

Rancang Bangun Sistem Multiplatform Text-to-Image berbasis Stable Diffusion dengan Fine-Tuning Dreambooth = Development of a Text-to-Image Multiplatform System Based on Stable Diffusion with Dreambooth Fine-Tuning

Samuel Nathaniel Halim, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920526244&lokasi=lokal>

Abstrak

Seiring perkembangan era digital, teknologi semakin mengintegrasikan diri ke dalam berbagai aspek kehidupan manusia, termasuk dalam pengembangan seni rupa. Salah satu hasilnya adalah pembentukan gambar oleh artificial intelligence (AI) dengan model Stable Diffusion yang dapat menghasilkan gambar dari kalimat teks atau prompt, disebut juga model text-to-image. Namun Stable Diffusion masih memiliki kelemahan, yaitu keterbatasan dalam menghasilkan gambar suatu subjek yang tidak dikenal karena tidak ada pada dataset pelatihan awal Stable Diffusion. Terdapat juga isu-isu lain yaitu penghasilan gambar yang kurang akurat, eror, mengandung bias, dan stereotyping sehingga memengaruhi kepuasan penggunaan Stable Diffusion. Oleh karena itu, diusulkan solusi berupa personalisasi dan kustomisasi, atau fine-tuning, model Stable Diffusion melalui model DreamBooth yang rilis pada Agustus 2022. Dalam penelitian ini, dikembangkan sebuah sistem multiplatform berupa sistem backend dan aplikasi mobile yang menyediakan fungsionalitas Stable Diffusion dan DreamBooth, agar dapat diakses oleh kalangan umum untuk membentuk gambar AI dengan kustomisasi tinggi. Untuk menguji gambar yang dihasilkan sistem, dan mengenal parameter pelatihan model apa saja yang dapat memengaruhi kualitas hasil gambar, dilakukan pengujian kuesioner responden berdasarkan parameter uji yaitu akurasi, kualitas pencahayaan, kualitas warna, jumlah eror, detail, dan nilai estetika dari 2 kelompok gambar AI hasil sistem. Kelompok pertama yaitu gambar yang dibentuk dengan 600 step pelatihan DreamBooth dan 24 step pembentukan Stable Diffusion, dan kelompok kedua dengan 1000 step DreamBooth dan 50 step Stable Diffusion. Hasil responden menunjukkan bahwa style gambar dan prompt yang digunakan juga berdampak terhadap pengaruh jumlah step pelatihan yang digunakan. Gambar yang menyerupai fotografi realistis dengan prompt sederhana, menunjukkan peningkatan nilai detail dan estetika mencapai 12,8% dan 4% seiring dengan peningkatan jumlah step pelatihan, dan penurunan 4,8% bagi gambar hasil prompt yang detail dan prompt gambar yang bersifat lukisan artistik, dan hasil netral untuk jumlah eror pada gambar dari setiap prompt.

..... As the digital era advances, technology is increasingly integrating itself into various aspects of human life, including the development of visual arts. One of the results is the generation of images by artificial intelligence (AI) using the Stable Diffusion model, which can produce images from texts or prompts, also known as a text-to-image model. However, Stable Diffusion still has its own limitations, particularly in generating images of unfamiliar subjects which are not present in Stable Diffusion's initial training dataset. Other issues include inaccuracies, errors, biases, and stereotyping in the generated images which affect user satisfaction in using Stable Diffusion. Therefore, a proposed solution is to personalize and customize a Stable Diffusion model using the DreamBooth model, released in August 2022. In this research, a multiplatform system comprising of a backend system and a mobile application was developed to provide Stable Diffusion and DreamBooth functionalities, making them accessible to the general public for generating highly customizable AI images. To evaluate the images produced by the system, and to find out

which model training parameters could affect the resulting images, a questionnaire survey was conducted based on evaluation parameters such as accuracy, lighting, color, amount of errors, level of detail, and general aesthetic value of 2 groups of AI images produced. The first group consists of images generated using 600 DreamBooth training steps and 24 Stable Diffusion inference steps, while the second group involves 1000 DreamBooth steps and 50 Stable Diffusion steps. The respondents results indicated that the image style and prompt used also have an impact on the effect of the number of training steps used. Images that resemble realistic photographs, generated using simpler prompts, showed an increase in detail and aesthetic values by 12.8% and 4% respectively, as the number of training steps increased. On the other hand, there was a 4.8% decrease in detail and aesthetic value for images generated from already detailed prompts and images with an artistic painting style. Meanwhile, the results were neutral for amount of error in the image for all prompts.