

# **Fake It Till You Make It**

## **Synthetic Data in Machine Learning**

Sebastian Dziadzio · GHOST Day Applied ML Conference · 21.04.2023



# Motivation



# Real data

Costly to collect

Legal & ethical issues

Limited diversity

Limited scale and quality

Difficult annotation

Task specific



# Synthetic data

Costly to build

Few legal & ethical issues

Controlled diversity

Large scale

Perfect annotation

Adaptable



# Domain gap

Refinement

Regularisation

Domain agnostic features

Domain randomisation

Photorealism at scale



# Domain gap

Refinement

Regularisation

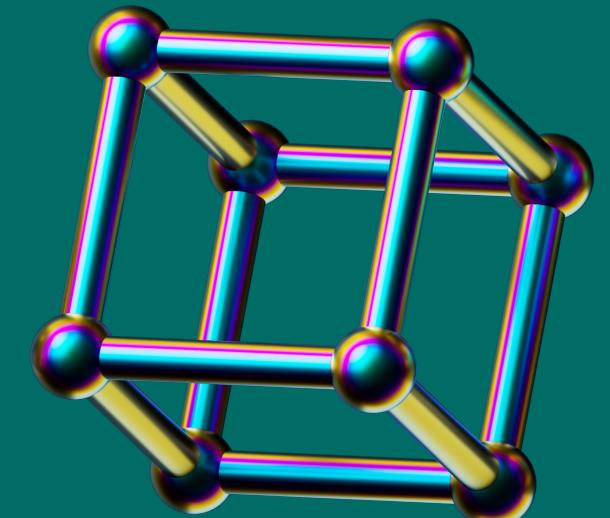
Domain agnostic features

Domain randomisation

Photorealism at scale

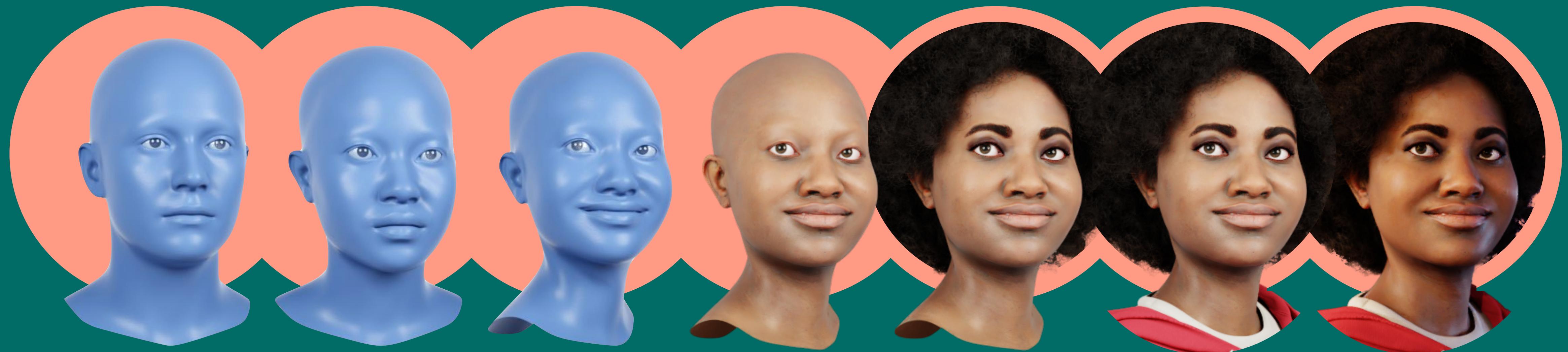


# Synthetic Faces





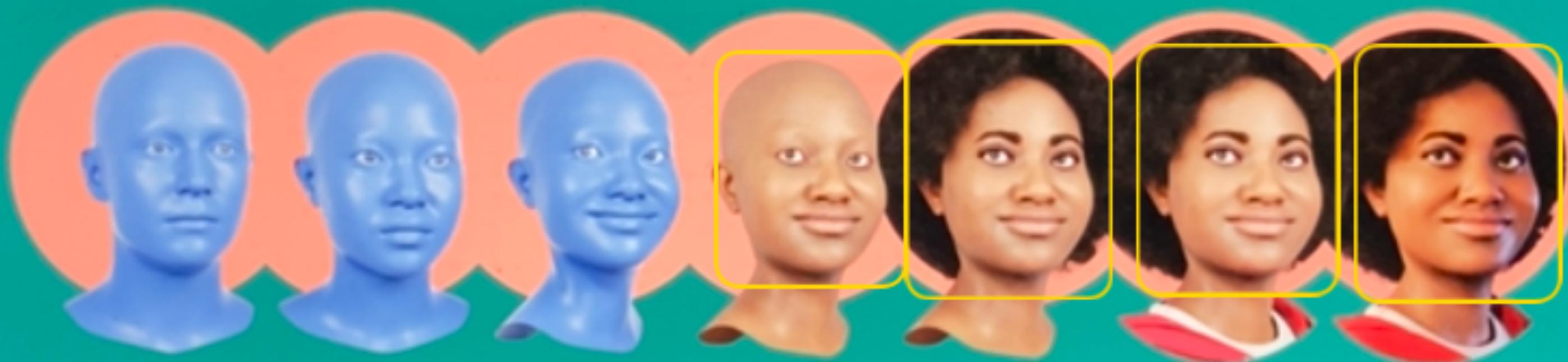
# The pipeline



Wood et al. 2021, Fake it till you make it: face analysis in the wild using synthetic data alone

Wood et al. 2021, Fake it till you make it: face analysis in the wild using synthetic data alone

