

To get to the point

Neural Network application to key-point
detection in radiographs

Physicians have to locate so called key-points e. g. for surgical procedures. Up to now, this was always done manually. In order to automate this process, innovative software was developed that uses artificial intelligence (AI) combining a clipping-window approach with the newly developed prediction shifting. The program can predict the key-points with a high degree of accuracy-making the AI as precise as a physician.

DER JUNGFORSCHER

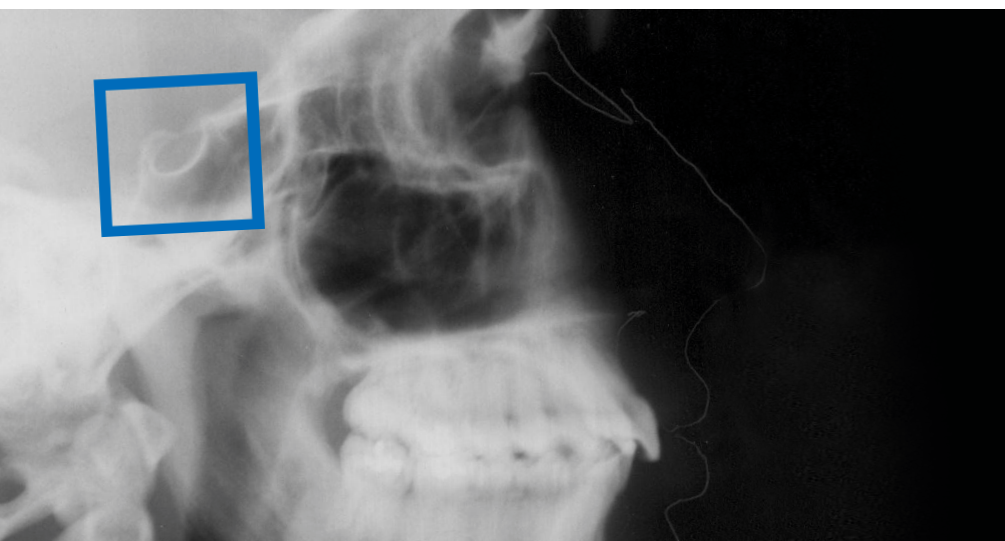


**Constantin Tilman Schott
(2003)**

Paul-Gerhardt-Schule
Dassel

Eingang der Arbeit:
6.10.2019

Arbeit angenommen:
6.12.2019



To get to the point

Neural Network application to key-point detection in radiographs

1. Introduction

The results of the X-ray image analysis provide the basis for medical diagnostics and treatment planning. Up to now, these analyses are performed manually, making them a drain on time and human resources, and an expensive task in the specialized medical field of X-ray image analysis. In light of these facts, this paper will explore the option of a standardized, computer-based analysis of X-ray images.

There are a variety of automated, computer-aided approaches used to extract information from images – finding “features” in images. The best results here are achieved by artificial intelligence in the form of convolutional neural networks (CNNs) [2]. Among other benefits, CNNs have

made it possible to achieve an extreme increase in efficiency and accuracy in automated detection of structures in recent years [3, 9]. This can be seen exemplarily at the MNIST database [6] which is used frequently to evaluate the quality of image recognition methods using the example of handwritten digit recognition.

These CNNs have already started being used to deal with classification tasks in X-ray diagnostics. For example, you can analyze a mammography for cancer foci [13]. There are already many examples of CNNs for chest x-rays, even commercial products (oxipit).

In addition to these classification tasks, in many X-ray image analyses (cranium,

hips etc.), specific structures have to be marked with precise points. This method is used in almost all orthopaedic or surgical X-ray analyses. The angles of these points in relation to each other and their distances from each other help the physician make a specific diagnosis. This process is known as predictive analytical key-point detection. The goal of this project is to use CNNs to automate this process in order to enable fully autonomous analysis in other areas of X-ray diagnostics over the long term. This involves applying various CNN structures (existing ones as well as self-developed ones) to the radiological analysis (X-ray image analysis) and drawing comparisons between them.

There are already AI methods that include localization (using facial recognition, for example) [14]. This paper discusses the applicability of these methods to the issue at hand here. In addition, new methods are being developed that allow for more accurate key-point detection.

In this paper the cephalometry (measurement of the cranium) is examined as a medical subfield, in which key-point detection is the mainly used and most important analysis technique. Cephalometry is used in areas such as orthodontic diagnostics and therapy planning, aesthetic surgery planning and, sometimes, in post-traumatic reconstructive surgery planning.

As a starting point, methods are developed in this paper to automatically locate the so-called Sella point, that was chosen for its great importance for the analysis. It is used as a datum point for almost all cephalometric analyses and is the center of the Sella turcica (Latin for “Turkish seat”). The Sella turcica is located in the centre of the cranium base and, as such, is a nearly constant datum point for these analyses, regardless of factors such as growth-related or traumatic changes to the bony viscerocranium.

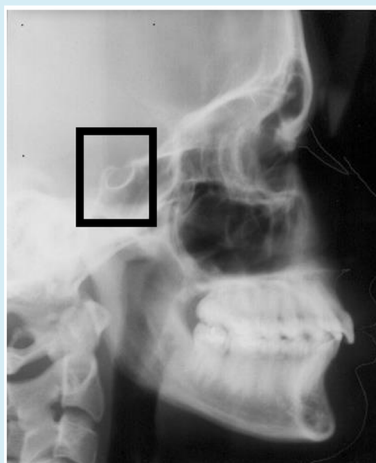


Fig. 1: Sella turcica in lateral cephalogram

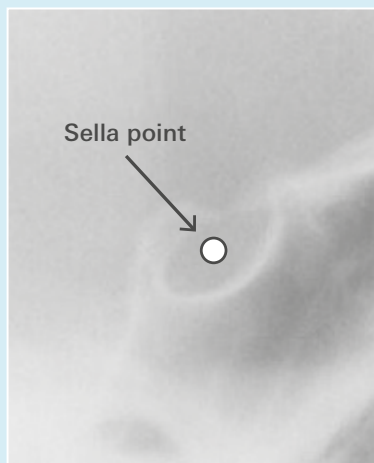


Fig. 2: Sella turcica with Sella point drawn in

After successfully automating this key-point-analysis of the Sella point, the developed method should be transferred on other important points of the cephalometric X-ray analysis.

2. Preliminary considerations and methodology

2.1 Problem statement

As mentioned above, the Sella turcica is used on an X-ray image of the side of the cranium (lateral cephalogram) as an application example. On the two-dimensional projection of the cranium using a lateral cephalogram, the Sella turcica can be identified as an oval opening upward. The Sella point (S-point) “is defined as the (geometric) center of the bony crypts of the Sella turcica” [15] (Fig. 1 and 2).

The boundaries of the structure must be identified individually from one lateral cephalogram to the next. The shape

of the Sella turcica also varies (Fig. 3). The purpose of automated analysis is to yield a reliable result, comparable to a professional medical analysis, in this complex initial situation.

