

Edge Computing for Having an Edge on Cancer Treatment: a Mobile App for Breast Image Analysis

Eleftherios Charteros

Department of Informatics

Athens University of Economics and Business

Athens, Greece

x.lefteris@aueb.gr

Iordanis Koutsopoulos

Department of Informatics

Athens University of Economics and Business

Athens, Greece

jordan@aueb.gr

Abstract—Edge computing has seen tremendous advances in recent years. This progress made it possible to develop mobile applications with greater computational needs that will be able to provide useful insights to users concerning both their habits and their health, by performing computations locally on the mobile device and thus keeping all the data and protecting their privacy. The purpose of this work is to provide a proof-of-concept of an image recognition and analysis app for patients that have undergone breast cancer. The patient takes a snapshot of her breast, and the app runs a Convolutional Neural Network (CNN) model and outputs a classification of breast cosmetic status. We show that it is possible to implement computationally heavy machine learning models in edge devices and to provide real-time status monitoring to users through image analysis done on images taken from the camera of their smartphone, without the photo leaving the device. The app module may enhance a larger-scale system that uses patient-sourced image data to test the effects of surgery or radiotherapy treatment on patients.

Index Terms—Edge Computing, Edge Analytics, On-Device Image Recognition

I. INTRODUCTION

Recently, the term edge computing was coined for the architectural paradigm in which computation, storage, and control are placed at the edge of the network, namely at home gateways, micro-servers or small cells but also on wearables, mobile devices and IoT devices such as sensors. Although the network edge is short of computational, bandwidth and storage resources compared to the resourceful cloud, the edge prioritizes agility over resources and is close to where information is generated, and where data analytics results are delivered. Edge computing has the potential to reduce response times, lower bandwidth usage and improve energy efficiency, while at the same time offering a higher Quality of Experience (QoE) to the end-user for the classes of services mentioned above. Chief among applications that will benefit from edge computing is mobile health. Such applications are expected to improve patient experience and treatment through mobile device and IoT-assisted community, clinical data collection,

This work was supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the First Call for H.F.R.I. Research Projects to support Faculty members and Researchers and the procurement of high-cost research equipment grant (Project Number: HFRI-FM17-352). I. Koutsopoulos acknowledges support from a GSRT Research Reinforcement grant and from the AUEB grant Original Scientific Publications 2019.

real-time patient monitoring and direct intervention of health-care practitioners.

In this work, we design and implement a mobile app that performs image analysis on images taken by the patient. Our app is focused on breast image analysis and can become an essential step towards assessing the post-treatment effects of cancer surgery or radiotherapy. In particular, our app analyzes breast images submitted by the patient on the spot and assesses the cosmetic status of the breast.

A. Motivation and Challenges

The development of a mobile app with on-device image processing has many challenges. Finding the breast boundary automatically without annotating the image is the first challenge that needs to be resolved for the application to be able to provide metrics. Therefore, an image analysis model is needed that can successfully detect and annotate the breast as well as the nipples of the patient. Using such a model creates another associated challenge, which is, creating a dataset on which it will be trained to identify cosmetic changes in unknown images. The second main challenge concerns mostly the application part, and it has to do with the deployment of the trained model on Android devices, as well as the inference of the model and fine-tuning of the results, for accurate metrics.

B. Contributions

Breast cancer is the most common type of cancer among women with an estimated incidence of more than 500,000 in 2018 [1]. Despite the large percentage of incidents, there does not exist a method for automatic breast image recognition and analysis. The reason is that, while recognition is a relatively simple task with modern models, when this task concerns not only recognizing small objects but also exact spatial locations and borders between embedded objects, the task becomes difficult. Thus, transformations, as well as pre-processing, are required, in order to have accurate predictions. With our work, we aim to solve the existing challenges of Breast Analysis Tool (BAT) [2] concerning privacy and objectivity and provide patients with a tool that can help them monitor themselves on a daily basis, without the need to regularly visit the hospital or the doctor. More specifically our contributions are as follows:

- We create a dataset with 140 different female breast images annotated with bounding boxes and masks.
- We configure and train an image recognition model for breast recognition that achieves 98% accuracy in detecting breasts and nipples.
- We use the breast image analysis model on edge devices with face-cropping and by performing only on-device computations without the use of the cloud, and therefore we cater for users privacy.
- We design and develop a mobile application that gives the ability to patients to monitor the cosmetic outcome of breast cancer surgery or radiotherapy, by implementing and further fine-tuning the image recognition model results.

