS	Classes	Context
Tiny YOLO	OID	test-dev	23.7	~4	80	Generic
YOLOv3-tiny	OID	test-dev	33.1	~4	80	Generic
HPTiny (ours)	OID	test-dev	37.0	~16	18	Specific
SqueezeDet [42]	OID	test-dev	32.4	~9	18	Specific
YOLO nano [43]	OID	test-dev	35.8	~12	18	Specific

5 CONCLUSION

We present in this work a compact model for the detection and segmentation of human body parts in uncontrolled environments. We approach that our neural network is capable of detecting body parts even in uncontrolled environments in real time, as well as being able to be executed in several devices in a universal way. The first results are satisfactory for the scope in which we work, even though they are not so close to the results of the base model. We also demonstrate that our methodology can be used in real-world applications.

In general, our model presented good results. It does not have high precision, like the ones used for detecting common objects, but we approach that this is related to its compact property. Our proposal offers a universal model for the detection of human body parts, which can be used as a basis for future improvements or comparisons in the work of other researchers. In the future, we plan to expand the model to support not only body parts, but regions as well. I would also like to use new approaches and variations in the layers of the network to provide a more accurate and efficient model. We plan to implement it for extracting masks instead of rectangles at the time of predictions. The model can be accessed through the repository https://claudemircasa.github.io/hptiny/.

REFERENCES

[1] W. Zhiqiang and L. Jun, "A review of object detection based on convolutional neural network," in 2017 36th Chinese Control Conference (CCC). IEEE, 2017, pp. 11 104–11 109.

[2] R. B. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," CoRR, vol. abs/1311.2524, 2013. [Online]. Available: http://arxiv.org/abs/1311.2524

[3] J. Dai, Y. Li, K. He, and J. Sun, "R-fcn: Object detection via region-based fully convolutional networks," in Proceedings of the 30th International Conference on Neural Information Processing Systems, ser. NIPS'16. USA: Curran Associates Inc., 2016, pp. 379–387. [Online]. Available: http://dl.acm.org/citation.cfm?id=3157096.3157139

[4] R. Girshick, "Fast r-cnn," in The IEEE International Conference on Computer Vision (ICCV), December 2015.

[5] S. Ren, K. He, R. B. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," CoRR, vol. abs/1506.01497, 2015. [Online]. Available: http://arxiv.org/abs/1506. 01497

[6] K. He, G. Gkioxari, P. Dolla´r, and R. B. Girshick, "Mask R-CNN," CoRR, vol. abs/1703.06870, 2017. [Online]. Available: http://arxiv.org/abs/1703.06870

[7] Z. Cai, Q. Fan, R. S. Feris, and N. Vasconcelos, "A unified multi-scale deep convolutional neural network for fast object detection," CoRR, vol. abs/1607.07155, 2016. [Online]. Available: http://arxiv.org/abs/1607.07155

[8] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," arXiv, 2018.

[9] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin et al., "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," arXiv preprint arXiv:1603.04467, 2016.

[10] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga et al., "Pytorch: An imperative style, high-performance deep learning library," in Advances in neural information processing systems, 2019, pp. 8026–8037.

[11] M. Coskun, A. Ucar, O¨ . Yildirim, and Y. Demir, "Face recognition based on convolutional neural network," in 2017 International Conference on Modern Electrical and Energy Systems (MEES). IEEE, 2017, pp. 376– 379.

[12] M. Sajjad, M. Nasir, F. U. M. Ullah, K. Muhammad, A. K. Sangaiah, and S. W. Baik, "Raspberry pi assisted facial expression recognition framework for smart security in law-enforcement services," Information Sciences, vol. 479, pp. 416–431, 2019.

[13] M. Coskun, A. Ucar, O. Yildirim, and Y. Demir, "Face recognition based on convolutional neural network," in 2017 International Conference on Modern Electrical and Energy Systems (MEES). IEEE, 2017, pp. 376– 379.

[14] L. P. e Silva, F. H. d. B. Zavan, O. R. Bellon, and L. Silva, "Nose based rigid face tracking," in Iberoamerican Congress on Pattern Recognition. Springer, 2018, pp. 556–563.

[15] V. R. R. Chirra, S. ReddyUyyala, and V. K. K. Kolli, "Deep cnn: A machine learning approach for driver drowsiness detection based on eye state." Revue d'Intelligence Artificielle, vol. 33, no. 6, pp. 461–466, 2019.

[16] ——, "Deep cnn: A machine learning approach for driver drowsiness detection based on eye state." Revue d'Intelligence Artificielle, vol. 33, no. 6, pp. 461–466, 2019.

[17] A. Jalal, A. Nadeem, and S. Bobasu, "Human body parts estimation and detection for physical sports movements," in 2019 2nd International Conference on Communication, Computing and Digital systems (C- CODE), 2019, pp. 104–109.

[18] P. Samangouei, V. M. Patel, and R. Chellappa, "Facial attributes for active authentication on mobile devices," Image and Vision Computing, vol. 58, pp. 181 – 192, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0262885616300828

[19] S. Yang, P. Luo, C.-C. Loy, and X. Tang, "From facial parts responses to face detection: A deep learning approach," in The IEEE International Conference on Computer Vision (ICCV), December 2015.

