

## B. Tech. Project **Handwritten Formulae Detection**

Shashwat Kathuria - B17CS050

Satya Prakash Sharma - B17CS048

19th December 2020

### **Outline**

- I. Aim of Project
- II. Implementation Steps
- III. Explanation of Model Implemented
  - A. Dataset (Normal & Synthetic) & Preprocessing
  - B. Training & Testing
  - C. Model
  - D. Training Strategy & Loss Function
  - E. Results
- IV. Conclusion
- V. Sample Outputs
- VI. References

### **Aim of Project**

- → **Detect various types of formulas** which can be present in images taken from PDF documents, like research papers.
- → Various challenges:
  - formulas can be of various types
  - can have different types of symbols inside them, with some having superscript and subscript
  - some having very different forms like some present inside a table, some written inside matrices, etc.
- → We aim to build a model to predict such formulas in PDFs and along with analyzing the working and results obtained.

Recall from the previous section that representation types play a special role in abstracting from the heterogeneity of representations. In line with [BW90a], we presume that these types actually have the form of a many sorted algebra. Formally, if  $r \in \mathbb{RP}$  is a representation type, then  $\Sigma_r$  is presumed to be the signature of the many sorted algebra that is associated to r. Usually, the signature of r will be of the form

#### $C_n = (S; f_1, f_2, ...)$

where S is the carrier set—the set of values / types that are already known, e.g. primitives in the JAVA programming language—and  $f_1, f_2, \ldots$  are functions. The domain of these functions correspond to tuples with elements from S, and the range corresponds to elements from S [BW00a]. Consider for example the case where r is the type ASCII. For this type, the carrier set has two elements, N (all natural numbers) and Caar (all available characters available in the character set). Furthermore, the signature holds two functions, Char: N — Char (takes a mumber n  $\in$  N as parameter and returns the nth character from an ASCII document), and Len: — N (returns the length of an ASCII document). In summar, the signature for ASCII is:

 $E_{ASCII} = \{\{N, Char\}; Char : N \rightarrow Char, Len : \rightarrow N\}$ 

In the case of dynamic resources, the state of the resource needs to be added to the signature of the algebra. As an example of a dynamic resource, let us consider a weather forecasting application. Let Sigmar<sub>C</sub> represent the signature corresponding to the weather forecasting applications FC. Let Lee be some domain of locations on earth (for instance GPS coordinates) and let FCState represent the state of the application. The signature for this application could then be:

 $\Sigma_{FC} = \{\{FCState, Loc, \mathbb{N}, ASCII\}, TodaysForecast : FCState \times Loc \times \mathbb{N} \rightarrow ASCII\}$ 

Note that setting the actual weather parameters, such as air pressure, temperatures, wind speed, etc. by means of which the application may compute the weather forecast are left out of this signature since this signature focuses solely on the information supply perspective.

Summary of the elementary concepts:

 $[TS, RP, FE, TP, \Sigma, HasType, Service, Representation, \sim, SubOf]$ 

23

### **Implementation Steps**

## Study about RCNN Model and Page Object Detection

Study about RCNN
Model and ICDAR 2017
Page Object Detection
Competition.

## Computerized Formulae Detection Model

Implement an RCNN model to detect computerized formulae in the ICDAR 2017 dataset.

### Study about Generation of Synthetic Datasets

Research about GANs and other possible techniques to generate a handwritten dataset with annotations.

### Generate Handwritten Formulae Dataset

Generate dataset with different types of handwritings and different types of alignments of formulae.

#### Handwritten Formulae Detection Model

Fine tune existing RCNN model to detect handwritten formulae using the generated dataset.

### **Dataset & Preprocessing**

#### I. Computerized Text Formulae

- → The ICDAR 2017 dataset contains of **1600** images having tables, formulas and figures annotations.
- → Out of 1600 images, about **910 images contain** formulas
- → Parse the coordinates of the formula regions by parsing the ground truth xml file.

As the capital market is supposed to be perfectly competitive, the rental rate of capital,  $r_t$ , adjusts in each period such as to clear the market:  $k_t = k_t^s = k_t^s = k_t^s$ . On the labour market, however, the fixed wage contract does usually not allow to clear the market.

