systems

MDPI

Article

Advanced Dance Choreography System Using Bidirectional LSTMs

Hanha Yoo [1] and Yunsick Sung [2,*]

[1] Department of Computer Engineering, Graduate School, Dongguk University-Seoul, Seoul 04620, Republic of Korea
[2] Division of AI Software Convergence, Dongguk University-Seoul, Seoul 04620, Republic of Korea
* Correspondence: sung@dongguk.edu; Tel.: +82-2-2260-3338

Abstract: Recently, the craze of K-POP contents is promoting the development of Korea's cultural and artistic industries. In particular, with the development of various K-POP contents, including dance, as well as the popularity of K-POP online due to the non-face-to-face social phenomenon of the Coronavirus Disease 2019 (COVID-19) era, interest in Korean dance and song has increased. Research on dance Artificial Intelligent (AI), such as artificial intelligence in a virtual environment, deepfake AI that transforms dancers into other people, and creative choreography AI that creates new dances by combining dance and music, is being actively conducted. Recently, the dance creative craze that creates new choreography is in the spotlight. Creative choreography AI technology requires the motions of various dancers to prepare a dance cover. This process causes problems, such as expensive input source datasets and the cost of switching to the target source to be used in the model. There is a problem in that different motions between various dance genres must be considered when converting. To solve this problem, it is necessary to promote creative choreography systems in a new direction while saving costs by enabling creative choreography without the use of expensive motion capture devices and minimizing the manpower of dancers according to consideration of various genres. This paper proposes a system in a virtual environment for automatically generating continuous K-POP creative choreography by deriving postures and gestures based on bidirectional long-short term memory (Bi-LSTM). K-POP dance videos and dance videos are collected in advance as input. Considering a dance video for defining a posture, users who want a choreography, a 3D dance character in the source movie, a new choreography is performed with Bi-LSTM and applied. For learning, considering creativity and popularity at the same time, the next motion is evaluated and selected with probability. If the proposed method is used, the effort for dataset collection can be reduced, and it is possible to provide an intensive AI research environment that generates creative choreography from various existing online dance videos.

Keywords: dance; choreography; system; learning; Bi-LSTM; AI

Citation: Yoo, H.; Sung, Y. Advanced Dance Choreography System Using Bidirectional LSTMs. *Systems* **2023**, *11*, 175. https://doi.org/10.3390/systems11040175

Academic Editor: Nirit Yuviler-Gavish

Received: 7 February 2023
Revised: 14 March 2023
Accepted: 23 March 2023
Published: 28 March 2023

1. Introduction

Not only in Asia, but also in Europe, South America, etc., the K-POP craze is spreading all over the world. The popularity of K-POP extends to Korean culture, art, and food, and among them, the worldwide craze for K-POP is showing remarkable development. The appeal of K-POP is diverse, including content, dance, singing, and fashion [1]. In addition to this popularity, K-POP dance is expanding to various contents in conjunction with the Social Network System (SNS) and the non-face-to-face social phenomenon of the Corona era [2]. In addition to non-face-to-face dance classes, various content developments and demands, such as the metaverse where you can meet popular singers, are important. Considering all types of dance genres, extensive research requires the skills and knowledge involved in each dance. For example, ballet requires a lot of time and effort to express a movement. It is necessary to consider the specialized and advanced movements included in each dance, but it takes a lot of research time to generalize and express them. The specific reason for K-POP choreography lies in the popularity of K-POP dance. K-POP does

not represent a specific genre, but it is a popularized choreography that does not require specific dance skills and knowledge because all genres are integrated. It may be seen as composed of fancy and complex movements from a non-professional's point of view, but for the popularization of K-POP choreography, it is composed of relatively easy movements compared to other dances that can be followed by both men and women of all ages while experts look gorgeous from the planning stage. There is a peculiarity of losing.

Recently, dance artificial intelligence has been developing along with the dance craze, and furthermore, the need for creative artificial intelligence has been highlighted. In particular, the role of dance artificial intelligence is important. Recently, in addition to the popularity of K-POP, research on artificial intelligence for artificial intelligence-based creative dance has been actively carried out [3]. AI choreographer research suggests a direction to create new creative dances based on the Full Attention Cross Modal Transformer (FACT) network to 3 Dimensions (3D) music [3]. To create new creative dances, they use 10 genre dance covers and camera multi-view video. This provides a transformative model that generates 3D motion related to music. However, in order to prepare the dance cover necessary for the creation of a creative dance, the motion of various dancers is required. This process causes problems, such as expensive input source datasets and the cost of switching to the target source to be used in the model. In particular, there is a problem to be considered when converting different motions between various dance genres. In order to solve this problem, future research is needed to promote creative dance in a new direction by minimizing the dependence of dancers in consideration of various genres and inducing cost reduction by enabling processing without using expensive motion capture devices.

This paper proposes an advanced system in virtual environments, which automatically generates a series of creative dance movements by identifying postures and gestures from K-POP dance videos previously collected and analyzing the identified postures and gestures based on bidirectional long-short-term memory (Bi-LSTM). Given that long-short-term memory (LSTM) performs better with small amounts of data than other temporal deep learning algorithms, such as transformer [4], LSTM is utilized for the proposed method considering the amount of input data. The proposed system consists of the Dance-Choreo-Input-Unit, the Dance-Choreo-Generation-Unit, and the Dance-Choreo-Output-Unit. The Dance-Choreo-Input-Unit defines a source video and the user. Here, a source video refers to a dance video used to define a dance posture, and the user refers to the main agent who is willing to obtain the result of dance choreography. The Dance-Choreo-Generation-Unit creates new choreography performed by a 3D dance character by analyzing a source video input and uses Bi-LSTM to facilitate the learning process of the proposed system based on the generated choreography. The Dance-Choreo-Output-Unit applies the final creative choreography derived through the previous learning process in various ways. In this paper, an Advanced Dance Choreography System is proposed and was verified with K-POP videos. In the experiment, the creation process of K-POP choreography was verified by generating creative choreography by obtaining posture from K-POP dance videos of human models or 3D characters in a virtual environment. This system limited the genres of choreography in consideration of the popular characteristics of K-POP choreography but plans to expand experiments by analyzing and applying other genres, complexity of genres, and difficulty within one genre in the future. To this end, exceptional circumstances that may occur when applied to the proposed system should be considered.

The advantages of the proposed system are as follows. First, it does not require the use of expensive devices, cameras, or sensors. Most existing studies on motion capture extracted movement data by using expensive devices, which caused limitations in data collection and use. On the other hand, the proposed system in this paper does not need to use expensive devices because it analyzes dance videos that are uploaded online or possessed by the user. Second, the proposed system can create new choreography regardless of genre. There are a variety of dance genres, such as performing arts, street dance, and dance sports, and each dance genre comprises sub-genres. For example, performing arts are divided into ballet, modern and contemporary dance, and jazz dance; street dance is divided into hip hop,

popping, waacking, krumping, locking, house dance, and breaking; and sports dance is divided into jive, tango, and the cha-cha-cha [3]. Since the proposed system in this paper generates postures and gestures by analyzing dance movements included in a source dance video, it can create new choreography based on diverse dance videos regardless of genre. Third, the proposed system reduces dependence on dance experts in the process of creating new dance. In the choreography creation process, choreographers generally rely on the capabilities and reputation of dancers. Accordingly, they encounter difficulties in inviting professional dancers and spend a long time completing the choreography creation process. To solve these problems, this paper presented a system that enables users to compose high-quality creative choreography by using only videos.

The proposed system in this paper is distinguished from existing choreography creation solutions for the following reasons. First, the proposed system can automatically establish data on postures and gestures based on a source dance video, whereas existing tools can establish data on postures and gestures only after inviting professional dancers and storing and collecting a great amount of data on their dance movements according to dance genres. Second, the proposed system does not require high-precision videos to derive key points. Existing choreography creation tools can use a video of the dance movements of a dancer only when this video satisfies several video recording conditions, such as lighting and distance between a camera and a dancer. Moreover, these tools cannot accurately distinguish dance movements from the background in the process of deriving key points from a source video. On the contrary, the proposed system in this paper can determine key points by using even a low-resolution video and thus contributes to reducing pressure on video recording. Third, the proposed system automatically extracts dance postures based on a source video. Users of existing choreography creation tools should prepare different types of hardware and sufficient space to record a video of dance movements based on expensive devices, including cameras, and to extract dance movements from the recorded video. Fourth, the proposed system encourages the supply of creative dance by lowering dependence on dance experts, unlike existing choreography creation tools that highly depend on the participation of dance experts to compose new choreography and that demand the effort and time of users in inviting dance experts.

The structure of this paper is as follows. Section 2 introduces relevant research, and Section 3 presents the proposed K-POP choreography creation system. Section 4 describes experiments conducted to verify the performance of the proposed system and indicates the analytic results of its performance. Finally, Section 5 derives conclusions on the proposed system.

2. Related Works

In this Section, this paper reviewed the existing system for choreography creation and specific movement estimation techniques. Existing systems for choreography creation include AI Choreographer [3], temporal deep learning algorithms [4], Everybody Dance Now [5], Bi-LSTM and Two-Layer LSTM for Human Motion Capture [6], Dance toolkit [7] and others. Among these systems, this paper focused on analyzing AI Choreographer and Everybody Dance Now. As for movement estimation techniques, it examined movement estimation techniques based on Bayesian probability and deep learning.

2.1. Frameworks for Choreography Creation

AI Choreographer [3] consists of AIST++, advanced industrial science and technology, a new multi-modal dataset of 3D dance movement and music, and the Full Attention Cross Modal Transformer (FACT) network for generating 3D dance movement conditioned on music. The AIST++ dataset contains 5.2 h of 3D dance movement in 1408 sequences and covers 10 dance genres based on videos recorded by cameras, which satisfied multi-view conditions according to the location shown in these videos. In general, the application of sequence models, such as transformers, to this dataset for the task of music conditioned 3D movement generation does not produce satisfactory 3D movement. In order to train the proposed model, data issues, motion capture data collection, and a highly instrumented

environment are required. There are also problems with datasets severely limited in the number of dances available, the variety of sequenced dancers and music, and reliable 3d motion recovery using multi-view video. To overcome this shortcoming, the developers of this solution adopted key changes in its architectural design and supervision. Moreover, they applied a deep cross-modal transformer block to the FACT model to train this model to predict several future movements. These changes serve as key factors in generating long sequences of realistic dance movement, which are well-attuned to the input music. People can compose movement patterns according to music beats and dance based on these patterns. This ability is considered an essential aspect of human behavior. Yet, even dancers, who can perform abundant dance movements, should receive professional training to create expressive choreography. Effective choreography creation methods are unlikely to be developed in that these solutions should be able to compose continuous movements by reflecting kinematic complexity to detect a non-linear relationship between music and movement based on calculation. Beyond these obstacles, the realistic 3D dance movement generation model indicated above can effectively learn a relationship between music and movement and generate sequences of dance movement which vary according to the input music.

Everybody Dance Now [5] presents Do as I Do, an effective method for retargeting a video. This method automatically transfers the movement of a person dancing in the source video to that of a target person. It is operated based on two videos, a source video where the movement of a person dancing in this video is transferred to a target person and a target video where the movement of a target person who appears in this video is morphed with the transferred movement. This method learns simple video-to-video translation to facilitate movement transfer between the source and target videos. Users of this system can generate numerous videos where an untrained amateur rotates his or her body like a ballerina or performs various ballet dance movements. Frame-based movement transfer methods should learn the mapping between two images to facilitate movement transfer between two videos. However, the Do as I Do approach can perform accurate frame-to-frame pose correspondence according to the body type and movement style of the target. The Everybody Dance Now solution also presents an image translation model between pose stick figures from the source image and the target image. Pose stick figures from the source are input into the trained model to obtain images of the target subject in the same pose as the source and to ultimately transfer movement from source to target. This system contributed to present a method that can generate results of transferring human movement.

2.2. Movement Estimation Techniques

It is recommended to reduce the number of sensors used to estimate movements in movement estimation techniques due to the high cost of wearable sensors. A few existing movement estimation algorithms based on Bayesian probability and K-means are used to estimate movements of body parts, which are not monitored, by considering movements directly measured by sensors [8,9]. Bayesian probability was first adopted in a movement estimation technique to estimate the movements of arms [10]. Two armbands were used to detect these movements, and the measured results were displayed as coordinate values in the direction of the target arm. Measurement data were classified according to angles ranging from -180° to 180° at the interval of 30°. The movement of the upper arm was estimated based on the maximum Bayesian probability between the angle of the upper arm in the initial place and that of the upper arm in the adjusted place in the direction of the movement. An armband represents the movement of an arm (the upper arm and forearm). Subsequently, an enhanced movement estimation technique based on Bayesian probability was developed. This method accurately estimates movements by determining a range of angles by using the minimum and maximum values of measured data instead of a fixed range of angles from -180° to 180°. This method was used to estimate the direction of the forearm based on the location of a hand connected to the forearm [11]. An armband was used to detect the direction of the forearm, and a VIVE controller was used to collect data on the location of the target

hand. This technique estimated the direction of the target forearm, which was not measured, based on the location of the hand detected and the Bayesian probability between the direction of the forearm measured and the location of the hand. Movement estimation techniques are appropriate for movement estimation based on previously defined behaviors. These techniques collect a great amount of data by using sensor-based wearable devices. However, the collected data only partially match the previously defined behaviors. This result indicates that an enormous dataset does not contribute to increasing the performance of movement estimation techniques based on Bayesian probability. To solve these problems, researchers have recently applied deep learning in movement estimation techniques to increase the performance of these techniques.

Deep learning is a machine learning technique that facilitates the processing of large-volume data and has thus been widely used in various fields. This technique is also the most widely used method for movement estimation based on its advantage of processing a considerable amount of data. Several researchers reported that human motion tasks based on deep learning algorithms derived the most excellent results compared to those derived from previous motion capture tasks not applying deep learning [12]. An existing paper developed a system based on Deep Neural Networks (DNNs) to accurately estimate a 3D pose from multi-view images [13]. MoDeep [14], a deep learning framework using motion features for human pose estimation, is a deep learning framework that was developed to estimate the 2D location of human joints based on features of movements in a video. A convolutional network architecture deals with color and movement features based on a sliding-window architecture. The input to the network is a 3D tensor containing an RGB image and its corresponding movement features, in optical flow, and the output is a 3D tensor comprising one response map for each joint. A Convolutional Neural Network (CNN) was trained to estimate unsupervised movements [15]. The input to this network is a pair of images, and a dense movement field can be produced on its output layer. This full CNN comprises 12 convolutional layers that can be regarded as two parts. In the first part, the CNN simply represents information on movements, which involves four down samplings. In the second part, the compact representation is used to reconstruct the movement field. Four upsamplings are involved in this process, and the movement of the movement can be estimated. MoDeep estimated human poses by using the FLIC-movement dataset [16], which comprises 5003 images collected from Hollywood movies, and additionally included movement features. Then, another paper presented a new way of training a CNN based on pairs of consecutive frames from the dataset UCF101 [17], a dataset of 101 human actions classes from videos in the wild. Both approaches estimated movements by using visual information on human movements included in a video. The goal of these approaches was to estimate movements in video frame sequences. Another paper proposed a method of investigating the performance of deep learning networks that use LSTM units to manage the sensory value of an Inertial Movement Unit (IMU) and to use sensory data [18]. This existing paper verified that the proposed approach based on machine learning detected the surface conditions of roads and the age-group of the subjects by analyzing sensory data collected from the walking behaviors of subjects.

