POLITECNICO DI TORINO
Repository ISTITUZIONALE

Privacy-preserving Social Distance Monitoring on Microcontrollers with Low-Resolution Infrared Sensors and CNNs

(Article begins on next page)

29 November 2022

Privacy-preserving Social Distance Monitoring on Microcontrollers with Low-Resolution Infrared Sensors and CNNs

Chen Xie*, Francesco Daghero*, Yukai Chen*, Marco Castellano†, Luca Gandolfi†,
Andrea Calimera*, Enrico Macii*, Massimo Poncino*, Daniele Jahier Pagliari*
*Politecnico di Torino, Turin, Italy, E-mail: name.surname@polito.it
†STMicroelectronics, Cornaredo, Italy, E-mail: name.surname@st.com

Abstract—**Low-resolution infrared (IR) array sensors offer a low-cost, low-power, and privacy-preserving alternative to optical cameras and smartphones/wearables for social distance monitoring in indoor spaces, permitting the recognition of basic shapes, without revealing the personal details of individuals. In this work, we demonstrate that an accurate detection of social distance violations can be achieved processing the raw output of a 8x8 IR array sensor with a small-sized Convolutional Neural Network (CNN). Furthermore, the CNN can be executed directly on a Microcontroller (MCU)-based sensor node.**

With results on a newly collected open dataset, we show that our best CNN achieves 86.3% balanced accuracy, significantly outperforming the 61% achieved by a state-of-the-art deterministic algorithm. Changing the architectural parameters of the CNN, we obtain a rich Pareto set of models, spanning 70.5-86.3% accuracy and 0.18-75k parameters. Deployed on a STM32L476RG MCU, these models have a latency of 0.73-5.33ms, with an energy consumption per inference of 9.38-68.57μJ.

Index Terms—**Social Distance, COVID-19, Edge Computing, Infrared Sensors, Convolutional Neural Networks**

I. INTRODUCTION AND RELATED WORKS

In the current COVID-19 pandemic, social distancing [1]–[3], together with extensive testing [4], [5] has been proven the most effective way to prevent the spread of infections, especially when effective treatments and vaccines were not yet available. This virus spread prevention measure, particularly critical for indoor environments such as offices, shops, factories etc [1], [2], will therefore continue to have an important role in this, and in possible future epidemics.

Several technical solutions have been proposed to monitor compliance with social distancing rules. One category is based on computing the distance among people using the Bluetooth or Wi-Fi transcievers available in smartphones [6], [7] or wearables [8], [9]. While effective, this approach requires the voluntary participation of users, hence not providing 100% safety guarantees. Another set of solutions monitors indoor spaces with cameras [10]–[14]. The videos are then processed with Machine Learning (ML) algorithms to locate people and compute their relative distance. While not requiring actions

from users, this approach poses critical privacy issues. In fact, workers may have concerns with the installation of a system that can not only monitor the distance among them, but also identify, record and track individuals, often in violation with privacy protection laws [15].

In this scenario, low-resolution infrared (IR) array sensors constitute an interesting alternative. Being capable of acquiring small-sized thermal images (typically 8x8 or 16x16 pixels) at a frame rate of $\approx$10 Frames Per Second (FPS) [16], these sensors can recognize basic shapes, without revealing privacy-sensitive details of an individual (facial features, clothes, hair style, etc). Furthermore, their limited power consumption, low cost, and low-resolution outputs has another positive implication for privacy. That is, it enables the implementation of social distance monitoring directly *on the sensing devices*, which are typically battery operated and equipped with resource-constrained Microcontrollers (MCUs) [17]–[21]. Executing the monitoring on end-nodes, in turn, enables privacy-preserving solutions in which social distance violations are signaled in real time, warning the responsible staff without ever transmitting and/or storing the collected data in the cloud.

The key question is then whether the data collected by low-resolution IR sensors can be effectively used for this task. In fact, while other authors have employed these sensors for various applications [20]–[34], to the best of our knowledge no previous work has considered them for social distancing.

