Temporal Image Forensics: Using CNNs for a Chronological Ordering of Line-Scan Data

Matthias Paulitsch[§], Andreas Vorderleitner[§], and Andreas Uhl[0000−0002−5921−8755]

Department of Computer Sciences, University of Salzburg, Salzburg, Austria
uhl@cosy.sbg.ac.at

Abstract. We propose to use CNNs for obtaining a temporal ordering of line-scanner data. Excellent age classification accuracy is achieved only in case network training and testing is done with image patches taken from consistent spatial locations, i.e. temporal features exploited are bound to specific positions in the image. With spatially consistent patches, up to 100% classification accuracy can be achieved, whereas with spatially varying patches the accuracy stagnates at around 54%. We have also noted a result dependency on image content and have found, that a consistent patch position relative to the scanning line is not sufficient for good results.

Keywords: temporal forensics · image age · chronological order · CNN.

1 Introduction

In temporal image forensics [3,4], the main objective is to create a chronological sequence of pieces of evidence. A chronological sequence of images can help to deduce a causal relationship between events, which can be important in a court trial, for example. In general, images can be ordered chronologically based on the acquisition time-stamp stored in the EXIF header. However, this time-stamp is very easy to manipulate and therefore not trustworthy. For this reason, methods are required that approximate the age of an image based on traces left during the image acquisition.

One source of such traces is the image sensor. More precisely, individual pixels can become defective. These defects are known as in-field sensor defects (i.e. they develop after the manufacturing process) and they accumulate over time. In-field sensor defects are due to cosmic radiation and lead to an offset in the dynamic range of the defective pixel. Since in-field sensor defects accumulate over time, the age of image under investigation can be deduced from the detected defects. Consequently, the state-of-the-art in this field currently relies on the analysis of detected pixel defects [3, 4], a process not necessarily very accurate, especially when image content exhibits many textured areas.

[§] The two authors contributed equally.

Overall, temporal forensics have to look at subtle changes in images, eventually caused by single pixels getting defective over time. From the perspective of the requirement to analyse such subtle image properties, the fields of steganalysis (i.e., broadly speaking, steganalysis aims at finding messages hidden by steganographic methods in data) and camera authentication / identification (typically done by analyzing sensor noise properties), respectively, are closely related to temporal forensics. While the technology used in temporal image forensics is currently based on model-based approaches, the field of camera / sensor recognition has already seen the successful application of convolutional neural networks (CNN) [1, 2, 6, 7, 10]. Also in the field of steganalysis, deep learning and CNNs already received much attention from academia. Multiple networks were developed with growing success [8,9,13–15]. CNNs succeeded in extracting complex statistical dependencies from images and also achieved to improve the detection accuracy as compared to traditional techniques.
Thus, for our aim in temporal image forensics, we focused on CNNs, in particular on networks trained on finding hidden messages in images. Due to the similarity of the tasks to be conducted, the architectures of state-of-the-art steganalysis networks served us as a model for our network architecture. Basically those networks consist of three main parts, beginning with a preprocessing, continuing with a convolution part and concluding with a classification module [13–15]. All those parts can also be found in our network architecture, discussed in section 2.1. Alike to steganalysis, our network should find hidden messages. In our context, however, the hidden message is any information about the age of an image. Information of that kind could be pixel defects, or other alterations occuring caused by the ageing process. This also includes information, not detectable for human eyes.

Techniques in temporal forensics currently rely on the analysis of single pixel defects [3,4] and focus on classical consumer camera (area) sensors. In this paper, we introduce the usage of convolutional neural networks for this task and consider image material acquired by a line-scanner.

In section 2 we describe the structure of our network and introduce a variety of different patch selection modes. We discuss accuracy results in section 3 and give an intuition on the development of the loss. Section 4 presents our concluding remarks.

2 Network

2.1 Architecture

Our network architecture is illustrated in Figure 1. It constitutes a modified version of AlexNet [5] and can be used as a single-channel or multi-channel network. Multiple steganalysis networks inspired the fundamental construction [8,9,13–15]. For the multi-channel approach, we were inspired by the implementation of [1]. Each channel has the same structure and is based on the CNN network of [6]. We introduced the possibility to use one or multiple channels in the network architecture to investigate the importance of spatial position in

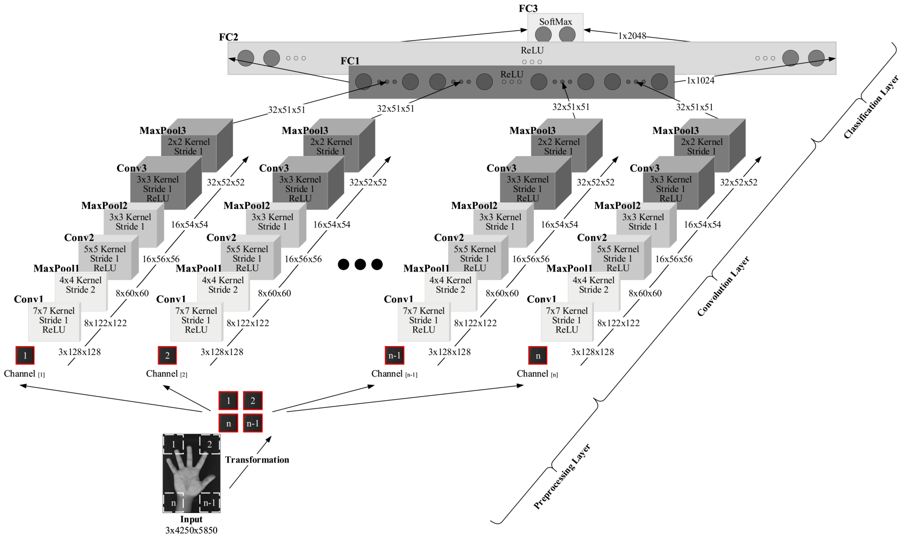


Fig. 1. Proposed CNN architecture.

images. In the multi-channel approach, input patches for each channel can be taken from an identical position in the image. Thus, each channel can be trained on its specific positional information by always receiving its input from the same image spatial location, while in the single-channel approach the position of the patches in the image used for training might change (depending on the patch selection strategy).

