Classification of incident-related images using machine learning

Primary Topic: IPCV, Secondary Topic: DM
Course: 2022-1B 201400174 – Group: 150 – Submission Date: 2023-02-05

Akram Sabik
École des Mines de Saint-Étienne x University of Twente
s.akram@student.utwente.nl

ABSTRACT

Natural disasters necessitate quick intervention to deal with human emergencies efficiently. With the emergence of social media, large amounts of images can be quickly accessed when an incident is happening and classifying them is primordial. While image classification tasks in machine learning have been studied thoroughly, recent advances show that combining feature extraction with deep learning models lead to better accuracy performance with less complex and computationally-expensive networks. In this work, feature extraction methods are combined with dimension reduction in order to create pipelines that generate feature vectors. These vectors are used as input to train simpler models for classifying incident-related images. By exploring the parameters of the pipelines, a better interpretability of the models and understanding of the tradeoff between data reduction & performance can be reached.

KEYWORDS

Computer Vision, Data Mining, Image Classification, Supervised Learning, Image Processing, Feature Extraction, Data Reduction

1 INTRODUCTION

The detection of natural disasters is primordial as they require quick intervention in order to control the damage, particularly in case of human emergencies. Collecting and analysing information is still a costly and often inefficient task as it requires a lot of manual effort. Satellite imagery has been used[3], however this method is still limited and requires absence of sky occlusion (clouds or smoke) to work, which are a normal occurrence for natural disasters. The paper [6] proposes a way to automate information processing about natural disasters from social media posts. The dataset Incidents1M is introduced: a large-scale, multi-label incident images dataset meant to train complex models to classify incident-related images. High-dimensional data, such as raw images, can be difficult to work with and can lead to problems such as the curse of dimensionality. This is due to the increased number of parameters and the complexity of the models required to process the data. Working with large datasets or deploying models in real-time applications costs computational resources and results in timely training process. One way to solve this is to use feature extraction techniques. By extracting relevant features from the images, the dimensionality of the data is reduced, making it more manageable and efficient to work with. Thus, the following research question will be addressed:

To what extent can feature extraction techniques help match the performance of state-of-the-art neural networks in image classification, while being less computationally expensive and more interpretable ?

2 BACKGROUND AND RELATED WORK

2.1 Background

2.1.1 Feature descriptors. A feature in computer vision is a region of interest in an image that is unique and easy to recognize. Features include things like, points, edges, and corners.

- **Scale-invariant feature transform (SIFT)** are designed to be robust to image rotation, scaling, and affine distortion. They are widely used in object recognition and image matching tasks because of their ability to identify distinctive local features in an image.
- **Gabor filter** is a linear filter used for texture analysis. A 2D Gabor filter is essentialy a Gaussian kernel function modulated by a sinusoidal plane wave. Gabor filter-based features are designed to capture the texture and shape information of an image. They are particularly useful for tasks where texture and shape information is important.

2.1.2 Bag of features (BoF). A Bag-of-features , or bag-of-visual-words is a method to represent the features of images. BoF is inspired by the bag-of-words model often used in the context of Natural Langage Processing. In the context of computer vision, BoF can be used to train classifiers or generative models. The usual steps are as follows: Feature Extraction, Vector Quantization (Creation of BoF often through clustering), Feature Vector Generation.

2.2 Related Works

[5] presents a class-negative loss to train a robust model for detecting incidents in the wild and demonstrate the model's effectiveness in identifying incidents in large collections of social media images. [6] present an extension of the Incidents dataset to a larger, multi-label dataset, known as the Incidents1M Dataset. The authors analyze the composition and potential biases in the data, perform incident detection experiments on social media images, and present preliminary experiments on the performance of the incident detection model. Regarding the use of SIFT feature extraction in machine learning, [4] presents a novel model, SIFT-CNN, that combines the strengths of SIFT descriptors and CNNs. SIFT descriptors are computed for every pixel in a single-channel image, generating a multi-channel SIFT image. A CNN is then trained to use these SIFT images as inputs. The SIFT-CNN outperforms CNNs trained directly on pixel values in various challenging tasks. Similarly, [1] presents a method for content-based image classification using SIFT for feature extraction, K-means clustering for feature grouping, BoF construction for image representation, and SVM for classification. This combination leads to a classification with an accuracy rate of about 90%. [2] demonstrated that, in comparison to traditional deep convolutional neural networks (DCNN), Gabor convolutionnal

networks (GCN) have superior performance in object recognition, require fewer network parameters to be learned, and have a more compact structure.

