CT-based Lymph Node Real-time Detection via Transfer Learning with YOLOv8

LIN RUIKAI

Supervisor: Dr. Andrew Makmur

A THESIS SUBMITTED FOR THE DEGREE OF MASTER OF SCIENCE IN BIOMEDICAL INFORMATICS

DIVISION OF BIOMEDICAL INFORMATICS

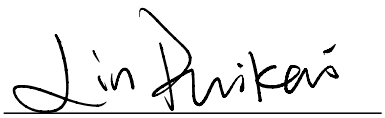
NATIONAL UNIVERSITY OF SINGAPORE

April 2023

Declaration

I hereby declare that this thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.



LIN RUIKAI

April 2023

Acknowledgments

I would like to express my deepest gratitude to my capstone project supervisor, Dr. Andrew Makmur, for his invaluable guidance and unwavering support throughout this project. His expertise and insightful feedback have been instrumental in shaping my research and methodology. I am truly grateful to him for his extensive support and helpful advice over time, and he has been an integral part of this research journey.

Additionally, I would like to extend my heartfelt thanks to the physicians from the National University Health System (NUHS) who participated in the annotation of the lymph node CT data. Without a large amount of accurately labeled lymph node images, this study would not have been able to explore lymph node detection in depth and achieve significant results. Their support has been critical in enabling me to pursue my research goals.

Moreover, I am deeply grateful to all professors from the Department of Biomedical Informatics, National University of Singapore, for their support and encouragement during my master's journey and capstone project. Their valuable feedback and insights have been instrumental in improving the quality of my work.

Finally, I would like to thank my family and friends for their unwavering love, understanding, and encouragement throughout my capstone project. Their support has been a constant source of inspiration and motivation, and I feel truly blessed to have them in my life.

Abstract

Lymph nodes are important immune organs in the human body, which can be detected and diagnosed using medical imaging technologies such as CT. Lymph node detection has significant application value in the field of medical imaging. Traditional lymph node detection methods rely on manual observation and analysis, which suffer from subjectivity, low efficiency, and high misdiagnosis rates. In recent years, with the development of deep learning technology, lymph node detection based on deep learning has gradually become mainstream. However, due to the scarcity of samples, data imbalance, and high diversity in lymph node detection tasks, the robustness and generalization ability of the model need to be improved.

This study explored methods and applications of lymph node CT image object detection. The characteristics and medical background of lymph node CT images were introduced, and a comprehensive overview of lymph node object detection and related computer vision models was provided. A lymph node detection dataset was annotated and released, and the feature expression capability of the dataset was enhanced through various preprocessing and data augmentation methods. Considering the characteristics of lymph node CT images, several famous object detection models, including YOLO, Faster-RCNN and SSD, were selected to conduct pre-experiments and comparative experiments. Finally, YOLO was chosen as the baseline model. The transfer

learning approach was adopted, and pre-trained YOLOv8n was used as the backbone. By combining with a series of optimization measures determined by a large number of ablation experiments, a powerful real-time lymph node detection model with strong generalization ability was trained. The model's mAP reached 98.2%, with a precision of 98.4% and a recall of 92.7%, providing prediction results comparable to the ground truth provided by human experts. The model was also deployed to an online platform, serving as a functional API to support various businesses. Finally, the study discussed how to integrate emerging ideas such as domain generalization, domain adaptation, label-efficient, and federated learning into lymph node target detection tasks to further improve the algorithm's performance and applicability.

In conclusion, this study provides a feasible pipeline solution for CT-based lymph node detection task, and also provides a reference for research in the field of medical object detection.

Keywords: lymph node detection, deep learning, CT image, YOLOv8, object detection

Contents

Chapter 1

1 Introduction

1.1 Research Background and Motivation

Lymph nodes are an essential component of the immune system and are distributed throughout the human body, playing crucial physiological functions [1]. In addition, lymph node stations are essential in the diagnosis, staging, and management of various malignancies, especially in the detection and assessment of lymph node involvement which are effective for determining appropriate treatment strategies and predicting patient outcomes [2]. Therefore, the accurate diagnosis of lymph node diseases is vital for human health. Computed Tomography (CT) imaging is a widely utilized medical imaging method and has become a crucial diagnostic tool for lymph node diseases. In lymph node CT image analysis, the detection of lymph node targets is a crucial task. It helps medical practitioners to swiftly and accurately identify the location and size of lymph nodes, providing significant guidance for disease diagnosis

and treatment. The supraclavicular, axillary, mediastinal, and inguinal lymph node stations are among the most important sites for evaluating lymphatic spread in cancer patients. These lymph nodes have characteristic CT imaging features that can provide important diagnostic information [3]. For instance, enlarged lymph nodes with irregular borders, heterogeneous enhancement, and central necrosis are indicative of malignancy.

However, the manual detection and assessment of lymph nodes in radiology images are time-consuming, labor-intensive, and prone to inter-observer variability. Radiologists need to meticulously evaluate multiple image slices to locate and assess the lymph nodes, which can be challenging due to their small size and variable shapes. Additionally, there are many distinct properties of lymph node CT images, including low contrast, image noise, etc. Inter-observer variability in lymph node detection can lead to inconsistencies in diagnosis and treatment planning [4].

An automated lymph node station detection system could potentially have a significant impact on cancer diagnosis and treatment. However, one potential challenge in developing an automated system is that lymph nodes can vary in size, shape, and location within the body. Therefore, the model would need to be able to generalize to a wide range of lymph node characteristics and be able to identify lymph nodes even when they are not in a typical location [3]. Another challenge could be in ensuring that the system is accurate and reliable enough

to be used in clinical settings [4-5]. This may require extensive validation and testing to ensure that the system performs consistently and does not produce false positives or false negatives. Overall, an automated lymph node station detection system has the potential to improve the efficiency and accuracy of cancer diagnosis and treatment. Therefore, this study hopes to overcome the above pain points, and develop an accurate and efficient lymph node detector.

1.2 Research Aims

The primary objective of this project is to develop an automated lymph node station detection system that can accurately identify and delineate the supraclavicular, axillary, mediastinal, and inguinal lymph node stations in radiology images. The proposed system should improve the efficiency and consistency of lymph node assessment in clinical settings, thereby reducing diagnostic errors and optimizing patient care. The system should be able to analyze CT images and detect lymph nodes based on their CT imaging features. Overall, this research aids in disease diagnosis and treatment planning.

1.3 Overview of Our Research

This study was conducted with support from the National University Health System (NUHS, Singapore). Firstly, we acquired CT from 30 patients and performed de-identification processing. Then, all CT images were marked with frames of four kinds of lymph nodes (region of interest, ROI) by experienced

radiologists, forming a new bounding-box dataset. Meanwhile, this study has done a lot of surveys, mainly focusing on medical image analysis, deep learning and computer vision, object detection, lymph node detector, etc., and formed two literature reviews. Next, we did extensive pre-experiments for model selection. In this part, the method and sequence of data preprocessing and data enhancement were selected according to the characteristics of the specific model, and ablation experiments were performed to confirm whether each link in the workflow promotes the performance of our model. After determining the backbone of the final model, we continued to improve the performance of the model through a series of optimization strategies. After long-term model training experiments, we proposed the final model workflow and pipeline, and deployed it on the client. Experimental results and evaluation results with discussion for all of the above experiments are also documented in this report. Finally, the advantages, disadvantages and limitations of this study were analyzed, and several guiding directions for further optimization in the future are proposed.

Figure 1 shows the complete research route, and the main content of each chapter in this report.

Chapter 1

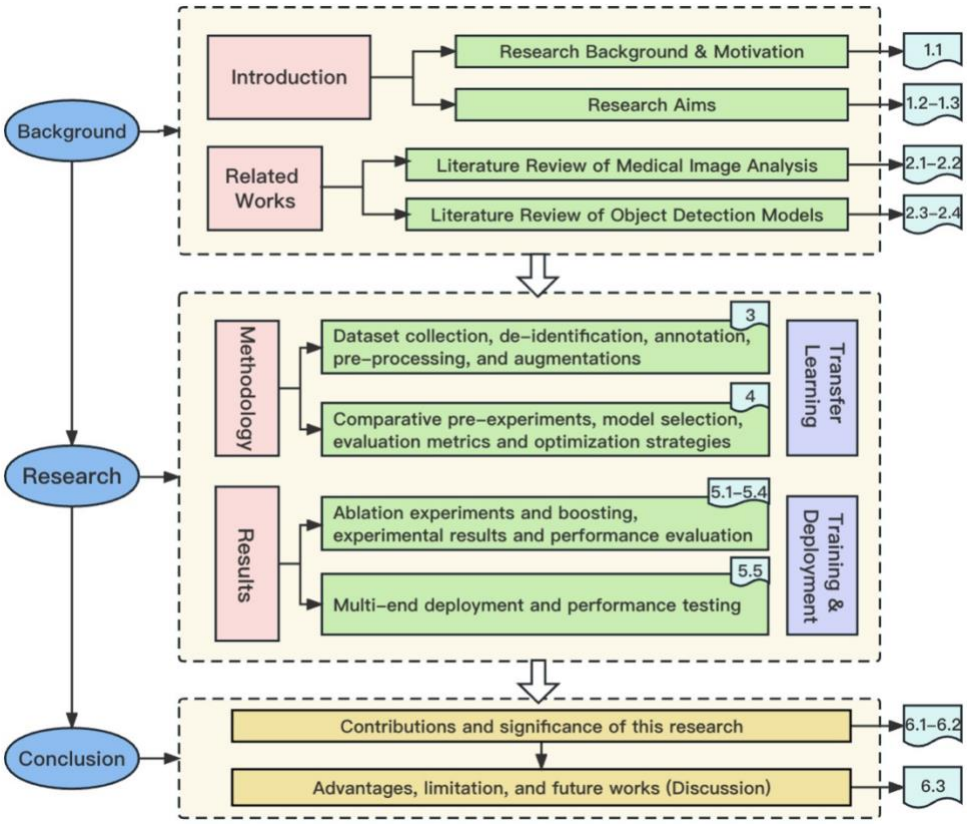


Figure 1: Research route.

