Project Defining Seminar

Synopsis

on

"DIGITAL WATERMARKING AND CAPTION GENERATOR APPLICATION USING DEEP NEURAL NETWORK"

Project Guide

Archana Dehankar

Projectees (7th Sem Sec 'A & B')

1. Tulsi Mundada (Roll No. 126)
2. Sakshi Kubde (Roll No. 210)
3. Pranali Kadukar (Roll No. 115)
4. Khushi Sahu (Roll No. 126)
5. Dnyananda Ittadwar (Roll No.204)

Priyadarshini College of Engineering Nagpur
Department of Computer Technology
Session 2022-23

INDEX

ABSTRACT

Data integrity and source origin authentication are essential topics for real-time multimedia systems. But the traditional methods are not very applicable to overcome the distortion introduced in multimedia data transportation. In this thesis some security mechanics are proposed, which rely on authentication rather than on encryption methods. The highly asymmetric architectures found in ubiquitous computing applications are exploited to provide a protection of the transmitted multimedia data by means of well-known digital watermarking techniques. Digital image watermarking is the process of embedding and extracting watermark covertly on a carrier image. Incorporating deep learning networks with image watermarking has attracted increasing attention during recent years. However, existing deep learning-based watermarking systems cannot achieve robustness, blindness, and automated embedding and extraction simultaneously. In this project, a fully automated image watermarking system based on deep neural networks is proposed to generalize the image watermarking processes. This project also proposes an image caption generator that will accept an image as an input and generate an English sentence as output by labelling the image's content using two optimization techniques such as beam search and greedy search.

Keywords: Image watermarking, robustness, deep learning, convolutional neural networks, caption generator.

INTRODUCTION

2.1 Introduction to Digital Watermarking:

A **digital watermark** is a kind of marker covertly embedded in a noise-tolerant signal such as audio, video or image data. It is typically used to identify ownership of the copyright of such signal. "Watermarking" is the process of hiding digital information in a carrier signal; the hidden information should, but does not need to, contain a relation to the carrier signal. Digital watermarks may be used to verify the authenticity or integrity of the carrier signal or to show the identity of its owners. It is prominently used for tracing copyright infringements and for banknote authentication.

Like traditional physical watermarks, digital watermarks are often only perceptible under certain conditions, e.g. after using some algorithm. [2] If a digital watermark distorts the carrier signal in a way that it becomes easily perceivable, it may be considered less effective depending on its purpose.[2] Traditional watermarks may be applied to visible media (like images or video), whereas in digital watermarking, the signal may be audio, pictures, video, texts or 3D models. A signal may carry several different watermarks at the same time. Unlike metadata that is added to the carrier signal, a digital watermark does not change the size of the carrier signal.

The needed properties of a digital watermark depend on the use case in which it is applied. For marking media files with copyright information, a digital watermark has to be rather robust against modifications that can be applied to the carrier signal. Instead, if integrity has to be ensured, a fragile watermark would be applied.

Both steganography and digital watermarking employ steganography techniques to embed data covertly in noisy signals. While steganography aims for imperceptibility to human senses, digital watermarking tries to control the robustness as top priority.

Since a digital copy of data is the same as the original, digital watermarking is a passive protection tool. It just marks data, but does not degrade it or control access to the data. One application of digital watermarking is source tracking. A watermark is embedded into a digital signal at each point of distribution. If a copy of the work is found later, then the watermark may be retrieved from the copy and the source of the distribution is known. This technique reportedly has been used to detect the source of illegally copied movies.

2.2 Introduction Caption Generation:

Automatically describing the content of images using natural language is a fundamental and challenging task. With the advancement in computing power along with the availability of huge datasets, building models that can generate captions [3] for an image has become possible. On the other hand, humans are able to easily describe the environments they are in. Given a picture, it's natural for a person to explain an immense amount of details about this image with a fast glance. Although great development has been made in computer vision, tasks such as recognizing an object, action classification, image classification, attribute classification and scene recognition are possible but it is a relatively new task to let a computer describe an image that is forwarded to it in the form of a human-like sentence. For this goal of image captioning, based on semantics of images should be captured here and expressed in the desired form of natural languages. It has a great impact in the real world, for instance by helping visually impaired people better understand the content of images on the web.

