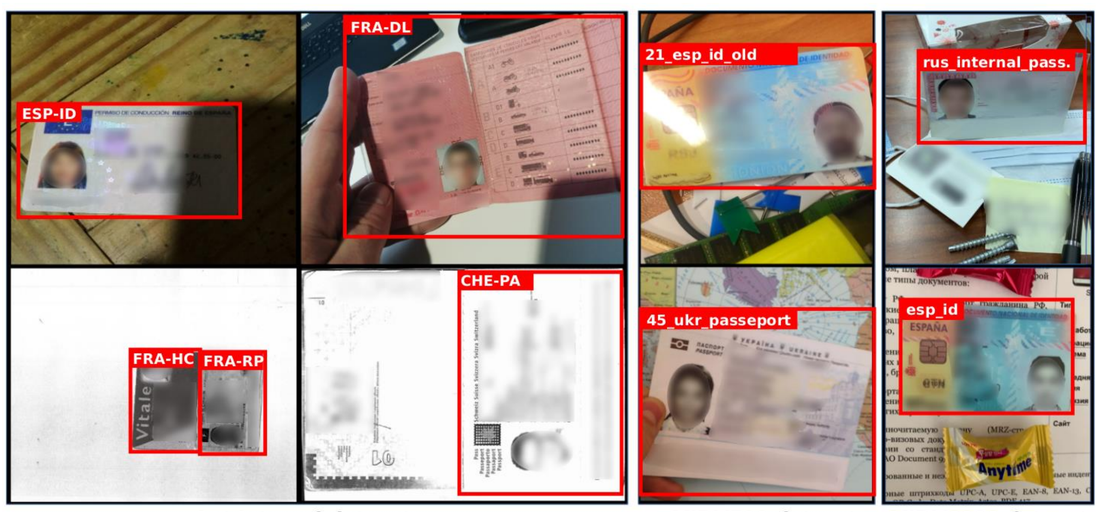
One-Shot Classification of ID Documents

Florian Arrestier, Guillaume Chiron and Ahmad Montaser Awal
AriadNEXT – Cesson-Sévigné, France
{florian.arrestier, guillaume.chiron, montaser.awal}@ariadnext.com

Objectives and Challenges

- Localize and **classify** ID documents in various capture conditions (smartphone, scans, webcam)
- Support new documents with only a few samples available.
- Support 1000+ identity document models.
- Diversity of document models and background.



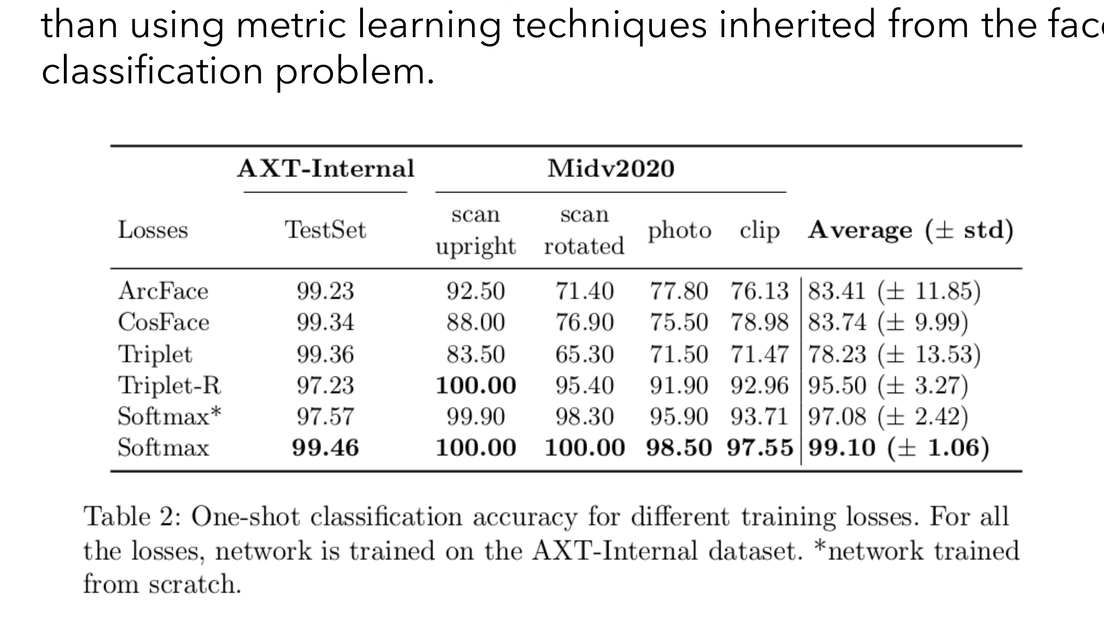
Axt-Internal dataset (industrial) | **MIDV500 dataset** (academic) | **MIDV2020 dataset** (academic)

Impact of training loss

- Using different losses for training reveals that training a network for pure classification using the Softmax loss yields better results than using metric learning techniques inherited from the face classification problem.

