ENIGMA: Egocentric Navigator for Industrial Guidance, Monitoring and Anticipation

Francesco Ragusa[1,2], Antonino Furnari[1,2], Antonino Lopes[3], Marco Moltisanti[3], Emanuele Ragusa[3], Marina Samarotto[3], Luciano Santo[3], Nicola Picone[4], Leo Scarso[4] and Giovanni Maria Farinella[1,2]

[1]*FPV@IPLAB, DMI - University of Catania, Italy*

[2]*Next Vision s.r.l. - Spinoff of the University of Catania, Italy*

[3]*Xenia Gestione Documentale s.r.l. - Xenia Progetti s.r.l., Acicastello, Catania, Italy*

[4]*Morpheos s.r.l. - Catania, Italy*

Keywords: Egocentric Vision, First Person Vision, Industrial Domain.

Abstract: We present ENIGMA (Egocentric Navigator for Industrial Guidance, Monitoring and Anticipation), an integrated system to support workers in an industrial laboratory. ENIGMA includes a wearable assistant which understands the worker's behavior through Computer Vision algorithms which 1) localize the operator, 2) recognize the objects present in the laboratory, 3) detect the human-object interactions which happen and 4) anticipate the next-active object with which the worker will interact. Furthermore, a back-end extracts high semantic information about the worker behavior to provide useful services and to improve his safety. Preliminary experiments were conducted showing good performance on the tasks of localization, object detection and recognition and egocentric human-object interaction detection considering the challenging industrial scenario.

1 INTRODUCTION

Understanding human behavior from an egocentric point of view allows to build an intelligent system able to support humans equipped with a camera (e.g., smartglasses, head-mounted, etc.) to achieve daily goals in different scenarios such as home environments (Damen et al., 2014), cultural sites (Farinella et al., 2019; Cucchiara and Bimbo, 2014), and industrial buildings (Colombo et al., 2019; Ragusa et al., 2021; Ragusa et al., 2022).

In particular, in the industrial scenario, localizing the users in an indoor workplace can be helpful in managing rescue situations such as fires or heartquakes guiding them to the closest emergency exit as well as detecting and recognizing objects in the surrounding environment allows to provide additional information on how to use them (i.e., automatic and continuous training). Moreover, recognizing human-object interactions can be useful to provide suggestions on how to execute a complex procedure of maintenance as well as to implement energy saving strategies. Furthermore, anticipating with which objects a worker will interact, allows to improve his safety in a factory, for example by notifying the user with an alert in case of a dangerous object.

Nowadays different systems have been developed to train workers for specific tasks using virtual (Osti et al., 2021) or augmented reality (Sorko and Brunnhofer, 2019) as well as to support them with remote assistance (Gurevich et al., 2012) guiding the local operator during the execution of procedural tasks (Rebol et al., 2021; Sun et al., 2021). Despite these systems provide help to workers in different manners, they suffer from several limitations due to the inability to understand the human behavior and the surrounding environment.

In this paper, we present ENIGMA (Egocentric Navigator for Industrial Guidance, Monitoring and Anticipation), an AI wearable assistant capable of supporting the workers of an industrial laboratory during the execution of complex tasks, providing suggestions on how to perform different maintenance and repairing procedures, improving their safety anticipating potential dangerous interactions and implementing energy saving strategies to reduce electricity consumption. To achieve the aforementioned goals, ENIGMA implements algorithms to localize workers in the industrial laboratory, to recognize the objects present in the surrounding environment and the interaction with them and to anticipate the next-active objects with which workers will interact from the egocentric point of view. Figure 1 shows the concept of the proposed AI assistant.

