



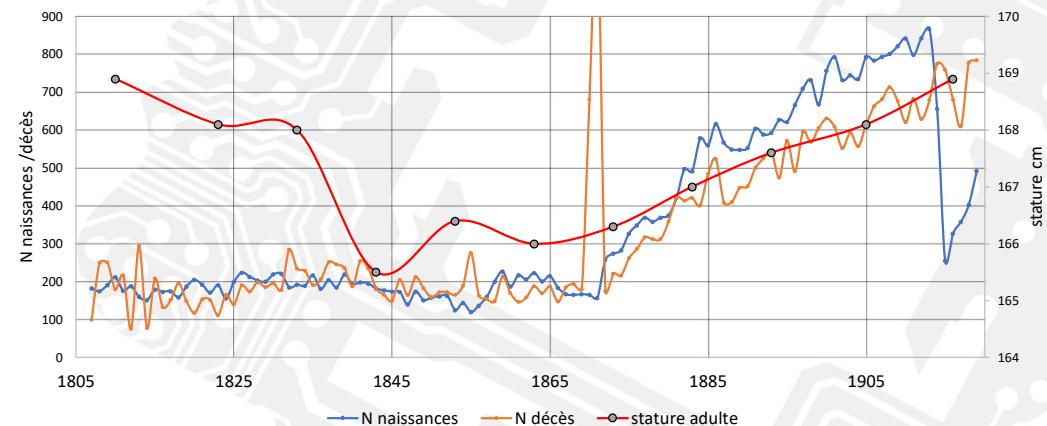
Belpop: a history-computer project to study the population of a town during early industrialization

Introduction



N of births (blue) and deaths (red) and stature of conscripts (red with points) in Belfort during early industrialization (Heyberger 2013 & Belpop project)

naissances & reconnaissances, décès (état civil) et stature des conscrits à Belfort
(1807-1919)



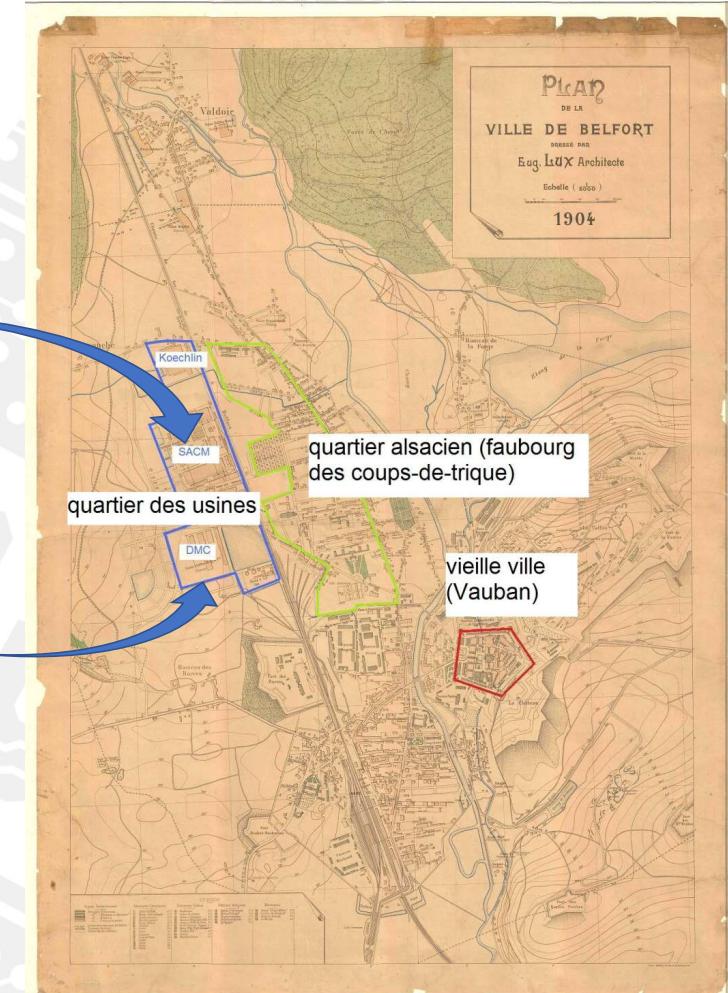
Introduction



one of the many barracks built after 1870

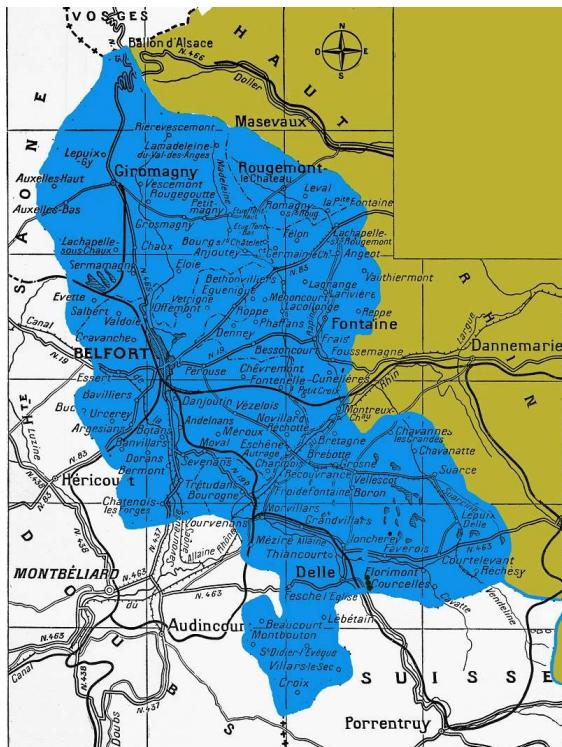


new factories (blue) and new working-class district (green) in Belfort (old town pre-dating 1870 in red)