Our breast image recognition work is destined to be part of a mobile or web app that would support the decision making of physicians. The app can also provide an explainable assessment of breast status to patients that have undergone surgery or radiotherapy treatment. The paper is organized as follows. In section II we present an overview of the literature. In section III the neural network model architecture, its implementation, and training. In section IV we describe the app and model testing, and in section V we discuss the metrics we used. Finally, section VI concludes the paper.

II. RELATED WORK

A. Edge Computing

Despite its importance, real-time analytics of massive IoT data is in its infancy. Very little emphasis has been placed on data analytics at the edge (see e.g. the surveys on data analytics in [3], [4]). Much focus has been given in the literature toward IoT [5]. A first attempt to prefetch training parameters at the edge and leverage them to facilitate image recognition is proposed in [6]. A distributed computing network architecture for implementing a deep neural network is proposed in [7], whereby the neural network computation task is partitioned among network nodes and the cloud, and the goal is to reduce communication and computation cost. The work [8] aims to tailor deep neural networks for operation onto mobile devices.

Our work complements this literature by performing analytics on the device with the help of neural networks for a specific problem, that of breast image analysis, where on-device analytics is of importance due to privacy concerns.

B. Related Apps

The only software currently available is the Breast Analysis Tool (BAT) [2] which concentrates on providing objective cosmetic metrics after breast surgery or breast radiotherapy. Although it analyzes and generates metrics of symmetry and aesthetics, critical issues arise. BAT is a desktop-based software, which means that the patient needs to regularly visit the doctor in order to analyze the progress of the surgery. Furthermore, for the analysis to take place, an image of the patient's breast is needed. This creates two more issues. First, to use the software, the doctor must annotate the image of the breast manually, and this makes the analysis of the software

subjective to the annotation of the doctor. Therefore the result for the same image may differ based on the annotations of different doctors. Second and most important, the image of the patient's breast is captured with the doctor's camera. This raises privacy concerns, as the patient does not know if the image is deleted or is used further without her consent.

C. Image Analysis

To develop an autonomous software that can detect and analyze breast surgery progress we need not only to detect the location of breasts and possible abnormalities but to also identify the exact spatial locations so that we extract various metrics concerning symmetry and aesthetics as well as to inform the patient if an abnormal result is found. There are multiple models that can perform general-purpose image analysis.

Single Shot Detector (SSD) [9] is an object detection model. It consists of multiple Convolutional Layers, and each of the last layers proposes a bounding box. The final prediction is the union of the bounding boxes produced for each object.

Faster R-CNN [10] is another model used for object detection. It uses a Convolutional Neural Network (CNN) and a Region Proposal Network (RPN) to suggest the location of objects in an image. After that, the suggestions are passed through Fully Connected Layers from which the class and the bounding box of the image are predicted.

Mask R-CNN [11] is a state-of-the-art instance segmentation model developed by the Facebook AI Research Team. It uses the same architecture as Faster R-CNN and extends it, by adding two extra Convolutional Layers that run parallel to the Fully Connected Layers in order to predict the mask. It is a deep neural network that can detect and separate different objects in an image or a video at the pixel level.

III. IMAGE RECOGNITION MODEL ARCHITECTURE AND IMPLEMENTATION

A. Neural Network Model Selection

We experiment with both SSD and Faster R-CNN models to be able to accurately detect breasts and nipples.

With SSD we achieve fast inference in Android due to the single network used without Region Proposals, but with medium accuracy. After obtaining the bounding boxes with the breasts, we use the Canny Edge detector [13] and Sobel filters in order to detect the border of breasts and abnormalities in appearance, but the result is discontinuous or missing borders due to shadow or color which leads us to reject this model.

With Faster R-CNN we obtain improved bounding boxes and, as with SSD, we use edge detection on them. Regardless of the improved localization, the results are inaccurate and the model is rejected as well.

