

Article

Not peer-reviewed version

# Object Detection in Laparoscopic Surgery: A Comparative Study of Deep Learning Models on a Custom Endometriosis Dataset

[Andrey Bondarenko](#) , [Vilen Jumutc](#) , [Antoine Netter](#) , Fanny Duchateau , [Henrique Mendonca Abrão](#) , Saman Noorzadeh , [Giuseppe Giacomello](#) , Filippo Ferrari , Nicolas Bourdel , [Ulrik Bak Kirk](#) , [Dmitrijs Blizņuks](#) \*

Posted Date: 25 December 2024

doi: 10.20944/preprints202412.2127.v1

Keywords: Endometriosis; Deep Learning; object detection; R-CNN









Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

## Article

# Object Detection in Laparoscopic Surgery: A Comparative Study of Deep Learning Models on a Custom Endometriosis Dataset

Andrey Bondarenko <sup>1</sup>, Vilen Jumutc <sup>1</sup>, Antoine Netter <sup>2</sup>, Fanny Duchateau <sup>2</sup>, Henrique Abrão <sup>3</sup>, Saman Noorzadeh <sup>4</sup>, Giuseppe Giacomello <sup>4,5</sup>, Filippo Ferrari <sup>4,6</sup>, Nicolas Bourdel <sup>4,7</sup>, Ulrik Bak Kirk <sup>8,9</sup>, Dmitrijs Bļizņuks <sup>10,\*</sup>

<sup>1</sup> Riga Technical University, Institute of Applied Computer Systems

<sup>2</sup> Department of Obstetrics and Gynecology, Marseille Hospital

<sup>3</sup> Gynecologic Division, Beneficência Portuguesa de São Paulo

<sup>4</sup> SurgAR, Clermont-Ferrand

<sup>5</sup> Department of Obstetrics and Gynecology, Istituto Ospedaliero Fondazione Poliambulanza, Brescia

<sup>6</sup> Department of Obstetrics and Gynecology, Gynecologic Oncology and Minimally Invasive Pelvic Surgery, International School of Surgical Anatomy (ISSA), IRCCS "Sacro Cuore - Don Calabria" Hospital, Negrar di Valpolicella, Verona

<sup>7</sup> Department of Clinical Research and Innovation, CHU Clermont Ferrand

<sup>8</sup> Department of Public Health, Aarhus University

<sup>9</sup> The Research Unit for General Practice, Aarhus

<sup>10</sup> Riga Technical University, Institute of Applied Computer Systems

\* Correspondence: dmitrijs.bliznuks@rtu.lv; Tel.: +(371)-26707961

**Abstract:** Laparoscopic surgery for endometriosis presents unique challenges due to the complexity and variability of lesion appearances within the abdominal cavity. This study investigates the application of deep learning models for object detection in laparoscopic videos, aiming to assist surgeons in accurately identifying and localising endometriosis lesions and related anatomical structures. A custom dataset was curated, comprising 332 video sequences and 354,906 frames. Of these, 17,560 frames were meticulously annotated by medical professionals. The dataset includes object-detection annotations for 10 object classes relevant to endometriosis, alongside segmentation masks for some classes. To address the object detection task, we evaluated the performance of three deep learning models - FasterRCNN, MaskRCNN, and YOLOv9 - under both stratified and non-stratified training scenarios. Experimental results demonstrated that stratified training significantly reduced the risk of data leakage and improved model generalization. The best-performing object detection model achieved high precision and recall across most classes. Despite these successes, the study also highlights the challenges posed by the weak annotations and class imbalances in the dataset, which impacted overall models performances. In conclusion, this study provides valuable insights into the application of deep learning for enhancing laparoscopic surgical precision in endometriosis treatment. The findings underscore the importance of robust dataset curation and advanced training strategies in developing reliable AI-assisted tools for surgical interventions. Future work will focus on refining the dataset and exploring more sophisticated model architectures to further improve detection accuracy.

**Keywords:** endometriosis; deep learning; object detection; R-CNN

## 1. Introduction

Endometriosis is a chronic condition affecting an estimated 5%-10% of women and adolescents of reproductive age (15-49 years). Recent data indicate an incidence rate of 2.37–2.49 per 1000 women per year, translating to an approximate prevalence of 6%-8% [1]. The condition frequently goes undiagnosed or misdiagnosed, with an average delay of 4-11 years between the onset of symptoms and definitive diagnosis. While endometriosis can occur at any age, it predominantly affects women between menarche and menopause, peaking between the ages of 25 and 45 years. Notably, the prevalence is higher among individuals experiencing infertility and chronic pelvic pain [2].

This disease is characterized by the presence of endometrial-like tissue outside the uterus, leading to debilitating symptoms and complications, including infertility. The complexity of endometriosis and the variability of lesion appearances often pose significant challenges for diagnosis and treatment. Laparoscopic surgery, a minimally invasive technique, is considered the gold standard for both the diagnosis and management of endometriosis. Despite its advantages, such as reduced recovery times and minimal scarring, the procedure is hindered by the intricate and variable nature of the disease, necessitating high levels of expertise to accurately identify and treat lesions.

Recent advances in artificial intelligence (AI) have shown promise in addressing some of these challenges. Deep learning models, particularly in object detection, offer the potential to assist surgeons in identifying and localizing endometriosis lesions and related anatomical structures during laparoscopic procedures. A custom dataset curated for this purpose, comprising 199 laparoscopic videos with over 205,725 frames and extensive annotations by medical professionals, was utilized to explore the application of various AI techniques. These approaches include well-known object detection algorithms such as FasterRCNN [3], MaskRCNN [4], and YOLOv9 [5], along with custom encoder-decoder segmentation networks - employed not to yield segmentation masks as the final output, but rather to facilitate bounding-box creation from the segmentation results.

The study emphasizes the importance of addressing challenges such as class imbalances through tailored augmentation techniques and loss functions. By leveraging these innovations, the research aims to establish a robust framework capable of adapting to the complexities of laparoscopic surgery. The integration of AI into surgical practice has the potential to enhance diagnostic precision, improve surgical outcomes, and ultimately contribute to better patient care.

