nnDetection: A Self-configuring Method for Medical Object Detection

Medical Imaging & Intelligent Reality Lab.
Convergence Medicine/Radiology,
University of Ulsan College of Medicine
Asan Medical Center

Michael Baumgartner, Paul F. Jaeger, Fabian Isensee, Klaus H. Maier-Hein
MICCAI 2021

Jihoon Jung

2023.02.16

Abstract

- medical object **detection** is of **high clinical relevance** because diagnostic decisions often **depend on** rating of **objects rather than pixels.**
- **Following nnU-Net's agenda**, in this work we systematize and automate the configuration process for **medical object detection.**
- We **demonstrate** the effectiveness of nnDetection on two public benchmarks, **ADAM** and **LUNA16**, and propose **11 further medical object detection tasks** on public data sets **for** comprehensive method **evaluation**.

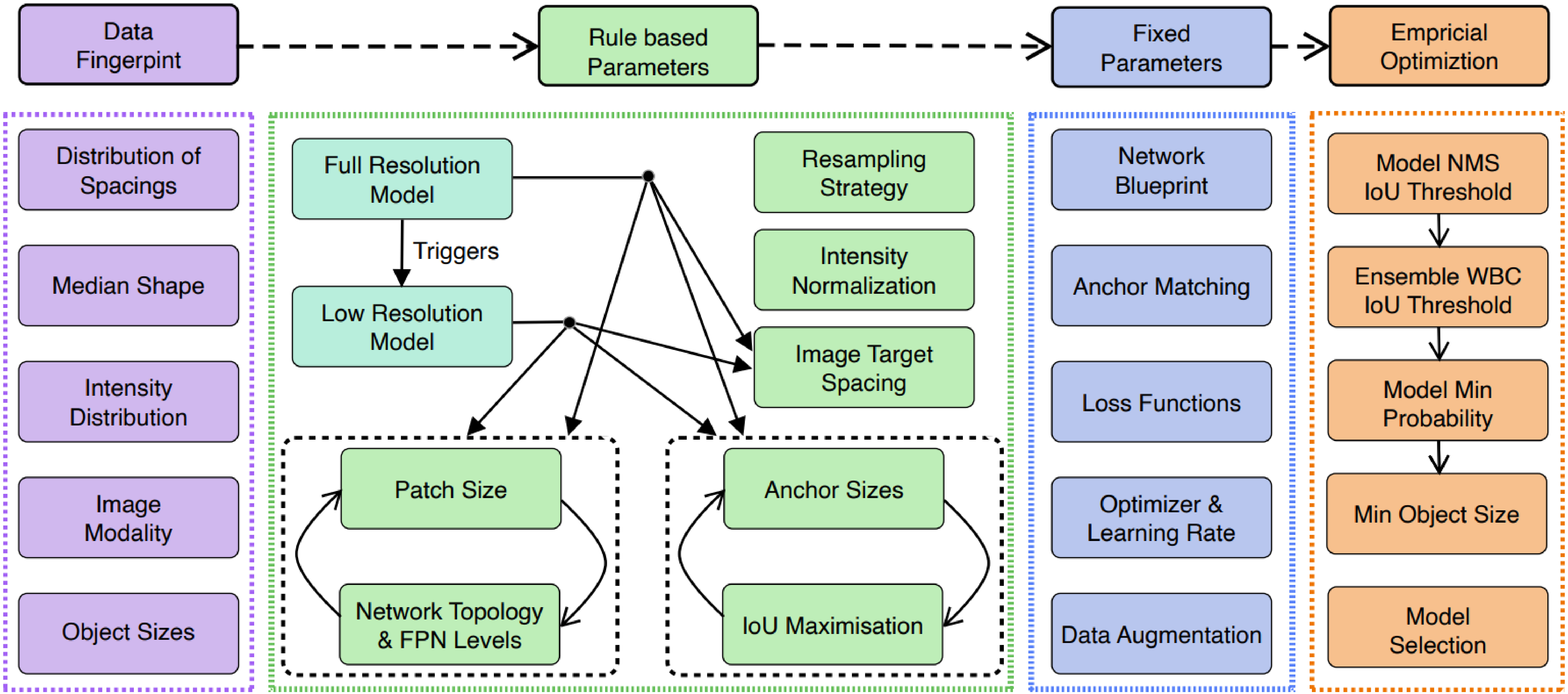
Introduction

- **semantic segmentation** remains the **predominant approach** in medical image analysis with 70% of biomedical challenges revolving around segmentation.
- **To be of diagnostic relevance**, however, in many use-cases segmentation methods require **ad-hoc postprocessing** that aggregates **pixel predictions to object scores.**
- Compared to a basic segmentation architecture like the U-Net, the set of **hyperparameters in a typical object detection architecture is extended by an additional detection head with multiple loss functions** including smart sampling strategies (“**hard negative mining**”), definition of size, density and location of prior boxes (“**anchors**”), or the consolidation of overlapping box predictions at test time (“**weighted box clustering**”).
- It further **aggravates** the already cumbersome and iterative process of method configuration, which currently **requires expert knowledge, extensive compute resources, sufficient validation data**, **and needs to be repeated on every new tasks** due to varying data set properties in the medical domain.

