applied sciences

MDPI

Article

Detection of Pediatric Femur Configuration on X-ray Images

Marta Drążkowska

Institute of Automatic Control and Robotics, Poznan University of Technology, Piotrowo 3a, 61-138 Poznan, Poland; marta.drazkowska@put.poznan.pl; Tel.: +48-61-665-2043

Abstract: In this paper, we present a fully automatic solution for denoting bone configuration on two-dimensional images. A dataset of 300 X-ray images of children's knee joints was collected. The strict experimental protocol established in this study increased the difficulty of post-processing. Therefore, we tackled the problem of obtaining reliable information from medical image data of insufficient quality. We proposed a set of features that unambiguously denoted configuration of the bone on the image, namely the femur. It was crucial to define the features that were independent of age, since age variability of subjects was high. Subsequently, we defined image keypoints directly corresponding to those features. Their positions were used to determine the coordinate system denoting femur configuration. A complex keypoint detector was proposed, composed of two different estimator architectures: gradient-based and based on the convolutional neural network. The positions of the keypoints were used to determine the configuration of the femur on each image frame. The overall performance of both estimators working in parallel was evaluated using X-ray images from the publicly available LERA dataset.

Keywords: biomechanics; image processing; machine learning; neural network; medical image

Citation: Drążkowska, M. Detection of Pediatric Femur Configuration on X-ray Images. *Appl. Sci.* **2021**, *11*, 9538. https://doi.org/10.3390/app11209538

Academic Editor: Fabio La Foresta

Received: 17 September 2021
Accepted: 11 October 2021
Published: 14 October 2021

1. Introduction

In recent years, robotics have largely influenced medical practices [1]. Artificial intelligence, miniaturization, and computer power all contribute to the widespread usage of robots in medicine. One of the branches benefiting from new structures and control systems are human motion aids, i.e., prostheses, orthosis, and rehabilitation manipulators.

The mechanisms reflect the individual patient's needs, which usually require acquisition of real medical data from the patient. Proper extraction of important features from the data is necessary to maintain the benefits of the aforementioned systems.

One example of such application is a robotic rehabilitation aid that is able to track the movement path of the real healthy joint. Source information can be obtained by acquisition of medical images (e.g., MRI, CT, X-ray). To automatically extract important features from medical image data, convolutional neural networks (CNNs) are commonly used [2]. The reason for choosing CNN to alleviate medically oriented problems is two-fold. First, due to the complex nature of the input data, the hand-engineered features are difficult to obtain. Second, CNN has found its application in plenty of medically oriented applications, giving satisfactory results in a broad range of solutions [2].

This paper presents results that are within this mainstream convention. We propose an automatic solution to detect key features, i.e., keypoints, on medical images. In this study, we considered children's X-ray images of knee joints in lateral view. The radiographs presented joints of different maturity levels, as all subjects underwent bone ossification. The ratio between bone structures and soft tissues is directly connected to the subject's age. Therefore the interpretation is more difficult in comparison to adult X-rays, where ossification is complete.

The main contribution of this paper can be summarized as follows:

1. Selection of image features that unambiguously define the configuration of the bone on the X-ray image, given the troublesome specification of the image data.

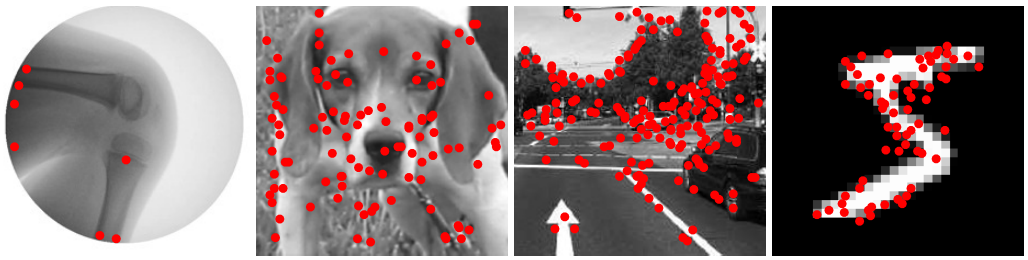
2. Selection of keypoint sets connected to those features.
3. Design of the optimal detection algorithm, which enables the proper estimation of keypoints on X-ray images; the proposed algorithm includes a specially tailored estimator consisting of adaptive threshold and deep CNN.
4. Proposition of the bone coordinate system directly corresponding to its configuration.

The accuracy of the proposed method is defined as a root mean squared error (RMSE) between the estimated configuration of the bone and the reference coordinate system. The overall accuracy of the presented method is evaluated on the publicly available LERA dataset [3]. The dataset consist of lower extremity radiographs of adults gathered by the Stanford University School of Medicine.

The contribution of this paper is threefold. From a medical perspective, we provide a complete solution to obtain femur configuration on two-dimensional X-ray images. From a robotics point of view, the trajectory of femur configuration could be used to define the kinematic model of the joint (to be incorporated in rehabilitation robotic aid). From an artificial intelligence point of view, we provide the optimal estimator architecture to solve a regression task for medical images.

1.1. Related Work

In recent years, CNN image processing has been successfully applied in many applications, e.g., road detection and face recognition. In the case of medical images, the input data possess less salient features than typical CNN input images. The example image frame, considered in this study, with speeded-up robust features (SURF) [4] denoted as red circles are presented in Figure 1a. Note the difference in feature number in contrast to example images from datasets used in different applications, presented in Figure 1b–d. As a side note, the SURF features are presented in Figure 1 for comparison reasons. Any other traditional gradient-based method of feature extraction would result in a similar result.



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

Figure 1. Example images with SURF features. (**a**) X-ray image; (**b**) Dogs vs. Cats [5]; (**c**) KITTI dataset [6]; (**d**) MNIST dataset [7].

Due to the complex (and unique) nature of the medical images, most CNN applications in image processing involve classification [8,9]. Since classification output is discrete (i.e., classes) it is considered less difficult than regression, where output is usually a real number (keypoint positions, segmentation, object detection, etc.). Although several CNN-based keypoint detection methods have been proposed in medical image analyses [10–12], it is still challenging to detect image keypoints.

Interestingly, several deep learning algorithms had been used on adult X-ray images [13–16]. Meanwhile, very little research was conducted for medical image data collected for children [17]. Plenty of reasons for this imbalance can be named, e.g., consent problems, complex nature of children's medical images (age dependency of visible structures, intra- and interpopulation variation).

Recently, individual studies have made attempts to apply CNN to solve regression tasks for children's medical images [18–20]. Nevertheless, there have been issues considering the lack of input data, as pediatric medical image datasets are rarely publicly available. To avoid the problem of limited training data, some deep learning based keypoint detection methods adopt local image patches as samples to perform regression for each of the patches

individually [21]. Those solutions are time consuming and require large computational costs, if each landmark is detected separately.

Alternative solutions use end-to-end learning strategies with entire images as input and the keypoint coordinates as output [22]. The keypoints can be represented as heatmaps [12], i.e., images where Gaussians are located at the position of the keypoints. Then, the task can be understood as image segmentation, with heatmaps being the target. This opens plenty of new possibilities, as many network architectures are designed for image segmentation, e.g., U-Net [23].

The complexity of pediatrics medical images, in comparison to adult ones, is specifically evident in knee radiographs. The images of younger patients have open growth plates, ossification center changes, and possess less characteristic radiographic landmarks [24]. For example, the contact points of knee joint surfaces [25] are not detectable in the X-ray images of young patients. Given this troublesome characteristic of input data, the task of keypoint detection is more demanding, which has to be encountered in the algorithm design.

