Actas de las XXXIX Jornadas de Automática, Badajoz, 5-7 de Septiembre de 2018
https://doi.org/10.17979/spudc.9788497497565.0991

COLOR SIFT DESCRIPTORS TO CATEGORIZE ILLEGAL ACTIVITIES IN IMAGES OF ONION DOMAINS

David Matilla[1], Víctor González-Castro[1,3], Laura Fernández-Robles[2,3], Eduardo Fidalgo[1,3], Mhd Wesam Al-Nabki[1,3]
[1] Departamento de Ingeniería Eléctrica y de Sistemas y Automática, Universidad de León
[2] Departamento de Ingeniería Mecánica, Informática y Aeroespacial, Universidad de León
[3] Researcher at INCIBE (Instituto Nacional de Ciberseguridad), León
{david.matilla, victor.gonzalez, l.fernandez, eduardo.fidalgo, mnab}@unileon.es

Abstract

Dark Web, i.e. the portion of the Web whose content is not indexed either accessible by standard web browsers, comprises several darknets. The Onion Router (Tor) is the most famous one, thanks to the anonymity provided to its users, and it results in the creation of domains, or hidden services, which hosts illegal activities. In this work, we explored the possibility of identifying illegal domains on Tor darknet based on its visual content. After crawling and filtering the images of 500 hidden services, we sorted them into five different illegal categories, and we trained a classifier using the Bag of Visual Words (BoVW) model. In this model, SIFT (Scale Invariant Feature Transform) or dense SIFT were used as the descriptors of the images patches to compute the visual words of the BoVW model. However, SIFT only works with gray-scale images; thus the information given by color in an image is not retrieved. To overcome this drawback, in this work we implemented and assessed the performance of three different variants of SIFT descriptors that can be used in color images, namely HSV-SIFT, RGB-SIFT and Opponent-SIFT, in the BoVW model for image classification. The obtained results showed the usefulness of using color-SIFT descriptors instead of SIFT, whereas in our experiments the latter achieved an accuracy of 57.52%, the HSV-SIFT descriptor achieved an accuracy up to 59.44%.

Keywords: SIFT, Image Classification, Tor, Cybersecurity, Machine Learning.

1 INTRODUCTION

Nowadays, the use of images is extremely common, e.g. more than 80M images are uploaded to Instagram every day. Whereas it is very convenient to have access to so many data easily, it has some drawbacks, since some images may contain unwanted, illegal or even hidden information, e.g. using steganography [1].

Instagram, Facebook or Flickr belong to the *Surface Web*, which is the part of the Web which is accessible by means of standard search engines. The rest of the Web, whose content cannot be indexed by these is called the *Deep Web*. The part of the *Deep Web* that can only be accessed through special browsers or a proxy server is called the *Dark Web*. The most famous example of networks in the *Darknet* is the Tor (The Onion Router) network[1], which is a project that makes possible to anonymize connections, thus making almost impossible to track where each communication comes from. This encourage its usage by cybercriminals, i.e. many types of illegal contents can be found, from drugs or guns stores to human trafficking or child sexual abuse [2]. Therefore, Law Enforcement Agencies (LEAs) monitor Tor, as well as other networks in the *Dark Web*, to find illegal contents and track down where they come from. However, carrying this out manually would be very time-consuming and a huge amount of resources would be needed. Therefore, it is important to provide LEAs with tools to detect illegal content automatically.

