

Master's Thesis

Gait Quality Assessment Using
Unsupervised Deep Learning Model
for Cerebral Palsy

Ginga Sumi

Division of Electrical and Electronic Engineering
Graduate School of Engineering
Mie University

Contents

1	Introduction	1
1.1	Background	1
1.2	Objective	2
2	Related Works	4
2.1	Gait Deviation Index (GDI)	4
2.2	Human Pose Estimation	6
2.3	Unsupervised Anomaly Detection	7
2.4	Gait Recognition	8
3	Experimental Materials	10
3.1	Materials	10
4	Method	12
4.1	Feature Engineering	12
4.2	Data Preprocessing	13
4.3	Outline of Proposed Method	13
4.4	Auto-Encoder Model	14
4.5	Memory-augmented Auto-Encoder Model	14
4.5.1	Mechanism of Memory-augmented Auto-Encoder	15
4.5.2	Objective Function for MemAE	17
4.6	Variational Auto-Encoder Model	18
4.6.1	Mechanism of Variational Auto-Encoder	18
4.6.2	Probabilistic Reconstruction Error	18
4.6.3	Regularizer for Objective Function	20
5	Experimental Results and Discussion	23
5.1	Experimental Set-Up	23
5.2	Implementation of Model	24
5.3	Result	25

6 Conclusion	33
6.1 Conclusion	33
6.2 Future Works	33
Acknowledgment	35
Reference	36
Publication List	39

List of Figures

1.1	Diagram of the treatment process for CP	2
1.2	Overall flow of the proposed method	3
2.1	Measurement method of the GDI	5
2.2	Human pose estimation	6
3.1	Time-series data used for the experiment	10
4.1	Illustration of creating the data	13
4.2	Network of the Auto-Encoder	14
4.3	Network of the Memory-augmented Auto-Encoder	15
4.4	Network of the Variational Auto-Encoder	19
4.5	Flow of the VAE in the proposed method	21
4.6	Comparison between the case of higher and lower standard deviation . .	22
5.1	Structure of the model	24
5.2	Gait score (reconstruction probability) vs. GDI	26
5.3	Examples of the input and reconstructed data in the VAE	27
5.4	Examples of the input and reconstructed data within the abnormal data	28
5.5	Comparison between the reconstruction probability and the reconstruction error	29
5.6	Structure of the model in the preliminary experiment	31

List of Tables

4.1	Description of key point used for proposed method	12
5.1	Number of dataset used for experiment	23
5.2	Comparison of Correlation Coefficient among Model	25
5.3	Difference of Results with Value of α	30
5.4	Result of Preliminary Experiments	30
5.5	Caption	32

Chapter 1

Introduction

1.1 Background

Cerebral Palsy (CP) is a movement disorder resulting from abnormal brain development or brain damage that occurs from prenatal to infancy [1]. The symptoms of CP vary among individual patients, mainly impairing body movement and muscle coordination. Since these conditions are not cured completely, the primary goal of its treatment is to sustain or improve the physical function needed for daily life. Therefore, continuous diagnosis and proper treatment are essential. To achieve this, treatment is conducted following the process in Fig. 1.1. In this process, the gait quality of the patient is assessed by experts at first. Subsequently, doctors diagnose the symptom and select the treatment such as physical therapy or medical therapy based on the assessment result. Moreover, the gait quality is assessed again to evaluate whether the conducted treatment was effective. This process is repeated to realize the continuous diagnosis and proper treatment. In particular, gait quality assessment has a significant role in the clinical environment [2]. A laboratory-based optical motion capture system is the gold standard for clinical gait quality assessment nowadays. It quantitatively assesses patients' motor function quality and helps medical decision-making. However, motion capture systems require significant burdens and highly trained personnel (e.g., physical therapists), which leads to preventing routine diagnoses. Moreover, collected data in the laboratory may fail to capture how patients move in natural situations [3].

Recent advances in human pose estimation using deep learning enable us to obtain the image-plane position of anatomical key points (e.g., ankle, knees, etc.) without a motion capture system [4–8]. Some research proposed the method using human pose estimation techniques to predict the clinical gait parameters from the given videos [9–11]. However, their methods require special equipment to collect labels (i.e., ground truth) of the training data. On the other hand, unsupervised anomaly detection methods can overcome the limitation of data collection [12–17]. Irregular events of human behavior or quantifying the gait quality can be realized by adopting unsupervised anomaly detection

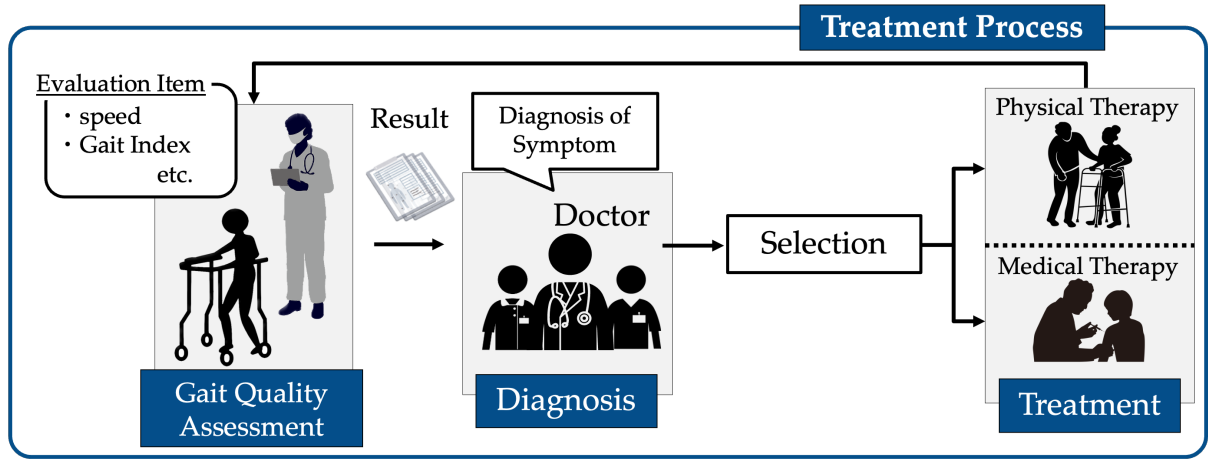


Figure 1.1: Diagram of the treatment process for CP

methods to the video anomaly detection domain.

1.2 Objective

This study aims to establish a method for estimating patients' gait quality without special equipment. The author proposed the method using deep learning to achieve the study objective. Fig. 1.2 shows the overall flow of the proposed method. In the proposed method, the only required thing for users is a standard camera (e.g., a smartphone). Initially, the key points are extracted from the video in chronological order using a DL-based human pose estimation technique. Subsequently, the author creates time-series data from the extracted key points and uses them for the DL-based gait quality estimation method. Finally, the gait score indicating gait abnormalities is obtained. In this thesis, the author used OpenPose [7] for human pose estimation and focused on establishing the gait quality estimation model using unsupervised deep learning method.

This thesis describes the methods for estimating the gait quality of CP patients from videos. The author will also discuss the future possibilities and problems of the methods and describe future developments.

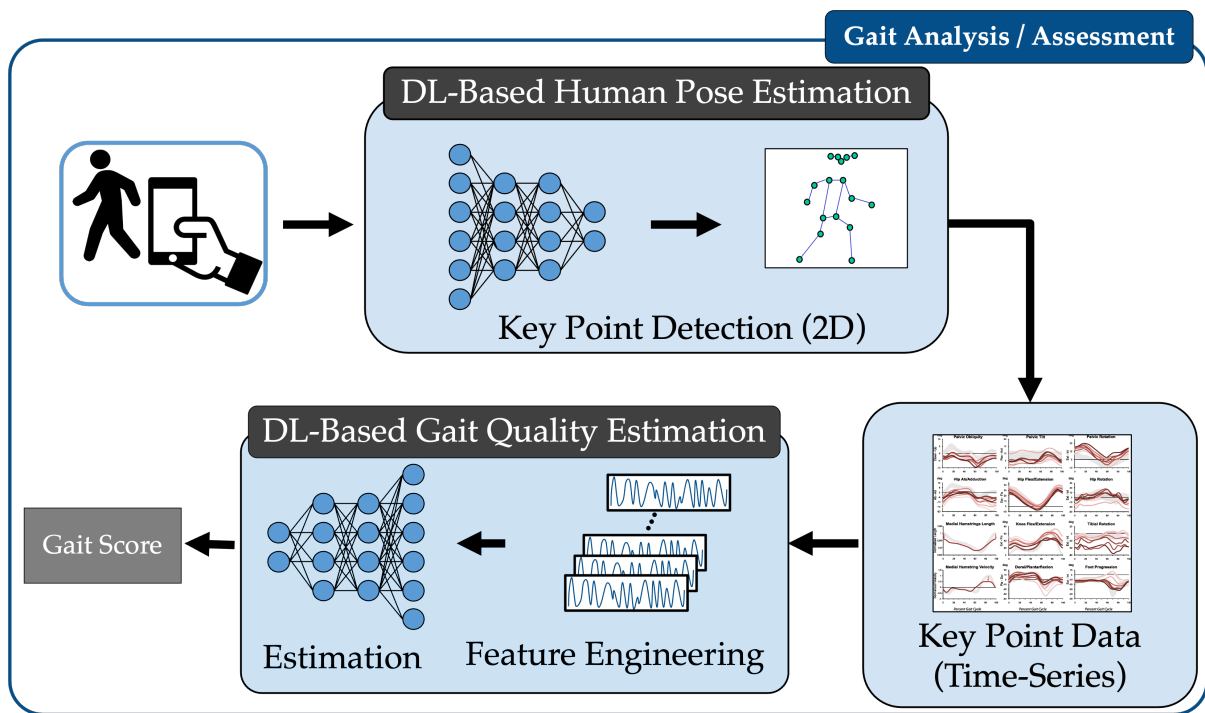


Figure 1.2: Overall flow of the proposed method

Chapter 2

Related Works

2.1 Gait Deviation Index (GDI)

The Gait Deviation Index (GDI) [18] is one of the clinical gait indexes to measure overall gait pathology. As shown in Fig. 2.1, the GDI is calculated from the joint data of the lower body obtained by optical motion capture. Each side of the body has its own unique GDI. The method used in calculating the GDI is motivated by the “eigenface” method. The eigenface is the method used for face identification systems.

In the method, a large collection of digitized faces is converted to vectors. This collection of vectors is compressed into a small number of the eigenvectors (called eigenfaces) by Principal Component Analysis (PCA). The eigenfaces that include the information of the original collection of faces are combined to create a reduced-order approximation of any given face. The similarity of one face to others is measured using each reduced-order approximation.

In the gait analysis, a set of kinematic quantities (digitized gait) collected by a laboratory-based optical motion capture system is used instead of digitized faces. The kinematic quantities (e.g., Pelvic and Hip angles, etc.) during a gait cycle ($9 \text{ angles} \times 51 \text{ points} = 459 \text{ data}$) are extracted for each side of the body and are converted to 459×1 gait vector \mathbf{g} . The gait vectors of all the subjects are combined to create the gait matrix $\mathbf{G} = (\mathbf{g}_1, \mathbf{g}_2 \dots)$. The Singular Value Decomposition (SVD) is performed and an optimal orthonormal basis $\{\hat{\mathbf{f}}_1, \hat{\mathbf{f}}_2, \hat{\mathbf{f}}_3, \dots, \hat{\mathbf{f}}_{459}\}$ (gait feature) is obtained. Given the gait features, an m -th order approximation of any gait vector \mathbf{g} can be computed as

$$\tilde{\mathbf{g}}^m = \sum_{k=1}^m c_k \hat{\mathbf{f}}_k \quad (2.1)$$

where the feature component c_k are

$$c_k = \mathbf{g} \cdot \hat{\mathbf{f}}_k. \quad (2.2)$$

The feature components is arranged as $\mathbf{c} = (c_1, c_2, \dots, c_m)$.

