electronics

MDPI

Article

Detection Model of Hangul Stroke Elements: Expansion of Non-Structured Font and Influence Evaluation by Stroke Element Combinations

Soon-Bum Lim [1], Jongwoo Lee [1], Xiaotong Zhao [2] and Yoojeong Song [3,*]

[1] Department of IT Engineering, Research Institute of ICT Convergence, Sookmyung Women's University, Cheongpa-ro 47-gil 100, Yongsan-gu, Seoul 04310, Republic of Korea
[2] School of Digital Arts, Dalian Neusoft University of Information, Dalian 116023, China
[3] School of Computer Science, Semyung University, Jecheon 27136, Republic of Korea
* Correspondence: yjsong@semyung.ac.kr

Abstract: With the increase of various media, fonts continue to be newly developed. In Korea, numerous 'Hangul' fonts are also being developed, and as a result, the need for research on determining the similarity between fonts is emerging. For example, when creating a document, the font to be used must be downloaded from each computing environment. However, this is a very cumbersome process. If there is a font that is not supported in the system, the above problem can be easily solved by recommending the most similar font that can replace it. According to this need, we conducted various prior studies for similar font recommendations. As a result, we developed a 'stroke element' that exists in each consonant and vowel in Korean font and developed a font recommendation model using a stroke element. However, there is a limitation in that the existing research was studied only for the structured fonts corresponding to the printed type. Additionally, the font size was not considered in the font recommendation. In this study, two experiments were conducted to expand the font recommendation model by supplementing the limitations of existing studies. First, in order to enable similar font recommendations based on the stroke element even in fonts with various shapes, the font was classified according to the shape, and the stroke elements in each classification were detected. Second, when the font sizes were different, the change in the font recommendations result based on the stroke element was analyzed. In conclusion, we found that it was necessary to find a plan to extract stroke elements for font recommendation of fonts that do not belong to standard fonts. In addition, since the influence of the stroke element varies depending on the size of the font, we propose a stroke element weight model that can be used for recommendation by reflecting it.

Keywords: Hangul stroke element; font similarity evaluation; weight calculation model; font recommendation

Citation: Lim, S.-B.; Lee, J.; Zhao, X.; Song, Y. Detection Model of Hangul Stroke Elements: Expansion of Non-Structured Font and Influence Evaluation by Stroke Element Combinations. *Electronics* **2023**, *12*, 383. https://doi.org/10.3390/electronics12020383

Academic Editors: Gwanggil Jeon and Chiman Kwan

Received: 25 November 2022
Revised: 25 December 2022
Accepted: 9 January 2023
Published: 12 January 2023

Correction Statement: This article has been republished with a minor change. The change does not affect the scientific content of the article and further details are available within the backmatter of the website version of this article.

1. Introduction

In Korea, the development of Korean fonts is actively progressing, and the number and form of fonts are increasing day by day due to the development of software specializing in Korean font design. As such, the design and development of fonts are in the spotlight as a new content marketing area, which proves that the influence of fonts is rapidly increasing [1,2]. However, while the importance of font influence is growing, research on recommending similar fonts for Korean fonts is very insufficient. Conventionally, font providers or systems store font information simply by the name of the font or the name of the font maker according to the criteria selected by the font company. Therefore, to find similar fonts or alternative fonts, it is necessary to use them passively, such as checking them one by one. However, it is impossible to check all the thousands of Korean fonts [3].

In addition, copyright-related problems are also being raised to determine whether the design of the produced font is like that of the existing font. This is a sensitive issue to font designers and is also a major problem in the development of the Korean font market [4]. To overcome the above situation, research is being conducted on the development of a system that recommends fonts based on the similarity between fonts. However, these problems can prevent copyright problems if there is a system that recommends fonts based on the similarity between fonts, and can easily perform tasks such as replacing broken fonts with similar fonts.

According to this need, we conducted various prior studies [5–7] for similar font recommendations. As a result, we developed a 'stroke element' that exists in each consonant and vowel in Korean font and developed a font recommendation model using a stroke element. However, there were deficiencies and limitations in previous studies, and we conducted in-depth research to supplement this and expand the contents of the study.

The first limitation is that all existing studies [5,6] have been studied only on formal fonts corresponding to printed materials. Formal fonts refer to structured fonts used in printed materials, and in this paper, they are called structured fonts. However, there are countless irregular fonts such as handwriting and graphic typefaces, and the font recommendation system should be able to recommend such irregular fonts. In this study, to make it possible to recommend the fonts with irregular shapes in the font recommendation system, we classified Korean fonts into three groups based on the form of the font: structured font (A), semi-structured font (B), and unstructured font (C). The classified Korean fonts are applied to a model that automatically extracts stroke elements of fonts for each group, and the stroke elements are automatically extracted. Finally, the accuracy of extraction of stroke elements for each group is compared with each other to determine whether it is possible to add to the font recommendation system.

