

Evaluating the autonomy of children with autism spectrum disorder in washing hands: a deep-learning approach

Daniele Berardini

Department of Information Engineering
Università Politecnica delle Marche
Ancona, Italy
dani.berard.db@gmail.com

Lucia Migliorelli

Department of Information Engineering
Università Politecnica delle Marche
Ancona, Italy
l.migliorelli@pm.univpm.it

Sara Moccia

Department of Information Engineering
Università Politecnica delle Marche
Ancona, Italy
s.moccia@univpm.it

Marcello Naldini

Project and order office manager
Cooperativa Sociale IL FARO
Macerata, Italy
marcello@ilfarosociale.it

Gioia De Angelis

Board Certified Assistant Behavior Analyst
Cooperativa Sociale IL FARO
Macerata, Italy
servizioaba@ilfarosociale.it

Emanuele Frontoni

Department of Information Engineering
Università Politecnica delle Marche
Ancona, Italy
e.frontoni@univpm.it

Abstract—Monitoring children with Autism Spectrum Disorder (ASD) during the execution of the Applied Behaviour Analysis (ABA) program is crucial to assess the progresses while performing actions. Despite its importance, this monitoring procedure still relies on ABA operators' visual observation and manual annotation of the significant events. In this work a deep learning (DL) based approach has been proposed to evaluate the autonomy of children with ASD while performing the hand-washing task. The goal of the algorithm is the automatic detection of RGB frames in which the ASD child washes his/her hands autonomously (*no-aid* frames) or is supported by the operator (*aid* frames). The proposed approach relies on a pre-trained VGG16 convolutional network (CNN) modified to fulfill the binary classification task. The performance of the fine-tuned VGG16 was compared against that of other CNN architectures. The fine-tuned VGG16 achieved the best performance with a recall of 0.92 and 0.89 for the *no-aid* and *aid* class, respectively. These results prompt the possibility of translating the presented methodology into the actual monitoring practice. The integration of the presented tool with other computer-aided monitoring systems into a single framework, will provide fully support to ABA operators during the therapy session.

Index Terms—Autism Spectrum Disorder, Applied Behaviour Analysis, Deep Learning, Fine-Tuning

I. INTRODUCTION

Autism Spectrum Disorder (ASD) is a severe permanent neurodevelopmental disorder with long-term and pervasive effects that affects 1 in 59 children, worldwide¹. Impaired communication skills, inability to socially interact, absence of emotional reciprocity, limited interests and repetitive behaviors are among the major adverse implications of ASD.

The treatment of children suffering from ASD mostly relies upon the Applied Behavioral Analysis (ABA) technique. The

¹<https://www.aap.org/en-us/Pages/Default.aspx>



Fig. 1. The acquisition set-up (yellow box), placed in the bathroom to record the sink, consisted of an Astra Mini S-Orbbec® RGB-D camera and a minipc Intel® NUC core i5. The blue boxes show the camera field of view.

ABA aims at modifying the child behaviour to make it functional to the tasks of everyday life (e.g., nutrition, personal hygiene, dressing, ...) and improving child ability to relate with others. The application of this technique at an early age allows to act effectively on the child's behavioral processes, leading to significant results [1]. The ABA therapy has generally three main phases: the first one is merely the observation of child behaviour and reaction to external stimuli. Subsequently, the ABA therapist analyses the behavioural reactions of the child.