Most works combining low-resolution IR sensors with ML focus on human activity recognition [23]–[32]. Some of them apply classical algorithms [23]–[26] while others propose Deep Learning (DL) approaches, based either on Convolutional Neural Networks (CNNs), Long-Short Term Memory (LSTMs), Gated Recurrent Units (GRUs), or a combination of the above (e.g., CNN-LSTM) [27]–[32]. Besides the selected algorithm, activity recognition solutions also differ in terms of: i) the number and position of the employed IR sensors, ii) the pre-processing applied to thermal images before feeding them to ML models, and iii) the type of recognized activities, which range from daily tasks such as walking, sitting, standing, etc [23], [25], [27], [29]–[31], to elderly people falls [24], [26], epilepsy-induced convulsions [32], and even yoga postures [28].

Another set of works based on IR thermal arrays focuses on

human presence detection [33], [34] and people counting [20]–[22]. Also in this case, both classical [22], [33] and DL solutions exist [20], [21], [34], and the latter are again based on CNNs, LSTMs and GRUs. In proper conditions, people counting can be transformed into social distance monitoring: if the IR sensor is positioned so that having more than a given number of people in the field of view corresponds to a violation of social distancing rules, then the monitoring algorithm can simply compute the people count and compare it with a threshold to trigger an alarm. However, existing IR array-based people counting implementations use relatively high resolution sensors (e.g., 24x32 [21] and 80x60 [20]), which do not guarantee the same level of privacy, and also induce more power consumption than low-resolution ones, both in the sensors themselves and in the computation part of the system. To our knowledge, the only people counting solution for low-resolution IR sensors (8x8 pixels) is the one proposed in [22], which is not based on ML. However, as shown in Section III, this approach obtains poor results when adapted to social distance monitoring.

In this work, we introduce the *first dedicated implementation* of a privacy-preserving social distance monitoring system on extremely low-resolution IR arrays. We use the same 8x8 sensor of [22], but we follow an end-to-end deep learning approach based on a CNN classifier. We explore different CNN architectures, which are then quantized and deployed on a commercial MCU for inference (the STM32L476RG by STMicroelectronics). With experiments on a new dataset, we show that we can obtain a rich set of Pareto-optimal solutions in the accuracy versus computation cost space. Specifically, all our CNNs meet real-time performance constraints, while reaching 70.5-86.3% balanced accuracy on the test set, occupying 64-139 kB of Flash memory (including code size) and consuming just 9.38-68.57μJ per inference, depending on the architecture. The balanced accuracy of our best CNN is 25% higher than the one obtained by the baseline method of [22].

II. PROPOSED METHOD

A. Dataset Collection

While there exist a few public datasets containing IR array sensor images, none of them fits the needs of our target application. In fact, most of them are relative to high resolution sensors (from 160x120 pixels to 640x480), for applications in robotics, autonomous driving, etc [35]–[38]. To the best of our knowledge, the only public dataset of low-resolution IR images is [39]. However, the latter is built for more complex human activity recognition tasks. Accordingly, samples are obtained combining the data from 3 different IR sensors, wall-mounted in different positions. In contrast, our goal is to build a social distance monitoring system based on a *single* IR array.

Therefore, we collected a brand new labelled dataset, specifically tailored for person counting and social distance monitoring, which is now available open source[1]. We decided to use ceiling-mounted sensors, since in indoor environments, the height of the ceiling is less variable than the size of the

[1] https://www.kaggle.com/francescodaghero/linaige

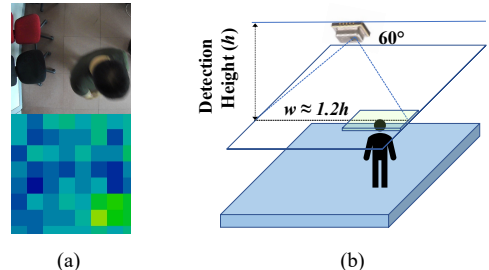| Session	N. Frames	Viol. [%]	N. People	Sensor Height [m]	Room
1	18124	34.3	0-5	2.4	A
2	1581	54.5	0-2	2.7	B
3	1519	6.4	0-3	2.4	A
4	196	5.6	0-2	2.4	A
5	2202	22.7	0-3	2.4	A
6	1850	32.3	0-3	2.7	C
Tot.	25472	32.5	-	-	-

Fig. 1. Example of optical and thermal frames (a) and Panasonic Grid-EYE detecton area calculation (b).

room. Consequently, the shape of a person seen from above in the thermal image does not change significantly for different environments. As in [39], IR frames are acquired using a Panasonic Grid-EYE (AMG8833), which outputs a 8x8 array, and has a view angle of 60° [16].

