Improving Smartphone-Based Dental Ceramic Manufacturer Classification Across Different Devices*

Johannes Schuiki[1] [0000-0002-6197-2327], Christof Kauba[1] [0000-0002-2716-1360], Heinz Hofbauer[1] [0000-0003-2969-1848], and Andreas Uhl[1] [0000-0002-5921-8755]

Department of Artificial Intelligence and Human Interfaces,
Paris Lodron University Salzburg, Salzburg, Austria
Corresponding author: jschuiki@cs.sbg.ac.at

Abstract. Driven by suboptimal results in cross-sensor manufacturer classification for dental ceramics, this work proposes performance enhancements by transitioning from traditional texture descriptors to deep learning architectures. Previous studies indicated that using raw images instead of ISP-processed data leads to only minor improvements and fails to fully resolve the domain shift in 1-to-1 cross-sensor scenarios. In contrast, we propose utilizing raw color filter array data, reordered into three channels by averaging the two green components. Our findings suggest that sample-wise normalization further mitigates inter-sensor variability. Furthermore, employing noise and denoising as data augmentation strategies produces measurable accuracy gains by forcing the model to learn sensor-agnostic features. Experimental results demonstrate that these targeted engineering enhancements significantly outperform the existing literature benchmarks in cross-sensor dental material manufacturer identification.

Keywords: manufacturer classification · dental ceramics · cross-sensor texture classification · object surface classification

1 Introduction

Intrinsic, non-invasive product authentication, i.e. confirming a products' origin or determining its material class without damaging the object and without the application of external markers etc. is often preferred for counterfeit product detection. An emerging application area is dental ceramics, particularly zircon oxide blocks used in the manufacturing of dental prostheses. In this context, two scenarios are of particular interest: i) forgery detection and ii) verification that a block inserted into a milling machine genuinely originates from a specific manufacturer, enabling rejection of materials from competing brands.

The ultimate goal is to be able to do non-invasive dental ceramic material classification using arbitrary consumer grade smartphones, without the need of a

* This work is supported by the Christian Doppler Laboratory for People and Object Surface Authentication.

previous calibration for each unseen device. Hence, after employing exactly one smartphone for capturing a training set (possibly without raw imaging capability), images from any smartphone in the wild without raw capability should be classified correctly. An on-clip macro lens is currently necessary, hence deemed acceptable. However, our previous work showed that this goal is by no means straight forward to achieve [10,12,13], especially if multiple distinct smartphone devices are utilized (cross-sensor scenario) [14].

This work is continuing our previous efforts and aims for improving the practical application of smartphone based material classification by introducing several engineering enhancements. Prior work relied on hand-crafted features (e.g., dense SIFT) and found limited utility in raw sensor data. In the present work, the previous hand-crafted feature extraction pipeline is replaced by a deep learning–based approach. Specifically, the proposed method employs re-trained deep neural network architectures operating directly on raw color filter array (CFA) data. To enable compatibility with standard network architectures, the raw sensor measurements are restructured into three-channel (RGB) representations. In addition, lightweight normalization procedures (in particular, white balancing according to metadata and sample-wise z-score normalization per channel) are applied, thereby aligning the model with real-world constraints where global dataset statistics are unavailable. Furthermore, the study demonstrates that noise-based data augmentation yields a further improvement in classification performance, altogether resulting in significant accuracy gains compared with prior work.

The rest of this paper is organized as follows: Section 2 summarizes related work on material classification and gives an overview on our previous work. Section 3 explains the experimental setup and justifies the proposed enhancements. In section 4, the results are presented along with some discussion. Finally, section 5 concludes this study.

2 Related Work

There is extensive work on smartphone based extrinsic product authentication, either based on so called copy detection patterns (CDPs) [18,19] or other types of printed QR codes [1,24]. However, our focus is on intrinsic non-invasive product authentication (or material classification), which alleviate the main drawback of the extrinsic ones, i.e. the need of applying any external properties to the product.

Several previous works have demonstrated that intrinsic product authentication or material classification is feasible using commodity smartphones. A prominent line of research relies on physical unclonable functions (PUFs) derived from surface micro-structures [21,3,18]. In [16], an efficient micro-structure orientation estimation technique is utilized which includes modeling of the light propagation path. In a related line of work, [15] demonstrate the detection of counterfeit liquid food products based on the characteristics of air bubbles formed when a bottle is flipped. Furthermore, [10] present an approach for drug packaging authentication using micro-texture images in an open-set scenario.

However, these previous works miss one important aspect in practical applications: cross-device product authentication, i.e. there are distinct smartphone devices used for training/enrollment and classification/authentication. While the cross-device scenario is not a complication for most applications such as biometric recognition or object detection, there is some evidence that the surface material classification is degraded in a cross-device scenario, e.g. in [11].

