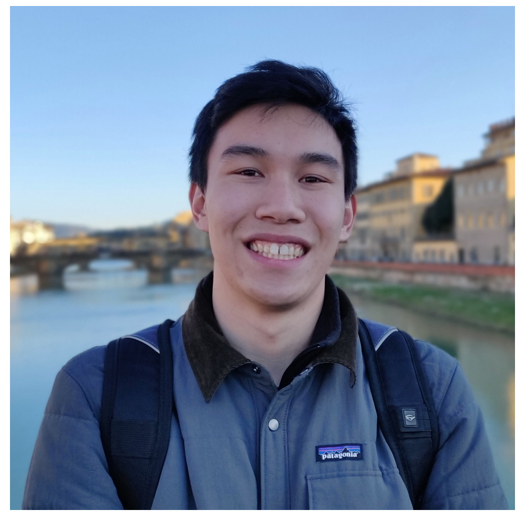
HandsOff: Labeled Dataset Generation with no Additional Human Annotations

CVPR 2023

Highlight Award

TUE-PM-370



Austin Xu
Georgia Tech

Mariya Vasileva
Kaiber

Achal Dave
Toyota Research Institute

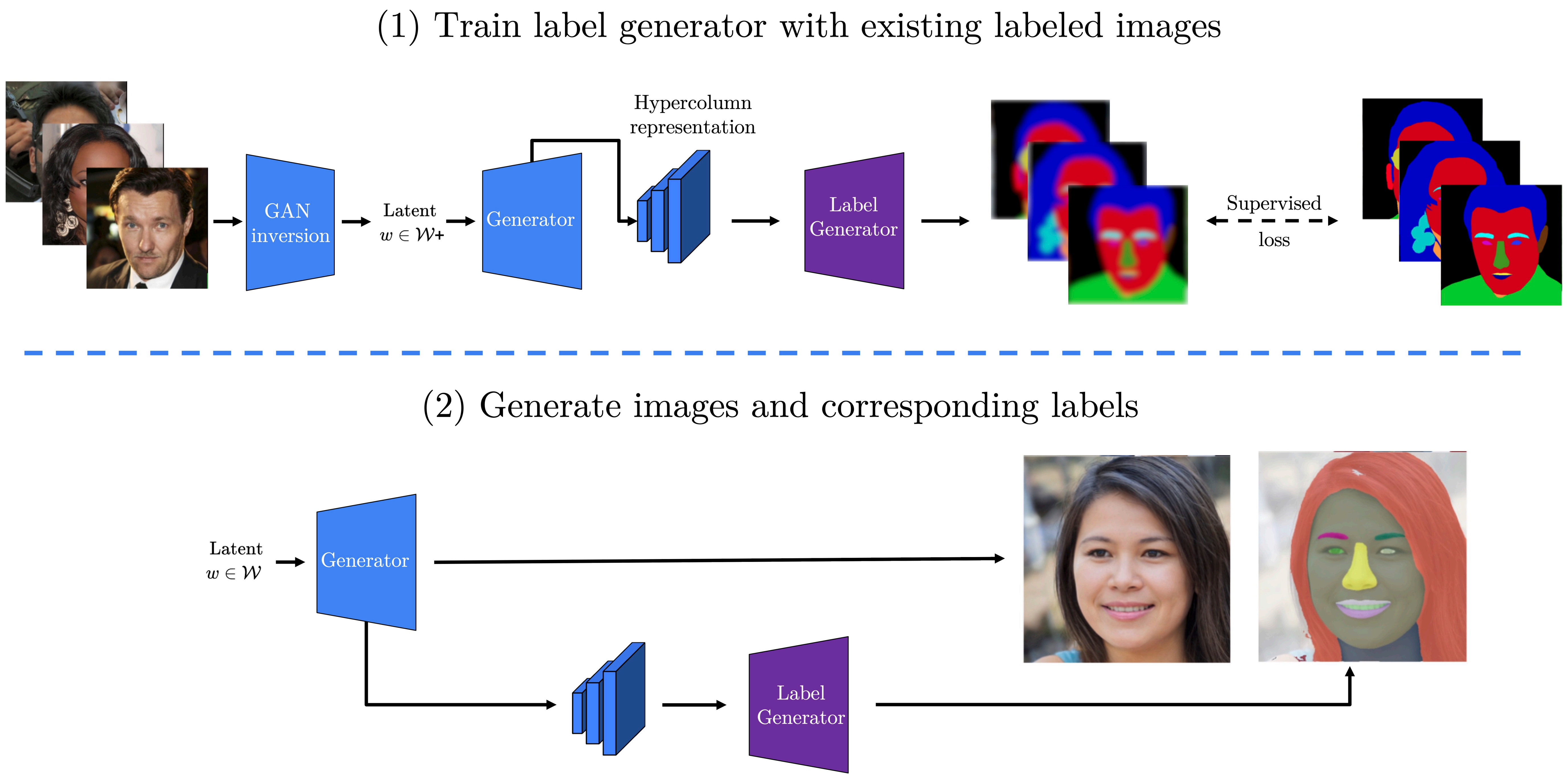
Arjun Seshadri
Amazon Style

Georgia Institute of Technology

amazon

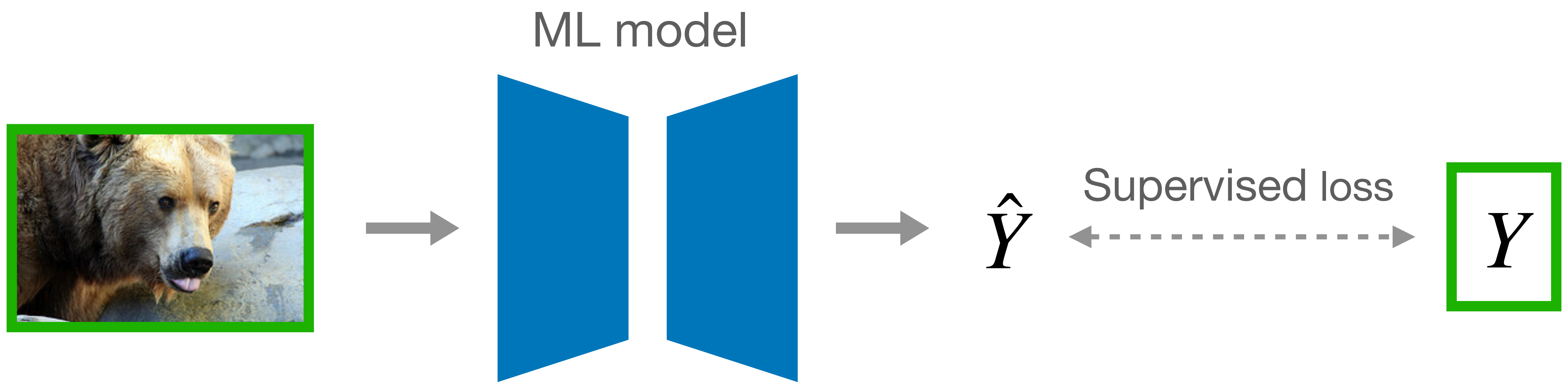
The HandsOff Framework

