Quality-Diversity Generative Sampling for Learning with Synthetic Data

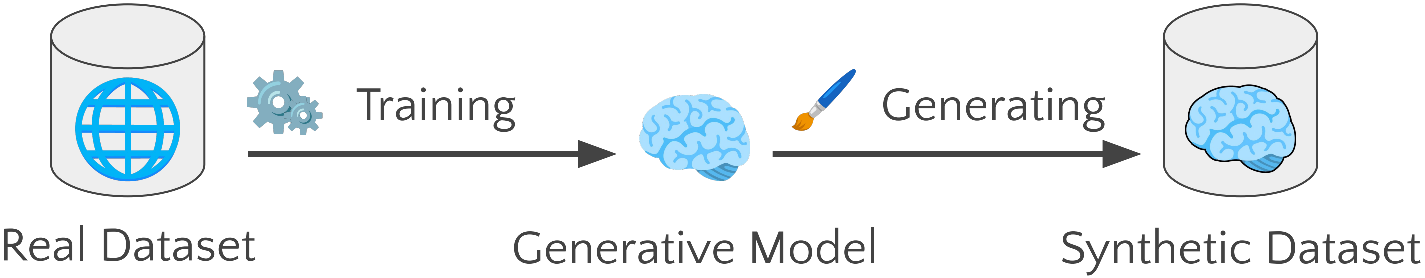
Allen Chang[1], Matthew C. Fontaine[1], Serena Booth[2], Maja J. Matarić[1], Stefanos Nikolaidis[1]

1 USC 2 MIT

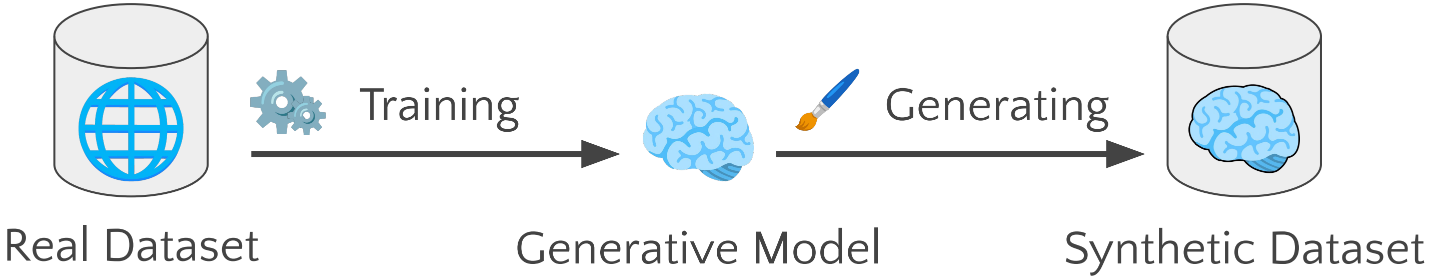
Paper 11791

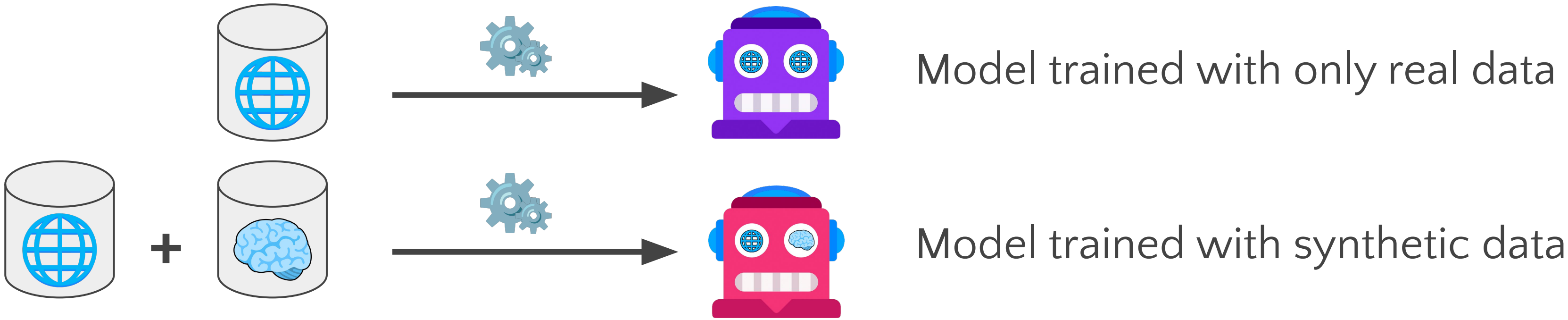
AAAI-24 / IAAI-24 / EAAI-24

Learning with Synthetic Data



Learning with Synthetic Data





More Accurate, Less Fair?

