BACHELOR THESIS EXPOSÉ

Optimizing perceived aesthetics of UIs using diffusion based generative models

MORITZ WÖRMANN

March 21, 2024

Variational Autoencoder

1 MOTIVATION AND BACKGROUND

Through tools like Figma [1] and co., the entry barriers for creating Apps and Websites with User Interfaces (UIs) for users to interact with decrease steadily. However, designing visually pleasing UIs still proves to be a complicated task, especially since these are highly subjective categories. This challenge becomes even more significant when considering the impact initial impressions of a UI can have on the users perception and further willingness to stay on the website or the mobile app [4]. Currently, designing a new User Interface (UI) is often a task for multiple teams with different professions, like graphic design and software engineering. While modern project management strategies like Scrum can alleviate the difficulties introduced by aligning and communicating stakeholder and user interests with the final product, they still rely heavily on good communication and the abundancy of time. Automating this task or provding assistance via automatic algorithms is therefore a worthwile subject as an "end-to-end" process of creating user-interfaces or at least optimising existing UIs can reduce time and effort.

Layout generation tasks describe the challenge of aligning different elements and components of user interfaces as well as controlling other parameters like font, font size and color in a visually pleasing way. While this task is evidently sufficiently difficult for humans, assigning this task to algorithms proves to be even more difficult as the challenge of defining what is considered visually pleasing for humans is not straight forward.

2 PROBLEM DESCRIPTION

The goal of this research is to provide a potential user with a fully usable pipeline in which they can input an image of an already existing UI and potentially provide additional instructions. This automated pipeline would be able to segment the **UI!** (**UI!**) into its components and rearrange them in a better or visually more pleasing way. This segmentation process thus functions as a transformation or mapping of the UI into a different space, which might even be called a latent space. An algorithm or model can operate in this space and retrieve feedback from a model, pretrained on a dataset in which users have been interrogated for their perceived aesthetics of user interfaces. It remains to be shown, if this classifier model can predict directly from this latent space or if the user interface has to be transformed out of this latent space again first (diffusion). Clearly, this transformation proves

1 https://www.figma.com/.

to be a an additional challenge as well as deciding the size and dimension of the latent space. This reasearch aims at exploring different approaches as to how these latent spaces can look with a focus to them being able to be used in a pipeline from the latent space to an aesthetics predictor without breaking autograd vectors in order to leverage common gradient-descent patterns for this task.

To solve these questions the following research questions will be answered:

RQ 1 How can UIs be segmented in a way such that the segmentation can be optimized and later reassembled?

RQ 2 How does an optimal latent space in which the User Interface Layout can be represented look? How can this space be used to optimize UIs after they have been segmented in order to maximize perceived aesthetic, measured by a pretrained classifier?

RQ 3 How can (accidental) adversarial attacks by the optimizer against the Aesthetic Predictor be avoided?

RQ 4 Do Diffusion Models provide advantages, either via Pix2Pix optimization or via a different latent space which represents the UI

RQ 5 How can all of these different approaches be constrained with user supplied input?

RQ 6 How can one of these approaches be packaged and supplied to a potential user in an end-to-end pipeline?

In the next sections, related work is portrayed, after which a structure for the final thesis is presented.

3 STATE-OF-THE-ART

The current state-of-the-art divides itself into three main parts: (1) Segmentation of UIs (2) Efforts to optimize UIs using non-diffusion based approaches (3) Efforts to optimize UIs using diffusion based approaches

3.1 *Segmentation of UIs*

3.2 *Efforts to optimize UIs using non-diffusion based approaches*

Recent research has shown impressive advancements while not relying on diffusion based approaches like in 2022: Kong et al. [8] which shows how a layout transformer model can be used to reliably generate missing attributes from their latent space, which is consisting of different elements, labeled with their category and their respective positioning on user interfaces. However such approaches often struggle with user inputs and constraints and are hard to control with fixed parameters, which include dictated positioning. A similar approach is presented in LayoutTransformer: Gupta et al. [5]

A different approach is the usage of a VAE (vae) like in Jiang et al. [7] which proposes segmenting the user interface into different regions first before then "filling out" these regions with other user interface segments in order to combat the challenges of high level relationships in user interfaces which are hard to process for these models. This research builds on Arroyo et al. [1] which initially proposed the usage of vaes for layout generation tasks. Such VAE approaches have also been explored in Xie et al. [13] and Patil et al. [11].

Still, non-VAE approaches also exist, leveraging advantages of Graph neural networks which allow for refinement of initial user controlled relationship definement like in H.-Y. Lee et al. [9].