- GAN-based synthetic dataset generating framework
- Trained with *a small number* of *real images* and their corresponding labels



We need *labeled training data!*

- Labeled training data underpins the success of contemporary machine learning (ML)
- Obtaining labels remains a bottleneck
    - Increasingly complex labels
    - Sheer volume of data



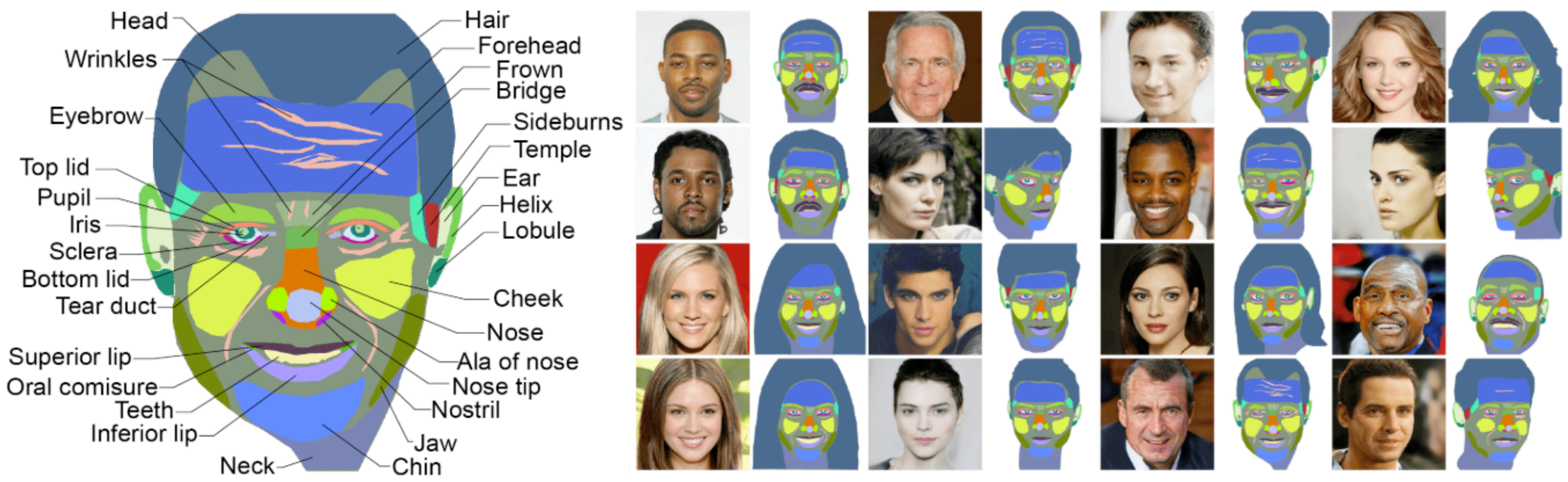
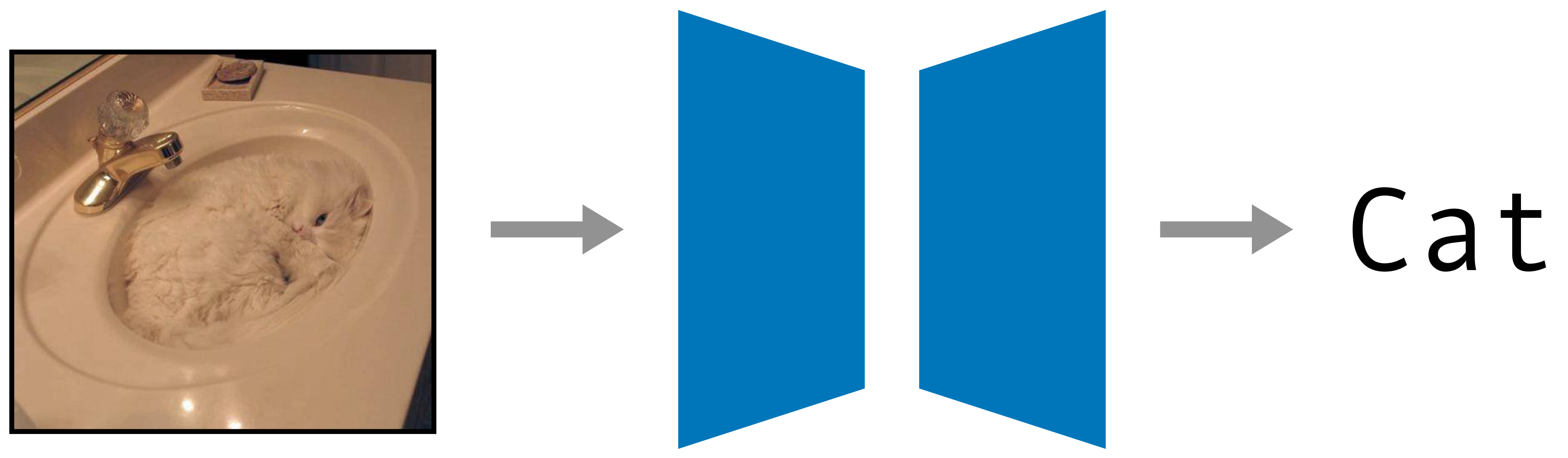


