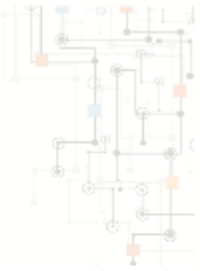


Centre de Recherche
en Numérique de Sfax

مركز البحث في الرقميات بصفاقس



**The International Conference on Intelligent Systems
and Pattern Recognition**
16-18 October Hammamet (Tunisia)

Document analysis and recognition : The Deep Learning Era

Yousri Kessentini

Digital Research center of Sfax

yousri.kessentini@crns.rnrt.tn

<http://yousri.fr.cr>



About me

□ Yousri Kessentini

- Assistant professor at **CRNS**
- Head of **DeepVision** research team
- More than 15 years experience in the document analysis & recognition
 - PHD in the university of Rouen
 - Post-doc Itesoft-LITIS
 - Researcher in CRNS



**Certified as an official instructor & ambassador from
NVIDIA Deep Learning Institute**



HomePage: <http://yousri.fr.cr>



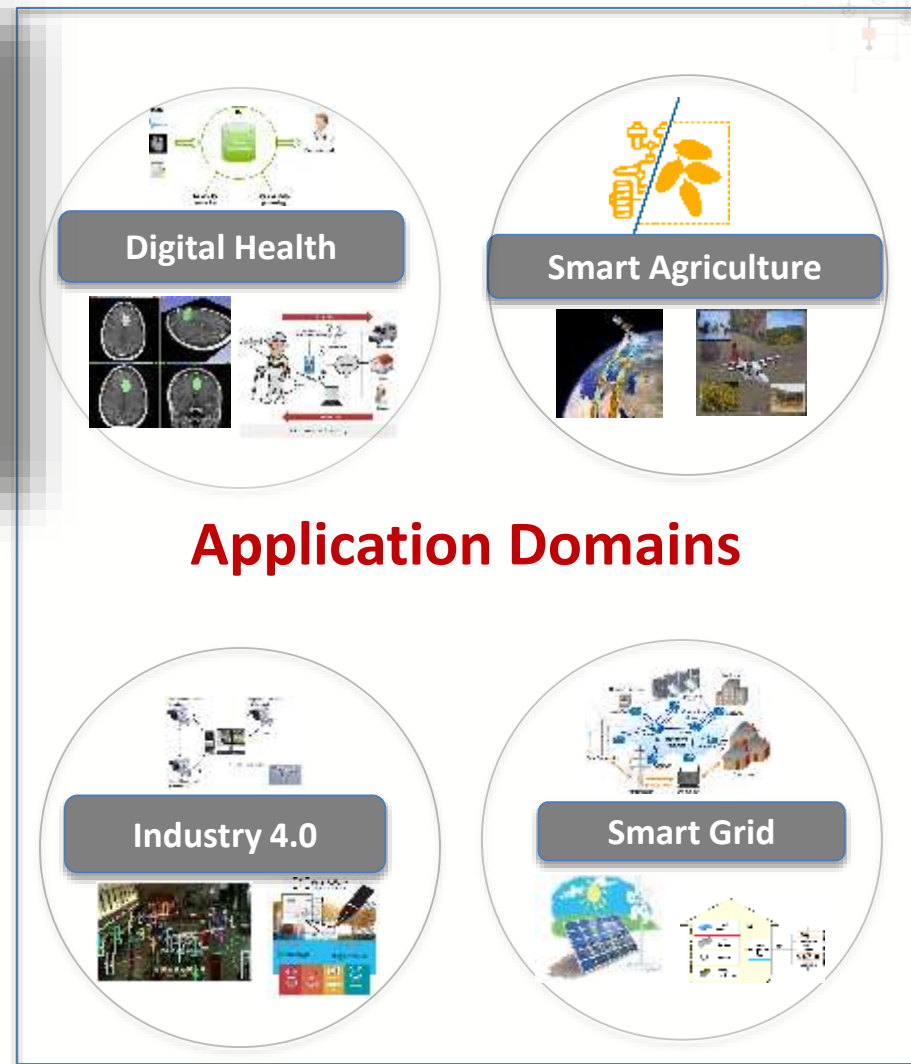
Digital research center of Sfax



- 11 000 m² / 165 researchers

Research Topics

1. Computer and distributed systems
2. Image and signal processing
3. Document and data analysis
4. Recognition of shapes and objects (3D)
5. Electronics and embedded systems
6. Computer Networks
7. Security of computer systems

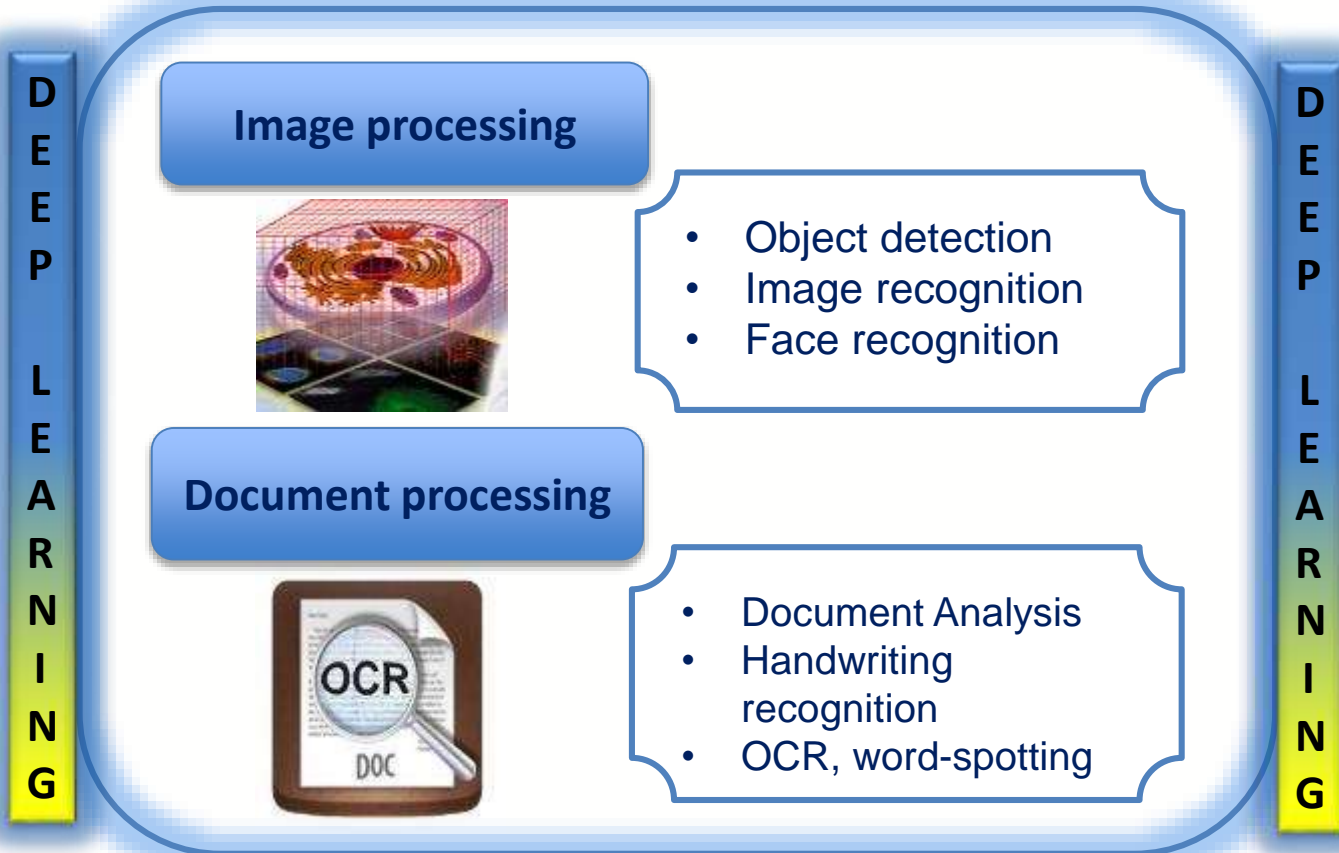


Research activities



DeepVision

Fields : Pattern recognition, Computer vision, Machine learning



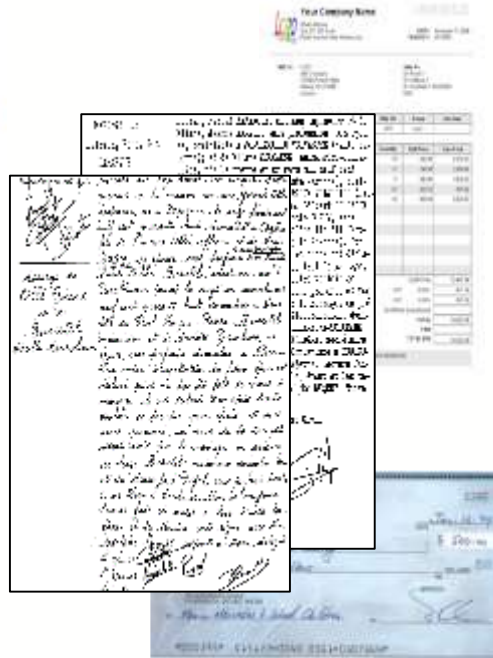
Introduction

A huge amount of documents (machine printed, handwritten)