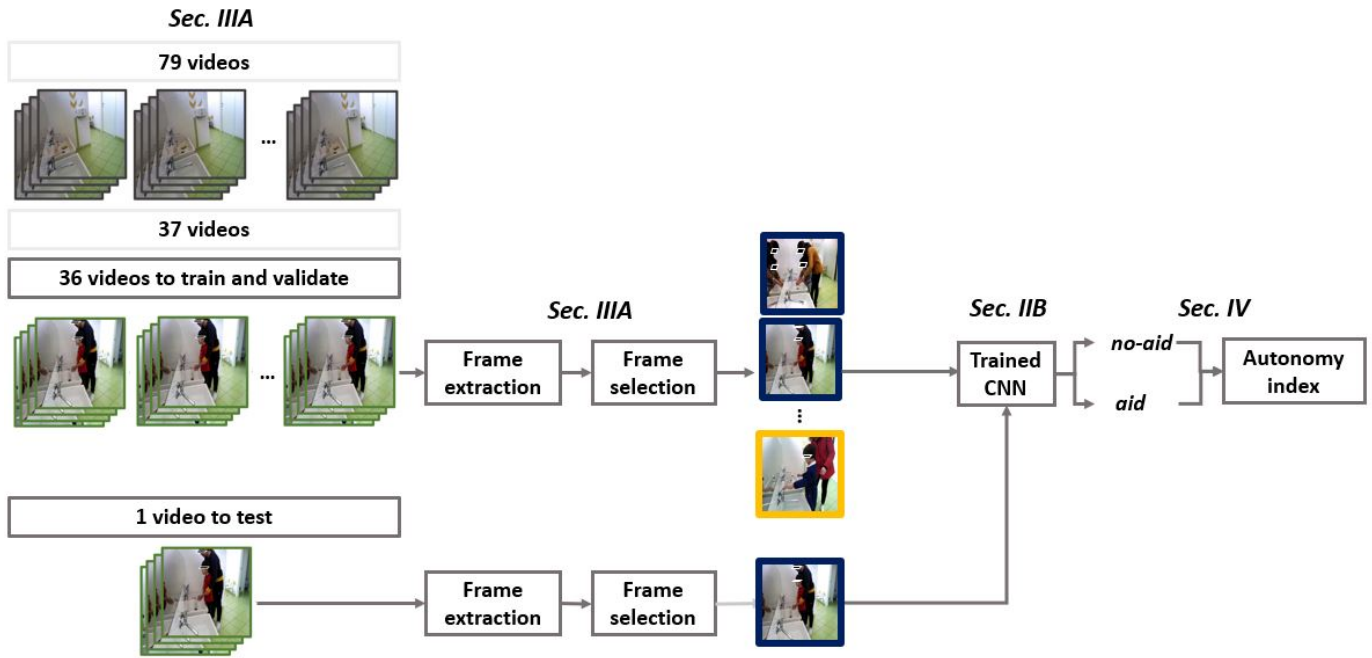


Fig. 2. Workflow of the proposed deep learning (DL)-based application to classify RGB frames in which the child washes his/her hands with the help of the operator (*aid*) or autonomously (*no-aid*). In grey the discarded video sequences (the ones without the subjects of interest), in green the video sequences with the ABA operators and the child during hand-washing. In the frame extraction step the video sequences have been divided in frames which were then selected and annotated in *no-aid* and *aid* class. The frames in which the child washes his/her hands autonomously (*no-aid* frames) are depicted in yellow. The frames in blue box are those in which the operator helps the child in washing his/her hands (*aid* frames).

Finally, the ABA operator draws up a program of specific and personalized exercises to modify the dysfunctional behavior of the child [2].

The child is constantly monitored to check the actual progress and take note, in the form of qualitative scales, of any encountered difficulty, which may require a variation in the ABA program. However, despite its importance, this monitoring procedure still heavily relies on either direct observation or revision of video recordings by the operators, both coupled with paper-and-pencil-rating scales [3]. This monitoring procedure, which includes both the child observation and behaviour evaluation, beside being time-consuming, is qualitative and may be prone to inaccuracies due to operator fatigue.

To attenuate this issue, some promising computer-assisted approaches have been proposed in literature. The majority of them is focused on diagnosing ASD. In [4] the authors propose a deep-learning (DL)-based algorithm that analyses eye movement patterns from video data to discriminate between children with diagnosed ASD and typically developing children. The work in [5] implements a DL framework that analyses video of commonly performed gestures (e.g., grasping a bottle).