## The pipeline



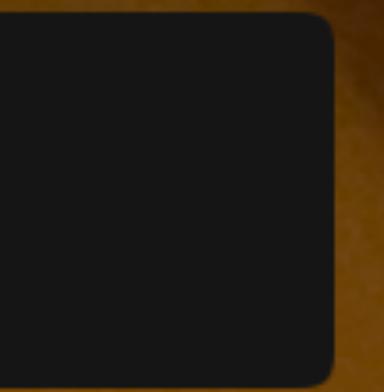
VIDEO

SLO-MO

PHOTO

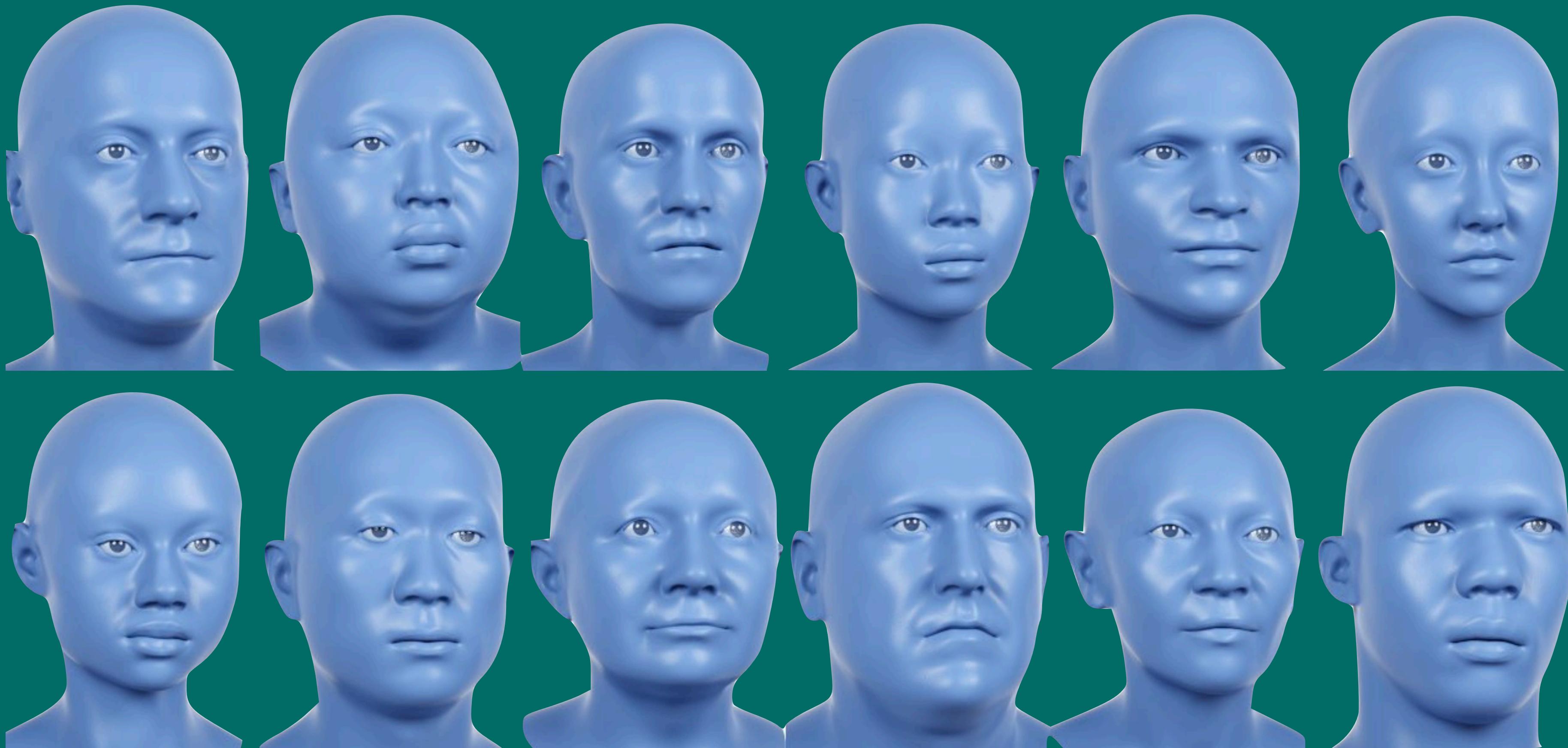
PORTRAIT

PANO



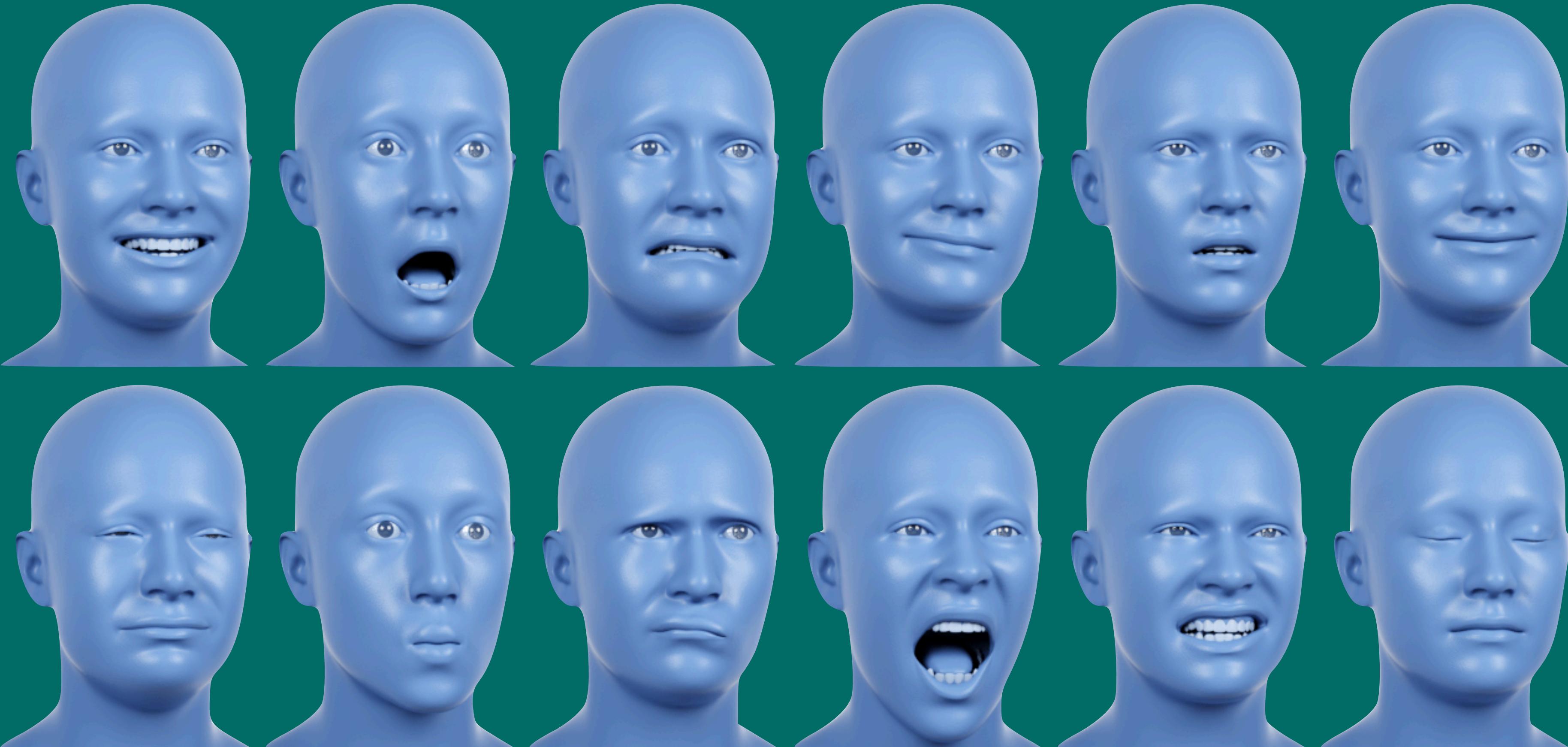
MacBook Pro

# Identity model



Wood et al. 2021, Fake It Till You Make It: Face Analysis in the Wild Using Synthetic Data Alone.

# Expression model



Wood et al. 2021, Fake It Till You Make It: Face Analysis in the Wild Using Synthetic Data Alone.

