DeepID Challenge of Detecting Synthetic Manipulations in ID Documents

Pavel Korshunov[1] Vidit[1] Amir Mohammadi[1] Christophe Ecabert[1]
Nevena Shamoska[2] Sébastien Marcel[1] Zeqin Yu[3] Ye Tian[3] Jiangqun Ni[3]
Lazar Lazarevic[4] Renat Khizbullin[4] Anastasiia Evteeva[4] Alexey Tochin[4]
Aleksei Grishin[4] Anjith George[1] Daniel DeAlcala[5] Tamás Endrei[5,6]
Javier Muñoz-Haro[5] Ruben Tolosana[5] Ruben Vera-Rodriguez[5] Aythami Morales[5]
Julian Fierrez[5] György Cserey[6] Hardik Sharma[7,10] Sachin Chaudhary[7,8]
Akshay Dudhane[9] Praful Hambarde[10] Amit Shukla[10] Prateek Shaily[7,8]
Jayant Kumar[7,8] Ajinkya Hase[10] Satish Maurya[10] Mridul Sharma[10]
Pallav Dwivedi[10]
[1]Idiap Research Institute, [2]PXL Vision, [3]Sun Yat-sen University, [4]Incode Technologies Inc.,
[5]Universidad Autonoma de Madrid, [6]Pázmany Péter Catholic University, [7]Reagvis Labs Pvt. Ltd.,
[8]UPES Dehradun, [9]MBZUAI, [10]IIT Mandi.

Abstract

An increase in AI based manipulations of ID document images threatens KYC systems widely used in online banking and other digital authentication services. DeepID challenge aimed to advance the research in the methods for detecting synthetic manipulations in ID documents. For that purpose a FantasyID dataset of both bona fide and manipulated fantasy ID cards was provided to the participants for training and tuning of their systems. Participating submissions were evaluated on a test set of FantasyID card created with both seen and unseen attacks, and on an out-of-domain private dataset of 20*K real ID documents containing both genuine bona fide and manipulated samples. The challenge included two tracks 1) a binary detection track to detect whether an ID document is manipulated or not and 2) a localization track, where the goal was to identify the manipulated regions of an ID document. The evaluations were based on the F1-score metric for both detection and localization track and the submissions were ranked based on the weighted average F1-score of FantasyID (with weight* 0.3*) and private (with weight* 0.7*) test sets. With more than* 100 *registrations in the challenge,* 26 *teams have participated and* 6 *of them managed to beat the provided TruFor baseline method in detection track and* 4 *teams in the localization track. Sunlight team from Sun Yat-sen University has won both tracks of the challenge and UAM-Biometrics has ranked best in the private dataset.*

(a) Bona fide

(b) Manipulated

Figure 1. Examples from FantasyID dataset.

1. Introduction

Identity (ID) documents play a critical role in modern identity verification and authentication systems, particularly in *Know Your Customer* (KYC) processes for online banking,

e-commerce, and other digital services. Recent advances in visual generative models, including generative adversarial networks (GANs) and diffusion-based methods, have made it increasingly easy to create highly realistic and convincing digital forgeries of ID documents. Such manipulations can alter faces, change text-based personal information, or modify document security features with minimal visual artifacts, raising serious concerns for the reliability of automated ID verification pipelines.

While a lot of research has been done in the domain of general-purpose manipulation detection in images and videos aka deepfakes detection [5, 6, 15, 19, 20, 22, 28, 29, 34, 35], the domain of ID document forgery detection presents unique challenges. Tampering in ID documents is often subtle and localized to small, high-value regions such as names, dates, identification numbers, or facial portraits. Additionally, there is lack of publicly available datasets of real ID documents due to privacy, ethical, and legal constraints, limiting the ability to develop and benchmark robust detection models [24, 25].

Therefore, we organized the **DeepID Challenge** at ICCV 2025, the first public competition dedicated to detecting synthetic manipulations in ID documents. The focus of the challenge was on *injection attacks* (as opposed to presentation attacks [10, 13, 14]), specifically:

- **Face-swapping manipulations**, where the document portrait is replaced using face-swapping methods.
- **Text inpainting manipulations**, where partial or full textual fields such as names, dates, or identification numbers are altered using generative techniques.

The competition targeted two core research problems in visual forensics: (1) binary classification of bona fide versus forged ID documents, and (2) pixel-level localization of manipulated regions.

Participants were provided with the *FantasyID* dataset [21] for training and tuning their models. After the competition, the dataset has been made public with most of the samples under CC By 4.0 license that allows commercial use[1]. FantasyID consists of 362 ID Cards generated from manually designed templates representing 13 different countries and languages, populated with fantasy personal data but real human faces from public datasets. The cards were printed using an Evolis Primacy 2 card printer and scanned with multiple devices (iPhone 15 Pro, Huawei Mate 30, Kyocera TASKalfa 2554ci). The participants were provided with 786 bona fide samples constituting train and validation sets. Manipulated versions were generated using two face-swapping (InSwapper [1] and Facedancer [27]), and two text-inpainting (Textdiffuser2 [4] and DiffSTE [16]) methods, resulting in $1,572$ forged samples. This dataset was designed to mimic the appearance and linguistic diversity of real ID documents while avoiding the use of actual personal information. For more details about the dataset refer to [21].

The evaluation phase used two separate test sets:

1. An **in-domain** test set, a different subset of FantasyID containing both known and unseen attack types.
2. An **out-of-domain** private test set, 20K real ID document images provided by PXL Vision[2], with forged versions created via face-swapping and text inpainting. This dataset was never released to participants and was used exclusively for hidden evaluation to assess cross-domain generalization.

The challenge had two tracks:

Track 1: Detection. Binary classification, where models output a confidence score (0 to 1) indicating whether the document is bona fide (closer to 1) or manipulated (closer to 0). Performance was measured using the F1-score with a 0.5 decision threshold.

Track 2: Localization. Pixel-wise identification of manipulated regions, evaluated using an aggregated per-image F1-score that equally weights bona fide and forged samples to mitigate class imbalance.

