systems

MDPI

Article

Harnessing Multi-Modal Synergy: A Systematic Framework for Disaster Loss Consistency Analysis and Emergency Response

Siqing Shan, Jingyu Su *, Junze Li, Yinong Li and Zhongbao Zhou

School of Economics and Management, Beihang University, Beijing 100191, China; shansiqing@buaa.edu.cn (S.S.); ljzsem@buaa.edu.cn (J.L.); liyinong@buaa.edu.cn (Y.L.); zhongbao@buaa.edu.cn (Z.Z.)
* Correspondence: sujingyu@buaa.edu.cn

Abstract

When a disaster occurs, a large number of social media posts on platforms like Weibo attract public attention with their combination of text and images. However, the consistency between textual descriptions and visual representations varies significantly. Consistent multi-modal data are crucial for helping the public understand the disaster situation and support rescue efforts. This study aims to develop a systematic framework for assessing the consistency of multi-modal disaster-related data on social media. This study explored how the congruence between text and image content affects public engagement and informs strategies for efficient emergency responses. Firstly, the Clip (Contrastive Language-Image Pre-Training) model was used to mine the disaster correlation, loss category, and severity of the images and text. Then, the consistency of image–text pairs was qualitatively analyzed and quantitatively calculated. Finally, the influence of graphic consistency on social concern was discussed. The experimental findings reveal that the consistency of text and image data significantly influences the degree of public concern. When the consistency increases by 1%, the social attention index will increase by about 0.8%. This shows that consistency is a key factor for attracting public attention and promoting the dissemination of information related to important disasters. The proposed framework offers a robust, systematic approach to analyzing disaster loss information consistency. It allows for the efficient extraction of high-consistency data from vast social media data sets, providing governments and emergency response agencies with timely, accurate insights into disaster situations.

Keywords: consistency; data fusion; deep learning; disaster loss information; social media

check for updates

Academic Editor: Ed Pohl

Received: 14 May 2025
Revised: 17 June 2025
Accepted: 19 June 2025
Published: 20 June 2025

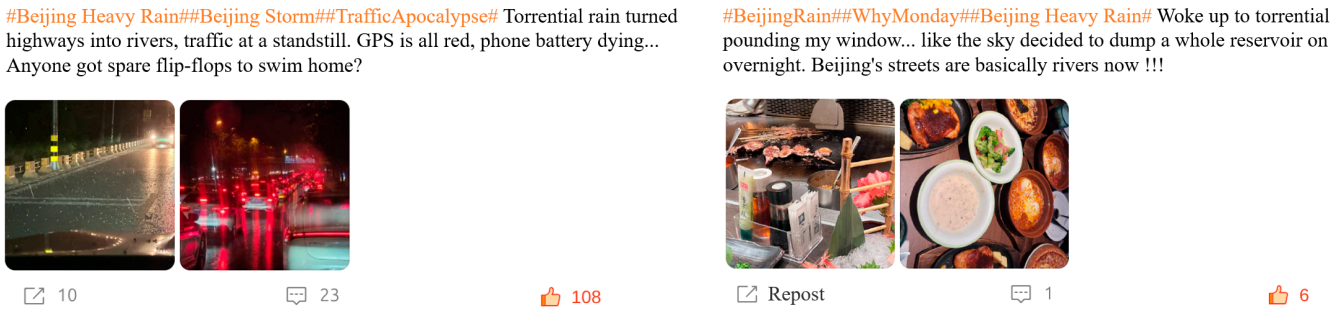
Citation: Shan, S.; Su, J.; Li, J.; Li, Y.; Zhou, Z. Harnessing Multi-Modal Synergy: A Systematic Framework for Disaster Loss Consistency Analysis and Emergency Response. *Systems* **2025**, *13*, 498. https://doi.org/10.3390/systems13070498

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In July 2023, affected by cold and warm air and typhoon Du Surui, heavy rains occurred in Beijing and Hebei. This sudden natural disaster has not only seriously affected the lives of local residents but also caused great damage to the infrastructure. After the disaster, social media became the main channel of information dissemination, and users shared the disaster situation by publishing information, including images and text [1]. Previous studies have shown that social media information with images can better represent the related situation of natural disasters than social media information without images [2]. However, the consistency of image and text information is very important for the accuracy and effectiveness of information. For example, Figure 1 shows two Weibo posts released during the Beijing rainstorm. The picture in Figure 1a is consistent with the information

conveyed by the text information, but there are differences in Figure 1b. This inconsistency of information may affect the public's understanding and response to disaster situations, mislead the public, and thus adversely affect disaster management.



(**a**) consistent information (**b**) inconsistent information.

Figure 1. Weibo examples of disaster loss information.

High graphic consistency means that the information transmitted by images and text is consistent, which helps to improve the credibility and communication effect of the information [3]. On the other hand, the inconsistency between the images and text may lead to misleading information and affect the effectiveness of the emergency response [4]. Although the existing disaster management systems (e.g., VGI and CrowdMap) [5,6] can collect information through crowdsourcing data, the research mostly focuses on single-mode data analysis, ignoring the potential of graphic information fusion, and lacks in-depth research on the consistency of multi-modal data, such as image and text content. At the same time, social concern is closely related to the rescue work. Extensive social concern can provide necessary resources, information, and support for the rescue operation, thus ensuring that the rescue work can be carried out quickly and orderly, effectively helping the affected people and reducing the losses caused by disasters. Previous studies mainly focus on the data level, lacking in-depth discussion on the influence mechanism and mode of consistency of social concern. Therefore, the purpose of this paper is to discuss how to evaluate the consistency of disaster loss images and text information on social media and analyze the impact of this consistency on social concern. By comprehensively analyzing the consistency of text and image data, the efficiency and effectiveness of obtaining information on social media channels are improved, which is more practical for guiding emergency responses.

In order to achieve the above goals, this paper used the CLIP (Contrastive Language-Image Pre-Training) deep learning model, data fusion technology, and the vectorization method to comprehensively use the multi-modal data obtained from the social media platforms. The CLIP model was chosen as the main research tool because of its excellent multi-modal task performance, especially in bridging the gap between visual and text information [7].

The framework of disaster loss information analysis based on image–text consistency proposed in this study has certain universality and can provide a reference for the loss evaluation of other similar disasters. Although the specific characteristics and influencing factors of different disaster events may be different, disaster events usually have common characteristics, such as casualties, infrastructure damage, disaster relief operations, vehicle damage, etc. These characteristics can be classified and quantified by similar methods. Therefore, when analyzing other disaster events, researchers only need to make appropriate adjustments according to the characteristics of specific events and research needs, and then researchers can apply the method of this study to other events. In this way, the

method of this study can better meet the assessment needs of different disaster events and provide more effective support for disaster management and emergency responses.

The practical significance of this study lies in providing a new perspective and tool for disaster management. By analyzing the consistency of multi-modal data, key disaster information can be identified and disseminated more accurately, thus optimizing the disaster information dissemination strategy. At the same time, it provides a new idea for policymakers; that is, by integrating text and image data, disaster losses can be evaluated more comprehensively, thus providing more accurate information support for the disaster response and recovery. This study further discusses how consistency affects the public's attention to disaster information, which provides a basis and theoretical support for the formulation of information dissemination strategies in disaster management. By optimizing the information dissemination strategy, the government and disaster management departments can guide the public to pay attention to disaster events more effectively and improve the public's disaster awareness and coping ability.

The main contributions of this study are as follows: (1) A disaster loss information extraction model is proposed from the perspective of social media so as to obtain information in time. (2) Image mode is introduced to systematically evaluate disaster losses, and a method to calculate the consistency between images and text is proposed, so as to use images and text information at the same time. (3) Explore the impact of the consistency of images and text on social concern so as to facilitate emergency management.

The rest of this paper is organized as follows: Section 2 summarizes the relevant literature, Section 3 introduces the research methods, Section 4 introduces the cases used in this paper and the results and finally summarizes and discusses this study.

2. Related Works

2.1. Application of Social Media Data in Disaster Management

With the development of the Internet and 5G technology, social media has become an important data source in the study of natural disasters [2]. Luo et al., based on machine learning and social network analysis methods, extracted theme, emotion, and network variables from a large-scale text data set related to COVID-19 and further analyzed the propagation characteristics and evolution law of network public opinion in crisis events [8]. Shan et al. established a physical damage dictionary covering six damage categories and extracted related topics and vocabulary from the disaster corpus through LDA (Latent Dirichlet Allocation) [9]. Hao and Wang further divided keywords into objects and description lists, which were used to identify social media information with disaster loss information [10]. Wu et al., through the case study of super typhoon Lichima in 2019, explored the algorithm by using Weibo data to mine the accurate location and spatial distribution of typhoon victims [11]. Shan et al. built a dynamic real-time disaster loss assessment model based on social media data and verified the model with the actual data of the Jiuzhaigou earthquake [12]. Xing et al. combined social media data and mobile phone signal data to evaluate the disaster impact of the Jiuzhaigou earthquake [13]. Xie et al. proposed a multi-label classification framework based on supervised comparative learning to deal with disaster information on social media [14]. Zhang and Ma developed a hybrid model enhanced by subject knowledge, which was applied to the emotion analysis of sudden natural disasters and showed excellent performance [15]. Li et al. trained a YOLO-based model to detect human parts in social media images so as to extract flood depth information and broaden the sources of flood severity information [16].

Although the existing research has made some progress in social media data mining and disaster event analysis, it still focuses on the text data level, and the research on

multi-modal data fusion is insufficient. Multi-modal data fusion can provide more comprehensive and accurate disaster information, but the related research is still limited.

2.2. Research on Data Fusion

In the era of information explosion, a large amount of data with different modes is created, collected, and processed, and there are more and more studies on data fusion. Chen et al. introduced a visual–text relevance-based adaptive method, fusing features for improved joint emotion analysis. The experimental results show that the fused features are more accurate than the single-mode classification, and the contribution of each single-mode is also obviously different in emotional expression [1]. Liu et al. proposed a focused attention model for cross-modal feature association [17]. Bryan-Smith et al. created a flood forecasting model using Transformer-based multi-modal inputs for severity assessment. In the context of disaster assessment [18], Cheng et al. built EADBM for earthquake-damaged building recognition [19]. Ge et al. used CycleGAN and incremental learning for efficient post-disaster building damage identification [20]. Wang et al. developed QuakeCityNet for UAV image building damage assessment [21], while Tasci et al. proposed the InCR network using remote sensing images for this task [22]. Islam et al. designed a UAV image classification method for flood area detection, in terms of vulnerability and emotion analysis [23]. Xing et al. pioneered a framework combining high-resolution remote sensing and street-view images to assess urban building flood vulnerability [24]. For multi-modal sentiment analysis, Huang et al. designed the text-centric fusion network TeFNA [25]. Pandey and Vishwakarma proposed VABDC-Net using visual–text attention [26]. Zeng et al. employed GCN and ensemble learning for fine-grained negative emotion recognition in COVID-19 data [27]. Lin et al. introduced the PS-Mixer, a polar and intensity vector mixed model based on the MLP-Mixer [28].

It can be seen that the era of big data provides more available data sources and richer application scenarios for the technical means of data fusion, and some research in the field of sentiment analysis has been carried out using multi-modal data. In the field of emergency management and disaster management, researchers have gradually begun to adopt the mode of multi-modal fusion to carry out related research, but in the experiment of data fusion, the number of studies considering the information consistency between image modality and text modality is still small.

2.3. Image–Text Consistency Evaluation

Image–text consistency assessment plays an important role in multi-modal data processing, especially in disaster management and humanitarian response scenarios. The existing research methods can be divided into two categories, according to semantic alignment: global alignment and local alignment. The global alignment method aims to build a common semantic space so that images and text embeddings can be mapped to each other in this space. By designing specific loss functions, these methods aim to make the representations of matched images and text closer in the common semantic space, thus enhancing their consistency [29]. Inspired by the wide application of the local correlation method in cross-modal tasks, the concept of local alignment is introduced into the matching process of images and text, and researchers are committed to learning the fine-grained correspondence between image areas and text words. This method aims to establish a more accurate and detailed cross-modal association by identifying and matching the specific visual elements in the image and the specific vocabulary in the text. For example, Liu et al. proposed a focus attention mechanism to give more weight to semantically related segments [30], while Diao et al. employed a similarity attention filter to deduce semantic links. While powerful, these methods often overlook the long-tail effect of rare, low-frequency content,

hindering comprehensive semantic exploration [31]. In applied disaster contexts, Rizk et al. created a multi-modal framework for energy-constrained devices analyzing social media tweets/images [32]. Mouzannar et al. developed a multi-modal DL model (Inception for images and CNN for text) to identify injury-related information [33]. Kumar et al. classified disaster information using LSTM and VGG16 [34]. Madichetty et al. utilized BERT and DenseNet to analyze tweets/images [35]. Beyond disaster management, image–text alignment impacts social media engagement. Ceylan et al. quantified Yelp text–image similarity via ML, finding that high consistency significantly increases comment usefulness by enhancing information processing ease [36]. Shin et al. demonstrated that integrating visual and textual features boosts likes/reposts, proving that both modalities significantly impact social media reach [37].

Graphic consistency has important application value in social media big data analysis, especially in disaster management and emergency responses. However, in the current research, there is still room for mining and analyzing the graphic consistency of the loss information of the massive multi-modal data of social media, and there is a research gap in the mechanism of consistency's influence on social concern.

2.4. Implications of This Study

Although the existing research has promoted the social media analysis of disasters, there is still a big gap in the comprehensive multi-modal field, image–text consistency evaluation, and the impact of consistency. This study directly solved these gaps. This paper proposes a comprehensive integration framework, which combines social media data fusion with image–text consistency evaluation. Specifically, the contribution of this paper lies in the following:

1. Build a multi-modal database: Establish a framework for effectively fusing text and image data so as to extract disaster loss information more richly and accurately.
2. Promote consistency measurement: Develop a qualitative and quantitative method and use accurate indicators and algorithms to robustly quantify image–text consistency in a complex disaster event environment.
3. Clarify the relationship between consistency and social concern: Through an empirical case study, systematically analyze how the change in graphic consistency affects social concern.

Through these innovations, this research not only improves the technical ability of social media analysis related to disasters but also provides important empirical insights on the role of information consistency in shaping public perception and response, which provides valuable guidance for future research and disaster management practice.

3. Methods

This section introduces the framework of the image–text consistency analysis model built in this study and how each step was realized.

3.1. Integral Model Framework

The framework of the image–text consistency analysis of disaster loss information is shown in Figure 2. Firstly, we collected data. With the keywords "Beijing rainstorm" and "Hebei rainstorm", this study collected the relevant Weibo posts from the Weibo platform, including Weibo text, images, and related Weibo features and user features. After cleaning and sorting out the data, this study constructed a multi-modal data set of Weibo disaster event images and text. Subsequently, this study obtained the classic CrisisMMD (Crisis Multi-Modal Dataset [38,39]), selected the flood disaster-related data, and labeled the downstream classification tasks accordingly, then trained the CLIP (Contrastive

Language-Image Pre-Training [7]) deep learning model and constructed three image classification models: disaster correlation, disaster loss category, and disaster loss severity. In terms of text data, this paper constructed two Chinese dictionaries, text loss category and text loss severity, by combining preset seed text and the way of expanding similar text in the Word2vec [40] word embedding model.

