Food Calorie Estimation using Convolutional Neural Network

V Balaji Kasyap
School of Engineering and Technology
CHRIST (Deemed to be University), India
kashyap.vellaluru@gmail.com

N. Jayapandian
School of Engineering and Technology
CHRIST (Deemed to be University), India
jayapandian.n@christuniversity.in

2021 3rd International Conference on Signal Processing and Communication (ICPSC) | 978-1-6654-2864-4/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICSPC51351.2021.9451812

***Abstract*—The modern world healthy body depends on the number of calories consumed, hence monitoring calorie intake is necessary to maintain good health. At the point when your Body Mass Index is somewhere in between from 25 to 29. It implies that you are conveying overabundance weight. Assuming your BMI is more than 30, it implies you have obesity. To get in shape or keep up the solid weight individuals needs to monitor the calorie they take. The existing system calorie estimation is to be happened manually. The proposed model is to provide unique solution for measuring calorie by using deep learning algorithm. The food calorie calculation is very important in medical field. Because this food calorie is provide good health condition. This measurement is taken from food image in different objects that is fruits and vegetables. This measurement is taken with the help of neural network. The tensor flow is one of the best methods to classify the machine learning method. This method is implementing to calculate the food calorie with the help of Convolutional Neural Network. The input of this calculated model is taken an image of food. The food calorie value is calculated the proposed CNN model with the help of food object detection. The primary parameter of the result is taken by volume error estimation and secondary parameter is calorie error estimation. The volume error estimation is gradually reduced by 20%. That indicates the proposed CNN model is providing higher accuracy level compare to existing model.**

Keywords— Convolutional Neural Network; Deep Learning; Food Classification; Food Detection; Pattern Recognition; Tensorflow;

I. INTRODUCTION

Food is the key of human's body. Nowadays more and more people care about the dietary intake since unhealthy diet leads to numerous diseases. A diet plan always needs to take into consideration the total number of calories to be consumed to maintain a fit and healthy life. Weight is an ailment and means you have an excess of muscle to fat ratio. Assuming your BMI is more than 30, it implies you have obesity [1]. Weight can have different reasons. One of these reasons is burning-through a lot calorie. Devouring an excessive number of calories implies that measure of calories that you are taking is greater than measure of calories you consume. The body stores the abundance calorie as muscle to fat ratio [2]. To get in shape or keep up the solid weight individuals needs to monitor the calorie they take. Yet, this interaction can be troublesome and tiring. Since individuals will in general dodge troublesome and tiring things, they regularly don't follow the amount they eat and this may prompt stoutness. Among these examinations, two fundamental variables of the precision change are object location calculation, volume and calorie assessment strategy. For instance, Support Vector Machine (SVM) is utilized for object discovery and characterization [3]. For volume and calorie, the assessment reference point is the thing that has the effect. Utilizing diverse reference points impacts the reasonableness of the application. Reason for this investigation is to make this following simpler. For this, we concoct a Machine Learning Base methodology [4]. With just two pictures (one is from the side and one is from the highest point) of the food and a solitary coin, individuals will actually want to know the calorie of the food that they are eating. In this investigation we discover and characterize the food and make an expectation about the volume of the food. At last, we figure the calorie of the food dependent on the volume that models have anticipated [5]. Nonetheless, we found that assessing the calories straightforwardly was giving us much precise outcomes. But, in most cases, unfortunately people face difficulties in estimating and measuring the amount of food intake due to the mainly lack of nutritional information, which includes manual process of writing down this information, and other reasons. As such, a system to record and measure the number of calories consumed in a meal is of a great help. Hence accurate prediction of food calorie is equally important in such cases. In the last three years, object classification and detection capabilities have dramatically improved due to advances in deep learning and convolutional networks [6]. Harnessing this technology to accurately classify and detect food objects is significantly essential for a healthy and fit life. But to always refer to the nutritional content in each food item is an extremely tedious task [7]. In this project, we use a deep learning-based fruit image recognition algorithm to improve the accuracy of dietary assessment and analyze each of the network architecture.

