



# Siamese Network applied to Authorship Verification

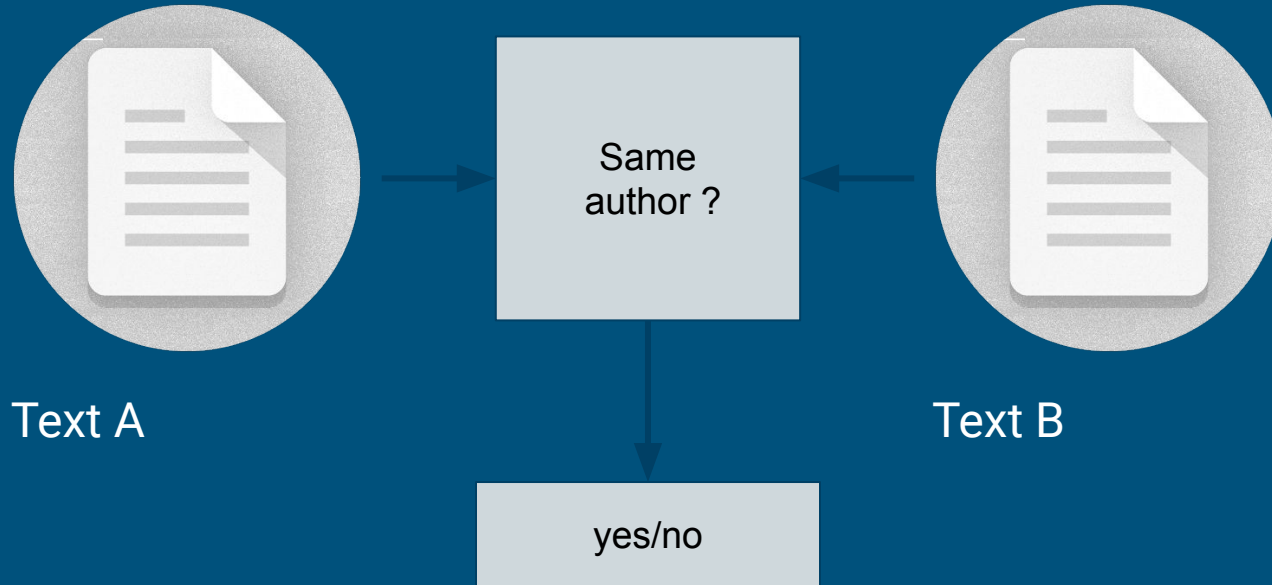
Emir Araujo-Pino\*  
Helena Gómez-Adorno\*\*  
Gibran Fuentes-Pineda\*\*

