

ARE GAN-BASED MORPHS THREATENING FACE RECOGNITION?

Eklavya Sarkar^{1,2}, Pavel Korshunov¹, Laurent Colbois^{1,3}, and Sébastien Marcel^{1,3}

¹Idiap Research Institute, Martigny, Switzerland

²École polytechnique fédérale de Lausanne, Switzerland

³University of Lausanne, Switzerland

{eklavya.sarkar, pavel.korshunov, laurent.colbois, sebastien.marcel}@idiap.ch

ABSTRACT

Morphing attacks are a threat to biometric systems where the biometric reference in an identity document can be altered. This form of attack presents an important issue in applications relying on identity documents such as border security or access control. Research in generation of face morphs and their detection is developing rapidly, however very few datasets with morphing attacks and open-source detection toolkits are publicly available. This paper bridges this gap by providing two datasets and the corresponding code for four types of morphing attacks: two that rely on facial landmarks based on OpenCV and FaceMorpher, and two that use StyleGAN 2 to generate synthetic morphs. We also conduct extensive experiments to assess the vulnerability of four state-of-the-art face recognition systems, including FaceNet, VGG-Face, ArcFace, and ISV. Surprisingly, the experiments demonstrate that, although visually more appealing, morphs based on StyleGAN 2 do not pose a significant threat to the state of face recognition systems, as these morphs were outmatched by the simple morphs that are based on facial landmarks.

Index Terms— Biometrics, Face Recognition, Vulnerability Analysis, Morphing Attack, StyleGAN 2

1. INTRODUCTION

After Ferrara *et al.* [1] showed that by using a morphed photo of two different people an adversary can circumvent passport registration process, morphing attacks and how to detect them received a lot of attention from academic, industrial, and security communities. The vulnerability of state-of-the-art (SOTA) face recognition systems (FR) and the threat such vulnerability poses to the security systems relying on recognition technologies led to the explosion of research work in this area.

Most of the work related to morphing attacks (MAs) focuses on their detection. Recently proposed techniques for morphing attack detection (MAD) include methods based on *so called* classical approaches using local binary patterns

(LBP) and support vector machines (SVM) [2], approaches rooted in image forensics that rely on photo response non uniformity (PRNU) function [3], deep neural networks specifically trained to detect morph images [4], and FR systems themselves serving as feature extractors for an support vector machine (SVM) classifier [5]. The National Institute of Standards and Technology (NIST) is now conducting independent evaluations of MAD technologies [6].

However the research in the area of morphing attacks and their detection suffers from a lack of datasets, evaluation protocols, and clear understanding of whether the latest face recognition systems are vulnerable to both ‘classical’ and the latest generative adversarial network (GAN)-based morphing attacks. So called ‘classical’ landmark-based morphing techniques are widely available, but the modern ones are rarely publicly released. Novel methods are often either proprietary, such as Combined Morphs [7], or are difficult to replicate from a published description without knowing the minutes technical details. Pre-generated databases of morphing attacks are therefore essential for biometrics research, yet only a few, like the Face Morph Image dataset [7] by Advanced Multimedia Security Lab, are publicly available.

Therefore, this paper provides the following contributions:

1. We provide an open source morphing tool¹ for generation of morphing attacks based on OpenCV [8], FaceMorpher [9], StyleGAN 2 [10], and our modified implementation of MIPGAN-II [11].
2. We provide publicly available datasets¹ with morphed images generated using the aforementioned techniques, including the latest GAN-based morphs, on the publicly available FERET [12] and Face Research Lab London (FRL) [13] datasets.
3. We conduct an extensive vulnerability assessment of the different morphing attacks images in our generated dataset against SOTA face recognition systems, used in

Work funded by the Swiss Center for Biometrics Research and Testing.