[20] H.-W. Chen and M. McGurr, "Moving human full body and body parts detection, tracking, and applications on human activity estimation, walking pattern and face recognition," in Automatic Target Recognition XXVI, F. A. Sadjadi and A. Mahalanobis, Eds., vol. 9844, International Society for Optics and Photonics. SPIE, 2016, pp. 213 – 246. [Online]. Available: https://doi.org/10.1117/12.2224319

[21] S. Yang, P. Luo, C.-C. Loy, and X. Tang, "From facial parts responses to face detection: A deep learning approach," in The IEEE International Conference on Computer Vision (ICCV), December 2015.

[22] W. Yang, W. Ouyang, H. Li, and X. Wang, "End-to-end learning of deformable mixture of parts and deep convolutional neural networks for human pose estimation," in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016.

[23] Y. Tian, P. Luo, X. Wang, and X. Tang, "Deep learning strong parts for pedestrian detection," in The IEEE International Conference on Computer Vision (ICCV), December 2015.

[24] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in Advances in Neural Information Processing Systems 28, C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, Eds. Curran Associates, Inc., 2015, pp. 91–99.

[25] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016.

[26] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017.

[27] G. A. Lima, D. T. Bravo and S. A. de Araújo, "Utilização de redes neurais convolucionais para a detecção de objetos em imagens aéreas adquiridas por drones," Brazilian Journal of Development, vol. 6, no. 7, pp. 50 702–50 713, 2020.

[28] Z. Liu, J. Zhu, J. Bu, and C. Chen, "A survey of human pose estimation: the body parts parsing based methods," Journal of Visual Communication and Image Representation, vol. 32, pp. 10–19, 2015.

[29] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in International Conference on Computer Vision & Pattern Recognition (CVPR '05), C. Schmid, S. Soatto, and C. Tomasi, Eds., vol. 1. San Diego, United States: IEEE Computer Society, Jun. 2005, pp. 886–893. [Online]. Available: https://hal.inria.fr/inria-00548512

[30] S. Bak, E. Corvee, F. Bremond, and M. Thonnat, "Person re- identification using spatial covariance regions of human body parts," in 2010 7th IEEE International Conference on Advanced Video and Signal Based Surveillance. IEEE, 2010, pp. 435–440.

[31] P. I. Wilson and J. Fernandez, "Facial feature detection using haar classifiers," Journal of Computing Sciences in Colleges, vol. 21, no. 4, pp. 127–133, 2006.

[32] M. M. Derakhshani, S. Masoudnia, A. H. Shaker, O. Mersa, M. A. Sadeghi, M. Rastegari, and B. N. Araabi, "Assisted excitation of activations: A learning technique to improve object detectors," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2019.

[33] A. Kuznetsova, H. Rom, N. Alldrin, J. Uijlings, I. Krasin, J. Pont-Tuset, S. Kamali, S. Popov, M. Malloci, T. Duerig, and V. Ferrari, "The open images dataset v4: Unified image classification, object detection, and visual relationship detection at scale," arXiv:1811.00982, 2018.

[34] I. Krasin, T. Duerig, N. Alldrin, V. Ferrari, S. Abu-El-Haija, A. Kuznetsova, H. Rom, J. Uijlings, S. Popov, S. Kamali, M. Malloci, J. Pont-Tuset, A. Veit, S. Belongie, V. Gomes, A. Gupta, C. Sun, G. Chechik, D. Cai, Z. Feng, D. Narayanan, and K. Murphy, "Openimages: A public dataset for large-scale multi-label and multi-class image classification." Dataset available from https://storage.googleapis.com/openimages/web/index.html, 2017.

[35] A. Vittorio, "Toolkit to download and visualize single or multiple classes from the huge open images v4 dataset," https://github.com/EscVM/ OIDv4 ToolKit, 2018.

[36] D. P. Papadopoulos, J. R. R. Uijlings, F. Keller, and V. Ferrari, "We don't need no bounding-boxes: Training object class detectors using only human verification," 2016.

[37] L. Pauly and D. Sankar, "Detection of drowsiness based on hog features and svm classifiers," in 2015 IEEE International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), 2015, pp. 181–186.

[38] Z. Orman, A. Battal, and E. Kemer, "A study on face, eye detection and gaze estimation," IJCSES, vol. 2, no. 3, pp. 29–46, 2011.

[39] B. Rehman, O. W. Hong, and A. T. C. Hong, "Hybrid model with margin-based real-time face detection and tracking," in Multi- disciplinary Trends in Artificial Intelligence, S. Phon-Amnuaisuk, S.-P. Ang, and S.-Y. Lee, Eds. Cham: Springer International Publishing, 2017, pp. 360–369.

[40] A. S. Kaitake and V. M. Suryawanshi, "Yawning detection to prevent road accidents," 2017.

[41] H. Ahamed, I. Alam, and M. M. Islam, "Hog-cnn based real time face recognition," in 2018 International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE), 2018, pp. 1–4.

[42] B. Wu, F. N. Iandola, P. H. Jin, and K. Keutzer, "Squeezedet: Unified, small, low power fully convolutional neural networks for real-time object detection for autonomous driving," CoRR, vol. abs/1612.01051, 2016. [Online]. Available: http://arxiv.org/abs/1612.01051

[43] A. Wong, M. Famouri, M. J. Shafiee, F. Li, B. Chwyl, and J. Chung, "YOLO nano: a highly compact you only look once convolutional neural network for object detection," CoRR, vol. abs/1910.01271, 2019. [Online]. Available: http://arxiv.org/abs/1910.01271