In our previous works we encountered this performance degradation of surface texture classification accuracy in the cross-device smartphone scenario as well. In [12] we investigated intrinsic product authentication for zircon oxide blocks (commonly used for dental ceramics) by capturing surface micro-structure images with a clip-on macro lens to and revealed problems especially if more than one type of smartphone was involved, raising concerns on the feasibility of cross-sensor material classification. As the PRNU [7] could be ruled out as a reason, these problems were attributed to a device-inherent signal consisting of a signal from the image sensor itself as well as an additional signal from the image processing tool-chain (ISP). In [13] we investigated if this problem vanishes if raw images are captured (to rule out the signal stemming from the ISP) and readily available software for minimal raw-to-rgb conversion is used. An additional denoising filter was applied to remove the PRNU signal. The results indicated that despite using minimally processed images (raw-like), cross-device classification performance was still suboptimal. This is a further evidence for a device-inherent signal, which is stronger in the ISP processed JPG images, but still present in the raw-like images. It can partly be attributed to the different physical structure of the image sensors in different smartphone models (color filter array). Hence, the hardware related part of the signal should be more similar and thus, have less influence for multiple instances of the same smartphone model. In [14] we investigated if the nature of this signal is instance/device or model specific. The results revealed that on the one hand the signal is strongly interdependent with the texture/image content and on the other hand, the results on pairs and triples of similar smartphone devices (same model) revealed that the main part of the signal is model and not instance/device specific. The final conclusion was that unrelated smartphones (different models) are the ones mainly causing problems.

This work is the continuation of our previous attempts towards smartphone-based cross-device material classification. We modernize the material classification pipelines in earlier works by employing deep learning models. Furthermore, we show that a) reordering CFA data into RGB data b) sample wise normalization and c) noise augmentation drastically improves cross-sensor classification performance for train-test pairs of unrelated devices.

3 Experimental Setup

This section describes the data used in this study, the cross-sensor evaluation protocol and model training.

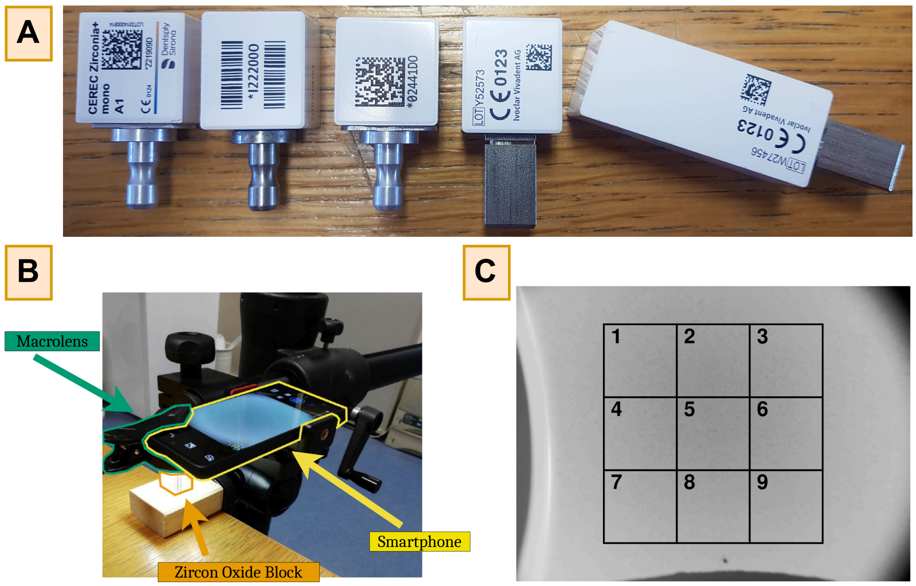

Fig. 1. A) Examples of dental ceramic blocks used as data source in this work. B) Image acquisition setup including a ceramic block, the on-clip macro lens and a fixated smartphone device. C) single acquired image showing distortions and indicating where patches are cropped.

3.1 Data

The dataset used within this study comprises images taken from top-sides of dental ceramic blocks via consumer grade smartphones. Because the smartphone camera sensors alone are currently not able to capture detailed surface texture of the material, a clip-on macro lens (Agritix WIDK-24X01 Xylorix Wood Identification Tool) was used. Images were taken at every corner of a dental ceramic block. From each image, a center piece is cropped because the macro lens introduces distortions at the image boundaries (such as pincusion and black regions where the lens did not fit the sensor exactly). Prior to cropping the center, the images were resized to a fixed dimension of 1037×1382 (assuming landscape orientation, prior 90° rotation otherwise) using bilinear interpolation to normalize the different resolutions of different smartphone models. This approach is justified because all smartphones used have the same resolution ratio and capture roughly the same field of view, which was tested using a small checkerboard pattern. Ultimately, 9 patches of size 256×256 are extracted in a 3×3 grid from every resized center crop. Thus, per ceramic block, 36 patches are available. Fig. 1 shows some exemplary dental ceramic blocks (A), depicts the data acquisition process (B) and gives an example why cropping a center piece is necessary (C). In total, the dataset comprises blocks from three manufacturers: *Ivoclar Vivadent* (M1, 16 blocks), *Dentsply Sirona* (M2, 6 blocks) and *3M* (M3, 10 blocks).

