biology and life sciences forum

MDPI

Proceedings

Automated Infield Grapevine Inflorescence Segmentation based on Deep Learning Models

Germano Moreira [1,*], Sandro Augusto Magalhães [2,3], Filipe Neves dos Santos [3] and Mário Cunha [1,3]

1 Faculty of Sciences, University of Porto, Rua do Campo Alegre s/n, 4169-007 Porto, Portugal; mccunha@fc.up.pt
2 Faculty of Engineering, University of Porto, Rua Dr. Roberto Frias s/n, 4200-465 Porto, Portugal
3 INESC TEC-Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência, Campus da FEUP, Rua Dr. Roberto Frias s/n, 4200-465 Porto, Portugal; sandro.a.magalhaes@inesctec.pt (S.A.M); fbsantos@inesctec.pt (F.N.d.S.)
* Correspondence: germano.f.moreira@inesctec.pt

Abstract: Yield forecasting is of immeasurable value in modern viticulture to optimise harvest scheduling and quality management. Traditionally, this task is carried out through manual and destructive sampling of production components and their accurate assessment is expensive, time-consuming, and error-prone, resulting in erroneous projections. The number of inflorescences and flowers per vine is one of the main components and serves as an early predictor. The adoption of new non-invasive technologies can automate this task and drive viticulture yield forecasting to higher levels of accuracy. In this study, different Single Stage Instance Segmentation models from the state-of-the-art You Only Look Once (YOLO) family, such as YOLOv5 and YOLOv8, were benchmarked on a dataset of RGB images for grapevine inflorescence detection and segmentation, with the aim of validating and subsequently implementing the solution for counting the number of inflorescences and flowers. All models obtained promising results, with the YOLOv8s and the YOLOv5s models standing out with an *F1-Score* of 95.1% and 97.7% for the detection and segmentation tasks, respectively. Besides, the low inference times obtained demonstrate the models' ability to be deployed in real-time applications, allowing for non-destructive predictions in uncontrolled environments.

Keywords: Computer Vision; Digital phenotyping; Object segmentation; Precision Viticulture; Yield forecasting

Citation: Moreira, G.; Magalhães, S.A.; dos Santos, F.N.; Cunha, M. Automated Infield Grapevine Inflorescence Segmentation based on Deep Learning Models. *Biol. Life Sci. Forum* **2023**, *1*, 0. https://doi.org/

Published:

1. Introduction

The world wine sector is a multi-billion dollar industry with a wide range of economic activities, representing a vital part of the global economy growth [1]. One crucial aspect of achieving optimal results in viticulture is the yield assessment – the anticipation of the quantity and quality of grapes that a vineyard will produce in a given season. Traditionally, it is carried out by measuring three main yield components, the number of bunches per vine, the number of berries per bunch and the mass of a berry, each one partly responsible for the season-to-season spatial yield variability [2]. One of the earliest assessments can be conducted during spring growth, as the formation of inflorescence primordia (flower buds) determines the potential number of bunches that the vine will produce, while the number of flowers formed on an inflorescence determines the potential number of berries on that bunch [3]. However, as these tasks are carried out manually and assessed by visual inspection, end up becoming expensive, time-consuming and error-prone, as they are repetitive and meticulous, ultimately becoming fatiguing and overly dependent on the operator's training and skills.

The synergy between viticulture and cutting-edge technology has given rise to transformative advancements, leading to more pragmatic and modern approaches, reshaping the sector landscape [4]. The most powerful and widely used technology in this area is computer vision (CV), employed to extract meaningful information of physical objects from images or videos [5]. The first approaches were based on more classic image processing and analysis techniques, focusing on counting the number of flowers per inflorescence [6–11]. It was therefore common to acquire images in controlled environments with artificial backgrounds, where the inflorescences were already detached from the plant. Thus, conventional methods are primarily constrained by the necessity to meticulously choose suitable algorithms for tasks like feature extraction, shape identification, and categorization, and often require a degree of control over the environment [12]. Recently, Deep Learning (DL) models have emerged as potent tools, having a massive impact on the development of CV algorithms, due to their capacity to unravel and deal with complex scenarios [13]. Regarding viticulture, the accessibility and visibility of different yield components are two major challenges that CV-endowed systems face. The rates of occlusion for both inflorescence and bunch exceed 50% by a significant margin [14]. DL models have made it possible to achieve non-destructive predictive models that can be used in uncontrolled environments, not only in terms of detecting and counting flowers per inflorescence, but also inflorescences per vine, since these are more robust, with better responses to occlusion and overlapping problems [15–24].

