AdSEE: Investigating the Impact of Image Style Editing on Advertisement Attractiveness

Liyao Jiang[1], Chenglin Li[1], Haolan Chen[2], Xiaodong Gao[2], Xinwang Zhong[2], Yang Qiu[2], Shani Ye[2] and Di Niu[1]

[1] Department of Electrical and Computer Engineering, University of Alberta [2] Platform and Content Group, Tencent

1. Introduction

Background

- Online and mobile advertisements are important elements in online platforms, e.g., e-commerce, and social media platforms.
- Many advertisements contain both visual and textual information to grab user attention.
- Previous studies find that appealing images lead to higher CTR and human faces in ads attract more user attention.

Questions

- Is there a linkage between advertisement popularity and image facial style editing?
- Can facial style editing (e.g. adjusting smile, hair, eye gaze direction, etc.) affect the popularity of online ads?

Our Approach

- We propose the AdSEE framework consisting of the CRP and GADE.
- We use a data science approach and verified the existence of the relationship between image style editing and ad popularity.

2. AdSEE Framework

We design the Advertisement Style Editing and Attractiveness Enhancement (AdSEE) framework, which consists of two parts:

- Click Rate Predictor (CRP), which takes an ad as input and predicts its averaged click rate.
- Genetic Advertisement Editor (GADE), which enhances the overall attractiveness of ads. Through editing the cover images with guidance from the CRP.

Examples of *ads* enhanced by AdSEE:

Example 1

***ad* text:** "What does a dimple on the forehead mean?"

Category: other

Example 2

***ad* text:** "How many steps can you think ahead in Go?"

Category: sports

2.1 Click Rate Predictor (CRP)

Our CTR predictor is the first to consider GAN-based facial latent as features, along with other image, text, and category features:

1. **Face Latent:** Persons are extracted with the SOLO instance segmentation model, then the Dlib Face Alignment algorithm extracts faces. The e4e Encoder (an inversion encoder for the StyleGAN2-FFHQ model) maps the extracted faces to latent codes.
2. **Image Embedding:** Pre-trained OpenImage and Sougou image embedding models convert cover images to embeddings.
3. **Text Embedding:** Using the pre-trained Bert-Chinese model to obtain high-quality text embeddings.
4. **Three Sparse Features:** The face count, ad category, and class labels (from the segmentation) are used as sparse features.

- All 6 features (3 sparse features, 3 dense features) are mapped to embeddings within the AutoInt model.
- We utilize the AutoInt model to learn both high-order and low-order feature interactions, which is based on the Self-Attention mechanism.
- The AutoInt model is selected through empirical evaluation of a series of SOTA recommender models.

2.2 Genetic Advertisement Editor (GADE)

- We perform **style editing** to the faces by perturbing the face latent code ***z*** along the latent editing directions ***n*** to change the style and semantics of the facial images (e.g. age, eye gaze direction, etc.)