2.2 Basic AI methodology

There are three basic possible application areas for the CNNs used in this paper:

1. Classification of images (e.g. binary classification → output between 0 and 1)
2. Key-point detection (→ output of a coordinate of the point(s))
3. Image segmentation (→ output of an image on which, for example, the detected features are drawn)

In this work, these three application options (sometimes in combination) are used to locate the Sella points.

On top of that, the following basic AI methods were used in this work in the experiments:

In order to reduce overfitting dropout was used. This involves randomly deactivating a fixed percentage of neurons in every learning step in the respective layer it is being applied to and optimizing the network without these neurons [12]. This distributes the learning process across more neurons and increases the probability of detecting relevant patterns and structures, and thus counteracts overfitting. The training loss converges more slowly as a consequence of dropout, since it is not the entire network that is learning simultaneously.

Furthermore, data-augmentation was applied:

In the lateral cephalogram analysis, the points to be found cumulate heavily. As a result, a prediction of the middle of the point cloud that is independent of the input might be enough for a low deviation with respect to the individual lateral cephalogram image (Fig. 4).

In the actual application example, the original images (without data augmentation) are shifted by a random value while making sure, that the Sella itself stays within the image boundaries (Fig. 5).

As a result, the network is forced to learn only the truly relevant correlations between inputs. If the network optimized to the augmented input is then tested on normal application examples without augmentation, the

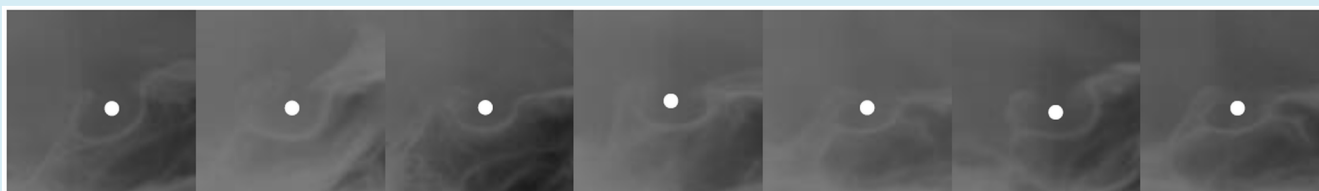


Fig. 3: Various Sella turcica structures with Sella point marked

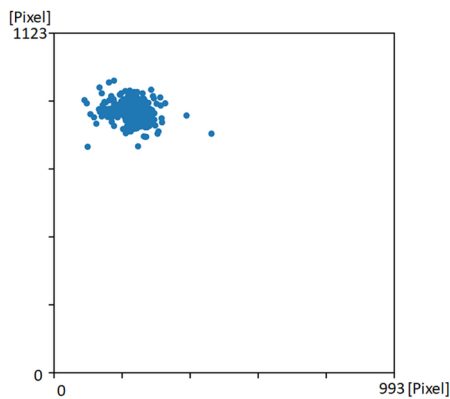


Fig. 4: Distribution of Sella points (without data augmentation)

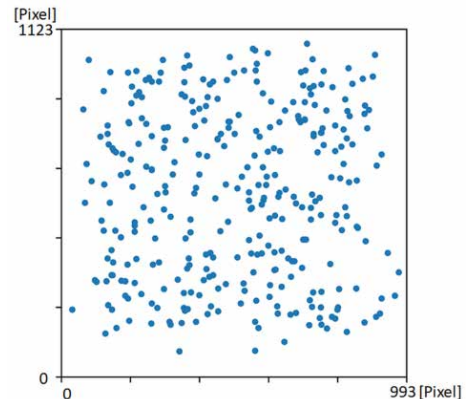


Fig. 5: Distribution of Sella points (with data augmentation)

network should detect them with a higher degree of accuracy compared to a net without image augmentation.

On top of that, histogram equalization was utilized. Lateral cephalogram images sometimes have significantly different contrast values. These values can be offset by a histogram equalization [5] (Fig. 6).

For all the nets the hyperparameters used were optimized in a random search [19].

These hyperparameters include the size of the batches the training dataset was split into, the number of convolutional layers (and their number of kernels and their respective sizes), the max-pooling layer kernel-sizes, the amount of dropout, the size of the fully-connected layers on top (if used), the loss function

and optimization function (with learning rate etc.). For all nets the Adam [8] optimization function and the ReLu activation function [4, 17] turned out to be optimal.

2.3 Tools

Programming was performed on a computer with Intel i7-8700-CPU, 32 GB DDR4-RAM and a Nvidia Geforce GTX 6000Ti with 6 GB RAM and with an Ubuntu 18.04 LTS operating system. The programming environment consisted of a Jupyter notebook [7] with browser-based server-client architecture, where the server used the local host environment via a loopback. The programming language used was “Python” combined with the “TensorFlow” library with GPU support and the “Keras” deep learning library with TensorFlow backend. The

calculated data were analyzed and graphically depicted using the Python Matplot library [18].

All the programs used for this paper are stored as a GitHub project at https://github.com/tinotil/DeepLearning_FRS_JuFo_2019.

2.4 Data

420 anonymized lateral cephalograms of the side of the cranium (FRS) analyzed by medical specialists are the basis of all the following analyses. The pictures themselves were taken from a private orthodontic doctor's office. The optimal points for each cephalogram are the geometric mean of the points of two doctors with an experience in this kind of x-ray image analysis of over 25 years.

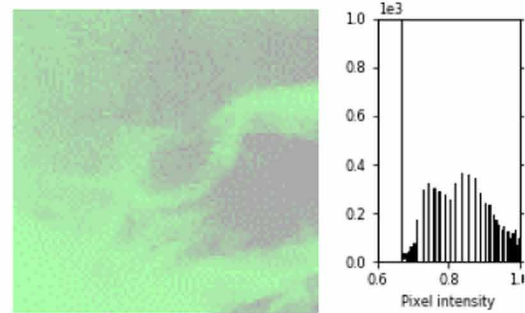
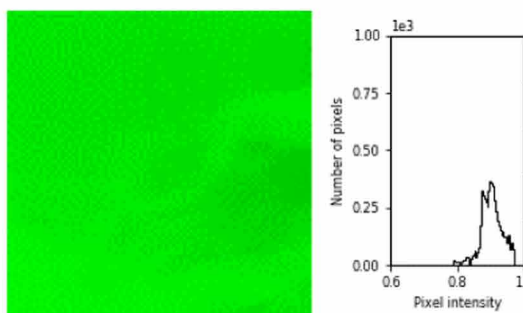


Fig. 6: Comparison of a Sella image with weak contrast before (left) and after a histogram equalization (right) and the associated pixel intensity histograms

