The BIR database – Identifying typographic emphasis in list-like historical documents

Anna Scius Bertrand, Simon Gabay, Ljudmila Petkovic, Juliette Janes, Caroline Corbières, Thibault Clérice

To cite this version:

Anna Scius Bertrand, Simon Gabay, Ljudmila Petkovic, Juliette Janes, Caroline Corbières, et al.. The BIR database – Identifying typographic emphasis in list-like historical documents. HIP@ICDAR21 - The 6th International Workshop on Historical Document Imaging and Processing, Sep 2021, Lausanne, Switzerland. 10.1145/3476887.3476913 . hal-03355683

HAL Id: hal-03355683
https://hal.science/hal-03355683

Submitted on 27 Sep 2021

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 - NonCommercial 4.0 International License

The BIR database – Identifying typographic emphasis in list-like historical documents

Anna Scius Bertrand*
name.surname@unige.ch
UniGE
Geneva, Switzerland
EPHE-PSL
Paris, France
HES-SO
Fribourg, Switzerland

Simon Gabay
Ljudmila Petković
name.surname@unige.ch
UniGE
Genève, Switzerland

Juliette Janes
Caroline Corbières
Thibault Clérice
name.surname@chartes.psl.eu
ENC-PSL
Paris, France

ABSTRACT

Layout analysis and optical character recognition have become traditional tasks for processing historical prints, but are now insufficient. Additional information is found in typographic emphasis, such as bold and italic letters. They carry semantic meaning (titles, emphasis. . .) and also outline the structure of the page (entries, sub-parts. . .). Retrieving such data is therefore crucial for information extraction and automatic document structuring. In this paper, we introduce the Bold-Italic-Regular (BIR) database, which contains 285 pages of scanned, list-like historical prints that have been annotated at word level with bold and italic emphasis. Baseline results are provided for word detection and style classification using state-of-the-art deep neural network models, highlighting promising possibilities, such as near-human performance for isolated word classification, but also demonstrating limitations for the task at hand.

CCS CONCEPTS

• **Applied computing** → **Arts and humanities**; **Optical character recognition**.

KEYWORDS

typographic information, historical print, object detection, style classification

ACM Reference Format:
Anna Scius Bertrand, Simon Gabay, Ljudmila Petković, Juliette Janes, Caroline Corbières, and Thibault Clérice. 2021. The BIR database – Identifying typographic emphasis in list-like historical documents. In *The 6th International Workshop on Historical Document Imaging and Processing (HIP '21), September 5–6, 2021, Lausanne, Switzerland.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3476887.3476913

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 the author(s) 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 '21, September 5–6, 2021, Lausanne, Switzerland
© 2021 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-8690-6/21/09. . . $15.00
https://doi.org/10.1145/3476887.3476913

1 INTRODUCTION

Thanks to the recent improvements of Handrwritten Text Recognition (HTR) engines [4, 24] and the creation of user-friendly interfaces [10, 14], retrieving the text from an image has become an increasingly easy task. But along with the text itself, (old) prints offer additional graphic information that convey a semantic meaning that is still too often forgotten. It is the case of the layout (*e.g.* a paragraph starts with a carriage return and, potentially, an indentation), but also of typographic emphasis (*e.g.* a title is in italic, such as a loan word). Identifying the latter is therefore crucial for digitisation tasks, may it be for traditional mining purposes (*e.g.* retrieving all the titles of works mentioned), but also automatic structuring of the text (*e.g.* encoding the logical structure thanks to physical and graphical hints found on the page).

The BIR database[1] introduced in this paper provides the research community with a benchmark dataset for developing and comparing methods for identifying **bold** and *italic* emphasis on scanned page images of list-like historical prints. In an experimental evaluation, we evaluate and compare state-of-the-art deep neural network models, including YOLOv5 [8] for word detection and MobileNetV2 [17] and Xception [1] for style classification. The obtained baseline results are promising but also demonstrate the difficulty of the task and the current limitations of the state of the art.

2 RELATED WORK

To be precise, "style" (*e.g.* italic, oblique. . .) or "weight" (*e.g.* bold, thin. . .) are attributes of possible "fonts" (*e.g.* Garamond, Arial. . .). When combined together we talk about a "typeface" (*e.g.* Garamond in bold vs Garamond in italic). These fonts are associated to "scripts" (*e.g.* latin for French or English, cyrillic for Russian or Serbian. . .). Very few researchers have addressed directly the question of typographic emphasis, but many have tackled the problem while dealing with a larger one – usually the identification of fonts or

[1] The BIR database is available via https://github.com/asciusb/BIR-database.

scripts, which are necessarily embodied in a typeface that can be in bold or italic.

