A system for verifying non-standard personal identity documents using deep learning

ICDAR2021-IJDAR Journal Track

A. Maurya, Amith P, S. Dubey, O. Manjrekar (*),
A. Furtado, K. Arya

Agenda

1. Introduction
2. The problem
3. Related work
4. Our solution
5. Experiments and results
6. Conclusion and Future Work
7. About the journal
8. References

Introduction

1. **Identity verification** is an essential requirement in many processes and procedures.
2. Examples of processes include banking, availing loans, car rental, student benefits and scholarships.
3. Non-trivial processes require a person to present **personal identity documents (ID)** for verification, provided by authorized organizations.
4. Verification is automatic (and remote) or manual.

Introduction

6. Recent research suggests methods for verifying **standard IDs** such as **passports, national identity cards,** etc.
7. These methods capture document image using mobile phone camera **[1]** and use computer vision techniques.
8. Some methods that exist:
    a. Extract text using OCR from any given image **[2] [3]**
    b. Identify class of the document **[4] [5] [6]** for querying stored layout **[7]**
    c. Verify image on the ID card with live-feed **[8]**

The problem

1. **Non-standard identity documents** are provided by local or regional organizations (e.g. student identity cards).
2. Student identity cards (an example of non-standard IDs) differ from school to school in:
   a. Form factor
   b. Layout
   c. Fonts
   d. Color
   e. Graphics (college emblem, etc.)
3. Besides noise due to capturing images using mobile phone camera

Fig. 1 Student identity cards of some engineering colleges in India

The problem

6. Previous methods won't necessarily work for non-standard identity cards such as **student identity cards** due to 3 reasons:
    a. **Large visual variation**
    b. **No layout information** thus making "matching" of textual information non-trivial
    c. **Uncontrollable conditions** when capturing the image from a mobile phone camera
7. Verification of Non-standard IDs has not been well explored

Related Work

Prior Work	Methodology Proposed	Problem
1. Amin, A., Fischer, S.: A document skew detection method using the hough transform. Pattern Analysis & Applications 2. Fang, X., Fu, X., Xu, X.: Id card identification system based on image recognition	1. Hough Transform, Projection Profile 2. Detects skew of within range of 45 degrees.	1. Performs well on scanned document with Plain Background 2. Fails when Image are flipped to 180 degrees
1. Usilin, S., Nikolaev, D., Postnikov, V., Schaefer, G.:Visual appearance based document image classification. 2. De las Heras, L.P., Terrades, O.R., Llados, J., Fernandez-Mota, D., Canero, C.: Use case visual bag-of-words techniques for camera based identity document classification. 3. Awal, A.M., Ghanmi, N., Sicre, R., Furon, T.: Complex document classification and localization application on identity document image 4. Kumar, J., Doermann, D.: Unsupervised classification of structurally similar document images.	1. Use National Emblem, face detection. 2. Bags of Word on ID like Passport, Driving Licence 3. Key Point Detection 4. Classifying Structural Similar Documents	1. College ID cards are non-standardized unlike Passport or DL. 2. College ID cards are also structurally different, different shape and size. 3. Key points are needed for every class.
1. Xu, J., Wu, X.: A system to localize and recognize texts in oriented id card images 2. Liem, H.D., Minh, N.D., Trung, N.B., Duc, H.T., Hiep,P.H., Dung, D.V., Vu, D.H.: Fvi: An end-to-end viet-namese identification card detection and recognition in images.	1. Haar Cascade Classifier to Detect text 2. Used EAST, CRAFT model for text Detection 3. Used STAR or CRNN model for text Recognition	1. EAST and CRAFT model have slow inference time as compared to new methods present. 2. STAR model did not perform well on our dataset.
1. Hall, P.A., Dowling, G.R.: Approximate string matching. 2. Tregoat,J.:An introduction to fuzzy string matching.	1. Previous work lacks in post processing of OCR Text	1. OCR data is prone to have spelling mistakes.

