Research and Development Project

Detection and segmentation of mathematical expressions in Leibniz manuscripts

Team IPI

Authors: Mamisoa RANDRIANARIMANANA
& Xianxiang ZHANG

Supervisors: Yejing XIE & Harold MOUCHERE

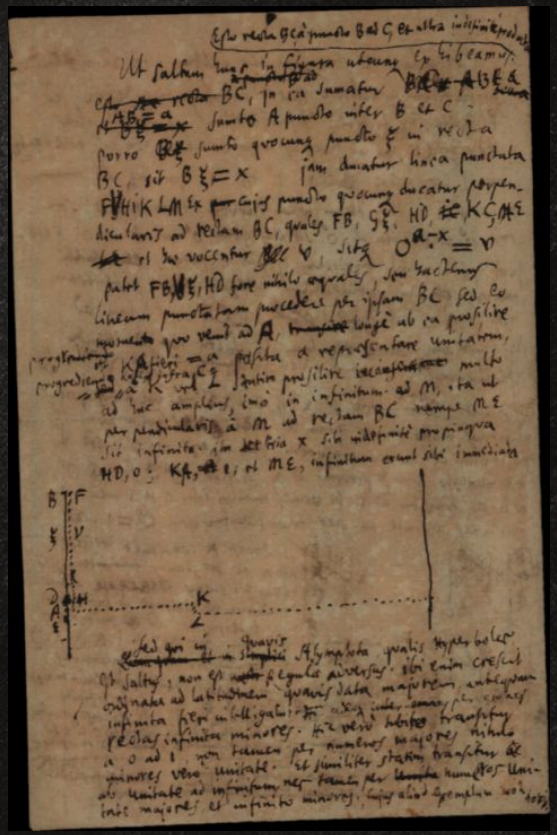
Coordinator: Philippe LERAY

Plan

Gottfried Wilhelm Leibniz (1646 - 1716)

- Philosopher and mathematician
- The last universal genius.
- Inventor of infinitesimal calculus.

Gottfried Wilhelm von Leibniz
Portrait by Andreas Scheits
Introduction aux suppléments Leibniz – L'Économie



Leibniz's manuscripts

- Treasures of historical knowledge
- gateways in the mind of a polymath.
- Bring together innovations in mathematics and philosophy.

Need

To transcribe and edit Leibniz's manuscripts

Constraints

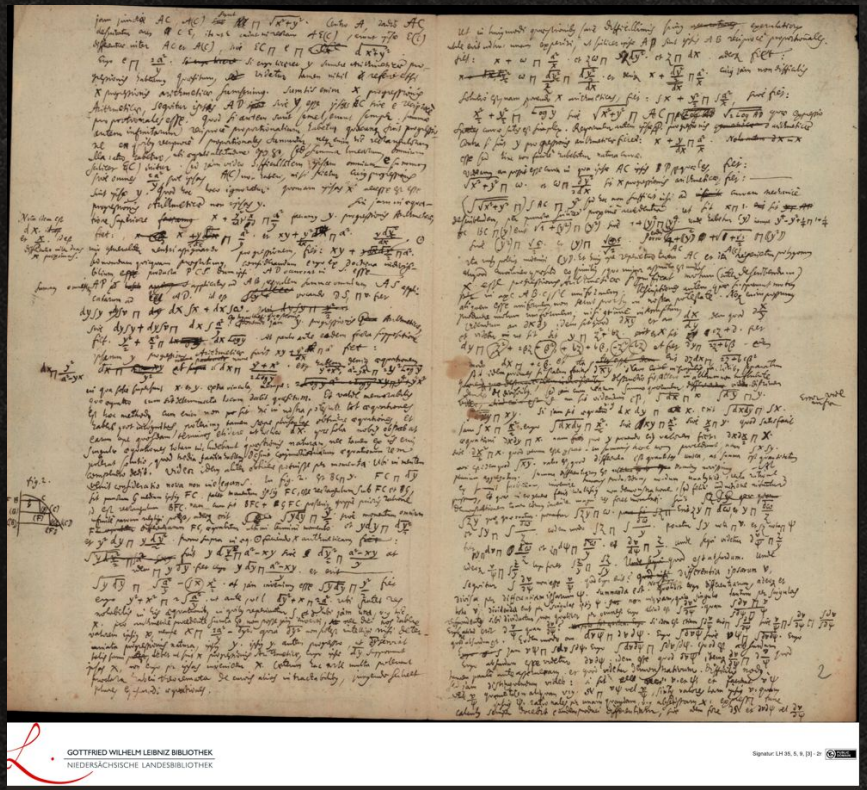
- Complex, laborious, and time-consuming manual transcriptions.
- Deciphering the content requires expertise in the field
- Inability of the currently used tools to transcribe and edit mathematical expressions and formulas

Problematic :

How to automatically detect and segment the mathematical expressions in Leibniz manuscripts?