Wage rigidities introduce frictions on the labour market and force market participations to carry out transactions off their optimal supply and demand schedule. Following the spirit of the search-and-matching literature?, we assume that these rigidities imply that actual employment (i.e. transactions) corresponds to a weighted average between labour supply and demand at the current wage:

$$n_t = \omega n_t^d + (1 - \omega) n_t^s$$

where  $\omega$  measures the degree to which employment is determined by labour demand and will play a key role in the interpretation of the model and its results. This equation indicates that actual employment can be a result of a matching process whereby not all desired transactions are carried out, but where - due to the probabilistic nature of the process - firms may end up hiring more than what their current needs are. This may also happen when, for instance, employment is negotiated, when firms hoard labour in downturns, employing more than the profit-maximizing level of workers or when some other real rigidities are present. For the moment, we leave the interpretation open and only notice that observed employment may not necessarily correspond to desired levels. We give a more detailed interpretation of the  $\omega$ -parameter as soon as we have estimated the model.

**Product markets.** The final good is produced by combining intermediate goods. This process is described by the following CES function:

$$Y_t = \left(\int_0^1 y_{it}^{\rho} di\right)^{\frac{1}{\rho}}$$
(2)

where  $\rho \in (-\infty, 1)$ .  $\rho$  determines the elasticity of substitution between the various inputs. The producers in this sector are assumed to behave competitively and to determine their

demand for each good,  $y_{ii}$ , by maximizing the static profit equation:

$$\max_{\{y_{it}^d\}_{i \in (0,1)}} P_t Y_t - \int_0^1 P_{it} y_{it} di$$

subject to (2). Given the general price index is supposed to remain constant and normalised to unity, the demand for intermediate goods depends only on the relative prices of intermediate goods,  $P_{\rm tt}$ , and the aggregate demand:

$$y_{it}^d = P_{it}^{\frac{1}{\rho-1}} Y_i$$

<sup>&</sup>lt;sup>1</sup>This may nevertheless happen if either the representative firm has perfect foresight of the sequence of technology shocks or the wage contract is arranged in the form of a contingency plan. Both will be excluded here; see Gong and Semmler (2006) for a discussion on this latter point.

<sup>&</sup>lt;sup>2</sup>We do not follow the precise set-up here, mainly for reasons of analytical simplicity.

### **Synthetic Dataset & Preprocessing**

#### II. Handwritten Text Formulae

- → Used IAM dataset having handwriting of various common types and CROHME dataset having mathematical handwritten formulae.
- → Designed a heuristic to generate a synthetic dataset consisting of 1000 handwritten pages in a randomized manner having the following features:
  - Having text and formulae regions with different types of handwritings
  - Formulae in separate regions and also in between text in different types of alignments
- → Parse the coordinates of the formula regions by parsing the ground truth of the generated images.

eyes there has sping up a cyclen amy ted by from Naparties tell: WHST CHERMANY - general and officerations. They was a terminary fell: certisation that west Grammy could not provine even a few homes PITANN-cith its will termits and every seaker, is becoming and abundant. The Nobes to the Big There and a memorandom fold the U.S. - They difficult to array a boson or overdulf them was, acided by the U.S. - the U.S. - form of special deposits with the Back of Enjand. I official not alcowed. The back say to wie out such a hypergeneral and officers.

-1000 x 31 U1 = a40+6

not alcowed. ... citics holded of one days in Europe the U.S. lell: East Garman dishulances on day 12, 17853. Author is underlined in The hig "sprease" wear that it is joy to be used I officials

 $A_{2K} = \frac{2RA_K}{2R + \sqrt{4R^2 + A_K^2}}$   $\lim_{n \to \infty} f(n) = 0$ 

 $\left[\frac{2}{5}\left(1+\frac{9}{4}\infty\right)^{\frac{3}{6}}\frac{4}{9}\right]_{0}^{4}$ 

with its untilorists and veverage seaker, is becoming at of officials Persian could like to see a characteristical to arrays a logu or overdules. entiration that west Generaly could not survive even a few hours PAHANCE-best about, in their Generally, he fore every one of the banks will habell. West about, in their Generally, he fore every one of their Generally BPITAIN.

