Predicting and Diagnosing User Engagement with Mobile UI Animation via a Data-Driven Approach

Ziming Wu[1], **Yulun Jiang**[2*], **Yiding Liu**[1*], **and Xiaojuan Ma**[1]
[1]Hong Kong University of Science and Technology, Hong Kong
[2]Wuhan University, China
zwual@connect.ust.hk, yljblues@whu.edu.cn, adamyd.liu@gmail.com, mxj@cse.ust.hk

ABSTRACT
Animation, a common design element in user interfaces (UI), can impact user engagement (UE) with mobile applications. To avoid impairing UE due to improper design of animation, designers rely on resource-intensive evaluation methods like user studies or expert reviews. To alleviate this burden, we propose a data-driven approach to assisting designers in examining UE issues with their animation designs. We first crowdsource UE assessments of mobile UI animations. Based on the collected data, we then build a novel deep learning model that captures both spatial and temporal features of animations to predict their UE levels. Evaluations show that our model achieves a reasonable accuracy. We further leverage the animation feature encoded by our model and a sample set of expert reviews to derive potential UE issues of a particular animation. Finally, we develop a proof-of-concept tool and evaluate its potential usage in actual design practices with experts.

Author Keywords
Mobile UI animation; user engagement; data-driven approach

CCS Concepts
•**Human-centered computing** → **User models; Graphical user interfaces;**

INTRODUCTION
Animation is a widely adopted design element in mobile user interface (UI). Defined as a visual change that is intentionally constructed within an interface [32], animation can expand the design space constrained by the physical form factor and technical features of mobile devices [24]. Initially introduced from cartoons [8], animation is proven to have the ability to improve mobile UI usability and efficiency [32, 25]. In recent years, researchers and designers have started to look more deeply into the role of animation in driving user engagement – a quality of user experience essential to user loyalty and long-term success [26] – in mobile interaction settings [1].

User engagement (UE) is defined as the level of a user's cognitive, affective, and behavioral investment when interacting with a digital system [39, 38]. Specifically, UE is an overarching construct that captures four dimensions of user experiences, *i.e.,* aesthetic appeal, focused attention, perceived usability, and reward [38]. Prior research suggests that the interface of an application compliments its content in producing an engaging experience; if the interface fails to engage users, the content may no longer matter [37]. Animation, as an integrated part of mobile UI, can potentially affect all four aspects of user engagement with the UI as a whole. For example, UI design with more animations is perceived as more aesthetic and engaging [18]. Likewise, a set of existing guidelines recommend using animation to notify UI status changes for improving website usability [25]. However, when embodied improperly, animation might pose a negative impact. Hong *et al.* highlight that only one animated item presented on a UI page can hold users' initial attention and thereby negatively affect their task performance and perception [22]. As such, having a means to assess user engagement with animation on mobile UI is essential for designers, especially given the limited guidelines available for mobile animation design.

Conventionally, designers evaluate their UI designs including the integrated animations through user studies with target users or design reviews with domain experts. While these evaluation methods are informative, they requires extensive efforts and time [56]. This issue motivates recent research efforts on generating automatic assessment of design to provide rapid, low-cost, and relatively reliable design feedback, especially in the early design stages [35]. There have existed computational models for measuring UI aesthetics [35], visual diversity [42], and brand perception [55], to name a few. Although effective in their designated domains, existing approaches may not be readily applied to the assessment of user engagement with mobile animation. First, the current methods only model user perception on a single static UI page without taking dynamic changes between consecutive UI pages into consideration [55]. Second, they tend to inspect only one aspect of user experience, which limits the comprehensiveness of the assessment [42]. Third, most existing works lack the capability of identifying reasons behind the prediction results generated by the model and thus shed limited light on design improvements [48].

∗ Work done during Y. Jiang and Y. Liu's internship at HKUST.

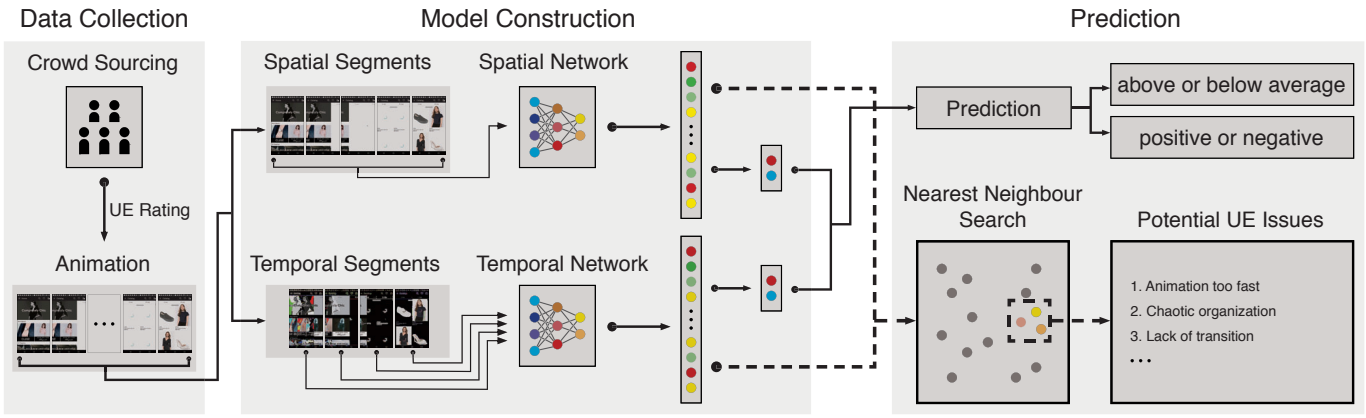

Figure 1. The overview of our study pipeline. We first crowdsource UE assessments of 1021 mobile UI animations. We then build a two-stream deep neural network model to predict UE levels of animation. We further leverage the animation feature encoded by our model and a sample set of expert reviews to derive potential UE issues of a poor animation design.

To bridge these gaps, we propose a data-driven approach to modeling overall as well as individual dimensions of user engagement with mobile UI animation. It can further automatically infer what aspects of an animation may potentially hinder user engagement. To this end, we first adapt a standard UE scale and collect a dataset by crowdsourcing user engagement assessments of 1021 mobile UI animations. We then train a deep convolutional neural network on our dataset to learn the features of mobile UI animation for predicting their UE levels. We consider UI elements (*e.g.,* UI images and controls) and their transitions as integrated parts of animation which may affect certain dimensions of UE with animation. Our method thereby incorporates those information for prediction by learning both the spatial and temporal attributes of animation simultaneously via a two-stream model. With our model, we achieve 71.4% precision on predicting "positive" UE (greater than the median of the UE scale) and 80.7% precision on predicting "above average" UE (greater than the mean of the crowdsourced UE scores).

Next, we take a step further to inform the possible animation design issues that can result in poor engagement, so that designers can reshape their products accordingly. Specifically, we examine a subset of the animations rated with low UE scores from our dataset with a group of UI/UX designers to identify the UE-related design issues of each animation. We calculate the similarity between a new animation and an animation from our subset using their feature representation encoded by our built neural network. We then associate the new animation with a list of potential UE-related design issues attached to its nearest neighbors according to the feature similarity.

Lastly, we design a proof-of-concept application which integrates our model to support designers and developers in investigating user engagement with their animation. We conduct interviews with five professional designers to envision the potential usage of this tool in actual mobile animation design settings, shedding light on the future development of more supportive design tools. To the best of our knowledge, this is the first work on computationally modeling user engagement with mobile UI animation[1].

Overall, this work makes three key contributions:

- We present the first data-driven approach to predicting user engagement with mobile UI animation by a large-scale crowd study and a computational model.
- We propose a novel deep neural network for learning the spatial and temporal information of mobile animation and predicting its user engagement level, which demonstrates a reasonably accurate performance. The model further allows us to infer potential UE-related design issues of typical animation, generating insights into UI design improvement.
- We develop a proof-of-concept design tool for assessing user engagement with mobile animation and conduct an informal study with designers to envision the use of our model in real design practices.

