IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 13, 2020, accepted September 30, 2020, date of publication October 13, 2020, date of current version November 12, 2020.

*Digital Object Identifier 10.1109/ACCESS.2020.3030723*

# Selective Content Removal for Egocentric Wearable Camera in Nutritional Studies

**MOHAMED ABUL HASSAN AND EDWARD SAZONOV, (Senior Member, IEEE)**
Department of Electrical and Computer Engineering, The University of Alabama, Tuscaloosa, Al 35401, USA

Corresponding author: Edward Sazonov (esazonov@eng.ua.edu)

This work was supported by the National Institute of Diabetes and Digestive and Kidney Diseases under Grant R01DK100796.

**ABSTRACT** Automatic Ingestion Monitor v2 (AIM-2) is an egocentric camera and sensor that aids monitoring of individual diet and eating behavior by capturing still images throughout the day and using sensor data to detect eating. The images may be used to recognize foods being eaten, eating environment, and other behaviors and daily activities. At the same time, captured images may carry privacy concerning content such as (1) people in social eating and/or bystanders (i.e., bystander privacy); (2) sensitive documents that may appear on a computer screen in the view of AIM-2 (i.e., context privacy). In this paper, we propose a novel approach based on automatic, image redaction for privacy protection by selective content removal by semantic segmentation using a deep learning neural network. The proposed method reported a bystander privacy removal with precision of 0.87 and recall of 0.94 and reported context privacy removal by precision and recall of 0.97 and 0.98. The results of the study showed that selective content removal using deep learning neural network is a much more desirable approach to address privacy concerns for an egocentric wearable camera for nutritional studies.

**INDEX TERMS** Privacy, egocentric wearable camera, bystander privacy, context privacy, lifelogging, monitoring of ingestive behavior, food intake, diet, nutritional studies.

## I. INTRODUCTION

Wearable sensor technology is an exponentially growing field, largely focused on health and wellness. Currently, the wearable sensor technology sector reports a global market value of 24.2 billion USD, with a projected growth of 100 billion USD by 2024 [1]. Consumer wearable sensors expand from smartwatches and wristbands to jewelry, glasses, and clothing. Each wearable sensor is providing the consumer with a unique aid, as illustrated in Fig. 1 and Fig. 2. A wearable sensor would commonly contain one or more of the following components: accelerometers, gyroscopes, GPS, physiological sensors, and cameras.

These components acquire personal information related to the motion, locations, vitals, and images of/from the wearer. The obtained data is used to provide aid in applications such as lifelogging, health monitoring, person tracking, leisure, and games/sports. However, continuous acquisition of personal information raises concerns towards the privacy of an individual. Surveys [2]–[4] have reported user privacy concerns such as privacy related to social implication, criminal abuse, facial recognition, access control, surveillance, and sousveillance (recording by wearable cameras), and speech disclosure. In the context of information technology

The associate editor coordinating the review of this manuscript and approving it for publication was Jenny Mahoney.

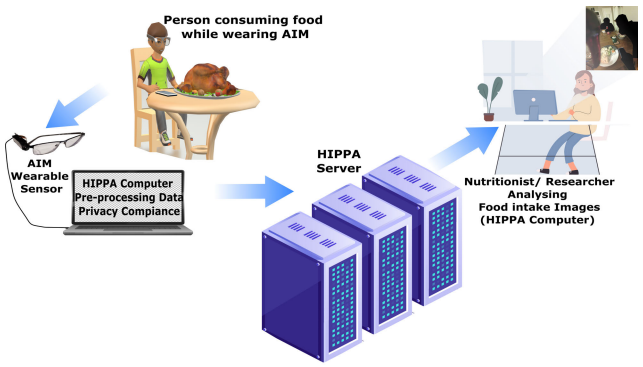



**FIGURE 1.** Overall Process of food intake analysis using AIM wearable sensor.

professionals, these privacy concerns are broadly classified into three groups: context privacy, bystander privacy, and external data sharing privacy [2], [5].

Context privacy is related to access control, location disclosure, and speech disclosure. For example, a wearable lifelogging sensor may acquire the wearer's speech, images of activities being performed, and location, which the wearer might not want to share with service providers. Bystander privacy is an issue of protecting the privacy of third parties in the surroundings of the wearer. Wearable sensors with microphone and camera might capture speech and images of

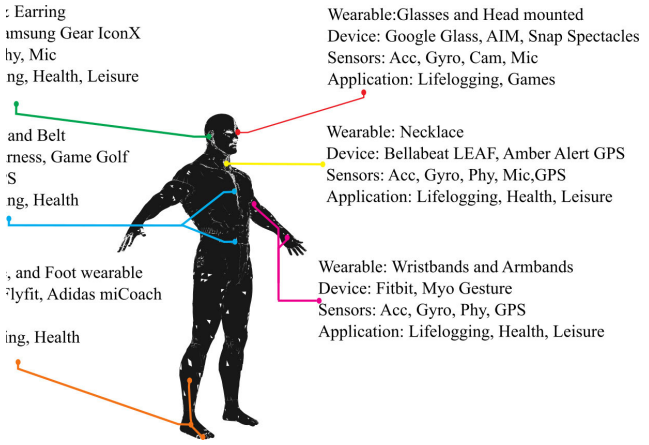



**FIGURE 2. Schematic review on wearable sensor technology, primary sensors, and application [2]. Acc-Accelerometer, Gyro-Gyroscope, Phy-Physiological sensor, Cam-Camera. Citations: Google Glass [6], AIM [12], Snap Spectacles [11], Bragi Dash [39], Samsung Gear IconX[40], Zephyr Bioharness [41], Bellabeat LEAF [42], Amber Alert GPS [43], BSXinsight [44], Addidas miCoach [45], Fitbit [46], MyoGesture [47].**

bystanders in the surroundings, leading to privacy concerns related to speech disclosure, facial recognition, surveillance, and sousveillance. External data sharing privacy is related to the security of the data that is forwarded to cloud service that store, analyzes, and provides feedback to the wearer.

