이학석사 학위논문

Cephalometric Landmarks Detection using Fully Convolutional Networks

(완전 컨볼루션 네트워크를 이용한 두부 계측 지표 탐색)

2017년 8월

서울대학교 대학원

계산과학 협동과정

박 수 범

Cephalometric Landmarks Detection using Fully Convolutional Networks

(완전 컨볼루션 네트워크를 이용한 두부 계측 지표 탐색)

지도교수 강 명 주

이 논문을 이학석사 학위논문으로 제출함

2017년 6월

서울대학교 대학원

계산과학 협동과정

박 수 범

박 수 범의 이학석사 학위논문을 인준함

2017년 6월

위 원 장 ____________________ (인)

부 위 원 장 ____________________ (인)

위 원 ____________________ (인)

Cephalometric Landmarks Detection using Fully Convolutional Networks

A dissertation
submitted in partial fulfillment
of the requirements for the degree of
Master of Science
to the faculty of the Graduate School of
Seoul National University

by

Subeom Park

Dissertation Director : Professor Myungjoo Kang

Interdisciplinary Program in
Computational Science and Technology
Seoul National University

August 2017

Abstract

In dentistry, quantitative cephalometry plays an essential role in the practice of medical care for patients. In this thesis, an automated landmark detection model is proposed using FCN (fully convolutional networks) with internally residual connections. The FCN model was trained to output an archery target shape heatmap when an image patch near the landmark was input. The image patches used for training were positioned and sized based on training data, and augmentation was performed. The cephalogram were used for training and testing used a publicly available datasets. SDR(Success detection rate) was used to evaluate the results. The trained models were evaluated using a test set and compared with previous studies. As a result, landmarks were detected with better accuracy than previous studies. The FCN model showed the potential for accurate landmark detection.

Key words: supervised learning, fully convolutional networks, residual networks, cephalometric landmarks, dental X-ray images
Student Number: 2015-22686

Contents

List of Figures

Chapter 1

Introduction

One of the major goals of dentistry is to resolve craniofacial discrepancies and functional and aesthetic complete tooth alignment. To achieve this, a detailed analysis should be done using a high resolution two-dimensional x-ray image of the head bone taken from the side. This analysis is called a lateral cephalometric analysis [1, 2, 3, 4, 5]. The information to be obtained from the analysis is anatomically defined landmarks and the angles and distances between them. However, manual analysis of the dentist requires a lot of time and can cause intra-observer variability [2]. Therefore, a stable and consistently automated end-to-end analysis model is needed.

There have been various studies for lateral cephalometric analysis. In particular, the International Symposium on Biomedical Imaging (ISBI) held in 2014 and 2015 challenged this problem [2, 3] and several approaches were published. Ibragimov et al. used Haar-like features to express the intensity of the landmark and to link it to a point detector using a random forest, the landmark is found by using random forest regression [6]. In the second stage, the landmark is modified by sparse shape composition model. The model of Chen et al. learned the visual characteristics of the image patches and the distance from the landmarks, and made a prediction model by voting the landmarks obtained from each patch [7]. Vandaele et al. solved this problem with each of the 19 landmark binary classification problems [8]. They used extremely randomized trees as pixel classifiers. Despite the wide variety of studies, no accurate model has yet been developed for use in clinics with an

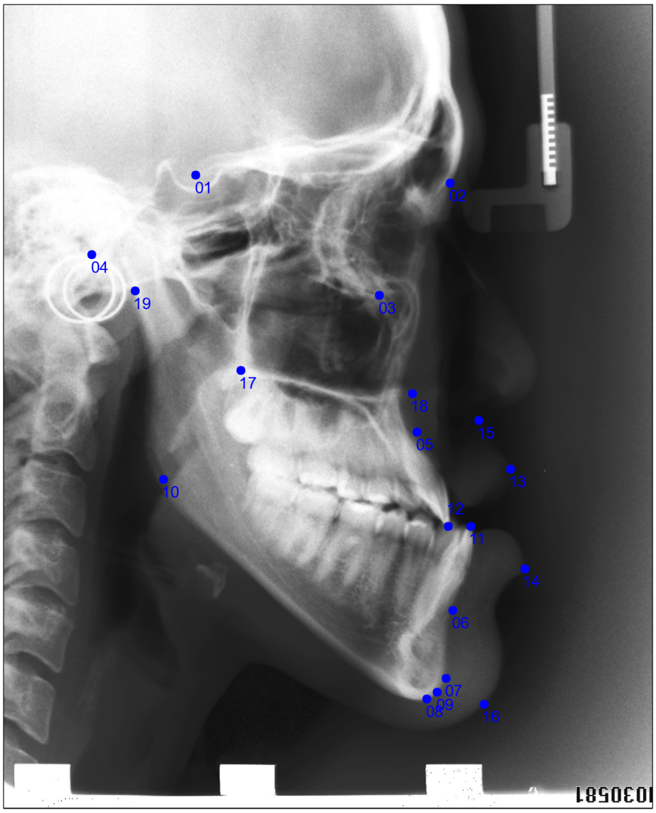

Figure 1.1: Example of cephalogram with 19 cephalometric landmarks positions

error of less than 2 mm [3].

The objective of this study is automatically detecting the cephalometric landmarks from lateral cephalogram using Convolutional Neural Networks [9]. In recent years, deep learning has outperformed existing algorithms in various areas. Especially since the Alexnet in ILSVRC in 2012 [10], CNN has been developed rapidly in image processing. CNN is a multi-layered perceptron model inspired by animal visual systems [9]. The color images given as input to the image processing problem are represented in a three-dimensional array inside the computer. For high-resolution images, one image is represented by many numbers. CNN has the characteristics of local connections and shared variables. This property allows spatial properties from images with few parameters. Therefore, CNN enables us to get specific information efficiently from images.

CNN has been widely applied to medical imaging [11], segmentation [12, 13], object/lesion detection [14, 15], image/exam classification [16], registration [17]. There are some papers that find landmarks in medical images. Payer et al. used CNN to find multiple landmark points. They first defined the location of the landmark as a heatmap using Gaussian [18]. Then, in the learning process, the landmark was estimated by learning the heatmap from the input image. Arık et al. solved the problem of cephalometric landmarks detection using CNN. Their idea is to find intensity appearance patterns for each landmark [19]. This method showed that the CNN-based method is better than random forest based methods.

In a similar vein, this paper trains the heatmap around the landmark using fully convolutional neural networks. The shape of the heatmap consisted of archery target, used to refer to a particular landmark while simultaneously learning the surrounding area. To objectively evaluate the performance of the proposed model, results were obtained using 400 publicly available images and well-defined landmarks by the dentist and compared with other state-of-the-art approaches.