The agricultural sector inherently complex and unstructured environment poses significant challenges that can hinder the performance of these solutions. While DL models have demonstrated great promise, the existing literature still exhibits notable weaknesses that warrant attention [12], related to poor dataset quality and size and the methodologies and detection frameworks employed may not be optimized for the unique challenges posed by agricultural settings. Therefore, this research aims to analyze the performance of different state-of-the-art YOLO model versions to detect and segment grapevine inflorescences. The implementation of these models can be beneficial, as they can perform feature extraction and object detection in a single step, consuming less time and potentially be used in real-time applications, as well as providing support for future tasks, such as counting flowers per inflorescence. The main contributions of this study are as follows: (i) Acquire and make publicly available datasets of labeled grapevine inflorescences images. (ii) Benchmark the results of DL models for detection and segmentation of inflorescences in different grape varieties and phenological stages.

2. Methods

2.1. Data Acquisition and Processing

A new RGB images dataset of grapevine inflorescences was collected throughout three grapevine phenological stages, according to the extended Biologische Bundesanstalt, Bundessortenamt und CHemische Industrie (BBCH) scale [25]: (i) BBCH Code 53 - Inflorescences clearly visible; (ii) BBCH Code 55 - Inflorescences swelling, flowers closely pressed together; and (iii) BBCH Code 57 - Inflorescences fully developed; flowers separating.

The images were acquired in an experimental vineyard of the Agrarian Campus of Vairão, of the Faculty of Sciences of the University of Porto (41°24’12.2 "N 2°10’26.5 "W), using a dual camera Xiaomi Redmi Note 7 smartphone with a resolution of 8000x6000 pixels. The dataset includes images of the following national and international grapevine varieties: Touriga Nacional (VIVC-12594); Barroca (VIVC-12462); Tinta Roriz (VIVC-12350); Cabernet Sauvignon (VIVC-1929); Viosinho (VIVC-13109); Trajadura (VIVC-12629).

Although colour is not a differentiating feature at this phenological stage, red and white grapevine varieties were considered mainly due to the differences they exhibit in terms of size and shape of the inflorescences. In addition, the images were collected in various lighting and perspective conditions, often presenting scenarios of occlusion and overlap of inflorescences by different structures, inherent to the plant (i.e., leaves, stems, trunks or other inflorescences) or to the vineyard trellis and training system itself (i.e., cordon or foliage wires), adding complex and varied visual information. A total of

539 images compose the dataset, which is publicly available on the open-access digital repository Zenodo: https://doi.org/10.5281/zenodo.8332171.

The high resolution of the images translates into a large amount of data to be processed by the DL models. Thus, the resolution of the images was decreased to 1254x1672 pixels, retaining the same aspect ratio without losing an excessive amount of relevant information for the models learning. Following this procedure, the images were manually annotated using the open-source Computer Vision Annotation Tool (see https://cvat.org/, accessed on 1 August 2023). Since it involves image segmentation, each annotation contains a bounding box around each object, representing its area, position, and class, and a segmentation mask that enables to associate each pixel within the bounding box to a particular class. The generated masks were used to produce YOLO format annotations.

To train and validate the different models, the images were divided into 3 sets: (i) Train (60%); (ii) Validation (20%); and (iii) Test (20%). Train and Validation sets were artificially increase through Albumentations [26], a Python library for image augmentation, generating new data points from the existing dataset. The image transform operations were carefully chosen to only generate realistic vineyard images, such as: (i) CLAHE, (ii) Emboss, (iii) Sharpen, (iV) ISO Noise, (v) Random Fog, (vi) Spatter, (vii) Random Brightness Contrast, (viii) Blur, (ix) Gaussian Noise, (x) Horizontal Flip, and (xi) Shift Scale Rotate. These operations were not only applied individually, but combinations were also made, thus totalling 59 transforms applied to each image of the two sets. After the augmentation procedure, the dataset's size increased to 26,027 images. The training and validation sets contained 19,500 and 6,420 images, respectively, while the test set was composed of 107 images.