Chapter 2

2 Related Works

2.1 Overview of Medical Related Background

Lymph nodes are a type of tissue in the human body and part of the lymphatic system [2]. They function to filter, store, and produce lymphocytes, and can clear waste and pathogens from the body by filtering blood and lymph fluid. Lymph nodes typically have a round or oval shape, with sizes ranging from 1 to 25 millimeters, and may vary in size and shape depending on factors such as location, age, and disease [1]. In medical diagnosis, lymph node examination is commonly performed to evaluate a patient's condition. Imaging techniques such as ultrasound, CT, or MRI may be used to examine the size, shape, and location of lymph nodes [5]. Lymph node detection is an important basis for radiologists to stage tumors and design treatment plans. The status and distribution of lymph nodes can provide critical information for tumor diagnosis and treatment, such as whether lymph node metastasis is present, the number, size, location, and

shape of lymph nodes. By analyzing lymph nodes, doctors can assess a patient's condition, develop personalized treatment plans, and predict patient prognosis [5].

CT, short for Computed Tomography, is a medical imaging technology that can acquire high-resolution structural images of the human body's internal organs. By utilizing X-rays and computer technology, CT scanners can perform multiple scans and image reconstructions in various directions, producing 3D images with high spatial resolution and contrast. CT scans can be used to guide surgical procedures and radiation therapy, improving treatment accuracy and effectiveness. Additionally, CT scans play a critical role in medical diagnosis and treatment and have become a common medical examination method. CT TAP (Computed Tomography Thoracoabdominal Pelvic) is a full-body application of CT technology that can evaluate various diseases and conditions, including the chest, abdomen, and pelvic area.

On CT imaging, lymph nodes typically appear as round or oval nodular shadows with sizes ranging from a few millimeters to several centimeters [6]. Their density is similar to that of surrounding soft tissues, but they may show different enhancement patterns on contrast-enhanced scans, such as peripheral or central enhancement. In addition, the CT image can reveal the tissue structure around lymph nodes and the relationship between lymph nodes and adjacent organs [5], which is important for diagnosing and treating lymph node diseases.

Chapter 2

DICOM, short for Digital Imaging and Communications in Medicine, is a universal storage standard for CT images and represents the field of digital imaging and medical communication. It is an international standard for medical images and related data, containing patient identity information, image acquisition parameters, and other metadata. Furthermore, DICOM defines communication protocols that enable different brands and types of medical equipment to communicate and exchange data [7]. DICOM's existence has facilitated the development of medical information technology, enabling medical images and related data to be shared and archived among different devices and medical centers.

2.2 Literature Review of Computer-aided Medical Image Analysis

In the mid-1960s, Lodwick [8] proposed a new concept of utilizing computer technology for medical image analysis, which he termed "Computer-aided Diagnosis" (CAD) in his famous paper "Computer-aided diagnosis in radiology: a research plan". However, due to various constraints such as technology and clinical concepts, CAD technology's development was extremely slow at that time. In the late 1970s and 1980s, early CAD systems appeared but belonged to the realm of "expert systems" [9], such as the well-known MYCIN expert system. It was not until the late 1980s that CAD began to be widely used in the

treatment and prognosis research of various diseases, thanks to the development of mathematics, statistics, data mining techniques, computer algorithms, and hardware computing power. Currently, the definition of "Computer-aided Diagnosis" by Wikipedia is "an interdisciplinary technology that combines artificial intelligence and computer vision elements with radiology and pathology image processing". In short, CAD is a technology that assists doctors in interpreting medical images and is a cross-disciplinary research direction between medicine and computer science.

With the rise of artificial intelligence in the 21st century, CAD has gradually spawned a new discipline or research field: medical image analysis. Scholars in this field are dedicated to using machine learning or deep learning, and even more advanced techniques to process medical images such as CT and MRI, to automatically annotate lesions in medical images, and to detect small tumors in human organs. In recent years, the emergence of computer technologies such as data mining, machine learning, deep learning, and even federated learning has greatly promoted the development of CAD. Scholars have begun to use algorithms to process medical images, and have achieved good results in clinical challenges such as breast cancer diagnosis, prostate cancer diagnosis, pathological brain detection, Alzheimer's disease, nuclear medicine etc. [10].

2.3 Overview of Object Detection in Medical Imaging

Chapter 2

Object detection is an important task in the field of computer vision [11], with the goal of automatically identifying the location and category of specific targets in images or videos. Medical target detection is an important research direction in the field of medical image processing. In recent years, with the development of deep learning and computer vision technology, research on object detection has made significant progress. This section will introduce the current state and development of object detection research.

(1) Overview of traditional approaches

The traditional approaches to object detection mainly include machine learning methods based on filters and feature extraction, segmentation-based methods, and classifier-based methods [12]. These methods require manual feature or rule design, as well as significant parameter tuning and computational resources.

(2) Overview of deep learning approaches

The deep learning-based methods have become the mainstream research direction in the field of object detection [11-12]. Currently, the main deep learning models include convolutional neural networks (CNNs), region-based convolutional neural networks (R-CNNs) [12], single-stage detectors (SSDs) [13], fast region-based convolutional neural networks (Fast R-CNNs) [14], faster region-based convolutional neural networks (Faster R-CNNs) [15], and You Only Look Once (YOLO) [16]. These models have demonstrated excellent performance in object detection tasks and have been widely applied in practical

applications. Commonly used deep learning-based object detection models can be subdivided into the following three types:

1. R-CNN series: including R-CNN, Fast R-CNN, Faster R-CNN, Mask R-CNN, etc. These methods first extract candidate regions, and then perform classification and localization on each candidate region.
2. YOLO (You Only Look Once) series: including YOLOv1 to YOLOv8, etc. These methods directly detect the whole image, and transform the detection and classification problems into regression problems.
3. SSD (Single Shot MultiBox Detector) series: including SSD, DSSD, MS-CNN, etc. This is a deep learning-based object detection algorithm, which directly detects multiple feature maps with different scales using convolutional neural networks.

(3) Challenges

At present, there are still some problems and challenges in object detection, mainly including the following aspects:

1. Inaccurate data annotation: Due to the need for manual annotation in medical image detection and the diverse shapes and complex locations of ROI, there is a certain degree of error and subjectivity in data annotation.
2. Insufficient samples: Currently available related datasets are limited, which poses some restrictions on the training and evaluation of medical image detection algorithms.
3. Poor generalization ability of models: Due to the diverse shapes and

complex locations of lesions, the generalization ability of algorithms is problematic.

4. Insufficient interpretability: The black-box nature of deep learning models makes it difficult to explain the reasons for algorithmic decisions, which poses some limitations on the clinical application of medical image detection.

2.4 Literature Review of Existing Lymph Node Detection Models

Ma et al. [17] proposed a multisource transfer learning CNN model for lymph node detection, which reduced the number of trainable parameters and prevented over-fitting by using an encoding process. Tekchandani et al. [18] proposed a fully convolutional network (FCN) based deep learning model for lymph nodes malignancy detection in CT images. Gao et al. [19] designed a faster region-based convolutional neural network (FR-CNN) for digging metastatic lymph nodes from gastric cancer. For the abundant spatial features, Peng et al. [20] employed a method named spatio-temporal context-based recurrent visual attention model (STRAM), their opinion that they inserted the Gaussian kernel based spatial attention mechanism supported them with a superior sensitivity in a low false positive rate per patient. In addition, some researchers focused on extending the convolutional neural network to three-

dimensional space to obtain higher dimensional morphological features. Zhou and Chen et al. [21] built a model that combined many-objective radiomics and 3-dimensional convolutional neural network (3D-CNN) to extract features. Zhao et al. [22] developed a cross-modal 3D neural network based on CT images and utilized prior clinical knowledge for accurate prediction of LN metastasis in clinical stage T1 lung adenocarcinoma. Debats et al [23] initial lymph node candidate regions were detected using a GentleBoost classifier trained on image features followed by a convolutional neural network (CNN) to reduce false positives. Multiple views were incorporated into the CNN training to provide 3D structure, and a sensitivity of 85% was achieved at the rate of 5–10 FP per image. Mathai et al. [24] used state-of-the-art detection neural networks to localize lymph nodes in T2 MRI scans acquired through a variety of scanners and exam protocols, and employ bounding box fusion techniques to reduce false positives (FP) and boost detection accuracy. They construct an ensemble of the best detection models to identify potential lymph node candidates for staging, obtaining a 71.75% precision and 91.96% sensitivity at 4 FP per image. Manjunatha et al [25] proposed an automated framework for LNs detection in order to obtain more accurate detection results with low false positives. It consists of two stages: candidate generation and false positive reduction. The first stage generates volumes of interest (VOI) of probable LN candidates using a modified U-Net with ResNet architecture to obtain high sensitivity but with the cost of increased false positives. The second-stage processes the obtained candidate LNs for false positive reduction using 3D

convolutional neural network (CNN) classifiers. This method yields sensitivities of 87% at 2.75 false positives per volume (FP/vol.) and 79% at 1.74 FP/vol.

