Holo-BLSD – A Holographic Tool for Self-training and Self-evaluation of Emergency Response Skills

Francesco Strada, Andrea Bottino, *Member, IEEE,*
Fabrizio Lamberti, *Senior Member, IEEE,* Giulia Mormando, and Pier Luigi Ingrassia

***Abstract*—In case of cardiac arrest, prompt intervention of bystanders can be vital in saving lives. Basic Life Support and Defibrillation (BLSD) is a procedure designed to deliver a proficient emergency first response. Developing skills in BLSD in a large part of the population is a primary educational goal of resuscitation medicine. In this context, novel computer science technologies like Augmented Reality (AR) and Virtual Reality (VR) can alleviate some of the drawbacks of traditional instructor-led courses, especially concerning time and cost constraints. This paper presents Holo-BLSD, an AR system that allows users to learn and train the different operations involved in BLSD and receive an automatic assessment. The system uses a standard manikin which is "augmented" by an interactive virtual environment that reproduces realistic emergency scenarios. The proposed approach has been validated through a user study. Subjective results confirmed the usability of the devised tool and its capability to stimulate learners' attention. Objective results indicated no statistical significance in the differences between the examiners' evaluation of users who underwent traditional and AR training; they also showed a close agreement between expert and automatic assessments, suggesting that Holo-BLSD can be regarded as an effective self-learning method and a reliable self-evaluation tool.**

***Index Terms*—Basic Life Support and Defibrillation (BLSD), self-learning, self-evaluation, Augmented Reality, user study.**

I. INTRODUCTION

IN a sudden cardiac arrest (SCA), the heart abruptly stops beating, halting the blood flow to the brain and other vital organs, eventually causing the death of the victim if not treated within minutes. In 2017, the estimated annual incidence of SCA victims in Europe and United States was, respectively, 300,000 and 347,000 [1], [2], confirming SCA as one of the major causes of death in adults in developed countries [3].

In case of out-of-hospital cardiac arrests, early recognition and intervention are critical for patient survival. In these situations, the combination of cardiopulmonary resuscitation (CPR) delivered by lay bystanders and the use of an automated external defibrillator (AED), can more than double the victim's chance of survival [4]. Unfortunately, only a minority of arrest victims receive bystander CPR, mainly due to the rescuers' fear or inability to perform this procedure [5]. Thus, increasing the percentage of population able to deliver proficient CPR in an emergency can help save lives [6].

To this end, a primary educational goal in resuscitation is to train laypeople in Basic Life Support and Defibrillation (BLSD), i.e. the sequence of operations aimed at recognizing a patient in cardiac arrest and managing the first response emergency procedures. Currently, the gold standard for BLSD training are instructor-led courses, which include theory lectures and demonstrations. In order to provide an experiential learning experience, these courses integrate different simulations, ranging from script-based role play (aimed at improving team's dynamics and performance [7]) to the use of high-fidelity mannikins allowing hands-on practice of several critical care procedures [8]. Regrettably, instructor-led courses have been reported to be time consuming and costly [9], [10].

An effective alternative, for both laypeople and healthcare providers, is represented by well-designed self-instruction contents. Multiple studies have shown that their learning outcomes (in terms of cognitive effort, skill performance at course conclusion, and skill decay) are comparable with that of traditional courses [3]. Among the BLSD self-instruction methods, the computer-based ones offer several interesting features. Besides enabling the development of compelling learning approaches that integrates multimedia and interactive contents, they can be deployed on mobile devices [11], [12] and can also encompass the use of manikins for hands-on learning (like in traditional courses) [13]. Students can complete the program at their own pace, repeat the training at will, and monitor their progresses through self-assessment features provided by the application. Furthermore, a computer-based method is cost-effective in many ways, as it does not require the physical presence of an instructor, it can reduce travel time and it allows developing the application once and using it for many learners.

Given the opportunities provided by computer-based education, recent advances in Virtual Reality (VR) and Augmented Reality (AR) fostered the introduction of these emerging computing technologies in a number of different fields [14], including BLSD learning. VR allows the development of experiential learning environments where different scenarios can be recreated, including the possibility to simulate (real) situations involving various risks for both the patient and the rescuers in a (virtual) safe environment. AR can be even more powerful for the scenario at hand, since it allows users to enjoy the learning experience in a real environment, which is "augmented" with a number of virtual elements that provide the context for the emergency scenario being simulated.

Based on these premises, in this paper we present Holo-

F. Strada, A. Bottino, F. Lamberti are with the Department of Computer and Control Engineering, Politecnico di Torino, Torino, Italy, e-mail: (see http://www.polito.it).

P. Ingrassia is with the Department of Translational Medicine, Universitá del Piemonte Orientale, Novara, Italy, e-mail: pierluigi.ingrassia@med.uniupo.it.

G. Mormando is with the Universitá degli Studi di Padova Scuola di Medicina e Chirurgia, Padova, Italy, e-mail: giulia.mormando@gmail.com.

Manuscript received xxx; revised yyy.

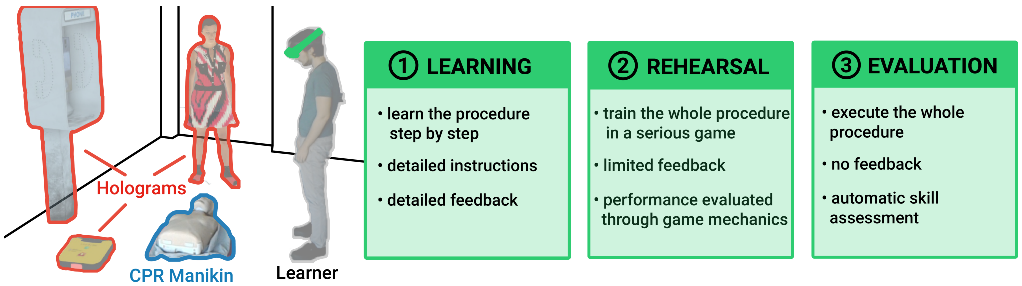

Fig. 1. Holo-BLSD: AR environment (left) and outline of the proposed learning path, including three modes (right).

BLSD, an AR tool for self-instruction training and self-assessment of BLSD skills. Holo-BLSD is an application for the Microsoft HoloLens headset that combines a standard low-cost CPR manikin with holographic interactive contents reproducing a realistic emergency scenario (Fig. 1, left). Within the application, learners can use natural gestures, body movements and spoken commands to perform their tasks. The rationale for choosing AR as technology is twofold. First, as far as training involves movements in a real environment, AR guarantees users higher confidence with respect to VR in performing their tasks. Second, the possibility to interact with digital elements in the real, physical environment can improve the knowledge transfer from the virtual to the real world.

The learning path of Holo-BLSD is divided into three different modes (Fig. 1, right): a *learning* mode, where users are guided step-by-step through the correct sequence of actions they have to perform to complete the procedures of interest; a *rehearsal* mode, envisioned as a serious game where users can train the intervention procedures they have learned in the previous phase; an *evaluation* mode, where trainees' BLSD skills are automatically assessed by the system. Users are free to repeat learning and rehearsal until they feel confident with the procedures and are ready to be evaluated.

To the best of our knowledge, this is the first project in the domain of BLSD that exploits AR to provide an all-in-one and complete self-learning, self-training and self-evaluation tool. The main features of Holo-BLSD can be summarized as follows:

- the sequence of managed actions can be tailored to different audiences (e.g., laypersons and healthcare providers), thus allowing the adaptation of the educational contents to the actual learners' need [6];
- its AR environment can be easily configured to simulate a variety of emergency scenarios (e.g., indoor, outdoor, with different hazards and victims' consciousness states);
- it guarantees hands-on learning through the use of a standard, low-cost CPR manikin;
- it helps learners to monitor their training progress by providing feedback on performed actions, and features an automatic self-assessment tool of the learned skills that reproduces traditional evaluation by limiting the impact of subjective and/or visual measurements;
- it includes a debriefing companion application that allows learners to review and critically analyze what they did.

The suitability of the proposed approach to BLSD training has been assessed through a user study that involved a panel of volunteers with no previous knowledge in BLSD and no previous experience with the HoloLens device. Experimental results demonstrated the usability of the proposed tool and highlighted its capability to stimulate learners' attention to levels similar to those achieved with traditional training. Furthermore, based on the comparison made between traditional/visual and automatic measurements, results also indicated that Holo-BLSD can be regarded as an effective learning method and a reliable self-evaluation tool.

The rest of the paper is organized as follows. After reviewing the state of the art (Section II), we detail the technical characteristics of Holo-BLSD (Section III) and the proposed learning path (Section IV). Then, we introduce the experimental design and setup (Section V), and we present the obtained results (Section VI). Finally, in Section VII we draw the conclusions and outline work envisioned for the future.

II. Related Work

Several VR and AR based approaches to CPR and BLS learning have been proposed in the literature. Mini-VREM [15] is a VR-enhanced manikin capable of rendering the main patient's vital signs and reactions to treatments. At the same time, the software provides real-time feedback on the quality of the chest compression during CPR by means of low-cost motion capture technologies based on the Microsoft Kinect sensor. A similar approach is pursued by RELIVE [16], a VR learning tool that leverages the Kinect to assess the correct arm pose and chest compression rate. Compared to the work reported in this paper, these approaches support only CPR training. Moreover, they merely provide a feedback on the performed actions, not a full assessment of learners' skills.

