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

# Integrating Visual Transformer and Graph Neural Network for Visual Analysis in Digital Marketing:

## Exploring and Predicting Advertising Effectiveness

Yingna Chao, Hoseo University, South Korea

Hongfeng Zhu, Hunan Institute of Science and Technology, China*

Yueding Zhou, Konkuk University, South Korea

## ABSTRACT

In today's digital economy, digital marketing has become a crucial means for businesses to drive growth and enhance brand exposure. However, with increasing competition, predicting and optimizing advertising effectiveness has become a pivotal component in formulating digital marketing strategies. In order to better understand ad creatives and deeply explore the information within them, this study focuses on integrating visual transformer (VIT) and graph neural network (GNN) methods. Additionally, the study leverages generative adversarial networks (GAN) to enhance the quality of visual features, aiming to achieve visual analysis, exploration, and prediction of advertising effectiveness in digital marketing. This approach begins by employing VIT, an emerging visual transformer technology, to transform image information into high-dimensional feature representations.

## KEYWORDS

Digital Marketing Analysis, Advertising Effectiveness Prediction, Visual Transformer, Graph Neural Network, Generative Adversarial Network

## INTRODUCTION

With the vigorous development of the internet and mobile technology, digital advertising is experiencing explosive growth in both scale and complexity (Desai, 2019). No longer confined to simple information delivery, modern digital advertising pursues higher levels of personalization, creativity, and intelligence to better cater to increasingly diverse and discerning user demands. In this wave of digitization, the prediction and analysis of advertising effectiveness have become particularly crucial.

DOI: 10.4018/JOEUC.342092

*Corresponding Author

Digital marketing, as a means of brand promotion, market expansion, and customer relationship management through digital technology and media, has become an indispensable component of corporate competitiveness. It boasts advantages such as low cost, high efficiency, quantifiability, and interactivity (Veleva & Tsvetanova, 2020) and is widely applied across various platforms such as the internet, mobile devices, and social media. As digital technology continues to innovate, digital marketing exhibits trends toward diversification, intelligence, and personalization, necessitating more flexible and precise advertising strategies.

In the complex ecosystem of digital marketing, predicting and analyzing advertising effectiveness is a critically important task. This task involves effectively leveraging user data, ad content, and platform features to accurately predict and assess key metrics of ads, such as click-through rates (CTRs), conversion rates, and revenue (Richardson et al., 2007). This not only provides crucial foundations for businesses to formulate advertising strategies and optimize decisions, but also empowers them to better understand and adapt to the ever-changing market environment.

In the core challenges of predicting and analyzing advertising effectiveness, the difficulty lies in delving into the rich information embedded in ad creatives, encompassing various forms such as text, images, and videos. Simultaneously, there is a need to comprehend the relationships between this information and complex factors such as user behavior, platform environment, and market competition. Advancements in this research domain will offer new insights and solutions for the future evolution of digital advertising.

In past studies, scholars have made significant progress in the field of predicting and analyzing advertising effectiveness. They have primarily focused their research efforts on the following areas:

### Data Mining Applications

In the early stages of predicting and analyzing advertising effectiveness, researchers actively explored the application of data mining techniques, particularly methods such as association rules and clustering analysis (Duran & Odell, 2013). Through these approaches, early researchers attempted to uncover latent relationships and user behavior patterns in advertising effectiveness data. The use of association rules enabled them to discover implicit associations in advertising placement data, thus identifying mutual influences among advertising elements (Lipianina-Honcharenko et al., 2022). For instance, a study proposed a two-stage ensemble algorithm based on cluster quality (Yan Chen, 2023). This algorithm first partitioned the dataset into several subsets using different clustering algorithms and then integrated the results of the subsets using ensemble learning methods, achieving a more optimal classification effect. This proposal provides valuable insights for our analysis of experimental results using ensemble learning methods.

### Role of Natural Language Processing

With the deepening of research, natural language processing (NLP) techniques have played an increasingly important role in predicting and analyzing advertising effectiveness. Pioneering researchers, through methods such as lexical analysis and syntactic analysis, have made the analysis of advertising text more profound and comprehensive. Through these techniques, researchers can extract more semantic information from ad copy, gaining a deeper understanding of the interaction between users and ads. This advancement has provided new perspectives for evaluating the quality of ad creatives, enabling researchers to more accurately measure the expressive effectiveness of ad text and optimize ad content to better meet user needs.

### Machine Learning in Advertising Optimization

Research centered around machine learning has played a crucial role in predicting and analyzing advertising effectiveness. This approach emphasizes learning from historical advertising effectiveness data to predict future advertising performance. Supervised learning (Jiang et al., 2020), unsupervised learning (James et al., 2023, Chapter 12), and other methods have emerged as focal points in advertising optimization research. Through supervised learning, researchers can build models that enable them to learn key features from existing labeled data, thereby predicting advertising effectiveness. Unsupervised learning allows researchers to discover latent patterns in unlabeled data, providing new insights for advertising strategies. The application of these methods offers more intelligent and precise means for optimizing advertising placement strategies.

## The Rise of Deep Learning

With the rapid development of deep learning technology, pioneers have begun to explore its application in predicting and analyzing advertising effectiveness. Convolutional neural networks (Li et al., 2021), generative adversarial networks (GANs), and other deep learning technologies have emerged, providing researchers with new avenues to extract more complex and profound features from the visual information and image content of advertisements. Through convolutional neural networks, researchers can more accurately capture visual features in advertising creatives, achieving a higher level of understanding of ad images. The introduction of GANs opens up possibilities for generating more realistic and creative advertising images. The emergence of deep learning technologies has significantly advanced advertising effectiveness analysis in image processing and visual recognition, laying a solid foundation for future advertising innovations.

Despite the significant contributions made by previous researchers in the field of advertising effectiveness prediction and analysis, their studies have certain limitations, such as data constraints, methodological singularity, overlooking temporal changes, and insufficient privacy protection. These limitations not only impact the model's generalization performance in practical applications but also hinder the comprehensive development and application of methods for predicting and analyzing advertising effectiveness.

This study aims to address the shortcomings of previous research by adopting a comprehensive approach to elevate the level of advertising effectiveness prediction and analysis. First, we enhance the understanding of visual information by introducing a visual transformer (ViT), transforming the image information of ad creatives into high-dimensional features, providing the model with richer feature representations. Second, the application of a graph neural network (GNN) not only enables the model to better consider the relationships between various elements in ad creatives but also provides more effective modeling for complex network structures. Lastly, the use of a GAN further strengthens the model's visual understanding of ad creatives, improving the quality and diversity of features. This comprehensive approach is expected to overcome the one-sidedness observed in traditional research and enhance a holistic understanding of advertising effectiveness. Our research not only addresses personalized targeting, creative matching, budget management, and privacy protection challenges in digital advertising more effectively but also represents a substantive step forward in the field of digital advertising.

In summary, within the rapidly evolving landscape of digital advertising, the innovative approach of this study not only offers new perspectives for addressing current challenges in the advertising industry but also imbues digital advertising with a deeper level of intelligence and precision. We believe that this research, by comprehensively considering user privacy and integrating multi-level feature extraction, will contribute to more sustainable and intelligent development in the field of digital advertising, providing business decision makers with more accurate means for predicting and analyzing advertising effectiveness. We firmly believe that this study not only addresses the shortcomings of previous research but also lays a solid foundation for the future development of the digital advertising industry, offering new ideas and methods for innovation and progress in the field.

The contributions of this paper can be summarized in three main aspects:

1. This paper introduces ViT technology into the field of digital marketing, transforming advertising materials into high-dimensional feature representations. This innovation achieves a more in-depth and comprehensive analysis of visual information. Leveraging the capabilities of ViT, we

effectively capture complex visual features in advertising materials, providing new perspectives and solutions for the analysis of advertising effectiveness in digital marketing.

2. In this study, we integrate GNN with ViT, constructing a graph structure for advertising materials to uncover relational information. Through the introduction of GNN, we not only consider the features of individual advertising elements but also comprehensively analyze the associative relationships among elements in advertising materials. This feature of the model enhances the overall understanding of advertising content, providing a more accurate foundation for predicting advertising effectiveness.
3. We further introduce GAN, enhancing the model's visual understanding of advertising materials by generating more realistic and distinctive image features. The application of GAN effectively improves feature expressiveness, enabling the model to better adapt to different advertising scenarios and achieve significant improvements in predicting advertising effectiveness. The introduction of GAN not only enhances model performance but also brings new methods and perspectives to the field of digital marketing.

## RELEVANT WORK

In the current era of the digital economy, digital marketing has become an indispensable means for businesses to stimulate growth and enhance brand exposure (Chaffey & Ellis-Chadwick, 2019). However, with the increasing intensity of market competition, predicting and optimizing advertising effectiveness has become particularly crucial, serving as an essential component in the formulation of digital marketing strategies. To gain a deeper understanding of advertising materials and unearth their intrinsic information, the comprehensive and profound analysis, exploration, and prediction of advertising materials using advanced visual analysis techniques have become imperative. This paper focuses on the integration of ViT and GNN methods, along with enhancing the quality of visual features through GAN, to achieve visual analysis, exploration, and prediction of advertising effectiveness in the field of digital marketing.

In the chapters that follow, we will first review the research progress in related fields, delving into the key findings and limitations of previous work. Through an in-depth analysis of existing literature, we will highlight some shortcomings of previous research, providing a solid background and clear motivation for the research methods we introduce in the subsequent chapters.

Over the past few decades, stereotypical gender portrayals in advertising have been a subject of extensive research. Grau & Zotos (2018) pointed out that the representation of gender in advertising is significantly influenced by changes in family and labor market role structures. The study emphasizes the historical evolution of gender stereotypes in advertising, particularly the gradual shift toward more positive portrayals of female roles. However, despite progress, gender stereotypes persist, providing an opportunity for in-depth analysis in our research. In the era of social media, the impact of advertising content on user engagement has become increasingly important. Through detailed analysis of Facebook data, Lee et al. (2018) found a positive correlation between incorporating brand personality-related content, such as humor and emotion, and user engagement. This finding underscores the importance of advertising content engineering, offering valuable lessons for designing attention-grabbing digital advertisements. Meanwhile, social media advertising is gaining more attention, but successfully attracting customers and stimulating purchase intentions remains a challenge for organizations. Alalwan (2018), found by constructing a conceptual model that factors such as performance expectations, hedonic motivation, interactivity, informativeness, and perceived relevance significantly influence purchase intentions. This study provides a deep understanding of the key influencing factors in social media advertising, offering substantial support for developing more targeted digital advertising strategies. With millennials and Generation Z becoming major purchasing forces, Munsch (2021) investigated how to formulate effective digital advertising strategies. It was found that short digital ads, pairing with music and humor, and utilizing social media influencers had positive effects on these two age groups. This study provides practical recommendations for crafting digital advertisements for different age groups, emphasizing the importance of innovation and clever use of digital media. In addition, Gharibshah et al. (2020) delved into the crucial steps of modeling user interests and behavior in online digital advertising. By introducing a deep learning framework, especially long short-term memory networks (LSTM), the study achieved accurate predictions of user clicks and clicks on specific ad campaigns. This approach provides a cutting-edge method to improve ad effectiveness predictions while considering time series and changes in user behavior. Ning (2024) proposed a multi-view image-language fusion method based on differentiable rendering techniques, capable of understanding zero-shot 3D shapes. This proposal offers a new perspective on using different types of networks for visual feature extraction. Finally, Singh et al. (2023) showed that the continuous changes in business conditions in digital marketing pose challenges to advertising strategies. The study proposes a reinforcement learning-based model that successfully addresses some key challenges faced by A/B testing and traditional machine learning methods in digital marketing campaigns. This contribution provides a forward-looking approach to improving experimental outcomes when facing complex problems and future trends, emphasizing the importance of continuous innovation and flexibility.

