

A Study on Novel Hand Hygiene Evaluation System using pix2pix

Fumiya Kinoshita^{*1}, Kosuke Nagano¹, Gaochao Cui², Miho Yoshii³, Hideaki Touyama¹

¹Information Systems Engineering, Graduate School of Engineering, Toyama Prefectural University, Toyama 939-0398 Japan

²Department of Electrical and Computer Engineering, Faculty of Engineering, Toyama Prefectural University, Toyama 939-0398 Japan

³Graduate School of Medicine and Pharmaceutical Sciences, University of Toyama, 930-0194 Japan

ARTICLE INFO

Article history:

Received: 21 January, 2022

Accepted: 09 March, 2022

Online: 18 March, 2022

Keywords:

Hand hygiene

Direct observation method

Generative adversarial networks (GAN)

pix2pix

Discriminant analysis

Mahalanobis distance

ABSTRACT

The novel coronavirus infection (COVID-19), which appeared at the end of 2019 has developed into a global pandemic with numerous deaths, and has also become a serious social concern. The most important and basic measure for preventing infection is hand hygiene. In this study, by photographing palm images of nursing students after hand-washing, using fluorescent lotion to conduct hand-washing training and as a black light, we developed a hand hygiene evaluation system using pix2pix, which is a type of the generative adversarial network (GAN). In pix2pix, the input image adopted was a black light image obtained after hand-washing, and the ground truth image was a binarized image obtained by extracting the residue left on the input image by a trained staff member. We adopted 443 paired-images after hand-washing as training models, and employed 20 images as verification images, which included 10 input images with 65% or more of the residue left, and 10 input images with 35% or less of the residue left in the ground truth images. To evaluate the training models, we calculated the percentage of residue left in the estimated images generated from the verification images, and conducted two-class discriminant analysis using the Mahalanobis distance. Consequently, misjudgment only occurred in one image for each image group, and the proposed system with pix2pix exhibited high discrimination accuracy.

1. Introduction

The novel coronavirus infection (COVID-19), which appeared at the end of 2019, has developed into a global pandemic with numerous deaths, and has also become a serious social concern. COVID-19 is transmitted in the community and develops into a nosocomial infection when an infected person visits the hospital. Because many elderly people and vulnerable patients are hospitalized in hospitals, the damage caused by nosocomial infections is enormous [1]. Developing a novel drug for COVID-19 will take a protracted period of time regardless of the promise of a possible therapeutic drug or the anticipated development of vaccines. Therefore, daily measures against this infection are crucial.

The most important and basic approach to preventing infection is hand hygiene. Hand hygiene is used as a general term that

applies to any form of hand-washing and hand disinfection during surgery, and it refers to washing off organic matter, such as dirt and transient bacteria, that cling to our hands [2]. Although hand-washing has been a cultural and religious practice for centuries, it is believed that the scientific evidence for hand hygiene in the prevention of human diseases emerged only in the early 19th century [3, 4]. The principle of hand hygiene is centered on reducing the dirt and bacteria accumulated on the hands of healthcare workers as much as possible, as well as keeping the hands of healthcare workers clean when they provide care to patients. SARS-CoV-2, the causative virus of COVID-19, invades the body through the mucous membranes of the eyes, nose, and mouth; however, the primary transmission routes that carry the virus to these areas are the fingers [5]. Therefore, keeping the hands clean ensures safety in medical care and nursing, including the safety of both the patient and medical staff. Conventionally, for hand hygiene methods, researchers refer to the guidelines issued by the Centers for Disease Control and Prevention (CDC) and

*Fumiya Kinoshita, Email: f.kinoshita@pu-toyama.ac.jp

World Health Organization (WHO) [6, 7]. These guidelines elucidate hand hygiene methods based on a large number of medical literature and research data, and are easily applicable in actual field situations as they also describe the recommended levels of application [8].

General hand hygiene evaluation methods include the indirect observation method for observing the amount of hand sanitizer, as well as the direct observation method employed for observing the timing for the hand-washing and disinfection method via direct visual inspection. The WHO guidelines recommend the direct observation method by trained staffs [7]. However, this method is limited because it requires trained staff, as well as a significant amount of time to observe several scenes. Additionally, because the direct observation method involves subjective evaluations by the staff, results may vary depending on the skill level of the staff [9]. The methods for the quantitative evaluation of hand hygiene include the palm stamp method, adenosine tri phosphate (ATP) wiping test, and glove juice method [10–13]. The palm stamp method is a method that adopts a special medium, in which the number of bacteria on a hand is visualized by pressing the palm against the special medium, before and after hand-washing. However, this method requires culturing the bacteria present on this hand, which is expensive and time-consuming. Next, the ATP wiping test is a method that analyzes the content of ATP in the living cells of animals, plants, microorganisms, etc. In this method, ATP wiped from the palm is chemically reacted with a special reagent to trigger light emission by the ATP. The amount of light emitted at this point (relative light units; RLU) is evaluated as an index of contamination. Because the ATP wiping test does not require a medium, it is not as expensive and time-consuming as the palm stamp method. However, it is limited because its values fluctuate according to the wiping approach employed with the wiping stick and the amount of strength applied to the wiping stick [14]. Finally, the glove juice method is an evaluation method for hand hygiene recommended by the FDA (U.S. Food and Drug Administration). In this method, the test subject puts on rubber gloves, and the sampling liquid and neutralizer are poured into the gloves. Then, the bacterial liquid in the rubber gloves is cultured, and the amount of bacteria on the hands of the test subject is measured to evaluate hand hygiene. However, the glove juice method is unsuitable for evaluations that focus on each part of the palm as this method evaluates the amount of bacteria present on the entire palm. Therefore, in this study, we developed a hand hygiene evaluation system for evaluating palm images after hand-washing based on the perspective of trained staffs using pix2pix, which is a type of generative adversarial network (GAN).