As a second limitation, since the number of stroke elements used for font recommendation is large, it was necessary to check whether all eight stroke elements were necessary or whether they could be partially applied because the performance was different depending on the combination. The font size was also not considered in the font recommendation. A previous study [7] automatically extracted eight characteristic stroke elements of Hangul and recommended similar fonts based on stroke elements. However, the font size is not considered and similar fonts are calculated, and ground-truth data is used to evaluate. The usability of similar font recommendation results is also not verified, so more accurate similar font calculation needs to be supplemented. The font is used in different sizes depending on the displayed medium, situation, and purpose, which makes it feel different in size depending on the user, although it is a similar font that is different in size. Therefore, it can be said that the ground-truth of previous studies that did not consider this situation is less reliable. Therefore, we made a new ground-truth that is constructed by presenting a user questionnaire configured by dividing the font size, and the reliability of ground-truth is secured through verification thereof. In addition, we would like to propose a similar font recommendation model considering the size of the letters.

The composition of this paper is as follows. Section 2 describes related works about font recommendation systems and previous studies. Section 3 describes the deep learning model used for stroke element detection. Section 4 outlines an experiment to solve the first limitations of the previous studies, and explains the model and data used to detect stroke elements and the extension of font classification. Section 5 describes the experiment to solve the second limitation, which is a newly conducted user study, ground-truth, stroke element impact evaluation, and stroke element impact by size. Section 6 describes the results and discussions of the experiments conducted in Sections 4 and 5, and finally describes the conclusions in Section 7.

2. Related Works

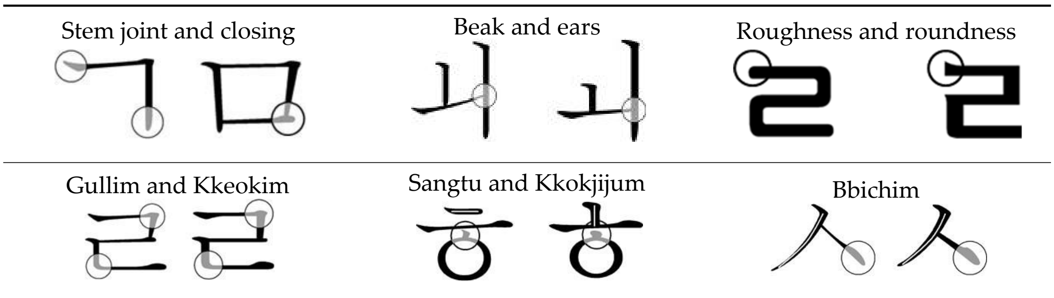
The latest computing technology, artificial intelligence, is widely used in research related to similar Hangul-recommended fonts. This section describes research related to Hangul stroke elements along with artificial intelligence technology used in font recommendation.

2.1. Stroke Element as a Criterion for Judging Font Similarity

2.1.1. A Study on the Detection of Korean Stroke Elements

In a previous study [5–7], we defined the design elements of Hangul and automatically detected them using a deep learning model. The criteria for determining the similarity between the types of Hangul are derived based on research conducted by the Korean Intellectual Property Office [8]. In this study, the morphological characteristics of Hangul and the classification system of Hangul type were summarized, and fonts with similar visual characteristics were classified for typography design experts. As a result, the criteria for determining the similarity of the Korean type were proposed. The criteria for determining the similarity based on the Korean font correspond to elements that can be characteristically found in consonants and vowels, and examples are shown in Table 1.

Table 1. Character-based similarity judgment criteria example. Each characteristic is a unique characteristic of the consonant shape of Hangul, which corresponds to the part shown in the picture.



Other related studies include the historical study of the change of type [9], the study of representative characters for efficient type design [10], and the study of shape discrimination factors [11]. Summarizing the shape characteristics mentioned in these studies, there are about 70, and among them, eight stroke elements that most intensively express structural information of Hangul were selected through K-means, PCA [12], and similarity measurement model analysis. The eight stroke elements are shown in Figure 1.