In our work, we use Mask R-CNN to extract from images the exact spatial locations of the breasts and nipples so as to subsequently calculate objective status metrics concerning symmetry and aesthetics for these areas, which we are unable to with the previously mentioned models. Its architecture, as seen in Figure 1, is an extension of Faster R-CNN [12]. Apart

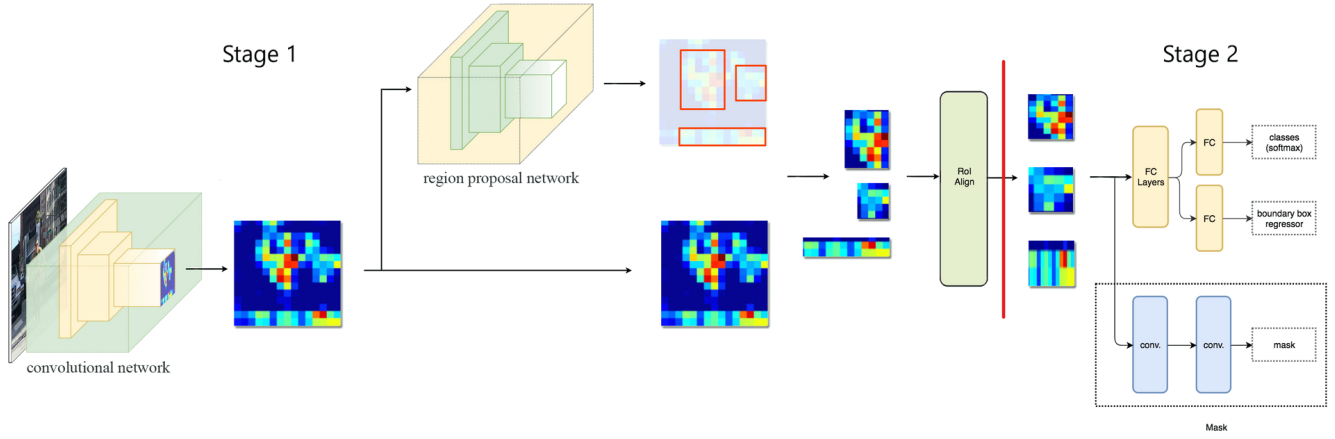


Fig. 1. Mask R-CNN architecture [12]. Stage 1 can be seen to the left of the red line and shows the CNN for feature extraction, the RPN for bounding box suggestion as well as the RoI align for extracting features from the RoIs suggested from the RPN. Stage 2 can be seen to the right of the red line and shows the FC layers for classification and bounding box regression with orange color and the FCN for mask detection with blue color.

from that, they differ in two main aspects that lead us to choose Mask R-CNN over Faster R-CNN.

The first one is that Mask R-CNN replaces RoI pooling, of Faster R-CNN, with RoI align. RoI pooling quantizes the RoI and extracts feature values that are not properly aligned with the input. This does not create large problems for classification and bounding box detection, but it has a large negative effect on mask detection. This problem is solved by RoI align as proposed in [11]. It is shown that with bi-linear interpolation the feature values extracted do not get quantized and they are properly aligned with the input, increasing both bounding box accuracy as well as mask detection. The second difference is that Mask R-CNN extends the second stage of detection by adding an FCN after RoI align for mask detection.

B. Mask R-CNN Architecture

Mask R-CNN's architecture consists of two stages as shown in Figure 1. In the first stage, the image is passed through a Convolutional Neural Network (CNN) where a feature map is extracted. The feature map is passed into a Regional Proposal Network (RPN) which proposes possible Regions of Interest (RoIs) and their bounding box locations in the image. In the second stage, using RoI align as described in [11], Mask R-CNN extracts features from the candidate RoIs. It then passes them through Fully Connected Layers and a Fully Convolutional Network (FCN) simultaneously in order to perform classification and bounding-box regression as well as mask detection. Each pixel of the RoI is mapped with a binary value, 1 if the pixel is part of the object and with 0 otherwise.

During inference, Mask R-CNN performs the same two stages with one difference. In the training phase of the model, both the bounding box detection and the mask detection branch are executed simultaneously. On the contrary, in the inference phase, when an image is fed to the model, it first performs the bounding box detection branch, and after that, the mask branch is applied only to the top- k bounding boxes predicted.

The configurable parameter k tells the model to find masks only for the top- k most confident bounding box predictions.

C. Tensorflow and Dependencies

For our implementation, we rely on Tensorflow Object Detection API [14] from where we use the Mask R-CNN model provided. All scripts essential for training, testing and exporting the trained model as well as for the Android interface for inference are found there. We use a few more tools like Anaconda to create a virtual environment where the dependencies are installed, Nvidia's CUDA to train the model with the GPU for faster training, as well as annotation tools to create ground truth for training. All the steps for the installation can be found in the tutorial provided for the API [15].

