



HAL
open science

Image Acquisition and Layout Analysis

Alix Chagué, Hippolyte Souvay

► **To cite this version:**

| Alix Chagué, Hippolyte Souvay. Image Acquisition and Layout Analysis. 2023. hal-04254223

HAL Id: hal-04254223

<https://inria.hal.science/hal-04254223v1>

Preprint submitted on 23 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



Image Acquisition and Layout Analysis

Alix Chagué¹, Hippolyte Souvay²
¹INRIA, ² UNIFR

From Source to Full Text: Workshop on Using Automatic Character Recognition (ATR)
September 8th 2023, DHIP



UNIVERSITÉ DE FRIBOURG
UNIVERSITÄT FREIBURG

Summary

1. Introduction: Important features of an image for HTR
2. Image processing for eScriptorium
3. Importing images to eScriptorium
4. Layout Analysis
5. Applying a segmentation model
6. Training
7. Bonus 1: Segmentation ontology
8. Bonus 2: Retraining a segmentation model



1. Introduction: Important features of an image for HTR

1.1. Image Definition

1.2. Image Resolution

1.3. Color

1.4. Take aways for HTR



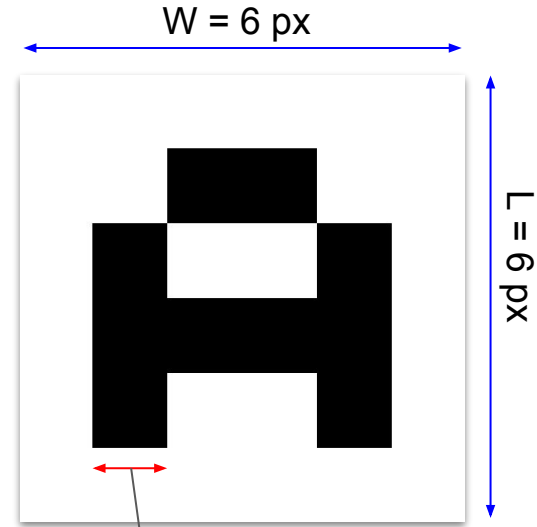
Features of an image 1/4

1.1. Definition

- Describes the amount of information in an image
- A digital image is made of pixels -> the definition says how many pixels
- Usually expressed as length by width, with pixels as a unit (L x W)

1.2. Resolution

- Transposition of the image size into real-life measures
- How many pixels to cover 1 inch -> DPI (dot per inch)



if resolution = 1 dpi
then 1px = 1 inch if printed



Features of an image 2/4

1.3. Color

0	0	0	0	0	0
0	0	1	1	0	0
0	1	0	0	1	0
0	1	1	1	1	0
0	1	0	0	1	0
0	0	0	0	0	0

binarized :

- 1 x L x W {0|1}
- 1 layer, poor information

255	255	255	255	255	255
255	255	0	0	255	255
255	0	255	255	90	255
255	0	130	90	90	255
255	0	255	255	90	255
255	255	255	255	255	255

grayscale :

- 1 x L x W {0...255}
- 1 layer, richer information



Features of an image 3/4

1.3. Color

255	255	255	255	255	255
255	255	0	0	255	255
255	0	255	255	90	255
255	0	130	90	90	255
255	0	255	255	90	255
255	255	255	255	255	255

RGB (Red, Green, Blue):

- 3 x L x W {0...255}
- 3 layers, rich information

RGBA (Red, Green, Blue, Alpha):

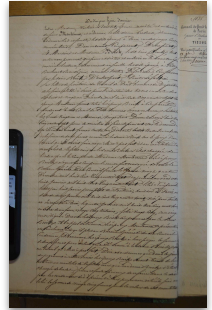
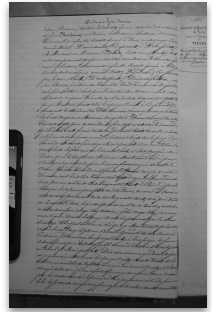
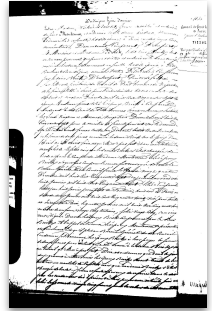
- 3 x L x W {0...255}
- 1 x L x W {0...100}
- 4 layers, rich information