# Hair and clothing



# Lighting



Wood et al. 2021, Fake It Till You Make It: Face Analysis in the Wild Using Synthetic Data Alone.

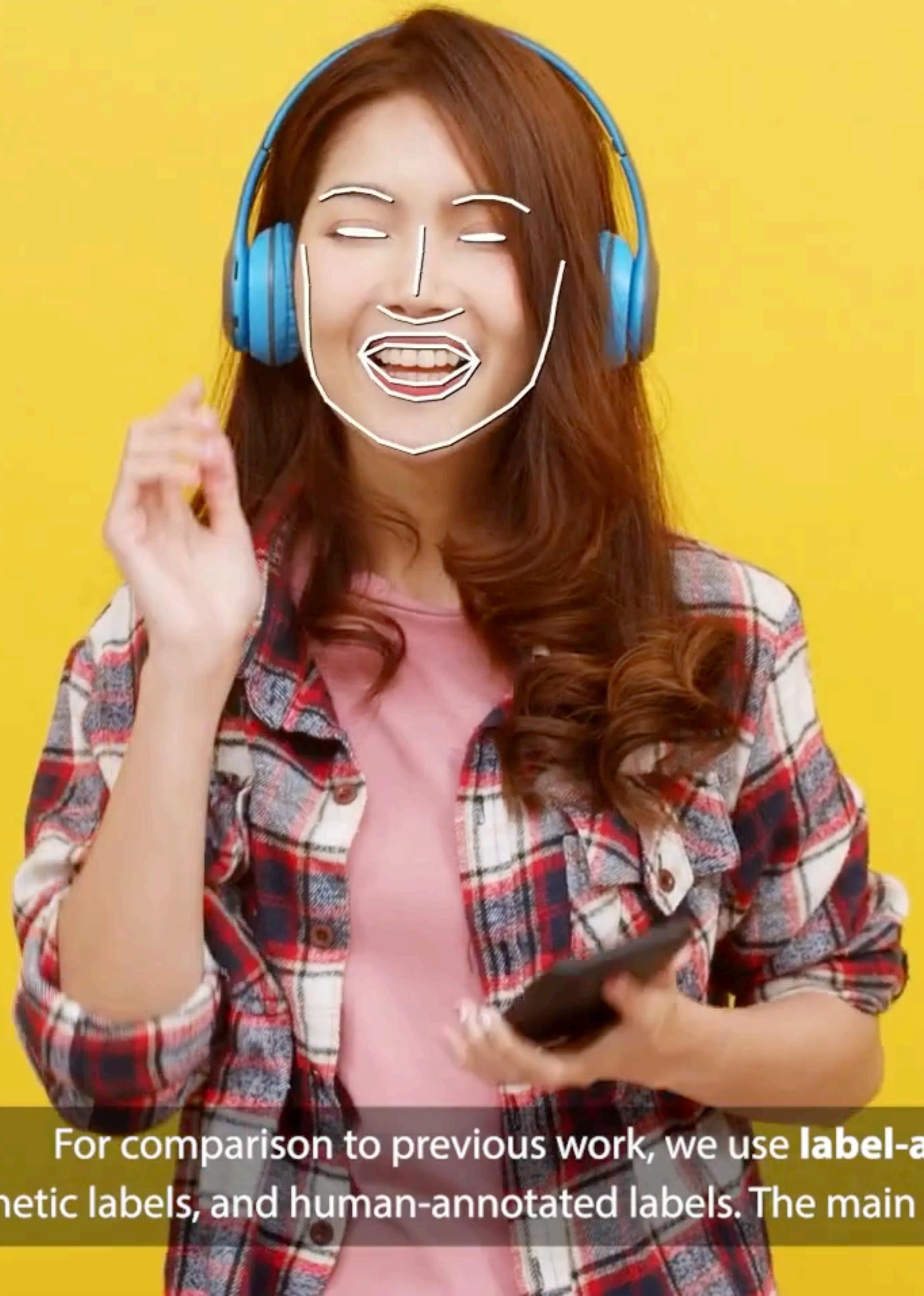


We render a training dataset of 100,000 synthetic face images. Here are some examples.



Here is an example of Face Parsing, trained with synthetic data alone.