2. Experimental Method

In the Department of Nursing, Faculty of Medicine, University of Toyama, education on hand hygiene is conducted for nursing students using a hand-washing evaluation kit (Spectro-pro plus kit, Moraine Corporation) [15]. In this process, students apply a special fluorescent lotion (Spectro-pro plus special lotion, Moraine Corporation) on their entire palms and wash their hands hygienically according to the guidelines recommended by WHO. Then, the students themselves evaluate their hand-washing skills by sketching the residue left under a black light. During this hand-washing training, the palm images adopted in this experiment were photographed and used for analysis. This experiment was

conducted after obtaining approval from the Toyama Prefectural University Ethics Committee.

A special imaging box (dimensions: 30 × 30 × 45 cm) was developed to merge the shooting environment for the palm images. Two LED fluorescent lights (RE-BLIS04-60F, Reudo Corp.) and two black lights (PL10BLB, Sankyo Denki) were installed inside the imaging box (Fig. 1). Although the black light was turned on while shooting the palm image, the entrance part was shaded with a blackout curtain, and then the picture was photographed using a digital camera (PowerShot G7 X Mark II, Canon). The number of pixels of the captured image was 5472 × 3072 px, and the parameters at the point of shooting, including resolution, F value, shutter speed, ISO sensitivity, and focal length were 72 ppi, 2.8, 1/60 s, 125, and 20 cm, respectively. For these parameters, the numerical values that are in focus on the palm without overexposing the image, even when shooting with the fluorescent light on, were selected. The procedure for obtaining the palm image is provided below.

- The test subject applies the fluorescent lotion on the entire palm and confirms that the fluorescent lotion is applied correctly under the black light.
- The test subject performs hygienic hand-washing using running water and soap according to WHO guidelines. After hand-washing, the test subject wipes off moisture with a paper towel.
- The test subject then places the palm of their right arm upwards in the imaging box, and the photographer takes one image each under the fluorescent and black lights (Fig. 2).

Via the hand-washing training, 463 palm images after hand-washing were obtained in this experiment.



Figure 1: Illustration of imaging box

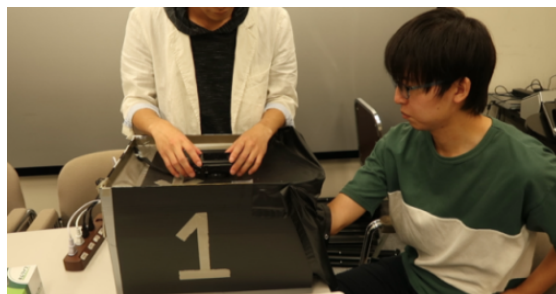


Figure 2: Experimental scene

3. Generation of training models using pix2pix

3.1. Characteristics of pix2pix

In recent years, machine learning technology has been remarkably improved and applied to various fields such as natural language processing and speech recognition, as well as imaging areas such as image classification, object detection, and segmentation [16]. Among these applications, generative adversarial networks (GANs) have been garnering considerable attention recently. GAN is an unsupervised learning method for neural networks proposed by [17]. It is characterized by its adoption of two convolutional neural networks called generator and discriminator in opposition. The generator learns to generate images such that generated images are not misidentified as images generated by the discriminator, and the discriminator learns such that the generated data are not misidentified as the training data. Accordingly, an image similar to the training data is ultimately generated. The pix2pix adopted in this study is also a type of GAN, as well as an image generation algorithm proposed by [18]. The pix2pix learns the relationship between images from two paired images: input and ground truth images, which generate an estimated image considering the relationship obtained when inputting the input image. In previous studies, pix2pix has been used to generate a wide variety of images, such as creating maps from aerial photographs and generating color images from black-and-white images [18–20]. Conventionally, it is necessary to design a network to address a specific challenge when generating images. However, in the proposed method, it is unnecessary to design a network for each problem as pix2pix is a highly versatile image generation algorithm. Implementation methods that can easily manage pix2pix are also published on the internet, and these methods can be simply adopted by preparing two paired images: the input and ground truth images. In this study, we implemented pix2pix according to the pix2pix-Tesorflo [21] repository on GitHub.

