International Journal for Housing Science and Its Applications
Publish August 6, 2025. Volume 46, Issue 3 Pages 7223-7239

https://doi.org/10.70517/ijhsa463619

A New Application Paradigm for Generating Creative Ideas for New Media Animation Advertisements by Improving GANs-VAEs Deep Models

Qian Zhou[1,*]
[1] School of Art, Anhui Wenda College of Information Engineering, Hefei, Anhui, 230000, China
Corresponding authors: (e-mail: zhouqian@mail.wenda.edu.cn).

Abstract In order to optimize the performance of generative adversarial networks on automatic advertisement image generation, this paper combines the variational self-encoder with generative adversarial networks, which consists of four parts: encoder network, decoder network, target-to-be-attacked network, and discriminator network to form a new adversarial sample generation method based on GANs, i.e., AdvAE-GAN model. To make the generated samples more clear and natural, the adversarial learning mechanism and similarity metric (PCE) are added to the AdvAE-GAN model. To obtain the performance of the model in diverse image coloring, multiple methods are elicited for subjective and objective qualitative evaluation and model complexity analysis, respectively. Combining the four standard datasets of AWA, CUB, SUN and FLO, zero-sample image recognition, generalized zero-sample learning experiments are carried out sequentially to derive the loss value curve of the model. The visual effects of animated advertisements generated by AdvAE-GAN model are rated using questionnaire research. For the product effect of animated advertisements generated by AdvAE-GAN model, the category diversity, design diversity, animation contour completeness, and image clarity indexes with scores above 7 account for 70.47%, 85.82%, 76.73%, and 84.02%, respectively. The animated advertisement generation model based on improved generative adversarial network is recognized by the market as well as the society and can be deepened.

Index Terms generative adversarial network, variational self-encoder, similarity metric, image colorization, zero sample image recognition

I. Introduction

Animation is a unique form of artistic expression, which, supported by information technology and digital technology, breaks the low interest of traditional animation and enhances animation creativity and communication presentation. In the further development process of animation industry, the integration of animation and advertising, undoubtedly inject new blood into the traditional advertising development mode [1], [2]. Creative generation of animation advertising products, enhance the visual effect of advertising, broaden the visual space, rich and flexible lens language to present a good model of animation advertising products [3], [4].

The biggest change in the new media era is that traditional hard advertisements are more and more disgusted by people, and personalized, experiential and diversified advertisements are increasingly accepted by the public [5]. The ability to use creativity in animated ad generation determines the brand effect, product function, and emotional experience. Animated advertisements give full play to the artistry, using a variety of forms of expression and powerful penetration of the product characteristics, more conducive to the audience's understanding of the product, animation elements are more able to capture the audience's eyeballs, and through the cartoon image to achieve a good interactive experience [6]-[9]. Compared with traditional advertising, animation advertising breaks the boundaries of the real space, creates a more imaginative and expressive story space, so that the audience is unconsciously immersed in the story, to achieve the purpose of advertising [10]-[12]. In addition, live-action advertising is affected by reality, while animation advertising does not have this problem, and the cost is lower, the effect is better, more competitive advantage [13]-[15]. Under the background of new media, the development of animation advertisement should keep abreast of the times, cater to the aesthetic and development needs of audience groups, enhance the interactivity of animation advertisement by taking advantage of advanced technology, improve people's sense of experience and sensation of animation advertisement, and then push the animation advertisement products to get good progress under the background of the new media era [16]-[20].

In the rapid development of information technology today, animation with its own unique advantages and advertising fusion, in the form of animated advertising in people's field of vision, for the development of the advertising industry has injected a new impetus. Literature [21] investigated and found that the interactive design of animation advertisement in e-sports games can effectively attract the audience's visual attention, which provides a theoretical basis for the placement strategy of animation advertisement in e-sports industry. Literature [22] analyzes the advantages of animated characters in the advertising industry, pointing out that animated characters that gradually replace advertising spokespersons attract more attention to the click rate in advertising communication, which is conducive to advertising communication. Literature [23] elaborated the process of mutual promotion between animation technology and advertising industry, and emphasized the importance of rational communication and emotional appeal balance in animation advertising. Literature [24] shows that animated forms of advertising content can effectively attract audience attention to improve the conversion rate and engagement of advertisements, in addition, animated advertisements play an important role in the advertising industry by promoting audience emotional engagement to improve brand image. Literature [25] compared the effects of animation and static banner ad design in the communication of offensive product advertisements, and found that animated advertisements effectively improved consumers' emotional attitudes toward offensive products, which provided certain insights for advertisers and marketers.

In the new media era, advertising design must be finely crafted thinking, and a large number of scholars have conducted in-depth research on advertising creative design methods for this purpose. Literature [26] found that deep learning technology can provide technical support for the intelligent interaction of computer-aided design models in the virtual reality environment, which improves the intelligent classification and recognition ability of advertisement design and promotes the further development of digital advertisement creative design. Literature [27] establishes a quality evaluation system for advertisement design with the synergy of computer-aided design and reinforcement learning, which can provide advertisement designers with a constant stream of creative inspiration, assist in generating more innovative and personalized advertisement design, and significantly improve the efficiency and quality of digital advertisement design. For animation advertisements, literature [28] introduces the use of ant colony optimization algorithm in the creative design process of animation public service advertisements, which combines Flash, 3DS Max and other technologies to improve the production level of animation advertisements in terms of content and form, and then enhance the overall communication effect of advertisements. Literature [29] developed a multimodal visual communication system (MVCS) model, which recognizes users' emotions and interests through multimodal sentiment analysis technology, dynamically adjusts the video content and playback mode of digital animated advertisements, which is conducive to improving user satisfaction and the communication effect of animated advertisements.

This paper analyzes the development trend of animation advertisement in the new media era and points out the importance of visual language design in animation advertisement. In view of the existing development of generative adversarial network in advertisement image synthesis, and the advantages of variational self-encoder in image and video resolution. In order to update the generation of adversarial samples, a new method of adversarial sample generation is proposed, namely AdvAE-GAN model. AdvAE-GAN model consists of four parts, namely, encoder network, decoder network, target-to-be-attacked network, and discriminator network.Describing the AdvAE-GAN model converts a single-phase process of adversarial sample generation to a two-phase implementation process. Analyze the index performance of AdvAE-GAN model proposed in this paper in terms of diversity image coloring effect, image recognition and image classification, and perform visual scoring of animated advertising images generated by AdvAE-GAN model.

II. Visual language of new media animated advertising

II. A. New Media Animated Advertising Development

Animation advertising is all or part of the use of animation expression in advertising, it is the use of animation technology in modern advertising, is the product of the combination of animation industry and advertising industry. It is a form of advertising that uses animated language to publicize and promote products [30].

In the new media era, the technology of the media, the context and mode of communication, and the audience's reasoning have all changed profoundly compared with the traditional media environment. This is for advertising communication, its ways and means are facing new thinking and choices, at this time to analyze the communication advantages of animation advertising will open up new ideas for advertising creation, and provide animation source for the sustainable development of the advertising industry. The rapid development of CG technology provides a combination of animation and advertising opportunities. With the development of new media technology, animation advertisement, with its own characteristics, is constantly adapting to the communication characteristics of new media while continuously tapping its interactivity. And the new forms of communication presented by new media make animation advertising in the new media era also has extremely powerful communication advantages. In the new media era, the flexible and changeable visual language of animation advertisement, the pervasive communication advantages, and the emotional continuity of respect for the audience are more conducive to timely grasp of the market situation, which is in line with the development needs of the times.

The communication advantages of animation advertisement in the new media era:

(1) Rich visual expression

Animation as a combination of audio-visual language and montage thinking art, and the combination of advertising highlights a very different artistic expression and narrative language with live shooting, creating a special visual experience. The characteristics and advantages of animated advertisements are also beneficial to the expansion of the advertising market. After integrating animation elements, advertisements can break through the barriers of language and culture, and realize cross-border and multi-language communication. In the creation of advertisements, animation expression should be fully considered with the product positioning, brand tone of the degree of coincidence, send sample makes the product in the first visual on the consumer consciousness to establish a unique brand impression different from similar products.