The data at our disposal

- 10 images of Leibniz original manuscripts
- 2 sections left and right per page



Characteristics of the data

- Variable distribution

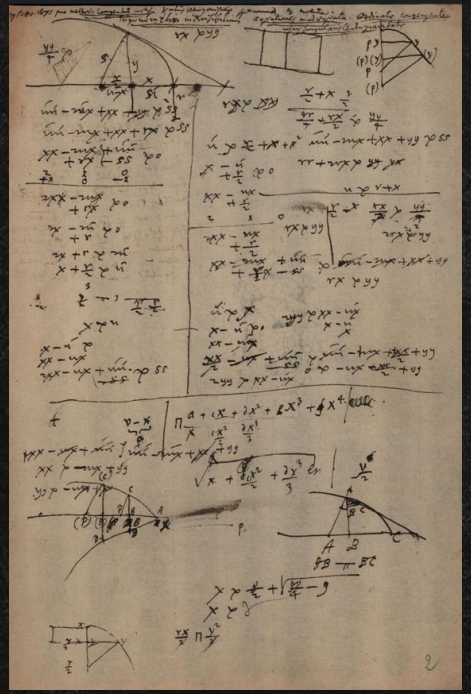


Fig : an image predominantly consisting of equations

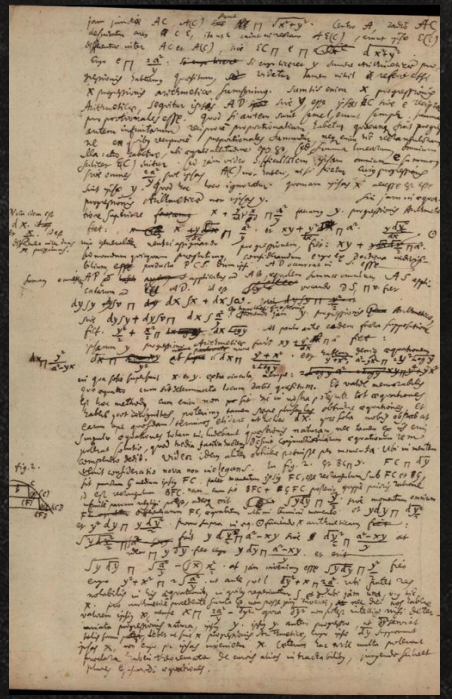


Fig : an image with both but more texts

Characteristics of the data

- Heterogeneous in content density

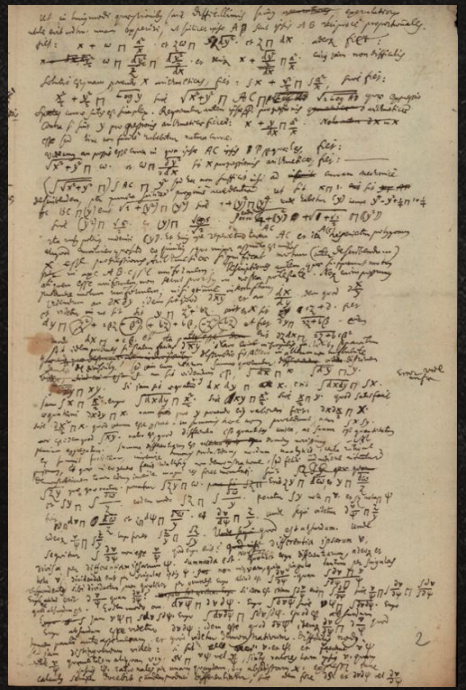


Fig : a page totally full

vs

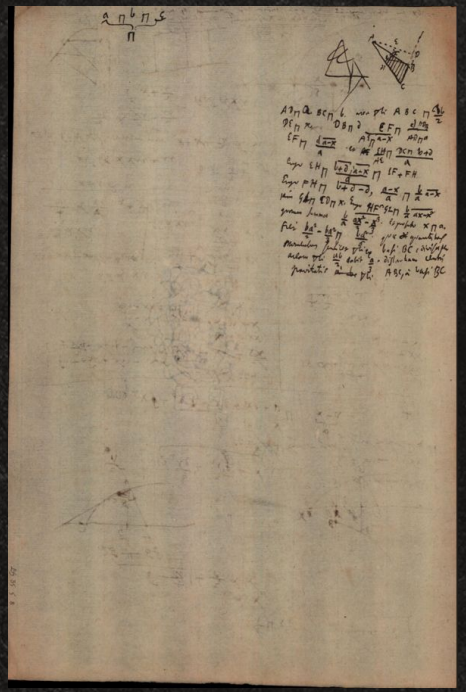


Fig : a page less than one-quarter full

Objectives Pursued on this project

- State-of-the-Art analysis emphasizing mathematical challenges in ancient manuscripts and data augmentation
- Data augmentation
- Development of detection and segmentation Techniques for mathematical expressions in Leibniz's Manuscripts
- Experimentation and evaluation

The objectives

Data Augmentation	Detection and Segmentation
❖ enough images for the data set	❖ identify the equation
❖ class balance	❖ identify the text
❖ inter-class and intra-class variability	❖ mark the area
❖ data consistency	❖ high flexibility
❖ low complexity and low time cost	❖ low time cost

State of art

Approaches for detection and segmentation

Approaches

- Based on Hidden Markov Model (HMM)
- Based on the Transformer
- Based on probabilistic indexing
- Based on convolutional neural networks(CNN)

Analysis

	CNN	HMM	Transformer	Probabilistic indexing
Flexibility	+++	+++	++	++
Simplicity	++	+	+	+
Efficiency	+++	++	+++	++
Adaptability	+++	++	+++	+++
Low time cost	++	+++	+	+
Low data	+++	+	+	++

Approaches for data augmentation

Approaches

- Data Warping
- Oversampling
- Copy Paste Data

Analysis

	Data Warping	Oversampling	CopyPaste
Structural variety	+++	++	+
Spatial variety	+++	+	+++
Style variety	+	+++	+
New content	+	+++	+
Inter and intra-class variety	+	+++	+++
Low Data requirement	+++	+	++
Low Time cost	+++	+	++
Low Complexity	+++	+	+++

Evaluation Approaches

Approaches

- Intersection Over Union
- Evaluation pixel by pixel
- Evaluation by Human Perception.

