Pacific Graphics 2019
C. Theobalt, J. Lee, and G. Wetzstein
(Guest Editors)

Volume 38 (2019), Number 7

Interactive Curation of Datasets for Training and Refining Generative Models

–Supplementary Material–

Wenjie Ye[1,2] Yue Dong[2] Pieter Peers[3]

[1]Tsinghua University
[2]Microsoft Research Asia
[3]College of William & Mary

This supplemental material lists additional ablation results, validations, and additional results produced by the GANs trained with curated datset.

1. Ablation Results for each Selection Criterion

Table 1 lists the ablation accuracy for each selection criteria separately. We toggle various combination of Query-by-Committee (QBC), allowing an *"undecided"* label (UL), using the disagreement distance (DD), and using parallel candidate selection and labeling (Parallel) to improve performance. The performance of each of the components are consistent for each of the cases.

Figure 1 showcases selected generated samples for each of the texture selection criteria.

2. Additional Numerical Validations

Based on the reference labels in CelebA [LLWT15], we synthesize additional selection criteria, and validate the performance of our system compared to a labeling on (an equal number of) randomly selected exemplars as well as compared to a reference classifier trained on the *full* dataset using the reference labels (Table 2). Similar as before, the accuracy of our interactive curation system is closer to the upperbound, and significantly better than random sampling.

Figure 2 showcases selected generated exemplars for each considered face selection criteria.

3. Additional Results

We showed that our framework can be used to remove unwanted samples with artifacts from a GAN. However, we can also use the same system for removing unwanted "features". Figure 3 shows an example of removing the "beard" features from generated samples.

References

[LLWT15] LIU Z., LUO P., WANG X., TANG X.: Deep learning face attributes in the wild. In *ICCV* (2015). 1

Figure 1: *Synthesized texture examples that follow the user's selection criteria used in the quantitative validation.*

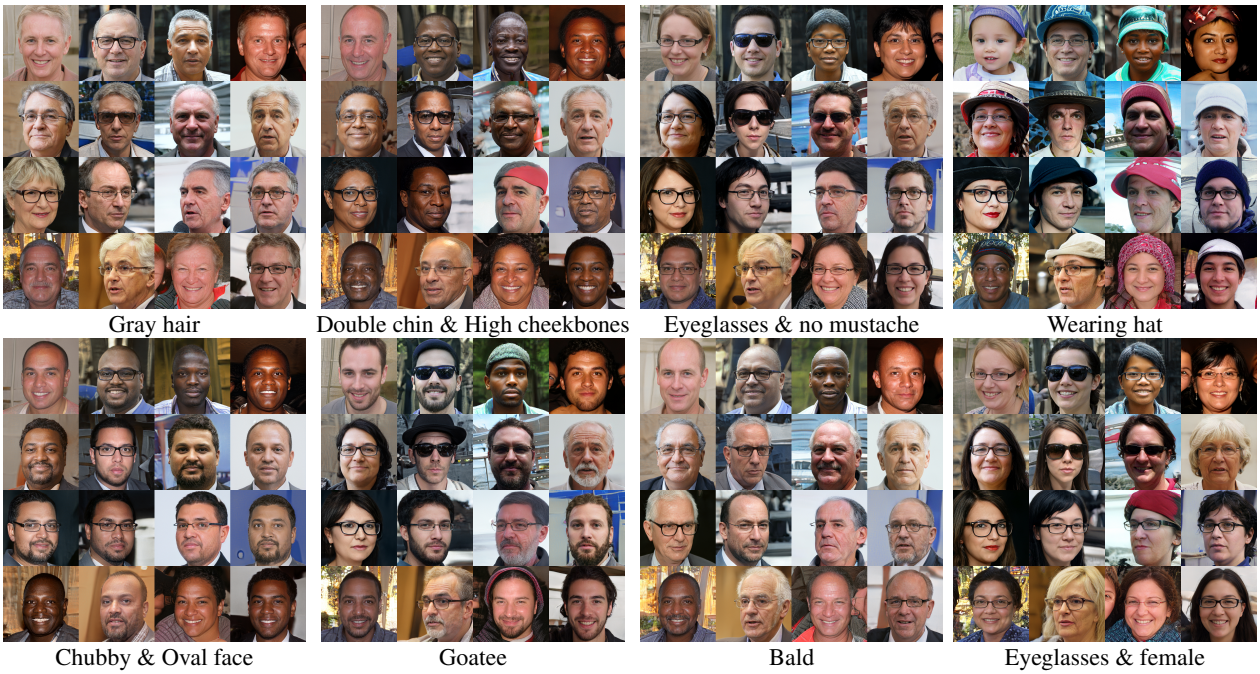

Figure 2: *Generated face examples that follow the user's selection criteria used in the quantitative validation.*

