Benchmarking Synthetic Data Generation Approaches for Eyeglasses Detection Algorithms Development

Dalius MATUZEVIČIUS, Henrikas GIEDRA

E-mail: {dalius.matuzevicius|henrikas.giedra}@vilniustech.lt,

Department of Electronic Systems, Vilnius Gediminas Technical University (VILNIUS TECH)

Introduction

The development of accurate eyeglasses detection algorithms relies on large, annotated datasets to train convolutional neural networks (CNNs). However, acquiring and labeling real-world images is often resource-intensive, posing challenges in scaling datasets to encompass diverse eyewear styles and lighting conditions. This research benchmarks the performance CNN-based eyeglasses detectors, when trained on genuine, semi-synthetic, and fully synthetic datasets. Synthetic datasets are computer-simulated, offering precise control over object diversity, pose, and labeling accuracy.

Aim

The aim of this research is to evaluate the effectiveness of synthetic datasets as alternatives to genuine datasets for the development of robust eyeglasses detection algorithms.

Methods

Two genuine image datasets (FFHQ[1] and CelebAMask-HQ[2]) and two synthetic image datasets (Face Synthetics[3] and StyleGAN2-generated[4]) were used to train the RetinaNet CNN model, utilizing four different backbones.

Table 1: Summary of the datasets used to develop the eyeglass detectors

Dataset	Number of Images	Number of Glasses	Resolution
FFHQ	70,000	16,039	1024×1024
CelebAMask-HQ	30,000	1545	1024×1024
Face Synthetics	100,000	14,303	512×512
StyleGAN2 generated	2664	2664	1024×1024

