KID34K: A Dataset for Online Identity Card Fraud Detection

Eun-Ju Park
Sungkyunkwan University
Suwon, Republic of Korea
eunju.park@g.skku.edu

Seung-Yeon Back
Sungkyunkwan University
Suwon, Republic of Korea
syon1203@g.skku.edu

Jeongho Kim
Korea Advanced Institute of Science and Technology
Daejeon, Republic of Korea
rlawjdghek@kaist.ac.kr

Simon S. Woo*
Sungkyunkwan University
Suwon, Republic of Korea
swoo@g.skku.edu

ABSTRACT

Though digital financial systems have provided users with convenient and accessible services, such as supporting banking or payment services anywhere, it is necessary to have robust security to protect against identity misuse. Thus, online digital identity (ID) verification plays a crucial role in securing financial services on mobile platforms. One of the most widely employed techniques for digital ID verification is that mobile applications request users to take and upload a picture of their own ID cards. However, this approach has vulnerabilities where someone takes pictures of the ID cards belonging to another person displayed on a screen, or printed on paper to be verified as the ID card owner. To mitigate the risks associated with fraudulent ID card verification, we present a novel dataset for classifying cases where the ID card images that users upload to the verification system are genuine or digitally represented. Our dataset is replicas designed to resemble real ID cards, making it available while avoiding privacy issues. Through extensive experiments, we demonstrate that our dataset is effective for detecting digitally represented ID card images, not only in our replica dataset but also in the dataset consisting of real ID cards. Our dataset is available at https://github.com/DASH-Lab/idcard_fraud_detection.

CCS CONCEPTS

• **Computing methodologies** → **Machine learning**; • **Security and privacy** → **Software and application security**.

KEYWORDS

Dataset, Neural networks, Identity card verification

ACM Reference Format:
Eun-Ju Park, Seung-Yeon Back, Jeongho Kim, and Simon S. Woo. 2023. KID34K: A Dataset for Online Identity Card Fraud Detection. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM '23), October 21–25, 2023, Birmingham, United Kingdom.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3583780.3615122

1 INTRODUCTION

The digital transformation of financial services has been accelerated, particularly since the COVID-19 crisis, resulting in a shift in how one accesses and utilizes these services. With smartphones, one can open bank accounts or apply for loans without visiting a physical branch. Although digital financial systems offer convenience in using banking or stock market services, they are required to be robust and secure to prevent one with malicious intent from misusing another one's identification information. Digital identity (ID) verification is a solution, mainly consisting of biometrics or artificial intelligence techniques, to ensure the safe use of financial services on mobile platforms. One of the most used methods for online ID verification is that mobile applications prompt users to take and upload a picture of their own ID cards to authenticate if the service requester is the same as the ID card owner. This method, however, encounters a significant issue where verification is processed even if a user takes a picture of another person's ID card displayed on a screen, or printed on paper.

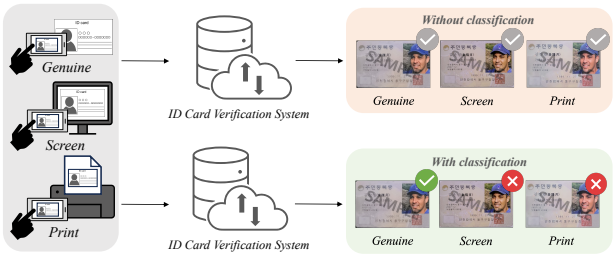

Figure 1: Two types of ID card verification systems. The financial service users can take photos of ID card images displayed on a screen or printed on paper. While the upper system does not distinguish those from genuine ID card images, the lower system does.

In order to tackle such fraudulent activities as shown in Figure 1, we present a novel dataset consisting of ID card images, taking into account three different user scenarios. In the first case, it is assumed that users take photos of their own ID cards, which we label as *genuine*. For the second and third scenarios, we assume that users take pictures of digitally represented ID cards, which we label *screen* and *paper*. It is challenging for human to distinguish between the three cases due to the similarity of these images. For better understanding, Figure 2 shows the examples corresponding to each case. Moreover, to make our dataset publicly accessible, we utilized ID cards that were *legitimately* produced to match the design template of real ID cards in South Korea. We provide two types of ID card images; registration cards, which are similar to green cards for U.S. citizens, and driver's license cards. The main contributions of this research are highlighted as follows:

- We present a novel dataset for ID card fraud detection. While previous research [6, 7, 12] that introduced datasets utilized real ID cards, which may give rise to privacy concerns, our

*Corresponding Author

CIKM '23, October 21–25, 2023, Birmingham, United Kingdom

ACM ISBN 979-8-4007-0124-5/23/10…$15.00
https://doi.org/10.1145/3583780.3615122

Figure 2: The samples of our KID34K dataset. The images in the first column are the genuine ID cards taken. The second and third columns show images taken of ID cards displayed on a screen and printed on paper. The face images and the text information in the ID card images represent all non-existent persons.

dataset is made up of replica ID cards with fake person faces, thereby alleviating any concerns pertaining to personal information.

- We conducted extensive experiments using CNN-based models to improve the accuracy of detecting digitally represented ID card images.
- We demonstrate that the real dataset and the proposed replica dataset (KID34K) have high similarity, thereby our dataset is feasible to deploy to the applications of real ID verification systems.

Note that we restrict our usage to the terms for this research to ensure clarity as follows:

Label. We distinguish three labels: genuine, screen, and paper. Genuine represents the cases where users take photos of real (plastic) ID cards. Screen and paper indicate that users take pictures of ID cards displayed on a screen and printed on paper.

Class. In this paper, we use two classes: real and fake. Images corresponding to "genuine" are classified as real, while images corresponding to "screen" and "print" are classified as fake.

Dataset. The KID34K dataset consists of images shown in Figure 2. To provide accessibility to the dataset without privacy issues, we produced replica IDs that resemble real ID cards. For comparison between our KID34K dataset and a dataset comprising images of real ID cards taken, we distinguish between the two datasets. The replica dataset refers to the KID34K dataset, while the real dataset indicates the dataset of real ID card images that are used in real-life.

2 RELATED WORK

There were several previous researches on ID card fraud detection. *Gonzalez et al.* [7] presented a Presentation Attack Detection (PAD) system for Chilean ID cards. They suggested various presentation attack methods including print, display, and composite. In subsequent research, they added plastic and synthetic to their presentation attack methods [6]. *Benalcazar et al.* [1] presented a GAN-based ID card image generation method for improving presentation attack detection. In order to enhance the size of the dataset, paper and screen textures were composited during the creation of the ID card images. Their experiment was carried out by using a 2-stage architecture based on MobileNet V2 [13] architectures. In contrast, our approach adopts simple model designs instead of complex architectures. This strategy not only simplifies the implementation process but also yields favorable outcomes. *Mudgalgundurao et al.* [12] proposed a pixel-wise supervision based on DenseNet [9] to detect presentation attacks of the printed on paper and displayed on a screen. They constructed a video-captured dataset consisting of German ID cards and residence permits. However, these studies have commonly refrained from sharing their datasets due to privacy concerns. Since we aim to address this issue by publicly releasing our dataset, we have created a publicly available dataset of Korean identification cards comprising three distinct labels: genuine, screen and print.

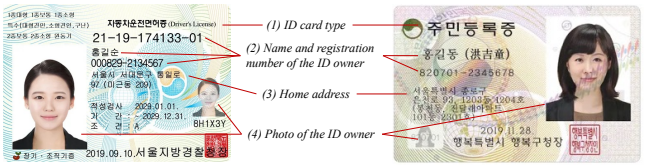

Figure 3: Real ID card samples issued by the Korean government [11, 16].

To create a publicly available dataset, we utilized the information of non-existent persons to generate ID card images that contain distinguishable information from actual registration cards. We generated our dataset by physically producing ID cards and capturing images of them in real-world scenarios. In addition to its practical utility, the KID34K dataset highlights an ability to effectively contribute to the classification performance on real ID card datasets.

3 THE KID34K DATASET

The KID34K dataset[1] is an image dataset that was created by producing replica ID cards and capturing images of the ID cards based on real scenarios. Our dataset consists of a total of 34,662 images of 82 ID cards. For 46 people who do not exist, we produced 37 registration cards and 45 driver's license cards. Of the 46 people, 36 have both types of ID cards. For the three labels shown in Figure 2, our dataset includes 13,746 genuine, 13,729 screen, and 7,187 print images. In this section, we describe how to make replica ID cards, how to take photos of the ID cards, and finally, the ethical considerations when creating the ID cards and the image dataset.

