Recommassance de Formes: Application ans Documents

-INSA de Rouen- ITI4 - EC Document -

Clément Chatelain

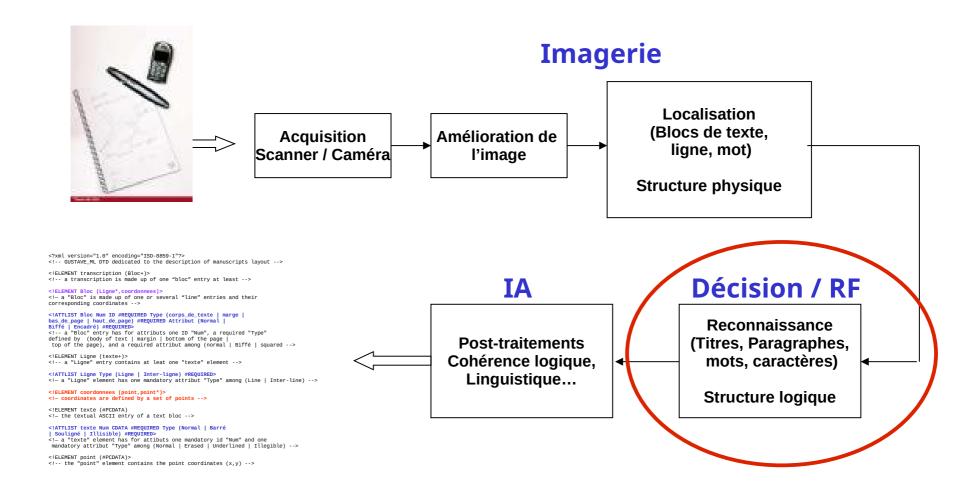
clement.chatelain@insa-rouen.fr

#### Plan du cours

- Introduction générale
- Introduction à la Reconnaissance de Formes
- Analyse et reconnaissance d'images de doc.
- Deep learning
- Retour sur le document

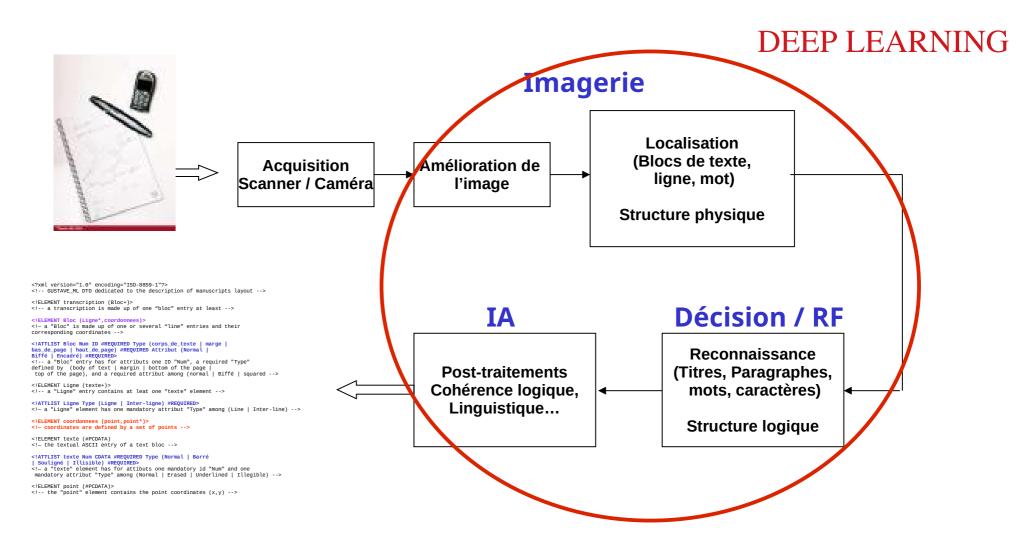
### Introduction Générale (old)

Des pixels ... au texte



### Introduction Générale (new)

Des pixels ... au texte



## RdF: application aux documents

Document papier → document numérisé → texte

#### Scanners industriels rapides





Jusqu'à 10 000 images A4 / heure saisies Dédiés pour le traitement automatique par lots en Noir et Blanc recto verso