II. STATE OF ART

The food classification, previous work was concentrated on basic machine learning algorithms like Random Forest and SVM with hand-tailored features. These methods are generally based on relative or spatial relationships of features. But these methods come with computational cost at a large scale. Random forest came up with an accuracy of 62% whereas SVM came up with an accuracy of 67% and our CNN model built with TensorFlow gave an accuracy of 97% [8]. Also, we show here how the model behaves with hyper parameter tuning by changing parameters like the learning rate and number of neurons in each hidden layer which govern how effectively a model behaves. The working model if this tensor flow is divided into different layer. That is sub category of L1 and L2 that will be helpful for optimizing the task. Tensorflow's Object detection API to detect multiple food items in each image and then using mathematical calculations to find the calorie content of food classes present in the image [9]. There are multiple flows of below approach such as Thumb, Simple Food Images, and Size assessment. So we utilize thumb for the volume expectation utilizing clients, the thumb is anything but a proficient way. Since other than the enlisted client whoever utilizes this than the precision will drop. Just Works on Simple Food Images, this methodology doesn't chip away at complex food varieties like soup, sandwich and so forth it just deals with straightforward food sources like apple, banana and so on [10]. I utilized a dataset of inexpensive food pictures dependent on the Pittsburgh Fast-Food Image Dataset. Mathworks image processing toolbox is utilized for separating highlights [11]. Absolute of 11,868 crude highlights separated from RGB portrayal of the picture. Crude highlights decreased utilizing Principal Component Analysis (PCA) and Information Gain (InfoGain) to 23 highlights [12]. Utilizing these highlights and Sequential Minimal Optimization (SMO) they arranged the food. Size expectation is finished by utilizing Random Forest in grams. At long last, the calorie of the food is anticipated utilizing multilayer perceptron. They investigate various sorts of portrayal of a picture, for example, Averaged RGB, Gray Scale, BW 0.7 and BW 0.5 however they get the best outcomes with RGB portrayal [13]. For the size assessment, the ground truth estimation of the food will be taken from its producer. On account of that this technique gives a low precision on the food sources from different producers. White Background, they utilized a white foundation for the food sources in this tasks dataset. Therefore, this venture isn't effective for day by day. A dataset of 2978 pictures of 19 distinctive food were taken. So I utilized Faster RCNN for object discovery [14]. Each jumping box that made by Faster R-CNN is arranged. Utilizing the GrabCut calculation for picture division I separated food into 3 classes' ellipsoid, segment and unpredictable [15]. Utilizing the coin as a kind of perspective point we can ascertain the volume of the food relying upon its shape. Realizing the volume mass can be handily determined utilizing the thickness of the food. With the mass is realized we can figure the calorie of the food.

III. Problem Statement

The approach to approach is from fragments of picture and through that anticipating the calorie of the food in the given picture. In this way, I utilized K-Mean bunching for the picture division and discover the food divides in the picture. For characterization Support Vector Machines (SVM) is utilized. The SVM model was prepared to arrange the food that given. SVM model took Color Features(10 classifications), Size Features(6 classifications), Shape Features(5 classes), Texture Features(5 classes) and group the food. Two pictures were required (one is from the side and other is from the top) which are contains food and the client's thumb for the volume assessment. So accuracy reached to 92.21% of exactness. This article is presenting different method of algorithms. First level random forest method is used and analyzing. The second major concept is support vector machines it is one of the convolutional methods to create different application program interface. We used LabelImage's graphical user interface to manually bound input bounding boxes on training images scraped from Google. The proposed model is train the different data set class, that classes can be category with food objects. This food object is detected with the help of racoon method. The tensor flow is creating own data set that will be training with the help of existing object. Evaluation is done using TensorBoard's. The last phase of the proposed model is calculating the individual food object class based on the input image. GrabCut is a calculation that separates an item to the closer view. In the event that it needs to rearrange, foundation in the picture is erased by GrabCut. Erasing implies in here is changing foundation pixels with dark. Initially, we recognize an item then we bound the article with jumping boxes. Utilizing jumping box, we put the pictures into GrabCut calculation. We can perceive what GrabCut do in our undertaking we utilized GrabCut calculation to check pixel recognized food sources. These pixels are the mathematical type of a picture. At that point we are utilizing them to gauge volume. The training set is inserted and calculated the value based on existing data set with the help of random forest method. It is likewise quite possibly the most utilized calculations, due to its effortlessness and variety (it tends to be utilized for both characterizations). This method is created in the concept of decision tree. In general decision tree is one of the best methods to find the exact value form that list. This decision tree is working in basic machine learning concept. We utilized this model for both assessing the volume and calories. The first approach was assessing the volume and duplicating this assessed volume with food's thickness and kcal (energy per grams). Notwithstanding, this methodology scaled up with volume blunder straightly so we attempted a subsequent methodology. With same arrangement of highlights, we attempted to appraise the complete calories straightforwardly which brought about a vastly improved mean error esteem. We will talk about both of these methodologies in the areas underneath.