3.3 *Efforts to optimize UIs using diffusion based approaches*

While not strictly related to UIs, research has already been done in the field of metric based optimization for Pix2Pix Diffusion models like in Deckers et al. [2] which shows that a simple gradient descent pipeline with a classifier at the end can be used to optimize a prompt embedding which is passed to a stable diffusion model.

Next to that is the different approach of not relying on Pix2Pix models but using a different autoencoder to get to the latent space from the UIs. Though not directly related to the actual diffusion, Deka et al. [3] already showcased an AutoEncoder which reduces the dimensions of a user interface layout to a 64-dimensional vector which can later be used to retrieve the layout representation again. However, this lacks closing the gap between a generated layout and the finished user interface, which is a vital part of this research. Still, this approach might provide useful insights if it were to be possible to use this AutoEncoder directly in a diffusion model.

4 PROPOSED APPROACH

As all of the mentioned research questions are somewhat related to the overall goal of this research, which is to optimize perceived aesthetics, a clear way to measure this metric is needed. As all of the approaches will rely on the usage of a common gradient descent pipeline on one way or another, this metric needs to measured in a way which doesn't break any autograd graphs. For simplicity, the same model will be used for all of the different approaches, which is the one presented in 20203: Leiva et al. [10]. However, a slight modification will be necessary as the provided pre-trained model is using the tensorflow [2] framework and therefore not compatible with the torch [3] autograd mechanisms used in this research. Thus, the model will be retrained on the same data using the same presented model architecture which should yield similar results to the ones presented in the research.

4.1 *Research Question 1*

Past Research like in Biplab Deka et. al: Rico [3] has shown that User Interface Segmentation is a task which is manageable by state of the art algorithms. It has been proven that optical segmentation into Text and Non-Text elements (by mere masking of the affected regions) can be used to train an AutoEncoder architecture which reliably reduces the dimension of the information in a user interface. As this research is exploring a similar question (transforming a user interface in a latent-like-space), a similar approach might provide favorable results for this task. It remains to be shown how significant the effects of different segmentation approaches are on the overall goal. The maturity and reliability of models like the one presented in the mentioned paper suggests that this effect may be minimal.

4.2 *Research Question 2*

As previously described, the main challenge in this domain will be finding an entire pipeline, which includes a latent space, a function which retrieves the user interface out of this latent space and a predictor which determines the aesthetics for this retrieved and modified user interface. One such space might just be a vector of coordinates where the segments are placed on the user interface.

Once such a space and pipeline has been found, it may be trivial to optimize the representation of the user interface by utilizing common gradient

2 https://www.tensorflow.org
3 https://pytorch.org

Table 1: Overview of Proposed Approach

Research Questions & Related Study Phase		
RQ 1: Requirements Engineering	RQ 2: Data Visualization	RQ 3: Development & Evaluation
Tasks		
○ Process Analysis ○ Stakeholder Identification ○ Requirements Elicitation ○ Use Cases ○ KPI ○ Relationship to User Experience (UX) Measures	○ Visualization design ○ Use Case Mapping ○ Data Dependencies	○ Join visualizations and Process ○ App Specifications & Prototype ○ Initial Evaluation ○ Iterative app development ○ Final evaluation & analysis
Expected Results		
○ Process Description ○ Stakeholder & Requirements Cluster ○ Prioritized Use Cases ○ Information Needs	○ Data pipeline (description) ○ Visualizations ○ Initial designs	○ Final app ○ Concrete Use Cases ○ Evaluation Results

descent patterns provided by major machine learning frameworks, provided that the whole pipeline actually converges.

4.3 *Research Question 3*

While this part of the research is arguably the most important one, it might also prove to be the most complex one. Keeping the pipeline from becoming too volatile or quickly "learning" how to exploit the aesthetics predictor and thus creating an adversarial task is a complex task. These exploits might lead to undesirable results in which user interfaces might show extreme changes for no apparent reason which might lead to a higher predict aesthetic but, are do in fact not show the same favorability during interrogation through humans.

Initially this might be mitigated by optimizing the predictor through a bigger and more complex model architecture and extending the datasets used for training it. However, this might only alleviate potential issues to a certain extent at which other techniques have to be explored such as restricting the latent space and adding a regularization or penalty on extreme changes.

4.4 *Research Question 4*

This research question can be broken up into two seperate tasks. While one, once again revolves around the task of finding a suitable latent space in which the user interface can be projected, the other one assumes that acceptable results can be achieved by solely relying on Pix2Pix approches like StableDiffusion [12]. This would most definetely involve finetuning the AutoEncoder in the StableDiffusion model to adapt to UIs. However this might prove to be a challenging approach as these models are notoriously hard to control, which has evolved into a seperate research field called prompt en-

gineering. It might even be unrealistic to assume that Pix2Pix optimization can even produce something remotely resembling a user interface, disregarding aesthetics.

