

A CONTROLLED DATASET OF PAIRED VISIBLE AND 365 NM ULTRAVIOLET 500 SAUDI RIYAL BANKNOTE IMAGES WITH SECURITY FEATURE ANNOTATIONS

Engr. Umar Said^{*1}, Sadiq Ali², Daniyal Ahmed Khan³, Muhammad Amin⁴, Shayan Riaz⁵,
Malik Muhammad Arif⁶

^{*1}Department of Artificial Intelligence, University of Engineering and Technology Peshawar, Pakistan

^{2,3}Department of Electrical Engineering, University of Engineering and Technology Peshawar, Pakistan

^{4,5,6}Department of Electrical Engineering, University of Engineering and Technology Peshawar, Pakistan

^{*1}umar.said.si@fia.gov.pk, ²sadiqali@uetpeshawar.edu.pk ³17pwmec4340@uetpeshawar.edu.pk

⁴muhammadamin0334@gmail.com, ⁵shayanriaz@uetpeshawar.edu.pk, ⁶engrmlikarif@uetpeshawar.edu.pk

DOI: <https://doi.org/10.5281/zenodo.22805707>

Keywords

banknote image dataset;
document forensics; ultraviolet
imaging; visible-ultraviolet pairs;
security-feature annotation; object
detection; controlled-access data;
reproducibility.

Article History

Received: 01 June 2026

Accepted: 23 August 2026

Published: 17 September 2026

Copyright @Author

Corresponding Author: *

Engr. Umar Said

Abstract

Reusable evaluation of automated secure-document analysis requires datasets that preserve imaging modality, feature semantics, spatial ground truth, and source lineage. This article presents a controlled resource of paired visible-light and 365 nm ultraviolet images of genuine and counterfeit 500 Saudi Riyal banknotes. Reconstruction of the archived YOLO-ready package identified 524 balanced filename-derived source keys (262 genuine and 262 counterfeit), each containing one original visible image and one original ultraviolet image, for 1,048 original modality images. The historical release contains 3,143 image-label pairs after augmentation; one expected visible descendant is absent, but its source key remains paired at the original-image level. A corrected experimental root was rebuilt by assigning source keys before augmentation: 420 keys were used for training, 52 for validation, and 52 for testing. The training set comprised 1 original and 2 images released images per source key and modality, while the validation set and test set consisted of 2,728 images of the original images. There are sixteen visible classes (eight expected features and eight replica classes), and four ultraviolet classes (two expected responses and two counterfeit responses). Valid class identifiers were found and found to be correct at the time of the audit, one-to-one image-label pairing was found and found to be correct at the time of the audit, there was no overlap of any source-key, augmentation locking was found and found to be correct at the time of the audit, and there were no exact-image duplication warnings found for any image with a cross-split except for 443 warnings that were retained for transparent review through perceptual-hash screening. Only annotation-leakability baselines were used with separate YOV8m detectors. On the held-out test set, VISIBLE16 achieved precision 0.9907, recall 0.9977, mAP50 0.9917, and mAP50-95 0.7958; UV4 achieved precision 0.9934, recall

1.0000, mAP50 0.9950, and mAP50-95 0.7829. The resource supports controlled research on localized security-feature detection; it does not establish operational forensic validity, physical-note independence beyond the recoverable source keys, or generalization to other denominations and document families.

INTRODUCTION

The examination of banknotes and other security documents is grounded in observable devices and responses rather than in appearance alone. The specific set of observable devices varies by document type but typically spans holographic elements, optically variable ink, watermark windows, security threads, microprinting, raised or intaglio printing, serial identifiers, and regions that respond under ultraviolet light. International specifications for machine-readable travel documents similarly treat visual and non-visible inspection as complementary parts of security evaluation [1]. Saudi banknotes also publish denomination-specific features for public verification [2]. These practices motivate machine-learning resources in which the unit of supervision is a named security feature with a recoverable image location.

Many banknote-learning studies instead assign a single denomination or authenticity label to a complete visible-light image. Such data are useful for recognition, but they do not support an audit of what local feature was observed, where it was observed, or whether an apparent replica was explicitly represented in the ground truth. Reviews of camera-based document analysis and banknote recognition have repeatedly identified sensitivity to sensor, acquisition, and layout conditions [3], [4]. Spectral and hyperspectral work shows that information outside ordinary visible imaging can be materially discriminative [5], [6]. A dataset intended for localized examination must therefore preserve modality, image lineage, feature class, and spatial annotation together. More broadly, recent multimodal AI research has emphasized the importance of integrating heterogeneous information while retaining interpretable evidence for decision support [7].

Paired visible and ultraviolet acquisition is especially relevant because a visually plausible imitation can behave differently under ultraviolet excitation. Existing smartphone studies demonstrate accessible visible-image classification [8], while ultraviolet and multimodal studies have evaluated region-proposal, detector, handcrafted, and fluorescence-based approaches [9, 10, 11]. Human-factors evidence also indicates that directing attention to named features can improve banknote authentication behavior [12]. These lines of work establish the importance of the sensing and feature-localization problem, but they do not by themselves provide a controlled paired resource that separately labels expected features and counterfeit counterparts.

This paper focuses on the data-resource aspect of this issue, through the case study of 500 bank notes of the Saudi Riyal. It does not aim to be comprehensive like an authentication system. It contains details for recoverable source units, acquisition evidence, visible and ultraviolet taxonomies, annotations in YOLO format, release reconstruction, grouped evaluation protocol, statistical characterization, integrity findings and controlled-access conditions. This article does not discuss specimen-level evidence fusion, GLOBAL20 representations, DINOv2 embeddings, downstream authenticity classification, calibration and examiner decision reports, which are all separate modeling efforts.

The contribution is fourfold. First, the resource pairs visible and 365 nm ultraviolet images through balanced filename-derived source keys. Second, it provides a 20-class modality-aware taxonomy that distinguishes eight expected visible features from eight replica counterparts and distinguishes two expected ultraviolet responses from two counterfeit responses. Third, it provides a verifiable lineage protocol which divorces the

historical 3,143-file release from a corrected 2,728-image group-split evaluation root, containing training-only descendants. Fourth, it provides validation of the datasets: Class Count, Box Geometry, Image-Quality Descriptors, Split Audits, Exact and Perceptual Duplicate Screening, and YOLO Modality-Specific Baselines. These contributions allow the paper to be independently evaluated as a dataset article instead of a replication of the two authentication architectures separately presented.

