2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 25-29, 2020, Las Vegas, NV, USA (Virtual)

Infant abnormal behavior classification through weakly supervised learning

YEJIN LEE, Kyekyung Kim*, Jaehong Kim, Seung-Ik Lee

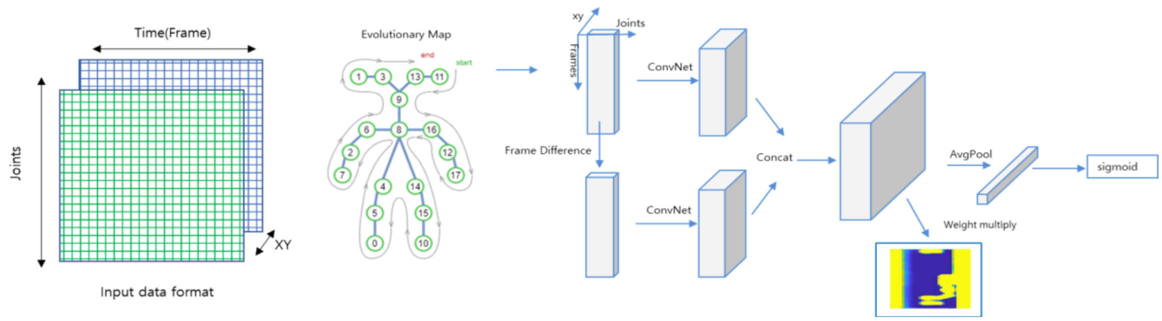

Figure 1. Infant abnormal behavior classification by fully convolutional network. On the back of the network, a class activation map can be created to check the information that has been weakly-supervised. The information is the abnormal presence of infants, video parts, and the body part where symptoms are expressed.

***Abstract*— Nowadays, caregivers are focusing on the health information of newborn infants. It is very important to discover a neurological disease in infants. Especially, Early diagnosis must be made, in the case of premature infants as the possibility of neurological diseases is high. There are many systems for monitoring infants, but few systems classify the status of infants through artificial intelligence analysis. In this paper, the normal and abnormal movements were classified to help diagnose neurological diseases of infants by using Fully Convolutional Network (FCN) according to the skeleton information of infants. Besides, only the labels of normal and abnormal can produce the following weakly-supervised three characteristics. 1) whether the infant's condition is normal or abnormal) 2) which segment of the video shows the infant's movement is abnormal, 3) which body part of the infant is found to be abnormal. The proposed network is simple and suitable for application to robot devices that observe infants in the home environment. This study was conducted using the skeleton database extracted from the video of infants.**

I. Introduction

Recently, the introduction of home robots, IoT robots, etc. has been on the rise as care for infants is needed in the home environment. As studies on infant monitoring have been conducted, infants can be observed 24 hours. However, most of the existing infant monitoring systems only have observational functions without intelligent analysis. Infants born prematurely are more likely to develop neurological diseases. Up to 18 percent of premature infants suffer from cerebral palsy (CP), and the total rate of neurological disorders is up to 45 percent. [1] Therefore, neurological disorders should be measured with close observation during the period of the newborn less than 28 weeks after childbirth. In this paper, introduces a deep learning method for intelligent analysis in determining neurological diseases in infants. Propose a system to assist the diagnosis of protectors and specialists.

II. Related Work

Research has continued to monitor the sleep or activity of vulnerable infants in the home environment. As infants need 24-hour observation, the need for care robots in home environments to help their guardians monitor is increasing. [2] proposed the results of tracking the movements and analyzing the behavior by attaching a sensor to the body of the infant. Although this played an important role in determining the abnormal of the infant, attaching a sensor to the infant's body involves a lot of discomforts. In [3], the movements of infants are tracked using three types of 3D information: x-coordinate, y-coordinate, and depth. This method requires a special camera with 3D support. Usually, the skeleton data used analysis methods such as RNN[4] or LSTM[5]. "Skeleton-Based Action Recognition With Convolutional Neural Networks" [6] proposed how to use CNN to analyze skeleton data. As a result, the number of ways to analyze Skeleton data using CNN has increased recently increasing.

In this paper, only x-coordinate, y-coordinate 2D information, which can be easily obtained with a regular camera, was used. Also, the structure of [6] was modified to suit the proposed study and the skeleton data was analyzed.

