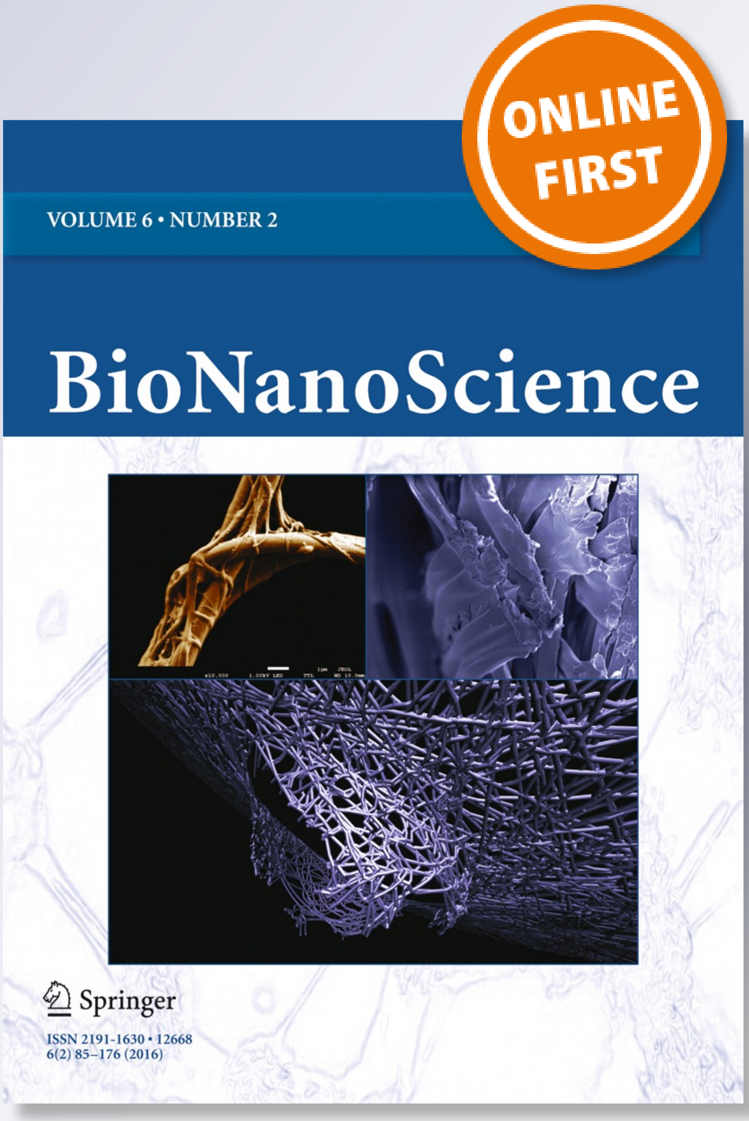
BioNanoScience

ISSN 2191-1630

BioNanoSci.
DOI 10.1007/s12668-016-0215-z



BioNanoSci.
DOI 10.1007/s12668-016-0215-z

Convolutional Neural Networks for Image Steganalysis

Dina Bashkirova[1]

Abstract Mathematical models based on human neuronal network behavior have recently become extremely popular and arouse interest as a solution of various computer vision problems. One of these models—Convolutional Neural Network—has been proven to be very efficient for object recognition problems and resembles principles of visual processing held by animal visual cortex. In this research, we propose a new approach to performing steganalysis on JPEG images using Convolutional Neural Networks. This approach allows to detect hidden embedding without computing features of an image predefined by empirical observations and obtain results comparable to state of the art methods of JPEG image steganalysis.

