



Signature detection as a way to recognise historical parish register structure

Solène Tarride, Aurélie Lemaitre, Bertrand B. Couasnon, Sophie Tardivel

► To cite this version:

Solène Tarride, Aurélie Lemaitre, Bertrand B. Couasnon, Sophie Tardivel. Signature detection as a way to recognise historical parish register structure. HIP 2019, Sep 2019, Sydney, Australia. pp.54-59, <10.1145/3352631.3352636>. <hal-02433964>

HAL Id: hal-02433964

<https://hal.science/hal-02433964v1>

Submitted on 18 Nov 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 CC0 1.0 - Universal - International License

Signature detection as a way to recognise historical parish register structure

Solène Tarride

solene.tarride@irisa.fr

Doptim, Univ Rennes, CNRS, IRISA

Rennes, France

Bertrand Couasnon

bertrand.couasnon@irisa.fr

Univ Rennes, CNRS, IRISA

Rennes, France

Aurélien Lemaitre

aurelie.lemaitre@irisa.fr

Univ Rennes, CNRS, IRISA

Rennes, France

Sophie Tardivel

sophie.tardivel@doptim.eu

Doptim

Rennes, France

ABSTRACT

This article deals with the analysis of pages of French parish registers from 16th to 18th century. These documents are structured in paragraphs called acts. Each act contains valuable demographic information that can be useful to genealogists willing to find information about their ancestors. The first step toward parish register analysis consists of delimiting each act. But these documents are so poorly-structured that the visual separation between the acts is not always clearly visible. One of the main visual indication of separation is the signature of the priest at the end of each act. In this work, we propose to train and compare several u-shaped neural networks for signature detection. We also propose a rule-based system for segmentation into acts and evaluate the impact of signature detection at act level.

CCS CONCEPTS

• Applied computing → Document analysis.

KEYWORDS

parish registers, neural networks, signature detection, structure analysis

1 INTRODUCTION

Parish registers - or church books - are documents in which were recorded the acts of baptisms, marriages and burials. From 1539, priests were required by law to keep these records written in French in order to save nominative and dated information about the population. This type of document is especially useful to people willing to find their ancestors. In France, parish registers are all the more important before 1792 since the French administration did not register the population before this date. Consequently, parish registers are a major source of demographic data.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

HIP '19, September 20–21, 2019, Sydney, NSW, Australia

© 2019 Association for Computing Machinery.

ACM ISBN 978-1-4503-7668-6/19/09...\$15.00

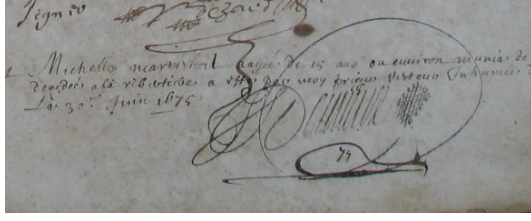
<https://doi.org/10.1145/3352631.3352636>



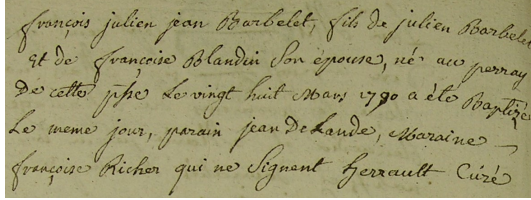
Figure 1: Pages from a French parish register from 1727 containing 5 acts on the left page and 3 acts on the right page

Recently, there has been growing interest in transforming these records into digital documents to automatically extract information. However this task is laborious : unlike civil records, parish registers are quite poorly-structured documents, as shown in Figure 1, and present many challenges relative to old documents such as bleed-through degradations, faded ink and contrast variations. But they also present artefacts related to the writer, like cross-outs, ink stains, small spacing and overlapping texts and signatures. All these characteristics make them hard to understand and process. Besides being challenging to read, their structure is also hard to recognise. Each page can contain up to ten acts. At first glance, one could believe that several elements allow for visualising the separation between each act: margin annotations, punctuation, vertical or horizontal spacing... However, these visual elements are not consistent among documents since they depend on the writer. Yet, the end of each act is very frequently signed by the priest, as shown in Figure 2a and Figure 2b. Sometimes, it is also signed by the witnesses, especially for marriages, like in Figure 2c. Consequently, signatures are the main graphical elements that could help to delimit the acts.