2.3. Virtual Reality

Virtual reality (VR) has been widely used as a tool to assist people by learning and simulating situations that are difficult to do in real life [19–21]. It is used in various ways, such as safety education and education in a game-like environment, and constitutes an interactive and immersive education environment. Virtual simulation minimizes constraints and contributes to the general domain.

Simulations and similar strategies are used in the study of common real-world problems and events. Virtual reality for dangerous situations can reliably acquire new skills and experiences, and is suitable for developing students' imagination, communication skills, and creativity. The main point of this study is to make the learning and education process exciting and expandable with entertainment, and to turn interest in the traditional learning

and training process into an interest in learning through user interaction with the help of a virtual environment. This study investigates the benefits of increasing participation by introducing a virtual environment to a small number of student participants in the COVID-19 pandemic situation.

First, additional training scenarios can be added and contain more complex tasks. Second, it can incorporate a robust scoring system by taking into account level completion times, allowing for greater user engagement. Third, additional gestures may be included to provide a more fluid and natural user experience. Finally, various environments that can be introduced in reality can be applied to the virtual environment as a learning experience.

2.4. Comparison with the Proposed System

AI Choreographer learns a relationship between music and movement and generates sequences of dance according to various types of music. It manages 5.2 h of 3D dance movement in 1408 sequences and covers 10 dance genres based on videos recorded by cameras, which satisfied multi-view conditions according to the location shown in these videos. In the dance sequence generation process, the learning performance of this framework is unlikely to increase due to the kinematic complexity between music and movement. Choreographers should collect a dataset of dance movements to design sequences of creative dance movements and should spend a great amount of time selecting skilled dance experts to work with them. Moreover, the choreography process incurs high costs related to the purchase and installation of hardware, such as expensive devices, equipment, and cameras for recording videos. Previous studies on AI-based choreography derived limited outcomes due to the difficulty in obtaining a dataset of dance movements and the high cost required for the establishment of hardware. To solve the aforementioned problems, this paper proposed an algorithm, which can perform creative choreography by minimizing hardware and software costs required to collect a dataset of dance movements, and a system applying the developed algorithm. The proposed system generates sequences of creative dance movements by using the movements of a dancing character, which were extracted from various dance videos. Based on the proposed system, this paper intends to contribute to developing the dance motion generation industry.

3. System for K-POP Choreography Creation

This paper presented a K-POP choreography creation algorithm based on Bi-LSTM and a system for K-POP choreography creation applying the developed algorithm. The proposed algorithm extracts postures from a dance video input as a source video and groups a series of postures as a gesture. Then, it uses words to represent each gesture and performs choreography based on the selected words. The aforementioned processes are divided into units, which form the system for K-POP choreography creation. This section describes the role, functions, and elements of each unit in detail.

3.1. Overview

Figure 1 shows the proposed system for K-POP choreography creation, which is operated based on the Dance-Choreo-Input-Unit, the Dance-Choreo-Generation-Unit, and the Dance-Choreo-Output-Unit. The Dance-Choreo-Input-Unit defines a source video and the user, and the Dance-Choreo-Generation-Unit extracts postures and gestures from the source video and converts the extracted movements to creative choreography. The Dance-Choreo-Output-Unit applies the result of creative choreography as input data.

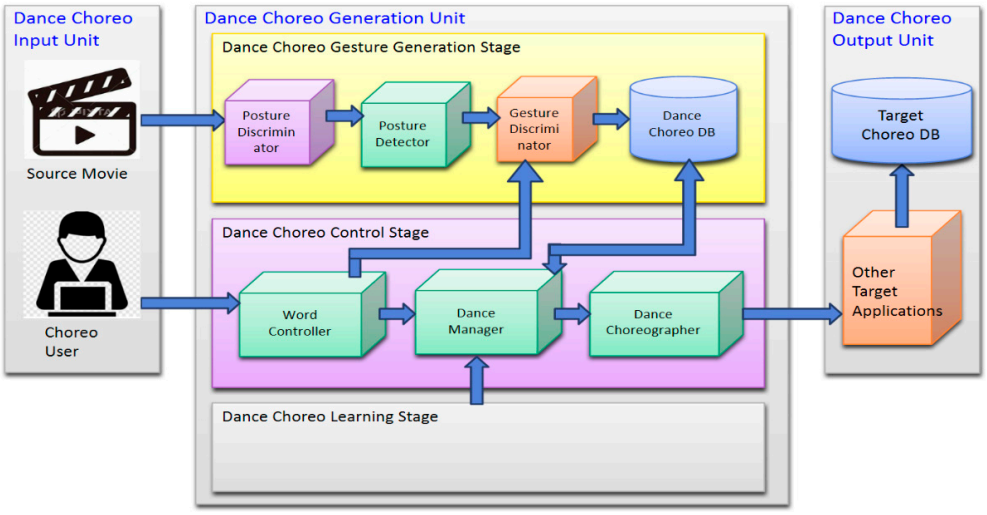


Figure 1. K-POP choreography system using Bi-LSTMs. System structure diagram showing the process of creating a creative choreographer through the system from a dance video.

The Dance-Choreo-Input-Unit is associated with a source video and the user. A source video is a dance video input to define dance postures. There are shooting conditions for the usable video for the dance video input. A fixed camera angle, the model of the video is a character or human model, and an object is also based on the video taken in the surrounding environment and the brightness of the model that can be distinguished. The minimum size of the model is retrieved through the bounding box and the environment and model are distinguished as regions. A user is a person who is willing to perform creative choreography and conduct the tasks of inputting, editing, and saving necessary words for creative choreography. There are conditions for the usable video for the dance video input. A fixed camera angle, the one model of the video is a character or human model, and an object are also based on the video taken in the surrounding environment and the brightness of the model that can be distinguished. The minimum size of the model is retrieved through the bounding box and the environment and model are distinguished by area. These constraints will be further explained in the video input area. The Dance-Choreo-Generation-Unit creates a new sequence of dance movements by using the source video input and a 3D dance character or a dancing character which shows dance movements. Specifically, the Dance-Choreo-Generation-Unit includes the following three stages: the Dance-Choreo-Gesture-Generation-Stage, the Dance-Choreo-Control-Stage, and the Dance-Choreo-Learning-Stage. In the Dance-Choreo -Gesture-Generation-Stage, the Dance-Choreo-Generation-Unit extracts postures from a source video, allocates numbers to each posture according to a certain order, and groups several postures to generate a gesture. The generated gesture is recorded in a DB of creative choreography called a Dance-Choreo-DB. The Dance-Choreo-Control-Stage consists of a word controller, which enables the user to input multiple words, a dance manager, which matches a gesture with a word based on the one-hot-encoding technique, and a dance choreographer, which performs choreography by using words provided by the user. In the Dance-Choreo-Learning-Stage, the Dance-Choreo-Generation-Unit operates the Bi-LSTM network to provide the customized result of creative choreography by analyzing the user's intention. In this stage, the Dance-Choreo-Generation-Unit rearranges gestures by analyzing a relationship among the selected gestures in the Dance-Choreo-Control-Stage. The Dance-Choreo-Output-Unit applies the result of creative choreography, which was derived from the previous learning process, in different ways. Various target applications can be used to input the result of creative choreography in a target DB of choreography. The result of creative choreography derived from the proposed system is applicable to a variety of fields, ranging from research to entertainment business and creative choreography. The Dance-Chore-Generation-Unit serves

as a core function for composing a sequence of creative dance movements. The following sections provide detailed descriptions of each stage of the Dance-Chore-Generation-Unit.

3.2. The Dance-Choreo-Gesture-Generation-Stage

The Dance-Choreo-Gesture-Generation-Stage is conducted to extract postures and gestures from a source video and store the extracted data in the Dance-Chore-DB. This stage is operated based on a posture discriminator, a posture detector, a gesture discriminator, and the Dance-Choreo-DB. The posture discriminator generates a posture by establishing 29 key points on an image of a dancing character. The posture detector automatically allocates numbers to the extracted postures according to a certain order and an index and determines the number of postures to be grouped as a gesture. The gesture discriminator groups multiple postures to generate a gesture. The Dance-Choreo-DB records the generated gesture.

3.2.1. Posture Discriminator

The posture discriminator generates postures by using an online dance video as a source video. Each posture is indicated in an image file and a script file containing 29 key points. Human pose estimate research is widely used for 3D characters and virtual machines, and among many studies, it is determined based on the thesis that proposes the number of key points that are unnecessarily large or too small to express motions in expressing dance motions [19]. In describing the human posture, the 29 skeletal points are suitable for expressing dance movements with minimal skeletal points such as the direction of the body, the direction of arms and legs, heels and toes, and fingers. The posture discriminator was designed to save up to 100 postures. Figure 2 shows two file types required to represent a posture.

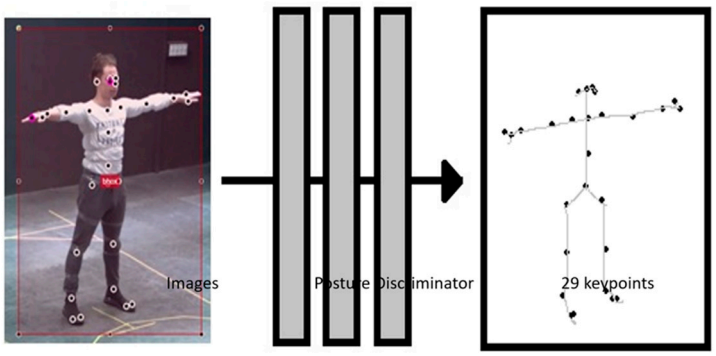


Figure 2. Human model with 29 key points for one posture, 29 key points extracted from a dance video through one posture.

3.2.2. Posture Detector

The posture detector automatically provides numerical indices to multiple postures generated by the posture discriminator. It receives the number of postures to be grouped, which is established by the user, to determine the number of postures that comprise a gesture. Figure 3a shows the postures of a dancing character including key points. Postures are indexed based on numbers ranging from 0 to 99. Figure 3b shows the process of receiving the number of postures to be grouped as input data to generate a group of postures, which is used to define a gesture. When the number of postures to be grouped is established, the posture detector determines the number of gestures by considering the number of the entire postures.

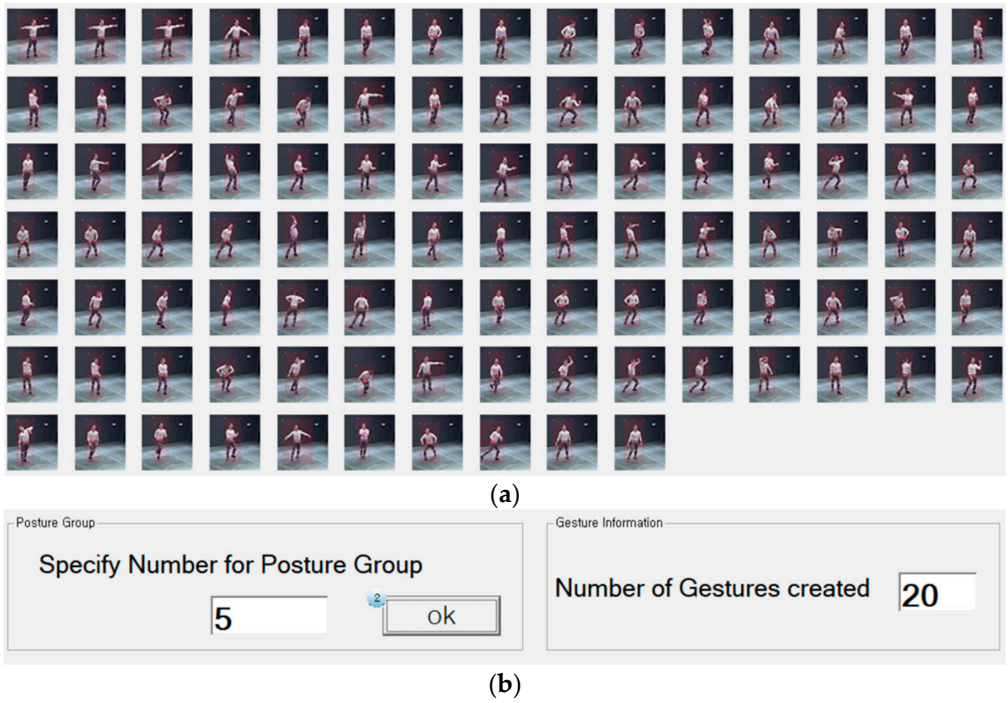


Figure 3. Two functions of posture detector, they should be listed as: (**a**) description of 100 postures in the first panel; (**b**) setting the number of grouped postures for gestures.

3.2.3. Gesture Discriminator

The gesture discriminator generates a gesture according to the number of postures grouped. Figure 4a shows a case where 100 postures are set to be divided into groups of five postures. In this case, 20 gestures are generated. Figure 4b shows a case where 100 postures are set to be divided into groups of ten postures. In this case, 10 gestures are generated.

The gesture discriminator performs the following steps. First, it generates a gesture by dividing consecutive postures according to the number of postures to be grouped. A red dotted line box located at the left in Figure 5 shows a case where gestures are generated according to the number of postures to be grouped. Each gesture includes a gesture number and multiple posture numbers. A word box is left blank without a word input. Second, each gesture is selectively established by words defined and delivered by the user, who inputs these words by using the word controller in the Dance-Choreo-Control-Stage. The word input step is described in the corresponding section indicated below. A red dotted line box located at the right in Figure 5 shows a case where a word for a gesture is established. In this step, each gesture contains a gesture number, a gesture word, and multiple posture numbers.

3.2.4. Dance-Choreo-DB

The Dance-Choreo-Database (DB) stores the gestures defined by the gesture discriminator for data collection and keeps records of a gesture number, a gesture word, and a posture group. Figure 6 shows a relationship between the posture detector, which represents the input of the gesture discriminator, and the Dance-Choreo-DB, which represents the output of the gesture discriminator. The No. 1 arrow shows a step where the gesture discriminator uses a group of postures defined by the posture detector to generate a gesture and designate a word for the generated gesture. The No. 2 arrow shows a step where the gesture defined by the gesture discriminator is stored in the Dance-Choreo-DB.

