

## *Retraction*

# **Retracted: Attention Feature Network Extraction Combined with the Generation Algorithm of Multimedia Image Description**

### **Advances in Multimedia**

Received 10 October 2023; Accepted 10 October 2023; Published 11 October 2023

Copyright © 2023 Advances in Multimedia. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

### **References**

- [1] B. Sun, "Attention Feature Network Extraction Combined with the Generation Algorithm of Multimedia Image Description," *Advances in Multimedia*, vol. 2021, Article ID 6484128, 8 pages, 2021.

## Research Article

# Attention Feature Network Extraction Combined with the Generation Algorithm of Multimedia Image Description

Beibei Sun 

*Department of Information Engineering, Zibo Vocational Institute, Zibo 255314, Shandong, China*

Correspondence should be addressed to Beibei Sun; 20130919@stu.nun.edu.cn

Received 23 July 2021; Accepted 31 August 2021; Published 28 September 2021

Academic Editor: Zhendong Mu

Copyright © 2021 Beibei Sun. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In view of the issue that the features of the images in the shallow layer cannot be fully utilized when the image description is generated and the target association of the image cannot be sufficiently obtained, a generation method for the description of the acquisition of attention images is put forward in this paper. The proportions of the features of images at various depths are autonomously assigned based on the content data of the language model, and the images thus generated are all pictures with image features with attention. In this way, the effect of description generation of images has been improved. After the testing of the database, the results indicate that the calculation method of the algorithm put forward in this paper is more accurate than the top-down multimedia image algorithm generated by a single attention.

## 1. Introduction

In recent years, the mobile Internet has developed rapidly, which has brought about a rich and colorful daily life to people. More image data have appeared in the hot headline messages on the network platform. If the content of each image is labeled based on the manual processing method, the cost will be relatively high. Therefore, labeling the content of images based on the intelligent methods has become the main research direction in the field of computer technology at present [1–3]. The features of the input images are classified, and the corresponding image content is generated intelligently, which is an effective way of cross-media association. The quality of images extracted intelligently is mainly dependent on the identification capacity of the target in the image and the correlation of the target image. The result of multimedia image is the conversion of multimedia image to text information by using multimedia encryption retrieval and robot question and answer, which can be applied in various fields such as assisting the education for children and guiding the blind, which play a relatively significant role in the research of multimedia image.

In view of the issues observed in the above analysis, an attention feature network extraction based on the

multimedia image description generation algorithm is put forward in this paper, mainly focusing on speeding up the multimedia image feature extraction, optimizing the multimedia image method, and generating the corresponding analysis model as the key content of the research. The features of multimedia image are extracted by using the target detection algorithm, and the multimedia image description generation algorithm is applied to ensure one-to-one training. In this way, the content of multimedia image can be analyzed, and the capacity of network learning can be improved quickly.

## 2. Image Description Generation Algorithm Based on the Attention Feature Extraction Network

In this paper, an image description generation algorithm combined with the image feature extraction attention mechanism is put forward. Through the adaptive weight distribution of image features at different depths, the target area of the output image features is enhanced in a way that the influence of the background area in the image on the foreground features can be limited [2, 4, 5]. As shown in Figure 1, the algorithm put forward in this paper includes

