Universität Ulm | 89069 Ulm | Germany

Faculty of Mathematics and Economics

Optimal Field-Of-View in Low-Dose X-Ray Lung Imaging

Master thesis at Ulm University.

Submitted for the degree of Master of Science (M.Sc.) in Mathematical Data Science.

Submitted by:

Alexander Ortlieb

alexander.ortlieb@uni-ulm.de

Matriculation number: 1160978

Thesis advisors:

Prof. Dr. Friedhelm Schwenker

Ulm University, Ulm

Prof. Dr. Michael Götz

University Hospital Ulm, Ulm

Supervisors:

M.Sc. Heiko Hillenhagen and Dr. Arthur Wunderlich

Ulm, November 1, 2024

Version November 1. 2024

© 2024 Alexander Ortlieb

This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0) License. To view a copy of this license, please visit https://creativecommons.org/licenses/by-nc-sa/4.0/ or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

Name: Alexander Ortlieb Matriculation number: 1160978

Declaration

I hereby declare that I wrote the master thesis independently and used no other aids than those cited. In each individual case, I have clearly identified the source of the passages that are taken word for word or paraphrased from other works.

Ulm, November 1. 2024

Alexander Ortlieb

Acknowledgements

Foremost, I would like to thank my advisor Prof. Dr. Michael Götz for the great opportunity to write my master thesis under him in the field of experimental radiology. I loved the process and the possibility to have a real impact on future patient care. I also want to thank Prof. Dr. Friedhelm Schwenker for lending a hand whenever I asked for thesis advice and for his support in making this thesis possible.

This work would not have been as good and fun without Heiko Hillenhagen, whom I want to thank from the bottom of my heart for his great supervision and close cooperation throughout the last six months. Furthermore I want to thank Dr. Arthur Wunderlich for his help with the clinical data, taking phantom X-ray images with me and always responding fast when I reached out with questions. My appreciation also goes out to Dominik Penka for providing the annotations for the clinical dataset. And last but not least, thank you to Luisa Gallée for helping me set up the computer and providing quick on site troubleshooting for me.

Abstract

This thesis addresses the challenge of minimizing patient overexposure during chest X-ray (CXR) imaging by automatically identifying the optimal field-of-view (FOV) from low-dose images. CXR imaging is crucial for diagnosing conditions such as pneumonia, tuberculosis, and lung cancer, but radiation exposure poses risks, especially in patients requiring frequent imaging. Although European guidelines recommend minimizing radiation while covering essential diagnostic areas, studies show CXRs have a high overexposure rates with an average of over 45 % overexposure. By simulating low-dose conditions through the analysis of low-dose anthropomorphic phantom images, this work investigates two methods for FOV detection: a segmentation-based approach for isolating the lung region and a keypoint detection-based method that utilizes anatomical landmarks. For comparing the two different methods metrics such as the Box IoU score, overreach, underreach and field size reduction were utilized. The segmentation-based approach showed a possible average field size reduction of 33.5 % while eliminating underreach for almost all test images. This combination of low-dose simulation and automated FOV detection offers a promising solution to enhance patient safety and compliance with field-size standards in radiology.

Keywords: Chest X-ray, Low-dose simulation, Field-of-view detection, Segmentation, Keypoint detection

Zusammenfassung

Diese Arbeit befasst sich mit der Herausforderung, die Strahlenüberexposition von Patienten bei Thorax-Röntgenaufnahmen zu minimieren, indem das optimale Sichtfeld automatisch aus Niedrigdosisbildern ermittelt wird. Thorax-Röntgenaufnahmen sind entscheidend für die Diagnose von Erkrankungen wie Pneumonie, Tuberkulose und Lungenkrebs, bergen jedoch insbesondere für Patienten mit häufigen Untersuchungen Risiken aufgrund der Strahlenbelastung. Europäische Richtlinien empfehlen, die Strahlenbelastung auf das notwendige diagnostische Mindestmaß zu reduzieren, dennoch zeigen Studien bei Patienten eine hohe Überexpositionsrate von durchschnittlich über 45 %. Durch die Simulation von Niedrigdosisbedingungen, die anhand der Analyse von Röntgenbilder eines anthropomorphischen Phantoms entstanden, untersucht diese Arbeit zwei Methoden zur Sichtfeld-Erkennung: einen segmentierungsbasierten Ansatz zur Isolierung der Lungenregion und eine auf anatomischen Fixpunkten basierende Fixpunkt-Erkennungsmethode. Zum Vergleich der beiden Ansätze wurden Metriken wie Box-IoU-Wert, Overreach, Underreach und Sichtfeldreduktion verwendet. Der segmentierungsbasierte Ansatz zeigte eine mögliche durchschnittliche Sichtfeldreduktion von 33,5 %, während Underreach für fast alle Testbilder eliminiert wurde. Die Kombination aus Niedrigdosissimulation und automatisierter Sichtfeld-Erkennung bietet eine vielversprechende Lösung zur Erhöhung der Patientensicherheit und Einhaltung von Richtlinien in der Radiologie.

Stichwörter: Thorax-Röntgenaufnahme, Simulation von Niedrigdosisbedingungen, Sichtfeldreduktion, Segmentation, Fixpunkt-Erkennung

Contents

List of Abbreviations **ix**

1 Introduction **1**

2 Background **3**
2.1 Research Design 3
2.2 Overview of existing research 3
2.3 Theoretical background 6
2.3.1 X-ray imaging 6
2.3.2 Convolutional neural network 8
2.3.3 U-Net 12
2.3.4 EfficientNet-B0 14
2.4 Datasets 17
2.4.1 Segmentation 17
2.4.2 Keypoint detection 18
2.4.3 Simulation 18
2.5 Data analysis techniques 19
2.5.1 Segmentation metrics 19
2.5.2 Loss metrics 25
2.6 Tools and technologies used 27

3 Methods and Experiments **29**
3.1 Simulation 29
3.2 Segmentation 33
3.3 Keypoint detection 37
3.4 Challenges and Solutions 40

4 Results **43**
4.1 Segmentation 43
4.2 Keypoint detection 47

4.3 Overreach and Underreach 50
4.4 Exposure Reduction 50

5 Discussion **57**
5.1 Key Findings 57
5.2 Implications of the Results 60
5.3 Limitations of the Study and possible improvements 62

6 Conclusion **63**
6.1 Summary of the Research 63
6.2 Recommendations for Future Work 63

List of Figures **65**

List of Tables **68**

Bibliography **71**

Abbreviations

AP	Anterior Posterior
BCE	Binary Cross-Entropy
CDF	Cumulative distribution function
CNN	Convolutional neural networks
CSV	Comma-separated values
CT	Computer Tomography
Cu	Copper
CXR	Chest X-ray/ Chest radiographs
DICOM	Digital Imaging and Communications in Medicine
FLOPs	Floating Point Operations
FN	False Negatives
FOV	Field of view
FP	False Positives
GPU	Graphics processing unit
HD95	Hausdorff distance (95th percentile)
IoU	Intersection-over-Union
MBConv	Mobile Inverted Bottleneck
maxFS	Maximal field size
minFS	Minimal field size
MONAI	Medical Open Network for Artificial Intelligence
mSv	Millisieverts
MSELoss	Mean Squared Error Loss
PA	Posterior Anterior
PNG	Portable Network Graphics
ReLU	Rectified Linear Unit

SE	Squeeze-and-Excitation
SGD	Stochastic Gradient Descent
std	Standard deviation
TN	True Negatives
TP	True Positives
μAs	microampere-seconds

1 Introduction

Medical imaging can be an essential part of diagnosing medical conditions, with chest X-rays being among the most frequently performed radiological procedures [1]. However, to prevent unnecessary amounts of exposure, protocols such as the European Commission guidelines on minimal field size and radiation exposure aim to minimize overexposure while ensuring essential diagnostic areas are covered [2]. In 2015, Tschauner et al. found that only 4 % of chest X-rays fully met these guidelines in pediatric radiology, with an average overexposure of 45 % [3]. Furthermore, the clinical X-ray image dataset on adults from the University Hospital Ulm used in this thesis shows similar characteristics with an average overexposure of 51 %. This finding affirms the urgent need for optimization techniques to improve compliance with field size standards to minimize unnecessary radiation exposure.

Thoracic imaging, particularly chest X-ray (CXR), is widely used to detect various lung conditions such as pneumonia, tuberculosis, and lung cancer. Its non-invasive nature, low cost, and broad availability make it a staple in routine medical examinations. Nonetheless, a key challenge in X-ray imaging is the patient's exposure to radiation - especially those who require frequent imaging to monitor their chronic conditions. Minimizing exposure is essential to mitigate potential risks to patients. This thesis explores an approach to reduce overexposure by automatically identifying the optimal field-of-view (FOV) based on low-dose X-ray images. The FOV refers to the necessary area selected for imaging to ensure maximum diagnostic utility. By utilizing this information, the collimator can be accurately adjusted to capture a standard-dose image over the defined optimal FOV, effectively reducing overexposure. Still, low-dose images suffer more prominently from shot noise due to the reduced photon count, which leads to a drop in image quality, complicating the identification of the optimal field-of-view. In this thesis, two distinct approaches were explored: a segmentation-based and a keypoint detection-based approach. The segmentation method accurately highlights the lungs from surrounding structures, while the keypoint detection method leverages clinical expertise to define the optimal FOV based on anatomical landmarks.

Low-dose X-ray images were simulated from standard-dose images. To simulate low-dose imaging conditions, X-ray images of a anthropomorphic phantom were captured to analyze the impact of reduced radiation exposure. A standardized anthropomorphic phantom, which replicates the human anatomy, allowed for the control and adjustment of dose levels and the examination of a range of doses. This helped to better understand the noise and image degradation characteristics of low-dose X-ray imaging. This analysis formed the basis for generating low-dose X-ray simulations and evaluating the performance of the segmentation-based and keypoint detection-based approaches for FOV detection under varying radiation conditions.

The first approach for optimal FOV detection was lung segmentation-based due to its effectiveness in defining regions of interest in thoracic images. The optimal FOV defined as the minimum FOV necessary to image the lungs. Automated segmentation techniques, including deep learning and traditional image processing, can isolate lung fields from complex backgrounds, and exclude irrelevant anatomical regions from the FOV.

The keypoint detection-based approach offered an alternative strategy that utilized clinical knowledge of anatomical landmarks, such as the location of the T1 and T12 spinal segments and the costophrenic angle. These keypoints were derived from the guidelines defined by the European Commission for thorax X-ray imaging to determine the optimal FOV for thoracic imaging. Clinicians typically use landmarks like the rib cage and collar bones to manually adjust window settings on X-ray systems, ensuring the inclusion of diagnostically relevant areas in the FOV. By integrating clinical data into an automated detection system, the aim was to develop a robust method capable of functioning effectively even in low-dose settings.

The novelty of this thesis lies in the combination of low-dose image simulation with the exploration of two distinct methods for optimal FOV detection. By investigating both segmentation-based and keypoint detection-based procedures, the goal was to determine the possible lowest dose for FOV detection with an acceptable accuracy and to provide a reliable system that can automatically select the optimal FOV from low-dose thoracic X-ray images, enhancing diagnostic accuracy while minimizing radiation exposure.

2 Background

2.1 Research Design

This thesis investigated methods to simulate low-dose X-ray images from standard-dose images by performing statistical analysis on phantom X-ray images captured at different radiation levels. An anthropomorphic phantom was used to understand how existing radiology software handle lower exposures. Based on this analysis, low-dose images were simulated for training machine learning models.

To detect the optimal field-of-view, two machine learning architectures were compared for the lung segmentation-based and keypoint detection-based approaches. First, models were tested on the original dataset, followed by a hyperparameter search to find the best configurations. Evaluation metrics included accuracy, Dice score, Hausdorff distance, box IoU score and reduction in FOV.

With the best hyperparameters, the models were then trained on simulated low-dose images. This provided insights into the models performance across different simulated data, assessing their robustness under varying conditions.

The research aimed to identify the most effective approach for detecting the optimal field-of-view in low-dose X-ray images to reduce radiation exposure.

2.2 Overview of existing research

The paper "European Guidelines for AP/PA chest X-rays: routinely satisfiable in a paediatric radiology division?", as already mentioned in the introduction, evaluated the adherence of a pediatric radiology division to the European guidelines for chest X-rays [3]. These guidelines emphasize minimizing radiation exposure while ensuring diagnostic accuracy. In the study an average overexposure of 45 % was measured.

Ronneberger et al. (2015) introduced the U-Net architecture in "U-Net: Convolutional Networks for Biomedical Image Segmentation" for medical image segmentation [4]. The U-Net uses a fully convolutional network structure that includes an encoder-decoder pathway with skip connections. This helps retain spatial information from the models downsampling process. A visualization of the architecture can be seen in Figure 2.1. This architecture demonstrated great performance in segmenting biomedical images, particularly for tasks like cell tracking, which involves identifying and following individual cells or groups of cells over time in a sequence of images to analyze their behavior, movement, and interactions. It allowed efficient training on small datasets while providing high accuracy in tasks requiring precise boundary delineation of objects within images, which is a vital component of lung image analysis.

Building on this, Novikov et al. (2017) extended the application of the U-Net architecture to lung segmentation tasks in the paper "Fully Convolutional Architectures for Multiclass Segmentation in Chest Radiographs" [5]. Using chest radiographs (CXR), the U-Net architecture was able to reliably perform lung field segmentation, helping with the identification of key anatomical structures within chest X-rays. This development showcased the U-Net's versatility and effectiveness in various medical image segmentation applications, ranging from organ delineation to disease localization. The architecture's ability to handle complex multiclass segmentation challenges further expanded its relevancy for medical imaging research.

ULTRA-X-COVID a model proposed in the paper "COVID-19 Detection via Ultra-Low-Dose X-ray Images Enabled by Deep Learning" presented a novel deep learning approach, designed to detect COVID-19 using ultra-low-dose X-ray images [6]. The paper utilized a modified U-Net architecture for denoising of the ultra-low-dose X-ray images and a residual network for further classification. The images were simulated from standard-dose images by adding white Gaussian noise to simulate electron noise and detector blur.

The paper "Automated segmentation of 2D low-dose Computer Tomography (CT) images of the psoas-major muscle using deep convolutional neural networks" investigated the use of U-Net architecture for automatic segmentation of the psoas-major muscle in low-dose CT images. This research aids the medical fields involved in muscle mass assessment and sarcopenia diagnosis [7]. Using five-fold cross-validation, the model achieved an intersection-over-union (IoU) accuracy of 86 % and a cross-sectional area (CSA) ratio of 103 %, proving competitive results with traditional methods that are effective for low-dose imaging applications.

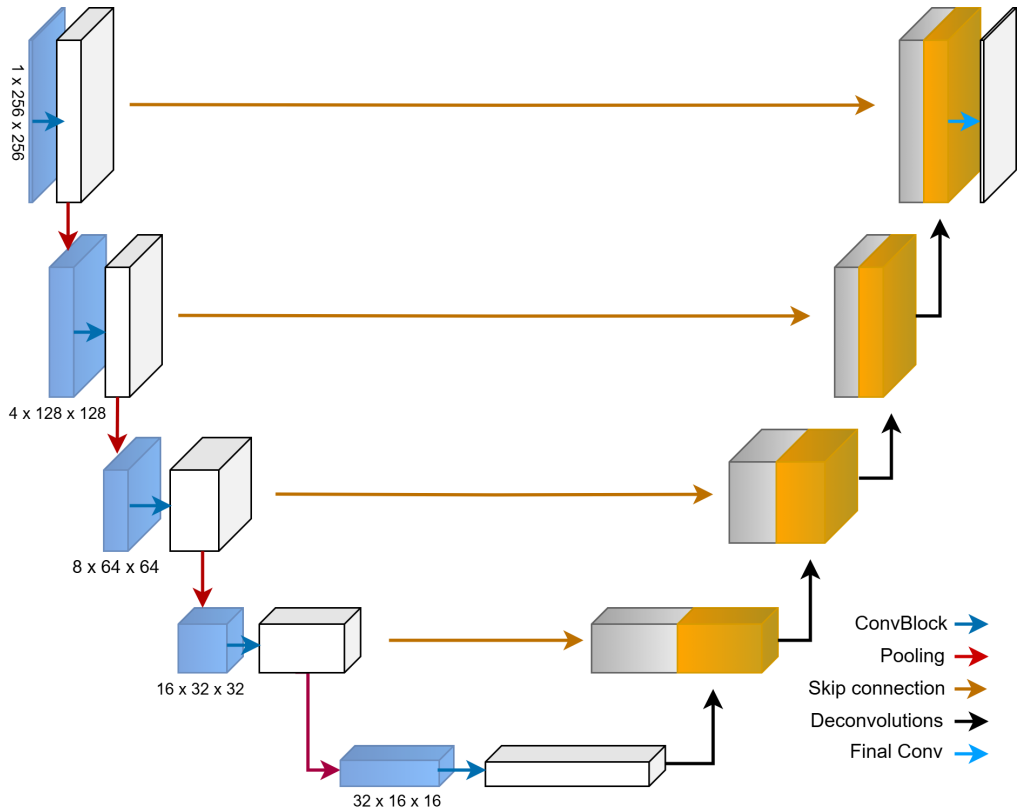

Figure 2.1: An example visualization of a U-Net architecture is shown, where each box represents the data and its dimensions. The downsampling process is illustrated by the numbers indicating the shape of the data at each stage.

Lastly, in the paper "Accurate Automated Keypoint Detections for Spinal Curvature Estimation," the authors proposed a method to improve the precision of keypoint detection for estimating spinal curvature [8]. This approach leveraged advanced machine learning techniques to accurately identify spinal landmarks, which are essential for evaluating curvature and diagnosing spinal conditions. Tested on clinical imaging datasets, the model demonstrated high accuracy, suggesting it could support more reliable and efficient spinal curvature assessments in medical settings. This work enhances existing methods, highlighting the impact of precise keypoint detection in improving diagnostic accuracy for spinal health evaluations.

These papers provided a strong foundation for the two approaches, lung segmentation and keypoint detection, presented in this thesis. No papers were found on the

topic of low-dose field-of-view detection in X-ray images, which gives rise to the research field of optimal field-of-view detection. In the future, the goal is to sustainably reduce patient radiation exposure during chest X-ray examinations and expand its application areas.

2.3 Theoretical background

2.3.1 X-ray imaging

X-rays are a form of high-energy electromagnetic radiation capable of penetrating most materials to varying degrees, depending on their density. X-ray imaging is primarily used in medical diagnostics to view the internal structures of the body, such as bones, tissues, and organs. Different body materials absorb X-rays differently, depending on their density. Dense structures like bones absorb more X-rays and thus appear lighter on the resulting image. Softer tissues, such as muscles or lungs, absorb fewer X-rays and show up in shades of gray or black. The X-rays that pass through the body are captured either on film or by digital detectors, creating an image that can be used for medical diagnosis [9].

An X-ray machine utilizes a filament inside a high vacuum tube. When heated the filament releases electrons. These electrons are accelerated toward a metal target, typically made of tungsten, using a high voltage potential difference. When the electrons collide with the target, they can knock loose an electron inside the lower orbits of the target metal atom. When this happens, an electron from a higher orbit drops into its place. The energy of an electron increases with its orbit, and as an electron transitions to a lower orbit, it releases part of the excess energy as a photon. X-ray photons have an energy level between 100 eV and 100 keV. These X-rays have a general direction that is determined by the angle of the target metal and are further restricted towards the object of interest by a collimator. This limits the spread of X-rays and the subsequent exposure to surrounding tissues.

Although very useful in healthcare, X-rays involve ionizing radiation, which can damage cells and DNA [10]. For this reason, experts recommend the exposure to X-rays should be minimized [11].

Imaging noise

Various types of noise can occur during the process of X-ray imaging. Firstly photon noise, also known as shot noise, refers to the statistical fluctuations in the number of photons detected by a sensor, and it arises from the discrete nature of light. Since light consists of individual photons, the number of photons detected over a given period can vary. This is especially the case when the signal is weak, such as in low-dose X-ray imaging. This type of noise is well described by the Poisson process, a mathematical model that governs events occurring independently over time, such as the arrival of photons at the detector [12].