RELATED WORK

UI Animation

Animation has recently become a common element in mobile UI [51]. The functions of animated design elements within mobile UIs range from giving the user an illusion of interacting with a naturally behaving object [28] to drawing user attention [33]. Chevalier *et al.* list a total of 23 different roles of UI animations, organized into the five categories *i.e.,* keeping in context, teaching aid, user experience, data encoding, and visual discourse [10]. While it is generally agreed within industry and research communities that animation can be a productive aspect of UI designs, different studies have different definitions of animation [36]. For instance, Kraft *et al.* define animation as a series of varying images presented dynamically according to user actions in ways that help the user perceive a continuous change over time [28] while Thomas *et al.* take animation as a feature describing the spatial movements of UI elements [49]. In our work, we adopt the definition of animation as a visual change that is intentionally constructed within

[1]The project website is at `http://home.cse.ust.hk/~zwual/aniUE`

an interface [32]. This definition excludes visual changes caused by errors or lack of design, as well as lengthy predefined animated content such as videos or splash screens. The definition encompasses both large-scale transitional animations between visual states (switching application screens) and smaller visual changes (icon bounce upon selection). The authors further give insight into design practices by emerging existing guidelines of different platforms regarding animation design [32]. As pointed out, however, there has been limited research attention paid to animation in mobile interface design.

User Engagement with Animation

User engagement (UE) measures the degree to which users become cognitively and affectively focused on media content [39]. It affects user loyalty, long term usage, retention, and eventually the success of an application [26]. O'brien *et al.* measure UE by four dimensions, including focused attention, aesthetic appeal, perceived usability, and reward factor[43]. Animation turns out to be associated with all four aspects.

First, animations can guide the user's attention by effectively explaining visual changes to the user interface [46, 20]. According to Robertson *et al.* , interactive animations reduce the user's cognitive load by shifting it to the human perceptual system [30]. It assists users in tracking elements and understanding visual changes on the screen by making the transition between visual states smoother [30]. Second, animations can affect the aesthetic appeal of the user interface [50]. For instance, UI designs with more animations are often perceived as more aesthetic [18]. By integrating animations, it can improve the hedonic quality of the user experience [49, 52]. Third, animations can increase the perceived usability of a user interface. According to Chang and Ungar, when the motion of user interface elements is realistic and convincing, users can concentrate more on the task itself rather than the mechanics of the interface [8]. Last, animations make the user experience more rewarding, and therefore encourage users to continue interacting with the interface. As Thomas *et al.* suggest, interfaces with well integrated and appropriately used animations are rated as more motivating than those without [49]. Yet, inappropriate uses of animation may instead result in decreased user engagement. For instance, inappropriate animations can be distracting and therefore draw attention away from the task at hand [49]. Likewise, animations may potentially appear childish and therefore drive users away [49]. Hence, while there are benefits to integrating animations in user interfaces, they must be used appropriately to generate a positive effect.

Computational Assessment of UI

There have been works on modeling user perception and the subjective feedback of user interfaces, *e.g.,* the judgment of aesthetics [35], visual diversity [4], and brand perception [55]. A typical way to achieve this is by compiling a set of visual descriptors that depict a UI page, such as color, texture, and organization [55]. Then they collect user perception data at scale and construct the corresponding prediction models. The hand-crafted features, however, may not be able to portray all the aspects of UI, leaving room for improvement. Alternatively, deep learning (DL) has demonstrated its decent performance on learning representative features based on large-scale data [29]. For example, Zhao *et al.* adopt a convolutional neural network to predict the look and feel of graphic designs. In the domain of mobile UI design, DL-based methods have been applied to predict the perceived tappability of interfaces [48] and user performance of menu selection [31]. Despite the vital role of UE in design, computational methods for modeling UE remain under-explored. Some relevant studies infer UE from logged interaction data [7] or from textual information of social media data [45], but have not investigated the effect of visual stimuli on UE. Meanwhile, none of the aforementioned works, whose targets are static UI pages, can be generalized for mobile animation as they do not consider dynamic UI changes. Moreover, while most of the existing studies only provide prediction scores, they lack the capability of offering detailed feedback to facilitate the communications with designers [44]. To bridge these gaps, we present a novel deep learning model which can predict UE level with mobile animation and further provide feedback on potential UE issues of animation design.

VALIDATING THE ADAPTED USER ENGAGEMENT SCALE

User Engagement Scale

The User Engagement Scale (UES) is a standard survey widely adopted to measure user engagement [38]. We employ the short version of UES, which consists of 12 statements measuring four dimensions of engagement. The 12-item scale is originally proposed for website design and tested on e-shopping experience. Though with a certain level of generality, not all of the current items are entirely suitable for evaluating mobile animation. Therefore, we adjust some of the items to fit our context. In particular, for measuring focused attention, we do not include any item measuring long-term experience due to the short duration of mobile animation. We instead replace such an item *"I lost myself in this experience."* with *"This animation holds my attention."* used by [54]. Further, we modify some wording of the original statements to fit our context. The detail of the adapted UES is shown in Table 1.

Study Description and Data Analysis

To provide a structural validation of the adapted UES, we deploy an online study on Amazon Mechanical Turk [2] with 378 crowd workers. The participants are restricted to US residents to avoid cultural difference. We first ask the participants about their basic demographic information including gender, age, and the experience with mobile application. Then, we assign seven mobile animations, randomly sampled from Rico[3] [13], to each participant to rate using the adapted UES questionnaire. The animations are sampled in a way that they cover the seven common types of animations [36]. We show the participants a starting page of an animation with the user interaction (e.g., a tap or scrolling) highlighted and then an animated gif of the consequent effect. Such setting is to mitigate possible effects introduced by other non-animation-related factors, e.g., the function of the app, individual behavior habits, etc. Also, presenting interaction flow with animation effects in the form of a video is a common practice in app store (e.g. Google Play) to attract potential users. Once respondents view an animation, they respond to 13 user engagement questions, including

[2]https://www.mturk.com/

[3]http://rico.interactionmining.org/

Dimension	Item	Factor	Cronbach's α	Mean	St.Dev.
Focused Attention (FA)	This animation held my attention.	0.643	0.739	2.084	1.093
	I was absorbed in this animation.	0.898			
	I was so involved in this experience that I lost track of time.	0.836			
Perceived Usability (PU)	I felt frustrated while interacting with this animation.	0.967	0.829	1.671	1.061
	I found this animation confusing to interact with.	0.815			
	I felt annoyed while interacting with this animation.	0.868			
Aesthetic Appeal (AA)	This animation was attractive.	0.796	0.858	2.491	0.916
	This animation was aesthetically appealing.	0.748			
	I liked the graphics and movement of this animation.	0.756			
Reward Factor (RF)	Using this animation was worthwile.	0.644	0.805	2.346	0.938
	My experience with the animation was rewarding.	0.794			
	I felt interested in this experience with the animation.	0.729			

Table 1. Details and statistic description of the user engagement scale adapted from [38].