Généralement dans l'industrie : images 200 ou 300 dpi compressées sans perte

## Introduction à la Reconnaissance de formes

- Une définition de la RdF:
  - J.P. Haton: « La RdF s'intéresse à la conception et à la réalisation de systèmes (matériels et logiciels) capables de percevoir, et dans une certaine mesure, d'interpréter des signaux captés dans le monde physique »
- Conception de systèmes automatiques qui reconnaissent les formes qu'on leur présente
- L'homme est le plus parfait des systèmes de RdF
- Copie des fonctions humaines :
  - Perception, analyse, interprétation
  - Capteur, extraction de caractéristiques, classification

### Un bébé est capable de reconnaitre cet animal ...



## Human Learning / Machine Learning

- Apprentissage
  - Règles difficiles à déterminer "à la main"
  - Comment apprendre à un enfant ce qu'est une vache ?
  - Apprentissage par l'exemple
- Notion de généralisation
  - Combien d'exemples ?
  - Réduction des signaux à l'essentiel
- Machine learning :
  - Apprendre, c'est généraliser

#### Plan du cours

- Introduction générale
- Introduction à la Reconnaissance de Formes
- Analyse et reconnaissance d'images de doc.
- Deep learning
- Retour sur le document

# Structure générale d'un système de RdF de documents (old)

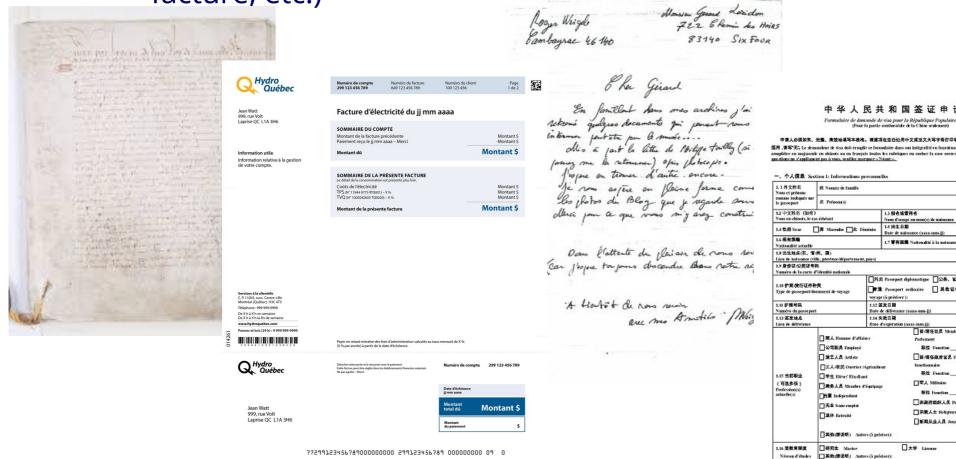
- 1. Prétraitements
  - Binarisation, localisation, débruitage, normalisation, etc.
- 2. Extraction des info pertinentes = les caractéristiques
  - Reconnaissance de la parole : analyse fréquentielle
  - Reconnaissance balles de couleur pour un robot : caractéristiques image : couleur, texture, rapport H/L, etc.
- 3. Représentation de ces informations en vue de leur classification
  - Vecteur, graphe, chaîne
- 4. Prise de décision
  - apprentissage/classification, combinaison/fusion
- 5. Vérification

# Structure générale d'un système de RdF de documents (new)

• 1. DEEP LEARNING.

## Analyse et Reconnaissance d'images de document

 i. Catégorisation automatique de documents (formulaire, courier, facture, etc.)



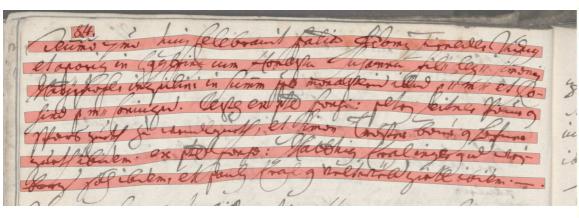
0863028# 98

::000 1 1 == 00 G ::

