J Inf Process Syst, Vol.19, No.3, pp.289~301, June 2023
https://doi.org/10.3745/JIPS.04.0274

ISSN 1976-913X (Print)
ISSN 2092-805X (Electronic)

JOURNAL OF INFORMATION PROCESSING SYSTEMS JIPS

Spam Image Detection Model based on Deep Learning for Improving Spam Filter

Seong-Guk Nam, Dong-Gun Lee, and Yeong-Seok Seo*

Abstract

Due to the development and dissemination of modern technology, anyone can easily communicate using services such as social network service (SNS) through a personal computer (PC) or smartphone. The development of these technologies has caused many beneficial effects. At the same time, bad effects also occurred, one of which was the spam problem. Spam refers to unwanted or rejected information received by unspecified users. The continuous exposure of such information to service users creates inconvenience in the user's use of the service, and if filtering is not performed correctly, the quality of service deteriorates. Recently, spammers are creating more malicious spam by distorting the image of spam text so that optical character recognition (OCR)-based spam filters cannot easily detect it. Fortunately, the level of transformation of image spam circulated on social media is not serious yet. However, in the mail system, spammers (the person who sends spam) showed various modifications to the spam image for neutralizing OCR, and therefore, the same situation can happen with spam images on social media. Spammers have been shown to interfere with OCR reading through geometric transformations such as image distortion, noise addition, and blurring. Various techniques have been studied to filter image spam, but at the same time, methods of interfering with image spam identification using obfuscated images are also continuously developing. In this paper, we propose a deep learning-based spam image detection model to improve the existing OCR-based spam image detection performance and compensate for vulnerabilities. The proposed model extracts text features and image features from the image using four sub-models. First, the OCR-based text model extracts the text-related features, whether the image contains spam words, and the word embedding vector from the input image. Then, the convolution neural network-based image model extracts image obfuscation and image feature vectors from the input image. The extracted feature is determined whether it is a spam image by the final spam image classifier. As a result of evaluating the F1-score of the proposed model, the performance was about 14 points higher than the OCR-based spam image detection performance.

Keywords

Classification, Deep Learning, Image Processing, Image SPAM, Obfuscated Feature, SPAM

1. Introduction

In modern society, due to the development of hardware devices, technologies, and services, most people can easily communicate with culture through social network service (SNS) using personal computers (PCs) or smart phones. Users' SNS usage time also showed a dramatic increase compared to before [1-4]. As increasing usage amount, a lot of content was generated. As a result, the contents used

Manuscript received January 14, 2021; first revision November 29, 2021; second revision January 19, 2022; third revision February 15, 2022; accepted February 26, 2022.

* **Corresponding Author:** Yeung-Seok Seo (ysseo@yu.ac.kr)
Dept. of Computer Engineering, Yeungnam University, Gyeongsan, Korea (sd05031@yu.ac.kr, dklee77@ynu.ac.kr, ysseo@yu.ac.kr)

to communicate with other users increased. However, contents making users that use the service feel uncomfortable also occurred. One of the contents that makes users who use the service feel uncomfortable is spam. Spam is information that the spammer sends for a specific purpose, but the recipient refuses to receive it because it does not want, is not of interest, or is unnecessary [5,6]. It is usually used for the purpose of advertising to many unspecified people. Now, SNS is a good place for spammers to disseminate spam content efficiently with a low cost. If users access the SNS service currently, they will be able to find the problem frequently. And spammers try various methods to disseminate spam content more effectively, one of which is disseminating spam with other formats than text. Currently, the trend of spam on SNS is changing from spam with a text format to creating and exposing image spam created by inserting text into images. As the form of spam content changes from a text format to an image format, the screen size of spam contents and the volume of spam data increased. Spam that disseminated randomly prevents users from using the service and causes problems that lead to poor service quality. Also, spammers do not consider the age of the user. It is also a big problem to expose bad content (adult, gambling, and drug) to people who use SNS, especially teenagers. As spam changes into visual content, it can also be seen frequently when sexually exciting images are added or used as backgrounds. Various attempts have been made to identify image spam or spam content. However, when the techniques were proposed, spam in email services became a social problem, and there are many studies focused on filtering spam in email systems. In the existing spam detection studies, there have been many attempts to analyze the contents using optical character recognition (OCR) for the detection of spam with image form. But spammers also attempted various modifications to spam images in order to neutralize the filtering of spam images with OCR. The development of OCR technology will enable flexible processing of even slightly modified images. The spammers will also disseminate image spam that neutralizes developed OCR through various methods.

Fig. 1. The image example derived by affine transform and Gaussian blurring.

For example, Fig. 1 shows affine transformation on X, Y and Gaussian blurring on the whole image. It was impossible to extract letters through OCR, but reading through the human's eye is possible. Because images can be modified by combining various techniques, as shown in the example, it is impossible to predict and cope with all of the methods of producing image spam. For this reason, it is not a good idea to attempt to identify image spam using only OCR. If the image is modified to a level that the machine cannot read, it is called an obfuscated image and defined as a spam image.

