Human or Machine? It Is Not What You Write, But How You Write It

Luis A. Leiva[1], Moises Diaz[2], Miguel A. Ferrer[3], Réjean Plamondon[4]

[1]Aalto University, Finland. name.surname@aalto.fi

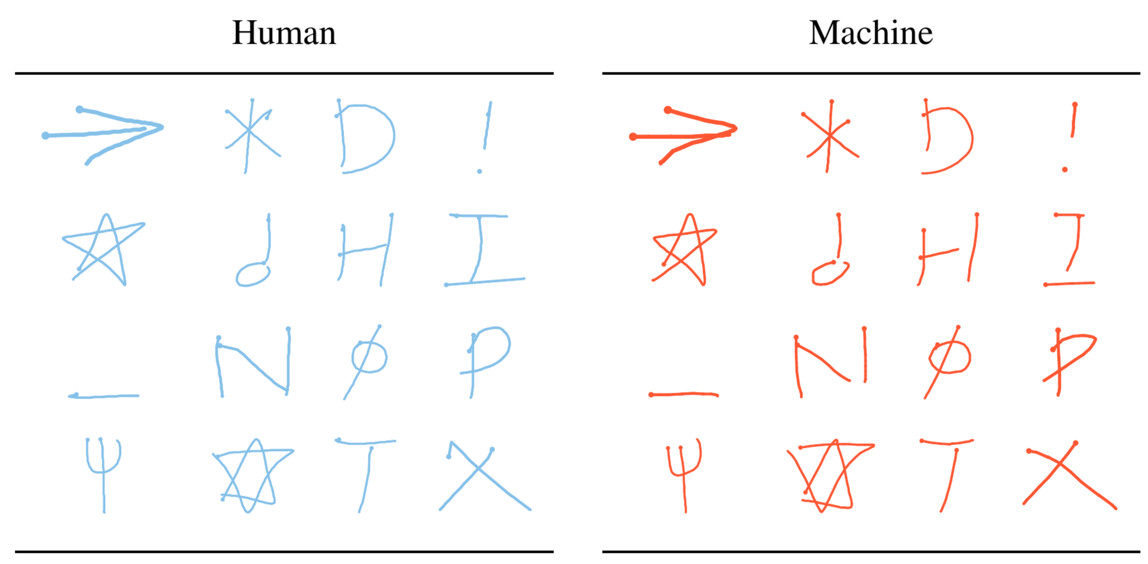
[2]Universidad del Atlantico Medio, Spain. moises.diaz@atlanticomedio.es

[3]IDeTIC Institute, Universidad de Las Palmas de Gran Canaria, Spain. miguelangel.ferrer@ulpgc.es

[4]Polytechnique Montréal, Canada. rejean.plamondon@polymtl.ca

Abstract

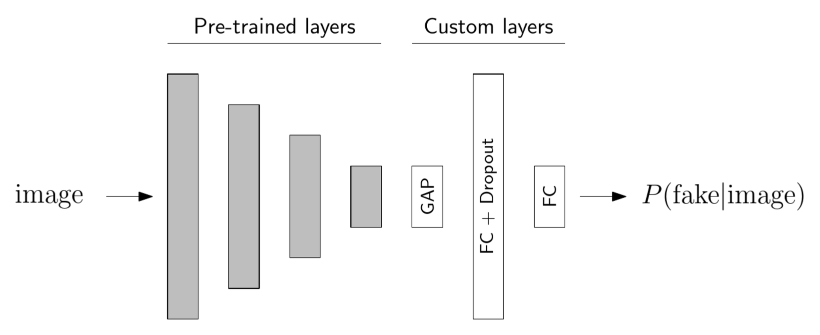
We study handwritten symbols (isolated characters, digits, gestures, and signatures) produced by humans and machines, and compare and contrast several deep learning models. We find that if symbols are presented as static images, they can fool state-of-the-art classifiers (near 75% accuracy in the best case) but can be distinguished with remarkable accuracy if they are presented as temporal sequences (95% accuracy in the average case). We conclude that an accurate detection of fake movements has more to do with *how* users write, rather than *what* they write. Our work has implications for computerized systems that need to authenticate or verify legitimate human users, and provides an additional layer of security to keep attackers at bay.