## 2. Related Work

### 2.1. Existing AI applications

Up until recent times research mostly focused on diagnosing endometriosis via ultrasound imaging or MRI imaging [6] or via biomarker or metabolite analysis (see review in [7]). Recent research tackles the problem of the endometrioma image classification [8] and detection [9]. These image analysis-based applications focus on using deep learning networks with convolutional backbones for object detection or segmentation tasks.

### 2.2. Review of Existing Datasets

Several datasets have been developed to support research in laparoscopic surgery image analysis, though most are not specifically tailored to the challenges posed by endometriosis.

Among the available data resources in this domain, some provide extensive pixel-level annotations of numerous anatomical structures extracted from surgical videos, ensuring a rich representation of organs and tissues across thousands of frames [10]. Others concentrate on a wide range of abdominal organs and vessels, offering large, high-quality annotations that are valuable for general surgical data science but lack targeted pathological focus [11]. There are resources specifically designed for particular conditions, such as endometriosis, which include annotated images that are directly relevant to this pathology [12]. However, the number of such annotated images often remains limited, and the specificity of covered lesions can restrict broader applicability. Another innovative approach combines real surgical footage with synthetic data generated through semantic image synthesis, improving segmentation performance and enabling extensive experimentation with cutting-edge models [13].

Despite these advancements, challenges remain. Some datasets focus narrowly on certain procedures or do not capture the unique complexities of endometriosis, while others rely heavily on synthetic data that may not fully translate into clinical practice. Each resource, therefore, presents a mix of outstanding features - such as comprehensive organ annotations or innovative data augmentation methods - and notable issues, including insufficient coverage of specific conditions, limited annotated samples, or potential gaps in real-world applicability. These datasets, while valuable, do not fully

address the unique challenges posed by endometriosis in laparoscopic surgery, particularly the need for comprehensive annotations across a wide variety of lesion types and surgical scenarios.

2.3. Limitations of Existing Datasets

The primary limitations of the existing datasets include small sample sizes, limited annotations, and a focus on either general anatomical structures or specific types of surgical procedures. For instance, while CholecSeg8K [10] and the Dresden Surgical Anatomy Dataset [11] offer detailed segmentations, they do not cover the pathological conditions relevant to endometriosis. On the other hand, GLEND A [12], while directly addressing endometriosis, is limited by the number of annotated frames and the variability of lesion types represented.

These limitations underscore the need for a new dataset that provides detailed, endometriosis-specific annotations across a broad range of surgical scenarios. Such a dataset would better support the development of AI models capable of assisting in the diagnosis and treatment of this complex condition.

3. Materials and Methods

The section introduces our custom endometriosis dataset and corresponding Deep Learning methodology to tackle the aforementioned object detection problem in laparoscopic images. First the dataset collection, composition and annotation details are described. Then machine and deep learning methodology along with experimentation setup are outlined.

3.1. Endometriosis Dataset

3.1.1. Overview

To address the complex challenge of object detection in laparoscopic videos for the treatment of endometriosis, we curated a comprehensive dataset specifically designed to support the development and evaluation of deep learning models. This dataset is meticulously tailored to facilitate the detection and classification of various endometriosis lesions and associated anatomical structures, providing a robust foundation for advancing AI-driven surgical tools.

3.1.2. Data Collection and Composition

The dataset comprises carefully chosen 332 laparoscopic videos, resulting in a total of 354,906 frames. Of these 17,560 frames have been meticulously annotated by experienced medical professionals. For the object detection task, annotations were focused on short sequences, typically consisting of 10–11 consecutive frames, while longer frames sequences were left unannotated, to ensure detailed and accurate labelling where it is most needed. This selective approach was adopted to effectively manage the large volume of data while capturing critical surgical moments and ensuring high-quality annotations.

The videos in the dataset were recorded at varying resolutions, reflecting the diverse conditions encountered during laparoscopic procedures. The majority of the videos were captured at a resolution of 1920 × 1080 pixels, providing high-definition imagery crucial for accurate annotation and model training. The distribution of video resolutions is summarized in Table 1.

Table 1. Video Resolutions.

Resolution	Number of Videos
1920 × 1080	193
1280 × 720	5
720 × 576	1

3.1.3. Annotation Details

Annotations within the dataset for object-detection tasks are provided in the form of bounding boxes and segmentation regions, covering 10 distinct classes relevant to endometriosis and other anatomical structures. These classes include various types of adhesions, endometriomas, and superficial lesions, which are critical for surgical decision-making. The distribution of annotated objects across these classes is detailed in Table 2.

Table 2. Class Distribution for Object-Detection.

Class Name	Class ID	Number of Annotated Objects
Adhesions Dense	0	1,424
Adhesions Filmy	1	537
Deep Endometriosis	2	700
Ovarian Chocolate Fluid	3	223
Ovarian Endometrioma	4	302
Ovarian Endometrioma[B]	4	382
Superficial Black	5	835
Superficial Red	6	642
Superficial Subtle	7	509
Superficial White	8	463

The varying frequency of these object classes highlights potential challenges in model performance, particularly concerning class imbalance. It is crucial to address this imbalance to ensure that the models can effectively detect both common and rare conditions, thus improving their robustness and generalizability in real-world scenarios.

3.1.4. Video Characteristics

The dataset includes videos of varying lengths, which contributes to the diversity and complexity of the data. This variation mirrors real-world surgical scenarios, where the duration of procedures can significantly differ depending on the complexity of the case and the surgeon’s expertise. The distribution of video lengths, measured in the number of frames, provides insight into the range of surgical cases represented within the dataset.

Frames were carefully extracted at key surgical moments to ensure that the annotated frames capture critical events and relevant anatomical structures. The high resolution and clarity of these frames provide a solid foundation for training and evaluating object detection models, facilitating the development of tools that can assist surgeons in making precise and timely decisions during surgery.

3.1.5. Data Organization and Accessibility

The dataset is organized into a well-structured directory system that enhances accessibility and usability. Each video is stored in a dedicated folder, containing both the raw video frames and the corresponding annotations. The annotations are stored separately with a clear naming convention that indicates the class and type of annotation (bounding box), allowing researchers to quickly locate and utilize the relevant data for their specific tasks. This organizational structure is designed to streamline the research process and ensure that the data can be efficiently integrated into various machine learning workflows.