Introduction

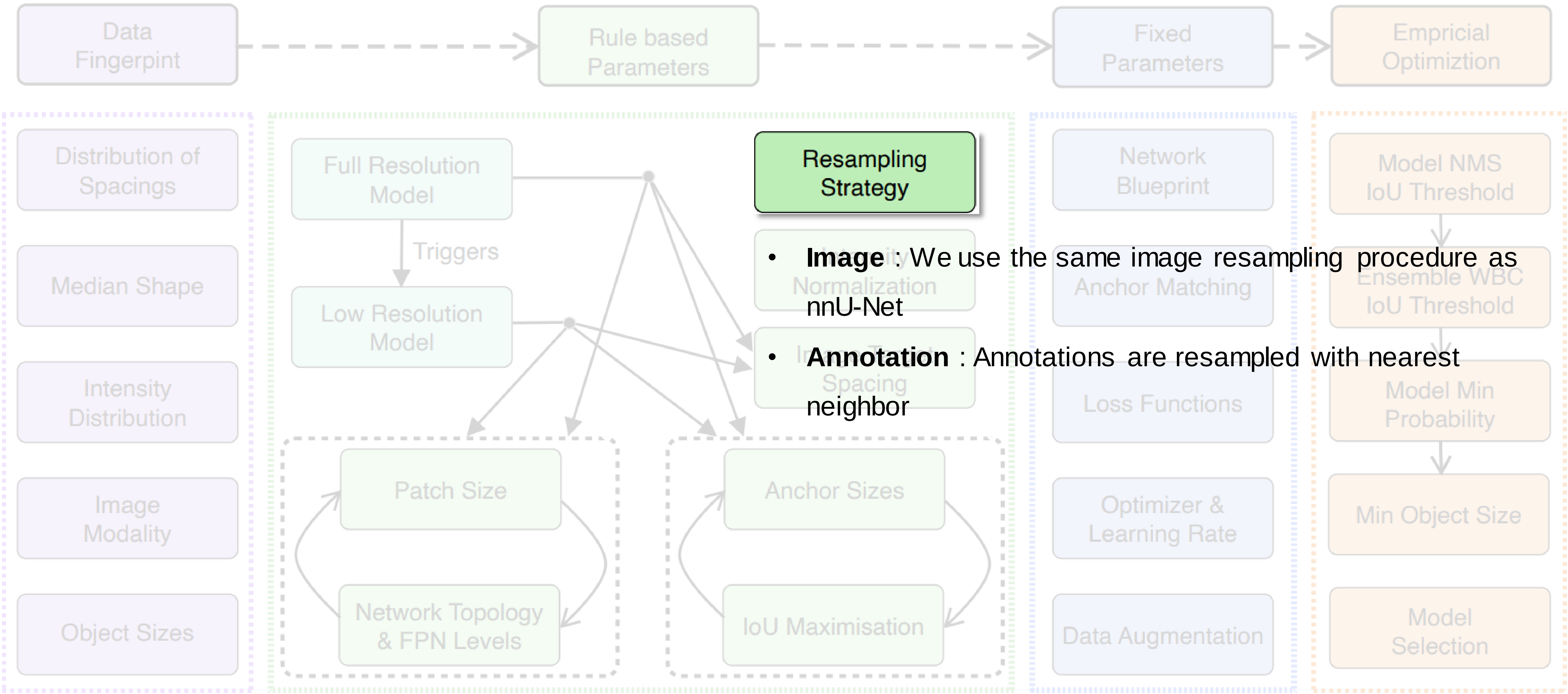
- In this work, we **follow the recipe of nnU-Net** to systematize and automate method configuration for medical object detection. Specifically, we identified **a novel set of fixed, rule-based, and empirical design choices** on a diverse development pool **comprising 10 data sets.**
- Without manual intervention, **nnDetection sets a new state of the art** on the nodule-candidate-detection task of the well-known **LUNA16** benchmark and achieves **competitive results on the ADAM leaderboard.**
- To address the current **lack of public data sets** compared to medical segmentation, we propose **a new large-scale benchmark totaling 13 data sets** enabling sufficiently diverse evaluation of medical object detection methods.
- To this end, we identified object detection tasks in **data sets of existing segmentation challenges** and **compare nnDetection against nnU-Net** (with additional postprocessing for object scoring) as a standardized baseline.
- We make **nnDetection** publicly **available as an out-of-the-box method for state-of-the-art object detection** on medical images, a framework for novel methodological work, as well as a standardized baseline to compare against without manual effort.

Method



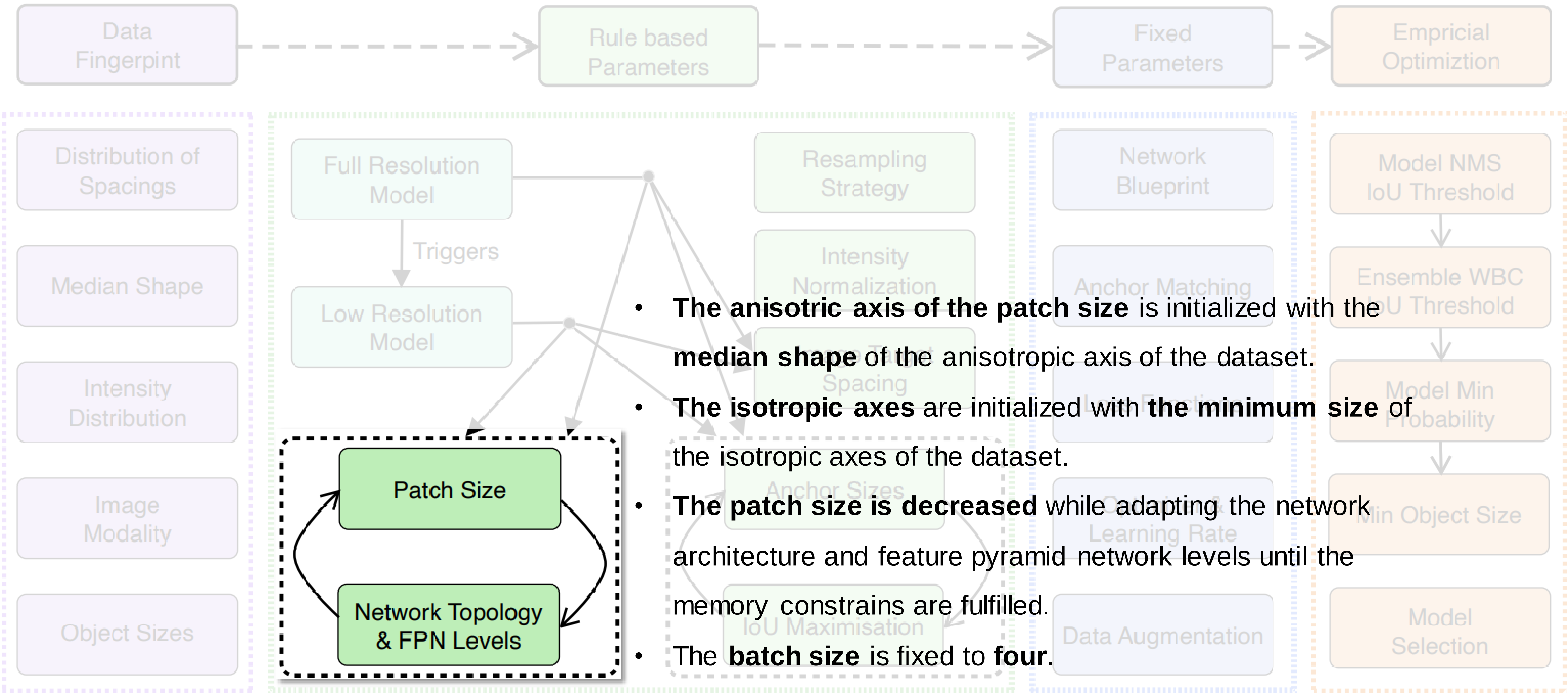
Data Fingerpint
Rule based Parameters
Fixed Parameters
Emprical Optimiztion
Distribution of Spacings
Median Shape
Intensity Distribution
Image Modality
Object Sizes
Full Resolution Model
Triggers
Low Resolution Model
Resampling Strategy
Intensity Normalization
Image Target Spacing
Patch Size
Network Topology & FPN Levels
Anchor Sizes
IoU Maximisation
Network Blueprint
Anchor Matching
Loss Functions
Optimizer & Learning Rate
Data Augmentation
Model NMS IoU Threshold
Ensemble WBC IoU Threshold
Model Min Probability
Min Object Size
Model Selection