Trained with synthetic data only



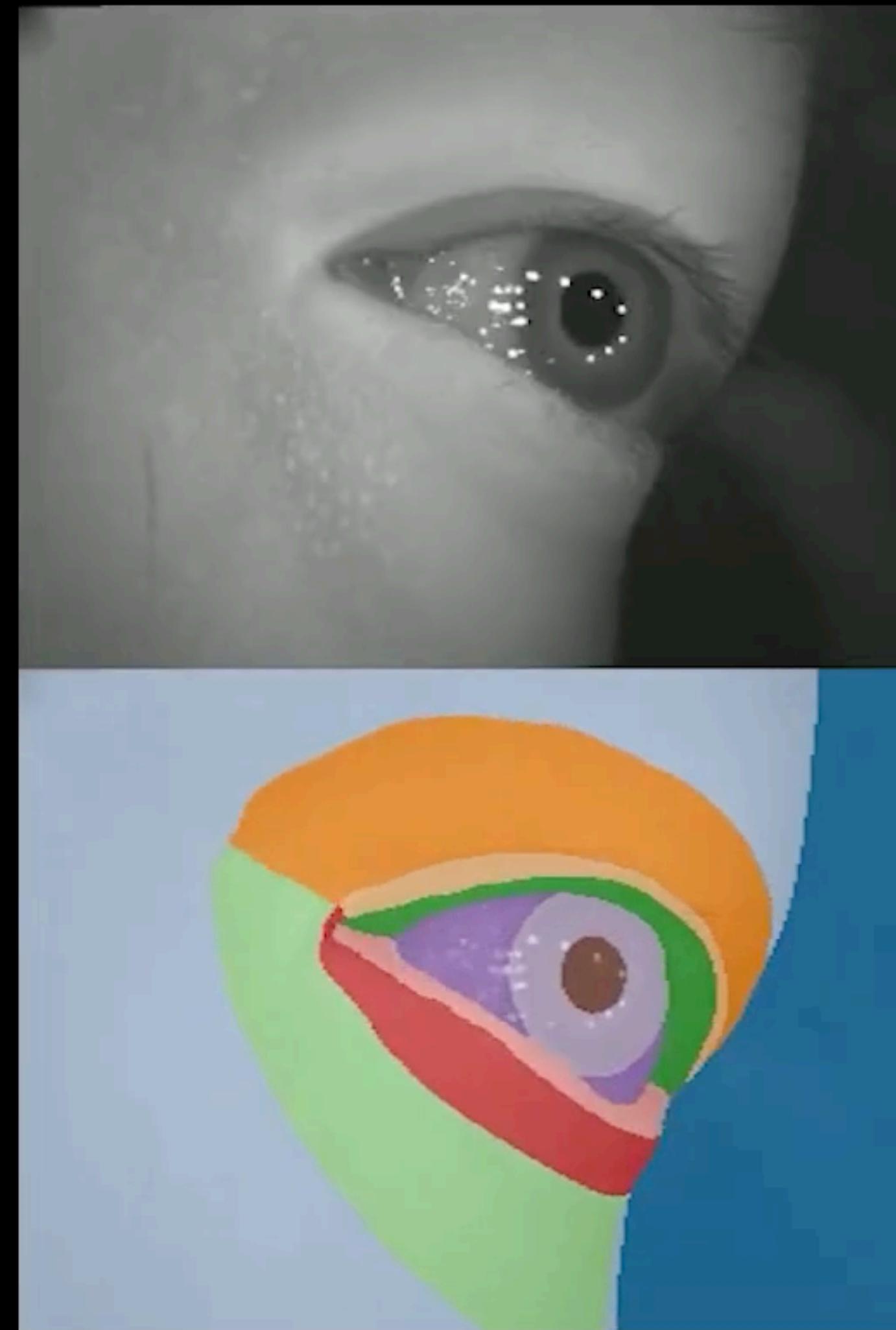
+ label adaptation



For comparison to previous work, we use **label-adaptation** to minimize the systematic differences between our synthetic labels, and human-annotated labels. The main difference is converting our 3D projected jawline into a 2D face outline.

Results on real eye tracking images,  
**trained with synthetic data only**

Synthetic eye tracking images

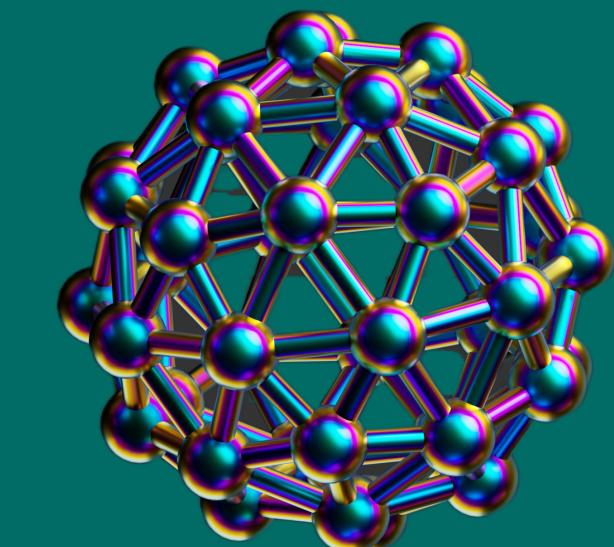


Eye tracking can be a key feature for virtual or augmented reality, but training data is difficult to acquire. Our faces look realistic close-up, so it is easy to set up a synthetic eye tracking camera and render training images..