3.1 ID Card Creation

The ID cards that we leveraged for generating the dataset were produced based on the template of Korean ID cards. Figure 3 shows two kinds of ID cards provided by the Korean government, in which the numbers in parentheses represent the following in order:

(1) The first indicates its ID card type. The left side is for a driver's license card and the right side is for a registration card.
(2) In the second area, the upper texts represent the name of the ID card owner. While driver's license cards only display the Korean name, registration cards include both the Korean name and the corresponding Chinese character name. The

[1] https://doi.org/10.5281/zenodo.8034016

numbers displayed beneath the name represent the registration number, which is similar to Social Security Number (SSN) in the U.S.
(3) The third text indicates the home address of the ID owner.
(4) Lastly, the two images on each ID card are the pictures of the ID card owner. The images within an ID card are identical but different in size. However, the registration cards issued before 2020 do not contain the small image.

Since ID cards involve the most private information of the ID owners, as listed above, we placed privacy-preserving as the top priority when producing ID cards. On top of that, we focused on our dataset being unbiased, thereby producing the ID cards to represent various ethnicities and ages. We handled the text and image information on ID cards in the following manner: For the text, we used arbitrary text for name, registration number, and home address to represent non-existent information. In registration cards, the Korean name does not match the Chinese character name in terms of pronunciation. As for the photos, we applied pictures of faces of non-existent humans for the ID cards by accessing [5]. The faces depict a wide range from children to the elderly with diverse ethnic backgrounds.

3.2 Dataset Acquisition

We acquired the KID34K image dataset by taking photos of the replica ID cards described in Section 3.1. We applied three real scenarios: taking photos of genuine ID cards, ID card images displayed on a screen, and ID card images printed on paper. We used a total of 8 digital devices, including tablets and monitors, for displaying ID images and 2 printers for printing ID images. All the images in our dataset were taken by 12 smartphones. We list the devices used for each case as follows:

For displaying: Galaxy Book, Galaxy Tablet S6, Galaxy Tap S6 Lite, Galaxy Tap Pro S, iPad Pro 4th, iPad Air 4th, iPad Air 5th, BenQ monitor

For printing: Samsung SL-X7400LX, HP LaserJet Pro MFP M477fdw

For taking images: iPhone 7, iPhone 12 Pro, iPhone 12 Pro Max, Galaxy S6, Galaxy S8, Galaxy S10e, Galaxy S20, Galaxy s21, Galaxy S22, Galaxy A53, Galaxy Note20, Galaxy Zflip3

Note that all the ID images featured in this paper contain a watermark with the text "SAMPLE", but our dataset does not.

3.3 Ethical Considerations

Our dataset is available for research purposes only and can be accessed through its assigned digital object identifier (DOI) on Zenodo [4]. All researchers involved in this study underwent Institutional Review Board (IRB) training provided by our institution and adhered to the data collection procedures outlined by the IRB. Real ID cards, registration cards and driver's licenses, were collected from participants who willingly expressed their desire to participate. Such real ID cards were utilized exclusively for the purpose of this study.

4 EXPERIMENT

In this section, we evaluate the performances of the classification models trained on a dataset consisting of real ID images only, and trained on both the real ID dataset and our KID34K dataset to show the effect of our dataset.

Table 1: Datasets used for the experiment.

Dataset	Trainset			Testset		
	Notation	Persons	IDs	Notation	Persons	IDs
KID34K	$\mathcal{K}_{train}$	35	70	$\mathcal{K}_{test}$	11	12
The first real	$\mathcal{R}^{I}_{train}{}^{\star}$ / $\mathcal{R}^{I}_{train}$	2 / 4	2 / 4	$\mathcal{R}^{I}_{test}$	2	2
The second real	$\mathcal{R}^{II}_{train}$	N/A	N/A	$\mathcal{R}^{II}_{test}$	3	3

4.1 Experimental Setup

Baselines. We used five models for the binary classification task: ResNet18 [8], ResNet34 [8], EfficientNet [15], MobileNet V2 [13], and DenseNet [9]. Besides, we employed the Discrete Fourier Transform (DFT) [14], which was used in prior studies [7].