Chapter 3

3 Dataset Preparation

3.1 Data Collection and De-identification

The data used in this study were collected from the National University Health System (NUHS), Singapore. Since the research took place in Singapore, we need to follow PDPA [26], which stands for Singapore's Personal Data Protection Act, a law aimed at protecting individuals' privacy. According to the law, organizations processing personal data must adhere to specific data protection guidelines. In our research, we have taken measures to comply with PDPA, such as obtaining consent from survey participants before collecting personal data, collecting only necessary data and ensuring its accuracy, protecting personal data confidentiality and security through encryption, sharing data with third parties only when necessary and taking reasonable measures to protect its confidentiality and security, and providing access, correction, and deletion of personal data upon request [27].

Chapter 3

We extracted a total of 30 chest CT scan images from 30 patients, each of which contained multiple lymph nodes, to form a lymph node detection dataset. These lymph nodes were annotated by specialist radiologists from NUHS, and the location coordinates of each lymph node were provided. The images in the dataset are of size 512x512 and each pixel value is in the range -1024 to 3071.

PDPA requires research involving personal data to be de-identified and anonymized to protect individuals' privacy [26-27]. Therefore, we take the following measures on all collected data, aiming to preserve the integrity and availability of the data while protecting individuals' privacy. Firstly, personal identifying information, such as names, addresses, and hospital information, is removed or replaced with anonymous identifiers from image data. This process is carried out by data administrators from hospitals or research institutions to ensure no personal identifying information is leaked. Secondly, de-identification is conducted to preserve data availability while protecting individuals' privacy. Common de-identification methods in medical image data include blurring or adding noise to reduce the risk of sensitive information leakage. Finally, anonymization is performed to remove all identifiable information from medical image data. Symmetric encryption or hashing methods are generally employed in medical image data to achieve the effect of anonymization. These methods irreversibly transform medical image data into a string of gibberish, thereby achieving the goal of anonymization.

Chapter 3

The results of section 3.1 are placed in a table "Patients' CT metadata after de-identification". The table contains metadata after de-identification such as patient ID, sex, age, study date, slice thickness, pixel spacing, manufacturer, etc.

3.2 Data Annotation and Format Conversion

The raw data was marked by a number of experienced experts through the V7 Darwin platform. The regions of interest in this study included the axillary, supraclavicular, mediastinal and hilar lymph nodes. Therefore, radiologists only framed these four types of lymph nodes, recorded the category and location information of each lymph node and saved them in annotation JSONs. After annotation, the images were uniformly scrutinized by a professional to ensure the accuracy and consistency of all annotations. The workflow of data labeling is shown in Figure 2(A).

The selection of the window width and window level of lymph node CT is usually adjusted according to the actual application requirements and the doctor's suggestion. Generally, the selection result will affect the display effect of the image, thereby affecting the doctor's diagnosis and judgment of the image. In this study, under the advice of professional radiologists, the window width was set to 350-400, and the window level was set to 50-70.

Chapter 3

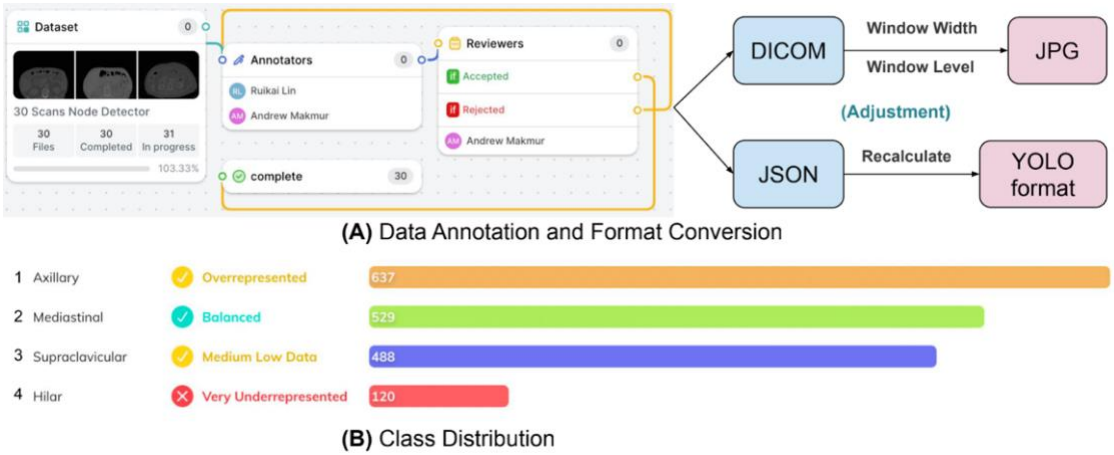


Figure 2: Workflow of data labeling, and class distribution.

As can be seen from Figure 2(B), the number of lymph nodes in the four categories is not balanced in our raw dataset. This may affect the generalization performance of the model. Therefore, we perform data augmentation on the training data, see section 3.4 for details.

After the annotation, we obtained 905 CT slices with lymph nodes and the type and location information of each lymph node. Since the original annotation file does not match the input format of the common object detection model, we further processed the original annotation file, see Figure 2(A) for details. Specifically, we converted all CT slices into JPG format image files under the premise of setting the optimal window width and window level. At the same time, we converted the type and location information in JSON format to YOLO format and stored them in txt files. Finally, we proportionally divided all training images into training set (70%, 634 total images), validation set (15%, 136 total images) and test set (15%, 135 total images). In order to let the model

learn the feature that "some input images may not contain lymph nodes" during the model training process, we also added 100 CT images that do not contain the above four types of lymph nodes and label information to the training set.

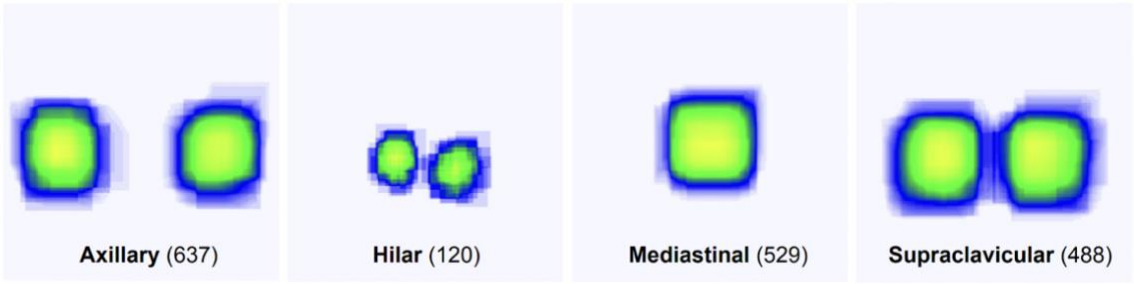


Figure 3: Annotation Heatmap (512x512 pixels).

Figure 3 shows the heatmap of the location of four kinds of lymph nodes. It is easy to see that the four kinds of lymph nodes have distinct characteristics respectively.

3.3 Data Pre-processing

In computer vision tasks, data preprocessing operations such as denoising, normalization, segmentation, and alignment are generally required to ensure data quality and consistency [28]. In the previous section, we have converted the CT from the DICOM format to the commonly used JPG format and stored it as a grayscale image; and converted the annotation file to the model-friendly YOLO format.

Next, we preprocessed and normalized all images: resize, auto-orient, etc. The

data was then normalized to scale pixel values into the same 512x512 range to improve model performance and reproducibility.

3.4 Data Augmentations

Data volume is a key factor in training machine learning models [12-16], so building a high-quality lymph node detection system can be more challenging when relatively little data is available [29]. Therefore, we employed data augmentation techniques such as rotation, flipping, cropping, etc. to augment our dataset. This can help the model better capture the different shapes, sizes and locations of lymph nodes and improve the generalization ability of the model [13]. Figure 4 shows the data augmentation workflow designed and used in this study. Figure 5 shows some typical examples of training set images after data augmentation.

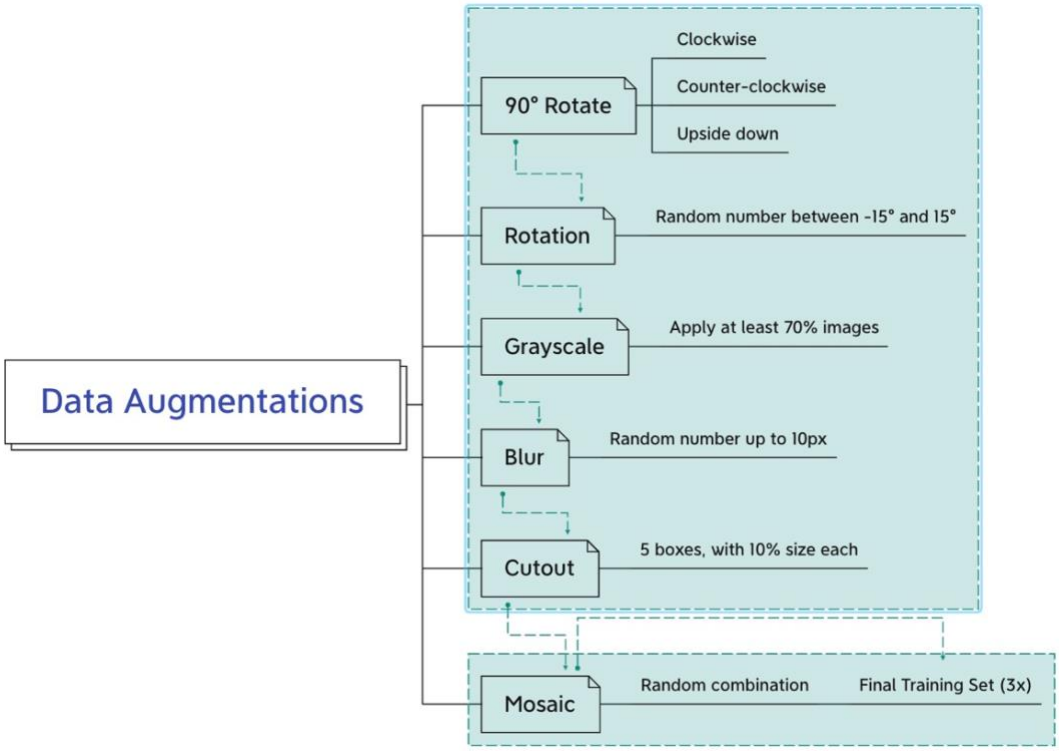


Figure 4: Workflow of data augmentations. The first five methods were randomly applied within the set threshold range, and finally mosaic processing was performed uniformly to obtain the final training set.

