Automatic image generation for application of advertisements generation based on diffusion models

Author: **Karol Piniarski**, Lingaro Group

Automatic Image generationfor defined group of clients

The paper presents an automatic image generation framework for marketing purposes. The main goal was to prepare high-quality advertising content for selected clients segment. For this purpose, the latest diffusion model architecture is used, which takes as input a text string describing a given customer segment. The text input is based on an analysis of content posted by users of the social networks.

Moreover, this framework allows to fine-tune the generative model to present a specific product on the generated advertisement.

In summary, the proposed framework is able to:
generate personalized advertising - for a given customer segment,
generate high-quality images - with super-resolution and reranking,
generate images of specific products - with additional model fine-tuning.

Fluence project

The proposed image generation framework is part of a larger Fluence project. The Fluence is an innovative analytical platform based on AI and machine learning principles that allows formulating marketing messages based on the analysis of content published by users of social networks. The platform is dedicated to entities planning and conducting marketing campaigns on the Internet and social media. Fluence significantly increases the effectiveness of marketing activities, allowing you to create, plan and settle promotional campaigns more effectively. This platform allows to segment clients, describe a specific group of customers and find product and brand features, as presented in the Figs. 1 and 2.

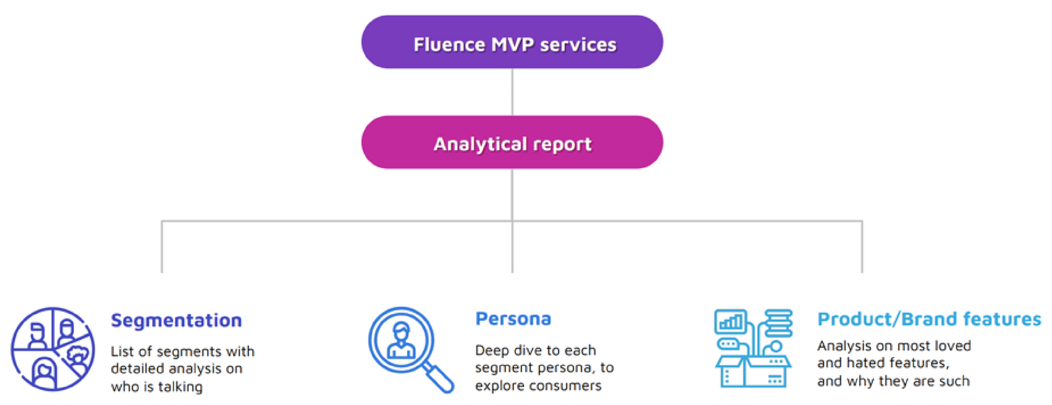

Figure 1. Natural language processing functions implemented in the Fluence project Figure

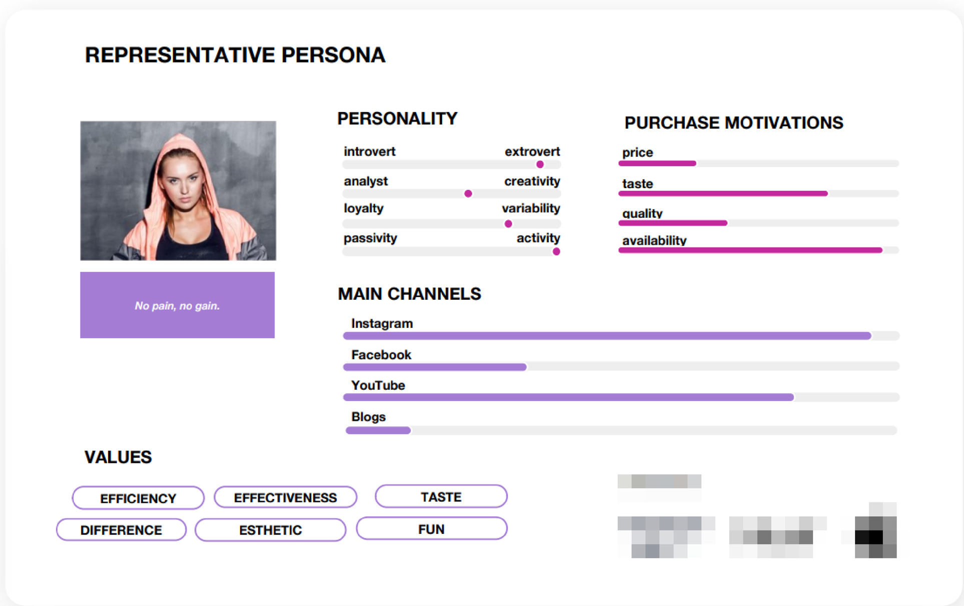

Figure 2. Illustrative example of information obtained for a defined customer segment