Chapter 2

Methodology

2.1 Convolutional Neural Networks

2.1.1 Basic Model

Convolutional Neural Network (CNN) is a deep learning model inspired by the animal visual system [20]. CNN has been widely used in image analysis and various models have been developed [21]. LeNet-5 is the well-known model that solves real-world problems using CNN for the first time [22]. This model consists of multi-layer artificial neural networks and consists of three types of layers. The three are the convolution layer, the pooling layer, and the fully connected layer. When an image comes in from the input layer, it passes through the convolution layer and the pooling layer repeatedly, and finally, through the fully connection layer, the result comes out. CNN is trained by the backpropagation algorithm. This algorithm updates parameters using chain rules and gradient descents [23].

Passing through the convolution layer in the equation:

$$z_{i,j,k}^{l} = {w_k^l}^T x_{i,j}^{l} + b_k^l$$

Where w_k^l and b_k^l are the weight vector and bias term of the k-th filter of the l-th layer respectively, and $x_{i,j}^l$ is the input patch centered at location (i, j) of the l-th layer. The weight filter is shared while sliding the image. This has the advantage of using a smaller weight when compared to a multi-layer perceptron, and is efficient in locating a particular feature across an image.

The pooling layer reduces the number of parameters by downsampling. It also has the invariance to local translation property. This means that even if the input changes slightly, the pooled result will not change. The methods of pooling include average pooling and max pooling.

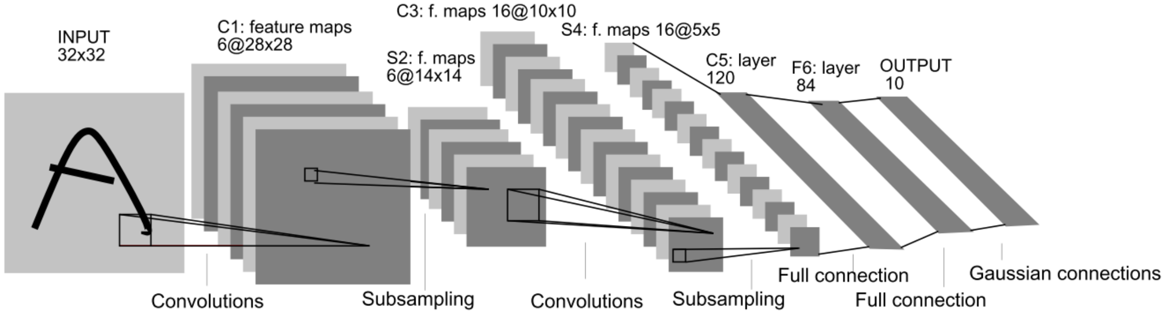

Figure 2.1: The architecture of the LeNet-5 network for digits recognition.

The LeNet-5 model does not have an activation function. However, in another paper, Lecun et al. wrote that the activation function was introduced to add nonlinearity to CNN [24]. Adding the nonlinear activation function $a(\cdot)$ to the output Z of the convolution layer is expressed as:

$$a^l_{i,j,k} = a(z^l_{i,j,k})$$

Commonly used nonlinear activations are sigmoid($f(x) = 1/(1 + e^{-x})$) [24], and hyperbolic tangent($f(x) = ((e^{2x} - 1)/(e^{2x} + 1))$ [24], ReLU($f(x) = max(x, 0)$) [25], and so on.

2.1.2 Fully Convolutional Networks

Fully Convolutional Networks(FCN) was developed to solve segmentation problems [26]. The output of the FCN is densely label maps. In the conventional CNN model, the back side is a fully connected layer. When connected to a fully connected layer, information about location or space is lost. The FCN consists of only the convolution layer and the pooling layer from the beginning to the end, so that the spatial information is not lost. In a FCN, the structure corresponding to a fully connected layer can be a 1x1 convolution layer. Also, since the fully connected layer has a fixed input size, the input image must have a fixed size. However, 1x1 convolution layer operate on arbitrary sized images.

When an image passes through several stages of convolution and pooling layer, the size of the feature map is reduced. In order to do pixel-by-pixel prediction, it is necessary to raise the reduced feature map again. The authors of the FCN paper train upsampling using deconvolution and skip connection.

There is a model called U-net which is well known for the segmentation problem of medical images [12].

2.1.3 Residual Networks

Let $H(x)$ be the specific stacked layers of a deep learning model, and let x be the given input. Suppose that $H(x)$ can represent a complicated function that we are trying to obtain. If so, then $H(x) - x$ can also represent the complicated function we want to obtain. If we add the input x to $H(x)$ in the network structure, the function we need to train is the residual function $F(x) = H(x) - x$. At this time, the expression power of F and H is the same by assumption, and it is generally easier to train F. With some additions, more stable training is possible.

The residual block is expressed as follows:

$$y = \mathcal{F}(x, \{W_i\}) + x$$

Where x and y are the input and output of the layer, respectively. $\mathcal{F}(x, \{W_i\})$ means a residual mapping to train.

Author of the Resnet paper proposed using a deeper bottleneck architectures in their paper [27]. The bottleneck design consists of three convolution layers with a kernel size of 1x1, 3x3, 1x1. The 1x1 layer serves to reduce or increase the number of filters and the 3x3 layer is the part that learns for the overall goal. This network configuration allows for efficient use of identity shortcuts while reducing the number of parameters needed to connect layers.

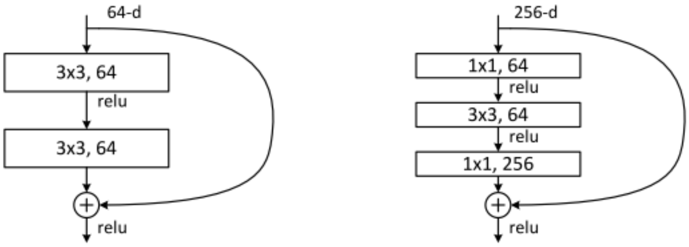

Figure 2.2: Residual building block $\mathcal{F}$. **Left**: normal structure, **Right**: bottleneck structure

2.1.4 Batch Normalization

When the problem is that the scale of the data itself is not important, we usually use it to normalize the data. Normalizing means changing the distribution of the data to mean zero and variance to one. However, even if the data is well preprocessed,in the learning process, the distribution of the data changes after the data passes through the artificial neural network. This problem is called covariate shift [28]. This loses the information that the original data has, which loses the accuracy of the model. Batch normalization(BN) is suggested as a way to solve this problem [29]. The BN process is to put the normalizing step after the layer. The statistics for normalization use average and variance of mini-batch when training, and moving average and moving variance when testing

The Batch Normalizing Transform is expressed as:

$\mu_\beta \leftarrow \frac{1}{m}\sum_{i=1}^{m} x_i$
$\sigma_\beta^2 \leftarrow \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu_\beta)^2$
$\hat{x}_i \leftarrow \frac{x_i - \mu_\beta}{\sqrt{\sigma_\beta^2 + \epsilon}}$
$y_i \leftarrow \gamma\hat{x}_i + \beta \equiv BN_{\gamma,\beta}(x_i)$

Where μ_β is the mini-batch mean and σ_β^2 is the mini-batch variance. The third is normalizing and the fourth is the scaling and shifting. γ and β are learnable parameters.

