Al-Nahrain Journal of Science
ANJS, Vol.23 (4), December, 2020, pp. 44–48

Color Model Based Convolutional Neural Network for Image Spam Classification

Ahmad Mahdi Salih* and Ban Nadeem Dhannoon

Department of Computer Science, College of Science, Al-Nahrain University, Baghdad, Iraq

Articles Information

Received:
06.09.2020
Accepted:
31.10.2020
Published:
01.12.2020

Keywords:
Convolutional neural network
E-mail
Image spamming
Color models

Abstract

For most of people, e-mail is the preferable medium for official communication. E-mail service providers face an endless challenge called spamming. Spamming is the exploitation of e-mail systems to send a bulk of unsolicited messages to a large number of recipients. Noisy image spamming is one of the new techniques to evade text analysis based and Optical Character Recognition (OCR) based spams filtering. In the present paper, Convolutional Neural Network (CNN) based on different color models was considered to address image spam problem. The proposed method was evaluated over a public image spam dataset. The results showed that the performance of the proposed CNN was affected by the color model used. The results also showed that XYZ model yields the best accuracy rate among all considered color models.

DOI: 10.22401/ANJS.23.4.08

*Corresponding author: ahmadtoday2@gmail.com

1. Introduction

Today, billions of people and devices are connected together through internet. One of extremely useful internet applications is E-mail. E-mail is a popular messaging system. Modern e-mail systems have many powerful advantages such as messaging with attached files and embedded images [1]. The popularity of e-mail makes it an attractive space for spamming. Spam is unwanted bulk e-mail [2]. Spams have several negative consequences such as network and human resources consumption [3]. Spams can become more dangerous by including scams, phishing links and offensive content, which may significantly threaten users' security and privacy [4, 5]. According to recent statistics, spam rate was approximately 56% of all e-mail traffic [6]. Several spam filters have been suggested to address spam problem. The 1st generation of spam is in text form, text analysis based on Natural Language Processing (NLP) techniques and machine learning algorithms provide powerful spam filters [7]. Spammers have to become more professional to avoid spam filters. One of spammers' new tricks was image spam. In image spamming, the fraudulent text is embedded in an image, resulting in the 2nd generation of spam [8]. The first reaction toward spam image was Optical Character Recognition (OCR). OCR is used to convert embedded text information into plaintext, and then text-based filters are used to recognize spam from legitimate e-mails [9]. To make OCR worthless, spammers use noisy images with a goal of making the image readable by the humans but un-readable by OCR [10]. Spam images have a number of unique visual features that distinguish them from ham images [11]. Image spamming literature refers to several filters have been recommended based on images analysis. The present paper aims to analyze the effect of color models on the performance of CNN in the context of image spamming. The paper also aims to yield a robust convolutional feature extraction and hence improving the performance of the existing image spam filters. Figure 1 presents examples of real world spam images.

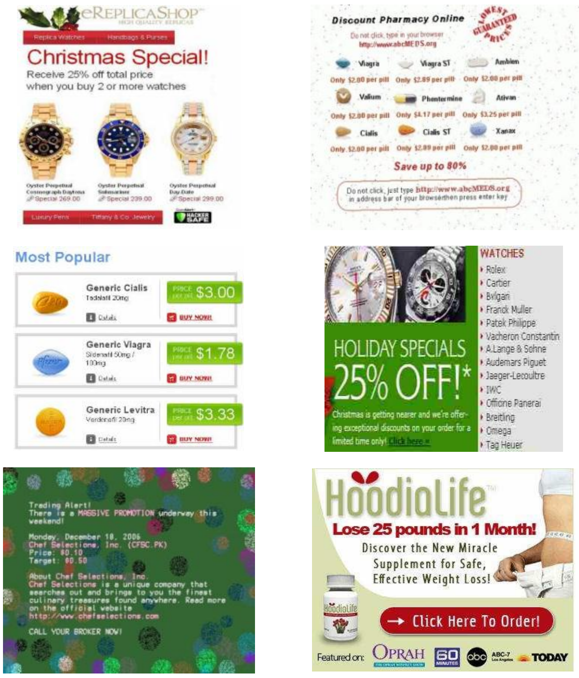


Figure 1. Image spam.