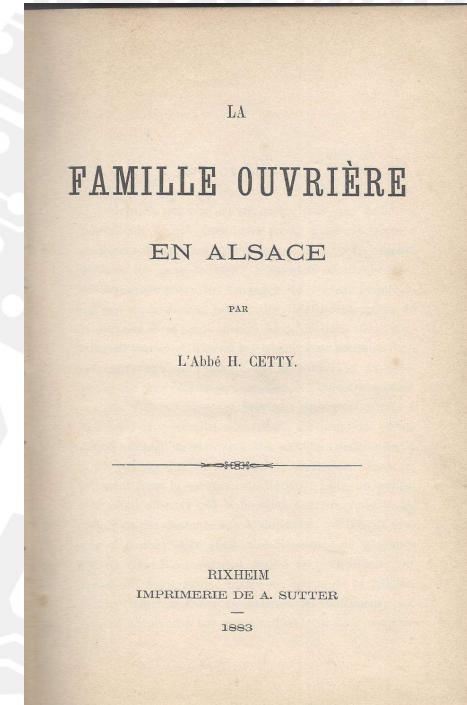


Introduction

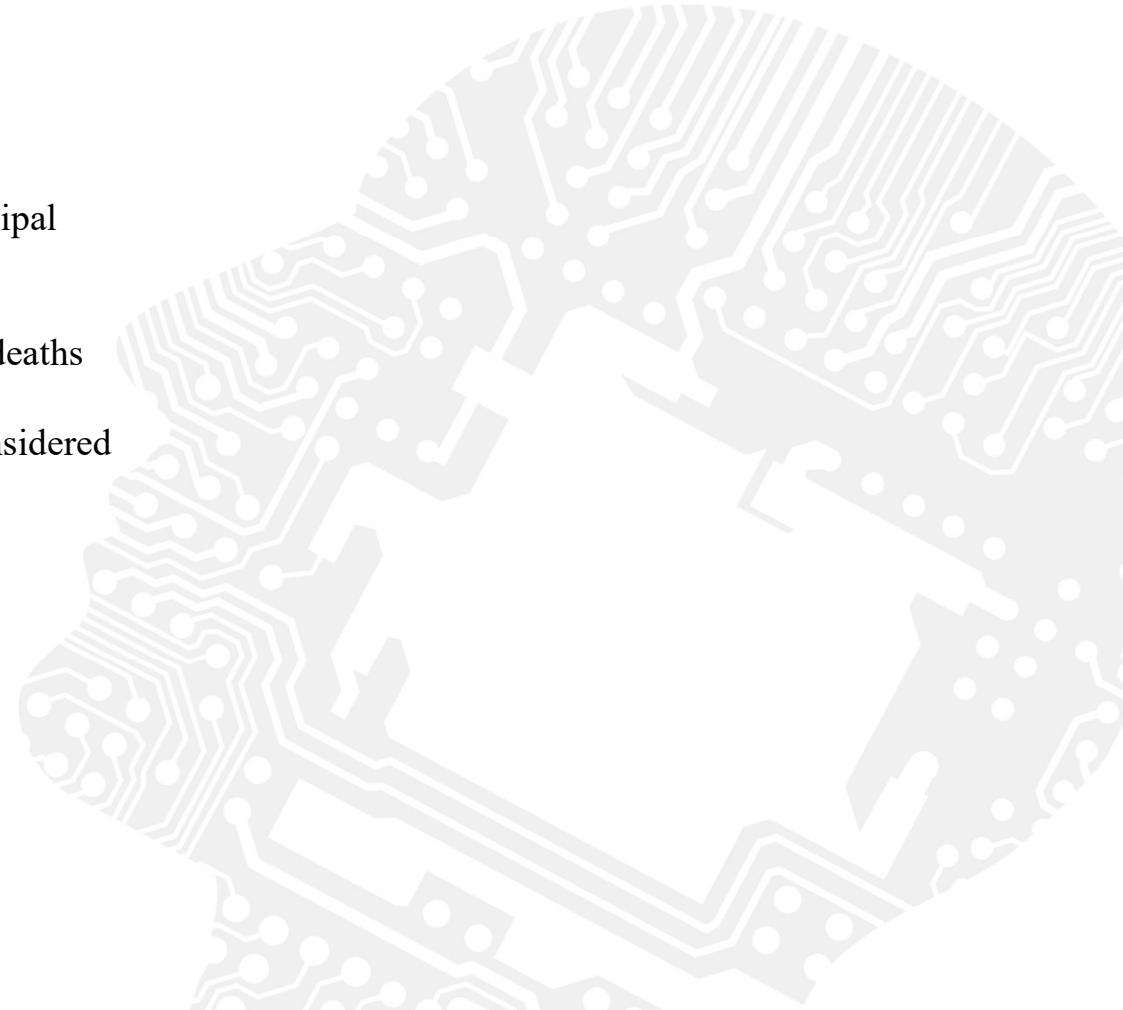
New border between Territoire de Belfort (blue) and Alsace (green)



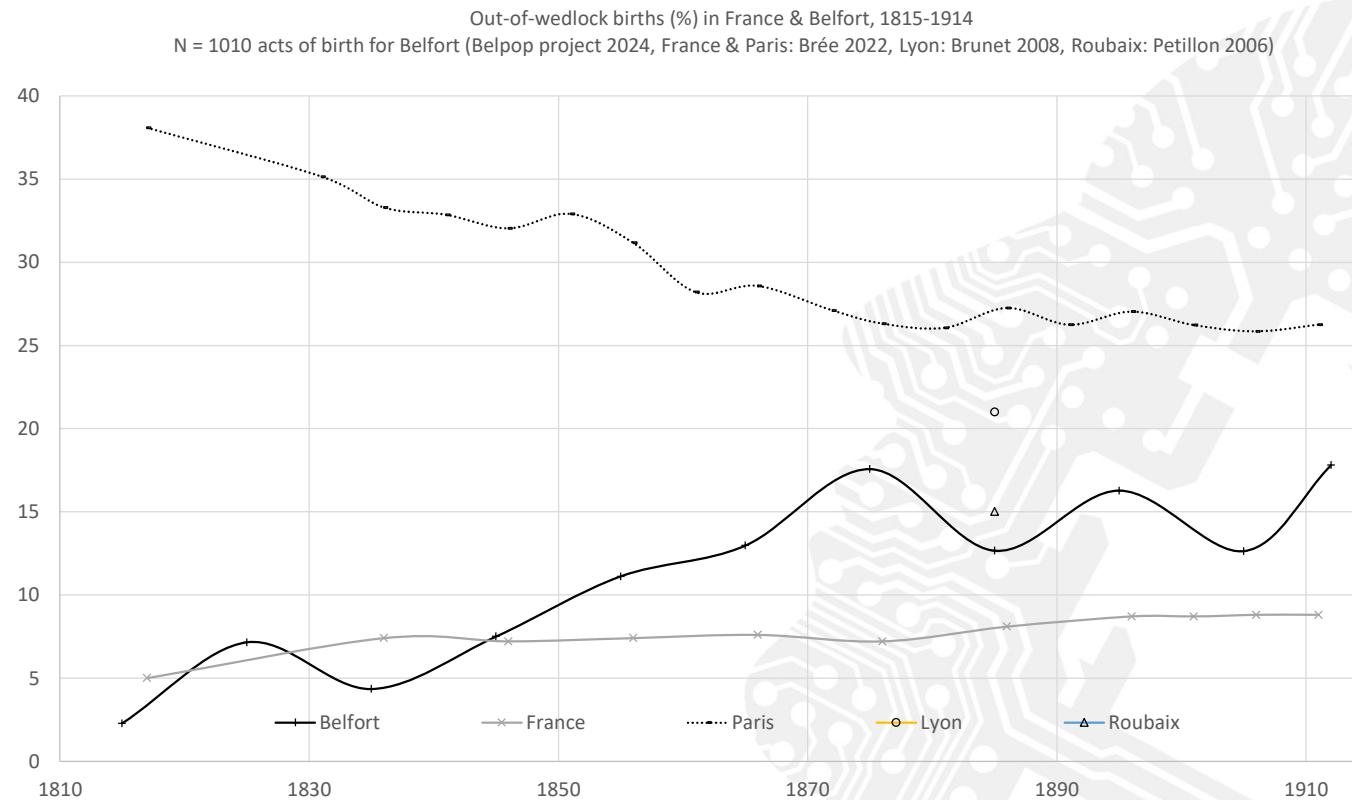
Workers « parisieren »
(Cetty 1881)



