International Journal for Housing Science and Its Applications
Publish August 3, 2025. Volume 46, Issue 3 Pages 467-477

https://doi.org/10.70517/ijhsa46333

Construction of Visualized Learning Resources for Secondary Mathematics Education Based on Deep Convolutional Neural Networks

Lei Zhang[1,*]
[1] School of Mathematics and Statistics, Hanshan Normal University, Chaozhou, Guangdong, 521041, China
Corresponding authors: (e-mail: zhang7388682@163.com).

Abstract Images play an important role in information communication and information preservation, which can meet the needs of the construction of visual learning resources in the education industry. This paper constructs a deep convolutional generative adversarial network model for text-generated images based on conditional enhancement and attention mechanism, which adopts a bidirectional long and short-term memory network to extract textual features, and enriches the feature information of the text through the conditional enhancement module. Subsequently, textual and visual features are fused and output in the generative adversarial network, and the detailed adjustment of the output features is accomplished based on the attention mechanism to generate educational visual image resources containing important features of textual descriptions. The text-to-generate-image method proposed in this paper obtained excellent IS scores on both CIFAR10 and CelebA datasets, with their mean values higher than 7. The visual learning resources, designed in conjunction with the content of the mathematics curriculum, help to enhance the multifaceted mathematical literacy and learning effectiveness of secondary school students, and have gained the approval of the majority of students, with an average satisfaction score of around 4 on the student questionnaire. The text-image generation method in this paper provides new ideas for the construction of visual learning resources for mathematics, which helps learners to better understand mathematics learning methods and concepts.

Index Terms Attention Mechanism, Deep Convolutional Generative Adversarial Network, Bidirectional Long and Short-term Memory Network, Visualized Learning Resource Construction

I. Introduction

The working points of the Ministry of Education in 2021 pointed out that the goal and task of actively promoting the construction of education informatization is to accelerate the high-quality development of education informatization and deepen the popularization action of network learning space application to comprehensively improve the information literacy of teachers and students [1]. The main points of the Ministry of Education in 2022 pointed out that in order to implement the strategy of education digitization, it is necessary to accelerate the digital transformation and intelligent upgrading of education by taking the demand as the guide, deepening the integration, constantly innovating, and taking the application as the driver [2]. The working points of the Ministry of Education in 2023 pointed out that it is necessary to accelerate the construction of new infrastructure for education, innovate the mode of supplying digital resources, promote the pilot demonstration of new modes in new fields of education informatization, and deepen the integration and innovation of information technology and education teaching [3]. The above policy indicates that the application of information technology to solve subject problems and improve teaching strategies has become an important and realistic topic. With the development of the times, the integration of information technology in teaching and exploring the development of a new type of "teaching" and "learning" has become a demand for the current development of education [4]. Obviously, the traditional teaching methods can no longer meet the current needs of information technology teaching. Therefore, following the development of informationized teaching has become an inevitable choice in education reform.

Visualization teaching is a teaching method that uses visual tools such as images, charts, animations, simulations and multimedia to help students better understand and learn knowledge [5]. Mathematics visualization teaching means that in the process of mathematics learning the originally abstract mathematical objects are presented to students using appropriate visual forms, which in turn helps students better understand mathematical concepts and master mathematical principles [6], [7]. Literature [8] suggests that visual learning resources can help students understand correlations in certain arithmetic functions in the classroom, and investigates the actual usability of visualization software based on students' feedback. Literature [9] proposes an innovative approach to teaching mathematics, which improves students' spatial reasoning and problem solving skills in multivariate environments by incorporating spatial visualization tools.GeoGbrae integrates mathematical tools such as geometry, algebra, calculus, and statistics, and provides intuitive interfaces and dynamic demonstration functions.Literature [10] relies on the GeoGbrae platform for teaching mathematics visualization that demonstrates the effectiveness of visual learning in mathematics through the case of quadrilateral properties. Similarly, literature [11] incorporates GeoGebra dynamic software into the mathematics classroom, leading to a shift in traditional mathematics education and highlighting visualization as a key competency in the mathematical process. Literature [12] considers teaching mathematics visualization as choosing appropriate visualization tools to transform abstraction into concrete and static into dynamic, further delineating visualization tools such as mind maps and mathematical software. From the existing studies, there are some limitations in defining the teaching of mathematics visualization, and these definitions are relatively monotonous and theoretical, which cannot fully reflect the real meaning of mathematics visualization in the discipline of mathematics.

This paper explores the application of a GAN-based text-generated image method in the construction of learning resources for mathematics visualization. The text-generated image generation method in this paper extracts text features through BiLSTM network and uses conditional enhancement to enrich the semantics of text features. In the generator, the text features and noise vectors are fused, the inputs are up-sampled to a block of residuals, an attention mechanism is introduced to adjust the high-resolution image features, and the convolutional layer outputs the generated image. The discriminator, on the other hand, performs the feature extraction of the image through a convolutional neural network, and reduces the resolution of the image features by downsampling the residual block. An adversarial loss network is designed to optimize the model. Comparative experiments and examples are used to verify the effectiveness of the model and the feasibility of its application in the construction of learning resources for mathematics visualization, respectively.