Ragusa, F., Furnari, A., Lopes, A., Moltisanti, M., Ragusa, E., Samarotto, M., Santo, L., Picone, N., Scarso, L. and Farinella, G.
ENIGMA: Egocentric Navigator for Industrial Guidance, Monitoring and Anticipation.
DOI: 10.5220/0011787900003417
In *Proceedings of the 18th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2023) - Volume 4: VISAPP*, pages 695-702
ISBN: 978-989-758-634-7; ISSN: 2184-4321

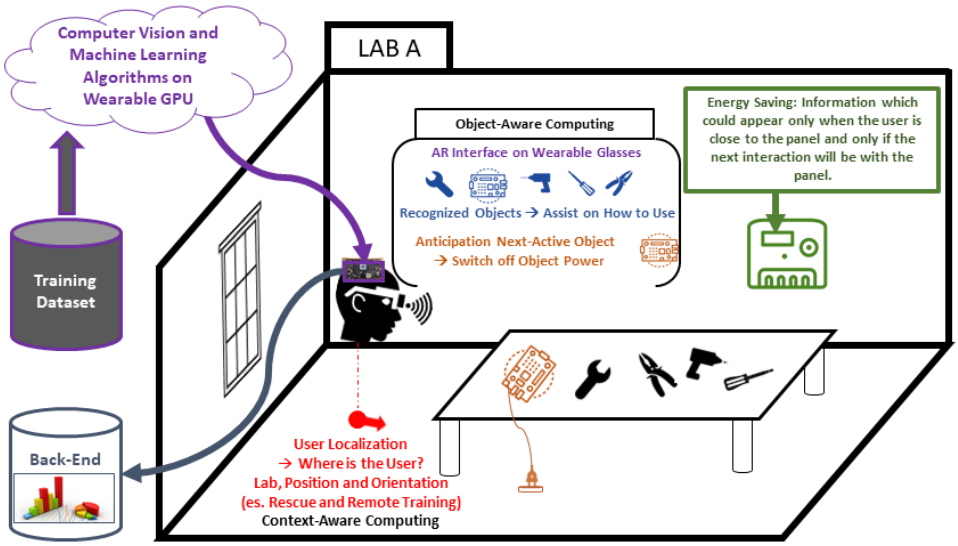

Figure 1: The concept of the proposed ENIGMA Assistant.

The proposed system has been tested in a laboratory which represents a realistic industrial scenario. In the considered laboratory there are 23 different objects both fixed such as electric panels, a power supply and a welding station and mobile such as screwdrivers and electric boards. In addition, there are different IoT devices installed in the sockets of the worktable and in the electric panel which allow to powering on and off the electricity of the tools connected to the sockets. In this laboratory, we have considered 8 contexts, 23 objects and 22 different human-object interactions. Preliminary experiments show that the proposed ENIGMA system achieves good performances on the tasks of Localization, Object Detection and Recognition and Egocentric Human-Object Interaction Detection while, future experiments will address the tasks of and Next-Active Objects Detection.

The reminder of the paper is organized as follows. Section 2 reports the related work. Section 3 presents the collected and labeled datasets. Section 4 describes the architecture of the ENIGMA system and explains the provided services. Section 5 reports the preliminary experimental results, whereas Section 6 concludes the paper.

2 RELATED WORK

Our work is related to different lines of research, including, visual localization, object detection and recognition and egocentric human-object interaction detection. The following sections discuss the relevant works belonging to the aforementioned research areas.

2.1 Visual Localization

Localization from egocentric images can be addressed considering both classification and camera pose estimation methods. In particular, classification based methods allow to localize the input image discretizing the space in cells and training a classifier which assigns the image to a cell. These cells can represent different generic areas (Torralba et al., 2003), daily-life environments (Furnari et al., 2018) or specific rooms of a museum (Ragusa et al., 2020). Instead, camera pose estimation methods establish correspondences between 2D pixels positions in the input image and 3D scene coordinates. This phase can be addressed using a matching algorithm or by regressing 3D coordinates from image patches (Brachmann and Rother, 2018; Taira et al., 2018). In this work, we focus on approaches based on both classification and camera pose estimation to give localization information to workers at different level of granularity.

2.2 Object Detection and Recognition

Object detection and recognition task has been tackled exploiting one-stage methods (Redmon and Farhadi, 2018) which prioritize the speed of detection over the accuracy of prediction, as well as two-stages approaches (Girshick, 2015), (Ren et al., 2015) which localize objects and classify them more precisely at higher computational time. Several works addressed the task of detecting and recognizing objects considering museums (Seidenari et al., 2017; Farinella et al., 2019) in which objects are represented by statues or artworks as well as industrial environments in which