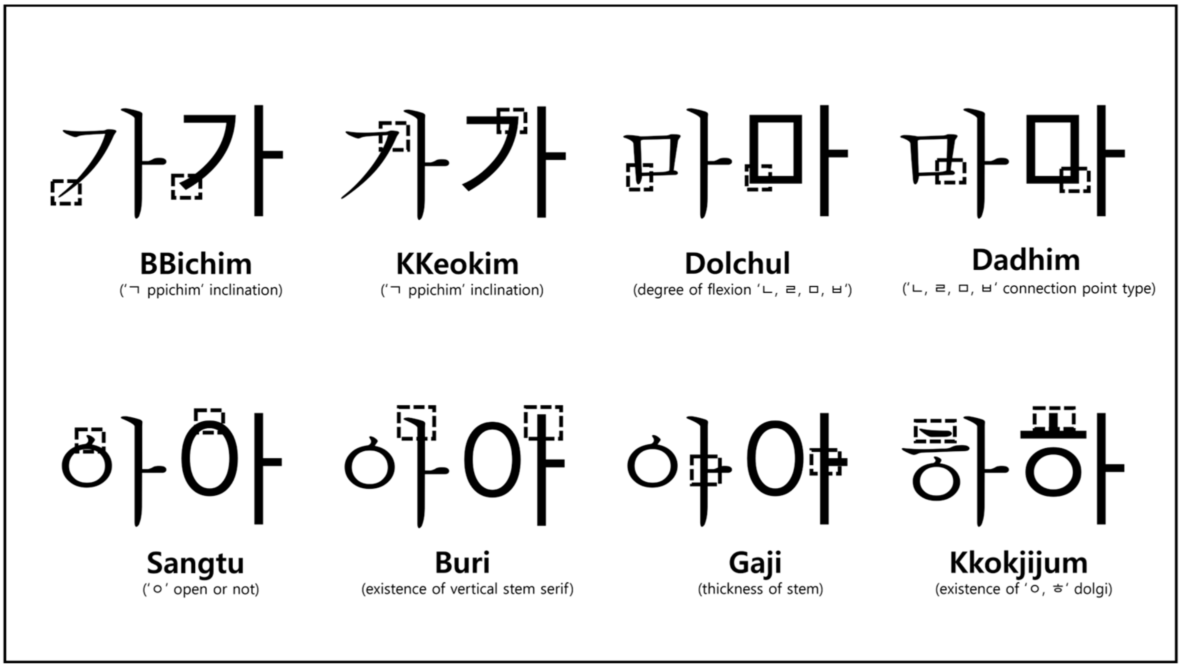


Figure 1. Eight representative stroke elements [7].

Eight stroke elements are automatically extracted through a deep learning-based Object Detection Model. After that, the extracted stroke element image is converted into a vector through the feature extraction step of the CNN model, and the structure thereof is shown in Figure 2. Through this process, images, which are high-dimensional data, are converted into vectors, which are low-dimensional data, and feature vectors of the stroke element are generated with image embedding that preserves the necessary information.

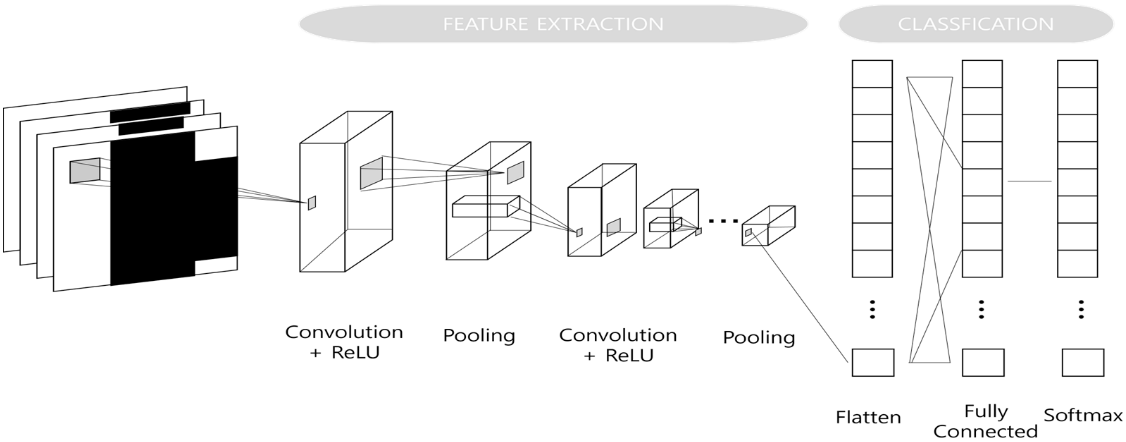


Figure 2. CNN Architecture including feature extraction and classification layer.

The stroke element feature vector is extracted for each font, and the most similar fonts are output in order through the cosine similarity calculation of each vector. Thereafter, a survey is conducted on the user to verify the similar font ranking derived from the deep learning model. Since determining the degree of similarity between fonts is a subjective problem, it should be compared with the rank of similar fonts that the user feels are similar. Therefore, in previous studies, the results of the survey were set to ground-truth to verify the results of similar font recommendations based on the stroke element.

