ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/332933805

A Study on Machine Learning for Steganalysis

Conference Paper · May 2019

DOI: 10.1145/3310986.3311000

CITATIONS
6

READS
5,970

1 author:

Ki-Hyun Jung
Kyungil University
89 PUBLICATIONS **1,326** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Development of Detection and Extraction Techniques on Digital Contents Forgery View project

Steganography in Emulab Environments View project

All content following this page was uploaded by Ki-Hyun Jung on 10 May 2019.

The user has requested enhancement of the downloaded file.

A Study on Machine Learning for Steganalysis

Ki-Hyun Jung
Department of Cyber Security, Kyungil University
50 Gamasil-gil, Hayang-eup, Gyeongsan, Gyeongbuk 38428 Korea
+82-53-600-5626
khanny.jung@gmail.com

ABSTRACT
Data security is very important when sensitive data are transmitted over the Internet. Steganography and steganalysis techniques can solve the problem of copyright, ownership, and detection malicious data. Steganography is to hide secret data without distortion and steganalysis is to detect the presence of hidden data. In this paper, steganography and steganalysis techniques are described together with machine learning frameworks to show that machine learning framework can be used to detect the secret data hiding in image using steganography algorithms.

Keywords
Steganography; Steganalysis; Machine Learning; Deep Learning.

1. INTRODUCTION
Steganography is to conceal the secret data within multimedia contents such as file, message, image, or video. Steganography is concerned with concealing the fact that the secret data is being sent covertly as well as concealing the contents of the secret data [1-2]. Steganalysis is the counter part of steganography that defined as the art of science of detecting the hidden secret data in cover objects. In other words, steganalysis is to detect secret data hidden using steganography, where steganalysis is to identify suspected packages, determine whether the secret data is embedded or not [3-5]. Machine learning is a field of artificial intelligence to provide the ability to learn without being programmed and deep learning is a subset of machine learning [6-8]. In this paper, steganalysis and machine learning techniques are explained and the process and possibility for steganalysis in various machine learning frameworks are described. Some datasets on stego-images are prepared and training model are tested.

2. STEGANOGRAPHY
Steganography can be categorized as what kind of criteria is used. In Figure 1, steganography techniques are divided according to multimedia data types and domains.

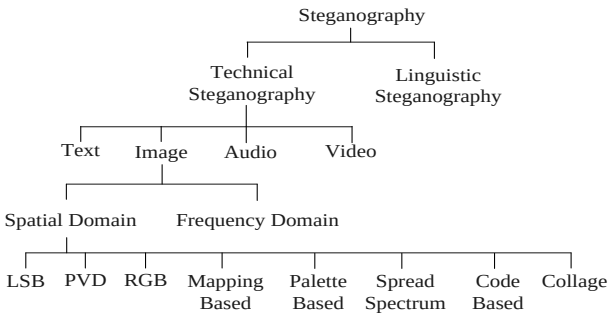


Figure 1. Classification of steganography.

In other way, steganography can be divided into reversible and irreversible data hiding techniques whether the cover object can be recovered, where the contents of secret data can exactly extracted. In irreversible data hiding, the least significant bits (LSB) replacement and pixel-value differencing (PVD) techniques are used in tradition. In reversible data hiding, difference expansion (DE), histogram shifting (HS) and prediction-error expansion (PEE) are known in the spatial domain [1-2]. Reversible data hiding techniques are used to solve the problem of lossless embedding in sensitive images such as military images, medical images, and artwork preservation [13-14].

3. STEGANALSYS
Steganalysis is to identify suspected data, determine hidden data, and recover the hidden data. Steganalysis can be divided into four categories: visual, structural, statistical, and learning steganalysis.

Visual steganalysis is to investigate visual artifacts in the stego-images, where try to catch visual difference by analyzing stego-images. Structural steganalysis looks into suspected signs in the media format representation since the format is often changed when the secret message is embedded. RS analysis and pair analysis are included in the structural steganalysis. Statistical steganalysis utilizes statistical models to detect steganography techniques. Statistical steganalysis can be divided into specific statistical and universal statistical steganalysis. Learning steganalysis also called blind steganalysis is one of universal statistical steganalysis since cover images and stego-images are used as training datasets.

Other classification of steganalysis can be divided into six categories as shown in Figure 2 [5]. It is depending on what kind of attacks a forensic examiner uses.