In this paper, we will further discuss privacy issues and solutions related to head-mounted, first-view egocentric cameras. Head-mounted wearable sensors (see Fig. 1) are often worn as eyewear. Google Glass [6], SenseCam [7], Epson Moverio [8], AIM [9], Vuzix [10], and Snap spectacle [11], may be used for health and wellness, lifelogging, leisure, and games. These wearable sensors comprise of either or a combination of several sensors (e.g., accelerometer, gyroscope, and microphone, and an egocentric (first-person view) camera. The camera is the most commonly used component for a head-mounted wearable sensor. Images captured from the egocentric camera stand as the primary source of information since the camera carries the wearer's visual perspective.

Automatic Ingestion Monitor V2 (AIM-2) is a sensor for monitoring food intake. The eating episodes are recognized from the signal of embedded accelerometer [12], prompting the egocentric camera to capture images of the food being eaten. However, individuals often eat in groups, or socially and studies [13] show that individuals are likely to use their mobile phone/computer while eating a meal (see Fig. 3).

Furthermore, in some studies AIM may be configured to capture images continuously throughout the day. Regardless of the mode of image collection, these egocentric images may contain (1) the surrounding environment, including people around the wearer; (2) sensitive information such as the content of the computer screens. The collected images may be a source of privacy leakage, raising issues of bystander privacy and context privacy. Specifically, in the United States, such images are recognized as information protected under the Health Insurance Portability and Accountability Act (HIPPA), Title II, Privacy Rule [14].

The images may contain HIPAA-protected information, such as account numbers on a computer screen, and full-face images of people. While the images are stored in a HIPAA-compliant repository and reviewed by certified nutritionists, who are not considered an explicit adversary, the risk of accidental privacy leakage still remains. Furthermore, the protected information cannot be shared. The HIPAA rule suggests deidentification of the protected information through removal of specific identifiers, such as names, phone and account numbers, and full-face photos.

Traditionally HIPPA compliance may be addressed by discarding the images with the privacy concerns [3] or scaling down the image to extremely low resolution [15]. However, in nutritional studies, discarding or scaling down the images may very likely result in complete or partial loss of information about food intake. Therefore, we opted to perform image redaction-based privacy protection by selectively removing content related to bystander privacy and context privacy [16] while preserving information related to food intake. The immediate goal is to eliminate the major sources of potential privacy leakage: screens and persons, as the first step to the eventual goal of achieving HIPAA-compliant de-identification of the images acquired in nutritional studies.

Therefore, we propose an automatic, selective content removal method to exclude bystander and context privacy content. The rest of the paper is as follows. Section II reviews current privacy content removal methods used for the head-mounted wearable sensor. Section III discusses the proposed approach used to address the privacy content removal from images of the head-mounted wearable sensor. Section IV discusses the results and analysis of the proposed approach for inappropriate content removal. Section V and VI derives the conclusion of the study, limitations, and future work.

## II. PRIVACY CONTENT REMOVAL METHODS

Privacy content removal fundamentally prevents information that an individual wants to be kept privet from being public [17]. Researchers in the field of data privacy protection for the context of image and videos as visual privacy protection. Visual privacy protection is widely performed as (i) intervention, (ii) blind vision, (iii) secure processing, (iv) redaction, (v) data hiding. An intuitive review of the privacy protection approaches can be found in [18]. We notice that the majority of the wearable sensor community adapted a redaction-based privacy protection approach. Here private information was concealed by modifying or removing sensitive image regions such as face, bodies, screens, etc. in the images being reviewed by nutritionists.

Traditionally, researchers in the wearable sensor community have opted for privacy content removal relies either on manual review or on automatic image classification. The wearer performs the manual review by removing images manually and discarding the privacy content images [19]. This is a tedious approach since the wearer may have to manually review hundreds to thousands of images daily to maintain

**FIGURE 3.** Illustration of an eating episode in which the individual consumes the meal in a group while using a mobile phone; prompting bystander, and context privacy concerns.

their privacy. The automatic image classification-based approach comes as a solution to the manual review approach.

At present, DeepAI [20], and Sightengine [21] performs automatic removal of content such as nudity detection, weapons, alcohol, drugs, offensive content, and hate signs by using deep convolution neural networks and discarding images. However, image removal is highly undesirable in food intake monitoring since it may lead to complete loss of information in many situations of social eating and eating in the presence of computer screens.

Researchers from the wearable sensor community attempted to address privacy issues. Thomaz *et al.* [22], proposed to address the bystander privacy issue by using a privacy saliency matrix approach. The authors used a combination of manual review by the wearer and image tagging to categorize the data into four quadrants, based on saliency, non-saliency, and privacy concern and non-concern. The bystander's faces were blurred by using a Haar cascade classifier. The privacy saliency matrix approach was mainly used to cluster the images prior to privacy content removal. The limitations of this method are the combination of manual review and image tagging approach, which is a tedious and time-consuming task. A privacy behavior study [23], on the wearable head-mounted sensors, reported that wearers prefer to manually control the image acquisition of the wearable camera to address privacy-related issues. This comes as the most straightforward solution; however, this approach brings in the human factor and the possibility of information loss.

