Latent Representation Generation for Efficient Content-Based Image Retrieval in Weather Satellite Images Using Self-Supervised Segmentation

Abdulhakin Mohamud Ismail
Kookmin University
iamismail028@gmail.com

Ha-Myung Park
Kookmin University
hmpark@kookmin.ac.kr

***Abstract*—With the rapid advancement of satellite sensor technology, a huge volume of high-resolution image data sets can now be acquired, and efficiently representing and recognizing scenes within these high-resolution images has emerged as a critical challenge. In this paper, we propose AES (Auto-Encoder after Segmentation), a method for learning a latent representation containing high-level semantics of meteorological satellite images. AES uses W-Net, an unsupervised segmentation model, to simplify the images. Our approach significantly improves CBIR performance in the domain of weather satellite images. Experimental results on satellite weather images demonstrate that our proposed method outperforms traditional auto-encoders in terms of mean squared error (MSE). A lower average MSE indicates stronger similarity between the retrieved and query images, suggesting that our method better represents the underlying similarities within the data. Additionally, our approach demonstrates higher recall values for similarity assessment and enhanced retrieval accuracy, showcasing its effectiveness in extracting meaningful features for content-based image retrieval in meteorological satellite images.**

I. Introduction

Weather satellite imagery plays a pivotal role in monitoring and understanding meteorological phenomena that influence our planet's climate and weather patterns. With the advancement of remote sensing imaging technology, satellite information now contains more abundant data, and the quality of satellite cloud images has made significant progress. These images contain valuable information crucial for weather forecasting and the early prediction of various atmospheric disturbances, including typhoons and hurricanes.

How effectively to store these images in a database and how efficiently to retrieve them from the database are challenging and emerging research topics in the field of remote sensing. Traditional satellite cloud image retrieval, which relies on associating images with file names and sensor parameters, fails to describe actual image contents, such as cloud shapes [1]. Additionally, [2] introduces a shape-based retrieval system for infrared satellite images. Despite improvements in satellite image retrieval, challenges persist in recognizing complex features in high-resolution meteorological satellite images.

This work was funded by the Korea Meteorological Administration Research and Development Program "Developing Intelligent Assistant Technology and Its Application for Weather Forecasting Process" under Grant (KMA2021-00123). Ha-Myung Park is the corresponding author.

To address these challenges, we propose an approach called AES (Auto-Encoder after Segmentation), which bridges this gap by utilizing unsupervised segmentation with W-Net [3] to simplify images and generate latent representations.

II. Method

As we work to improve the way we find and retrieve weather satellite images, we ask ourselves, 'How does segmentation help in generating latent representations for content-based image retrieval in the domain of weather satellite images?' By addressing this question, we introduce a two-stage approach named Auto-Encoder after Segmentation, illustrated in Fig. 1. First, we utilize a segmentation model (W-Net) to process weather satellite images. This generates segmented images that simplify the original ones. Then, these segmented images are fed into our embedding model, enabling the generation of meaningful latent representations. The lower part of the architecture is dedicated to the auto-encoder. It is divided into an encoder (left side) and a corresponding decoder (right side). Our architecture consists of 10 convolutional layers, structured into 5 models in both the encoder and decoder.

III. Experiment

In this section, we present an empirical evaluation of four models: AES, Auto-Encoder, Y-Net, and W-Net. Our experiment focuses on addressing the following key questions:

Q1. Model Performance: Does the similarity between the latent representations generated by each model reflect the true image similarity well? Also, how well do the images found based on the similarity of the latent representations match the images found with the true image similarity? **(Section III-A)**

Q2. Qualitative Evaluation: Given an image of an anomaly, such as a tropical cyclone, how well do the models find images of similar weather phenomenon? **(Section III-B)**

A. Model Performance

In our experimental evaluation, we compare the performance of four models: AES, Auto-Encoder, Y-Net, and W-Net. The primary focus is on their ability to retrieve similar images from the training data set using pixel-by-pixel ground truth comparison. We utilize four key metrics for this evaluation: average mean square error (MSE), standard deviation (SD),