Few literature exists on computer-aided systems to support ABA operators in monitoring children with ASD during the therapy session. This may be attributed to the lack of publicly available observational databases. In [6] and [7], stereotypical motor movements of children with ASD, are detected by analysing data from wearable sensors (e.g., accelerometers data) via standard machine learning (ML) and DL meth-

ods, respectively. Unlike DL, ML-based approaches, require a handcrafted-feature extraction step. This procedure, which could be performed either manually or via specific feature extraction algorithms, may be computational expensive, limiting the translation of such application in the actual monitoring practice. More in general, the use of wearable sensors may alter the behavior of the monitored child, especially for the youngest ones.

To offer all the possible support to the operators in monitoring children with autism during the ABA therapy, in this work we propose a DL-based application to analyze images collected from an RGB-D camera. In accordance with our clinical partners, we decided to analyse, among the basic autonomies, that of hand-washing, so we placed the camera over the bathroom sink as showed in Fig 1 and deepened in Sec. II-A. The camera records the ABA operator intent on teaching the child what to do to wash the hands, autonomously. The proposed DL algorithm aims to detect, from an RGB frame, whether the child is washing hands autonomously or with the support from the ABA operator. Then, based on the prediction of the algorithm, an intuitive washing-hand autonomy index is calculated.

This paper is organized as follows: Sec. II describes the DL-approach to classify if the child washes his/her hands with (*aid*) or without (*no-aid*) the support of the operator, Sec. III presents the experimental protocol to validate the presented approach. The experimental results are reported in Sec. IV and discussed in Sec. V, Sec. VI concludes the paper.

TABLE I

DATASET DESCRIPTION: WE USED THE ANNOTATED FRAMES FROM 36 VIDEOS TO TRAIN AND VALIDATE THE ARCHITECTURE (70% OF FRAMES TO TRAIN AND 30% OF FRAMES TO VALIDATE), WHILE THE ANNOTATED FRAMES OF 1 VIDEO WERE USED TO TEST THE ARCHITECTURE.

Training set		Validation set		Test set	
<i>no-aid</i>	<i>aid</i>	<i>no-aid</i>	<i>aid</i>	<i>no-aid</i>	<i>aid</i>
3124	3470	1339	1488	201	118
36 videos				1 video	
9 children				1 child	

II. METHODS

In this section, we first explain the rationale of the case study we examined (Sec. II-A). We describe the network architecture in Sec. II-B and the training strategy implemented to accomplish our classification task in Sec. II-C. Finally we introduce the Hand-Washing autonomy index (II-D). The workflow of the proposed DL approach is showed in Fig. 2

A. Data acquisition protocol: the hand-washing case of study

Personal autonomy skills are certainly one of the elements that mostly affect the quality of life of the child with ASD: being independent from assistance for personal needs might change the future of these children and the way they relate to the environment [8]. In accordance with our clinical partners, we decided to monitor the autonomy of hand-washing which, among the basic autonomies, is fundamental for the safety of the person, for ameliorating social integration and to strengthen the child's self-esteem.

As showed in Fig. 1, to accomplish our goal, we placed an RGB camera on the corner of the bathroom of the ABA centre² to film over the sink. The acquisition set-up, which consisted of an RGB-D camera (Astra Mini S-Orbbec®) and a minipc Intel® NUC core *i5*, was installed to be imperceptible and to not distract the child during the therapy.

A custom-built python script was implemented to automatically acquire concatenated video sequences of 5 minutes each³. After gaining the authorization by the children legal guardians, the acquisition sessions were carried out using a digital programmable timer, for one month, six hours per 5 days (from Monday to Friday).

B. Network architecture

To classify the selected frames in *aid* and *no-aid* we decided to implement VGG16 network as a trade-off between low model complexity and good predictive power.