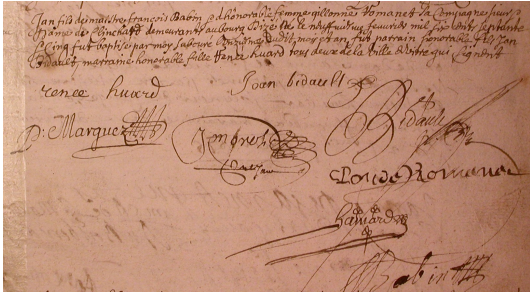
Yet, signature detection is not easy in such documents. First of all, signature appearances are variable as shown in Figure 2. Each writer has a unique signature making the signature detection task hard for an automatic system. Moreover, it can be hard to detect some signatures without any information on the meaning of each word. For instance in Figure 2b, the two last words are actually a



(a) Act followed by a stylized signature



(b) Act followed by a non-stylized signature



(c) Act followed by a group of signatures

Figure 2: Illustration of different kinds of signatures found in parish registers

signature but could easily be interpreted as words by non-French-speaking readers. Second, their localisation is not regular. If most signatures are likely found at the end of a text line (e.g in the right side of the page), some can be found under the act in the middle of the page or in the left side of the page.

In this paper, we aim to delimit each act in these records using signatures as separators. Isolating an act is the first step towards church record analysis and would allow to extract more meaningful information at this level, such as dates and names.

2 RELATED WORKS

To the best of our knowledge, the challenge of finding signatures in historical books has not been addressed yet. This can be easily explained since this task is very specific to our problematic. However, many similar issues have been successively tackled in historical documents: layout analysis, text line segmentation, baseline detection and ornament/decoration detection.

In the last literature, many neural-network based methods managed to significantly improve state-of-the-art performances [2] [11], especially u-shaped networks. U-Nets were originally introduced by Ronneberger et al. [10] for biomedical segmentation. It is now widely used for semantic segmentation in many fields of research and industry, including document image segmentation. The principle of U-Net is the following: first, an encoder network takes the

image to be processed as an input and outputs a feature map, i.e. a representation of the input in the feature space. Second, a decoder network that is symmetrical to the encoder takes that feature map and outputs the closest match to the actual input.

Recently, Oliveira et al. [9] introduced dhSegment, a u-shaped neural network that was successively used to solve five different tasks relative to document image segmentation, including ornament detection which is very similar to our task. Promising results were obtained using few training data. One of the advantages of dhSegment is that it uses a pre-trained model (e.g ResNet-50 [6], VGG-16 [12]) as an encoder so that it takes advantage of the high level features learned on the ImageNet dataset [1]. Such a strategy also significantly reduces the training time.

Grüning et al. [5] propose several improvements to the U-Net for baseline detection. They designed a U-Net with residual blocks, called RU-Net. Residual blocks increase the representative power of the network. The authors also introduced ARU-Net, which is RU-Net designed with a spatial attention mechanism. Finally, they proposed LARU-Net, which is ARU-Net with a MDLSTM layer at the lowest resolution. Taking advantage of data augmentation strategies, they managed to train these models from scratch with few training data. Results show that the ARU-Net architecture performs significantly better than the others for text-line segmentation. In the following, we refer to all these improved U-Net architectures to {LARU}-Net.

We choose to train both dhSegment [9] and {LARU}-Net [5] for the signature detection task. This decision is motivated by three main reasons:

- both networks perform well when trained on similar tasks;
- they do not require a lot of training data;
- both frameworks are freely available on Github.

Another interesting approach is presented in [8] where a rule-based grammar is used to find the location of handwritten fields in pre-printed records. Since the documents share a similar layout, the authors design a logical description of the contents of the collection of documents in order to find a region of interest based on other visual indicators (line segments, keywords...). In the end, they were able to successively detect 97.2% of the fields.

This paper aims to take advantage of both approaches. It is organised as follows: first we describe the objective of the article as well as the methods used, then we describe our database and the experiments performed, finally we evaluate our strategies and comment on the obtained results.

3 METHOD FOR ACT DETECTION

3.1 Objective

Since signatures are key elements regarding the document structure, our first goal is to achieve pixel-wise segmentation of the signature class. We also want our model to be able to recognise signatures using few training data. Finally, as our goal is to segment pages into acts, we have designed a rule-based description grammar to delimit the acts from the location of the signatures. We evaluate our system at the pixel level and at the act level by comparing the acts obtained when using ground truth and predicted signatures as an input.