3.2. Pre-processing of training data

The input image adopted for pix2pix in this study was a black light image obtained after hand-washing, while the adopted ground truth image was a binarized image obtained by extracting the residue left from the input image by a trained staff from a faculty in the Department of Nursing. The pre-processing procedure for the paired image is presented below.

- Dirt, such as the adhesion of fluorescent lotion to the floor surface of the imaging box, may affect the analysis. Therefore, using the image captured under fluorescent light, only the right arm part was extracted from the image captured under the black light. In contour extraction, MATLAB's image processing toolbox was employed to perform edge detection and basic morphology via Sobel's method [22]. Fig. 3a presents an image captured under the fluorescent light, while Fig. 3b illustrates a binarized image in which the outline of the arm area is extracted from Fig. 3a. In addition, the arm area is painted white and the rest of the area is painted black. Fig. 3c presents the results of superimposing Fig. 3b on the image captured under black light.
- In this system, the parts from the arm to the wrist of the black light image with the background (painted in black) is excluded

to focus on the residue left in the palm area (Fig. 4a). It is necessary to change the length-width pixel ratio of the input image of pix2pix to 1:1. Accordingly, the black pixels were evenly arranged at the top and left-right corners of the image such that the number of pixels of all black light images became 4746×4746 px. After that, the residue left was extracted visually from the black light image by the trained staff. In the extraction of the residue left, the brightness value of the black light image was targeted, and binarization image processing was conducted such that the areas the trained staff felt were left unwashed are displayed in black (Fig. 4b). Furthermore, only the residue left is depicted as black in the actual ground truth image because the training model did not include the contour extraction of the palm.

The above-mentioned processing method was carried out on all 463 captured images. Among these images, 20 images, which include 10 input images with 65% or more of the residue left and 10 input images with 35% or less of the residue left in the ground truth images, were randomly extracted to evaluate the training models, and these images were adopted as verification images. When generating the training models, the paired images were reduced to an image size from 4746×4746 px to 256×256 px. The other parameters considered, which include batch size and learning rate values were 10 and 0.00002, respectively. In addition, in the 443 paired images adopted in the training models, a bias exists in their percentage of residue left in the ground truth images. Therefore, we also investigated the training models when a data expansion technique (data augmentation) is applied, such that the percentage of residue left in the ground truth images is uniform [23].

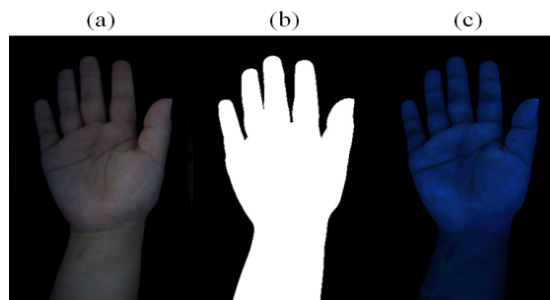


Figure 3: Pre-processing method of input image

- (a) Image captured under fluorescent light, (b) Binarized image, in which arm area is extracted from (a), (c) Image captured under black light, in which every area other than the arm area is painted in black using (b)

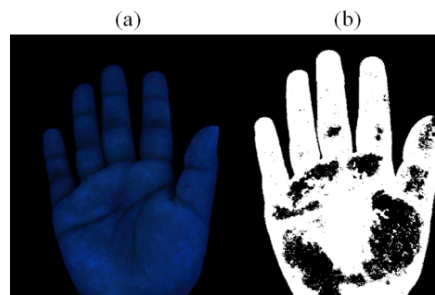


Figure 4: Method for generating ground truth image

- (a) Input image (brightness tripled for the paper), (b) Binarized image, in which the residue left is extracted from the input image

4. Results

In this experiment, 463 palm images were obtained, and among them, 20 were used as verification images, such that the remaining 443 paired images were adopted as pix2pix training models. In addition, a histogram was created from the 443 ground truth images, in which the horizontal and vertical axes represent the percentage of residue left (in increments of 10%) and number of images, respectively. Furthermore, we also examined the training models at the point where a negligible bias appeared in the percentage of residue left in the ground truth images. In data augmentation, which is a data expansion technique, three processes, including enlargement, reduction, and translation, were conducted on the palm images. In both the enlargement and reduction processes, the size of the palm was altered from 85% to 115% in 5% increments. In the translation process, the position of the palm was moved horizontally by 5 px and 10 px, respectively. Fig. 5 presents a histogram of the percentage of residue left adopted in the training models, and the number of images. Fig. 5a presents a histogram of the number of images present in the 443 paired images while Fig. 5b presents a histogram of 1800 paired images after data augmentation.