Method



- **Image** : We use the same image resampling procedure as nnU-Net
- **Annotation** : Annotations are resampled with nearest neighbor

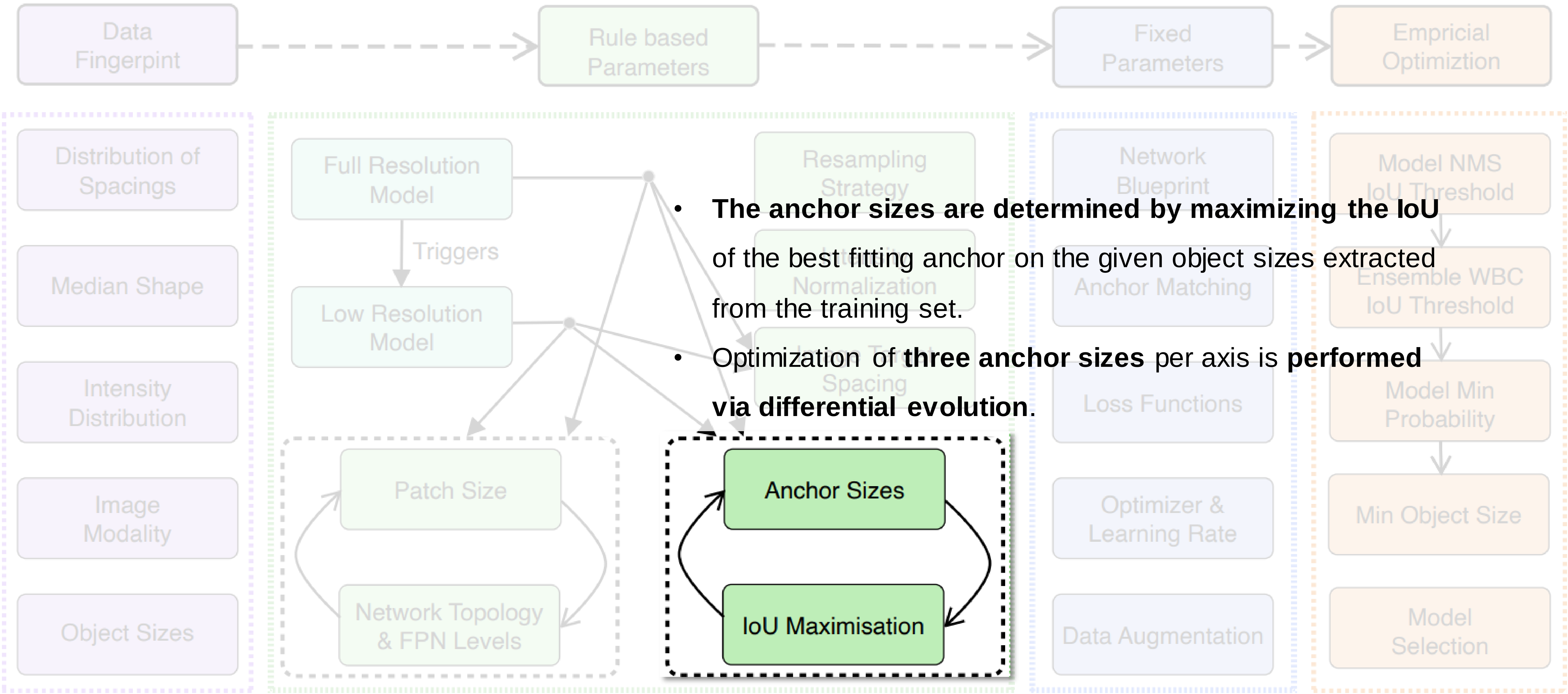
Method



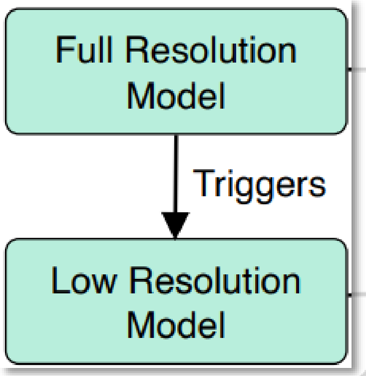
- **The anisotric axis of the patch size** is initialized with the **median shape** of the anisotropic axis of the dataset.
- **The isotropic axes** are initialized with **the minimum size** of the isotropic axes of the dataset.
- **The patch size is decreased** while adapting the network architecture and feature pyramid network levels until the memory constrains are fulfilled.
- The **batch size** is fixed to **four**.

Method

- **The anchor sizes are determined by maximizing the IoU** of the best fitting anchor on the given object sizes extracted from the training set.
- Optimization of **three anchor sizes** per axis is **performed via differential evolution**.



Method



- **The low resolution** configuration will be triggered if the **99.5 percentile of object sizes** along any axes exceeds the patch size of the full resolution model.
- **If the low resolution** configuration is **triggered**, the target spacing along each axes **will be increased by two** to incorporate more contextual information.

Method

- **Retina U-Net** with an encoder which consists of plain convolutions, **ReLU** and **instance normalization** blocks.
- The detection heads used for anchor **classification and regression consist of three convolutions with group norm.**

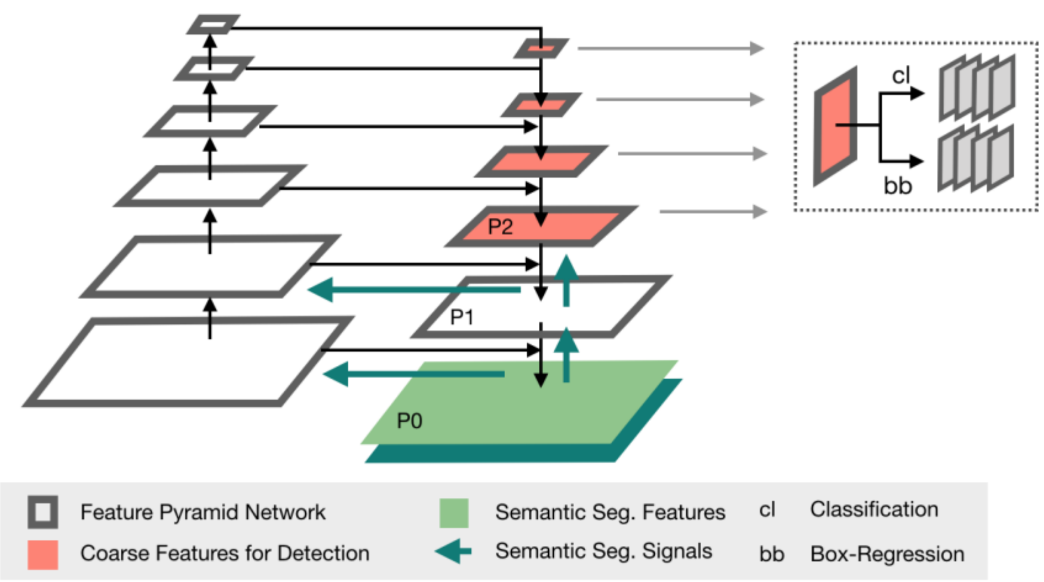


Figure 2. The Retina U-Net architecture in 2D.

