Development of an Automatic Correction System for Japanese Calligraphy Beginners

Yoshida Nozomu

Master's Program in Informatics
Degree Programs in Comprehensive Human Sciences
Graduate School of Comprehensive Human Sciences
University of Tsukuba

March 2023

Development of an Automatic Correction System for Japanese Calligraphy Beginners

Name: Yoshida Nozomu

In this study, the effectiveness of using existing fonts and machine learning to evaluate and improve the quality of calligraphy characters with block style was demonstrated through the visualization of Shodo correction. A font survey and the implementation of stroke-level and balance-level classification methods allowed for the development of a proof-of-concept prototype for an auto-correction system, which also exhibited some similarities to human correction. The research has the potential for practical applications in Shodo instruction and practice. It offers two contributions: (1) the evaluation of 99 kaisho fonts through two surveys involving 40 participants, including 26 Dan-holders, with the top and bottom four fonts serving as positive and negative examples, and the provision of insights that may be applied to fields such as computer science, education, and digital humanities, and (2) the creation of a prototype Shodo auto-correction system utilizing machine learning and achieving high accuracy rates of 94% and 98% for stroke and balance classification, respectively, based on teacher data obtained from font questionnaires, with the classification processes also presented in a manner that is easily understandable to humans.

Main Academic Advisor: Yoichi OCHIAI
Secondary Academic Advisor: Hiroko TERASAWA

Contents

List of Figures

Chapter 1

Introduction

This chapter describes the background of this research, such as introducing some basic concepts on Shodo (Japanese calligraphy) and research motivation. Also, a description of the key idea for the visualization of the Shodo correction by machine learning approach is included. Finally, some contributions of this research are listed.

1.1 Background

This section describes the Shodo concept, instruction process, and research motivation for this research.

1.1.1 Terminology: What is Shodo?

Japanese calligraphy (書道, shodō) also called shūji (習字) is a form of calligraphy, or artistic writing, of the Japanese language. For a long time, the most esteemed calligrapher in Japan had been Wang Xizhi, a Chinese calligrapher from the 4th century, but after the invention of Hiragana and Katakana, the Japanese unique syllabaries, the distinctive Japanese writing system developed and calligraphers produced styles intrinsic to Japan. The term shodō (書道, "way of writing") is of Chinese origin as it is widely used to describe the art of Chinese calligraphy during the medieval Tang dynasty [1]. Here is a quote from an explanation of calligraphy in a guidebook for beginners to experience calligraphy written by the Japan Calligraphy Education Foundation, a major calligraphy organization [2]: "It is believed that Chinese characters (kanji in Japanese) were introduced to Japan in the second half of the 4th century. At around the same time, shufa (Chinese calligraphy) was incorporated into Japanese culture, and studying calligraphy became proof of one's knowledge and education. Japan established its own unique form of calligraphy by the 9th century. Even though printing and information technology has developed significantly, handwriting is still highly valued in modern-day Japan. It is incorporated into street signs, advertisements, restaurant menus, nameplates, letters, certificates, school curricula, enrichment lessons, and art. Most of these are written with ink brushes, not pens. Why do the Japanese value brush-writing so much? One of the reasons is that people have always viewed those acquainted with calligraphy as wise or educated. But perhaps the biggest reason is that the Japanese recognize that handwriting can convey the writer's emotions, feelings, and character. The popularity of handwriting-style fonts, emoticons, and emoji imply that the Japanese find it difficult to express one's true feelings and emotions with printed words. In this sense, one of the greatest appeals of Japanese culture is that calligraphy and handwriting are deeply rooted in its society and in the people's daily lives."

1.1.2 Instruction Process in Shodo

Although the instruction process in Shodo differs depending on 会派/書派/塾, there is a representative training flow to learn Shodo: (1) Observation of the model: In this step, the student observes and learns basic calligraphy techniques and strokes by looking at a model of proper calligraphy; (2) Writing actually on the paper; The student then practices writing on paper using the techniques and strokes learned in the previous step. (3) Receiving feedback and correction from the instructor: The student receives feedback and correction on their writing from the instructor to improve their techniques and strokes. (4) Repeating steps 1 to 3: The student repeats these steps multiple times to practice and improve their calligraphy skills.

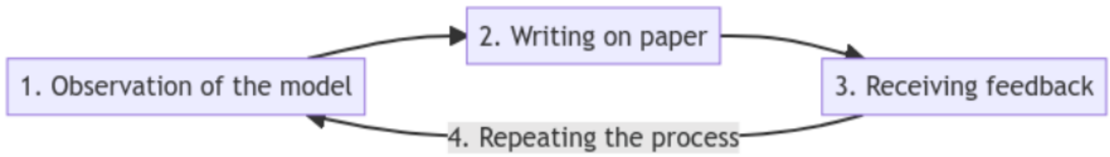

Figure 1.1: Basic practice process of Shodo: Repeat the process of observing a model, actually writing with a brush, and receiving corrections from the teacher to recognize areas for correction.

1.1.3 Research Motivation

The idea to automate calligraphy corrections came from the observation that corrections often follow certain patterns, both subjectively and objectively. The motivation for this automation originated from the author's experience learning calligraphy from two masters over the course of eight years. During this time, the author noticed typical patterns in the feedback given on submitted writing homework, particularly regarding common elements in kanji characters. This led the author to think that it might be possible to program a system to automate at least some of the correction processes, even if it could not fully replace the current process of human correction. This automation aims to capture the essence of the correction process and apply it more objectively. Basic stroke (基本点画) shown in fig.1.2 refers to the basic way of drawing Japanese characters in calligraphy. Basic stroke includes brush strokes and dot combinations. Basic stroke is the most fundamental aspect of learning calligraphy and is essential for learning how to write Chinese characters. In addition, mastering basic strokes lays the foundation for creating works of art using calligraphy. In the sense that all characters are made up of combinations and applications of basic strokes, it can be said that basic strokes are objective patterns unique to calligraphy.

1.2 Proposal of This Research

This section describes the research question (how I connect my motivation for Shodo correction automation to the research), the expected result, and the key idea for the visualization of the Shodo correction by machine learning approach.

Research Question

I am interested in exploring the feasibility of using a machine learning-based auto-correction system to replace the traditional one-on-one correction process typically used in Japanese calligraphy instruction. Specifically, I want to determine whether the system can accurately and effectively identify and visualize corrections in the learner's Shodo handwriting and to what extent it can replace the existing heuristic correction process from the perspective of both accuracy (both individual classification and overall proportion) and visibility (glance-easiness) as perceived by the learners. I am also interested in understanding how well the system performs compared to human instructors in providing feedback and guidance to learners and whether it can help learners improve their Shodo handwriting skills more efficiently and effectively.

Key idea for visualization of the shodo correction: machine learning approach

The technical challenge in this study is the incorporation of fonts as a dataset for a machine learning-based auto-correction system that can identify and visualize specific corrections in the input Shodo image provided by the learner. The system uses machine learning algorithms to analyze positive and negative examples of font data to determine the correct and incorrect parts of the learner's input image of their writing work. The goal is to provide the learner with visual feedback on their writing, highlighting areas where they may need to improve their technique to produce more accurate and visually pleasing Shodo handwriting.

1.3 Contributions

The main contributions of this study are the following two:

(1) Evaluation of kaisho fonts with various perspectives

Two surveys were conducted to evaluate 99 kaisho fonts with 40 participants, including 26 Dan-holders. Among the 99 fonts, 27 selected fonts are evaluated with multiple perspectives by ten Dan-holders. Although this study only applied the evaluation result of the top four and bottom four fonts as positive/negative examples, there are unused insights from the survey result that would be utilized in other research such as computer science (e.g., kaisho font-based machine learning approach including generation/classification of the character), education (e.g., Utilization of fonts data as a model in Shodo learning), and digital humanities (e.g., This study: Auto-correction in Shodo.)

(2) Development a Proof-of-concept prototype of Shodo auto correction

An automatic correction system was constructed to visualize areas where learners should practice more by machine learning using positive and negative example fonts carefully selected through two font questionnaires as teacher data. The stroke classification model and the balance judgment model achieved a high accuracy of 94% and 98%, respectively, in classifying positive and negative examples. The basis for the classification was visualized in a form understandable to humans.

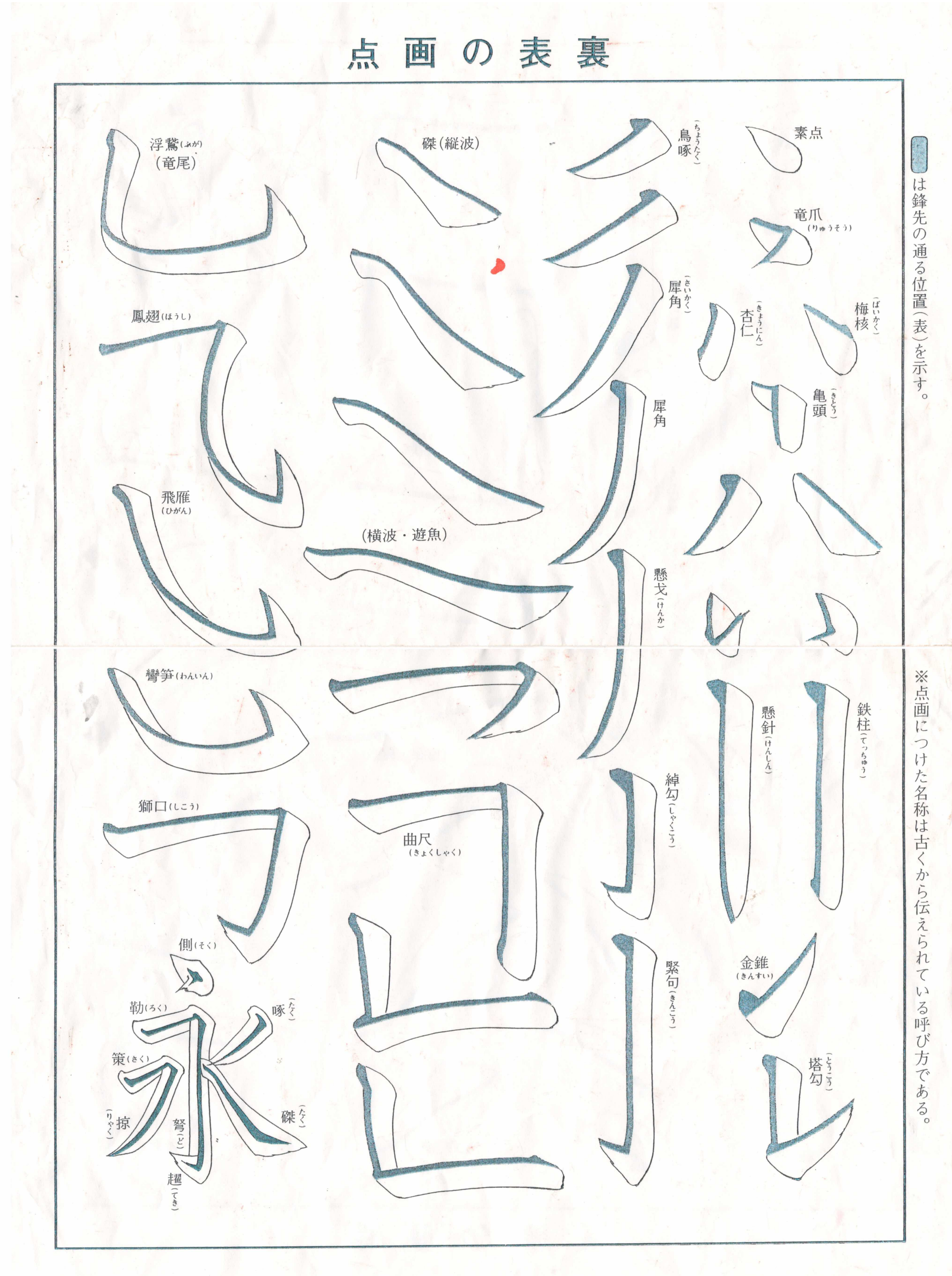

Figure 1.2: Example of basic stroke: In Japanese calligraphy, the concept of basic stroke (基本点画) refers to the fundamental building blocks of kanji characters. Each kanji is made up of a number of strokes, which are the lines and brush movements that make up the character. These strokes are arranged in a specific order, and the correct order is important for creating legible and aesthetically pleasing characters. Learning the basic strokes is an important part of learning Japanese calligraphy, as it allows the calligrapher to create the various kanji characters with precision and grace. It is also important to understand the structure and meaning of kanji characters, as the strokes can often provide clues to their pronunciation and meaning. Source from the Japan Calligraphy Education Foundation.

Chapter 2

Related Work

There are related works on Calligraphy (including Shodo) learning support systems such as evaluation by computer, learning mobile/VR interfaces, development of haptic feedback devices, and so on. This section describes related works divided into several categories.

2.1 Generation and Representation of Character

Berio et al. devised a technique for automatically breaking up a font's glyphs into a collection of overlapping and intersecting strokes, which can then be used to generate variations, stylizations, and animations in various artistic or design-focused styles [3]. Seah et al. proposed a modeling and representation method for brushstroke and animation based on disk B-spline curves (DBSC) [4]. Lin et al. present an autonomous robotic writing system for Chinese calligraphy that is enabled by the proposed automatic stroke matching and generation mechanisms, allowing the robot to effectively learn to write any Chinese character in a style that is sampled by a small number of handwritten Chinese characters with a specific target writing style [5].

2.2 Auto Evaluation of character

Fang et al. proposed a design of the Qi Gong calligraphy learning system with a computer by simulating the actual writing process stroke by stroke, including the construction of a Qi font vector library with 3755 characters, decomposition of the character by corner detection algorithm, and evaluation of the stroke through graphical user interface [6]. Sun et al. proposed a number of aesthetic feature representations and fed them into Artificial Neural Networks [7]. Moreover, a Chinese Handwriting Aesthetic Evaluation Database (CHAED) is also built with 1000 Chinese handwriting images. Experimental results demonstrate that the proposed AI system with its original database for the aesthetic evaluation of Chinese calligraphy provides a comparable performance with human evaluation. Xu et al. use an area-based evaluation approach to automatically grade the visual appearance of calligraphic writing, where the grading result of our method closely resembles human aesthetic opinions. They also treat stroke decomposition and grading visualization through the implementation [8]. Wang et al. developed a stroke-by-stroke calligraphy evaluation system with area-based scoring using a vector data structure, and a skeleton-based reference [9]. It is faster and more versatile than the raster format evaluation that was mainstream in previous research.

Since the input is assumed to be on an electronic tablet, the evaluation of non-digital calligraphy is omitted. In the process of developing a PC-based system to support the practice of brush copying evaluation methods, Miyaji et al. surveyed and reviewed instructional books and manuals on brush copying and organized the evaluation items (brush stroke usage, letter shapes, letter arrangement, etc.) [10]. Han et al. suggested an interactive calligraphic grading system that employs fuzzy inference and image processing to assess the quality of written characters with advice for improvement, allowing users to study and practice Chinese calligraphy at home [11]. However, the system feedback is not visualized so users can't figure out the correction point intuitively. Nobe et al. proposed a method for the aesthetic evaluation of characters in distance education using penmanship and calligraphy [12]. In actual correction, evaluation is not conducted by focusing only on the general characteristics of the characters but also on the individual dots that make up the characters. In this study, they created character evaluation data for each character type and developed a program that can also evaluate the details of individual characters. Shin et al. proposed a pen-tablet-based calligraphy system for the oriental brush writing with the sensory calligraphy method, so-called Yongzi-Bafa in Chinese, Yongza-Palbop in Korean, Eiji-Happo in Japanese [13]. The system compares each stroke's feature point to the reference data, which includes 13 feature points from one character's Chinese name, Young. The system then provides the user with the analysis and comparison of model and input data. Since the system only covers "Young" ("永") characters, the system lacks flexibility in practice.

2.3 Visualization of Correction

Bando et al. and Nishioka et al. have developed an application that performs simple corrections on a smartphone by superimposing the input work, and the model image [14, 15]. Also, the process of extracting Chinese character strokes requires correcting any errors, and visualization and adaptive correction methods were proposed to facilitate this process [16]. These methods visualize the extracted strokes using color, brightness, saturation, and the degree to facilitate manual human correction of errors. Also, a method for visualizing the results of matching Chinese character strokes using a multi-level hierarchy based on features and information is proposed [17]. The hierarchy includes colors, symbols, and numbers and effectively simplifies the process, and improves efficiency. The mind calligraphy system is an interactive calligraphy tool that visualizes the writer's emotions in real-time through animations and color palettes [18]. It uses brain wave data to classify the writer's emotions into four categories. It has been shown to increase the writer's interest in calligraphy, help them understand the connection between calligraphy and emotions, and provide a new and interactive experience of calligraphy.