Smartphone	Abbrev.	Image Resolution
iPhone 13 1	i13-1	3024 × 4032
iPhone 13 2	i13-2	3024 × 4032
iPhone XS 1	iXS-1	3024 × 4032
iPhone XS 2	iXS-2	3024 × 4032
iPhone X	iX	3024 × 4032
iPhone 11	i11	3024 × 4032
iPhone 13 Pro	i13P	3024 × 4032
Google Pixel 4a 1	GP4-1	3024 × 4032
Google Pixel 4a 2	GP4-2	3024 × 4032
Google Pixel 4a 3	GP4-3	3024 × 4032
Google Pixel 1	GP1	3036 × 4048
LG Nexus 5 1	LGN5-1	2448 × 3264
LG Nexus 5 2	LGN5-2	2448 × 3264
Nokia 6.1 1	N61-2	3456 × 4608
Nokia 6.1 2	N61-1	3456 × 4608
Huawei P20 Lite	P20	3456 × 4608
Huawei P30 Pro	P30	2736 × 3648
Xiaomi Mi A3	Xia	3000 × 4000
Samsung Galaxy A52	SGA52	3468 × 4624

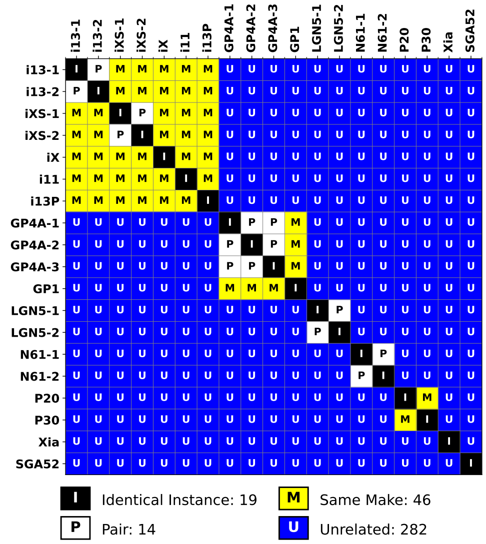

Fig. 2. (left) Overview of employed smartphone devices, their abbreviations used within this study and their default image resolution. (right) Overview of all devices used in the experiments, arranged in a grid. Rows represent the training device, while columns indicate the device used for evaluation. Each cell denotes the corresponding comparison category and the number of comparisons per setting.

Thus, for every smartphone 576, 216 and 360 ceramic patches are available per manufacturer, respectively. The data was split into a train and test set such that images from 2 fixed blocks (corresponding to 72 patches per manufacturer) were held out during training. In addition to "normal" (ISP) images, all images were captured in raw format. Since native smartphone camera applications typically restrict raw capture, specialized software was utilized: OpenCamera for Android devices and Halide Mark II for iOS. In total, data were acquired using 19 distinct smartphone instances as listed in Fig. 2 (left).

3.2 Cross-sensor evaluation protocol

In this work, we focus on investigating one-to-one cross-sensor performance, where a model is trained on the training data from a single sensor and evaluated on the test data from a different sensor. To systematically analyze generalization across devices, we define four comparison settings:

1. **Identical Instance**: The model is both trained and evaluated on data from the same device instance.
2. **Pair**: The model is evaluated on data from a similar device, i.e., the same device model but a different physical instance.
3. **Same Make**: The model is evaluated on data from a device produced by the same manufacturer (e.g., iPhone or Huawei), but of a different model.

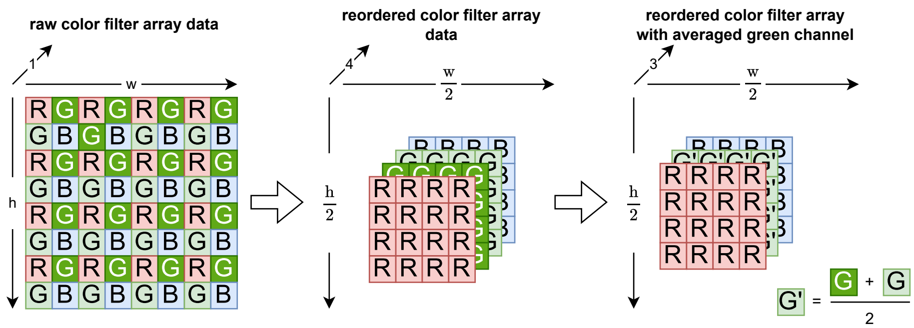

Fig. 3. Illustration of custom CFA-to-RGB re-structuring.

4. **Unrelated**: The model is evaluated on data from a device produced by a different manufacturer.

Because prior work has demonstrated better performance in paired settings, this study focuses primarily on the more challenging unrelated comparison scenario. An overview of all devices and their corresponding comparison categories is provided in Figure 2 (right), which also summarizes the number of comparisons per setting. Note that train–test configurations are not necessarily symmetric, and thus both directional combinations are evaluated.