Since obfuscated images interfere with the text extraction function of OCR, the performance of text-based spam image detection techniques can be greatly reduced. Therefore, it is necessary to respond to the obfuscated image from which text information cannot be extracted, and additional information that can recognize spam only with image information will be needed.

Therefore, in this paper, to improve the performance of the existing spam filter that uses only OCR, we propose a deep learning-based spam image detection model for improving the performance of spam filters. The proposed model improves the identification performance of spam images by adding three sub-models to the existing OCR and Apache SpamAssassin [7] based spam detection model. The added three sub-models consist of an obfuscated image classifier, a spam image feature extractor based on deep learning, and a spam text feature extractor based on word embedding. The result value of each sub-model is merged into one feature vector and transmitted to the spam image detection model to output the final result.

Following the introduction in Section 1, this paper introduces related research in Section 2 and the proposed model in Section 3. Next, in Section 4, the performance evaluation of the proposed technique is carried out, and the conclusion is drawn in Section 5.

2. Related Work

Various studies have been conducted from the past to identify spam composed of images. Biggio et al. [8] assesses that any component that interferes with OCR may be spam. In order to detect components that interferes with OCR, the image was analyzed at a low level and noise around the character area was attempted. In [9], in order to cope with various classic image transformations used to neutralize OCR technology, spam images collected in database were detected near-duplicate, and spam images were identified through similarity between images. Barbar and Ismail [10] mainly focused on identifying spam mail through the authentication system process of the email service in order to solve the spam problem that occurs in the email service. In this study, there was also an image spam identification process through OCR. In order to improve the precision of the OCR reading result, the result obtained from the context analysis spell check was used when determining that there was a problem with the OCR result. In the study of Fatichah et al. [11], using several layers of convolution neural network (CNN) and many data sets, features of spam images were extracted and spam was identified through neural networks without manual processes. Although various studies have been conducted to identify image spam before, there have been no studies that can flexibly cope with obfuscated images. The absence of a process to check obfuscation has a problem of creating a loophole in image spam identification and reducing performance.

Recently, spam detection studies from various perspectives have been conducted with a dazzlingly growing artificial intelligence (AI) technique-based approach. Imam and Vassilakis [12] conducted the learning of OCR models based on effective and accurate scene text detector (EAST) and convolutional recurrent neural Network (CRNN) to extract Arabic included in spam images, and classified spam words in embedded Arabic images. There was also a study conducted using features of images, not text. Singh [13] performed image classification using 38 features of the image, such as image resolution, average pixel, and histogram properties, as inputs of neural network and deep neural network (DNN), and compared with the classification results of CNN models using images as inputs. In addition, there was a challenge to improve the performance of spam detection models in an environment where training datasets were insufficient. Rao and Gopalapillai [14] showed the possibility of spam image classification

models with less training and computing power through transfer learning of large-scale image classification models such as VGG16, VGG19, and MobileNet. Fan and Yang [15] suggested a method for expanding insufficient data sets for training. Among the data augmentation methods, k-means clustering was applied to change the image size to automatically find a good size and ratio for training. In addition, the accuracy of image classification models was analyzed using the augmented dataset. Sharmin et al. [16] constructed an instrument dataset using edge detecting technology to expand insufficient datasets, and analyzed the spam image classification performance of support vector machine (SVM), multilayer perceptron, and CNN. Recent studies described above have shown that AI-based spam detection is very useful and valuable, and predict that the performance of spam detection will continue to improve due to the development of AI technology in the future.

3. Proposed Spam Image Detection Model

This section describes the obfuscated spam image detection model to improve the spam filter. Fig. 2 shows an overview of the proposed model. The input image is transmitted to four sub-models. Among them, the first spam classifier that processes text and the spam text feature extraction model receive preprocessed text by OCR rather than images. The first spam classifier is a technology that applies the existing OCR and spam assay, and outputs the first spam classification result as “True” or “False” from the input text. Spam text feature extractor converts spam features from input text into word embedding vectors and outputs them. The obfuscated image classifier detects whether the input image is distorted by an external operation and outputs whether it is obfuscated as “True” or “False.” Finally, the spam image feature extractor extracts the spam feature from the input image as a vector and outputs it. The output of each sub-model is merged into one vector, and these are passed to the final spam classifier to determine whether it is a final spam image.

