Journal of Organizational and End User Computing
Volume 36 • Issue 1

The Intelligent Advertising Image Generation Using Generative Adversarial Networks and Vision Transformer:
A Novel Approach in Digital Marketing

Hang Zhang, School of Media and Design, Hangzhou Dianzi University, China

Wenzheng Qu, School of Informatics, University of Edinburgh, UK

Huizhen Long, SHTM, Hong Kong Polytechnic University, China & Guangdong University of Finance and Economics, China

Min Chen, School of Business, Wenzhou University, China*

ABSTRACT

With the continuous evolution of digital marketing, the generation of advertising images has become crucial in capturing user interest and enhancing advertising effectiveness. However, existing methods face limitations in meeting the diverse and creative demands of advertising content, necessitating innovative algorithms to improve advertising generation outcomes. In addressing these challenges, this study proposes a deep learning algorithm framework that cleverly integrates a generative adversarial network and an VGG-based visual transformer model to enhance the effectiveness of advertising image generation. Systematic experimentation shows that the model proposed in this article achieves an AUC metric value of more than 0.7 on several datasets. The results of the experiments demonstrate that the novel algorithm significantly improves the attractiveness of advertising content, particularly showcasing substantial benefits in website operations during online evaluation experiments.

KEYWORDS

advertising content generation, digital marketing, e-commerce, Sequence Generative Adversarial Networks (SeqGAN), Vision Transformer (ViT), Visual Geometry Group (VGG)

INTRODUCTION

In today's digital era, advertising has evolved from traditional broadcast-style promotional methods to more personalized and precise marketing strategies (Kim et al., 2022). The driving force behind this transformation is the rise of personalized advertising, representing a novel approach to deeply understanding consumer needs, behaviors, and preferences. Against this backdrop, personalized advertising research has become a prominent topic in the field of marketing, as businesses

DOI: 10.4018/JOEUC.340932 *Corresponding Author

increasingly recognize the effectiveness of attracting and retaining target audiences by meeting individualized needs. Personalized advertising research aims to construct more intelligent and personalized advertising communication systems by leveraging advanced technologies such as big data (L. Li & Zhang, 2021; Zhu, 2021), artificial intelligence (Ford et al., 2023), and deep learning (X. Liu, 2023). This research has made significant progress, providing businesses with powerful tools to gain a deeper understanding of consumers and create more personalized experiences in advertising communication.

Central to this research is the precise generation of personalized advertising content. Researchers focus on developing systems capable of generating advertising content (Lopes & Casais, 2022) based on the unique characteristics of each user, achieved through the analysis of users' historical behaviors (Noor et al., 2022), preferences, and social media activities (Jacobson & Harrison, 2022). This personalized generation not only encompasses textual information, but also includes multimedia elements such as images (Ramesh et al., 2022), videos (Singer et al., 2022), and audio (Kreuk et al., 2022), thereby enhancing the perceptibility and engagement of advertisements. With the flourishing development of social media and e-commerce platforms, consumers leave increasingly extensive digital footprints, offering richer data sources for personalized advertising. Consequently, researchers are committed to building more intelligent algorithms to accurately analyze and predict consumer behaviors, providing a more reliable foundation for the precise dissemination of personalized advertising (Chen et al., 2022; Feng & Chen, 2022).

As a successful case of intelligent advertising content generation technology, IBM Watson Advertising represents the application of IBM Watson artificial intelligence technology in the digital advertising domain, aiming to provide more intelligent and personalized advertising solutions. Leveraging cognitive computing and artificial intelligence, IBM Watson Advertising possesses the capability to comprehend, learn, and adapt to user behavior patterns. Through in-depth analysis of user behaviors and preferences, the system can generate more personalized and engaging advertising content. Additionally, IBM Watson Advertising integrates big data analytics, extracting insights from vast datasets to assist advertisers in gaining a better understanding of their target audience and adjusting advertising strategies. The platform's real-time optimization allows advertisers to obtain real-time advertising performance data during campaigns and dynamically optimize based on this information to enhance the effectiveness of advertisement placements. Finally, while ensuring advertising personalization, IBM Watson Advertising places a strong emphasis on user privacy protection, ensuring compliance with relevant privacy regulations and policies throughout the advertising delivery process.

The application of deep learning in personalized image generation for advertising marks a profound transformation in the field (X. Liu, 2023), injecting new vitality into personalized advertising communication. This technological introduction not only enhances advertising effectiveness and user interactivity, but also provides unprecedented opportunities for brands and advertisers. Firstly, the utilization of deep learning technology makes advertising image generation more intelligent and precise. Through extensive training and iterative optimization of deep learning models, the system can gradually learn and comprehend the individual differences, preferences, and purchasing behaviors of users. This personalized understanding enables targeted fulfillment of each user's needs, thereby increasing advertisement click-through rates and conversion rates. Secondly, deep learning opens up new possibilities for the creative expression of advertising images. Traditional advertising content is confined to static images and simple text, while deep learning technology can create more captivating and diversified advertisement formats, including dynamic images based on user interests and personalized video segments. This transforms advertising from a mere conveyer of information into a more narrative and artistic interactive experience.

The robust generalization capability of deep learning algorithms is another advantage in personalized image generation for advertising tasks. Even in complex and dynamic market environments, deep learning models can extract universal patterns and trends from large-scale data,

enabling advertisers to better adapt to market changes and maintain the timeliness and attractiveness of advertising content. The application of deep learning in personalized image generation for advertising not only equips brands and advertisers with more powerful and intelligent tools, but also endows users with a richer and more engaging advertising experience. This technological application signifies a closer alignment of advertising communication with individual user needs, serving as a driving force for the development of the advertising industry and heralding the advent of the personalized digital advertising era.

The following are five deep learning models commonly used for personalized advertisement content generation:

1. Recurrent Neural Networks (RNN) (Hasan et al., 2023): RNN is a type of neural network suitable for processing sequential data. Its core idea is to share weights at different time steps, enabling the network to consider contextual information within the sequence. The output at each time step serves as the input for the next. In the context of advertisement content generation, RNN can be employed to handle user historical behavior sequences, such as click records and browsing history. RNN can capture the sequential nature of user behavior, better understanding the evolution of interests and personalized needs (Shen et al., 2019).
2. Long Short-Term Memory (LSTM) (Birim et al., 2022): LSTM is an improvement over traditional RNN, introducing gated mechanisms, including forget gates, input gates, and output gates, to capture long-term dependencies more effectively. This allows LSTM to better handle long sequential data. In advertisement generation, LSTM can better capture users' long-term interests and behavioral evolution, contributing to a more accurate generation of personalized advertisement content.
3. Convolutional Neural Networks (CNN) (Yu et al., 2022): CNN is primarily used for processing image data, extracting features through convolutional operations. In the advertising context, it can be employed to analyze advertisement images or process visual information related to advertisements. In advertisement generation, CNN can be used to generate more attractive and personalized advertisement images by learning image features, enhancing the visual appeal of advertisements.
4. Generative Adversarial Networks (GAN) (Madarasingha et al., 2022): GAN consists of a generator and a discriminator, employing adversarial learning to generate realistic data. The generator aims to create samples realistic enough, while the discriminator endeavors to differentiate between real and generated data. In advertisement content generation, GAN can be used to generate more creative and unique advertisement creatives, enhancing the personalization and attention-grabbing aspects of advertisements.
5. Attention Mechanism (Pingan et al., 2022): Attention mechanism allows the model to focus more on important parts when processing sequential data. By assigning different weights to inputs, the model can more selectively attend to different parts of the input sequence. In advertisement content generation, attention mechanisms can be used to pay finer attention to key elements in the user behavior sequence, enhancing the personalization of generated advertisement content to better cater to user interests.

This study proposes a framework for personalized advertising image generation. The framework initially employs VGG (Wang et al., 2022) for feature extraction from advertising background images and utilizes a ViT (Xiao Xiong et al., 2022) model for selecting advertising background images. Subsequently, the framework incorporates the SeqGAN model (W. Li et al., 2022) to distill the product's textual information and extract the product's slogan. Finally, the framework employs a genetic algorithm (Sui Zhen et al., 2022) to optimize the layout of elements within the image based on the principle of minimizing overlap between key elements. Following these steps, the study conducts both offline and online evaluations to assess the effectiveness of the proposed framework.

The three main contributions of this study are:

1. Collaborative multi-model generation: This study innovatively employs a collaborative multi-model approach, combining VGG for background image feature extraction, the ViT model for image selection, and the SeqGAN model for generating product slogans. This comprehensive utilization of different models contributes to enhancing the diversity and quality of generated advertising images.
2. Integration of textual information for tagline generation: By introducing the SeqGAN model, this study innovatively integrates textual information from products into the generation of advertising images. The application of the SeqGAN model enables the framework to extract key elements from product text information, leading to the generation of compelling and distinctive product slogans and thereby infusing more personalized content into advertising images (Zhang et al., 2022).
3. Genetic algorithm-based layout optimization (Sohail, 2023): Another innovative aspect of this study is the incorporation of a genetic algorithm to optimize the layout of elements within advertising images. By considering the principle of minimizing overlap between key elements, this genetic algorithm-based layout optimization method ensures visual appeal in advertising images, ultimately enhancing user receptiveness to the advertisements.