While there have been numerous studies delving into the realm of digital advertising, there still exist some shortcomings and gaps, such as the need for quantifying and evaluating content in ad materials, optimizing and improving ad content, accurately predicting ad effectiveness, and optimizing and adjusting ad strategies. In this domain, one study proposed a few-shot relation extraction model based on an attention mechanism (Ji Bonan, 2023). This model utilizes attention mechanisms for learning relation representations and is trained within a meta-learning framework, effectively enhancing the model's generalization ability in situations with limited samples. That model provides a reference for predicting ad effectiveness in scenarios with limited samples.

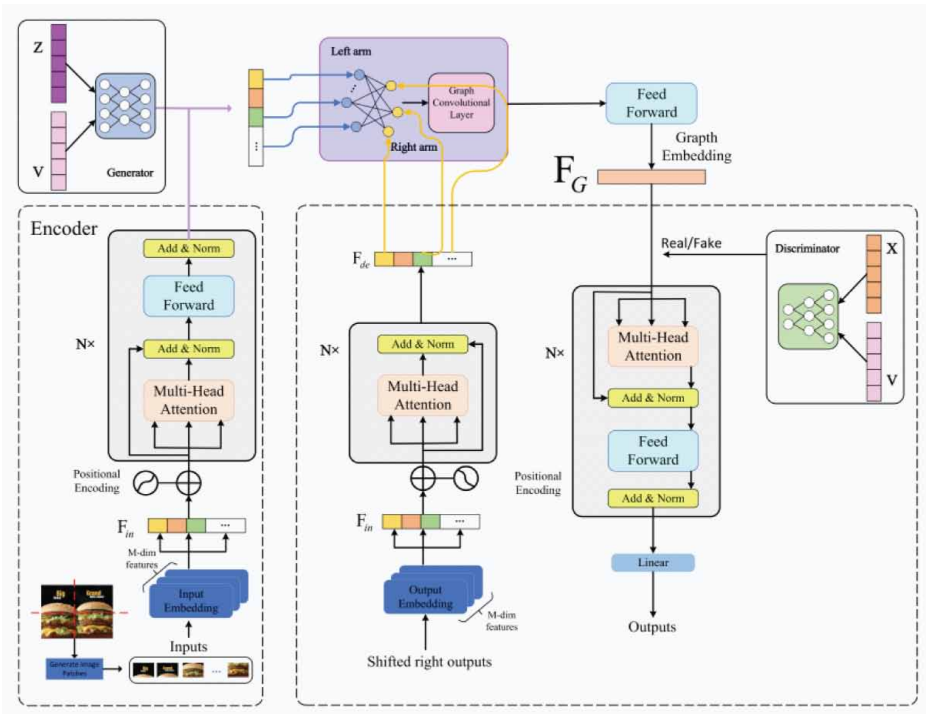
To address these research gaps, our study introduces a method that integrates ViT with GNN and enhances visual features' quality through GAN, thereby achieving visual analysis, exploration, and prediction of ad effectiveness in digital marketing. Unlike previous studies, we leverage ViT technology for high-dimensional feature representation of ad materials, integrate GNN to extract relationship information within the materials, and introduce GAN to enhance feature expression. This combination makes our model more comprehensive and accurate in understanding ad content. Through a series of experiments, our model demonstrates significant advantages in ad effectiveness prediction, offering a new visual analysis paradigm for the field of digital marketing.

In summary, this study, through summarizing previous work and analyzing limitations, successfully reveals research focal points and challenges in the domain of digital advertising. Through innovative integration of various advanced technologies, we provide digital marketing decision makers with a more powerful and accurate tool, promoting the intelligent and refined development of digital advertising strategies. In the future, we will further optimize the model and explore broader application scenarios, contributing more to the development of the digital marketing field.

## METHOD

This chapter will provide a detailed overview of the three key methods employed in our study: ViT, GNN, and GAN. The clever integration of these three advanced technologies forms our unique and efficient algorithmic framework, offering a novel visual analysis paradigm for ad effectiveness analysis in digital marketing. In the following discussion, we will systematically explain the principles and applications of these methods, presenting a clear and in-depth understanding for readers. The overall algorithmic framework is illustrated in Figure 1, depicting the collaborative interactions of these three methods in our research.

**Figure 1. Overall algorithm framework**



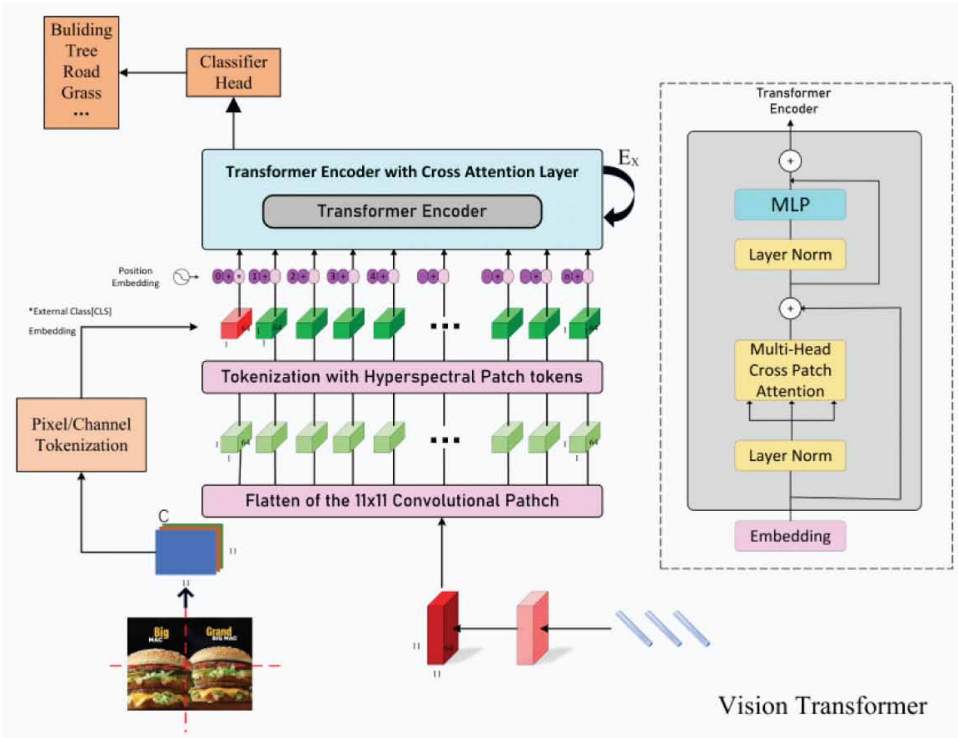

## ViT Model

The ViT model is an approach that applies transformer architecture to image classification. It can be pre-trained on large-scale datasets and then exhibits effective transfer learning in downstream tasks (Han et al., 2022). The core idea of ViT involves dividing the input image into several fixed-size patches. Each patch is projected into a fixed-length vector, and, when combined with positional encoding, forms a sequence that is input into a standard transformer encoder for feature extraction. Finally, the output corresponding to a special token is used as the representation of the image, followed by a multilayer perceptron (MLP) for classification (Almeida, 2020). In our study, ViT is introduced to convert advertising materials into high-dimensional feature representations, thus providing richer input features for the subsequent GNN and GAN. The structure of ViT is depicted in Figure 2:

The input image has dimensions $H \times W \times C$, where $H$ and $W$ represent the height and width of the image, and $C$ is the number of channels in the image. The image is divided into $N$ patches of size $P \times P \times C$, where $N = \frac{HW}{P^2}$. Each patch undergoes a linear transformation layer, resulting in a $D$-dimensional vector, known as the patch embedding. To retain positional information of patches in the image, we incorporate positional encoding, a $D$-dimensional learnable vector added to the patch embedding. Additionally, to obtain the classification result for the image, a special token, denoted as $\left[CLS\right]$, is added at the beginning of the sequence. Its initial value is also a $D$-dimensional learnable vector. Consequently, the final input sequence has dimensions $\left(N+1\right) \times D$.

The input sequence undergoes $L$ ViT blocks, each ViT block consisting of the following steps:

**Figure 2. ViT model**




- Layer Normalization (LN): Normalize each vector in the input sequence, setting its mean to 0 and variance to 1, facilitating accelerated training and improving generalization ability.
- Multi-Head Self-Attention (MHSA; Niu et al., 2021): For each vector in the input sequence, calculate its correlation with other vectors to obtain an attention matrix. Then, based on the attention matrix, perform a weighted sum on the input sequence to obtain a new sequence. To enhance the model's expressive power, we can divide the input sequence into multiple subspaces, perform self-attention on each subspace, and concatenate the results. Specifically, for an input sequence $X \in R^{(N+1)\times D}$, we first map it into query (Q), key (K), and value (V) vectors, denoted as $Q = XW_Q, K = XW_K, V = XW_V$, where $W_Q, W_K, W_V \in R^{D\times d_k}$ are learnable weight matrices, and $d_k$ is the dimension of each head, satisfying $d_k \times h = D$ where $h$ is the number of heads. Next, we split Q, K, and V into $h$ sub-vectors, denoted as $Q_i, K_i, V_i \in R^{(N+1)\times d_k}$, where $i$, j ranges from 1 to $h$. For each sub-vector, we calculate the output of self-attention as follows:

$$\text{Attention}\left(Q_i, K_i, V_i\right) = \text{softmax}\left(\frac{Q_i K_i^T}{\sqrt{d_k}}\right)V_i$$

Finally, we concatenate the outputs of the $h$ self-attention heads and multiply the result by a learnable weight matrix $W_O \in R^{D\times D}$ to obtain the output of MHSA, namely:

$$MHSA\left(Q,K,V\right)=Concat\left(Attention\left(Q_1,K_1,V_1\right),\ldots,Attention\left(Q_h,K_h,V_h\right)\right)W_O$$