Our solution

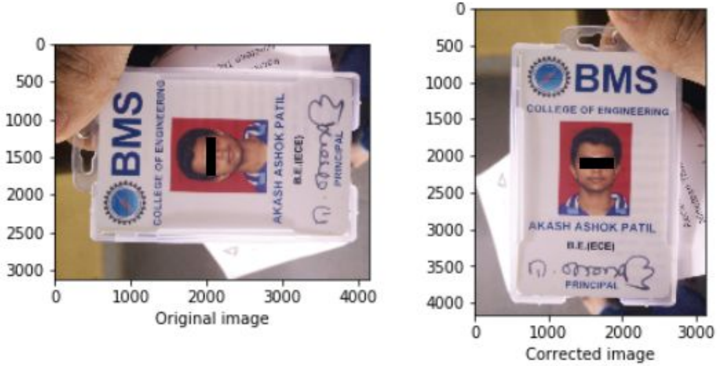
The a complete pipeline with following stages

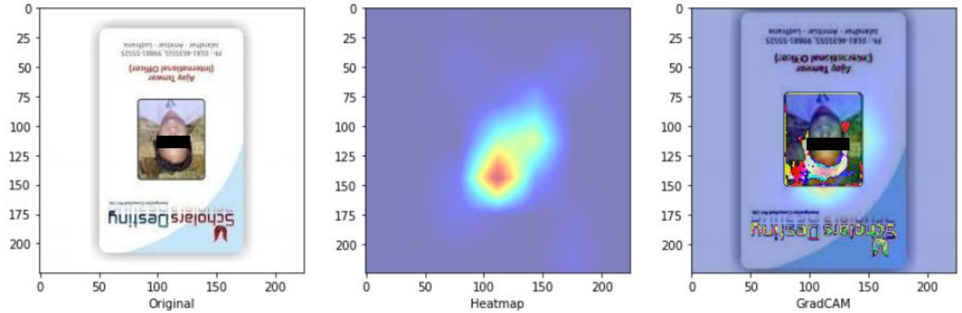
1. Orientation Correction (Rotate-Net)
2. Image Classification
3. Text Extraction
    a. Detection
    b. Recognition
4. String Matching (verify with reference information)

1. Orientation Correction (Rotate-Net)

A DCNN based model to automatically correct the Image Orientation.

Model Name	Architecture used	Training parameter	Output classes
Rotate Net	Resnet-50 Pre-trained on ImageNet dataset	Adam Optimizer trained with 30 epochs, Trained on Id cards and face images.	[0, 90, 180, 360]



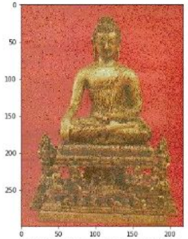


2. Image Classification

A DCNN based model to classify college ID cards.

Model Name	Architecture used	Training parameter	Output classes
College ID card Classifier	Resnet-50, with last three layer changed as [256, 128, 4] nodes Pre-trained on ImageNet dataset	Adam Optimizer trained with 50 epochs, Trained on Id cards, adhaar card, pan card, other	[College_ID, Adhaar_card, Pan_card, Others]