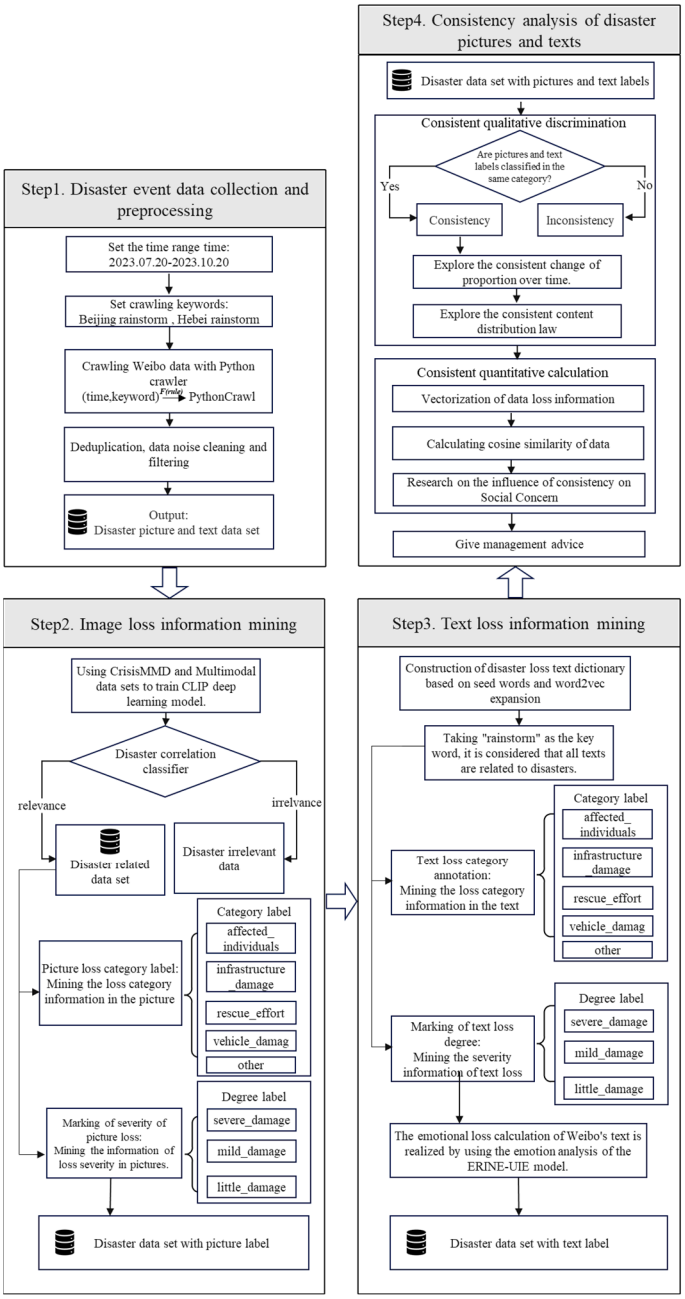


Figure 2. Framework for the consistent analysis of disaster loss information.

Then, this paper classified the loss of image data on Weibo, used the above-mentioned image loss classification model and text loss dictionary to distinguish and calculate the loss categories and severity of the image data and text data, and obtained the quantitative distribution of different loss categories and the severity of the Weibo images and text related to the Beijing rainstorm and the Hebei rainstorm, as well as the fluctuation trend of the disaster loss categories and severity categories over time.

Finally, based on data fusion, this paper evaluated the image consistency of the loss information of rainstorm and flood disasters contained in Weibo data and explored the

impact of image consistency on social concern. Firstly, according to the experimental classification results of the disaster loss of image data and text data, each category was qualitatively discriminated, and the image–text data pair with a consistent disaster loss category and disaster loss severity was obtained. Further, this paper used the vectorization method to calculate the cosine similarity between the image and text vectors and realized the qualitative calculation of image–text consistency. Finally, this paper studied the influence mechanism of image–text consistency on social concern.

3.2. Step 2: Image Loss Information Mining

CLIP is a multi-modal learning model proposed by OpenAI in 2021 [7]. The CLIP model was chosen because it has unique advantages in integrating visual and text information and can seamlessly understand and relate these two modes. Different from the traditional single-modal model, the pre-training framework of CLIP is based on a large-scale image–text corpus, which provides a strong foundation for research tasks and allows us to fine-tune with the least extra data to ensure that the model can effectively adapt to specific data sets. Compared with other models, CLIP is pre-trained on a huge image–text corpus, which enables it to be well summarized across fields and tasks, especially in the field of disaster emergency management. In addition, CLIP's architecture supports efficient fine-tuning and adaptation to specific data sets, making it a universal choice. Compared with traditional CNN and RNN, CLIP provides a more integrated and efficient method to process visual and text information. This integration is very important for this research because it enables us to take advantage of the complementary advantages of the two modes to obtain more accurate and comprehensive results. In order to ensure the rationality of the selected hyperparameters, this paper conducted extensive comparative experiments and finally selected a group of parameters with the best performance through tests of various parameter combinations. The CLIP deep learning model was built and trained based on the paddlepaddle platform. The Python version was 3.10 in the experimental environment, and the deep learning framework was Paddle2.5. The experiments in this paper were all conducted on a Tesla V100 GPU with a 16G memory. In this paper, the CLIP_vit_base_patch16_224 model was constructed based on the paddlepaddle platform. The model used Vision Transformer (ViT) architecture as a part of the image encoder and adopted the patch size of 16 $\times$ 16 pixels and the input image size of 224 $\times$ 224 pixels. The model used AdamWDL(Adam improved variant of Adam optimizer, also known as AdamW, which treats the weight attenuation and gradient update separately) as the optimizer of the model objective function, the model loss function was the cross-entropy loss function (CELoss), and the number of warm-up steps (WarmUp Epochs, rounds of training at a small learning rate in the initial training stage) was 5. The training set and verification set were divided by 5-fold cross-validation, with 80% of the data set as the training set and 20% as the verification set.

In the construction of the model training data set, this paper selected the Srilanka floods image data in the CrisisMMD data set to construct the CLIP model training data set.

The CrisisMMD data set is a multi-modal Twitter data set, which contains thousands of manually marked tweets and images collected during seven serious natural disasters around the world in 2017. The types of natural disasters in the data set include earthquakes, hurricanes, floods, and other disaster categories. There are 1022 images in the srilanka_floods image data set in the CrisisMMD data set, and the images in the data set all have corresponding research tasks for the three image classification tasks in this work, namely, image disaster correlation, image loss category, and image loss severity. In the experiment of the correlation of image disasters and the severity of image loss, the classification category of this experiment was consistent with the labeling results of the data

set, while in the experiment of the image loss category (called the humanitarian category in the data set), considering the relevant situation of urban rainstorm and flood disasters in this paper, this paper unified three categories related to human beings in the data set, including "injured or dead human beings", "affected human beings", and "missing or found human beings".

3.3. Step 3: Text Loss Information Mining

This section introduces the construction of the text loss information classification model. Specifically, it includes the construction of the disaster loss categories classification model and the disaster loss severity classification model. The model construction in this part was carried out by the dictionary method.

3.3.1. Classification Algorithm of Text Loss Categories

Firstly, this paper constructed five dictionaries of storm and flood disaster losses, which were the same as those of images, including "affected individuals", "infrastructure and utility damage", "vehicle damage", "rescue volunteer or donation effort", and "other relevant information". The dictionary of each category contained a series of texts related to the disaster loss category.

In the process of constructing a dictionary, this paper adopted a method of combining preset seed texts, using the Word2vec model to calculate cosine similarity to expand similar texts. Word2vec is a word vector generation model widely used in the field of natural language processing, which can map words into a low-dimensional vector space, thus capturing the semantic similarity between words [40]. Word2vec's efficiency and robustness in generating word embeddings made it an ideal choice for the dictionary expansion task. While other models like GloVe and FastText have their merits, Word2vec's simplicity and speed were particularly advantageous given the scale of our text data. Moreover, Word2vec's cosine similarity measure provided a reliable and interpretable way to identify semantically similar words, which was essential for the dictionary construction process. Firstly, this paper selected some seed texts to form the initial dictionary of each category. These seed texts were selected from some dictionaries in the field of emergency management, combined with the understanding and analysis of the loss categories that may be involved in Weibo text data and the professional knowledge of each loss category. Next, this paper used the word2vec word embedding model to further expand the dictionary. Using the word2vec similarity of the initial selected seed texts, texts with high similarity to the initial seed texts were calculated. If the cosine similarity between these texts and the seed texts exceeded the preset threshold, it was considered that the word was related to the seed texts, so it was added to the dictionary. Finally, the texts in each loss category in the dictionary were determined through manual screening and review. Finally, a Chinese dictionary of disaster loss categories (Appendix A Table A1) was formed.

The pseudo-code of the algorithm for classifying the disaster loss categories of text data is shown in Appendix A. In this paper, it was considered that a Weibo text may contain a variety of loss category information, so when the algorithm was set, it was set as a multi-classification rule; that is, a Weibo text could be classified into multiple loss categories. When the number of disaster category characteristic texts contained in a text was greater than or equal to two, the text was considered to be classified into the corresponding category.

3.3.2. Classification Algorithm of Text Loss Severity

In the work of quantifying the degree of disaster loss in Weibo texts, this paper constructed a dictionary of disaster loss severity. The dictionary contained Chinese texts related to disaster losses and their corresponding loss scores, and the loss scores increased

from 1 to 10, indicating that the severity information contained in the texts in the dictionary increased. In the process of dictionary construction, this paper combined the preset initial seed texts, the Word2vec word vector model to calculate the cosine similarity and expand the similar texts of seed texts, as well as the method of manual verification and discrimination.

Firstly, this paper selected some seed texts to form the initial part of the dictionary of loss severity. The determination of seed texts referred to some texts in the loss dictionary of relevant research papers in the field of emergency management and combined them with the understanding and analysis of the losses that may be involved in the Weibo text data, as well as the professional knowledge of disaster losses. Each seed word was given an initial loss score, which was set according to its severity in describing disaster losses. Subsequently, this paper used the word2vec word embedding model to further expand the dictionary. By calculating the word2vec similarity of the initial seed texts, some texts with high similarity to the seed texts were found. If the cosine similarity between these texts and the seed texts exceeded the preset threshold, they were considered to be related to the seed texts and added to the dictionary. New texts were given loss scores corresponding to their similarity, and these scores were adjusted and determined by comparing the initial seed texts. Finally, through manual screening and auditing, the accuracy and rationality of each word and its loss score in the dictionary were ensured. After this series of steps, this paper finally formed a Chinese dictionary for quantifying and calculating the degree of disaster losses in social media texts (Appendix A Table A2).

When designing the classification algorithm of text loss severity, this paper referred to the definition and classification standard of loss severity in the field of disaster management. This paper combined the following two factors to determine these thresholds. On the one hand, this paper had in-depth discussions with experts in the field of disaster management, and according to their experience in judging the severity of disaster losses, this paper divided the losses into three main levels: serious losses, moderate losses, and slight or no losses. On the other hand, this paper made a preliminary analysis of a large amount of Weibo text data and counted the frequency and distribution of loss-related words in the text. By analyzing these data, it was found that the score distribution of loss severity presents a certain law. For example, a text with a high score usually contains many lost words with a high score, while a text with a low score mainly contains some words with a low score. Based on these analysis results, this paper chose 15 as the threshold of "serious loss", 5 as the threshold of "moderate loss", and 0 as the threshold of "slight or no loss". At the same time, in order to ensure the validity and rationality of these thresholds, this paper verified and adjusted them many times on the actual data set. This paper used a Weibo text data set with marked loss severity and calculated the classification accuracy, recall, and F1 score by comparing the classification results of the algorithm with those of experts. After several iterations, this paper found that the current threshold setting can balance the accuracy and recall of classification and achieve high classification performance. The pseudo-code of the algorithm for calculating the disaster loss severity of text data by using the loss severity dictionary is shown in Appendix A.

3.4. Step 4: Consistency Analysis

This section uses the tags identified between picture data and text data to study the consistency of data loss information and realizes the qualitative matching and quantitative calculation of the consistency.

3.4.1. Qualitative Discrimination Algorithm of Image–Text Consistency

In this part, the image–text consistency analysis and discrimination were qualitative discrimination first; that is, matching the results obtained from the three classification experiments of image data and text data in disaster correlation, the disaster loss category, and disaster loss severity. When the image data and the corresponding text data belonged to the same category, it was considered that, under the qualitative classification task, the loss information carried by the image data and the text data was consistent. The pseudo-code algorithm based on data fusion for image–text consistency discrimination is shown in Appendix A.

3.4.2. Quantitative Discrimination Algorithm of Image–Text Consistency

After realizing the qualitative discrimination and matching of the consistency of rainstorm and flood disaster loss information, this paper calculated the consistency of rainstorm and flood disaster loss information quantitatively. Because the qualitative matching in the previous article took a picture as the minimum granularity, the quantitative calculation in this section was optimized. In the process of quantitative calculation, the algorithm in this paper comprehensively considered all the loss information contained in a Weibo text and its corresponding images, including loss category information and loss severity information. In the algorithm, firstly, the loss information contained in the text and image was vectorized, and each text was encoded as an 8-dimensional vector, and all the images of the text were also encoded as an 8-dimensional vector. The first five dimensions of the vector contained the loss category information, and the last three dimensions contained the loss severity information. For Weibo text data, a text may only contain one kind of loss severity information, but it may contain multiple pieces of loss category information. For Weibo image data, a Weibo image may only contain one kind of loss severity information, but for the loss category information, this paper found that there are a certain number of Weibo images in the process of data set construction and data labeling, which may contain two or more kinds of loss information, such as both the affected people and the vehicle loss information. If the secondary loss information contained in it was ignored, it may have an impact on the final calculation of the degree of image–text consistency. Therefore, in this paper, the main loss category information and the secondary loss category information contained in the image were considered in the vectorization process of the image loss category information. On the technical realization level, in the loss category classification experiment, the threshold of secondary loss information was set, and the category with the second highest probability output by the deep learning model and the probability value greater than the threshold of secondary loss information was regarded as the secondary loss category information of the image, and the secondary loss category information of the image was also included in the final quantitative calculation algorithm and analysis framework of image–text consistency. Therefore, the information of disaster loss categories contained in the image data can be obtained more fully, which makes the subsequent work of the quantitative calculation of the degree of consistency between images and text more accurate. The pseudo-code of the algorithm to realize the quantitative calculation of image–text consistency is shown in Algorithm 1.