Features of an image 4/4

1.4. Takeaways for images and HTR

- Resolution is an important parameter for readability by annotators and engines
- Color mode can also have an impact
- Poor images are not great (not enough information)
- Extra rich images are not great either (too heavy to process)
- Binarization is not necessary with Kraken
- PNG and JPEG are good formats
- TIF is not necessary



2. Image preprocessing for eScriptorium

2.1. Optimizing images for eScriptorium

2.2. Possible pre-processing



Image preprocessing for eScriptorium 1/2

2.1. Optimizing images for eScriptorium:

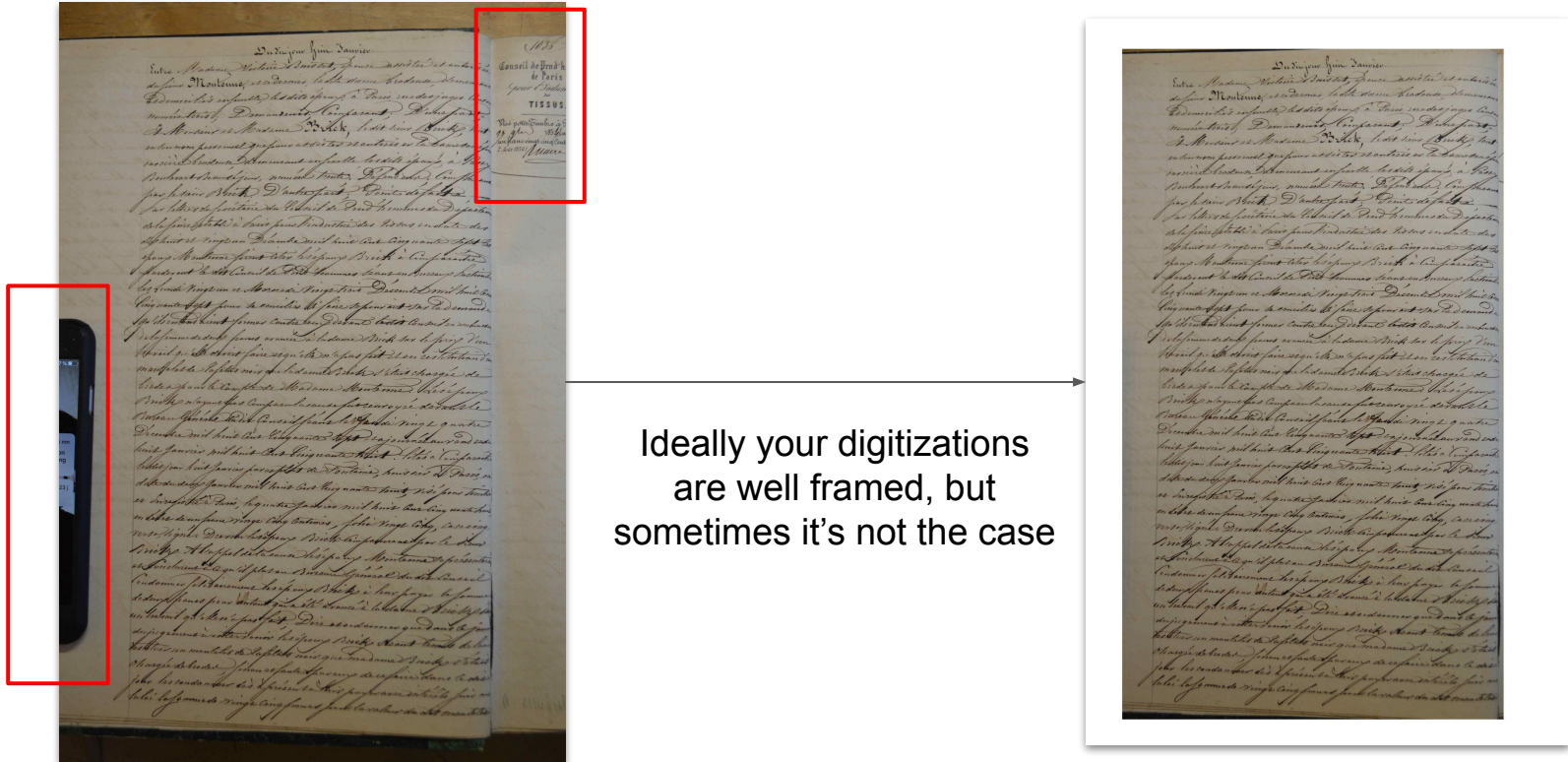
- Optional and can be time consuming (trade-off preprocessing/impact)
- But can reduce the time spent correcting the segmentation
- But can increase your chances to get a better transcription