2. Literature Review

Several studies have been published to address image spam. The present section discusses the recent studies in the field of image spamming.

Annadatha *et al.* [12] proposed an image spam classification method based on image processing. The proposed method utilized a variety of image descriptors as well as image file properties. Using Support Vector Machines (SVM), best result was an accuracy score of 97.25% on ISH dataset (the dataset is discussed in section 5).

Chavda *et al.* [13] suggested image spam classification method based on a large set of image features. In this study, SVM have been applied to a set of 41 image features. Recursive Feature Elimination (RFE) based on SVM weights was used as feature selection technique. The researchers were able to obtain an accuracy score of 97% on ISH dataset.

Kumar *et al.* [14] recommended a convolution neural network for image spam classification. The proposed network has three convolutional layers, three max-pooling windows and a dropout layer to avoid over-fitting. The suggested method was trained and evaluated over ISH dataset and yielded a classification accuracy of 91.7%.

Srinivasan *et al.* [15] explored image spam detection based on deep features. The deep features are extracted using convolutional layers. After that, different machine learning algorithms were experimented. The researchers considered several datasets and their best results on ISH dataset were obtained by using linear SVM with a classification accuracy of 98.1%.

3. Methodology

This section presents an overview of the color models and deep neural network algorithm considered in the present paper.

3.1 Color models

The purpose of using different color models is to obtain image features that are as invariant as possible to shading, shadows, contrast and illumination. The six considered color models are listed in the following subsections.

3.1.1 RGB model

Digital Images are usually in the RGB color space, one channel for each primary color (Red, Green, and Blue) [16]. Although RGB color model is useful for color displaying, it is not suitable for image analysis. This is because there is a high correlation among RGB channels.

3.1.2 Chrominance-luminance color models (HSV, YCbCr, XYZ, LAB and YUV)

Chrominance-luminance color space takes the advantage of the separation between the chrominance components and the luminance/brightness component (illumination invariant) [17]. This makes such color models desirable for image analysis particularly for the aim of image spam classification where the fluctuations in the brightness component in spam images (mostly computer generated) is low as compared to that of natural images.

3.2 Convolutional neural network

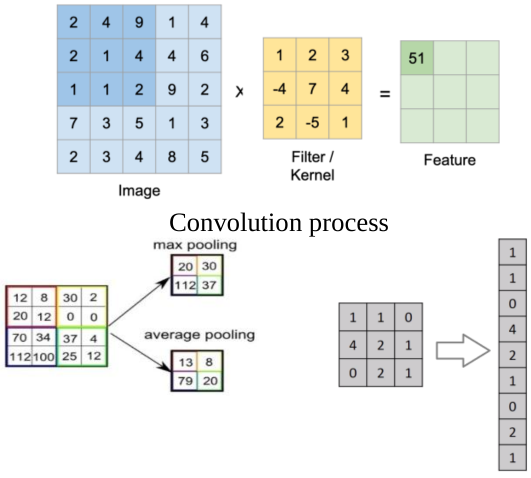
Neural networks are the biologically inspired learning algorithms that make the computer machine to learn from data. Deep learning is a class of neural networks that uses multiple hidden layers to gradually extract high level features from the raw input [18]. Convolutional Neural Network is the deep learning algorithm for image processing and computer vision applications. Instead of using hand-define filters/operators to obtain various image features (i.e., shape, texture, color, etc.), deep learning and especially Convolutional Neural Network takes a different approach to extract features from an image, In CNN, image features are automatically and progressively learned from the training process, thus the primary advantage CNN is that it allows skipping the feature extraction step and instead focusing on training the network to learn convolutional filters [19]. Basically CNN consists of four components as shown in Figure 2:

-**Convolutional layer:** its purpose is image feature extraction via kernels parameters optimization.

-**Pooling layer:** its purpose is dimensionality reduction via down sampling process.

-**Flatten layer**: its purpose is to convert the resulting matrices into a 1D vectors.

-**Fully connected layer:** a feed forward neural network and its purpose is classification.



Convolution process

pooling process Flatten process

Figure 2. CNN components.