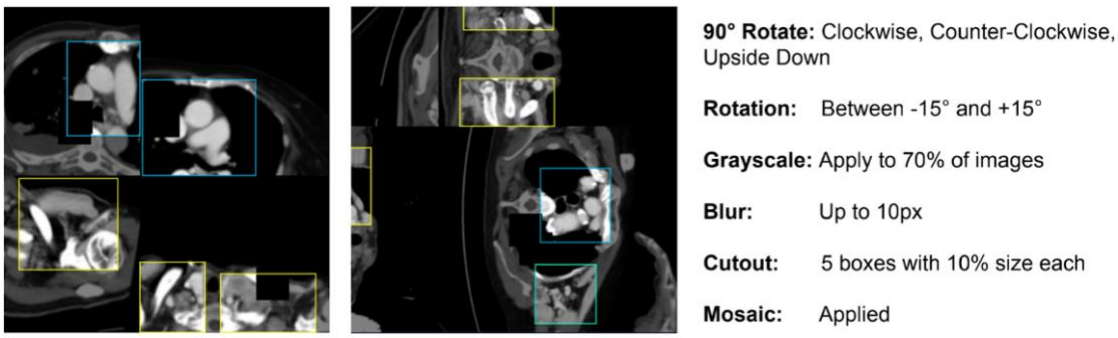


Figure 5: Cases after data augmentation transformation.

It is worth noting that we only perform data augmentations on the original training set, not the validation and test sets. This is to ensure the accuracy of the

test results. As shown in Figure 6, the final training set contains 1902 images containing a large amount of lymph node features, and the number of images is three times larger than the original one.

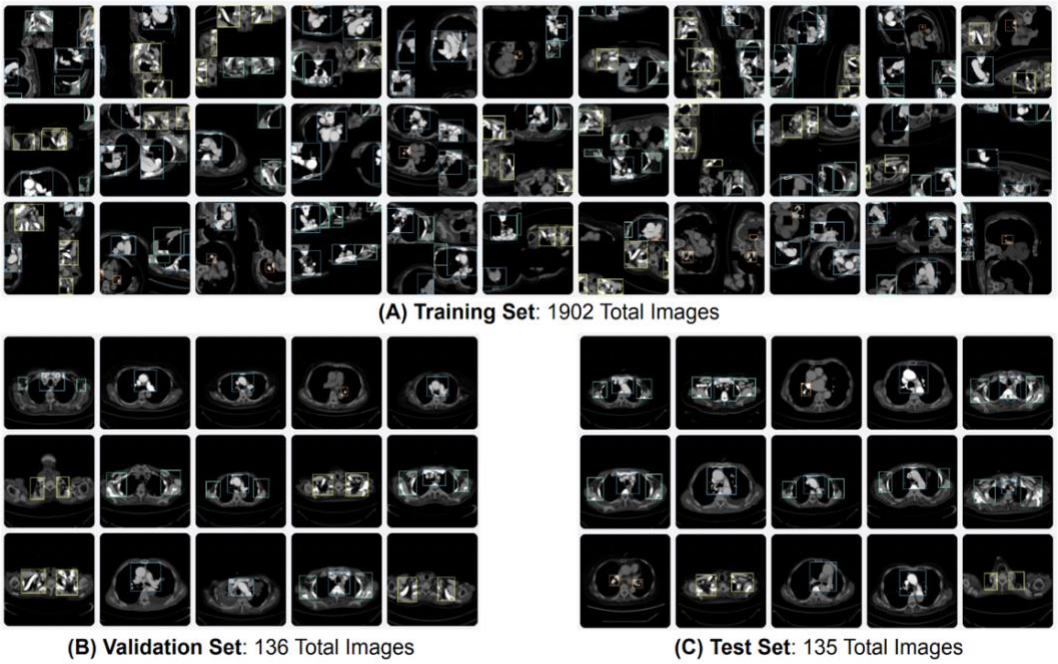


Figure 6: Training set, validation set and test set after data preprocessing and augmentation.

Chapter 4

4 Methodology of Algorithms

4.1 Data Collection and De-identification

Transfer learning is a technique to transfer the knowledge learned from one or multiple source domains to a target domain [30]. In the task of lymph node detection, transfer learning can help us improve the performance of the algorithm by utilizing existing lymph node datasets or datasets from related tasks [31].

For training object detection algorithms, a sufficiently large annotated dataset is generally required. Therefore, under the strict condition of only having 30 lymph node CT scans, in order to train a powerful and applicable lymph node detector, commonly used medical imaging datasets such as LUNA16 and NIH Chest X-ray dataset can be used to train and fine-tune these algorithms, which are then applied to the limited target dataset. These datasets may not be highly

correlated with lymph nodes, but they are commonly used medical imaging datasets that can be used to train and evaluate lymph node object detection algorithms. Specifically, these datasets can be used to train some basic feature extraction and pattern recognition abilities of the algorithm (especially for common features of medical imaging), such as shape, texture, edge, etc., which are helpful for the final lymph node object detection task. For example, models pre-trained on the ImageNet dataset, such as ResNet, VGG, Inception, can be transferred to the lymph node detection task [30]. The pre-trained model can be loaded, and only the last few layers of the model need to be fine-tuned to adapt to the lymph node detection task. These models have been trained on large-scale image datasets and have learned some common features, which can converge faster and generalize better to new datasets.

4.2 Baseline Model Selection and Architecture Design

In order to select the best baseline model, we conducted a large number of comparative experiments or architecture comparisons on many well-known target detection models, such as Faster-RCNN, SSD, RetinaNet [32], YOLO and so on. Among them, in the multi-dimensional evaluation of real-time, accuracy, and robustness, the comprehensive performance of the YOLO-based model is the best.

The You Only Look Once (YOLO) detection algorithm [16] is widely used by researchers in the field. YOLOv1 creatively proposed to directly regress the

specific location and class information of the object detection, first scaling the image to 448×448×3, dividing each image into an average of 7×7 grids, and each grid is responsible for predicting the object with its center point falling within that grid. Next, GoogLeNet is used to extract the feature map, and finally the extracted feature map is input into two fully connected layers, resulting in a 7×7×30 feature map output. This detection method greatly reduces the computational complexity compared to Faster-RCNN and significantly improves the detection speed, but with only 2 prediction boxes per grid, it selects only the prediction box with high IoU as output. In cases of dense multiple objects or small targets, it can only detect one, resulting in poor detection performance. To solve the low recall rate and inaccurate positioning of YOLOv1, YOLOv2 [33] introduced methods such as anchor boxes, high-resolution classifiers, and convolutional with anchor boxes. Multi-scale training strategy was used in training YOLOv2 to improve detection speed, and YOLOv2 was designed with only convolutional and pooling layers in the Darknet-19 network structure. YOLOv2 achieved significant improvements in identifying categories, accuracy, speed, and location accuracy. YOLOv3 [34] used residual networks to extract image features and adopted the feature pyramid structure to achieve multi-scale prediction, balancing detection accuracy and speed. YOLOv4 [35] was optimized in various aspects such as data processing, backbone network, network training, activation function, and loss function, making it have a lower training threshold and achieve better detection results under limited GPU resources. YOLOv5 [35-36] made minor

changes to YOLOv4, expanding the CSP structure in the backbone structure to the Neck structure and using SFP instead of SPP structure. The biggest feature of YOLOv5 is its memory footprint of only 27MB (compared to 244MB of YOLOv4), but its detection performance for small objects is very poor. YOLOv6 [37] replaced the backbone with the more efficient EfficientRep, and the Neck is built based on Rep and PAN, while the Head is similar to YOLOX [38]. YOLOv7 [39] algorithm adopts strategies such as extended efficient long-range attention network (E-ELAN), model scaling for concatenated-based models, and convolutional reparameterization to achieve a good balance between detection efficiency and accuracy. YOLOv8 [40] is the latest algorithm in the YOLO series, with faster detection accuracy and speed. It is an improvement on the YOLOv5 version, replacing the C3 module in the backbone section of YOLOv5 with the C2f module to achieve further lightweighting. In the neck section, the convolutional structure in the upsampling stage of PAN-FPN in YOLOv5 is removed, and in the head section, decoupled-head is adopted. In the label assign section, anchor-free thinking is adopted instead of anchor-based thinking.

Therefore, we chose YOLOv8 as the baseline model. In order to realize the idea of transfer learning, we poured in the pre-training weight of YOLOv8n as the weight basis. This pre-trained model has been trained on large-scale data sets such as COCO and ImageNet, and has strong processing capabilities for basic graphic features such as shape, texture, and edge. The model contains 225

layers, 3157200 parameters, 3157184 gradients and 8.9 GFLOPs. The optimizer was SGD (lr=0.01) with parameter groups 57 weight(decay=0.0), 64 weight(decay=0.0005), and 63 bias. Our model backbone is shown in Table 1.