- Residual Connection (Rassil et al., 2022): To avoid gradient vanishing and expedite convergence, we add a residual connection after the output of MHSA; i.e., we add the input sequence to the output sequence, resulting in a new sequence.
- LN: The new sequence undergoes LN again in preparation for the subsequent MLP.
- MLP: For each vector in the new sequence, a two-layer fully connected network is applied for non-linear transformation, namely

$$MLP\left(x\right)=GELU\left(xW_1+b_1\right)W_2+b_2$$

where $x\in R^D$ is the input vector, $W_1\in R^{D\times D_{FF}},W_2\in R^{D_{FF}\times D},b_1\in R^{D_{FF}},b_2\in R^D$ are learnable parameters, $D_{FF}$ is the dimension of the intermediate layer, typically four times $D$, namely, $D_{FF}=4D$, and $\mathrm{GELU}$ is an activation function defined as:

$$GELU\left(x\right)=x\Phi\left(x\right)=x\frac{1}{2}\left[1+erf\left(\frac{x}{\sqrt{2}}\right)\right]$$

where $\Phi\left(x\right)$ is the cumulative distribution function of the standard normal distribution, and $\mathrm{erf}\left(x\right)$ is the error function defined as:

$$erf\left(x\right)=\frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt$$

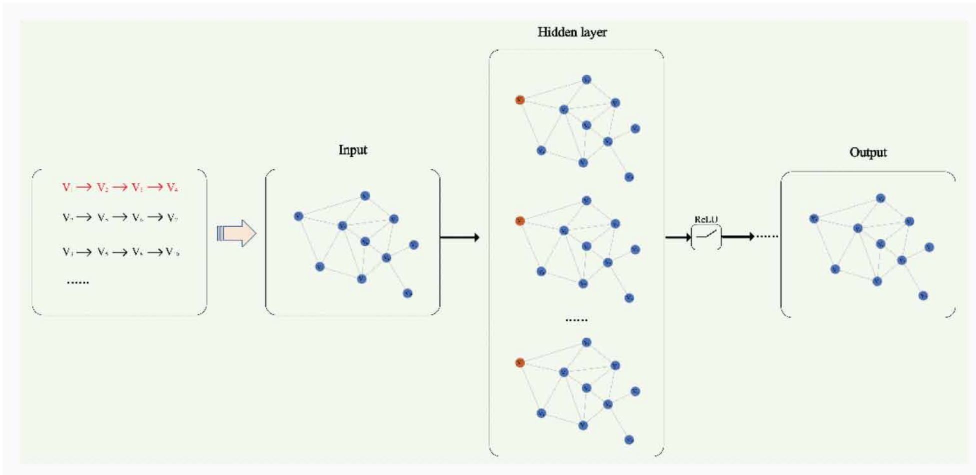
After $L$ ViT blocks, we obtain a sequence of dimension $\left(N+1\right)\times D$, where the first vector corresponds to the [CLS] token, and its output is considered as the representation of the image, denoted as $z_L^0$. To obtain the classification result for the image, we append an MLP after $z_L^0$, namely:

$$Classifier\left(z_L^0\right)=soft\max\left(GELU\left(z_L^0W_C+b_C\right)W_H+b_H\right)$$

where $W_C\in R^{D\times D_C}$, $W_H\in R^{D_C\times C}$, $b_C\in R^{D_C}$, $b_H\in R^C$ are learnable parameters, $D_C$ is the dimension of the intermediate layer, and $C$ is the number of classes. The $\mathrm{softmax}$ function is a normalization function that ensures each element of the output vector is in the range $\left(0,1\right)$ and sums to 1, representing the probability of each category.

By introducing the ViT model, we gain a more comprehensive understanding of the visual information in advertising materials, providing richer input features for subsequent GNNs and GANs. Next, we will focus on the design and application of GNNs to explore relationship information more deeply in advertising materials.

**Figure 3. GNN**




## GNN

A GNN is a type of neural network designed for learning from graph-structured data. It aims to extract and discover features and patterns within graph-structured data to fulfill various graph learning tasks, including clustering, classification, prediction, segmentation, and generation (Zhou et al., 2020). The fundamental idea behind GNN is to update node feature representations by propagating and aggregating information between nodes, thereby capturing the topological structure and relationships within the graph. In our research, GNN is applied to model the graph structure of advertising materials, aiming to unearth complex relational information within the materials and enhance the overall understanding of advertising content. The GNN is illustrated in Figure 3.

The fundamental framework of a GNN is the message passing mechanism, wherein each node updates its state by receiving and sending messages, thereby integrating its own features with those of its neighbors. The message passing mechanism can be expressed as follows:

$$h_v^{(k)} = \phi^{(k)}\left(h_v^{(k-1)}, ?_{u \in \mathcal{N}(v)}\, \psi^{(k)}\left(h_u^{(k-1)}, h_v^{(k-1)}, e_{uv}\right)\right)$$

where $h_v^{(k)}$ represents the state vector of node $v$ after the $k$ -th iteration, $\phi^{(k)}$ and $\psi^{(k)}$ are learnable functions, ? is an aggregation operation such as summation, averaging, maximum, etc., $\mathcal{N}(v)$ denotes the set of neighbors of node $v$ , and $e_{uv}$ represents the edge features between nodes $u$ and $v$ . Initially, $h_v^{(0)}$ can be the feature vector of node $v$ or a random vector. After $K$ iterations, each node's state vector contains information up to $K$ -th order neighborhoods, which can be used for subsequent tasks.

Different GNN models differ mainly in the design of the message passing mechanism, i.e., how to define $\phi^{(k)}$ and $\psi^{(k)}$ . In this study, we adopt the graph convolutional network (GCN; Zhang et al., 2019) as our GNN model, which is a simple and effective model for graph-based convolutional operations. The message passing mechanism of GCN can be expressed as follows:

$$h_v^{(k)} = \sigma\left(\sum_{u\in\mathcal{N}(v)\cup\{v\}} \frac{1}{\sqrt{d_v d_u}} W^{(k)} h_u^{(k-1)}\right)$$

where $\sigma$ is a non-linear activation function such as ReLU, $d_v$ represents the degree of node $v$, and $W^{(k)}$ is the weight matrix for the $k$-th layer. It can be observed that the message passing mechanism of GCN involves normalizing the adjacency matrix, multiplying it with the node feature matrix, and then applying a non-linear transformation to achieve information propagation on the graph.

To optimize the parameters of the GNN, we need to define a loss function that measures the difference between the model's predictions and the ground truth. In this study, our objective is to predict the CTR of advertising materials, i.e., the probability that a user clicks after seeing an ad. Therefore, we can frame this problem as a binary classification problem, in which a user either clicks or does not click. For binary classification problems, a commonly used loss function is the cross-entropy loss (Zhang & Sabuncu, 2018), which can be expressed as

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}\left(y_i \log \widehat{y_i} + \left(1-y_i\right)\log\left(1-\widehat{y_i}\right)\right)$$

where $N$ is the number of training samples, $y_i$ is the true label of the $i$-th sample, and $\widehat{y_i}$ is the predicted probability for the $i$-th sample. Cross-entropy loss measures the difference between true labels and predicted probabilities, and the loss is smaller when predicted probabilities are closer to the true labels.

To minimize the cross-entropy loss, we can use gradient descent (Haji & Abdulazeez, 2021) or its variants, such as stochastic gradient descent, Adam optimizer, etc., to update the parameters of the GNN. The basic idea of gradient descent is to update parameters along the negative gradient direction of the loss function with a certain step size, gradually reducing the value of the loss function until reaching a local optimum. The update formula for gradient descent is

$$\theta^{(t+1)} = \theta^{(t)} - \eta\nabla_\theta\mathcal{L}$$

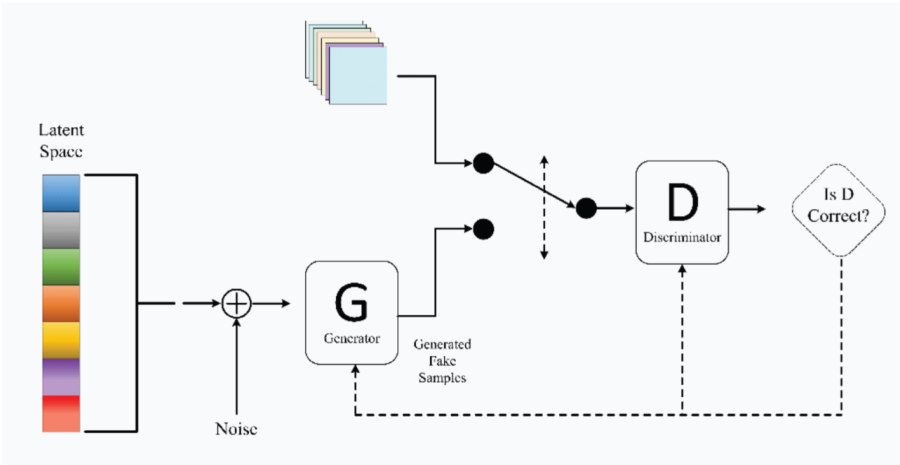
where $\theta$ represents the parameters of the GNN, $\eta$ represents the learning rate, $\nabla_\theta\mathcal{L}$ represents the gradient of the loss function with respect to the parameters, and $t$ represents the number of iterations.

By introducing the GNN, we can comprehensively consider the relationships between elements in advertising materials, enhancing the overall understanding of advertising content. Next, we will discuss the crucial role of GAN in this study. The next subsection will elaborate on the application of GANs, aiming to generate more realistic and distinctive image features, further improving the visual understanding capability of advertising materials.

## GAN

A GAN is a model that leverages a game-theoretic process between a generator and a discriminator to generate and evaluate data such as images, text, or speech (Creswell et al., 2018). The GAN consists of a generator and a discriminator. The generator's task is to generate data, such as images or text, from a random noise vector. The discriminator aims to distinguish between real and generated data, providing a probability value. The generator and discriminator engage in a competitive process,

**Figure 4. GAN**




where the generator attempts to deceive the discriminator and the discriminator strives to identify generated output. Through alternating training of these two networks, the generator gradually learns the distribution of real data, generating more realistic data, and the discriminator improves its ability to discriminate. In our study, the introduction of GAN aims to enhance the visual understanding of advertising materials by generating more realistic and distinctive image features. The structure of the GAN is illustrated in Figure 4.