3 DATASET PRESENTATION

The Incidents1M dataset [6] is a multi-label dataset of incident-related images across multiple locations. It contains multi-label category annotations and has been created to facilitate incident analysis and encourage the use of computer vision for humanitarian aid. Here, we use a small subset of around 7000 images and only keep a single label for incidents. There is a total of 12 labels. Fig.8 presents an image example of each label. Fig.1 presents a distribution of each label across the dataset.

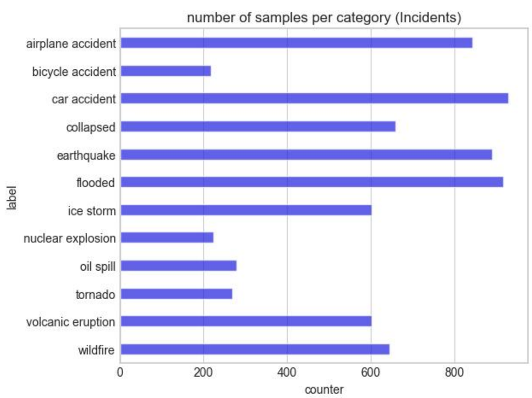

Figure 1: Categories Histogram

4 METHODOLOGY

4.1 Roadmap

Fig.3 presents the general roadmap. Both descriptors have a dedicated feature extraction pipeline. In order to get closer to an optimal pipeline for each pipepline, the roadmap (Fig.3) is used to first create a baseline set for both SIFT and GABOR features. Then, additional sets are created with varying parameters in order to study their effects on the chosen metrics.

The additional sets are compared based on the accuracy of the logistic regression, the accuracy of the random forest, and the total explained variance resulting from two-dimensional PCA.

All sets go through pre-processing steps: All values are scaled by removing the mean and scaling to unit variance. Data augmentation is performed with random over-sampling in order to generate 1000 samples for each class. Finally, a 80/20 train/test split is performed where classes remain well balanced.

The logistic model is trained with 5-fold cross-validation using the LBFGS solver, and 500 max iterations. The random forest is trained with 50 estimators.

4.2 SIFT features extraction

4.2.1 SIFT features roadmap. Fig.4 presents the pipeline for SIFT feature extraction. An example image for each step helps visualize the process from start to finish. Fig.9 is an example of SIFT features extraction. Fig.2 is a histogram of the number of descriptor that is generated across the dataset (See Fig.4, (1)). Some images have almost no SIFT descriptor while other have up to 1400. On average, an image generates 490 descriptors. In order to process the dataset in a homogenous way, all images should be described by the same number of descriptors. In order to do so, 30 descriptors are picked at random for each image for the baseline set (See Fig.4, (2))

4.2.2 SIFT features parameters exploration. The current SIFT extraction has 3 main parameters:

- Number of randomly picked descriptors
- Bag of Features Size
- N-Closest Clusters

In order to determine how each parameter affects the performance of the SIFT features, SIFT sets with a different number of randomly-picked descriptors and number of closest clusters are compared with the baseline set independently. The tested number of randomly picked descriptors are 100, 200 and 400. The tested number of randomly picked descriptors are 2, 5 and 10.

4.3 GABOR features extraction

4.3.1 GABOR features roadmap. Fig.5 presents the pipeline for GABOR feature extraction. Fig.10 presents the 24 kernels that are used in the baseline set. Each kernel has its own set of parameters and will extract different information from the image. The chosen parameters for the GABOR kernels are:

- Kernel size: 9
- λ: $\frac{\pi}{4}$, $\frac{2\pi}{4}$,
- θ: 0, $\frac{1}{4\pi}$
- ψ: 0
- σ: 1,2,3
- γ: 0.05, 0.5

4.3.2 GABOR features parameters exploration. The current GABOR extraction has 2 main parameters:

- Number of parameter combinations (number of kernels)
- Number of bins per image

More parameter combinations allows to extract more information, but requires more computational power and time. Binning the filtered images reduces the dimensionality and makes the model more robust by preventing overfitting, however, it causes loss of information to some extent. The goal is to test whether it is preferable to:

- use less GABOR filters & keep more information from each filtered image ?
- use more GABOR filters & keep less information from each filtered image ?

In order to determine how each parameter affects the performance of the GABOR features, GABOR sets with varying number of bins per image and number of used kernels are compared with the baseline set independently. The tested bin sizes are 16, 32 and 64. The tested number of kernels is 42. The parameters are similar to the baseline kernels with additional θ values of $\frac{1}{2\pi}$ and $\frac{3}{4\pi}$ which results in 48 unique combinations.

