ALASKA2: IMAGE STEGANALYSIS COMPETITION. A FINAL PROJECT PROPOSAL

Ilia Ilmer
Graduate Center CUNY
Department of Computer Science
iilmer@gradcenter.cuny.edu

May 16, 2020

ABSTRACT

In May of 2020, Troy University of Technology organized a competition on Kaggle.com titled "ALASKA2: Detect secret data hidden within digital images". The goal of the competition is to predict presence of hidden data in a particular image. In this project, we will identify reliable and promising convolutional neural network based models that will help me in solving the task. I will define the task as a multi-class classification problem to simplify the approach. In the process of solving the problem we will apply various deep learning training techniques in order to maximize the score.

1 Introduction

Steganalysis is a scientific discipline that studies various forms of data in order to determine whether or not a secret message is concealed in that data. It allows communication through means that are indistinguishable from regular exchanges of information. Researchers working on discovering hidden information through steganalysis use setups that do not necessarily resemble real-life occurrences. The aspects of message encoding such as methods and parameters used are often known in advance, the data is clean, taken from a single resource using the same camera settings for all images.

In real life, message decoding is a much harder task to solve. This competition aims to simulate a realistic scenario in which 3 types of encoding are used. A dataset 75,000 unique 512 by 512 color images have been used as a cover for secret messages. The data were obtained from different sources with different quality settings.

There are three unique ways the message is encoded into the image for us, JMiPOD [1], JUNIWARD [4], and UERD [2]. Each method is applied to all 75,000 original covers creating in total 225,000 images that have encodings of different types. Encoding is done using the JPEG compression algorithm. More specifically, during the JPEG compression, the message is encoded through DCT coefficients and is hidden from the viewer.

The JUNIWARD algorithm is exploiting the linearity of discrete cosine transform and linear transformation of Gaussian random variables. In order to minimize detectability, the algorithm solves an optimization problem with respect to pixel values obtained from inverse DCT.JMiPOD method utilizes the approach of bitwise encoding. The image is converted to binary codes and the message is encoded into these codes bitwise. Finally, UERD also utilizes DCT coefficients during encoding.

The following report is organized as follows. We formulate a machine learning problem presented to us as a multi-class classification task in section 2. In section 3 we perform simple data exploration and analysis. In section 4, we describe the model selection algorithm and hyperparameter tuning as well as the training setup and tricks. In that same section we also report the results of the training and the leaderboard score obtained so far. Finally, we list further ways of improvement in the final section.

2 Task Formulation

Let us outline the initial idea for the project and possible improvements. As the competition deals with images, we should take advantage of reliable neural network architectures available as parts of PyTorch or Tensorflow packages. Advanced architectures can to catch intricate differences between an image with one encoding or another. I would like to pose the problem as a 4-class classification task in the following manner.

Consider a dataset of 300,000 images of which there are 4 classes

$$\begin{cases} 0 - \text{cover}, \\ 1 - \text{jmipod}, \\ 2 - \text{juniward}, \\ 3 - \text{uerd}. \end{cases}$$

The classes arise naturally in this setting and, furthermore, they are perfectly balanced and we need not worry about imbalanced data when sampling. We will then train a neural network $\mathcal{F}$ on this collection of images $\mathcal{X} = \{\mathrm{x}_k : k = 0, 1, 2, 3\}$ to maximize the key metric of the competition: area under ROC curve with weights. As per the competition requirements, the submission's true positive rate values between 0 and 0.4 have weight 2 while the rest carry weight 1.

To summarize, we must maximize the weighted-AUC metric on a 4-class classification problem in this setup for the competition.

3 Data Exploration

Let us discuss the data in question. We are studying 300,000 RGB images of size 512 by 512 pixels. For 75,000 of those images no hidden message is encoded, on the other three images, the encoding corresponds to one of three types, JMiPOD [1], JUNIWARD [4], or UERD [2]. Each encoding utilizes the JPEG compression process and encodes the information into the DCT coefficients. This allows the message to remain hidden so much so that the difference is not reflected in the spatial domain.

In the competition description [6], the organizers provide some basic information about the data. We know that all steganalysis algorithms are applied with the same probability and that the average message length is 0.4 bits per DCT coefficient.

3.1 Examples of data

4 Model Selection

4.1 Deep Neural Networks

The first model we consider is the EfficientNet [5] model which is most widely used by the competitors. This network beats state of the art results in classification and was created through an automatic architecture search algorithm. There are several versions of the network with increasing complexity and computational requirements. In this work we will experiment with a pre-trained EfficientNet-B0 [5]. The network's implementation is supplied as a PyTorch based Python package available for install. The implementation provided also utilizes Swish activation function. This function was first introduced in [3] where authors considered multiple candidate-functions. Swish activation function is defined as follows

$$Swish(x) = x \cdot \sigma(x),$$

where σ is the sigmoid function.

4.2 Hyperparameter choosing

4.3 Training setup

1. *Gradient accumulation:*
2. *YCrCb color model:*
3. *Network*

4.3.1 Metric

5 Results

6 Conclusion

References

[1] Rmi Cogranne, Quentin Giboulot, and Patrick Bas. “Steganography by Minimizing Statistical Detectability: The cases of JPEG and Color Images.” In: *ACM Information Hiding and MultiMedia Security (IH&MMSec)*. 2020.

[2] L. Guo et al. “Using Statistical Image Model for JPEG Steganography: Uniform Embedding Revisited”. In: *IEEE Transactions on Information Forensics and Security* 10.12 (2015), pp. 2669–2680.

[3] Prajit Ramachandran, Barret Zoph, and Quoc V Le. “Searching for activation functions”. In: *arXiv preprint arXiv:1710.05941* (2017).

[4] X. Song et al. “Steganalysis of adaptive JPEG steganography using 2D Gabor filters”. In: *Proceedings of the 3rd ACM workshop on information hiding and multimedia security*. 2015, pp. 15–23.

[5] Mingxing Tan and Quoc V Le. “Efficientnet: Rethinking model scaling for convolutional neural networks”. In: *arXiv preprint arXiv:1905.11946* (2019).

[6] Troyes University of Technology. *ALASKA2 Image Steganalysis*. https://www.kaggle.com/c/alaska2-image-steganalysis/data. 2020.