In the original VGG16 implementation [9], the input 224×224 RGB image is processed through 13 convolutional (conv) layers which act as features extractors. Each conv block has filters with a quite small receptive field (3×3 pixels) and is activated by a rectified linear unit (ReLU) activation function. Every two or three convolutional blocks (depending on the network depth), max pooling layers are used to progressively

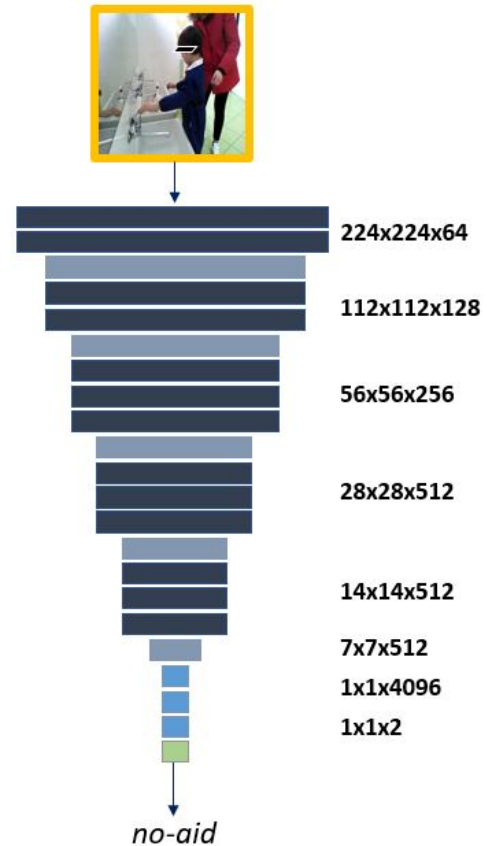


Fig. 3. VGG16 neural network to classify *aid* and *no-aid* frames. In black convolutional (conv) layers, in grey max pooling layers, in light-blue fully connected layers and in green softmax layer.

reduce the spatial size of the feature map. Max pooling aims to lower the amount of training parameters, to reduce the computational complexity and consequently the risk of overfitting.

The network ends with 3 fully-connected layers with 4096, 4096, and 1000 neurons, respectively, separated by dropouts to reduce the effects of overtraining of the neural network. The last fully connected layer is followed by a softmax layer, used to predict the probability of the image to belong to each class of the Imagenet dataset⁴, the natural-image dataset used to train originally VGG16.

To accomplish our binary classification task, we replaced the last fully connected layer with a fully connected layer with 2 neurons (Fig. 3).

C. Training strategy

To train our model we adopted fine-tuning methodology. This procedure allows to migrate the knowledge learned by VGG16 during the training on Imagenet to our binary classification, reducing the risk of overfitting as the features extracted from ImageNet database are very generic [10].

²<https://www.ilfarosociale.it/>

³<https://github.com/roccopietrini/pyOniRecorder>

⁴<http://www.image-net.org/>

In our model, we initialized the weights of the conv blocks and the connections between neurons in the first two fully-connected layers, with the weights of Imagenet, while the last fully-connected layer was initialized with the standard *Glorot* initialization.

D. Hand-Washing autonomy index

The Hand-Washing autonomy index (HWI) (Eq.1) is computed from network predictions as:

$$HWI = \frac{|N|}{|A| + |N|} \quad (1)$$

where $A := \{a \mid \text{child is aided in frame } a\}$ is the set of *aid* frames and $N := \{n \mid \text{child is not aided in frame } n\}$ is the set of *no-aid* frames.

This index is provided to ABA operators to quantify the child’s level of autonomy during the hand-washing task. Evaluating the trend of this index over time would allow the operators to evaluate the progress of the child in performing the task.

III. EXPERIMENTAL ANALYSIS

This section details: the dataset subdivision in training, validation and testing set (Sec. III-A) and the algorithm-related technical specifications (Sec. III-B). In Sec. III-C we present the ablation study and the comparison against other DL architectures. The metrics used to evaluate our model are summarized in Sec. III-D. Finally, in Sec. III-E, we deal with the predictions of “challenging frames”.

A. Dataset specification

The dataset used in this work consisted of 115 RGB video sequences of 5 minutes each. The camera frame rate was 30 frames per second with image size of 640x480 pixels.

Among the 115 video sequences, we first manually selected those where the child and the ABA operator were in the camera field of view at least in one video frame, resulting in a collection of 37 videos. Considering that the hand-washing action is characterized by low dynamics [11], we decided to extract, for each of the 37 videos, 1 frame every 6, obtaining 1080 frames per video.