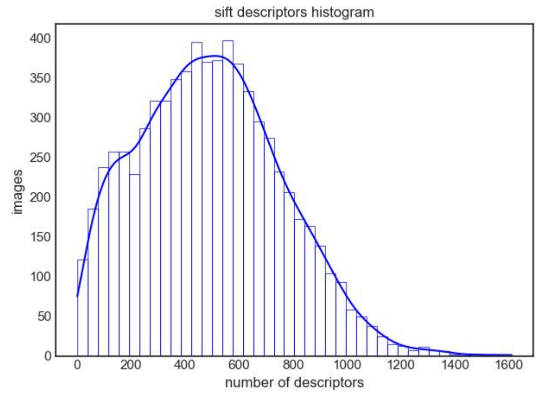

Figure 2: Sift Descriptors Histogram

5 RESULTS

5.1 Baseline set

Fig.11 presents the confusion matrix for both models for SIFT and GABOR baseline sets. Fig.6 depicts the performance of the baseline set along with the influence of the parameters. Both SIFT and GABOR baseline sets perform with low accuracy for logistic regression compared to random forest, which is expected. The explained variance between both descriptors is very different, with about 4% of explained variance for SIFT descriptors compared to 51% for GABOR descriptors as observed in Fig.7. The SIFT PCA plot doesn't permit at all to differentiate each class as they are all gathered in the center. It is easier to distinguish each class for the GABOR plot, although it remains a somewhat difficult task because of the number of categories.

5.2 Parameter study

5.2.1 *SIFT.*

Fig.6 presents the parameters influence on all models performance. The number of randomly-picked SIFT descriptors barely changes the final accuracy on both logistic regression and random forest models(+2%), which means that generating the cluster with 30 descriptor from each image is enought. Knowing can help save computational power. There is however a slight improvement (+10%) for the PCA's total explained variance. It can be observed that classes are further apart around the center, however it is still not enough compared to GABOR's performance

The performance is not improved when a SIFT descriptor is described by more than its closest cluster in the histogram. However, the explained variance increases more significantly, which makes it easier to distinguish each category. The total explained variance is still low.

5.2.2 *GABOR.*

The number of bins per image doesn't affect the random forest final accuracy but enhances the logistic regression's performance significantly (+20%). The increasing number of bins also leads to longer train and test time for the models (The process can take up to 4 times as long). The explained variance is also slightly lower with higher bin numbers.

The number of kernels doesn't impact accuracy or explained variance for both models, but leads to higher processing time (4 times as long as for the baseline set)

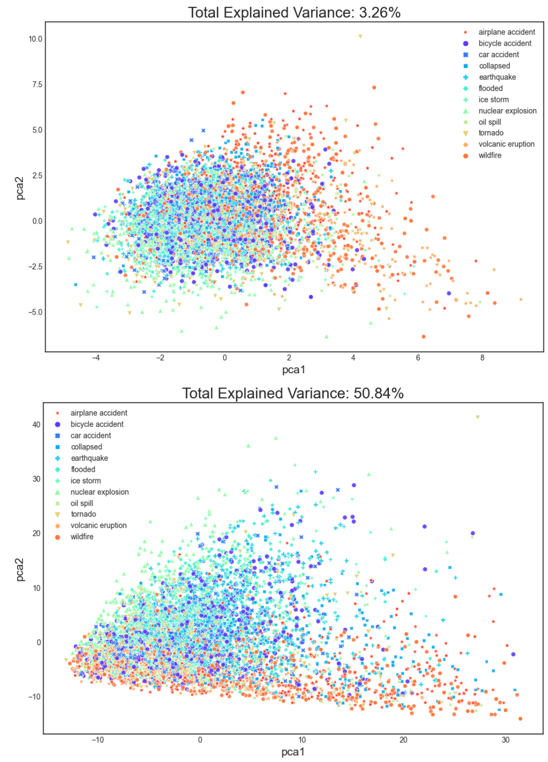

Figure 7: 2d PCA plot for SIFT and GABOR baseline sets. (See Fig.12)

6 DISCUSSION

6.1 Levels of abstraction in similar incident categories

Fig.11 presents the confusion matrix for both models for SIFT and GABOR baseline sets. It can be observed that pairs of similar categories are often misclassified. This is happening for less performing models (SIFT logistic regression) but also for more performant ones (GABOR random forest). The most obvious ones being the pairs *Wildfire & Volcanic eruption*, *Ice Storm & Flooded*, *Collapsed & Earthquake*. This is expected as images from each pair will have several visual similarities. For example, the model at a low level of abstraction can learn how to recognize fire or water when presented an image. However, more clues are needed to actually recognize the context. Reaching this higher level of abstraction is the biggest challenge in this method. The reason convolutional neural networks work so well for this task is because the convolutional layers manage to learn features with increasing abstraction. As it can be observed in Fig.12 in the GABOR PCA plot, the fire-related pair seems to form a horizontal line toward the bottom of the plot while the water-related pair is more scattered in the upper half. This proves that the random forest is indeed able to learn these low-level

