International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 17 (2025) pp. 1-15

https://doi.org/10.70917/ijcisim-2025-0271
Article

An Analysis of the Spatial Layout of Housing on the Presentation of Characters' Inner Conflicts and Social Contexts in Film Narratives Based on Image Processing

Xuanyu Yu *

College of Art and Design, Xi'an Mingde Institute of Technology, Xi'an, Shaanxi, 710124, China;
murfish2024@126.com

Abstract: The study explores how the spatial layout of housing therein has an impact on character conflict and social context from the perspective of movie narrative context. Image processing technology is used to automatically analyze the movie frames, and a generative adversarial network model DAE-GAN based on dynamic hierarchical semantic perception is constructed, which can automatically capture spatial features from the movie frames. The experiment proves the excellent performance of DAE-GAN, which can accurately label 179 good keywords out of 385 keywords, with a 68.28% checking completeness rate and 82.55% checking accuracy rate. Especially for words with clear structure, the check accuracy rate reaches more than 94%. For the abstract atmosphere words such as "dim" and "small", the recognition effect is also much better than other models. The empirical analysis based on 200 subjects shows that the layout of housing significantly affects the narrative, with cramped space being the most likely to trigger character conflict, with a rating of 4.24±1.14, and the social background being more prominent, with a rating of 4.04±1.17, because crowded environments naturally bring out the issues of class and resources, followed by complex spaces, while spacious and conventional layouts are relatively bland. ANOVA showed that the effects of cramped and complex layout were significant, but open space did not have a significant effect on either.

Keywords: image processing; movie narrative; DAE-GAN; housing spatial layout; character internal conflict; social context

1. Introduction

The visual style of the spatial layout of the film is a key factor in the realization of its aesthetic effect, the interpretation of the film is a combination of specific content and presentation, whether the film can effectively mobilize the aesthetic passion of the audience is a prerequisite for the success of a work [1-3]. Behind the visual pleasure, displaying the inner spirit of the objective object is the materialization of the artist's emotion and aesthetic consciousness [4]. On the basis of fully understanding the content and theme of the film, the film designer, through his unique spatial visual imagination, carries out the overall layout and mobilization of the screen, and designs and sculpts the specific details, so that it finally becomes the audience's aesthetic object appearing in front of the screen [5-6]. The spatial layout of the housing of the movie narrative powerfully demonstrates the inner conflict of the characters and the social background through the visual language.

The idea that inner conflict is instructive to characters' behavior, that is, the complex conflicts within characters that influence their behavior and guide their actions, can be traced back to the level of the spiritual world [7]. The spiritual world is the world generated by the conscious activities of human beings, which is strongly subjective and is formed by the combination of the conscious activities of the characters in the play and the results resulting from their activities [8]. Hegel's study of conflict suggests three different conflict situations [9]: first, conflict caused by physical or natural conditions. The second is a conflict of the mind caused by natural conditions. The third is a split caused by differences in

mindfulness, which is the really important conflict.

Image processing, which focuses on compression, enhancement, restoration, matching, description, and recognition of images, is a technology that uses computers or other digital devices to process and manipulate captured images, converting image information into digital information [10-12]. Its application has been involved in various industries, such as agriculture, architecture, medicine, etc., especially in recent years, the popular Internet of Things relies on image processing technology, and its application in the field of film includes, but is not limited to, the processing of movie special effects, restoration of old movies, etc. [13-15]. The application of image processing technology to the analysis of the spatial layout of housing in film narratives, due to the ability to accurately locate the overall spatial situation and fast recognition speed, will enable the analysis and evaluation of the results of the spatial layout of housing to be significantly improved in terms of efficiency and accuracy, and to help the researcher to carry out the analysis of the inner conflict of the film characters and the display of the social background [16-17].

With the purpose of interpreting the role played by architectural space in films, the study innovatively uses image processing technology as an entry point to analyze the spatial narratives of films and the hidden character conflicts and social backgrounds behind them. The study focuses on the spatial layout of housing in film narratives, starting from the most basic sense of scale and exploring how the director utilizes the camera to let the audience perceive the grandeur or oppression of the space. Then, in the contextual construction of architectural space in the film, it analyzes how a single visual impression of the building is sublimated into a context rich in deep meaning. In particular, it analyzes how the imagery evolves and deepens the theme in the film through the technique of progressive repetition. In order to capture and analyze these spatial imagery more objectively, a Dynamic Hierarchical Semantic Perception Generative Adversarial Network model (DAE-GAN) is constructed on this basis. The model mainly consists of three modules: multi-granularity semantic representation, initial image generation, and dynamic repainter. This enables it to automatically identify multi-level features such as spatial scale, material texture, and light mood in complex movie scenes, and associate these visual elements with semantic words like "oppressive", "warm", and "distant". Finally, all the previous perceptions and analyses are put into practice in terms of specific narrative expression techniques, explicitly pointing out how the housing space in the film has an impact on the characters' conflicts and the social context through materials, spatial structures and constructive practices.

2. Research on the integration mechanism of architectural space and movie narratives

2.1. Perceptual spatialization - the resonance of architectural scale and cinematic narrative

Architectural scale refers to the internal visual feeling of people for the size and depth of architectural space, which is the relative psychological perception of architectural space and its details, not only the quantitative description of its physical size. The difference in the sense of scale can often be expressed as a rich visual impression of grandeur and magnificence, crampedness and narrowness, roughness and thickness, and smallness and delicacy. Therefore, the concept of scale echoes with the spatial perception of film often carries a human-centered humanistic concern, emphasizing the individual's sense of presence at a particular moment. True architecture is the exchange of experience feeling and meaning between the materially constructed space and the subject's spiritual space. The dynamic expressiveness shared by architecture and cinema reveals how both communicate authentic emotional experiences through the uniqueness of the medium. The narratives in many sequences of cinematic works, driven by spatial representations at different architectural scales, display an impressive tension of visual presentation. This paper analyzes the inner conflict of the characters and the social background presentation behind the refraction.

2.2. The construction of mood in movie architectural space

Architecture itself is a kind of imagery, a material existence and spiritual preservation, symbolizing mankind's understanding and expression of space and time, and containing symbols and connotations of a specific cultural and historical background; it is not only the existence of physical space or a functional spatial carrier, but also a carrier of spirit and emotion and a crystallization of thought and art. Regarding the four elements of heaven, earth, man and god in the settlement, it can also be understood as the concept of “four in one”, which means that man discovers himself when he settles down, and this settlement can determine his life and way of existence. In this way, the movie creates a sense of space and atmosphere through the careful design of the architectural sets and camera angles, making the architecture part of the

movie's mood, and making it a real settlement and existence.

