Enhanced Security Framework for Enabling Facial Recognition in Autonomous Shuttles Public Transportation During COVID-19

Dimitris Tsiktsiris, Antonios Lalas, Minas Dasygenis, Konstantinos Votis, Dimitrios Tzovaras

Enhanced Security Framework for Enabling Facial Recognition in Autonomous Shuttles Public Transportation during COVID-19 *

Dimitris Tsiktsiris[1,2] [0000−0001−6475−5865], Antonios Lalas[1,2,*] [0000−0002−5337−161X], Minas Dasygenis[2] [0000−0002−2180−9752], Konstantinos Votis[1] [0000−0001−6381−8326], and Dimitrios Tzovaras[1] [0000−0001−6915−6722]

[1] Information Technologies Institute, Center for Research and Technology Hellas, 6th km Xarilaou - Thermi, Thessaloniki, 57001, Greece
{tsiktsiris,lalas,kvotis,dimitrios.tzovaras}@iti.gr
[2] Department of Electrical and Computer Engineering, University of Western Macedonia, Kozani, 50100 Greece
mdasyg@ieee.org

Abstract. Autonomous Vehicles (AVs) can potentially reduce the accident risk while a human is driving. They can also improve the public transportation by connecting city centers with main mass transit systems. The development of technologies that can provide a sense of security to the passenger when the driver is missing remains a challenging task. Moreover, such technologies are forced to adopt to the new reality formed by the COVID-19 pandemic, as it has created significant restrictions to passenger mobility through public transportation. In this work, an image-based approach, supported by novel AI algorithms, is proposed as a service to increase autonomy of non-fully autonomous people such as kids, grandparents and disabled people. The proposed real-time service, can identify family members via facial characteristics and efficiently ignore face masks, while providing notifications for their condition to their supervisor relatives. The envisioned AI-supported security framework, apart from enhancing the trust to autonomous mobility, could be advantageous in other applications also related to domestic security and defense.

Keywords: Autonomous Vehicles · Public Transportation · Neural Networks · Image Processing · Security

1 Introduction

The concept of autonomous mobility as a service (MaaS) [5] is progressively adopted by public transportation systems. However, transitioning to fully autonomous public vehicles in the real world is not a seamless process and has

* Supported by the European Union's Horizon 2020 Research and Innovation Programme Autonomous Vehicles to Evolve to a New Urban Experience (AVENUE) under Grant Agreement No 769033.

several obstacles that derive mainly from the safety concerns of the passengers [2] [14]. These perceptions of traffic safety and in-vehicle security have a significant impact on the acceptance of the overall concept of autonomous public transportation. The prospective passengers fear several possible instances that could arise in case there is no driver or staff in the shuttle. More indicatively:

- Passengers feeling discomfort traveling alone during night-time.
- Parents not being able to know if their kids have reached their destination safely.
- Caregivers not being able to track passengers with dementia or other health issues.

To address the aforementioned concerns on social and personal safety and security into the vehicle, certain measures need to be implemented. For example, third parties monitoring the route of minors or passengers with health issues could make their route much easier and less frightening. This may be followed by appropriate notifications and/or instructions to the third party, while the vehicle may also implement respective actions. We propose a service to increase autonomy of non-fully autonomous people such as kids, grandparents and disabled individuals. This service will both ensure carers or family members that their beloved ones are safe while commuting around the city and increase confidence to the non-fully autonomous people to use public transport knowing that their family can "be with them". To achieve such functionality, we rely on one-shot (or single-shot) facial recognition techniques capable of identifying or verifying a person from a video frame [4] [17]. In general, facial recognition, uses computing techniques for the identification of human faces in an image or a video, and then proceeds to measure specific facial characteristics. This information is later combined to create a facial signature, or profile. On the other hand, when used for facial verification, a frame from the camera footage is compared to the recorded facial signature. However, as these profiles are based on mathematical models as a result of the relative positions of their facial features, anything that can lead to reduced visibility of key characteristics, such as the nose, mouth and chin, intervenes with facial recognition [6]. Since the beginning of COVID-19 pandemic, people began to increasingly wear masks as they prevent people from getting and spreading the virus [15] and it seems obvious that algorithms designed to analyze faces and facial features will be less accurate if part of the face is concealed.

