논문 23-48-12-12 The Journal of Korean Institute of Communications and Information Sciences '23-12 Vol.48 No.12
https://doi.org/10.7840/kics.2023.48.12.1633

Automatic Classification of Disaster Images Based on Deep Learning

Hojun Song◆, Dong-hun Lee*,
Han-gyul Baek*, Byungjun Bae**,
Sang-hyo Park°

ABSTRACT

Recently, persistent catastrophic issues and advancements in science and technology have increased the need for disaster research. In this study, we propose a deep learning-based framework that distinguishes disaster images from large-scale datasets, thereby providing access to disaster-related image data. To construct an accurate dataset for our framework, disaster images were manually collected and labeled from various open datasets. Image generation and augmentation techniques were used to supplement the insufficient training dataset and enhance the classifier training of our classification framework. We built a classification framework that demonstrates over 99% accuracy in classification experiments using open datasets.

Key Words : Disaster Image Data, Image Classification, Image Augmentation, Image Generation

Ⅰ. Introduction

In recent years, significant advancements have been made in the fields of image retrieval and captioning, primarily through transformer models[1] and contrastive language-image pretraining learning frameworks[2]. The development of these models[1,2] has been applied to various fields such as medical care, economy, and transportation. However, their application in the disaster field has been minimal.

For optimal learning in a specific field, a sufficient quantity of related data is essential[3]. The major challenges in the disaster field are the lack of specific disaster-related data and the difficulty of extracting proper disaster images from large-scale daasets. Specifically, an insufficient and low quality of image data for model training leads to overfitting issues. Therefore, conducting research related to disaster situations requires not only the collection of adequate disaster images, but also the labor-intensive and time-consuming task of clearly classifying and labeling them.

To address the challenges in collecting labeled disaster research data, we propose an automatic disaster image classification framework designed to create and expand disaster image datasets. Our classification framework efficiently classified disaster data from large-scale datasets, as evidenced by the experimental results. The disaster datasets obtained using our framework are widely used in disaster research.

Ⅱ. Proposed method

2.1 Dataset

Images related to disaster keywords were collected from well-known benchmark datasets, such as MS-COCO[4], Open Images Dataset V4[5], and Flickr30k[6], and supplemented with images sourced from the web. We defined representative keywords related to disaster situations as fire, earthquake, tempest, thunderstorm, fog, and snowstorms. Annotation captions and disaster scenario keywords were used for seperating disaster images from benchmark datasets. However, incorporated images that lack relevance to disaster situations inevitably

※ This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (2022-0-00083, Development of customized disaster media service platform technology for the vulnerable in disaster information awareness).
◆ First Author : School of Computer Science and Engineering, Kyungpook National University, hojunking@knu.ac.kr, 학생회원
° Corresponding Author : School of Computer Science and Engineering, Kyungpook National University, s.park@knu.ac.kr, 정회원
* School of Computer Science and Engineering, Kyungpook National University, s.park@knu.ac.kr, 정회원
** Media Research Division, Electronics and Telecommunications Research Institute, 1080i@etri.re.kr, 정회원
논문번호 : 202309-071-B-LU, Received September 4, 2023; Revised September 18, 2023; Accepted September 28, 2023

require manual classification for refinement purposes. To address the limitations of the existing datasets, we employed web crawling techniques that target images corresponding to representative disaster-related keywords.

Preprocessing was performed to standardize the dimensions of the collected disaster images by resizing them to a consistent resolution of 512 × 512 pixels. Subsequently, as shown in Figure 1, we employed a range of augmentation techniques to address the overfitting issues resulting from limited data, including contrast adjustment, image rotation, and horizontal flipping[7]. Finally, we refined the images by cropping them to 256 × 256 pixels and focused on extracting essential information from the central regions of the images.

In addition, a disaster image was created using a stable diffusion model[8], which is a text-to-image generation model that has recently shown excellent performance.

The text required for image generation was obtained using captions related to disaster keywords from the MS-COCO[4], Open Images[5], and Flickr30k[6] annotations. This study sets itself apart from conventional image classifier training methodologies by leveraging generated disaster image data as a training resource.

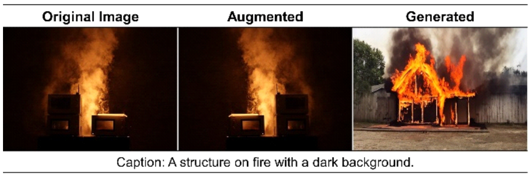

Fig. 1. Example of an original collected disaster image, alongside its augmented and generated versions.

2.2 Framework

As shown in Figure 2, our framework comprises preprocessing, a model architecture, and an auto-labeling process. We utilized disaster image data sourced from both benchmark datasets and the web, augmented and generated images as inputs for the model architecture. Our approach is distinct from traditional image classifier training because it leverages disaster images generated through a text-to-image model within the training data, emphasizing the key differentiation from conventional methodologies.

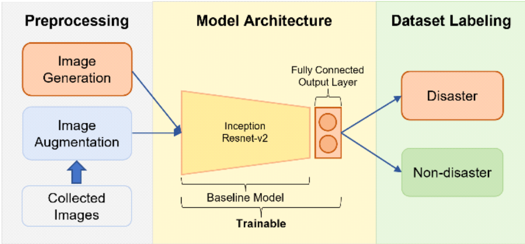

Fig. 2. Proposed Process of Disaster Image Classification Framework.