D. Dataset Creation

After setting up Tensorflow, in order to use the Mask R-CNN for breast detection and analysis, the model needs to be retrained on breast images so that it successfully recognizes them. We handcraft a synthetic dataset by collecting freely distributed images of female breasts available on the internet. Thus, the software we create stands mostly as a proof of concept that verifies the effectiveness of our tool. The dataset we create consists of 140 images; 120 of these are used for training and 20 for testing the accuracy of our model. To train the model, ground truth labels are needed to learn how to detect the spatial location of the breast in each image. To create ground truth, we use two different pieces of software for annotating the dataset.

E. Labeling Procedure

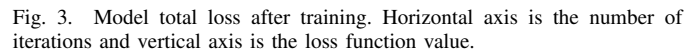
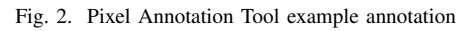
1) *Labellmg*: This tool [16] allows us to draw bounding boxes around the objects we want to detect. We annotate the objects we want to detect and give them specific labels (breast and nipple in our case). After finishing the annotation in each image, the software exports it in a .XML format containing the

2) *Pixel Annotation Tool*: Bounding boxes are not enough for Mask R-CNN to detect the exact spatial locations. Thus we use the Pixel Annotation Tool [17] which allows us to draw the ground truth mask specifying the exact pixels that are part of the breast, as seen in Figure 2. After finishing the annotation in each image, the software exports a .PNG file that contains the mask of each object we annotated.

F. Model Configuration and Training

G. Model Embedding for Android Environment

Mask R-CNN passes the aligned RoI through a Fully Convolutional Network which returns for each RoI a fixed size mask (in our case the mask is $15 \times 15 \times \text{ImageDimension}$ for every possible class as coded in the configuration file of the model). This mask is resized during inference to the size of the image. This task is easy in Python due to the computational



IV. ANDROID APP DESIGN

The developed Android application focuses on making the task of breast status assessment convenient for the patient, while at the same time preserving her privacy. It has only one screen which allows the patient to take a picture of their breast. After that, it gives her the ability to evaluate her breast status without the need for manual annotations or corrections. During the image capturing part, the user needs to capture an image containing both the face and the breasts by holding the mobile device at the height of the neck. A face detection algorithm from Android ML Kit is then used to crop the user's face from the image, and after that use the rest of the image for breast analysis. Capturing the face and removing it from the image is crucial both to ensure that the breasts are captured properly

as well as for privacy since we can ensure that it will never be visible. So even in the case of a data leak, the identity of the user is protected. If the face is detected and cropped the image is displayed and the analysis starts with the user's command. The app passes the image to the trained Mask R-CNN model which analyzes it and returns the predictions with their corresponding masks.

B. Trained Model Inference

The exported .PB graph of the model can be inserted to Android by placing it in the assets folder together with the LABELS.TXT containing the classes we want to predict. To get the inference results from the model, we use the provided Object Detection API java file [14] which we extend to return apart from bounding boxes, the mask for each object predicted. To get the inference results from the model, due to the reshaping of the mask that we described above we rescale the image to a size of 512 x 512 and feed it to the network. We notice, however, that the results are not very good, because the image is stretched at either height or width depending on the capture, resulting in the model being unable to infer properly the distorted image. The solution we resort to is to calculate a scaling factor with which we rescale the largest dimension of the image to 512 pixels. After that, we add to the smallest dimension black pixels to achieve the 512 pixels size, in order to keep the spatial properties of the image after compression. This way, the image keeps the size and shape of the breast in the scaled image, resulting in accurate predictions.

C. Detection Analysis

The model inference results are the spatial locations corresponding to the bounding box of the prediction, as well as the mask of each one of them. After getting the results from the model, the bounding box areas that represent the breasts or the nipples are parsed, and information is stored about five different metrics.

The metrics are: 1) *breast contour*, 2) *breast mass*, 3) *euclidean distance from the centre of the neck between the two nipples*, 4) *horizontal* and 5) *vertical distance between the nipples and the conceivable line from the centre of the neck*, as described in the BAT paper [2].

Then for each metric, the graphical result is presented together with the metric results for the two breasts. It is important to emphasize that the prediction alone is not accurate without fine-tuning the result. Since nipples, as well as most abnormalities, are inside the breast, the model finds it difficult to recognize the exact border between them, leaving the area surrounding them empty, without mask, meaning that we lose accuracy concerning breast metrics. We solve this problem by detecting the pixels that are on the inner side of the breast mask border and are not part of any object detected, and by marking them as being part of the breast mask. After the analysis is complete, the graphical representation of the analysis, as well as the mask of the image, are scaled back to the original size and presented to the user as seen in Figure 4.