2.4 Other HCI research on Shodo

"Spring-Pen" is a research paper that describes a system for reproducing the softness of materials using 3D printing. The system utilizes a 3D printer that can print a spring-like structure in addition to solid objects. By adjusting the shape and stiffness of the spring, the researchers were able to create objects with a wide range of softness, from hard to soft. The system could reproduce the softness of various materials, including silicone rubber and foam, with good accuracy. The authors suggest that this technology could be used to create custom-designed soft objects for a range of applications, such as robotics and haptic interfaces [19, 20]. Muranaka et al. describe the development and evaluation of a calligraphy learning system using virtual reality (VR) technology [21]. Conventional tablet-type pens do not provide a sense of writing pressure. Therefore, they developed a pressure pen input device that closely resembles the sensation of writing pressure. They also developed an automatic animation generation process using 3D computer graphics. Using this animation, practitioners can experience the subtle writing quality of calligraphy as seen through the eyes of a calligraphy teacher. Yang et al. simplified the generation of writing trajectories to an optimization problem that considers only the width of the stroke as a function of its height (z-axis) [22]. Based on the skeleton extracted from each stroke, the width at each sampling point is solved and converted to a smooth trajectory using dynamic programming and a Gaussian process model. In this way, the robot can learn to write any Kanji character directly from the image input within half a minute. Brozkova attempted to comprehensively discover where computers could be beneficial in personal instruction of calligraphy from an HCI perspective [23]. Specifically, they found the most suitable human-computer interaction device for computerized calligraphy and created a tutoring program using the device. Shichinohe et al. introduced an augmented calligraphy system that aims to support the self-learning process of calligraphy learners through feedback [24]. Body posture is a very important factor in writing well. However, it is difficult to maintain correct posture without an assistant. Therefore, they developed a system that monitors the learner's posture with a web camera and notifies the learner when the posture becomes poor. Hira et al. proposed a writing system using a pen-type haptic device that enables human-computer interaction through touch operation and forces feedback of the pen to realize a virtual writing system [25]. In this system, a reaction force is applied to the virtual paper, and the handwriting was recorded according to the position and pressure of the pen tip during virtual writing. Suzuki et al. proposed a virtual calligraphy brush model for haptic devices that takes into account the brush's ground contact area, compression force, and friction force to enable realistic drawing with haptic devices [26]. The implemented VR calligraphy system was evaluated through experiments, showing that the proposed model can realize a realistic sensation of calligraphy. Li et al. proposed a system that can imitate a specified Kanji character with a virtual brush according to the movement and strength of the user's finger that touches the cell phone screen [27]. In this case, a calligraphy imitation system based on the virtual brush and scale-invariant feature transform is used to learn a feature library of many kanji characters. Experimental results show that the system can accurately imitate Chinese characters and efficiently match accurate images. A method of controlling brush grasping and stroke movement in calligraphy using a multi-fingered hand robot is proposed by Tsutsumi et al [28]. The proposed system realized a brush's grasping and stroke control, although the quality of the beginning and ending strokes remained a problem. Shimada et al. proposed a virtual calligraphy system drawable with a Chinese brush that users can write directly on the screen with a brush that has a built-in 3D position sensor and experience virtual calligraphy under a working environment that is almost identical to that of real-world calligraphy, including an actual brush, virtual half-paper, and inkstone [29]. The system is calibrated to generate handwriting exactly where the brush strokes touch and the handwriting is generated based on actual measurements of the strokes made with the real brush strokes. The system also monitors the spatial behavior of the brush strokes and expresses visual effects such as ink drips and splashes. I/O Brush is a drawing tool designed for young children to explore the colors, textures, and movements of everyday materials [30]. It looks like a regular paintbrush but has a camera and sensors inside that allow it to pick up and draw with the colors and textures of the materials it touches.

2.5 Position of This Study

A summary of the related studies mentioned above yields the following findings. (1) few studies deal with real calligraphy: Many studies on calligraphy use tablet devices for input, and there are few studies that evaluate actual calligraphy works using actual brushes and paper. (2) Visualization is rough: Although there are studies on correcting real calligraphy, they are limited to simple corrections, such as overlapping input images with an example image, and cannot be said to visualize actual corrections. (3) Rule-based Evaluation with low applicability: Many existing evaluations of calligraphy are rule-based. However, characters are diverse, and few visualize evaluations with large-scale data. In other words, no research in Japan or China uses machine learning-based large-scale data, corresponds to real calligraphy, and visualizes the evaluation of calligraphy. The correction system proposed in this study does not assume tablet-based input but rather supports image-based input, making it applicable to real-world calligraphy.

Chapter 3

System Design

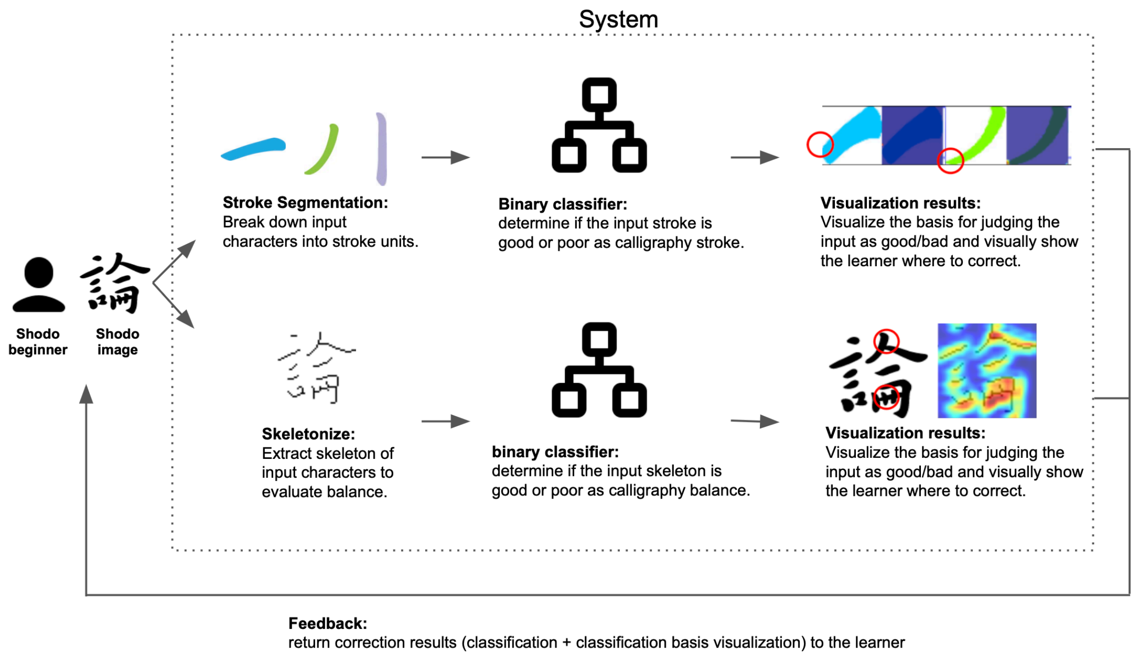

Figure 3.1: Overview of system design with ML approach for classification and visualization: The system corrects calligraphy in terms of stroke units and overall balance. A) For stroke correction: 1) input an image of calligraphy; 2) segment the input image into stroke units; 3) have the input strokes evaluated by a binary classifier that distinguishes between good and bad strokes; 4) visualize the features that the classification results are based on; 5) feedback the results to the user. B) For whole balance correction: 1) input an image of calligraphy; 2) preprocess the input image and extract a skeleton representation; 3) use the binary classifier to determine the quality of the input strokes; 4) visualize the feature values that the classification result is based on; 5) feedback the results to the user.

[0] CNN image in the figure: https://towardsdatascience.com/understanding-input-and-output-shapes-in-convolution-network-keras-f143923d56ca, accessed on 2022.12.25

3.1 Factorization of Shodo correction: perspectives and method

When we break Shodo correction down into its constituent parts, there are two main factors: (1) individual quality of basic stroke and (2) overall proportion as an entire character. Accordingly, these two perspectives are important for reproducing the correction process. On the other hand, the correction in Shodo describes the good part (which means keep on going in the next practice) and the bad part (which means try to modify in the next practice) for each character. Accordingly, it's natural for visualization of the correction to show the actual good/bad parts of the character.

3.2 Overview of system design

Therefore, the correction system aims to feedback visualization results of the good/bad part in the input character for these two perspectives: individual strokes and overall proportion. To visualize, this study created a classification model of good and bad players by CNN and visualized the basis for the classification as shown in fig.3.1.

3.3 Classification design

"In this research, a machine learning approach, rather than a rule-based approach, is adopted to classify the input characters as good or bad for the following three reasons: 1) ML-compatibility in terms of data volume: There is a large amount of font data available for machine learning as training data (e.g. In Japanese, a single font typically includes about 7600 characters, though this may vary depending on the font[1]). 2) ML-compatibility in terms of pattern recognition: There are typical patterns that emerge when human instructors correct Shodo (e.g. There are common components for Shodo known as 'Kihon Tenkaku,' which means basic strokes. These strokes are commonly found in a lot of characters with typical correction patterns.). 3) Rule-based-incompatibility: Correction is a subjective process (two people's corrections often differ to some extent, even if they both have an instructor license in Shodo), so it is difficult to define a rules-based algorithm."

3.4 Visualization design

Numerous visualization techniques can be employed to gain insight into the decision-making processes of machine learning models. Some common examples include:

- Activation maps: These visualizations highlight the most influential features for making a prediction by displaying the activations of neurons in different layers of the model.
- Saliency maps: These visualizations depict the regions of the input that are most consequential for making a prediction by emphasizing pixels with a high impact on the model's output.
- Feature maps: These visualizations illustrate the features learned by different model layers, providing insight into what the model is looking for when making a prediction.

[1] https://www.dynacw.co.jp/support/support_faq_detail.aspx?qid=377, accessed on 17th, Nov

- Layer-wise relevance propagation (LRP): This technique breaks down the prediction of a model into the contributions of individual neurons, enabling users to understand how the model utilizes various features in its prediction.

In comparison to these other techniques, Grad-CAM has several notable advantages. It is easy to implement and provides clear, interpretable visualizations that facilitate an understanding of the model's decision-making process. Additionally, it is model-agnostic and can be applied to a wide range of CNN architectures. Grad-CAM belongs to the category of activation maps, which visualize the activations of neurons in different layers of a model and highlight the most relevant features for making a prediction. Specifically, Grad-CAM generates a heatmap by utilizing the activations of neurons in the final convolutional layer of a CNN to show the regions of the input that are most pertinent to the prediction. Saliency maps, feature maps, and LRP are all distinct visualization techniques that can be used to gain insight into the workings of machine learning models. Saliency maps highlight pixels in the input that significantly impacts the model's output, while feature maps display the features learned by different model layers. LRP decomposes a model's prediction into the contributions of individual neurons, enabling users to understand how the model utilizes various features in its prediction. "The two essential steps for visualizing correction are: 1) having the classifier determine whether the input image of written Shodo is of positive (well-written) or negative not (well-written) quality and 2) visualizing the basis for this judgment. It is crucial that these two approaches are applied to both individual strokes and the overall proportion of the character. "

Chapter 4

Font survey for data construction

This chapter describes how the author collected, selected, and preprocessed the font as training data to make ML-model for classification.

4.1 Summary of the selection

An automatic correction system necessitates classification to discern whether the input character constitutes a positive or negative example. In this study, the author implemented a machine learning approach rather than a rule-based approach for this classification, as outlined in the "System Design" section. Two surveys were conducted: (1) the first survey was administered to 30 participants, comprising individuals with and without familiarity with Shodo, to identify potential training data candidates, and (2) the second survey was conducted with a group of seasoned practitioners (ten people of Dan holders) to verify the training data. As a result of these surveys, eight fonts (four positive example fonts and four negative example fonts) were selected as the final training data for stroke-level classification. Four fonts (two positive example fonts and two negative example fonts) were selected as the final training data for stroke-level classification as shown in fig4.9. The whole selection process is shown in fig.4.1, and the 99 kaisho fonts surveyed were samples from a website[31] that compiles Japanese fonts.

4.2 Adopted font style: kaisho (block style)

There are five basic styles commonly used in Japan for writing Kanji: They are tensho (seal style), reisho (scribe's style), and kaisho (block style) in chronological order, which all appeared in China before the end of the fourth-century [32]. In this paper, our target writing style is Kaisho, which looks closest to the original Chinese characters, and is easiest to read. Kaisho is the first style most Shodo beginners usually learn and practice. The target characters for use in training are the educational Kanji characters specified by the Japanese Ministry of Education, Culture, Sports, Science, and Technology.

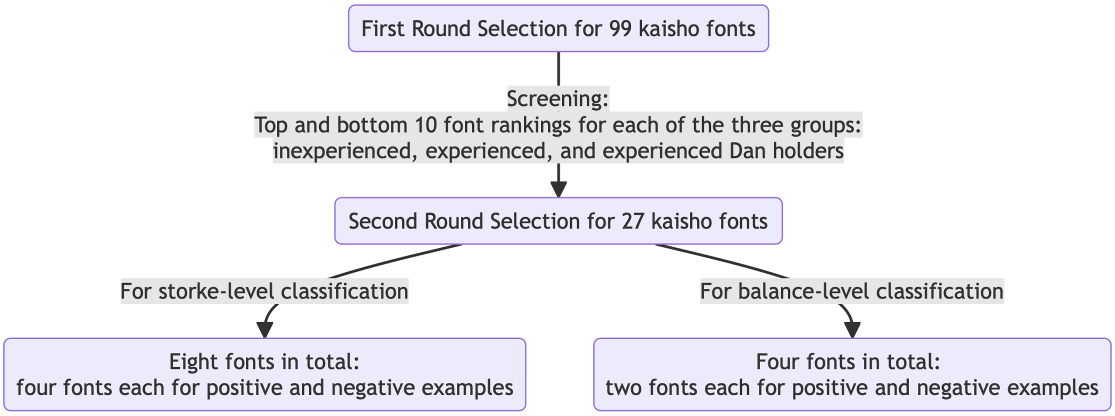

Figure 4.1: Overview of font survey process: Two surveys were conducted: (1) the first survey was administered to 30 participants to identify potential training data candidates, and (2) the second survey was conducted with a group of seasoned practitioners to confirm the training data.

4.3 First round selection: 27 fonts selected from 99 fonts

4.3.1 Selection attributes and perspectives

99 Kaisho fonts sample are collected from an online website called *The Japanese Font library for macOS and Windows* [31]. For screening purposes, the first survey was conducted to select Kaisho fonts which are used as positive and negative examples. Participants collected from social media (e.g., Facebook and Twitter) were asked to answer the evaluation of 99 fonts at google forms online. The evaluation required a single five-level judgment for each font, where the evaluation used a five-level Likert scale (1: Bad, 2: Slightly bad, 3: Normal, 4: Slightly good, and 5: Good) based on the question of "How much do you want to use the designated font as the model when you practice Kaisho on the brush?".

The table 4.4 presents the description of the first font survey. The aim of the study is to assess which fonts are suitable as positive and negative examples of kaisho fonts in brush calligraphy. The methodology employed involved evaluating 99 block letter fonts according to a five-point scale (ranging from poor to excellent). The evaluation criterion was whether the font would be suitable as a model for brush calligraphy. The study period was from August 7th to August 12th, 2022, spanning a total of six days. Recruitment for the study was conducted via social media, using an online questionnaire administered through Google Forms. There were 30 respondents, all of whom were Japanese. The age distribution of the respondents included one person under ten years of age, two teenagers, ten people in their twenties, three people in their thirties, seven people in their forties, four people in their fifties, and three people who did not provide an answer. The degree distribution of the respondents included 24 experienced individuals, 16 graded individuals, and six people who were either inexperienced or had only learned in a class. The group affiliations of the respondents included Japanese Characters, the Japanese Calligraphy Academy, the Japanese Pen Calligraphy Research Association, the Tokyo Calligraphy Education Association, the Japanese Art Academy, the Yomiuri Calligraphy Exhibition, Kumon, and the Independent Calligraphy Group (previously).

Table 4.1: Description of first font survey

Step	Description
Purpose	Investigate suitable fonts as positive and negative examples of kaisho fonts in brush calligraphy
Method	Evaluation of 99 block letter fonts in 5 grades (5 grades = poor, slightly poor, normal, slightly good, good)
Evaluation criteria	"Would you like to use this font as a model for brush calligraphy when you practice calligraphy?"
Period	2022/08/07 - 2022/08/12 (6 days)
Recruitment	SNS (online questionnaire via Google Forms)
Respondents	30 people (all Japanese)
Recruitment	SNS (online questionnaire through Google Forms)
Age distribution	Under 10 years old: 1 person / Teens: 2 people / 20s: 10 people / 30s: 3 people / 40s: 7 people / 50s: 4 people / No answer: 3 people
Degree distribution	Experienced: 24 people, Graded: 16 people, Unexperienced / Learned in class: 6 people
Associations	日本習字 / 日本書道学院 / 日本ペン習字研究会 / 東京書道教育会 / 日本書芸院 / 読売書法展 / くもん / 独立書人団

4.3.2 Result

The result is shown in the fig.4.2 to fig.4.7. Many fonts, both positive and negative examples, were ranked in duplicate by all three respondent groups. However, some fonts were ranked by only one respondent group (e.g., positive example #7 Arphic Kaisho Medium within the 16 Dan-holders respondent group ranking).

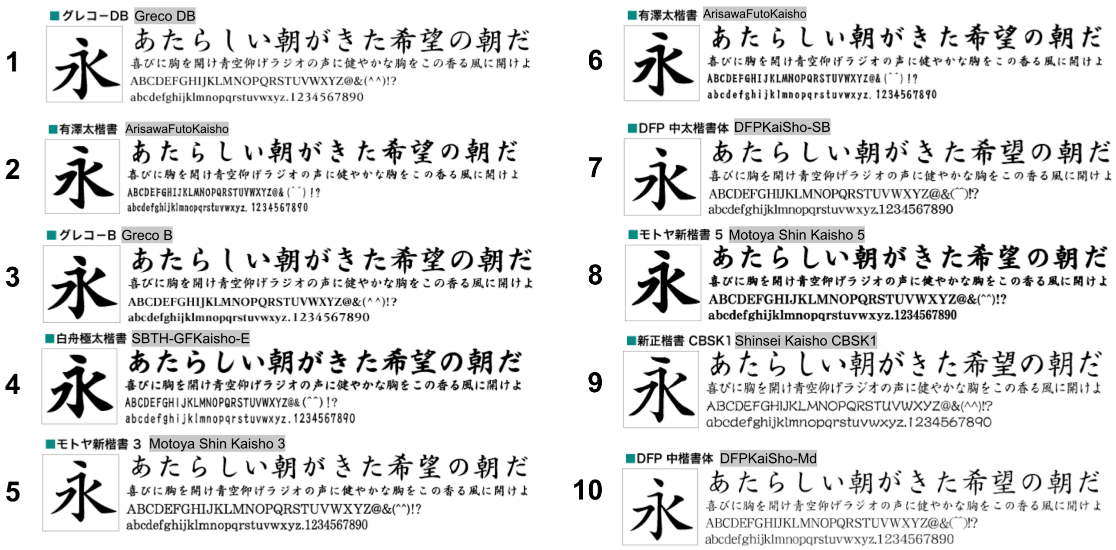

Figure 4.2: Positive example candidates ranking for 30 respondents (ascending-order)

Table 4.2: Candidate fonts of positive example: Fonts ranked in the top 10 in at least one of the three attribute groups (all respondents (including inexperienced respondents), experienced respondents, and only experienced respondents) in the first font survey.