### Training & Testing

- → Selective search to obtain region proposals of the images.
- → Intersection over Union (IoU) of the proposed region with the annotated formula region used to add images to the training set as labelled or not.
- → Limit the number of positive and negative samples entering the model, to an equal of 30 per page.
- → Because there are many white spaces inside some formulas, the loU is set to >= 0.6 for formulae region, and <= 0.05 for non formulae region.
- → Splitting Ratio: 80:20



#### Model

- → The model implemented is Regions with Convolutional Neural Networks (RCNN).
- → We use VGG 16 Model because it performs well with specific tasks, along with two unit softmax layer, because our prediction is binary in nature.
- → We make some **modifications** on the images before passing to the model like **horizontal flip, vertical flip and rotation** to increase the dataset.

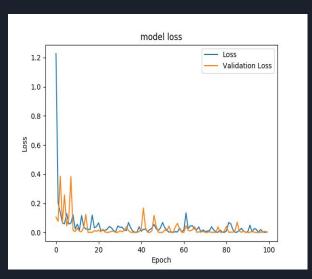
Layer (type)	Output	Shape	Param #
input_1 (InputLayer)	(None,	224, 224, 3)	Θ
block1_conv1 (Conv2D)	(None,	224, 224, 64)	1792
block1_conv2 (Conv2D)	(None,	224, 224, 64)	36928
block1_pool (MaxPooling2D)	(None,	112, 112, 64)	0
block2_conv1 (Conv2D)	(None,	112, 112, 128)	73856
block2_conv2 (Conv2D)	(None,	112, 112, 128)	147584
block2_pool (MaxPooling2D)	(None,	56, 56, 128)	Θ
block3_conv1 (Conv2D)	(None,	56, 56, 256)	295168
block3_conv2 (Conv2D)	(None,	56, 56, 256)	590080
block3_conv3 (Conv2D)	(None,	56, 56, 256)	590080
block3_pool (MaxPooling2D)	(None,	28, 28, 256)	0
block4_convl (Conv2D)	(None,	28, 28, 512)	1180160
block4_conv2 (Conv2D)	(None,	28, 28, 512)	2359808
block4_conv3 (Conv2D)	(None,	28, 28, 512)	2359808
block4_pool (MaxPooling2D)	(None,	14, 14, 512)	Θ
block5_conv1 (Conv2D)	(None,	14, 14, 512)	2359808
block5_conv2 (Conv2D)	(None,	14, 14, 512)	2359808
block5_conv3 (Conv2D)	(None,	14, 14, 512)	2359808
block5_pool (MaxPooling2D)	(None,	7, 7, 512)	Θ
flatten (Flatten)	(None,	25088)	0
fcl (Dense)	(None,	4096)	102764544
fc2 (Dense)	(None,	4096)	16781312
dense 1 (Dense)	(None,	2)	8194

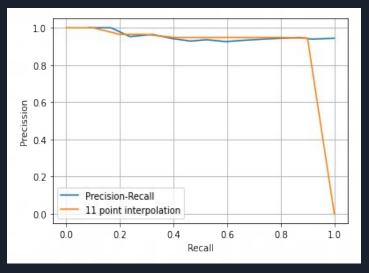
Total params: 134,268,738 Trainable params: 126,633,474 Non-trainable params: 7,635,264

### Training Strategy & Loss Function

- → Adam for optimizer because it is efficient in use and as it does stochastic optimization. The learning rate is set to 0.001.
- → Loss and accuracy are kept track of in each epoch, for 100 epochs with 10 steps per epoch and update the model parameters only if the loss value decreases.
- → Categorical cross entropy is used as loss because our prediction is categorical and binary, whether or not there is a formula (one hot encoding).
- → For prediction, we get proposed regions using selective search and then obtain the model prediction, which if it is >= 0.75, we consider the proposed region as having formula, otherwise not.

### Results





- → Loss and accuracy are kept track of in each epoch, for 100 epochs with 10 steps per epoch and update the model parameters only if the loss value decreases.
- → Average Precision: 87.38%, calculated using 11 Point Interpolation Technique

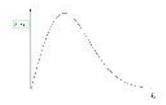
### Conclusion

- → The model we have implemented is able to identify formulae in computerized as well as handwritten forms in the images.
- → Some regions were formulas get selected inside text are when there are equations/formulas inside the text itself, or inside tables or figures.
- → Average Precision of our Handwritten Formulae Detection is: 87.38%
- → Code is available at:

https://drive.google.com/drive/folders/1melLTux5SQs13MkyC tkehlHKo5U3 947?usp=sharing

2.4 Properties describet apical Crises, police





Power by the management of comm-

Policia hide history

When J is the function countries operates  $\lambda \lambda$ , the stance L form the single displaced in form  $2\lambda$ . For an array with the expectation by which is a consequence of the stance of the s

In an income as one way of the energy of the state of  $\frac{df}{dt}$ . If t is the stable f is the stable f is the stable f in product f and f is the stable f.

