InsetGAN for Full-Body Image Generation

Anna Frühstück, Krishna Kumar Singh, Eli Shechtman, Niloy J.Mitra, Peter Wonka, Jingwan Lu

جامعة الملك عبدالله للعلوم والتقنية
King Abdullah University of Science and Technology

Adobe Research

UCL

CVPR JUNE 19-24 2022
NEW ORLEANS • LOUISIANA

Synthesis of Full-Body Humans

When synthesizing human images at high-resolution, StyleGAN2 can generate **plausible bodies** but is **unable to capture fine details** like high-quality faces, shoes, or accessories.

A variety of synthesized humans using a vanilla StyleGAN2 generator. Most results on this poster are shown cropped due to space constraints.

To address these artifacts, InsetGAN achieves superior quality by utilizing **specialized, independently trained generators** to **improve important regions** (e.g. faces or shoes) of a canvas generator (e.g. human body) through **joint latent space exploration**.

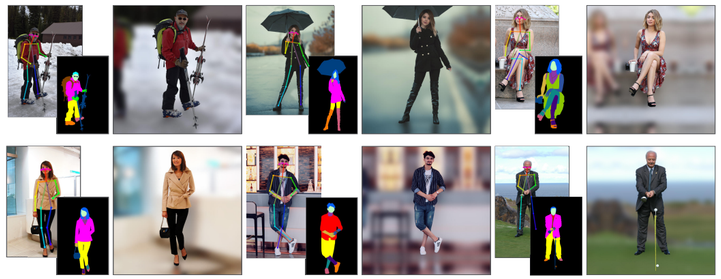
Our approach achieves coherent and seamlessly merged output images of full-body humans, leveraging the power of each individual generator.

To generate a seamlessly merged canvas and inset, we define the inset region $\mathcal{B}$ in the canvas and **set of losses** that ensure **high-level content coherence** L_{coarse} as well as **edge correspondence** L_{border}. Depending on the application, we may use additional constraints and regularizers. We optimize both latent spaces to generate our improved output images.

Faces generated using a specialized face generator trained on face crops of the same dataset. Note the higher quality and diversity of the faces.

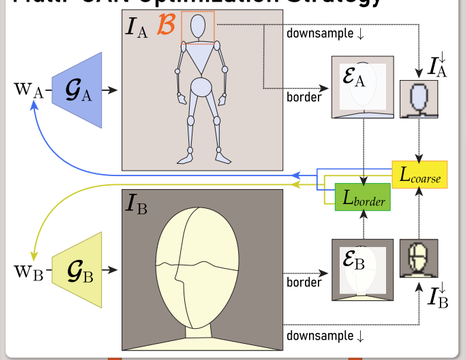
Dataset Curation

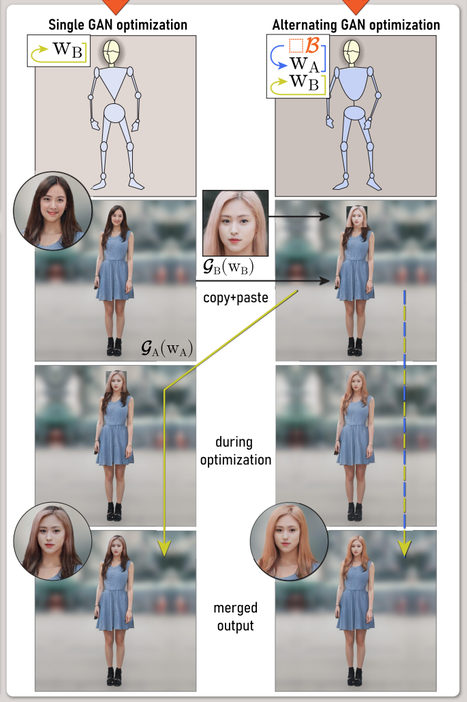
The **quality, diversity and alignment of the dataset** plays a crucial role in the ability of the generator to capture challenging domains such as the human body from head to toe.
We trained our StyleGAN2 network on a dataset from images of humans in the wild in a multitude of poses and items of clothing. Each human subject is aligned within their picture based on their upper body axis and the background is heavily blurred, yielding a dataset of around 83K images at 1024×1024px.



Humans within the source photographs are aligned and cropped based on their pose. The background remains for context, but is heavily blurred.

Multi-GAN Optimization Strategy



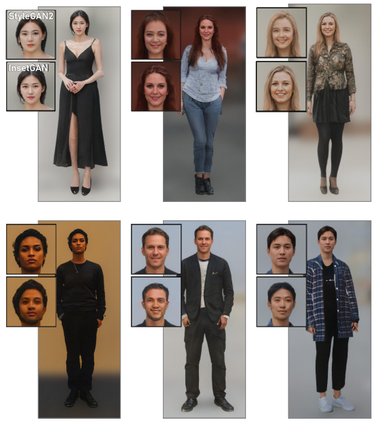


Results

We showcase a variety of different applications of the InsetGAN method. Our technique has the potential to be generalized to other domains where specialized generators can improve challenging image regions within a target image.

GAN Artifact Removal

InsetGAN can **improve our generated humans' faces** while preserving each person's identity.



Multimodal Conditional Generation

Given an **input face**, we can synthesize **several plausible bodies** with different poses and clothing. The input identity is faithfully preserved.

We can also generate **multimodal faces for a given input body**, if the face coherence L_{coarse} is enforced at a low resolution. A generator trained on the **DeepFashion** [1] dataset synthesized the bodies in these results. The faces are generated from the FFHQ dataset.

[1] Ziwei Liu et al.: DeepFashion: Powering Robust Clothes Recognition and Retrieval with Rich Annotations, CVPR 2016

Face/Body Composition

Joint optimization of three latent spaces: body, face, and shoes

Our approach is able to **seamlessly unite multiple specialized generators by jointly optimizing all their latent spaces**. Here, we simultaneously improve face and shoe areas of these generated human images by using a face and shoe generator independently trained on crops from the same dataset as the human bodies. The generated bodies, faces, and shoes look plausible and are merged without visible transitions.

Evaluation

FID Score ▽

	t=0.7 BODY	t=0.7 FACE	t=0.4 BODY	t=0.4 FACE
REFERENCE	26.67	**27.14**	71.90	66.61
INSETGAN	**25.33**	31.61	**69.58**	**61.57**

Precision & Recall △

	BODY PRECISION	BODY RECALL	FACE PRECISION	FACE RECALL	
REFERENCE	0.69	**0.32**	0.79	**0.25**	t=0.7
INSETGAN	**0.83**	0.30	**0.89**	0.16	t=0.7
REFERENCE	0.92	0.03	0.92	**0.05**	t=0.4
INSETGAN	**0.93**	**0.04**	**0.95**	0.01	t=0.4

PROJECT PAGE
afruehstueck.github.io/insetgan