II. Construction of Visual Learning Resources for Secondary Mathematics Education

The images generated by Generative Adversarial Networks can be well integrated into the development mode of mathematics visualization learning resources, and the images generated by Generative Adversarial Networks can be used as supplementary materials to complement each other with the traditional mathematics textbooks and the image resources in online courses to form a complete and multi-level mathematics visualization learning resource base.

II. A.Generating Adversarial Networks

Generative Adversarial Network [13] (GAN) is the most applied and recognized method with the best results for deep learning in image generation. It has no restriction on the network structure, and there is no restriction on the dimensionality of the generated data, and it can generate data with high dimensionality, which is very friendly to the learning and generation of high-dimensional data. Figure 1 shows the network structure of GAN.

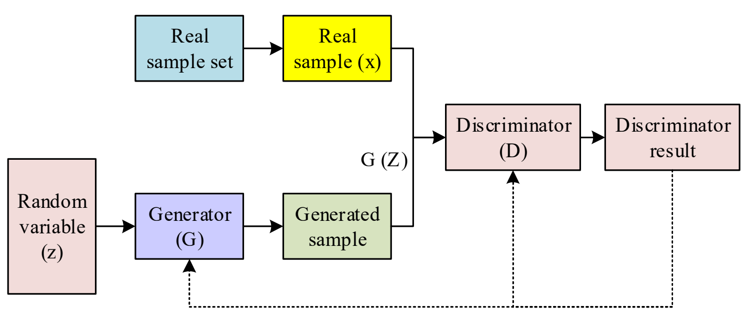

Figure 1: Network structure of GAN

GAN can be formally represented by the following formula for the optimization process of the generator and discriminator in the learning and training process, which is essentially a very large and very small two-player game process, and the following formula can represent its objective function:

$$\min_G \max_D V(D,G) = E_{x\sim P_{data(x)}}\left[\log D(x)\right] + E_{Z\sim P_Z(x)}\left[\log\left(1-D\left(G(z)\right)\right)\right] \quad (1)$$

In the formula, $P_{data(x)}$ represents the data distribution of the real sample, $P_{Z(x)}$ represents the distribution of the generated data, the generator is constantly learning, the purpose of training is to make the objective function minimized, the discriminator is constantly trained to maximize the objective function, the two are constantly adjusted to optimize, and ultimately the generator generates samples similar to the real data. Since the training process is a process of maximizing and minimizing, and the two are contradictory, it is difficult to train together. So the model adopts the way of fixing one of the modules and training the other. For example, when training the discriminator, the generator needs to be fixed, and at this time the objective function is shown in the following equation.

$$\max_D V\left(D,G\right)=E_{x\sim P_{data(x)}}\left[\log D(x)\right] + E_{Z\sim P_Z(x)}\left[\log\left(1-D\left(G\left(z\right)\right)\right)\right] \tag{2}$$

When training the discriminator, it is desired that the distribution of the real data is as large as possible, while the generated data is as small as possible, and the purpose of the training is to maximize $V\left(D,G\right)$, so the second term of the formula uses $1-D(G(z))$, which is the optimization objective of the discriminator during training.

And the generator fixes the discriminator during training. The objective of training is to minimize the function with the aim of making the output of the generated data after the discriminator as large as possible, close to 1. The objective function for training the generator is shown in the following equation.

$$\min_G V\left(D,G\right)=E_{Z\sim P_{Z(x)}}\left[\log\left(1-D\left(G\left(z\right)\right)\right)\right] \tag{3}$$

Through the formula can be seen throughout the GAN training process, the real data and generated data sent to the discriminator, after the discrimination of the discriminator, so that the output objective function is maximized, and the generator to random noise as input, and ultimately generate samples close to the real data, but the generator's objective function is determined by the discriminator, according to the feedback of the input of the discriminator to adjust the update of its own parameters.

Generative Adversarial Networks play an important role in the field of image processing by virtue of their good image feature extraction capabilities, greatly promoting the development of image processing technology and solving many problems related to images in the industry. But Generative Adversarial Network GAN still has some problems. First is the convergence problem, before reaching equilibrium, the model will oscillate, resulting in non-convergence. Secondly, there is the problem of model collapse, which occurs when the samples lack diversity.

II. B.Deep Convolutional Generative Adversarial Networks

In order to solve these problems and improve the model performance, Deep Convolutional Generative Adversarial Networks DCGAN [14] came into the limelight.DCGAN is a direct extension of GAN by introducing Convolutional Neural Networks (CNNs) in the generator and the discriminator, where the discriminator uses convolutional layers to recognize the features of the image, and the generator uses transposed convolutional layers to reconstruct the image and thus generate the image.