Algorithm 1. Quantitative Calculation of Image–Text Consistency

Input: Weibo text loss information and corresponding map loss information.
Procedure:

```
Initialize Vector_Text [8] = [0, 0, 0, 0, 0, 0, 0, 0]
Initialize Vector_Picture [8] = [0, 0, 0, 0, 0, 0, 0, 0]
Function ExtractFeatures(DataSource):
Initialize Vector [8] = [0, 0, 0, 0, 0, 0, 0, 0]
For each data entry in DataSource:
Vector [0] = feature_affected_individuals
Vector [1] = feature_infrastructure_and_utility_damage
Vector [2] = feature_vehicle_damage
Vector [3] = feature_rescue_volunteering_or_donation_effort
Vector [4] = feature_other_relevant_information
Vector [5] = feature_severe_damage
Vector [6] = feature_mild_damage
Vector [7] = feature_little_damage
Return Vector
Vector_Text = ExtractFeatures(Textdata)
Vector_Picture = ExtractFeatures(Picturedata)
Initialize vector_dot_product = 0
Initialize norm_Vector1 = 0
Initialize norm_Vector2 = 0
For i from 1 to 8:
vector_dot_product = vector_dot_product + (Vector1[i] * Vector2[i])
norm_Vector1 = norm_Vector1 + (Vector1[i] * Vector1[i])
norm_Vector2 = norm_Vector2 + (Vector2[i] * Vector2[i])
norm_Vector1 = sqrt(norm_Vector1)
norm_Vector2 = sqrt(norm_Vector2)
If norm_Vector1 is 0 or norm_Vector2 is 0:
Return 0 (to handle the case of zero vectors)
CosineSimilarity = vector_dot_product/(norm_Vector1 * norm_Vector2)
```

Output: the consistency of loss category and severity between Weibo text and the accompanying pictures.

4. Case Analysis

This section selects the cases of the Beijing rainstorm and Hebei rainstorm and introduces the case background, data, and experimental results in detail. Based on the econometric model, this paper explores the influence mechanism of image–text consistency on social attention.

4.1. Data Acquisition and Cleaning

From 27 July 2023, affected by cold and warm air and typhoon Du Surui, heavy rainfall occurred in most parts of Hebei Province, with an average rainfall of 146.2 mm, which lasted for a long time, from 8:00 on 27 July to 8:00 on 2 August, for nearly 144 h. From 29 July, due to the influence of typhoon Du Surui's residual circulation and subtropical high, typhoon Khanun's water vapor transport and topography, disastrous torrential rain occurred in Beijing and its surrounding areas. From 20:00 on 29 July to 7:00 on 2 August, the average rainfall in Beijing reached 331 mm, and the rainfall within 83 h was 60% of the annual average rainfall. The severe rainstorm and flood disasters in Beijing and Hebei

Province have prompted a lot of attention and heated discussion among users on Weibo. Based on the Python Scrapy framework, this paper used crawler to crawl information on Weibo and used "Beijing rainstorm" and "Hebei rainstorm" as keywords, respectively, and the time span was three months, from 20 July to 20 October 2023, to collect data on Weibo. The collected Weibo information included Weibo IDs, Weibo user nicknames, Weibo's text content, Weibo's corresponding topics, Weibo's IP address, Weibo's release time, Weibo images, Weibo videos, and Weibo likes, retweets, and comments. This study collected 11,646 pieces of text data, 30,155 pieces of image data, and 106 pieces of video data obtained by the keyword "Beijing rainstorm". With the keyword "Hebei rainstorm", 11,288 pieces of text data, 29,511 pieces of image data, and 57 pieces of video data were obtained. Based on the collected Weibo data, the text, image, and video multi-modal loss information data sets of the 2023 Beijing rainstorm and Hebei rainstorm were constructed.

After constructing the data set, detailed data cleaning work was carried out for the text content to ensure the quality and relevance of the data. Firstly, regular expression matching technology was used to remove the irrelevant contents in the text, such as "@ user", advertising links, irrelevant tags (#), and other noise information that may interfere with the text analysis. These contents usually do not contain information directly related to disaster events, so removing them was helpful to improve the purity of the data. Secondly, non-Chinese characters, including letters and special symbols, were deleted from the text. This processing step was helpful to reduce the noise data and ensure that the subsequent text analysis could focus on the disaster-related information in the Chinese context more accurately. Considering that Weibo's text may contain emoticons, which can often convey information and emotion, this paper did not delete them directly. Using artificial regular matching rules and Python's emojiswitch library, emoji symbols were replaced by corresponding text descriptions. This method not only preserved the emotional information conveyed by the emoticons but also made text data more suitable for subsequent natural language processing and analysis. Finally, in order to ensure the authenticity and reliability of the research results, this paper strictly verified and screened the crawled Weibo data. The reliability of social media data may be affected by false information, exaggerated information, and non-disaster-related posts, so this paper paid special attention to these issues in the process of data cleaning. In addition, in order to further verify the reliability of the data, this paper also conducted a manual audit and multi-source verification. After data cleaning, some data were randomly selected for manual review to ensure the effectiveness and accuracy of the cleaning process. For the key information, this paper verified its authenticity by comparing Weibo posts from multiple sources. Through the above cleaning steps, this study effectively removed these potential interference factors and ensured the high quality and representativeness of the data.

4.2. Data Description

This paper made a preliminary exploration and visualization of the obtained Weibo text data. According to the Weibo text crawled by the Beijing rainstorm and Hebei rainstorm, the line charts of the quantitative changes in the Weibo texts of the Beijing rainstorm and Hebei rainstorm over time are shown in Figure 3.

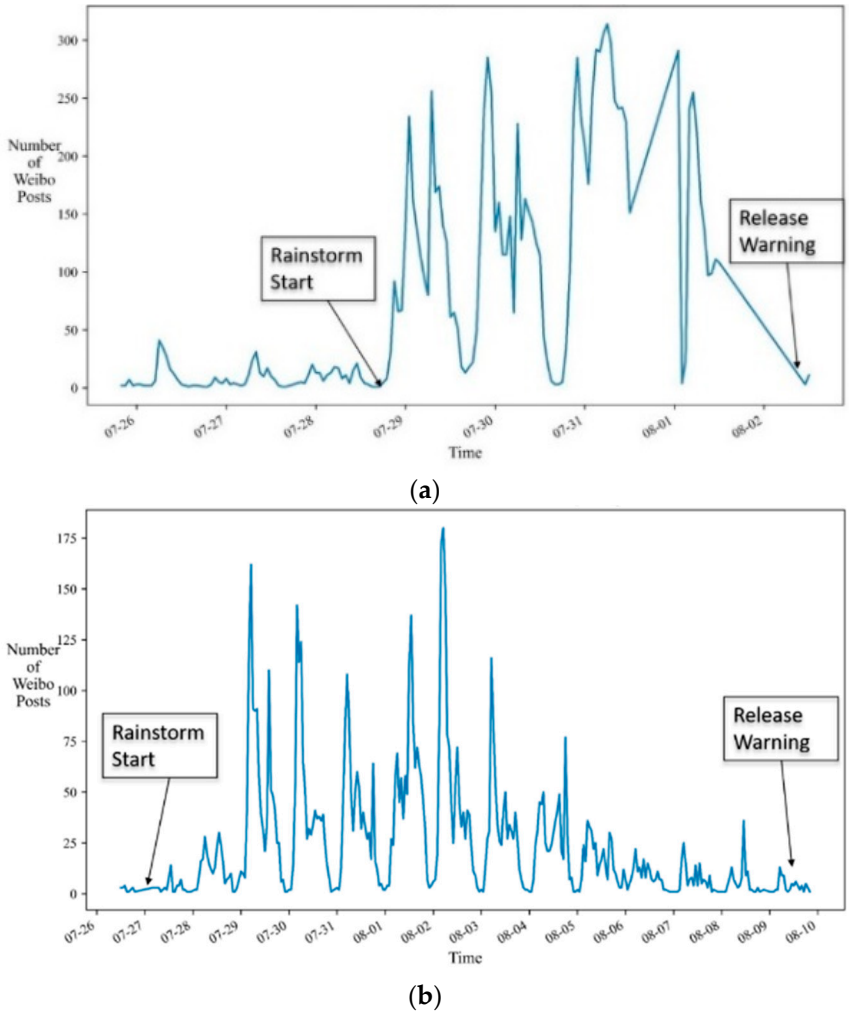


Figure 3. Variation trend of Weibo text quantity of rainstorm in Beijing (**a**) and Hebei (**b**).

The line chart shows that Beijing-related Weibo posts on the rainstorm surged rapidly on 29 July, reaching multiple daily peaks of nearly 300 posts between 29 July and August 2. Peak volumes were identified using the raw hourly counts of user-generated content. The highest peak exceeded 300 posts on the morning of 31 July. Following this period, posts gradually decreased in the afternoon of 1 August and dropped sharply after the Beijing rainstorm warning ended on 2 August. In Hebei Province, related Weibo posts increased slowly starting on 27 July, peaking at over 150 posts on 29 July. From 29 July to 5 August, post numbers exhibited significant diurnal fluctuation, frequently reaching daily peaks during the day before declining rapidly at night. After 5 August, the daily volume decreased markedly, with daily peaks around 25 posts. Posts subsided further after Hebei lifted its rainstorm warning on 10 August.

4.3. Classification Results of Loss Information

This section introduces the classification results of the experiment. It includes the classification results of disaster correlation and the classification results of the loss category and loss severity.

4.3.1. Classification Results of Disaster Correlation

In order to realize the classification of Weibo images, this paper built three classifiers based on the CLIP deep learning model, which were used to judge the disaster correlation, disaster loss category, and disaster loss severity of images. The CLIP model was trained

by using the classic CrisisMMD and multi-modal data sets in the field of disaster management. These data sets provided a wealth of labeled samples, covering the types and losses of floods and other disasters. After the training, this paper inputted the collected Weibo image data into the model, extracted features through the model's image encoder, and matched them with disaster-related text descriptions so as to obtain the classification results quickly and accurately. This method made full use of the multi-modal ability of CLIP and ensured the efficiency and accuracy of classification.

The experimental results of disaster-related image classification are shown in Figure 4. It can be seen that the ratio of disaster-related images to non-disaster-related images on Weibo is about 1:5. This result is consistent with the results of the direct observation of the images collected on Weibo; that is, most of the images on Weibo related to rainstorms did not contain disaster-related information. The content review mechanism of social media platforms such as Weibo may restrict the dissemination of some disaster-related information, so the number of disaster-related images may be smaller than the actual situation. However, based on this, we can also know that even if users on Weibo mentioned the rainstorm-related information in their texts, the corresponding images of Weibo often did not include rainstorm-related images.

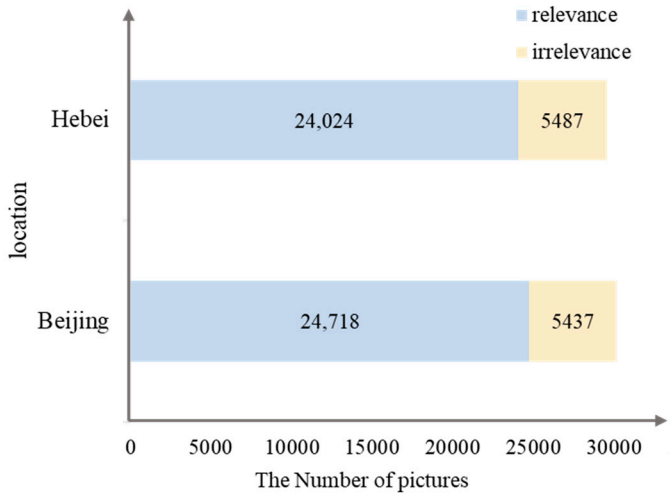


Figure 4. Comparison of disaster-related images in Beijing and Hebei.

4.3.2. Classification Results of Disaster Loss Category and Loss Severity

The experimental results of disaster loss information classification are shown in Figure 5. Based on the classification results of the disaster loss categories of the Weibo images, it can be seen that, except for other disaster-related information categories that have not been clearly classified yet, among the rainstorm-related images on Weibo, the number of images in the disaster relief action category is the highest; that is, people are more willing to share the disaster relief action-related scenes in the rainstorm disaster on the Weibo platform, and the number of Weibo images sharing information about damaged facilities and affected people ranks second or third, while the number of images sharing information about vehicles damaged in the rainstorm is the lowest. As can be seen from the figure, the image categories showing serious losses are the highest, the image categories with moderate losses are in the middle, and the image categories with slight or no losses are the lowest.

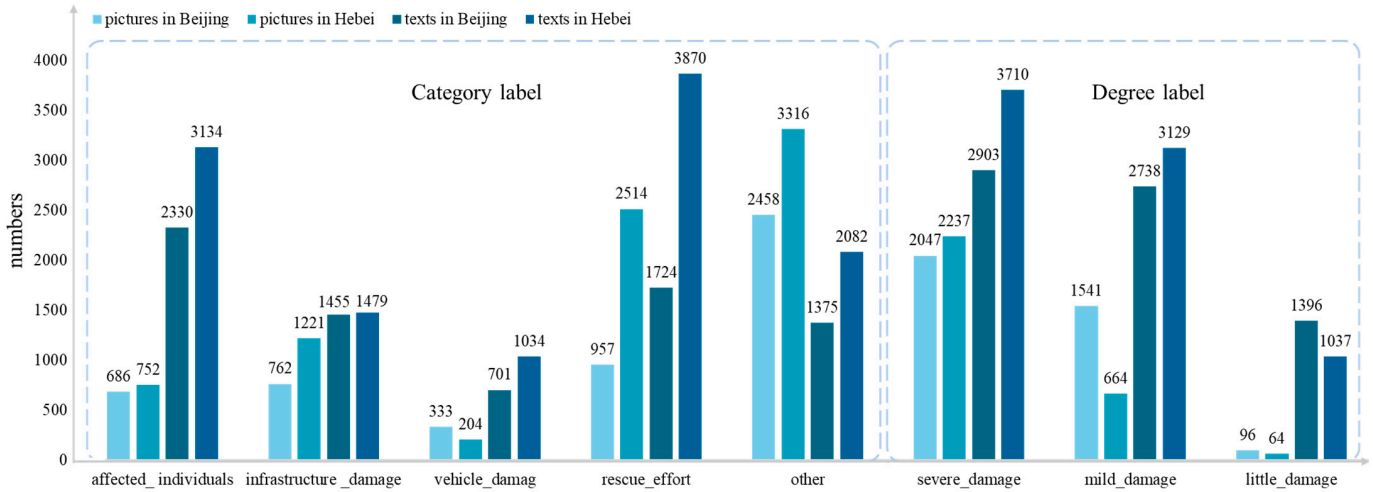


Figure 5. Quantitative distribution of images and texts of disaster loss attributes.