2.2. Model's Training and Inference

To correctly identify grapevine inflorescences, four YOLO models were benchmarked, since they have a strong reputation for its accuracy and speed, which is beneficial for live inference tasks and real-time applications: (i) YOLOv5n; (ii) YOLOv5s; (iii) YOLOv8n; and (iv) YOLOv8s. The models were pre-trained with Microsoft's COCO (Common Objects in Context) dataset [27] and through transfer learning, a fine-tune was performed to detect and segment grapevine inflorescences. Training sessions ran for 20 epochs, with a batch size of 16. PyTorch [28] was employed for the training and inference tasks, using a NVIDIA GeForce RTX 4060 graphics processing unit (GPU) with 8 GBs of available memory.

In segmentation tasks, a mask is predicted. A successful prediction is one which maximizes the overlap between the predicted and true objects. The two main metrics used to assess a "correct prediction" are the *Intersection over Union* (IoU) and *F1-Score*. Additionally, the metrics used by the Pascal VOC challenge [29], *Precision × Recall curve* and *Average Precision* (AP), were chosen to better benchmark the DL models. A key step in the models inference is the optimisation of the confidence threshold. For this purpose, a cross-validation technique was used. the F1-Score was computed for all the confidence thresholds, in the validation set, from 0% to 100%, into steps of 1%. The confidence threshold that optimises the F1-Score was selected and then the models were evaluated on the test set, considering a IoU >= 90%.

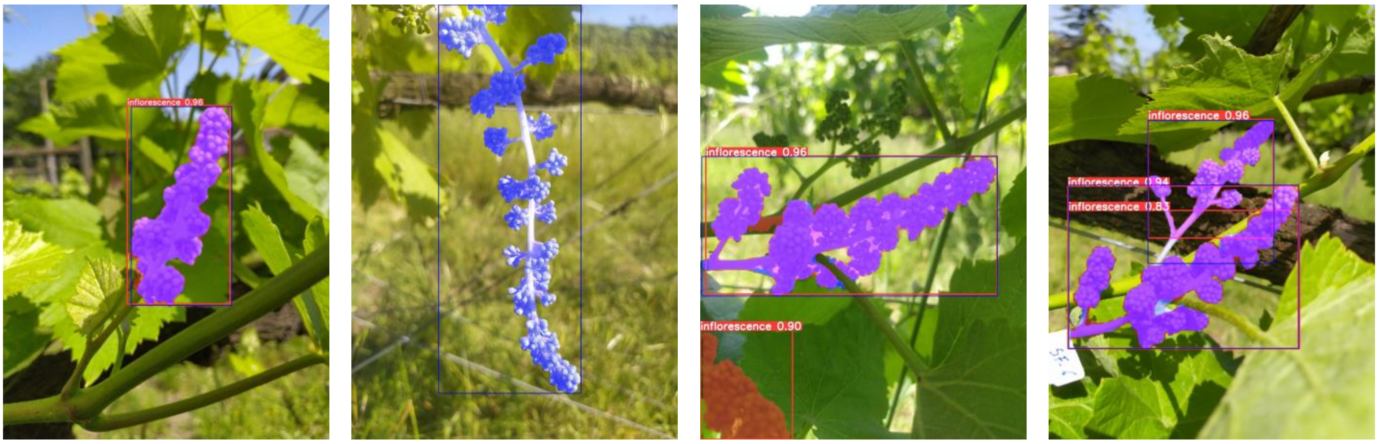
3. Results and Discussion

The models required defining the best confidence threshold that maximises the F1-Score before evaluating their performance. Usually, higher thresholds increases *Precision*, the percentage of correct detections, but decreases *Recall*, the ability to detect all relevant objects. Table 1 shows the results across the different metrics. The confidence threshold values presented lead to the best balance between *Precision* and *Recall* and all four models found their best F1-score above 65%, with the highest belonging to the YOLOv8s model at 82.7%. Overall, the results for the four models are encouraging and very similar, all above 90%. YOLOv8s has the best performance with regard to the location of objects in the image ($F1_{Box}$ = 95.1%), however the YOLOv5s outperformed all the other models in terms of the segmentation mask's quality ($F1_{Mask}$ = 97.7%). Another important factor when it comes to real-time applications is the inference time. Both YOLOv5 models are faster at detecting and segmenting than their YOLOv8 counterparts, which is to be expected given the size of the models.