¹<https://gitlab.idiap.ch/bob/bob.paper.icassp2022.morph.generate>

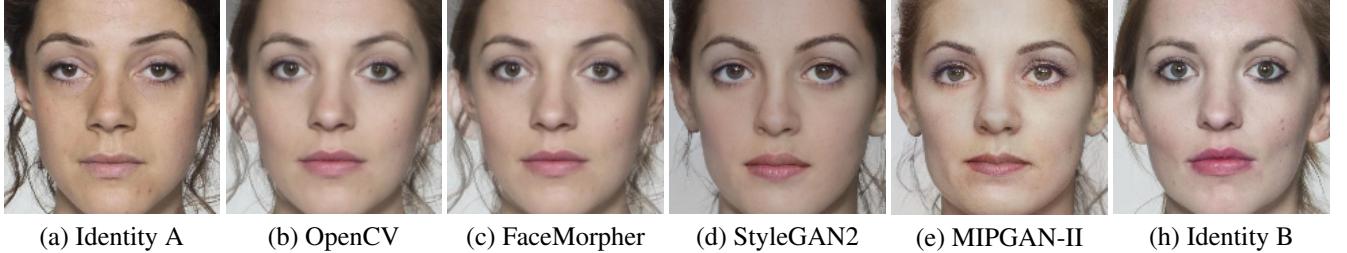


Fig. 1. Different types of generated morphed images from two identities in the FRLL dataset.

some of the latest morphing attacks vulnerability studies [14]. We specifically use the FaceNet [15] VGG-Face [16], ArcFace [17], and ISV [18] systems, which are pre-trained on ‘clean’ bona fide databases.

We also highlight that the majority of existing work on Morphing Attacks (MAs) only consider the ‘typical’ scenario where the morphs are used to attack the enrollment process of face recognition. This paper, to the best of our knowledge is amount the the first to evaluate two scenarios: i) when morphs attack the the enrollment process, and ii) when they are used to attack the probing process, which is similar to a presentation attack [19].

2. MORPH GENERATION

In this section, we present the datasets with bona fide faces and the different tools, including GAN-based, that we use to generate the morphing images for the vulnerability experiments.

2.1. Datasets

We used FERET [12] and FRLL [13] datasets of facial images to generate the morphs. FERET was selected because it is the *de facto* standard datasets commonly used in papers on morphing attack detection [14, 20] and it has large number of images of different identities. The FRLL dataset is also ideal for creating morphing attacks because it contains close-up frontal face images of very high visual quality and 1350×1350 resolution, shot under *uniform* illumination with large varieties in ethnicity, pose, and expression. Each face is annotated using 189 facial landmarks, which is notably a very high number, as typical landmarks detectors provide no more than 68-70 landmarks. The main limitation of FRLL dataset, compared to FERET, is the limited number 102 of different identities with 53 males and 49 females.

For each dataset, we select bona fide (or original) face pairs for morph generation by following the existing protocols used in previous work. For FERET, we follow the protocols used in the work by Scherhag *et al.* [14] that were kindly provided by the authors. For FRLL dataset, we follow the

protocols used in AMSL Face Morph Image dataset by Neuert *et al.* [7]. Using these protocols (essentially, which facial image pairs to morph), we generated morphs using four different methods: based on OpenCV, based on FaceMorpher, based on StyleGAN 2, and a modified MIPGAN-II [11].

2.2. Morphing Tools

As representatives of the ‘classical’ landmark-based morphing tools, we provide two commonly used open source face morphing algorithms. First one is the **OpenCV**-based algorithm, referred throughout the paper as *OpenCV*, which is an adaptation of an open-source implementation [8] for morphing faces using 68-point annotator from Dlib library [21]. Face landmarks are obtained for each of the two bona fide source images and are used to form Delaunay triangles, which are in-turn warped and alpha blended.

FaceMorpher [9], referred to as *FaceMorpher*, is another open-source landmark-based morphing algorithm, but with the STASM [22] landmark detector instead. Both algorithms create morphs with noticeable ghosting artefacts for all three datasets, as the region outside the area covered by these landmarks is simply averaged.

Following the advances in generative adversarial networks (GANs), there were attempts to generate morphed images using a GAN instead of landmark-based methods [23, 24]. In this paper, we adapted the latest **StyleGAN 2** [10], referred throughout the paper as *StyleGAN2*, to develop a morphing algorithm which can generate high resolution realistic looking faces with no noticeable artifacts. The StyleGAN 2 was pre-trained on the FFHQ dataset introduced in [25].