Thus, the first approach, finding a different latent space could show improved results. For this, some research has already been done, like in 2023: Hui et al. [6] where latent space effectively only holds information about the layout of a user interface.

4.5 *Research Question 5*

Constraining the generated layouts and UIs has been the subject of multiple research efforts [9]. While most of these efforts rely on giving the constraints at the start of the pipeline, e.g. developing relationships and going from there on to the user interfaces, another approach could be to penalize a model/pipeline for a undesirable results which may include user defined constraints. It would then be entirely up to the model to grasp these constraints and work them into the predictions.

4.6 *Research Question 6*

Following the described approach, the structure shown in fig. 1 is proposed for the thesis.

Figure 1: Proposed Structure

1. Introduction
   - 1.1 Motivation
   - 1.2 Problem Statement
   - 1.3 Structure of the Work
2. Background
   - 2.1 User Experience
   - 2.2 Usability Evaluation
   - 2.3 In-Vehicle Information Systems
   - 2.4 Requirements Engineering
   - 2.5 Creativity Techniques - Design Thinking
3. State-of-the-Art and Related Work
   - 3.1 User Experience Evaluation
   - 3.2 Data-Analysis and Visualization
   - 3.3 Usages in the Automotive Domain
   - 3.4 Comparison of existing Approaches
4. Proposed Method and Implementation
   - 4.1 Requirements Engineering
   - 4.2 Data Visualizations
   - 4.3 Final Application
5. Evaluation
   - 5.1 Evaluation Method
   - 5.2 Use Cases
   - 5.3 Interview Guidelines
   - 5.4 Questionnaires
6. Results
7. Discussion
   - 7.1 Threats to Validity
   - 7.2 Future Work

[1] Diego Martin Arroyo, Janis Postels, and Federico Tombari. Variational transformer networks for layout generation, 2021.

[2] Niklas Deckers, Julia Peters, and Martin Potthast. Manipulating embeddings of stable diffusion prompts, 2023.

[3] Biplab Deka, Zifeng Huang, Chad Franzen, Joshua Hibschman, Daniel Afergan, Yang Li, Jeffrey Nichols, and Ranjitha Kumar. Rico: A mobile app dataset for building data-driven design applications. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, UIST '17, page 845–854, New York, NY, USA, 2017. Association for Computing Machinery.

[4] Maria Douneva, Rafael Jaron, and Meinald T. Thielsch. Effects of Different Website Designs on First Impressions, Aesthetic Judgements and Memory Performance after Short Presentation. *Interacting with Computers*, 28(4):552–567, 06 2016.

[5] Kamal Gupta, Justin Lazarow, Alessandro Achille, Larry Davis, Vijay Mahadevan, and Abhinav Shrivastava. Layouttransformer: Layout generation and completion with self-attention, 2021.

[6] Mude Hui, Zhizheng Zhang, Xiaoyi Zhang, Wenxuan Xie, Yuwang Wang, and Yan Lu. Unifying layout generation with a decoupled diffusion model, 2023.

[7] Zhaoyun Jiang, Shizhao Sun, Jihua Zhu, Jian-Guang Lou, and Dongmei Zhang. Coarse-to-fine generative modeling for graphic layouts. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(1):1096–1103, Jun. 2022.

[8] Xiang Kong, Lu Jiang, Huiwen Chang, Han Zhang, Yuan Hao, Haifeng Gong, and Irfan Essa. Blt: Bidirectional layout transformer for controllable layout generation, 2022.

[9] Hsin-Ying Lee, Lu Jiang, Irfan Essa, Phuong B Le, Haifeng Gong, Ming-Hsuan Yang, and Weilong Yang. Neural design network: Graphic layout generation with constraints, 2020.

[10] Luis A. Leiva, Morteza Shiripour, and Antti Oulasvirta. Modeling how different user groups perceive webpage aesthetics. *Universal Access in the Information Society*, 22(4):1417–1424, Nov 2023.

[11] Akshay Gadi Patil, Omri Ben-Eliezer, Or Perel, and Hadar Averbuch-Elor. Read: Recursive autoencoders for document layout generation, 2020.

[12] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models, 2021.

[13] Yuxi Xie, Danqing Huang, Jinpeng Wang, and Chin-Yew Lin. Canvasemb: Learning layout representation with large-scale pre-training for graphic design. In *Proceedings of the 29th ACM International Conference on Multimedia*, pages 4100–4108, October 2021.