- Preservation
- Storage
- Access to contents



Administrative Documents



Historical Documents



→ High Cost of manual processing

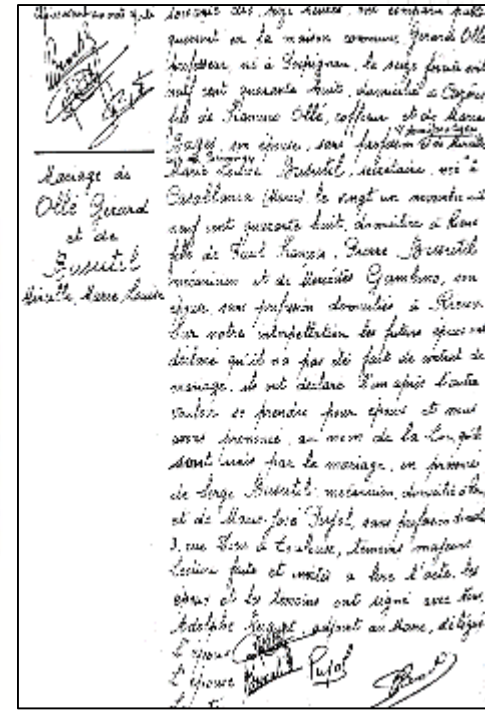
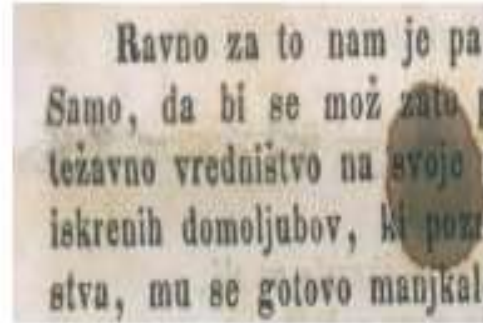
PayStream Advisors research reveals the average **cost of manually processing** an invoice can be as high as \$20, versus \$4 for automated invoice



Introduction

Automatic document processing

- Classification
- Document Enhancement
- Writer/script Identification
- Layout analysis
- Recognition
- Spotting ...

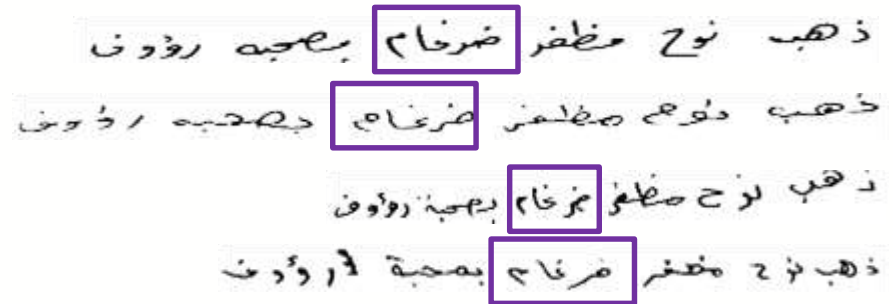


Difficulties

- Unstructured data
- Deformations, noise
- Different handwriting styles
- Segmentation problem

Handwritten characters 'e', 'u', 'u' in blue, green, and purple.

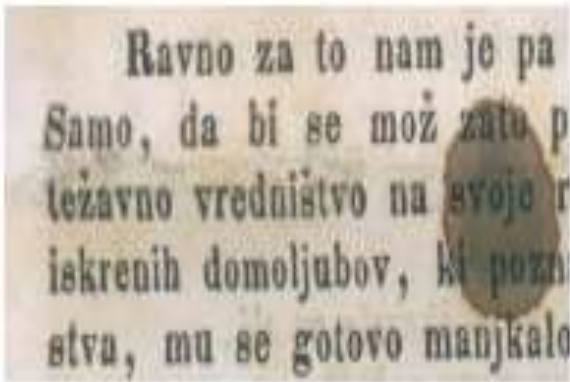
Handwritten word 'documentation' in multi-colored letters.



Document Recognition process



Original document



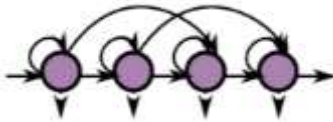
Document enhancement

Ravno za to nam je pa
Samo, da bi se mož, zato p
težavno vredništvo na svoje r
iskrenih domoljubov, ki pozni
stva, mu se gotovo manjkalo

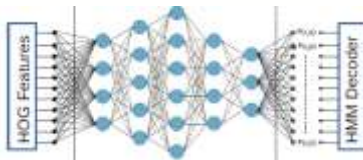
Line segmentation

Ravno za to nam je pa
Samo, da bi se mož, zato p
težavno vredništvo na svoje r
iskrenih domoljubov, ki pozni
stva, mu se gotovo manjkalo

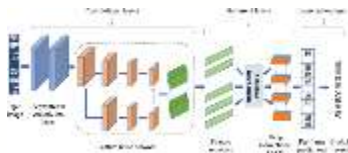
HMM



NN-HMM



Deep Learning



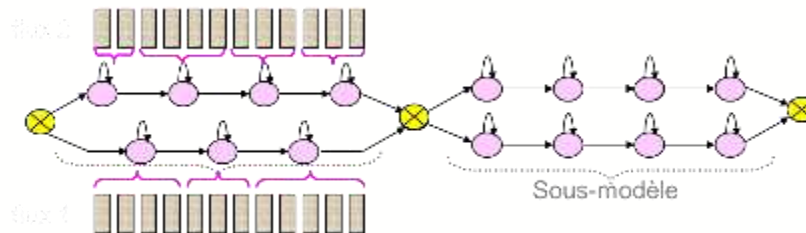
Script identification

Document recognition

Preprocessing
Feature Extraction
Optical Model
Langage Model

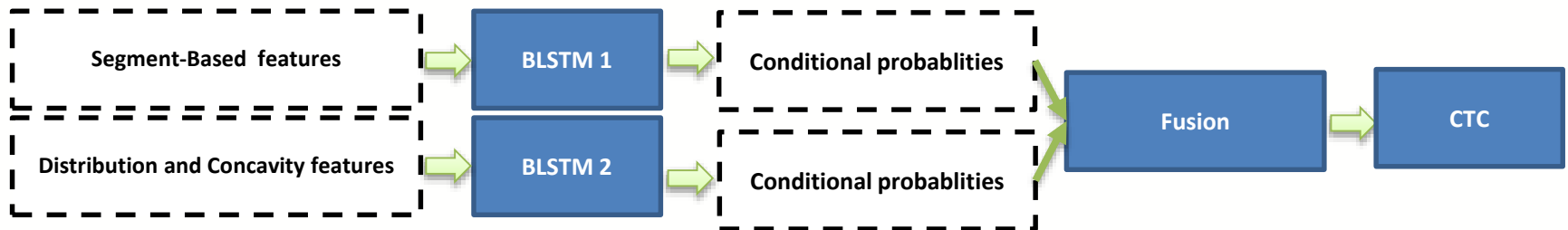
Handwriting recognition before deep Learning

❑ GMM-HMM with carefully chosen features



Y. Kessentini, T. Paquet, A. Benhamadou. Off-Line Handwritten Word Recognition Using Multi-Stream Hidden Markov Models. Pattern recognition Letters (PRL), Vol 30, Issue 1, pp. 60-70, January 2010.