In the rest of this paper, the authors introduce the recently related work in the next section, including personalized advertising, text summarization, and the Visual Transformer Model. The subsequent section presents the proposed methods: overview, VGG and ViT-based advertisement background image selection, SeqGAN-based tagline generation, and layout generation based on genetic algorithm. The following section introduces the experimental part, including practical details, comparative experiments, and an ablation study. The next section includes a conclusion. The final section includes an outlook (Ye & Zhao, 2023; Ye et al., 2023).

RELATED WORK

Personalized Advertising

Personalized advertising (Chandra et al., 2022) is a crucial strategy in the digital marketing domain, tailoring promotional content based on users' interests, behaviors, and characteristics to enhance the effectiveness of advertisements and user engagement. To date, researchers have made significant progress in the field of personalized advertising, covering multiple aspects:

1. Application of deep learning in personalized advertising: In recent years, deep learning technologies, such as neural networks, convolutional neural networks (CNN) (Ning et al., 2024), and recurrent neural networks (RNN) have been widely applied in personalized advertising. These models are capable of handling large-scale and complex user data, extracting higher-level abstract features to more accurately understand user interests and behavior patterns.
2. Temporal data analysis: With the continuous accumulation of user behavioral data, the analysis of temporal data has become increasingly important for researchers. By considering the evolving process of user behavior, researchers can better comprehend changes in user interests, thereby improving the timeliness and precision of personalized advertising.
3. Cross-platform personalization: With the increase in user activities across multiple platforms, researchers are increasingly focused on achieving cross-platform personalized advertising. This involves integrating user data from different platforms to provide a consistent and comprehensive personalized experience.

4. Integration of multimodal information: In addition to traditional text and image information, researchers are exploring methods to integrate multimodal information in personalized advertising. This includes diverse forms of information such as audio, video, and social media content, creating a more varied and captivating advertising experience.
5. Privacy protection and ethical issues: As concerns about user privacy grow, researchers have conducted extensive studies on balancing personalized demands and privacy protection in personalized advertising. Ethical considerations have become a focal point of research, encompassing transparency, fairness, and user autonomy in advertising.

Text Summarization

Text summarization is a critical task in the field of natural language processing, aiming to automatically generate concise summaries of input text while preserving key information (Widyassari et al., 2022). Currently, research in the field of text summarization primarily focuses on the following aspects:

1. Extractive summarization and abstractive summarization: Researchers extensively study two main approaches in text summarization: extractive summarization and abstractive summarization. Extractive summarization constructs summaries by selecting key sentences or phrases directly from the original text, while abstractive summarization employs natural language generation techniques to create new summary content from scratch. Each approach has its strengths and weaknesses, and researchers are working to overcome their limitations to enhance the quality and fluency of summaries.
2. Application of deep learning in text summarization: With the continuous evolution of deep learning technologies, researchers have started to employ various deep learning models, such as recurrent neural networks (RNN), long short-term memory networks (LSTM), transformers, etc., to improve the effectiveness of summarization. These models can capture long-range dependencies and abstract semantic information, resulting in more accurate and fluent generated summaries.
3. Multimodal summarization: In addition to handling textual information, researchers are exploring methods for processing multimodal data in text summarization. This includes integrating various forms of data such as images, audio, and video to generate more comprehensive and enriched summaries. The study of multimodal summarization provides effective solutions for dealing with increasingly diverse multimedia information.
4. Application of reinforcement learning in abstractive summarization: The application of reinforcement learning in abstractive summarization has gained widespread attention. By introducing reinforcement learning frameworks, researchers can optimize abstractive models to better adjust and learn based on the quality of feedback of the generated summaries.
5. Domain-specific summarization: Researchers are also attempting to develop tailored summarization methods for different domains. These methods take into consideration the language characteristics and information requirements specific to particular domains and aim to enhance the relevance and practicality of summaries.

Visual Transformer Model

The Visual Transformer Model represents a significant research direction in the field of computer vision. Originally introduced by Google, this model builds upon the success of Transformer models in natural language processing and has been adapted for image processing tasks. The core concept of the Visual Transformer Model involves treating an image as a sequence and leveraging the attention mechanism of Transformers to capture both global and local relationships within the image.

To date, researchers have extensively explored the structure and design of the Visual Transformer Model. They have investigated various combinations of hyperparameters such as the number of attention heads, layers, and channels to optimize the model's performance across different tasks.

Additionally, several enhanced variants and modules of the Transformer have been proposed to cater to different types of image data and tasks. Similarly to Transformers in natural language processing, the Visual Transformer Model benefits from pre-training techniques. Researchers conduct pre-training on large-scale image datasets, enabling the model to learn universal image features. Subsequently, fine-tuning on specific tasks enhances its performance. This approach has yielded significant achievements in various computer vision tasks, including image classification, object detection, and semantic segmentation (Yang et al., 2022).

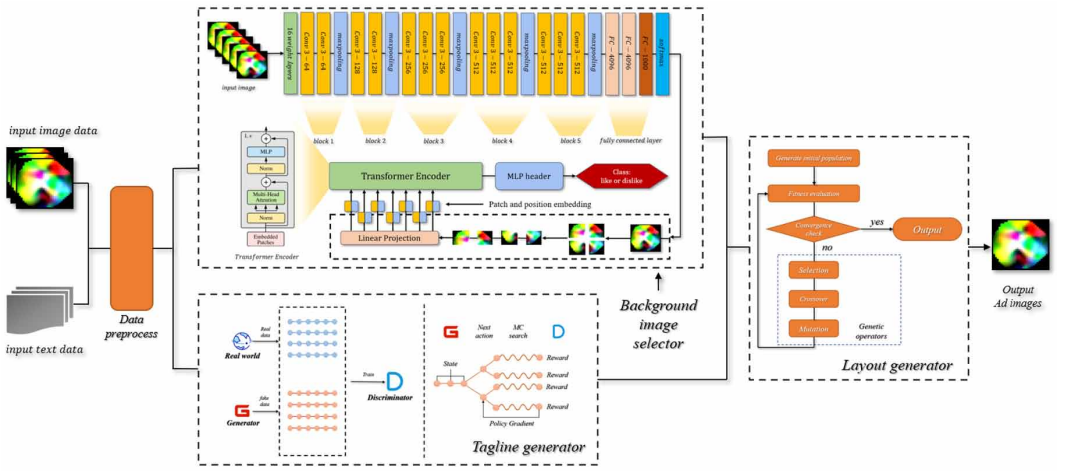
The success of the Visual Transformer Model extends beyond the realm of images and includes applications in cross-modal tasks. Researchers have applied this model to tasks involving the association between images and text, such as image captioning and visual question answering. This cross-modal application broadens the scope of the Visual Transformer Model, making it a powerful tool for handling multimodal data. Furthermore, researchers have focused on domain adaptation and transfer learning techniques to address the model's performance in specific domains or tasks. By transferring knowledge effectively between domains, and adapting the model from one domain to another, improvements in generalization and adaptability are achieved (Chen & Zhang, 2023).

METHOD

Overview

This study proposes an innovative framework for personalized advertising image generation aimed at enhancing advertising effectiveness. The framework begins by leveraging the VGG model for feature extraction from the advertising background images and then uses the ViT model to select suitable advertising background images. Subsequently, the SeqGAN model is employed to generate the product tagline by synthesizing textual information, thereby enhancing the expressiveness of the advertising images. Finally, the framework utilizes a genetic algorithm to select the optimal layout for the elements within the image based on the principle of minimizing overlap among key elements. This ensures a rational arrangement of advertising image elements. The modeling principle is shown in Figure 1.

Figure 1. The principle of this framework



VGG and ViT-Based Advertising Image Selection

The proposed framework first utilizes the VGG model for feature extraction from product images (Ning et al., 2024). In this study, the choice to use the VGG model for feature extraction is motivated by the following reasons:

1. The VGG model adopts a relatively simple yet deep convolutional neural network structure, consisting of multiple convolutional and pooling layers. This depth allows VGG to better capture low-level features such as textures, shapes, and edges in images. In the context of advertising image generation, these low-level features are crucial for generating creative and appealing advertisement content.
2. The VGG model progressively abstracts features in images through the layer-wise stacking of convolutional operations. This layer-wise processing endows VGG with sensitivity to features at different levels in the image, effectively capturing information from low to high-level features. For the task of advertising image generation, this multi-level sensitivity contributes to capturing rich information in images, enhancing the diversity and creativity of generated advertisement content.
3. The VGG model has been pre-trained on large-scale image classification tasks. Leveraging a pre-trained VGG model allows the utilization of learned general features, enabling better adaptation to the advertising image generation task. This transfer learning approach enhances the model's performance, especially in scenarios with limited data.