EMERGENZA [17] is a VR system that allows to simulate different first-aid scenarios and procedures, including BLS. Learners' performance can be assessed both on-line, by analyzing the actions performed in the virtual environment, and off-line, in a debriefing session based on the logs recorded during the training session. Compared to the devised Holo-BLSD application, this system focuses only on the training phase and does not provide a built-in support for learning. A more complete approach is proposed by LISSA (LIfe Support Simulation Application) [18], a desktop VR application that manages self-learning and self-assessment of BLSD procedures. The main difference with Holo-BLSD is the lack of a self-training part. Moreover, the self-assessment includes feedback aimed to help learners in their activities (opposite to the solution discussed in the current work, where no feedback is provided) and, most importantly, its reliability has not been validated through comparison with expert ratings.

As for AR, one of the first solutions to be proposed was ARLIST (Augmented Reality for Life Support Training, [19]), which exploits projector-based AR to augment a standard manikin with vital signs and facial expressions. More complex implementations of the approach adopted in ARLIST, which include also feedback for CPR rate and depth (based on sensors embedded in the manikin) as well as CPR trainee's posture (by means of a RGB-D camera), were discussed in [20] and [21]. The main limitations of these solutions are that they provide only real-time feedback on the CPR quality, and lack any self-instruction module. HoloCPR [22] is an AR application that exploits the Microsoft HoloLens to provide contextual information and real-time instructions for delivering CPR; however, AR is only used to overlay hints on the actions to perform and no evaluation of the learners' skills is actually considered. Given the recent capabilities (and performance) of mobile devices, developers also started to create AR apps for smartphones and tablets. An example is My Cardiac Coach [23], an app that combines interactive lessons with an AR-based CPR training system allowing users to practice CPR on a virtual victim and obtain feedback on their performance.

It is worth noting that some of the above works embed game mechanics and game elements within the learning process [16], [17], [18], [23]. Motivations for this choice build on the recognized value in education of Serious Games (SG), i.e., games that do not have entertainment as their primary purpose [24], [25], [26]. SGs provide an educational setting where entertainment and instruction are seamlessly integrated. Furthermore, they have the capability to let users feel immersed in the intended setting, providing them with an experience in which attention is fully turned to the desired contents [27], [28] (which, in turn, creates the ideal situation for learning to happen [29]). These are the reasons that guided the design of the rehearsal mode of Holo-BLSD as a SG.

Table I summarizes the main features of all the above methods, by indicating the problem(s) addressed, the possible use of a virtual environment to boost the sense of presence (and the device used to present it to the user), the support for self-learning, -training and -assessment phases (or the lack of it), the availability of an experimental comparison between automatic and expert ratings, i.e., of a validation, the use of manikins, the interactions considered (voice, gestures or mouse & keyboard), the need for external devices and the use of game mechanics within the learning process.

Based on the review performed, Holo-BLSD appears to be the only tool to provide a complete coverage of the various steps involved in the users' learning path, i.e., receiving instruction on the main actions to perform, train the whole procedure(s), assess the achieved skills and critically review performance in a debriefing session. It is also the only AR tool to provide context for the learning experience through a virtually-recreated emergency scenario, and to further boost users' immersion though the use of various natural interaction techniques. Furthermore, and most importantly, to the best of our knowledge, it is the only tool for which a confirmation of the reliability of automatic assessment has been obtained through a comparison, on a common rubric, of computer-based and traditional (that is, performed by instructors via visual inspection) evaluations. As a final note, it does not require the use of any external sensor or device, which makes it extremely easier to deploy and use.

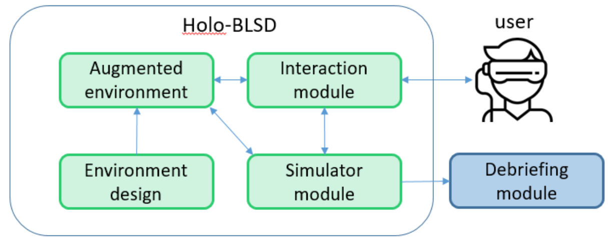

Fig. 2. The architecture of Holo-BLSD.

III. Application and Environment Design

Holo-BLSD has been envisaged as an application that should allow users experiment the BLSD procedures in their entirety and in different scenarios (e.g., hospital, street, mall, and so on). As a further requirement, it should be able to target different potential learner categories, such as laypeople and healthcare providers, each with its own peculiar learning requirements. Furthermore, various clinical cases should be simulated, including, e.g., a conscious victim requiring no intervention or an unconscious victim showing vital signs, which only requires the learner to call the emergency medical service (EMS).

Thus, Holo-BLSD has been designed to be flexible enough to support the different configurations, scenarios and action sequences mentioned above, as well as to facilitate the introduction of new elements, thus enabling future extensions (e.g., to support emergency team training).

The overall logical architecture of the system is shown in Fig. 2. The simulator module is responsible of managing all the activities involved in the specific learning scenario and learning mode selected. The session parameters can be selected through a configuration menu within the application. The user performs his or her activities in an augmented world that can generate different scenarios at varying complexity and that can be adapted to the actual physical location where the training is delivered (through an integrated editor). The interaction module is responsible for managing all the interactions between the learner and the augmented environment. Rehearsal and evaluation sessions can be recorded to be later analyzed in a debriefing step. The application has been created in Unity.

The rest of the section presents some of the technical aspects of the above modules, focusing in particular on the design of the simulator, of the AR environment and of the User Interface (UI). Presentation makes reference to the simulated scenario exploited for the experimental observations, which involved volunteers with no previous knowledge of BLSD; for this reason, the considered procedures include a modified version of CPR named continuous chest compression (CCC), which does not require interrupting the chest compression for rescue breathing. CCC is easier to perform than standard CPR, and

TABLE I
COMPARISON OF CPR/BLSD APPLICATIONS.

Method	Problem addr.	Virtual envir.	Self-learn.	Self-train.	Self-assess.	Exper. valid.	Manikin	Voice int.	Gesture int.	Mouse & keyb.	Ext. dev.	Game mech.
VR approaches												
MINI-VREM	CPR	no	no	yes	CPR rate/depth	no	yes	no	no	no	yes	no
RELIVE	CPR	yes (PC)	no	yes	CPR rate/depth	no	yes	no	no	yes	yes	yes
EMERGENZA	BLSD	yes (PC)	no	yes	Limited feedback	no	no	no	yes	yes	yes	yes
LISSA	BLSD	yes (PC)	yes	no	yes	no	yes	no	no	yes	yes	yes
AR approaches												
ARLIST	CPR	no	no	yes	no	no	yes	no	no	no	no	no
Park et al. [20]	CPR	no	no	yes	CPR rate/depth	no	yes	no	no	no	yes	no
Higashi et al. [21]	CPR	no	no	yes	CPR quality	no	yes	no	no	no	yes	no
HoloCPR	CPR	no	yes	no	no	no	yes	no	yes	no	no	no
MyCardiacCoach	CPR	yes (mobile)	yes	no	CPR rate	no	no	no	yes	no	no	yes
HoloBLSD	BLSD	yes (HMD)	yes	yes	yes	yes	yes	yes	yes	no	no	yes

was likely to be more acceptable for a layperson since it does not require mouth-to-mouth breathing [30].

In this emergency scenario, an adult is lying on the floor. The learner should first check the scene safety (eventually removing potential hazards such as wet, sharp objects or broken glasses), then evaluate the responsiveness of the victim or his or her vital signs (e.g., moving, coughing or breathing). Afterwards, the learner should call the local emergency number for getting professional support and ask a bystander, if present, to get an AED; otherwise, he or she should get one by himself or herself only if within reach. If an AED is not readily available, the learner should start chest compression immediately. When the AED becomes available, it should be operated by first switching it on, then plugging and placing the pads, and finally checking the victim's heart rhythm. Upon machine instruction, the learner must control that no one is touching the victim, deliver one electric shock and then restart chest compression until the AED suggests delivering another shock. This cycle must be repeated until the victim starts to breathe or the EMS arrives.

A. Simulator Module

The BLSD procedures, irrespective of the actual clinical case and simulation scenario, require to carry out certain actions in a specific order. The completion of an action involves the interaction with the environment (and objects in it), the possible bystanders and/or the victim.

In order to keep into account the adaptability and extensibility constraints previously mentioned, in designing the simulator we modeled dependencies between actions as a directed graph, where nodes represent individual actions and edges correspond to dependency requirements. The simulator is then responsible for the action scheduling. Each action is implemented as a single software module, and its execution flow is a function of the user's interactions with the virtual environment as well as of internal and external events. To further gain in flexibility when defining the action graphs of the procedures, we included the possibility to use composite nodes, which gather various sub-nodes implementing different algorithms for managing them (e.g., sequential or parallel execution and loop management). An example of action graph is shown in Fig. 3.

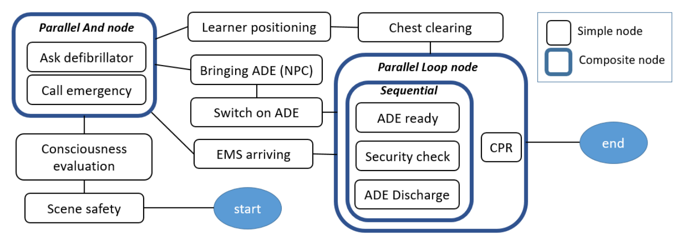

Fig. 3. Action graph for the clinical case corresponding to a patient unconscious and not showing vital signs. Boxes with a thin outline represent simple nodes, boxes with a thick outline are composite nodes. Start and end nodes represent, respectively, the beginning and end of the procedure.