At the couper point of the edge  $E_{ij}$  if and  $E_{ij}$  will have a again appaired that of  $E_{ij}$  because in the coupling of the couper point of the couper  $E_{ij}$  and  $E_{ij}$  is a constant of the  $E_{ij}$  because the couper  $E_{ij}$  and  $E_{ij}$  is smaller than each top quadrate which larger than compared the coupling of the threshold matter.  $E_{ij}$  is the coupling of the form of the  $E_{ij}$  and  $E_{ij}$  is the coupling of the coupli

Product and deling Will take [10] and Will be and other [11] the other in the design of the design processing that could be the only product with the product of the design of the other product of the design of the other product of the other

#### 3.3 Properties of the anisotropic diffusion equation



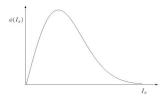


Figure 3.2: The flux as a function of the gradient

First we have to define the flux  $\phi$  as:

$$\phi(I_x) = C(I_x)I_x \qquad (3.5)$$

When C is the function defined in equation 3.3, the flux will have the shape depicted in figure 3.2. For an unconstrained diffusion process the flux would be a straight line, whereas the anisotropic diffusion equation causes the flux to decrease for high gradient values. Equation 3.2 can now be re-written as:

$$\frac{\partial I}{\partial x} = \frac{\partial}{\partial x} \phi(I_x) = \phi'(I_x)I_{xx}$$
(3.6)

To see that this equation can enhance edges, we have to look at  $\frac{\partial f_x}{\partial r}$ , that is obtained by taking the derivative with respect to x of the previous equation:

$$\frac{\partial I_x}{\partial x} = \phi''(I_x)I_{xx}^2 + \phi'(I_x)I_{xxx}$$
(3.7)

At the steepest point of the edge  $L_{xx}=0$  and  $L_{xx}$  will have a sign opposite to that of  $I_x$ , see figure 3.3. This means that if  $\phi'$  is greater than zero edges will be smoothed and if  $\phi'$  is smaller than zero edges will be enhanced. As can be seen in figure 3.2 the slope of  $\phi$  is smaller than zero for gradient values larger than some value dependent of the threshold value K, in the case that C is the function given in equation 3.3 the critical value is  $K/\sqrt{2}$ . If the parameter K is chosen close to its critical value, given an edge with a known gradient, the anisotropic diffusion will take less time to enhance or blur the edge, since the flux is largest for the critical value. This does make the process more sensitive to weak points in the edge however. If somewhere along the edge the gradient is smaller than the critical value for the gradient, the intensity will "leak" away from the object at that point.

Further analysis by Whitaker [45] and Whitaker and Pizer [46] shows that the anisotropic diffusion equation can produce edges that consist of several plateau's with small edges in between resembling a staircase, and therefore this property of the anisotropic diffusion equation is called the staircasing effect. This effect happens if the gradient is higher than a certain

### **Computerized Formulae Detection Sample Results**

#### 1. SEPTICS TO LENGTH TO THE PROPERTY OF THE PR

The incomediate for grant beams the markets has been deaded at Markets and Committee of the Committee of the

$$J(\varphi) = \int_{\mathbb{R}^{n}} J(\underline{\Psi}(\underline{x}, \underline{x})) dx, \qquad (6)$$

where the Part II is defined for the property of the control of th

#### A first decomplished and call lists the fundamental and

displacement (Marson, et al., confidence part of Marson, and at 11 february part of the first produced of the second of the seco

 In this properties on properties on again angle of auditors y because Construction of African state

The first considerable is a class of detects and the experimental constraints of the experimental constraints and experimental constraints of the experimental

The root of the company of the control of the contr



[8] J. S. Bert, C. Lee, a Problem in thoughtness described intelligence for many an appropriate in 2006 resistance spin and described in the control of the company of pages.

