

Thesis Title	Investigating and Generating the Evaluation Method Between Human Fashion Aesthetic and Generative AI
Author	Hsi Yeh Wang
Degree	Master of Science (Digital Transformation Technology)
Advisor	Asst. Prof. Worasak Rueangsirarak, Ph. D.
Co-Advisor	Surapong Uttama, Ph. D.

ABSTRACT

AI drawing tools set off a revolutionary trend in the field of image creation. However, no clear and appropriate evaluation standard to rank AI graphics in fashion exists. In addition, most fashion industry insiders have never used AI tools before. This research aims to evaluate whether AI-generated images could satisfy fashion designer's needs by comparing automatic and human evaluations, and see if there is a need to create a new evaluation method. Therefore, in the first part, AI-generated fashion datasets with 25 images using Leonardo AI were created, and a survey was conducted to check how the experts ranked the AI images. Automatic evaluation methods, such as FID and Clip scores of each picture were measured to observe the correlation with human evaluation. The result showed that the correlation coefficient between expert and FID scores is only 0.30, while the correlation coefficient between expert and CLIP scores is 0.05. In other words, human evaluation and automatic evaluation are not so related and both have insufficiencies. Automatic evaluation is unable to provide judgments on fashion and aesthetics. The evaluations of different experts vary greatly due to the subjective consciousness and cannot provide fair and objective standards. Thus, it is necessary to create a new evaluation method that can evaluate the generated image in both fashion and AI aspects.

In the second part, based on the conclusion of the first part, it is necessary to create a new evaluation with the general public aspect. Therefore, the research addresses these issues by (1) creating a dataset of AI-generated fashion images labeled with customer rankings and (2) establishing a new evaluation method from the perspectives of the general public and the market. Two CNN models were trained: one for regression predicting continuous aesthetic scores and another for classification categorizing images into discrete rating intervals. Performance evaluation revealed that the regression model constrained by clip techniques maintained the original distribution of data, while the classification model provided a benchmark for design indicators. The study concludes that addressing data imbalance and applying augmentation techniques yielded significant results. Specifically, the regression model achieved an RMSE of 0.866 while the classification model attained an accuracy of 93%.

Keywords: Generative AI, Fashion Design, AI Evaluation Method, Convolutional Neural Network, Classification, Regression