DCGAN, as an unsupervised training model, adds the idea of convolutional operation to the network model, and the convolutional network is used as the main structure to extract the features of the training set, which is an end-to-end learning model, and the discriminator of DCGAN is a typical CNN classifier, but unlike traditional convolutional neural networks, the discriminator does not use the fully-connected layer, and cancels the pooling layer, so that the feature blurring caused by the pooling layer is avoided and the feature information is guaranteed. This can avoid the feature blurring caused by the pooling layer and ensure the clarity of feature information. As for the generator, the convolutional layer is replaced by a transposed convolutional layer for image generation. In addition, batch normalization (BN) is used in both the generator and the discriminator to ensure that the model can correctly learn the mean and variance of the data during training.

II. B. 1) Convolutional layers

The convolutional layer is the core of a neural network, and convolution is the process of doing an inner product operation between a window of data and a convolutional kernel. The convolutional layer contains multiple convolutional kernels inside, the convolutional kernel, also known as a filter, is a neuron with weights. A convolution kernel is essentially a two-dimensional matrix, and during training, the input is scanned regularly to extract data features, such as color, contour, etc. of the image. It is worth mentioning that different convolution kernels can perform different tasks such as filtering, image sharpening etc. During the convolution process, in order to avoid losing information when scanning to the edge of the image, the matrix is expanded by padding the edges of the original image before the convolution operation, and the value of the padding is usually taken as 0. In addition, the distance that the convolution window slides on the image each time is called the step size, and multiple convolution kernels stacked on top of each other make up a convolutional layer.

The convolutional layer is composed of multiple convolutional kernels, and the convolution process is not a one-time feature extraction of the input, but rather a weighted average calculation of multiple regions, i.e., local feature extraction, and then the combination of the extracted features to return to the output corresponding region. Convolutional networks use local sensing and weight sharing mechanism when performing convolution operation, compared with fully connected networks, this method mechanism can effectively reduce the number of parameters of the network and improve the computational speed of the model.

II. B. 2) Transposing Convolutional Layers

Transpose convolution, whose process is also a convolution operation, can be understood as the inverse process of convolution, but not simply reverse convolution, because the inverse convolution is a reduction of the value obtained after convolution, and the convolution process of the neural network on the input features are calculated, through the mapping of the new value obtained, the process is not reversible, so the transpose convolution is used here, whose main purpose is to be used for upsampling, and for the general convolution operation, the After convolution operation, the feature map size will be reduced, but for generating image task, it is necessary to restore the feature map to its original size, transpose convolution operation accomplishes this task.

II. C.Text Generation Image Method Based on Deep Convolutional Generative Adversarial Networks

II. C. 1) System framework

The overall network architecture of deep fusion generative adversarial network based on conditional enhancement and attention mechanism is shown in Fig. 2, and the whole network model consists of text processing network and generative adversarial network.

The text processing network consists of text encoder and conditional enhancement module. The bidirectional long and short-term memory network (BiLSTM) is designed as the text encoder to perform feature extraction on the text, and the conditional augmentation module (CA) further enriches the semantic information of the text.

The generative adversarial network consists of a generator G and a discriminator D. The generator consists of an upsampled residual block, an attention mechanism and a convolutional layer. The discriminator consists of a convolutional layer and a downsampling residual block, the generated image from the generator is used for feature extraction through the convolutional layer, the downsampling residual block downsamples the features and fuses the textual features to discriminate between the generated image and the real image, and combines with the MA-GP loss to design the antagonistic loss function for the evaluation of the network, and then provides feedback for updating the parameters of the generator to generate a higher-quality image.

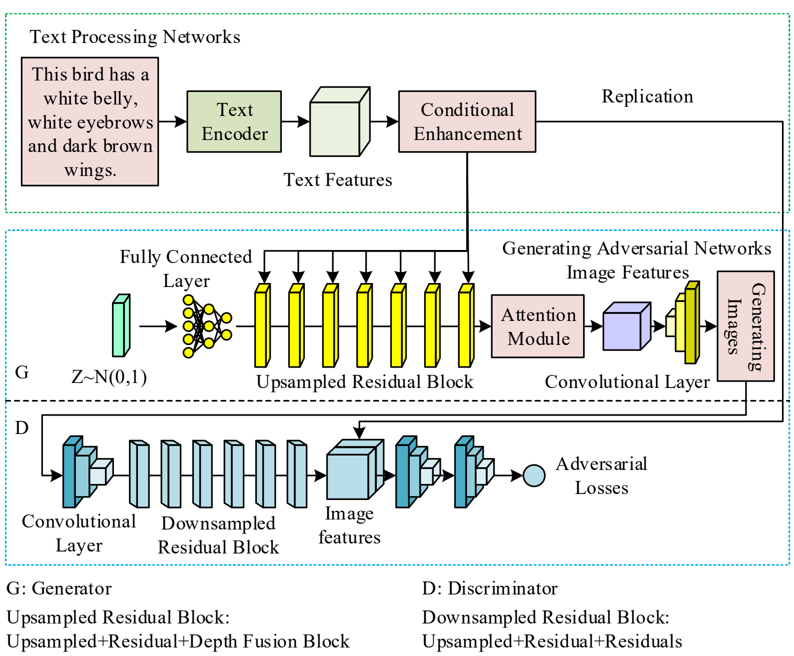

Figure 2: Deep fusion generation against network structure

II. C. 2) Fundamentals

The text encoder in the text processing network uses a bi-directional long short-term memory network whose goal is to learn textual feature representations from utterances by feeding the words in the utterance sequentially into a BiLSTM network [15]. The network consists of a number of long and short-term memory modules that capture bi-directional semantic information, memorizing valid information while discarding information that needs to be forgotten.