Predicted Class : Adhaar_Card


Predicted Class : Other


Predicted Class : Other


Predicted Class : College_ID


Predicted Class : College_ID


Predicted Class : Pan_card

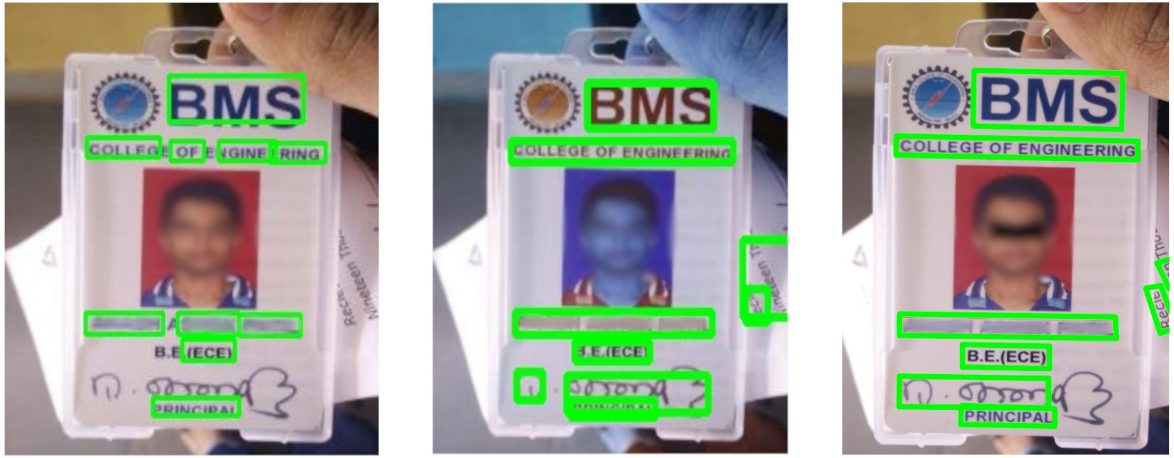

Predicted Class : Adhaar_Card


Predicted Class : College_ID

3. Text Extraction

1. Text Detection

a. Compared EAST, CRAFT and DBNET.
b. Found DBNET to have best inference time and better results.



(a) (b) (c)

Output of (a) EAST text detector, (b) CRAFT text detector and (c) DBNet-50 text detector

3. Text Extraction

2. Text Recognition

a. Compared STARNet and CRNN.

b. Performance of STAR was not satisfactory for our dataset.

(a)

{'BMS': 0.9833903312683105, 'COLLEGEOFENGMEERING': 0.618780255317688, 'and': 0.013286435976624489, 'is': 0.02926507405936718, [redacted]: 0.3788427412509918, 'B.E.(ECE)': 0.938913881778717, 'of': 0.05884723365306854, 'PRINCIPAL': 0.9993307590484619}

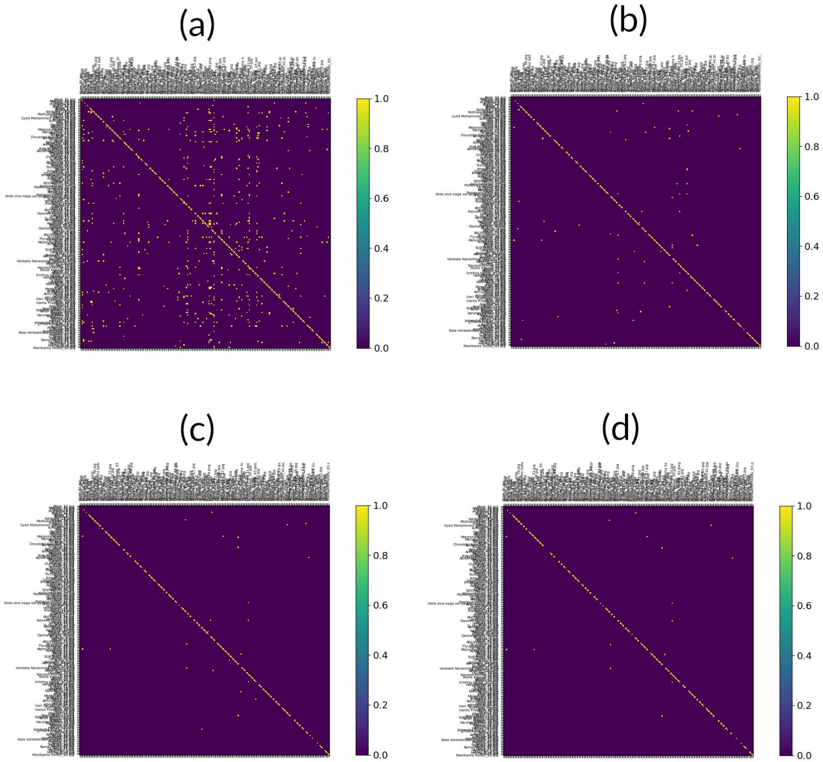
(b)

