VML-HP: Hebrew paleography dataset

Ahmad Droby[1], Daria Vasyutinsky Shapira[1], Irina Rabaev[2], Berat Kurar Barakat[1], and Jihad El-Sana[1]

[1]Ben-Gurion University of the Negev , [2]Shamoon College of Engineering

Introduction

We present our research on automatic classification of Hebrew manuscripts into fourteen categories according to the script types and graphical modes. To train a deep neural network, we compiled a dataset of manuscripts where all of these categories are present.

The margins between categories of writing styles are sometimes fuzzy and overlap on visual appearances level. To categorize the document, paleographers examine the visual appearance of the handwriting as well as the codicological data, e.g., the media on which the document was written. Since we are working with digital images only, we are unable to utilize the codicological data. We hypothesize that hard-labeling may not be the ideal way for training the deep-learning model to recognize the writing category. Therefore, for each page image, we decided to add an additional level of llabelling - a soft label. The soft label is a label vector, where each element indicates the similarity of the document's script to a specific script type or mode.

An expert in Hebrew paleography manually annotated the soft label for each document.

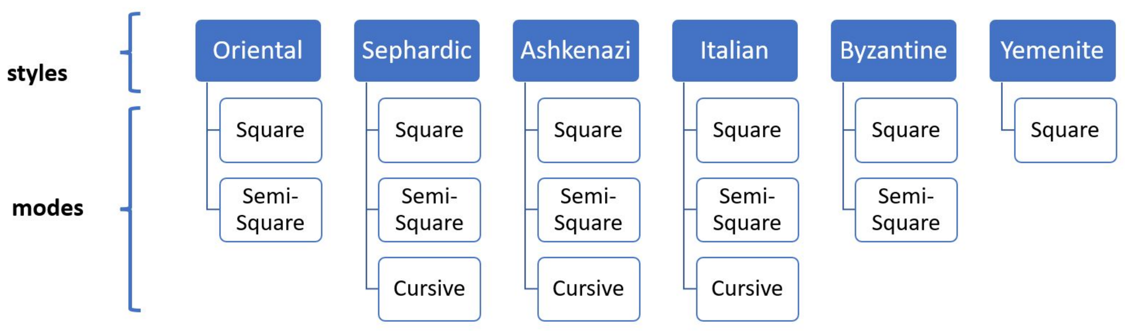

Figure 1. Hebrew script styles and modes; not all regional styles have cursive or semi-square mode.

VML-HP-ext

we present an extended VML-HP-ext (Visual Media Lab - Hebrew Paleography Extended) dataset.

Compared to the first version, the extended dataset includes sample pages from three times more manuscripts. Every manuscript was carefully selected by our team's paleographer. The majority of the manuscripts used in this dataset are kept in the National Library of Israel, the British Library, and the Bibliotheque nationale de France. Almost all manuscripts in the Oriental square script belong to the National library of Russia (we used b/w microfilms from the collection of the Institute for Microfilmed Hebrew Manuscripts at the National Library of Israel).

We only included pages with one script type and one script mode per page. For example, Sephardic square only, and not main text in Sephardic square and comments in Sephardic cursive.

Table 1: Summary of the extended VML-HP-ext dataset. Some scripts do not have semi-cursive or cursive modes. Mss=manuscripts, pp = pages.

Type	Mode: Square		Semi-Square		Cursive	
	#Mss	#pp	#Mss	#pp	#Mss	#pp
Ashkenazi	14	56	12	48	12	48
Byzantine	7	49	12	48	-	-
Italian	5	50	11	44	5	50
Oriental	15	45	11	44	-	-
Sephardic	15	45	16	48	12	48
Yemenite	24	92	-	-	-	-

Hard-label classifiication

We trained and evaluated several architectures on the extended dataset.

The models were trained until convergence using 50K patches extracted from pages in the train set. The model was trained using binary cross entropy loss function. The patches were extracted using the patch generation method proposed in our previous work[1], which extracts patches with uniform text scale and on average 5 lines in each patch.

Table 3: Evaluation results of several classification models on blind test set of the extended dataset.