1.10 护黑旅行证件转类		□ 計选 Passeport ordinaire □ 其他证件(推误明) Autres documents or	
Type de passeport/document de vayage		voyage (à prédiser):	
1.11 护理号码		1.12 節度日間	
Numéro du passeport		Date de déliveance (auna-mm-jj)	
1.13 高发地点		1.14 失敗日間	
Lieu de délivrance		Date d'expiration (xxxx-nam ji)	
1.65 当前职业 《可选多项》) Profession(s) actuache(s)	開入 Homme d'affaire    公司販売 Engled    富工人売 Article    工人売販 Overlet Agricultur  学生 New Educatut    最秀人長 Membre d'éptings    国		■ 新教皇安長 Neutre de Dulcurent Anden messke d Partenouti 現在 Function □ 職務性就實度 Functionative on foundon' Anden foundomie 取在 Function □ 年代 Paration □ 非政策的基本 及 Personal of ONG □ 宗教人士 Religious □ 解於基本及 Personaliste' Tour méda

## Analyse et Reconnaissance d'images de document

- ii. "Document Layout Analysis" (DLA):
  - Identification des paragraphes, des images, des tableaux,
  - Segmentation de documents/paragraphes en lignes







Extraction d'information

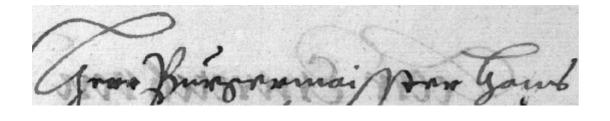
## Analyse et Reconnaissance d'images de document

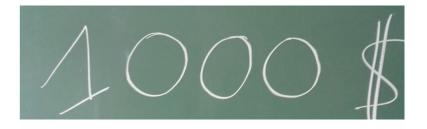
- iii. Reconnaissance de texte
  - Reconnaissance de caractères
  - Reconnaissance de mots / séquences numériques
  - Reconnaissance de lignes, de paragraphes



En effet, la vitesse ne doit pas être

mon sensemil le medlen,





## Analyse et Reconnaissance d'images de document

Tout ceci, avec du "deep learning"

 En 2017, l'immense majorité des articles présentés à la conférence ICDAR étaient basé sur du deep learning (https://dblp1.uni-trier.de/db/conf/icdar/icdar2017.html)

→ brève présentation du deep learning

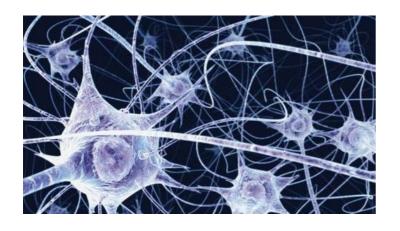
#### Plan du cours

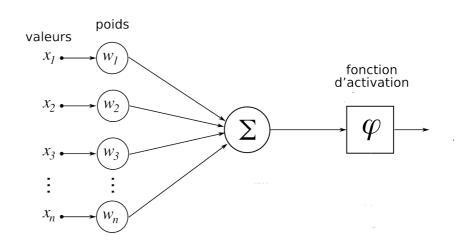
- Introduction générale
- Introduction à la Reconnaissance de Formes
- Analyse et reconnaissance d'images de doc.
- Deep learning
- Retour sur le document

# Apprentissage Profond, AKA "deep learning"

- Modèle d'apprentissage statistique
- Méthodes anciennes (~ 80's/90's), remises en lumière en 2006
- Aujourd'hui, IA = Deep learning
- Particularités
  - Modèles conséquents (gros en RAM) mais rapides en décision
  - Souvent plusieurs millions de paramètres
  - Apprentissage long et difficile
  - Nécessite de très nombreux exemples <u>annotés</u> pour l'apprentissage
  - Performances à l'état de l'art dans tous les domaines de vision par ordinateur / IA / reconnaissance de formes

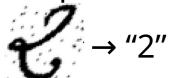
- Analogie avec le cerveau, modélisation du neurone biologique :
  - Somme pondérée + non linéarité

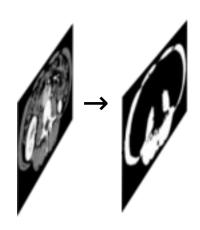




 Mise en réseau de nombreux neurones pour modéliser des fonctions mathématiques très complexes

Exemple de fonctions à apprendre :







"man in black shirt is playing guitar."



"girl in pink dress is jumping in



"construction worker in orange safety vest is working on road."



"black and white dog jumps over



"two young girls are playing with lego toy."