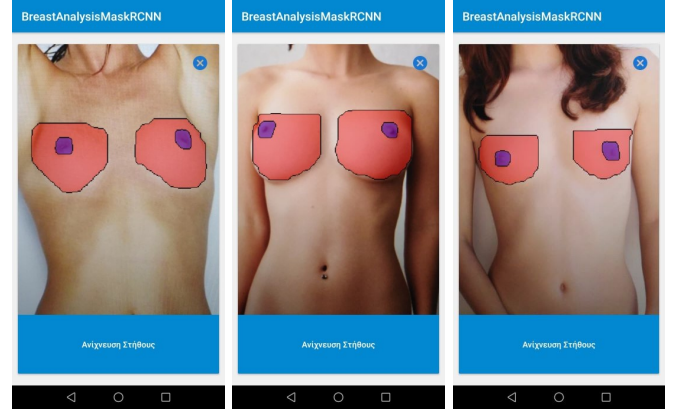


Fig. 4. Breast Detection results. In the left image, the detection result is visible for a woman with small breasts. In the middle image, the detection result is visible for a woman with medium breasts. In the right image, the detection result is visible for a woman with large breasts.

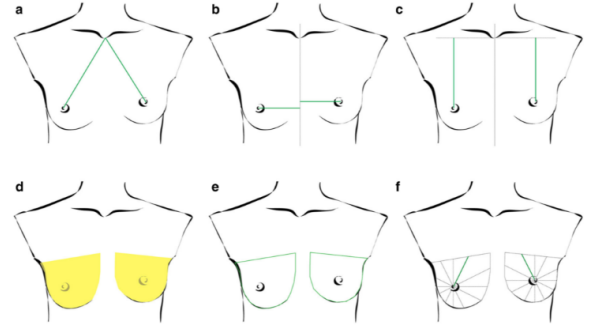


Fig. 5. BAT software metrics

V. PERFORMANCE EVALUATION

A. Prediction Performance and Mask Accuracy

The trained model achieves very good results, detecting a large percentage of the breast and nipples for all breast sizes as seen in Figure 4. The training images are annotated by ourselves and by viewing the annotations provided in the BAT paper, meaning that the shape of prediction (horizontal line at top of the breast mask) is connected to the way we annotated the images. The model is tested through images captured from the screen of another device due to the lack of patients and by female volunteers who are willing to test our app. Identification of breasts and nipples has 98% accuracy with the mask prediction finding correctly about 90-95% of the breast surface in both cases. When images are captured from another devices' screen, the border curves of the mask may not always be smooth due to the visibility of pixels on the device which adds noise to the image and reduces prediction accuracy. Both identification and mask prediction has better results when tested by volunteers.

The analysis metrics are calculated in pixels, and not in cm, on the scale of 512 x 512 pixels. We use the metrics provided in the BAT paper as seen in Figure 5 [2] except the one labeled as *f*, since the points selected from the circumference of the breast

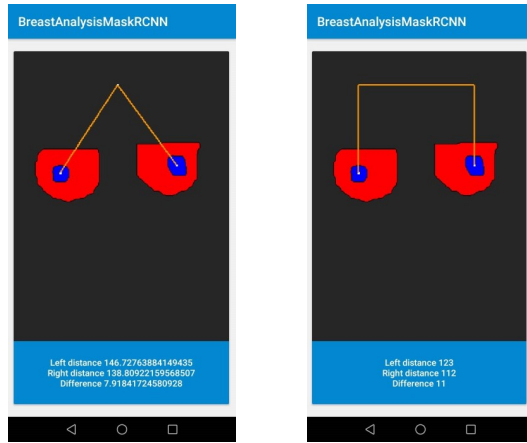


Fig. 6. Analysis metrics based on BAT metrics a and c as seen in Figure 5

are not mentioned. The values calculated are the absolute difference between the pixels of the left and the right breast, as shown in Figure 6. Due to the absence of a regularization metric that would correlate the distance at which the image is taken with the actual size of the breast, the results may vary, based on the distance at which the picture is taken. This means that objects captured within short or long distances from the mobile devices' camera occupy a larger or smaller part of the image respectively. Thus, in our case, the breast would seem larger or smaller than what it is and occupy a varying number of pixels, based on the distance captured, making like this the analysis return unstable results. The final evaluation of the breast is an aggregate of the metrics as described in [2].