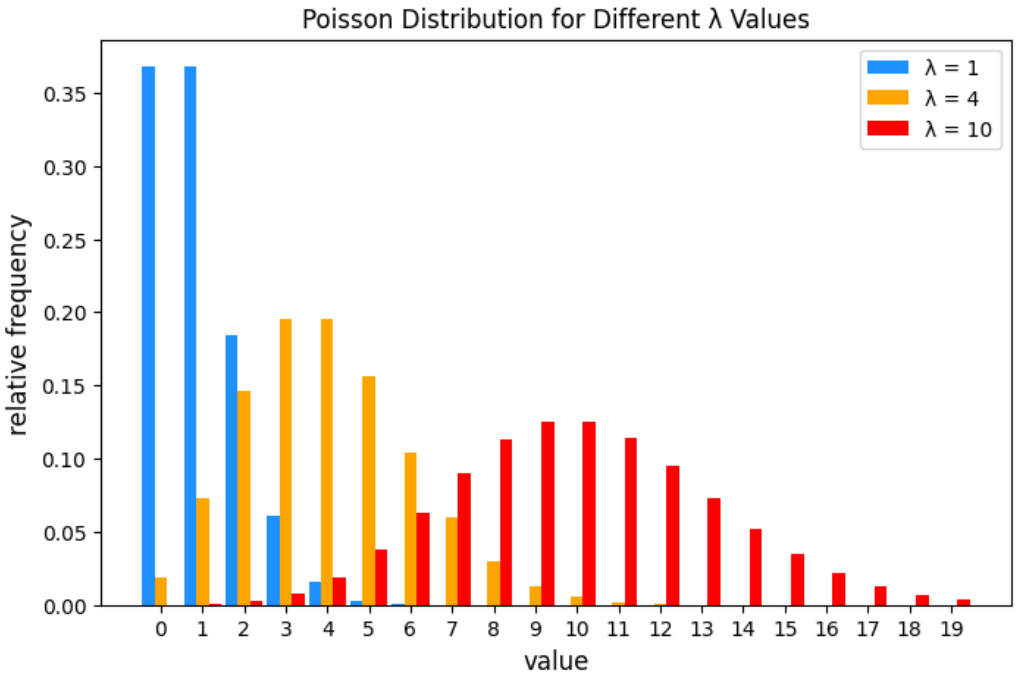

Figure 2.2: Example of different Poisson distributions.

In a Poisson process, the probability of detecting a certain number of photons in a given time period follows a Poisson distribution, where the variance of the distribution equals its mean. Different Poisson distributions can be seen in Figure 2.2. This means that as the number of detected photons decreases, the relative noise (variance) becomes more significant when compared to the signal. In low-light or low-dose imaging conditions, the smaller number of photons results in increased fluctuations, making photon noise more noticeable. The Poisson nature of photon counting makes

this noise highly relevant in low-dose X-ray imaging.

Secondly, electronic noise refers to random fluctuations in electrical signals within an imaging system, typically caused by imperfections and thermal activity in the sensor's components. Electronic noise can arise from a variety of sources, including the random motion of electrons (thermal noise or Johnson-Nyquist noise) and variations during the readout process. Electronic noise is generally modeled as Gaussian noise, meaning it follows a normal distribution where most noise values cluster around a mean and fewer extreme values.

This type of noise is independent of the intensity of the incoming signal, unlike photon noise, and is inherent to the hardware itself. It can affect the accuracy and quality of the captured images, especially in low-light situations or long exposure times, where the noise becomes more noticeable relative to the signal.

Histogram matching

Histogram matching, also known as histogram specification, is a process in image processing where the pixel intensity distribution (histogram) of a source image is adjusted to match the histogram of a reference image. This transformation is based on the cumulative distribution functions (CDFs) of both images. This process is done separately for each channel of the images.

For each intensity value in the source image, the corresponding value in the reference image is determined by matching their cumulative distribution function values. The closest intensity in the reference image's CDF is identified for each pixel intensity in the source image. This mapping modifies the pixel intensity distribution in the source image to align with that of the reference image.

2.3.2 Convolutional neural network

In this work, Convolutional Neural Networks (CNNs) were entailed for the segmentation and the keypoint detection tasks. Segmentation involves partitioning an image into different segments or object classes on a pixel-wise basis and keypoint detection consists of locating specific and distinct points or positions within an image or a video frame [13].

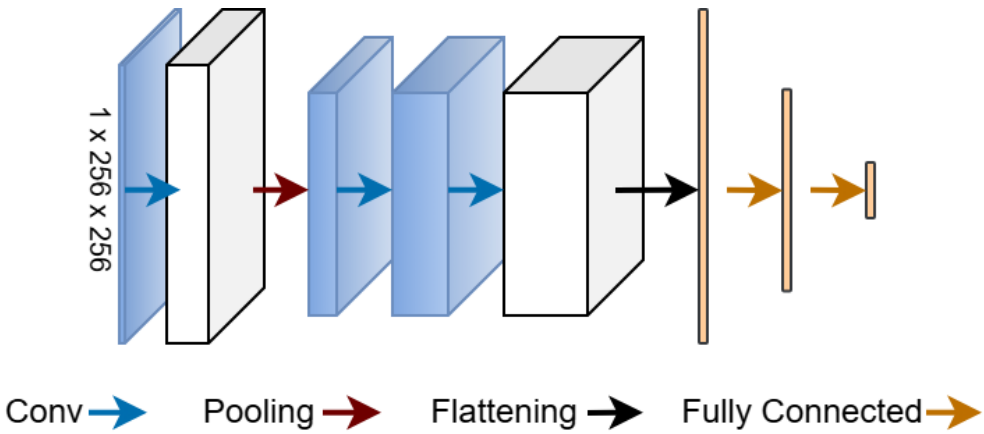

Figure 2.3: Example visualization of a CNN architecture.

Convolution

Convolution is a mathematical operation of two functions yielding another function. The term convolution also refers to the operation and the resulting output. Convolution is used to process input data, for example images, by highlighting specific features like edges, textures, and patterns. The general mathematical definition of convolution, where f and g are functions of the same domain is as follows [14].

$$(f * g)(t) := \int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau$$

This definition can be discretized for image convolution

$$\omega * I(x,y) := \sum_{i=0}^{a}\sum_{j=0}^{b} \omega(i,j)I(x-i,y-j)$$

where I is the input image and ω is the filter kernel. a and b describe the dimensions of the image.

In a CNN, convolution involves applying a small matrix called a filter or kernel to the input matrix. The kernel, typically smaller than the input matrix (e.g., 3 × 3 or 5 × 5), slides over the input matrix, moving a certain number of steps at a time, known as the stride. At each position, the filter performs an element-wise multiplication with the corresponding elements of the input matrix and sums these products to produce a

single value in the output feature map. This process repeats across the entire input.

Two important hyperparameters in convolutional layers are padding and stride. Padding refers to adding extra rows and columns (typically zeros) around the input image to preserve the spatial dimensions after the convolution operation. There are two main types of padding: valid padding, where no padding is added, resulting in a smaller output, and same padding, where padding is added to ensure the output has the same dimensions as the input. This is useful when the goal is to maintain the same output size as the input size.

By adjusting the stride and padding, the size of the output can be controlled, which allows the network to focus on different scales of patterns in the input image. A stride of 1 means the filter moves one pixel at a time, while a stride of 2 skips every other pixel, resulting in downsampling.

For example, a 5 × 5 input matrix and a 3 × 3 filter with a stride of 1 and no padding, this results in an output feature map of size 3 × 3, which is smaller than the input due to the lack of padding.

Batch normalization

Batch normalization normalizes the inputs of each layer by adjusting the input values to have zero mean and unit variance during training. This reduces internal covariate shift. This stabilization improves training speed and gradient flow, making deep networks converge faster [15].

Batch normalization is typically applied before the activation function because normalizing inputs before activation ensures a consistent range of values. This helps prevent issues like vanishing or exploding gradients, ensuring balanced activations and better gradient propagation, leading to faster, more stable learning.

After normalization, the network learns two parameters for each layer: the scale γ that scales the normalized output and the shift β that shifts the normalized output.

The transformation for each feature in the batch is given by:

$$\hat{x} = \frac{x - \mu_{\text{batch}}}{\sqrt{\sigma^2_{\text{batch}} + \epsilon}} \tag{2.1}$$

$$y = \gamma \hat{x} + \beta \tag{2.2}$$

where:

- x is the input to the layer,
- μ_{batch} and σ_{batch} are the mean and standard deviation (std) of the batch,
- $\hat{x}$ is the normalized input,
- y is the final output after scaling and shifting,
- ϵ is a small constant added for numerical stability.

Activation function

After the convolution operation, an activation function, typically the Rectified Linear Unit (ReLU) function, was applied to introduce non-linearity into the network. Other activation functions, such as sigmoid or tanh, are used less frequently in CNNs due to their slower convergence and potential gradient-related issues.

$$\mathrm{ReLU}(x) = \max(0, x) \tag{2.3}$$

The ReLU activation function replaces all negative values in the feature maps with zero, keeping only positive values. This non-linear transformation allows the network to learn more complex patterns and relationships, beyond simple linear combinations of input data [16]. It is preferred in most CNNs because it helps avoid the vanishing gradient problem, which can hinder the training of deep networks.

Pooling layer

The pooling layer is responsible for downsampling spatial dimensions of the feature maps, reducing the number of parameters and computational cost. The most common type of pooling used in CNNs is max pooling, where the maximum value is selected from a small window (typically 2 × 2) in the feature map. This operation reduces the spatial size of the feature maps while preserving the most important features.

By downsampling the feature maps, pooling layers help the network become more invariant to small translations or distortions in the input image. This invariance is particularly useful for tasks such as object recognition, where the position of objects in the image can vary. Pooling also helps prevent overfitting by reducing the number of parameters in the network.

Fully connected layer

After several layers of convolutions and pooling, the feature maps are typically flattened into a 1D vector and passed to one or more fully connected layers. These fully connected layers are similar to those used in traditional feedforward neural networks, where each neuron is connected to every neuron in the previous layer.

The fully connected layers are responsible for making the final prediction by combining the high-level features extracted by the convolutional and pooling layers. For example, in an image classification task, the final fully connected layer outputs a vector of class probabilities, one for each class, using a softmax activation function.

2.3.3 U-Net

The U-Net is a convolutional neural network architecture that is widely used for image segmentation tasks [17]. It follows a symmetrical encoder-decoder structure, allowing it to capture both high-level contextual information and fine-grained spatial details. The architecture is designed to output pixel-level predictions, making it ideal for tasks that require precise localization of objects within an image. The U-Net consists of two main parts: a contracting path (encoder) that downsamples the input image, and an expanding path (decoder) that upsamples it back to its original resolution. Between the encoder and decoder is a bottleneck layer which represents the deepest part of the network. This layer captures the most abstract and compressed representation of the input, allowing the network to focus on the most essential features before the decoder reconstructs the output.

Contracting path (encoder)

The contracting path, or encoder, is responsible for capturing high-level features from the input image. It consists of a series of convolutional blocks, each followed by a max-pooling operation. Each convolutional block typically consists of two consecutive 3×3 convolution layers, each followed by a ReLU activation function. A batch normalization operation can be added before or after the activation function. After the convolution block, a 2×2 max pooling operation is applied to reduce the spatial dimensions by half, which helps the network focus on larger patterns and reduces computational cost. As the network progresses through the contracting path, the spatial resolution of the feature maps decreases while the number of feature channels

increases. This allows the network to learn more abstract, high-level representations of the input image while also reducing its spatial resolution.

Bottleneck

In the middle (bottom) of the U-Net, the bottleneck layer connects the encoder and decoder. This layer represents the most compressed representation of the input image, where spatial resolution is at its lowest, and feature channels are at their highest. The bottleneck consists of a convolution block, and no pooling takes place to ensure that no downsampling occurs. The bottleneck captures the most abstract features from the image, providing a bridge between the encoding and decoding paths.

Expanding path (decoder)

The expanding path, or decoder, is responsible for reconstructing the spatial resolution of the input image and providing precise segmentation results. The decoder upsamples the feature maps from the bottleneck back to the original image resolution. Each upsampling step is typically performed using transposed convolutions (also known as deconvolutions), which increase the spatial dimensions of the feature maps. At each level of the decoder, the upsampled feature maps are concatenated with the corresponding feature maps from the contracting path via skip connections. This ensures that the decoder has access to both high-level contextual information from the bottleneck and low-level spatial details from the encoder. After concatenation, a convolution block is applied to refine the feature maps at each step of upsampling.

Skip connections

Skip connections are a crucial part of the U-Net architecture. They connect the feature maps from the contracting path directly to the corresponding feature maps in the expanding path. These connections allow the network to retain important spatial information that might otherwise be lost during downsampling. By concatenating feature maps from the encoder to the decoder, the network can combine high-level, abstract features with low-level, detailed features. This improves the precision of the segmentation, particularly for fine structures and boundaries.

Final output layer

The final layer of the U-Net is a 1×1 convolution that refines the output feature maps from the decoder to the desired number of output channels. In binary segmentation

tasks, the output will be a single channel, where each pixel represents the probability of belonging to the foreground or background. For multi-class segmentation, the output will have as many channels as the number of classes, and each pixel will be assigned a probability distribution over the different classes. A sigmoid activation is applied in the binary case, while a softmax activation is used for multi-class segmentation.

Used U-Net architecture

Our U-Net architecture applied in this thesis is shown in Figure 2.1. As previously described, the encoder comprises four convolutional blocks, each with batch normalization applied before the activation function. These convolutional blocks increase the number of feature channels and were included in the hyperparameter search. The feature channel expansion, as depicted in Figure 2.1, follows the pattern [4, 8, 16, 32], with max pooling layers reducing the spatial size of each channel by a factor of four after each block. The bottleneck and decoder follow the structure outlined earlier.

Used CNN architecture

The CNN utilized in the keypoint detection-based method is illustrated in Figure 2.3. It begins with a convolutional layer followed by a max pooling layer, which reduces the spatial dimensions. Subsequently, two additional convolutional layers are applied without further spatial reduction, focusing on increasing the number of feature channels and enhancing information extraction. The resulting output is then flattened and fed into a fully connected neural network composed of two layers.

2.3.4 EfficientNet-B0

EfficientNet-B0 was introduced by Mingxing Tan et al. [18] and is a compact and efficient deep learning model designed for image classification, focusing on balancing accuracy and computational cost. It uses compound scaling, a method that scales the model's depth (number of layers), width (number of channels), and input image resolution together. This ensurs the network grows in a balanced way to improve performance without overly burdening the required computation.

The architecture starts with a convolutional layer that extracts basic features from the input image, followed by a series of Mobile Inverted Bottleneck (MBConv) blocks.

These blocks are inspired by MobileNetV2 and consist of:

- Depthwise separable convolutions, which reduce computation by splitting the convolution into two steps: first filtering each channel separately, then combining them,
- Inverted residuals, which expand the number of channels during processing but project them back down to fewer channels after,
- Squeeze-and-Excitation (SE) layers, which improve the network's focus by reweighting features based on their importance.

After passing through multiple MBConv blocks, the network flattens the information from the feature maps into a single vector. This is followed by a fully connected layer with dropout that returns the calculated keypoint.

EfficientNet-B0 stands out by using fewer parameters than other large models, thanks to its careful scaling and efficient block design, making it a great choice for tasks that need both accuracy and speed.

Adam optimizer

The Adam optimizer is a popular and highly effective algorithm for training deep learning models and was developed by Diederik P. et al. [19]. It combines the strengths of two other optimizers: AdaGrad, known for handling sparse gradients well, and RMSProp, which is effective with noisy data. Adaptive Moment Estimation (Adam) adjusts the learning rates for each parameter individually by estimating the first and second moments (mean and variance) of the gradients.

Adam maintains two moving averages: one for the gradients (first moment) and another for the squared gradients (second moment). These moving averages are used to calculate adaptive learning rates for each parameter. A key feature of Adam is its bias correction mechanism, which helps prevent slow convergence at the beginning of training by adjusting the moving averages early on.

Adam is widely preferred in deep learning because of its adaptability and efficiency. It often converges faster than optimizers like Stochastic Gradient Descent (SGD) and is less sensitive to hyperparameters like the learning rate. Additionally, Adam is well-suited for tasks with noisy data or sparse gradients, making it effective in diverse applications such as natural language processing and computer vision.

GradScaler

The `GradScaler` is a utility designed to enhance the stability and efficiency of mixed-precision training, especially on NVIDIA GPUs with Tensor Cores. Mixed-precision training combines 16-bit (`float16`) and 32-bit (`float32`) floating-point operations, reducing memory use and speeding up training. However, using `float16` directly can lead to numerical instability, such as underflow or overflow, with extreme gradient values.

`GradScaler` mitigates this by dynamically adjusting the loss during backpropagation. It scales the loss by a factor before computing gradients, ensuring they remain within a stable range. After the gradients are calculated, they are scaled back by dividing them by the same factor. If the gradients overflow, the scaler reduces the factor for the next iteration to maintain numerical stability.

The key advantage of `GradScaler` is that it enables the speed benefits of mixed-precision training while maintaining accuracy and stability. It is particularly useful for models with widely varying gradient magnitudes, as it preserves the precision of smaller gradients while taking advantage of `float32`'s faster computations [20].

Learning rate

The learning rate controls the size of the steps taken during training when updating the model's weights in response to the calculated gradient of the loss function. It can be fixed for the whole training duration or different functions can be applied that change the learning rate during the training process. One of those functions is `StepLR`.

`StepLR` reduces the learning rate by a fixed factor after a set number of epochs. It relies on two main parameters: step size, which specifies how many epochs to wait before reducing the learning rate, and gamma, which is the multiplier applied to the learning rate. This scheduler can help improve convergence by gradually lowering the learning rate during training. For example, with a step size of 10 and gamma of 0.1, the learning rate will drop to 10 % of its current value every 10 epochs.

2.4 Datasets

2.4.1 Segmentation

For the task of segmentation, the Montgomery County CXR Set and the Shenzhen Hospital CXR Set with lung segmentation masks were used [21]. Annotations were probided by Rajaraman et al. (2021) [22]

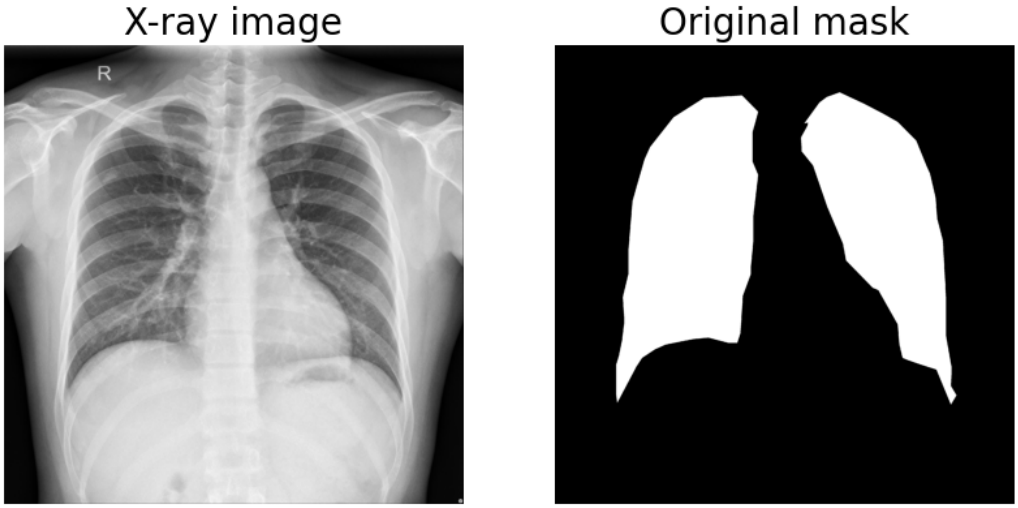

Figure 2.4: Example of the Shenzhen Hospital CXR dataset; Left: X-ray image; Right: correspondent binary lung segmentation mask.

The Montgomery County CXR Set includes 138 chest X-ray images obtained from the TB Control Program at the Department of Health and Human Services in Montgomery County, Maryland, USA. In this collection, 80 images are labeled as normal and 58 show tuberculosis-related abnormalities. All data has been stripped of any identifying information and features left and right lung masks in PNG format. Additionally, the dataset provides resized images with a resolution of 1024 × 1024, annotations agreed upon by two radiologists, and associated radiology readings.