The number of LSTM modules is determined by the number of words, taking three words as an example, the BiLSTM model is composed of three forward LSTM modules and three backward LSTM modules. The forward hidden vector features $\{h_{L0}, h_{L1}, h_{L2}\}$ are obtained from $LSTM_L$, and the backward hidden vector features $\{h_{R0}, h_{R1}, h_{R2}\}$ are obtained from $LSTM_R$ and the two are concatenated as $\{[h_{L0}, h_{R0}],[h_{L1}, h_{R1}],[h_{L2}, h_{R2}]\}$, i.e., the text features $\phi_t = \{h_0, h_1, h_2\}$.

When the number of text features is small, the conditional enhancement module can enrich the semantic information of the text features to produce more text features, which makes the generated images more diverse. In order to obtain richer text features $\tilde{c}$, the initial text features ϕ_t are firstly input into the fully connected layer to generate the Gaussian distribution $N\left(\mu_0(\phi_t), \sum_0(\phi_t)\right)$ in μ_0 and σ_0, the mean function $\mu_0(\phi_t)$ and the diagonal covariance function $\sum_0(\phi_t)$ are the functions of the text feature ϕ_t. The $\tilde{c}$ is obtained from Eq:

$$\tilde{c} = \mu_0 + \sigma_0 \otimes \varepsilon \tag{4}$$

Generative Adversarial Networks contain two main components, the generator and the discriminator. Both are trained alternately and are continuously improved and optimized as they compete with each other.

The generator has two inputs. One is the text features obtained from the text processing network, and the other is sampled from the same standard normal distribution as ε to obtain a random noise vector $Z \sim N(0,1)$ with the same dimension of $100\times1\times1$. The generator consists of an upsampled residual block, a convolutional attention module [16] (CBAM), and a convolutional layer (with a convolutional kernel size of 3×3).

The upsampling residual block is set to seven layers, in which textual features are progressively deepened and fused with visual features to obtain high-resolution image features. The upsampling residual block consists of upsampling, residual network and depth fusion block.

Through the affine layer, the image features and text features are better fused. If the input text feature is $\tilde{c} \in R^{B\times C\times H\times W}$, the output of the multilayer perceptron MLP is a vector of size C. The affine layer employs two single-hidden multilayer perceptron MLPs to predict the channel scale parameter γ and shift parameter β of the linguistic condition from the textual feature $\tilde{c}$, respectively. As in Eq:

$$\begin{aligned} \gamma &= MLP_1(\tilde{c}) \\ \beta &= MLP_2(\tilde{c}) \end{aligned} \tag{5}$$

A scaling operation is performed on $\tilde{c}$, using a scaling scale parameter of γ, and a shifting operation is performed on $\tilde{c}$, using a scaling scale parameter of β. The affine transformation is as in Eq:

$$AFF(x_i \mid c) = \gamma_i \cdot x_i + \beta \tag{6}$$

The high-resolution image features that have been up-sampled with a residual block are fed into the CBAM attention module, which employs a simple and efficient forward convolutional neural network attention mechanism. It processes the input features through convolutional operations so that the features can be refined in both channel and spatial dimensions. This processing helps to improve the performance and accuracy of the model.The CBAM model consists of two sub-modules, the channel attention mechanism and the spatial attention, connected in a sequential manner. The two are combined by means of sequential linking. Input feature $F \in R^{C\times H\times W}$ into the CBAM module, and output one-dimensional channel attention feature $F' \in R^{C\times1\times1}$ and two-dimensional spatial attention feature $F'' = R^{1\times H\times W}$ in sequence. The total process is as in Eq:

$$\begin{aligned} F' &= M_C(F) \otimes F \\ F'' &= M_S(F') \otimes F \end{aligned} \tag{7}$$

where $\otimes$ denotes multiplication by elements, F' is the intermediate feature after the channel attention mechanism M_C, and F'' is the refined feature after the spatial attention mechanism M_S.

To compute the channel attention features, the spatial information of the maximally pooled and average pooled features is firstly obtained, and the maximally pooled features and average pooled features with different spatial context descriptors are generated: $F_{\max^c} R^{C\times1\times1}$ and $F_{avg}^{c} \in R^{C\times1\times1}$. The two descriptive features are fed into a multilayer perceptual machine to generate the corresponding 1D channel attention features, respectively. The shared module in this chapter includes the multilayer perceptual machine (MLP) and the corresponding hidden layers. To reduce the number of parameters of the network, the activation size of the hidden layer is set to $R^{\frac{C}{r}\times1\times1}$, where the reduction rate r is a tuned parameter. By element-wise summation of the output of the shared network with a specific activation function, the corresponding feature vector M_C can be generated.