The BN has several advantages that affect the learning process. Firstly, it reduces internal covariant shift. Secondly, it enables higher learning rates because it prevents exploding or vanishing gredient phenomena. Thirdly, it does not need a dropout process because it normalizes the model. Fourthly, it prevents saturating, so it allows you to use saturating nonlinear functions.

2.2 Cephalometric Landmark Detection

The problem of finding landmarks in a cephalogram is to find 19 anatomically important points in a given profile of facial bones. To automate this with a computer program, we need to create a generalized algorithm to find landmarks. The medical image given as input consists of a large number of pixel values and the output is to get 19 coordinates. Our problem can be regarded as a function estimation problem connecting these inputs and outputs.

Formally, our goal is to find the following function:

$$g : [0, 255]^{H\times W\times 3} \Rightarrow \{(x_1, y_1), ..., (x_{19}, y_{19})\}$$

In the above equation [0,255] means the range of pixel values, H and W mean the height and width of the image respectively, and 3 means RGB of color. It is difficult to find landmarks because a given cephalogram is so large in image size. It should solve the problem by cropping the image into a small area.

Assume that the data given in the training set has sufficient general information. Since cephalograms are taken at a certain distance and at a fixed angle, it is possible to assume that each landmark is within a certain range, even when considering the diversity of human face bones. So we can use the range of each landmark in the training set to determine the crop area to find the landmark. In this way, the amount of computation for finding a particular landmark can be reduced when compared to using the entire image. The cropped area should be large enough to find landmarks, and smaller with respect to computational cost. After cropping the image, the problem that needs to be solved turns into finding a particular landmark in a small area. Even if the model is created using CNN, it is difficult to find the landmark coordinates directly from the input image. To indirectly solve this problem, the model turned into a problem of obtaining an archery target heatmap centered on a landmark.

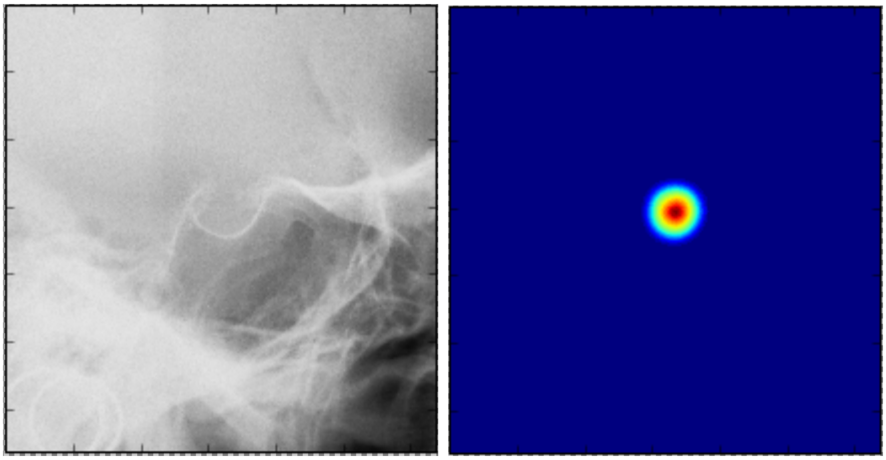

Figure 2.3: **Left**: Cropped image, **Right**: Corresponding landmark heatmap

Chapter 3

Experiment

3.1 Dataset

3.1.1 Description of Datasets

The data used in this experiment was provided in the Grand Challenges in Dental X-ray Image Analysis of IEEE International Symposium on Biomedical Imaging 2015. Cephalometric radiographs data were collected from 400 patients between the ages of 6 and 60 years. The cephalograms were acquired in TIFF format with Soredex CRANEXr Excel Ceph machine (Tuusula, Finland) and Soredex SorCom software (3.1.5, version 2.0), and the image resolution was 1935 × 2400 pixels. Two experienced dentists manually obtained cephalometric landmarks and averaged them to obtain ground truth data [3].

The distribution of the training set in Fig 3.1 shows that no landmark is shown on the top and left side of the image. And we can see that each landmark is scattered to some extent.

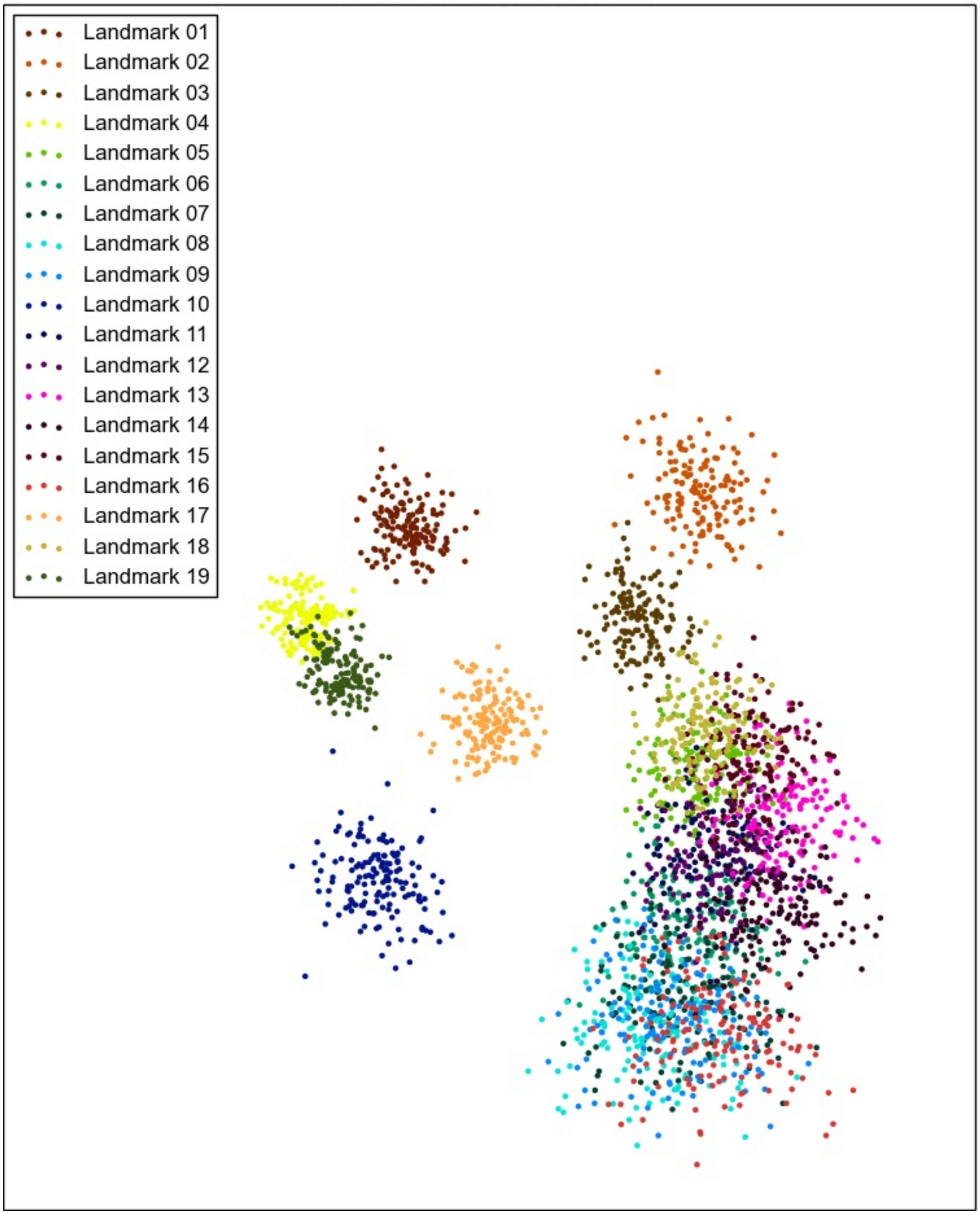

Figure 3.1: Distribution of training set(150 patients)