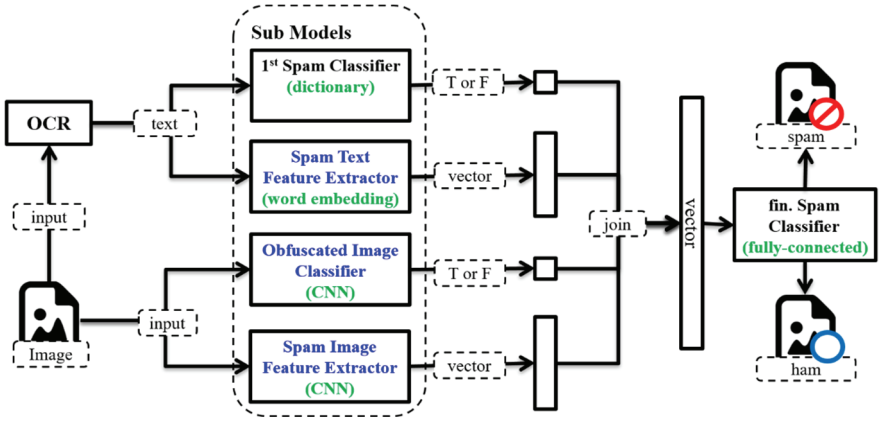


Fig. 2. Summary of proposed spam image detection model.

3.1 First Spam Classifier

The first spam classifier is a spam image classifier based on Apache SpamAssassin [7]. The classifier receives text extracted from the OCR and detects if there is a word related to spam. It extracts a list of

words in English by using a regular expression (Regex) for the text output from the OCR that received the image. Then, the output word list is compared with the pre-configured spam word dictionary to evaluate whether the word is related to spam and output whether the input image is spam or not. The spam word dictionary was constructed by extracting words from regular expressions that perform string filtering from the source code provided by Apache SpamAssassin.

3.2 Spam Text Feature Extractor

Currently, spam is transmitted in a variety of ways and is aimed at a wider target, making it very difficult to collect all the words used in spam, and spam filters using only spam dictionaries are not perfect. Therefore, the spam text features extractor extracts features of spam text using word embedding to improve the performance of spam filters based on spam dictionaries. Fig. 3 shows the process of the spam text features extractor. First, the text extracted from the OCR is converted into a token in the form of a word through Tokenizer. Each token is converted into an integer corresponding to each word through the vocabulary of TorchText [17] and then transferred to the embedding bag of TorchText. The embedding bag generates and outputs embedding vectors of each word.

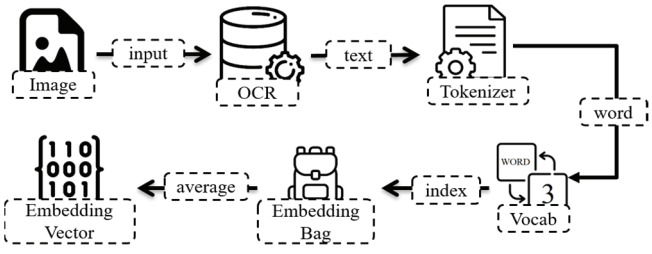


Fig. 3. Spam text feature extractor.

3.3 Obfuscated Image Classifier

The obfuscated image classifier detects obfuscated images that interfere with text recognition of OCR. Affine transformation [18], Gaussian blurring [19], distortion transformation, and completely automated public touring test to tell computers and Humans Apartment (CAPTCHA [20]) are representative methods of obfuscating images. To train the obfuscated image classifier, an obfuscated image data set was created using the image distortion techniques described above using OpenCV2 [21]. In addition, an obfuscated image classifier was created by training the dataset generated in the CNN model. Fig. 4 shows the work process of the obfuscated image classifier.

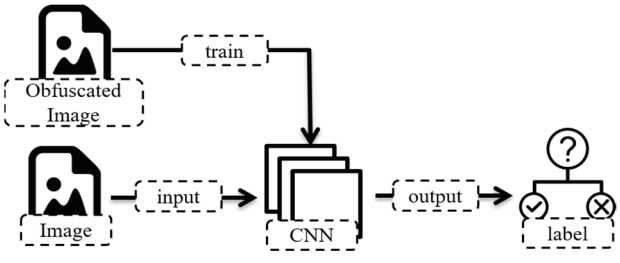


Fig. 4. Obfuscated image classifier.

3.4 Spam Image Feature Extractor

Text-based spam filters do not perform well unless accurate text is provided. Therefore, in order to overcome the limitations of text-based spam filters, information other than text must be able to be utilized to the fullest. The spam image feature extractor described in this section extracts various features of spam from images and convolution layers and pooling layers within the CNN model and outputs them as vectors. The CNN for the spam image feature extractor consists of four convolutional layers and Max pooling layers, and rectified linear unit (ReLU) activation function. The model is shown in Fig. 5.