An estimated image was generated from the verification image using each training model, and the percentage of residue left was calculated accordingly (Fig. 6). Fig. 6a presents the percentage of residue left in the 443 training models while Fig. 6b presents the percentage of residue left in the training models after data augmentation. Subsequently, a two-class discriminant analysis was performed using the Mahalanobis distance as the percentage of residue left in the estimated image [24]. Consequently, the misjudgment in the training models without data augmentation occurred in two images for each image group, and the misjudgment in the training models with data augmentation occurred in one image for each image group. Examples of the input, ground truth, and estimated images adopted for evaluation in each training model are presented in Figs. 7–9 (note that the input image is tripled in brightness for this paper). Fig. 7 presents an example in which the percentage of residue left in the estimated image was correctly discriminated in both training models. Fig. 8 illustrates an example in which the percentage of residue left in the estimated image was solely correctly discriminated in the training model after data augmentation. Finally, Fig. 9 shows an example in which the percentage of residue left in the estimated image could not be correctly discriminated in both training models.

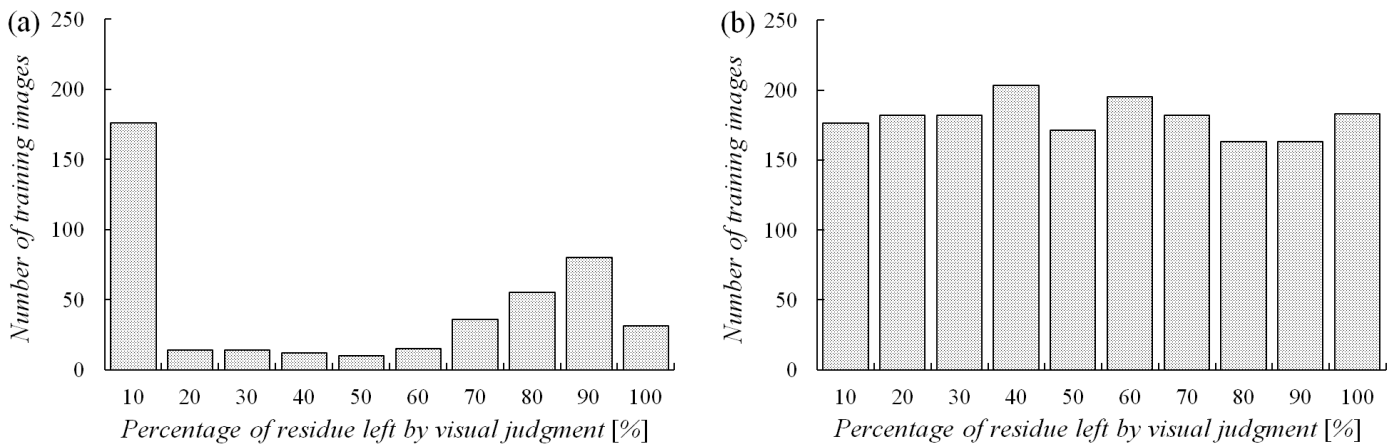


Figure 5: Histogram of percentage of residue left in the ground truth image in each training model

(a) Training model without data augmentation, (b) Training model with data augmentation

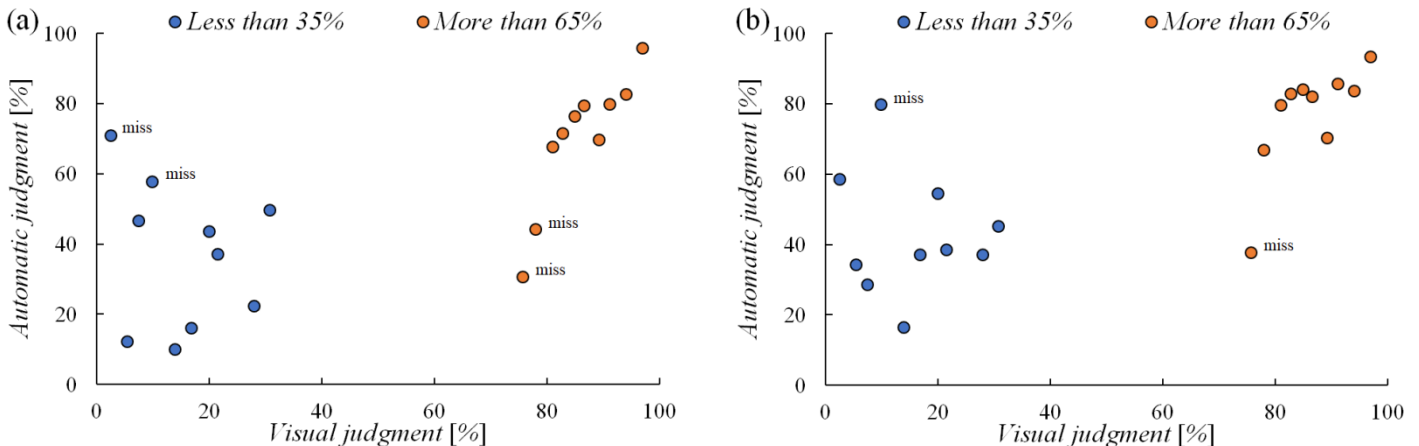


Figure 6: Percentage of residue left in the estimated image generated from the verification image

(a) Training model without data augmentation, (b) Training model with data augmentation