In this paper, we present an image-based approach, supported by novel AI algorithms, as an end-to-end service to increase the COVID-19 safety rules adherence of the passengers inside an autonomous shuttle. The proposed real-time service can identify family members via facial characteristics and effectively ignore face masks. The main contributions of this work can be summarized as follows:

- We propose a service to increase confidence of non-fully autonomous people such as kids, grandparents and disabled individuals in order to use autonomous public transportation.

– We present an end-to-end service based on deep learning, for automated facial recognition in autonomous shuttles.
– We introduce techniques based on attention to mitigate the occlusion issues introduced by face masks during the COVID-19 pandemic.

2 Related Work

There are multiple methods in which facial recognition systems work, but in general, they work by comparing selected facial features from a given image with faces within a database. It is also described as a Biometric Artificial Intelligence based application that can uniquely identify a person by analyzing patterns based on the person's facial textures and shape. Many face recognition techniques require multiple data of the subject in the training dataset, in order to correctly identify the face of a person. From our perspective this is not possible as we rely in one single input image of the subject. In order to be able to overcome this problem, we are using one shot (or single shot) facial recognition algorithms.

Over the last years, Deep Learning algorithms have achieved great success in the field of facial recognition, as deep features are quite common in occlusion conditions, showing better performance over shallow features. In [10], the authors proposed a Dynamic Feature Matching approach, combining a Fully Convolutional Network (FCN) with a Sparse Representation-based Classification (SRC) to recognize face parts of random size. Finally, after the last pooling layer, the deep features were linearly presented using gallery feature maps. In another approach of [16], the authors established a mask dictionary by computing the difference between the top convolution features with or without occlusion. After that, the damaged features were discarded by querying the mask dictionary. However, this method cannot be easily adopted, as it requires paired images. On the other hand, Duan et. al. [8] proposed an end-to-end BoostGAN model for profile facial recognition with occlusion. In this approach, firstly, a non-occluded image was developed by the occluded image and, then, it was used for refined facial recognition. Nonetheless, this is also a hard method to reproduce, especially in cases of large area occlusions, like face masks.

Based on the reviewed literature, we have concluded that, even though the aforementioned approaches show the latest progress of Deep Learning technology in facial recognition with occlusion, most of the work done is not suitable for mask facial recognition in real life, when the key discriminating features of the nose, mouth and chin are completely damaged.

3 Methodology

A high-level overview of the service is depicted in Figure 1. The first layer of sensors connects to the Hardware Abstraction Layer (HAL). The HAL implements the IP and the USB protocol supporting IP and USB cameras respectively but also can request raw data by the API endpoints in order to perform face recognition. The input data is converted and transformed in a compatible format and

passed into the analytics algorithms. The result is then transferred via the API endpoints into the cloud. The user has access to the data and acts accordingly. A new passenger can be enrolled to the service using a single image of his face which will be stored in a database. Using this as the reference image, the network will calculate the similarity for any new instances presented to it.

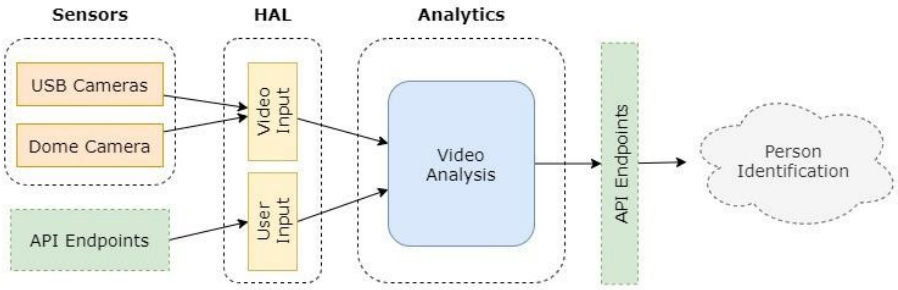

Fig. 1. High-level overview of the proposed service

As for the video analysis, facial recognition techniques identify human faces in images or videos by measuring specific facial characteristics. The extracted information is later combined to create a facial signature, or profile. On the other hand, when used for facial verification, a frame from the camera footage is compared to the recorded profile. In our architecture, a Multi-task Cascaded Convolutional Network (MTCNN) [18] receives the input frame in order to extract and align facial images. The facial images are then pre-processed and passed into a feature extractor (CNN backbone) linked with the explainable cosine (xCos) module [13] that features an explainable cosine metric.