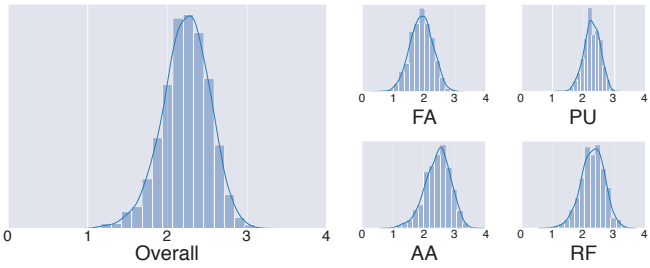

Figure 2. The user engagement score distribution of the collected animation. The score is in 5-point Likert scale from strongly disagree to strongly agree (0 being "strongly disagree").

a pair of quality control questions to filter out unreliable responses, by indicating their level of agreement on a 5-point Likert scale from "0. strongly disagree" to "4. strongly agree". We screen out submissions which: 1) provide contradictory answers to quality control questions, or 2) present answers in some systematic patterns such as "0 1 2 3 4". To eliminate potential bias caused by prior experience, in the following analysis, we exclude the responses from workers who reported to have used the associated applications before. Eventually, we secure the ratings from 215 participants (136 males and $age_{mean} = 30.98$), which meets the recommended sample size (i.e., greater than 200 participants) for factor analysis [23].

Bartlett's test shows the significant factorability of the collected data ($\chi^2_{df} = 332.48, p < .001$), and the average Kaiser-Meyer-Olkin factor adequacy measure indicates good suitability (overall $MSR = 0.83$). To test the validity of the items, we conduct confirmatory factor analysis (CFA) using structural equation modeling. The Cronbach's alpha values for all items are above the recommended 0.7 value, indicating good reliability. We thus conclude that the adapted 12-item scale is valid for measuring user engagement with mobile animation.

MODELING MOBILE ANIMATION ENGAGEMENT

Data Preparation

For constructing the computational model, we collect a large scale of user engagement data regarding mobile animation measured by the adapted UES.

Data Collection

We perform a similar crowdsourcing experiment as described in the previous section (Sec.3.2). The difference in this experiment is that we randomly sample another 1021 animations from the apps in Rico dataset. Rico contains apps of diverse quality (with different user ratings) from Google Play. The animations chosen from different apps cover seven common animation types [36]. Following the practice described in Sec.3.2, each animation is rated by at least five crowd workers on the adapted user engagement scale (as shown in Table 1). Each worker is assigned to three animations. The animation duration ranges from 1s to 12s. After removing unreliable responses based on the aforementioned criteria, we eventually secure 5106 individual assessments from 1702 participants (970 males and $Age_{mean} = 32.18$). For each animation, we aggregate the scores from different raters by majority voting.

Crowd Data Analysis

Figure 2 illustrates the distribution of overall user engagement scores across the selected animations (Mean = 2.28, SD = 0.64). Among the four dimensions of user engagement, FA has the lowest average score (Mean = 1.94, SD = 0.67) whereas AA is the most positively rated (Mean = 2.51, SD = 0.51). The scores of PU (Mean = 2.31, SD = 0.70) and RF (Mean = 2.34, SD = 0.49) sit in the middle.

We further examine whether the raters' perception of UE with animation is consistent. We adopt the intraclass correlation coefficient (ICC) [5], a standard measure for quantifying the degree to which a fixed number of raters have consistent judgments. The result (ICC1k = .708; 95% conf. interval is .699 to .716; F(11231, 44928) = 3.424; p<.001) suggests a good consistency level in user rating on the measurement according to a standard guideline [11]. This demonstrates a high degree of reliability of the crowdsourced data, although there still exists a certain level of disagreement. The statistics also shows that the participants have the highest agreement on AA (ICC1k= .714) and the lowest consistency in PU ratings (ICC1k = .650), indicating a more compatible perception of aesthetics than on the other three dimensions.

Model Construction

To predict user engagement with mobile UI animations, we build a deep-learning-based computational model. An

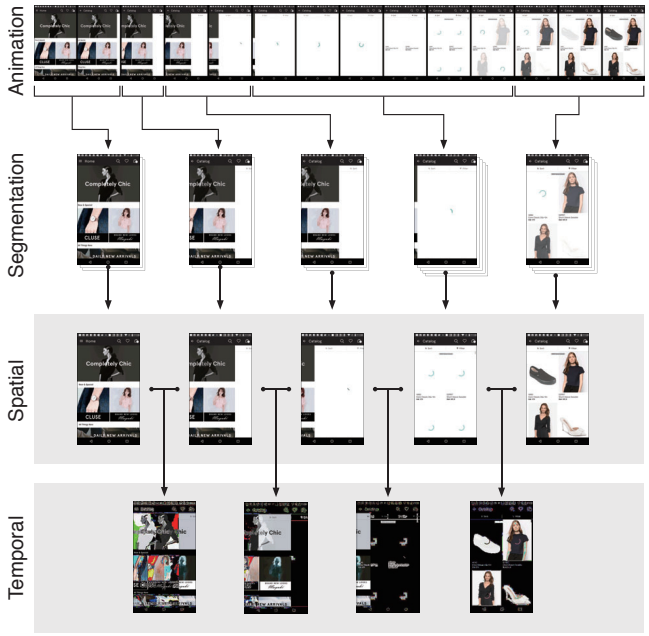

Figure 3. An illustration of our semantic temporal segment strategy. In this example, our algorithm divides a typical loading animation into five sequential segments according to their frame similarity. The result segments serve as the input of the spatial network while their differences between consecutive frames serve as the temporal input.

overview of the model is presented in Figure 4. Overall, we use a two-stream neural network, adapted from an existing method for video analysis [47], which accepts spatial and temporal descriptors of a given mobile UI animation as input and produces the prediction of its user engagement level as output.

Network Input: Spatial and Temporal Descriptors

Inspired by [47], we utilize a two-stream neural network structure which decouples the source stimuli into spatial and temporal representations for better modeling performance. More specifically, we design our model to accept the two following features of mobile UI animations as input: 1) an encoding of spatial features such as color and layout, and 2) an encoding of temporal features that describe the dynamic changes within and between UI pages.

The original TSN method developed for video data [47], however, cannot be readily applied to model UI animation for several reasons. First, prior works [47, 53] use a randomly sampling strategy that randomly selects video segments as the model inputs. When applied to mobile UI animations, this sampling strategy can lead to significant information loss as it might miss some of the key frames that are influential on user engagement perception. Second, a UI animation is usually composed of a series of actions of different durations. For example, a sign-in animation commonly consists of a long loading sequence followed by a short page switch. A random sequence sampling method is incapable of capturing such duration variations, which have the potential to significantly impact on user engagement [24]. Moreover, those randomly sampled segments may not capture notable transitions between key frames while prior research has indicated the critical role of transition in shaping user engagement [25].

To tackle these challenges, we introduce a semantic temporal segmentation strategy to extract the informative spatial and temporal features of a mobile UI animation. In particular, we segment the animation based on frame similarity. Each segment represents a subsequence of the animation with its duration measured by the number of frames it contains. Our algorithm for segmenting the sequences evolves from K-Means clustering. Upon the original K-Means algorithm, we impose an additional constraint to ensure that each segment is composed of consecutive frames and all segments maintain sequential order. Specifically, for a given mobile UI animation $A = \{a_1, a_2, ..., a_n\}$, where a_i denotes the i^{th} frame of A, we segment it into K segments $\{S_1, S_2, ..., S_k\}$. The average of the frames denotes the center of each segment. In each iteration of our algorithm, we traverse all the frames sequentially, comparing the distances between each frame and the segment centers to sort the frames into separate segments. In this way, two consecutive frames will be assigned to different segments if they have a large pixel-wise difference. Empirically we find that good initialization is essential for convergence. Therefore, we use K-Means clustering to initialize all our segment centers.

We use the centers of the resultant segments as our spatial input features; meanwhile, we take the RGB differences between the resultant segment centers as our temporal input. While optical flow is used as temporal feature for video data in the original TSN paper [47], we adopt RGB-diff since we empirically find that it works better than optical flow for animation in our task. An illustration of our feature encoding procedure is shown in Figure 3. The first row of this figure shows the raw animation frames with the second row illustrating the generated segments. The third row presents the center of each segment, which together serves as the input of the spatial network. Meanwhile, their differences, as displayed in the fourth row, will be taken as the input of the temporal network. In this presented example, the segmentation strategy captures different actions of the animation. The first screenshot depicts the initial page, the second and third capture a sliding behavior, whereas the fourth one illustrates a loading process, and the last one presents the post-loading action.