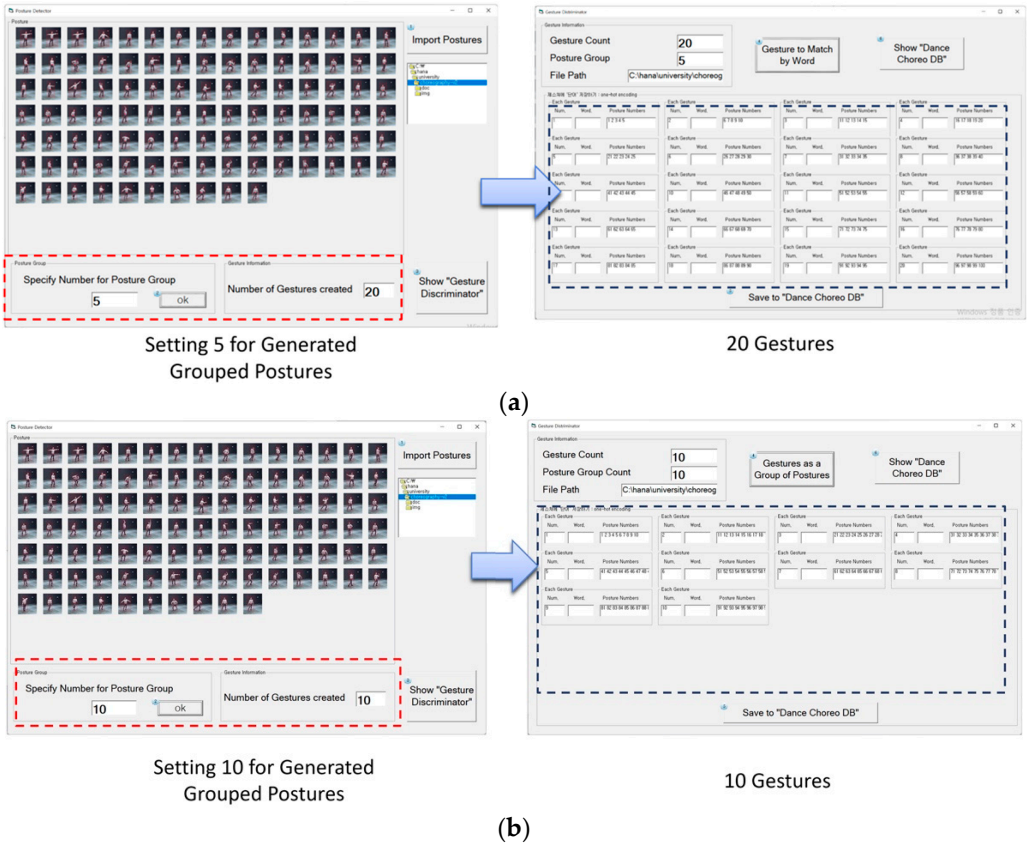


Figure 4. Thirty generated gestures by grouping with postures, they should be listed as: (**a**) 20 generated gestures by grouping 5 postures; (**b**) 10 generated gestures by grouping 10 postures.

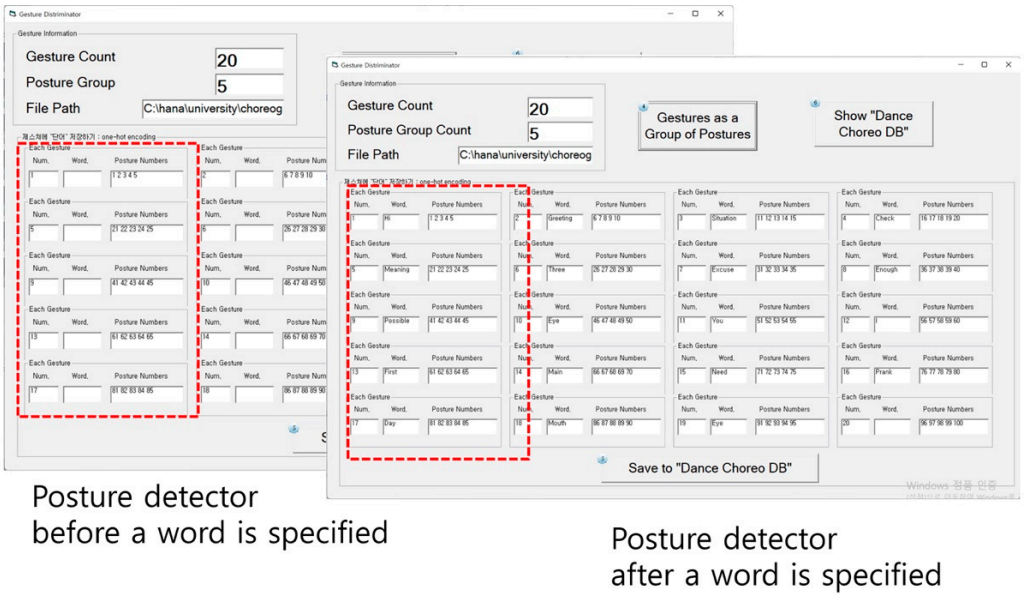


Figure 5. Two case of gesture discriminator with or without user controlling.

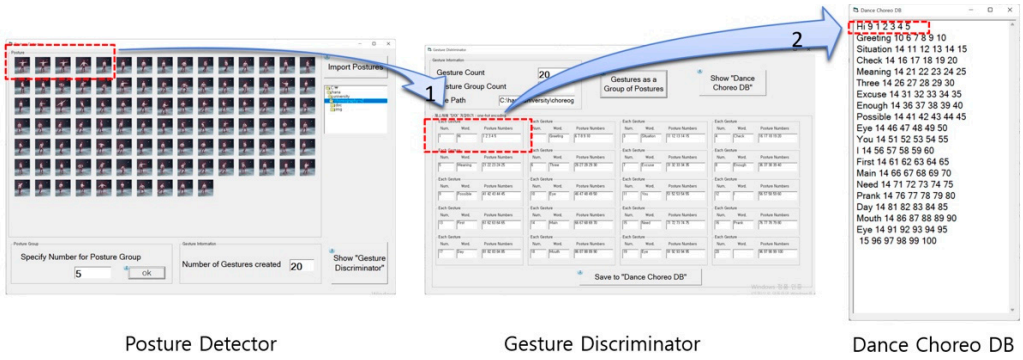


Figure 6. Information of Dance-Choreo-Database saving to word, the process of grouping 5 postures, matching them with a single word gesture, and saving them in the DB.

3.3. Dance-Choreo-Control-Stage

In the Dance-Choreo-Control-Stage, the Dance-Choreo-Generation-Unit edits the initial result of creative choreography stored by the user intervention and outputs the revised result of creative choreography. This stage is operated based on a word controller, a dance manager, and a dance choreographer. The word controller receives multiple words input by the user and generates gesture words. The dance manager detects words input by the user from the Dance-Choreo-DB and matches these words with posture groups. The word and posture numbers can be edited according to the user's intention to expand a range of creations for choreography. The dance choreographer generates creative choreography through the aforementioned step.

3.3.1. Word Controller

The word controller receives a series of words from the user or a producer who is willing to obtain the result of choreography. The user inputs words by using a keyboard, and the word controller extracts words to be used to compose a sequence of creative dance movements. The extracted words are used by the gesture discriminator and the dance manager as follows. The gesture discriminator allocates a word determined by the word controller to an automatically generated gesture to interconnect them. The dance manager searches a word determined by the user in the Dance-Choreo-DB and selectively matches a group of postures, which is established to generate the corresponding gesture, with the searched word. Figure 7 shows a screenshot of the word controller containing words randomly input.

Figure 8 shows the step where gesture words are transferred to the gesture discriminator and the dance manager. The left image of Figure 8 shows the step where the word controller analyzes a series of words input by the user and divides these words into multiple words. The right images of Figure 8 show the gesture discriminator and dance manager called the word controller. The No. 1 arrow indicates a step where the word controller calls the gesture discriminator to convert the word determined by the word controller to a gesture word and record the converted word. The No. 2 arrow indicates a step where the dance manager detects the word determined by the user from the Dance-Choreo-DB and extracts posture numbers included in the corresponding gesture.

3.3.2. Dance Manager

The dance manager detects gesture words stored in the Dance-Choreo-DB and lists posture numbers included in the corresponding gesture. The word controller extracts multiple words from a series of words input by the user according to the user's intention. When the Dance-Choreo-DB includes a gesture word that corresponds to the extracted word, the dance manager extracts several posture numbers included in the corresponding gesture word. Creative choreography consists of a series of gestures, each of which contains several posture numbers. The proposed system is trained based on the Bi-LSTM technique

to automatically control the level of creativity, preference, and difficulty. Figure 9 shows a step where the dance manager detects and arranges several posture numbers from multiple gesture words. The No. 1 arrow indicates a step where the dance manager detects a gesture word from the Dance-Choreo-DB. The No. 2 arrow indicates a step where the dance manager detects several posture numbers that belong to the corresponding gesture word recorded in the Dance-Choreo-DB. When the user clicks the Learn button marked by the No. 3 red dotted line box, the proposed system learns the level of dance creativity, preference, and difficulty by using the Bi-LSTM technique and updates the arrangement of posture numbers.

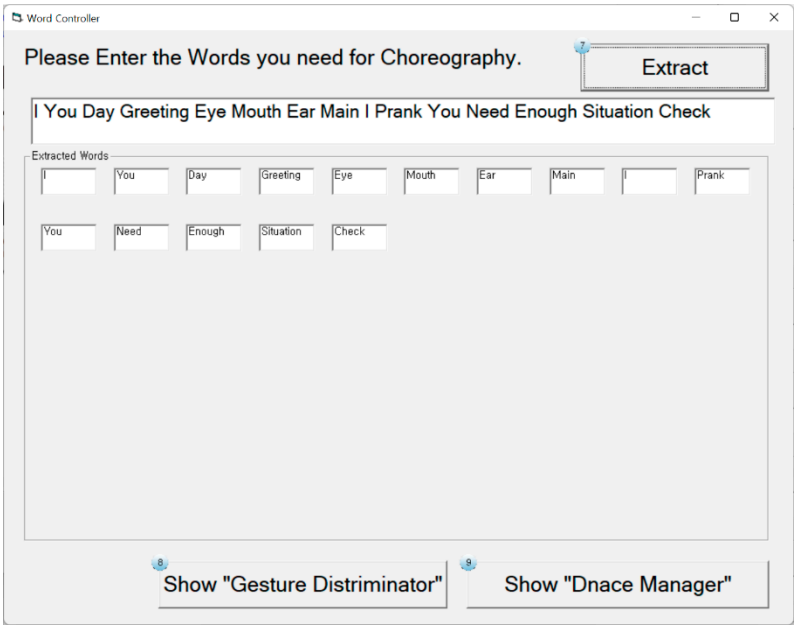


Figure 7. Screenshot of word controller, the process of dividing the words entered by the user into words to deliver them as a single gesture.

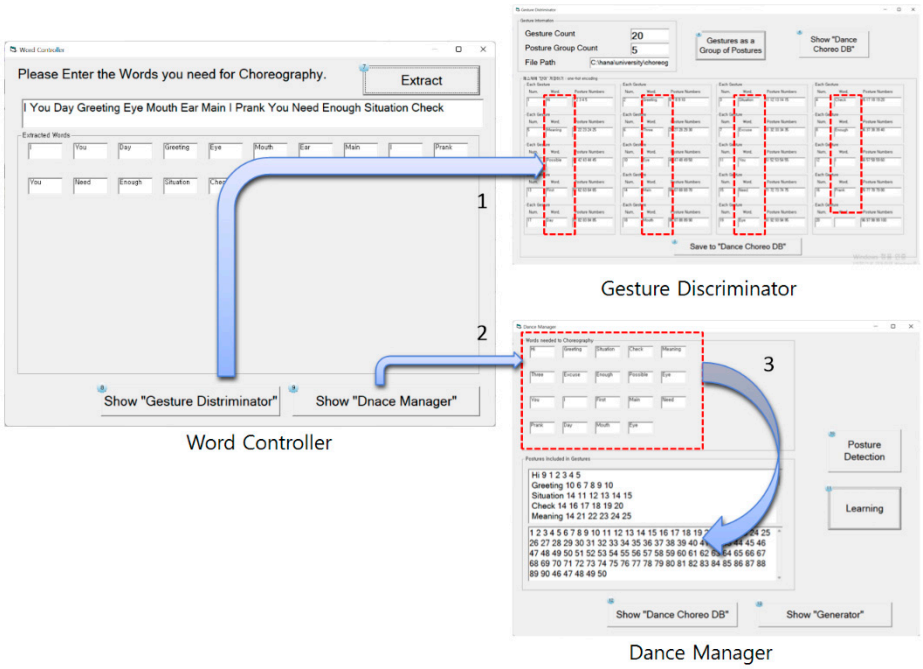


Figure 8. Two cases following to gesture discriminator or dance manager from user controller.

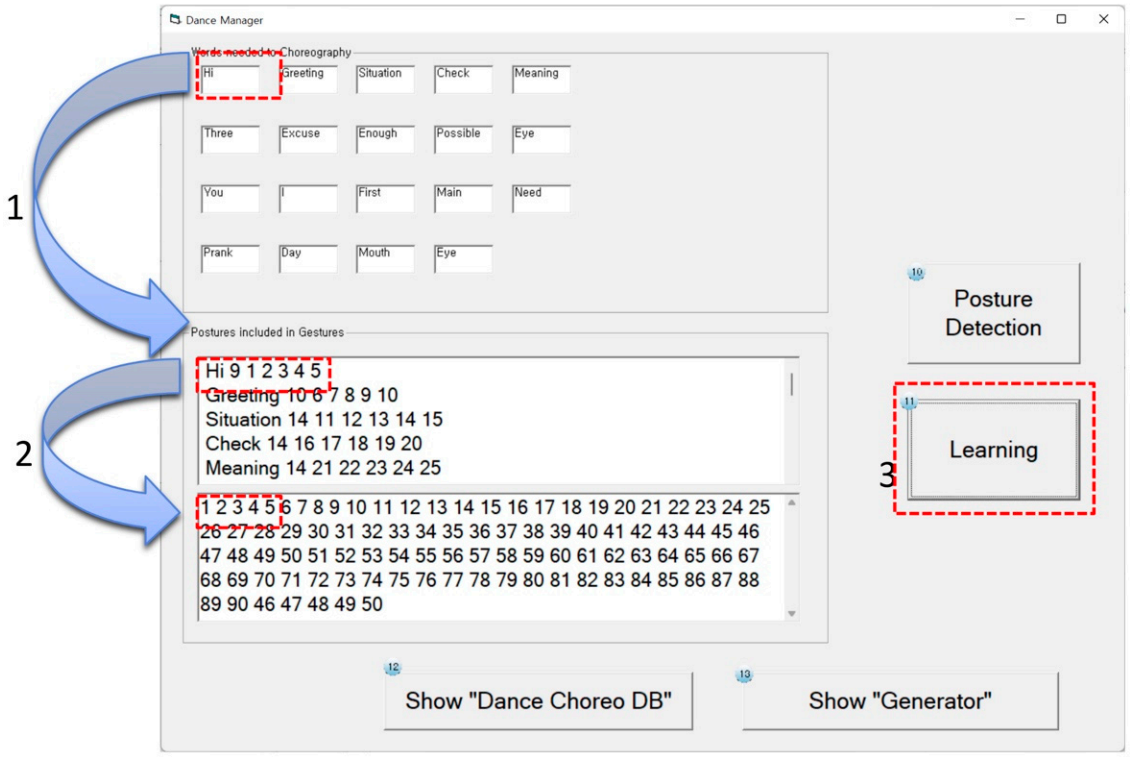


Figure 9. Arrangement of posture numbers from gesture word using Bi-LSTM, a process of showing multiple posture bundles corresponding to multiple gestures and the learned results.