Related Work And Dataset Positioning

Early camera-based and banknote-recognition systems relied on color, texture, edges, keypoints, and sensor-specific descriptors [3], [4]. Modern object detectors, beginning with the unified YOLO formulation [13], made it practical to learn localized visual categories from bounding boxes. Nevertheless, detector capability does not remove the need for domain-appropriate annotations. In secure-document analysis, an unlabeled ultraviolet anomaly or a single whole-image label cannot identify which device produced the decision.

Visible-light banknote classification has been studied under smartphone acquisition [8], and ultraviolet imaging has been used in counterfeit-

money detection [9], [10]. Multimodal fluorescence systems further demonstrate that material response can complement ordinary appearance [11]. Hyperspectral approaches provide richer spectral sampling [5], [6], but such instrumentation and data structures differ from the paired smartphone-visible and 365 nm ultraviolet setting considered here. The present resource does not claim spectral calibration; it instead provides a reproducible two-modality benchmark under its documented controlled domain.

Document datasets such as MIDV-2020 emphasize identity-document analysis, layout, and recognition [14], whereas SIDTD provides synthetic fraud variants for identity and travel documents [15]. Historical-document dataset surveys address recognition and transcription rather than genuine-versus-replica security devices [16]. The resources are relevant in their respective contexts but cannot be matched with the taxonomy of banknote labels used in this site. Since no one modality, local annotation, or scope of authenticity is inherently superior to the others, the current resource is placed in the context of its modality, local annotation, and scope of authenticity in Table I.

TABLE I. Positioning against representative banknote and document resources

Resource	Modality	Localization	Authenticity scope	Dataset distinction
Pham et al. [8]	Smartphone visible	Whole-note	Genuine/counterfeit classification	No paired UV or named local feature labels
Chowdhury et al. [9]	UV	Study-specific regions	Fake-money detection	No paired visible branch or released 20-class taxonomy
Chhabra and Goyal [10]	Visible and UV	Detected security features	Two-stage banknote decision	System study rather than this paired source-key resource
Taheri et al. [11]	Multimodal fluorescence	Material/security features	Document authentication	Different sensing materials and study design
MIDV-2020 [14]	Visible	Layout and text regions	Identity-document analysis	Not counterfeit banknote feature annotation
SIDTD	Synthetic	Document/text	Synthetic fraud	Does not contain physical

Resource	Modality	Localization	Authenticity scope	Dataset distinction
[15]	visible IDs	manipulation		visible/UV banknote pairs
This work	Visible plus 365 nm UV	20 localized feature classes	Expected and replica/counterfeit responses	Controlled 500 SAR domain with restricted access

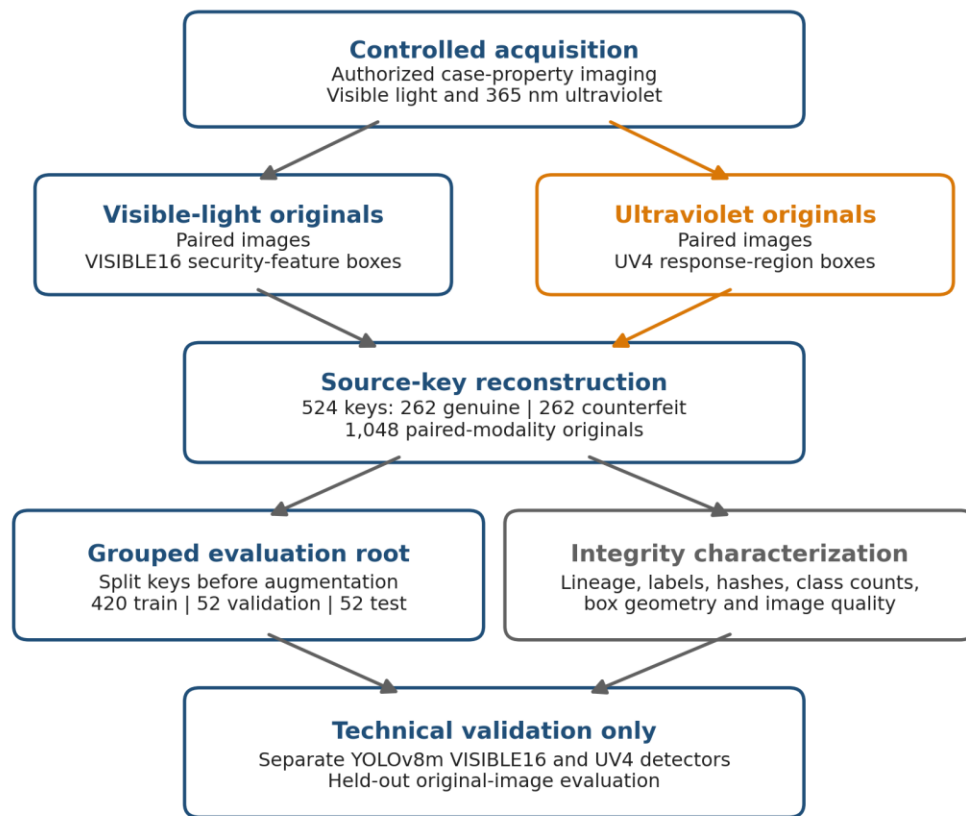
Materials And Methods

A. Study Scope And Evidentiary Provenance

The empirical case study concerns the 500 Saudi Riyal denomination. Genuine images represent authentic reference notes, while counterfeit images originate from notes held as case property in counterfeit-currency matters registered with the Federal Investigation Agency of Pakistan. The corresponding author is an Investigation Officer. Access for imaging was obtained through the case-property custodian, the Moharar Malkhana. No case number, accused-person information, seizure narrative, investigation identifier, or specimen-level legal metadata is included in the research package. The counterfeit branch therefore represents real case-property material

rather than synthetic printouts, while the published resource remains detached from operational case adjudication.

The available notebook inputs do not contain an independent physical-note registry. Accordingly, this paper does not describe the 524 recoverable keys as a re-verified count of physically independent banknotes. A source key is the pairing and split unit inferred from the authenticity prefix and base identifier in the archived filenames. This definition is used consistently in the abstract, tables, figures, and results. It supports reproducible lineage within the released files, but it cannot replace a contemporaneous physical-specimen register.



Resource boundary: no GLOBAL20, DINOv2, specimen classifier, or examiner decision report

Fig. 1. Dataset construction, audit, and technical-validation workflow. The boundary at the bottom separates this dataset article from downstream modeling work.