Model Architecture

Figure 4 presents our model architecture. We leverage a two-stream framework to predict user engagement with mobile UI animations. One of our network streams processes the spatial descriptor of animation, and the other processes the temporal descriptor. Their output are fused at the end as the final prediction result.

The network stream for processing spatial features consists of a 34-layer network, using a standard ResNet architecture [19] followed by an FC layer and Softmax activation and finally makes a loss-level prediction. The temporal network stream also contains a similar structure. Different from the spatial network, we include an attention mechanism in between the 34-layer ResNet and the FC layer. It is for allowing the model to attend to the more informative parts of the input [2]. With the attention mechanism, the operation on the extracted temporal

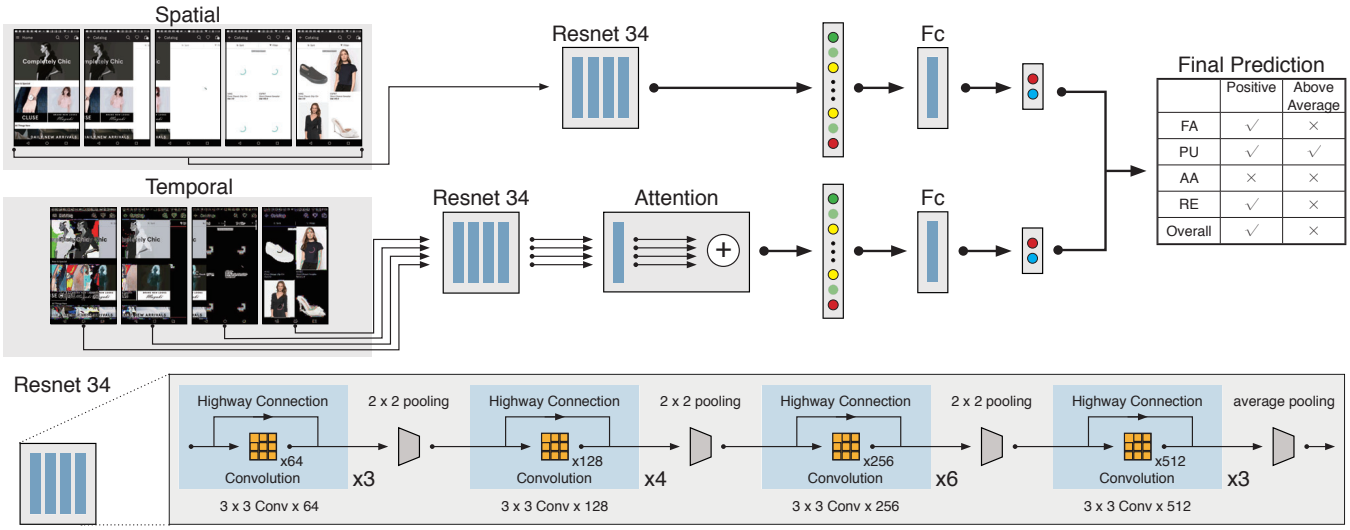

Figure 4. The overview of our proposed two-stream network for predicting UE with mobile UI animation. One of the Resnet network streams processes the spatial descriptor of an animation, and the other processes the temporal descriptor. Their output are fused at the end as the final result.

features can be represented as follows:

$$Out = H(G([F(d_1), l_1], [F(d_2), l_2], ..., [F(d_{K-1}), l_{K-1}])) \quad (1)$$

d_i denotes the extracted temporal features, which is equivalent to the difference between the segment centers of S_i, and S_{i+1}. l_i denotes the length of the corresponding segments. F denotes the ResNet feed-forward operation. G takes in the ResNet output plus the duration $[F(d_i), l_i]$ and combines all the segments with an attention mechanism in order to get a comprehensive interpretation of the UI animation sequence. H is an FC layer with Softmax activation that makes a loss-level prediction based on the outputs of the attention mechanism.

Let $f_i = [F(d_i), l_i]$ denote the ResNet output plus duration corresponding to the ith segment. The attention mechanism will produce a weight combination of the f_is as follows:

$$G(f_1, f_2, ..., f_K) = \sum_{i=1}^{K} \alpha_i f_i, \text{ where } \alpha_i = \frac{exp(wf_i)}{\sum_{j=1}^{K} exp(wf_j)} \quad (2)$$

w denotes a weight, which will be learned in the training process. With the attention mechanism, the network is able to assign a higher weight to the more critical temporal features.

Model Training

We first filter out the animations with inconsistent crowdsourcing ratings (the standard deviation is above top 5%). This leaves us with 997 animations in our dataset. We label all animations with a rating above the mean score as "above average" and those below as "below average". We also label all animations with a rating above the median of the UE scale as "positive" and below as "negative" to provide another classification of UE level. 71% of the animations is labeled as "positive". These two classifications provide two different evaluation perspectives. The former indicates whether the UE of a given animation is generally positive in terms of the scale, whereas the latter shows whether it is better than half of the stimuli in the entire dataset.

We randomly split the whole dataset into a training set (60%), a validation set (20%), and a testing set (20%). Each animation is resized to 280×158. To augment the training data size for avoiding overfitting, we randomly crop each animation into the size of 224×126 and perform horizontal frame flipping. The Softmax activation function we include as part of the last layer generates a predicted probability. The probability indicates how likely the model perceives the animation as "above average" (or "positive"). The spatial and temporal streams are trained separately by minimizing the cross-entropy loss between the predicted values and our labels. We also incorporate a triplet loss to boost the loss convergence [21]. After training, we fuse the two network streams together by averaging their output. For loss optimization, we use the Adam optimizer [27] with a learning rate of 0.001 and a batch size of 64. A dropout ratio of 50% and weight decay with rate 0.0001 are applied to alleviate model overfitting. We build our model using Pytorch [40] and train it on two Nvidia GTX 1080 Ti GPUs. The training process is terminated at the epoch when the models achieve the optimal performance on the validation set. Then the trained models are evaluated on the test set.

Model	Precision	Recall	F-score
TSN-Spatial	63.0%	59.4%	61.2%
TSN-Temporal	57.9%	62.3%	60.0%
Our-Spatial	67.3%	62.3%	64.7%
Our-Temporal	68.1%	60.4%	64.0%
TSN-Fusion	60.2%	69.8%	64.6%
Our-Fusion	**71.4%**	**70.0%**	**70.7%**

Table 2. Performance of proposed models compared to the baseline (TSN [47]) on predicting whether UE with an animation is above average

Model Evaluation

In Table 2 and 3, we compare our models to the original two-stream network (TSN) [47] in terms of predicting whether UE with an animation is "above average" and whether UE with an animation is "positive", respectively. To quantify the model

Model	Precision	Recall	F-score
TSN-Spatial	79.4%	71.8%	75.4%
TSN-Temporal	78.9%	57.7%	66.7%
Our-Spatial	79.7%	92.3%	87.0%
Our-Temporal	79.0%	91.7%	84.9%
TSN-Fusion	79.3%	73.3%	78.0%
Our-Fusion	**80.7**%	**96.8**%	**88.1**%

Table 3. Performance of proposed models compared to the baseline (TSN [47]) on predicting whether UE with an animation is positive

performance, we provide the standard metrics of precision, recall, and F-score ($\frac{2*precision*recall}{precision+recall}$) [14].