The VGG16 used in this study is one variant within the VGG model series, introduced by Simonyan and Zisserman (2014). It is a classical Convolutional Neural Network designed for image classification tasks. The structural overview of the VGG16 model (Ahsan et al., 2022) employed in this research is as follows:

1. Input layer: The input for VGG16 are RGB images with a resolution of 224×224 pixels. Preprocessing operations are applied in this study to resize the images to the required 224×224 dimensions.
2. Convolutional layers: VGG16 consists of 13 convolutional layers, each with a 3×3 kernel size, a stride of 1, and a padding of 1. ReLU activation functions are used to introduce non-linearity.
3. Pooling layers: Between every two consecutive convolutional layers, VGG16 includes a max-pooling layer. These pooling layers aim to reduce the spatial resolution of the feature maps while retaining essential features.
4. Fully connected layers: Following the convolutional layers, VGG16 has three fully connected layers, each containing 4,096 neurons. The output of these fully connected layers is activated using ReLU.
5. Output layer: The final layer of VGG16 is a fully connected layer with 1,000 neurons, corresponding to the 1,000 categories in the ImageNet dataset. The output is processed through the Softmax activation function for image classification purposes.

The proposed model then leverages the user's historical image browsing records, utilizing a Vision Transformer Model to select an advertising image background that is most suitable for the current user and product in the present scenario. The following provides an overview of the layers in the ViT model within the proposed framework:

1. Input representation: The feature maps of e-commerce product images outputted from the VGG model are partitioned into fixed-size blocks, each block measuring 32×32 pixels. These blocks

are linearly embedded into one-dimensional vectors, which are then concatenated into a sequence, serving as the input to the ViT.

2. Embedding layer: The input sequence undergoes an embedding layer, embedding the vectors of each block into a high-dimensional space of dimension 768.
3. Transformer encoder: This ViT model comprises 12 Transformer encoder layers, with each layer containing 12 attention heads. The dimension of each head is 64. Consequently, the total output dimension of self-attention is $12\times12\times64=9,216$.
4. MLP layer: Following each attention head, ViT utilizes a Multi-Layer Perceptron structure, including two fully connected layers. These MLP layers map the attention output to a higher dimension.
5. Classification head: The final output passes through a classification head, generating the model's ultimate classification results. In the proposed framework, the background image for the automatically generated advertisement is chosen from the classification results.

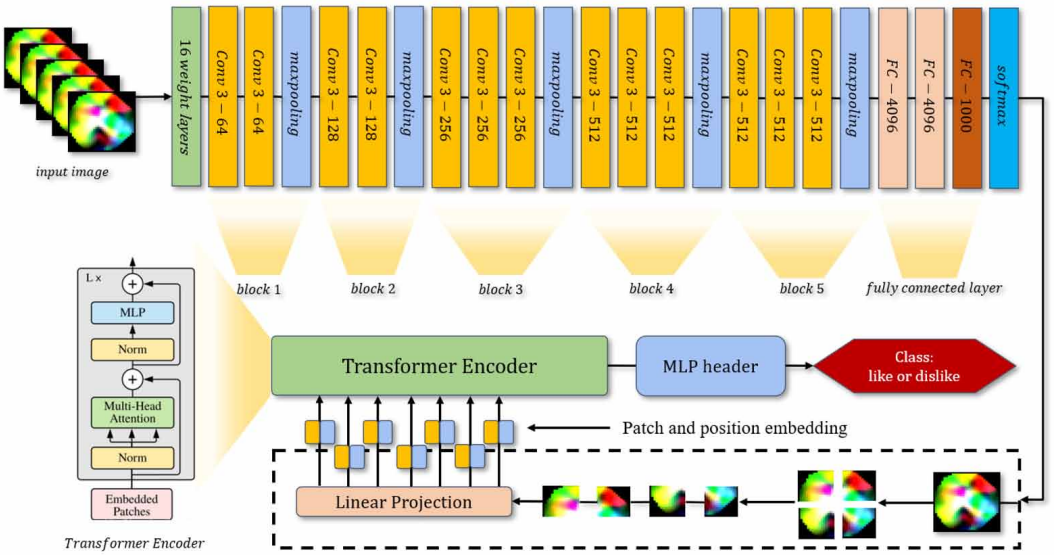
The principle of the VGG-ViT-based background image selection algorithm for advertisements is shown in Figure 2.

SeqGAN-Based Tagline Generation

Sequential Generative Adversarial Networks (SeqGAN) is a variant of generative adversarial networks designed for sequence generation tasks. It incorporates reinforcement learning to train the generator, enabling it to produce sequences that closely resemble real distributions. This is particularly useful in tasks such as text summarization. The following outlines the principles of SeqGAN when applied to tagline generation:

1. Generator (Generator): SeqGAN features a generator implemented using Recurrent Neural Networks (RNNs) or Long Short-Term Memory Networks (LSTMs). The generator progressively generates a sequence x_t based on the input noise signal z and the previously generated word x_{t-1}.

Figure 2. The principle of the VGG-ViT-Based background image selection algorithm



$$x_t \sim P\left(x_t | x_{t-1}, z; , _G\right) \tag{1}$$

2. Discriminator (Discriminator): SeqGAN includes a discriminator implemented using Convolutional Neural Networks or Recurrent Neural Networks. The discriminator assesses the authenticity of sequences.

$$D\left(x; , _D\right) \in \left[0,1\right] \tag{2}$$

3. Training SeqGAN: The generator aims to deceive the discriminator by generating realistic sequences, while the discriminator endeavors to distinguish between real and generated sequences. The two loss functions are

   Generator Loss Function:

$$J_G\left(, _G\right) = -E_{x\sim P_{\text{data}}}\left[D\left(x; , _D\right)\right] \tag{3}$$

   Discriminator Loss Function:

$$J_D\left(, _D\right) = -E_{x\sim P_{\text{data}}}\left[D\left(x; , _D\right)\right] - E_{x\sim P_{\text{pen}}}\left[1 - D\left(x; , _D\right)\right] \tag{4}$$

4. Reinforcement Learning: SeqGAN introduces reinforcement learning, where the reward signal comes from an external evaluator that assesses the generated sequences. The reward signal is used to adjust the generator's parameters to enhance the quality of the generated sequences. In this study, the external evaluator provides a reward signal based on ROUGE scores.

$$J_{RL}\left(, _G\right) = E_{x\sim P_{\text{pen}}}\left[R\left(x\right)\right] \tag{5}$$

5. Overall Optimization of SeqGAN: By jointly training the generator and discriminator, the objective is to maximize the expected reward of the tagline generator.

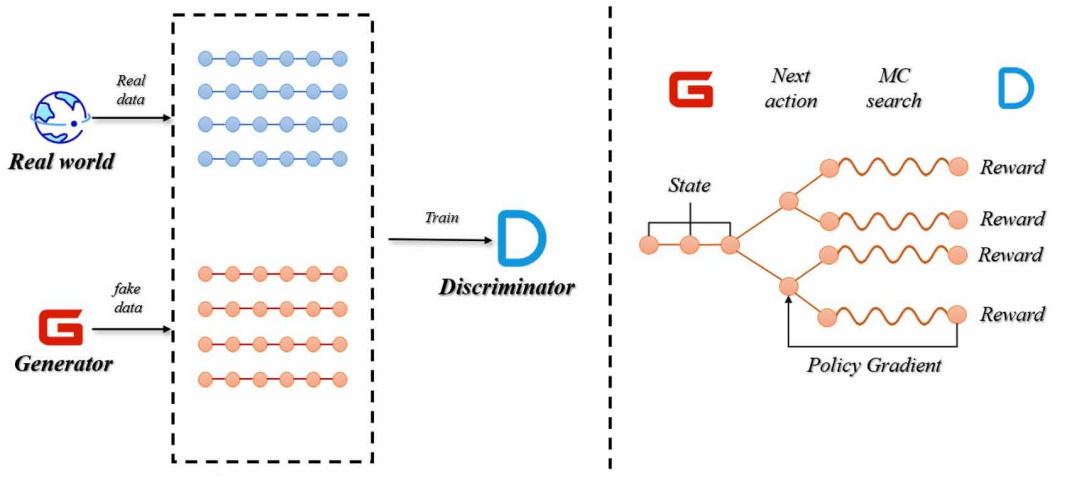
$$\max_{, G} J_G\left(, _G\right) + »J_{RL}\left(, _G\right) + »'J_D\left(, _D\right) \tag{6}$$

» and »′ are hyperparameters that balance the generator loss, reinforcement learning loss, and discriminator loss.

SeqGAN, by incorporating reinforcement learning, enables the generator to adapt better to the task of generating long sequences for tagline generation.

The principle of SeqGAN-based tagline generation is shown in Figure 3.