1.2. Problem Statement

Bone configuration on each image frame can be understood as its orientation and position, i.e.,

$$\boldsymbol{g} = \begin{bmatrix} \theta \; x \; y \end{bmatrix}^\top, \tag{1}$$

where θ denote the orientation of the bone and x, y stand for its position. To define the configuration of the bone on each image, we assume keypoints $\boldsymbol{k}_j$, $j = 1, \ldots, f$. With each image, we correlate the keypoint set:

$$K \triangleq \begin{bmatrix} \boldsymbol{k}_1 \; \boldsymbol{k}_2 \; \ldots \; \boldsymbol{k}_f \end{bmatrix} = \begin{bmatrix} x_1 \; x_2 \; \ldots \; x_f \\ y_1 \; y_2 \; \ldots \; y_f \end{bmatrix}, \; K \in \mathbb{R}^{2 \times f}, \tag{2}$$

where x_j and y_j are the coordinates of j-th keypoint $\boldsymbol{k}_j$. For each image frame, we assume that it is possible to obtain valid femur configuration $\boldsymbol{g}$ from the selected keypoint set. Therefore, we can state that

$$\boldsymbol{g} = \phi_g(K), \quad \phi_g : \mathbb{R}^{2 \times f} \to \mathbb{R}^3, \tag{3}$$

where ϕ_g is a transformation from keypoint matrix (2) to the bone configuration (1). To avoid the ambiguity of bone configuration, we assume that the transformation ϕ_g should be equal for all image frames.

Since ground truth is not available for a considered problem, the reference configuration of the femur will be based on manually marked keypoints. Therefore, we distinguish the configuration of the femur obtained by manually denoted keypoints (with subscript m) and the configuration obtained by estimated positions of keypoints (with subscript e), i.e.,

$$\boldsymbol{g}_m = \begin{bmatrix} \theta_m \; x_m \; y_m \end{bmatrix}^\top, \quad \boldsymbol{g}_e = \begin{bmatrix} \theta_e \; x_e \; y_e \end{bmatrix}^\top. \tag{4}$$

The accuracy of the proposed method will be evaluated as a difference between manual and estimated femur configuration. The proposed solution, depicted in Figure 2, consists of the following steps. First, in the initialization phase, the set of image features that correspond to the femur configuration is chosen. Second, femur features are manually annotated on each image frame by a medical expert. The keypoints are treated as a reference and the goal of the proposed estimation algorithm is to reflect those positions as closely as possible. Two different estimation techniques are chosen in this specific scenario, i.e., tailored adaptive threshold estimator and CNN-based estimator. The overall performance of the estimator is evaluated in the test set, consisting of previously unused medical image frames.

The configuration of the femur is used as a registration coordinate system, to remove the influence of femur configuration changes on the overall kinematics of the joint. If suc-

ceeding, the estimation algorithm could be used as an initialization phase of knee joint kinematic model evaluation.

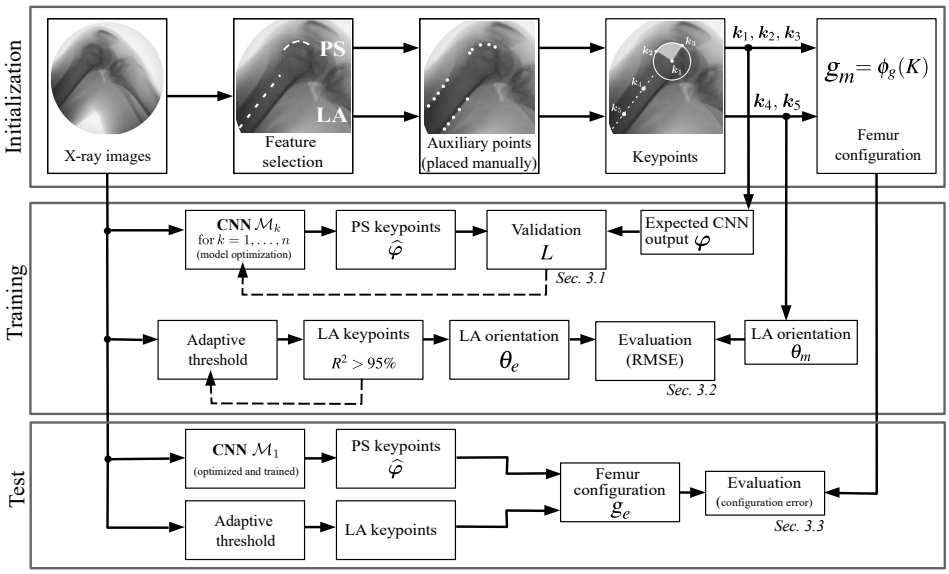


Figure 2. The three stages of the proposed algorithm with schematic representation of each step.

2. Materials and Methods

The method presented in this paper is divided into three stages, as depicted in Figure 2. Each will be discussed briefly. The initialization step, described in Section 2.1, consist of selection of the features that unambiguously determine the configuration of the femur on the image data. Each feature is then annotated by manually marked keypoints. This stage results in femur configuration g_m corresponding to each image frame. This step is not repeated for new, unseen image data.

In the training stage, the image-configuration pairs from the initialization stage are used to select the optimal structure of the estimators. The manually denoted keypoints are used as a reference during the evaluation. This part is described in Sections 2.2 and 2.3 while the results are gathered in Sections 3.1 and 3.2.

The test stage evaluates the performance of the trained estimators. New X-rays, representing new subjects are examined. The performance is evaluated as a difference between the estimated femur configuration g_e and the reference g_m. This step is described in Section 3.3.

2.1. Initialization

In this study, 14 subjects were examined, 12 of which were orthopedic patients averaging 10 years (5–18), 9 female, and 6 male. The legal guardians of all subjects gave informed consent to participate in this study approved by the Bioethics Committee of Poznan University of Medical Sciences (resolution 699/09). The remaining two subjects were 25-year-old healthy adults (one female and one male).

Static image frames were recorded for a non-weight bearing passive movement in a horizontal plane using a fluoroscopy system (Philips BV Libra C-Arm, 1008 px × 576 px resolution). Lateral view frames were gathered for each subject for different angular positions of tibia, whereas the femur was fixed manually. Several selected image frames are presented in Figure 3. Note that, more than one image frame was taken for each subject.

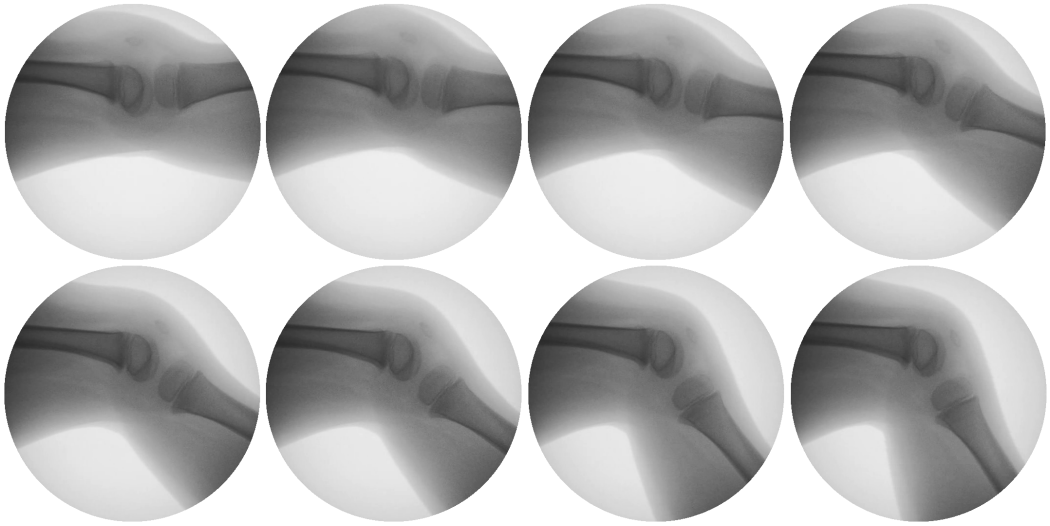


Figure 3. Example image frames of one subject. Images were adjusted for visualization purposes.