Data Fingerprint

Distribution of Spacings

Median Shape

Intensity Distribution

Image Modality

Object Sizes

Fixed Parameters

Network Blueprint

Anchor Matching

Loss Functions

Optimizer & Learning Rate

Data Augmentation

Emprical Optimiztion

Model NMS IoU Threshold

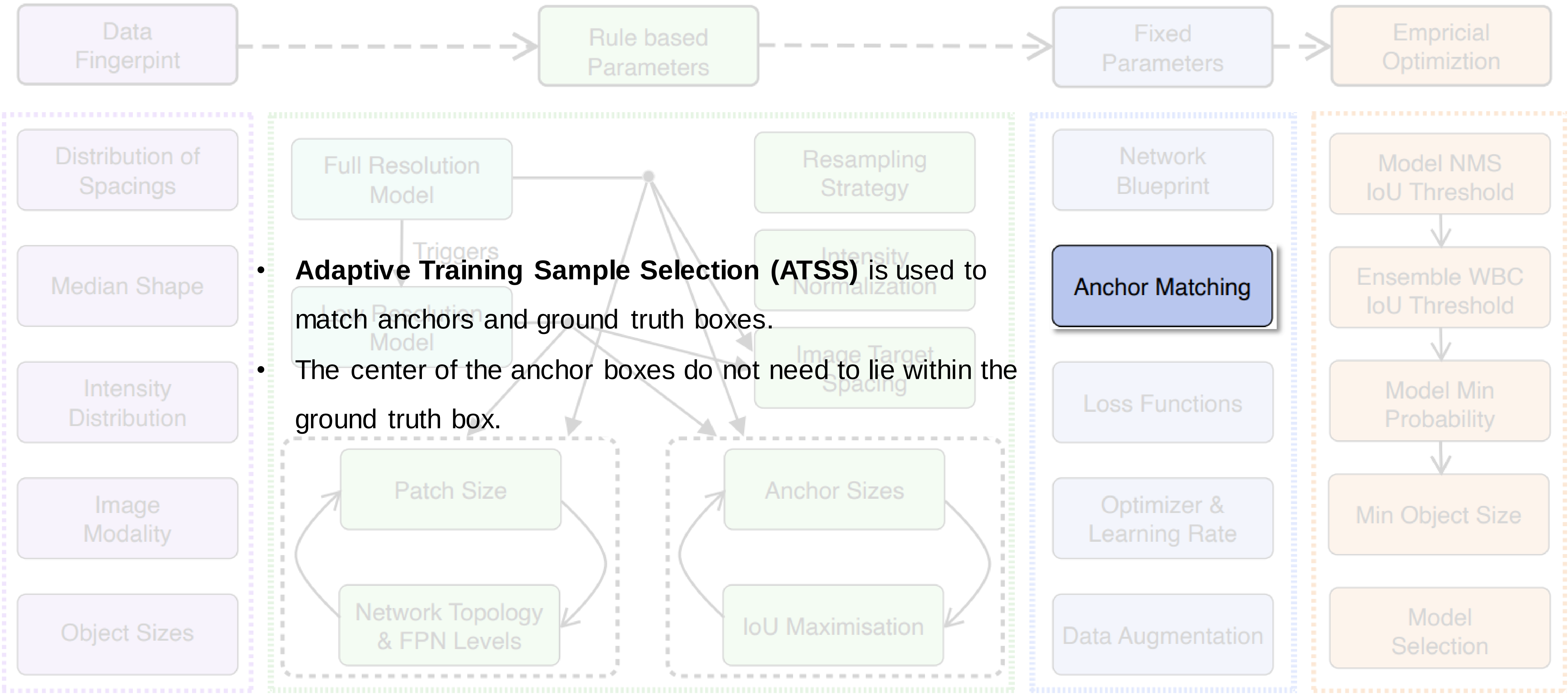
Ensemble WBC IoU Threshold

Model Min Probability

Min Object Size

Model Selection

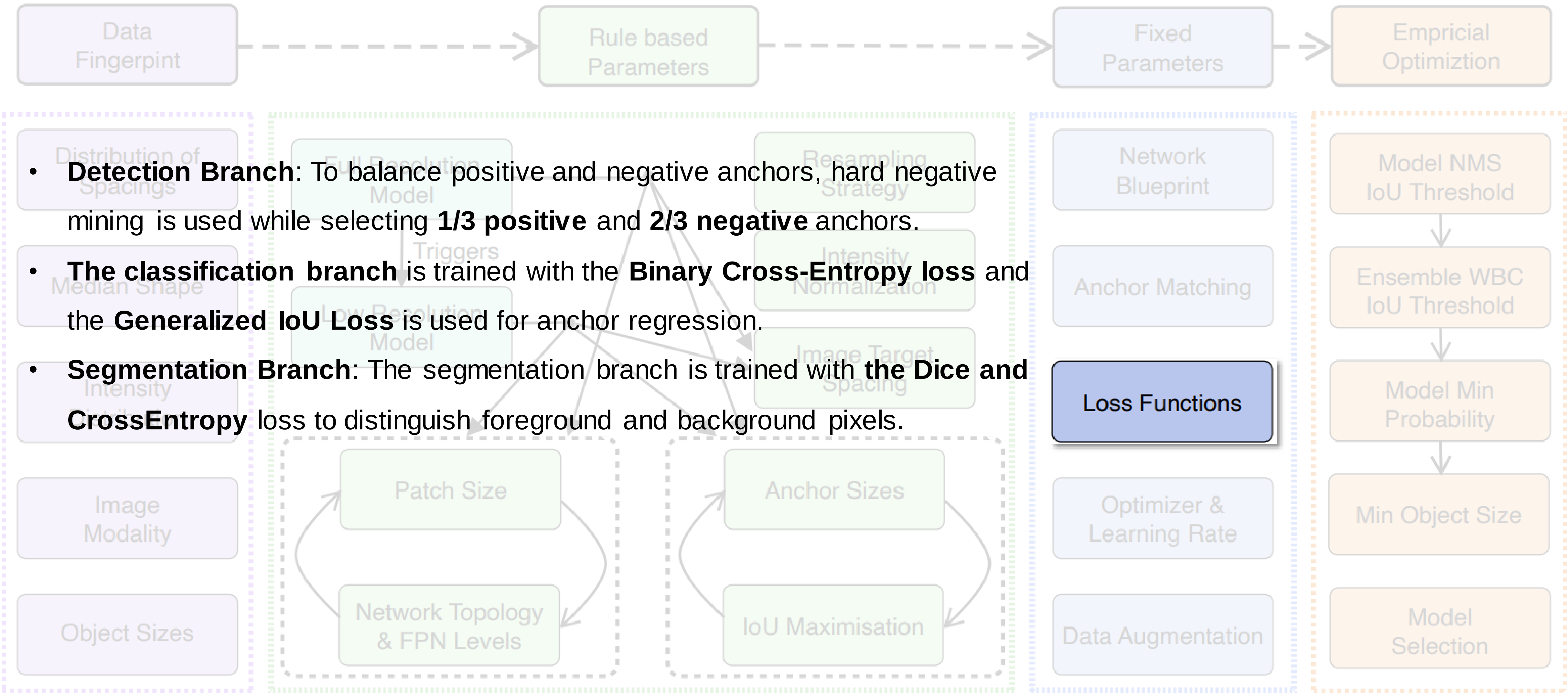
Method



- **Adaptive Training Sample Selection (ATSS)** is used to match anchors and ground truth boxes.