For both tracks, the final ranking was computed as a weighted aggregate F1-score:

$$\text{Aggregate F1} = 0.3 \times \text{F1}_{\text{FantasyID}} + 0.7 \times \text{F1}_{\text{Private}},$$

making the performance on the more challenging out-of-domain private dataset more important. All submissions were evaluated under identical conditions on an air-gapped GPU server (RTX 3090, 24 GB) using Docker containers. A baseline method, *TruFor* [12], was provided to participants along with reference Docker code.

The DeepID Challenge attracted over 100 registered participants, with 26 teams successfully submitting docker containers that were evaluated. Only six teams scored higher than the TruFor baseline in the detection track, and four teams in the localization track (see Tab. 1). The *Sunlight* team from Sun Yat-sen University, China achieved the highest scores in both tracks by significantly outperforming every other team on the FantasyID test set. On the private dataset, UAM-Biometrics achieved the highest scores, showing a better out-of-domain generalization capabilities of their approach.

In this paper, we present an overview of the DeepID Challenge, describe the winning approaches that overpass the baseline and provide more in-depth analysis of the results on both FantasyID and the private datasets focusing on the results for different types of manipulations.

[1] https://www.idiap.ch/paper/fantasyid/

[2] https://www.pxl-vision.com/

Team	Affiliations	Country	Tracks
Sunlight	3	China	Both
Incode	4	USA	Detection
AG (Anjith George)	1	Switzerland	Both
UAM-Biometrics	5,6	Spain	Both
hardik	7,8,9,10	India	Detection
Reagvis	7,8	India	Detection
VISION	10	India	Localization

Table 1. The teams whose submissions scored above the TruFor baseline in either detection, localization, or both tracks. Affiliations refer to the author names in the title page.

#	Team	t_f	F1_f	t_p	F1_p	Agg F1
1	**Sunlight**	0:26	**0.991**	1:50	0.719	**0.801**
2	Incode	0:45	0.868	4:10	0.753	0.788
3	AG	0:56	0.958	1:27	0.711	0.785
4	UAM-Biometrics	0:51	0.712	1:18	**0.788**	0.765
5	Hardik	0:59	0.839	1:28	0.658	0.712
6	Reagvis	1:09	0.816	2:12	0.663	0.709
7	*Baseline*	*0:44*	*0.807*	*1:09*	*0.662*	*0.706*

Table 2. Detection performance of winning teams in DeepID.

#	Team	t_f	F1_f	t_p	F1_p	Agg F1
1	**Sunlight**	0:24	**0.784**	1:51	0.716	**0.737**
2	UAM-Biometrics	0:51	0.620	1:18	**0.757**	0.716
3	VISION	1:02	0.612	1:30	0.738	0.700
4	AG	8:30	0.686	11:34	0.662	0.669
5	*Baseline*	*0:44*	*0.590*	*1:09*	*0.627*	*0.616*

Table 3. Localization performance of winning teams in DeepID.

2. Winning Teams

F1 scores ranking the winning teams (see Tab. 1) compared with TruFor baseline are shown in Tab. 2 for detection track and Tab. 3 for localization track. Columns t_f and F1_f show inference times and F1 scores on FantasyID and t_p and F1_p on the private set. Full results can be viewed on DeepID website[3]. In the following, we summarize the top teams approaches focusing on the architectural choices, training strategies, fusion techniques, and domain adaptations.

Method	Data	Detection	Localization	Average
MVSS-Net	13k	0.533	0.187	0.360
TruFor	876k	0.746	0.626	0.686
Re-MTKD	60k	0.758	0.637	0.697

Table 4. Zero-shot performance (F1 score) of existing IFDL methods on the FantasyID validation set.

[3] https://deepid-iccv.github.io/

2.1. Sunlight Team

The *Sunlight* team from Sun Yat-sen University proposed a two-stage training pipeline (see Fig. 2) based on the Reinforced Multi-teacher Knowledge Distillation (Re-MTKD) framework [36].

Preliminary analysis. The Sunlight team conducted an initial analysis to understand the domain-specific characteristics of the ID document images in the challenge. They observed that real and tampered images differ in JPEG compression quality (QF 95 vs. QF 75), and that the images feature structured layouts, clean backgrounds, and concentrated text regions. Such properties pose challenges for general-purpose image forgery detection and localization (IFDL) models, particularly in modeling compression artifacts and document-specific textures.

They also evaluated the zero-shot performance of several state-of-the-art IFDL methods, including MVSS-Net [5], TruFor [12], and their own Re-MTKD [36]. As shown in Tab. 4, Re-MTKD achieved the highest F1 scores in both detection and localization despite being trained on fewer samples, indicating strong cross-domain generalization.

Proposed pipeline. In **Stage 1** (left part of Fig. 2), the team reused the Cue-Net student model pretrained using Re-MTKD framework that has shown a strong generalization across various natural image forgery datasets. Teacher models, each focusing on a specific manipulation type (e.g., splicing, copy-move, inpainting), were trained on the datasets corresponding to each manipulation type, such as CASIAv2 [9] for copy-move, Fantastic Reality [18] for splicing, and GC [32] for inpainting. The student model was optimized using a combination of soft distillation loss $\mathcal{L}_{\text{soft}}$ and hard supervision loss $\mathcal{L}_{\text{hard}}$, which combines segmentation, classification, and edge-aware losses to enhance detection accuracy and localization precision.

In **Stage 2** (right part of Fig. 2), the pretrained model was fine-tuned on the FantasyID dataset to capture ID-specific characteristics such as structured layouts, subtle tampering traces, and compression-induced artifacts. Training was done on 512×512 cropped patches with randomized JPEG compression (QF in $[70, 100]$) applied to simulate the diverse compression artifacts observed in the provided FantasyID dataset. During fine-tuning, only the hard supervision loss $\mathcal{L}_{\text{hard}}$ is optimized to ensure stable domain adaptation in the absence of teacher's guidance. At inference, whole-image processing was used to avoid resizing artifacts. This approach[4], when trained for 16 epochs, achieved first place in detection track and, when trained for 31 epochs, first place in localization track.