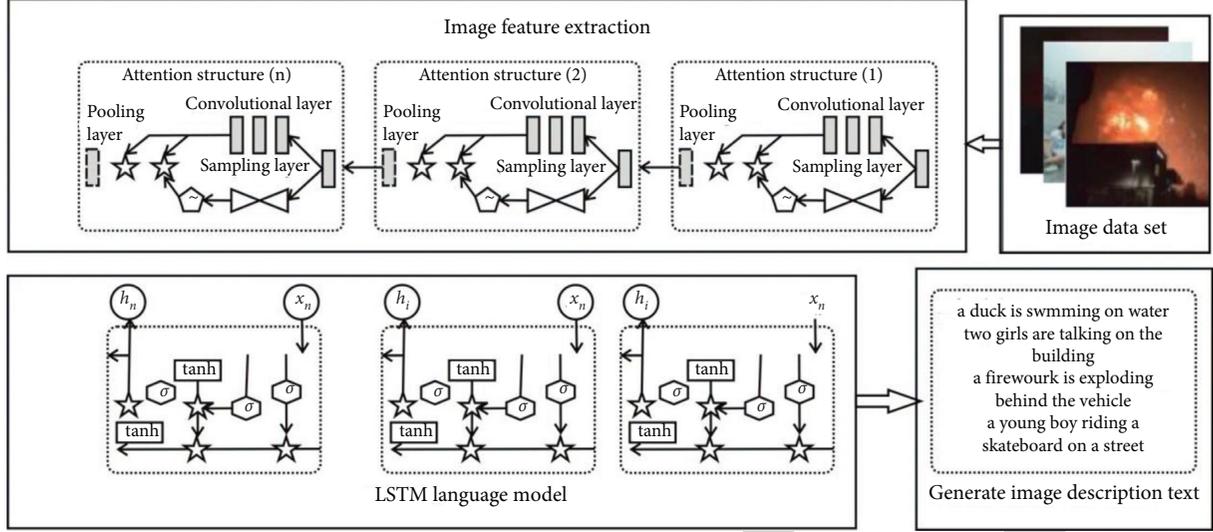


FIGURE 1: The image description generation algorithm of the concentrated feature extraction network.

two parts: the image feature extraction based on attention and the language generation.

**2.1. Extraction of the Image Feature.** In image detection algorithms based on feature, the extraction of features is the first step that is highly crucial. The features extracted in this paper are the characteristics in the input images. The feature extraction process can be divided into two steps: the attention feature network extraction and the feature link. The feature refers to a set of pixels in which the gray level of the surrounding pixels is changed step by step and the roof is also changed accordingly. In the presence of noise, the feature pixels detected by the attention feature network extraction operator are isolated or only small continuous segments in general. For the purpose of obtaining the continuous features, it is necessary to connect the feature pixels to the boundary of the constituent area. In the Sigmoid activation function layer, the feature map is normalized within (0, 1), and the output results are shown in the following equations:

$$V_t = \text{CNN}(X_{t-1}), \quad (1)$$

$$a_c = \tanh((W_s \times V_t + b_s) \oplus W_{hs} \times h_{t-1}), \quad (2)$$

$$M(x_{i,c}) = \frac{1}{1 + \exp(-a_c)}, \quad (3)$$

where  $x_{i,c}$  stands for the input feature map,  $c$  stands for the number of attention structure layers,  $W_s$ ,  $W_{hs}$ , and  $b_s$  stand for the linear transformation parameters to be learned, and  $V_t$  stands for the convolution of the output feature of the previous attention structure, which is used as the input of the next attention structure.

The output  $M(x_{i,c})$  of the sampling branch is multiplied by the bit relative to the output  $F(x_{i,c})$  of the primary branch, and each pixel of the output  $A(x_{i,c})$  from primary branch is processed by concentration weighting (Figure 2).

The output of the attention structure is shown in the following equation:

$$A(x_{i,c}) = M(x_{i,c}) \otimes F(x_{i,c}), \quad (4)$$

where  $\otimes$  stands for the multiplication of the counterbits.

The attention module is conducive to enhancing the crucial part of the feature map in each layer, and the superposition of the multilayer attention structure results in a significant decline in the performance of the model. The output of the sampling branch is normalized based on the Sigmoid function. In addition, if it is aligned with the primary branch, the part for the feature value of the layer is suppressed. If multiple attention structures are stacked for the subsequent calculation, the feature value of each pixel in the feature map that is finally output may become lower as a result. For the purpose of resolving the problem described above, the focus structure output is aligned with the position of the primary branch based on the positions of the sampling branch and the primary branch, and the output result of the attention structure is shown in the following equation:

$$A(x_{i,c}) = (M(x_{i,c}) \otimes F(x_{i,c})) \oplus F(x_{i,c}), \quad (5)$$

where  $\oplus$  stands for the addition of the counterbits.

Based on the feature map obtained through the fitting of the primary branch convolutional neural network, the attention features of the sampled branch outputs are combined. Through the primary branch output features, the essential features have been increased, and the nonessential feature  $A(x_{i,c})$  has been suppressed. As a result, the semantic information contained in the output features of the concentrated structure in each layer  $F(x_{i,c})$  is identically mapped compared with the semantic information contained in the output features of the primary branch. With the increase in the attention structure, the further intention based on the model is conducive to the extraction of the target.

