19th International Conference on Modeling & Applied Simulation
17th International Multidisciplinary Modeling & Simulation Multiconference

ISSN 2724-0037 ISBN 978-88-85741-48-5 © 2020 The Authors.
DOI: 10.46354/i3m.2020.mas.016

Automatic Modeling Method of Support Behavior for Work Support System Based on Variational Deep Enbedding–Generative Adversarial Networks

Kohjiro Hashimoto[1,*], Tadashi Miyosawa[1] and Tetsuyasu Yamada[1]

[1]Suwa University of Science, Japan, 5000-1 Toyohira, Chino, Nagano, 391-0292, Japan

[*]Corresponding author. Email address: k-hashimoto@rs.sus.ac.jp

Abstract

In Japan, small and medium-sized enterprises often handle people-centered work that cannot be automated. However, However, the shortage of work trainers has become serious problem with the declining birthrate and aging population, and an education system for novice workers is needed. By the way, Head Mounted Display(HMD) can present instruction information on the wearer's field of view in real time. Therefore, it is expected that even novice worker will be able to perform complicated tasks by designing appropriate instruction information to be presented. However, it is difficult for designer to design the support behavior of the system, when there are many patterns of work should be supported. Therefore, it is necessary to establish a technique can generate automatically the support behavior of the system. In this paper, we propose a deep learning model for automatically generating the support behavior of the system. Here, the work of repairing a laptop computer is taken as a working example. Then, it is assumed that the HMD presents the next work location as the instruction information. The proposed model can automatically generates a work process model and instruction information should be presented.

Keywords: Application of AR, Human support system, Deep learning

1. Introduction

A shortage of workers associated with declining birthrate and a growing proportion of elderly people becomes a serious social issue in Japan. In particular, medium-sized companies have taken on human-centric tasks that can not be mechanized, and have a huge challenge of shortage and fostering of engineer. Therefore, several studies on technology transfer skills and work support system have been conducted(Patrice Humblot-Nino, Oscar Sandoval-Gonzalez, Ignacio Herrera-Aguilar and Daniel Rangel-Penuelas 2017, Masao Sugi, Hisato Nakanishi, Masataka Nishino, Yusule Tamura, Tamio Arai, and Jun Ota 2010, Kazuyoshi Tagawa, Hiromi Tanaka, Masayu Komori, Yoshimasa Kurumi, and Shigehiro Morikawa, 2012).

We have developed a work support system by using HMD(Head Mounted Display). In the repair factory of notebook computer, the more than 90 kinds of notebook computer are dealt. However, worker must repair many kinds of notebook computers by himself according to a shortage of workers. Therefore, it is necessary to become skilled in the work in order to do work effectively. According to this reason, we have developed the work support system by using HMD for repair work of notebook computer(Kohjiro Hashimoto, Tadashi Miyosawa, and Mai Higuchi, 2018). Fig.1 shows an example of support behavior of our system. Our developed system presents work location and explanatory text of work procedure. As this figure shows, the HMD can realize a work support

timely because it can overlap the digital information with a field of wearer's view. Therefore, even if user is beginner at the task, the device can lead user in achievement of task. In fact, it has been reported that the work support system by using HMD improves the work efficiency compared on using of the paper-based manual(Reiko Takizuka, Haruhisa Kato, Hiromasa Yanagihara, and Masaru Sugano, 2016). And several support systems for piping construction(Koichiro Shiraishi, and Kohei Matsuo, 2015), and patrol of electrical power distribution equipment(Hirohiko Sagawa, Hiroto Nagayoshi, Harumi Kiyomizu, and Tsuneya Kurihara, 2015) have been studied as the application examples.

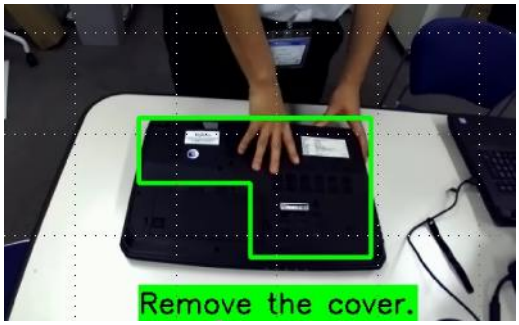

Figure 1. An example of support behavior of our developed system

However, when there are many work patterns should be supported, the many support behaviors must be designed by designer of the system. In fact, designer must spend much time and effort to design support behavior for work of repairing for more than 90 kinds of notebook computer. According to this reason, an automatic generation technique of support behavior of system is proposed in this paper.