(a) Original GAN (b) Improved GAN

Figure 3: *Example of removing unwanted features from the generated samples. (a) Original GAN with unwanted "beard" features. (b) Improved GAN refined from the original GAN without the unwanted feature (i.e., no beard).*

	TAR		
	FAR 0.01	FAR 0.05	FAR 0.1
Wood			
Low Contrast			
Random	0.566	0.825	0.913
QBC	0.562	0.839	0.929
QBC + UL	0.587	0.873	0.947
QBC + DD	0.659	0.888	0.958
QBC + UL + DD	0.670	0.903	0.963
Our + Parallel	0.661	0.896	0.961
Hue Cold			
Random	0.699	0.851	0.882
QBC	0.929	0.949	0.952
QBC + UL	0.978	0.993	0.997
QBC + DD	0.938	0.952	0.953
QBC + UL + DD	0.960	0.994	0.998
Our + Parallel	0.920	0.944	0.947
Horizontal			
Random	0.742	0.935	0.978
QBC	0.862	0.980	0.995
QBC + UL	0.898	0.990	0.997
QBC + DD	0.882	0.985	0.996
QBC + UL + DD	0.922	0.992	0.998
Our + Parallel	0.889	0.985	0.996
Directional			
Random	0.212	0.457	0.593
QBC	0.307	0.546	0.677
QBC + UL	0.349	0.588	0.718
QBC + DD	0.285	0.537	0.691
QBC + UL + DD	0.401	0.656	0.780
Our + Parallel	0.380	0.651	0.771
Manually Marked			
Random	0.540	0.710	0.786
QBC	0.676	0.776	0.805
QBC + UL	0.850	0.886	0.903
QBC + DD	0.907	0.932	0.943
QBC + UL + DD	0.963	0.985	0.991
Our + Parallel	0.971	0.989	0.994
Metal			
High Contrast			
Random	0.717	0.898	0.950
QBC	0.849	0.933	0.957
QBC + UL	0.861	0.939	0.963
QBC + DD	0.897	0.962	0.979
QBC + UL + DD	0.905	0.964	0.982
Our + Parallel	0.909	0.964	0.980
Stone			
Hue Cold			
Random	0.725	0.829	0.865
QBC	0.562	0.603	0.675
QBC + UL	0.670	0.711	0.724
QBC + DD	0.784	0.823	0.853
QBC + UL + DD	0.860	0.909	0.913
Our + Parallel	0.773	0.827	0.845

Table 1: *Ablation study by enabling/disabling various combinations of: query-by-committee (QBC), allowing an* "undecided" *label (UL), using the disagreement distance (DD), and using parallel candidate selection and labeling (Parallel) to improve performance.*

	TAR					
FAR	0.001	0.01	0.02	0.05	0.1	0.2
Gray hair						
Random	0.126	0.403	0.528	0.714	0.833	0.921
Our	0.145	0.648	0.750	0.862	0.929	0.969
All	0.187	0.624	0.774	0.921	0.975	0.997
Double chin & High cheekbones						
Random	0.011	0.088	0.141	0.272	0.404	0.577
Our	0.078	0.228	0.307	0.466	0.647	0.806
All	0.083	0.372	0.508	0.727	0.854	0.945
Eyeglasses & No Mustache						
Random	0.346	0.785	0.842	0.924	0.958	0.983
Our	0.509	0.947	0.965	0.982	0.987	0.991
All	0.565	0.971	0.983	0.990	0.993	0.995
Wearing hat						
Random	0.261	0.628	0.747	0.839	0.914	0.964
Our	0.495	0.837	0.894	0.939	0.968	0.987
All	0.626	0.931	0.967	0.981	0.987	0.994
Chubby & Oval face						
Random	0.011	0.038	0.065	0.115	0.179	0.305
Our	0.118	0.210	0.271	0.366	0.450	0.561
All	0.038	0.248	0.363	0.565	0.744	0.908
Goatee						
Random	0.078	0.282	0.464	0.674	0.804	0.911
Our	0.072	0.426	0.577	0.776	0.880	0.943
All	0.137	0.554	0.713	0.907	0.973	0.996
Bald						
Random	0.128	0.518	0.664	0.882	0.948	0.965
Our	0.251	0.735	0.846	0.920	0.962	0.979
All	0.274	0.837	0.913	0.979	0.993	0.998
Eyeglasses & Female						
Random	0.252	0.565	0.688	0.798	0.861	0.918
Our	0.700	0.927	0.959	0.968	0.981	0.991
All	0.830	0.950	0.978	0.994	0.994	0.997

Table 2: *A comparison of TAR scores on different tasks of face selection criterion with different labeling strategies: labeling* 600 *random selected candidates, labeling* 600 *candidates selected with our interactive system, and using* all *reference labels over the whole dataset.*