Synthetic Data from Diffusion Models Improves ImageNet Classification

Shekoofeh Azizi, Simon Kornblith, Chitwan Saharia*, Mohammad Norouzi*, David J. Fleet
Google Research, Brain Team †

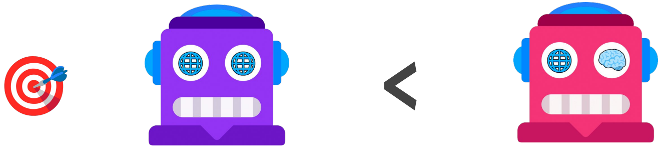
StableRep: Synthetic Images from Text-to-Image Models Make Strong Visual Representation Learners

Yonglong Tian[1,*] **Lijie Fan**[1,2,*] **Phillip Isola**[2] **Huiwen Chang**[1] **Dilip Krishnan**[1]
[1]Google Research, [2]MIT CSAIL, *equal contribution

phillip_isola

"Often you have a bunch of mappings X→Y, Y→Z, etc and you want other mappings implied by these.

A simple approach is to use the given mappings to sample training data for the implied mappings …"



Models with synthetic training data can be more accurate

Models with synthetic training data can be ~~more fair~~

More Accurate, **Less Fair?**

Synthetic Data from Diffusion Models Improves ImageNet Classification

Shekoofeh Azizi, Simon Kornblith, Chitwan Saharia*, Mohammad Norouzi*, David J. Fleet
Google Research, Brain Team †

StableRep: Synthetic Images from Text-to-Image Models Make Strong Visual Representation Learners

Yonglong Tian[1,*] Lijie Fan[1,2,*] Phillip Isola[2] Huiwen Chang[1] Dilip Krishnan[1]
[1]Google Research, [2]MIT CSAIL, *equal contribution

 phillip_isola

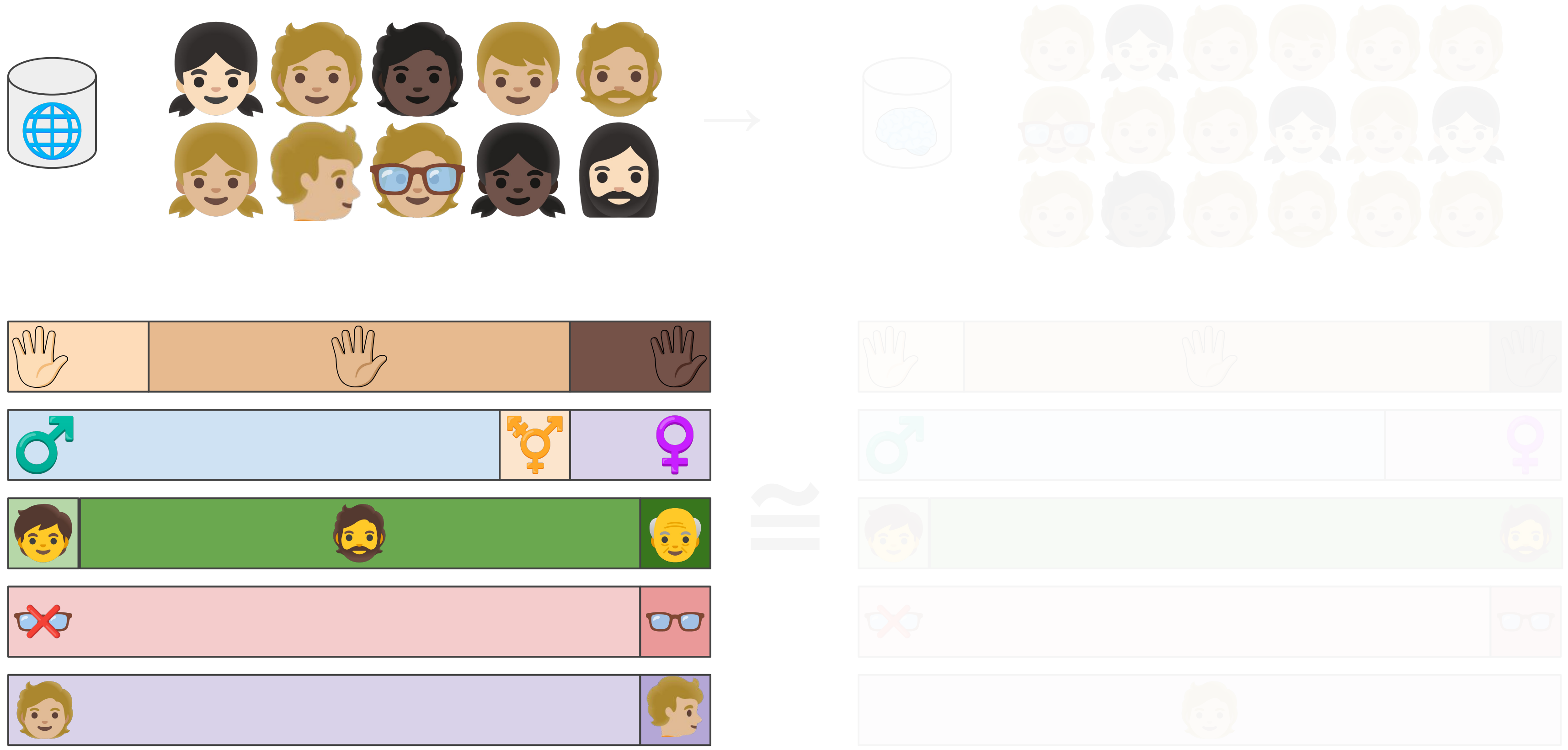
"Often you have a bunch of mappings X→Y, Y→Z, etc and you want other mappings implied by these.