{ 'BMS': 0.9932370185852051,"COLLEGEOFENGINEERING": 0.9946616888046265, "uaaauN": 0.8223435878753662,[redacted]": 0.988580584526062, "B.E. (ECE)": 0.920192301273346,"PRINCIPAL": 0.9925515651702881}

Output of text recognition model (a) STAR and (b) CRNN

4. String Matching using custom Algorithm

1. Exact string matching is not possible as OCR information can have spelling mistakes.
2. We build custom Algorithm, which was weighted sum of college name (60), candidate's name (35), and Department name (5).
3. Different threshold value was applied and results are presented below.



Confusion matrices for the custom fuzzy string matching algorithm over different thresholds (a) 60, (b) 70, (c) 75 and (d) 80

Experiments and Results

Datasets

1. **Training and validation dataset**
   a. ID Cards from e-Yantra MOOC = 1800 images
   b. Mined from web (Aadhar, PAN and Others) = 1700 images
   c. CSV file for verification provided by MOOC admin
2. **Testing**
   a. ID Cards from e-YRC 21 = 8493 images
   b. CSV file for verification provided by e-YRC admin

Experiments and Results (Stages)

1. **Individual stage performance** is evaluated on **validation subset** of "**Training and validation dataset**".
2. Stages
    a. Image classification - 98.8% **Accuracy**
    b. Orientation correction - 91% **Accuracy**
    c. Off-the shelf text detection (accuracy is not reported)
    d. Matching (200 random images at threshold 75)
        i. Accuracy - 99.80%
        ii. Precision - 85.294%
        iii. Recall - 72.864%
        iv. F1 score - 0.78591

Experiments and Results (E2E)

1. End to end performance is evaluated on **“Testing dataset” (8493 images).**
2. System specs
   a. Intel(R) Core(TM) i7-7700 CPU @ 3.60GHz
   b. RAM 16 GB
   c. GTX 1080Ti 12 GB
3. Time per ID card
   a. 4381 images / 169m 7s => 2.31 seconds per image

Experiments and Results (E2E)

1. As per manual annotations, out of 8493 images
    a. 7884 (92.83%) are “**Accepts**”
    b. 609 (7.171%) are “**Rejects**”
2. As the data is skewed, performance is reported on original (8493) as well as randomly downsampled dataset (1218).
3. Accuracy is not reliable.
4. We use Precision, Recall, F1 score, Sensitivity, Specificity, G-Mean and PR AUC (average precision).

Table **1** Confusion matrix for original dataset (8493)

	Actual Positive	Actual Negative
Predicted Positive	575 (True Positive)	1214 (False Positive)
Predicted Negative	34 (False Negative)	6670 (True Negative)

Table **2** Confusion matrix for downsampled dataset (1218)

	Actual Positive	Actual Negative
Predicted Positive	575 (True Positive)	93 (False Positive)
Predicted Negative	34 (False Negative)	516 (True Negative)