3.1 Data collection and preprocessing

We used the MS1M-ArcFace [9] dataset for training our network and the LFW [11] for the testing. The two datasets were augmented via pre-processing in order to include face masks using the MaskTheFace tool [3]. The data were artificially generated due to the limited available datasets that include face masks and are suitable for the face verification task. For the sake of simplicity, we name the synthetic datasets MS1M-ArcFace+M and LFW+M, respectively.

3.2 Network pipeline

As current face verification models use fully connected layers, spatial information is lost along with the ability to understand the convolution features in a human sense. To address this obstacle, the plug-in xCos module is integrated as described below.

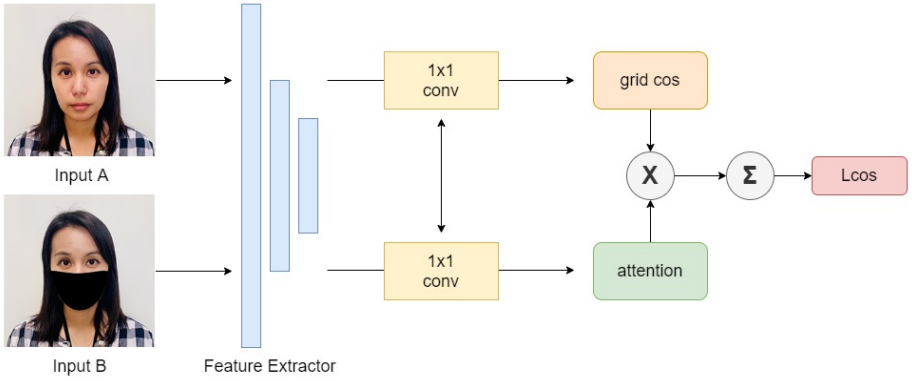

Fig. 2. Network Pipeline: The two input images are preprocessed and passed into the backbone CNN for feature extraction along with the plug-in xCos module.

Input The two input images are preprocessed and passed into the feature extractor. The Input A is the database image while the Input B is the image cropped from the video stream.

Backbone We implement the same CNN feature extractor as in ArcFace [7]. However, in order to employ the xCos module, the last fully connected layer and the previous flatten layer are replaced with an 1x1 convolutional layer.

Lcos calculation (xCos) Patch-wise cosine similarity is multiplied by the attention maps and then summed to calculate the Lcos.

3.3 Results

Table 1 shows the testing accuracy on the original MS1M-ArcFace, LFW and the artificially created MS1M-ArcFace+M and LFW+M, which contain additional masked samples. As we can see, despite the minimal loss caused by the explainability module on the original datasets, a significant improvement of about 6.2% has been accomplished in our augmented datasets which contain face masks.

Table 1. Accuracy comparison between methods and different datasets indicate a significant improvement of about 6.2% in our augmented datasets which contain face masks.

Method	**Training Dataset**	**Testing Dataset**	**Masks**	**Accuracy**
ArcFace	MS1M-ArcFace	LFW	No	99.83%
ArcFace-xCos	MS1M-ArcFace	LFW	No	99.35%
ArcFace+M	MS1M-ArcFace+M	LFW+M	Yes	68.33%
ours, ArcFace-xCos+M	MS1M-ArcFace+M	LFW+M	Yes	74.52%

Figure 3 also highlights the improvements made over the original ArcFace+M. The maps are generated (a) by the original ArcFace+M while (b) by our improved model. The cosine similarity between the two faces is negative both in (a) and (b) on the bottom half of the faces as the mask portrays different characteristics such as shape and color. However, the attention map in our improved (b) model indicated that the network focuses more on the upper half characteristics around the eyes and the nose.