IV. Proposed Methodology

Convolutional Neural Network Model Architecture is Classifier Netw. So here we have used ECUSTFD which is a food image dataset which also holds the records of food volumes and masses. It provides labeled images, so we didn't have to handle labeling part. It will give real mass and volume for each food in image. We have used tensorflow's Object detection API to detect food items from image. Also along with that we have also used Random Forest and SVM (Support Vector Machine) with CNN (Convolutional Neural Networks). Keras is also used. ECUSTFD contains 12 varieties of food- apple, banana. Labelled images are used to train object detection models. CNN is built using sequential, convolutional, Pooling, Full Connection. So these are the steps that are followed to complete CNN. CNN Model is giving us 92 percent accuracy. Accuracy has been achieved through several parameters mainly on optimizer, activation function and initializer. Activation function controls the network model so that it can learn more from training dataset. It will give several prediction that model can try. Here we are using RMSprop which is used as an optimizer. Optimizer will make the difference between Algorithm Converging and uploading. Accuracy can vary with respect to learning rate and number of neurons during hyper-parameter tuning. In case we have multiple foods in an image, we will count and will multiple with their calorie amount. Then finally we have to add all the calories to get total calorie. I utilized ECUSTFD (ECUST Food Dataset) in the SeeFood. ECUSTFD is a free open food picture dataset. In ECUSTFD, food sources volume and mass records are given, just as one coin is utilized as the alignment object. ECUSTFD has 19 sorts of food: apple, banana, bread, bun, donut, egg, singed mixture wind, grape, lemon, litchi, mango, moon cake, orange, peach, pear, plum, kiwi, sachima, tomato. There is additionally blend object that is two food varieties in a single picture however we didn't utilize them. There are 2978 pictures in the dataset however we erase blend pictures and we utilized 2870 pictures of 19 unique food varieties. The dataset contains a top view and a side perspective on photographs. ECUSFTD gives other data to each picture as follows such as annotation. In annotation, we didn't manage marking pictures on the grounds that ECUSFTD has the pictures that are named. We utilize marked pictures while preparing object location models. These marked pictures have 2 bouncing box name, one for food and one for coin. In Mass and volume our dataset gives genuine mass and volume to every food in the picture. The pictures are taken various conditions. For instance, some photographs were assumed in a dim position without utilizing a blaze light. Another model is foundation, in most of photographs have utilized 2 distinct plates, there isn't in other. The coin is utilized for the reference point. We utilize this to assess volume. The keras is used to create initial value with different layer. The weight is calculated with neurons and it can be measure the accuracy level. The activation function measurement is also important factor. This layer is activated with different object and sub class.

V. Result and Discussion

The proposed model is to performing hyper parameter tuning on neurons v/s learning rate. This model is to know learning rate affects training process. If learning rate is too low then neural network will not learn that will lead to test error and in case of high learning rate. There might be a random oscillation and divergence in training process. Tensorflow's object detection API trains a model from scratch. Once the model is trained, it will provide Meta and index files which contain trained data of metadata. It will give annotations, flow of data and other related information's. We have used Occam's razor principle. It is generally used to combat over fitting. This States that any simple solutions are having more chances to be correct as compare to complex. Also no of neurons used will not show any affect the performance of model. Properly use hyper parameters and investing lot of resources, while training will give the more accuracy. Table 1 is to be elaborated image count for different fruits. So here the count is indicating number of fruits that is detected through image processing. The table is showing dataset image count.