2.2. Possible pre-processing:

- Adjust the image resolution
- Adjust the contrast
- Invert color (clear text on dark background)
- Crop the image to isolate the page and exclude the background
- Split double pages



Image preprocessing for eScriptorium 2/2



3. Import images to eScriptorium

You are now ready to import your images into eScriptorium.

To import image files, open your document and switch to the image dashboard. You can now either:

- Drag and drop the image files onto the rectangle with a blue border that is reserved for this purpose.
- Click on the same rectangle and browse the files on your computer.

We will not demonstrate it but: you can also import PDF files, or use the IIIF protocol if provided by the institution that hosts a document online.



3. Import images to eScriptorium

The screenshot displays the eScriptorium web application interface. At the top, the logo 'eScriptorium' is visible along with navigation links for 'Home' and 'Contact'. On the right side of the header, there are links for 'My Projects', 'My Models', and a user profile 'Hello Alix'. Below the header, a navigation bar contains tabs for 'Description', 'Ontology', 'Images', 'Edit', 'Models', and 'Reports'. The 'Images' tab is currently selected. The main content area features a large dashed blue box with the text 'Drop images here or click to upload.' centered inside. Below this area, there are several interactive buttons: 'Select all', 'Unselect all', and 'Selected 0/3' on the left; 'Import' and 'Export' in the center; and 'Train', 'Binarize', 'Segment', and 'Transcribe' on the right. At the bottom, three image preview windows are shown, each containing a sample of handwritten text. The first window is labeled '1' and has a zoom level of 100%. The second window is labeled '2' and has a zoom level of 100%. The third window is labeled '3' and has a zoom level of 100%. A red rectangle highlights the 'Images' tab and the upload area. A yellow circle is present in the bottom right corner of the slide.

4. Layout Analysis

Zone and Lines for Layout Analysis:

Layout analysis is rendered with 2 mains elements:

- Zone: a block (empty or not)
- Line: a line of text (not a sentence)
 - polygon (mask)
 - baseline

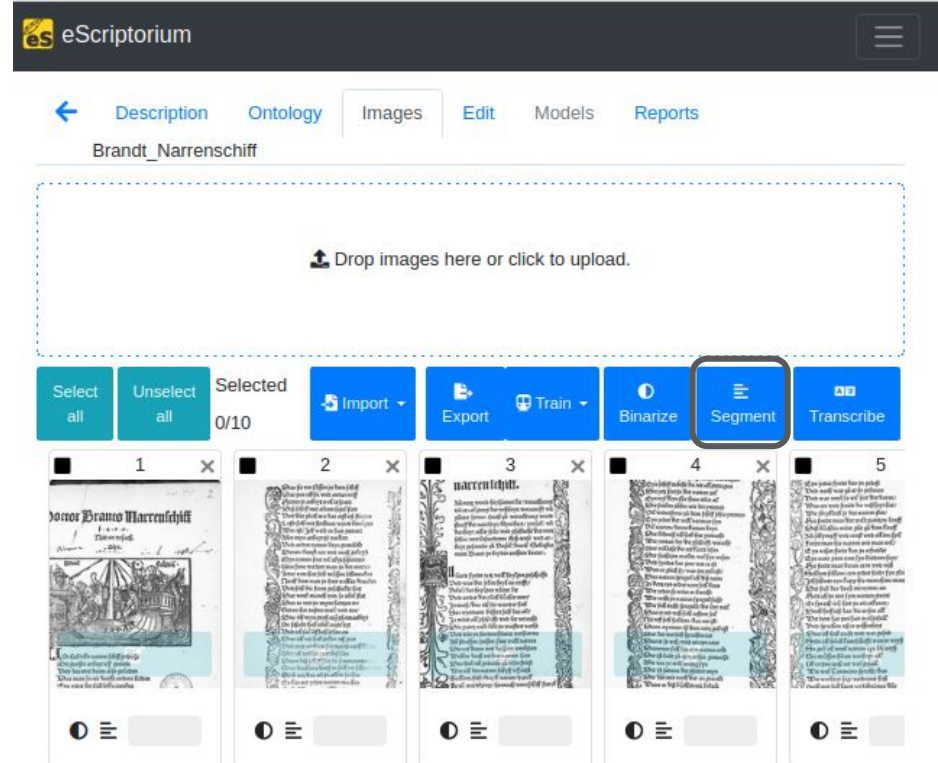
Besides describing the layout, zones helps have a logical reading order.