Figure 3. The principle of SeqGan-Based tagline generation



Layout Generation-Based on Genetic Algorithm

For the placement of product trademarks and taglines in advertising background images, this study employs a genetic algorithm for determination. The specific placement steps are as follows:

1. Step 1: Utilize object detection algorithms to identify key elements in the background image, representing the position of each key element as a bounding box.
2. Step 2: Take the positions and sizes of the bounding boxes of key elements, along with the random initial positions and sizes of the product's trademark and taglines, as input data and feed them into a genetic algorithm.
3. Step 3: The optimization objective of this genetic algorithm is to minimize the overlap between elements, as overlap among elements can impact the advertising effectiveness of the image. The minimization objective function is defined as follows:

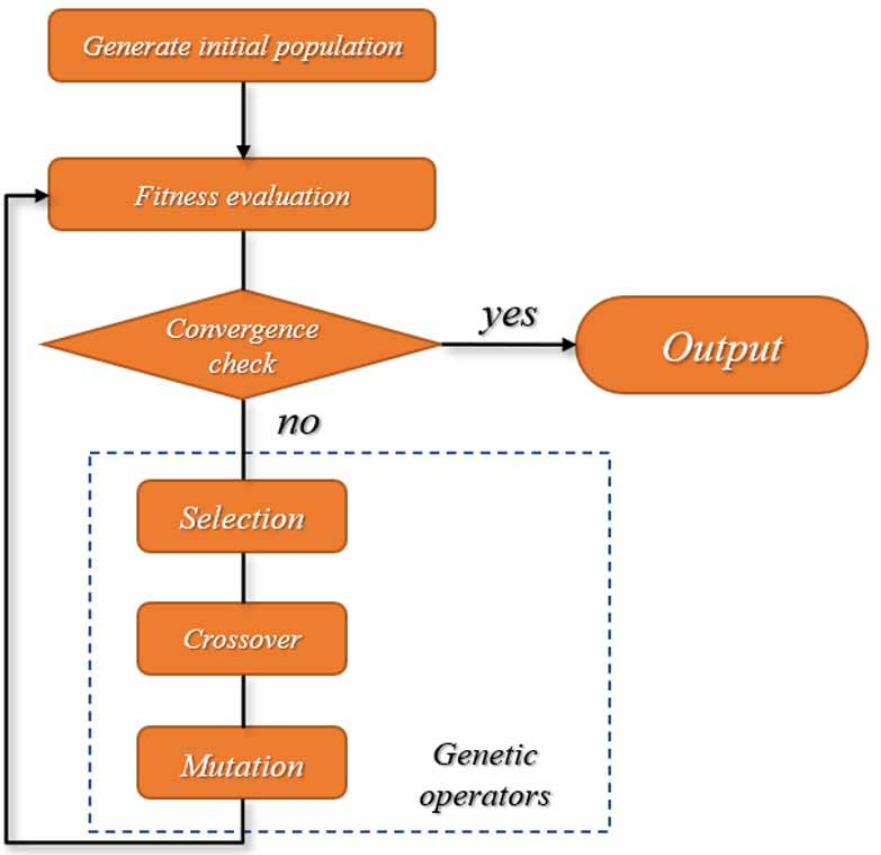
$$Overlap\% = \frac{Area_{overlap}}{Area_{total}} \tag{7}$$

where the Euclidean distance is used for the distance between elements:

$$\begin{aligned} Distance\left(i,j\right) &= \sqrt{\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2} \\ \forall\left(i,j\right) i &\neq j \\ i,j \in \left\{tagline, trademark, person, flower \ldots\right\} \end{aligned} \tag{8}$$

The principle of the layout generation algorithm is shown in Figure 4.

Figure 4. The principle of the layout generation algorithm



EXPERIMENT

Experimental Design

To evaluate the effectiveness of the proposed advertising image generation algorithm, comparative experiments were conducted against six baseline models using data from five e-commerce website datasets. The comparative experiments consisted of two parts aiming to assess the performance and robustness of the ViT algorithm in the task of advertising image background classification and the effectiveness of the SeqGAN-based tagline generation algorithm.

For the evaluation of the VGG-ViT advertising background image selection algorithm, the experiments involved selecting the browsing records of 100 website users from each e-commerce site. The images from these records were used as the experimental dataset, which was split into training and testing sets in a 7:3 ratio. The performance of VGG-ViT's advertising background image selection was evaluated by comparing it with six baseline models.

In the assessment of the SeqGAN-based tagline generation algorithm, the experiments utilized the introductory and tagline texts of 100 products from these five e-commerce websites. The model's tagline generation performance was evaluated by comparing it with six baseline models.

Finally, this study conducted online evaluations of the framework and analyzed the results obtained from the online assessments.

The following is a description of the hardware and software environment for the experiment.

1. Hardware Environment:
   a. GPU: NVIDIA Tesla V100, 4 blocks
   b. CPU: Intel Xeon Gold 6,248
   c. Number of cores: 20 cores
   d. Memory: 256GB
   e. Storage: 1TB SSD
2. Software Environment:
   a. Operating system: Ubuntu 18.04 LTS
   b. Deep learning frameworks: TensorFlow 2.3 and PyTorch 1.7
   c. GPU driver: NVIDIA CUDA Toolkit 11.1
   d. cuDNN library: CUDA Toolkit version
   e. Python version: Python 3.8

Parameter setting:

1. VGG Model:
   a. Learning Rate: 0.001, Batch Size: 64, Epochs: 50, Weight Decay: 0.0001, Convolutional Kernel Size: 3x3, Optimization Algorithm: AdamW
2. ViT Model:
   a. Learning Rate: 0.0001, Batch Size: 32, Epochs: 30, Transformer Blocks: 12, Transformer Hidden Layer Dimension: 768, MLP Hidden Layer Dimension: 2048, Dropout Rate: 0.1, Optimization Algorithm: AdamW

The metrics used in this study to evaluate the selection of background images for advertisements are accuracy, precision, recall, f1-score, and AUC:

1. Accuracy: Accuracy is the ratio of correctly predicted instances to the total instances. In the context of a binary classification task, the confusion matrix is used, where:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP+TN+FP+FN}} \tag{9}$$

2. Precision: Precision is the ratio of correctly predicted positive observations to the total predicted positives. It is calculated as:

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{10}$$

3. Recall: Also known as sensitivity or true positive rate, recall is the ratio of correctly predicted positive observations to all actual positives. It is given by:

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{11}$$

4. F1-score: The F1-score is the harmonic mean of precision and recall, balancing the trade-off between precision and recall:

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{12}$$

5. AUC: The AUC is computed based on the True Positive Rate (Recall) and False Positive Rate, visualized in the ROC curve. AUC represents the area under this curve, indicating the model's ability to distinguish between positive and negative instances.

The metrics used in this study to evaluate tagline generation are BLEU-2, BLEU-3, BLEU-4, and BLEU-5.

BLEU (Bilingual Evaluation Understudy) is:

$$\text{BLEU-n} = \text{BP} \times \exp\left(\frac{1}{n}\sum_{i=1}^{n} \log\left(p_i\right)\right) \tag{13}$$

Here, p_i is the precision for n-grams, and BP (brevity penalty) is used to penalize short generations. The BLEU-n scores represent the precision of n-grams in the generated sequence compared to the reference sequence.

Datasets and Baseline Models

The data sets used in this study comes from five e-commerce websites:

1. Taobao_behavior dataset: This dataset can be downloaded from this address:
   a. https://www.kaggle.com/datasets/edwardctt/taobao-behavior
2. Jingdong dataset: This dataset can be downloaded from this address:
   a. https://www.kaggle.com/datasets/pwang001/jjingdong-contest-dataset/data
3. Amazon consumer behavior dataset: This dataset can be downloaded from this address:
   a. https://www.kaggle.com/datasets/swathiunnikrishnan/amazon-consumer-behaviour-dataset
4. EDA: Online C2C fashion store – user behavior: This dataset can be downloaded from this address:
   a. https://www.kaggle.com/code/jmmvutu/eda-online-c2c-fashion-store-user-behaviour
5. ASOS e-commerce dataset: This dataset can be downloaded from this address:
   a. https://www.kaggle.com/datasets/trainingdatapro/asos-e-commerce-dataset-30845-products

The six baseline models used in the experiments are:

1. Model 1 (F. Liu & Guo, 2022) proposed an image personalized recommendation algorithm integrated CNN, fuzzy k-means, and Poincare map model.
2. Model 2 (Hui et al., 2022) proposed an image personalized recommendation algorithm integrated knowledge graph and collaborative filtering algorithm.
3. Model 3 (Paul et al., 2022) proposed a novel Dynamic Collaborative Filtering with aesthetic approach, which leverages aesthetic features of clothing images into a multi-objective pairwise ranking to capture consumer aesthetic taste at a specific time through adversarial training.
4. Model 4 (Xu & Sang, 2022) proposed a recommendation model based on integrated multiple personalized recommendation algorithms: random forest, gradient boosting decision tree, and eXtreme gradient boosting.
5. Model 5 (Cao, 2022) proposed a two-dimensional correlation-based personalized recommendation algorithm for e-commerce network information.

6. Model 6 (Indira et al., 2022) proposed an innovative optimization-driven deep residual network integrated CNN, Elephant Herding Feedback Artificial Optimization (EHFAO), and k-means algorithms.

Among them, Models 1 and 6 are relatively similar and are based on deep learning, Models 2 and 3 are improved collaborative filtering methods, and Models 4 and 5 are based on machine learning.

VGG-ViT Model Evaluation

Experiment One: Multi-Model Comparative Experiments on the Taobao Behavior Dataset

The results of comparative experiments of the VGG-ViT model on the Taobao behavior dataset with six baseline models are shown in Table 1.

The results of this experiment are shown in chart form in Figure 5.