$$z_i' = z_i + \alpha_i n$$

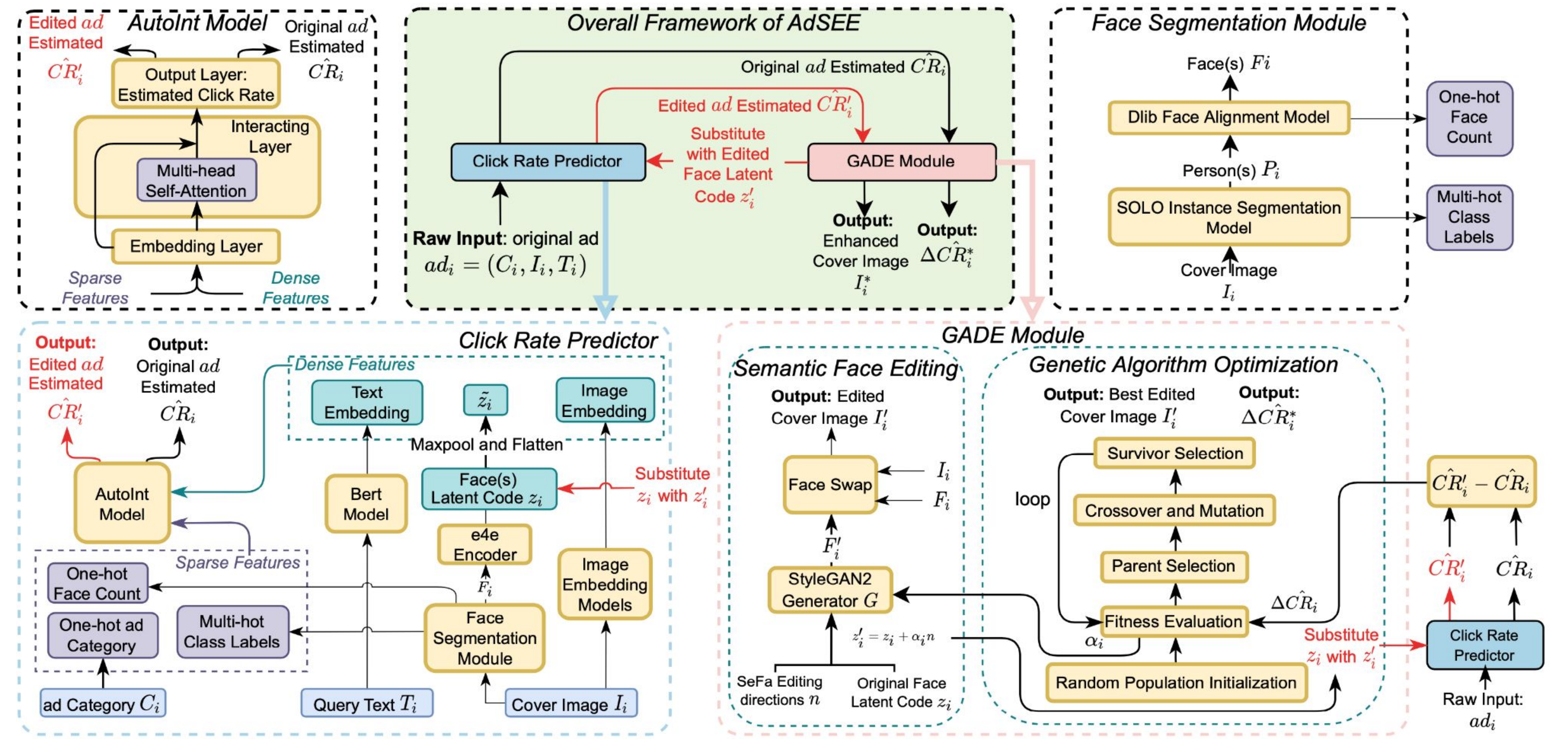
- We use the SeFa semantic factorization method to identify a set of *q* **latent editing directions *n***. SeFa method applies eigen-decomposition to find the most important semantic concepts, e.g., smile, age, etc.

$$n = \{n_p\}_{p=1}^{q}$$

- Different **editing intensity coefficients α** leads to different editing result, thus different predicted CR. The GADE module **finds the optimal facial editing coefficients α*** (i.e. the edit with the highest projected increase in CR) with genetic algorithm guided by Click Rate Predictor feedback.

- We use the StyleGAN2-FFHQ image generator *G* to **generate the edited face *F'*** from the edited latent code ***z'***. Utilizing the Face Swap algorithm from the OpenCV and Dlib functions to replace the original face *F* with edited face ***F'***.

$$F_{i,j}' = edit(F_{i,j}) = G(z_{i,j}') = G(z_{i,j} + \alpha_{i,j} n) \qquad I_i' = Swap(F_i, F_i', I_i)$$



3. Evaluation

Offline Evaluation

- Compare the proposed set of features (Combined) with other baselines with different types of features
- The results show the benefits of the proposed combined set of multi-modal features.