3.1.6. Considerations and Challenges

While this dataset provides a valuable resource for the development of object detection models in laparoscopic surgery, it poses certain challenges. One of the primary challenges is the variability in interpretation during the annotation process, despite being conducted by experienced medical professionals. This can lead to potential inconsistencies in the labels, which may affect model performance. Moreover, the proprietary nature of the dataset limits its broader distribution, which may restrict its



use in the wider research community. This limitation could pose challenges for collaborative research efforts and the reproducibility of findings.

Another significant challenge is the class imbalance within the dataset. Certain lesion types are underrepresented, which could bias the models toward detecting more common conditions at the expense of rarer ones. Addressing this imbalance through techniques such as data augmentation, class re-weighting, or synthetic data generation will be crucial for developing models that are both robust and generalizable across diverse surgical scenarios.

### 3.2. Methods

In this section, we describe the methodology employed for training and evaluating object detection models on our laparoscopic endometriosis dataset. The primary focus is on comparing the performance of FasterRCNN [3], MaskRCNN [4,14], and YOLOv9 [5] object detection models using both stratified and non-stratified data splits [15]. We also discuss the unique challenges posed by the dataset, such as annotation inconsistencies and class imbalance, and how we addressed them.

### 3.3. Data Preprocessing

The laparoscopic endometriosis dataset consists of high-resolution videos, with the frames mostly captured at  $1920 \times 1080$  resolution. To preserve as much data as possible we opted for the usage of this resolution for MaskRCNN and FasterRCNN models. But for the YOLOv9 model we have downsampled images to  $640 \times 640$  pixels as it is a default model resolution which would produce a near real-time inference.

Bounding boxes were resized proportionally to match the new image dimensions, and stratified data splits [15] were applied to ensure that all classes were represented equally in the train, validation, and test sets. Stratification was achieved using the Python's iterative-stratification package which supports various data splitting strategies for multi-label scenarios. To prevent any data leakage we have split videos given multiple labels per example into train/test/validation sets, thus entire video frames from the train set videos are never appearing in the validation and test sets. A failure to prevent common data leakages can pave the way to unsatisfactory results (undetected model overfitting) [16]. We opted for exploring effects of a data leakage thus non-stratified data splits were created (having a data leak in between train and validation datasets, but not in train/test or validation/test sets).

### 3.4. Object Detection Models

We employed three object detection models - FasterRCNN, MaskRCNN, and YOLOv9 - to identify and classify endometriosis lesions and other relevant anatomical structures.

1. **FasterRCNN:** We used a FasterRCNN model with a ResNet50 backbone, initialised with pre-trained weights from the ImageNet1K\_V2 model in PyTorch's torchvision model zoo. The model has 60.27M parameters and was trained with heavy augmentations to handle a high variability in the dataset.
2. **MaskRCNN:** This model [14], also based on the ResNet50 backbone, initialised with pretrained weights from the ImageNet1K\_V2 model, was trained using bounding box-based segmentation masks. Due to the lack of precise segmentation maps, we filled bounding boxes to create rectangular masks, which impacted the models performance. Copy-paste augmentation was not applied to MaskRCNN. The MaskRCNN model was employed in an unconventional manner to generate bounding boxes. This approach was hypothesized to be effective based on the characteristics of the dataset and the nature of the annotated tissue classes. One of the tissue classes in the dataset was provided with a few hundred precise segmentation masks, which were utilized to generate bounding boxes for that specific class, ensuring consistency in annotation and model input format. Additionally, two classes in the dataset, originally annotated using bounding boxes, suffered from a significant inclusion of background elements. This issue stemmed from the physical appearance of the tissues, which often resembled string-like structures and were

predominantly located at oblique angles. Consequently, the bounding boxes captured substantial portions of the surrounding background, reducing annotation precision. By using the MaskRCNN model to generate bounding boxes based on segmentation masks, it was hypothesized that the quality and relevance of the annotations for these tissue classes would be improved. This method aimed to ensure that the annotations better represented the target structures while minimizing unnecessary background information, thus enhancing model training and performance.

3. **YOLOv9:** We utilized the largest [YOLOv9e](#) model, known for its extensive architecture and high performance on object detection tasks. The model was pretrained on the COCO dataset [17] and supports  $640 \times 640$  resolution inputs. It was fine tuned on our dataset with the default [Ultralytics augmentations](#).

Each model was evaluated on the stratified and non-stratified data splits to compare their performances in the scenarios with varying degrees of a data leakage.

### 3.5. Training Strategy

#### 3.5.1. Stratified vs. Non-Stratified Splits

To ensure robust results, we performed experiments with both stratified and non-stratified splits. In stratified splits, frames from the same video were placed in the same subset (train, validation, or test) to prevent data leakage. In non-stratified splits, train and validation frames were randomly sampled from the same videos, while test frames were held out from separate videos. This allowed us to compare performance in a more realistic setup and evaluate the risk of overfitting.

#### 3.5.2. Augmentation Techniques

Given the limited variability in the dataset (199 videos), heavy augmentations were crucial to improving model performance and preventing overfitting. For FasterRCNN and MaskRCNN augmentations included Resize, Normalise, CenterCrop, and RandomFlip. Additionally, we implemented a custom augmentation pipeline, which resizes bounding boxes accordingly to maintain label consistency. The most important augmentation used for FasterRCNN only was so-called Copy-Paste augmentation [18] used by the authors of YoloV9 model. For YOLOv9 model we used default Ultralytics augmentations, including rotations, flips, mosaic, and colour augmentations. Generally speaking, the most advanced augmentations are considered to be the Copy-Paste one (used in our experiments with FasterRCNN) and the Mosaic augmentation, which extends the MixUp augmentation [19].