2.1.2. Classification System According to the Characteristics of Hangul Shape

The Korean font shape classification system standardized by the Telecommunication Technology Association (TTA) [13] focusing on the glyph design of Korean fonts was classified as shown in Table 2. In previous studies [5–7,14], printing was selected from the following classifications to select representative fonts to be used for data collection and verification in the process of researching similar font recommendation techniques, and representative fonts were selected by thickness for more diverse shapes. Table 2 specifies the fonts used in the experiment in previous studies. This is a limited dataset in the development of a stroke element detection model using only one of the four major classification items. Therefore, in this paper, we evaluate the performance of a model that can detect stroke elements by addressing unstructured fonts that are not included in previous studies. This is essential currently at this time when there are more unstructured fonts than printed fonts due to the diversification of fonts.

Table 2. Whether to use font by classification based on Korean font shape classification standards (TTA) in previous study.

1st Tier Classification Criteria	Classification Name	Used/ Unused
InSwae Geulssi (Printing)	Structured Font	Used
Son Geulssi (Script)	Semi-structured Font	Unused
SaeGim Geulssi (Inscriptional)	Unstructured Font	Unused
Kita Geulssi (Miscellaneous)		Unused

2.2. Similar Hangul Recommendation System

2.2.1. Image-Based Font Recommendation System

WhatTheFont [15,16] is a system that recommends similar English fonts based on images using deep learning technology, and Sandollcloud [17] is a system that recommends Korean fonts in the same way. This system is currently the most popular similar font recommendation method using deep learning. However, in recommending fonts, fonts are recommended in units of images without considering the structural characteristics of Hangul. It is necessary to consider the fact that the design of the font is very characteristic, and existing research simply calculates the similarity of the entire letter and does not reflect the shape and structural properties of Hangul, so the satisfaction of accurate font recognition or similar font recommendation results is low. Hangul has various shapes and structural features, and fonts should be recommended based on the stroke element reflecting this characteristic.

Meanwhile, there are studies [18,19] that recommend similar fonts after automatic font identification. These provide similar font recommendation results as a web-based system, recognizing fonts from images and then extracting features based on CNN. Thereafter, similar fonts are recommended by calculating Euclidean distances between feature vectors. However, they do not reflect the characteristics of the font's shape, such as the stroke element. Simple experimental results [20] demonstrate that font recommendations using stroke elements perform better than the entire image. Thus, using our research, similar font recommendations that better reflect features are possible.

2.2.2. Limitations of Existing Services

Currently serviced font recommendation systems and related studies have several limitations. First, since deep learning is performed using the entire image, it is difficult to reflect detailed features. In this regard, there is also a related study that recommends similar fonts based on stroke elements to produce better performance. Second, some services are made only for English or Hangul fonts. So, it is difficult to find similar fonts for various languages. In this study, we intend to develop a similar font recommendation model based on stroke elements that can be applied to other language fonts by using stroke elements. Since stroke elements exist not only in Hangul but also in various languages, wide expansion is possible based on this study.

3. Expanded Experiments on Semi-Structured and Unstructured Font

This section describes the deep learning model used in the research. It is necessary to automatically extract eight stroke elements for the recognition of stroke elements in a letter image. In this case, Faster-RCNN was used among deep learning models.

Faster-RCNN

There are several ways to check the similarity of images. Generally, the full image of the text is used to judge similar fonts. However, after extracting the stroke element, it is possible to determine the similarity more efficiently by using the image of the stroke element [20]. Based on this, we tried to extract the stroke element and detect and extract the stroke element for fonts with irregular shapes other than the existing standard fonts.

Previously, we have attempted to study the detection of stroke elements using a support vector machine (SVM) [21] to detect the stroke element in the character image. However, the accuracy was significantly reduced during the detection process. Therefore, stroke element detection should use a deep learning-based object detection model, which is a deeper neural network. We detected stroke elements by fine-tuning the Faster R-CNN Inception-V2 model [22] among deep learning-based object detection models. Faster R-CNN has the highest accuracy compared to other models and is superior in detection accuracy for small-sized objects in particular. A faster R-CNN structure is shown in Figure 3.