Image classification has shown to be a useful tool to categorize images automatically. Many descriptors can be found in the literature, such as texture descriptors [3], shape descriptors [4], adaptive texture descriptors [5] or color descriptors [6]. The Bag of Visual Words (BoVW) model has shown an excellent performance to carry out image classification [7, 8]. This model is originally inspired by the Bag of Words model used for text analysis. It extracts local descriptors as low-level image features, which are quantized as *visual words*) with the help of a *visual dictionary* and then analyzes the distributions of these visual words to describe the image content. Scale Invariant Feature Transform (SIFT) [9] has been widely along with the BoVW model to characterize the visual words [10, 11]. Regardless BoVW has been outperformed by other methods, such as Spatial Pyramid Matching [12] or Convolutional Neural Networks (CNN) [13], it is still being used in recent researches. Lazebnik et al. [11] used bag-of-features approach with SIFT descriptors over Harris-affine keypoints [14]. Fidalgo et al. proposed Compass

[1] https://www.torproject.org/

Radius Estimation for Improved Image Classification (CREIC) [15], a method based on Edge-SIFT descriptors to estimate an optimized radius value for the compass operator to compute Edge-SIFT descriptors. Later, they used CREIC method to characterize images related to illegal activities in the Tor network, presenting a new dataset named TOIC (TOr Image Categories) [16], which contains images from Tor domains labelled as *drugs*, *weapons*, *money*, *personal ID* or *credit cards*, depending on the category of its corresponding domain in the DUTA (Darknet Usage Text Addresses) dataset [2, 17]. Later, they proposed AutoBlur and two variants of Semantic Attention Region Filtering (SARF) methods [18], which create richer dictionaries to boost the performance of the BoVW model, which also used dense SIFT to extract low-level features from the image.

Whereas SIFT is very common as low-level descriptor in BoVW, it works only on gray-level images, thus missing the information that color may provide for characterizing them. Recently, several neural networks techniques have been proposed to overcome the shortage of SIFT [19, 20, 21], but they are out of the scope of this paper.

In this work we are assessing the performance of three variants of SIFT, namely, HSV-SIFT, RGB-SIFT, and Opponent-SIFT, for color images [6] for classifying TOIC dataset images. At the same time, we will compare their performance with the one obtained by traditional SIFT descriptors. The application of this work would help the Law Enforcement Agencies in classifying the graphical content in Tor hidden services.

The rest of the paper is organized as follows: Section 2 describes the methods that we have used for low-level image description (i.e. the color SIFT methods). The experiments carried out to assess the color-SIFT descriptors are explained in Section 3, as well as their results and discussion. Finally, the conclusions and future work lines are shown in Section 4.

2 METHODOLOGY

In this paper, the classification of the images has been carried out using the Bag of Visual Words (BoVW) [7, 8] model. This model describes the contents of an image by means of the frequency of occurrence of some visual elements contained on it. To accomplish that, small patches around keypoints of the images are described by means of a local descriptor, e.g. SIFT, to get the so-called *visual words*. Using a set of images representative of the classes that we want to classify, we create a *dictionary* by clustering the visual words extracted from the images. Thereafter, in order to represent each image, its visual words are quantized with the help of the dictionary in order to get the frequency of their occurrences along the image. Once the images are globally described, we use a Support Vector Machine (SVM) classifier [22] to characterize them.

In order to describe the low-level image features (i.e., the keypoint descriptors) we use the classical SIFT [9] descriptors, calculated from gray-level images, and also some of the color descriptors based on SIFT proposed by van de Sande et al. [6]. The goal is to assess the usefulness of the information captured by the color of the image in the low-level descriptors.

SIFT is a well-known state-of-the-art descriptor, so we will not include details about it. We address the reader interested in more details to [9]. In this work we have not used any keypoint detector to select the points to extract the SIFT descriptors from. Instead, we have used a dense sampling of keypoints, i.e. dense-SIFT, using the VLFeat library [23]. We have set both the parameters of the sampling density (i.e., *step*) and scale of the extracted descriptors (i.e., *size*) in dense-SIFT to 7, as it was done in [15].

2.1 COLOR SIFT DESCRIPTORS

All the color SIFT descriptors referred to in this Section (i.e., HSV-SIFT -Figure 1-, RGB-SIFT and Opponent-SIFT) are based on the classical SIFT descriptor. We will refer only to the description part, thus not addressing the SIFT keypoint detection.