The Shenzhen Hospital CXR Set is comprised of chest X-rays from Shenzhen No.3 Hospital in Guangdong, China. There are 326 normal images and 336 abnormal images, each exhibiting tuberculosis-consistent signs. The dataset, in PNG format, also provides images with a resolution of 1024 × 1024, annotations, and radiology readings for a subset of 68 resized images.

2.4.2 Keypoint detection

The images provided in the dataset, by the University Hospital Ulm, have been acquired during regular patient examinations. The set contains 498 CXRs of which 259 are taken while the patient was standing, while the remaining 259 are taken from laying positions. The data was provided in DICOM format. Separate CSV files where provided containing keypoints annotations describing the costophrenic angle, T1 and T12/L1 vertebrae, and the coordinates describing the minimal field size (minFS) and the tolerance window (maxFS).

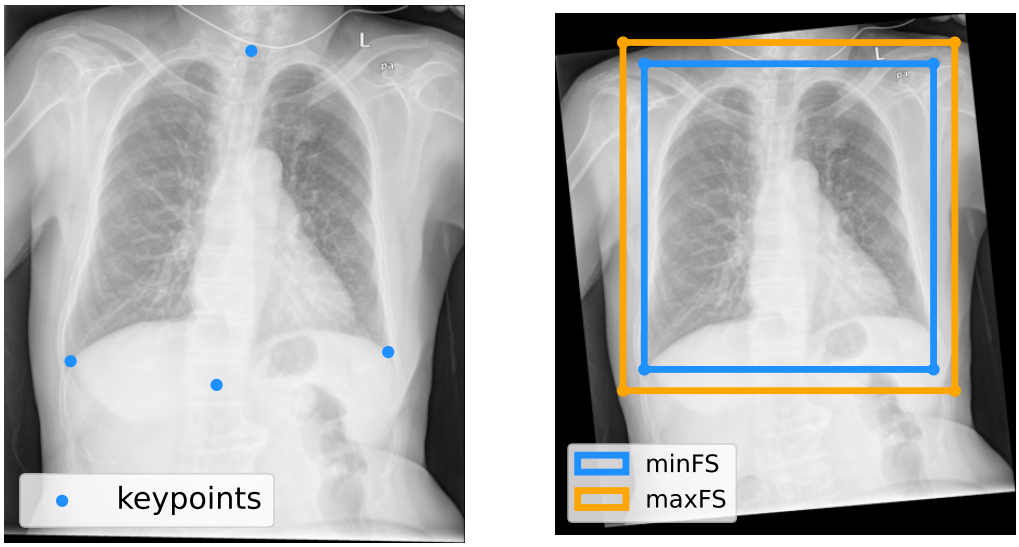

Figure 2.5: Representational example of keypoint detection dataset; Left: X-ray image with annotated keypoints, T1, T12 vertebrae and both costophrenic angle; Right: X-ray image with minimum field size and maximum field size described in the European commission guidelines.

Figure 2.5 shows an exemplary image and keypoints, as images from the original dataset cannot be shown due to copyright and patient confidentiality reasons.

2.4.3 Simulation

X-ray images of an anthropomorphic phantom were take in different dose intensities. With an an additional copper (Cu) filter of 0.3 mm and an aluminum block with 4 cm thickness placed into the beam of radiation, radiation was further decreased. These

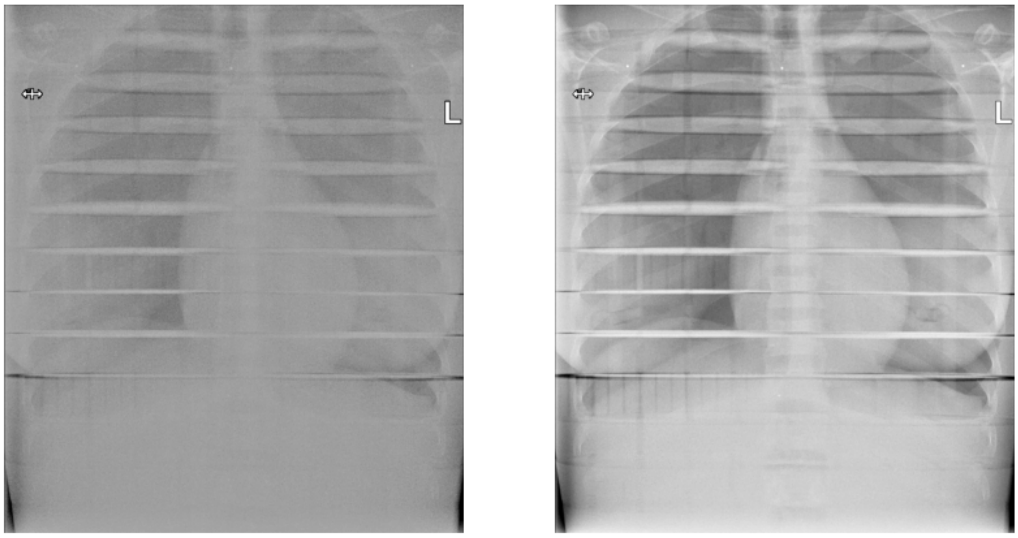

Figure 2.6: Example of phantom X-ray images; Left: X-ray image with 760 μAs; Right: image with 2920 μAs.

images were used for the analysis of noise and image quality degradation. The set consists of 18 images with a radiation exposure from 650 up to 4120 microampere-seconds (μAs). Two examples can be seen in Figure 2.6. The quantity ampere-seconds in X-ray imaging is referred to as the exposure or radiation exposure. It measures the total amount of X-ray radiation generated during an exposure, directly influencing the image quality and the radiation dose received by the patient. A standard exposure for a chest X-ray ranges from 2 to 3 mAs.

2.5 Data analysis techniques

2.5.1 Segmentation metrics

For the analysis, multiple metrics were deployed for better evaluation of the robustness of the models.

In segmentation tasks, the performance of an algorithm can be evaluated using several accuracy metrics that reflect how well the predicted segmentation matches the

ground truth. The following sections describe the common metrics used: overall accuracy, Dice Similarity Coefficient, Hausdorff distance at the 95th percentile, an adapted Intersection over Union metric for the resulting crop box and the Over- and Underreach of the resulting crop.

Accuracy

The accuracy of a segmentation algorithm is the ratio between the correctly classified pixel (or voxel) and the total number of pixel.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{2.4}$$

where:

- TP (True Positives): Pixel correctly identified as part of the object (e.g., lung),
- TN (True Negatives): Pixel correctly identified as not part of the object,
- FP (False Positives): Pixel incorrectly identified as part of the object,
- FN (False Negatives): Pixel incorrectly identified as background.

A visualization of these definitions can be seen in Table 2.1

		Prediction value	
		True	**False**
Ground truth	**True**	True Positive	False Negative
	False	False Positive	True Negative

Table 2.1: Confusion matrix visualizing definition of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN).

While accuracy provides a general indication of performance, it can be misleading when the data is imbalanced (i.e., when the object of interest occupies a small region

relative to the background). For a more nuanced evaluation, additional metrics like the Dice score and Hausdorff distance are typically used.

Dice similarity coefficient (Dice Score)

The Dice Similarity Coefficient (Dice Score) is a widely used metric for assessing the overlap between the predicted segmentation and the ground truth. It is defined as:

$$\text{Dice} = \frac{2 \cdot |X \cap Y|}{|X| + |Y|} \tag{2.5}$$

where:

- X is the set of pixel (or voxel) in the predicted segmentation,
- Y is the set of pixel (or voxel) in the ground truth segmentation,
- $|X \cap Y|$ is the number of pixel common to both X and Y (true positives & true negatives).

The Dice score ranges from 0 to 1, with:

- Dice $= 1$ indicating perfect overlap between the prediction and ground truth.
- Dice $= 0$ indicating no overlap at all.

This metric is sensitive to both over-segmentation and under-segmentation, making it particularly useful for evaluating medical image segmentation.

Hausdorff distance (95th percentile)

The Hausdorff distance (HD) is a metric that measures the distance between the boundaries of the predicted segmentation and the ground truth. Specifically, the Hausdorff distance quantifies the maximum distance between any point on the predicted boundary and the closest point on the ground truth boundary, and vice versa. It is defined as:

$$\text{HD}(X, Y) = \max(h(X, Y), h(Y, X)) \tag{2.6}$$

where:

$$h(X, Y) = \max_{x \in X} \min_{y \in Y} d(x, y) \tag{2.7}$$

and:

- X and Y are the sets of boundary points of the predicted segmentation and the ground truth, respectively,
- $d(x, y)$ is the Euclidean distance between two points x and y.

However, the Hausdorff distance can be overly sensitive to outliers. Therefore, the Hausdorff distance at the 95th percentile (HD95) is often used to provide a more robust evaluation. This version measures the 95th percentile of the distances between the boundaries, effectively ignoring extreme outliers while capturing the overall boundary discrepancy.

The HD 95 % is computed by ranking the distances between the boundaries in ascending order and selecting the 95th percentile distance, thus offering a less sensitive but still reliable assessment of boundary accuracy.

Box Intersection over Union

Intersection over Union (IoU) is a metric used to evaluate the accuracy of an object detection by measuring the overlap between two regions (e.g. segmentation masks) the predicted region generated by the model and the ground truth region that represents the actual location of the object. IoU is calculated as the ratio of the intersection area to the union area of these two regions, providing a value that ranges from 0 to 1. An IoU of 1 indicates perfect overlap between the predicted and ground truth regions, while an IoU of 0 indicates no overlap at all. By providing a clear and intuitive measure of how well prediction aligns with the ground truth, IoU is a great metric for object detection.

For finding the optimal FOV, we adopted the idea of the Intersection over Union Score, which is similar to the Dice Score, and defined as follows:

$$\mathrm{IoU} = \frac{|X \cap Y|}{|X \cup Y|} \tag{2.8}$$

where:

- X is the set of pixel in the predicted segmentation,
- Y is the set of pixel in the ground truth segmentation.

This metric was applied to the bounding boxes generated by creating the smallest possible boxes that encapsulate the ground truth and the predicted data points re-

spectively. This metric was applied to the bounding boxes that resulted from the smallest possible bounding box around the given and predicted data points, such that all data points are included in the box. This represents the FOV derived from the data, giving a direct measure of how acceptable the resulting FOV would be.

Overreach and underreach in field-of-view detection

For the quest of predicting the optimal FOV, it is also important to look at the resulting overreach and underreach. In the context of medical imaging, particularly in optimal FOV detection, the terms overreach and underreach describe how well the detected regions captures the anatomical area of interest.

Overreach occurs when the detected or predicted FOV includes not only the region of interest (e.g., the lungs in thoracic X-rays), but also irrelevant surrounding areas. This can result in the inclusion of non-diagnostic regions and can potentially obscure important details.

Overreach can be quantified as the proportion of pixel or area within the predicted FOV that extend beyond the optimal FOV. Figure 2.7 gives a visualization of the areas for better understanding.

$$\text{Overreach} = \frac{\text{Area of False Positives}}{\text{Total Area of Predicted FOV}} \tag{2.9}$$

Underreach refers to the failure of the predicted FOV to fully capture the region of interest, thus excluding diagnostically relevant areas. This may lead to missed pathological regions and decreased diagnostic accuracy.

Underreach is measured as the proportion of pixel or area from the optimal FOV that are excluded from the predicted FOV. It can be described as the ratio of false negatives to the optimal FOV area:

$$\text{Underreach} = \frac{\text{Area of False Negatives}}{\text{Total Area of optimal FOV}} \tag{2.10}$$

A high underreach value indicates that important areas of the anatomical region are missing from the predicted FOV, which could lead to the omission of critical diagnostic information.

The goal of FOV detection is to minimize both overreach and underreach. Achieving a perfect match between the predicted FOV and the ground truth minimizes the

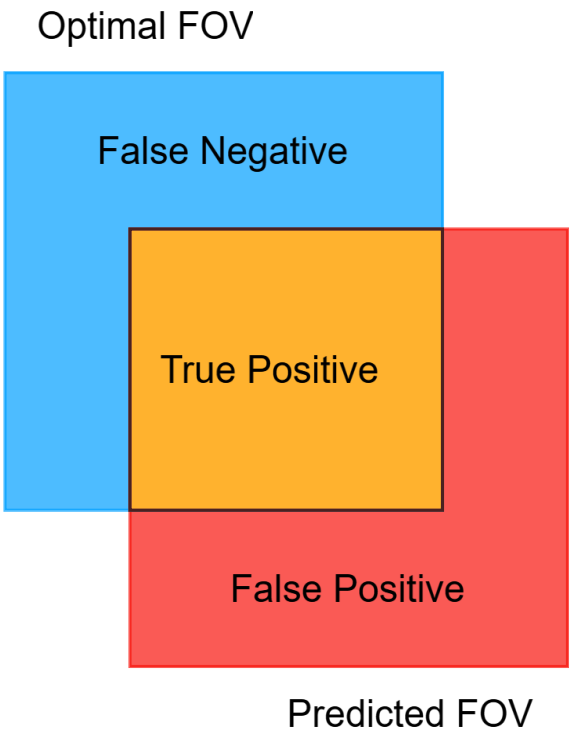

Figure 2.7: Visualization for components of overreach and underreach.

inclusion of irrelevant areas (overreach) and the exclusion of important regions (underreach).

Exposed area reduction

Radiation exposure from a X-ray or CT examination is typically measured in units of effective dose, which quantifies the overall risk of exposure to ionizing radiation. The effective dose is expressed in millisieverts (mSv) and accounts for both the amount of radiation absorbed by the body and the sensitivity of different tissues to radiation.

In practice, instead of manually calculating effective dose for each X-ray, radiologists and technicians often refer to standardized values based on typical X-ray protocols. For a standard chest X-ray, the effective dose is usually around 0.1 mSv, which is considered very low compared to other medical imaging techniques like CT scans, where the effective dose can be up to 7 mSv for a chest CT according to [23]. These

exposures can be compared to the average background radiation one is exposed to due to natural sources. For example the average effective dose received by a person in a year is 2.4 mSv, according to the International Atomic Energy Agency (IAEA) [24].

Calculating the reduction in exposure using an optimal field-of-view is more complex than just considering the reduction in area or pixel values from the X-ray images. Therefore, in this thesis, the area reduction percentage is estimated by comparing the areas of the original image and the predicted view using:

$$\text{reduction} = 1 - \frac{\text{predicted FoV}}{\text{original FoV}} \tag{2.11}$$

2.5.2 Loss metrics

BCEWithLogitsLoss

The BCEWithLogitsLoss is a commonly used loss function in binary classification tasks, especially in machine learning models such as neural networks. It combines a Sigmoid activation function and the Binary Cross-Entropy (BCE) loss in a numerically stable way. This is useful because instead of applying the Sigmoid function separately and then passing the output to the BCE loss, BCEWithLogitsLoss performs these operations in one step, providing better numerical stability and reducing the risk of issues like vanishing gradients.

Binary Cross-Entropy Loss is used to measure the difference between the predicted probability and the actual label (ground truth). For a single binary classification, it is defined as:

$$\text{BCE}(p, y) = -\left[y \cdot \log(p) + (1 - y) \cdot \log(1 - p)\right] \tag{2.12}$$

where:

- p is the predicted probability for the positive class (output of the Sigmoid function),
- y is the ground truth label (1 for positive class, 0 for negative class).

The Sigmoid activation function was applied to the raw logits (unscaled values predicted by the model) to transform them into a probability value between 0 and 1. The

Sigmoid function is defined as:

$$\sigma(z) = \frac{1}{1 + e^{-z}} \tag{2.13}$$

where z is the raw logit (the unscaled model output).

The BCEWithLogitsLoss combines the Sigmoid function and BCE loss in a more stable form. Instead of applying the Sigmoid function directly to the logits and then computing the Binary Cross-Entropy loss, this loss function performs both steps in one operation. The resulting loss function is:

$$\text{BCEWithLogitsLoss}(z, y) = \max(z, 0) - z \cdot y + \log\left(1 + e^{-|z|}\right) \tag{2.14}$$

where:

- z is the raw logit (unscaled model output before applying the Sigmoid function),
- y is the ground truth label (1 for positive class, 0 for negative class).

This formulation avoids potential numerical instabilities when calculating the logarithms and exponentials by keeping the intermediate values stable, particularly for large magnitudes of z.

MSELoss

The MSELoss (Mean Squared Error Loss) is a commonly used loss function for regression tasks, where the goal is to predict continuous values. It measures the average of the squared differences between the predicted values and the actual target values. The MSELoss is simple to compute and provides a clear measure of how far off the predictions are from the true values.

The Mean Squared Error (MSE) loss is defined as:

$$\text{MSE}(y_{\text{pred}}, y_{\text{true}}) = \frac{1}{n} \sum_{i=1}^{n} \left(y_{\text{pred}}^{(i)} - y_{\text{true}}^{(i)}\right)^2 \tag{2.15}$$

where:

- $y_{\text{pred}}^{(i)}$ is the predicted value for the i-th data point,
- $y_{\text{true}}^{(i)}$ is the ground truth (target) value for the i-th data point,

- n is the number of data points.

The MSELoss measures the overall deviation between the predicted values and the true values. A lower value of MSE indicates that the predicted values are close to the true values, while a higher value indicates a larger error.

2.6 Tools and technologies used

Python version 3.10.12 was utilized for the analysis and model programming [25], particularly for training and evaluation, due to familiarity with this language from previous courses like the mandatory Data Lab curriculum and the lecture on 3D Deep Learning, which provided hands-on experience with PyTorch. This included tasks such as data preprocessing, dataset creation, model construction, training, evaluation, and hyperparameter tuning. Torch version 2.1.0 was employed for the implementation [26], along with TorchMetrics 1.4.0 for metric functions applicable to TorchTensors. The Medical Open Network for Artificial Intelligence (MONAI) U-Net was loaded via the monai 1.3.1 package and the EfficientNet via TorchVision 0.16.0. To speed up the process, Cuda 12.4 was used to execute operations simultaneously, leveraging the NVIDIA GeForce RTX 3060 GPU's parallel architecture to significantly speed up computations.

Data augmentation and pre-processing were handled using TorchVision and Albumentations 1.4.12, which allowed easy augmentation of both images and corresponding masks or keypoints with identical transformations. To ensure Albumentation's functionality, the PIL package was used for loading image files, particularly for segmentation tasks. The keypoints were stored in a CSV file that was loaded and keypoints extracted with pandas 2.2.2.

For the creation of architecture diagrams, draw.io Diagrams & Whiteboards 24.8.0 was used and for the creation of all figures, matplotlib 3.9.0 and seaborn 0.13.2 were used.

The code used in this thesis can be found here: `https://github.com/Turadon/Master-thesis`

3 Methods and Experiments

3.1 Simulation

The simulation step is a very important part of this thesis, as model results will only be meaningful if the simulation is representational of real low-dose images. First, an anthropomorphic phantom was imaged with different doses. The dataset was described in Section 2.4.3 and example images can be seen there in Figure 2.6.

The phantom X-ray images were analyzed through visualization techniques and the distribution of their gray-scale pixel values was plotted. The image contrast, using the mean and standard deviation (std) of each distribution was assessed. Examples can be seen in Figure 3.1

Histogram matching is an image processing method used to modify the contrast of an image to resemble that of a reference image. This approach was applied to replicate the contrast change observed at a lower-dose by matching the original standard-dose image with a lower-dose phantom X-ray, as demonstrated in Figure 3.2.

In the analysis of Gaussian noise, homogeneous areas within the phantom images were examined, and the mean pixel values and standard deviation were calculated. These values served as approximations for the standard deviation in the Gaussian noise generation process. As a result, different noise parameters were derived for each reference image. Figure 3.3 provides two example regions for this noise evaluation.

The first tests were conducted with the reference image with the lowest dose of 650 μAs, which was captured using the lowest possible settings in the imaging software. An added 0.3 mm Cu-filter and an extra aluminum block with approximately 4 cm thickness was placed directly behind the tube into the path of radiation. For reference, a standard-dose image has a approximate dose between 2-3 mAs. The addition of the aluminum block reduces the effective dose by roughly 60 %, additionally the addition of the Cu-filter and reduction to 650 μAs roughly reduces radiation by one third. This results in a approximate effective dose of 0.013 mSv.