English	**Original Japanese**
Greco B	グレコー B
Motoya regular Block Style 5	モトヤ正楷書 5
Greco DB	グレコー DB
Motoya Sei Kaisho 3	モトヤ正楷書 3
Motoya Shin Kaisho 3	モトヤ新楷書 3
JTC Namiki Tokubuto Kaisho	JTC ナミキ特太楷書
Motoya Shin Kaisho 5	モトヤ新楷書 5
ArisawaFutoKaisho	有澤太楷書
SBTH-GFKaisho-E	白舟極太楷書
Arphic Kaisho Medium	AR 楷書体 M
Shinsei Kaisho CBSK1	新正楷書 CBSK1
DFPKaiSho-Md	DFP 中楷書体
A-OTF	欧体楷書
DFPKaiSho-B	DFP 太楷書体
DFPKaiSho-SB	DFP 中太楷書体

Table 4.3: Candidate fonts of negative example: Fonts ranked in the bottom 10 in at least one of the three attribute groups (all respondents (including inexperienced respondents), experienced respondents, and only experienced respondents) in the first font survey.

English	**Original Japanese**
Soei Futo Kaisho 11*	創英太楷書体 11
DFPGanKaiSho W7	DFP 顔楷書 W7
DFPGanKaiSho W9	DFP 顔楷書 W9
DFKKaiShoA Std W5	DFP 華康楷書体 A W5
DFPGanKaiSho W5	DFP 顔楷書 W5
HGGyokoku	HG 行刻
DFKKaiShoB Std W5	DFP 華康楷書体 B W5
DFPDanKaiSho-W5	DFP 談楷書 W5
HGHakushuGokubutoKaishotai	HG 白洲極太楷書体
DFP Shinsotai*	DFP 新宋体
Soei kaisho 4*	創英楷書体 4
DF Kyo Sui Japanese W3	DF 鏡水 std W3

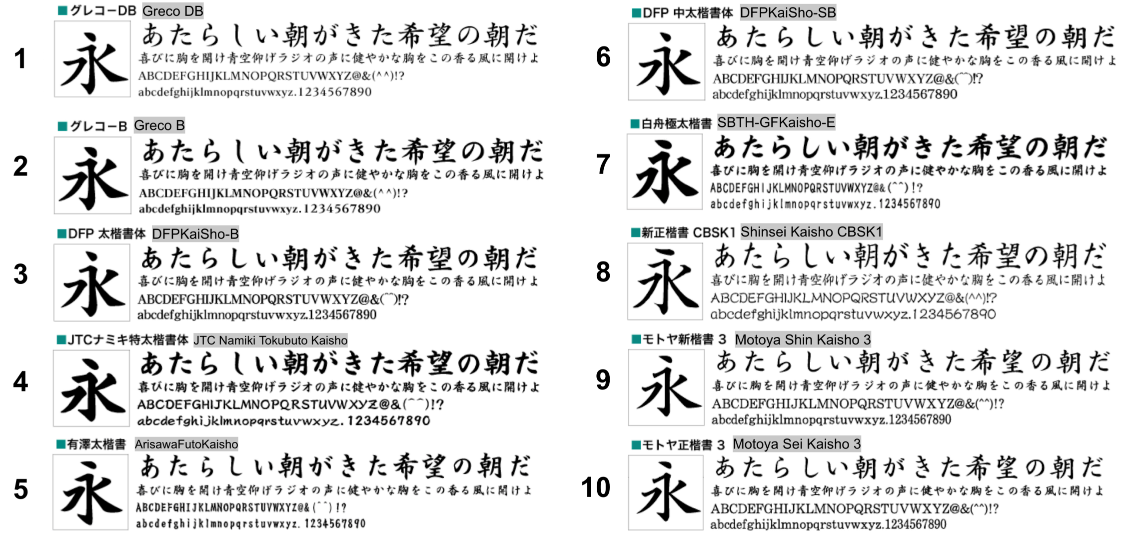

Figure 4.3: Positive example candidates ranking for 24 experienced respondents (ascending-order)

4.4 Second round selection: 8 fonts selected from 27 fonts

4.4.1 Selection attributes and perspectives

After the screening in the first round survey, the second round survey was conducted to evaluate 27 fonts more precisely. All participants are Dan holders collected from social media (e.g., Facebook and Twitter) and the contact list of the author asked to answer the evaluation of 27 fonts at google forms online. The evaluation required the judgment for each font, where the evaluation used a five-level Likert scale (1: Bad, 2: Slightly bad, 3: Normal, 4: Slightly good, and 5: Good) based on the seven perspectives (1: General impression, 2: brush-stokes for handling a brush, 3: Shape of the character, 4: Balance of the character, 5: Momentum, 6: Dynamism, 7: Stretched), where the 2-7th perspectives are evidenced from the Paper *Scoring Criterion − referenced Evaluation from Criticism for Calligraphy Works* by Miyaji et. al [10].

Objective: To evaluate 27 standard fonts from 7 perspectives on a 5-point scale to determine desirable fonts for brush calligraphy Means: Each font will be evaluated using the following criteria: stroke usage/strokes, character shape, balance, momentum, strength, stretch, and overall evaluation on a scale ranging from poor to good Period: August 23, 2022, to September 16, 2022 Recruitment: Participants will be recruited via email and social media Survey method: An online survey will be conducted using Google Forms Respondents: 10 Dan-holders including previous round respondents Age distribution: 40s (6 people), 50s (2 people), 60s (1 person), 70s (1 person) Distribution of ranks: 1st dan (1 person), 3rd dan (1 person), provisional 3rd dan (1 person), 4th dan (1 person), 5th dan (1 person), 7th dan (1 person), 8th dan or higher (4 people) Associations：日本習字、日本書鏡院、橘書道会、日本書道学院、璞社

Figure 4.4: Positive example candidates ranking for 16 Dan-holders (ascending-order)

4.4.2 Result

As a result of second-round surveys, eight fonts (four positive exemplars and four negative exemplars) were selected as the final training data for stroke-level classification, and four fonts (two positive exemplars and two negative exemplars) were selected as the final training data for stroke-level classification, as depicted in fig4.9.

Training fonts for stroke-level classification

Greco B (#4), MotoyaSeikaiStd-W5 (#10), Greco DB (#3), and JTC Namiki Tokubuto Kaisho (#4) were selected as the positive example fonts for the stroke classification model (top three in Overall evaluation ranking and top three in Six perspectives ranking). The four selected fonts are (the top three in the Overall evaluation ranking and the top three in the Six perspectives ranking). Four fonts were selected as negative example fonts: DF Kyo Sui Japanese W3 (#23), DFP Shinsotai (#21), HGHakushuGokubutoKaishotai (#25), and DFPDanKaiSho-W5 (#22) (the bottom three in the Overall evaluation ranking and the bottom three in the Six perspectives ranking).

Training fonts for balance-level classification

Greco B (#4) and MotoyaSeikaiStd-W5 (#10) were selected as the positive example fonts for the balanced classification model (top two in the Balance ranking). DF Kyo Sui Japanese W3 (#23) and HHakushuGokubutoKaishotai (#25) were selected as negative example fonts (the bottom two in the Balance ranking).

Distribution of ratings for each perspective

The distribution of ratings for each perspective was investigated. In the fig.4.11, the vertical axis represents the mean value of the responses of all evaluators for font evaluation (5 grades = 1: poor, 2: slightly poor, 3: normal, 4: slightly good, 5: good). Referring to the

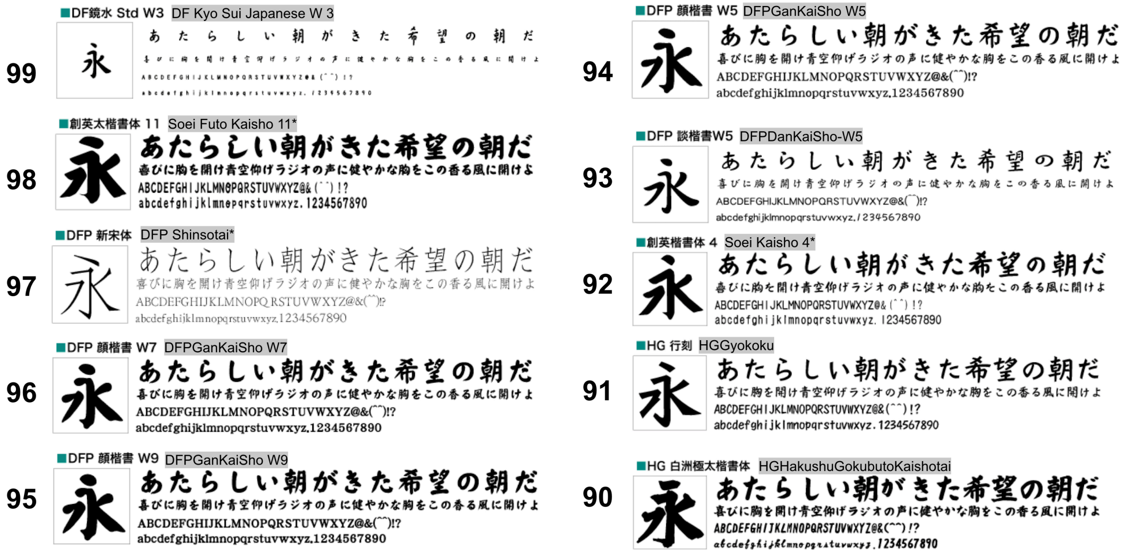

Figure 4.5: Negative example candidates ranking for 30 respondents (descending-order)

interquartile range, it can be seen that, in general, stroke and brushwork, character shapes, and balance, which belong to Skills, have a wide range of evaluations and are prone to be evaluated differently depending on the font. Among them, stroke and brushwork had the greatest variance. On the other hand, momentum, strength, and stretch, which belong to spirituality, are relatively less variable in their evaluations. Among them, stretch had the smallest variance.

Weights of rating for six perspectives

The rating weights for six perspectives were investigated to determine how much evaluators prioritize each perspective when evaluating fonts. In fig.4.12, the vertical axis represents the evaluation value, with 1 being "not prioritized" and 5 being "prioritized." Referring to the interquartile range, there is a tendency for stroke and brushwork, and character shapes to be the most prioritized, followed by balance. The three perspectives of spirituality, momentum, strength, and stretch tend to be less prioritized than the three perspectives of skills, stroke and brushwork, character shapes, and balance.

Weights of the parts in the example sentence

The importance of each part of the sample sentence presented for font evaluation was investigated. In the fig.4.13, the vertical axis is the evaluation value with "not considered" as 1 and "considered" as 5. The character "永" written in large letters on the left was overwhelmingly considered the most important, with an average value exceeding 4.5. Next, the sample sentence at the top right was given the second highest priority, with an average value of 3.5. It can be seen that the smallest Japanese sample sentence and capital/lowercase alphabet were almost not considered.

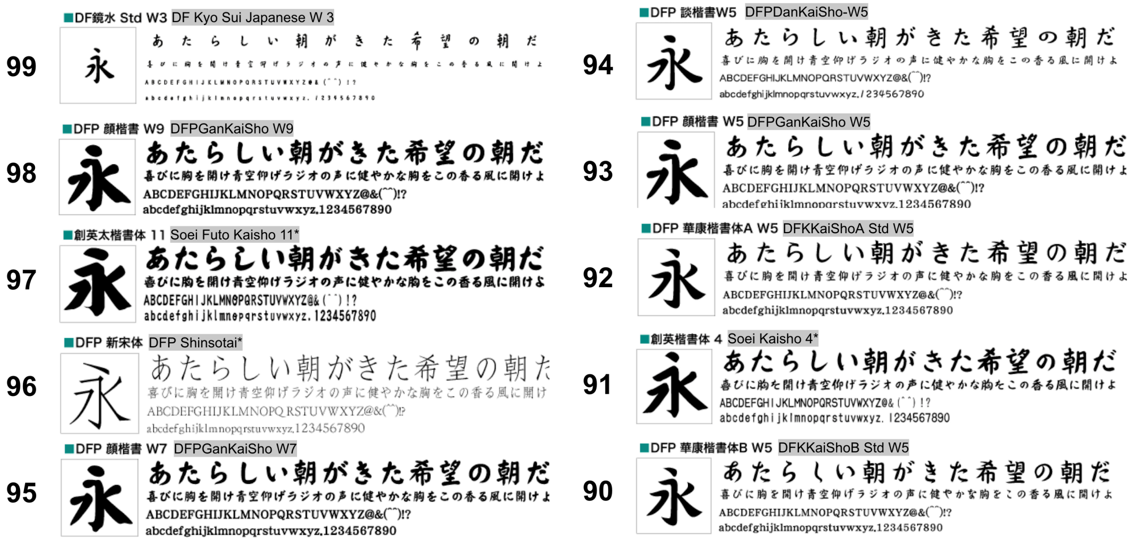

Figure 4.6: Negative example candidates ranking for 24 experienced respondents (descending-order)

Correlation of six perspectives and overall rating

The fig.4.14 represents the correlation of scores: 6-Perspective Total - Overall Rating. It can be seen that there is a correlation between the total score of 6 perspectives and the overall rating. The fig.4.15 also represents the correlation of rankings: 6-Perspective Total - Overall Rating. It can be seen that there is a correlation between the rank of the total score of the 6 perspectives and the rank of the overall rating. What can be inferred from these two figures is that the evaluation based on the six perspectives divided into objective elements is interesting in that it is linked to the overall evaluation, which includes subjective elements.

Correlation of six perspectives and balance rating

The fig.4.16 represents the correlation of scores: 6-perspective total - balance. It can be seen that there is a correlation between the total score of 6 perspectives and the balance rating. The fig.4.17 also represents the correlation of rankings: 6-perspective total - balance rating. It can be seen that there is a correlation between the ranking of the total score of the 6 perspectives and the balance ranking. What can be inferred from these two figures is that the evaluation of 6 points with various evaluation elements is interesting in that it is linked to the evaluation of a single point called balance. In other words, it can be said that fonts that are more likely to receive high evaluations often have excellent balance.

Figure 4.7: Negative example candidates ranking for 16 Dan-holders (descending-order)

Table 4.4: Description of second font survey

Objective	To evaluate 27 standard fonts from 7 perspectives on a 5-point scale to determine desirable fonts for brush calligraphy
Method	Each font will be evaluated using the following criteria: stroke usage/strokes, character shape, balance, momentum, strength, stretch, and overall evaluation on a scale ranging from poor to good
Evaluation criteria	"1. Stroke usage/stroke movement: the perspective of the stroke size/stroke end/stroke beginning/stroke end/stroke; 2. Character shape: size, grain, and overall character shape; 3. Balance: balance between the character and space / whether the character is balanced or not; 4. Momentum: Is the character vigorous?; 5. Strength: Is the character bold/strong/sturdy?; 6: Spontaneity: Is it spontaneous / is the size appropriate / is it written spontaneously?; 8. Overall evaluation: Overall evaluation refers to a subjective judgment, such as "whether or not you would like to use this font as a model for your own calligraphy."
Period	2022/08/23 - 2022/09/16 (24 days)
Recruitment	SNS (online questionnaire via Google Forms)
Respondents	30 people (all Japanese)
Recruitment	Email and SNS (online questionnaire through Google Forms)
Age distribution	40s (6 people), 50s (2 people), 60s (1 person), 70s (1 person)
Degree distribution	Experienced: 1st dan (1 person), 3rd dan (1 person), provisional 3rd dan (1 person), 4th dan (1 person), 5th dan (1 person), 7th dan (1 person), 8th dan or higher (4 people)
Associations	日本習字、日本書鏡院、橘書道会、日本書道学院、璞社

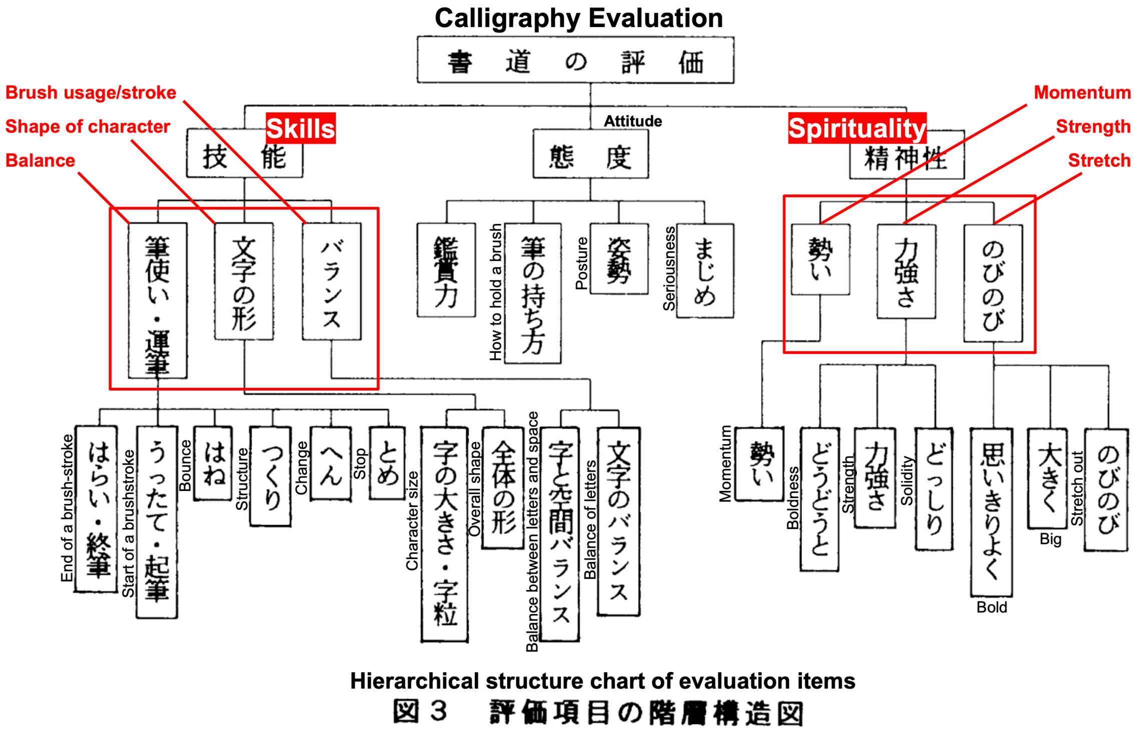

Figure 4.8: Hierarchical Diagram of Evaluation Items for Calligraphy[33] (English translation attached by the author): The evaluation required the judgment of fonts using a 5-point Likert scale based on 7 perspectives. The Likert scale ranges from "Bad" to "Good," The perspectives include general impression, brush strokes, character shape, balance, momentum, dynamism, and stretched.