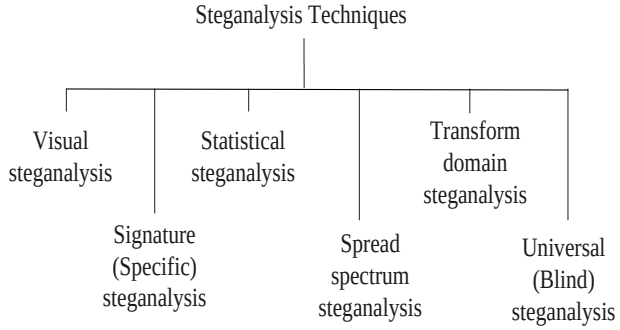


Figure 2. Classification of steganalysis.

In Figure 2, universal or blind steganalysis techniques are based on detecting the secret messages without regard to steganography techniques. Comparing with other steganalysis techniques, universal steganalysis technique is very difficult to find extraction

features, where machine learning techniques are often used to build, train, and evaluate models.

4. MACHINE LEARNING

Machine learning is one of artificial intelligence researches which is a computer-based method of learning as the ability of human brains [6-9]. Machine learning can be divided into three main categories: supervised learning, unsupervised learning and reinforcement learning. Supervised learning is to learn a function that maps an input to an output by giving input-output pairs, which is used often in speech recognition, spam detection and object recognition. Unsupervised learning is the task of learning from test data that has not been labeled, classified or categorized. Main application of unsupervised learning is in the field of cluster analysis, principal component analysis, vector quantization and self-organization. Reinforcement learning is concerned with how to take actions to maximize some notion of cumulative reward. Reinforcement learning is used in robotics, investment decisions, and inventory management to learn actions to be performed. Deep learning is a part of machine learning which is based on learning data representations.

To develop machine learning algorithms, many frameworks are used such as TensorFlow, Theano, Keras, Caffe, Torch, Deep Learning 4j, MxNet, CNTK, Lasagne, BigDL and so on. The following section explains some of frameworks in current.

In this paper, three frameworks are explained bellows since scikit-learn, TensorFlow, and Keras were used to test stego-images.

4.1 Scikit-learn

Scikit-learn framework is a free machine learning library for python that provides classification, regression and clustering algorithms [10]. Scikit-learn provides supervised and unsupervised learning algorithms by Python interface that is built upon the SciPy. The library is focused on modeling data. For unsupervised learning, scikit-learn provides various clustering and decomposition algorithms that are simple to use.

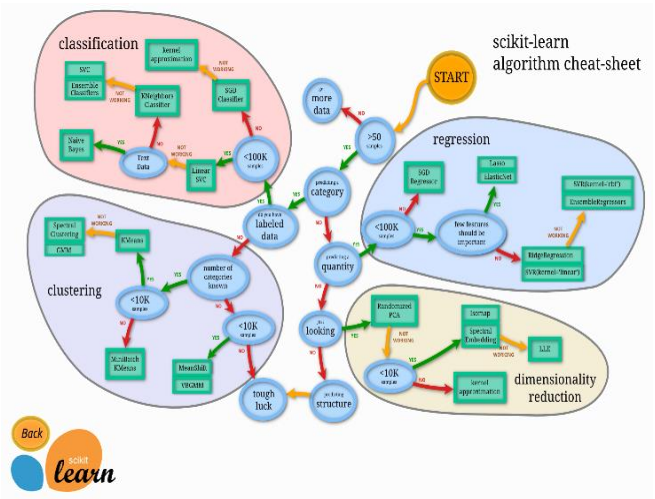


Figure 3. Flowchart of scikit-learn algorithm.

4.2 TensorFlow

TensorFlow is an open library to allow deployment of computation across a various platforms like as CPUs, GPUs, and TPUs [11]. The runtime library is a cross-platform and the C API separates user level code in different languages from the core runtime shown in Figure. 4. The runtime library contains over 200 standard operations including mathematical, array manipulation, control flow, and state management operations. TensorFlow is good for deep learning applications.

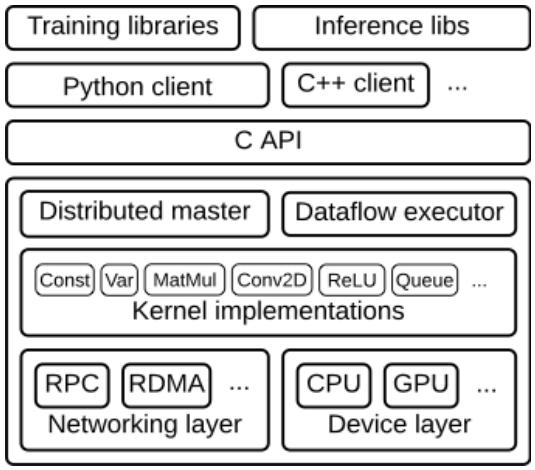


Figure 4. TensorFlow architecture.