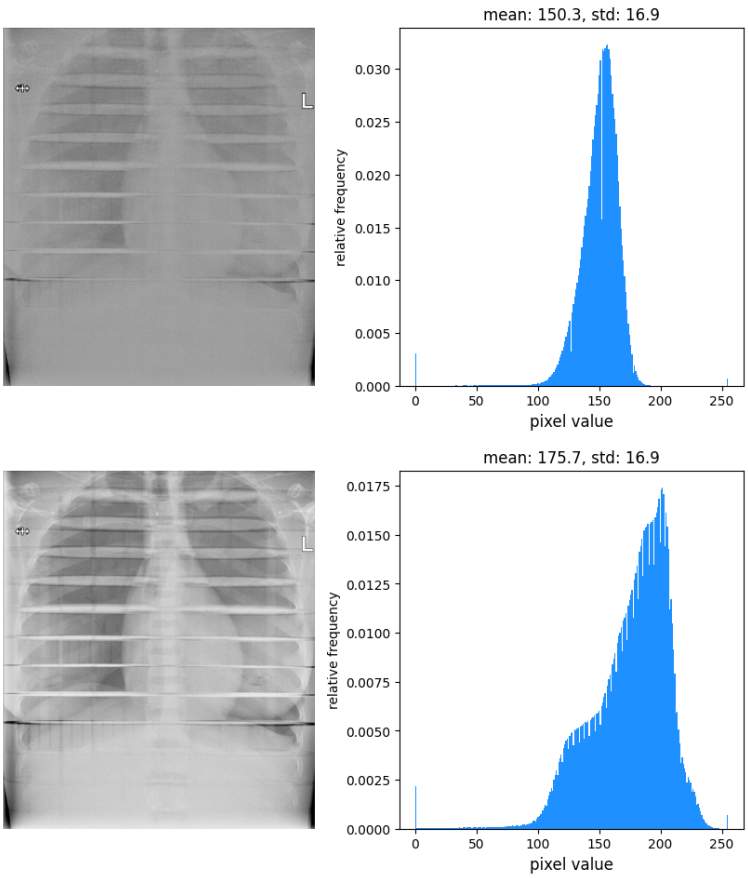

Figure 3.1: Example of phantom X-ray images with histograms; Left: X-ray images; Right: histogram of gray pixel values corresponding to images on the left with mean and std.

To further explore dose reduction, a statistical analysis of the phantom images was performed to estimate potential lower-dose references. This can be seen in Figure 3.4. For histogram matching, only the pixel value distribution mattered, not the visible content of the image. As the phantom image histograms tend to follow a normal distribution at lower-doses (as seen in Figure 3.1), a generated image with a normal distribution was used as a reference for histogram matching. For even lower-dose simulations, different Gaussian noise intensities were added.

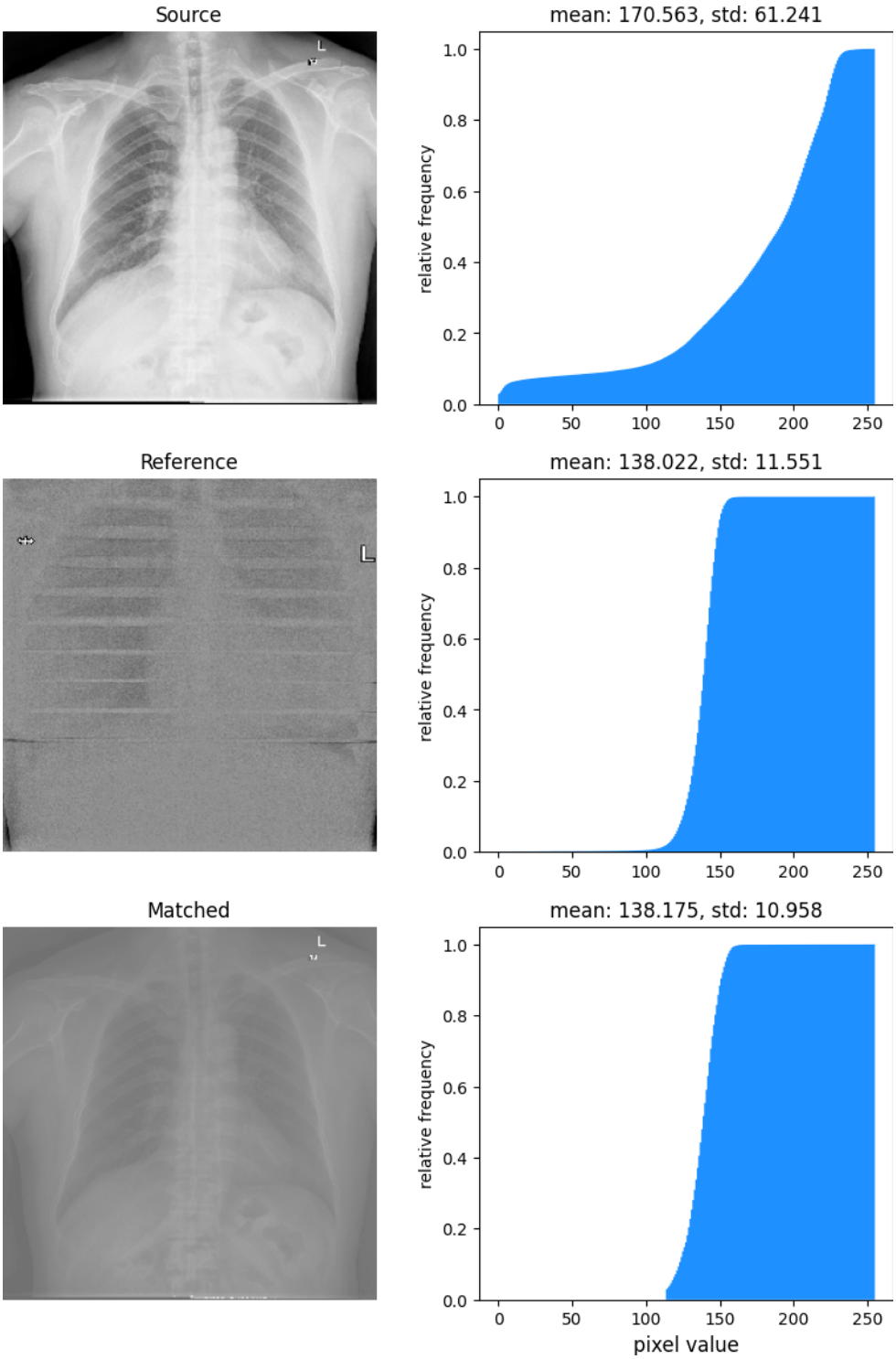

Figure 3.2: Example of histogram matching with original X-ray Image, phantom X-ray images as reference and matched image on the left. On the right are the corresponding histograms.

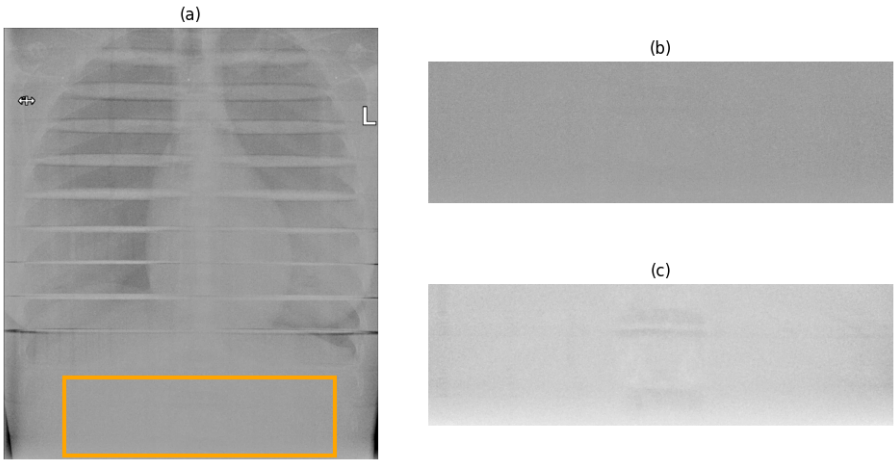

Figure 3.3: (a) Phantom X-ray image with 760 µAs with highlighted selected area for noise analysis; (b) Selected area for noise analysis of image (a); (c) Selected area for noise analysis of another image with 2920 µAs.

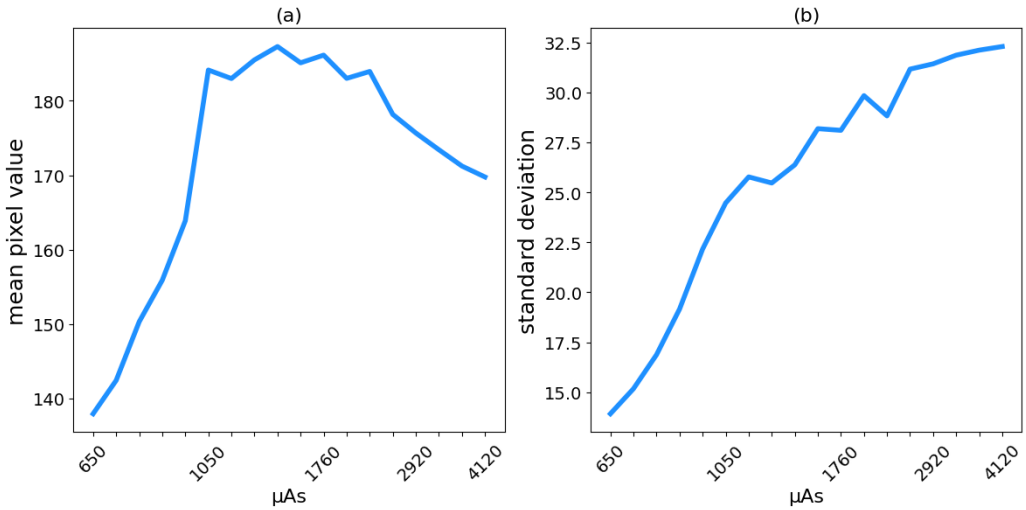

Figure 3.4: (a) Mean pixel value of gray-scale phantom X-ray images of dataset 2.4.3, the x-Axis represents the 18 images with their exposure in µAs;
(b) Standard deviation of phantom X-ray images.

3.2 Segmentation

Data Loader

For the task of segmentation, the dataset from Section 2.4.1 was used with a train-test-validation split of 70 % - 15 % - 15 %. A data loader was created to load and augmentate the X-ray images with its corresponding segmentation mask. For data augmentation, the training data was resized, black padding was added around the edges, and a random rotation between -15° and $+15^\circ$ was applied with a 50 $\%$ probability. This is demonstrated in Figure 3.5.

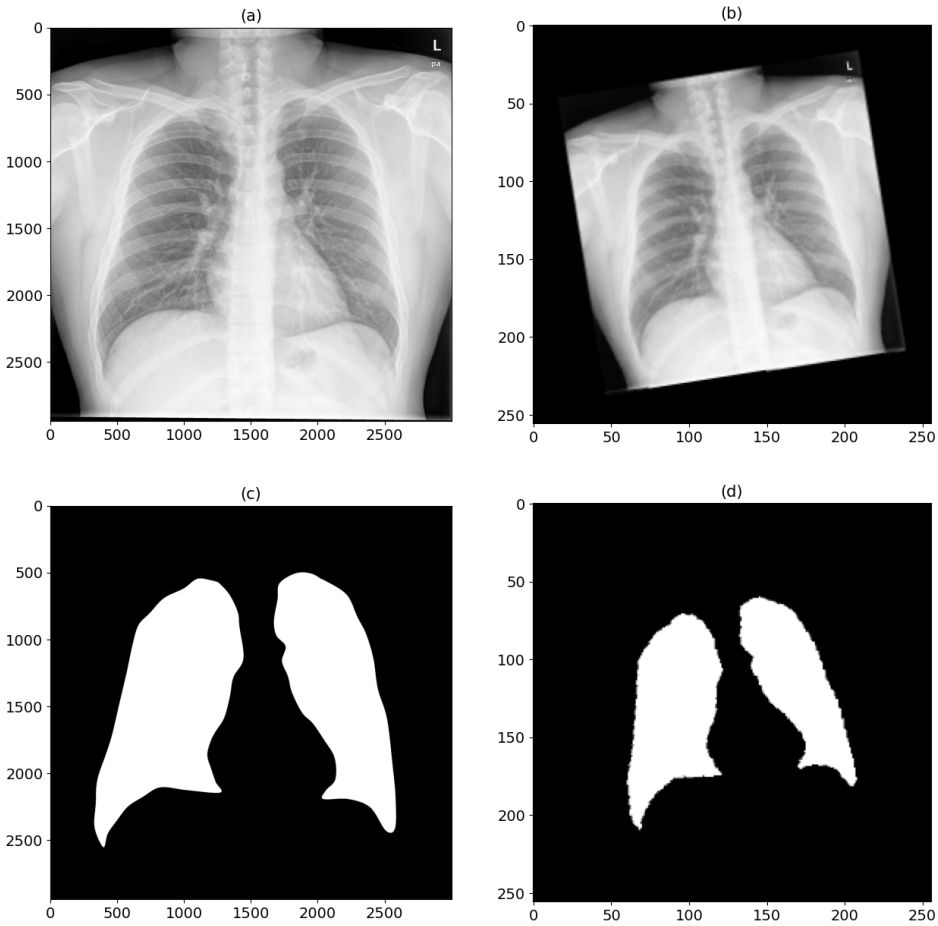

Figure 3.5: (a) Original X-ray image; (b) Augmented X-ray image; (c) Original segmentation mask; (d) Augmented segmentation mask

Padding was used to address segmentation issues, especially for images where the lungs and their masks were near the edge, which made it harder to accurately define the boundary between the lungs and surrounding tissue. Furthermore, the images were scaled to the interval from 0 to 1 to enhance the learning process.

Training

In this thesis, two different U-Net models were compared. Firstly a basic U-Net model was built as described in Section 2.3.3, and secondly the U-Net from the MONAI package, built for medical segmentation tasks, was used [27]. To be able to compare the two different models, the package weights and biases was utilized to track the training process and log their input variables and configurations. This made hyper-parameter search easier to analyse and configurations easier to keep track of. The `BCEWithLogitsLoss` loss function was used, and for the backpropagation the Adam optimizer and the `GradScaler` were applied to enhance training.

The models were trained with the standard-dose images in order to compare them under ideal conditions and to perform hyperparameter search over a varying number of feature channel reductions. For the convolutional feature extraction of the U-Net, 4 convolutional blocks were used. The following different sizes of feature channel expansions were tested: [4, 8, 16, 32], [8, 16, 32, 64] and [16, 32, 64, 128] and image sizes of: 128 x 128, 240 x 240 ,and 256 x 256 resolutions were examined. Normally image sizes of powers of 2 are used in convolution because the max pooling layer reduces the image width and height by a factor of 2. The 240 x 240 image size was included, as it was the recommended size for the MONAI U-Net.

Training on simulated data

With the best performing configuration of [16, 32, 64, 128] feature channels during the contraction path and an image size of 256 x 256 pixel, the two models were trained on the two different low-dose simulations approaches. To simulate the image with Gaussian noise, the data loader included the phantom reference image of 650 μAs for live histogram matching and data augmentation as described in Section 3.2. The lowest quality image was initially selected as the reference to assess the models performance. For one approach, Gaussian noise was introduced. The noise level was estimated from a heterogeneous area within the reference image, as described in Section 3.1, providing an estimated noise standard deviation. For each pixel, noise was applied by sampling from a normal distribution with a mean of 0 and the estimated

standard deviation, this noise was then added to the pixel values. To simulate photon noise, the matched images were used as inputs for pixel-wise Poisson sampling, yielding simulated pixel values. As before, the images were scaled to the range [0,1].

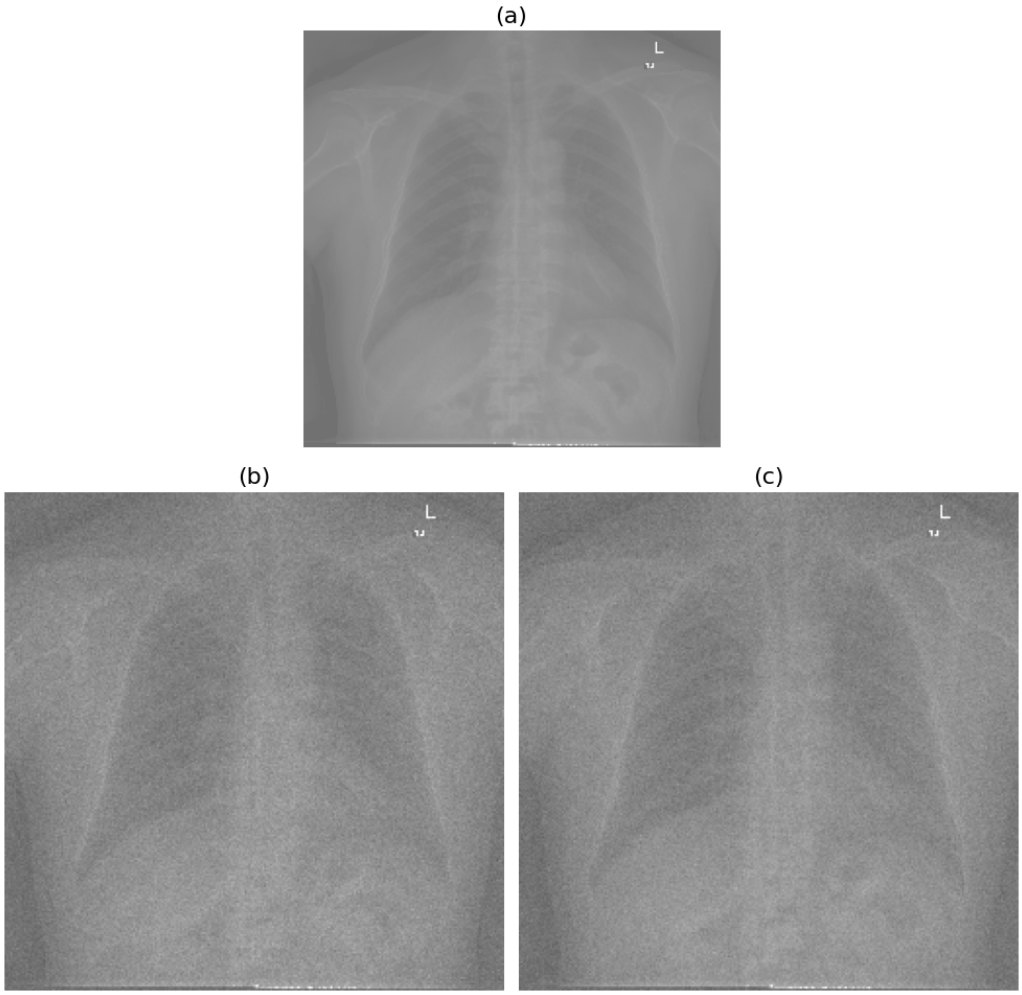

Figure 3.6: (a) X-ray image after histogram matching with reference image of 650 μAs, (b) Poisson simulated image with image (a) as input for sampling; (c) Matched image with added Gaussian noise.

The two different approaches in simulation, Gaussian and Poisson (Figure 3.6 provides an example) were used in training, and to speed up the training process, the model weights from the standard-dose trainings were loaded.