According to the knowledge of previous study, it is necessary to construct recognition model of work object, work process and work locations. In general, it is necessary to decide the image feature and statistical model by heuristic way in order to construct these models. On the other hand, in our propose method, these models are generated automatically from work video based on a deep learning technique. Concretely, out proposed model structure is consisted of Single Shot MultiBox Detector(SSD) and Variational Deep Embedding - Generative Adversarial Networks(VaDE-GAN). These deep learning model with convolutional layer can extract the image features of target to recognize through learning. Therefore, detection of image features of work object, work process and work locations, and modeling them are realized from work video by using this characteristic of deep learning model.

In this paper, the effectiveness of the propose method for part removal work process of notebook computer is evaluated through the experiments.

2. Precondiction

Fig.2 shows the outline of the supposed system of work support in this paper. A worker works a given task wearing HMD. Work video can be obtained from fixed camera on the ceiling, and is sent to the computer. The computer recognizes the current and next work process, and decides support information. The decided support information is sent to HMD, and overlapped in a field of worker's view. In this paper, the function that performs processing from recognition of work process to determination of support information is called support behavior model. And this modeling method of support behavior is proposed.

This paper is target at removal work of parts in notebook computer. In this work, a worker intervention and a state change of work object are occurred alternately. Fig.3 shows the flow of removal work of HDD(Hard Disk Drive) as an example of removal work of part in notebook computer. Note that, it is assumed that the state change of work object is occurred by worker intervention. The worker intervention can be detected by using Leaf motion sensor which is sensor that can recognize person's hand and fingers. Moreover, the work process does not branch.

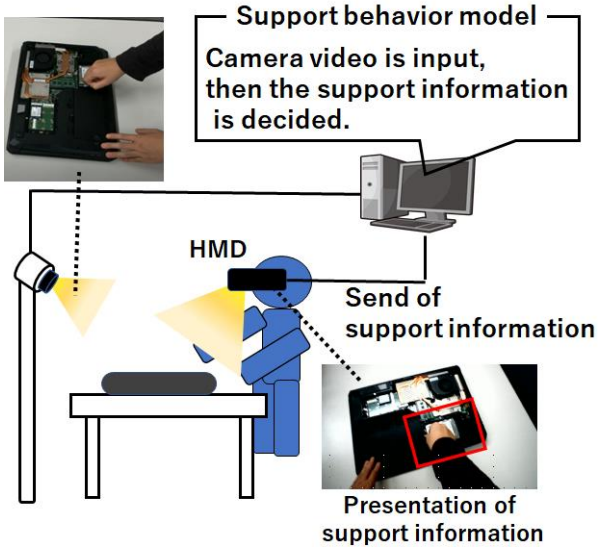

Figure 2. An expected work support system by using HMD

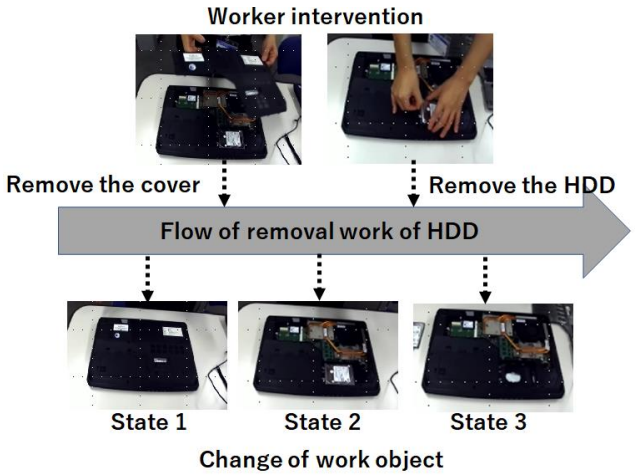

Figure 3. Flow of the removal work of Hard Disk Drive as an example of target work

3. Modeling Method of Support Behavior

3.1. Approach

As Fig.3 shows, the state of work object changes according to execute of work process. Here, the state sequence of work object can be regarded as work process. Moreover, change region of work object under state transition can be regarded as work locations. Therefore, if the state sequence of work object and current state of work object can be recognized, next state of work object can be predicted. Moreover, the work locations can be detected by estimating the change region between current and next state of work object. In this paper, a model structure that can process the above functions is proposed as the support behavior model.