2.2.1. Connotation of architectural imagery highlights movie themes

The theory of spatial perception deals with how humans understand and interact with space through sensory experience and cognition. Architectural elements are not only background decorations, but also highlight the inherent themes of a movie. Architectural styles and structures can convey the context of time and guide the viewer into a particular historical or cultural environment; the cultural soul of a building is rooted in the cultural roots in which it is situated. Architecture can become the stage for a movie, providing a unique space and context. Movies support and emphasize the emotional atmosphere and themes of the film by choosing different architectural environments. Architectural elements can become symbols in a movie. A particular architectural style, structure or location may represent a specific concept on a cultural, historical or social level, and these symbols infuse the movie with deeper symbolism through the choice of architecture and the way it is presented.

2.2.2. Repeated progression of architectural imagery deepens the mood of the movie

The instantaneous generation of artistic imagery in the movie illuminates the obscured history and the reality of existence, and this artistic reality can strengthen the richness and profoundness of the movie realm through the “progressive repetition” means of expression of imagery. First of all, the progression of imagery can be carefully constructed through the reproduction of different scenes, objects or symbols. These preliminary images not only create a unique atmosphere for the movie, but also effectively stimulate the audience's curiosity and aesthetic expectation. Secondly, through the development of the plot, the progression of imagery can lead the audience into the deeper level of the story. Repeated or mutated imagery can project new meanings in different scenes and deepen the audience's understanding of the movie's theme. This process of progression is like a jigsaw puzzle that gradually reveals a larger and more complex picture. At the same time, the progression of imagery is also realized through the film's visual language, sound effects and the director's narrative techniques. Specifically, this includes the use of camera movement, editing techniques, sound changes and other technical means, so that the imagery does not only remain at the static level, but evolves organically with the passage of time and the development of the plot. Eventually, the progression of imagery reaches its peak as the movie moves toward a climax or resolution of the conflict. At this point, the audience has a deeper understanding of the imagery that appeared earlier and is more easily moved by the movie. The progressive and deepened imagery becomes a key element of the movie's mood, infusing the story with richer emotion and reflection.

2.3. Dynamic Hierarchical Semantic-Aware Generative Adversarial Network Model (DAE-GAN)

On the basis of the contextual construction in the previous section, in order to explore more deeply how architectural imagery is semantically expressed and emotionally transmitted in film narratives, a Dynamic Hierarchical Semantic Perception Generative Adversarial Network model (DAE-GAN) is introduced to realize the transformation from image features to semantic perception.

This section first introduces the formal definition of the text to image generation task, including specific task input as well as output definitions. Given a textual description, a realistic and graphically semantically consistent image, denoted as $I = G(T)$, needs to be generated. Where $T = \left\{ T_j \mid j = 0,1,\ldots,l-1 \right\}$ denotes the textual description consisting of l words, $G(\cdot)$ is the image generator to be learned, and I is the generated image. In order to improve the quality of the generated images, the following two key issues need to be addressed:

In order to make the image look more realistic, how to guarantee the realism of the generated image from the overall image to the local details? How to ensure the consistency of graphical semantics from the overall to the local? Focusing on the above problems, in order to improve the quality of the generated images, especially to improve the image detail learning ability, this chapter proposes the Dynamic Hierarchical Semantic Aware Generative Adversarial Network model (DAE-GAN), whose model structure is schematically shown in Fig. 1.

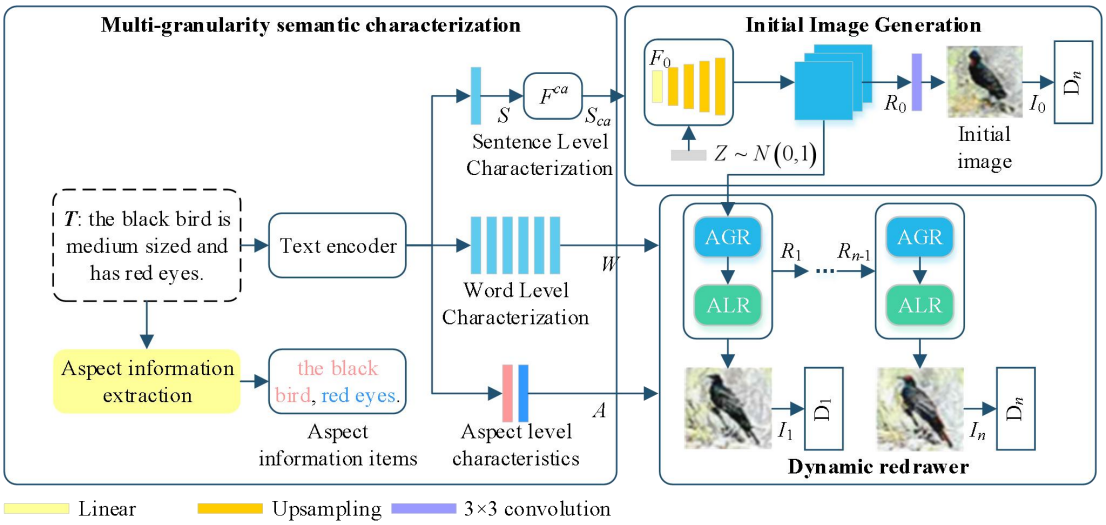

Figure 1. The DAE-GAN model structure.

The DAE-GAN model contains three main modules. Among them, (1) Multi-Granularity Semantic Representation: comprehensively characterize the textual semantics from multiple granularities, including sentence-level, word-level and aspect-level; (2) Initial Image Generation: generate a low-pixel image based on sentence-level features and random noise vectors; (3) Dynamic Redrawer: dynamically refine the image generated in the initial stage from global and local perspectives, respectively, in the way of hierarchical modeling, which is also is the focus of the research in this chapter. The specific technical details of the DAE-GAN model are described in detail next.

2.3.1. Multi-Granularity Semantic Representation

A comprehensive understanding of text semantics plays a crucial role in text to image generation tasks. Previous approaches mainly extract text features from sentence level and word level. However, they ignore the aspect-level information embedded in text descriptions. Aspect information refers to the use of a number of words rather than a single word to describe a specific part or feature of a thing. The granularity of aspect level information happens to be between the sentence level and the word level, which is very important to help synthesize the details of an image, and thus should attract more attention. In this paper, we propose a comprehensive characterization of textual features at multiple granularities, including sentence level, word level, and aspect level. Specifically, a Long Short-Term Memory (LSTM) network is employed to extract the semantic encoding of the textual description $\mathbf{T}$, which can be formalized as follows:

$$s, \mathbf{W} = \mathrm{LSTM}(T) \tag{1}$$

where $T = \{T_j \mid j = 0,1,\ldots,l-1\}$ consists of l words. $\mathbf{W} = \{\mathbf{W}_j \mid j = 0,1,\ldots,l-1\} \in \mathcal{R}^{l \times d_w}$ denotes the matrix of word-level features extracted from each time step of the LSTM network, d_w refers to the dimension of the textual input representation. The $s \in \mathcal{R}^{d_w}$ denotes the vector of sentence-level representations obtained from the hidden state of the last time step of the LSTM.

Further, a conditional augmentation (CA) method is used to augment the training data and avoid overfitting by resampling the input sentence vectors from an independent Gaussian distribution. Specifically, in this paper, the CA method is used to augment sentence-level features, and the process can be formalized as follows:

$$s_{ca} = F^{ca}(s) \tag{2}$$

Here, F^{ca} denotes the CA function. s_{ca} is the sentence-level semantic representation enhanced by CA.