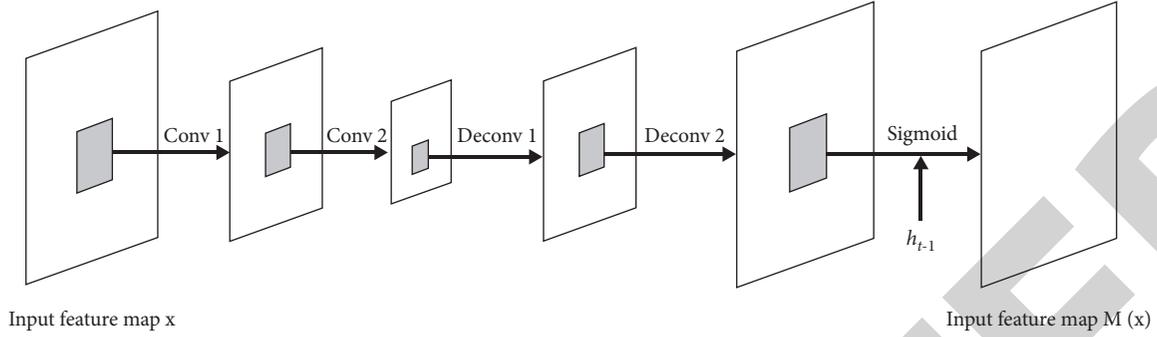


FIGURE 2: Branch structure of sampling.

**2.2. Language Generation Model.** LSTM is used as the basic unit of the language generation model, and the structure of the language model is shown in Figure 3 as the following.

In the initialization layer network of the input layer of the image feature  $A(x_{i,1})$  in the LSTM, as the output of the first attention structure, the input image feature is projected to the initialization implicit layer of the dimension  $d$  through linear transformation and ReLU activation function:

$$h'_0 = \text{ReLU}(W_0 \times A(x_{i,1}) + b_0), \quad (6)$$

where  $W_0$  and  $b_0$  stand for the parameters of the linear transformation to be learned.

The multiscale features extracted from the image are input into each layer of the LSTM in turn. The hidden layer of the  $n-1$  layer of the language model, that is, the  $h_{n-1}$  vector  $W_{\text{input}}$ , is combined with the image feature  $A(x_{i,c})$  output from the last layer in the attention structure and then input to the last layer of the LSTM language model.

$$h_t = \text{LSTM}(h_{n-1}, A(x_{i,c}), W_{\text{input}}). \quad (7)$$

The dimension output from the last layer of the LSTM maps the implicit layer of  $d$  to a vector of dimension  $m$ , in which  $m$  stands for the number of words in the dictionary. The LSTM model is selected through the Softmax layer, and the word with the highest probability in the output at each time is connected to the descriptive sentence and used as the final output result of the model.

$$\bar{y}_t \sim p_t = \text{Softmax}(h_t, W_{\text{input}}). \quad (8)$$

The cross entropy that is commonly used in the generation task based on the image description is trained as a loss function based on the model, and the form of the loss function is shown in the following equation:

$$\text{Loss}(\theta) = - \sum_{t=1}^T \ln(p_\theta(\bar{y}_t | y_{1:t-1})), \quad (9)$$

where  $y_{1:T}$  and  $\theta$  generate the actual word sequence and the image description of the target description, as well as the parameters for the decoder in the model, respectively.  $(p_\theta(\bar{y}_t | y_{1:t-1}))$  stands for the probability that the words are output based on the LSTM language model.

In Algorithm 1, the image description generation process based on the target image feature extraction network is described.

### 3. Experiment and Result Analysis

**3.1. Data Set and Experimental Environment.** The data set used in this paper is the data set of MSCOCO2014 [6–8]. The data set of MSCOCO can be used to identify multimedia image and multimedia image segmentation and create multimedia image description and other tasks. The data set includes the category of the objects contained in the multimedia image, the outline coordinates of the objects, the boundary coordinates, and the description of the multimedia image content, and the encryption description of each multimedia image has a minimum of five types. In this manual, the method of dividing the data set is used to divide the data set into a training set, a verification set, and a test set, which respectively contain 1123287 and 5000 multimedia image encryption. Although the length of all multimedia image description texts in the analysis data set has been described with reference to Figure 4 as shown in the following, the length of the description text is concentrated between 9 and 16 words. In addition, in the experiment, a list of vocabulary is created through the application of 16 or less sentences.

