Object features and face detection performance: Analyses with 3D-rendered synthetic data

Jian Han
University of Amsterdam
Amsterdam, The Netherlands
Email: j.han@uva.nl

Sezer Karaoglu
3DUniversum
Amsterdam, The Netherlands
Email: s.karaoglu@3duniversum.com

Hoang-An Le and Theo Gevers
University of Amsterdam
Amsterdam, The Netherlands
Email: {h.a.le, th.gevers}@uva.nl

***Abstract*—This paper is to provide an overview of how object features from images influence face detection performance, and how to select synthetic faces to address specific features. To this end, we investigate the effects of occlusion, scale, viewpoint, background, and noise by using a novel synthetic image generator based on 3DU Face Dataset. To examine the effects of different features, we selected three detectors (Faster RCNN, HR, SSH) as representative of various face detection methodologies. Comparing different configurations of synthetic data on face detection systems, it showed that our synthetic dataset could complement face detectors to become more robust against features in the real world. Our analysis also demonstrated that a variety of data augmentation is necessary to address nuanced differences in performance.**

I. Introduction

Face detection is one of the most studied topics in the field of computer vision. It plays a fundamental role in basically all face related applications. Face detection requires first to determine whether there is a face in an image or a video and then to return the precise location of the face. A number of effective face detection systems have been rapidly emerged in recent years. It is impractical to evaluate on every recently proposed detection system, we are fortunate that most of the leading approaches shared a common methodology [1]. Deep network based object detection system can be roughly classified as two categories: 1) Two step face detector is the most representative detector based on deep network. Region proposal is the fundamental step for all these kind of methods. The first stage normally proposes candidate bounding boxes. The second stage, features are extracted from each candidate box for the following classification and bounding-box regression tasks. 2) One step face detector doesn't need a region proposal unit before classification. Its light-weighted structure is more time efficient but less accurate [2].

Face detection still confronts challenges from features in scale, head pose, expression, facial occlusion and illumination. Existing dataset for face detection like FDDB [3], MAFA [4] and Wider Face [5], the majority of data normally belong to limited range of variations. The faces did not sufficiently represent extreme poses, scale or heavy occlusion, to train a robust detector against all potential variations. Previous researchers designed different face detectors to address specific types of features in real-world situations. The rapid development of deep learning essentially relies on the availability of large-scale annotated datasets. Collecting and annotating real-world datasets with different attributes is unpractical. It is also difficult to fully control the imaging variations in such datasets, or to avoid errors during annotation process. A bias from ground truth may lead to far-reaching impact in deep network.

Data augmentation deals with aforementioned issue by artificially inflating the training set with label preserving transformations [6, 7]. A variety of data augmentation methods have shown effectiveness in face related tasks [8, 9]. In this paper, we aim to address the issue by using synthetic data, as complementary to real data, to create fully controlled conditions with automatic and error-less annotation. We develop a synthetic data generator based on 3D face models. The 2D face synthesis process contains varied viewpoint, scale, illumination, occlusion and background. We manipulated all these features or attributes in 3D scenes, to make the rendered images more realistic than direct manipulation on 2D. With the help of synthetic data, we are able to systematically investigate the effects of different features. Then the face detectors are trained on the combination of real data and synthetic dataset to address the features. Based on our experiments, we also identified some potential deficiencies of the current face detection systems. Our paper can also be an example to analyze other detectors.

Our contributions are:

- We provide a 2D face synthetic data generator with manipulated features (on pose, scale, background, illumination, and occlusion), which enables specified examinations of face detector performances.
- We conduct detailed analyses between feature and performance, which can be a guide to compare performances of other face detectors.
- Our analyses also reveal some weaknesses of the current face detectors and suggest using synthetic data for future improvement on robustness.

II. Related work

A. Face detection

Face detection can be considered as a special case of object detection. Two thorough survey related to object detection can be found in [10] and [1]. Most face detectors are designed