According to the results of the text loss classification of the Beijing rainstorm and Hebei rainstorm, it can be seen that there are some differences between the text loss category distribution and the image loss category distribution. In the classification of loss categories in the text, the number of disaster-stricken people categories and disaster relief action categories is the highest, which is obviously more than the other categories. Similarly to the image loss category, the number of vehicle damage categories is the lowest. Among the Weibo texts related to the rainstorm in Beijing, the category of "affected people" is the highest, and the number of disaster relief operations and damage to infrastructure and public facilities ranks second and third. In Weibo texts related to the rainstorm in Hebei, the category of "disaster relief action" is the highest, and the category of "affected people" is the second highest. It can be seen that in the Weibo texts related to rainstorm and flood disasters, Weibo users will convey more information related to disaster relief operations and affected people. As can be seen from the figure, the image categories showing serious losses are the highest, the image categories with moderate losses are in the middle, and the image categories with slight or no losses are the lowest.

Figure 6 depicts the trend of the images of disaster loss severity during the Beijing rainstorm. On 29 July, around noon, pictures categorized as "other relevant information" surged to a peak of about 60 images. This initial wave coincided with the storm's onset and likely reflects heightened public attention. By noon on 30 July, the focus shifted, and disaster rescue operation pictures became dominant, peaking at about 80 images. This aligns with the documented escalation of relief efforts. The following day (31 July, afternoon), several categories experienced smaller, concurrent peaks: "affected individuals," "infrastructure and utility damage", and "rescue volunteering or donation effort", with each recorded peak at around 40 images. Meanwhile, the "Other relevant information" pictures spiked again to a significant 100 images. This period suggests a shift in public attention towards assessing specific impacts like infrastructure damage and affected populations, while established relief efforts continued and broader discussions of losses intensified. The change trend of the number of Weibo images and texts of disaster loss information in Beijing and Hebei is shown in Appendix B.

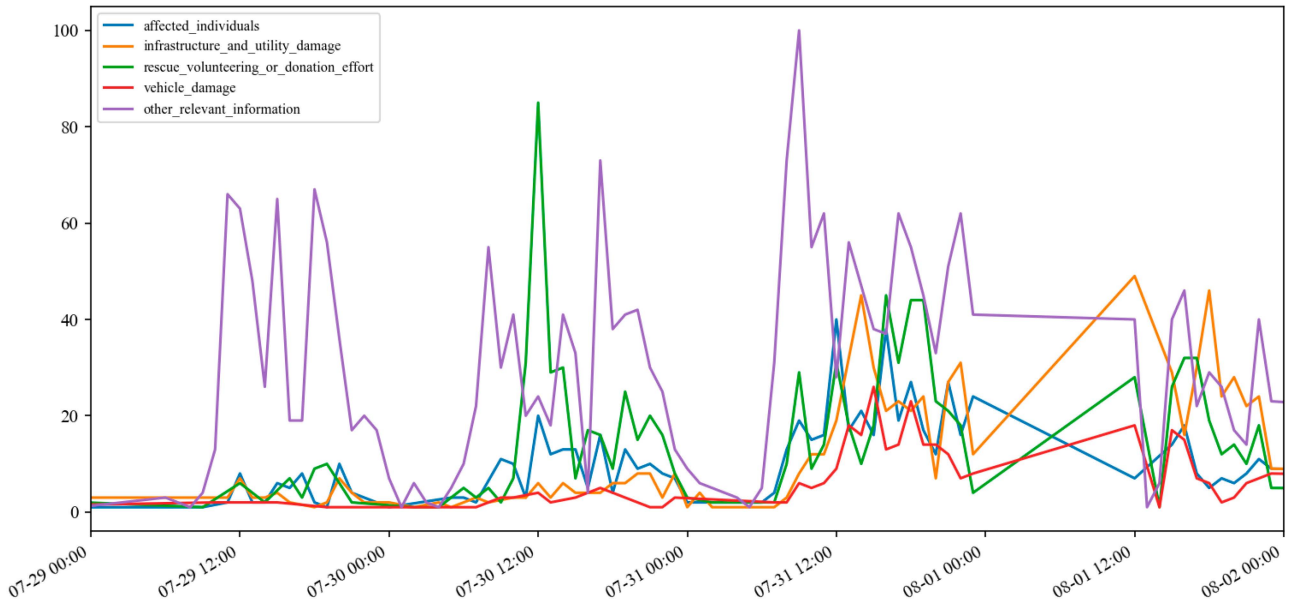


Figure 6. Variation trend of images of loss severity in Beijing.

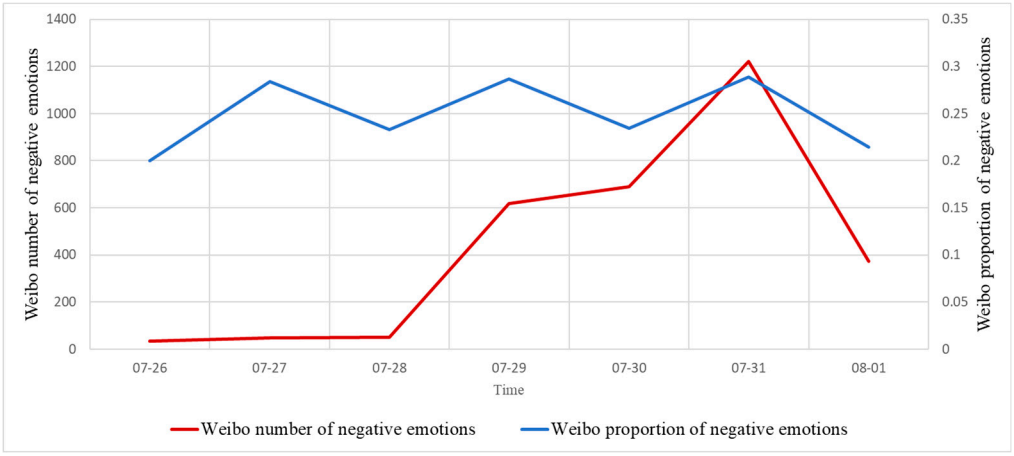
4.4. Mental and Emotional Loss Calculation

In the work of disaster loss assessment, in addition to assessing the economic and property losses suffered by the disaster areas, the mental and emotional losses of the people in the disaster areas are equally important and also an extremely important part of disaster loss assessment. This section used the emotion analysis technology to analyze the Weibo text data released by users on Weibo during the disaster so as to quantify and understand the emotional loss of users on Weibo. Through the systematic analysis of Weibo's text, this paper can identify the emotional changes in users after the disaster, and this information is of great significance to the formulation of psychological intervention and support policies after the disaster. The method in this paper can not only reveal the individual's emotional state but also reflect the overall trend of social emotions at the macrolevel, providing data support for disaster response and recovery.

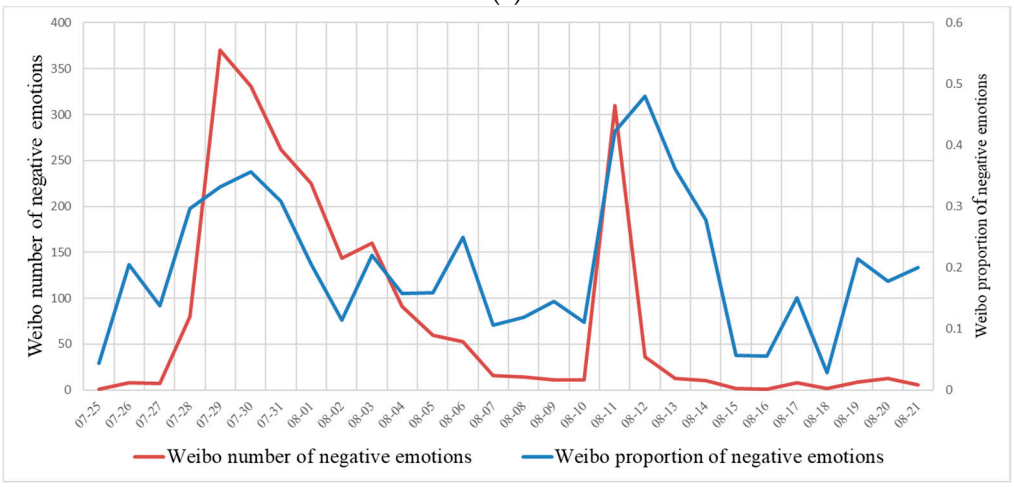
ERNIE-UIE (Enhanced Representation through Knowledge Integration–Universal Information Extraction) is a state-of-the-art information extraction framework developed by Baidu Research [41]. It unifies multiple IE tasks into a single prompt-based paradigm through structured knowledge injection during pre-training. The model establishes new benchmarks for Chinese text understanding by incorporating knowledge-enhanced representations into a unified generative architecture. In this paper, the ERINE-UIE model was used to analyze the emotion of Weibo's text. For a Weibo text, the model identified the corresponding emotional polarity of the text and calculated the emotional intensity of the corresponding emotional polarity. Among them, the polarity of emotion is divided into positive emotion and negative emotion, the range of the emotional intensity score is $[-1,1]$; the emotional intensity value of negative emotion is negative, and the emotional intensity value of positive emotion is positive.

Figure 7 shows distinct patterns in negative sentiment on Weibo for the Beijing and Hebei rainstorms. For Beijing, negative Weibo volume increased sharply starting on 29 July, peaking at over 1200 posts on 31 July. However, the proportion of negative posts remained remarkably stable throughout the event, consistently ranging between 20% and 30% of total rainstorm-related posts, despite the surge in absolute numbers. In contrast, Hebei's negative Weibo presented greater fluctuations in both volume and proportion. The volume rose initially on 28 July, reaching a first peak exceeding 350 posts on 30 July, followed by a gradual decline. A second wave then crested on August 11 with over 300 posts

before subsiding to low volumes. Crucially, the proportion of negative posts in Hebei fluctuated significantly in tandem with these volume changes. Initially, during the main peak (28–31 July), the negative proportion exceeded 30% daily. As volumes lessened, the proportion decreased to around 20%. It then spiked dramatically alongside the second volume peak, reaching over 40% on 11–12 August—the highest sentiment intensity recorded throughout the Hebei rainstorm period.



(**a**)



(**b**)

Figure 7. The changing trend of negative emotions in Beijing (**a**) and Hebei (**b**).

4.5. Analysis on the Consistency of Disaster Information

This section analyzed the classification results of consistency. It included the consistency classification results of loss category and loss severity, the credibility analysis of consistency, and the research on the influence mechanism of consistency on social concern.

4.5.1. Analysis on the Consistency of the Disaster Loss Category and Loss Severity

The classification of disaster loss information with picture–text consistency is shown in Figure 8. Category analysis reveals rescue efforts as the predominant focus in both regions, substantially exceeding vehicle damage and other categories. Severity assessment demonstrates a pronounced emphasis on severe damage. Little damage is negligible. Hebei has more consistent content than Beijing in most cases. It can be found that the relative distribution of image and text modes, category, and severity in the two regions is almost the same. The primary variation lies in Hebei's higher volume of rescue-related content compared to Beijing.

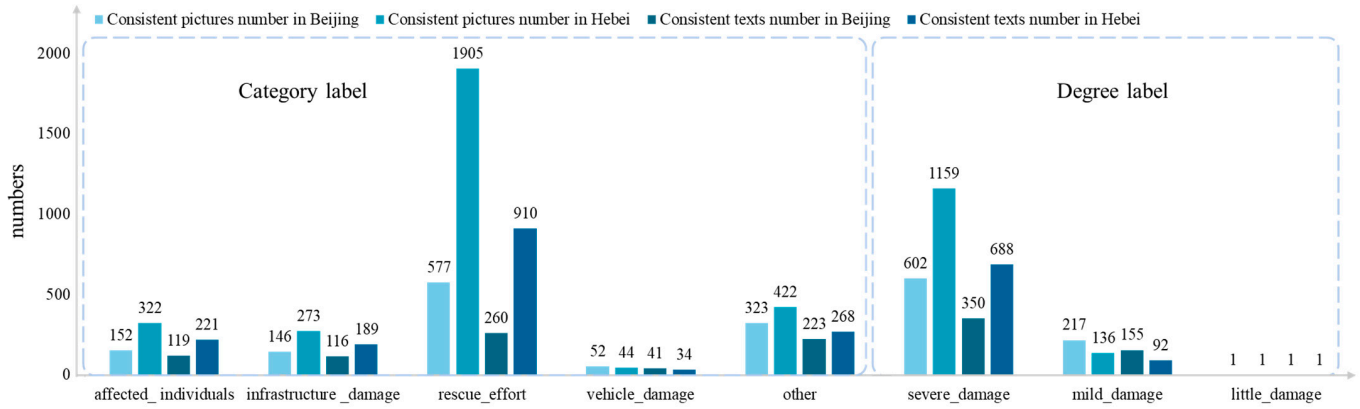


Figure 8. Classification of disaster loss information with consistent image and text.

4.5.2. Credibility Analysis of Consistency

In order to ensure the credibility and sufficiency of the consistency evaluation index, this study introduced the expert evaluation method. Five experts from disaster management, emergency management, and sociology were invited to participate in the assessment. These experts have rich experience in disaster information processing and multi-modal data analysis. We provided each expert with a set of multi-modal data samples after preliminary quantitative evaluation and asked them to rate the consistency of the data according to the following criteria:

1. Category consistency: Assess whether the image matches the disaster category described in the text.
2. Semantic consistency: Assess whether the text and images convey the same or similar information when describing disaster events.
3. Overall consistency: Comprehensively consider the comprehensive factors, such as category and semantics, and score the overall consistency of the multi-modal data.

Experts use a scoring system of 1 to 5, where 1 means "completely inconsistent" and 5 means "highly consistent". In order to ensure the objectivity and consistency of the evaluation, we briefly trained the experts before the evaluation and introduced the evaluation standards and processes in detail. In addition, we also provided several samples for experts to do preliminary exercises and calibration.

After the experts completed the evaluation, we collected their scoring data and compared them with the above-mentioned qualitative and quantitative calculation results. By calculating the average and standard deviation of the expert scores, we can evaluate the explanatory power of the indicators proposed in this paper.

By comparison, it was found that the Pearson correlation coefficient between the expert's evaluation results and the consistency index of this paper is 0.78, which indicates that the expert's score is consistent with the consistency index calculated in this paper in most cases. The reliability of the proposed method was verified.

4.5.3. Analysis on the Influence of Consistency on Social Concern

After realizing the quantitative calculation of the degree of graphic consistency, this paper explored the effect of graphic consistency in Weibo posts on Weibo's social attention based on the econometric model. According to the interpretation level theory, information consistency theory, and dual processing model, this paper holds that text and image information with consistent graphics and text can provide specific and detailed loss information, which is more likely to attract users' interaction and attention. Consistency between the loss information in a Weibo's text and image creates a concrete and detailed scene. This vividness reduces the psychological distance, which in turn enhances users'

emotional resonance and cognitive participation. According to the theory of information consistency, consistent information can reduce cognitive conflicts, make information processing smoother, increase users' goodwill, and thus promote information sharing and discussion. In addition, the dual processing model shows that information with the same image and text can effectively attract users' attention, persuade them to process information at a deeper level, form a deeper memory, and improve the retention rate and communication efficiency of information.