The experimental environment is established based on the Linux-based Pytorch  $h$  deep learning frame, which supports GPU computing, and equipped with NVIDIA CUDA 8.0 + cuDN-V5.1 deep learning library to accelerate GPU computing. The test software is Python 2.7. The hardware configuration used for training and testing is Intel Xeon CPU E5-2650. v3@2.30 GHz processor, equipped with the NVidia TITAN XP graphics card.

**3.2. Scoring Criteria.** The evaluation criteria for the generation of description by the existing multimedia image algorithms include manual subjective extraction evaluation and objective quantitative evaluation [9, 10]. In the subjective evaluation, the encryption of the output multimedia image is observed manually, and the quality of the encrypted description of the multimedia image is evaluated accordingly. At present, the most common objective quantitative scoring methods include the following: BLEU (Bilingual Evaluation Understudy), ROUGE\_L (longest common subsequence-based Recall-Oriented Understudy for Gisting Evaluation), METEOR (Metric for Evaluation of Translation with Explicit Ordering), and CIDEr (Consensus-Based Image Description Evaluation). In

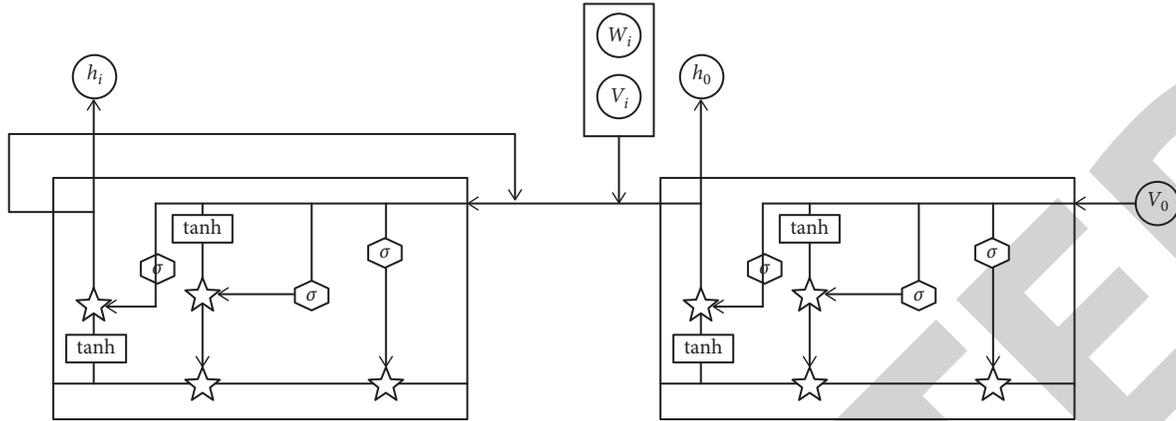


FIGURE 3: Structure of the language model.

Input: The image data set and the Wiki text data set are input.

Output: The image feature description text is output. The following steps are taken for each image in the data set:

Step1. The image feature  $V_1$  of the first layer is extracted;

Step2. The image feature of this layer is transferred to the first layer of the LSTM for the initialization of  $h_0$ ;

Step3. The image feature  $V_i$  of the  $i$ th layer is extracted;

Step4 The word vector  $W_{input}$ , the hidden layer  $h_{y-1}^n$  of the previous layer of LSTM, and the image feature  $V_i$  are input into the next layer of LSTM, and the next output word is calculated accordingly;

Step5. The loss "Loss" is calculated based on the cross entropy, and the parameters are adjusted according to the feedback;

Step6. Return to Step3 until the output is <END> or the maximum length of the sentence is reached;

Step7. Return the image description text.

ALGORITHM 1: The network image description is extracted and generated based on the feature of the attention image.