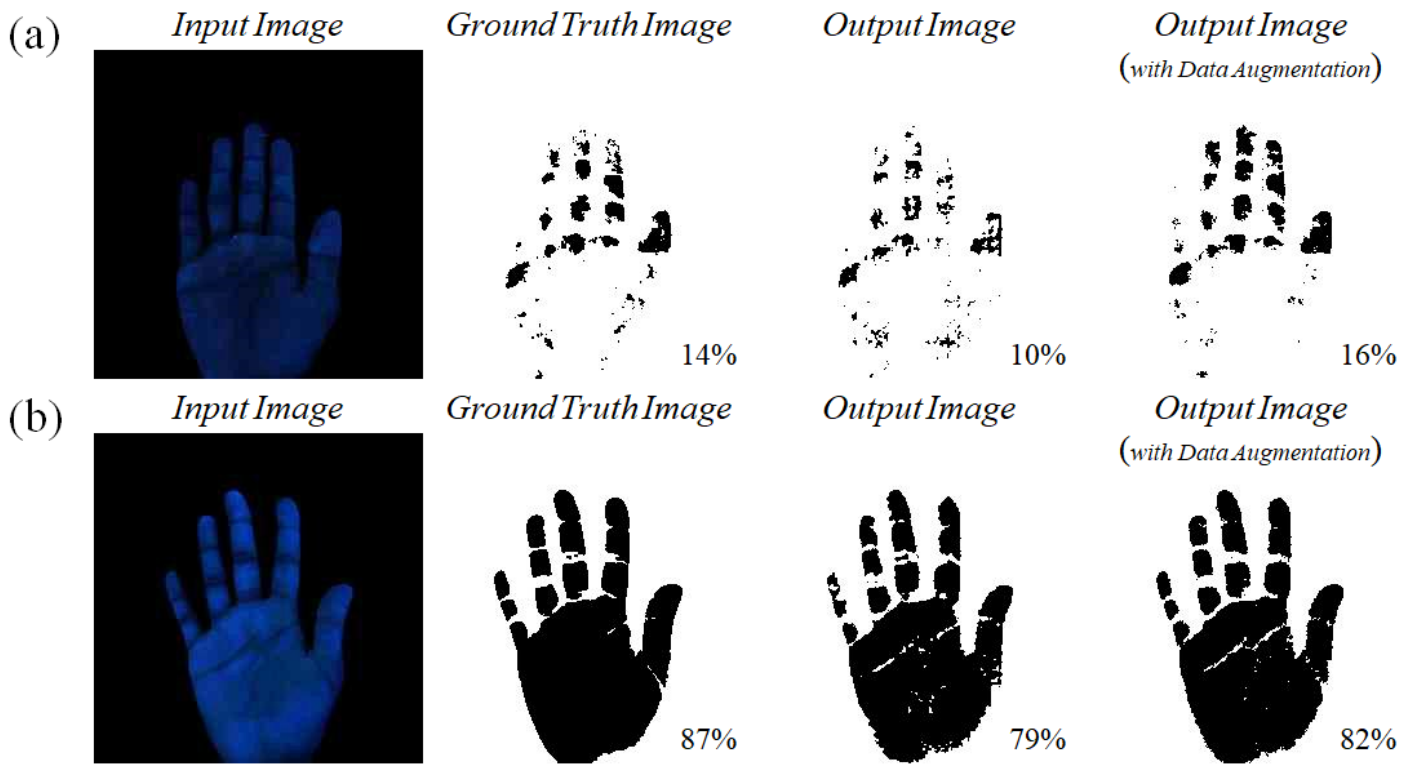


Figure 7: Examples of verification images correctly discriminated in both training models
 (a) Input image with 35% or less of residue left, (b) Input image with 65% or more of residue left