\*Posgrado en ciencias e ingeniería en computación

\*\*Instituto en investigaciones en matemáticas aplicadas y en sistemas

# Authorship Verification

---



# Previous work

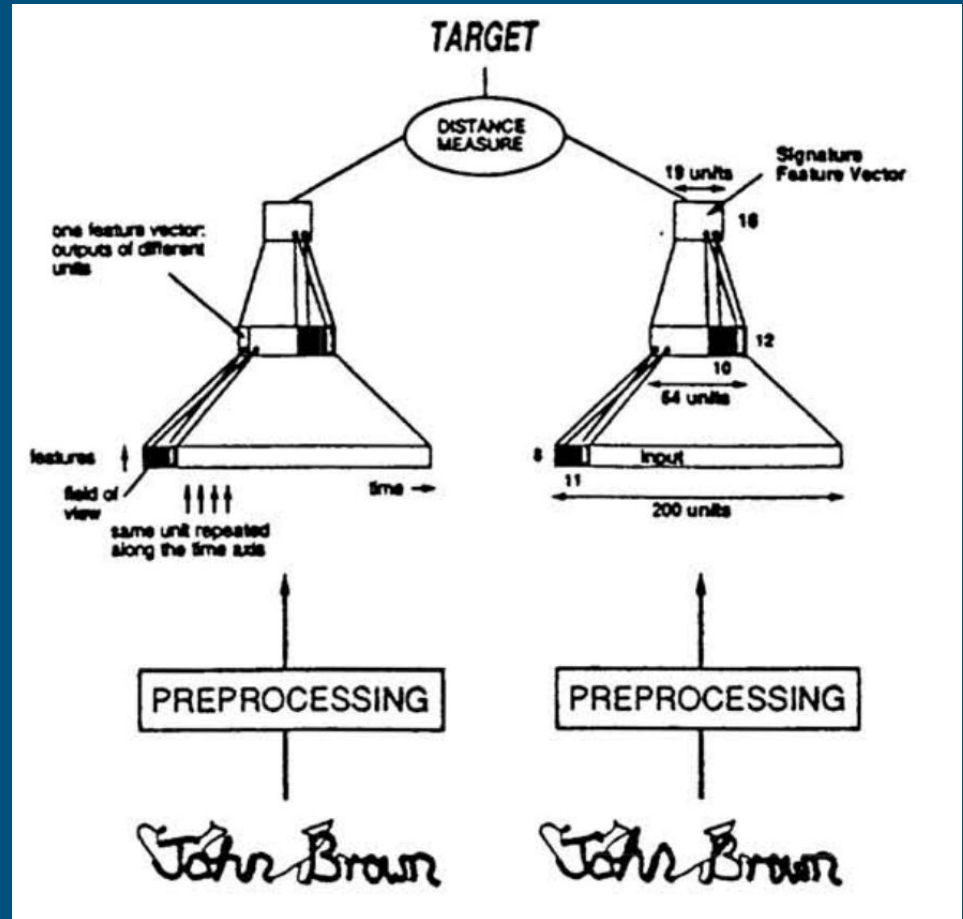
---

The three most used algorithms submitted to authorship verification PAN previous edition (2015), were SVM, random forest, and genetic algorithms.

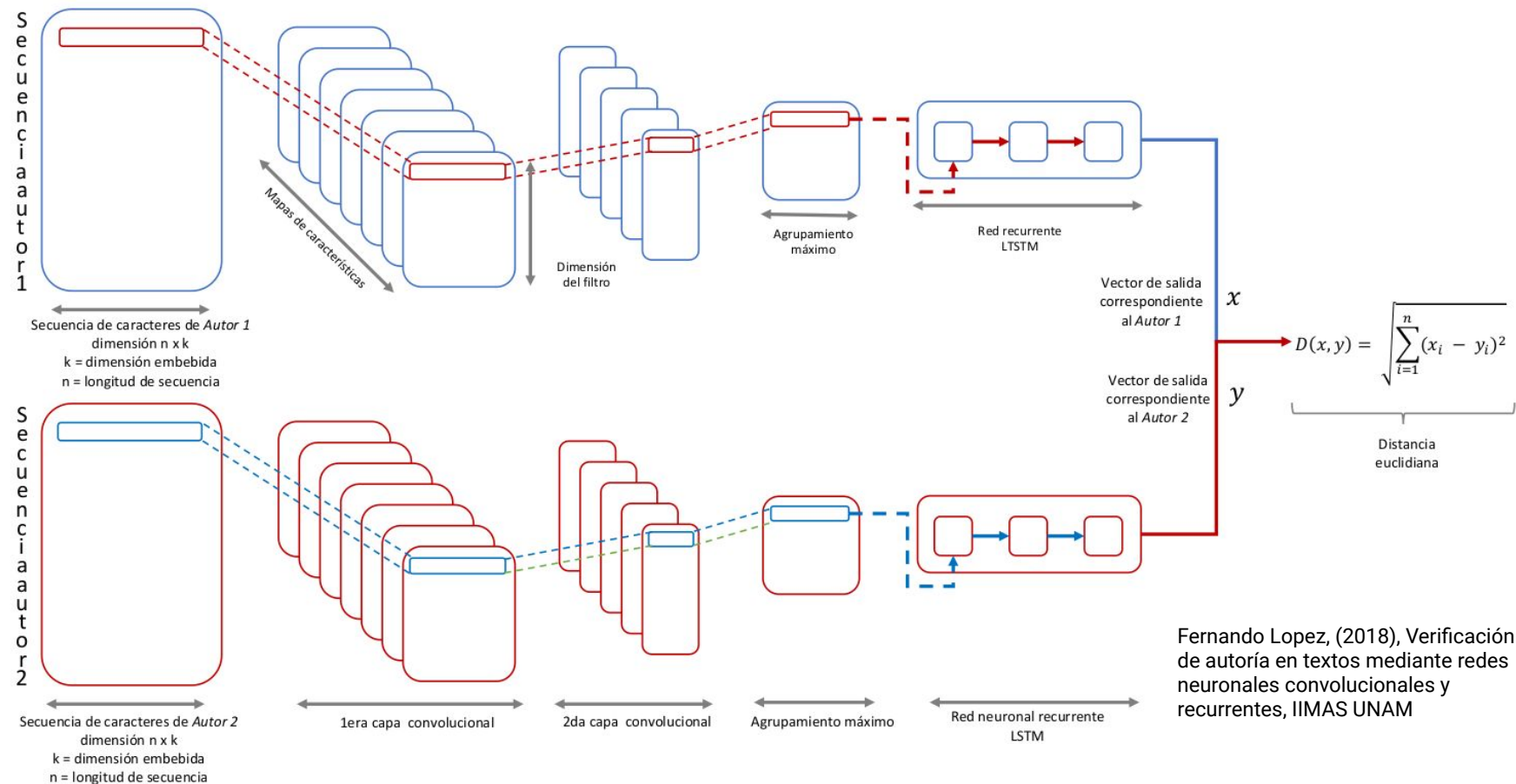
One of the latest PAN shared tasks of Authorship Attribution was proposed in 2019 and only one work was submitted with a neural network approach.