3.2 Presentation of the networks used

We choose to train dhSegment and {LARU}-Net to detect signatures. We use the code provided by the authors. Both networks were introduced to extract text baselines in old documents. The two architectures are both u-shaped networks - meaning that they consist of a contracting path followed by a symmetrical expansive path. Figure 3 shows the structure of U-Net as it was originally proposed by Ronneberger et al. [10].

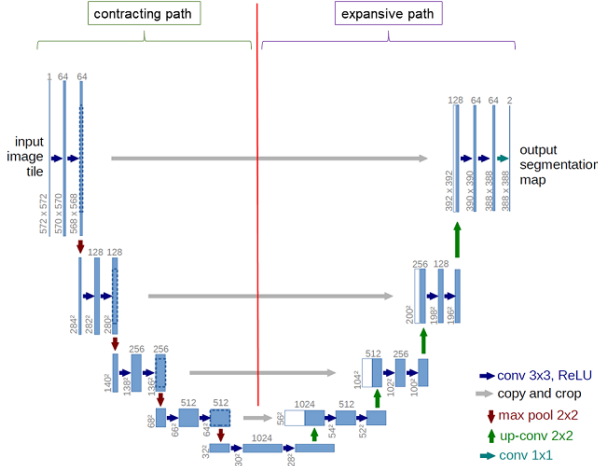


Figure 3: U-Net architecture [10], ©Springer

3.2.1 dhSegment. dhSegment is based on the U-Net architecture except that its contracting path consists of a pre-trained model e.g ResNet or VGG. Consequently, the training time is significantly reduced. The authors also observed that using pre-trained weights was also beneficial to the learning process since the model appeared less sensitive to outliers. For our task, we use a ResNet as a contracting path since it gives better results. Weights are initialised using Xavier initialisation [4] and optimised using Adam optimiser [7]. dhSegment can take as input images of any size as long as they fit into memory. To this end, images are resized so that the total number of pixels lies between 6×10^5 and 10^6 . Images can also be cropped into patches (we used patches of size 300×300) in order to allow batch training. A margin was added to avoid any border effects. Data augmentation is applied to all images: rotation and scaling. The authors claim that these parameters can be applied to most tasks. Just note that the input resolution of the images must be chosen carefully to fit the task.

3.2.2 {LARU}-Net. {LARU}-Net refers to 3 different improvements to the U-Net architectures proposed by Grüning et al. in [5]:

- RU-Net: U-Net with residual blocks
- ARU-Net: RU-Net with attention mechanism
- LARU-Net: ARU-Net with a MDLSTM layer

We refer the reader to [5] for a detailed description of the different architectures. We choose to compare U-Net, RU-Net, ARU-Net and LARU-Net to build our method. Each model is trained from scratch with the following parameters. Input images can be of any size. They are normalized and downscaled by a factor of 2, 3 or 4 depending of their original size. Xavier initialisation [4] is used and the weights

are optimised using Adam [7]. Due to memory limitations, batch size is fixed to 1.

3.3 Rule based-grammar

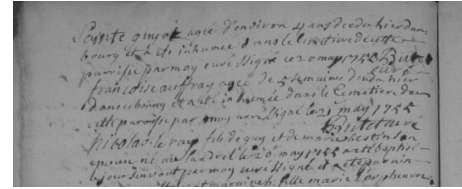
We design a logical description of the documents at the collection level. Our main goal would be to delimit each act in parish registers. Signatures are essential since they mark the end of each act. Another key visual aspect of the acts lies in the text lines. Information about the localisation of signatures and text lines in every documents allow us to design a logical description of parish registers to delimit the acts. This idea has been presented in [8] where the authors combine the results of an automatic analysis with a logical description of the contents of the documents.

3.3.1 Description. The designed system takes as an input the localisation of text lines and signatures, and outputs the bounding box of each act. Text lines were extracted using the baseline detector proposed in [5] which is freely available on Github.

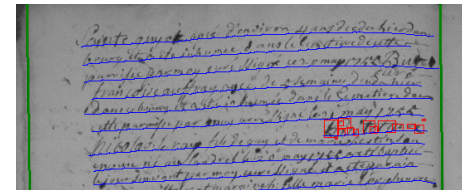
We use the DMOS-PI method to implement logical rules for this collection. The main outline of the algorithm is the following:

- Detect the pages in the image,
- In each page, look for a group of signatures. Look for text lines upwards. This combination of text lines and signatures forms an act.
- Repeat until finding all the acts in the page.