Propositions

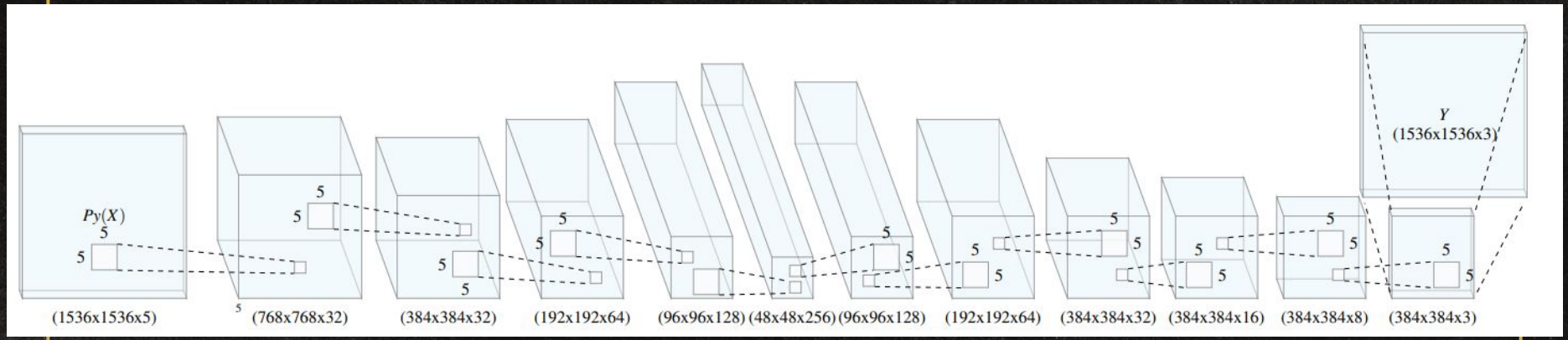
Propositions

Data Augmentation	Copy Past Augmentation
Detection and Segmentation	CNN

Data augmentation

1. Annotation
2. Reading Dataset and collecting Annotation Elements: Texte or Equation
3. Randomly select a specified number of text and equation elements
4. Placing Elements onto a background image without overlap + adjusting positions to avoid collision and maintaining within image bounds.
5. Updating the annotation: revising coordinates and other relevant metadata in the annotation structure
6. Generating and save annotated Images

Detection and Segmentation: FCNN

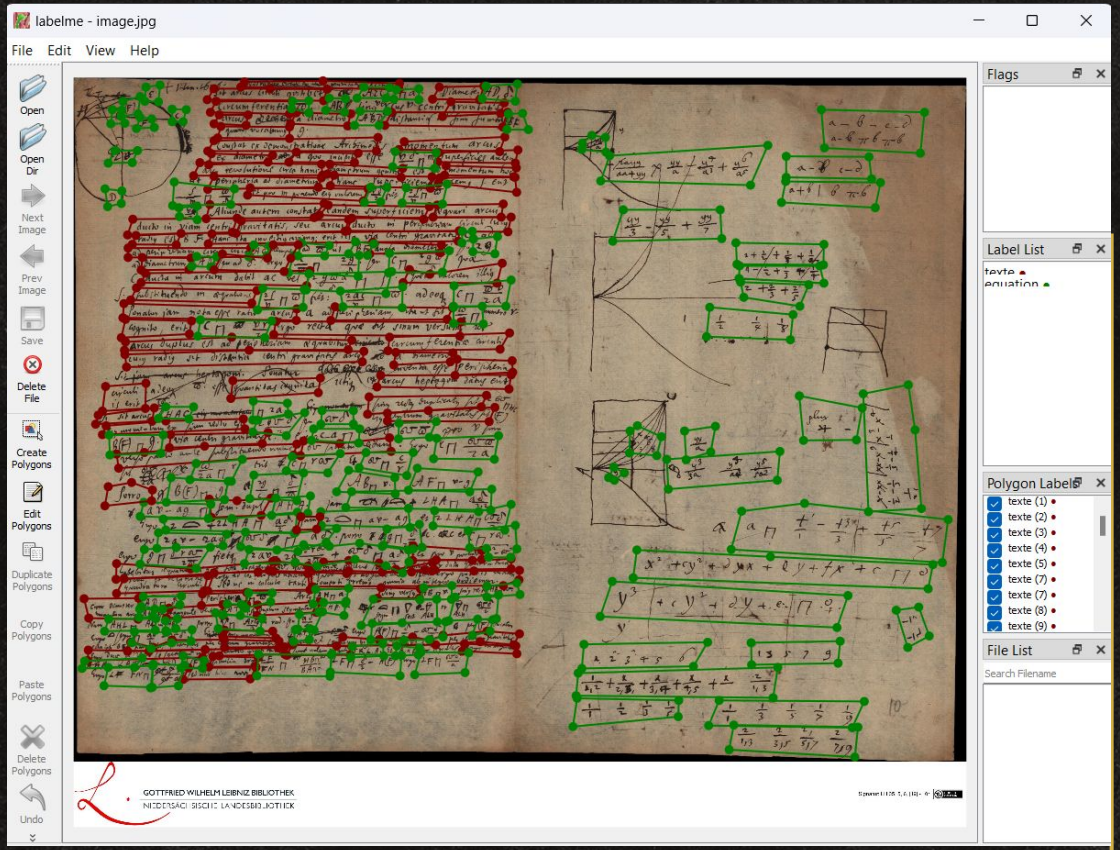


Source : Roman-Jimenez G, Viard-Gaudin C, Granet A, et al. Transfer Learning for Structures Spotting in Unlabeled Handwritten Documents using Randomly Generated Documents[C]//International Conference on Pattern Recognition Applications and Methods. 2018.