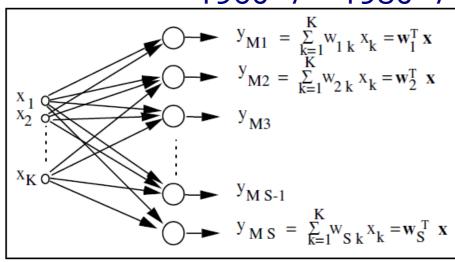


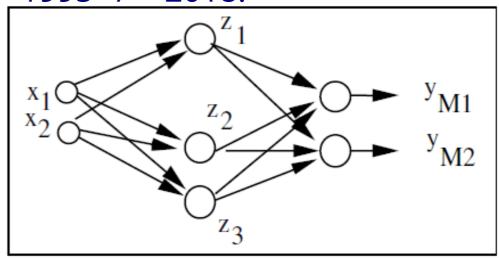
"young girl in pink shirt is swinging on swing."

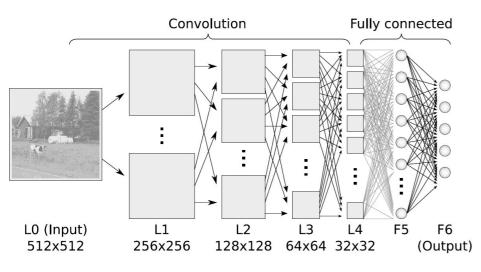
Juon serveuir le medleur, → "Je vous prie de ..."

• Le plus souvent, organisation en couches

1960 / 1980 / 1995 / 2018:

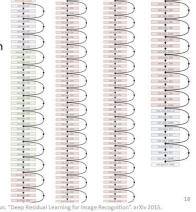






#### **Network Design**

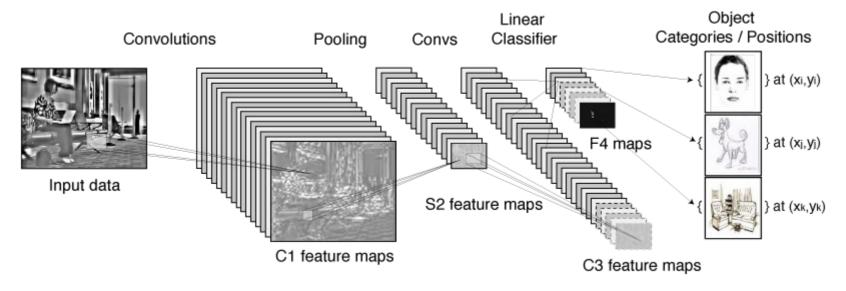
- ResNet-152
  - Use bottlenecks
  - ResNet-152(11.3 billion FLOPs) has lower complexity than VGG-16/19 nets (15.3/19.6 billion FLOPs)



- L'apprentissage consiste à régler les poids
- Sortie désirée pour une entrée donnée (nécessite les sorties = l'annotation !)
- Apprentissage par l'algo de la "rétropropagation du gradient"
- Algorithme:
  - Initialiser les poids au hasard
  - TANT QUE critère d'arret non satisfait, FAIRE :
    - POUR chaque exemple de la base d'app:
      - Propager les carac
      - Estimer l'erreur en sortie
      - Rétropropager l'erreur et corriger les poids
    - FINPOUR
  - FIN TANT QUE
- Tester sur une base différente de la base d'app

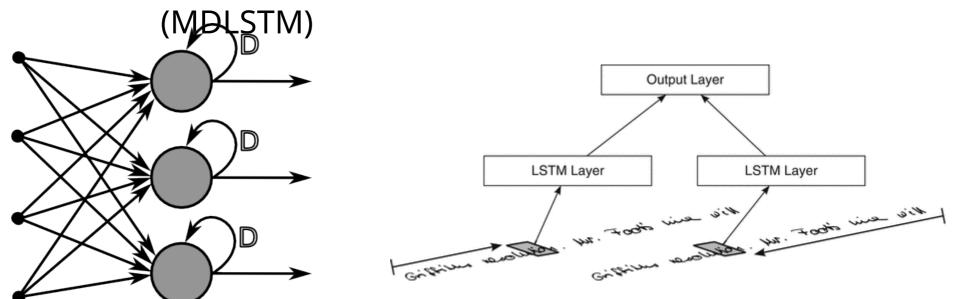
## Type de couches

- Couches "convolutionnelles": dédiées au traitement des images (CNN)
  - Léger (= peu de paramètres)
  - Permet de caractériser les images
  - Transfer Learning possible