Figure adapted from [1]

[1] Yuxuan Zhang, Huan Ling, Jun Gao, Kangxue Yin, Jean- Francois Lafleche, Adela Barriuso, Antonio Torralba, and Sanja Fidler. DatasetGAN: Efficient Labeled Data Factory with Minimal Human Effort. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.

What if we could have access to an *infinite* pool of labeled data?

- More *generalizable* and *accurate* models
  - The more data, the merrier! Helps close the generalization gap
  - Access to *labeled examples of edge cases* can help us tame the long-tail.



The DatasetGAN paradigm

DatasetGAN [1] trains a *label generator* from a small number of *labeled GAN generated images* (<50)

- Starting from a latent code, generate
  - An image, which is then manually annotated with pixel-wise labels, Y
  - A pixel-wise image representation S
- Then, train a *label generator* with the (S, Y) pairs

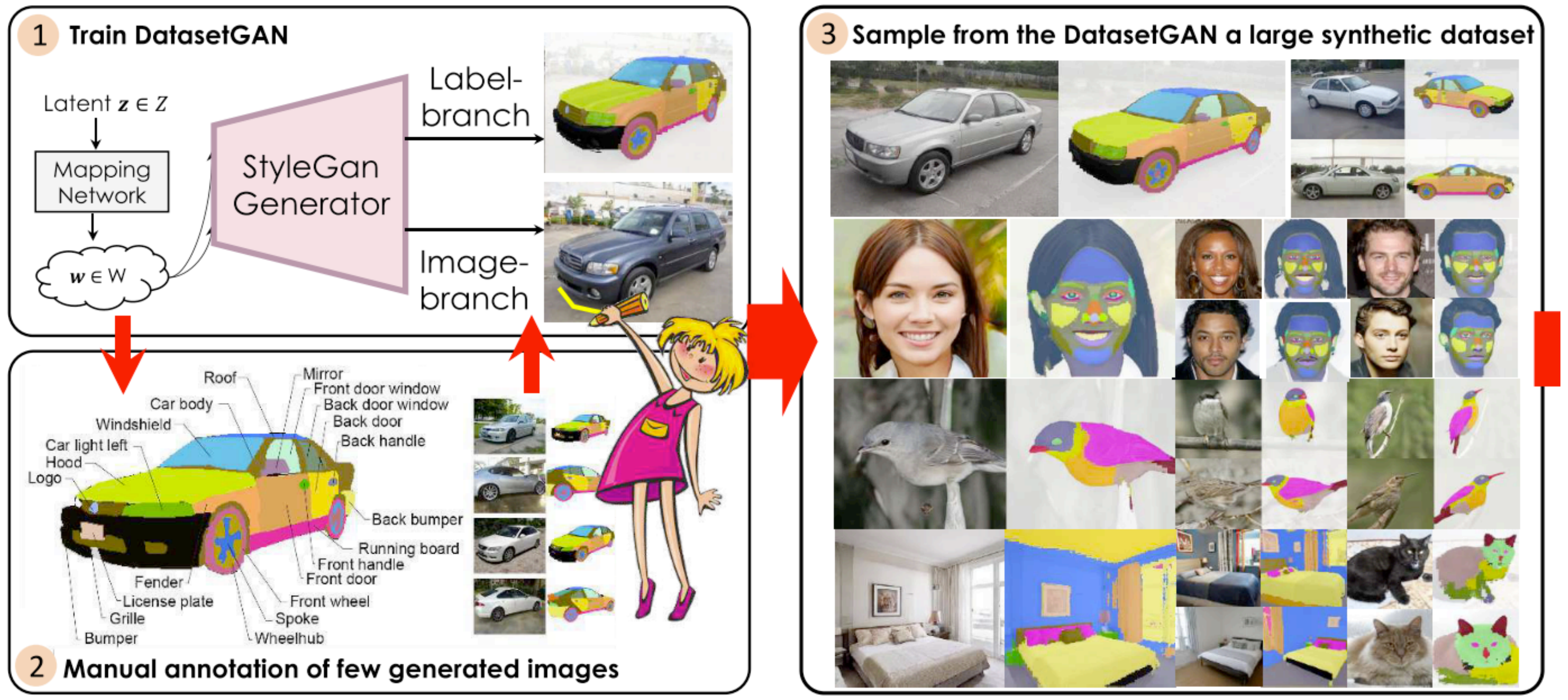


Figure adapted from [3]

[1] Yuxuan Zhang, Huan Ling, Jun Gao, Kangxue Yin, Jean- Francois Lafleche, Adela Barriuso, Antonio Torralba, and Sanja Fidler. DatasetGAN: Efficient Labeled Data Factory with Minimal Human Effort. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.