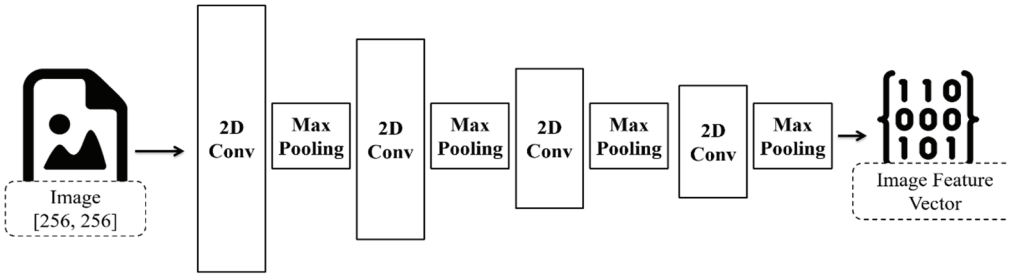


Fig. 5. Spam image feature extractor.

3.5 Final Spam Classifier

Each output feature vector of the four sub-models described above is merged into one vector. The merged feature vector includes both features of spam text extracted from images and features of spam images. Vectors are delivered to the final spam classifier consisting of fully connected to detect spam images.

4. Evaluation

In this section, the performance of the proposed model is evaluated. We present the following three research questions for systematic evaluation.

- RQ1. Is it possible to improve the performance of existing OCR-based spam filter by using the proposed spam text feature extractor?
- RQ2. How badly does obfuscated images affect OCR-based spam detection models?
- RQ3. Do each sub-model practically contribute to improving detection performance?

For detailed evaluation of the proposed model, evaluation models are defined as shown in Table 1 below. The baseline (first spam classifier) is a model that combines OCR and Apache SpamAssassin. STF is a model that merges the spam text feature extractor and the final spam classifier. SIF is a model that combines a spam image feature extractor, an obfuscated image classifier, and a final spam classifier. Finally, Base_STF and Base_SIF are defined by combining baseline, STF, and SIF, and proposed is the proposed model of this study as a model that merges all the defined models.

Table 1. Definition of models for evaluation

	Description
Text based model	
Baseline	OCR + Apache SpamAssasin [7] (First spam classifier)
STF	OCR + Spam text feature extractor + Final spam classifier
Base_STF	Baseline + STF
Image based model	
SIF	Spam image feature extractor + Obfuscated image classifier + Final spam classifier
Base_SIF	Baseline + SIF
Text & image model	
Proposed	Baseline + STF + SIF + Final spam classifier

For model training, we collected 3,217 spam images related to sales, drugs, gambling, and adult material from Image Spam dataset [22] and 1,211 ham images from Unsplash [23]. Table 2 below defines the data set to be used for the experiment. N_IMG is a dataset obtained by collecting and merging ham (non-spam) and spam images. For image obfuscation, O_IMG applies image distortion using affine transform, Gaussian blur, and distortion transform to random 40% samples among spam images of N_IMG.

Table 2. Definition of datasets

Datasets	Description
N_IMG	Normal spam image [22] + Ham image [23]
O_IMG	Obfuscated spam image + Ham image

Also, PyTorch was used as a tool for spam feature extraction and deep learning implementation, Tesseract-OCR [24] was used to extract text from images, and OpenCV2 was used to generate obfuscated images. Finally, accuracy, F1-score, precision, and recall were used as evaluation methods.

4.1 The Answer for RQ1

To evaluate how much the word embedding-based spam text feature extractor improves the performance of the existing spam image detection method compared to the existing model, the performance between text-based models is evaluated. Fig. 7 below shows the performance of baseline, STF, and Base_STF for the N_IMG dataset. N_IMG is a general data set that does not apply obfuscation, and the performance of baseline and STF can be compared.

Fig. 7(a) shows the performance for spam image detection. All models achieved scores of 91% or higher in all evaluations. However, the baseline had very unbalanced recall and precision, resulting in a relatively low F1-score. On the other hand, STF showed a stable score of 96% or higher in all evaluations. Base_STF is also affected by baseline and shows unbalanced performance. However, the performance of precision almost reached 99%. As a result of the experiment on spam image detection, it was verified that the word embedding-based spam text feature extractor can improve the performance of the existing OCR-based model. Fig. 7(b) shows the evaluation of ham image (non-spam image) detection. Even in ham image detection, baseline showed unbalanced performance and STF showed stable performance. In the end, STF had the highest F1-score of about 90%. Fig. 7(c) is the evaluation of accuracy and macroF1-score. MacroF1-score is the average of F1-score for spam image detection and F1-score for ham image

detection, and generally means overall performance. Because STF showed stable performance in all evaluations, it achieved the highest accuracy (about 94%) and macroF1-score (about 93%). As a result of the analysis of Fig. 7, it was verified that STF can improve the performance of the existing OCR-based spam image detection model.