Font ID (#)	English Font name	Original Font Name	総合評価順位 Overall evaluation ranking	総合評価 Overall evaluation	６観点合計順位 Six perspectives ranking	６観点合計 Six perspectives total	バランス順位 Balance ranking	バランス点数 Balance evaluation
4	Greco B	グレコーB	1	3.8	1	22.5	1	3.9
10	MotoyaSeikaiStd-W5	モトヤ正楷書５	2	3.6	2	21.8	2	3.5
3	Greco DB	グレコーDB	3	3.5	7	18.8	9	3.5
9	MotoyaSeikaiStd-W3	モトヤ正楷書３	4	3.5	9	18.4	6	3.5
11	MotoyaShinkaiStd-W3	モトヤ新楷書３	5	3.5	5	19.4	2	3.5
8	JTC Namiki Tokubuto Kaisho	JTCナミキ特太楷書	6	3.4	3	20.7	2	3.4
12	MotoyaShinkaiStd-W5	モトヤ新楷書５	7	3.4	4	19.9	2	3.3
15	ArisawaFutoKaisho	有澤太楷書	8	3.4	6	19.3	7	3.3
13	SBTH-GFKaisho-E	白舟極太楷書	9	3.2	8	18.5	14	3.2
14	Arphic Kaisho Medium	AR楷書体M	10	3.2	11	17.3	7	3.1
2	Shinsei Kaisho CBSK1	新正楷書CBSK1	11	3.1	12	16.5	13	3.1
5	DFPKaiSho-Md	DFP中楷書体	12	3.1	10	17.7	10	3
1	A-OTF	欧体楷書	13	2.9	14	15.7	12	2.9
7	DFPKaiSho-B	DFP太楷書体	14	2.9	13	16.4	15	2.7
6	DFPKaiSho-SB	DFP中太楷書体	15	2.7	15	15.3	10	2.5
27	Soei Futo Kaisho 11*	創英太楷書体11	16	2.1	16	15.2	16	2.2
19	DFPGanKaiSho W7	DFP顔楷書W7	17	2	18	14.1	19	2.2
20	DFPGanKaiSho W9	DFP顔楷書W9	18	2	19	14.1	16	2.1
16	DFKKaiShoA Std W5	DFP華康楷書体A W5	19	1.9	21	11.2	23	2
18	DFPGanKaiSho W5	DFP顔楷書W5	20	1.9	20	13.7	18	2
24	HGGyokoku	HG行刻	21	1.8	22	11.2	21	1.9
17	DFKKaiShoB Std W5	DFP華康楷書体B W5	22	1.7	24	10.6	23	1.9
22	DFPDanKaiSho-W5	DFP談楷書W5	23	1.7	25	9.9	21	1.8
25	HGHakushuGokubutoKaishotai	HG白洲極太楷書体	24	1.7	17	14.3	26	1.8
21	DFP Shinsotai*	DFP新宋体	25	1.6	26	9.6	19	1.6
26	Soei kaisho 4*	創英楷書体４	26	1.5	23	11.1	25	1.5
23	DF Kyo Sui Japanese W3	DF鏡水 std W3	27	1.3	27	8.1	26	1.5

Figure 4.9: Result of the second questionnaire: The fonts colored orange are the final positive and negative example fonts selected (some fonts were used for stroke classification only and others for both stroke and balance classification training). Greco B (#4), MotoyaSeikaiStd-W5 (#10), Greco DB (#3), and JTC Namiki Tokubuto Kaisho (#4) were selected as positive example fonts. DF Kyo Sui Japanese W3 (#23), DFP Shinsotai (#21), HGHakushuGokubutoKaishotai (#25), and DFPDanKaiSho-W5 (#22) were selected as negative example fonts. Note that Soei kaisho 4, which was within the adoption zone for negative example fonts, was not adopted due to inaccessibility. Soei kaisho 4 (#26) was not selected as a negative example font due to inaccessibility. Also, it should be noted that while all eight fonts were utilized for training the stroke-level classification model, only four fonts were utilized for training the balance classification model. This decision was made because two of the negative example fonts, DFP shinsotai and DFP dankaisho W5, received relatively high scores in terms of balance, making them unsuitable as negative examples for training the balance classification model. Also, note that fonts with "*" at the end of the English font name have no confirmed official English name and are the author's English translation of the original Japanese font name.

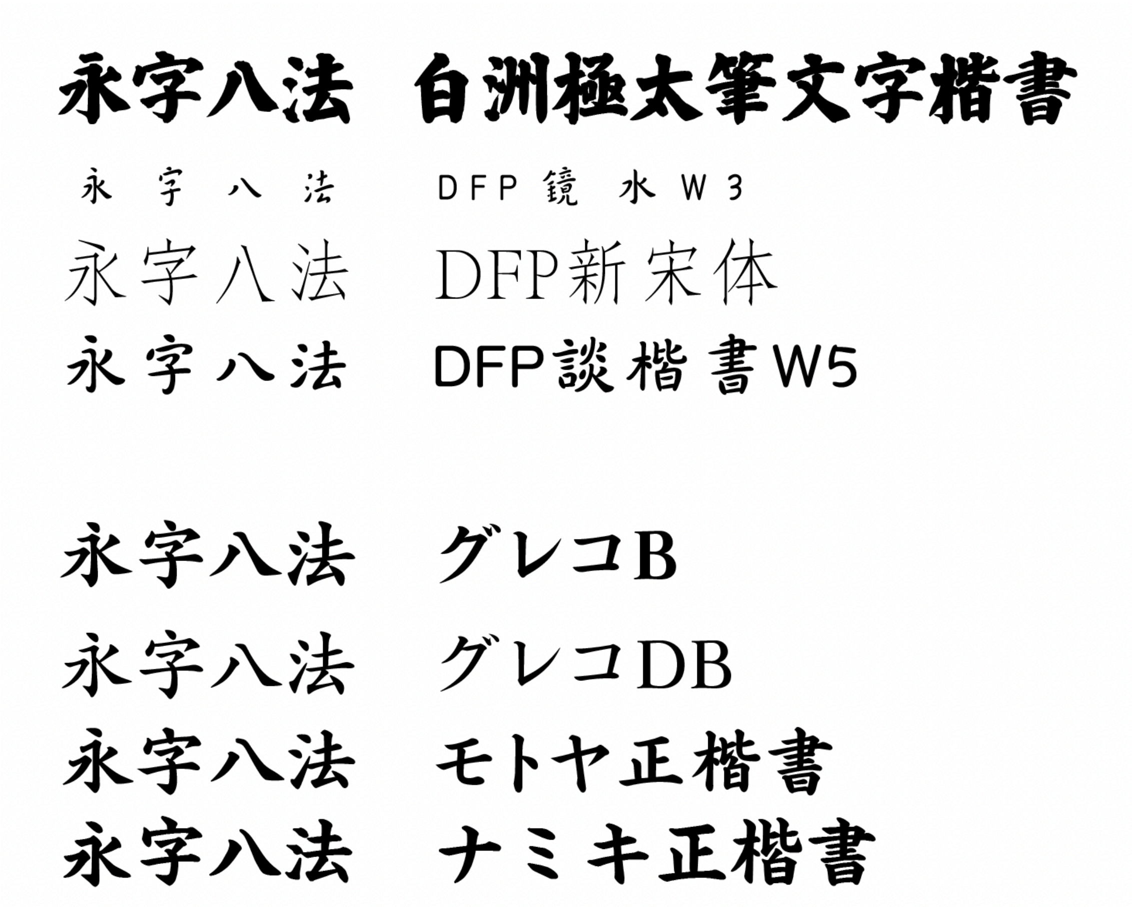

Figure 4.10: Positive and negative example fonts: In the figure, the top four are negative example fonts, and the bottom four are positive example fonts. From top to bottom, HGHakushuGokubutoKaishotai (#25), DF Kyo Sui Japanese W3 (#23), DFP Shinsotai (#21), DFPDanKaiSho-W5 (#22), Greco B (#4), Greco DB (#3), MotoyaSeikaiStd-W5 (#10), and JTC Namiki Tokubuto Kaisho (#4).

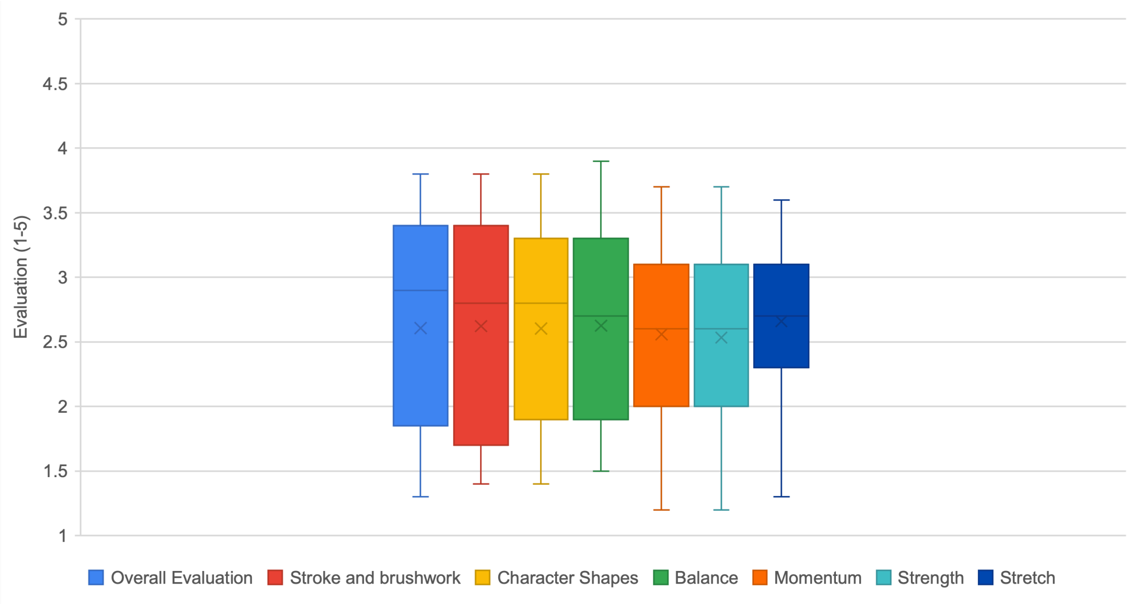

Figure 4.11: Box plot of seven perspectives: the vertical axis represents the mean value of the responses of all evaluators for font evaluation (5 grades = 1: poor, 2: slightly poor, 3: normal, 4: slightly good, 5: good).

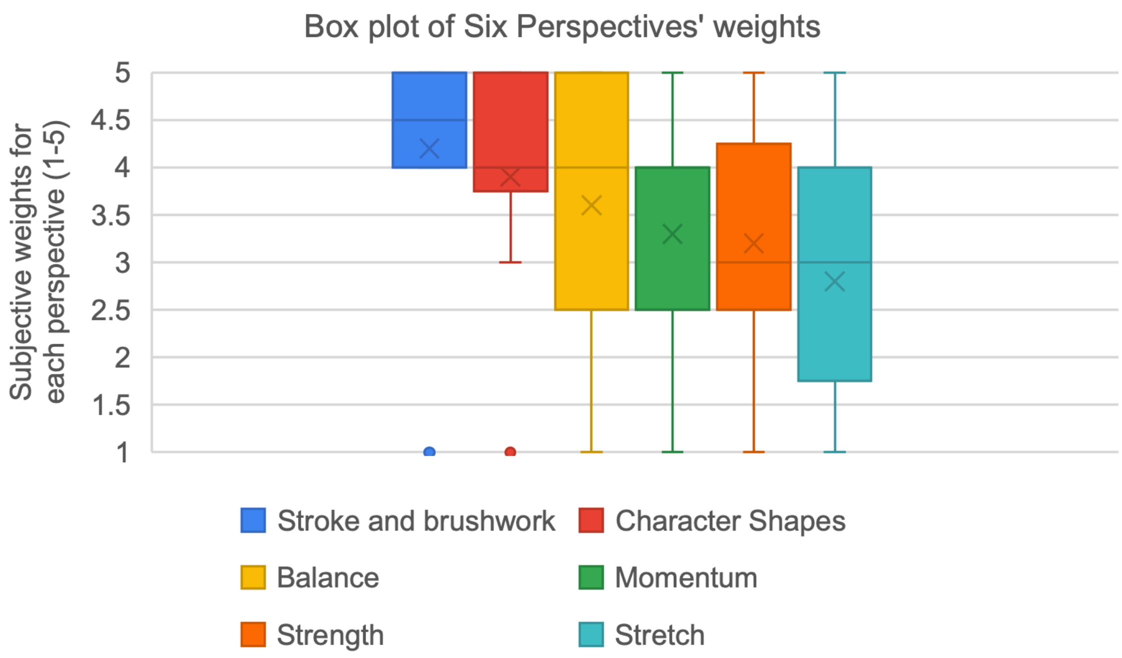

Figure 4.12: Box plot of six perspectives: The respondents were asked to rate the importance of the survey on a 5-point scale, with 1 indicating "not prioritized" and 5 indicating "prioritized"

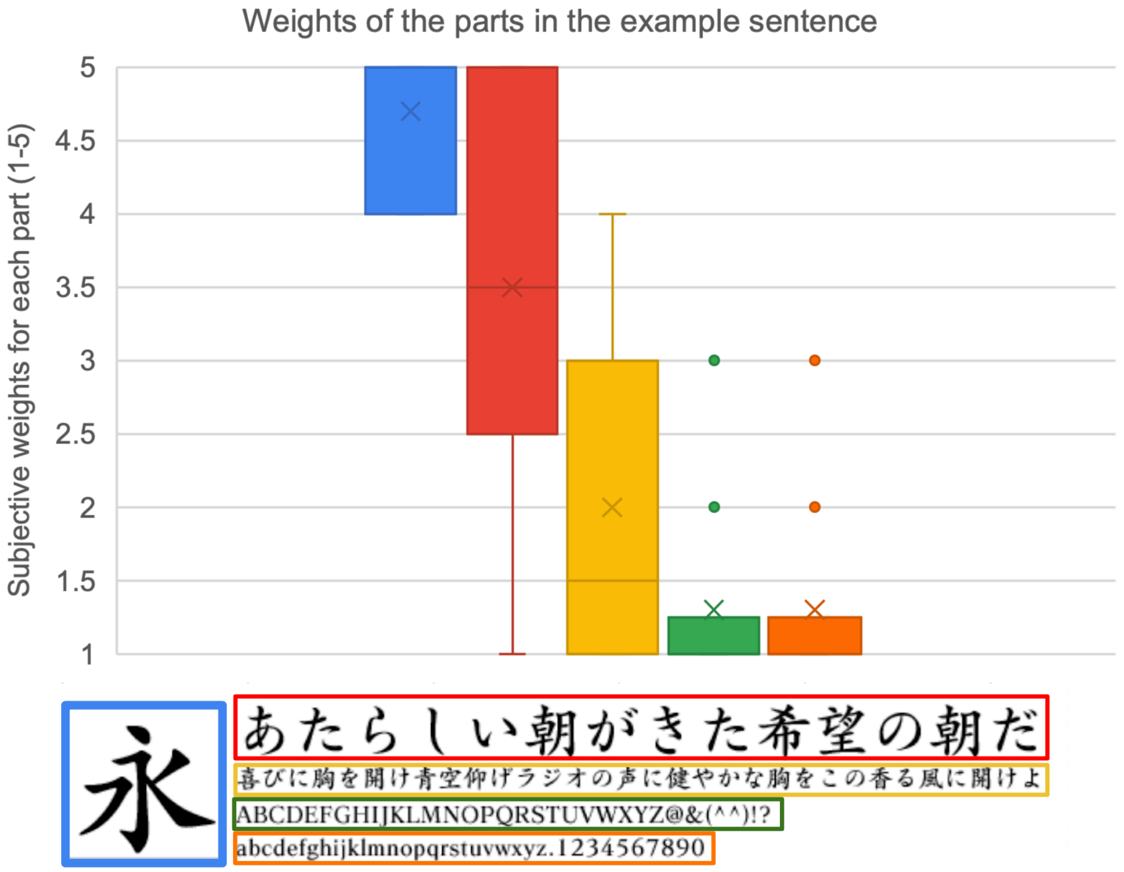

Figure 4.13: Weights of example sentence: the vertical axis is the evaluation value with "not considered" as 1 and "considered" as 5.