### 3.6. Experimental Setup

In our experiments, we used several models with consistent training setups. The FasterRCNN model was trained with a batch size of 8 and the Adamax optimizer (learning rate of 0.0001, betas set to (0.9, 0.999), epsilon  $1e-08$ , and no weight decay) over 125 epochs. Similarly, the MaskRCNN model used a batch size of 8, with the Adam optimizer [20] (learning rate of 0.0005) for 125 epochs. Finally, the YOLOv9 model was trained with a batch size of 16, using the Adam optimizer with default values for 5 epochs.

#### 3.6.1. Evaluation Metrics

To evaluate the performance of our object detection models, we utilized several key metrics:

1. **Precision:** measures the proportion of correctly predicted positive instances among all predicted positives.
2. **Recall:** reflects the model's ability to identify true positives (actual lesions).
3. **mAP50:** evaluates localization accuracy at an Intersection over Union (IoU) threshold of 0.50.
4. **mAP50-95:** assesses mean Average Precision over a range of IoU thresholds (0.50 to 0.95).
5. **Fitness:** a combined metric that balances precision and recall, often represented by the F-1 Score.

For all metrics, we report the mean values alongside standard deviations to provide insights into the consistency and stability of the models' performance across different data splits (stratified and

non-stratified). The precision, recall, and mAP scores allow us to assess the strengths and weaknesses of each model in detecting and segmenting endometriosis lesions, while the fitness metric gives an overall perspective on the model’s reliability.

3.7. Challenges and Limitations

Several challenges were encountered during training. Overfitting was the main problem for all models. YOLOv9 showed signs of overfitting after just 5 epochs, likely due to the small dataset size in relation to the complexity of the model. To mitigate this, we decided against using Transformer-based models (e.g. ViT [21]) due to their propensity for overfitting on small datasets. Despite minimizing both training and validation loss during the training, MaskRCNN struggled with making accurate predictions and eventually the model stopped producing any predictions at all. We hypothesize that the use of bounding box-generated segmentation masks may have led the model to avoid predictions as a means of reducing loss. Introducing more complex augmentations or adjusting the loss function to place less emphasis on minor segmentation misalignments could improve the performance.

4. Results

4.1. Main Results

The performance of YOLOv9 and FasterRCNN is analyzed across both stratified and non-stratified splits, with particular attention to precision, recall, and mIoU scores. In general, FasterRCNN outperformed YOLOv9 in terms of overall precision and recall, with YOLOv9 showing more variability, especially in the stratified splits. Precision-recall and F1 score curves highlight these trends, with FasterRCNN maintaining high precision and recall across the train, validation, and test sets. The main findings can be outlined in Figures 3-6.

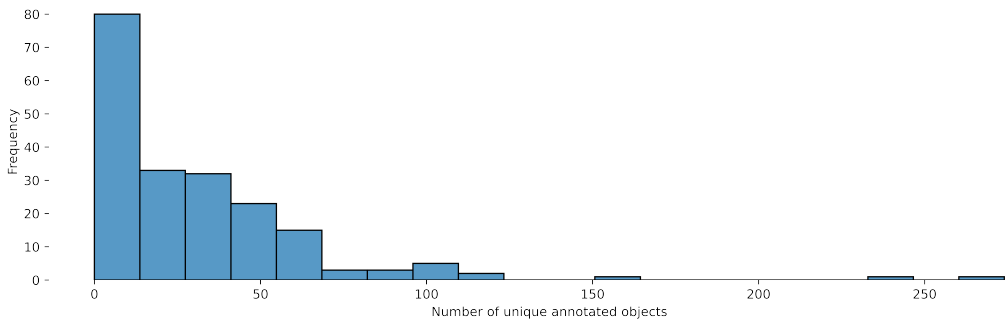


Figure 1. Object-Detection distribution of the number of unique annotated objects per video.

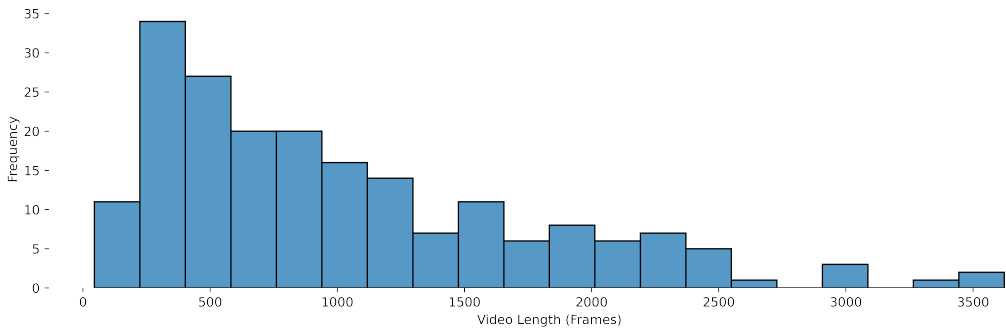


Figure 2. Object-Detection distribution of video lengths (in number of the frames).