The architecture is structured in three parts, namely the preprocessing layer, the convolution layer and the classification layer. Starting with the **Preprocessing Layer (PP)**, where images are fed into the network. Depending on the chosen patch selection mode the input image is transformed into a single cropped image or into multiple cropped images of size (img-w × img-h), (e.g. 128 × 128). Further details can be found in section 2.2. All resulting images are fed sequentially into the single-channel or one by one into each of the multi-channels. The **Convolution Layer (CV)** consists of a single-channel or multiple single-channels (multi-channel). Each channel starts with a 2D convolution at a kernel size of 7×7 and stride 1, followed by a Rectified Linear Unit (ReLU) function. Then, a 2D max pooling with a kernel size of 4×4 and stride 2 is applied. Every channel consists of three such blocks where the kernel size of the convolution is reduced by two and the kernel size of the max pooling is reduced by one at every block. The stride of the max pooling is then set to one for the following blocks. The image size gets reduced by each block while the feature size increases. After the last block the feature size is 32 and the image size is 51×51. This results in an output size of $32 \times 51 \times 51$ for each channel. The **Classification Layer (CL)** is featured with three Fully Connected (FC) sub-layers. A linear transformation is applied in the first two sub-layers, followed by a ReLU function while the last sublayer also applies a linear transformation, but followed by a softmax function with an output size of two. This output specifies, if an image is classified as Old, or New. The input size of the first sub-layer is ch × (32 × 51 × 51) where ch denominates the number of channels. 1024 output features are produced in this process. The input size of all FC sub-layers, except the first of the three, corresponds to the output size of their preceding FC sub-layer.

2.2 Patch Selection Strategies

We followed different strategies for selecting certain image patches from scanner images, which are fed into the CNN eventually. All selection strategies described below can be applied to both versions of the network (multi-channel and single-channel), unless it is explicitly stated otherwise. Recall that running the network in multi-channel mode implies that a certain channel always receives its input from the same patch position in the image. This aims at maintaining the spatial consistency, in order to investigate the importance of spatial position coherence when investigating ageing effects. When using different patch selection strategies, the accuracy of the classification varies, as discussed in more detail in section 3. All patch selection schemes use a unified patch size. (i.e. 128 × 128).

Crop Five (CF). CF takes crops in the exact following order from the left/right top, right/left bottom and from the center position of the scanner image, marked in red in Figure 2a. Furthermore, the single crop (SC) positions represent a selection scheme on their own, applied only in single-channel mode. For example, the center mode uses only the center patch of an image combined with the single-channel mode. The SC versions are denominated by LT (Left Top), RT (Right Top), RB (Right Bottom), LB (Left Bottom), and C (Center).

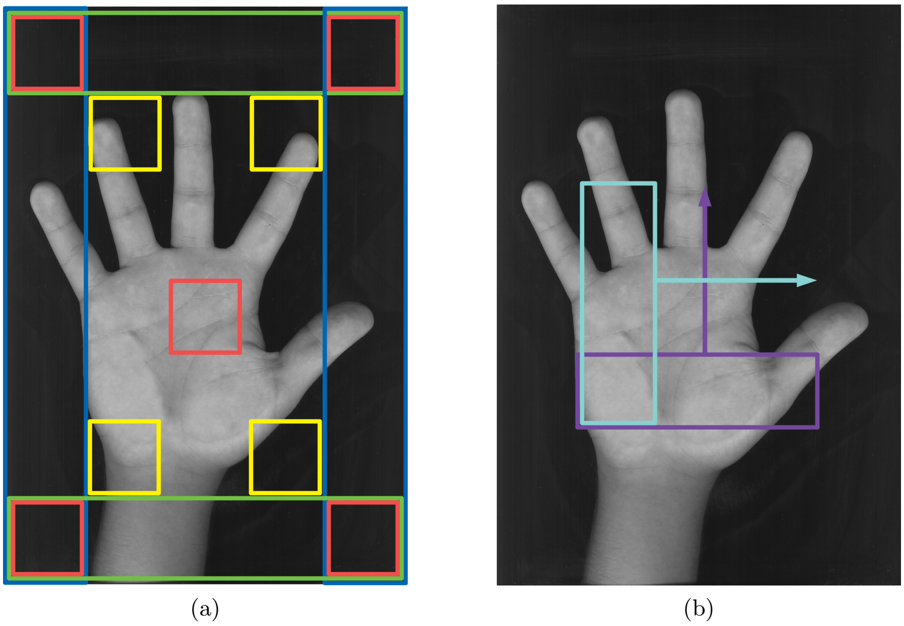
Crop Ten (CT). CT uses the same patch positions as CF and in addition also uses the horizontally flipped version of every crop, again illustrated by the red crops in Figure 2a.

Spiral Eight (SE). SE uses the four corner crop positions, alike to CF. Additionally it uses the corners of the "inner" image of size $(\text{width} - 2 \cdot \text{blocksize}) \times (\text{height} - 2 \cdot \text{blocksize})$. SE results from Figure 2a when combining all yellow patches with the red patch positions, without the center patch.

