UWash: Promoting Hand Hygiene with Smartwatches in Daily Life

Anonymous submission

Abstract

Hand hygiene is one of the most efficient daily actions to prevent infectious diseases, such as Influenza, Malaria, and skin infections. We have been suggested to wash our hands under professional guidelines to prevent virus infection. However, several surveys show that very few people, even during the recent COVID-19 pandemic, follow this suggestion. Thus we propose UWash, a wearable solution with smartwatches, to assess handwashing procedures for the purpose of raising users' awareness and cultivating habits of high-quality handwashing. We address the task of handwashing assessment from readings of motion sensors similar to the action segmentation problem in computer vision, and propose a simple and lightweight two-stream UNet-like network to achieve it effectively. Experiments over 51 subjects show that UWash achieves an accuracy of 92.27% on handwashing gesture recognition, <0.5 *seconds* error on onset/offset detection, and <5 *points* error on gesture scoring in the user-dependent setting, and keeps promising in the user-independent evaluation and the user-independent-location-independent evaluation. UWash even performs well on 10 random passersby in a hospital 9 months later. UWash is the first work that scores the handwashing quality by gesture sequences and is instructive to guide users in promoting hand hygiene in daily life.

Introduction

Hand hygiene is an efficient and effective approach to preventing various infectious diseases, e.g., colds and flu, enterovirus infections, and skin infections. During the recent COVID-19 pandemic, we conducted a questionnaire on handwashing knowledge and practices over 505 subjects across 26 provinces in China. The questionnaire shows that 96.04% of subjects have heard of World Health Organization (WHO) guidelines or 7-step guidelines but only 34.65% of subjects follow the guidelines, similar to the situation reported in two other surveys conducted in Germany (Mieth et al. 2021) and in Nigeria (Wada and Oloruntoba 2021). This situation may become more severe as the pandemic eases. Thus, it is critical to raise people's awareness and cultivate habits of high-quality handwashing in daily life.

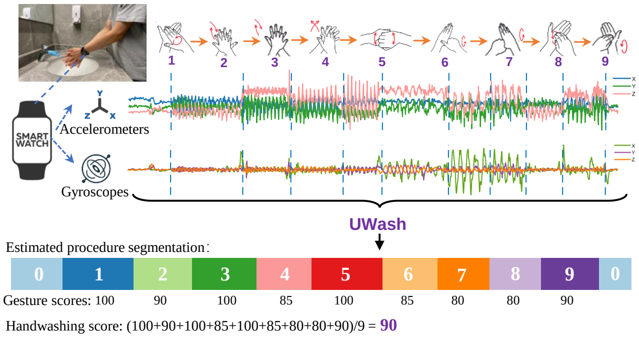


Figure 1: UWash utilizes accelerometers and gyroscopes of smartwatches to access handwashing procedures. With daily feedback, users can check their handwashing procedures and promote their handwashing gestures accordingly.

There are many automatic solutions to detect people's handwashing activity, however, these solutions are not applicable to promoting handwashing quality in daily life due to several limitations. (1) We may wash our hands at home, workplaces, restaurants, etc. However, most of the existing solutions, such as alcohol sensors (Edmond et al. 2010), pressure sensors (Kinsella, Thomas, and Taylor 2007), cameras (Haque et al. 2017), mmWave radars (Khamis et al. 2020), Bluetooth beacons (Zhong, Kanhere, and Chou 2016), RFID cards (Samyoun et al. 2021), and smart dispenser (Cao et al. 2021) only work at fixed positions where systems are deployed. (2) Many existing works simply count the occurrence of handwashing or report the duration of the entire process (Samyoun et al. 2021; Mondol and Stankovic 2020; Wang et al. 2020). This could remind users to wash hands, but cannot help users to promote their handwashing quality. (3) Without considering the limitations of RFWash (Khamis et al. 2020) working at fixed positions, RFWash can report the duration of each gesture shown in Fig. 1. However, from the reported duration, it is difficult for ordinary users to determine which handwashing gesture needs to be improved.

In this paper, we propose UWash. UWash leverages the inertial measurement unit (IMU) of smartwatches, i.e., accelerometers and gyroscopes for handwashing sensing. UWash adapts U-Net (Ronneberger, Fischer, and Brox 2015), a well-known pixel-wise classification network, to align every sampling point of the IMU readings with a handwashing gesture category. Fig. 1 shows an example of sample-wise gesture classification on a handwashing procedure, with which UWash is able to detect the start and end of handwashing, and estimate the duration of each gesture. To further assess the quality of each gesture from the estimated duration, we first carefully select 12 Youtube videos posted by medical workers that describe WHO handwashing guidelines. Then we take these videos as references and compute the recommended duration of each gesture. At last, we compare the estimated duration with the recommended duration to obtain the scores of each gesture as well as the entire handwashing procedure, with which, users can check their handwashing procedures and promote handwashing quality accordingly in daily life.

The technical novelty of UWash lies in the synergy achieved by picking the right techniques such as U-Net (Ronneberger, Fischer, and Brox 2015), Pyramid Pooling Module (Zhao et al. 2017) and channel-wise attention (Hu, Shen, and Sun 2018), and proposing a two-stream U-Net-like network for the two modal inputs from accelerometers and gyroscopes of smartwatches. The proposed methods maintain the following four main features.