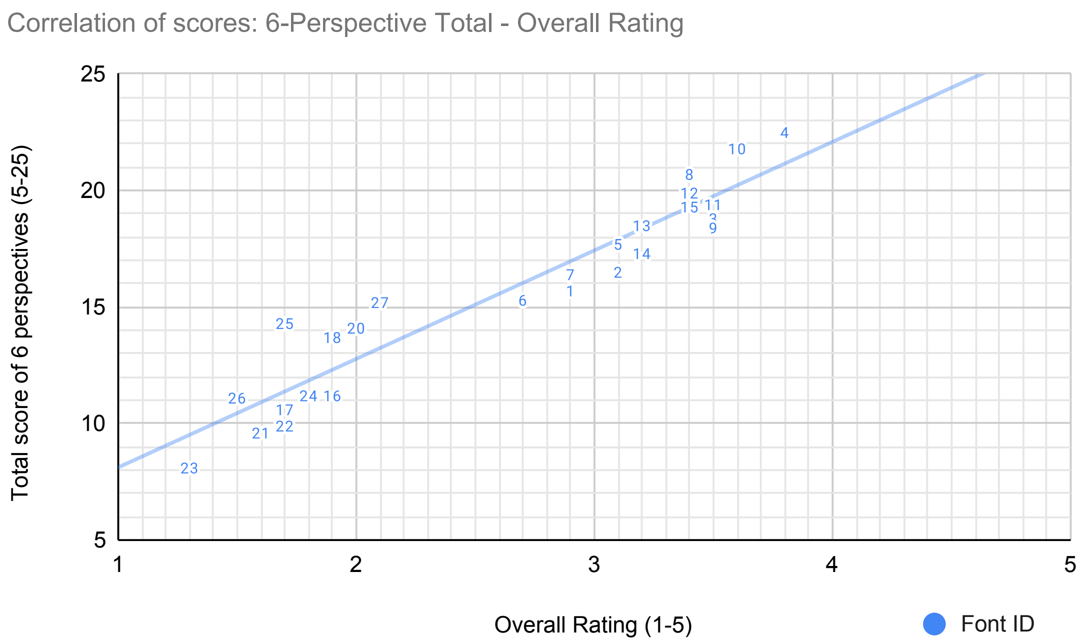

Figure 4.14: Correlation of scores: 6-Perspective Total - Overall Rating: There is a distinct correlation.

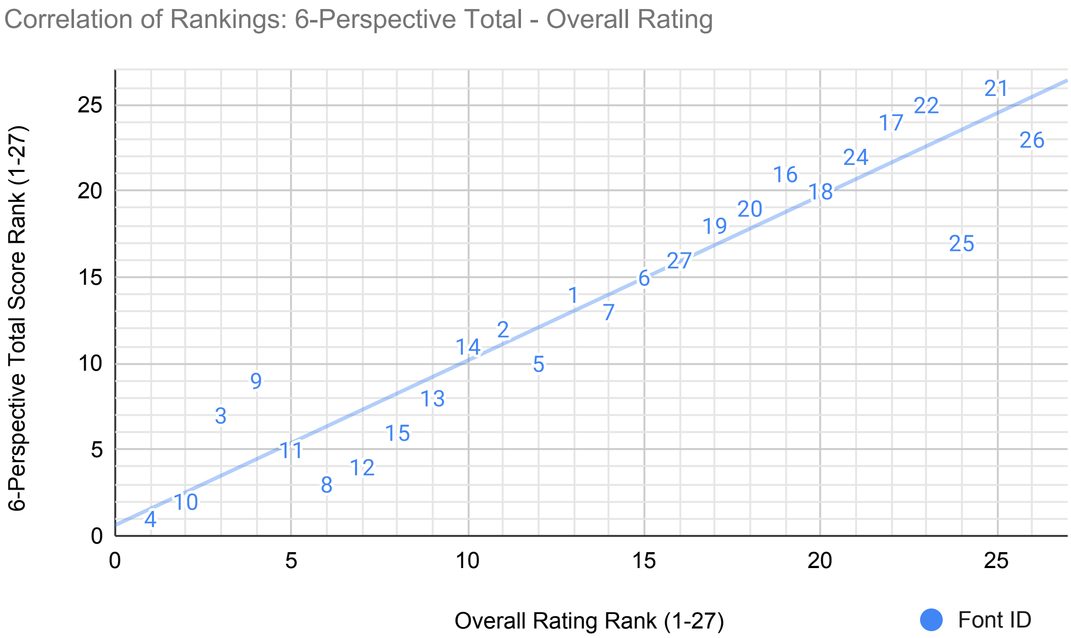

Figure 4.15: Correlation of Rankings: 6-Perspective Total - Overall Rating: There is a distinct correlation.

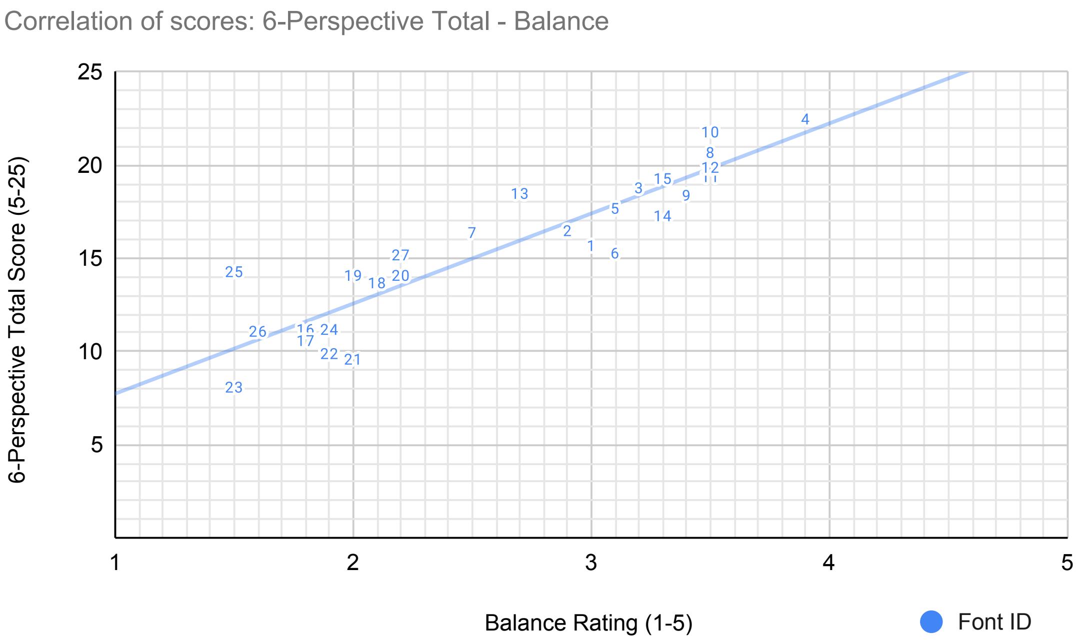

Figure 4.16: Correlation of scores: 6-Perspective Total - Balance: There is a distinct correlation.

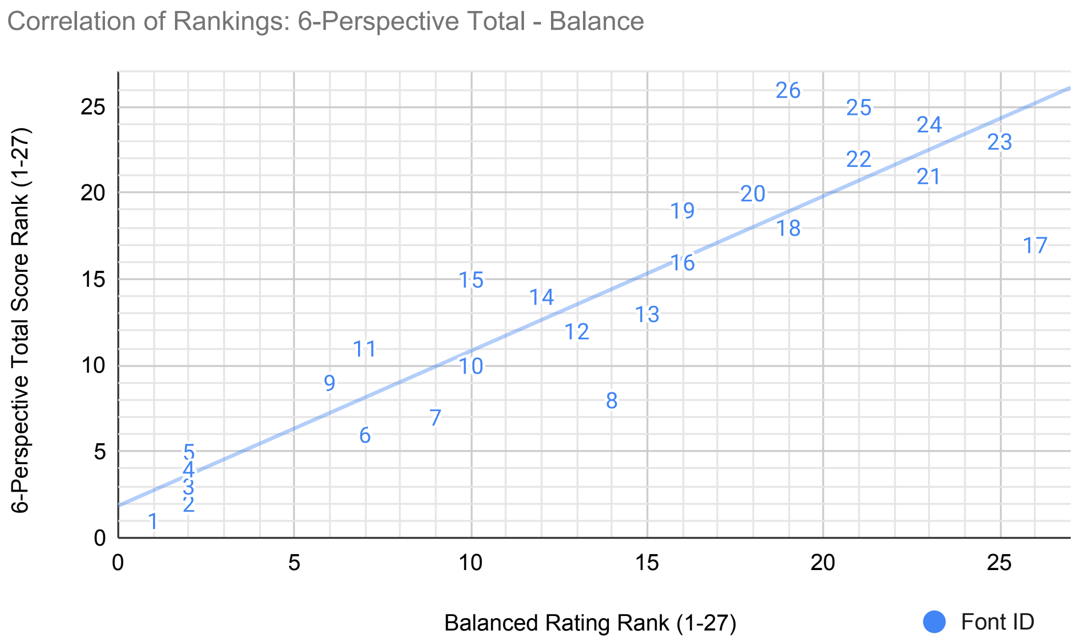

Figure 4.17: Correlation of Rankings: 6-Perspective Total - Balance: There is a distinct correlation.

Chapter 5

Implementation

5.1 Overview of implementation

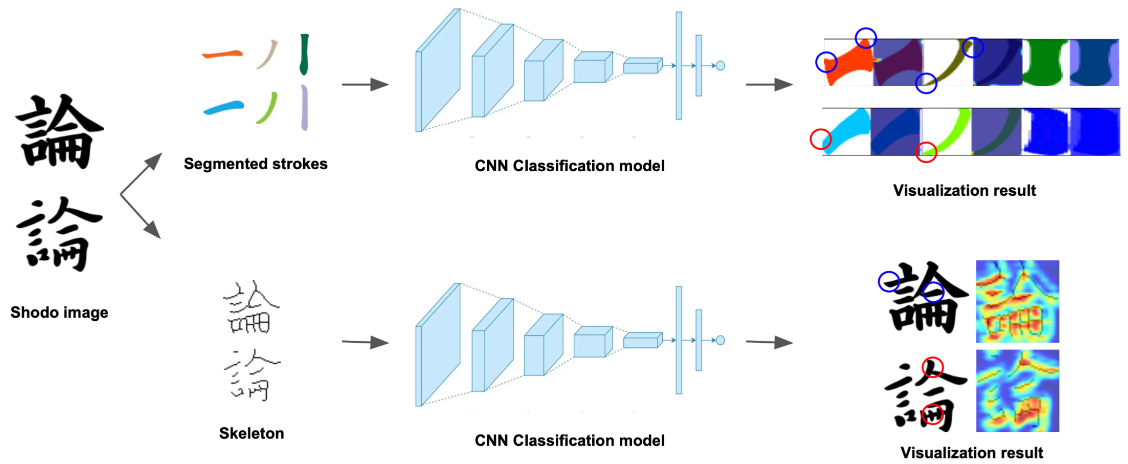

Figure 5.1: Overview of implementation with ML approach for classification and visualization: The system corrects calligraphy in terms of stroke units and overall balance. The top and bottom of each diagram section correspond to positive and negative examples, respectively. A) For stroke correction: 1) input an image of calligraphy; 2) segment the input image into stroke units; 3) have the input strokes evaluated by a binary classification model[2]that distinguishes between good and bad strokes; 4) visualize the features that the classification results are based on. B) For whole balance correction: 1) input an image of calligraphy; 2) preprocess the input image and extract a skeleton representation; 3) use the binary classification model to determine the quality of the input strokes; 4) visualize the feature values that the classification result is based on; The CNN model for both stroke and balance classification consists of 4 convolution layers and 2 fully-connected layers. Grad-CAM was used for visualization, and for both stroke and balance classification, the gradient through the fourth last convolution layer was used for visualization.

A system that uses a CNN to classify the skill level of calligraphy is a type of image recognition task that uses a Convolutional Neural Network (CNN) to extract features from images and use those features to classify them. To do this, first, collect images of calligraphy fonts that have been surveyed for skill level using a questionnaire. This training data will include images of both skilled and unskilled calligraphy fonts. Next, train a CNN. CNNs

[2]CNN image in the figure: https://towardsdatascience.com/understanding-input-and-output-shapes-in-convolution-network-keras-f143923d56ca, accessed on 2022.12.25

can extract features from images by repeating convolutional and pooling layers. These features are then used to make a final classification through fully connected layers. The CNN model for both stroke and balance classification consists of 4 convolution layers and 2 fully-connected layers. Once the training is complete, the CNN will be able to accept new images of calligraphy fonts as input and use the features extracted from those images to determine whether they are skilled or unskilled. In addition, the classification results using a CNN can be visualized to make the basis of the classification more understandable. For example, highlighting parts of an image that the CNN placed particular importance on can make it easier to understand why it was determined to be negative. In this study, grad-CAM was used for visualization, and for both stroke and balance classification, the gradient through the fourth last convolution layer was used for visualization.

5.2 Preprocessing

The training fonts' input font images were preprocessed for the following classification and visualization processing. Preprocessing consisted of three parts: Stroke segmentation which segment the overall character into individual strokes (*individual stroke data is only used for the stroke classification model. Whereas overall character data is used for the proportion classification model), normalization of the image size, and adding margin around the image. For stroke segmentation, illustrator and its third-party script[3] were used.

5.3 Stroke Classification Model Construction and Visualization

The system needs to let the classifier judge whether the input image of written Shodo is good or bad based on two perspectives of (1) individual stroke and (2) overall character. This section treats (1) individual stroke; it explains how to build the stroke classification model and visualize the evidence for the classification result. There are two steps to implement the correction process of individual stroke: Classification and visualization. (*Note that these two steps are also applied to the correction process of proportion in 5.4.)

5.3.1 Classification model by CNN

Classification is needed since the system wants to judge whether the input image of written Shodo is good or bad. A machine learning approach, especially a Convolutional Neural Network (CNN) with selected font data of positive/negative examples was adopted to build a classification model. CNN is adopted since this network specializes in pattern recognition in the image.

5.3.2 Visualization by Grad-CAM

Visualization is needed since the system wants the user to understand the judgment basis of the classification. To visualize the judgment basis of the classification, the simplest way is to colorize the input stroke image depending on the goodness/badness, where the classification model makes the decision whether the input image is good or bad. In this

[3] MojiDisassembler, an Illustrator script that disassembles text: https://sppy.stars.ne.jp/mojidisassembler

research, I adopted a visualization framework called Grad-CAM[34]. Grad-CAM produces visual explanations for decisions from CNN-based models.

5.4 Proportion Classification Model Construction and Visualization

The system needs to let the classifier judge whether the input image of written Shodo is good or bad based on two perspectives of (1) individual stroke and (2) overall character. This section treats (2) overall proportion; it explains how to build the proportion classification model and visualize the evidence for the classification result. There are two steps to implement the correction process of proportion: Classification and visualization. (*Note that these two steps are also applied to the correction process of stroke classification in 5.4.)

5.4.1 Classification model by CNN

Classification is needed since the system wants to judge whether the input image of written Shodo is good or bad. A machine learning approach, especially a Convolutional Neural Network (CNN) with selected font data of positive/negative examples is adopted to build a classification model. CNN is adopted since this network specializes in pattern recognition in the image. Also, the Deep learning framework Keras is adopted for implementation. When it comes to thinking about the proportion evaluation, the judgment of the model should only take into account proportion/balance without the biased weight/boldness of the character. Therefore, we first skeletonize the image data of the character before it is used as training data. The actual implementation process is as follows: (1) Skeletonize the overall image of the character as a positive/negative example; (2) Build the model by CNN with the Keras framework. 1006 character images for four fonts (two positive and two negative examples) are used as training data.

5.4.2 Visualization by Grad-CAM

Visualization is needed since the system wants the user to understand the judgment basis of the classification. To visualize the judgment basis of the classification, the simplest way is to colorize the input skeleton image depending on the goodness/badness, where the classification model decides whether the input image is good or bad. Grad-CAM[34] is adopted as same as stroke-level visualization in sec.5.4.

Chapter 6

System Evaluation

This chapter describes system capability, evaluation perspectives, actual evaluation, and its result.

6.1 System Capability

In this study, we developed a beginner-friendly calligraphy automatic correction system with two main functions.

6.1.1 Stroke Classification and visualization

The first function is the stroke-level classification of good/bad calligraphy and visualization of the basis for this classification. Here, the definition of good/bad classification is the binary classification model's judgment of the fonts ranked in the top/bottom 4 for overall evaluation in the calligraphy font survey. The definition of visualization of the basis for this classification is a color-coded map of the parts deemed good/bad in terms of overall evaluation using Grad-cam, sorted by degree of impact.

6.1.2 Proportion Classification and visualization

The second function is the classification of good/bad balance in calligraphy and visualization of the basis for this classification. Here, the definition of good/bad classification is the binary classification model's judgment of the fonts ranked in the top/bottom 2 for balance in the calligraphy font survey. The definition of visualization of the basis for this classification is a color-coded map of the parts deemed good/bad in terms of balance using Grad-CAM, sorted by degree of impact.