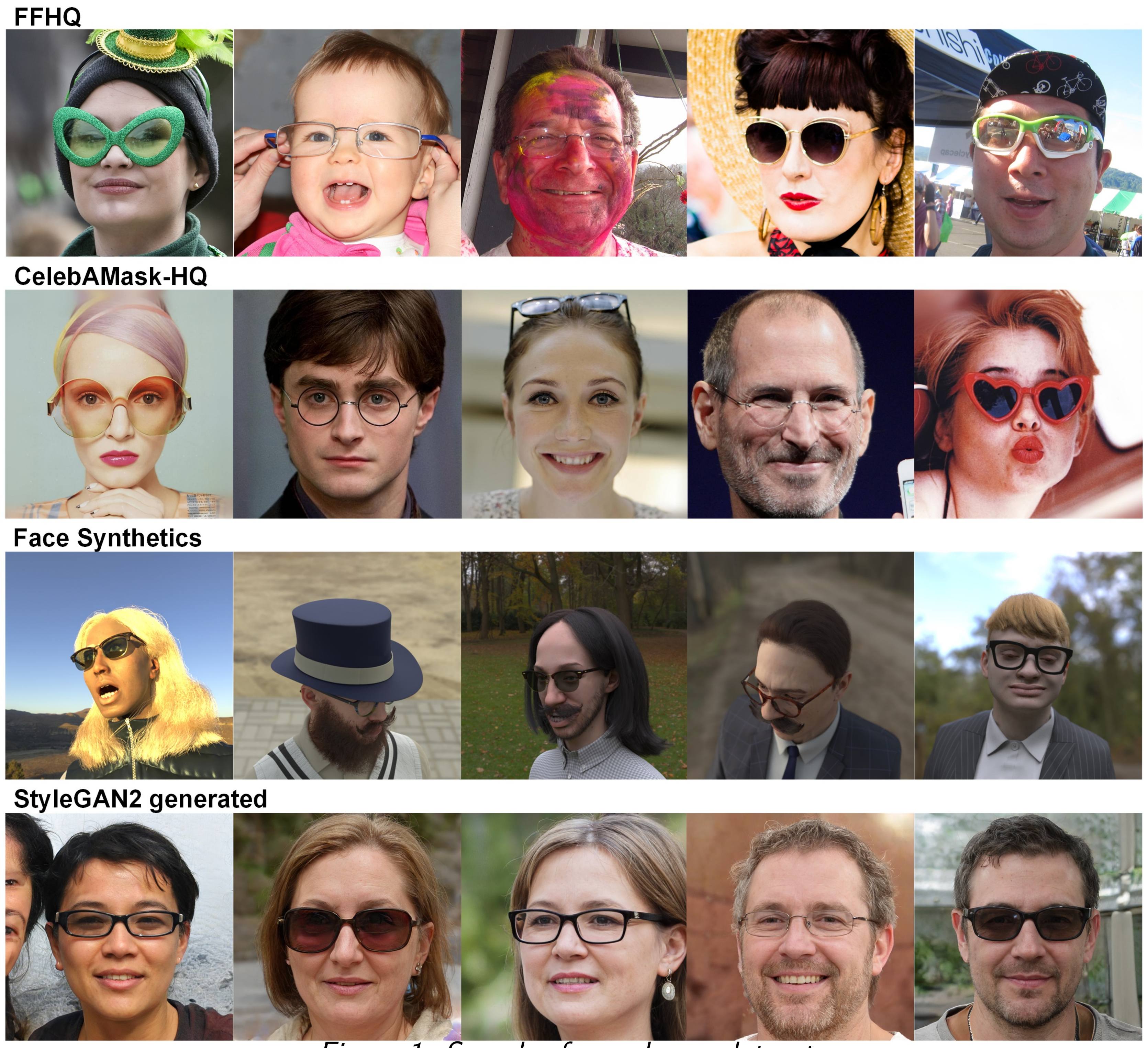

Figure 1: Samples from glasses datasets

[1]https://github.com/NVlabs/ffhq-dataset
[2]https://github.com/switchablenorms/CelebAMask-HQ
[3]https://github.com/microsoft/FaceSynthetics
[4]https://github.com/NVlabs/stylegan2

Table 2: Summary of the backbones

Backbone[1]	Parameters[2]	Depth[3]	Size (MB)[4]	Input[5]	Latency CPU (ms)[6]	Latency GPU (ms)[6]
YOLOv8xs	1.28M	27	4.87	512^2	484	24.8
YOLOv8s	5.09M	27	19.42	512^2	1070	33.2
YOLOv8m	11.87M	39	45.29	512^2	2210	44.9
MobileNetV3s	939.12K	34	3.58	512^2	479	29.3

[1] Backbone type; [2] Number of backbone's parameters; [3] number of convolutional layers; [4] size of the weights file; [5] input size (resolution); [6] time per inference step on CPU/GPU evaluated by averaging 30 batches of size 32, and 10 repetitions (CPU: Intel Core i7 12700K Processor; RAM: 128 GB; GPU: NVIDIA GeForce RTX 4090, 24 GB; platform: Windows Subsystem for Linux (WSL) 2).

Results

Table 3: Comparing real and synthetic images for the development of eyeglass detection algorithms

		Test Data			
		FFHQ Glasses		CelebAMask-HQ	
Model	Training Data	AP[1]	AR[1]	AP	AR
YOLOv8xs	FFHQ Glasses	0.86	0.89	0.83	0.87
	CelebAMask-HQ	0.77	0.83	0.79	0.84
	Face Synthetics	0.76	0.79	0.76	0.80
	StyleGAN2 gen.	0.79	0.84	0.76	0.81
YOLOv8s	FFHQ Glasses	0.87	0.90	0.84	0.88
	CelebAMask-HQ	0.78	0.84	0.79	0.85
	Face Synthetics	0.78	0.81	0.78	0.82
	StyleGAN2 gen.	0.80	0.85	0.78	0.83
YOLOv8m	FFHQ Glasses	0.88	0.91	0.84	0.88
	CelebAMask-HQ	0.79	0.84	0.81	0.85
	Face Synthetics	0.79	0.82	0.78	0.82
	StyleGAN2 gen.	0.81	0.85	0.79	0.83
MobileNetV3s	FFHQ Glasses	0.83	0.87	0.80	0.84
	CelebAMask-HQ	0.74	0.80	0.76	0.82
	Face Synthetics	0.74	0.78	0.75	0.79
	StyleGAN2 gen.	0.74	0.79	0.69	0.74

[1] Average Precision and Average Recall. AP and AR are averaged over multiple Intersection over Union (IoU) values (0.50:0.05:0.95.)

Discussion

- Synthetic images demonstrate significant potential as viable alternatives to genuine datasets for training ML models.
- Synthetic datasets provide flexibility in image generation and annotation while maintaining detection accuracy comparable to models trained on genuine datasets.
- CNN architectures (YOLOv8 and MobileNetV3) trained on synthetic datasets achieved detection performance with minimal degradation compared to those trained on real high-quality datasets, or even outperformed real datasets with lower diversity. This suggests that synthetic data can effectively capture the variability necessary for robust model training.
- Synthetic data generation methods are less resource-intensive than collecting and annotating large real-world datasets. This scalability enables researchers to create extensive datasets encompassing diverse eyewear styles and lighting conditions without prohibitive costs.
- The findings underscore the potential of synthetic data generation as a generalizable solution for vision tasks, particularly in scenarios where annotated real-world datasets are limited or expensive to obtain.
- Further exploration of synthetic data augmentation techniques and their integration with real-world datasets could enhance performance across a broader spectrum of computer vision tasks. Additionally, refining synthetic data generation methods to mimic real-world complexities more accurately could further improve detection accuracy and robustness.

VILNIUS TECH
Department of Electronic Systems

This project has received funding from the Research Council of Lithuania (LMTLT), agreement No S-ITP-24-12.