In this era of reading pictures, the more novel the visual language of the advertisement, the more it can attract the attention of the public. The continuous improvement of new media technology makes the animation advertisements innovative in the production method, the strange creative inspiration, the sky is the limit of creative expression through the support of new media technology to create a vivid picture, so that the advertisements present dynamic and flexible three-dimensional modeling, crisscrossing visual space, vivid and detailed realism to create a realm of visual wonders and so on, so that the visual expression of the animation advertisements has become rich and colorful. It makes the audience experience a different sensory impact, reduces the production cost of animation advertisement, and at the same time lays the foundation for animation advertisement to adapt to the development of new media. The use of animation language elements in advertising and dilute the marketing color, increase the fun of advertising, and promote the dissemination of advertising.

(2) Outstanding media adaptability

The new communication methods presented by the media provide a diversified platform for the dissemination of animation advertisements. The integration of media provides a broad communication channel for animation advertisement, so that animation advertisement can interact with network media, combine with mobile new media, and be associated with interactive TV media. It can be seen that animation advertising has outstanding media adaptability.

(3) Emotional continuity of the audience

With the arrival of the new media era, the production methods and dissemination channels of advertisements are changing, and the audience has full autonomy to choose, while the resistance to advertisements is also increasing. Animated advertisements, with their unique visual language forms of expression, greatly enhance the creative thinking space of the advertisement, and satisfy the audience's emotional appeal to the advertisement. Animation ads with a kind of unrestrained creative expression, humorous narrative language, with exaggerated expressions, tension of the action settings, so that the ads with a kind of light visual experience into the audience's field of vision.

(4) Timely control of the market situation

With the intensification of homogenization competition, market segmentation is the best way to communicate effectively with the target audience, and specific media have their inherent user groups. The animation advertisement is based on new media, which has a huge audience base and relatively clear division of audience groups. This provides a clear direction for the positioning of animation advertisement and ensures the communication effect of animation advertisement.

II. B.Visual Language Design in Animated Advertising

The biggest difference between animation visual language and visual language is that visual language includes all the visually visible images, whether they are flat newspapers or three-dimensional buildings, concrete objects or abstract creations, as long as the images give the viewers emotional communication or exchange, then they can all be included in the ranks of visual language. The “animation visual language” must be the content of the “movement” for the purpose of the “painting” form to show the work to the audience to convey feelings and ideas.

II. B. 1) Breaking the limitations of advertising itself

Since advertisement is a movie made by using real actors and real scenes, these real factors will inevitably lead to certain limitations, such as the deformation of the actors, the exaggeration of the movements, the transformation of the scenes, etc. Animation, on the other hand, presents the audience with an unrealistic virtual world in which the artist reorganizes some elements after exaggerating them. Animation, on the other hand, presents the audience with an unrealistic virtual world, in which the artist expresses some of the elements in the world through reorganization after exaggerating and deforming them, so that the audience can have a sense of novelty and empathy for the field of animation. Here the role is equivalent to the real actors in the advertisement, but it is not limited by real people, the shape also has the characteristics of diversity, can be anthropomorphic or anthropomorphic, a chair, a lamp can be molded into the role of the film, where the role of the character molding is good or bad directly affects the degree of success of the film, if the character's characteristics are outstanding, unforgettable, the audience will have a deep memory.

Combining the visual language of animation with advertising, the publicity effect will be significantly improved, on the contrary, the bland character design is difficult to increase the interest of the audience, and the effect of attracting the audience's attention will be greatly weakened.

The use of animation visual language in the lens can be changed at will, eliminating the many inconvenient factors in the reality of shooting. In the scene layout, if you need to change or destructive practices, the use of animation to produce and real scenes compared to reduce many unnecessary troubles. In short, the visual language of animation used in advertising, animation advertising to add color, bring more shocking visual experience for the audience, while bringing considerable revenue for advertising.

II. B. 2) Enhance the effect of real vividness of the picture

All films are the inspiration of the artist in real life, but also a high degree of generalization and distillation of the real world, for the audience to convey a deep emotional truth. Animation advertising is also based on this plus more complex creation. In the process of creation, the plot, characters, actions and other elements of the movie are exaggerated and deformed to the extent that the real movie can't show. Bold exaggeration and strange deformation techniques are also the characteristics and essence of animated films. Adding animation to film and television commercials is a special and novel form of expression, and its visual language also changes with the addition of animation. For example, the plot, lens, color, character, modeling, light and shadow, action and other elements, are very different from the pure film and television advertising film. Greatly enhance the advertising point of view, for people to convey a certain idea or bring entertainment effects, so that the plot of the ads have order, for the audience to bring a real and lively picture visual effects.

III. Visual effect optimization technique for animated advertisement based on improved GANs

III. A. Advertisement image synthesis based on generative adversarial networks

The most widely used area of Generative Adversarial Networks (GANs) in image processing is image synthesis, and the use of this technique can effectively reduce the acquisition cost of animated advertisements [31]. The automatic generation of animated advertisements based on GANs highlights a broad market application prospect in the commercial field to meet the batch production needs of online and offline merchants.

III. A. 1) Network structure

Generative Adversarial Networks, as a generative model based on deep learning techniques, are one of the most promising approaches on fitting complex data distributions in recent years.

The network contains at least two modules: generator model (G) and discriminator model (D). The task of the generator model is to fit the features of the target domain. In the case of the image super-resolution task, the generator generates high-definition images with similar features and resolution of the target image domain. The task of the discriminant model is to discriminate whether the input image is real or not, and in the case of the image super-resolution task, the main purpose is to discriminate whether the input image is a real high-definition image or a false super-resolution image generated by the generator.

In the training optimization process, the generator's goal is to make the generated image fool the discriminator, so that the image generated by itself is mistaken by the discriminator as a real image. The goal of the discriminator is to discriminate the input image as accurately as possible, and the two are trained against each other to optimize the final output by improving each other's abilities in the game.

Generative Adversarial Networks are widely used in the fields of image super-resolution, image restoration, image style migration, image generation, etc. due to the fact that they can learn how to infer and generate images, and can effectively fit some complex data distributions and generate high-quality images, videos, audios, etc., and have achieved good results.

III. A. 2) Confrontation training process

The training process of the generative adversarial network is shown in Fig. 1, by first fixing the generator, i.e., turning off the backpropagation of the generator model so that its parameters are no longer transformed. Only the backpropagation of the discriminator model is turned on, and backpropagation and parameter updates are performed on the discriminator alone. Then fix the discriminator and update the generator, i.e., turn off the backpropagation of the discriminator model and perform backpropagation and parameter update for the generator alone. The discriminator and generator are continuously trained separately using this procedure to make both sides more and more capable.

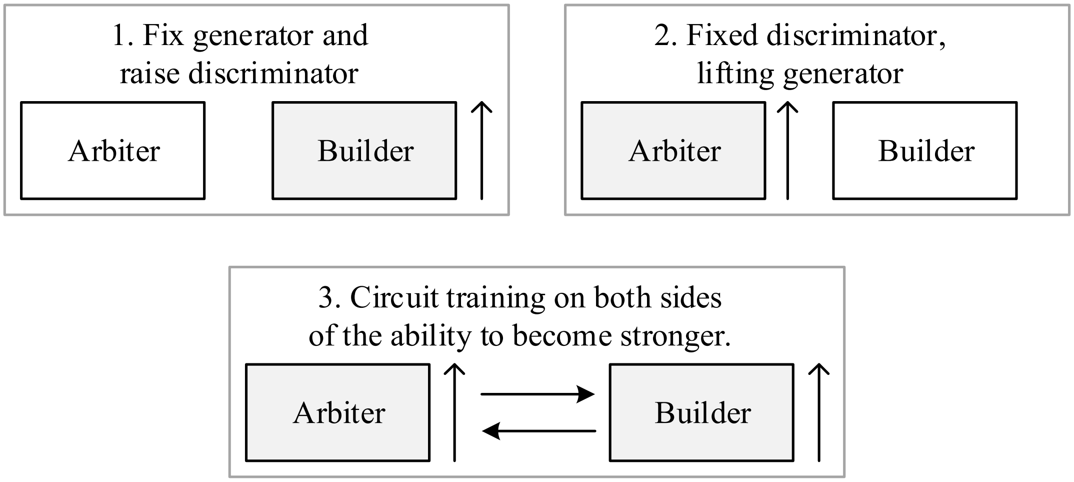


Figure 1: The training process of generating a confrontation network

The optimization objective function for the entire generative adversarial network is as shown in equation (1):

$$\min_{G}\max_{D} E_{x\sim p_{duta}}[\log D(x)]+E_{z\sim p_z}\left[\log(1-D(G(z)))\right] \tag{1}$$

where the optimization objective of the generator is to make the following objective formula drop its stochastic gradient:

$$\nabla_{\theta_g}\frac{1}{m}\sum_{i=1}^{m}\log(1-D(G(z^{(i)}))) \tag{2}$$

where $z^{(i)}$ is the input and m is the batch size.