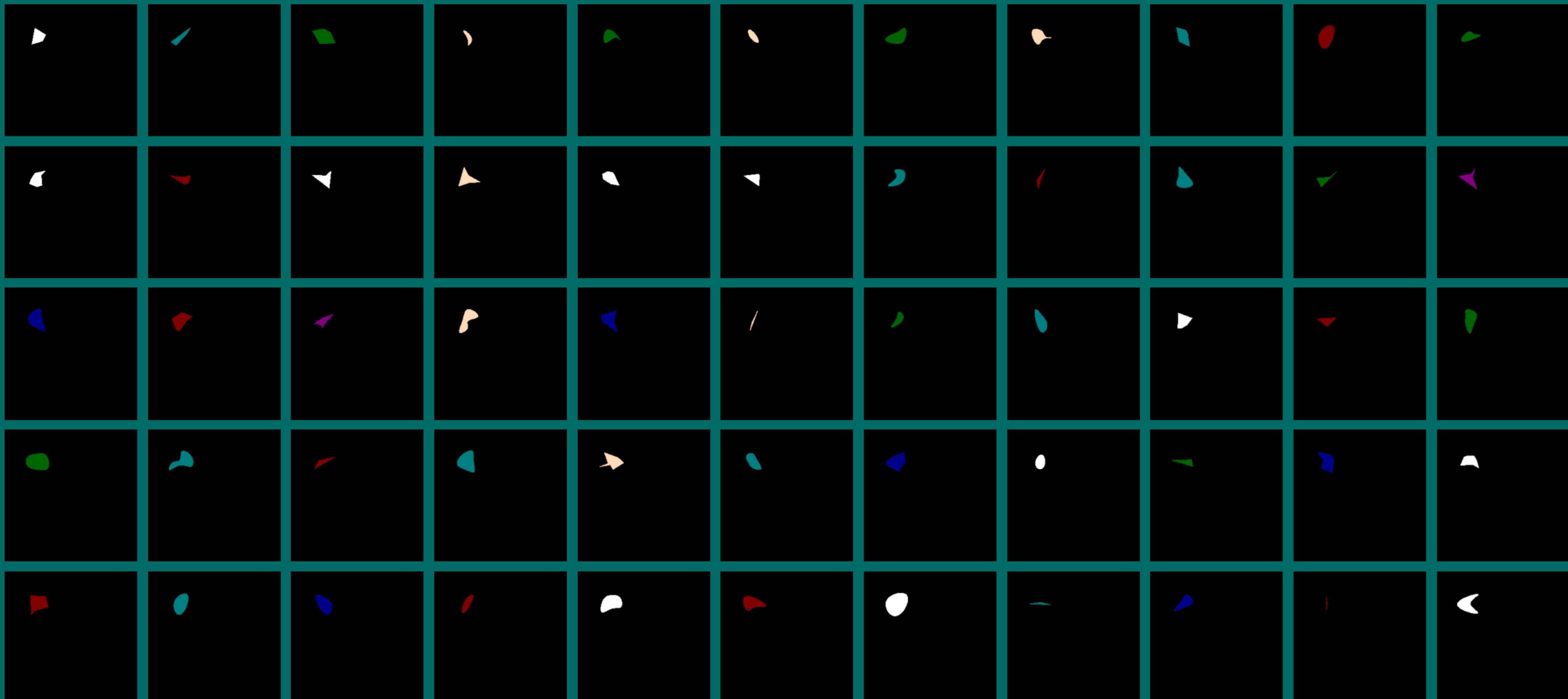


Here are some results of predicting ten times as many landmarks as usual, using synthetic training data.