Y. Lee, K. Kim, J. Kim and S. Lee are with the Human Robot Interaction Research Lab, Electronics and Telecommunications Research Institute, Daejeon, Republic of Korea ({yejin1001, kyekyung, jhkim504, the_silee}@etri.re.kr).
* Corresponding author

III. THE INFANT EXPERIMENTAL DATASET

Infant data used in this study used 81 videos of 'normal' youtube skeleton data and 15 videos of 'abnormal' clinical skeleton data released in [7]. This is a set of data collected by the medical institution The Children's Hospital of Philadelphia (CHOP) with the criteria of Bayley Infant Neurodevelopmental Screener (BINS) [2]. In [2], infant data was refined through pre-processing after extracting an infant's point using OpenPose[8]. Experimental data are 2D skeleton data with 18 points per infant, consisting of 'normal' and 'abnormal' labels for neurological diseases.

IV. INFANT ABNORMAL BEHAVIOR CLASSIFICATION NETWORK

In this paper, we proposed a network to classify the actions of 2D Skeleton data by applying the CNN (Convolutional Neural Networks) method. The proposed network is as follows. The network of "Skeleton-Based Action Recognition With Convolutional Neural Networks"[6] was transformed as in Figure1. Instead of erasing the Skeleton Transformer block of [6], the Evolution map of [9] was adapted to the infant's joint as appropriate for this study and used as input data to enhance physical connectivity between the joints. It also eliminated the Fully Connected Network and replaced it with the Fully Convolution Network (FCN) in order not to lose the Local context information.

$J = (x, y)$ expressed as , in point coordinates. $S = (J_1, J_2, \dots, J_N)$. N is the joint number of a person's skeleton. The existing method used differential values from the very previous frame, but in the proposed network, the differential values from the five frames were used to take into account temporal connectivity. The last classification used flatten and changed the activation function softmax to sigmoid to distinguish between 'normal' and 'abnormal'.

A. Weakly supervised abnormal cam

In this paper, Class Activation Map (CAM) [10] was interpreted in a new way. The x-axis of the CAM is used as a body joint, and the y-axis is applied as a frame. Through this, we have detected a section of the infant's abnormal time and weakly supervised information. With only the label of normal and abnormal, the following three facts can be found. First, whether the infant is normal or abnormal. Second, what part of the body, not the joint, is abnormal. Thirdly, what part of the video is abnormal.

Figure2 shows the CAM of normal infants and the CAM of abnormal infants. In the figure, yellow is the part where the infant becomes active in the feature map of the network when it shows normal movement. Blue is the part that activates when an infant shows abnormal movement.

First, the x-axis of the CAM is the order of the evolutionary map. The rule of the joint area (joint number) means R to the right and L to the left. A 1d vector is generated in the order of each joint number and used as input data.

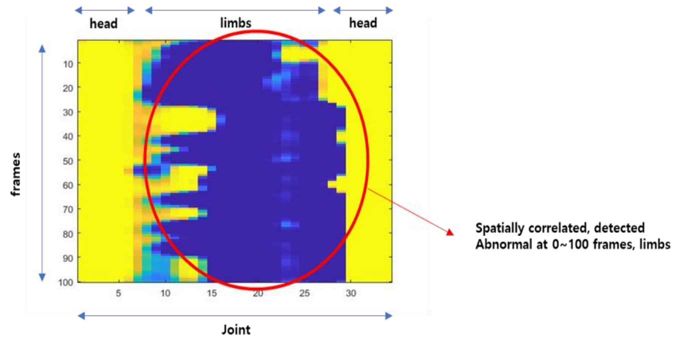

Figure 2. The x-axis is the infant body joint number, and the y-axis is the frame. The blue activated part is the limbs, and the infant's body is spatial and temporal correlated.

Evolutionary map order = {Rear (11), Reye (13), Nose (9), Neck (8), Rshoulder (12), Lelbow (12), Rwrist (12), Rshoulder (12), Neck (8), Rhip (14), Rankee (15), Rankle (10), Rknee (15), Rknee (15), Rknee (15), Rknee (15), Nose (9), Reye (13), Rear (11)}

When using the evolutionary map, the CAM can be checked by the unit of the body part that is spatially correlated, not by the joint of the infant's body. Figure 2. is the result of abnormal detection in the area of the limb of the infant for 0 to 100 frames.