The DatasetGAN paradigm: drawbacks?

DatasetGAN relies on *manual annotations* …

- Limits types of labels that can be synthesized
- Practical drawbacks: labeling infrastructure, start-up costs, etc.

… of GAN generated images

- Does not allow for manual curation of training data
- Training on labeled GAN images results in lower quality labels [2]

[2] Dmitry Baranchuk, Andrey Voynov, Ivan Rubachev, Valentin Khrulkov, and Artem Babenko. Label-Efficient Semantic Segmentation with Diffusion Models. In *International Conference on Learning Representations (ICLR)*, 2022.

Can we overcome the two drawbacks?

DatasetGAN relies on *manual annotations* …

- Limits types of labels that can be synthesized
- Practical drawbacks: labeling infrastructure, start-up costs, etc.

… of *GAN generated images*

- Does not allow for manual curation of training data
- Training on labeled GAN images results in lower quality labels [2]

We want a method that uses *existing annotations* …

- Enables more complex label generation (e.g., continuous values)
- No associated start up cost for labeling infrastructure

… of *real images*

- Unlocks a larger pool of candidates for training images
- Higher quality training images leads to better labels

[2] Dmitry Baranchuk, Andrey Voynov, Ivan Rubachev, Valentin Khrulkov, and Artem Babenko. Label-Efficient Semantic Segmentation with Diffusion Models. In *International Conference on Learning Representations (ICLR)*, 2022.

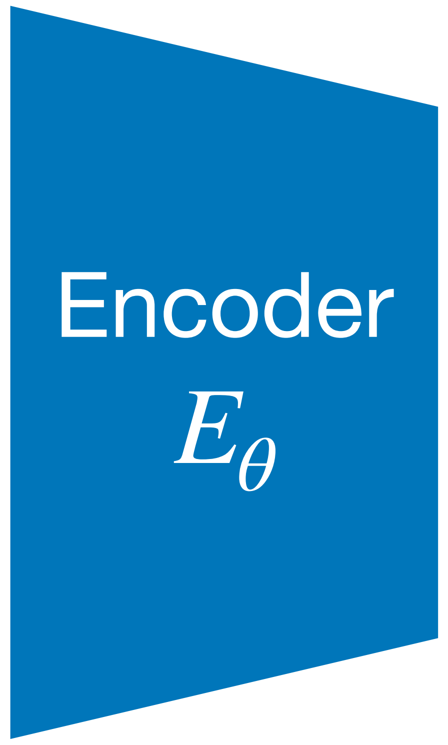
GAN inversion: connecting *real labeled images* to dataset generation

Given a pre-trained generator G and a similarity-based loss $\mathscr{L}$ (e.g., LPIPS)

Encoder-based

Train an encoder to minimize similarity-based loss

X



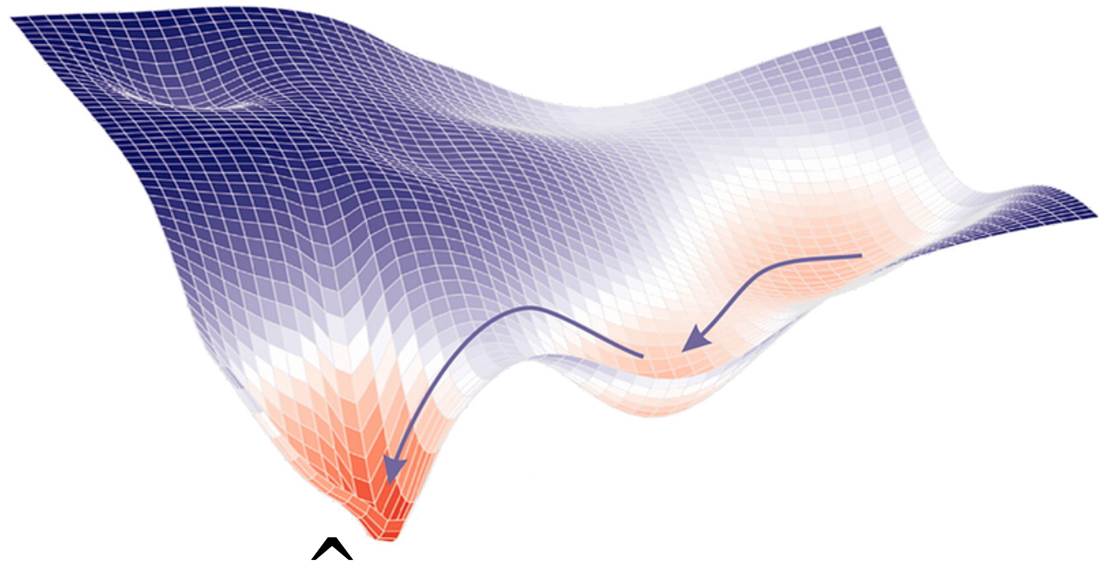
$\rightarrow \hat{w}$

$$\arg\min_{\theta} \mathscr{L}\left(E_\theta(X), G\left(E_\theta(X)\right)\right)$$