3.3.3. Dance Choreographer

The dance choreographer analyzes gestures generated by the dance manager and creative choreography updated by the Bi-LSTM technique. The purpose of the proposed system is to generate creative choreography through the aforementioned processes. Figure 10 shows a screenshot of a video of a character dancing based on the creative choreography derived from the proposed system. The No. 1 part indicates records of posture numbers learned by the proposed system, the No. 2 part indicates a necessary controller for playing the video, and the No. 3 part indicates an area that displays the creative choreography derived by the proposed system.

3.4. Dance-Choreo-Learning-Stage

In the Dance-Choreo-Learning-Stage, the proposed system learns the allocated gestures according to conditions defined by the user, such as the level of dance creativity, preference, and difficulty, to expand the range and use the scale of new choreography created. The proposed system establishes 29 key points for each posture and determines the level of difficulty of dance by comparing coordinate movement values and the range and location of the distribution of dots between postures. Specifically, it receives coordinates of key pointers as input data and learns coordinate movement values and the range and location of the distribution of dots based on the input data to determine the level of difficulty of dance. Figure 11 is the Bi-LSTM structure diagram. The posture numbers included in the gesture are input as experimental data. It shows the process of generating prediction data through learning of Bi-LSTM and generating it as training data.

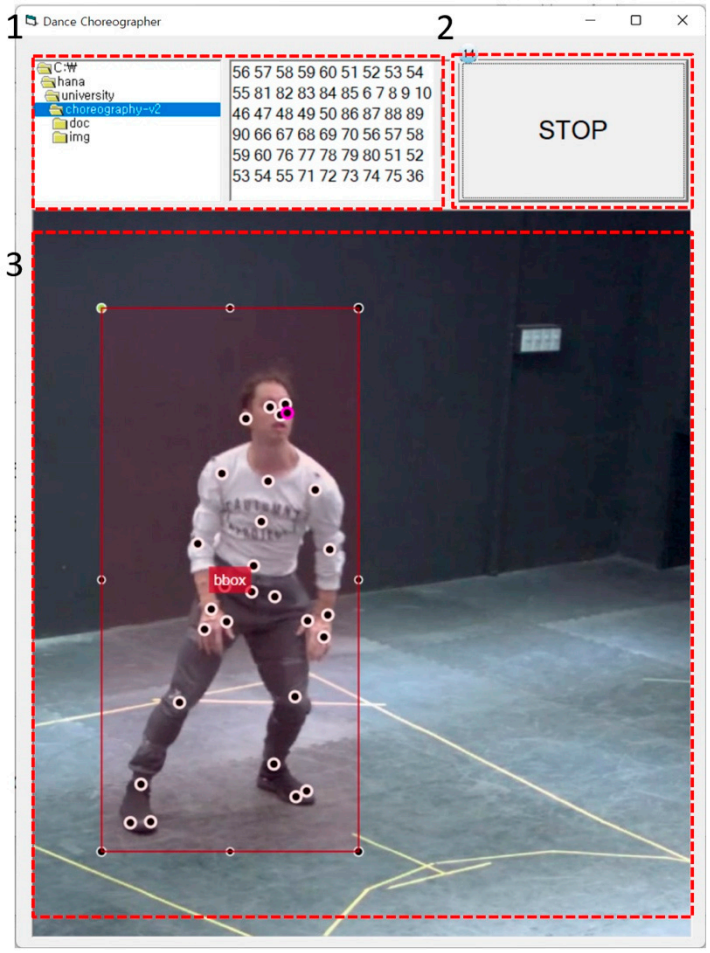


Figure 10. Execution of generated dance choreography including Bi-LSTM.

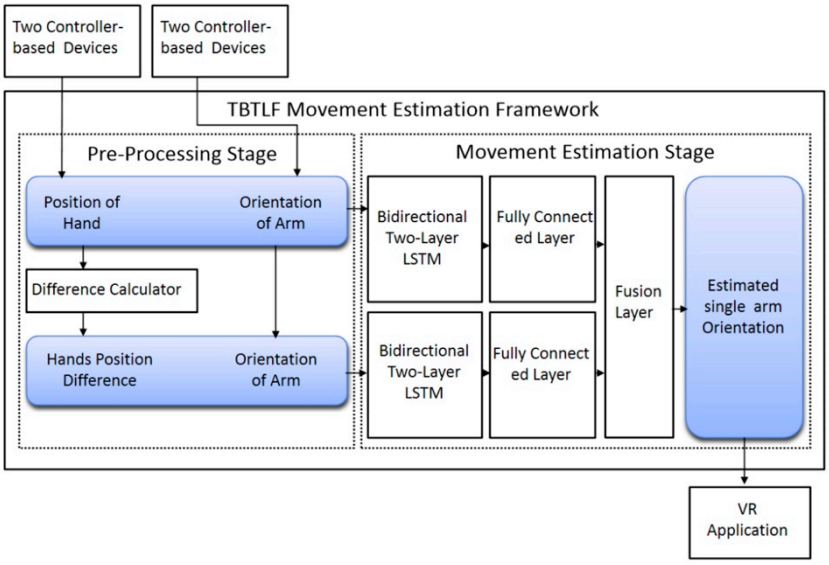


Figure 11. Overview of the Bi-LSTM, the Bi-LSTM structure diagram that generates prediction data with experimental data as input and generates it as training data.

Figure 12 shows the location of dots displayed when key points in a posture are moved to those of another posture. Figure 12a is an example of a representation of a dance movement that shows a high level of difficulty in dance. This dance movement is analyzed

particularly based on red dots, which indicate the ring fingers of the right and left hands among 29 key points. The x and y values at the center of the bounding box (bbox) are established as (0, 0). The ring finger of the right-hand moves from (−9, 8) to (−9, 9), and the ring finger of the left-hand moves from (9, 8) to (−4, 6). This result indicates that the right-hand shows a wider range of movement than the left-hand. The movement gap of the ring finger of the right-hand is 1, and that of the ring finger of the left-hand is 13. The proposed system determines weight based on the learning result. Figure 12b is an example of dance movement that shows a low level of difficulty in dance. The ring finger of the right-hand moves from (−9, 8) to (−6, 8), and the ring finger of the left-hand moves from (9, 8) to (3, 8). The movement gap of the ring finger of the right-hand is 3, and that of the ring finger of the left-hand is 6. The proposed system determines weight based on the Bi-LSTM learning result.

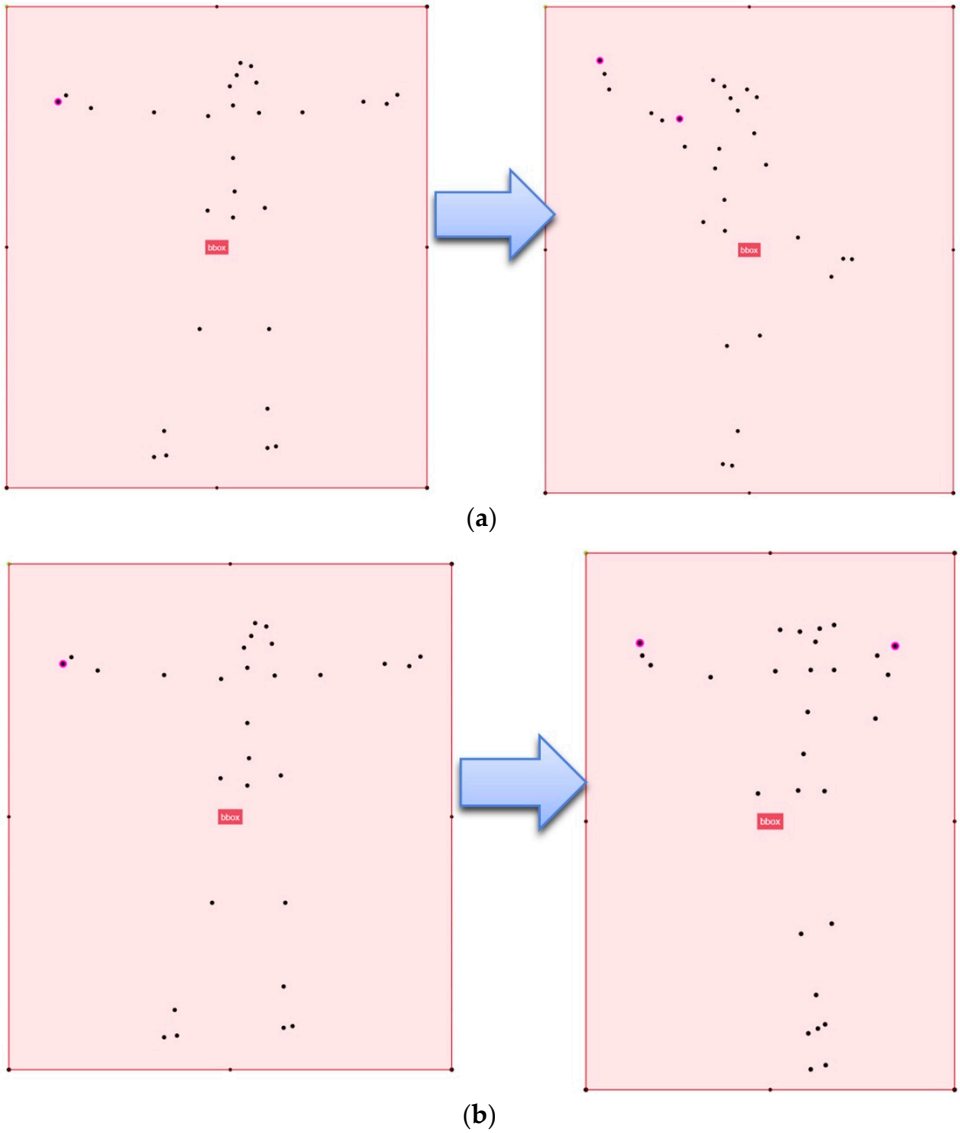


Figure 12. Dance level of the key point position, they should be listed as: (**a**) high case of key point movement range; (**b**) low case of key point movement range.

3.5. Flow Chart of Creative Choreography Data

The proposed system captures the main dance movement from a source video and saves it as an image file. Then, it establishes 29 key points, which indicate the frame of a dancing character or a 3D character, on the image file and saves the result as a script file. It combines both files to generate a posture, and up to 100 postures can be generated. The

posture detector receives several postures, and the user determines the number of postures to be grouped in the process. The posture detector designates posture groups by grouping postures, and these posture groups are transferred to the gesture discriminator. The posture group can be defined as a gesture and recorded in the Dance-Choreo-DB. In the following processes, the user stores gesture words and generates creative choreography. The word controller receives multiple words input by the user. The input words are sequentially recorded as gesture words, and these gesture words are stored in the Dance-Choreo-DB for the dance manager. When gesture words are already stored in the Dance-Choreo-DB, the dance manager arranges posture numbers that belong to the corresponding gesture. The dance choreographer receives the arranged posture numbers and generates creative choreography based on the data. Figure 13 shows the data flow chart of the proposed system for creative K-POP choreography generation.

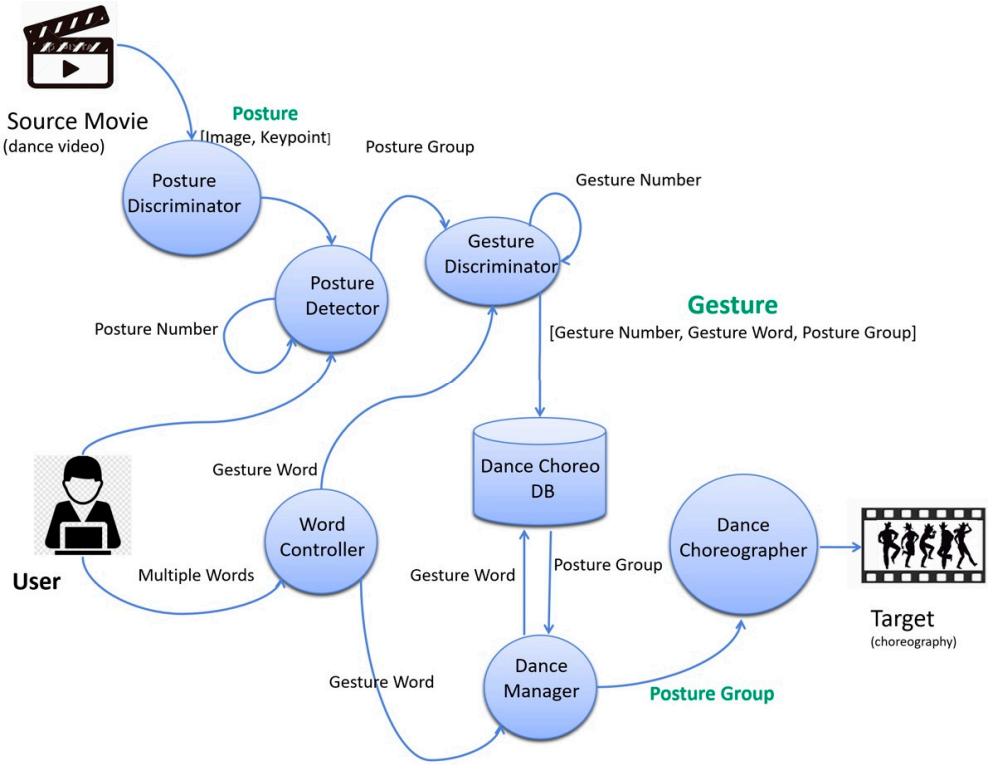


Figure 13. Data flow chart of K-POP choreography system, a data flow chart showing the relationship between input and output of each internal function in the process of creating creative choreography from experimental data.

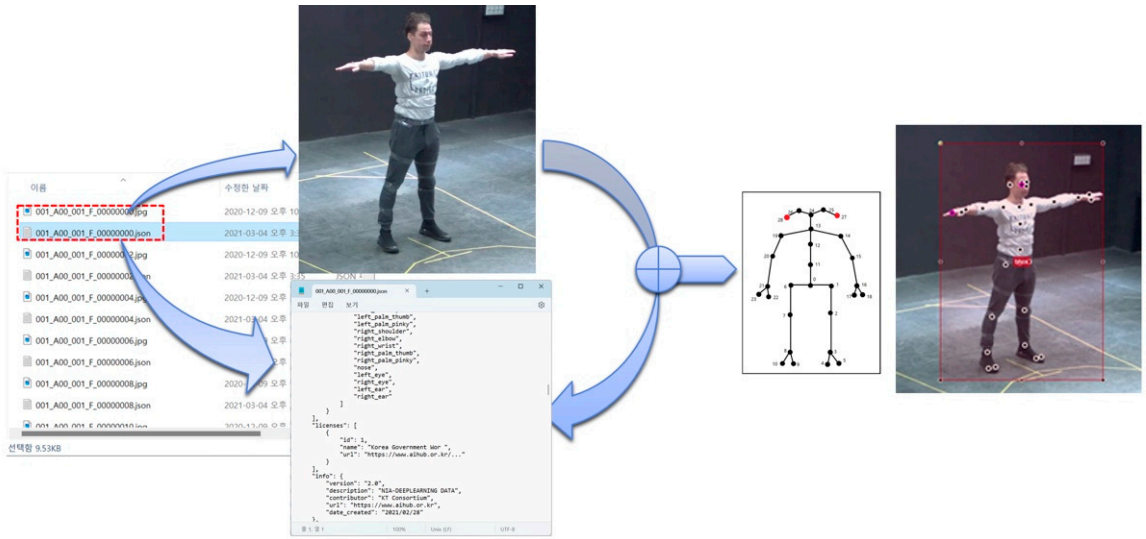
4. Experiments and Results

This section describes the environment established to verify the proposed system and presents experimental results obtained from the aforementioned environment. This paper conducted experiments to analyze sequences of creative dance movements generated by the proposed system.

