The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-2/W9-2025
ISPRS Intl. Workshop “Photogrammetric and computer vision techniques for environmental and infraStructure monitoring, Biometrics and Biomedicine”
PSBB25 , 9–11 June 2025, Moscow, Russia

An Approach to Searching for Hieroglyphs in Chinese Manuscript Archives Based on Morphological Analysis

Mingchuan Xu[1], Leonid Mestetskiy[2]

[1] Faculty of Computational Mathematics and Cybernetics, Moscow State University - Moscow, Russia - xu.mingchuan@cs.msu.ru
[2] Faculty of Computational Mathematics and Cybernetics, Moscow State University - Moscow, Russia - mestlm@mail.ru

Keywords: Hieroglyph, Metagraph, Continuous Skeleton, Graph Similarity, Efficient Recognition

Abstract

Searching for Chinese characters in large archives of handwritten documents using a single sample in a query is a so-called single recognition task. The total number of different characters in the archives of ancient Chinese handwritten texts is estimated at several tens of thousands, which creates substantial obstacles for machine-learning based methods due to complexity of constructing training datasets. This article proposes an effective method for recognizing and searching for characters based on a direct comparison of the form of the query character with the characters from the file. We proposed a method for constructing a hieroglyph model in the form of a planar geometric graph. A measure is proposed for assessing the similarity and difference of generally non-isomorphic geometric graphs of hieroglyphs by solving assignment problem and a recognition method based on this measure. Computational experiments with large databases of handwritten hieroglyphs confirmed the effectiveness of the proposed approach by achieving comparable results to deep-learning methods, while can be implemented with conventional modern computers without accelerators. In addition, the method is fully interpretable, which is important for understanding and adjusting the recognition process, as well as for further development of the proposed approach.

1. Introduction

Optical character recognition is a technology that converts digital images of handwritten or printed text into machine text for further storage and processing. In recent years, advanced methods (Du et al., 2020)(Li et al., 2023) for working with images of printed texts have been developed, but only a small part of the research is devoted to the recognition of texts from handwritten archives (Melnyk et al., 2019)(Xu et al., 2022)(Chen et al., 2021). At the same time, ancient hieroglyphs in handwritten texts have great research value as a humanitarian heritage that records historical culture, language, and the development of science. Information search and query navigation in a huge number of ancient documents always remain one of the central problems in the study of ancient texts. However, the application of modern recognition methods using machine learning encounters great difficulties in preparing training data for ancient Chinese characters. Firstly, a very large amount of data is required for training, since the number of symbols of the ancient Chinese script is very large. Secondly, the variability of writing styles and scripts seriously complicates the recognition process. Thirdly, the labor costs of annotating symbols are very high even for a specialist in ancient hieroglyphs. Recognizing symbols and extracting key information from multiple pages of ancient manuscripts can take several hours or even days.

Various methods have been developed to solve the problem of recognizing ancient hieroglyphs, particularly for the one-shot learning scenario. Methods using deep learning demonstrate good results and high recognition speed (Liu et al., 2022)(Li et al., 2020). However, they require a huge amount of training data to achieve sufficient recognition accuracy. Obtaining this data is very difficult in real-world conditions, since many hieroglyphs are very rare in most ancient documents and finding them for inclusion in the training set is a big problem. In addition, both training and practical application require modern accelerators, the cost of their use is relatively high. Modern neural networks do not provide the ability to interpret the underlying recognition process. But understanding the motives of the solution formed in the network is very important for researchers, since it helps to better understand the structure of the symbols and ways to improve the recognition algorithm.

Thus, the problem of searching for symbols in documents that match a given query symbol, but without long and labor-intensive preliminary training, is relevant. In pattern recognition, the problem of using a single sample for recognition is called one-shot learning. In this paper, we propose an efficient method for one-shot recognition of ancient Chinese characters based on constructing a measure of their shape similarity according to topological and geometric criteria. This method maintains high recognition accuracy compared to neural network-based methods, but has the advantages of no need for training, low resource requirements, and good interpretability.