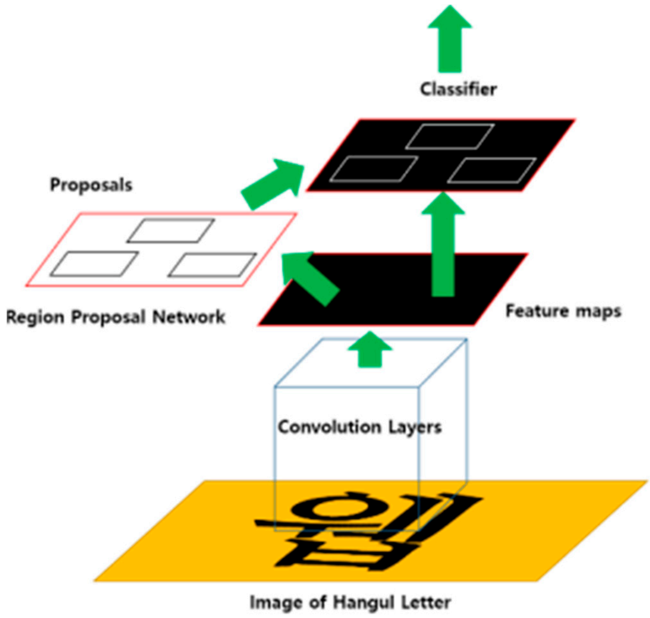


Figure 3. Faster R-CNN structure.

Faster R-CNN [23] is a CNN (Convolution Neural Network)-based deep learning object detection network, which has been improved since the R-CNN model [24] in 2014. The biggest feature of Faster R-CNN is that RPN (Region Proposal Networks) is added instead of the previous method of the Selective Search process, showing a great time-saving effect. Therefore, Faster R-CNN has significantly improved accuracy and speed compared to previous models by utilizing an RPN network that detects object candidate regions and a Fast R-CNN network that extracts and classifies features of object candidates. The RPN network operates in a sliding window method and calculates and detects candidate areas of objects around. Each sliding window has an anchor box with various sizes and aspect ratios, so it can flexibly detect candidate regions with high accuracy in terms of candidate region size and deformation.

4. Expanded Experiments on Semi-Structured and Unstructured Font

In this section, an extension experiment into semi-structured and unstructured fonts is described. Section 4.1 describes the model used, the extension of the letters data to be training, and unstructured fonts. Section 4.2 explains the stroke element extraction performance for unstructured fonts based on the font shape classification performed in Section 4.1.

4.1. Representative Character Extensions and Semi-Structured/Unstructured Fonts

In this experiment, a deep learning model was trained using the same eight stroke elements as in the previous study [5]. However, since there is a possibility of overfitting by performing the detection of stroke elements in simple letters, to compensate for this, the character images used in deep learning were expanded using various consonant-vowel combinations. The representative letters used in previous studies are four letters: ' 가', ' 마', ' 아', and ' 하', and the combination of letters is very monotonous. For example, ' 가' can be said to be a two-form letter consisting of the consonant ' ㄱ' and vowel ' ㅏ'.

On the other hand, there are six forms of letters that can be combined in Korean, including Chosung (initial consonant), Jungsung (middle vowel), and Jongsung (final consonant) letters. Figure 4 is a combination of six forms of letter combinations that can be combined into ' ㄱ' and ' ㅇ'.

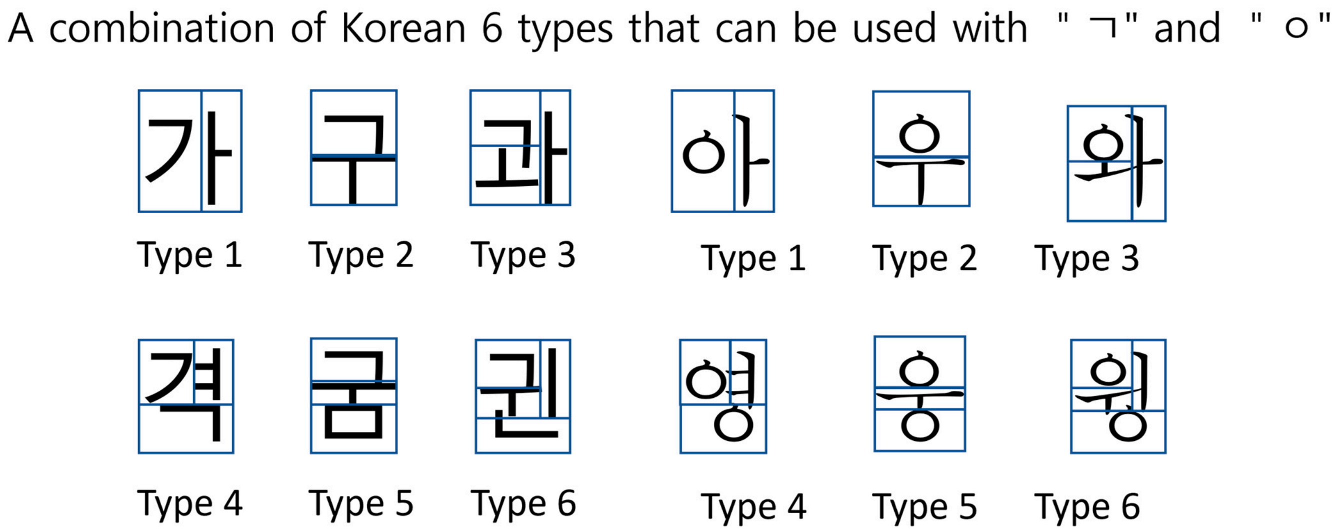


Figure 4. Six types of Korean alphabet combinations.