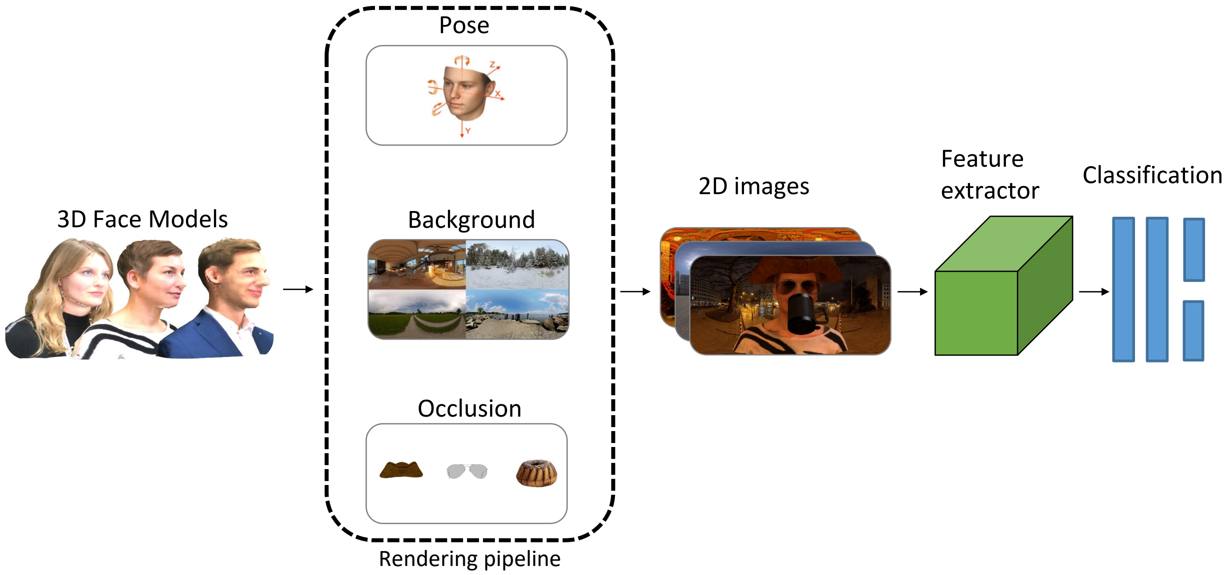

Fig. 1. An overview of render pipeline. We manipulate pose, background, occlusion, and illumination on original 3D models, and then render 3D models into 2D images.

to address specific characteristics in real-world scenarios, in terms of, for example, scale [11–15], occlusion [4, 16], pose [17] or lighting condition [18]. A certain face detector may be only suitable for datasets with corresponding characteristics and is almost impossible to be robust against features of all datasets.

We will briefly discuss several typical detectors and the features they mainly deal with. To handle large variations of scale, Hybrid Resolutions (HR) face detector is designed to detect faces with extreme scale by using contextual information and an image pyramid [19]; Single Stage Headless (SSH) detector [20] is a fast one-step detector based on scale-invariant design. To detect occluded faces, a local linear embedding method is used to reduce noise and recover the lost cues from occlusion in [4]; Face Attention Network [16] uses a special attention network with reduced background information and data augmentation to address face occlusion.

B. *Face specific data augmentation*

Basic geometric and photometric data augmentation methods, like flipping, rotation, resizing, cropping, color jittering, have been widely used in deep learning based face applications. Detailed survey about face specific data augmentation can be found in [21, 22]. Previous research converges to support the effectiveness of synthetic data in improving the performance in face related applications [23–25]. Masi et al. [26] introduce face appearance variations with pose, shape and expression for effective face recognition. Lv et al. [8] propose multiple data augmentation for face recognition, including synthetic variations for hairstyle, glasses, poses and illumination.

Then, the problem shifts to the generation of synthetic face images. Face editing includes shape morphing [27, 28], relighting [29, 30], pose normalization [31–33], and expression modification [34, 35]. GAN-based methods can provide realistic results of facial attribute manipulation [36, 37] but is yet bounded with the limitation of its training images. The training images merely cover a narrow range of variations, and cause some artifacts in generation.

III. Problem Formalization

In order to investigate the influences of object features systematically, we generate synthetic face data targeting specific feature for face detectors. In section III-A, we introduce the influence of several major object features on face detectors. Then we provide basic information about face detectors in our experiments in section III-B. In section III-C, we explain how we synthesize face images based on 3D face models.

A. *Challenging Object features*

We will briefly discuss several features which have major effect on face detection performance.

Pose could significantly change the appearance of faces. Extreme pose can lead to heavy occlusion or skewed aspect ratio of face bounding boxes [33].

Scale is very challenging to deep network based modern object detectors. For example, the features between a 10px tall face and a 1000px tall face are essentially different [19]. Pyramid architecture and multi-scale inference are currently the common approaches to detect faces of extreme scales [38].

Context information plays a fundamental role in providing the precise location of faces. Normally, surrounding regions of faces provide complementary information on object appearance and high-level features [39]. However, faces in unconstrained settings may be surrounded or occluded by different distracting objects.

Facial occlusion decreases information available for detection and introduces additional noise. Facial occlusion can be divided into two different categories: landmark occlusion and heavy occlusion. Landmark occlusion means that only a few landmarks like eyes or mouths are occluded, while most parts of the faces are still visible. In contrast, heavy occlusion represents situations where more than half of the face is missing due to occlusion, image border or extreme pose. It is most challenging when the occlusion comes from other faces. A detector may identify several partially occluded faces as one face [4].