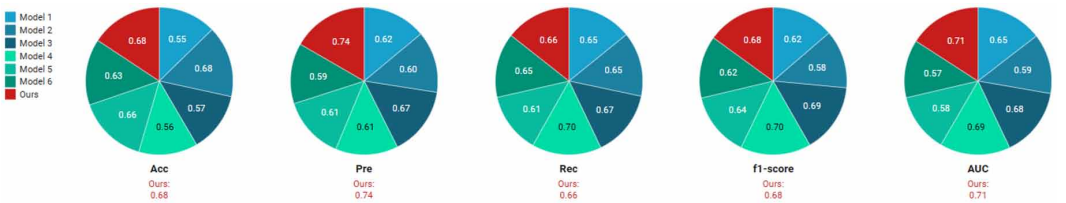
Upon analyzing the results of Experiment 1, there are four main reasons why the VGG-ViT outperforms the six baseline models:

1. The VGG-ViT model integrates the strengths of both VGG and Vision Transformer architectures, demonstrating excellent performance in capturing low-level and high-level visual features. This comprehensive feature representation is crucial for accurate image classification.
2. The VGG-ViT model leverages pre-trained weights from both VGG and Transformer components, facilitating robust feature extraction and enabling the model to recognize complex patterns in advertising background images.
3. The adaptability of the VGG-ViT model to diverse visual elements in advertising backgrounds is significantly enhanced. Its architecture allows for the identification and interpretation of subtle visual patterns in advertising images, thereby improving accuracy.
4. The VGG-ViT model's combination of local and global features ensures a holistic understanding of image content. This balanced representation enables the model to differentiate subtle variations in advertising images.

Table 1. The Results of Experiment 1

	Acc	Pre	Rec	F1-Score	AUC
Model 1	0.55	0.62	0.65	0.62	0.65
Model 2	0.68	0.60	0.65	0.58	0.59
Model 3	0.57	0.67	0.67	0.69	0.68
Model 4	0.56	0.61	0.70	0.70	0.69
Model 5	0.66	0.61	0.61	0.64	0.58
Model 6	0.63	0.59	0.65	0.62	0.57
The Authors'	0.68	0.74	0.66	0.68	0.71

Figure 5. The results of experiment one in chart form



Experiment Two: Multi-Model Comparative Experiments on the Jingdong Dataset

The results of the comparative experiments of the VGG-ViT model on the Jingdong dataset with six baseline models are shown in Table 2.

The results of this experiment are shown in chart form in Figure 6.

From the results of Experiment 2, it is evident that the VGG-ViT model exhibits superior classification performance on the JD.com dataset, primarily due to two key factors:

1. The diversity of features in the JD.com dataset likely emphasizes the significance of comprehensive feature representation, thereby magnifying the value of the VGG model.
2. JD.com advertising images may possess a more varied range of styles and elements, where the adaptability of the VGG-ViT model likely plays a greater role in identifying and interpreting these diverse visual elements, consequently enhancing accuracy.

Experiment Three: Multi-Dataset Comparative Experiments

The results of the comparative experiments of the VGG-ViT model in the rest of the three datasets are shown in Table 3.

The results of this experiment are shown in chart form in Figure 7.

Table 2. The results of experiment three

	Acc	Pre	Rec	F1-Score	AUC
Model 1	0.57	0.57	0.65	0.62	0.67
Model 2	0.67	0.63	0.62	0.57	0.63
Model 3	0.60	0.61	0.55	0.69	0.69
Model 4	0.61	0.58	0.70	0.62	0.57
Model 5	0.60	0.60	0.70	0.68	0.55
Model 6	0.63	0.60	0.55	0.59	0.64
The Authors'	0.72	0.75	0.67	0.70	0.72

Figure 6. The results of experiment two in chart form

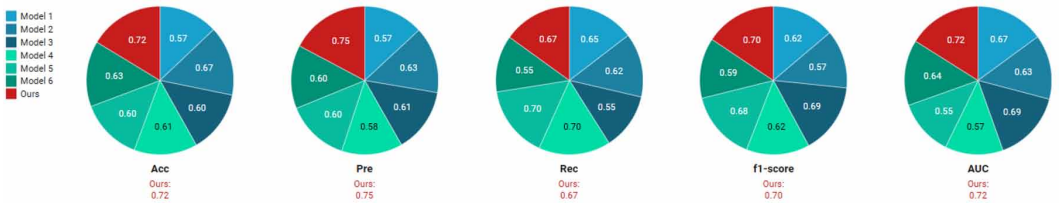


Table 3. The results of experiment three

	Acc	Pre	Rec	F1-Score	AUC
Amazon	0.73	0.66	0.67	0.67	0.66
EDA	0.71	0.67	0.77	0.71	0.69
ASOS	0.71	0.69	0.72	0.66	0.65

Figure 7. The Results of experiment three in chart form

It is noteworthy that the VGG-ViT model demonstrates outstanding performance across all three datasets. The following analysis delves into the factors contributing to its commendable performance on different platforms:

1. Generalization capability from transfer learning: The VGG-ViT model benefits from the generalization capability facilitated by transfer learning, utilizing pre-trained weights from both VGG and Transformer components. The design of transfer learning aids in its ability to recognize patterns and features unique to the advertising backgrounds of each platform.
2. Robustness from comprehensive feature representation: The VGG-ViT model's capacity for comprehensive feature representation enables it to effectively capture both low-level and high-level visual features. This versatility ensures the model's adaptability to variations in image styles and content across the Amazon, eBay, and ASOS datasets.
3. Effective adaptation to diverse styles: The VGG-ViT model effectively interprets the diverse styles and elements present in advertising images on Amazon, eBay, and ASOS. Its adaptability to different visual contexts contributes to its robust performance across the three datasets.

SeqGAN Model Evaluation

Experiment Four: Multi-Model Comparative Experiments on Taobao Dataset

The results of the comparative experiments of the SeqGAN model on the Taobao dataset with six baseline models are shown in Table 4.

The results of this experiment are shown in chart form in Figure 8.

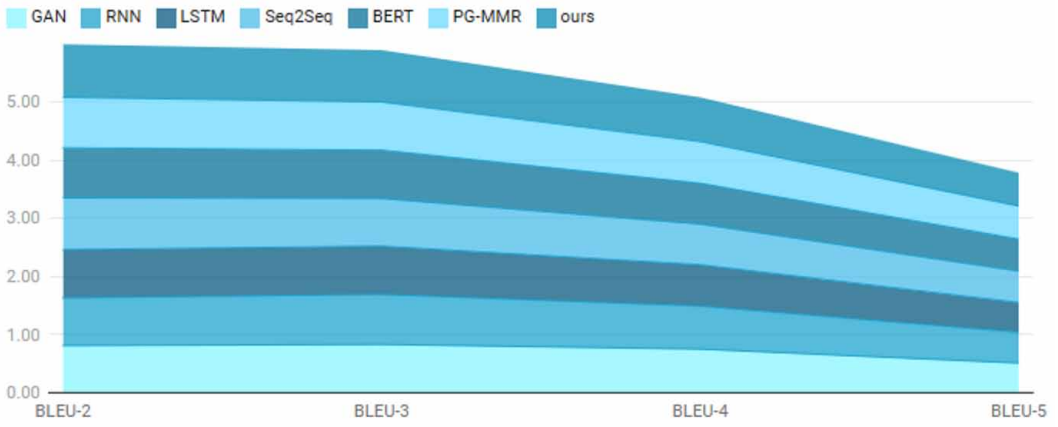
From the experimental results, the advantages of SeqGAN may manifest in the following aspects:

1. Diversity and creativity: SeqGAN, by emphasizing the diversity of the generator, facilitates the generation of more creative slogans, demonstrating greater flexibility compared to other models.

Table 4. The results of experiment four

	BLEU-2	BLEU-3	BLEU-4	BLEU-5
GAN	0.81	0.83	0.75	0.51
RNN	0.82	0.86	0.74	0.53
LSTM	0.83	0.83	0.71	0.51
Seq2Seq	0.89	0.82	0.70	0.54
BERT	0.86	0.83	0.70	0.55
PG-MMR	0.87	0.83	0.72	0.57
The Authors'	0.91	0.89	0.76	0.57

Figure 8. The results of experiment four in chart form



2. Long-term dependency: The GAN framework contributes to addressing issues related to long-term dependencies, enabling SeqGAN to better capture contextual information within the text, resulting in slogans that are more coherent and accurate.
3. Advantages of adversarial training: The adversarial training process of the Generative Adversarial Network (GAN) imparts stability to SeqGAN during training, mitigating potential instability issues that certain models may encounter.

Experiment Five: Multi-Model Comparative Experiments on Jingdong Dataset

The results of the comparative experiments of the SeqGAN model on the Jingdong dataset with six baseline models are shown in Table 5.

The results of this experiment are shown in chart form in Figure 9.