B. Inference Time

The inference time on a mobile device is around 20 seconds for a low-end device like Huawei Y6 Prime and around 12-14 seconds on a high-end device like Huawei Mate 20 Pro. The energy consumption of the application based on the metrics provided by Android Studio is medium when using the camera and the face detection algorithm to crop the image, and high when getting the inference results of the model for recognition and analysis.

VI. CONCLUSION

This work is the first step and a proof-of-concept validation that a breast status assessment solution actually works on the device, as it combines the state-of-the-art instance segmentation model Mask R-CNN with the widespread use of mobile devices. We believe that with the appropriate enhancement, our software can be of use both to patients and to doctors. Our application delivers fast results to the patients, it caters for patient privacy and above all, it enables them to monitor the effects of their treatment and the cosmetic changes of it without the need of visiting regularly their doctor. On the other hand, it can turn out to be a powerful tool for doctors to monitor their patients and the time evolution of their treatment

and to find correlations between possible treatments and results based on the analysis results.

The next step is to turn this into a full-fledged architecture, systematically collect and train datasets and perform a more thorough performance assessment. We leave these for future work. Our future improvements concern both the image segmentation model as well as the Android application. More specifically, we aim to experiment with other image detection models to improve both accuracy and speed, and implement our work as part of a mobile or web app system that would help physicians in decision making concerning breast treatment progress after surgery.

REFERENCES

- [1] "Breast cancer facts," 2018. [Online]. Available: <https://europadonna.org/breast-cancer-facts/>
- [2] W. Krois, A. K. Romar, T. Wild, P. Dubsy, R. Exner, P. Panhofer, R. Jakesz, M. Gnant, and F. Fitzal, "Objective breast symmetry analysis with the breast analyzing tool (bat): improved tool for clinical trials," *Breast cancer research and treatment*, vol. 164, no. 2, pp. 421–427, 2017.
- [3] P. Patel, M. I. Ali, and A. Sheth, "On using the intelligent edge for iot analytics," *IEEE Intelligent Systems*, vol. 32, no. 5, pp. 64–69, 2017.
- [4] S. K. Sharma and X. Wang, "Live data analytics with collaborative edge and cloud processing in wireless iot networks," *IEEE Access*, vol. 5, pp. 4621–4635, 2017.
- [5] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of things: A survey on enabling technologies, protocols, and applications," *IEEE communications surveys & tutorials*, vol. 17, no. 4, pp. 2347–2376, 2015.
- [6] U. Drolia, K. Guo, and P. Narasimhan, "Precog: Prefetching for image recognition applications at the edge," in *Proceedings of the Second ACM/IEEE Symposium on Edge Computing*, 2017, pp. 1–13.
- [7] S. Teerapittayanon, B. McDanel, and H.-T. Kung, "Distributed deep neural networks over the cloud, the edge and end devices," in *2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 2017, pp. 328–339.
- [8] X. Ran, H. Chen, Z. Liu, and J. Chen, "Delivering deep learning to mobile devices via offloading," in *Proceedings of the Workshop on Virtual Reality and Augmented Reality Network*, 2017, pp. 42–47.
- [9] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *European conference on computer vision*. Springer, 2016, pp. 21–37.
- [10] 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*, 2015, pp. 91–99.
- [11] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961–2969.
- [12] J. Hui, "Image segmentation with mask r-cnn," Feb 2019. [Online]. Available: https://medium.com/@jonathan_hui/image-segmentation-with-mask-r-cnn-eb6d793272
- [13] J. Canny, "A computational approach to edge detection," *IEEE Transactions on pattern analysis and machine intelligence*, no. 6, pp. 679–698, 1986.
- [14] J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama *et al.*, "Speed/accuracy trade-offs for modern convolutional object detectors," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7310–7311. [Online]. Available: https://github.com/tensorflow/models/tree/master/research/object_detection
- [15] "Tensorflow object detection api tutorial." [Online]. Available: <https://tensorflow-object-detection-api-tutorial.readthedocs.io/en/latest/>
- [16] Tzutalin, "labelimg," 2015. [Online]. Available: <https://github.com/tzutalin/labelImg>
- [17] Abreheret, "Pixelannotationtool," 2017. [Online]. Available: <https://github.com/abreheret/PixelAnnotationTool>
- [18] vijendra1125, "Custom mask-rcnn using tensorflow object detection api," 2014. [Online]. Available: <https://github.com/vijendra1125/Custom-Mask-RCNN-using-Tensorflow-Object-detection-API>