Imaging Evidence And Processing Resolution

Images were reportedly captured with the rear camera of an iPhone 11 Pro Max mounted on a stand. Visible-light and 365 nm ultraviolet acquisitions were performed with a fixed working arrangement, and the ultraviolet source was supported separately to reduce hand motion and large pose changes. The reported native capture resolution was 4032 x 3024 pixels. Apple documents a 12-megapixel wide camera and standard exposure, framing, and focus controls for the device family [17], [18]. These specifications support the device description, but they do not reconstruct settings that were not preserved in the original capture logs.

Not all EXIF information, ultraviolet irradiance, radiometric calibration, camera-to-document distance, lamp-to-document distance, shutter speed, ISO, white balance, or AE/AF-lock state were maintained across the images in this archived notebook. These amounts are not therefore imputed. The experimental images included in the audited outputs were all stored in portrait format (1,350 x 1,800 pixels) and rotated variants (1,800 x 1,350 pixels) were all standardized to a maximum working 4 side of 1,800 pixels, which was the YOLO-ready and corrected format. Throughout the manuscript the difference between native and processed learning resolution in native capture is preserved.

C. Release Reconstruction And Experimental Units

The historical archive contains 3,143 image-label pairs: 1,571 visible and 1,572 ultraviolet. A canonical parser removed file extensions and augmentation suffixes, interpreted rn and ru as genuine visible and ultraviolet images, interpreted fn and fu as counterfeit visible and ultraviolet images, and combined authenticity with the numerical base identifier. Parsing succeeded for all 3,143 image records. The procedure recovered 524 paired source keys, balanced between 262 genuine and 262 counterfeit keys. Every key contained exactly one

original visible image and one original ultraviolet image, yielding 1,048 original modality images.

The historical release was built around 1 original and 2 descendants per source key and modality, to generate 3,144 files. The audit found 3,143 files, since there are only two files in the visible branch for counterfeit source key 00066 instead of three. The original members of the visible and ultraviolet groups are still left, so paired-source reconstruction is not affected. Some keys had some originals and descendants in separate release partitions in the historical release partition. For evaluation, the archive was therefore treated as an input resource rather than as a valid experimental split.

TABLE II. Audited dataset cardinality and experimental units

Data layer	Count	Interpretation
Historical YOLO-ready release	3,143 image-label pairs	1,571 visible; 1,572 UV; contains originals and descendants
Recovered pairing layer	524 source keys	262 genuine; 262 counterfeit; filename-derived
Original modality layer	1,048 images	One visible and one UV original per source key
Corrected training branch	2,520 images	420 keys x 2 modalities x 3 variants
Corrected validation branch	104 images	52 keys x 2 original modalities
Corrected test branch	104 images	52 keys x 2 original modalities
Corrected evaluation root	2,728 images	All descendants restricted to the training split

A source key is a filename-derived pairing and split unit; it is not presented as an independently re-verified physical-note identifier.

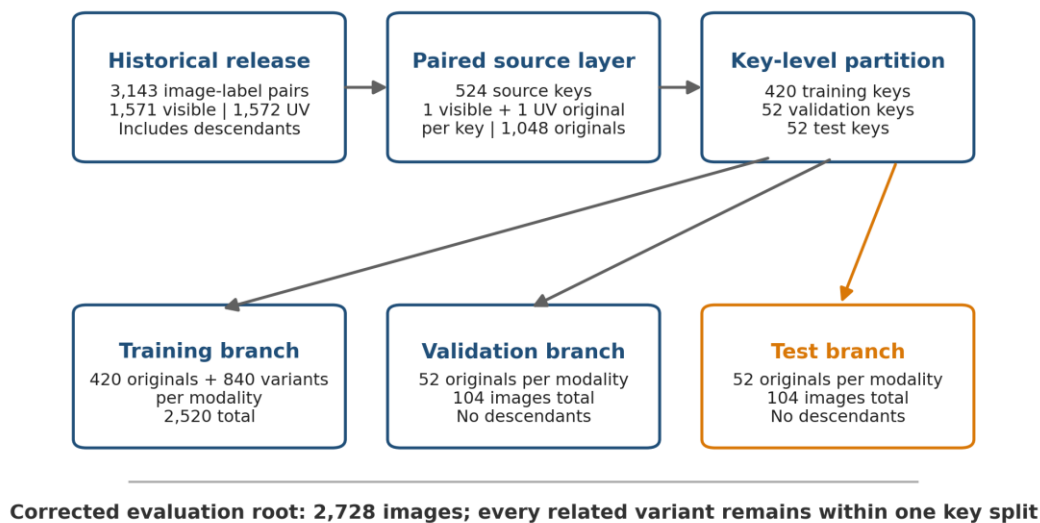


Fig. 2. Separation of the historical release, canonical paired layer, and corrected group-split evaluation root.

D. Security Feature Taxonomy And Annotation

Labelling was used to place YOLO-format bounding boxes around named security-feature regions [19]. Each annotation record contains the class identifier and normalized x center, y center, width, and height. The visible detector vocabulary contains sixteen classes: eight expected features and eight semantically paired replica counterparts. The ultraviolet vocabulary contains four classes: two expected responses and two counterfeit-response regions. Visible and ultraviolet identifiers are local to their respective detector configurations; the combined conceptual taxonomy contains twenty semantic classes.

The design clearly shows the fake representations as such, as opposed to assuming that they are just background. The distinction is significant if the replica is in a plausible location but has different visual or ultraviolet characteristics. Valid class ranges and normalized coordinates were checked for bounding boxes. Boxes were augmented with the image transformations and, during the augmentation, clipped to the image boundary and filtered if the visible area was below the minimum configured value of 0.20. No fully completed second annotator's study is provided and inter-annotator class agreement and box IoU are therefore reported as a limitation.

TABLE III. VISIBLE16 and UV4 taxonomy in the corrected group-split root

Modality	ID	Class name	Semantic type	Instances
Visible	0	OVI feature	Expected/genuine	661
Visible	1	Watermark window	Expected/genuine	646
Visible	2	Security thread	Expected/genuine	679
Visible	3	Microtext band	Expected/genuine	668

Modality	ID	Class name	Semantic type	Instances
Visible	4	Hologram patch	Expected/genuine	654
Visible	5	Portrait area	Expected/genuine	680
Visible	6	Raised printing area	Expected/genuine	684
Visible	7	Serial number	Expected/genuine	681
Visible	8	Replica OVI feature	Replica/counterfeit response	639
Visible	9	Replica watermark window	Replica/counterfeit response	632
Visible	10	Replica security thread	Replica/counterfeit response	667
Visible	11	Replica microtext band	Replica/counterfeit response	653
Visible	12	Replica hologram patch	Replica/counterfeit response	601
Visible	13	Replica portrait area	Replica/counterfeit response	670
Visible	14	Replica raised printing area	Replica/counterfeit response	678
Visible	15	Replica serial number	Replica/counterfeit response	667
UV	0	UV feature region 1	Expected/genuine	686
UV	1	UV feature region 2	Expected/genuine	678
UV	2	Counterfeit UV feature region 1	Replica/counterfeit response	683
UV	3	Counterfeit UV feature region 2	Replica/counterfeit response	681

Counts include training originals and descendants plus original-only validation and test images.