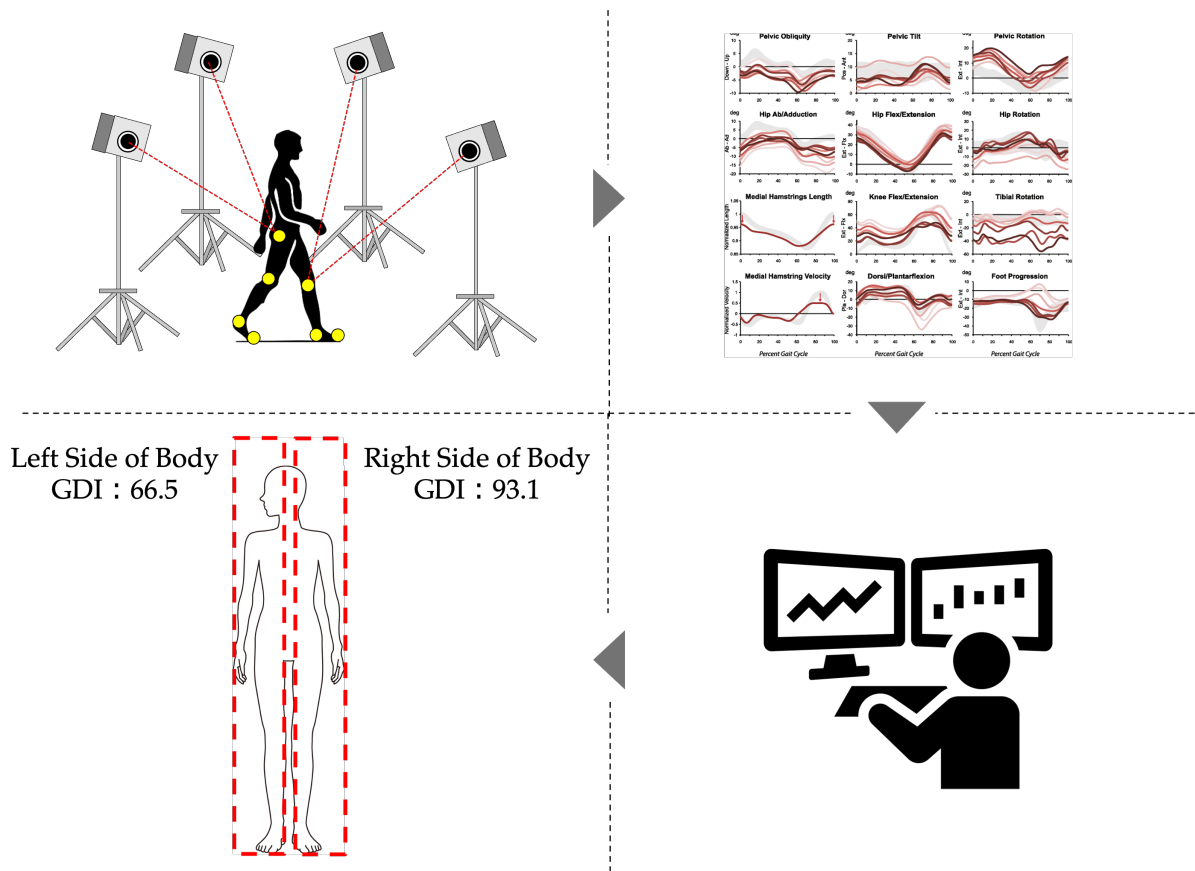


Figure 2.1: Measurement method of the GDI

Given average of the feature components obtained from the gait of a control group (e.g., typically developing - or TD - children) \mathbf{c}^{TD} , the distance ($d^{\alpha, \text{TD}}$) of a target child α from the TD gait is calculated using the Euclidean distance.

$$d^{\alpha, \text{TD}} = \|\mathbf{c}^{\alpha} - \mathbf{c}^{\text{TD}}\| \quad (2.3)$$

The raw GDI is defined by

$$\text{GDI}_{\text{raw}}^{\alpha} = \ln(d^{\alpha, \text{TD}}). \quad (2.4)$$

For interpretability, the z-score of subject α is calculated as

$$z\text{GDI}_{\text{raw}}^{\alpha} = \frac{\text{GDI}_{\text{raw}}^{\alpha} - \text{Mean}(\text{GDI}_{\text{raw}}^{\text{TD}})}{\text{S.D.}(\text{GDI}_{\text{raw}}^{\text{TD}})} \quad (2.5)$$

where $\text{Mean}(\text{GDI}_{\text{raw}}^{\text{TD}})$ and $\text{S.D.}(\text{GDI}_{\text{raw}}^{\text{TD}})$ are the sample mean and standard deviation of the control group, respectively. As a result, the GDI of the target child α is defined by the following equation.

$$\text{GDI}^{\alpha} = 100 - 10 \times z\text{GDI}_{\text{raw}}^{\alpha} \quad (2.6)$$

If the obtained score GDI^{α} is 100 or higher, the target child has no gait pathology. For every ten drops in the GDI, the target child's gait deviates by a standard deviation from the gait of a control group. In other words, the target child's gait is regarded as a severe case when the GDI is smaller than 100. The GDI has a clear trend that varies with the disease severity of CP, making it suitable for assessing gait quality.

2.2 Human Pose Estimation

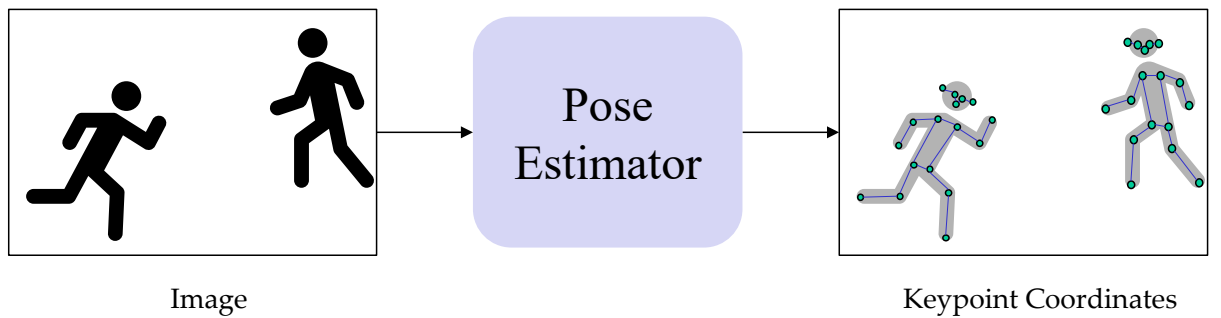


Figure 2.2: Human pose estimation

Human pose estimation in the computer vision domain refers to the task of extracting the location of key points (joints of a person) from the image. Convolutional Neural Networks (CNNs) are used exclusively for human point estimation. This is the current trend in image-related tasks in general and has resulted in high accuracy. The mainstream

method of key point estimation does not directly estimate the coordinates of the key points but estimates the probability that the key points exist in certain pixels by using a heatmap. The heatmap is estimated for each key point, and the coordinates with the highest probability are finally used as the coordinates of the key point.

Key points are estimated in 2D or 3D. In the case of 2D, (x, y) coordinates of key points in pixels are presented originating from the upper left corner of the image. In the case of 3D, (x, y, z) coordinates of key points are presented by estimating the depth direction. The ground truth for the training data is also required to have the number of dimensions that match the dimensions to be estimated. It is more challenging to collect training data for 3D than 2D. In addition, the accuracy is not high when estimating 3D coordinates due to the computational complexity.

In the method of key points estimation, there are two approaches: top-down and bottom-up. In the top-down approach, the person is first detected in the given image, and then the pose is estimated for each detected person. Therefore, more computing resources are required when estimating the pose of multiple people. On the other hand, the bottom-up approach estimates the coordinates of key points in the given image, and then groups key points to form the pose for each person. Since key points are estimated at once in this method, it works fast in estimating the pose of multiple persons. In addition, this approach is conducted in end-to-end learning from the given image to key points output. However, the calculated accuracy is not often reliable when the distance from the camera and the scale are different from the setting of collecting the training data.

To be available in crowded situations, Bottom-up networks have been actively studied. Moreover, almost all of the current representative pose estimation networks are bottom-up approaches.

2.3 Unsupervised Anomaly Detection

Anomaly detection refers to the identification of events or objects that do not conform to expected behavior. This task has been applied in various domains, such as the manufacturing industry, network security, and video surveillance. However, abnormal events are unbounded in real applications and the definition of abnormality is sometimes ambiguous. It is almost infeasible to collect all kinds of abnormal events and tackle the task with a classification method. From these backgrounds, unsupervised anomaly detection has been getting attention.

Unsupervised anomaly detection is detecting the anomaly data using deep learning without the labeled (i.e., ground truth). In unsupervised anomaly detection, an effective solution is to feed the normal data to the model to learn its features. Most existing

methods train the model by reconstructing the normal data. By using the trained model, the anomaly data can be detected as an outlier. Based on this approach, some researchers have proposed the method using the Auto-Encoder (AE) type model [12–17]. The AE is the neural network with a nonlinear dimensionality reduction mechanism and can detect the subtle anomaly data [17]. In addition, models with the extended structure of the AE (e.g., AE incorporating Recurrent Neural Network or Variational Auto-Encoder) have been proposed, and the research that applies these models to video surveillance is also over the active areas.

2.4 Gait Recognition

Gait recognition is a task to identify people by their walking [19]. “Gait” is usually defined as how a person walks, including the individuality of the human body. To obtain the gait, video capture with a camera is often used. Therefore, this is a technical domain of computer vision. It is getting popular because it can be accomplished even with low-quality video using uncomplicated devices (e.g., smartphones) and personal information (e.g., face or fingerprints) are not required. In addition, it is challenging to imitate how a person walks. In the field of computer vision, there has been a great deal of progress in the use of deep learning methods, and deep learning is used in the most advanced techniques in the field of gait recognition. In order to evaluate a person’s gait, obtaining a body representation of the person from the video is required. There are two major paradigms for body representation, i.e., appearance-based methods using contours and model-based methods using skeletons.

Model-based Methods In the method, the skeleton is used to represent the human body. The skeleton is defined by the representative points of the joints of the human body (i.e., keypoints) and their connections. The skeleton is extracted using a technique called pose estimation, which estimates the coordinates of each keypoint from the image. Pose estimation is performed for all frames of the walking video, and time-series data of keypoint coordinates are obtained. This is used for gait parameter estimation. It is robust to the clothing and equipment of the subject but heavily depends on the image quality and the accuracy of the pose estimation.

Appearance-based Methods In this method, silhouettes are used to represent the body. The silhouette is obtained by estimating the boundary line (contour) between the person and the background and binarizing the inside and outside of the contour. The computational cost for extracting the silhouette is, therefore, very small. The same silhouette extraction is performed for all video frames, and a time series of silhouette

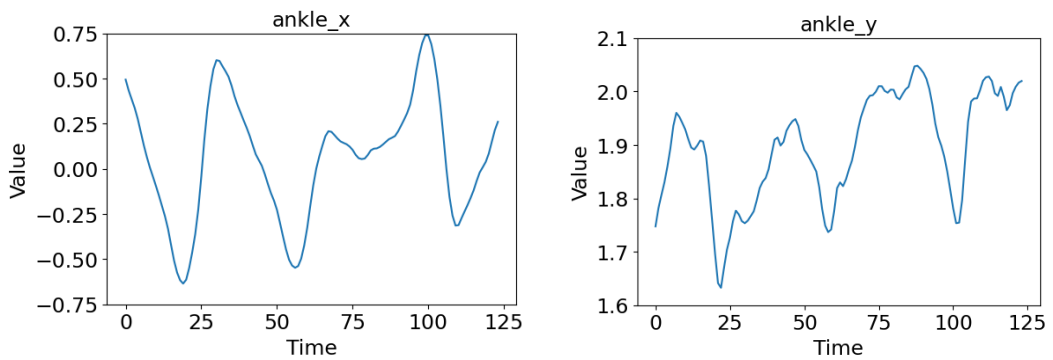
images is obtained. This image sequence, either directly or after conversion, such as Gait Energy Image (GEI) [20] or Gait Entropy Image (GEnI) [21], is used for gait parameter estimation. On the other hand, the silhouette extraction is affected by the target's clothing and ornaments, and information about the target's orientation in the front-back direction is missing. Nevertheless, it is a more popular method because it does not require coordinate estimation, and its accuracy is not affected. It is an end-to-end method with the image sequence intact.