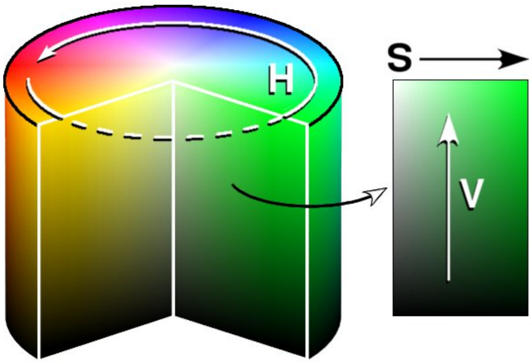
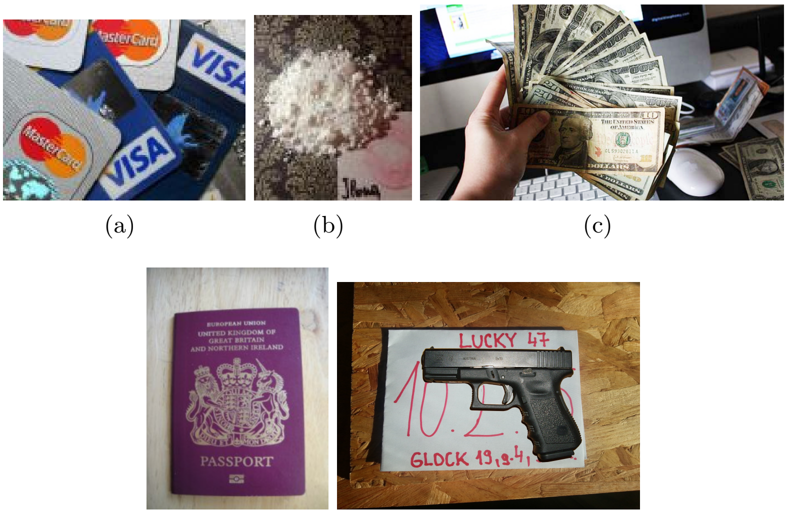


Figure 1: Representation of the HSV color space as a cylinder. The illustration has been created by Eric Pierce (source: `https://commons.wikimedia.org/wiki/File:HSV_cylinder.jpg`)



(a) (b) (c)

(d) (e)

Figure 2: Examples of a TOIC image per class: (a) Credit card (b) Drugs (c) Money (d) Personal ID and (e) Weapons

2.1.1 HSV-SIFT

In order to get the HSV-SIFT descriptor of a point [24], we calculate the SIFT descriptors of each channel of the HSV color space, i.e. Hue (H), Saturation (S) and Value (V) (see Figure 1). After the descriptors are calculated, they are concatenated, thus creating a vector with 384 features, i.e., 128 per channel [25].

2.1.2 RGB-SIFT

In the case of the RGB-SIFT descriptor [26], SIFT descriptors are calculated independently for each of the RGB channels. Thereafter, the 128-feature vectors of each channel are concatenated, so that the RGB-SIFT descriptor has 384 features.

2.1.3 Opponent-SIFT

In order to compute the Opponent-SIFT descriptors [27], it is necessary to transform the original RGB image into the opponent color space. Let R_i, B_i and G_i be the red, green and blue color values of the i-th pixel, respectively. The values (O_1, O_2, O_3) of that pixel in the opponent color space are given by the equation (1).

$$\begin{pmatrix} O_1 \\ O_2 \\ O_3 \end{pmatrix} = \begin{pmatrix} \frac{R_i - G_i}{\sqrt{2}} \\ \frac{R_i + G_i - 2B_i}{\sqrt{6}} \\ \frac{R_i + G_i + B_i}{\sqrt{3}} \end{pmatrix} \tag{1}$$

O_1 and O_2 capture the color information while O_3 represents the intensity information [28]. Hereafter, for each keypoint with coordinates (x, y), Opponent-SIFT describes that point in all the channels in the opponent color space. Therefore, the Opponent-SIFT descriptor of each keypoint has 384 features (i.e. 128 per channel).