2. Proposed Approach

The proposed approach to constructing a comparison metric includes the following elements:

1. constructing planar geometric graphs that describe the topological and geometric structure of hieroglyphs;
2. generating informative feature descriptions of the graph structure;
3. a measure for assessing the similarity and difference of feature descriptions.

The proposed solution is based on constructing a morphological model of a hieroglyph in the form of a geometric graph. This

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLVIII-2-W9-2025-283-2025 | © Author(s) 2025. CC BY 4.0 License.

graph is called a hieroglyph metagraph. The metagraph is constructed based on the transformation of the original digital image of the hieroglyph. To determine the similarity and difference of hieroglyphs, a measure of comparison of metagraphs by topological and geometric features is proposed. This measure is used to solve the problem of searching for hieroglyphs for navigation in large manuscript archives.

Construction of a geometric model of a hieroglyph is based on the methods of computational geometry, which ensures the mathematical correctness of the model, the stability of computational procedures and the high computational efficiency of algorithms for comparing and searching for hieroglyphs. Computational experiments to test the proposed solution are conducted on a large-scale dataset of ancient Chinese characters. The experiments demonstrate the effectiveness and efficiency of the proposed approach, as well as good interpretability of the results. A digital image of a text consisting of hieroglyphs looks like a binary image in which the characters are represented by black dots on a white background. In the example of Fig. 1, some hieroglyphs appear several times. However, they do not match when superimposed on each other. The task of comparison is to determine their similarity, despite the discrepancy between their pixel matrices.

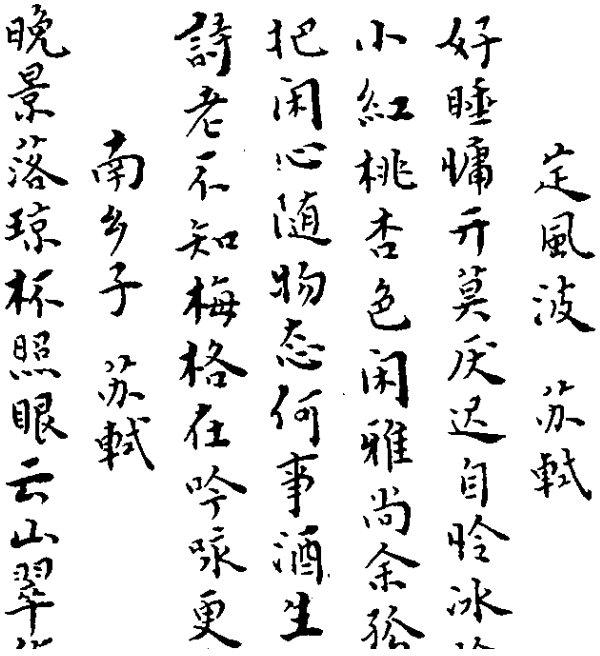

Figure 1. A sample of scanned binarized Chinese historical handwritten manuscript.

2.1 Construction of Hieroglyph Meta-graph

The construction of the hieroglyph metagraph is based on the use of the medial representation of the image shape (Siddiqi and Pizer, 2008). The medial representation consists of two components: the medial axis and the radial function. The medial axis of the figure, also called the skeleton, is the set of centers of all circles inscribed in the figure. The radial function is specified at each point of the skeleton and is equal to the radius of the maximum empty circle with the center at this point. The advantage of such a model lies, first of all, in its strict mathematical definition, high accuracy, and the computational efficiency of the algorithms.