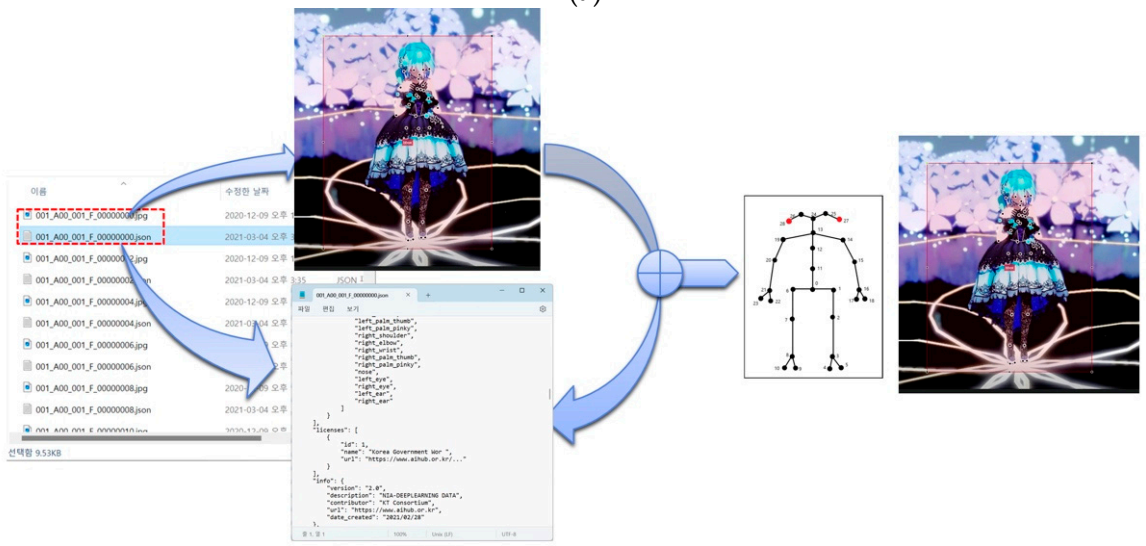
4.1. Experimental Environment

Dance videos uploaded online can be used as source videos in this paper. However, this paper limitedly used dance videos of a dancing character provided on AI HUB, an integrated AI platform, and dance videos of 3D animation characters, which were designed in this paper, because of copyright issues. In this paper, two types of experiments were carried out. To generate a posture, the proposed system generated an image file of a character and a script file where 29 key points were established in a bbox to represent the skeleton of the character. Figure 14a shows an image file of a character and a script

file containing the skeleton of the character. Figure 14b shows the process of generating a posture by using an image file of a 3D animation character and a script file.



(**a**)



(**b**)

Figure 14. Script execution result, they should be listed as: (**a**) one posture image file and one script file with 29 key point information for human model; (**b**) one posture image file and one script file with 29 key point information for 3D animation character.

In Figure 15, numbers from 0 to 28 represent the location of the skeleton of a character. Numbers in ascending order refer to the middle buttocks, left buttock, left knee, left ankle, left big toe, left little toe, right buttock, right knee, right ankle, right big toe, right little toe, waist, chest, neck, left shoulder, left elbow, left wrist, left thumb, left ring finger, right shoulder, right elbow, right wrist, right thumb, right ring finger, nose, left eye, right eye, left ear, and right ear, respectively.

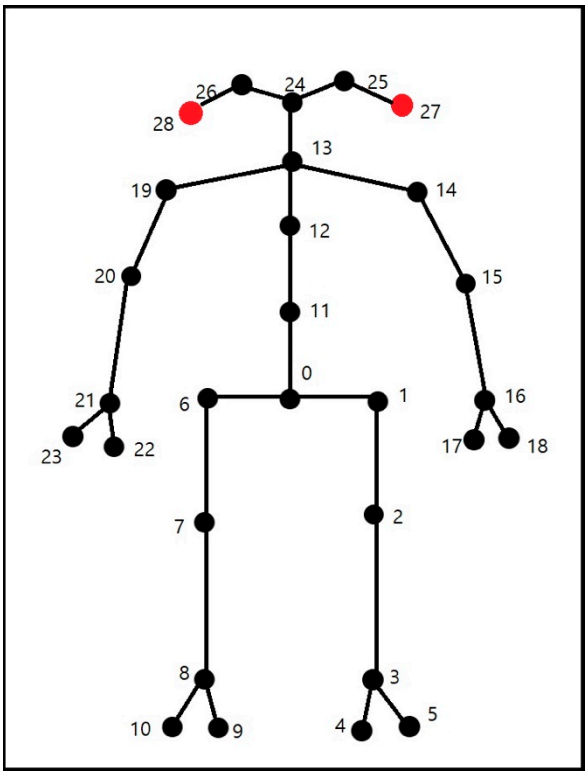


Figure 15. 29 Key points to express the human model.

Figure 16 shows a description of a posture stored in a script file. This description includes information on a posture, such as the category, 29 key points, licenses, coordinates of key points, 3D, videos, the name of an image file, images, and actors. This paper limited the maximum number of postures to be generated to 100 and thus included 100 script files (*.json).

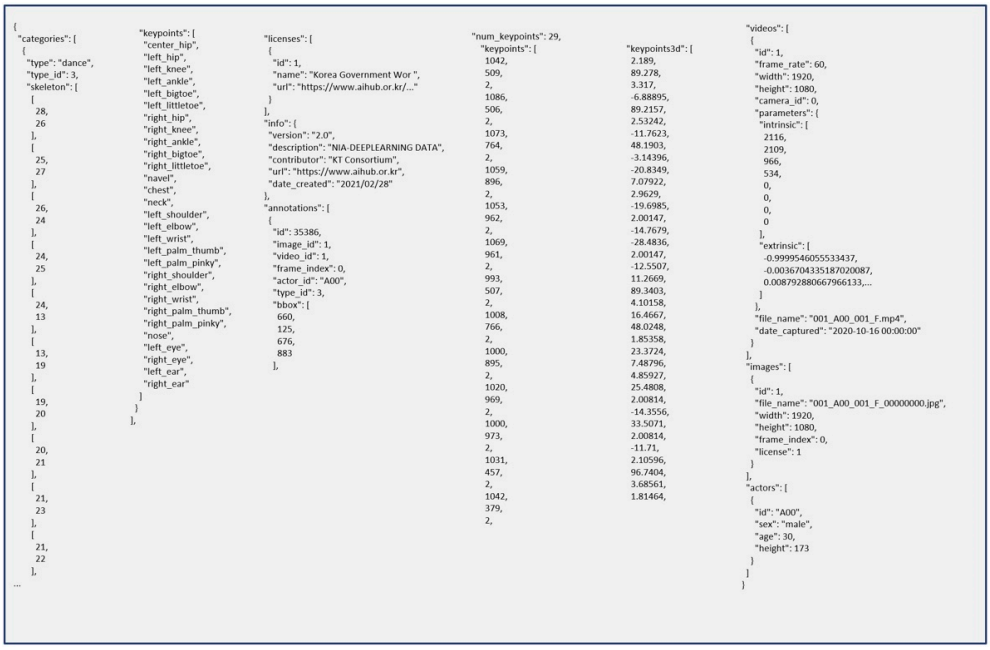


Figure 16. 29 Key point information written to one script file.

4.2. Experimental Results

Figure 17 shows several postures that the proposed system derived by analyzing a video of a dancing animation character, which was received as input data, in an experiment.

Everglow is the name of the K-POP dance group that released Adios, a song that was used in the dance video, and dance movements presented in this video. This animation video is 4 min and 38 s long and 64 MB in size.

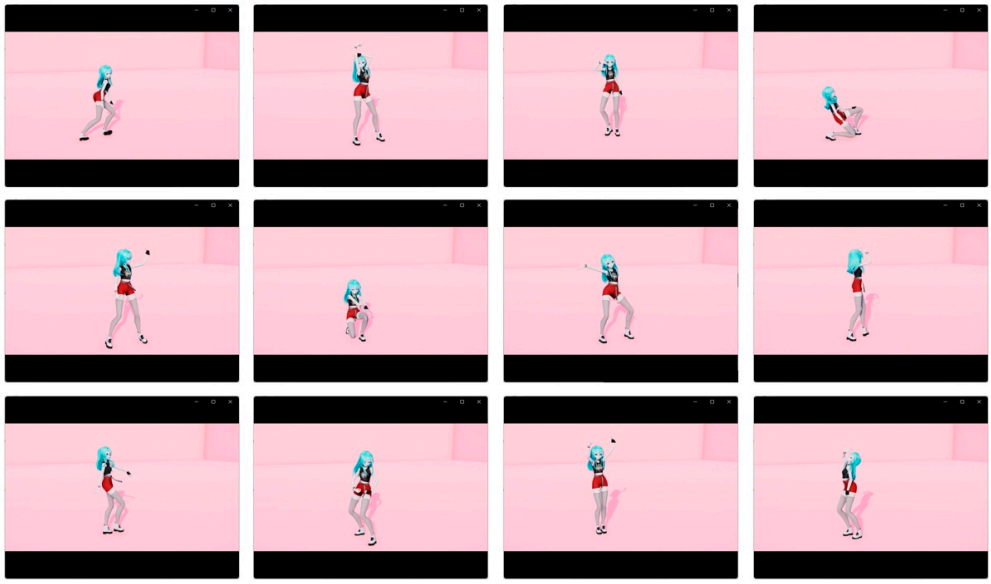


Figure 17. Representative dance video frames of the song, "Adios", by K-POP dance team "Everglow".

A creative choreography creator or producer, who had experience in using creative choreography generation systems, was selected as the target user of the proposed system. Figure 18 shows an example of words that the user input to generate choreography. To establish words input by the user as gesture words, data of these words were transferred to the gesture discriminator.

Hi Greeting Situation Check Meaning Three Excuse Enough Possible Eye You I First Main Need Prank Day Mouth Eye There

Figure 18. Words entered by the user in the gesture discriminator.

Figure 19 shows another example of words that the user input. The word controller determines multiple words to be used, and the dance manager detects gestures based on the data transferred.

I You Day Greeting Eye Mouth Ear Main I Prank You Need Enough Situation Check

Figure 19. Data passed as a dance manager.

Figure 20 shows a process where the posture discriminator stored 100 postures, which were extracted from the dance video for the song Adios input as the source video, as image files. This tool generates script files containing 29 key points on each image.

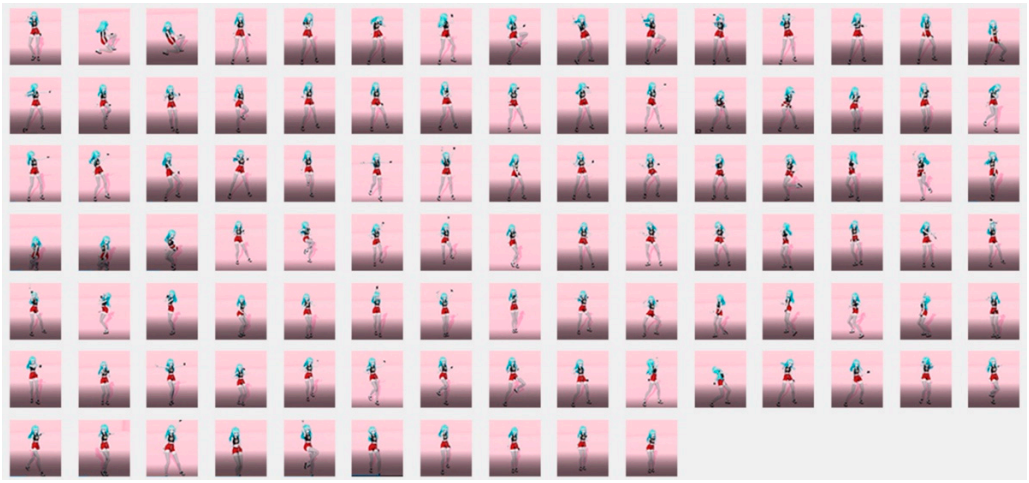


Figure 20. One hundred posture images with 29 key points.

Figure 21 shows samples of a few enlarged posture images among 100 posture images, each of which contain 29 key points.

Figure 21. Samples of 100 posture images with 29 key points.

Figure 22 shows samples of script files that store information on 29 key points established for each posture.

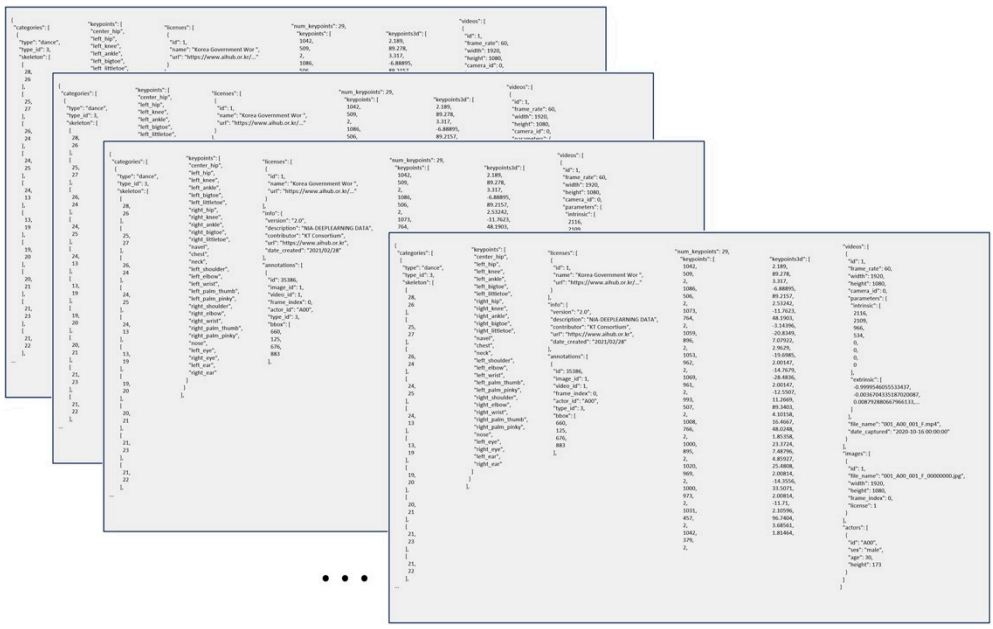


Figure 22. One hundred script files with 29 key points.