Training on lower-dose estimation

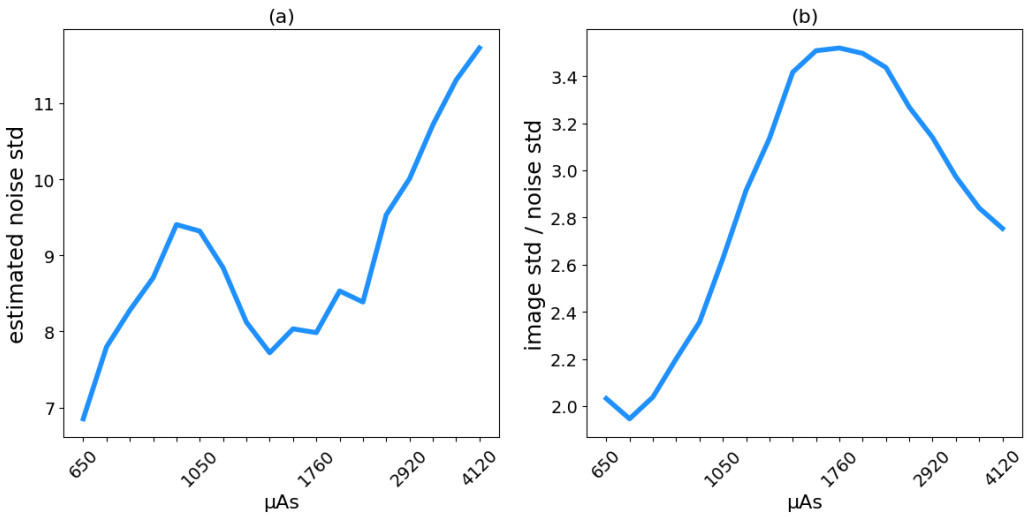

Figure 3.7: (a) Standard deviation of estimated noise as described in 3.1, x axis represents images with increasing exposure in µAs; (b) Ratio between image standard deviation to noise standard deviation.

Further testing of the model's capabilities involved simulating potential lower-dose X-rays by analyzing the trends in contrast and noise from the phantom images. The analysis was extended to doses lower than those achievable with the current software. This was done by plotting the mean and standard deviation, that are visualized in Figure 3.4, and the estimated noise standard deviation, and the ratio between the standard deviation and the estimated noise (noise standard deviation), which can be seen in Figure 3.7. As expected, lower µAs values corresponded to worse image quality.

Using these insights, different image standard deviations and noise ratios were tested to determine how accurately the model could perform with increasingly degraded image quality. Two example simulations are visualized in Figure 3.8.

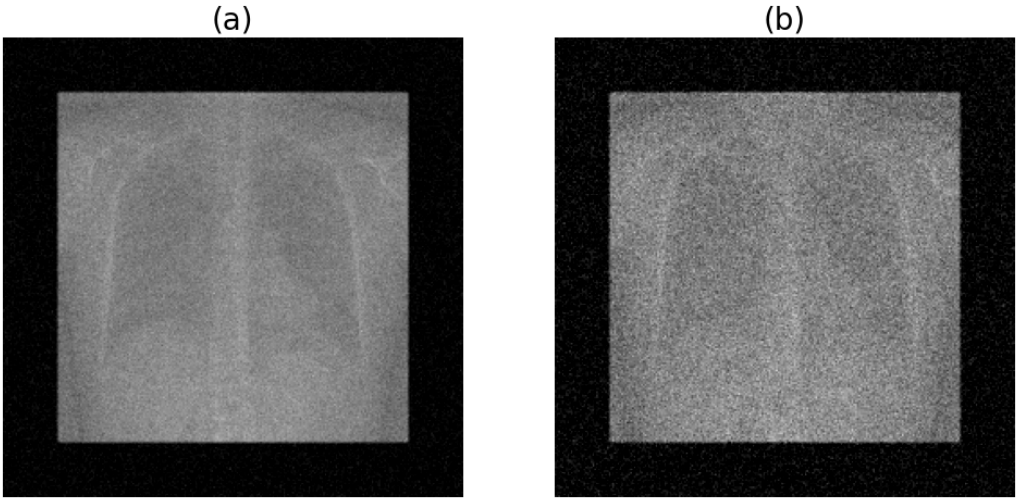

Figure 3.8: (a) Example simulation with image standard deviation of 10 and a standard deviation ratio of 1; (b) Example simulation with image standard deviation of 10 and a standard deviation ratio of 0.5.

3.3 Keypoint detection

Data Loader

For the task of keypoint detection, the dataset from Section 2.4.2 was used, the images in Digital Imaging and Communications in Medicine (DICOM) format were loaded, and the images with their corresponding pixel interval were extracted and saved in PNG format for easier loading during the training process. The keypoints were provided in a separate CSV file, and the corresponding X-ray images were connected to their keypoints by their DICOM filename. The new paths were saved in a separate column, such that the CSV file could be used to manage all the necessary information. A train-test-validation split of 70 % - 15 % - 15 % was used and the data loader was created, similar to the segmentation data loader, to load and augmentate the X-ray images with its corresponding keypoints. For data augmentation, the training data was augmented identically to the segmentation-based approach. Here, padding was used to address vanishing keypoints that where close to the image edge and were lost during rotation. An example of this can be seen in Figure 3.10.

Additionally, this helped improve accuracy in general for keypoints located close to the images edge. Further, the images were scaled to the interval from 0 to 1 to enhance

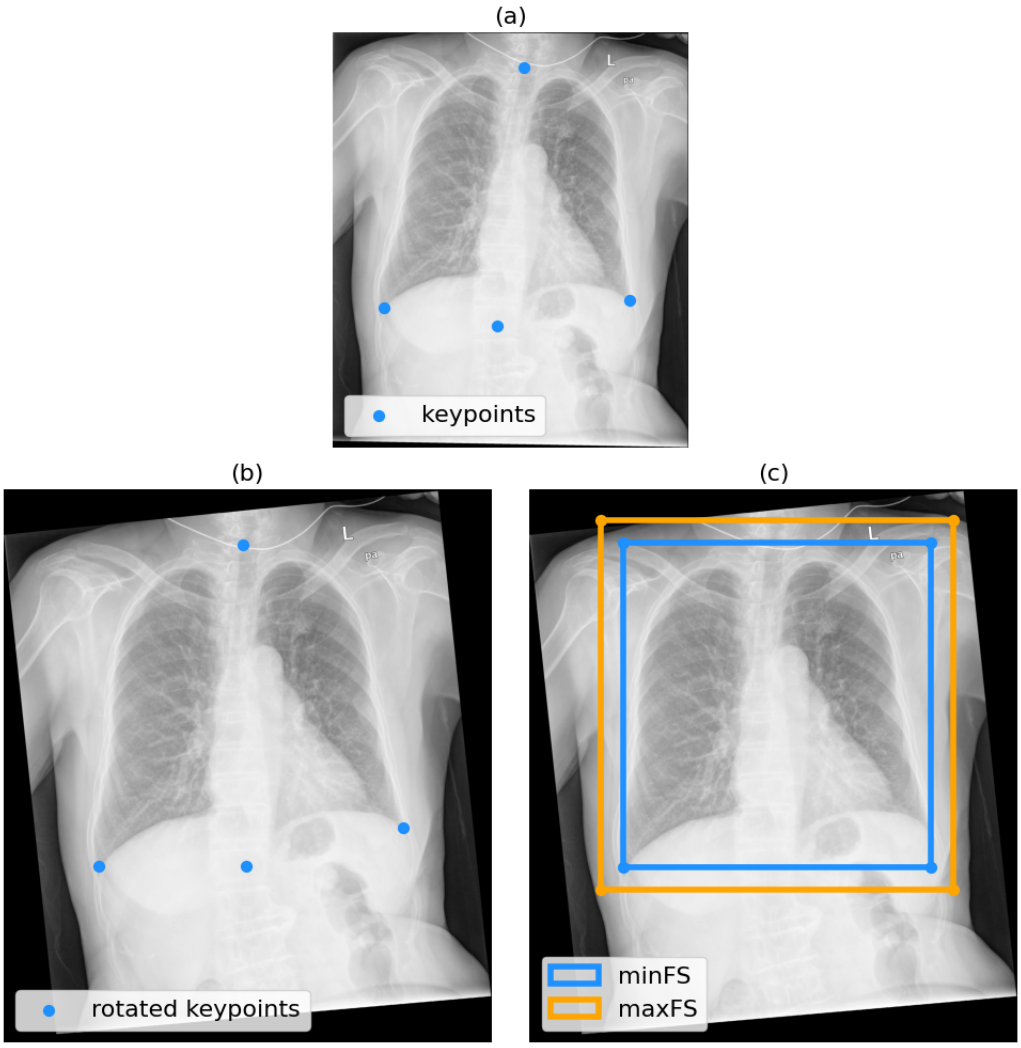

Figure 3.9: (a) X-ray image with annotated keypoints; (b) X-ray image with keypoints rotated such that the spine aligns vertically; (c) X-ray image with minFS and maxFS.

the learning process.

Training

For the keypoint detection-based approach the CNN described in Section 2.3.3 and the EfficientNet-B0 from Section 2.3.4 were deployed. The training process for the

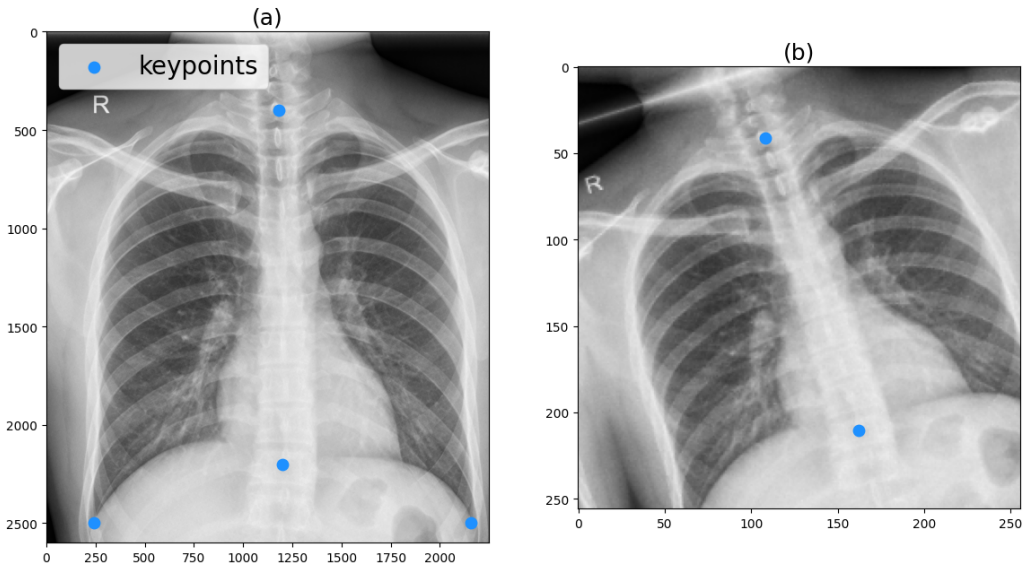

Figure 3.10: This image is made up due to copyright reason but represents difficulties of simple rotation; (a) Example X-ray image with keypoints; (b) Augmented X-ray image, showing that keypoints are lost due to rotation.

keypoint detection-based approach followed a similar methodology to the segmentation-based approach previously described. In the keypoint detection process the `MSELoss` was utilized. During hyperparameter search, the models were not changed like they were during the segmentation-based approach. Here the following learning rates 0.001 and 0.0001 were tested. Another configuration was using the `GradScaler` and the application of learning rate decay with a reduction of 50 % every 5 epochs by the `StepLR` function mentioned in Section 2.3.4 in order to enhance the learning process.

Training on simulated data

With the best performing configuration using a learning rate of 0.001, the `GradScaler` and no learning rate decay function models were trained on the low-dose simulated data as described before in the U-Net Section 3.1.

Fusion of Segmentation and Keypoint detection

An additional brief experiment focused on the combination of the segmentation-based approach with the keypoint detection by the EfficientNet-B0, which takes a 3 channel input. To enhance keypoint detection, the segmentation-based approaches were

used as backbone for the keypoint detection. The backbone provided a predicted segmentation mask that was utilized as input. The input for the EfficientNet-B0 consisted of the X-ray image in the first channel, for the second channel the predicted mask was added pixel-wise to the X-ray image and as third channel the predicted segmentation mask. This configuration was tested on the standard-dose images of the keypoint detection dataset and on the simulated datasets with Gaussian and Poisson simulated noise.

Overreach, underreach and overexposure

At the end the segmentation-based approaches and the EfficientNet-B0-based approach were further analyzed by calculating the overreach, underreach and overexposure of the resulting FOV prediction of the simulated X-ray images. This was done by taking the resulting bounding box around the predicted data. Separately a tolerance was added by extending the calculated bounding box in each direction by 10 pixel values. This addition was motivated by the maxFS of the keypoint detection dataset and the goal was to see by how much this reduces underreach. An example of both bounding boxes around the predicted segmentation can be seen in Figure 3.11

3.4 Challenges and Solutions

DICOM file missing image pixel range

DICOM files store not only medical images but also essential metadata, including patient information (e.g., name, ID), image acquisition details (e.g., date, time, modality), and medical settings. The X-ray image is represented as a pixel array, with values typically ranging from 0 to 4080. The window center and window width are parameters used to adjust the contrast and brightness of the image. These parameters allow radiologists to optimize the display of grayscale pixel values, enhance the visualization of different tissues or structures. Usually, these values are precalculated by imaging software and stored in the DICOM file.

For half of the keypoint dataset, this information was missing. To address this issue, a suitable range for all files with missing data had to be determined. Fortunately, the necessary pixel values were generally in the same range, so a fixed range was applied to extract the X-ray images from the DICOM files. As a result, the contrast

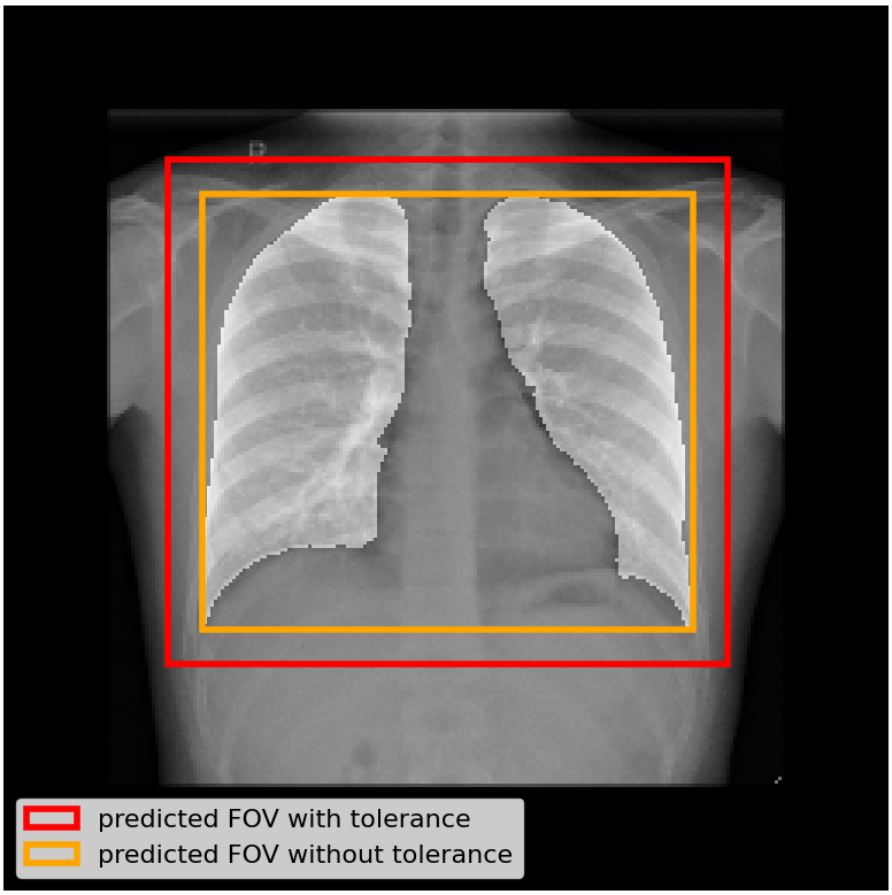

Figure 3.11: Example field-of-view from predicted segmentation mask highlighted in the standard-dose X-ray image. The FOV is presented without added tolerance and with an added tolerance of 10 pixel into each direction.

of the images varied, effectively serving as an augmentation that contributes to the robustness of the approach.

Single pixel influence on FOV detection

During the segmentation-based approach the predicted mask can have some smaller disconnected areas, which is visually represented in Figure 3.12, that influence the resulting bounding box (FOV prediction) around the predicted mask and with this the Box IoU score. To improve the field-of-view prediction, the predicted mask was separated into connected pixel areas. From those areas only the biggest two, the

assumed left and right lung, were used for calculating the bounding box.

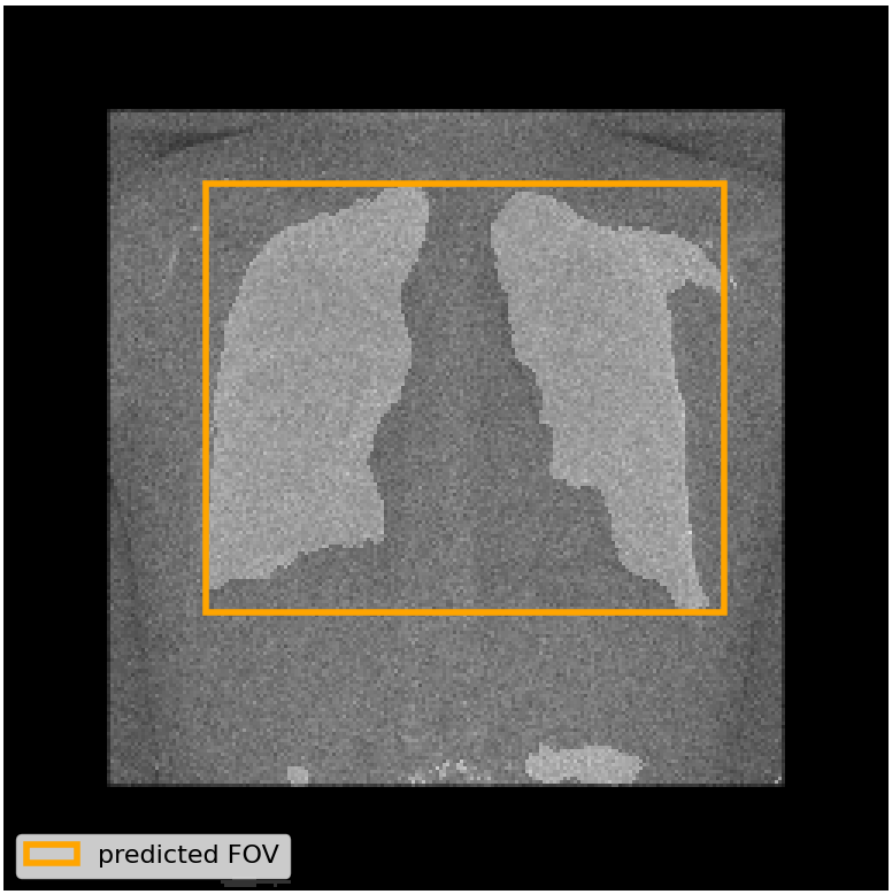

Figure 3.12: Example image that shows an adjusted bounding box. Predicted segmentation was highlighted on top of the simulated X-ray image. The adjustment was achieved by selecting only the two largest segmented areas. This minimizes the impact of smaller detected regions, for examples those that can be seen at the bottom of the image.

4 Results

In this section, for the analysis of the results, each configuration was initiated randomly, while others utilized pretrained states. Each configuration training was performed five times. The mean of the resulting metrics for each configuration was calculated to get an average evaluation.

4.1 Segmentation

Hyperparameter search

The results of the hyperparameter search, as described in Section 3.2, can be seen in Table 4.1 for our U-Net model and in Table 4.2 for the MONAI U-Net. Despite the resulting accuracy of both approaches being similar the Box IoU score and the Dice score were better for our U-Net approach when compared between same configurations. The best results for both approaches were achieved with an image resolution of 256 × 256 pixel and the downsampling feature channel numbers of [16, 32, 64, 128].

Features	Image size	Accuracy	Box IoU
[16, 32, 64, 128]	**256 × 256**	**0.983 ± 0.003**	**0.967 ± 0.02**
[16, 32, 64, 128]	240 × 240	0.983 ± 0.0	0.961 ± 0.003
[16, 32, 64, 128]	128 × 128	0.975 ± 0.0	0.957 ± 0.0
[8, 16, 32, 64]	256 × 256	0.982 ± 0.0	0.956 ± 0.005
[8, 16, 32, 64]	240 × 240	0.982 ± 0.0	0.955 ± 0.006
[8, 16, 32, 64]	128 × 128	0.974 ± 0.0	0.944 ± 0.008
[4, 8, 16, 32]	256 × 256	0.981 ± 0.001	0.950 ± 0.006
[4, 8, 16, 32]	240 × 240	0.980 ± 0.001	0.946 ± 0.005
[4, 8, 16, 32]	128 × 128	0.972 ± 0.0	0.938 ± 0.004

Table 4.1: Our U-Net model hyperparameter search metrics.