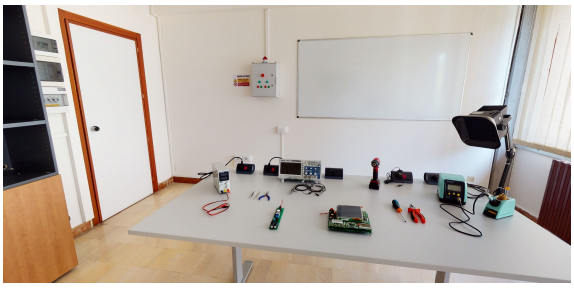

Figure 2: The industrial-like laboratory where the ENIGMA system has been tested.

tiny and small objects need to be recognized (Ragusa et al., 2021; Ragusa et al., 2022). The proposed ENIGMA system leverages state of the art object detectors to recognize objects present on the worktable of the considered industrial laboratory. Specifically, our system depends on the Faster-RCNN object detector (Ren et al., 2015).

2.3 Egocentric Human-Object Interaction Detection

Previous works focused on the Human-Object Interaction (HOI) detection task considering third person view. The authors of (Gupta and Malik, 2015) were the first to explore the HOI task annotating the COCO dataset (Lin et al., 2014) with verbs. (Gkioxari et al., 2018) proposed a method which detects and localizes humans and objects present in the scene, analyzes each human-object pair using a heat map to represent their relationship as well as to estimate the verb which describes it. The aforementioned problem has been studied also from the first point of view. The authors of (Nagarajan et al., 2019) studied the problem of understanding how to interact with an object, learning human-object interaction "hotspots" from egocentric videos. (Nagarajan et al., 2020) introduced a model to capture primary spatial zones of an environment and the possible activities which could happen there (i.e. environment affordance). The authors of (Shan et al., 2020) proposed an hand-centric method which classifies objects into the *active* and *passive* classes depending if they are or not involved in an interaction. A few works addressed this task considering an industrial domain. The authors of (Ragusa et al., 2021; Ragusa et al., 2022) studied human-object interaction releasing the MECCANO dataset while people building a toy model of a motorcycle. (Leonardi et al., 2022) studied egocentric human-object interaction exploiting both synthetic and real images in an industrial environment.

In the proposed system we adopted an hand-centric method based on a standard object detector

Figure 3: The 8 contexts of the industrial laboratory.

(Ren et al., 2015) to detect and recognize hands and objects in the scene as well as to understand their relationship considering the overlap between bounding boxes.

3 EXPERIMENTAL LABORATORY AND DATASETS

Our system has been tested in an industrial context. Specifically, we set up a laboratory (as shown in Figure 2) in which there are 23 different objects such as an electric screwdriver, a welding station and electrical boards as well as there is an electrical panel which allows powering on and off the sockets placed in the worktable. We have collected and labeled two different datasets of egocentric videos useful to design the services which compose the ENIGMA system: 1) Localization, 2) Object Detection and Recognition, 3) Egocentric Human-Object Interaction and 4) Next-active Object Detection.

3.1 Localization Dataset

We acquired 62 videos using using Hololens 2 device with a resolution of 2272x1278 at 30 frame per second. We extracted 55824 frames which have been divided into Training, Validation and Test sets considering 39437, 4394 and 11993 frames respectively. We labeled the dataset exploiting a Structure from Motion (SfM) approach using the open source software COLMAP[1], obtaining for each frame the 2D/3D positions and its orientation. The collected dataset is useful to assess the performances of algorithms for both punctual and contextual localization. In particular, for

[1]https://github.com/colmap/colmap

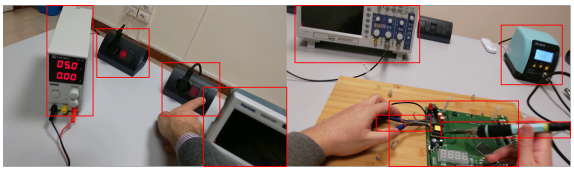

Figure 4: Examples of images annotated with bounding boxes around the objects.

the contextual localization we divided the laboratory into 8 cells which represents 8 different contexts: *C1 Lab Door, C2 Panel A, C3 Panel B, C4 Fire Extinguisher, C5 Workbench, C6 Cabinet, C7 Lab* (which represents where the user is in the lab but is not in any of the other cells) and *C8 Out of the Lab*. We assigned each frame to the correspondent cell considering its 2D position. Figure 3 shows the 8 contexts in the industrial laboratory.