Figure 23 shows the process where the posture detector represented 100 dance posture images, each of which was represented by 29 key points. At this time, the user determined the number of postures to be grouped to represent a gesture. The posture detector calculated the number of gestures to be generated based on the set number of postures to be grouped. For example, if a group of postures is established to include five postures, 20 gestures will be generated.

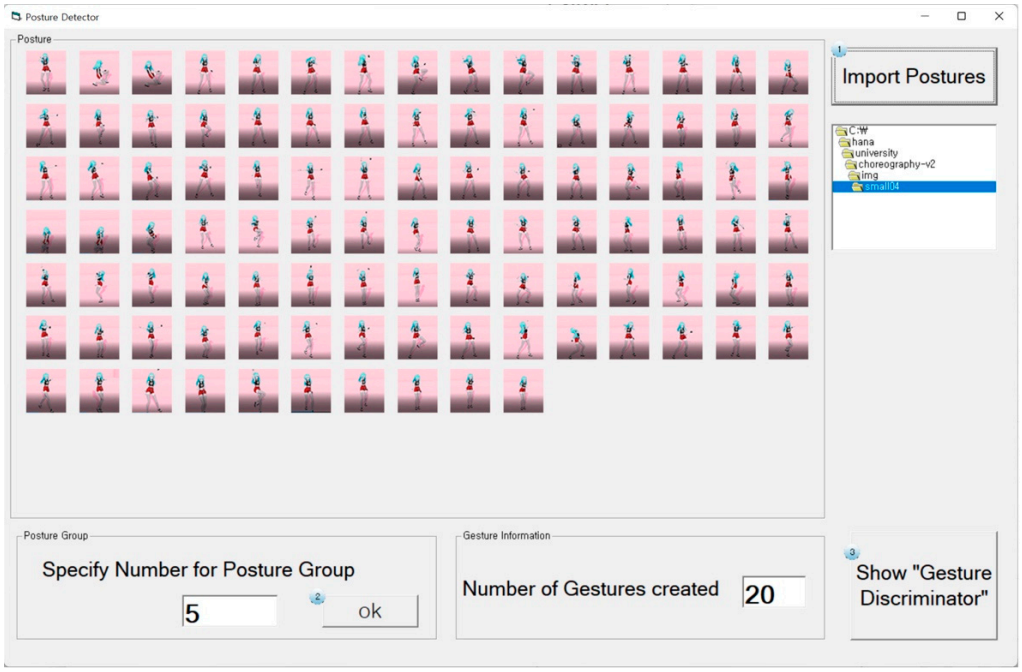


Figure 23. Posture discriminator to process the 100 posture images.

Information on gestures generated by the posture detector was transferred to the gesture discriminator. Figure 24 shows information transferred in the form of [a gesture number and multiple posture numbers].

[1 1 2 3 4 5] [2 6 7 8 9 10] [3 11 12 13 14 15] [4 16 17 18 19 20] [5 21 22 23 24 25] [6 26 27 28 29 30] [7 31 32 33 34 35] [8 36 37 38 39 40] [9 41 42 43 44 45] [10 46 47 48 49 50] [11 51 52 53 54 55] [12 56 57 58 59 60] [13 61 62 63 64 65] [14 66 67 68 69 70] [15 71 72 73 74 75] [16 76 77 78 79 80] [17 81 82 83 84 85] [18 86 87 88 89 90] [19 91 92 93 94 95] [20 96 97 98 99 100]

Figure 24. The words entered by the user are converted into the corresponding gesture numbers.

Figure 25 shows a process where the gesture discriminator determined a group of postures as a gesture. The gesture designation process is divided into a case where a gesture word was determined and a case where a gesture word was not determined. In the case where a gesture word was not determined, the gesture discriminator indicated a gesture number and a posture group generated by the posture detector. In the case where a gesture word was determined, the gesture word previously input by the user was applied to the corresponding gesture.

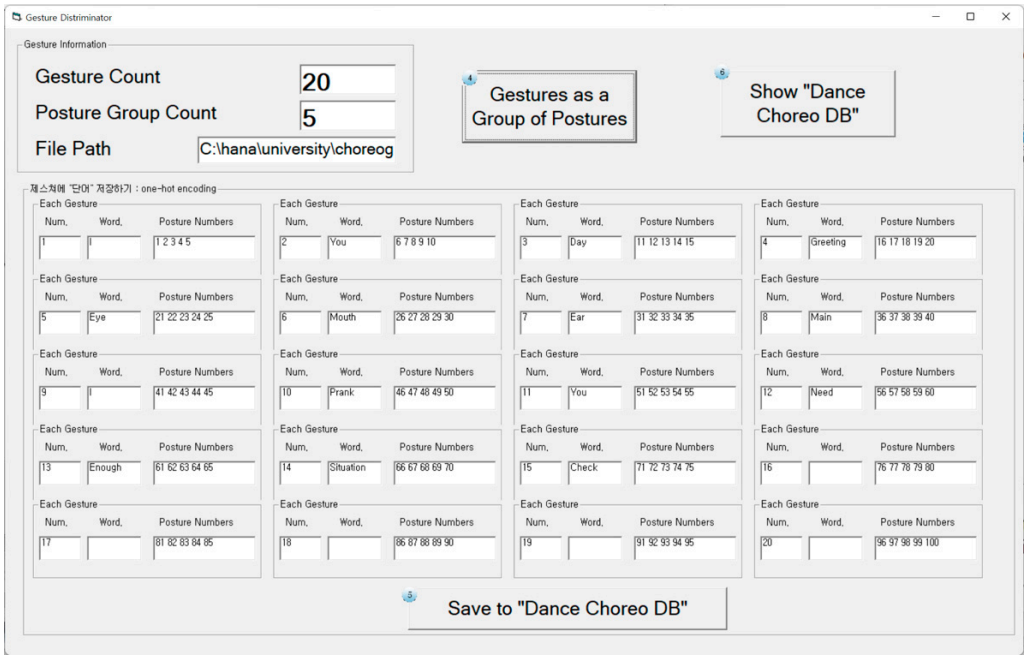


Figure 25. Gesture discriminator screenshot.

As shown in Figure 26, the gesture discriminator transferred information to the Dance-Choreo-DB, which stores the transferred information, or outputs it. The output data was based on the form of [a gesture number, a gesture word, multiple posture numbers]. The output data was formed based on the K-POP dance video for Everglow's song Adios used in this paper.

[1 Hi 1 2 3 4 5] [2 Greeting 6 7 8 9 10] [3 Situation 11 12 13 14 15] [4 Check 16 17 18 19 20] [5 Meaning 21 22 23 24 25] [6 Three 26 27 28 29 30] [7 Excuse 31 32 33 34 35] [8 Enough 36 37 38 39 40] [9 Possible 41 42 43 44 45] [10 Eye 46 47 48 49 50] [11 You 51 52 53 54 55] [12 I 56 57 58 59 60] [13 First 61 62 63 64 65] [14 Main 66 67 68 69 70] [15 Need 71 72 73 74 75] [16 Prank 76 77 78 79 80] [17 Day 81 82 83 84 85] [18 Mouth 86 87 88 89 90] [19 Ear 91 92 93 94 95] [20 There 96 97 98 99 100]

Figure 26. Dataset stored in Dance-Choreo-DB.

Figure 27 shows a file generated based on data transferred from the gesture discriminator and stored in the Dance-Choreo-DB.

The Dance-Choreo-DB stores gestures that contain designated gesture words and those which do not contain designated gesture words. This DB stores gestures without words designated by the gesture discriminator, as shown in Figure 28a, and those with words designated by the user, as shown in Figure 28b.

The word controller extracts words from words input by the user by classifying words based on spacing. Figure 29 shows elements of the word controller. Extracted words are designated as gestures. The word controller extracted lyrics from the K-POP song and detected, and designated functions for gestures and random words input by the user were applied. Based on the applied data, gestures stored in the Dance-Choreo-DB were detected. Figure 29a shows the processes where the word controller extracted lyrics and where the user stored words. As for input data, the word controller extracted the following 20 words from the lyrics of the K-POP song Adios released by Everglow: Hi, Greeting, Situation, Check, Meaning, Three, Excuse, Enough, Possible, Eye, You, I, First, Main, Need, Prank,

Day, Mouth, Ear, and There. Figure 29b shows a process where the user inputs random words for a creative purpose on the word controller. For example, the user formed a new word arrangement by combining previously-stored words as follows: [I You Day Greeting Eye Mouth Ear Main I Prank You Need Enough Situation Check Greeting Hi].

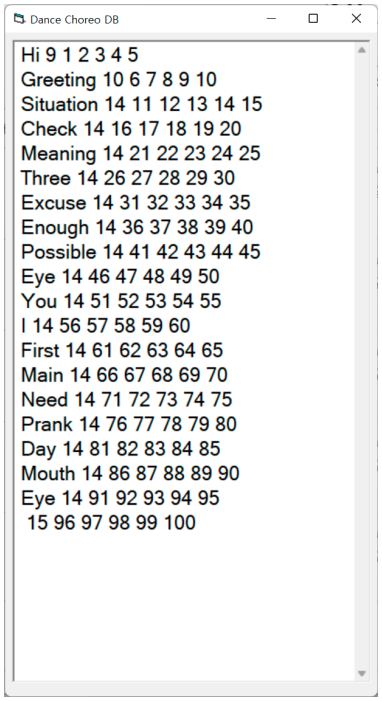


Figure 27. Dataset of Dance-Choreo-DB.

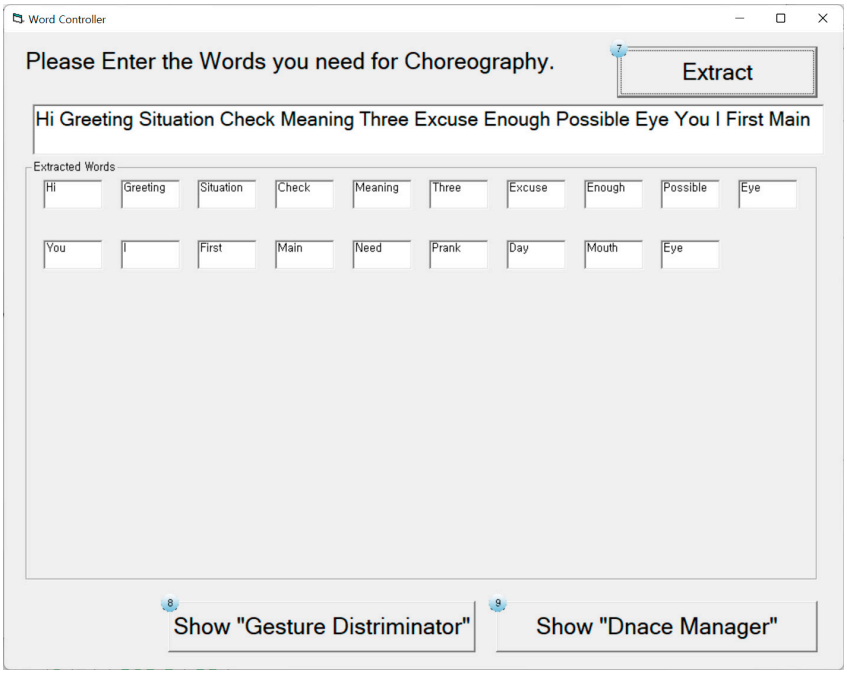
[1 1 2 3 4 5] [2 6 7 8 9 10] [3 11 12 13 14 15] [4 16 17 18 19 20] [5 21 22 23 24 25] [6 26 27 28 29 30] [7 31 32 33 34 35] [8 36 37 38 39 40] [9 41 42 43 44 45] [10 46 47 48 49 50] [11 51 52 53 54 55] [12 56 57 58 59 60] [13 61 62 63 64 65] [14 66 67 68 69 70] [15 71 72 73 74 75] [16 76 77 78 79 80] [17 81 82 83 84 85] [18 86 87 88 89 90] [19

(**a**)

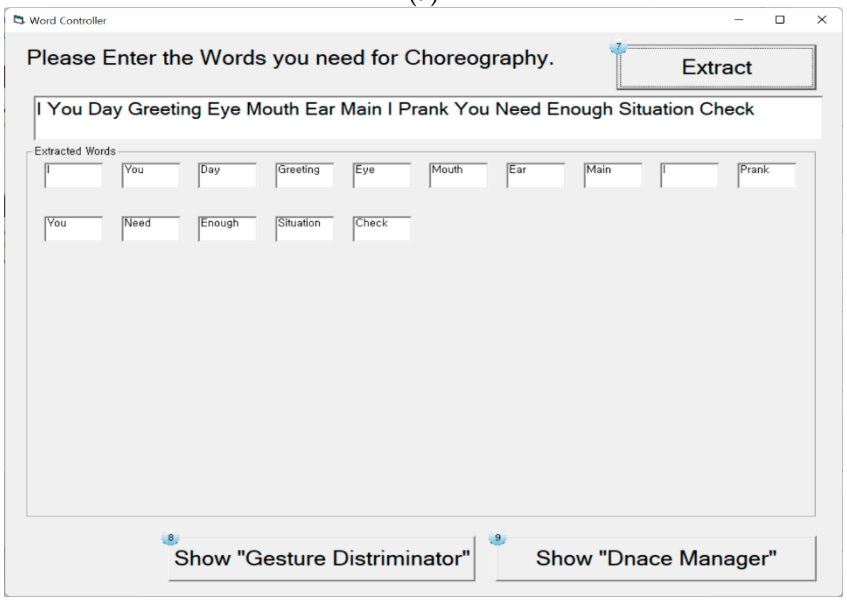
[1 Hi 1 2 3 4 5] [2 Greeting 6 7 8 9 10] [3 Situation 11 12 13 14 15] [4 Check 16 17 18 19 20] [5 Meaning 21 22 23 24 25] [6 Three 26 27 28 29 30] [7 Excuse 31 32 33 34 35] [8 Enough 36 37 38 39 40] [9 Possible 41 42 43 44 45] [10 Eye 46 47 48 49 50] [11 You 51 52 53 54 55] [12 I 56 57 58 59 60] [13 First 61 62 63 64 65] [14 Main 66 67 68 69 70] [15 Need 71 72 73 74 75] [16 Prank 76 77 78 79 80] [17 Day 81 82 83 84 85] [18 Mouth 86 87 88 89 90] [19 Ear 91 92 93 94 95] [20 There 96 97 98 99 100]

(**b**)

Figure 28. Two cases stored in the Dance-Choreo-DB, they should be listed as: (**a**) output data: gestures without words; (**b**) output data: gestures with words.



