A deep learning approach to patch-based image inpainting forensics

Xinshan Zhu, Yongjun Qian, Xianfeng Zhao, Biao Sun, Ya Sun

PII: S0923-5965(18)30534-4
DOI: https://doi.org/10.1016/j.image.2018.05.015
Reference: IMAGE 15395

To appear in: *Signal Processing: Image Communication*

Received date : 6 November 2017
Revised date : 11 April 2018
Accepted date : 28 May 2018

Please cite this article as: X. Zhu, Y. Qian, X. Zhao, B. Sun, Y. Sun, A deep learning approach to patch-based image inpainting forensics, *Signal Processing: Image Communication* (2018), https://doi.org/10.1016/j.image.2018.05.015

A Deep Learning Approach to Patch-based Image Inpainting Forensics

Xinshan Zhu[a,b], Yongjun Qian[a], Xianfeng Zhao[b], Biao Sun[a,*], Ya Sun[a]

[a] *School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China*
[b] *State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China*
Phone: +86-15122580213

Abstract

Although image inpainting is now an effective image editing technique, limited work has been done for inpainting forensics. The main drawbacks of the conventional inpainting forensics methods lie in the difficulties on inpainting feature extraction and the very high computational cost. In this paper, we propose a novel approach based on a convolutional neural network (CNN) to detect patch-based inpainting operation. Specifically, the CNN is built following the encoder-decoder network structure, which allows us to predict the inpainting probability for each pixel in an image. To guide the CNN to automatically learn the inpainting features, a label matrix is generated for the CNN training by assigning a class label for each pixel of an image, and the designed weighted cross-entropy serves as the loss function. They further help to strongly supervise the CNN to capture the manipulation information rather than the image content features. By the established CNN, inpainting forensics does not need to consider feature extraction and classifier design, and use any postprocessing as in conventional forensics methods. They are combined into the unique framework and optimized simultaneously. Experimental results show that the proposed method achieves superior performance in terms of true positive rate, false positive rate and the running time, as compared with state-of-the-art methods for inpainting forensics, and is very robust against JPEG compression and scaling manipulations.

Keywords: Inpainting, forensics, convolutional neural network, loss function, compression

1. Introduction

With the rapid development of computer and Internet technologies, digital images have been widely used in many aspects of social life, such as news, advertisement, publications, etc.. However, due to the availability of powerful image processing software, the images can be accessed, copied and edited more easily than ever before. The reliability and authenticity of image information have been a main concern in the digital multimedia era. As such a promising technique, digital image forensics has attracted considerable attention and seen numerous applications recently. The basic idea is to extract the unique artifacts introduced by an image operation to decide whether an image once underwent such an operation. A great effort has been spent by researchers in developing various forensics techniques on JPEG compression [1, 2], histogram equalization [3], median filtering [4], resizing [5], copy move forgery [6, 7] and so on [8].

As a new and effective image editing technique, image inpainting aims to effectively repair damaged or removed image regions in a visually plausible manner using the known image information. It is an important topic in the field of computer vision and image processing, and has made great progress in the past few years. Applications of image inpainting include image restoration (e.g., scratch or text removal), image coding and transmission (recovery of missing blocks), photo-editing (object removal), virtual restoration of digitized paintings (crack removal), etc. [9].

Figure 1: An illustrative example for object removal by image inpainting: original image (left) and inpainted image (right).

However, inpainting could be also exploited to modify image content with malicious motives, which brings the crisis of confidence in image content. Consequently, the inpainting forensics, i.e., the detection of inpainting operation, is proposed to deal with this problem. In a practical scenario for inpainting forensics, a host image is first altered by a forger with the inpainting operation. For instance, in Fig. 1, an image is modified by removing an object in it using inpainting operation. Then, the inpainted image could undergo other image processing operations before forensics, e.g., JPEG compression, which inevitably cause the inpainting information loss. Last, an investigator receives the possibly distorted inpainted image, and detects the existence of inpainting by forensic techniques. This process is depicted in Fig. 2. Generally, there are two major forensic tasks to be considered: is the input image inpainted by a forger, and if yes, where is the inpainted region? By the inpainting region information, the investigator knows exactly

*Corresponding author.
Email address: {sunbiao,xszhu}@tju.edu.cn (Ya Sun)