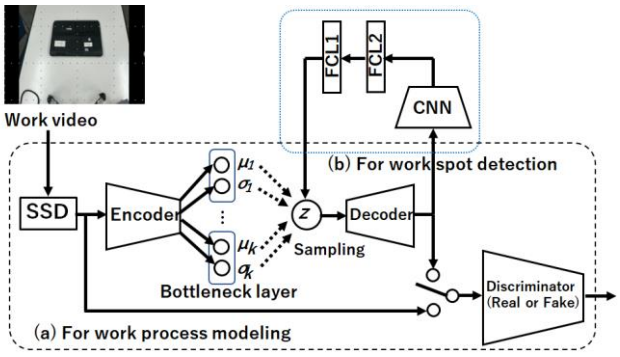

Figure 4. The proposed model architecture

Fig.4 shows the architecture of proposed model. This model is composed of two model structures. The model structure shown in Fig.4(a) is the structure to generate work process model. This model is called as work process model in this paper. On the other hand, the model structure shown in Fig.4(b) is the structure to detect work locations. This model is called as work location detector in this paper.

3.2. Work process model

The model shown in Fig.4(a) is composed of Single Shot MultiBox Detector(SSD), VaDE(Variational Deep Embedding), and GAN(Generative Adversarial Networks). SSD is deep learning model to detect a object. In this paper, when the work video is input to support behavior model, the rectangle region of the notebook computer is detected based on SSD(Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg, 2016). Then, the detected rectangle region image is input to VaDE.

VaDE is one of the image generation models based on deep learning model(Zhuxi Jiang, Yin Zheng, Huachun Tan, Bangsheng Tang, and Hanning Zhou, 2017). Fig.5(a) shows the structure of VaDE. VaDE is composed of Encoder layer which converts input image to low-dimensional feature, Bottleneck layer which expresses the converted feature space, and Detector layer which converts the converted feature to image. This model is trained to generate same image as input image from Decoder layer. Then, the feature distribution for input image is generated in Bottleneck layer. Here, the feature value of Bottleneck layer is called as latent variable. When similar images input to VaDE, these images are converted to latent variables as near value. Therefore, when the images which expresses each state of work object input to VaDE, it is considered that the feature distributions of each state of work object are generated in the feature space. Moreover, the state sequence of work process can be expressed by estimating the transition paths of the feature distributions.

As Fig.5(b) shows, VaDE can model the feature distributions based on GMM(Gaussian Mixture model). Therefore, unsupervised classification of states of work object can be executed. Note that, the number K of clusters is set by designer. Moreover, when the work video is input to VaDE, the transition path among clusters can be obtained. Here, the transition network of clusters can be regarded as direct graph. Therefore, the similar clusters are combined based on Markov Cluster Algorithm, and finally work process model is generated as Fig.5(c) shows.

GAN is one of the image generation models based on deep learning model(Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil, Aaron Courville, and Yoshua Bengio, 2014). In this model, Discriminator layer is added for the Decoder layer. Discriminator layer judges whether true or fake for the generated image from Decoder layer. And, Decoder layer is learned to judge true in Discriminator layer. Finally, Decoder layer can generate image with high quality. The generated images from VaDE tend to become blurred. This is because that this model has probabilistic distribution in Bottleneck. Therefore, the image generation with high quality is realized by adding GAN to VaDE in the propose model. In this paper, this model is call as

VaDE-GAN.