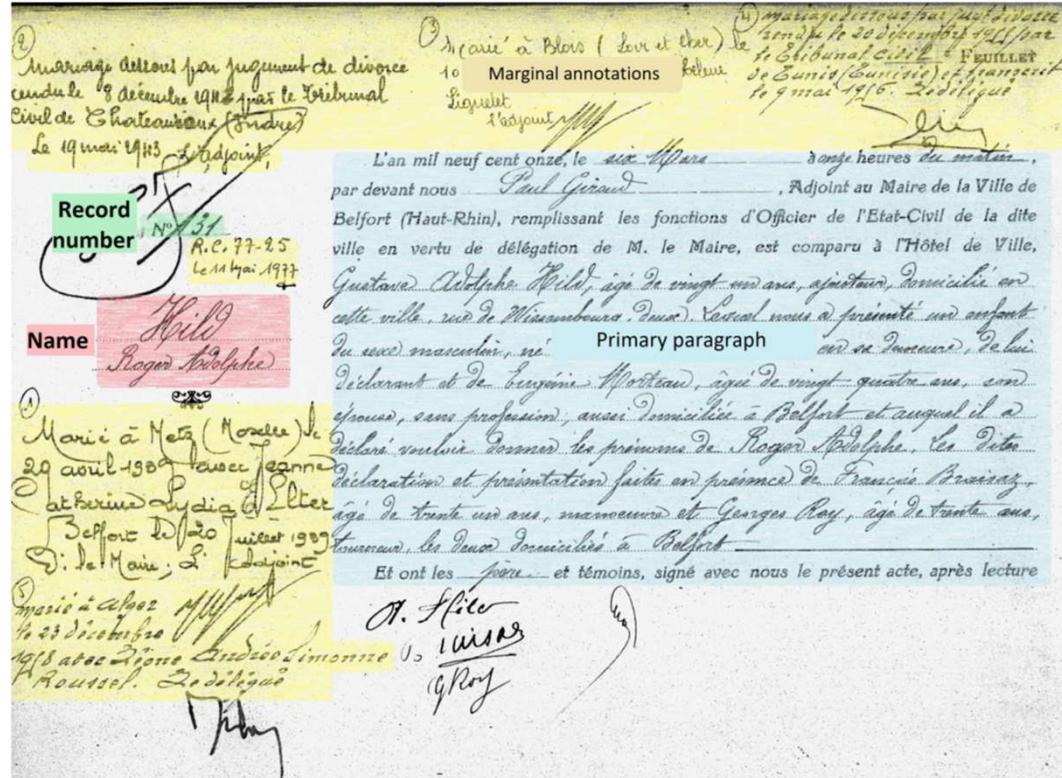
- birth and death certificates already scanned by the Belfort municipal archives for the years 1807 to 1919
- approximately 40,000 records for births and 38,000 records for deaths
- 2 random samples of just over 1,000 acts for each of the sets considered



Introduction



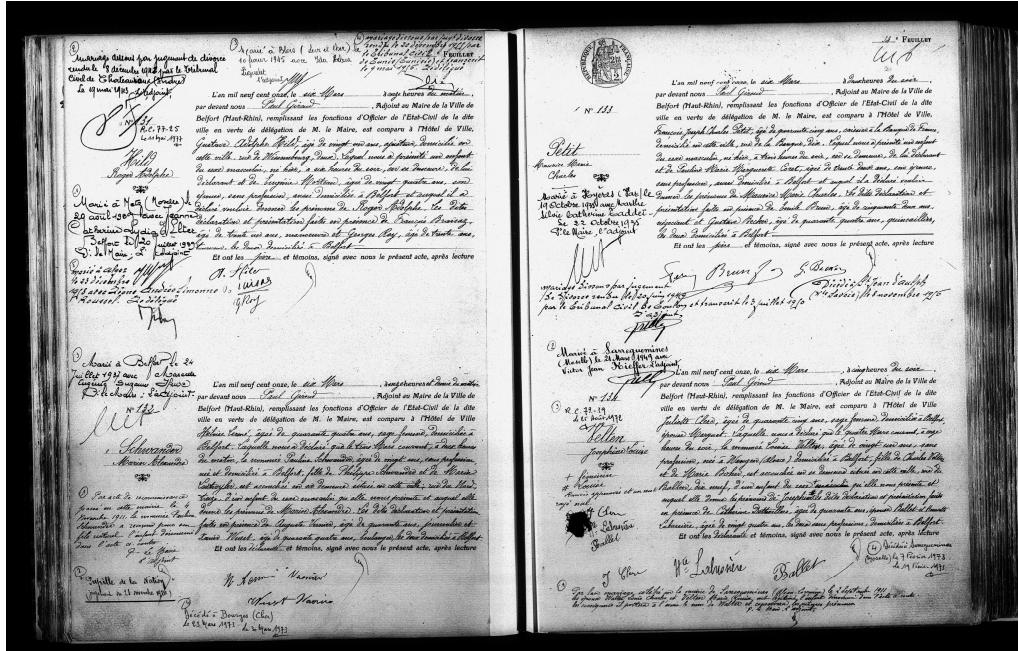
Our dataset



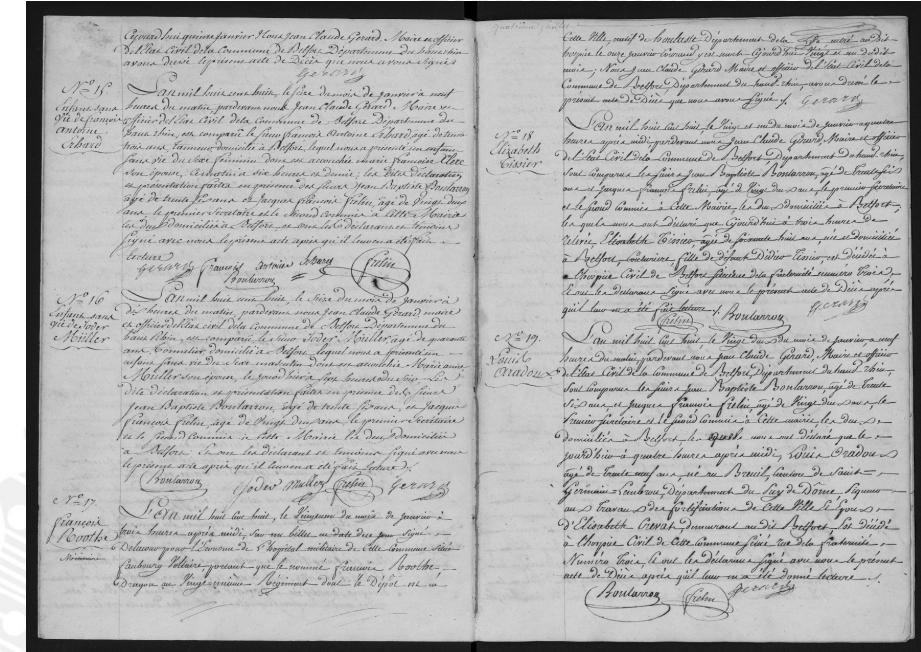
A sample record from Belfort civil registers of birth with annotated components [1]