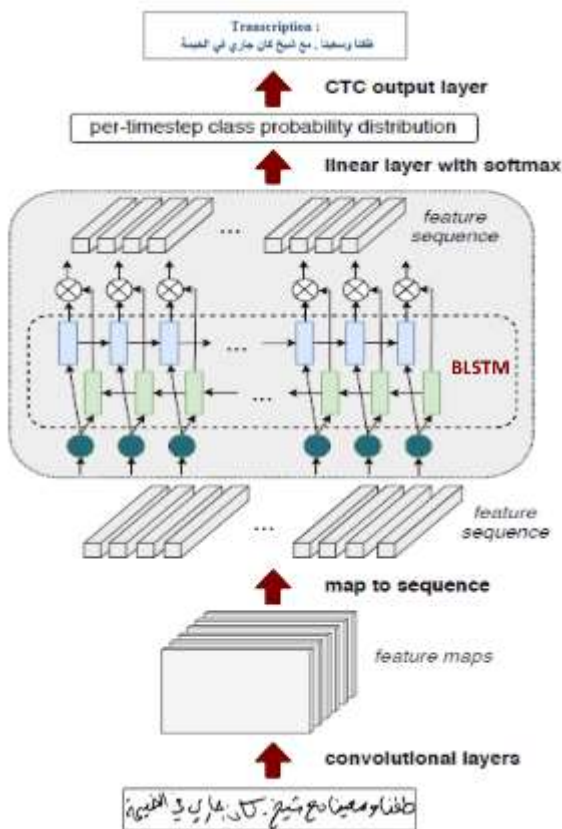
❑ Combination of BLSTMs at different levels



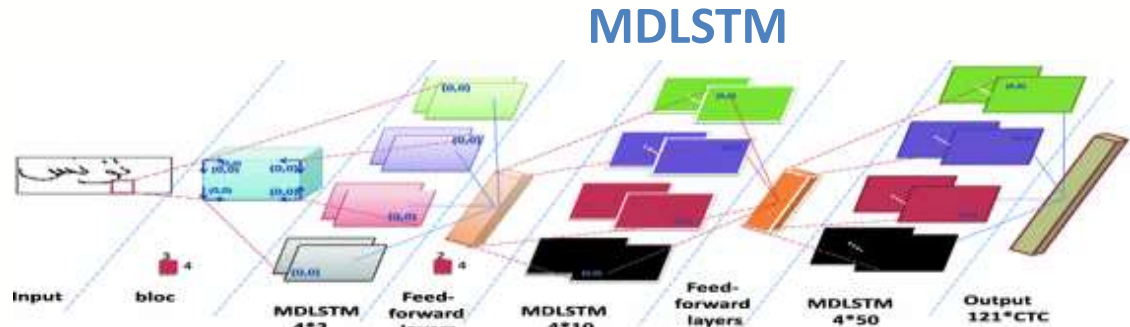
S. Khamekhem Jemni, Y. Kessentini, S. Kanoun, Offline Arabic Handwriting Recognition Using BLSTMs Combination. 13th international workshop on document analysis systems (DAS), pp 31-36, Vienna, Austria, 2018.

Deep Learning arrived

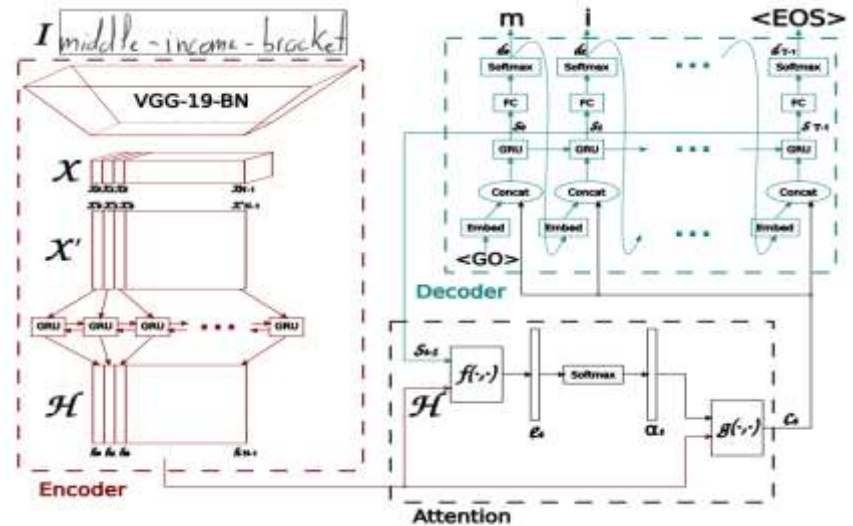
- Applied directly to the pixel of the raw text line image
 - BLSTM → MDLSTM → CRNN → Seq2Seq ...



CRNN : CNN+BLSTM



Seq2Seq model





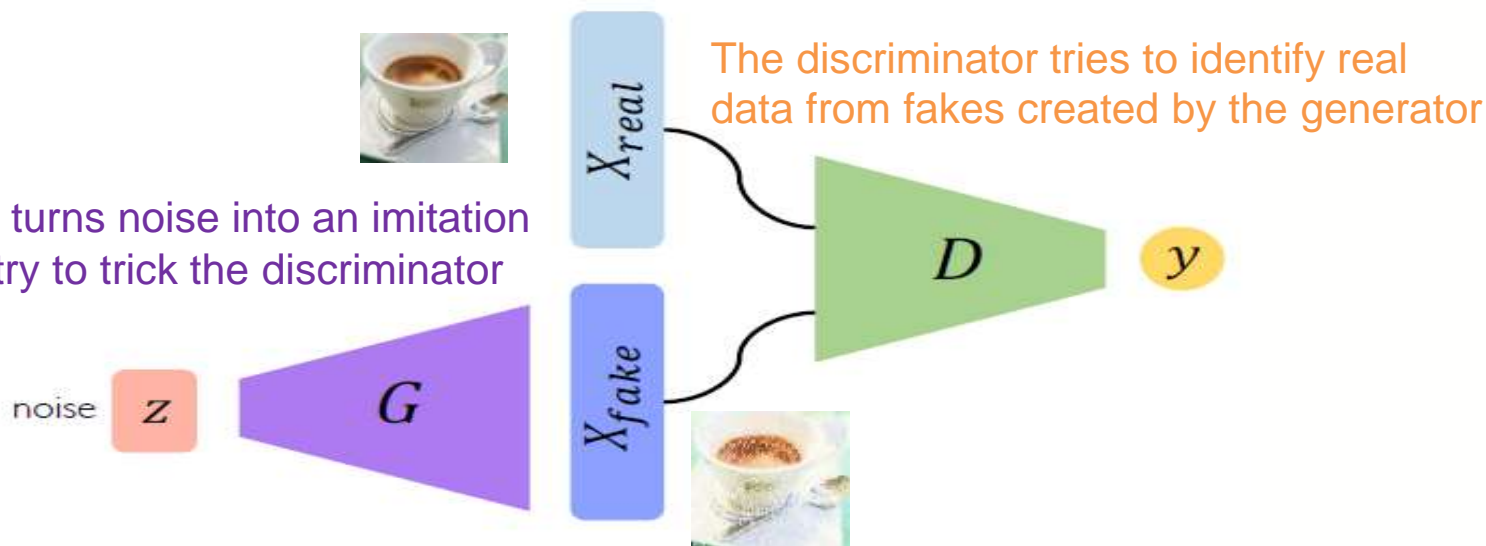
HIGH PERFORMANCE WITH DL BUT...

How to recognize degraded documents ?

Components of GANs



The generator turns noise into an imitation of the data to try to trick the discriminator



Train GAN jointly via minimax game:

- The discriminator tries to maximize its classification accuracy
- The generator tries to minimize the discriminator's classification accuracy

$$\min_G \max_D V(D, G) \quad V(D, G) = \underbrace{\mathbb{E}_{x \sim p(x)} [\log D(x)]}_{\text{Maximized by } D} + \underbrace{\mathbb{E}_{z \sim q(z)} [\log(1 - D(G(z)))]}_{\text{Minimized by } G}$$





DE-GAN : Document Enhancement

Ravno za to nam je pa
Samo, da bi se mož zato p
težavno vredništvo na svoje r
iskrenih domoljubov, ki pozn
stva, mu se gotovo manjkalo

Original

Ravno za to nam je pa
Samo, da bi se mož zato p
težavno vredništvo na svoje r
iskrenih domoljubov, ki pozn
stva, mu se gotovo manjkalo

Ground truth

Ravno za to nam je pa
Samo, da bi se mož zato p
težavno vredništvo na
iskrenih domoljubov, ki pozn
stva, mu se gotovo manjkalo

Otsu [16]

Ravno za to nam je pa
Samo, da bi se mož zato p
težavno vredništvo na svoje r
iskrenih domoljubov, ki pozn
stva, mu se gotovo manjkalo

Niblack [50]

Ravno za to nam je pa
Samo, da bi se mož zato p
težavno vredništvo na svoje r
iskrenih domoljubov, ki pozn
stva, mu se gotovo manjkalo

Sauvola et al. [51]

Ravno za to nam je pa
Samo, da bi se mož zato p
težavno vredništvo na svoje r
iskrenih domoljubov, ki pozn
stva, mu se gotovo manjkalo