Chapter 3

Experimental Materials

3.1 Materials

The dataset of CP patients' gait published by Kidziński's study [11] was analyzed in the experiments. The dataset includes 667 videos of 435 pediatric patients diagnosed with CP at Gillette Children's Specialty Healthcare. Some patients were recorded multiple times, but these recordings were conducted on different days. The video data were not publicly available due to restrictions on sharing patient health information. Therefore, the dataset the author used was the processed dataset to a de-identified form by Gillette Children's Specialty Healthcare. The processed data is CSV files recording (x, y) coordinates of 25 anatomical key points estimated by OpenPose [7] in chronological order. The author created the time-series data from CSV files shown in Fig. 3.1 and used them in the experiments. Note that, the values of the time-series data in Fig 3.1 were post-processed, thus they differ from the actual values in CSV files.



(a) x -coordinate of Ankle

(b) y -coordinate of Ankle

Figure 3.1: Time-series data used for the experiment

In the original videos, a patient's gait was recorded by a camera that was positioned 3-4 meters away from their right side. The resolution and frame rate of the video were

640×480 and 29.97 fps, respectively. Each patient’s video had about 500 frames, corresponding to around 16 seconds.

In addition, the dataset has some metrics (e.g., GDI [18], walking speed, and cadence) as ground-truth. These metrics were computed from optical motion capture data. Note that these motion capture data were collected at the same visit as the videos, though not simultaneously. In the experiments, the GDI was used as ground-truth to evaluate the proposed method.

Chapter 4

Method

4.1 Feature Engineering

Coordinates of key points (e.g., ankle and knee, etc.) and hand-engineered features were used as features. Table 4.1 shows all features used for the proposed method. The OpenPose estimates the position of 25 key points in the body. From all prediction of OpenPose, plural key points were selected as features, and time-series data was created of all frames. Since the position of key points was estimated in (x, y) coordinates, two time-series data were created for one key point. In addition, the author created hand-engineered features (e.g., the angle formed by the ankle, knee, hip, etc.) used in clinical gait quality assessment. The coordinate and angle of the elbow were added to features because the movement of the elbow, wrist, or shoulder might related to the gait variability. Each feature was created for each left and right side of the body as gait quality is assessed for each side.

Table 4.1: Description of key point used for proposed method

Features	
(x, y) -Coordinate of Ankle	(x, y) -Coordinate of Knee
(x, y) -Coordinate of Hip	(x, y) -Coordinate of Big Toe
(x, y) -Coordinate of Elbow	Angle Formed by Ankle, Knee, and Hip
Angled Formed by Big Toe, Ankle, and Knee	Angle Formed by Wrist, Elbow, and Shoulder
Distance of Left and Right Ankle	Distance of Big Toe and Ankle

4.2 Data Preprocessing

Fig. 4.1 shows how to create the experimental materials. In this approach, features shown in Tab 4.1 were extracted across 124 frames, and one data was created including $124 \text{ (time length)} \times 15 \text{ (number of key points)}$ data points. In general, the gait cycle varies among individuals. Therefore, even if the two sets of data have the same number of frames, the gait behavior included in each data is different. To mitigate this effect on the model, 124 frames were extracted to overlap while shifting the 31 frames. Thus, letting \mathbf{X} denote one input data containing 500 frames, segments were constructed as $\mathbf{X}[0 : 124], \mathbf{X}[31 : 155], \dots, \mathbf{X}[372 : 496]$. Also, training data can be increased.

In the data preprocessing, the author applied a one-dimensional unit-variance to smooth the skeleton trajectories and linear interpolation to fill the missing observations. In addition, (x, y) image-plane coordinates of each time series were normalized. In this process, the subtraction of each key point from the center hip position was calculated and then the subtraction value was divided by the Euclidean distance between the hip and the knee, respectively. This operation mitigates the effect of the body size difference among individuals on the proposed method.

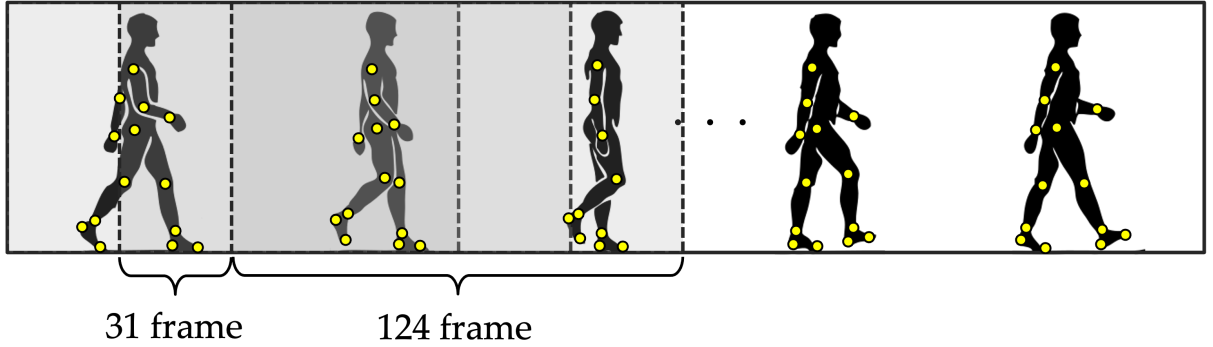


Figure 4.1: Illustration of creating the data

4.3 Outline of Proposed Method

To assess the gait quality of a patient, a method to capture the deviation of data from standard cases was required. This idea has also been often applied to unsupervised anomaly detection, and dimensionality reduction models (e.g., Auto-Encoder) were proposed [12,13,15,17,22]. The dimensionality reduction models are one of the deep learning models that learn the characteristics of the given data by reconstructing it. Their method followed the approach; the model was trained with only normal data and calculated a reconstruction capability as an anomaly score. This strategy was mainly based on the

assumption that the reconstruction capability of the trained model became lower when anomaly data were given. By taking their approach, the advantage where this method can be used in an unsupervised manner can be obtained. Since it is difficult to collect data and label in practical situations, the unsupervised method is useful. In this thesis, the author developed plural models and validated their effectiveness against the task. The following sections describe the details of each model.

4.4 Auto-Encoder Model

Auto-Encoder (AE) is often used for feature extraction or anomaly detection. Fig. 4.2 shows the structure of the AE. The AE mainly consists of the encoder, latent space, and decoder. The given data is converted to the lower dimensional latent space by the encoder, and the converted latent vector(s) is recovered to the original data by the decoder. In the training phase, the AE is trained to reconstruct the data close to the training data. In other words, the trained AE with the normal data outputs significant reconstruction error when the abnormal data is given. In the experiment, the absolute value between input and output data was calculated as the reconstruction error. This value was used as the gait score.

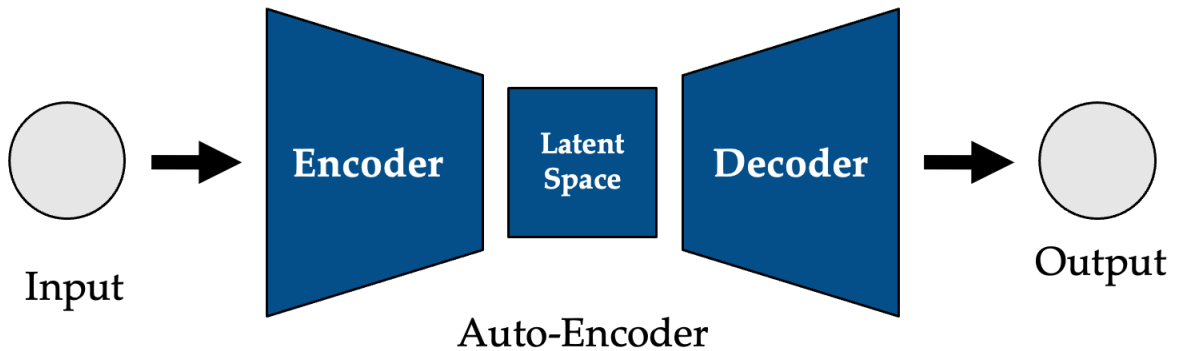


Figure 4.2: Network of the Auto-Encoder

4.5 Memory-augmented Auto-Encoder Model

The AE is often used for any task since its mechanism is simple. However, the AE can sometimes reconstruct abnormal data. In other words, the deviation of the abnormal data from normal data sometimes cannot be captured. Dong *et al.* focused on this point and proposed the Memory-augmented Auto-Encoder (MemAE) to overcome this issue [23]. In this thesis, the author also used MemAE and validated its effectiveness when estimating the deviation in patients' gait.

4.5.1 Mechanism of Memory-augmented Auto-Encoder

The MemAE is the model to reconstruct the input close to the training data. Fig. 4.3 shows the outline of Memory-augmented Auto-Encoder. The memory module is adopted to the latent space in the MemAE. As described in 4.4, the encoder converts the given data to the lower dimensional latent space (i.e., feature representation $\mathbf{z} \in \mathbb{Z}$). In the MemAE, the feature representation \mathbf{z} is regarded as a query since it is used to retrieve memory items similar to itself. Therefore, the encoder also is regarded as a query generator. The decoder is used to reconstruct the given data from the feature representation $\hat{\mathbf{z}} \in \hat{\mathbb{Z}}$. Therefore, given an input \mathbf{x} , the encoder and decoder are trained as follows:

$$\mathbf{z} = f_e(\mathbf{x}; \theta_e), \quad (4.1)$$

$$\hat{\mathbf{x}} = f_d(\hat{\mathbf{z}}; \theta_d) \quad (4.2)$$

where $\hat{\mathbf{x}}$, θ_e , and θ_d denote the reconstructed input and the parameters of encoder $f_e(\cdot)$ and decoder $f_d(\cdot)$, respectively. In other words, the decoder in MemAE takes $\hat{\mathbf{z}}$, rather than receiving the output \mathbf{z} from the encoder (i.e., $\hat{\mathbf{z}} \neq \mathbf{z}$) unlike the decoder in the AE.