Blur and low resolution usually impede face detectors from retrieving available information. In some practical applications, images may be distorted in collection, storage, or transmission, leading to degraded quality of images [40]. In some extreme cases, mere outline of faces can be identified.

B. Face detectors

We provide an overview about the face detectors in our experiment. A summary of how face detectors differ in their performances can be found in supplementary material.

Faster RCNN is the most representative object detector based on deep network. However, its initial design does not have additional settings targeted at challenging features [41] in face detection task.

Single Stage Headless (SSH) [20] detector is an extremely fast one-step face detector, designed to be scale invariant. To accelerate the inference process, it has a light-weighted structure. This strategy jeopardizes the detector's performance when confronted with other potential variations.

Hybrid Resolutions (HR) face detector [19] has good performance on tiny face detection by using wide-range contextual information and testing on multiple resolutions. Its architecture resembles RPN [41] and uses both feature pyramid and image pyramid. However, HR face detector is extremely sensitive to tiny distracting objects from background. HR also heavily relies on contextual information to locate faces. For faces with limited information (e.g., heavily occluded, extremely small or blurry), complex background could hinder precise detection. Even though HR can sometimes perform well when dealing with blurry or extreme pose, it is insufficiently robust in the detection of occluded faces, especially when occlusion stems from other faces.

C. Face synthesis

Here we give a brief introduction on how we rendered images from 3D face models. Our synthetic data generation is based on a new 3D face dataset called 3DU Face Dataset. It has 700 3D face mesh models with high-resolution texture of 435 different individuals. Some people have multiple records taken at different times. Most models of this dataset are taken in varying conditions. For future research and applications, each model is annotated by humans with 50 landmarks.

An overview of our render pipeline can be found in Fig. 1. The rendering pipeline is built on Blender. To change the viewpoints, we rotated the model with different Euler angles with the camera staying in the same position. The parameters of pitch, yaw and roll are selected randomly within different ranges. For face scale variation, the distance between camera and face models is selected randomly from a uniform distribution. The ground truth for face detection is generated from 3D landmarks. The annotation policy is the same as in Wider Face. Our rendering pipeline can also be applied on other 3D face models.

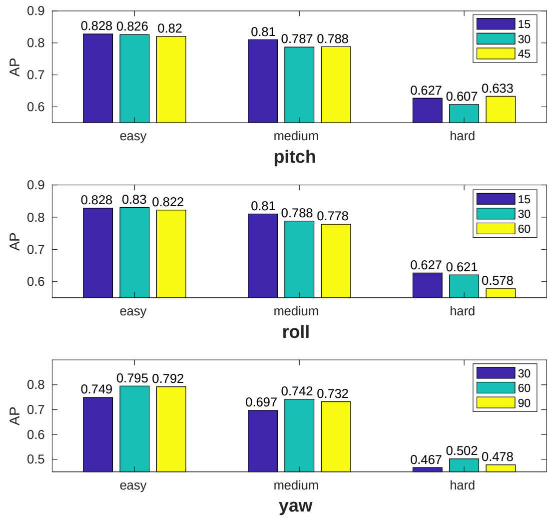

Fig. 2. Performance comparison on pose variation. Across pitch, yaw, and roll, different colors represent the rotated degree, as labeled at the top right (e.g., "15" means "-15" to "15").

In the current research, we consider face occlusion as a crucial factor in face detection. Common way to add occlusion is to crop face images. However, cropping reduces the information of faces and cannot provide reasonable noise as in real occlusion samples. We randomly add different 3D objects like sunglasses, hats, and helmets in the 3D scenes before we start rendering. With the anchoring of landmarks, all the objects can be placed in a reasonable location to simulate the landmark occlusion. Face region has been divided into three different parts of head, eye and mouth, to simulate occlusion. We can generate more than 1000 different combinations of occlusion for each model.

IV. Experiments

A. Experimental setup

We conduct experiments to systematically investigate how configuration of data augmentation influences performance of face detection. We test on advanced face detection benchmarks MAFA, UFDD and Wider Face. The face detection methods include Faster RCNN, SSH and HR. Despite the common practices of training and testing on the same datasets, here we first train the face detectors on synthetic data and then test on real data. We validate on subset of real data training part. After comparing the performance on real data with different rendering parameters, we attain a suitable configuration for one specific dataset. We use these augmented synthetic data to improve performance on real data. The metric for all the experiment is AP (average precision). We keep the original parameters and settings including data augmentation for each

TABLE I
THREE FACE DETECTION BENCHMARKS AND THEIR CHALLENGING CHARACTERISTICS.