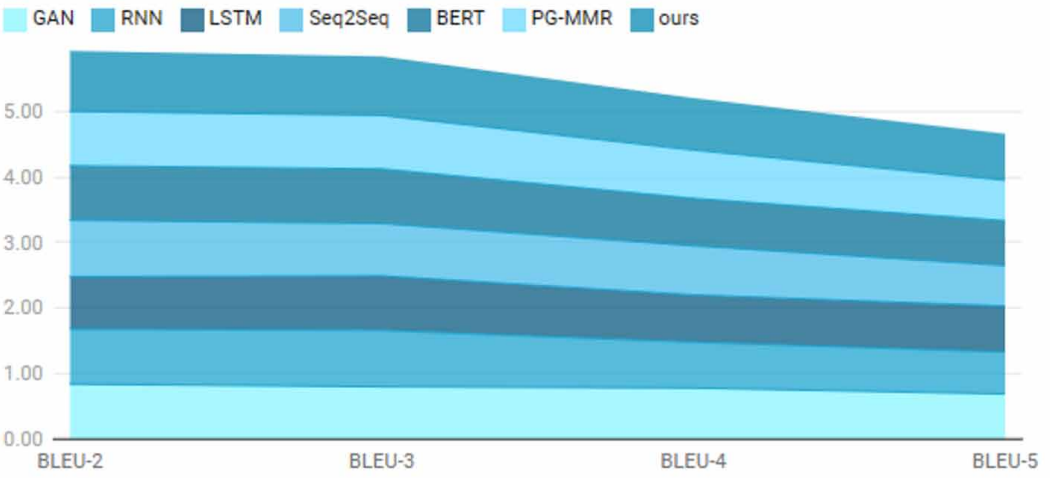
Based on the experimental results, the heightened advantages of the SeqGAN model when transitioning from the Taobao dataset to the JD.com dataset can be attributed to several factors:

1. The JD.com dataset possesses unique characteristics in terms of product descriptions and advertising language. SeqGAN, emphasizing the diversity of the generator, proves beneficial in capturing and adapting to the specific nuances and stylistic variations present in the JD.com dataset.

Table 5. The results of experiment five

	BLEU-2	BLEU-3	BLEU-4	BLEU-5
GAN	0.84	0.80	0.78	0.69
RNN	0.84	0.86	0.70	0.64
LSTM	0.80	0.83	0.72	0.70
Seq2Seq	0.86	0.80	0.75	0.62
BERT	0.83	0.83	0.72	0.68
PG-MMR	0.83	0.82	0.74	0.62
The Authors'	0.92	0.90	0.79	0.71

Figure 9. The results of experiment five in chart form



2. The GAN framework's ability to handle long-term dependencies becomes crucial when dealing with the nuanced differences in JD.com product descriptions. SeqGAN excels at capturing contextual information, enabling it to generate slogans that are not only diverse and creative but also contextually relevant to the intricacies of the JD.com platform.
3. In the JD.com dataset, the adversarial training process of SeqGAN contributes to its enhanced stability during training. This robustness becomes particularly important when addressing the various changes and challenges within the JD.com dataset, providing SeqGAN with a clear advantage over potential instability issues faced by other models.
4. SeqGAN's architecture enables it to adapt to the domain-specific language and content found in JD.com product descriptions. The model comprehends and creatively synthesizes slogans that align with the JD.com context, resulting in more attention-grabbing and contextually relevant outputs.

Ablation Study Results and Analysis

In order to analyze the functionality of each module in the VGG-ViT and SeqGAN models, two ablation experiments were conducted to present each module in the two models and compare the transformations of the model performance.

Ablation Experiment One: Ablation Study of VGG-ViT Model

The results of this ablation experiment are shown in Table 6.

The results of this ablation experiment are shown in chart form in Figure 10.

Through the removal of individual components and the comparative analysis of model performance, the authors have drawn conclusions that substantiate the significant roles played by both VGG and ViT modules in image classification tasks.

Firstly, the authors examined the model performance variation upon the removal of the VGG component. The results demonstrated a pronounced decline in performance, particularly in the capture of low-level image features. This affirms the critical role of the VGG module in image feature extraction. The VGG module, leveraging its deep convolutional structure, effectively captures low-level features such as textures, shapes, and edges in images, providing the model with rich information about image details.

Subsequently, the authors conducted an analysis of the model performance after removing the ViT component. The outcomes indicated a similar performance degradation, especially in handling global relationships and complex structures. The ViT module, facilitated by its self-attention mechanism, adeptly captures global information in images, contributing to the understanding of relationships and context between different regions in the image. ViT plays a pivotal role in comprehending the overall content of images and addressing long-range dependencies.

Ablation Experiment Two: Ablation Study of SeqGAN Model

The results of this ablation experiment are shown in Table 7.

The results of this ablation experiment are shown in chart form in Figure 11.

Table 6. Results of ablation experiment one

VGG	ViT	Acc	Pre	Rec	F1-Score	AUC
✓	✓	0.68	0.74	0.66	0.68	0.71
✓	O	0.55	0.67	0.55	0.61	0.59
O	✓	0.61	0.72	0.60	0.67	0.66

Figure 10. The results of ablation experiment 1 in chart form

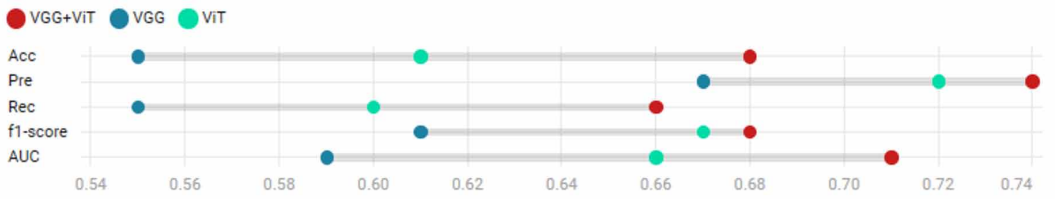
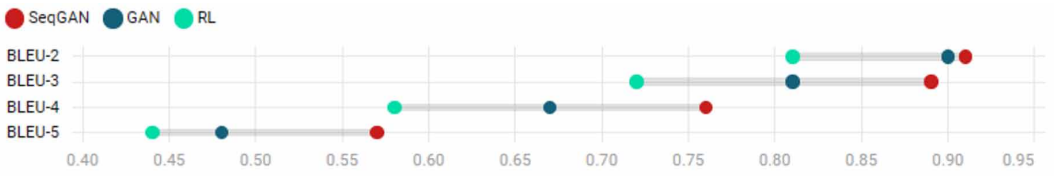


Table 7. Results of ablation experiment two

GAN	RL	BLEU-2	BLEU-3	BLEU-4	BLEU-5
✓	✓	0.91	0.89	0.76	0.57
✓	O	0.90	0.81	0.67	0.48
O	✓	0.81	0.72	0.58	0.44

Figure 11. The results of ablation experiment two in chart form



Through the systematic removal of individual modules and the comparative analysis of model performance, the authors have drawn conclusions that substantiate the crucial roles played by both the GAN and reinforcement learning modules in text generation tasks.

Initially, the authors observed the model performance variation upon the removal of the GAN module. The results indicated a significant decrease in the diversity and creativity of generated text, manifesting as a shift towards more singular and less imaginative text generation. This confirms the pivotal role of the GAN module in SeqGAN. GAN, through its adversarial training mechanism, incentivizes the generator to produce more diverse and creative text, aligning it more closely with the distribution of real-world text.

Subsequently, the authors conducted an analysis of the model performance after removing the reinforcement learning module. The outcomes suggested a similar decline in performance, particularly in the coherence and accuracy of generated text. The reinforcement learning module, through its reward mechanism, guides the generator to produce text sequences closer to the target, thereby enhancing the quality of generated text to align with expected grammatical and semantic rules.

Online Evaluation

In order to evaluate the effectiveness of this model in real-world scenarios, an online evaluation experiment was designed to evaluate the impact of this model on the click-through rate and sales of an e-commerce website before and after its use. The results of the experiment are shown in Table 8.

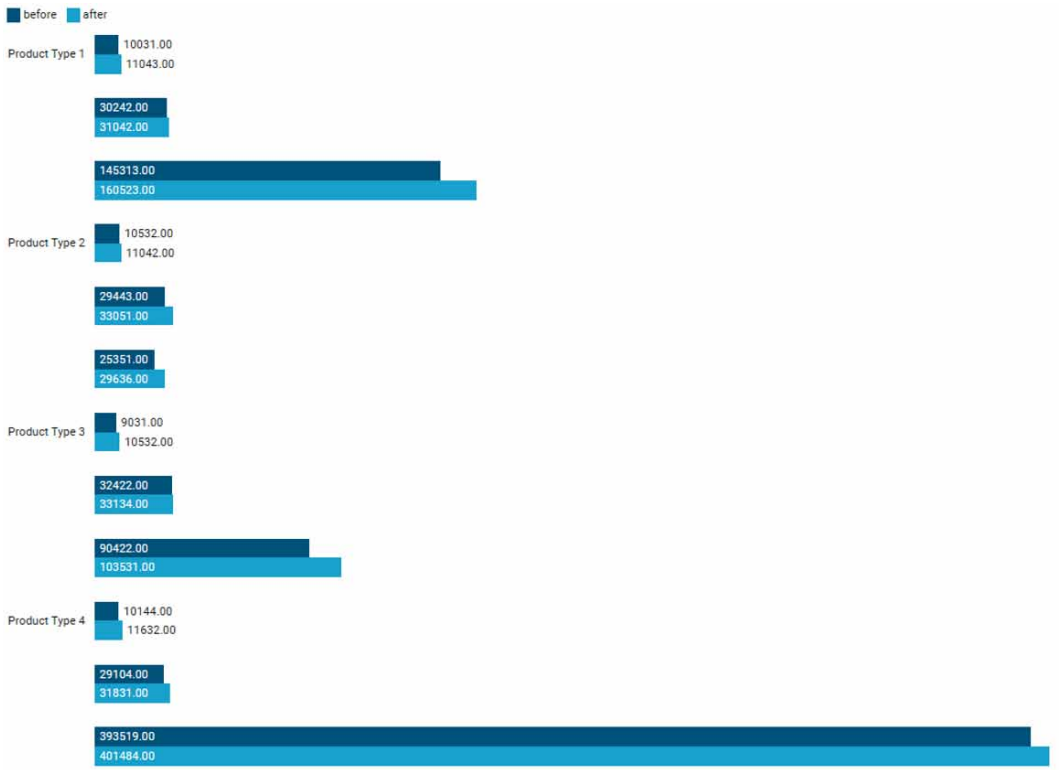
The results of online evaluation are shown in chart form in Figure 12.