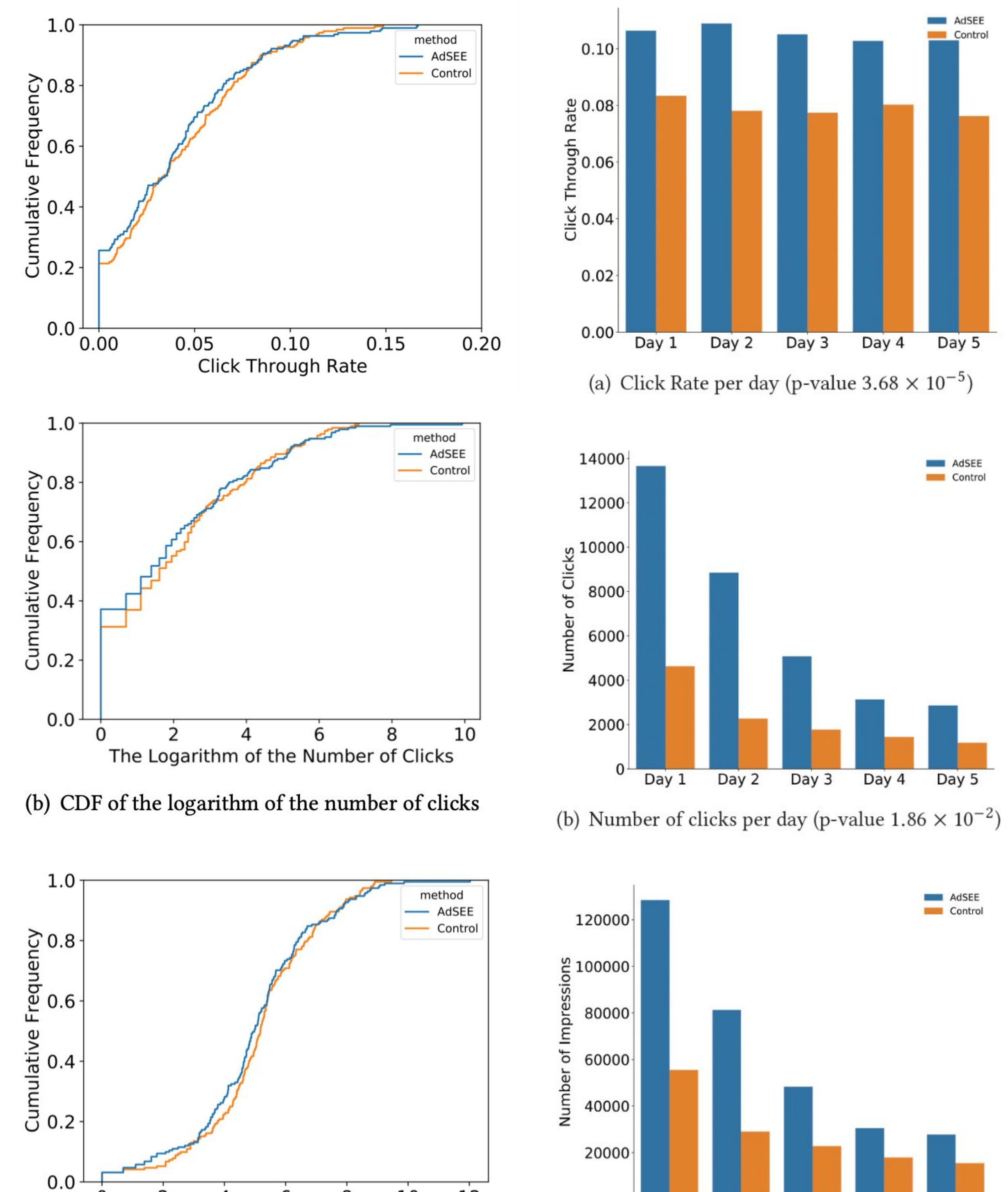
Model	Feature Type	MAE ↓	MAPE ↓	NDCG@10 ↑	NDCG@50 ↑
CRP-NIMA	Image Quality	0.0299	0.7456	0.2764	0.3917
CRP-OpenImage	Image Embedding	0.0295	0.7258	0.5950	0.5551
CRP-Sogou	Image Embedding	0.0299	0.7429	0.5095	0.5175
CRP-e4e	Face Latent Code	0.0306	0.7663	0.5149	0.5204
CRP	Combined	**0.0262**	**0.6542**	**0.6854**	**0.7337**

Model	Feature Type	Spearman's rho ↑	Kendall's tau ↑
CRP-NIMA	Image Quality	0.3634	0.2480
CRP-OpenImage	Image Embedding	0.3941	0.2696
CRP-Sogou	Image Embedding	0.3613	0.2464
CRP-e4e	Face Latent Code	0.2954	0.2003
CRP	Combined	**0.5122**	**0.3609**

Online Evaluation

- We perform **online A/B test for 5 days** over the QQ Browser mobile app:

1. The AdSEE edited group has **higher click rates** which show a significant increase in attractiveness to users.
2. In addition, the **higher number of impressions** shows the production recommender believes AdSEE-edited ads are more relevant and pushed to users more often.



(a) Click Rate per day (p-value 3.68×10^{-5})

(b) CDF of the logarithm of the number of clicks

(b) Number of clicks per day (p-value 1.86×10^{-2})

(c) CDF of the logarithm of the number of impressions

(c) Number of impressions per day (p-value 2.11×10^{-2})

4. Conclusion

- We designed a **Click Rate Predictor** that takes multi-modal (i.e. text and image) features including facial latent embeddings from a StyleGAN2 model.
- Our proposed **GADE module** can efficiently search for optimal editing directions with the genetic algorithm and feedback from CRP. In our paper, we provide insights on the attractive semantic editing direction to the users, e.g. facing downward, smile, feminine features.
- Through the **online experiments**, we verified the **existence of a correlation** between facial style editing and click rates in online ads.