Table 1: Backbone of our pre-trained YOLOv8 model.

	Params	**Module**	**Arguments**
1	464	Conv	[3, 16, 3, 2]
2	4672	Conv	[16, 32, 3, 2]
3	7360	C2f	[32, 32, 1, True]
4	18560	Conv	[32, 64, 3, 2]
5	49664	C2f	[64, 64, 2, True]
6	73984	Conv	[64, 128, 3, 2]
7	197632	C2f	[128, 128, 2, True]
8	295424	Conv	[128, 256, 3, 2]
9	460288	C2f	[256, 256, 1, True]
10	164608	SPPF	[256, 256, 5]
11	0	upsampling.Upsample	[None, 2, 'nearest']
12	0	Concat	[1]
13	148224	C2f	[384, 128, 1]
14	0	upsampling.Upsample	[None, 2, 'nearest']
15	0	Concat	[1]
16	37248	C2f	[192, 64, 1]
17	36992	Conv	[64, 64, 3, 2]
18	0	Concat	[1]

19	123648	C2f	[192, 128, 1]
20	147712	Conv	[128, 128, 3, 2]
21	0	Concat	[1]
22	493056	C2f	[384, 256, 1]
23	897664	Detect	[80, [64, 128, 256]]

4.3 Evaluation Metrics

In order to better measure the performance of the algorithm, Precision, Recall, F1-score, IoU and mAP are used as evaluation metrics for lymph node detection in this study. As shown in Equation 4-1 for Precision, it indicates the number of true lymph nodes as a percentage of the total number of detections. Equation 4-2 shows Recall, which indicates the proportion of the number of correctly detected lymph nodes to the total number of lymph nodes in the test results. Equation 4-3 shows the F1-score, which combines Precision and Recall, and represents the summed average of Precision and Recall, with higher F1-score indicating more effective detection performance. The higher the F1-score, the more effective the detection performance is. TP indicates positive samples with positive prediction, TN indicates negative samples with negative prediction, FP indicates negative samples with positive prediction, and FN marks positive samples with negative prediction.

$$precision = \frac{TP}{TP + FP} \qquad (4-1)$$

$$Recall = \frac{TP}{TP + FN} \qquad (4-2)$$

$$F1\_score = \frac{2(Precision \times Recall)}{Precision + Recall} \quad (4-3)$$

The PR curve is also an important index used to measure the detection performance, its horizontal coordinate is Recall and vertical coordinate is Precision, by dynamically updating the coordinate points corresponding to Precision and Recall at different thresholds, a smooth curve is used to connect all the points to form the PR curve, when the curve is closer to the upper right, it means that its network performance is better. The area is AP, and AP indicates the recognition accuracy for a category, as shown in Equation 4-4. mAP, on the other hand, is expressed as the average of AP of several different categories, as shown in Equation 4-5, and m indicates the total number of categories.

$$AP = \int_0^1 p(r)dr \quad (4-4)$$

$$mAP = \frac{1}{m}\sum_1^m AP_i \quad (4-5)$$

IoU denotes the intersection of the detection frame and the real annotation IoU denotes the ratio of the intersection of the detection frame and the real annotation frame region to the concurrent set [40], as shown in Equation 4-6, is used to measure the accuracy of the location of the detection frame, the larger the IoU, the more accurate the result of the target detection. Where Y denotes the formal labeled box region and Y' denotes the detection box region. The ratio of intersection and concatenation of frame regions, as shown in Equation 4-6, is used to measure the positional accuracy of the detection frame, and the larger the IoU, the more accurate the result of target detection. Where Y denotes the

formal labeled box region and Y' denotes the detection box region.

$$IoU = \frac{Y \cap Y'}{Y \cup Y'} \qquad (4-6)$$

4.4 Model Optimization

If the performance of the model on the test set is unsatisfactory, the data augmentation techniques, model architecture, hyperparameter combinations, and other aspects can be further optimized and improved, and ablation experiments can be used to determine whether certain tricks are helpful for improving model performance [35,38]. By iteratively upgrading the model, a lymph node detection model with strong generalization ability and robustness can be ultimately obtained. In this study, several optimization techniques, such as learning rate decay, regularization, and data augmentation, were employed to further improve model performance [34-35]. Finally, the model was extensively trained and fine-tuned to improve its performance and generalization ability.

Chapter 5

5 Results

5.1 Experiment Environment

The GPU used in the experiment is NVIDIA A40 (GPU Memory: 48 GB GDDR6), the CPU is Intel Xeon Gold 6226R CPU @2.90GHz, the CUDA version is 11.4, the main deep learning framework is PyTorch 1.90, and the Python version is python3.9. The operating system is Ubuntu 22.04.

5.2 Ablation Experiments

The ablation experiment is an experiment that removes certain components or operations in the model to evaluate the impact of these components or operations on the performance of the model [41]. When training a YOLO model on a custom dataset, ablation experiments can be used to improve the training performance.

Chapter 5

In addition to the data augmentations already mentioned in section 3.4, we also conducted a number of ablation experiments from the following three perspectives. We hope that by systematically varying these components of the YOLO model and evaluating their impact on the performance, it is possible to identify the optimal combination of techniques and parameters for our lymph nodes dataset, and improve the training performance of the YOLO model.

(1) Network architecture: Modify the network architecture by changing the number of layers, the size of the feature maps, or the number of filters in each layer [35]. Evaluate the impact of each modification on the model's performance.

(2) Hyperparameters: Adjust the hyperparameters such as the learning rate, momentum, weight decay, and batch size. Evaluate the impact of each modification on the model's performance.

(3) Post-processing: Modify the post-processing algorithm by adjusting the confidence threshold or the non-maximum suppression (NMS) threshold. Evaluate the impact of each modification on the model's performance.

5.3 Our Model

The final model was trained using the best model backbone and hyperparameter combination determined through ablation experiments with the pre-trained

YOLOv8n. The mAP-epoch curve is shown in Figure 7. It can be seen intuitively that the training effect of the model is very good, and the mAP reaches 98.2% when the training round is 100 rounds. No overfitting and other abnormalities occurred. In addition, Figure 7 also shows the application of transfer learning ideas to make the model fit faster.

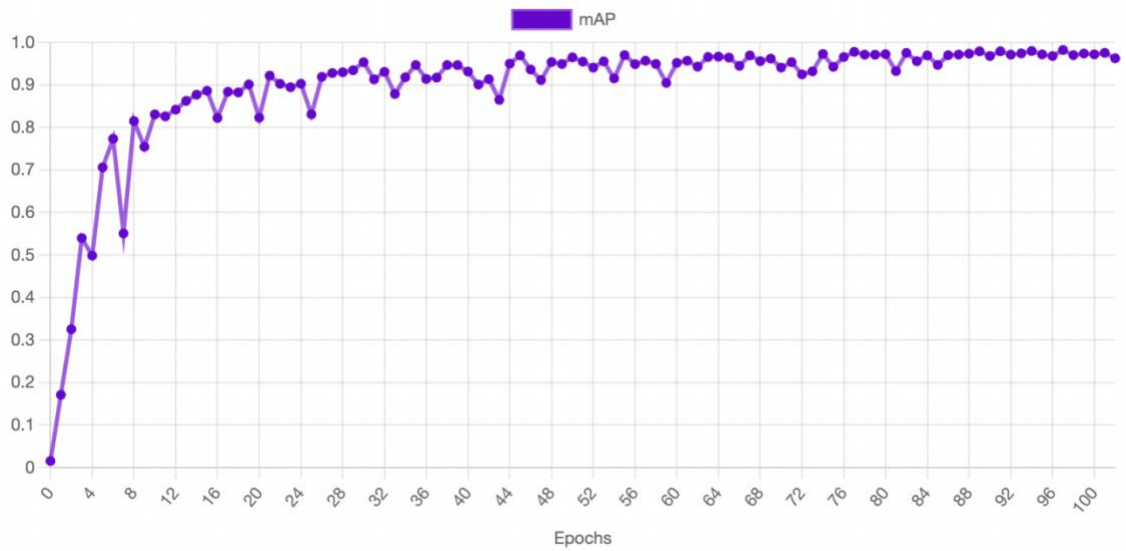


Figure 7: mAP-epoch curve of our model. This final model reached a mAP of 98.2 when it was trained for about 100 epochs.

Our final model has a strong generalization ability after training, and its precision equals 98.4%, and recall equals 92.7%. As shown in Figure 8, our model shows very good performance on all metrics.