In short, the channel attention features are computed as in Eq:

$$\begin{aligned} M_c(F) &= \sigma\Big(MLP\big(AvgPool(F)\big)+MLP\big(MaxPool(F)\big)\Big) \\ &= \sigma\Big(W_1\big(W_0\big(F_{avg}^{c}\big)\big)+W_1\big(W_0\big(F_{\max}^{c}\big)\big)\Big) \end{aligned} \tag{8}$$

where σ denotes the $sigmoid$ activation function, $AvgPool(F)$ and $MaxPool(F)$ denote average pooling and maximum pooling, + denotes element summation, and the weights W_0 and W_1 of the MLP are shared by the two inputs, where $W_0 \in R^{\frac{C}{r}\times C}$, and $W_1 \in R^{\frac{C}{r}\times C}$.

To compute the spatial attention mechanism, an average pooling operation is first applied along the channel dimensions to obtain the average features as $F_{avg}^{s} = R^{1\times H\times W}$, and then a maximum pooling operation is applied to obtain the maximally pooled features as $F_{\max^s} = R^{1\times H\times W}$, which are converted to the average features by using a standard convolutional layer to connect them and perform a convolution operation, and then use the activation function to generate the attention feature M_S in 2D space. The computation of spatial attention is shown in Eq:

$$\begin{aligned} M_S(F'') &= \sigma\Big(f^{7\times7}\big(\big[AvgPool(F');MaxPool(F')\big]\big)\Big) \\ &= \sigma\Big(f^{7\times7}\big(\big[F_{avg}^{s};F_{\max}^{s}\big]\big)\Big) \end{aligned} \tag{9}$$

The inputs to the discriminator are the generated image of the generator and the real image in the dataset. The discriminator consists of a downsampled residual block and a 2D convolutional layer.

II. C. 3) Training process and network loss function

The software equipment configurations used for the experiments in this chapter are Inteli7 CPU, NVDIA GeForce RTX3060 GPU. Using Pytorch as the deep learning framework for the experiments in this chapter. The network is optimized using Adam.

The network loss function is as in Eq:

$$\begin{aligned} L_D =& L_{Text\ Matching\ Real\ Images} + L_{Text\ Matching\ Generated\ Images} \\ &+ L_{Text\ Mismatch\ Losses} + L_{MA\text{-}GP} \\ &- E_{x\sim P_r}\left[\min(0,-1+D(x,\tilde{c}))\right] \\ &- \frac{1}{2}E_{G(z)\sim P_g}\left[\min(0,-1-D(G(z),\tilde{c}))\right] \\ &- \frac{1}{2}E_{x\sim P_{mis}}\left[\min(0,-1-D(x,\tilde{c}))\right] \\ &+ kE_{x\sim P_r}\left[\left\|\nabla_x D(x,\tilde{c})\right\| + \left\|\nabla_{\tilde{c}} D(x,\tilde{c})\right\|^p\right] \\ L_G =& -E_{G(z)\sim P_g}\left[D(G(z),\tilde{c})\right] \\ &+ D_{KL}\Big(N\big(\mu_0(\tilde{c}),\textstyle\sum_0(\tilde{c})\big)\Big\|N(0,1)\Big) \end{aligned} \tag{10}$$

L_G denotes the generator loss function, which is also computed using the function expectation. D_{KL} is the KL scatter of the standard Gaussian distribution $N\left(\mu_0(\tilde{c}),\sum_0(\tilde{c})\right)$ and the standard normal distribution $N(0,1)$, and the KL scatters of $\mu_0(\tilde{c})$, $\sum_0(\tilde{c})$ are the mean function, diagonal covariance function about the text feature $\tilde{c}$, respectively. This regularization term is for further smoothing the conditional augmentation module in the generator to avoid overfitting.

II. D.Web Image Generation Capability Analysis

II. D. 1) Experimental setup

For the training phase of this model, the CIFAR10 dataset and CelebA dataset are trained separately.In this chapter, experiments are conducted on the CIFAR10 and CelebA datasets, and the Resnet50 model, which is pre-trained on Imagenet, is used as the feature extraction network. The number of parameters for this model is 156M, the initial learning power is set to 0.0001, and a batch is 64 samples, which contains both generated and real samples. The training is carried out by training the generator and discriminator alternately, training the discriminator once for every 5 times the generator is trained. The training process for this experiment was done in linux environment using NVIDIA CUDA utilizing graphics card to accelerate the training.

II. D. 2) Comparative experimental results

In this experiment, IS was used to evaluate the generative ability of the network and as a comparison, WGANGP was also trained on the same dataset.Consistent with WGAN, the Adam optimizer was used in the training to update the network with the parameters of β1=0.5.β2=0.999, the learning rate was set to 0.0001, and the number of iterative rounds was 40,000.

The changes in IS scores for training this model and WGANGP on CIFAR10 dataset and CelebA dataset are shown in Fig. 3 and Fig. 4 respectively.