Losses	AXT-Internal TestSet	Midv2020 scan upright	Midv2020 scan rotated	Midv2020 photo	Midv2020 clip	**Average (± std)**
ArcFace	99.23	92.50	71.40	77.80	76.13	83.41 (± 11.85)
CosFace	99.34	88.00	76.90	75.50	78.98	83.74 (± 9.99)
Triplet	99.36	83.50	65.30	71.50	71.47	78.23 (± 13.53)
Triplet-R	97.23	**100.00**	95.40	91.90	92.96	95.50 (± 3.27)
Softmax*	97.57	99.90	98.30	95.90	93.71	97.08 (± 2.42)
Softmax	**99.46**	**100.00**	**100.00**	**98.50**	**97.55**	**99.10 (± 1.06)**

Table 2: One-shot classification accuracy for different training losses. For all the losses, network is trained on the AXT-Internal dataset. *network trained from scratch.



(a) Triplet loss (b) Triplet-R loss (c) Softmax loss (d) ArcFace loss

Fig. 3: Scatter plots showing the cosine similarity to the most similar example from the same and different classes of the AXT-Internal test dataset, for 4 different losses.

Generalization Capabilities

- Training on a varied dataset yield better results on unseen datasets.

Train \ Test		AXT-Internal TestSet	Midv2020 scan rotated	Midv2020 scan upright	Midv2020 photo	Midv2020 clip	Midv500 * All	Midv500 ** TestSet
AXT-Internal	100%	**99.80**	100.00	100.0	**98.80**	**98.09**	**99.70**	**99.71**
	75%	96.20	100.00	100.0	97.70	97.55	99.57	99.54
	50%	92.52	100.00	99.70	95.80	96.96	99.22	99.29
	25%	84.20	98.60	98.00	91.10	94.87	98.59	98.87
Midv500*	All	71.21	99.80	97.70	94.30	96.21	99.96†	99.94†
Midv500**	TrainSet	49.00	79.70	79.10	58.40	58.19	88.79	87.55
Midv2020	All split	66.11	100.00†	100.00†	99.00†	99.75†	94.35	95.43

Table 4: Cross dataset one-shot classification accuracies. * filtered using [13] criterion, ** filtered using [23] criterion. † test samples overlap with training samples, values only serves as references of "ideal classification accuracy".

- Increasing the number of training classes help to better generalize to an increasing number of reference models.

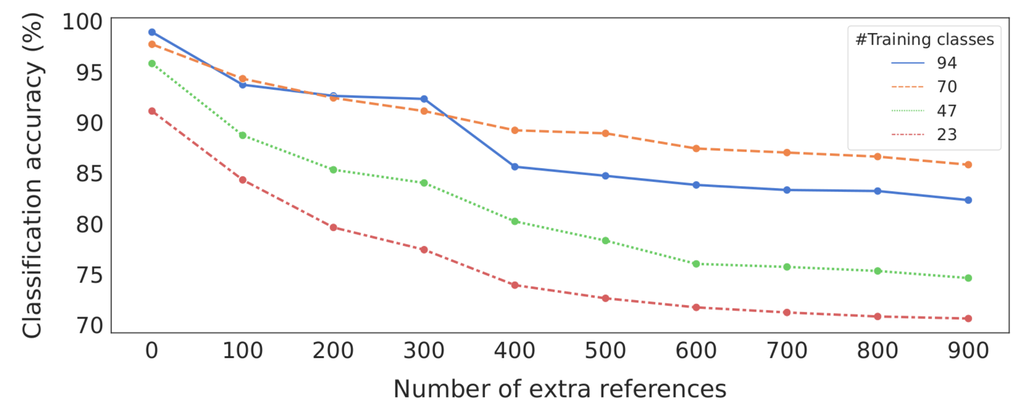


Fig. 4: One-shot classification accuracy in regard to the number of reference classes on the Midv2020 photo dataset.

Comparison with SOTA

- Our end-to-end localization + classification pipeline outperforms the current state-of-the-art approaches on all acadamic and private datasets.

Method \ Dataset	AXT-Internal TestSet	Midv2020 scan rotated	Midv2020 scan upright	Midv2020 photo	Midv2020 clip	Midv500 * All	Midv500 ** TestSet
RFDoc [23]	-	-	-	-	-	-	93.46
SURF + Filters [5]	-	100.00 [9]	100.00 [9]	95.10 [9]	64.38 [9]	97.20 [15]	-
Beblid256 [9]	-	100.00	99.90	98.20	81.75	-	92.78 [23]
Beblid512 [9]	-	100.00	**100.00**	**98.70**	84.48	-	93.51 [23]
EffDet + Mnasnet [13]	94.98	-	-	-	-	93.91	-
EffDet + Ours	**99.34**	**100.00**	99.40	96.10	**95.78**	**99.50**	**99.47**

Table 5: Results of the proposed one-shot classification approach and other published classification methods, a - means that no results for this particular dataset was available.

Conclusion

- **Diversity matters:** The greater diversity in the axt-internal private dataset compared to the academic datasets show greater generalization capabilities on unseen datasets.
- **Simplicity is always a good thing:** Surprisingly, on simple use cases like this one, the simple softmax loss outperforms modern alternatives tailored for face classification.