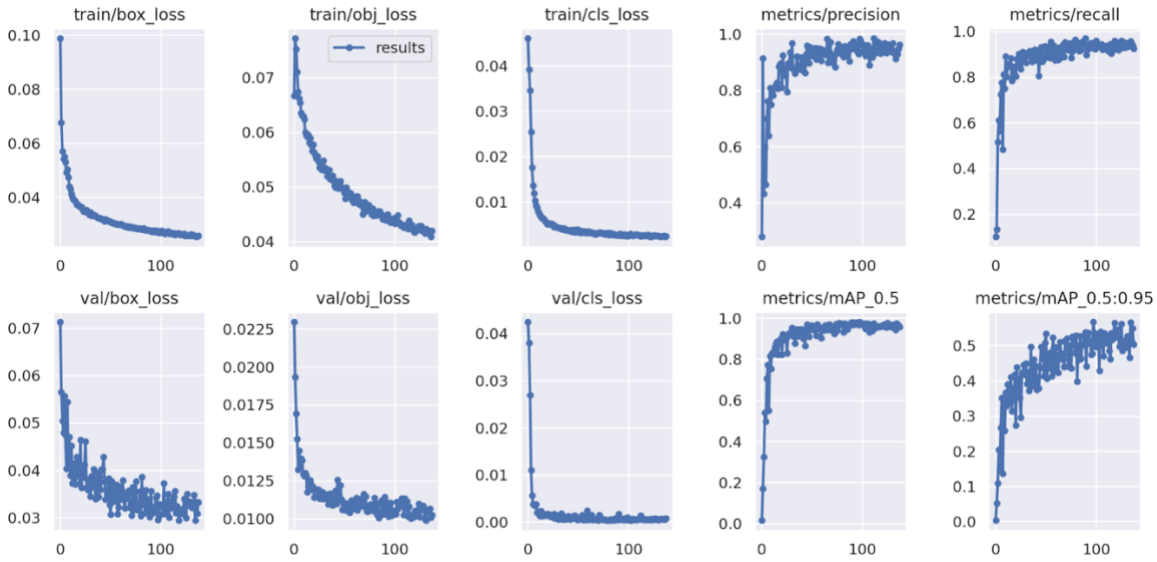


Figure 8: Box loss, object loss, class loss of training and validation sets from our final model training process; and the precision curve, recall curve, mAP_0.5 curve, and mAP_0.5:0.95 curve.

Table 2 shows the average precision values by class. These data calculated on the test set show that our model has a very strong generalization ability and can be applied in real-world clinical scenarios.

Table 2: Average Precision by Class.

	Axillary	Hilar	Mediastinal	Supraclavicular	**All**
Validation Set	97%	99%	100%	98%	**98%**
Test set	95%	92%	99%	95%	**95%**

Figure 9 shows the effect of using our model to label some test set CT images for lymph nodes. The left and right sides are the ground truth and model

prediction, representing human experts and our model, respectively. This once again illustrates the strong performance and practical value of the model we proposed.

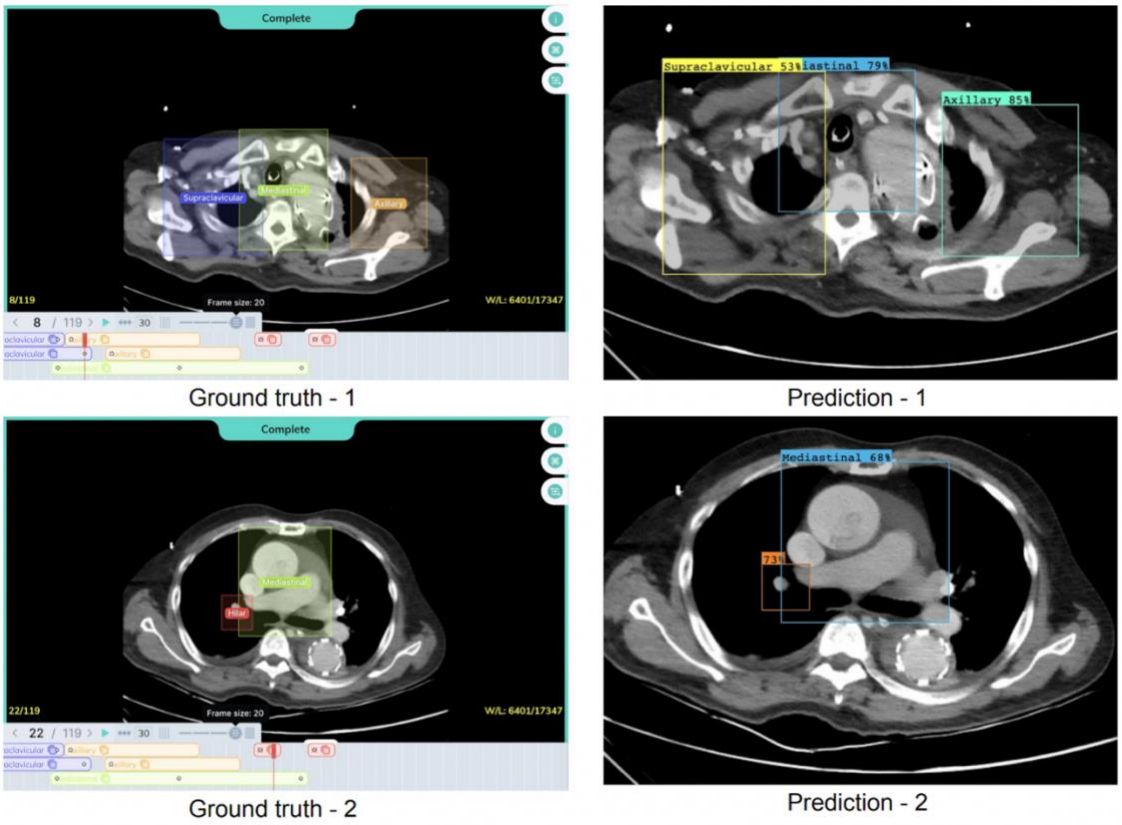


Figure 9: The performance of our model on the external test set. The left side is the ground truth (human expert level), and the right side is the prediction result of the model.

5.4 Interpretation and Analysis

Although various metrics and use tests have shown near-perfect results, we still found and summarized some existing problems.

Firstly, in Figure 8, mAP_0.5:0.95 means average mAP over different IoU thresholds, from 0.5 to 0.95, step 0.05 (0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95). From the mAP_0.5 curve, it can be found that the model can detect and frame lymph nodes in almost all correct positions. But from the mAP_0.5:0.95 curve, the position of the prediction box is not particularly accurate. This shows that the model still has room for improvement. After analysis, we found that this problem may be caused by the low quality of the annotation of the dataset itself (i.e., the location of the data was not recorded along the edge of the lymph nodes when the data was annotated). This is also a future optimization direction proposed in this paper, which will be further explained in section 6.3.

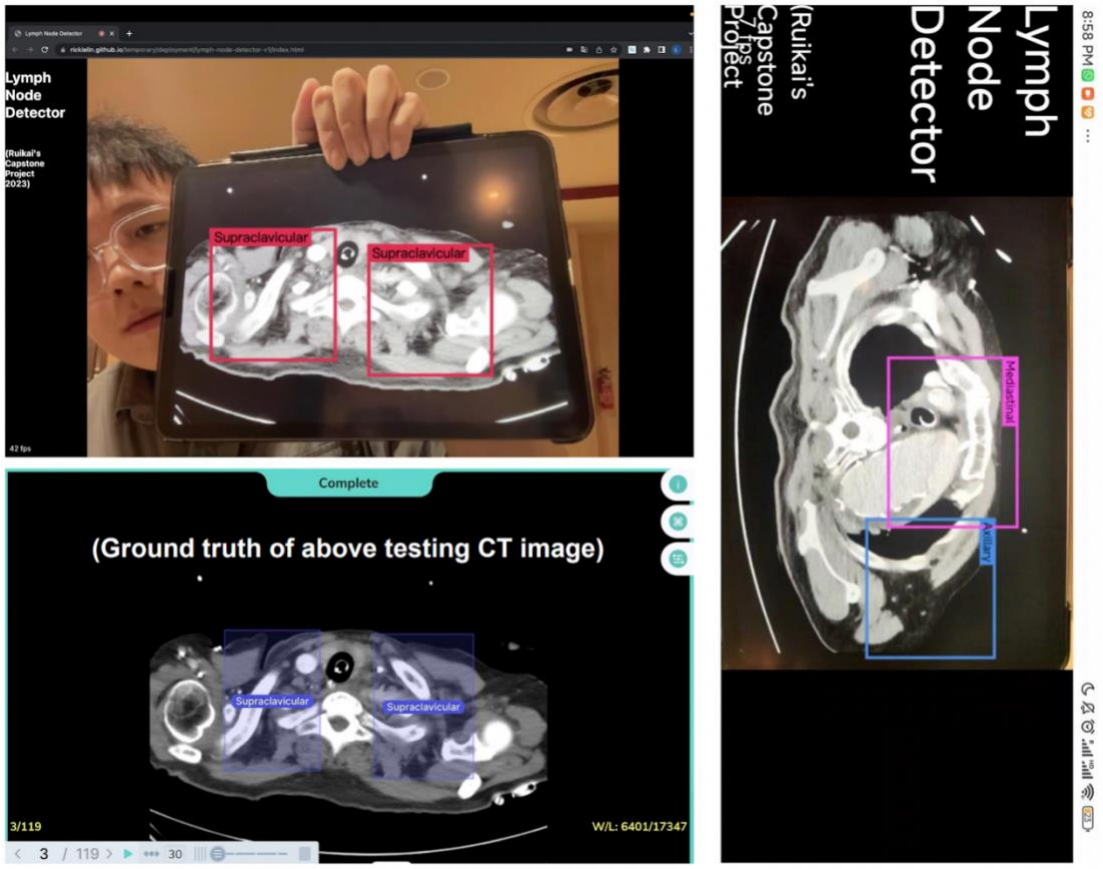
Secondly, our model has high requirements on the quality of input CT images. In other words, only when the input CT image is 512x512 pixels and the settings of the optimal window width and window level can the model show the best performance. This suggests that we should add strict preprocessing operations to the final pipeline, such as adjusting the new input image to an appropriate size, window width and window level, so that the model can detect it more stably.