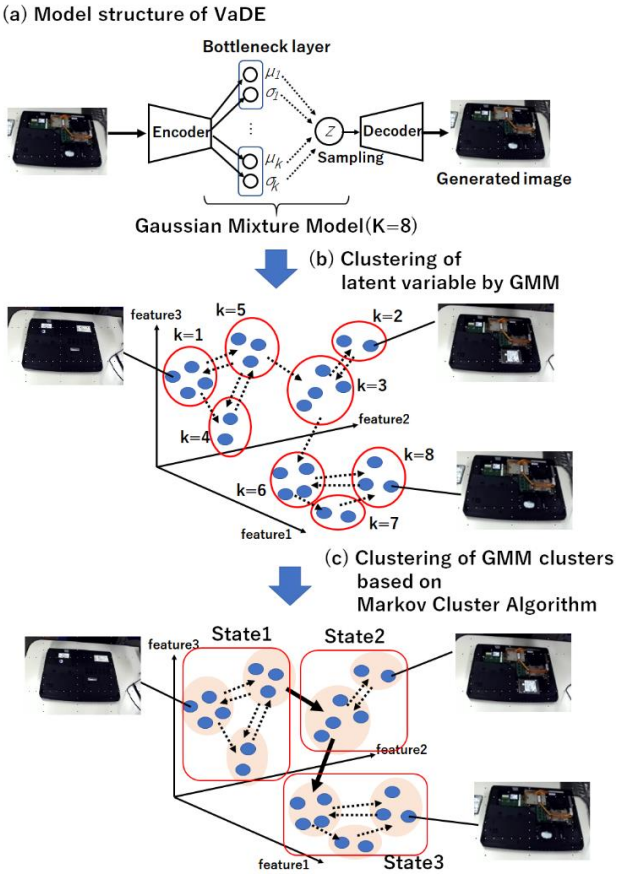

Figure 5. The structure of VaDE and modeling flow of work process

3.3. Work location detector

As Fig.3 shows, the state of work object changes according to execute of work process. Here, the change region of work object under state transition can be regarded as work locations. In this model, the work locations are detected by using CNN and Grad-CAM(Gradient-weighted Class Activation Mapping). CNN is used for recognition of state label of work object. On the other hand, Grad-CAM is a method that image region which influences recognition result of CNN visualizes by heat map(Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhrov Batra, 2016). In this model, the visualized heat map by Grad-CAM is regarded as work locations. The model shown in Fig.4(b) is generated the following way. When the image of work object is input to VaDE, the latent variable which is output from Bottleneck layer, state label of work object which is recognized based on work process model, the image which is generated from VaDE-GAN are obtained. Then, CNN in Fig.4(b) is trained by using the generated image from VaDE-GAN as input data and the state label of work object as output data. Moreover, FCL in Fig.4(b) is trained by using the state label of work object as input data and the latent variable as output data.

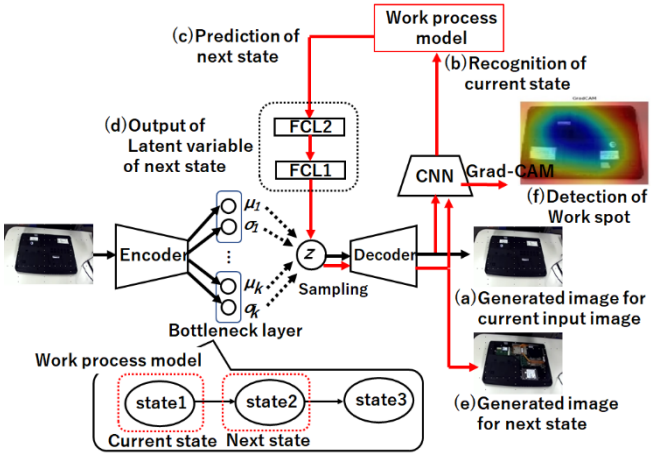

Figure 6. The flow of detection of work locations based on the propose model

The flow of detection of work spots is described as follows. At first step, the image which expresses current state of work object is generated from VaDE by inputting the work video as Fig.6(a) shows. Then, the generated image is input to CNN and the recognition result is obtained as current state label of work object as Fig.6(b) shows. Then, the next state label of work object is estimated based on the work process model by inputting the obtained current state label of work object as Fig.6(c) shows. Then, the estimated next state label of work object is input to FCL, and the latent variable corresponding to the next state is obtained as Fig.6(d) shows. Then, the obtained latent variable is input to Decoder on VaDE, and the image which expresses next state of work object is obtained. Finally, the obtained image is input to CNN, and the heat map based on Grad-CAM is obtained as Fig.6(f) shows.

4. Experiment

4.1. Experimental Setup

One repair notebook computer is prepared as work object. In this experiment, the part removal work process for notebook computer is modeled based on the proposed method. There are three type work process, Battery, HDD, Fan removal work process. Fig.7, 8, and 9 show notebook computer's states for each work process. Note that, each state of the notebook computer called as Normal, Battery, Cover, HDD, Fan Cover, and Fan as these figures show. The Battery removal work process is composed of one process. And the state is transited in order of Normal, and Battery. The HDD removal work process is composed of two processes. And the state is transited in order of Normal, Cover, and HDD. The Fan removal work process is composed of three processes. And the state is transited in order of Normal, Cover, Fan Cover, and Fan. In this experiment, each work video is obtained, and the support behavior model for each work process is generated by training the obtained work video based on the proposed method.