As mentioned earlier, aspectual information is important for the details of the generated images. However, it is not easy to accurately identify and extract aspectual information from each textual description because the focus and presentation structure of each textual utterance is generally different. Therefore, this paper solves this problem with the help of syntactic structure. Specifically, the NLTK tool is first used to do lexical annotation for each text description. Then, according to the characteristics of

different datasets, corresponding rules (i.e., regular expressions) are designed to extract aspect information. Based on this, the aspect information $\{asp_i \mid i = 0,1,\ldots,n-1\}$ can be obtained. Next, LSTM is utilized to extract the aspect features, which are formally represented as follows:

$$\mathbf{A} = \mathrm{LSTM}\left(\{\mathbf{asp}_i \mid i = 0,1,2,\ldots,n-1\}\right) \tag{3}$$

where $\mathbf{A}$ denotes the aspect level features of the textual description and n is the number of extracted aspect information.

2.3.2. Initial Image Generation

According to the general approach, a low-pixel image is first generated in the initial stage. The initial image I_0 is generated using an augmented sentence level representation $\mathbf{s}_{ca}$ and a random noise $\mathbf{z}$. where $z \sim N(0,1)$ is a vector sampled from a normal distribution. In mathematical form, $\mathbf{R}_0$ is used to denote the image features generated at the initial stage:

$$\mathbf{R}_0 = F_0\left(\mathbf{s}_{ca}, \mathbf{z}\right) \tag{4}$$

where F_0 denotes the initial stage image generator, which consists of one fully connected layer and four upsampled layers.

2.3.3. Dynamic redrawer

This paper is the first work to introduce aspectual information embedded in a given text sentence to the task of generating images from text. Therefore, how to incorporate aspectual information into the image refinement stage is an important challenge that needs to be addressed urgently. To this end, this chapter proposes a newly-participated dynamic redrawer (ADR) to refine images by incorporating aspectual information embedded in sentences. Specifically, an Attention Mechanism-based Global Refinement (AGR) module is designed to utilize fine-grained word-level features for global refinement, and an Aspect-aware Local Refinement (ALR) module is designed to utilize aspect-level features for local refinement. By dynamically calling the two modules AGR and ALR alternately, image details can be refined from global and local perspectives, respectively. In the next section, the technical details of the AGR and ALR modules are described in detail using the i th refinement operation step as an example.

Fig. 2 shows a schematic diagram of the dynamic redrawer structure. Among them, Fig. (A) shows the schematic diagram of global refinement (AGR) based on attention mechanism, and Fig. (B) shows the schematic diagram of aspect-aware local refinement (ALR).

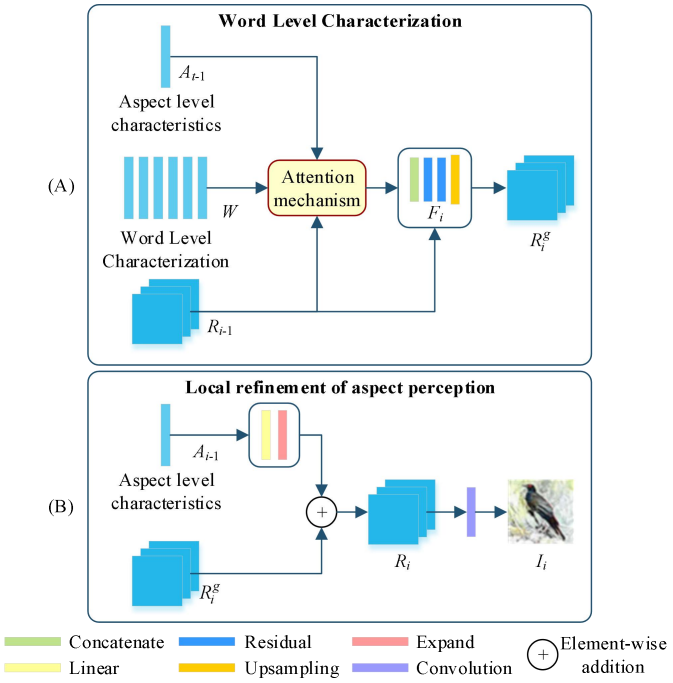

Figure 2. Dynamic Redrawer Structure.

(1) Global refinement of attention

In order to generate a realistic and graphically semantically consistent image, it is essential to refine the image from a global perspective using fine-grained features. Therefore, based on the initial image, AGR is designed for global refinement.

Specifically, word-level text features are mainly used here for global refinement. This process takes into account the different importance of each word for the current refinement step. The previous work focuses on updating the word-level features by selecting the important words using the image features from the previous step through the attention mechanism. This section further utilizes a combination of image features and aspect level features to enhance the word level features. The process can be formally represented as follows:

$$\begin{aligned} \mathbf{R}_i^g &= F_i\left(\mathbf{R}_{i-1}, \mathbf{W}_i^g\right), i = 1, 2, \ldots, n \\ \mathbf{W}_i^g &= \sum_{j=0}^{l-1}\left(\mathbf{W}_j \mathbf{U}\right)\alpha_{i,j} \\ \alpha_{i,j} &= \operatorname{softmax}\left(\left(\mathbf{W}_j \mathbf{U} + \mathbf{A}_{i-1}\mathbf{V}\right)\mathbf{R}_{i-1}\right) \end{aligned} \tag{5}$$

where $\mathbf{R}_i^g \in \mathcal{R}^{d_r \times N_i}$ denotes the globally refined augmented image features, which are generated from the previous stage image features $\mathbf{R}_{i-1} \in \mathcal{R}^{d_r \times N_i}$ and the word level features obtained based on the attention mechanism are fused to generate. N_i is the size of the image feature $\mathbf{R}_i^g$ at step i. $F_i(\cdot,\cdot)$ is the image feature mapping transformation function. $\mathbf{W}_i^g \in \mathcal{R}^{d_r \times N_{t-1}}$ denotes the global features based on the attention mechanism. $\alpha_{i,j}$ denotes the attention weight coefficients. $\mathbf{U} \in \mathcal{R}^{d_u \times d_r}$ and $\mathbf{V} \in \mathcal{R}^{d_w \times d_r}$ are learnable weight matrices used to incorporate the word-level features $\mathbf{W}$ and aspect-level features $\mathbf{A}$ into a semantic space that is harmonized with visual features.

(2) Local Refinement of Aspect Perception

In the previous section, it has been described how to use word-level features to refine images from a global perspective. However, the enhancement of some specific image details has not been fully accomplished. As mentioned in the previous section, the aspect information embedded in the text description is important for synthesizing the corresponding local image details. Therefore, ALR is further designed to refine the image from a local perspective using aspect level features.

