IMAGINE Dataset: Digital Camera Identification Image Benchmarking Dataset

Jarosław Bernacki, Rafał Scherer

Czestochowa University of Technology, Czestochowa, Poland

Introduction

- Digital camera identification based on images:
  - Popular research field in digital forensics, many articles have been published in the last 15 years;
  - A huge number of imaging devices allowing for capturing images in an easy way;
  - Many algorithms for camera identification, but not so many image datasets to benchmark them.

Contribution

- We propose the IMAGINE dataset for benchmarking camera identification algorithms which includes a number of images coming from modern imaging devices;
- We benchmark proposed dataset with a set of modern camera identification methods and experimentally show that the IMAGINE dataset allows for a reliable testing of such methods.

Existing Datasets

- Dresden Image Database (2010):
  - A great dataset including tens of images coming from a number of cameras, but equipped with charge-coupled device (CCD) imaging sensors which now are obsolete (replaced by CMOS sensors);
- MICHE (2015), VISION (2017), UNIFI (2018):
  - Small datasets containing not representative number of devices;
- Social media (Flickr, 500px, etc):
  - Hard to get a representative number of images from the same camera; images are often manipulated.

IMAGINE Dataset Description

- Available online:

  `https://kisi.pcz.pl/imagine`
- 2500 images, 55 devices (digital single lens reflex/mirrorless, compact cameras, smartphones/tablets, drones);
- JPG images coming directly from cameras – not edited in any software; cameras set to default shooting mode with default white balance;
- Different imaging sensor sizes: from "full frame" ($\mathbf{36} \times \mathbf{24}$ mm) to small sensors ($\mathbf{3.2} \times \mathbf{2.4}$ mm) utilized in mobile devices.

Utilized Devices

The IMAGINE dataset includes images coming from the following devices (please note that this is not a complete list):

- Canon EOS: 1D X Mark II, 5D Mark IV, 6D Mark II, M3, M5, 90D, R, R5, R6, RP;
- Nikon: D5, D6, D500, D610, D750, D810, D850, D3100, Z6, Z6 II, Z7, Z7 II;
- Sony: A1, A7R III, A7S, A9;
- Many other devices (including mobile) by Apple, DJI, Fujifilm, LG, Nokia, Samsung, Yuneec.

Experimental evaluation – setup

- We experimentally check the classification accuracy for the following methods: an algorithm by Lukás (2006)* and CNNs presented by Bondi, Tuama, Mandelli, Kirchner & Johnson (2016-2020) with the use of the IMAGINE dataset;
- We also analyze results on CNNs training accuracy with IMAGINE and Dresden Image Database (further named shortly: Dresden) datasets.

* Although presented in 2006, this algorithm is still considered as the most efficient.

Experimental evaluation (1/4)

- We use the standard accuracy (ACC) measure:

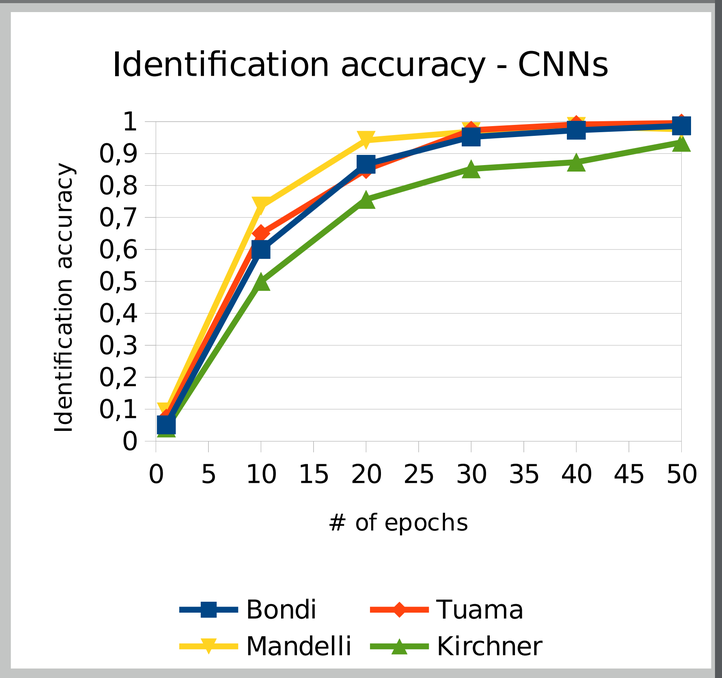
$$\mathbf{ACC} = \frac{\mathbf{TP} + \mathbf{TN}}{\mathbf{TP} + \mathbf{FP} + \mathbf{FN} + \mathbf{TN}}$$

  - TP/TN stands for true positive/true negative;
  - FP/FN stands for false positive/false negative.