With this approach, the procedures related to different clinical cases can be simply modeled as different action graphs, and targeting a novel audience (e.g., security officers in a chemical plant, etc.) can be addressed by implementing additional and target-specific actions (e.g., managing the spill of toxic substances).

B. Augmented Environment

The augmented environment manages all the virtual objects that are necessary to create the context of the simulated scenario. The environment can be also populated with non playing characters (NPCs) that can take different roles in the simulation. In order to improve the user's sense of presence, high-quality and realistic holographic contents were used. For the same purpose, the avatars included in the environment were animated using motion-capture data, and were provided with the capability to have voice-based and realistic (though limited) social interactions with the user.

The possibility to deploy the system in different real environments requires to adapt the placement of virtual elements in the emergency scenario to the actual setting. To this purpose, Holo-BLSD includes an environment design feature that exploits the HoloLens's *spatial anchors*, i.e., geometric descriptors that allow to register one or more persistent holograms with the real-world surfaces reconstructed by the device.

Holo-BLSD provides two environment design modes. The first is an interactive placement mode (Fig. 4), which leverages basic HoloLens features and offers the designer an easy and quick way to reconfigure the virtual environment at the cost of a coarse object placement. A second design mode has

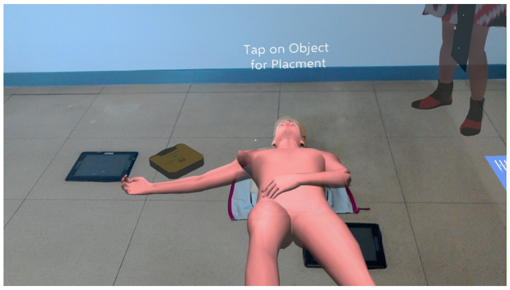

Fig. 4. An example of the interactive placement of virtual objects.

Fig. 5. Highlighting of a gaze-selected, or -targeted, object (the white dot is the gaze cursor).

been implemented ad-hoc for situations requiring precise and accurate anchor placement. This mode consists in elaborating offline (in a 3D editor) the geometric reconstruction of the real environment obtained from the HoloLens, then storing the result in a file that is processed by the application at startup.

In the emergency scenario, the CPR manikin is augmented by superimposing to it the full body of the victim in AR. Currently, rather than registering the victim's body with the manikin (which would require the availability of an extra tracking device, reducing portability and flexibility), the manikin is manually aligned with the virtual body in the real world. Any possible misalignment introduced during the execution of the procedures can then be easily adjusted at the beginning of a new session.

C. Interaction Module

In Holo-BLSD, users are engaged with an (augmented) environment which is experienced through the HoloLens device. The user controls both the camera position and the navigation inside the virtual environment with his or her own movements. Optical sensors are used to compute spatial mapping and positional information (via inside-out tracking) as well as to perform gesture recognition. A microphone array is used to capture the user's voice.

It should be noted that working with a holographic headset for the first time may represent a barrier for the users, and newbies often need significant practice and training before they get comfortable with this type of devices. This is a relevant issue that needed to be addressed in a careful way in our case. Users should be enabled to operate the system efficiently in the shortest time possible. In this way, they can spend their mental effort in learning and performing the actions required during the simulated emergency procedure, rather than in trying to understand how the AR system works.

Thus, the design of the UI was subject to the following interwoven constraints, all aimed at boosting intuitiveness and maximizing the User eXperience (UX). First, users should be provided with a limited set of interaction modalities in order to reduce their cognitive load[1]. Second, the system and the UI design should foster the learnability and memorability of the interface. Third, every type of interaction should include some sort of prompt and clear feedback from the UI; feedback is necessary to (i) inform the users when an interaction is available, (ii) allow them to predict the result of this interaction, and (iii) notify the success/failure in performing a task.

According to the aforementioned constraints, the set of required interactions was limited to object selection & dragging and to voice interaction. These interactions build upon the ability to select the object to interact with. Selection is controlled by the user's gaze. An highlight was added to any interactable object to signal users when it is "active", i.e., selected (Fig. 5).

Actions can be taken on the selected object either by means of a tapping gesture (an "air tap") or by pressing the button of the Clicker, a peripheral device of the HoloLens. In both cases, a selection event is triggered and proper per-object audio and/or visual feedback is provided. As an example, a specific sound underlines when a broken glass is selected to be removed from the scene (as requested by the procedure).

Users can also interactively move objects around the environment by dragging them. Dragging is performed as follows: once the object has been selected, its position is "hooked" along the current gaze direction at a suitable distance from the user; the drag is then completed by "clicking" with the air tap or with the Clicker. As an option, it is also possible to add a target in the environment to indicate where the object should be dropped. This is done, for instance, to indicate in learning mode where to place AED pads.

The last interaction considered is based on voice and it is used in two different ways. The first one is to recognize phrases in simulated dialogues, which have to be managed, for instance, when the user calls the EMS or asks a bystander (NPC) to get an AED. This modality is first activated by selecting a voice-responsive object (e.g., a telephone box or a bystander). When the volume of the user's voice is higher than a suitable threshold, the beginning of a phrase is detected; when the volume returns below the threshold, the end of the phrase is marked. Although extremely simple, this method is indeed effective for the task at hand. The second use of the voice is to recognize two simple keywords, "yes" and "no", which are used to reply to direct questions asked by the system (i.e., to check if the victim is conscious or shows vital signs). Using only these two simple keywords makes sure that the user's commands are interpreted unambiguously. In both cases, a small volume meter in the UI indicates when voice interaction is enabled; proper per-object feedback is provided when voice interaction is completed.

[1]In UX design, the term cognitive load refers to the amount of mental resources that is required to operate a system.

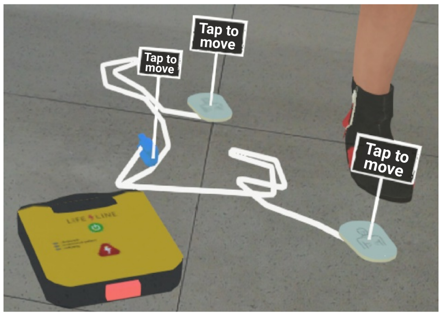

Fig. 6. Virtual AED: Tooltips attached to the defibrillator paddles.

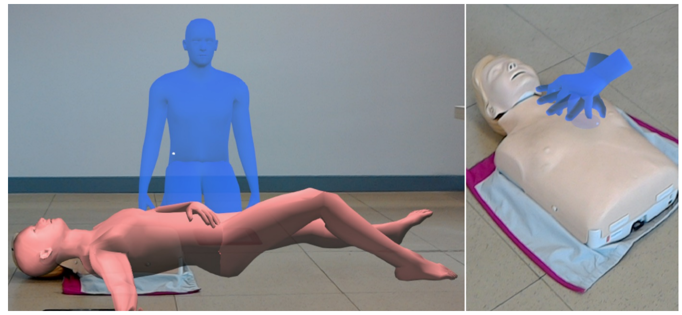

Fig. 7. Visual hints indicating the correct learner's position (left) and hands placement on the upper chest (right) when executing the CPR.

D. Interaction Training

Users can get acquainted with the interaction system through a training session. All interaction types are introduced individually (and incrementally) in the following order: targeting, air tap, Clicker, drag, voice interaction. Users first receive detailed instructions on the actual interaction method they are going to experience through voice and visual clues (animated GIFs). Then, they are requested to try each interaction at least two times. Every interaction element is characterized by an icon, which is introduced in this step and then used in the learning phase as a reminder of the types of interaction required to complete an action.

Users can repeat the interaction training until they become confident with the various methods. Since starting any session requires users to push a button in the main menu, which is an ability they might not have learned yet, if the system does not detect any user interaction within a predefined time interval, it automatically starts a short voice tutorial and then asks the learner if it should activate the interaction training session.

IV. Holo-BLSD Learning Path

As said, the learning path of Holo-BLSD includes three modes, which are designed to allow learners to (i) receive step–by–step instructions on the procedure they have to perform (learning mode), (ii) train on that procedure in the context of a serious-game (rehearsal mode), and (iii) get an automatic assessment of their BLSD skills (evaluation mode). In the following, the three modes are described in detail. A debriefing phase is also discussed.

A. Learning Mode

In a learning session, users are guided through the various actions of the given emergency procedure. Each action is introduced by visual and audio hints, aimed at explaining learners what they have to do, why, which are the objects they have to interact with and through which mechanics.

Objects and hints are introduced one at a time, with the aim to keep users' cognitive load low. Objects are presented as interactable (a glow effect is added when they are gazed) and provide consistent feedback in response to users' operations. Icons and other graphic signs (e.g., tooltips) are exploited as remainders of the interaction required by/allowed on a given object (Fig. 6). When dragging needs to be used on an object, a virtual target is displayed to indicate where it can be dropped. Since interaction is situated in a full 360° space, voice prompts were integrated to encourage users to look around and explore the whole environment, thus helping them to find possible off-screen objects. As an example, when an AED has to be requested and no actions are recorded within a certain time interval, the learner is guided to look in the direction where a bystander can be found.