V. EXPERIMENTS

The proposed network was used to classify Abnormal infants. After going through three convolution layers of the convolution block, create a CAM through weight multiply. The study was conducted using 45 normal learning data, 16 normal skeleton videos, 40 normal test data, and 14 normal skeleton videos. The Optimizer used Adam. The epochs proceeded 12 times. The learning rate is 1e-4.

A. Classification accuracy- comparing with baseline approaches

In this section, we studied and compared under the same conditions with the existing "CNN Skeleton Based Action Recognition" network presented in [6]. The classification of existing networks resulted in an accuracy of 88.89%. As a result of classifying into the proposed network, 94.44% accuracy was obtained, and 5.55% performance improvement was confirmed.

B. Weakly supervised CAM

According to the paper [11], the normal infant shows Fidgety Movements (FMs). FMs usually appear from about 9 weeks or 16 to 20 weeks. It includes the neck, shoulders, wrists, hips, ankles, etc., especially marked on the hips and ankles. Normal infants show FMs when they are awake and they show

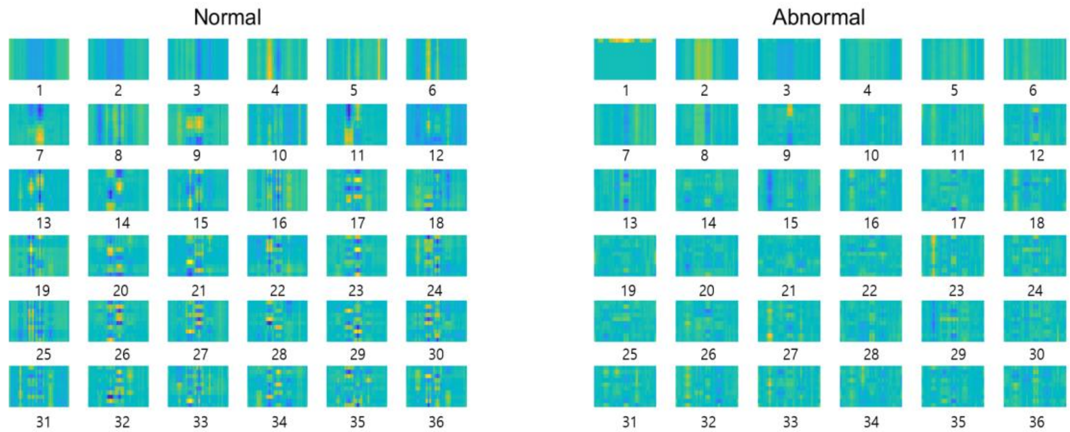

Figure 3. Data of the infant is applied with a singular value decomposition. The y-axis movement of the skeleton was considered. From about the seventh basis, the difference between normal and abnormal neurological acting appears.

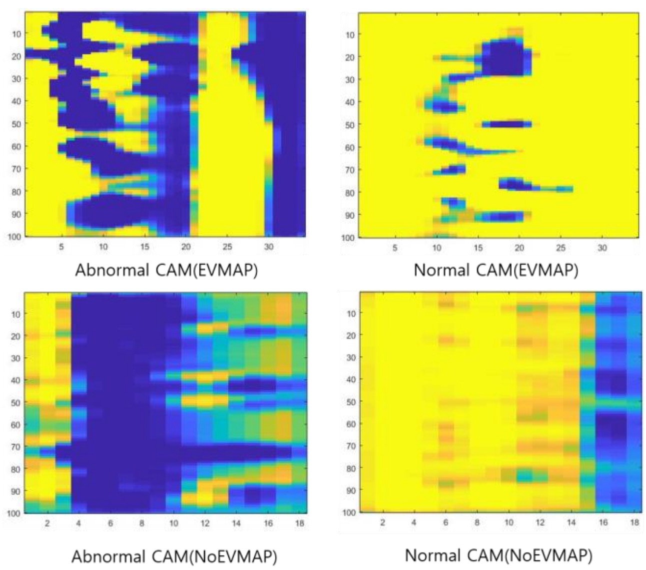

Figure 4. Class activation map of infant skeleton moving of normal and abnormal. The CAM is weakly-supervised information. The first row indicates the applied evolution map. The second row indicates no applied evolution map.

sustained soft movements. On the other hand, abnormal infants lack FMs. Or speed, large in size, or convulsions. In this paper, this data without separate physical guidance, the difference between normal and abnormal could be identified.