Line Mode (LM-X). LM-X starts by cutting the scanner images into (scan-width× img-h) sized slices, resulting in several lines. These are treated as images from now on. The patches are then selected from the produced line-images. X denominates the number of patches taken from a line. Two example line-images produced by LM-X are illustrated in green in Figure 2a.

Column Mode (CM-X). CM-X represents the opposite approach to LM-X, creating new images by cropping the scanner images into (img-w × scan-height) sized slices, resulting in columns. The patches are then again taken from the produced column-images. X denominates the number of crops, taken from a column. This behaviour is shown in blue in Figure 2a. Note that the scan direction is from top to bottom in the images displayed. Thus, patches extracted from column-images have the identical relative position to the scan-line, and thus may be considered to have the same spatial position.

Center Block Mode (CBM-X). CBM-X is applied in combination with either CM-X, or LM-X. All patches are taken from the center area of the scanner image. Depending on the combined patch selection mode, either parts of centered columns, or centered lines are used as patches. Again X denominates the number of patches, taken from a line/column. This procedure aims at using only patches mostly consisting of scanned hand parts. The approximately covered center area is illustrated in Figure 2b, where CBM + LM-X is colored purple and CBM + CM-X is colored cyan.



(a)

(b)

Fig. 2. (a) Different crop positions used in patch selection modes, (b) CBM crop positions.

3 Experiments

3.1 Data-Set

We use a data-set that has been used before in the context of investigating ageing effects in hand-related biometrics [11, 12]. We collected a database of high-resolution human palmar handprints (see Figure 2a and Figure 2b) from 28 members in our lab with 443 hand images in the first session (Old) captured in November 2007 and 164 hands in the second session (New) captured in October/November 2012, i.e. exhibiting a time lapse of approximately 5 years between recordings adhering to the same strict recording protocol (users were allowed to wear rings or watches and occupy an arbitrary position on the scanner as long as fingers did not touch each other). During the 5 years lapse, no images were taken with the scanner. Image acquisition was done using the same instance of a flatbed HP 3500c scanner for both sessions, recording a 216 × 297 millimeters area at 500 dpi resolution (resulting in 4250 × 5850 pixels sized images). Due to the limited illumination capabilities of traditional flatbed scanning devices, environmental surrounding light was shielded by a scanning box with a round hole for hand insertion.

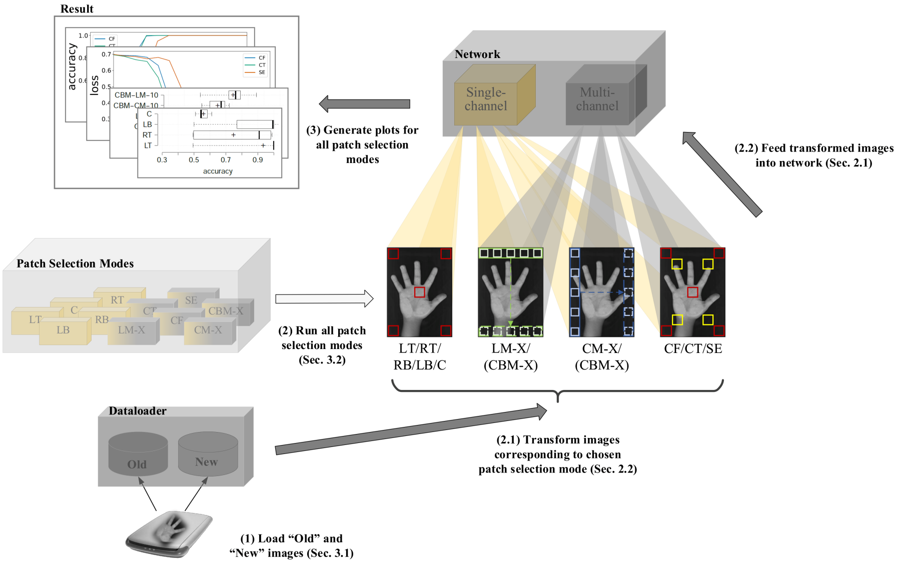

Result
accuracy
loss
CF
CT
SE
CBM–LM–10
CBM–CM–10
C
LB
RT
LT
0.3
0.5
0.7
0.9
accuracy
(3) Generate plots for all patch selection modes
Network
Single-channel
Multi-channel
(2.2) Feed transformed images into network (Sec. 2.1)
Patch Selection Modes
RT
SE
C
CT
CBM-X
LT
RB
CF
LB
LM-X
CM-X
(2) Run all patch selection modes (Sec. 3.2)
LT/RT/ RB/LB/C
LM-X/ (CBM-X)
CM-X/ (CBM-X)
CF/CT/SE
Dataloader
Old
New
(2.1) Transform images corresponding to chosen patch selection mode (Sec. 2.2)
(1) Load "Old" and "New" images (Sec. 3.1)

Fig. 3. Experimental workflow.