Offline Classification

"Convolutional Neural Nets"

Our CNN models were built on top of existing pre-trained network architectures, which we extended and fine-tuned for our classification task. *Conceptual representation:*



CNN-based sequence classifiers – $1-GDS dataset

System	Precision	Recall	F-measure	Accuracy	AUC
VGG16	68.50	68.48	68.45	68.48	75.39
ResNet50	68.97	68.95	68.95	68.95	76.34
Xception	71.15	71.03	70.98	71.03	79.58
DenseNet	71.41	71.41	71.41	71.41	78.74
Inception	**75.09**	**74.69**	**74.57**	**74.69**	**82.82**
Custom CNN	74.32	74.28	74.27	74.28	82.50

Online Classification

"Recurrent Neural Nets"

Single input feature: velocity

$$v_i = \frac{\sqrt{\Delta x_i^2 + \Delta y_i^2}}{t_i - t_{i-1}}$$

RNN-based sequence classifiers – $1-GDS dataset

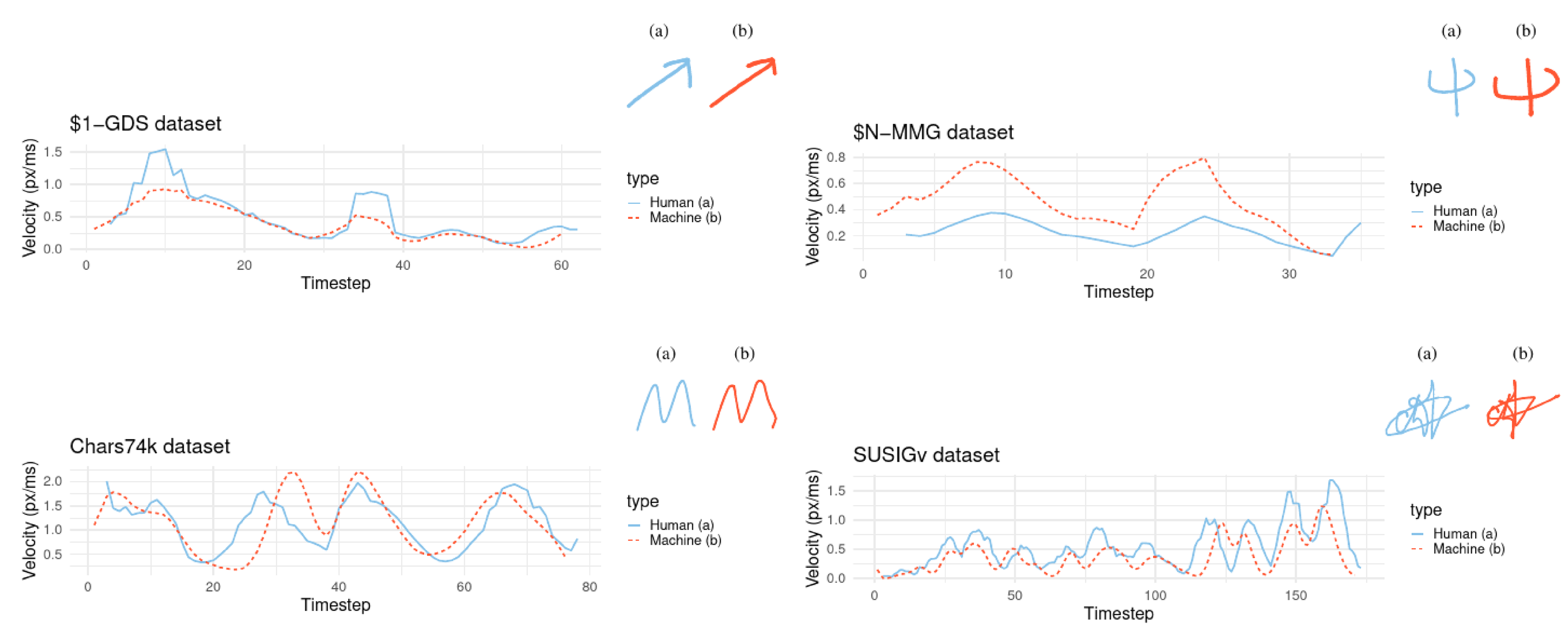
System	Precision	Recall	F-measure	Accuracy	AUC
1NN-DTW	85.32	83.97	83.80	83.97	83.88
Vanilla RNN	94.93	94.51	94.49	94.51	94.46
LSTM	95.66	95.27	95.25	95.27	97.45
Bi-LSTM	95.14	94.73	94.72	94.73	96.98
GRU	**95.78**	**95.39**	**95.38**	**95.39**	**98.20**
Bi-GRU	95.62	95.20	95.19	95.20	97.76

