International Conference on Artificial Reality and Telexistence
Eurographics Symposium on Virtual Environments (2020)
A. Kulik and M. Sra (Editors)



# An Automated Virtual Receptionist for Recognizing Visitors and Assuring Mask Wearing

S. Zehtabian [1] S. Khodadadeh[1] K. Kim[2] G. Bruder[2] G. Welch[2] L. Bölöni[1] D. Turgut[1]

[1]Department of Computer Science, University of Central Florida, Orlando, FL, USA
[2]Institute for Simulation and Training, University of Central Florida, Orlando, FL, USA

**Abstract**

*Intelligent virtual agents have many societal uses, specifically in situations in which the presence of real humans would be prohibitive. In particular, virtual receptionists can perform a variety of tasks associated with visitor and employee safety, e.g., during the COVID-19 pandemic. In this poster, we present our prototype of a virtual receptionist that employs computer vision and meta-learning techniques to identify and interact with a visitor in a manner similar to that of a real human receptionist. Specifically we employ a meta-learning-based classifier to learn the visitors' faces from the minimal data collected during a first visit, such that the receptionist can recognize the same visitor during follow-up visits. The system also makes use of deep neural network-based computer vision techniques to recognize whether the visitor is wearing a face mask or not.*

**CCS Concepts**

• ***Computing methodologies*** → ***Intelligent agents; Object identification;***

## 1. Introduction

Thanks to recent advances in intelligent virtual agent technologies [NBB*19], agents have shown their uses for a wide range of social situations in which the presence of real humans would be prohibitive. This is in particular true for the current COVID-19 pandemic, which introduced the need for new social norms, such as maintaining social distancing and wearing a mask, which could be ensured in public spaces by using virtual receptionists.

Detecting and identifying a visitor is an example of a "simple" and yet essential skill for a receptionist. A more advanced skill related to the COVID-19 pandemic would be for virtual receptionists to screen visitors to reduce the risk of infection, e.g., by encouraging social distancing and mask-wearing. Such virtual receptionists, in particular when realized via commercial-off-the-shelf (COTS) projectors/screens and cameras, are attractive even for comparatively small companies/organizations that would not usually deploy a human receptionist. Virtual receptionists require integrating several software components, including face recognition, natural language processing, rendering, and a dialogue system.

In terms of functionality, it is more important than ever for virtual receptionists to be able to recognize repeat visitors; for instance, recognizing whether a person entering the space had a positive (or negative) COVID-19 test the week before can inform how the receptionist routes that person. Finally, the protocols and requirements are changing rapidly, which require receptionists to adapt quickly to such changes. These desired adaptations can only be accomplished with learning algorithms that can be quickly tuned with minimal data. In this work, we describe the design and implementation of an embodied virtual receptionist aiming to satisfy the aforementioned requirements during the COVID-19 pandemic.

## 2. The Embodied Virtual Receptionist Prototype

**Embodied Virtual Character and Natural Interface.** We designed a humanoid virtual character using Adobe Fuse Character Creator and animated it with Mixamo. We performed a pilot test with available COTS hardware and found an appropriate choice in the use of a projected virtual character that allows us to create a life-sized receptionist on any wall in a public space. The setup is shown in Figure 1. The code for the central controller of our system is available at: https://github.com/sharare90/ARVR-LabAgent-Controller.

**Face Recognition and Re-identification.** We compared the performance of two state-of-the-art publicly available libraries: Adam Geitgey's Face Recognition library (AGFRL) built on top of dlib [Kin09] and FaceNet [SKP15]. These libraries use large neural networks with more than 20 million parameters. To achieve the same performance with smaller networks, we suggest applying meta-learning that can adapt to new situations with a few examples.

The most important performance metrics of our system are the speed and accuracy of the face recognition. For the speed metric, we report the system response for the compared algorithms, averaged over five runs. For the accuracy metric, we consider the accuracy of the system when encountering new users the system has not seen before. For fair comparisons, we applied pre-trained weights on VGGFACE2 [CSX*18] whenever

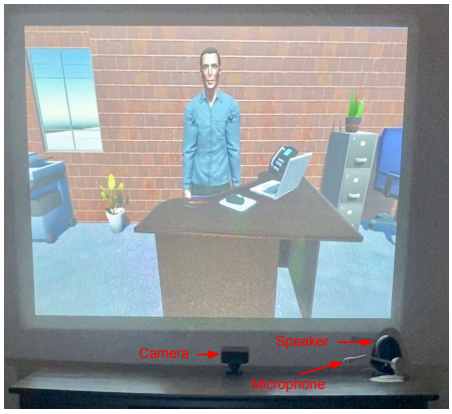


**Figure 1:** *The embodied virtual receptionist prototype.*


**Figure 2:** *Results of the mask wearing classifier.*

**Table 1:** *The average accuracy (with a 95% confidence interval), the average evaluation time per image on CelebA dataset, and the number of parameters for different methods.*