5. Applying a Segmentation model

Go to your documents **Image** Dashboard

- Select the images you want to segment.
- Click on the “Segment” button.
- You have to choose:
 - The segmentation model you want to use.
 - The features you want to recognise.
 - The orientation of your text.



Foreword on the segmentation interface

- Accessible in “Edition” page
- Activate the “Segmentation Panel”
- It’s easy to get lost, so take your time and pay attention to where you click/zoom
- **Escape** to stop drawing
- **Control + Z** to cancel an action
- Reset zoom with **Control + Backspace** (or use the button on the left)
- **Right click + hold** on the image to move once zoomed in!
- Many possible shortcuts, shown when hovering over an image or when you click on “?”

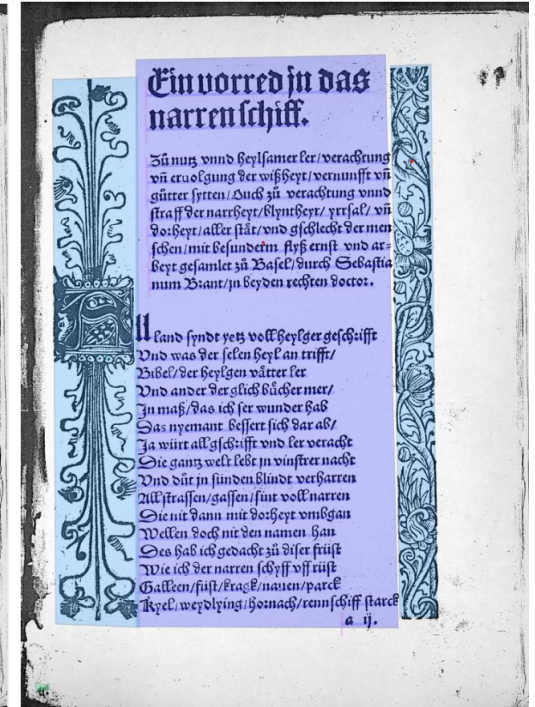
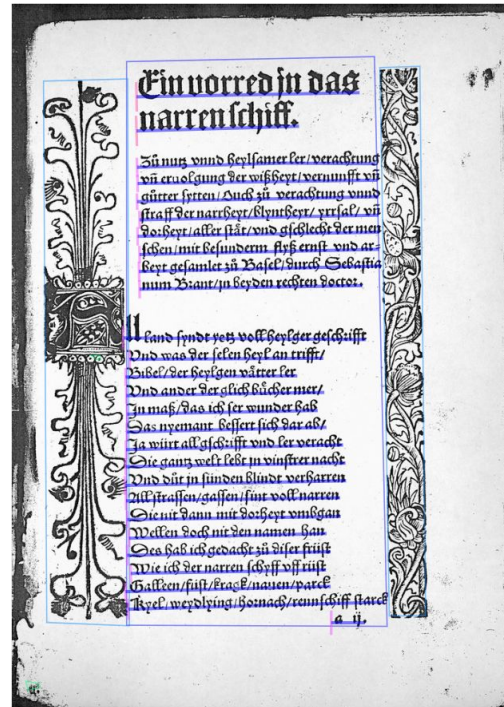


Manual correction of the segmentation output 1/2

Go to your Image Dashboard:

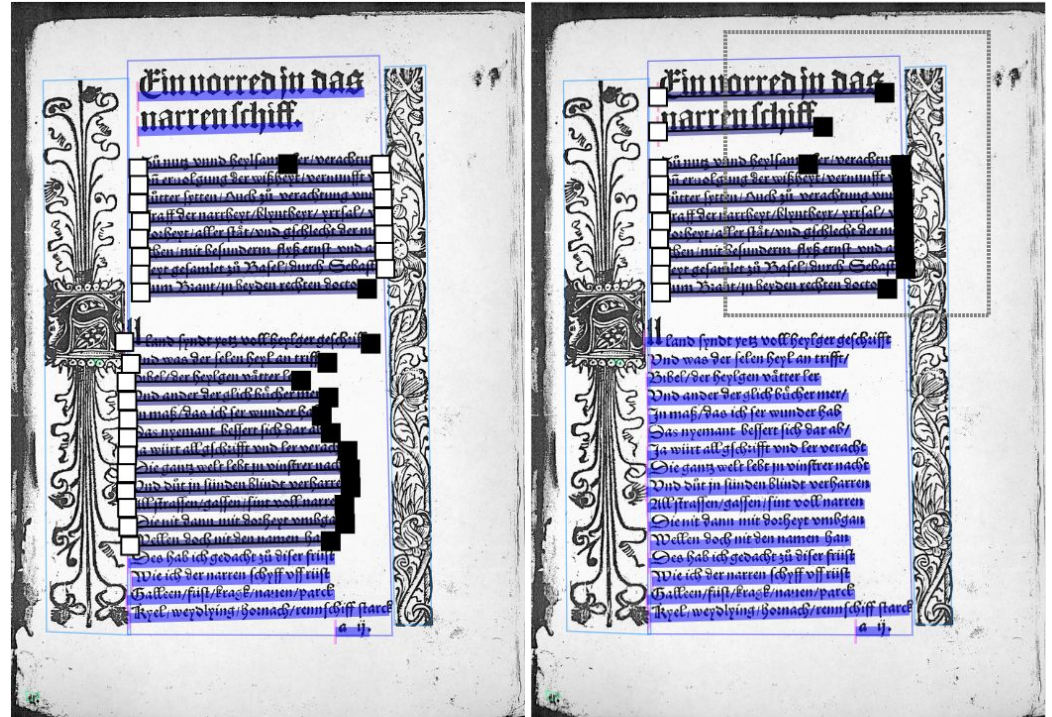
- Click the **Switch to Region** button (or press the **R** key).

Each mode allows you to work on its features (lines or regions).



Manual correction of the segmentation output 2/2

- To select a single line or region, **left-click** on it.
- To select multiple lines or regions:
 - Hold **Shift** and **left-click** each feature you want so select, or:
 - Hold **Shift** and drag **left-click** to select all features in the selected zone.



Exercise : The *Ship of Fools*, by Sebastian Brandt

- Identification on the server: <https://escriptorium.inria.fr>
 - Login : apinche_formation - password : training1234
- Open the document Brandt_Narrenschiff.
- Apply the default (blla.mlmodel) segmentation model and the incunable_seg_retrain model.
- Establish your segmentation standards to create a training corpus.
- Do the required corrections in order to obtain a clean training corpus.
- Optional: apply SegmOnto typing



7. Controlled vocabulary to describe zones and lines

To improve the reusability of training data and the collaboration between research projects using HTR, the community has developed an **ontology** for describing the layout: [Segmonto](https://segmonto.github.io/) (<https://segmonto.github.io/>).

Examples of types available in SegmOnto:

- “MainZone”, “TitlePageZone”, “GraphicZone”
- “DefaultLine”, “HeadingLine”

These types can be associated with logical elements (zones/lines)



8. Retraining of a segmentation model

You can retrain a segmentation model to increase its efficiency on the layout of your document:

- Select the already segmented and corrected images on which you want to retrain your model.
- Click on the **Train** button and select **Segmenter** from the drop-down menu.
- In the **New Model** box, chose a name for your your retrained model.
- In the **Or select an existing one** box, chose the model you want to retrain.

You can also decide to simply overwrite your model, but beware, we recommend that you export and save your model to a safe place first.



Evaluating a segmentation model

- To date, no application has been developed to evaluate segmentation models.
- Do it and try again, see for yourself if the results look better than before.



References

- Simon Gabay, Ariane Pinche, Kelly Christensen, Jean-Baptiste Camps, Nicola Carboni, *SegmOnto, A Controlled Vocabulary to Describe the Layout of Pages*, version 0.9, Genève/Lyon/Paris, 2023, <https://segmonto.github.io/>