- **Lightweight.** UWash is with 0.0012 GFlops, 0.099M Params, and 496KB model file size. UWash only requires 3% computation of MobileNetV3-small (Howard et al. 2019), a typical mobile-device-oriented model.
- **Simple.** UWash addresses the handwashing assessment problem as a sample-wise gesture classification problem. To make U-Net suitable for time-serial IMU readings, we simply replace its 2D operations, e.g., convolution and pooling, with 1D versions.
- **Effective.** Extensive evaluation results show UWash performs well in multiple settings, e.g., user-dependent, cross-user, cross-location, and cross-time.
- **One model for multiple tasks.** UWash models can automatically detect the start and end of the handwashing procedure, estimate the duration of each gesture, and score each gesture as well as the entire procedure.

The main contributions of this paper are as follows.

(1) We propose UWash to automatically assess handwashing procedures. We novelly regard the handwashing assessment task as a sample-wise gesture classification task, and design a U-Net variant to achieve it well.

(2) We leverage a simple approach to obtain a standard of handwashing following WHO guidelines. With the standard, UWash is the first work that can score the handwashing quality to guide users to improve their handwashing techniques.

(3) We collect a dataset with 51 subjects and 5 locations. Besides, we collect an additional dataset 9 months later in a hospital. We conduct extensive evaluation in settings of user-dependent, cross-user, cross-location, and cross-time. All datasets and codes are committed to release later.

Related Work

Hand hygiene has already been crucial to preventing healthcare-associated infections in hospitals. Human mandatory audits are applied to improve healthcare workers' compliance with the WHO guidelines. However, this approach is labor-intensive, time-consuming, and costly. Wearable devices such as wristbands (Li et al. 2018; Mondol and Stankovic 2020), armbands (Wang et al. 2020), and smartwatches (Mondol and Stankovic 2015; Samyoun et al. 2021; Cao et al. 2021) are also proposed to monitor handwashing in recent years. However, wearing wristbands and armbands in daily life is an extra interaction for users, limiting widespread use. Considering this, smartwatches are ideal platforms for monitoring handwashing procedures. Unfortunately, current smartwatch-based work cannot detect the onset/offset of handwashing, requiring to work along with Bluetooth sensors on dispensers (Mondol and Stankovic 2015; Cao et al. 2021) or to be awakened manually (Samyoun et al. 2021), where the former limits the use places and the latter harms the use frequency. Besides, we believe it will help people to improve handwashing techniques if a monitoring system can score handwashing procedures. Thus we propose UWash to achieve this.

UWash is realized by predicting the gesture category over every sample of continuous IMU readings. Generally, there are two main schemas for continuous gesture recognition over time-serial readings. (1) Top-down. This schema first segments the readings into slices according to the assumption that readings of two contiguous gestures are different and can be segmented, then applies gesture recognition on each slice. However, segmenting two contiguous gestures through the readings is challenging and always results in errors as explained in (Khamis et al. 2020). One commonly used strategy to relieve this is to set a pause gesture between two gestures to make the corresponding readings easy to segment (Ding et al. 2015). However, this strategy will raise great inconvenience to users. (2) Bottom-up. This schema leverages a sliding window. All samples in the window would be categorized as the same gesture with methods such as Hidden Markov Models (Li et al. 2018), Dynamic Time Warping (Akyazi et al. 2017), RNNs (Hou et al. 2019), CNNs (Zhang et al. 2019; Perslev et al. 2019), etc. As the window slides, entire readings will be categorized. This schema bypasses the segmenting problem in the top-down schema. However, since all samples in a window are categorized as the same gesture, this schema has an intrinsic false classification when the window spans readings of two contiguous gestures. To tackle the intrinsic false of the bottom-up schema, we apply UNet-like networks to conduct sampling-point-wise classification over each window.

Methods

Insights from Observation

The handwashing procedures vary from person to person, even from time to time for the same person, leading to diverse handwashing motion sequences. Conventional action recognition approaches use sliding windows with a data-dependent stride to crop clips from sequences. Then gesture classification is conducted clip-by-clip. To facilitate understanding, here we call the size of sliding windows the field of

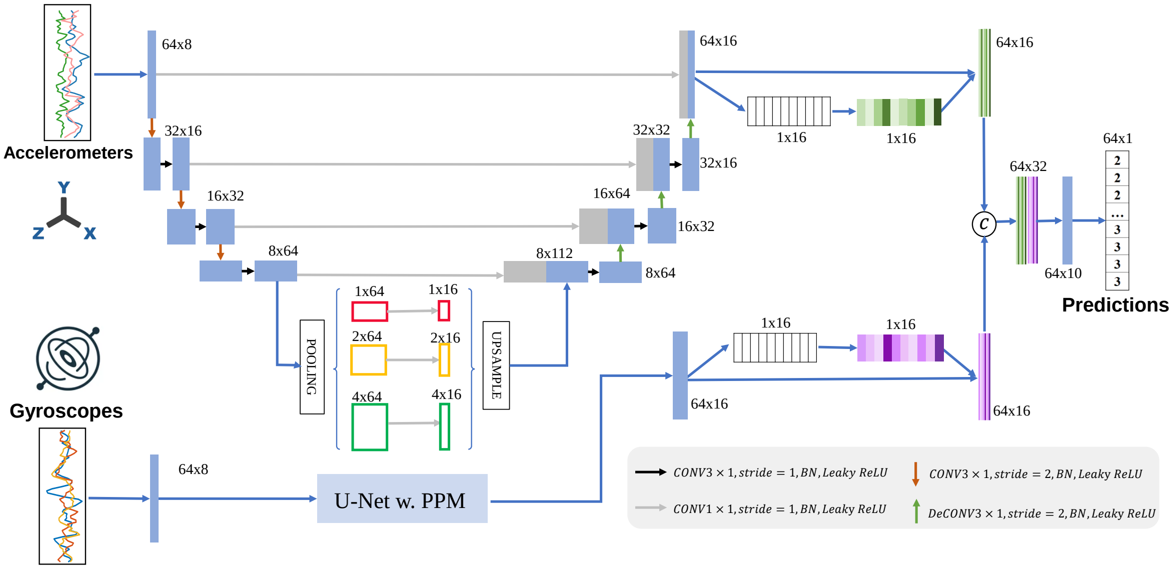


Figure 2: Two-stream U-Nets take data from two modality sensors, i.e., accelerometers and gyroscopes, as inputs, respectively. Feature maps from two streams are further concatenated in high-level layers for sample-wise gesture recognition.