Zarepour *et al.* [24] proposed a context privacy-preserving framework for a lifelogging sensor. The authors used a combination of human activity recognition, ambient environment detection, and sensitive subject detection modules to remove context privacy concerns. The human activity recognition module classified accelerometer data to understand human activity (i.e., walking, sitting, running).

Whereas, the ambient environment detection was performed to detect the environment (i.e., indoor/outdoor). This human activity recognition and ambient environment detection module were used to detect contextual cues. Based on the contextual cues, the authors performed sensitive subject detection to remove context privacy concerns by using an array of best-fitting-deep learning and computer vision algorithms such as AlexNet, FastRCNN, and Viola-Jones detector. The authors reported an accuracy of 70% during testing on 300 images.

Muchen *et al.* [25], proposed a privacy-preserving approach for the head-mounted wearable sensor using multi-sensor information from a mobile phone and smartwatch. Here the authors used the multi-sensor information to identify privacy concern scenarios and trigger the wearable camera.

## III. METHOD

Existing privacy content removal methods use semi-automated approaches by combining manual review, image tagging methods, and performing a multi-sensor fusion. The semi-automatic approach brings in human factors; whereas, the multisensory approach depends on the availability of the multiple sensors. Automated privacy content removal methods based solely on images would be much more desirable. Scene understanding using semantic segmentation is a widely used approach to identify contents within the image. Selectively identifying specific content within an image would improve privacy content recognition and the removal process.

Our proposed method aims to perform image redaction-based privacy protection by selectively removing privacy content (object and people) on free-living image data captured by AIM-2. We use a semantic segmentation-based approach to selectively remove bystander and context privacy content from images. Semantic segmentation performs a pixel-level classification of the content, compared to the object detection approach using bounding boxes. Object detection methods may capture the food content and background information within the bounding box used for content removal. This would lead to the loss of information related to the eating episode.

Multiple semantic segmentation models such as Unet [26], FCN [27], DeconvNet [28], SegNet [29] were studied. We chose to use SegNet, due to its ability to provide accurate semantic segmentation while operating with a simple architecture at high computational efficiency, and low memory usage. Computational efficiency is an essential aspect of the practical usage of this method since the AIM-2 is meant to operate throughout the day. The participants from the

study have shown an active usage period between 12.5 to 18.5 hours, capturing about 3000 to 4500 images per day.

### A. SegNet

SegNet comprises of 109 layers, following an encoder-decoder architecture. The primary emphasis of SegNet is to use the encoder layer for feature extraction and performing semantic segmentation at decoder layers. Therefore, SegNet incorporated the architecture and the weights of the convolution base from the VGG16 [30]. The VGG16 is already been trained on the ImageNet database comprised of 1000 classes for over 14 million images [31]. As mentioned in the previous sections, wearable sensors capture images from various environments with a great diversity of content (see Fig. 5). Hence, using pre-trained weights is highly advantageous for the feature extraction.

Therefore, SegNet naturally inherits 13 pre-trained convolution layers, branched among five max-pooling operators. Each convolution and pooling stage comprises a combination of convolution filters,a batch normalization layer, and a ReLU activation layer to extract feature maps at each encoder stage. The $2 \times 2$ max-pooling operators are an essential aspect of the network as the max-pooling indices capture the locations of the maximum feature value in each pooling window and forward the information to the decoder layer to improve boundary delineation of the semantic segmentation.

The decoder networks upsample the input feature map at each decoder stage using the pre-known max-pooling indices at each corresponding encoder stage. Thus, resulting in 13 trainable convolution layers. The convolution layers are brunched among a combination of Max Unpooling layers, batch normalization layers, and ReLU activation layers. Here the final encoded feature map of the input image is passed to the decoder Max Unpooling layer along with the Max Pooling indices to produce a dense feature map. The process of decoding is followed for the 5 Max Unpooling layers.

Finally, a multi-channel feature map, similar to the initial feature map of the encoder stage is generated. A trainable, SoftMax layer followed by a pixel classification layer is used to classify the pixels of the image to corresponding classes: person/bystander privacy, screen/content privacy, and background.

### B. SENSOR SYSTEM

The sensor system used for the method development is Automatic Ingestion Monitor, version 2 (AIM-2), which is a second-generation egocentric wearable sensor used for monitoringof diet and eating behavior (Fig. 4). The AIM-2 may capture periodic images only during food intake or during the whole day an image per 15 seconds or about 5760 images per day. AIM comprises of five main components: inputs; 5-megapixel CMOS image sensor and 3D accelerometer, STM32 processing unit and an FPGA based frame buffer, and micro SD-based storage unit. The camera sensor is a 5-megapixel CMOS image sensor with a wide-angle lens to capture a broader field of view.

**FIGURE 4.** AIM-2, an egocentric wearable camera with food intake detection sensors.

**FIGURE 5.** Illustration of the diversity of the dataset from a free-living experiment using AIM-2. (Faces and screens are blacked out).

### C. DATA

We used data from fifteen volunteers (9 males and 6 females) aged 18 to 33 years old. The study was approved by the Institutional review board of the University of Alabama. The participants signed an informed consent form prior to the experiments and were compensated for their participation. The experiment was conducted in two parts: a controlled laboratory experiment (1 day) and a free-living experiment (1 day).