So, to make our image caption generator model, we will be merging CNN-RNN architectures. Feature extraction from images is done using CNN. We have used the pre-trained model Exception. The information received from CNN is then used by LSTM for generating a description of the image. However, sentences that are generated using these approaches are usually generic descriptions of the visual content and background information is ignored. Such generic descriptions do not satisfy in emergent situations as they, essentially replicate the information present in the images and detailed descriptions regarding events and entities present in the images are not provided, which is imperative to understanding emergent situations.

LITERATURE SURVEY

1. **Uchida et al. [1]** previously introduced the first approach for embedding watermarks into a deep neural network through which he embedded the data into the weights of DNN. Hence, it believes that the stolen prototypes can be nearby accessible to obtain the group of parameters, which is not possible since most deep learning models are used as online assistance and it will be troublesome to immediately procure access to model parameters, particularly for the stolen prototypes. As DNN models are broadly deployed and become more critical, they are frequently targeted by enemies. Enemies can steal the model and build a copy of AI service.

2. **Rouhani et al. [2]** presented a general watermarking framework that works in both the white box and the black box environment. The embedded bit watermark is embedded in the activation cards of the selected layer (s) of the target DNN by integrating two additional regularization loss terms, binary cross-entropy loss, and GMM agent loss. To activate cards, a deep Gaussian Mixed Model (GMM) is used. During DNN training, WM-specific regularization loss conditions are integrated to match the activations and encode the WM information. Later in the extraction step, the WM key frames are transferred to the model to trigger a marked activation. The BER is calculated and the result of these activation watermarks is restored. This has an advantage over the Uchida [6] because it gives both data and model-dependent. At the same time, it is robust against attacks with fine-tuning, overwriting and parameter setting. Since it is both data and model-dependent but more computation.

3. **Merrer et al. [3]** suggest creating controversial patterns ifthe opponent succeeded in the attack, the correspondingsamples are called "true opponents”. In this scheme, the set ofopposite samples is used as the WM key set to change thedecision boundary of the target neural network. While attacksfail, images l watermark key is a complete combination oftrue and false opponents. The model is queried in thedetection phase of watermark and the null hypothesis is tested,this is done via the Binary distribution. If the watermark keyis less than the threshold, the watermark should be present inthe model. It is practical from the point of view of

detectionand embedding, but the accuracy can be lowered afterembedding watermarks, achieves a very high false-positiverate and is not very robust against attacks.

4. Images incorrectlyclassified by the model are suggested by **Yossi et al. [4]** these misclassified images known as backdoor images andrandom labels as a watermark key that is set to embed thewatermark with a zero bit that is DNN compliant. Thecomparable basis of the key label is randomly selected fromall classes except the real label and the predictable realsample. First, the threshold is set by the binomial distributionand then compared to the watermark trigger set to detect thewatermark. This scheme provides better accuracy and a highdetection rate but is sensitive to water removal attacks.

5. **Zhanget al. [5]** introduced the three different key generationschemes content-based, noise-based and unrelated-basedimages respectively. Pre-train model is accurately adjustedwith watermark keys and in the detection phase, the usersends the watermark key to the DNN model. This is theexternal service provider that ultimately submits thethreshold to classify the Boolean decision. He introducedthree different methods for generating a watermark key, butnot well for giving a continuous performance on the standarddatasets.

6. **Rouhani et al. [6]** in the goal, DNN investigatedthe rarely inhabited space to regard the random images andlabels as a watermark set. The watermark key used foridentifying the key pairs and for Watermark embeddingfine-tuning the target DNN model. The final WM key set isthe intersection of the keys that are correctly predicted by thehighlighted model and incorrectly predicted by the unmarkedmodel. In the multinomial distribution of the detection phasefor the output prediction and does the examination of theimaginary hypothesis with the definitive Watermark key set.It gives a high detection rate with a low percentage of falsealarms but the generation of keys generates higher overheadcosts. Look at fake enemies that preserve the model'saccuracy for specific data.