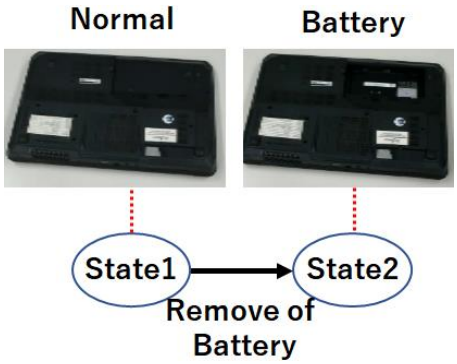

Figure 7. State transition of Battery removal work process

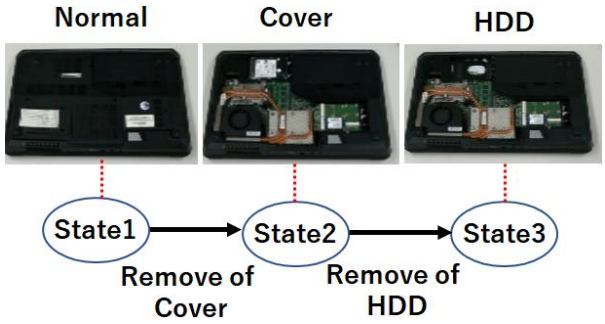

Figure 8. State transition of HDD removal work process

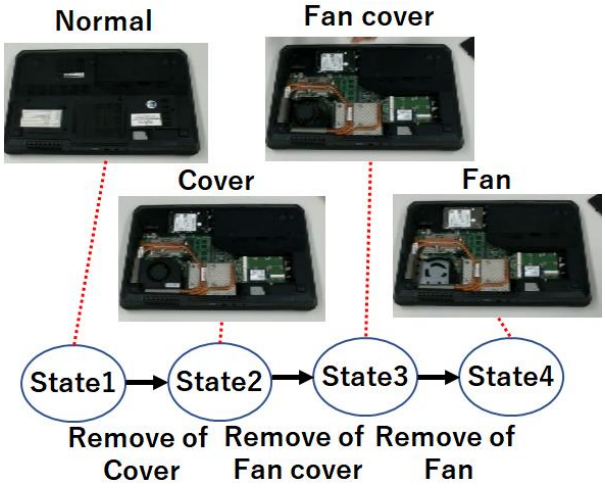

Figure 9. State transition of FAN removal work process

4.2. Evaluation of the Generated Network of State Transition

Table.1 shows the composition of the generated state transition network for the Battery removal work process based on the proposed method. The value of K which expresses mixture gaussian model's parameter is set 6, 8, 10, and 15. State label shows the clustering result based on Markov Cluster Algorithm, and each label is described in order of state transition. On the other hand, cluster label shows the cluster number of clustering result based on VaDE with mixture gaussian model. As this table shows, it is confirmed that each state transition network is composed of two states even if the value of K changes. Fig.10 shows the visualization of the generated state transition network for K=15. In this figure, each latent variable of VaDE for learning images is plotted in 2-dimensional feature space based on multidimensional scaling method, and plotted points are colored different of state label in different colors. As this figure shows, it is confirmed that two distributions are generated. In this figure, each generated image by VaDE for cluster label 1, 6, 9, and 13 is shown. As this result shows, state label 1 and 2 express Normal and Battery state. According to these results, it is confirmed that the Battery removal work process is modeled based on the proposed method.

Table 1. Generated state transition network for each value of K(Battery removal work process)

Value of K	State label	Cluster label
6	State1	0, 2, 4
	State2	1, 3, 5
8	State1	1, 4, 6
	State2	0, 2, 3, 5, 7
10	State1	1, 4, 6
	State2	0, 2, 3, 5, 7,8, 9
15	State1	1, 2, 3, 6, 7, 8, 12
	State2	0, 4, 5, 9, 10, 11, 13, 14

Table 2. Generated state transition network for each value of K(HDD removal work process)

Value of K	State label	Cluster label
6	State1	0, 2, 5
	State2	1, 3, 4
8	State1	0, 2, 4, 7
	State2	1, 3, 5, 6
10	State1	2, 3, 5, 8, 9
	State2	0, 1, 4, 6, 7
15	State1	1, 4, 6, 7, 8, 9, 11
	State2	5, 10, 14
	State3	0, 2, 3, 12, 13

Table 3. Generated state transition network for each value of K(FAN removal work process)

Value of K	State label	Cluster label
6	State1	0
	State2	1, 2, 3, 4, 5
8	State1	2, 6
	State2	0, 1, 3, 4, 5, 7
10	State1	3, 6, 8
	State2	0, 4, 9
	State3	1, 2, 5, 7
15	State1	0, 3, 7, 10
	State2	4, 5, 9, 11, 13, 14
	State3	1, 2, 6, 8, 12

Table.2 shows the composition of the generated state transition network for HDD removal work process based on the proposed method. As this table shows, when K is 6, 8, 10, the generated state transition network is composed of two states. On the other hand, when K is 15, the generated state transition network is composed of three states. Fig.11 show the visualization of the generated state transition network for K=15. As Fig.11 shows, it is confirmed that Normal, Cover, and HDD states are expressed by state label 1, 2, and 3.

Table.3 shows the composition of the generated state transition network for Fan removal work process based on the proposed method. As this table shows, the state transition network with 4 states could be not

generated. Fig.12 shows the visualization of the network state transition for K=15. Both Cover and Fan Cover state are expressed by state label 2. This is because that the variation of image information from Cover to Fan Cover state is few. According to these results, network's states are generated based on different of image information in the proposed method. Therefore, it is difficult to apply to a work process that captured images are not varied.