For predicting "above average" UE, our proposed spatial-only network and temporal-only network both outperform the single-stream baselines for all three metrics. Compared to the single stream models, our fusion model achieves an even higher precision, recall, and F-score (71.4%, 70.0%, and 70.7% respectively). It also significantly outperforms the baseline fusion model which only reaches a precision of 60.2%, a recall of 69.8%, an F-score of 64.6%. Additionally, we compare the temporal attention models with and without the incorporation of the segment duration. The results shows that the one with segment duration performs better (2.6% improvement on F-score), indicating the benefit of animation timing on UE prediction. Likewise, for predicting "positive" UE with animation, our proposed spatial-only and temporal-only network stream models both beat the single-stream baselines on all three metrics. Our fused model also performs better than the baseline fused model across the board. Meanwhile, the precision of predicting the "negative" class is 93.1%. In all, we show that user perception of UE with mobile animation can be computationally predicted with a reasonably accurate performance. Moreover, the comparative results indicate the effectiveness of our proposed sequential sampling strategy and attention mechanism in capturing animation characteristics for inferring UE levels. While the average fusion brings significant improvement for our proposed models, such effect on TSN baselines is much weaker. It is possibly because with the poorer performance of TSN, the outcome contains more noise and thus achieves less improvement after fusion.

We also investigate our model's performance across the four UE dimensions: focused attention (FA), perceived usability (PU), aesthetic appeal (AA), and reward factor (RF). We train

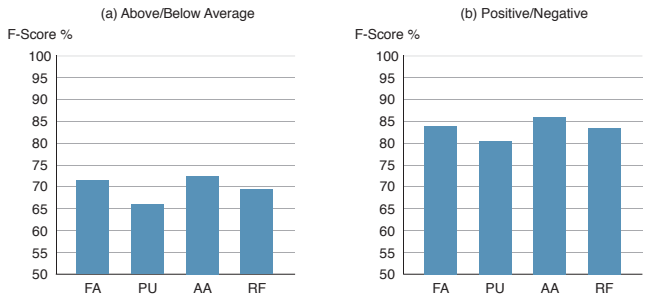

Figure 5. Our fusion model's F-score performance across different UE dimensions.

and evaluate our model on each of the four dimensions using the same settings detailed in the *Model Training* section. The results are presented in Figure 5 for predicting "above average" and "positive" UE. It shows that in both settings, our fusion model yields the highest F-score on predicting Aesthetic Appeal (AA) and the lowest for Perceived Usability (PU). One possible explanation is that the user ratings we collect for the AA of an animation are more consistent than those of the PU, as reported in the previous section (Section 4.1.2). Given more varied and ambiguous data, it is naturally more difficult for the a computational model to make accurate predictions. Another possible explanation is that while AA is perceived purely from the appearance of an animation, PU is also related to user interactions [18]. Since our model makes predictions based entirely on spatial and temporal features extracted from animations, users' experience with actually interacting with the interface, which is a component of PU, may be more difficult to encode than just its visual appearance.

To better understand our model's effectiveness on characterizing mobile animations, we visualize the features of the animations in our dataset embedded by our model. Specifically, we collect the model's 512-dimension output before the last FC layer as the encoded feature representation of each animation. We then project them onto a 2-D space using the tSNE method [34], an optimization-based dimensionality reduction technique. tSNE is suitable for visualizing high-dimensional data as it can reliably represent the distances between data points [34]. The result is shown in Figure 6. Examining the data points, we find that similar types of animations tend to appear in close proximity to one another. For instance, the points representing loading animations generally tend to be neighbors while locating far apart from the points for sliding animations. This suggests that the feature representation of the animations generated by the model may potentially be used to differentiate between distinct types of animations.

IDENTIFYING POTENTIAL UE RELATED DESIGN ISSUES

From designers' perspective, just being able to assess whether an animation is engaging may not be insightful enough to get informed on how to reshape their animation design. Therefore, we propose to automatically identify the potential UE issues of a poor animation design. The most straightforward method to achieving this goal is by analyzing the features that contribute to the model. However, this means is yet to be practically effective due to the black-box nature of deep learning models [16]. Alternatively, we can train another model to predict the reasons leading to poor UE with animation or mine user reviews from app store to infer UE issues. However, these methods both require a large dataset of high-quality expert reviews or online app comments. Instead of the above options, we adopt a weakly supervised method which utilizes the feature representations encoded by our UE prediction model and the data from a small scale expert interview. It is based on the manifold assumption that the animations sharing similar features in constructed high-level space bare similar strength and weakness [57]. Overall, we first conduct expert interviews to identify the reasons why animation from a subset of our collected data has poor user engagement. Based on the feedback, we compile a list of potential issues for each animation in the

Focused Attention (FA)	Potential UE Issue	Count	Percentage	Quote
animation	redundant animation	49	11.98%	The topmost navigation bar does not have to move.
	chaotic organization	22	5.38%	It looks chaotic when the last page slides left while the new slides up.
	animation too slow	14	3.42%	The keyboard slides in too slowly.
	animation too fast	13	3.18%	The animation is too fast to see clearly.
	animation not salient enough	3	0.73%	The animation is too subtle to notice.
component	component not salient enough	26	6.36%	The loading ring is hard to notice, making users lose patience.
	distracting component	11	2.69%	The jumping top right icon is distracting.
page	overwhelming information	11	2.69%	There are too much information in the same page.
	redundant page contents	3	0.73%	There is no need to show contents of previous page in the background.
Perceived Usability (PU)	**Potential UE Issue**	**Count**	**Percentage**	**Quote**
animation	misused animation	31	7.58%	The animation does not clearly indicate the hierarchy between pages.
	lack of feedback on operation	8	1.96%	There is no clear feedback after clicking a button of the side menu.
	improper movement	6	1.47%	Pages of the same hierarchy should move in unified directions.
	abnormal sequence	3	0.73%	The order in which the menu and prompts appear is very strange.
component	misleading or confusing components	15	3.67%	The icon of 'pen' is incorrect.
	absence of components	14	3.42%	The loading page is twinkling ,lack of place holding component.
	redundant component	12	2.93%	It is not necessary to click 'back' to the previous page.
	improper location of components	11	2.69%	The alert message should not appear on the top of the screen.
	overlapping components	6	1.47%	The navigation bar slides in to overlap another button.
page	ineffective layout of one page	16	3.91%	Unclear hierarchical relationship between the front and the back pages.
	unclear indication between pages	7	1.71%	It is better to show where you came from when you quit.
	inconsistent layouts among pages	3	0.73%	The page layouts of different pages are too inconsistent.
Aesthetic Appeal (AA)	**Potential UE Issue**	**Count**	**Percentage**	**Quote**
animation	not aesthetically appealing	7	1.71%	The fade out animation is not good-looking here.
	unrealistic animation	3	0.73%	The shadow of clicked button should become smaller instead of bigger.
component	improper graphic style of components	4	0.98%	The shadow of components is too heavy.
page	inconsistent graphic styles	6	1.47%	The graphic styles among pages are inconsistent.
	improper color design	6	1.47%	The color of the button is too saturated and extreme.
	misalignment	3	0.73%	The alignment is weird. It is better to be lefft aligned.
Reward Factor (RF)	**Potential UE Issue**	**Count**	**Percentage**	**Quote**
animation	animation not smooth enough	36	8.80%	The animation of the tab bar below is not very smooth.
	sudden change without animation	34	8.31%	Pages look isolated joined by a sudden black page.
	animation too long	6	1.47%	The loading animation is too long.
	lack of novelty	4	0.98%	This can be more special since it is not on the primary user activity path.
	animation too short	3	0.73%	The movment distance is too short.
component	absence of components	7	1.71%	The title disappeared in the new page.
page	too blank and lack of design	4	0.98%	The pages without content can also be designed (such as adding an illustration).
	misleading interaction logic	2	0.49%	The dead-end page is disappointing.

Table 4. The summary of the UE related issues given by five expert designers.