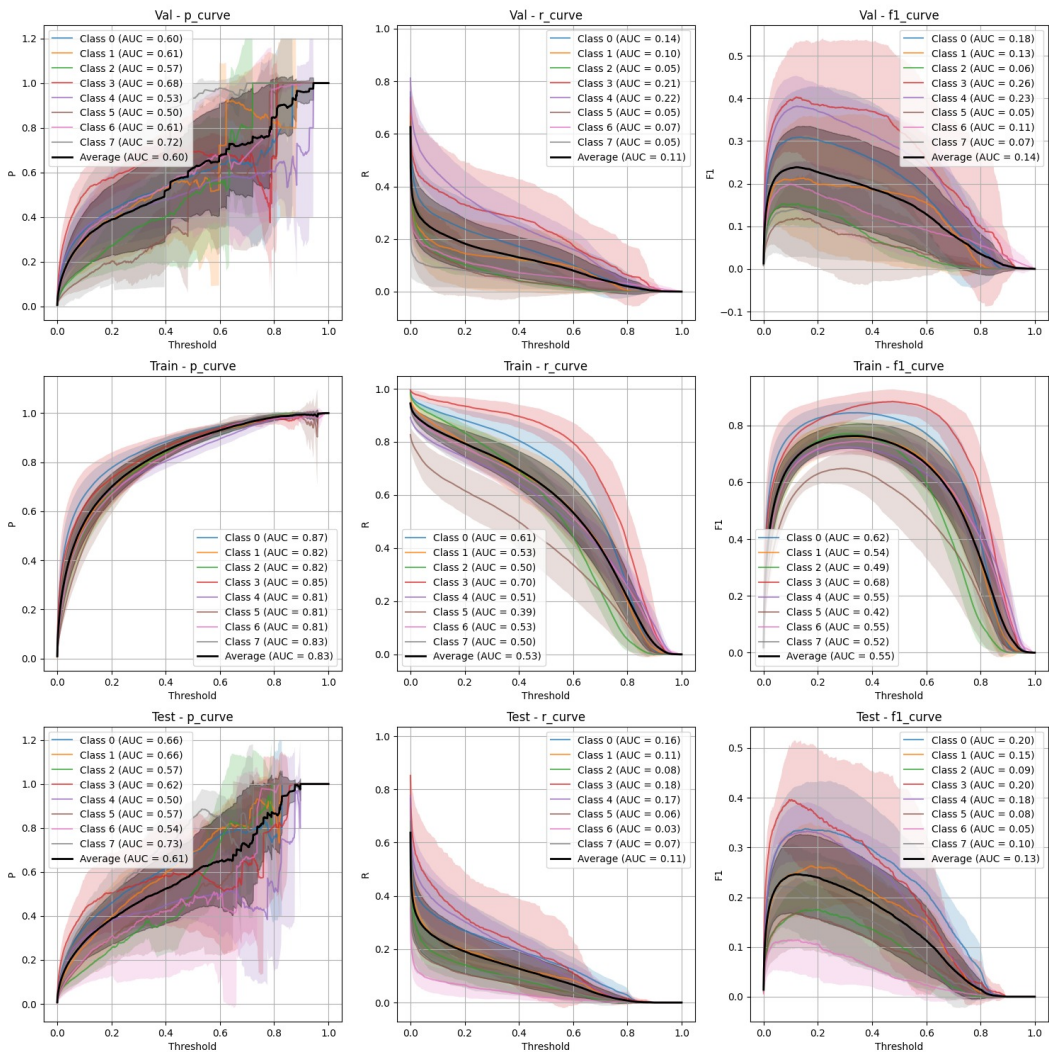


Figure 3. YOLOv9 Stratified Train/Val/Test Precision, Recall, and F-1 Curves.

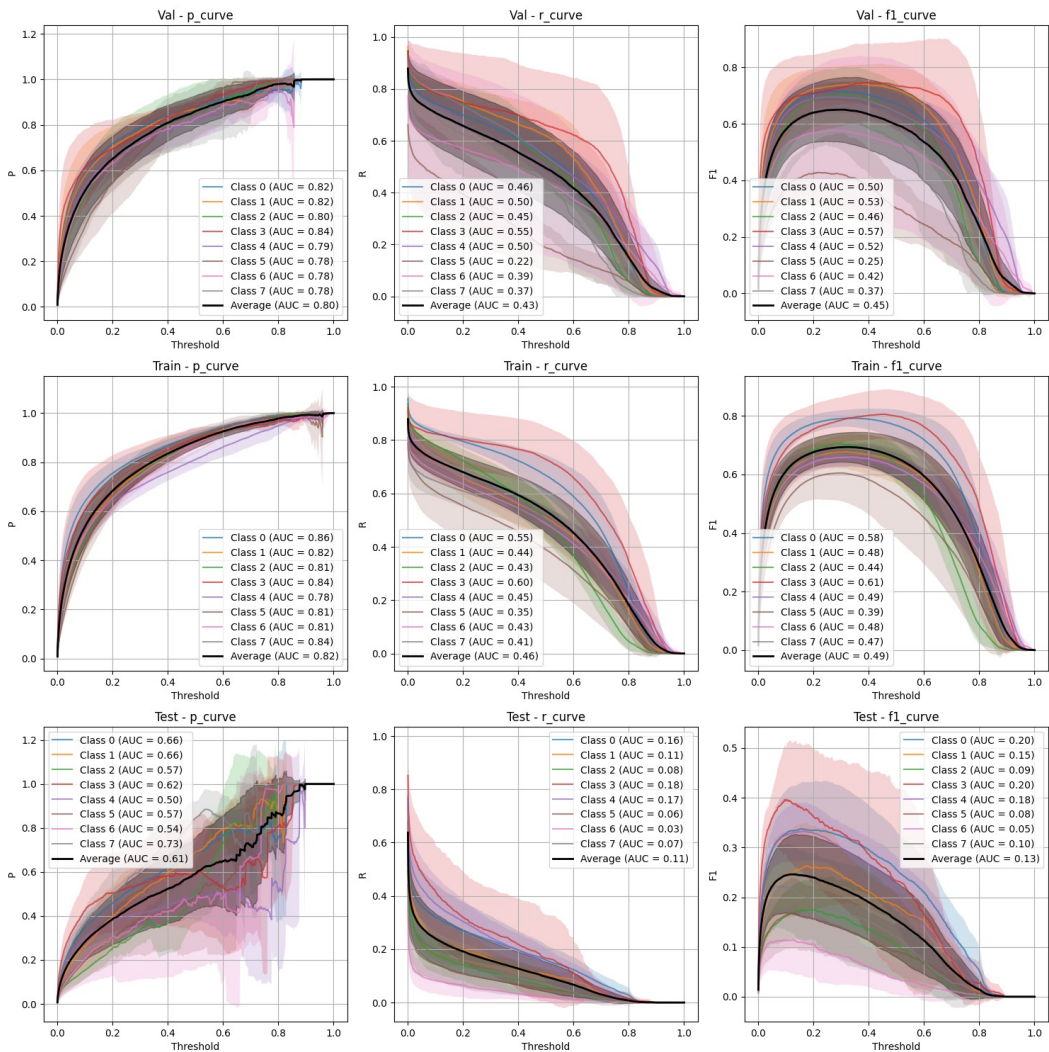


Figure 4. YOLOv9 Non-Stratified Train/Val/Test Precision, Recall, and F-1 Curves.