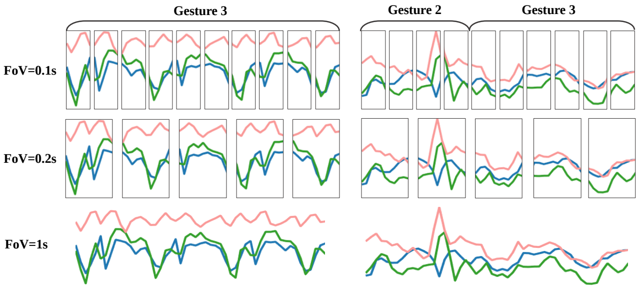


Figure 3: Dilemma. Handwashing patterns emerge in a larger field of view (left). However, in window-grained classification, a larger field of view may lead to a large classification error when handwashing gestures switch (right). Therefore, we apply the image semantic segmentation method for sample-wise classification in the window to bypass the dilemma.

view (FoV), which represents the receptive field that handwashing gesture recognition models can see at one time.

Selecting FoV in the conventional FoV-wise recognition schema is a work of dilemma. As shown in Fig. 3, the left three sub-figures are an example of recorded accelerometer sequences with the FoV of 0.1 *seconds*, 0.2 *seconds*, and 1 *second*, respectively. Let us see the upper-left first, we can hardly tell the gesture category from one single sequence, for these 10 sequences are quite different. If we take a twice-wider FoV of 0.2 *seconds*, the tendency of the sequences becomes much more clear, e.g., sequences in the 1st and 4th windows are similar; and those in the 2nd and 5th windows are similar. However, the blue/green curves in the 1st/4th window increase, while the blue/green curves in the 2nd/5th window decrease. This leads to an ambiguity in handwashing gesture recognition. Furthermore, if we take an even wider FoV of 1 *second*, the unified and periodical patterns in the sequence finally emerge, with which we can easily make an accurate handwashing gesture recognition.

- Pro. Larger FoV facilitates handwashing recognition.

However, a larger FoV may also cause a larger error. As shown in Fig. 3, the right three sub-figures demonstrate a procedure of gesture switching from gesture 2 (G2) to gesture 3 (G3), where G2 and G3 occupy the duration in 40% and 60% respectively. As G3 occupies the dominant duration, gesture recognition models tend to classify the sequences as G3, resulting in a recognition error of 40%.

- Con. Larger FoV may lead to larger recognition errors.

Our solution to the dilemma is to borrow the image semantic segmentation schema (Long, Shelhamer, and Darrell 2015) to the handwashing gesture recognition task. Instead of outputting one single gesture category for one entire FoV, the semantic segmentation schema predicts the gesture category for every single sample in an FoV. This schema takes advantage of a large FoV and makes fine-grained recognition for every sampling point in the FoV, bypassing the dilemma on FoVs naturally.

Deep Learning Model

U-Net is a widely used pixel-wise classification architecture in the image semantic segmentation task. We simply replace 2D convolutions, deconvolutions, and poolings in U-Net with 1D versions to conduct sample-wise gesture recognition on the 1D time-serial sensory recordings of smartwatches. The network architecture is shown in Fig. 2, reasons for several proposed adaptation explained next.

(1) Two-stream. Accelerometers and gyroscopes of smartwatches measure linear accelerations and angular accelerations respectively, describing two physical quantities

in different scales. To properly leverage data of the two modalities, we have to do data normalization before merging them for the later task. To conduct automatic modality normalization, we feed raw accelerometer data and raw gyroscope data into the two streams of U-Nets, respectively. We also apply Batch Normalization (Ioffe and Szegedy 2015) and Leaky Rectified Linear Unit (Maas et al. 2013) to facilitate the normalization between the two modalities.

(2) Pyramid Pooling Module(PPM) (Zhao et al. 2017). PPM is proved efficient to harvest features across different field of views. Therefore, we apply it at the middle of U-Nets. Before fed into PPM, feature maps are with size of 8×64, where 8 and 64 represent temporal dimension and channel dimension, respectively. We use three average pooling operations with window/stride sizes of 8, 4, and 2 on the feature maps, generating three outputs with the size of 1×64, 2×64, and 4×64, respectively. We then use convolutions with the kernel size of 1×1 to reduce the channel to be of 16, for the channel balance between the 3 pooling outputs and the input feature maps. Further, we upsample the pooling outputs and concatenate them with input feature maps, outputting feature maps with the size of 8×112.

(3) Squeeze-and-Excitation Module (Hu, Shen, and Sun 2018). Though features learned from the two-stream U-Nets are considered to be normalized, we still believe their contributions to the final gesture recognition are always not equal. For example, as shown in Fig. 1, accelerometers are more sensitive than gyroscopes for G1, while gyroscopes are more sensitive for G6. To re-weight the contribution on handwashing gesture recognition of the dual-modal sensors, we apply Squeeze-and-Excitation Modules on the learned features. After that, we concatenate the re-weighted features along the channel dimension for gesture recognition.

We use Pytorch 1.9.0 to implement the network. The initial learning rate is 0.001. The batch size is 16k. We use Cross-Entropy losses and Adam (Kingma and Ba 2014) to optimize the network. We train the network for 500 epochs.