Technically, this part fuses the aspect level feature $\mathbf{A}_{i-1}$ with the globally refined image feature $\mathbf{R}_i^g$, and this operation is accomplished by bitwise summation, which is formally represented as follows:

$$\mathbf{R}_i = \mathbf{R}_i^g + \left[\mathbf{A}_{i-1}\mathbf{V}\right] \star N_i, i = 1, 2, \ldots, n, \tag{6}$$

where the operation $A_i \star N_i = [A; A; \ldots; A]$ denotes the repetition of the splicing N_i times for A_i. Finally, in order to generate a realistic image, a convolution function of 3×3 is introduced to transform the refined features $\mathbf{R}_i$ into an image I_i in the i th refinement step of the ADR module. Overall, the AGR and ALR modules are applied alternately. Meanwhile, aspect level features are dynamically considered in each refinement step of ADR.

2.4. Narrative thinking expression of architectural space

Supported by the DAE-GAN model, we further return to the nature of architectural space as a narrative medium. In the following, we will look at the dimensions of material, structure and construction to illustrate how architectural space conveys the inner and social messages of characters through narrative thinking.

Architectural narrative is to give experiential and essential meaning to architectural space, and to convey architectural discourse in a narratological way. Studying the expression of architecture in narrative context can harmonize the contradiction between people and architecture, site and architecture, express social culture, human history and geography with architecture, and use the constituent elements of architecture as a way of narrative, which can better express and understand the real connotation of architecture.

A complete architectural narrative contains the designer's pre-creation, the expression of architectural objects and the user's experience. The narrative mechanism of architecture can be divided into translation, intervention and reconstruction, and the medium of architectural narrative can be spatial form, material

use, construction practice, light and shadow design and other elements, combined with the narrator's creation and design strategy and the receiver's behavioral activities and emotional experience, to form a complete architectural narrative system, and the receiver is able to accept the information conveyed by the building and the space in the most familiar and easy-to-understand way.

2.4.1. Material Use Narrative Expression

The selection of different materials and the different use of the same material will produce different expression effects, and will also be affected by the architect's secondary creation, the material has a sense of place, in addition to its apparent material information, there is also a subjective experience of cognition, which is an important factor affecting the atmosphere of the space. The material intuitively plays a role in people's tactile and visual senses, and stimulates people's reasoning and imagination based on their experience and memory, so the material is endowed with symbolic intention, which in turn stimulates the viewer's emotion.

Different materials have different colors and textures, and the designer selects and redesigns the materials with the theme emotion and spatial tone of the building. The expression of the material includes the style determined by the properties of the material and the processing of craftsmen's skills and techniques. In this way, the spatial tone of the building is endowed with emotional factors, mobilizing visitors' psychological perception and emotional changes in the architectural space. Wood, bamboo, earth, stone, brick, concrete, artificial plates, metal, glass, these common building materials have their own material properties, and the emotional tone embedded in the material is an important part of the narrative of the architectural space. For example, wood makes people feel close to nature, concrete makes people feel solemn and dignified, and glass makes users feel the psychological implication of transparency and openness.

2.4.2. Narrative representation of spatial structure

The architectural space is built as a scene for narrative through spatial contextualization, and the spatial scene is the ontology of the narrative. The narrative of architectural space has a great relevance to the expression of material, the choice of style and the arrangement and organization of space.

Visitors' behavioral events and the space in which they are located form a unit plot, and the tandem nature of the path space forms a multifaceted spatial plot. Architectural spatial narrative can be divided into two categories, one is the architectural spatial narrative through the performance of the theme scene, this category is mainly through the space of the scenery, materials and construction to create a narrative theme; the other is the pre-arranged spatial narrative of the functional events, this kind of narrative is the behavior of the space user and the activities of the arrangement, through the presetting of the functional events to guide the human behavioral activities, such as the exhibition order of the museum For example, the order of exhibition and theme display in exhibition museums. Visitors experience the architectural space in multiple directions and levels as a result of changes in the location of activities. The interaction between the building and the visitors stimulates their memorized life experience through the visitors' perception, which transcends the boundaries of time and space and embodies the narrative meaning of the space.

2.4.3. Narrative representation of constructed practices

Architecture is the basis for a building to maintain its material nature, and the combination of the building's foundation, walls, roof, floor, beams and columns into a complete building needs to ensure the integrity of the architecture, and each link is interconnected to form a complete narrative medium.

The narrative of architectural construction is reflected in its structure or construction, which is the expression of local experience based on craftsmanship and technology, and the construction itself is intuitively perceived as a narrative text, which expresses the regional and cultural characteristics of the building through the selection of the design and detailed treatment of the construction and tells the connotation of its semantics. Architecture includes walls, floor slabs, roofs and other basic enclosing structures, emphasizing the surface presentation or practice techniques. Roofing forms such as flat, sloped, and gabled roofs, through the operation of elevation and eaves; wall construction forms such as masonry, curtain walls, and so on. Through the logical organization of the structure, the operation of the wall such as window and hole opening, and the operation of the floor such as hollow and staggered floors, there is a different expression of the form of the building.

3. Experiments on Automatic Image Semantic Labeling Based on DAE-GAN

3.1. Comparative Experiments on Automatic Labeling Algorithms

3.1.1. Experimental setup

In order to facilitate the comparison of the advantages and disadvantages of automatic annotation algorithms, this paper adopts the self-constructed experimental data, which contains 5000 movie clip frame images intercepted from 100 representative Chinese and foreign movies, covering a wide range of housing types such as apartments, villas, shantytowns, traditional courtyards, and so on. It is divided into training set and test set according to the ratio of 4:1. Each image is manually labeled with multiple tags by three annotators independently, and a total of 385 valid keywords are finally integrated, which cover multiple dimensions such as spatial structure and material.

In order to quantitatively evaluate the image semantic annotation performance of DAE-GAN algorithm in this paper, all the keywords contained in the training set are utilized as the query to retrieve the images, and then the average checking rate Recall and checking rate Precision are calculated. During the retrieval process, if the keyword of the image annotation result contains the query keyword, the image is returned as the query result. The manual annotation of images is used as a criterion to evaluate the relevance of the query. The check accuracy rate is the number of correctly retrieved images divided by the number of all relevant images. The accuracy rate is the number of correctly retrieved images divided by the number of images returned from the search.

3.1.2. Comparative experimental analysis

Four popular annotation models, Co-occurrence, Translation, L-VM and FACMRM, are selected to be compared with the dynamic hierarchical semantic-aware DAE-GAN model based on this paper, and 20 frequently occurring recognition keywords are taken as the object of the study, and Fig. 3 and Fig. 4 show the recognition effects of the five models on the dataset for these 20 keywords, respectively.