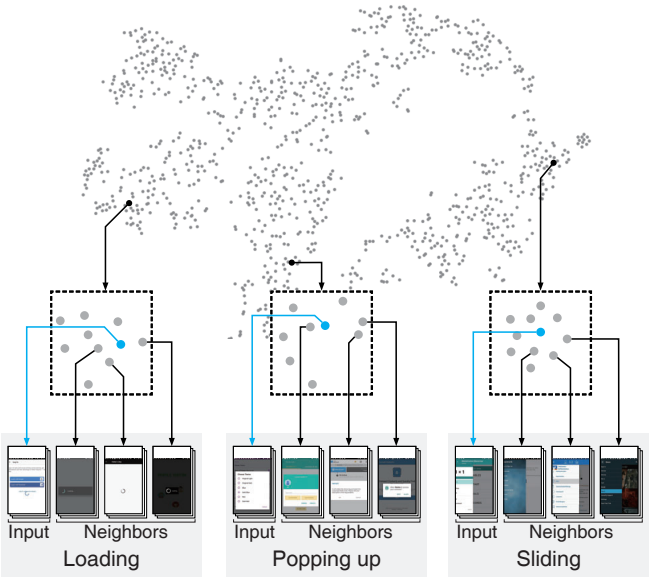

Figure 6. tSNE visualization of the embedded animation features. For a given animation, its neighbors tend to be the similar type of animations while those distant from it are often significantly different animations.

subset. Given a new animation with poor user engagement, we then associate the animation with a list of potential issues using nearest neighbor search in terms of feature similarity.

Expert Interviews

We conduct semi-structured online interviews with five professional designers. They are recruited from technology companies via personal social media or word of mouth and are required to have at least two-year industrial experience with UI/UX design projects. The interview with each participant lasts around two hours. From our original dataset, we sample 100 animations with low user engagement scores. In the interview, we present these animations together with their crowdsourced UE scores one by one to the participants, and ask them to comment on the potential reasons why the animations were not engaging to users. We perform thematic analysis [6] on the interview transcripts and compile the feedback into a list of potential UE issues for the sampled animations. A summary of the final themes is presented in Table 4.

Specifically, we organize the identified UE issues around the four dimensions of UE (*i.e.,* FA, PU, AA, and RF), and then into three sub-categories *animation*, *component*, *page*. *Animation* refers to UE issues associated with the animation itself (*e.g.,* speed and direction of movement). *Component* refers

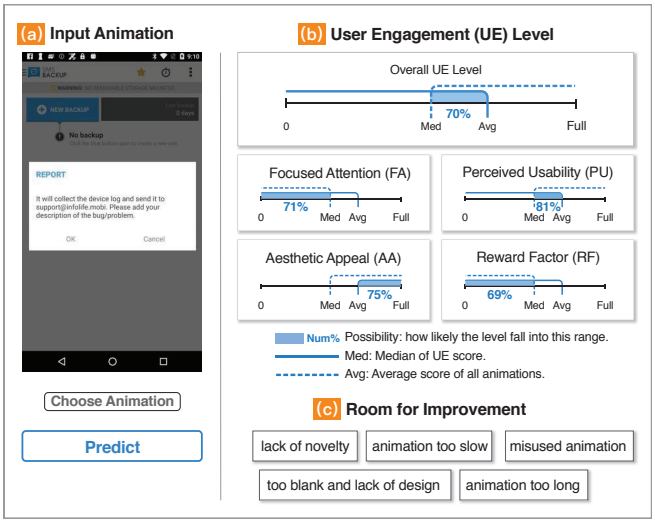

Figure 7. The interface of AniLens. Designers can upload an animation on the left panel (a). Then the corresponding results including the predicted UE level (b) and potential UE related design issues (c) will show up on the right.

to UE issues related to the UI components (*e.g.,* the graphic style of the button), whereas the UE issues associated with a UI page or the transition between UI pages fall into *Page*. In Table 4, we sort the UE issues listed within each subcategory according to their frequency mentioned by experts. Across the three subcategories, *animation* related UE issues are discussed the most by the expert participants. The most frequently mentioned issues in the FA, PU, AA, and RF are "redundant animation," "misused animation," "not aesthetically appealing enough," and "animation not smooth enough" respectively. More details can be found in Table 4.

Inferring Potential UE Issues

Given a new animation with a low predicted UE level, we embed it into the feature representation generated by our model. The feature encoding is the 512-dim output before the last FC layer of the network. With the feature vector, we search the closest neighbor examples in the sample set based on L2 distance and return the top UE related issues associated with the neighbors. To evaluate the effectiveness of this approach, we perform a five-fold cross validation across the collected review dataset with 80% training and 20% validation data. The results show that our method can achieve 60.6% accuracy for Top-5 performance, which significantly outperforms random guessing (14.3%). This reveals that the introduced method can effectively infer the potential UE-related issues of animation with a reasonable accuracy. As such, our model has the potential to help designers improve their animation designs. Despite the presented effectiveness of our method, we are aware that the selected datapoints may not be sufficient enough to cover all the cases. In future studies, it is worth evaluating how our method performs with an extended dataset.

A DESIGN TOOL AND ITS EVALUATION

We develop AniLens, a proof-of-concept web application with our built-in model for designers and developers to assess their mobile animation design. The interface is implemented in Javascript supported by a Python back-end server. It takes a mobile animation in GIF format as input and outputs the model's predicted engagement results (whether "positive" and whether "above average") on the overall UE as well as each of the four dimensions. If the animation receives a "negative" or "below average" assessment, AniLens further generates a list of five UE issues potentially associated with the given animation. The interface of AniLens is shown in Figure 7.

To attain overall feedback of AniLens and envision its potential use in actual design settings, we conduct semi-structured interviews with another five mobile UI/UX designers (D1-D5). Following the same process mentioned in Sec.5.1, we recruit the participants who have at least two-year mobile UI/UX design experience in industrial companies via personal invitation or word of mouth. During the interview, we walk them through AniLens and collect informal feedback on its two main features: predicting UE levels of an animation and providing suggestions on fixing the animation design predicted to have low UE. We also ask the designers to envision how they would utilize such a tool in actual design practice. Overall, the designers are excited about AniLens and share their expectation of improvement. Particularly, we identify the following themes through thematic analysis [6].

Potential Usage in Design Practice

The designers respond positively towards the two features provided by AniLens. For example, D2 expresses his interest in the tool that, *"I like this idea of getting automatic feedback. It would be very inspiring for me to know how users might perceive my animation designs, especially for the ones I'm not confident about"*. D3 adds that, *"I would love to use the tool for quick sanity check to see if the design has some basic faults."* The designers further mention that they can easily foresee various uses of AniLens in their design practices. First, AniLens could help them investigate different dimensions of animations for a comprehensive diagnosis, *e.g., "in a navigation app design, perceived usability is more essential than appeal. So with the tool, I would be more strict on usability levels while its aesthetics just needs to meet the baseline."*(D4) Second, the results could be used as a source for design validation. *" When I want to convince clients or managers of my animation design, it could be extremely helpful to show them the positive prediction results."*(D1) Third, the tool could aid designers in fixing their problematic animation by prioritizing the possible UE issues. *"With an unshaped animation design and a tight project deadline, I can get quick hints from the tool on which aspect should be primarily focused on. I will first look at the worst dimension in the prediction and find the related issues from the suggestion list."*(D5).