A simple approach is to use the given mappings to sample training data for the implied mappings …"

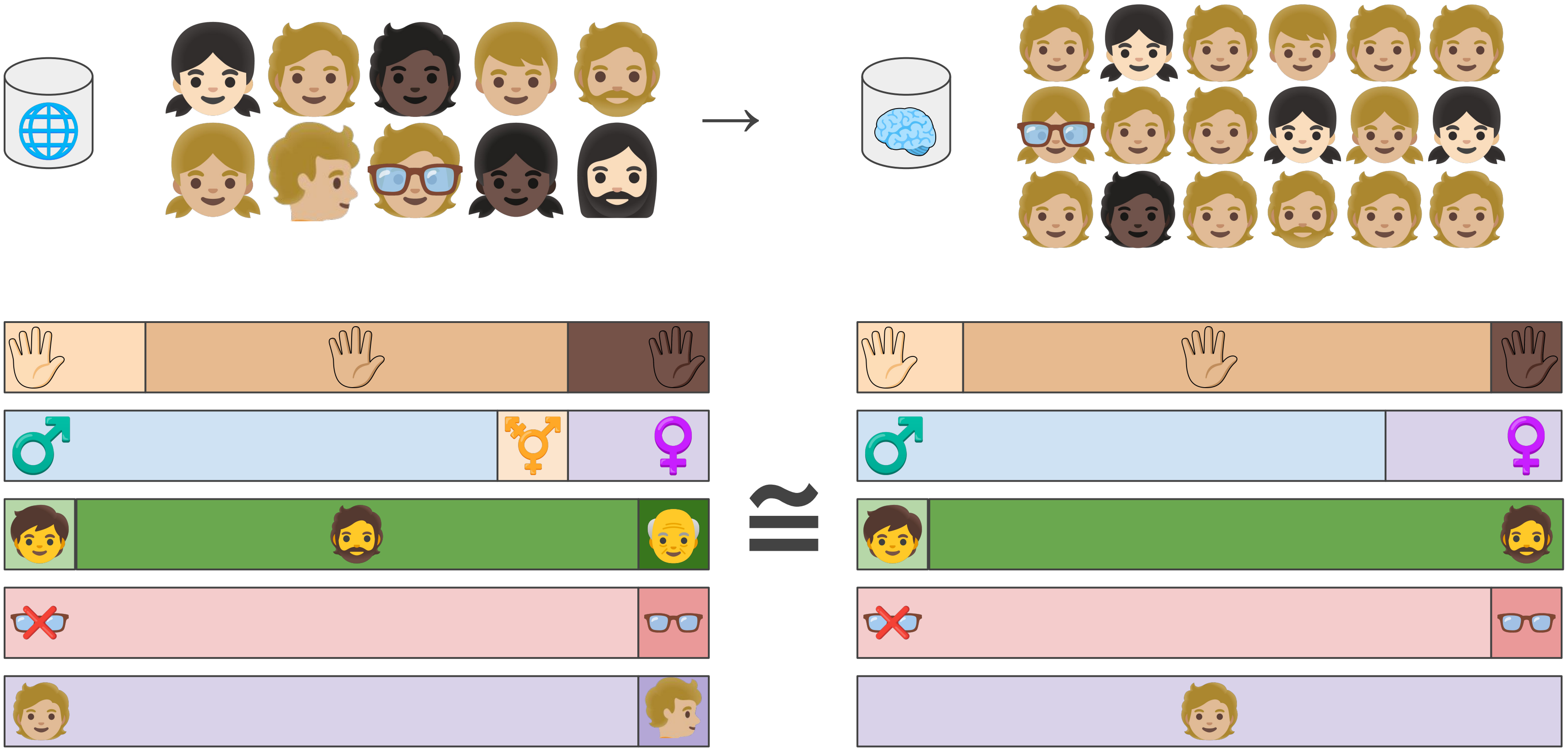
Models with synthetic training data can be more accurate

Models with synthetic training data can be more fair?