| Algorithm | # of Parameters | Time (seconds) | Accuracy (%) |
|---|---|---|---|
| MAML | **4.5 M** | $\mathbf{6.34 \times 10^{-6}}$ | $90.55 \pm 0.50$ |
| ProtoNets | **4.5 M** | $\mathbf{8.25 \times 10^{-6}}$ | $\mathbf{95.17 \pm 0.38}$ |
| FaceNet | 23.5 M | $1.28 \times 10^{-4}$ | $\mathbf{96.28 \pm 0.30}$ |
| AGFRL | 21.8 M | 1.25 | $90.05 \pm 0.59$ |

we used an off-the-shelf method, and also trained all meta-learning approaches on VGGFACE2. For tests, we leveraged the CelebA [LLWT15] dataset. Our experimental protocol follows meta-learning benchmarks' evaluation. We reported 95% confidence interval for the accuracy over 1000 different classification tasks (5000 faces). We show that we can use much smaller networks for face recognition by leveraging meta-learning techniques such as MAML [FAL17] and ProtoNets [SSZ17]. We build on the publicly available code by Khodadadeh et al. [KBS19]. See Table 1 for the results. Our implementation is available at https://github.com/siavash-khodadadeh/MetaLearning-TF2.0/tree/master/models/face-recognition.

**Face Mask Classification.** Determining whether the incoming users are wearing a mask or not is an important component of our prototype. We achieve this by training a classifier using transfer learning, a technique that uses knowledge gained from solving a different but related problem. We report the accuracy of our algorithm on a publicly available dataset for face mask classification (MaskML: https://makeml.app/datasets/mask). The dataset contains 853 images of 4072 faces belonging to two classes: with mask, without mask or mask worn incorrectly. We split this dataset into train, validation, and test with 80%, 10%, and 10% ratio, respectively. The code is available at: https://github.com/sharare90/Face-Mask-Classifiaction . The accuracy on train, validation, and test data are 95.47%, 89.93%, and 89.95%, respectively. See Figure 2 for a set of images from with and without mask classes with their accuracy. Note that this is not in conflict with face re-identification. Here, we only classify whether the user is wearing a mask or not. In face re-identification, we aim to re-identify the same person in subsequent frames regardless of whether they are wearing a mask or not. Recognizing the same person with and without a mask is not the purpose of this work and introduces new challenges that we leave for future research.

## 3. Conclusions

We designed a virtual receptionist with basic and advanced skills related to recognizing visitors and determining whether or not they comply with COVID-19 related mask-wearing rules. We implemented functionalities for face recognition and face mask detection. We also trained two meta-learning algorithms on a large, diverse face dataset and showed that it is possible to use a much smaller network that can adapt to unseen situations with 15 to 20 times faster than the current state of the art approaches.

**Acknowledgement** This material includes work supported in part by the National Science Foundation under Collaborative Award # 1800961, 1800947, and 1800922 (Dr. Ephraim P. Glinert, IIS) to the University of Central Florida, University of Florida, and Stanford University respectively; the Office of Naval Research under Award # N00014-17-1-2927 (Dr. Peter Squire, Code 34); and the AdventHealth Endowed Chair in Healthcare Simulation (Prof. Welch). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the supporting institutions.

## References

[CSX*18] CAO Q., SHEN L., XIE W., PARKHI O. M., ZISSERMAN A.: Vggface2: A dataset for recognising faces across pose and age. In *Proc. of FC* (2018), pp. 67–74. 1

[FAL17] FINN C., ABBEEL P., LEVINE S.: Model-agnostic meta-learning for fast adaptation of deep networks. In *Proc. of ICML* (2017), pp. 1126–1135. 2

[KBS19] KHODADADEH S., BÖLÖNI L., SHAH M.: Unsupervised meta-learning for few-shot image classification. In *Proc. of NeurIPS* (2019), pp. 10132–10142. 2

[Kin09] KING D. E.: Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research (JMLR)* (2009), 1755–1758. 1

[LLWT15] LIU Z., LUO P., WANG X., TANG X.: Deep learning face attributes in the wild. In *Proc. of ICCV* (2015). 2

[NBB*19] NOROUZI N., BRUDER G., BELNA B., MUTTER S., TURGUT D., WELCH G.: A Systematic Review of the Convergence of Augmented Reality, Intelligent Virtual Agents, and the Internet of Things. In *Artificial Intelligence in IoT* (2019). 1

[SKP15] SCHROFF F., KALENICHENKO D., PHILBIN J.: FaceNet: A unified embedding for face recognition and clustering. In *Proc. of CVPR* (2015), pp. 815–823. 1

[SSZ17] SNELL J., SWERSKY K., ZEMEL R.: Prototypical networks for few-shot learning. In *Proc. of NeurIPS* (2017), pp. 4077–4087. 2