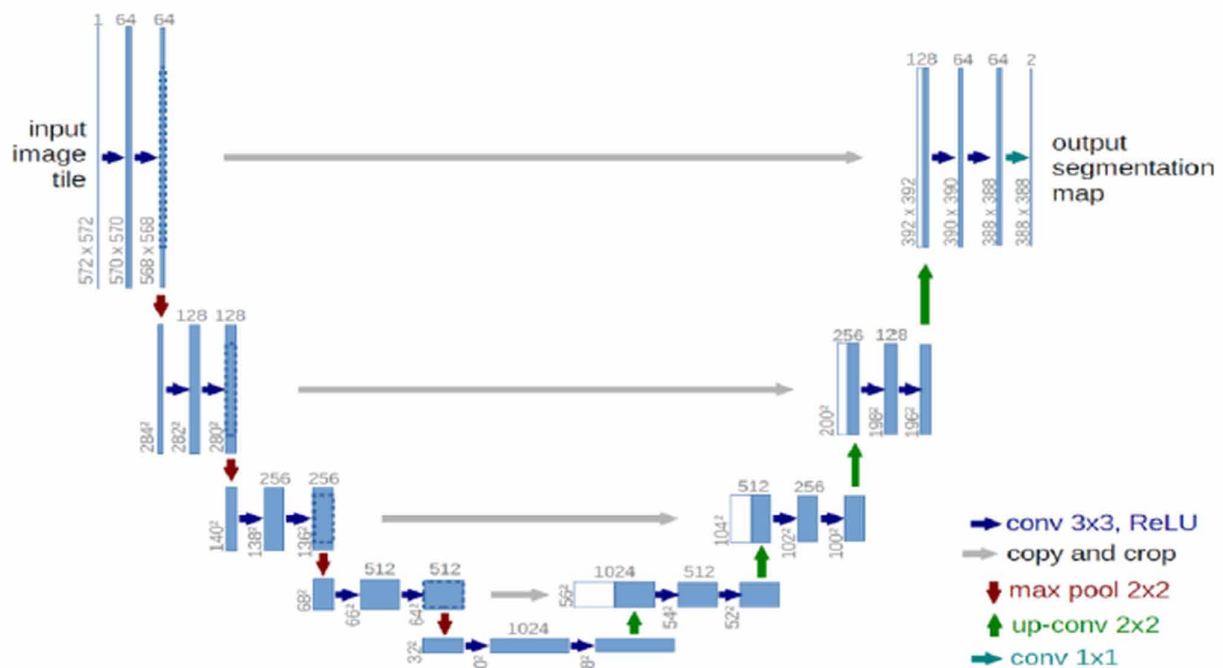


Fig. 7: Example structure of the U-Net (according to Ronneberger et al. [10])

300 of the images were part of the training data set, 100 made up the validation data set and 20 made up the test data set. While this assignment remains the same for a learning cycle (first to last epoch), the images are randomly re-assigned to these data sets (cross-validation) before each subsequent learning cycle (weights reset).

The original image has a size of 993×1123 pixel with 150 dpi.

The pieces of data listed below each represent the mean values of five learning cycles (5-fold CV).

3. Evolution of approaches

3.1 Image segmentation: Heatmap approach

A heatmap method is used as the first approach for automated localization of the Sella point on a lateral cephalogram. This process is based on facial recognition CNNs, where heatmaps

enable the detection of multiple faces at once [14].

In medical applications comparable approaches are also already used in the Biomedical Image Segmentation and feature Localization with e.g. so-called U-Nets [10]. Such a U-Net structure was used as a first approach. In this structure, another part with up-convolutions (expansive path) was attached to the classic CNN (contracting path). For finding the S-point, the lateral cephalograms were used as the input and a corresponding image matrix with the maximum activation at the optimal point was used as the output. In addition, the pixels around the optimal point were activated to the effect of a statistical standard distribution in order to enable gradual convergence to the optimal point. The optimal scaling factor of this Gaussian was experimentally determined and then fixed for all further experiments. This creates what is known as the heatmap for training. The point of maximum activation of the output-heatmap gets

interpreted as the predicted S-point.

As shown in [Figure 7](#), a multi-channel feature map is created in the left half (CNN), which then gets turned back into an output image through up-sampling and other convolutions in the right half.

3.2 Key-point detection network using entire lateral cephalogram

3.2.1 Pre-trained key-point detection network

As a second approach a CNN was used and its output feature maps are interpreted by a DenseNet, which itself outputs a 2×1 vector - the predicted S-point coordinate.

The deep learning library Keras offers different pre-trained CNNs whose weights have already been optimized to certain applications, and are therefore able to detect basic structures reliably. During an initial test, a pre-trained

VGG16 network [11] was used with a two-layer DenseNet.

The coordinates of the points are standardized for all key-point detections. In doing so, each coordinate value was divided by the width of the respective image edge (x-coordinate/width of the image, y-coordinate/height of the image). As a result, the expected outputs were limited to a range between 0 and 1. Here, only the DenseNet and the top two layers of the VGG16 network were trained. This was done to prevent overwriting weights of the lower layers that had already been optimized and to reduce the training time. The lateral cephalogram images were used as input data and the standardized coordinates of the S-points were used as output. As the VGG16 network uses RGB images as input the grayscale channel values were duplicated to generate a three-channel picture. These input images were also rescaled using PIL (Python Image Library) in order to fit the VGG16 input dimensions.

Dropout was utilized as needed increasing from 0.2 up to 0.7.

3.2.2 Custom key-point detection network

In the next step, the pretrained VGG16 network was replaced by a custom trained, smaller network with only 4 convolutional layers and two dense layers on top.

In this example data augmentation was used extensively varying between a random shift over 3/4 or 1/2 of the output range.

In order to enable a learning process the pictures had to be down-sampled to an input size of 150×150 pixels from an original size of 993×1123 pixels. A larger input cannot be implemented in this key-point detection network due to hardware limitations.

3.3 Clipping window approach

An improved approach for precisely detecting the S-point is based on the process of Ciresan et al. [1]. The image is

evaluated on two levels using a clipping window approach. First, a classification CNN classifies selected partial images according to the categories “Sella” or “not Sella”. In the second step, a key-point detection network sets the S-point on the classified images of a Sella turcica.

Since only a section of the lateral cephalogram image is input (partial image size 150×150 pixels) into a network, down-sampling no longer needs to be applied to the input lateral cephalogram image. As a result, a lateral cephalogram image that is 750×750 pixels can be used as an initial input, thereby increasing the information density.

3.3.1 Classification network

A simple CNN with only 3 convolutional layers and two interpreting Dense layers was used as a classification network. In this process, an output of 0 was assigned for Sella and 1 for non-Sella. Therefore, a sigmoid function was used as an activation function of the output layer. Since all outputs greater

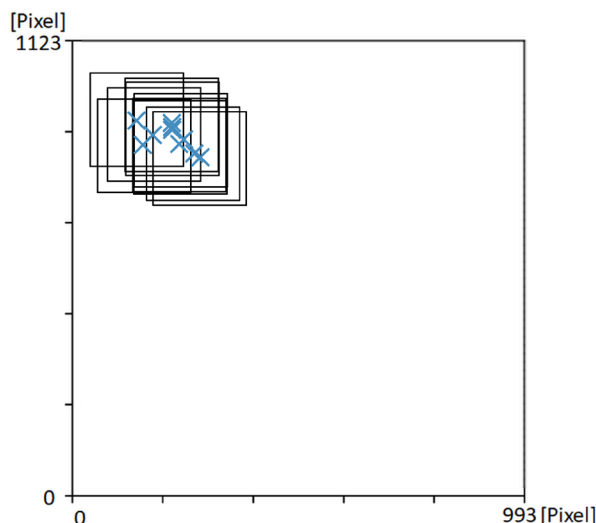


Fig. 8: Example of generating partial images based on the Sella points that have already been identified in the entire lateral cephalogram (with 10 points for purposes of simplification)