6.2 Evaluation perspectives

An evaluation was conducted with three perspectives as follows:

1. **Verification of correct classification of inputs in both non-training-used-but-same fonts and different fonts**: An experiment to verify that the system can accurately classify inputs. In other words, it is the verification of the correct classification of test data (unseen characters of the same font type as those used in training.): An experiment to verify that the system can also accurately classify input

test data that were not included in its training data, though the font type is same. Input samples of calligraphy that the system has not previously seen and verify that it can correctly classify them.

2. **Evaluation of the accuracy of visualization as a basis for classification**: An experiment to evaluate the accuracy of the system's basis for its classification. Input samples of well-written and poorly-written calligraphy and verify that the system correctly classifies each as good or bad.

6.3 Verification of correct classification of inputs

An experiment to verify that the system can accurately classify inputs. In other words, verification of the correct classification of the fonts which are not used in training but the same type as the training font: An experiment to verify that the system can accurately classify different styles of calligraphy. Input samples of calligraphy in different styles and verify that the system can correctly classify each style.

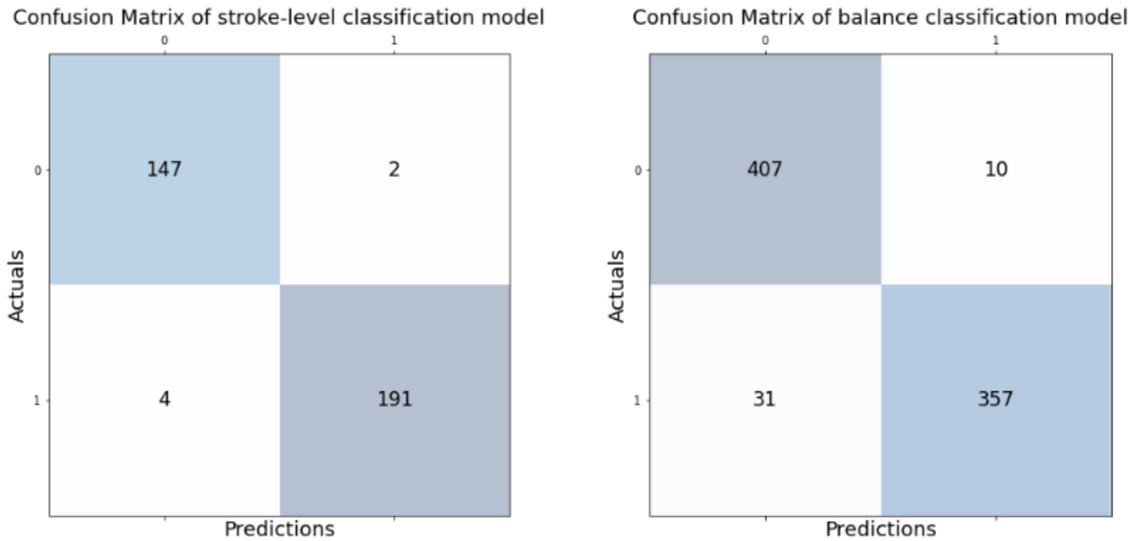

Figure 6.1: Confusion Matrix of two models: stroke-level classification model with 94% accuracy (left) and balance-level classification model with 98% accuracy (right). Here, 0 means well-written (good) and 1 means not well-written (bad).

6.3.1 Verification of correct classification of unseen characters of the same font type as those used in training.

An experiment to verify that the system can also accurately classify inputs that were not included in its training data but in the same font used in training. Input samples of calligraphy that the system has not previously seen and verify that it can correctly classify them. It was verified that both stroke evaluation and balance judgment models could correctly classify unseen characters of the same font type as those used in training.

6.4 Simple verification of an ability to produce similar visualization results to those of calligraphy teachers/Dan-holders

A simple validation was conducted to verify that the classification results are similar to those of Dan-holders and calligraphy teachers. Specifically, the system and the human calligrapher (visualization results) were tested to see how similar the two types of fonts classified as good (positive) and bad (negative) by the balanced classification model were to three Dan-holders (one of whom is a master calligrapher). Dan-holders were asked to circle in red/blue the letters that they felt were particularly bad/good (the number and shape of the circles were up to them) for those that the system judged to be negative/positive. The results are shown in fig.6.2.

It is important to note that the system visualization results are evaluated only from the perspective of balance, while the human correctors evaluate the strokes from the perspective of overall evaluation points. In some cases, the system did not give a high evaluation to the "stroke," but in the case of the stroke classification model visualization, it is highly likely that the system places importance on the "stroke" (in this verification, since the human corrections were made by looking at the entire stroke, not just the individual strokes, the system's classification was also based not on the stroke classification model but on the entire character (the balance of the character)). (Since human corrections are made by looking at the whole character rather than stroke by stroke, the system classification was compared using a balanced classification model that evaluates the balance of the entire character, not a stroke classification model.)

Also, the fact that the visualization of the judgment basis of the balance classification model and the human corrections do not match in some cases suggests that in human corrections, the quality of each stroke unit, which is an element other than balance, may be an important point in the correction.

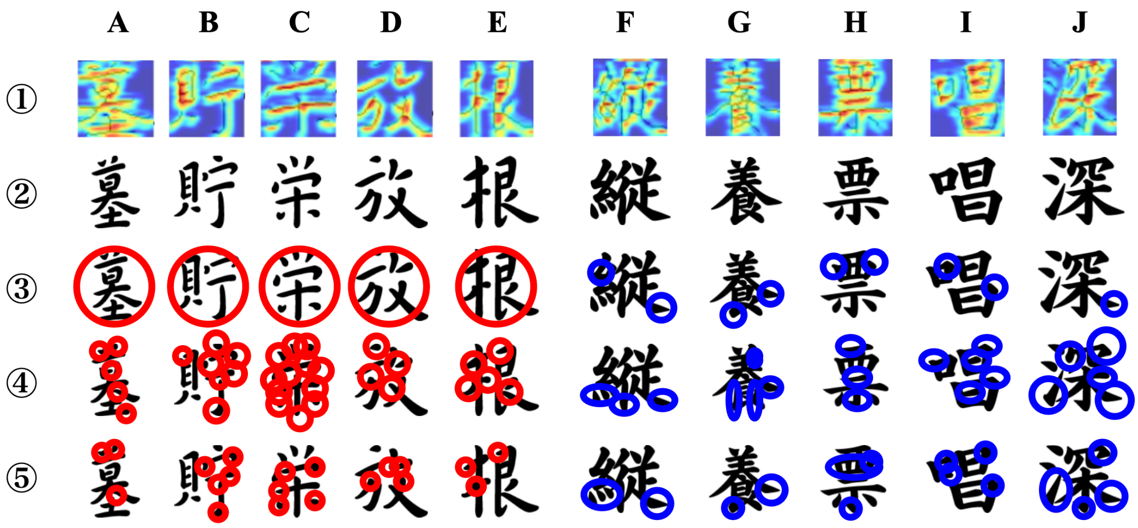

Figure 6.2: Comparison of the system and human corrections: Specifically, the degree of similarity between the system and human corrections (visualization results) for two fonts classified as bad (A-E)/good (F-J) by the balanced classification model was verified with the help of three calligraphers Dan-holders (one of whom is a master calligrapher). Columns: (1): results of visualization, (2): original font image, (3): correction by Dan-holder P1 (calligraphy instructor), (4): correction by Dan-holder P2, (5): correction by Dan-holder P3. Dan-holders were asked to circle in red/blue the letters that they felt were particularly bad/good for the letters that the system judged to be negative/positive (the number and shape of the circles were up to them). <3> A-E: P1 judged all of the letters to be particularly bad and circled the entire letter. <3, 4, 5> F: In common, the corrections of the Dan-holders students gave high marks to the last stroke, the "Harai (Lower Right Sweeping)". Visualization of the system also rated the "Harai (Lower Right Sweeping)" highly, but other parts of the system (especially the turning section) were rated highly as well. <3, 4, 5> G/J: In common, the corrective actions of the Yudansha highly evaluated the last stroke, "Harai (Lower Right Sweeping)" but other parts (especially the turning section) were also highly evaluated. The system visualization did not rate "Harai (Lower Right Sweeping)", particularly highly.

Chapter 7

Discussion

In this chapter, the followings are described: (1) Feedback from the questionnaire, (2) Validity of used font data (and segmentation), and (3) Validity of visualization.

7.1 Validity of font data for training model

7.1.1 Feedback from the questionnaire

I solicited opinions and feedback as free text in the first round of the survey (a total of 30 people responded to 99 block fonts). Some of them are divided into categories, and their feedback is shared in this section.

Feedback on how to improve the survey process system

The total number of responses to the fonts was large, therefore the following responses were received. "I had a gestalt collapse after about the 80th one(P4)" and "It was like a gestalt collapse. (P15)" Also, since the survey was conducted on Google forms, some pointed out the visibility of the surveyed fonts. "If this was somewhat bad, then maybe what we saw was normal. I know it is difficult to do it again, but I thought it would be nice to see a list. (p. 22)" and "Since I viewed the survey on my cell phone, the text was very small, so I am unsure if I could give accurate answers. (P30)"

Comments on the content of the survey (especially the reasons for their answers)

Feedback on the content of the survey, especially regarding the reasons for their answers, included the following: "Besides Ei(永), of course, but I also personally decided to look quite a bit at the differences in the way Bou(望) was written (P5)." "I think I personally preferred the ones that were neat and clean without any quirks in the form of *tome*, *hane*, and *harai* (P4)." "I felt that a certain degree of thinness tended to score higher, and the thicker the line, the lower the score (p. 9)." "There were many fonts that I felt did not have strong enough lines or were incorrect. Some of the mistakes were not acceptable even if the rest of the font was perfect, while others were acceptable if the rest was fine (p. 30)." "The authors gave high marks to fonts that reproduced the shoulder-to-arm movements that occur in the act of writing with the human right hand (p. 18)." "It is also possible that I gave favorable ratings to fonts that I liked in terms of balance (how well the letters looked) (p. 23)."

Comments on the Current State of Research and Calligraphy

We also received feedback on the current state of this research and calligraphy. "The font is somewhat unreasonable to consider it as a model for calligraphy. We sometimes see fonts distributed by schools as name models for elementary school students at the beginning of their calligraphy year, perhaps due to a lack of instructors, but they are all completely useless as calligraphy models (p. 18)."

Other Comments

"I remember that I liked the thick lines of the calligraphy teacher in my calligraphy class, so I felt that the models in school calligraphy class were too thin and different (p. 23)." "I was surprised that there are so many fonts. I was surprised at how many fonts there are. Do you want it to be unique, or do you want it to be easy to read for everyone? Fonts play an important role in commercial logos. I enjoyed looking at it (p. 30)."

7.1.2 Validity of stroke segmentation

For stroke segmentation, we used a dedicated character segmentation software that operates within Illustrator. However, not all the characters were segmented correctly. In some places where characters were present, the segmentation was insufficient (for example, stroke segmentation was not performed on strokes that should have been divided into two) or overly segmented (a single stroke was divided into two strokes). Incomplete strokes (strokes that have not been correctly segmented) can be divided into two types: (1) incomplete strokes that have a positive effect on essential learning (contribute to the learning of brush calligraphy-like characteristics) and (2) incomplete strokes that have a negative effect on essential learning (do not contribute to the learning of brush calligraphy-like characteristics). (1) refers to incomplete strokes that are established as strokes of more than one basic stroke unit. For example, it refers to cases where two basic stroke units are connected. (2) refers to incomplete strokes with less than one basic stroke unit. Strokes are divided into three categories: starting stroke, stroke process, and ending-stroke, and if any of these are missing, they will not be a characteristic stroke of calligraphy. Therefore, strokes that include these three elements are likely to be noisy when learning the characteristics of the corresponding font. The data quality can be improved by removing incomplete strokes corresponding to (2) during learning, improving classification accuracy, and visualization of higher-quality correction.

7.2 Validity of classification

7.2.1 Increase of kinds of training font

We are considering increasing the number of types of training fonts used to create the models. Specifically, we are currently using 8 types of fonts for the stroke judgment model and 4 types of fonts for the balance judgment model, but we believe that by increasing the types of fonts for negative and positive examples, we can use a more general-purpose classifier to distinguish between good and bad strokes. We believe that by increasing the number of negative and positive example fonts.

Figure 7.1: Stroke segmentation of GrekoB by moji-disassembler: Relatively correct stroke segmentation is achieved, but there are some areas where the segmentation is not perfect in dense areas. Strokes are smooth.

Figure 7.2: Stroke segmentation of HGHakushuGokubutoKaishota by moji-disassembler: From the looks of it, stroke segmentation is correctly achieved in about half of the strokes, but there are a few areas that are not segmented correctly. The strokes are jagged in places and have more of a brush stroke feel.

7.2.2 Introduction of scoring

In addition to visualization, a scoring system could be introduced to convey quantitative feedback to users. For example, the degree of how good/poor the letter is can be presented by simultaneously presenting the classification score. Another way to convey results would be to present the sum of the stroke score and the balance score to make the use of corrections more intuitive to the user.

7.2.3 Accuracy of perspective

In this study, we corrected the classifications from two perspectives, the stroke perspective and the balance-level perspective. Section 3.1 explains the rationale for choosing these perspectives. The extent to which these perspectives actually influence the overall evaluation of the characters can be shown in two ways, by focusing on the explicit and latent perspective weights, respectively. The explicit perspective weights were obtained from the second round of the questionnaire (for the 10 Dan-holders), when the respondents were asked, "How much importance did you give to each of the evaluation perspectives? (on a 5-point scale), the weight of each perspective is the respondent's average of the weight of each perspective. It can be seen that the stroke perspective is closely related to "stroke use and movement" and the balance-level perspective is closely related to the two "letter shapes" and "balance," but these items recorded higher weights than the other perspectives related to mentality. The weights by potential perspective were inferred by looking at the correlations between the total 6 perspectives - general evaluation (fig. 4.15) and the total 6 perspectives - balance-level perspective (fig. 4.17) at the time of the second questionnaire, respectively.

7.3 Validity of visualization

7.3.1 Restoration of image proportions during stroke-level visualization

Although the stroke images in which corrections are visualized are converted to 1:1 50*50 pixel squares, the original proportions vary from stroke to stroke. For example, horizontal strokes have a horizontal ratio, while vertical strokes have a vertical ratio. We believe that visualization can be made more user-friendly by restoring the strokes to the original strokes before presenting the visualization results.

7.3.2 Stroke-level visualization

Fig.7.3 demonstrates the result of visualizing the underlying foundation for evaluating strokes deemed satisfactory in the stroke classification model (excerpt from Figure X). In this figure, the visualization process exhibits a red reaction at the onset and termination of strokes, as well as at the inflections in the strokes. This implies that the starting point, ending point, and curvature points of strokes, which are regarded as crucial in calligraphy as shown in fig.7.4, serve as the basis for determining praiseworthy strokes.

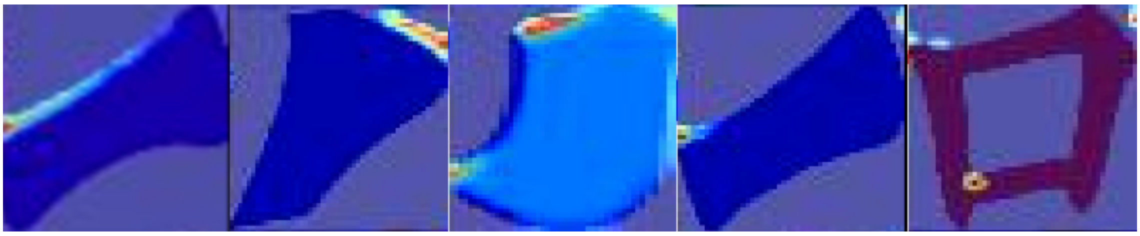

Figure 7.3: Visualization of the basis for evaluating strokes judged satisfactory in the stroke classification model. The beginning stroke is rated highly for the two strokes on the left and the second on the right. The beginning and ending strokes are highly rated for the stroke in the middle. In the rightmost set of strokes, the third stroke, the beginning stroke (or the turning stroke, if you look at it differently), is highly evaluated.

7.3.3 Balance-level visualization

As depicted in fig.7.5, the results of the Grad-CAM visualization illustrate the evidence of block letters rated as negative in terms of balance. In contrast to the visualization of stroke classification, the direct visualization of the color map for balance classification may not afford the learner an understanding of where to concentrate their efforts in correcting the issue. Therefore, visualization methods that highlight specific areas that merit particular attention are also considered.

