Apart from In-field Sensor Defects, are there Additional Age Traces Hidden in a Digital Image?

13th IEEE International Workshop on Information Forensics and Security
Montpellier, France, 7-10 December, 2021

Robert Jöchl, Andreas Uhl

Department of Computer Sciences, University of Salzburg, Austria

December 9, 2021

Outline

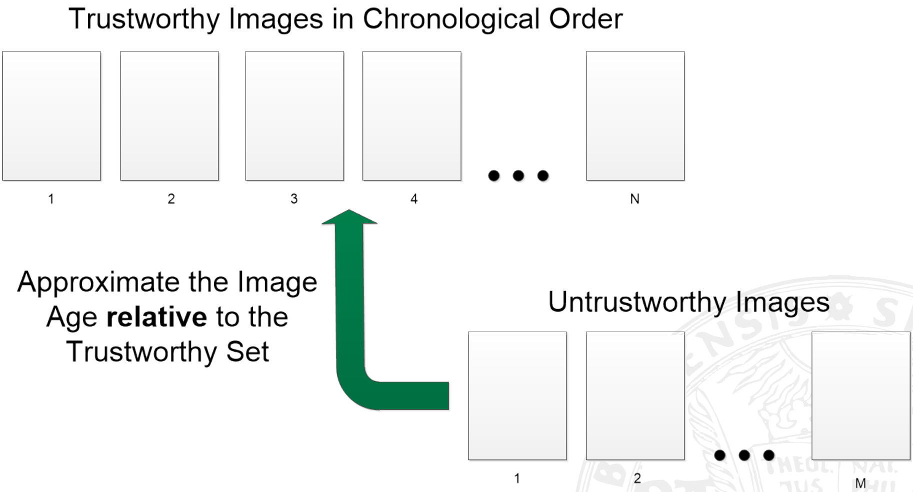


Figure: Overview image age approximation.

In-Field Sensor Defects

In-Field Sensor Defects:

- Develop after the manufacturing process and accumulate over time.
- Are due to cosmic radiation [1].
- Spread to the neighboring pixels because of preprocessing (e.g., demosaicing).

Defect model,

$$F(I) = I + IK + \tau D + c. \tag{1}$$

Figure: In-field sensor defects extracted from captured dark-field images.

Defect-Based Image Age Approximation

Methods for image age approximation based on the presence of sensor defects:

- A maximum likelihood approach introduced by Fridrich et al.[2].
- We propose to utilize traditional machine learning techniques (*i.e.,* a 'Naive Bayes Classifier' and a 'Support Vector Machine') in [3].

⇒ A limitation of both methods is that the defect locations have to be known in advance.

A CNN learns the classification features used.

- Ahmed et al.[4] utilize two well-known CNN architectures (*i.e.*, the AlexNet and GoogLeNet) to approximate the age of a digital image.
- The authors suggest that the features learned are not dependent on a certain image block, since the networks are trained on several non-overlapping image patches.

Questions

- *How relevant is the exact position of a strong in-field sensor defect?*
- *Apart from strong in-field sensor defects, are there additional age traces hidden in a digital image?*
- *Are the learned features position invariant?*

We have systematically investigated these questions.

Used CNN Approach

- Analogy to Image Steganalysis $\rightarrow$ detection of a weak signal.
- A recent approach is the Steganalysis Residual Network (SRNet) published by Boroumand et al. in [5].
  - Based on the residual learning principle [6].
  - The idea is that the residual mapping $F(x) = H(x) - x$, forces the network to preserve the weak embedded stego signal.

$$x = \begin{cases} c + 0, \text{cover} \\ c + m, \text{stego} \end{cases} \tag{2}$$

$\rightarrow$ Since m is a small signal, it can be effectively mapped by $F(x)$.