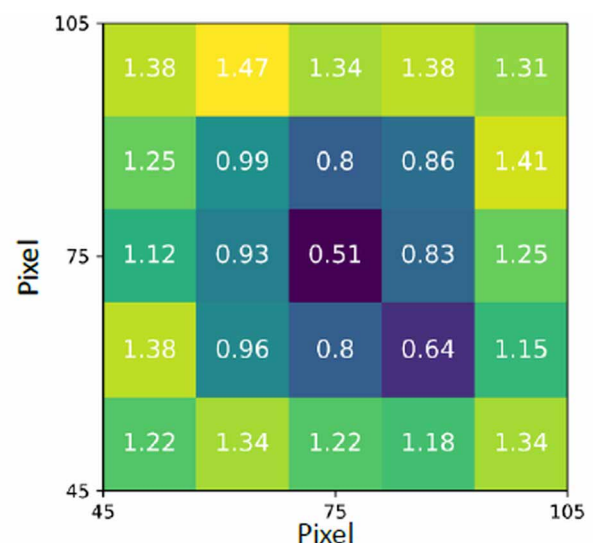


Fig. 9: Deviations of predictions from the expected values in mm for 20 test Sella images, depending on the position of the ground truth point on the Sella partial image (image center: 75|75 pixels)

than 0.5 can be interpreted as 1 and all those less than 0.5 can be interpreted as 0 in the binary classification of this case example, the model is not as susceptible to overfitting. For learning purposes, the corresponding Sella image was generated from each lateral cephalogram with the set Sella point as the middle point. The non-Sella images were able to be generated by randomly selecting partial images from the X-ray images. The edges of the non-Sella images adhered to a minimum distance of 75 pixels from the set Sella point.

To ensure practical use of the network, a method must be developed that divides the entire input lateral cephalogram into partial images that are then divided into categories of “Sella” and “non-Sella” by the classification network. There are a variety of search patterns that can be used to accomplish this. A square with a side length that is 1/5 of the total image height is used as the size of these images.

The a priori knowledge of the accumulation of S-points on the lateral cephalogram images (Fig. 4) can be used to generate pictures of a high probability of containing the Sella.

First, all the manually set Sella points that have already been analyzed once are used as centres of the generated images, which are then fed into the classification network as input to be classified (Fig. 8).

As already mentioned, the classification network uses a sigmoid function as an activation function of the output layer. Unlike functions such as the heaviside step function, the advantage here is that differences in equivalence between the input and the learned patterns are depicted. Despite this, a clear distinction (0 or 1) is favored through a larger gradient (major change in each learning step) in the value range between 0 and 1. Therefore, in the clipping window approach, the network can assess the degree of similarity to the Sella that the input image depicts. Accordingly,

only the image with the highest degree of equivalence to a Sella is ultimately selected.

If none of the partial images output using the described search method yield an equivalence of at least 99.9 %, a simple overlapping sliding window process (fixed offset in x- and y-directions) is used.

3.3.2 Key-point detection network using Sella partial image

Again, a key-point detection network outputs the exact coordinate of the Sella point. This time, however, it is not operating on the entire lateral cephalogram, but only on the partial image containing the Sella structure, that has been positively classified by the classification net.

Image augmentation, an even distribution of optimal points in the output area, is absolutely necessary for the training set, since the images were created for the training data set based on the optimal point. Without image augmentation, the S-point would always be in the center of the image.

3.3.3 Key-point detection network using Sella partial image with prediction shifting

A consideration of the average error, depending on the position of the Sella on the partial image (based on the optimal points), results in the following distribution (Fig. 9).

The prediction becomes more accurate the more centrally the Sella is located. Between the edge areas and the center, the deviations can be seen to be reduced by more than half. This can be correlated with the higher information density of the central areas (see “valid padding”), as well as better visibility of the relevant structures. To make use of this fact, the same net was used to make two separate prediction. Between the predictions the

partial image was recentered, with the previously predicted Sella point as the new centre. Only the second prediction then is the final prediction of the network. The network used is the same as in 3.3.2.

As a second measure the coordinate was split into two separate scalars for the x and the y coordinate that are being predicted by separate networks.

4. Results

4.1 Image segmentation: Heatmap approach

Learning is not possible for this kind of network in this application example. The U-Net learns to output a completely non-activated output, by adapting to the desired output. Similar results can be seen when using different heatmap generation methods (other than U-Nets).

The heatmap approach is a quantitative key-point detection. This means that the focus is on detecting multiple structures (e.g. faces) simultaneously, during which the accuracy of an individual prediction is less relevant. During X-ray image key-point detection, each point only needs to be detected and located once, but with very high accuracy, which is not possible using this approach.

Therefore, qualitative key-point detection methods must be developed and tested.

4.2 Key-point detection network using entire lateral cephalogram

4.2.1 Pre-trained key-point detection network

The quality of this approach can be seen in Fig. 10, where the deviation of the validation data set from that of the training data set is shown. The diagram shows clear overfitting. The deviation of the training data set converges to zero,

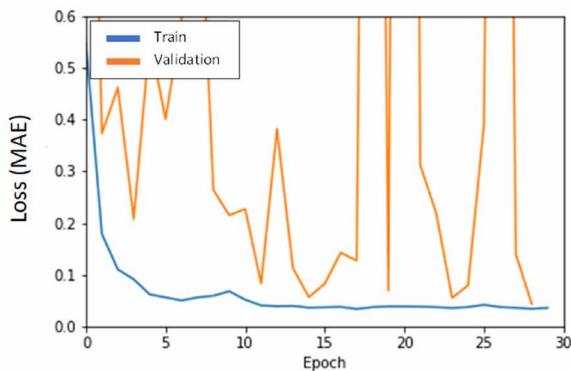


Fig. 10: Training and validation loss over epochs (pre-trained key-point detection network; no dropout)

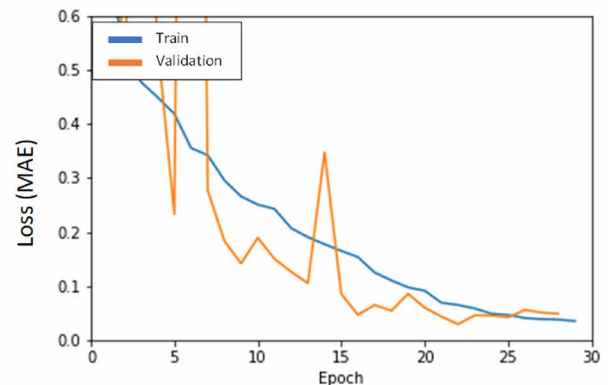


Fig. 11: Training and validation loss over epochs (pre-trained key-point detection network; dropout of 0.2)

although the deviation of the validation data set fluctuates significantly and does not converge. This means that the network can set the learned points in the training data set very precisely but fails at new, unseen tasks. An average deviation of 23 % of the image size (34.5 mm) in the test data shows similar discrepancies as the validation data set.