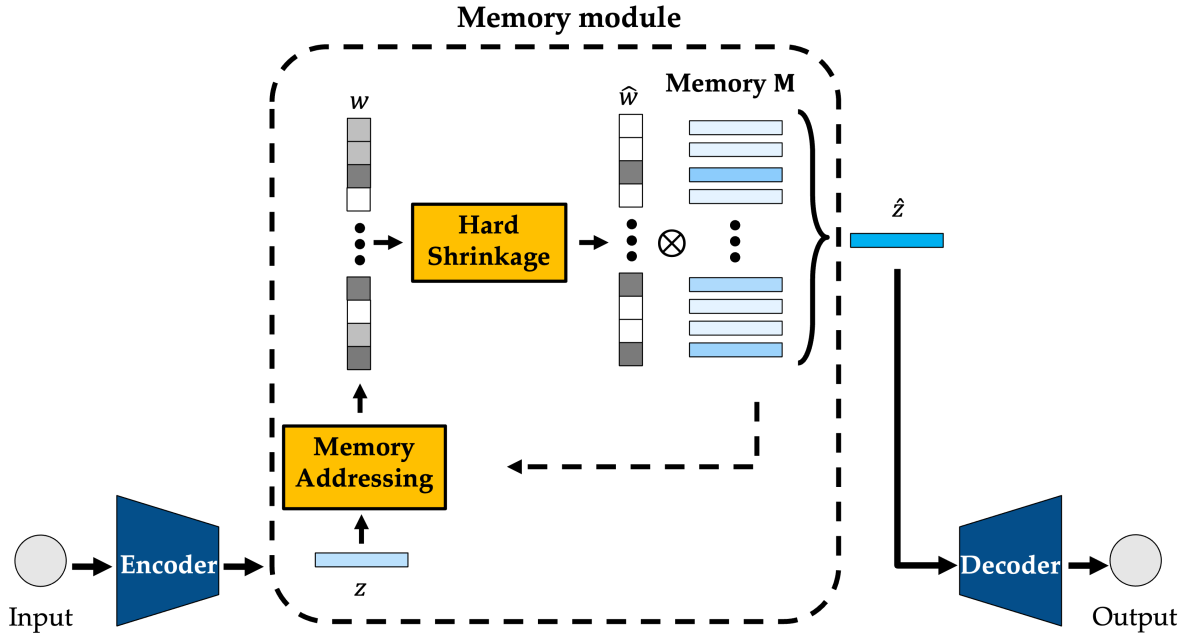


Figure 4.3: Network of the Memory-augmented Auto-Encoder

The memory module mainly consists of two parts: the memory to record and the attention-based addressing operator. The memory records the informative feature patterns to represent the training data. The memory is designed as a matrix $\mathbf{M} \in \mathbb{R}^{S \times D}$ containing S real-valued vectors of fixed dimension D . The hyper-parameter S can be

seen as the maximum capacity of the memory, with larger capacities preserving more patterns. However, the excessively large capacity makes the training of the MemAE difficult since the number of feature representations required for the data should be small. The author discussed S through the experiments, as a result, $S = 500$ was a desirable result. The D is the dimension of the feature representation \mathbf{z} (i.e., $\mathbb{Z} = \mathbb{R}^D$) $\forall i \in [S]$ and \mathbf{m}_i denote the number of row and the i -th row vector of the matrix M , the \mathbf{m}_i is a memory item in the memory M . Note that $[S]$ is the set of integers from 1 to S .

The attention-based addressing operator calculates the relevant memory items in the memory. Given a query \mathbf{z} (i.e., latent variable), $\hat{\mathbf{z}}$ is calculated using a memory addressing vector $\mathbf{w} \in \mathbb{R}^{1 \times S}$ and the memory M as

$$\hat{\mathbf{z}} = \mathbf{w}M = \sum_{n=1}^S w_n \mathbf{m}_n \quad (4.3)$$

where \mathbf{w} is the vector with positive value entry elements that sum to one, and it is calculated based on the similarity between a query \mathbf{z} and memory items \mathbf{m}_i . The w_i means the i -th entry elements of \mathbf{w} . In other words, $\hat{\mathbf{z}}$ is obtained as the weighted sum of the memory items \mathbf{m}_i . In particular, w is computed via a softmax operation so that the sum of w_i is 1:

$$w_i = \frac{\exp(d(\mathbf{z}, \mathbf{m}_i))}{\sum_{j=1}^S \exp(d(\mathbf{z}, \mathbf{m}_j))} \quad (4.4)$$

where $d(\mathbf{z}, \mathbf{m}_i)$ is a similarity measurement. In this thesis, $d(\mathbf{z}, \mathbf{m}_i)$ was defined as

$$d(\mathbf{z}, \mathbf{m}_i) = \sqrt{\sum_{j=1}^D (z_j - m_{ij})^2} \quad (4.5)$$

where z_j and m_{ij} are j -th entry element of \mathbf{z} and \mathbf{m}_i , respectively. Although the original paper proposed cosine similarity as a similarity measurement, Euclidean distance showed a better performance for the dataset.

However, if many memory items are retrieved from the memory, the expressive power of the latent space becomes relatively high due to a complex combination of the memory items. This leads that the MemAE can reconstruct the abnormal data. To mitigate this issue, the hard shrinkage operation for w is applied as follows.

$$\hat{w}_i = h(w_i; \lambda) = \begin{cases} w_i & (\text{if } w_i > \lambda) \\ 0 & (\text{otherwise}) \end{cases} \quad (4.6)$$

where \hat{w}_i is the i -th entry elements after the hard shrinkage operation and λ is the shrinkage threshold. Since w_i can be used only when it is greater than λ , Eq. (4.6) makes \mathbf{w} a sparse matrix. However, since Eq. (4.6) is the discontinuous function, implementing

the back-propagation for training the MemAE is difficult. Therefore, the hard shrinkage operation is rewritten using the continuous Rectified Linear Unit (ReLU) activation function defined as

$$\hat{w}_i = \frac{\max(w_i - \lambda, 0) \cdot w_i}{|w_i - \lambda| + \varepsilon}. \quad (4.7)$$

In the above equation, $\max(w_i - \lambda, 0)$ means ReLU activation. The ε is a very small positive scalar, and it was set as 1.0×10^{-12} . As hyper-parameters λ increases, the \mathbf{w} becomes a sparser matrix. However, if λ is excessively large, the MemAE does not learn well. In this thesis, a better result was obtained when λ is 0.0025. After applying the hard shrinkage operation shown in Eq. (4.7), $\hat{\mathbf{w}}$ is normalized by letting $\hat{w}_i = \hat{w}_i / \|\hat{\mathbf{w}}\|_1, \forall i$, where $\hat{\mathbf{w}}$ is the memory addressing vector \mathbf{w} after applying the hard shrinkage operation.

By applying these above operations, the MemAE is encouraged to record the informative feature representations and reconstruct the data close to training data.

4.5.2 Objective Function for MemAE

Given a dataset $\{\mathbf{x}^n\}_{n=1}^N$ containing N data, let $\hat{\mathbf{x}}^n$ denote the reconstructed data corresponding to each input data. AE is trained to minimize the reconstruction error on each data:

$$R(\mathbf{x}^t, \hat{\mathbf{x}}^t) = \|\mathbf{x}^t - \hat{\mathbf{x}}^t\|_2^2, \quad (4.8)$$

where the ℓ_2 -norm is used to measure the reconstruction error. In addition to the reconstruction error, a sparsity regularizer on $\hat{\mathbf{w}}$ is added in the MemAE. As discussed above, the sparsity of $\hat{\mathbf{w}}$ leads that latent space has more informative feature representations of training data. To further promote this, a sparsity regularizer on $\hat{\mathbf{w}}$ was minimized during training. Note that let $\hat{\mathbf{w}}^t$ denote the memory addressing vector for each data \mathbf{x}^t , the entropy of $\hat{\mathbf{w}}^t$ is used as a regularizer since $\hat{\mathbf{w}}$ is positive value and $\|\hat{\mathbf{w}}\|_1 = 1$:

$$E(\hat{\mathbf{w}}^t) = \sum_{i=1}^N -\hat{w}_i \cdot \log(\hat{w}_i). \quad (4.9)$$

To summarize, the objective function for training the MemAE is the objective function of AE plus a regularizer on $\hat{\mathbf{w}}$ like

$$L(\theta_e, \theta_d, \mathbf{M}) = \frac{1}{T} \sum_{t=1}^T (R(\mathbf{x}^t, \mathbf{y}^t) + \alpha E(\hat{\mathbf{w}}^t)), \quad (4.10)$$

where the hyper-parameter α means the weight of the regularizer for the objective function, and it was set to 0.0002 in this thesis. Also, the memory \mathbf{M} is updated to optimize the objective function using backpropagation. In the backward pass, only the gradients for the memory items with non-zero addressing weights w_i can be updated.

4.6 Variational Auto-Encoder Model

4.6.1 Mechanism of Variational Auto-Encoder

Variational Auto-Encoder (VAE) is for modeling the data generation process using neural networks [24]. Let us consider a dataset $\mathbf{X} = \{\mathbf{x}^{(i)}\}_{i=1}^N$, and assume that data \mathbf{x} is generated from unobserved latent variables \mathbf{z} , the VAE takes the form of the AE shown in Fig. 4.4 to estimate the generation process. The encoder infers the approximate posterior $q_\varphi(\mathbf{z}|\mathbf{x})$ from the input \mathbf{x} , and the decoder generates the reconstruction data $\hat{\mathbf{x}}$ of the input from the approximate posterior $q_\varphi(\mathbf{z}|\mathbf{x})$. The encoder and decoder is modeled by the neural network. The objective function of VAE is evidence lower bound (i.e., ELBO) of a dataset. The DNN parameters φ and θ are optimized to maximize the ELBO as follows:

$$L(\theta, \varphi; \mathbf{x}^{(i)}) = \mathbb{E}_{q_\varphi(\mathbf{z}|\mathbf{x}^{(i)})} [\log p_\theta(\mathbf{x}^{(i)}|\mathbf{z})] - D_{KL}(q_\varphi(\mathbf{z}|\mathbf{x}^{(i)}) \parallel p(\mathbf{z})) \quad (4.11)$$

where $p_\theta(\mathbf{x}|\mathbf{z})$ and $p(\mathbf{z})$ denote the likelihood of the data \mathbf{x} given the latent variable \mathbf{z} and the prior distribution of the latent variable \mathbf{z} , respectively. In general, the prior $p(\mathbf{z})$ is defined as a Gaussian distribution. Thus, the encoder estimates the mean and standard deviation of the distribution. The first term of equation (4.11) is the expectation, and the second term is the KL-divergence between the approximate posterior $q_\varphi(\mathbf{z}|\mathbf{x})$ and the prior $p(\mathbf{z})$. By maximizing the ELBO, the reconstruction data $\hat{\mathbf{x}}$ and approximate posterior $q_\varphi(\mathbf{z}|\mathbf{x})$ will be close to the input \mathbf{x} and prior $p(\mathbf{z})$, respectively. In other words, the model will become the generation model. In practice, since determining the expectation (integration calculations) in equation (4.11) is intractable, the first term of equation (4.11) is calculated via the Monte Carlo methods as follows:

$$\mathbb{E}_{q_\varphi(\mathbf{z}|\mathbf{x}^{(i)})} [\log p_\theta(\mathbf{x}^{(i)}|\mathbf{z})] \cong \frac{1}{L} \sum_{l=1}^L \log p_\theta(\mathbf{x}^{(i)}|\mathbf{z}^{(l)}), \mathbf{z}^{(l)} \sim q_\varphi(\mathbf{z}|\mathbf{x}) \quad (4.12)$$

However, since the equation 4.12 makes the computational graph of VAE disconnected, the random variable \mathbf{z} is computed via the reparameterization trick as follows:

$$\mathbf{z} = \mu + \sigma \odot \varepsilon, \varepsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \quad (4.13)$$

ε are random numbers given from a standard normal distribution.