The experimental results demonstrate that the adoption of the advertising image generation framework proposed in this study indeed significantly enhances the operation of the website. Prior to the replacement, the website's advertising image generation module may have had limitations, failing to fully meet user expectations and attractiveness. However, upon implementing the algorithm proposed in this study, a substantial increase in page click-through rates and product order quantities was observed. This suggests that the new framework for generating advertising images is more

Table 8. Online evaluation experiment results

	Product Type 1			**Product Type 2**		
	UV	**PV**	**GMV( )**	**UV**	**PV**	**GMV( )**
before	10,031	30,242	145,313	10,532	29,443	25,351
after	11,043	31,042	160,523	11,042	33,051	29,636
	Product Type 3			Product Type 4		
	UV	PV	GMV( )	UV	PV	GMV( )
before	9,031	32,422	90,422	10,144	29,104	393,519
after	10,532	33,134	103,531	11,632	31,831	401,484

Figure 12. The results of online evaluation in chart form



effective in capturing user interest, prompting them to click on advertisements more frequently and make purchases.

Upon deploying the authors' algorithm, a significant enhancement was observed in terms of page click-through rates and product order quantities. This observation suggested that the novel framework for generating advertising images not only piqued user interest, but also prompted users to engage more frequently with the advertisements and make actual purchases. This improvement can be interpreted as a substantial contribution of the advertising image generation module to the website's operational success.

The positive user responses to the newly generated advertising images led to heightened user engagement, subsequently increasing the page click-through rates. More importantly, the increase in product order quantities further substantiated the effectiveness of the new algorithm in stimulating user purchasing interest and facilitating tangible transactions.

Through online evaluation experiments, the authors explicitly demonstrated the positive impact of the advertising image generation algorithm proposed in this study on website operations, providing substantial evidence for the enhancement of advertising content quality and increased user interaction. This holds significant implications for businesses aiming to optimize advertising strategies, improve conversion rates, and enhance user experiences.

CONCLUSION

This study aims to explore and propose an innovative advertising image generation algorithm to enhance the attractiveness and user interactivity of advertisements. The study introduces a deep

learning-based advertising image generation framework that integrates generative adversarial networks and the Visual Transformer Model with an attention mechanism. The framework is designed to maintain image diversity and creativity during the generation process while capturing both global and local features to better meet the diverse needs of advertising content.

The innovation of this study is manifested in multiple aspects. Firstly, the authors introduce a previously unparalleled deep learning algorithmic framework that ingeniously integrates Generative Adversarial Networks (GAN) and Visual Transformer Models. GAN, with its mechanism of adversarial training, achieves diversity and creativity in generating images. Simultaneously, the Visual Transformer Model enables the model to focus on specific regions within input data, facilitating the more effective capture of both global and local features. This distinctive combination empowers the authors' algorithm to address the requirements of diversity and creativity in advertising image generation. Through GAN, it achieves vivid and creative expressions of advertising content, while the attention mechanism of the Visual Transformer precisely captures both global and local features. In summary, the innovation of this study lies not only in the introduction of advanced deep learning technologies but also in the unique integration of these technologies. This integration provides a novel and efficient approach to advertising image generation, allowing the authors' algorithm to effectively balance the demands for diversity and creativity.

To validate the effectiveness of the proposed algorithm, a series of experiments were conducted, including two sets of comparative experiments, two ablation experiments, and an online evaluation experiment. In the ablation experiments, the significance of each module within SeqGAN and VGG-ViT was verified by selectively removing different components and comparing model performance. In the online evaluation, the advertising image generation module of a running website was replaced, leading to a significant increase in page click-through rates and product order quantities. This not only confirms the effectiveness of the new algorithm in improving user engagement and promoting actual transactions, but also provides strong support for businesses seeking to optimize advertising strategies and enhance user experiences.

Overall, the advertising image generation algorithm proposed in this study represents a significant improvement over traditional methods. By incorporating deep learning techniques, the authors have successfully elevated the quality of advertising content, increased user interactivity, and provided an effective means for businesses to achieve more significant results in digital marketing. In the future, the authors will further deepen their research, exploring more advanced technologies and methods to adapt to the continuous development and innovation in the field of digital marketing.

OUTLOOK

Despite significant advancements in the advertising image generation algorithm proposed in this study, there remains a notable limitation, namely the insufficient validation of the algorithm's generalization performance for generating advertising content in specific domains. While the current algorithm performs exceptionally well on specific datasets, it may exhibit decreased performance when faced with advertisements from different domains or themes. To further enhance the algorithm's generalization capabilities, the next step in improvement involves introducing more diverse training data that encompass advertising images from various domains and themes. Additionally, considering the specific requirements of certain domains, targeted model fine-tuning strategies can be explored to better adapt to the specific domain needs in advertising content generation.

Another significant limitation in this study is the insufficient integration of multimodal information during the algorithm's generation process. The current algorithm primarily focuses on image generation, overlooking the crucial role of textual information in the advertising domain. To comprehensively represent advertising content, future improvements could include the effective handling and integration of textual information. Designing a joint generative model capable of simultaneously processing both image and text information is a potential avenue for improvement.

Such enhancement not only amplifies the expressive power of advertising content, but also holds promise for improving users' overall understanding and acceptance of advertisements.

Another significant drawback of the model lies in the insufficient integration of multimodal information during the algorithm's generation process. The current algorithm primarily focuses on image generation, while in the field of advertising, textual information holds equal importance. To present advertising content more comprehensively, the next improvement could involve introducing effective processing and integration of textual information. Considering the design of a joint generation model that simultaneously handles both image and textual information may better reflect the diverse visual and language elements in advertisements. This enhancement not only enhances the expressive power of advertising content, but is also expected to increase user understanding and acceptance of advertisements.

Furthermore, this study faces certain limitations in training data, potentially resulting in the model's inadequate adaptability to specific styles or elements. A key aspect of the next improvement is expanding the training dataset to cover a more extensive range of advertising image samples, enhancing the model's learning of various styles and elements. Introducing more diverse data will contribute to improving the model's generalization capabilities, enabling it to perform better in a broader range of advertising content generation tasks.

Lastly, the current algorithm may encounter efficiency issues during the advertising image generation process, especially when dealing with large-scale data. The next improvement can focus on optimizing the computational performance of the algorithm by adopting more efficient model architectures or parallel computing strategies. Additionally, considering optimizations at the hardware level, such as utilizing GPUs or distributed computing resources, could accelerate the training and inference processes of the algorithm. This enhancement will increase the algorithm's practicality, making it more suitable for real-world applications.

FUNDING STATEMENT

This research was supported by the Humanity and Social Science Youth Foundation of the Ministry of Education of China [grant number 21YJC760102] and the Zhejiang Key Research Base of Philosophy and Social Sciences "Academy of Art and Chinese Studies, China Academy of Art" [grant number 23CAA13]. This research was also supported by the project of "Study on the security risk warning mechanism of the cross-border flow of sensitive data of enterprises in Zhejiang in the era of digital economy" (2024 Philosophy and Social Science Project of Zhejiang Province). This research was also supported by the project of "One Core, Two Rings and Four Modules, Graduate Personnel Training Mode in Economics and Management From the Perspective of Digital Economy," the 14th Five-Year Plan Graduate Teaching Reform Project of Zhejiang Province in 2022 [grant number Zhejiang Position Office (2023) No. 1, Serial Number 335].

REFERENCES

Ahsan, M. M., Uddin, M. R., Farjana, M., Sakib, A. N., Momin, K. A., & Luna, S. A. (2022). Image data collection and implementation of deep learning-based model in detecting Monkeypox disease using modified VGG16. *arXiv preprint arXiv:2206.01862*.