Table 1: Dataset image count for different fruits

CATEGORY	COUNT
POMEGRANATE	37
BANANA	58
CARROT	57
KIWI	100
GUAVA	79
EGGPLANT	58
CUSTARD APPLE	41
TOMATO	75

Each image was manually labeled for multiple fruits per image. This creates an XML per image that basically has the filename, path and the coordinates of the boxes with their labels so that the model can understand where the object of interest is present. Secondly, each xml is converted into csv for easy tensor vector creation. Tensor Vectors are generated into a TFRecord file format; which is nothing but a simple record-oriented binary format; that the tensorflow application uses for training data. This approach makes it easier to mix and match data for network architectures as all the features are in a field format. This is then used in Tensorflow's object detection API in accordance with SSD_MobileNet_V1_COCO dataset's weights to train a model from scratch. We trained the model for 1200 epochs with the training loss falling to 1 mAP (mean average precision). The trained model generates a set of meta files and index files that contains the metadata of trained data and a serialized MetaGraphDef protocol buffer. This file describes the data flow, annotations and pipeline related information. These files are finally used to generate a frozen inference graph. We used tensorboard for loss metrics and to view other evaluations visually. In case, of multiple items in an image, we simply count the number of objects of each food class in the image and then multiple each with their respective calorie amount to finally summing up the individual counts to get the total calories in the image. Further analysis was conducted on topics like which model results in the best accuracy, how the loss curve looks like, and how the optimizers like RMSprop or Adam optimizes the model. After we got high accuracy on finding coin and food in given images, we saved our object detection model. Then we obtain numerical form of images with using saved model. We added extra knowledge from our data set such as real volume, real density, per calorie in gram and so on. Complete mean volume blunder is 21.06. Our volume blunder result is in Table 2 for each class of food. A large portion of the food varieties' assessed volumes are generally acceptable. Some food is not like kiwi, moon cake and mango. This high blunder is most likely because of our powerlessness to get great information from GrabCut. These numerical form names are in column. We used several images' numerical form for training and 862 images' numerical form for testing in our estimation methods.

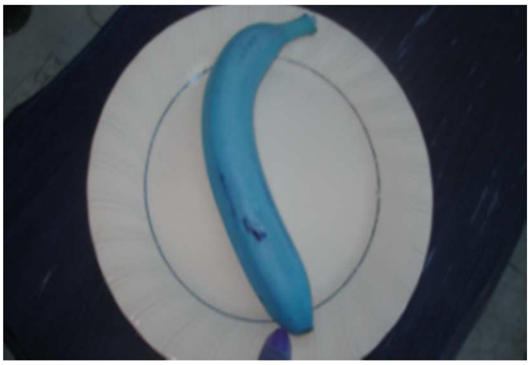

Figure 1: Object Detection through Image Processing

Table 2: Volume Estimation Error Result

Food Type	Estimation Food Count	Mean Volume	Mean Estimation Volume	Volume Error (%)
apple	38	321.32	327.03	1.78
egg	12	51.67	52.26	1.15
lemon	15	96.67	97.62	0.99
orange	28	222.14	226.89	2.14
peach	14	102.14	100.61	-1.5
plum	31	111.29	110.55	-0.66
kiwi	9	160.0	127.94	-20.04
tomato	21	183.81	182.79	-0.56
bread	8	150.0	143.93	-4.05
grape	13	304.62	304.29	-0.11
mooncake	19	52.11	46.99	-9.81
sachima	17	152.35	152.61	0.17
banana	17	174.71	174.03	-0.38
bun	11	236.36	224.55	-5.0
doughnut	26	198.08	197.97	-0.06
Fireddough twist	13	70.0	70.22	0.31
litchi	11	42.73	40.65	-4.86
mango	34	100.88	109.54	8.58
pear	18	242.22	251.9	4.0

We utilized every one of the pictures' handled mathematical structure and afterward when we decide best K. Discovering the estimation of best K isn't simple. We attempt from 1 to 100 for K worth. We ascertain RMSE (Root Mean Square Error) an each estimation of K. At that point we analyzed these RMSEs we determined. We discovered the estimation of K that has least RMSE esteem. The best estimation of K is 7 for K closest neighbor technique. By and large, are two food sources. Least RMSE is the most comparable food. Subsequently when we gauge the volume of food we select closest 7 food varieties that are mathematical the most comparable food varieties. Our assessing volume is the normal estimation of these 7 food varieties genuine volume.