Table I summarizes the division of the dataset into training, validation and testing sets. Among the frames extracted from 36 videos, 70% were used to train the network and 30% to validate it. The annotated frames of the remaining video were used to test the network.

Starting from the 1080 frames in each video, we manually discarded frames with no one in the camera field of view, and with subjects performing actions different than hand-washing (e.g., dish-washing). Supported by the ABA operators, we assigned each of the remaining frames to the *aid* and *no-aid* class using a custom-built annotator⁵.

10 children and 6 ABA operators took part in the study.

B. Parameter settings

Prior feeding the network, the frames were resized to 224 x 224 pixels, for size compatibility with the pretrained network. To train fine-tuned VGG16 we used the binary cross-entropy loss minimized in 50 epochs with stochastic gradient descent (SGD) as optimizer and a learning rate equal to 0.0005 decayed by a factor of 2 every 10 epochs.

The batch size was set to 64 as a trade-off between memory requirement and training convergence. The best weight configuration among epochs for each model was retrieved according to the highest Accuracy (*Acc*) on the validation set:

$$Acc = \frac{\sum_k^i TP_k}{n}, i \in C : [aid, no-aid] \quad (2)$$

Where: C represents the classes set, TP_k represents the correctly classified sample and n is the total number of samples.

All our analyses were performed using the PyTorch⁶, a Python library, on a 12 GB Nvidia Tesla K80.

C. Ablation study and comparison with other architectures

As an ablation study the performance of the fine-tuned VGG16 was compared against the VGG16 trained from scratch. In the VGG16 trained from scratch we initialized the weights of the conv blocks with *He* initialization while the fully-connected layers were initialized with the standard *Glorot* initialization. We also tested the performance of ResNet50, both fine-tuned with Imagenet pretrained weights and trained from scratch. Considering the structure and the depth of ResNet50 we chose to minimize the binary cross-entropy loss with Adam optimizer setting the learning rate to 0.0001.

For all the architectures, the batch size and the number of epochs, were set to 64 and 50, respectively.

The final model was chosen, among the 4 architectures, as the one with the highest *Acc* in the test set.

D. Performance assessment

To assess the performance of the networks, we computed the classification *Acc* (Eq.2), Precision ($Prec_i$) (Eq. 3), Recall (Rec_i) (Eq. 4), and f1-score ($f1_i$) (Eq. 5) for the i -th class, with $i \in C : [aid, no-aid]$.

$$Prec_i = \frac{TP_i}{TP_i + FP_i} \quad (3)$$

$$Rec_i = \frac{TP_i}{TP_i + FN_i} \quad (4)$$

$$f1_i = \frac{2 \times Prec_i \times Rec_i}{Prec_i + Rec_i} \quad (5)$$

where TP_i , FP_i , FN_i are the correctly classified samples, the false positives and the false negatives for the i -th class, respectively.

⁵<https://github.com/roccopietrini/pyMultipleImgAnnot>

⁶<https://pytorch.org/>

TABLE II

RESULTS OF THE VGG16 AND THE RESNET50, BOTH WITH FINE-TUNING TECHNIQUE AND FROM SCRATCH. RESULTS WERE EVALUATED IN TERMS OF: CLASS-SPECIFIC CLASSIFICATION PRECISION ($Prec_i$), RECALL (Rec_i), F1-SCORE ($f1_i$), FOR $i \in [aid, no-aid]$ AND CLASSIFICATION ACCURACY (Acc).

	$Prec$		Rec		$f1$		Acc
	$no-aid$	aid	$no-aid$	aid	$no-aid$	aid	
ResNet50 trained from scratch	0.64	0.63	0.99	0.04	0.77	0.08	0.64
VGG16 trained from scratch	0.76	0.80	0.93	0.50	0.83	0.62	0.77
fine-tuned ResNet50	0.91	0.76	0.84	0.86	0.87	0.81	0.85
fine-tuned VGG16	0.93	0.86	0.92	0.89	0.92	0.88	0.91



Fig. 4. Example of challenging frames: on the left side the child is aided by the operator (blue box), on the right side the child washes his hands autonomously (yellow box).