3.2 Procedural Dataset

This dataset consists in 8 egocentric videos acquired with a Microsoft Hololens 2 device while 7 different subjects performed test and repair procedures on electrical boards in the industrial laboratory. The 8 videos have been acquired with a resolution of 2272x1278 with a framerate of 30 fps. We manually annotated human-object interactions selecting the first frame in which the hand of the subject touches an object and the frame after the hand releases it and assigning a verb which describes the interaction: 1) *Take*, 2) *Release*, 3) *Contact* and 4) *De-contact*. Moreover, for each frame we annotated both the objects which are involved in the human-object interaction (*active objects*) and all the other objects. In particular, we annotated each object with *(x, y, w, h, c, a)* tuple where *(x,y,w,h)* represent the 2D coordinates of the bounding box, *c* indicates the object class considering a total of 23 object classes and *a* indicates if the object is involved in the current interaction or not. Following this procedure we labeled 20000 objects. Figure 4 shows some examples of the annotated frames.

4 ARCHITECTURE AND SERVICES

In this Section, we first discuss the architecture of the proposed system (Section 4.1), then we present the services implemented by ENIGMA (Section 4.2).

4.1 Architecture

Figure 5 shows the whole architecture of the proposed ENIGMA system which is composed of 4 main components:

- **Wearable Devices:** devices such as smartglasses (i.e. Microsoft Hololens 2) are provided to the operator in the industrial laboratory. Repair and testing activities are shown on the screen through Augmented Reality. Moreover, images and videos are acquired from the point of view of the subject and sent to the Artificial Intelligence Inference Engine via a dedicated A/V Message Broker;
- **AI Inference Engine:** high performance multi core processing unit specifically designed for AI tasks. This engine executes AI algorithms on a dedicated GPU in order to process egocentric videos and address operator localization, object detection and recognition, human-object interactions and next-active object detection;
- **BI Logic Engine:** collects information from AI inferences and IoT sensors status to take decisions considering the behavior of the operator (e.g., turn off the electric power in case of electrical risk). Messages are exchanged through a message broker specifically designed for short messages protocols;
- **IoT Devices**: sensing and controlling periphery of the ENIGMA system. They continuously check the status of electronic devices connected to the sockets (e.g., oscilloscope, power supply).

Furthermore, the system relies on different standard modules which enable 1) the communication between modules (message broker), fast-access storage (IMDB), persistent storage (non-relational database), REST API services (web server - python modules) and administration, remote control and analytics (web application).

4.2 Services

This Section presents the services implemented by ENIGMA:

- **Augmented Reality.** The ENIGMA system uses Augmented Reality in order to provide additional and meaningful information while keeping the operator's hands free to work safely. Alarms and warnings will be shown onto the holographic glass and the whole interface can be controlled using vocal commands.
- **Localization.** This information is used to provide suggestions and alerts through the Augmented