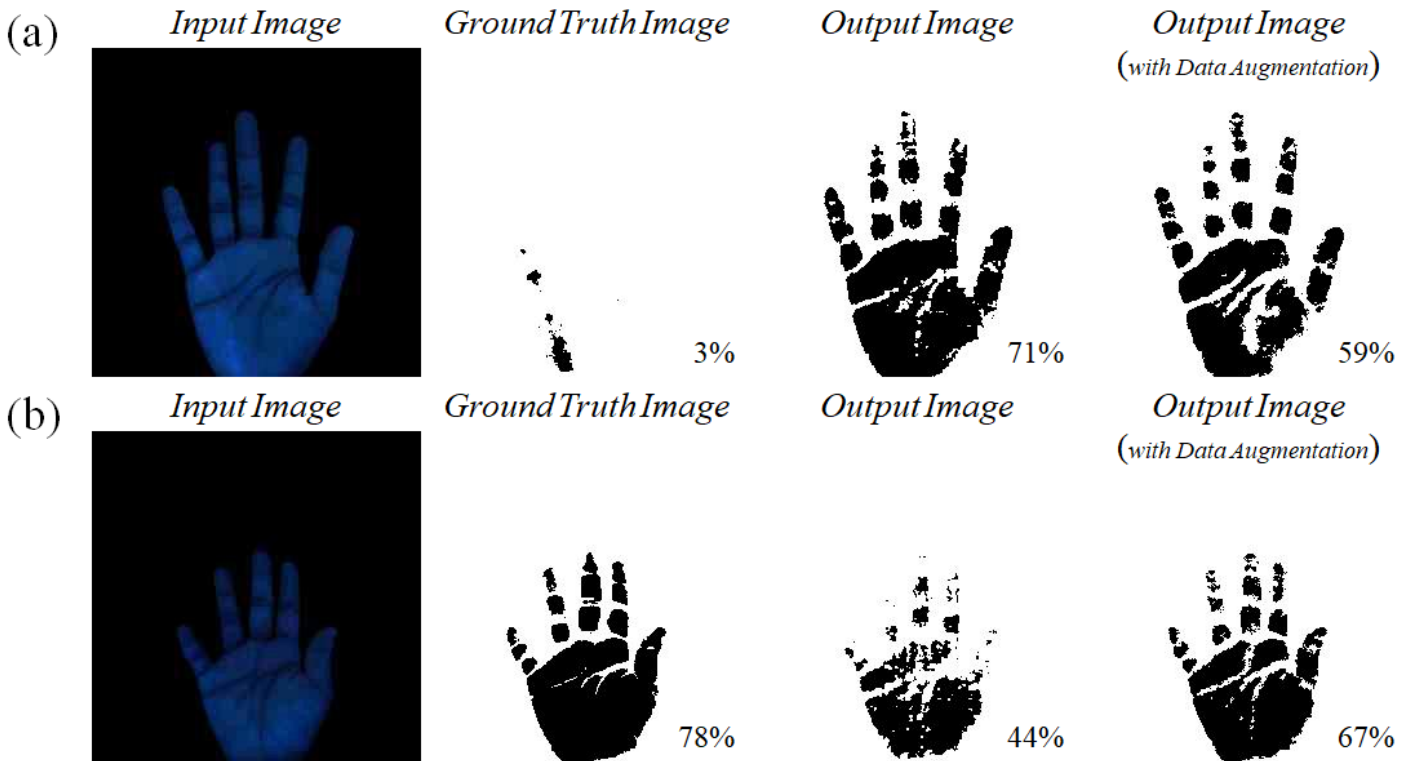


Figure 8: Examples of verification images that were correctly discriminated only in the training model with data augmentation
 (a) Input image with 35% or less of residue left, (b) Input image with 65% or more of residue left