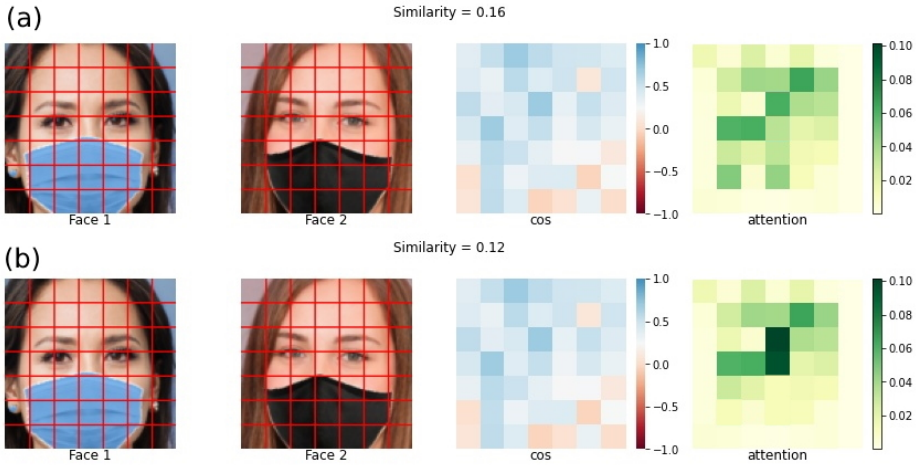

Fig. 3. Input and maps generated by (a) the original ArcFace+M and (b) our improved model. Although, the cosine (cos) similarity is negative both in (a) and (b) on the bottom half of the faces, the attention map in our improved (b) model indicated that the network focuses more on the upper half characteristics around the eyes and the nose, ignoring the region covered by the mask.

Fig. 4. An experimental real-world demonstration running on the NVIDIA Jetson AGX Xavier. The proposed system is able to correctly identify the two passengers of the autonomous shuttle, among a database of 20 people.

Figure 4 showcases an experimental real-world scenario on a autonomous shuttle, running on the NVIDIA Jetson AGX Xavier [1]. The proposed system is able to correctly identify the two passengers among a database of 20 people.