- The center of the anchor boxes do not need to lie within the ground truth box.

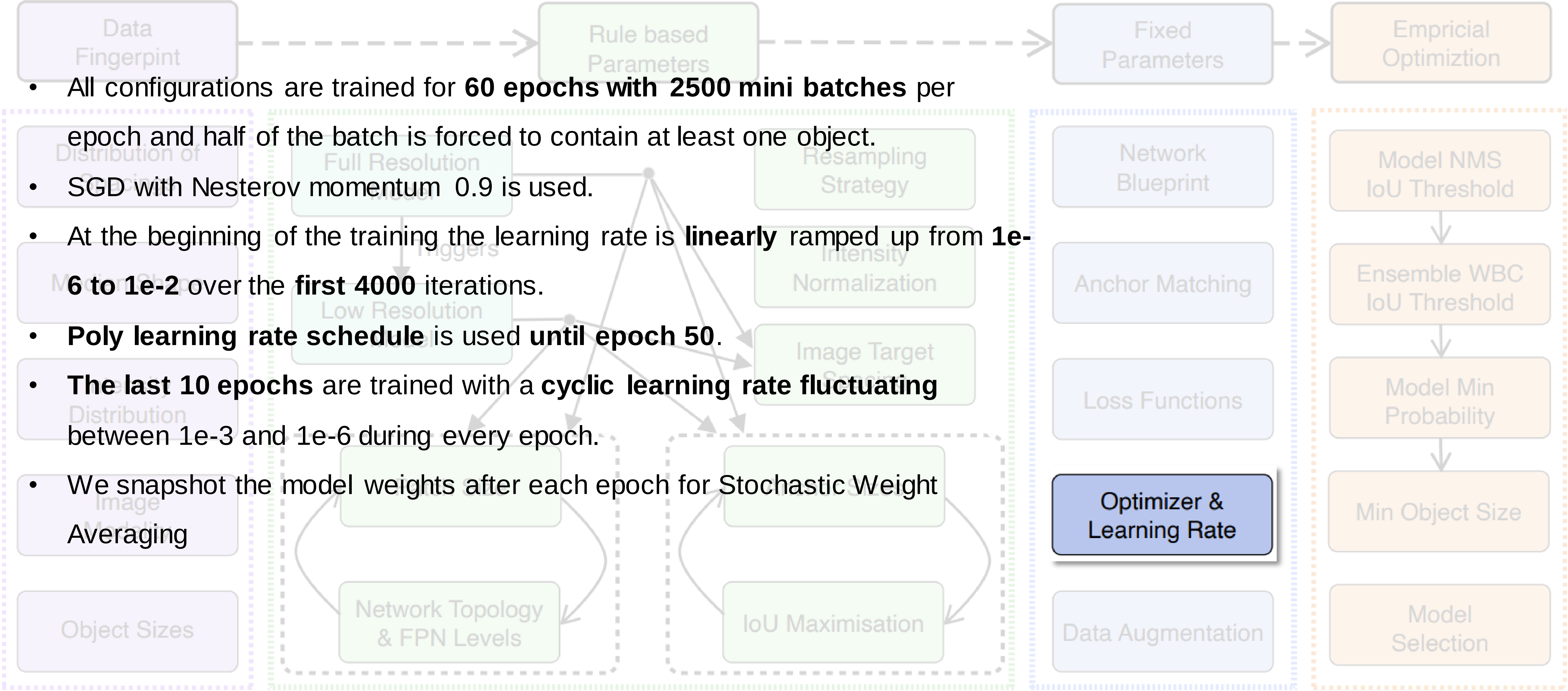
Method

- **Detection Branch**: To balance positive and negative anchors, hard negative mining is used while selecting **1/3 positive** and **2/3 negative** anchors.
- **The classification branch** is trained with the **Binary Cross-Entropy loss** and the **Generalized IoU Loss** is used for anchor regression.
- **Segmentation Branch**: The segmentation branch is trained with **the Dice and CrossEntropy** loss to distinguish foreground and background pixels.

