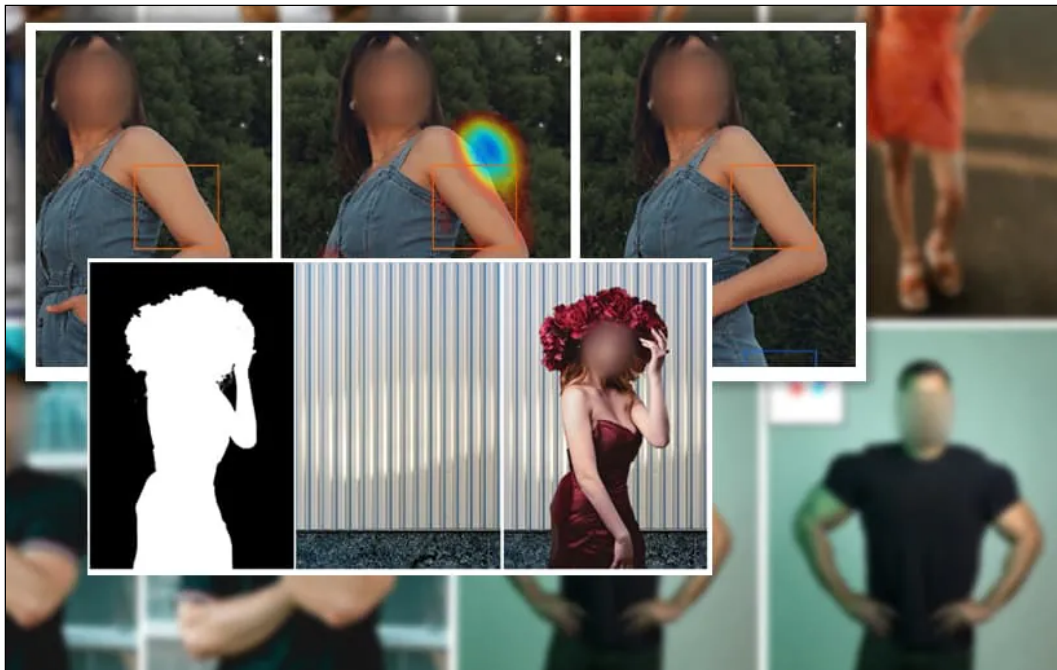
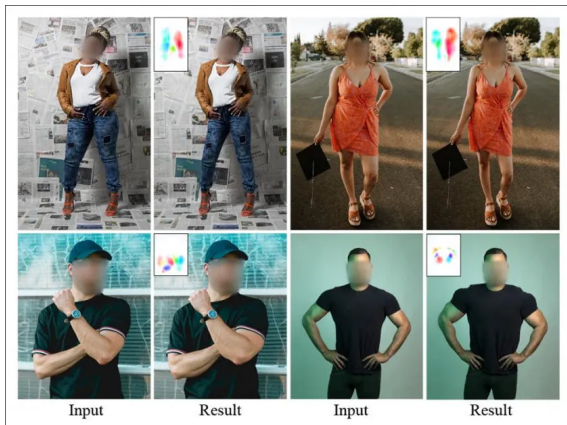


Faking 'Better' Bodies With AI

Originally published at <https://www.unite.ai/faking-better-bodies-with-ai/>

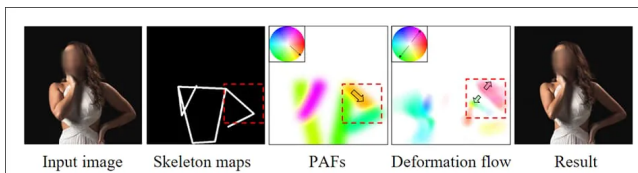


New research from the Alibaba DAMO academy offers an AI-driven workflow for automating the reshaping of images of bodies – a rare effort in a computer vision sector currently occupied with [face-based manipulations](#) such as deepfakes and GAN-based [face editing](#).



Inset in 'result' columns, the generated attention maps which define the areas to be amended. Source: <https://arxiv.org/pdf/2203.04670.pdf>

The researchers' architecture uses skeleton pose estimation to tackle the greater complexity that image synthesis and editing systems face in conceptualizing and parametrizing existing body images, at least to a level of granularity that actually allows meaningful and selective editing.