Post Smoothing Methods

In testing, given a test time-series, we conduct gesture segmentation with the sliding window size of 64 and the *stride of 64*. An example result is shown in Fig. 4, which demonstrates some false classification. Thus, we further apply two post smoothing methods on the initial outputs.

(1) Multiple Test Voting (MTV). We conduct gesture recognition over the test time-series via the sliding window size of 64 with the *stride of 1*, resulting in multiple outputs on each sample. For each sample, we take the mode of all its outputs as its final recognition result.

(2) The Mode Filter (TMF). For each sample, we use the output mode of its nearest 128 samples as its final recognition result, a Mode Filter with window size of 128 and stride of 1.

A post-smoothed example is visualized in Fig. 4, which shows MTV and TMF can effectively reduce jitter errors and improve the sample-wise handwashing gesture recognition, numerical results reported in Evaluation Section.

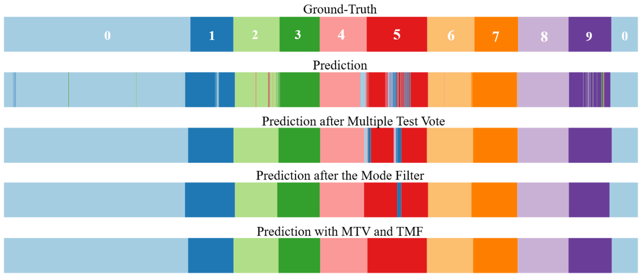


Figure 4: A post smoothing example. UWash can do sample-wise recognition well directly, while jitter errors cause the majority of the false recognition. To reduce these errors, we propose post smoothing methods including multiple test voting and the mode filter, which are simple but effective.

Handwashing Scoring

Scoring handwashing procedures can help users to improve their handwashing techniques accordingly. In this paper, we score handwashing procedures by considering the duration of each gesture with respect to the WHO guidelines. To obtain the guideline-recommended duration, we collect 60+ online videos that describe WHO guidelines and carefully select 12 of them as references, ignoring those with slow play, fast play, over-detailed explanation, etc. Table 1 shows the recommended duration of each handwashing gesture in these videos. In addition, we remove the maximum and minimum of each gesture and compute the average as the professional handwashing duration.

We follow two empirical assumptions. (1) Since each gesture emphasizes cleaning one part of hands, we assume each gesture is equally important. (2) The quality of cleaning under each gesture increases linearly with its duration, and the perfect quality is reached and saturated when the duration is equal to or greater than the professional duration. Given these assumptions, we score 9 handwashing gestures via

$$Score = \sum_{i=1}^{9} \frac{100}{9} * min(1, \frac{D_i^e}{D_i^p}) \qquad (1)$$

where $\frac{100}{9}$ is the peak score of each gesture, to match the first assumption; D_i^p and D_i^e represents the estimated professional duration and estimated of the i-th gesture, respectively; $min(1, \frac{D_i^e}{D_i^p})$ is to match the second assumption.

This handwashing scoring strategy works perfectly along with the sample-wise gesture classification method. That is, as shown in Fig. 4, UWash scores the handwashing procedures with the estimated duration of each gesture no matter the gesture conducting sequences. Besides, if users miss certain sub-actions in a handwashing session, UWash will give zero scores to those sub-actions accordingly.

Evaluation

We use Samsung Gear Sport smartwatch to record data of motion sensors and corresponding timestamps (Fomichev

No.	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**Avg.**
G1	~~3~~	4	4	4	5	6	4	5	3	~~9~~	7	7	4.9
G2	~~1.5~~	3.5	3	4.5	3	3.5	4.5	3.5	3.5	~~6.5~~	4.5	3	3.65
G3	~~1.5~~	3.5	3	4.5	3	3.5	4.5	3.5	3.5	~~6.5~~	4.5	3	3.65
G4	4	4	3	5	6	5	6	6	~~11~~	10	5	~~2~~	5.4
G5	~~2~~	4	3	5	5.5	3.5	5	3.5	3.5	~~8.5~~	4	3	4
G6	~~2~~	3	2	~~5.5~~	4.5	2.5	4	3.5	3.5	4	5	2.5	3.45
G7	~~2~~	3	2	~~5.5~~	4.5	2.5	4	3.5	3.5	4	5	2.5	3.45
G8	3	3	~~2.5~~	4.5	5.5	3.5	~~6.5~~	3	4	6	5	3.5	4.1
G9	3	3	~~2.5~~	4.5	5.5	3.5	~~6.5~~	3	4	6	5	3.5	4.1

Table 1: Professional duration of gestures is summarized from the recommendation of 12 online videos posted by healthcare institutions or experts. UWash compares the estimated procedure with the professional duration to score users' handwashing techniques.

et al. 2019). To increase the diversity of external conditions such as the type of hydrants, sinks, dispenses, etc., with IRB approval, we collect data at 5 buildings on a campus, i.e., a teaching hall, a laboratory hall, a cafeteria, a dormitory, and a library. At each building, we randomly recruit 10 passersby (11 at the laboratory hall) as participants and train them to wash their hands with WHO guidelines.

To act like the daily handwashing procedures, participants were asked, with smartwatches normally worn left, to conduct activities including walking to the sink, washing hands, walking out of the restroom, while other activities such as wetting hands with water, applying soap, drying hands with a towel are not mandatory, depending on their behaviors. We denote gestures in WHO guidelines as category 1 to category 9, and all other activities as category 0. Every participant repeats the procedure 5 times, which is a tolerable number and would not cause discomfort. Along with the sensor data, we also use codes from (Wang et al. 2019) to record videos on the participants' hands and corresponding timestamps. We watch the synchronized video streams to label gestures on motion sensory data. Further, we segment motion sensory records with the sliding window size of 64 and the stride of 1, leading to a simple ×64 data augmentation.