Our dataset

- The complexity of our dataset lies in different aspects

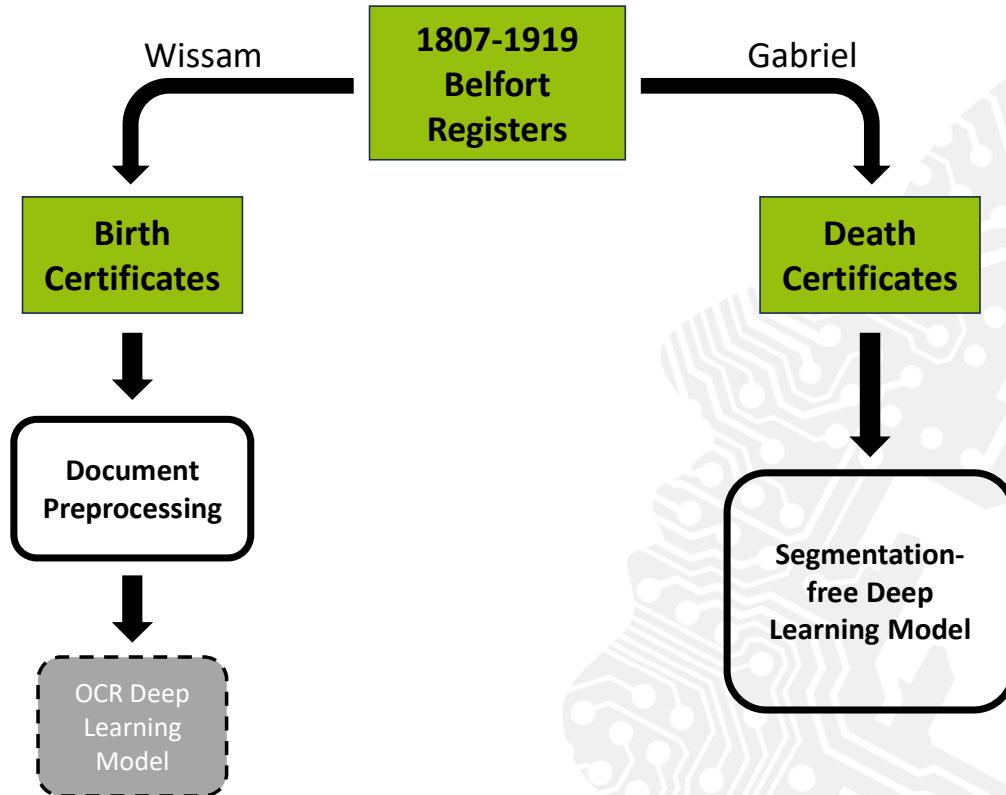


Sample birth record from 1911



Sample death record from 1808

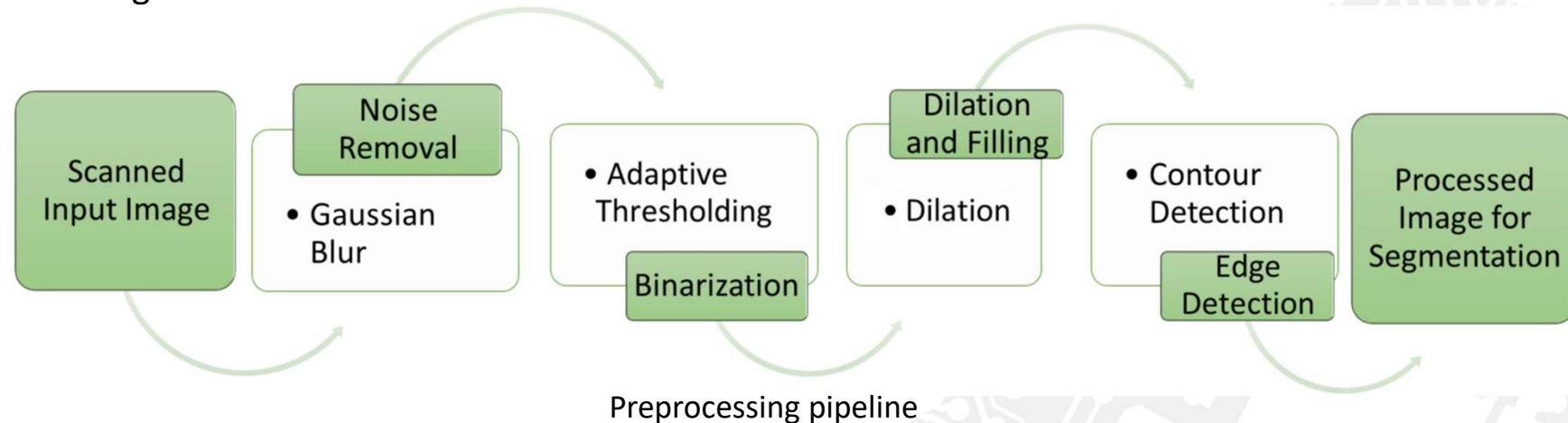
Computer Vision strategies





Birth Certificates Transcription

- Preprocessing



L'an mil neuf cent onze, le Trois Avril à Dix heures et demie du matin, par devant nous Paul Goriot, Adjoint au Maire de la Ville de Belfort (Haut-Rhin), remplissant les fonctions d'Officier de l'Etat-Civil de la dite ville en vertu de délégation de M. le Maire, est comparu à l'Hôtel de Ville, Henri Louis Rudolphe Steinmann, âgé de vingt-neuf ans, ingénieur, domicilié à Nancy (Vaudreuil et Marville), également muni d'un acte de naissance masculin, né le premier Avril courant, à Belfort dans le arrondissement de cette ville, de la Châtaigneraie, trente deux. De lui déclarant et de François Marie Rose Marion, âgé de vingt-deux ans, professeur, domicilié à Nancy, marié récemment à Belfort et auquel il a déclaré tenir de son père le prénom de Jean Marie Pierre à la date de déclaration et présentation faite au juge de paix de Belfort, âgé de vingt-six ans, agent d'assurances, les deux domiciliés à Belfort. Et ont les père et témoins, signé avec nous le présent acte, après lecture



Grayscale record

L'an mil neuf cent onze, le Trois Avril à Dix heures et demie du matin, par devant nous Paul Goriot, 2, Adjoint au Maire de la Ville de Belfort (Haut-Rhin), remplissant les fonctions d'Officier de l'Etat-Civil de la dite ville en vertu de délégation de M. le Maire, est comparu à l'Hôtel de Ville, Henri Louis Rudolphe Steinmann, âgé 5 de vingt-neuf ans, ingénieur, Domicilié à Nancy (Vaudreuil et Marville), également muni d'un acte de naissance masculin, né le premier Avril 6, à Belfort dans le 7 arrondissement, de cette ville, de la Châtaigneraie, 8, trente deux. De lui déclarant et de François Marie Rose Marion, âgé de vingt-deux ans, professeur, domicilié à Nancy, marié récemment à Belfort et auquel il a déclaré tenir de son père le prénom de Jean Marie Pierre à la date de déclaration et présentation faite au juge de paix de Belfort, âgé de vingt-six ans, agent d'assurances, les deux domiciliés à Belfort. Et ont les père et témoins, signé avec nous le présent acte, après lecture