5.5 Model Deployment and Testing

In addition to the most stable direct input of CT images for the model to detect lymph nodes, this study also tried real-time video detection of lymph nodes to

reflect the strong robustness and high stability of our model, and to explore the feasibility of practical scenarios such as surgical video analysis and real-time medical image analysis in XR glasses.

We encapsulated our model algorithm, and then deployed it on the webpage (https://rickielin.github.io/temporary/deployment/lymph-node-detector-v1/index.html). The webpage interface and usage effect are shown in Figure 10.



(A) PC client test with GT image **(B)** Mobile client test

Figure 10: Software interface (GUI) and usage examples. (A) is a PC client, (B) is a mobile client. The running effect may be affected by factors such as

ambient light.

Figure 11 shows the software architecture of our Lymph Node Detector. It can be used as an API to connect with related foreground businesses. It is worth mentioning that we found that when the model calls the real-time video of the camera as input, it will be greatly disturbed by the ambient light.

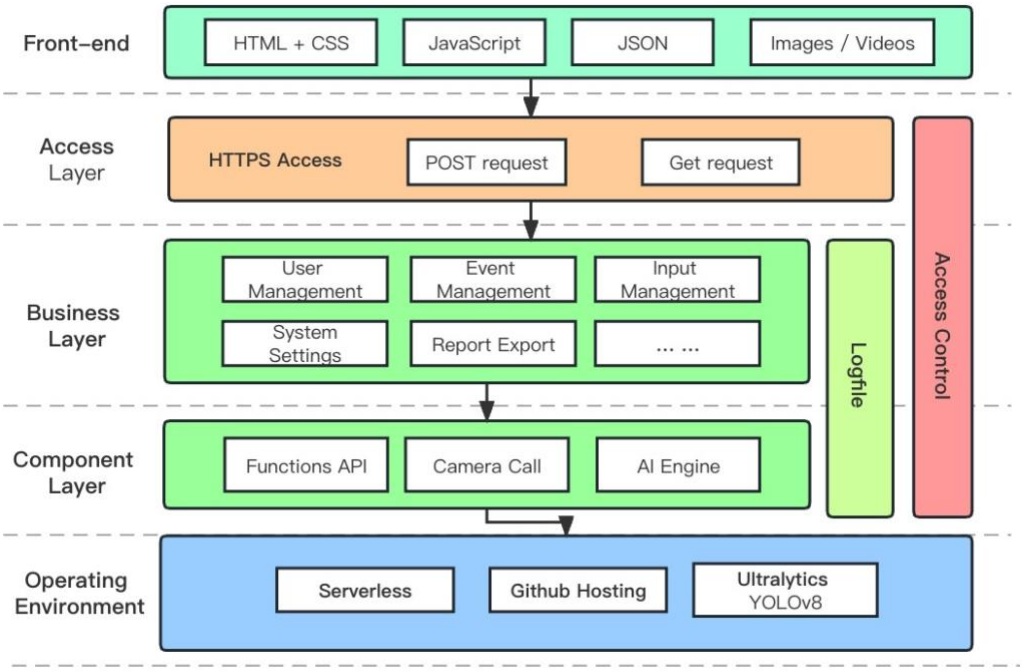


Figure 11: Software architecture of the Lymph Node Detector.

Chapter 6

6 Discussion and Conclusion

6.1 Summary of Research Contributions

1. Collect and organize existing research papers on medical image analysis and lymph node object detection, and conduct a comprehensive review and analysis of existing research results and methods.
2. An object detection dataset for lymph nodes was labeled and released with the help of radiologists. The dataset has been converted to YOLO format for a range of object detection models.
3. We explored and implemented 7 kinds of data augmentations with excellent effects, which made the original training set 3 times larger. Meanwhile, the feature expression ability of the training set was greatly increased through mosaic operation, which had a great positive effect on model training.
4. A model for lymph node object detection is proposed, which is trained

based on the idea of transfer learning. The model can perform real-time object detection for four types of lymph nodes (axillary, hilar, mediastinal, and supraclavicular) in CT images, and the experimental results prove that it has strong robustness and generalization ability.

5. Use metrics such as precision, recall, and mAP to evaluate model performance from multiple perspectives, and compare the model prediction results with the ground truth provided by human experts. In addition, the effectiveness and superiority of our algorithm are proved by comparing with other existing algorithms.
6. We released an online lymph node object detection system that will be able to automatically detect and mark the location of lymph nodes in CT images. The system is real-time, accurate and reliable, and will be able to improve the accuracy and speed of diagnosis.

6.2 Significance of Our Research

Our research is of high value and significance in real-world medical scenarios, as detailed below.

1. Improving the Accuracy and Efficiency of Diagnosis and Treatment: Traditional lymph node detection methods rely on manual judgment and measurement by physicians, which are subject to human error and subjectivity. However, an automated lymph node detection system can reduce the burden on physicians and improve the accuracy and efficiency of

diagnosis and treatment.

2. Enhancing Tumor Evaluation and Treatment Planning: Lymph nodes are one of the main pathways for tumor metastasis, and the detection and evaluation of lymph nodes are of great significance for the diagnosis and treatment planning of tumors. An automated lymph node detection system can more accurately detect and evaluate lymph nodes, thus improving tumor evaluation and treatment planning.

In conclusion, the automated lymph node detection model and system has very important value, which can improve the efficiency and quality of medical treatment, improve the evaluation of tumors and the formulation of treatment plans, and also contribute to the progress and development of medical research.

6.3 Limitation and Future Works

This study conducted a comprehensive investigation and review of the existing related research, and proposed a powerful lymph node detection model on this basis. However, during the research process, we also gradually discovered some limitations and possible improvement directions. The details are as follows.

1. Insufficient High-Quality Data: In this study, the labeled boxes in the training set only roughly delineate the location of the lymph nodes, without drawing boxes along the edges of each lymph node. This can have a significant impact on the accuracy and stability of the algorithm. If the

labeled boxes are not drawn along the edges of the lymph nodes, it may be difficult for the algorithm to accurately detect and locate the lymph nodes. In addition, if the labeled boxes are too large or too small, this can also affect the performance of the algorithm. Therefore, before using these labeled data for training, it is necessary to review and correct the labeling to ensure its accuracy and consistency. If there is not enough accurately labeled data, semi-supervised or unsupervised methods can be used to increase the available data and improve the algorithm's performance. Semi-supervised learning utilizes unlabeled data to enhance the training of supervised learning models. Specifically, a pre-trained model is used to generate predictions for unlabeled data, and these prediction results are treated as labels to train a new model. This approach can to some extent address the problem of insufficient data and improve the model's performance. In addition, as the lymph node dataset is derived from only 30 CT scans, future collaborations with other medical institutions or data providers can be considered to increase the scale and diversity of the dataset.

2. To further improve performance, domain adaptation methods [42] in transfer learning can be used, such as Domain-Adversarial Neural Networks (DANN) or Adversarial Discriminative Domain Adaptation (ADDA). These methods establish mapping relationships between source and target domains to reduce domain differences. During model training, the idea of domain generalization can be employed to mix medical image datasets from different sources, forming a larger and more diverse dataset to train deep

learning models with improved generalization performance. Additionally, to enhance the adaptability to the target domain, domain adaptation methods can be used to fine-tune the model in the target domain, further improving its performance.

3. The explainability of the proposed deep learning model is limited [43]. In medical image analysis, model explainability is crucial as doctors need to understand how the model makes its decisions. To enhance the model's explainability, specific explainability methods such as visualization techniques and attention mechanisms should be adopted to improve its credibility and reliability.
4. Multimodal data can be combined for lymph node detection in the future. In addition to CT images, other modalities such as MRI images can be used for lymph node detection. Exploring fusion methods for multimodal data [44] can improve the accuracy and robustness of lymph node detection.
5. Clinical information can also be integrated into lymph node detection. In addition to image data, clinical information is an important reference factor in lymph node detection. Researchers can explore integrating clinical information into deep learning models to improve the accuracy and practicality of the entire lymph node detection system.
6. Consider introducing federated learning [45] algorithms, allowing different medical institutions to share models and achieve better model performance while protecting data privacy and security. Using federated learning can also solve the problem of insufficient data collection. In medical image analysis,

many diseases are rare, so data from certain specific cases may be very valuable, but insufficient data collection prevents comprehensive research. Through federated learning, data from different regions and institutions can be merged, increasing data sample diversity and quantity to improve the model's generalization ability and accuracy.

Looking to the future, we will continue to explore optimization methods for lymph node target detection, constantly improve the model's performance and generalization ability, and expand our research scope to a wider range of medical image analysis fields, including medical image diagnosis and medical decision support. At the same time, we will focus on more practical application scenarios, exploring how to better translate research results into practical applications, and contribute to the development of the medical field.