Optimization-based

Directly optimize similarity-based loss

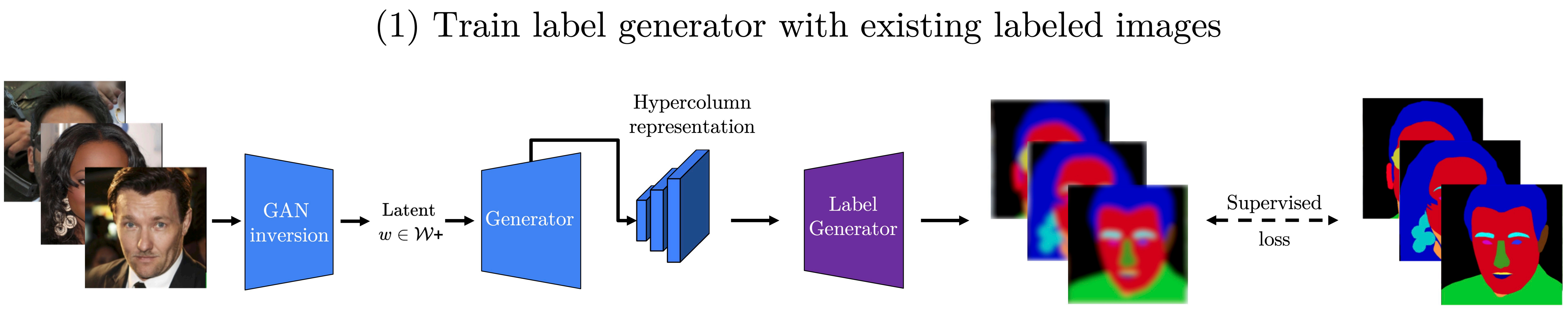
X

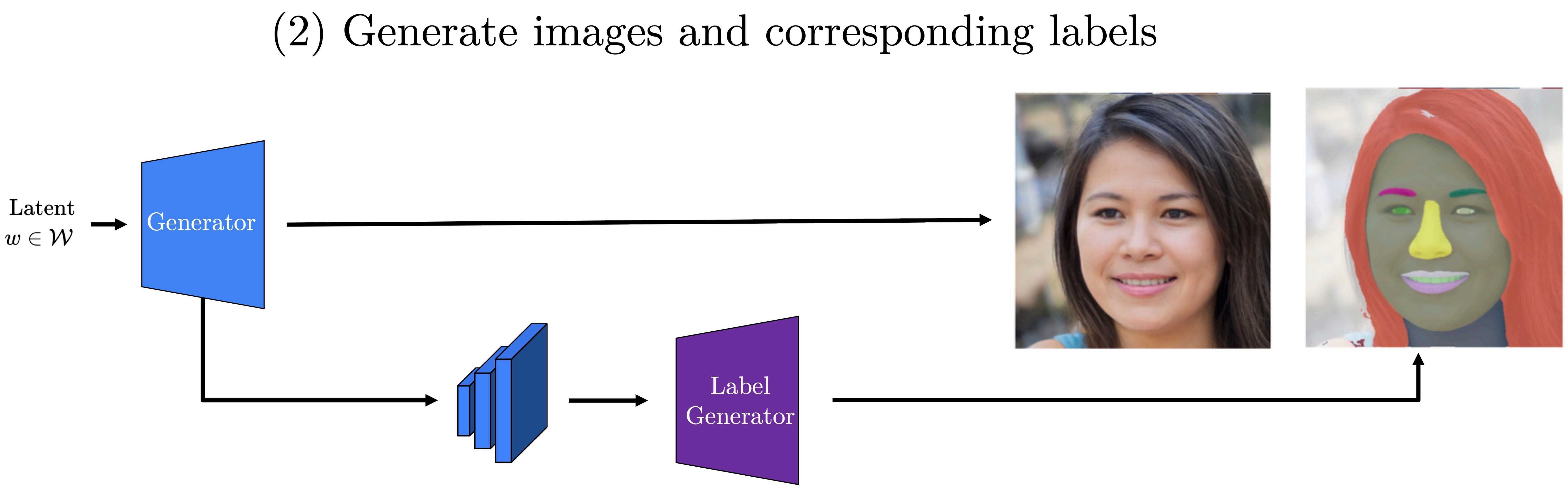


$\hat{w}$

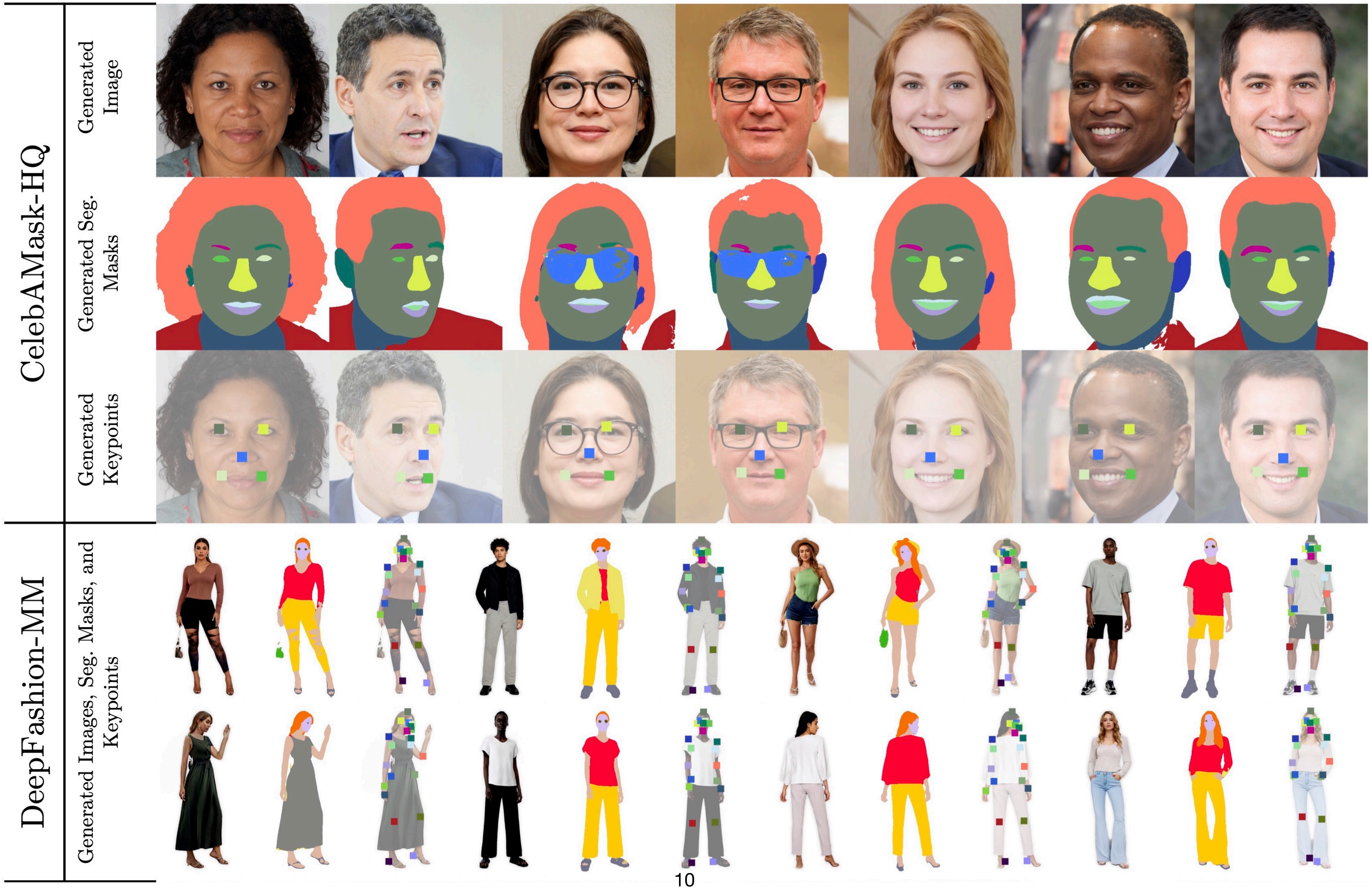
$$\arg\min_{w} \mathscr{L}\left(X, G(w)\right)$$