Bluring & Binarization

Contour detection for text line extraction

Birth Certificates Transcription

- Skew correction

V Delfort et auquel il a déclaré voulois donner le prénom de



V Delfort et auquel il a déclaré voulois donner le prénom de

Skewness correction example [1]

Birth Certificates Transcription

- XML File Generator

```
<-<page Name="1E43_0003.jpg" Type="Double page" ImageHeight="4195" ImageWidth="5896">
  -<ReadingOrder>
    -<RegionReflIndexed Index="1" Type="Title-number"/>
    -<RegionReflIndexed Index="2" Type="Title-name"/>
    -<RegionReflIndexed Index="3" Type="Paragraph"/>
    -<RegionReflIndexed Index="4" Type="Title-number"/>
    -<RegionReflIndexed Index="5" Type="Title-name"/>
    -<RegionReflIndexed Index="6" Type="Paragraph"/>
    -<RegionReflIndexed Index="7" Type="Paragraph"/>
  -</ReadingOrder>
  -<TextRegion Index="1" Type="Title-number">
    <Coords Points="(388, 1311, 146, 141)">
    -<TextEquiv>
      <Unicode>N°3</Unicode>
    -</TextEquiv>
  -</TextRegion>
  -<TextRegion Index="2" Type="Title-name">
    <Coords Points="(111, 1437, 576, 282)">
    -<TextEquiv>
      <Unicode>Denise Brunschwick</Unicode>
    -</TextEquiv>
  -</TextRegion>
  -<TextRegion Index="3" Type="Paragraph">
    <Coords Points="(687, 1326, 2200, 1476)">
    -<TextEquiv>
      -<Unicode>
        L'an mil huit cent vingt neuf, le trois du mois de Janvier, à onze heures du matin, par devant nous a présenté un enfant du sexe féminin né hier matin à huit heures, de lui déclarant et d David Levy, âgé de vingt huit ans, les deux marchands drapiers, domiciliés à belfort. Et ont
      -</Unicode>
    -</TextEquiv>
    -<TextLine ID="0">
      <Coords Points="(687, 1326, 2200, 163)">
      -<TextEquiv>
        -<Unicode>
          L'an mil huit cent vingt neuf, le trois du mois de Janvier,
        -</Unicode>
      -</TextEquiv>
    -</TextLine>
    -<TextLine ID="1">
      <Coords Points="(687, 1501, 2200, 67)">
      -<TextEquiv>
        -<Unicode>
          à onze heures du matin, par devant nous François-Joseph
        -</Unicode>
      -</TextEquiv>
    -</TextLine>
  -</TextRegion>
```

The diagram illustrates the hierarchical structure of the XML file. It uses orange brackets to group elements by level:

- Record number:** Groups the first two elements under the `<TextRegion Index="1" Type="Title-number">` and `<TextRegion Index="2" Type="Title-name">` tags.
- Record name:** Groups the third element under the `<TextRegion Index="3" Type="Paragraph">` tag.
- Paragraph level:** Groups the fourth element under the `<TextRegion Index="4" Type="Title-number"/>` tag.
- Text line level:** Groups the fifth element under the `<TextRegion Index="5" Type="Title-name"/>` tag.

Example of a generated XML file

Birth Certificates Transcription

- Metrics and results :

$$IoU = \frac{|\text{predicted mask} \cap \text{ground truth mask}|}{|\text{predicted mask} \cup \text{ground truth mask}|}$$

$$DR = \frac{O2O}{N_{\text{gt}}}$$

$$RA = \frac{O2O}{N_{\text{pred}}}$$

$$FM = 2 \times \frac{DR \times RA}{DR + RA}$$

Component	Accuracy			
	IoU	DR	RA	FM
Primary paragraphs	93.3%	97%	95.2%	95.9%
Marginal annotations	77%	83%	82%	82.5%
Saint Gall	98.3%	98%	98%	98.04%
RIMES	96.6%	95%	94%	94.4%

Accuracy results on different datasets

System	Accuracy (%)
DOCsumo	97
Ocelus	99
Transkribus	97
Our method	95.9

Accuracy comparison of this method with commercial AI systems

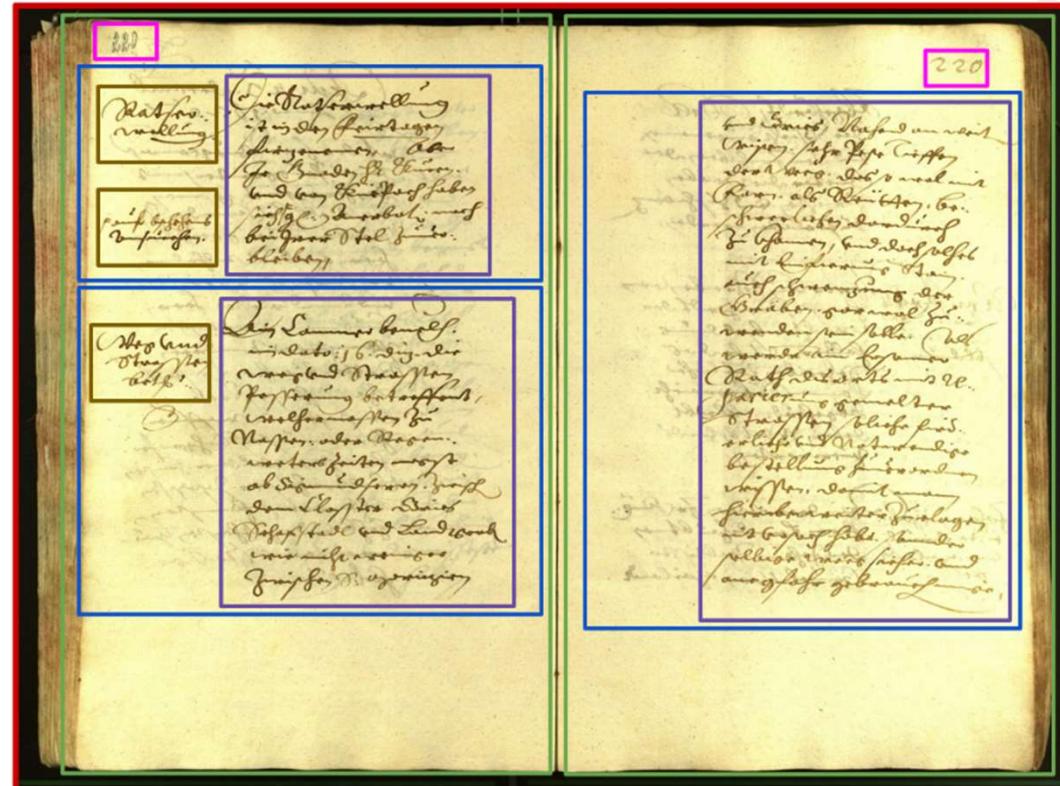
Death Certificates Transcription



Passage en revue du moral des troupes.