At the same time, combined with the theory of crisis management and crisis communication, in the crisis situation, information with consistent images and text can provide a clearer and more credible explanation, which is helpful to reduce misunderstanding and panic during the disaster and stabilize the audience's mood quickly. According to the theory of crisis management, when an organization faces a potential or actual crisis, it needs to manage and control stakeholders' cognition and response to the crisis through effective information dissemination. Information with consistent images and text not only reduces the fuzziness and uncertainty of information but also improves the transparency and credibility of information. It also evokes the emotional resonance of users through the joint action of vision and text, helping the audience to better understand the seriousness of the crisis, and at the same time, making the public feel the concern and support of the organization. In addition, according to the theory of crisis communication, consistency and transparency are the key factors to establish and maintain public trust. Information with consistent images and text can enhance the credibility of information and improve the trust of the audience. In the dual processing model, information with the same image and text can help the audience better understand and remember the key information through multi-sensory stimulation and improve the information retention rate and communication efficiency. This consistency can also stimulate the interactive behavior of users, increase the exposure of information, and promote the discussion and information diffusion within the community, thus helping organizations to understand the public's needs and concerns more quickly and adjust their crisis response strategies in time.

To sum up, this paper puts forward the core hypothesis that Weibo posts with consistent images and text can significantly enhance the social concern on Weibo.

(1) Regression analysis

This paper constructed a multiple linear regression model to explore the influence of image–text consistency on the social concern of rainstorm-related Weibo posts on the Weibo platform. The model was set as shown in Formula (1):

$$Attention = \theta_0 + \theta_1 Consistency + \theta_2 Controls + \mu, \quad (1)$$

The dependent variable of the model, Attention, is the social attention gained by a Weibo post. In this paper, the sum of the three indicators of Weibo's likes, comments, and reposts is taken as the measure of the social attention gained by the Weibo post. The core explanatory variable is the consistency of the Weibo graphic loss information calculated in this paper, and Controls is the control variable in the model. As for the control variables, this paper refers to the research of Cai et al. [42] and selects the relevant features of the text, such as the emotional intensity of a Weibo text, the length of a Weibo text, the number of topics in a Weibo text, the number of words in disaster loss categories, and the number of words in disaster loss severity in the Weibo text. Referring to the research of Shin et al. [37], we selected the related characteristics of the image, such as the number of images for each Weibo post, the storage space of the image, the warm and cold colors of the image, and the number of faces in the image. And the personal characteristics of the bloggers on Weibo [42], including the number of fans of bloggers on Weibo and whether the IP addresses of bloggers on Weibo are located in the disaster-stricken area. For the relevant

characteristics of the text data, this paper used the pre-trained ERNIE-UIE emotion analysis model to calculate the emotional polarity and emotional intensity of Weibo's text, counted the number of all characters in Weibo's text to characterize the text length, and measured the loss category information and loss severity information contained in Weibo's text by using the number of words matched with those in Weibo's text in the disaster loss category dictionary and disaster loss severity dictionary. For the relevant characteristics of image data, this paper used the computer vision method to judge the color tone of the image, calculated the number of faces in the image based on the Haar cascade classifier, and counted the storage space of the image as a symbol of image clarity. Descriptive statistics of each variable in the model are shown in Table 1.

Table 1. Overall distribution of variable data.

	Average	Variance	Minimum	Maximum
Social concern	85.14	2,098,644.77	0	110,812
Consistency	0.08	0.05	0	1
Emotional intensity of text	0.70	0.06	0.25	1
Text length	277	151,381.69	3	6217
Number of pictures	2.61	6.25	1	9
Picture size (KB)	283	179,725.12	1.3	6756.26
Number of fans	2,879,555	1.1881×10^{14}	0	1.55×10^{8}
The color of the picture warm and cold	0.62	0.11	0	1
Number of words of loss severity	2.72	20.7	0	141
Number of words in loss category	6	158.76	0	296
Number of faces	1.5	5.29	0	53
Weibo topic number	1.65	2.37	0	47
Blogger IP is located in the affected area	0.31	0.21	0	1

In order to ensure the independence of variables in the model, this study conducted multiple collinearity tests. By calculating the variance expansion factor (VIF), we found that the VIF values of all the variables are less than 2.5, indicating that there was no multicollinearity in the explanatory variables. Because some variables have heteroscedasticity, this study chose the robust least squares method to measure them to improve the accuracy of model coefficient estimation. The regression results of the econometric model are shown in Table 2. The results of the econometric model show that the consistency of the images and text has a significant positive impact on Weibo's social concern, and the results meet the previous assumptions. At the same time, it is observed that the control variables, such as the emotional intensity of the text, the number of pictures, the number of fans, and the location of bloggers' IP in the disaster-stricken areas, also have a significant improvement effect on Weibo's social concern, while the model estimation coefficient of text length is negative and significant, indicating that the increase in text length will inhibit Weibo's social concern, and the lengthy Weibo will reduce readability and improve the difficulty of obtaining information.

(2) Robustness test

In order to ensure the reliability and validity of the research conclusion, this study changed the calculation method of the core explanatory variables, and based on the CLIP model, calculated the similarity between a Weibo text and a picture on the overall granularity of the Weibo text and images. Specifically, this study calculated the similarity between image data and text data based on cn_cnlip (Chinese clip) and Clip-vit-large-patch14 models suitable for Chinese scenes. During the experiment, the input of the model was a complete Weibo text and its corresponding Weibo image, and the output was the semantic similarity between the Weibo text and the whole picture. At the same time, in order to make the output similarity results meet the requirements of the specific scenario described in this

paper, this study further improved the existing neural network structure of the model and added a batch normalization regularization layer before the final sigmoid layer, so that the values propagated forward by the neural network can be standardized and normalized, and the numerical distribution can be standardized without destroying the existing pre-training model structure, so that the sigmoid layer can work normally and output more realistic results. The output result of the model was 0-1, in which the numerical value tends to 1, which means that the graphic similarity was high, while the graphic similarity was low. Because this paper focuses on the consistency of graphic loss information, and the graphic similarity directly calculated by using the CLIP model can be considered the similarity of the whole graphic, not limited to the dimension of loss information, the results were fine-tuned based on the classification results of the previous graphic loss information. If the previous graphic loss information labels did not match, the graphic similarity was assigned to 0, while if the previous graphic loss information labels matched, the graphic similarity was assigned to the similarity calculated based on the CLIP model. In this paper, the similarity of image and text calculated by the fine-tuned CLIP model was used as a replacement variable for the consistency of image and text loss information in order to test the robustness of the measurement model. The model regression results of the robustness test by replacing the core explanatory variables are shown in Table 3.

Table 2. Regression results of the influence of graphic consistency on social concern on Weibo.

	Social Concern	**VIF**
Consistency	0.798 ***	1.11
	(7.740)	
Emotional intensity of text	1.729 ***	1.16
	(18.373)	
Text length	−0.255 ***	2.14
	(−10.025)	
Number of pictures	0.122 ***	1.10
	(14.506)	
Picture size	0.316 ***	1.43
	(14.593)	
Number of fans	0.316 ***	1.45
	(53.711)	
The color of the picture	0.013	1.06
	(0.205)	
Number of words of loss severity	0.004	2.22
	(0.633)	
Number of words in loss category	−0.006 **	2.18
	(−2.341)	
Number of faces	−0.006	1.25
	(−0.642)	
Weibo topic number	−0.049 ***	1.06
	(−3.305)	
Blogger IP is located in the affected area	0.712 ***	1.44
	(13.289)	
N	22,922	
R2	0.082	
F	297.13	

*** $p < 0.01$, ** $p < 0.05$.

Table 3. Robustness test of substitution variables.

	Social Concern
Consistency	0.856 ***
	(7.924)
Emotional intensity of text	1.724 ***
	(18.307)
Text length	−0.251 ***
	(−9.893)
Number of pictures	0.123 ***
	(14.682)
Picture size	0.313 ***
	(14.461)
Number of fans	0.315 ***
	(53.470)
Number of words of loss severity	0.005
	(0.770)
Number of words in loss category	−0.007 ***
	(−2.683)
Weibo topic number	−0.050 ***
	(−3.351)
Blogger IP is located in the affected area	0.715 ***
	(13.353)
N	22,922
R^2	0.081
F	298.801

*** $p < 0.01$.

According to the regression results, it can be seen that after replacing the core explanatory variables, the degree of graphic consistency still has a significant positive impact on Weibo's social concern, so it can be proved that the conclusion of this paper is robust.

In order to further verify the robustness of the research results, this paper used the method of reducing control variables to test the robustness. In the original model, this paper included several control variables to capture the factors that may affect the dependent variables as comprehensively as possible. However, too many control variables may introduce noise, which will affect the explanatory ability of the model and the stability of the results. Therefore, we adopted the method of reducing control variables, only kept the core explanatory variables and main control variables, and conducted a regression analysis again to observe the robustness of the results. The experimental results are shown in Table 4.

Table 4. Robustness test of reduced control variables.

	Social Concern
Consistency	0.960 ***
	(8.840)
Emotional intensity of text	1.071 ***
	(11.575)
Number of pictures	0.122 ***
	(13.760)
N	22,922
R^2	0.018
F	137.851

*** $p < 0.01$.

By reducing the control variables, this paper finds that the significance and signs of core explanatory variables are consistent under different model settings. This shows that even if there are few control variables, the influence of the core explanatory variables on the dependent variables is still significant, and the direction remains unchanged. This result further verifies the robustness of the core explanatory variables in the model and enhances the confidence in the research conclusions.

In order to further verify the robustness of the research results, this paper also adopted the robustness test method of shortening the time window. The sample period selected by the original regression was from 20 July to 20 October 2023. However, the actual duration of the rainstorm event was less than three months. According to the records of the meteorological department, the rainstorm event occurred on 26 July 2023. Before 12 August, the discussion on this rainstorm and the disasters caused by it was most concentrated on social media. On 12 August 2023, the Government Information Office held a press conference to inform people about flood control and disaster relief. Since then, with the advancement of rescue work and the shift in public concerns, the relevant discussions have gradually decreased. Therefore, ending the sampling period on 12 August 2023 can help avoid the interference of discussions unrelated to this rainstorm in the later period. In order to ensure the robustness of the analysis results, this paper adjusted the sample period from 26 July to 12 August 2023 and re-conducted the regression analysis. The experimental results are shown in Table 5.

Table 5. Robustness test of sample selection.

	Social Concern
Consistency	0.787 ***
	(6.968)
Emotional intensity of text	1.320 ***
	(13.176)
Text length	−0.000 *
	(−1.788)
Number of pictures	0.139 ***
	(14.663)
Picture size	0.000 **
	(2.350)
Number of fans	0.000 ***
	(37.687)
Number of words of loss severity	0.031
	(0.441)
Number of words in loss category	0.001
	(0.169)
Weibo topic number	0.001
	(0.309)
Blogger IP is located in the affected area	0.001
	(0.079)
N	20,375
R^2	0.084
F	155.481

*** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

Using the robustness test based on sample selection, this paper finds that the significance and signs of the core explanatory variables are consistent under different sample conditions. This shows that even when the sample range is adjusted, the influence of the core explanatory variables on the dependent variables is still significant, and the direction remains unchanged. This result further verifies the robustness of the core explanatory variables in the model.

(3) Heterogeneity analysis

Considering that different users on Weibo have different numbers of fans, and after the emotion analysis, Weibo's text was divided into two types of emotional polarity, positive emotion and negative emotion, so this paper made a heterogeneity analysis on the number of fans of different users on Weibo and the positive and negative emotional tendencies of Weibo's text. Among them, whether the number of fans exceeded 10,000 was the criterion. Users with more than 10,000 fans were considered to belong to a group with more fans, while users with fewer than 10,000 fans were considered to belong to a group with fewer fans. The emotional polarity of Weibo's text was characterized by the discriminant results output by the pre-trained ERNIE-UIE model, including positive emotions and negative emotions. The results of the grouping regression based on the differences in the number of fans and the emotional polarity of the text are shown in Table 6.

Table 6. Heterogeneity analysis.

	(1) More Fans	(2) Less Fans	(3) Positive Emotion	(4) Negative Emotion
Consistency	0.929 ***	0.344 *	0.884 ***	0.255
	(7.677)	(1.768)	(7.404)	(1.213)
Emotional intensity of text	1.995 ***	1.144 ***	1.914 ***	1.937 ***
	(16.372)	(7.965)	(16.636)	(9.955)
Text length	−0.403 ***	−0.123 ***	−0.266 ***	−0.197 ***
	(−10.627)	(−3.421)	(−9.271)	(−3.441)
Number of pictures	0.160 ***	0.106 ***	0.116 ***	0.161 ***
	(14.454)	(8.371)	(12.552)	(7.645)
Picture size	0.337 ***	0.263 ***	0.294 ***	0.377 ***
	(11.978)	(7.978)	(11.492)	(9.128)
Number of fans	0.601 ***	0.137 ***	0.326 ***	0.298 ***
	(45.603)	(10.205)	(49.262)	(23.795)
The color of the picture	0.027	−0.122	−0.132 *	0.432 ***
	(0.344)	(−1.207)	(−1.800)	(3.374)
Number of words of loss severity	0.012	−0.048 **	0.010	−0.041 **
	(1.612)	(−2.447)	(1.232)	(−2.414)
Number of words in loss category	−0.005 **	−0.008	−0.008 ***	0.026 ***
	(−2.000)	(−1.351)	(−2.859)	(3.121)
Number of faces	0.034 **	−0.026 **	−0.003	−0.008
	(2.142)	(−2.185)	(−0.292)	(−0.387)
Weibo topic number	−0.052 **	−0.114 ***	−0.066 ***	0.069 **
	(−2.283)	(−5.867)	(−3.691)	(2.036)
Blogger IP is located in the affected area	1.088 ***	0.372 ***	0.699 ***	0.713 ***
	(14.821)	(4.892)	(11.513)	(6.309)
N	14,855	8067	17,184	5738
R^2	0.114	0.074	0.086	0.074
F	269.795	61.625	249.011	63.363

*** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

According to Table 6, it can be seen that the consistency of the core explanatory variables is significantly different among groups with different numbers of fans and emotional polarity. Specifically, for groups with a large number of fans and a small number of fans, the consistency of the graphics and text has a significant positive impact on Weibo's social attention, but the coefficients and significance of the two groups are different. In groups with a large number of fans, the coefficients and significance of the explanatory variables are higher. The results show that the consistency of graphics and text content has a more significant role and effect in enhancing social attention. For accounts with a large number

of fans, the content published by them can quickly attract and maintain a high level of social attention because of its high initial visibility and broad audience base. According to the theory of information communication in crisis communication, the information of high-impact accounts is not only transmitted quickly but also easily regarded as authoritative and credible by the public, thus enhancing the acceptance and communication efficiency of information. In addition, such accounts often form a close social network with other influential accounts, further amplifying the effect of information dissemination. Therefore, in a crisis situation, accounts with large fan bases can convey key information more effectively, enhance the public's trust, and quickly gain extensive social attention through the content of graphic consistency. In contrast, although accounts with a small number of fans can attract some attention through high-quality graphic content, their influence is relatively limited due to the lack of sufficient initial exposure and communication channels.