Using a dropout of 0.2 (deactivation of every fifth neuron), it was possible to make noticeable improvements to the overfitting problem (Fig. 11). The predictions of the test data set were made with an average deviation of 11% of the image size (16.5 mm) using this new model.

Despite the dropout, we also have a relatively high fluctuation of validation loss in this adjusted model. If the dropout is increased even further, the deviation can also be further reduced with a very high dropout up to a certain point.

At a dropout of 0.5 (50 % of all neurons are deactivated in every learning step), the deviation was able to be reduced to 7 % of the image size (10.5 mm), and to 6 % of the image size (9 mm) with a dropout of 0.7.

Such a positive reaction to high dropout speaks to the fact that a simpler network would be able to achieve better results

for the task at hand than the complex one used previously.

4.2.2 Custom key-point detection network

The learning progress of the custom net can be seen in (Fig. 12).

This simplified and more specialized model learns more quickly because fewer parameters have to be optimized (30 seconds for 30 epochs instead of 2.30 minutes for 30 epochs with the pre-trained model). Here, an average deviation of just 3.2 % of the image size (4.8 mm) can be identified with the test data set. Using additional dropout, overfitting can also be minimized here.

A dropout of 0.2 prevents overfitting here as well, and allows the model to achieve an even lower average test deviation of 2.8 % of the image size (4.2 mm).

The average deviation of predictions of the test data set (3 % of image size, 4.5 mm) differs from the average deviation of predictions of the validation data set (2 % of image size, 3 mm).

The specific Data-Augmentation used on the custom network did not improve the results.

By randomly shifting the points onto an area that is 3/4 of the image area, only a test deviation of 8 % of the image size

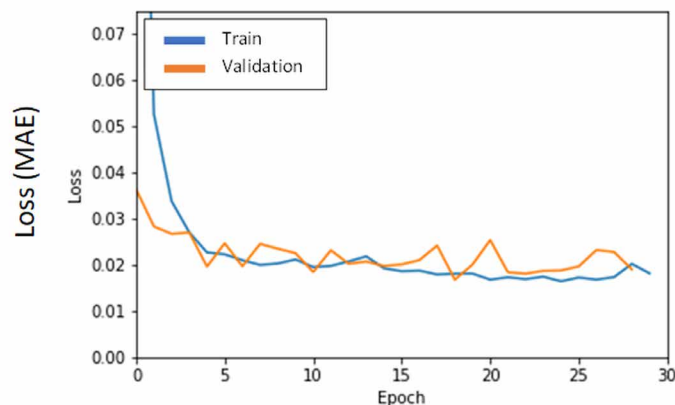


Fig. 12: Training and validation loss over epochs (custom key-point detection network; no dropout)

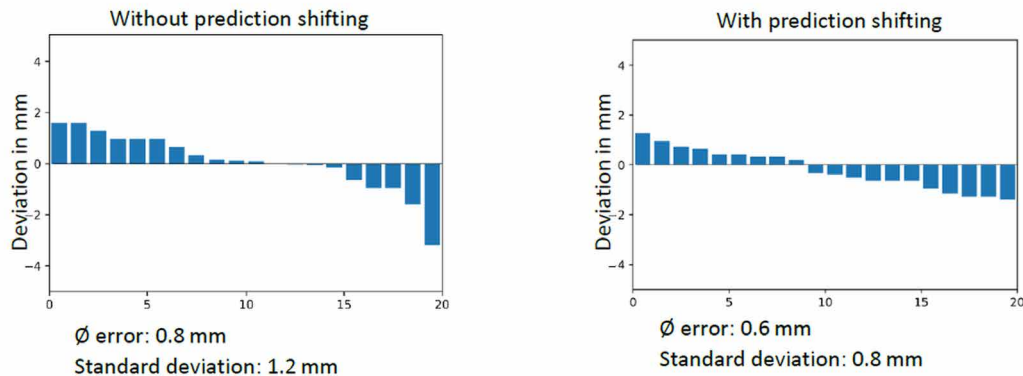


Fig. 13: Waterfall plot of prediction errors for 20 Sella partial images (test data set)

(12 mm) could be achieved, and 7 % (10.5 mm) at 1/2 the image area.

4.3 Clipping window approach

4.3.1 Classification network

After adjusting the network structure and the dropout, the network achieves an accuracy of 98.8 % correct classification for lateral cephalogram partial images after 150 epochs.

Further improvement to 99.5 % correct classification for lateral cephalogram partial images was achieved using histogram equalization.

Using the clipping window search algorithm the Sella structure was identified every time in each of the 20 test images and an image of a complete Sella was outputted each time.

4.3.2 Key-point detection network using Sella partial image

After a total of 60 training epochs, an average deviation of the test data set of 2.5 % of the partial image size (0.8 mm) was achieved. While this approach exhibits a low average deviation, the scattering of the losses is high (at maximum 10 %, 3.2 mm) with a standard deviation of 1.2 mm.

4.3.3 Key-point detection network using Sella partial image with prediction shifting

Fig. 13 compares the approach of 4.2.3 (left) and of 4.3.3 (right).

Both networks described 4.3.2 and 4.3.3 are applied to the Sella partial images created using the classification network described in 4.3.1 from the whole X-ray images.

The classification process (4.3.1) consequently extracts Sella partial images with a size of 150×150 pixels from the entire input lateral cephalogram which itself has a

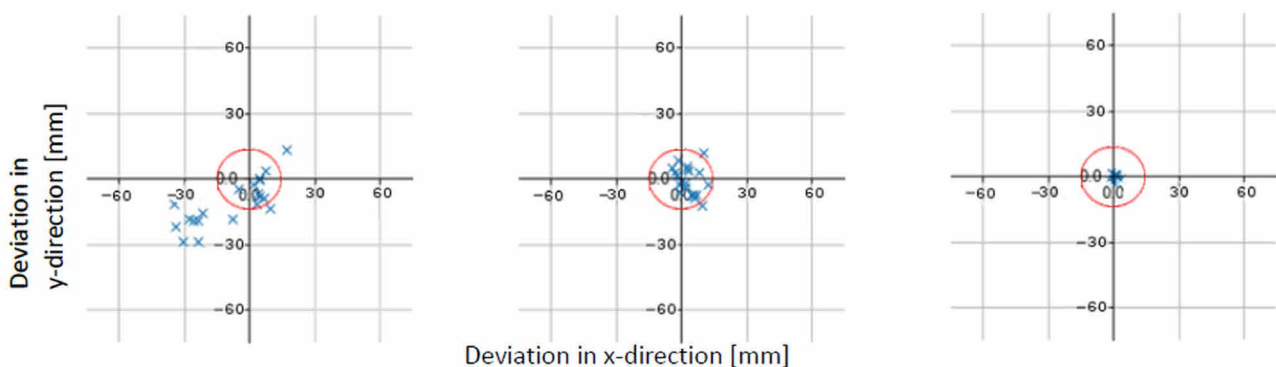


Fig. 14: Deviation of predictions with respect to the points set by specialists:
 Left: Pre-trained key-point detection network (3.2.1)
 Middle: Custom key-point detection network (3.2.2)
 Right: Clipping window approach with prediction shifting (3.3)
 Red circle: Average Sella size

standardized size of 750×750 pixels. The exact middle point of the Sella now can be determined on these partial images by the key-point detection net (4.3.3) with an average deviation of 1.8 % and maximum 5 % of the partial image size. In order to make this approach comparable to the prior ones the error should not be correlated to the Sella partial image but to the entire lateral cephalogram (as in 4.1 and 4.2), leading to an average deviation of 0.36 % of the image size and a maximum deviation of 1 % of the image size. Based on an original lateral cephalogram as it is presented to the doctor, this corresponds to an average deviation of 0.6 mm and a maximum deviation of 1.5 mm.