Feature	MAFA	UFDD	Wider
landmark occlusion	✓	✓	✓
complex background			✓
extreme pose			✓
extreme scale		✓	✓
heavy occlusion	✓	✓	✓
blur	✓	✓	✓
extreme illumination		✓	✓
misleading objects		✓	✓

detector. More details of face detectors and rendering process can be found in supplementary material.

B. Datasets

We briefly introduce three face detection datasets used in our experiments and their features. In Table I, we listed and compared features of these face detection benchmarks, which are derived from their official introductions. For all three datasets, we follow official settings for splitting train and test set.

MAFA is a representative dataset of facial occlusion, which is mainly composed of various level of occlusions [4]. To exclude the interference of pose, MAFA only includes a narrow range of head poses. Faces are labeled as "Ignore" if they are very difficult to be detected.

UFDD contains faces in different weather conditions and other challenging features concerning lens impediments, motion blur and defocus blur [42]. Additionally, it has a collection of distracting images to enhance difficulty. In UFDD, the most challenging part is extreme lighting condition and blur.

Wider Face has been the most demanding benchmark for face detection pipeline till now [5]. It includes diverse events with a variety of backgrounds. The massive number of faces included has extreme poses, exaggerated expressions, heavy occlusion and extreme lighting conditions. All these features, especially scale, are difficult to handle for most face detectors. Table II shows the basic characteristics of faces, irrespective of invalid faces, in the Wider Face validation partition. Successively, the easy partition only has large faces; the medium partition additionally contains medium faces; and the hard partition includes the whole dataset.

TABLE II
FACE SCALE INFORMATION OF THE VALIDATION SET IN WIDER FACE. WE DISTINGUISH THREE FACE CATEGORIES BASED ON HEIGHT AND WIDTH. PROPORTION INFORMATION REPRESENTS THE PERCENTAGE OF FACES THAT FITS WITHIN THE SCALE INTERVAL.

Partition	Large	Medium	Tiny
Height	50-400 (96.6%)	30-50 (99%)	10-30 (99%)
Width	20-300 (96.3%)	10-70 (99.7%)	8-20 (95%)
Number	7211	6108	18636

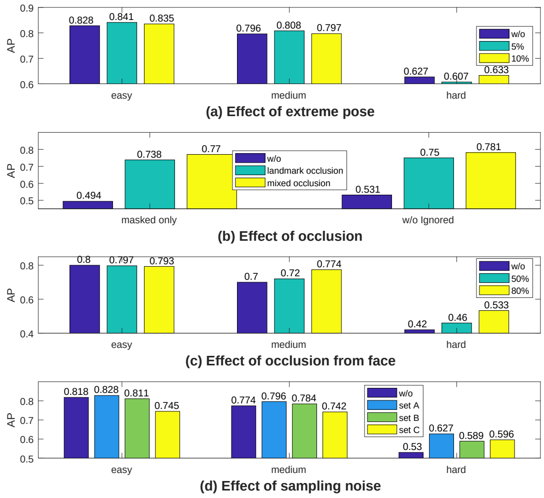

Fig. 3. Performance comparison on other features. Only (b) tests on MAFA test set, while the remaining are on Wider Face validation set. (a) shows the results after adding small-portion extreme pose into training dataset; (b) shows the results of adding different types of occlusion. "w/o Ignored" means face with label "Ignored" are not included; (c) shows the results of adding occlusion from other faces. (d) shows the results of adding different noise from down-sampling or up-sampling;

V. EVALUATION ON OBJECT FEATURES

Based on our own rendering pipeline, we are able to generate all kinds of synthetic datasets with fully controlled configuration. We can investigate in-depth the effect of various data augmentations. Compared with other detectors, the architecture of HR is specially designed for face detection. HR would be a suitable option to reveal all the effects of different data augmentations. The following experiments in this section are all based on HR. For all experiments, we study one feature at the time. The other features are kept the same. The comparison about the basic settings of rendering process is in supplementary material.

Effect of pose: We investigate the effect of head pose on Wider Face because it has a wide range of pose. To this end, we set different ranges for pitch, roll and yaw. As shown in Fig. 2, a narrow range of head pose provides better performance. Despite some extremes, most faces only have a small range of pose. When head pose becomes too extreme, performance starts degrading. Then we add different portion of face images with extreme orientation to the training dataset in Fig. 3 (a). A lower ratio of extreme data boosts the performance on "easy" and "medium" faces.