The proposed examination protocol possesses few limitations. Undoubtedly, the quality and the quantity of information present in the input image data are limited and below modern medical data acquisition standards. However, poor quality constitutes a scientific challenge to overcome. Thus, the proposed algorithm should alleviate the issue of problematic input data. In this particular scenario, the following aspects of the examination protocol had to be taken into consideration:

1. Minimization of the subjects' fatigue during examination (femur was fixed manually, not firmly; thus, the configuration of femur g_i was not static);
2. Minimization of the radiation level during examination (certain radiation-free techniques, e.g., magnetic resonance imaging, were not allowed for a given study; subjects with the Ilizarov apparatus, screws);
3. The difference of visible bone outlines on images of subjects of different ages (bone formation and growth occurs gradually up to 23 years old);
4. Subjects with normal and abnormal knees had to be examined (the pathology largely influences the bone structure).

Given the problems stated above, we propose that the configuration of the femur is defined by two features, namely the patellar surface (PS) and the long axis (LA) of the femur, as presented in Figure 4. Notably, the chosen features are redundant, but the redundancy is intentional. The bone image is a two-dimensional projection of the three-dimensional structure on the fluoroscopic screen; thus, the visible bone outline cannot be treated as a rigid body. It is possible that the out of plane rotation of the bone could be interpreted as bone deformation (The assumption was made that the rotation around the sagittal axis, i.e., out of plane rotation, is limited.). It must be encountered in the proper selection of keypoints corresponding to the chosen features.

LA can be defined as the middle line of the femoral shaft and, therefore, can be obtained by clearly visible borders of the femur shaft (Figure 4). Detection of keypoints denoting LA may be completed by traditional gradient-based image processing. On the other hand, keypoints on PS are ambiguous. The surroundings of PS are greatly age-dependent, and the border between the bone and soft tissue is untraceable. Using traditional image keypoint detectors may be invalid in this particular case. Therefore, we propose dividing the task of keypoint detection into two, i.e.,

- Keypoints corresponding to the LA of the femur will be estimated using traditional gradient-based methods, as described in Section 2.3;
- Keypoints corresponding to the PS of the femur will be estimated using CNN, as described in Section 2.2.

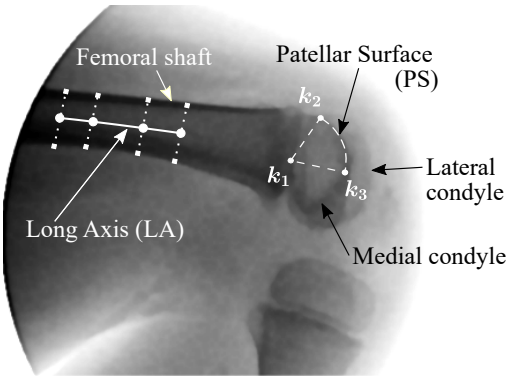


Figure 4. X-ray image frame with assigned features of the femur. Original image was adjusted for visualization purposes.

What is worth pointing out, the feature selection is a part of the initialization stage of the algorithm, as presented in Figure 2. The features will remain equal for all subjects evaluated by the proposed algorithm. Only the positions of keypoints on image data will change.

The following procedure is proposed to obtain keypoints on each image. Each image frame is presented on screen and a medical expert denotes auxiliary points manually on the image. For LA, there are 10 auxiliary points, 5 for each bone shaft border, and PS is determined by 5 auxiliary points (see Figure 2 for reference). The auxiliary points are used to create the linear approximation of LA, and the circular sector approximating the PS (as denoted in Figure 4). Five keypoints $\boldsymbol{k}_1, \ldots, \boldsymbol{k}_5$ are automatically denoted on LA and PS, as shown in Figure 2.

The set of keypoints, given by Equation (2), constitutes the geometric parameters of important features of the femur, and is necessary to calculate the configuration of the bone on each image. In this work, the assumption was made that the transformation (3) exists. As stated before, a visible bone image cannot be considered a rigid body; therefore, the exact mapping between keypoints from two image frames may not exist for a two-dimensional model. Therefore, we propose to define femur configuration as presented in Figure 5.

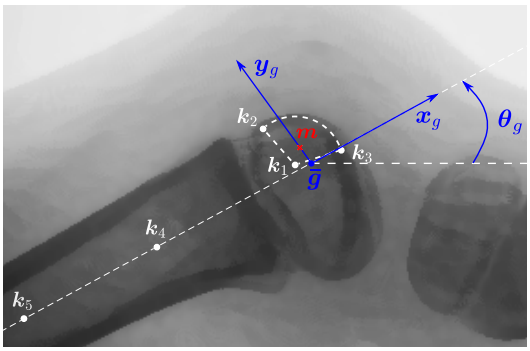


Figure 5. Keypoints of the femur and corresponding femur coordinate system.

The orientation of the bone $\boldsymbol{\theta}_g$ is defined merely by the LA angle. On the other hand, the origin of the coordinate system of femur configuration $\bar{\boldsymbol{g}}_i$ is defined using both, LA and PS. Assume $\boldsymbol{m}$ is a centroid of PS, then we can state that $\boldsymbol{m} = \begin{bmatrix} m_x & m_y \end{bmatrix}^\top = \frac{1}{3}(\boldsymbol{k}_1 + \boldsymbol{k}_2 + \boldsymbol{k}_3)$. Accordingly, $\bar{\boldsymbol{g}}_i$ is a point on LA, which is the closest to $\boldsymbol{m}$. Assuming the previously stated reasoning, it is possible to obtain the transformation ϕ_g from Equation (3) as

$$\phi_g = \begin{bmatrix} \operatorname{atan2}\left(\frac{y_4-y_5}{x_4-x_5}\right) \\ \frac{\left(\frac{y_4-y_5}{x_4-x_5}\right)m_y+m_x-\left(\frac{y_4-y_5}{x_4-x_5}\right)\left(y_5-x_5\left(\frac{y_4-y_5}{x_4-x_5}\right)\right)}{1+\left(\frac{y_4-y_5}{x_4-x_5}\right)^2} \\ \frac{\left(\frac{y_4-y_5}{x_4-x_5}\right)^2 m_y+\left(\frac{y_4-y_5}{x_4-x_5}\right)m_x+y_5-x_5\left(\frac{y_4-y_5}{x_4-x_5}\right)}{1+\left(\frac{y_4-y_5}{x_4-x_5}\right)^2} \end{bmatrix}. \quad (5)$$

2.2. Training Stage: CNN Estimator

The CNN estimator is designed to detect the positions of three keypoints $\boldsymbol{k}_1$, $\boldsymbol{k}_2$, and $\boldsymbol{k}_3$. Those keypoints correspond to PS, which is located in the less salient region of the X-ray image. The correctly designed estimator should assign keypoints in the positions of the manually marked keypoints. For example, for every image frame, the expected output of CNN is given by

$$\boldsymbol{\varphi} = [\boldsymbol{k}_1\ \boldsymbol{k}_2\ \boldsymbol{k}_3]^\top \in \mathbb{R}^6. \quad (6)$$

First, X-ray images with corresponding keypoints described in the previous section were preprocessed to constitute valid CNN data. The work-flow of this part is presented in Figure 6. Note that, all of the presented transformations are conducted simultaneously on images and corresponding keypoint positions. Thus, keypoints reflect the configuration of PS on the source image.