To represent a variety of real-world eating scenarios, we only used the data from the free-living experiment. The participants were asked to wear the AIM-2 egocentric wearable sensor for the entire day. The actual wear time was between 12.5 to 18.5 hours. The participants were asked to follow their daily routine activities and have at least one meal. Furthermore, the free-living experiment did not limit the participants from any social/personal interaction and activities. At the end of the experiment, participants reviewed and removed any images that did not want to be included in the dataset. Even though the images been removed manually, the free-living experiment generated a dynamic dataset containing social/personal interaction and activities at public/personal spaces including bystanders and the usage of screens (see Fig. 5).

We randomly selected 400 images from each participant to generate our dataset comprised of 6000 images.

Images that were distorted by motion blur, overexposure, and idle view (i.e., the view when the participant is not wearing glasses/ after removing the glasses and placing somewhere else) were excluded. The randomly selected images were annotated to 3 classes: person, screen, and background. The persons and screens were annotated. The remaining pixels (i.e., non-person/screen) of the image were grouped as background pixels.

The annotation was performed by pixel-wise labeling using the MATLAB annotation tool known as the Image Labeler. Annotation for the person was defined as an entire human or any human body part occluded or non-occluded within the frame of the image. The definition for screens expanded over any digital screens: mobile phones, desktop monitors, televisions. We annotated 3781 images with screen labels containing entire, or partially occluded mobile phones/ desktop monitor/ television. We annotated 4378 images with labels for single person/ multiple persons/ partially occluded person.

We supplemented our dataset with publicly available datasets. We used the pixel annotated data from the CamVid database [29], SUN RGB-D database [32], and ADE20K database [33]. Here we relabeled the annotations of these public databases to use classes: person, screen and relabeled all the remaining pixels to the background.

AIM-2 is built with a wide-angle lens to capture a broader field of view and captures images in portrait orientation. The wide-angle lens projects a barrel distortion to the captured image. The barrel distortions and portrait orientation would contribute to reducing the efficiency of SegNet as its unlikely that the encoder base was previously trained with distorted images. Therefore, we performed a $-90^0$ image rotation and a barrel correction. We calculated the focal length, camera optical center, and the radial distortion coefficients of the lens to rectify the barrel distortion [34]. Furthermore, we carried out an image enhancement step by performing a histogram normalization step to improve the intensity distribution of the image.

#### 1) TRAINING AND VALIDATION

A transfer learning strategy was adopted to train the SegNet semantic segmentation model. We froze the encoder base and trained the 13 convolution layers in the decoder base along with the SoftMax layer and pixel classification layer (see Fig. 5). Stochastic gradient descent with momentum as the optimization algorithm to train the decoder layers. A momentum of 0.4 with a fixed learning rate of 0.01 was defined for the optimization. A class weighted cross-entropy loss approach was used to train the pixel classification layer to overcome the unbalanced data distribution in the training dataset.

We used 60% of our self-collected AIM-2 images for training along with the complete data from CamVid database, SUN RGB-D database, and ADE20K database, amounting to a total of 36846 training images. We further enhanced the number of training images to 331,614 in our training dataset by augmenting the data using 7 augmentation techniques: image rotation by $\pm$ 25-degree, image shift in horizontal axis, and vertical axis by $\pm$30 pixels, horizontal image flip, image scaling by a factor of 1.25, and 1.5, and brightness shift by a factor of 0.5 to 1.5. The training was carried out on a GeForce RTX 2070 GPU using a mini-batch size of 8 images for 80 epochs. The data was shuffled at each epoch.

During the validation phase, we adopted a holdout validation approach. We held 40% of the self-collected AIM-2 images from the self-collected dataset and was used for the validation.

The results of the validated images were evaluated using measures of accuracy, MeanBFScore, and IoU for each class [29]. The accuracy (see Eq. 1) shows the percentage of pixels that are correctly identified to the respective class. MeanBFScore (see Eq. 4), also known as the boundary F1 contour matching score shows how well are the boundaries of each class is aligned with the ground truth boundary. The intersection of the union (IoU) (see Eq. 5) shows a measure in the convergence between the pixel area of each class to the ground truth.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

$$\text{MeanBFScore} = \text{mean}\left(\frac{2 \times \text{Precision} \times Recall}{\text{Precision} + Recall}\right) \tag{4}$$

$$IoU = \frac{AreaofOverlap}{AreaofUnion} \tag{5}$$

where TP is true positive, TN is true negative, FP is false positive, and FN is a false negative. We evaluate the performance of the privacy concerning content (i.e., bystander and context privacy) removal by using precision (see Eq. 2) and recall (see Eq. 3) metrics. Here we considered bystander and context privacy content as objects by inserting a bounding box around the object silhouette in each image channel corresponding to the classes: person/bystander privacy and screen/context privacy. Hence the classification of the bounding box was used to estimate the precision and recall.

### D. SELECTIVE REMOVAL OF PRIVACY CONTENT

The images were processed with the trained SegNet semantic segmentation model (see Fig. 6). SegNet generates a semantically segmented tri-channel pixel classification map highlighting the pixels related to the person, screen, and background. Here the pixels in the first channel are classified as a person, pixels in the second channel are classified as screens, and pixels in the third channel are classified as background. We perform bystander and context privacy removal by nullifying the pixels in the first and second channels and preserving the pixels in the third channel. The updated pixel classification map is superimposed to the original image to generate the image with privacy content removed (see Fig. 7(d)).