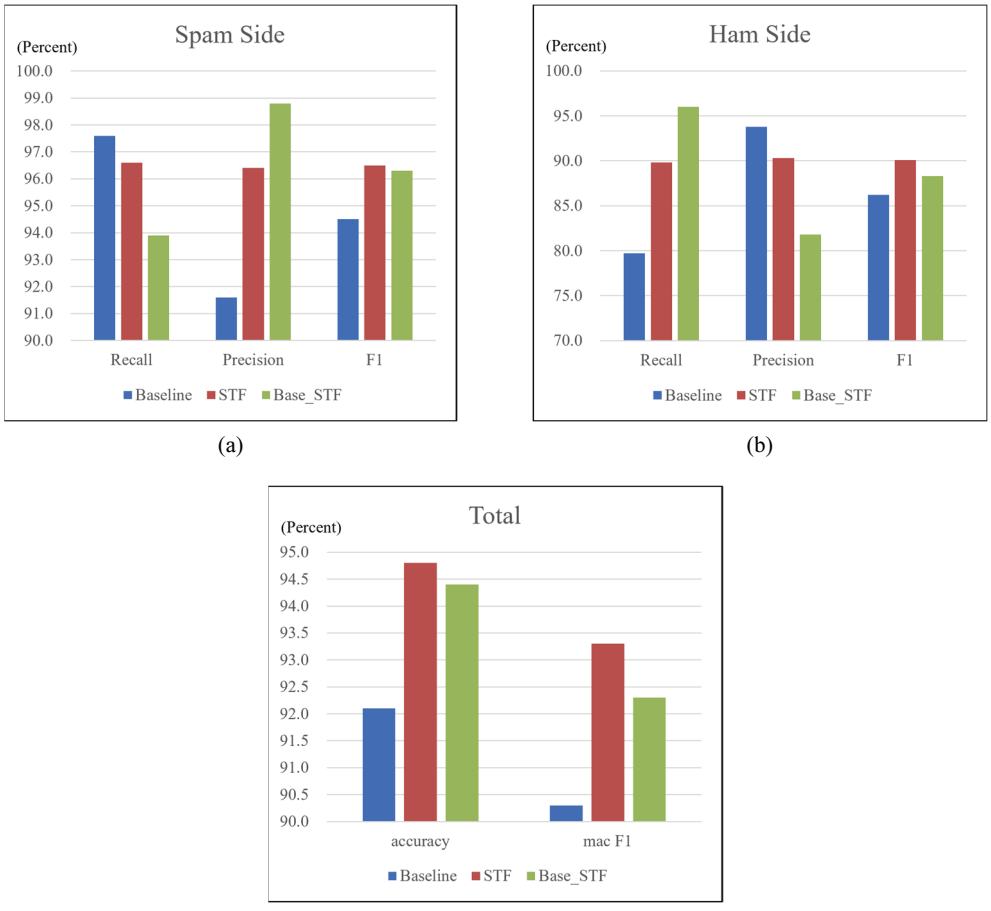


Fig. 7. Evaluation of text-based models for N_IMG dataset: (a) spam-side, (b) Ham-side, and (c) total evaluations.

4.2 The Answer for RQ2

Fig. 8 shows the performance evaluation of the text-based spam image detection model for O_IMG with obfuscated images. The obfuscated image deteriorates the performance of OCR because image distortion that interferes with text extraction is applied. Therefore, compared to the previous experiment, it is confirmed that the overall performance was decreased. Through the comparison between Fig. 7(c) and Fig. 8(c), it was shown that the classification performance of the model (O_IMG based classifier) that trained the dataset with the obfuscated images decreased compared to the model (N_IMG based classifier) that did not. In particular, the baseline showed a sharp performance decrease in precision. In the baseline that trained O_IMG, accuracy and macroF1-score decreased by 13%, and in STF and

base_STF, accuracy and macroF1-score decreased by 10%. Through these experimental results, it was verified that obfuscated images hinder text extraction from images and further adversely affect spam detection. In this evaluation, base_STF and STF showed almost similar performance, and although there was a sharp decrease in performance due to obfuscated images, STF showed better performance than baseline.