DE-GAN

DE-GAN : Watermark removal



Watermarked images	Ground truth	Predicted images
<p>concept of evolution, or transposition as it was known in Paris at the time, and he tried—generally, it seems, successfully—to humiliate any colleagues who advanced the theory. Curiously, it was the very same child that led him to discover extinction that made evolution seem to him so preposterous, an affair as unlikely as levitation.</p> <p>As Cuvier liked to point out, he put his faith in anatomy; this was what had allowed him to distinguish the bones of a mammoth from those of an elephant and to recognize as a giant salamander what others took to be a bear. At the heart of his understanding of anatomy was a notion he termed “correlation of parts.” By this he meant that the components of an animal all fit together and are optimally designed for its particular way of life. Thus, for example, a carnivore will have an intestinal system suited to digesting flesh, at the same time, its jaws will</p> <p>be constructed for devouring prey; its claws for tearing and tearing it, the teeth for cutting and dividing to split the entire system of its incisor or cuspids, for piercing and cutting it to serve organs for digesting its victims.</p> <p>Conversely, an animal with hooves must necessarily be an herbivore, since it has “no means of seizing prey.” It will have “teeth with a flat crown, to grind seeds and grasses,” and a jaw capable of lateral motion. Were any one of these parts to be altered, the functional integrity of the whole would be destroyed. An animal that was born with, say, teeth or teeth organs that were somehow different from its parents’ would not be able to survive, let alone give rise to a whole new kind of creature.</p> <p>In Cuvier’s day, the most prominent proponent of transposition was his senior colleague at the Museum of Natural History, Jean-Baptiste Lamarck, according to Lamarck, there was a force—the “power of life”—that pushed organisms to become increasingly complex. At the same time, animals and also plants often had to cope with changes in their environment. They did so by adjusting their habits; these new habits, in turn, produced physical modifications that were then passed down to their offspring. Birds that sought prey in lakes spread out their toes when they hit the water, and in this way eventually developed webbed feet and</p> <p>e; thus, for example, a carnivorous animal must necessarily be an herbivore, since it has “no means of seizing prey.” It will have “teeth with a flat crown, to grind seeds and grasses,” and a jaw capable of lateral motion. Were any one of these parts to be altered, the functional integrity of the whole would be destroyed. An animal that was born with, say, teeth or teeth organs that were somehow different from its parents’ would not be able to survive, let alone give rise to a whole new kind of creature.</p> <p>In Cuvier’s day, the most prominent proponent of transposition was his senior colleague at the Museum of Natural History, Jean-Baptiste Lamarck, according to Lamarck, there was a force—the “power of life”—that pushed organisms to become increasingly complex. At the same time, animals and also plants often had to cope with changes in their environment. They did so by adjusting their habits; these new habits, in turn, produced physical modifications that were then passed down to their offspring. Birds that sought prey in lakes spread out their toes when they hit the water, and in this way eventually developed webbed feet and</p> <p>e; thus, for example, a carnivorous animal must necessarily be an herbivore, since it has “no means of seizing prey.” It will have “teeth with a flat crown, to grind seeds and grasses,” and a jaw capable of lateral motion. Were any one of these parts to be altered, the functional integrity of the whole would be destroyed. An animal that was born with, say, teeth or teeth organs that were somehow different from its parents’ would not be able to survive, let alone give rise to a whole new kind of creature.</p> <p>In Cuvier’s day, the most prominent proponent of transposition was his senior colleague at the Museum of Natural History, Jean-Baptiste Lamarck, according to Lamarck, there was a force—the “power of life”—that pushed organisms to become increasingly complex. At the same time, animals and also plants often had to cope with changes in their environment. They did so by adjusting their habits; these new habits, in turn, produced physical modifications that were then passed down to their offspring. Birds that sought prey in lakes spread out their toes when they hit the water, and in this way eventually developed webbed feet and</p>	<p>concept of evolution, or transposition as it was known in Paris at the time, and he tried—generally, it seems, successfully—to humiliate any colleagues who advanced the theory. Curiously, it was the very same child that led him to discover extinction that made evolution seem to him so preposterous, an affair as unlikely as levitation.</p> <p>As Cuvier liked to point out, he put his faith in anatomy; this was what had allowed him to distinguish the bones of a mammoth from those of an elephant and to recognize as a giant salamander what others took to be a bear. At the heart of his understanding of anatomy was a notion he termed “correlation of parts.” By this he meant that the components of an animal all fit together and are optimally designed for its particular way of life. Thus, for example, a carnivore will have an intestinal system suited to digesting flesh, at the same time, its jaws will</p> <p>be constructed for devouring prey; its claws for tearing and tearing it, the teeth for cutting and dividing to split the entire system of its incisor or cuspids, for piercing and cutting it to serve organs for digesting its victims.</p> <p>Conversely, an animal with hooves must necessarily be an herbivore, since it has “no means of seizing prey.” It will have “teeth with a flat crown, to grind seeds and grasses,” and a jaw capable of lateral motion. Were any one of these parts to be altered, the functional integrity of the whole would be destroyed. An animal that was born with, say, teeth or teeth organs that were somehow different from its parents’ would not be able to survive, let alone give rise to a whole new kind of creature.</p> <p>In Cuvier’s day, the most prominent proponent of transposition was his senior colleague at the Museum of Natural History, Jean-Baptiste Lamarck, according to Lamarck, there was a force—the “power of life”—that pushed organisms to become increasingly complex. At the same time, animals and also plants often had to cope with changes in their environment. They did so by adjusting their habits; these new habits, in turn, produced physical modifications that were then passed down to their offspring. Birds that sought prey in lakes spread out their toes when they hit the water, and in this way eventually developed webbed feet and</p> <p>e; thus, for example, a carnivorous animal must necessarily be an herbivore, since it has “no means of seizing prey.” It will have “teeth with a flat crown, to grind seeds and grasses,” and a jaw capable of lateral motion. Were any one of these parts to be altered, the functional integrity of the whole would be destroyed. An animal that was born with, say, teeth or teeth organs that were somehow different from its parents’ would not be able to survive, let alone give rise to a whole new kind of creature.</p> <p>In Cuvier’s day, the most prominent proponent of transposition was his senior colleague at the Museum of Natural History, Jean-Baptiste Lamarck, according to Lamarck, there was a force—the “power of life”—that pushed organisms to become increasingly complex. At the same time, animals and also plants often had to cope with changes in their environment. They did so by adjusting their habits; these new habits, in turn, produced physical modifications that were then passed down to their offspring. Birds that sought prey in lakes spread out their toes when they hit the water, and in this way eventually developed webbed feet and</p>	<p>concept of evolution, or transposition as it was known in Paris at the time, and he tried—generally, it seems, successfully—to humiliate any colleagues who advanced the theory. Curiously, it was the very same child that led him to discover extinction that made evolution seem to him so preposterous, an affair as unlikely as levitation.</p> <p>As Cuvier liked to point out, he put his faith in anatomy; this was what had allowed him to distinguish the bones of a mammoth from those of an elephant and to recognize as a giant salamander what others took to be a bear. At the heart of his understanding of anatomy was a notion he termed “correlation of parts.” By this he meant that the components of an animal all fit together and are optimally designed for its particular way of life. Thus, for example, a carnivore will have an intestinal system suited to digesting flesh, at the same time, its jaws will</p> <p>be constructed for devouring prey; its claws for tearing and tearing it, the teeth for cutting and dividing to split the entire system of its incisor or cuspids, for piercing and cutting it to serve organs for digesting its victims.</p> <p>Conversely, an animal with hooves must necessarily be an herbivore, since it has “no means of seizing prey.” It will have “teeth with a flat crown, to grind seeds and grasses,” and a jaw capable of lateral motion. Were any one of these parts to be altered, the functional integrity of the whole would be destroyed. An animal that was born with, say, teeth or teeth organs that were somehow different from its parents’ would not be able to survive, let alone give rise to a whole new kind of creature.</p> <p>In Cuvier’s day, the most prominent proponent of transposition was his senior colleague at the Museum of Natural History, Jean-Baptiste Lamarck, according to Lamarck, there was a force—the “power of life”—that pushed organisms to become increasingly complex. At the same time, animals and also plants often had to cope with changes in their environment. They did so by adjusting their habits; these new habits, in turn, produced physical modifications that were then passed down to their offspring. Birds that sought prey in lakes spread out their toes when they hit the water, and in this way eventually developed webbed feet and</p> <p>e; thus, for example, a carnivorous animal must necessarily be an herbivore, since it has “no means of seizing prey.” It will have “teeth with a flat crown, to grind seeds and grasses,” and a jaw capable of lateral motion. Were any one of these parts to be altered, the functional integrity of the whole would be destroyed. An animal that was born with, say, teeth or teeth organs that were somehow different from its parents’ would not be able to survive, let alone give rise to a whole new kind of creature.</p> <p>In Cuvier’s day, the most prominent proponent of transposition was his senior colleague at the Museum of Natural History, Jean-Baptiste Lamarck, according to Lamarck, there was a force—the “power of life”—that pushed organisms to become increasingly complex. At the same time, animals and also plants often had to cope with changes in their environment. They did so by adjusting their habits; these new habits, in turn, produced physical modifications that were then passed down to their offspring. Birds that sought prey in lakes spread out their toes when they hit the water, and in this way eventually developed webbed feet and</p>