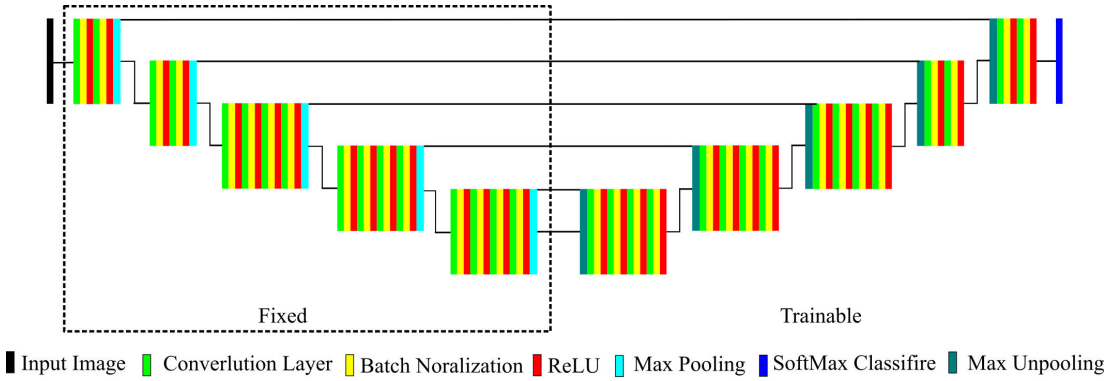



**FIGURE 6.** Illustration of the SegNet, encoder-decoder based deep convolution neural network architecture.

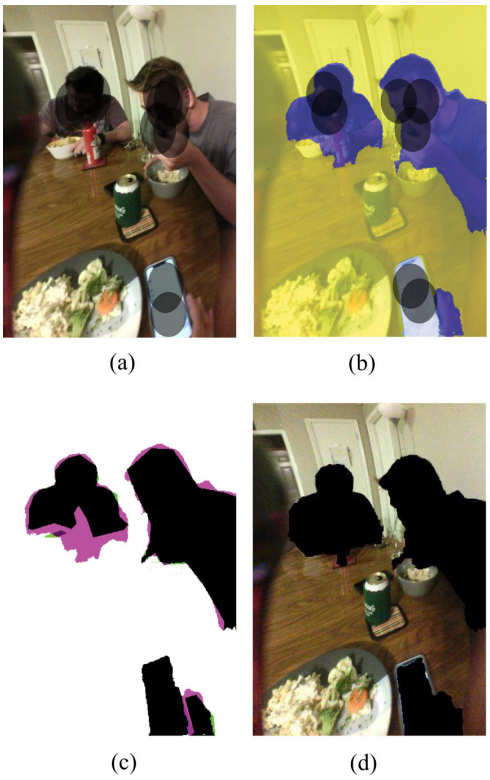



**FIGURE 7.** Accuracy of bystander and content privacy removal. (a) the input image, (b) output of the pixel classification layer: yellow is the background, purple is the person, and blue is the screen. (c) amount of overlap between classes, illustrating the rate of IoU. (d) privacy content removed output.

## IV. RESULTS AND DISCUSSION

The results for the bystander privacy removal are tabulated in Table 1, showing high precision and recall of 0.87 and 0.94 as an average of all the test subjects. The high recall rate suggests that our method is sensitive to detecting bystanders in the diverse dataset containing bystanders in social/individual interaction, and activities at public/personal spaces. Furthermore, our method was also able to remove body parts that were occluded.

This was important since the bystanders and the wearer's privacy could be violated either by the full figure or part of the body. The accuracy of 0.87 was reported for pixel-wise classification and removal of the bystander. However, our method reported a lower IoU and MeanBF; this is mainly due to the convergence of the classification region and boundary with the ground truth region and boundary (Fig. 7(c)).

This is mostly due to the boundary delineation of the semantic segmentation approach. SegNet performs a process of capturing the locations of the maximum feature value in each pooling window and passing the information to the decoder layer to improve boundary delineation. However, we noticed that the SegNet did not segment within the boundary of the bystander and often smear at edges.

The results related to context privacy removal are tabulated in Table. 1. The context privacy removal results showed high precision and recall of 0.97 and 0.98 as an average of all the test subjects. The high recall rate suggests that our method is sensitive to detecting digital screen from mobile phone screens, laptop screens, and, television screens at privet and public spaces. Partially occluded screens were also detected and were removed to preserve context privacy.

Therefore, the overall accuracy of pixel classification and removal for context privacy was 0.90. The method also reported a reasonable convergence of the classified region with ground truth by reporting an IoU of 0.77. The improvement in IoU for context privacy would have been due to the ridged structure of the objects (see Fig. 8(c)).

However, context privacy removal reported a lower meanBF of 0.55. The boundary misalignment between the classified region and the actual region is the leading cause of the lower meanBF. Furthermore, it was noticed that ReLU layers activated in much coherence to the brightness intensity in comparison to other features (see Fig. 8(b)).

Therefore, it could have contributed to smearing the boundary of the screen (see Fig. 8(d)). We also noticed an odd incident when the mobile screen was classified as a bystander. Here the picture of a person was displayed on the screen, and our method removed region as a bystander instead of content (see Fig. 9(b)). This is mainly due to the training examples in which the wearer's mirror reflection was labeled as the bystander (see Fig. 9(a)).