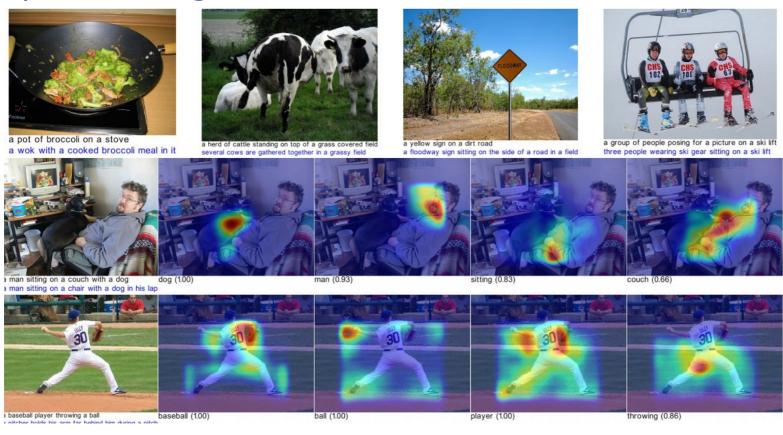
## Type de couches

- Couches récurrentes : dédiées au séquences à 1 ou plusieurs dimensions (RNN, LSTM)
  - Traitement par fenêtre glissante
  - Permet de mémoriser le passé (RNN, LSTM) ...
  - ... et le Futur (BLSTM) ... ou d'autres dimensions



## Type de couches

• Couches d'attention : pour apprendre à faire le focus sur une sous partie du signal d'entrée



 Couches classiques (denses) : couches de décision : classification ou régression

- Empilement des couches (CNN, LSTM, dense, etc.)
- One algo to beat them all: Backprop
- Légos : empilement de couches variées, adaptées au signal d'entrée/de sortie
- Il est désormais courant d'obtenir des architectures à plusieurs dizaines de couches.
  - M de paramètres
  - Difficile à apprendre
  - Demande beaucoup de données ...

• 2012 : imagenet



- 14M d'images annotées selon 5000 classes
- ImageNet Large Scale Visual Recognition Challenge (ILSVRC)
- Modèles appris rendus publics

#### Transfert Learning

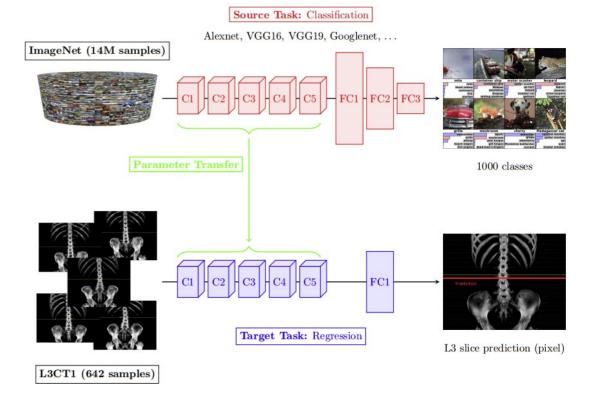
- Idée : un modèle ayant "vu" 14 millions d'images généralise bien.
- On peut transférer sa "connaissance" à un domaine proche
  - Prendre les couches convolutionnelles "basses" préapprises
  - Adapter la ou les dernières couches de décision
  - Apprendre sur un dataset métier plus petit de manière à apprendre les dernières couches = Fine Tuning

#### Réutilisation des représentations

Transfert Learning

#### Fonctionne avec:

- Images (imagenet)
- Texte (wikipedia)
- Séries temporelles (lag llama)
- Etc.



Aujourd'hui TL généralisé: perf ++, solutionne en grande partie le besoin en données, démocratisation de l'IA (Vous vous servez du transfer learning tous les jours ...)