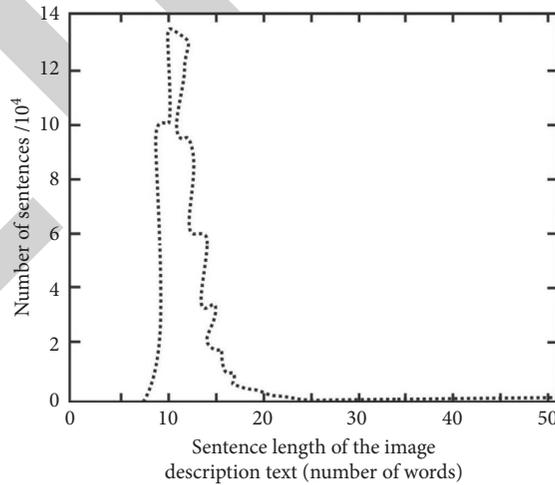


FIGURE 4: Distribution of the lengths of multimedia image description in the MSCOCO data set.

this paper, the results of the experiment are evaluated based on the criteria described above.

**3.3. Settings of the Main Parameters.** For the purpose of further illustrating the validity and evaluation speed of the continuity index put forward in this paper, the methods proposed in the article are compared (the parameters

include the distance threshold  $D = 10$  (number of pixels) and  $\text{Alpha} = \text{number of features}/20$ ). The high threshold of the images selected in the experiment is decreased when the vertical contrast is selected. The computer used in the experiment is equipped with a 2.30 GHz Intelli5 processor with 4 GB of memory, based on the data acquired in the algorithm implemented by Halcon and VSL 2010. In addition, the calculation time is the time used to execute the algorithm

100 times. From the experimental results of the comparison of the methods in this article, it can be observed that the two methods can both be used to evaluate the continuity level of the detection results of different image features, but the method put forward in this paper can distinguish the images more effectively. In addition, the evaluation time of the method put forward in this paper is less than that of the method described in literature [11–14]. Hence, it has significant advantages in the application.

Through several experiments described in the section above, it can be observed that the evaluation method put forward in this paper has reflected the continuity of the feature image effectively, which indicates that the method has the performance advantage in a way that it is suitable for human subjective cognition and has relatively short calculation time. The combination of the feature cam area and the feature length is selected as an indicator to measure the continuity of the feature segment, which has also the capacity to expand the space of the feature segment and is highly sensitive to the reflection of the fracture feature. In addition, the contribution of the length of the characteristic segment to the continuity is also taken into consideration. The bulge area and feature length are taken as the indicators to assess the continuity of the bulge feature. In this way, not only is the continuity description of the feature image accurate but also the calculation is simple and highly efficient, which has saved a lot of calculation time.

**3.4. Experimental Method.** For the purpose of verifying the influence of the feature of attention degree in the algorithm put forward in this paper on the effect of multimedia image description, the method based on single attention LSTM + ATT<sub>topdown</sub> is compared with the method based on attention multiscale fusion in this paper.

Through the method described in the above section, the algorithm model put forward in this paper is trained. The CIDEr score of the proposed algorithm is up to 1.154, and the BLEU score is up to 0.804. The objective quantitative evaluation method is used to evaluate the results of the algorithm in this paper, and the comparison score is shown in Table 1, where B@1, B@4 are the abbreviations of BLEU-4.BLEU-4. In the same data set and under the same training conditions, the objective quantitative evaluation of the multimedia image description generated by the algorithm put forward in this paper is relatively high.

After the model training of the algorithm put forward in this paper is completed, the focused search method is used to verify the quality of the proposed algorithm. With the increase in the cluster setting value, the evaluation of the encrypted description of the multimedia image generated by the model will be increased as well. No overfitting phenomenon is observed during the training of the model. When the focus is set to 3, the highest evaluation is obtained based on the model; and then as the focus is increased, the evaluation is no longer increased. When the focus is set to 3, the final evaluation of the model is shown in Table 2 as the following. From Table 2, it can be seen that, in the same data set and under the same training conditions, the objective

quantitative scores for the multimedia image description generated in the algorithm put forward in this paper by using the cluster search have been improved in different widths.