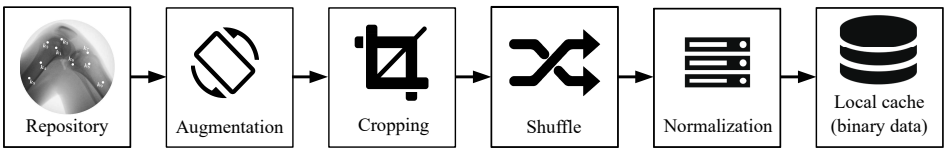


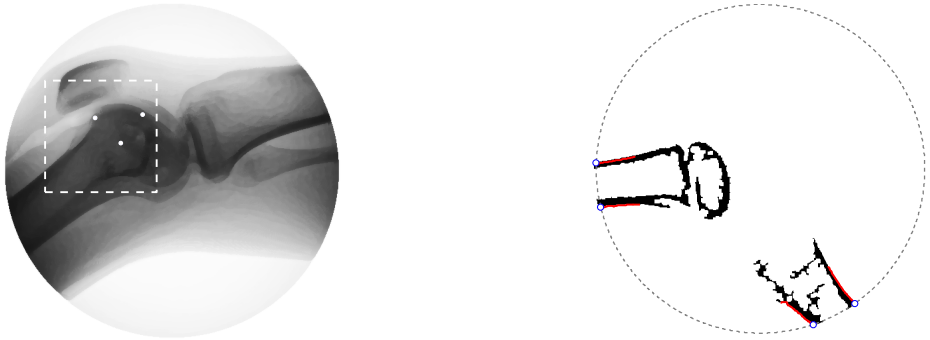
Figure 6. Generation of CNN learning sets.

As a first stage, due to the small dataset size, the original data were augmented with typical image transformations (rotation, translation, scale, reflection, contrast change [26]). Second, image frames were cropped to size 178 × 178 px. The smaller resolution was selected as a trade off between hardware requirements (memory limitation) and minimizing the loss of information. The example of cropping operation is presented in Figure 7a. The position of the cropping window was chosen randomly with the assumption that it contained all of the keypoints.

The third step consists of shuffling data to avoid local minima in the learning process. Note that, after shuffling, the input and output pair remains the same. Finally, the images are normalized to unify the significance of each input feature on the output.

The learning data are sequentially divided between the train and development sets, as described in Table 1. Note that images of one subject constitute exclusively one of the sets.

To evaluate the performance of CNN architecture, a separate test set is formed. In this study, a slice of the publicly available LERA dataset [3] is used, consisting of knee joint images in the lateral view. The whole dataset consists of 182 images of different joints of the upper and lower limb, collected between 2003 and 2014. Note that the dataset includes radiographs varying in size and quality; therefore, a proper preprocessing and standardization of resolution is needed.



(**a**) (**b**)

Figure 7. Visualization of certain preprocessing stages of the algorithm. (**a**) The whole X-ray image with cropped window (dashed line) and keypoints (circle) of PS. (**b**) Adaptive thresholded X-ray image with fluoroscopic lens (dotted line), points $\boldsymbol{p}_{p_1}$ and $\boldsymbol{p}_{a_1}$ (round marker), and set of points $\boldsymbol{p}_p$ and $\boldsymbol{p}_a$ (red line). Images were preprocessed for visualization purposes.

Table 1. Gathered data sets for CNN training.

Learning Set	Learning Examples		Number of Subjects
	Original	Augmented	
Train	318	12,000	12
Development	32	1200	2
Test [1]	44	44	44
Overall	394	13,244	58

[1] The test set comprises of the LERA dataset [3] images. Only images of the knee joint were selected from the dataset.

This study focuses on classic feedforward networks, i.e., without feedback connections. It is assumed that the values of the weights and biases are trained in the stochastic gradient descent learning process. The chosen optimization criterion is given by mean squared error value

$$L \triangleq \widetilde{\boldsymbol{\varphi}}^{\top}\widetilde{\boldsymbol{\varphi}}, \quad \widetilde{\boldsymbol{\varphi}} \triangleq \widehat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}, \tag{7}$$

where $\widehat{\boldsymbol{\varphi}}$ is the estimated output of CNN and $\boldsymbol{\varphi}$ is the expected output of CNN given by Equation (6). Note that, contrary to most medical image oriented CNN scenarios, here CNN is designed to solve regression task, i.e., keypoint coordinates are given in real numbers.

Importantly, the loss function (7) gradient is calculated with a modified backpropagation process, i.e., ADAptive Moment estimation [27]. Due to the large complexity of the considered problem, CNN architecture, as well as learning parameters, will be optimized. The optimal network architecture, among different possible structures, will ensure the lowest loss function value (7). The optimization procedure is described in Appendix A.

We acknowledge that collected datasets (Table 1) are limited in size. Nevertheless, we are convinced that proper regularization will assure an appropriate learning process. Following the work presented in [26], we have applied:

1. L_2 Regularization: enforcing a sparsity constraint on CNN parameters;
2. Dropout [28]: randomly omitting units during the training process (denoted as DO(p), where $p \in (0, 1)$ stands for a probability of dropout);
3. Batch normalization [29] (denoted as BN): normalization of every layer's output, normalization parameters are trained together with other CNN parameters;
4. Early stopping: avoiding long-term training of models with high loss function value and/or overfitting the training data. If after 50 epochs of learning, the minimum of loss function value for training and validation set is higher than 11 px^2 or after 150 epochs is higher than 6 px^2, the training process is stopped.

2.3. Training Stage: Gradient-Based Estimator

As stated previously, border points of bone shafts can be used to calculate the LA keypoints. Assume a set of points on each border of the bone shaft

$$\boldsymbol{p}_p \triangleq \{p_{p_k}\}_{k=1}^{\lambda}, \quad \boldsymbol{p}_a \triangleq \{p_{a_k}\}_{k=1}^{\lambda}, \tag{8}$$

where λ is a predefined number of auxiliary keypoints to be assigned on each side of the border, for the posterior and anterior side, respectively. The first point in each set represents an intersection between the bone shaft and the fluoroscopic lens, as presented in Figure 7b.

The proposed algorithm of LA keypoints estimation is described as Algorithm 1. The steps will be discussed briefly. First, the input image is converted to binary by the adaptive threshold technique [30] with a randomly chosen window size s. Second, the binary image is opened to cancel all small objects. The points of remaining objects that intersect with the fluoroscopic lens are detected (round markers in Figure 3) and are used as initial conditions to the Moore–Neighbor tracing algorithm [31]. An eight pixel neighborhood is chosen and border points $\boldsymbol{p}_p$ and $\boldsymbol{p}_a$ are extracted. Each border is approximated using simple linear regression [32]. Both linear approximations are verified in terms of the coefficient of determination, denoted as R^2. If both bone borders are estimated correctly, i.e., the linear regression of points can be correctly approximated by a straight line, R^2 is high and the algorithm passes to the next step. If the condition is not satisfied, a different adaptive threshold window size s is chosen, and all of the steps are repeated for a new binary image.

Algorithm 1: LA keypoint estimation.