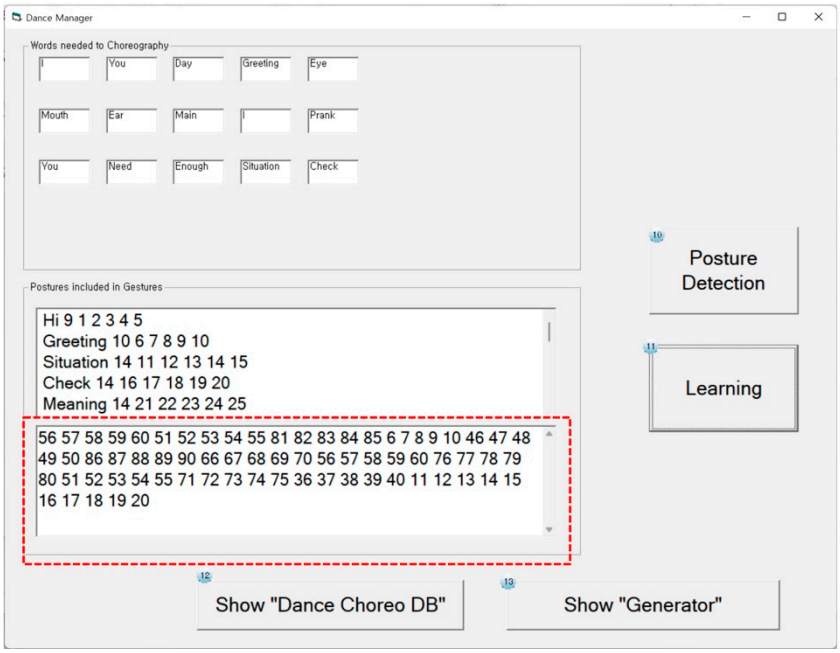
(**a**)



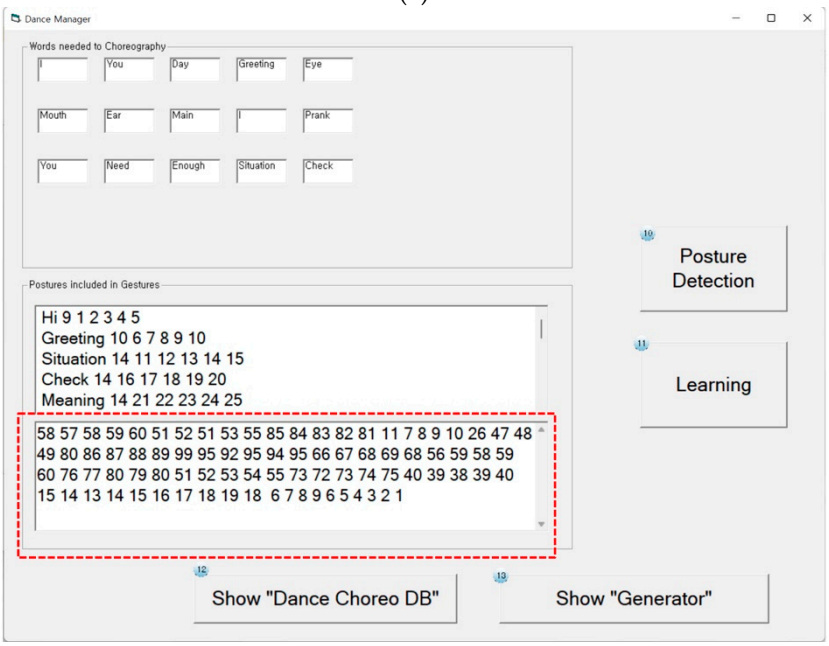
(**b**)

Figure 29. Word controller screenshot, they should be listed as: (**a**) process of passing the user-specified word to gesture discriminator; (**b**) the designated word process for dance choreography generation.

The dance manager searched words input by the word controller in the Dance-Choreo-DB and detected gestures based on the search result. Figure 30 shows the processes of the dance manager. The proposed system learned the level and preference of gesture by using the Bi-LSTM network. The word controller indicated necessary input words for dance sequence creation: [I You Day Greeting Eye Mouth EarMain I Prank You Need Enough Situation Check Greeting Hi].



(**a**)



(**b**)

Figure 30. Dance choreography screenshot, they should be listed as: (**a**) before Bi-LSTM learning; (**b**) learning through Bi-LSTM for difficulty level of posture, preference, and level of dance.

The dance manager transferred creative choreography generated before the learning process and after the learning process to the dance choreographer. Figure 31a shows a type of gesture data stored before the learning process, and Figure 31b shows a type of gesture data stored after the learning process.

[12 I 56 57 58 59 60] [11 You 51 52 53 54 55] [17 Day 81 82 83 84 85] [2 Greeting 6 7 8 9 10] [10 Eye 46 47 48 49 50] [18 Mouth 86 87 88 89 90] [19 Ear 91 92 93 94 95] [14 Main 66 67 68 69 70] [12 I 56 57 58 59 60] [16 Prank 76 77 78 79 80] [11 You 51 52 53 54 55] [15 Need 71 72 73 74 75] [8 Enough 36 37 38 39 40] [3 Situation 11 12 13 14 15] [4 Check 16 17 18 19 20] [2 Greeting 6 7 8 9 10] [1 Hi 1 2 3 4 5]

(**a**)

[12 I 58 57 58 59 60] [11 You 51 52 51 53 55] [17 Day 85 84 83 82 81] [2 Greeting 11 7 8 9 10] [10 Eye 26 47 48 49 80] [18 Mouth 86 87 88 89 99] [19 Ear 95 92 95 94 95] [14 Main 66 67 68 69 68] [12 I 56 59 58 59 60] [16 Prank 76 77 80 79 80] [11 You 51 52 53 54 55] [15 Need 73 72 73 74 75] [8 Enough 40 39 38 39 40] [3 Situation 15 14 13 14 15] [4 Check 16 17 18 19 18] [2 Greeting 6 7 8 9 6] [1 Hi 5 4 3 2 1]

(**b**)

Figure 31. Two examples of applying learning in dance manager, they should be listed as: (**a**) output data: choreography words before learning; (**b**) output data: choreography words after learning.

The dance choreographer received creative choreography generated after the learning process from the dance manager. Then, it arranged posture numbers that belonged to the gesture word, which was learned by the proposed system, and which matched the combination of words input by the user. Finally, it displayed postures images, which corresponded to the detected posture numbers, at the interval of a second. Figure 32 shows images of the creative choreography result derived from the proposed system. The first image is a posture image detected from the dance video of an AI HUB's dancing character or that of an animation character designed in this paper. The second image shows 29 key points that represent the body of a dancing character on AI HUB or an animated character designed in this paper.

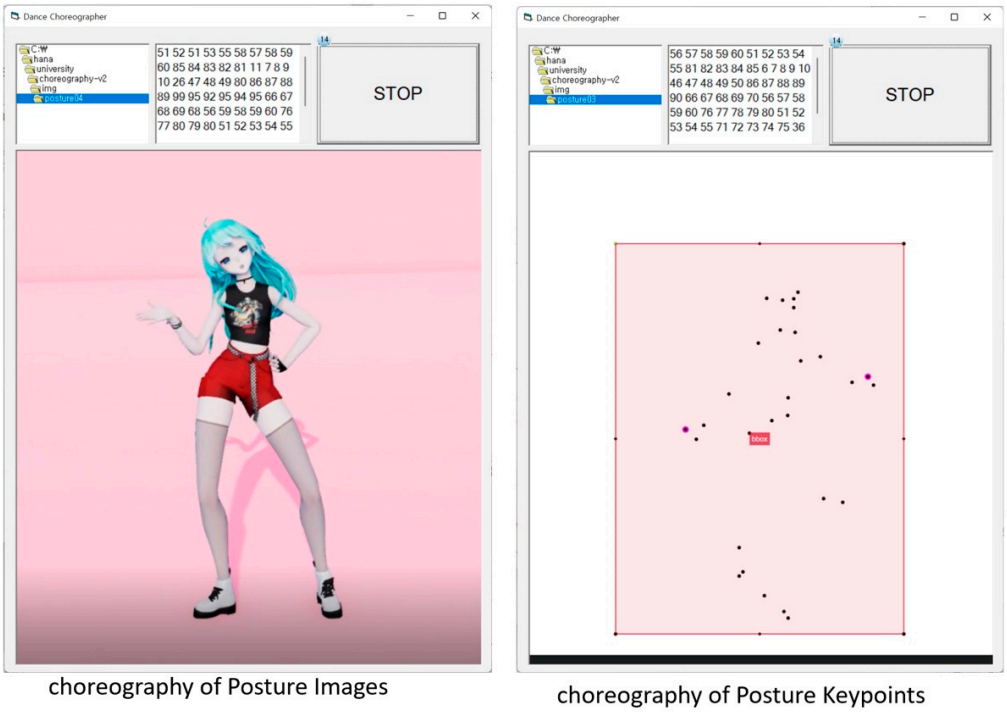


Figure 32. Dance choreography screenshot.

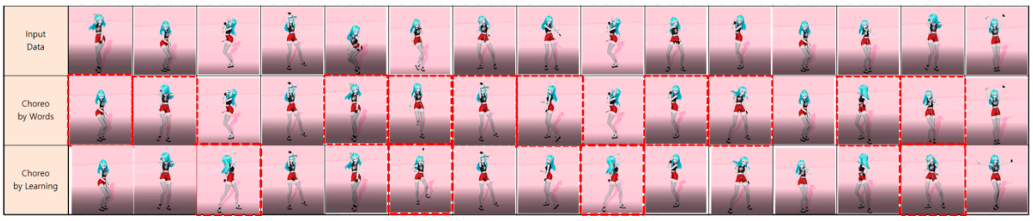
Figure 33 shows a list of postures of choreography that came out through the creative choreography generation system. Creative choreography finally derived by the dance choreographer can be applied to other target applications. The corresponding data consist of posture numbers, including 29 key points that represent the body of a posture.

[58 57 58 59 60] [51 52 51 53 55] [85 84 83 82 81] [11 7 8 9 10] [26 47 48 49 80] [86 87 88 89 99] [95 92 95 94 95] [66 67 68 69 68] [56 59 58 59 60] [76 77 80 79 80] [51 52 53 54 55] [73 72 73 74 75] [40 39 38 39 40] [15 14 13 14 15] [16 17 18 19 18] [6 7 8 9 6] [5 4 3 2 1]

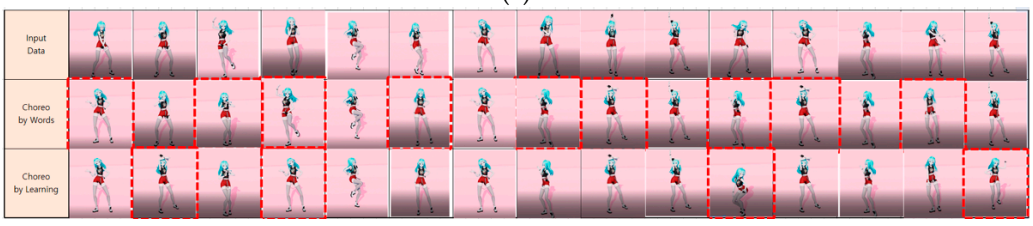
Figure 33. Trained posture group data executed by the dance choreographer.

4.3. Experimental Results

For the experiment, 100 postures are saved as input data, including images and 29 key points. One gesture is created in a group of five, and choreography is created with a combination of words input by the user. One posture is designated as 1 s, and choreography creates 100 postures or 20 gestures based on 100 s. Choreography is created through the data input by the user, which is called input data. At this time, a choreography is created based on the input data. When the same word is repeated several times, the same motion is repeated several times to create the choreography. In this creation data, a certain posture is converted into another posture in consideration of the characteristics of originality. These data are called creative data. The choreography created in this way is the learned creative choreography. The following figure compares the images of input data, creative data, and learning data. This study experiments with 10 choreography samples among recent choreography. The experimental method is divided into two main parts. The first is to create several choreographies through the choreography of one song, and the second is to create and experiment with each choreography in several songs. Figure 34 shows the results of creating two or more creative choreographies through BlackPink-PinkVenom choreography. Figure 35 shows the choreography results generated through two samples: NewJeans-hype boy and LESSERAFIM-antifragile choreography.



(**a**)



(**b**)

Figure 34. Creative choreography created based on BlackPink-PinkVenom choreography: Input data stored based on K-POP choreography, creative data combined by user definition, learning data that has gone through the learning process: (**a**) creative product, (**b**) another creation product.

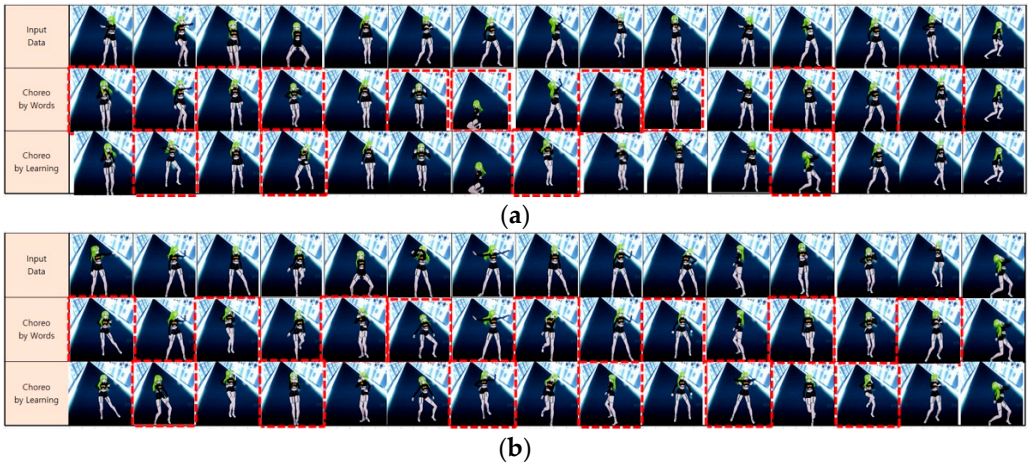


Figure 35. Creative choreography created based on K-POP choreography: (**a**) NewJeans-hype boy creative product (**b**), LESSERAFIM-antifragile creative product.

In the case of creating multiple choreography through one song, the process of creating multiple pieces through multiple songs is compared. Looking at the results of the graph distribution of the input data, the index changes are jagged and irregular. In this case, it may be creative or original, but the probability of occurrence between posture and posture is low, and the connection is insufficient. It can be seen that the creative data adjusts the trend of change based on the input data, reduces the curvature through the learning data, and creates a popular expression method. Figure 36a,b are the graphs of the two creations that will come out through the BlackPink-PinkVenom choreography, and Figure 36c is the graph of the choreography created with the NewJeans-hype boy song.

The hyper parameters set when learning Bi-LSTM are shown in Table 1 and each function is explained. sequence_length designates one posture value as 5, and number_class is the total number of postures. hidden_size is the vector size when learning with the LSTM hidden set in the neural network, number_layer is the number of LSTM layers, and set to 2 to configure Bi-LSTM. When batch_size data are processed, 128 are processed together at a time, and the difference is averaged to update the model. lt sets the weight when training the update, and the remaining variables are used when optimizing the model update.

The proposed system generated a posture by using an image obtained from a dance video and 29 key points and represented creative choreography based on an arrangement of postures that belonged to gestures that it learned. Postures including key points were displayed in three types. The first type shows the posture of an AI HUB's dancing character, including key points. The second type shows a posture of a 3D animation character designed in this paper. The third type shows an image of the skeleton extracted from the AI HUB's dancing character or 3D animation character designed in this paper. Based on the entered choreography as the correct answer, the choreography is created by recombining the generated gestures. Change the existing choreography by learning the relationship between posture and posture in the posture group that composes the gesture. The evaluation criterion is the probability of creating a new choreography creatively through the input choreography, but at the same time, learning the connection between poses and making the next move. The three cases created through this process are shown in Figure 37.