3.1.2 Image Cropping

To execute the learning process, the entire image should be cropped into areas around each landmark. The cropping range was obtained by finding the minimum rectangular area containing each landmark in the training set and enlarging it by 200 pixels on each side(up to the end of the whole image). The reason why the surrounding area is large is because it is determined that the information is important for finding a specific landmark. And considering that the landmark in the test set would be an exceptional distance from the existing distribution. In Fig 3.2, unlike Fig 3.1, only the area with the landmark is visualized except the top and the left side (Horizontal range=(400px,1850px),Vertical range = (600px,2400px)).

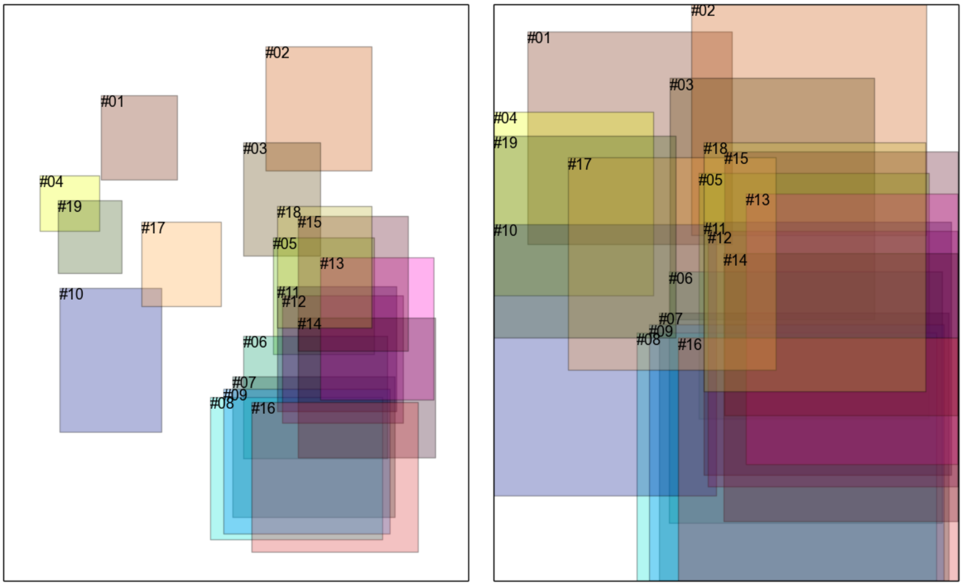

Figure 3.2: **Left**: Minimal rectangle containing landmarks in training dataset, **Right**: Expanded rectangle for cropping

3.1.3 Data Augmentation

Data augmentation is a method of artificially enlarging data by applying label-preserving transformations based on collected data [10]. When compared to other image processing problems using deep learning, the 150 training sets are too small. Few datasets can easily overfit. It is known that data augmentation can reduce overfitting [10].

Different data types (eg; text or voice data) are difficult to apply augmentation, but various augmentation methods are available for images [30]. We need to create a dataset that is different from the original, without losing the characteristics of the original dataset. In this paper, augmentation is performed by using “Crop each side”, “Add pixel value”, “Multiply pixel value” and “Rotation”. Each method was applied simultaneously in a random range. The crop range is from 0 to 10 pixels, the add range is from -10 to 10, the multiply range is from 0.9 to 1.1, and the rotation range is from $-3°$ to $3°$. It has increased the number of data by ten times.

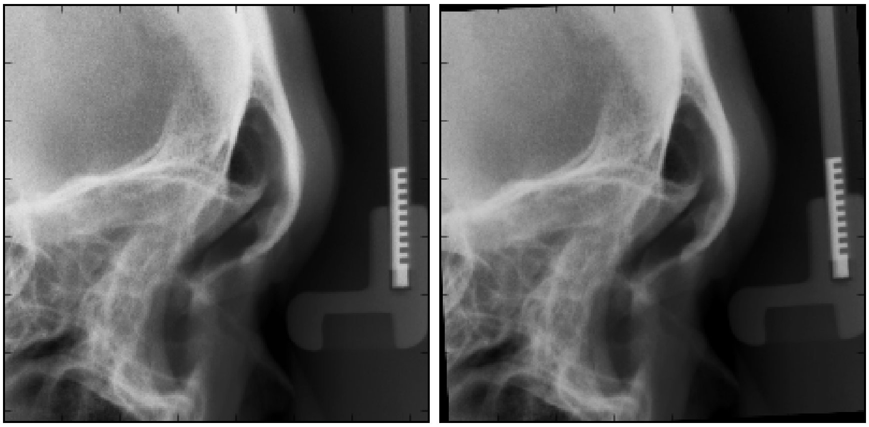

Figure 3.3: Augmentation example(Landmark 2) **Left**: Original image, **Right**: Artificial image

3.2 Model Architecture

The input of the proposed model is the cropped image defined in section 3.1.2 and the output is the archery target shape heatmap image of the same size. The overall model structure is FCN and goes through 14 convolution stages. Fig.3.4 shows the overall structure. This image is visualized for the model finding landmark 1, but all landmarks share the same structure.