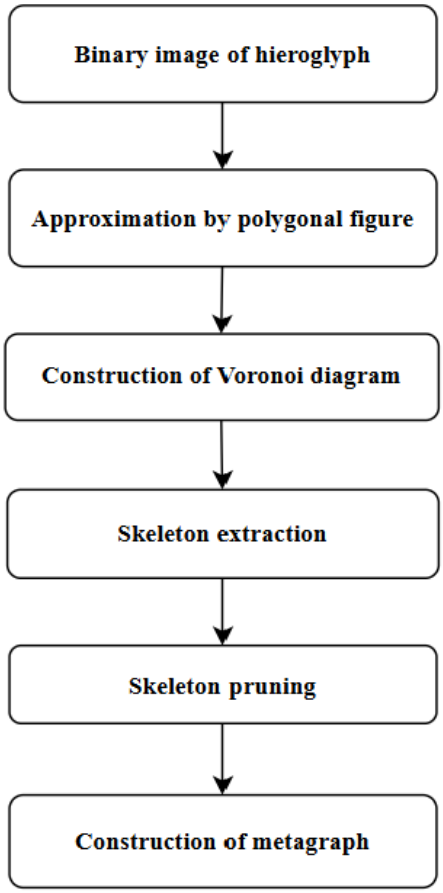

Figure 2. Sequence of operations in constructing a hieroglyph metagraph based on the transformation of a digital binary image

The general sequence of actions in constructing a metagraph is shown in Fig. 3. It includes the following steps:

1. Approximation of the binary image of a hieroglyph by a polygonal figure.
2. Construction of the internal Voronoi diagram and the skeleton of the polygonal figure.
3. Construction of the metagraph of the hieroglyph as a subgraph of the skeleton.

The choice of polygonal figures for approximating the shape of hieroglyphs is due to the possibility of using efficient algorithms of computational geometry to construct their medial representation based on Voronoi diagrams.

The approximating polygonal figure can consist of several connectivity components, each of which is a polygon with polygonal holes (Fig. 3d). Approximation of a hieroglyph by a polygonal figure is shown in Fig. 3a-3d. For the original raster image of a hieroglyph (Fig. 3b), it is necessary to search, trace and approximate all the boundary contours of the figure (Mestetskiy and Semenov, 2008).

The constructed polygons for all the boundary contours form an approximating polygonal figure for the image of the hieroglyph (Fig. 3d).

The next step of the proposed approach is to construct an internal Voronoi diagram (DV) of the resulting polygonal figure (Fig. 3e). To do this, the boundary polygons of the figure are divided into subsets called sites. Sites are all vertices and all sides of the boundary polygons. The internal DV of a polygonal figure is a partition of the set of points of the figure into loci - subsets of points closest to one of the sites of the figure. To construct the internal DV, we use the algorithm (Mestetskiy and Koptelov, 2024).

The loci of the DV intersect at the points of their boundaries. In Fig. 3e, these intersections are shown by red lines. This set of lines is called the edges of the DV. It can be considered as a planar geometric graph. The skeleton is a subgraph of the DV, it includes almost all the edges of the DV, except for the edges incident to the loci of the concave vertices of the polygonal figure. Thus, the skeleton can be obtained based on the cutting off of these edges of the DV (Fig. 3f).