This combination of six forms can better reflect shape elements that may appear differently depending on the position between consonants and vowels than ' 가', ' 마', ' 아', and ' 하' made up of simple two-form combinations, and it is easier to detect stroke elements in complex characters. Accordingly, letters with uniform consonants and vowels with six forms and morphemes were expanded to representative letters, which are six letters as shown in Figure 5 below.

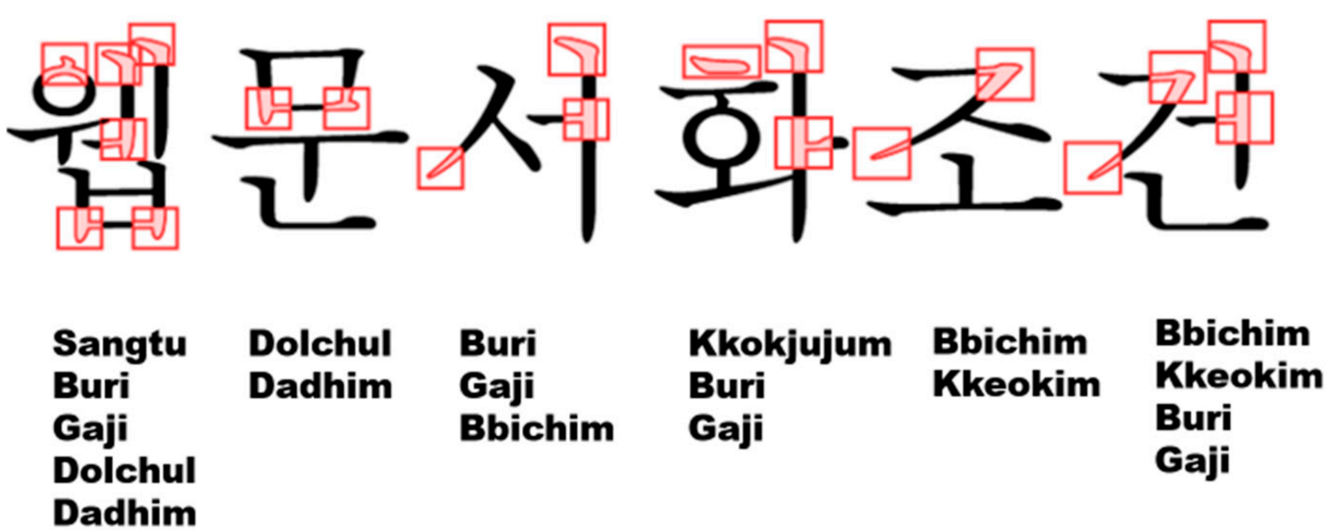


Figure 5. Representative Letter used in training object detection model. In the figure, the stroke elements of each Hangul letter are shown in red, and you can see the stroke elements that can be found in each letter. The above stroke elements are used for training the object detection model.

Each stroke element defined and training in Figure 5 confirms the extraction performance from different types of letters as shown in Figure 6 below. The letter below is the old name of Hangul, and it has evenly spaced stroke elements, so it was used in an experiment.

Figure 6. Target Letter for stroke element object detection. In the figure, the stroke elements of each Hangul letter are shown in red, and you can see the stroke elements that can be found in each letter. The above stroke elements are used to detect stroke elements.

Since the fonts used in previous studies are composed of only printing fonts, they are not suitable for detecting stroke elements and recommending fonts to be made later. Therefore, in addition to the existing printing fonts, the experiment was expanded by adding fonts such as amorphous fonts, design fonts, and handwriting fonts. This expanded font group was grouped into a total of three, A, B and C, and A group features a structured font including the existing printing font, and the B group is divided into a semi-structured font with similar characteristic elements to the regular font even if there is a change in font thickness or vowel deformation. Other handwritten fonts, dot fonts, and fonts with stroke elements deviating from the general stroke element type category were classified into Group C. The classified groups are shown in Table 3.

Table 3. Structured (A), semi-structured (B), and unstructured (C) font groups according to font classification criteria.