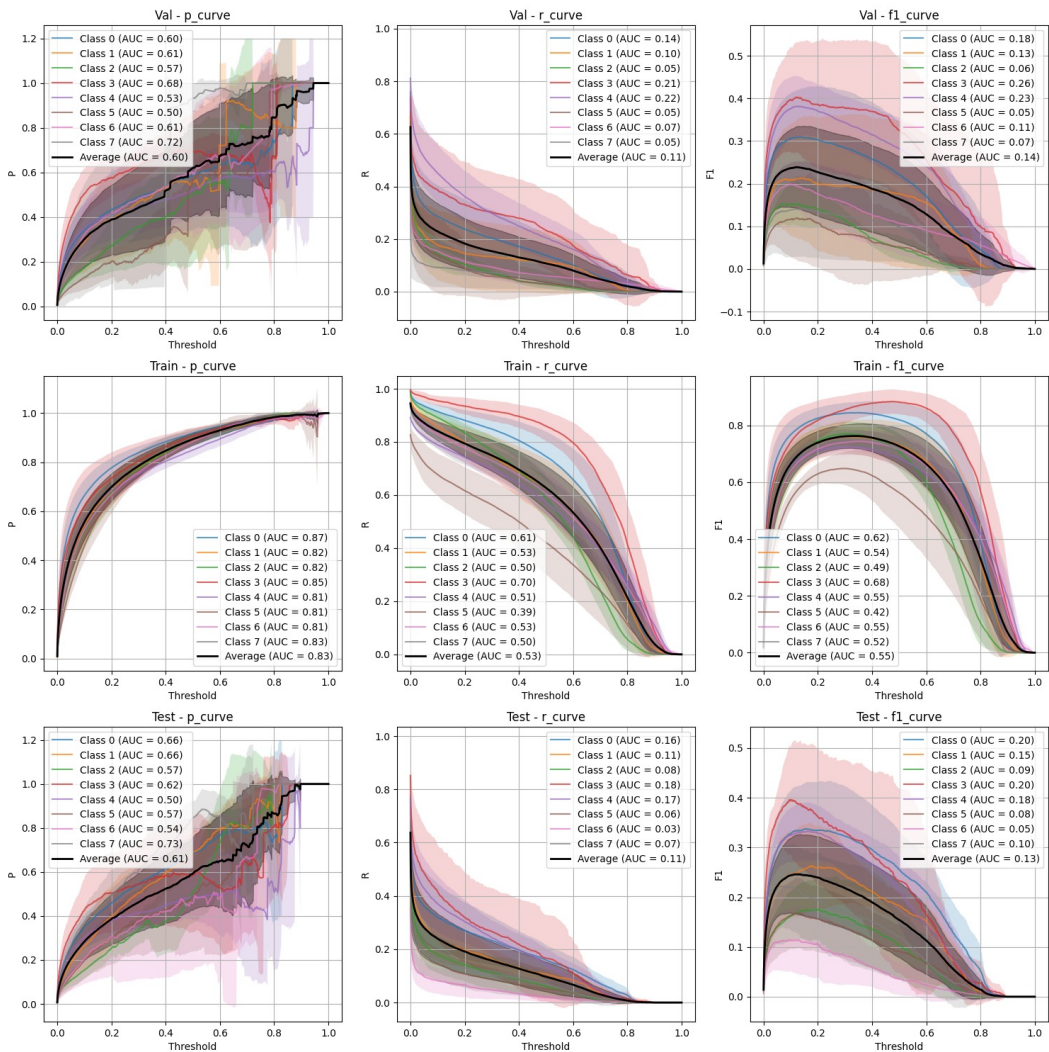


Figure 5. FasterRCNN Stratified Train/Val/Test Precision, Recall, and F-1 Curves.