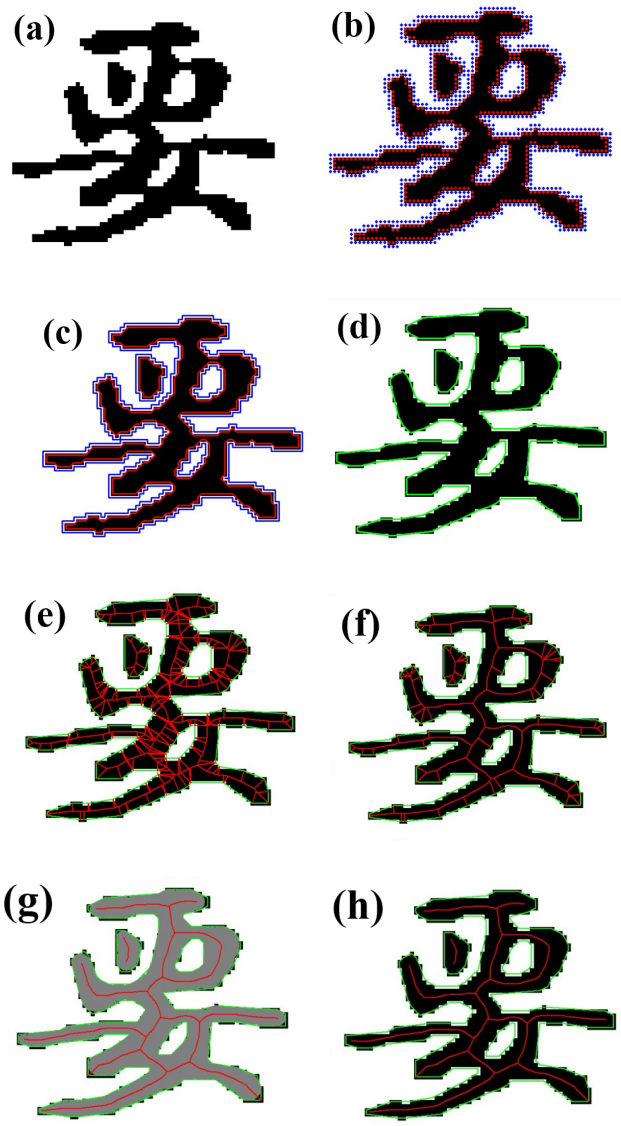

Figure 3. An example of constructing a hieroglyph metagraph for a digital image. From top left to bottom right the figures are enumerated as 3a-3h.

Next, a metagraph of the hieroglyph is constructed, which is a subgraph of the skeleton. The process of extracting a subgraph is reduced to successively cutting off part of the edges of the skeleton. This process is called pruning. The use of the proposed continuous model allows us to obtain a mathematically rigorous pruning criterion based on the concept of the silhouette of a skeletal subgraph (Mestetskiy and Koptelov, 2024). The silhouette of a skeletal subgraph is the union of all inscribed circles of a figure with centers at the vertices and edges of the subgraph. The general idea of pruning is to select the minimal subgraph in the skeleton whose silhouette differs from the polygonal figure by no more than a given threshold value in the Hausdorff metric. It is necessary to sequentially cut off those terminal edges of the subgraph, upon removal of which the Hausdorff distance between the subgraph silhouette and the polygonal figure does not exceed a given threshold. The pixel size can be chosen as a threshold value for images of hieroglyphs. An example of a subgraph obtained as a result of pruning is shown in Fig. 3f-3g.

The metagraph of a hieroglyph is a planar graph whose vertices are a subset of the vertices of the DV, and whose edges consist of chains of DV edges. The skeleton subgraph obtained as a result of the pruning contains vertices of degrees 1, 2, and 3. The set of metagraph vertices is formed based on two rules.

1. All vertices of degrees 1 and 3 of this subgraph are considered metagraph vertices
2. Some subset of vertices of degree 2 are also declared metagraph vertices to account for data on the shape of metagraph edges. These are vertices of maximum curvature in edge chains connecting vertices of degrees 1 and 3.

The edges of the metagraph are all the skeleton chains connecting the selected vertices. The shape of the chain is determined by the position of the intermediate vertices of degree 2.

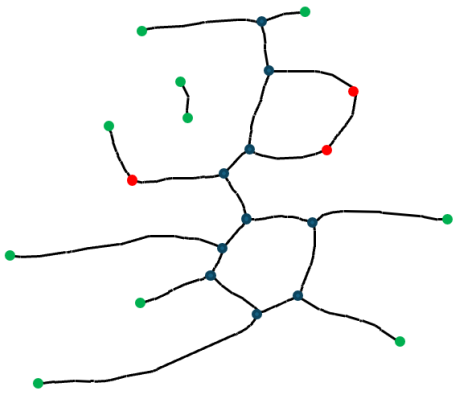

Figure 4. Metagraph vertices: first (green), second (red) and third (blue) degrees

2.2 Metric of Character Similarity

The measure of similarity and difference of hieroglyphs is built on the basis of comparison of their metagraphs. From this point of view, we see one of the advantages of the proposed method: simple but effective construction of features, compared to methods based on deep learning, which require a relatively large amount of resources for application. In addition, the entire process is transparent and interpretable, which provides more opportunities for manipulation of feature generation.