To gather the data, we setup a system based on a Raspberry Pi 3 equipped with the Grid-Eye and a standard USB camera, pointing in the same direction. We collected synchronized frames from the camera and sensor at 10 FPS, in different indoor environments including offices, research laboratories and corridors, in 6 different sessions, asking up to 5 volunteers to stand, walk, etc, in the area under the sensor. Table I reports the detailed information of each session, whereas Figure 1a shows an example of optical and IR frames.

We used a semi-automatic approach to speed-up the labeling of IR data. Namely, we applied a pre-trained Mask R-CNN [40] to the optical images, detecting and counting the number of people in each frame, and associating the same people count to the corresponding IR frame. Human labellers were then provided with the two images (optical and thermal) and asked to confirm or correct the people count suggested by Mask R-CNN. Besides coping with Mask R-CNN misclassifications, human checking was also needed because of the difficulty of exactly matching the alignment and viewing angles of the camera and IR array. Due to such imperfect matching, some IR frames showed the heat profile of person (close to a corner) which did not appear on the corresponding optical frame. Human labellers were given the possibility of annotating such "hard-to-label" frames, which were then excluded from the training and testing of the proposed model and of the state-of-the-art comparison.

B. Social Distance Monitoring Solution

Given the view angle of the IR sensor, the width w of the (square) field of view can be calculated with simple trigonometry. As explained in [16] and shown in Figure 1b, $w \approx 1.2h$ where h is the distance between the sensor and the detected object. For the range of sensor heights present in our

dataset (see Table I), and assuming that people heads (the main sources of heat detected by the sensor) are at least 1.5m above the floor level, w spans the range $[1.08 : 1.44]$ m. Consequently, the field of view diagonal, i.e., the maximum distance between two in-frame objects is in the range $[1.53 : 2.04]$ m.

Given that the typical recommendation for social distancing is to maintain at least 2m from the closest person [2], we can conclude that *any frame containing more than 1 person* corresponds to a violation. Accordingly, we model our task as a *binary classification*, where the goal is to predict whether a frame contains 2 or more people. This approach can be extended to larger spaces by combining multiple IR arrays placed in different positions on the roof [16]. Note that, while this formulation simplifies the problem with respect to predicting the exact people count, it is still significantly more complex than simple presence detection [33], [34], i.e. distinguishing between no person and 1 or more people. For instance, Figure 2 shows that two nearby people can produce a single hot area in the frame, which can be easily misclassified as a single person.
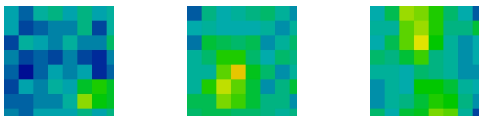

Fig. 2. Examples of IR array frames corresponding to 1 person (left), 2 people, close to each other (center), two people, far from each other (right).

Convolutional Neural Networks (CNNs) [41] are known to be amongst the most effective DL models to analyze visual imagery. Therefore, we design a set of simple CNN classifiers to implement the proposed IR array-based social distance monitoring. The common template of the models is inspired by classic CNNs such as LeNet-5 [41], and is shown in Figure 3. It includes two Convolutional (Conv) layers with Rectified Linear Unit (ReLU) activation, one Max Pooling layer, and two Fully Connected (FC) layers. The first FC layer uses a ReLU activation and has a hidden size of 64, while the second one has a sigmoid activation and a single output neuron, producing the probability of a distancing rule violation.

Starting from this template, we perform an extensive architectural exploration. Different networks are obtained eliminating some of the layers enclosed in dashed boxes in Figure 3. Specifically, we consider architectures with : i) one or two Conv layers, ii) one or two FC layers, iii) with or without Max Pooling. We also vary the number of *channels* in each of the two convolutional layers considering values $\{8, 16, 32, 64\}$.