In groups with different emotional polarities in Weibo's texts, the effect of graphic consistency on Weibo's social concern is different. On Weibo, where the text is positive, the content of graphic consistency can also significantly enhance Weibo's social concern, while on Weibo, where the text is negative, the effect of graphic consistency on Weibo's social concern is not significant. From the perspective of crisis management and crisis communication, positive emotional information helps to alleviate the public's anxiety during the crisis, enhance the credibility and affinity of organizations or individuals, and then promote the positive dissemination of information. According to the theory of emotional resonance in crisis communication, the content of positive emotions is more likely to arouse the emotional resonance of the public and stimulate their motivation to share and support, thus expanding the coverage of information. On the contrary, the content of negative emotions, even if the images and text are consistent, may reduce their appeal by causing public panic, dissatisfaction, or resistance. In the crisis situation, negative information was widespread, and overemphasizing negative emotions may aggravate the psychological burden of the public and lead to information overload, thus weakening the positive effect of graphic consistency. Therefore, in order to effectively enhance social concern in a crisis, we should pay attention to the use of positive emotional graphic content to establish a positive public image and trust relationship.

(4) Moderating effect analysis

After exploring the robustness and heterogeneity of the influence of consistency on Weibo's social concern, this paper further explored the moderating effect in order to provide optimization strategies and guidance for the official media and individual users to release relevant information on social media platforms during the disaster. This paper holds that the correlation between Weibo's text and images will adjust the influence of the consistency of text and images on Weibo's social concern. According to the theory of crisis management, transparency and responsibility are the key principles in crisis responses, and concrete images and text descriptions containing disaster-related information can enhance the transparency of information and facilitate public understanding, thus enhancing trust during disasters. Therefore, this paper expected that the number of images on Weibo and the information related to disaster relief contained in Weibo's texts can positively adjust the influence of the consistency of images and text on the social concern on Weibo, that is, by enhancing the transparency of information, reducing the cognitive burden, and improving the consistency and reliability of disaster-related information, thus enhancing the social concern on Weibo. In this section, the number of images on Weibo, the information related to disaster relief operations contained in Weibo texts, and the consistency of images and text were constructed. The regression results of the Weibo graphics and other characteristics as adjustment variables are shown in Table 7.

Table 7. Analysis on the moderating effect of graphic features on Weibo.

	(1)	(2)
Consistency	0.558 ***	0.296 **
	(3.391)	(2.007)
Number of pictures	0.116 ***	0.125 ***
	(13.070)	(14.670)
Consistency *Number of pictures	0.075 *	
	(1.909)	
The text contains information on disaster relief operations		0.128 *
		(1.859)
Consistency * The text contains information on disaster relief operations		0.817 ***
		(3.865)
Emotional intensity of text	1.722 ***	1.650 ***
	(18.293)	(17.166)
Text length	−0.257 ***	−0.252 ***
	(−10.071)	(−9.852)
Picture size	0.316 ***	0.318 ***
	(14.585)	(14.676)
Number of fans	0.316 ***	0.314 ***
	(53.709)	(53.234)
Number of words in loss category	−0.006 **	−0.009 ***
	(−2.344)	(−3.442)
Weibo topic number	−0.049 ***	−0.051 ***
	(−3.335)	(−3.362)
Blogger IP is located in the affected area	0.710 ***	0.724 ***
	(13.239)	(13.499)
N	22,922	22,922
R^2	0.081	0.084
F	274.446	259.658

*** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

According to the regression results, it can be seen that the number of pictures on Weibo with consistent images and text and the information related to disaster relief actions contained in the text can play a positive role in regulating the influence of the consistency of images and text on the social concern on Weibo. According to the theory of visual communication, images can provide intuitive and concrete information and enhance the perceptibility and attractiveness of information. In crisis situations, people tend to receive and process visual information because images can quickly convey complex situations and emotions and reduce obstacles to text understanding. When the content of Weibo contains more images, these images can not only supplement the text description but also attract more attention through visual impact. Especially during the crisis, images can show key information, such as the scene situation and the progress of rescue, so that the public can understand the development of the situation more intuitively, thus increasing the authenticity and credibility of the information. In addition, the use of multiple images can also build a coherent storyline to help the audience better understand and remember information and promote the sharing and dissemination of information. Therefore, the increase in the number of images enhances the visual effect and story of the information, further magnifies the positive effect of the consistency of the images and text, and improves the social attention of Weibo.

For the information related to disaster relief operations contained in Weibo's text, according to the transparency principle and responsibility theory in crisis communication, it is helpful to report the specific measures and progress of disaster relief operations openly and transparently, which will help to establish and maintain the credibility of organiza-

tions or individuals. In the crisis situation, the public's demand for information is particularly urgent, and they want to know what specific actions government agencies, non-governmental organizations, or other interested parties are taking to deal with the crisis. When Weibo described the specific disaster relief action in detail, it not only showed the responsibility and action of the information publisher but also provided a clear source of information for the public, reducing uncertainty and panic. In addition, disaster relief information usually has high news value and social value, which makes it easy to attract the attention of the media and the public and then lead to more discussion and sharing. Therefore, the information related to disaster relief operations contained in the text enhances the influence of graphic consistency by improving the relevance and practicability of the information and further enhances the social concern on Weibo.

(5) Quantile regression

In order to further verify the robustness of the benchmark regression results, this paper used the quantile regression method for supplementary testing. Considering that the dependent variables usually have skewed distribution characteristics, the mean regression results may be disturbed by extreme values. Quantile regression can reflect the influence of core variables on typical observations more stably by estimating the effects of the dependent variables on different conditional quantiles (τ = 0.25, 0.5, 0.75) and is insensitive to outliers. The results in Table 8 show that the consistency is significantly positive at τ = 0.5 and τ = 0.75, and the effect scale increases with the increase in the quantile. The variable failed the significance test at τ = 0.25. It can be seen that the promotion of graphic consistency to social attention focuses on Weibo, which is of medium and high concern but has limited influence on Weibo, which has low communication power. This stratification effect may be one of the reasons why the benchmark OLS regression R^2 is relatively low; that is, the mean regression fails to fully capture the heterogeneous influence of core variables at different response levels, which weakens the overall explanatory power of the model. This discovery has practical enlightenment, and improving the consistency of images and text is more effective for Weibo (such as official government accounts), which has a certain communication foundation. For Weibo posts with low influence, priority should be given to solving other communication bottlenecks (such as initial exposure). Future research can further explore content optimization strategies at different communication levels.

Table 8. Quantile regression results.

	(1) **QR (τ = 0.25)**	**(2)** **QR (τ = 0.5)**	**(3)** **QR (τ = 0.75)**
Consistency	0.157	0.560 ***	0.716 ***
	(0.469)	(6.196)	(6.993)
Emotional intensity of text	0.316	0.806 ***	0.708 ***
	(1.059)	(10.015)	(7.772)
Text length	−0.000	−0.000 ***	0.000
	(−0.252)	(−2.610)	(1.146)
Number of pictures	0.219 ***	0.108 ***	0.097 ***
	(7.887)	(14.329)	(11.422)
Picture size	0.001 ***	0.000 **	−0.000
	(5.108)	(2.100)	(−0.505)
Number of fans	0.000 ***	0.000 ***	0.000 ***
	(15.423)	(37.593)	(30.685)
The color of the picture	−0.026	0.033	0.053
	(−0.126)	(0.593)	(0.845)
Number of words of loss severity	−0.002	0.002	0.008
	(−0.106)	(0.297)	(1.090)

Table 8. *Cont.*

	(1) QR (τ = 0.25)	(2) QR (τ = 0.5)	(3) QR (τ = 0.75)
Number of words in loss category	0.001	0.001	−0.004
	(0.104)	(0.370)	(−1.598)
Number of faces	0.020	−0.001	0.001
	(0.557)	(−0.080)	(0.066)
Weibo topic number	−0.017	−0.042 ***	−0.031 **
	(−0.392)	(−3.539)	(−2.297)
Blogger IP is located in the affected area	0.096	0.126 ***	−0.048
	(0.598)	(2.918)	(−0.985)
N	22,922	22,922	22,922

*** $p < 0.01$, ** $p < 0.05$.

5. Conclusions and Discussion

Mining loss information from the massive multi-modal data of social media and screening out multi-modal data with loss information and highly consistent images and text are very important for disaster management research and emergency management departments to respond quickly and make correct decisions. This study put forward a disaster loss information extraction model from the perspective of social media and introduced multi-modal data to evaluate the disaster loss, aiming at improving the efficiency and accuracy of disaster loss information identification. In order to ensure the transparency and repeatability of the research, this paper recorded and described the details of model construction, data crawling, pretreatment, and the experimental environment in detail. The CLIP model architecture, the experimental environment configuration, and the data set this study rely on were all public information, and the data set can be obtained through public channels (https://crisisnlp.qcri.org/crisismmd.html, accessed on 10 May 2025). In order to obtain the data from Weibo, this study used the crawler tool written by Python to collect data and described the data processing process in detail, including key steps such as data cleaning and feature extraction. All data dictionaries are also open and can be copied in the process. The openness and transparency of this information ensures that other researchers can replicate experiments and improve models.

This study found that the Weibo posts of disaster relief actions were themost in the loss category, and most of the posts of disaster loss severity showed serious losses. This study further explored the influence of consistency on social concern. In this study, social concern refers to the degree of user interaction with specific content on social media platforms, which is measured by the number of reposts, comments, and likes on Weibo. These indicators reflect the public's attention and participation in disaster-related information and are important quantitative indicators for evaluating social attention. This study found that high consistency will significantly enhance social concern. This discovery can be explained by the construction level theory (CLT). According to the theory of interpretation level, concrete and detailed information is more likely to prompt users' interaction and attention because it is closer to users' psychological distance. On social media, information with consistent images and text can provide more specific and detailed disaster loss information, which is easier to be understood and accepted by users. In addition, the information consistency theory also supports the view that consistent information can enhance the credibility and persuasiveness of information, thus improving users' attention to and interactivity with information. At the same time, combined with the theory of crisis management and crisis communication, in the crisis situation, information with consistent images and text can provide a clearer and more credible explanation. During a disaster, such clear and credible information is essential to reduce misunderstanding and panic. Infor-

mation with consistent images and text can help the public better understand the disaster situation, thus quickly stabilizing the audience's emotions and providing support for the effective communication and decision-making of emergency management departments.

Therefore, this study puts forward the following policy suggestions:

- Actively use social media data for disaster early warning and response. The government and disaster management departments need to use advanced data mining technology to monitor disaster-related information on social media platforms (such as Weibo, WeChat, etc.) in real time to find and respond to potential disaster risks in time. At the same time, according to the types of disasters and public demand, targeted information dissemination strategies are formulated to improve the dissemination effect of information.
- Pay attention to the important role of data fusion in disaster loss assessment and improve the degree and quality of data fusion. By integrating multi-modal data, disaster losses can be evaluated more comprehensively and accurately, and more reliable decision support can be provided for disaster response and resource allocation. Therefore, the government needs to establish a cross-departmental and cross-platform data sharing mechanism, optimize the data processing process, and ensure the accuracy and consistency of data.
- Use consistency to improve social concern. In the dissemination of disaster information, the consistency of information is very important to enhance social attention, and social attention is an important factor to measure the quality of disaster relief work. Therefore, the government and disaster management departments should ensure the consistency of graphic information in content, semantics, and vision when releasing disaster information. For example, formulate unified standards and guidelines for information dissemination to avoid information conflicts or misleading information. Establish a strict information review mechanism to correct and clarify inconsistent or false information in a timely manner. Through these measures, the effect of disaster information dissemination can be effectively improved, and the public's attention and understanding of disaster events can be enhanced, thus better supporting disaster management and emergency responses.

This study provides a new perspective and method for multi-modal data analysis in the field of disaster management, but there are still the following practical limitations and future research directions:

- Limitation in data source diversity: The conclusion is mainly based on the Weibo data of rainstorms in a specific area, which limits the universality and representativeness of the results. In the future, we plan to integrate diversified data sources, such as GIS data, meteorological data, and official reports, to build a more comprehensive disaster assessment view.
- Limitation in model generalization: The applicability of the model in different regions and disaster types (such as earthquakes and typhoons) needs to be verified. In the future, we plan to collect and analyze more extensive regional and disaster type data to improve the generalization ability and adaptability of the model.
- Limitation in fusion depth: The current fusion of the data layer and decision layer may not fully tap the complementary information between modes. In the future, we plan to explore feature-level fusion technology to enhance the fine-grained recognition ability of the model for disaster information.

The framework of the consistent analysis of disaster loss information put forward in this study provides a powerful tool for disaster management practice. The emergency management department can quickly screen out the loss information with high consistency

between images and text so as to evaluate the disaster impact more effectively and make rescue plans and resource allocation. The data-driven method proposed in this paper is also helpful to reduce the uncertainty of decision-making and improve the timeliness and effectiveness of disaster responses. At the same time, the research on the influence mechanism of the consistency of images and text on social concern reveals the importance of information presentation to the public response, which provides theoretical support for the formulation of disaster information dissemination strategies. By optimizing the information dissemination strategy, we can better guide the public to pay attention to and participate in disaster management and enhance the overall coping abilities of society. This also provides a new perspective for disaster education and public awareness, which is helpful to improve public participation and satisfaction.

Author Contributions: Conceptualization, S.S., J.S. and J.L.; methodology, J.S. and J.L.; validation, Y.L. and Z.Z.; data curation, J.L.; writing—original draft preparation, J.S. and J.L.; writing—review and editing, Y.L. and Z.Z.; visualization, J.S. and J.L.; supervision, S.S.; funding acquisition, S.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (No.72071010).

Data Availability Statement: The original contributions presented in the study are included in the article; further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

The classification dictionaries of the disaster loss category, severity, and pseudo-code of algorithms are shown in the following table.

Table A1. Classification dictionary of disaster loss categories.