The oldest approach to address typographic emphasis is the one of optical font recognition (OFR). Researchers have first tried to solve it using a font model base of several hundred known fonts and a multivariate Bayesian classifier [26]. Further studies have associated typographic emphasis with a texture. On the one hand, a simple weighted Euclidian distance has been used to classify fonts of synthetic data, after extracting features such as spatial frequency and orientation with multi-channel Gabor filters [25]. On the other hand, local binary patterns have been used with a synthetically generated database of arabic text images [12]. Most recent approaches obviously use neural networks, may they be LSTM [20] or CNN [19].

In the past few years, in parallel to OFR, script recognition has benefited from numerous studies [22], some of which may well be useful for the identification of typographic emphasis. It is particularly the case of experiments using the neural network implementation of OCR engines to identify multiple scripts at text-line level [23], a strategy used successfully for recognizing semantico-typographic classes (dictionary entry, antiqua, fraktur, author's name and letter-spacing) in a real historical use case: Daniel Sander's *Wörterbuch der Deutschen Sprache* [15].

3 THE BIR DATABASE

In the following, we comment on the general context of list-like historical prints, describe the specific document collection that constitutes the Bold-Italic-Regular (BIR) database, and provide details on how the scanned pages were annotated.

3.1 List-Like Historical Prints

During the past decade, list-like documents have drawn considerable attention from digital humanists, may they be art historians [9], historians [3], linguists [16], philologists [7]... Such documents have the particularity to use typographic emphasis to organise a rather dense information. Italic is used to identify professions in phone directories (cf. fig. 1), examples in dictionaries (cf. fig. 2) or title of works in manuscript catalogues (cf. fig. 3). Bold tends to have a more unified usage, and delimits the subject of the entry. Additional styles, such as small capitals, can also be found, as shown in the first example of the *Annuaire-Almanach.*

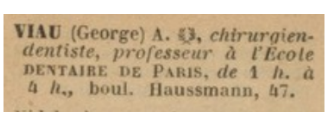
VIAU (George) A. ✠, *chirurgien-dentiste, professeur à l'Ecole* DENTAIRE DE PARIS, *de 1 h. à 4 h.*, boul. Haussmann, 47.

Figure 1: *Annuaire-almanach du commerce, de l'industrie, de la magistrature et de l'administration*, 1894, p. 1272, ark:/12148/bpt6k9732740w/f1498.

On top of additional mining options, typographic emphasis can be used to structure documents that we would like to encode. Bold is not only the topic of an entry, but the

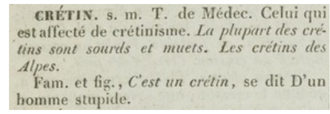
CRÉTIN. s. m. T. de Médec. Celui qui est affecté de crétinisme. *La plupart des crétins sont sourds et muets. Les crétins des Alpes.*
Fam. et fig., *C'est un crétin*, se dit D'un homme stupide.

Figure 2: *Dictionnaire de l'Académie française - sixième édition*, Paris: Firmin Didot, 1835, vol. 1 (A-K.), p. 450, ark:/12148/bpt6k12804289/f490.


24. La Rochefoucauld (François VI, duc de), le célèbre auteur des *Maximes*. L. aut. sig., à Mlle de Scudéry. Paris, 22 août... 2 gr. p. pl. et quart in-4. Cachets et soies. Très-belle lettre. 200 »

Figure 3: *Catalogues de lettres autographes, manuscrits, documents historiques, etc.*, Paris: Auguste Laverdet, N°1, avril 1856, p. 5, ark:/12148/bpt6k9687751c/f17.

most efficient way to locate its beginning, and italic the only solution to locate specific passages. Without the information that some characters are in bold font, it is impossible to differentiate the first line (*24. La Rochefoucauld*) from the third (*2 gr...*) in the fig. 3. Similarly, it is impossible to distinguish the definition from the examples in the fig. 2, or the biographical sub-part (in green) from the philological one (in blue) in the fig. 4, and reconstitute with precision the entry in XML for further exploration [7] .