As far as we know, there are only 3 deep learning approaches to solve authorship problems. These are the Fernando López work, Chakaveh Saedi network, and last years approach.

# Siamese Networks



# Convolutional and recurrent neural network (Lopez 2018)

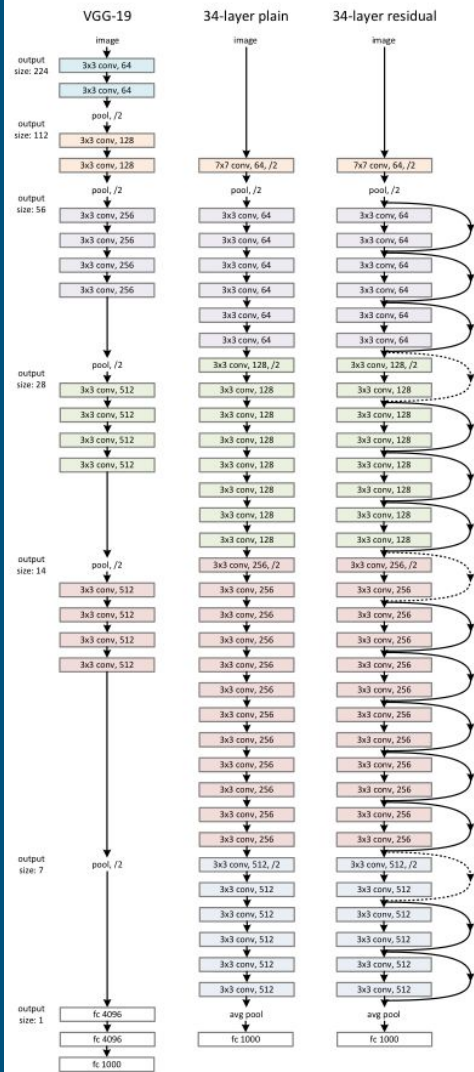


# Resnet

Won the 1st places in several tracks in ILSVRC & COCO 2015 competitions:

- ImageNet detection
- ImageNet localization
- COCO detection
- COCO segmentation

Kaiming He, et al., (2015), Deep Residual Learning for Image Recognition.