We further build two variants of each CNN architecture, differing in the processed input. The first network is fed with a a *single* thermal frame X_t, and learns to predict whether the corresponding people count y_t is greater than 1 (social distance violation). The second variant is fed with a *sliding window* of W consecutive frames $(X_t, ..., X_{t+W-1})$, and is trained to detect social distance violations in the *last* frame of the window (i.e., $y_{t+W-1} > 1$). The corresponding input tensor shapes are (8,8,1) and (8,8,W) respectively, i.e., the W frames are passed to the CNN as different *input channels*. The two types of input are depicted in the left of Figure 3

as A and B respectively. For clarity, a single input sample of each type is enclosed by a purple box. The second CNN variant has the advantage of having access to *past* frames, which can be used, for instance, to detect that a single heat source corresponds to two nearby people (see Figure 2), based on their movement. On the other hand, this variant has a higher computational complexity, which is particularly critical for in-sensor inference. In our experiments, we use $W = 8$ for a fair comparison with the algorithm of [16], which uses a window of 8 frames for background subtraction.

As a further optimization step, to reduce the memory occupation and inference latency/energy of our CNNs, all parameters and intermediate activations are *quantized* to 8-bit integers [42].

C. Training Protocol

We trained our CNNs on Session 1 data (see Table I) and tested them on all other sessions. This *per-session* split ensures that test samples are taken either in a different room with respect to training data, or in the same room but at a different day/time. In contrast, a purely random split in which consecutive frames from the same session can end up in different data subsets would lead to an overly simplified and unrealistic version of the problem. Considering all combinations of CNN architectures and input shapes, we trained a total of 96 networks. Each model has been trained 5 times with different random seeds, for a maximum of 500 epochs, with early stopping after 10 non-improving epochs. We used a binary cross-entropy loss function and the Adam optimizer, with an initial Learning Rate (LR) of $1 \cdot 10^{-3}$. The LR is reduced by a factor 0.3 on plateau, with a patience of 5 epochs. Once the floating point model reaches convergence, we quantize it and then apply quantization-aware training [42] with the same protocol, except for the initial LR, which is set to $5e^{-4}$. To deal with the class imbalance of the target dataset (see Table I), we apply sample weights to the loss function, equal to the inverse of the class probabilities. After training all CNNs, we select those that are in the Pareto front as detailed in Section III.

III. RESULTS

We trained and quantized our CNNs with Keras/TensorFlow 2.0 [43], and we used the X-CUBE-AI v6.0 toolchain [44] to convert trained models into optimized C code for inference. Given the class imbalance of the target dataset, we mainly evaluate the performance of our models in terms of *balanced accuracy* (Bal. Acc.), i.e. the arithmetic mean of sensitivity and specificity. However, we also report the Area under the ROC (ROC-AUC), the F1-Score (F1), and the standard micro-average accuracy (Acc.) [45]. For all metrics, we report mean $\pm$ standard deviation on the 5 training runs.

The hardware-independent computational cost of the CNNs, used to extract Pareto curves, is measured in terms of model size (number of parameters) and number of Multiply-and-Accumulate (MAC) operations. Models deployed on the MCU are then further evaluated in terms of inference latency, energy consumption, and memory occupation.

We mainly compare against the algorithm of [22], which is, to our knowledge, the only person counting solution on
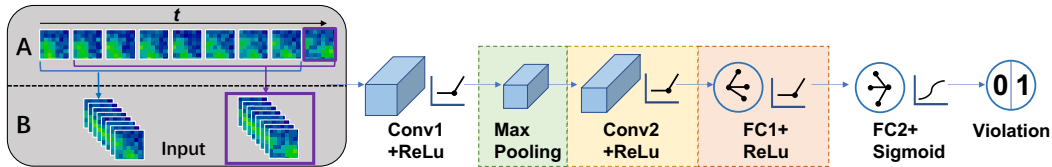
Fig. 3. CNN Architecture and input generation.

TABLE II
DETAILED EVALUATION AND DEPLOYMENT RESULTS OF THREE SELECTED ARCHITECTURES ON STM32L4 @ 80MHZ (12.8 MW AVERAGE POWER).
ABBREVIATIONS: 1/8F = 1-FRAME/8-FRAME INPUT, CN = CONVOLUTION WITH N CHANNELS, P = MAX POOLING, FC = FULLY-CONNECTED.