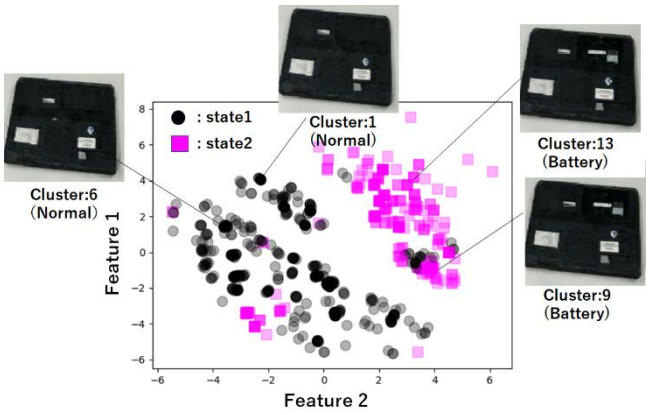

Figure 10. Visualization of the generated state transition network for Battery removal work process(K=15)

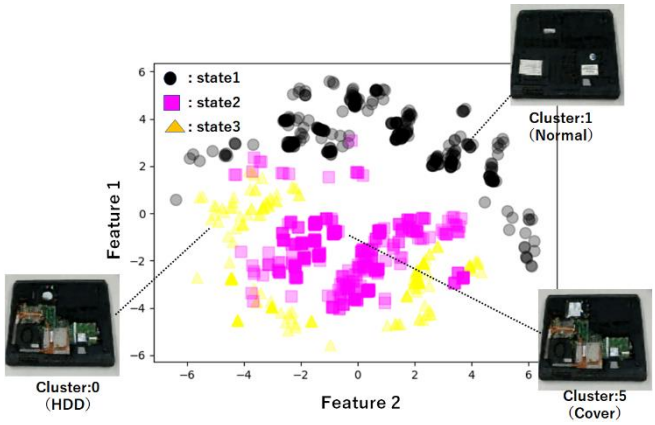

Figure 11. Visualization of the generated state transition network for HDD removal work process(K=15)

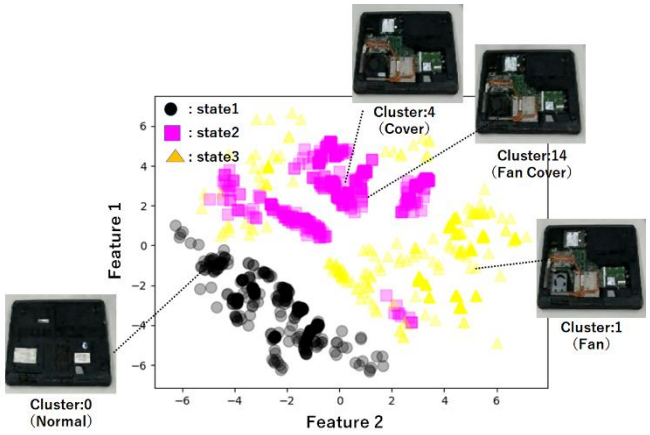

Figure 12. Visualization of the generated state transition network for FAN removal work process(K=15)

4.3. Accuracy Evaluation of Work Location Detection

Next, the usefulness of the proposed method is evaluated by comparing on detected region of work location based on proposed model and the correct region. Note that parameter K on VaDE is set 15.

Fig.13, Fig.14, and Fig.15 show the detected work location based on the proposed model for each work process. Fig.13 is detection result of work location based on the proposed model for Battery removal work process. The work location is presented by heat map based on Grad-CAM. In this map, it is regarded that blue color's region is high degree of influence for work location. On the other hand, the top of figure shows the correct region of work location. As this figure shows, it is confirmed that close region to correct region is detected as work location based on proposed model. Fig.14 shows the detection result of work location for HDD removal work process. And Fig.15 shows the detection result of work location for Fan removal work process.

Next, the detected region based on the proposed model is evaluated by using recall and precision value. Note that, the discriminant analysis method is applied to the heat map of Grad-CAM. And the region of high-value class is regarded as the detected region based on the proposed method. Table.4 and 5 show the calculation result of recall and precision value. As these tables show, it is confirmed that high recall value is calculated, and low precision value is calculated. This means that the detected region based on the proposed model includes the correct region and unrelated region. Heat map of Grad-CAM is calculated on feature map of convolutional layer of CNN. This feature map is scaled down compared on input image. In the other hand, heat map is obtained by scaling the distribution of influence degree calculated on the feature map up. Therefore, the heat map based on Grad-CAM is occurred the lack of position information for work location. According to this reason, it is difficult to detect the accurate region of work location based on the Grad-CAM. For this problem, the expanded method of Grad-CAM, such as Score-CAM(Wang Haofan, Du Mengnan, Yang Gan, and Zhang Zijian, 2019) and GAIN(Guided Attention Inference Networks)(Kunpeng Li, Ziyan Wu, Kuan-Chuan Peng, Jan Ernst, and Yun Fu, 2018) has been proposed. Moreover, the change region detection method, such as ChangeNet(Ashley Varghese, Jayavardhana Gubbi, Akshaya Ramaswamy, and P. Balamuralidhar, 2018), has been proposed. By using these methods, it is considered that the detection accuracy will be improved.