3.3 Image Representations and Preprocessing

In addition to ISP-processed images, this work considers three alternative image representations derived directly from the raw sensor data. First, we employ a custom CFA-to-RGB conversion approach: To extract the raw CFA values from the raw file, the *rawpy* library is utilized. The single-channel CFA data is reshaped into a four-channel (RGGB) half-resolution representation. Subsequently, the two green channels are averaged, resulting in a three-channel RGB image. Figure 3 illustrates this custom CFA-to-RGB conversion process. In addition, minimal adjustments are applied using sensor-specific black level, white level saturation, and white balance parameters obtained from the image metadata. Beyond direct access to CFA data, *rawpy* provides functionality to obtain minimal-processed RGB images from the raw data. We employ two variants of this processing, which differ in how white balancing is performed. In the first variant, white balance is automatically estimated, while in the second it is derived from camera metadata. We refer to these approaches as *rawpy-autowb* and *rawpy-camerawb*, respectively.

Deep learning models typically benefit from appropriately normalized input data. Visual inspection of our dataset further suggests that normalization is beneficial, as considerable variation in color distribution and illumination is present across devices and ceramic manufacturers. Figure 4 shows representative ISP

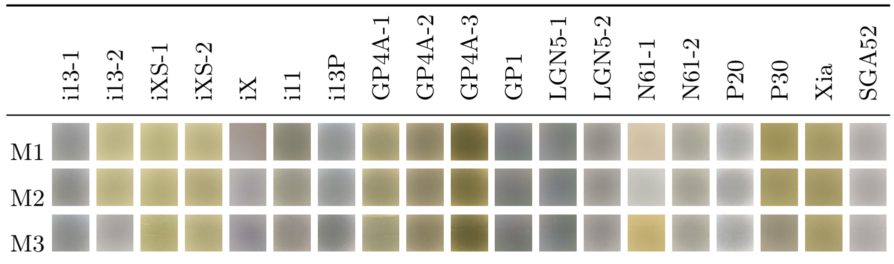

Fig. 4. ISP examples of images across all smartphone devices and manufacturers used in this work. This overview suggests a high variation in illumination and color and motivates per sample normalization.

images, highlighting these differences. To account for this variability, we employ per-sample normalization using channel-wise z-score normalization,

$$x' = \frac{x - \mu}{\sigma},$$

where μ and σ are computed per channel for each individual sample. Note that this would also be preferred for an inference-in-the-wild scenario, as dataset statistics are not necessarily available for an arbitrary sample. However, since sample-wise normalization eliminates brightness information, this strategy would fail for any material which differs solely in its brightness but not in texture. Additionally, we conduct an ablation study to assess the impact of different normalization strategies. The evaluated configurations are as follows:

- **no-norm**: No normalization applied during training and testing,
- **z-norm train-set**: Training on device A using normalization parameters computed from its training set, with the same parameters also used for the test set of device B,
- **z-norm set-wise**: Training on device A using normalization parameters from its training set, and testing using parameters computed from the test set of device B,
- **z-norm sample-wise**: Sample-wise normalization.

Results of this study are reported in Section 4.2.

3.4 Classification models and training configuration

In this study, we evaluate a range of classification models that vary in architectural complexity and number of trainable parameters. Specifically, we analyze model performance in relation to model size, with the goal of identifying a good trade-off. This aspect is particularly interesting in the context of smartphone image data, where deployment on low-resource devices systems is desirable.

Network model architectures are taken from the PyTorch Image Models [22] collection, which encompasses a large variety of state-of-the-art model architectures. In particular, ten different architectures are employed: MobileNetv4 [9], EfficientNet [17], TinyNet [4], DenseNet-121[6], ResNet-18 [5], MobileViT xs and xxs[8], FastViT t8 and s12 [20] and TinyViT 5m[23]. Models are loaded with pre-trained weights and re-trained with no frozen parameters.

We used a single set of hyperparameters for training of all models. This can be justified because the intra-device test performance (set *identical instance*) was checked to confirm that the model learns properly. We observe almost perfect scores for all configurations and models. The models were trained using the AdamW optimizer with a learning rate of 5×10^{-4} and a weight decay of 0.05. Training was performed for 25 epochs using the cross-entropy loss function. An exponential learning rate scheduler with a decay factor of $\gamma = 0.9$ was applied to adjust the learning rate over time. To address class imbalance, a `WeightedRandomSampler` was employed during training.

Preliminary experiments indicated that incorporating noise augmentation improves cross-sensor robustness. Therefore, a fixed noise augmentation strategy is applied during all training runs. In particular, random horizontal and vertical flips are applied to all samples. In addition, we employ a probabilistic noise pipeline: with 50% probability, samples are kept unchanged; with 25% probability, a denoising step is applied using `denoise_tv_chambolle` [2] from the *scikit-image* library; and with 25% probability, additional noise is added in the form of scaled uniform noise, defined as $\frac{\mathcal{U}(0,1)-0.5}{20}$ (i.e., $\pm 2.5\%$ of the dynamic range). To assess the impact of this augmentation, we additionally report results without noise augmentation in section 4.2.