1st Tier Classification Criteria	Group	Example
InSwae Geulssi (Printing)	A (Structured)	윤탈명조 SM신명조 나눔고딕 타이포씨고딕 나눔명조
Son Geulssi (Script)	B (Semi-Structured)	HS새마을체 마포나루체 J해바라기 도스고딕
SaeGim Geulssi (Inscriptional)	C (Unstructured)	정선아리랑뿌리체 HS겨울눈꽃체 정규직으로 발걸음 1 정묵바위체
Kita Geulssi (Miscellaneous)		

The data generated through this process consisted of 1020 in group A, 486 in group B, and 672 in group C, 80% of each group was training data and 20% was test data. Using the deep learning framework TensorFlow Object Detection API [25], we went through the process of fine-tuning the already learned Faster-RCNN-Inception-V2-COCOO model [26]. All three groups conducted approximately 40,000 trains until the error was lowered.

4.2. Performance Evaluation of the Stroke Element Detection in Semi/Unstructured Fonts

Performance evaluation of the stroke element detection model was performed using a mean average precision (mAP) index using precision and recall for each stroke element, and the indicators and formulas are shown in Table 4 and Equations (1) and (2).

$$Recall = \frac{TP}{TP + FN} = \frac{TP}{All\ Ground\ Truth} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} = \frac{TP}{All\ Ground\ Truth} \tag{2}$$

Table 4. Precision Recall.

Ground Truth	Predict Result	
	Positive (+)	**Negative (−)**
Positive (+)	TP (True Positive)	FN (False Negative)
Negative (-)	FP (False Positive)	TN (True Negative)

Intersection over Union (IoU) evaluation indicators were used to verify that the results detected by the model were correctly detected. IoU is an evaluation method of measuring the area of the stroke element detected by the model in the image and the area of the ground-truth in which the location of the stroke element is directly indicated in the image and judging it as True Positive when the area of the overlapping intersection is 0.5 or more.

As a result of the experiment, the AP value of the model learned from group A, which is a regular font, is 60.54% and the AP value of the automatic extraction model of the stroke element learned from group B, is 62.36%. The AP value of the automatic extraction model of the stroke element learned with the amorphous font C group was 33.70%, which showed significantly lower results compared to the A and B groups. This result can be found in Table 5.

Table 5. Performance of the Stroke Element Detection Model for Structured, Semi-structured, and Unstructured Font.

	Group A	**Group B**	**Group C**
Gaji	92.31	100.00	52.08
Kkeokim	73.57	75.00	48.61
Kkokjijum	100	87.50	37.50
Sangtu	30.77	18.75	6.25
Dadhim	14.62	19.20	6.25
Dolchul	19.23	28.12	12.50
Bbichim	53.85	82.81	41.59
Buri	100	87.50	64.82
mAP	60.54%	62.36%	33.70%

5. An Experiment for Evaluating the Influence of Stroke Elements by Character Size

In this section, we will conduct an experiment to verify the reliability of the existing user evaluation ground-truth and to see if there is a difference in the influence of Hangul font's stroke elements depending on the size of the letter. First, Section 5.1 describes a new ground-truth built to validate the ground-truth used in existing studies. In Section 5.2, an experiment is conducted to verify the new ground-truth. Section 5.3 describes the need for a model of stroke elements weight by font size.

5.1. Ground-Truth through User-Based Font Similarity Evaluation

In the previous study [5–7,14], thirteen representative Hangul fonts were selected and font similarity was evaluated for users, and the evaluation results were detailed and used as ground-truth to verify the results of deep learning-based font recommendation. However, since the ground-truth data constructed in this way was used without sufficient validity verification, re-verification is necessary.

This study attempted to verify the validity of the ground-truth data used in the existing user evaluation by re-evaluating the user font similarity ranking for four of the thirteen fonts used in the previous study.

The four-font selections to verify similarity through the survey showed remarkable font characteristics in terms of the font's structure, such as stroke elements, skeleton, and thickness, and primarily selected fonts that can represent several Korean fonts. To this end, the standard for the Korean font shape classification system (TTAK.KO10.0906-Part2) [13] was referenced, and four fonts were finally selected: 'HY 견명조', ' 나눔스퀘어', ' 바탕체', and ' 산돌이야기M'. The above four fonts are used as reference fonts for the user font similarity ranking survey, and examples of the four fonts are fonts marked with red borders in Figure 7. The remaining comparative fonts can be checked together.

