

A model that uses human prompts and sketches to generate realistic fashion images

April 25 2023, by Ingrid Fadelli



The team's multimodal garment designer framework can generate a new fashion image from text, human-drafted key-points, and a garment sketch. Credit: Baldrati et al

Artificial intelligence (AI) recently started making its way into many creative industries, for instance, in the form of tools for digital artists, architects, interior designers and image editors. In these contexts, AI can automate processes that are tedious or time consuming, while also potentially inspiring artists and facilitating their creative process.

Researchers at University of Florence, University of Modena and Reggio

Emilia and University of Pisa recently set out to explore the potential of AI models in [fashion design](#). In a paper pre-published on *arXiv*, they introduced a new computer vision framework that could help fashion designers to visualize their designs, by showing them how they might look on the [human body](#).

Most past studies exploring the use of AI in the [fashion industry](#) focused on computational tools that can recommend garments similar to those selected by a user or models that can show online customers how garments would look on their body (i.e., virtual try-on systems). This team of Italian researchers, on the other hand, set out to develop a framework that could support the work of designers, showing them how garments they designed might look in real-life, so that they can find new inspiration, identify potential issues and alter their designs if needed.

"Differently from previous works that mainly focused on the virtual try-on of garments, we propose the task of multimodal conditioned fashion image editing, guiding the generation of human-centric fashion images by following multimodal prompts, such as text, human body poses, and garment sketches," Alberto Baldrati, Davide Morelli and their colleagues wrote in their paper.

"We tackle this problem by proposing a new architecture based on latent diffusion models, an approach that has not been used before in the fashion domain."

Instead of using [generative adversarial networks](#) (GANs), artificial neural network architectures often used to generate new texts or images, the researchers decided to create a framework based on latent diffusion models or LDMs. As they are trained in a compressed and lower-dimensional latent space, LDMs can create high-quality synthetic images.

While these promising models have been applied to many tasks that require the generation of artificial images or videos, they have rarely been used in the context of fashion image editing. Most previous works in this area introduced GAN-based architectures, which generate lower quality images than LDMs.

Most existing datasets for training AI models on fashion design tasks only include low-resolution images of clothing and do not include the information necessary to create fashion images based on text prompts and sketches. To effectively train their model, Baldrati, Morelli and their colleagues thus had to first update these existing datasets or create new ones.

"Given the lack of existing datasets suitable for the task, we also extend two existing fashion datasets, namely Dress Code and VITON-HD, with multimodal annotations collected in a semi-automatic manner," Baldrati, Morelli and their colleagues explained in their paper. "Experimental results on these new datasets demonstrate the effectiveness of our proposal, both in terms of realism and coherence with the given multimodal inputs."

In initial evaluations, the model created by this team of researchers achieved very promising results, creating realistic images of garments on human bodies inspired by human sketches and specific text prompts. Their model's source code and the multimodal annotations they added to the datasets will soon be released on GitHub.

In the future, this new model could be integrated in existing or new software tools for fashion designers. It could also inform the development of other AI architectures based on LDMs for real-world creative applications.

"This is one of the first successful attempts to mimic the designers' job

in the creative process of fashion design and could be a starting point for a capillary adoption of diffusion models in [creative industries](#), oversight by human input," Baldrati, Morelli and their colleagues conclude in their paper.

More information: Alberto Baldrati et al, Multimodal Garment Designer: Human-Centric Latent Diffusion Models for Fashion Image Editing, *arXiv* (2023). [DOI: 10.48550/arxiv.2304.02051](https://doi.org/10.48550/arxiv.2304.02051)

© 2023 Science X Network

Citation: A model that uses human prompts and sketches to generate realistic fashion images (2023, April 25) retrieved 25 September 2023 from <https://techxplore.com/news/2023-04-human-prompts-generate-realistic-fashion.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.