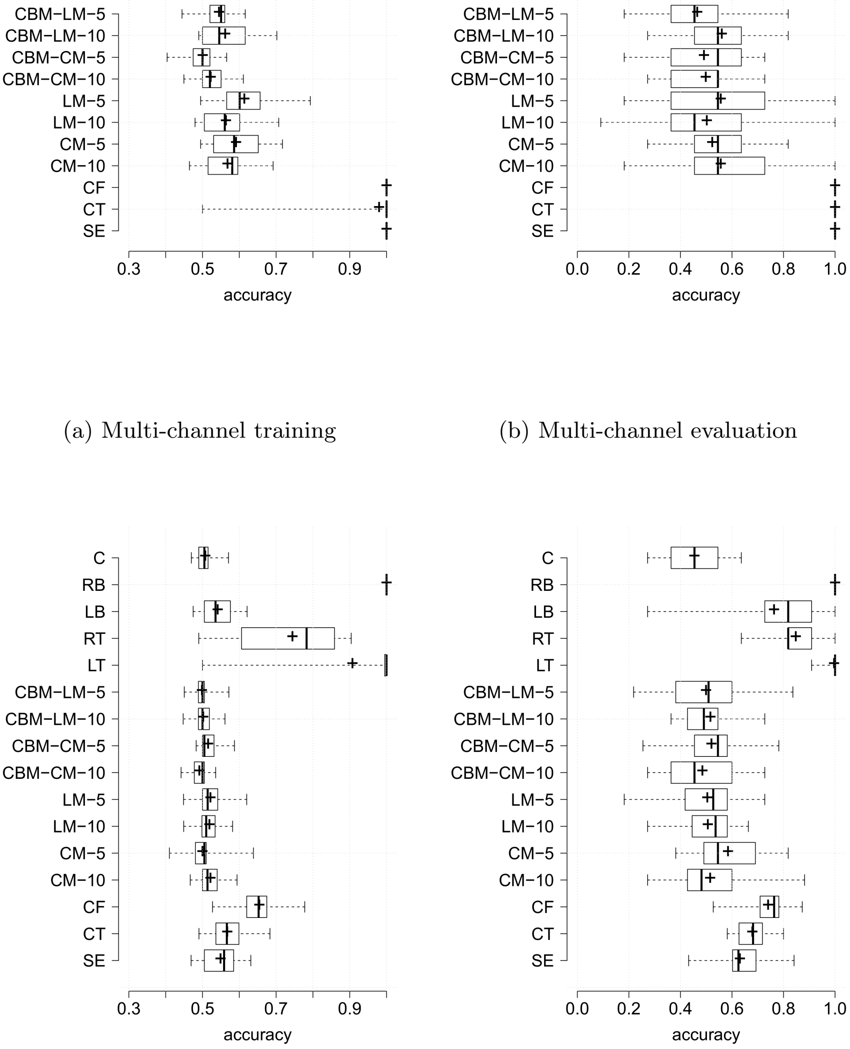
3.2 Experimental Setup

We conduct our experiments on a NVIDIA TITAN X (Pascal) 12 GB, with CUDA version 10.1. For all experiments a stochastic gradient descent optimizer with an initial learning rate of 0.01 and a momentum of 0.9 is used alike to [5]. Unlike the previous article, we set the decay to 0.005, as seen in [6]. We varied the values for learning rate, momentum, and decay, but found out that the network converged best with the selected values. The same learning rate is applied to all layers. Additionally, the bias is disabled for all convolutional layers, but remains enabled for the linear layers. No dropout is used. The weights follow a zero-mean normal distribution with a standard deviation of 0.08. We use an equal number of 164 images from the data-set for each of the two classes. The data-set is then divided into a training set, a validation set and an evaluation set with a division of 64%, 16% and 20% for a 5-fold cross-validation (5-CV). The validation set is used as feedback regarding to overfitting during the training process and the evaluation set serves for eventually testing the classification accuracy on data, which the CNN has never seen before. The batch size is set to 11 except for those experiments which are running in patch selection modes CF, CT, SE, LM-X, CM-X, CBM-X with explicitly deactivated multi-channel mode. Here the batch size is multiplied by the number of patches of the chosen selection mode. We use 19 batches for the training, four batches for the validation and five for the evaluation in each epoch. The 5-CV is repeated five times (i.e. 25 runs) with a fixed number of 15 epochs and the main results are plotted in Figure 4. For each 5-CV repetition, the data distribution of the data-set is random. Training the network with these parameters in multi-channel mode, CT selection mode and patches of size 128×128 takes about 60 minutes, with convergence within 10 epochs. Whereas in CF selection mode it takes about 30 minutes. We did not put our focus on training performance, therefore, undoubtedly this can be improved with some effort. In addition to the 5-CV with a fixed number of 15 epochs, we ran another 5-CV for some specific patch selection modes with 100 epochs (long-run) for three times to ensure the accuracy results. Figure 3 presents the workflow of our study.

3.3 Classification Accuracy

In this subsection we discuss the achieved classification accuracy with respect to the used patch selection mode, single- and multi-channel setup, respectively, and discuss the results. Furthermore, we observe the development of the loss in relation to the trend of the accuracy. In the following, it is assumed that all patch selection modes are executed in multi-channel mode, unless the operation in single-channel mode is expressly pointed out.

The best training accuracies are achieved by the CF, CT and SE selection modes. As early as the second epoch, the loss decreases until it converges to 0.3 at epoch 8. While the loss is decreasing the accuracy increases up to 1.0. This behavior can be observed in Figure 5a and Figure 5b. Those three patch selection modes also achieve the best evaluation accuracies, which exactly reflect the results of the training. Figure 4 and Figure 12 show the training and evaluation accuracies of all experiments, which are structured according to their patch selection modes. In Figure 4a and Figure 4b we recognize that CF, CT and SE achieve the same training and evaluation accuracy. To clarify the importance of the spatial position of the patches, we apply the same experiment in the dedicated single-channel mode (where a single network is trained using patches from different spatial locations). The results are shown in Figure 4c and Figure 4d. The single-channel versions of CF, CT and SE perform clearly worse than in multi-channel mode. We can observe that the training and evaluation accuracy



(a) Multi-channel training

(b) Multi-channel evaluation

(c) Single-channel training

(d) Single-channel evaluation

Fig. 4. Boxplot of the resulting training and evaluation accuracy for the five 5-CV runs in single-channel and multi-channel mode.

clearly do not reach the same level as the multi-channel version. Thus, we may clearly state that the position information of the patches at least contributes to high image age classification accuracy.