3 EXPERIMENTS AND RESULTS

In this section, we present the dataset of images used for evaluating the method presented in Section 2, the details of the experimental setup, the results obtained and a discussion.

3.1 DATASET

In this work we extended the TOr Image Categories (TOIC) dataset proposed by [16]. Figure 2 shows an example of an image from TOIC per class.

We extended TOIC dataset[2] [16] using a Tor image crawling and a filtering strategy. TOIC dataset was created with the intention of collecting a set of images from Tor domains. The crawler designed for capturing the text content of the Darknet Usage Text Addresses (DUTA) dataset[3] [2] was modified in order to download the image resources that the Tor domains contained. The images were assigned the same labels as the labels of the text contents of the Tor domains. Initially, TOIC contained 698 images in five categories: Drugs (both legal and illegal), Weapons, Money (counterfeit money), Personal ID (counter-

[2] http://pitia.unileon.es/varp/node/445

[3] http://pitia.unileon.es/varp/node/447

feit personal IDs such as driving licences, IDs and passports) and Credit Cards (counterfeit credit cards), as shown in Table 1. In this work, we crawled 500 new Tor domains and download a total of 783 new images. Only 15 images were unique and belonged to these categories. The rest of images were manually discarded. Therefore, the new version of TOIC contains 713 images distributed as shown in Table 1.

Table 1: Extended TOIC dataset. Number of images per category in the version of TOIC dataset used in this work.

Category	Original TOIC	New images	Extended TOIC
Drugs	369	4	373
Weapons	154	0	154
Money	74	1	75
Personal ID	49	1	50
Credit Cards	52	9	61
Total	698	15	713

3.2 IMPLEMENTATION DETAILS

For SIFT, we set the step and the size equals to 7, following [29]. With regard to BoVW, we used k-means [30] algorithm to cluster the visual words. We evaluated the method for different number of visual words, specifically we tested 250, 500, 750, 100, 1500 and 2048 visual words. The latter value of visual words, i.e. 2048, was chosen following the settings where TOIC dataset was published [16]. We used a hard assignment approach [31] to represent the images with feature vectors. We trained a Support Vector Machine classifier (SVM) [32] in order to classify new images into the selected categories.

We divided the dataset in two subsets: training that contains 70% of the images, and test that comprises the rest. We calculated the accuracy of the classification as the number of correctly classified images divided by the total number of images. We repeated the experiments five times, randomly selecting different sets of training and test to compensate for possible effects of random sub-sampling. We report the results as the average and standard deviation of the accuracy over the five runs.

3.3 RESULTS AND DISCUSSION

We represented the results as a confusion matrix in Figure 3. It is noticeable that the categories Drugs, Money and Weapons yielded better results than Weapons and Credit Cards. Specifically, Credit Cards can be easily mistaken with any other category apart from Drugs.

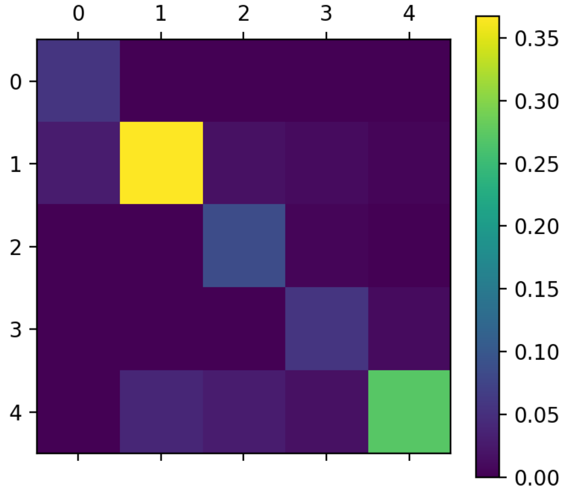


Figure 3: Confusion matrix. The rows of the matrix represent the instances in a predicted class whereas the columns represent the instances in an actual class. Classes 0 to 4 correspond to Credit Cards, Drugs, Money, Personal ID and Weapons, respectively.