Experimentation

Data Annotation

- LabelMe as Tool
- Two labels:
  - Equation
  - Texte
- Unannotated ambiguous situation



Data Annotation

- Variable distribution per image
- Means of the proportion of texts considering only the black pixels : 55%
- Means of the proportion of equations considering only the black pixels: 45%

Image	Texte	Equation
image1_g.jpg	0.3099	0.6901
image1_d.jpg	0.7948	0.2052
image2_g.jpg	0.4764	0.5236
image2_d.jpg	0.6792	0.3208
image3_g.jpg	0.9818	0.0182
image3_d.jpg	0.1632	0.8368
image4_g.jpg	0.2791	0.7209
image4_d.jpg	0.6929	0.3071
image5_g.jpg	0.7159	0.2841
image5_d.jpg	0.0000	1.0000
image6_g.jpg	0.1300	0.8700
image6_d.jpg	0.7200	0.2800
image7_g.jpg	0.5488	0.4512
image7_d.jpg	0.6659	0.3341
image8_g.jpg	0.4223	0.5777
image8_d.jpg	0.2100	0.79
image9_g.jpg	0.8874	0.1126
image9_d.jpg	0.8184	0.1816
image10_g.jpg	0.7844	0.2156
image10_d.jpg	0.710	0.390
Moyennes	**0.55**	**0.45**

Fig : Pixel-based quantification of texts and equations in manuscripts without background.

Data augmentation and setting up database

DataSet	Original images	Details
Training DataSet	image1_left, image1_right image2_left, image2_right Image3_left, image3_right	300 images with balanced equations and texts proportions

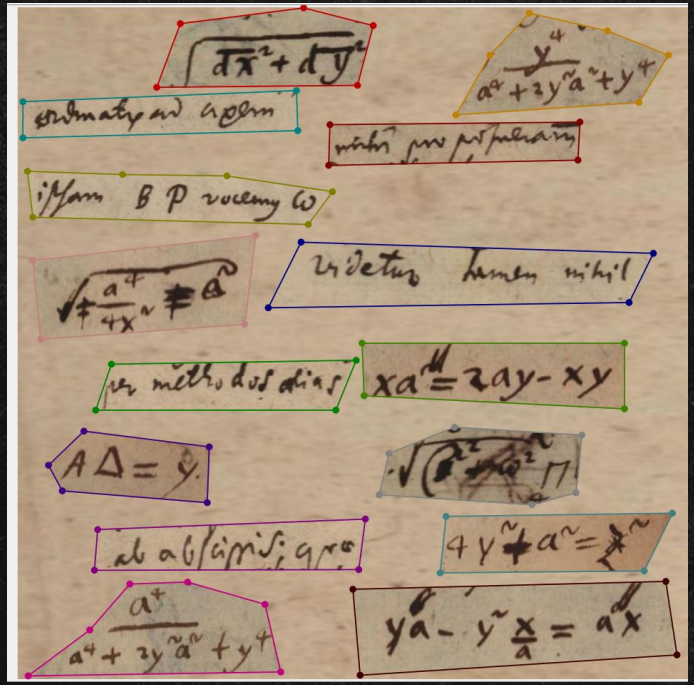


Fig : Example of training image

Data augmentation and setting up database

DataSet	Original images	Details
Training DataSet	image1_left, image1_right image2_left, image2_right image3_left, image3_right	300 augmented images with 50% equations and 50 % text 0 original data
Validation DataSet	image4_left, image4_right image5_left, image5_right image6_left, image6_right	300 augmented images with 50% equations and 50 % text 0 original data
Test DataSet	image7_left, image7_right image8_left, image8_right image9_left, image9_right image10	70 augmented images with various proportion of equations and text 3 original data

Training

➢ Use of pre-trained model

Source : Roman-Jimenez G, Viard-Gaudin C, Granet A, et al. Transfer Learning for Structures Spotting in Unlabeled Handwritten Documents using Randomly Generated Documents[C]//International Conference on Pattern Recognition Applications and Methods. 2018.
https://github.com/GeoTrouvetout/CIRESFI