Model	Bal. Acc. [%]	ROC-AUC	Acc. [%]	F1	Size[B]	Mem.[B]	MACs	Energy[μJ]	Latency[ms]	Architecture
MinSize	70.5±3.6	0.83±0.04	75.6±0.9	0.56±0.05	0.18k	63.2k	2.6k	9.38	0.728	1F-C8-P-FC
MaxAcc-1%	85.3±2.0	0.93±0.04	86.4±1.9	0.77±0.03	4.9k	68.2k	7.3k	12.93	1.005	1F-C8-P-FC-FC
MaxAcc	86.3±2.2	0.99±0.01	87.6±1.8	0.78±0.03	74.7k	138.8k	157k	68.57	5.328	8F-C32-FC-FC
[22]	61	-	73	0.41	-	33.5k	-	53.09	4.125	-

low-resolution IR sensors. We execute the algorithm of [22] as is, and simply convert the person count into a social distance violation whenever 2 or more people are detected in a frame.

A. Architectural Exploration

Fig. 4 shows the Pareto curves generated by different CNN architectures in terms of balanced accuracy versus number of parameters and MAC operations respectively. The two curves refer to the single-frame and windowed input variants respectively (A and B in Figure 3), and each point represents a different architectural configuration (type and number of layers, number of channels, etc). Dots correspond to the mean accuracy over 5 trainings, while colored bands identify the ± standard deviation range. The horizontal dashed line shows the accuracy obtained by the baseline method of [22]. Note that the model size and MAC curves contain in general *different* CNNs.

The results clearly show that all proposed CNNs significantly outperform the baseline comparison in terms of accuracy. In particular, the single-frame networks achieve very high accuracy (> 80%) even with just around 1k parameters and less than 10k MACs. Windowed networks are significantly more costly, especially in terms of number of operations, yet they are able to obtain the overall best accuracy (86.3%), in a configuration with 75k parameters, and requiring 157k MACs.

In general, we found that both types of models achieve better performance using a small number of channel in convolutional layers (architectures with 64 channels are not in the Pareto front). This is probably due to the relatively small and simple dataset, for which large number of features lead to over-fitting.

B. Deployment

We selected 3 architectures from the parameters Pareto curve (indicated by red dots in Figure 4) and deployed them on the target MCU. The detailed results are shown in Table II. The selected CNNs are: i) the most accurate one, ii) the smallest model that achieves a < 1% balanced accuracy drop with respect to the best, and iii) the smallest overall. We also compiled and profiled the code of [22] on the same MCU for comparison. For our CNNs, the table reports both the quantized model size in Bytes (Size column), as well as the total occupied Flash memory, also including code size (Mem). As shown, our CNNs outperform [22] in all considered accuracy metrics. Note
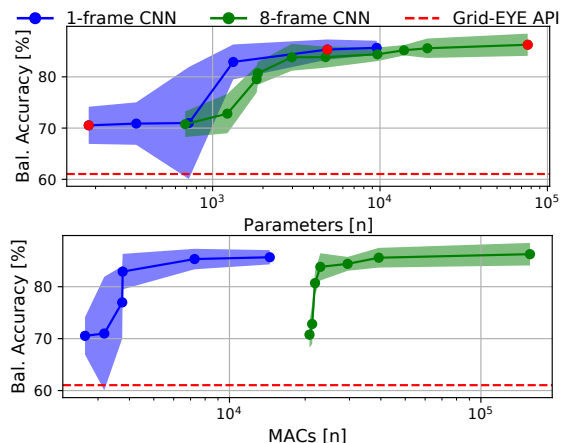

Fig. 4. Pareto Curves of balanced accuracy vs. parameters and n. of operations.