The faces are cropped to obtain the same landmark alignment as in the FFHQ dataset. The images are then projected into the \mathcal{W} space of StyleGAN 2 by optimizing the input latent style vector that is fed to the generator network, such that it minimizes the perceptual loss between the generated and real image [10]. Once an associated latent vector has been computed for each of the source images, morphs can be generated by linearly interpolating between two latent vectors, and feeding the interpolated vector back into the generator.

This technique yields very realistic looking morphs without visual artefacts, however, since StyleGAN does not have

any information about the identities in bona fide images, there is no guarantee that the resulted morph is actually a blend of these identities (see the example in Figure 1(d) for an idea).

For this purpose, we implemented a modified version of the recent MIPGAN-II technique [11], referred throughout the paper as *MIPGAN-II*, which improves on the StyleGAN 2 morphs, by further optimizing the interpolated latent vector with four additional weighted losses, which help to preserve the identity information and structural correspondence of the two bona-fide images. The main difference in our versions of MIPGAN is that we use the pre-trained VGGFace model with the ResNet50 backbone as feature extractor in the identity loss, instead of a pre-trained embedding extractor with ResNet50 as backbone using the ArcFace loss.

Both, StyleGAN2 and MIPGAN-II GAN-based morphs require the projected images to be at a high resolution (1024×1024 after cropping), and work better with an uniform background, which makes the FRLL dataset particularly appropriate. A side note observation of using GAN-based techniques for generating morphs is that it is equally easy to generate high-quality morphs for smiling expressions as it is for the neutral faces, which is not possible with typical landmark-based tools.

Using four (two ‘classical’ and two GAN-based) morphing tools, we generate 529 morphs per each tool for FERET datasets using the same protocols as in [14], and 1222 morphs per tool images for FRLL, following the morph generation protocol defined in ASML Face Morph Image dataset. It is to be noted that the protocols insure not to morph across genders and ethnicities, and only interpolate images if neither or only one of the two subjects is wearing glasses.

3. EVALUATION PROTOCOLS

3.1. Face Recognition Systems

To evaluate vulnerability of face recognition against morphing attacks, we used publicly available pre-trained FaceNet [15], ArcFace [17], and VGG-Face [16] architectures. We used the last fully connected layers of these networks as features and the cosine distance as a classifier. For a given test face, the confidence score of whether it belongs to a reference model is the cosine distance between the average reference feature vector and the feature vector of a test face. These systems are the state of the art recognition systems with Facenet showing 99.63% [15], ArcFace – 99.53 [17], and VGG-Face – 98.95% [16] accuracies on the labeled faces in the wild (LFW) dataset.

We also used an inter-session variability (ISV) based face recognition [18], pre-trained on the MOBIO [26] dataset, as a ‘classical’ baseline. The DCT features computed on overlapping blocks of 40×40 were used for the ISV-based system of 512 Gaussian mixture models (GMMs) and 160 dimensional subspace.

3.2. Evaluation Metrics

In a verification process, the user attempting to authenticate presents a biometric probe and a claimed identity, and can be classified into one of the following 3 categories. *A) Genuine user* (BF): probe and claimed identity both correctly belong to the user. *B) Zero-effort impostor* (BF): probe belongs to the user, but the claimed identity corresponds to a different enrolled user. *C) Morph attack impostor* (MA): probe matches the claimed identity but does not correspond to the user.

The verification performance is typically evaluated with the following metrics.

- *False Match Rate (FMR)* [20]: proportion of zero-effort impostors that are falsely authenticated.
- *False Non-Match Rate (FNMR)* [20]: proportion of genuine users which are falsely rejected.
- *Mated Morph Presentation Match Rate (MMPMR)* [27]: proportion of morphs attacks impostors accepted by the face recognition system.

3.3. Evaluation scenarios

In general, there are two main scenarios under which a face recognition system is evaluated: a bona fide (BF) scenario where both the reference and probes images as genuine, so there are no attacks and the system is assumed to perform under the conditions it was designed for; and the morphing attack (MA) scenario when morphs are introduced to the face recognition with a malicious intent to spoof the recognition. There are also two variants of MA scenario, when a morphed image can be either used as a reference, i.e., FR system is hijacked during enrollment process (typical morphing attack scenario), or a morphed image is used as a probe, which is similar to presentation attack scenario [19].

