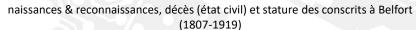


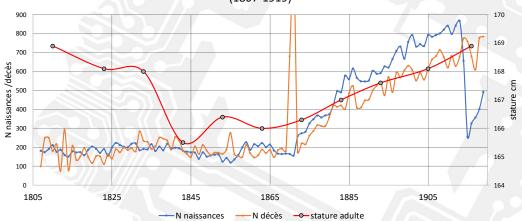




Belfort & Alsace-Moselle in white

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















Introduction

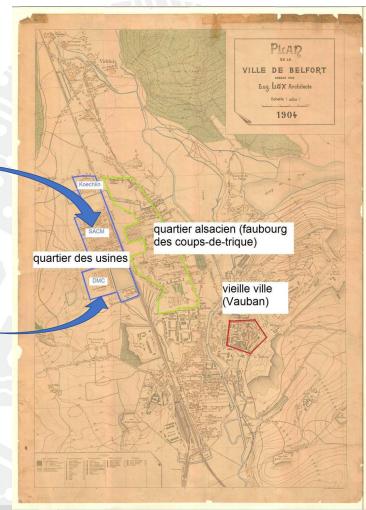




one of the many barracks built after 1870



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







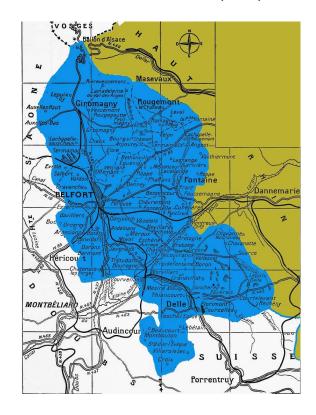




RÉPUBLIQUE INTRODUCTION Liberte legalité Praternité



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



Workers « parisieren » (Cetty 1881)













RÉPUBLIQUE FRANÇAISE Liberté Égatité Production



- ➤ 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





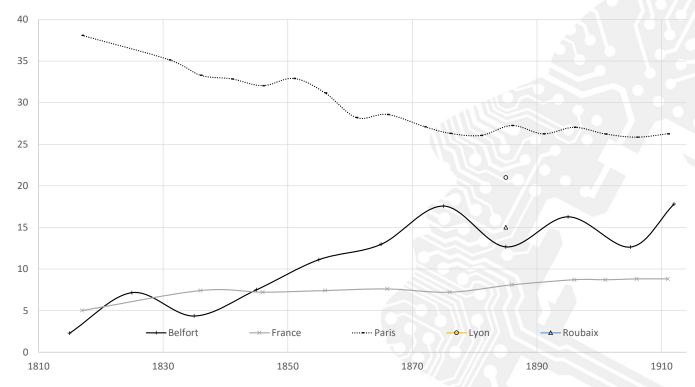




RÉPUBLIQUE FRANÇAISE Liberte fruiteraité



Out-of-wedlock births (%) in France & Belfort, 1815-1914 N = 1010 acts of birth for Belfort (Belpop project 2024, France & Paris: Brée 2022, Lyon: Brunet 2008, Roubaix: Petillon 2006)















②	Maria politica de Blors (box et elm) le rende le 20 de empte reffert
manage allow for juguent	Marginal annotations De Guned Guneste) of Francisco
Civil de Chateausaux Frances	Librard Signal Signal Mill
Le 19 mai 4/13 Litadjoint	
	L'an mil neuf cent onze, le six Mars donze heures du matin ,
Record /	par devant nous Saul Giraud , 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,
number R.C. 77-25	quatore Abeliha Mill, ago 2a ringt un ans, aportano, americalis en
Me on	
Name Spill	catte villa, rive de Missamboura Deux . La scal mois a presente un arefact
Roger Bolphe	Que se marculai, ne Primary paragraph an se Domeure , De lui
(A)	Dicharant at de benjaine Marteau, agui de vingt guestre aux, con
Marie à Metz (Moselee) le	eprouse, exer profession, ansei demicibile a Balfort et augual il a
20 anil 1980 lauser Jeanne	Declar's melvie Borner les primme de Ragar Galfahe, Les Dites
Catherine Lydia & Plie	declaration at presentation fails an presence de Trancis d'arrivag.
D: le Maire: 2 Edogaint	Tournand, les Denois Possessi les a Balfert
O C. TE COME, OF THE PERSON OF	Et ont les et témoins, signé avec nous le présent acte, après lecture
mario à celor Ally	of file h
16 23 dicomifro VI dioce din	1 Lucio)
Roussel. Zedeligue	(gRoy
, Juny	