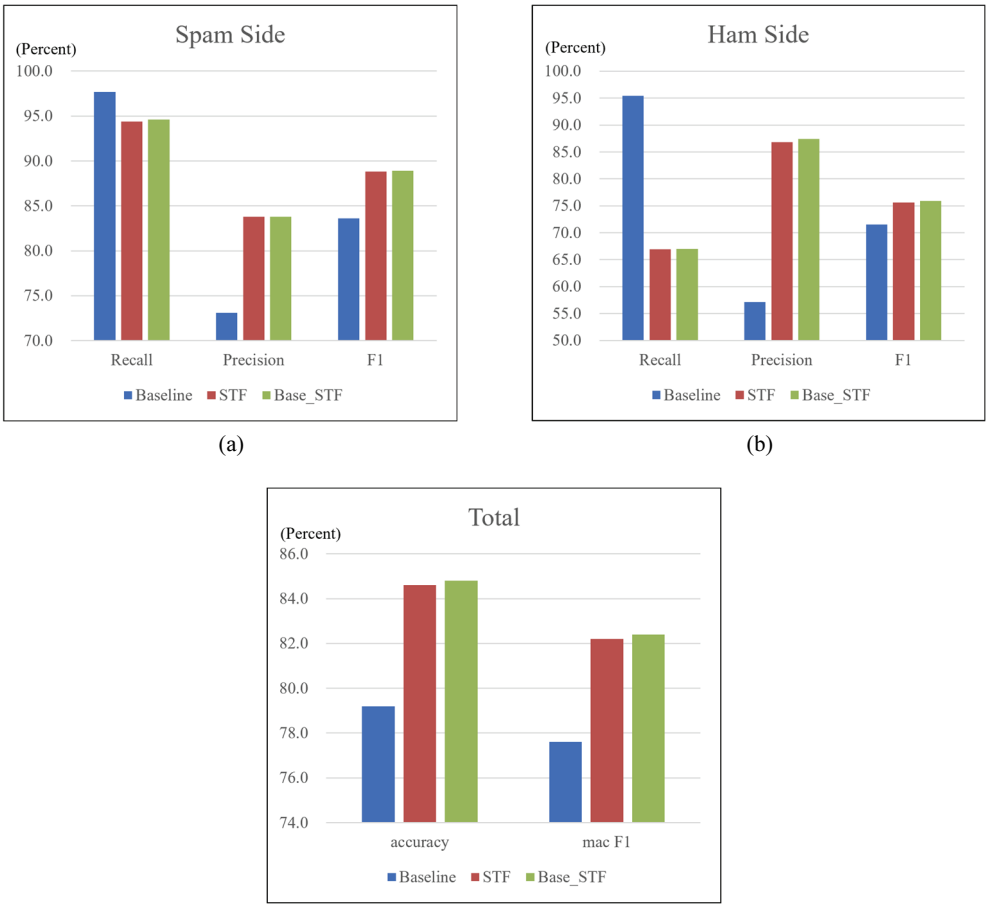
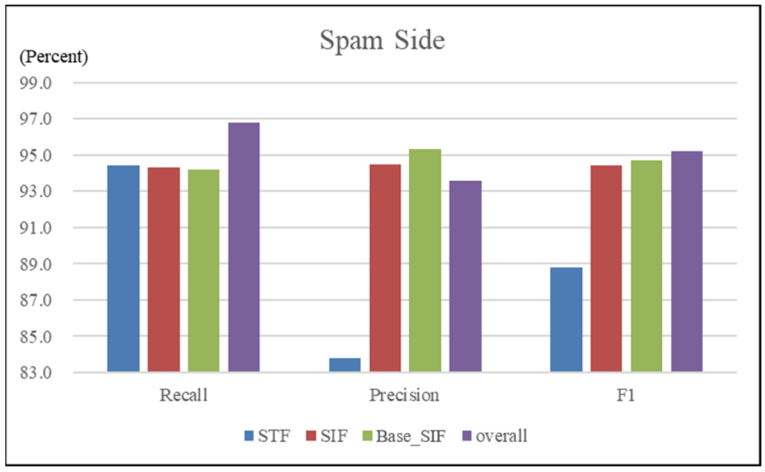


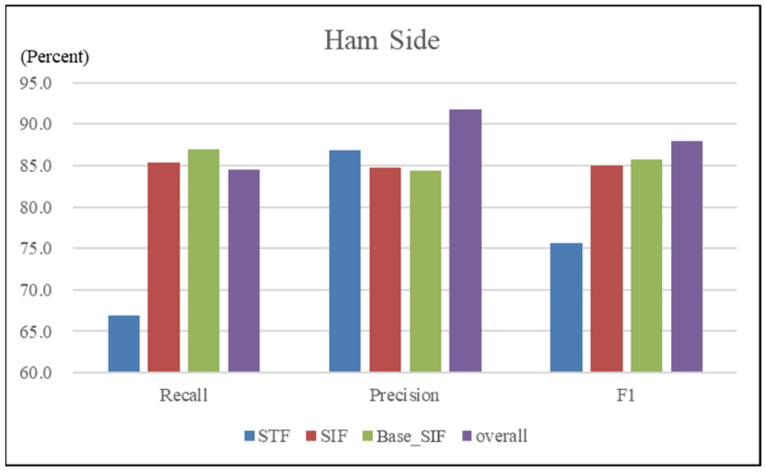
Fig. 8. Evaluation of text-based models for O_IMG dataset: (a) spam-side, (b) Ham-side, and (c) total evaluations.