4 Results and Discussion

This section presents experimental results. All experiments in this section were run five times. Therefore, reported results in tables are always averaged over 282 (number of unrelated comparisons) * 5 (runs) = 1410 values. The accompanied subscript value is the minimum over the 1410 values.

4.1 One-to-one cross-device performance using proposed changes

Table 1 shows results for the cross-sensor experiments using only *unrelated* comparisons. The mode here was sample-wise normalization and active noise augmentation. Models are divided into two blocks based on their architecture: convnets (first block) and vision transformers (second block). Within each block, rows are sorted according to number of parameters in ascending order. As expected, the general trend indicates that models with more capacity (i.e. higher number of parameters) tend to achieve higher accuracy values compared to low-capacity models. Because we are dealing with a three-class problem and a balanced test set, 33% and 66% minimum accuracy likely indicates difficulties with two or one classes, respectively, in some runs. Based on overall performance and number of trainable model parameters, the `tiny_vit_5m_224` architecture is used for further experiments

Table 1. Comparison of various classification network architectures. Main values show averaged accuracies over all unrelated pairs and 5 runs (=1410 values). Accompanying subscript value indicates the minimum cross-sensor accuracy value observed. Models are divided into two blocks based on their architecture: convnets (first block) and vision transformers (second block). Sorted according to number of parameters in ascending order.

timm name	#Params Mio.	ISP	rawpy camera-wb	rawpy auto-wb	custom raw CFA to RGB
mobilenetv4_conv_small	3.77	$0.877_{0.37}$	$0.914_{0.33}$	$0.920_{0.34}$	$0.969_{0.60}$
efficientnet_b0	5.29	$0.919_{0.40}$	$0.936_{0.34}$	$0.938_{0.37}$	$0.981_{0.70}$
tinynet_a	6.19	$0.935_{0.50}$	$0.958_{0.49}$	$0.955_{0.41}$	$0.979_{0.56}$
densenet121	7.98	$0.950_{0.47}$	$0.930_{0.33}$	$0.945_{0.33}$	$0.986_{0.66}$
resnet18	11.69	$0.908_{0.39}$	$0.906_{0.33}$	$0.907_{0.33}$	$0.979_{0.66}$
mobilevit_xxs	1.27	$0.957_{0.54}$	$0.964_{0.37}$	$0.953_{0.34}$	$0.979_{0.56}$
mobilevit_xs	2.32	$0.955_{0.49}$	$0.951_{0.33}$	$0.946_{0.33}$	$0.980_{0.59}$
fastvit_t8	4.03	$0.936_{0.35}$	$0.955_{0.35}$	$0.959_{0.36}$	$0.985_{0.67}$
tiny_vit_5m_224	5.39	$0.943_{0.44}$	$0.962_{0.49}$	$0.964_{0.42}$	$0.985_{0.67}$
fastvit_s12	9.47	$0.956_{0.56}$	$0.968_{0.35}$	$0.963_{0.36}$	$0.985_{0.66}$

For further analysis, we look at score distributions per comparison set and per image modality. In addition, we compare the results to the traditional feature based results from [14]. Fig. 5 shows the score distributions using boxplots. Shown are comparison sets *paired*, *same make* and *unrelated*, each in a different row. On the left, ISP data is shown while plots on the right depict the raw variants. In [14], the raw data was transformed to RGB data using a minimal-ISP pipeline using a software named darktable (abbreviated as DT). The results from prior work are shaded in gray. For comparison, the average value (also indicated using a green triangle per boxplot) is extended to the current results using a dashed line. The results for *paired* were already high in the prior work for both, ISP and raw representations, however, for the ISP data the average accuracy could be improved from 94.1% to 99.8%. For comparison set *same make*, accuracy values could also be measurably improved for all image modalities, especially for the raw ones with only a few outliers left. For the *unrelated* comparisons, major improvements in average accuracy from 64.3% to 94.3% for ISP and from 72.4% to 98.5% for the custom raw representation can be seen.

However, despite these improvements in accuracy for the unrelated comparisons, there is still a number of outliers present for every image modality. To further investigate these outliers, we look at the cross-sensor accuracy results on a per-pair basis. Fig. 6 shows the confusion plots for all image representations. Reported per cell is the minimum value over all five runs. For brevity, results are shown only as one digit together with coloring. Asterisks indicate 100% accuracy. From this overview, we can observe two things: First, the accuracy values increasingly approach perfect scores as we transition from ISP to the custom raw variant with the rawpy variants being somewhere in-between.