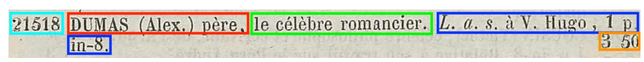
21518 DUMAS (Alex.) père, le célèbre romancier. *L. a. s.* à V. Hugo, 1 p. in-8. 3 50

Figure 4: *Librairie autographe Ancienne et autographe* Paris: Jacques Charavay, N°156, avril-mai 1867, p. 6

```
<item n="21518" xml:id="CAT_001_e21518">
   <num type="lot">21518</num>
   <name type="author">Dumas (Alex.) père,</name>
   <trait>
      <p>le célèbre romancier</p>
   </trait>
   <desc xml:id="CAT_001_e21518_d1">L.a.s. à V. Hugo,
    1 p., in-8.</desc>
   <measure commodity="currency" unit="FRF">3 50</measure>
</item>
```

Figure 5: TEI modelling of fig. 4.

3.2 Document Collection

The BIR database contains a subset of 285 pages come from exhibition [21] and sale catalogues [5], for which we have strong evidences that using bold and italic information increase the precision of the extraction and the structuring [6]. Carefully selected excerpts have been taken in the documentation of the *Artl@s* and the *Katabase* projects, *i.e.* mainly 19th French catalogues, which remains our primary target. Other types of documents (*e.g.* a 19th Latin lexicon) or in other languages than French (*e.g.* the São Art Paulo Biennale) have been added to diversify the data, as well as more recent catalogues (20^{th} and 21^{th} c.). Tab. 1 lists the number of pages considered for the BIR database, according to different categories.

19^{th} prints	220	Exhibition catalogues	143
20^{th} prints	67	Manuscript catalogues	111
French documents	258	Other documents	33
Foreign documents	29		

Table 1: Number of pages according to different categories

3.3 Ground Truth Annotation

The words of the original database were segmented and classified according their style with the ABBYY FineReader software [18]. The word segmentation was excellent but the style classification was far less successful. Each sentence was manually annotated with HTML tags to identify bold and italic words and the transcription was improved.

To conduct our experiments it was necessary to align the style information contained in the sentences with the locations of each word. Due to the manual correction of the transcription at the sentence level, the number of words contained in the sentence regularly differed from the number of words in the ABBYY FineReader output.

To solve this problem, a string edit distance algorithm was used to find the closest matching words, in order to correctly transfer the typographic emphasis. Finally, the ground truth file was manually inspected and corrected by adding, merging, and splitting word bounding boxes, and correcting the style information. Pages containing no text and those requiring too much manual editing were excluded. Table 2 lists some basic statistics of the resulting BIR database.

Documents	35
Pages	285
Words	88,019
- bold	2,106
- italic	5,745
- regular	80,168

Table 2: BIR database.

4 BASELINE RESULTS

Several experiments have been conducted on the BIR database to establish baseline results with methods from the current state of the art. Two tasks are considered, word detection and style classification.

4.1 Word Detection

Word detection aims at localizing words on a scanned page. We perform this task with a fully-convolutional object detection network, which takes a whole page image as input and provides bounding boxes around the detected words as output.

The YOLO [13] (You Only Look Once) model is considered, which has been a pioneering architecture for one-stage object detection and remained competitive over the years both in terms of speed and accuracy. The architecture consists of a convolutional *backbone*, followed by a feature pyramid *neck* that combines the extracted features at different scales, which are then processed by the *head* that computes two loss functions, one for bounding box regression and one for bounding box classification. In our case, three style classes are considered: bold, italic, and regular. Experiments are performed with the PyTorch-based YOLOv5 [8] version using the medium-sized YOLOv5m model with 21.4 million parameters pre-trained on the COCO [11] database.

Image preprocessing includes downscaling to a height of 1024 pixels, keeping the same aspect ratio, in order to fit the input images into the GPU memory. Furthermore, random scaling, translation, and rotation operations are applied during training to augment the number of training samples and improve the generalizability of the model.

4.2 Style Classification

Style classification aims at determining the style of an individual word image that has already been localized on the scanned page. The BIR database distinguishes bold, italic, and regular words.

