Final Year Project Interim Report

Early Diagnosis of Scoliosis in Children from RGB-D Images Using Deep Learning

Supervisor
Dr. Kenneth M.K. Wong

Team:
Huang Siyi (3035335349)
Li Gengyu (3035331886)

February 2, 2020

Department of Computer Science, the University of Hong Kong

Abstract

Scoliosis is a medical condition where a person's spine has a sideways curve. It usually occurs before puberty and prevents the lungs from functioning normally in severe cases. Scoliosis is typically diagnosed from X-ray images, but diagnostic X-rays increase the risk of developmental problems and cancer in those exposed. This project aims to obviate children's X-ray exposure in the early diagnosis of scoliosis by generating X-ray images of children's backs from the corresponding harmless RGB-D images rather than capturing the X-ray images directly from the children. An RGB-D image of a back combines a standard RGB image with its depth image, and the depth image consists of the distance between each point on the back and the camera to track the spine. In this project, a deep learning model will be built and trained to synthesize X-ray images of children's backs from the corresponding RGB-D images. At this point, our team has completed software development and equipment setup for data collection. We also built and trained a deep learning model based on HRNet to detect landmark locations of the backs on the RGB-D images for further synthesis. In the next semester, we will continue to train this model with more data provided. Moreover, another deep learning model based on GAN will be built and trained to synthesize X-ray images of the children's backs from the corresponding RGB-D images and the predicted landmarks.

Table of Contents

List of Figures

List of Tables

Abbreviations

RGB-D image: a combination of an RGB image and its corresponding depth image.

RGB image: an image channel that defines red, green, and blue color components for each pixel.

Depth image: an image channel in which each pixel relates to a distance between the image plane and the object.

AIS: Adolescent Idiopathic Scoliosis

CNN: Convolutional Neural Network

GAN: Generative Adversarial Network

HRNet: High-Resolution Network

AKDK: Microsoft Azure Kinect DK

DK hospital: The Duchess of Kent Children's Hospital at Sandy Bay

1. Introduction

This section introduces the background information, objective, and breakdown structure of this project.

1.1 Background Information on Scoliosis

Scoliosis is a medical condition where a person's spine has a sideways curve. Adolescent Idiopathic Scoliosis (AIS) is the most typical scoliosis that occurs in children at the onset of puberty [1]. Researchers found that the prevalence of AIS was up to 5.2% among adolescents, and the extension of AIS may result in severe complications from injuries to the heart and lungs [2]. The early diagnosis of scoliosis in children is thus fundamental for precluding the severe consequences and guiding the early treatments. The current golden standard for AIS assessment is to measure the Cobb Angles from X-ray images. A Cobb Angle is the curvature of a spine in degrees, and the curvature is currently estimated by detecting and visualizing each vertebra from the X-ray image. A noteworthy drawback of this Cobb Angle measurement is its reliance on X-ray images. X-ray devices expose patients to radiation, and the radiation can increase the risk of developmental problems and cancer in those exposed. Overall, X-ray images are indispensable in the current diagnosis of scoliosis, but the process of capturing X-ray images is harmful to patients, especially children.

1.2 Project Objective

Given the significance of the early diagnosis of scoliosis in children and the adverse effects of the current diagnostic method, our team proposed to generate X-ray images of children's backs from the corresponding RGB-D images to obviate children's X-ray exposure. A deep neural network, which takes an RGB-D image of a child's back as an input to produce a corresponding X-ray image, is expected to be built and trained throughout the whole project. An RGB-D image combines an RGB image with its depth image in which each pixel relates to a distance between the image plane and the corresponding point on the object in the RGB image. A traditional RGB image only contains the information on the surface of the back as it is two-dimensional. With the addition of depth information, an RGB-D image becomes a valid solution in this project because the depth information adds a third dimension to the original RGB image so that the spine can be detected, and its curvature can be measured. The current diagnostic method will remain the same except that the Cobb Angle measurement will base on the synthetic X-ray image instead of the image taken directly by an X-ray machine. In addition to the spine curve, the synthetic X-ray images may provide more medical information like real X-ray images. Children can, therefore, avoid radiation exposure in the early diagnosis of scoliosis, but still receive reliable diagnosis and treatments.