Directions for Tool Improvement

Our participants point out several directions for improving AniLens. Two of them (D2 and D5) hope that the tool could also take in a batch of animations from an application as a whole and generate an overall UE prediction. It is for ensuring the high *"consistency and smoothness of consecutive animations"* (D3). Such capacity could be achieved by detecting

individual animations from the batch and aggregating their predicted UE levels. Meanwhile, D5 demands more fine-grained prediction results. While AniLens only provides the predicted UE levels, D5 hopes to get precise UE value as it can facilitate direct comparison between two animations falling into the same UE level. We could extend AniLens by integrating another prediction model for regressing UE value, which however, requires vast high-quality crowd data with balanced UE score distribution. Moreover, D4 wishes AniLens to present some examples that have the same UE issues as the input animation. Because adding such examples can allow more intuitive understanding of the identified issues than presenting textual information alone, avoiding *"improper interpretation of the results"*. To achieve so, we could augment AniLens in a way that it will display several similar animation examples from the expert annotated pool labelled with the same issue when a user clicks on one of the suggested UE issues. Besides, D1 and D2 mention that in actual practice, an animation can be saved in different formats, such as animated images in GIF, a sequence of images in JPEG, and even video formats (*e.g.,* AVI and MPEG4). They suggest that AniLens should afford different formats of animation as input, which could be done by incorporating a format conversion function into AniLens.

DISCUSSION & FUTURE WORK

In addition to the usage as an assessment tool mentioned in the previous section, our computational model can also be applied in many other contexts. For instance, our model can automatically suggest a high-quality reference set of exemplary animations, so that designers do not need to manually screen out bad animation examples when searching for design materials online. Additionally, we see the development of our model as a step towards automatic generation of UI animation designs. The prediction results of our model can potentially serve as a design criterion in the objective function of a generative model. Moreover, our methodology for building a UE predication model for mobile animation is easily extensible to other design domains where animation plays a critical role (*e.g.,* web design and AR/VR applications).

Although our model is reasonably accurate, we acknowledge that there is still room for improvement to support wider application scenarios. First, our model may not factor in all the characteristics of mobile animation which can affect user engagement. For example, our current model cannot well capture semantic information contained within UI. While prior research has demonstrated, for instance, that images of human faces have an effect on user engagement [3], our current model cannot handle such indicators and has lower precision in this case. Likewise, as discussed earlier in *Model Evaluation* subsection, our model only considers spatial and temporal features of a mobile animation, while perceived usability also depends on the user's interactions with the interface. One potential solution to this issue would be by incorporating more relevant features, *e.g.,* face descriptors and logged user interaction data, into our engagement prediction model. Second, despite promising performance with the embedded 34-layer ResNet, in the future, we can consider trying other common network architectures (*e.g.,* the deep recurrent neural nets prior research has shown decent performance on time-series sequential data [15]). Further, as noted in the data collection section, the UE ratings we collected are not always consistent. Although we aggregate the judgments from five raters to alleviate potential subjective biases, there are other possible means to deal with inconsistent data for improving model performance. For example, in addition to using a 5-point Likert scale, we may ask crowd workers to perform pairwise comparisons between UI designs in order to avoid individual scale differences [12]. We could also investigate whether existing methods for handling noisy and subjective data, such as bootstrapping [41] and co-teaching [17], can boost the performance of our model.

There are also several directions for future work. Besides user engagement, there exist other qualities that define user experience with mobile animations, such as time perception [24] and brand perception [55]. We might apply the same methodology to computationally model these qualities. Besides, mobile UI animations in our current dataset are all from Android applications. We are interested in exploring whether our model can be generalized to other platforms, such as iOS applications. Moreover, we simulate the experience with animation by presenting participants the animation effect caused by an user interaction instead of asking them to actually interact with the animation. Although such setting is to mitigate possible effects introduced by non-animation-related factors, we plan to investigate the potential gap between the simulated and actual experience. Last but not least, prior work has highlighted that the use of design tools varies across designers with different level of expertise [9]. We speculate that novices will rely more on our tool compared to experts. Studying such behavior differences in using our tool can guide us to provide more customized support. It would be insightful to evaluate Anilens in the actual design settings with both experts and novices.

CONCLUSION

In this paper, we present a data-driven approach to assisting designers in examining user engagement (UE) with their mobile UI animation designs. We first crowdsource a UE assessment dataset of 1021 animations based on a validated UE scale. Then based on the collected data, we design and train a deep model that performs reasonably well in predicting the level of UE with mobile UI animations.To further aid designers in improving their mobile UI animations, we allow the automatic identification of potential UE related issues of an animation by utilizing the animation feature encoded by our model and data from small-scale expert interviews. Finally, we build AniLens, a proof-of-concept tool for designers and developers to assess and improve UI animation designs. We evaluate the potential usage of AniLens in real-world design settings with five professional UI/UX designers, who respond positively. Overall, our work can benefit designers by enabling automatic animation design assessment and provide the implications on the future development of computational design tools.

ACKNOWLEDGMENTS

This work is supported by the Research Grants Council of the Hong Kong Special Administrative Region, China under Grant No.: C6030-18G.

REFERENCES

[1] 2016. UX animation: When to use it, when to lose it. https://techbeacon.com/app-dev-testing/ux-animation-when-use-it-when-lose-it. (Febuary 2016).

[2] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473* (2014).

[3] Saeideh Bakhshi, David A. Shamma, and Eric Gilbert. 2014. Faces Engage Us: Photos with Faces Attract More Likes and Comments on Instagram. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 965–974. DOI: http://dx.doi.org/10.1145/2556288.2557403

[4] Nikola Banovic, Antti Oulasvirta, and Per Ola Kristensson. 2019. Computational Modeling in Human-Computer Interaction. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, W26.

[5] John J Bartko. 1966. The intraclass correlation coefficient as a measure of reliability. *Psychological reports* 19, 1 (1966), 3–11.

[6] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. DOI: http://dx.doi.org/10.1191/1478088706qp063oa

[7] Jonathan Carlton, Andy Brown, Caroline Jay, and John Keane. 2019. Inferring User Engagement from Interaction Data. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*. ACM, New York, NY, USA, Article LBW1212, 6 pages. DOI: http://dx.doi.org/10.1145/3290607.3313009

[8] Bay-Wei Chang and David Ungar. 1995. Animation: from cartoons to the user interface. (1995).

[9] Aline Chevalier and Melody Y Ivory. 2003. Web site designs: Influences of designer's expertise and design constraints. *International Journal of Human-Computer Studies* 58, 1 (2003), 57–87.

[10] Fanny Chevalier, Nathalie Henry Riche, Catherine Plaisant, Amira Chalbi, and Christophe Hurter. 2016. Animations 25 years later: New roles and opportunities. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*. ACM, 280–287.

[11] Domenic V Cicchetti. 1994. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological assessment* 6, 4 (1994), 284.

[12] Rene V Dawis. 1987. Scale construction. *Journal of Counseling Psychology* 34, 4 (1987), 481.

[13] Biplab Deka, Zifeng Huang, Chad Franzen, Joshua Hibschman, Daniel Afergan, Yang Li, Jeffrey Nichols, and Ranjitha Kumar. 2017. 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)*. ACM, New York, NY, USA, 845–854. DOI: http://dx.doi.org/10.1145/3126594.3126651

[14] Cyril Goutte and Eric Gaussier. 2005. A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In *European Conference on Information Retrieval*. Springer, 345–359.

[15] Alex Graves, Abdel-rahman Mohamed, and Geoffrey Hinton. 2013. Speech recognition with deep recurrent neural networks. In *2013 IEEE international conference on acoustics, speech and signal processing*. IEEE, 6645–6649.

[16] David Gunning. 2017. Explainable artificial intelligence (xai). *Defense Advanced Research Projects Agency (DARPA), nd Web* (2017).

[17] Bo Han, Quanming Yao, Xingrui Yu, Gang Niu, Miao Xu, Weihua Hu, Ivor Tsang, and Masashi Sugiyama. 2018. Co-teaching: Robust training of deep neural networks with extremely noisy labels. In *Advances in neural information processing systems*. 8527–8537.