Fig. 3. Project annotation examples supplied in the original manuscript: expected visible features, expected ultraviolet responses, replica visible features, and counterfeit ultraviolet responses.

E. Grouped Partitioning And Training Only Augmentation

Source keys were split before construction of the modeling root. A deterministic seed of 42 and stratification by authenticity produced 420 training keys (210 genuine and 210 counterfeit), 52 validation keys (26 and 26), and 52 test keys (26 and 26). All files derived from one key were locked to the same split. Validation and test retained only the paired original images. Training retained one original and two released descendants for each modality and key, producing 1,260 visible and 1,260 ultraviolet training images.

The notebook supports reuse, repair, and regeneration modes. The executed package used

reuse_release, preserving the available released descendants rather than silently generating a new dataset. Table IV records the explicit augmentation policy encoded in the notebook. Visible images received conservative photometric perturbations without an additional geometric stage in the explicit dataset transform. The ultraviolet branch used controlled geometric and photometric transformations but omitted hue and saturation changes to avoid deliberate recoloring of fluorescence semantics. Detector training also retained the Ultralytics run defaults recorded in the training log; these are distinct from the dataset descendants described here.

TABLE IV. Explicit dataset augmentation policy encoded with Albumentations [20]

Branch	Transformation	Configured range	Probability
Visible	HSV	hue +/-8, saturation +/-12, value +/-12	0.30
Visible	Gaussian noise	sigma 5-25 pixel levels	0.20
Visible	Motion blur	kernel limit 3	0.15
Visible	CLAHE	library default parameters	0.10
UV	Horizontal flip	paired box propagation	0.40
UV	Random 90-degree rotation	paired box propagation	0.20
UV	Affine	scale 0.92-1.08; translation +/-0.04; rotation +/-6 degrees; shear +/-5 degrees	0.35
UV	Brightness/contrast	brightness -0.15 to 0.25; contrast -0.10 to 0.20	0.45
UV	Gaussian noise	sigma 5-20 pixel levels	0.20
UV	Motion blur	kernel limit 3	0.12
UV	CLAHE	library default parameters	0.08

Bounding boxes used YOLO coordinates, clipping, and minimum visibility 0.20. The executed package reused released descendants.

F. Integrity And Statistical Characterization

The audit checked folder structure, one-to-one image-label correspondence, parser validity, visible-ultraviolet pairing, class ranges, normalized box coordinates, source-key overlap, augmentation locking, and cryptographic image hashes. MD5 was used to detect byte-identical images across partitions. Difference hashing with Hamming distance at most three was used only as a sensitive warning screen for perceptual similarity; it was not treated as proof that two files represented the same physical note. This distinction matters in a fixed-layout banknote dataset, where different notes can share nearly identical low-frequency appearance.

Class counts and box geometry were recomputed from the corrected group-split label records.

Pearson chi-square tests assessed departure from equal class counts within each modality and association between split and class. Wilson intervals were used in the archived resource characterization because they remain well behaved for binomial proportions [21]. OpenCV routines supported grayscale conversion and image-quality calculations [22]. Image brightness was defined as the mean 8-bit grayscale value, contrast as grayscale standard deviation, and sharpness as Laplacian variance. These measures characterize the processed images; they are not substitutes for radiometric or forensic image-quality validation.

G. Technical Validation Protocol

Separate VISIBLE16 and UV4 detectors were trained as dataset-validation baselines using YOLOv8m [23]. The purpose was to test whether the supplied annotations could be learned under the corrected grouping policy, not to establish a final authenticity classifier. Models were initialized from the pretrained checkpoint and trained for 30 epochs at 640 x 640 input size with batch size 16, cosine learning-rate scheduling, patience 20, seed 42, deterministic mode, and four configured workers. The optimizer was set to auto and resolved to AdamW with learning rate 0.00125 and momentum 0.9 in the recorded run. Training used an NVIDIA Tesla T4, Python

3.13.15, PyTorch 2.11.0+cu128, and Ultralytics 8.4.133.

Evaluation was performed separately on the original-only validation and test partitions. Precision, recall, average precision at intersection-over-union 0.50 (AP50), and COCO-style average precision averaged over thresholds 0.50 to 0.95 (AP50-95) were retained [24]. Aggregate mAP values are macro-averages over detector classes. Per-class AP was collected from the same trained checkpoints. Because the package contains a single principal seed for this dataset article, the reported detector results are point estimates rather than confidence intervals over repeated training runs.

TABLE V. Baseline detector configuration

Item	Recorded configuration
Detector	YOLOv8m; separate VISIBLE16 and UV4 runs
Input size	640 x 640
Epochs / batch	30 / 16
Initialization	Ultralytics pretrained YOLOv8m checkpoint
Optimizer	Auto; resolved to AdamW, learning rate 0.00125, momentum 0.9
Schedule	Cosine learning rate; patience 20
Reproducibility	Seed 42; deterministic mode; four configured workers
Runtime	Tesla T4; Python 3.13.15; PyTorch 2.11.0+cu128; Ultralytics 8.4.133

Results**A. Dataset Reconstruction And Split Integrity**

All 3,143 historical image filenames were parsed successfully, and every recovered source key contained an original member in both modalities. The corrected root contained 2,728 images and corresponding labels. Its training branch contained 2,520 images, while validation and test each contained 104 originals. No augmented images were present in validation or test. Source-key intersections for train-validation, train-test, and validation-test were all zero.

Of the 2,728 images, 2,596 different MD5 values were found throughout the corrected root. There

were no cross-split boundaries in terms of cross byte-identical image and the 129 repeated hash groups were restricted to their assigned split. Difference hashing found 2609 close pairs (443 of which were cross-split warnings) at Hamming distance 3 or less. The warning, however, is not repeated in the case of a common background; this is not because it creates a stronger claim of independence of the physical notes, but because of the expected visual similarity. The passed critical checks and the caution retained are reported in Table VI.