The optimization objective of the discriminator is to make the following objective equation rise its stochastic gradient:

$$\nabla_{\theta_d}\frac{1}{m}\sum_{i=1}^{m}\left[\log D(x^{(i)})+\log(1-D(G(z^{(i)})))\right] \tag{3}$$

III. B. Variational Self-Encoder Based on Image and Video Super-Resolution Tasks

Variational autocoders (VAEs) are an important class of generative models proposed by improving the hidden layer representation based on traditional autocoders. Combining the features of deep learning and statistical learning, they generate new data by modeling the probability distribution of the data through neural networks, so that the generated data can represent the distribution of the original data as accurately as possible, and are widely used in the field of image generation [32].

In image and video super-resolution tasks, VAEs can generate multiple possible HR reconstructed images for the same LR image by learning the latent probability distribution of the LR input. And in the absence of texture details, their powerful generation capability is utilized to narrow the information gap between the LR image and the HR image, providing flexibility and versatility for super-resolution reconstruction.

III. B. 1) Hidden variable modeling

Assuming that dataset $X=\{x\}$ consists of N independently and identically distributed random variables, theoretically sampling from the distribution $p(x)$ of input data x will yield data similar to the original data, however $p(x)$ is usually unknown. In other words, if there exists a generative model that can produce data from the training set, then samples similar to the training data can be generated by the model.VAE Introducing the hidden variable z, assume that the data in the dataset X are all generated by z undergoing a stochastic process, which can be obtained according to Bayes' theorem:

$$p_\theta(x)=\int p_\theta(x\mid z)p_\theta(z)dz \tag{4}$$

where $p_\theta(x\mid z)$ is called the decoder of the VAE and represents the likelihood function of generating data x from the hidden variables z, which is uniquely determined by parameter θ, and $p_\theta(z)$ represents the probability distribution of z, also called the prior distribution, which is usually assumed to be $p_\theta(z)=N(0,I)$. The goal of the VAE is to train a generative model of the form of the $p_\theta(x)=\int p_\theta(x\mid z)p_\theta(z)dz$, which, by optimizing parameter θ, allows the decoder $p_\theta(x\mid z)$ to maximize the probability of generating samples that are similar to the original data, and thus maximize generative probability $p_\theta(x)$.

III. B. 2) Fundamentals of VAEs

VAE works by mapping the input data to the hidden space through an encoder, and then feeding the sampled hidden variables into a decoder to generate the reconstructed data. For each sample x, if there exists a $p_\theta(z\mid x)$ that obeys a Gaussian distribution, then ideally the hidden variables z sampled by $p_\theta(z\mid x)$ should be decoded to produce a data that is as similar as possible to x. Since the mean and variance of the true posterior $p_\theta(z\mid x)$ are usually difficult to obtain computationally and need to be trained with the help of an encoder $q_\phi(z\mid x)$ that approximates the posterior distribution, the VAE uses an encoder $q_\phi(z\mid x)$ consisting of two neural networks to fit its mean and variance. In fact, VAE constructs a corresponding Gaussian distribution for each input and reconstructs the data by sampling this distribution, and introduces KL scatter to describe the difference between $q_\phi(z\mid x)$ and $p_\theta(z\mid x)$, making $q_\phi(z\mid x)$ as close as possible to the ideal $p_\theta(z\mid x)$ by minimizing the KL scatter. Eq:

$$KL(q_\phi(z\mid x)\,\|\,p_\theta(z\mid x))=E\left[\log q_\phi(z\mid x)\right]-E_{q_\phi(z|x)}\left[\log p_\theta(z\mid x)\right] \tag{5}$$

According to Bayes' formula:

$$p_\theta(z\mid x)=\frac{p_\theta(z)p_\theta(x\mid z)}{p_\theta(x)} \tag{6}$$

Thus equation (5) is further expressed as:

$$\begin{aligned}&KL(q_\phi(z\mid x)\,\|\,p_\theta(z\mid x))\\&=E\left[\log q_\phi(z\mid x)\right]-E_{q_\phi(z|x)}\left[\log\frac{p_\theta(z)p_\theta(x\mid z)}{p_\theta(x)}\right]\\&=E\left[\log q_\phi(z\mid x)\right]-E_{q_\phi(z|x)}\left[\log p_\theta(z)\right]\\&-E_{q_\phi(Z|X)}\left[\log p_\theta(x\mid z)\right]+E_{q_\phi(Z|X)}\left[\log p_\theta(x)\right]\end{aligned} \tag{7}$$

The collation leads to the core formula for VAE:

$$\begin{aligned}&E_{q_\phi(Z|X)}[\log p_\theta(x)]-KL(q_\phi(z\mid x)\,\|\,p_\theta(z\mid x))\\&=E_{q_\phi(Z|X)}[\log p_\theta(x\mid z)]+E_{q_\phi(Z|X)}[\log p_\theta(z)]-E[\log q_\phi(z\mid x)]\\&=E_{q_\phi(Z|X)}[\log p_\theta(x\mid z)]-KL(q_\phi(z\mid x)\,\|\,p_\theta(z))\end{aligned} \tag{8}$$

The goal of the VAE is to maximize the generation probability $p_\theta(x)$ and minimize $KL(q_\phi(z\mid x)\,\|\,p_\theta(z\mid x))$. Since the KL scatter is non-negative, then this goal translates into maximizing Eq. Maximizing the lower bound on the evidence:

$$\mathrm{L}(x;\theta,\phi):=E_{q_\phi(Z|X)}[logp_\theta(x\mid z)]-KL(q_\phi(z\mid x)\,\|\,p_\theta(z)) \tag{9}$$

In this case, maximizing this objective function can be achieved by minimizing $KL(q_\phi(z\mid x)\,\|\,p_\theta(z))$. That is, making $q_\phi(z\mid x)$ as close as possible to the true distribution $p_\theta(z)$, this term is the optimization objective term of the encoder, while the previous term of the objective function represents the decoder part of the VAE.

Theoretically, a z needs to be sampled from $p_\theta(z|x)$ before training the decoder, and then the generated data is used for optimization to continuously improve the generation quality of the model.

Although $p_\theta(z|x)$ is Gaussian distributed, its mean μ and variance σ are calculated by the encoder $q_\phi(z|x)$ and should be sampled from $N(\mu,\sigma^2 I)$ when sampling.

Whereas this sampling operation is a stochastic process and is not a function of the distribution parameters, the sampling operation is not derivable in backpropagation, and training the model cannot be optimized using gradient descent, so the re-parameterization trick is used, whereby a z from $N(\mu,\sigma^2 I)$ is transformed into a ε from $N(0,I)$, and then transformed by the parameter $z=\mu+\varepsilon\times\sigma$ to get the same result as that obtained by sampling from $N(\mu,\sigma^2 I)$. The z constructed in this way allows the gradient to be passed back to the encoder through the mean and variance, and thus the entire VAE model becomes trainable. The next step is simply to combine the encoder and decoder to maximize the objective function through an optimization algorithm, which allows the VAE to be continuously optimized to generate data similar to the original data.