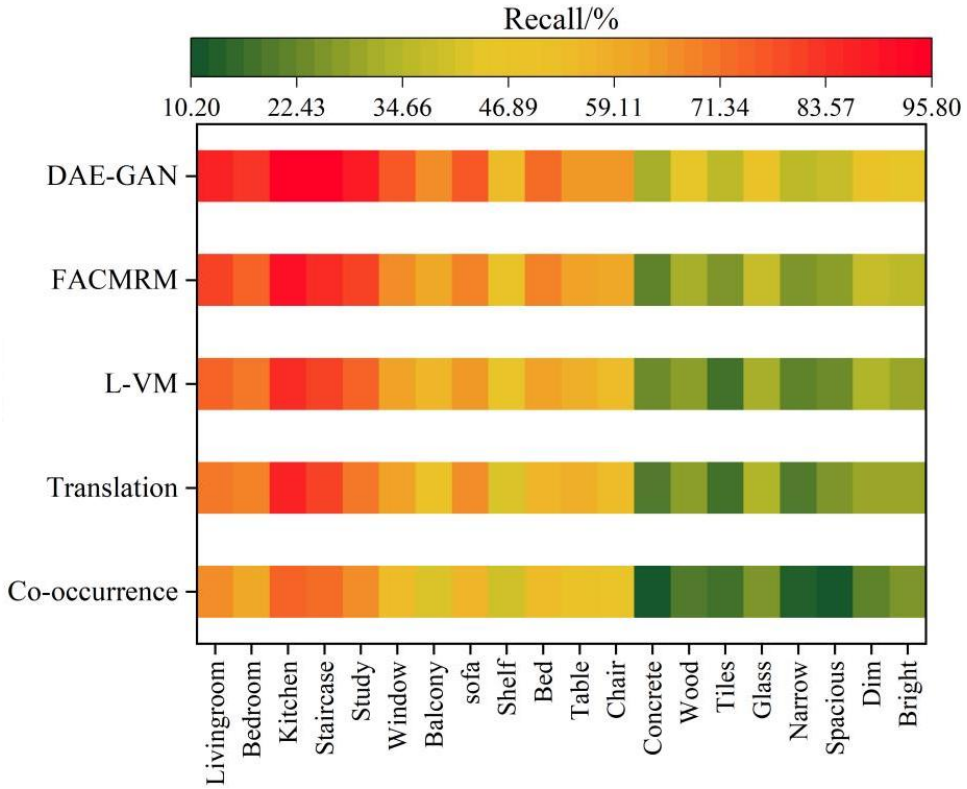

Figure 3. The recognition Recall of 20 keywords of 5 models on the dataset.

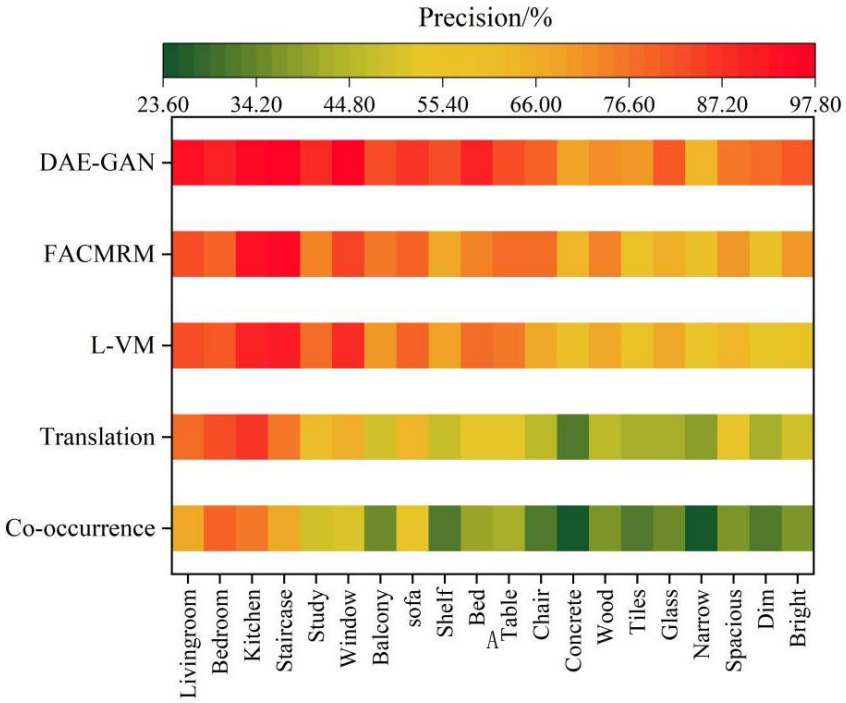

Figure 4. The recognition Precision of 20 keywords of 5 models on the dataset.

DAE-GAN has shown obvious advantages in the top 20 keyword recognition tasks with the highest occurrence frequency. Specifically, for spatial functional keywords with clear structures such as "kitchen" and "staircase", the recognition foundation of all models is quite good. The dynamic hierarchical semantic perception model based in this paper can maintain a Recall rate of over 93%, and the precision rate even reaches an astonishing 94% to 97%. This shows that DAE-GAN not only recognizes the key items clearly in the image, but also finds them with great accuracy. In contrast, for material attribute keywords such as "concrete", "wood", and "ceramic tiles", traditional models like Co-occurrence and Translation seem somewhat inadequate, with recall rates generally lower than 30%. Although the performance of DAE-GAN in these difficulties is not very high either, However, the recall rate is almost twice that of the Co-occurrence, confirming the model's unique advantages in multi-granularity semantic perception, especially in fine-grained feature capture.

For the atmosphere keywords of residential Spaces such as "narrow", "spacious" and "dim", given that these words are completely abstract in meaning, understanding them requires more context and overall perception. Therefore, the recognition recall and accuracy of this model on these adjectives are much higher than other models, for example, for the recognition of “dim”, the detection rate is 47.01%, and the detection rate is 77.33%.

Keywords that have a check-perfect rate greater than 40% and a check-accuracy rate greater than 60% are referred to as good keywords. Good keywords have better annotation effect on the whole test image set. In order to compare the advantages and disadvantages of the algorithms more comprehensively, the five annotation models are derived to calculate the number of good keywords as well as the average checking completeness and checking accuracy of the 385 keywords, respectively. In general, as the check accuracy rate increases, the check precision rate decreases accordingly. In order to consider the two together, F1 metrics are also introduced for the comprehensive evaluation of the models. The experimental results of the comparison of good keywords among the labeled models are shown in Table 1. In order to show more clearly the relationship between the number of good keywords recognized by the model and the comprehensive index F1, the bubble diagrams of the five models are also plotted as shown in Fig. 5, with the vertical coordinate indicating the number of good keywords recognized, and the size of the bubbles indicating the F1 index.

Table 1. The comparative experimental results of good keywords.