Death Certificates Transcription

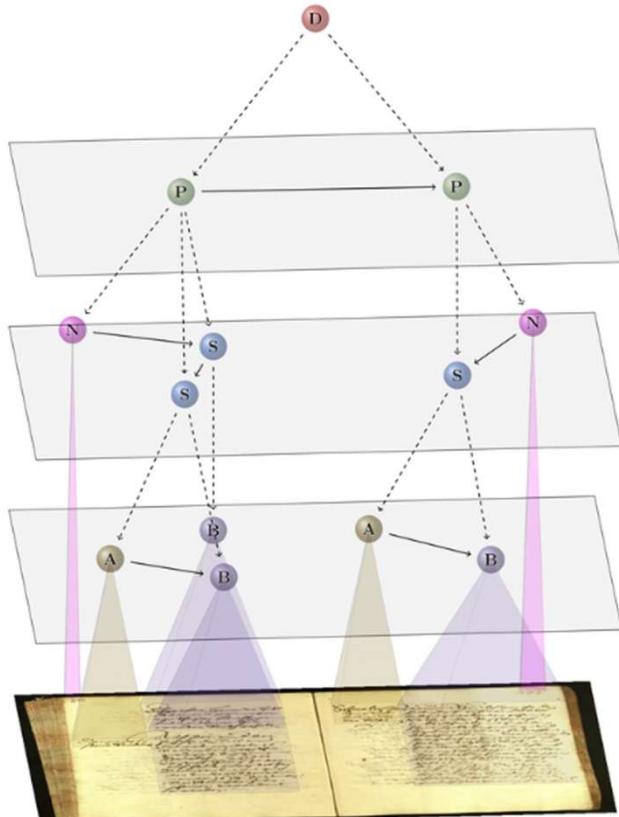
- The Document Attention Network [2] :



DAN segmentation results from READ 2016

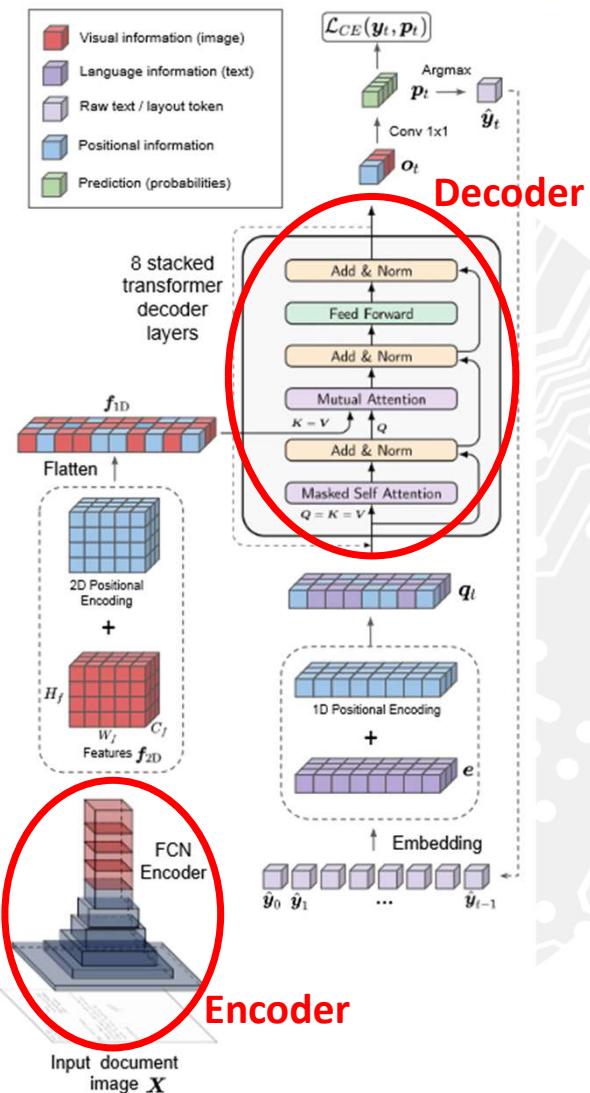
Death Certificates Transcription

- The Document Attention Network [2] :



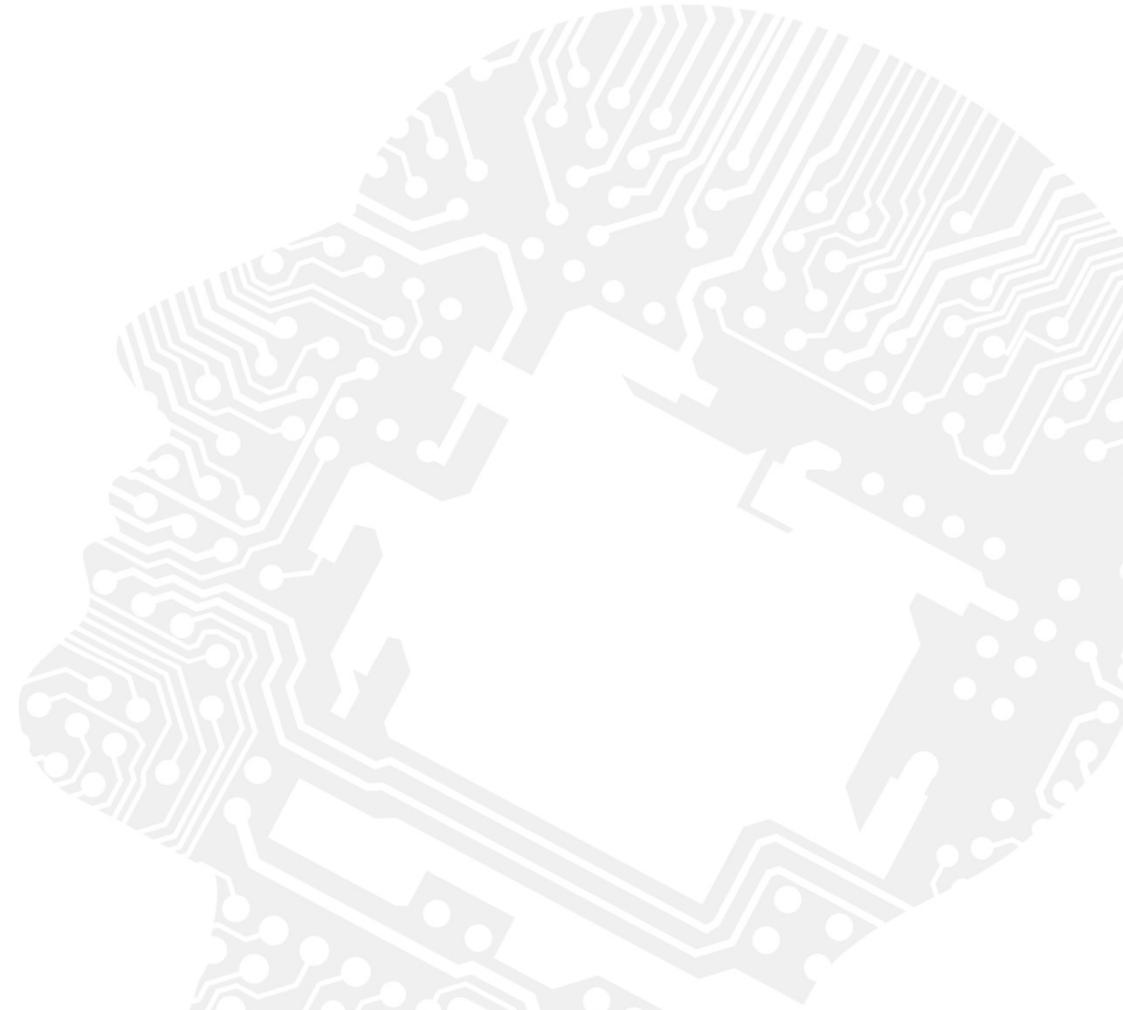
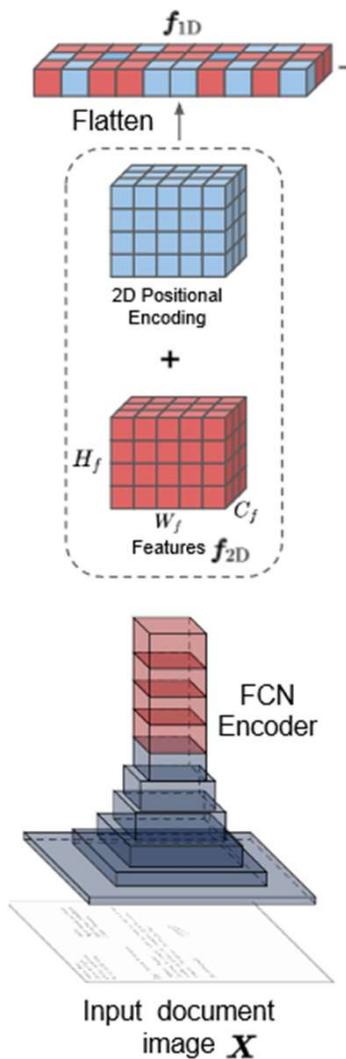
<document>
<page>
<page_number> 204
</page_number>
<section>
<body> Schgrafer, [...] gehal
</body>
</section>
<section>
<annotation> Genneral [...] Raitung
</annotation>
<body> Auf den: [...] werden,
</body>
</section>
</page>
<page>
<page_number> 204
</page_number>
<section>
<annotation> Schmalz. [...] bet:
</annotation>
<body> Verer [...] dar
</body>
</section>
</page>
</document>