4.3 Keras

Keras is a high-level neural networks library written in Python and capable of running on top of TensorFlow, CNTK, or Theano [12]. The core data structure is a model to organize layers. Keras contains neural network building blocks such as layers, objectives, activation functions, optimizers to implement with image and text data easier.

5. STEGANALYSIS TESTS

5.1 Steganalysis Flowchart

The process to apply the machine learning for steganography analysis is shown in Figure 5. It consists of four steps: data collecting, model building, training and evaluating.

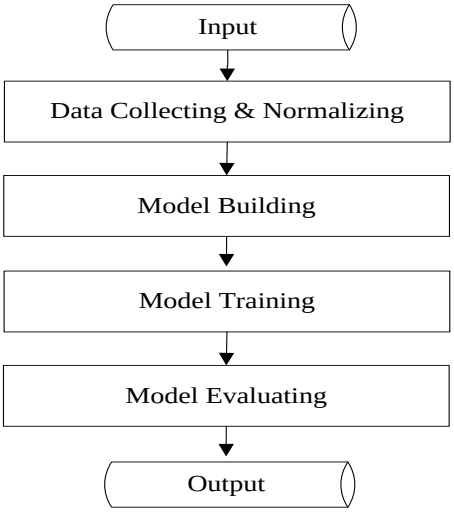
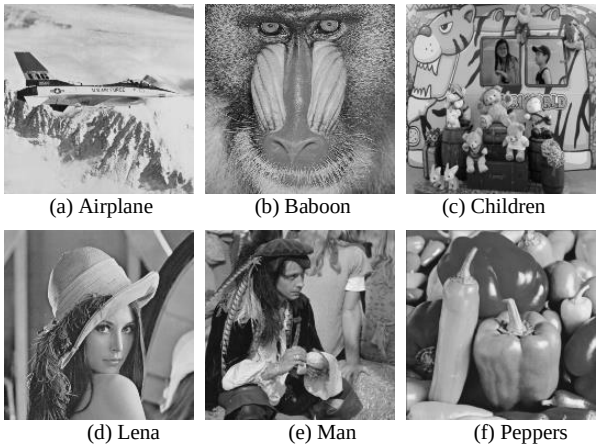


Figure 5. Machine learning process for steganalysis.

Many training data and test data should be collected and normalized to increase accuracy. In particular, vectorization must be performed through feature extraction from images since images are used in steganography and steganalysis in general. In model building, training, and evaluating steps, various algorithms can be utilized depending on machine learning framework.

5.2 Preparing Data

For the simple test, cover images and stego-images were prepared that embedded the secret data with 3-bit least significant bits replacement. The secret data was generated by random function. Datasets can be prepared by combining images and the secret data in various ways.

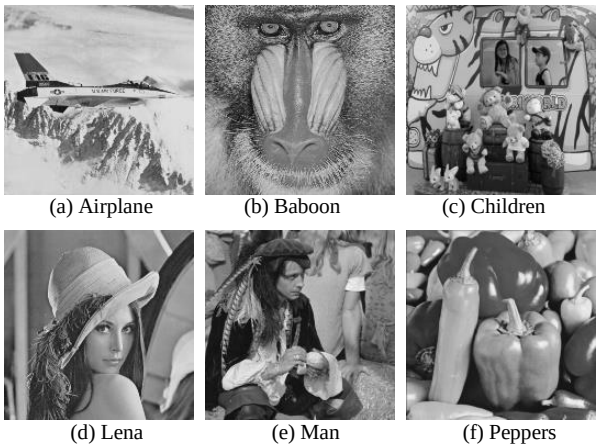


(a) Airplane (b) Baboon (c) Children

(d) Lena (e) Man (f) Peppers

Figure 6. Training over images.

The cover images in Figure 6 and stego-images in Figure 7 are used training datasets and test datasets randomly.



(a) Airplane (b) Baboon (c) Children

(d) Lena (e) Man (f) Peppers

Figure 7. Training stego-images.

For test images, the measurements embedding capacity and visual image quality PNSR, Q index on average for 3-bit least significant bits substitution are shown in Table. 1. Test images are converted into python data using Numpy array formats. Training and test datasets are prepared by embedding the secret data by random, and separated by scikit-learn library.

Table 1. Measurements of stego-images for 512x512 images

Images	Embedding capacity (bits)	PSNR (dB)	Q index
Airplane	786,432	35.75	0.7153
Baboon	786,432	35.70	0.9772
Children	786,432	35.67	0.8314
Lena	786,432	35.70	0.8320
Man	786,432	35.66	0.9167
Peppers	786,432	35.69	0.8668

5.3 Training Model

Keras and TensorFlow were used to construct models, to train models, and to evaluate the outputs. As shown in Figure 8, training process is consisted of convolution, pooling, convolution, pooling, and dense for input data to obtain outputs.