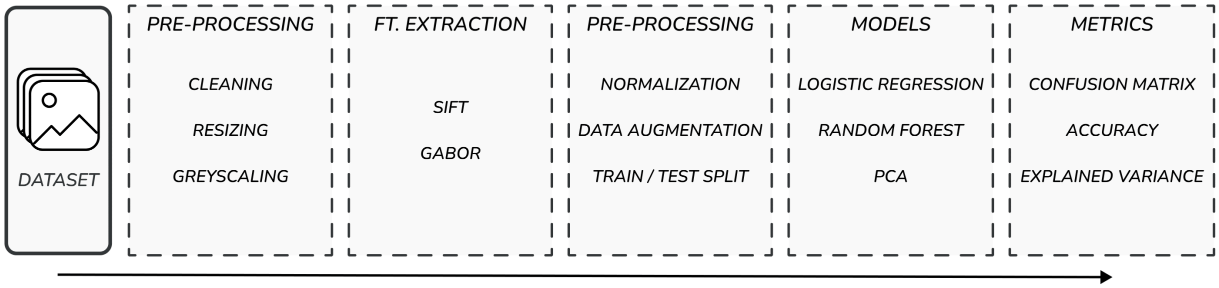

Figure 3: Project Roadmap. First, the images go through standard pre-processing. Next, SIFT and GABOR features are extracted from each image resulting in feature sets. Each set is then appropriately pre-processed to be fed into the models. Finally, metrics are used to assess the performance of the models and allow comparison between the generated feature sets.

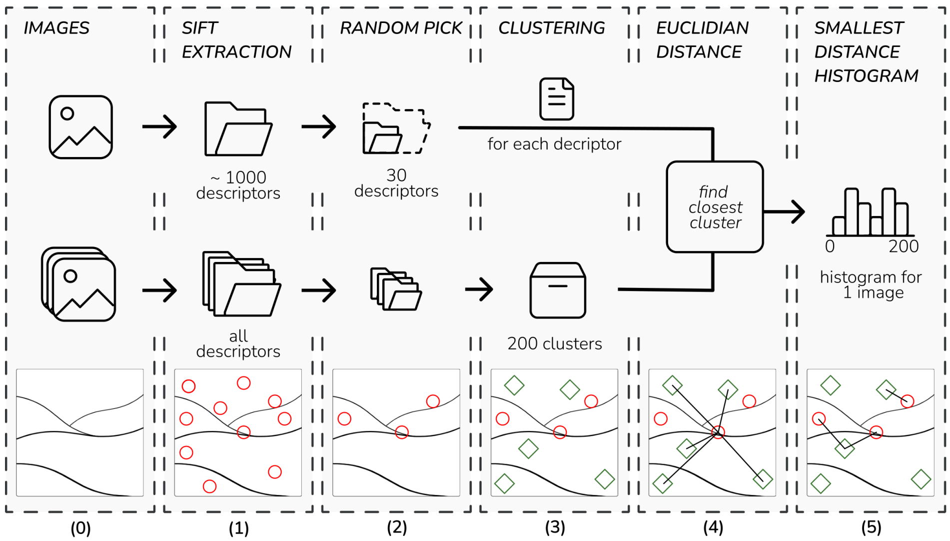

Figure 4: SIFT extraction Roadmap. (1) An image of the dataset generate an average of 490 descriptors. (2) 30 (baseline) descriptors are then randomly selected for each image. (3) All the selected descriptors from all images are clustered into a boF of 200 descriptors. (4) For each image, the 1-closest (baseline) cluster to each of its descriptors is determined based on euclidian distance (5) A histogram of how many times each cluster descriptor comes up during the previous step represents the final feature vector.

features but because of that they also remain difficult to tell apart from each other as seen in the confusion matrices. On the other hand, the low performance of SIFT features can be explained by two reasons. One reason might be that describing all images with 200 cluster isn't enough to learn the low-level features, and more clusters could solve this. Another reason could be that the final histogram isn't populated enough. The PCA values are significantly smaller in the SIFT plot compared to the GABOR plot. Furthermore, according to Fig.6, enhancing the number of closest SIFT cluster does enhance the explained variance from 3.5% to 15.6%. These two reasons also help explain the lower accuracy when using SIFT descriptors instead of GABOR descriptors.