Model	Avg. Precision	Avg. Recall	Avg. F1-score	Accuracy
DenseNet	58	53	53	53
AlexNet	55	52	52	51
VGG19	60	57	56	56
ResNet50	**63**	**60**	**59**	**60**
SqueezeNet	58	55	54	55

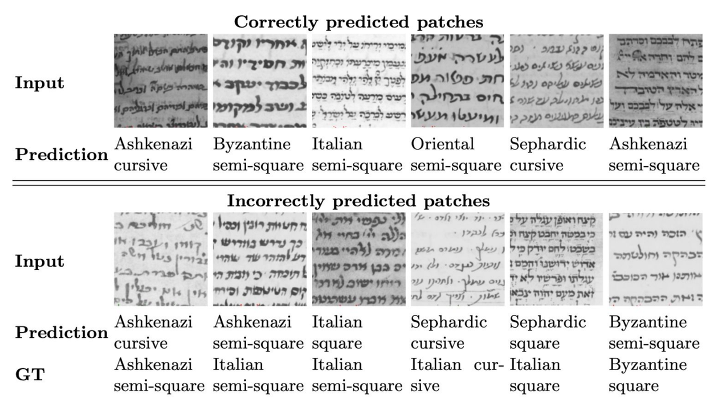

Fig. 2: Sample results from the ResNet50 classification model

Soft-Label regression

In a soft-labeling scheme, we label each manuscript using a vector of size eight. The first six elements of the vector express the degree of similarity of the manuscript to belong to certain regional type (Ashkenazi, Italian, Sephardic, Oriental, Byzantine and Yemenite) and the last two elements are the degrees of similarity to certain graphical mode, square and cursive (similar values for both square and cursive indicate the semi-square mode). Similar to the previous experiment, we extracted 50K patches and assign each patch a vector of probability values corresponding to a regional and graphical types.

We trained a regression model with a ResNet50 backbone on the mentioned 50K patches with mean squared error loss function. The model was trained until convergence, which happened after 10 epochs.

The trained regression model achieved RMSE of about 0.24. Although, this might give us an indication that the model give good results (as can be seen in Figure 3, it is not very meaningful and does not show how this model compare against other classification methods. Therefore, arose a need to convert the predicted soft-label to hard-labels. Next, we explore two different conversion methods:

Table 6: Evaluation results of the regression model with maximum score class conversion method.

Label	Square P	Square R	Square F1	Semi-square P	Semi-square R	Semi-square F1	Cursive P	Cursive R	Cursive F1
Ashkenazi	1.00	0.79	0.88	0.32	0.40	0.36	0.29	0.09	0.14
Byzantine	0.27	0.06	0.10	0.24	0.81	0.37		-	
Italian	0.00	0.00	0.00	0.22	0.66	0.33	0.18	0.18	0.18
Oriental	0.88	0.61	0.72	0.17	0.07	0.09		-	
Sephardic	0.98	0.36	0.52	0.32	0.64	0.43	0.99	0.15	0.25
Yemenite	0.83	0.31	0.45		-			-	
Average	P 0.50	R 0.37	F1 0.34		**Accuracy**	0.37			

P: precision, R: recall, F1: F1-score

Maximum score class assignment

The label is determined by taking the regional style and graphical mode with the maximum score unless both, the square and cursive, scores are under a predefined threshold T (we set T = 0.3), in which case the graphical mode is determined to be as semi-square.

Nearest neighbour label conversion

This approach utilizes the soft and hard labels in the training set. It calculates the distances between the predicted labels and the soft-labels in the training set and converts each predicted soft-label to the nearest hard-label in the train set. Figure 5 presets sample results of this conversion.