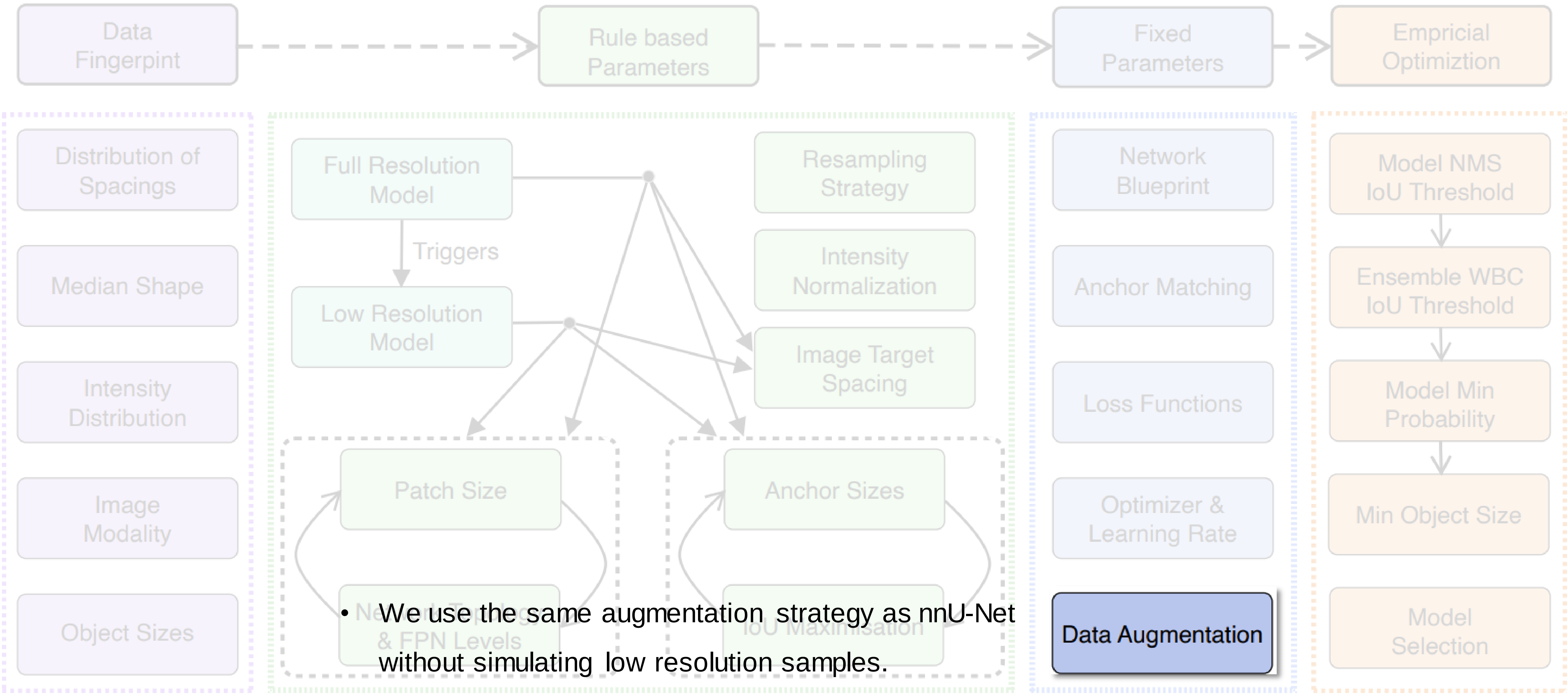


Method

- All configurations are trained for **60 epochs with 2500 mini batches** per epoch and half of the batch is forced to contain at least one object.
- SGD with Nesterov momentum 0.9 is used.
- At the beginning of the training the learning rate is **linearly** ramped up from **1e-6 to 1e-2** over the **first 4000** iterations.
- **Poly learning rate schedule** is used **until epoch 50**.
- **The last 10 epochs** are trained with a **cyclic learning rate fluctuating** between 1e-3 and 1e-6 during every epoch.
- We snapshot the model weights after each epoch for Stochastic Weight Averaging