7. **S. Albawi et al. [7]** in this paper one of the most popular deep neural networks is the Convolutional Neural Network (CNN) is explained. There are multiple layersin

CNN; such as convolutional layer, & nonlinearity layer, &pooling layer and fully-connected layer as well. The CNNhas an excellent performance in machine learning problems and one of the most common algorithms.

8. **S. Hochreiter et al. [8]** in this paper Sepp Hochreiter explain about the deep neural network algorithm long short team Memory (LSTM). LSTM is local in both space aswell as in time; the computational complexity is per time of step and also the weight pattern representation. In comparison to other algorithm LSTM leads to many more successful runs, and learn much faster. It's even solve complex, artificial long time lag tasks that have never been solved by previous recurrent network

9. **O. Vinyals et al. [9]** the fundamental problem in artificial intelligence that connects computer vision and Natural language processing is automatically describing the content of an image. In this paper, A.L systematically analyze a deep neural networks based image caption generation method. Here an image is given as the input, and the method as outputin the form of sentence in English describing the content of the image. They analyze three components of the method: convolutional neural network (CNN), recurrent neural network (RNN) and sentence generation. This model analyze image and generate more trial and relevant words for images.

10. **D. S. Whitehead t al. [10]** current image captioning approaches generate descriptions which lack specific information, such as named entities that are involved in the images. HereDi Lu, Spencer Whitehead had proposed a very new task which generates descriptive image captions, given images as input. A simple solution to this problem that we are proposing is that we will train a CNN-LSTM model so that it can generate a caption based on the image.

OBJECTIVES

The objective behind developing an automatic digital watermark generator is to provide facility to the users that they can encrypt their documentations or media files and send them to any other person without fear of hacking.

The objective of our project is to develop a web based interface for users to get the description of the image and to make a classification system in order to differentiate images as per their description. It can also make the task of SEO easier which is complicated as they have to maintain and explore enormous amounts of data.

PROJECT METHODOLOGY

5.1 Digital Watermarking:

Digital watermarking is being used in numerous applications. Most of the current applications are devoted to copyright protection.

It has the following purposes in general:

(1) Covert Communications: These are mainly applications of steganography. In military and intelligence applications, people would like to send messages to each other without being detected.

(2) Authentication: Sometimes it is necessary to verify the authenticity of input data, i.e., to determine whether the data are original, fake, or the altered version of the original. For authentication purposes, fragile watermarks seem to be a good solution; a properly designed fragile watermarking algorithm should be able to detect any alterations.

(3) Identification of Ownership: Robust watermarking algorithms are developed to identify the ownership of digital media. In this kind of applications, a movie producer selling its products in digital formats is subject to copyright piracy. In such situations, original producers would like to have legally proof that they are the real owners. A well-designed robust watermarking scheme is a possible solution to these cases. Additionally, to prevent the unauthorized users from playing the digital products, some media player manufacturers consider adding the watermarking detection facilities in their products. In such a scenario, media players would play the input clips only if they successfully detect the watermark of the company and confirm the authorization of the consumers. This is similar to DRM (short for digital right management) solutions except that watermarking is more resistant to media processing.

