Abstract

In the midst of rapid advancements in data retrieval and large-scale generative model technologies, the paradigms of artistic endeavors, including writing and painting, are constantly evolving. Within the realms of Computer Graphics (CG) and Computer Vision (CV), these technological strides have substantially bolstered the efficiency of image design processes. Historically, designers would hone their artistic prowess through rigorous practice, paired with an immersion in classical artworks to cultivate their aesthetic judgment.

Generative models are able to produce complex, high-quality images from simple inputs. However, it is crucial to demystify the notion that such advancements inherently equate to significant increases in designers' productivity. While the algorithms behind AI generative models are indeed potent and efficient, their value to users is contingent upon their alignment with the designers' needs. From my perspective, the value that current image-generative models provide to designers is not commensurate with the models' performance capabilities. The majority of these models operate as end-to-end 'black boxes', making it challenging for users to achieve desired outcomes through straightforward inputs. To address this issue, this thesis discusses the following approach: 1) From the perspective of developing an algorithm: enhance the algorithm design of the generative model to allow for a variety of input modalities. This would not only improve model performance but also enrich user interaction with the algorithm. Recent advancements in multimodal generative models exemplify the capability to produce images through various inputs such as text prompts, image references, and spatial guidance, including sketches and semantic maps 2) From the perspective of interaction with models: Offer expansive interactive and editing capabilities during the creative process. By doing so, the generative model becomes a tool for exploration and refinement, rather than a one-shot solution. 3) From the perspective of the design process: It is imperative to engage with the designer's workflow, understanding the specific needs at each stage of creation. Generative model algorithms should be tailored to address these needs, ensuring that the tools developed are not just technologically advanced, but also contextually relevant.

As Picasso aptly noted, "A painting is not thought out in advance. While it is being done, it changes as one' s thoughts change. And when it' s finished, it goes on changing, according to the state of mind of whoever is looking at it." Thus, creation is an exploration, with designers iteratively reflecting on and drawing inspiration from intermediate outputs until satisfaction is achieved. Contemporary generative models and data retrieval techniques have yet to fully encapsulate this ethos. In this dissertation, we conceptualize the creative trajectory as an Ambiguous-to-Concrete continuum. Taking full-body human figure design as a case study, we break down the conventional

figure painting workflow into three separate stages. For each stage, we identify the unique requirements of the user and introduce new data retrieval or generative modeling techniques specifically designed to best assist the user in achieving their creative vision. These stages include (In sequential order):

- **Posture Initialization:** At the initial stage where user intent is still forming and exploration is needed, we introduce a 'global-to-local' retrieval scheme for 3D motion data. Instead of traditional skeletal sketching, users draw trajectories for specific joints to retrieve and refine snippets of motion data, choosing keyframes that will guide further design steps. This allows users to view the motion data from different angles within a 3D space, enabling them to select the desired pose, particularly when aiming to depict dynamic actions such as dancing, thus offering users a broader range of referencesand choices.
- **Outfit Selection:** This phase often entails sifting through a plethora of attire references to pinpoint desired designs, coupled with iterative refinements. Considering the intricacies of fabric depiction demand profound sartorial design expertise, our solution offers an 'image-guided' generative model. This approach omits the drawing input stage, allowing users to concentrate on attire's alignment with the posture and the overall character portrayal.
- **Facial and Detail Depiction:** With the overarching attire and posture solidified, users typically possess clearer intentions regarding intricate details, especially facial attributes like hairstyle and expressions. Our solution here is a high-fidelity 'sketch-guided' generative model, ensuring the output closely mirrors the input while maintaining consistency in non-edited areas.

Through rigorous experiments, we validate the efficacy of our phase-specific interactive pipelines, benchmarking them against state-of-the-art (SOTA) counterparts on analogous tasks. The empirical results affirm that our pipeline adeptly navigates each phase of the Ambiguous-to-Concrete spectrum, offering meaningful design support. We posit that our methodologies and concepts are not confined to human figure design but are readily applicable across diverse design scenarios. As such, the frameworks and insights proffered herein can serve as foundational pillars for subsequent inquiries and innovations in the design research landscape.

Keywords: Character Image Generation, Ambiguous to Concrete, Data Retrieval, Generative Models, Artistic Creation.