Ground truth



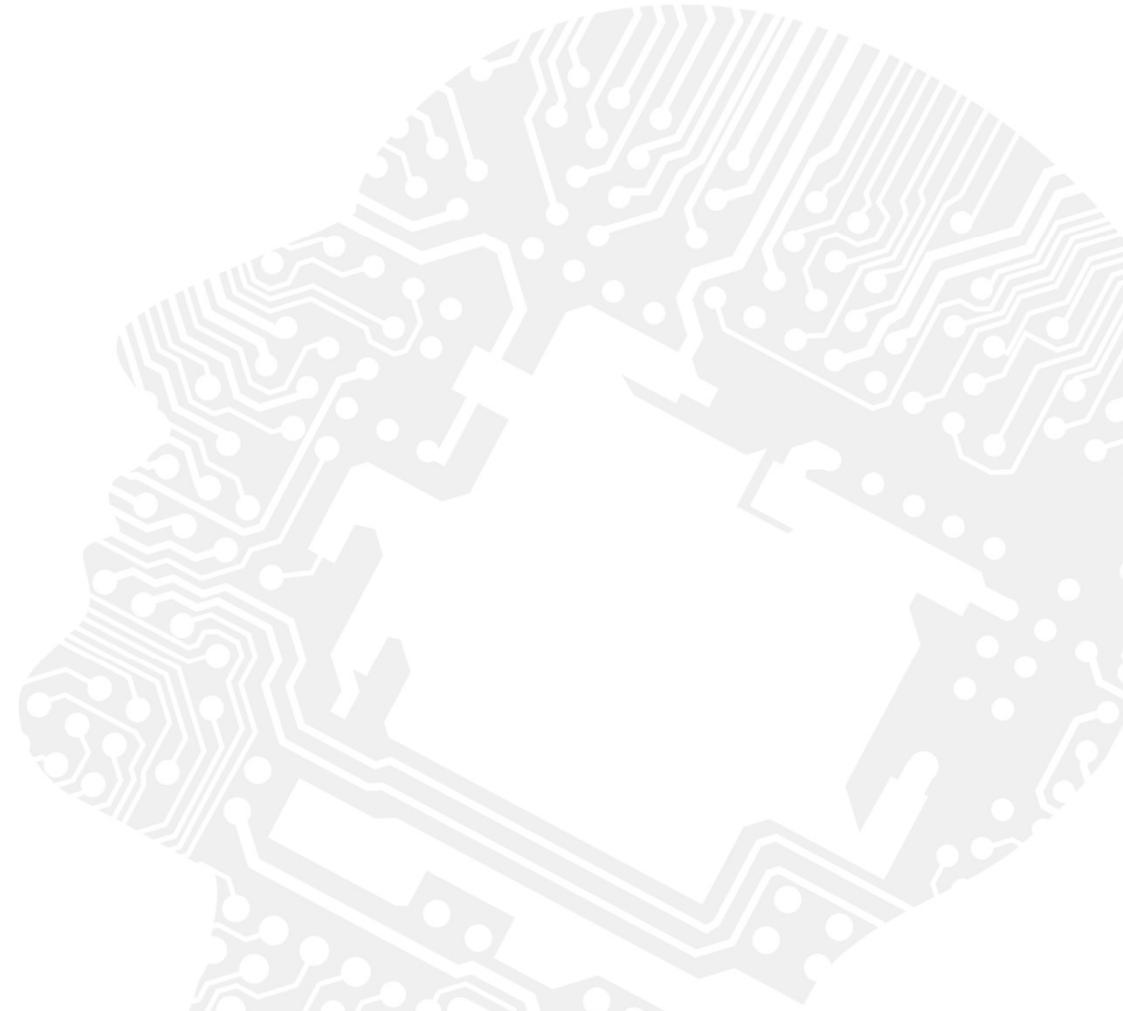
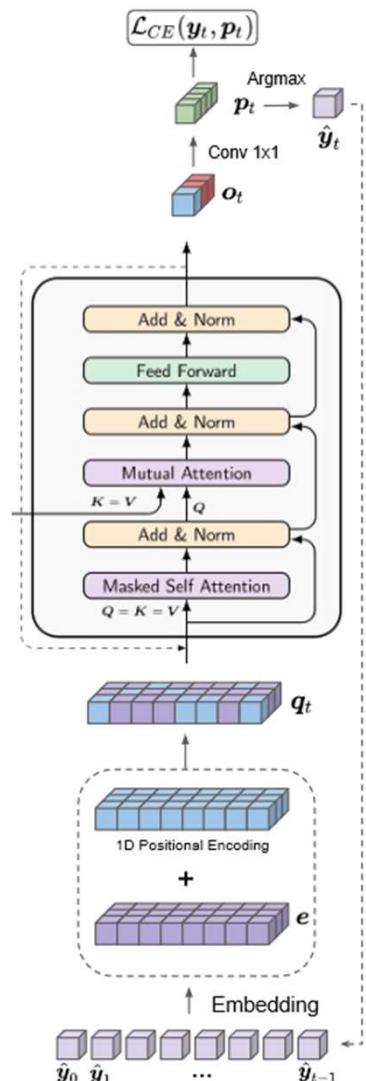
Death Certificates Transcription