As mentioned earlier, AIM-2 is intended to monitor individual eating habits. Therefore, selectively removing privacy content was essential to retain information related to food intake. Dietitians and nutritionists often use the images of

**TABLE 1. Validation results of bystander privacy and context privacy removal.**

| Subject | Bystander Privacy | | | | | Context Privacy | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Accuracy | IoU | MeanBF | Precision | Recall | Accuracy | IoU | MeanBF | Precision | Recall |
| Subject 10 | 0.82 | 0.66 | 0.52 | 0.81 | 0.92 | 0.89 | 0.75 | 0.53 | 0.96 | 0.98 |
| Subject 11 | 0.92 | 0.64 | 0.55 | 0.92 | 0.97 | 0.91 | 0.78 | 0.51 | 0.98 | 0.97 |
| Subject 12 | 0.86 | 0.67 | 0.67 | 0.81 | 0.92 | 0.87 | 0.76 | 0.57 | 0.95 | 0.94 |
| Subject 13 | 0.93 | 0.70 | 0.62 | 0.98 | 0.94 | 0.89 | 0.78 | 0.57 | 0.98 | 1.00 |
| Subject 14 | 0.90 | 0.69 | 0.52 | 0.92 | 0.95 | 0.91 | 0.79 | 0.55 | 0.96 | 0.99 |
| Subject 15 | 0.79 | 0.62 | 0.67 | 0.78 | 0.92 | 0.92 | 0.78 | 0.60 | 0.96 | 0.98 |
| **Mean** | **0.87** | **0.66** | **0.59** | **0.87** | **0.94** | **0.90** | **0.77** | **0.55** | **0.97** | **0.98** |

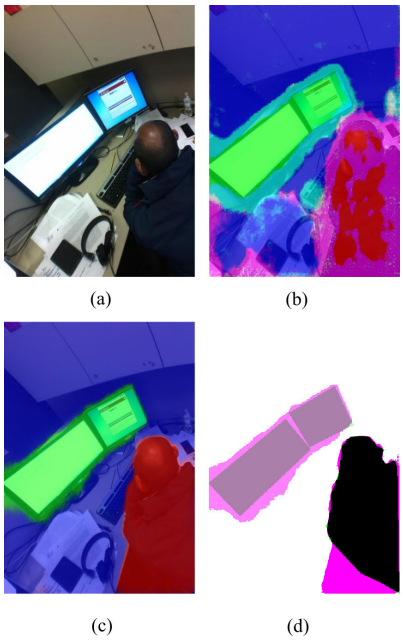



**FIGURE 8. Accuracy of context privacy removal, (a) input image, (b) output of the final ReLU activation layer, decoder1_relu_1, (c) output of the softmax classification layer and (d) amount of overlap between classes.**

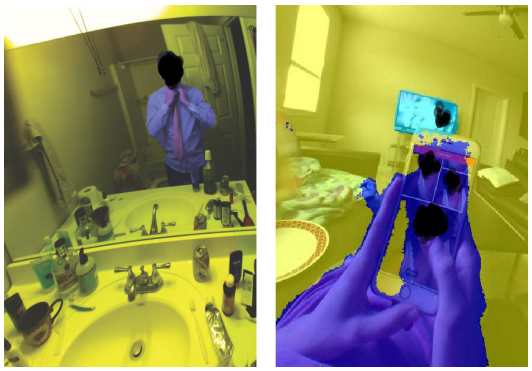


**FIGURE 9. Illustration of bystander removal instead of context removal. (a) the mirror reflection of a person as a training example annotated purple as a person and yellow as background. (b) pixel classification result of the test image, purple as a person, blue as a screen and yellow as background.**

the food intake to determine food content and the eating environment. Therefore, we estimated the background pixel classification accuracy.

Our proposed method removed privacy content while preserving 0.99 of the background accurately. Considering the application of ingestion monitoring, the critical process of privacy content removal is performed as a preprocessing operation. Therefore, we also estimated the computational efficiency of the proposed method. We tested the method on the test dataset containing data of 6 participants. The proposed method computed efficiently at a computational speed of 8.93 images per second.''

## V. CONCLUSION

This paper addressed an important problem of privacy concern for the egocentric camera sensor. Image information related to bystander privacy and context privacy was identified as the most significant issue for imaging-based wearable sensors. A fully automatic proposed method was developed to remove privacy content selectively from an egocentric wearable sensor. Semantic segmentation based deep convolution neural network was adopted to remove privacy content accurately while preserving background information. We used about to 331,614 images comprised of augmented data from 9 participates of the self-collected data and publicly available data for training. The proposed method was validated on data of 6 participants. The validation results reported that the proposed method removed bystander privacy with precision and recall of 0.87 and 0.94 while removing context privacy with precision and recall of 0.97 and 0.98.

## VI. LIMITATIONS

Our proposed method was able to protect sensitive information of individuals participating in the study by selectively removing sensitive content while preserving information related to food intake. This would potentially enable nutritionists to analyze diet and ingestive behavior. However, the sensitive information in the images from these ingestion monitoring studies are not entirely protected and could be compromised by adversarial attacks [35]–[37], where the individuals and their behavioral patterns may potentially be identified based on the object silhouette in the repeated food images.

While this is a limitation, the main goal of the proposed method is to improve the acceptance and compliance with device wear in nutritional studies, where the participants

may be concerned that the device captures private information during meals. The proposed privacy protection method should alleviate these concerns, which we plan to test in future studies. A limitation of the proposed method is that it provides privacy protection for off-line, post-data-collection review of the images of nutritional studies. Ongoing development of embedded hardware accelerators for deep neural networks, such as Google's Edge TPU [38], may enable eventual deployment of these algorithms directly on the wearable device for on-line execution.