Keywords Convolutional neural network · Stegananlysis · Image processing

1 Introduction

The goal of passive steganalysis is to detect the hidden embedding of information in a digital object. Images, especially in JPEG domain, are one of the most popular type of data representation in the Internet, so creating a steganalytical tool that predicts the presence of hidden message in JPEG images with an appropriate accuracy is a highly pressing task. The most successful approach for solving this problem includes two main parts: building representative image model and applying a machine learning algorithm to perform classification on characteristics obtained from that image model. It is worth mentioning that the choice of image characteristics used for steganalysis almost entirely determines the quality of classification. The most advanced and efficient image models (e.g., JPEG Rich Model [1], CHEN [2], LIU [3], etc.) use statistics of values of DCT coefficients. In particular, JPEG Rich Model includes joint statistics over all DCT coefficients and integral joint statistics over the dependent coefficients in the image. However, using these empirically driven models may imply loss of information that is useful for detecting hidden embedding by a steganographic method.

In this paper, we propose a new approach to image steganalysis that will overcome the problem mentioned above by using Convolutional Neural Networks (CNN) [4] to obtain both representative and comprehensive image model and efficient classifier without precomputing statistical values of an image.

1.1 Problem Formulation

For the sake of simplicity, let us call the image that is not being changed by a staganographic algorithm as *cover* and the changed one as *stego*. Let us denote C indicates as a set of all possible digital objects, M indicates as a set of all possible messages that could be hidden in digital objects, and

✉ Dina Bashkirova
drbashkirova@yandex.ru

[1] Institute of Computer Mathematics and Information Technologies, KFU, 35 Kremlyovskaya, Kazan 420008, Russia

Published online: 02 August 2016

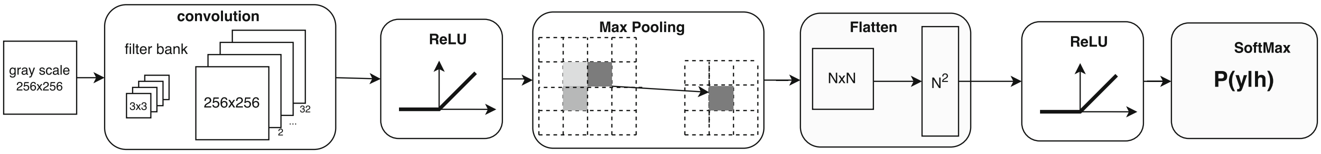


Fig. 1 Architecture of Convolutional Neural Network

K indicates a set of all possible steganographic keys. Then the problem of steganographic embedding can be written as follows:

$$Embed : C \times M \times K \to C.$$

Message retrieval is then described as

$$Extract : C \times K \to M,$$
$$Extract(Embed(c, k, m), k) = m.$$

Passive staganalysis problem is formally captured using the mapping:

$$Detect : C \to \{cover, stego\}. \quad (1)$$

In other words, mapping *Detect*(c) can be thought of as a binary classifier, thus it can be found using wide variety of machine learning algorithms.

As shown in [5], visual cortex consists of hierarchy of simple, complex, and hyper-complex cells forming a so-called visual field; simple cells detect low-level features, such as lines and edges, whereas complex cells receive impulses from simple cells and are able to detect more complicated patterns like shapes. This principle was taken as a basis for the architecture of CNN model, which has shown astonishing results in image processing and pattern recognition.

This model was applied in order to solve the problem of passive steganalysis, i.e., approximating the mapping 1. The main goal was to achieve high classification accuracy without predefining specific image characteristics. This was achieved by training a CNN to compute significant features using convolutional layers in contrast to the popular steganalytical approaches that are based on statistical image modal representation.

2 Materials and Methods

The experiments were run on two image databases: standard database BOSSbase 1.01 [8] with 10,000 images and realistic Microsoft Coco [9] database with 200,000 images. These images were converted to the unified format: 256 × 256 grayscale JPEG images with quality factor 0.75. Information embedding was performed using nsF5 [10] algorithm with 0.1 bit per non-zero AC coefficient(bpac).

In order to compare proposed method with the existing approaches in steganalysis, there was implemented a program tool for image classification using JPEG Rich Model and Random Forest [6] algorithm for classification as reported in [7].