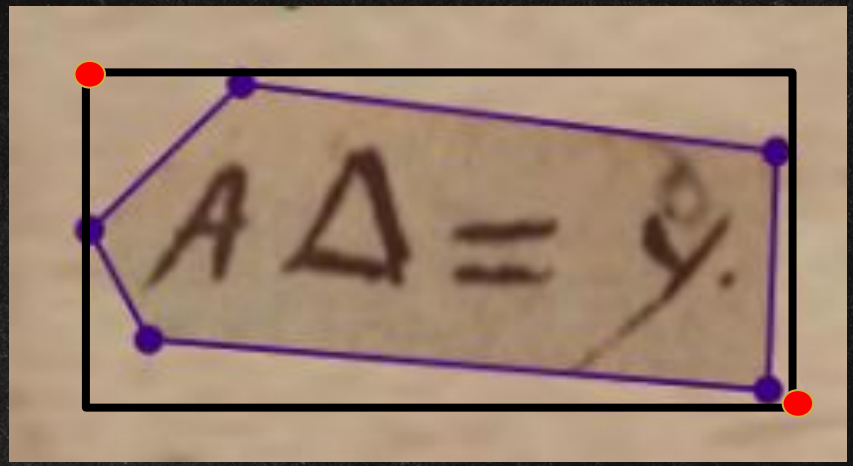
➢ Adjusting settings

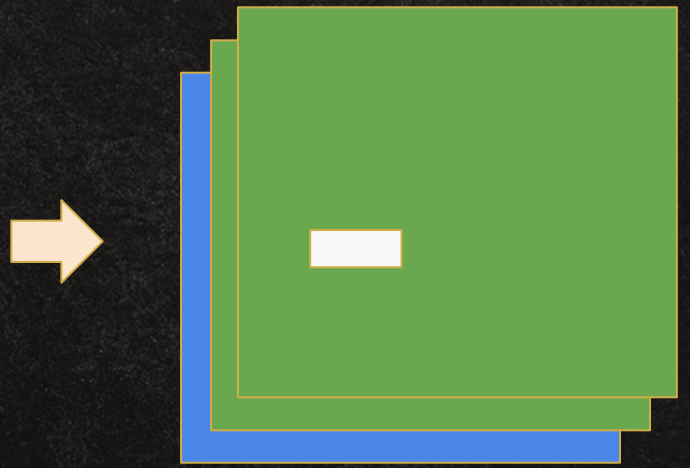
Background
Digit
Number
Character
Word

Background
Equation
Equation
Text
Text

Training

Input image and csv

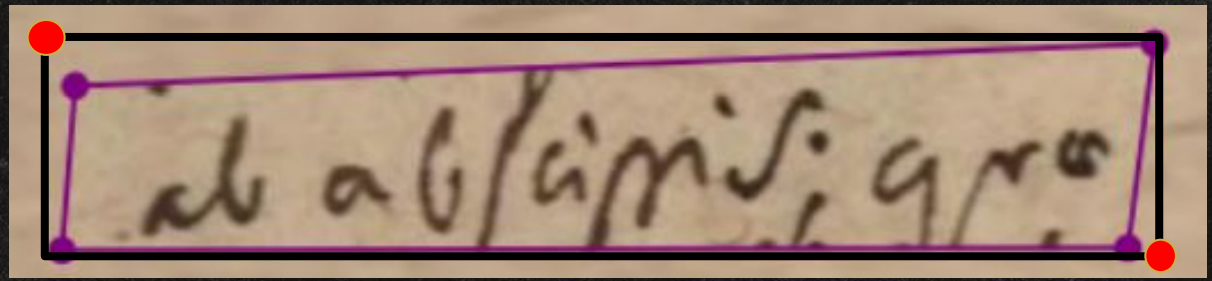


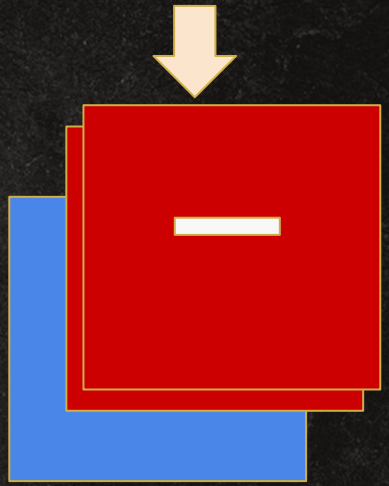


mask_structures[0, min_y:max_y, min_x:max_x] = 0
mask_structures[1, min_y:max_y, min_x:max_x] = 1
mask_structures[2, min_y:max_y, min_x:max_x] = 1

Training

Input images and csv

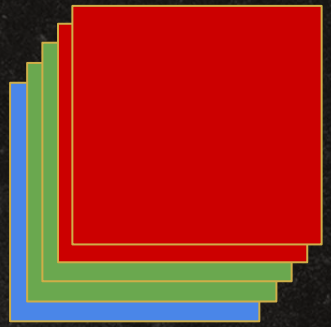




```
mask_structures[0, min_y:max_y, min_x:max_x] = 0
mask_structures[3, min_y:max_y, min_x:max_x] = 1
mask_structures[4, min_y:max_y, min_x:max_x] = 1
```

Training

Input images and csv

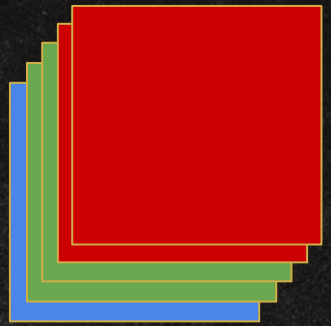


Probability structures:

Count the number of pixels of each type
Calculate pixel ratio for each type
Calculate the probability of each type

Training

Loss function



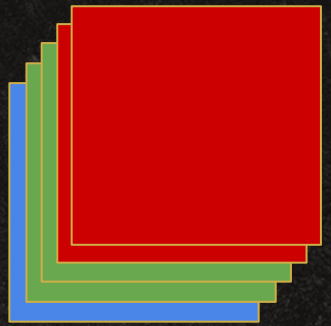
Cross Entropy (CE) loss

$$C(Y,S) = -\sum_{i}^{N}\sum_{k=0}^{2}\frac{1}{p_i^k}(s_i^k \cdot log(y_i^k)),$$

S : structure map
Y : the map to be built
p : The probability that a pixel belongs to type k,
k = 0, 1, 2