1.3 Project Workflow

This project contains three main phases. The workflow of this project is illustrated in Figure 1. The project is currently at the end of the second phase. During the inception phase, our team implemented an interface to control the camera AKDK for capturing RGB-D images of children's backs. Research assistants are responsible for taking RGB-D images and labeling landmarks, which are some predefined critical points of a patient's back on the images in DK hospital. In the second phase, our team built and trained a deep learning model based on HRNet to perform landmark detection that estimates the critical points of the patient's backs on the RGB-D images. In the third phase, another deep learning model will be built and trained to synthesize X-ray images from the detected landmarks in the second phase and the original RGB-D images. We expect that the predicted landmarks in the second phase will increase the accuracy and reliability of the model in the third phase. Eventually, this model is supposed to generate X-ray images of the children's backs from the corresponding RGB-D images.

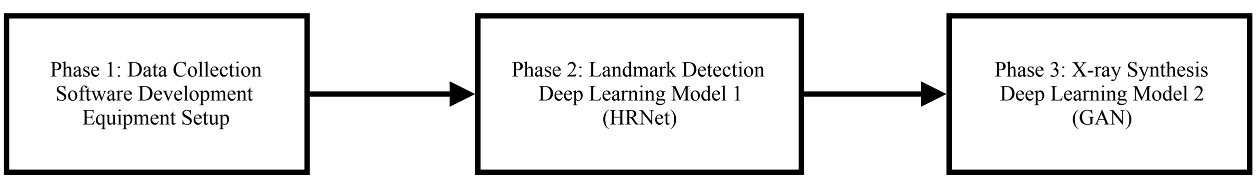

Figure 1: The flow chart of this project.

In the remainder of this report, separate sections introduce different contents related to this project. Previous works to which this project refers are listed in section 2. Section 3 demonstrates primary methods used in this project, while section 4 explains our experiments and results.

Section 5 outlines the project progress, section 6 suggests a tentative plan for future development, and section 7 concludes the whole report.

2. Related Works

This section presents previous works that form the academic basis of this project.

2.1 Cobb Angle Measurement

This project aims to generate X-ray images on which the Cobb Angle relies since it is the current golden standard for the diagnosis of scoliosis. A Cobb Angle is defined as the greatest angle at a particular region of the vertebral column when measured from the superior endplate of a superior vertebra to the inferior endplate of an inferior vertebra [3]. The measurement of a Cobb Angle is visualized in Figure 2. In the medical field, a person whose Cobb Angle exceeds 10 degrees is scoliotic [4]. The current diagnosis of scoliosis extensively relies on measuring the Cobb Angle from X-ray images. The Cobb Angle is preferably measured while standing, since lying down decreases Cobb angles by some degrees. In this project, patients stand in front of AKDK according to instructions to capture the RGB-D images of the backs, which ensures the project's conformity to the Cobb Angle measuring standard.

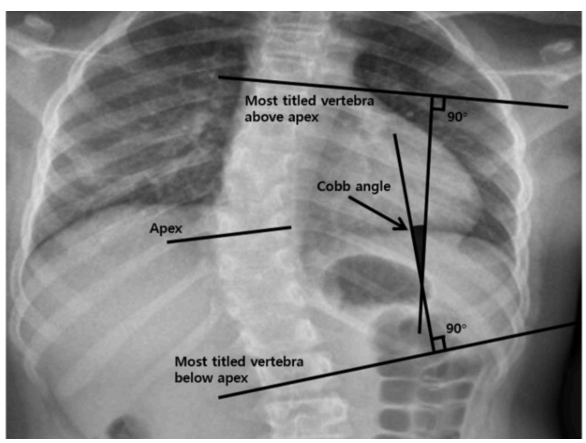

Figure 1: A Cobb Angle measurement.