In the BLSD training, learning how to perform a correct chest compression is crucial; thus, this action received a particular attention. Learners are first instructed to position themselves in the correct way. This goal is accomplished with the help of an avatar, which shows the learners how to kneel beside the victim's upper chest and to place their hands in the proper place (Fig. 7). Audio and visual instructions explain all the details of an effective chest compression. An audio feedback is provided for every chest compression detected by the system, and a metronome at 110 BPM helps learners to keep the right rhythm. Learners also receive a feedback on the actual quality of the cardiac massage in terms of compression frequency, which is displayed on a virtual panel placed in front of them. The method used to monitor the chest compression is discussed in the Appendix. It is worth noting that compression frequency data are relevant for all the Holo-BLSD modes; besides serving to provide users with an immediate feedback on the correctness of their operations in the learning and in the rehearsal modes, they are used to measure and automatically evaluate the learners' performance in both the rehearsal and evaluation modes.

The simulated clinical case (victim unconscious, conscious or showing vital signs) is generated randomly at every given session with a non-uniform distribution favoring the most complex case (victim unconscious not showing vital signs, which is proposed in 70% of the sessions). This way, learners are forced to experiment how to deal with different emergency situations characterized by varying difficulty.

B. Rehearsal Mode: A BLSD Serious Game

In this mode, users can train the procedures they have learnt in the previous mode. Given the relevance of serious games in educational contexts (which has been discussed in Section II), rehearsal has been designed according to such principles. In the devised game, users' (players') goal is to maximize their score by completing in the proper way the sequence of actions required by the considered procedure.

The choices made in designing the game leverage results reported in [31]; authors of that work analyzed the effectiveness, for the purpose of learning BLS procedures, of two game design patterns, namely a timer pattern and a score pattern. The timer pattern enforces a time limit for completing a given task. The score pattern defines the score as a numerical representation of the player's success in the game. Experimental results showed that the combination of these two patterns had positive outcomes in terms of both users' knowledge gain and UX [31].

Within the game, players can freely perform any action, but they cannot benefit from the visual and audio aids available in the learning mode. The game logic manages an action timer and checks action completion according to an internal scheduler. Players collect points when an action is completed in a suitable time and in the proper order, as well as when they perform actions in the correct way (e.g., when they execute a chest compression at the right frequency). If the completion timer for a particular action expires, the player looses a certain amount of points per second. During the game, players can make mistakes characterized by a different severity, which are treated either as warnings or errors and promptly reported by the system. As an example, starting the chest compression before having secured the scene is treated as an error, whereas doing it prior to asking for an AED raises a warning. Warnings simply cause a loss of points, whereas errors cause the termination of the game session (thus ensuring that actions are performed only in the correct way).

The current score and the timer are displayed on a panel placed in the virtual environment, while audio and visual cues highlight specific game events. For instance, an alarm sound signals the timer expiration, whereas another sound informs the players when new points are obtained. At the end of the game, players can see their placement in the overall ranking; this feature has been added to foster competitive behaviors (i.e., to challenge players to beat their colleagues' scores), which have been proved to be beneficial to learning.

C. Evaluation Mode

The aim of this mode is to enable the assessment of learners' BLSD skills. Trainees are presented with the same emergency situation experienced during learning and rehearsal sessions, and are asked to complete all the required actions without any audio or visual help from the system. As in the rehearsal mode, learners are free to execute actions in any order, with the difference that in the evaluation mode warnings and errors contribute (negatively) to the final assessment without compromising the session (which is not terminated).

Evaluation is based on the assessment form reported in Table II (which will be discussed in Section V-B), whose items are rated based on actual users' activity. In the experiments performed in this work, the clinical case always encompassed an unconscious, non-breathing victim, in order to evaluate the learners on the most complex and challenging case available. Nonetheless, the evaluation mode can, in principle, consider any of the clinical cases and scenarios available.

D. Session Debriefing

The Holo-BLSD tool includes a debriefing companion application, which relies on the analytics collected during the rehearsal and evaluation sessions. The availability of a debriefing step is extremely relevant for knowledge retention, since it helps learners to reflect on what they did, get insights from their experience and make meaningful connections with the real world, thus enhancing transfer of knowledge and skills. Even when results are not as successful as the learners hoped, debriefing can still promote active learning by helping them to analyze mistakes made and explore alternative solutions [32].

The Holo-BLSD debriefing application is a server component that runs on a PC and allows for the visualization of an interactive Gantt-like chart showing the begin and end of each action, as well as all the events related to it (e.g., the removal of a debris, the placement of an AED pad, the detected chest compressions and so forth). The chart can be synced with a video recording of the session, captured from the Hololens: this feature can be used to visually inspect a particular action in detail and to possibly create a repository of training events.

V. Experiments

In order to analyze the suitability of the proposed tool, a user study was performed by involving 58 volunteers selected among Health-care Nursing first year students of the University of Eastern Piedmont in Novara, Italy. Volunteers underwent either a traditional instructor-led course or a self-training delivered by the developed AR-based tool. The main aim was to investigate:

- to what extent the devised AR-based training course can stimulate learners' attention compared to traditional training;
- the usability of the devised tool from multiple perspectives, including, among others, the learnability of the interaction techniques, the realism of the simulation and the flexibility of the experience;
- how results obtained by learners in the AR training course compare to those obtained with traditional training;
- whether automatic assessment of learners' performance made by the tool is consistent with the evaluation performed by examiners through visual inspection.

In the following, the methodology adopted for the study will be first introduced, by also presenting the objective and subjective (questionnaire-based) metrics devised to analyze the above aspects. All the questionnaires and video material related to the experiments are available for download[2].

A. Methodology

All the volunteers were first provided with an introduction to the BLSD procedure, which was delivered by a medical instructor with the support of ad hoc video contents. Afterwards, volunteers were randomly split in two groups, later referred to as the "traditional training" and the "AR training" groups.

[2] http://tiny.cc/holo-blsd

1) Traditional training group: this group included 29 volunteers aged between 19 and 47 (M = 21.34, STD = 5.59), 5 males and 24 females. One of the volunteers had a previous knowledge of the BLSD procedure and was not included in the evaluation.

Learners were organized in small groups of 2-3 people. First, an instructor showed them how to carry out the whole procedure, which includes the 21 actions listed in Table II. All possible situations were illustrated (i.e., conscious, unconscious but breathing and unconscious, non-breathing victim). Afterwards, the instructor invited the learners to individually carry out all the actions, by answering possible questions and correcting them as needed. This phase was repeated two times. Then, each learner was asked to perform the procedure alone, without any instructor's feedback. In case of a severe error, the learner was stopped and asked to repeat the session until the procedure was successfully completed.

Finally, the evaluation was started. As said, the clinical case simulated always encompassed an unconscious, not breathing victim. This way, the whole BLSD procedure could be assessed, including the use of the defibrillator. During the evaluation, the learners were requested to carry out the BLSD procedure autonomously. An examiner observed the operations without interacting with the learners and assigned a score to every action in the list (evaluation criteria are reported in Table II and will be discussed in detail in Section V-B).

2) AR training group: this group included 29 volunteers, aged between 19 and 34 (M = 20.76, STD = 2.85), 13 males and 16 females. One of them had a previous knowledge of the BLSD procedure and was not included in the evaluation. Concerning technology awareness, 6 of them have had a previous experience with VR (3 of them using hand controllers); similarly, 6 of them have used already AR applications (4 on mobile devices, 1 on wearable devices and 1 on both the devices). Out of the 28 volunteers involved in the evaluation, only 26 filled in the questionnaire used to collect subjective measurements.

Learners were first given time to familiarize with the HoloLens by running the interaction training (Section III-D). The session was repeated twice to let learners get acquainted with both the air tap gesture and the Clicker device. Then, they were allowed to choose the interaction means they preferred.

After that, volunteers were invited to run the learning session, where the actual BLSD training is delivered (Section IV-A). Differently than in the traditional training, there was no interaction with a human agent. Learners were allowed to execute this session at least twice. In fact, it is not uncommon for users who experiment AR for the first time to simply disregard provided instructions while they explore the virtual world. During the session, different clinical cases (conscious victim, unconscious breathing and unconscious non-breathing) were randomly selected. It was guaranteed that each volunteer experimented at least once the clinical case considered in the evaluation, which was the unconscious non-breathing case like in the traditional training.

At the conclusion of the learning session, volunteers were asked to engage in the rehearsal session (Section IV-B). Like in the traditional session, high-severity errors terminated the session and learners had to repeat rehearsal until they were able to successfully complete a session that contemplates the evaluation case.

Finally, volunteers entered the evaluation session, where they had a six-minute time limit to complete the procedure; afterwards, the session was automatically terminated. Like with the traditional training group, during the evaluation session an examiner observed the learners and assessed their performance against the score sheet in Table II.

Additionally, the same evaluations were collected automatically by the AR-based tool for most of the actions (as it will be shown in Section V-B). This way, it was possible to study to what extent the devised tool is able to replicate the examiner's evaluation and, hence, serves as a reliable self-assessment tool.

B. Evaluation Metrics

Evaluation encompassed both objective and subjective measurements. Metrics adopted are reported in the following subsections, whereas results obtained are discussed in Section VI.

1) Objective measurements: as introduced in Section V-A, learners' performance was evaluated in objective terms against the score sheet in Table II. In particular, each action in the BLSD procedure could be assigned a score from 0 to 2, with 0 meaning that the action was not started, 1 that the action was started but not completed (when a partial completion is possible), and 2 that the action was completed.