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







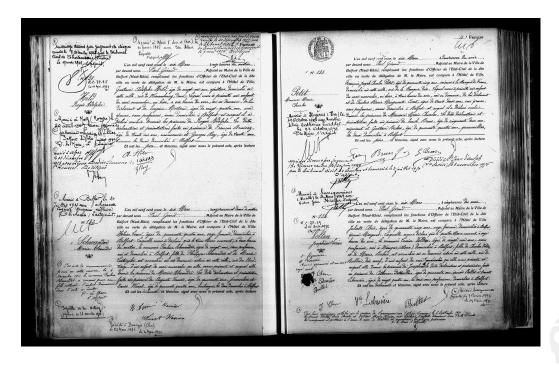




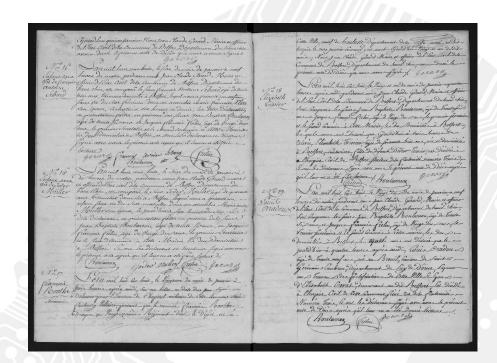
RÉPUBLIQUE Our dataset



• The complexity of our dataset lies in different aspects



Sample birth record from 1911



Sample death record from 1808





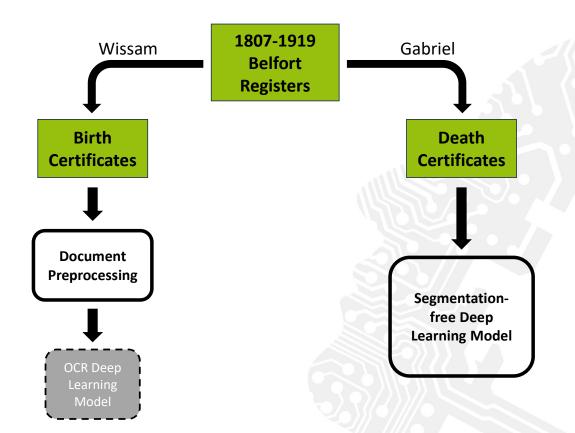






RÉPUBLIQUE Computer Vision strategies Liberte Equitien Française Française



























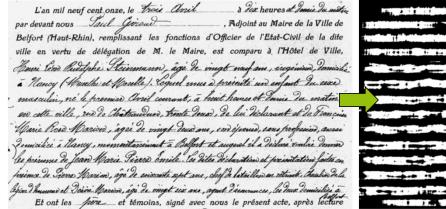




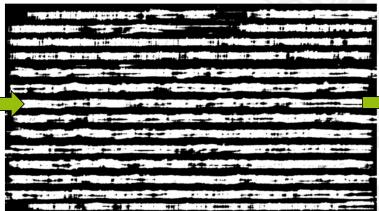
Preprocessing

Dilation Noise and Filling Removal Scanned Adaptive Contour Processed Input Image **Thresholding** Dilation Detection Gaussian Image for Blur Segmentation Edge Binarization Detection