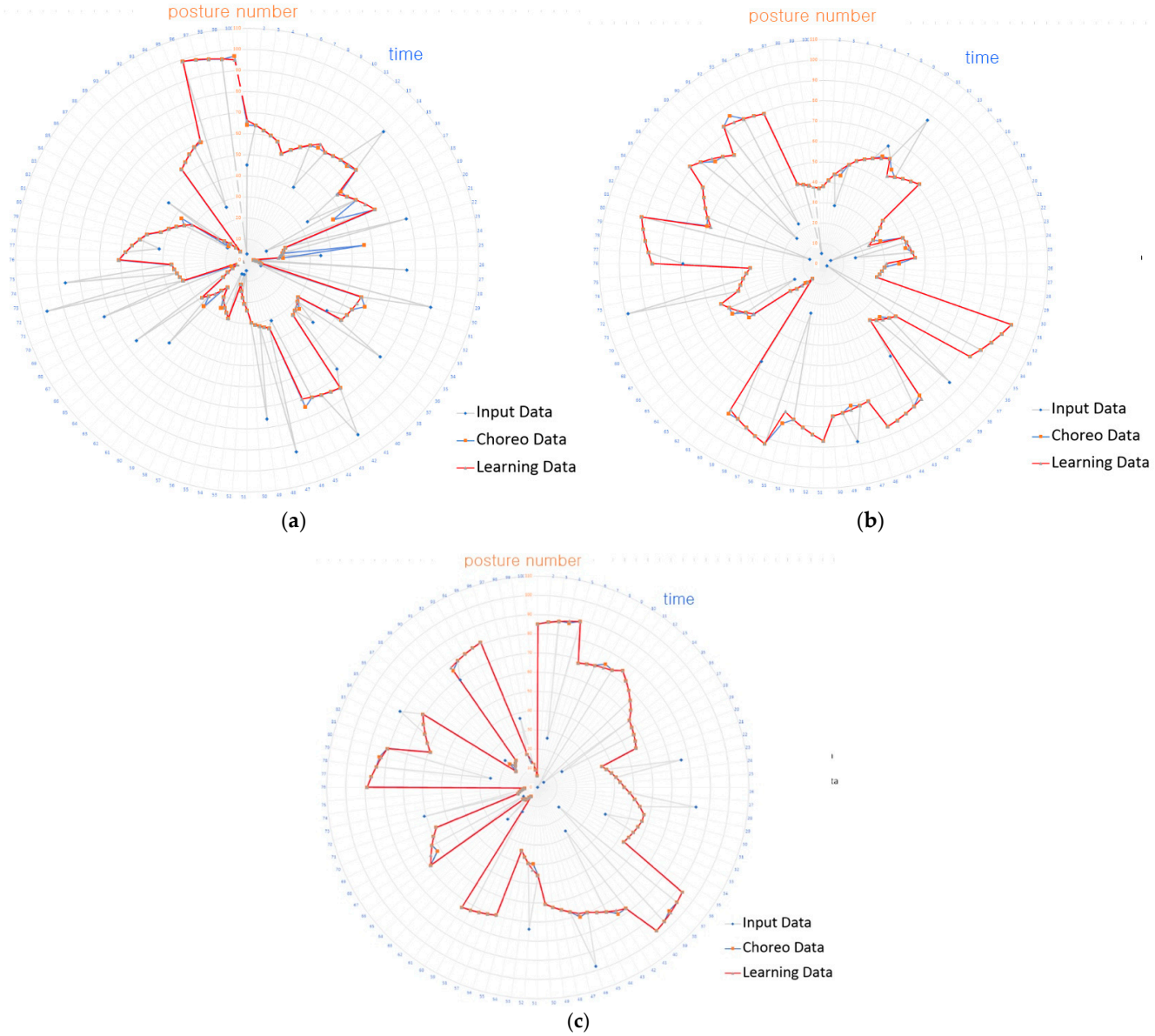


Figure 36. Issues of input data, creation data, and storage data: (**a**,**b**) BlackPink-PinkVenom creators, (**c**) NewJeans-hype boy another creator.

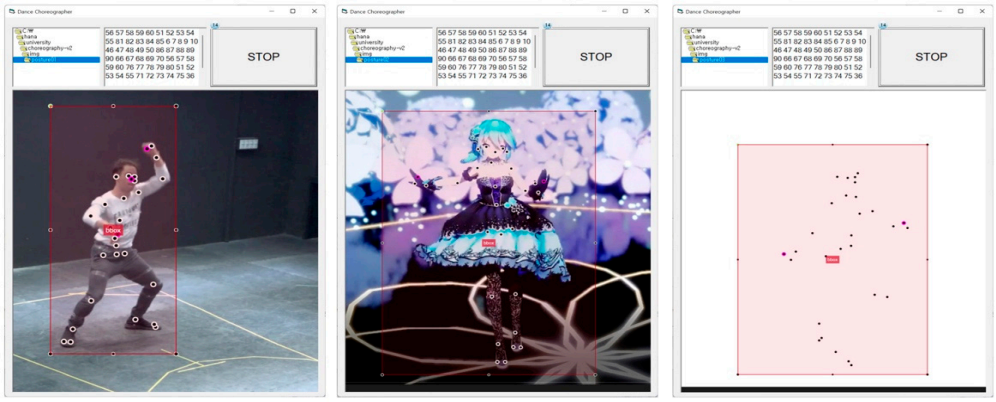


Figure 37. 3 Types of dance choreography results; human model, 3d character, human skeleton.

Table 1. Hyper Parameter: Learning parameter settings for setting up Bi-LSTM.

sequence_length	5	This is the length of the input sequences that the network will operate on. In other words, if the network is processing text data, it will work with sequences of 5 words at a time.
number_class	100	This is the number of classes that the network is attempting to classify the input into. For example, if the network is trying to classify images, there might be 100 different types of objects that the network is trying to identify.
hidden_size	512	This is the number of neurons in each hidden layer of the network. The higher the hidden size, the more complex the network can be.
number_layer	2	This is the number of layers in the network. The more layers, the more complex the network can be.
batch_size	128	This is the number of samples that the network processes at a time during training. A higher batch size can lead to more efficient training.
dropout	0.1	This is the dropout rate, which is the probability that each neuron will be randomly dropped out during training. Dropout is a regularization technique that can prevent overfitting.
Lr	1×10^{-4}	This is the learning rate, which determines how quickly the network will adjust its weights during training. A higher learning rate can lead to faster training, but can also cause the network to overshoot optimal weights.
adam_weight_decay	0.01	This is the weight decay parameter used by the Adam optimizer. Weight decay is another regularization technique that can prevent overfitting.
adam_beta1	0.9	This is the beta1 parameter used by the Adam optimizer. It determines the decay rate for the moving average of the gradient.
adam_beta2	0.98	This is the beta2 parameter used by the Adam optimizer. It determines the decay rate for the moving average of the squared gradient.

5. Conclusions

This paper proposed a technique and a system based on the technique of automatically generating continuous K-POP creative choreography by deriving postures and gestures based on K-POP dance videos collected in advance and analyzing them with bidirectional LSTM (Bi-LSTM). Based on the K-POP dance video, the posture was extracted and a series of postures were defined as gestures. Each gesture was expressed in words to create the entire choreography. This series of processes was divided into units to form a K-POP choreography creation system. The proposed system consists of a Dance-Choreo-Input-Unit, a Dance-Choreo-Generation-Unit, and a Dance-Choreo-Output-Unit.

The Dance-Choreo-Input-Unit includes the source movie and the user. First, the source movie is an input dance video for defining the dance posture. Second, the user is the user who wants to dance the creation and is the subject who inputs, edits, and stores the creative content. The Dance-Choreo-Generation-Unit takes the dance video input and creates a new choreography with a 3D dance character in three stages. First, The Dance-Choreo-Gesture-Generation-Stage extracts the postures from the source movie, assigns them numbers in a series of sequences, and groups multiple postures into a single gesture. The gestures you create are recorded in the Dance-Choreo-Database. Second, the Dance-Choreo-Control-Stage is divided into a word controller in which the user enters multiple words, a creative dance manager that matches gestures with words, and a dance creator that creates choreography with sentences given by the user. Finally, the Dance-Choreo-Learning-Stage operates the Bi-LSTM Network to analyze the user's intent and provide customized choreography creation. It analyzes and rearranges the correlation of the gestures selected in Dance-Choreo-Control-Stage. The Dance-Choreo-Output-Unit is the process of utilizing the choreography generated by learning and applying it in various ways. The research can be used not only for research, but also for a wide range of purposes, from the entertainment business to creative choreography production. In the Dance-Choreo-Output-Unit, the created choreography is input to the Target-Choreo-Database with various Other-Target-Applications. Key point-based creative choreography can be used as an image translation model between target images.

In order to dance choreography, hardware costs such as motion capture equipment that extracts motion through the human skeleton or expensive cameras for filming are usually expensive. In addition, software costs such as reliance on dance experts and large

databases are also required to extract dance motions. Due to the difficulty in collecting these dance datasets and hardware limitations, dance AI research has been limited. As a result, research is actively being carried out, but it does not achieve much results due to the constraints of the pre-production case. From this point of view, the goal of this paper is to propose effective algorithms and a system based on low-cost datasets to promote the development of artificial intelligence. Based on various dance videos existing online, creative choreography is created using movements extracted through shooting video characters. As a result, research and development necessary for direct choreography creation artificial intelligence will be carried out in earnest. This paper proposed a system for generating creative choreography under conditions that minimize the cost of hardware and software in dataset collection. These results will serve as a starting point for various research such as the relationship between music and dance and the learning of created choreography, which will contribute to the expansion and development of the K-POP market through agreements with entertainment for creative choreography.

Future research tasks will expand to research that induces various creative choreography through one-to-many and many-to-many relationships by reducing the uncertainty of words and gestures. At this time, when selecting the next gesture from one gesture, it is determined through learning, and it is learned to connect the dance movements naturally. In addition, it will develop into a study that creates novel choreography by defining the relationship between originality and popularity by adjusting the genre and style suggestions through object setting according to the camera angle and consideration of the genre of dance.

Author Contributions: Conceptualization, H.Y. and Y.S.; methodology, H.Y. and Y.S.; software, H.Y. and Y.S.; validation, H.Y. and Y.S.; writing—original draft, H.Y.; writing—review and editing, Y.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Joonhee, K. Analysis of Research Trends on K-pop—Focused on Research Papers from 2011 to 2018. *Cult. Converg.* **2019**, *41*, 461–490.
2. Sol, K.; Jun, K. Exploring the Changes and Prospects of the Street Dance Environment caused by COVID-19. *J. Korean Soc. Sport. Sci.* **2021**, *30*, 257–272.
3. Li, R.; Yang, S.; Ross, D.A.; Kanazawa, A. Ai choreographer: Music Conditioned 3D Dance Generation with AIST++. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 10–17 October 2021; pp. 13401–13412.
4. Wang, Q.; Li, B.; Xiao, T.; Zhu, J.; Li, C.; Wong, D.F.; Chao, L.S. Learning deep transformer models for machine translation. *arXiv* **2019**, arXiv:1906.01787.
5. Umino, B.; Soga, A. Automatic Composition Software for Three Genres of Dance Using 3D Motion Data. In Proceedings of the 17th Generative Art Conference, Rome, Italy, 17–19 December 2014; pp. 79–90.
6. Chan, C.; Ginosar, S.; Zhou, T.; Efros, A. Everybody Dance Now. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 5933–5942.
7. Jordan, B.; Devasia, N.; Hong, J.; Williams, R.; Breazeal, C. A Toolkit for Creating (and Dancing) with AI. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtual, 2–9 February 2021; Volume 35, pp. 15551–15559.
8. Sminchisescu, C. 3D Human Motion Analysis in Monocular Video: Techniques and Challenges. In *Human Motion*; Springer: Dordrecht, The Netherlands, 2008; pp. 185–211.
9. Guo, H.; Sung, Y. Movement Estimation using Soft Sensors based on Bi-LSTM and Two-Layer LSTM for Human Motion Capture. *Sensors* **2020**, *20*, 1801. [CrossRef] [PubMed]
10. Yim, S. Suggestions for the Independent Body in the era of Artificial Intelligence Choreography. *Trans-* **2022**, *12*, 1–19.
11. Pan, Z.; Yu, W.; Yi, X.; Khan, A.; Yuan, F.; Zheng, Y. Recent Progress on Generative Adversarial Networks (GANs): A survey. *IEEE Access* **2019**, *7*, 36322–36333. [CrossRef]
12. Huang, Y.; Kaufmann, M.; Aksan, E.; Black, M.J.; Hilliges, O.; Pons-Moll, G. Deep Inertial Poser: Learning to Reconstruct Human Pose from Sparse Inertial Measurements in Real Time. *ACM Trans. Graph. (TOG)* **2018**, *37*, 1–15. [CrossRef]

13. Iashin, V.; Rahtu, E. Multi-modal Dense Video Captioning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 958–959.
14. Jain, A.; Tompson, J.; LeCun, Y.; Bregler, C. MoDeep: A Deep Learning Framework Using Motion Features for Human Pose Estimation. In Proceedings of the Computer Vision–ACCV 2014: 12th Asian Conference on Computer Vision, Singapore, 1–5 November 2014; Revised Selected Papers, Part II 12. Springer International Publishing: Berlin/Heidelberg, Germany, 2014; pp. 302–315.
15. Ionescu, C.; Papava, D.; Olaru, V.; Sminchisescu, C. Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *36*, 1325–1339. [CrossRef] [PubMed]
16. Jain, A.; Zamir, A.R.; Savarese, S.; Saxena, A. Deep Learning on Spatio-temporal Graphs. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 5308–5317.
17. Soomro, K.; Zamir, A.R.; Shah, M. UCF101: A dataset of 101 human actions classes from videos in the wild. *arXiv* **2012**, arXiv:1212.0402.
18. Jia, Y.; Johnson, M.; Macherey, W.; Weiss, R.J.; Cao, Y.; Chiu, C.C.; Wu, Y. Leveraging Weakly Supervised Data to Improve end-to-end Speech-to-text Translation. In Proceedings of the ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 7180–7184.
19. Khan, N.; Muhammad, K.; Hussain, T.; Nasir, M.; Munsif, M.; Imran, A.S.; Sajjad, M. An Adaptive Game-Based Learning Strategy for Children Road Safety Education and Practice in Virtual Space. *Sensors* **2021**, *21*, 3661. [CrossRef] [PubMed]
20. Tsuchida, S.; Fukayama, S.; Hamasaki, M.; Goto, M. AIST Dance Video Database: Multi-Genre, Multi-Dancer, and Multi-Camera Database for Dance Information Processing. *Int. Soc. Music Inf.* **2019**, *1*, 6.
21. Zhang, J.; Chen, Z.; Tao, D. Towards high performance human keypoint detection. *Int. J. Comput. Vis.* **2021**, *129*, 2639–2662. [CrossRef]