The fundamental principles of GAN can be mathematically expressed as

$$\min_{G}\max_{D} V\left(D,G\right) = E_{x\sim p_{data}\left(x\right)}\left[\log D\left(x\right)\right] + E_{z\sim p_{z}\left(z\right)}\left[\log\left(1 - D\left(G\left(z\right)\right)\right)\right]$$

where $V\left(D,G\right)$ represents the value function of the two networks, $D\left(x\right)$ is the output probability of the discriminator for real data $x$, $G\left(z\right)$ is the output data of the generator for random noise vector $z$, $D\left(G\left(z\right)\right)$ is the output probability of the discriminator for generated data $G\left(z\right)$, $E$ denotes expectation, $p_{data}\left(x\right)$ is the distribution of real data, and $p_{z}\left(z\right)$ is the distribution of random noise vectors. The meaning of this formula is that the discriminator aims to maximize its ability to distinguish between real and generated data, i.e., maximize $V\left(D,G\right)$, and the generator aims to minimize the discriminator's ability to distinguish generated data, i.e., minimize $V\left(D,G\right)$. When the two networks reach Nash Equilibrium (Mazumdar et al., 2019), the generator can recover the distribution of real data, and the discriminator cannot differentiate between real and generated data.

To optimize the parameters of the GAN, we need to define a loss function to measure the competitive results of the two networks. In this study, we use the original loss function of the GAN, which is the negative value of the value function $V\left(D,G\right)$, as the loss function for both networks. Specifically, the loss function for the generator is:

$$\mathcal{L}_{\mathcal{G}} = -E_{z\sim p_{z}\left(z\right)}\left[\log D\left(G\left(z\right)\right)\right]$$

The loss function for the discriminator is

$$\mathcal{L}_{\mathcal{D}} = -E_{x \sim p_{data}(x)}\left[\log D\left(x\right)\right] - E_{z \sim p_z(z)}\left[\log\left(1 - D\left(G\left(z\right)\right)\right)\right]$$

To minimize the generator's loss function, we need to maximize the discriminator's output probability for generated data, making the generated data as close to real data as possible. To minimize the discriminator's loss function, we need to maximize the discriminator's output probability for real data and simultaneously minimize its output probability for generated data, making the discriminator distinguish between real and generated data as much as possible.

To update the parameters of the GAN, we can use gradient descent or its variants, such as stochastic gradient descent, the Adam optimizer, etc., to update the parameters of both networks. The basic idea of gradient descent is to update parameters along the negative gradient direction of the loss function with a certain step size, gradually reducing the value of the loss function until reaching a local optimum. The update formula for gradient descent is:

$$\theta^{(t+1)} = \theta^{(t)} - \eta \nabla_{\theta} \mathcal{L}$$

where $\theta$ represents the network parameters, $\eta$ is the learning rate, $\nabla_{\theta}\mathcal{L}$ is the gradient of the loss function with respect to the parameters, and $t$ represents the iteration times.

By introducing the GAN, we can enhance the model's visual understanding of advertising materials through adversarial learning, generating more realistic and distinctive image features.

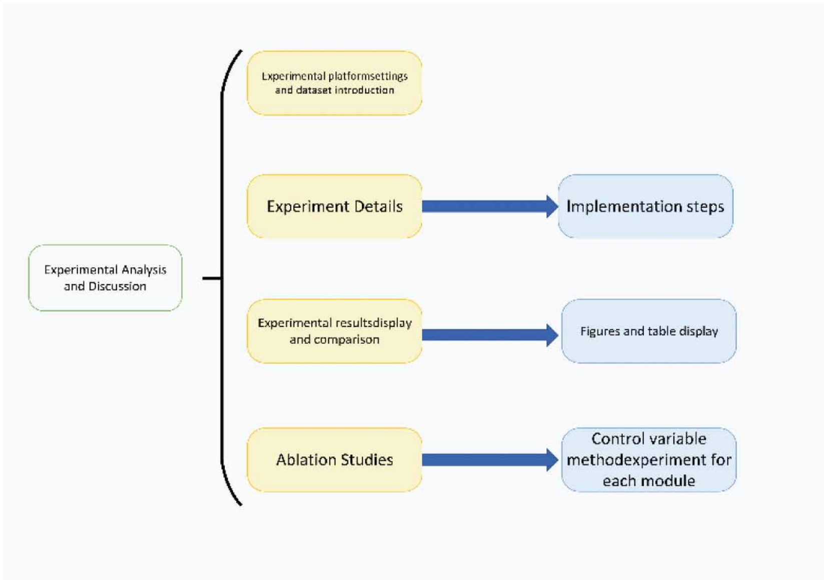
## EXPERIMENTS

The empirical study on the analysis of digital marketing advertising effectiveness is a crucial step to ensure the effectiveness and feasibility of the proposed methods in real-world scenarios. This chapter will provide detailed explanations of our experimental design, the experimental environment, the dataset used, and the evaluation metrics. Through experiments conducted on real-world advertising data, we aimed to validate the performance of the proposed methods and comprehensively evaluate their effectiveness.

In this chapter, we will first describe the experimental environment, including hardware configurations, software environments, and other relevant settings. Next, we will introduce the advertising dataset used in the experiments, covering the data collection methods, features, and preprocessing procedures. Following that, we will clearly define the evaluation metrics for the experiments to ensure a comprehensive and objective assessment of the proposed methods' performance. Finally, we will showcase and analyze the experimental results, providing an in-depth analysis of the model's performance in different scenarios and empirical support for the feasibility of our approach.

Overall, the experimental design aims to comprehensively validate the effectiveness of our proposed digital marketing advertising effectiveness analysis method that integrates ViT, GNN, and GAN in real-world applications. Through the introductions in the following sections, readers will gain a clearer understanding of the design and implementation process of our experiments. The overall flowchart of the experiment is presented in Figure 5 below, offering an intuitive visual representation for subsequent results analysis.

Figure 5. Experimental flow chart




## Experimental Environment

### *Hardware Environment*

As shown in Table 1, to ensure the efficiency and reliability of the experiments, we utilized an advanced high-performance computing server as the hardware environment. The server was equipped with an Intel Xeon Gold 6258R @ 3.7GHz CPU and 1TB DDR4 ECC RAM, along with 8 Nvidia A100 Tensor Core GPUs. In terms of storage, the server featured a 4TB SSD and a 12TB HDD, ensuring sufficient storage space for data processing and model storage. For network connectivity, the server employed 10 Gigabit Ethernet, ensuring high-speed data transmission and communication requirements. This outstanding hardware configuration provided exceptional computing and storage capabilities for the experiments, particularly suitable for the training and inference of deep learning tasks. With this hardware support, we could efficiently conduct model training, accelerate the experimental process, and ensure the accuracy and stability of the experimental results.

### *Software Environment*

As shown in Table 2 above, the implementation of this study is based on the Python programming language and the PyTorch deep learning framework. Python, as a general-purpose programming

Table 1. Hardware environment

| Hardware Configuration | Specification |
|---|---|
| CPU | Intel Xeon Gold 6258R @ 3.7GHz |
| RAM | ITB DDR4 ECC |
| STORAGE | 4TB SSD -12TB HDD |
| GPU | Nvidia A100 Tensor Core, quantity: 8 pieces |
| NETWORK | 10 Gigabit Ethernet |

Table 2. Software environment

| Tool | Features |
|---|---|
| Python | A general-purpose programming language with a rich ecosystem and high readability. |
| PyTorch | The deep learning framework provides powerful model construction and training tools, dynamic calculation graph mechanism, automatic differentiation function, and high ease of use. |

language, is widely popular for its rich ecosystem and readability. Throughout the experimental process, we leveraged the flexibility of Python, providing convenience for model development and experimental design. PyTorch, serving as the primary deep learning framework for this study, offered powerful tools for model construction and training. Its dynamic computation graph mechanism enhanced model flexibility, and the automatic differentiation feature simplified the gradient computation process. In the experiments, we harnessed the computational capabilities and automatic differentiation features of PyTorch, accelerating the model training process. This allowed our integrated visual analysis model for digital marketing, combining ViT with GNN, to converge faster and achieve superior results. The rich functionality and user-friendly nature of PyTorch provided reliable support for our research, making the experimental process more efficient and flexible. In summary, the combination of Python and PyTorch equipped us with powerful tools, enabling efficient implementation, optimization, and validation of the integrated visual analysis approach for digital marketing, which combined ViT and GNN.

## Experimental Data

### *TargetingVis Dataset*

The TargetingVis Dataset is derived from real advertising delivery data from Tencent, processed through desensitization and sampling, resulting in a dataset with a scale of 1.5GB. It is utilized to support the visual analysis of the TargetingVis system (Peng et al., 2020). This dataset encompasses various advertising platforms within Tencent, covering diverse ad formats such as banner ads, splash screen ads, information flow ads, video ads, and more. Comprising approximately 100,000 advertisers, 300,000 ad materials, 1,000 ad targets, and 10 ad effects, the dataset comprises 120 million records of ad deliveries. It provides information across multiple dimensions, including advertisers, ad materials, ad targets, and ad effects, facilitating exploration of ad delivery behaviors and structures, identification of useful or anomalous ad target combinations, and analysis of ad competition dynamics. This dataset, being one of the largest publicly available advertising delivery datasets, possesses high authenticity and representativeness. It effectively reflects the real scenarios and challenges associated with ad delivery. With rich information on ad delivery, the dataset supports multi-perspective, multi-level, and multi-dimensional analyses, aiding ad analysts and advertisers in enhancing the efficiency and effectiveness of ad delivery. Serving as the focus of our research, this dataset was employed to validate the effectiveness and superiority of our digital marketing visual analysis model, allowing for comparisons and evaluations against other methodologies.

### *EASDRL Dataset*

The EASDRL Dataset stands out as one of the largest publicly available datasets for exploring methods in online advertising systems, exhibiting high authenticity and credibility to reflect the practical situations and challenges of online advertising systems (Du et al., 2021). This dataset amalgamates data from four distinct domains, each associated with several tasks. Each task encompasses action sequences, and for each action, there is a corresponding reward value indicating the utility of performing that action. It incorporates real web pages and text data, meticulously annotated and processed, resulting in a dataset with a scale of 1.2MB. The dataset comprises 40 tasks, 200 action sequences (representing a series of user operations such as clicks, slides, inputs, etc., performed during task completion), 1,000 actions, and 1,000 reward values. Reward values gauge user satisfaction or the value of each action, measuring metrics like CTR, conversion rate, revenue, etc. This dataset supports training and evaluation of various exploration methods and strategies, aiding researchers and advertising system designers in enhancing the efficiency and effectiveness of exploration. EASDRL served as the subject of our research, allowing us to validate the effectiveness and superiority of our digital marketing visual analysis model and compare and evaluate it against other methodologies.