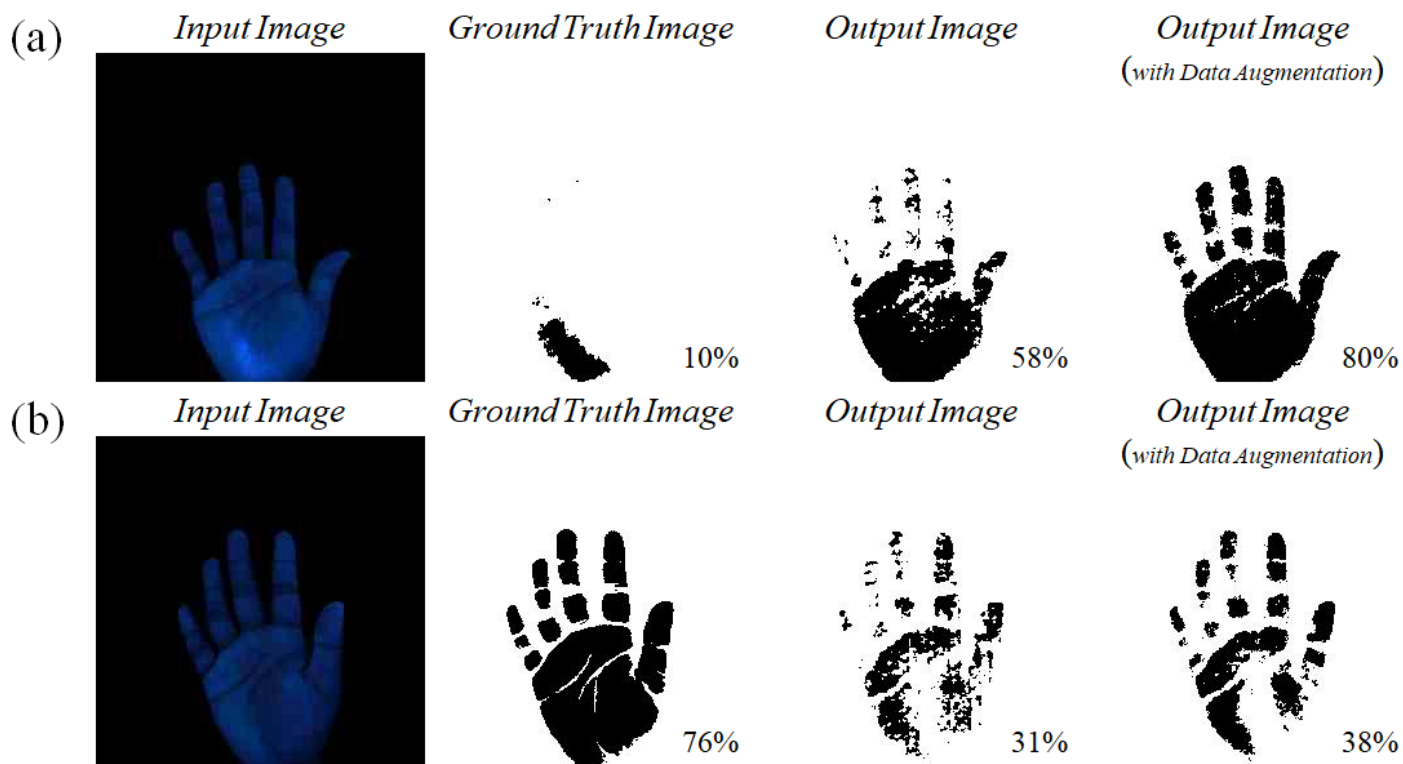


Figure 9: Examples of verification images that could not be correctly discriminated in both training models

(a) Input image with 35% or less of residue left, (b) Input image with 65% or more of residue left

5. Discussions

WHO guidelines recommend direct observation by trained staffs as a general hand hygiene evaluation method. However, certain factors have been identified as limitations to this method, such as the requirement to secure trained staff and the amount of time required to observe several scenes in this method. In addition, because the direct observation method involves subjective evaluations by the staff, results obtained may vary depending on the skill level of the staff. Therefore, based on the perspective of trained staffs, we developed a hand hygiene evaluation system in this study to evaluate palm images after hand-washing using pix2pix, which is a type of generative adversarial network (GAN). The input image adopted for pix2pix was a black light image obtained after hand-washing, and the ground truth image employed was a binarized image obtained by extracting the residue left from the input image by a trained staff. Regarding the generated estimated image, an image similar to the ground truth image was generated when the verification image was input to the pix2pix training model. In addition, the discrimination accuracy of the verification image was improved in the training models via the data augmentation method. The percentage of residue left was 14% in the ground truth image of Fig. 7a, whereas it was 10% in the estimated image. In contrast, the percentage of residue left for the estimated image was improved to 16% via the data augmentation method. Similarly, the percentage of residue left was 87% in the ground truth image of Fig. 7b, whereas it was 79% in the estimated image. The percentage of residue left in the estimated image was improved to 82% via the data augmentation method

Fig. 8 presents an example in which the percentage of residue left in the estimated image was only correctly discriminated in the training model after data augmentation. In both Figs. 8a and b, the percentage of residue left in the estimated image exhibited a tendency to approach the percentage of residue left in the ground truth image via the data augmentation method. In this pix2pix method, it was verified that the learning rate is improved by eliminating the bias in the percentage of residue left in the ground truth image of the training models. In contrast, Fig. 9 presents an example in which the percentage of residue left in the estimated image could not be discriminated correctly in both training models. In Fig. 9b, the percentage of residue left in the estimated image exhibited a tendency to approach the ground truth image via the data augmentation method; however, in Fig. 9a, it exhibited a tendency to move away from the ground truth image. There are two possible reasons for this phenomenon. First, the palm image in Fig. 9a might have been insufficiently shaded at the time of shooting. After inserting the right arm, a blackout curtain was adopted to shade the entrance at the time of shooting. However, the parts without the light emission of the fluorescent lotion were emphasized owing to insufficient shading, which may be ascertained to be the residue left. Second, it is possible that the training model could not be optimally created in the training data with a small percentage of residue left. In the training data with a significant percentage of residue left, the entire palm is painted black, such that the effect of residue left is negligible owing to these parts. However, in the training data with a negligible percentage of residue left, the position of the residue left differs for each part, which indicates the possibility of insufficient training. Fig. 6 presents the percentage of residue left in the verification and estimated images. In the image group with a

significant percentage of residue left, the variation in the data is negligible, and the estimated image often takes a value approximate to that of the ground truth image. However, in the image group with a negligible percentage of residue left, the variation in the data is significant and the values are also scattered. In other words, it is necessary to further increase the number of data used for training data with a small percentage of residue left. However, in the two-class discriminant analysis with the Mahalanobis distance, the misjudgment of both image groups occurred in one image each, and the analysis exhibited high discriminant accuracy.

6. Conclusions

Several infection control measures have been adopted in the medical field to address the unprecedented infectious disease called the novel coronavirus. Among them, the most important and basic approach to preventing infection is hand hygiene. Conventionally, for hand hygiene methods, researchers refer to the guidelines issued by CDC and WHO. However, the direct observation method, which is a subjective method, has been proposed as the gold standard for hand hygiene evaluation. Therefore, in this study, we developed a hand hygiene evaluation system using pix2pix as a simple and quantitative hand hygiene evaluation method. In this system, using the ground truth image of pix2pix from the perspective of trained staff, we addressed challenges such as securing trained staff and the protracted time required for evaluation, which are the limitations to the conventional direct observation method. Furthermore, in the two-class discriminant analysis for the presence or absence of residue left, misjudgment occurred in one image each for both image groups, and the analysis exhibited high discriminant accuracy. In the future, we will aim to apply the proposed system as the primary screening for hygienic hand-washing by further increasing the training data and improving discrimination accuracy. In addition to the medical field, we will apply this system in the educational field by investigating methods that do not require fluorescent lotion or black light.