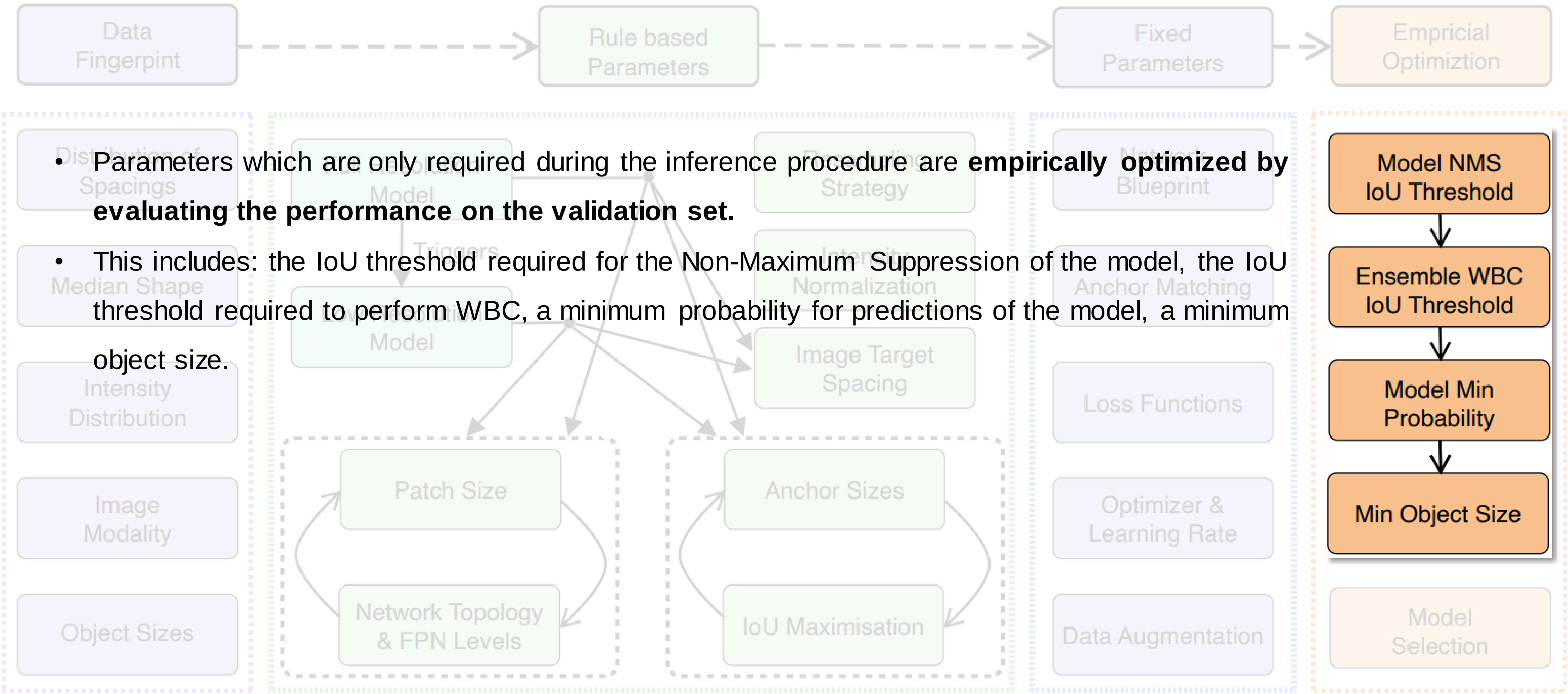


Method



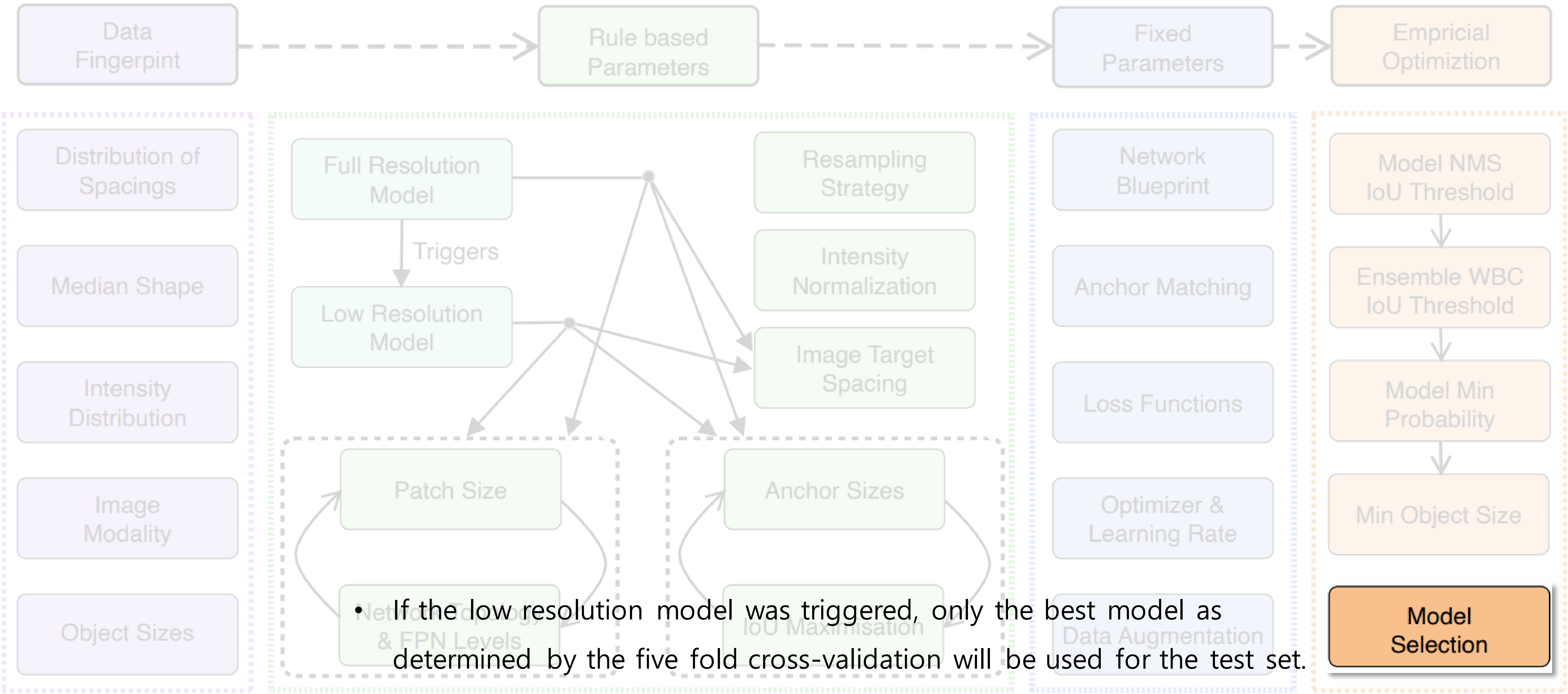
- We use the same augmentation strategy as nnU-Net without simulating low resolution samples.

Method



- Parameters which are only required during the inference procedure are **empirically optimized by evaluating the performance on the validation set.**
- This includes: the IoU threshold required for the Non-Maximum Suppression of the model, the IoU threshold required to perform WBC, a minimum probability for predictions of the model, a minimum object size.

Method



- If the low resolution model was triggered, only the best model as determined by the five fold cross-validation will be used for the test set.

nnDetection development

- To achieve automated method configuration in medical object detection, we roughly **follow the recipe** outlined in **nnU-Net.**
- **Development** was performed on **a pool of 10 data sets** (see supplementary material).
- **Fixed Parameters**
    a. We opt for **Retina U-Net** as our architecture template, which builds on the simple RetinaNet **to enable leveraging of pixel-level annotations**.
    b. we **discarded** the requirement as to which **the center point of selected anchors needs to lie inside the ground truth box** because, as we found it often resulted in **the removal of all positive anchors for small objects.**
- **Rule based Parameters**
    a. For as many of the remaining decisions as possible, we **formulate** explicit dependencies **between the Data Fingerprint and design choices** in the form of interdependent heuristic rules.
    b. We iteratively **maximize the intersection over union** (IoU) between anchors and ground-truth boxes.
    c. We found **performing** this optimization **on the training split** instead of the validation split led to **more robust** anchor configurations **due to a higher number of samples**.

nnUNet as an object detection baseline

- **nnUNet Basic**

  nnUNet Basic reflects the common approach to aggregating pixel predictions: Argmax is applied over softmax predictions, followed by **connected component analysis per foreground class**, and finally an object score per component is obtained as the **maximum pixel softmax score of the assigned category**.
- **nnUNet plus**

  To ensure the fairest possible comparison, we enhance the baseline by empirically choosing the following postprocessing parameters based on the training data for each individual task: **Replacement of argmax by a minimum threshold on the softmax** scores to be assigned to a component, **a threshold on the minimum number of pixels per object**, and the choice of the aggregation method (max, mean, median, 95% percentile).