The HandsOff Framework

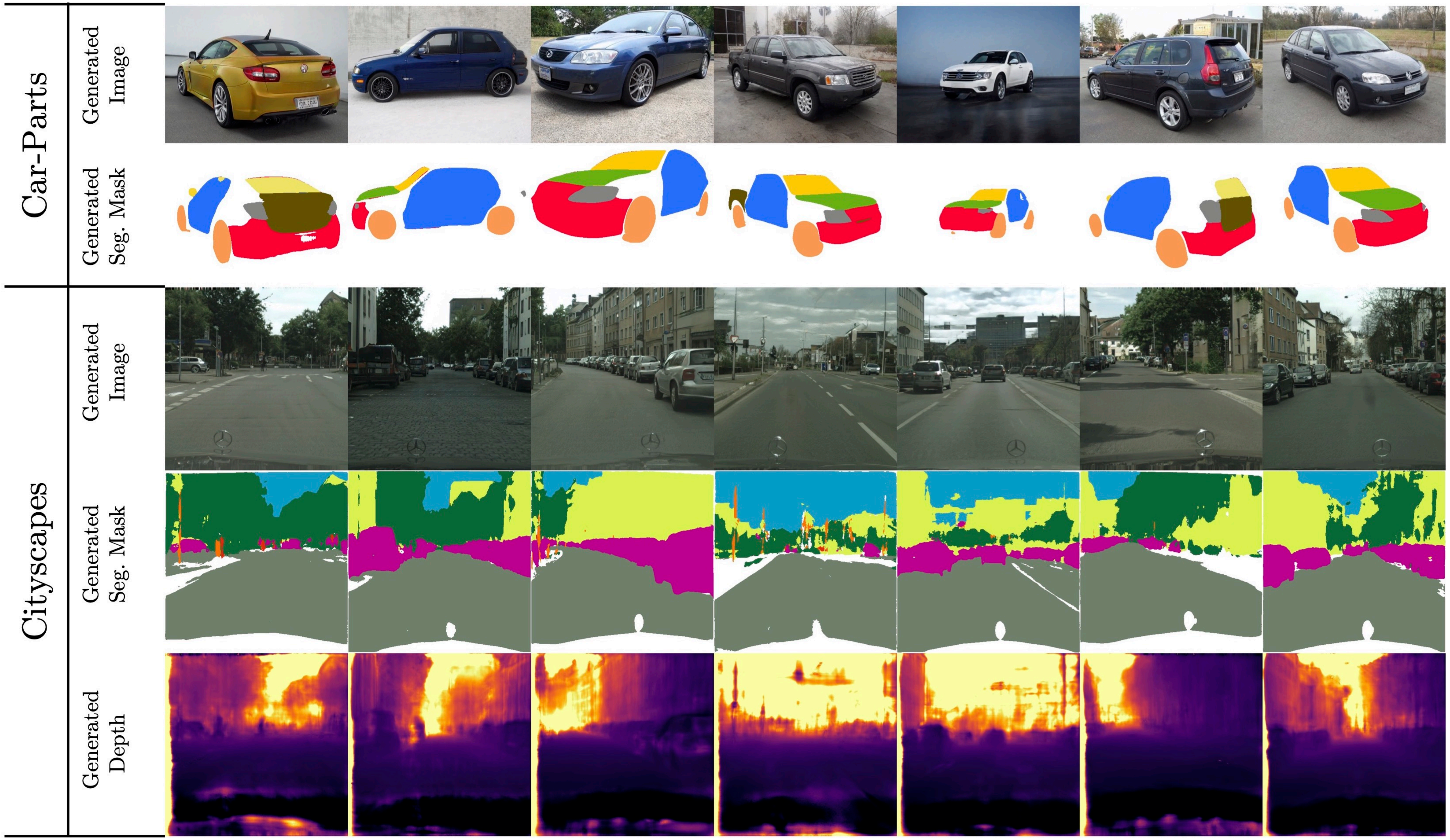




Synthesized Labels



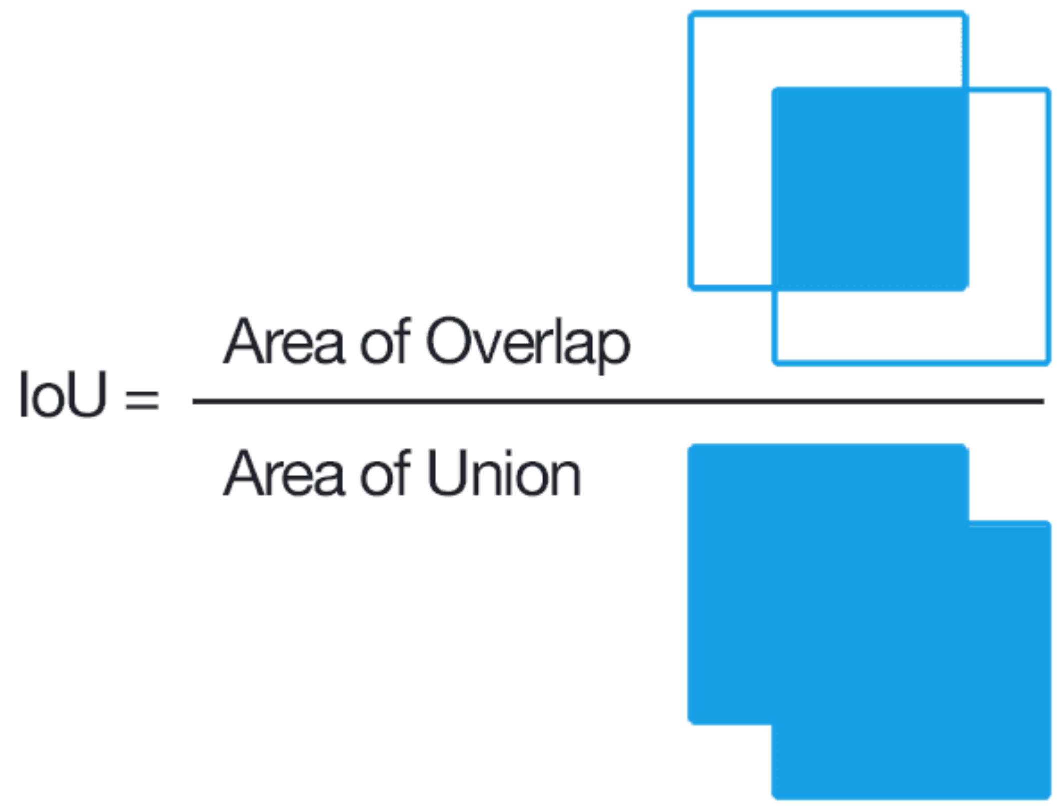
Synthesized Labels



How do we evaluate the quality of our synthetic data?

A numerical measure of dataset quality:

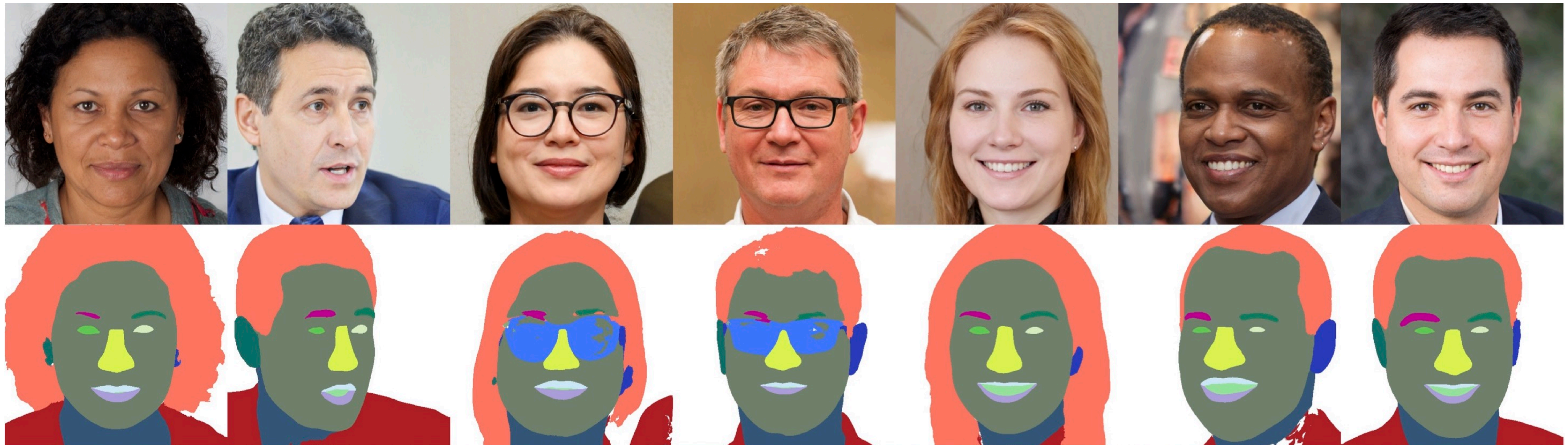
- Train a **downstream network** on the synthesized data
- Evaluate trained downstream network on **real labeled test images**
  - Segmentation: mean Intersection-over-Union (mIOU, ↑)
  - Keypoints: Percentage of Correct Keypoints (PCK, ↑)
  - Depth: Mean Squared Error (MSE, ↓)