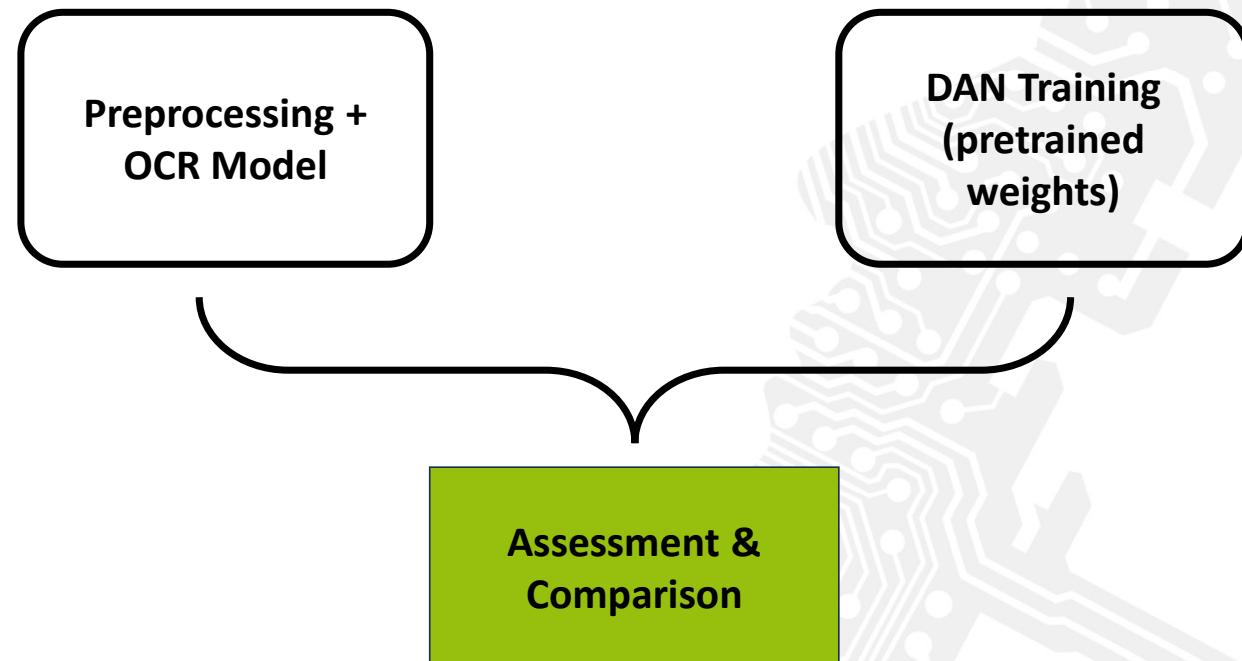
- The FCN Encoder:



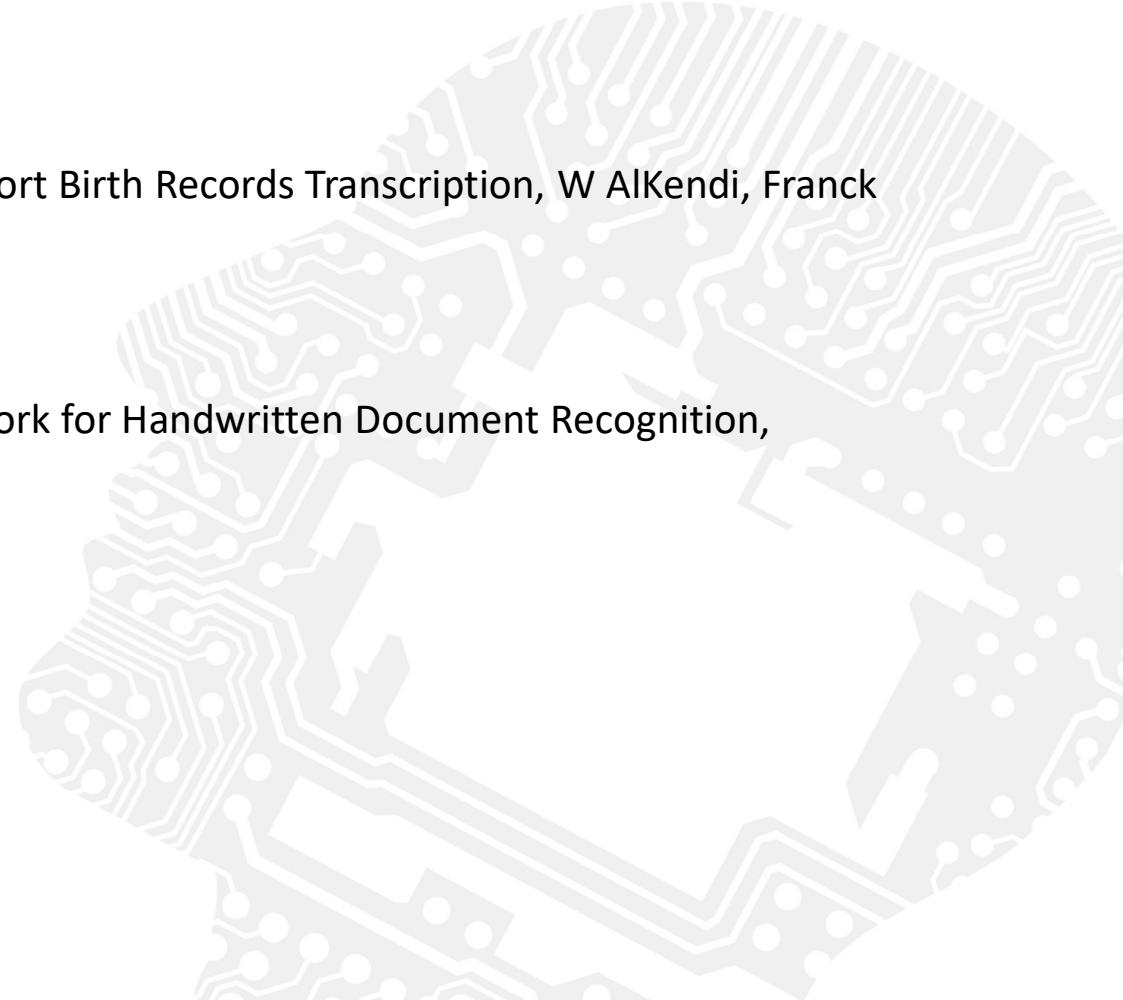
Death Certificates Transcription

- The Transformer :





- [1] Preprocessing and Data Structuring Technique for Belfort Birth Records Transcription, W AlKendi, Franck Gechter, Laurent Heyberger, Christophe Guyeux, 2024
- [2] DAN : a Segmentation-free Document Attention Network for Handwritten Document Recognition, D. Coquenet, C. Chatelain, T. Paquet, 2022



Thank you for your attention

Any questions ?