E. Performance assessment on challenging frames

To further validate the final classification model, we decided to test its performance on challenging frames selected from the original test set. We refer to challenging frames as those in which the child and the operator were close to each other. Samples of challenging frames are shown in Fig. 4.

To quantitatively evaluate the proximity of the operator and child, we first detected them in each frame. For detection, we used the well-known FASTER-RCNN, which was pre-trained on the large-scale COCO⁷ dataset for natural-image detection tasks. At prediction time, in this work, only the bounding boxes associated with the *person* class were retrieved.

We then computed, as quantitative index of proximity, the Overlap Ratio (OR) among the two bounding boxes (i.e., the child and operator one). The OR was defined as:

$$OR = \frac{area(P \cap K)}{\min(area(P), area(K))} \quad (6)$$

where P and K identified the bounding box of the ABA operator and the child, respectively.

Due to the positioning of our acquisition set-up (next to a mirror), before computing the OR index, the FASTER-RCNN predictions were post-processed to delete the boxes corresponding to the person detected in the mirror. We computed the Euclidean distance of each up-left corner of the predicted bounding box from the origin of the image reference frame. This allowed to select the rightmost bounding boxes excluding those of the mirror.

Then to identify challenging frames we computed the boxplot of the OR for the *aid* and *no-aid* class and we excluded all the frames with OR lower than the minimum of the boxplot

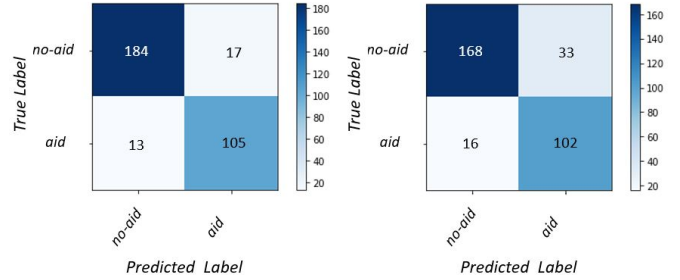


Fig. 5. Confusion matrix for fine-tuned VGG16 (on the left side) and fine-tuned ResNet50 (on the right side): the two best performing models.

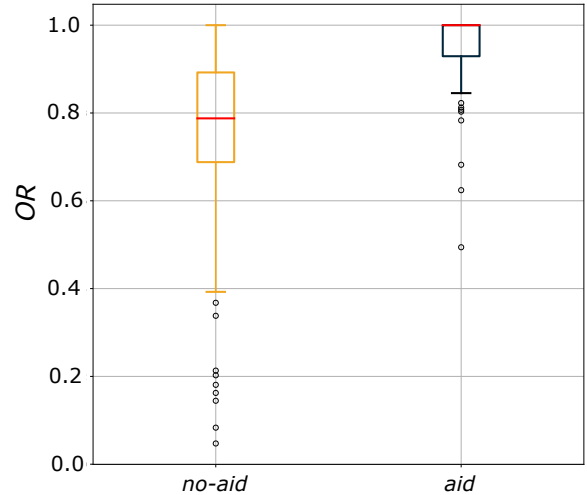


Fig. 6. Boxplot of the Overlap Ratio (OR) for the *no-aid* class in yellow and the *aid* class in blue. Median values of the two distributions are shown in red.

(i.e., the lowest data point excluding any outliers) of the *aid* class.

IV. RESULTS

Table II summarizes the results achieved by VGG16 and ResNet50, both fine-tuned and trained from scratch. The two networks trained from scratch achieved the lowest performance when compared with their counterparts trained with fine-tuning technique, with extremely unbalanced values of per-class metrics. ResNet50 trained from scratch achieved the worst results with imbalanced values of Rec for the *no-aid* and *aid* of 0.99 and 0.04, respectively. VGG16 trained from