[4] https://github.com/ZeqinYu/ICCV-DeepID2025-Sunlight

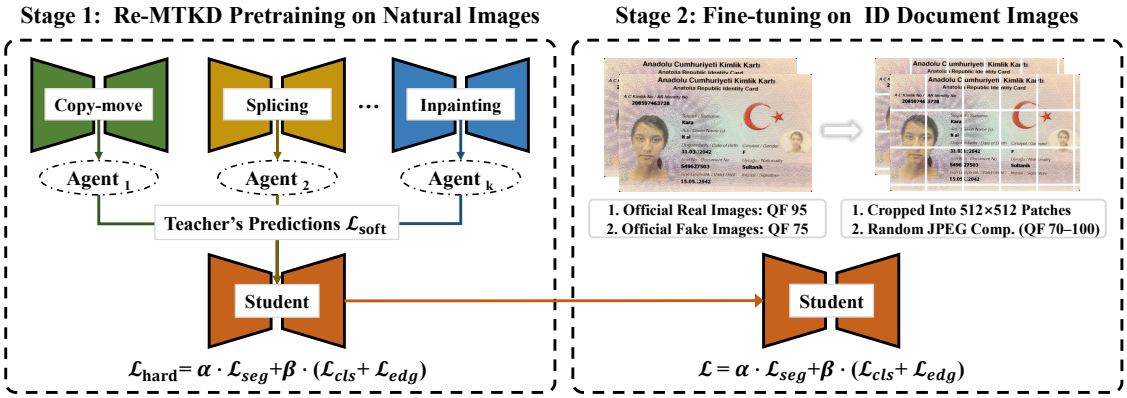


Figure 2. Overview of Sunlight team's two-stage training pipeline.

2.2. Incode Team

The *Incode* team (USA) developed an ensemble combining their proprietary transformer-based fraud detection models with several open-source architectures, including TruFor [12]. Their in-house transformer-based models were initially trained on ∼100K ID document images, including 20K manipulated samples with various portrait and text alterations. For the challenge, these models were fine-tuned on the FantasyID dataset to adapt to synthetic forgeries. The final submission fused outputs from the proprietary models and an adapted TruFor variant, with parameters tuned to reduce false rejections. This ensemble achieved second highest performance on both the in-domain and out-of-domain datasets in the detection track.

2.3. AG Team

The *AG* team (Switzerland) submitted *EdgeDoc*, a hybrid model combining TruFor [12] with a custom architecture derived from EdgeFace [11]. Due to the limited availability of public training data, a pretrained TruFor model was leveraged to extract NoisePrint maps. These maps, along with the original images, served as inputs to train a custom EdgeDoc model. The EdgeDoc model was designed to output both a segmentation mask and a binary classification score, enabling both localization and detection of image manipulations. Training was performed on the public training subset of the FantasyID dataset. To adapt the EdgeFace backbone for this task, it was modified into a U-Net-style segmentation network. The original EdgeFace architecture, which combines convolutional and transformer components, allows for effective patch-level interaction. Given the small size of the training set, a lightweight version of the architecture was adopted to improve efficiency, allowing for full-resolution image processing while maintaining speed. During inference, the outputs from both the EdgeDoc and TruFor models were fused, specifically, their scores and localization masks to compute the final prediction score. This approach achieved a third place and fourth places in detection and localization tracks.

2.4. UAM-Biometrics Team

The *UAM-Biometrics* team (Spain) proposed *TruForID*, an adaptation of TruFor [12] with a SegFormer backbone [33]. In general, the base model was fine-tuned using the train split of the FantasyID dataset, improving the optimizer and learning rate scheduler, and incorporating knowledge distillation. Since TruFor, like many state-of-the-art methods [8], performs localization-based detection, the TruForID model was submitted to both detection (scoring fourth place) and localization (scoring second place) tracks.

Since TruFor model itself had a good performance (over 80% F1-score) on the FantasyID dataset, it meant FantasyID cards were created applying digital techniques (i.e., face swapping and text inpainting) over the real ID samples, introducing distinguishable fingerprints/traces in the images. Therefore, fine-tuning the TruFor model on FantasyID starting from the pre-trained weights would preserve part of this knowledge.

Starting from pretrained TruFor weights, the model was fine-tuned using 512×512 patches, as commonly done in the forgery detection literature [2, 12, 30] and, to improve generalization, extensive augmentations including scaling, compression, horizontal flipping, brightness, and contrast adjustments. Two patches were selected from each ID: one is chosen randomly, and the second one is selected from 30 candidates as a random sample among the 15 farthest from the first. The patches were treated as independent samples, effectively doubling the training set size. TruForID model was trained using the AdamW optimizer [23] combined with a cosine learning rate scheduler with warm-up to improve training dynamics. Additionally, a similarity-preserving knowledge distillation [31] was incorporated during training, where the baseline TruFor model was defined as the teacher and the proposed TruForID as the student. During training, distillation loss was computed

as the cosine similarity between the bottleneck values of the SegFormer encoder for both models. This distillation-based loss encourages the model to retain generalizable features while adapting to the new task.

2.5. Hardik Team

The *Hardik* team (India) developed a Region-Aware ID Card Forgery Detection system enhancing TruFor [12] with targeted region analysis. Automatic segmentation identified face regions, text fields, and date fields. Each was analyzed independently: faces for texture and lighting inconsistencies (15% weight in the final score), typography uniformity and digit pattern consistency (25% weight in the score), and edges for boundary consistency (10% weight). These scores were fused with the baseline TruFor output (50% weight in the final score). No additional training or tuning was done, so original baseline TruFor was used. Instead, the lightweight computer vision techniques and statistical measures were used in combination with TruFor score. The weights used for statistics and TruFor scores were determined through empirical testing on validation samples from the fantasy ID dataset. The method achieved an aggregate F1 of 0.712 compared to pure TruFor baseline 0.706, with notable improvement on the FantasyID set, with 0.839 F1 score compared to 0.807 baseline.