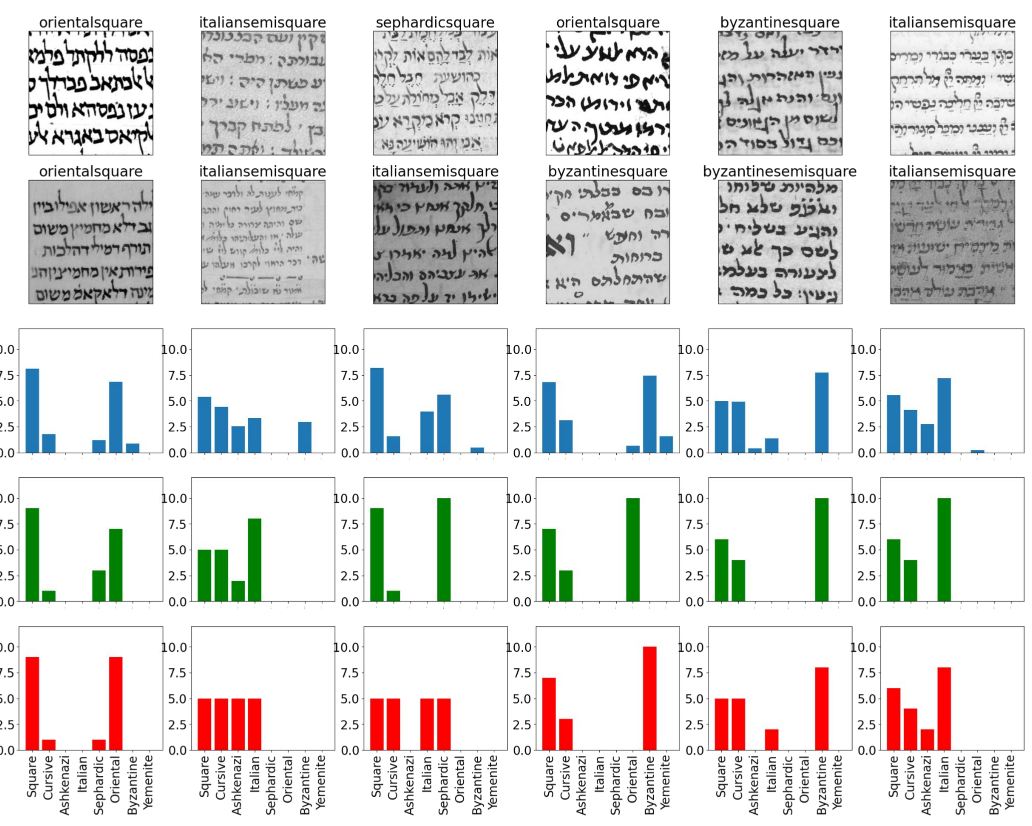

Figure 5. Sample results of regression model with the nearest neighbor label conversion method. Top row: input patch with its ground-truth label. Second row: The nearest neighbour of the input patch. Third row: The predicted label of the input patch. Fourth row: The ground-truth soft-label of the input patch. Bottom row: The ground-truth soft-label of the nearest neighbour patch.

Table 7: Evaluation results of the regression model with the nearest neighbor label conversion method.

Label	Square P	Square R	Square F1	Semi-square P	Semi-square R	Semi-square F1	Cursive P	Cursive R	Cursive F1
Ashkenazi	0.98	0.57	0.72	0.43	0.67	0.52	0.50	0.01	0.03
Byzantine	0.03	0.01	0.01	0.25	0.87	0.38		-	
Italian	0.00	0.00	0.00	0.23	0.64	0.33	0.39	0.21	0.27
Oriental	0.99	0.65	0.79	0.29	0.06	0.10		-	
Sephardic	0.99	0.37	0.54	0.49	0.73	0.58	1.00	0.01	0.02
Yemenite	0.88	0.63	0.73		-			-	
Average	P 0.53	R 0.40	F1 0.37		**Accuracy**	0.40			

P: precision, R: recall, F1: F1-score

Conclusions

We trained and evaluated several classification models on the hard-labeling configuration. ResNet50 topped the list with an accuracy of 60%. In addition, we experimented with soft-labeling, training a regression model to predict the similarity values of each image to each geographical and graphical type. Since such a model cannot be directly compared with regular hard-label classification, we proposed and evaluated two methods that convert soft labels to hard labels. We conclude that while the soft-labeling provides more information about the script style, e.g., how square or cursive it is, using the regression model with the conversion methods does not reach the accuracy of the models trained using hard-labeling.

References

1. Droby, A., Kurar Barakat, B., Vasyutinsky Shapira, D., Rabaev, I., El-Sana, J.: Vml-hp: Hebrew paleography dataset. In: International Conference on Document Analysis and Recognition. pp. 205–220 (2021)