⁷<http://cocodataset.org/#home>

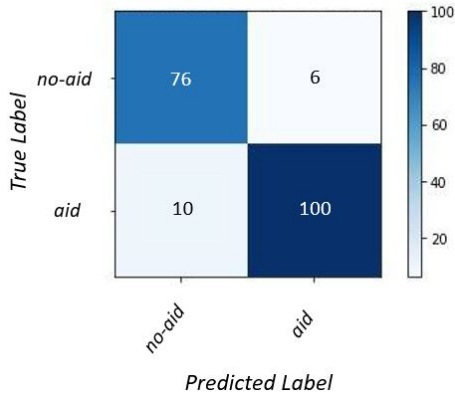


Fig. 7. Confusion matrix of the fine-VGG16 tested on challenging frames

scratch achieved slightly better results with *Rec* of 0.93 and 0.50 for the *no-aid* and *aid* class, respectively. The results highlighted that both the architectures trained from-scratch are more confident in predicting the *no-aid* class with respect to the *aid* one.

The confusion matrices of the two best performing models (i.e., fine-tuned VGG16 and ResNet50) are shown in Fig. 5. Both the models did not outperform in predicting one class with respect to the other. The fine-tuned VGG16 achieved slightly better performance when compared to fine-tuned ResNet50, with higher values of *Prec*, *Rec* and *f1* for both the *no-aid* class (0.93, 0.92 and 0.92, respectively) and the *aid* class (0.86, 0.89 and 0.88, respectively), with an overall *Acc* equal to 0.91. For the VGG16 the predicted *HWI* was equal to 0.62 (actual *HWI*=0.63).

Boxplots of the *OR* for the *aid* and *no-aid* class are depicted in Fig. 6. The minimum of the boxplot for the *aid* class (0.82) was used as threshold to select challenging frames. These 192 frames, resulting from the thresholding, were the ones with *OR* greater than the threshold and have been used to further validate the performance of the fine-tuned VGG16. The confusion matrix of the fine-tuned VGG16 tested on challenging frames is shown in Fig. 7.

V. DISCUSSION

The ABA therapy is an applied science based on experimental behavior analysis with the aim of improving the dysfunctional behaviours of the child with ASD to make them functional to the everyday life tasks. During the therapy session the operators need to constantly observe the child and take paper-and-pencil-rating-scales to evaluate his/her progress and eventual difficulty. To support the ABA therapists during their actual practice, in this work we developed a DL-based application to monitor children with ASD while performing the hand-washing task. By analysing RGB frames, the presented DL model detected whether the child washed hands autonomously (*no-aid* class) or supported by the ABA operator (*aid* class).

We decided to implement VGG16 using fine-tuning technique as trade-off between model complexity and accuracy in

predictions. This model was compared against VGG16 and ResNet50 trained from scratch and fine-tuned ResNet50. Both the performance of the architectures trained from scratch were unsatisfactory as the network poorly predicted the *aid* class. Fine-tuning technique has improved the performance of both the ResNet50 and VGG16 with respect to their corresponding trained from scratch. Hence, fine-tuning allowed to migrate the knowledge of the training on the large-scale ImageNet dataset to our classification task improving the networks generalization ability. However, fine-tuned VGG16 showed better performance in terms of classification *Acc* and per-class metrics. This may be due to the relatively simple and shallow structure (16 layers) of VGG16 coupled with a small-size dataset.

To further validate fine-tuned VGG16, we decided to test its performance on the most challenging frames among the testing set. These frames were quantitatively identified as the ones in which the ABA operator and the child were close to each other, even if the child performed the hand-washing task autonomously. When tested on challenging frames, fine-tuned VGG16 achieved encouraging results in predicting both the *aid* and *no-aid* class (with only 6 out of 76, and 10 out of 100 misclassified frames for the *no-aid* and *aid* class, respectively). This suggested that the network was able to extract information from the frame of interest which did not solely rely on the spatial distance between operator and child, proving to be suitable for the task.