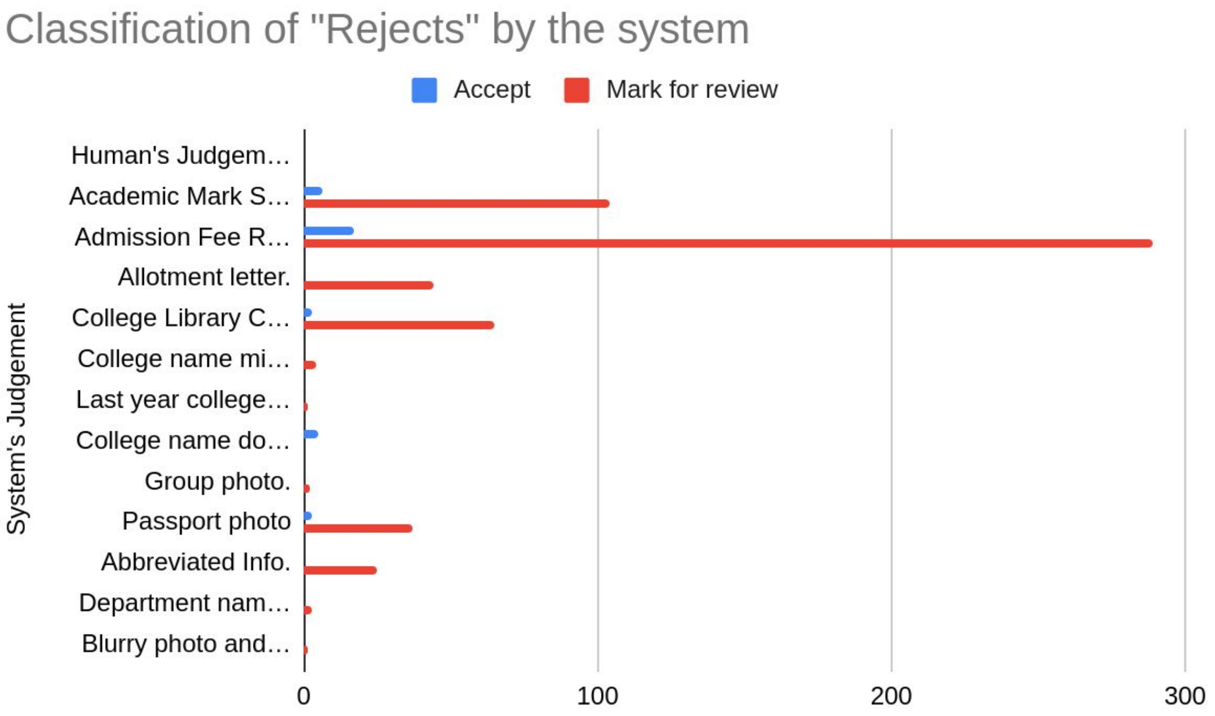


Fig. 2 Classification of Rejects (609 images) by the system

Table **3** Performance metrics calculated on original and downsampled dataset

Sr. No.	Metric	Formula	Original	Downsampled
1	Sensitivity	TP / (TP + FN)	94.417%	94.417%
2	Specificity	TN / (TN + FP)	84.602%	84.729%
3	G-Mean	Sqrt(sensitivity * specificity)	0.89375	0.89442
4	Precision	TP / (TP + FP)	32.141%	86.078%
4	Recall	**Same as sensitivity**	94.417%	94.417%
5	Accuracy	TP + TN / (TP + FP + TN + FN)	85.306%	89.573%
6	F1 score	2 * recall * precision / (recall + precision)	0.47957	0.90054