Effect of occlusion: We test two different kinds of face occlusion respectively. The first kind is from other objects except faces. MAFA concentrates on occluded face images, so we test on MAFA test set in three different settings: baseline condition with no occlusion, landmark occlusion setting, and mixed occlusion setting (including landmark and heavy occlusion). As shown in Fig. 3 (b), the performance on MAFA test set improves drastically after adding occlusion in the synthetic training dataset. HR becomes more robust after training on synthetic faces with landmark and heavy

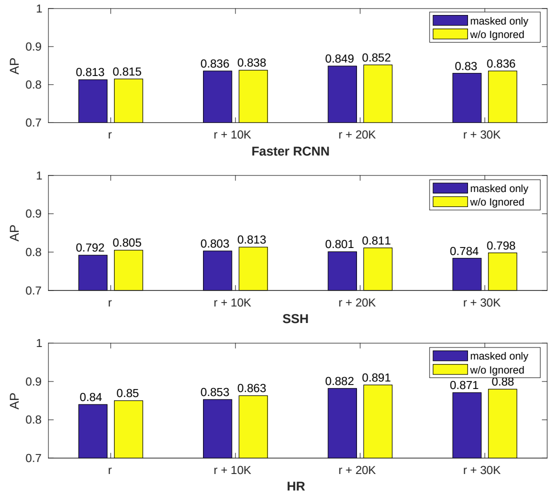

Fig. 4. Performance comparison on different sizes of synthetic data in training on MAFA test set with different detectors.

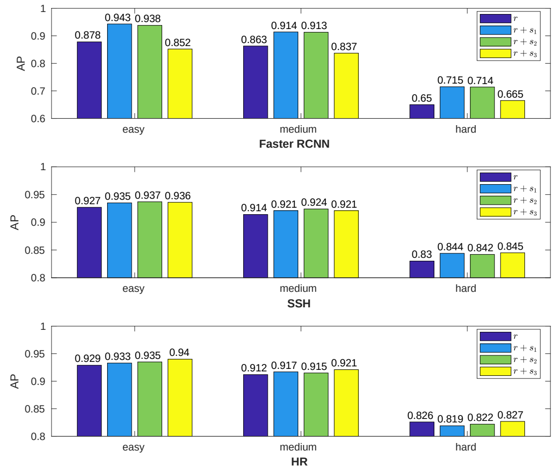

Fig. 5. Performance comparison on different data augmentations on Wider Face validation set with different detectors.

occlusions. Some occlusion examples can be found in Fig. 1.

The second kind of occlusion is from other faces or human body parts. We choose Wider Face validation set to test the effect. This is because many images in Wider Face are group pictures, and some image may have hundreds of overlapped faces. We only set the threshold for the overlap between synthetic faces, to avoid other large faces to cover the tiny faces. As shown in Fig. 3 (c), after adding occlusion from other faces, the performance of HR for hard level in Wider validation set improves substantially.

Effect of noise: Every benchmark has its own configurations when established. Wider Face dataset has a bias during its collection process: the original images downloaded from search engine are resized to one predetermined width 1024 pixel, which causes every image to have noise from down-sampling or up-sampling. Therefore, we first render images with multiple resolutions (as Set A, B and C as below) , and then re-size them to one fixed resolution (1024×768). Set A includes various high-resolution images (4096, 3072, 2048). Set B has various high- and low-resolution images (4096, 3072, 2048, 512, 256, 128). And Set C has various low-resolution images (512, 256, 128). We demonstrate the influence of noise at all the difficulty levels of Wider Face in Fig. 3 (d). The performance at all the difficulty levels of Wider Face has been improved, especially for tiny faces. Set A achieved the best performance because its pattern resembles tiny faces in Wider Face.

VI. Improve performance through synthetic data

In this part, we show how to use our synthetic data to improve performance on real datasets. We train on a combination of Wider Face and synthetic data and then test on other dataset. Visualization of our detection results can be found in Fig. 6.

A. Evaluation on MAFA

We only use mixed face occlusion (that is landmark and heavy occlusion as in Section V). The synthetic images for data augmentation follow the setting of MAFA training set as closely as possible. We presume the training data size would have an effect on the performance. As shown in Fig. 4, the performances of different detectors improve to some extent with the increase of synthetic data. We do note that when we add more synthetic data, the performance saturates and drops, however. Our tentative explanation is the inherent bias in our synthetic data.

B. Evaluation on UFDD

We show the influence of our data augmentation in Table III. Three different synthetic sets are combined with real data to improve detectors' performance. "r" denotes the detector is trained on real data. "$r+s_1$" denotes the detector is trained on a combination of real data and synthetic data set s_1. s_1 is our basic settings for rendering with light occlusion. s_2 combines s_1 with extra occlusion from other faces in the render process. s_3 adds additional blurry results from down-sampled high resolution images into s_2. More information can be found in supplementary material. After merging synthetic data and real data together, the performance of Faster RCNN, which was trained on real data, has been improved significantly on $r+s_1$ and $r+s_2$. Given that Faster RCNN was not trained on different scales, the noise of s_3 impedes it performance. As for SSH, its architecture and parameters heavily relies on Wider Face. The features in UFDD is too difficult for its light-weight structure. Unsurprisingly, its performance becomes saturated after being trained on real data. After adding synthetic data, its performance is even worse than Faster RCNN. Faces in UFDD are not very challenging to HR; the performance therefore only changes slightly after using our data augmentation.