The number of reference and probe images for each evaluation scenario is summarized in Table 1.

Table 1. Number of images in different evaluation scenarios.

Dataset	Morphs as	BF	MA	Impostors
FERET	References	529	791	418,439
	Probes	791	529	418,439
FRLL	References	91	584	1,984
	Probes	584	91	4,153

It is also to be noted that we did not split datasets into training, development, and test subsets but used each whole dataset as one single test set, as all used FR systems were pre-trained on other databases. Furthermore, we choose the decision threshold to compute MMPMR value for MA scenario based on FMR value computed in the bona fide scenario, thus removing the need for a development set.

4. EXPERIMENTAL RESULTS

Table 2 summarizes the results of the vulnerability assessment of the several face recognition systems (described in section 3.1) under the different morphing attack scenarios (as explained in section 3.3). The MMPMR metric is calculated by setting the decision threshold at FMR=0.1% in the bona fide scenario.

Table 2. MMPMR @ FMR = 0.1%
(Morphs as references — Morphs as probes) [%]

Tools	FRS	FRL	FERET
OpenCV	FaceNet	83.3 — 72.0	41.1 — 40.6
	Arcface	59.8 — 73.8	34.6 — 35.2
	VGG	39.7 — 48.6	22.0 — 21.0
	ISV	59.8 — 97.8	44.8 — 58.4
FaceMorpher	FaceNet	64.5 — 68.2	39.9 — 40.3
	Arcface	57.6 — 75.3	34.1 — 34.8
	VGG	23.4 — 47.1	20.5 — 18.3
	ISV	56.1 — 96.1	42.6 — 56.5
StyleGAN2	FaceNet	5.9 — 11.0	1.6 — 1.3
	Arcface	9.8 — 18.3	2.4 — 2.5
	VGG	3.0 — 9.1	2.0 — 1.5
	ISV	9.2 — 43.6	2.7 — 3.4
MIPGAN-II	FaceNet	47.2 — 62.7	32.9 — 32.3
	Arcface	32.0 — 46.5	26.0 — 25.1
	VGG	15.9 — 30.4	14.5 — 13.2
	ISV	3.6 — 23.7	7.3 — 9.6

The results in Table 2 reveal a number of interesting observations. The StyleGAN2-morphs do not pose a significant threat to the state of the art face recognition systems, compared to landmark-based morphs, despite being of higher visual quality, and with very few ghosting artefacts. This likely occurs because the original pixels of both contributing images are still present in the features after landmark-based morph-generation pipeline is applied, and are later picked up during face recognition, thus successfully fooling the FR systems. Conversely, the StyleGAN pipeline conserves no pixel traces of the original contributing subjects, other than the positions of the facial landmarks, as it generates the morphed image by interpolating the projected vectors in the \mathcal{W} latent space. The interpolated vector fed back through the synthesis network does not contain the features of both identities, and instead is perceived as a new, different identity altogether.

This is further proved when the MIPGAN-II morphs which purposefully use four additional losses to further optimize the generated morph in an attempt to conserve the identities of the two source subjects: the vulnerability is significantly higher than with naive linear-interpolation in the StyleGAN \mathcal{W} space. However, our implementation of MIPGAN-II lead to twice as low MMPMR rates compared

to the numbers reported in original MIPGAN-II paper [11]. Such significant disparity in results is puzzling but a few elements could contribute to it:

A) It appears that rather than using the \mathcal{W} space of StyleGAN 2 for generating the morphs, as we did in our implementation, the authors of MIPGAN-II paper [11] instead used the $\mathcal{W}+$ [28] latent space of StyleGAN, as they seem to describe their latent projections as a concatenation of 18 different 512-dimensional w vectors, one for each layer of the StyleGAN architecture. We believe that using \mathcal{W} space is more reasonable as operating in $\mathcal{W}+$ does not guarantee a visual realism of resulted morphs. However, this point is hard to verify, since the authors of MIPGAN-II [11] did not release their code and their paper is not very explicit about this aspect.