DE-GAN : Deblurring



Blurred images	Ground truth	DE-GAN
<p>on. This is similar to the Hara & King experiences (such as frustration due to communication breakdowns) with the system caused major s. For one week during what wa discussions the system was unav</p> <p>tudents were able to view the techni re are always technical glitches wh mity to discuss issues related to the ppreciated that external study can b al issues made the experience less c</p>	<p>on. This is similar to the Hara & King experiences (such as frustration due to communication breakdowns) with the system caused major s. For one week during what wa discussions the system was unav</p> <p>tudents were able to view the techni re are always technical glitches wh mity to discuss issues related to the ppreciated that external study can b al issues made the experience less c</p>	<p>on. This is similar to the Hara & King experiences (such as frustration due to communication breakdowns) with the system caused major s. For one week during what wa discussions the system was unav</p> <p>tudents were able to view the techni re are always technical glitches wh mity to discuss issues related to the ppreciated that external study can b al issues made the experience less c</p>
<p>... objects, we will be able to en sions. In the AIPL, which we discus encapsulate performatives in them. ' create new contexts from existing o municative act CA x to an agent B, objects, where each object is bound criteria, temporal properties, constra in the language. We view this collec</p> <p>4.2 Contexts in AIPL</p> <p>The approach of using intensional g a conservative extension of Lucid b cid [1]. In our approach, the name</p>	<p>... objects, we will be able to en sions. In the AIPL, which we discus encapsulate performatives in them. ' create new contexts from existing o municative act CA x to an agent B, objects, where each object is bound criteria, temporal properties, constra in the language. We view this collec</p> <p>4.2 Contexts in AIPL</p> <p>The approach of using intensional g a conservative extension of Lucid b cid [1]. In our approach, the name</p>	<p>... objects, we will be able to en sions. In the AIPL, which we discus encapsulate performatives in them. ' create new contexts from existing o municative act CA x to an agent B, objects, where each object is bound criteria, temporal properties, constra in the language. We view this collec</p> <p>4.2 Contexts in AIPL</p> <p>The approach of using intensional g a conservative extension of Lucid b cid [1]. In our approach, the name</p>
<p>... the tracking mode, the estimate $\hat{\tau}$ f to the optimum sampling phase τ_0</p> <p>Following approaches of [2] an likelihood function for a given sequen is determined by</p> $p(z \tau) = \left(\frac{1}{\pi\sigma^2}\right)^K \exp\left(-\frac{1}{\sigma^2}\right)$ <p>Taking the natural logarithm of obtain the log-likelihood function</p> $L(\tau) = \ln\left[\left(\frac{1}{\pi\sigma^2}\right)^K\right] - \frac{1}{\sigma^2}$	<p>... the tracking mode, the estimate $\hat{\tau}$ f to the optimum sampling phase τ_0</p> <p>Following approaches of [2] an likelihood function for a given sequen is determined by</p> $p(z \tau) = \left(\frac{1}{\pi\sigma^2}\right)^K \exp\left(-\frac{1}{\sigma^2}\right)$ <p>Taking the natural logarithm of obtain the log-likelihood function</p> $L(\tau) = \ln\left[\left(\frac{1}{\pi\sigma^2}\right)^K\right] - \frac{1}{\sigma^2}$	<p>... the tracking mode, the estimate $\hat{\tau}$ f to the optimum sampling phase τ_0</p> <p>Following approaches of [2] an likelihood function for a given sequen is determined by</p> $p(z \tau) = \left(\frac{1}{\pi\sigma^2}\right)^K \exp\left(-\frac{1}{\sigma^2}\right)$ <p>Taking the natural logarithm of obtain the log-likelihood function</p> $L(\tau) = \ln\left[\left(\frac{1}{\pi\sigma^2}\right)^K\right] - \frac{1}{\sigma^2}$

Source code: <https://github.com/dali92002/DE-GAN>



HIGH PERFORMANCE WITH DL BUT...

Recognition still depends of linguistic resources

How to handle OOV words ?

Out-of-vocabulary words

- ❑ Handwriting recognition systems usually rely on static dictionaries
 - Small lexicon: Low coverage rate
 - Big lexicon: High computing time + high confusion between words
- Full coverage of these dictionaries is generally not achieved

Out Of Vocabulary word (OOV)

توظيف عدد أكبر من خريجي الجامعات

IN Vocabulary word (IV)

The word « الجامعات »
Does not exist in the
vocabulary

Word lexicon driven recognition

توظيف أكبر عدد من خريجي الجامع

The most similar word from
the dictionary

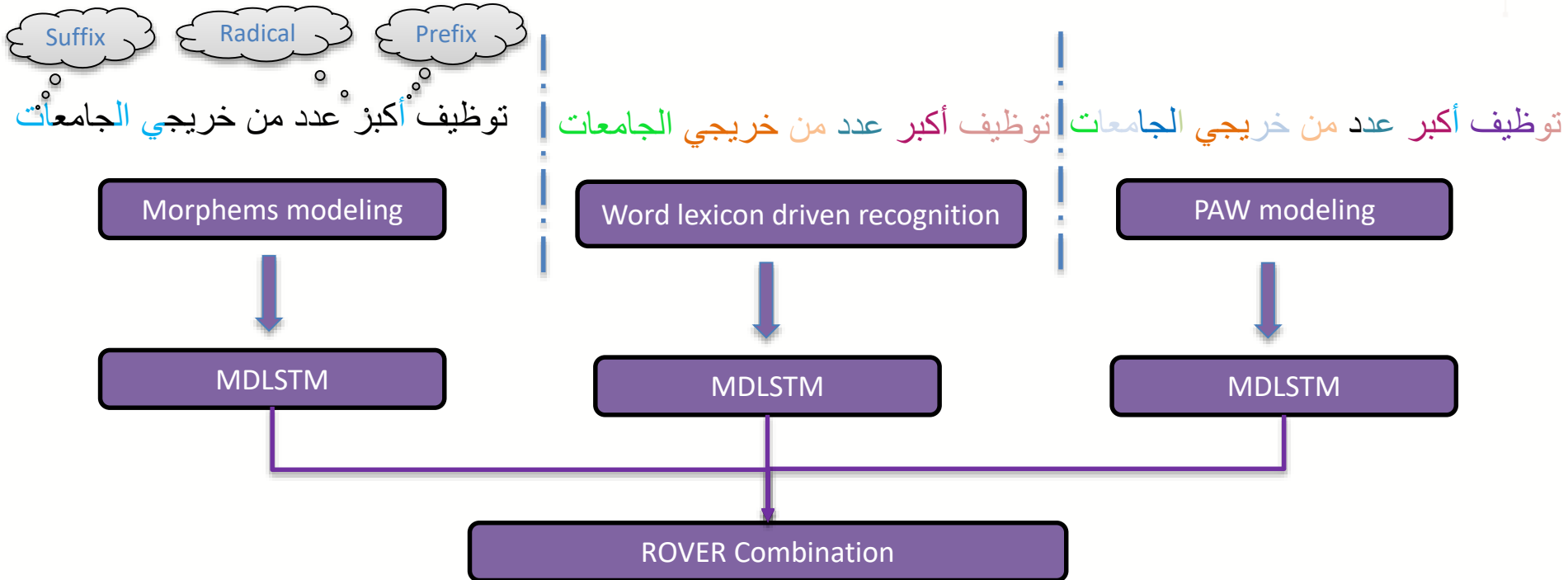
Dictionary

توظيف
توظف
أكبر
عدد
من
خريجي
الجامع
.....

« الجامعات »

How to recognize OOV words ?

Solution 1 : Combining different sub-word modeling



S. Khamekhem Jemni, **Y. Kessentini**, S. Kanoun, Improving Recurrent Neural Networks for Offline Arabic Handwriting Recognition by combining different Language Models, International Journal of Pattern Recognition and Artificial Intelligence. November 2019.