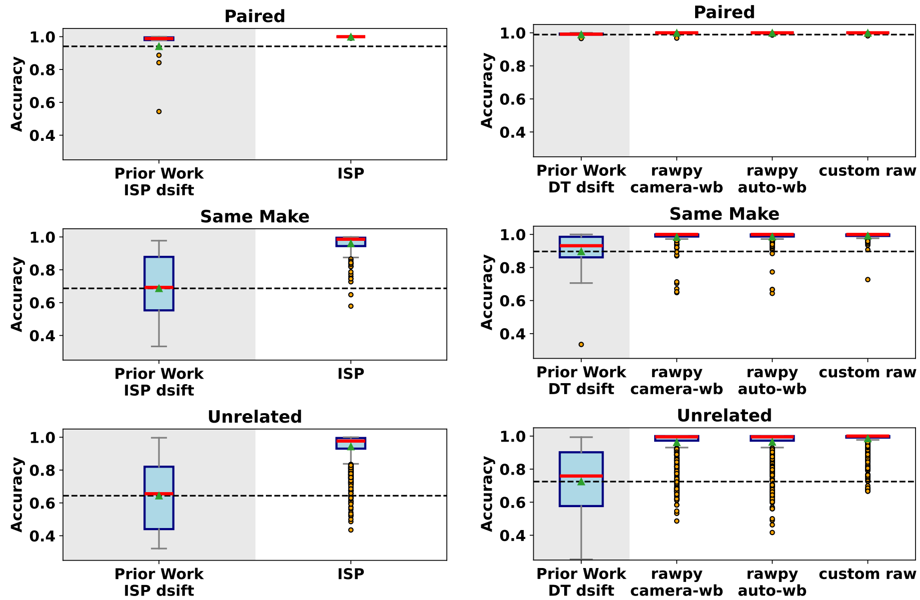

Fig. 5. Score distributions for the comparison settings *pair*, *same make*, and *unrelated*, evaluated using the TinyViT-5M model and visualized as boxplots. The left column presents results for ISP data, while the right column shows results for raw modalities. Gray shaded regions indicate scores as reported in prior work using traditional feature-based classification.

This underlines the functionality of the custom raw structure. Second, looking only on the heatmap in the lower right, i.e. for custom raw, we can see that the low values are caused mostly by one smartphone, namely the Huawei P20 Lite device. We can derive that using raw data and some aimed pre-processing steps allows one-to-one cross-sensor manufacturer classification for dental ceramics for most (unrelated) devices.

4.2 Ablation studies: Effect of sample normalization and noise augmentations

To study the effect of different normalization schemes and the effect of noise augmentation, two ablation studies are presented in this section. Table 2 presents the results for the various set normalization approaches. Compared to the other modalities, rawpy-autowb achieves quite high values regardless of the normalization used. This indicates that the automatic white balance already does a heavy per-sample normalization. For the other three columns, we can see that no normalization and normalization parameters derived from the training device leads to relatively low accuracy. While the per-set normalization is measurably higher, it appears to be still an inferior strategy compared to sample-wise normalization.

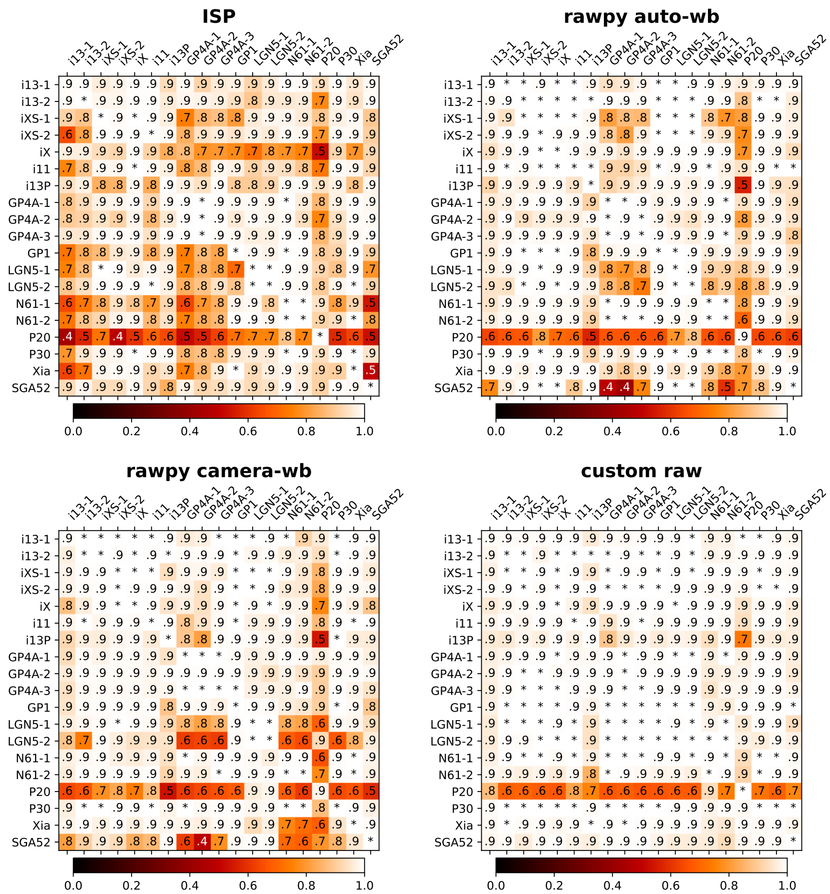

Fig. 6. Minimum scores over five runs on a per-pair cross-sensor basis shown for all four image modalities used in this study. Asterisks indicate 100% accuracy.