4. The Proposed Method

The proposed method consists of two steps image preprocessing and convolutional neural network. Python was used to create the proposed method, Scikit-learn package for color model transformations and Tensor-Flow package to implement the convolutional neural network. The general diagram of the proposed method is illustrated in Figure 3.

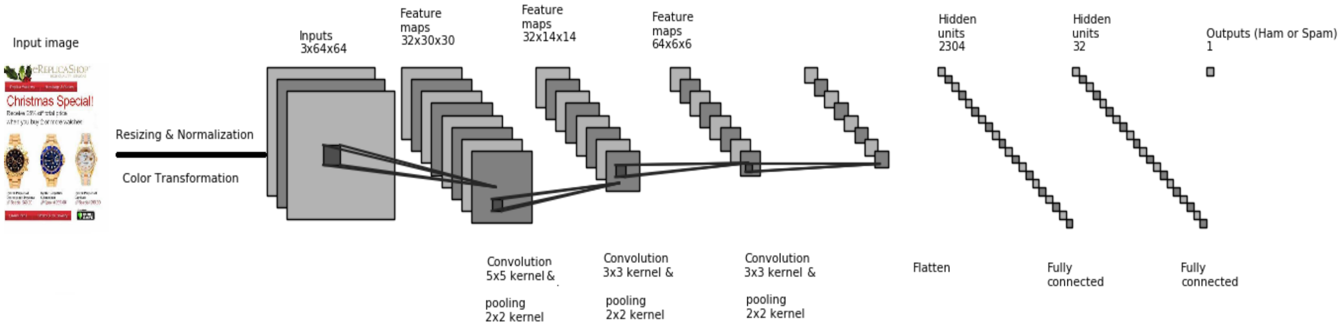


Figure 3. The proposed detection model.

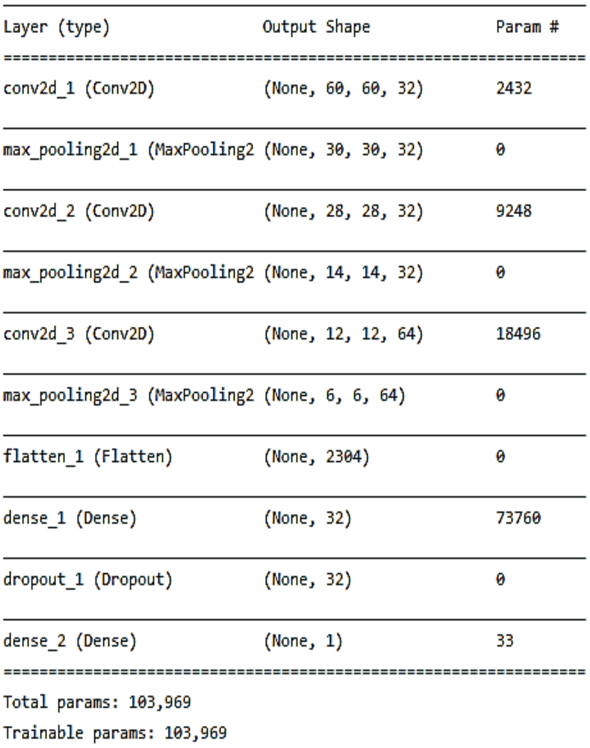
4.1 Preprocessing

The preprocessing step aims to prepare the input image into convolutional layers. All images are normalized into 0-1 range and resized into 64×64 dimensions. The preprocessing step also includes transforming images into different color models.

4.2 Convolutional neural network

The proposed CNN model for image spam detection has input layer, three convolutional layers, three max-pooling layers, a flatten layer, a drop out unit and two dense fully connected layers. The following configuration was used:

The first convolution layer uses a 5×5 kernel size and 32 nodes (kernel). The second convolution layer uses a 3×3 kernel size and 32 nodes and the last convolutional layer having 64 nodes with 3×3 kernel size. The convolutional stride is one pixel without space padding. Rectified Linear Unit (ReLU) is used as activation function for all convolutional layers, and each convolutional layer is connected to a 2×2 Max-Pooling layer which down samples the input data into half of its original dimension. Adaptive Moment Estimation (Adam) is employed as an optimization algorithm with initial learning-rate = 0.001. Cross entropy loss is utilized as a cost function. A dropout unit with 0.25 rate is used for regularization (i.e. avoid over fitting). Early Stopping is used to stop the training after 10 epochs with no improvement. In the final two dense fully-connected layers, ReLU and sigmoid were used as activation functions respectively. A general description of the proposed convolutional neural network architecture, output shape and number of parameters for each layer is provided in Figure 4.



```
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_1 (Conv2D)            (None, 60, 60, 32)        2432
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 30, 30, 32)        0
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 28, 28, 32)        9248
_________________________________________________________________
max_pooling2d_2 (MaxPooling2 (None, 14, 14, 32)        0
_________________________________________________________________
conv2d_3 (Conv2D)            (None, 12, 12, 64)        18496
_________________________________________________________________
max_pooling2d_3 (MaxPooling2 (None, 6, 6, 64)          0
_________________________________________________________________
flatten_1 (Flatten)          (None, 2304)              0
_________________________________________________________________
dense_1 (Dense)              (None, 32)                73760
_________________________________________________________________
dropout_1 (Dropout)          (None, 32)                0
_________________________________________________________________
dense_2 (Dense)              (None, 1)                 33
=================================================================
Total params: 103,969
Trainable params: 103,969
```

Figure 4. Proposed CNN summary.

5. ISH Dataset

ISH dataset is publicly available at North Western University web page [20]. It was proposed by the researchers of a study titled “Image Spam Hunter” [21]. The images were collected from the real world e-mails. The dataset consists of 929 spam and 810 non-spam images.

6. Classification Results

The dataset splitting was 75% for training and the remaining for testing, and the proposed CNN was trained for 50 epochs with a batch size of 50. Table 1 shows the accuracy scores of the different color models used for image spam classification. Since the RGB is the mostly used color model, it was considered as the benchmark for the comparisons with other color models performance. From Table 1, RGB obtained better accuracy than LAB and same as HSV. On the other hand, RGB experiment did not obtain accuracy as good as YCbCr, XYZ, and YUV. The XYZ color model performed the best among all color models with an accuracy score of 98.4%. In comparison to related works, XYZ model based CNN was able to improve the performance as the best accuracy previously obtained was 98% on ISH data set. Figures 5 and 6 show accuracy and loss rate of the proposed XYZ-CNN model over 50 epochs.

Table 1. Accuracy rates for different color model.

Color model	Accuracy rate
RGB	97.4
HSV	97.4
YCbCr	98
XYZ	**98.4**
LAB	85.3
YUV	97.8

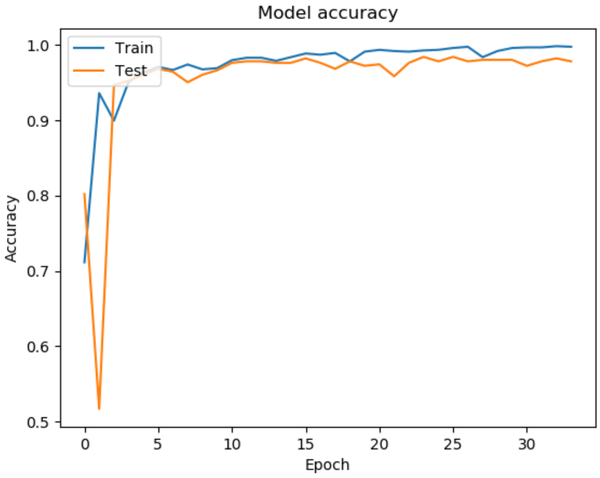


Figure 5. The classification rate of XYZ-CNN model.