Features	Image size	Accuracy	Box IoU
[16, 32, 64, 128]	**256 × 256**	**0.980 ± 0.001**	**0.951 ± 0.004**
[16, 32, 64, 128]	240 × 240	0.979 ± 0.001	0.950 ± 0.003
[16, 32, 64, 128]	128 × 128	0.973 ± 0.003	0.945 ± 0.002
[8, 16, 32, 64]	256 × 256	0.975 ± 0.006	0.932 ± 0.002
[8, 16, 32, 64]	240 × 240	0.976 ± 0.006	0.934 ± 0.004
[8, 16, 32, 64]	128 × 128	0.970 ± 0.0	0.931 ± 0.003
[4, 8, 16, 32]	256 × 256	0.967 ± 0.003	0.895 ± 0.006
[4, 8, 16, 32]	240 × 240	0.969 ± 0.001	0.897 ± 0.011
[4, 8, 16, 32]	128 × 128	0.965 ± 0.003	0.912 ± 0.007

Table 4.2: MONAI U-Net hyperparameter search metrics.

Segmentation-based approach on simulated low-dose images

The best configuration, utilizing downsampling of [16, 32, 64, 128] and a resolution of 256 × 256 pixel, was determined by the above hyperparameter search. The approaches were trained on the Gaussian and Poisson simulated images from Section 3.1. Both models performed better on the Gaussian data with our U-Net model outperforming the MONAI U-Net on both simulation types with a higher Box IoU score.

Simulation type	Model	Accuracy	Box IoU	HD95
Gaussian	**Our U-Net**	**0.977 ± 0.002**	**0.914 ± 0.02**	**7.307 ± 0. 442**
Gaussian	MONAI U-Net	0.972 ± 0.001	0.851 ± 0.001	14.080 ± 0.877
Poisson	Our U-Net	0.971 ± 0.013	0.892 ± 0.062	9.029± 3.939
Poisson	MONAI U-Net	0.969 ± 0.002	0.841 ± 0.01	14.562 ± 1.16

Table 4.3: Segmentation-based metrics for low-dose X-ray simulations.

Estimated lower-doses simulation

To further test the capabilities of the two approaches, references with an even lower-dose were estimated, and stronger Gaussian noise was added, by increasing the noise standard deviation. Two example lower-dose simulations can be seen in the last chapter in Figure 3.8. The resulting accuracy and Box IoU score for our U-Net approach can be seen in Figure 4.1 and Figure 4.2 respectively. The accuracy decreased with stronger noise and the Box IoU score decreased even stronger.

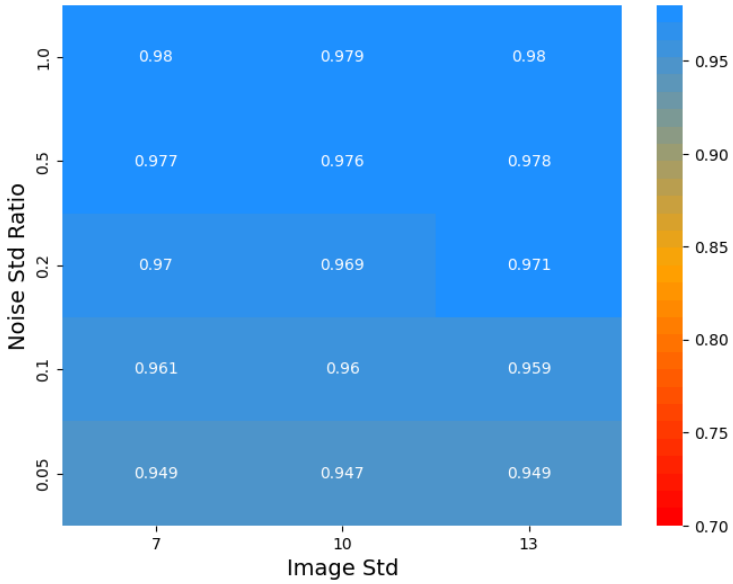

Figure 4.1: Heatmap presenting accuracy comparing image standard deviation against noise standard deviation ratio in our U-Net-based approach.

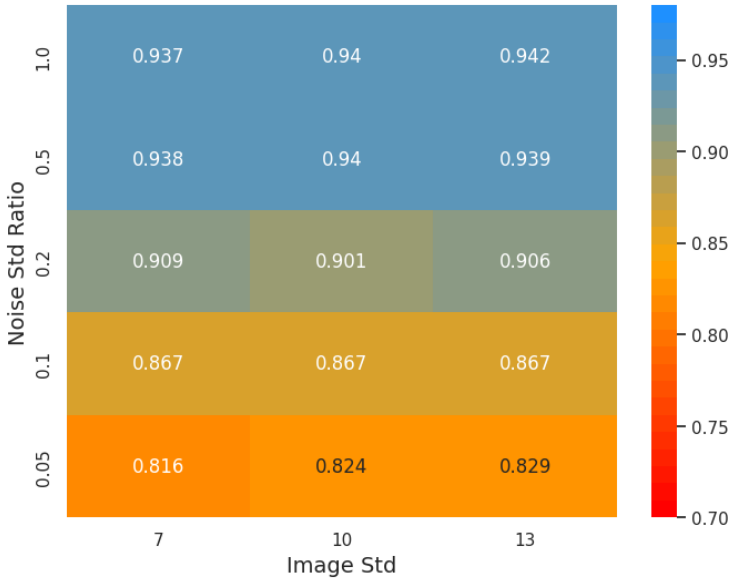

Figure 4.2: Heatmap presenting Box IoU comparing image standard deviation against noise standard deviation ratio in our U-Net-based approach.

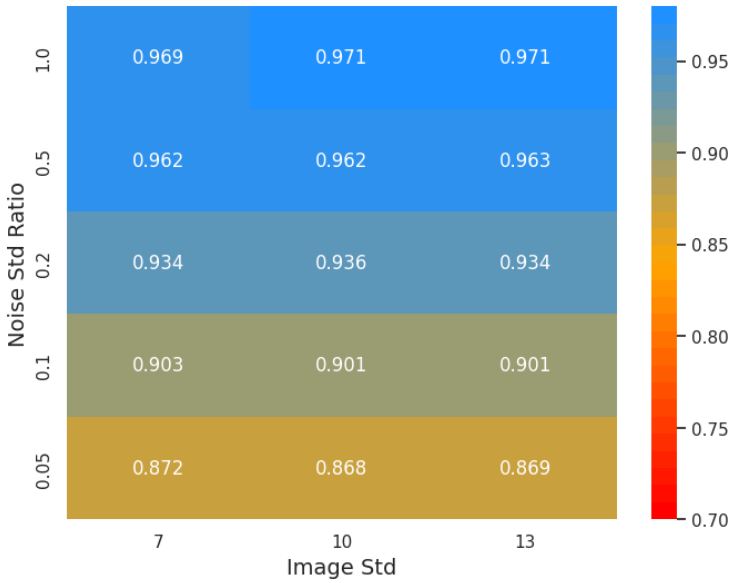

Figure 4.3: Heatmap presenting accuracy comparing image standard deviation against noise standard deviation ratio in MONAI U-Net-based approach.

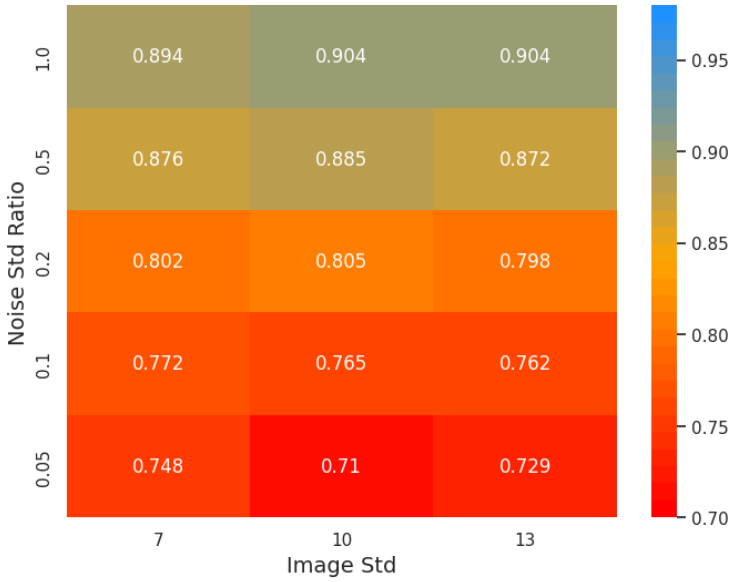

Figure 4.4: Heatmap presenting Box IoU comparing image standard deviation against noise standard deviation ratio in MONAI U-Net-based approach.

The results for the MONAI U-Net can be seen in Figure 4.3 and Figure 4.4. As before, the MONAI U-Net performed worse than our U-Net approach because both metrics decreased stronger with an increase in noise intensity.

4.2 Keypoint detection

Hyperparameter search

The results of the hyperparameter search as described in Section 3.3 can be seen in Table 4.4 for our CNN model-based approach and in Table 4.5 for the EfficientNet-B0-based approach. The EfficientNet-B0 performed much better than our CNN model in all regards. It can be seen that a constant learning rate of 0.001 performed better than an initial lower learning rate or a reduction of learning rate during training. Additionally the use of the `GradScaler` increased performance too.

Learning rate	`StepLR`	`GradScaler`	Box IoU
0.001	**No**	**Yes**	**0.775 ± 0.015**
0.001	Yes	Yes	0.775 ± 0.0
0.0001	No	Yes	0.775 ± 0.0
0.0001	Yes	Yes	0.775 ± 0.0
0.0001	No	No	0.773 ± 0.003
0.0001	Yes	No	0.772 ± 0.006
0.001	Yes	No	0.761 ± 0.01
0.001	No	No	0.759 ± 0.03

Table 4.4: Our CNN-based hyperparameter search metrics.

Keypoint detection-approach on simulated low-dose images

The two approaches were trained on the Gaussian and Poisson simulated images from Section 3.1 with a learning rate of 0.001 and the use of the `GradScaler`. The resulting Box IoU score can be seen in Table 4.6. Here the EfficientNet-B0-based approach also performed much better than our CNN and was slightly better for Poisson simulated images than Gaussian simulated images. For this reason only the EfficientNet-B0-based approach was tested on the estimated lower-doses.

Learning rate	`StepLR`	`GradScaler`	Box IoU
0.001	**No**	**Yes**	**0.901 ± 0.006**
0.001	Yes	Yes	0.893 ± 0.003
0.001	Yes	No	0.893 ± 0.003
0.001	No	No	0.891 ± 0.009
0.0001	No	No	0.885 ± 0.006
0.0001	No	Yes	0.882 ± 0.003
0.0001	Yes	No	0.845 ± 0.004
0.0001	Yes	Yes	0.843 ± 0.001

Table 4.5: EfficientNet-B0-based hyperparameter search metrics.

Simulation type	Model name	Box IoU
Poisson	**EfficientNet-B0**	**0.903 ± 0.007**
Poisson	Our CNN	0.798 ± 0.022
Gaussian	EfficientNet-B0	0.893 ± 0.009
Gaussian	Our CNN	0.833 ± 0.012

Table 4.6: Keypoint detection-based metrics for low-dose X-ray simulation.

Keypoint detection with continued simulated images

As for the segmentation-based approach lower-doses were estimated and increasing Gaussian noise was added after histogram matching. The resulting Box IoU score for the EfficientNet-B0 can be seen in Figure 4.5.

Fusion segmentation in keypoint detection

The fusion of segmentation- and keypoint detection-based approaches described in Section 3.3, was tested on the standard-dose images and on the low-dose simulated images with both segmentation approaches as backbone. The results for the standard-dose can be seen in Table 4.7 and the low-dose simulation results for Gaussian and Poisson simulation can be seen in Table 4.8. Both segmentation approaches performed roughly the same, and the addition of the segmentation-based approaches performed overall worse than the EfficientNet-B0 alone on only the X-ray images without the addition of segmentation masks.

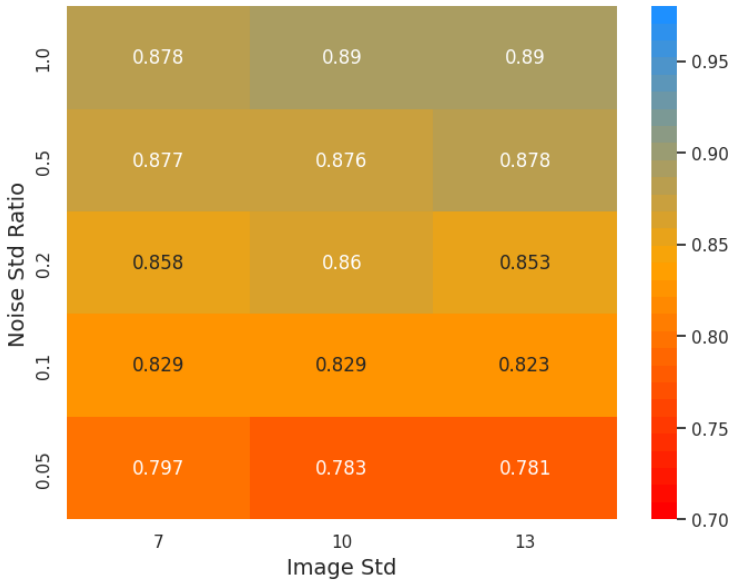

Figure 4.5: Heatmap representing Box IoU of image standard deviation against noise standard deviation ratio in EfficientNet-B0 approach.

Backbone approach	Box IoU
Our U-Net	0.886 $\pm$ 0.011
MONAI U-Net	0.878 $\pm$ 0.016

Table 4.7: Fusion of segmentation- and keypoint detection-based approach metrics for standard-dose images. Segmentation masks were calculated by the segmentation-based approaches.

Simulation type	Backbone approach	Box IoU
Gaussian	Our U-Net	0.896 $\pm$ 0.005
Gaussian	MONAI U-Net	0.890 $\pm$ 0.004
Poisson	Our U-Net	0.905 $\pm$ 0.006
Poisson	MONAI U-Net	0.900 $\pm$ 0.012

Table 4.8: Fusion of segmentation- and keypoint detection-based approach metrics for simulated low-dose images. Segmentation masks were calculated using segmentation-based approaches.

4.3 Overreach and Underreach

As discussed in Section 3.3, overreach and underreach are crucial metrics for evaluating field-of-view detection. While reducing overexposure can be beneficial, it must be balanced to ensure that no vital areas are omitted from the resulting X-ray image. For this the three approaches: our U-Net, MONAI U-Net and EfficientNet-B0 were analysed for the low-dose Gaussian simulation with the phantom X-ray reference of 650 μAs described in Section 3.1 and Section 3.3. The best results can be seen for our U-Net model in Figure 4.6 with an added tolerance of 10 pixel to represent the maxFS as described in Section 3.3. This addition resulted in almost no underreach for the test set.

Adding a tolerance increases both overreach and exposure. With an added tolerance, the MONAI U-Net approach managed to reduce underreach to zero for nearly all test images, only two significant outlier remained, as shown in Figure 4.7.

The EfficientNet-B0-based approach for the keypoint dataset did not perform as well as the segmentation-based approaches, as seen in Figure 4.8, while the added tolerance still reduced underreach significantly.

4.4 Exposure Reduction

For our U-Net-based approach, which demonstrated the strongest overall performance, the FOV reduction was analyzed to estimate the corresponding reduction in exposure. The reduction in FOV percentage was calculated, as shown in Figure 4.9. The resulting average reduction was 50 % without added tolerance and 33.5 % with an added tolerance. The same analysis was repeated for the EfficientNet-B0 for the keypoint detection dataset. The reduction in FOV can be seen in Figure 4.10. The average reduction was 42 % without added tolerance and 25 % with an added tolerance of 10 pixel in each direction.

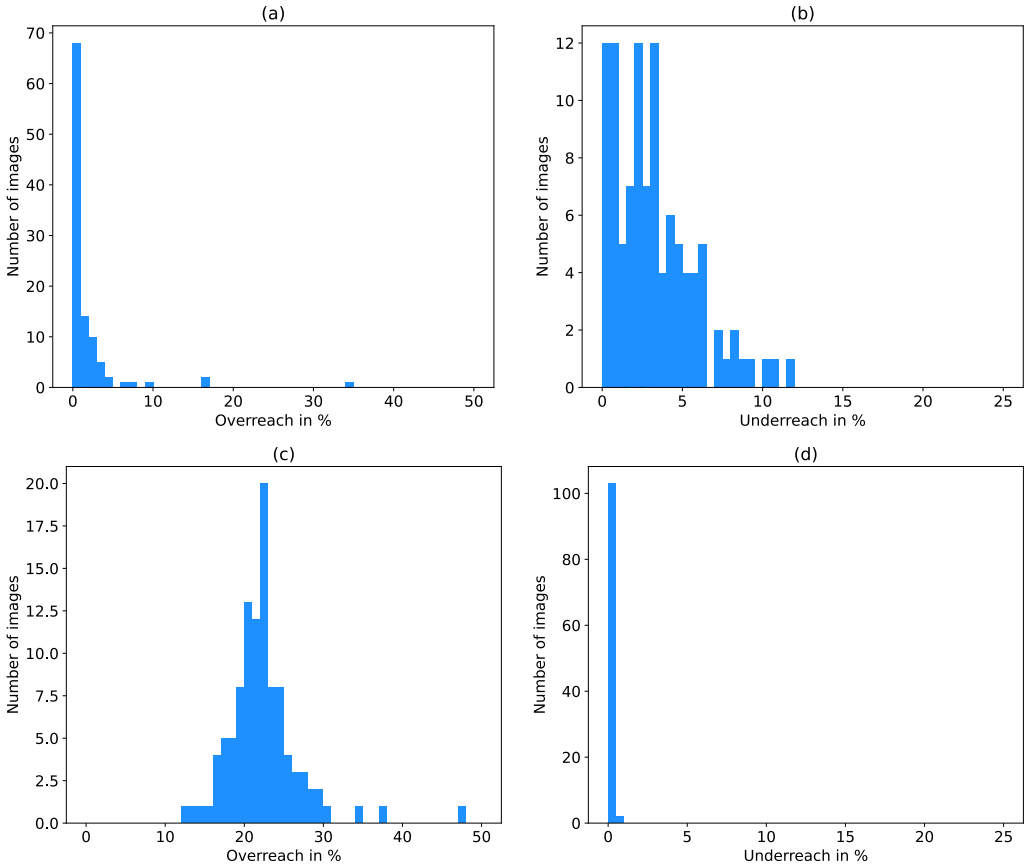

Figure 4.6: Results for our U-Net-based approach on Gaussian simulated low-dose X-ray images; (a) Overreach without tolerance; (b) Underreach without added tolerance; (c) Overreach with tolerance; (d) Underreach with tolerance.