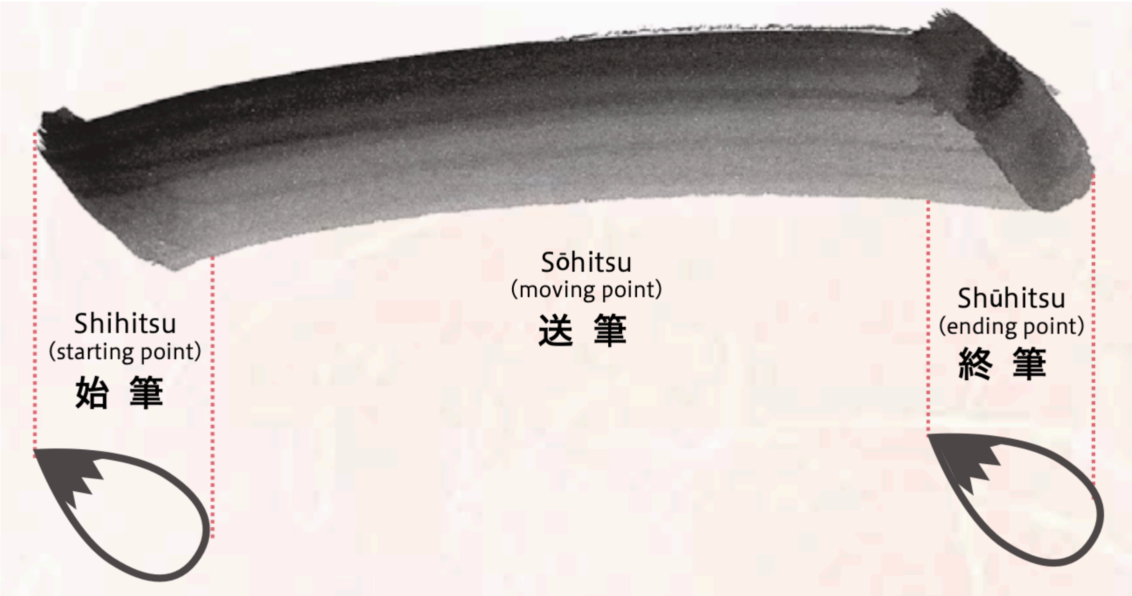

Figure 7.4: *Shihitsu*, *sohitsu*, and *shuhitsu* of horizontal stroke. Image from Japanese Calligraphy Experience Book by japan Calligraphy Education Foundation[2]

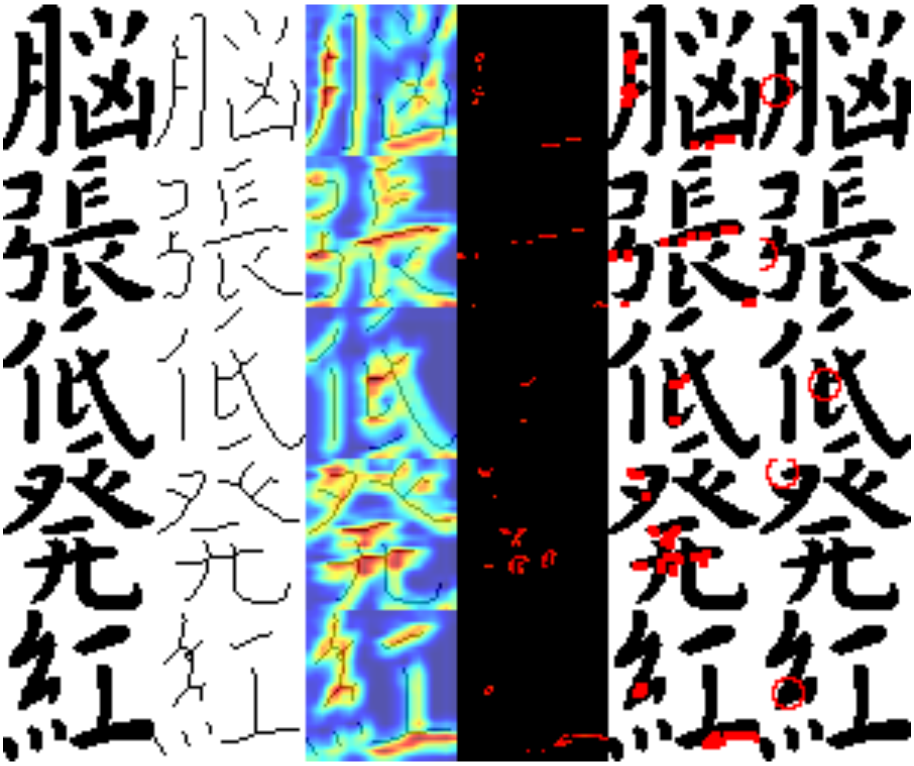

Figure 7.5: Results of Grad-CAM visualization of the rationale for the block calligraphy characters rated as poor in the balance-level perspective. From left to right: (1) Original (2) Skeleton (3) Color map output by Grad-CAM (4) Image with only the areas particularly affected by classification extracted (5) Multiple areas particularly affected by classification overlaid with the original image as red dots (6) Single areas particularly affected by classification as red circles (6)Overlaid image with the original image with red dots in a single area that has a strong influence on classification

Chapter 8

Future Work

8.1 Mobile Application as an Integrated Learning Support System

With the proposed auto-correction technique proposed in this study, an integrated learning support system for Shodo beginners could be a mobile application that helps users learn and practice the art of Shodo. The app could provide users with a variety of features to support their learning, including step-by-step tutorials and lessons on how to write the different characters and strokes used in Shodo correctly; an interactive practice area where users can try writing the characters and strokes on their own, with the app providing real-time feedback and corrections to help them improve their technique; A reference library of characters and strokes, including examples of how they should be written and their meanings; A community forum or messaging system where users can connect with other learners and experts to share tips and ask for guidance; Integration with a digital pen or stylus, allowing users to practice writing on a tablet or other device with a screen; Overall, such an app could serve as a comprehensive resource for anyone interested in learning and improving their skills in Shodo.

8.2 Application for other Penmanship Culture

There are many other penmanship cultures and traditions from around the world that could also be supported by the proposed system. Some examples might include the following:

Chinese calligraphy: This art form uses a brush and ink to write Chinese characters in a visually expressive and aesthetically pleasing way. Korean calligraphy: Also known as "hanja," Korean calligraphy is similar to Chinese calligraphy in many ways, but with its own set of characters and styles. Arabic calligraphy: This art form involves writing the Arabic alphabet and other texts in a beautiful and ornate style using a pen or brush. Western calligraphy: This encompasses a wide range of styles and traditions, including Gothic, Italic, Copperplate, and more. Overall, there are many different penmanship cultures and traditions that could be supported with the proposed technique, and such an app could be a valuable resource for anyone interested in learning and practicing these art forms.

8.3 Application to other written style

The correction algorithm proposed in this study, which consists of classification and visualization, could be applied to any typeface. However, each script has its own unique characters and writing style, so it is necessary to be able to recognize them accurately. The system is expected to accurately recognize the script to apply the automatic correction system to a Shodo work. This will require a dataset representing the script, which is a collection of images or character data representing the script. This dataset helps to train a model that can recognize the script.

Chapter 9

Conclusion

In conclusion, our study on the visualization of Shodo correction demonstrated the effectiveness of using a machine learning approach to evaluate and improve the quality of Kaisho calligraphy characters with block style. By conducting a font survey and implementing both stroke-level and balance-level classification methods, the development of a proof-of-concept prototype for an auto-correction system was realized. The proposed visualization design also showed some similarities to human correction. Overall, this research has the potential for practical applications in Shodo instruction and practice. This research has two contributions: (1) This research conducted two surveys to assess 99 kaisho fonts with 40 participants, including 26 Dan-holders. Ten Dan-holders evaluated a total of 27 selected fonts from various perspectives. The evaluation results of the top four and bottom four fonts were utilized as positive and negative examples. Still, the survey results also contained unused insights that could be applied to diverse fields such as computer science, education, and digital humanities. (2) A proof-of-concept prototype of a Shodo auto-correction system was developed using machine learning. The stroke classification model and balance judgment model had high accuracy rates of 94% and 98%, respectively, in classifying positive and negative example fonts selected through two font questionnaires as teacher data. The basis for the classification was also visualized in a way that is comprehendible to humans.

Future endeavors include (1) augmenting the quantity and diversity of training fonts to enhance the adaptability of the classification model, (2) examining visualization and scoring techniques for learners, (3) investigating correspondence with more detailed human corrections, (4) extending to Japanese fonts beyond Kaisho (block style) and to foreign penmanship cultures, and (6) constructing a comprehensive application that manages everything from camera capture to visualization feedback for offline calligraphy.

Acknowledgement

I would like to express my heartfelt gratitude to my advisor, Yoichi Ochiai, for providing crucial guidance and direction throughout my research endeavors and for showing me the path toward becoming a researcher. Additionally, I deeply appreciate Hiroko Terasawa, who graciously supported me and served as a co-advisor in my research endeavors. Tatsuki Fushimi, who aided me in analyzing font survey data and proofreading the paper, was also invaluable. Special thanks are also due to Naruya Kondo, who played a key role in the technical implementation of machine learning in this research. Furthermore, I received valuable input and guidance from Akiho Ogawa, Kenta Yamamoto, and Xanat Vargas Meza in the initial stages of planning the automated proofreading system. Chinatsu Ozawa provided support with research tools, and I also thank Jacob Zhang, Kazuya Izumi, and Atsushi Shinoda for their technical input. Manabu Nakano from Pixie Dust Technologies, Inc. was also greatly helpful. I am grateful to the members of the lab who consistently supported me and made my research experience enjoyable. Furthermore, I am grateful to all participants who took the time to respond to our survey. I am also thankful to Keigo Yoshinaga, who always supports me in the overseas calligraphy distance teaching project, and all participants in the project. Finally, I want to express my deep appreciation to my family, particularly my mother, who provided valuable calligraphy-related advice as a calligraphy instructor, and to my beloved dog, Chibi, who always comforts me.

References

[1] KS Vincent Poon. Vincent's calligraphy - preserving authentic traditional chinese culture in the west, 2013-2022. Accecced on Dec 20, 2022.

[2] Japan Calligraphy Education Foundation. *Japanese Calligraphy Experience Book*. 2018.

[3] Daniel Berio, Frederic Fol Leymarie, Paul Asente, and Jose Echevarria. Strokestyles: Stroke-based segmentation and stylization of fonts. *ACM Transactions on Graphics (TOG)*, 41(3):1–21, 2022.

[4] Hock Soon Seah, Zhongke Wu, Feng Tian, Xian Xiao, and Boya Xie. Artistic brush-stroke representation and animation with disk b-spline curve. In *Proceedings of the 2005 ACM SIGCHI International Conference on Advances in computer entertainment technology*, pages 88–93, 2005.

[5] Gan Lin, Zhihua Guo, Fei Chao, Longzhi Yang, Xiang Chang, Chih-Min Lin, Changle Zhou, V Vijayakumar, and Changjing Shang. Automatic stroke generation for style-oriented robotic chinese calligraphy. *Future Generation Computer Systems*, 119:20–30, 2021.

[6] Fang Cao, Zhongke Wu, Pengfei Xu, Mingquan Zhou, and Xuefeng Ao. A learning system of qi gong calligraphy. In *Proceedings of the 14th Global Chinese Conference on Computers in Education (GCCCE' 10)*, 2010.

[7] Rongju Sun, Zhouhui Lian, Yingmin Tang, and Jianguo Xiao. Aesthetic visual quality evaluation of chinese handwritings. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*, 2015.

[8] Songhua Xu, Hao Jiang, Francis CM Lau, and Yunhe Pan. Computationally evaluating and reproducing the beauty of chinese calligraphy. *IEEE intelligent systems*, 27(03):63–72, 2012.

[9] Mengdi Wang, Qian Fu, Xingce Wang, Zhongke Wu, and Mingquan Zhou. Evaluation of chinese calligraphy by using dbsc vectorization and icp algorithm. *Mathematical Problems in Engineering*, 2016, 2016.

[10] 宮地功. Ahp による毛筆書写の評価支援システムの開発. **岡山理科大学情報処理センター研究報告**, 1996.

[11] Chin-Chuan Han, Chih-Hsun Chou, and Chung-Shiou Wu. An interactive grading and learning system for chinese calligraphy. *Machine Vision and Applications*, 19(1):43–55, 2007.

[12] 野辺昌史, 藪田直紀, 鈴木雅人, 山下晴雨, 市村洋, et al. ペン字・書道を事例とした遠隔教育における文字の美的評価の一手法. **第 *65* 回全国大会講演論文集**, 2003(1):353–354, 2003.

[13] Jungpil Shin, Toshiki Okuyama, and Keunsoo Yun. Sensory calligraphy learning system using yongzi-bafa. In *Ifost*, volume 2, pages 128–131. IEEE, 2013.

[14] 伴藤信一郎 and 西田茂生. 書道自動添削システムの開発. 第 27 回「学生会員卒業研究発表講演会」, 2020.

[15] 西岡亜美 and 西田茂生. 書道自動添削スマートフォンアプリケーションの開発. 第 28 回「学生会員卒業研究発表講演会」, 2020.

[16] Hao Bai and Xiwen Zhang. Adaptive correcting strokes extracted from chinese characters in digital ink of non-native writers based on comprehensive visualization. In *2017 International Conference on Computer Network, Electronic and Automation (ICCNEA)*, pages 124–128. IEEE, 2017.

[17] Hao Bai. Visualizing results of digital ink stroke matching in chinese based on multi-perceived hierarchy. In *IOP Conference Series: Materials Science and Engineering*, volume 490, page 042032. IOP Publishing, 2019.

[18] Chao-Ming Wang and Yu-Chen Chen. Design of an interactive mind calligraphy system by affective computing and visualization techniques for real-time reflections of the writer ’ s emotions. *Sensors*, 20(20):5741, 2020.

[19] Kengo Tanaka, Taisuke Ohshima, and Yoichi Ochiai. Spring-pen: reproduction of any softness with the 3d printed spring. In *SIGGRAPH Asia 2017 Posters*, pages 1–2. 2017.

[20] Kengo Tanaka, Kohei Ogawa, Tatsuya Minagawa, and Yoichi Ochiai. Design method of digitally fabricated spring glass pen. In *ACM SIGGRAPH 2018 Studio*, pages 1–2. 2018.

[21] Noriaki Muranaka, Takafumi Yamamoto, and Shigeru Imanishi. A calligraphy mastering support system using virtual reality technology and its learning effects. *IEEJ Transactions on Fundamentals and Materials*, 123(12):1206–1216, 2003.

[22] Yufan Yang, Weiming Chen, Lingxiang Zhou, Bote Zheng, Wei Xiao, Yanwei Huang, and Zhenglong Sun. A hybrid control framework teaching robot to write chinese characters: from image to handwriting. In *2021 IEEE 17th International Conference on Automation Science and Engineering (CASE)*, pages 1161–1166. IEEE, 2021.

[23] Lucie Brozková. Education program for calligraphy using hci devices. *Faculty of Informatics, Masaryk Univesity*, 2008.

[24] Takahiro Shichinohe, Tetsuo Yamabe, Takahiro Iwata, and Tatsuo Nakajima. Augmented calligraphy: experimental feedback design for writing skill development. In *Proceedings of the fifth international conference on Tangible, embedded, and embodied interaction*, pages 301–302, 2010.

[25] 山岡信介 and 平俊男. 918 力覚デバイスを用いた仮想毛筆書字による技能的動作の特徴抽出 (gs-16 mems (2)). In **関西支部講演会講演論文集** *2009.84*, pages _9–18_. 一般社団法人 日本機械学会, 2009.

[26] 鈴木祐一, 井口寧, and 堀口進. 力触覚提示装置を用いた書道用筆モデルの提案. **日本バーチャルリアリティ学会論文誌**, 10(4):573–580, 2005.

[27] Zhu Li, Li Jun, Hu Tao, and Xiang Jun. Calligraphy imitation system based on virtual brush and scale-invariant feature transform. In *2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing*, pages 1936–1941. IEEE, 2015.

[28] 堤勇介, 中島明, and 早川義一. 多指ハンドロボットを用いた書道における筆の把持及び運筆の制御. In **自動制御連合講演会講演論文集 第** *54* **回自動制御連合講演会**, pages 210–210. 自動制御連合講演会, 2011.

[29] 島田英之, 島田恭宏, 大倉充, et al. 毛筆で筆記可能な仮想書道システムの開発. **情報処理学会論文誌**, 47(12):3392–3401, 2006.

[30] Kimiko Ryokai, Stefan Marti, and Hiroshi Ishii. I/o brush: drawing with everyday objects as ink. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 303–310, 2004.

[31] Wabunfont. The japanese font liblary for macos and windows, 2001-2010. Accecced on Dec 14, 2022.

[32] S. Sato, G.A. Roshi, and A.O. Sato. *Shodo: The Quiet Art of Japanese Zen Calligraphy, Learn the Wisdom of Zen Through Traditional Brush Painting*. Tuttle Publishing, 2014.

[33] 宮地功, 清水誠一, 岸誠一, and 梶浦文夫. 書道作品についての評語の得点化の試み. **教育情報研究**, 8(1):11–19, 1992.

[34] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626, 2017.