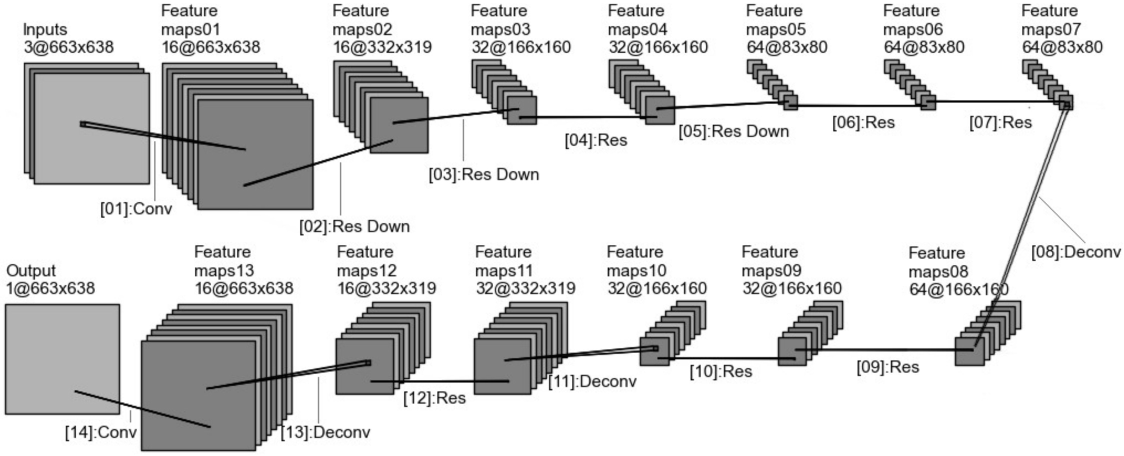
Inputs 3@663x638
Feature maps01 16@663x638
Feature maps02 16@332x319
Feature maps03 32@166x160
Feature maps04 32@166x160
Feature maps05 64@83x80
Feature maps06 64@83x80
Feature maps07 64@83x80
[01]:Conv
[02]:Res Down
[03]:Res Down
[04]:Res
[05]:Res Down
[06]:Res
[07]:Res
[08]:Deconv
Output 1@663x638
Feature maps13 16@663x638
Feature maps12 16@332x319
Feature maps11 32@332x319
Feature maps10 32@166x160
Feature maps09 32@166x160
Feature maps08 64@166x160
[09]:Res
[10]:Res
[11]:Deconv
[12]:Res
[13]:Deconv
[14]:Conv

Figure 3.4: The structure of the proposed model (for Landmark 1)

There are four types of convolution connections that construct the proposed model. The four types are “Conv”, “Res Down”, “Res”, “Deconv”. The “Conv” is the basic convolution layer. It is located at the beginning and end of the proposed model. The initial conv layer has a filter size of 25x25, which is good for experimentally large filter sizes. The last conv layer has a 1x1 filter and gives the result. The “Res Down” is the residual down block. It has four convolution layers internally. It is basically a bottleneck structure [27]. The feature map that comes into the input passes through the three conv layers, and passes a single conv layer at the same time. The three conv layers have filter sizes of 1x1, 5x5, and 1x1, respectively. The first 1x1 layer works to reduce the number of filters. The 5x5 conv layer has a stride of 2, which reduces the size of the feature map in half. The last 1x1 layer again increases the number of filters. In the conv layer only, the filter size is 1x1 and the stride is 2 so that the size of feature map is halved without greatly deforming the previous stage feature.

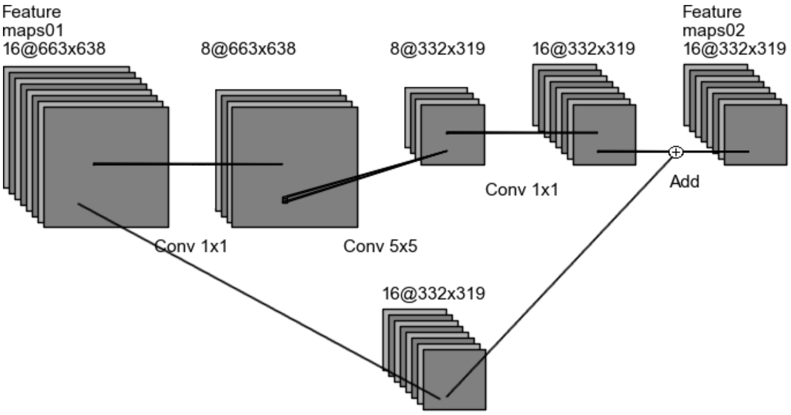

Figure 3.5: The structure of “[02]:Res Down” in figure 3.4

The “Res” is a residual block. Like “Res Down”, it has a bottleneck structure and passes through conv layers with 1x1, 5x5, and 1x1 filters. The difference with “Res Down” is that the size of the feature map does not change, so we use identity mapping for the residual structure.

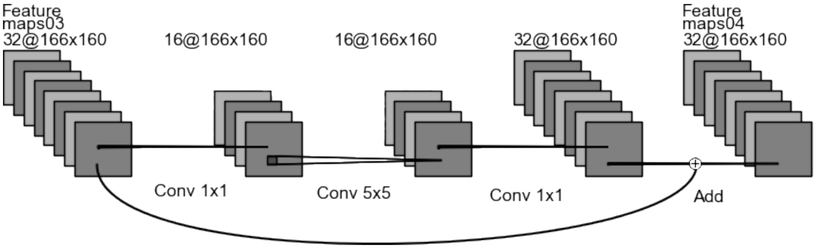

Figure 3.6: The structure of “[04]:Res” in figure 3.4

The “Deconv” is the deconvolution layer. It makes the reduced feature map size larger again. With the conv layer at the beginning of the model, the filter size was experimentally determined to be 25. All conv layers used in the model have Batch Normalization layer attached, and the ReLU function is used for activation. Also, zero padding is performed so that the conv operation does not change the feature map size.

3.3 Cost Function

The commonly used cost function is the following L1,L2 loss :

$$\mathcal{L}_{L1} = \frac{1}{m}\sum_{i=1}^{m}|y - \hat{y}|$$
$$\mathcal{L}_{L2} = \frac{1}{m}\sum_{i=1}^{m}(y - \hat{y})^2$$

In order to learn the network that outputs the heatmap, the cost function was not enough for such loss. In the heatmap image, there is a value only in the landmark and its vicinity, and the rests are all zero. Therefore, when learning by only L2 loss, all values of the output image are often estimated to be zero. In order for the model to learn the heatmap well, we need to add a term to the cost function that is only affected by the heatmap. The product of the heatmap image and the predicted image is regarded as a loss function, and the objective is achieved by maximizing the product. When the ground truth heatmap image is $\hat{H}$ and the prediction image is H, this cost is expressed as follows.

$$\mathcal{L}_{prod} = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(H_{i,j} \times \hat{H}_{i,j})$$

Since $-\mathcal{L}_{prod}$ was effective in initial training(minimization process), the first epoch only used product loss function. Then $\mathcal{L}_{L1}$ term was added to the $-\mathcal{L}_{prod}$, and a normalization term was added to prevent overfitting. The following Cost function was used.

$$\mathcal{L} = -\mathcal{L}_{prod} + \lambda_{L1}\mathcal{L}_{L1} + \lambda_{reg}||w||_2$$

In the above loss function, λ is the weight of each term. In the experiment, λ_{L1} is 2×10^{-2} and λ_{reg} is 10^{-7}.

3.4 Training

For given 150 training dataset, augmentation was performed using the method given in Section 3.1.3 to create 1500 dataset, which were used for training. In the training process, some landmarks resulted in models being over-fitting to training sets easily, and learning at some landmarks was slow. The solution of this problem is to apply the model to the testset at the end of each epoch, and then averaging the resulting heatmap to obtain the final heatmap. Averages the output heatmap from epoch 2. Since then, training accuracy were good at 87%. The reason for this is that every time the epoch ends, the model changes slightly. Therefore, it is anticipated that using the average of the heatmap will produce an ensemble effect. Training was performed up to 5 epochs, with a training accuracy of 97%. In the learning process, testset1 was used as a validation set to select the model. The optimization algorithm uses ADAM [31] and the initial learning rate is 10^{-3}.