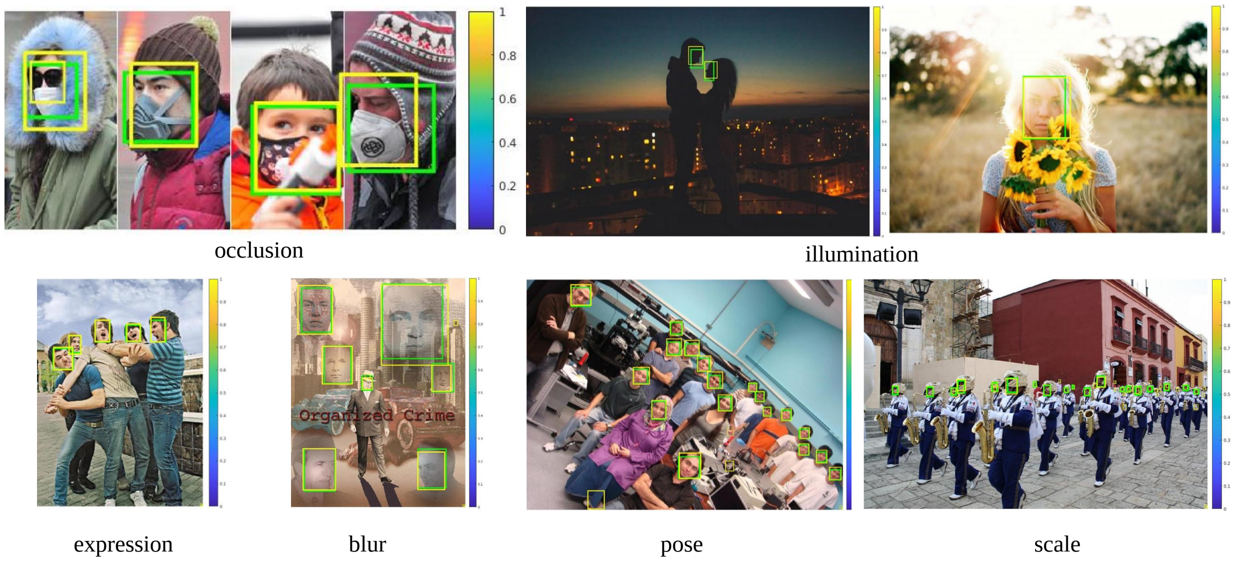

Fig. 6. Qualitative results on different features of real dataset. We visualize examples of each feature. The green bounding boxes are ground truth. The bounding boxes with other colors are predictions with different confidence intervals.

TABLE III
PERFORMANCE COMPARISON ON DIFFERENT DATA AUGMENTATIONS ON UFDD TEST SET WITH DIFFERENT DETECTORS.

Detectors	r	$r+s_1$	$r+s_2$	$r+s_3$
Faster RCNN	0.64	0.745	0.742	0.64
SSH	0.681	0.682	0.672	0.674
HR	0.721	0.733	0.721	0.736

C. Evaluation on Wider Face

We still use the same setting as in Section V to perform data augmentation for Wider Face. The performance comparisons are showed in Fig. 5. Faster RCNN is a generic object detector without multi-scale inference, so it is supposed to generate fewer predictions than HR and SSH. After we add synthetic data, the performance substantially improves on all difficult levels. The performances of HR and SSH nearly saturate after being trained on real data. Although their architectures aimed at Wider Face dataset, our synthetic data can still improve performance to a certain extent.

VII. Analysis

A. Analysis of object features

In general, proper amount of well-structured synthetic data can be a good complement to real data in training face detectors. The settings of synthetic data need to be similar with configurations of real data. If targeting at a single feature, the increase of data quantity cannot yield a consistent improvement on performances. For a more complex dataset like Wider Face, synthetic face images are generated with a comprehensive combination of several features.

The advantage of synthetic data is that the variations in dataset can be fully controlled in different practical situations. Dataset could be adjusted depending on specific requirements. Although, admittedly, there is always a domain gap between synthetic data and real data, synthetic data can provide large-scale dataset with annotation conveniently and precisely. Our results on multiple challenging benchmarks with different advanced detectors highlight the applicability of synthetic data as complement to real data, to equip face detectors against various features.