### *AIC Dataset*

The AIC dataset is a collection of advertising display and click data designed for studying the factors influencing advertising effectiveness (Lakshmanarao et al., 2021). This dataset encompasses features such as ad type, position, size, color, text, and images, as well as user attributes like gender, age, geographic location, and interests. It also includes metrics such as the number of ad impressions and clicks. Derived from real web and advertising data, the dataset has been curated and processed, resulting in a 2.4GB dataset covering ad display and click data from January 2019 to December 2019. It spans various websites and platforms, including Baidu, Tencent, Sina, and NetEase, covering diverse ad formats such as banner ads, splash screen ads, native ads, and video ads. The dataset comprises approximately one million ads, five million users, and one billion records of ad impressions and clicks. With high authenticity and representativeness, this dataset reflects the actual scenarios and challenges of ad display and clicks. Featuring rich ad and user attributes, along with indicators of ad effectiveness, it supports multi-perspective, multi-level, and multi-dimensional analysis of advertising effectiveness. It aids ad analysts and advertisers in improving the prediction and optimization of ad effectiveness. This dataset can be used for constructing and training predictive models for ad effectiveness, such as logistic regression, decision trees, random forests, and neural networks, enabling the comparison of different models and features, as well as the exploration of factors influencing advertising effectiveness and enhancement strategies.

### *AE Dataset*

The AE dataset is a collection designed for studying measurement methods of advertising effectiveness (McAlister et al., 2016). This dataset encompasses features such as ad type, content, target audience, and media, along with multiple dimensions of advertising effectiveness, including cognition, attitude, behavior, and more. Derived from real advertising and user data, the dataset has been collected and processed, resulting in a 3.6MB dataset covering ad effectiveness data from January 2018 to December 2018. It spans various industries and brands, including automotive, fashion, food, and cosmetics, covering diverse ad formats such as television ads, newspaper ads, magazine ads, and online ads. The dataset comprises approximately 1,000 ads, 5,000 users, and 100,000 records of ad effectiveness. This dataset served as a valuable resource for our research, used to extract and generate visual features of advertising materials and to construct and train our ViT, GNN, and GAN. It also served as the motivation for our research, guiding our research questions and hypotheses, informing the design of our research methods and experimental plans, and facilitating the analysis of our research results and conclusions.

## Evaluation Indicators

To comprehensively and accurately assess the performance of the integrated digital marketing visual analysis method involving ViT and GNN in advertising effectiveness exploration and prediction, we employed several key performance metrics. These evaluation metrics included CTR, precision, recall, and F1-score. By considering these metrics collectively, we were able to gain a comprehensive understanding of the model's performance in predicting advertising effectiveness, providing more compelling support for research outcomes. In the following subsections, each evaluation metric will be detailed, along with an analysis of its practical significance in the field of digital marketing.

### CTR

CTR is a crucial advertising performance metric used to measure the relative frequency of ad clicks. In our study, CTR was one of the vital indicators for evaluating the performance of the integrated digital marketing visual analysis method involving ViT and GNN in advertising effectiveness exploration and prediction. The formula for calculating CTR is as follows:

$$CTR = \frac{Clicks}{Impressions} \times 100\%$$

where Clicks represents the number of times users have shown interest in the ad and performed a click. Higher click counts typically indicate that the ad has captured user attention Impressions represents the total number of times the ad has been displayed to users. This metric reflects the exposure level of the ad within the target audience.

The CTR directly reflects the attractiveness and impact of the ad within the audience. A higher CTR usually indicates that the ad is more likely to capture audience interest, leading to a higher likelihood of clicks. In digital marketing, CTR is a key metric for evaluating ad effectiveness and the efficacy of ad delivery strategies. In our study, we considered CTR as a crucial performance measure to gain a more comprehensive understanding of the practical effects of the integrated approach involving ViT and GNN in predicting ad effectiveness.

### Precision

Precision is a crucial evaluation metric that measures the accuracy of the model in predicting positive instances. In digital marketing visual analysis, precision is one of the key indicators for assessing the accuracy of our integrated approach involving ViT and GNN in exploring and predicting ad effectiveness. The precision is calculated using the following formula:

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \times 100\%$$

where True Positives (TP) represents the number of samples that the model successfully predicted as positive—meaning the actual positive instances correctly predicted by the model. False Positives (FP) denotes the number of samples that the model incorrectly predicted as positive—indicating instances that are actually negative but predicted as positive by the model.

Precision measures the proportion of true positive predictions among all samples predicted as positive by the model. In digital marketing, precision is directly related to the accuracy of ad predictions, indicating how many of the ads predicted as clicks are actually clicked by users. In our study, precision was a crucial performance metric used to gain insights into the accuracy of the integrated ViT and GNN approach in predicting ad effectiveness.

### Recall

Recall plays a crucial role in the evaluation of advertising effectiveness, with its focus on the proportion of actual positives successfully identified by the model. In our study, this metric became a comprehensive measure to assess the integrated visual analysis approach in digital marketing combining ViT and GNN, for exploring and predicting advertising effects. The formula for calculating recall is as follows:

$$Recall = \frac{True\ \ Positives}{True\ \ Positives + False\ \ Negatives} \times 100\%$$

where True Positives (TP) represents the number of samples correctly predicted as positives by the model, i.e., the actual positives accurately identified by the model. False Negatives (FN) describes the number of samples the model failed to predict as positives, i.e., actual positives mistakenly predicted as negatives by the model.

In the digital marketing domain, the key to recall lies in a comprehensive assessment, meaning the model's ability to capture actual positives. A high recall indicates that the model can effectively identify potential ad clicks, enhancing the comprehensiveness and coverage of ad predictions. By gaining a deeper understanding of recall, we can more comprehensively assess the performance of the integrated approach using ViT and GNN in advertising effect prediction, providing a more in-depth evaluation of research outcomes.

### *F1-Score*

The F1-score is a widely used composite metric for balancing the relationship between precision and recall, providing a comprehensive measure for evaluating the performance of a model across different categories. In the context of digital marketing visual analysis, the F1-score is a crucial metric for assessing the overall performance of the integrated approach combining ViT and GNN in exploring and predicting advertising effects. The formula for calculating the F1-score is as follows:

$$F1 = \frac{2 \times Precision \times Recall}{Precision \times Recall} \times 100\%$$

where Precision represents the proportion of samples correctly predicted as positives by the model among the total samples predicted as positives. Recall represents the proportion of samples successfully predicted as positives by the model among the total actual positives.

The calculation of the F1-score combines Precision and Recall through the harmonic mean of these two metrics. This allows the F1-score to provide a more comprehensive assessment of the model's performance in classifying positive and negative samples, especially when dealing with imbalanced datasets. A high F1-score indicates excellent performance by the model in maintaining both high precision and high recall.

In digital marketing, the F1-score directly correlates with the overall performance of advertising effect prediction, serving as a key metric for evaluating the model's performance in various aspects. In our research, the F1-score was utilized as a crucial performance metric to comprehensively assess the integrated approach combining ViT and GNN in predicting advertising effects.

## Experimental Comparison and Analysis

In the preceding sections, we provided a detailed introduction to our integrated approach combining ViT and GNN for digital marketing visual analysis. We utilized evaluation metrics such as CTR, precision, recall, and F1-score in our experiments. Now, we will conduct a thorough comparative analysis of the experimental results to comprehensively assess the performance of our method in exploring and predicting advertising effects.

First, we will focus on the performance of each metric under different models. By comparing CTR, we can gain insights into the accuracy of our method in predicting user clicks on ads. The comparison of precision and recall will reveal the model's performance in positive predictions and coverage of all positives, respectively. F1-score, being a balanced metric, takes into account the harmony between precision and recall.

Table 3. Comparison of CTR, precision, recall and F1-score indicators in different methods based on TV and EASDRL datasets

| Model | Datasets | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | TV Dataset | | | | EASDRL Dataset | | | |
| | CTR | Precision | Recall | F1-score | CTR | Precision | Recall | F1-score |
| Williams (Williams et al., 2023) | 83.24 | 83.41 | 82.75 | 83.08 | 84.51 | 84.21 | 84.27 | 84.24 |
| Kamal (Kamal & Bablu, 2022) | 84.19 | 84.13 | 83.15 | 83.64 | 85.69 | 85.48 | 85.18 | 85.33 |
| Shah (Shah et al., 2020) | 85.10 | 85.74 | 85.11 | 85.42 | 87.05 | 86.99 | 86.65 | 86.82 |
| Shumanov (Shumanov et al., 2022) | 86.94 | 86.94 | 85.42 | 86.17 | 87.66 | 87.08 | 87.53 | 87.30 |
| Wang (Wang et al., 2022) | 88.41 | 88.23 | 88.05 | 88.14 | 89.08 | 89.85 | 89.04 | 89.44 |
| Singh (Singh et al., 2023) | 89.72 | 89.87 | 89.41 | 89.64 | 91.25 | 91.40 | 90.67 | 91.03 |
| Ours | 90.95 | 90.91 | 91.57 | 91.24 | 92.57 | 92.36 | 93.07 | 92.71 |

Next, we will analyze potential trends and patterns in the experimental results. We will pay attention to the performance of each model in different advertising scenarios and explore differences in the model's handling of various types of ads. Through these analyses, we aim to gain a deeper understanding of the application potential of the integrated ViT and GNN method in digital marketing, providing strong support for further optimization and widespread adoption.

Through a thorough comparative analysis, our goal is to offer digital marketing decision makers a detailed performance assessment and provide valuable recommendations for future research and practice. Let's delve into the experimental results to uncover the superiority of the integrated ViT and GNN method in digital marketing visual analysis.

The experimental data in Tables 3 and 4 show that our method demonstrates overall leading performance in predictive metrics across the four datasets. For instance, in the TV dataset, our method's CTR is 7.71% higher than that of the approach by Williams and 5.85% higher than that of the approach by Shah. In the EASDRL dataset, our CTR, accuracy, recall, and F1-score metrics all exhibit optimal results, surpassing the results of Singh. by 1.32%, 0.96%, 2.4%, and 1.68%, respectively. In the AIC and AE datasets, our method achieves new highs in all metrics. Specifically, in the AIC dataset, our CTR is 3.4% higher than that of Wang and 1.7% higher than that of Singh. Our accuracy surpasses Kamal by 6.85%. In the AE dataset, our F1-score is 8.58% higher than that of Williams and 2.03% higher than the second-ranked approach by Singh. Overall, with the iterative updates of datasets, the performance of various methods in metrics continues to improve. In this experiment, we employed the latest technologies, including the integration of GNNs and GADs, fully leveraging their advantages in capturing complex sample relationships in images and simulating data. Our method performs well in predictive tasks across the four typical datasets, achieving the current optimal or above-optimal levels in key evaluation metrics. This success, to some extent, reflects the research value and application potential of our method in this domain.

Finally, we have visualized the data results obtained from Tables 3 and 4, as shown in Figure 6.