Two fully-convolutional network architectures are used for establishing baseline results, namely MobileNetV2 [17] and Xception [1]. They were chosen with respect to their excellent classification performance on ImageNet [2] and to include both a smaller (MobileNetV2: 3.5 million parameters) and a larger (Xception: 22.9 million parameters) architecture. Experiments are performed with a Keras-based implementation and ImageNet pre-trained parameters. We drop the top layer, perform global average pooling to reduce the spatial dimensions, add a dense layer with 100 neurons, dropout, and rectified linear unit (ReLu) activation, and finish with a softmax classification layer. The architectures are illustrated in Figure 6.

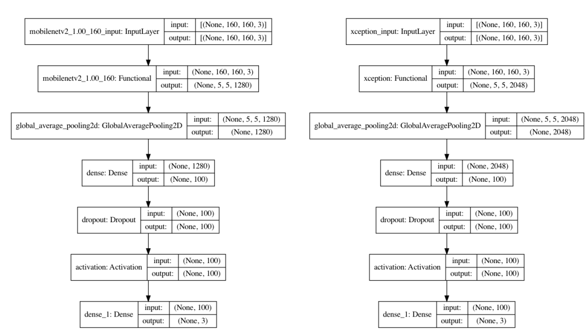

Figure 6: Network architecture for style classification using MobileNetV2 and Xception, respectively.

Image preprocessing includes resizing to (160, 160) pixels and normalizing the RGB inputs to (-1,1) to match the pre-training condition.

As shown in Table 2, there is a significant class imbalance among the three classes, with *regular* being the majority class and *bold* and *italic* being the minority classes. We consider two strategies to alleviate this class imbalance:

- Balancing via over-sampling and under-sampling. We define a fixed amount of n training samples for each class, under-sample the majority class, randomly selecting n samples, and over-sample the minority classes, repeating the samples until n is reached.
- Using class weights for the loss function. We weight the loss of the minority classes with factor $\frac{m}{m'}$ where m is the number of majority samples and m' is the number of minority samples, in order to give more emphasis to the minority samples during training.

4.3 Experimental Setup

Two datasets splits are defined for experimental evaluation.

- **TVT.** In this standard setup, the 285 pages are randomly distributed into three distinct sets for training the neural networks (50%), validation of metaparameters (25%), and testing of the final system performance (25%).
- **CV5.** In this cross-validation setup, the 35 documents are split into five distinct parts, each containing 7 documents. Five cross-validations are performed using three parts for training, one part for validation, and one part for testing, thus allowing to test the generalization capability of a trained network to an unseen document.

For word detection, the TVT setup is considered, training the YOLOv5m network with its default fine-tuning hyperparameters 100 epochs on the training set until convergence. This base model is then further fine-tuned on an augmented dataset, randomly scaling, translating, and rotating each page hundred times. The augmented model is trained 20 epochs until convergence, which takes about 1 hour on two Titan RTX cards.

The detection performance is evaluated with respect to precision (*#correct detections / #detections*), recall (*#correct detections / #words*), and F1-score (harmonic mean of precision and recall), considering a detection as correct if its intersection over union (IoU) with the ground truth box is larger than 50%.

For style classification, an initial experiment is conducted with the MobileNetV2 for the TVT setup. Afterwards, more detailed experiments are conducted for the CV5 setup, comparing MobileNetV2 with the larger Xception architecture and the two strategies for coping with class imbalance, *class balancing* and *class weights* respectively. The networks are trained with categorical cross-entropy loss and an adaptive Adam learning rate 50 epochs until convergence, which takes about 50 minutes on a Titan RTX card.

The best model is selected with respect to its accuracy on the validation set and then evaluated on the test set. Precision, recall, and F1-score are calculated for each style individually (bold, italic, regular), as well as their macro-average, i.e. the average of the evaluation measures without weighting with the number of samples from each class.

Table 3: Word detection results on the TVT test set.

	Precision	Recall	F1-Score
YOLOv5m	0.93	0.90	0.91

4.4 Results

Table 3 presents the detection performance achieved with the YOLO model on the TVT test set and Figure 7 illustrates some exemplary detection results. The object detection network is able to retrieve words with recall and precision over 90%, which demonstrates the feasibility of the approach but clearly leaves room for improvements. Typical errors include missing words (document in the middle) and errors due to ink bleed-through (document on the right).