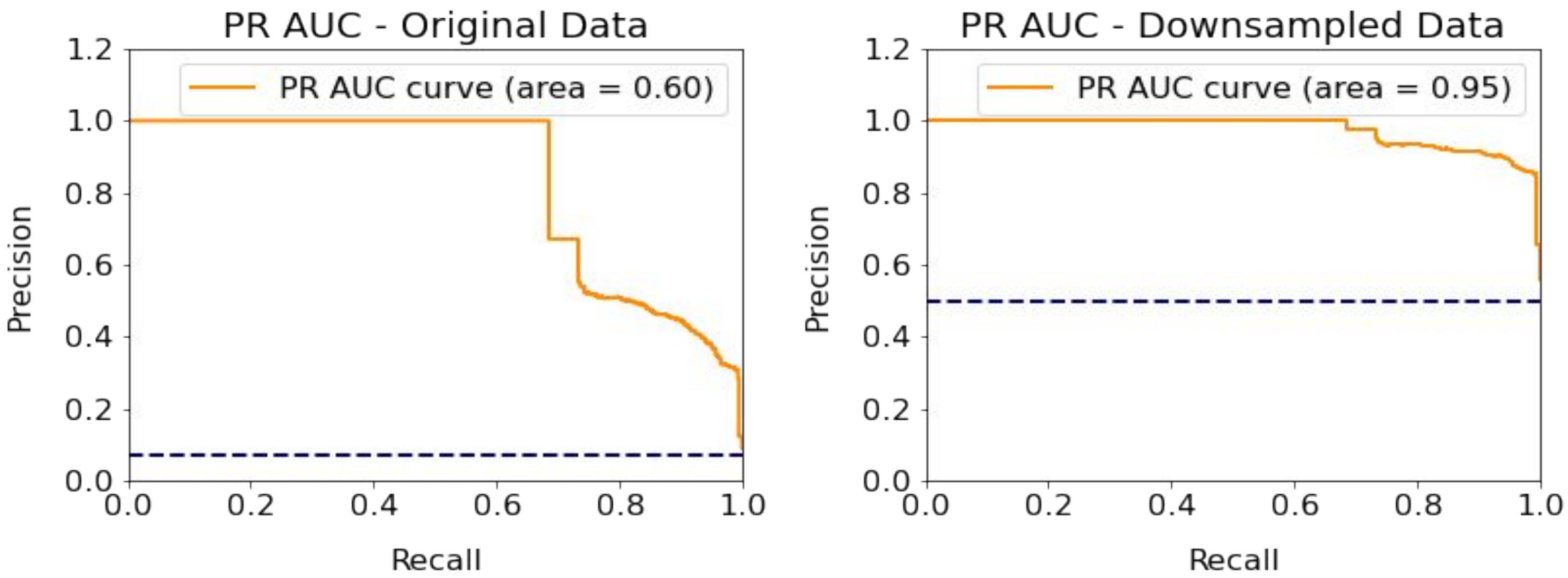


Fig. 3 PR curves and PR AUC for (a) original dataset and (b) downsampled dataset

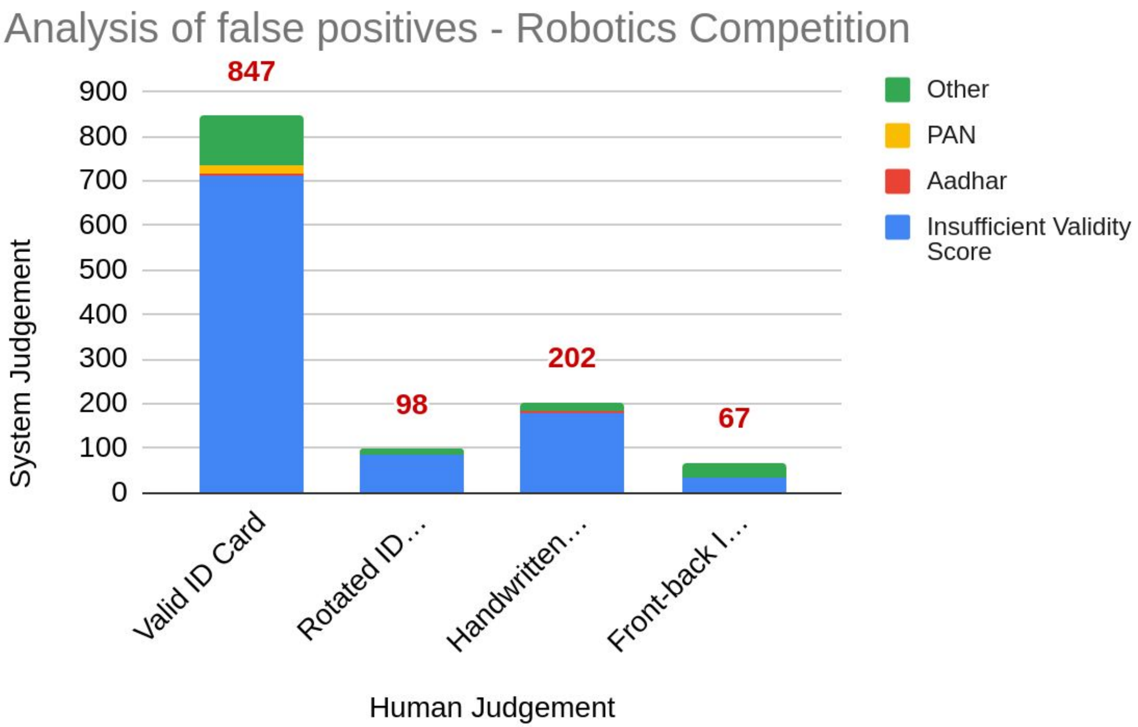


Fig. 4 Analysis of false positives on the original testing dataset

Conclusion and Future Work

1. The study described a system with superior performance on a not well explored problem of verifying non-standard ID cards.
2. Future directions for research:
   a. Querying entities such as name, address, phone number from ID cards.
   b. A way to fetch, differentiate and validate dates present on student identity cards.
   c. Handling handwritten text.
   d. Robust ID card classification by checking presence of faces inside ID cards.
   e. Orientation correction for finer angles.

ICDAR-IJDAR Journal Track

The paper has been submitted to S.I. of IJDAR known as ICDAR-IJDAR journal track. Following are the rankings:

1. ICDAR
   a. H Index 68
   b. A2 (Qualis)
   c. A (Core 2020)
2. IJDAR
   a. H Index 47
   b. Q2
   c. SJR 2019 - 0.49

References

1. Liang, J., Doermann, D., Li, H.: Camera-based analysis of text and documents: A survey (2005). DOI 10.1007/ s10032-004-0138-z. URL https://link.springer.com/ article/10.1007/s10032-004-0138-z
2. Pratama, M.O., Satyawan, W., Fajar, B., Fikri, R., Hamzah, H.: Indonesian id card recognition using convolutional neural networks. Proceeding of the Electrical Engineering Computer Science and Informatics 5(1), 178–181 (2018)