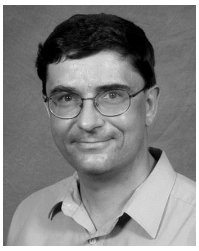
Although the proposed method operated satisfactorily on unseen data from ongoing studies, we noticed instances of inaccurate segmentation that erased the food images due to the person sitting next to the food. This is mainly caused by the generalization problem of the deep learning neural network. However, this is not a major concern for the related nutritional analysis, as many images are captured during any given meal and instances of improper segmentation are sufficiently rare. We propose to overcome this limitation by fine-tuning the model with the negative results from future studies that will provide additional data for retraining the models.

## REFERENCES

[1] J. Hayward, "Wearable technology 2018-2028: Markets, players, forecasts," IDTechEx, Cambridge, U.K., Tech. Rep. 49, 2018.
[2] A. J. Perez and S. Zeadally, "Privacy issues and solutions for consumer wearables," *IT Prof.*, vol. 20, no. 4, pp. 46–56, Jul. 2018.
[3] V. G. Motti and K. Caine, "Users' privacy concerns about wearables: Impact of form factor, sensors and type of data collected," in *Proc. Int. Conf. Financial Cryptogr. Data Secur.*, vol. 8976, 2015, pp. 231–234.
[4] A. M. Al-Busaidi and L. Khriji, "Wearable wireless medical sensors toward standards, safety and intelligence: A review," *Int. J. Biomed. Eng. Technol.*, vol. 14, no. 2, pp. 119–147, 2014.
[5] W. Tang, J. Ren, and Y. Zhang, "Enabling trusted and privacy-preserving healthcare services in social media health networks," *IEEE Trans. Multimedia*, vol. 21, no. 3, pp. 579–590, Mar. 2019.
[6] (2018). *Googleglass IDTechx*. [Online]. Available: https://en.wikipedia.org/wiki/Google_Glass
[7] L. Dubourg, A. R. Silva, C. Fitamen, C. J. A. Moulin, and C. Souchay, "SenseCam: A new tool for memory rehabilitation?" *Revue Neurologique*, vol. 172, no. 12, pp. 735–747, Dec. 2016.
[8] (2018). *Epson Moverio*. [Online]. Available: https://epson.com/moverio-augmented-reality
[9] J. M. Fontana, M. Farooq, and E. Sazonov, "Automatic ingestion monitor: A novel wearable device for monitoring of ingestive behavior," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 6, pp. 1772–1779, Jun. 2014.
[10] Vuzix. (2018). *Vuzix M100 Smart Glasses*. Accessed: Nov. 14, 2018. [Online]. Available: https://www.vuzix.com/Products/m100-smart-glasses
[11] (2018). *Snapspectacles*. [Online]. Available: https://www.spectacles.com/
[12] M. Farooq and E. Sazonov, "Accelerometer-based detection of food intake in free-living individuals," *IEEE Sensors J.*, vol. 18, no. 9, pp. 3752–3758, May 2018.
[13] F. Bellisle, A. M. Dalix, and G. Slama, "Non food-related environmental stimuli induce increased meal intake in healthy women: Comparison of television viewing versus listening to a recorded story in laboratory settings," *Appetite*, vol. 42, no. 2, pp. 175–180, 2004.
[14] *Health Information Privacy*. Accessed: Aug. 6, 2020. [Online]. Available: https://www.hhs.gov/hipaa/index.html
[15] M. S. Ryoo, B. Rothrock, C. Fleming, and H. J. acnd Yang, "Privacy-preserving human activity recognition from extreme low resolution," in *Proc. 31st AAAI Conf. Artif. Intell. (AAAI)*, 2017, pp. 4255–4262.
[16] J. Padilla-López, A. Chaaraoui, F. Gu, and F. Flórez-Revuelta, "Visual privacy by context: Proposal and evaluation of a level-based visualisation scheme," *Sensors*, vol. 15, no. 6, pp. 12959–12982, Jun. 2015.
[17] S. Ribaric, A. Ariyaeeinia, and N. Pavesic, "De-identification for privacy protection in multimedia content: A survey," *Signal Process., Image Commun.*, vol. 47, pp. 131–151, Sep. 2016.
[18] J. R. Padilla-López, A. A. Chaaraoui, and F. Flórez-Revuelta, "Visual privacy protection methods: A survey," *Expert Syst. Appl.*, vol. 42, no. 9, pp. 4177–4195, Jun. 2015.
[19] P. Kelly, S. J. Marshall, H. Badland, J. Kerr, M. Oliver, A. R. Doherty, and C. Foster, "An ethical framework for automated, wearable cameras in health behavior research," *Amer. J. Preventive Med.*, vol. 44, no. 3, pp. 314–319, Mar. 2013.
[20] (2018). *DeepAI Nudity Detection*. [Online]. Available: https://deepai.org/machine-learning-model/nsfw-detector
[21] Sightengine. (2018). *The Leading Image and Video Moderation API*. [Online]. Available: https://sightengine.com/
[22] E. Thomaz, A. Parnami, J. Bidwell, I. Essa, and G. D. Abowd, "Technological approaches for addressing privacy concerns when recognizing eating behaviors with wearable cameras," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput. (UbiComp)*, 2013, pp. 739–748.
[23] R. Hoyle, R. Templeman, S. Armes, D. Anthony, D. Crandall, and A. Kapadia, "Privacy behaviors of lifeloggers using wearable cameras," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput. (UbiComp Adjunct)*, 2014, pp. 571–582.
[24] E. Zarepour, M. Hosseini, S. S. Kanhere, and A. Sowmya, "A context-based privacy preserving framework for wearable visual lifeloggers," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops (PerCom Workshops)*, Mar. 2016, pp. 1–4.
[25] M. Wu, P. H. Pathak, and P. Mohapatra, "Enabling privacy-preserving first-person cameras using low-power sensors," in *Proc. 12th Annu. IEEE Int. Conf. Sens., Commun., Netw. (SECON)*, 2015, pp. 444–452.
[26] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*. Munich, Germany: Springer, 2015.
[27] M. Lin, Q. Chen, and S. Yan, "Network in network," 2013, *arXiv:1312.4400*. [Online]. Available: https://arxiv.org/abs/1312.4400
[28] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional Networks," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 818–833.
[29] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
[30] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556
[31] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[32] C. Zheng, J. Wang, W. Chen, and X. Wu, "Multi-class indoor semantic segmentation with deep structured model," *Vis. Comput.*, vol. 34, no. 5, pp. 735–747, 2018.
[33] B. Zhou, H. Zhao, X. Puig, S. Fidler, A. Barriuso, and A. Torralba, "Scene parsing through ADE20K dataset," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 633–641.
[34] V. Raju and E. Sazonov, "Processing of egocentric camera images from a wearable food intake sensor," in *Proc. SoutheastCon*, Huntsville, AL, USA, 2019, pp. 1–6, doi: 10.1109/SoutheastCon42311.2019.9020284.
[35] M. Saini, P. K. Atrey, S. Mehrotra, and M. Kankanhalli, "W3-privacy: Understanding what, when, and where inference channels in multi-camera surveillance video," *Multimedia Tools Appl.*, vol. 68, no. 1, pp. 135–158, Jan. 2014.
[36] A. Senior, *Protecting Privacy in Video Surveillance*. London, U.K.: Springer, 2009.
[37] Y. Chang, R. Yan, D. Chen, and J. Yang, "People identification with limited labels in privacy-protected video," in *Proc. IEEE Int. Conf. Multimedia Expo*, Toronto, ON, Canada, 2006, pp. 1005–1008, doi: 10.1109/ICME.2006.262703.
[38] *Edge TPU*. Accessed: Jun. 1, 2020. [Online]. Available: https://cloud.google.com/edge-tpu
[39] (2018). *Bragi Dash*. [Online]. Available: https://bragi.com/products/thedashpro
[40] (2018). *Samsung Gear IconX*. [Online]. Available: https://www.samsung.com/us/support/owners/product/gear-iconx
[41] J. Hailstone and A. E. Kilding, "Reliability and validity of the Zephyr BioHarness to measure respiratory responses to exercise," *Meas. Phys. Edu. Exercise Sci.*, vol. 15, no. 4, pp. 293–300, Oct. 2011.