III. C. A New Adversarial Sample Generation Method AdvAE-GAN

III. C. 1) Problem description

The goal of the AdvAE-GAN model in this paper is to generate adversarial samples. Specifically, an encoder is used to compress the original sample x with added noise into smaller feature maps f_x, and a decoder receives the input feature maps f_x and reconstructs them into adversarial samples x_{adv_pre} and adversarial perturbations x_{per}. To further improve the aggressiveness and migration ability of the adversarial samples, new adversarial samples x_{adv} are further constructed by overlaying explicit perturbations x_{per} on the original adversarial samples x_{adv_pre}, mathematically:

$$f_x=\mathrm{Encoder}(x+x_{noise}) \tag{10}$$

$$(x_{adv_pre},x_{per})=\mathrm{Decoder}(f_x) \tag{11}$$

$$x_{adv}=x_{adv_pre}+x_{per} \tag{12}$$

Also, to ensure that the generated adversarial samples can effectively fool the target network, the following conditions must be satisfied:

$$T(x_{adv\_pre})\neq T(x) \tag{13}$$

$$T(x_{adv})\neq T(x) \tag{14}$$

$$sim(x,x_{adv_pre})>\delta \tag{15}$$

where $sim(\cdot,\cdot)$ represents a similarity metric function. In this method, a new similarity metric criterion-Pixel Cross Entropy (PCE), i.e., $sim(x,x_{adv})=-PCE(x,x_{adv})$, is defined to measure the similarity between the original sample x and the generated antagonistic sample x_{adv}, where a larger value of PCE denotes a larger similarity value sim between the two inputs, and a smaller value denotes a smaller value of similarity.The definition of PCE is given by Eq. (16), and δ is the minimum positive similarity threshold value. Eq:

$$\mathrm{PCE}(x,x_{adv})=\sum_{i=1}^{m}\sum_{j=1}^{n}x^{ij}\times[-log\,(x_{adv}^{ij})]+(1-x^{ij})\times[-log\,(1-x_{adv}^{ij})] \tag{16}$$

where x^{ij} denotes the value of a pixel point on the feature map of size $m\times n$. In the experiment, a normalization operation is performed on each input sample before training, i.e., the value of each pixel is converted to the interval [0, 1].

III. C. 2) Modeling

Adversarial samples are generated using AdvAE-GAN, which consists of four components: encoder network (Encoder), decoder network (Decoder), target-to-be-attacked network (T), and discriminator network (Dis).

AdvAE-GAN converts the single-stage adversarial sample generation process into two stages.

In stage 1, noise is added to the original clean samples and fed into the encoder, where the noise has the same number of channels and size as the original clean samples and the distribution of each pixel value obeys a standard

Gaussian distribution. The encoder compresses the input into a feature map f_x, i.e., $f_x = Encoder(x + x_{noise})$. The decoder accepts the feature map f_x and reconstructs it into counter samples x_{adv_pre} and counter perturbations x_{per} corresponding to the original samples x.

In stage II, in order to find a larger domain of adversarial attacks and to improve the aggressiveness and migration of the adversarial samples, x_{per} is cropped to a certain small extent, and the cropped x_{per} is overwritten onto x_{adv_pre} to further construct the adversarial sample x_{adv}, and the cropping operation $Clip(x_{per}, d)$ is defined in (17), where $d \in [0,1]$. i. e.:

$$\mathrm{Clip}(x_{per}, d) = d \cdot \frac{x_{per} - \min(x_{per})}{\max(x_{per}) - \min(x_{per})} \tag{17}$$

In order to ensure that the antagonistic samples generated in the previous stage have low perceptual dissimilarity to the original clean samples, PCE defined in (16) is used as a similarity metric criterion to constrain the size of each pixel value of the generated antagonistic samples. Obviously, as a constraint term, it is desired that PCE is as small as possible. For descriptive convenience, the reconstruction loss term L_{res} defined in Eq. (18) is introduced:

$$L_{res} = PCE(x, x_{adv}) \tag{18}$$

In order to improve the quality of the generation of stage 2 adversarial samples, a discriminator is used to identify whether the input samples are original clean samples or generated adversarial samples. At the end of training, the discriminator is unable to discriminate the input well and the adversarial loss of the discriminator is given by equation (19):

$$T_{GAN} = \varphi E_x logDis(x) + (1-\varphi) E_x log(1 - Dis(x_{adv})) \tag{19}$$

In the expression $Dis(\cdot)$ denotes the discriminator function and φ is a hyperparameter whose value lies in the interval (0, 1). The purpose of the discriminator is to distinguish between the original clean sample x and the generated adversarial sample x_{adv}, thus ensuring that the generated adversarial sample has the advantage of being natural and clear. The final two-stage generated adversarial samples x_{adv_pre} and x_{adv} are fed into the target network T. The attack effectiveness of the adversarial samples is determined by the loss function (20)), where a smaller value of the loss function indicates a better attack effectiveness. Namely:

$$L_{adv} = \alpha E_{x_{adv_{pre}}} L_T(x_{adv\_pre}, t) + \beta E_{x_{adv}} L_T(x_{adv}, t) \tag{20}$$

where $L_T(\cdot;)$ denotes the categorical cross-entropy loss function. $L_T(x_{adv\_pre}, t)$ denotes the cross-entropy loss value between the prediction result of the target network T for x_{adv_pre} and the category t, $L_T(x_{adv}, t)$ and $L_T(x_{adv\_pre}, t)$ are the same, E_r denotes the expectation value, and the two items on the right hand side of expression (20) are expanded in detail as expressions (21), (22):

$$E_{x_{adv_{pre}}} L_T(x_{adv\_pre}, t) = -\frac{1}{n} * \sum_{k=1}^{n} \sum_{y=1}^{c} t[y] * \log(T_y(x_{adv_{pre}}[k])) \tag{21}$$

$$E_{x_{adv}} L_T(x_{adv}, t) = -\frac{1}{n} * \sum_{k=1}^{n} \sum_{y=1}^{c} t[y] * \log(T_y(x_{adv}[k])) \tag{22}$$

where n denotes the number of samples in the training set and c denotes the number of classes of samples. Here, $T_y(x)$ is used to denote the confidence level of the target network to categorize the samples into category y, $x[k]$ denotes the index of the samples as k, and t denotes a one-hot coding vector for the target class. Ultimately, the total loss function of the AdvAE-GAN model is given by expression (23). Namely:

$$L(E, D, Dis) = L_{GAN} + \alpha L_{res} + \beta L_{adv} \tag{23}$$

where α and β are two hyperparameters. At the same time, $\alpha, \beta \in [0,1]$. It is optimal to achieve a certain balance between the reconstruction loss and the counter loss for a certain total loss.

IV. Analysis of the effectiveness of the adversarial sample generation algorithm

IV. A. Diversified image coloring effects

(1) Model Deployment

The computer hardware configuration is as follows: Linux operating system, AMD EPYC7601 processor (64G), NVIDIA GeForce RTX 3080 graphics card (10G). The computer software environment is as follows: Python 3.8.10, Pytorch 10.0, CUDA 11.1.

The dataset used in this paper is from Places365 Standard and contains a total of 365 unique scene classes. For each scene class, there are 5000 images in the training set and 100 images in the test set. In this paper, 10 scene types: “Valley”, “Sky”, etc. are randomly selected for automatic grayscale image coloring experiments. The Adam optimizer is used to update the network parameters, the learning rate is set to 0.0003, the batch size is set to 3, and the iteration period is set to 500.

(2) Comparison method

Pix2Pix+noise is the introduction of noise as a condition to generate diverse results based on Pix2Pix.

MUNIT is an unsupervised translation method for image-to-image diversification. The method learns two main components to achieve translation between images: an inter-domain converter and an intra-domain converter.

BicycleGAN is a classical diverse image translation method. It learns the mapping relationship between potential codes and outputs through two sub-networks to achieve diverse translation between images.

BigColor is a diversified coloring method using a generative prior that draws on the network design of BigGAN as well as improved spatial feature maps to excel on field images with complex structures.

IV. A. 1) Subjective qualitative evaluation

Pix2Pix+noise, MUNIT, BicycleGAN, BigColor, Our approach are selected as comparison algorithms.

In this paper the Mean Opinion Score (MOS) method is used as a subjective evaluation metric as a way to quantitatively assess the perceptual quality of the generated images. Color Reproducibility represents how well the colorized image reproduces the color of the reference image or reference color block. Detail consistency represents the degree of coloring perfection of the image in details, and overall effect represents the coloring effect of the image as a whole.

The coloring performance of each method on the grayscale image is shown in Figure 2. From the figure, it can be seen that the method of this paper outperforms the other methods and achieves the highest score. The AdvAE-GAN model in this paper scores more than 2.8 in the three dimensions of color reproducibility, detail consistency, and overall effect, especially the degree of coloring refinement on the image details reaches 3.43, which is the optimal result for the coloring refinement of image details.