**3.5. Comparison of Experimental Results.** The multimedia image description generated in the algorithm model put forward in this paper is evaluated by using the objective quantitative scoring methods. Among them, the comparison results of mRNN, GoogleNIC, DeepVS, ATT-FCN (Attention model on Fully Connect Network), ATT-FCN, ERD (Encode Review and Decode), MSM (the other algorithms such as Multimodal Similarity Model) are shown in Table 3 as the following.

After the model training is completed, the multimedia image in the test set (as shown in Figure 5) is selected for testing. The description of each image is shown in Table 4 as the following (as shown in Figures 5 and 5(a), based on the algorithm put forward in this paper, not only can the people and phones be recognized, but also the relationship between the people and phones can be further illustrated; that is, the people are making a phone call). In Figure 5(b), the character in the image is described in detail based on the algorithm put forward in this paper; that is, the character in the image is a kid. In addition, the position of the character in the figure is described; that is, the character is on the street. In Figure 5(c), the relative positional relationship between the train and the river is described correctly. In Figure 5(d), the direction in which the character is skiing down the slope is illustrated in detail. In Figure 5(e), a parasol is detected. In Figure 5(f), more than one person in the multimedia image are correctly identified. It can be observed from the results described above that the description generated by the algorithm put forward in this paper can present the details in the multimedia image more accurately and effectively.

**3.6. Experimental Analysis.** The task of the multimedia image description can be divided into two main parts. The first part is the feature extraction of multimedia images. The second part is the establishment of a language model based on the features of multimedia images. In accordance with the experimental results of this paper, the improvement in the effect of the model multimedia image description put forward in this paper is mainly attributed to the following points:

- (1) In the extraction of features from multimedia images, the capacity of feature extraction of multimedia images can affect the final experimental results. The FasterR-Conn target detection model is taken as a feature extraction model, and the superior capacity of content detection in multimedia image is used to extract the features from the multimedia images. The results suggest that it can improve the description effect of the final multimedia image. In the experiment, the selected target detection model has demonstrated the Faster R-NN model in the pre-training with the most excellent performance, which

TABLE 1: Evaluation contrast of two generation algorithms for the multimedia image description after optimization by SCST.

Algorithm	B@1	B@4	METEOR	ROUGEL	CIDEr
LSTM + ATT <sub>topdown</sub>	0.799	0.362	0.267	0.567	1.146
Algorithm put forward in this paper	0.824	0.381	0.289	0.588	1.174

TABLE 2: Contrast after the addition of two generation algorithms for the multimedia image description to concentrated search.

Algorithm	B@1	B@4	METEOR	ROUGEL	CIDEr
LSTM + ATT <sub>topdown</sub>	0.801	0.367	0.269	0.57	1.159
Algorithm put forward in this paper	0.805	0.365	0.273	0.57	1.168

TABLE 3: Comparison of the scores obtained by different generation algorithms for multimedia image description.

Algorithm	B@1	B@4	METEOR	ROUGEL	CIDEr
mRNN	0.67	0.241	—	—	—
GoogleNIC	—	0.278	0.238	—	0.855
Hard-Attention	0.718	0.251	0.231	—	—
Adaptive	0.742	0.333	0.267	—	1.085
DeepVS	0.625	0.231	0.196	—	0.66
LSTM + ATT <sub>topdown</sub>	0.8	0.367	0.269	0.569	1.158
SCST: Att2in	—	0.314	0.261	0.543	1.013
ATT-FCN	0.709	0.305	0.244	—	—
ERD	—	0.299	0.241	—	0.895
MSM	0.73	0.326	0.252	—	0.986
Algorithm put forward in this paper	0.804	0.365	0.273	0.569	1.167



FIGURE 5: Test of the multimedia image selected for the comparison of algorithm effects. (a) Life image (chat on the phone). (b) Sports image (skateboarding on the street). (c) Traffic image (a train by the river). (d) Sports image (skiing). (e) Life image (beach). (f) Lifestyle image (cooking).

is used to extract the features from the multimedia images and improve the experimental effect.

- (2) The increase in the number of hidden nodes in the LSTM layer of the cyclic neural network decoding

model and the number of hidden nodes in the attention structure can improve the effect of the algorithm put forward in this paper effectively. In addition to cluster search, it can also improve the model effect. In the

TABLE 4: Comparison of the description details between the algorithm put forward in this paper and the single attention algorithm.