Solution 2 : OOV words detection and recovery using dynamic dictionaries



Idea:

- Keep the IV words recognized by the word lexicon driven recognition
- Recover OOV word using dynamic lexicons built from large text corpora

Line Image

بدأت قوافله الحجاج المثر آخر تلبية

Word lexicon driven recognition

بدأت نوافل الحجاج حاجة أثر آخر تلبية

PAW lexicon driven recognition

بدأت قوافل الحجاج حاج أثر آخر تلبية

Morphemes lexicon driven recognition

بدأت نوافل الحجاج حاج أثر آخر تلبية

OOV Detection stage



Results

- KHATT dataset
- Train : 9475 Test : 2007
- Lexicon : 18933 words, 11,46% OOV

Systems	WER
Hamdani et al. [2]	26.80
BenZeghiba et al. [1]	30.9
BenZeghiba [3]	34.3
Our approach	20.83





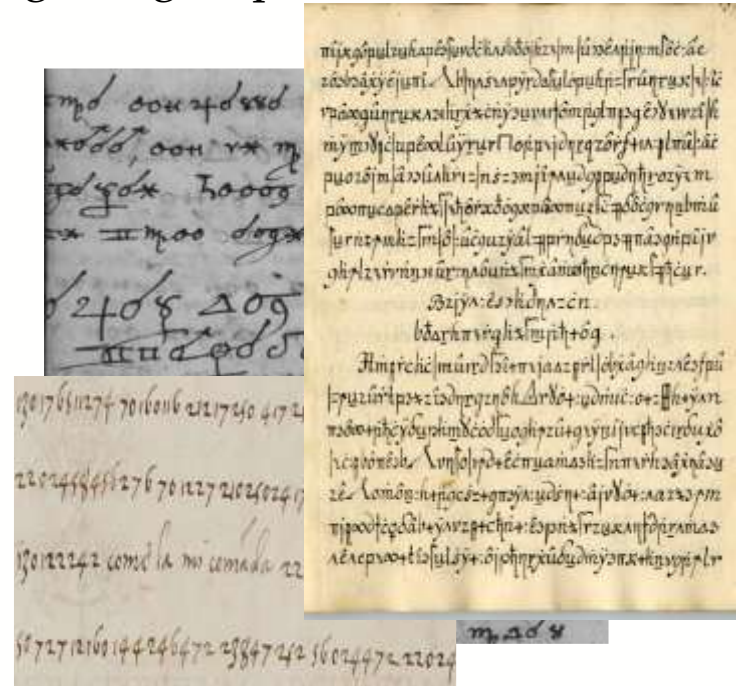
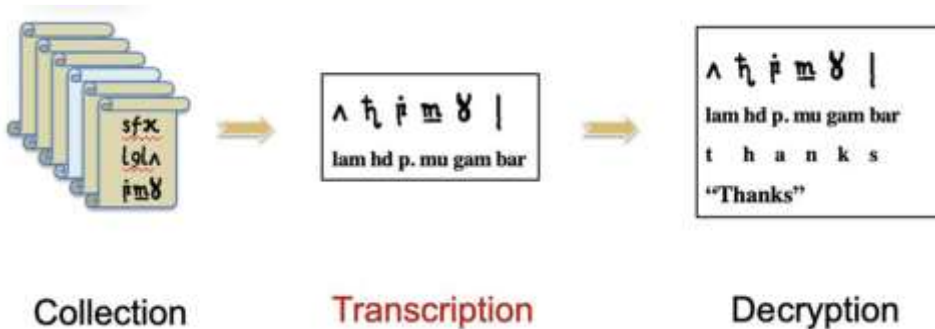
HIGH PERFORMANCE WITH DL BUT...

What about cases where you don't have labeled data?

Recognition of ciphered manuscript

- Around 1% of documents in archives contain encrypted text:
 - Diplomatic correspondence, secret societies/religious groups...

- DECODE/DECRYPT project:
 - To develop resources & tools for automatic decryption of enciphered documents from early modern times.

