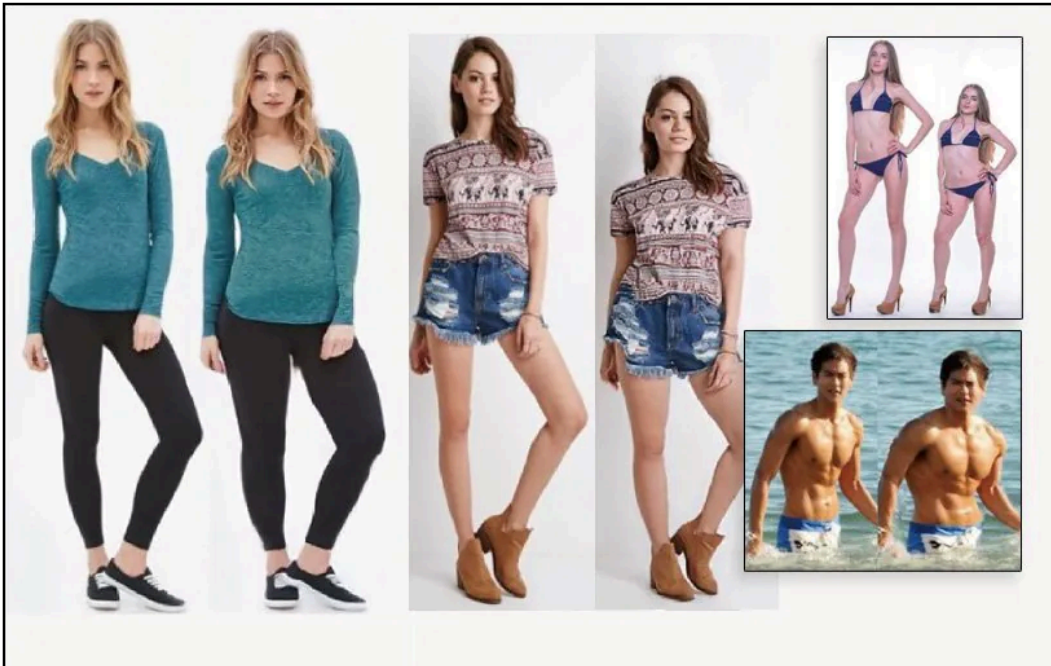
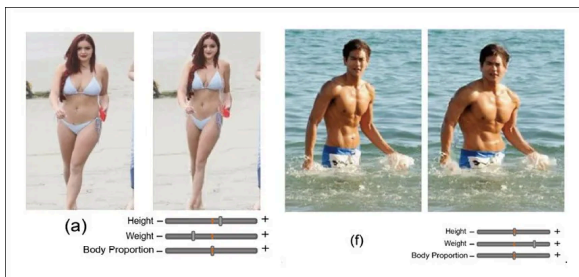


Reshaping Human Body Types With AI

Originally published at <https://www.unite.ai/reshaping-human-body-types-with-ai/>



A new research collaboration from China offers a novel method of reshaping the human body in images, by the use of a coordinated twin neural encoder network, guided by a parametric model, that allows an end-user to modulate weight, height, and body proportion in an interactive GUI.

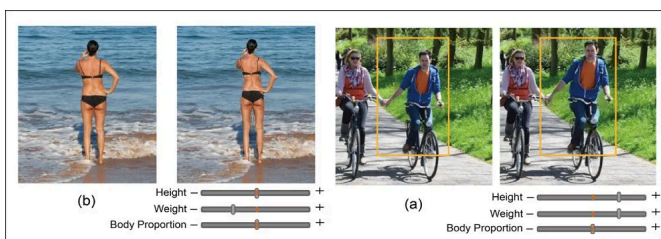


Parametrized modulation of body shape, with sliders altering the three available features. Source: <https://arxiv.org/pdf/2203.10496.pdf>

The work offers several improvements over a [recent similar project](#) from Alibaba, in that it can convincingly alter height and body proportion as well as weight, and has a dedicated neural network for 'inpainting' the (non-existent) background that can be revealed by 'slimmer' body images. It also improves on a notable [earlier parametric method](#) for body reshaping by removing the need for extensive human intervention during the formulation of the transformation.