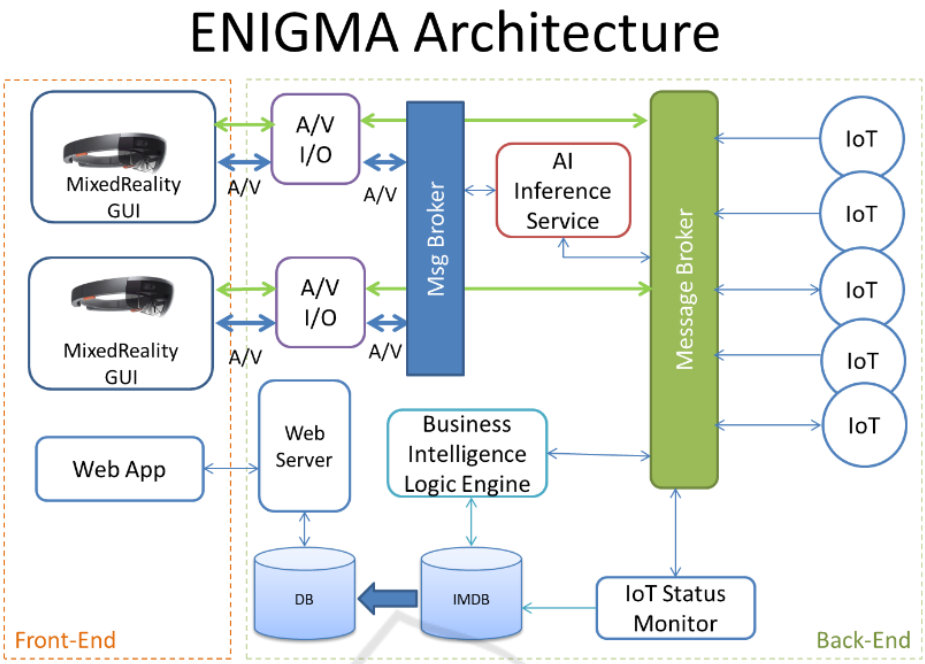

Figure 5: Overall architecture of ENIGMA system which is composed of 4 main components.

Reality, as well as to show the position of the operator in the web application, in order to log and monitor his activities.

- **Object Detection.** Recognizing the objects present in the surrounding environment allows the system to provide useful information to the worker about the objects such as time of usage for maintenance purposes or suggestions on how to use that specific object.
- **Human-Object Interaction.** This service allows to check the correctness of the procedures which the operator is doing, and also to implement energy saving strategies and tools preservation in an automatic way.
- **Next-Active Object.** Predicting which object is going to be used in the near future can prevent dangerous interactions, such as touching powered devices or tools. These information are used to alert the operator and to improve his safety.
- **Context-based Services.** A laboratory equipped with the ENIGMA system can be used as training space for people who needs to improve their skills. The capabilities of the system, allows also remote training and assistance, ensuring safety and health of both operator and instructor.
- **Authentication, Accounting and Administration.** The ENIGMA system relies on an administrative backend that provides control and monitoring of the system itself and also to the workers and their activities. A web application, enables the supervisors and the administrators to assign the activities to the operators and to monitor the status of the work. Moreover, the operator can consult his state of work using the same interface.
- **Visual Analytics.** An overall view of the laboratory, with all the devices and the operator, is available through a web based interface. The system reports useful information, such energy consumption, peaks, detected alerts, and so on, as well as provides an historical log of the events that happened in the laboratory.
- **Energy Saving.** The proposed system allows to optimize energy savings within laboratory due to the analysis provided by the energy consumption sensors present in the IOT devices installed in the laboratory. The system can warn operators or autonomously deactivate working tools that are not necessary for the current task allowing to save energy due to any worker oversights.
- **Safety.** The system is able to detect and provide solutions to the following safety risks:
  - *High/Low Voltage*: working with high and low voltage electrical boards exposes the operator to the electrical risks, due to the fact that the board could be under current, or to the residual electricity stored in the capacitors. The system is able to alert workers with visual and acoustic alarms before the board is touched and to

Table 1: The results obtained by the proposed system in the task of contextual localization.

Contexts	Validation	Test
C1 Door	0.936	0.584
C2 Panel A	0.986	0.831
C3 Panel B	0.985	0.727
C4 Fire Extinguisher	0.980	0.870
C5 Workbench	0.937	0.689
C6 Cabinet	0.964	0.242
C7 Lab	0.981	0.752
C8 Out of the Lab	0.960	0.294
Average	**0.975**	**0.647**

turn off the electrical current preventing electrical shocks;

- *Break from Work*: work's regulations prescribe to take breaks at regular time intervals. ENIGMA can monitor how long the operator has been working continuously, and suggest when it's time to take a break.
- *Safety Procedures*: specific tools and equipment need specific procedures and usage modalities. The system provides hints and suggestions and checks that the procedures are respected, improving the safety of workers.

5 PRELIMINARY RESULTS

We tested our ENIGMA system to assess the performances of localization (contextual and punctual), object detection and recognition and egocentric human-object interaction detection tasks, which represent three of the main cores of the whole system. Table 1 reports the results of the context-based localization system based on TripletNet (Hoffer and Ailon, 2015) for the feature extraction phase and a K-NN with $K = 1$ to assign the correct context.

Table 2 reports the results of the punctual localization task also based on TripletNet (Hoffer and Ailon, 2015) considering both 6 degrees (Table 2-top) and 3 degrees (Table 2-bottom) of freedom. We reported the mean and the median errors considering position (meters), quaternion rotation (degrees) and Euler angles (degrees).