Data augmentation. We employed two representative augmentation techniques: CutMix [18] and Cutout [3]. CutMix is a commonly used approach in image classification tasks, which involves blending or replacing regions of one image with corresponding regions from another image to enhance the model's performance. Meanwhile, Cutout involves randomly masking out square regions within an image, occluding specific portions and encouraging the model to focus on other relevant features. We followed the augmentation configuration specified in the CutMix, which included horizontal flip, lighting, and jittering as augmentations.

Dataset. In the experiment, we used three datasets: the KID34K dataset, and two real datasets. For the two real datasets, we collected a total of nine ID cards from nine volunteers and split the ID cards into two sets. Table 1 shows the notation, the number of individuals, and the count of ID cards corresponding to each dataset, where a maximum of two ID cards-a registration and a driver's license card-correspond to a single person. For the first real dataset, we utilized both its subset and the entire set in the experiment, so we distinguished the two cases as $\mathcal{R}^{I}_{train}{}^{\star}$ and $\mathcal{R}^{I}_{train}$. The second real dataset was used solely for model testing. We split the training and test set of the datasets according to the number of persons so that the baselines can only consider the IDs belonging to a person as either *seen* or *unseen*. The total number of images in the first real dataset is 8,699, consisting of 3,337 genuine, 3,175 screen, 2,187 print images. For the second real dataset, the total number of images is 240, comprising 96 genuine, 96 screen, 48 print images.

Classification. Given the three labels, which are genuine, screen, and print, it is possible to consider a three-class classification. However, our focus is on addressing the practical scenario where the system requires to distinguish whether an image of an ID card is genuine or not. Therefore, we apply binary classification, which classifies images taken of genuine vs. digitally represented ID cards.

Implementation details. We measured the classification performances of the baselines on our dataset by using the models pretrained on the ImageNet dataset [2]. We resized the input images to 512×800 (H×W). All training and test were performed with a batch size of 16, a learning rate of 1e-3, and the cosine annealing with warm up scheduler [10].

4.2 Results

Performance. Table 2 presents the performance of three different scenarios: 1) training on the real dataset with the application of DFT, 2) training on the real dataset without DFT, and 3) training on a

Table 2: Performance of the models. For augmentation methods, α only involves the horizontal flip while β encompasses horizontal flip, lighting, and jittering. Each cell indicates the test results on the KID34K, the first real dataset, and the second real dataset in order.