[42] (2018). *Bellabeat LEAF*. [Online]. Available: https://www.bellabeat.com/products/leaf-urban
[43] (2018). *Amber Alert GPS*. [Online]. Available: https://amberalertgps.com/
[44] (2018). *BSXinsight*. [Online]. Available: https://www.bsxinsight.com/
[45] (2018). *Addidas Micoach*. [Online]. Available: https://www.adidas.com/de/micoach/
[46] (2018). *Fitbit*. [Online]. Available: https://www.fitbit.com/ionic
[47] (2018). *Myo Gesture Control Armband*. [Online]. Available: https://www.robotshop.com/en/myo-gesture-control-armband-black.html

**MOHAMED ABUL HASSAN** received the B.E. degree (Hons.) in electrical and electronic engineering from the University of East London, in 2012, and the M.Sc. and Ph.D. degrees in electrical and electronic engineering from University Technology Petronas, in 2015 and 2018, respectively. He worked as a Research Scientist with the Institute of Health Monitoring and Analytics, from 2017 to 2018, and a Postdoctoral Associate with CLAWS groups at the University of Alabama. His research interests include applying computer vision and linear and non-linear machine learning-based methods to address real-world problems in video analytics, wearable sensors, and interdisciplinary research with a particular focus on remote health monitoring applications. He serves as a reviewer for several journals, including IEEE, Elsevier, and other publications.



**EDWARD SAZONOV** (Senior Member, IEEE) received the Diploma degree in systems engineer from the Khabarovsk State University of Technology, Russia, in 1993, and the Ph.D. degree in computer engineering from West Virginia University, Morgantown, WV, in 2002. He is currently a Professor with the Department of Electrical and Computer Engineering, University of Alabama, Tuscaloosa, AL, and the Head of the Computer Laboratory of Ambient and Wearable Systems. His research interests span wearable devices, sensor-based behavioral informatics, and methods of biomedical signal processing and pattern recognition. Devices developed in his laboratory include a wearable sensor for objective detection and characterization of food intake (AIM – Automatic Ingestion Monitor); a highly accurate physical activity and gait monitor integrated into a shoe insole (SmartStep); a wearable sensor system for monitoring of cigarette smoking (PACT); and others. His research has been supported by the National Institutes of Health, National Science Foundation, National Academies of Science, as well as by state agencies, private industry and foundations. He serves as an Associate Editor for several journals, including IEEE, Frontiers and other publications.