Table 1. Detection and Segmentation results with the test set considering optimized confidence thresholds. (P = *Precision*; R = *Recall*; F1 = *F1-Score*)

Model	Confidence Threshold (%)	P_{Box} (%)	R_{Box} (%)	$F1_{Box}$ (%)	P_{Mask} (%)	R_{Mask} (%)	$F1_{Mask}$ (%)	Speed (ms)
YOLOv5n	76.1	93.5	91.7	92.6	96.3	94.5	95.4	4.5
YOLOv5s	67.8	93.8	96.3	95.0	96.4	99.1	97.7	9.6
YOLOv8n	73.0	92.8	94.9	93.8	95.5	97.8	96.6	6.8
YOLOv8s	82.7	93.0	97.2	95.1	94.7	99.1	96.9	12.3

To better understand the performance of the models and, above all, identify flaws and areas of improvement, it is essential to analyse the images from the test set. The strong performance is evident in all the models (a), but it is clear that the results could have been better had it not been for some errors, such as non-detections (b), detections of non-annotated inflorescences (c) and multiple detections of the same inflorescence (d), as Figure 1 illustrates.



(a) (b) (c) (d)

Figure 1. Detection and segmentation of grapevine inflorescence test set samples: (a) correct detection (YOLOv5n), (b) missed detection (YOLOv5s), (c) detections of non-annotated inflorescences (YOLOv8n) and (d) multiple detections of the same inflorescence (YOLOv8s). Red bounding boxes represent the models predictions and blue bounding boxes represent the groundtruth annotations.

To understand the relevance of the results obtained, it becomes essential to compare them with the current literature. To the authors' knowledge, all the models evaluated outperformed the existing literature, as far as inflorescence segmentation is concerned, with the advantage of using a robust dataset under uncontrolled conditions. Certain studies have taken the approach of capturing images at night using artificial light, allowing for greater homogeneity, trying to extract the complexity provided by the background. These are the cases of Palacios *et al.* [20] and Rahim *et al.* [22], who through the SegNet (VGG19) and Mask-RCNN models obtained *F1-Scores* of 93.0% and 94.3%, respectively. However, it should be noted that the images were taken at a longer distance, which makes the task of detection and segmentation more difficult. The scarcity of images is also a problem, with the majority of works presenting datasets with less than 10,000 images. Rudolph *et al.* [16], for example, tested a AlexNet-based FCN on just 10 images, achieving a *mean IoU* of 76.0%.

All in all, the results presented are hopeful about the success of detecting and segmenting inflorescences, but drawbacks such as the low robustness of the datasets and the poor specification of the evaluation metrics need to be addressed in order to take the next step towards automating these tasks.

4. Conclusions

In this paper, four pre-trained YOLO models were benchmarked in grapevine inflorescence detection and segmentation. One dataset of inflorescence images was acquired under uncontrolled conditions for that purpose.

The results obtained were promising, with all models achieving F1-Scores above 90%. The YOLOv8s and YOLOv5s models stood out, achieving a $F1\text{-}Score_{Box}$ of 95.1% and a $F1\text{-}Score_{Mask}$ of 97.7%, for the detection and segmentation tasks respectively. Allied to this performance, the low inference times recorded (under 13ms), where the Yolov5s model showed the best trade-off, prove the suitability of these models for deployment in real-time applications and the ability to support algorithms capable of counting flowers in the field in a non-destructive way, allowing for more accurate and robust sampling and forecasting.

In perspective, future work should go through: (i) enlarge the dataset with images from farther distances, to be able to infer the number of inflorescences per vine ; (ii) evaluate the performance of these models in real-time conditions in a vineyard and (iii) incorporate these models into a framework that allows the subsequent counting of the flower number per inflorescence.

Author Contributions: Conceptualisation, G.M., S.A.M., F.N.d.S. and M.C.; data curation, G.M.; funding acquisition, F.N.d.S.; investigation, G.M.; methodology, G.M.; project administration, F.N.d.S.; software, G.M. and S.A.M.; supervision, M.C. and F.N.d.S.; validation, M.C. and F.N.d.S.; visualisation, G.M.; writing—original draft, G.M.; writing—review and editing, S.A.M, M.C. and F.N.d.S. All authors read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The data presented in this study are openly available in the digital repository Zenodo: "GVXmi | Grapevine Inflorescence Datatset" https://doi.org/10.5281/zenodo.8332171.