# Other Datasets



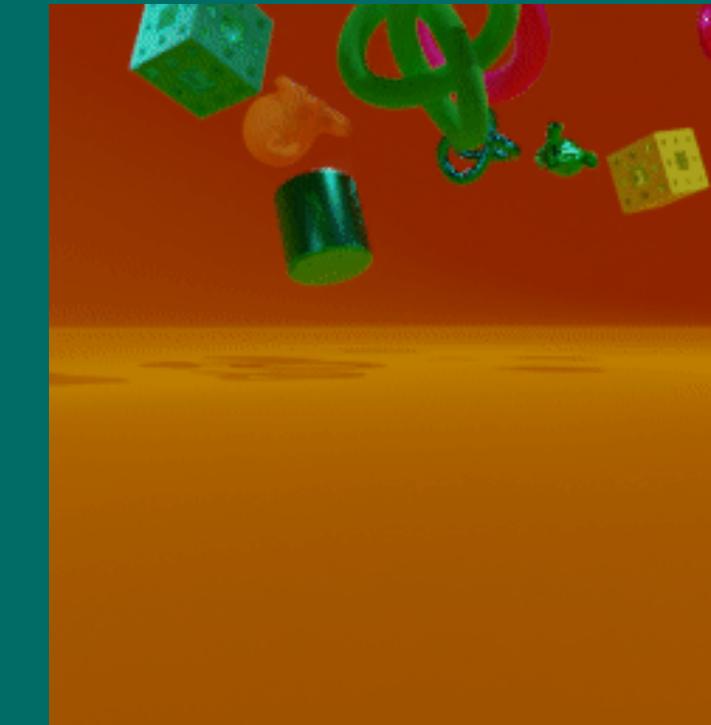
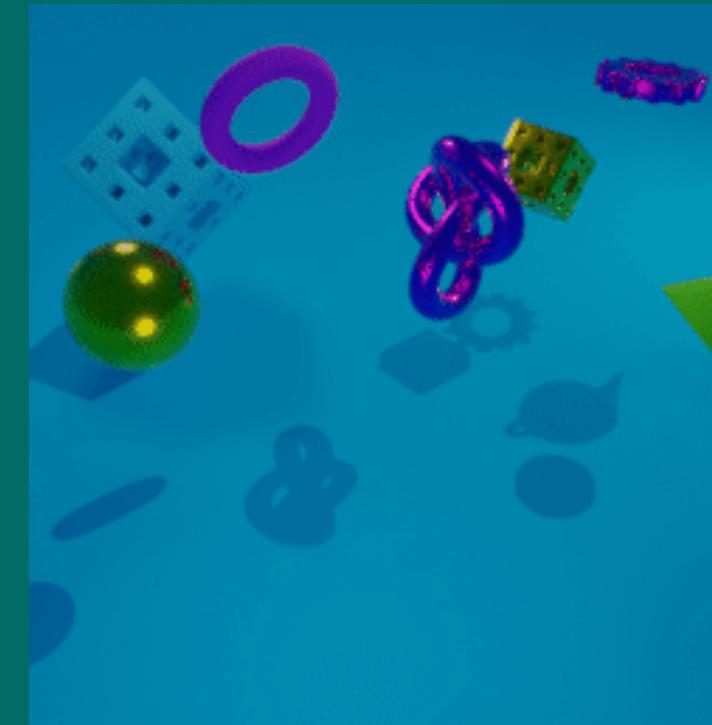
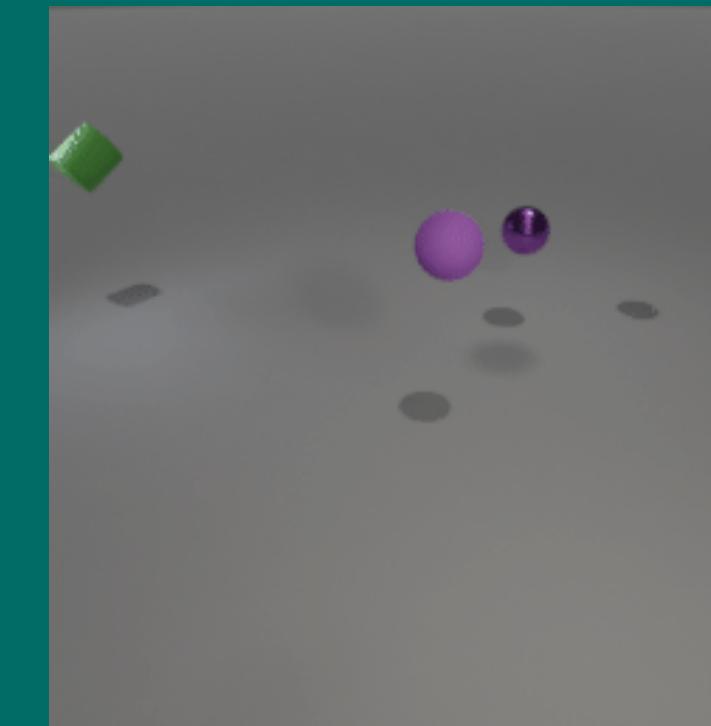