TABLE VI. Integrity audit of the reconstructed and corrected dataset

Audit item	Outcome	Interpretation
Historical image filenames parsed	3,143 / 3,143	Pass
Paired source keys	524 / 524	Pass
Authenticity balance	262 genuine / 262 counterfeit	Pass
Corrected image-label records	2,728 / 2,728	Pass
Validation/test descendants	0	Pass
Cross-split source-key overlap	0	Pass
Out-of-range class identifiers	0	Pass
Cross-split exact MD5 duplicates	0	Pass
Repeated MD5 groups within assigned splits	129	Retained; no cross-split occurrence
dHash close pairs	2,609 total; 443 cross-split	Warning for review, not proof of identity
Independent physical-note register	Unavailable in notebook inputs	Limitation

Annotation Distribution And Geometry

The corrected root contained 10,560 visible boxes and 2,728 ultraviolet boxes. Visible class counts ranged from 601 to 684; ultraviolet counts ranged from 678 to 686. Equal-count chi-square tests did not show strong class imbalance for VISIBLE16 (chi-square = 11.139, $p = 0.743$)

or UV4 (chi-square = 0.050, $p = 0.997$). Split-by-class tests likewise showed no detectable distribution shift for visible labels (chi-square = 1.039, p approximately 1.000) or ultraviolet labels (chi-square = 0.004, p greater than 0.999). These tests describe label counts and do not establish population representativeness.

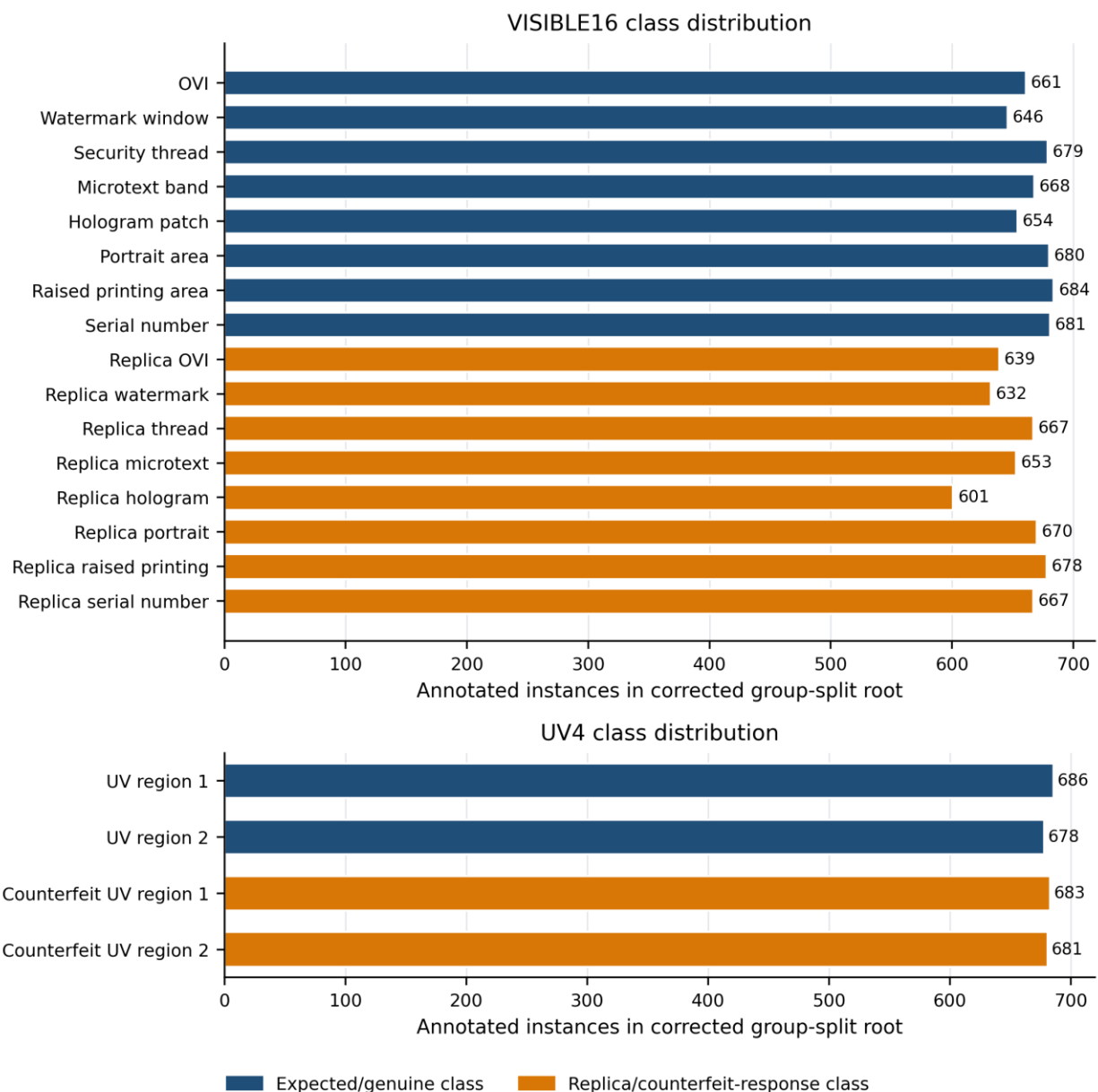


Fig. 4. Corrected group-split class-instance distributions. Expected and counterpart classes are shown separately for visible and ultraviolet branches.

Visible boxes occupied a median normalized area of 0.01999 (5th-95th percentile: 0.00771-0.05397), with median aspect ratio 1.548. Ultraviolet boxes were somewhat larger, with median area 0.02348 (0.01487-0.04078) and median aspect ratio 1.756. The center distribution reflects the fixed security-feature layout of the denomination rather than generic object locations. Consequently, high localization

scores must be interpreted within the controlled layout and acquisition domain.

C. Processed Image Characteristics

Processed visible images were brighter and had higher grayscale contrast and Laplacian sharpness than ultraviolet images, as expected from the two illumination regimes. Training distributions were wider because they include descendants;

validation and test contain originals only. These summary statistics are useful for detecting gross distribution drift in later extensions, but they do

not measure spectral fidelity, microtext readability, or forensic sufficiency.