that the ROC-AUC is not reported for [22] since this method outputs a deterministic 0/1 value, and not a probability score. In terms of bal. acc., we outperform [22] by 9.5-25.2%, depending on the selected CNN. Moreover, MinSize and MaxAcc-1% are respectively 5.6x and 4.1x faster and more energy efficient than [22]. Although the total memory occupation of our models is larger, this is mostly due to the large code size of X-CUBE-AI (≈60kB). Moreover, all three CNNs easily fit in the 1MB Flash of the target MCU. Finally, note that, mainly because we rely on a very low-resolution 8x8 sensor, our MinSize/MaxAcc CNNs are 60.6x/8.3x faster than the networks proposed in [21], and strikingly 235000x/32000x more energy efficient than the method of [20], i.e., the only other two works proposing MCU deployments of IR array-based ML tasks. However, it must be underlined that, besides using different inputs, [20], [21] also solve a slightly different task (person counting), making a fair comparison with our work impossible.

IV. CONCLUSIONS

We have proposed a new method to implement privacy-preserving social distance monitoring in indoor spaces, using a low-resolution IR array sensor and an end-to-end deep learning approach. Our results show that, with compact and energy-efficient CNNs, an accurate social distance monitoring can be implemented directly on MCU-based sensor nodes. Future

works will include the application of sub-byte quantization to further reduce models sizes and complexity.

REFERENCES

[1] C. Sun and Z. Zhai, "The efficacy of social distance and ventilation effectiveness in preventing covid-19 transmission," *Sustainable Cities and Society*, vol. 62, p. 102390, 2020.

[2] Z. Vokó and J. G. Pitter, "The effect of social distance measures on covid-19 epidemics in europe: an interrupted time series analysis," *GeroScience*, vol. 42, no. 4, pp. 1075–1082, 2020.

[3] M. Greenstone and V. Nigam, "Does social distancing matter?" *University of Chicago, Becker Friedman Institute for Economics Working Paper*, no. 2020-26, 2020.

[4] World Health Organization, "Laboratory testing strategy recommendations for covid-19," 2020.

[5] P. Berchialla et al, "To swab or not to swab? the lesson learned in italy in the early stage of the covid-19 pandemic," *Applied Sciences*, vol. 11, no. 9, 2021.

[6] M. E. Rusli et al, "Mysd: A smart social distancing monitoring system," in *2020 8th International Conference on Information Technology and Multimedia (ICIMU)*, 2020, pp. 399–403.

[7] L. Li et al, "Smartdistance: A mobile-based positioning system for automatically monitoring social distance," in *IEEE INFOCOM*, 2021, pp. 1–10.

[8] A. O. Cunha et al, "Design and development of a wearable device for monitoring social distance using received signal strength indicator," in *WebMedia '20*, 2020, p. 57–60.

[9] Y. Kobayashi et al, "A system for monitoring social distancing using microcomputer modules on university campuses," in *ICCE-Asia*, 2020, pp. 1–4.

[10] I. Ahmed et al, "A deep learning-based social distance monitoring framework for covid-19," *Sustainable Cities and Society*, vol. 65, p. 102571, 2021.

[11] D. Yang et al "A vision-based social distancing and critical density detection system for covid-19," *Sensors*, vol. 21, no. 13, 2021.

[12] T. Alhmiedat and M. Aborokbah, "Social distance monitoring approach using wearable smart tags," *Electronics*, vol. 10, no. 19, 2021.

[13] S. Jethani et al, "Surveillance system for monitoring social distance," in *Advanced Computing*, D. Garg, K. Wong, J. Sarangapani, and S. K. Gupta, Eds. 2021, pp. 100–112.

[14] Y. Niu et al, "Monocular pedestrian 3d localization for social distance monitoring," *Sensors*, vol. 21, no. 17, 2021.

[15] P. Barsocchi et al, "Covid-19 & privacy: Enhancing of indoor localization architectures towards effective social distancing," *Array*, vol. 9, p. 100051, 2021.

[16] Panasonic, "Panasonic High Performance Grid-EYE Sensors Reference Specification," 2019.

[17] F. Daghero et al, "Ultra-compact binary neural networks for human activity recognition on risc-v processors," in *ACM CF'21*, 2021, p. 3–11.

[18] M. Risso et al, "Robust and energy-efficient ppg-based heart-rate monitoring," in *IEEE ISCAS*, 2021, pp. 1–5.