6.2 Interpretability

The choice of parameters of GABOR kernels for the baseline set aims to cover most cases but since there are many parameters to GABOR kernels, there are multiple relevant combinations of kernels. Studying which kernels are most efficient for a classification decision can help bring insight to which features are more important in the inputs and can help achieve better interpretation. According to Fig.6, more kernels isn't necessarily better as it doesn't increase the final accuracy of the models. However, enhancing the number of bins per image significantly enhances the logistic regression's accuracy (+20%). This makes sense because more bins means more information available for the model to learn. This is particularly interesting as this model is interpretable and can be used to find explanations on its learning process that are comprehensible for humans. With close accuracies, the model can also act as a surrogate model to explain a black-box model such as the random forest.

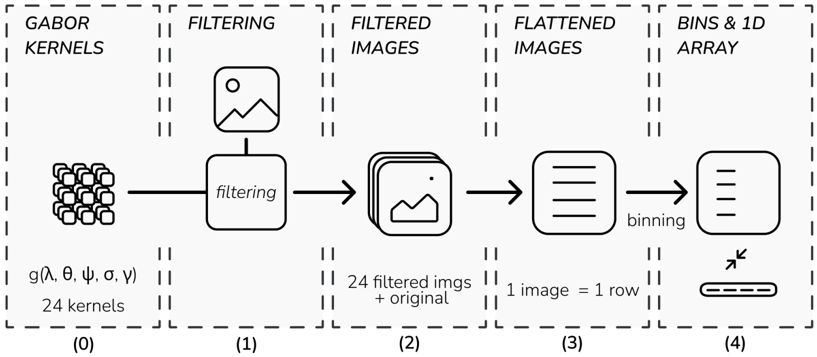

Figure 5: GABOR extraction Roadmap.(1) Each image of the dataset is filtered using the 24 GABOR kernels (baseline). (2) Each image generates 24 new images, each one different from one another. (3) Each image is then flattened to 1 dimension, resulting in an array where each row represents one image. (4) Each row is binned into 8 values (baseline) and put next to one another. The final 1D array represents the feature vector for the image.

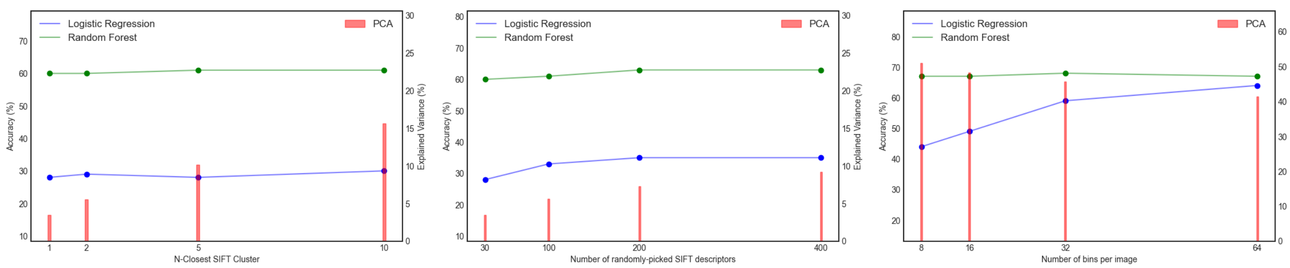

Figure 6: Parameters' influence on model performance. Left y-axis: Accuracy (%); Right y-axis: Explained Variance (%)

6.3 Improvements

Data reduction is an important aspect regarding how the research question is approached in this paper and is present under different forms: Randomly picking a fixed amount of SIFT descriptors, clustering them into a BoF, binning filtered GABOR images to a fixed value... While data reduction is an efficient way to reduce computational cost, it also causes loss of information which can make learning more difficult for models. This tradeoff needs to be balanced and can be further studied in order to reach better performance on the models. Experimenting with the number of cluster in the SIFT pipeline can help bring more insight on how to get better performance with this descriptor. An implementation that could help lose less valuable information while performing data reduction would be to replace the random SIFT descriptors pick by an encoder that would create a latent space. A PCA plot showing only the previously mentioned similar category pairs could also improve readability and interpretation of the results. Finally, concatenating SIFT and GABOR with the now known optimal parameters could produce an even more performant set for the classification task.

7 CONCLUSIONS

Combining feature extraction techniques with deep learning models has proved to be efficient in image classification tasks and reached remarkable results. This paper tried to bring this approach to incident-related images by focusing especially on the tradeoff between data reduction and information as well as intrepretability of the models instead of final accuracy. The exploration of the parameters of the two pipelines brings insight on what matters most when it comes to data reduction and the challenges of learning high level features in simpler models.