Acknowledgments: The authors would like to acknowledge the scholarship number "2022.09726.BD", funded by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e Tecnologia.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mekouar, M.A. Food and Agriculture Organization of the United Nations (FAO). *Yearbook of International Environmental Law* **2020**, *31*, 326–340.
2. Carrillo, E.; Matese, A.; Rousseau, J.; Tisseyre, B. Use of multi-spectral airborne imagery to improve yield sampling in viticulture. *Precision agriculture* **2016**, *17*, 74–92.
3. Clingeleffer, P.R.; Martin, S.; Dunn, G.; Krstic, M. Crop development, crop estimation and crop control to secure quality and production of major wine grape varieties: a national approach: final report to Grape and Wine Research **2001**.
4. Arnó, J.; Casasnovas, M.; Ribes-Dasi, M.; Rosell-Polo, J. Review. Precision Viticulture. Research topics, challenges and opportunities in site-specific vineyard management. *Spanish Journal of Agricultural Research* **2009**, *7*. https://doi.org/10.5424/sjar/2009074-1092.
5. Tardaguila, J.; Stoll, M.; Gutiérrez, S.; Proffitt, T.; Diago, M.P. Smart applications and digital technologies in viticulture: A review. *Smart Agricultural Technology* **2021**, *1*, 100005. https://doi.org/https://doi.org/10.1016/j.atech.2021.100005.
6. Diago, M.P.; Sanz García, A.; Millan, B.; Blasco, J.; Tardaguila, J. Assessment of Flower Number Per Inflorescence in Grapevine by Image Analysis Under Field Conditions. *Journal of the science of food and agriculture* **2014**, *94*. https://doi.org/10.1002/jsfa.6512.
7. Millan, B.; Aquino, A.; Diago, M.P.; Tardaguila, J. Image analysis-based modelling for flower number estimation in grapevine. *Journal of the science of food and agriculture* **2016**, *97*. https://doi.org/10.1002/jsfa.7797.