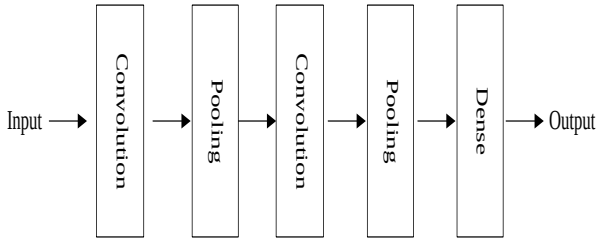


Figure 8. Training process.

Datasets are normalized before modeling, where value is set from 0 to 255 for each R, G, and B pixels.

5.4 Experimental Results

As machine learning process as described in Figure 5, pseudo code to implement is shown in Figure 9, where Keras framework was used to test stego-images.

```
Read images;
Normalize input images and covert into datasets;
Separate training and test data;
Build model;
  model.Sequential();
  model.add(Convolution2D() ...);
  model.add(Activation() ...);
  model.add(MaxPooling() ...);
  model.add(Dropout() ...);
  ...
  model.add(Flatten() ...);
  model.add(Dense() ...);
  model.add(Activation() ...);
  model.add(Dropout() ...);
  model.add(Dense() ...);
  model.add(Activation() ...);
  model.compile();
Train model;
  model.fit();
Evaluate output;
  model.evaluate();
```

Figure 9. Pseudo algorithm.

Accuracy of machine learning tests is not satisfactory, but it can be improved when many sample datasets are prepared.

Table 2. Results of machine learning

Loss	5.34341287612915
Accuracy	0.6666666865348816

In this paper, the process and possibility for various machine learning frameworks are considered, so the accuracy was not important. To increase the accuracy, many datasets have to be prepared and normalized.

6. CONCLUSION AND DISCUSSION

In this paper, various machine learning frameworks have been analyzed to show the possibility to steganography analysis. Prediction results and accuracy would be different depending on which framework was using and how many datasets were used. A

new model will be developed with efforts to improve accuracy by preparing various datasets to apply machine learning techniques to steganalysis.

7. ACKNOWLEDGMENTS

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(NRF-2018R1D1A1A09081842), Brain Pool program funded by the Ministry of Science and ICT through the National Research Foundation of Korea(No. 2018H1D3A2065993) and Ministry of Culture, Sports and Tourism(MCST) and from Korea Copyright Commission in 2018(2018-f_drm-9500).

8. REFERENCES

[1] Khan, A., Siddiqa, A., Munib, S., and Malik, S. A. 2014. A recent survey of reversible watermarking techniques, *Information Sciences* 279 (2014), 251-272.

[2] Subhedar, M. S. and Mankar, V.H. 2014. Current status and key issues in image steganography: a survey, *Computer Science Review* 13 (2014), 95-113.

[3] Nissar, A., Mir, A. H. (2010). Classification of steganalysis techniques: a study, *Digital Signal Processing* 20 (2010), 1758-1770.

[4] Cho, S., Cha, B. H., Gawecki, M., and Kuo, C. C. 2013. Block-based image steganalysis: algorithm and performance evaluation, J. Vis. Commun. Image R. 24 (2013), 846-856.

[5] Karampidis, K., Kavallieratour, E., and Papadourakis, G. 2018. A review of image steganalysis techniques for digital forensics. *Journal of Information Security and Applications* 40 (2018), 217-235.

[6] Musumeci, F. et al. 2018. An overview on application of machine learning techniques in optical networks. *Computer Science, Cornell University Library* (Oct. 2018), 1-27. https://arxiv.org/abs/1803.07976

[7] Lee, J. H., Shin, J., and Realff, M. J. 2018. Machine Learning: overview of the recent progresses and implications for the process systems engineering field. *Computer and Chemical Engineering* 114 (Oct. 2017), 111-121.

[8] Schmidhuber, J. 2015. Deep learning in neural networks: an overview. *Neural Networks* 61 (Oct. 2014), 85-117.

[9] Machine learning, https://en.wikipedia.org/

[10] Scikit-learn, https://scikit-learn.org/

[11] TensorFlow, https://www.tensorflow.org/

[12] Keras, https://keras.io

[13] Jung, K.H., 2016. A survey of reversible data hiding methods in dual images, *IETE Technical Review* 33 (2016), 441-452.

[14] Jung, K.H., 2018. A survey of interpolation-based reversible data hiding methods, Multimedia Tools and Applications 77 (2018), 7795-7810.

View publication stats