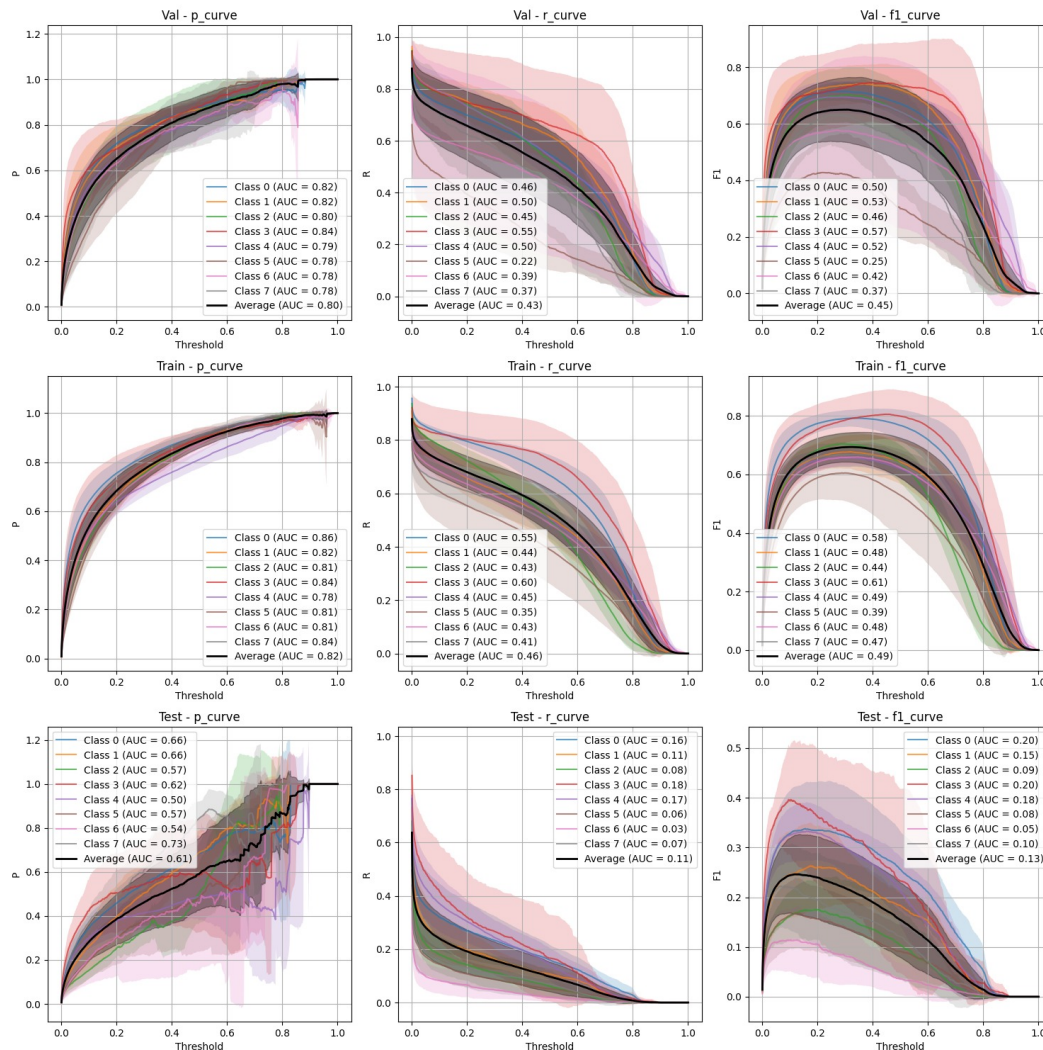


Figure 6. FasterRCNN Non-Stratified Train/Val/Test Precision, Recall, and F-1 Curves.

### YOLOv9 Performance

1. **Stratified split** (Figure 3): YOLOv9 showed lower precision and recall compared to FasterRCNN, with substantial variation in the stratified scenario. The F-1 curves demonstrate challenges in maintaining a balance between precision and recall, particularly on the validation and test sets.
2. **Non-Stratified split** (Figure 4): YOLOv9 improved its performance in the non-stratified case, as reflected in higher precision, recall, and mAP scores. The non-stratified F-1 curves also show better consistency across training and validation sets.

### FasterRCNN Performance

1. **Stratified split** (Figure 5): FasterRCNN demonstrated superior performance across all metrics in the stratified scenario, with precision exceeding 0.97 and relatively stable recall. The mAP50 and mAP50-95 scores indicate that the model was able to detect and segment objects with high accuracy.
2. **Non-Stratified split** (Figure 6): While the performance of FasterRCNN remained high in the non-stratified scenario, slight differences were observed in precision and recall across the training, validation, and test datasets. F-1 curves for FasterRCNN consistently showed better generalization and balance across all splits compared to YOLOv9.



Comparison of Test Performance Metrics

Table 3 summarizes the test performance metrics for both YOLOv9 and FasterRCNN, highlighting the key differences between stratified and non-stratified training splits. FasterRCNN consistently outperforms YOLOv9 in both setups, particularly in terms of precision and mAP scores.

**Table 3.** Test Performance Metrics with standard deviations for YOLOv9 and FasterRCNN (Stratified and Non-Stratified data splits).

Model / Split	Precision	Recall	mAP50	mAP50-95	Fitness
FasterRCNN / Stratified	0.9811±0.0084	0.7083±0.0807	0.8185±0.0562	0.7345±0.0554	0.7429±0.0555
FasterRCNN / Non-Stratified	0.9787±0.0107	0.7076±0.0957	0.8162±0.0647	0.7309±0.0612	0.7395±0.0615
YOLOv9 / Stratified	0.5504±0.1864	0.3580±0.2701	0.4599±0.2503	0.2767±0.1877	0.2951±0.1939
YOLOv9 / Non-Stratified	0.6458±0.1662	0.4742±0.2193	0.5771±0.2113	0.3622±0.1656	0.3837±0.1701

5. Discussion of Challenges and Future Directions

Both YOLOv9 and FasterRCNN faced challenges in handling class imbalance and potential overfitting, particularly in the stratified splits. YOLOv9’s performance was more sensitive to these factors, especially in terms of recall, where it missed a significant number of true positives. FasterRCNN, while more consistent, also faced challenges in detecting smaller or more complex objects. Overfitting was an issue for YOLOv9, with validation loss increasing after just a few epochs. It’s worth noting that class 8 - "Superficial White" was rather poorly represented in the training set and the model was not able to make any predictions out of it - thus it is not present on Figures 3-6. We observed data leakage in YOLOv9 model more severely than in FasterRCNN, probably due to a smaller YOLOv9 model size and lower resolution. It failed to build robust visual features and their hierarchy and thus overfitted more severely and underperformed in comparison to FasterRCNN-ResNet50 fine-tuned model.

To address all the aforementioned challenges, future work should explore data augmentation techniques as well, such as oversampling or synthetic data generation for underrepresented classes. Moreover, advanced model architectures, such as multi-target architectures using a mixture of datasets and tasks could improve performance as well as pave the way for the transformer-based approaches (due to increase of the acquired training set size), which could further enhance the generalization and performance.

6. Conclusions

This study provides evaluation of deep learning models for object detection in laparoscopic surgery, focusing on endometriosis-specific challenges. Among the tested models, FasterRCNN demonstrated best performance in both stratified and non-stratified setups, showcasing its potential as a robust tool for detecting complex anatomical structures and lesions in laparoscopic videos. YOLOv9, while competitive, exhibited variability and a higher susceptibility to overfitting, emphasizing the importance of data quality and model architecture. The challenges identified, including class imbalance and data leakage, underline the critical need for rigorous dataset preparation and advanced augmentation techniques. Future work will focus on overcoming these limitations by incorporating additional data, exploring advanced model architectures, and refining augmentation strategies. Ultimately, these efforts aim to bridge the gap between experimental results and real-world surgical applications, paving the way for AI-driven improvements in endometriosis diagnosis and treatment.

**Author Contributions:** The main conceptualization, method development, code, evaluation and paper writing was done by Andrey Bondarenko and Vilen Jumutc jointly. Resources and initial review were provided by Dmitrijs Bliznuks. Dataset was provided and curated by the authors from SurgAR, Marseille Hospital and Beneficência Portuguesa de São Paulo. Finally, the supervision, project administration and funding acquisitions were done by Dmitrijs Bliznuks and Ulrik Bak Kirk.