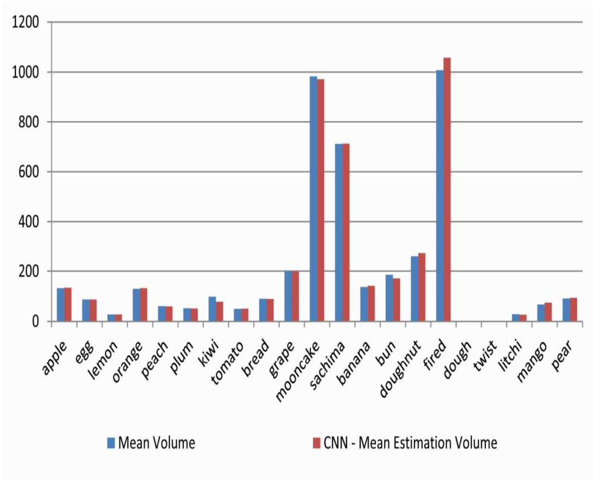

Figure 2: Comparison of Volume Estimation

The figure 2 is to compare the volume estimation in mean value. At the point when we saw these outcomes, it amazed us on the grounds that the outcomes are superior to we anticipate. We get the complete calories utilizing the assessed volume. On the off chance that we duplicate the volume by the food's thickness it will give an assessment about its mass. The proposed model increase this mass with kcals (Joule per grams) to acquire the all-out calories. We can see the exhibition of the model. We saw that a portion of the classes have higher mistake than others like grape and terminated batter wist.

Table 3: Calorie Estimation Error

Food Type	Estimation Food Count	Mean Calorie	Mean Estimation Calorie	Calorie Error (%)
apple	38	131.79	134.24	1.86
egg	12	86.96	87.06	0.12
lemon	15	27.29	27.38	0.32
orange	28	129.12	131.8	2.08
peach	14	60.05	59.4	-1.09
plum	31	51.73	51.38	-0.66
kiwi	9	98.46	78.1	-20.68
tomato	21	49.06	49.85	1.6
bread	8	89.66	88.87	-0.88
grape	13	199.97	199.06	-0.45
mooncake	19	982.53	971.82	-1.09
sachima	17	711.76	712.62	0.12
banana	17	137.18	141.38	3.06
bun	11	185.8	170.52	-8.22
doughnut	26	259.63	272.14	4.82
fireddough twist	13	1007.84	1057.79	4.96
litchi	11	28.19	26.76	-5.09
mango	34	66.86	73.78	10.35
pear	18	90.59	93.89	3.64

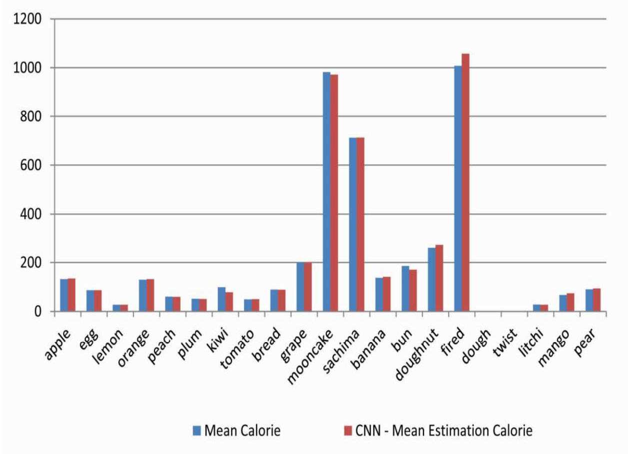

Figure 3: Comparison of Calorie Estimation

This is brought about by the abnormality of our informational collection and the unpredictable design of the food. Our second technique for calorie assessment was straightforwardly assessing the calories with same arrangement of highlights. This methodology decreases the root mean square mistake 32.52 to 30.37.We calculate total calories using estimated volume in mathematical calorie calculation formulas. $C = c \times \rho \times v$. C will estimate calorie where c is calories per gram, ρ is average density of food and v is our estimated volume. Total mean calorie error is 45.79. Our calorie error result is in Table 3 for each different food.

VI. Conclusion

This proposed method is to create food recognition and detection while using several algorithms. Those algorithms are CNN, Random forest, SVM to get better accuracy and we obtained it. We have used a food image dataset which is publicly available. CNN was used for the image recognition. Also we trained the models using information from dataset. Also accuracy has further improved through optimization, hyper parameter tuning. We have written a function which determines calories based on the fruit detected by taking in consideration the average calorie value of that fruit. We will finish up our work contrasting our outcomes and the benchmark work regarding volume assessment since the outcomes for calorie assessments are not partaken in the paper. In the event that we utilize 30% for testing Random Forest model is the most ideal model for our concern with the mean mistake of 13.12 though KNN has a mean blunder of 21.06. We can see the correlation between benchmark results and both of our models. As we can see both of our models beats the benchmark work. Since volumes determined with math equations in our standard work, some natural product that are near amazing shape like lemon has higher precision than our models, anyway this is simply restricted to food sources with ellipsoid shapes. At the point when we look at our volume assessment techniques, we can find that irregular woods model is marginally outflanking the KNN model. The motivation behind why the K Nearest Neighbors technique is however great as the Random Forest strategy seems to be that the informational index and the quantity of highlights that we use is little and in this manner the model doesn't experience the ill effects of revile of dimensionality. Since our informational collection is moderately little we can grow our informational index which will diminish the mistake considerably more as a future work. Additionally eliminating the imbalanced conveyance in our informational index will diminish the blunder too.