REFERENCES

[1] Li, Q. and Wang, X. (2018). Image classification based on sift and svm. In *2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS)*, pages 762–765.

[2] Luan, S., Chen, C., Zhang, B., Han, J., and Liu, J. (2018). Gabor convolutional networks. *IEEE Transactions on Image Processing*, 27(9):4357–4366.

[3] Skakun, S., Kussul, N., Shelestov, A., and Kussul, O. (2014). Flood hazard and flood risk assessment using a time series of satellite images: A case study in namibia. *Risk Analysis*, 34(8):1521–1537.

[4] Tsourounis, D., Kastaniotis, D., Theoharatos, C., Kazantzidis, A., and Economou, G. (2022). Sift-cnn: When convolutional neural networks meet dense sift descriptors for image and sequence classification. *Journal of Imaging*, 8(10).

[5] Weber, E., Marzo, N., Papadopoulos, D. P., Biswas, A., Lapedriza, A., Ofli, F., Imran, M., and Torralba, A. (2020). Detecting natural disasters, damage, and incidents in the wild.

[6] Weber, E., Papadopoulos, D. P., Lapedriza, A., Ofli, F., Imran, M., and Torralba, A. (2022). Incidents1m: a large-scale dataset of images with natural disasters, damage, and incidents.

APPENDIX

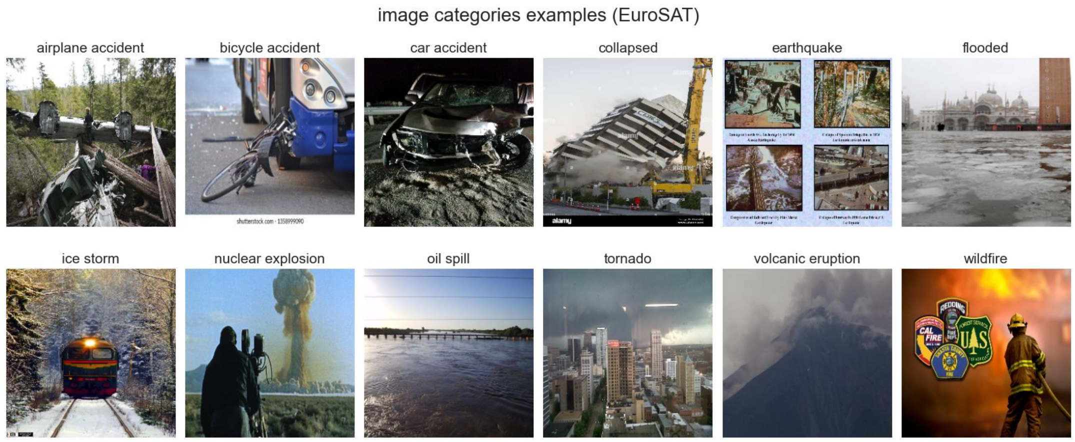

Figure 8: 1 image from each category of the dataset

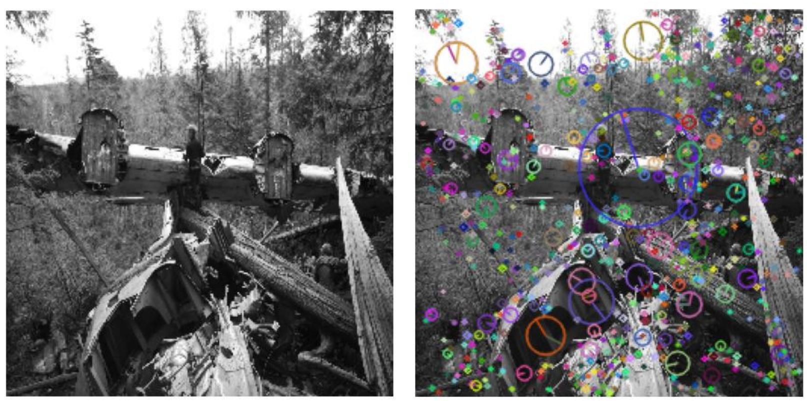

Figure 9: Left: Image Example (Greyscale). This image belongs to the category *Airplane Accident*. Right: Image on the left with extracted SIFT descriptors. Each circle represents a descriptor. A descriptor is characterised by its size (scale) and direction (rotation).