2.6. Reagvis Team

The *Reagvis* team (India) augmented the TruFor [12] baseline with three lightweight, domain-specific modules. First, a **chronological sanity gate** used a Donut OCR head [17] to extract and normalize date-of-birth, issue, and expiry fields, applying six logical checks, including chronological order, plausible age, no future dates, and that the implied age is plausible. Cards failing these checks were immediately flagged as forged. Second, for remaining cases when the card either passes the gate or OCR was inconclusive, **hand-crafted anomaly cues** were computed: a 32×32 patch-level Canny edge-density score [3] and a HOG-based font-consistency score [7] to detect copy-paste artifacts. Finally, in a **fixed-weight fusion** step, the inverted TruFor confidence (so that a value of 1 denotes a pristine document) was combined with the edge and font scores as $S = 0.6\, S_{\text{TruFor}} + 0.2\, S_{\text{edge}} + 0.2\, S_{\text{font}}$. No training beyond the available TruFor baseline was done. Fusion weights were selected via grid search on the FantasyID dataset maximising F1 while maintaining deterministic behaviour. This ensemble improved the leaderboard F1 from 0.706 to 0.709 while adding negligible runtime overhead.

2.7. VISION Team

The *VISION* team (India) focused on preserving high-frequency forensic cues by increasing the input resolution to 1024×1024, compared to 512×512 in the TruFor [12] baseline. The hypothesis was that downscaling removes subtle artifacts such as texture mismatches and unnatural edges, reducing detection efficacy. The higher-resolution inputs allowed TruFor model to better exploit these details, leading to improved localization accuracy. This came at the cost of longer inference times (1:30:46 vs. 1:09:49 on the private set), but yielded an aggregate F1 score of 0.700 in the localization track, surpassing the baseline which was at 0.616.

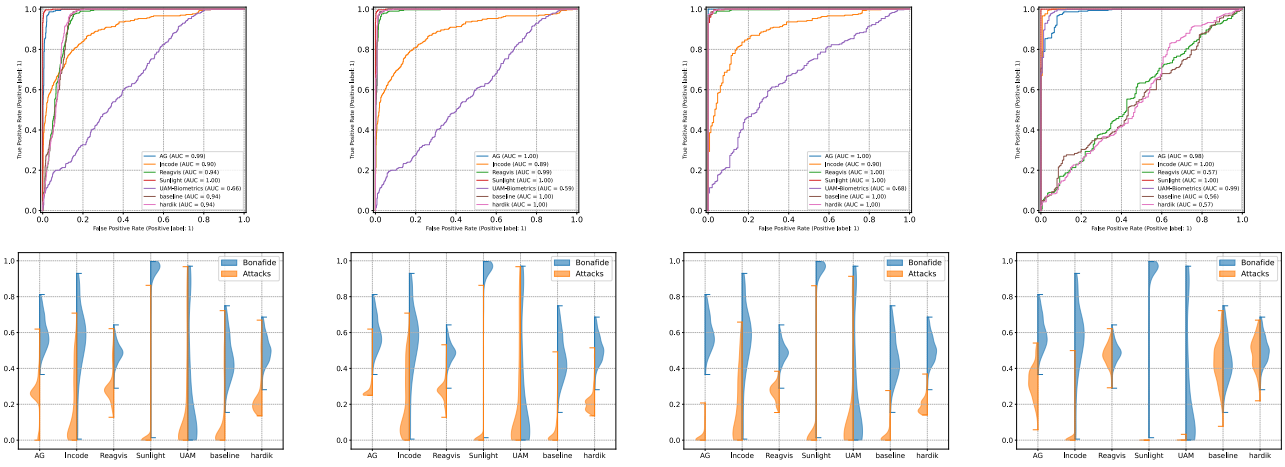
3. In-depth Analysis of the Results

We analyze the performance of the top submissions on our test set. This set consists of FantasyID [21] and a private set of real ID documents provided by PXL Vision. It is designed to test the generalization of methods on *unseen* manipulations and *unseen* card type scenarios.

We provide participants with the FantasyID [21] train-val set and use its test set to benchmark all submissions. This set consists of manipulations that are only text or face region specific. (a) **FA1** is text-only manipulation created by fine-tuning Textdiffuser2 [4] on the train-val set. Additional post-processing is done to subvert the copy-paste artifacts. (b) **FA2** is text manipulation (same as FA1) performed on card templates different from the train-val set. (c) **FA3** is face manipulation on unseen card templates. We refer to all of these manipulations combined as **FA-all**