B. Analysis of face detectors

Based on our detection results, we analyze the performance of three detectors respectively. Faster RCNN is an object, instead of face, detector. It does not adjust settings of anchors for any face detection benchmarks, and has much fewer predictions without multi-scale inference. Despite that, our synthetic data augmentation substantially improves its performance on multiple challenging datasets.

SSH is the representative of one-step face detector. Even though SSH is a face-targeted detector, adding our synthetic data augmentation cannot help it outperform Faster RCNN in most detection tasks except in hard level of Wider Face. Specialized in scale, SSH sometimes has imperfect performance when encountering other features. Detectors have a trade-off between speed and performance. SSH pursues fast speed in inference process so that its light weight architecture cannot handle other features.

Of crucial importance, although HR face detector already has excellent performance in terms of all kinds of features, our synthetic data still boosts its performance. However, HR has a drawback that is extremely sensitive to tiny distracting objects given its tiny-face-targeted architecture. In general, HR generates more false positives than other detectors. The design of HR restricts the generalization on normal faces. Visualization and analysis for false positive examples can be found in supplementary material.

VIII. Conclusion

In this paper, we proposed an experimental comparison of main characteristics that influence face detection performance. We customized synthetic dataset to address specific types of features (scale, pose, occlusion, blur, etc.), and systematically investigated the influence of different features on face detection performance. Through our analyses, we also identified some potential deficiencies of the current face detection architectures. To conclude, there are often challenging features in real-world face detection. By providing an overview of the relationship between object features and face detection performances, we hope to assist researchers to choose more appropriate synthetic data when addressing challenging real-life variations.

References

[1] J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama *et al.*, "Speed/accuracy trade-offs for modern convolutional object detectors," in *IEEE CVPR*, 2017.

[2] S. Zafeiriou, C. Zhang, and Z. Zhang, "A survey on face detection in the wild: past, present and future," *Computer Vision and Image Understanding*, vol. 138, pp. 1–24, 2015.

[3] V. Jain and E. Learned-Miller, "FDDB: A benchmark for face detection in unconstrained settings," Technical Report UM-CS-2010-009, University of Massachusetts, Amherst, Tech. Rep., 2010.

[4] S. Ge, J. Li, Q. Ye, and Z. Luo, "Detecting masked faces in the wild with lle-cnns," in *IEEE CVPR*, 2017.

[5] S. Yang, P. Luo, C.-C. Loy, and X. Tang, "Wider face: A face detection benchmark," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 5525–5533.

[6] J. Wang and L. Perez, "The effectiveness of data augmentation in image classification using deep learning," *Convolutional Neural Networks Vis. Recognit*, p. 11, 2017.

[7] L. Taylor and G. Nitschke, "Improving deep learning with generic data augmentation," in *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 2018, pp. 1542–1547.

[8] J.-J. Lv, X.-H. Shao, J.-S. Huang, X.-D. Zhou, and X. Zhou, "Data augmentation for face recognition," *Neurocomputing*, vol. 230, pp. 184–196, 2017.

[9] I. Masi, A. T. Tran, T. Hassner, G. Sahin, and G. Medioni, "Face-specific data augmentation for unconstrained face recognition," *International Journal of Computer Vision*, vol. 127, no. 6-7, pp. 642–667, 2019.

[10] D. Hoiem, Y. Chodpathumwan, and Q. Dai, "Diagnosing error in object detectors," in *European conference on computer vision*. Springer, 2012, pp. 340–353.

[11] Z. Yang and R. Nevatia, "A multi-scale cascade fully convolutional network face detector," in *Pattern Recognition (ICPR), 2016 23rd International Conference on*. IEEE, 2016, pp. 633–638.

[12] S. Zhang, X. Zhu, Z. Lei, H. Shi, X. Wang, and S. Z. Li, "S^3FD: Single shot scale-invariant face detector," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 192–201.

[13] Z. Hao, Y. Liu, H. Qin, J. Yan, X. Li, and X. Hu, "Scale-aware face detection," in *The IEEE Conference on Computer Vision and Pattern Recognition*, vol. 3, 2017.

[14] X. Tang, D. K. Du, Z. He, and J. Liu, "PyramidBox: A context-assisted single shot face detector," in *Proceedings of the European Conference on Computer Vision*, 2018, pp. 797–813.

[15] S. Zhang, L. Wen, H. Shi, Z. Lei, S. Lyu, and S. Z. Li, "Single-shot scale-aware network for real-time face detection," *International Journal of Computer Vision*, pp. 1–23, 2019.

[16] J. Wang, Y. Yuan, and G. Yu, "Face attention network: An effective face detector for the occluded faces," *arXiv preprint arXiv:1711.07246*, 2017.