	Datasets for training	ResNet18	ResNet34	DenseNet-121	MobileNet V2	EfficientNet
with DFT	$\{\mathcal{R}^I_{train}{}^{\star}\}$	83.62 / 62.08 / 55.42	84.53 / 62.5 / 58.19	72.48 / 60. / 53.47	84.6 / 60. / 53.59	87.08 / 60. / 53.44
with DFT	$\{\mathcal{R}^I_{train}\}$	88.19 / 60. / 56.27	88.02 / 60. / 53.81	76.04 / 60. / 53.45	87.98 / 60. / 53.52	90.92 / 60. / 53.44
without DFT	$\{\mathcal{R}^I_{train}{}^{\star}\}$	67.28 / 97.87 / 87.92	67.33 / 95.63 / 94.17	66.16 / 97.94 / 90.	64.7 / 96.26 / 87.5	68.58 / 87.88 / 93.33
without DFT	$\{\mathcal{R}^I_{train}\}$	70.33 / 99.86 / 80.42	67.95 / 99.76 / 90.83	68.18 / 99.48 / 84.58	66.09 / 99.65 / 73.5	69.47 / 99.06 / 93.33
CutMix + β	$\{\mathcal{R}^I_{train}{}^{\star}\} \cup \{\mathcal{K}_{train}\}$	96.47 / 92.25 / 92.5	95.12 / 92.67 / 95.42	94.79 / 94.87 / 96.67	91.5 / 90.05 / 85.83	94.95 / 88.09 / 88.75
Cutout + β	$\{\mathcal{R}^I_{train}{}^{\star}\} \cup \{\mathcal{K}_{train}\}$	95.73 / 95.88 / 95.83	95.57 / 93.82 / 90.42	94.94 / 96.12 / 95.83	93.35 / 93.54 / 94.17	95.94 / 96.33 / 96.67
β	$\{\mathcal{R}^I_{train}{}^{\star}\} \cup \{\mathcal{K}_{train}\}$	96.06 / 94.38 / 90.83	94.73 / 93.22 / 91.67	95.41 / 96.3 / 95.83	95.78 / 92.7 / 88.33	95.8 / 96.09 / 93.33
α	$\{\mathcal{R}^I_{train}{}^{\star}\} \cup \{\mathcal{K}_{train}\}$	94.95 / 96.37 / 97.5	94.05 / 95.49 / 95.42	93.69 / 96.19 / 95.83	96.48 / 95.21 / 95.83	96.75 / 98.43 / 96.67
CutMix + β	$\{\mathcal{R}^I_{train}\} \cup \{\mathcal{K}_{train}\}$	96.24 / 98.71 / 97.5	94.37 / 97.45 / 98.75	95.27 / 98.43 / 99.17	94.93 / 97.07 / 97.5	93.97 / 98.04 / 97.92
Cutout + β	$\{\mathcal{R}^I_{train}\} \cup \{\mathcal{K}_{train}\}$	96.16 / 98.67 / 94.17	95.21 / 98.43 / 98.33	96.39 / 99.09 / 99.17	96.19 / 97.59 / 99.17	96.3 / 99.34 / 95.
β	$\{\mathcal{R}^I_{train}\} \cup \{\mathcal{K}_{train}\}$	96.15 / 98.57 / 95.	95.15 / 98.57 / 99.58	93.95 / 98.6 / 97.5	96.59 / 97.87 / 97.08	96.22 / 99.3 / 91.67
α	$\{\mathcal{R}^I_{train}\} \cup \{\mathcal{K}_{train}\}$	95.37 / 98.57 / 95.42	96.57 / 98.95 / 97.92	96.39 / 99.09 / 99.17	97.25 / 98.18 / 97.5	95.68 / 99.06 / 96.25

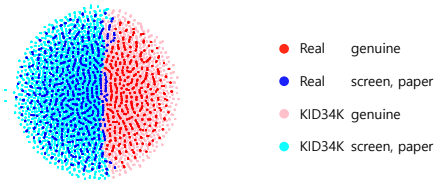

Figure 4: t-SNE [17] visualization.

combination of our KID34K dataset and the real dataset employing various augmentation techniques.

We find out that the models with DFT generally show lower performance than the models without DFT. This demonstrates that our test sets are more challenging than previous studies [7] that showed high performance by applying DFT. For the models without DFT, the test results on $\mathcal{R}^I_{test}$ show superior performance, while they show lower performance on $\mathcal{R}^{II}_{test}$, indicating the presence of unstable generalization when solely training on $\mathcal{R}^I_{train}$. On the other hand, when both $\mathcal{K}_{train}$ and $\mathcal{R}^I_{train}$ are employed for training, the performance shows enhanced accuracy and stability for both $\mathcal{R}^I_{test}$ and $\mathcal{R}^{II}_{test}$.

Moreover, in most of the experimental results, we discover that the accuracy of $\mathcal{K}_{test}$ is analogous to that of $\mathcal{R}^I_{test}$ and $\mathcal{R}^{II}_{test}$, implying that our KID34K dataset is feasible to deploy to the applications of real ID verification systems.

Discussion. In the field of image classification tasks, CutMix is commonly employed to enhance the robustness and generalization capability of models. However, our experiments revealed that only applying horizontal flip to our dataset resulted in the best performance. This can be attributed to the inherent similarities observed among the classes in our task, which could potentially result in the loss of significant features and patterns when applying cropping or replace processes involved in CutMix and Cutout.

Our experiments also demonstrated that training on $\mathcal{R}^I_{train}$ yielded better results compared to using $\mathcal{R}^I_{train}{}^{\star}$. This provides evidence that the performance improves as the quantity of data increases, highlighting the positive correlation between the amount of data and the model's generalization performance. Finally, Figure 4 visualizes the model evaluation results on our KID34K dataset and the real dataset. This shows that there is little difference between the real and the proposed replica (KID34K) ID cards distributions.