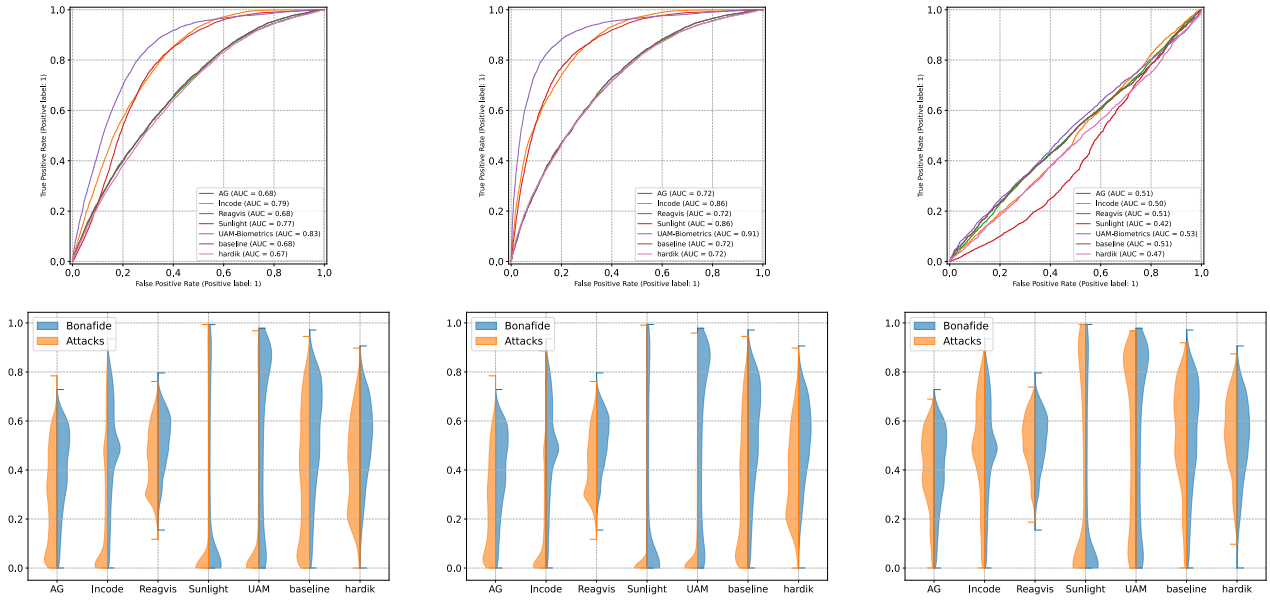
The private set consists of real ID documents, and we create the following manipulations: (a) using the off-the-shelf model we swap faces using Facedancer [27], and Inswapper [1]; text are swapped using DiffSTE [16], and Textdiffuser2 [4]; **PA1** is a text-only manipulation using DiffSTE; **PA2** is face and text manipulation using Facedancer and Textdiffuser2; **PA3** is face-only manipulation using Inswapper. We refer to them as *no-finetuning* manipulations, as pre-trained models are used *as is*. (b) We create more sophisticated text manipulation by finetuning Textdiffuser2 on private data, followed by Poisson blending [26]. **PA4-7** are created by changing a single text field birthdate, expirydate, lastname, and firstname, respectively. We refer to them as *finetuned* manipulations and all of the combined manipulations are refered to as **PA-all**.

Detection. All submissions, except for the Sunlight team, were either extension of TruFor or used it directly, because it performs well on inpainting/copy-paste forgeries. Fig. 3 illustrates the detection performance of different methods on the FantasyID test set. For a detection threshold of 0.5, Sunlight achieves the best performance. The score distribution shows that, except for Sunlight, all other methods require threshold tuning to reach a good performance on the test set. Except for Incode and UAM-Biometrics, all other teams perform well on the text-only manipulations, FA1, and FA2. In the face-only manipulation, FA3, Reagvis



(a) *FA-all* (b) *FA1- text* (c) *FA2- text* (d) *FA3- face*

Figure 3. **FantasyID Detection Results**: ROC plots (top) and Score distributions (bottom); a–d show performance of the challenge submissions on different kinds of manipulations. (a) Overall performance on the FantasyID (b–c) is text-only manipulation, and (d) contains only face changes.

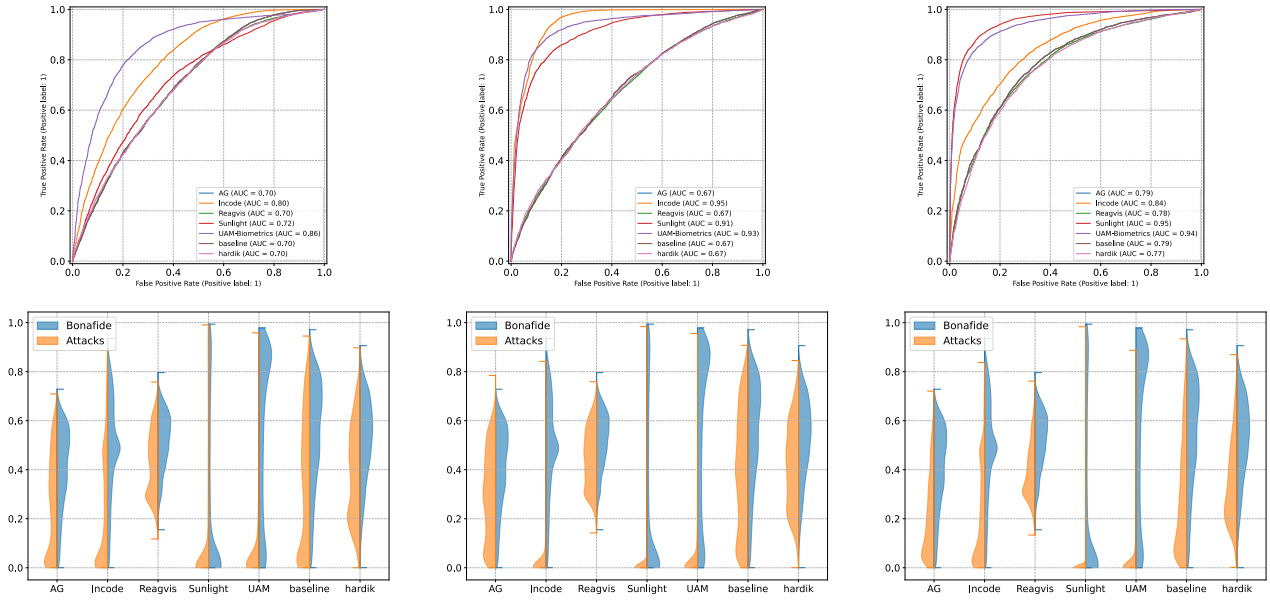


(a) *PA-all* (b) *No finetuning* (c) *Finetuned*

Figure 4. **Private Set Detection Results**: ROC plots (top) and Score distributions (bottom); (a) Overall performance on the private set (b) text and face manipulation with off-the-shelf models (c) text manipulation with finetuned model and poisson blending.