8. Aquino, A.; Millan, B.; Gaston, D.; Diago, M.P.; Tardaguila, J. vitisFlower®: Development and Testing of a Novel Android-Smartphone Application for Assessing the Number of Grapevine Flowers per Inflorescence Using Artificial Vision Techniques. *Sensors* **2015**, *15*, 21204–21218. https://doi.org/10.3390/s150921204.
9. Aquino, A.; Millan, B.; Gutiérrez, S.; Tardáguila, J. Grapevine flower estimation by applying artificial vision techniques on images with uncontrolled scene and multi-model analysis. *Computers and Electronics in Agriculture* **2015**, *119*, 92–104. https://doi.org/https://doi.org/10.1016/j.compag.2015.10.009.
10. Liu, S.; Li, X.; Wu, H.; Xin, B.; Tang, J.; Petrie, P.R.; Whitty, M. A robust automated flower estimation system for grape vines. *Biosystems Engineering* **2018**, *172*, 110–123. https://doi.org/https://doi.org/10.1016/j.biosystemseng.2018.05.009.
11. Tello, J.; Herzog, K.; Rist, F.; This, P.; Doligez, A. Automatic Flower Number Evaluation in Grapevine Inflorescences Using RGB Images. *American Journal of Enology and Viticulture* **2019**, [https://www.ajevonline.org/content/early/2019/09/12/ajev.2019.19036.full.pdf]. https://doi.org/10.5344/ajev.2019.19036.
12. Mohimont, L.; Alin, F.; Rondeau, M.; Gaveau, N.; Steffenel, L.A. Computer Vision and Deep Learning for Precision Viticulture. *Agronomy* **2022**, *12*. https://doi.org/10.3390/agronomy12102463.
13. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural networks* **2015**, *61*, 85–117.
14. Victorino, G.; Braga, R.; Santos-Victor, J.; Lopes, C.M. Yield components detection and image-based indicators for non-invasive grapevine yield prediction at different phenological phases. *Oeno One* **2020**.
15. Grimm, J.; Herzog, K.; Rist, F.; Kicherer, A.; Töpfer, R.; Steinhage, V. An adaptable approach to automated visual detection of plant organs with applications in grapevine breeding. *Biosystems Engineering* **2019**, *183*, 170–183. https://doi.org/https://doi.org/10.1016/j.biosystemseng.2019.04.018.
16. Rudolph, R.; Herzog, K.; Toepfer, R.; Steinhage, V. Efficient identification, localization and quantification of grapevine inflorescences and flowers in unprepared field images using Fully Convolutional Networks. *Vitis -Geilweilerhof-* **2019**, *58*, 95–104. https://doi.org/10.5073/vitis.2019.58.95-104.
17. Khokher, M.R.; Liao, Q.; Smith, A.L.; Sun, C.; Mackenzie, D.; Thomas, M.R.; Wang, D.; Edwards, E.J. Early Yield Estimation in Viticulture Based on Grapevine Inflorescence Detection and Counting in Videos. *IEEE Access* **2023**, *11*, 37790–37808. https://doi.org/10.1109/ACCESS.2023.3263238.
18. Pahalawatta, K.; Fourie, J.; Parker, A.; Carey, P.; Werner, A. Detection and classification of opened and closed flowers in grape inflorescences using Mask R-CNN. In Proceedings of the 2020 35th International Conference on Image and Vision Computing New Zealand (IVCNZ), 2020, pp. 1–6. https://doi.org/10.1109/IVCNZ51579.2020.9290720.
19. Rahim, U.; Mineno, H.; Tomoyoshi, U. Comparison of Grape Flower Counting Using Patch-Based Instance Segmentation and Density-Based Estimation with Convolutional Neural Networks Comparison of grape flower counting using patch-based instance segmentation and density-based estimation with Convolutional Neural Networks **2021**.
20. Palacios, F.; Bueno, G.; Salido, J.; Diago, M.P.; Hernández, I.; Tardaguila, J. Automated grapevine flower detection and quantification method based on computer vision and deep learning from on-the-go imaging using a mobile sensing platform under field conditions. *Computers and Electronics in Agriculture* **2020**, *178*, 105796. https://doi.org/https://doi.org/10.1016/j.compag.2020.105796.
21. Jaramillo, J.; Vanden Heuvel, J.; Petersen, K.H. Low-Cost, Computer Vision-Based, Prebloom Cluster Count Prediction in Vineyards. *Frontiers in Agronomy* **2021**, *3*. https://doi.org/10.3389/fagro.2021.648080.
22. Rahim, U.; Utsumi, T.; Mineno, H. Deep learning-based accurate grapevine inflorescence and flower quantification in unstructured vineyard images acquired using a mobile sensing platform. *Computers and Electronics in Agriculture* **2022**, *198*, 107088. https://doi.org/10.1016/j.compag.2022.107088.
23. Buayai, P.; Yok-In, K.; Inoue, D.; Leow, C.S.; Nishizaki, H.; Makino, K.; Mao, X. End-to-End Inflorescence Measurement for Supporting Table Grape Trimming with Augmented Reality. In Proceedings of the 2021 International Conference on Cyberworlds (CW), 2021, pp. 101–108. https://doi.org/10.1109/CW52790.2021.00022.
24. Du, W.; Zhu, Y.; Li, S.; Liu, P. Spikelets detection of table grape before thinning based on improved YOLOV5s and Kmeans under the complex environment. *Computers and Electronics in Agriculture* **2022**, *203*, 107432. https://doi.org/https://doi.org/10.1016/j.compag.2022.107432.
25. Meier, U. *Growth stages of mono-and dicotyledonous plants*; Blackwell Wissenschafts-Verlag, 1997.
26. Buslaev, A.; Iglovikov, V.I.; Khvedchenya, E.; Parinov, A.; Druzhinin, M.; Kalinin, A.A. Albumentations: Fast and Flexible Image Augmentations. *Information* **2020**, *11*. https://doi.org/10.3390/info11020125.
27. Lin, T.Y.; Maire, M.; Belongie, S.; Bourdev, L.; Girshick, R.; Hays, J.; Perona, P.; Ramanan, D.; Zitnick, C.L.; Dollár, P. Microsoft COCO: Common Objects in Context, 2015, [arXiv:cs.CV/1405.0312].
28. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library, 2019, [arXiv:cs.LG/1912.01703].
29. Everingham, M.; Gool, L.; Williams, C.K.; Winn, J.; Zisserman, A. The Pascal Visual Object Classes (VOC) Challenge. *Int. J. Comput. Vision* **2010**, *88*, 303–338. https://doi.org/10.1007/s11263-009-0275-4.