Experiments and Results

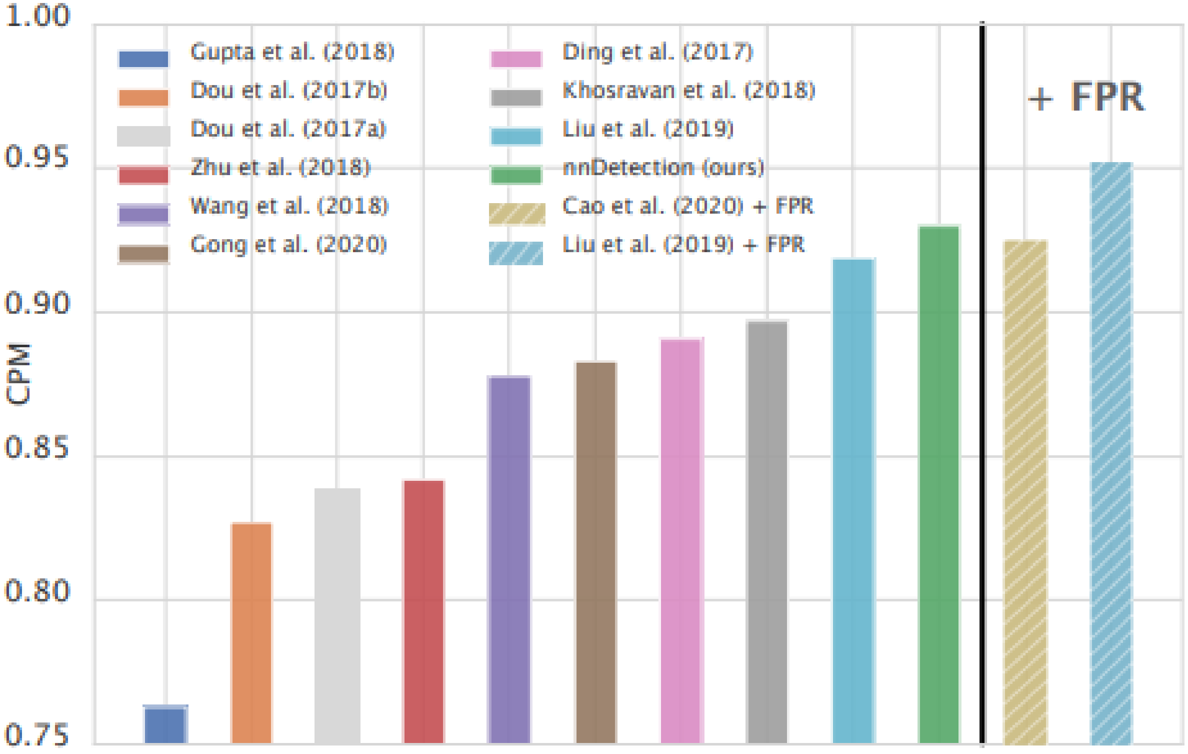
- **Proposed benchmark for medical object detection :**
  - ✓ We argue these aspects directly translate to medical object detection and thus propose a new benchmark based a diverse pool of **13 existing data sets.**
  - ✓ we identified object **detection tasks in 5 data sets** of existing segmentation challenges (where we **focus on detecting tumors and consider organs as background**, see supplementary material for details)
  - ✓ we performed connected component analysis and **discarded** all objects with a diameter **less than 3mm.**
- **Data sets.**
  - ✓ Out of the 13 data sets, we used 10 for development and validation of nnDetection.

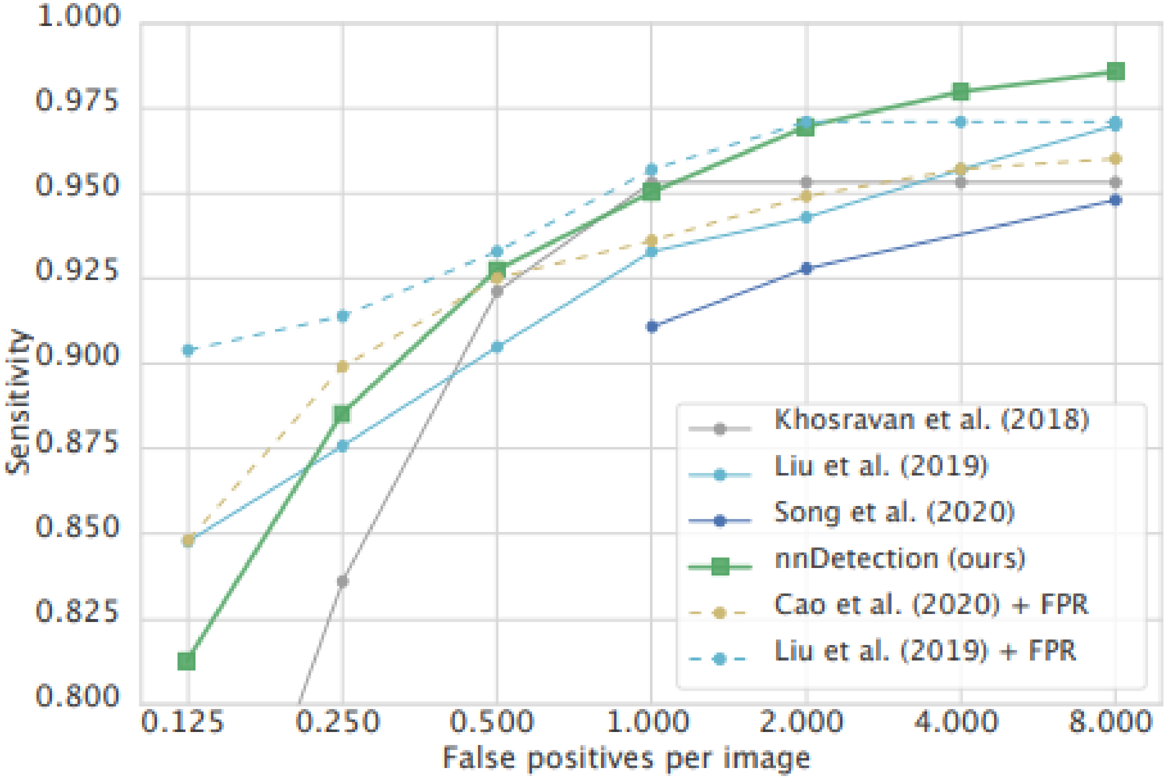
    Training – CADA(chest-CT / covid19) , LIDC-IDRI(chet-CT Lung), RibFrac(chest X-ray), Kits19(abdomen CT kidney tumor)

    Validation – ProstateX(MRI prostate), ADAM(abdomen and pelvis CT), Medical Segmentation Decathlon Liver, Pancreas, Hepatic Vessel and Colon

    Test - The test - LUNA16(chest-CT), and TCIA Lymph-Node(CT, PET/CT)

Experiments and Results





Experiments and Results

Collaborators

Medical
Imaging
Intelligent
Reality
Lab

Cardiac

June-Goo Lee
Gyu-Jun Jeong
Tae-Won Kim
Ji-Hoon Jung

Pathology

Hyunjeong Go, Gyuheon Choi
Gyungyub Gong, Dong Eun Song

Cardiology

Jaekwan Song, Jongmin Song
Young-Hak Kim

Anesthesiology

Sung-Hoon Kim, Eun Ho Lee

Neurology

Dong-Wha Kang, Chongsik Lee
Jaehong Lee, Sangbeom Jun
Misun Kwon, Beomjun Kim, Sun Kwon,
Eun-Jae Lee

Surgery

Beom Seok Ko, JongHun Jeong
Songchuk Kim, Tae-Yon Sung

Gastroenterology

Jeongsik Byeon, Kang Mo Kim, Do-hoon Kim

Emergency Medicine

Dong-Woo Seo

Pulmonology and Critical Care Medicine

Yoen-mok Oh, Sei Won Lee, Jin-won Huh