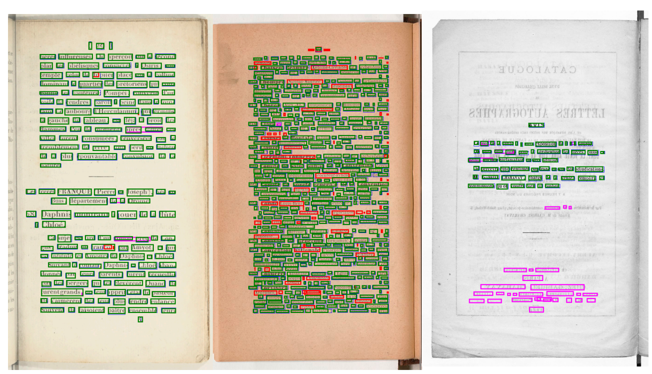

Figure 7: Word detection using YOLOv5m. Ground truth is marked in blue, detected words in green, missing words in red, and insertion errors in magenta.

Table 4 shows the style classification results with the MobileNetV2 model on the TVT test set. Overall, the achieved macro-average of 90% F1-Score is a promising result, especially when compared with ABBYY FineReader, which fails almost completely for style classification. While the retrieval of regular text is nearly perfect, only three out of four bold words are retrieved and the recall of italic words is also below 90%, illustrating the difficulty of the task even when training samples of the same document collection are available. Table 5 provides more details on the misclassifications. There is no confusion between bold and italic but both means of typographic emphasis are frequently confused with regular text.

Finally, the transfer of style classification to unseen document collections is assessed in the CV5 cross-validation setup. Table 6 shows the macro-average of the F1-score for MobileNetV2 and Xception, as well as for the class balancing

Table 4: Style classification results with MobileNetV2 on the TVT test set.

	Precision	Recall	F1-Score
Regular	0.98	0.99	0.99
Bold	0.93	0.76	0.84
Italic	0.92	0.85	0.89
Macro-Average	0.95	0.87	0.90

Table 5: Style confusion matrix for MobileNetV2 on the TVT test set.

	Regular	Bold	Italic
Regular	19,435	35	75
Bold	143	459	0
Italic	160	0	909

Table 6: Style classification results on the CV5 test sets in terms of macro-average of the F1-score. The best result is highlighted in bold font for each cross-validation.

	CV1	CV2	CV3	CV4	CV5
MobileNetV2	0.80	**0.77**	**0.76**	**0.71**	**0.60**
Xception	**0.93**	0.70	0.74	0.70	0.59
Class Balance	0.76	0.64	0.69	0.67	0.54
Class Weights	0.71	0.59	0.74	0.69	0.57

and class weight strategies when applied to MobileNetV2. In this scenario, the overall performance drops significantly, demonstrating how difficult it is to transfer knowledge on typewritten emphasis from one printed document collection to another. In four out of five cases the MobileNetV2 architecture achieves the best result. The basic strategies to cope with class imbalance are not able to improve the results.

It is also interesting to notice the relatively large variance of the results among the five cross-validations, highlighting the fact that each document collection has its specific properties and challenges.

4.5 Comparison with Human Performance

In a final experiment, we have investigated how the performance of the automatic system compares with human performance if the human is only presented with individual, cropped out word images without their context. For this purpose, we have randomly selected 1,000 words from the test set of the TVT setup, according to a non-uniform distribution: 325 regular words, 425 bold words, and 250 italic words. A human expert, who has created ground truth at page level, was asked to classify the individual word images by putting them into three distinct folders. The human was made aware of the fact that the word distribution is not uniform, without providing any hints about their real distribution, to avoid a bias when distributing the words into the different folders.

Table 7: Style classification results on the 1,000 words test set in terms of F1-score.

	Regular	Bold	Italic
MobileNetV2	0.82	0.86	0.92
Human expert	0.85	0.87	0.95

Table 8: Style confusion matrix for MobileNetV2 on the 1,000 words test set.

	Regular	Bold	Italic
Regular	323	0	2
Bold	103	322	0
Italic	35	0	215

Table 9: Style confusion matrix for the human expert on the 1,000 words test set (excluding 27 non-annotated samples).

	Regular	Bold	Italic
Regular	303	9	4
Bold	93	331	0
Italic	17	0	216

To our surprise, the human expert refused to classify 27 images because they were near-empty or of low quality. These samples have been excluded from the evaluation of the human performance.