Appendix

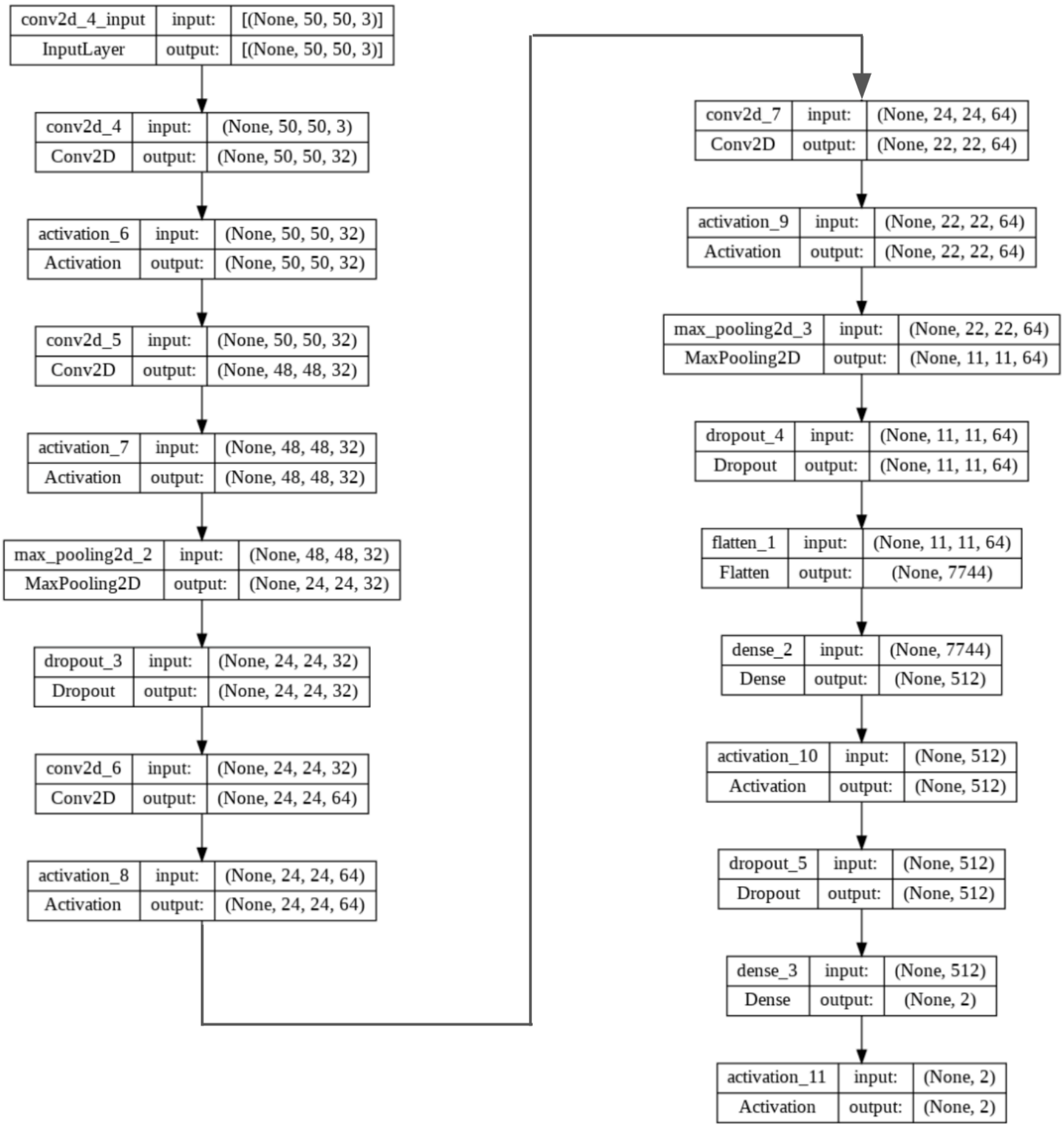

Figure 1: Summary of stroke-level classification model

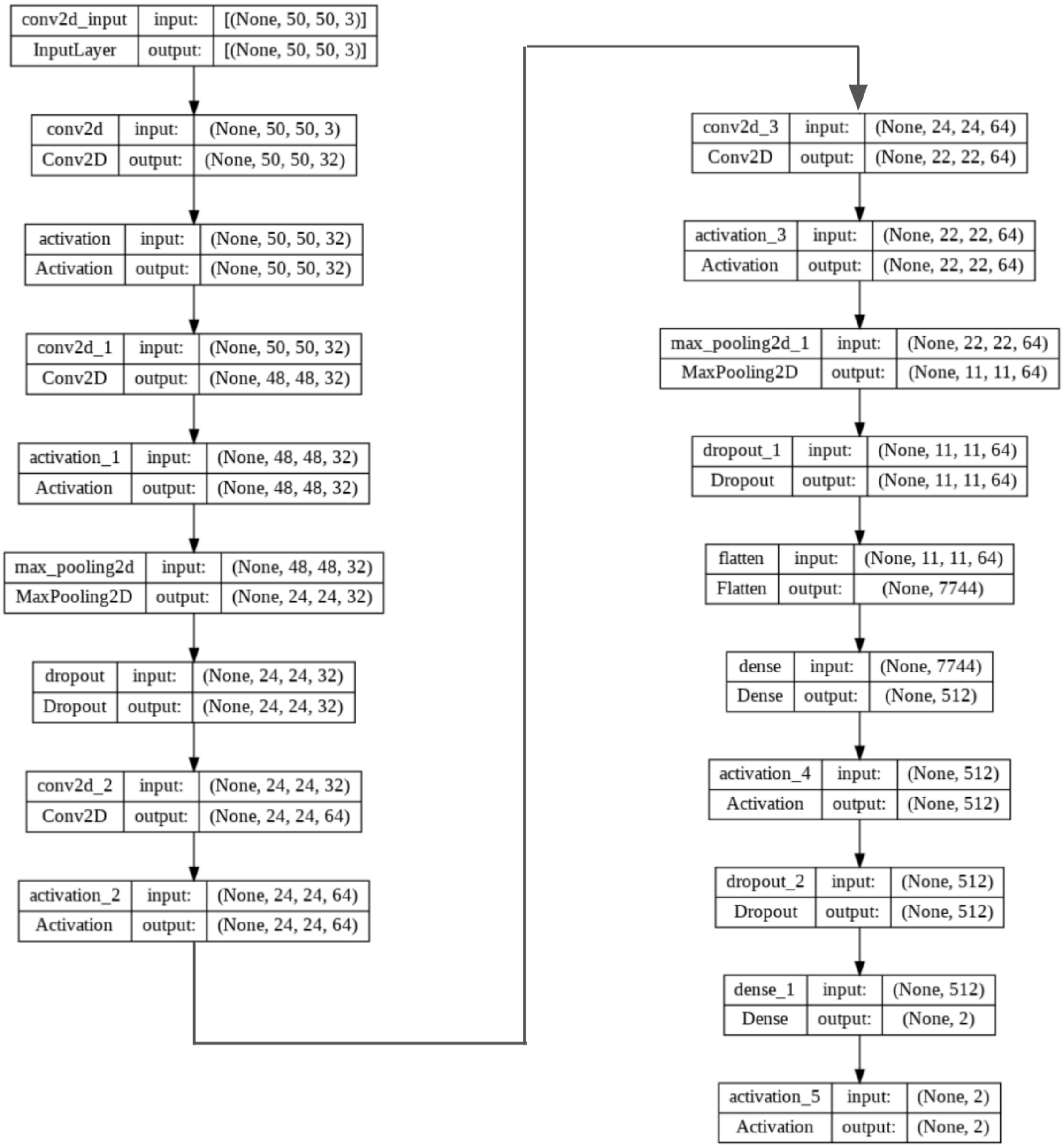

Figure 2: Summary of balance classification model

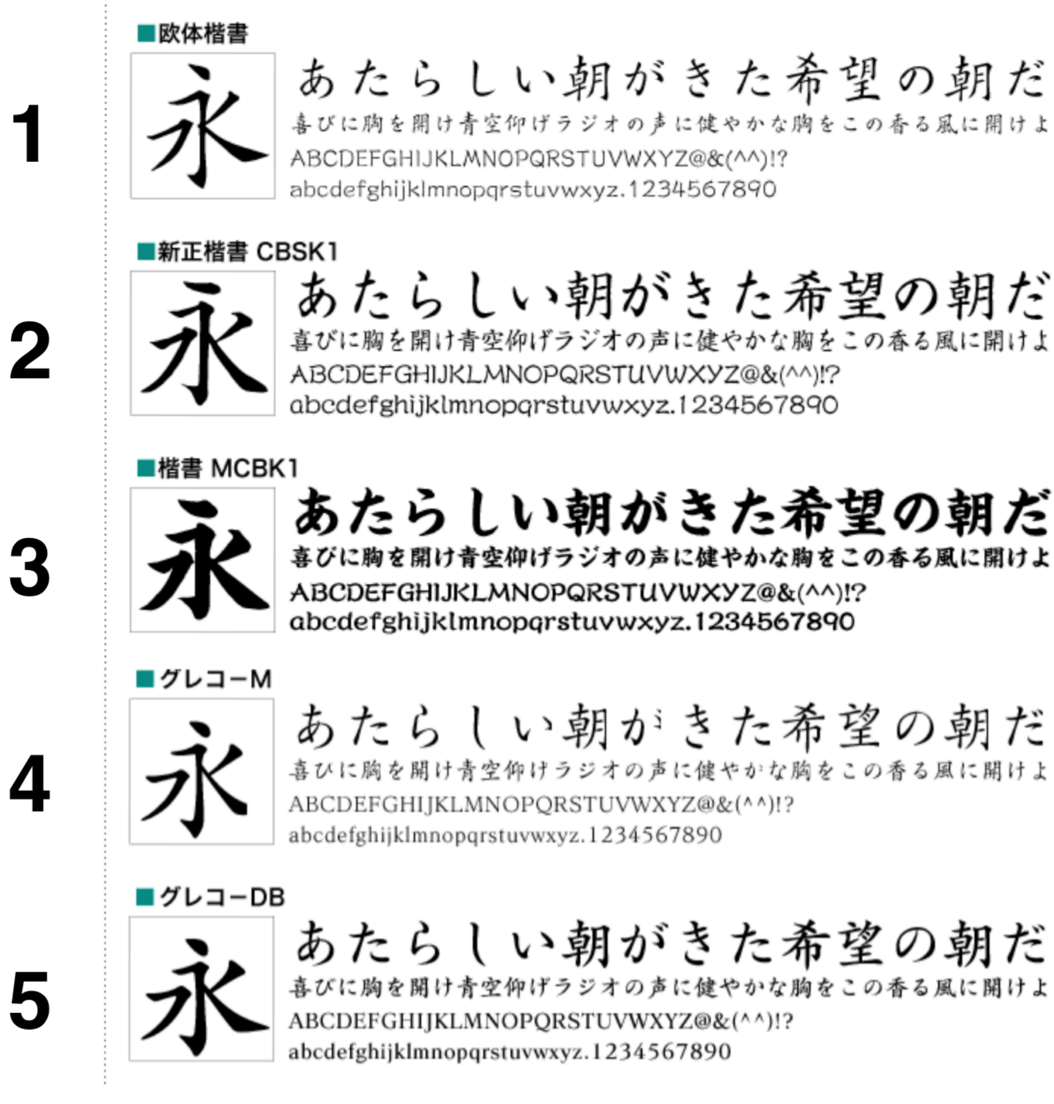

Figure 3: Fonts covered by the first round survey: 1-5

Figure 4: Fonts covered by the first round survey: 6-10

Figure 5: Fonts covered by the first round survey: 11-15

Figure 6: Fonts covered by the first round survey: 16-20

Figure 7: Fonts covered by the first round survey: 21-25

Figure 8: Fonts covered by the first round survey: 26-30

Figure 9: Fonts covered by the first round survey: 31-35

Figure 10: Fonts covered by the first round survey: 36-40

Figure 11: Fonts covered by the first round survey: 41-45

Figure 12: Fonts covered by the first round survey: 46-50

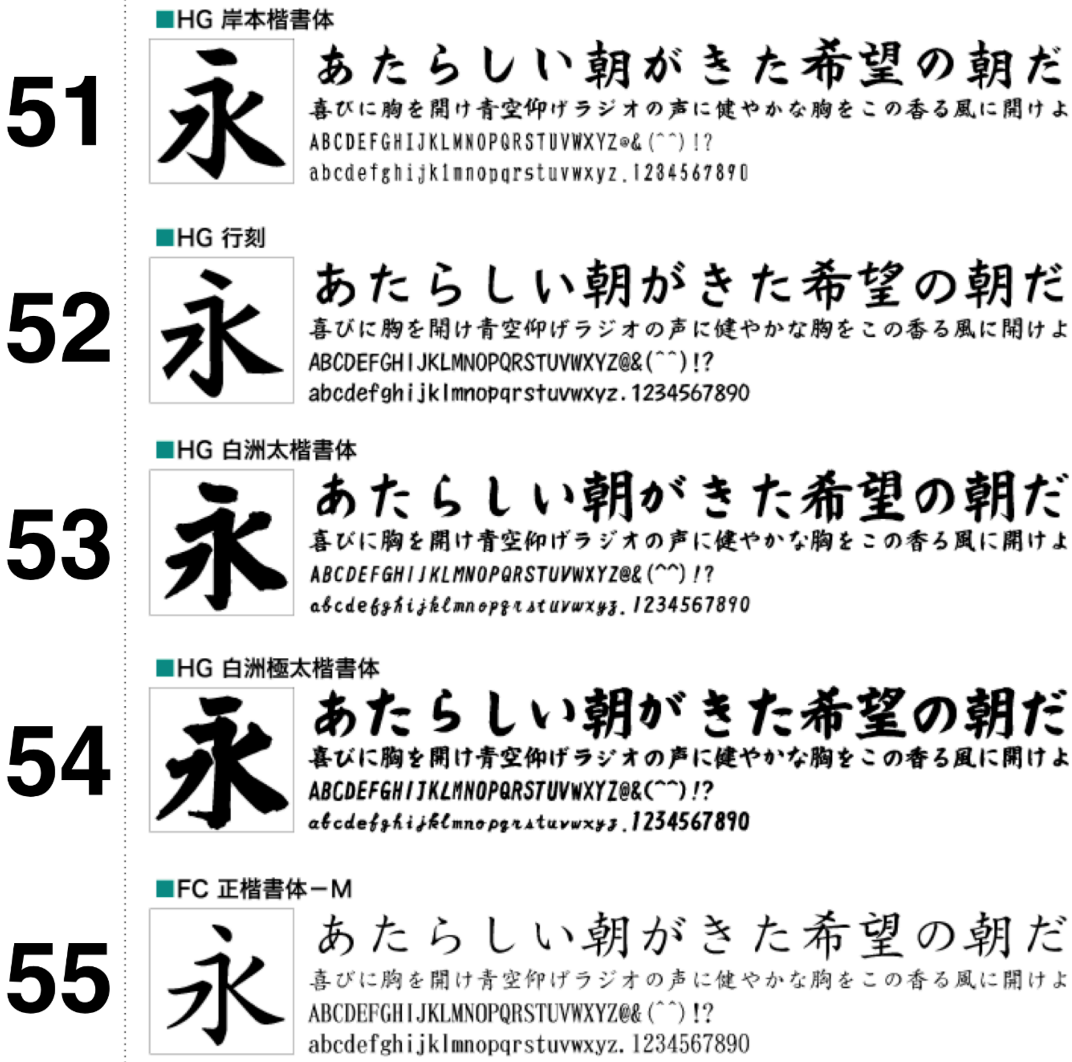

Figure 13: Fonts covered by the first round survey: 51-55

Figure 14: Fonts covered by the first round survey: 56-60

Figure 15: Fonts covered by the first round survey: 61-65

Figure 16: Fonts covered by the first round survey: 66-70

Figure 17: Fonts covered by the first round survey: 71-75

Figure 18: Fonts covered by the first round survey: 76-80

Figure 19: Fonts covered by the first round survey: 81-85

Figure 20: Fonts covered by the first round survey: 86-90

Figure 21: Fonts covered by the first round survey: 91-95

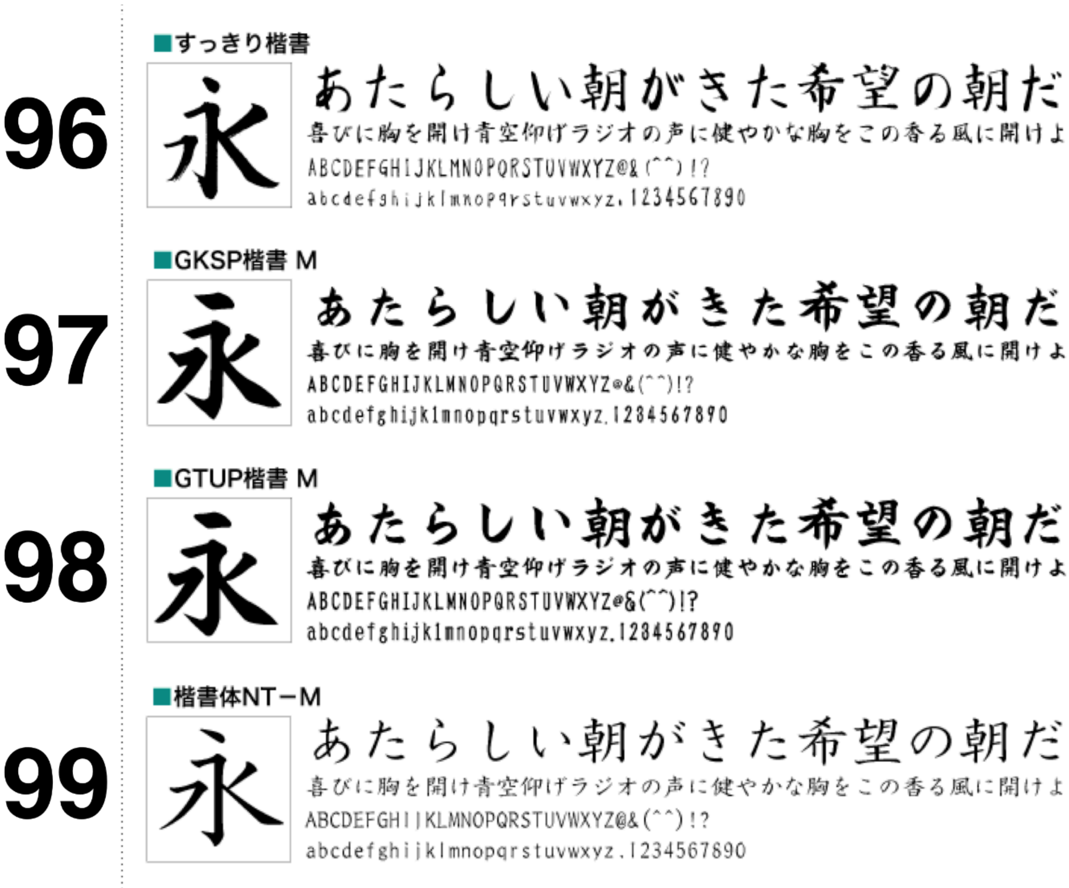

Figure 22: Fonts covered by the first round survey: 96-99