Theano [11] and Keras [12] python libraries were used in order to define and train a convolutional neural network. The implemented CNN (see Fig. 1) consists of the following parts : two convolutional layers with 32 3×3 filters combined with nonlinear ReLU layers, 2×2 MaxPooling layer, 0.25 Dropout layer (drops out 25 % of randomly selected inputs), Dense (fully connected) layer, ReLU layer, 0.5 Dropout layer, Dense layer with two neurons and a SoftMax output unit.

Full 256 × 256 matrix of DCT-coefficients of an image was given as an input to CNN.

3 Results and Discussion

By training the described above convolutional neural network using given images, there was created a steganalytic classifier that established 98.7 % accuracy on test set, which is comparable to and even slightly higher than the result that was shown by classic image statistics based classifier. Figure 2 represents the ROC curve of the trained CNN on test set.

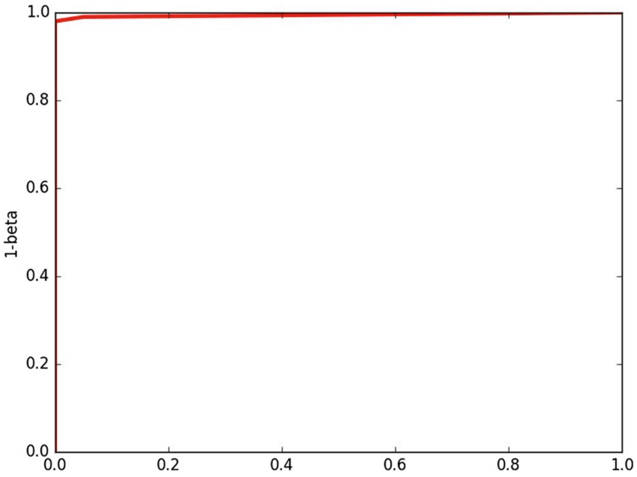


Fig. 2 ROC-curve of CNN applied to testing set

From those results, we conclude that deep learning approach can significantly simplify the process of creating a steganalytic tool with no compromise in its efficiency. Such promising result arouses interest in further research in this area. In particular, the problem of determining the most efficient architecture of convolutional network for image steganalysis needs further investigation.

4 Conclusion

It was shown that using convolutional neural networks for digital image steganalysis can both eliminate the need to select a complex image model and allow to obtain high accuracy in prediction of hidden embedding.

References

1. Kodovskỳ, J., & Fridrich, J. (2012). Steganalysis of JPEG images using rich models. *Media Watermarking Security, and Forensics*, 8303.
2. Chen, C., & Shi, Y.Q. (2008). JPEG image steganalysis utilizing both intrablock and interblock correlations. In *IEEE International Symposium on Circuits and Systems, IEEE, 2008*.
3. Qingzhong, L. (2011). Steganalysis of DCT-embedding based adaptive steganography and YASS. In *Proceedings of the thirteenth ACM multimedia workshop on Multimedia and security, ACM*.
4. Krizhevsky, A., Sutskever, I., & Hinton, G.E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*.
5. Hubel, D.H., & Wiesel, T.N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, *160.1*, 106–154.
6. Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news 2.3*, 18–22.
7. Vojtech, H., & Fridrich, J. (2014). Challenging the doctrines of JPEG steganography, IST/SPIE Electronic Imaging. International Society for Optics and Photonics.
8. Bas, P., Filler, T., & Pevn, T. (2011). *Break Our Steganographic System: The Ins and Outs of Organizing BOSS. International Workshop on Information Hiding*: Springer Berlin Heidelberg.
9. Lin, T.-Y., et al. (2014). Microsoft coco: Common objects in context. In *European Conference on Computer Vision. Springer International Publishing*.
10. Fridrich, J., Pevn, T., & Kodovsk, J. (2007). Statistically undetectable jpeg steganography: dead ends challenges, and opportunities. In *Proceedings of the 9th workshop on Multimedia & security. ACM*.
11. Team, The Theano Development, & et al (2016). Theano: A Python framework for fast computation of mathematical expressions. arXiv preprint. arXiv:1605.02688.
12. Franois, C. (2015). Keras: Deep learning library for theano and tensorflow.