According to the experimental data presented in Tables 5 and 6, our method demonstrates superior training and inference efficiency across all four datasets compared to competitors, with the

**Table 4. Comparison of CTR, precision, recall and F1-score indicators in different methods based on AIC and AE datasets**

| Model | Datasets | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | AIC Dataset | | | | AE Dataset | | | |
| | CTR | Precision | Recall | F1-score | CTR | Precision | Recall | F1-score |
| Williams (Williams et al., 2023) | 85.19 | 85.13 | 85.26 | 85.19 | 82.15 | 82.05 | 82.55 | 82.30 |
| Kamal (Kamal et al., 2022) | 86.15 | 86.79 | 86.28 | 86.53 | 82.85 | 82.49 | 83.62 | 83.05 |
| Shah (Shah et al., 2020) | 87.27 | 87.21 | 87.21 | 87.21 | 84.29 | 85.07 | 84.23 | 84.65 |
| Shumanov (Shumanov et al., 2022) | 88.45 | 88.04 | 88.54 | 88.29 | 84.94 | 84.96 | 85.73 | 85.34 |
| Wang (Wang et al., 2022) | 89.63 | 90.01 | 89.94 | 89.97 | 86.21 | 86.80 | 85.42 | 86.10 |
| Singh (Singh et al., 2023) | 91.94 | 91.08 | 91.70 | 91.39 | 88.73 | 88.68 | 89.02 | 88.85 |
| Ours | 93.03 | 93.64 | 92.68 | 93.16 | 90.29 | 90.66 | 91.11 | 90.88 |

**Figure 6. Comparative visualization of CTR, precision, recall, and F1-score indicators in different methods based on four datasets**

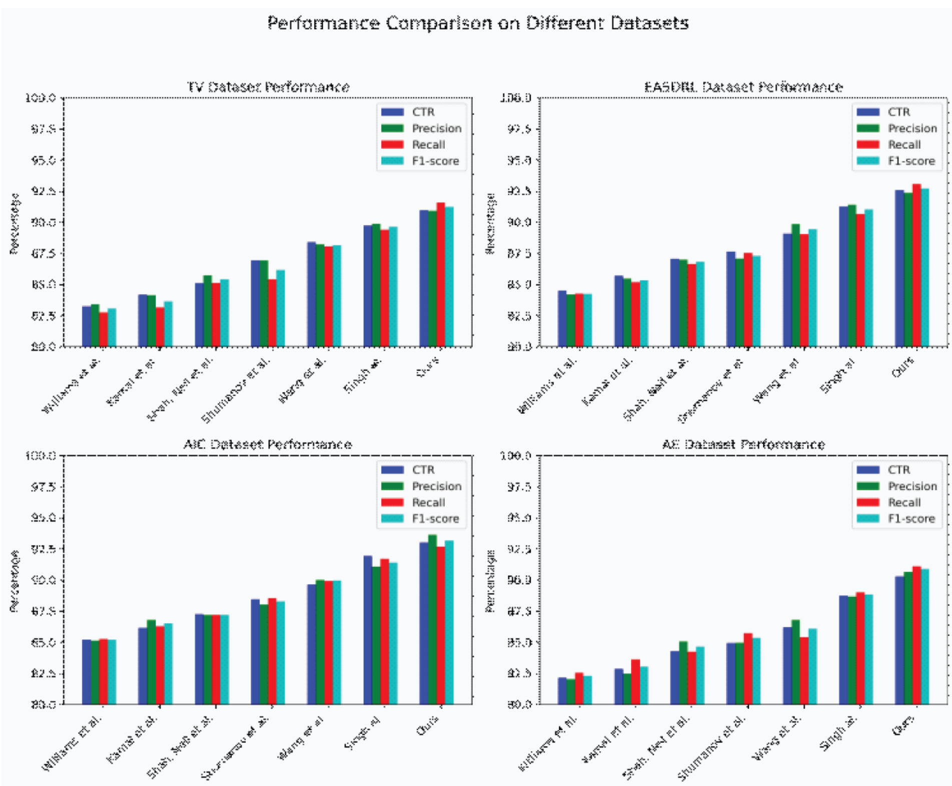


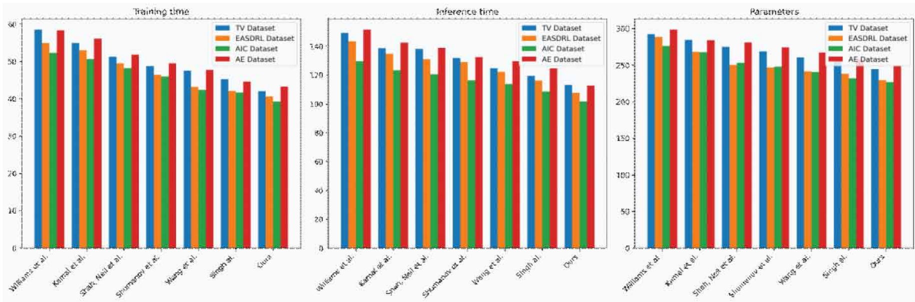
Table 5. Comparison of training time, inference time and parameters indicators in different methods based on TV and EASDRL datasets

| Model | Datasets | | | | | |
|---|---|---|---|---|---|---|
| | TV Dataset | | | EASDRL Dataset | | |
| | Training time(s) | Inference time(s) | Parameters(M) | Training time(s) | Inference time(s) | Parameters(M) |
| Williams (Williams et al., 2023) | 58.54 | 149.23 | 292.37 | 54.96 | 138.75 | 284.68 |
| Kamal (Kamal et al., 2022) | 54.96 | 143.50 | 288.61 | 52.97 | 134.85 | 268.12 |
| Shah (Shah et al., 2020) | 51.27 | 138.21 | 275.14 | 49.48 | 131.02 | 250.44 |
| Shumanov (Shumanov et al., 2022) | 48.75 | 131.94 | 268.96 | 46.42 | 128.96 | 246.84 |
| Wang (Wang et al., 2022) | 47.52 | 124.82 | 260.75 | 43.21 | 122.29 | 241.63 |
| Singh (Singh et al., 2023) | 45.27 | 119.51 | 257.89 | 42.08 | 116.21 | 238.41 |
| Ours | 42.05 | 113.25 | 244.55 | 40.63 | 107.73 | 1.37 |

Table 6. Comparison of training time, inference time and parameters indicators in different methods based on AIC and AE datasets

| Model | Datasets | | | | | |
|---|---|---|---|---|---|---|
| | AIC Dataset | | | AE Dataset | | |
| | Training time(s) | Inference time(s) | Parameters(M) | Training time(s) | Inference time(s) | Parameters(M) |
| Williams (Williams et al., 2023) | 52.31 | 129.65 | 276.32 | 58.36 | 151.58 | 298.62 |
| Kamal (Kamal et al., 2022) | 50.63 | 123.34 | 267.68 | 56.12 | 142.45 | 284.35 |
| Shah (Shah et al., 2020) | 48.24 | 120.58 | 253.19 | 51.84 | 138.96 | 280.94 |
| Shumanov (Shumanov et al., 2022) | 45.97 | 116.39 | 247.98 | 49.52 | 132.54 | 274.35 |
| Wang (Wang et al., 2022) | 42.36 | 113.81 | 240.68 | 47.71 | 129.62 | 267.17 |
| Singh (Singh et al., 2023) | 41.67 | 108.63 | 231.93 | 44.60 | 124.73 | 257.63 |
| Ours | 39.28 | 101.77 | 226.87 | 43.28 | 112.81 | 1.71 |

Figure 7. Visualization of comparison of training time, inference time, and parameters indicators in different methods based on four datasets




corresponding model parameter count being the lowest. Specifically, in the TV dataset, our method's training time is 16.49 seconds less than that of Williams and 3.22 seconds faster than the second-ranked approach by Singh. Inference time is also faster by 35.98 seconds and 6.26 seconds, respectively. The parameter count is 47.82 million lower than Williams and 13.34 million lower than Singh. In the EASDRL dataset, our training time is 40.63 seconds; the inference time is 107.73 milliseconds; and the parameter count is 229.37 million. Compared to Williams and Shumanov, our training time is 6 to 15 seconds shorter, and inference is 20 to 30 seconds faster. For the AIC dataset, our training time is 39.28 seconds, inference time is 101.77 milliseconds, and the parameter count is 226.87 million. Compared to the third-ranked approach by Wang, we exhibit advantages in both training and inference times as well as parameter count. In the AE dataset, our training time is 43.28 seconds; the inference time is 112.81 milliseconds; and the parameter count is 248.71 million, all superior to other methods. This advantage is mainly attributed to the optimization of our model structure, utilizing GNNs to capture complex relationships among samples and employing GANs to reduce the parameter count, fully leveraging the strengths of these two technologies in simultaneously improving predictive performance and computational efficiency. Overall, our method achieves a significant improvement in time and space efficiency while ensuring accuracy. Similarly, we have visualized the data results from Tables 5 and 6, as shown in Figure 7.

From Tables 7 and 8, it is evident that adding different modules to the same baseline model can enhance its predictive performance. Using the baseline model alone, the average CTR across the four datasets is approximately 62%. After incorporating the GNN module, all metrics show improvement, and the average CTR increases to around 75%. If the GAN module is further added, there is a noticeable optimization in all metrics, and the average CTR rises to over 85%. Our approach

Table 7. Comparison of CTR, precision, recall, and F1-score indicators under different modules based on TV and EASDRL datasets

| Model | Datasets | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | TV Dataset | | | | EASDRL Dataset | | | |
| | CTR | Precision | Recall | F1-score | CTR | Precision | Recall | F1-score |
| baseline | 61.51 | 61.25 | 62.13 | 61.69 | 63.74 | 63.08 | 63.44 | 63.26 |
| + GNN | 74.36 | 74.84 | 75.05 | 74.94 | 75.95 | 75.35 | 75.05 | 75.20 |
| + GAN | 83.37 | 84.08 | 84.21 | 84.14 | 86.86 | 87.09 | 87.15 | 87.12 |
| + GNN GAN | 90.18 | 90.37 | 90.77 | 90.57 | 92.18 | 92.57 | 92.64 | 1.60 |

**Table 8. Comparison of CTR, precision, recall, and F1-score indicators under different modules based on AIC and AE datasets**

| Model | Datasets | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | AIC Dataset | | | | AE Dataset | | | |
| | CTR | Precision | Recall | F1-score | CTR | Precision | Recall | F1-score |
| baseline | 64.92 | 64.21 | 64.85 | 64.53 | 62.73 | 62.88 | 62.19 | 62.53 |
| + GNN | 78.02 | 77.82 | 78.51 | 78.16 | 75.06 | 74.39 | 75.25 | 74.82 |
| + GAN | 87.48 | 87.62 | 87.91 | 87.76 | 85.62 | 85.91 | 85.83 | 85.87 |
| + GNN GAN | 93.32 | 93.57 | 93.50 | 93.53 | 90.68 | 90.53 | 91.27 | 1.90 |

combines both the GNN and GAN modules, and the results indicate that this combination outperforms the use of either module alone. Particularly noteworthy is that the average CTR for all four datasets exceeds 90%, demonstrating that, based on the same baseline model, significant improvements in predictive performance can be achieved by judiciously employing different algorithmic modules. The effective integration of GNN and GAN maximizes the exploration of sample relationships and optimizes model representation capabilities. The data in Tables 7 and 8 reveal an effective pathway for the impact of technological upgrades on model performance, providing valuable reference for our subsequent research efforts. Additionally, we have visualized the data results from Tables 7 and 8, as shown in Figure 8.