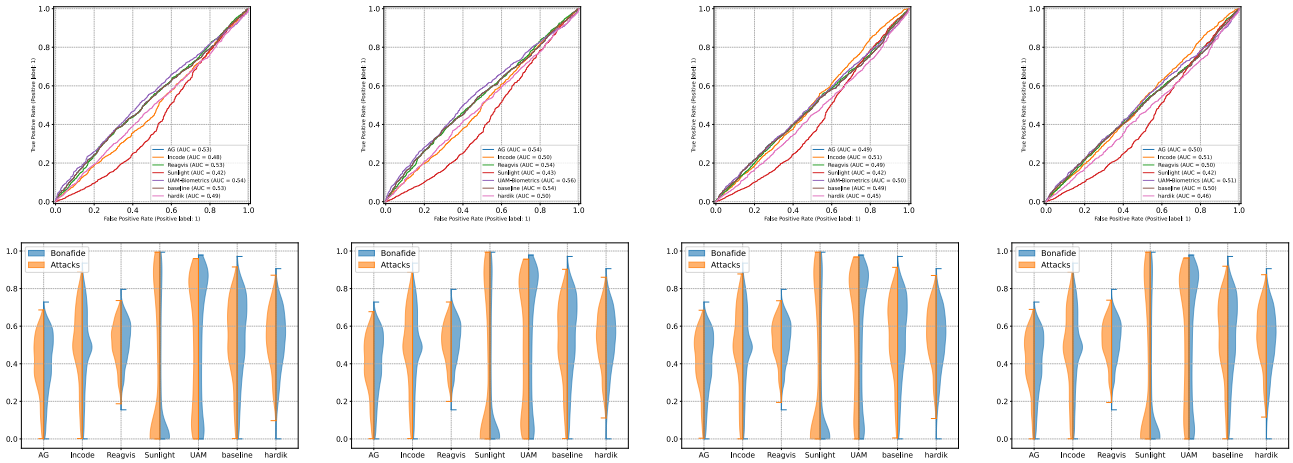
and Hardik demonstrate near random performance. Since TruFor baseline underperforms on this manipulation, it suggests that models from these teams heavily rely on TruFor. Interestingly, FA2 and FA3 contain unseen Fantasy-style card designs, but most of the methods show a good generalization when trained with the FantasyID train set. FantasyID [21] dataset serves as a challenging benchmark with diverse ID card templates and various forms of manipulations.

The private set serves as a benchmark for out-of-distribution performance, as they are real-world ID cards. Here, UAM-Biometrics performs the best, followed by Incode. Sunlight, the best performer on FantasyID, falls behind in this challenging setup. Incode might have seen such real-world IDs in its proprietary training dataset, leading to better generalization than on FantasyID. However, on this private set, all the methods achieve F_1 lower than on FantasyID, see Tab. 2, illustrating the domain gap between the



(a) *PA1- text* (b) *PA2- text and face* (c) *PA2- face*

Figure 5. **Private Set Detection Results for No-finetuning manipulations** ROC plots (top) and Score distributions(bottom): Manipulation created using pretrained models. (a) text-only manipulation using pretrained DiffSTE [16] (b) face manipulated using Facedance [27] and text using Textdiffuser2 [4] (c) face-only manipulation using off-the-shelf Inswapper.

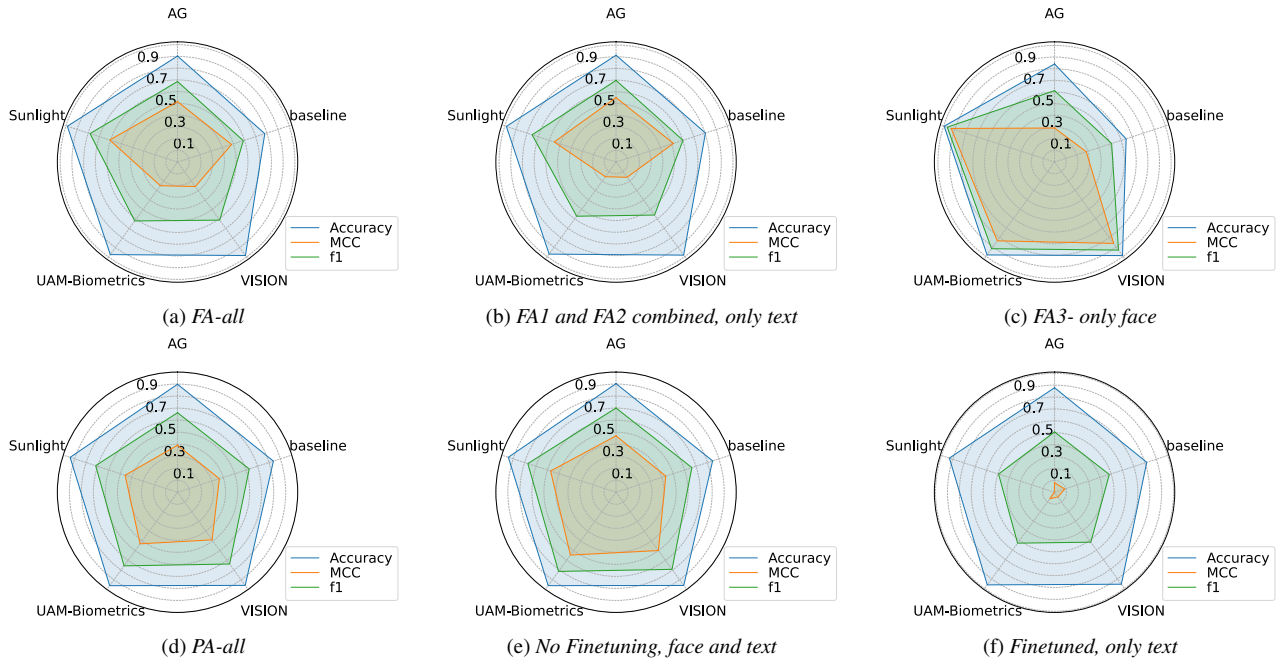


(a) *PA4- birthdate* (b) *PA5- expirydate* (c) *PA6- lastname* (d) *PA7- firstname*

Figure 6. **Private Set Detection Results for Finetuned manipulations** ROC plots (top) and Score distributions (bottom): Text manipulation created by finetuning Textdiffuser2 [4] and poisson blending [26]. We modify the following fields (a) birthdate (b) expirydate (c) lastname (d) firstname. It is a challenging detection problem as the regions modified are small.

two datasets. As shown in Fig. 4, for most of the methods, the score distributions are highly overlapping. *Finetuned* text manipulation serves as a challenging manipulation in which none of the methods performs better than chance.