Table 3 shows results for omitting noise augmentation during model training. We can see that for every modality, the average accuracy is measurably higher compared to the results where noise augmentation was disabled. To validate our chosen noise strength of 2.5% of the dynamic range, we additionally tested 1% and 5%. As demonstrated in Table 3, variations around this hyperparameter yield only marginal performance differences. It can be derived that both tested improvements contribute measurably in one-to-one cross-sensor performance for unrelated device pairs.

5 Conclusion

In this work, we described a number of engineering enhancements with the goal of improving cross-device performance for dental ceramics manufacturer classi-

Table 2. Cross evaluation of normalization approaches with ISP and raw-variants. Values depict the average cross sensor accuracy on the test sets, where only unrelated pairs are considered.

Norm	ISP	rawpy camera-wb	rawpy auto-wb	custom raw CFA to RGB
no norm	$0.589_{0.16}$	$0.746_{0.31}$	$0.925_{0.41}$	$0.621_{0.15}$
z-norm train-set	$0.562_{0.17}$	$0.753_{0.26}$	$0.932_{0.40}$	$0.607_{0.12}$
z-norm set-wise	$0.884_{0.35}$	$0.922_{0.29}$	$0.959_{0.48}$	$0.924_{0.33}$
z-norm sample-wise	$0.943_{0.44}$	$0.962_{0.49}$	$0.964_{0.42}$	$0.985_{0.67}$

Table 3. Evaluation of the effect of noise augmentation (25% chance denoising, 25% chance of additional noise).

Noise Augmentation	Noise Strength	ISP	rawpy camera-wb	rawpy auto-wb	custom raw CFA to RGB
no	–	$0.886_{0.33}$	$0.943_{0.33}$	$0.951_{0.34}$	$0.965_{0.36}$
yes	± 1 %	$0.935_{0.38}$	$0.966_{0.34}$	$0.965_{0.35}$	$0.988_{0.64}$
yes	± 2.5 %	$0.943_{0.44}$	$0.962_{0.49}$	$0.964_{0.42}$	$0.985_{0.67}$
yes	± 5 %	$0.950_{0.48}$	$0.966_{0.39}$	$0.968_{0.50}$	$0.986_{0.65}$

fication. Experimental results indicate that using custom reordering of raw data combined with metadata preprocessing such as white balancing, per-sample normalization and appropriate augmentation, one-to-one cross-sensor classification is mostly possible. Future work will include validation on data from other domains where non-invasive object surface authentication via smartphone imagery is desired. Beyond that, future work will investigate robust approaches to convert ISP data back into raw format to eliminate the need for additional capturing software and extend the approach to smartphones without raw capturing capabilities.

Data availability

The data are publicly available at https://zenodo.org/records/20762122. Additionally, example code is available at https://gitlab.cosy.sbg.ac.at/wavelab/crosssurf-ceramic-code.

References

1. Cai, S., Zhao, L., Chen, C.: Open-set product authentication based on deep texture verification. In: Peng, Y., Hu, S.M., Gabbouj, M., Zhou, K., Elad, M., Xu, K. (eds.) Image and Graphics. pp. 114–125. Springer International Publishing, Cham (2021)
2. Chambolle, A.: An algorithm for total variation minimization and applications. Journal of Mathematical Imaging and Vision **20**(1), 89–97 (Jan 2004). https://doi.org/10.1023/B:JMIV.0000011325.36760.1e