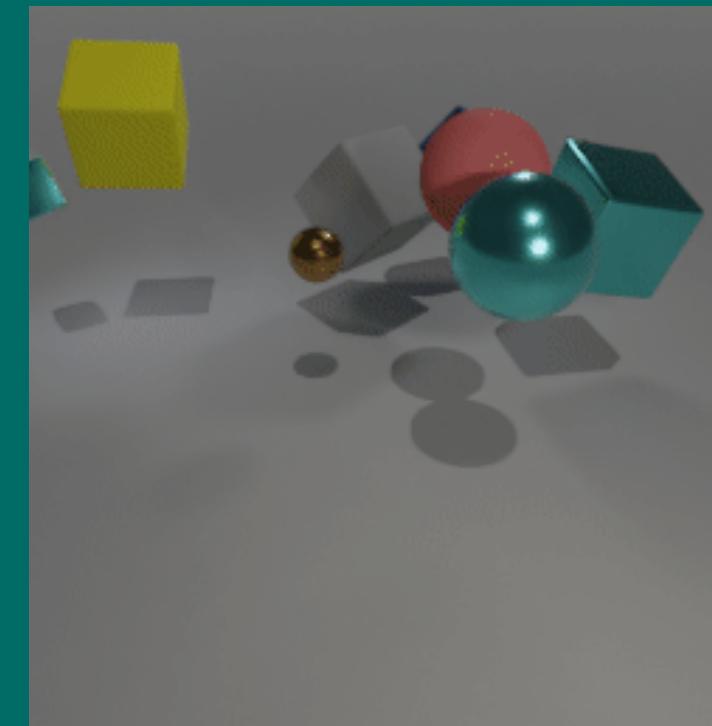
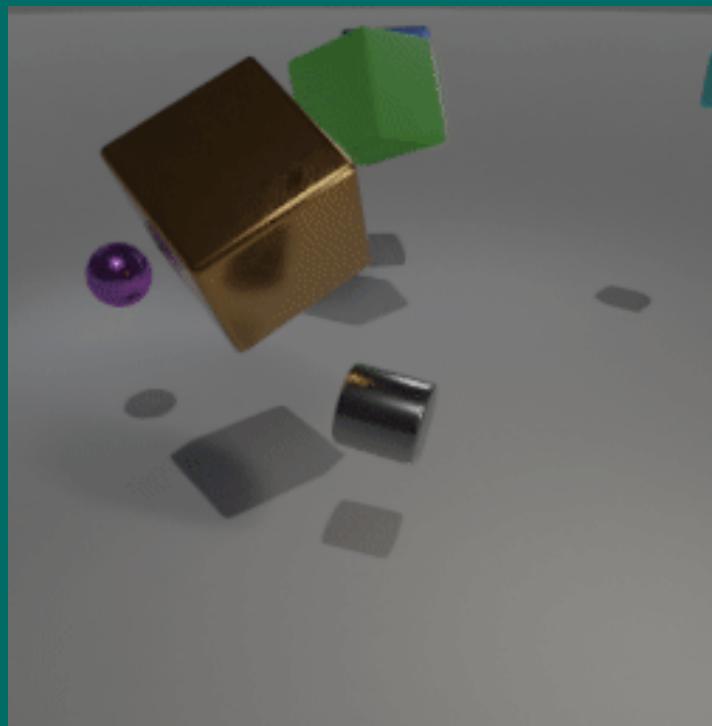
# Continual Disentanglement