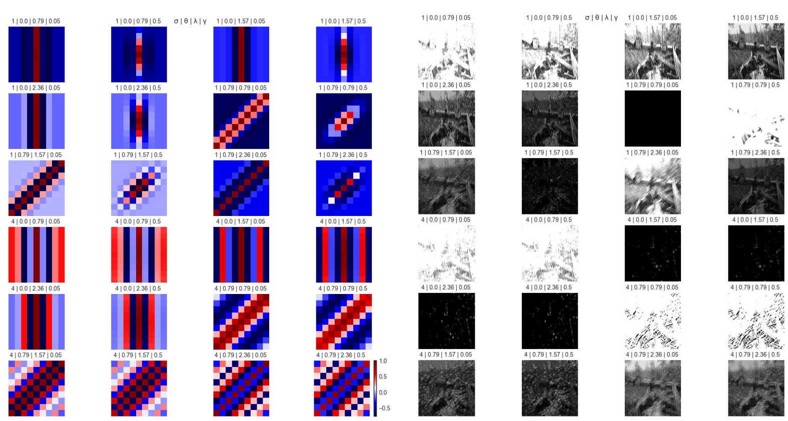

Figure 10: Left: Baseline set of 24 GABOR Kernels. The color map allows to easily distinguish negative values (blue) and positive values (red). Each kernel is created with a different set of parameters which allows it to pick up on different features of the filtered image. Right: Left image in Fig.9 filtered with the 24 baseline GABOR kernels on the left.

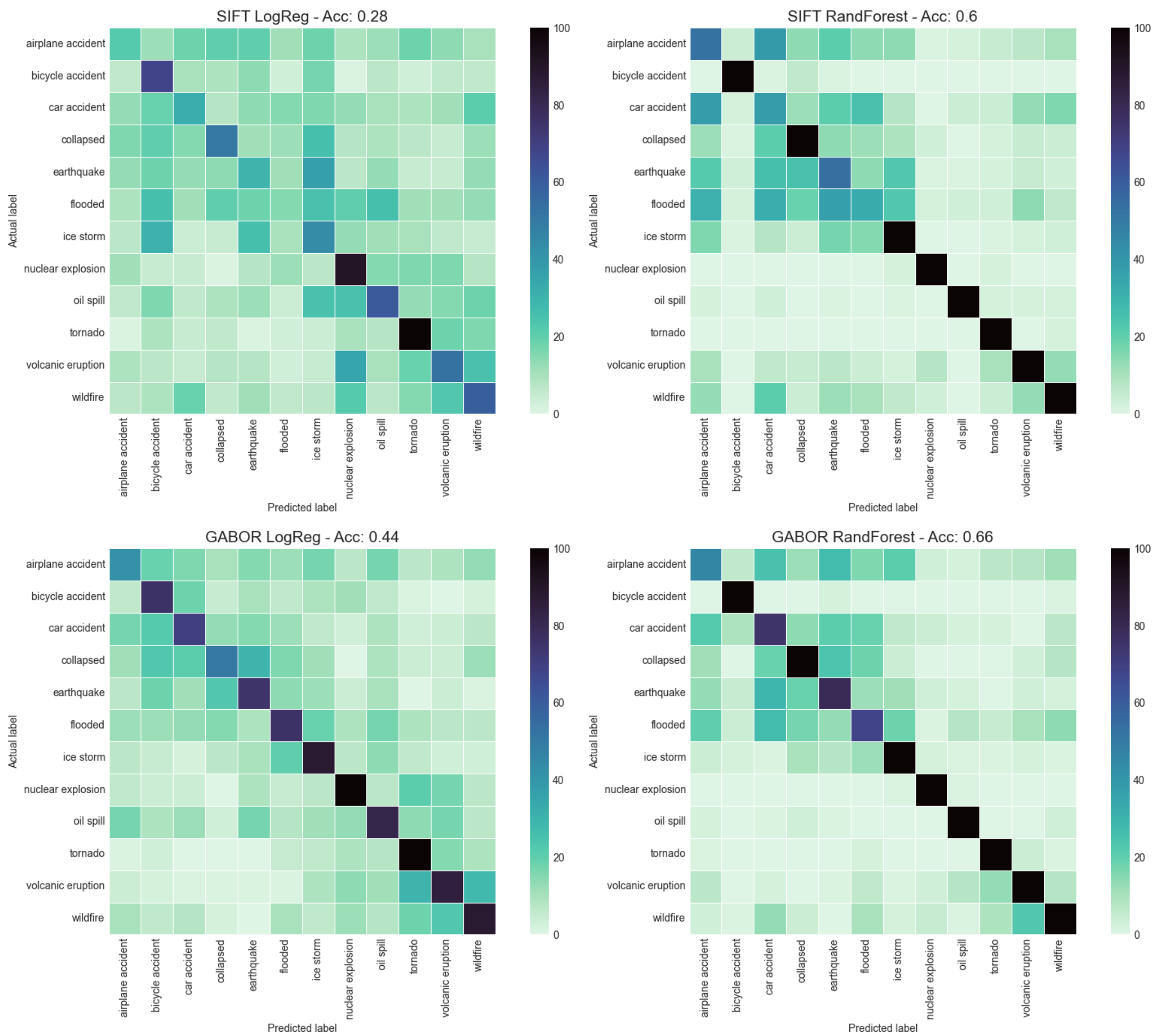

Figure 11: Confusion Matrix for logistic regression and random forest models for SIFT and GABOR baseline sets