Serial number of the multimedia image	Description based on the single attention algorithm	Description of the algorithm put forward in this paper
Figure 5(a)	A woman is holding a cell phone	A woman is talking on a cell phone
Figure 5(b)	A person doing a trick on a skateboard	A young boy riding a skateboard on a street
Figure 5(c)	A train on the tracks of a river	A train on the tracks next to a river
Figure 5(d)	A person riding a snowboard in the snow	A man riding a snowboard down a snow covered slope
Figure 5(e)	A group of people sitting on the beach	A group of people with umbrellas on the beach
Figure 5(f)	a man is preparing food in a kitchen	A group of people preparing food in a kitchen

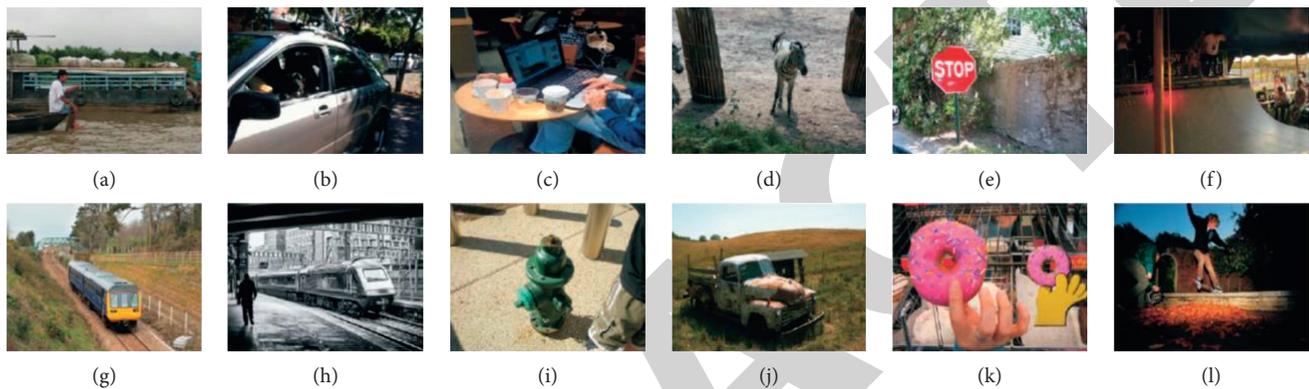


FIGURE 6: Test multimedia image selected for the testing of the algorithm put forward in this paper. (a) Folk style images (a fisherman). (b) Traffic image (a small car). (c) Life images (work and study). (d) Animal image (zebra). (e) Traffic image (a traffic sign). (f) Sports image (the sign of a skate park). (g) Traffic image (a train that is travelling). (h) Traffic image (a train station). (i) Public facility image (fire). (j) Environmental image (a dilapidated truck). (k) Life image (donut). (l) Sports image (skateboard stunt).

TABLE 5: Descriptive effect of the algorithm put forward in this paper in the test of multimedia image.

Serial number of the multimedia image	Description of the algorithm put forward in this paper
Figure 6(a)	A man sitting on a boat in the water
Figure 6(b)	A dog sitting in the driver's seat of a car
Figure 6(c)	A man sitting at a table with a laptop computer
Figure 6(d)	A zebra standing in the dirt of a fence
Figure 6(e)	A stop sign on the side of a road
Figure 6(f)	A group of people riding skateboards at a skate park
Figure 6(g)	A yellow and blue train travelling down the tracks
Figure 6(h)	A black and white photo of a train station
Figure 6(i)	A green fire hydrant on the side of a sidewalk
Figure 6(j)	An old truck parked in a field of grass
Figure 6(k)	A person holding a pink donut with sprinkles
Figure 6(l)	A man is doing a trick on a skateboard

process of adjusting the parameters described above, the algorithm put forward in this paper and the multimedia image description effect of LSTM + ATTtopdown have both been improved.

- (3) Through the application of the method for the feature network extraction, the significance information of the relationship between objects in the multimedia image can be obtained as a whole. At the same time, information such as the number or color of the

objects in the multimedia image can be obtained from the details.