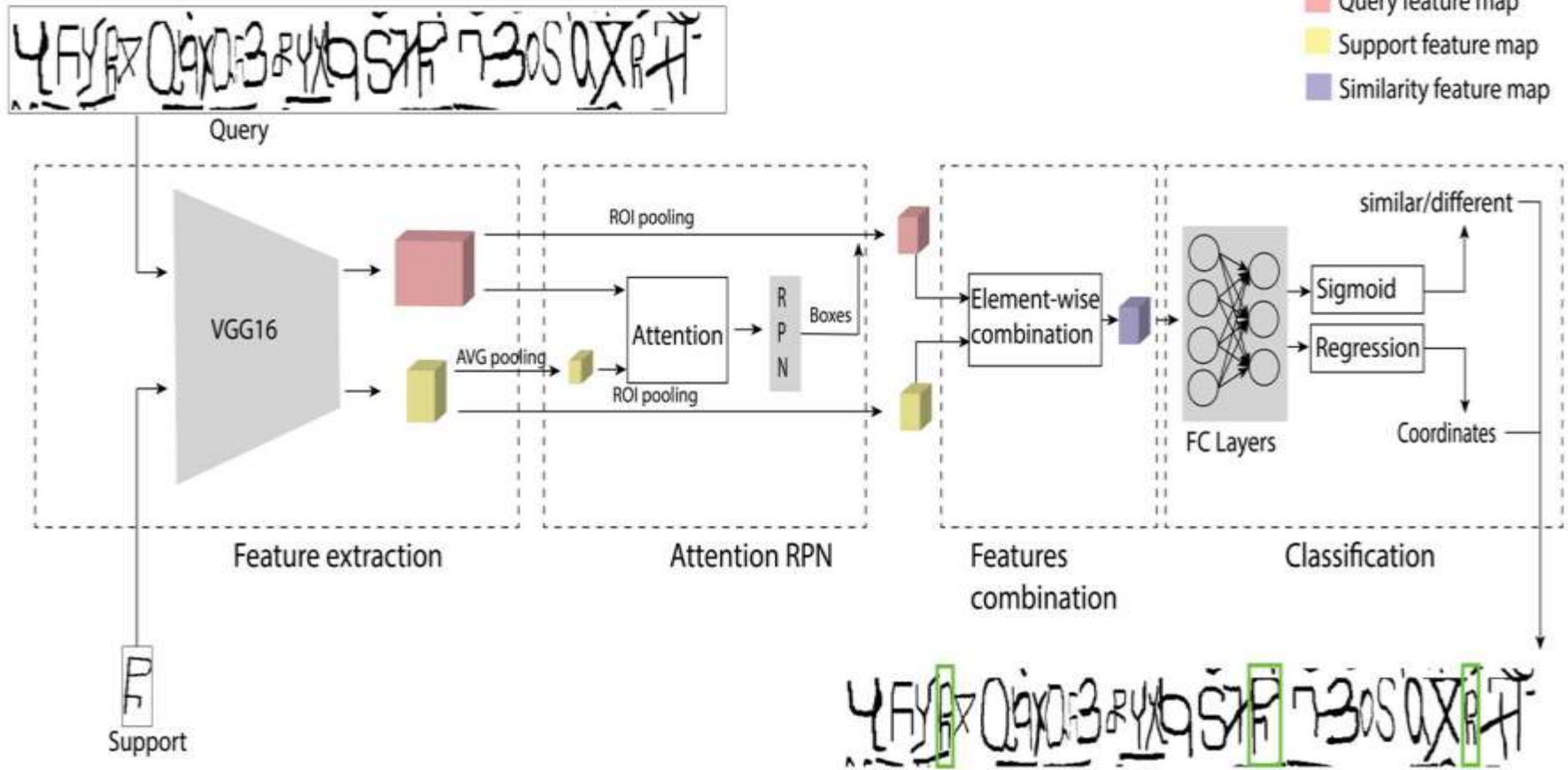


Problem : Very few (or none) labeled data to train

Few-shot Learning for Historical Ciphered Manuscript Recognition



Symbol detection step

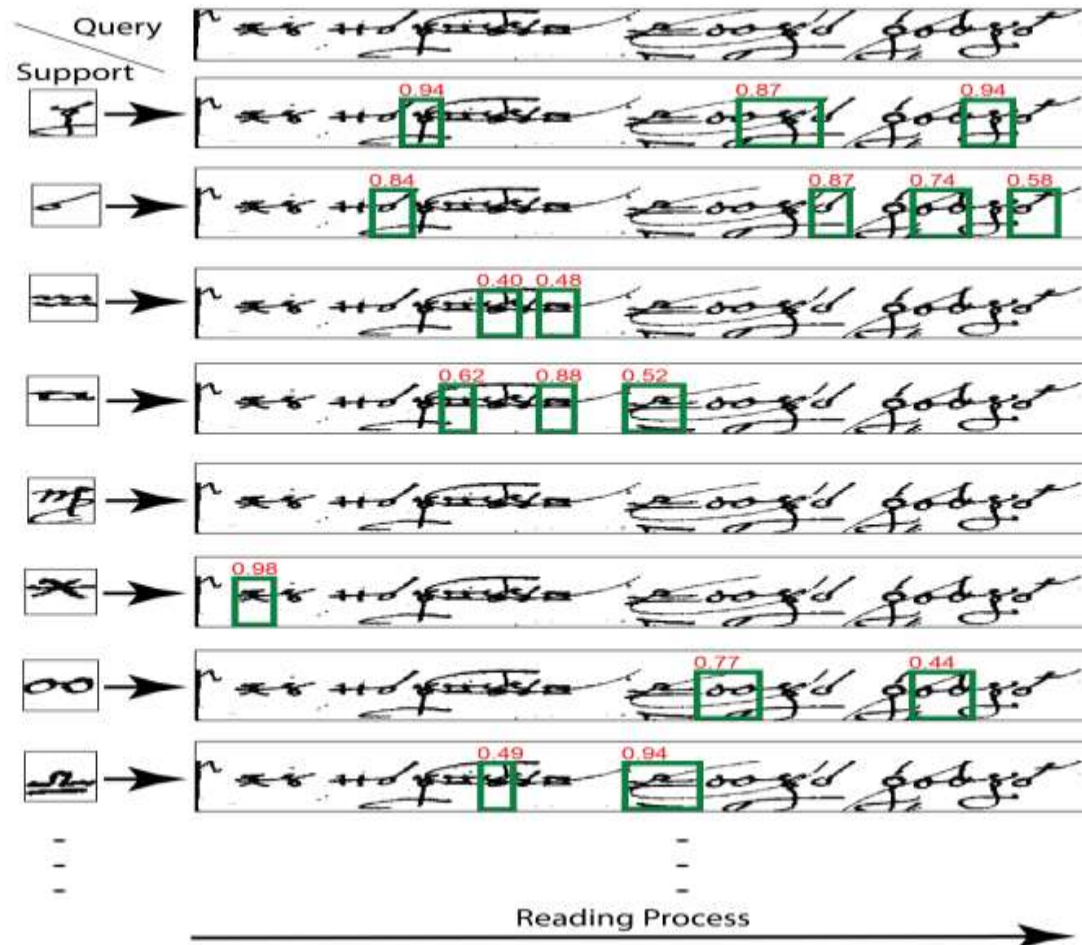


M.A.Souibgui, A.Fornés, Y.Kessentini, C.Tudor. A Few-shot Learning Approach for Historical Encoded Manuscript Recognition, International Conference on Pattern Recognition (ICPR), 2020.

Few-shot Learning for Historical Ciphered Manuscript Recognition



Text recognition step



Few-shot Learning for Historical Ciphpered Manuscript Recognition

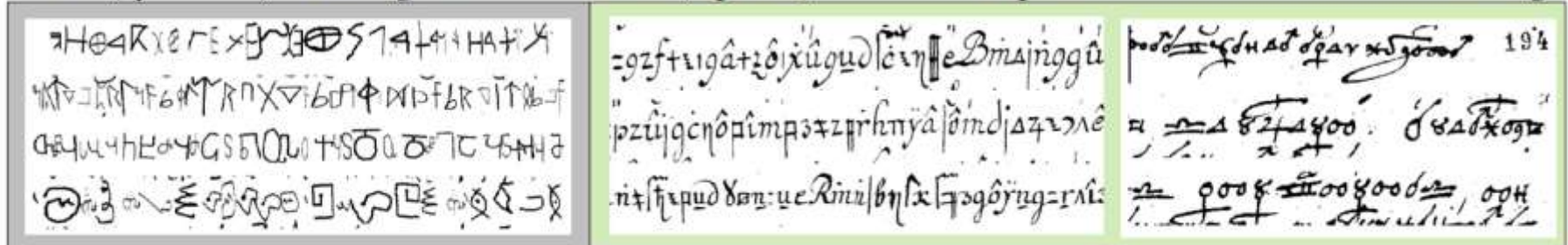


Train (synthetic): Omniglot

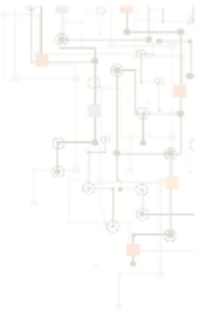
Test (ciphers):

Copiale

Borg



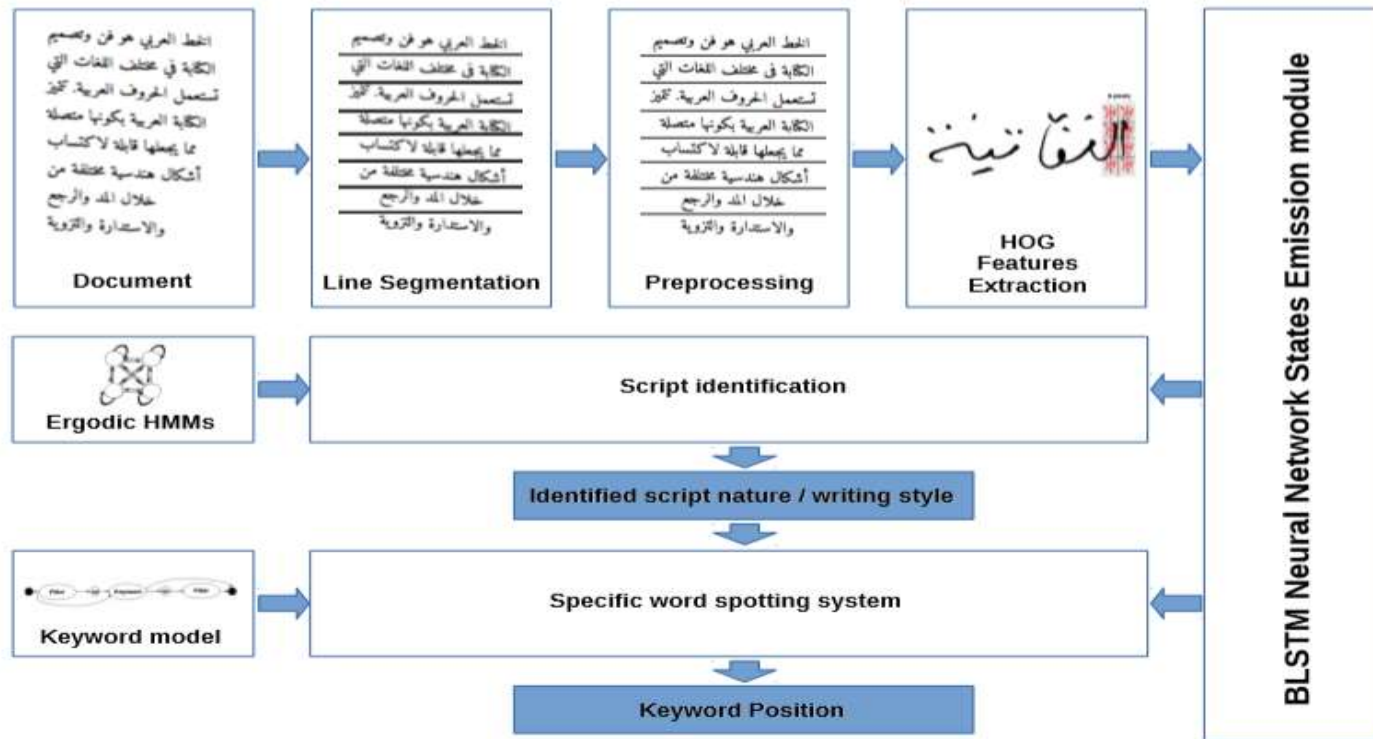
Method	Copiale			Borg		
	Training Pages	Shots	SER	Training pages	Shots	SER
LSTMs	25	-	0.11	7	-	0.55
	34	-	0.10	9	-	0.52
	42	-	0.07	11	-	0.45
Few-shot	2	1	0.10	2	1	0.17
		5	0.10		5	0.18