In all, the data acquisition process involves 51 participants and 5 locations, resulting in a dataset with 804991 IMU clips with the length of 64.

User-Dependent Results

We first evaluate UWash on all participants under the user-dependent setting. For each participant, we use IMU clips corresponding to the first 4 handwashing procedures as the training set, and the last ones as the test set, leading to the training and test set with clips of 643971 and 161020, respectively. In this paper, the training set and the test set have no overlap in all evaluations.

(1) Overall Results, Ablation Study, and SOTA. Table 2 shows the sample-wise classification accuracy, which is computed via Equation 2.

$$Accuracy = \frac{\sum_{p=1}^{51}\sum_{i=1}^{N_p} I(S^*_{p,i} == S_{p,i})}{\sum_{p=1}^{51}\sum_{i=1}^{N_p}} \qquad (2)$$

	Accuracy ↑	mPrecison ↑	mRecall ↑	mF1 ↑
UWash	86.31%	84.92%	84.48%	0.84
+MTV	91.10%	90.08%	89.68%	0.89
+TMF	91.13%	90.17%	89.45%	0.89
+MTV+TMF	**92.27%**	**91.26%**	**90.85%**	**0.91**
U-Net	82.25%	80.98%	80.23%	0.81
+TS	83.36%	81.96%	81.43%	0.82%
+TS+SE	84.16%	82.76%	81.87%	0.82%
+TS+SE+PPM	86.31%	84.92%	84.48%	0.84
MobileNetV3-s	82.95%	80.50%	79.88%	0.80%
ResNet-18	84.53%	82.40%	81.87%	0.82%

Table 2: Results on all participants. The results show that UWash performs well and can be further enhanced with post-smoothing methods of multiple test voting (MTV) and the mode filter (TMF). Moreover, UWash is only with ∼2.6-3.2% parameters of MobileNetV3-small and ResNet-18, much more lightweight. ↑ means the larger the better.

where N_p represents the length of the test series of the p-th participant; $S_{p,i}$ and $S^*_{p,i}$ represent the ground-truth and the prediction on the i-th sample of the p-th participant, respectively; $I(S^*_{p,i} == S_{p,i})$ outputs 1 if UWash recognizes correctly on $S_{p,i}$, otherwise 0. The table shows that UWash achieves the sample-wise classification accuracy of 86.31% directly. With simple post smoothing, i.e., multiple test voting (MTV) and the mode filter (TMF), UWash can eventually achieve an accuracy of 92.27%.

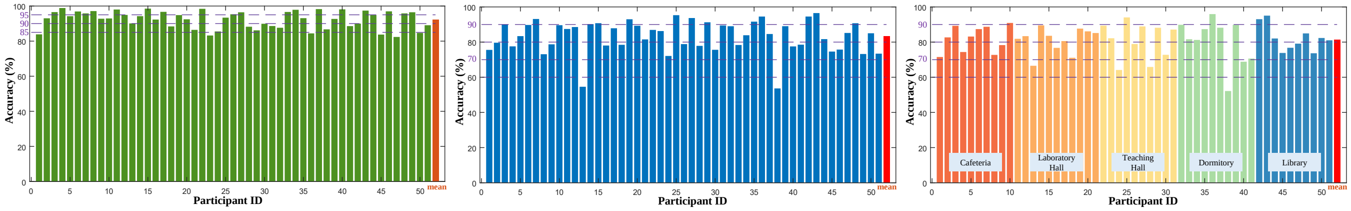
We also report the ablation study on our adapted methods, i.e., Two-Stream (TS), Squeeze-and-Excitation (SE) , and Pyramid Pooling Module (PPM) in Table 2. The table shows that these methods can effectively improve the accuracy of one-stream vanilla UNet by >4%.

We further apply representative MobilenetV3-small and ResNet-18 on the test clips to conduct clip-wise gesture classification. We take the clip-wise result as the result of all samples in the clip, and compute the sample-wise accuracy via Eq.2. Table 2 shows that UWash is ∼2-4% higher than MobilenetV3-small and ResNet-18. More importantly, MobilenetV3-small and ResNet-18 on our task are with 3.056M and 3.851M parameters respectively, while UWash is only with 0.099M parameters, ∼2.6-3.2% of these two networks, much more lightweight.

Table 3 compares our UWash with several recent work in view of accuracy, number of evaluated subjects, and granularity. UWash has been evaluated over the largest number of subjects, and achieves competitive accuracy with work that conducts clip-wise as well as sample-wise gesture classification.

Next, we are going to show the performance of UWash+MTV+TMF from the other four perspectives.

(2) Performance on Gestures. We show the confusion matrix of UWash+MTV+TMF on 10 gestures (9 handwashing gestures + 1 background) in Fig. 7. Though the data of these 10 gestures are not quite balanced (29.2% of background), UWash works well for all gestures, especially for the 2nd, 3rd, 6th, and 7th gestures. Besides, we find two types of false recognition are dominant. The first happens



(a) (b) (c)

Figure 5: (a) Accuracy over participants. (b) Cross-Participant Accuracy. (c) Cross-Participant-Cross-Location Accuracy.