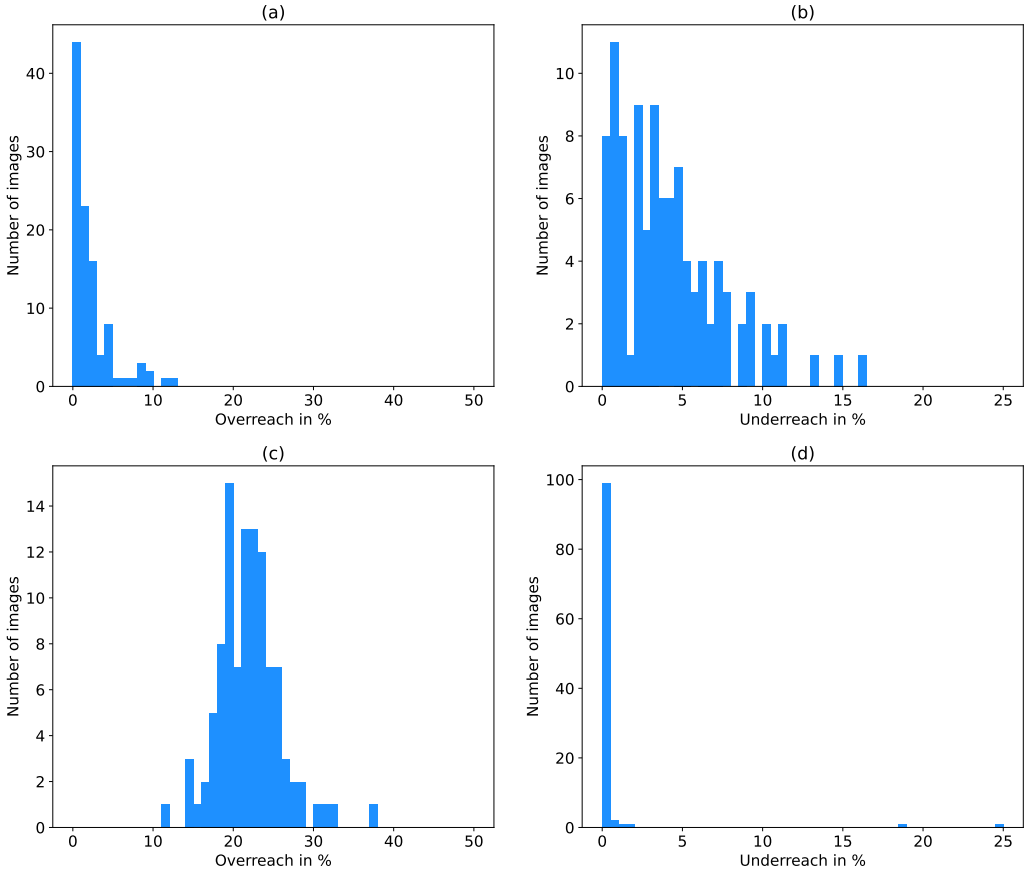

Figure 4.7: Results for MONAI U-Net-based approach on Gaussian simulated low-dose X-ray images; (a) Overreach without tolerance; (b) Underreach without added tolerance; (c) Overreach with tolerance; (d) Underreach with tolerance.

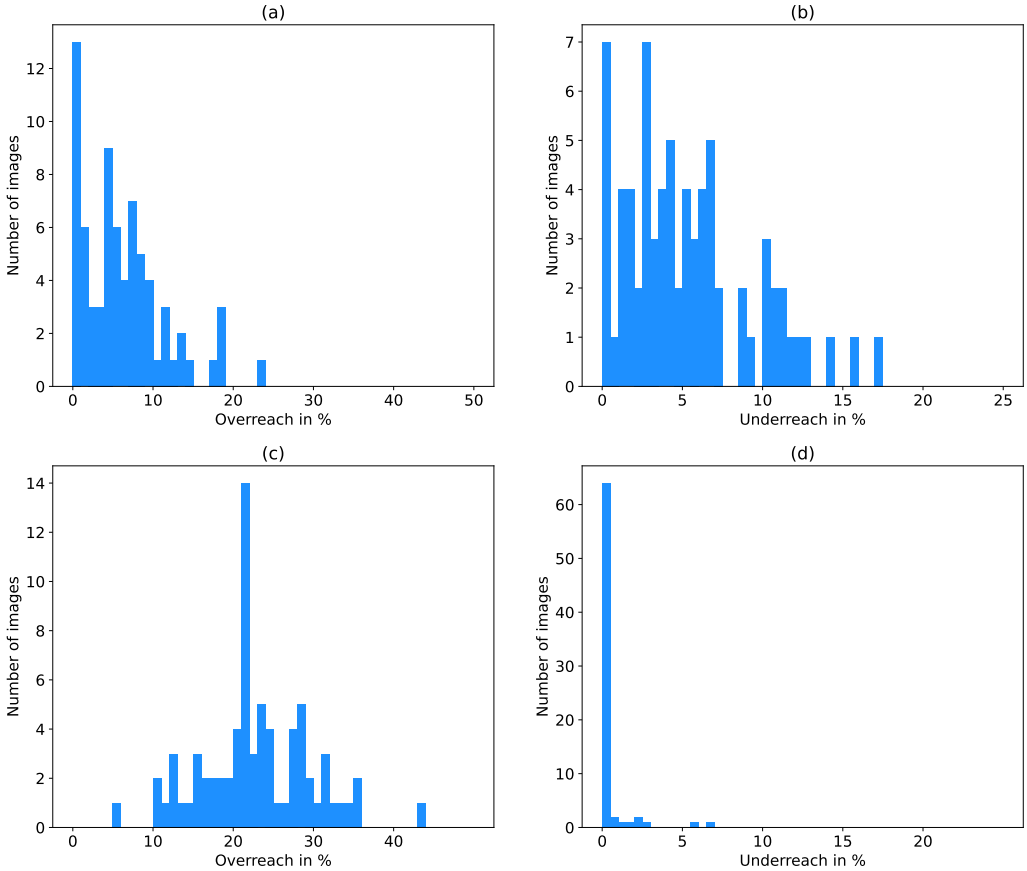

Figure 4.8: Results for EfficientNet-B0-based approach on Gaussian simulated low-dose X-ray images; (a) Overreach without tolerance; (b) Underreach without added tolerance; (c) Overreach with tolerance; (d) Underreach with tolerance.

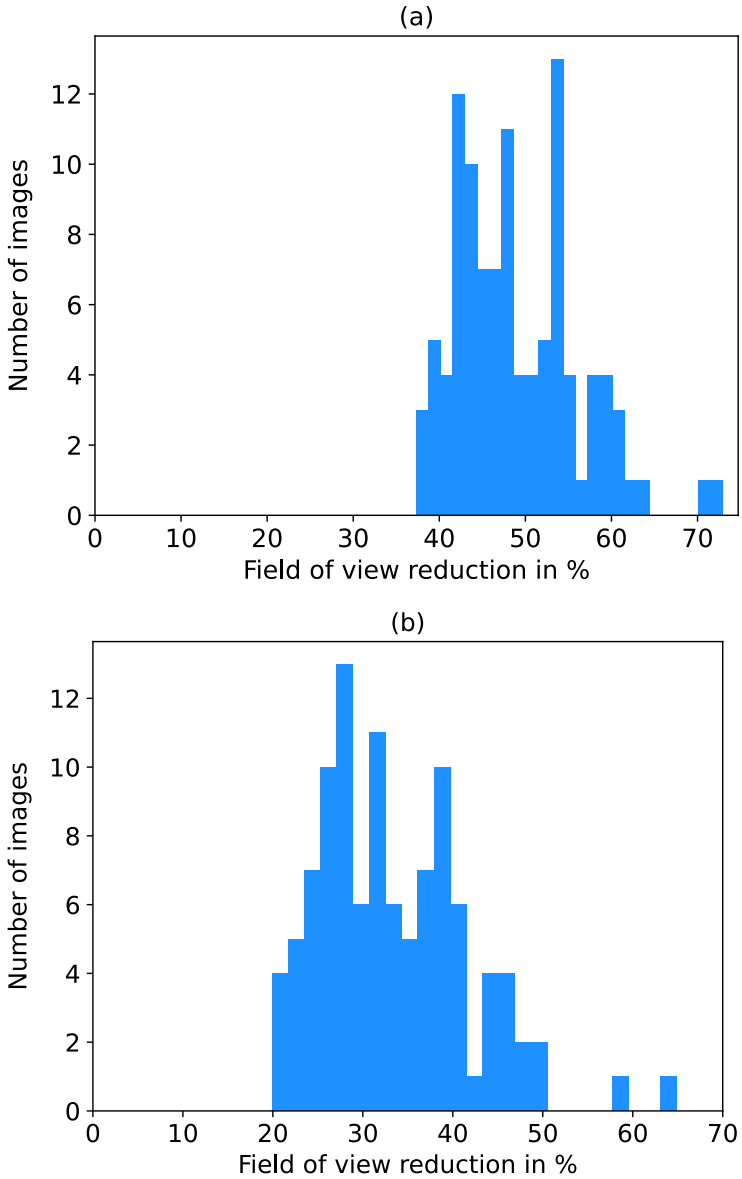

Figure 4.9: Reduction in field-of-view for our U-Net-based approach of Gaussian simulated low-dose X-ray images; (a) Reduction without added tolerance to the resulting bounding box (average 50 %); (b) Reduction with an added tolerance of 10 pixel in each direction (average 33.5 %).

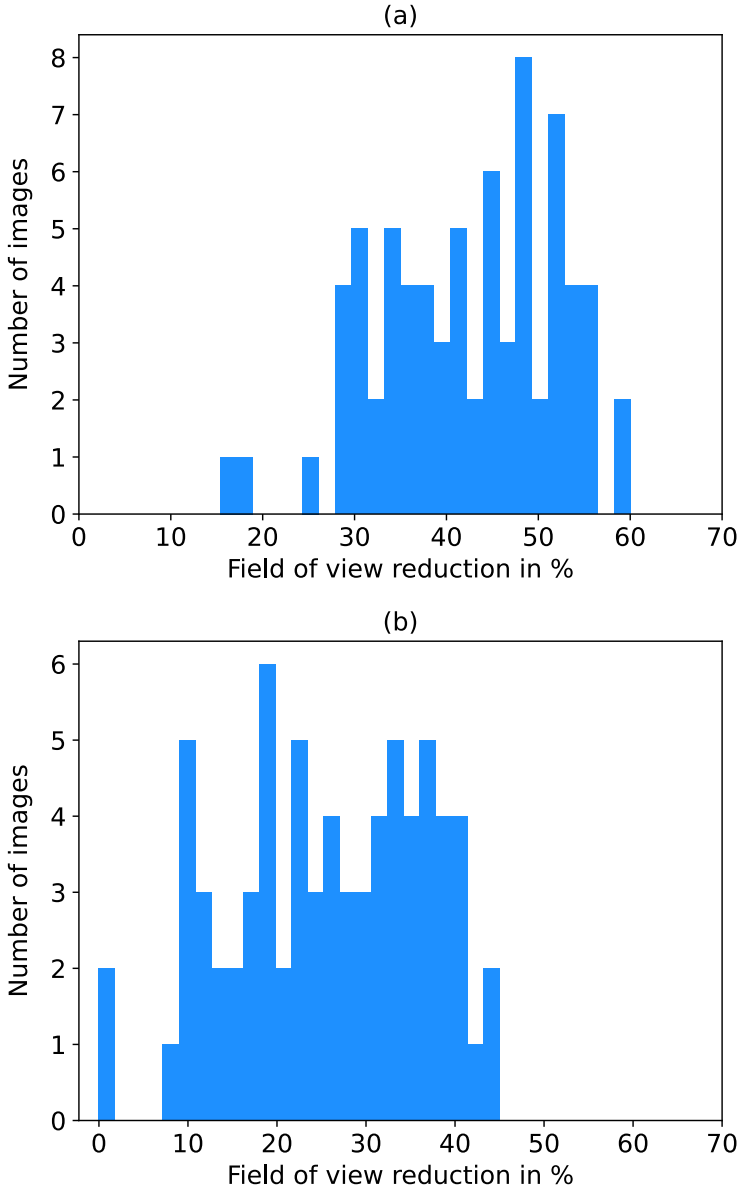

Figure 4.10: Reduction in field-of-view for the EfficientNet-B0-based approach of Gaussian simulated low-dose X-ray images; (a) Reduction without added tolerance to the resulting bounding box (average 42 %); (b) Reduction with an added tolerance (average 25 %).

5 Discussion

5.1 Key Findings

When comparing the segmentation-based approach to the keypoint detection-based approach, it's essential to remember that each method was trained and tested on separate datasets. For the segmentation-based approaches our U-Net performed better than the MONAI U-Net. A possible reason for the better performance could be that it had much more parameters and Floating Point Operations (FLOPs)(which can be seen in Table 5.1). A higher number of parameters enables the model to capture more information, thereby potentially improving accuracy. This might also account for the performance observed with hyperparameter settings that allowed for higher information content.

Model	Parameters (in millions)	FLOPs (in millions)
Our U-Net	1.942	3.448
MONAI U-Net	0.162	0.515
EfficientNet-B0	4.553	0.409
Our CNN	67.206	0.690

Table 5.1: Parameter count and FLOPs for our U-Net, MONAI U-Net and EfficientNet-B0 approaches.

The simulation of low-dose X-ray images using phantom X-ray images as reference did not only use theoretical noise addition, but also took real contrast change into account for lower-dose imaging, strengthening the reliability of the results. The results of our U-Net model on estimated lower-doses are shown in the last chapter in Figure 4.1 and Figure 4.2. These results appear unexpectedly good, given the extent of image manipulation involved. Further analysis of the simulated lower-dose image, with image standard deviation of 10 and noise ratio of 0.2, showed that by contrast manipulation, the contours of the lung can still be made out by eye. This manipulation

is visualized in Figure 5.1. A visualization is displayed with a pixel range of 256 grayscale values, and this kind of detail cannot be visible when displaying the images. X-ray imaging detectors have a wider range of values to store the resulting X-ray detection. This means that FOV detection can also take place on a wider range of values.

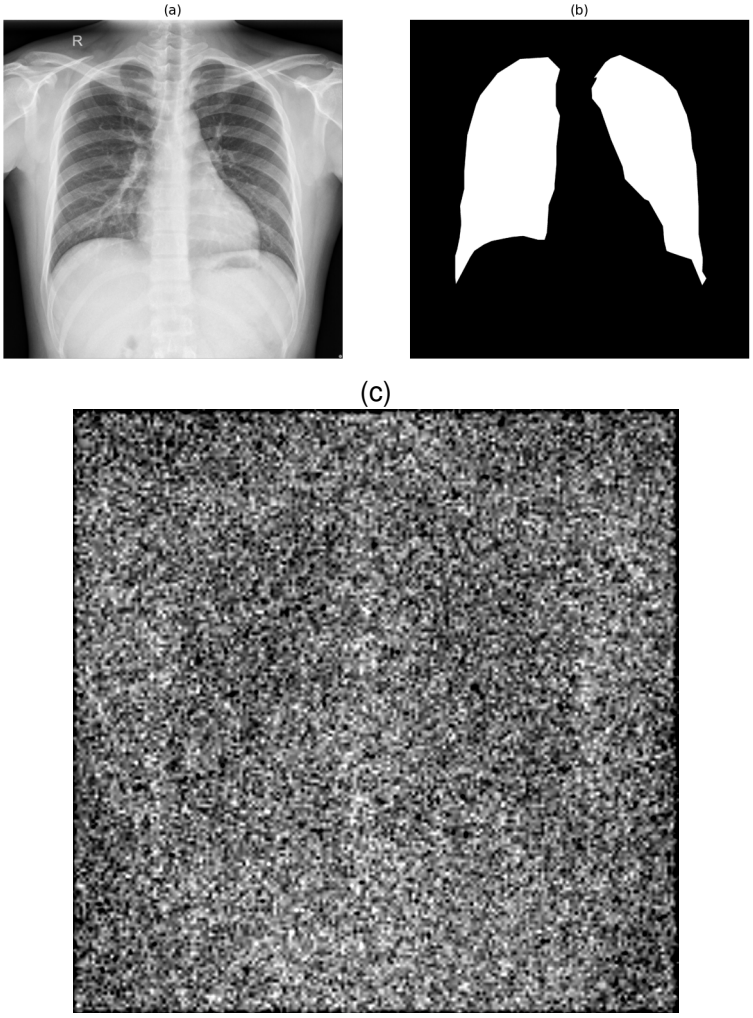

Figure 5.1: Example of lower estimated dose; (a) Original X-ray image; (b) segmentation mask for image (a); (c) Contrast manipulated simulated lower dose image with an reference image standard deviation of 10 and a noise standard deviation ratio of 0.2, which results in added Gaussian noise with a standard deviation of 50.

In contrast to the segmentation-base approaches, for the keypoint detection-based approaches, the parameter count was higher for our CNN approach. The reason for this could be the dense layers after the convolutional layers or that the architecture of the EfficientNet-B0 is specifically designed to be more efficient. As a result, it performed much better than our simple CNN approach.

The fusion-based approach performed worse than the EfficientNet-B0 approach. This could be explained by the variance in datasets and the fact that the segmentation-based backbones were trained on a different dataset and the varying contrast in the keypoint detection dataset. The X-ray images when used alone provided better information about the position of the keypoints than the images with addition of the segmentation masks. Therefore, images with the predicted segmentation masks resulted in a worse overall outcome.

The analysis of overreach and underreach with the addition of a tolerance representing the maxFS showed that with the addition of tolerance, our U-Net based approach was able to eliminate underreach for all except for two test images. The images both had a underreach of around 0.6 %, which is very low. The MONAI U-Net and EfficientNet-B0 approaches, were able to reduce underreach significantly. Underreach was deemed unacceptable, as this implied that regions of interest would not be imaged during an examination. The result of our U-Net-based approach with added tolerance reduced the exposed area by an average of 33 %.

The limitation to the two biggest segmentation areas mentioned in Section 3.4, where the predicted FOV only included the two biggest segments, also had a negative effect on the FOV detection in more difficult images. Two examples are illustrated, firstly in Figure 5.2, where the predicted mask was broken of into smaller parts, resulting in underreach due to the limitation to the two biggest segments. Secondly, the patient in Figure 5.3 exhibited a reduced left lung size or a condition causing the lung to appear smaller. Due to the segmented area between the lungs being bigger than the left lung the predicted FOV is detected wrong and causes underreach.

The effective exposure reduction between the actually exposed area and the minFS of the keypoint detection dataset was estimated by an expert to be between 30 % to 40 %. In Figure 5.4 the FOV reduction by the minFS for all images can be seen. The average area reduction is 32 %. Comparing these results to the predicted FOV reduction by the EfficientNet-B0-based approach in Section 4.4 shows that the keypoint-detection-based approach has strong potential and with further improvement it can help clinics significantly reduce overexposure.

5.2 Implications of the Results

The keypoint detection dataset was motivated by the European guidelines for X-ray imaging, giving this approach more importance and relevance in future research. In general, it can be concluded the datasets and the predicted FOV approaches showed great potential to reduce overexposure.

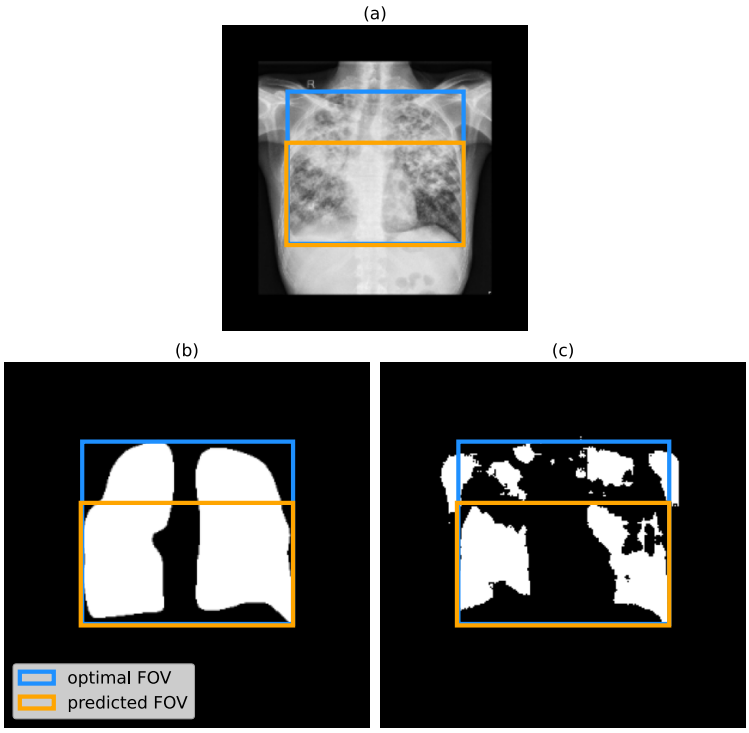

Figure 5.2: Example segmentation of a Gaussian simulated image by the MONAI U-Net; (a) Standard-dose image before low-dose simulation, showing strong signs of pulmonary tuberculosis; (b) Original segmentation mask of X-ray image (a); (c) Predicted segmentation and resulting FOV by only accounting for the two biggest segments, showing a negative effect due to poor segmentation.