TABLE VII. Processed image-quality descriptors in the corrected root

Modality	Images	Brightness mean +/- SD	Contrast mean +/- SD	Median sharpness	Dark fraction	Bright fraction
Visible	1364	117.73 +/- 21.29	69.16 +/- 5.48	1495.5	0.0114	0.0241
Uv	1364	69.16 +/- 18.99	43.28 +/- 6.22	565.7	0.0137	0.0007

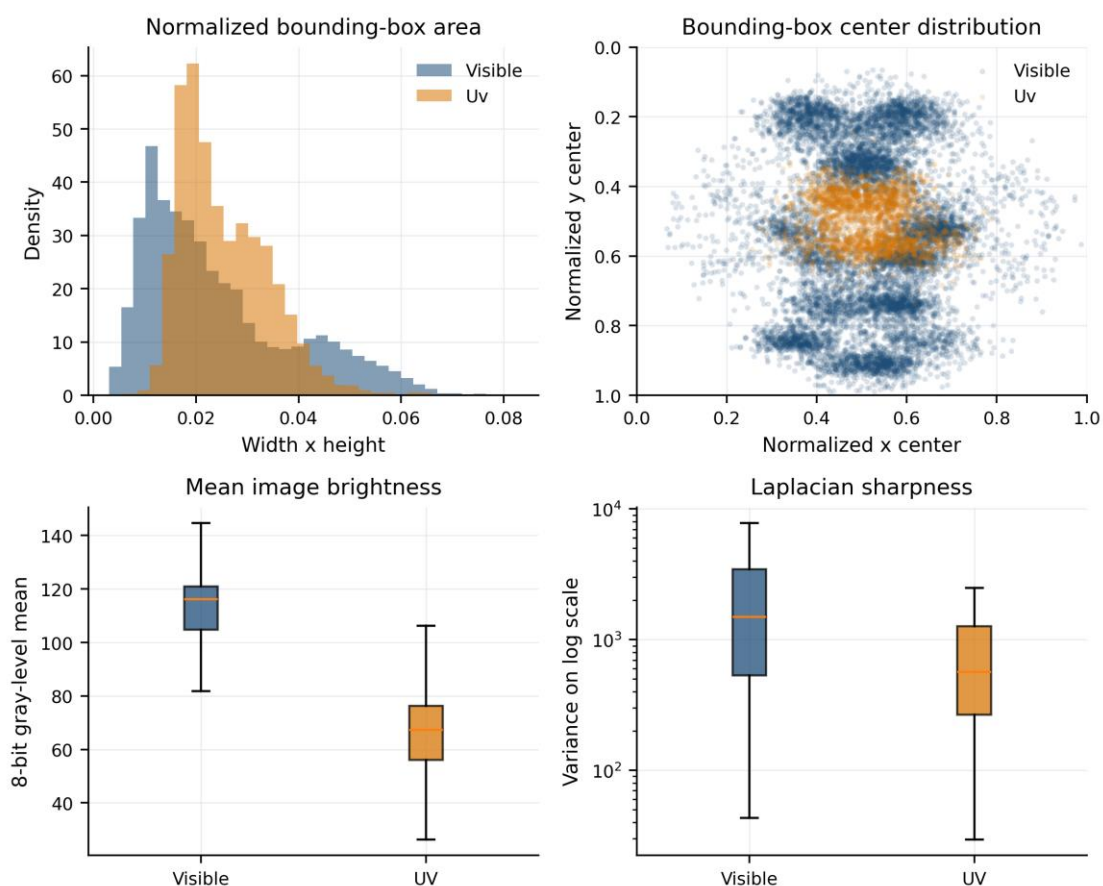


Fig. 5. Spatial annotation and processed image-quality characterization. Sharpness is displayed on a logarithmic scale because augmentation produced a long-tailed distribution.

D. Baseline Annotation Learnability

Both modality-specific detectors learned the named feature regions under the corrected split. On held-out originals, VISIBLE16 reached mAP50 0.9917 and mAP50-95 0.7958; UV4 reached 0.9950 and 0.7829. The difference between mAP50 and mAP50-95 is consistent

with less sharply bounded regions such as fluorescence, holographic zones, watermark windows, raised printing, and thin serial or microtext regions. High AP50 is evidence that the labels are learnable in this domain; it is not evidence of field robustness, cross-denomination validity, or authenticity certification.

TABLE VIII. YOLOv8m technical-validation results

Dataset	Split	Precision	Recall	mAP50	mAP50-95	Parameters
VISIBLE16	Val	0.9932	0.9998	0.9950	0.8036	25,849,024
VISIBLE16	Test	0.9907	0.9977	0.9917	0.7958	25,849,024
UV4	Val	0.9927	1.0000	0.9950	0.7893	25,842,076
UV4	Test	0.9934	1.0000	0.9950	0.7829	25,842,076

Per-class AP50 was 0.995 for fourteen of the sixteen visible test classes; security thread and microtext band were 0.965 and 0.972, respectively. At the stricter AP50-95 measure, visible classes ranged from 0.643 for replica serial number to 0.868 for replica raised printing. All

four ultraviolet classes achieved AP50 0.995, while AP50-95 ranged from 0.664 for counterfeit UV region 1 to 0.845 for expected UV region 1. These differences identify boundary-localization targets for future annotation refinement.