Birim, S., Kazancoglu, I., Mangla, S. K., Kahraman, A., & Kazancoglu, Y. (2022). The derived demand for advertising expenses and implications on sustainability: A comparative study using deep learning and traditional machine learning methods. *Annals of Operations Research*, 1–31. doi:10.1007/s10479-021-04429-x PMID:35017781

Cao, E. (2022). A personalised recommendation algorithm for e-commerce network information based on two-dimensional correlation. *International Journal of Autonomous and Adaptive Communications Systems*, *15*(4), 345–360. doi:10.1504/IJAACS.2022.127411

Chandra, S., Verma, S., Lim, W. M., Kumar, S., & Donthu, N. (2022). Personalization in personalized marketing: Trends and ways forward. *Psychology and Marketing*, *39*(8), 1529–1562. doi:10.1002/mar.21670

Chen, M., Liu, Q., Huang, S., & Dang, C. (2022). Environmental cost control system of manufacturing enterprises using artificial intelligence based on value chain of circular economy. *Enterprise Information Systems*, *16*(8-9), 1268–1287. doi:10.1080/17517575.2020.1856422

Chen, M., & Zhang, L. (2023). The econometric analysis of voluntary environmental regulations and total factor productivity in agribusiness under digitization. *PLoS One*, *18*(9), 0291637. doi:10.1371/journal.pone.0291637 PMID:37708182

Feng, Z., & Chen, M. (2022). Platformance-based cross-border import retail e-commerce service quality evaluation using an artificial neural network analysis. *Journal of Global Information Management*, *30*(11), 1–17. doi:10.4018/JGIM.306271

Ford, J., Jain, V., Wadhwani, K., & Gupta, D. G. (2023). AI advertising: An overview and guidelines. *Journal of Business Research*, *166*, 114124. doi:10.1016/j.jbusres.2023.114124

Hasan, A. H., Anbar, M., & Alamiedy, T. A. (2023). Deep learning approach for detecting router advertisement flooding-based DDoS attacks. *Journal of Ambient Intelligence and Humanized Computing*, *14*(6), 7281–7295. doi:10.1007/s12652-022-04437-0

Hui, B., Zhang, L., Zhou, X., Wen, X., & Nian, Y. (2022). Personalized recommendation system based on knowledge embedding and historical behavior. *Applied Intelligence*, *52*(1), 1–13. doi:10.1007/s10489-021-02363-w

Indira, D., Markapudi, B. R., Chaduvula, K., & Chaduvula, R. J. (2022). Visual and buying sequence features-based product image recommendation using optimization based deep residual network. *Gene Expression Patterns*, *45*, 119261. doi:10.1016/j.gep.2022.119261 PMID:35817289

Jacobson, J., & Harrison, B. (2022). Sustainable fashion social media influencers and content creation calibration. *International Journal of Advertising*, *41*(1), 150–177. doi:10.1080/02650487.2021.2000125

Kim, J. J., Kim, T., Wojdynski, B. W., & Jun, H. (2022). Getting a little too personal? Positive and negative effects of personalized advertising on online multitaskers. *Telematics and Informatics*, *71*, 101831. doi:10.1016/j.tele.2022.101831

Kreuk, F., Synnaeve, G., Polyak, A., Singer, U., Défossez, A., Copet, J., Parikh, D., Taigman, Y., & Adi, Y. (2022). Audiogen: Textually guided audio generation. *arXiv preprint arXiv:2209.15352*.

Li, L., & Zhang, J. (2021). Research and analysis of an enterprise e-commerce marketing system under the big data environment. [JOEUC]. *Journal of Organizational and End User Computing*, *33*(6), 1–19. doi:10.4018/JOEUC.20211101.oa15

Li, W., Wang, X.-B., & Xu, Y. (2022). Recognition of CRISPR off-target cleavage sites with SeqGAN. *Current Bioinformatics*, *17*(1), 101–107. doi:10.2174/1574893616666210727162650

Liu, F., & Guo, W. (2022). Personalized recommendation algorithm for interactive medical image using deep learning. *Mathematical Problems in Engineering*, *2022*, 2022. doi:10.1155/2022/2876481

Liu, X. (2023). Deep learning in marketing: A review and research agenda. *Artificial Intelligence in Marketing*, *20*, 239–271. doi:10.1108/S1548-643520230000020014

Lopes, A. R., & Casais, B. (2022). Digital content marketing: Conceptual review and recommendations for practitioners. *Academy of Strategic Management Journal*, *21*(2), 1–17.

Madarasingha, C., Muramudalige, S. R., Jourjon, G., Jayasumana, A., & Thilakarathna, K. (2022). VideoTrain++: GAN-based adaptive framework for synthetic video traffic generation. *Computer Networks*, *206*, 108785. doi:10.1016/j.comnet.2022.108785

Ning, E., Wang, Y., Wang, C., Zhang, H., & Ning, X. (2024). Enhancement, integration, expansion: Activating representation of detailed features for occluded person re-identification. *Neural Networks*, *169*, 532–541. doi:10.1016/j.neunet.2023.11.003 PMID:37948971

Noor, U., Mansoor, M., & Shamim, A. (2022). Customers create customers!–Assessing the role of perceived personalization, online advertising engagement and online users' modes in generating positive e-WOM. *Asia-Pacific Journal of Business Administration*.

Paul, A., Wu, Z., Liu, K., & Gong, S. (2022). Robust multi-objective visual bayesian personalized ranking for multimedia recommendation. *Applied Intelligence*, *52*(4), 1–12. doi:10.1007/s10489-021-02355-w

Pingan, Q., Yuan, L., & Ruixue, S. (2022). Image caption description generation method based on reflective attention mechanism. *The International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery*. IEEE.

Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Chen, M. (2022). Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125*, *1*(2), 3.

Shen, C.-W., Min, C., & Wang, C.-C. (2019). Analyzing the trend of O2O commerce by bilingual text mining on social media. *Computers in Human Behavior*, *101*, 474–483. doi:10.1016/j.chb.2018.09.031

Singer, U., Polyak, A., Hayes, T., Yin, X., An, J., Zhang, S., Hu, Q., Yang, H., Ashual, O., & Gafni, O. (2022). Make-a-video: Text-to-video generation without text-video data. *arXiv preprint arXiv:2209.14792*.

Sohail, A. (2023). Genetic algorithms in the fields of artificial intelligence and data sciences. *Annals of Data Science*, *10*(4), 1007–1018. doi:10.1007/s40745-021-00354-9

Wang, S.-H., Khan, M. A., & Zhang, Y.-D. (2022). VISPNN: VGG-inspired stochastic pooling neural network. *Computers, Materials & Continua*, *70*(2), 3081–3097. doi:10.32604/cmc.2022.019447 PMID:35615529

Widyassari, A. P., Rustad, S., Shidik, G. F., Noersasongko, E., Syukur, A., Affandy, A., & Setiadi, D. R. I. M. (2022). Review of automatic text summarization techniques & methods. *Journal of King Saud University. Computer and Information Sciences*, *34*(4), 1029–1046. doi:10.1016/j.jksuci.2020.05.006

Xiong, Xiao, X. W., Wang Hongtao, Su Pan, & Gao Sihua. (2022). Chinese caption of fine-grained images based on Transformer. *Journal of Jilin University Science Edition*, *60*(5), 1103–1112.

Xu, L., & Sang, X. (2022). E-commerce online shopping platform recommendation model based on integrated personalized recommendation. *Scientific Programming*, *2022*, 2022. doi:10.1155/2022/4823828

Yang, Y., Jiao, L., Liu, X., Liu, F., Yang, S., Feng, Z., & Tang, X. (2022). Transformers meet visual learning understanding: A comprehensive review. *arXiv preprint arXiv:2203.12944*.

Ye, S., Yao, K., & Xue, J. (2023). Leveraging empowering leadership to improve employees' improvisational behavior: The role of promotion focus and willingness to take risks. *Psychological Reports*, 00332941231172707. doi:10.1177/00332941231172707 PMID:37092876

Ye, S., & Zhao, T. (2023). Team knowledge management: How leaders' expertise recognition influences expertise utilization. *Management Decision*, *61*(1), 77–96. doi:10.1108/MD-09-2021-1166

Yu, T., Jin, Z., Liu, J., Yang, Y., Fei, H., & Li, P. (2022). Boost ctr prediction for new advertisements via modeling visual content. *2022 IEEE International Conference on Big Data (Big Data)*. IEEE. doi:10.1109/BigData55660.2022.10020786

Zhang, H., Fan, L., Chen, M., & Qiu, C. (2022). The impact of SIPOC on process reengineering and sustainability of enterprise procurement management in e-commerce environments using deep learning. *Journal of Organizational and End User Computing*, *34*(8), 1–17. doi:10.4018/JOEUC.306270

Zhen, Sui, Z. T., Wu Tao, & Chen Huarui. (2022). Storage optimization of three-dimensional warehouse based on multi population space mapping genetic algorithm. *Jilin Daxue Xuebao (Lixue Ban), Journal of Jilin University*, *60*(1).

Zhu, D. (2021). Research and analysis of a real estate virtual e-commerce model based on big data under the background of COVID-19. [JOEUC]. *Journal of Organizational and End User Computing*, *33*(6), 1–16. doi:10.4018/JOEUC.20211101.oa28