2.2 Landmark Detection with HRNet

The Hight-Resolution Network (HRNet) used in the second phase of this project for landmark detection on RGB-D images was developed by Ke et al. for obtaining strong high-resolution presentations of a pixel-labeled or region-labeled image that stores and represents an image in pixel or region level [8]. When performing facial landmark detection, the resulting high-resolution representations produced by HRNet were proved to be not only strong but also spatially precise. The HRNet can be considered as a reliable and promising framework in pixel-level classification tasks according to its lower error rates of landmark detection outcomes on the standard datasets WFLW, AFLW, COFW, and 300W [8]. The HRNet, therefore, is expected to give a satisfactory performance in the landmark detection tasks on RGB-D images in this project since they are pixel-level classification tasks as well.

2.3 X-ray Images Generation with GAN

GAN, which will be the primary framework in the third phase of this project to generate X-ray images, is a hybrid deep neural framework used extensively in machine learning problems. Brain et al. successfully generated X-ray images from the surface geometry of backs with GAN [9]. Although their input image type is different from that of this project, the digital representations of these two types of images are similar. Their successful generation of X-ray images from another type of image with GAN provides a head start into this project, and the architecture of their model will be useful guidance for the model design and implementation in the third phase of this project.

3. Methodology

This section demonstrates the primary techniques used in the second and third phases of this project, respectively.

3.1 High-Resolution Network (HRNet)

An HRNet is essentially a CNN-based framework proposed by Ke et al. to learn reliable high-resolution representations of an image [8]. It maintains high-resolution representations of an input image by connecting high-to-low resolution convolutions in parallel and produces strong and spatially precise high-resolution presentations by repeatedly conducting multi-scale fusions across parallel convolutions [8]. The underlying architecture of an HRNet is shown in [8, Fig. 3] that contains four stages. The first stage consists of basic high-resolution convolutions. The remaining stages repeat two-resolution, three-resolution, and four-resolution blocks. Repeated information is exchanged across parallel convolutions to increase the learning capability of the model. Moreover, the capacity of this multi-resolution convolution is fully explored by aggregating outputs from all stages to produce the outcome. Therefore, the HRNet is a robust framework for landmark detection on RGB-D images in the second phase. So far, our team has built a deep learning model based on HRNet to perform landmark detection on RGB-D images. The model was also trained with 300 RGB-D images and validated with 75 RGB-D images.

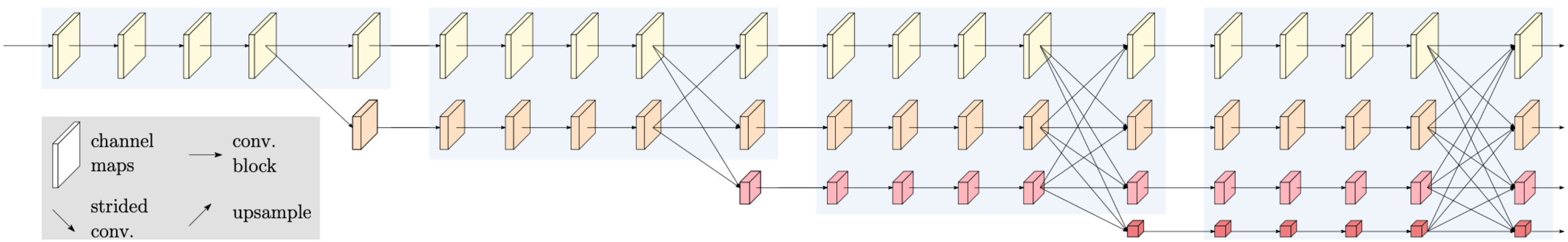

Figure 2. [8] The underlying architecture of an HRNet.