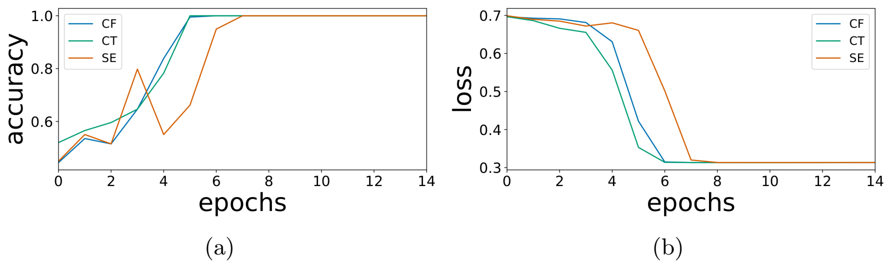
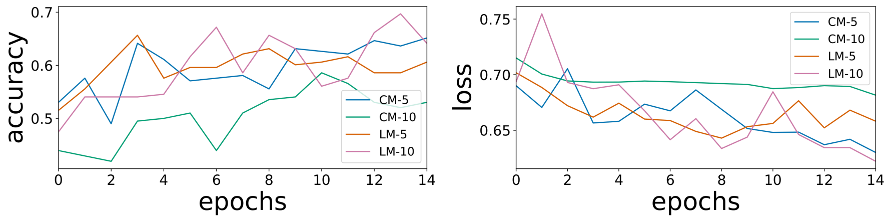


Fig. 5. (a) Comparing the training accuracies of the three best patch selection modes CT, CF and SE in multi-channel mode from an arbitrary sample run, (b) Comparing the training convergence of the three best patch selection modes CT, CF and SE in multi-channel mode from an arbitrary sample run.

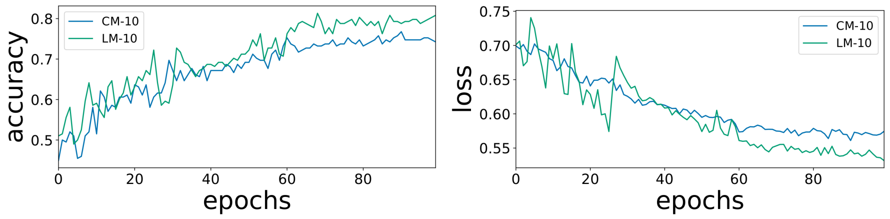
Figure 6a shows the training accuracy of the CM-X and LM-X selection modes. The purpose of this experiment is to find out whether it is sufficient to use the patches from the same relative position to the scan-line, but not identical in the overall image. Patches hereby always come from the same position in the row or column, but the position of the entire row or column slides over the original scanner image. As a result, the absolute position of the patches changes continuously, but its row/column-position always stays the same. A recurring cycle is created, but of course different than in CF, CT and SE. The training and evaluation accuracies are far behind the results of CT, CF and SE. In single-channel mode, the accuracies tend to be similar, but CT, CF, and SE are still better. The loss also shows no signs of convergence over the entire epochs, as can be seen in Figure 6b. Regarding the LM-X selection modes, it can be stated that these produce slightly better results than the CM-X modes, as can be seen in Figure 4. The long-runs of the LM-X and CM-X in Figure 7 show a convergence of the loss, but the results as explained before do not change. Overall, it turns out that the same relative position to the scan-line is not sufficient for decent classification (instead, a fixed position in the image is required to successfully train a network), even stronger, as we do not identify a clear difference between CM-X and LM-X, the identical position relative to the scan-line does not matter.

Furthermore, we tested a variant of LM-X, which only uses the top five lines of a scanner image to see whether the hand areas of the image are causing the poor accuracy. Those five lines mostly consist of black background. However, the results are almost exactly the same as using the entire lines of the scanner image. The same applies to a variant of CM-X, which only uses the rightmost 5 columns. For both, LM-X and CM-X, this holds for the multi-channel as well as the single-channel versions (results are not shown).



(a) (b)

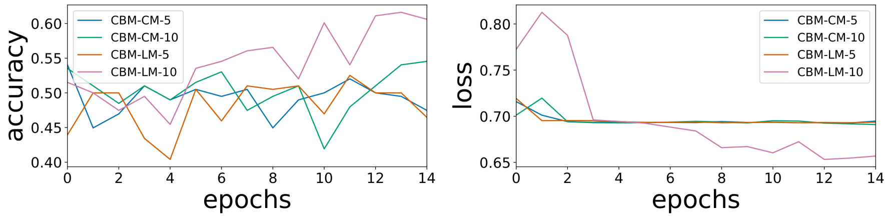
Fig. 6. (a) Comparing the training accuracies of CM-X and LM-X patch selection in multi-channel mode from an arbitrary sample run, (b) Comparing the training convergence performance of CM-X and LM-X patch selection in multi-channel mode from an arbitrary sample run.