The difficulty of comparison is that metagraphs of images of the same hieroglyph often turn out to be non-isomorphic graphs. This occurs due to natural differences in the outline of handwritten characters, deformation of ancient manuscripts due to long-term storage, as well as distortions when scanning documents. Deformations and distortions are expressed in the fact that some strokes stick together on the obtained digital images, or gaps appear in them. Therefore, the topological criterion of metagraph similarity is insufficient for comparing hieroglyphs. The proposed measure of metagraph similarity is based on the geometric criterion of the proximity of metagraph vertices.

To assess the similarity of two metagraphs, their normalization is performed in the first place, which consists of adjusting the metagraphs by size and position. Formally, normalization is performed by shifting and scaling the metagraphs to superimpose them on each other. After that, the similarity is assessed based on the selection of the best matching of the metagraph vertices by the distances between the vertices. The choice of normalization method could drastically influence algorithm performence. Various methods of normalization could be implemented, including scaling vertices coordinates by their centroid, while in this article we stretch vertices horizontally and vertically by mapping coordinates to square frame based on maximum and minimum values. In this way, we reduce the influence of different frame sizes of metagraphs when assessing their similarity. Example of our normalization procedure is shown in Fig.5.

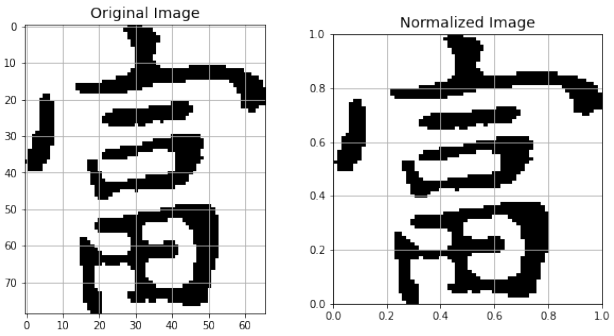

Figure 5. Normalization: frame character images into $[0,1] \times [0,1]$ square.

Two metagraphs are compared by superimposing them on each other and establishing a correspondence between the metagraph vertices. There are several requirements for the correspondence. First, for each vertex of one metagraph, at least one corresponding vertex in the other metagraph is selected. Second, it is preferable that the corresponding vertices have the same degrees. The task is to select from all admissible variants of vertex correspondence one in which the sum of the distances between the corresponding vertices is the smallest and at the same time the degrees of these vertices are, if possible, the same. Such a choice is implemented based on the optimization model of assignments.

Let’s denote:

- U, V - metagraphs defined by their nodes $\{u_1, \ldots, u_m\}$ and $\{v_1, \ldots, v_n\}$,
- $D(U,V)$ - the distance between metagraphs U and V,
- p_i, q_j - degrees of nodes u_i and v_j,
- λ_{ij} - distance between nodes u_i and v_j,
- h_{ij} - cost of difference in degree of nodes u_i and v_j,

$$h_{ij} = \begin{cases} 1, & \text{if } p_i = q_j \\ H \cdot |p_i - q_j|, & \text{if } p_i \neq q_j \end{cases}$$

- H - penalty when choosing a match between a pair of nodes of different degrees,
- x_{ij} - matrix of matching nodes of metagraphs,

$$x_{ij} = \begin{cases} 1, & \text{if vertices } u_i \text{ and } v_j \text{ correspond to each other} \\ 0, & \text{otherwise} \end{cases}$$

The selection of the best match of metagraph vertices is described by the optimization model of linear programming - the assignment problem (Kunwar and Sapkota, 2022):

$$\begin{aligned} &D(U,V) = \min_{x_{ij}} \frac{1}{m+n} \sum_{i=1}^{m} \sum_{j=1}^{n} \lambda_{ij} h_{ij} x_{ij}, \\ &s.t. \sum_{i=1}^{m} x_{ij} \geq 1, \sum_{j=1}^{n} x_{ij} \geq 1 \end{aligned} \tag{1}$$

Based on the solution of this optimization problem, the measure of difference between two metagraphs is estimated. The smaller the value of $D(U,V)$, the less the metagraphs differ, and the higher the similarity between the hieroglyphs. The advantage of the proposed model is that it reduces the problem to constructing the best matching of a finite small set of metagraph vertices. In this case, any metagraphs can be formally compared, both of similar and completely different hieroglyphs. The measure of difference can always be obtained based on the solution of the assignment problem.