3.2 Generative Adversarial Network (GAN)

GAN is a class of machine learning systems invented by Ian Goodfellow et al. in 2014, containing two neural networks contesting with each other, namely a generator and a discriminator [10]. Given a training set, this technique learns to generate new data with the same statistics as the training set. Since its invention, GAN has been adopted in many popular research fields in machine learning, such as image reconstruction and image synthesis, and it has yielded satisfactory results. In this project, GAN will be used as a basic framework in the third phase, and project-specific configurations will be included according to the input and output requirements of this project.

4. Experiments and Results

This section gives the experiments our team performed with HRNet and RGB-D images in the second phase and evaluates the results.

4.1 Learning Rate and Learning Scheduler of HRNet

The learning rate is a crucial hyperparameter that determines the performance of a deep learning model. Our team carried out several experiments on the CS GPU farm by tuning the learning rate of HRNet to find the model with the least loss on convergence. Pattern [A| E1, E2 | E3] denotes the experiment configuration with the initial learning rate being A and learning rate scheduler set at epoch E1 and E2 with total epochs being E3.

Initially, we trained a model with [0.0001 || 30], and its loss converged at epoch six at the level of around 4.78. According to this initial experiment, we chose six different initial learning rates around 0.0001. For each value, we set 3 different schemes of learning rate schedulers. In each experiment, the minimum losses during both the training and validation process measure the performance of the model. The results are shown in Table 1.

Experiments	Configurations	Loss Trend	Training Results	Validating Results
EX1	[0.00001 | 7, 15 | 30]	Decrease overall, does not converge	4.34	4.78
EX2	[0.00001 | 15, 25 | 30]	Decrease overall, does not converge	4.22	4.67
EX3	[0.00001 | 7, 25 | 30]	Decrease overall, does not converge	4.28	4.79
EX4	[0.00005 | 7, 15 | 30]	Decrease overall, converge at 14	4.04	4.11
EX5	[0.00005 | 15, 25 | 30]	Decrease overall, converge at 27	3.94	4.01
EX6	[0.00005 | 7, 25 | 30]	Decrease overall, converge at 14	4.04	4.05
EX7	[0.0001 | 7, 15 | 30]	Decrease overall, converge at epoch 19	**3.89**	**3.97**
EX8	[0.0001 | 15, 25 | 30]	Decrease overall, converge at epoch 19	3.98	4.01
EX9	[0.0001 | 7, 25 | 30]	Decrease overall, converge at epoch 27	**3.95**	**3.99**
EX10	[0.0005 | 7, 15 | 30]	Decrease overall, converge at 24	4.10	4.12
EX11	[0.0005 | 15, 25 | 30]	Decrease overall, converge at 28	4.46	4.80
EX12	[0.0005 | 7, 25 | 30]	Decrease overall, converge at 26	4.21	4.38
EX13	[0.001 | 7, 15 | 30]	Decrease overall, does not converge	6.01	7.43
EX14	[0.001 | 15, 25 | 30]	Gradient explode at epoch 12	N/A	N/A
EX15	[0.001 | 7, 25 | 30]	Decrease overall, converge at epoch 29	6.08	7.89
EX16	[0.005 | 7, 15 | 30]	Gradient explode at epoch 5	N/A	N/A

EX17	[0.005 | 15, 25 | 30]	Gradient explode at epoch 5	N/A	N/A
EX18	[0.005 | 7, 15 | 30]	Gradient explode at epoch 5	N/A	N/A

Table 1. Experiment Results.

According to the experiment results, 0.0001 was the best initial learning rate because of its average lower minimum loss on the training and validation dataset. However, the learning rate scheduler impacted differently on different experiments with various initial learning rates. Since configuration [0.0001 | 7, 15 | 30] had the minimum training and validation loss, we could estimate an interval where the best learning rate scheduler located. The scheduler at epoch 7 had a good impact on the performance of the model, as the model seems to converge at epoch 7 in the initial experiment. The predicted landmarks of this model on one sample RGB image and its depth image are visualized in Figure 3 and Figure 4, separately.