Bias in Real Data → Bias in Synthetic Data



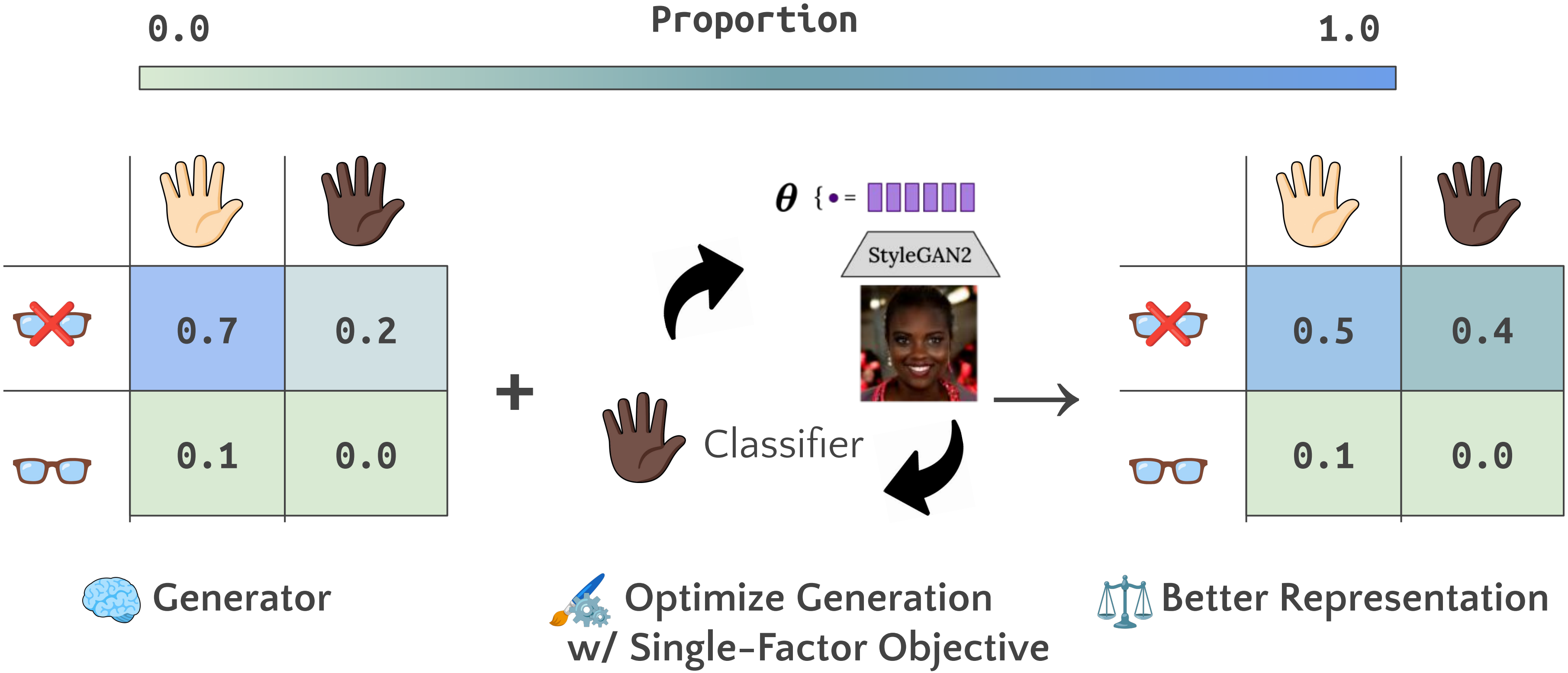
Bias in Real Data → Bias in Synthetic Data



Bias in Real Data → Bias in Synthetic Data

	Skin Tone			Age		Gender		
	Light	*Mixed*	*Dark*	*Young*	*Old*	*Masc*	*Andro*	*Fem*
FFHQ Dataset	73.1	17.1	09.8	64.9	35.1	44.2	03.9	51.9
StyleGAN2	69.8	20.8	09.4	63.0	37.0	44.0	04.2	51.8

Prior Work: How to Protect Diversity?