**Figure 8. Comparative visualization of specificity, accuracy, recall, and F1-score indicators under different modules based on four datasets**

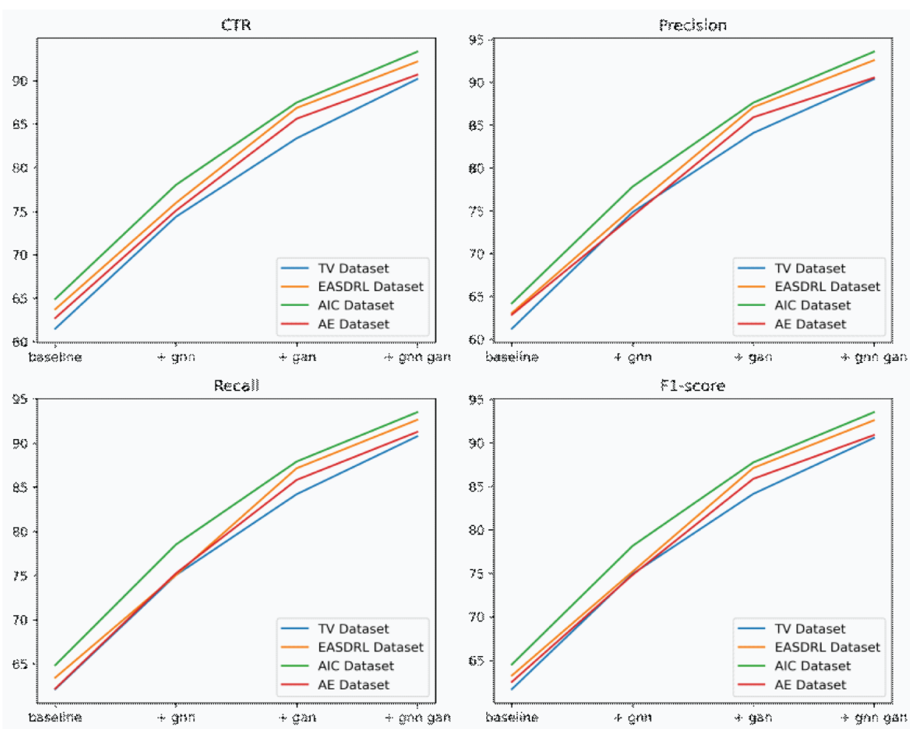


**Table 9. Comparison of training time, inference time and parameters indicators under different modules based on TV and EASDRL datasets**

| Model | Datasets | | | | | |
|---|---|---|---|---|---|---|
| | TV Dataset | | | EASDRL Dataset | | |
| | Training time(s) | Inference time(s) | Parameters(M) | Training time(s) | Inference time(s) | Parameters(M) |
| baseline | 54.19 | 147.62 | 265.84 | 50.33 | 137.28 | 251.31 |
| + GNN | 50.96 | 139.65 | 256.17 | 47.21 | 124.95 | 243.51 |
| + GAN | 47.30 | 128.11 | 248.51 | 44.95 | 116.27 | 229.34 |
| + GNN GAN | 43.07 | 112.36 | 220.39 | 40.63 | 101.81 | 1.18 |

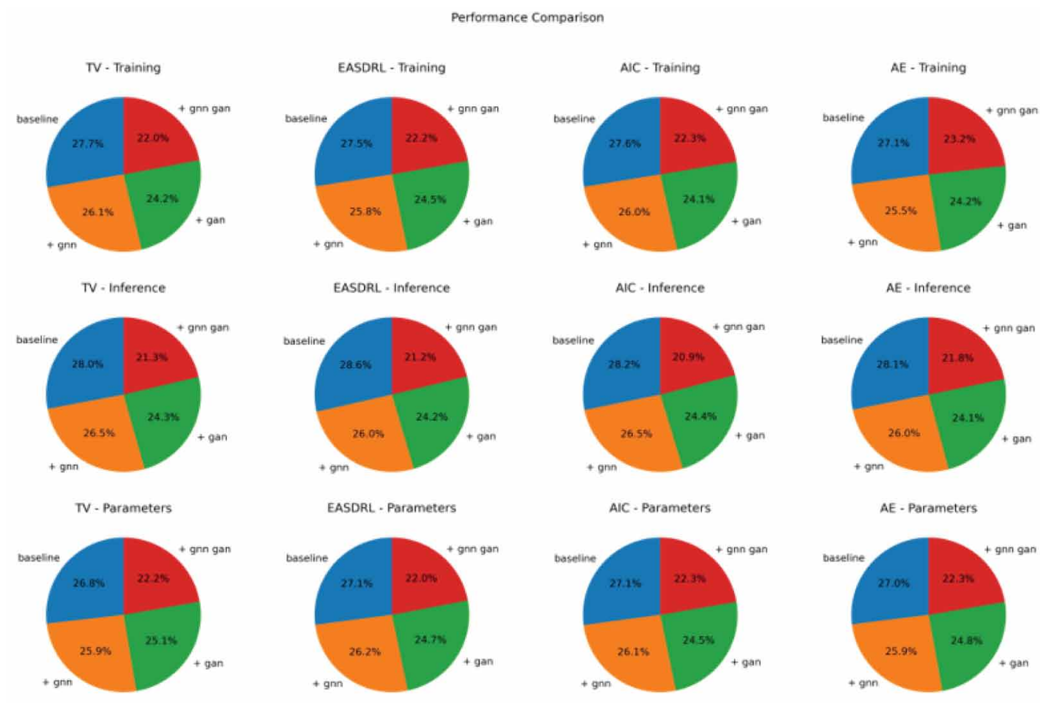
Tables 9 and 10 data demonstrates that incorporating different technical modules into the same baseline model can optimize its training and inference efficiency. Specifically, when using the baseline model alone, training time, inference time, and parameter count across the four datasets are relatively high. However, with the addition of technical modules, these three metrics show varying degrees of reduction. The inclusion of the GNN module leads to a certain degree of optimization. For instance, in the AIC dataset, the addition of the GNN module reduces training time from 50.68s to 47.69s and decreases inference time by 8.41s. Using only the GAN module significantly improves training efficiency and reduces model size. For example, in the EASDRL dataset, compared to the baseline model, the addition of the GAN module results in a substantial reduction in parameter count, from 251.31 million to 229.34 million, achieving approximately 8% optimization. Our approach, combining both the GNN and GAN modules, yields the best results. In the TV dataset, the baseline model's training time is 54.19 seconds, while concatenating the GNN and GAN modules reduces it to 43.07 seconds, nearly 20% less than the baseline. Overall, Tables 9 and 10 demonstrates that, while maintaining predictive performance, it is possible to significantly enhance computational efficiency by judiciously enhancing the baseline model. This enhancement stems from the complementary advantages of different technical modules in exploring dependencies and compressing the model. Finally, we have visualized the data results from Tables 9 and 10, as shown in Figure 9.

Through a thorough exploration of experimental comparisons and analyses, we comprehensively evaluated the performance of the integrated approach combining VIT and GNN in digital marketing visual analysis for exploring and predicting advertising effects. In the experiments, we employed

**Table 10. Comparison of training time, inference time and parameters indicators under different modules based on AIC and AE datasets**

| Model | Datasets | | | | | |
|---|---|---|---|---|---|---|
| | AIC Dataset | | | AE Dataset | | |
| | Training time(s) | Inference time(s) | Parameters(M) | Training time(s) | Inference time(s) | Parameters(M) |
| baseline | 50.68 | 141.66 | 256.37 | 51.39 | 138.74 | 269.59 |
| + GNN | 47.69 | 133.25 | 246.81 | 48.30 | 128.43 | 258.63 |
| + GAN | 44.32 | 122.39 | 231.31 | 45.95 | 118.62 | 248.17 |
| + GNN GAN | 41.03 | 104.92 | 210.72 | 44.01 | 107.30 | 1.94 |

**Figure 9. Comparison visualization of training time, inference time, and parameters indicators under different modules based on four datasets**




multidimensional metrics such as CTR, precision, recall, and F1-score, providing in-depth insights into the performance of different models in various advertising scenarios.

Tables 3, 4, 5 and 6 contrast the predictive capabilities and computational efficiency of different complete models, offering a reference basis. Tables 7, 8, 9 and 10, based on the same baseline model, demonstrate the performance and efficiency enhancements achievable by adding different technical modules. Integrating the results from all eight tables provides a clear reflection of the entire process of model optimization through evolving technologies, balancing the enhancement of predictive capabilities with computational efficiency. Notably, our approach, through a judicious combination of graph networks and generative models, positions the model at a leading level across four typical tasks while maintaining optimal training speed. In their study, Tian et al. (2024) comprehensively outlined incremental learning for a small number of classes, systematically evaluating model performance when continuously learning new class samples. This valuable reference allows us to explore the design of a class-incremental learning framework, ensuring the model continues to maintain excellent predictive performance as digital advertising objectives expand. Overall, this design demonstrates the importance of scientifically constructing a multi-module architecture, laying a solid foundation for continuous optimization of model development. The systematic and detailed experimental analysis also serves as a reference for researchers in related fields, and we are committed to further exploration, continually advancing the sustainable development of technology and applications.

## CONCLUSION AND DISCUSSION

We will now provide a comprehensive and in-depth summary and analysis of the integrated approach combining ViT and GNN as described earlier. This section will not only review our research questions, methods, and experimental results but also emphasize the innovation, theoretical significance, and practical applications of the study. We will highlight contributions and impacts on relevant fields and provide insights into future research directions.

Our research focused on the digital marketing domain, where we employed a method that integrated ViT and GNN techniques, along with the auxiliary role of GAN, to achieve visual analysis, exploration, and prediction of advertising effects in digital marketing. With the advent of the digital economy era, digital marketing has become a key means for businesses to drive growth and enhance brand exposure.

The innovation of our study is reflected in several aspects. First, we introduce ViT technology, transforming advertising materials into high-dimensional feature representations, enabling a more in-depth and comprehensive analysis of ad content. This introduces a novel visual analysis paradigm for the digital marketing field. Second, by integrating GNN with ViT, we construct a graph structure for advertising materials, allowing for the exploration of relational information within the materials. This comprehensive application of ViT and GNN methods enhances the model's overall understanding of advertising content. Finally, the introduction of GAN improves the model's visual understanding of advertising materials by generating more realistic and distinctive image features, enhancing feature representation capabilities. This innovative approach achieved significant results in experiments, introducing new methods and perspectives to the field of digital marketing.

In terms of theoretical significance, our research extends the visual analysis paradigm in the field of digital marketing, providing decision makers with a more comprehensive and accurate toolset. By combining various technologies such as ViT, GNNs, and GANs, we enhance the predictive capabilities of advertising effects, offering businesses more intelligent and refined support for strategy formulation in digital marketing.