Table 3 shows the results for the object detection and recognition task. We report the Average Precision (AP) for each of the 23 object classes. We also computed the mean Average Precision (mAP) measure with an *Intersection over Union (IoU)* of 0.5 (mAP@50). We obtained an mAP@50 over the Test set of 73.41%. Results suggest that the system is able to recognize well large objects such as *oscilloscope* or the *socket* obtaining a mAP of 90.12% and 90.27%

Table 2: The results obtained by the proposed system in the task of punctual localization considering both 6 (top) and 3 (bottom) degrees of freedom.

	Validation		Test	
Errors	**Avg**	**Median**	**Avg**	**Median**
Position	0.034	0.012	0.787	0.406
Quaternion	36.28	01.50	29.82	15.49
X angle	1.095	0.517	7.130	4.934
Y angle	0.756	0.315	5.116	3.732
Z angle	1.874	0.798	25.889	10.902

	Validation		Test	
Errors	**Avg**	**Median**	**Avg**	**Median**
Position	0.031	0.011	0.769	0.386
Angle	1.874	0.798	25.889	10.902

Table 3: The results obtained by the object detector on the industrial laboratory.

Object Category	AP	Object Category	AP
Power Supply	80.18	Working Area	90.18
Oscilloscope	90.12	Welder Base	88.82
Welder Station	89.87	Socket	90.27
Electric Screwdriver	81.45	Left Red Button	100.00
Screwdriver	58.73	Left Green Button	100.00
Pliers	79.18	Right Red Button	81.82
Welder Probe Tip	50.63	Right Green Button	90.91
Oscilloscope Probe Tip	51.72	Power Supply Cables	41.34
Low Voltage Board	88.53	Ground Clip	44.84
High Voltage Board	61.44	Battery Charger Connector	15.91
Register	71.07	Electric Panel	89.77
Electric Screwdriver Battery	51.72		

respectively, whereas it has trouble to recognize small objects such as the *Power Supply Cables* (mAP of 41.34%) or the *Battery Charger Connector* (mAP of 15.91%).

Furthermore, we addressed the egocentric human-object interaction detection task detecting all the active objects involved in the interactions. We trained an hand-centric method based on a standard object detector (Ren et al., 2015) on 6 videos of the procedural dataset and tested on the 2 remaining videos. We choose the active objects filtering all the detected objects considering the minimum distance between the centers of hands and objects bounding boxes. We evaluated the performance computing the mAP@50 and the mean Average Recall (mAR) obtaining a value of 36.80% and 28.31% respectively.

6 CONCLUSION

We have presented ENIGMA, a wearable system able to assist workers in an industrial laboratory providing information about the surrounding environment as well as improving the safety of workers and implementing energy saving strategies. Preliminary experiments show good performance considering localization, object detection and recognition and egocen-

tric human-object interaction tasks. Future work are related to the improvement of these services as well as the integration of the next-active object detection service.

ACKNOWLEDGEMENTS

This research has been supported by Next Vision s.r.l., by the project MISE - PON I&C 2014-2020 - Progetto ENIGMA - Prog n. F/190050/02/X44 – CUP: B61B19000520008 and by MEGABIT - PIAno di inCEntivi per la RIcerca di Ateneo 2020/2022 (PIACERI) – linea di intervento 2, DMI - University of Catania.

REFERENCES

Brachmann, E. and Rother, C. (2018). Learning less is more - 6d camera localization via 3d surface regression. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Colombo, S., Lim, Y., and Casalegno, F. (2019). Deep vision shield: Assessing the use of hmd and wearable sensors in a smart safety device. In *ACM PETRA*.

Cucchiara, R. and Bimbo, A. D. (2014). Visions for augmented cultural heritage experience. *IEEE MultiMedia*, 21(1):74–82.

Damen, D., Leelasawassuk, T., Haines, O., Calway, A., and Mayol-Cuevas, W. (2014). You-do, i-learn: Discovering task relevant objects and their modes of interaction from multi-user egocentric video. In *BMVC*.