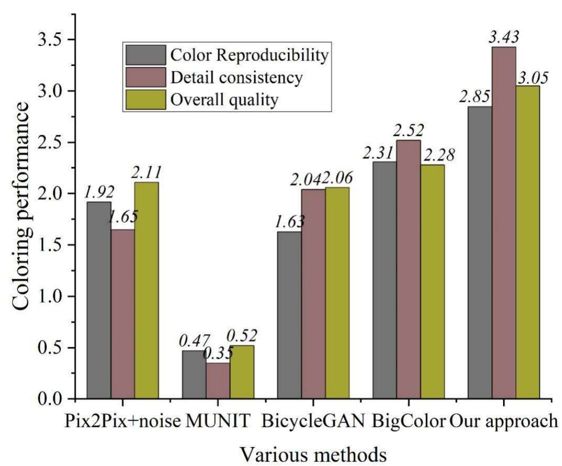


Figure 2: The color of the method in the gray image

IV. A. 2) Objective quantitative evaluation

FID metric and LPIPS metric are chosen to evaluate each method objectively.FID metric is commonly used to measure the quality of generated images.The smaller the value of FID,the better the quality of image generation.LPIPS metric is used as an image quality assessment metric.The larger the LPIPS metric,the better the internal diversity of the translation result,and the worse the opposite is.LPIPS metric is used as a kind of image quality assessment metric.LPIPS metric is used as a kind of image quality assessment metric.

The quantitative results of the different comparison methods are shown in Table 1, from which it can be seen that the method Pix2Pix+noise is the best in terms of FID metrics, which indicates that the statistical properties of the feature space between its generated image and the real image are the most similar. Its LPIPS metric is the worst, indicating that it generates the worst diversity. This is mainly due to the fact that the method does not work for generating diverse color images through noise control alone.

The two metrics of method MUNIT, although indicating that it generates the best diversity, can be seen from the subjective results that the generated content has deviated far from the real image, so the results of this metric are not informative.

Method BicycleGAN achieves a lower FID and higher LPIPS, indicating that it generates images of higher quality as well as a certain degree of diversity.

Method BigColor achieved higher FID and lower LPIPS values, and performs poorly in generating diversity when the categories of small-scale datasets are similar.

The method proposed in this paper achieves good results with FID metrics and LPIPS metrics of 42.3256 and 0.3155+/-0.0028 respectively compared to other comparative methods, keeping the authenticity of the generated images high while achieving better diversity results.

Table 1: The quantitative results of different comparison methods were compared

Method	FID	LPIPS
Pix2Pix+noise	23.5263	0.1354+/-0.0002
MUNIT	96.4112	0.6253+/-0.0018
BicycleGAN	45.0208	0.3296+/-0.0011
BigColor	37.6691	0.2362+/-0.0003
Our approach	42.3256	0.3155+/-0.0028

IV. A. 3) Model complexity analysis

In order to study the model performance in depth, this section analyzes and compares the model complexity and efficiency of various methods. For testing time, all methods are uniformly set to generate 30 different color images corresponding to each test image, as well as retaining the same image post-processing operations and using the same testing machine to ensure a fair and consistent comparison environment. The results of parameter size comparison of different methods are shown in Table 2.

The number of model parameters and the number of floating-point operations per second of this method are 50.12M and 30262.54M, respectively, compared with the MUNIT method, which has the lowest number of model parameters, and the BicycleGAN method, which has the lowest number of floating-point operations per second. In this paper, the method is more advantageous in terms of the quality and diversity of the generated images, and the test time is very similar, both of them have a certain degree of real-time.

Table 2: Comparison of parameters of different methods

Method	Parameters (M)	FLOPs (M)	Test time (s/iter)
Pix2Pix+noise	56.93	18263.64	0.0464
MUNIT	18.42	80316.79	0.0102
BicycleGAN	43.68	15050.42	0.0218
BigColor	479.34	300326.13	0.8629
Our approach	50.12	30262.54	0.0423

IV. B. Zero Sample Image Recognition

(1) Datasets

The proposed method is evaluated on four mainstream datasets, Caltech-UCSD Birds-200-2011 (CUB), SUN Attribute (SUN), Animals with Attributes 1 (AWA1), and Oxford Flowers (FLO).

The AWA1 dataset is a medium-sized coarse-grained multi-animal dataset consisting of 50 classes of animals.

The CUB dataset is a medium-sized fine-grained bird dataset consisting of 200 bird species.

The SUN dataset is a medium-sized fine-grained scene type dataset consisting of more than 700 different scenes.

The FLO dataset is a medium-scale fine-grained flower type dataset consisting of 102 different fine-grained flower classes.

(2) Network parameterization

Similar to previous methods, 2048-dimensional features output from the top-level average pooling unit of ResNet-101 were used as visual features for the baseline dataset. The encoder E and the decoder Dec (generator G) and the semantic decoder SED are two-layer fully connected (FC) networks with 4096 hidden units, and the discriminator D is a FC network with 1024 hidden units. The de-redundancy module M is a fully connected layer with ReLU activation.

The complexity of the model was calculated based on the network structure, and the computational complexity of the baseline method F-VAE-GAN is 10.0 MB, and that of the method in this paper is 12.0 MB.

Based on empirical multiple experiments, the classification weight λ_{cls} was set to 0.01, the cyclic consistency weight λ_R was set to 0.001, the weight parameter λ_m for selecting the de-redundancy loss, the perceptual loss λ_{pd}, and the threshold b were set to be all initialized to 0.01, and the dimension of the non-redundancy feature space f was initialized to 1024. And the parameters of the various parts of the network were optimized using Adam's optimizer (β_1=0.3, β_2=0.9), the number of iteration epochs for this experiment is 800.

(3) Evaluation Metrics

For the ZSL task, the Top1 class average accuracy T1 is used as the metric, i.e., the Top1 recognition accuracy of each unseen class is counted first, and then the class average accuracy T1 is calculated by solving the mean value.Since the test set of the GZSL task consists of visible and unseen classes. The Top1 class average accuracy T1 is first calculated for each class of both, denoted by S and U, respectively. Then, the reconciled mean H is calculated as a comprehensive evaluation metric to assess the performance of the model under the GZSL task.

IV. B. 1) Zero Sample Image Recognition

In order to validate the effectiveness of the proposed method, experiments were conducted on four standard datasets, AWA, CUB, SUN and FLO. And it is compared with 10 more advanced classical algorithms for attribute-based zero-sample recognition: SYNC, F-CLSWGAN, SE-GZSL, LisGAN, Cycle-WGAN, Zero-VAEGAN, LsrGan, DE-VAE, OCD, and f-VAEGAN.

The accuracy results (%) for zero-sample image recognition on the four datasets are shown in Figure 3.

The classification accuracy of this paper outperforms the compared methods on the CUB, SUN, FLO, and AWA1 datasets, with recognition accuracies of 66.89%, 73.24%, 75.69%, and 77.65%, respectively.

Compared to f-VAEGAN with no fine-tuning of the CNN backbone network, the AdvAE-GAN model in this paper outperforms by 4.35%, 7.86%, 7.25%, and 7.46% on the CUB, SUN, FLO, and AWA1 datasets, respectively, with the greatest improvement on the SUN dataset. The experimental results demonstrate the feasibility of the method in this paper on the zero-sample image recognition task. The reason for this may lie in the fact that by adding a de-redundancy module, the interference of redundant information on classification is reduced, and the classification discriminative of pseudo-features is effectively improved.

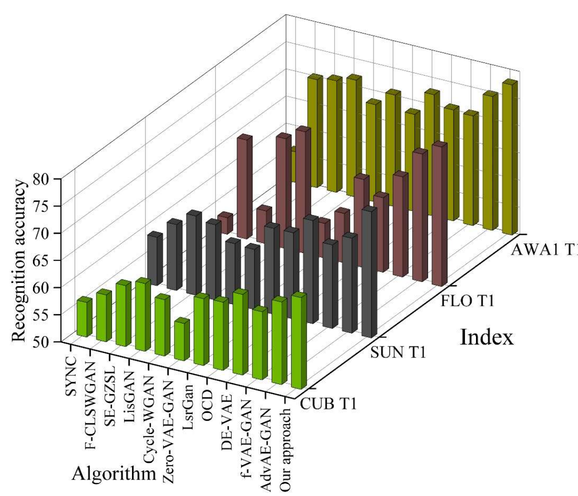


Figure 3: The accuracy of the zero sample image recognition in four data sets

IV. B. 2) Ablation experiments

In this paper, we first add semantic decoders to the VAEGAN framework in order to construct semantic circular consistency loss, and use this model as the baseline of this paper for ablation experiments.