Training

Optimizer

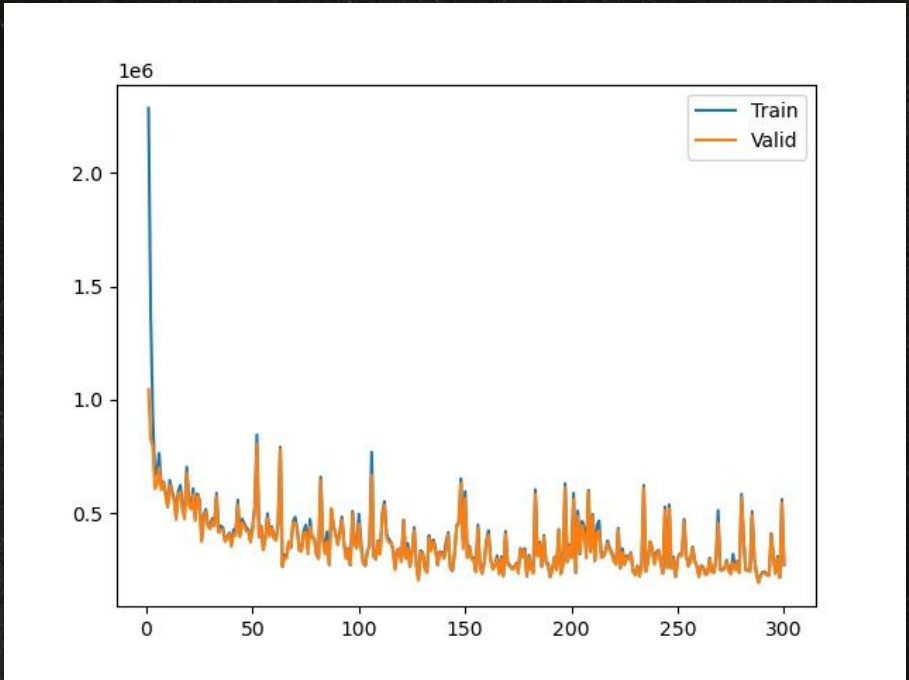


Adam optimization algorithm
(Adaptive Moment Estimation)

Often used in gradient descent in deep learning

Used to minimize the loss function

Result



Epoches : 300

The training loss dropped significantly and stabilized.

Image Processing

Output for model trained by 300 images without pre-trained model

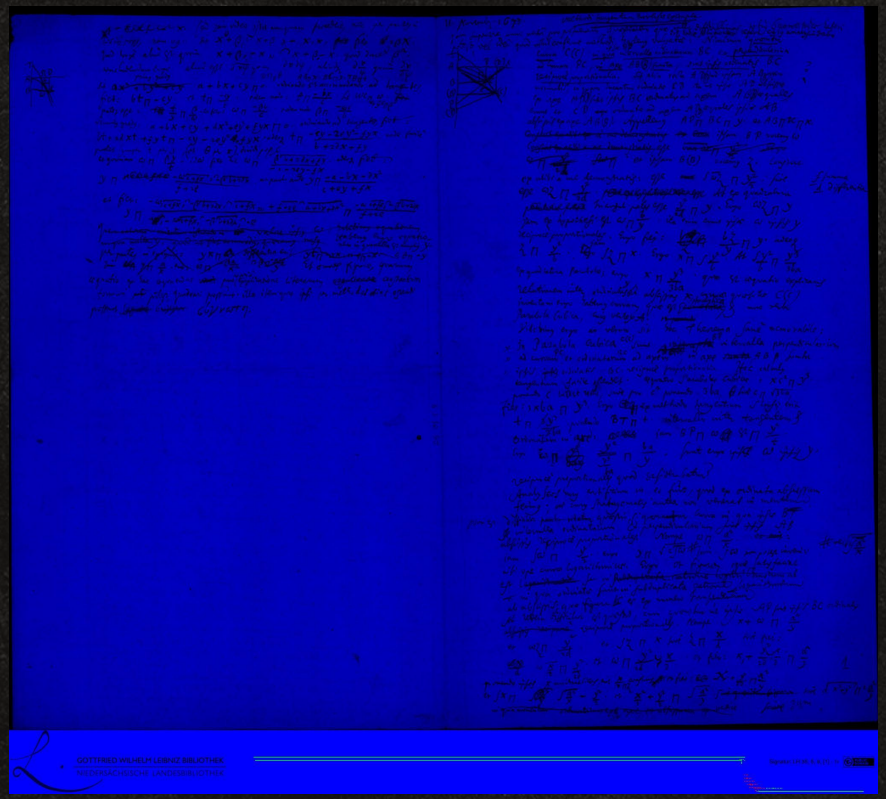


Image Processing

Output for model using pre-trained model and 300 images as training dataset on a generated image