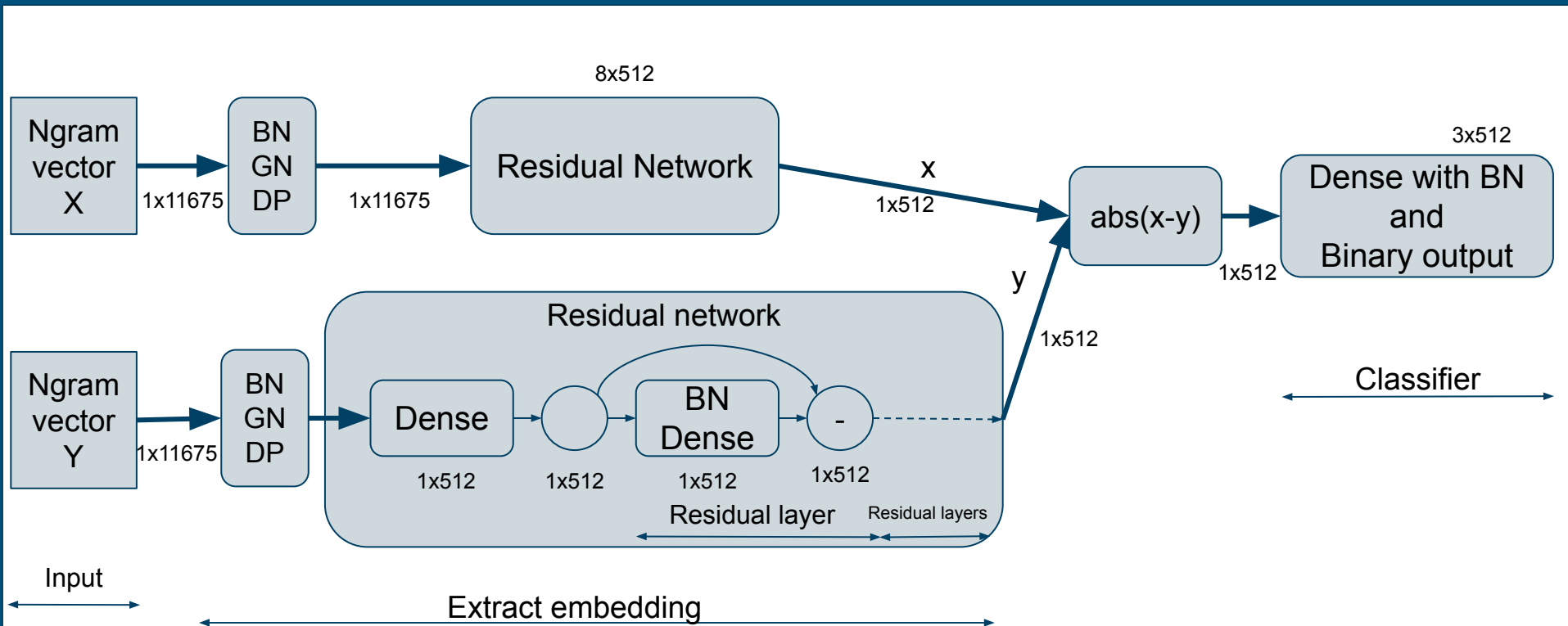


# Methodology

---

1. Data analysis
  - a. Statistical measures
2. Feature extraction
  - a. N-gram tuning: Min document frequency, Max document frequency, Sample length
3. Training
  - a. Network tuning
4. Evaluation

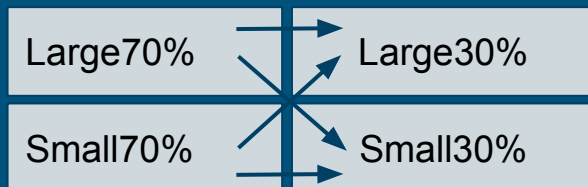
# Submitted Network Architecture





# Results

Training Dataset	Validation Dataset	AUC	C@1	f_05_u	F1	Mean
Large(70%)	Small(30%)	0.964	0.882	0.858	0.894	0.899
Large(70%)	Large(30%)	1.000	0.993	0.990	0.993	0.994
Small(70%)	Small(30%)	1.000	0.987	0.981	0.988	0.989
Small(70%)	Large(30%)	0.823	0.748	0.773	0.745	0.772



# PAN 2020 Official Results

---

<b>Training Dataset</b>	<b>AUC</b>	<b>C@1</b>	<b>f_05_u</b>	<b>F1</b>	<b>Mean</b>
<b>Large</b>	<b>0.859</b>	<b>0.751</b>	<b>0.745</b>	<b>0.800</b>	<b>0.789</b>
<b>Small</b>	<b>0.874</b>	<b>0.770</b>	<b>0.762</b>	<b>0.811</b>	<b>0.804</b>

<https://pan.webis.de/clef20/pan20-web/author-identification.html>



Thank you

Emir Araujo  
emiraraujoing@gmail.com

# References

---

- Efstathios Stamatatos, (2016), Authorship Verification: A Review of Recent Advances, University of the Aegean, Karlovassi, Greece
- Fernando Lopez, (2018), Verificación de autoría en textos mediante redes neuronales convolucionales y recurrentes, IIMAS UNAM
- Mike Kestemont, et al., (2018), Overview of the Author Identification Task at PAN-2018 Cross-domain Authorship Attribution and Style Change Detection
- Efstathios Stamatatos, et al., (2015), Overview of the Author Identification Task at PAN
- Chakaveh Saedi and Mark Dras ,(2020),Siamese Networks for Large-Scale Author Identification,Department of Computing, Sydney, Australia