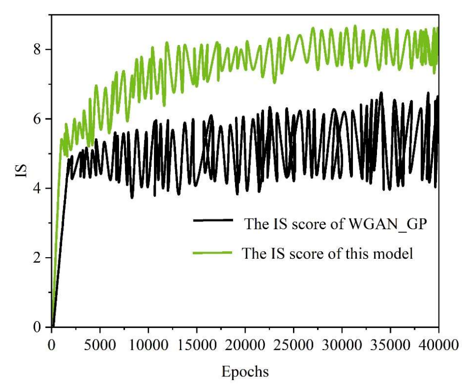

Figure 3: On the CIFAR-10 data set

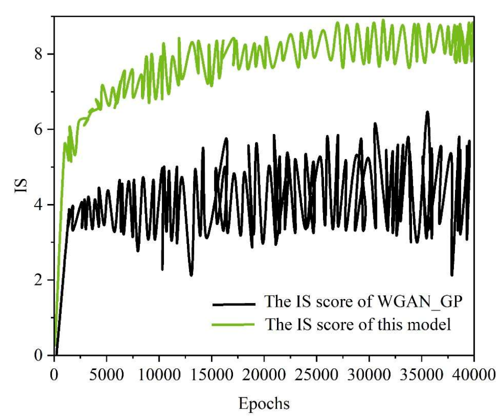

Figure 4: On the CelebA data set

The effectiveness of the present model in generating images exceeds that of WGAN on all datasets, and with the same number of training rounds, the present model has a very good performance in generating quality.The IS scores have high values with a small range of fluctuation, and the average value of the IS scores on CIFAR10 dataset is about 7.99, which is higher than that of the IS scores of WGANGP by about 1.5 points. The present model also achieves better results on celeba, proving the adaptability of the image generation method based on conditional enhancement and attention mechanism on different data and on. Due to the relatively homogeneous features of the face and the small difference in distribution, the effect of the present model on the improvement of the quality of the generated images is limited.

III. Results and analysis of the construction of visualized mathematics learning resources

III. A. Models for the development of visual learning resources

The visual learning resource development model is shown in Figure 5, which is generally divided into four stages: pre-analysis, resource structure design, resource specific development, and resource testing and evaluation, and is gradually improved through continuous feedback and correction, and cyclic repetition. Pre-analysis includes analysis of learning needs and hardware environment. Resource structure design mainly includes the overall architecture, interface design, resource presentation, visual representation and other aspects. In the resource development stage, the preparation and production of media resource materials will be carried out for specific requirements, while the overall design and development of resources will be completed. When resources are integrated, the principles of teaching and learning should be followed to make the articulation between resources more reasonable. After the design, production and implementation of learning resources, it is necessary to evaluate their effects and make subsequent improvements based on the feedback, which can also provide rich experience for the construction of resources in the future.

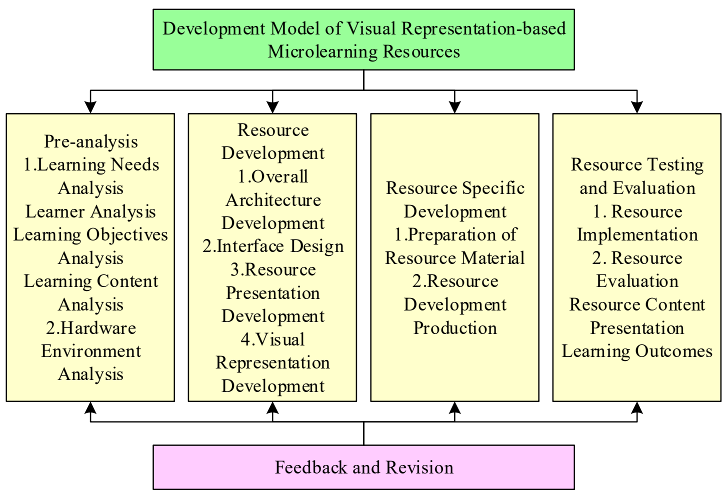

Figure 5: The development model of microlearning resources

III. B. Experimental analysis of visual learning resources

In this study, 50 first-year students from a middle school in City A were selected to test the effectiveness of the application of visual learning resources, and to test whether the use of visual learning resources would have an impact on students' different dimensions of mathematics learning ability and mathematics achievement.

The experimental data were collected mainly from students' pre and post-test scores and regular scores, student questionnaires on visualized learning resources and interviews. The results of the study are analyzed as follows. The learning effectiveness of the students was examined by analyzing their pre and post-test scores and six usual scores in the mathematics course. Table 1 shows the results of the comparison of the pre and post-test scores, the mean of the pre-test scores is 49.12 and the mean of the post-test scores is 78.45, which indicates that the students' math scores have improved significantly. Among the results of the paired-sample t-test of the pre and post-tests, the difference in means test $t = 19.004$, $p = 0.000 < 0.05$, indicating that there is a significant difference between the students' pre and post-test math scores.

Table 2 shows the descriptive statistics of the six usual grades, with a minimum score of 80 and a maximum score of 97, with a mean value of about 85 or more, indicating an excellent level of learning in terms of students' mastery of various topics in mathematics through visual learning resources.