(a) (b)

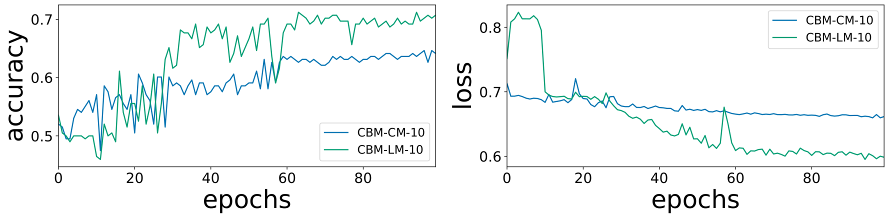
Fig. 7. (a) Comparing the training accuracies of CM-X and LM-X patch selection in multi-channel mode from an arbitrary long-run sample, (b) Comparing the training convergence performance of CM-X and LM-X patch selection in multi-channel mode from an arbitrary long-run sample.

In contrast to these variants of LM-X and CM-X, the patch selection modes CBM-CM-X and CBM-LM-X intend to use only the hand area of the scanner image. This evaluates the effects of the hand area on the accuracy of the patch selection modes LM-X and CM-X. Figure 8a and Figure 8b present the accuracy and the loss of the training. In multi-channel mode we observe a decrease of the resulting training accuracy, while the evaluation accuracy is largely the same. In single-channel mode the training and evaluation accuracy decreases slightly, compared to the LM-X and CM-X. Thus, the line- and column oriented patch selection strategies do not turn out to be competitive to those approaches, where the spatial location of patches used in training and evaluation is fixed. Figure 9 shows the convergence of the loss up to epoch 100. Comparing the long-runs of the CBM-CM-X, CBM-LM-X to the CM-X, LM-X patch selection modes the relation does not change.



(a) (b)

Fig. 8. (a) Comparing the training accuracies of CBM-CM-X and CBM-LM-X patch selection modes in multi-channel operation, (b) Comparing the training convergence performance of CBM-CM-X and CBM-LM-X patch selection modes in multi-channel operation, all from the same arbitrary sample run.



(a) (b)

Fig. 9. (a) Comparing the long-run training accuracies of CBM-CM-X and CBM-LM-X patch selection modes in multi-channel operation, (b) Comparing the training convergence performance of CBM-CM-X and CBM-LM-X patch selection modes in multi-channel operation, all from the same arbitrary long-run sample.

Finally, we look at the single patch results - the SC selection modes show very different results for the training as well as for the evaluation accuracy. Where RB always has an accuracy of 1.0, the C selection mode performs worst in both cases (so it seems that the image content - skin texture in this case - has a clearly negative impact on age classification results). See Figure 13 for a comparison of “Old” and “New” background patches - a stripe pattern in orthogonal direction to the scan direction is visible, with stronger and clearer patterns for the "New" class (a zoom into patches is shown). LB, RT, and LT vary widely when it comes to training accuracy. In terms of evaluation accuracy, they do not vary that much and achieve better results, but do not reach the performance of RB. The comparison between training accuracy and loss for all SC patch selection modes is shown in Figure 10a and Figure 10b. Due to the extreme instability of the resulting accuracy between training and evaluation, the SC selection modes, with the exception of RB, are not sufficient enough to make a reliable classification. The long-runs of the SC selection modes, shown in Figure 11, reflect those results. Increasing the size of the data-set possibly has a positive impact on the reliability and the performance of the SC selection modes.

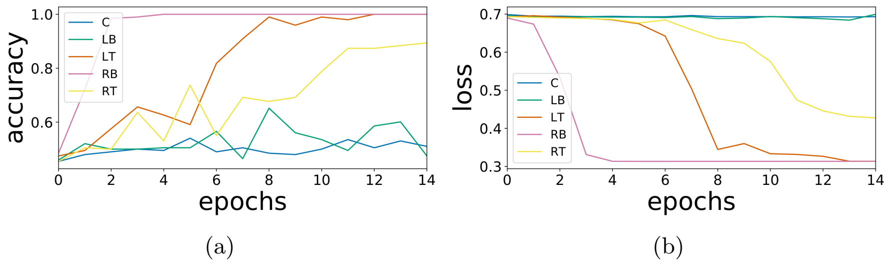


Fig. 10. (a) Comparing training accuracies of the five SC patch selection modes in single-channel operation, (b) Comparing the training convergence performance of the five SC patch selection modes in single-channel operation, all from the same arbitrary sample run.

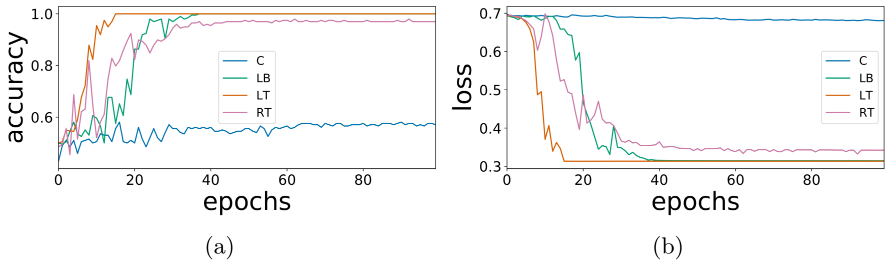
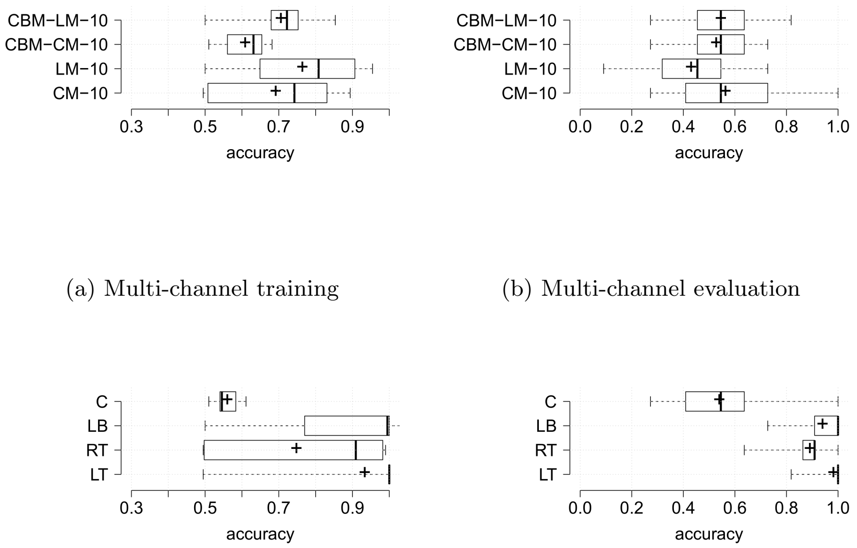