4.3 The Answer for RQ3

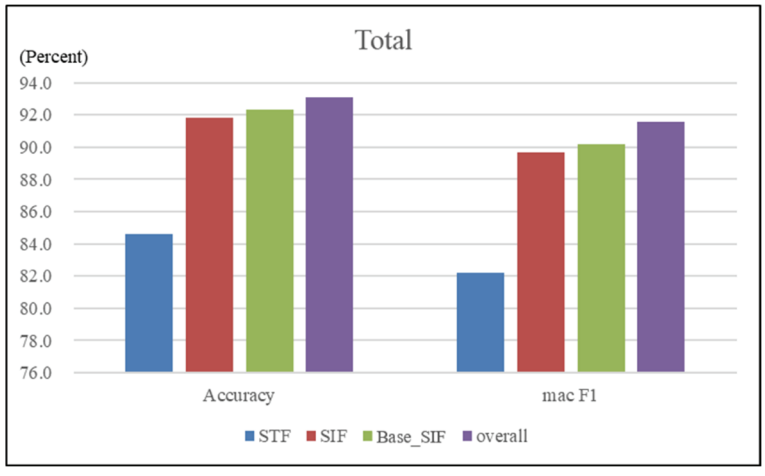
In this section, we evaluated the performance of STF and image-based models to evaluate the performance of image-based models. Fig. 9 shows the performance of each model for the O_IMG dataset. In the performance evaluation of obfuscated spam image detection shown in Fig. 9(a), image-based models show stable and high performance compared to STF. Image-based models showed very strong performance in detecting obfuscated spam images. In particular, the proposed model achieved very high performance over 95%.



(a)



(b)



(c)

Fig. 9. Evaluation of image-based and STF models for O_IMG dataset: (a) spam-side, (b) Ham-side, and (c) total evaluations.

In the ham image detection shown in Fig. 9(b), compared to the image-based models STF, the performance was very good, and the performance of the proposed model was the highest in the ham image detection. The macroF1-score of the comprehensive evaluation shown in Fig. 9(c) also reached about 91%. The superiority of the proposed model was verified by significantly improving the performance (about 14%) compared to the baseline macroF1-score (about 77%) for O_IMG. Fig. 9(c) showed an improvement in overall performance in accuracy, F1-score, and macroF1-score through comparison of the proposed models. Through this experiment, it is proved that the features extracted by each sub-model are actually contributing to performance improvement.

5. Conclusion

In this paper, we point out the limitations of the existing OCR-based spam image detection model and propose a deep learning-based spam image detection technique to improve performance and supplement vulnerabilities. To evaluate the identification performance of the proposed method, spam images and ham images were collected from Image Spam dataset and Unsplash, and the accuracy of the classifier, F1-score, precision, and recall were set as evaluation criteria. As a result of the evaluation, the proposed model improved the macroF1-score by 14% or more compared to the existing technique, and it was verified that it was very good in an environment where text extraction was difficult. Through the three questions presented in evaluation, it was shown that the four sub-models practically contribute to the spam image detection performance.

Although encouraging results have been obtained from our experiments, the proposed method still requires further work. The optimization of model hyperparameters and learning speed is insufficient. Moreover, it is difficult to respond to all attack patterns because new ways of spammers attacking the proposed model continue to evolve. In order to cope with this issue, it is necessary to add a routine of periodically updating the model by learning on a new attack pattern. Furthermore, because this study has focused only on spam image detection, it is necessary to improve the performance for ham image detection.

Acknowledgement

This research was results of a study on the "HPC Support" Project, supported by the Ministry of Science and ICT' and NIPA. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. NRF-2020R1I1A3073313).

References

[1] M. Malekshahi Rad, A. M. Rahmani, A. Sahafi, and N. Nasih Qader, "Social Internet of Things: vision, challenges, and trends," *Human-centric Computing and Information Sciences*, vol. 10, article no. 52, 2020. https://doi.org/10.1186/s13673-020-00254-6

[2] J. Salminen, M. Hopf, S. A. Chowdhury, S. G. Jung, H. Almerekhi, and B. J. Jansen, "Developing an online hate classifier for multiple social media platforms," *Human-centric Computing and Information Sciences*, vol. 10, article no. 1, 2020. https://doi.org/10.1186/s13673-019-0205-6

[3] Z. Zhang, J. Jing, X. Wang, K. K. R. Choo, and B. B. Gupta, "A crowdsourcing method for online social networks security assessment based on human-centric computing," *Human-centric Computing and Information Sciences*, vol. 10, article no. 23, 2020.

[4] J. H. Park, S. Rathore, S. K. Singh, M. M. Salim, A. E. Azzaoui, T. W. Kim, Y. Pan, and J. H. Park, "A comprehensive survey on core technologies and services for 5G security: Taxonomies, issues, and solutions," *Human-centric Computing and Information Sciences*, vol. 11, article no. 3, 2021. https://doi.org/10.22967/HCIS.2021.11.003

[5] S. Rathore, J. H. Park, and H. Chang, "Deep learning and blockchain-empowered security framework for intelligent 5G-enabled IoT," *IEEE Access*, vol. 9, pp. 90075-90083, 2021.

[6] S. Rathore and J. H. Park, "A blockchain-based deep learning approach for cyber security in next generation industrial cyber-physical systems," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 8, pp. 5522-5532, 2021.