Table 1: The results of the students' math(n=50)

	Number	average	Standard deviation	T value
Pre-test	60	49.12	7.98	19.004*
Post-test	60	78.45	7.56	

Table 2: The description of the average performance(n=50)

Theme	Minimum value	Maximum value	Mean	Standard deviationStandard deviation
Mathematics and formulas	86.00	97.00	91.00	3.25
Equation and inequality	80.00	93.00	90.50	2.57
Function	81.00	93.00	90.80	2.44
The nature of the image	81.00	93.00	90.00	3.56
Change of image	81.00	94.00	89.50	3.01
Statistics and probability	85.00	94.00	89.00	2.35

The correlation matrix between the six usual grades is shown in Table 3, and the analysis of the correlation matrix between the usual grades and the learning effectiveness (usual total grades) of the six themes shows that the correlation coefficients of "the nature of the image", "the change of the image", "equation and inequality" and "function" all reached a significant level of 0.01, and the correlation coefficient between "statistics and probability" and learning effectiveness was low.

Table 3: The correlation matrix of six normal grades

	1	2	3	4	5	6	Learning results
1.Mathematics and formulas	1						0.461**
2.Change of image	0.295*	1					0.715**
3.Equation and inequality		0.455**	1				0.688**
4.Function		0.377**	0.329*	1			0.651**
5.The nature of the image		0.305*	0.495**	0.334*	1		0.669**
6.Statistics and probability	-0.281*					1	0.105

Table 4 shows the results of the survey on the benefits of visual learning resources on learning, students generally agree that they can be viewed repeatedly, provide mathematics learning support materials, change the pace of playback, but also improve information retrieval skills, learn to correctly cite the literature, and help to complete the exercises more efficiently and so on, and the mean of students' agreement for the repeatable viewing reaches 4.39, but there is room to strengthen the aspects of stimulating the interest of the students, and the mean of students' agreement for this aspect is only 3.56. The mean value of students' agreement is only 3.56.

Table 4: The benefits of learning resources for learning(n=50)

Content	Average	Standard deviation
The is can be repeated	4.45	0.68
Provide courses for learning materials	4.39	0.55
Can change the schedule	4.38	0.68
Help improve information retrieval	4.33	0.61
Learn to quote the literature correctly	4.25	0.67
Help with higher efficiency	4.22	0.55
Help improve information management	4.18	0.54
Learn to make search	4.13	0.68
Better preparation for course testing	4.03	0.69
10. Flexible mastery of learning time	3.95	0.64
Help improve information evaluation	3.84	1.03
Arouse learning interests	3.56	0.77

The study designed the questionnaire based on the analysis above and the proposed model. The questionnaire includes 5 latent variables, each latent variable contains 4 observed variables, a total of 20 question items, in the form of a five-level Likert scale, and the options "strongly disagree", "disagree", "uncertain", "agree" and "agree very" are scored on a scale of 1 to 5 respectively. Before the questionnaire was officially launched, a small area of testing was conducted, and a total of 30 students in the "Mathematics" course in the second year of junior high school were surveyed. SPSS 19.0 was used to analyze the reliability of the questionnaire and its dimensions, and the overall reliability of the questionnaire was 0.977. Through the reliability analysis, it can be seen that the reliability of the questionnaire as a whole and all dimensions is higher than 0.7, and the reliability is good.

The questionnaire was formally implemented for students with experience of using online visual learning resources. The questionnaire was distributed to the users of mathematics courses on the visual learning resource development platform. The questionnaires were distributed to junior students who had just completed a math course at the university. The questionnaires were distributed through a link generated by “Questionnaire Star”, and the survey period was 7 days, with a total of 500 questionnaires returned, of which 450 were valid questionnaires. The data were analyzed using SPSS 19.0 and Amos 21.0. Table 5 shows the descriptive statistical analysis of the questionnaire results.

As can be seen from the table, all the questions scored above or close to 4, indicating that the learners' expectations of mathematics learning resources are high in these dimensions. The learning support dimension has the highest score, indicating that the support of learning by the resources is the most concerned factor for the learners.

Table 5: Questionnaire survey results describe statistical analysis

Latent variable	Observed variable	Mean value	Standard deviation	Mean value	Standard deviation
Content characteristics	Information input	4.13	0.845	17.17	3.541
	comprehensibility	4.26	0.879		
	fun	4.45	0.895		
	correlation	4.33	0.785		
Cognitive load	Information content	3.85	1.025	16.11	3.442
	Mission guidance	4.15	1.112		
	Picture content	3.85	0.954		
	Custom options	4.26	0.987		
Visual attention	Content mark	4.18	0.587	16.8	3.385
	Text prompt	4.24	0.598		
	Image prompt	4.25	0.845		
	Occurrence frequency	4.13	0.896		
Social feeling	Teacher image	3.98	0.954	16.78	3.426
	Contextual creation	3.89	0.577		
	Thinking interaction	4.55	0.569		
	Man-machine interaction	4.36	0.951		
Learning support	Autonomous learning support	4.24	0.922	17.56	3.445
	Exploratory learning support	4.15	0.913		
	Fragmentation chemistry support	4.59	0.845		
	Personalized learning support	4.58	0.831		

IV. Conclusion

In this paper, based on deep convolutional neural network, we design a text-generated image method based on conditional enhancement and attention mechanism, and construct a model for visualizing mathematical learning resources.