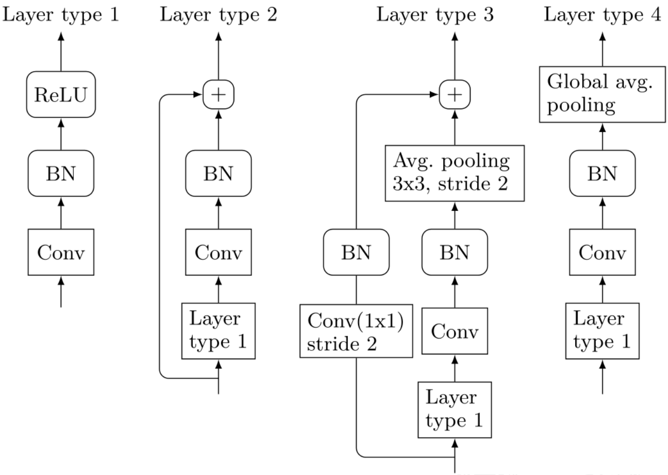


Figure: Overvie of the SRNet layer types [5].

CNN Architecture

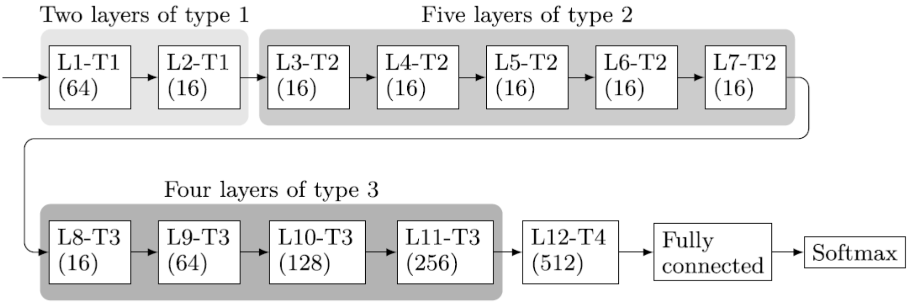


Figure: Layer sequence SRNet [5].

Cropping Methods

roi-crop: extract a small region around each defect (i.e. 32 $\times$ 32)

(a) defect positions (b) roi crops

Figure: Roi-crop example.

Cropping Methods

roi-crop-rp: defect position varies inside the 32×32 region.

rand-roi-crop-rp: extract a 256×256 region at a random position where the resulting patch contains at least one defect.

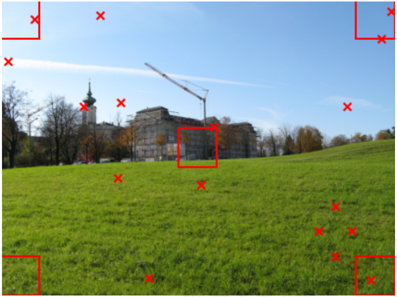
rand-crop: random 256×256 crop completely independet of the exact defect locations.

five-crop-fusion: train five different SRNets each of them with a different fixed image patch (i.e. 256×256).

five-crop: train a single network with all five image patches.

five-crop-ro: apply data augemntation in form of random rotation additionally.

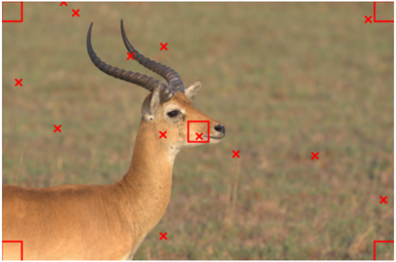
(a) Nikon



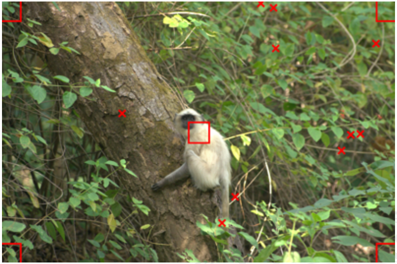
(b) Canon

Figure: Five-crop and defect locations.

Cropping Methods



(c) Pentax K5



(d) Pentax K5II

Figure: Five-crop and defect locations.

(e) Sony

Figure: Five-crop and defect locations.

Dataset

We consider a binary classification problem.