#### Plan du cours

- Introduction générale
- Introduction à la Reconnaissance de Formes
- Analyse et reconnaissance d'images de doc.
- Deep learning
- Retour sur le document

# Reconnaissance de caractères imprimés (OCR)

- Pas toujours si facile ...
- CNN
- Problème ~ résolu
- Erreur < 1% généralement</li>

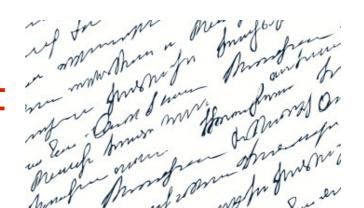


En effet, la vitesse ne doit pas être

## Reconnaissance de caractères manuscrits

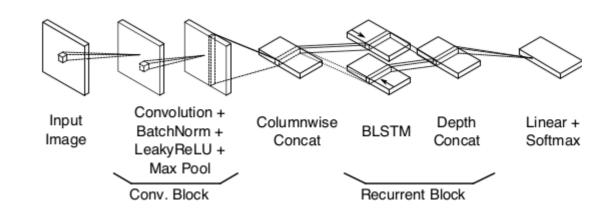
- Problème plus difficile
- CNN
- Problème ~ résolu ...
- ... mais problématique peu réaliste!
- Les caractères manuscrits sont rarement isolés ...





#### Reconnaissance de mots manuscrits

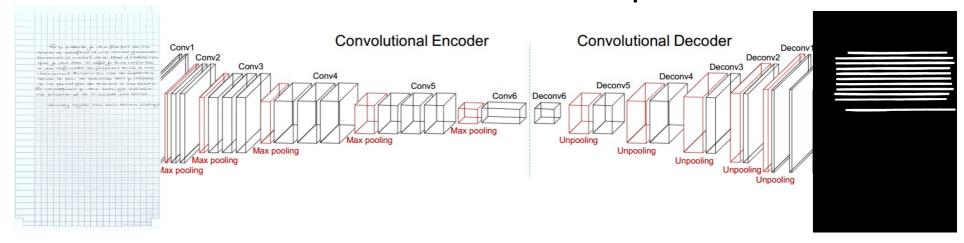
CNN + LSTM



- Taux d'erreur intéressants (WER ~ 5-25% selon les problèmes)
- … mais problématique toujours peu réaliste!
  - À part certains car particuliers (chèques, adresses), les mots manuscrits sont rarement isolés ...
  - → nécessité de traiter des documents complets!

### **Document Layout Analysis**

- Segmentation en lignes ou en zones
- FCN
  - Architecture convolutionnelles pures



- Léger, s'adapte à la taille de l'image

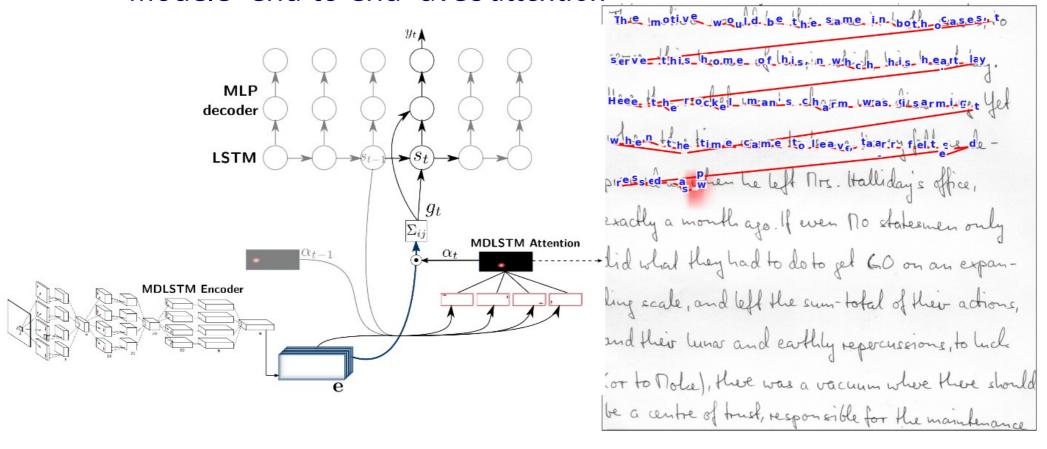
## Reconnaissance de lignes de texte

- Segmentation en ligne nécessaire
- Généralement, 2 étapes :
  - 1) Reconnaissance optique (CNN, LSTM):
    - "j ' ai malencontrousement marchésur le chion de ma vosrre. cele ci , tenant brauemryp à son chien , a éle parthaliérement afestée par l 'acaident et me demande un dédammagement de taille"
  - 2) Exploitation de connaissances linguistiques
    - Au niveau caractères
    - Au niveau mots
    - Utilisation de lexiques