Tables 7-9 demonstrate that the automatic system reaches near-human performance for isolated word classification. While it is possible that neural networks may outperform humans for this task in the future, it also highlights that taking contextual information from the whole document page into account is expected to be a key element to improve the performance in the future.

5 CONCLUSION

The BIR database and its baseline results for word detection and style classification demonstrate the difficulty of identifying typewritten emphasis in historical prints. Given the importance of style emphasis for document analysis and understanding, we hope that our research database can contribute to the development of novel methods in this field. A promising line of research is related to building combined word localization, recognition, and style classification models. Furthermore, data augmentation, synthesis, and balancing strategies are expected to play a central role when developing more robust systems that generalize well to unseen document collections. Going beyond isolated word classification by including more contextual information is expected to be of fundamental importance for improving the performance.

ACKNOWLEDGMENTS

This work benefited from funding from the Center of Excellence Jean Monnet IMAGO (École normale supérieure) and the support of Prof. Béatrice Joyeux-Prunel.

REFERENCES

[1] François Chollet. 2017. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1251–1258.

[2] Jia Deng, Wei Dong, Richard Socher, Li-Jia 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*. Ieee, 248–255.

[3] Isabella di Lenardo, Raphaël Barman, Albane Descombes, and Frédéric Kaplan. 2019. Repopulating Paris: massive extraction of 4 Million addresses from city directories between 1839 and 1922. In *Digital Humanities 2019 Conference Abstracts*. Alliance of Digital Humanities Organizations (ADHO), Utrecht, The Netherlands. https://dev.clariah.nl/files/dh2019/boa/0878.html

[4] Andreas Fischer, Marcus Liwicki, and Rolf Ingold (Eds.). 2020. *Handwritten Historical Document Analysis, Recognition, and Retrieval — State of the Art and Future Trends*. World Scientific.

[5] Simon Gabay, Ljudmila Petkovic, Alexandre Bartz, Matthias Gille Levenson, and Lucie Rondeau Du Noyer. 2021. Katabase: À la recherche des manuscrits vendus. In *Humanistica 2021*. Humanistica, Rennes, France. https://hal.archives-ouvertes.fr/hal-03066108

[6] Simon Gabay, Lucie Rondeau Du Noyer, Matthias Gille Levenson, Ljudmila Petkovic, and Alexandre Bartz. 2020. Quantifying the Unknown: How many manuscripts of the marquise de Sévigné still exist?. In *Digital Humanities DH2020 (DH2020 Book of Abstracts)*. ADHO, Ottawa, Canada. https://hal.archives-ouvertes.fr/hal-02898929

[7] Simon Gabay, Lucie Rondeau Du Noyer, and Mohamed Khemakhem. 2020. Selling autograph manuscripts in 19th c. Paris: digitising the Revue des Autographes. In *Atti del IX Convegno Annuale AIUCD. La svolta inevitabile: sfide e prospettive per l'Informatica Umanistica (Quaderni di Umanistica Digitale)*. Associazione per l'Informatica Umanistica e la Cultura Digitale, Milan, Italy, 113–118.

[8] Glenn Jocher et al. 2021. ultralytics/yolov5: v4.0 - nn.SiLU() activations, Weights & Biases logging, PyTorch Hub integration. https://doi.org/10.5281/ZENODO.4418161

[9] Béatrice Joyeux-Prunel and Olivier Marcel. 2016. Exhibition Catalogues in the Globalization of Art. A Source for Social and Spatial Art History. *Artl@s Bulletin* 4, 2 (2016), 80–104. https://docs.lib.purdue.edu/artlas/vol4/iss2/8

[10] B. Kiessling, R. Tissot, P. Stokes, and D. Stokl Ben Ezra. 2019. eScriptorium: An Open Source Platform for Historical Document Analysis. In *2019 International Conference on Document Analysis and Recognition Workshops (ICDARW)*, Vol. 2. IEEE Computer Society, Los Alamitos, CA, USA, 19–19. https://doi.org/10.1109/ICDARW.2019.10032

[11] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. 2014. Microsoft COCO: Common Objects in Context. In *Proc. 13th European Conf. on Computer Vision (ECCV)*. 740–755.