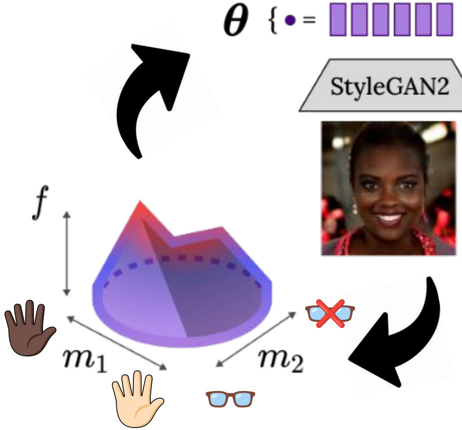


Our Approach: Quality-Diversity Generative Sampling (QDGS)

0.0 Proportion 1.0

	✋🏻	✋🏿
❌👓	0.7	0.2
👓	0.1	0.0

\+



→

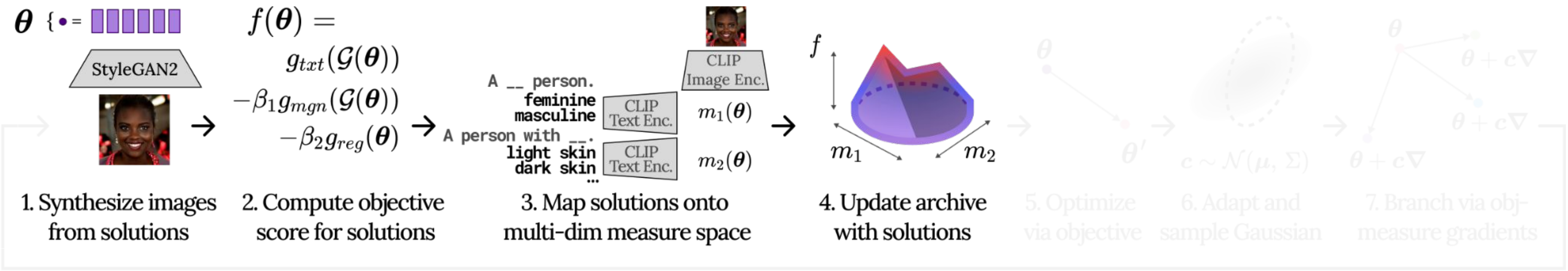
	✋🏻	✋🏿
❌👓	0.3	0.3
👓	0.2	0.2

🧠 Generator

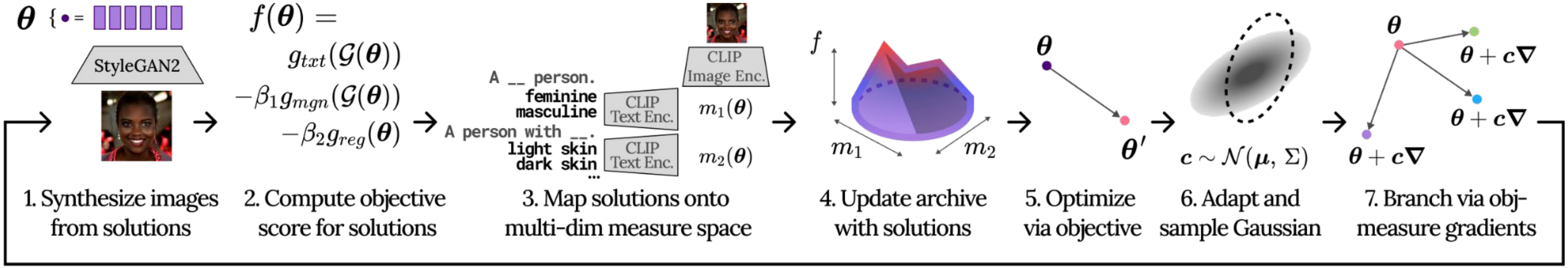
Optimize Generation w/ Quality-Diversity