The effectiveness of the text-generated image method based on conditional enhancement and attention mechanism proposed in this paper is demonstrated through comparative experiments, which can enhance the quality of the network-generated images, and obtain better IS scores than WGANGP on both CIFAR10 and CelebA datasets, and IS scores are about 1.5 points higher than those of WGANGP, which indicates that the quality of the images generated by this paper's method is clearer.

In the empirical experiments, the visual math learning resource development model can be used as a supplemental math learning material for secondary school students, which can help to improve their learning outcomes and mathematical literacy. Students' mathematical literacy in the areas of image changes, equations and inequalities, functions, and properties of images were significantly improved, laying a good foundation for the development of their academic ability in mathematics. In the questionnaire survey, students' satisfaction with all the question items of the visualized mathematics learning resources development model was around 4 points, which again verified the rationality of the visualized mathematics learning resources development model designed in this paper.

The visualized mathematics learning resources development model in this paper not only improves the design method of visual learning resources based on visual representations, but also combines scientific theoretical guidance with teaching practice in order to be able to pay attention to the learning situation and learning effects of mathematics learners in a timely manner, which further enhances the students' mathematics learning ability and learning interest.

Funding

This work was supported by the Research on the Demand System for Online Teaching Content Resources in Primary and Secondary School Mathematics Based on the Satisfaction Evaluation Model (GD22CJY24), a 2022 annual regular project under the 14th Five-Year Plan for Philosophy and Social Sciences in Guangdong Province.

References

[1] Yan, S., & Yang, Y. (2021). Education informatization 2.0 in China: Motivation, framework, and vision. ECNU Review of Education, 4(2), 410-428.

[2] Zhou, L., Meng, W., Wu, S., & Cheng, X. (2023). Development of Digital Education in the Age of Digital Transformation: Citing China's Practice in Smart Education as a Case Study. Science Insights Education Frontiers, 14(2), 2077-2092.

[3] Wang, Y., Lin, H., Sun, L., Zhang, D., Wang, J., & She, L. (2020, December). Research on the Deep Integration of Modern Information Technology and Five Education. In 2020 International Symposium on Advances in Informatics, Electronics and Education (ISAIEE) (pp. 241-244). IEEE.

[4] Jacka, L. (2023). Flipping the focus: an innovative design strategy to support technology integration in teacher education. International Journal of Technology Enhanced Learning, 15(4), 397-411.

[5] Ziatdinov, R., & Valles Jr, J. R. (2022). Synthesis of modeling, visualization, and programming in GeoGebra as an effective approach for teaching and learning STEM topics. Mathematics, 10(3), 398.

[6] Nardi, E. (2014). Reflections on visualization in mathematics and in mathematics education. Mathematics & mathematics education: searching for common ground, 193-220.

[7] Presmeg, N. (2020). Visualization and learning in mathematics education. Encyclopedia of mathematics education, 900-904.

[8] Svitek, S., Annuš, N., & Filip, F. (2022). Math Can Be Visual—Teaching and Understanding Arithmetical Functions through Visualization. Mathematics, 10(15), 2656.

[9] Medina Herrera, L. M., Juárez Ordóñez, S., & Ruiz-Loza, S. (2024, February). Enhancing mathematical education with spatial visualization tools. In Frontiers in Education (Vol. 9, p. 1229126). Frontiers Media SA.

[10] Schäfer, M. (2021). Manipulatives as mediums for visualisation processes in the teaching of mathematics. Mathematics Teaching and Professional Learning in sub-Sahara Africa, 5-21.

[11] Dockendorff, M., & Solar, H. (2018). ICT integration in mathematics initial teacher training and its impact on visualization: the case of GeoGebra. International Journal of Mathematical Education in Science and Technology, 49(1), 66-84.

[12] Kadunz, G., & Yerushalmy, M. (2015). Visualization in the Teaching and Learning of Mathematics. In The Proceedings of the 12th International Congress on Mathematical Education: Intellectual and attitudinal challenges (pp. 463-467). Springer International Publishing.

[13] Shilong Fan. (2019). Research on Deep Learning Energy Consumption Prediction Based on Generating Confrontation Network.. IEEE Access,7,165143-165154.

[14] Wang Ru & Gong Daqing. (2025). Abnormal Behavior Recognition Algorithm in Small Sample Scenarios for a Utility Tunnel Project Based on DCGAN. Journal of Construction Engineering and Management,151(6),

[15] Zhao Dengfeng,Tian Chaoyang,Fu Zhijun,Zhong Yudong,Hou Junjian & He Wenbin. (2025). Multi scale convolutional neural network combining BiLSTM and attention mechanism for bearing fault diagnosis under multiple working conditions. Scientific Reports,15(1),13035-13035.

[16] Yanjie Qi,Huibin Liu & Xiaomin Ji. (2025). Multi-focus image fusion algorithm based on adaptive connection and hybrid convolution attention. Computers and Electrical Engineering,123(PC),110236-110236.