Preprocessing pipeline



Grayscale record



Bluring & Binarization

L'an mil neuf cent onze, le Frais avil à Lex heures at Jamie du mitige par devant nous Jaid Girout , 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, Jani leva Redalpha Iteinmann, and de ingt nay am, inginian somicilie a Manoy (Mouther at Moulla). Loquet wour a presente un aufant In sace asculir, ne le promise avril courants, à heut haves et Somie du matin on colle ville, reve de Chateauder, trent Jeux, de lui léclarant et le Françoise Marie Rosa Marion, agai de mingt and me, can épousa, sans profession, ausai Pomicilice a Many, mammitaniment a Datest et augual il a deilare rombie danna les primeres de Jean Marie Tierre berile les lites déclaration et presentation prisma de Sierre Marian, aga de soixante moltons, chef de lataillon un retraite de afin I knowe at Deini Marion, aga de wingt see Ine, agust desennences, les leux de Et ont les pere et témoins, signe avec nous le présent acte, après lecture

Contour detection for text line extraction











RÉPUBLIQUE Birth Certificates Transcription



Skew correction

Belfort et auguel il a declare vouloir donner to premary do



Resfort et auguel il a déclaré vouloir donnes le prénons de

Skewness correction example [1]













XML File Generator

```
<page Name="1E43_0003.jpg" Type="Double page" ImageHeight="4195" ImageWidth="5896">
                    <KeadingOrder>
                      <RegionRefIndexed Index="1" Type="Title-number"/>
                      <RegionRefIndexed Index="2" Type="Title-name"/>
                      <RegionRefIndexed Index="3" Type="Paragraph"/>
                                                                             Reading
                      <RegionRefIndexed Index="4" Type="Title-number"/>
                                                                               order
                      <RegionRefIndexed Index="5" Type="Title-name"/>
                      <RegionRefIndexed Index="6" Type="Paragraph"/>
                      <RegionRefIndexed Index="7" Type="Paragraph"/>
                     </ReadingOrder>
                   -<TextRegion Index="1" Type="Title-number">
                      <Coords Points="(388, 1311, 146, 141)"/>
    Record
                     -<TextEquiv>
   number
                        <Unicode>N°3</Unicode>
                      </TextEquiv>
                     </TextRegion>
                    <TextRegion Index="2" Type="Title-name">
                      <Coords Points="(111, 1437, 576, 282)"/>
                                                                     Record
                     -<TextEquiv>
                        <Unicode>Denise Brunschwick</Unicode>
                                                                      name
                      </TextEquiv>
                     </TextRegion>
                    <TextRegion Index="3" Type="Paragraph">
                      <Coords Points="(687, 1326, 2200, 1476)"/>
                     -<TextEquiv>
Paragraph
                       -<Unicode>
                          L'an mil huit cent vingt neuf, le trois du mois de Janvier, à onze heures du matin, pardevant
   level
                          nous a présenté un enfant du sexe féminin né hier matin à huit heures, de lui déclarant et d
                          David Levy, agé 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>
                                                                                   Text line
                      </TextLine>
                                                                                     level
                     -<TextLine ID="1">
                        <Coords Points="(687, 1501, 2200, 67)"/>
                       -<TextEquiv>
                         -<Unicode>
                            à onze heures du matin, pardevant nous François-Joseph
                           </Unicode>
                        </TextEquiv>
```

Example of a generated XML file













• Metrics and results:

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

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

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

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

C	Accuracy			
Component	IoU	\mathbf{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











RÉPUBLIQUE FRANÇAISE Death Certificates Transcription









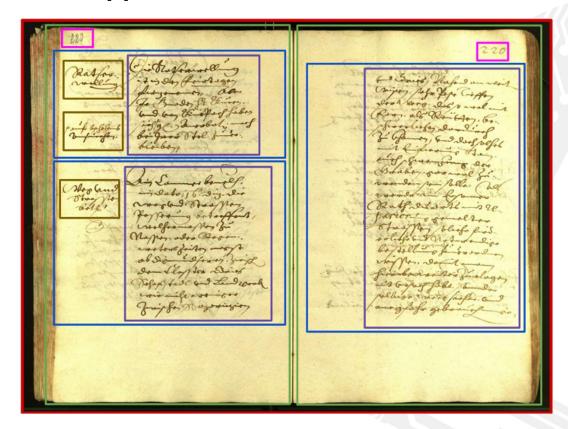








• The Document Attention Network [2]:



DAN segmentation results from READ 2016





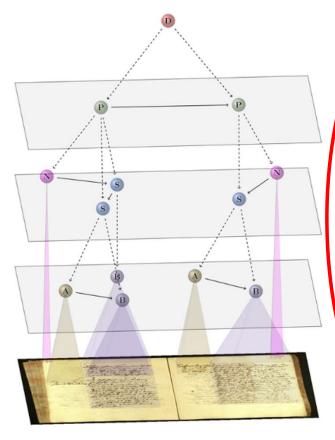




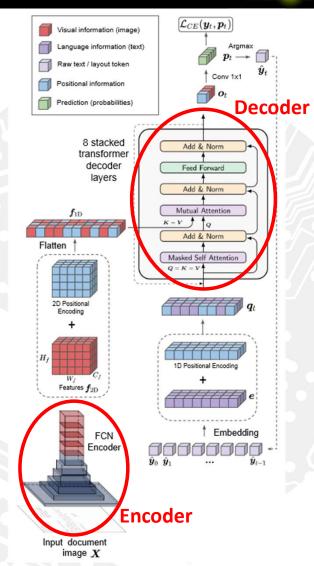




• The Document Attention Network [2]:



```
page>
  <page_number>
204
  </page_number>
  <section>
    <body>
      Schgrafer, [...] gehalt
    </body>
  </section>
  <section>
    <annotation>
      Genneral [...] Raitūng
    </annotation>
    <body>
      Aūf den: [...] werden,
    </body>
  </section>
</page>
<page>
  <page_number>
    204
  </page number>
  <section>
    <annotation>
      Schmalz. [...] bet:
    </annotation>
    <body>
      Verer [...] dara
     </body>
  </section>
</page>
 ocument>
                    Ground truth
```









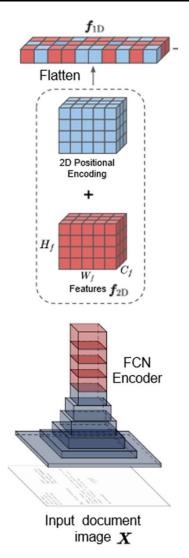




RÉPUBLIQUE FRANÇAISE Death Certificates Transcription



• The FCN Encoder:









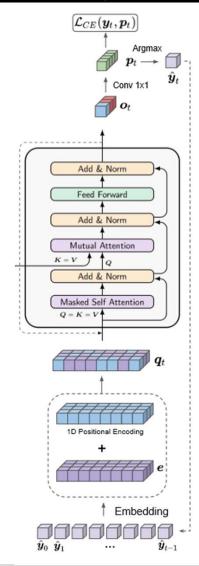




RÉPUBLIQUE FRANÇAISE Liberte française Protection Correction C



• The Transformer :



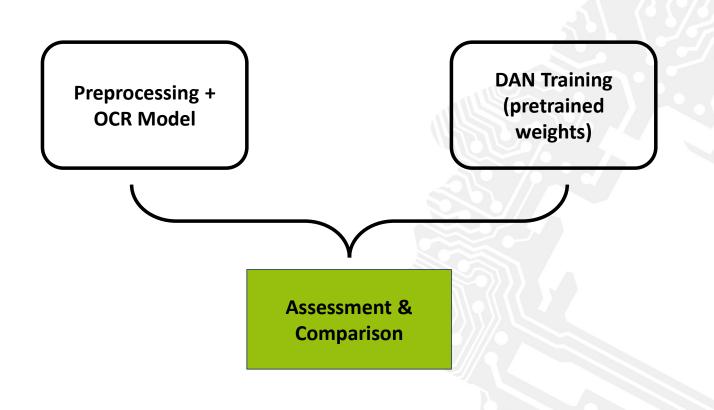






















• [1] Preprocessing and Data Structuring Technique for Belfort Birth Records Transcription, W AlKendi, Franck Gechter, Laurent Heyberger, Christophe Guyeux, 2024

• [2] DAN: a Segmentation-free Document Attention Network for Handwritten Document Recognition, D. Coquenet, C. Chatelain, T. Paquet, 2022