Result: $\boldsymbol{k}_4$, $\boldsymbol{k}_5$
Input: X-ray image;
Set $\boldsymbol{s} = 11 + 2 \cdot j, \quad j \in \{0, \ldots, 15\}$;
Set $\lambda = 50$, $flag = 0$;
while *flag* = 0 **do**
  Randomly select window size $s \in \boldsymbol{s}$;
  Binarize image with s [30];
  Cut small objects (binary image opening);
  Find initial points p_{p_1} and p_{a_1} on binary image;
  Trace $\boldsymbol{p}_p$ and $\boldsymbol{p}_a$ using [31];
  Calculate linear regression of $\boldsymbol{p}_p$ and $\boldsymbol{p}_a$ with [32];
  if $R^2 > 95\%$ **then**
    flag = 1;
  end
end
Calculate $\boldsymbol{p}_{LA} = (\boldsymbol{p}_a + \boldsymbol{p}_p)/2$;
Set $\boldsymbol{k}_4$ and $\boldsymbol{k}_5$ on linear regression of $\boldsymbol{p}_{LA}$;

As soon as bone borders are estimated correctly, the mid-points between the borders are obtained. Their linear regression comprises the LA line. Keypoints $\boldsymbol{k}_4$ and $\boldsymbol{k}_5$ are assigned as random points on a linear approximation of LA.

3. Results

In this section, we present the estimation results of both LA and PS. First, the optimal CNN architecture is discussed and the results of PS keypoint detection are presented, according to scheme from Figure 2. Second, the results of the proposed gradient-based method of the LA keypoint estimation are described. At the end of the section, the overall performance of the two combined methods of estimation is presented. The results are compared with the configuration of the femur obtained by manually marked keypoints.

3.1. PS Estimation

As a result of training over 200 networks with different architectures, the one ensuring the minimum loss function value (7) was chosen. The network architecture is presented in Figure 8. The optimal CNN architecture [26] consists of 15 layers, 10 of which are convolutional. The size of the last layer represents the number of network outputs, i.e., the coordinates of keypoints $\boldsymbol{k}_1$, $\boldsymbol{k}_2$, $\boldsymbol{k}_3$.

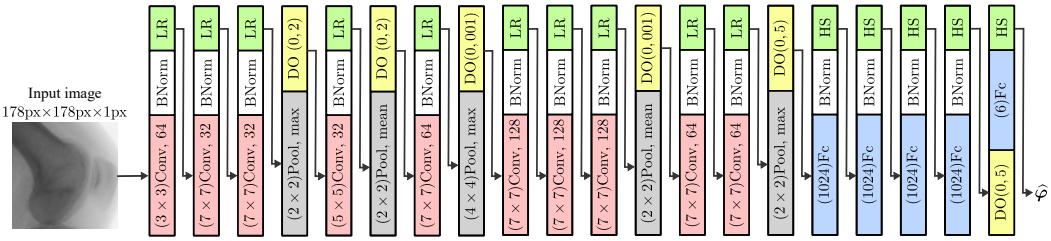


Figure 8. The optimal CNN architecture. Each rectangle represents one layer of CNN. The following colors are used to distinguish important elements of the network: blue (fully connected layer), green (activation functions, where HS stands for hard sigmoid, and LR denotes leaky ReLU), pink (convolution), purple (pooling), white (batch normalization), and yellow (dropout).

After 94 epochs of training, the early stopping rule was met and the learning process was terminated. The loss function of development set was equal to 8.4507 px^2. The results for all learning sets are gathered in Table 2.

Table 2. CNN loss function (7) values for different learning sets.

Learning Set	Proposed Solution	U-Net [23] (with Heatmaps)
Train	7.92 px^2	9.04 px^2
Development	8.45 px^2	10.31 px^2
Test	6.57 px^2	6.43 px^2

Loss function values for all learning sets are within acceptable range, given the overall complexity of the assigned task. The performance was slightly better for the train set in comparison to the development set. This feature usually correlates to overfitting of train data. Fortunately, low test set loss function value clarified that the network performance is accurate for previously unknown data.

Interestingly, test set data achieved the lowest loss function value, which is not common for CNNs. There may be several reasons for that. First, X-ray images used during training were of slightly different distribution than those from the test set. The train set consisted of images of children varying in age and, consequently, of a different knee joint ossification level, whereas the test set included adult X-rays. Second, train and development sets were augmented using typical image transformations, to constitute a valid CNN learning set (as described in Table 1). The corresponding loss function values in Table 2 are calculated for augmented sets. Some of the image transformations (randomly selected) resulted in high contrast images, close to binary. Consequently, those images were validated with high loss function value, influencing the overall performance of the set. On the other hand, the test set was not augmented, i.e., X-ray images were not transformed before the validation.

The optimization of the *hyperparameters* of CNN, as described in Appendix A, improved the process of network architecture tuning, in terms of processing time as well as low loss function value (7). The optimal network architecture (optimal in the sense of minimizing the assumed criterion (7)) consists of convolution layers with different window sizes, for convolution and for pooling layers. It is not consistent with the widely popular heuristics of small window sizes [33]. In this particular scenario, small window sizes in

CNN resulted in higher loss function or exceeded the maximum network size limited by the hardware restrictions.

Several regularization techniques were implemented, enabling the long-term learning process and avoiding overfitting of the goal function. For instance, the probability of dropout was high, especially in the deep layers of the network. Additionally, the most effective activation function was leaky ReLU [34]. The other well-known and widely popular activation function ReLU was also considered, nevertheless, it was Leaky ReLU that was chosen in all network layers.

Interestingly, the pooling layer type in this optimal network architecture alternates between mean and max pooling. Therefore, after each convolution layer, the pooling layer sharpens the features (max) or smoothing them (mean).

As an additional evaluation of the proposed algorithm, we compare its performance with an alternative solution. Based on studies [12] we apply U-Net [23] to regress heatmaps corresponding to keypoints $\boldsymbol{k}_1, \ldots, \boldsymbol{k}_3$. Keypoints heatmaps were created centering normal distribution at keypoint positions, normalized to maximum value of 1, with standard deviation equal to 1.5.

Original U-Net architecture [23] was used in this comparison. Note that, the input image is grayscale with resolution 572 px $\times$ 572 px; therefore, the whole X-ray image, within the limits of the fluoroscopic lens, is fed to the network.

The results of applying U-Net on X-ray images considered in this study are gathered in Table 2. It is evident that our proposed solution guaranteed lower loss function values in comparison with U-Net. Admittedly, U-Net performance was superior for images in the test set, but the difference is neglectable.

3.2. LA Estimation

The overall result of the LA estimation for all subjects from train and development sets (as described in Table 1) are gathered in Figure 9. Test set results will be discussed in the next section. Since no significant translational errors were noticed, only LA orientation errors are presented. The LA orientation error is considered as a difference between the angle θ_m, obtained from manually marked keypoints (using Equation (5)) and orientation θ_e obtained from estimated keypoints (using Algorithm 1).