Figs. 5 and 6 show a breakdown of the team's performance on different types of manipulations. The manipulations shown in Fig. 5 are the same as in the training set of FantasyID. The poorer performance of all the methods here, is largely due to the domain shift from clean FantasyID to real-world capture of IDs using a hand-held camera. The detection task gets even more difficult with *finetuned* text manipulations applied on small text fields as illustrated by Fig. 6. None of the methods generalize to this manipulation, as manipulation is in small image regions and poisson blending [26] suppresses copy-paste artifacts, which TruFor-based methods are good at detecting.



(a) *FA-all* (b) *FA1 and FA2 combined, only text* (c) *FA3- only face*

(d) *PA-all* (e) *No Finetuning, face and text* (f) *Finetuned, only text*

Figure 7. **Localization Results** on FantasyID (top) and Private Set (bottom) for different kinds of manipulations. We plot the performance of different teams across manipulations on Accuracy, F_1 score, and Matthews Correlation Coefficient(MCC). Since the localization task is pixel-wise prediction, it is highly imbalanced, as manipulated pixel are far less than un-manipulated ones. Therefore, for images with manipulation we compute MCC. All the teams fail to achieve a decent performance w.r.t the MCC metric.

Localization. This task involves providing a per-pixel classification score. As this is a fine-grained task, methods need to identify and localize all the manipulated regions. Usually, the image has a few manipulated pixels compared to the size of the entire image. This makes the localization task highly imbalanced, and we need the right metric to account for it. Fig. 7 shows the performance of different teams on w.r.t Accuracy, F_1, and Matthews Correlation Coefficient (MCC). We can see that it is easy to achieve a high accuracy score by simply predicting all pixels as pristine. Therefore, we use different metrics to evaluate the methods.

We compute MCC only for images with manipulation, as MCC is always 0 for bonafide images. None of the methods achieves good performance for the localization task w.r.t. the MCC metric, which provides a balanced assessment under data imbalance. The performance is even worse for manipulations limited to small image regions, i.e., text, for both the FantasyID and Private set. For finetuned manipulations of the private set, the MCC score is close to 0 for all of the methods. This highlights the challenging nature of the localization task.

We used the F_1 score to rank the teams on the leaderboard[3], as we can account for the predictions on both bonafide and manipulated images and it is less affected by the data imbalance than accuracy. Sunlight is the best overall and UAM-Biometrics is the best on the private set.

4. Conclusion

DeepID challenge features a real-world use case of forgery detection in ID cards. Most of the teams built upon TruFor, except for the **Sunlight** team, where they used a multiple teacher-student training method to achieve top spot on the leaderboard. **UAM-Biometrics** shows the best performance on the private set by following knowledge distillation using TruFor as the teacher model. Overall generalization to unseen manipulations and unseen card types remains challenging; none of the teams achieved top spot on both the FantasyID and private sets. The localization task remains difficult, as methods fail to achieve a high score when manipulated regions are small. We hope this challenge motivates more research on ID document forgery detection.

Acknowledgements

Organizers (Idiap) were funded by InnoSuisse 106.729 IP-ICT. Sunlight team was funded by National Natural Science Foundation of China (grants U23B2022 and U22A2030) and Guangdong Major Project of Basic and Applied Basic Research (grant 2023B0303000010). UAM-Biometrics team was funded by M2RAI (PID2024-160053OB-I00 MICIU/FEDER), Cátedra ENIA UAM-VERIDAS en IA Responsable (NextGenerationEU PRTR TSI-100927-2023-2), and PowerAI+ (SI4/PJI/2024-00062).

References

[1] Inswapper. https://github.com/deepinsight/insightface/blob/master/examples/in_swapper/README.md. 2, 5

[2] Arian Bakhtiarnia, Qi Zhang, and Alexandros Iosifidis. Efficient High-Resolution Deep Learning: A Survey. *ACM Computing Surveys*, 56(7):1–35, 2024. 4

[3] John Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6):679–698, 1986. 5

[4] Jingye Chen, Yupan Huang, Tengchao Lv, Lei Cui, Qifeng Chen, and Furu Wei. Textdiffuser-2: Unleashing the power of language models for text rendering. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 386–402. Springer, 2024. 2, 5, 7

[5] Xinru Chen, Chengbo Dong, Jiaqi Ji, Juan Cao, and Xirong Li. Image manipulation detection by multi-view multi-scale supervision. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 14185–14193, 2021. 2, 3

[6] Davide Cozzolino, Giovanni Poggi, Riccardo Corvi, Matthias Nieãner, and Luisa Verdoliva. Raising the Bar of AI-generated Image Detection with CLIP. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 4356–4366, 2024. 2

[7] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 886–893 vol. 1, 2005. 5

[8] Amir Etefaghi Daryani, Mahdieh Mirmahdi, Ahmad Hassanpour, Hatef Otroshi Shahreza, Bian Yang, and Julian Fierrez. IRL-Net: Inpainted region localization network via spatial attention. *IEEE Access*, 11:115677–115687, 2023. 4

[9] Jing Dong, Wei Wang, and Tieniu Tan. CASIA image tampering detection evaluation database. In *Proceedings of the IEEE China Summit and International Conference on Signal and Information Processing (ChinaSIP)*, pages 422–426. IEEE, 2013. 3

[10] Meiling Fang, Marco Huber, Julian Fierrez, et al. SynFacePAD 2023: Competition on face presentation attack detection based on privacy-aware synthetic training data. In *Proceedings of the IEEE International Joint Conference on Biometrics (IJCB)*, 2023. 2

[11] Anjith George, Christophe Ecabert, Hatef Otroshi Shahreza, Ketan Kotwal, and Sébastien Marcel. EdgeFace: Efficient face recognition model for edge devices. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 6(2):158–168, 2024. 4