	Number of good keywords	Recall	Precision	F1
Co-occurrence	14	31.13%	44.20%	37.67%
Translation	39	48.21%	55.33%	51.77%

L-VM	55	53.99%	71.45%	62.72%
FACMRM	76	56.27%	74.05%	65.16%
DAE-GAN	179	68.28%	82.55%	75.42%

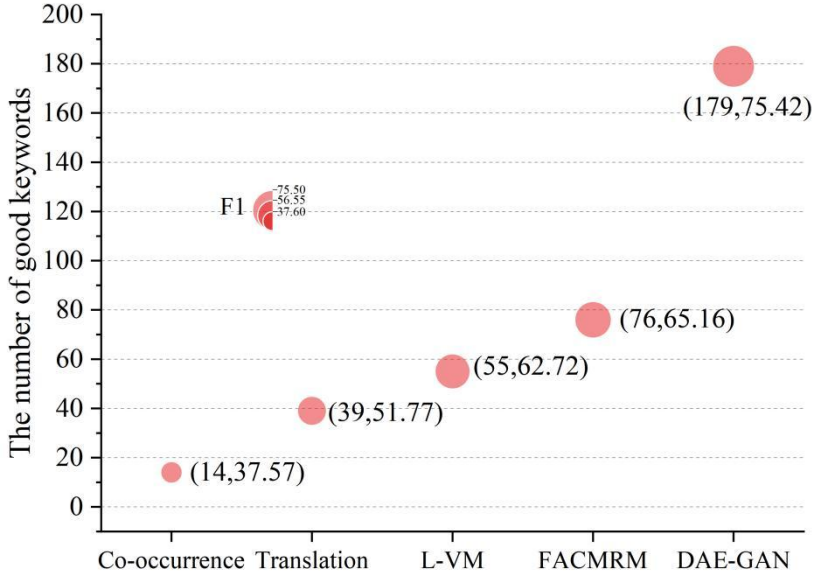

Figure 5. Key word count - F1 bubble chart.

The DAE-GAN model has the highest and largest bubbles, indicating that it is optimal in both number and quality of recognitions. Out of 385 keywords in the self-constructed dataset, DAE-GAN recognizes 179 good keywords in one go, which is 2.36 times more than the 76 of the second-place FACMRM model. This means that DAE-GAN has greatly expanded the scope of annotated semantics, and many abstract words that were previously beyond the power of the model are included in the scope of DAE-GAN model analysis.DAE-GAN's Recall and Precision reached 68.28% and 82.55%, respectively, with an F1 score of 75.42%, which once again verifies that DAE-GAN is able to capture most of the relevant information about the architectural images of the movie while ensuring the high quality of the returned results.

3.2. Discussion of weighting parameters

The visual vocabulary distribution and text vocabulary distribution weights of each image are different, 10 images with different topics are randomly selected in the dataset, and 10 sample images are tested to show the effect of the weighting parameter τ on the annotation performance, and the specific results are shown in Table 2. In order to show the relationship between the information entropy of each image, the weighting parameter, and the final annotation accuracy rate more clearly, the article also draws a ternary scatterplot of the three As shown in Fig. 6, the three sides of the triangle underneath represent the normalized visual vocabulary distribution and the information entropy of the text vocabulary and its weight parameter respectively, and the vertical axis is the labeling accuracy rate.

Table 2. The annotation performance under different weight parameters for 10 images.

Image	Information entropy		Weight parameters	Annotation precision
	Visual vocabulary distribution	Textual vocabulary distribution		
1	3.2	4.5	0.52	67.86%
2	2.8	3.9	0.58	70.65%
3	5.4	4.1	0.19	73.62%
4	0.6	1.2	0.91	71.65%
5	1.85	2.3	0.78	62.52%
6	4.2	3.8	0.31	58.49%
7	2.5	3.1	0.63	53.22%
8	3.9	4.6	0.42	76.98%

9	1.2	2.1	0.85	65.81%
10	4.8	3.4	0.25	67.86%

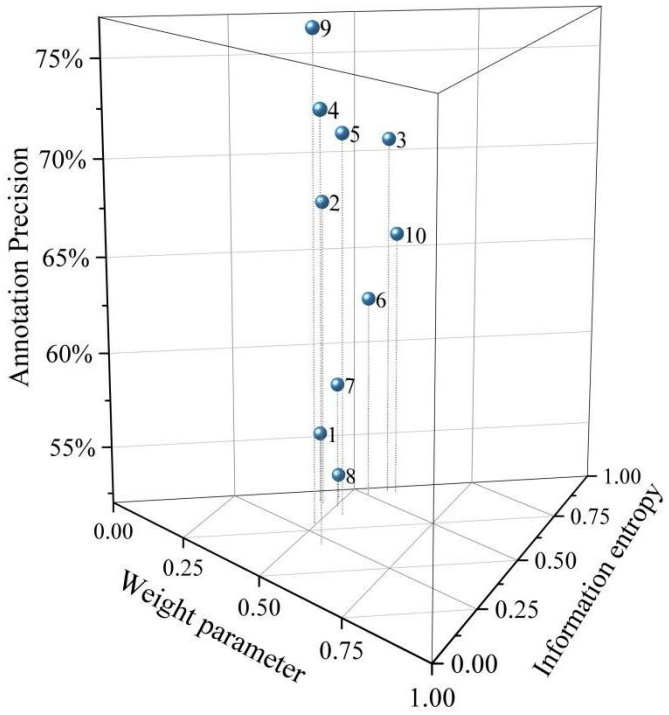

Figure 6. Ternary scatter plot of Information entropy-Weight parameters-Precision.

A large number of experiments with the DAE-GAN labeling model on the dataset show that the semantic annotation performance of the images is better when the information entropy of the visual vocabulary distribution is less than 2, which suggests that the visual modality component in the fused distribution plays a major role in the image semantic learning process. This is verified in images 4, 5, and 9, whose information entropy H of visual vocabulary distribution is 0.60, 1.85, and 1.20, respectively, and the model obtains high weight parameter τ, which is 0.91, 0.78, and 0.85, respectively, while the labeling accuracy also reaches a high level, which is 73.62%, 71.65%, and 76.98%, respectively. For images with concise and clear visual content, the model can effectively rely on the visual modal data, thus achieving excellent labeling performance.

If the information entropy is greater than 4 and the performance of the annotation model is still better, the weight of the visual modal component in the fusion distribution is lower and the textual modal component plays a larger role. For example, images 3, 6, and 10, whose H is 5.4, 4.2, and 4.8, respectively, the weight parameter τ of the model is correspondingly lower, which is 0.19, 0.31, and 0.25, respectively, and their annotation accuracies are 70.65%, 62.52%, and 65.81%, respectively.

When the value of information entropy is between 2 and 4, the labeling performance is not very satisfactory. For example, images 1, 2, 7, and 8 have visual vocabulary distribution information entropy H between 2.5 and 3.9. Their τ values are distributed in the interval of 0.42 to 0.63, and their labeling accuracies are 55.65%, 67.86%, 58.49%, and 53.22%, respectively, which are the lowest among all samples. It is because images with information entropy of visual vocabulary distribution between 2 and 4 have strong content complexity. Therefore, it is difficult to determine the contribution of each modal data by simple weight taking, and thus difficult to learn the exact semantics contained in each image.

4. Influence of the spatial layout of housing on the presentation of characters' internal conflicts and social contexts

Chapter 3 verifies the superiority of the DAE-GAN-based model in the experiments of automatic image semantic annotation, followed by an empirical examination of the specific impact of the spatial layout of housing on the narrative effect.