3.5 Result

3.5.1 Evaluation Approaches

In order to measure the performance of the landmark detection model, there are measurement methods. The radial error R is the Euclidian distance, the distance between the predicted and actual coordinates. Using this, the mean and standard deviation of the error are calculated as follows.

$$\text{Mean Radial Error(MRE)} = \frac{\sum_{i=1}^{N} R_i}{N}$$

$$\text{Standard Deviation (SD)} = \sqrt{\frac{\sum_{i=1}^{N}(R_i - MRE)^2}{N-1}}$$

An important measure for this problem is the success detection rate. The ground truth is not the range of the landmark, but the landmark coordinates taken by the medical doctor. Thus, if the error between the estimated coordinate and the correct position is below a certain distance, the estimated coordinates are correct.

The success detection rate is defined as follows:

$$p_z = \frac{\#\{j : ||L_{pred}(j) - L_{gt}(j)|| < z\}}{\#\Omega} \times 100\%$$

In the above equation, L_{pred} and L_{gt} mean predicted landmark and ground truth landmark, respectively. z means the precision range for the evaluation. $j \in \Omega$, and $\#\Omega$ means the number of data [2, 3].

3.5.2 Landmarks Detection Results

After training the proposed model with the given 150 dataset, the learned model was tested with 250 data. The output from the proposed CNN model is a heatmap. However, in the given problem, what we really need to obtain is the coordinates of the landmark, not the heatmap, so we must get the landmark from the heatmap.

In this experiment, coordinates were obtained using a weighted average for coordinates greater than 90% of the highest value in the heatmap. Figure 3.7 shows the input image and the predicted heatmap image. It is observed that the heatmap predicted according to the landmark shows a different form. Landmark 2,17 was predicted to be a narrow region, while landmark 4 was predicted to be a relatively wide region. We can also see the benefits of averaging heatmaps. In the first line of the picture, the mispredicted bright blue part of the heatmap shows dimmedness in the averaged heatmap. The resulting heatmap of the fifth epoch in the picture for landmark 4 on the second line of the figure makes an incorrect guess, but the average of the heatmap makes the correct guess.

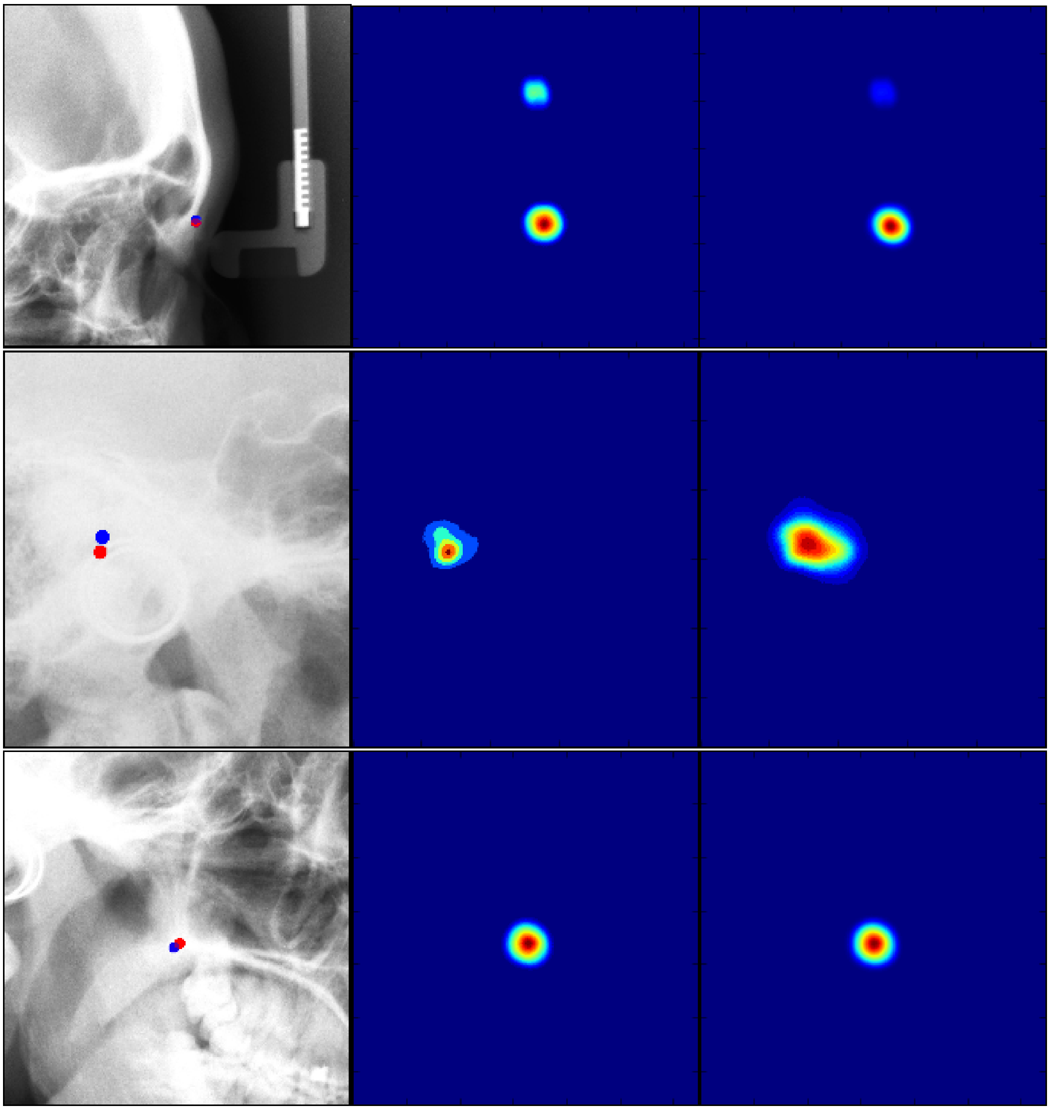

Figure 3.7: Results of landmarks detection at L2,L4,L17. **Left**: Input image, ground truth(blue), predicted(red), **Center**: Resulting heatmap at epoch 5, **Right**: Average of heatmap from 2 to 5 epoch