Results with our GRU classifier – All dataset

Dataset	Precision	Recall	F-measure	Accuracy	AUC
$1-GDS	95.78	95.39	95.38	95.39	98.20
$N-MMG	87.41	86.98	86.94	86.98	92.07
Chars74k	97.06	97.04	97.04	97.04	99.30
SUSIGv	93.68	93.35	93.34	93.35	95.43

Online Visual Examples

Velocity profile examples from our evaluated datasets, describing how a handwriting movement "unfolds" over time. A moving average filter of size 3 is applied to remove artificial jitter introduced by the input device. For each human movement, a synthetic version is generated with the ΣΛ model and plotted together with their human counterpart.



Synthetic and human samples are visually similar but the synthetic velocity profiles are smoother than their human counterparts.

A. Models Complexity

Summary of the complexity of our models, informed by the usual proxy metrics in deep learning.

System	Params	FLOPS	Memory
VGG16	16M	33M	81M
ResNet50	32M	63M	367M
Xception	30M	58M	336M
DenseNet	26M	51M	302M
Inception	30M	60M	347M
Custom CNN	72K	289K	1.8M
Vanilla RNN	41K	40K	148K
LSTM	41K	161K	507K
GRU	31K	120K	392K
Bi-LSTM	83K	322K	1M
Bi-GRU	63K	241K	765K

- "Params": the number of trainable model weights
- "FLOPS" *(Floating Point Operations Per Second)*: the number of multiply-and-accumulate operations
- "Memory": the model operational footprint.

B. Models Performance

The performance of all the RNNs over all our evaluated datasets.

Dataset	Model	Precision	Recall	F-measure	Accuracy	AUC
$1-GDS	Vanilla	94.93	94.51	94.49	94.51	94.46
	LSTM	95.66	95.27	95.25	95.27	97.45
	Bi-LSTM	95.14	94.73	94.72	94.73	96.98
	GRU	**95.78**	**95.39**	**95.38**	**95.39**	**98.20**
	Bi-GRU	95.62	95.20	95.19	95.20	97.76
	1NN-DTW	85.32	83.97	83.80	83.97	83.88
$N-MMG	Vanilla	77.94	77.65	77.59	77.65	83.59
	LSTM	81.84	80.08	79.79	80.08	81.25
	Bi-LSTM	85.89	85.66	85.63	85.66	90.20
	GRU	87.41	86.98	86.94	86.98	92.07
	Bi-GRU	**87.52**	**87.01**	**86.96**	**87.01**	**92.14**
	1NN-DTW	71.28	62.02	57.51	62.02	62.18
Chars74k	Vanilla	91.31	90.34	90.28	90.34	91.64
	LSTM	92.92	92.90	92.90	92.90	98.46
	Bi-LSTM	95.43	95.27	95.26	95.27	99.25
	GRU	**97.06**	**97.04**	**97.04**	**97.04**	99.30
	Bi-GRU	96.60	96.60	96.60	96.60	**99.49**
	1NN-DTW	92.55	91.52	91.47	91.52	91.53
SUSIGv	Vanilla	84.66	84.57	84.56	84.57	90.28
	LSTM	65.98	65.78	65.62	65.78	72.48
	Bi-LSTM	88.37	87.41	87.32	87.41	92.21
	GRU	93.68	93.35	93.34	93.35	95.43
	Bi-GRU	**95.00**	**94.68**	**94.67**	**94.68**	**97.29**
	1NN-DTW	72.65	65.16	61.82	65.16	64.85

GRU is our main RNN classifier since it is a simpler architecture.
Similar performance between GRU and Bi-GRU models.