These scores could be modified to account for errors made by the learner. In general, errors correspond to actions that are started earlier than expected. When this is the case, if the action was completed its score is lowered from 2 to 1 (if action was not started or not completed, score remains unchanged). As an example, suppose that the learner calls the EMS before securing the scene. Since the first action follows the second in the procedure, the score of the two actions is lowered by 1 even though they have been both completed.

Actions concerning the CPR were scored in a different way. In particular, for actions 7 and 16 the timely start of chest compression was evaluated: if the compression was initiated within 30 seconds after the conclusion of the previous action, 2 points were assigned, otherwise 0. For actions 8 and 17, the quality of the chest compression rate was considered, and score was assigned based on the distance between the observed/measured rate and the advised one.

It is worth recalling that BLSD actions were evaluated using the same metrics by both the examiner (through visual inspection) and the AR-based tool (in an automatic way). However, as explained in the Appendix, at present the AR-based tool is not able to measure the CPR compression depth and chest recoil and, thus, it cannot assess actions 9–10 and 18–19 (marked by † in Table II). Hence, when comparing the examiner's and the AR tool's evaluations, these actions were not considered and the maximum score that could be reached was 34 (17×2).

The learner passes the examination if he or she obtains at least 60% of the available points (i.e., 21 or higher) and no major error was made. Major errors are as follows: none of the CPR phases was ever started; security procedure was

TABLE II
LIST OF ACTIONS OF THE BLSD PROCEDURE AND EVALUATION CRITERIA.

Action	Evaluation criteria
1. Scene safety	In the scene there are always 3 objects to remove. If no object is removed by the learner, 0 points; 1–2 objects removed, 1 point; 3 objects removed (action completed), 2 points.
2. LOC evaluation	Learner is requested to assess level of consciousness by shaking the victim and ask if he or she is all right with loud voice. Action is completed only if both the operations have been accomplished: for instance, calling the victim without shaking him or her (or vice versa) means that action is not completed (1 point).
3. Vital signs evaluation	Learner has to observe the victim at least for five seconds to see whether he or she is breathing or not (2 points), otherwise the action is considered as not started (0 points).
4. Emergency call	Action is assigned 2 points is the call is completed. If the learner makes errors in informing the medical services about victim's LOC and vital signs, score is reduced to 0.
5. Get AED	If the learner asks bystanders for the defibrillator, 2 points, otherwise 0 points (the lack of the defibrillator will make it impossible for the learner to start and complete steps 12–15 and 20–21).
6. Clear chest	If the learner clears victim's chest from the obstructing arm (which blocks CPR), 2 points, otherwise 0 points.
7. 1st CPR – Start	If compression is initiated within 30 second after having asked for the defibrillator, 2 points, otherwise 0 points.
8. 1st CPR – Rate	Score is assigned based on average compression rate (BPM): 95 $<$BPM $<$125, 2 points; 80 $<$BPM $<$95 or 125 $<$BPM $<$140, 1 point, BPM $<$80 or BPM $>$140, 0 points.
9. 1st CPR – Depth†	If the compressions are 5 cm deep, 2 points, otherwise 0 points. Not assessed by the AR tool.
10. 1st CPR – Expansion†	If the chest returns in a neutral position after each compression, 2 points, otherwise 0 points. Not assessed by the AR tool.
11. AED turned on	If the defibrillator is turned on, 2 points, otherwise 0 points.
12. Paddles placed	If the defibrillator's paddles are placed correctly on the victim's chest, 2 points, otherwise 0 points.
13. Paddles plugged in	If the paddles' connector is plugged in to the defibrillator, 2 points, otherwise 0 points.
14. 1st security protocol	Defibrillation security protocol requests to move away from the victim, use loud voice to ask bystander to do the same, look around to make sure that nobody approaches the victim. The AR tool only assesses the last two operations: if both of them are performed, 2 points; if just one of them is performed, 1 point; otherwise, 0 points.
15. 1st defibrillation	If the defibrillator is discharged after having been invited to do that by the device (ready signal), 2 points. If the defibrillator is never discharged, 0 points.
16. 2nd CPR – Start	If compression is initiated within 30 second after having discharged the defibrillator, 2 points, otherwise 0 points.
17. 2nd CPR – Rate	Same as for action 8.
18. 2nd CPR – Depth†	Same as for action 9.
19. 2nd CPR – Expansion†	Same as for action 10.
20. 2nd security protocol	Same as for action 14.
21. 2nd defibrillation	Same as for action 15.

never started; emergency call was never made; defibrillator was never requested; defibrillator was never switched on; paddles were never placed on the victim's body nor plugged into the defibrillator; defibrillator was never discharged.

2) Subjective measurements: the suitability of the devised AR-based tool (and its contents) for self-training in the context of BLSD was analyzed through a questionnaire based on the Instructional Materials Motivation Survey (IMMS) [33]. This questionnaire was delivered to learners in both the AR training and traditional training groups. To this aim, statements in the original IMMS were slightly adapted to match the way training was actually delivered and to the material used.

A usability questionnaire was then delivered to the sole users of the AR-based tool to identify possible issues and drive future developments in the field. The questionnaire was organized in four sections. The first two sections analyzed usability in broad terms by considering the System Usability Scale (SUS) [34] and the five attributes defined by Nielsen [35], i.e., learnability, efficiency, memorability, (possibility to recover from) errors and satisfaction. The third section explored ergonomics aspects concerning the interaction with the device through statements derived from the ISO 9241-400 standard [36]. The fourth section explored in detail a number of usability aspects concerning virtual/synthetic environments defined in [37]. In particular, the questionnaire focused on user input (gaze, gestures and voice), sense of immersion/presence, system output (display), user guidance and help, consistency, simulation fidelity, flexibility, error correction/handling, i.e., robustness, and overall usability. Aspects in [37] related to functionality were not considered since they were addressed already in previous sections.

Lastly, a further questionnaire was delivered again only to the learners in the AR training group to assess their perception of the gamified learning approach adopted in the rehearsal session.

VI. RESULTS

In this section, the results of the subjective evaluation concerning learners' motivation levels, usability of the AR-based tool and suitability of the gamified learning approach will be discussed first. Afterwards, objective measurements collected by examiners in the two groups will be compared, and agreement between manual and automatic evaluations will be determined.

A. Subjective Results

The IMMS includes 36 statements organized in four sub-scales, which are aimed to investigate learners' motivation levels based on several principles of instructional design, i.e., attention (12 statements), relevance (nine statements), confidence (nine statements) and satisfaction (six statements). With the exception of 10 reverted statements, the higher the score the learner gives to a statement, the higher his or her motivational score is. Statements are evaluated on a scale from 1 (strong disagreement) to 5 (strong agreement).

TABLE III
LEARNER'S MOTIVATION LEVELS FOR THE TWO GROUPS: OVERALL RESULTS (NUMBER OF OCCURRENCES AND PERCENTAGES).

Motiv. level	Scores	Trad. training	AR training
High	4.21–5.00	18 (64,29%)	13 (50.00%)
Medium-high	3.41–4.20	9 (32,14%)	12 (46.15%)
Medium	2.60–3.40	1 (3,57%)	1 (3.85%)
Medium-low	1.80–2.59	0 (0%)	0 (0%)
Low	1.00–1.79	0 (0%)	0 (0%)

A scale reliability test was first conducted on the overall IMMS scale (36 statements). A standardized Cronbach Alpha equal to 0.876 ($n = 28$) and 0.919 ($n = 24$, 2 excluded) was calculated, respectively, for the traditional and the AR training groups, suggesting a good to excellent reliability of the results.

Overall motivation levels are reported in Table III on a 5-interval scale from low to high. For all learners except one in both groups, motivation was from medium-high to high (summing up to 96.43% and 96.15% of the respondents for, respectively, the traditional and AR training groups). These numbers indicate that both training methods were largely able to positively stimulate learners' motivation.

Averaged results for individual IMMS sub-scales (attention, relevance, confidence and satisfaction) are summarized in Table IV. Analyzing data using unpaired t-tests, the only differences that can be considered as statistically significant ($p < 0.05$) are those related to the confidence and satisfaction perspectives. In particular, AR learners reported a confidence and satisfaction lower than that of users in the traditional training group. These findings could be explained by the fact that the questionnaire asked respondents to focus only on course material. However, learners in the AR training group had to face the difficulties posed both by the BLSD contents and the use of a new tool and, in some cases, it was difficult for them to isolate content- from technology-related aspects.

Although the differences are not significant, slightly higher results can also be observed in the other two sub-scales for learners in the traditional training group.

For instance, considering the attention perspective and focusing on statistically significant statements, it can be noticed that learners in the AR training group found that "the amount of repetition caused them to get bored sometimes" and said that "the amount of information was so high that it was irritating". These outcomes are not surprising. In fact, the AR-based tool is currently designed to present material always in the same way (in terms of both visual and audio contents), independent of the fact that concepts have been assimilated or not. Moreover, in the experiments, learners were requested to carry out the BLSD procedure several times with only slight modifications (concerning victim's conditions and the use of gamification elements). This fact indeed contributed at making them perceive contents as repeated.

Similar considerations hold for the relevance dimension. For instance, higher scores assigned by learners in the AR training group to statements like "there are explanations or examples of how people use the knowledge of this course" can be easily explained by the lack of such contents in the AR-based procedure (whereas they could be provided by the instructor in the traditional training course, at least upon learners' request).