Farinella, G. M., Signorello, G., Battiato, S., Furnari, A., Ragusa, F., Leonardi, R., Ragusa, E., Scuderi, E., Lopes, A., Santo, L., and Samarotto, M. (2019). Vedi: Vision exploitation for data interpretation. In *ICIAP*.

Furnari, A., Battiato, S., and Farinella, G. M. (2018). Personal-location-based temporal segmentation of egocentric video for lifelogging applications. *Journal of Visual Communication and Image Representation*, 52:1–12.

Girshick, R. (2015). Fast R-CNN. In *ICCV*.

Gkioxari, G., Girshick, R. B., Dollár, P., and He, K. (2018). Detecting and recognizing human-object interactions. *CVPR*, pages 8359–8367.

Gupta, S. and Malik, J. (2015). Visual semantic role labeling. *ArXiv*, abs/1505.04474.

Gurevich, P., Lanir, J., Cohen, B., and Stone, R. (2012). Teleadvisor: a versatile augmented reality tool for remote assistance. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*.

Hoffer, E. and Ailon, N. (2015). Deep metric learning using triplet network. In Feragen, A., Pelillo, M., and Loog, M., editors, *Similarity-Based Pattern Recognition*, pages 84–92. Springer International Publishing.

Leonardi, R., Ragusa, F., Furnari, A., and Farinella, G. M. (2022). Egocentric human-object interaction detection exploiting synthetic data.

Lin, T. Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., and Dollár, P. (2014). Microsoft coco: Common objects in context.

Nagarajan, T., Feichtenhofer, C., and Grauman, K. (2019). Grounded human-object interaction hotspots from video. In *ICCV*, pages 8687–8696.

Nagarajan, T., Li, Y., Feichtenhofer, C., and Grauman, K. (2020). Ego-topo: Environment affordances from egocentric video. *ArXiv*, abs/2001.04583.

Osti, F., de Amicis, R., Sanchez, C. A., Tilt, A. B., Prather, E., and Liverani, A. (2021). A vr training system for learning and skills development for construction workers. *Virtual Reality*, 25:523–538.

Ragusa, F., Furnari, A., Battiato, S., Signorello, G., and Farinella, G. M. (2020). EGO-CH: Dataset and fundamental tasks for visitors behavioral understanding using egocentric vision. *Pattern Recognition Letters*.

Ragusa, F., Furnari, A., and Farinella, G. M. (2022). Meccano: A multimodal egocentric dataset for humans behavior understanding in the industrial-like domain.

Ragusa, F., Furnari, A., Livatino, S., and Farinella, G. M. (2021). The meccano dataset: Understanding human-object interactions from egocentric videos in an industrial-like domain. In *IEEE Winter Conference on Application of Computer Vision (WACV)*.

Rebol, M., Hood, C., Ranniger, C., Rutenberg, A., Sikka, N., Horan, E. M., Gütl, C., and Pietroszek, K. (2021). Remote assistance with mixed reality for procedural tasks. *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 653–654.

Redmon, J. and Farhadi, A. (2018). Yolov3: An incremental improvement. *CoRR*, abs/1804.02767.

Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. In *NeurIPS*, pages 91–99.

Seidenari, L., Baecchi, C., Uricchio, T., Ferracani, A., Bertini, M., and Bimbo, A. D. (2017). Deep artwork detection and retrieval for automatic context-aware audio guides. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 13(3s):35.

Shan, D., Geng, J., Shu, M., and Fouhey, D. (2020). Understanding human hands in contact at internet scale. In *CVPR*.

Sorko, S. R. and Brunnhofer, M. (2019). Potentials of augmented reality in training. *Procedia Manufacturing*.

Sun, L., Osman, H. A., and Lang, J. (2021). An augmented reality online assistance platform for repair tasks. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 17:1 – 23.

Taira, H., Okutomi, M., Sattler, T., Cimpoi, M., Pollefeys, M., Sivic, J., Pajdla, T., and Torii, A. (2018). Inloc: Indoor visual localization with dense matching and view

synthesis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Torralba, A., Murphy, K. P., Freeman, W. T., and Rubin, M. A. (2003). Context-based vision system for place and object recognition. In *Proceedings of the Ninth IEEE International Conference on Computer Vision - Volume 2*, ICCV '03, page 273. IEEE Computer Society.