Imager	Class 1	Class 2
Nikon E7600	212 (2005)	320 (2019/2020)
Canon PS A720IS	669 (2008/2009)	331 (2019/2020)
Pentax K5	386 (2013/2014)	362 (2019)
Pentax K5II	465 (2014)	255 (2019/2020)
Sony DSC-P8	369 (2004)	476 (2008)

Table: Overview of images per class and device.

Dataset

The ‘Northumbria Temporal Image Forensics (NTIF)’ database [7]:

- We select images from two devices, a Canon IXUS115HS-1 (NTIF Canon) and a Fujifilm S2950-1 (NTIF Fujifilm).
- The first 5 timeslots are considered the first class (2014), and the timeslots 21-25 represent the second class (2015).
- No strong in-field sensor defects could be found.

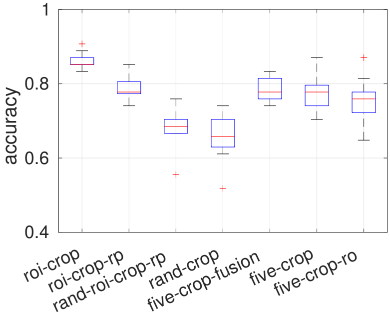
Experiments

SRNet training parameters.

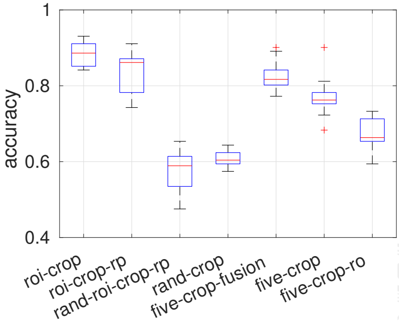
- In general, the training parameters are defined according the definitions in [5].
- The class with fewer samples is oversampled during training.
- Performance evaluation based on the classification accuracy,

$$\mathrm{acc} = \frac{1}{n} \sum_{i}^{n} I[\hat{y} = y]. \tag{3}$$

Experimental Results



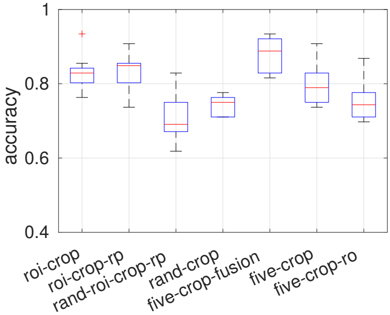
(a) Nikon



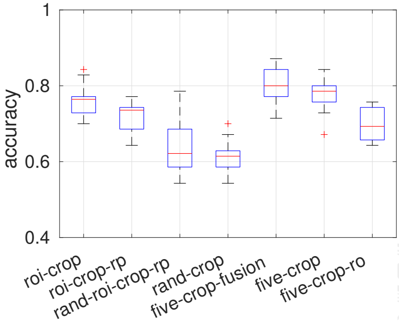
(b) Canon

Figure: Boxplot of the resulting prediction accuracy (10 runs).

Experimental Results



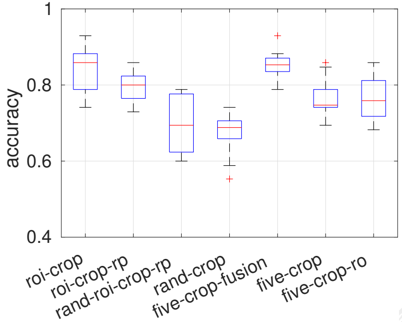
(c) Pentax K5



(d) Pentax K5II

Figure: Boxplot of the resulting prediction accuracy (10 runs).

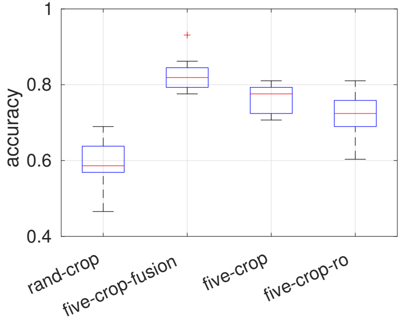
Experimental Results



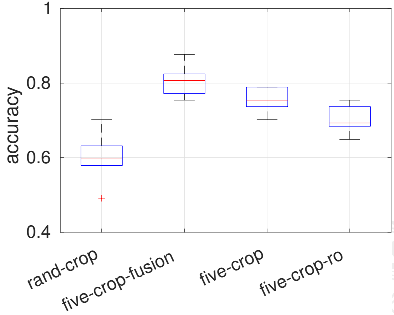
(e) Sony

Figure: Boxplot of the resulting prediction accuracy (10 runs).