In ZSL and GZSL tasks, two standard datasets, AW1 and FLO, are selected for the experiments, and the results of the ablation experiments are shown in Table 3.

f-VAEGAN+R denotes the baseline network of this paper, baseline network+pd denotes the addition of perceptual distance loss on top of the baseline, baseline network+m denotes the addition of a de-redundancy module on top of the baseline, baseline network+cls+m denotes the addition of a classification regularization term on top of the previous one, and similarly stacks up the corresponding losses sequentially, baseline network+pd+cls+m denotes the proposed model.

Since the AdvAE-GAN model proposed in this paper contains multiple modules, the impact of each module on the performance is analyzed by ablation experiments on the AWA1 and FLO standard datasets in the ZSL and GZSL tasks.

In this paper, the baseline network improves the S, U and H classification accuracies on both AWA1 and FLO datasets compared to the f-VAEGAN method after using the cyclic consistency loss, and it improves more on the FLO dataset, which suggests that it is more sensitive to this loss. Base line+ L_{pd} + L_m improves the T1 and H values by 0.01% and 3.66% on AWA1 dataset compared to the baseline.

It can be obtained through several combination experiments that the best recognition accuracy is obtained by introducing the loss of de-redundant modules, the loss of perceptual distance regularization and classification regularization terms simultaneously into the baseline method, this is due to the fact that the AdvAE-GAN model in this paper combines the advantages of all the three, so the recognition accuracy is higher.

Table 3: Ablation experiment results

Method	AWA1				FLO			
	ZSL		GZSL		T1		GZSL	
	T1	S	U	H	T1	S	U	H
f-VAE-GAN	72.35	68.89	59.62	65.12	68.75	74.69	57.05	66.51
Base line(f-VAE-GAN+ L_R)	70.21	70.43	61.57	65.79	68.23	81.42	60.78	69.65
Base line+ L_{pd}	72.36	71.21	63.45	68.09	70.42	80.72	62.34	70.96
Base line+ L_m	72.22	76.26	60.23	68.79	71.25	84.65	62.98	71.24
Base line+ L_{cls} + L_m	72.36	74.58	62.84	68.78	72.15	85.62	62.38	72.06
Base line+ L_{pd} + L_m	74.15	72.63	63.85	68.98	72.68	79.62	66.96	72.22
Base line+ L_{pd} + L_{cls} + L_m	75.62	76.58	65.96	70.35	73.97	80.24	68.39	73.48

IV. C. Zero sample image classification task

IV. C. 1) Experimental results of generalized zero-sample learning

The division of the dataset for this experiment was performed according to the division of the generalized zero-sample classification task, following the default division of the dataset into visible and unseen classes.

AwA dataset 40 classes are used for training and validation, 10 classes are used for testing the accuracy obtained corresponds to the unseen class image classification accuracy acc_U. And some images of 13 classes are selected from 40 classes of visible classes for validation their accuracy corresponds to the visible class image classification accuracy acc_S. acc_H is the reconciled average accuracy of both.

The CUB dataset is validated by selecting partial images of 50 classes from 150 visible classes.

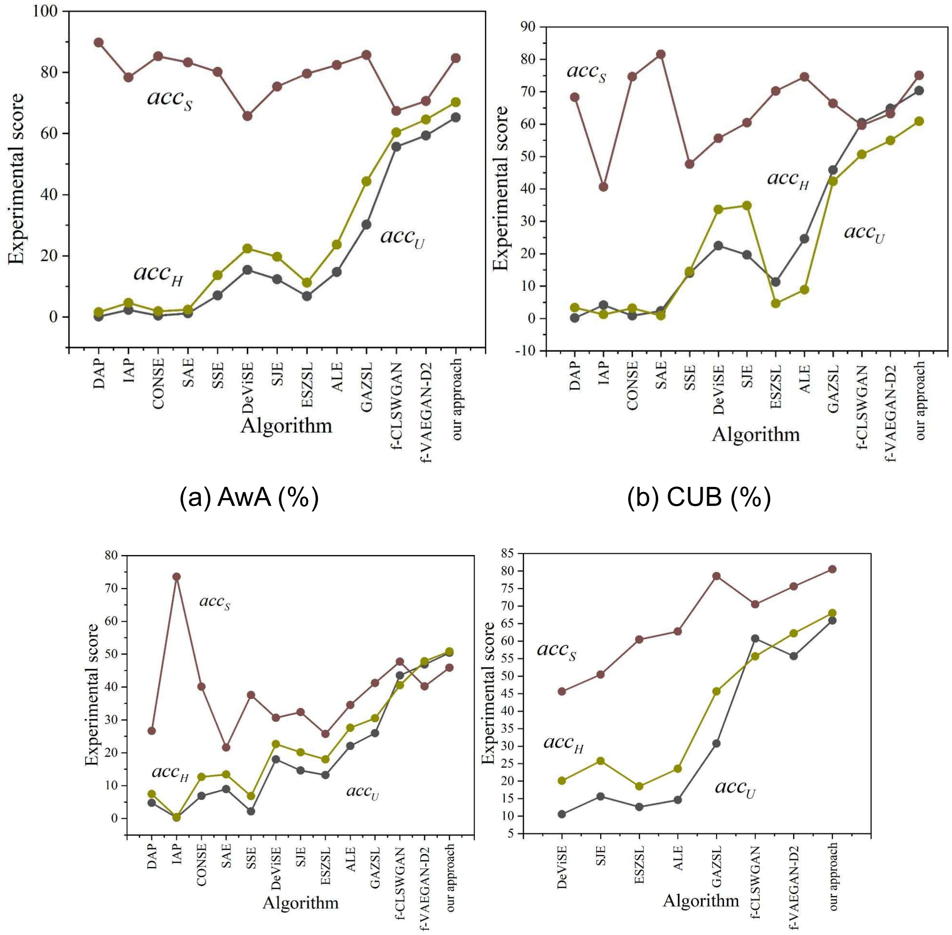
The FLO dataset is validated by selecting some images of 20 classes from 82 visible classes.

The SUN dataset, on the other hand, is validated by selecting partial images of 65 classes from 645 visible categories.

The experimental results of applying this paper's AdvAE-GAN model to the above four datasets for generalized zero-sample learning are shown in Fig. 4, and Figs. (a) to (d) show the experimental results of the generalized zero-sample learning of this paper's algorithmic model on the AwA dataset, CUB dataset, SUN dataset, and FLO dataset, respectively.

From the figure, it can be seen that the unseen class image classification accuracy acc_U of this paper's model in generalized zero-sample learning is higher than that of the existing zero-sample learning model, which indicates that the generative performance of this paper's model is useful for this task.

The traditional attribute-based approach has better performance in the visible class image classification accuracy acc_S, and in terms of the reconciled average accuracy acc_H, the model in this paper improves its performance in all four datasets compared with the existing model.The AwA dataset, the CUB dataset, the SUN dataset, and the FLO dataset improve their performance by 5.65%, 5.91%, 2.98%, and 5.78%, respectively. . The results of the comparison experiments show that the model in this paper also has some feasibility and superiority in generalized zero-sample learning.



(a) AwA (%)

(b) CUB (%)

(c) SUN (%)

(d) FLO (%)

Figure 4: The results of the generalized zero-sample study in this paper

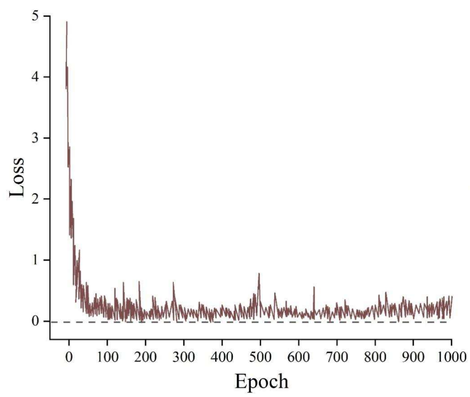
IV. C. 2) Model analysis and visualization

In this chapter, the model performance will be discussed based on the training results of the model, and the quality of the generated images.