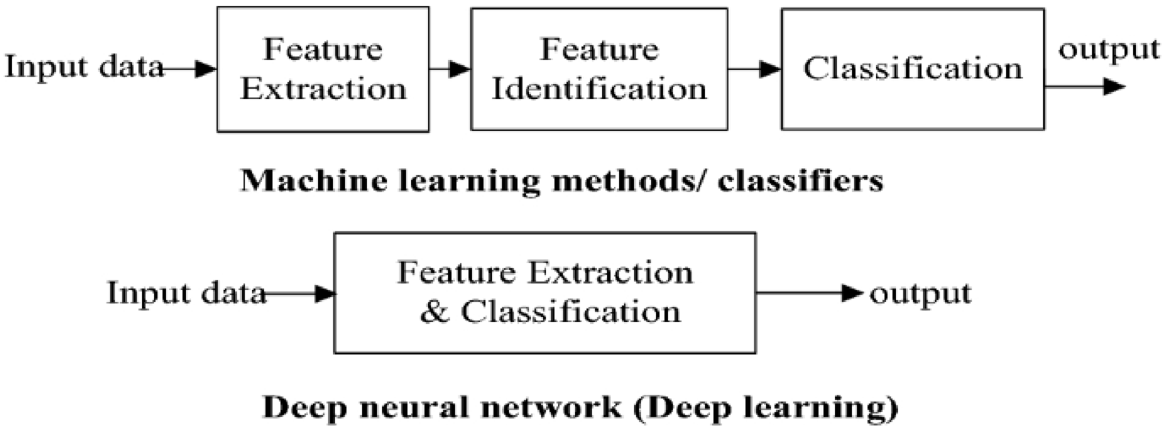

Input data
Feature Extraction
Feature Identification
Classification
output
Machine learning methods/ classifiers
Input data
Feature Extraction & Classification
output
Deep neural network (Deep learning)

Figure 5.1.1 Proposed Model of Digital Watermarking

5.2 Caption Generator:

The caption generator's main aim is to generate an English sentence or caption for an input automatically. The main challenge of the system is to generate correct captions for the given input. This project proposes a caption generator that will accept an input and generate an English sentence as output by labelling the content using two optimization techniques such as beam search and greedy search. The system takes the pre-trained deep learning convolutional neural network (CNN) architecture VGG16 model for learning the input features, uses long short-term memory (LSTM) for learning the text features, and combines the content's result with an LSTM to generate a caption for the input. We use the LSTM model to generate text or sentences or captions for the given input content. This model was developed to build a caption generator by implementing the convolutional neural network with long short-term memory. The pre-trained VGG16 is used to extract features from the given image. LSTM works as a decoder to generate sentences or captions for the content. This model is trained so that if the input image is given to the model, it will generate captions or sentences describing the input content.

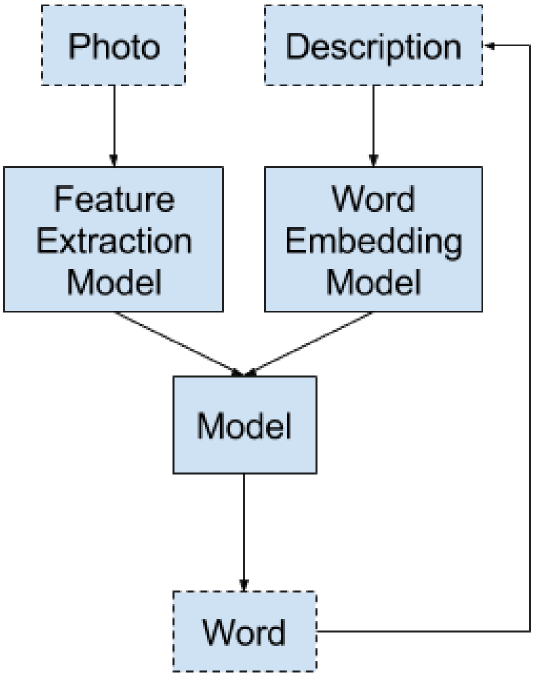


Figure 5.2: Proposed Model of Caption Generator

The proposed model of Image Caption Generator is as shown in the above figure 1.Here in this model, input image is given & then a convolutional neural network is used to create a dense feature vector as shown in figure. This dense vector, also called an embedding, this vector can be used as input into other algorithms, and it generates suitable caption for given image as output. For an image caption generator, this embedding becomes a representation of the image and used as the initial state of the LSTM for generating meaningful captions, for the image.

5.2.1 Implementation of the System:

Here we will discuss the implementation of the system.

5.2.1.1. Object Detection

Objects are detected from the image with the help of CNN Encoder.

5.2.1.2 Sentence Generation

By using LSTM, sentences are generated. Each predicted word is employed to get subsequent words. Using these words, appropriate sentence is formed with the help of optimal beam search. Here, Softmax function will be used for prediction of word.

5.2.1.3 Deployment

The final project will be deployed using Tkinter which is Python based GUI. It is the standard Python Interface for developing GUI’s.