[19] M. Risso et al, "Pruning In Time (PIT): A Lightweight Network Architecture Optimizer for Temporal Convolutional Networks," in *58th ACM/IEEE DAC*, 2021, pp. 1015-–1020.

[20] A. Gomez et al, "Thermal image-based cnn's for ultra-low power people recognition," in *ACM CF'18*, 2018, p. 326–331.

[21] A. Metwaly et al, "Edge computing with embedded ai: Thermal image analysis for occupancy estimation in intelligent buildings," in *ACM INTESA*, 2019, p. 1–6.

[22] P. Industry, "Grid-eye application note on social distancing. people detection and tracking with ceiling mounted sensors," 2020.

[23] S. Mashiyama et al, "Activity recognition using low resolution infrared array sensor," in *IEEE ICC*, 2015, pp. 495–500.

[24] A. Hayashida et al, "The use of thermal ir array sensor for indoor fall detection," in *IEEE SMC*, 2017, pp. 594–599.

[25] L. Tao et al, "Home activity monitoring using low resolution infrared sensor," *arXiv preprint arXiv:1811.05416*, 2018.

[26] Z. Liu et al, "Fall detection and personnel tracking system using infrared array sensors," *IEEE Sensors Journal*, vol. 20, no. 16, pp. 9558–9566, 2020.

[27] T. Kawashima et al, "Action recognition from extremely low-resolution thermal image sequence," in *14th IEEE AVSS*, 2017, pp. 1–6.

[28] M. Gochoo et al, "Novel IoT-based privacy-preserving yoga posture recognition system using low-resolution infrared sensors and deep learning," *IEEE Internet of Things Journal*, vol. 6, no. 4, pp. 7192–7200, 2019.

[29] L. Tao et al, "3d convolutional neural network for home monitoring using low resolution thermal-sensor array," in *3rd IET TechAAL*, 2019, pp. 1–6.

[30] C.-S. Shih et al, "Multiple-image super-resolution for networked extremely low-resolution thermal sensor array," in *IEEE SenSys-ML*, 2020, pp. 1–6.

[31] I. Morawski and W.-N. Lie, "Two-stream deep learning architecture for action recognition by using extremely low-resolution infrared thermopile arrays," in *SPIE IWAIT 2020*, 2020, pp. 164–169.

[32] O. Hanosh et al, "Convulsive movement detection using low-resolution thermopile sensor array," in *IEEE/CVF CVPR Workshops*, 2020, pp. 300–301.

[33] A. Trofimova et al, "Indoor human detection based on thermal array sensor data and adaptive background estimation," *Journal of Computer and Communications*, vol. 05, pp. 16–28, 03 2017.

[34] C. Herrmann et al, "CNN-based thermal infrared person detection by domain adaptation," in *SPIE Autonomous Systems: Sensors, Vehicles, Security, and the Internet of Everything*, 2018, pp. 38 – 43.

[35] D. Olmeda et al, "Pedestrian detection in far infrared images," *Integrated Computer-Aided Engineering*, vol. 20, no. 4, pp. 347–360, 2013.

[36] FLIR, "Free flir thermal dataset for algorithm training," 2018.

[37] S. Hwang et al, "Multispectral pedestrian detection: Benchmark dataset and baselines," in *IEEE CVPR*, 2015.

[38] R. E. Rivadeneira et al, "Thermal image super-resolution challenge - PBVS 2020," in *IEEE/CVF CVPR Workshops*, 2020.

[39] Y. Karayaneva et al, "Infrared human activity recognition dataset - Coventry-2018," IEEE Dataport. DOI: 10.21227/baja-1j59, 2020.

[40] K. He et al, "Mask R-CNN," in *IEEE ICCV*, 2017, pp. 2980–2988.

[41] Y. Lecun et al, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[42] B. Jacob et al, "Quantization and training of neural networks for efficient integer-arithmetic-only inference," in *IEEE CVPR*, 2018.

[43] M. Abadi et al, "TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems", https://www.tensorflow.org/, 2015.

[44] STMicroelectronics, X-CUBE-AI, AI expansion pack for STM32CubeMX, https://www.st.com/en/embedded-software/x-cube-ai.html.

[45] J. D. Kelleher et al, "Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies". MIT Press, 2015.