In the Figure 4, the evolution map, number 11 to 20 on the x-axis, is a joint that is the order of the right hip, right leg, left hip, and left leg. In the case of normal infants, movements are shown in the joints 11 to 20, and the movements are regular. In the case of abnormal infants, movements are either lacking movement or movements in areas other than joints 11 to 20 are detected. It can also be seen that the pattern of motion is convulsed.

In addition, the Figure3. applied the Singular Value Decomposition (SVD) method to all the data of the infant to identify the critical basis in the data. The x-axis is the joint number and the y-axis is the frame. The following figure takes into account the skeleton y-axis movement. Of the total 36 basis. Positive directional movement for yellow, and the negative directional movement for blue. In the case of normal in Figure3 motion was shown at the joints 11 to 20. However, in the case of abnormal, the pattern shows a lack of movement. This shows that normal infants show FMs, and abnormal infants lack the Fidgety movement. Through these facts, we can confirm that weakly supervised cam correctly classified infants.

VI. CONCLUSION

The proposed methods showed 5.55% action classification performance improvement compared to the existing network. Besides, according to the classification performance, we have identified information that has been weakly-supervised. You can find out the body part and the segment of the video where the problem with the abnormal infant was found. This means that you can get a lot of information without additional learning. Its benefits are suitable for hardware such as mobile robot devices. This can obtain multiple information using a very simple network.

This study can be used to help with quick diagnosis by closely monitoring infants and informing caregivers or doctors of possible infant risks. Rapid diagnosis can prevent greater neurological disorders from occurring during the growth period. In the future, the skeleton detector will be going to improve, further enhancing the skeleton reliability. It will also secure more infant data to better diagnose diseases.

ACKNOWLEDGMENT

This research was financially supported by the Ministry of Trade, Industry and Energy (MOTIE) and Korea Institute for Advancement of Technology (KIAT) through the National Innovation Cluster R&D program (P0006710_Development of Intelligent Platform for Living Safety Service in the Complex Living Space.

REFERENCES

[1] L.Adde, J. L. Helbostad, A. R. Jensenius, G. Taraldsen, & R. Støen, "Using computer-based video analysis in the study of fidgety movements". Early human development 85.9, 2009.

[2] L. Meinecke, N. Breitbach-Faller, C. Bartz, R. Damen, G. Rau, & C. Disselhorst-Klug, "Movement analysis in the early detection of newborns at risk for developing spasticity due to infantile cerebral palsy." Human movement science, 25.2, 2006.

[3] N. Hesse, A. S. Schröder, W. Müller-Felber,C. Bodensteiner, M. Arens, & U. G. Hofmann, "Body pose estimation in depth images for infant motion analysis." In 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2017.

[4] D. E. Rumelhart, G. E. Hinton, & R. J. Williams, "Learning representations by back-propagating errors." nature, 323(6088),1986.

[5] S. Hochreiter, & J. Schmidhuber, "Long short-term memory.", Neural computation, 9(8), 1997.

[6] C. Li, Q. Zhong, D. Xie, & S. Pu, "Skeleton-based action recognition with convolutional neural networks." In 2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), 2017.

[7] C. Chambers, N. Seethapathi, R. Saluja, H. Loeb, S. Pierce, D. Bogen, ... & K. P. Kording, "Computer vision to automatically assess infant neuromotor risk". BioRxiv, 756262, 2019.

[8] Z. Cao, G. Hidalgo, T. Simon, S. E. Wei, & 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.

[9] M. Liu, & J. Yuan, "Recognizing human actions as the evolution of pose estimation maps". In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition ,2018

[10] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, & A. Torralba, "Learning deep features for discriminative localization." In Proceedings of the IEEE conference on computer vision and pattern recognition, 2016.

[11] C. Einspieler, R. Peharz, & P. B. Marschik, "Fidgety movements–tiny in appearance, but huge in impact." Jornal de Pediatria, 92.3, 2016.