4.6.2 Probabilistic Reconstruction Error

Most of the methods using the VAE for anomaly detection use the reconstruction error as anomaly score (i.e., gait score in the research). However, it is difficult to reconstruct the value itself of input data that is high intra-class variation and noisy. Also,

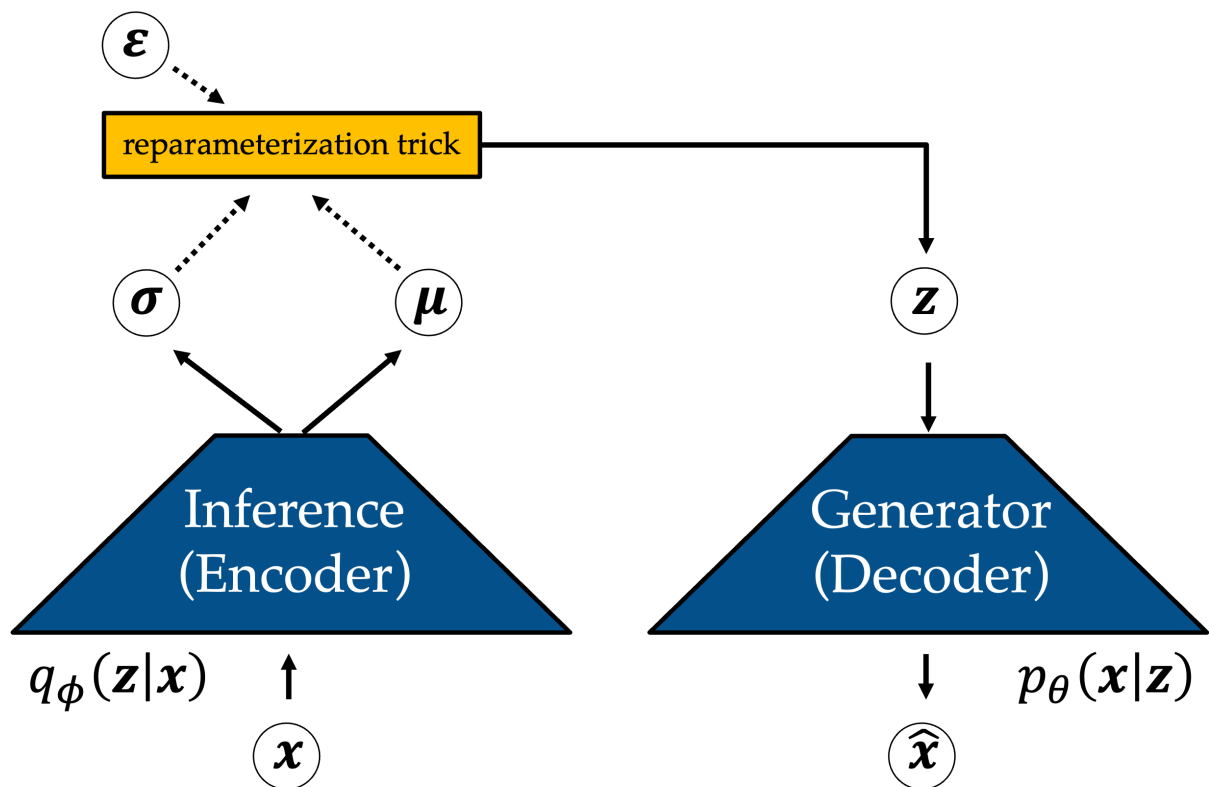


Figure 4.4: Network of the Variational Auto-Encoder

the reconstruction error cannot account for the difference in the scale of values for each feature. In other words, the feature with a small value has a smaller impact on the gait score although all key points should be evaluated fairly. To alleviate this problem, a stochastic framework is applied to the gait score [25]. In this method, the decoder defines a Gaussian distribution in the original input variable space, estimating the mean and $\hat{\mu}$ variance $\hat{\sigma}$ parameters for each data point. In the training, the DNN parameters of the decoder are updated to maximize the likelihood of the training data given the predicted distribution. In the test, the probability that the test data exists in the predicted distribution is calculated. In this time, the lower probability is obtained when the abnormal data. Since probabilistic metrics is used, all features are evaluated on an equal scale. To summarize, the model is trained to predict the area where the training data is likely to exist rather than the value itself, and its probability is used as the gait score.

In addition to the above problem, most of the methods using the VAE do not consider the variability of the generation process. In the Monte Carlo method shown in equation (4.12), $L = 1$ is defined in general. Therefore, the decoder disregards the variability of the distribution although the latent variable space is a distribution. In this method, L number of latent variables \mathbf{z} are sampled from the approximate posterior, and the average of all probabilities is used as an anomaly score (called reconstruction probability). The overall flow of the VAE applying the above operation is shown in Fig. 4.5.

4.6.3 Regularizer for Objective Function

It is assumed that capturing the abnormality in the VAE would be difficult when the estimated variance of the distribution is higher. Fig. 4.6 shows the difference (i.e., gait abnormality) in gait scores between normal and abnormal data when higher and lower standard deviation is estimated by the VAE. Trained VAE is expected to predict a distribution for abnormal data similar to that of normal data. Therefore, suppose a blue distribution with the same standard deviation $\sigma_{y^{(i)}}$ sigma and mean $\mu_{y^{(i)}}$ for both normal data $x^{(i)}$ and abnormal data $\hat{x}^{(i)}$ shown in Fig. 4.6. As described in 4.6.2, given the normal data $x^{(i)}$, a higher probability is obtained. On the other hand, a lower probability is obtained given the abnormal data $\hat{x}^{(i)}$. This probability is the gait score in the proposed method. However, the probability (i.e., gait score) could be obtained despite the abnormal data when a higher standard deviation is obtained, as shown in Fig. 4.6 (a). This would make the distinction between normal and abnormal data ambiguous. It is required to suppress the increase in the standard deviation, as shown in Fig. 4.6 (b).

Therefore, the regularizer $\alpha \sum_{i=1}^K \sigma_{y^{(i)}}$ was added to the original objective function of the VAE. α and K are the weight of the regularizer and the number of data points. It is expected that a lower standard deviation could be obtained by minimizing the regularizer.

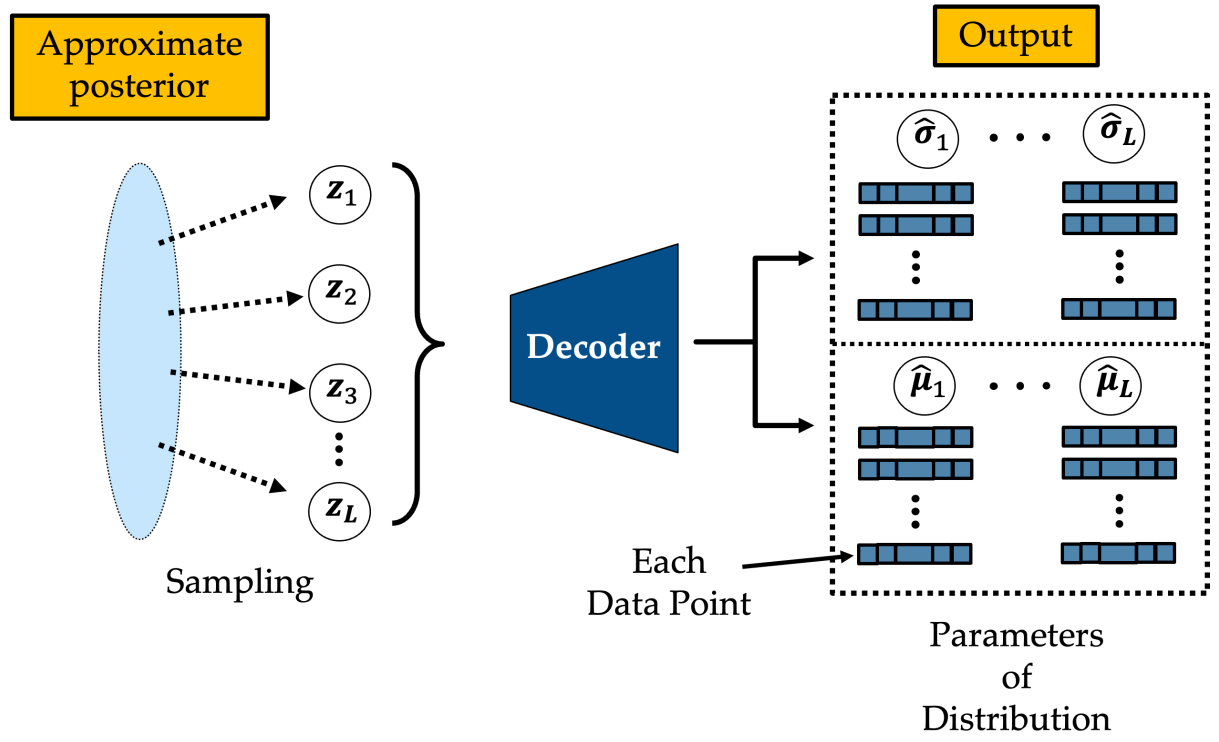
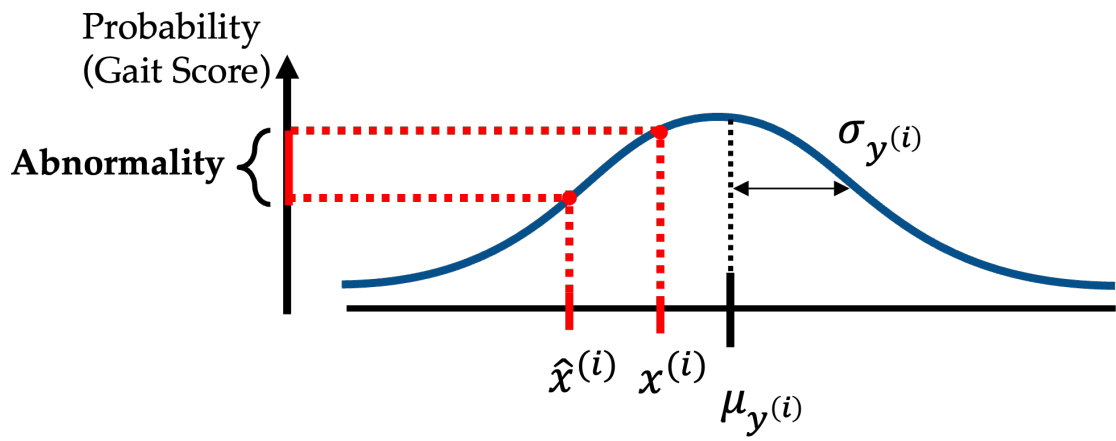
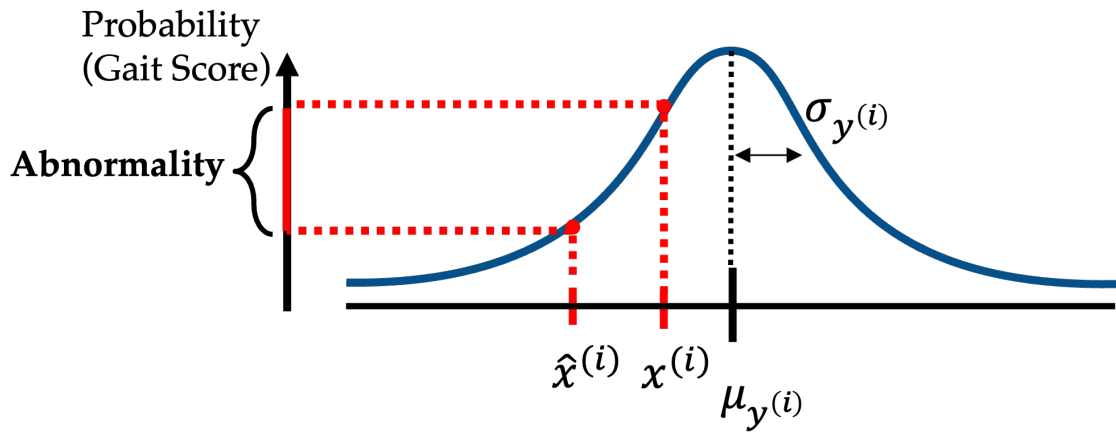


Figure 4.5: Flow of the VAE in the proposed method



(a) Higher Standard Deviation



(b) Lower Standard Deviation

Figure 4.6: Comparison between the case of higher and lower standard deviation

Chapter 5

Experimental Results and Discussion

5.1 Experimental Set-Up

In this experiment, the data whose GDI was above 90 was defined as normal data and divided into the training, validation, and test sets. Also, the data whose GDI was less than 90 were defined as abnormal data and added to the test set. Table 5.1 shows the number of patients, videos, and data for each dataset. In the training phase, each model was trained for 100 epochs using the training set. In the test phase, the model of the epoch where the loss of the validation set is minimal was used. Subsequently, the test set was presented to the model and the gait scores were obtained from the model.

Table 5.1: Number of dataset used for experiment

Dataset	Role	Patients	Video	Data
Normal ($GDI \geq 90$)	Train	226	296	4589
	Validation	30	36	523
	Test	25	34	526
Abnormal ($GDI < 90$)	Test	349	503	7535

The correlation coefficient between the obtained gait score and the GDI was calculated as the evaluation metric. If the calculated correlation coefficient is high, it means the obtained gait score from the proposed method can be used in the practical situation instead of the GDI. The author developed a plural model described in 4 and compared the results.