Experimental evaluation (2/4)

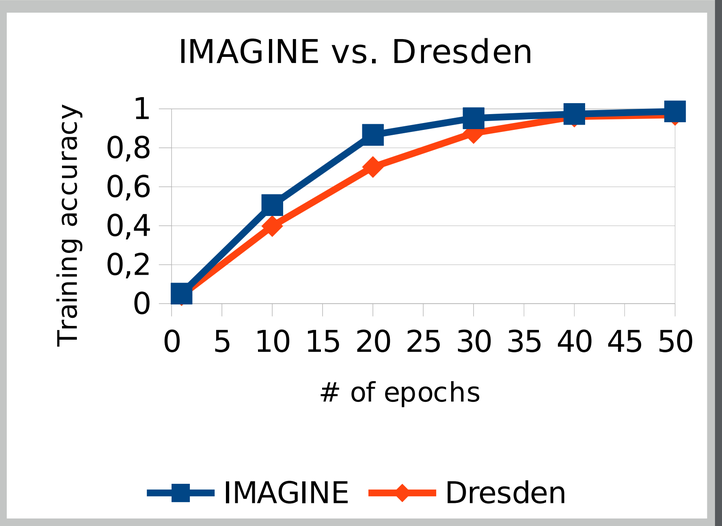
Identification accuracy for 50 epochs:

- Comparable for nearly all tested CNNs (only Kirchner & Johnson achieved a bit lower results).



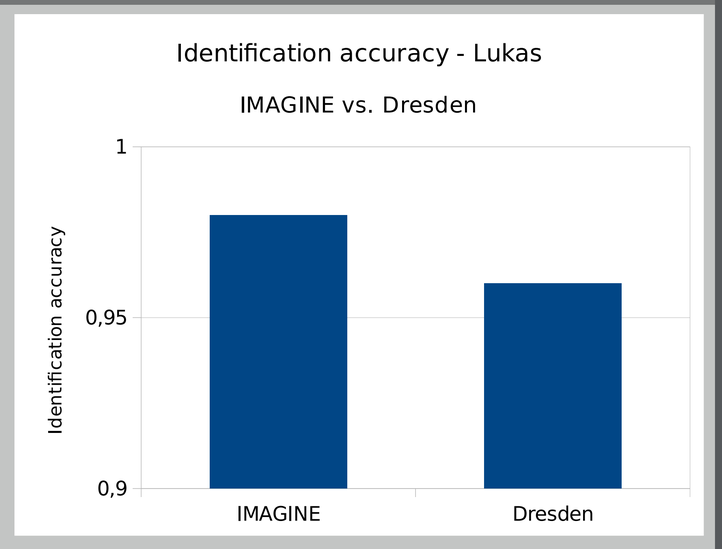
Experimental evaluation (3/4)

- Training accuracy for 20 epochs:
  - IMAGINE: 80.0%
  - Dresden: 70.0%
- Training accuracy for 30 epochs:
  - IMAGINE: 95.0%
  - Dresden: 85.0%
- Training accuracy for 40 epochs:
  - IMAGINE: 98.0%
  - Dresden: 95.0%



Experimental evaluation (4/4)

- Identification accuracy on Lukás algorithm:
  - IMAGINE: 98.0%
  - Dresden: 96.0%



Conclusion

- We have proposed an IMAGINE dataset for benchmarking digital camera identification algorithms;
- The dataset includes a representative number of images coming from modern imaging devices;
- The dataset may be used for testing different camera identification methods;
- Experimental evaluation confirmed the reliability of our dataset.

Contact with authors: jaroslaw.bernacki@pcz.pl, rafal.scherer@pcz.pl