4.4 Comparison of approaches

Finally, if the three processes examined in this paper are compared based on the absolute deviations for the respective coordinates of the Sella point (Fig. 14), the increase in accuracy can be seen clearly, as has been shown in the course of this project. The average deviation of 6 % (9 mm) in the optimized first approach was reduced to 0.36 % of the image size (0,6 mm) in the third process.

The following three diagrams (Fig. 14) shows the deviation of predictions of the respective networks in the x- and y-direction. Here, the custom, combined network from 3.3 with

prediction shifting, as tested on the test data set, is clearly the most precise.

4.5 Application on other points

To test the transferability of the developed method, it was tested on the incisal point. This is another important point to be placed on the foremost tip of the first incisor (Fig. 15, 16).

The mean deviation of 0,65 mm and maximum deviation of 1,4 mm is very comparable to the results seen on the original application on the Sella point.

5. Discussion

In this project, different methods were developed and compared for automatically carrying out key-point detection on X-ray images using neural networks. The research used the example of the Sella point on lateral cephalograms of the side of the cranium. Three different approaches were compared.

A heatmap approach, as is used in facial recognition, was introduced in the first step. This network structure is particularly well-suited for detecting multiple points at the same time, but less so for precisely setting individual key points.

During X-ray image key-point detection, the stimulus density (i.e. the

number of activation cumulations on the heatmap) is very low (singular cumulation). Therefore, learning is not possible for this kind of network in this application example. The U-Net learns to output a completely non-activated output, by adapting to the desired output. Due to the low stimulus density, the parts of the desired heatmap that are non-activated are too large. These problems could be addressed by enlarging the distribution radius of activations, although this inevitably would make the output less specific. The same problem also applies to possible heatmap approaches without using a U-Net structure. However, this kind of method could help in other radiological tasks, such as detecting carious areas, since this involves quantitative key-point detection.

A key-point detection network was used in the next step. The reduction of complexity of the output allowed predictions that achieved an average deviation of just 4.2 mm with the redesigned and custom network. A key-point detection network is useful for finding individual, specific points. Due to the limitations of the available hardware, the size of the lateral cephalogram images needed to be reduced from the original 993×1123 to 150×150 pixels.

Further important localization information is lost due to the above



Fig. 15: Incisor in lateral cephalogram

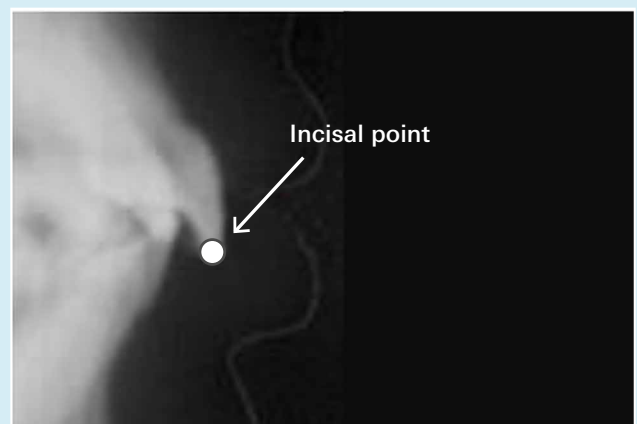


Fig. 16: Incisor with Incisal point drawn in

mentioned max pooling and valid padding. However, these methods were necessary in this application in order for the CNN error to converge while keeping the number of optimized parameters within the bounds set by the hardware.

This sets a upper limit on the possible accuracy of key-point localization. This type of network has shown itself to be suitable for finding points in this research, but it exhibits significant precision problems in prediction.

Two neural networks that build on each other were used as a third approach. First, a classification CNN extracts the partial image from the entire image containing the Sella. Only once this partial image is extracted does a key-point detection network set the precise S-point on it. This limits the influence of the loss of localization information of the key-point detection network on final prediction.

This approach is also based on the way a doctor works, first by searching for the most likely search area for the Sella structure, and then (after detecting the Sella) approximating the S-point.

Lastly the method of prediction shifting was used based on the fact that a more central position of the relevant structure leads to a more accurate prediction.

It could be speculated that this is correlated with the higher information density of the central areas (see “valid padding”) as well as better visibility of the relevant structures in the centre.

Using the clipping window and prediction shifting, the divide and conquer paradigm was successfully applied.

The goal of developing a workable procedure for predictive analytic key-point detection was achieved using the clipping window approach and the prediction shifting developed here. The

Sella point on which these methods were tested was able to be located with an average deviation of 0.6 mm.

After a reliable way to automatically detect the S-point has been developed, this method can be applied to other relevant points of cephalometric X-ray analysis.

Unlike some algorithms or programs which are not capable of learning, a neural network structure is not tied to the original application area. Instead, it can be transferred to other problems using the same solution approach with little to no change in accuracy.

This transferability of this approach was proven by testing it on the incisal point. The deviation only differs by less than 0,01 mm from the results of the Sella point which shows the applicability of the method to different points.

In the context of a manual evaluation by a specialist, Segner and Hasund write the following about the reproducibility of manual point localizations: “It generally will not be possible to reproduce the measured values of the first evaluation with an accuracy of 0.1 or 0.5 mm, no matter whether the evaluation was done (by a doctor) digitally or by hand” [16].

The deviation of the neural network from the points set by professionals that was determined in this paper is thus comparable to the deviation that occurs between two medical evaluations. The developed method thus enables practical results for radiologic evaluation to be achieved.

However, the existing results thus far only pertain to the example application of the methods on two points.

In localization of other important points on a lateral cephalogram as well, certain features have to be detected and correlated in order to approximate the respective position.

This ability is especially important in localizing the Sella point, since this is an imaginary point. This means that the point is not set on a real structure, but rather has to be constructed based on the surrounding structure of the Sella turcica. This makes localization difficult and demonstrates the ability of the model to not only detect features, but also to correlate different detected features to each other.

The application to the incisal point has furthermore shown a certain degree of transferability, in particular to a non-imaginary point which is to be placed on an actual depicted structure.

The actual transferring of this method to other points can be executed in further research.

Outside of lateral cephalogram analysis, there are a few other X-ray image evaluations that have so far applied manual key-point detection. This includes all orthopaedic, bone-related malpositions, i.e. malpositions of the hip, knee or spine. The transferability of the process developed here can also be tested in this field.