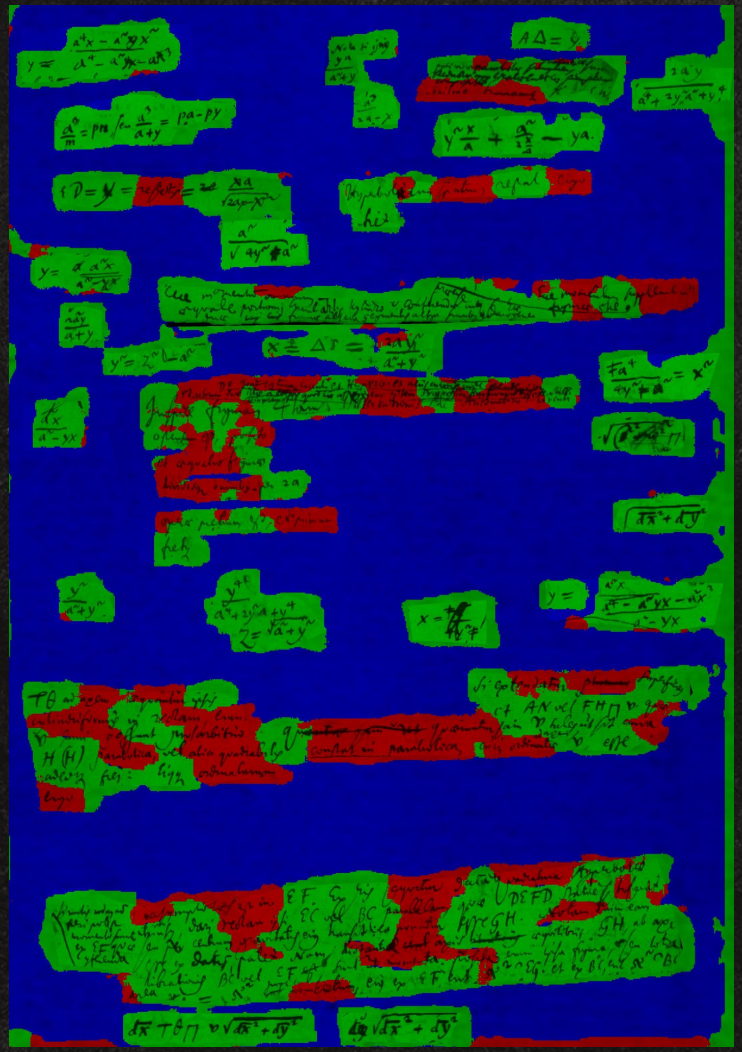
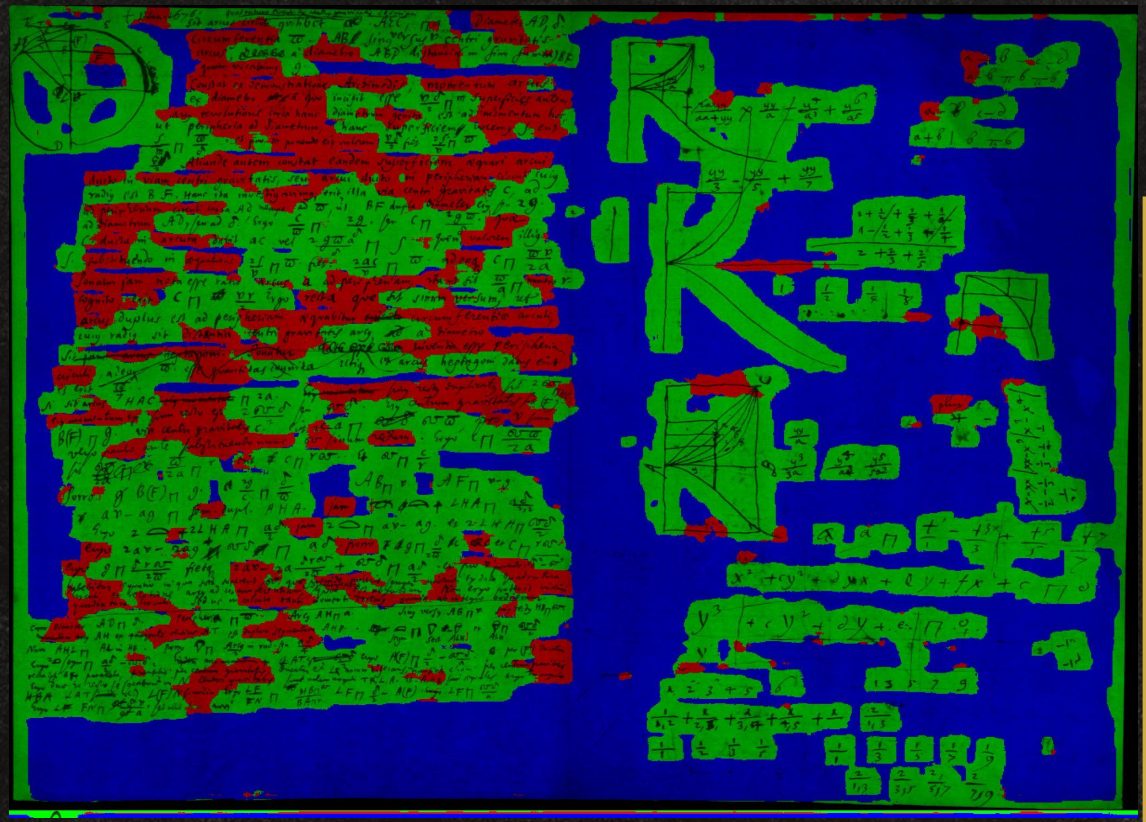