5. Conclusions

In this paper, we focused on the time and effort to design the support behavior of work support system by using HMD. The more increasing the number of work contents, work processes, and work parts, the more increasing time and effort to design the support

system. Therefore, we considered that it is effective to develop the technique that support behavior of work support system can be generated automatically. In this paper, we proposed an automatic generation method of support behavior of the system based on the VaDE-GAN.

In the proposed method, The deep learning model is constructed composed of work process model and work location detector. The work process model can model the state sequence of work object. And the work location detector can detect the change region between current and next state of work object as work location.

In the experiment, the generated network of state transition based on the proposed method was evaluated by comparing the actual state transition for removal work process of notebook computer. As the result of experiment, it is confirmed that Battery and HDD removal work process could be modeled based on the proposed method. However, Fan removal work process could not be modeled correctly based on the proposed model. This is because that the variation of image information for state transition was low. Therefore, it is difficult to apply to a work process that captured images are not varied. It is necessary to improve the obtained learning data to increase amount information. For example, learning data is obtained at three-dimensional image information by using depth sensor and polarization camera.

Moreover, the usefulness of the proposed method for removal work process of notebook computer is evaluated by comparing on detected region of work location based on the proposed method and the correct location. As the result of the experiment, it was confirmed that close region to correct region is detected as work location based on proposed model. However, the accurate region can not be detected. This is because that the obtained heat map based on Grad-CAM is occurred the lack of position information for work location. Therefore, it is necessary to apply the expand method of Grad-CAM, such as Score-CAM and GAIN or detection methof of change region, such as ChangeNet.

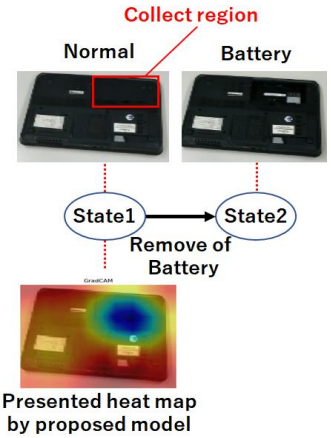

Figure 13. Obtained heat map as work location based on Grad-CAM on the proposed model for Battery removal work process.

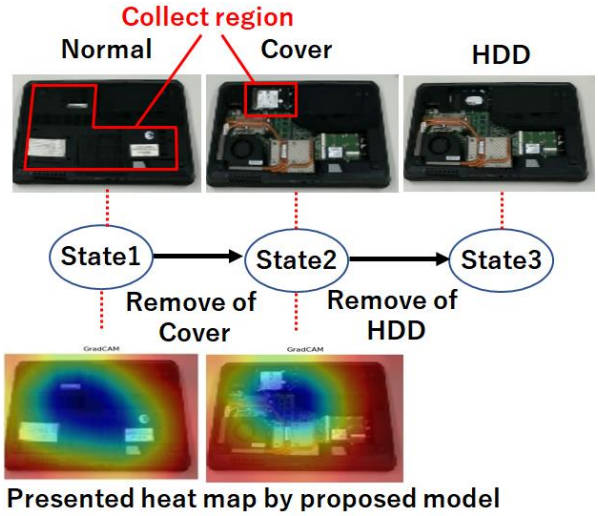

Figure 14. Obtained heat map as work location based on Grad-CAM on the proposed model for HDD removal work process.

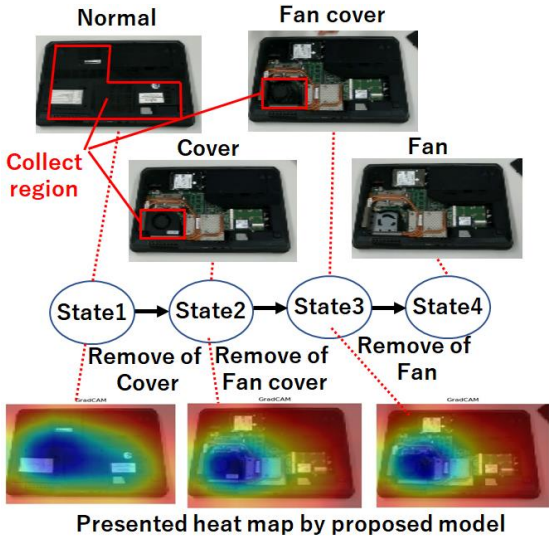

Figure 15. Obtained heat map as work location based on Grad-CAM on the proposed model for FAN removal work process.