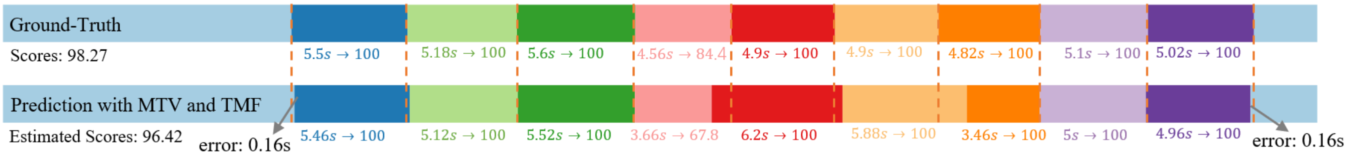


Figure 6: Visualization example of start/end detection and handwashing scoring.

Work	Accuracy	#Subject	Granularity
WristWash (Li et al. 2018)	95%	6	clip-wise
Awash (Cao et al. 2021)	92.94%	8	clip-wise
iWash (Samyoun et al. 2021)	92∼98%	14	clip-wise
RFWash (Khamis et al. 2020)	85%	10	sample-wise
Uwash(Ours)	92.27%	51	sample-wise

Table 3: Comparison with existing work in view of accuracy, number of evaluated subject, and granularity.

between the background and gestures of 1 and 9. We think this is because the background contains diverse activities such as wetting hands with water, applying soap, and drying hands with a towel, which may largely increase the difficulty to be classified correctly with its post-activity (G1) or pre-activity (G9). The other false happens between two successive gestures. As the start/end time of every gesture is annotated manually, the discordance across different labeling workers, participants, and locations may cause false between successive gestures.

(3) Performance on Participants. For the p-th participant, we use Equation 3 to compute the accuracy.

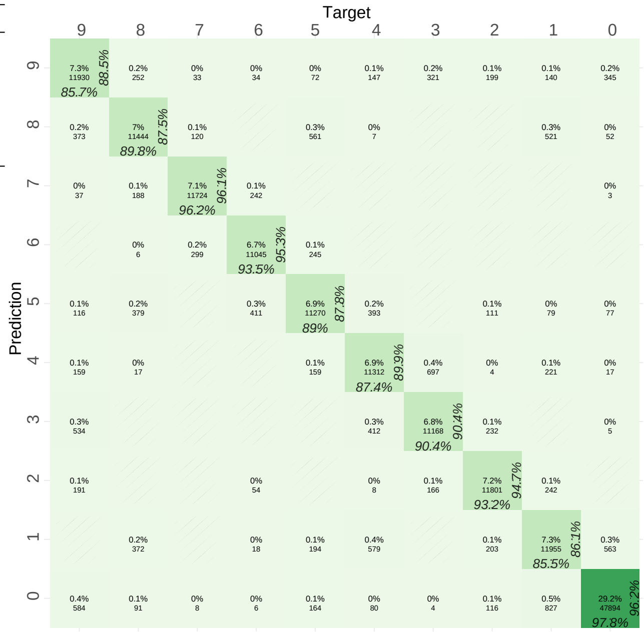
$$Accuracy_p = \frac{\sum_{i=1}^{N_p} I(S^*_{p,i} == S_{p,i})}{N_p} \tag{3}$$

where symbols share the same meanings with Equation 2.

Fig. 5 (a) shows that the accuracy of 33 participants is over 95%; only 6 of them have the accuracy of less than 85%; the average accuracy is 92.33% (orange bar). The results demonstrate that, for the seen participants, UWash achieves sample-wise handwashing gesture recognition effectively.

(4) Start/End Detection. Deserved to reiterate, UWash can automatically detect the handwashing start/end time, not requiring to work along with additional sensors (Samyoun et al. 2021; Zhong, Kanhere, and Chou 2016) or to be awakened manually (Samyoun et al. 2021). We use Equation 4 to compute the start time detection error.

$$Error_s = |t^*_s - t_s| \tag{4}$$



Prediction \ Target	9	8	7	6	5	4	3	2	1	0
9	7.3% 11930 88.5% 85.7%	0.2% 252	0% 33	0% 34	0% 72	0.1% 147	0.2% 321	0.1% 199	0.1% 140	0.2% 345
8	0.2% 373	7% 11444 87.5% 89.8%	0.1% 120		0.3% 561	0% 7			0.3% 521	0% 52
7	0% 37	0.1% 188	7.1% 11724 96.1% 96.2%	0.1% 242						0% 3
6		0% 6	0.2% 299	6.7% 11045 95.3% 93.5%	0.1% 245					
5	0.1% 116	0.2% 379		0.3% 411	6.9% 11270 87.8% 89%	0.2% 393		0.1% 111	0% 79	0% 77
4	0.1% 159	0% 17			0.1% 159	6.9% 11312 89.9% 87.4%	0.4% 697	0% 4	0.1% 221	0% 17
3	0.3% 534					0.3% 412	6.8% 11168 90.4% 90.4%	0.1% 232		0% 5
2	0.1% 191			0% 54		0% 8	0.1% 166	7.2% 11801 94.7% 93.2%	0.1% 242	
1		0.2% 372		0% 18	0.1% 194	0.4% 579		0.1% 203	7.3% 11955 86.1% 85.5%	0.3% 563
0	0.4% 584	0.1% 91	0% 8	0% 6	0.1% 164	0% 80	0% 4	0.1% 116	0.5% 827	29.2% 47894 96.2% 97.8%

Figure 7: Confusion Matrix of Gesture Classification. UWash works well for all gestures, especially for the 2nd, 3rd, 6th, and 7th gestures (zoom in for a better view).

where t^*_s and t_s represent the detection and the ground-truth of the start time of a test handwashing procedure, respectively; $|\cdot|$ is to compute the absolute value. Similarly, we use $|t^*_e - t_e|$ to compute the end time detection error. We report the detection error Mean and standard deviation (SD) of the test set in Table 4. The table shows that Means and SDs are all within 1 second, indicating UWash can detect handwashing events correctly and stably.