Automatic Image generation

Text-to-image models gives unprecedented freedom to control creation of images through natural language. Diffusion models have emerged as a new, state-of-the-art family of deep generative models. They broke the long domination of Generative Adversary Networks (GANs) in the difficult task of image generation, and demonstrated the potential use of various applications of computer vision. Over the past year, numerous architectures have been developed (as presented in Fig. 3) [9], the most important and effective diffusion models are: Glide [3], Dalle2 [7], Imagen [8] and Stable Diffusion [4].

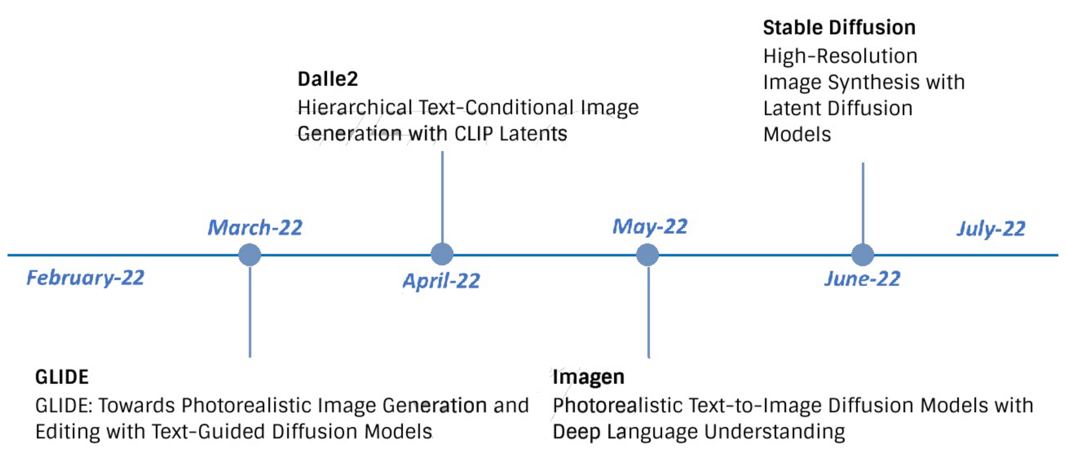

Figure 3. Timeline presenting the latest generative diffusion model architectures

Innovative framework

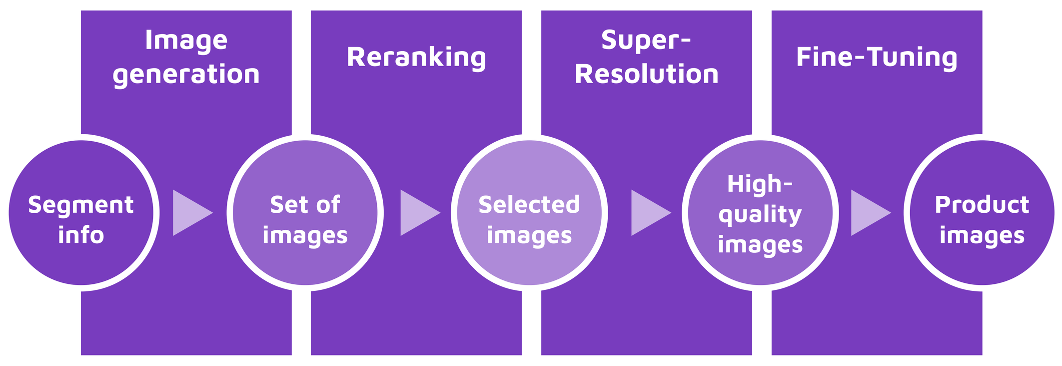

Figure 4. Structure of the proposed framework

In order to automatically generate advertisements adapted to a given customers segment, an innovative framework has been proposed, which includes many steps:

- **image generation** - with Stable Diffusion, [4]
- **image reranking** - based on CLIP model [5],
- **super-resolution** - based on SwinIR model [2],
- **fine-tuning** - in embedding space based on Textual Inversion [1] and with unfrozen model weights based on DreamBooth [6] approach [8].