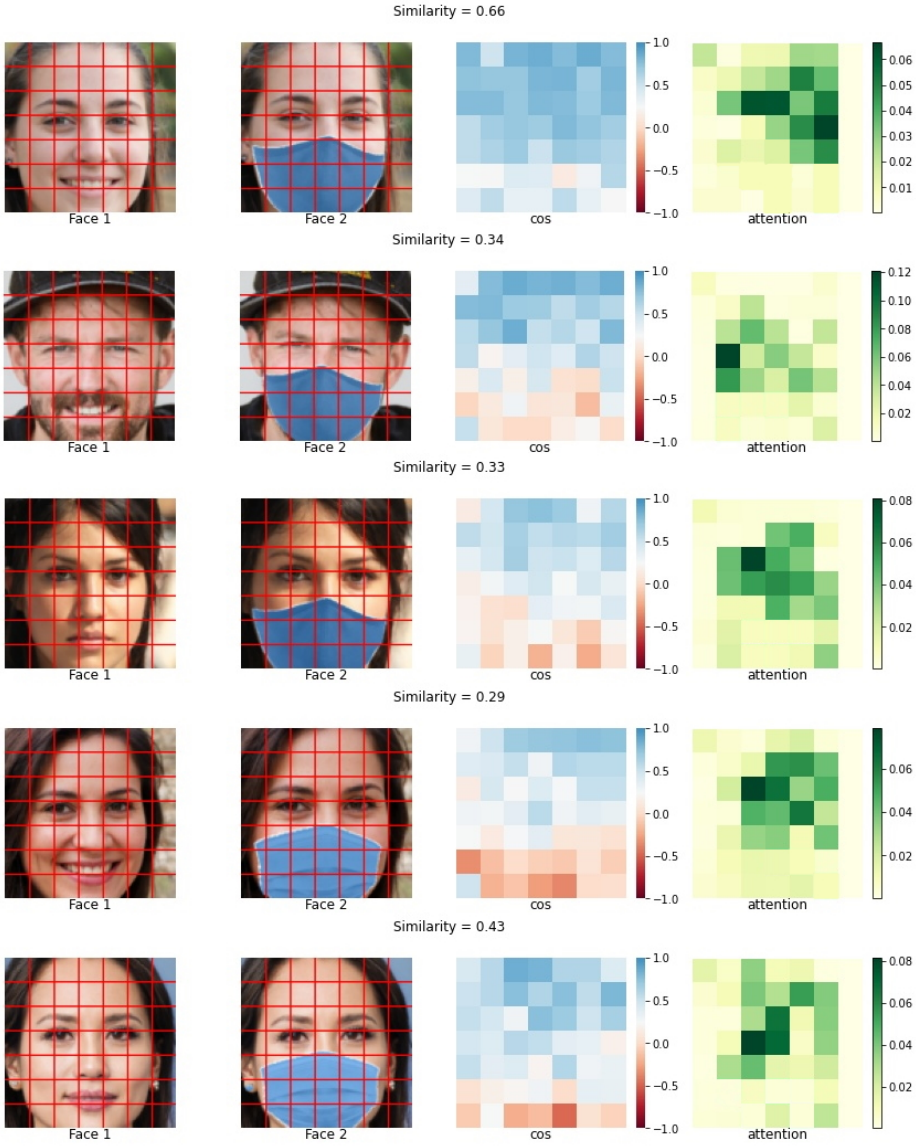

Fig. 5. Additional results of our improved model. Facial images are artificially generated using StyleGAN [12] and post-processed to include masks using MaskTheFace [3]

Figure 5 shows additional results of our improved model. The attention maps verify that the bottom parts of the face are efficiently ignored while the upper face characteristics have the most significant impact on calculating the similarity score.

The system was designed as a flexible and end-to-end service accessible by an Android mobile application. In Figure 6, the architecture of the proposed system is depicted. A new passenger can be enrolled to the service via the mobile application, using a single image of his face. The image will be processed and stored temporarily in a database on the cloud platform. By using this as the reference image, the network calculates the similarity for any new instance presented to it from the shuttle. The network can predict a similarity score in one shot and inform the client via an API call and relevant notification through the mobile phone. The proposed approach can be further extended to support homeland security surveillance infrastructures in order to mitigate domestic security risks. In this context, it is envisioned to establish a highly adaptable security framework capable of leveraging the capabilities of autonomous public transport operators as well as law enforcement agencies.

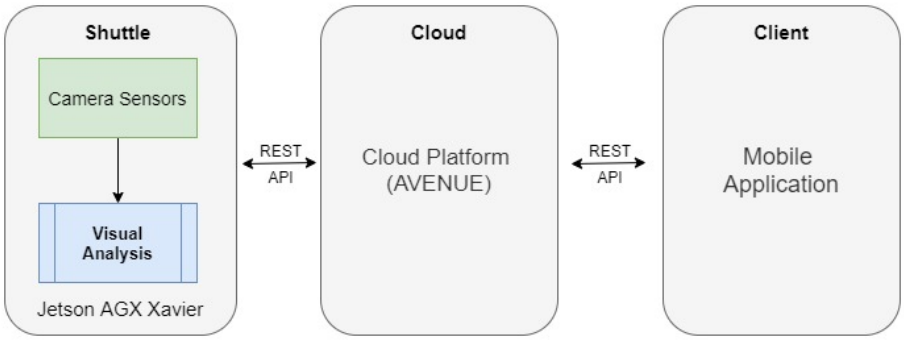

Fig. 6. The architecture of the proposed system. A new passenger is enrolled by the client. A single image of his face will be processed and stored temporarily in a database and the network will calculate similarities with current passengers. The client is informed via an API call and relevant notification through the mobile phone.

The solution was build upon several power consumption constraints as we target an autonomous platform. The system operates on the edge and draws power from the vehicle batteries; therefore, it needs to be efficient and conserve energy. Several optimizations were applied in order to reduce the energy footprint on the vehicle's battery life. We optimized the network using TensorRT but also offloaded some convolutional layers to the more energy efficient Deep Learning Accelerator (DLA) engines provided by the Nvidia Jetson. Moreover, the integration with other vehicle services that are able to report the passengers

on board helped us to selectively perform inference and further reduce the power consumption to a total average of 6.43 Watts.