5 CONCLUSION

We proposed a novel dataset specifically designed to verify the authenticity of the ID cards uploaded by users to the online ID card verification system. Our dataset included genuine ID card images, as well as images of ID cards displayed on screen and printed on paper. To ensure privacy while still allowing accessibility to the dataset, we created replica ID cards that resemble real ID cards. We conducted extensive experiments to evaluate the performance of our proposed dataset. These experiments involved training and testing on various subsets of the dataset, including both genuine ID card images and digitally represented ID card images. Through comprehensive experiments and analysis of the results, we were able to validate the effectiveness and robustness of our dataset in accurately classifying and detecting instances of identity misuse. We will distribute this dataset for research purposes only, with the aim of facilitating advancements and innovations in the field. By providing access to this dataset, we seek to foster collaborations and encourage researchers to explore novel approaches in addressing challenges related to ID card fraud detection.

ACKNOWLEDGMENTS

We thank Kwansik Yoon, Yuna Seo, and Minki Hong at Samsung SDS for creating this dataset for this research. This work was partly supported by Institute for Information & communication Technology Planning & evaluation (IITP) grants funded by the Korean government MSIT: (No. 2022-0-01199, Graduate School of Convergence Security at Sungkyunkwan University), (No. 2022-0-01045, Self-directed Multi-Modal Intelligence for solving unknown, open domain problems), (No. 2022-0-00688, AI Platform to Fully Adapt and Reflect Privacy-Policy Changes), (No. 2021-0-02068, Artificial Intelligence Innovation Hub), (No. 2019-0-00421, AI Graduate School Support Program at Sungkyunkwan University), and (No. RS-2023-00230337, Advanced and Proactive AI Platform Research and Development Against Malicious deepfakes).

REFERENCES

[1] Daniel Benalcazar, Juan E Tapia, Sebastian Gonzalez, and Christoph Busch. 2023. Synthetic ID Card Image Generation for Improving Presentation Attack Detection. *IEEE Transactions on Information Forensics and Security* 18 (2023), 1814–1824.

[2] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*. 248–255. https://doi.org/10.1109/CVPR.2009.5206848

[3] Terrance DeVries and Graham W Taylor. 2017. Improved regularization of convolutional neural networks with cutout. *arXiv preprint arXiv:1708.04552* (2017).

[4] European Organization For Nuclear Research and OpenAIRE. 2013. Zenodo. https://doi.org/10.25495/7GXK-RD71

[5] This Person Does Not Exist. 2021-2023. Random Face Generator(This Person Does Not Exist). https://this-person-does-not-exist.com/en

[6] Sebastian Gonzalez and Juan Tapia. 2023. Improving Presentation Attack Detection for ID Cards on Remote Verification Systems. *arXiv preprint arXiv:2301.09542* (2023).

[7] Sebastian Gonzalez, Andres Valenzuela, and Juan Tapia. 2020. Hybrid two-stage architecture for tampering detection of chipless id cards. *IEEE Transactions on Biometrics, Behavior, and Identity Science* 3, 1 (2020), 89–100.

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[9] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. 2017. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4700–4708.

[10] Naoki Katsura. 2020. pytorch-cosine-annealing-with-warmup. https://github.com/katsura-jp/pytorch-cosine-annealing-with-warmup.

[11] KoROAD. 2018-2023. Driver's License Acquisition Process in South Korea Governed by the Government of South Korea. https://www.safedriving.or.kr/main.do

[12] Raghavendra Mudgalgundurao, Patrick Schuch, Kiran Raja, Raghavendra Ramachandra, and Naser Damer. 2022. Pixel-wise supervision for presentation attack detection on identity document cards. *IET Biometrics* 11, 5 (2022), 383–395.

[13] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. 2018. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4510–4520.

[14] L. Tan and J. Jiang. 2013. *Digital Signal Processing: Fundamentals and Applications*. Academic, Cambridge, MA, USA. 87–136 pages.

[15] Mingxing Tan and Quoc Le. 2019. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*. PMLR, 6105–6114.

[16] The Korean Ministry of the Interior and Safety. 1998-2023. South Korea's ID Card. https://mois.go.kr/frt/sub/a06/b06/IDCard/screen.do

[17] Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of machine learning research* 9, 11 (2008).

[18] Sangdoo Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, and Youngjoon Yoo. 2019. Cutmix: Regularization strategy to train strong classifiers with localizable features. In *Proceedings of the IEEE/CVF international conference on computer vision*. 6023–6032.