⚖️ Better Representation

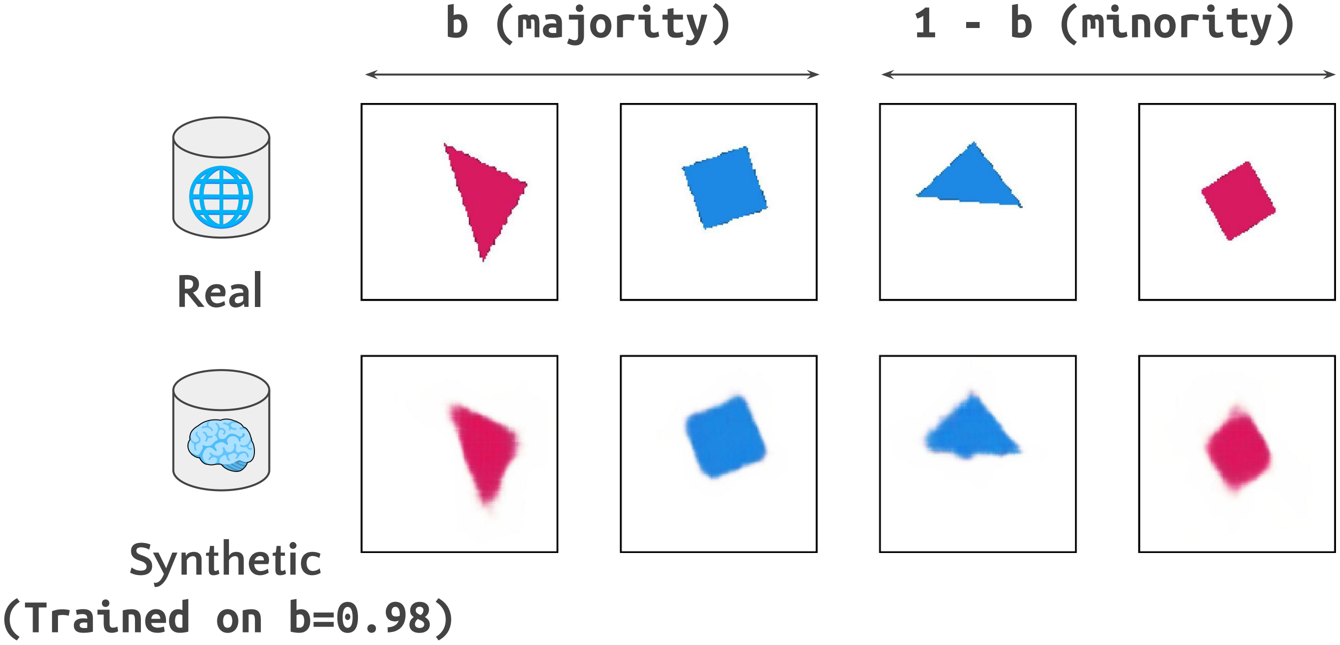
QDGS: Creating Balanced & Intersectional Data