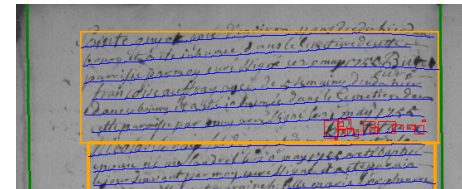
Result for one act is presented in Figure 4c.



(a) Original image



(b) Input signatures (red) and baselines (blue)



(c) Output: bounding box of the acts (orange)

Figure 4: Act construction using baselines and signatures

Applying this logical description algorithm using baselines and ground truth signatures outputs a pseudo ground truth of the acts.

3.3.2 Limit of the description. Logical rules were designed to produce a correct delimitation of the acts that are signed. However, while most of the acts are signed, some of them are not. For instance, in Figure 5a, there are actually 3 acts, however two of them are not signed. These acts are ignored to be eventually incorporated in the next signed act, as shown in Figure 5b.

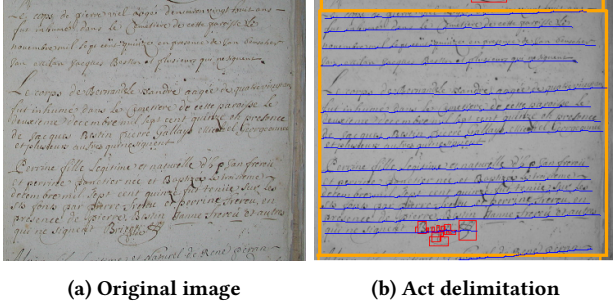


Figure 5: Act construction when some acts are not signed

4 EXPERIMENTS

4.1 Description of the database

Our database is provided by Archives Départementales d’Ille-Et-Vilaine (35, France). It contains 340,000 images of documents from 54 cities from 16th to 18th century.

Obtaining training data is time-consuming since each signature must be delineated. We have arbitrarily selected 200 images from this database such that each city is represented, and that it spans the whole period. We used VGG Image Annotator [3] to draw polygons around each signature (Figure 9a). Two ground truth images are obtained :

- Figure 6c: pixels belonging to polygonal zones are set to 255, others are set to 0.
- Figure 6d: pixels belonging to polygonal zones are set to their value in the corresponding binary image (Figure 9b), other pixels are set to 0.

For each image, we tried to capture the signatures without including surrounding elements (text, ink stain...) in order to limit the noise induced by the binarisation algorithm. Ground truth images could be improved by precisely delineating the signature outlines, however this would require a significant amount of time.

4.2 Signature detection

We choose to compare two Neural Network architectures: dhSegment and {LARU}-Net where {LARU}-Net refers to all the models presented by Grüning et al. [5] (e.g U-Net, RU-Net...).

4.2.1 Data augmentation. Both [9] and [5] propose strategies of data augmentation. However, they were specifically designed for baseline detection. For signature detection, we chose to only apply transformations that do not alter the symmetry of the documents. All images are rotated with $r \in [-0.2, 0.2]$ rad and scaled by a random factor $s \in [0.2, 0.5]$ to speed-up computations. These transformations are applied on-the-fly, and double the training set size.

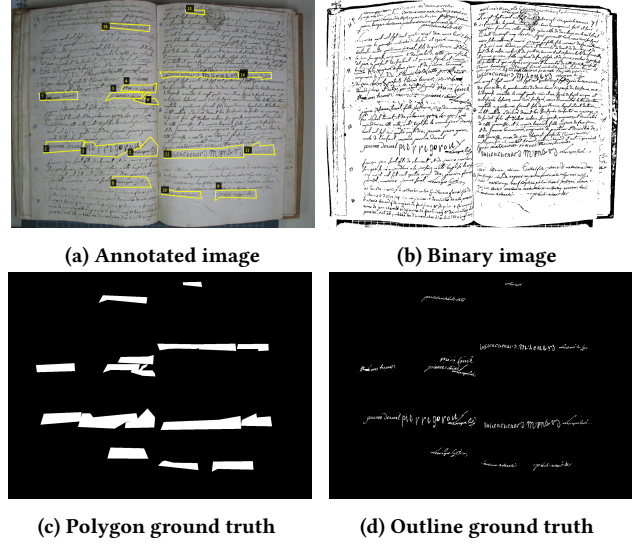


Figure 6: Polygonal zones are drawn around each signatures. Ground truth is obtained by combining the binary image and the polygonal zones.