2.3 Efficient Recognition: Pipeline

The general procedure for single recognition of ancient Chinese characters is based on the above-described methods for obtaining character metagraphs and calculating the measure of difference between these metagraphs. The input of the procedure is a query with a character image, as well as a file representing a large set of character images. The output is a ranked list of characters from the file, ordered by increasing measure of their difference with the query character. It is assumed that this list of candidates is then placed on the desk of a researcher of ancient handwritten texts. Considering that a file can contain thousands of characters, such a recognition procedure can reduce the time it takes a researcher to perform routine search operations in ancient Chinese handwritten archives by several orders of magnitude.

To perform mass search queries, the construction of character metagraphs from a file must be performed in advance during preprocessing. Preprocessing includes preliminary segmentation and binarization of images, after which the metagraph construction operations described above in Section 3 are performed. The resulting metagraphs are ready for use in search queries.

Each search query includes the construction of a metagraph for the image of the query hieroglyph and its subsequent comparison with the metagraphs of the hieroglyphs from the file. The result in the form of a ranked list of hieroglyphs in ascending

order of difference with the query is formed after a full cycle of comparing the query with all hieroglyphs in the file.

3. Experiment results

To compare the qualities of different algorithms with our developed method, computational experiments were conducted. According to (Liu et al., 2022), we solve single recognition problems in a sample of images with a total of 550 samples per problem, based on which we calculate the classification accuracy. Each sample consists of m samples, of which one is positive and $m-1$ are negative. In addition, one query example is attached to each sample, with which all samples in the sample are compared. If the positive sample is in the first place of the ranked list in ascending order of the calculated measure values, then the classification is considered successful for this sample. We chose $m = 5$ and $m = 20$ according to (Liu et al., 2022). The final classification accuracy is calculated as follows:

$$\text{Accuracy} = \frac{\text{Number of samples with successful classification}}{\text{Total number of samples}}$$

Ancient Handwritten Character Database (CASIA-AHCDB) (Xu et al., 2019) is a dataset for character recognition research. The dataset contains over 2.2 million annotated character samples from 10,350 classes. The character samples are collected from over 12,000 pages of annotated Chinese ancient handwritten documents. According to different document sources, the database is mainly divided into two sub-databases: Complete Library in Four Sections (Style 1) and Ancient Buddhist Scripts (Style 2). To ensure fair comparison with existing approaches, we test our results on the test set of Style 1 Basic Category.

All experiments are conducted on a modern 8x-core laptop with 11th generation Intel(R) Core(TM) i5-11300H @ 3.10 GHz, 16 GB RAM. This configuration compares favorably with advanced methods with neural networks using multiple GPU accelerators. The experimental results are presented in Table 1. It is noticeable that although our proposed method is an algorithmic approach by design, the classification accuracy reaches results comparable to or even exceeds the latest advanced methods when using multiple configurations. This confirms the effectiveness of our methods.

Table 1. Accuracy of one shot classification (%)

Method	Category	5-way	20-way
Siamese Network	Neural Network	95.00	-
MED(eul)	Neural Network	96.64	94.52
Ours	**Rule-based**	**96.91**	**95.10**

We leveraged basic multi-processing to speed up similarity metrics calculation. From Table 2 we observe linear acceleration effect when increasing number of CPU cores from 1 to 4, but fail to scale acceleration further due to possible increasing multi-processing overhead and CPU overloads. To balance the efficiency and resource consumption, we decided to use four CPU cores as main configuration.