**Funding:** This project has received funding from The European Union’s Horizon 2020 Research and Innovation Programme "FEMaLe" under Grant agreement 101017562.



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

## References

1. Hummelshoj, L.; Prentice, A.; Groothuis, P. Update on Endometriosis: 9th World Congress on Endometriosis, 14–17 September 2005, Maastricht, the Netherlands. *Women's Health* **2006**, *2*, 53–56, [<https://doi.org/10.2217/17455057.2.1.53>]. <https://doi.org/10.2217/17455057.2.1.53>.
2. Beata, S.; Szyłło, K.; Romanowicz, H. Endometriosis: Epidemiology, Classification, Pathogenesis, Treatment and Genetics (Review of Literature). *International Journal of Molecular Sciences* **2021**, *22*, 10554. <https://doi.org/10.3390/ijms221910554>.
3. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2015**, *39*. <https://doi.org/10.1109/TPAMI.2016.2577031>.
4. He, K.; Gkioxari, G.; Dollar, P.; Girshick, R. Mask R-CNN. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2018**, *PP*, 1–1. <https://doi.org/10.1109/TPAMI.2018.2844175>.
5. Wang, C.Y.; Yeh, I.H.; Liao, H.y., YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information; 2024; pp. 1–21. [https://doi.org/10.1007/978-3-031-72751-1\\_1](https://doi.org/10.1007/978-3-031-72751-1_1).
6. Foti, P.; Palmucci, S.; Vizzini, I.; Libertini, N.; Coronella, M.; Spadola, S.; Caltabiano, R.; Iraci Sareri, M.; Basile, A.; Milone, P.; et al. Endometriosis: clinical features, MR imaging findings and pathologic correlation. *Insights into Imaging* **2018**, *9*. <https://doi.org/10.1007/s13244-017-0591-0>.
7. Sivajohan, B.; Elgendi, M.; Menon, C.; Allaire, C.; Yong, P.; Bedaiwy, M. Clinical use of artificial intelligence in endometriosis: a scoping review. *npj Digital Medicine* **2022**, *5*, 109. <https://doi.org/10.1038/s41746-022-00638-1>.
8. Nifora, C.; Chasapi, M.K.; Chasapi, L.; Koutsojannis, C. Deep Learning Improves Accuracy of Laparoscopic Imaging Classification for Endometriosis Diagnosis. *Journal of Clinical and Medical Surgery* **2024**, *4*, 1137–1145. <https://doi.org/10.52768/2833-5465/1137>.
9. Leibetseder, A.; Schoeffmann, K.; Keckstein, J.; Keckstein, S. Endometriosis detection and localization in laparoscopic gynecology. *Multimedia Tools and Applications* **2022**, *81*. <https://doi.org/10.1007/s11042-021-11730-1>.
10. Hong, W.; Kao, C.; Kuo, Y.; Wang, J.; Chang, W.; Shih, C. CholecSeg8k: A Semantic Segmentation Dataset for Laparoscopic Cholecystectomy Based on Cholec80. *CoRR* **2020**, *abs/2012.12453*, [[2012.12453](https://arxiv.org/abs/2012.12453)].
11. Carstens, M.; Rinner, F.; Bodenstedt, S.; Jenke, A.; Weitz, J.; Distler, M.; Speidel, S.; Kolbinger, F. The Dresden Surgical Anatomy Dataset for Abdominal Organ Segmentation in Surgical Data Science. *Scientific Data* **2023**, *10*. <https://doi.org/10.1038/s41597-022-01719-2>.
12. Leibetseder, A.; Kletz, S.; Schoeffmann, K.; Keckstein, S.; Keckstein, J., GLEND: Gynecologic Laparoscopy Endometriosis Dataset; 2019; pp. 439–450. [https://doi.org/10.1007/978-3-030-37734-2\\_36](https://doi.org/10.1007/978-3-030-37734-2_36).
13. Yoon, J.; Hong, S.; Hong, S.; Lee, J.; Shin, S.; Park, B.; Sung, N.; Yu, H.; Kim, S.; Park, S.; et al., Surgical Scene Segmentation Using Semantic Image Synthesis with a Virtual Surgery Environment; 2022; pp. 551–561. [https://doi.org/10.1007/978-3-031-16449-1\\_53](https://doi.org/10.1007/978-3-031-16449-1_53).
14. Fujita, H.; Itagaki, M.; Hooi, Y.K.; Ichikawa, K.; Kawano, K.; Yamamoto, R. Detector Algorithms of Bounding Box and Segmentation Mask of a Mask R-CNN Model. *ArXiv* **2020**, *abs/2010.13783*.
15. Figueiredo, R.B.D.; Mendes, H.A. Analyzing Information Leakage on Video Object Detection Datasets by Splitting Images Into Clusters With High Spatiotemporal Correlation. *IEEE Access* **2024**, *12*, 47646–47655. <https://doi.org/10.1109/ACCESS.2024.3383047>.
16. Apicella, A.; Isgrò, F.; Prevete, R. Don't Push the Button! Exploring Data Leakage Risks in Machine Learning and Transfer Learning, 2024, [[arXiv:cs.LG/2401.13796](https://arxiv.org/abs/2401.13796)].
17. Lin, T.; Maire, M.; Belongie, S.J.; Bourdev, L.D.; Girshick, R.B.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft COCO: Common Objects in Context. *CoRR* **2014**, *abs/1405.0312*, [[1405.0312](https://arxiv.org/abs/1405.0312)].
18. Ghiasi, G.; Cui, Y.; Srinivas, A.; Qian, R.; Lin, T.Y.; Cubuk, E.D.; Le, Q.V.; Zoph, B. Simple Copy-Paste is a Strong Data Augmentation Method for Instance Segmentation. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* **2020**, pp. 2917–2927.
19. Zhang, H.; Cissé, M.; Dauphin, Y.; Lopez-Paz, D. mixup: Beyond Empirical Risk Minimization. *ArXiv* **2017**, *abs/1710.09412*.
20. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *CoRR* **2014**, *abs/1412.6980*.

21. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *ArXiv* **2020**, *abs/2010.11929*.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.