4.2.2 Experiments. We train different architectures in order to compare them. To this end, we perform two main experiments.

The first main experiment consists in a binary classification problem where we use the ground truth images of the outline of signatures (Figure 6d). Background corresponds to class 0 and signature outlines to class 1. In this task, the two classes are exclusive so the softmax cross entropy loss function is used.

Our second experiment is based on the method presented in [5] that suggests that using multiple classes actually helps the training. In the same way, we use 3 classes for signature detection: signatures (Figure 6d), surrounding area (e.g polygonal zones in Figure 6c) and background. For this task, a pixel can belong to more than one class (e.g signature and polygonal zone) so the sigmoid cross entropy loss function is used.

We use the Adam optimizer [7] and Xavier initialisation [4] for all experiments. {LARU} architectures are trained for 250 epochs. dhSegment is trained for 100 epochs since a pre-trained Resnet is used as a contracting path.

4.2.3 Training. We use 5-cross-validation which means that our dataset is divided into 5 subsets. We run the training 5 times, using 3 subsets for training (120 images), 1 for validation (40 images) and 1 for testing (40 images). The main advantages is that every image is at one point in the testing subset so we get an estimation of the test error on all 200 images. It should be noted that we followed the recommendations of the original articles for the number of epochs. {LARU}-Net is trained from scratch, which explain why more epochs are needed.

All experiments run on Nvidia RTX 2080 Ti GPU. Each fold takes between 2 and 10 hours depending on the model.

4.2.4 Post-processing. The output probability map is threshold with a threshold of 0.5. We do not apply further post-processing technique.

5 EVALUATION STRATEGIES

We propose to evaluate the models on their ability to build the acts from their signature predictions. We also evaluate the models at the pixel level in order to compare our results to other tasks. Results at the pixel level show how well the image pixels are correctly classified while results at the act level show how well the acts are detected using the signature predictions. All scores are summarised in Table 1.

5.1 At the pixel level

We compute several metrics to compare the performance of the tested models. Note that we only compute scores for pixels belonging to signatures, e.g for the class pictured in Figure 6d.

We use 5 cross validations which means that every image has been tested exactly one time. All scores presented in Table 1 are averaged on all test images except for 6 images that contain no signature. Three scores are presented :

- Recall: a high recall means most of the signature pixels are detected.
- Precision: a high precision means that more relevant signature pixels than irrelevant ones are detected.
- F1-score: the harmonic mean of recall and precision.

In Figure 7, we depict some results obtained for the best network on 3 images: 2 containing signatures with high and low F1-score and one containing no signature. For each image, we zoom on one specific predicted signature or group of signatures to give an illustration of the segmentation quality. We observe that the model sometimes struggles to differentiate text and signatures.



Figure 7: Error visualisation for signatures taken from: an image with a high F1-score ($F1 = 0.73$) (1), an image with a low F1-score ($F1 = 0.35$) (2), an image with no signature (3). True positives appear in green, true negatives in black, false positives in blue and false negatives in red.

5.2 At the act level

We obtain a pseudo-ground-truth of the acts by applying the grammatical description to the ground truth signatures. Now, we will apply these same rules using the signature predictions, and compare the acts obtained with the pseudo-ground truth.

We use the criteria introduced in the PASCAL VOC challenge 2012: a prediction is considered positive if $IoU > t$. In some images

there are more than one detected act overlapping a ground truth act. For those cases the detection with the highest IoU is taken while the others are counted as false positives.

For most object detection tasks, the threshold t is set to 0.5. However, this threshold is too permissive for our case since we mainly observe vertical variations: a predicted act could overlap two ground truth acts and still be accepted as true positive. We choose to set a stricter threshold $t = 0.65$ in order to avoid this issue. Figure 8 compares two predicted acts with different thresholds, showing that $t = 0.5$ is too permissive while $t = 0.65$ appears reasonable.