TABLE IV
LEARNER'S MOTIVATION LEVELS FOR THE TWO GROUPS: RESULTS FOR INDIVIDUAL SUB-SCALES (MEAN VALUES AND STANDARD DEVIATIONS REPORTED). ROWS MARKED WITH "*" SHOW A STATISTICALLY SIGNIFICANT DIFFERENCE ($p < 0.05$) AMONG THE GROUPS

Sub-scale	p-value	Trad. training	AR training
Attention	0,061	4.51 (0.52)	4.25 (0.47)
Relevance	0,230	4.39 (0.40)	4.24 (0.82)
Confidence*	0,013	4.15 (0.42)	3.83 (0.48)
Satisfaction*	0,045	4.57 (0.54)	4.23 (0.66)

TABLE V
STATEMENTS IN THE SUS TOOL (MEAN VALUES AND STANDARD DEVIATIONS REPORTED).

Statement	Score
I think that I would like to use this system frequently	3.89 (1.05)
I found the system unnecessarily complex	1.67 (0.92)
I thought the system was easy to use	3.89 (0.93)
I think that I would need the support of a technical person to be able to use this system	2.78 (1.09)
I found the various functions in this system were well integrated	3.63 (0.93)
I thought there was too much inconsistency in this system	1.48 (0.75)
I would imagine that most people would learn to use this system very quickly	3.44 (1.19)
I found the system very cumbersome to use	1.74 (0.90)
I felt very confident using the system	3.67 (1.00)
I needed to learn a lot of things before I could get going with this system	2.04 (1.09)

Despite these differences, it shall be noticed that, for all the scales, average levels in the AR training group are in the medium-high interval (according to the 5-interval scale used in Table III), which confirms the important level of motivation that the proposed tool can guarantee.

As previously introduced, usability was assessed through a questionnaire (delivered only to the AR training group) that included four sections based on (i) SUS, (ii) Nielsen's attributes of usability, (iii) the ISO 9241-400 standard about man-machine ergonomics and (iv) the VRUSE tool. All items had to be scored on a scale from 1 to 5 (with the same meaning of the IMMS tool).

Concerning SUS (first section), learners were asked to score the 10 items in Table V. The normalized result (with odd items reverted) equal to 72.03 in the 0–100 range can be regarded as an indication of the usability of the designed tool (according to [38], a score above 68 shall be considered as above average).

Similar conclusions can be drawn by considering Nielsen's attributes of usability (second section). Mean scores were as follow: learnability 3.63 (SD = 1.01), efficiency 3.93 (SD = 0.73), memorability 3.89 (SD = 1.01), possibility to recover from errors 3.04 (SD = 0.85) and satisfaction 4.44 (SD = 1.01). The low score assigned to the possibility to recover from errors can be explained by the fact that learners were generally not allowed to recover from mistakes made with the AR tool. In fact, these mistakes could correspond to errors in the BLSD procedure, which had to be recorded and evaluated as explained in Section V-A.

As for ergonomics (third section), learners were asked to evaluate the interaction with the HoloLens based on the four

TABLE VI
STATEMENTS FROM THE ISO 9241-400 STANDARD (MEAN VALUES AND STANDARD DEVIATIONS REPORTED).

Statement	Score
The wearable device is very cumbersome / heavy	2.15 (1.13)
Mental effort required to operate the device is very high	2.41 (1.05)
Physical effort required to operate the device is very high	2.26 (0.86)
I would feel comfortable using the system for long times	3.07 (1.04)

TABLE VII
CATEGORIES FROM THE VRUSE TOOL (MEAN VALUES AND STANDARD DEVIATIONS OF THE OVERALL SATISFACTION REPORTED).

Category	Score
User input (gaze and taps)	3.58 (0.76)
User input (voice)	3.77 (0.76)
Sense of immersion/presence	3.54 (0.81)
System output (display)	3.96 (0.65)
User guidance and help	4.00 (0.78)
Consistency	3.81 (0.83)
Simulation fidelity	3.89 (0.93)
Flexibility	3.73 (0.78)
Error correction/handling and robustness	3.58 (0.81)
Overall usability	4.36 (0.64)

statements in Table VI. Scores are in the medium-low to medium range, suggesting that developments are still required in the field of head-mounted AR devices.

The interaction with synthetic and virtual environments (fourth section) was analyzed through statements, adapted from VRUSE, belonging to 10 categories. For each category, learners were asked to additionally express their overall level of satisfaction on a scale from 1 (poorly satisfied) to 5 (very satisfied). Categories and overall scores are given in Table VII.

According to the scale used in Table III, all the categories received medium-high to high scores. This fact indeed represents a further confirmation of the appreciation and the usability level reached by the devised AR tool. Despite that, interesting insights for driving future developments can be obtained from the statements that obtained the lower scores in each category (although it is worth observing that worst scores were in the 2.60–3.40 medium range).

With respect to user input, learners found that the modality based on finger gestures was not ideal for interacting with virtual elements. Nevertheless, this fact was not particularly critical, since learners were allowed to choose between gesture- and Clicker- based interaction. However, similar concerns were raised also for voice interaction. In fact, learners stated that they did not feel to have always the right control on what they wanted to do while interacting with the system, mainly because of language recognition/understanding issues.

Learners stated that being immersed in the virtual experience was important for completing the assigned task, confirming the importance of using AR to create a simulated scenario. Regrettably, they also indicated that the characteristics of the screen (its limited field of view, in particular) partially reduced their sense of immersion (confirming that further advancements in wearable AR technology are needed).

Concerning system output, learners rated graphics quality as appropriate for the task, though realism was not judged as particularly high. They also stated to be able to read and understand the information displayed by the AR application. None of them experienced motion sickness or eye fatigue.

Although they did not find it difficult to learn how to use the system and to use it, learners said they needed external help while using it (notwithstanding, it is worth recalling that they had to carry out the learning, rehearsal and evaluation sessions without any human intervention).

Learners found that, in general, the system behaved in a consistent way, but in some cases the meaning of visual and audio cues was not as straightforward as they were expecting.

Concerning simulation, they found that the BLSD procedure was simulated with the appropriate fidelity and that simplifications that were possibly implemented did not impact their performance. However, they felt that virtual elements did not always behave and move in a natural way.

Scores regarding flexibility are not very high. In fact, learners felt that interaction modality sometimes interfered with their activity. Thus, they had to adapt their behavior to the system, and this fact sometimes prevented them to achieve exactly the intended result. They also found that the system lacks shortcuts to perform given operations.

With respect to error correction/error handling and robustness, learners lamented the fact that the system is not adopting strategies able to prevent them to make silly mistakes and that it is difficult to recover from errors. Although strategies could be devised to deal with errors deriving from interaction issues, it is worth recalling that means to recover from procedural errors were intentionally avoided (see discussion above).

Lastly, considering overall usability, learners found that the system's responsiveness partially impacted on their performance (time is needed to advance in the simulation, to react to user's interactions, etc.). Learners also stated that the system does not presently do all what they would expect. Indeed, motivations for this result could be identified in the issues discussed for previous categories.

As for the various problems highlighted in the previous paragraphs, it is worth saying that a possible way to tackle them could be to introduce in the AR tool some of the features that have been considered positively by the traditional training group. Examples of such features are the possibility to interact with a question-and-answer mechanism (e.g., mimicking the presence of a human instructor through conversational agents) and to let the learners tailor training to their actual needs (e.g., by selecting particular actions in the procedure to experiment with), or the adoption of different ways to present the same content (to avoid repetitions), and so on.

The analysis of the last questionnaire, whose nine items are reported in Table VIII with assigned scores, suggests a positive users' perception of the devised gamification approach. Learners appreciated the approach pursued to deliver the intended contents and considered it effective and fun. They also felt that education and entertainment aspects were well balanced, and judged the elements introduced to foster competitiveness (i.e., the score and the leaderboard) in a positive way.

Concluding, the subjective assessment of Holo-BLSD confirmed that the proposed AR-based learning path is able to stimulate learners' attention to levels similar to those achieved

TABLE VIII
STATEMENTS USED TO ASSESS PERCEPTION OF THE GAMIFIED APPROCH FOR TRAINING PURPOSES (MEAN VALUES AND STANDARD DEVIATIONS).

Category	**Score**
I found it fun to use a game to learn intended content	4.31 (0.93)
I would have preferred a different learning modality, based on the presence of an instructor	2.08 (1.29)
I would have preferred a different learning modality, based on books, notes, slides, etc.	1.73 (1.15)
The designed game is a valid tool to learn intended content	4.19 (0.80)
The possibility to compare my score with other learners scores made me try to improve my results	3.92 (1.06)
The possibility to play the game several times made me better understand the correct sequence in which actions have to be performed	4.08 (0.89)
I found the tool more a game than a system suitable for training	2.69 (1.23)
Trying to improve my results in the game let me learn intended content better	4.04 (1.04)
The presence of a timer stimulated me to quickly carry out required actions	3.42 (1.27)
I found the timer a stressful element	1.88 (1.20)

with traditional training. Results also demonstrated the usability of the devised tool. Nevertheless, the same results also allowed us to identify aspects that shall be considered to enhance the suitability of the proposed AR-based BLSD training in terms of both contents and technology (e.g., avoiding repetitions, improving interaction, etc.).