2375-9356/24/$31.00 ©2024 IEEE
DOI 10.1109/BigComp60711.2024.00078

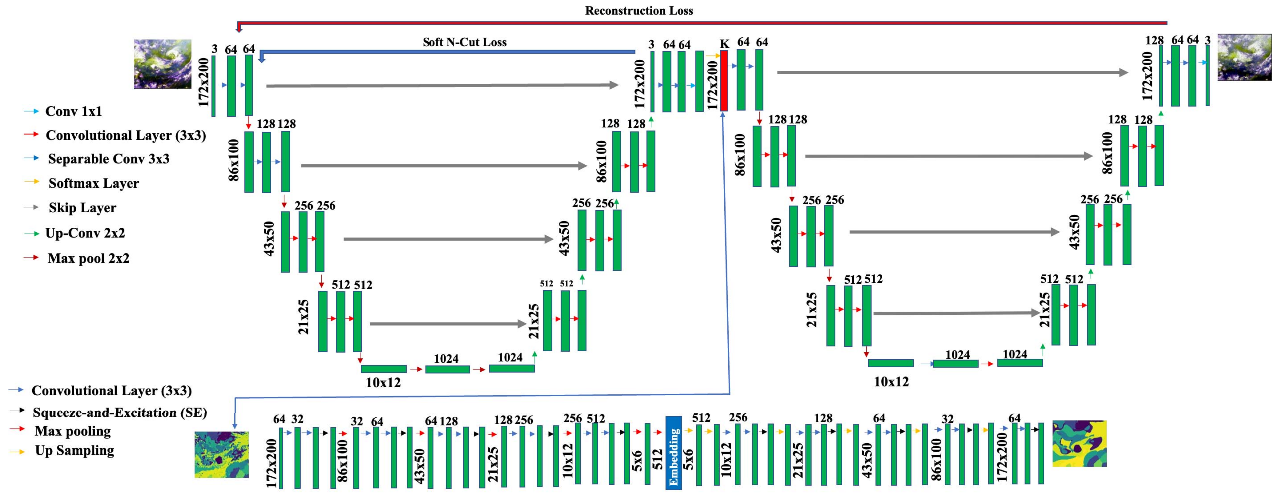

Fig. 1: Auto-Encoder after Segmentation.

correlation coefficient, and recall. For average MSE loss, we calculate the similarity between each query image in the test data set and its top 10 similar images within the training data set. A lower average MSE loss indicates a higher degree of similarity between the query image and its retrieval images. Additionally, we employ the standard deviation as a key measure to understand the variability in the difference between the latent representation MSE loss of models and the ground truth (pixel by pixel) MSE loss. We calculate residuals by subtracting pixel-by-pixel MSE losses from latent representation MSE losses in each model. A lower standard deviation suggests a more consistent level of similarity in retrieval performance. Furthermore, we consider the correlation coefficient to assess the linear relationship between the latent representation MSE losses predicted by the models and the ground truth (pixel by pixel) MSE losses. A higher correlation coefficient indicates a stronger linear relationship. Moreover, we evaluate the recall values of the models by assessing their ability to identify similar images for a given query image among the top 10. The obtained results are presented in Table I, where **AES** exhibits.

TABLE I: Model Performance Comparison

Model	MSE	SD	Corr.Coeff	Recall@10
W-Net	0.1050	0.0419	0.3273	5%
Y-Net	0.1654	0.0394	-0.6068	0.13%
Auto-Encoder	0.0987	0.0309	0.5440	28%
AES	**0.0503**	**0.0072**	**0.9654**	50%
Pixel-by-Pixel	0.0459	0.0	1	-

B. Qualitative Evaluation

In this section, we present a qualitative evaluation of our content-based image retrieval (CBIR) approach, focusing on the performance and effectiveness of four models: AES, Auto-Encoder, W-Net, and Y-Net are capable of retrieving similar weather phenomenon images from the training data set based on a given query image, as illustrated in Fig. 2. During evaluation, we utilize a query image captured by a satellite on October 4, 2018, at 18:00, featuring a tropical cyclone.

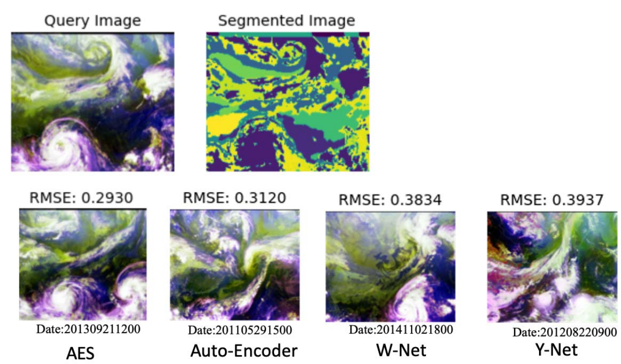

Fig. 2: Retrieval Images.

Our model excels at identifying similar weather phenomenon within retrieved images, outperforming other models in this regard.

IV. Conclusion

In this study, we introduce AES (Auto-Encoder after Segmentation), a novel representation learning method for efficient content-based image retrieval in weather satellite images. AES consistently outperforms alternatives, with a lower average MSE loss of 0.0503, a standard deviation of 0.0072, and a high correlation coefficient of 0.9654. AES model achieves a 50% recall rate in the top 10 retrieval images, compared to W-Net, Y-Net, and Auto-Encoder, which achieve 5%, 0.13%, and 28%, respectively.

References

[1] F. Dell'Acqua and P. Gamba, "Query-by-shape in meteorological image archives using the point diffusion technique," *IEEE transactions on geoscience and remote sensing*, vol. 39, no. 9, pp. 1834–1843, 2001.

[2] A. Ma and I. K. Sethi, "Local shape association based retrieval of infrared satellite images," in *Seventh IEEE International Symposium on Multimedia (ISM'05)*. IEEE, 2005, pp. 7–pp.

[3] X. Xia and B. Kulis, "W-net: A deep model for fully unsupervised image segmentation. arxiv 2017," *arXiv preprint arXiv:1711.08506*.