The running time demonstrated in Table 2 indicates that searching key characters in databases with 10,000 samples could be done in several minutes with low-resource requirements, for example, an ordinary laptop, while the computing resources required for reaching such running speed using deep-learning methods are significantly increased. This provides wider possibility and availability for historical documents researchers to automate their routine work.

Table 2. Images Processing Speed v.s. Number of Used Processors

# CPU cores	Processing Average Speed (times/s)
1	11.16
2	19.27
4	**39.54**
8	41.69

4. Conclusion

We present a new strategy for efficient one-shot recognition of ancient Chinese characters. The advantages of the proposed method include the methodology of using meta-graph information based on morphological analysis, a new procedure combined with linear programming, low resource requirements, good interpretability, and the possibility of generalizing the methodology to other series of recognition problems.

References

Chen, J., Li, B., Xue, X., 2021. Zero-shot Chinese character recognition with stroke-level decomposition. *arXiv preprint arXiv:2106.11613*.

Du, Y., Li, C., Guo, R., Yin, X., Liu, W., Zhou, J., Bai, Y., Yu, Z., Yang, Y., Dang, Q. et al., 2020. PP-OCR: A practical ultra lightweight ocr system. *arXiv preprint arXiv:2009.09941*.

Kunwar, R., Sapkota, H. P., 2022. An Introduction to Linear Programming Problems with Some Real-Life Applications. *European Journal of Mathematics and Statistics*, 3(2), 21-27. https://www.ej-math.org/index.php/ejmath/article/view/108.

Li, H., Ren, X., Lv, Y., 2020. One-shot chinese character recognition based on deep siamese networks. Y. Jia, J. Du, W. Zhang (eds), *Proceedings of 2019 Chinese Intelligent Systems Conference*, Springer Singapore, Singapore, 742–750.

Li, M., Lv, T., Chen, J., Cui, L., Lu, Y., Florencio, D., Zhang, C., Li, Z., Wei, F., 2023. Trocr: Transformer-based optical character recognition with pre-trained models. *Proceedings of the AAAI conference on artificial intelligence*, 37number 11, 13094–13102.

Liu, X., Gao, W., Li, R., Xiong, Y., Tang, X., Chen, S., 2022. One shot ancient character recognition with siamese similarity network. *Scientific Reports*, 12(1), 14820. https://doi.org/10.1038/s41598-022-18986-z.

Melnyk, P., You, Z., Li, K., 2019. A high-performance CNN method for offline handwritten Chinese character recognition and visualization. *Soft Computing*, 24(11), 7977-7987. http://dx.doi.org/10.1007/s00500-019-04083-3.

Mestetskiy, L. M., Koptelov, D. A., 2024. Constructing the Internal Voronoi Diagram of Polygonal Figure Using the Sweepline Method. *Programming and Computer Software*, 50(4), 292-303. https://doi.org/10.1134/S0361768824700105.

Mestetskiy, L., Semenov, A., 2008. Binary image skeleton - continuous approach. *Proceedings of the Third International Conference on Computer Vision Theory and Applications - Volume 1: VISAPP, (VISIGRAPP 2008)*, INSTICC, SciTePress, 251–258.

Siddiqi, K., Pizer, S., 2008. *Medial Representations: Mathematics, Algorithms and Applications*. 1st edn, Springer Publishing Company, Incorporated.

Xu, X., Yang, C., Wang, L., Zhong, J., Bao, W., Guo, J., 2022. A sophisticated offline network developed for recognizing handwritten Chinese character efficiently. *Computers and Electrical Engineering*, 100, 107857. https://www.sciencedirect.com/science/article/pii/S0045790622001495.

Xu, Y., Yin, F., Wang, D.-H., Zhang, X.-Y., Zhang, Z., Liu, C.-L., 2019. Casia-ahcdb: A large-scale chinese ancient handwritten characters database. *2019 International Conference on Document Analysis and Recognition (ICDAR)*, 793–798.