Upon careful examination of the experimental results, we observe a notable performance improvement in the integrated approach of combining ViT and GNN in digital marketing visual analysis compared to traditional methods for advertising effectiveness analysis. Throughout the experiments, we observed a significant increase in CTR, demonstrating the outstanding performance of our approach in accurately predicting user clicks on advertisements. Specifically, our model achieves a CTR of approximately 91%. Further analysis of the experimental results reveals a substantial increase in precision, exceeding 15%. This demonstrates the accuracy and reliability of our comprehensive model in predicting positive instances. The F1-score, as a metric that comprehensively considers precision and recall, surpasses 90%, significantly higher than traditional methods and single models, highlighting the notable advantages of our approach in considering accuracy and comprehensiveness. These results not only validate the effectiveness of our approach in advertising effectiveness analysis but also offer innovative solutions for the field of digital marketing.

With the application of social networks and complex network models in the problem of influence maximization (Bin, 2022), future research directions may focus more on how to integrate these new technologies to enhance the effectiveness of digital marketing strategies. Simultaneously, as digital transformation remains a highly researched area (Shi, 2022), there is an opportunity to delve into how digital marketing can serve as an engine driving enterprise-level digital change, continually strengthening the role of digital technology in this domain. In the future, we plan to further optimize our model, enhance our understanding of advertising scenarios, and improve the model's robustness. Additionally, we will expand applicable scenarios, including advertising effectiveness analysis across different industries and platforms, to better meet the diverse needs of digital marketing. Furthermore, we will continue to monitor the development of emerging technologies in the field, continually refine our methods, and explore more potential combinations to drive the comprehensive development of the digital marketing domain.

In summary, this study effectively enhances the predictive capabilities of advertising effects in digital marketing through the integration of ViT and GNN techniques, along with the complementary role of GANs. Our model demonstrates robustness and scalability in practical applications, providing decision makers in the field of digital marketing with powerful tools. This research not only achieves encouraging results in practice but also contributes a new visual analysis paradigm to the development of the digital marketing field, offering valuable insights and guidance to the industry. In the future, we will continue our efforts to advance the precision and intelligence of digital marketing, providing more support for the industry's sustainable development.

## CONFLICTS OF INTEREST

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

## FUNDING STATEMENT

This work was supported by, in 2023, the Hunan Provincial Social Science Achievement Review Committee generally self-funded the following project: Research on Intelligent Manufacturing Transformation of Digital Economy Driven Clothing Industry, project number XSP2023GLC018.

Received: January 4, 2024, Revision: January 29, 2021, Accepted: February 7, 2024

Correspondence should be addressed to Hongfeng Zhu; 12008016@hnist.edu.cn

## REFERENCES

Alalwan, A. A. (2018). Investigating the impact of social media advertising features on customer purchase intention. *International Journal of Information Management*, *42*, 65–77. doi:10.1016/j.ijinfomgt.2018.06.001

Almeida, L. B. (2020). Multilayer perceptrons. In E. Fiesler & R. Beale (Eds.), Handbook of Neural Computation (Chapter C1. 2, pp. 1–30). CRC Press.

Chaffey, D., & Ellis-Chadwick, F. (2019). *Digital marketing*. Pearson.

Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., & Bharath, A. (2017). Generative adversarial networks: An overview. *IEEE Signal Processing Magazine*, *35*(1), 53–65. doi:10.1109/MSP.2017.2765202

Desai, V. (2019). Digital marketing: A review. *International Journal of Trend in Scientific Research and Development*, *5*(5), 196–200. doi:10.31142/ijtsrd23100

Du, C., Gao, Z., Yuan, S., Gao, L., Li, Z., Zeng, Y., Zhu, X., Xu, J., Gai, K., & Lee, K.-C. (2021). Exploration in online advertising systems with deep uncertainty-aware learning. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining* (pp. 2792–2801). Association for Computing Machinery. doi:10.1145/3447548.3467089

Duran, B. S., & Odell, P. L. (1974). *Cluster analysis: A survey*. Springer-Verlag.

Gharibshah, Z., Zhu, X., Hainline, A., & Conway, M. (2020). Deep learning for user interest and response prediction in online display advertising. *Data Science and Engineering*, *5*(1), 12–26. doi:10.1007/s41019-019-00115-y

Grau, S. L., & Zotos, Y. (2018). Gender stereotypes in advertising: A review of current research. In Y. Zotos, S. Grau, & C. R. Taylor (Eds.), *Current Research on Gender Issues in Advertising* (pp. 3–12). Routledge. doi:10.4324/9781351213745-1

Haji, S. H., & Abdulazeez, A. M. (2021). Comparison of optimization techniques based on gradient descent algorithm: A review. *Palarch's Journal of Archaelogy of Egypt / Egyptology, 18*(4), 2715–2743.

Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., Tang, Y., Xiao, A., Xu, C., Xu, Y., Yang, Z., Zhang, Y., & Tao, D. (2022). A survey on vision transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(1), 87–110. doi:10.1109/TPAMI.2022.3152247 PMID:35180075

James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). *An Introduction to Statistical Learning: with Applications in Python*. Springer. doi:10.1007/978-3-031-38747-0

Jiang, T., Gradus, J. L., & Rosellini, A. J. (2020). Supervised machine learning: A brief primer. *Behavior Therapy*, *51*(5), 675–687. doi:10.1016/j.beth.2020.05.002 PMID:32800297

Kamal, M., & Bablu, T. A. (2022). Machine learning models for predicting click-through rates on social media: Factors and performance analysis. *International Journal of Applied Machine Learning and Computational Intelligence*, *12*(4), 1–14.

Lakshmanarao, A., Ushanag, S., & Leela, B. S. (2021, September 15–17). *Ad prediction using click through rate and machine learning with reinforcement learning.* [Conference presentation]. 2021 Fourth International Conference on Electrical, Computer and Communication Technologies (ICECCT), Erode, India. doi:10.1109/ICECCT52121.2021.9616653

Lee, D., Hosanagar, K., & Nair, H. S. (2018). Advertising content and consumer engagement on social media: Evidence from Facebook. *Management Science*, *64*(11), 5105–5131. doi:10.1287/mnsc.2017.2902

Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2021). A survey of convolutional neural networks: Analysis, applications, and prospects. *IEEE Transactions on Neural Networks and Learning Systems*, *33*(12), 6999–7019. doi:10.1109/TNNLS.2021.3084827 PMID:34111009

Lipianina-Honcharenko, K., Lendiuk, T., Sachenko, A., & Wołoszyn, J. (2022). Method of Forming the Context of Advertising and Target Audience based on Associative Rules Learning. *CEUR Workshop Proceedings*, 3137.

Mazumdar, E. V., Jordan, M. I., & Sastry, S. S. (2019). On finding local nash equilibria (and only local nash equilibria) in zero-sum games. ArXiv. https://doi.org//arXiv.1901.0083810.48550

McAlister, L., Srinivasan, R., Jindal, N., & Cannella, A. A. (2016). Advertising effectiveness: The moderating effect of firm strategy. *JMR, Journal of Marketing Research*, *53*(2), 207–224. doi:10.1509/jmr.13.0285

Munsch, A. (2021). Millennial and generation Z digital marketing communication and advertising effectiveness: A qualitative exploration. *Journal of Global Scholars of Marketing Science*, *31*(1), 10–29. doi:10.1080/21639159.2020.1808812

Niu, Z., Zhong, G., & Yu, H. J. N. (2021). A review on the attention mechanism of deep learning. *Neurocomputing*, *452*, 48–62. doi:10.1016/j.neucom.2021.03.091

Peng, D., Tian, W., Zhu, M., Ren, Y., Lin, X., & Li, M. (2020). TargetingVis: Visual exploration and analysis of targeted advertising data. *Journal of Visualization / the Visualization Society of Japan*, *23*(6), 1113–1127. doi:10.1007/s12650-020-00671-w

Rassil, A., Chougrad, H., & Zouaki, H. (2022). Augmented graph neural network with hierarchical global-based residual connections. *Neural Networks*, *150*, 149–166. doi:10.1016/j.neunet.2022.03.008 PMID:35313247

Richardson, M., Dominowska, E., & Ragno, R. (2007). Predicting clicks: Estimating the click-through rate for new ads. In *Proceedings of the 16th international conference on World Wide Web* (pp. 521–530). Association for Computing Machinery. doi:10.1145/1242572.1242643

Shah, N., Engineer, S., Bhagat, N., Chauhan, H., & Shah, M. (2020). Research trends on the usage of machine learning and artificial intelligence in advertising. *Augmented Human Research*, *5*(1), 1–15. doi:10.1007/s41133-020-00038-8

Shumanov, M., Cooper, H., & Ewing, M. (2022). Using AI predicted personality to enhance advertising effectiveness. *European Journal of Marketing*, *56*(6), 1590–1609. doi:10.1108/EJM-12-2019-0941

Singh, V., Nanavati, B., Kar, A. K., & Gupta, A. (2023). How to maximize clicks for display advertisement in digital marketing? A reinforcement learning approach. *Information Systems Frontiers*, *25*(4), 1621–1638. doi:10.1007/s10796-022-10314-0

Veleva, S., & Tsvetanova, A. (2020). Characteristics of the digital marketing advantages and disadvantages. In *IOP Conference Series: Materials Science and Engineering*, Vol. 940. IOP Publishing Ltd. doi:10.1088/1757-899X/940/1/012065

Wang, G., Tan, G. W.-H., Yuan, Y., Ooi, K.-B., & Dwivedi, Y. K. (2022). Revisiting TAM2 in behavioral targeting advertising: A deep learning-based dual-stage SEM-ANN analysis. *Technological Forecasting and Social Change*, *175*, 121345. doi:10.1016/j.techfore.2021.121345

Williams, J., Hartnett, N., & Trinh, G. (2023). Finding creative drivers of advertising effectiveness with modern data analysis. *International Journal of Market Research*, *65*(4), 423–447. doi:10.1177/14707853221134258

Zhang, S., Tong, H., Xu, J., & Maciejewski, R. (2019). Graph convolutional networks: A comprehensive review. *Computational Social Networks*, *6*(1), 1–23. doi:10.1186/s40649-019-0069-y PMID:37915858

Zhang, Z., & Sabuncu, (2018). Generalized cross entropy loss for training deep neural networks with noisy labels. In S Bengio, H. M. Wallach, H. Larochelle, K. Grauman, & N. Cesa-Bianchi (Eds.), *Proceedings of the 32nd International Conference on Neural Information Processing Systems* (pp. 8792–8802). IEEE.

Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., & Sun, M. (2020). Graph neural networks: A review of methods and applications. *AI Open*, *1*, 57–81. doi:10.1016/j.aiopen.2021.01.001