Fig. 11. (a) Comparing the long-run training accuracies of all except RB SC patch selection modes in single-channel operation, (b) Comparing the training convergence performance of all except RB SC patch selection modes in single-channel operation, all from the same arbitrary long-run sample.

To conclude this section we want to point out that the confusion matrices of two different modes, namely RT and LM-10, show different error patterns (see Figure 14). After looking at three arbitrary sample runs it can be observed that RT tends to classify "New" images as "Old", whereas LM-10 does not show any apparent pattern. This holds true for both, a run of 15 epochs and a long-run.



(a) Multi-channel training

(b) Multi-channel evaluation

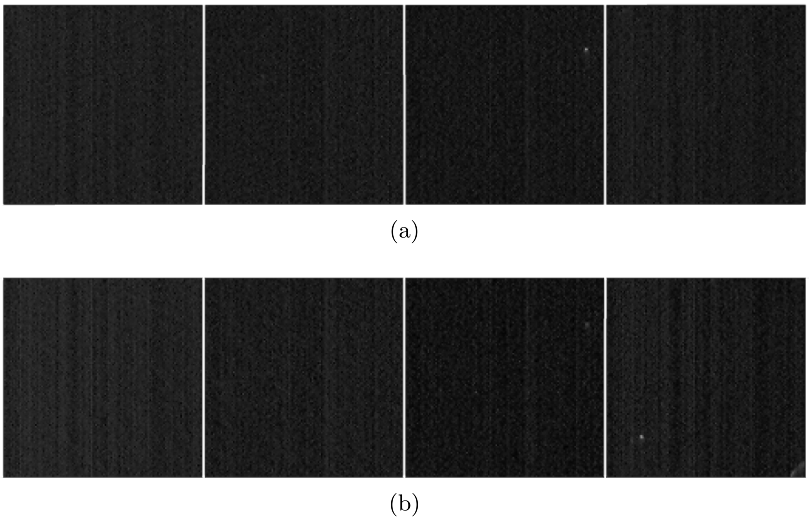
(c) Single-channel training

(d) Single-channel evaluation

Fig. 12. Boxplot of the resulting training and evaluation accuracy for the three 5-CV long-runs in single-channel and multi-channel mode.

4 Conclusion

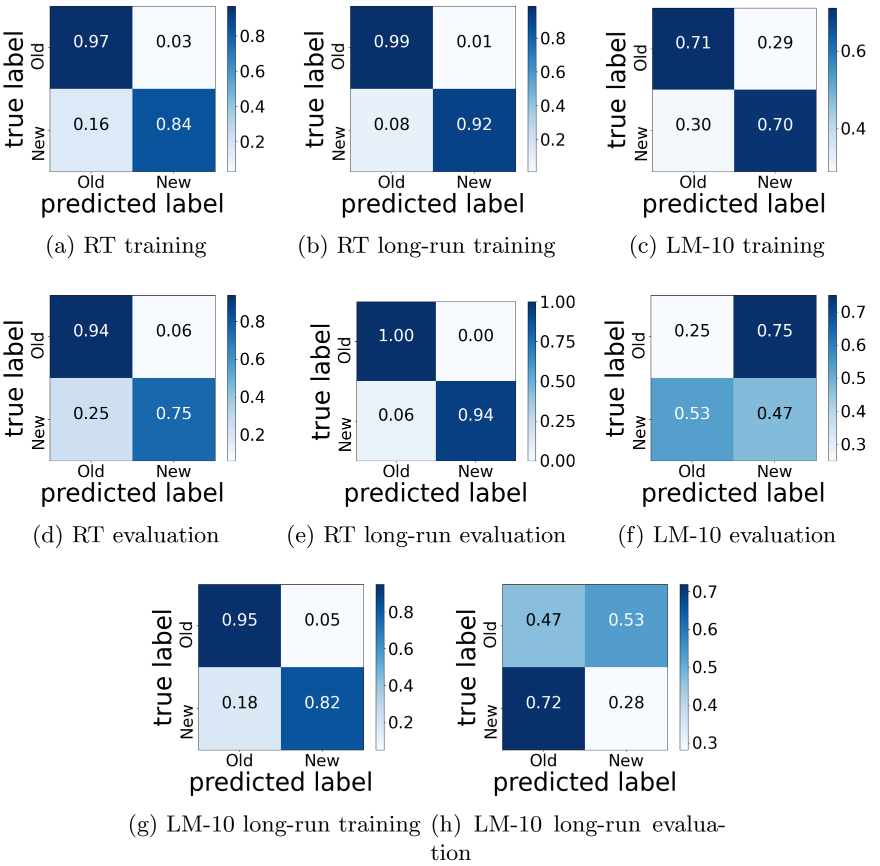
We introduced CNNs for classifying the age of scanner images. Even with a relatively small data-set, the CNN accomplished to reach high classification accuracies. We showed that maintaining the spatial consistency of patches that are fed into the network induces a clearly positive effect on the classification accuracy (i.e.: median: 100% consistent vs 54% inconsistent). This was mainly done by exploiting the benefits of multi-channel networks over single-channel networks for classifying the age of scanner images. Overall, a single network has to be trained and tested with patches from the same spatial location in the image to achieve high classification accuracy - this implies that features used for classification are local features bound to specific positions on the sensor. The single-channel versions of the network also occasionally showed decent results, but overall with a apparently lower reliability than the multi-channel versions (i.e.: highest median: 100%, lowest median: 50% of classification accuracies). Further, we were not able to exploit the line-scan nature of our data to reduce the required spatial consistency to a consistency of relative position to the scan-line. Based on our findings we see big potential for CNNs for future utilization in the field of temporal image forensics, however, a clear result dependency on image content has to be stated (i.e. uniform background leads to best results observed).