For the purpose of further testing the effect of the algorithm put forward in this paper, the multimedia image in Figure 6 is selected and encrypted, as shown in the following. In addition, the algorithm model established in this paper is used to generate the multimedia image description, and the results are shown in Table 5. From these examples, it can be observed that natural language descriptions of multimedia

image can be generated with high accuracy based on the algorithm put forward in this paper.

#### 4. Conclusions

In this paper, based on the calculation method for the generation of image descriptions acquired for the attention features, a stacked attention architecture is adopted to add the content of the language model to the image characteristics, acquire the features of network images, and import each image feature into the LETM language model to generate the image description. Compared with the other calculation methods, the image descriptions based on the algorithm put forward in this paper have obtained relatively good results in the evaluation. Through the research, it is verified that the network model with different scales of attention in combination of multimedia image description has achieved very excellent results, and the application of multiple attentions can improve the performance of the multimedia picture encryption processing and the position relations significantly.

#### Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

#### Conflicts of Interest

The author declares no conflicts of interest.

#### References

- [1] J. Ma, H. Wan, J. Wang, H. Xia, and C. Bai, "An improved one-stage pedestrian detection method based on multi-scale attention feature extraction," *Journal of Real-Time Image Processing*, no. 4, pp. 1–14, 2021.
- [2] C. Liu, L. Huang, Z. Wei, and W. Zhang, "Subtler mixed attention network on fine-grained image classification," *Applied Intelligence*, pp. 1–14, 2021.
- [3] E. T. Nykaza, A. P. Boedihardjo, Z. Wang et al., "Deep learning for unsupervised feature extraction in audio signals: monaural source separation," *Journal of the Acoustical Society of America*, vol. 140, no. 4, p. 3424, 2016.
- [4] N. Brancati, G. D. Pietro, D. Riccio, and M. Frucci, "Gigapixel histopathological image analysis using attention-based neural networks," *IEEE Access*, vol. 99, p. 1, 2021.
- [5] C. Cao, K. Sun, and W. Liu, "A novel bit-level image encryption algorithm based on 2d-licm hyperchaotic map," *Signal Processing*, vol. 143, pp. 122–133, 2017.
- [6] Y. Chen, "Sports sequence images based on convolutional neural network," *Mathematical Problems in Engineering*, vol. 2021, Article ID 3326847, 14 pages, 2021.
- [7] K. F. Wang, S. Zhuang, and X. R. Zhao, "Jpeg image encryption algorithm based on three-dimensional multi-chaotic system," *Applied Mechanics and Materials*, vol. 734, pp. 554–557, 2015.
- [8] M. Xu, C. Zou, R. Liang, W. Li, and Q. Wang, "Feature extraction and recognition of auditory attention in middle latency response from single electrode," *Journal of Southeast University (Natural Science Edition)*, vol. 47, no. 3, pp. 432–437, 2017.
- [9] X. Chen, J. Li, Y. Zhang, Y. Lu, and S. Liu, "Automatic feature extraction in x-ray image based on deep learning approach for determination of bone age," *Future Generation Computer Systems*, vol. 26, no. 6, pp. 1–20, 2020.
- [10] N. Ståhl, G. Mathiason, G. Falkman, and A. Karlsson, "Using recurrent neural networks with attention for detecting problematic slab shapes in steel rolling," *Applied Mathematical Modelling*, vol. 70, no. 4, pp. 365–377, 2019.
- [11] K. Qian, L. Tian, Y. Liu, X. Wen, and J. Bao, "Image robust recognition based on feature-entropy-oriented differential fusion capsule network," *Applied Intelligence*, vol. 20, no. 2, pp. 506–515, 2020.
- [12] S. Halagowda, S. K. Lakshminarayana, and S. Lakshminarayana, "Image encryption method based on hybrid fractal-chaos algorithm," *International Journal of Intelligent Engineering and Systems*, vol. 10, no. 6, pp. 221–229, 2017.
- [13] L. Sergey, M. Vasily, K. Innokentiy, and K. Victor, "A spiking neural network in semg feature extraction," *Sensors*, vol. 15, no. 11, pp. 27894–27904, 2015.
- [14] X. Dai, J. Cheng, Y. Gao et al., "Deep belief network for feature extraction of urban artificial targets," *Mathematical Problems in Engineering*, vol. 2020, Article ID 2387823, 13 pages, 2020.