B) The pre-trained StyleGAN model in [11] was fine-tuned on their test dataset (FRGCv2 [29]), which made the generated morphs to appear visually and structurally very similar to the original images in the dataset. This type of ‘trick’ clearly would increase the chance of the morphs to be more threatening to face recognition, which was tested on the same dataset.

Table 2 also demonstrates that the more accurate face recognition system is the more vulnerable it is to morphing attacks, which is also in line with the findings reported for presentation attacks [30]. This trend is especially evident when we compare a more accurate and deeper FaceNet architecture with VGG for all databases and types of morphs.

We can also notice that the results for the scenario when morphs attack the enrollment process (see ‘morphs as references’ sub-columns in Table 2) are very similar to the results for the scenario when morphs are used as probes (see ‘morphs as probes’ sub-columns) for the FERET morphs database. However, in the case of FRL, the face recognition systems are clearly more vulnerable to the scenario when morphs are used as probes. It means that the quality of original images used to create morphs may lead to more threatening morphs in the presentation attack scenarios, rather than when attacking FR systems from the inside.

5. CONCLUSION

In this paper, we assess the level of vulnerability of existing face recognition systems, based on VGG-Face, ArcFace, and FaceNet neural network models, against four morphing attacks, including two ‘classical’ morphs based on facial landmarks and two based on StyleGAN 2. The results demonstrate that ‘classical’ morphs still are of the highest threat to the face recognition while GAN-based morphs, despite their higher visual appeal, do not pose as much of a threat to automated system. We also note that the face recognition systems that are better at recognition are also more vulnerable to morphing attacks. We publicly release the databases we generated and used, and provide all tools for generating morphs and running the evaluation experiments as an open source package.