References

[1] Winter-Jensen, M., Afzal, S., Jess, T., Nordestgaard, B. G., & Allin, K. H. Body mass index and risk of infections: a Mendelian randomization study of 101,447 individuals. European journal of epidemiology, 35(4), 347-354. (2020)

[2] Bai, C., Liu, T., Xu, J., Ma, X., Huang, L., Liu, S., & Gu, X. Effect of high calorie diet on intestinal flora in LPS-induced pneumonia rats. Scientific reports, 10(1), 1-12. (2020)

[3] Sree, S. R., Vyshnavi, S. B., & Jayapandian, N. Real-world application of machine learning and deep learning. In 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT) (pp. 1069-1073). IEEE. (2019)

[4] Natarajan, J. Cyber Secure Man-in-the-Middle Attack Intrusion Detection Using Machine Learning Algorithms. In AI and Big Data's Potential for Disruptive Innovation (pp. 291-316). IGI Global. (2020)

[5] Morquecho-Campos, P., de Graaf, K., & Boesveldt, S. Smelling our appetite? The influence of food odors on congruent appetite, food preferences and intake. Food Quality and Preference, 85, 103959.(2020)

[6] Pal, S. K., Pramanik, A., Maiti, J., & Mitra, P. (2021). Deep learning in multi-object detection and tracking: state of the art. Applied Intelligence, 1-30. (2021)

[7] Fahira, P. K., Rahmadhani, Z. P., Mursanto, P., Wibisono, A., & Wisesa, H. A. Classical Machine Learning Classification for Javanese Traditional Food Image. In 2020 4th International Conference on Informatics and Computational Sciences (ICICoS) (pp. 1-5). IEEE. (2020)

[8] Grattarola, D., & Alippi, C. Graph Neural Networks in TensorFlow and Keras with Spektral [Application Notes]. IEEE Computational Intelligence Magazine, 16(1), 99-106. (2021)

[9] Talukdar, J., Gupta, S., Rajpura, P. S., & Hegde, R. S. Transfer learning for object detection using state-of-the-art deep neural networks. In 2018 5th International Conference on Signal Processing and Integrated Networks (SPIN) (pp. 78-83). IEEE. (2018)

[10] Bakke, A. J., Carney, E. M., Higgins, M. J., Moding, K., Johnson, S. L., & Hayes, J. E. Blending dark green vegetables with fruits in commercially available infant foods makes them taste like fruit. Appetite, 150, 104652.(2020)

[11] Zheng, L., Lawlor, B., Katko, B. J., McGuire, C., Zanteson, J., & Eliasson, V. Image processing and edge detection techniques to quantify shock wave dynamics experiments. Experimental Techniques, 1-13. (2020)

[12] Asante-Okyere, S., Shen, C., Ziggah, Y. Y., Rulegeya, M. M., & Zhu, X. Principal component analysis (PCA) based hybrid models for the accurate estimation of reservoir water saturation. Computers & Geosciences, 145, 104555. (2020)

[13] Huynh-The, T., Hua, C. H., & Kim, D. S. Encoding pose features to images with data augmentation for 3-D action recognition. IEEE Transactions on Industrial Informatics, 16(5), 3100-3111. (2019)

[14] Vo, H. V., Pérez, P., & Ponce, J. Toward unsupervised, multi-object discovery in large-scale image collections. In European Conference on Computer Vision (pp. 779-795). Springer, Cham. (2019)

[15] Grinvald, M., Furrer, F., Novkovic, T., Chung, J. J., Cadena, C., Siegwart, R., & Nieto, J. Volumetric instance-aware semantic mapping and 3D object discovery. IEEE Robotics and Automation Letters, 4(3), 3037-3044. (2019)