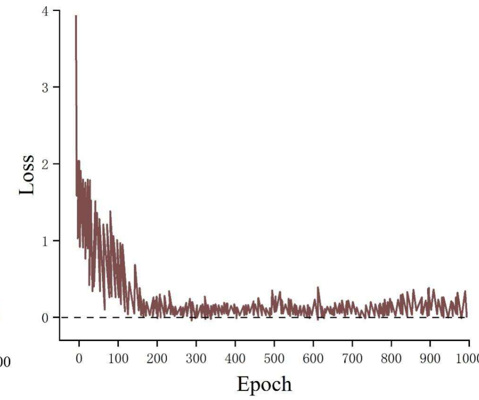
The training results of the model based on Aw A and CUB datasets are shown in Fig. 5, Figs. (a) and (b) show the training results based on Aw A dataset, and the training results based on CUB dataset, respectively.

It can be seen from the figure that the loss value of the model decreases with the increase of the number of training times and finally reaches a stabilized state. The loss value of the training based on the CUB dataset, although there are some fluctuations in the training to about 100 cycles, but with the growth of the training cycle finally shows a steady decrease trend, and converges to a stable state in the training to 200 cycles. Therefore, it can be visualized from the figure that the AdvAE-GAN model in this paper has good convergence.

The image features generated by the AdvAE-GAN model in this paper are visualized and saved to get the corresponding images. The generative model AdvAE-GAN in this paper pays more attention to the details of the target's color as well as contour compared to the original model, such as generating images with colors and shapes that are closer to the real image. As a result, the generated samples have strong authenticity and match their corresponding classification characteristics. Although some images also have some details missing, from the perspective of the overall image generation quality, the model in this paper has a better performance than other models in sample generation, which in turn enables the classification network to learn more effective information to obtain a better zero-sample classification model.



(a) Training results based on the AwA data set



(b) Training results based on CUB data sets

Figure 5: Model training results based on AwA and CUB data sets

V. AdvAE-GAN Generation for Visual Quality Evaluation of Animated Advertisements

In order to subjectively evaluate the image quality generated by the AdvAE-GAN model in the paper, and obtain the human visual intuition. In this chapter, the expert scoring method is adopted, and a questionnaire survey is conducted for the group with the education of postgraduate students and above, and a total of 100 questionnaires on animated advertisement generation with improved generative adversarial network and variational self-encoder are issued to the researchers, and 98 valid answers are received. Among the respondents, 50 of them, accounting for about 51%, were researching in the direction of advertising visual design. Art and design 30, accounting for about 31%. Animation production category 10, accounting for 10%, and computer category 8, accounting for about 8%.

V. A. Questionnaire setting

The questionnaire research is divided into two parts.

The first part scores the effect of AdvAE-GAN-based image synthesis of animated advertisement products, with 3 questions, and 4 image evaluation indexes for each question, namely, category diversity, design diversity, animation outline completeness, and image clarity.

The second part scores the effect of animated advertisement layout based on AdvAE-GAN, with 4 questions and 3 evaluation indexes, namely, layout reasonableness, layout aesthetics, and typographic neatness. Each topic has two images for control, and the evaluator is required to give the layout effect score by observing the images. Image a is the animated advertisement layout image generated by AdvAE-GAN model, which contains several elements such as image, text, background, title, etc., which are represented by different color layout components. Image b is the advertisement effect image demonstrated according to the layout, which is used to visualize the layout and facilitate the evaluator to score.

V. B. Data analysis

V. B. 1) Evaluation of visual quality of animated advertising products

Through the research, the importance of the evaluation indicators of animation advertising product image synthesis is divided into seven levels, from low to high for the sequential order of, extremely unimportant, unimportant, slightly unimportant, general, slightly important, important, extremely important, respectively, with a score of 1-7 instead.

The evaluation of the importance of product image synthesis indicators is shown in Table 4. From the observations in the table, it can be seen that the importance of the four indicators of category diversity, design diversity, animation contour completeness, and image clarity is located in the range of 0.82-0.94, which indicates that the experimentally selected indicators are representative of the evaluation of experimental effects. Among them, the importance of image clarity is the highest, 0.94, and the importance of category diversity is the lowest, 0.82, indicating that the clarity of image synthesis in animated advertising products is more important in comparison. While comparing the animated advertisement design, the importance of category index is slightly lower, indicating that consumers are more concerned about the diversity of animated advertisement design presented in the advertisement.

Table 4: Evaluation of product image synthesis index

Index	Rating scores (%)							Importance
	0	1	2	3	4	5	6	
Class diversity	0.00	0.00	3.35	10.63	22.36	35.68	27.98	0.82
Design diversity	0.00	0.00	6.81	3.52	21.52	53.07	15.08	0.86
Animation outline completeness	0.00	0.00	4.50	8.76	15.59	31.24	39.91	0.88
Image resolution	0.00	0.00	0.00	4.21	17.17	24.96	53.66	0.94

The effect scores of the animated advertisement products synthesized by AdvAE-GAN model are shown in Table 5, and the mean scores of the four evaluation indexes, in descending order, are: design diversity (8.11), image clarity (7.87), contour completeness (7.63), and category diversity (7.58).

Category diversity, design diversity, animation contour completeness, and image clarity scored above 7 (i.e., rated as “good” and “very good”) in 70.47%, 85.82%, 76.73%, and 84.02%, respectively. The favorable rate is obviously higher.

Table 5: AdvAE-GAN model synthesis animation advertising effect evaluation

Index	Grading (%)					Average score
	1-2	3-4	5-6	7-8	9-10	
Class diversity	0.00	10.52	19.01	60.34	10.13	7.58
Design diversity	0.00	3.66	10.52	50.68	35.14	8.11
Animation outline completeness	0.00	8.50	14.77	62.43	14.30	7.63
Image resolution	0.00	2.69	13.29	67.21	16.81	7.87

V. B. 2) Evaluation of visual layout effect of animated advertisement

The questionnaire was set up with four groups of animated advertisements, and the evaluator scored each group of advertisements according to their visual effects at three levels: layout rationality, layout aesthetics, and typographic neatness. The scoring level is the same as that of the first part of the questionnaire, and the score of 1-2 is rated as very bad layout effect, 2-4 is bad, 5-6 is average, 7-8 is good, and the score of 9-10 is very good effect. The breakdown of each group's rating is shown in Table 6.

The percentages of the above four groups of animated advertisements rated as “good” and “very good” in the layout rationality index are 81.95%, 81.20%, 78.68% and 71.91% in turn. The above values fluctuate around 78.44%, with the highest rate of 81.95%, indicating that the layout of the animated ads generated by the AdvAE-GAN model is excellent. 4 groups of animated ads fluctuate in the overall values of layout aesthetics and neatness of layout to 68.33% and 82.51%, among which the most reasonable layout and the highest degree of neatness of the layout is the first group of images, with a mean value of 8.0% and a mean value of 8.0% for the layout reasonableness. The average value of the layout aesthetics score is 8.12 and the average value of the layout neatness is 8.67.

Table 6: The scores of each group were detailed

Group	Index	Grading (%)					Mean
		1-2	3-4	5-6	7-8	9-10	
First set	Layout rationality	0.00	0.00	18.05	53.14	28.81	8.12
	Layout aesthetics	0.00	0.00	24.69	45.87	29.44	7.83
	Layout uniformity	0.00	0.00	13.96	56.37	29.67	8.67
Second group	Layout rationality	0.00	3.56	15.24	52.63	28.57	8.07
	Layout aesthetics	0.00	6.78	33.52	36.24	23.46	6.76
	Layout uniformity	0.00	2.96	12.75	50.11	34.18	8.42
Third group	Layout rationality	0.00	8.57	12.75	56.89	21.79	7.86
	Layout aesthetics	0.00	9.16	23.75	42.75	24.34	7.11
	Layout uniformity	0.00	3.15	20.97	42.36	33.52	7.71
Fourth group	Layout rationality	0.00	3.57	24.52	42.57	29.34	7.43
	Layout aesthetics	0.00	1.15	27.63	49.16	22.06	7.38
	Layout uniformity	0.00	3.64	12.52	56.78	27.06	8.35

VI. Conclusion

In this paper, we take advantage of the algorithmic advantage of generative adversarial network to synthesize advertisement images, and propose a new adversarial sample generation method to update the visual generation effect of animated advertisements and optimize the visual expression of animated advertisements by improving the generative adversarial network.

(1) The proposed new generation method of adversarial samples, i.e., AdvAE-GAN model, is respectively analyzed for diversified image coloring effect, zero-sample image recognition analysis and zero-sample image classification accuracy determination.