Table 2 and Figure 4 show the results obtained in terms of mean and standard deviation of the accuracy yielded over the five runs for different number of visual words. We achieved the best results with HSV-SIFT no matter the number of visual words considered. The yielded accuracy with this method is importantly higher than with the rest, which obtain similar results. The best overall result is obtained with HSV-SIFT and 1500 words, yielding an accuracy of 59.44 ± 2.02%. We can conclude that the use of a combination of SIFT descriptors computed in different color spaces can outperform the SIFT descriptors of the gray scale images for TOIC dataset. The standard deviations are usually below 3%, which makes the results consistent the conclusions drawn.

4 CONCLUSIONS

In this work, we have used the Bag of Visual Words (BoVW) model to classify images crawled from domains of the the Tor network into the illegal activity according to the domain they belong to. Specifically, we have extended the TOIC (TOr Image Categories) dataset [16] with 15 new images extracted from the crawling of 500 new Tor domains, therefore using 713 images. In the BoVW model, SIFT (Scale Invariant Feature Transform) or dense SIFT has been used as low-level descriptors (i.e. the descriptors of the images patches or *visual words*), but SIFT operates on gray level images. In this work we have implemented and as-

Table 2: Results, in percentage, as mean and standard deviation of the accuracy over five runs for different amounts of visual words and the four proposed methods. In bold, we highlight the best results per amount of visual words. The best overall result appears underlined.

No. Visual Words	Descriptors			
	SIFT	HSV-SIFT	RGB-SIFT	Opponent-SIFT
250	42.94±0.40	**49.04±1.09**	42.60±0.64	43.16±0.50
500	42.37±1.24	**49.83±3.23**	42.03±0.65	42.71±0.51
750	46.87±2.00	**53.33±2.14**	49.38±4.22	47.57±2.51
1000	49.49±1.30	**53.56±5.44**	48.04±2.25	51.07±3.06
1500	53.44±1.86	<u>**59.44±2.02**</u>	51.48±1.81	53.67±2.00
2048	57.52±1.29	**58.07±1.68**	56.44±2.31	56.82±2.16

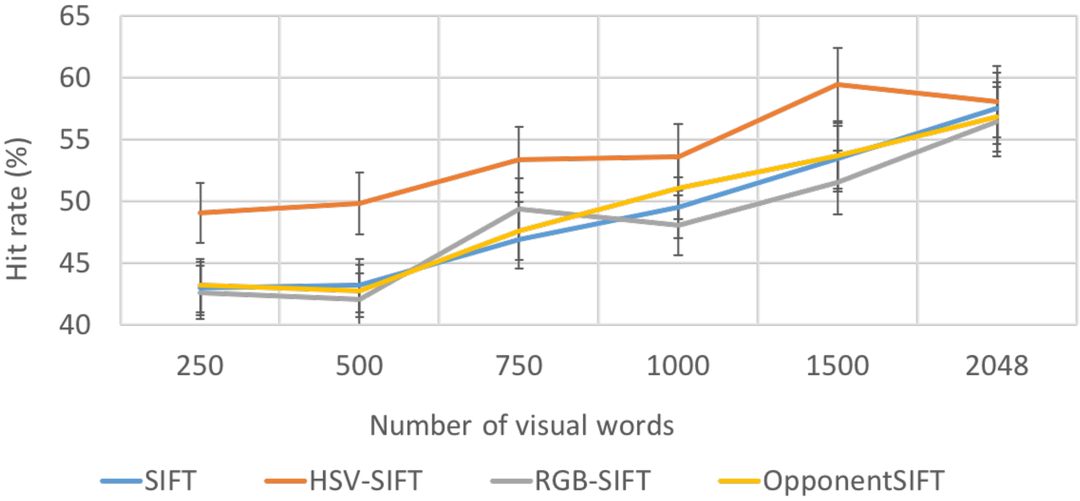


Figure 4: The color lines represent the mean of the accuracy over five runs (hit rates) for different amounts of visual words and the four proposed methods. The vertical lines represent the standard deviations.