Estimated skeleton maps help to individuate and focus attention on areas of the body likely to be retouched, such as the upper arm area.

The system ultimately enables a user to set parameters that can change the appearance of weight, muscle mass, or weight distribution in full-length or mid-length photos of people, and is able to generate arbitrary transformations on clothed or unclothed body sections.



Left, the input image; middle, a heat-map of the derived attention areas; right, the transformed image.

The motivation for the work is the development of automated workflows that could replace the arduous digital manipulations undertaken by photographers and production graphics artists in various branches of the media, from fashion to magazine-style output and [publicity material](#).

In general, the authors acknowledge, these transformations are usually applied with 'warp' techniques in Photoshop and other traditional bitmap editors, and are almost exclusively used on images of women. Consequently, the custom dataset developed to facilitate the new process consists mostly of pictures of female subjects:

'As body retouching is mainly desired by females, the majority of our collection are female photos, considering the diversity of ages, races (African:Asian:Caucasian = 0.33:0.35:0.32), poses, and garments.'

The [paper](#) is titled *Structure-Aware Flow Generation for Human Body Reshaping*, and comes from five authors associated with Alibaba's global DAMO academy.

Dataset Development

As is usually the case with image synthesis and editing systems, the architecture for the project required a customized training dataset. The authors commissioned three photographers to produce standard Photoshop manipulations of apposite images from stock photography site Unsplash, resulting in a dataset – titled *BR-5K** – of 5,000 high quality images at 2K resolution.

The researchers emphasize that the objective of training on this dataset is not to produce 'idealized' and generalized features relating to an index of attractiveness or desirable appearance, but rather to extract the central feature mappings associated with professional manipulations of body images.

However, they concede that the manipulations ultimately reflect transformative processes that map a progression from 'real' to a preset notion of 'ideal':

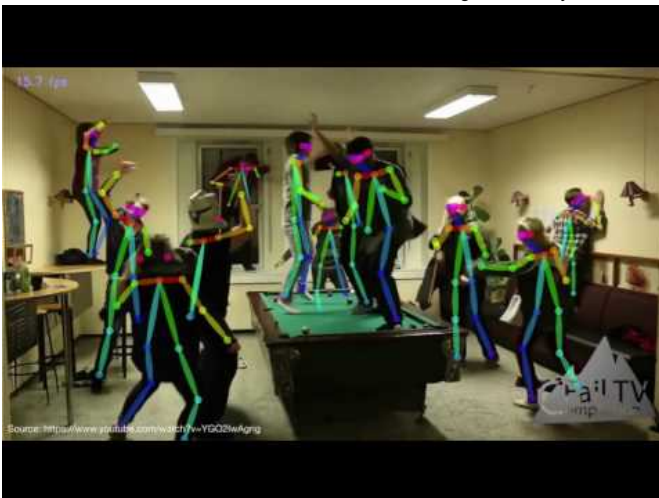
'We invite three professional artists to retouch bodies using Photoshop independently, with the goal of achieving slender figures that meet the popular aesthetics, and select the best one as ground-truth.'

Since the framework does not deal with faces at all, these were blurred out before being included in the dataset.

Architecture and Core Concepts

The system's workflow involves feeding in a high resolution portrait, downsampling it to a lower resolution that can fit into the available computing resources, and extracting an estimated skeleton-map pose (second figure from left in image below), as well as Part Affinity Fields (PAFs), which were [innovated](#) in 2016 by The Robotics Institute at Carnegie Mellon University (see video embedded directly below).

Realtime Multi-Person 2D Human Pose Estimation using Part Affinity Fields, CVPR 2017 Oral



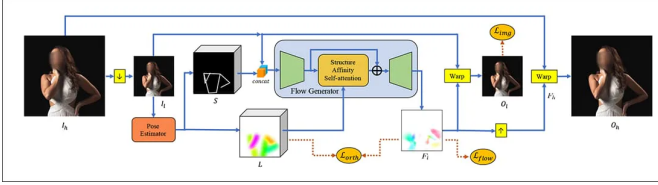
Part Affinity Fields help to define orientation of limbs and general association with the broader skeletal framework, providing the new project with an additional attention/localization tool.