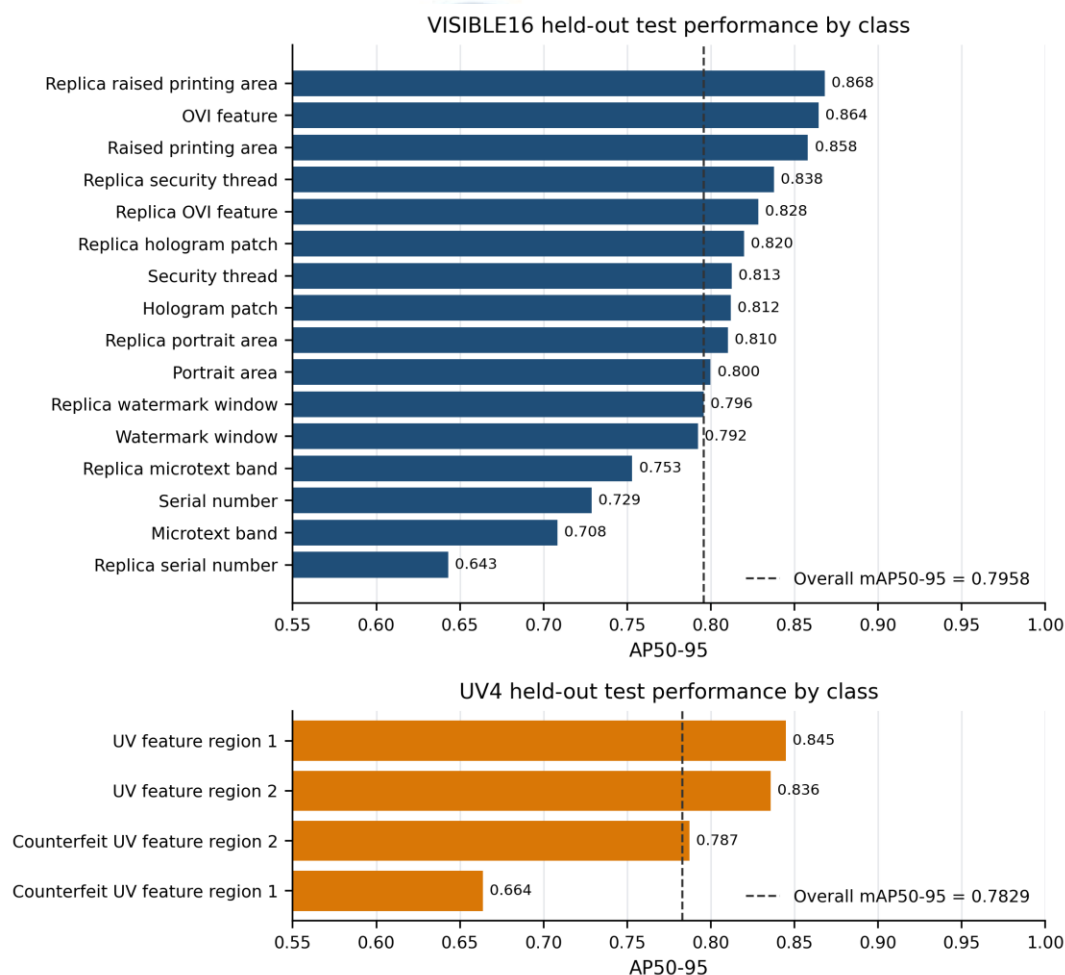


Fig. 6. Held-out test AP50-95 by class for VISIBLE16 and UV4. Dashed lines show the corresponding overall mAP50-95.

Discussion**A. Resource Contribution**

The principal contribution is a recoverable, modality-aware annotation resource rather than a new detector. The value of the release lies in the conjunction of paired visible and ultraviolet originals, explicit expected-versus-counterpart feature semantics, normalized spatial labels, lineage reconstruction, training-only descendants, and transparent integrity warnings. A whole-note authenticity label cannot represent these properties. The taxonomy allows later systems to analyze errors by feature family and modality, while the source-key structure makes it possible to prevent known descendants from crossing the evaluation boundary.

The paper also resolves a cardinality problem that could otherwise inflate apparent sample size. The historical release count of 3,143 is an image-label count, not a count of independent sources. The 1,048 original images are two modality members of 524 filename-derived keys. The corrected 2,728-image evaluation root includes descendants only for the 420 training keys. Reporting these layers separately prevents augmentation, modality, and source units from being treated as interchangeable observations.

B. Interpretation Of The Baseline Results

Only technical validation is reported for the detector results. Where classes are in stable positions, background geometry and imaging geometry are alike, and expected and counterpart categories are visually consistent, a near-perfect AP50 can be found in a controlled fixed-layout domain. The tighter the mAP50-95, the more useful it is to indicate the precision of the boundary and meaningful class variation. The outcomes are therefore in favor of learnability of annotations, rather than extensive forensic use. Images taken with external photos of various phones, backgrounds, ultraviolet sources, wear, folds, occlusion, fake image capture methods and denominations are still required.

The difference-hash audit reinforces this caution. Although no exact image crossed a split, 443 cross-split close-pair warnings were found. Some

are plausibly explained by common denomination layout and acquisition background, but the unavailable physical-note registry prevents a definitive identity-level exclusion. Future releases should assign immutable physical specimen identifiers at acquisition, preserve full EXIF and session logs, and capture calibration targets. A prospective external test set should then be collected under different sessions and hardware.

C. Reuse Potential

Within its access conditions, the resource can support modality-specific security-feature detection, expected-versus-replica class analysis, visible-ultraviolet registration research, feature co-occurrence analysis, robustness testing, and interpretable error reporting. The annotation vocabulary should not be transferred unchanged to passports, identity cards, visas, driving licenses, or certificates. Those documents require new expert-defined taxonomies and independent acquisition and validation. What is reusable across document families is the data protocol: paired modality capture, named local features, explicit counterpart labels, source-first partitioning, training-only augmentation, and auditable lineage.

D. Limitations

- The empirical domain contains one denomination and a controlled acquisition arrangement; population diversity and field conditions are not represented.
- The available inputs do not include an independent physical-note register, so the 524 source keys cannot be externally re-verified as 524 physically independent banknotes.
- Complete EXIF histories, ultraviolet irradiance, spectral calibration, geometric measurements, and session-level capture logs were not preserved.
- The historical release contains one fewer visible descendant than the nominal 3,144-file design and originally placed related variants across folders; the corrected evaluation root addresses the split problem without rewriting the historical archive.

- Difference hashing retained cross-split near-similarity warnings; fixed layout may explain many warnings, but physical identity cannot be decided by perceptual hash alone.
- No completed independent second-annotator agreement study is available. Class agreement, Cohen's kappa, and box IoU should be added in a future release.
- The baseline contains one principal training seed and no external dataset. Detector results are therefore point estimates within this domain.
- Counterfeit case-property imagery is security-sensitive and cannot be released without institutional authorization and access controls.

E. Separation From The Companion Modeling Study

This dataset article and the companion modeling study answer different questions. The present paper asks whether the resource is documented, paired, annotated, auditable, statistically characterized, and learnable under a corrected split. The companion study asks how named detections can be converted into specimen-level evidence representations and downstream decisions. To preserve that separation, this manuscript does not report GLOBAL20, localized or whole-image DINOv2, handcrafted baselines, patch classifiers, authentication accuracy, calibration, ablations, selective prediction, or examiner-readable reports. The relationship should be disclosed to both journal editors, and whichever manuscript becomes citable first should be referenced in the other.

Data Availability And Ethical Use

The counterfeit images originate from FIA case-property material and contain security-sensitive content. The full image resource is therefore not described as an unrestricted public dataset. Subject to institutional authorization, controlled access may be considered for eligible researchers on the basis of institutional affiliation, a lawful research proposal, agreed use restrictions, and case-property confidentiality requirements. The intended controlled package comprises the authorized image subset, YOLO labels, class

definitions, split assignments, lineage manifest, machine-readable data card, audit tables, and the reproducible Colab notebook. No public repository DOI is claimed in this manuscript.