Segmentation Mask Quality

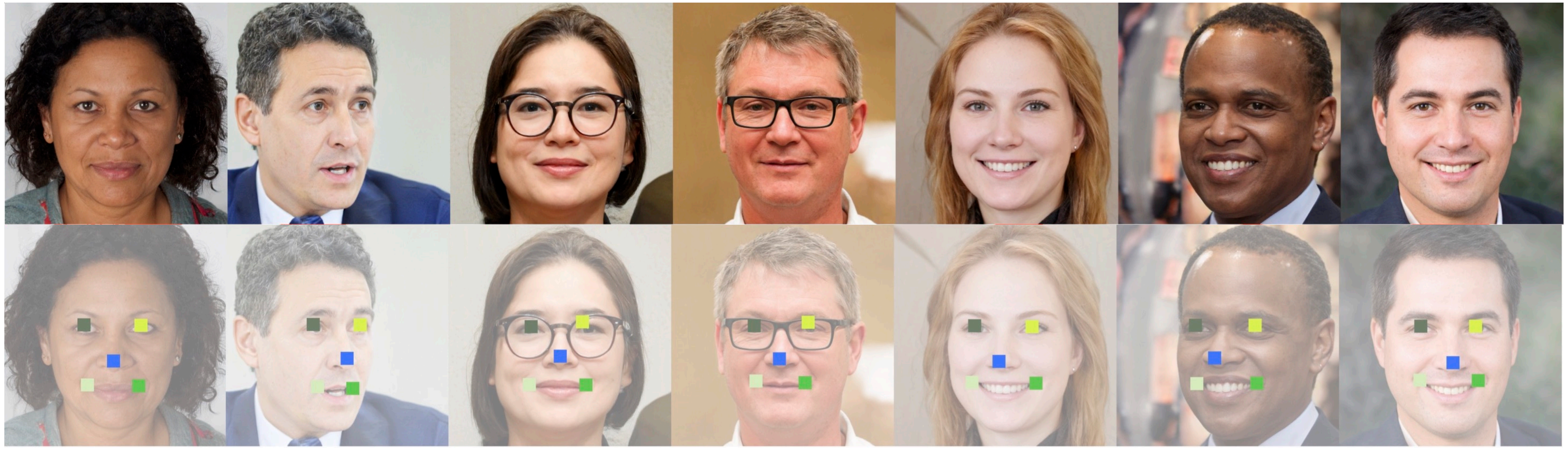
Segmentation performance measured in mIOU

	# labeled images	CelebAMask-HQ 8 classes	Car-Parts 10 train	DeepFashion-MM 8 classes	DeepFashion-MM 10 classes	Cityscapes 8 classes
DatasetGAN	16	0.7013	×	×	×	×
EditGAN	16	0.7244	0.6023	×	×	×
Transfer Learning	16	0.4575	0.3232	0.5192	0.4564	0.4954
HandsOff (Ours)	16	**0.7814**	**0.6222**	**0.6094**	**0.4989**	**0.5510**
Transfer Learning	50	0.6197	0.4802	0.6213	0.5559	0.5745
HandsOff (Ours)	50	**0.7859**	**0.6679**	**0.6840**	**0.5565**	**0.6047**



Keypoint and Depth Map Quality

	# labeled images	CelebAMask-HQ PCK-0.1 ↑	CelebAMask-HQ PCK-0.05 ↑	CelebAMask-HQ PCK-0.02 ↑	DeepFashion-MM PCK-0.1 ↑	DeepFashion-MM PCK-0.05 ↑	DeepFashion-MM PCK-0.02 ↑	Cityscapes-Depth mNMSE ↓	Cityscapes-Depth RMSE ↓	Cityscapes-Depth RMSE-log ↓
Transfer Learning	16	78.96	42.06	7.32	91.24	83.52	48.21	0.4022	18.12	2.75
HandsOff (Ours)	16	**97.19**	**76.36**	**17.44**	**94.19**	**88.48**	**70.22**	**0.2553**	**14.52**	**1.64**
Transfer Learning	50	90.88	61.75	12.30	91.24	83.52	48.20	0.2525	15.07	3.01
HandsOff (Ours)	50	**97.71**	**79.99**	**19.10**	**95.41**	**90.89**	**74.02**	**0.1967**	**13.01**	**1.58**



Thank you! Questions?

- Contact: axu@gatech.edu
- Website: austinxu87.github.io/handsoff