After preprocessing, Images with a resolution of 256 × 256 pixels were used as inputs for the model architecture. The architecture's baseline model utilized Inception-ResNet-v2[9], which is known for its exceptional image classification performance in 2017. A fully connected layer comprising of two nodes was added for non-disaster and disaster image classification. The training process involved fine-tuning all layers, including the baseline model and the additional layer. This process ultimately results in the labeling of non-disaster and disaster images.

Ⅲ. Experiments

Our experiments were conducted on Ubuntu 20.04 LTS using NVIDIA RTX 3070 and NVIDIA RTX 3060TI GPUs in parallel.

Table 1 lists the disaster image datasets used in this study. Excluding text-to-image generated images, we divided the 2,055 images into an approximately 4:1 train-test split. The initial 554 disaster images of the benchmark datasets were physically augmented to 7,174 images, and 3,437 images were generated using the diffusion model, resulting in 10,057 disaster images for model training. For non-disaster images, we collected 60,000 images from benchmark datasets. To tune the learning process, by employing k-fold cross-validation, the total training dataset of 70,057 images was divided into a 4:1 ratio, testing 20,000 random non-disaster images from Open Images[5] and 400 designated disaster images.

Table 1. Disaster Image Dataset

	Train	Test	Total
Benchmark Dataset	554	80	2,135
Web	1,181	320	
Generated	1,782	-	1,782
Gathered + Generated	3,437	400	3,837
Benchmark Augmented	7,174	-	7,174
Total	10,057	400	10,457

Table 2. Results (Accuracy-1, Precision, Recall, F1 Score)

	Accuracy	Precision	Recall	F1 Score
Test Data	98.93%	99.65%	99.26%	87.31%
MS-COCO[4]	99.76%	99.98%	99.79%	76.81%
Open Image[5]	99.24%	99.99%	99.24%	52.12%
Flickr30k[6]	99.59%	100%	99.59%	65.92%
Benchmark Average	99.38%	99.90%	99.47%	70.54%

Table 2 presents the performance of the proposed classifier for disaster image classification on the test and benchmark datasets. Test data were produced using the method described in Table 1. Each benchmark dataset included all disaster images collected from the datasets for training and testing purposes. High performance in terms of accuracy was observed because of the small proportion of disaster images within these extensive datasets. The macro F1 score was comparatively lower than the other metrics, which was attributed to the imbalance between the non-disaster and disaster images in the test dataset. Consequently, Open Images Dataset V4[5], which has the most significant imbalance, had the lowest F1 score. The average F1 score of the benchmark datasets was lower than that of the test dataset, possibly due to the test dataset contraining a number of disaster-related images.

Ⅳ. Conclusion

The study of disaster situations requires consistent research across various fields. In line with this, we anticipate that our disaster classification framework for collecting disaster image data will play a crucial role. By streamlining the collection and labeling of disaster data, this classifier aids in advancing deep learning applications, including image retrieval and captioning. Our future research will focus on automatic disaster classification based on various keywords, aiming to promote its application in the disaster field through advancements in deep learning.

References

[1] L. Zhou, Y. Zhou, J. J. Corso, R. Socher, and C. Xiong, “End-to-end dense video captioning with masked transformer,” in *Proc. IEEE Conf. CVPR*, pp. 8739-8748, Utah, USA, Jun. 2018. (https://doi.org/10.48550/arXiv.1804.00819)

[2] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, and G. Krueger, “Learning transferable visual models from natural language supervision,” *ICML*, pp. 8748-8763, Feb. 2021. (https://doi.org/10.48550/arXiv.2103.00020)

[3] T. Kim, J. Bang, J. Seo, and K. Sohn, “Object detection for Fire Disaster Situation Recognition,” in *Proc. Annual Conf. of KIPS 2022(*ACK 2022), pp. 426-428, Chuncheon, Korea, Nov. 2022. (https://doi.org/10.3745/PKIPS.y2022m11a.426)

[4] T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollar, and C. L. Zitnick, “Microsoft COCO: Common objects in context,” *Computer Vision–ECCV 2014*, pp. 740-755, Zurich, Switzerland, Sep. 2014. (https://doi.org/10.1007/978-3-319-10602-1_48)

[5] A. Kuznetsova, H. Rom, N. Alldrin, J. Uijlings, I. Krasin, J. Pont-Tuset, S. Popov, M. Malloci, A. Kolesnikov, T. Duerig, and V. Ferrari, “The open images dataset v4: Unified image classification, object detection, and visual relationship detection at scale,” *Int. J. Comput. Vision*, vol. 128, no. 7, pp. 1956-1981, Mar. 2020. (https://doi.org/10.1007/s11263-020-01316-z)

[6] P. Young, A. Lai, M. Hodosh, and J. Hockenmaier, "From image descriptions to visual denotations: New similarity metrics for semantic inference over event descriptions," *Trans. Assoc. for Computational Linguistics*, vol. 2, pp. 67-78, Feb. 2014. (https://doi.org/10.1162/tacl_a_00166)

[7] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, pp. 1-48, Jul. 2019. (https://doi.org/10.1186/s40537-019-0197-0)

[8] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proc. IEEE/CVF Conf. CVPR*, pp. 10684-10695, New Orleans, USA, Jun. 2022. (https://doi.org/10.48550/arXiv.2112.10752)

[9] C. Szegedy, S. Ioffe, V. Vanhouccke, and A. Alemi, "Inception-v4, inception-ResNet and the impact of residual connections on learning," in *Proc. AAAI Conf. Artificial Intell.*, vol. 31, no. 1, Feb. 2017. (https://doi.org/10.48550/arXiv.1602.07261)