5.2 Implementation of Model

The key building block of each model was implemented by the 1-D convolutional neural network. The network architecture was designed based on a structure of the U-net shown in Fig. 5.1. Note that the proposed method did not use the skip connection. This is because the model was used to compress and decompress the data. The encoder consisted of 6 convolution layers followed by Batch Normalization and rectified linear unit (ReLU). In addition, every 2 layers were followed by an average pooling layer. Each convolutional layer had 256 filters and a filter length of 4. The final layer of the encoder is the flattening block (Flatten) which changes the two-dimensional vector outputted from a convolutional block to a one-dimensional latent vector (50-dimensional). The decoder structure was an inverse of the encoder, i.e., built from flatten, transposed convolutional, convolutional, batch normalization, and ReLU layer. These parameters were determined empirically.

For optimizer, Adam was used with a learning rate of 0.001 as the optimizer. all models were implemented using PyTorch (ver. 1.13.0).

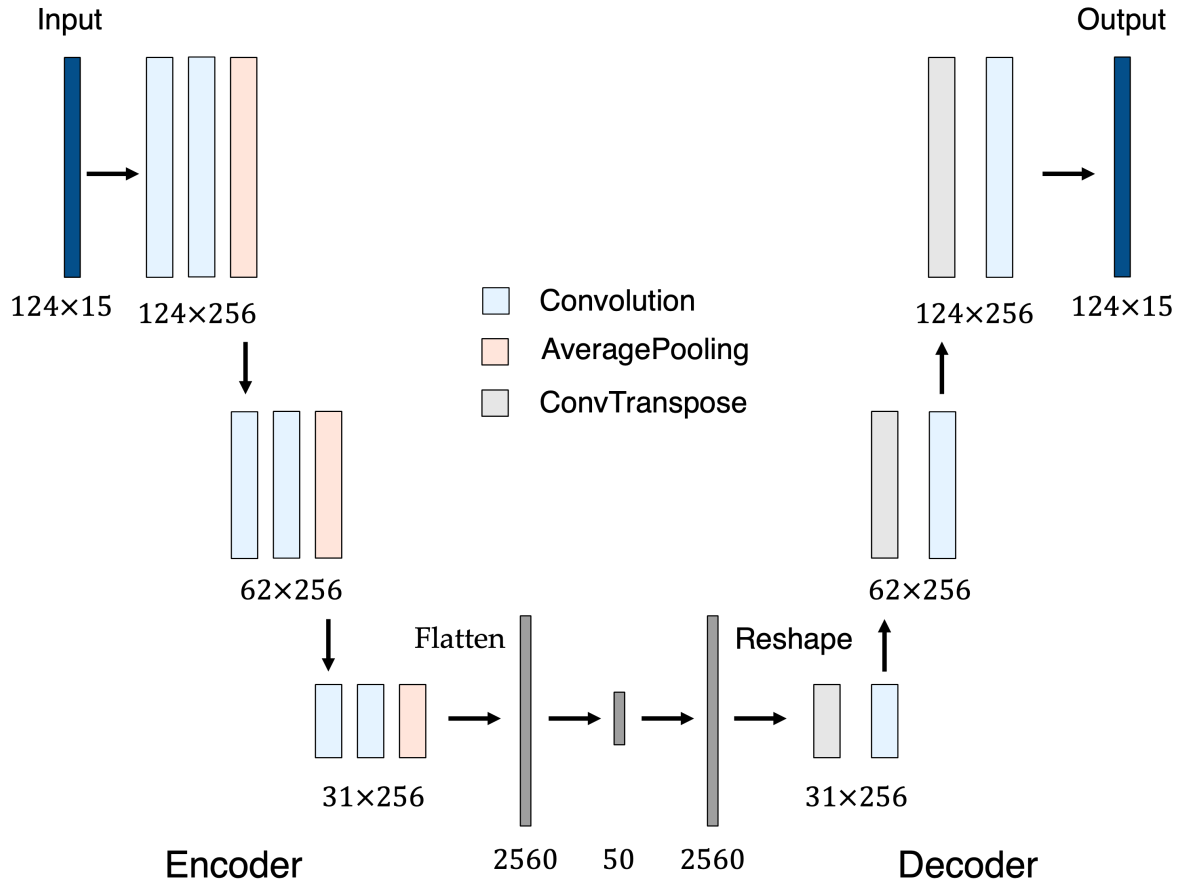


Figure 5.1: Structure of the model

5.3 Result

Table 5.2 shows the results of the experiments. The correlation coefficient of the method using the reconstruction error becomes a negative value. In this study, the relationship between the reconstruction error and the GDI is in the opposite direction. In other words, a data sample with a small reconstruction error is regarded as normal in the proposed method while a large value is a normal case in the GDI. The absolute value of the correlation coefficient is essential to compare methods.

Table 5.2: Comparison of Correlation Coefficient among Model

Model	Correlation coefficient
AE	-0.469
MemAE	-0.409
VAE (Reconstruction Error)	-0.449
VAE (Reconstruction Probability)	0.514
VAE (Reconstruction Probability + Regularizer)	0.533

The method using VAE (reconstruction Probability + regularizer) outperformed other methods. Fig. 5.2 shows the relationship between the gait score (Reconstruction Probability) and the GDI. In Fig. 5.2, red and blue dots denote the training and test data samples, respectively. When focusing on the test dataset (blue dots), it is observed that the normal data with $GDI > 90$ exhibit high gait scores. On the other hand, the number of data with a lower gait score increases as the GDI decreases. This relationship suggests that the proposed method can capture the gait deviation from the data.

Fig. 5.3 shows the example of the graph the author plotted the input data and reconstructed data (output data) of the VAE(reconstruction Probability + regularizer). Note that values within the reconstructed data differ from the actual values since they were sampled from Gaussian distributions predicted by the VAE. From the top, the figure shows the time series of hip x -coordinates, the angle formed by the big toe, ankle, and knee, and the angle formed by the ankle, knee, and hip. As shown in the figure, the abnormal data could not be reconstructed well compared with the normal data. This gap between input and output data was used as the gait abnormality in the proposed method.

Fig. 5.4 shows data within abnormal data, each having different GDI. In the case of lower GDI, the reconstructed data is closer to the input data compared to the case of higher GDI. This supposed that the proposed method could also capture the abnormality of the gait.

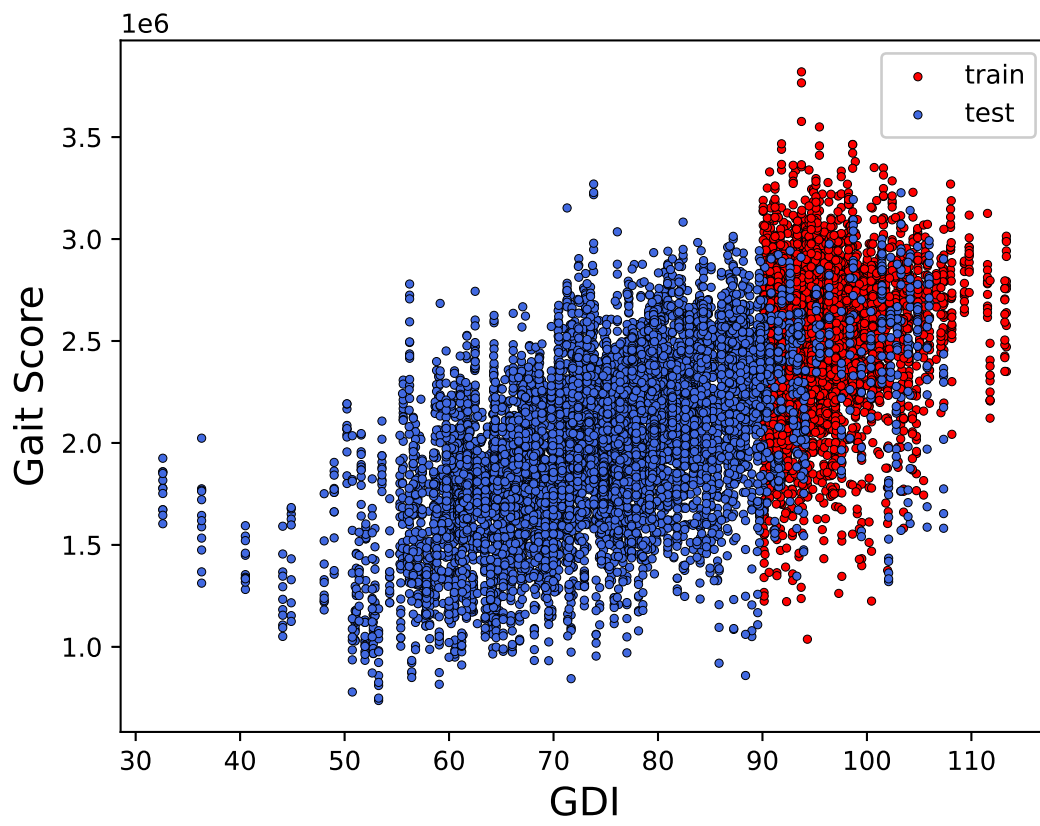
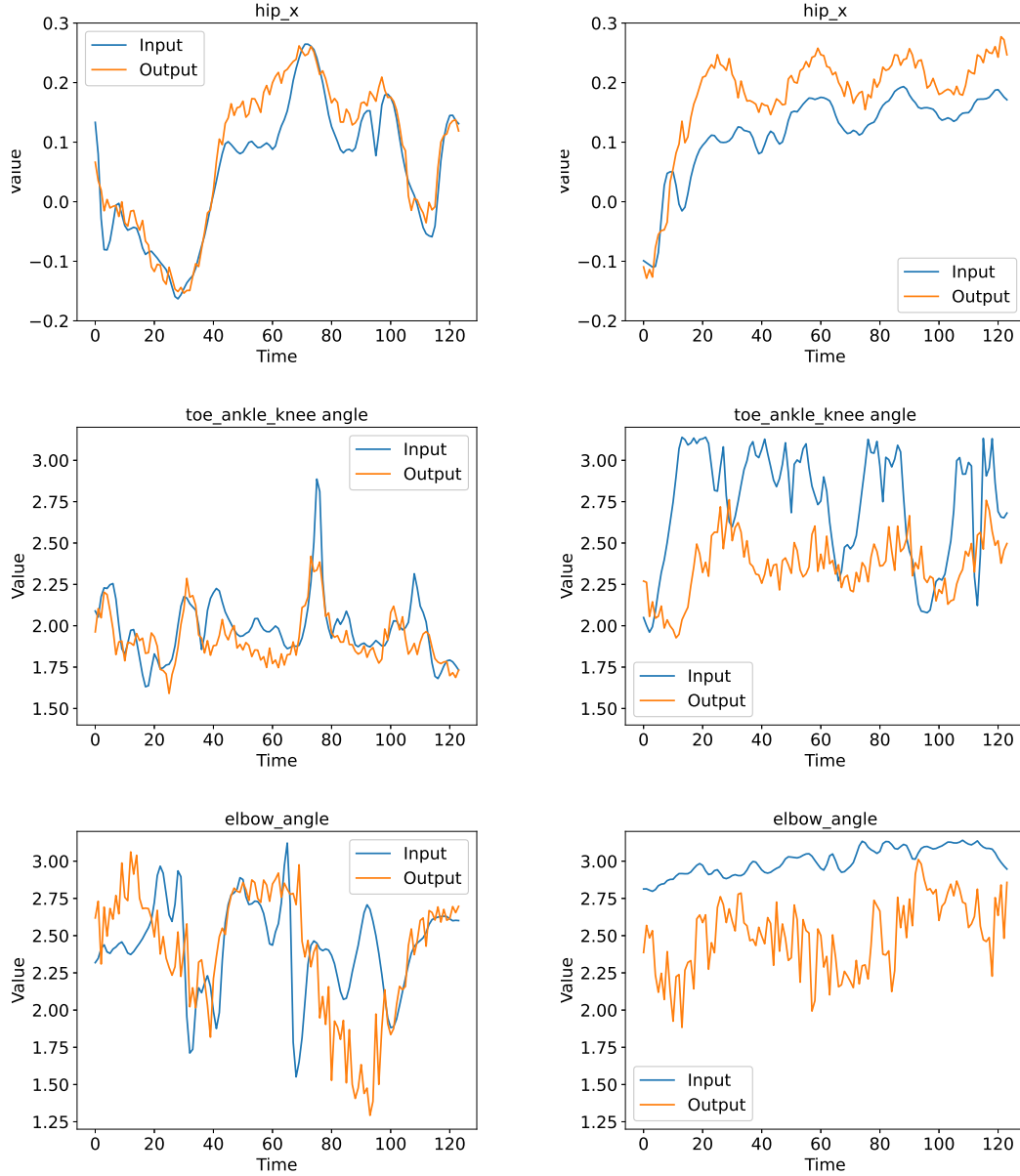