Table 3.1: Landmarks detection result for testset1

Landmarks	*SDR(2.0mm)*	*SDR(2.5mm)*	*SDR(3.0mm)*	*SDR(4.0mm)*	*MRE*	*SD*
1	98.00%	98.00%	98.00%	98.00%	10.81	42.64
2	72.67%	74.00%	78.00%	89.33%	18.85	22.14
3	62.00%	70.67%	82.00%	90.67%	19.84	14.40
4	52.00%	56.00%	58.00%	66.00%	34.90	36.22
5	54.67%	64.67%	75.33%	87.33%	21.88	15.54
6	66.00%	78.00%	84.00%	90.00%	21.22	29.02
7	91.33%	94.67%	96.66%	99.33%	9.24	8.10
8	89.33%	94.00%	98.00%	98.67%	10.58	9.67
9	92.67%	97.33%	98.00%	98.67%	8.49	8.82
10	31.33%	42.67%	49.33%	66.67%	36.54	34.27
11	86.67%	90.67%	95.33%	96.00%	10.89	13.87
12	94.67%	96.00%	96.66%	97.33%	10.18	49.94
13	95.33%	96.67%	98.67%	99.33%	8.61	8.21
14	96.00%	98.67%	100.00%	100.00%	7.50	5.14
15	90.67%	94.00%	96.67%	98.00%	13.09	32.57
16	60.67%	68.67%	78.00%	86.67%	20.35	15.15
17	90.00%	93.33%	96.67%	100.00%	10.79	7.60
18	68.00%	79.33%	83.33%	90.00%	21.37	43.26
19	55.33%	59.33%	67.33%	79.33%	26.19	29.25
Average	76.18%	81.40%	85.79%	91.12%	16.91	22.41

In the result of table 3.1, the SDR of the landmark 1 is as high as 98%, and the standard deviation value is considerably large. This is because there are few cases where forecasts are very wrong.

Table 3.2: Landmarks detection result for testset2

Landmarks	*SDR(2.0mm)*	*SDR(2.5mm)*	*SDR(3.0mm)*	*SDR(4.0mm)*	*MRE*	*SD*
1	98.00%	98.00%	98.00%	98.00%	13.65	55.97
2	85.00%	91.00%	95.00%	97.00%	12.01	13.65
3	4.00%	8.00%	15.00%	30.00%	48.35	24.92
4	58.00%	69.00%	76.00%	83.00%	31.73	49.79
5	52.00%	65.00%	69.00%	84.00%	23.95	17.07
6	7.00%	8.00%	15.00%	36.00%	52.96	32.11
7	100.00%	100.00%	100.00%	100.00%	6.81	4.20
8	97.00%	99.00%	100.00%	100.00%	7.61	4.62
9	100.00%	100.00%	100.00%	100.00%	5.60	3.81
10	46.00%	55.00%	64.00%	81.00%	27.62	25.50
11	87.00%	91.00%	96.00%	97.00%	10.13	11.14
12	97.00%	97.00%	98.00%	98.00%	7.02	11.90
13	16.00%	35.00%	65.00%	93.00%	27.38	7.69
14	51.00%	64.00%	81.00%	94.00%	21.57	10.59
15	89.00%	95.00%	99.00%	100.00%	11.35	6.94
16	74.00%	80.00%	85.00%	92.00%	17.37	21.08
17	83.00%	88.00%	93.00%	99.00%	16.13	30.54
18	80.00%	86.00%	92.00%	96.00%	14.69	11.38
19	67.00%	80.00%	82.00%	88.00%	35.07	81.01
Average	67.95%	74.16%	80.16%	87.68%	20.58	22.31

Unusually, Table 3.2 shows that the mean error is large at landmarks 13, but the standard deviation is quite small. This means that there is some consistent deviation between the data and the forecast. Since the SDR value in test1 is 95%, the characteristics of the data may be different. Also in landmark 3 and 6, the results of testset1 and testset2 are different, which is not clear because of differences in datasets or lack of versatility of models.

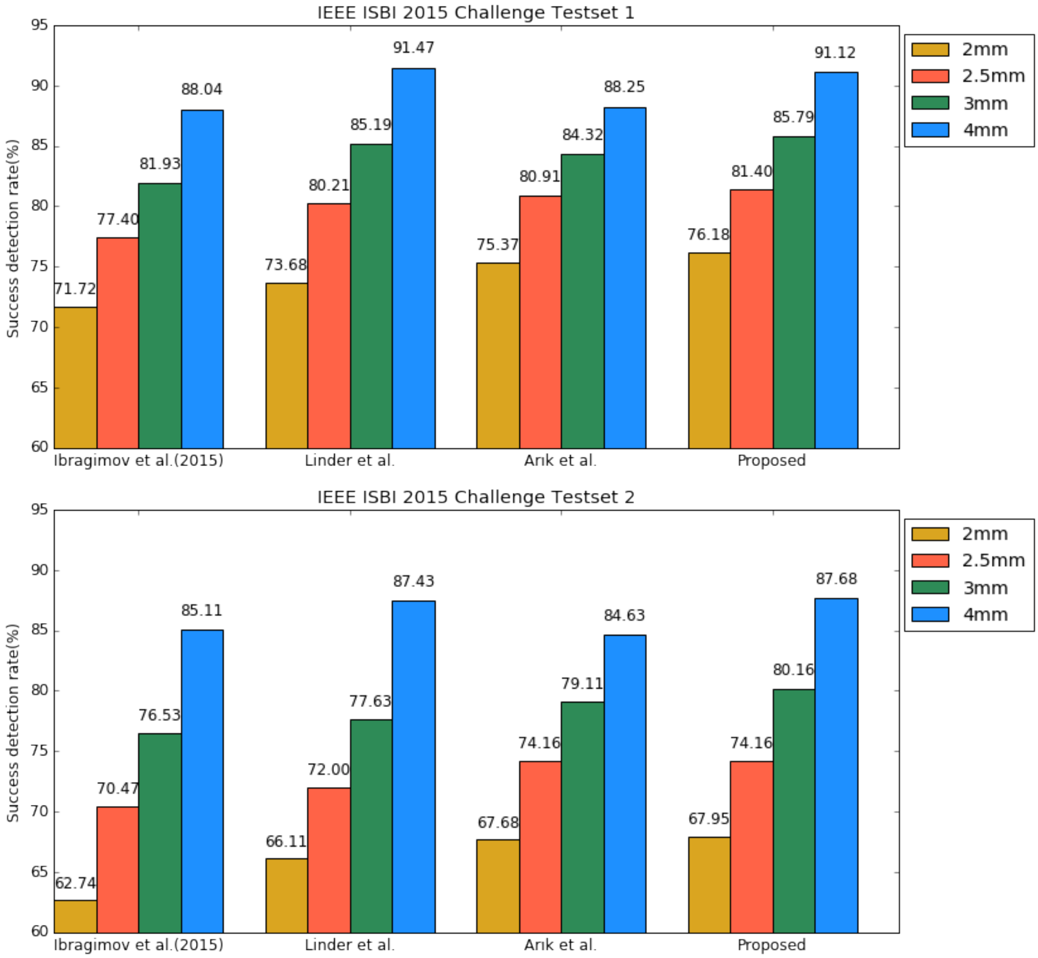

Figure 3.8: Success detection rates comparisons were made for the proposed model and the benchmark model. The test set is IEEE ISBI 2015 Challenge Datasets.

Chapter 4

Conclusion

The purpose of this thesis is to construct fully automated quantitative cephalometry as FCN model. The proposed model is trained to generate an archery target shape at the location of anatomical landmarks. The image patch, which is the input of the model, has a patch size based on the data for each landmark. Experiments were conducted to find landmarks by applying the proposed model to the published data. SDR of a medically available 2mm range is better than previous studies. However, some parts of the wrong prediction caused a considerable error. It would be better to add constraints that take into account biological characteristics. Also, given enough data, the model is expected to perform better. Since the proposed model is generic enough not to depend on this problem, it can be used for other anatomical landmark detection problems.