(5) Scoring Results. UWash is the first work to score handwashing following the WHO guidelines. We use methods described in Section to compute handwashing scores

	Start	End	Scoring
Mean ↓	0.49*s*	0.23*s*	4.0pts
SD ↓	0.58*s*	0.30*s*	4.3pts

Table 4: Error in Start/End Detection and Handwashing Scoring. ↓ means the less the better.

	CP	CP-CL	CP-CL-CT
Accuracy↑	83.34%	81.45%	82.08%
Mean(start error)↓	0.43*s*	0.48*s*	0.39*s*
SD(start error)↓	0.45*s*	0.53*s*	0.20*s*
Mean(end error)↓	0.23*s*	0.36*s*	0.14*s*
SD(end error)↓	0.33*s*	0.44*s*	0.09*s*
Mean(score error)↓	12.74pts	15.21pts	14.3pts
SD(score error)↓	8.01pts	9.26pts	5.23pts

Table 5: Cross-domain results. ↑ means the higher the better. ↓ means the less the better. CP, CP-CL, and CP-CL-CT are for Cross-Participant, Cross-Participant-Cross-Location, and Cross-Participant-Cross-Location-Cross-Time.

on the test set. Further, we compute the Mean and SD of scoring errors against ground truth. As shown in Table 4, the mean and the SD are less than 5 points, indicating UWash can score handwashing well.

In Fig. 6, we visualize how UWash works on a test series of the 26th participant. UWash outputs sample-wise gesture classification, with which we can detect the start/end time, estimate the duration of gestures, score gestures, and score the whole handwashing procedure. With UWash, users can promote their handwashing practice with the estimated scores accordingly in daily life.

Cross-Domain Results

(1) Cross-Participant. We train UWash with data of 50 out of 51 participants and test the trained model with data of the remaining participant. We conduct this leave-one-participant-out process over 51 participants respectively to evaluate the cross-participant performance. As shown in Fig. 5 (b), accuracy among some left-out participants is more discrete than those in the user-dependent setting, shown in Fig. 5 (a), e.g., the 13th and the 38th participants wash their hands with micro hand movement, which does not meet the WHO guidelines and results in an accuracy of <60%. In this case, UWash could remind users to improve their handwashing techniques. The mean accuracy in the cross-participant setting is 83.34% since the personalized gestures of these individuals are not included in the training phase.

Table 5 shows that UWash detects the start/end time of handwashing well even in the cross-participant setting, with errors of <0.5 *seconds*, revealing that UWash can effectively distinguish handwashing gestures and other activities. However, the handwashing scoring performance drops largely, repeatedly indicating the performance of gesture classification on unseen users highly depends on how well they wash their hands following WHO guidelines.

(2) Cross-Participant-Cross-Location. We use data from 4 out of 5 locations to train UWash and test the trained model on the remaining one location. We conduct this leave-one-location-out process over 5 locations respectively. Since recruited participants have no overlap between different locations, the leave-one-location-out process also leads to the evaluation in the cross-participant-cross-location setting.

The experimental results are shown results in Fig. 5 and Table 5. In this setting, the performances have similar characteristics to those in the cross-participant setting, e.g., the accuracy is more discrete; the 13th and the 38th participants have the lowest accuracy; the mean accuracy drops to 81.45%; the start/end time detection is achieved well.

We further find that the performance in the cross-participant setting is slightly better than those in the cross-participant-cross-location setting. We think this is because we use data from 50 participants to train UWash each time in the former setting, while in the latter setting, we use only data from 40 or 41 participants to train UWash each time.

We think the largest obstacle to UWash for wider-spread use is caused by cross-participant since the performance in the cross-participant-cross-location setting has a small decay compared to those in the cross-participant setting. Thus, we could enlarge the training set by recruiting more participants, which is not an impossible problem for the industry.

(3) Cross-Participant-Cross-Location-Cross-Time.

Nine months after the data collection, we further randomly recruit 10 passersby in a hospital to wash their hands 5 times, and follow the same process to obtain a new dataset. We apply the trained model in user-dependent results on the new dataset to evaluate the cross-time performance.

Table 5 shows the cross-time results. The average sample-wise gesture recognition accuracy is 82.08%. The average error in the start and end detection of handwashing are 0.39*s* and 0.14*s* respectively. The average standard deviation in the start and end detection of handwashing are 0.20*s* and 0.09*s* respectively. The average scoring error and scoring standard deviation are 14.30 points and 5.23 points respectively. We find that these average values are at the same level as in the Cross-Participant-Cross-Location evaluation listed in Table 5. These results demonstrate that UWash is stable and promising across time. Note that the new dataset is from the new participants, new location, and new date, results of this evaluation provide us strong confidence in performance if UWash is promoted to large-scale real-world use.

Conclusion

We present UWash, a handwashing assessment system, to raise people's awareness of handwashing in daily use and adherence to the WHO handwashing guidelines. UWash takes the readings of accelerometers and gyroscopes of smartwatches as inputs, feeds the inputs into a two-stream U-Net, and outputs sample-wise gesture recognition results effectively. UWash can detect handwashing start/end time, estimate the duration of every handwashing gesture, and score gestures as well as the entire procedure following WHO guidelines. Experimental results over 51 participants show that UWash works well in the user-dependent, cross-participant, and cross-participant-cross-location settings. Moreover, UWash is lightweight, only 496KB, with great potential to be deployed on edge devices in the future.

References

Akyazi, O.; Batmaz, S.; Kosucu, B.; and Arnrich, B. 2017. SmokeWatch: A smartwatch smoking cessation assistant. In *2017 25th Signal Processing and Communications Applications Conference (SIU)*, 1–4. IEEE.