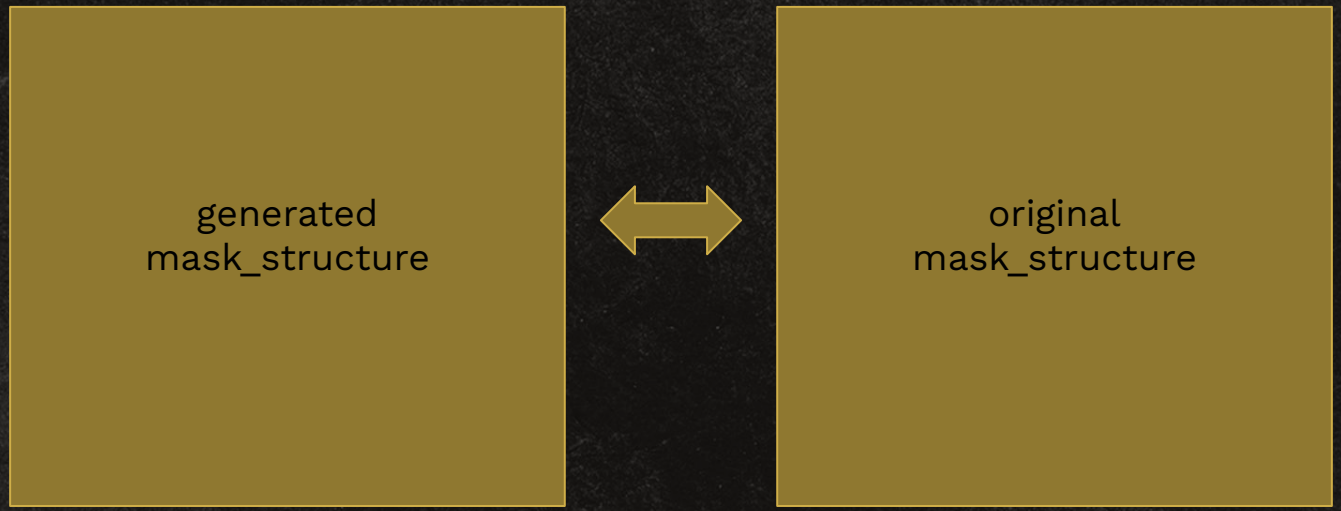
Image Processing

Output for model using pre-trained model and 300 images as training dataset on a real image of Leibniz Manuscript



Evaluation

Comparison of The mask_structure generated by the model prediction and the mask_structure created by hand



Evaluation

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

=> the proportion of actual positives correctly identified

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

=> the proportion of actual negatives correctly identified

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}$$

=> the proportion of all true results in the data

$$\text{MCC} = \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}}$$

=> correlation coefficient between the observed and predicted binary classifications

Evaluation

Image	Evaluation
Generated images	Sensitivity: 0.56 Specificity: 0.78 Accuracy: 0.71 MCC: 0.34
Image of real manuscript of Leibniz	Sensitivity: 0.43 Specificity: 0.71 Accuracy: 0.61 MCC: 0.14

Conclusion of the experimentation

- MCC > 0 : predictions of the model are correlated with the actual data.
- Better result for the generated image than the real images
- Reliability in identifying negatives (specificity) but struggles more with identifying positives accurately in real manuscript images.

Project Management

❖ Gantt Diagram

Activity	Temporality																				
	S43	S44	S45	S46	S47	S48	S49	S50	S51	S52	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11
Subject Selection																					
Information Gathering and Reformulation																					
Literature Review																					
Analysis and Comparison of Methods																					
Presentation and Deliverable 1																					
Detailed Design																					
Validation of Choice																					
Data Annotation																					
Data Preparation																					
Implementation of Data Augmentation Method																					
Data Augmentation																					
Database Development																					
Correction and Improvement of Data Augmentation Method																					
Implementation of the Solution for Detection and Segmentation																					
Model Training																					
Correction and Improvement of the Solution																					
Testing																					
Implementation of the Evaluation Method																					
Evaluation																					
Correction and Improvement of the Evaluation Method																					
Results Analysis																					
Project Finalization																					
Documentation																					
Final presentation and Final Deliverable																					
Acceptance																					
Time spent on the project Mamisoa	06:30	04:00	14:30	12:30	18:00	15:00	15:00	15:00	15:00	05:00	15:00	13:00	12:00	14:00	06:00	10:00	20:00	17:00			
Time spent on the project Xianxiang	06:30	10:00	11:20	11:00	18:30	15:00	13:30	15:00	12:30	12:30	13:30	12:30	15:30	10:30	15:30	15:00	20:00	17:00			

Effective hours:
MR: **212 h**
XZ: **230 h**

MR & XZ Mamisoa (MR) Xianxiang (XZ) Not done

Conclusion

Conclusion

- **Project Overview: Work achieved and contributions**
  - **Immersive Initial Phase: Understanding Leibniz Manuscripts and the approach to resolve the problematic**
  - **Pioneering Solutions: Leveraging Convolutional Neural Networks for Enhanced Detection**
  - **Strategic Data Augmentation**
  - **Research Impact: Elevating Manuscript Analysis with AI-Driven Insights**

Conclusion

- **Research directions**
  - **Exploring Larger Datasets: Enhancing Model Precision**
  - **Focus on Ink Pixels: Refining Data Analysis Clarity**
  - **Advanced Learning Methods: Expanding Leibniz Manuscripts Study to recognition, transcription and edition**

Thank you for your attention