5.3 Algorithms Used:

Convolutional Neural Network CNN:

CNN is a subfield of Deep learning and specialized deep neural networks used for the recognition and classification of images. It is used to process the data represented as a 2D matrix like images. It can deal with scaled, translated, and rotated imagery. It analyzes the visual imagery by scanning them from left to right and top to bottom and extracting relevant features from that. Finally, it combines all the features for image classification.

Long short-term memory (LSTM):

Being a type of RNN (recurrent neural network), LSTM (Long short-term memory) is capable of working with sequence prediction problems. It is mostly used for the next word prediction purposes, as in Google search our system is showing the next word based on the previous text. Throughout the processing of inputs, LSTM is used to carry out the relevant information and to discard non-relevant information. To build an image caption generator model we have to merge CNN with LSTM. We can drive that:

Image Caption Generator Model (CNN-RNN model) = CNN + LSTM.

- CNN- To extract features from the image. A pre-trained model called Xception is used for this.
- LSTM- To generate a description from the extracted information of the image.

5.4 Pre – requisites:

- Python:

Python is commonly used for developing websites and software, task automation, data analysis, and data visualization. Since it's relatively easy to learn, Python has been adopted by many non-programmers such as accountants and scientists, for a variety of everyday tasks, like organizing finances.

- Deep Learning:

Deep learning is a subset of machine learning, which is essentially a neural network with three or more layers. These neural networks attempt to simulate the behavior of the human brain—albeit far from matching its ability—allowing it to “learn” from large amounts of data.

- Natural Language Processing (NLP):

NLP algorithms are typically based on machine learning algorithms. Instead of hand-coding large sets of rules, NLP can rely on machine learning to automatically learn these rules by analyzing a set of examples (i.e. a large corpus, like a book, down to a collection of sentences), and making a statistical inference.

EXPECTED OUTCOMES

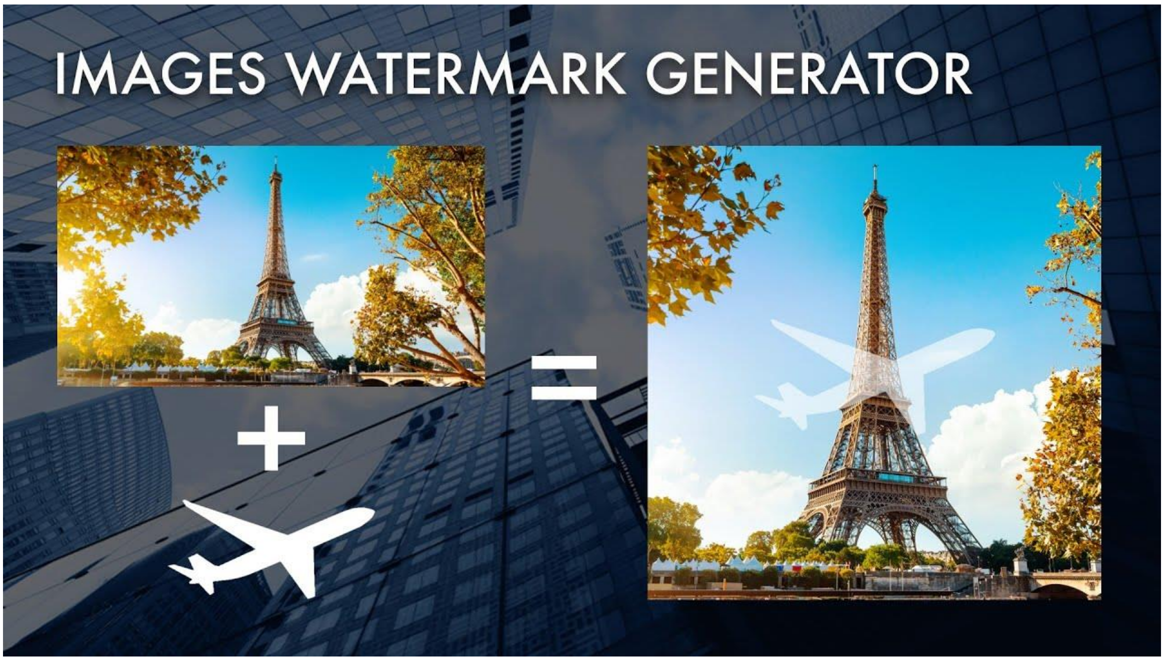
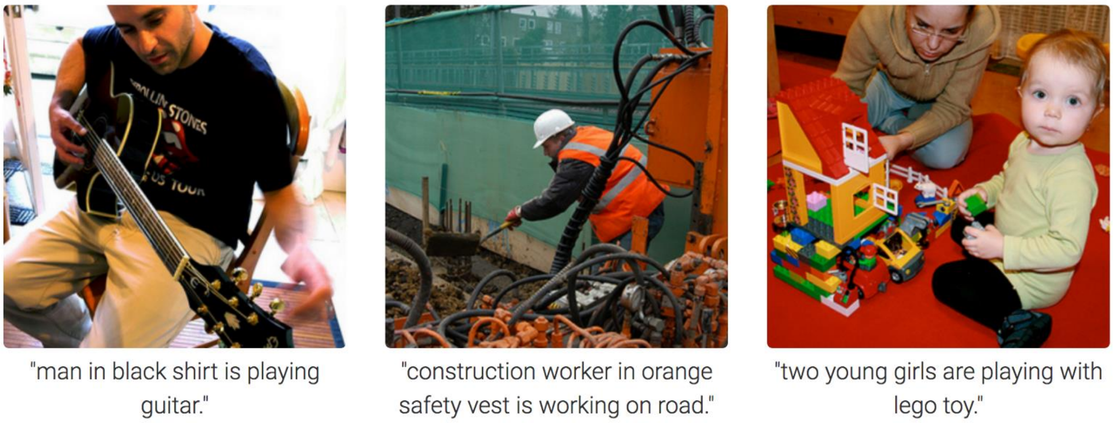