Cao, Y.; Chen, H.; Li, F.; Yang, S.; and Wang, Y. 2021. Awash: handwashing assistance for the elderly with dementia via wearables. In *IEEE INFOCOM*.

Ding, H.; Shangguan, L.; Yang, Z.; Han, J.; Zhou, Z.; Yang, P.; Xi, W.; and Zhao, J. 2015. Femo: A platform for free-weight exercise monitoring with rfids. In *ACM Sensys*.

Edmond, M.; Goodell, A.; Zuelzer, W.; Sanogo, K.; Elam, K.; Bearman, G.; et al. 2010. Successful use of alcohol sensor technology to monitor and report hand hygiene compliance. *Journal of Hospital Infection*, 76(4): 364–365.

Fomichev, M.; Maass, M.; Almon, L.; Molina, A.; and Hollick, M. 2019. Perils of zero-interaction security in the Internet of Things. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3(1): 1–38.

Haque, A.; Guo, M.; Alahi, A.; Yeung, S.; Luo, Z.; Rege, A.; Jopling, J.; Downing, L.; Beninati, W.; Singh, A.; et al. 2017. Towards vision-based smart hospitals: a system for tracking and monitoring hand hygiene compliance. In *Machine Learning for Healthcare Conference*, 75–87. PMLR.

Hou, J.; Li, X.-Y.; Zhu, P.; Wang, Z.; Wang, Y.; Qian, J.; and Yang, P. 2019. Signspeaker: A real-time, high-precision smartwatch-based sign language translator. In *ACM MobiCOM*.

Howard, A.; Sandler, M.; Chu, G.; Chen, L.-C.; Chen, B.; Tan, M.; Wang, W.; Zhu, Y.; Pang, R.; Vasudevan, V.; et al. 2019. Searching for mobilenetv3. In *Proceedings of the IEEE/CVF international conference on computer vision*, 1314–1324.

Hu, J.; Shen, L.; and Sun, G. 2018. Squeeze-and-excitation networks. In *CVPR*.

Ioffe, S.; and Szegedy, C. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *ICML*.

Khamis, A.; Kusy, B.; Chou, C. T.; McLaws, M.-L.; and Hu, W. 2020. RFWash: a weakly supervised tracking of hand hygiene technique. In *ACM Sensys*.

Kingma, D. P.; and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kinsella, G.; Thomas, A.; and Taylor, R. 2007. Electronic surveillance of wall-mounted soap and alcohol gel dispensers in an intensive care unit. *Journal of Hospital Infection*, 66(1): 34–39.

Li, H.; Chawla, S.; Li, R.; Jain, S.; Abowd, G. D.; Starner, T.; Zhang, C.; and Plötz, T. 2018. Wristwash: towards automatic handwashing assessment using a wrist-worn device. In *Proceedings of the 2018 ACM international symposium on wearable computers*, 132–139.

Long, J.; Shelhamer, E.; and Darrell, T. 2015. Fully convolutional networks for semantic segmentation. In *CVPR*.

Maas, A. L.; Hannun, A. Y.; Ng, A. Y.; et al. 2013. Rectifier nonlinearities improve neural network acoustic models. In *ICML*.

Mieth, L.; Mayer, M. M.; Hoffmann, A.; Buchner, A.; and Bell, R. 2021. Do they really wash their hands? Prevalence estimates for personal hygiene behaviour during the COVID-19 pandemic based on indirect questions. *BMC public health*, 21(1): 1–8.

Mondol, M. A. S.; and Stankovic, J. A. 2015. Harmony: A hand wash monitoring and reminder system using smart watches. In *proceedings of the 12th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*.

Mondol, M. A. S.; and Stankovic, J. A. 2020. HAWAD: Hand Washing Detection using Wrist Wearable Inertial Sensors. In *2020 16th International Conference on Distributed Computing in Sensor Systems (DCOSS)*, 11–18. IEEE.

Perslev, M.; Jensen, M.; Darkner, S.; Jennum, P. J.; and Igel, C. 2019. U-time: A fully convolutional network for time series segmentation applied to sleep staging. *Advances in Neural Information Processing Systems*, 32.

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, 234–241. Springer.

Samyoun, S.; Shubha, S. S.; Mondol, M. A. S.; and Stankovic, J. A. 2021. iWash: A smartwatch handwashing quality assessment and reminder system with real-time feedback in the context of infectious disease. *Smart Health*, 19: 100171.

Wada, O. Z.; and Oloruntoba, E. O. 2021. Safe reopening of schools during COVID-19: an evaluation of handwash facilities and students' hand hygiene knowledge and practices. *European Journal of Environment and Public Health*, 5(2).

Wang, C.; Sarsenbayeva, Z.; Chen, X.; Dingler, T.; Goncalves, J.; and Kostakos, V. 2020. Accurate measurement of handwash quality using sensor armbands: Instrument validation study. *JMIR mHealth and uHealth*, 8(3).

Wang, F.; Zhou, S.; Panev, S.; Han, J.; and Huang, D. 2019. Person-in-WiFi: Fine-grained person perception using WiFi. In *ICCV*.

Zhang, Y.; Zhang, Z.; Zhang, Y.; Bao, J.; Zhang, Y.; and Deng, H. 2019. Human activity recognition based on motion sensor using u-net. *IEEE Access*, 7: 75213–75226.

Zhao, H.; Shi, J.; Qi, X.; Wang, X.; and Jia, J. 2017. Pyramid scene parsing network. In *CVPR*.

Zhong, H.; Kanhere, S. S.; and Chou, C. T. 2016. WashInDepth: Lightweight hand wash monitor using depth sensor. In *Proceedings of the 13th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, 28–37.