# Autonomous driving



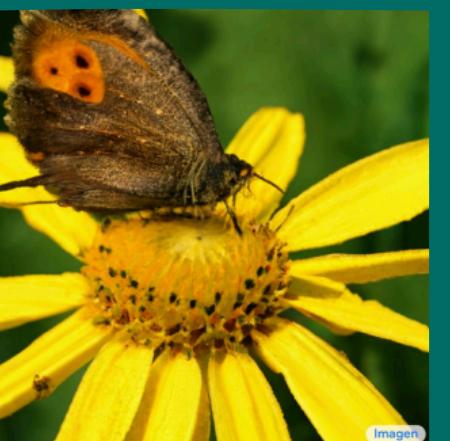
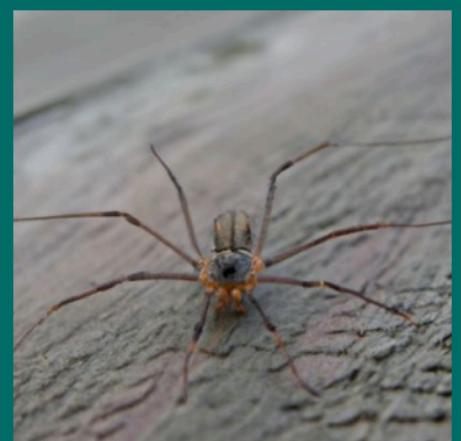
# Object tracking



# Image classification



bittern bird



harvestman



leonberger



schipperke

# Tabular Data

☰ README.md

🔗



**synthcity**

A library for generating and evaluating synthetic tabular data..

 Open in Colab  Tests Fast Python passing  Tests Full Python failing  Tutorials passing  docs passing arXiv 2301.07573

# Thank you!

 @sbdzdz