Titled *NeuralReshaper*, the new architecture fits a parametric 3D human template to a source image, and then uses distortions in the template to adapt the original image to the new parameters.

The system is able to handle body transformations on clothed as well as semi-clothed (i.e. beachwear) figures.

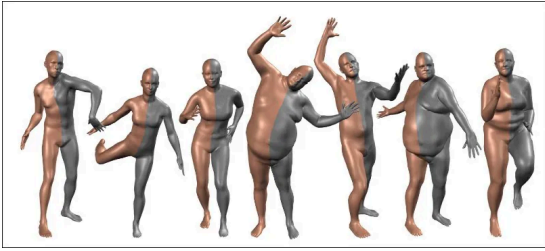


Transformations of this type are currently of intense interest to the [fashion AI](#) research sector, which has produced a number of StyleGAN/CycleGAN-based and general neural network platforms for [virtual try-ons](#) which can adapt available clothing items to the body shape and type of a user-submitted image, or otherwise help with visual conformity.

The [paper](#) is titled *Single-image Human-body Reshaping with Deep Neural Networks*, and comes from researchers at Zhejiang University in Hangzhou, and the School of Creative Media at the City University of Hong Kong.

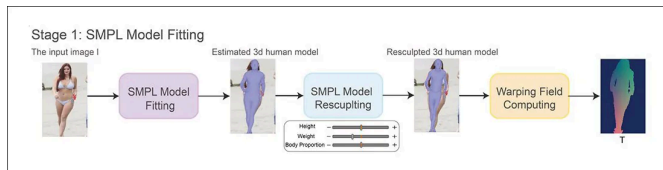
SMPL Fitting

NeuralReshaper makes use of the Skinned Multi-Person Linear Model (SMPL) [developed](#) by the Max Planck Institute for Intelligent Systems and renowned VFX house Industrial Light and Magic in 2015.



SMPL Parametric humans from the 2015 Planck/ILM collaboration. Source: <https://files.is.tue.mpg.de/black/papers/SMPL2015.pdf>

In the first stage of the process, an SMPL model is generated from a source image to which body transformations are desired to be made. The adaptation of the SMPL model to the image follows the [methodology](#) of the Human Mesh Recovery (HMR) method proposed by universities in Germany and the US in 2018.



The three parameters for deformation (weight, height, body proportion) are calculated at this stage, together with a consideration of the camera parameters, such as focal length. 2D keypoints and generated silhouette alignment provide the enclosure for the deformation in the form of a 2D silhouette, an additional optimization measure that increases the boundary accuracy and allows for authentic background inpainting further down the pipeline.

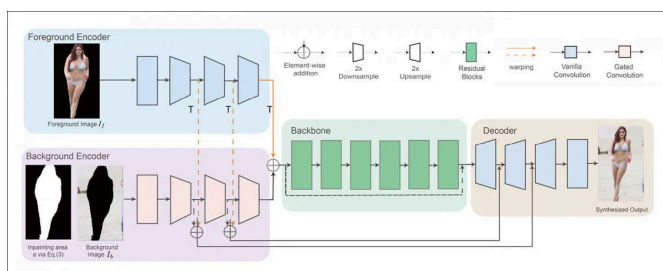


SMPL fitting stages: left, the source image; second, the optimization result obtained from the method outlined in 2016 [research](#) led by the Max Planck Institute for Intelligent Systems; third, a direct inference result from the pre-trained model for [End-to-end Recovery of Human Shape and Pose](#); fourth, the results obtained after optimization of the 2D keypoints; and finally, fifth, the completed fit after silhouette optimization (see above).

The 3D deformation is then projected into the architecture's image space to facilitate a dense warping field that will define the deformation. This process takes around 30 seconds per image.

NeuralReshaper Architecture

NeuralReshaper runs two neural networks in tandem: a foreground encoder that generates the transformed body shape, and a background encoder that focuses on filling in 'de-occluded' background regions (in the case, for instance, of slimming down a body – see image below).