The structure of the framework is shown in Fig. 4. The input text for generation is based on the description of clients segment. The illustrative examples of the generated images are presented in Figs. 5, 6 and 7.

Generation examples

Figure 5. Image generation results for customer segment: coffee enthusiast

Figure 6. Image generation results for customer segment: cooking enthusiast

Figure 7. Image generation results for customer segment: healthy vegan

Fine-tuning of generative model

Text-to-image models offer a large variety of generated results and are well suited for generating advertising content (as can be seen in Figs. 5, 6 and 7.). However, the pre-trained model will not know the appearance of a the product and its label. Therefore, it is necessary to fine-tune the model to a different product each time. For this purpose, image sets of various products have been prepared, examples of which are presented in Fig. 8.

Two state-of-the-art methods of fine-tuning were used to personalize Stable Diffusion model: Textual Inversion [1] in embedding space and DreamBooth [6] with unfrozen weights of the model. Both methods only require a few product images to learn the model properly.

Figure 8. Illustrative examples from milk and jam sets of images

Fine-tuning in embedding space

In the first approach, new "words" in the embedding space of a frozen text-to-image model were trained with Textual Inversion [1]. The training procedure was performed for several product images and for 15k iterations.

After training, these new "words" can be composed into natural language sentences to better guide the generative model. Results of generation are presented presented in Fig. 9.

Figure 9. Results from generation after fine-tuning in embedding space

Fine-tuning with unfrozen weights

The second approach is based on DreamBooth [6]. In this case, the fine-tuning process is performed for model with unfrozen weights. In addition, training is performed with a very low learning rate and additional model-generated class images (to reduce over-fitting). The product appearance is embedded in the output domain of the model and then can be used to generate images with a different scene context. Exemplary generation results are presented in Fig. 10.

Figure 10. Results from generation after fine-tuning with unfrozen weights

Conclusions

The proposed framework for automatic image generation allows for generation of high-quality, personalized advertising for a given customers segment. With proper fine-tuning, it is also possible to generate images with defined product. It was noticed, that fine-tuning with unfrozen weight performs better for most cases that fine-tuning in embedding space. The current process of improving the generation procedure includes the correction of faces and texts generated in the images. For this, an additional adaptation and fine-tuning of the model is performed.

References

[1] Rinon Gal, Yuval Alaluf, Yuval Atzmon, and et al. An image is worth one word: Personalizing text-to-image generation using textual inversion, 2022.
[2] Jingyun Liang, Jiezhang Cao, Guolei Sun, and et al. Swinir: Image restoration using swin transformer, 2021.
[3] Alex Nichol, Prafulla Dhariwal, Aditya Ramesh, and et al. Glide: Towards photorealistic image generation and editing with text-guided diffusion models, 2021.
[4] Alex Nichol, Prafulla Dhariwal, Aditya Ramesh, and et al. High-resolution image synthesis with latent diffusion models, 2021.
[5] Alec Radford, Jong Wook Kim, Chris Hallacy, and et al. Learning transferable visual models from natural language supervision, 2021.
[6] Nataniel Ruiz, Yuanzhen Li, Varun Jampani, and et al. Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation, 2022.
[7] Chitwan Saharia, William Chan, Saurabh Saxena, and et al. Photorealistic text-to-image diffusion models with deep language understanding, 2022.
[8] Chitwan Saharia, William Chan, Saurabh Saxena, and et al. Photorealistic text-to-image diffusion models with deep language understanding, 2022.
[9] Anwaar Ulhaq, Naveed Akhtar, and Ganna Pogrebna. Efficient diffusion models for vision: A survey, 2022.

This project is supported by the National Centre for Research and Development with grant number: POIR.01.01.01-00-0062/20.