VI. CONCLUSIONS

In this paper, we presented a DL-based application aimed at classifying if the child with ASD washes his/her hands autonomously or is supported by the ABA operator. The achieved results suggested that DL can be successfully exploited in the field of ASD children monitoring and evaluation. However, to fully support the operators in quantifying the progress achieved by the children who underwent ABA therapy, it is acknowledged that further research is required. The proposed methodology can be coupled with a pose-estimation framework [12], and other base autonomies (i.e., the tooth brushing) could be investigated. In addition to the gestural functions, the pitch of the voice could also be analyzed to assess communication skills [13]. Moreover, all the computer-aided solutions, relevant to the monitoring of the child with autism, could be included in a single framework, to fully support the operators during their actual practice [14].

ACKNOWLEDGMENT

This work was supported by “Cooperativa Sociale IL FARO”⁸ and is part of the project “COMEACASA” (POR Marche FESR 2014/2020).

REFERENCES

- [1] L. A. Livingston, E. Colvert, S. R. S. Team, P. Bolton, and F. Happé, “Good social skills despite poor theory of mind: exploring compensation in autism spectrum disorder,” *Journal of Child Psychology and Psychiatry*, vol. 60, no. 1, pp. 102–110, 2019.

⁸<https://www.ilfarosociale.it/>

- [2] G. R. Mayer and B. Sulzer-Azaroff, *Applying behavior-analysis procedures with children and youth*. Holt, Rinehart and Winston, 1977.
- [3] D. Granpeesheh, J. Tarbox, and D. R. Dixon, "Applied behavior analytic interventions for children with autism: a description and review of treatment research," *Annals of Clinical Psychiatry*, vol. 21, no. 3, pp. 162–173, 2009.
- [4] J. Xie, L. Wang, P. Webster, Y. Yao, J. Sun, S. Wang, and H. Zhou, "A two-stream end-to-end deep learning network for recognizing atypical visual attention in autism spectrum disorder," *arXiv preprint arXiv:1911.11393*, 2019.
- [5] A. Zunino, P. Morerio, A. Cavallo, C. Ansuini, J. Podda, F. Battaglia, E. Veneselli, C. Becchio, and V. Murino, "Video gesture analysis for autism spectrum disorder detection," in *2018 24th International Conference on Pattern Recognition (ICPR)*. IEEE, 2018, pp. 3421–3426.
- [6] U. Großekathöfer, N. V. Manyakov, V. Mihajlović, G. Pandina, A. Skalkin, S. Ness, A. Bangerter, and M. S. Goodwin, "Automated detection of stereotypical motor movements in autism spectrum disorder using recurrence quantification analysis," *Frontiers in neuroinformatics*, vol. 11, p. 9, 2017.
- [7] N. M. Rad, A. Bizzego, S. M. Kia, G. Jurman, P. Venuti, and C. Furlanello, "Convolutional neural network for stereotypical motor movement detection in autism," *arXiv preprint arXiv:1511.01865*, 2015.
- [8] C. P. Johnson, S. M. Myers *et al.*, "Identification and evaluation of children with autism spectrum disorders," *Pediatrics*, vol. 120, no. 5, pp. 1183–1215, 2007.
- [9] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [10] Y.-X. Wang, D. Ramanan, and M. Hebert, "Growing a brain: Fine-tuning by increasing model capacity," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2471–2480.
- [11] J. L. Cook, S.-J. Blakemore, and C. Press, "Atypical basic movement kinematics in autism spectrum conditions," *Brain*, vol. 136, no. 9, pp. 2816–2824, 2013.
- [12] S. Moccia, L. Migliorelli, V. Carnielli, and E. Frontoni, "Preterm infants' pose estimation with spatio-temporal features," *IEEE Transactions on Biomedical Engineering*, 2019.
- [13] E. Ambrosini, M. Caielli, M. Milis, C. Loizou, D. Azzolino, S. Damanti, L. Bertagnoli, M. Cesari, S. Moccia, M. Cid *et al.*, "Automatic speech analysis to early detect functional cognitive decline in elderly population," in *2019 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2019, pp. 212–216.
- [14] E. Frontoni, A. Mancini, M. Baldi, M. Paolanti, S. Moccia, P. Zingaretti, V. Landro, and P. Misericordia, "Sharing health data among general practitioners: The nu. sa. project," *International Journal of Medical Informatics*, vol. 129, pp. 267–274, 2019.