[17] X. Shi, S. Shan, M. Kan, S. Wu, and X. Chen, "Real-time rotation-invariant face detection with progressive calibration networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2295–2303.

[18] M. Zhou, H. Lin, S. S. Young, and J. Yu, "Hybrid sensing face detection and registration for low-light and unconstrained conditions," *Applied optics*, vol. 57, no. 1, pp. 69–78, 2018.

[19] P. Hu and D. Ramanan, "Finding tiny faces," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 951–959.

[20] M. Najibi, P. Samangouei, R. Chellappa, and L. Davis, "SSH: Single stage headless face detector," in *IEEE ICCV*, 2017.

[21] L. Li, Y. Peng, G. Qiu, Z. Sun, and S. Liu, "A survey of virtual sample generation technology for face recognition," *Artificial Intelligence Review*, vol. 50, no. 1, pp. 1–20, 2018.

[22] X. Wang, K. Wang, and S. Lian, "A survey on face data augmentation," *arXiv preprint arXiv:1904.11685*, 2019.

[23] M. Osadchy, Y. Wang, O. Dunkelman, S. Gibson, J. Hernandez-Castro, and C. Solomon, "Genface: Improving cyber security using realistic synthetic face generation," in *International Conference on Cyber Security Cryptography and Machine Learning*. Springer, 2017, pp. 19–33.

[24] I. Abbasnejad, S. Sridharan, D. Nguyen, S. Denman, C. Fookes, and S. Lucey, "Using synthetic data to improve facial expression analysis with 3d convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1609–1618.

[25] A. Kortylewski, A. Schneider, T. Gerig, B. Egger, A. Morel-Forster, and T. Vetter, "Training deep face recognition systems with synthetic data," *arXiv preprint arXiv:1802.05891*, 2018.

[26] I. Masi, A. T. Tran, T. Hassner, J. T. Leksut, and

G. Medioni, "Do we really need to collect millions of faces for effective face recognition?" in *European Conference on Computer Vision*. Springer, 2016, pp. 579–596.

[27] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3d faces," in *Proceedings on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co., 1999, pp. 187–194.

[28] K. Scherbaum, J. Petterson, R. S. Feris, V. Blanz, and H.-P. Seidel, "Fast face detector training using tailored views," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 2848–2855.

[29] Y. Wang, L. Zhang, Z. Liu, G. Hua, Z. Wen, Z. Zhang, and D. Samaras, "Face relighting from a single image under arbitrary unknown lighting conditions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 11, pp. 1968–1984, 2009.

[30] Z. Shu, E. Yumer, S. Hadap, K. Sunkavalli, E. Shechtman, and D. Samaras, "Neural face editing with intrinsic image disentangling," in *Computer Vision and Pattern Recognition, 2017 IEEE Conference on*. IEEE, 2017, pp. 5444–5453.

[31] T. Hassner, S. Harel, E. Paz, and R. Enbar, "Effective face frontalization in unconstrained images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 4295–4304.

[32] X. Zhu, Z. Lei, J. Yan, D. Yi, and S. Z. Li, "High-fidelity pose and expression normalization for face recognition in the wild," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 787–796.

[33] J. Yim, H. Jung, B. Yoo, C. Choi, D. Park, and J. Kim, "Rotating your face using multi-task deep neural network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 676–684.

[34] J. Thies, M. Zollhofer, M. Stamminger, C. Theobalt, and M. Nießner, "Face2Face: Real-time face capture and reenactment of RGB videos," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2387–2395.

[35] A. Pumarola, A. Agudo, A. M. Martinez, A. Sanfeliu, and F. Moreno-Noguer, "Ganimation: Anatomically-aware facial animation from a single image," in *Proceedings of the European Conference on Computer Vision*, 2018, pp. 818–833.

[36] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, "Stargan: Unified generative adversarial networks for multi-domain image-to-image translation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8789–8797.

[37] W. Shen and R. Liu, "Learning residual images for face attribute manipulation," in *2017 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2017, pp. 1225–1233.

[38] B. Singh and L. S. Davis, "An analysis of scale invariance in object detection snip," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 3578–3587.

[39] Z. Chen, S. Huang, and D. Tao, "Context refinement for object detection," in *Proceedings of the European Conference on Computer Vision*, 2018, pp. 71–86.

[40] Y. Zhou, D. Liu, and T. Huang, "Survey of face detection on low-quality images," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. IEEE, 2018, pp. 769–773.

[41] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91–99.

[42] H. Nada, V. A. Sindagi, H. Zhang, and V. M. Patel, "Pushing the limits of unconstrained face detection: a challenge dataset and baseline results," *arXiv preprint arXiv:1804.10275*, 2018.