(a)

(b)

Fig. 13. Comparison between the corner crops of CF of a sample image from (a) "Old" class, (b) "New" class. (i.e. from left to right: LT, RT, RB, LB)



(a) RT training (b) RT long-run training (c) LM-10 training

(d) RT evaluation (e) RT long-run evaluation (f) LM-10 evaluation

(g) LM-10 long-run training (h) LM-10 long-run evaluation

Fig. 14. Confusion matrices from arbitrary sample runs out of the different patch selection modes and run lengths.

References

1. Bayar, B., Stamm, M.C.: Augmented convolutional feature maps for robust cnn-based camera model identification. In: 2017 IEEE International Conference on Image Processing (ICIP). pp. 4098–4102 (2017). https://doi.org/10.1109/ICIP.2017.8297053
2. Bondi, L., Baroffio, L., Güera, D., Bestagini, P., Delp, E.J., Tubaro, S.: First steps toward camera model identification with convolutional neural networks. IEEE Signal Processing Letters **24**(3), 259–263 (2017). https://doi.org/10.1109/LSP.2016.2641006
3. Fridrich, J., Goljan, M.: Determining approximate age of digital images using sensor defects. In: Memon, N.D., Dittmann, J., Alattar, A.M., III, E.J.D. (eds.) Media Watermarking, Security, and Forensics III. vol. 7880, pp. 49 – 59. International Society for Optics and Photonics, SPIE (2011). https://doi.org/10.1117/12.872198
4. Jöchl, R., Uhl, A.: A machine learning approach to approximate the age of an digital image. LNCS (2020 (to appear))
5. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. Commun. ACM **60**(6), 84–90 (May 2017). https://doi.org/10.1145/3065386
6. Marra, F., Poggi, G., Sansone, C., Verdoliva, L.: A deep learning approach for iris sensor model identification. Pattern Recognition Letters **113**, 46–53 (2018). https://doi.org/10.1016/j.patrec.2017.04.010, integrating Biometrics and Forensics
7. Mendes Júnior, P.R., Bondi, L., Bestagini, P., Tubaro, S., Rocha, A.: An in-depth study on open-set camera model identification. IEEE Access **7**, 180713–180726 (2019). https://doi.org/10.1109/ACCESS.2019.2921436
8. Reinel, T.S., Raúl, R.P., Gustavo, I.: Deep learning applied to steganalysis of digital images: A systematic review. IEEE Access **7**, 68970–68990 (2019). https://doi.org/10.1109/ACCESS.2019.2918086
9. Tan, S., Li, B.: Stacked convolutional auto-encoders for steganalysis of digital images. Signal and Information Processing Association Annual Summit and Conference (APSIPA) pp. 1–4 (2014). https://doi.org/10.1109/APSIPA.2014.7041565
10. Tuama, A., Comby, F., Chaumont, M.: Camera model identification with the use of deep convolutional neural networks. In: Proceedings of Workshop on Information Forensics and Security (WIFS'16). Abu Dhabi, United Arab Emirates (2016). https://doi.org/10.1109/WIFS.2016.7823908
11. Uhl, A., Wild, P.: Age factors in biometric processing. IET pp. 153–170 (2013)
12. Uhl, A., Wild, P.: Experimental evidence of ageing in hand biometrics. In: Proceedings of the International Conference of the Biometrics Special Interest Group (BIOSIG'13). pp. 39–50. Darmstadt, Germany (2013)
13. Xu, G., Wu, H.Z., Shi, Y.Q.: Structural design of convolutional neural networks for steganalysis. IEEE Signal Processing Letters **23**(5), 708–712 (2016). https://doi.org/10.1109/LSP.2016.2548421
14. Yedroudj, M., Comby, F., Chaumont, M.: Yedroudj-net: An efficient cnn for spatial steganalysis. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 2092–2096 (2018). https://doi.org/10.1109/ICASSP.2018.8461438
15. Zhang, R., Zhu, F., Liu, J., Liu, G.: Depth-wise separable convolutions and multi-level pooling for an efficient spatial cnn-based steganalysis. IEEE Transactions on Information Forensics and Security **15**, 1138–1150 (2020). https://doi.org/10.1109/TIFS.2019.2936913