#### ASSESSMENT OF SHORT OF

Different between compared in a paint and only account of decident and a second of the second of the

a construction of the self-transfer to a sign of the State of the self-transfer of the self-t

#### Mr. To enhancing Contains M. auto-

the alignment of the contract of the contract

#### A STREET, TO

As a consequent of the policy of the second of the consequent of the consequent of the consequence of the co

as a superconnection of the contract of the co

exceleration a definite embalar a familie de melece de

Agreed by the manufacture of the second of the second second

#### 3. A SEQUENCE OF UNSCENTED TRANSFORMS

The unscented transform approximates the transformation  $Y=\{X\}$  of a Gaussian random variable by  $\sim$  surprise  $\sim$  Gaussian random variable. That is perfectly reasonable for linear and approximately linear transforms. In the presence of considerable nonlinearities, however, the Gaussian fit might not well represent the true distribution. Therefore, Alspach and Soronson [3] proposed to approximate the distribution  $p(x) = \mathcal{N}(x, \mu_X, \Sigma_X)$  of X as a mixture

$$p(x) \approx m(x) \triangleq \sum_{k=1}^{K} c_k \underbrace{\mathcal{N}(x, \mu_X^{(k)}, \Sigma_X^{(k)})}_{=p(x|k)}$$
 (11)

of Gaussian distributions with weights  $c_a$ , means  $\mu_A^{(k)}$  and covariances  $\Sigma_a^{(k)}$ . Then X can be transformed by transforming the individual mixture components, i.e. the conditional variables X/k. If the can be chosen small enough for f to be approximately linear for each of the transforms and the distribution of Y is well approximated by

$$p(y) \approx \sum_{k=1}^{K} c_k \text{UT}\{p(x|k), f\},$$
 (12)

where  $\mathrm{UT}[p(z|k), f]$  denotes the unscented transform of the Gaussian random variable X[k] with respect to the function f. The six-ture parameters in (11) are typically chosen as to minimize the mean squared error to the true distribution. In order to do that, the Gaussians are arranged on an equidistant grid with equal covariance matrices. Then the mixture weights are optimized [3]:

#### 3.1. Adapting the Level of Detail to the Nonlinearities

Replacing a Gaussian by an equidistant grid of Gaussians, as in [3]. can be regarded as increasing the level of detail in a uniform fashion. In contrast to that, we propose here to adapt the level of detail an uniform fashion. In contrast to that, we propose here to adapt the level of detail according to the nonlinearities, by that we mean keeping fewer Gaussians in relatively linear regions, where the transformation is accurate, more Gaussians in nonlinear regions, where the approximation error is higher. That can be achieved by performing the Gaussians in component alwain great destination of X- the variable to be transformed. Then, after an initial unscented transform, the mixture, and thereby the transform, is iteratively refined by:

- splitting that mixture component with the highest degree of nonlinearity into two Gaussians that have half the weight;
- repeating the unscented transform for the split components and re-evaluating their degrees of nonlinearity.

The iteration is stopped when either the degree of nonlinearity has dropped below a certain threshold or when a certain number of mixture components has been reached. After, the complexity of the Gaussian mixture approximation of  $p_{Y}(y)$  can be reduced by Gaussian mixture reduction techniques [9].

In this work, splitting was performed as described in Appendix B of [4]. For stacked variables (see Section 2.2), only the variable  $\chi^{(j)}$  with the highest degree of nonlinearity  $\eta^{(j)}$  was split. The other variables  $\chi^{(i)}, l \neq j$  remained unchanged.



Fig. 2. Splitting a Gaussian. The picture to the left shows the original Gaussian distribution; the pictures in the middle and to the right show the maximum and the mixture of the split components, respectivals.

#### 3.2. Splitting Priority

Selecting the mixture component to be split based only on its degree of nonlinearity can nonetimer scut in repeated splits of components whose weights are getting increasingly smaller. That might be suboptimal, as components with a very low weight represent only a small amount of probability mass and thereby do not contribute much to the transformation. Hence, we replace the splitting criterion from the previous section—the component's degree of nonlinearity (s)<sup>1</sup>—by the splitting priority jo<sup>1</sup>, which we define as geometric interpolation between the component's mixture weight and its degree

$$\rho^{(k)} \triangleq \text{pow}(c_k, \beta) \cdot \text{pow}(\eta^{(k)}, 1 - \beta),$$
 (13)

where  $\beta$  is the interpolation weight and where pow(a, b) denotes the b-th power of a, used here for disambiguation from superscript (k).