The validity of the movie materials used in the experiment is first examined to ensure that they accurately represent the four spatial types of spaciousness, crampedness, regularity, and complexity. Subsequently, descriptive statistics and analysis of variance (ANOVA) were conducted on the recovered

questionnaire data to reveal the differences in the ability of different spatial types in triggering the characters' inner conflicts and social background presentation.

4.1. Material validity analysis

The material was derived from 238 films that have been more widely popular since 2000 as a sample for analysis. In order to judge the validity of the material, 50 university students (25 men and 25 women) were invited to score the attributes of the type of housing spatial layout in the movie material on a 7-point scale, first judging to which type of spatial layout the housing spatial layout in the movie belongs, and then rating the space on that type. The data obtained were statistically analyzed to produce the results of the validity test. The results of the validity test of the housing space layout type manipulation material are shown in Figure 7.

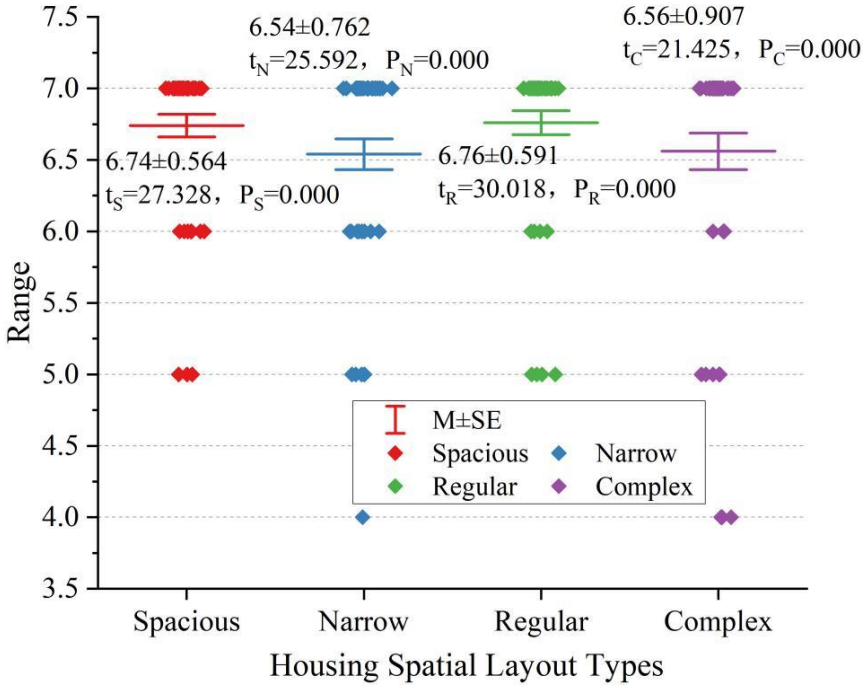

Figure 7. Validation of the Effectiveness for Housing Spatial Layout Types.

By using a one-sample t-test (with a test value of 4 at the midpoint of the 7-point scale) on the ratings of each character type, it was found that Spacious ($t=28.328$, $p<0.01$); Cramped ($t=25.592$, $p<0.01$); Conventional ($t=30.018$, $p<0.01$); and Complex ($t=21.425$, $p<0.01$). This shows that the subjects did not misunderstand the spatial layout of the housing in the material. Therefore the material is valid.

4.2. Descriptive and ANOVA results analysis

The independent variables in this experiment were four different spatial layouts of housing (spacious, cramped, conventional and complex types) and the dependent variables were both the inner conflict of the characters and the social background presentation. The degree of embodiment of the character's inner conflict and social context presentation was quantified according to a 5-point Likert scale (1-no embodiment, 2-slightly embodied, 3-average embodiment, 4-embodied; 5-very embodied). Two hundred subjects were recruited for the questionnaire.

4.2.1. Descriptive statistics results

The 200 subjects' ratings of the characters' internal conflict and social context presentation in different residential space types are shown in Figure 8.

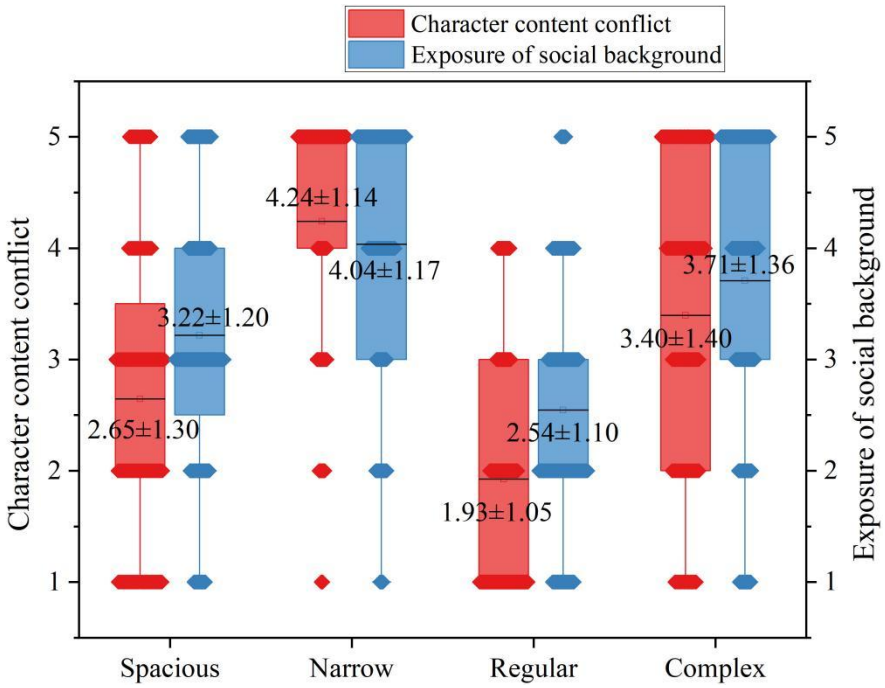

Figure 8. Descriptive analysis of different types of residential spaces.

In terms of the characters' inner conflicts, the 200 subjects rated the cramped space as 4.24±1.14, indicating that the subjects generally believed that a cramped and crowded living environment was the most likely to trigger character conflicts, and the average score for the complex space was 3.40, whose complexity was also considered to be easy to reflect the characters' inner conflicts; in contrast, the conventional space had the lowest character conflict score, only 1.93±1.05, indicating that people generally believed that ordinary space was less likely to carry intense conflicts in the narrative; and the spacious space was in the middle of the range, with an average score of 2.65±1.30. 1.05, suggesting that ordinary spaces are generally perceived as less likely to carry intense conflicts in narratives; while spacious spaces are at an intermediate level, with an average score of 2.65±1.30.

In the dimension of social context presentation, cramped-type space also received the highest score of 4.04±1.17, probably because cramped space can easily reflect social stratification, lack of resources and other problems. Complex spaces also have a higher degree of reflection of social context, with a score of 3.71±1.36. The scores of spacious and conventional spaces in this category are 3.22 and 2.55 respectively, and although these two more conventional spatial layouts can also convey social information, their impact is not as strong as the first two. Overall, the unconventionality of the spatial layout (cramped, complex) had a greater impact than the conventionality (spacious, regular) in reinforcing character conflict and presenting the social context.