[12] Fabrizio Guillaro, Davide Cozzolino, Avneesh Sud, Nicholas Dufour, and Luisa Verdoliva. TruFor: Leveraging all-round clues for trustworthy image forgery detection and localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 20606–20615, 2023. 2, 3, 4, 5

[13] A. Hadid, N. Evans, S. Marcel, and J. Fierrez. Biometrics systems under spoofing attack: an evaluation methodology and lessons learned. *IEEE Signal Processing Magazine*, 32 (5):20–30, 2015. 2

[14] Javier Hernandez-Ortega, Julian Fierrez, Aythami Morales, and Javier Galbally. Introduction to presentation attack detection in face biometrics and recent advances. In *Handbook of Biometric Anti-Spoofing: Presentation Attack Detection and Vulnerability Assessment*, pages 203–230. Springer, 2023. 3rd Ed. 2

[15] Anubhav Jain, Pavel Korshunov, and Sébastien Marcel. Improving generalization of deepfake detection by training for attribution. In *Proceedings of the IEEE International Workshop on Multimedia Signal Processing (MMSP)*, pages 1–6, 2021. 2

[16] Jiabao Ji, Guanhua Zhang, Zhaowen Wang, Bairu Hou, Zhifei Zhang, Brian Price, and Shiyu Chang. Improving diffusion models for scene text editing with dual encoders, 2023. arXiv:2304.05568 [cs]. 2, 5, 7

[17] Geewook Kim, Teakgyu Hong, Moonbin Yim, Jeongyeon Nam, Jinyoung Park, Jinyeong Yim, Wonseok Hwang, Sangdoo Yun, Dongyoon Han, and Seunghyun Park. OCR-free document understanding transformer. In *Proceedings of the European Conference on Computer Vision (ECCV)*, page 498–517, 2022. 5

[18] Vladimir V Kniaz, Vladimir Knyaz, and Fabio Remondino. The point where reality meets fantasy: Mixed adversarial generators for image splice detection. *Advances in Neural Information Processing Systems (NeurIPS)*, 32, 2019. 3

[19] Pavel Korshunov and Sébastien Marcel. Improving generalization of deepfake detection with data farming and few-shot learning. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2022. 2

[20] Pavel Korshunov, Haolin Chen, Philip N. Garner, and Sébastien Marcel. Vulnerability of automatic identity recognition to audio-visual deepfakes. In *Proceedings of the IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–10, 2023. 2

[21] Pavel Korshunov, Amir Mohammadi, Vidit, Christophe Ecabert, and Sébastien Marcel. FantasyID: A dataset for detecting digital manipulations of ID-documents. In *Proceedings of the IEEE International Joint Conference on Biometrics (IJCB)*, 2025. 2, 5, 6

[22] Myung-Joon Kwon, In-Jae Yu, Seung-Hun Nam, and Heung-Kyu Lee. CAT-Net: Compression artifact tracing network for detection and localization of image splicing. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 375–384, 2021. 2

[23] Ilya Loshchilov and Frank Hutter. Decoupled Weight Decay Regularization. In *Proceedings of the International Conference on Learning Representation (ICLR)*, 2017. 4

[24] Javier Muñoz-Haro, Ruben Tolosana, Ruben Vera-Rodriguez, Aythami Morales, and Julian Fierrez. FakeIDet: Exploring patches for privacy-preserving fake ID detection. In *Proceedings of the IEEE International Joint Conference on Biometrics (IJCB)*, 2025. 2

[25] Javier Muñoz-Haro, Ruben Tolosana, Ruben Vera-Rodriguez, Aythami Morales, and Julian Fierrez. Privacy-aware detection of fake identity documents: Methodology,

benchmark, and improved detection methods (FakeIDet2). *arXiv preprint*, 2025. 2

[26] Patrick Pérez, Michel Gangnet, and Andrew Blake. Poisson image editing. In *ACM SIGGRAPH 2003 Papers*, page 313–318, 2003. 5, 7

[27] Felix Rosberg, Eren Erdal Aksoy, Fernando Alonso-Fernandez, and Cristofer Englund. FaceDancer: Pose-and occlusion-aware high fidelity face swapping. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 3454–3463, 2023. 2, 5, 7

[28] Ruben Tolosana, Christian Rathgeb, Ruben Vera-Rodriguez, Christoph Busch, Luisa Verdoliva, et al. Future trends in digital face manipulation and detection. In *Handbook of Digital Face Manipulation and Detection: From DeepFakes to Morphing Attacks*, pages 463–482. Springer, 2022. 2

[29] Ruben Tolosana, Sergio Romero-Tapiador, Ruben Vera-Rodriguez, Ester Gonzalez-Sosa, and Julian Fierrez. Deepfakes detection across generations: Analysis of facial regions, fusion, and performance evaluation. *Engineering Applications of Artificial Intelligence*, 110:104673, 2022. 2

[30] Konstantinos Triaridis and Vasileios Mezaris. Exploring Multi-Modal Fusion for Image Manipulation Detection and Localization. In *Proceedings of the International Conference on MultiMedia Modeling (MMM)*, page 198–211, 2024. 4

[31] Frederick Tung and Greg Mori. Similarity-Preserving Knowledge Distillation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019. 4

[32] Haiwei Wu and Jiantao Zhou. IID-Net: Image inpainting detection network via neural architecture search and attention. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(3):1172–1185, 2021. 3

[33] Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M Alvarez, and Ping Luo. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. In *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*, 2021. 4

[34] Qichao Ying, Hang Zhou, Zhenxing Qian, Sheng Li, and Xinpeng Zhang. Learning to immunize images for tamper localization and self-recovery. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(11):13814–13830, 2023. 2

[35] Jiwen Yu, Xuanyu Zhang, Youmin Xu, and Jian Zhang. Cross: Diffusion model makes controllable, robust and secure image steganography. *Advances in Neural Information Processing Systems*, 36:80730–80743, 2023. 2

[36] Zeqin Yu, Jiangqun Ni, Jian Zhang, Haoyi Deng, and Yuzhen Lin. Reinforced multi-teacher knowledge distillation for efficient general image forgery detection and localization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 995–1003, 2025. 3