	Chinese	English	Chinese	English
affected_activities	活动	Activity	减少	Reduce
	取消	Cancel	供应链	Supply Chain
	出行	Travel	工作	Work
	受阻	Hindered	远程	Remote
	交通	Traffic	办公	Office
	中断	Interrupt	网络	Network
	停工	Shutdown	教育	Education
	停课	Suspend	在线	Online
	延迟	Delay	灾民	Disaster Victims
	聚会	Gathering	学习	Study
	户外	Outdoor	医疗	Medical
	受限	Restricted	紧张	Tense
	休闲	Leisure	民众	Citizens
	娱乐	Entertainment	群众	Crowd
	运输	Transport	受灾	Affected
	困难	Difficulty	人员	Personnel
	市场	Market	人身	Personal
	关闭	Close	安全	Safety
	商业	Business	受伤	Injured
	运营	Operation	伤亡	Casualty
	中止	Terminate	死亡	Death
	旅游	Tourism	失踪	Missing
	限制	Limit	遇难	Perish
	公共	Public	受困	Trapped

Table A1. *Cont.*

	Chinese	English	Chinese	English
damage_infrastructure	服务	Service	被困	Stranded
	楼房	Building	积水	Waterlogging
	房子	House	房屋	House
	房屋	House	被淹	Flooded
	建筑	Structure	设施	Facilities
	桥梁	Bridge	不通	Inaccessible
	道路	Road	供电	Power supply
	公路	Highway	中断	Interruption
	倒塌	Collapse	供水	Water supply
	受损	Damaged	中断	Interruption
	损毁	Destroyed	燃气	Gas
	冲毁	Washed out	泄漏	Leak
	堤坝	Levee	排水	Drainage
	决口	Breach	系统	System
	水管	Water pipe	故障	Failure
	破裂	Ruptured	交通	Traffic
	电网	Power grid	信号	Signal
	电力	Electricity	故障	Failure
	中断	Interruption	轨道	Rail
	通讯	Communication	交通	Traffic
	网络	Network	停摆	Paralyzed
	故障	Failure	基础	Basic
	道路	Road	铁路	Railway
loss_of_property	车	Vehicle	受损	Damaged
	汽车	Car	进水	Water ingress
	卡车	Truck	丢失	Lost
	电动车	Electric vehicle	事故	Accident
	摩托车	Motorcycle	被毁	Destroyed
	轿车	Sedan	损坏	Damaged
	公交车	Bus	亏损	Loss
	救护车	Ambulance	保险	Insurance
	出租车	Taxi	理赔	Claim
	货车	Freight truck	财产	Property
	自行车	Bicycle	贬值	Depreciation
	小轿车	Compact car	赔偿	Compensation
	停车	Parking	故障	Failure
	泊车	Parking	维修	Repair
	车库	Garage	积水	Waterlogging
	车辆	Vehicle	泡水	Water-soaked
	被淹	Flooded	浸水	Immersed
	淹没	Submerged	熄火	Stall
	损失	Loss	失灵	Malfunction
	沉没	Sunk	失控	Out of control
disaster_relief	紧急	Emergency	救助	Assistance
	救援	Rescue	灾后	Post-disaster
	疏散	Evacuation	重建	Reconstruction
	转移	Relocation	受困	Trapped
	抢险	Emergency response	心理	Psychological
	救灾	Disaster relief	援助	Aid
	搜救	Search and rescue	物流	Logistics
	失踪	Missing	支援	Support
	人员	Personnel	基础	Basic
	安置	Placement	设施	Facilities
	受灾	Affected	恢复	Recovery
	群众	Public	公共	Public

Table A1. *Cont.*

	Chinese	English	Chinese	English
other_losses	捐赠	Donation	秩序	Order
	物资	Supplies	应急	Emergency
	志愿者	Volunteer	管理	Management
	行动	Action	救灾	Disaster relief
	医疗	Medical	训练	Training
	生态	Ecology	问题	Issue
	破坏	Destruction	经济	Economic
	动物	Animal	影响	Impact
	农作物	Crop	旅游业	Tourism
	歉收	Poor harvest	教育	Education
	自然	Nature	创伤	Trauma
	资源	Resource	人际	Interpersonal
	损失	Loss	关系	Relationship
	环境	Environment	紧张	Tension
	污染	Pollution	社区	Community
	文化	Culture	支持	Support
	遗产	Heritage	减少	Reduce
	受损	Damaged	治安	Public security
	社会	Society	压力	Stress
	秩序	Order	焦虑	Anxiety
	安全	Safety	恐惧	Fear
	威胁	Threat	抑郁	Depression
	心理	Psychological	综合症	Syndrome
	健康	Health	失眠	Insomnia

Table A2. Classification dictionary of disaster loss severity.

	Chinese	English	Chinese	English
10	死亡	Death	溺亡	Drown
	尸体	Corpse	致命	Lethal
	死	Dead	丧生	Perish
	去世	Pass away	牺牲	Sacrifice
	致命的	Fatal	殉职	Die in the line of duty
	遇难	Perish	失去生命	Lose life
	遇难者	Victim		
9	伤亡	Casualty	伤亡人员	Casualty personnel
	受伤	Injured	伤亡者	Casualty
	失踪	Missing	伤亡情况	Casualty situation
	失联	Out of contact	下落不明	Missing
	受害	Affected	伤病	Injury and illness
	重伤	Seriously injured	埋压	Buried
	伤者	Wounded	毁灭	Annihilate
8	紧急情况	Emergency situation	灾难性	Catastrophic
	事故	Accident	险情	Perilous situation
	灾难	Disaster	洪灾	Flood disaster
	悲剧	Tragedy	危险	Dangerous
	突发事件	Sudden event	受灾	Affected by disaster
	灾害	Calamity	惨烈	Terrible
	灾情	Disaster situation		
7	断电	Power outage	终止	Terminate
	毁坏	Ruin	瘫痪	Paralyze

Table A2. *Cont.*

	Chinese	English	Chinese	English
	崩塌	Collapse	搁置	Shelve
	摧毁	Destroy	被困	Trapped
	严重	Severe	停运	Cease operations
	混乱	Chaos	停工	Stop work
	风暴	Storm	围困	Besiege
	损毁	Damage	内涝	Urban flooding
	破损	Broken	被淹	Flooded
	暂停	Suspend		
6	破坏	Destruction	撞击	Collision
	残骸	Debris	严重	Severe
	垃圾	Garbage	破裂	Rupture
	碎片	Fragment	破碎	Shattered
	冲击	Impact	毁坏	Ruin
	重创	Severe blow		
5	限制	Restriction	中断	Interruption
	延迟	Delay	延误	Delay
	停止	Stop	短缺	Shortage
	暂停	Suspend	缺水	Water shortage
	关闭	Close	取消	Cancel
	停滞	Stagnation	紧缺	Scarce
4	淹没	Submerge	威胁	Threaten
	失败	Failure	破坏	Destroy
	阻挡	Block	袭击	Attack
	噩梦	Nightmare	冲击	Impact
	妨碍	Obstruct	障碍	Obstacle
3	加剧	Intensify	绝望	Despair
	极端	Extreme	崩溃	Collapse
	暴力	Violence	强降雨	Heavy rainfall
	强烈	Intense	极值	Extreme value
	猛烈	Fierce	危急	Critical
	剧烈	Violent		
2	突然	Suddenly	恐怖	Terrifying
	骤然	Abruptly	恐惧	Fear
	急剧	Sharply	惊恐	Panic
	突如其来	Come suddenly	惊慌	Flustered
	突发	Sudden	突降	Sudden drop
	突变	Sudden change	警惕	Be on guard
	害怕	Afraid	预警	Early warning
1	悲伤	Sadness	伤心	Heartbroken
	痛苦	Pain	揪心	Heart-wrenching
	困扰	Distress	哽咽	Sobbing
	悲痛	Grief	担忧	Worry
	难过	Upset		

Algorithm A1: Classification of Text Data Disaster Loss Category

Input: Weibo text

Procedure:

```
dict_categories = {
"affected individuals": ["suffer disaster", "common people", "..."],
"infrastructure and utility damage": ["house", "collapse", "..."],
"vehicle damage": ["vehicle", "Flooded", "..."],
```

Algorithm A1 *Cont.*

```
"rescue volunteering or donation effort": ["urgent", "rescue", "..."],
"other relevant information": ["ecology", "sabotage", "..."] }
stop_texts = ["is", "the", "..."]
texts = tokenize(text)
texts = [word for word in texts if word not in stop_texts]
category_counts = {category: 0 for category in dict_categories.keys()}
for word in texts:
for category, category_texts in dict_categories.items():
if word in category_texts:
category_counts[category] += 1
end if
end for
end for
categories = [category for category, count in category_counts.items() if count > 1]
```

Output: The category of disaster losses to which Weibo was judged to belong.

Algorithm A2: Text Loss Severity Calculation

Input: Weibo text
Procedure:

```
loss_dict = {10: ["die", …],
9: ["casualties", "injured", …],
8: ["accident", "disaster",…],
# ... Omit more scores and corresponding words.
1: ["sad", "difficult",…] }
thresholds = {" Serious loss ": 15, "Medium loss": 5, "Slight or no loss": 0}
def classify_text(text):
texts = segment_text(text)
total_score = 0
for word in texts:
for score, lexicons in loss_dict.items():
if word in lexicons:
total_score += score
end for
end for
if total_score >= thresholds["Serious loss "]:
return "Serious loss"
else if total_score >= thresholds["Medium loss"]:
return "Medium loss"
else if total_score > thresholds["Slight or no loss"]:
return "Slight or no loss"
else:
return "irrelevant"
end if
```

Output: the loss severity score corresponding to the text and its loss severity category

Algorithm A3: Qualitative Discrimination of Image Consistency

Input: Weibo text and corresponding image.
Procedure:

```
damage_categories = ["affected individuals", "infrastructure and utility damage", "vehicle damage", "rescue volunteering or donation effort", "other relevant information"]
severity_levels = ["Severe damage", "Mild damage", "Little or no damage"]
def is_consistent(picture, text):
picture_damage_category = classify_damage(picture)
picture_severity_level = classify_severity(picture)
text_damage_category = classify_text_damage(text)
text_severity_level = classify_text_severity(text)
damage_consistent_count = 0
severity_consistent_count = 0
if picture_damage_category == text_damage_category:
then damage_category is consistent
damage_consistent_count += 1
if picture_severity_level == text_severity_level:
then damage_level is consistent
severity_consistent_count += 1
return damage_consistent_count, severity_consistent_count
damage_consistent_count, severity_consistent_count = is_consistent (picture, text)
print("Number of consistent pictures and texts of loss categories:", damage_consistent_count)
print("Severity of loss Number of consistent pictures and texts:", severity_consistent_count)
```

Output: The number of text and picture loss categories and severity in Weibo is consistent.

Appendix B

The temporal distribution of rainstorm-related images in Hebei exhibits distinct phases (Figure A1). From 29 July to 1 August, other disaster loss imagery recurrently peaked at about 80 images, likely reflecting intense public engagement with early warnings, weather updates, and official announcements during the initial event phase. Subsequent infrastructure damage and relief operation imagery peaked at about 50 images on 2–3 August, coinciding with emerging disaster impacts and corresponding public discourse on road/bridge failures. A pronounced surge in disaster relief operations imagery (about 100 images) occurred on 5 August, signaling large-scale response mobilization as the rainfall subsided. Post-peak declines with minor fluctuations were observed across categories, while vehicle damage imagery remained consistently low throughout the event.

The temporal patterns of disaster severity imagery in the Beijing rainstorm reveal marked divergence between impact levels (Figure A2). Images depicting serious and medium loss exhibited substantial fluctuation throughout the event, peaking recurrently during 29–30 July afternoons and showing secondary surges on 31 July–1 August. Conversely, imagery of slight/no loss maintained stable low volumes with negligible fluctuation. This distribution asymmetry likely stems from selective public documentation, prioritizing high-impact scenarios and the potential misclassification of minor damage as "indistinguishable" cases in model inference.

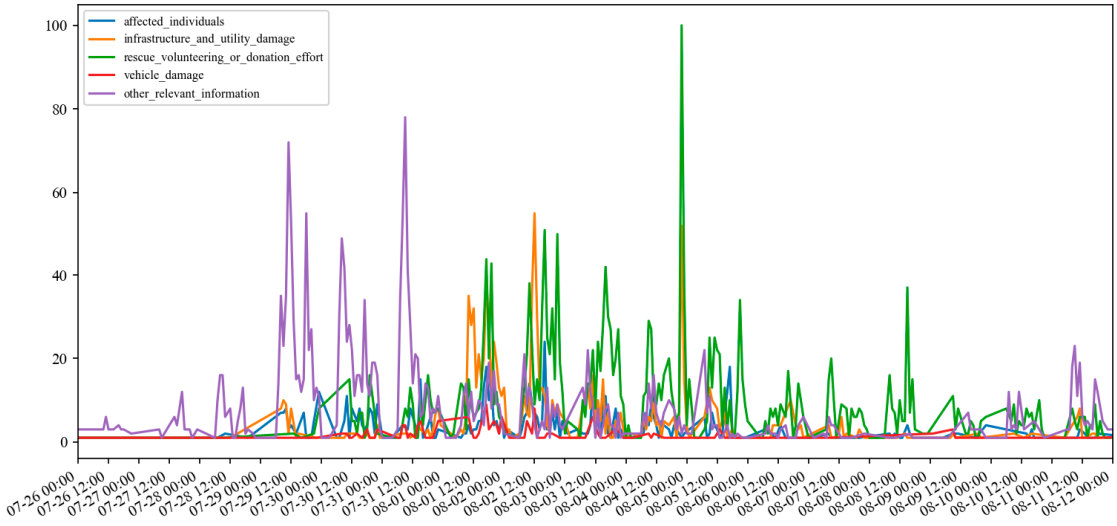


Figure A1. Variation trend of images of loss severity in Hebei.

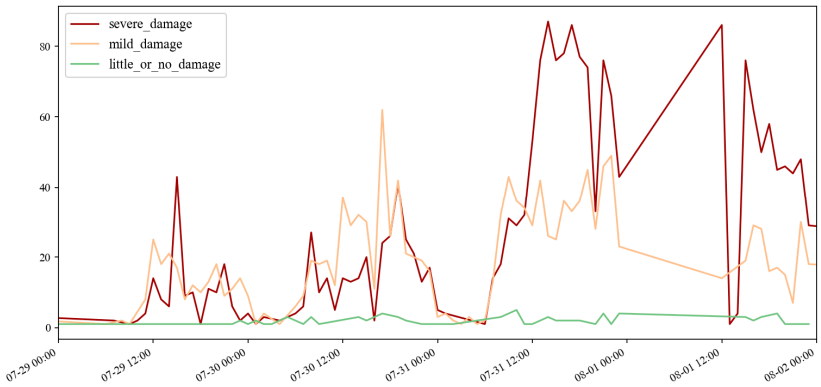


Figure A2. Variation trend of images of loss categories in Beijing.

An analysis of damage severity imagery trends during the Hebei rainstorm (26 July–12 August 2023) reveals distinct temporal patterns (Figure A3). Serious damage imagery exhibited the most significant variation: negligible until 28 July, increased from 29 July, and peaked between 2 and 6 August. Following a rapid decline after 7 August, levels stabilized minimally. Moderate damage imagery showed similar progression but lower magnitude: initial low counts until 28 July, moderate increases from 29 July, secondary peaks (3–4 August) likely reflecting expanded geographical impacts, followed by post-7 August stabilization. Slight/no damage imagery consistently persisted at negligible levels throughout the monitoring period with minimal fluctuation.