Table 4. Calculation result of recall

		Recall	
	State	Average	Standard deviation
Battery	State1	0.94	0.03
HDD	State1	0.97	0.05
	State2	0.65	0.02
FAN	State1	0.80	0.01
	State2	0.74	0.08
	State3	0.78	0.03

Table 5. Calculation result of precision

		Precision	
	State	Average	Standard deviation
Battery	State1	0.52	0.03
HDD	State1	1.00	0.05
	State2	0.11	0.05
FAN	State1	0.96	0.02
	State2	0.24	0.02
	State3	0.24	0.01

Acknowledgements

This research is supported by Takahashi Industrial and Economic Research Foundation, and Takano Science Foundation.

References

Patrice Humblot-Nino, Oscar Sandoval-Gonzalez, Ignacio Herrera-Aguilar, Daniel Rangel-Penuelas, (2017), Approach on a new methodology for skills transfer using a parallel planar robot with visuo-vibrotactile feedback. Proceedings of 14th International Conference in Electrical Engineering, Computing Science and Automatic Control, October 20-22, Mexico City, Mexico.

Masao Sugi, Hisato Nakanishi, Masataka Nishino, Yusule Tamura, Tamio Arai, Jun Ota, (2010), Development of Deskeork Support System using Pointing Gesture Interface, Journal of Robotics and Mechatronics, Vol.22, No.4, pp.430-438.

Kazuyoshi Tagawa, Hiromi Tanaka, Masayu Komori, Yoshimasa Kurumi, and Shigehiro Morikawa, (2012), A Visiohaptic Surgery Training System for Laparoscopical Techniques, Japanese Journal for Medical Virtual Reality, Vol.10, No.1, pp.11-18.

Kohjiro Hashimoto, Tadashi Miyosawa, Mai Higuchi, (2018), DEVELOPMENT AND EVALUATION OF WORK SUPPORT SYSTEM BY AR USING HMD, Proceedings of the International Conference of the Virtual and Augmented Reality in Education Workshop, pp.28-33.

Reiko Takizuka, Haruhisa Kato, Hiromasa Yanagihara, Masaru Sugano, (2016), Usefulness of operation support system by using AR technique, Technical report of Information Processing Society of Japan, pp.1-6, No.11.

Koichiro Shiraishi, Kohei Matsuo, (2015), Piping installation support system using augmented reality, Transactions of the Japan society of mechanical engineers, Vol.81, No.825.

Hirohiko Sagawa, Hiroto Nagayoshi, Harumi Kiyomizu, Tsuneya Kurihara, (2015), Hands-free AR Work Support System Monitoring Work Progress with Point-cloud Data Processing, Proceedings of IEEE International Symposium on Mixed and Augmented Reality, pp.172-173.

Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. Berg, (2016), SSD:Single Shot MultiBox Detector, In: European Conference on Computer Vision, arXiv:1512.02325v5, pp.21-37, 2016.

Zhuxi Jiang, Yin Zheng, Huachun Tan, Bangsheng Tang, and Hanning Zhou, (2017), Variational Deep Embedding:An Unsupervised and Generative Approach to Clustering, proceedings of the 26th International Joint Conference on Artificial Intelligence, pp.1965-1972, 2017.

Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil, Aaron Courville, Yoshua Bengio, (2014), Generative Adversarial Nets, Proceedings of Neural Information Processing Systems, pp.2672-2680, 2014.

Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, Dhrov Batra, (2016), Grad-CAM:Visual Explanations from Deep Networks via Gradient-based Localization, Proceedings of International Conference on Computer Vision, pp.618-626, 2016.

Wang Haofan, Du Mengnan, Yang Gan, Zhang Zijian, (2019), Score-CAM:Improved Visual Explanations Via Score-Weighted Class Activation Mapping, arxiv.org/abs/1910.01279, 2019.

Kunpeng Li, Ziyan Wu, Kuan-Chuan Peng, Jan Ernst, Yun Fu, (2018), Tell Me Where to Look: Guided Attention Inference Network, Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.9215-9223, 2018.

Ashley Varghese, Jayavardhana Gubbi, Akshaya Ramaswamy, and P. Balamuralidhar, (2018), ChangeNet: A Deep Learning Architecture for Visual Change Detection, Proceedings of the European Conference on Computer Vision, pp.129-145, 2018.