The simulation of low-dose X-ray images using phantom X-ray images as reference should provide a good representation of real low-dose X-ray images. The estimated lower-doses simulation might not be a representation of real ultra low-dose X-ray images, but still show the capabilities of both approaches and should be further pursued.

Additionally, not useful optimal FOV detection results, such as the two examples in Figure 5.2 and Figure 5.3 above, need to be able to be detected, such that the technician can intervene and manually adjust the collimation.

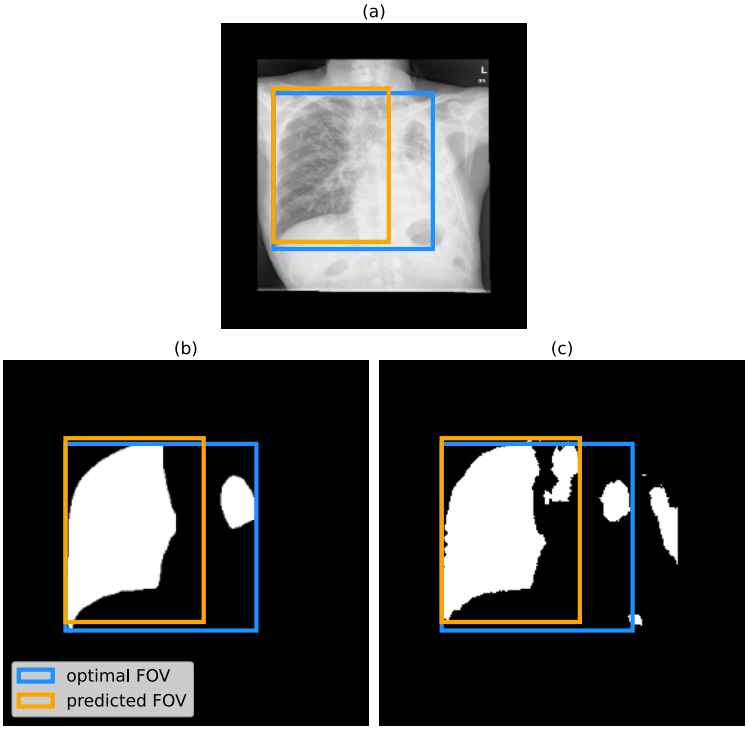

Figure 5.3: Example segmentation of a Gaussian simulated image by the MONAI U-Net; (a) Standard-dose image before low-dose simulation showing the abnormal anatomy of the patient; (b) Original segmentation mask of X-ray image (a) showing a small left lung segment; (c) Predicted segmentation and resulting FOV by only accounting for the two biggest segments, showing a negative effect due to miss segmentation of a patch between the two lungs.

For the live use of optimal FOV detection during X-ray imaging, a fast calculation time is also important because if the evaluation takes too long, the patient is more likely to move. The average evaluation time was calculated for our U-Net, MONAI U-Net and EfficientNet-B0 returning a predicted field-of-view in 1.1 ms, 0.6 ms and 2.7 ms respectively. This shows that the evaluation time should have no impact on the application. To further increase accuracy and reduce overexposure other models could be tested, for example a Residual U-Net.

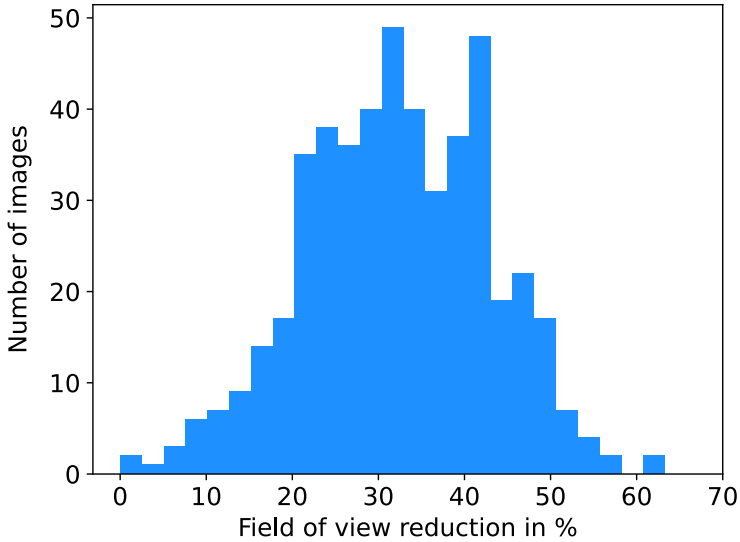

Figure 5.4: Percentage of FOV reduction in the keypoint detection dataset. Reduction from actually exposed area to optimal FOV.

5.3 Limitations of the Study and possible improvements

Overall it can be said, the difference between the segmentation dataset and the keypoint-detection dataset limits the comparison between the two approaches. The segmentation dataset has less variance than the keypoint-detection dataset. For example, pacemakers or pulmonary edema, which is fluid in the lungs, cause the area to absorb more radiation, leaving a brighter area on the resulting X-ray images. This could have affected the results of both approaches. To improve this effect, a dataset with lung segmentation masks and annotated keypoints could be used to directly compare the results.

The need to simulate low-dose X-ray images also weakens the validity of the results. With real low-dose X-ray images, no estimation would have been needed and the results would show the real possibilities in reduction of overexposure.

For medical studies, it is always difficult to obtain a large dataset for a variety of reasons such as ethical and information security reasons. This limits the variance in data and reduces the significance of the performed research.

6 Conclusion

6.1 Summary of the Research

In an effort to reduce overexposure in chest X-ray examinations, two different machine learning-based approaches were tested on simulated low-dose X-ray images to detect the optimal FOV that includes the region of interest while minimizing overexposure. Not only were good results achieved with standard dose images, but additionally simulated images with an exposure of roughly one eighth (0,013 mSv) of the standard dose yielded comparable results. The combination of low-dose image simulation and FOV detection by segmentation of the lung or anatomical keypoint detection showed promise.

6.2 Recommendations for Future Work

Suggested next steps include a detailed calculation of exposure reduction achieved through the FOV prediction. To assess the feasibility of automating FOV adjustments, one could evaluate the process duration for low-dose imaging, FOV prediction, automated adjustment, and standard-dose imaging. A follow-up might involve building a proof-of-concept to further validate this approach.

Additional steps might include using a single dataset with annotated keypoints and lung segmentation masks, enabling direct comparison of both presented approaches or instead expanding one dataset to improve robustness. Testing the two approaches on real low-dose X-ray images would be crucial to validate the approach in clinical settings. Even without fully automated FOV detection and collimation adjustment, technicians could use low-dose images to manually adjust the collimator for cases were estimating the region of interest proves difficult.

Lastly, the approach to reduce exposure during X-ray examination does not only have to focus on chest X-ray imaging. As long as the region of interest is well defined

or a better estimation of the region of interest is needed, a similar deep learning approach could be developed to reduce overexposure. This approach could also be used on other imaging modalities like CT scans, and because a CT scan has a higher exposure than a X-ray examination its resulting radiation reduction for patients could be higher. There are many possibilities to further improve the presented approaches and even more patients can benefit when these methods are applied to a range of radiological fields.

List of Figures

2.1 An example visualization of a U-Net architecture is shown, where each box represents the data and its dimensions. The downsampling process is illustrated by the numbers indicating the shape of the data at each stage. 5
2.2 Example of different Poisson distributions. 7
2.3 Example visualization of a CNN architecture. 9
2.4 Example of the Shenzhen Hospital CXR dataset; Left: X-ray image; Right: correspondent binary lung segmentation mask. 17
2.5 Representational example of keypoint detection dataset; Left: X-ray image with annotated keypoints, T1, T12 vertebrae and both costophrenic angle; Right: X-ray image with minimum field size and maximum field size described in the European commission guidelines. 18
2.6 Example of phantom X-ray images; Left: X-ray image with 760 μAs; Right: image with 2920 μAs. 19
2.7 Visualization for components of overreach and underreach. 24

3.1 Example of phantom X-ray images with histograms; Left: X-ray images; Right: histogram of gray pixel values corresponding to images on the left with mean and std. 30
3.2 Example of histogram matching with original X-ray Image, phantom X-ray images as reference and matched image on the left. On the right are the corresponding histograms. 31
3.3 (a) Phantom X-ray image with 760 μAs with highlighted selected area for noise analysis; (b) Selected area for noise analysis of image (a); (c) Selected area for noise analysis of another image with 2920 μAs. . . . 32
3.4 (a) Mean pixel value of gray-scale phantom X-ray images of dataset 2.4.3, the x-Axis represents the 18 images with their exposure in μAs; (b) Standard deviation of phantom X-ray images. 32
3.5 (a) Original X-ray image; (b) Augmented X-ray image; (c) Original segmentation mask; (d) Augmented segmentation mask 33

3.6 (a) X-ray image after histogram matching with reference image of 650 μAs, (b) Poisson simulated image with image (a) as input for sampling; (c) Matched image with added Gaussian noise. 35

3.7 (a) Standard deviation of estimated noise as described in 3.1, x axis represents images with increasing exposure in μAs; (b) Ratio between image standard deviation to noise standard deviation. 36

3.8 (a) Example simulation with image standard deviation of 10 and a standard deviation ratio of 1; (b) Example simulation with image standard deviation of 10 and a standard deviation ratio of 0.5. 37

3.9 (a) X-ray image with annotated keypoints; (b) X-ray image with keypoints rotated such that the spine aligns vertically; (c) X-ray image with minFS and maxFS. 38

3.10 This image is made up due to copyright reason but represents difficulties of simple rotation; (a) Example X-ray image with keypoints; (b) Augmented X-ray image, showing that keypoints are lost due to rotation. 39

3.11 Example field-of-view from predicted segmentation mask highlighted in the standard-dose X-ray image. The FOV is presented without added tolerance and with an added tolerance of 10 pixel into each direction. . 41

3.12 Example image that shows an adjusted bounding box. Predicted segmentation was highlighted on top of the simulated X-ray image. The adjustment was achieved by selecting only the two largest segmented areas. This minimizes the impact of smaller detected regions, for examples those that can be seen at the bottom of the image. 42

4.1 Heatmap presenting accuracy comparing image standard deviation against noise standard deviation ratio in our U-Net-based approach. 45

4.2 Heatmap presenting Box IoU comparing image standard deviation against noise standard deviation ratio in our U-Net-based approach. 45

4.3 Heatmap presenting accuracy comparing image standard deviation against noise standard deviation ratio in MONAI U-Net-based approach. 46

4.4 Heatmap presenting Box IoU comparing image standard deviation against noise standard deviation ratio in MONAI U-Net-based approach. 46

4.5 Heatmap representing Box IoU of image standard deviation against noise standard deviation ratio in EfficientNet-B0 approach. 49

4.6 Results for our U-Net-based approach on Gaussian simulated low-dose X-ray images; (a) Overreach without tolerance; (b) Underreach without added tolerance; (c) Overreach with tolerance; (d) Underreach with tolerance. 51

4.7 Results for MONAI U-Net-based approach on Gaussian simulated low-dose X-ray images; (a) Overreach without tolerance; (b) Underreach without added tolerance; (c) Overreach with tolerance; (d) Underreach with tolerance. 52

4.8 Results for EfficientNet-B0-based approach on Gaussian simulated low-dose X-ray images; (a) Overreach without tolerance; (b) Underreach without added tolerance; (c) Overreach with tolerance; (d) Underreach with tolerance. 53

4.9 Reduction in field-of-view for our U-Net-based approach of Gaussian simulated low-dose X-ray images; (a) Reduction without added tolerance to the resulting bounding box (average 50 %); (b) Reduction with an added tolerance of 10 pixel in each direction (average 33.5 %). . . . 54

4.10 Reduction in field-of-view for the EfficientNet-B0-based approach of Gaussian simulated low-dose X-ray images; (a) Reduction without added tolerance to the resulting bounding box (average 42 %); (b) Reduction with an added tolerance (average 25 %). 55

5.1 Example of lower estimated dose; (a) Original X-ray image; (b) segmentation mask for image (a); (c) Contrast manipulated simulated lower dose image with an reference image standard deviation of 10 and a noise standard deviation ratio of 0.2, which results in added Gaussian noise with a standard deviation of 50. 58

5.2 Example segmentation of a Gaussian simulated image by the MONAI U-Net; (a) Standard-dose image before low-dose simulation, showing strong signs of pulmonary tuberculosis; (b) Original segmentation mask of X-ray image (a); (c) Predicted segmentation and resulting FOV by only accounting for the two biggest segments, showing a negative effect due to poor segmentation. 60

5.3 Example segmentation of a Gaussian simulated image by the MONAI U-Net; (a) Standard-dose image before low-dose simulation showing the abnormal anatomy of the patient; (b) Original segmentation mask of X-ray image (a) showing a small left lung segment; (c) Predicted segmentation and resulting FOV by only accounting for the two biggest segments, showing a negative effect due to miss segmentation of a patch between the two lungs. 61
5.4 Percentage of FOV reduction in the keypoint detection dataset. Reduction from actually exposed area to optimal FOV. 62

List of Tables

2.1 Confusion matrix visualizing definition of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). 20

4.1 Our U-Net model hyperparameter search metrics. 43
4.2 MONAI U-Net hyperparameter search metrics. 44
4.3 Segmentation-based metrics for low-dose X-ray simulations. 44
4.4 Our CNN-based hyperparameter search metrics. 47
4.5 EfficientNet-B0-based hyperparameter search metrics. 48
4.6 Keypoint detection-based metrics for low-dose X-ray simulation. 48
4.7 Fusion of segmentation- and keypoint detection-based approach metrics for standard-dose images. Segmentation masks were calculated by the segmentation-based approaches. 49
4.8 Fusion of segmentation- and keypoint detection-based approach metrics for simulated low-dose images. Segmentation masks were calculated using segmentation-based approaches. 49

5.1 Parameter count and FLOPs for our U-Net, MONAI U-Net and EfficientNet-B0 approaches. 57

Bibliography

[1] Bundesamt für Strahlenschutz. *X-ray diagnostics: Frequency and radiation exposure of the German population*. en. URL: https://www.bfs.de/EN/topics/ion/medicine/diagnostics/x-rays/frequency-exposure.html#:~:text=The%20frequency%20of%20X%2Dray%20examinations%20in%20Germany%20was%20on,patients%20with%20statutory%20health%20insurance..

[2] European Commission et al. *European guidelines on quality criteria for diagnostic radiographic images*. Publications Office, 1996.

[3] Sebastian Tschauner et al. „European Guidelines for AP/PA chest X-rays: routinely satisfiable in a paediatric radiology division?“ In: *European radiology* 26 (May 2015). DOI: 10.1007/s00330-015-3836-7.

[4] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. 2015. arXiv: 1505.04597 [cs.CV]. URL: https://arxiv.org/abs/1505.04597.

[5] Alexey A. Novikov et al. „Fully Convolutional Architectures for Multiclass Segmentation in Chest Radiographs“. In: *IEEE Transactions on Medical Imaging* 37.8 (2018), pp. 1865–1876. DOI: 10.1109/TMI.2018.2806086.

[6] Isah Salim Ahmad et al. „COVID-19 Detection via Ultra-Low-Dose X-ray Images Enabled by Deep Learning“. In: *Bioengineering* 10.11 (2023). ISSN: 2306-5354. DOI: 10.3390/bioengineering10111314. URL: https://www.mdpi.com/2306-5354/10/11/1314.

[7] Fumio Hashimoto et al. „Automated segmentation of 2D low-dose CT images of the psoas-major muscle using deep convolutional neural networks“. en. In: *Radiol. Phys. Technol.* 12.2 (June 2019), pp. 210–215.

[8] Kailin Chen et al. „Accurate Automated Keypoint Detections for Spinal Curvature Estimation“. In: *Computational Methods and Clinical Applications for Spine Imaging*. Ed. by Yunliang Cai et al. Cham: Springer International Publishing, 2020, pp. 63–68. ISBN: 978-3-030-39752-4.

[9] Sangeetha Prabhu et al. „Production of X-RAYS using X-RAY Tube“. In: *Journal of Physics: Conference Series* 1712.1 (Dec. 2020), p. 012036. DOI: 10.1088/1742-6596/1712/1/012036. URL: https://dx.doi.org/10.1088/1742-6596/1712/1/012036.

[10] M.E. Lomax, L.K. Folkes, and P. O'Neill. „Biological Consequences of Radiation-induced DNA Damage: Relevance to Radiotherapy“. In: *Clinical Oncology* 25.10 (2013). Advances in Clinical Radiobiology, pp. 578–585. ISSN: 0936-6555. DOI: https://doi.org/10.1016/j.clon.2013.06.007. URL: https://www.sciencedirect.com/science/article/pii/S0936655513002471.

[11] Erica L. Mitchell and Patricia Furey. „Prevention of radiation injury from medical imaging“. In: *Journal of Vascular Surgery* 53.1, Supplement (2011). Radiation Safety in Vascular Surgery, 22S–27S. ISSN: 0741-5214. DOI: https://doi.org/10.1016/j.jvs.2010.05.139. URL: https://www.sciencedirect.com/science/article/pii/S0741521410017313.

[12] Ajay Kumar Boyat and Brijendra Kumar Joshi. „A Review Paper: Noise Models in Digital Image Processing“. In: *CoRR* abs/1505.03489 (2015). arXiv: 1505.03489. URL: http://arxiv.org/abs/1505.03489.

[13] V.K. Ayyadevara and Y. Reddy. *Modern Computer Vision with PyTorch: Explore deep learning concepts and implement over 50 real-world image applications*. Packt Publishing, 2020. ISBN: 9781839216534. URL: https://books.google.de/books?id=GfILEAAAQBAJ.

[14] Ian J. Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. http://www.deeplearningbook.org. Cambridge, MA, USA: MIT Press, 2016.

[15] Sergey Ioffe and Christian Szegedy. „Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift“. In: *CoRR* abs/1502.03167 (2015). arXiv: 1502.03167. URL: http://arxiv.org/abs/1502.03167.

[16] Herman Jaramillo and Andreas Rüger. *Machine Learning for Science and Engineering Volume I: Fundamentals*. Society of Exploration Geophysicists, 2023, p. 201. DOI: https://doi.org/10.1190/1.9781560803898.fm.

[17] Reza Azad et al. *Medical Image Segmentation Review: The success of U-Net*. 2022. arXiv: 2211.14830 [eess.IV]. URL: https://arxiv.org/abs/2211.14830.

[18] Mingxing Tan and Quoc V. Le. „EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks“. In: *CoRR* abs/1905.11946 (2019). arXiv: 1905.11946. URL: http://arxiv.org/abs/1905.11946.

[19] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. 2017. arXiv: 1412.6980 [cs.LG]. URL: https://arxiv.org/abs/1412.6980.

[20] Syed Ahmed. *What every user should know about mixed precision training in pytorch*. July 2022. URL: https://pytorch.org/blog/what-every-user-should-know-about-mixed-precision-training-in-pytorch/.

[21] Stefan Jaeger et al. „Two public chest X-ray datasets for computer-aided screening of pulmonary diseases“. In: *Quantitative imaging in medicine and surgery* 4.6 (2014), p. 475.

[22] Sivaramakrishnan Rajaraman et al. „Improved Semantic Segmentation of Tuberculosis—Consistent Findings in Chest X-rays Using Augmented Training of Modality-Specific U-Net Models with Weak Localizations“. In: *Diagnostics* 11.4 (2021). ISSN: 2075-4418. DOI: 10.3390/diagnostics11040616. URL: https://www.mdpi.com/2075-4418/11/4/616.

[23] Eugene C. Lin. „Radiation risk from medical imaging“. In: *Mayo Clinic Proceedings* 85.12 (2010), pp. 1142–1146. DOI: 10.4065/mcp.2010.0260.

[24] *Radiation in Everyday Life | IAEA — iaea.org*. https://www.iaea.org/Publications/Factsheets/English/radlife. [Accessed 21-10-2024].

[25] Guido Van Rossum and Fred L. Drake. *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace, 2009. ISBN: 1441412697.

[26] Adam Paszke et al. „Automatic differentiation in PyTorch“. In: (2017).

[27] M. Jorge Cardoso et al. „MONAI: An open-source framework for deep learning in healthcare“. In: (Nov. 2022). DOI: https://doi.org/10.48550/arXiv.2211.02701.