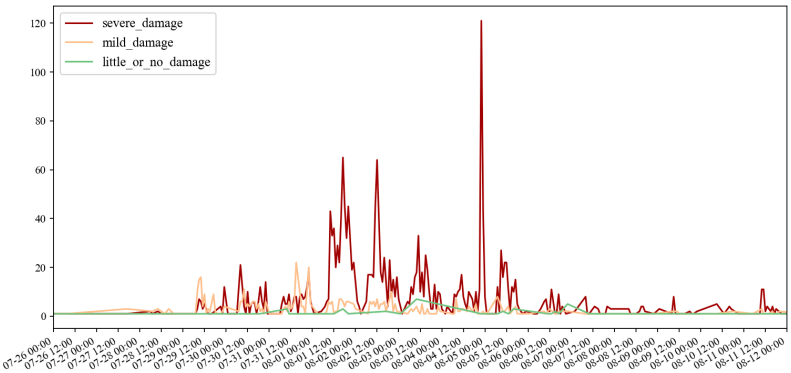
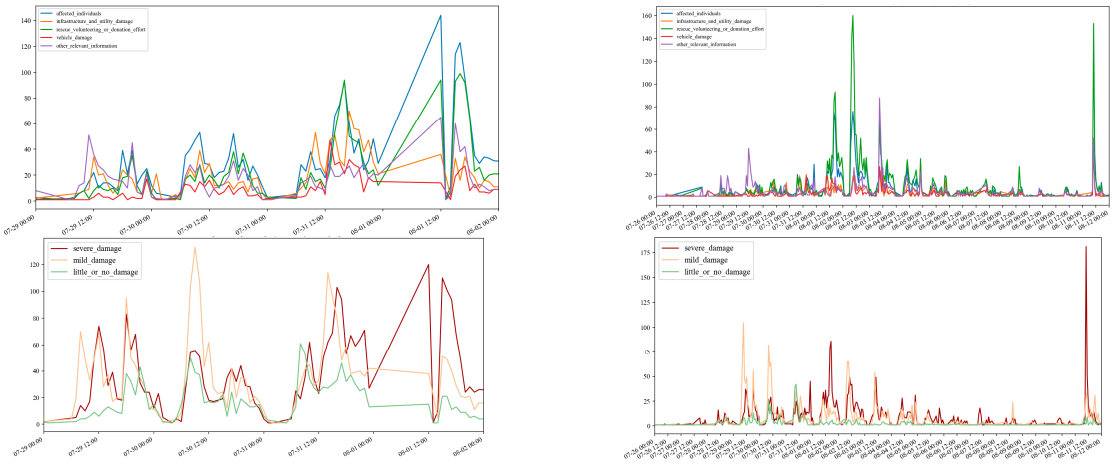


Figure A3. Variation trend of images of loss categories in Hebei.

With the passage of time, the change trend of the number of Weibo texts of disaster loss information in Beijing and Hebei is shown in Figure A4. According to the change trend of the text quantity of the rainstorm disaster loss category in Beijing, it can be seen that the peak values of text data related to the rainstorm disaster in Beijing are mostly reached in two categories, affected people and disaster relief operations, and the peak fluctuation of disasters has a daily trend, with the highest peak reached on 1 August, and the category is related to affected people, with the peak number of Weibo texts exceeding 140. However, there is a big difference between Hebei and Beijing in the quantitative change trend of rainstorm disaster loss categories. The highest peak in the whole time period is disaster relief action, which was reached between 31 July and 3 August. The highest peak number of texts on Weibo exceeded 1000, and the category of people affected by the storm also reached its peak, and the highest peak number of texts on Weibo exceeded 500. It can be seen that the period from 31 July to 3 August is the peak period of disaster-related information on Weibo, Hebei Province, and most disaster-related Weibo categories are related to disaster relief operations.

According to the line chart of the text quantity of rainstorm disaster loss severity in Beijing, it can be seen that the text quantity of the different levels of severity has obvious daily fluctuation characteristics. On 29 July, the number of texts with medium loss and serious loss reached the peak, respectively, and at noon on 30 July, the number of texts with medium loss reached the peak, with the peak number exceeding 120. Subsequently, on 31 July and 1 August, the number of severely damaged and moderately damaged texts still fluctuated obviously. At noon on 1 August, the number of severely damaged texts reached two peaks, and the peak number of texts was about 120. Then the number gradually decreased. According to the line chart of the text quantity of rainstorm disaster loss severity in Hebei Province, it can be seen that the fluctuation of the text quantity of severe loss and medium loss also has obvious daily characteristics from 29 July to 4 August. On 29–30 July, the medium loss first reached its peak, with the number of peak texts exceeding 100, and then the heavy loss and medium loss texts alternately reached their peak values, with the number of peak texts all within 100. After 5 August, the severity of text loss tends to decrease gradually, fluctuating in a small range, with a peak number of about 20.

On 12 August, serious losses reached the peak of all time periods, and the number of peaks exceeded 175. Subsequently, the number of texts of all kinds of loss severity decreased to fade.



(**a**) (**b**)

Figure A4. Variation trend of disaster information texts in Beijing (**a**) and Hebei (**b**).

References

1. Chen, D.; Su, W.; Wu, P.; Hua, B. Joint multimodal sentiment analysis based on information relevance. *Inf. Process. Manag.* **2023**, *60*, 103193. [CrossRef]
2. Tang, J.; Yang, S.; Wang, W. Social media-based disaster research: Development, trends, and obstacles. *Int. J. Disaster Risk Reduct.* **2021**, *55*, 102095. [CrossRef]
3. Wu, Z.; Chen, L.; Song, Y. A Model for Classifying Emergency Events Based on Social Media Multimodal Data. In Proceedings of the International Work-Conference on Artificial Neural Networks, Ponta Delgada, Portugal, 19–21 June 2023; pp. 316–327.
4. Sørensen, K. Lack of alignment in emergency response by systems and the public: A Dutch disaster health literacy case study. *Disaster Med. Public Health Prep.* **2022**, *16*, 25–28. [CrossRef] [PubMed]
5. Goodchild, M.F. Citizens as sensors: The world of volunteered geography. *GeoJournal* **2007**, *69*, 211–221. [CrossRef]
6. Min, S.; Ahuja, R.; Liu, Y.; Zaidi, A.; Phu, C.; Nocera, L.; Shahabi, C. CrowdMap: Spatiotemporal Visualization of Anonymous Occupancy Data for Pandemic Response. In Proceedings of the 29th International Conference on Advances in Geographic Information Systems, Beijing, China, 2–5 November 2021; pp. 630–633.
7. Radford, A.; Kim, J.W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J. Learning transferable visual models from natural language supervision. In Proceedings of the International Conference on Machine Learning, Online, 18–24 July 2021; pp. 8748–8763.
8. Luo, H.; Meng, X.; Zhao, Y.; Cai, M. Exploring the impact of sentiment on multi-dimensional information dissemination using COVID-19 data in China. *Comput. Hum. Behav.* **2023**, *144*, 107733. [CrossRef]
9. Shan, S.; Zhao, F.; Wei, Y.; Liu, M. Disaster management 2.0: A real-time disaster damage assessment model based on mobile social media data—A case study of Weibo (Chinese Twitter). *Saf. Sci.* **2019**, *115*, 393–413. [CrossRef]
10. Hao, H.; Wang, Y. Leveraging multimodal social media data for rapid disaster damage assessment. *Int. J. Disaster Risk Reduct.* **2020**, *51*, 101760. [CrossRef]
11. Wu, K.; Wu, J.; Li, Y. Mining typhoon victim information based on multi-source data fusion using social media data in China: A case study of the 2019 Super Typhoon Lekima. Geomatics. *Nat. Hazards Risk* **2022**, *13*, 1087–1105. [CrossRef]
12. Shan, S.; Zhao, F.; Wei, Y. Real-time assessment of human loss in disasters based on social media mining and the truth discovery algorithm. *Int. J. Disaster Risk Reduct.* **2021**, *62*, 102418. [CrossRef]
13. Xing, Z.; Zhang, X.; Zan, X.; Xiao, C.; Li, B.; Han, K.; Liu, Z.; Liu, J. Crowdsourced social media and mobile phone signaling data for disaster impact assessment: A case study of the 8.8 Jiuzhaigou earthquake. *Int. J. Disaster Risk Reduct.* **2021**, *58*, 102200. [CrossRef]
14. Xie, S.; Hou, C.; Yu, H.; Zhang, Z.; Luo, X.; Zhu, N. Multi-label disaster text classification via supervised contrastive learning for social media data. *Comput. Electr. Eng.* **2022**, *104*, 108401. [CrossRef]
15. Zhang, X.; Ma, Y. An ALBERT-based TextCNN-Hatt hybrid model enhanced with topic knowledge for sentiment analysis of sudden-onset disasters. *Eng. Appl. Artif. Intell.* **2023**, *123*, 106136. [CrossRef]
16. Li, J.; Cai, R.; Tan, Y.; Zhou, H.; Sadick, A.-M.; Shou, W.; Wang, X. Automatic detection of actual water depth of urban floods from social media images. *Measurement* **2023**, *216*, 112891. [CrossRef]
17. Liu, Y.; Li, Z.; Zhou, K.; Zhang, L.; Li, L.; Tian, P.; Shen, S. Scanning, attention, and reasoning multimodal content for sentiment analysis. *Knowl. -Based Syst.* **2023**, *268*, 110467. [CrossRef]
18. Bryan-Smith, L.; Godsall, J.; George, F.; Egode, K.; Dethlefs, N.; Parsons, D. Real-time social media sentiment analysis for rapid impact assessment of floods. *Comput. Geosci.* **2023**, *178*, 105405. [CrossRef]
19. Cheng, M.-Y.; Khasani, R.R.; Citra, R.J. Image-based preliminary emergency assessment of damaged buildings after earthquake: Taiwan case studies. *Eng. Appl. Artif. Intell.* **2023**, *126*, 107164. [CrossRef]
20. Ge, J.; Tang, H.; Yang, N.; Hu, Y. Rapid identification of damaged buildings using incremental learning with transferred data from historical natural disaster cases. *ISPRS J. Photogramm. Remote Sens.* **2023**, *195*, 105–128. [CrossRef]
21. Wang, Y.; Jing, X.; Cui, L.; Zhang, C.; Xu, Y.; Yuan, J.; Zhang, Q. Geometric consistency enhanced deep convolutional encoder-decoder for urban seismic damage assessment by UAV images. *Eng. Struct.* **2023**, *286*, 116132. [CrossRef]
22. Tasci, B.; Acharya, M.R.; Baygin, M.; Dogan, S.; Tuncer, T.; Belhaouari, S.B. InCR: Inception and concatenation residual block-based deep learning network for damaged building detection using remote sensing images. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *123*, 103483. [CrossRef]
23. Islam, M.A.; Rashid, S.I.; Hossain, N.U.I.; Fleming, R.; Sokolov, A. An integrated convolutional neural network and sorting algorithm for image classification for efficient flood disaster management. *Decis. Anal. J.* **2023**, *7*, 100225. [CrossRef]
24. Xing, Z.; Yang, S.; Zan, X.; Dong, X.; Yao, Y.; Liu, Z.; Zhang, X. Flood vulnerability assessment of urban buildings based on integrating high-resolution remote sensing and street view images. *Sustain. Cities Soc.* **2023**, *92*, 104467. [CrossRef]
25. Huang, C.; Zhang, J.; Wu, X.; Wang, Y.; Li, M.; Huang, X. TeFNA: Text-centered fusion network with crossmodal attention for multimodal sentiment analysis. *Knowl. Based Syst.* **2023**, *269*, 110502. [CrossRef]

26. Pandey, A.; Vishwakarma, D.K. VABDC-Net: A framework for Visual-Caption Sentiment Recognition via spatio-depth visual attention and bi-directional caption processing. *Knowl. -Based Syst.* **2023**, *269*, 110515. [CrossRef]
27. Zeng, Z.; Sun, S.; Li, Q. Multimodal negative sentiment recognition of online public opinion on public health emergencies based on graph convolutional networks and ensemble learning. *Inf. Process. Manag.* **2023**, *60*, 103378. [CrossRef]
28. Lin, H.; Zhang, P.; Ling, J.; Yang, Z.; Lee, L.K.; Liu, W. PS-mixer: A polar-vector and strength-vector mixer model for multimodal sentiment analysis. *Inf. Process. Manag.* **2023**, *60*, 103229. [CrossRef]
29. Wang, Z.; Liu, X.; Li, H.; Sheng, L.; Yan, J.; Wang, X.; Shao, J. Camp: Cross-modal adaptive message passing for text-image retrieval. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 5764–5773.
30. Liu, C.; Mao, Z.; Liu, A.-A.; Zhang, T.; Wang, B.; Zhang, Y. Focus your attention: A bidirectional focal attention network for image-text matching. In Proceedings of the 27th ACM International Conference on Multimedia, Nice, France, 21–25 October 2019; pp. 3–11.
31. Diao, H.; Zhang, Y.; Ma, L.; Lu, H. Similarity reasoning and filtration for image-text matching. In Proceedings of the AAAI Conference on Artificial Intelligence, Online, 2–9 February 2021; pp. 1218–1226.
32. Rizk, Y.; Jomaa, H.S.; Awad, M.; Castillo, C. A computationally efficient multi-modal classification approach of disaster-related Twitter images. In Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing, Limassol, Cyprus, 8–12 April 2019; pp. 2050–2059.
33. Mouzannar, H.; Rizk, Y.; Awad, M. Damage Identification in Social Media Posts using Multimodal Deep Learning. In Proceedings of the ISCRAM, Rochester, NY, USA, 4–7 November 2018.
34. Kumar, A.; Singh, J.P.; Dwivedi, Y.K.; Rana, N.P. A deep multi-modal neural network for informative Twitter content classification during emergencies. *Ann. Oper. Res.* **2022**, *319*, 1–32. [CrossRef]
35. Madichetty, S.; Muthukumarasamy, S.; Jayadev, P. Multi-modal classification of Twitter data during disasters for humanitarian response. *J. Ambient Intell. Humaniz. Comput.* **2021**, *12*, 10223–10237. [CrossRef]
36. Ceylan, G.; Diehl, K.; Proserpio, D. Words meet photos: When and why photos increase review helpfulness. *J. Mark. Res.* **2024**, *61*, 5–26. [CrossRef]
37. Shin, D.; He, S.; Lee, G.M.; Whinston, A.B.; Cetintas, S.; Lee, K.-C. Enhancing social media analysis with visual data analytics: A deep learning approach. *MIS Q. Manag. Inf. Syst.* **2020**, *44*, 1459–1492. [CrossRef]
38. Ofli, F.; Alam, F.; Imran, M. Analysis of social media data using multimodal deep learning for disaster response. *arXiv* **2020**, arXiv:2004.11838 2020.
39. Alam, F.; Ofli, F.; Imran, M. Crisismmd: Multimodal twitter datasets from natural disasters. In Proceedings of the International AAAI Conference on Web and Social Media, Palo Alto, CA, USA, 25–28 June 2018.
40. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient estimation of word representations in vector space. *arXiv* **2013**, arXiv:1301.3781 2013.
41. Lu, Y.; Liu, Q.; Dai, D.; Xiao, X.; Lin, H.; Han, X.; Sun, L.; Wu, H. Unified structure generation for universal information extraction. *arXiv* **2022**, arXiv:2203.12277 2022.
42. Cai, M.; Luo, H.; Meng, X.; Cui, Y.; Wang, W. Influence of information attributes on information dissemination in public health emergencies. *Humanit. Soc. Sci. Commun.* **2022**, *9*, 1–22. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.