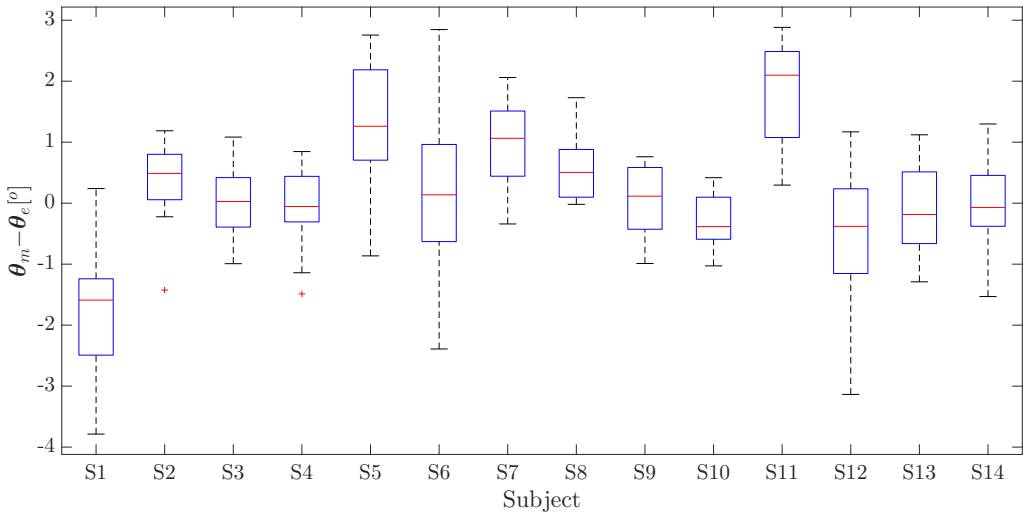


Figure 9. RMSE between the estimated and reference femur orientation.

The accuracy is defined by a root mean square error (RMSE). The red line in Figure 9 represents the median of the data, whereas the blue rectangles represent the interquartile range (between the first and third quartiles). The dashed line represents the data outside of this range, with several outliers denoted as red plus sign. The error median fits within

range $(-1.59°, 2.1°)$. The interquartile range for all subjects is relatively low, and the error rates are close to median values, therefore the diversity of error values is low.

The estimation of the LA orientation is of decent precision. The absolute value of orientation angle is lower than 4° for all image frames. The highest error corresponds to those image frames, which were slightly blurry and/or the bone shaft was just partially visible. Given the overall quality of the images, the error is negligible.

What is worth pointing out, Algorithm 1 resulted in a valid outcome after only one iteration, for most of the image frames. Therefore, the initial empirically chosen image window size $s = 25$ was reasonable for plenty of image frames. Nevertheless, 8 out of 14 subject images were thresholded with different window sizes. According to the adaptive thresholding technique, smaller window sizes were chosen for clear object borders, whereas bigger window sizes for more blurry images. Different s values reflect the differences in image quality and the bone age of each subject.

3.3. Femur Configuration Estimation (Test Stage)

In this section, we present the combined performance of both the LA and PS estimator, to evaluate the femur configuration on each X-ray image frame. Both estimators were designed and tuned using images from train and development sets, according to the description in Table 1. We assume that no further changes will be made in the architecture as well as parameter values of both estimators, once the training phase is finished.

In the test stage, we will evaluate the performance of the estimators on new data, not used during training, i.e., included in the test set.

Remember that, the reference configuration of the femur $\boldsymbol{g}_m$ is calculated from positions of manually marked keypoints. The same set of transformations (5) is applied to both manually denoted and estimated keypoints, to calculate the configuration. The overall performance of the algorithm is defined as a difference between $\boldsymbol{g}_m$ and $\boldsymbol{g}_e$. The results for each configuration element separately are presented in Figure 10.

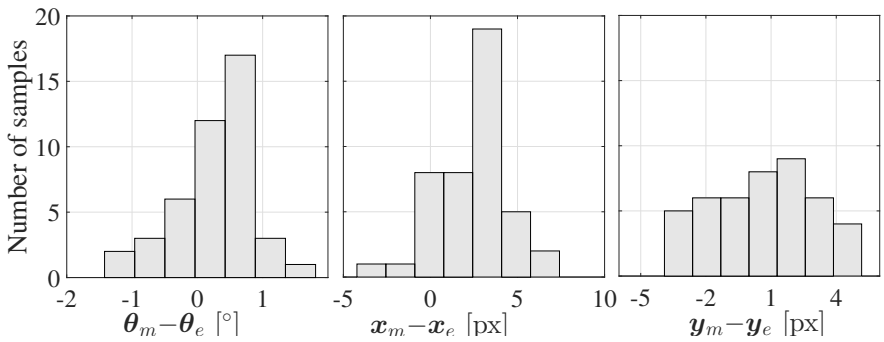


Figure 10. Femur configuration estimation results.

Position error is defined in pixels, whereas orientation is given in degrees. Note that the orientation error $(\theta_m - \theta_e)$ is purely dependent on the performance of the gradient-based estimator and the results correspond to the values presented in Figure 9. Therefore, the estimator detects LA keypoints on new image data with similar accuracy to the one observed in the training stage.

Position error combines the inaccuracies of both estimators, nevertheless proposed redundancy of keypoint selection causes slight robustness to those errors. Estimation errors of both position components of femur configuration is limited. The overall performance is satisfactory, given the size of the input image.

Interestingly, the femur coordinate center was swiped to the left $(x_e < x_m)$ on most X-ray image data, in comparison to manually denoted configuration. It could be interpreted as a systematic error of the estimator and could be canceled out in the forthcoming validations. However, the sources of error may be connected to the reference configuration, which is calculated for manually placed keypoints. This assumption could lead to the remark that CNN actually performed better than the human operator.

The results achieved by the proposed algorithm of femur configuration detection cannot be compared with any alternative solutions. The femur coordinate system proposed in this study was not incorporated in any outgoing or previous studies. Other authors proposed different representations [35,36], but those do not apply for this specific image data. As far as the author's knowledge is concerned, there are no alternative configuration detectors of the pediatric femur bone in the lateral view.

4. Discussion

In this work, we specified the feature set that unambiguously determines femur configuration, the defined corresponding image keypoints, and we constructed femur coordinate system derived from those features. Subsequently, we proposed the fully automatic keypoint detector. The performance of the algorithm was evaluated on new data and achieved satisfactory results.

The proposed set of features reflected the strict examination protocol and is only valid for two-dimensional image data. Admittedly, modern acquisition systems enable more informative image data (e.g., MRI). Then, image processing is less demanding, and higher accuracy can be obtained for the detection and/or classification task. The main motivation of our work was to change the balance between data acquisition and image processing. Therefore, we used lower quality image data (still present in plenty of medical facilities) but simultaneously lowered the fatigue of specific and fragile group of subjects, considered in this study. This forced us to design a more sophisticated and complex image processing algorithm.

Our image processing algorithm consisted of two estimators. One of them was based on CNN, and contrary to widely popular hand-engineering, we proposed to optimize network architecture automatically. The optimization algorithm accelerated largely the process of *hyperparameter* tuning. What is worth noticing, in the optimization process, at least 10 network architectures resulted in similar loss function values. We can explicitly state that the given estimation problem can be solved via CNN.

Both keypoint estimators work in parallel, and their result is used to evaluate the configuration of the femur. Each image frame is processed separately; therefore, no prior information is used to determine femur configuration. The important feature of this solution is that the error does not accumulate for images of one sequence, i.e., corresponding to one subject.

The main benefit of both estimators is the end-to-end learning pattern. In general, this type of solution processes the input image data faster and with lower computational costs than, e.g., image patch based evaluation [21]. Admittedly, the accuracy of the method is lower than for projects where three-dimensional data are available alongside two-dimensional data [37,38]. However, it is the input data quality responsible for this outcome, not the method itself.

Additionally, if three-dimensional data are not available, the segmented bone image may not be directly connected to the actual bone configuration. For example, out of plane rotation will influence the shape greatly. Therefore, simple segmentation methods [37] cannot be applied in this study.

The proposed algorithm of keypoint detection results in a decent accuracy, similar to [39,40]. Given the troublesome characteristics of images, we believe it is a success.