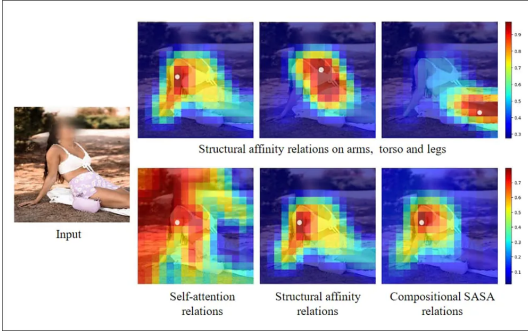
From the 2016 Part Affinity Fields paper, predicted PAFs encode limb orientation as part of a 2D vector that also includes the general position of the limb. Source: <https://arxiv.org/pdf/1611.08050.pdf>

Despite their apparent irrelevance to the appearance of weight, skeleton maps are useful in directing the final transformative processes to parts of the body to be amended, such as upper arms, rear, and thighs.

After this, the results are fed to a Structure Affinity Self-Attention (SASA) in the central bottleneck of the process (see image below).



The SASA regulates the consistency of the flow generator that fuels the process, the results of which are then passed to the warping module (second from right in the image above), which applies the transformations learned from training on the manual revisions included in the dataset.



The Structure Affinity Self-Attention (SASA) module allocates attention to pertinent body parts, helping to avoid extraneous or irrelevant transformations.

The output image is subsequently upsampled back to the original 2K resolution, using processes not dissimilar to the standard, 2017-style deepfake architecture from which popular packages such as DeepFaceLab have since been derived; the upsampling process is also common in GAN editing frameworks.

The attention network for the schema is modeled after *Compositional De-Attention Networks* ([CODA](#)), a 2019 US/Singapore academic collaboration with Amazon AI and Microsoft.

Tests

The flow-based framework was tested against prior flow-based methods [FAL](#) and Animating Through Warping ([ATW](#)), as well as image translation architectures [Pix2PixHD](#) and [GFLA](#), with [SSIM](#), [PSNR](#) and [LPIPS](#) as evaluation metrics.

Method	SSIM $\uparrow$	PSNR $\uparrow$	LPIPS $\downarrow$	User Study $\uparrow$
Baseline	0.8339	24.4916	0.0823	N.A.
FAL	0.8261	24.1841	0.0837	14.4%
ATW	0.8316	24.6332	0.0805	9.8%
pix2pixHD	0.7271	21.8381	0.2800	3.6%
GFLA	0.6649	21.4796	0.6136	1.8%
Ours	0.8354	24.7924	0.0777	70.4%

Results of initial tests (arrow direction in headers indicates whether lower or higher figures are best).

Based on these adopted metrics, the authors' system outperforms the prior architectures.



Selected results. Please refer to the original PDF linked in this article for higher resolution comparisons.

In addition to the automated metrics, the researchers conducted a user study (final column of results table pictured earlier), wherein 40 participants were each shown 30 questions randomly selected from a 100-question pool relating to the images produced via the various methods. 70% of the respondents favored the new technique as more 'visually appealing'.

Challenges

The new paper represents a rare excursion into AI-based body manipulation. The image synthesis sector is currently far more interested either in generating editable bodies via methods such as Neural Radiance Fields (NeRF), or else is fixated on exploring the latent space of GANs and the potential of autoencoders for facial manipulation.

The authors' initiative is currently limited to producing changes in perceived weight, and they have not implemented any kind of inpainting technique that would restore the background that's inevitably revealed when you slim down a picture of someone.

However, they propose that portrait matting and background blending through textural inference could trivially solve the problem of restoring the parts of the world that were formerly hidden in the image by human 'imperfection'.



A proposed solution for restoring background that's revealed by AI-driven fat reduction.

** Though the preprint refers to supplemental material giving more details about the dataset, as well as further examples from the project, the location of this material is not made available in the paper, and the corresponding author has not yet responded to our request for access.*

First published 10th March 2022.

Writer on machine learning, domain specialist in human image synthesis. Former head of research content at Metaphysic.ai, until its dissolution into DNEG's Brahma.ai.

Portfolio site: martinanderson.ai

Contact: martin@martinanderson.ai



Discover More