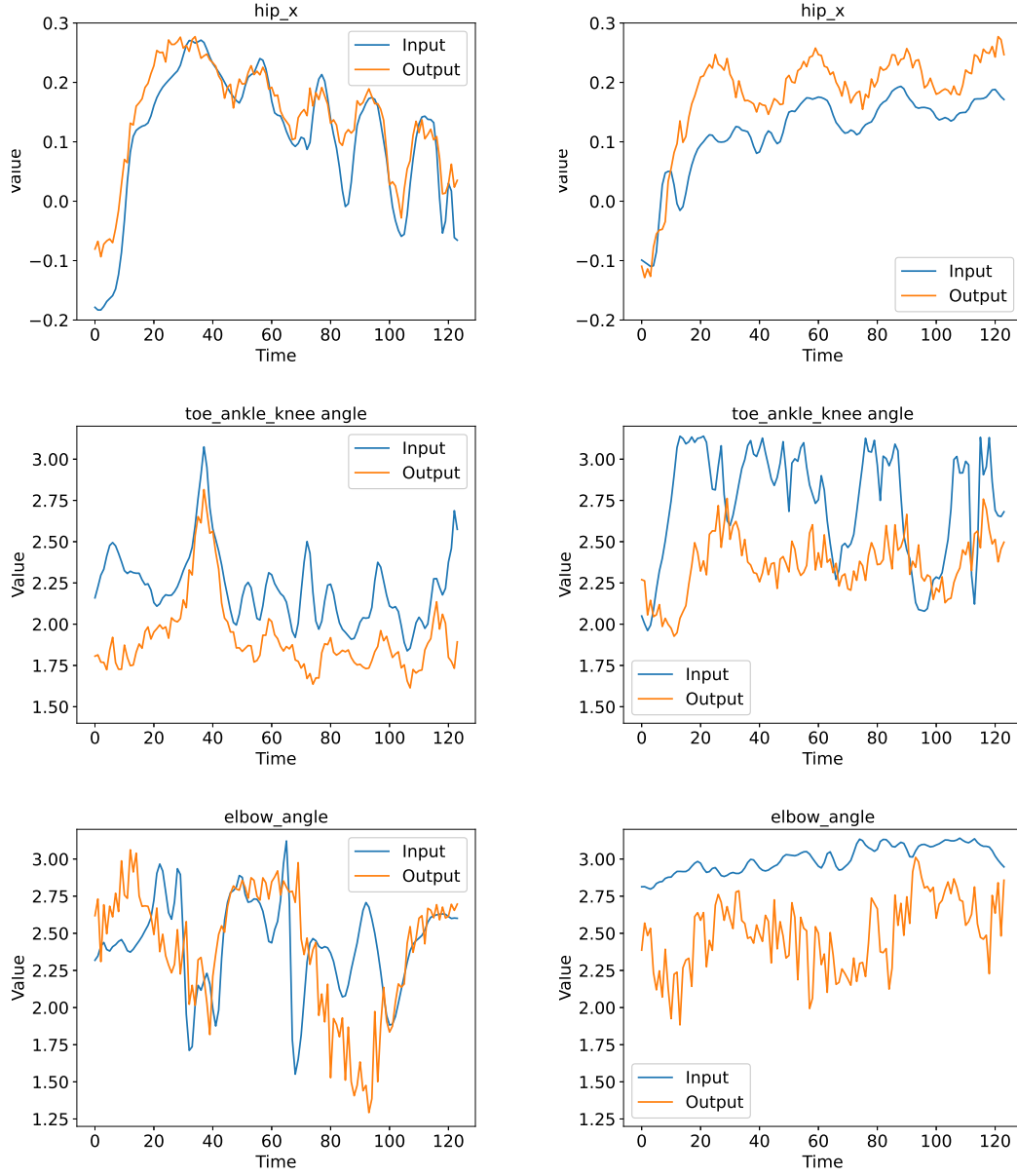
Figure 5.2: Gait score (reconstruction probability) vs. GDI



(a) Normal (GDI= 93.3)

(b) Abnormal (GDI= 44.9)

Figure 5.3: Examples of the input and reconstructed data in the VAE



(a) Abnormal (GDI= 71.7)

(b) Abnormal (GDI= 44.9)

Figure 5.4: Examples of the input and reconstructed data within the abnormal data

The result of the VAE (Reconstruction Probability) was better than that of the VAE (Reconstruction Error). That's why the VAE using reconstruction error could reconstruct well even when the input value was the outlier. Fig. 5.5 shows the comparison of the abnormal data reconstructed by VAE (Reconstruction Probability + Regularizer) and VAE (Reconstruction Error). The VAE using reconstruction error could reconstruct the abnormal data well. The VAE using reconstructed error could reconstruct even outlier input values as it predicted the input values themselves. On the other hand, the VAE (Reconstruction Probability) could not reconstruct abnormal data well as it predicted the area where the normal data is likely to exist. Therefore, it is assumed that the VAE (Reconstruction Probability) could capture the gait abnormality.

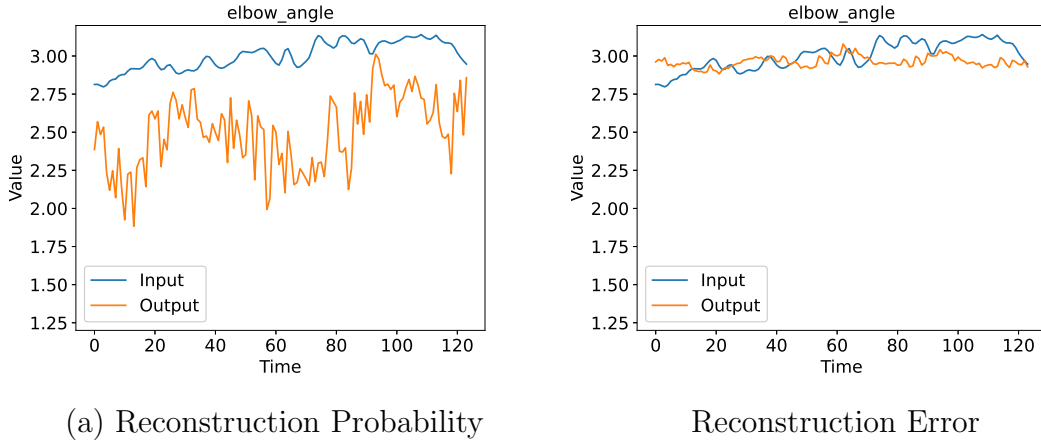


Figure 5.5: Comparison between the reconstruction probability and the reconstruction error

The effect of the regularizer can be recognized by comparing the VAE (Reconstruction Probability) and the VAE (Reconstruction Probability + Regularizer). Tab 5.3 shows the result when varying the value of α . The correlation coefficient increased as α approaches 1. As α becomes greater than 1, the correlation coefficient decreases. When α becomes larger, the standard deviation would be smaller, i.e., the VAE predicts the value itself rather than the distribution. The VAE lost the characteristic of reconstruction probability, and then such results were obtained.

The result of the AE was better performance than the one of the VAE (Reconstruction Error). In general, the VAE estimates obscure outputs since it has a distribution in the latent space. Therefore, it is assumed that the VAE reconstructed data obscurely, regardless of whether it was normal or abnormal. This leads that the model could not capture the gait deviation.

The result of the MemAE was the worst in the experiments. However, the MemAE outperformed the AE and VAE in the preliminary experiments shown in Tab 5.4. In

Table 5.3: Difference of Results with Value of α

Condition	Correlation Coefficient
no Regularizer	0.514
$\alpha = 0.1$	0.519
$\alpha = 1$	0.533
$\alpha = 2$	0.527
$\alpha = 3$	0.510

the preliminary experiments, the parameter size of the latent space, and decoder was small shown in Fig.5.6. In general, a larger parameter size in the reconstruction process improves the reconstruction capability, i.e., the model could reconstruct the abnormal data well. Therefore, this leads to the loss of the advantage that the MemAE reconstructs the data close to the training data in the experiments. That was why the result in Tab 5.4 was obtained.

Model	Correlation coefficient
AE	-0.375
VAE	-0.406
MemAE	-0.432

Table 5.4: Result of Preliminary Experiments

Moreover, the correlation coefficients between the reconstruction error of each feature in AE and GDI were calculated. Tab 5.5 shows the result. As shown in Tab 5.5, the largest absolute value of the correlation coefficient was the feature of the angle formed by the ankle, knee, and hip. On the other hand, the smallest absolute value was the feature of the angle formed by the wrist, elbow, and shoulder. This result is acceptable because GDI is calculated based on the information about the joint angle in the lower body. In addition, the same experiment using features except for the angle formed by the wrist, elbow and shoulder was performed. However, the correlation coefficient was reduced. This result suggests that even features that have weak relevance to the gait quality estimation on their own would contribute to improving the overall estimation accuracy.