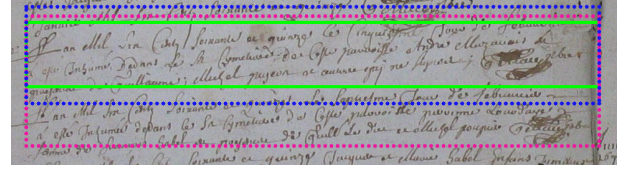


Figure 8: Comparison between different thresholds. The green box represents a ground-truth act, the blue one is a prediction with $IoU = 0.65$, the pink one is a prediction with $IoU = 0.5$ that overlaps two ground truth acts.

Once the predicted acts have been associated with ground truth acts, several metrics can be computed. We have chosen to compute the precision, recall and average precision - which are frequently used metrics for object detection. The average precision (AP) is the area under the precision/recall curve with precision decreasing. All results are presented in Table 1.

One example of ground truth versus detection is depicted in Figure 9. On the detection image, a false positive signature leads to the creation of a false positive act. In this case, none of the two detected acts sufficiently overlap the ground truth so they are counted as 2 false positives. The ground truth is missed and will be counted as false negative.

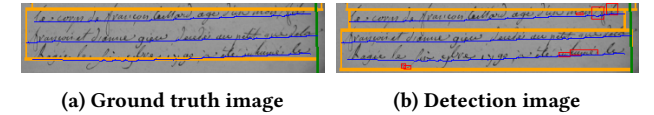


Figure 9: Comparison between ground truth acts and corresponding detection

5.3 Discussion

The aim of our method is to segment registers into acts by using signatures as separators. Our approach allows us to recover 80% of the acts using the LARU-Net architecture with 3 classes. Not only this architecture performs better than the others, but the training is also more stable over the 5 folds. We also show that data augmentation significantly improves the predictions for this model. More generally, we observe that using three classes instead of two definitely improves the performance on the signature class.

As expected, we observe a correlation between the scores at the pixel and act levels. Scores at the act level are higher than at the pixel level because our rule-based grammar is able to construct the acts from a few detected signature pixels, regardless of their precise delineation. This result is encouraging considering the difficulty of the task. However, the obtained scores for signature segmentation are low as compared to performance obtained by dhSegment and

Architecture	Nb of classes	Epochs	Pixel level			Act level		
			Precision	Recall	F1-score	Precision	Recall	AP (%)
dhSegment (no patch)	2	100	0.36 ± 0.04	0.24 ± 0.01	0.28 ± 0.02	0.60	0.47	28
dhSegment (no patch)	3	100	0.37 ± 0.03	0.31 ± 0.05	0.33 ± 0.03	0.60	0.48	29
dhSegment (patches 300x300)	2	100	0.45 ± 0.04	0.22 ± 0.05	0.28 ± 0.04	0.62	0.43	27
dhSegment (patches 300x300)	3	100	0.37 ± 0.07	0.39 ± 0.09	0.35 ± 0.02	0.59	0.43	26
U-Net	2	250	0.37 ± 0.10	0.45 ± 0.09	0.38 ± 0.10	0.42	0.47	25
U-Net	3	250	0.33 ± 0.10	0.54 ± 0.09	0.37 ± 0.09	0.48	0.47	25
RU-Net	2	250	0.32 ± 0.10	0.48 ± 0.12	0.34 ± 0.11	0.39	0.43	23
RU-Net	3	250	0.36 ± 0.11	0.49 ± 0.12	0.38 ± 0.11	0.46	0.51	29
ARU-Net	2	250	0.59 ± 0.09	0.43 ± 0.11	0.46 ± 0.10	0.74	0.63	48
ARU-Net	3	250	0.62 ± 0.02	0.47 ± 0.08	0.50 ± 0.06	0.77	0.70	54
LARU-Net	2	250	0.59 ± 0.11	0.43 ± 0.08	0.47 ± 0.09	0.73	0.58	42
LARU-Net	3	250	0.63 ± 0.03	0.48 ± 0.07	0.52 ± 0.05	0.79	0.80	65
LARU-Net (no data augmentation)	3	250	0.50 ± 0.10	0.17 ± 0.09	0.23 ± 0.12	-	-	-

Table 1: Performance of various architectures averaged on test images. On the 200 test images, 6 containing no signatures have been discarded to compute the pixel level results. Standard deviation over the 5 folds is shown for pixel level performance. All 200 images are used to compute the act level results, which corresponds to more than 1500 acts.