Figure 6.1 Digital Watermark Generator


"man in black shirt is playing guitar."

"construction worker in orange safety vest is working on road."

"two young girls are playing with lego toy."

Figure 6.2 Image Caption Generator

REFERNCES

1. Y. Uchida, Y. Nagai, S. Sakazawa, and S. Satoh, “Embeddingwatermarks into deep neural networks,” in Proc. the 2017 ACM onInternational Conference on Multimedia Retrieval, pp. 269-277.

2. Theano Development Team, “A python framework for fast computation of mathematical expressions,” 2016.arXiv e-prints, abs/1605.02688

3. E. L. Merrer, P. Perez, and G. Tredan, “Adversarial frontier stitching for ´remote neural network watermarking,” arXiv preprint, arXiv:1711.01894, 2017

4. N. Aloysius and M. Geetha, “A review on deep convolutional neural networks,” in Proc. 2017 International Conference on Communicationand Signal Processing (ICCSP), Chennai, 2017, pp. 0588-0592.

5. J. Zhang, Z. Gu, J. Jang, H. Wu, M. P. Stoecklin, H. Huang, and I. Molloy, “Protecting intellectual property of deep neural networks with watermarking,” in Proc. the 2018 on Asia Conference on Computer and Communications Security, pp. 159–172.

6. B. D. Rouhani, H. Chen, and F. Koushanfar, “Deepsigns: A generic watermarking framework for IP protection of deep learning models,” arXiv preprint, arXiv:1804.00750, 2018.

7. S. ALBAWI and T. A. MOHAMMED, "Understanding of a Convolutional Neural Network,"in ICET, Antalya, 2017

8. S. Hochreiter, "LONG SHORT-TERM MEMORY," Neural Computation, December 2019

9. O. Vinyals, A. Toshev, S. Bengio and D. Erhan,"A Neural Image Caption Generator," CVPR 2015 Open Access Repository, vol. Xiv, 17 November 2020.

10. D. S. Whitehead, L. Huang, H. and S.-F. Chang, "Entityaware Image CaptionGeneration,"inEmpirical Methods in Natural Language Processing, Brussels, 2018.