#### 3.3. Transforming Gaussian Mixtures

The adaptive level of detail approach from Section 3.1 can easily be extended to transforming Gaussian mixture random variables, simply by initializing the procedure with a Gaussian mixture instead of a sinele Gaussian.

#### 4. EXPERIMENTS

The proposed method was evaluated by performing a simulation, in which the distribution  $p_{\nu}(y)$  of noisy speech V was to be approximated from the distribution  $N(x_{1}, y_{N}, y_{N})$  of clean speech X, given the distribution  $N(x_{1}, y_{N}, y_{N})$  of noise N was well as a nonlinear interaction function in the speech feature (logarithmic MeI spectra) domain:

$$y = \underbrace{\log(\exp(x) + \exp(n))}_{-f(x,n)}.$$
 (14)

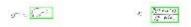
This transformation is the central point of all speech feature enhancement approaches that estimate the distribution of noise with the expectation maximization (EM) algorithm [1]. Morivated by the fact that frequency bands can be reated independently if the Gaussians are assumed to have diagonal covariance matrices – as is quite common in automatic speech recognition – we simulate the transformation for only one dimension. In the simulation, speech had a Gaussian distribution with mean 3.3 and variance 3.0. The smallout on the mean 3.3 and variance 3.0. The small consistence of the simulation of the

A box of the common description and the

<sup>&</sup>lt;sup>1</sup>A linear transform of a Gaussian random variable always results in a Gaussian random variable.

### Handwritten Formulae Detection Sample Results

down is on daywood point with it with the forms and it the open. could deput my low rend . " high the upon the winter all with blood. als send the system - and would it will more and activity at the system paid when and fell a sum of rightnesseen who their built with you " 10.2000 motions to und it whending divine of their istinct leavings of my like is the same of several the try " on a police ties so down for a yould deline with a doubt of "



delet and cripted and adoptions coulding "my by " whater is made.

then he plant that he was only a well with on the sample to the light efficient up in it and is to over their did just of different and efficient by in the weather to me theter that just of the field the I as impact of the street of this just people "former and today" activations competensive that will very law was from in an 199 by to comply they would be present to that they in his the said treft. at the aperatection later who would be land righer the millipship with he will the system - and wast it will gave and making they believe marger and all the abbilliant stight that weller while made stor and fel a me of eighteness who ther land watered waying

. The jackale buryese quoted from Main his past. & move to stop Mr. Goubbell front and Malle.



operate all the opposition by terming attentions saying the Texted Bureau of hurshigation had, I PRESIDENT KENNEY is mady to get thoughting. I



Whatighon goods that whow once undendor is to be made at a meeting of Labour !



Africans the owner majority they are people. no a parel declar sin faller mountain majects. the new knowledge administration in his post the 

Then negro is her. Robot atoms Epile in Floring. he and not known wholes knowica wonder. Hay I ringing the Ferteral Bureau of huroligation had I Sir long is violently profiting making morgans.



But representative of Sir Roylanocostine. And he indicated book him Governmentschanger. I week used to me home least able to bent how the consider was monstroughbest German Garaning. corried on in financing ". In his post The Hills where , our -. " PRESIDED! WERNEDY LOBOR THE HE This surprise present our a shorp about Them quickly round out that this is money another the Occaminate to get them to pay a regular. Ma the chance go from Sir loy werenity we mercican was wany different fudamentally the with Pollater, of Virginia- and Johan when a J the new purchase administration projects, Then PRESIDENT KENNEDY bodayourrish on susbertle

#### References

- → Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks: Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun
  - https://arxiv.org/abs/1506.01497
- → ICDAR 2017 Competition on Page Object Detection: Liangcai Gao, Xiaohan Yi, Zhuoren Jiang, Leipeng Hao, Zhi Tang
  - https://ieeexplore.ieee.org/document/8270162
- → Datasets
  - ♦ ICDAR 2017 Dataset: https://mega.nz/file/6QlwGaAb#BKf962iBlfeL7oEqaVnDC4K3F47zrqtaU12O CJlcbTw
  - IAM Dataset: <a href="https://fki.tic.heia-fr.ch/databases/iam-handwriting-database">https://fki.tic.heia-fr.ch/databases/iam-handwriting-database</a>
  - CROHME Dataset: <a href="https://www.isical.ac.in/~crohme/CROHME">https://www.isical.ac.in/~crohme/CROHME</a> data.html

# Thank You