3. Diephuis, M., Voloshynovskiy, S., Holotyak, T.: Sketchprint: Physical object microstructure identification using mobile phones. In: 2015 23rd European Signal Processing Conference (EUSIPCO). pp. 834–838 (2015). https://doi.org/10.1109/EUSIPCO.2015.7362500
4. Han, K., Wang, Y., Zhang, Q., Zhang, W., Xu, C., Zhang, T.: Model rubik's cube: twisting resolution, depth and width for tinynets. In: Proceedings of the 34th International Conference on Neural Information Processing Systems. NIPS '20, Curran Associates Inc., Red Hook, NY, USA (2020)
5. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE CVPR. pp. 770–778 (2016). https://doi.org/10.1109/CVPR.2016.90
6. Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (July 2017)
7. Lukas, J., Fridrich, J.J., Goljan, M.: Digital camera identification from sensor pattern noise. IEEE Transactions on Information Forensics and Security **1**(2), 205–214 (2006)
8. Mehta, S., Rastegari, M.: Mobilevit: Light-weight, general-purpose, and mobile-friendly vision transformer. In: International Conference on Learning Representations (2022)
9. Qin, D., Leichner, C., Delakis, M., Fornoni, M., Luo, S., Yang, F., Wang, W., Banbury, C., Ye, C., Akin, B., Aggarwal, V., Zhu, T., Moro, D., Howard, A.: Mobilenetv4: Universal models for the mobile ecosystem. In: Leonardis, A., Ricci, E., Roth, S., Russakovsky, O., Sattler, T., Varol, G. (eds.) Computer Vision – ECCV 2024. pp. 78–96. Springer Nature Switzerland, Cham (2025)
10. Schraml, R., Debiasi, L., Kauba, C., Uhl, A.: On the feasibility of classification-based product package authentication. In: IEEE Workshop on Information Forensics and Security (WIFS'17). p. 6. Rennes, FR (December 2017). https://doi.org/10.1109/WIFS.2017.8267659
11. Schraml, R., Debiasi, L., Uhl, A.: Real or fake: Mobile device drug packaging authentication. In: Proceedings of the 6th ACM Workshop on Information Hiding and Multimedia Security. p. 121–126. IH&MMSec '18, Association for Computing Machinery, New York, NY, USA (2018). https://doi.org/10.1145/3206004.3206016
12. Schuiki, J., Kauba, C., Hofbauer, H., Uhl, A.: Cross-sensor micro-texture material classification and smartphone acquisition do not go well together. In: Proceedings of the 11th International Workshop on Biometrics and Forensics (IWBF'23). pp. 1–6. Barcelona, Spain (2023). https://doi.org/10.1109/IWBF57495.2023.10157739
13. Schuiki, J., Kauba, C., Hofbauer, H., Uhl, A.: Limiting factors in smartphone-based cross-sensor microstructure material classification. In: 22th International Workshop on Digital-forensics and Watermarking (IWDW'23). Springer LNCS, vol. 14511, pp. 33–47. Jinan, China (2023). https://doi.org/10.1007/978-981-97-2585-4_3
14. Schuiki, J., Kauba, C., Hofbauer, H., Uhl, A.: On the influence of texture strength and camera signal on micro-texture surface classification. In: Proceedings of the 12th International Workshop on Biometrics and Forensics (IWBF'24). pp. 1–6. Enschede, Netherlands (2024). https://doi.org/10.1109/IWBF62628.2024.10593842
15. Sun, B., Tan, S.R.X., Ren, Z., Chan, M.C., Han, J.: On utilizing smartphone cameras to detect counterfeit liquid food products. In: Proceedings of the 20th Annual International Conference on Mobile Systems, Applications and Services. p. 551–552. MobiSys '22, Association for Computing Machinery, New York, NY, USA (2022). https://doi.org/10.1145/3498361.3538779

16. Sun, Y., Liao, X., Liu, J.: An efficient paper anti-counterfeiting method based on microstructure orientation estimation. In: ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 2525–2529 (2021). https://doi.org/10.1109/ICASSP39728.2021.9415114
17. Tan, M., Le, Q.: EfficientNet: Rethinking model scaling for convolutional neural networks. In: Chaudhuri, K., Salakhutdinov, R. (eds.) Proceedings of the 36th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 97, pp. 6105–6114. PMLR (06 2019)
18. Taran, O., Tutt, J., Holotyak, T., Chaban, R., Bonev, S., Voloshynovskiy, S.: Mobile authentication of copy detection patterns: how critical is to know fakes? In: 2021 IEEE International Workshop on Information Forensics and Security (WIFS). pp. 1–6 (2021). https://doi.org/10.1109/WIFS53200.2021.9648398
19. Taran, O., Tutt, J., Holotyak, T., Chaban, R., Bonev, S., Voloshynovskiy, S.: Mobile authentication of copy detection patterns. EURASIP J Inf Secur **2023**(1), 4 (Jun 2023)
20. Vasu, P.K.A., Gabriel, J., Zhu, J., Tuzel, O., Ranjan, A.: Fastvit: A fast hybrid vision transformer using structural reparameterization. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 5785–5795 (October 2023)
21. Voloshynovskiy, S., Diephuis, M., Beekhof, F., Koval, O., Keel, B.: Towards reproducible results in authentication based on physical non-cloneable functions: The forensic authentication microstructure optical set (famos). In: 2012 IEEE International Workshop on Information Forensics and Security (WIFS). pp. 43–48 (2012). https://doi.org/10.1109/WIFS.2012.6412623
22. Wightman, R.: Pytorch image models. https://github.com/rwightman/pytorch-image-models (2019)
23. Wu, K., Zhang, J., Peng, H., Liu, M., Xiao, B., Fu, J., Yuan, L.: Tinyvit: Fast pretraining distillation for small vision transformers. In: Computer Vision – ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXI. p. 68–85. Springer-Verlag, Berlin, Heidelberg (2022). https://doi.org/10.1007/978-3-031-19803-8_5
24. Yan, Y., Zou, Z., Xie, H., Gao, Y., Zheng, L.: An IoT-based anti-counterfeiting system using visual features on QR code. IEEE Internet of Things Journal **8**(8), 6789–6799 (2021). https://doi.org/10.1109/JIOT.2020.3035697