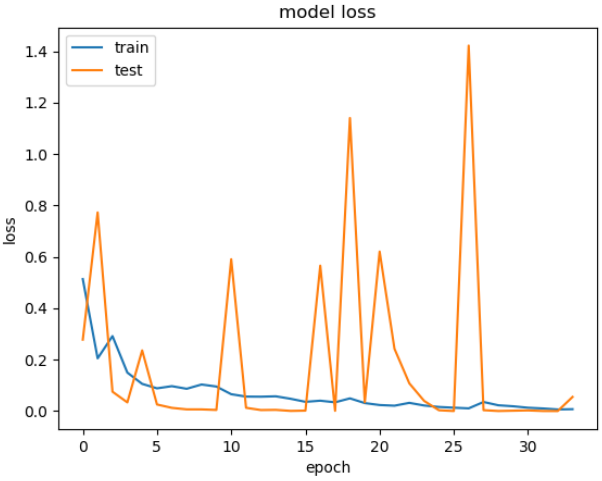


Figure 6. The loss rate of XYZ-CNN model.

7. Conclusion and Future Work

Differentiate spam images from non-spam ones is a challenging task.In this paper, various color models were considered to perform convolutional neural network for the purpose of image spam classification. From the results, different color models obtained different performances. The color model that most fitted the image spam classification was the XYZ as it yields the highest scores among all experimented color models. In comparison to previous works, our obtained accuracy scores improved over the previous works on the same dataset. The present work can help e-mail service providers to select the suitable color model when developing deep-learning based image spam detection. As a future work, image data augmentation based on color models can be used to increase the amount of training images by creating various color modeled versions of the original RGB images.

References

[1] Kurose, J. and K. Ross; "Computer Networking: A Top Down Approach"; Addison-Wesley; 6th edition; 2013.
[2] Cormack, G. V.; "Email spam filtering:A systematic review"; Now Publishers ; pp. 335-455; 2008.
[3] M. Prince; "Clustering-based spam image filtering considering fuzziness of the spam image"; Int. J. Adv. Comput. Sci. Appl.;p: 269-270; 2016.
[4] Gangavarapu,T., Jaidhar,C. "Applicability of machine learning in spam and phishing email filtering: review and approaches", Arti. Intell. Rev.; p. 1-63; 2020.
[5] Kigerl, A.; "Spam-Based Scams. The Palgrave Handbook of International Cybercrime and Cyberdeviance"; pp. 877-897; 2020.
[6] Dada, E. G.; "Machine learning for email spam filtering: review, approaches and open research problems"; pp. 1-23; 2019.
[7] Kumar, J., S. Taterh, and D. Kamnthania; "Study and Comparative Analysis of Various Image Spamming Techniques"; in Soft Computing: Theories and Applications; Springer; pp.351-365; 2018.
[8] Dhavale, S. V.; "Advanced image-based spam detection and filtering techniques"; IGI Global; 2017.
[9] Attar, A., R. M. Rad, R. E. Atani; "A survey of image spamming and filtering techniques"; Arti. Intell. Rev.; p.71-105; 2013.
[10] Dhahi, E. H., S. A. Ali, and M. A. Naser; "Text Region Extraction for Noisy Spam Image"; in Cognitive Informatics and Soft Computing; Springer; p. 225-233; 2020.
[11] Singh, A. P.; "Image Spam Classification using Deep Learning"; Master report; San Jose State University; 2018.
[12] Annadatha, A. and M. Stamp; "Image spam analysis and detection"; Journal of Computer Virology and Hacking Techniques"; p. 39-52; 2016.
[13] Chavda, A., et al.; "Support Vector Machines for Image Spam Analysis". in ICETE;2018.

[14] D.Kumar, Soman K. P.; "DeepImageSpam: Deep Learning based Image Spam Detection"; 2018.

[15] Srinivasan, S., et al.; "Deep convolutional neural network based image spam classification"; in 6th Conference on Data Science and Machine Learning Applications (CDMA); IEEE; 2020.

[16] Gonzalez, R. C.; "Digital Image Processing"; Pearson; fourth ed. 2018.

[17] Zhang, D.; "Fundamentals of Image Data Mining"; Springer; 2019.

[18] Wani, M. A., et al.; "Advances in Deep Learning"; Springer; 2020.

[19] Rosebrock, A.; "Deep Learning for Computer Vision with Python: ImageNet Bundle"; 2017.

[20] Gao, Y. IMAGE SPAM HUNTER; 2008.

[21] Gao, Y.; "Image spam hunter"; in IEEE International Conference on Acoustics; Speech and Signal Processing, IEEE: Las Vegas, NV, USA; 2008.