4.2.2. Analysis of variance

Using different housing space types as independent variables, ANOVA was conducted to analyze the characters' inner conflict and social background presentation, respectively, and the results are shown in Table 3.

Table 3. Analysis of variance on the variance of spatial types.

Dependent variable	Independent variable	Sum of Squares	Mean Square	F	P
Character-content conflict	Spacious	5.217	5.217	2.894	0.312
	Narrow	128.635	128.635	71.352	0.000
	Regular	22.894	22.894	12.701	0.000
	Complex	61.328	61.328	34.025	0.000
Social background presentation	Spacious	7.352	7.352	3.127	0.085
	Narrow	95.417	95.417	40.583	0.000
	Regular	9.834	9.834	4.182	0.104
	Complex	18.926	18.926	8.049	0.002

The analysis of variance in Table 3 shows that cramped space (F=71.352, p<0.001), conventional space (F=12.701, p<0.001), and complexity space (F=34.025, p<0.001) all reflect significant effects in the dimension of characters' inner conflicts. Only spacious space failed the test of significance, F=2.894, p=0.312, which suggests that open spatial layout does not have a significant impact in stimulating character conflict. Turning to the dimension of social context presentation, cramped-type space still maintains the strongest explanatory power (F=40.583, p<0.001), suggesting that crowded spatial layout serves as a visual representation that highlights social problems. The complex type of space also demonstrated clear statistical significance (F=8.049, p=0.002), suggesting that complex living environments are equally capable of conveying social messages. Both spacious (F=3.127, p=0.085) and regular (F=4.182, p=0.104) spaces failed the test of significance, suggesting that the open and ordinary spatial layout is not directly linked to the presentation of the social context, and that positive layouts, on the contrary, have a lesser impact on the characters' internal conflicts and the presentation of the social context.

5. Conclusion

The study confirms that the spatial layout of housing does have a quantifiable influence mechanism in movie narratives through the construction and experimental validation of the DAE-GAN model.

In the task of automatic image annotation, the DAE-GAN model is very effective in recognizing the keywords of material attributes. Taking the keyword “concrete” as an example, the traditional Co-occurrence model's detection rate is only 10.22%, while DAE-GAN reaches 32.46%. For abstract ambient words, DAE-GAN's search rate for “dim” reaches 47.01%, with an accuracy rate of 77.33%, which is much higher than that of other comparative models. This shows that the multi-granular semantic perception mechanism of DAE-GAN model can effectively solve the problem of recognizing fine-grained features and abstract concepts.

The weighting parameter experiment further verifies that the DAE-GAN model can dynamically adjust the multimodal fusion strategy. When the information entropy of visual vocabulary distribution is lower than 2, the model relies on visual modality, and the weighting parameter reaches 0.78-0.91, with the labeling accuracy over 71%; while the information entropy is higher than 4, the model shifts to the dominance of textual modality, and the weighting parameter decreases to 0.19-0.31, when it can still maintain the accuracy rate of 62.52%-70.65%.

The type of housing layout significantly affects narrative expression. Cramped type of space scored the highest on character conflict score and social context presentation with 4.24 and 4.04, respectively. Analysis of variance (ANOVA) showed a significant effect (F=71.352/40.583, p<0.001).

About the Author

Xuanyu Yu (born March 1990), male, Han Chinese, native of Xi 'an, Shaanxi Province, holds a master's degree. He is currently a lecturer at Xi’an Mingde Institute of Technology, specializing in the study of realistic visual narrative strategies.

ReferencesGibbs, J. (2015). The life of mise-en-scène: visual style and British film criticism, 1946–78. In The life of mise-en-scène. Manchester University Press.

2. Tarvainen, J., Sjöberg, M., Westman, S., Laaksonen, J., & Oittinen, P. (2014). Content-based prediction of movie style, aesthetics, and affect: Data set and baseline experiments. IEEE Transactions on Multimedia, 16(8), 2085-2098.
3. Isik, A. I., & Vessel, E. A. (2021). From visual perception to aesthetic appeal: Brain responses to aesthetically appealing natural landscape movies. Frontiers in Human Neuroscience, 15, 676032.
4. Silvia, P. J., & Berg, C. (2011). Finding movies interesting: How appraisals and expertise influence the aesthetic experience of film. Empirical Studies of the Arts, 29(1), 73-88.
5. Cook, R. F. (2011). Correspondences in visual imaging and spatial orientation in dreaming and film viewing. Dreaming, 21(2), 89.
6. Hallam, J. (2010). Film, space and place: researching a city in film. New review of film and television studies, 8(3), 277-296.
7. Stolorow, R. D. (2013). Toward a pure psychology of inner conflict. In Progress in Self Psychology, V. 1 (pp. 193-201). Routledge.
8. Hart, T. (2010). The secret spiritual world of children: The breakthrough discovery that profoundly alters our conventional view of children's mystical experiences. New World Library.
9. Blake, W. (2014). Hegel: Conflict and Order. Tragedy and Theory: The Problem of Conflict Since Aristotle, 23.
10. Dastres, R., & Soori, M. (2021). Advanced image processing systems. International Journal of Imagining and Robotics, 21(1), 27-44.

11. Razzak, M. I., Naz, S., & Zaib, A. (2017). Deep learning for medical image processing: Overview, challenges and the future. Classification in BioApps: Automation of decision making, 323-350.
12. McQuin, C., Goodman, A., Chernyshev, V., Kamentsky, L., Cimini, B. A., Karhohs, K. W., ... & Carpenter, A. E. (2018). CellProfiler 3.0: Next-generation image processing for biology. PLoS biology, 16(7), e2005970.
13. Zheng, W. (2023). Current Technologies and Applications of Digital Image Processing. Journal of Biomedical and Sustainable Healthcare Applications, 3(1), 013-023.
14. Cheng, Y., & Wang, Y. (2022). Movie Special Effects Processing Based on Computer Imaging Technology. Scientific Programming, 2022(1), 1384589.
15. Kumar, N., Kumar, K., & Kumar, A. (2022, February). Application of internet of things in image processing. In 2022 IEEE Delhi Section Conference (DELCON) (pp. 1-5). IEEE.
16. Jordan, G., Petrik, A., De Vivo, B., Albanese, S., Demetriades, A., Sadeghi, M., & GEMAS Project Team. (2018). GEMAS: Spatial analysis of the Ni distribution on a continental-scale using digital image processing techniques on European agricultural soil data. Journal of geochemical exploration, 186, 143-157.
17. Szabó, K. Z., Jordan, G., Petrik, A., Horváth, Á., & Szabó, C. (2017). Spatial analysis of ambient gamma dose equivalent rate data by means of digital image processing techniques. Journal of Environmental Radioactivity, 166, 309-320.