6. REFERENCES

- [1] M. Ferrara, A. Franco, and D. Maltoni, “The magic passport,” in *IEEE International Joint Conference on Biometrics*, 2014, pp. 1–7.
- [2] L. Spreeuwiers, M. Schils, and R. Veldhuis, “Towards robust evaluation of face morphing detection,” in *2018 26th European Signal Processing Conference (EUSIPCO)*, 2018, pp. 1027–1031.
- [3] U. Scherhag, L. Debiasi, C. Rathgeb, C. Busch, and A. Uhl, “Detection of face morphing attacks based on prnu analysis,” *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 1, no. 4, pp. 302–317, 2019.
- [4] Clemens Seibold, Wojciech Samek, Anna Hilsmann, and Peter Eisert, “Accurate and robust neural networks for face morphing attack detection,” *Journal of Information Security and Applications*, vol. 53, pp. 102526, 2020.
- [5] L. Wandzik, G. Kaeding, and R. V. Garcia, “Morphing detection using a general- purpose face recognition system,” in *2018 26th European Signal Processing Conference (EUSIPCO)*, 2018, pp. 1012–1016.
- [6] Mei L. Ngan, Patrick J. Grother, Kayee K. Hanaoka, and Jason M. Kuo, “Face recognition vendor test (frvt) part 4: Morph - performance of automated face morph detection,” Tech. Rep., National Institute of Standards and Technology, 2020.
- [7] T. Neubert, A. Makrushin, M. Hildebrandt, C. Kraetzer, and J. Dittmann, “Extended stirtrace benchmarking of biometric and forensic qualities of morphed face images,” *IET Biometrics*, vol. 7, no. 4, pp. 325–332, 2018.
- [8] Satya Mallick, “Face morph using opencv — c++ / python,” March 2016.
- [9] Alyssa Quek, “Facemorpher,” January 2019.
- [10] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, “Analyzing and improving the image quality of StyleGAN,” in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 8107–8116.
- [11] Haoyu Zhang, Sushma Venkatesh, Raghavendra Ramachandra, Kiran Raja, Naser Damer, and Christoph Busch, “Mipgan—generating strong and high quality morphing attacks using identity prior driven gan,” *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 3, no. 3, pp. 365–383, 2021.
- [12] P.Jonathon Phillips, Harry Wechsler, Jeffery Huang, and Patrick J. Rauss, “The feret database and evaluation procedure for face-recognition algorithms,” *Image and Vision Computing*, vol. 16, no. 5, pp. 295 – 306, 1998.
- [13] Lisa DeBruine and Benedict Jones, “Face Research Lab London Set,” 5 2017.
- [14] U. Scherhag, C. Rathgeb, J. Merkle, and C. Busch, “Deep face representations for differential morphing attack detection,” *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 3625–3639, 2020.
- [15] F. Schroff, D. Kalenichenko, and J. Philbin, “Facenet: A unified embedding for face recognition and clustering,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 815–823.
- [16] Omkar M. Parkhi, Andrea Vedaldi, and Andrew Zisserman, “Deep face recognition,” in *British Machine Vision Conference*, 2015.
- [17] Jiankang Deng, Jia Guo, Xue Niannan, and Stefanos Zafeiriou, “Arcface: Additive angular margin loss for deep face recognition,” in *CVPR*, 2019.
- [18] R. Wallace, M. McLaren, C. McCool, and S. Marcel, “Inter-session variability modelling and joint factor analysis for face authentication,” in *2011 International Joint Conference on Biometrics (IJCB)*, 2011, pp. 1–8.
- [19] ISO/IEC FDIS 30107-3:2017, “Information technology — Biometric presentation attack detection — Part 3: Testing and reporting,” Standard, International Organization for Standardization, Geneva, Switzerland, 09 2017.
- [20] Raghavendra Ramachandra and Christoph Busch, “Presentation attack detection methods for face recognition systems: A comprehensive survey,” *ACM Comput. Surv.*, vol. 50, no. 1, Mar. 2017.
- [21] Davis E. King, “Dlib-ml: A machine learning toolkit,” *Journal of Machine Learning Research*, vol. 10, pp. 1755–1758, 2009.
- [22] S. Milborrow and F. Nicolls, “Active Shape Models with SIFT Descriptors and MARS,” *VISAPP*, 2014.
- [23] S. Venkatesh, H. Zhang, R. Ramachandra, K. Raja, N. Damer, and C. Busch, “Can gan generated morphs threaten face recognition systems equally as landmark based morphs? - vulnerability and detection,” in *2020 8th International Workshop on Biometrics and Forensics (IWBFF)*, 2020, pp. 1–6.
- [24] Pavel Korshunov and Sébastien Marcel, “Vulnerability of face recognition to deep morphing,” in *International Conference on Biometrics for Borders*, oct 2019, pp. 1–5.
- [25] T. Karras, S. Laine, and T. Aila, “A style-based generator architecture for generative adversarial networks,” in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 4396–4405.
- [26] Chris McCool, Sébastien Marcel, Abdenour Hadid, Matti Pietikäinen, Pavel Matějka, Jan Černocký, Norman Poh, Josef Kittler, Anthony Larcher, Christophe Lévy, Driss Matrouf, Jean-François Bonastre, Phil Tresadern, and Timothy Cootes, “Bi-modal person recognition on a mobile phone: using mobile phone data,” in *IEEE ICME Workshop on Hot Topic in Mobile Multimedia*, 2012.
- [27] U. Scherhag, A. Nautsch, C. Rathgeb, M. Gomez-Barrero, R. N. J. Veldhuis, L. Spreeuwiers, M. Schils, D. Maltoni, P. Grother, S. Marcel, R. Breithaupt, R. Ramachandra, and C. Busch, “Biometric systems under morphing attacks: Assessment of morphing techniques and vulnerability reporting,” in *2017 International Conference of the Biometrics Special Interest Group (BIOSIG)*, 2017, pp. 1–7.
- [28] Rameen Abdal, Yipeng Qin, and Peter Wonka, “Image2stylegan: How to embed images into the stylegan latent space?,” in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 4431–4440.
- [29] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, Jin Chang, K. Hoffman, J. Marques, Jaesik Min, and W. Worek, “Overview of the face recognition grand challenge,” in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)*, 2005, vol. 1, pp. 947–954 vol. 1.
- [30] A. Mohammadi, S. Bhattacharjee, and S. Marcel, “Deeply vulnerable: a study of the robustness of face recognition to presentation attacks,” *IET Biometrics*, vol. 7, no. 1, pp. 15–26, 2018.