sessed the performance of three different variants of SIFT descriptors extracted from color images, i.e. HSV-SIFT, RGB-SIFT and Opponent-SIFT, in the BoVW model for the task of classifying the images of the extended TOIC (i.e. the original TOIC with the 15 new crawled images). Whereas the best accuracy using the classical SIFT have been obtained using a visual dictionary of 2048 visual words (i.e., 57.52 ± 1.29%). This accuracy is overcome by all the color SIFT variants. However, by far the best result has been obtained using HSV-SIFT using a dictionary of 1500 visual words (i.e. 59.44±2.02%). This demonstrates the usefulness of using low-level descriptors that represent the information given by color in the image.

For future work, we plan to implement and assess more Color-SIFT descriptors in different datasets used for image-classification.

Acknowledgement

This research is supported by the INCIBE grant “INCIBEI-2015-27359” corresponding to the “Ayudas para la Excelencia de los Equipos de Investigación avanzada en ciberseguridad” and also by the framework agreement between the University of León and INCIBE (Spanish National Cybersecurity Institute) under Addendum 22.

References

[1] Bin Li, Zhongpeng Li, Shijun Zhou, Shunquan Tan, and Xiaoling Zhang. New Steganalytic Features for Spatial Image Steganography Based on Derivative Filters and Threshold LBP Operator. *IEEE Transactions on Information Forensics and Security*, 13(5): 1242–1257, 2018.

[2] Mhd Wesam Al Nabki, Eduardo Fidalgo, Enrique Alegre, and Iván de Paz Centeno. Classifying illegal activities on tor network based on web textual contents. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, Valencia, Spain, 04/2017 2017. Association for Computational Linguistics, Associa-

tion for Computational Linguistics.

[3] Loris Nanni, Alessandra Lumini, and Sheryl Brahnam. Survey on LBP based texture descriptors for image classification. *Expert Systems with Applications*, 39(3):3634–3641, 2012.

[4] M.T. García-Ordás, E. Alegre-Gutiérrez, V. González-Castro, and R. Alaiz-Rodríguez. Combining shape and contour features to improve tool wear monitoring in milling processes. *International Journal of Production Research*, 2018.

[5] V. González-Castro, E. Alegre, O. García-Olalla, L. Fernández-Robles, and M.T. García-Ordás. Adaptive pattern spectrum image description using euclidean and Geodesic distance without training for texture classification. *IET Computer Vision*, 6 (6), 2012.

[6] Koen E A van de Sande, T Gevers, and Cees G M Snoek. Evaluating Color Descriptors for Object and Scene Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9):1582–1596, 2010.

[7] Sivic and Zisserman. Video Google: a text retrieval approach to object matching in videos. In *Proceedings Ninth IEEE International Conference on Computer Vision*, pages 1470–1477 vol.2. IEEE, 2003.

[8] Gabriella Csurka, Christopher R. Dance, Lixin Fan, Jutta Willamowski, and Cédric Bray. Visual categorization with bags of keypoints. In *In Workshop on Statistical Learning in Computer Vision, ECCV*, pages 1–22, 2004.

[9] David G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[10] Florent Perronnin, Jorge Sánchez, and Thomas Mensink. Improving the fisher kernel for large-scale image classification. In *European conference on computer vision*, pages 143–156. Springer, 2010.

[11] Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce. Affine-invariant local descriptors and neighborhood statistics for texture recognition. In *ICCV*, pages 649–655, 2003.

[12] Jianchao Yang, Kai Yu, Yihong Gong, and T. Huang. Linear spatial pyramid matching using sparse coding for image classification. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1794–1801, June 2009.

[13] M. Paulin, M. Douze, Z. Harchaoui, J. Mairal, F. Perronin, and C. Schmid. Local convolutional features with unsupervised training for image retrieval. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 91–99, 2015.