The algorithm model AdvAE-GAN in this paper is able to achieve a better level of color reproducibility, detail consistency and overall effect in three dimensions, with scores of 2.85, 3.43 and 3.05 in the three dimensions, respectively, and the FID index and the LPIPS index are both at the middle level, which is capable of balancing the authenticity of the image and the diversity.

Compared with other classical algorithms for zero-sample recognition based on attributes, the classification accuracy of this paper's method model AdvAE-GAN outperforms the comparative methods on CUB, SUN, FLO and AWA1 datasets, and the recognition accuracy stays above and below 70%.

The training results of AdvAE-GAN model on AwA and CUB datasets show that the AdvAE-GAN model designed in this paper has good convergence. And the image generated by AdvAE-GAN model is more realistic compared with the image generated by the original model, which preserves the details of the target image such as color and contour, and is closer to the target image.

(2) The visual quality score of the animated advertisement product and the visual layout effect score of the animated advertisement have achieved excellent results, which proves that the AdvAE-GAN model, which combines the improved generative adversarial network and the variational self-encoder designed in this paper, can be further utilized in the production of animated advertisements.

Funding

This work was supported by Textbook Construction Project of Anhui Provincial Quality Engineering (Fund No.: 2023jcjs170); Support Program for Outstanding Young Talents in Anhui Provincial Universities (Fund No.: gxyq2019143); New Major Quality Improvement Project of Anhui Provincial Quality Engineering (Fund No.: 2023xjzlts099).

References

[1] Li, F., & Wang, Z. (2021, July). The Value and Significance of Digital Media Interactive Technology in Animation Design. In 2021 International Conference on Education, Information Management and Service Science (EIMSS) (pp. 13-16). IEEE.

[2] Kathuria, A. (2022). THE USE OF ANIMATION IN ADVERTISING AND MARKETING. NeuroQuantology, 20(5), 5345.

[3] Li, T., & Zhang, X. (2022). Research on the characteristics of the animation industry and the innovation of marketing and communication strategies in the data era. Academic Journal of Business & Management, 4(1), 50-54.

[4] Kurniawan, E. (2022). Animation, New Habit Solution In Creative TV Ads Content to Survive from Endemic Covid-19. Journal of Art, Film, Television, Animation, Games and Technology, 1(1), 1-11.

[5] Tan, L. O., Balaya, S. A. P., & Aaron Goh, P. H. (2024). Creative struggle in Malaysia: a case study of the advertising, film, and animation industry sectors. Creative Industries Journal, 17(1), 75-87.

[6] Nazley, N. M., Effendy, I. Z. M. I., Jayakumar, M., & Herrow, M. F. M. (2024). Caught In the Net: Narrative Style 2d Animation on Advertising Awareness Campaign on The Effects of Social Media Towards Adolescent Development. Multidisciplinary Applied Research and Innovation, 5(3), 86-93.

[7] Ijaz, N., & Saddiqa, A. (2022). Animated Advertisement in Pakistan: Its Impact as an Emerging Medium in Pakistani Media. Multicultural Education, 8(7).

[8] Wu, R. (2021, November). Innovation and application of animation art design based on digital media background. In 2021 2nd International Conference on Information Science and Education (ICISE-IE) (pp. 1530-1534). IEEE.

[9] Hushain, J., Nihalani, M., Sharma, D., & Kant, K. (2023). The Advantage of Animated Advertisements in Today's Era. International Journal of Innovative Science and Research Technology, 8(1), 431-436.

[10] Trisemarawima, I. N. Y., Prayoga, I. K. M., Jayanegara, I. N., & Yasa, I. W. A. P. (2024). INFORMATION MEDIA PUBLIC SERVICE ADVERTISEMENT ABOUT VCT BASED ON 2D ANIMATION. Journal of Aesthetics, Design, and Art Management, 4(1), 20-29.

[11] Wang, X., & Zhong, W. (2024). Evolution and innovations in animation: A comprehensive review and future directions. Concurrency and Computation: Practice and Experience, 36(2), e7904.

[12] Liu, X. (2022). Animation special effects production method and art color research based on visual communication design. Scientific Programming, 2022(1), 7835917.

[13] Vijayakrishnan, V., Harikrishnan, D., Babu, D., & Kochi, A. V. V. (2018). Marketing strategy in advertisements using animated characters. International Journal of Pure and Applied Mathematics, 119(12), 2841-2852.

[14] Solanki Sandip, P., & Sheth Bhagyashree, H. (2016). Effectiveness of Advertisements: A Study on Comparative Analysis of Celebrity-endorsed Advertisements versus Animated-character-endorsed Advertisements for Children. Rupkatha Journal on Interdisciplinary Studies Humanities, 8(1).

[15] Shuja, K., Ali, M., Anjum, M. M., & Rahim, A. (2018). Effectiveness of animated spokes-character in advertising targeted to kids. Shuja, K., Ali, M., Mehak Anjum, M., & Rahim, A.(2018). Effectiveness of Animated Spokes Character in Advertising Targeted to Kids. Journal of Marketing Management and Consumer Behavior, 2(2), 31-47.
[16] Hong, W., Cheung, M. Y., & Thong, J. Y. (2021). The impact of animated banner ads on online consumers: A feature-level analysis using eye tracking. Journal of the Association for Information Systems, 22(1).
[17] Yang, X. (2022, April). Application of digital media technology in animation design. In International Conference on Multi-Modal Information Analytics (pp. 947-951). Cham: Springer International Publishing.
[18] Chen, Y., & Yang, X. (2023). The Application of Digital Animation in Advertising Creativity in the New Media Environment. Probe-Media and Communication Studies, 5(2).
[19] Yuanliang, W., & Zhe, Z. (2024). Integration effect of artificial intelligence and traditional animation creation technology. Journal of Intelligent Systems, 33(1), 20230305.
[20] Peng, M. (2024). The application of digital media technology in the post-production of film and television animation. Media and Communication Research, 5(2), 129-134.
[21] Seo, Y. N., Kim, M., Lee, D., & Jung, Y. (2018). Attention to eSports advertisement: effects of ad animation and in-game dynamics on viewers' visual attention. Behaviour & Information Technology, 37(12), 1194-1202.
[22] Goel, D., & Upadhyay, R. (2017). Effectiveness of use of animation in advertising: a literature review. International Journal of Scientific Research in Network Security and Communication, 5(3), 146-159.
[23] Cook, M., & Thompson, K. M. (2019). Introduction to animation and advertising. Animation and Advertising, 1-51.
[24] Naik, M., Suryawanshi, Y., & Atre, A. (2023). UNLEASHING THE POWER OF ANIMATION IN MARKETING: INSIGHTS AND IMPLICATIONS. The Online Journal of Distance Education and e-Learning, 11(2).
[25] Rakrachakarn, P., & Moschis, G. P. (2015). Is animation effective for advertising offensive products? Insights from an internet advertising experiment. International Journal of Electronic Marketing and Retailing, 6(4), 259-276.
[26] Zhang, B. (2025). Development and implementation of a deep learning-based smart evaluation system for animated films. Journal of Computational Methods in Sciences and Engineering, 14727978251314558.
[27] Liang, Y., Deng, R., Lin, W., Deng, R., Zhu, X., & Yu, L. (2025). Modeling and Reinforcement Learning Assessment System for Quality Improvement of Advertising Design. Computer-Aided Design & Applications, 21, 188-200.
[28] Liu, H. (2023). Creative design and technical analysis of animation public service advertisement based on ant Colony optimization algorithm. International Journal of Digital Multimedia Broadcasting, 2023(1), 5059665.
[29] Fang, J., & Gong, X. (2023). Application of visual communication in digital animation advertising design using convolutional neural networks and big data. Peerj Computer Science, 9, e1383.
[30] Fang Jingyi & Gong Xiang. (2023). Application of visual communication in digital animation advertising design using convolutional neural networks and big data..PeerJ. Computer sciencee1383-e1383.
[31] William Villegas Ch,Alexandra Maldonado Navarro & Araceli Mera Navarrete. (2025). Using Generative Adversarial Networks for the synthesis of emotional facial expressions in virtual educational environments. Intelligent Systems with Applications200479-200479.
[32] Hiroki Kawabe,Kentaro Yaji & Yuichiro Aoki. (2025). Data-driven multifidelity topology design with multi-channel variational auto-encoder for concurrent optimization of multiple design variable fields. Computer Methods in Applied Mechanics and Engineering117772-117772.