B. Objective Results

Objective results collected during the experiments are summarized in Table IX. The first column reports the actions of the BLSD procedure. The second and third columns tabulate scores assigned by the examiners in, respectively, the traditional (TE) and AR (AE) training groups (mean values and standard deviations reported). Statistical significance of the differences between TE and AE scores is measured using unpaired t-tests ($p = 0.05$). The fourth column provides calculated p-values. The fifth column (AT†) reports the automatic scores assigned by the tool in the AR course (actions that cannot be assessed so far are marked with † in the first column). The sixth column reports p-values calculated on the comparison between examiner's scores for the traditional training course (TE†) and automatic scores for the AR training group (AT†). Finally, the last column provides inter-rater agreement between the examiner (AE†) and the tool (AT†) when scoring the AR training group. These agreements were calculated using weighted Cohen's k. The last two rows of the table report the overall evaluation results, considering or not the actions that cannot be assessed automatically by the tool.

As a first comment, it can be seen that the average overall examiner scores of the two groups (second and third columns) are rather close (39.48 for the traditional training group, 37.07 for the AR training group, on a maximum score of 44), and that their difference is not statistically significant. This finding suggests that the learning outcomes achieved by the instructor-led and the AR-based courses are overall comparable, which is also a possible indication of the effectiveness of the proposed approach as a learning tool.

However, it is also worth observing that standard deviation is much higher for AR learners. This difference could be due to difficulties that (some of the) learners may have experienced in interacting with the AR tool, e.g., due to missed recognition of gesture and voice inputs. Another explanation could be the higher complexity for examiners to judge learners' performance. In fact, the execution of many actions required learners to interact with virtual elements and, even though examiners were allowed to see the point of view of the learners, delays due to data transmission made it difficult in some cases to fully appreciate their actual behavior. Lastly, this result may also indicate a lower capability of the AR tool to level learners abilities, which could be due, among others, to the lack of mechanisms for adapting contents and their presentation to learners' actual needs.

Similar considerations can be made when analyzing the results of individual actions. The cases in which differences are significant are only four, namely LOC and vital signs evaluation, start of first defibrillation and of second CPR (rows marked with * in the fourth column). In some cases, differences can be explained again with the difficulty for examiners to judge learners' operation. For instance, in the LOC evaluation, learners were expected to shake the victim and call him or her loud; in some cases examiners judged the force applied or the voice level used as not appropriate. However, examiners could not rely on any quantitative information in the assessment. Furthermore, differently than in the traditional training group, users of the AR tool did not had a ground truth, since they had not seen the instructor execute those actions. In the evaluation of the victim's vital signs, learners' have to observe the victim for five seconds; even using measurement instruments, examiners tended to approximate actual time. While performing the defibrillation, in some cases examiners were not able to determine whether actions had been performed in the proper way due to the subtle movements involved or fine precision requested (e.g., in plugging paddles' connector, or placing paddles on the victim's chest). With respect to CPR, it was quite difficult for the examiner to determine when to start measuring time. In the above situations, the AR tool was able to assign a truly objective score, based on measurements collected by internal sensors and rules defined in the application (e.g., on interactions performed, threshold levels passed, and so on).

A support for these explanation can come from comparing the examiner's scores in the traditional course (TE†) and the AR tool scores (AT†). It can be seen that only the differences concerning the first defibrillation and the second CPR remain significant (sixth column, rows marked with *). However, also this result can be easily explained, since the AR tool is not able to assign a score for the second CPR if defibrillation was not completed (CPR time is measured from discharge).

It is also worth observing that, for both courses, there are a number of actions that received low scores. This outcome is particularly relevant, since it suggests those parts of the learning path that should be improved, no matter how the course is going to be delivered. Other interesting insights can be obtained about the effectiveness of Holo-BLSD as a self-evaluation tool. Comparing the overall scores assigned in

TABLE IX
OBJECTIVE RESULTS FOR INDIVIDUAL ACTIONS AND FOR THE WHOLE BLSD PROCEDURE.

Action	Trad.–Ex. (TE)	AR–Ex. (AE)	p (TE–AE)	AR–Tool (AT†)	p (TE†-AT†)	k (AE†-AT†)
1. Scene safety	1.97 (0.19)	2.00 (0.00)	0.322	1.97 (0.19)	1.000	0.926
2. LOC evaluation	2.00 (0.00)	1.86 (0.35)	0.039*	1.93 (0.26)	0.161	–0.101
3. Vital signs evaluation	1.97 (0.19)	1.72 (0.45)	0.011*	1.93 (0.26)	0.562	–0.124
4. Emergency call	2.00 (0.00)	2.00 (0.00)	–	2.00 (0.00)	–	1.000
5. Get AED	1.86 (0.35)	1.90 (0.41)	0.732	1.90 (0.41)	0.732	1.000
6. Clear chest	1.93 (0.26)	1.97 (0.19)	0.561	1.90 (0.31)	0.647	0.055
1st CPR – Start	2.00 (0.00)	1.90 (0.31)	0.078	1.93 (0.37)	0.326	0.374
1st CPR – Rate	1.90 (0.31)	1.79 (0.41)	0.285	1.69 (0.66)	0.135	0.329
1st CPR – Depth†	1.83 (0.38)	1.76 (0.51)	0.564	–	–	–
1st CPR – Expansion†	2.00 (0.00)	1.90 (0.41)	0.179	–	–	–
AED turned on	1.97 (0.19)	1.86 (0.44)	0.249	1.83 (0.47)	0.149	0.877
Paddles placed	1.97 (0.19)	1.83 (0.47)	0.146	1.79 (0.62)	0.161	0.699
Pladdles plugged in	1.86 (0.44)	1.86 (0.44)	1.000	1.93 (0.37)	0.977	0.651
1st security	1.86 (0.35)	1.79 (0.56)	0.576	1.72 (0.65)	0.320	0.836
1st defibrillation	2.00 (0.00)	1.66 (0.77)	0.019*	1.72 (0.70)	0.043*	0.869
2nd CPR – Start	2.00 (0.00)	1.66 (0.77)	0.013*	1.66 (0.72)	0.023*	0.879
2nd CPR – Rate	1.76 (0.51)	1.48 (0.78)	0.118	1.45 (0.74)	0.021*	0.620
2nd CPR – Depth†	1.62 (0.56)	1.59 (0.73)	0.841	–	–	–
2nd CPR – Expansion†	1.93 (0.37)	1.66 (0.72)	0.072	–	–	–
2nd security	1.45 (0.74)	1.38 (0.90)	0.751	1.48 (0.87)	0.299	0.873
2nd defibrillation	1.62 (0.78)	1.52 (0.87)	0.635	1.52 (0.87)	0.055	1.000
All actions	39.48 (2.50)	37.07 (7.07)	0.088	–	–	–
All actions except †	32.11 (2.28)	30.34 (5.60)	–	30.17 (5.41)	0.109	0.794

the AR training group by the examiner (AE†) and the tool (AT†), it can be observed that mean values are comparable (30.17 vs. 30.34, with no significant difference), like standard deviations. This finding is confirmed by a Cohens k value equal to 0.794 (last column), which suggests a quite high inter-rater agreement. When considering individual actions, it can be noticed that there are a number of situations for which inter-rater agreement is lower. Based on discussion above, this result can be largely explained with the difficulty in providing objective results for such actions.

With the aim to summarize this discussion, several insights can be obtained from the analysis of the objective results. First, by comparing scores assigned by human examiners in the traditional and AR-based courses, it was observed that learners achieved comparable results, thus confirming the suitability of the devised self-learning tool for training considered skills. Second, since scores automatically computed by the tool were found to be largely consistent with those assigned by the (human) examiners to the same learners, it can be concluded that the proposed system can also be regarded as a reliable instrument for self-assessment.

VII. CONCLUSION

This paper presented Holo-BLSD, an AR based self-learning and self-evaluation tool that does not require professional instructor intervention, thus helping to maximize learning results while reducing associated costs. As major features, the system offers realistic tactile feedback through a CPR manikin, a virtual scenario that can be easily reconfigured to generate many different situations, including extreme and dangerous ones, and a flexible software architecture that allows to easily manage different target audiences and clinical cases.

The suitability of the proposed approach to BLSD training was assessed through a user study that involved a panel of volunteers. The main experimental results indicate that Holo-BLSD can provide both a learning experience similar to instructor-led training and a reliable automatic assessment of the learner's performance. Results also demonstrated the usability of the devised tool and the capability to stimulate learners' attention to levels similar to those achieved with traditional training. The above findings suggest that Holo-BLSD could be used as a cost-effective learning tool for supplementing traditional training and for possibly replacing it when appropriate.

Future work will include the development of several features that could further improve the proposed tool. As for the CPR quality assessment, methods capable of analyzing compression depth and chest recoil will be investigated, possibly without resorting to external sensors or devices. The use of conversational agents will also be explored, in order to mimic the presence of a human instructor and provide a more functional dialog-based interaction while executing the procedure. Finally, there are plans to make the application consider both co-located and remote emergency team training.

ACKNOWLEDGMENTS

Activities were partially supported by VR@POLITO. The authors wish to thank LogosNet and Fernando Salvetti, for kindly providing the devices used in development and evaluation steps, as well as Filippo Gabriele Pratticò and prof. Gabriele Garnero for contributing to the evaluation.

REFERENCES

[1] A. Timmis and al., "European society of cardiology: Cardiovascular disease statistics 2017," *European Heart Journal*, vol. 39, no. 7, pp. 508–579, 2018.

[2] E. J. Benjamin and al., "Heart disease and stroke statistics: A report from the american heart association," *Circulation*, vol. 137, no. 12, 2018.