C. GRU Robustness

We train our GRU it on different splits of the original training data
The model is fine-tuned on 20 % of the training data

Dataset	Split	Train	Test	Prec.	Recall	F1	Acc.	AUC
$1-GDS	10%	1056	9504	95.37	94.91	94.89	94.91	96.61
	20%	2112	8448	95.39	94.92	94.91	94.92	97.49
	40%	4224	6336	95.47	95.03	95.02	95.03	97.52
	80%	8448	2112	96.03	95.69	95.68	95.69	98.18
	99%	10455	106	96.48	96.23	96.22	96.23	98.00
$N-MMG	10%	1919	17277	85.88	84.19	84.01	84.19	86.88
	20%	3840	15356	87.81	87.26	87.21	87.26	92.32
	40%	7678	11518	87.75	87.20	87.16	87.20	92.28
	80%	15356	3840	88.19	87.45	87.38	87.45	92.39
	99%	19004	192	87.31	86.98	86.95	86.98	93.21
Chars74k	10%	675	6083	91.47	91.44	91.43	91.44	96.76
	20%	1351	5407	97.28	97.28	97.28	97.28	99.57
	40%	2703	4055	98.57	98.57	98.57	98.57	99.87
	80%	5407	1351	97.87	97.86	97.85	97.86	99.88
	99%	6690	68	98.58	98.53	98.53	98.53	99.91
SUSIGv	10%	376	3384	84.35	84.13	84.11	84.13	91.38
	20%	752	3008	93.40	93.35	93.35	93.35	96.32
	40%	1504	2256	92.92	92.69	92.68	92.69	95.80
	80%	3008	752	95.75	95.48	95.46	95.48	97.47
	99%	3722	38	100.0	100.0	100.0	100.0	100.0

Notice that when training on 99 % of the data, the model:

1. has almost full knowledge of the data distribution
2. is tested on a smaller number of samples

D. Effect of Input Device

$N-MMG dataset contains stylus and finger input data.
Train: A single type of data, stylus or finger
Test: with both types of input data

Train	Test	Precision	Recall	F-measure	Accuracy	AUC
Stylus	Stylus	83.27	79.31	78.75	79.31	86.88
Stylus	Finger	93.21	92.29	92.25	92.29	97.25
Finger	Finger	95.47	95.24	95.24	95.24	97.03
Finger	Stylus	79.55	78.82	78.73	78.82	85.19

- Better results when tested on finger-only samples
- Poor quality on the stylus samples on this dataset
- As consequence: lower-than-usual performance when tested on these stylus samples.

Final Remarks

1. Liveness detection problem via handwriting symbols (isolated characters, digits, gestures, and signatures)
2. Classification through deep learning architectures (Convolutional Neural Network for off-line images and Recurrent Neural Network for on-line sequences x, y, t)
3. State-of-the-art result in off-line and on-line. Remarkable improvements in on-line sequences.
4. TL;DR: Accurate detection of fake movements has more to do with *how* users write, rather than *what* they write.

Future Work

1. What is real sequence? is it the obtained by the acquisition device? Special attention to the mathematical procedure to generate synthetic samples
2. Is there is an adequate type of handwriting to train the networks? Increasing the number of handwriting specimens and its typology (more symbols, signatures, characters and so on)
3. Liveness detection regarding the effect of using multi-device acquisition

References

[1] L. A. Leiva and F. Álvaro μcaptcha: Human interaction proofs tailored to touch-capable devices via math handwriting. *Int. J. Hum. - Comput. Interact.*, vol. 31, no. 7, 2012.

[2] J. Galbally, R. Plamondon, J. Fierrez, and J. Ortega-García Synthetic on-line signature generation. Part II: Experimental validation *Pattern Recognit.*, vol. 45, no. 7, 2012.

[3] U. Bhattacharya, R. Plamondon, S. Chowdhury, P. Goyal, and S. Parui, A sigma-lognormal model based approach to generating large synthetic online handwriting samples databases *Int. J. Doc. Anal. Recogn.*, vol. 20, no. 71, 2017.

[4] A. Acien, A. Morales, J. Fierrez, and R. Vera-Rodriguez, BeCAPTCHA-Mouse: Synthetic mouse trajectories and improved bot detection In *arXiv 2005.00890*, 2020.

[5] L. A. Leiva, D. Martín-Albo, and R. Plamondon, Gestures a go go: Authoring synthetic human-like stroke gestures using the kinematic theory of rapid movements *IEEE Transactions on Information Forensics and Security*, ACM Trans. Intell. Syst. Technol., vol. 7, no. 2, 2016.

[6] C. D. Stefano, G. Guadagno, and A. Marcelli A saliency-based segmentation method for online cursive handwriting