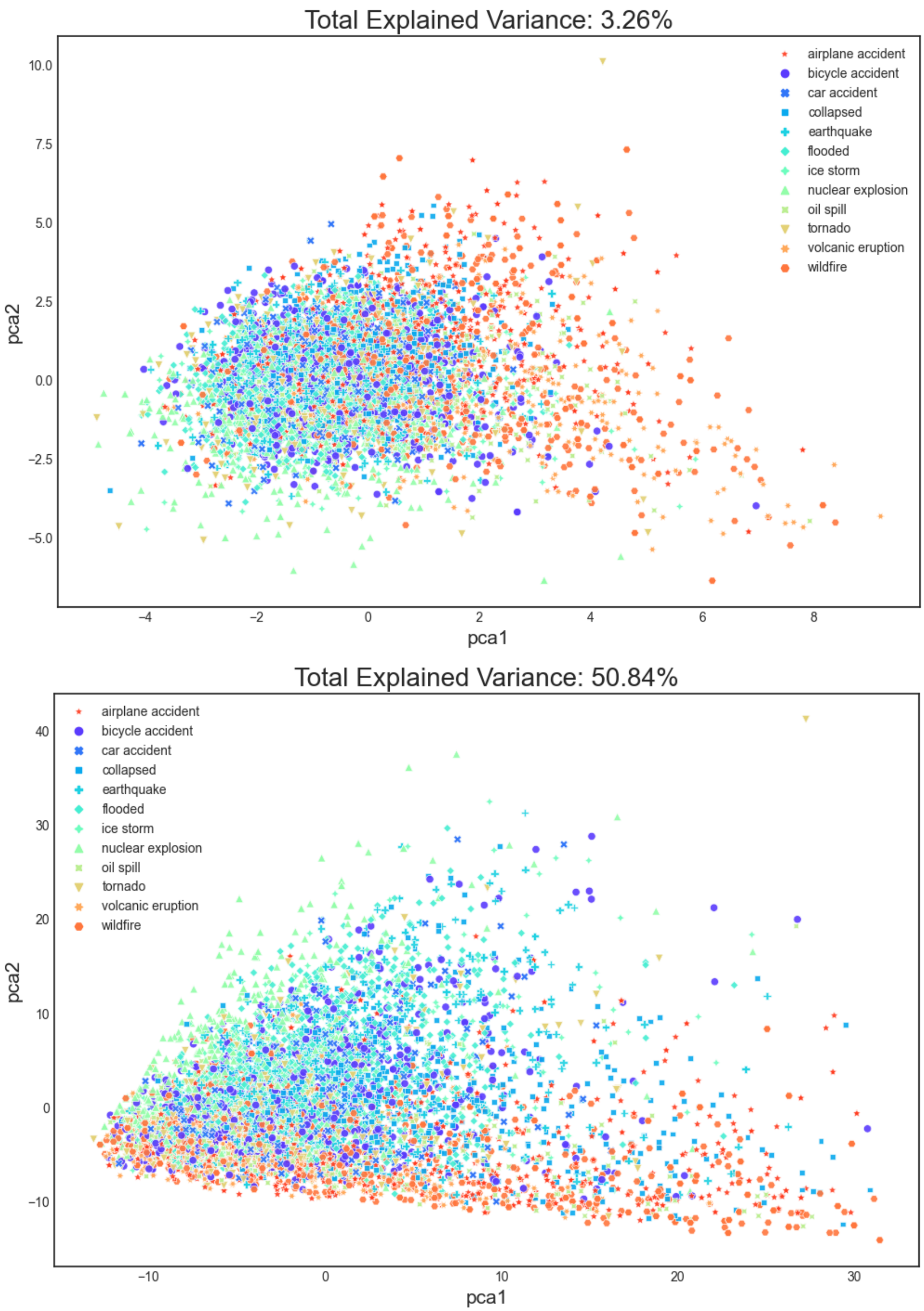

Figure 12: 2d PCA plot for SIFT and GABOR baseline sets.