The whole algorithm of femur configuration detection resulted in a reliable outcome even for images of different distributions than training data. The train and development sets were mostly pediatric images. Two healthy adult subjects were introduced to increase the generality of the proposed solution. On the other hand, the test set was composed of merely adult subjects' images. In the future, it would be beneficial to validate the algorithm on a dataset composed of children's X-rays.

An important aspect of this work is the lack of ground truth in medical image data. The reference values used in this study were influenced by human error. Obtaining reliable reference data for keypoint detection still remains an open problem.

Funding: This research was partially supported by the statutory grant no. 0211/SBAD/0321.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Bioethics Committee of Poznan University of Medical Sciences (resolution 699/09).

Informed Consent Statement: Informed consent was obtained from legal guardians of all subjects involved in the study.

Acknowledgments: I would like to acknowledge Paweł Koczewski for invaluable help in gathering X-ray data and choosing the proper femur features that determined its configuration.

Conflicts of Interest: The author declares no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CNN	convolutional neural networks
CT	computed tomography
LA	long axis of femur
MRI	magnetic resonance imaging
PS	patellar surface
RMSE	root mean squared error

Appendix A

In this work, contrary to frequently used hand engineering, we propose to optimize the structure of the estimator through a heuristic random search in a discrete space of *hyperparameters*. The *hyperparameters* will be defined as all CNN features selected in the optimization process. The following features are considered as *hyperparameters* [26]: number of convolution layers, number of neurons in each layer, number of fully connected layers, number of filters in convolution layer and their size, batch normalization [29], activation function type, pooling type, pooling window size, and probability of dropout [28]. Additionally, the batch size $\boldsymbol{X}$ as well as the learning parameters: learning factor, cooldown, and patience, are treated as *hyperparameters*, and their values were optimized simultaneously with the others.

What is worth noticing—some of the *hyperparameters* are numerical (e.g., number of layers), while the others are structural (e.g., type of activation function). This ambiguity is solved by assigning individual dimension to each *hyperparameter* in the discrete search space. In this study, 17 different *hyperparameters* were optimized [26]; therefore, a 17-th dimensional search space was created. A single architecture of CNN, denoted as $\mathcal{M}$, is featured by a unique set of *hyperparameters*, and corresponds to one point in the search space.

The optimization of the CNN architecture, due to the vast space of possible solutions, is achieved with the tree-structured Parzen estimator (TPE) proposed in [41]. The algorithm is initialized with n_s start-up iterations of random search. Secondly, in each k-th iteration the *hyperparameter* set $\mathcal{M}_k$ is chosen, using the information from previous iterations (from 0 to $k-1$). The goal of the optimization process is to find the CNN model $\boldsymbol{\mathcal{M}}$, which minimizes the assumed optimization criterion (7).

In the TPE search, the formerly evaluated models are divided into two groups: with low loss function (20%) and with high loss function value (80%). Two probability density functions are modeled: G for CNN models resulting with low loss function, and Z for high loss function. The next candidate $\mathcal{M}_k$ model is chosen to maximize the Expected Improvement (EI) ratio, given by:

$$EI(\mathcal{M}_k) = \frac{P(\mathcal{M}_k \in G)}{P(\mathcal{M}_k \in Z)}. \tag{A1}$$

TPE search enables evaluation (training and validation) of $\mathcal{M}_k$, which has the highest probability of low loss function, given the history of search. The algorithm stops

after predefined n iterations. The whole optimization process can be characterized by Algorithm A1.

Algorithm A1: CNN structure optimization

Result: $\boldsymbol{\mathcal{M}}$, $\boldsymbol{L}$
Initialize empty sets: $\boldsymbol{L} = \varnothing$, $\boldsymbol{\mathcal{M}} = \varnothing$;
Set n and $n_s < n$;
for $k = 1$ **to** $n\_startup$ **do**
  Random search $\mathcal{M}_k$;
  Train $\mathcal{M}_k$ and calculate L_k from (7);
  $\boldsymbol{\mathcal{M}} \cup \mathcal{M}_k$; $\boldsymbol{L} \cup L_k$;
end
for $k = n_s$ **to** n **do**
  Sort $\boldsymbol{\mathcal{M}}$ is increasing order of $\boldsymbol{L}$;
  Model density $G(\mathcal{M}_j)$, $j = 1, \ldots, 0.2k$ with TPE;
  Model density $Z(\mathcal{M}_j)$, $j = 0.2k, \ldots, k$ with TPE;
  Choose $\mathcal{M}_k$ with maximum EI (A1);
  if $\mathcal{M}_k \notin \boldsymbol{\mathcal{M}}$ **then**
    Train $\mathcal{M}_k$ and calculate L_k from (7);
    $\boldsymbol{\mathcal{M}} \cup \mathcal{M}_k$; $\boldsymbol{L} \cup L_k$;
  end
end

The presented optimization returns the set of the evaluated model architectures:

$$\boldsymbol{\mathcal{M}} \triangleq \{\mathcal{M}_k\}_{k=1}^{n}, \tag{A2}$$

together with corresponding loss function values

$$\boldsymbol{L} \triangleq \{L_k\}_{k=1}^{n}, \tag{A3}$$

in the increasing order. The first set of *hyperparameters*, i.e., $\mathcal{M}_1$ assures the minimum value of loss function, at least among all of the $\mathcal{M}_k$ that are stored in $\boldsymbol{\mathcal{M}}$. As a side note, the TPE-based search algorithm enables covering the whole search space for a large number of iterations n. In this particular scenario, for 17-th dimensional search space, the number of all possible CNN architectures is approximately equal to $20{,}592 \times 10^{12}$. Covering the whole search space would be a very time consuming task. With the TPE search algorithm, we are able to accelerate the optimization process.

References

1. Gyles, C. Robots in medicine. *Can. Vet. J.* **2019**, *60*, 819–820. [PubMed]
2. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.; Ciompi, F.; Ghafoorian, M.; Van Der Laak, J.A.; Van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [CrossRef] [PubMed]
3. LERA—Lower Extremity RAdiographs Dataset. Available online: https://aimi.stanford.edu/lera-lower-extremity-radiographs-2 (accessed on 1 April 2019).
4. Bay, H.; Tuytelaars, T.; Van Gool, L. SURF: Speeded Up Robust Features. In *Computer Vision—ECCV 2006*; Leonardis, A., Bischof, H., Pinz, A., Eds.; Springer: Berlin/Heidelberg, Germany, 2006; pp. 404–417.
5. Parkhi, O.M.; Vedaldi, A.; Zisserman, A.; Jawahar, C.V. Cats and dogs. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 3498–3505.
6. Geiger, A.; Lenz, P.; Stiller, C.; Urtasun, R. Vision meets Robotics: The KITTI Dataset. *Int. J. Rob. Res.* **2013**, *32*, 1231–1237.
7. MNIST Handwritten Digit Database. Available online: http://yann.lecun.com/exdb/mnist/ (accessed on 14 September 2021).
8. Fonseca, A.U.; Vieira, G.S.; Soares, F.A.A.M.N.; Bulcão-Neto, R.F. A Research Agenda on Pediatric Chest X-Ray: Is Deep Learning Still in Childhood? *arXiv* **2020**, arXiv:2007.11369.
9. Liu, X.; Song, L.; Liu, S.; Zhang, Y. A Review of Deep-Learning-Based Medical Image Segmentation Methods. *Sustainability* **2021**, *13*, 1224. [CrossRef]