Bibliography

[1] A. Kaur and C. Singh, “Automatic cephalometric landmark detection using Zernike moments and template matching,” *Signal, Image Video Process.*, vol. 9, no. 1, pp. 117–132, 2015.

[2] C. W. Wang, C. T. Huang, M. C. Hsieh, C. H. Li, S. W. Chang, W. C. Li, R. Vandaele, R. Mar??e, S. Jodogne, P. Geurts, C. Chen, G. Zheng, C. Chu, H. Mirzaalian, G. Hamarneh, T. Vrtovec, and B. Ibragimov, “Evaluation and Comparison of Anatomical Landmark Detection Methods for Cephalometric X-Ray Images: A Grand Challenge,” *IEEE Trans. Med. Imaging*, vol. 34, no. 9, pp. 1890–1900, 2015.

[3] C. W. Wang, C. T. Huang, J. H. Lee, C. H. Li, S. W. Chang, M. J. Siao, T. M. Lai, B. Ibragimov, T. Vrtovec, O. Ronneberger, P. Fischer, T. F. Cootes, and C. Lindner, “A benchmark for comparison of dental radiography analysis algorithms,” *Med. Image Anal.*, vol. 31, pp. 63–76, 2016.

[4] J. T. L. Ferreira and C. d. S. Telles, “Evaluation of the reliability of computerized profile cephalometric analysis.,” *Braz. Dent. J.*, vol. 13, no. 3, pp. 201–204, 2002.

[5] V. Grau, M. Alcañiz, M. C. Juan, C. Monserrat, and C. Knoll, “Automatic Localization of Cephalometric Landmarks,” *J. Biomed. Inform.*, vol. 34, no. 3, pp. 146–156, 2001.

[6] B. Ibragimov, B. Likar, F. Pernuš, and T. Vrtovec, “Automatic Cephalometric X-Ray Landmark Detection by Applying Game Theory and Random Forests,” *Mach. Learn.*, 2014.

[7] C. Chen, W. Xie, J. Franke, P. A. Grutzner, L. P. Nolte, and G. Zheng, "Automatic X-ray landmark detection and shape segmentation via data-driven joint estimation of image displacements," *Med. Image Anal.*, vol. 18, no. 3, pp. 487–499, 2014.

[8] P. Geurts, "Automatic Cephalometric X-Ray Landmark Detection Challenge 2014 : A machine learning tree-based approach," 2014.

[9] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," 2016.

[10] A. Krizhevsky, I. Sutskever, and H. Geoffrey E., "ImageNet Classification with Deep Convolutional Neural Networks," *Adv. Neural Inf. Process. Syst. 25*, pp. 1–9, 2012.

[11] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A Survey on Deep Learning in Medical Image Analysis," no. 1995, 2017.

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," pp. 1–8, 2015.

[13] F. Milletari, N. Navab, and S. A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," *Proc. - 2016 4th Int. Conf. 3D Vision, 3DV 2016*, pp. 565–571, 2016.

[14] J. Kawahara, A. Bentaieb, and G. Hamarneh, "Deep Features to Classify Skin Lesions," vol. 6, pp. 6–9.

[15] W. Shen, M. Zhou, F. Y. B, C. Yang, and J. T. B, "Multi-scale Convolutional Neural Networks for Lung Nodule Classificatio," vol. 9123, pp. 588–599, 2015.

[16] J. Antony, K. McGuinness, N. E. O. Connor, and K. Moran, "Quantifying Radiographic Knee Osteoarthritis Severity using Deep Convolutional Neural Networks," 2016.

[17] G. Wu, M. Kim, Q. Wang, Y. Gao, S. Liao, and D. Shen, "Unsupervised deep feature learning for deformable registration of MR brain images," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8150 LNCS, no. PART 2, pp. 649–656, 2013.

[18] C. P. B, D. Stern, H. Bischof, and M. Urschler, "Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016," vol. 9902, pp. 230–238, 2016.

[19] S. Ö. Arık, B. Ibragimov, and L. Xing, "Fully automated quantitative cephalometry using convolutional neural networks," 2017.

[20] L. D. Le Cun Jackel, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, B. L. Cun, J. Denker, and D. Henderson, "Handwritten Digit Recognition with a Back-Propagation Network," *Adv. Neural Inf. Process. Syst.*, pp. 396–404, 1990.

[21] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, and G. Wang, "Recent Advances in Convolutional Neural Networks," pp. 1–37, 2015.

[22] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2323, 1998.

[23] R. Hecht-Nielsen, "Theory of the Backpropagation Neural Network," *Proc. Int. Jt. Conf. Neural Networks*, vol. 1, pp. 593–605, 1989.

[24] Y. A. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller, *Efficient BackProp*, vol. 0. 2012.

[25] V. Nair and G. E. Hinton, "Rectified Linear Units Improve Restricted Boltzmann Machines," *Proc. 27th Int. Conf. Mach. Learn.*, no. 3, pp. 807–814, 2010.

[26] E. Shelhamer, J. Long, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, 2017.

[27] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in *2016 IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 770–778, 2016.

[28] H. Shimodaira, “Improving predictive inference under covariate shift by weighting the log-likelihood function,” *J. Stat. Plan. Inference*, vol. 90, no. 2, pp. 227–244, 2000.

[29] S. Ioffe and C. Szegedy, “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,” 2015.

[30] h. Alexander Jung, “Imgaug,” 2017.

[31] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” pp. 1–15, 2014.

국문초록

치의학에서 계량적인 두부계측은 환자를 진단하거나 치료 하는데 필수적인 역할을 한다. 본 학위논문에서는 내부적으로 레지듀얼 연결을 가지고 있는 FCN(fully convolutional networks)모델을 이용해 자동화된 랜드마크 탐색모델을 제안한다. 제안된 FCN 모델은 랜드마크 근처의 이미지 패치가 입력되면, 양궁 과녁 모양의 히트맵을 출력 하도록 학습되었다. 학습과 테스트에 사용된 측면 두부 영상은 공개적으로 이용가능한 데이터셋을 이용했다. 학습에 사용된 이미지 패치는 트레이닝 데이터에 기반해 위치와 크기가 정해졌고, 인공적으로 데이터를 증가시켰다. 결과를 평가하는 방법으로 SDR(Success detection rate)을 사용했다. 테스트 셋을 이용해 학습된 모델을 평가하고 기존연구와 비교했다. 결과적으로 랜드마크가 기존의 연구보다 더 좋은 정확도로 탐지되었다. FCN모델이 정확하게 랜드마크 탐지하는데 가능성이 있음을 보였다.

주요어휘: 지도학습, 완전 컨볼루션 네트워크, 레지듀얼 네트워크, 두부 계측 지표, 치과 엑스레이 영상
학번: 2015-22686