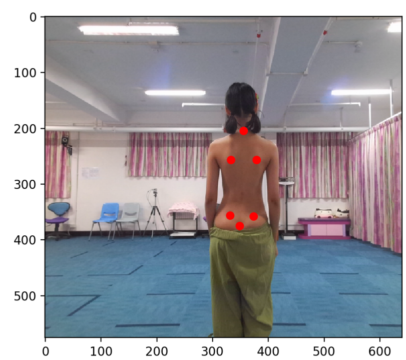

Figure 3. Predicted Landmarks on the RGB Image.

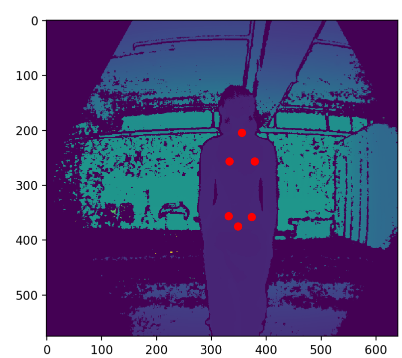

Figure 4. Predicted Landmarks on the Depth Image.

5. Project Progress

This section identifies the current status of this project and points out two main limitations of this project.

5.1 Current Status

In the first phase, our team implemented an interface to control AKDK and capture the RGB-D images. We finished this equipment setup process by the end of October 2019. Meanwhile, we also identified the project scope, completed a detailed project plan, built a project website, and confirmed the landmark locations for future labeling and detection. In the second phase, we identified the HRNet as the basic framework for landmark detection after relevant researches. Our team, therefore, built a deep learning model based on HRNet to perform landmark detection on RGB-D images. So far, we have collected 375 RGB-D images from DK hospital. Our team has trained the model based on 300 RGB-D images, with some data augmentation techniques, and validated the model with the remaining 75 images. Table 5. gives a detailed timetable for the previous stages.

Date	Progress	Deliverable
9/1/2019 - 9/30/2019	Identified project scope. Planned for the project.	A detailed project plan. The project website.
10/1/2019 - 10/31/2019	Developed an interface to control AKDK. Set up AKDK to capture RGB-D images.	A program for controlling AKDK. Captured RGB-D images.
11/1/2019 - 12/31/2019	Built a deep learning model based on HRNet.	A basic structure of the HRNet.
1/1/2020 - 2/2/2020	Trained and adjusted the HRNet	A trained and tested HRNet A detailed interim report

Table 2. The previous stages in this project.

5.2 Data Availability Problem

The most critical problem of this project is the limited availability of data. The data collection process in DK hospital started after the inception phase, and the estimated amount of data collected every week is around 50. However, the previous deep learning projects that succeeded in obtaining a reliable and precise deep neural network trained their models with thousands of data. Therefore, the shortage of data is a prominent obstacle to this project. Our team proposed to perform data augmentation on the existing data that involves rotation, flipping, translation, and depth-offset to produce more data for training purposes so that the model can be more robust.

5.3 Data Inconsistency Problem

Another noticeable limitation of this project is the inconsistency between the same patient's X-ray image and the RGB-D image. In this project, the RGB-D images taken by AKDK will serve as input data for the HRNet to detect landmarks in the second phase and for GAN to synthesize X-ray images in the third phase. The patients' X-ray images will be ground truth for GAN. The main concern is that these two types of images are not taken at the same time for the same patients, which may result in the discrepancy between a patient's relative positions in the RGB-D image and that in the X-ray image. Currently, this problem is mitigated by giving instructions to patients when taking the images so that the same patient's relative positions in these two images are almost the same.