3. Aprillian, H.D.D., Purnomo, H.D., Purwanto, H.: Utilization of optical character recognition technology in reading identity cards. International Journal of Information Technology and Business 2(1), 38–46 (2019)
4. Usilin, S., Nikolaev, D., Postnikov, V., Schaefer, G.: Visual appearance based document image classification. In: 2010 IEEE International Conference on Image Processing, pp. 2133–2136 (2010). DOI 10.1109/ ICIP .2010 .5652024
5. De las Heras, L.P., Terrades, O.R., Llados, J., Fernandez Mota, D., Canero, C.: Use case visual bag-of-words techniques for camera based identity document classification. In: 2015 13th International Conference on Document Analysis and Recognition (ICDAR), pp. 721–725. IEEE (2015)
6. Awal, A.M., Ghanmi, N., Sicre, R., Furon, T.: Complex document classification and localization application on identity document images. In: 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), vol. 1, pp. 426–431. IEEE (2017)

References

7. Simon, M., Rodner, E., Denzler, J.: Fine-grained classification of identity document types with only one example. In: 2015 14th IAPR International Conference on Machine Vision Applications (MVA), pp. 126–129 (2015). DOI 10.1109/MVA.2015.7153149
8. Hanh Vu, Thuyen Nguyen Hoang, and Nguyen Van Nam "An efficient deep face matching method for ID and selfie photos: SIRFace", Proc. SPIE 11519, Twelfth International Conference on Digital Image Processing (ICDIP 2020), 115191Z (12 June 2020); https://doi.org/10.1117/12.2573327

Thank you!