Experimental Results



(f) NTIF Canon



(g) NTIF FujiFilm

Figure: Boxplot of the resulting prediction accuracy (10 runs).

Experimental Results

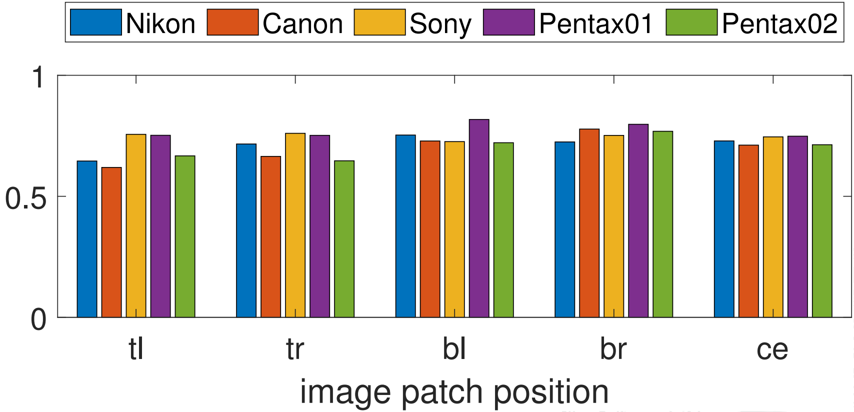


Figure: Average softmax output of the correct class.

Conclusion

- If the CNN is focused on single-pixel defects, the exact position of the defect within the image patch is relevant.
- The presence of strong in-field sensor defects is irrelevant for training the SRNet in the five-crop fusion scenario, implying other age traces are hidden in a digital image.
- The continuous accuracy decrease when reducing the positional dependencies (comparing the ‘five-crop-fusion’, ‘five-crop’ and ‘five-crop-ro’) indicates that these revealed age traces are not position invariant.

References

[1] A. J. Theuwissen, “Influence of terrestrial cosmic rays on the reliability of ccd image sensors part 1: Experiments at room temperature,” *IEEE Transactions on Electron Devices*, vol. 54, no. 12, pp. 3260–3266, 2007.

[2] J. Fridrich and M. Goljan, “Determining approximate age of digital images using sensor defects,” in *Media Watermarking, Security, and Forensics III* (N. D. Memon, J. Dittmann, A. M. Alattar, and E. J. D. III, eds.), vol. 7880, pp. 49 – 59, International Society for Optics and Photonics, SPIE, 2011.

[3] R. Joechl and A. Uhl, “A machine learning approach to approximate the age of a digital image,” in *Digital Forensics and Watermarking: 19th International Workshop, IWDW 2020, Melbourne, VIC, Australia, November 25–27, 2020, Revised Selected Papers*, pp. 181–195, 02 2021.

[4] F. Ahmed, F. Khelifi, A. Lawgaly, and A. Bouridane, “Temporal image forensic analysis for picture dating with deep learning,” in *2020 International Conference on Computing, Electronics Communications Engineering (iCCECE)*, pp. 109–114, 2020.

[5] M. Boroumand, M. Chen, and J. Fridrich, “Deep residual network for steganalysis of digital images,” *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 5, pp. 1181–1193, 2018.

[6] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[7] F. Ahmed, F. Khelifi, A. Lawgaly, and A. Bouridane, “The ‘northumbria temporal image forensics’ database: Description and analysis,” in *2020 7th International Conference on Control, Decision and Information Technologies (CoDIT)*, vol. 1, pp. 982–987, 2020.

Thank you for your attention!