6. Outlook

Based on these results, the next goal in this process is the automated evaluation of other points based on the described method of the clipping window with prediction shifting. To do so, new data sets are required that must be created by hand.

The long-term goal here is being able to perform completely autonomous X-ray image evaluation (possibly with diagnosis) in cephalometry based on key-point detections by applying this method.

Here, only the angles and distances between the individual points need to be calculated in the last step. The appropriate software already exists for this purely algebraic task. However,

at the moment, it only produces calculations based on points set by hand.

An opportunity for the direct implementation of the system in an application environment is thereby given.

In the next step, transferring the methods to similar X-ray image evaluations (hip, knee, or spine X-ray images) is to be tested.

Any real-world implementation of this kind of automation will certainly still (currently) need the supervision and review of a specialist, but the evaluation can be accelerated and standardized. That is an advantage for both the doctor and for the patient. This application area is an example of a beneficial human-machine collaboration.

In following projects, image preprocessing could be examined in greater detail. In this paper, histogram equalization was applied to compensate for contrast differences. Other methods include edge detection or thresholding, for example. In both methods, irrelevant data can be discarded before beginning training, which enables faster and more reliable detection by the network. This can reduce the necessary network depths and, as a result, the calculation effort as well. On the other hand, preprocessing images too heavily can destroy important structures on some images. Individualized preprocessing tailored to each image increases the effort needed for learning — a compromise needs to be found.

7. Summary

The goal of this research was to develop methods for automated key-point detection on X-ray images, using the example of Sella points within lateral cephalogram of the side of the cranium with the aid of artificial intelligence, specifically the use of convolutional neural networks.

This project succeeded in identifying the Sella point with an average deviation to the manually set points of less than 1 % of the image size and an absolute deviation of 0.6 mm. This is within the spread range for specialist evaluation.

Using the method developed by combining a clipping window process, a classification network, two key-point detection networks and prediction shifting, the method has already been successfully tested on an additional point of cephalometric analysis. The transferability has been proven.

Acknowledgements

I would like to express my thanks for the support from Dr. Carsten Winkler (mathematics and physics teacher, PGS Dassel) for the critical examination of my paper as well as Drs. Angrit and Thomas Schott — my parents — for providing the anonymized lateral cephalograms and the manual, professional setting of the Sella points.

Sources and bibliography

- [1] Ciresan, D.C., Gambardella, L.M., Giusti, A., Schmidhuber, J.: Deep neural net-works segment neuronal membranes in electron microscopy images. In: NIPS. pp. 2852–2860 (2012)
- [2] Ciresan, D., Meier, U., Schmidhuber, J.: Multi-column deep neural networks for image classification. In: Institute of Electrical and Electronics Engineers (IEEE), p. 3642–3649. (2012)
- [3] Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE, Conference on Computer Vision and Pattern Recognition (CVPR) (2014)
- [4] Hahnloser, R., Seung, H.S.: Permitted and Forbidden Sets in Symmetric Threshold-Linear Networks In: Neural Computation archive, Volume 15, Issue 3, pp. 621–638 (2003)
- [5] Abdullah-Al-Wadud M., Hasanul Kabir Md., Ali Akber Dewan M., Chae O.: A Dynamic Histogram Equalization for Image Contrast Enhancement, In: IEEE Transactions on Consumer Electronics (Volume: 53 , Issue: 2 , May 2007)
- [6] <http://yann.lecun.com/exdb/mnist/> (retrieved on 2018-11-12, 17:14)
- [7] <https://jupyter.org/> (retrieved on 2021-07-1, 15:00)
- [8] Kingma, D., Ba, J.: Adam: A Method for Stochastic Optimization In: ICLR (2015)
- [9] Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks, 2012 In: NIPS'12 Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1, p. 1097–1105 (2012)
- [10] Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation In: Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, LNCS, Vol.9351: p. 234–241 (2015)
- [11] Simonyan, K., Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition In: ICLR 2015 (2015)
- [12] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout. A Simple Way to Prevent Neural Networks from Overfitting. In: Journal of Machine Learning Research. Volume 15, No. 2. Department of Computer Science University of Toronto, p. 1929–1958 (2014)
- [13] Platania, R., Shams, S., Yang, S., Zhang, J., Lee, K., Park, S: Automated Breast Cancer Diagnosis Using Deep Learning and Region of Interest Detection (BC-DROID) In: Proceedings of ACM-BCB'17, August 20-23, 2017, Boston
- [14] Do Nhu, T., Kim, S.H. , Yang, H.J., Lee, G.-S.: Face Tracking with Convolutional Neural Network Heat-Map, 2018, Conference: The 2nd International Conference on Machine Learning and Soft Computing (ICMLSC 2018), At Phu Quoc, Vietnam
- [15] Segner, D., Hasund A.: Individualisierte Kephalometrie, Dietmar Segner Verlag, Hamburg, 1998, 3rd Edition, p. 14
- [16] Segner, D., Hasund A.: Individualisierte Kephalometrie, Dietmar Segner Verlag, Hamburg, 1998, 3rd Edition, p. 42
- [17] Leshno M., Lin V. Ya., Pinkus A., Schocken Sh.: Multilayer feedforward networks with a nonpolynomial activation function can approximate any function, Neural Networks. 6 (6), pp. 861–867, Jan 1993
- [18] Hunter, J.D.: Matplotlib: A 2D graphics environment, Computing in Science & Engineering, Volume 9, p. 90–95, 2007
- [19] Bergstra J., Bengio Y.: Random Search for Hyper-Parameter Optimization In: Journal of Machine Learning Research. 13, pp. 281–305, 2012

Publiziere auch Du hier!

Forschungsarbeiten von
Schüler/Inne/n und Student/Inn/en

In der Jungen Wissenschaft werden Forschungsarbeiten von SchülerInnen, die selbstständig, z.B. in einer Schule oder einem Schülerforschungszentrum, durchgeführt wurden, veröffentlicht. Die Arbeiten können auf Deutsch oder Englisch geschrieben sein.

Wer kann einreichen?

SchülerInnen, AbiturientInnen und Studierende ohne Abschluss, die nicht älter als 23 Jahre sind.

Was musst Du beim Einreichen beachten?

Lies die [Richtlinien für Beiträge](#). Sie enthalten Hinweise, wie Deine Arbeit aufgebaut sein soll, wie lang sie sein darf, wie die Bilder einzureichen sind und welche weiteren Informationen wir benötigen. Solltest Du Fragen haben, dann wende Dich gern schon vor dem Einreichen an die Chefredakteurin Sabine Walter.

Lade die [Erstveröffentlichungserklärung](#) herunter, drucke und fülle sie aus und unterschreibe sie.

Dann sende Deine Arbeit und die Erstveröffentlichungserklärung per Post an:

Chefredaktion Junge Wissenschaft

Dr.-Ing. Sabine Walter
Paul-Ducros-Straße 7
30952 Ronnenberg
Tel: 05109 / 561508
Mail: sabine.walter@verlag-jungewissenschaft.de

Wie geht es nach dem Einreichen weiter?