6. Future Plan

The last phase of this project generates X-ray images of children's backs from the corresponding RGB-D images, and it is supposed to be finished by the end of April 2020. A detailed timetable is given in Table 7. The Cobb Angle will be measured based on the synthetic X-ray image instead of the X-ray image taken directly by an X-ray machine. The third phase is the one that delivers the final product of this project. GAN will be used as a basic framework to produce a project-specific generative model. The model will input RGB-D images and estimated landmarks from the second phase to synthesize the corresponding X-ray images. It will also take real X-ray images of the children's backs as ground truth to perform deep learning. Like any other deep neural networks, the amount of data determines the performance of this model. Therefore, our team will also perform data augmentation to increase the accuracy and reliability of this model. The two deep learning models in this project will be finalized before the final presentation.

Date	Progress	Deliverable
2/3/2020 - 3/1/2020	Build a deep leaning model based on GAN.	A runnable model for X-ray synthesis.
3/1/2020 - 3/31/2020	Train and adjust the model.	A trained and tested model for X-ray synthesis.
4/1/2020 - 4/19/2020	Finalize the two deep learning models.	Two finalized deep learning models.
4/20/2020 - 4/24/2020	Prepare for the final presentation.	A final presentation for the whole project.

Table 3. A detailed plan for X-ray synthesis in the third phase and the final presentation.

7. Conclusion

In conclusion, this project aims to generate X-ray images of children's backs from the corresponding RGB-D images using deep learning in the early diagnosis of scoliosis. The project contains three phases, and the major deliverables are two deep learning models for landmark detection and X-ray synthesis based on HRNet and GAN, respectively. So far, our team has completed all the preliminary work for data collection. An HRNet was built and trained for landmark detection on RGB-D images in the second phase. We will continue to train this model with more data provided in the future. Another deep learning model for X-ray synthesis based on GAN will be built and trained at the later stage.

References

[1] Weinstein, S.L., Dolan, L.A., Cheng, J.C., Danielsson, A., Morcuende, J.A., 2008. Adolescent idiopathic scoliosis. The Lancet 371, 1527–1537.

[2] Asher, M.A., Burton, D.C., 2006. Adolescent idiopathic scoliosis: natural his- tory and long term treatment effects. Scoliosis 1, 2.

[3] Brian D. Coley., 2013. Caffey's Pediatric Diagnostic Imaging (12 ed.). Elsevier Health Sciences.

[4] K. Lau. The Complete Scoliosis Surgery Handbook for Patients: An In-Depth and Unbiased Look Into What to Expect Before and During Scoliosis Surgery. Health In Your Hands. P89.

[5] V. Bonnet, T. Yamaguchi, A. Dupeyron, S. Andary, A. Seilles, P. Fraisse, and G. Venture., 2016 Application of an RGBD augmented C-arm for minimally invasive scoliosis surgery assistance.

[6] S. Habert, U. Eck, P. Fallavollita, S. Parent, N. Navab, and F. Cheriet., 2017 Automatic Estimate of Back Anatomical Landmarks and 3D spine curve From a Kinect Sensor

[7] R. Choi, K. Watanabe, H. Jinguji, N. Fujita, Y. Ogura, S. Demura, T. Kotani, K. Wada, M. Miyazaki, H. Shigematsu, and Y. Aoki. CNN-based Spine and Cobb Angle Estimator Using Moire Images. In IIEEJ, 2017.

[8]K. Sun, Y. Zhao, B. Jiang, T. Cheng, B. Xiao, D. Liu, Y. Mu, X. Wang, W. Liu, and J. Wang. High-Resolution Representations for Labelling Pixels and Regions. In arXiv, 2019.

[9] B. Teixeira, V. Singh, T. Chen, K. Ma, B. Tamersoy, Y. Wu, E. Balashova, and D. Comaniciu. Generating Synthetic X-ray Images of a Person from the Surface Geometry. In CVPR, 2018.

[10] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative Adversarial Nets. In arXiv: 1406.2661, 2014.