[12] Anguelos Nicolaou, Fouad Slimane, Volker Maergner, and Marcus Liwicki. 2014. Local Binary Patterns for Arabic Optical Font Recognition. In *2014 11th IAPR International Workshop on Document Analysis Systems*. IEEE, Tours, France, 76–80. https://doi.org/10.1109/DAS.2014.71

[13] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. 2016. You only look once: Unified, real-time object detection. In *Proc. Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*. 779–788.

[14] Christian Reul, Dennis Christ, Alexander Hartelt, Nico Balbach, Maximilian Wehner, Uwe Springmann, Christoph Wick, Christine Grundig, Andreas Büttner, and Frank Puppe. 2019. OCR4all—An Open-Source Tool Providing a (Semi-)Automatic OCR Workflow for Historical Printings. *Applied Sciences* 9, 22 (2019), 4853. https://doi.org/10.3390/app9224853

[15] Christian Reul, Sebastian Göttel, Uwe Springmann, Christoph Wick, Kay-Michael Würzner, and Frank Puppe. 2019. Automatic Semantic Text Tagging on Historical Lexica by Combining OCR and Typography Classification: A Case Study on Daniel Sander's Wörterbuch der Deutschen Sprache. In *Proceedings of the 3rd International Conference on Digital Access to Textual Cultural Heritage (DATeCH2019)*. Association for Computing Machinery, Brussels, Belgium, 33–38. https://doi.org/10.1145/3322905.3322910

[16] Ana Salgado and Rute Costa. 2020. O projeto Edição Digital dos Vocabulários da Academia das Ciências:o VOLP-1940. *Revista da Associação Portuguesa de Linguística* 7 (2020), 275–294. https://doi.org/10.5281/zenodo.4453139

[17] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. 2018. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4510–4520.

[18] Andrey Shapenko, Vladimir Korovkin, and Benoit Leleux. 2018. ABBYY: the digitization of language and text. *Emerald Emerging Markets Case Studies* (2018).

[19] Fouad Slimane, Rolf Ingold, and Jean Hennebert. 2017. ICDAR2017 Competition on Multi-Font and Multi-Size Digitally Represented Arabic Text. In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, Vol. 01. IEEE, Kyoto, Japan, 1466–1472. https://doi.org/10.1109/ICDAR.2017.239 ISSN: 2379-2140.

[20] Dapeng Tao, Xu Lin, Lianwen Jin, and Xuelong Li. 2016. Principal Component 2-D Long Short-Term Memory for Font Recognition on Single Chinese Characters. *IEEE Transactions on Cybernetics* 46, 3 (mar 2016), 756–765. https://doi.org/10.1109/TCYB.2015.2414920 Conference Name: IEEE Transactions on Cybernetics.

[21] Barbara Topalov, Simon Gabay, Caroline Corbières, Laurent Romary, Béatrice Joyeux-Prunel, and Lucie Rondeau du Noyer. 2021. Automating Artl@s - extracting data from exhibition catalogues. In *EADH'21 Book of abstracts*. European Association for Digital Humanities (EADH), Krasnoyarsk, Russia.

[22] Kurban Ubul, Gulzira Tursun, Alimjan Aysa, Donato Impedovo, Giuseppe Pirlo, and Tuergen Yibulayin. 2017. Script Identification of Multi-Script Documents: A Survey. *IEEE Access* 5 (2017), 6546–6559. https://doi.org/10.1109/ACCESS.2017.2689159

[23] Adnan Ul-Hasan, Muhammad Zeshan Afzal, Faisal Shafait, Marcus Liwicki, and Thomas M. Breuel. 2015. A sequence learning approach for multiple script identification. In *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, Tunis, Tunisia, 1046–1050. https://doi.org/10.1109/ICDAR.2015.7333921

[24] Christoph Wick, Christian Reul, and Frank Puppe. 2018. Calamari - A High-Performance Tensorflow-based Deep Learning Package for Optical Character Recognition. *CoRR* abs/1807.02004 (2018). http://dblp.uni-trier.de/db/journals/corr/corr1807.html#abs-1807-02004

[25] Yong Zhu, Tieniu Tan, and Yunhong Wang. 2001. Font recognition based on global texture analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23, 10 (2001), 1192–1200. https://doi.org/10.1109/34.954608 Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.

[26] Abdelwahab Zramdini and Rolf Ingold. 1998. Optical font recognition using typographical features. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20, 8 (aug 1998), 877–882. https://doi.org/10.1109/34.709616 Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.