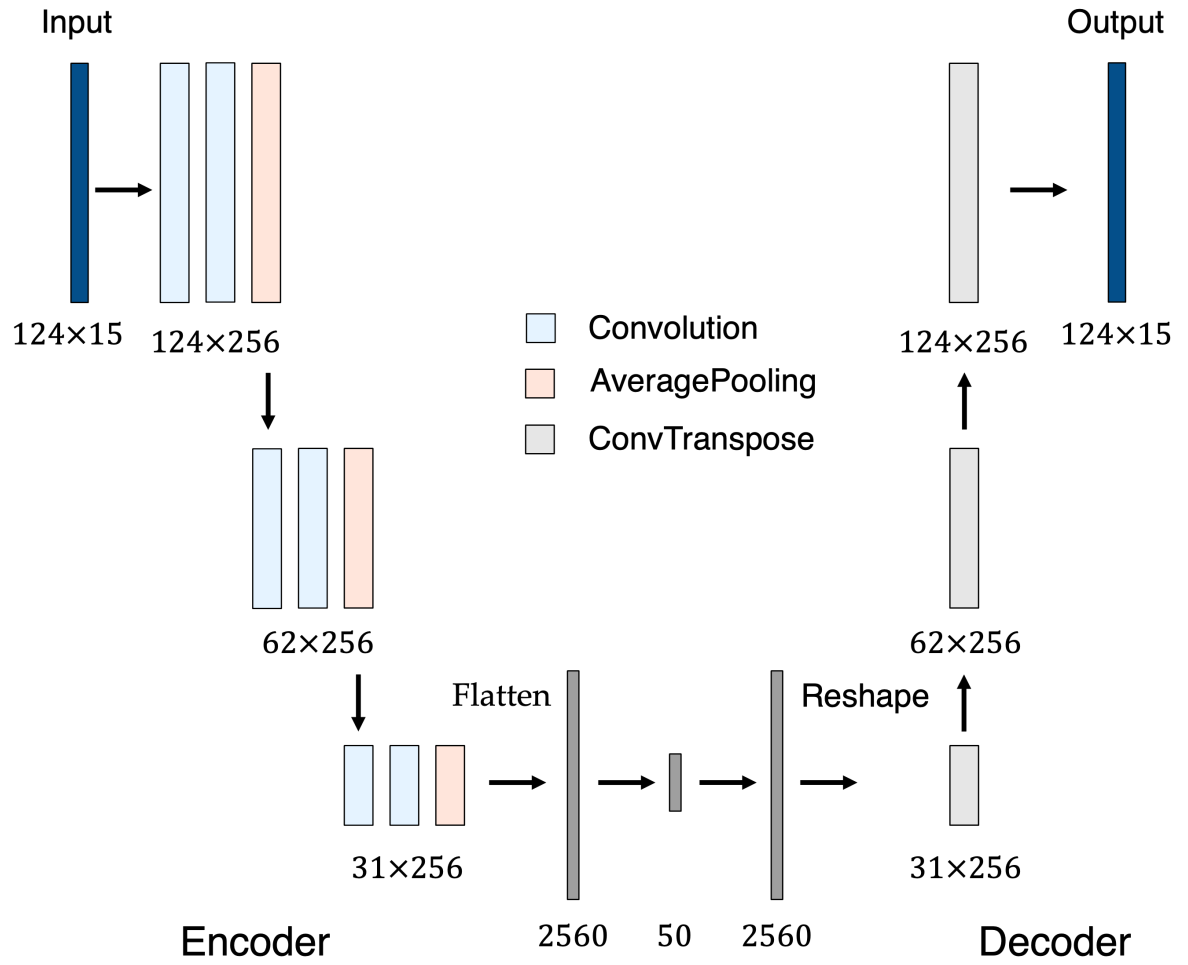


Figure 5.6: Structure of the model in the preliminary experiment

Feature	Correlation Coefficient
x -Coordinate of Ankle	-0.288
y -Coordinate of Ankle	-0.366
x -Coordinate of Knee	-0.401
y -Coordinate of Knee	-0.299
x -Coordinate of Hip	-0.230
y -Coordinate of Hip	-0.268
x -Coordinate of Big Toe	-0.328
y -Coordinate of Big Toe	-0.309
x -Coordinate of Elbow	-0.272
y -Coordinate of Elbow	-0.206
Angle Formed by Ankle, Knee, and Hip	-0.443
Angled Formed by Big Toe, Ankle, and Knee	-0.385
Angle Formed by Wrist, Elbow, and Shoulder	-0.132
Distance of Left and Right Ankle	-0.246
Distance of Left and Right Ankle	-0.279

Table 5.5: Caption

Chapter 6

Conclusion

6.1 Conclusion

This study aimed to establish a method to estimate patients' gait quality with only a standard camera in the pediatric Cerebral Palsy cohort. In this thesis, the authors addressed the challenge of estimating the the gait quality of CP patients using unsupervised deep learning-based models. In the experiment, the correlation between the gait score estimated by the model and the pathological gait index for CP was confirmed. The proposed method could be used in the practical situation of the gait assessment for CP. In particular, the method using the VAE and the reconstruction probability outperforms the compared methods.

The proposed method takes the approach using unsupervised deep learning-based models. All you need to train the model is the gait of an able-bodied person. This point is an advantage over other methods because it is difficult to collect the data and label patients. In addition, the proposed method could be used for other movement disorders or dementia.

6.2 Future Works

There is the possibility of practical application by improving the proposed method since the gait score moderately relevant to the gait pathological index was obtained. However, the obtained result is insufficient and suggests many issues left to use the proposed method in a practical situation. For example, the training data used in conducted experiments was the gait of patients with CP. The proposed method is supposed to use the gait of an able-bodied person for training the model. Although the gait of patients with relatively mild symptoms was used for the training data, the patients had symptoms at least somewhere in the body, which also varies for individuals. In other words, the training data contains features that should not be included in the normal gait, and thus models could not learn the features of the normal gait well. It is required to collect

the gait data of an able-bodied person in a similar environment to the data used in this thesis in the future. Also, many deep neural network architectures have been proposed to learn the characteristics of the data efficiently. The author has to investigate which architecture is appropriate for the proposed method.

Acknowledgement

First of all, I really would like to express my deepest gratitude to Prof. Hiroharu Kawanaka at the Graduate School of Engineering, Mie University who offered continuing support and constant encouragement.

I am also grateful to Balaji Iyer, Prof. Bruce J. Aronow, and Prof. V. B. Surya Prasath at Cincinnati Children's Hospital Medical Center, USA. They provided a lot of technical help and encouragement. Thanks to their help and encouragement, I was able to make good progress in my project. Prof. Surya gave me a lot of technical and language help in my research project and writing papers.

Also, the results shown here are in whole based upon the data published by Gillette Specialty Healthcare and downloaded from <https://simtk.org/projects/video-gaitlab>.

Reference

- [1] K. W. Krigger, “Cerebral palsy: An overview.” *American Family Physician*, vol. 73, no. 1, 2006.
- [2] J. R. Gage and T. F. Novacheck, “An update on the treatment of gait problems in cerebral palsy,” *Journal of Pediatric Orthopaedics B*, vol. 10, no. 4, pp. 265–274, 2001.
- [3] L. Carcreff, C. N. Gerber, A. Paraschiv-Ionescu, G. De Coulon, C. J. Newman, K. Aminian, and S. Armand, “Comparison of gait characteristics between clinical and daily life settings in children with cerebral palsy,” *Scientific reports*, vol. 10, no. 1, p. 2091, 2020.
- [4] A. Toshev and C. Szegedy, “DeepPose: Human pose estimation via deep neural networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1653–1660.
- [5] K. Sun, B. Xiao, D. Liu, and J. Wang, “Deep high-resolution representation learning for human pose estimation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 5693–5703.
- [6] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. G. Yong, J. Lee *et al.*, “Mediapipe: A framework for building perception pipelines,” *arXiv preprint arXiv:1906.08172*, 2019.
- [7] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, “Realtime multi-person 2d pose estimation using part affinity fields,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 7291–7299.
- [8] Y. Xu, J. Zhang, Q. Zhang, and D. Tao, “Vitpose: Simple vision transformer baselines for human pose estimation,” *Advances in Neural Information Processing Systems*, vol. 35, pp. 38 571–38 584, 2022.
- [9] K. Sato, Y. Nagashima, T. Mano, A. Iwata, and T. Toda, “Quantifying normal and parkinsonian gait features from home movies: Practical application of a deep learning-based 2d pose estimator,” *PloS one*, vol. 14, no. 11, p. e0223549, 2019.

- [10] K.-D. Ng, S. Mehdizadeh, A. Iaboni, A. Mansfield, A. Flint, and B. Taati, “Measuring gait variables using computer vision to assess mobility and fall risk in older adults with dementia,” *IEEE journal of translational engineering in health and medicine*, vol. 8, pp. 1–9, 2020.
- [11] L. Kidzinski, B. Yang, J. L. Hicks, A. Rajagopal, S. L. Delp, and M. H. Schwartz, “Deep neural networks enable quantitative movement analysis using single-camera videos,” *Nature Communications*, vol. 11, no. 1, p. 4054, 2020.
- [12] R. Morais, V. Le, T. Tran, B. Saha, M. Mansour, and S. Venkatesh, “Learning regularity in skeleton trajectories for anomaly detection in videos,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 11 996–12 004.
- [13] J. Sun, X. Wang, N. Xiong, and J. Shao, “Learning sparse representation with variational auto-encoder for anomaly detection,” *IEEE Access*, vol. 6, pp. 33 353–33 361, 2018.
- [14] W. Liu, W. Luo, D. Lian, and S. Gao, “Future frame prediction for anomaly detection—a new baseline,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 6536–6545.
- [15] H. Park, J. Noh, and B. Ham, “Learning memory-guided normality for anomaly detection,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 14 372–14 381.
- [16] T.-N. Nguyen and J. Meunier, “Anomaly detection in video sequence with appearance-motion correspondence,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.
- [17] M. Sakurada and T. Yairi, “Anomaly detection using autoencoders with nonlinear dimensionality reduction,” in *Proceedings of the MLSDA 2014 2nd workshop on machine learning for sensory data analysis*, 2014, pp. 4–11.
- [18] M. H. Schwartz and A. Rozumalski, “The gait deviation index: a new comprehensive index of gait pathology,” *Gait & posture*, vol. 28, no. 3, pp. 351–357, 2008.
- [19] C. Wan, L. Wang, and V. V. Phoha, “A survey on gait recognition,” *ACM Computing Surveys (CSUR)*, vol. 51, no. 5, pp. 1–35, 2018.
- [20] J. Han and B. Bhanu, “Individual recognition using gait energy image,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 2, pp. 316–322, 2005.

- [21] K. Bashir, T. Xiang, and S. Gong, “Gait recognition using gait entropy image,” 2009.
- [22] Y. Zhao, B. Deng, C. Shen, Y. Liu, H. Lu, and X.-S. Hua, “Spatio-temporal autoencoder for video anomaly detection,” in *Proceedings of the 25th ACM international conference on Multimedia*, 2017, pp. 1933–1941.
- [23] D. Gong, L. Liu, V. Le, B. Saha, M. R. Mansour, S. Venkatesh, and A. v. d. Hengel, “Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 1705–1714.
- [24] D. P. Kingma and M. Welling, “Auto-encoding variational bayes,” *arXiv preprint arXiv:1312.6114*, 2013.
- [25] J. An and S. Cho, “Variational autoencoder based anomaly detection using reconstruction probability,” *Special lecture on IE*, vol. 2, no. 1, pp. 1–18, 2015.

Publication List

Journal Papers

- (1) 鷺見銀河, 北島巧海, 川中普晴, Balaji Iyer, V. B. Surya Prasath, Bruce J. Aronow, "教師なし深層学習モデルを用いた脳性麻痺患者のための歩行機能評価手法の提案と基礎的検討 (A Study on Gait Quality Assessment for Cerebral Palsy Using Unsupervised Deep Learning Model)" 知能と情報 (Journal of Japan Society for Fuzzy Theory and Intelligent Informatics), Vol.36, No.1, 2023.

International Conferences

- (1) Ginga Sumi, Hiroharu Kawanaka, Balaji Iyer, V. B. Surya Prasath, and Bruce J. Aronow, "A Study on Deep Learning for Gait Assessment without Special Equipment", *13th International Symposium for Sustainability by Engineering at Mie University (Research Area C)*, 2023.
- (2) Ginga Sumi, Balaji Iyer, V. B. Surya Prasath, Hiroharu Kawanaka, and Bruce J. Aronow, "Gait Analysis for Cerebral Palsy Using Memory-Augmented Auto-Encoder Model", *Joint 6th International Conference on Imaging, Vision & Pattern Recognition (IVPR) & 11th International Conference on Informatics, Electronics & Vision (ICIEV)*, 2023. (Excellent Paper Award)
- (3) Ginga Sumi, Balaji Iyer, Kitajima Takumi, Hiroharu Kawanaka, V. B. Surya Prasath, and Bruce J. Aronow, "Gait Quality Assessment for Cerebral Palsy Using Variational Auto-Encoder", *24th International Symposium on Advanced Intelligent Systems, (ISIS2023)*, 2023.

Domestic Conferences

- (1) 鷺見銀河, 青木リュウジ, 川中普晴, V. B. Surya Prasath, Bruce J. Aronow, "深層学習を用いた脳性麻痺患者のための運動機能評価 (A Study on Movement Function Evaluation for Cerebral Palsy using Deep Learning)", 令和4年度電気・電子・情報関係学会東海支部連合大会, 2022.
- (2) 鷺見銀河, 北島巧海, 川中普晴, "深層学習を用いた歩行機能評価の特徴量に関する一検討 (A Study on Feature Variable of Gait Assessment using Deep Neural Network)", 令和5年度電気・電子・情報関係学会東海支部連合大会, 2023.

- (3) 鷺見銀河, 川中普晴, 北島巧海, V. B. Surya Prasath, Bruce J. Aronow, "Auto-Encoder を用いた脳性麻痺患者のための歩行機能評価手法に関する一検討 (A Study on Gait Analysis for Cerebral Palsy Using Auto-Encoder)", 第39回ファジィシステムシンポジウム (FSS2023), 2023. (優秀発表賞)
- (4) 鷺見銀河, 北島巧海, 川中普晴, Balaji Iyer, V. B. Surya Prasath, Bruce J. Aronow, "Variational Auto-Encoder を用いた歩行機能評価の再構成データの分布におけるばらつき抑制に関する一検討", 2023 年度日本生体医工学会東海支部大会, 2023.