Conflict of Interest

The authors declare no conflict of interest.

References

- [1] T. Kikuchi, "COVID-19 outbreak: An elusive enemy," *Respiratory Investigation*, **58**(4), 225–226, 2020, doi: 10.1016/j.resinv.2020.03.006.
- [2] H. Misao, "Hand hygiene concept: unchanging principles and changing evidence," *The Journal for Infection Control Team*, **11**(1), 7–12 2016.
- [3] J. D. Katz, "Hand washing and hand disinfection," *Anesthesiology Clinics of North America*, 2004, doi: 10.1016/j.atc.2004.04.002.
- [4] A. G. Labarraque, "Instructions and Observations concerning the use of the Chlorides of Soda and Lime," *The American Journal of the Medical Sciences*, 1831, doi: 10.1097/00000441-183108150-00021.
- [5] K. Yano, "Novel coronavirus (COVID-19) infection measures," *Infection and Antimicrobials*, **23**(3), 176-181, 2020.
- [6] J. M. Boyce, D. Pittet, "Guideline for hand hygiene in health-care settings: Recommendations of the Healthcare Infection Control Practices Advisory Committee and the HICPAC/SHEA/APIC/IDSA Hand Hygiene Task Force," *American Journal of Infection Control*, 2002, doi: 10.1067/mic.2002.130391.
- [7] World Health Organization, "WHO guidelines on hand hygiene in health care," World Health, 2009.
- [8] M. Ichinohe, "Basics of hand hygiene technique: how to choose a method that suits the on-site situation," *Journal for Infection Control Team*, **11**(1), 19–25, 2016.
- [9] V. Erasmus et al., "Systematic Review of Studies on Compliance with Hand Hygiene Guidelines in Hospital Care," *Infection Control & Hospital*

- Epidemiology*, **31**, 283–294, 2010, doi: 10.1086/650451.
- [10] T. Kato, "Systematic Assessment and Evaluation for Improvement of Hand Hygiene Compliance," *Journal of Environmental Infections*, **30** (4), 274–280, 2015.
- [11] M. Sato, R. Saito, "Nursing Students' Knowledge of Hand Hygiene and Hand Hygiene Compliance Rate during On-site Clinical Training," *Japanese Journal of Environmental Infections*, **34**(3), 182–189, 2019.
- [12] U.S. Food and Drug Administration (FDA), "Guidelines for effectiveness testing of surgical hand scrub (global juice test)," *Federal Register*, **43**, 1242–1243, 1978.
- [13] Y. Hirose, H. Yano, S. Baba, K. Kodama, S. Kimura, "Educational Effect of Practice in Hygienic Handwashing on Nursing Students: Using a device that allows immediate visual confirmation of finger contamination," *Environmental Infections*, **14**(2), 123–126, 1999.
- [14] K. Murakami, S. Umesako, "Study of Methods of Examination by the Full-hand Touch Plate Method in Food Hygiene," *Japanese Journal of Environmental Infections*, **28**(1), 29–32, 2013, doi: 10.4058/jsei.28.29.
- [15] M. Yoshii, K. Yamamoto, H. Miyahara, F. Kinoshita, H. Touyama, Consideration of homology between the visual sketches and photographic data of nursing student's hand contamination, *International Council of Nursing Congress*, 2019.
- [16] T. Shinozaki, "Recent Progress of GAN: Generative Adversarial Network," *Artificial Intelligence*, **33**(2), 181–188, 2018, doi: 10.11517/jjsai.33.2_181
- [17] I. J. Goodfellow et al., "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, 2672–2680, 2014, doi: 10.3156/jsoft.29.5_177_2.
- [18] P. Isola, J. Y. Zhu, T. Zhou, A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 5967–5979, 2017, doi: 10.1109/CVPR.2017.632.
- [19] Y. Taigman, A. Polyak, L. Wolf, "Unsupervised cross-domain image generation," 2017.
- [20] M. Y. Liu, T. Breuel, J. Kautz, "Unsupervised image-to-image translation networks," in *Advances in Neural Information Processing Systems*, 701–709, 2017.
- [21] <https://github.com/tensorflow/docs/blob/master/site/en/tutorials/generative/pix2pix.ipynb>
- [22] <https://jp.mathworks.com/help/images/detecting-a-cell-using-image-segmentation.html?lang=en>
- [23] C. Shorten, T. M. Khoshgofaar, "A survey on Image Data Augmentation for Deep Learning," *Journal of Big Data*, **6**(1), 2019, doi: 10.1186/s40537-019-0197-0.
- [24] G. J. McLachlan, "Mahalanobis distance," *Resonance*, **4**(6), 20–26, 1999, doi: 10.1007/bf02834632.