The publication materials exclude case identifiers, accused-person information, seizure details, investigation narratives, and other legal metadata. The resource is intended for lawful research in document analysis, robustness, and forensic decision support. It is not intended for the reproduction of security devices, operational adjudication, unauthorized commercial deployment, or training intended to facilitate counterfeiting. Any external transfer remains subject to the relevant institutional and case-property controls.

Reproducibility Package

The supplied notebook reconstructs the archive, audits and remaps labels, parses source keys, verifies paired modalities, applies a deterministic 420/52/52 key split, builds the corrected modeling root, confirms training-only descendants, calculates class and geometry summaries, computes image-quality descriptors, trains the two YOLOv8m baselines, evaluates validation and test partitions, exports per-class AP, freezes the software environment, and writes a machine-readable data card. The results archive retains the image-lineage manifest, dataset and label tables, audit summaries, plot hashes, and model-weight hashes. Optional model-size and resolution ablations are present in the notebook but were skipped in the executed package and are not reported as completed results.

Conclusion

This paper documents a controlled paired visible and 365 nm ultraviolet banknote-image resource with feature-level annotations for a 500 Saudi Riyal case study. The reproducible audit recovered 524 balanced filename-derived source keys, 1,048 paired original modality images, and a 3,143-file historical YOLO-ready release. The corrected source-key-first protocol yielded 2,728 evaluation images, in which only descendants of the original images were used for training, while

only original images were used for validation and test. Both the VISIBLE16 and UV4 taxonomies have explicit expected and counterpart classes, integrity tests set up valid labels, augmentation locking, zero source-key overlap and zero cross-split exact duplicate classes while maintaining the evidentiary limitation of near duplicate warnings. The test mAP50 for visible and ultraviolet branches of YOLOv8m baselines were 0.9917 and 0.9950 respectively, demonstrating learnability of annotations in the controlled domain. The resource does not provide independence of physical-note beyond the archived keys, across-document generalization or operational forensic validity. It's defensible contribution is the dataset protocol and auditable resource layer that future external and document specific studies can be built upon.

Acknowledgment

The authors acknowledge the institutional case-property custodian for permitting controlled access to the counterfeit specimens used for research imaging. The acknowledgment does not imply institutional endorsement of any operational use or unrestricted release of the dataset.

REFERENCES

- [1] International Civil Aviation Organization, "Machine Readable Travel Documents, Part 2: Specifications for the Security of the Design, Manufacture and Issuance of MRTDs," ICAO Doc 9303, Montreal, 2021.
- [2] Saudi Central Bank, "Sixth issue - security features of the five hundred riyal note," 2026. [Online]. Available: https://www.sama.gov.sa/en-US/Currency/Pages/Security_Feature.aspx. [Accessed 14/09/2026].
- [4] J. W. Lee, H. G. Hong, K. W. Kim, and K. R. Park, "A survey on banknote recognition methods by various sensors," *Sensors*, vol. 17, no. 2, p. 313, 2017.
- [5] E. B. Brauns and R. B. Dyer, "Fourier transform hyperspectral visible imaging and the nondestructive analysis of potentially fraudulent documents," *Applied Spectroscopy*, vol. 60, no. 8, pp. 833-840, 2006.
- [6] J. Shit and V. M. Manikandan, "Role of hyperspectral imaging in forensic science," *Algorithms*, vol. 19, no. 8, p. 629, 2026.
- [7] D. A. Khan, S. Ali, A. N. Khan, H. Y. Ahmed, M. Zeghid and S. A. Alqahtani, "Explainable Swin transformer with clinically constrained ECG-vital signs fusion for cardiovascular disease detection from real-world ECG images," *Scientific Reports*, 2026.
- [8] T. D. Pham, C. Park, D. T. Nguyen, G. Batchuluun, and K. R. Park, "Deep learning-based fake-banknote detection for the visually impaired people using visible-light images captured by smartphone cameras," *IEEE Access*, vol. 8, pp. 63144-63161, 2020.
- [9] A. A. Chowdhury, A. Das, D. Karmaker, and K. K. S. Hoque, "Evaluation of deep learning models on UV ink: A fake money detection scheme with RPN," in *2nd International Conference on Computing Advancements*, 2022.
- [10] P. Chhabra and S. Goyal, "A two-phase deep learning model for counterfeit detection of Indian banknotes using YOLO-NAS and UV imaging for visually impaired people," *Evergreen*, vol. 12, no. 3, pp. 1633-1653, 2025.
- [11] M. Taheri, C. F. Moreno-Garcia, C. Stewart, C. Fernandez, and R. Prabhu, "AI-enhanced imaging and multimodal detection of rare-earth fluorescence-based security features for document authentication and border control," in *Proceedings of SPIE*, 2025.

- [12] F. C. P. van der Horst, J. Snell, and J. Theeuwes, "Enhancing banknote authentication by guiding attention to security features and manipulating prevalence expectancy," *Cognitive Research: Principles and Implications*, vol. 6, p. 73, 2021.
- [13] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016.
- [14] K. B. Bulatov et al., "MIDV-2020: A comprehensive benchmark dataset for identity document analysis," *Computer Optics*, vol. 46, no. 2, pp. 252-270, 2022.
- [15] C. Boned et al., "Synthetic dataset of ID and travel documents," *Scientific Data*, vol. 11, p. 1356, 2024.
- [16] K. Nikolaidou, M. Seuret, H. Mokayed, and M. Liwicki, "A survey of historical document image datasets," *International Journal on Document Analysis and Recognition*, vol. 25, no. 4, pp. 305-338, 2022.
- [17] Apple Inc., "iPhone 11 Pro - technical specifications," 2019. [Online]. Available: <https://support.apple.com/en-us/111879>. [Accessed 14 09 2026].
- [18] Apple Inc., "Use iPhone camera tools to set up your shot," 2026. [Online]. Available: <https://support.apple.com/guide/iphone/set-up-your-shot-iph3dc593597/ios>. [Accessed 14 09 2026].
- [19] Tzutalin, "LabelImg," GitHub repository, 2015.
- [20] A. Buslaev et al., "Albumentations: Fast and flexible image augmentations," *Information*, vol. 11, no. 2, p. 125, 2020.
- [21] E. B. Wilson, "Probable inference, the law of succession, and statistical inference," *Journal of the American Statistical Association*, vol. 22, no. 158, pp. 209-212, 1927.
- [22] G. Bradski, "The OpenCV library," vol. 25, no. 11, pp. 120-125, 2000.
- [23] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics YOLO," GitHub repository, 2023.
- [24] T.-Y. Lin et al., "Microsoft COCO: Common objects in context," in *Computer Vision - ECCV 2014*, 2014.