QDGS: Creating Balanced & Intersectional Data

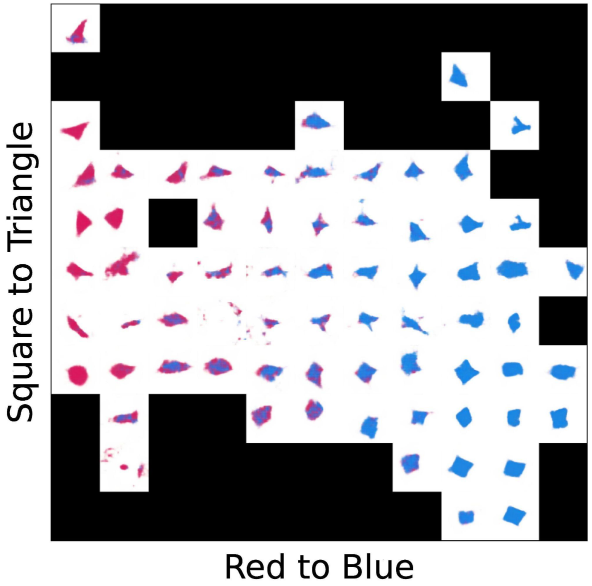


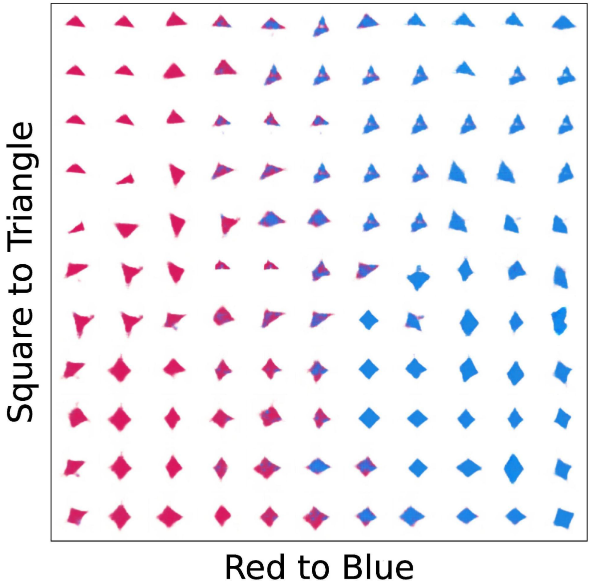
Toy Domain: Color-Biased Shapes



5 Versions: $b \in \{0.98, 0.95, 0.90, 0.85, 0.80\}$

Spread w.r.t. CLIP

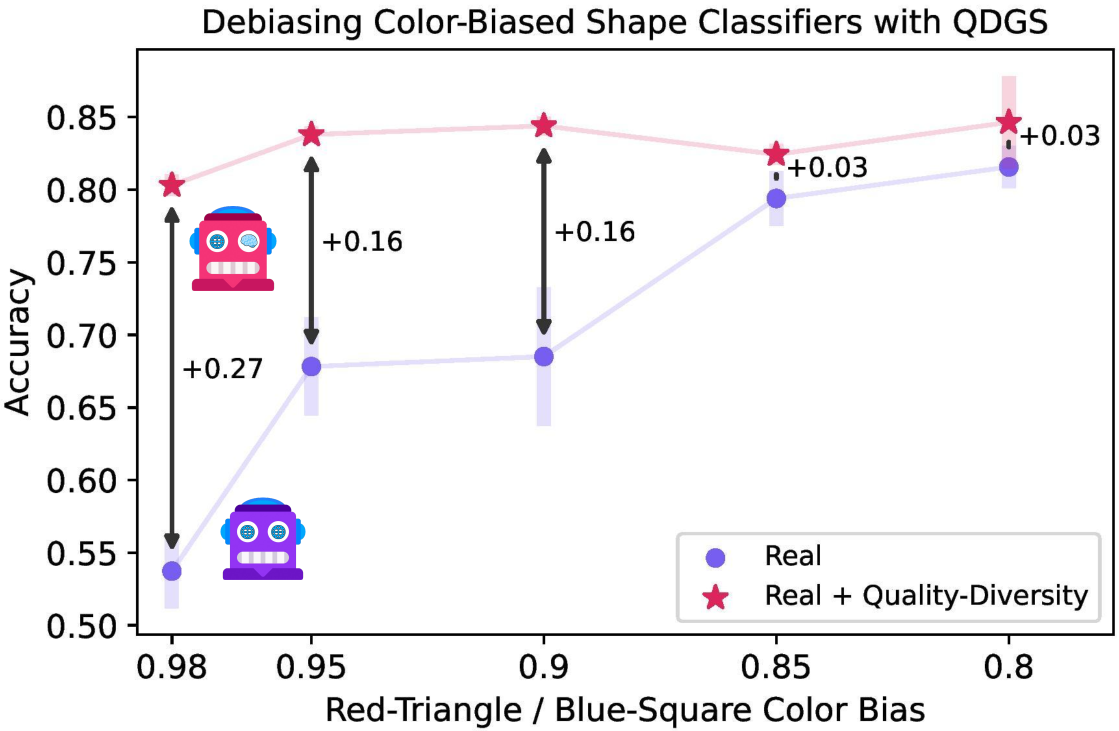




🧠 Generator

🧠 Generator + QDGS

Repairing Biases in Classifiers



QDGS on StyleGAN2

1. Generate Synthetic Pretraining Datasets

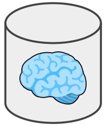
Rand15	Rand50
QD15 (Ours)	**QD50 (Ours)**

2. Pretrain + Train Facial Recognition Classifiers

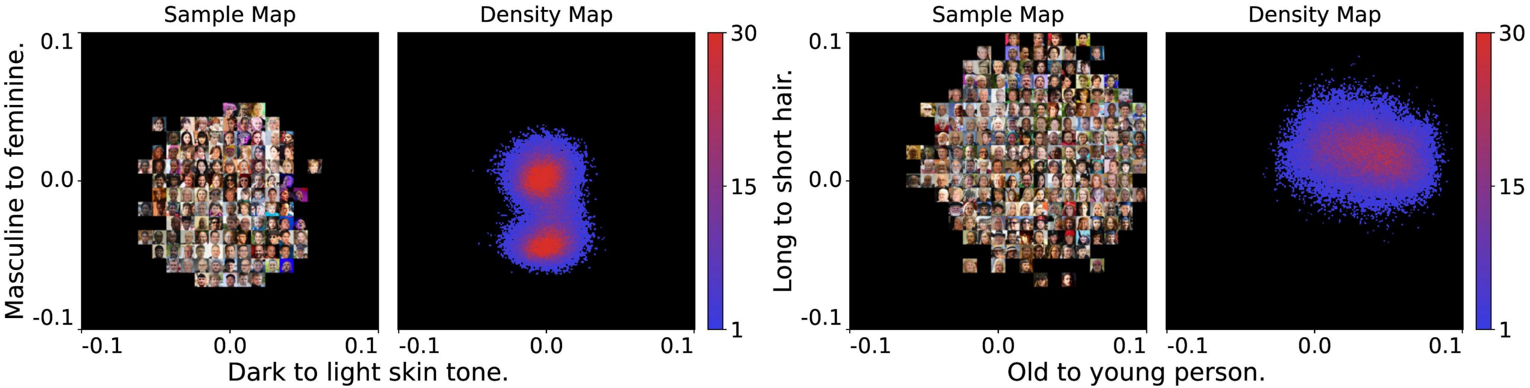
Real	Real
Real + Rand15	Real + Rand50
Real + QD15 (Ours)	**Real + QD50 (Ours)**