[18] Jan Hartmann, Alistair Sutcliffe, and Antonella De Angeli. 2008. Towards a Theory of User Judgment of Aesthetics and User Interface Quality. *ACM Trans. Comput.-Hum. Interact.* 15, 4, Article 15 (Dec. 2008), 30 pages. DOI: http://dx.doi.org/10.1145/1460355.1460357

[19] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[20] Jeffrey Heer and George Robertson. 2007. Animated transitions in statistical data graphics. *IEEE transactions on visualization and computer graphics* 13, 6 (2007), 1240–1247.

[21] Elad Hoffer and Nir Ailon. 2015. Deep metric learning using triplet network. In *International Workshop on Similarity-Based Pattern Recognition*. Springer, 84–92.

[22] Weiyin Hong, James YL Thong, and Kar Yan Tam. 2007. How do Web users respond to non-banner-ads animation? The effects of task type and user experience. *Journal of the American Society for Information Science and Technology* 58, 10 (2007), 1467–1482.

[23] Matt C Howard. 2016. A review of exploratory factor analysis decisions and overview of current practices: What we are doing and how can we improve? *International Journal of Human-Computer Interaction* 32, 1 (2016), 51–62.

[24] Jussi Huhtala, Ari-Heikki Sarjanoja, Jani Mäntyjärvi, Minna Isomursu, and Jonna Häkkilä. 2010. Animated UI Transitions and Perception of Time: A User Study on Animated Effects on a Mobile Screen. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 1339–1342. DOI: http://dx.doi.org/10.1145/1753326.1753527

[25] Kevin Keeker. 1997. Improving web site usability and appeal. *Retrieved May* 19 (1997), 2002.

[26] Young Hoon Kim, Dan J Kim, and Kathy Wachter. 2013. A study of mobile user engagement (MoEN): Engagement motivations, perceived value, satisfaction, and continued engagement intention. *Decision Support Systems* 56 (2013), 361–370.

[27] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[28] Jonas F Kraft and Jörn Hurtienne. 2017. Transition animations support orientation in mobile interfaces without increased user effort. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM, 17.

[29] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2017. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM* 60, 6 (May 2017), 84–90. DOI: http://dx.doi.org/10.1145/3065386

[30] Brenda Laurel and S Joy Mountford. 1990. *The art of human-computer interface design*. Addison-Wesley Longman Publishing Co., Inc.

[31] Yang Li, Samy Bengio, and Gilles Bailly. 2018. Predicting human performance in vertical menu selection using deep learning. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 29.

[32] Daniel Liddle. 2016a. Emerging Guidelines for Communicating with Animation in Mobile User Interfaces. In *Proceedings of the 34th ACM International Conference on the Design of Communication (SIGDOC '16)*. ACM, New York, NY, USA, Article 16, 9 pages. DOI: http://dx.doi.org/10.1145/2987592.2987614

[33] Daniel Liddle. 2016b. Emerging guidelines for communicating with animation in mobile user interfaces. In *Proceedings of the 34th ACM International Conference on the Design of Communication*. ACM, 16.

[34] Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of machine learning research* 9, Nov (2008), 2579–2605.

[35] Aliaksei Miniukovich and Antonella De Angeli. 2015. Computation of Interface Aesthetics. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 1163–1172. DOI: http://dx.doi.org/10.1145/2702123.2702575

[36] David Novick, Joseph Rhodes, and Wervyn Wert. 2011. The communicative functions of animation in user interfaces. In *Proceedings of the 29th ACM international conference on Design of communication*. ACM, 1–8.

[37] Heather L O'Brien. 2011. Exploring user engagement in online news interactions. *Proceedings of the American society for information science and technology* 48, 1 (2011), 1–10.

[38] Heather L O'Brien, Paul Cairns, and Mark Hall. 2018. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. *International Journal of Human-Computer Studies* 112 (2018), 28–39.

[39] Jeeyun Oh and S Shyam Sundar. 2016. User engagement with interactive media: A communication perspective. In *Why Engagement Matters*. Springer, 177–198.

[40] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. 2017. Automatic differentiation in PyTorch. (2017).

[41] Scott Reed, Honglak Lee, Dragomir Anguelov, Christian Szegedy, Dumitru Erhan, and Andrew Rabinovich. 2014. Training deep neural networks on noisy labels with bootstrapping. *arXiv preprint arXiv:1412.6596* (2014).

[42] Katharina Reinecke and Krzysztof Z. Gajos. 2014. Quantifying Visual Preferences Around the World. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 11–20. DOI: http://dx.doi.org/10.1145/2556288.2557052

[43] George G Robertson, Stuart K Card, and Jock D Mackinlay. 1993. Information visualization using 3D interactive animation. *Commun. ACM* 36, 4 (1993), 56–72.

[44] Ruth Rosenholtz, Amal Dorai, and Rosalind Freeman. 2011. Do predictions of visual perception aid design? *ACM Transactions on Applied Perception (TAP)* 8, 2 (2011), 12.

[45] Holly M Rus and Linda D Cameron. 2016. Health communication in social media: message features predicting user engagement on diabetes-related Facebook pages. *Annals of behavioral medicine* 50, 5 (2016), 678–689.

[46] Céline Schlienger, Stéphane Conversy, Stéphane Chatty, Magali Anquetil, and Christophe Mertz. 2007. Improving users' comprehension of changes with animation and sound: An empirical assessment. In *IFIP Conference on Human-Computer Interaction*. Springer, 207–220.

[47] Karen Simonyan and Andrew Zisserman. 2014. Two-stream convolutional networks for action recognition in videos. In *Advances in neural information processing systems*. 568–576.

[48] Amanda Swearngin and Yang Li. 2019. Modeling Mobile Interface Tappability Using Crowdsourcing and Deep Learning. *arXiv preprint arXiv:1902.11247* (2019).

[49] Bruce H Thomas and Paul Calder. 2001. Applying cartoon animation techniques to graphical user interfaces. *ACM Transactions on Computer-Human Interaction (TOCHI)* 8, 3 (2001), 198–222.

[50] Noam Tractinsky, Ohad Inbar, Omer Tsimhoni, and Thomas Seder. 2011. Slow Down, You Move Too Fast: Examining Animation Aesthetics to Promote Eco-driving. In *Proceedings of the 3rd International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '11)*. ACM, New York, NY, USA, 193–202. DOI: http://dx.doi.org/10.1145/2381416.2381447

[51] Marcus Trapp and René Yasmin. 2013. Addressing animated transitions already in mobile app storyboards. In *International Conference of Design, User Experience, and Usability*. Springer, 723–732.

[52] Barbara Tversky, Julie Bauer Morrison, and Mireille Betrancourt. 2002. Animation: can it facilitate? *International journal of human-computer studies* 57, 4 (2002), 247–262.

[53] Limin Wang, Yuanjun Xiong, Zhe Wang, Yu Qiao, Dahua Lin, Xiaoou Tang, and Luc Van Gool. 2016. Temporal segment networks: Towards good practices for deep action recognition. In *European conference on computer vision*. Springer, 20–36.

[54] Jane Webster and Hayes Ho. 1997. Audience engagement in multimedia presentations. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems* 28, 2 (1997), 63–77.

[55] Ziming Wu, Taewook Kim, Quan Li, and Xiaojuan Ma. 2019. Understanding and Modeling User-Perceived Brand Personality from Mobile Application UIs. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 213, 12 pages. DOI: http://dx.doi.org/10.1145/3290605.3300443

[56] Anbang Xu, Shih-Wen Huang, and Brian Bailey. 2014. Voyant: Generating Structured Feedback on Visual Designs Using a Crowd of Non-experts. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '14)*. ACM, New York, NY, USA, 1433–1444. DOI: http://dx.doi.org/10.1145/2531602.2531604

[57] Xiaojin Zhu, Zoubin Ghahramani, and John D Lafferty. 2003. Semi-supervised learning using gaussian fields and harmonic functions. In *Proceedings of the 20th International conference on Machine learning (ICML-03)*. 912–919.