## Reconnaissance de lignes de texte

Sans segmentation en ligne (prémices)

Modèle "end-to-end" avec attention



## Aujourd'hui

- Modèles profonds préappris dédiés aux images de Documents :
  - Donut (Document Understanding Transformer) encoder pré appris de documents

https://huggingface.co/docs/transformers/model\_doc/donut

- Nougat (fb, "Neural Optical Understanding for Academic Documents") https://facebookresearch.github.io/nougat/
- DAN (Document Attention Network) démo https://github.com/FactoDeepLearning/DAN
- DANiel (DAN + combinaison avec LLM)

#### Conclusion

- Techniques matures
- Besoin industriels plutôt en déclin :
  - B2B de + en + numérique
  - Courier perso en baisse
  - Chèques en baisse
  - Écriture manuscrite en baisse d'une manière générale ...

- En revanche besoins importants sur :
  - Le patrimoine : documents historiques, Genese d'une oeuvre (Bovary), etc.
  - L'archivage : documents financiers, archives du CHU, etc.

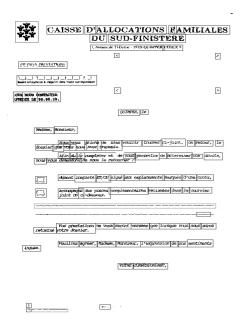
### Annexes

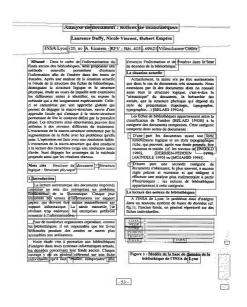
## Applications industrielles

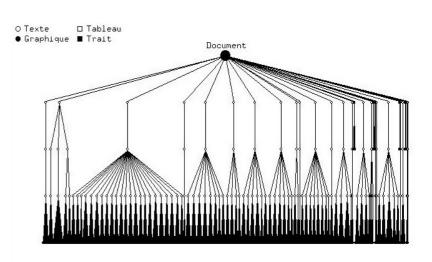
 1985 : premières machines de tri postal grâce à la lecture de caractères imprimés



 1992 : traitement des documents structurés formulaires, ouvrages imprimés de bonne qualité, etc.







## Applications industrielles

 1995 : lecture d'adresses postales manuscrites bureau distributeur

MR Edward J. 7438 Claar Run Drive Warrenton, VA 20187 Ste F 15 rue Boudeville 21 Thibaud 31000 Toulouse TOM NBC TV 30 ROCKEFELLER PLAZA NEW YORK NY 10112

1995 : lecture automatique de chèques



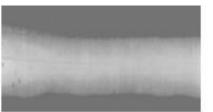
### Applications industrielles

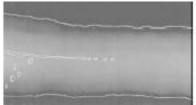
- 1999 : Recensement
  - 90 millions de formulaires
  - Collaboration La poste-Bull-Jouve
  - 97% des caractères reconnus
  - 1% d'erreur

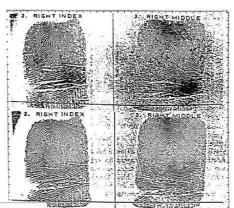


## De nombreuses applications ...

- Banque : reconnaissance de signature
- Robotique : Guidage de robots
- Météorologie : détection de phénomènes à partir d'images satellites
- Police : reconnaissance d'empreintes, de visages
- Industrie : détection de défauts, automatisation

















## De nombreuses applications ...

- Médical : détection de pathologie à partir d'ECG/EEG, imagerie : aide à la décision
- Militaire : détection de cibles
- Sécurité automobile : détection d'obstacles
- Traitement du signal : reconnaissance de la parole, de l'écriture
- etc.



Odrio 7 ou glithographies de Balsac. elle sont petites

ai fait photographies un tres

beau postel (Grature) de Gust

qui est au music de Eours où

il y a aussi un dessin de Boulangol