[3] J. Yeung, D. Okamoto, J. Soar, and G. D. Perkins, "AED training and its impact on skill acquisition, retention and performance – A systematic review of alternative training methods," *Resuscitation*, vol. 82, no. 6, pp. 657 – 664, 2011.

[4] G. Ritter, R. A. Wolfe, S. Goldstein, J. Landis, C. Vasu, A. Acheson, R. Leighton, and S. V. Medendrop, "The effect of bystander cpr on survival of out-of-hospital cardiac arrest victims," *American Heart Journal*, vol. 110, no. 5, pp. 932 – 937, 1985.

[5] S. Girotra, S. van Diepen, B. K. Nallamothu, M. Carrel, K. Vellano, M. L. Anderson, B. McNally, B. S. Abella, C. Sasson, and P. S. Chan, "Regional variation in out-of-hospital cardiac arrest survival in the United States," *Circulation*, vol. 133, no. 22, pp. 2159–2168, 2016.

[6] R. Greif, A. S. Lockey, P. Conaghan, A. Lippert, W. D. Vries, and K. G. Monsieurs, "European resuscitation council guidelines for resuscitation 2015: Section 10. Education and implementation of resuscitation," *Resuscitation*, vol. 95, pp. 288 – 301, 2015.

[7] S. P. Chung, J. Cho, Y. S. Park, H. G. Kang, C. W. Kim, K. J. Song, H. Lim, and G. C. Cho, "Effects of script-based role play in cardiopulmonary resuscitation team training," *Emergency Medicine Journal*, vol. 28, no. 8, pp. 690–694, Sep. 2010.

[8] W. C. Mundell, C. C. Kennedy, J. H. Szostek, and D. A. Cook, "Simulation technology for resuscitation training: A systematic review and meta-analysis," *Resuscitation*, vol. 84, no. 9, pp. 1174–1183, 2013.

[9] W. de Vries, M. Schelvis, I. Rustemeijer, and J. J. Bierens, "Self-training in the use of automated external defibrillators: The same results for less money," *Resuscitation*, vol. 76, no. 1, pp. 76 – 82, 2008.

[10] E. Skura, "Pros and cons of first aid training?" *Canadian Medical Association Journal*, vol. 182, no. 12, pp. E549–E550, 2010.

[11] Stone Meadow Development LLC, "Resuscitate! CPR AED & Choking," 2010. [Online]. Available: https://itunes.apple.com/us/app/resuscitate-cpr-aed-choking/id363393502?mt=8

[12] Duke Health, "Duke CPR," 2015. [Online]. Available: https://itunes.apple.com/us/app/duke-cpr/id1022440083?mt=8

[13] Y. Kwon, S. Lee, J. Jeong, and W. Kim, "Heartisense: A novel approach to enable effective basic life support training without an instructor," in *CHI '14*, 2014, pp. 431–434.

[14] R. Shewaga, A. Uribe-Quevedo, B. Kapralos, and F. Alam, "A comparison of seated and room-scale virtual reality in a serious game for epidural preparation," *IEEE TETC*, pp. 1–1, 2018.

[15] F. Semeraro, A. Frisoli, C. Loconsole, F. Bann, G. Tammaro, G. Imbriaco, L. Marchetti, and E. L. Cerchiari, "Motion detection technology as a tool for cardiopulmonary resuscitation (CPR) quality training: A randomised crossover mannequin pilot study," *Resuscitation*, vol. 84, no. 4, pp. 501 – 507, 2013.

[16] F. Semeraro, A. Frisoli, C. Loconsole, N. Mastronicola, F. Stroppa, G. Ristagno, A. Scapigliati, L. Marchetti, and E. Cerchiari, "Kids (learn how to) save lives in the school with the serious game relive," *Resuscitation*, vol. 116, pp. 27–32, 2017.

[17] A. Ferracani, D. Pezzatini, L. Seidenari, and A. Del Bimbo, "Natural and virtual environments for the training of emergency medicine personnel," *Univ. Access in the Inf. Society*, vol. 14, no. 3, pp. 351–362, 2015.

[18] V. Wattanasoontorn, M. Magdics, I. Boada, and M. Sbert, "A Kinect-based system for cardiopulmonary resuscitation simulation: A pilot study," in *Serious Games Development and Applications*, M. Ma, M. F. Oliveira, S. Petersen, and J. B. Hauge, Eds., Berlin, 2013, pp. 51–63.

[19] F. Pretto, I. H. Manssour, M. H. I. Lopes, and M. S. Pinho, "Experiences using augmented reality environment for training and evaluating medical students," in *2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, 2013, pp. 1–4.

[20] N. Park, Y. Kwon, S. Lee, W. Woo, and J. Jeong, "Projected AR-based interactive cpr simulator," in *Virtual, Augmented and Mixed Reality. Systems and Applications*, R. Shumaker, Ed., Berlin, 2013, pp. 83–89.

[21] E. Higashi, K. Fukagawa, R. Kasimura, Y. Kanamori, A. Minazuki, and H. Hayashi, "Development and evaluation of a corrective feedback system using augmented reality for the high-quality cardiopulmonary resuscitation training," in *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2017.

[22] D. G. Rodrigues, J. G. Johnson, and N. Weibel, "Real-time guidance for cardiopulmonary resuscitation in mixed reality," 2018.

[23] American Heart Association, "My Cardiac Coach," 2017. [Online]. Available: https://play.google.com/store/apps/details?id=cc.aha.cardiaccoach

[24] D. R. Michael and S. L. Chen, *Serious games: Games that educate, train, and inform*. Muska & Lipman/Premier-Trade, 2005.

[25] R. Dörner, S. Göbel, W. Effelsberg, and J. Wiemeyer, *Serious Games: Foundations, Concepts and Practice*. Springer, 2016.

[26] M. Callaghan, M. Savin-Baden, N. McShane, and A. G. Eguluz, "Mapping learning and game mechanics for serious games analysis in engineering education," *IEEE Transactions on Emerging Topics in Computing*, vol. 5, no. 1, pp. 77–83, 2017.

[27] N. Whitton, *Learning with digital games: A practical guide to engaging students in higher education*. Routledge, 2009.

[28] M. Prensky, *Digital game-based learning*. McGraw Hill, 2001.

[29] C. M. Barrio, M. Muoz-Organero, and J. S. Soriano, "Can gamification improve the benefits of student response systems in learning? an experimental study," *IEEE Transactions on Emerging Topics in Computing*, vol. 4, no. 3, pp. 429–438, 2016.

[30] A. L. Blewer, M. Leary, E. C. Esposito, M. Gonzalez, B. Riegel, B. J. Bobrow, and B. S. Abella, "Continuous chest compression cardiopulmonary resuscitation training promotes rescuer self-confidence and increased secondary training," *Critical Care Medicine*, vol. 40, no. 3, pp. 787–792, 2012.

[31] S. Kelle, R. Klemke, and M. Specht, "Effects of game design patterns on basic life support training content," *Educational Technology & Society*, vol. 1, no. 16, pp. 275–285, 2013.

[32] R. Garris, R. Ahlers, and J. E. Driskell, "Games, motivation, and learning: A research and practice model," *Simulation & Gaming*, vol. 33, no. 4, pp. 441–467, 2002.

[33] J. M. Keller, "Development and use of the arcs model of instructional design," *Journal of Instr. Dev.*, vol. 10, no. 3, pp. 2–10, 1987.

[34] J. Brooke, "SUS – A quick and dirty usability scale," *Usability Evaluation in Industry*, vol. 189, pp. 4–7, 1996, iSBN: 9780748404605.

[35] J. Nielsen, *Usability Engineering*. Academic Press, 1993.

[36] *ISO 9241-400 Ergonomics of human-system interaction – Part 400: Principles and requirements for physical input devices*, 2007.

[37] R. S. Kalawsky, "Vruse – A computerised diagnostic tool for usability evaluation of virtual/synthetic environment systems," *Applied Ergonomics*, vol. 30, no. 1, pp. 11–25, 1999.

[38] J. Brooke, "Sus: A retrospective," *Journal of Usability Studies*, vol. 8, no. 2, pp. 29–40, 2013.

Francesco Strada received the MSc degree in Digital Media from Politecnico di Torino. He is currently a PhD student at the Department of Control and Computer Engineering of the same university. His research focuses on VR, AR, HCI, and serious games.

Andrea Bottino is currently an associate professor at Department of Control and Computer Engineering of Politecnico di Torino and head of the CGV research group of the same university. His current research interests include computer vision, machine learning, HCI, computer graphics, VR, AR, and virtual heritage. He is a member of the IEEE.

Fabrizio Lamberti is currently an associate professor at the Department of Control and Computer Engineering of Politecnico di Torino. He has published more than 150 papers in books, journals, and conference proceedings in the areas of computer graphics, HCI, intelligent systems, and educational computing. He is serving as Associate Editor of IEEE TC, IEEE TETC, IEEE TLT and IEEE TCE. He is a senior member of the IEEE and IEEE CS.

Giulia Mormando is an Emergency Medicine Doctor. She is currently a PhD student at Università di Padova. Her research focuses on healtcare simulation.

Pier Luigi Ingrassia started in Italian NHS hospitals as resident in Anaesthesiology and Intensive Care and developed great interest in Emergency Medicine and Disaster response. In his current role at Università del Piemonte Orientale, he is leading the Simulation Centre in Health Science and doing research in the field of simulation-based learning and teaching.