[7] Apache SpamAssassin [Online]. Available: https://spamassassin.apache.org.

[8] B. Biggio, G. Fumera, I. Pillai, and F. Roli, "Image spam filtering using visual information," in *Proceedings of the 14th International Conference on Image Analysis and Processing (ICIAP)*, Modena, Italy, 2007, pp. 105-110.

[9] Z. Wang, W. K. Josephson, Q. Lv, M. Charikar, and K. Li, "Filtering image spam with near-duplicate detection," in *Proceedings of the 4th Conference on Email and Anti-Spam (CEAS)*, Mountain View, CA, 2007.

[10] A. Barbar and A. Ismail, "Image spam detection using FENOMAA technique," in *Artificial Intelligence and Applied Mathematics in Engineering Problems*. Cham, Switzerland: Springer, 2020, pp. 347-364.

[11] C. Fatichah, W. F. Lazuardi, D. A. Navastara, N. Suciati, and A. Munif, "Image spam detection on instagram using convolutional neural network," in *Intelligent and Interactive Computing*. Singapore: Springer, 2019, pp. 295-303.

[12] N. Imam and V. Vassilakis, "Detecting spam images with embedded Arabic text in twitter," in *Proceedings of 2019 International Conference on Document Analysis and Recognition Workshops (ICDARW)*, Sydney, Australia, 2019, pp. 1-6.

[13] A. P. Singh, "Image spam classification using deep learning," Master's thesis, San Jose State University, San Jose, CA, 2018.

[14] S. Rao and R. Gopalapillai, "Effective spam image classification using CNN and transfer learning," in *Computational Vision and Bio-Inspired Computing*. Cham, Switzerland: Springer, 2020, pp. 1378-1385.

[15] A. Fan and Z. Yang, "Image spam filtering using convolutional neural networks," *Personal and Ubiquitous Computing*, vol. 22, pp. 1029-1037, 2018.

[16] T. Sharmin, F. Di Troia, K. Potika, and M. Stamp, "Convolutional neural networks for image spam detection," *Information Security Journal: A Global Perspective*, vol. 29, no. 3, pp. 103-117, 2020.

[17] TorchText [Online]. Available: https://pytorch.org/text/stable/index.html.

[18] OpenCV, "Affine Transformations tutorials," 2019 [Online]. Available: https://docs.opencv.org/2.4/doc/tutorials/imgproc/imgtrans/warp_affine/warp_affine.html.

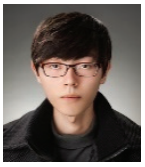
[19] OpenCV, "Smoothing images tutorial: Gaussian Blurring," 2019 [Online]. Available: https://docs.opencv.org/2.4/doc/tutorials/imgproc/gausian_median_blur_bilateral_filter/gausian_median_blur_bilateral_filter.html.

[20] CAPTCHA, "CAPTCHA: Telling Humans and Computers Apart Automatically," 2010 [Online]. Available: http://www.captcha.net/.

[21] G. Bradski, “The OpenCV library,” *Dr. Dobb's Journal*, vol. 25, no. 11, pp. 120-125, 2000.
[22] M. Dredze, R. Gevaryahu, and A. Elias-Bachrach, “Image Spam Dataset,” 2007 [Online]. Available: https://www.cs.jhu.edu/~mdredze/datasets/image_spam/.
[23] Unsplash [Online]. Available: https://unsplash.com/.
[24] Tesseract Open-Source OCR Engine, “Tesseract-OCR,” 2023 [Online]. Available: https://github.com/tesseract-ocr/tesseract.

Seong-Guk Nam **https://orcid.org/0000-0002-6197-3645**

He received the B.S. degree in computer engineering from Yeungnam University, Korea, in 2021. He is currently a M.S. student in the Department of Computer Engineering, Yeungnam University, Korea. His research interests include AI, multimedia, and software engineering.



Dong-Gun Lee **https://orcid.org/0000-0001-6792-4572**

He received the M.S. degree in computer engineering from Yeungnam University, Korea, in 2021. He is currently a doctoral student in the Department of Computer Engineering, Yeungnam University, Korea. His research interests include software engineering, open-source software, software defect prediction, and edge computing.

Yeong-Seok Seo **https://orcid.org/0000-0002-5319-7674**

He received the B.S. degree in computer science from Soongsil University, Korea, in 2006, and the M.S. and Ph.D. degrees in computer science from Korea Advanced Institute of Science and Technology (KAIST), Korea, in 2008 and 2012, respectively. From September 2012 to December 2013, he was a postdoctoral researcher in KAIST institute for Information and Electronics. From January 2014 to August 2016, he was a senior researcher in Korea Testing Laboratory (KTL), Korea. He is currently an assistant professor in the Department of Computer Engineering, Yeungnam University, Korea. His research interests include software engineering, artificial intelligence, Internet of Things (IoT), and data mining.