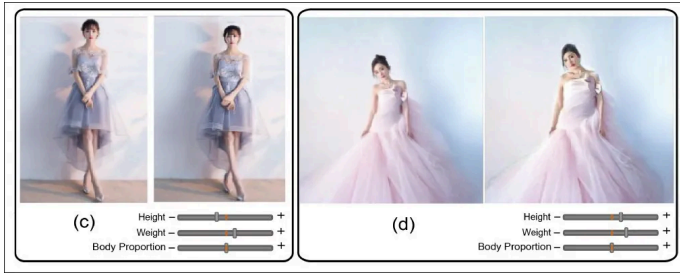
The U-net-style framework integrates the output from the two encoders' features before passing the result to a unified encoder which ultimately produces a novel image from the two inputs. The architecture features a novel warp-guided mechanism to enable integration.

Training and Experiments

NeuralReshaper is implemented in PyTorch on a single NVIDIA

1080ti GPU with 11gb of VRAM. The network was trained for 100 epochs under

the Adam optimizer, with the generator set to a target loss of 0.0001 and the discriminator to a target loss of 0.0004. The training occurred on a batch size of 8 for a proprietary outdoor dataset (drawn from [COCO](#), [MPII](#), and LSP), and 2 for training on the [DeepFashion](#) dataset.



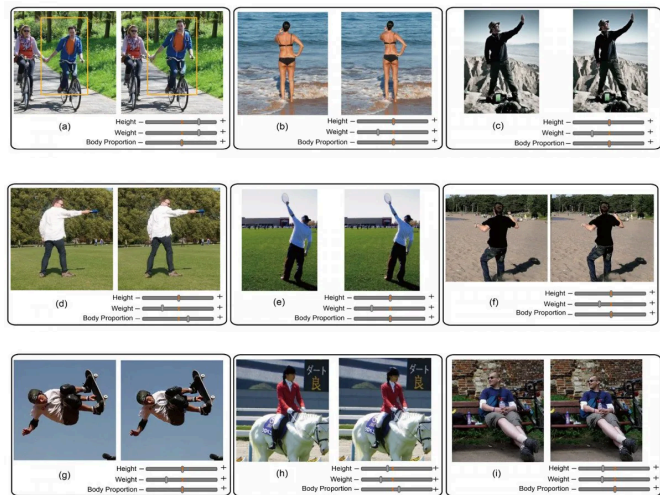
On the left, the original images, on the right, the repositioned output of NeuralReshaper.

Below are some examples exclusively from the DeepFashion dataset as trained for NeuralReshaper, with the original images always on the left.



The three controllable attributes are disentangled, and can be applied separately.

Transformations on the derived outdoor dataset are more challenging, since they frequently require infilling of complex backgrounds and clear and convincing delineation of the transformed body types:



Parametric Necessity

As the paper observes, same-image transformations of this type represent an ill-posed problem in image synthesis. Many transformative GAN and encoder frameworks can make use of paired images (such as the diverse projects designed to effect [sketch>photo](#) and [photo>sketch](#) transformations).

However, in the case at hand, this would require image pairs featuring the same people in different physical configurations, such as the ‘before and after’ images in diet or plastic surgery advertisements – data that is difficult to obtain or generate.

Alternately, transformative GAN networks can train on much more diverse data, and effect transformations by seeking out the [latent direction](#) between the source (original image latent code) and the desired class (in this case ‘fat’, ‘thin’, ‘tall’, etc.). However, this approach is currently too limited for the purposes of fine-tuned body reshaping.

Neural Radiance Fields ([NeRF](#)) approaches are much further advanced in full-body simulation than most GAN-based systems, but remain scene-specific and resource intensive, with currently very limited ability to edit body types in the granular way that NeuralReshaper and prior projects are trying to address (short of [scaling the entire body down](#) relative to its environment).

The GAN’s latent space is hard to govern; VAEs alone do not yet address the complexities of full-body reproduction; and NeRF’s capacity to consistently and realistically remodel human bodies is still nascent. Therefore the incorporation of ‘traditional’ CGI methodologies such as SMPL seems set to continue in the human image synthesis research sector, as a method to corral and consolidate features, classes, and latent codes whose parameters and exploitability are not yet fully understood in these emerging technologies.

First published 31st 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