References

[1] Krishnamurty, A.T., & Turley, S.J. (2020). Lymph node stromal cells: cartographers of the immune system. *Nature Immunology, 21*, 369-380.
[2] Pereira, E.R., Kedrin, D., Seano, G., Gautier, O., Meijer, E.F., Jones, D., Chin, S.M., Kitahara, S., Bouta, E.M., Chang, J., Beech, E., Jeong, H., Carroll, M.C., Taghian, A.G., & Padera, T.P. (2018). Lymph node metastases can invade local blood vessels, exit the node, and colonize distant organs in mice. *Science, 359*, 1403 - 1407.
[3] Uthaman, S., Kim, H.S., Revuri, V., Min, J., Lee, Y., Huh, K.M., & Park, I. (2018). Green synthesis of bioactive polysaccharide-capped gold nanoparticles for lymph node CT imaging. *Carbohydrate polymers, 181*, 27-33 .
[4] Weisman, A.J., Kieler, M.W., Perlman, S.B., Hutchings, M., Jeraj, R., Kostakoglu, L., & Bradshaw, T.J. (2020). Convolutional Neural Networks for Automated PET/CT Detection of Diseased Lymph Node Burden in Patients with Lymphoma. *Radiology. Artificial intelligence, 2 5*, e200016 .
[5] Liu, J., Wang, Z., Shao, H., Qu, D., Liu, J., & Yao, L. (2017). Improving CT detection sensitivity for nodal metastases in oesophageal cancer with combination of smaller size and lymph node axial ratio. *European Radiology, 28*, 188-195.
[6] Rufini, V., Garganese, G., Ieria, F.P., Pasciuto, T., Fragomeni, S.M., Gui, B., Florit, A., Inzani, F., Zannoni, G.F., Scambia, G., Giordano, A., & Collarino, A. (2021). Diagnostic performance of preoperative [18F]FDG-PET/CT for lymph node staging in vulvar cancer: a large single-centre study. *European Journal of Nuclear Medicine and Molecular Imaging, 48*, 3303 - 3314.
[7] Rutherford, M.W., Mun, S.K., Levine, B.A., Bennett, W.C., Smith, K., Farmer, P., Jarosz, Q., Wagner, U., Freyman, J., Blake, G., Tarbox, L., Farahani, K., & Prior, F.W. (2021). A DICOM dataset for evaluation of medical image de-identification. *Scientific Data, 8*.
[8] Lodwick, G.S. (1966). Computer-aided diagnosis in radiology. A research plan. *Investigative radiology, 1 1*, 72-80 .
[9] Lauritzen, S.L., & Spiegelhalter, D.J. (1990). Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the royal statistical society series b-methodological, 50*, 415-448.
[10] Litjens, G.J., Kooi, T., Bejnordi, B.E., Setio, A.A., Ciompi, F., Ghafoorian, M., Laak, J.V., Ginneken, B.V., & Sánchez, C.I. (2017). A survey on deep learning in medical image analysis. *Medical image analysis, 42*, 60-88 .
[11] Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., & Zagoruyko, S. (2020). End-to-End Object Detection with Transformers. *ArXiv, abs/2005.12872*.
[12] Dai, J., Li, Y., He, K., & Sun, J. (2016). R-FCN: Object Detection via Region-based Fully Convolutional Networks. *ArXiv, abs/1605.06409*.

References

[13] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S.E., Fu, C., & Berg, A.C. (2015). SSD: Single Shot MultiBox Detector. *European Conference on Computer Vision*.

[14] Girshick, R.B. (2015). Fast R-CNN. *2015 IEEE International Conference on Computer Vision (ICCV)*, 1440-1448.

[15] Ren, S., He, K., Girshick, R.B., & Sun, J. (2015). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 39*, 1137-1149.

[16] Redmon, J., Divvala, S.K., Girshick, R.B., & Farhadi, A. (2015). You Only Look Once: Unified, Real-Time Object Detection. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 779-788.

[17] Ma, Y., & Peng, Y. (2020). Lymph node detection method based on multisource transfer learning and convolutional neural network. *International Journal of Imaging Systems and Technology, 30*, 298 - 310.

[18] Tekchandani, H., Verma, S., & Londhe, N.D. (2020). Mediastinal lymph node malignancy detection in computed tomography images using fully convolutional network. *Biocybernetics and Biomedical Engineering, 40*, 187-199.

[19] Gao, Y., Zhang, Z., Li, S., Guo, Y., Wu, Q., Liu, S., Yang, S., Ding, L., Zhao, B., Li, S., & Lu, Y. (2019). Deep neural network-assisted computed tomography diagnosis of metastatic lymph nodes from gastric cancer. *Chinese Medical Journal*.

[20] Peng, H., & Peng, Y. (2020). Spatio-temporal context based recurrent visual attention model for lymph node detection. *International Journal of Imaging Systems and Technology, 30*, 1220 - 1242.

[21] Chen, L., Zhou, Z., Sher, D.J., Zhang, Q., Shah, J.L., Pham, N., Jiang, S.B., & Wang, J. (2018). Combining many-objective radiomics and 3D convolutional neural network through evidential reasoning to predict lymph node metastasis in head and neck cancer. *Physics in Medicine & Biology, 64*.

[22] Zhao, X., Wang, X., Xia, W., Li, Q., Zhou, L., Li, Q., Zhang, R., Cai, J., Jian, J., Fan, L., Wang, W., Bai, H., Li, Z., Xiao, Y., Tang, Y., Gao, X., & Liu, S. (2020). A cross-modal 3D deep learning for accurate lymph node metastasis prediction in clinical stage T1 lung adenocarcinoma. *Lung cancer, 145*, 10-17 .

[23] Debats, O.A., Litjens, G.J., & Huisman, H.J. (2019). Lymph node detection in MR Lymphography: false positive reduction using multi-view convolutional neural networks. *PeerJ, 7*.

[24] Mathai, T., Lee, S., Elton, D.C., Shen, T.C., Peng, Y., Lu, Z., & Summers, R.M. (2021). Detection of Lymph Nodes in T2 MRI Using Neural Network Ensembles. *MLMI@MICCAI*.

[25] Manjunatha, Y., Sharma, V., Iwahori, Y., Bhuyan, M.K., Wang, A., Ouchi, A., & Shimizu, Y. (2023). Lymph node detection in CT scans using modified U-Net with residual learning and 3D deep network. *International Journal of Computer Assisted Radiology and Surgery, 18*, 723 - 732

[26] Ong, E. (2019). Singapore Report: Data Protection in the Internet.

References

[27] Greenleaf, G. (2016). Singapore Starts Privacy Enforcement: Fines for Lax Security.

[28] Lin, R., Ma, J., Yu, H., & Zhang, Y. (2021). Accurate recognition method of plant leaves based on multi-feature fusion. *Other Conferences*.

[29] Papakipos, Z., & Bitton, J. (2022). AugLy: Data Augmentations for Robustness. *ArXiv, abs/2201.06494*.

[30] Pan, S.J., & Yang, Q. (2010). A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering, 22*, 1345-1359.

[31] Shin, H., Roth, H.R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D.J., & Summers, R.M. (2016). Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. *IEEE Transactions on Medical Imaging, 35*, 1285-1298.

[32] Tan, L., Huangfu, T., Wu, L., & Chen, W. (2021). Comparison of RetinaNet, SSD, and YOLO v3 for real-time pill identification. *BMC Medical Informatics and Decision Making, 21*.

[33] Redmon, J., & Farhadi, A. (2016). YOLO9000: Better, Faster, Stronger. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 6517-6525.

[34] Redmon, J., & Farhadi, A. (2018). YOLOv 3 : An Incremental Improvement.

[35] Bochkovskiy, A., Wang, C., & Liao, H.M. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. *ArXiv, abs/2004.10934*.

[36] Jocher, G.R., Stoken, A., Borovec, J., NanoCode, Chaurasia, A., TaoXie, Changyu, L., Abhiram, Laughing, tkianai, yxNONG, Hogan, A., lorenzomammana, AlexWang, Hájek, J., Diaconu, L., Marc, Kwon, Y., Oleg, wanghaoyang, Defretin, Y., Lohia, A., ah, M., Milanko, B., Fineran, B., Khromov, D.P., Yiwei, D., Doug, Durgesh, & Ingham, F. (2021). ultralytics/yolov5: v5.0 - YOLOv5-P6 1280 models, AWS, Supervise.ly and YouTube integrations.

[37] Li, C., Li, L., Jiang, H., Weng, K., Geng, Y., Li, L., Ke, Z., Li, Q., Cheng, M., Nie, W., Li, Y., Zhang, B., Liang, Y., Zhou, L., Xu, X., Chu, X., Wei, X., & Wei, X. (2022). YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications. *ArXiv, abs/2209.02976*.

[38] Ge, Z., Liu, S., Wang, F., Li, Z., & Sun, J. (2021). YOLOX: Exceeding YOLO Series in 2021. *ArXiv, abs/2107.08430*.

[39] Wang, C., Bochkovskiy, A., & Liao, H.M. (2022). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *ArXiv, abs/2207.02696*.

[40] Terven, J.R., & Esparza, D.M. (2023). A Comprehensive Review of YOLO: From YOLOv1 to YOLOv8 and Beyond. *ArXiv, abs/2304.00501*.

[41] Zheng, B., Jiang, G., Wang, W., Wang, K., & Mei, X. (2014). Ablation experiment and threshold calculation of titanium alloy irradiated by ultra-fast pulse laser. *AIP Advances, 4*, 031310.

[42] Ganin, Y., & Lempitsky, V.S. (2014). Unsupervised Domain Adaptation by Backpropagation. *ArXiv, abs/1409.7495*.

References

[43] Adadi, A., & Berrada, M. (2018). Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access, 6*, 52138-52160.
[44] Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C.A., Zagar, M., Hoffman, P.J., Stoeckius, M., Papalexi, E., Mimitou, E.P., Jain, J., Srivastava, A., Stuart, T., Fleming, L.B., Yeung, B.Z., Rogers, A.J., McElrath, J.M., Blish, C.A., Gottardo, R., Smibert, P., & Satija, R. (2020). Integrated analysis of multimodal single-cell data. *Cell, 184*, 3573 - 3587.e29.
[45] Li, T., Sahu, A., Talwalkar, A., & Smith, V. (2019). Federated Learning: Challenges, Methods, and Future Directions. *IEEE Signal Processing Magazine, 37*, 50-60.