4 Conclusions

In this paper, we present a deep learning approach deployed as a real-time service in order to increase autonomy of non-fully autonomous people such as kids, grandparents and disabled individuals. The service can identify the passengers of an autonomous vehicle via facial characteristics to provide a sense of safety to the passenger and his family relatives via appropriate notifications. Legal issues related to GDPR and privacy concerns could be solved using dedicated consent forms of the passengers. To address the occlusion issues introduced by face masks during COVID-19, we propose a model based on ArcFace-xCos trained and evaluated on artificially generated datasets that include face masks. Experimental results indicated that the network yields better accuracy versus the original ArcFace due to the attention on the upper face characteristics such as eyes, eyebrows and nose. The system was deployed as a real-time service on the Jetson AGX Xavier. Future work will support real-time notifications to inform the relatives about the density of the passengers and the number of the people near their relative. We believe that our research will be a stepping stone towards increased AI-based safety and security in the autonomous public transport and other application domains as well, i.e. homeland security and defense. It will also significantly contribute in advancing the trust of the passengers to a human driverless transport system during the pandemic and beyond.

References

1. NVIDIA Jetson AGX Xavier. https://developer.nvidia.com/embedded/jetson-agx-xavier-developer-kit, accessed: 2021-03-28
2. Anania, E.C., Rice, S., Walters, N.W., Pierce, M., Winter, S.R., Milner, M.N.: The effects of positive and negative information on consumers' willingness to ride in a driverless vehicle. Transport policy **72**, 218–224 (2018)
3. Anwar, A., Raychowdhury, A.: Masked face recognition for secure authentication (2020)
4. Azeem, A., Sharif, M., Raza, M., Murtaza, M.: A survey: Face recognition techniques under partial occlusion. Int. Arab J. Inf. Technol. **11**(1), 1–10 (2014)
5. Cruz, C.O., Sarmento, J.M.: "mobility as a service" platforms: A critical path towards increasing the sustainability of transportation systems. Sustainability **12**(16), 6368 (2020)
6. Damer, N., Grebe, J.H., Chen, C., Boutros, F., Kirchbuchner, F., Kuijper, A.: The effect of wearing a mask on face recognition performance: an exploratory study. In: 2020 International Conference of the Biometrics Special Interest Group (BIOSIG). pp. 1–6. IEEE (2020)
7. Deng, J., Guo, J., Xue, N., Zafeiriou, S.: Arcface: Additive angular margin loss for deep face recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4690–4699 (2019)

8. Duan, Q., Zhang, L.: Look more into occlusion: realistic face frontalization and recognition with boostgan. IEEE transactions on neural networks and learning systems (2020)
9. Guo, Y., Zhang, L., Hu, Y., He, X., Gao, J.: Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In: European conference on computer vision. pp. 87–102. Springer (2016)
10. He, L., Li, H., Zhang, Q., Sun, Z.: Dynamic feature learning for partial face recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7054–7063 (2018)
11. Huang, G.B., Mattar, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. In: Workshop on faces in'Real-Life'Images: detection, alignment, and recognition (2008)
12. Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., Aila, T.: Analyzing and improving the image quality of StyleGAN. In: Proc. CVPR (2020)
13. Lin, Y.S., Liu, Z.Y., Chen, Y.A., Wang, Y.S., Lee, H.Y., Chen, Y.R., Chang, Y.L., Hsu, W.H.: xcos: An explainable cosine metric for face verification task. arXiv preprint arXiv:2003.05383 (2020)
14. Ngan, M.L., Grother, P.J., Hanaoka, K.K.: Ongoing face recognition vendor test (frvt) part 6a: Face recognition accuracy with masks using pre-covid-19 algorithms (2020)
15. Organization, W.H., et al.: Advice on the use of masks in the context of covid-19: interim guidance, 6 april 2020. Tech. rep., World Health Organization (2020)
16. Song, L., Gong, D., Li, Z., Liu, C., Liu, W.: Occlusion robust face recognition based on mask learning with pairwise differential siamese network. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 773–782 (2019)
17. Wang, L., Li, Y., Wang, S.: Feature learning for one-shot face recognition. In: 2018 25th IEEE International Conference on Image Processing (ICIP). pp. 2386–2390. IEEE (2018)
18. Zhang, K., Zhang, Z., Li, Z., Qiao, Y.: Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Processing Letters **23**(10), 1499–1503 (2016)