{LARU}-Net for baseline detection. Several factors can explain this poor performance:

- **Image quality:** contrast variation, blur, uneven lightning...
- **Ground truth quality:** the signature ground truth is dependant on the binarisation and can therefore be noisy, especially on poorly contrasted images.
- **Training settings:** only 120 images are used for training, which corresponds to approximately 900 signatures.
- **Task complexity:** signatures are by nature very similar to surrounding text since they come from the same writer. A signature can look like a word (Figure 7c) and a word can look like a signature (Figure 7e).

6 CONCLUSION AND OUTLOOK

In this article, we have tried to recognise the structure of parish registers (e.g the acts) based on signature detection. To this end, we designed a logical description of the collection that takes the localisation of signatures as an input and outputs the acts. We also trained and compared two state-of-the-art neural networks to detect signatures in parish registers. Using this approach, we were able to correctly detect 80% of the acts. We believe that our results could be improved by taking into account other visual indicators (spacing, margin annotations, recurrent keywords...) that could help delimiting the acts in the logical description. We plan to train a neural network to recognise all these visual indicators. This would allow us to build more reliable acts. The logical rules have allowed us to build a pseudo-ground truth of the acts. We plan to take advantages of this pseudo-ground-truth by training a neural network to directly recognise the structure of the acts in order to compare the results with our current method. Finally, we plan to use a metric based on the text lines contained in the acts rather than the overlap in order to associate a predicted act with a ground truth.

Being able to automatically delimit the acts of parish registers would be a first step towards further record analysis. This would allow to extract useful information such as names, places and dates and, more importantly, to index the acts in order to ease the search for ancestors.

ACKNOWLEDGMENTS

The database used for this work was provided by Les Archives Départementales d’Ille-et-Vilaine.

REFERENCES

- [1] J. Deng, W. Dong, R. Socher, L. Li, Kai Li, and Li Fei-Fei. 2009. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*. 248–255. <https://doi.org/10.1109/CVPR.2009.5206848>
- [2] Markus Diem, Florian Kleber, Stefan Fiel, Tobias Grüning, and Basilios Gatos. 2017. cBAD: ICDAR2017 Competition on Baseline Detection. *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)* 01 (2017), 1355–1360.
- [3] A. Dutta, A. Gupta, and A. Zissermann. 2016. VGG Image Annotator (VIA). <http://www.robots.ox.ac.uk/vgg/software/via/>. Version: 2.0.6.
- [4] Xavier Glorot and Yoshua Bengio. 2010. Understanding the difficulty of training deep feedforward neural networks. In *In Proceedings of the International Conference on Artificial Intelligence and Statistics. Society for Artificial Intelligence and Statistics*.
- [5] Tobias Grüning, Gundram Leifert, Tobias Strauß, and Roger Labahn. 2018. A Two-Stage Method for Text Line Detection in Historical Documents. *CoRR* abs/1802.03345 (2018). arXiv:1802.03345 <http://arxiv.org/abs/1802.03345>
- [6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. Deep Residual Learning for Image Recognition. *CoRR* abs/1512.03385 (2015). arXiv:1512.03385 <http://arxiv.org/abs/1512.03385>
- [7] Diederik P. Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*. <http://arxiv.org/abs/1412.6980>
- [8] Aurélie Lemaitre, Jean Camillerapp, Cérés Carton, and Bertrand B. Couasnon. 2018. A combined strategy of analysis for the localization of heterogeneous form fields in ancient pre-printed records. *International Journal on Document Analysis and Recognition* 21(4), 269–282 (July 2018). <https://doi.org/10.1007/s10032-018-0309-y>
- [9] Sofia Ares Oliveira, Benoit Seguin, and Frédéric Kaplan. 2018. dhSegment: A generic deep-learning approach for document segmentation. *CoRR* abs/1804.10371 (2018). arXiv:1804.10371 <http://arxiv.org/abs/1804.10371>
- [10] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. *CoRR* abs/1505.04597 (2015). arXiv:1505.04597 <http://arxiv.org/abs/1505.04597>
- [11] F. Simistira, M. Bouillon, M. Seuret, M. Wäijrsch, M. Alberti, R. Ingold, and M. Liwicki. 2017. ICDAR2017 Competition on Layout Analysis for Challenging Medieval Manuscripts. In *14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, Vol. 01. 1361–1370. <https://doi.org/10.1109/ICDAR.2017.223>
- [12] K. Simonyan and A. Zisserman. 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. In *International Conference on Learning Representations*.