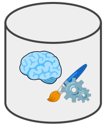
3. Evaluate ⚖️ Fairness, 🎯 Accuracy

Spread w.r.t. CLIP

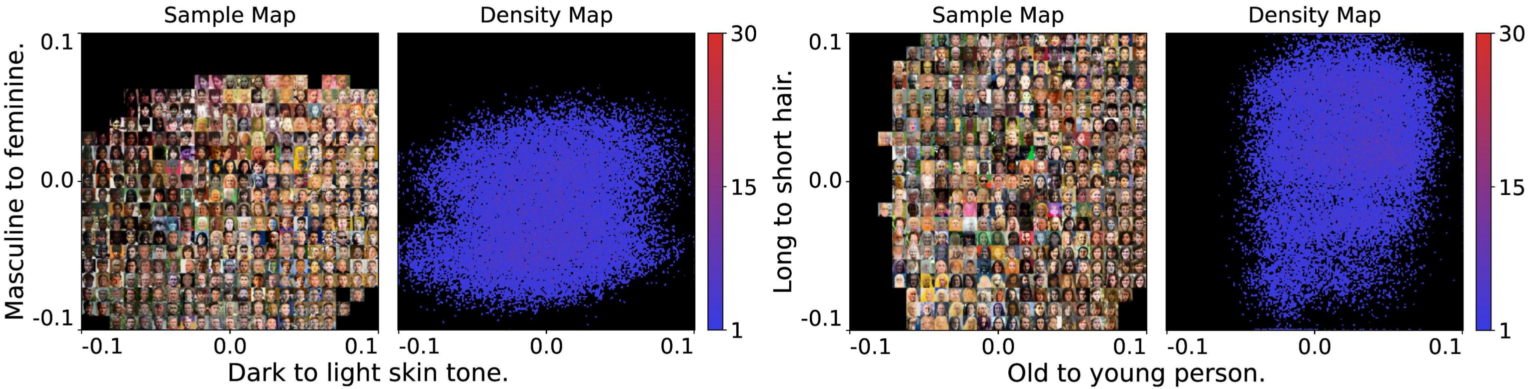


Rand50





QD50



Spread w.r.t. Categorical Labels

	Skin Tone			Age		Gender		
	Light	*Mixed*	*Dark*	*Young*	*Old*	*Masc*	*Andro*	*Fem*
FFHQ Dataset	73.1	17.1	09.8	64.9	35.1	44.2	03.9	51.9
StyleGAN2	69.8	20.8	09.4	63.0	37.0	44.0	04.2	51.8
StyleGAN2 + QDGS	**56.8**	**18.1**	**25.2**	**54.3**	**45.7**	**51.1**	**04.3**	**44.6**

Repairing Biases in Classifiers

Pretraining	Dark-skinned	Light-skinned	DI
None	88.08±0.07	94.05±0.06	93.65±0.10
Rand50	88.69±0.10	**94.67±0.07**	93.68±0.10
QD50	**88.94±0.07**	**94.62±0.10**	**93.99±0.06**

Effects on Accuracy

Pretraining	LFW	CFPFP	CPLFW	CALFW	AgeDB	AVG
None	99.38±0.02	95.14±0.06	90.24±0.09	93.53±0.05	94.33±0.06	94.52±0.03
Rand50	99.45±0.02	95.70±0.04	90.91±0.08	93.60±0.04	94.82±0.05	94.89±0.03
QD50	99.50±0.01	95.72±0.03	90.94±0.06	93.71±0.06	94.72±0.07	94.92±0.02

Limitations of QDGS

→ QDGS aims for uniform representations, rather than proportionate representations

→ Language prompts should be carefully designed to avoid linguistic biases

→ Desired attributes must be represented to a sufficient standard in the data used to train the generative model

Takeaways

→ QDGS uses QD optimization + text prompts to create balanced synthetic training datasets

→ Better fairness improvement, similar accuracy improvement