10. Zheng, Y.; Liu, D.; Georgescu, B.; Nguyen, H.; Comaniciu, D. 3D deep learning for efficient and robust landmark detection in volumetric data. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015.
11. Suzani, A.; Rasoulian, A.; Seitel, A.; Fels, S.; Rohling, R.N.; Abolmaesumi, P. Deep learning for automatic localization, identification, and segmentation of vertebral bodies in volumetric MR images. In Proceedings of the Medical Imaging 2015: Image Perception, Observer Performance, and Technology Assessment, Orlando, FL, USA, 25–26 February 2015.
12. Payer, C.; Stern, D.; Bischof, H.; Urschler, M. Regressing heatmaps for multiple landmark localization using CNNs. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Athens, Greece, 17–21 October 2016.
13. Khalid, H.; Hussain, M.; Al Ghamdi, M.A.; Khalid, T.; Khalid, K.; Khan, M.A.; Fatima, K.; Masood, K.; Almotiri, S.H.; Farooq, M.S.; et al. A Comparative Systematic Literature Review on Knee Bone Reports from MRI, X-rays and CT Scans Using Deep Learning and Machine Learning Methodologies. *Diagnostics* **2020**, *10*, 518.
14. Bayramoglu, N.; Tiulpin, A.; Hirvasniemi, J.; Nieminen, M.T.; Saarakkala, S. Adaptive segmentation of knee radiographs for selecting the optimal ROI in texture analysis. *Osteoarthr. Cart.* **2020**, *28*, 941–952. [CrossRef]
15. Lundervold, A.S.; Lundervold, A. An overview of deep learning in medical imaging focusing on MRI. *Z Med. Phys.* **2019**, *29*, 102–127. [CrossRef]
16. Cerveri, P.; Belfatto, A.; Manzotti, A. Predicting Knee Joint Instability Using a Tibio-Femoral Statistical Shape Model. *Front. Bioeng. Biotechnol.* **2020**, *8*, 253. [CrossRef]
17. Zheng, Q.; Shellikeri, S.; Huang, H.; Hwang, M.; Sze, R.W. Deep Learning Measurement of Leg Length Discrepancy in Children Based on Radiographs. *Radiology* **2020**, *296*, 152–158. [CrossRef]
18. Iglovikov, V.I.; Rakhlin, A.; Kalinin, A.A.; Shvets, A.A. Paediatric Bone Age Assessment Using Deep Convolutional Neural Networks. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2018; Volume 11045, pp. 300–308.
19. Koitka, S.; Demircioglu, A.; Kim, M.S.; Friedrich, C.M.; Nensa, F. Ossification area localization in pediatric hand radiographs using deep neural networks for object detection. *PLoS ONE* **2018**, *13*, e0207496.
20. Chen, R.; Ma Y.; Liu L.; Chen N.; Cui Z.; Wei G.; Wang W. Semi-supervised Anatomical Landmark Detection via Shape-regulated Self-training. *arXiv* **2021**, arXiv:2105.13593.
21. Song, Y.; Qiao, X.; Iwamoto, Y.; Chen, Y.-W. Automatic Cephalometric Landmark Detection on X-ray Images Using a Deep-Learning Method. *Appl. Sci.* **2020**, *10*, 2547. [CrossRef]
22. Yeh, Y.C.; Weng, C.H.; Huang, Y.J.; Fu, C.J.; Tsai, T.T.; Yeh, C.Y. Deep learning approach for automatic landmark detection and alignment analysis in whole-spine lateral radiographs. *Sci. Rep.* **2021**, *11*, 7618. [CrossRef] [PubMed]
23. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Navab N., Hornegger J., Wells W., Frangi A., Eds.; Springer: Cham, Switzerland, 2015; pp. 234–241.
24. Chisholm, C.; Mak, D.; Thyagarajan, M. Imaging Paediatric Joint Effusions: Techniques, Findings and Pitfalls. In Proceedings of the Congress of European Society of Musculoskeletal Radiology, York, UK, 18–20 June 2015; p. 0124.
25. Wijdicks, C.A. Radiographic Identification of the Primary Medial Knee Structures. *J. Bone Jt. Surg.* **2009**, *91*, 521–529. [CrossRef]
26. Drazkowska, M.; Gawron, T.; Kozlowski, K. Application of Convolutional Neural Networks to Femur Tracking in a Sequence of X-ray Images. In Proceedings of the 7th IEEE International Conference on Biomedical Robotics and Biomechatronics (Biorob), Enschede, The Netherlands, 26–29 August 2018; pp. 49–54.
27. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. In Proceedings of the International Conference on Learning Representations, Banff, AB, Canada, 14–16 April 2014.
28. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
29. Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv* **2015**, arXiv:1502.03167.
30. Bradley, D.; Roth, G. Adapting Thresholding Using the Integral Image. *J. Graph. Tools* **2007**, *12*, 13–21. [CrossRef]
31. Gray, L. A Mathematician Looks at Wolfram's New Kind of Science. *Not. Am. Math. Soc.* **2003**, *50*, 200–211.
32. Yan, X.; Su, X.G. *Linear Regression Analysis: Theory and Computing*, 1st ed.; World Scientific Publishing Company: Singapore, 2009.
33. Stanford Computer Science Class CS231n: Convolutional Neural Networks for Visual Recognition. Available online: http://cs231n.github.io/ (accessed on 22 October 2020).
34. Maas, A.L.; Hannun, A.Y.; Ng, A.Y. Rectifier nonlinearities improve neural network acoustic models. In Proceedings of the International Conference on Machine Learning, Atlanta, GA, USA, 16–21 June 2013.
35. Rood, J.E.; Stuart, T.; Ghazanfar, S.; Biancalani, T.; Fisher, E.; Butler, A.; Hupalowska, A.; Gaffney, L.; Mauck, W.; Eraslan, G.; et. al. Toward a Common Coordinate Framework for the Human Body. *Cell* **2019**, *179*, 1455–1467. [CrossRef]
36. Fischer, M.C.M.; Grothues, S.A.G.A.; Habor, J.; de la Fuente, M.; Radermacher, K. A robust method for automatic identification of femoral landmarks, axes, planes and bone coordinate systems using surface models. *Sci. Rep.* **2020**, *10*, 20859. [CrossRef] [PubMed]

37. Wu, J. 2D-3D Registration of Knee Joint from Single Plane X-ray Fluoroscopy Using Nonlinear Shape Priors. Ph.D. Thesis, University of Tennessee, Knoxville, TN, USA, 2016.
38. Ambellan, F.; Tack, A.; Ehlke, M.; Zachow, S. Automated segmentation of knee bone and cartilage combining statistical shape knowledge and convolutional neural networks: Data from the Osteoarthritis Initiative. *Med. Image Anal.* **2019**, *52*, 109–118. [CrossRef] [PubMed]
39. Feng, Z.H.; Kittler, J.; Awais, M.; Huber, P.; Wu, X.J. Wing loss for robust facial landmark localisation with convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018.
40. Bier, B.; Aschoff, K.; Syben, C.; Unberath, M.; Levenston, M.; Gold, G.; Fahrig, R.; Maier, A. Detecting Anatomical Landmarks for Motion Estimation in Weight-Bearing Imaging of Knees. In *MLMIR 2018: Machine Learning for Medical Image Reconstruction*; Lecture Notes in Computer Science; Knoll F., Maier A., Rueckert D., Eds.; Springer: Cham, Switzerland, 2018; Volume 11074.
41. Bergstra, J.S.; Bardenet, R.; Bengio, Y.; Balázs, K. Algorithms for Hyper Parameter Optimization. In *Advances in Neural Information Processing Systems 24*; Shawe-Taylor, J., Zemel, R.S., Bartlett, P.L., Pereira, F., Weinberger, K.Q., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2011; pp. 2546–2554.