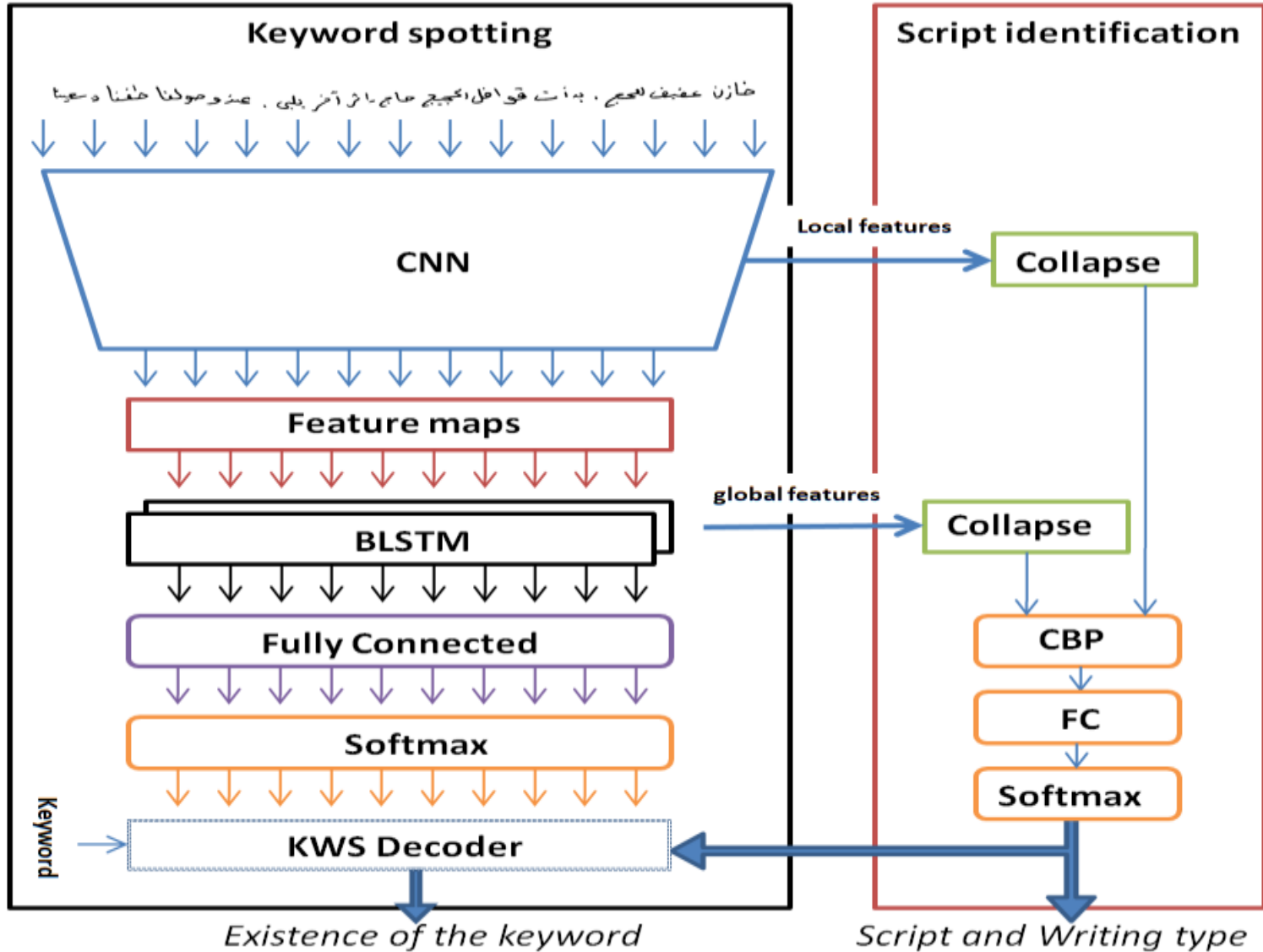
BUT...
**DO WE NEED TO RECOGNIZE ALL THE
DOCUMENT ?**

Keyword spotting

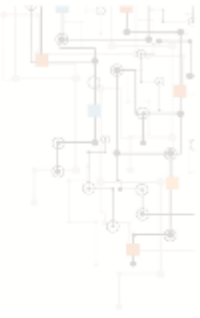
- Finding all instances of a query word that exist in a scanned document image, without fully recognizing it.
 - Document indexing and retrieval, routing, categorization...



Multi-task learning



Results



Script identification results (Multi-task Learning)

Script	Handwritten French	Handwritten Arabic	Printed Arabic
Handwritten French	100%	0%	0%
Handwritten Arabic	0%	99.9%	0.1%
Printed Arabic	0%	0.11%	99.89%

Keyword spotting results (100 keywords MAP)

System	Handwritten French	Handwritten Arabic	Printed Arabic
Multi-Task	89%	92.87%	98.29%
BLSTM-HMM	70.12%	84.76%	82.15%
HMM	51.9%	49.1%	61.8%





BUT...
IS RECOGNITION ENOUGH?

Recognition → Understanding

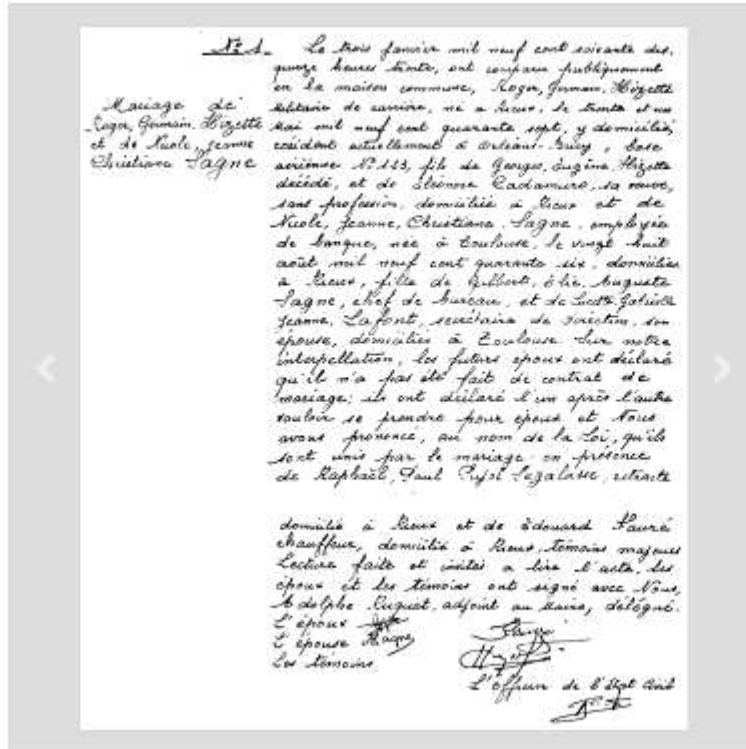
❑ READ project

- Information Extraction: Transcription + Semantic Recognition (NER)

❑ Death, birth and marriage records

Files

1970M0001.tif



INFORMATION EPOUX

Nom Epoux

Eliette

Prénom Epoux

Roger German

Date Naissance

31-5-1947

Ville

Rieux

Nom Père

Eugène

Prénom Père

Georges

Nom Mère

Laure,

Prénom Mère

Eléonore

INFORMATION EPOUSE

Nom Epouse

Lane

Prénom Epouse

Nicole Jeanne Christiane



Conclusion

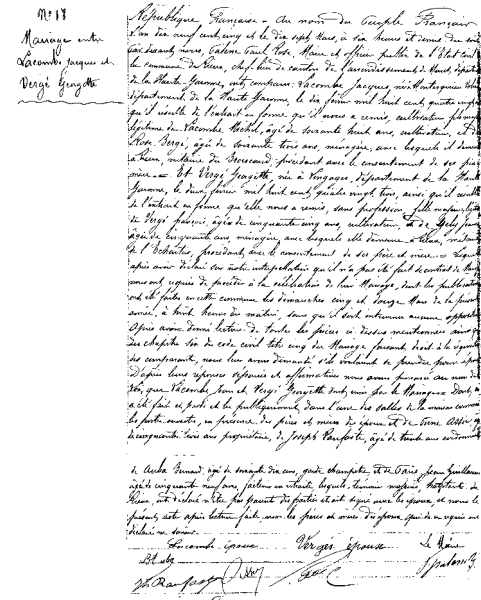
- Whenever I say that I am a researcher in Document Analysis & Recognition, people say :

I thought it was a solved problem !

Real data, realistic problem



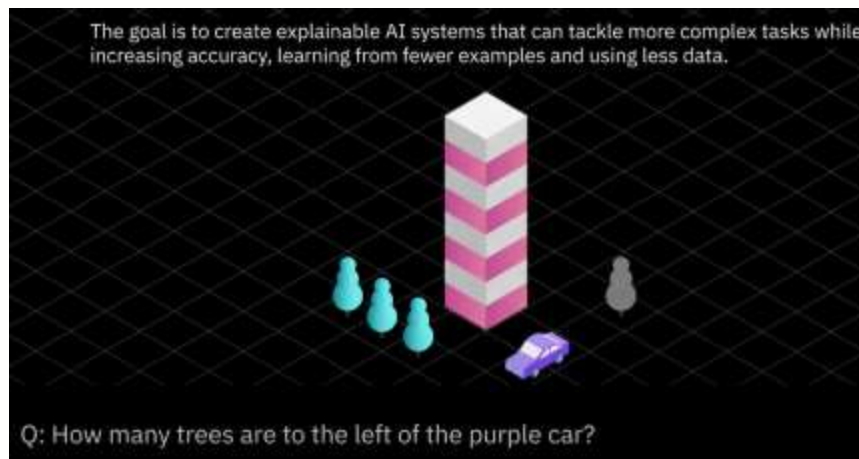
Real data, real problem



Very difficult to reach, even for industrial solution providers

The future

- ❑ Full-page Document recognition
- ❑ There will be a new approach to OCR
 - Seq2Seq, Transformers, Graph Neural Networks...
- ❑ **Labeled data** is an important limitation
 - Data generation, DomainAdaptation, Few/zero-shotLearning...
- ❑ Neuro-Symbolic AI
 - Represents the causal and compositional processes behind perceptual observations





The International Conference on Intelligent Systems and Pattern Recognition

16-18 October Hammamet (Tunisia)

Thank you!

Yousri Kessentini
Digital research center of Sfax

yousri.kessentini@crns.rnrt.tn
<http://yousri.fr.cr>