Die Chefredakteurin sucht einen geeigneten Fachgutachter, der die inhaltliche Richtigkeit der eingereichten Arbeit überprüft und eine Empfehlung ausspricht, ob sie veröffentlicht werden kann (Peer-Review-Verfahren). Das Gutachten wird den Euch, den AutorInnen zugeschickt und Du erhältst gegebenenfalls die Möglichkeit, Hinweise des Fachgutachters einzuarbeiten.

Die Erfahrung zeigt, dass Arbeiten, die z.B. im Rahmen eines Wettbewerbs wie **Jugend forscht** die Endrunde erreicht haben, die besten Chancen haben, dieses Peer-Review-Verfahren zu bestehen.

Schließlich kommt die Arbeit in die Redaktion, wird für das Layout vorbereitet und als Open-Access-Beitrag veröffentlicht.

Was ist Dein Benefit?

Deine Forschungsarbeit ist nun in einer Gutachterzeitschrift (Peer-Review-Journal) veröffentlicht worden, d.h. Du kannst die Veröffentlichung in Deine wissenschaftliche Literaturliste aufnehmen. Deine Arbeit erhält als Open-Access-Veröffentlichung einen DOI (Data Object Identifier) und kann von entsprechenden Suchmaschinen (z.B. BASE) gefunden werden.

Die Junge Wissenschaft wird zusätzlich in wissenschaftlichen Datenbanken gelistet, d.h. Deine Arbeit kann von Experten gefunden und sogar zitiert werden. Die Junge Wissenschaft wird Dich durch den Gesamtprozess des Erstellens einer wissenschaftlichen Arbeit begleiten – als gute Vorbereitung auf das, was Du im Studium benötigst.

Richtlinien für Beiträge

Für die meisten Autor/Inn/en ist dies die erste wissenschaftliche Veröffentlichung. Die Einhaltung der folgenden Richtlinien hilft allen – den Autor/innen/en und dem Redaktionsteam

Die Junge Wissenschaft veröffentlicht Originalbeiträge junger AutorInnen bis zum Alter von 23 Jahren.

- Die Beiträge können auf Deutsch oder Englisch verfasst sein und sollten nicht länger als 15 Seiten mit je 35 Zeilen sein. Hierbei sind Bilder, Grafiken und Tabellen mitgezählt. Anhänge werden nicht veröffentlicht. Deckblatt und Inhaltsverzeichnis zählen nicht mit.
- Formulieren Sie eine eingängige Überschrift, um bei der Leserschaft Interesse für Ihre Arbeit zu wecken, sowie eine wissenschaftliche Überschrift.
- Formulieren Sie eine kurze, leicht verständliche Zusammenfassung (maximal 400 Zeichen).
- Die Beiträge sollen in der üblichen Form gegliedert sein, d. h. Einleitung, Erläuterungen zur Durchführung der Arbeit sowie evtl. Überwindung von Schwierigkeiten, Ergebnisse, Schlussfolgerungen, Diskussion, Liste der zitierten Literatur. In der Einleitung sollte die Idee zu der Arbeit beschrieben und die Aufgabenstellung definiert werden. Außerdem sollte sie eine kurze Darstellung schon bekannter, ähnlicher Lösungsversuche enthalten (Stand der Literatur). Am Schluss des Beitrages kann ein Dank an Förderer der Arbeit, z. B. Lehrer und Sponsoren, mit vollständigem Namen angefügt werden. Für die Leser kann ein Glossar mit den wichtigsten Fachausdrücken hilfreich sein.
- Bitte reichen Sie alle Bilder, Grafiken und Tabellen nummeriert und zusätzlich als eigene Dateien ein. Bitte geben Sie bei nicht selbst erstellten Bildern, Tabellen, Zeichnungen, Grafiken etc. die genauen und korrekten Quellenangaben an (siehe auch [Erstveröffentlichungserklärung](#)). Senden Sie Ihre Bilder als Originaldateien oder mit einer Auflösung von mindestens 300 dpi bei einer Größe von 10 · 15 cm! Bei Grafiken, die mit Excel erstellt wurden, reichen Sie bitte ebenfalls die Originaldatei mit ein.
- Vermeiden Sie aufwendige und lange Zahlentabellen.
- Formelzeichen nach DIN, ggf. IUPAC oder IUPAP verwenden. Gleichungen sind stets als Größengleichungen zu schreiben.
- Die Literaturliste steht am Ende der Arbeit. Alle Stellen erhalten eine Nummer und werden in eckigen Klammern zitiert (Beispiel: Wie in [12] dargestellt ...). Fußnoten sieht das Layout nicht vor.
- Reichen Sie Ihren Beitrag sowohl in ausgedruckter Form als auch als PDF

ein. Für die weitere Bearbeitung und die Umsetzung in das Layout der Jungen Wissenschaft ist ein Word-Dokument mit möglichst wenig Formatierung erforderlich. (Sollte dies Schwierigkeiten bereiten, setzen Sie sich bitte mit uns in Verbindung, damit wir gemeinsam eine Lösung finden können.)

- Senden Sie mit dem Beitrag die [Erstveröffentlichungserklärung](#) ein. Diese beinhaltet im Wesentlichen, dass der Beitrag von dem/der angegebenen AutorIn stammt, keine Rechte Dritter verletzt werden und noch nicht an anderer Stelle veröffentlicht wurde (außer im Zusammenhang mit **Jugend forscht** oder einem vergleichbaren Wettbewerb). Ebenfalls ist zu versichern, dass alle von Ihnen verwendeten Bilder, Tabellen, Zeichnungen, Grafiken etc. von Ihnen veröffentlicht werden dürfen, also keine Rechte Dritter durch die Verwendung und Veröffentlichung verletzt werden. Entsprechendes [Formular](#) ist von der Homepage www.junge-wissenschaft.ptb.de herunterzuladen, auszudrucken, auszufüllen und dem gedruckten Beitrag unterschrieben beizulegen.
- Schließlich sind die genauen Anschriften der AutorInnen mit Telefonnummer und E-Mail-Adresse sowie Geburtsdaten und Fotografien (Auflösung 300 dpi bei einer Bildgröße von mindestens 10 · 15 cm) erforderlich.
- Neulingen im Publizieren werden als Vorbilder andere Publikationen, z. B. hier in der Jungen Wissenschaft, empfohlen.

Impressum

[JUNGE]
wissenschaft



Junge Wissenschaft

c/o Physikalisch-Technische
Bundesanstalt (PTB)
www.junge-wissenschaft.ptb.de

Redaktion

Dr. Sabine Walter, Chefredaktion
Junge Wissenschaft
Paul-Ducros-Str. 7
30952 Ronnenberg
E-Mail: sabine.walter@verlag-jungewissenschaft.de
Tel.: 05109 / 561 508

Verlag

Dr. Dr. Jens Simon,
Pressesprecher der PTB
Bundesallee 100
38116 Braunschweig
E-Mail: jens.simon@ptb.de
Tel.: 0531 / 592 3006
(Sekretariat der PTB-Pressestelle)

Design & Satz

Sabine Siems
Agentur „proviele werbung“
E-Mail: info@proviele-werbung.de
Tel.: 05307 / 939 3350