[14] Krystian Mikolajczyk and Cordelia Schmid. An affine invariant interest point detector. In *European conference on computer vision*, pages 128–142. Springer, 2002.

[15] E. Fidalgo, E. Alegre, V. González-Castro, and L. Fernández-Robles. Compass radius estimation for improved image classification using Edge-SIFT. *Neurocomputing*, 197:119–135, 2016. ISSN 18728286. doi: 10.1016/j.neucom.2016.02.045.

[16] Eduardo Fidalgo, Enrique Alegre, Victor González-Castro, and Laura Fernández-Robles. Illegal activity categorisation in DarkNet based on image classification using CREIC method. In *Advances in Intelligent Systems and Computing*, volume 649, pages 600–609, 2018.

[17] Wesam Al-Nabki, Eduardo Fidalgo, Enrique Alegre, and Victor Gonzalez-Castro. Detecting Emerging Products in Tor Network Based on K-Shell Graph Decomposition. *III Jornadas Nacionales de Investigación en Ciberseguridad (JNIC)*, 1:24–30, 2017.

[18] Eduardo Fidalgo, Enrique Alegre, Victor GonzÃ¡lez-Castro, and Laura FernÃ¡ndez-Robles. Boosting image classification through semantic attention filtering strategies. *Pattern Recognition Letters*, 112:176 – 183, 2018. ISSN 0167-8655. doi: https://doi.org/10.1016/j.patrec.2018.06.033.

[19] Emmanuel Maggiori, Yuliya Tarabalka, Guillaume Charpiat, and Pierre Alliez. Convolutional neural networks for large-scale remote-sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55 (2):645–657, 2017.

[20] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[21] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *AAAI*, volume 4, page 12, 2017.

[22] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.

[23] Andrea Vedaldi and Brian Fulkerson. Vlfeat. *Proceedings of the international conference on Multimedia - MM '10*, 3(1):1469, 2010.

[24] Koen Van De Sande, Theo Gevers, and Cees Snoek. Evaluating color descriptors for object and scene recognition. *IEEE transactions on pattern analysis and machine intelligence*, 32 (9):1582–1596, 2010.

[25] A. Bosch, A. Zisserman, and X. Mu noz. Scene Classification Using a Hybrid Generative/Discriminative Approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(4):712–727, 2008.

[26] Abhishek Verma, Chengjun Liu, and Jiancheng Jia. New colour sift descriptors for image classification with applications to biometrics. *International Journal of Biometrics*, 3(1):56–75, 2010.

[27] Matthew Brown and Sabine Süsstrunk. Multi-spectral sift for scene category recognition. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 177–184. IEEE, 2011.

[28] Koen EA van de Sande, Theo Gevers, and Cees GM Snoek. Color descriptors for object category recognition. In *European Conference on Color in Graphics, Imaging and Vision*, pages 378–381. Society for Imaging Science and Technology, 2008.

[29] L. Xie, Q. Tian, M. Wang, and B. Zhang. Spatial pooling of heterogeneous features for image classification. *IEEE Transactions on Image Processing*, 23(5):1994–2008, May 2014. ISSN 1057-7149. doi: 10.1109/TIP.2014.2310117.

[30] J. A. Hartigan and M. A. Wong. A k-means clustering algorithm. *JSTOR: Applied Statistics*, 28(1):100–108, 1979.

[31] Gabriella Csurka, Christopher R. Dance, Lixin Fan, Jutta Willamowski, and Cédric Bray. Visual categorization with bags of keypoints. In *In Workshop on Statistical Learning in Computer Vision, ECCV*, pages 1–22, 2004.

[32] Marti A. Hearst. Support vector machines. *IEEE Intelligent Systems*, 13(4):18–28, July 1998. ISSN 1541-1672. doi: 10.1109/5254.708428. URL `http://dx.doi.org/10.1109/5254.708428`.