HY견명조 (HY GyeonMyungjo)	HY신명조 (HY ShinMyungjo)	굴림체 (GulimChe)	나눔명조 (NanumMyungjo)	나눔스퀘어 (NanumSquare)
돋움체 (DodumChe)	맑은고딕 (MalguenGothic)	바탕체 (BatangChe)	빙그레체 (BinguraeChe)	산돌독수리 Bold (SandolDosuri Bold)
산돌이야기 M (SandolIyagi M)	타이포도담체 (TypoDodamChe)	휴먼옛체 (HumanYetche)		

Figure 7. Types of representative fonts used in the experiment.

The composition of the Korean font similarity evaluation questionnaire is presented in three forms: paragraph, sentence, and word, and the representative fonts that are thought to be like the four reference fonts are written in order of similarity ranking. In this paper, the standard font was subdivided into small (paragraphs), medium (sentences), and large (words) based on the general usage of each letter size to determine the effect of letter size on similarity.

The example sentences used in the survey are shown in Table 6, and the example sentences are presented as images of the same size regardless of the size of the letters. A total of 41 survey respondents were asked to respond to information on experience, font interest, age, and gender in font-related fields with basic surveys. The survey results are converted into similarity scores out of five for use in individual stroke factor influence ranking calculation models and are used as Ground-Truth for font recommendation models. The Ground-Truth constructed because of evaluating user font similarity by character size is shown in Tables 7–9.

Table 6. User Survey Examples.

Character Size	Example Sentence
Small (7 pt)	나랏말씀이 중국과 달라 문자끼리 서로 맞지 아니하다. 이런 까닭으로 어리석은 백성이 이르고자 할 바가 있어도 마침내 자신의 뜻을 펴지 못하는 사람이 많으니라. 내 이를 위하여 가엾게 여겨새로 스믈 여덟자를 만드노니 모든 사람으로 하여금 쉽게 익혀 날마다 쓰기에 편안케 하고자 할 따름이니라.
Medium (10 pt)	한글은 세종대왕이 창제한 한국어의 문자이며, 자음 14자, 모음 10자이다.
Large (32 pt)	훈민정음

Table 7. Ground-truth as a result of a small font survey.

	HY 견명조	나눔스퀘어	바탕체	산돌이야기M
HY 견명조	5	2.3251	4.4013	3.5982
HY 신명조	4.5033	2.5478	4.7251	2.8462
굴림체	2.0818	4.0181	2.6459	2.9754
나눔명조	4.3846	2.6597	4.4026	3.2909
나눔스퀘어	2.3251	5	2.655	3.4245
돋움체	2.641	3.9727	3.0727	3.2105
맑은고딕	2.6317	4.1663	2.8636	3.4242
바탕체	4.4013	2.655	5	3.359
빙그레체	1.7181	3.4895	2.0163	3.0233
산돌독수리Bold	2.7156	2.1636	2.1196	2.6037
산돌이야기M	3.5982	3.4245	3.3590	5
타이포도담체	2.2587	3.3636	1.9138	3.1909
휴먼옛체	1.796	2.0536	1.5315	2.0239

Table 8. Ground-truth as a result of a medium font survey.

	HY 견명조	나눔스퀘어	바탕체	산돌이야기M
HY 견명조	5	2.3283	4.4013	3.3683
HY 신명조	4.7161	2.7364	4.8404	2.6651
굴림체	2.0536	3.6234	2.7529	2.8836
나눔명조	4.459	2.9	4.4856	2.9
나눔스퀘어	2.3283	5	2.6597	3.67406
돋움체	2.9636	4	3.1727	3.1364
맑은고딕	2.8555	4.2284	3.0699	3.5455
바탕체	4.40133	2.6597	5	2.9455
빙그레체	1.8205	3.3732	1.8299	3.387
산돌독수리Bold	2.6037	2.0163	2.2	2.3876
산돌이야기M	3.3683	3.6741	2.9455	5
타이포도담체	1.8951	2.9674	1.7371	3.5272
휴먼옛체	1.7925	2.2703	1.6527	2.2028

Table 9. Ground-truth as a result of a large font survey.

	HY 견명조	나눔스퀘어	바탕체	산돌이야기M
HY 견명조	5	2.4422	4.5765	2.8583
HY 신명조	4.796	2.4273	4.8403	2.4825
굴림체	2.0536	3.6253	2.4354	2.8091
나눔명조	4.2018	2.8727	4.4058	2.7622
나눔스퀘어	2.4422	5	2.7646	3.6741
돋움체	2.8091	4.2018	3.1636	3.2545
맑은고딕	2.8741	4.3	2.8091	3.5636
바탕체	4.5765	2.7646	5	2.9455
빙그레체	1.9604	3.1545	2.1002	3.5818
산돌독수리Bold	2.6224	2.1530	2.244	2.3727
산돌이야기M	2.8583	3.6741	2.7762	5
타이포도담체	2.0443	3.0792	1.9044	4.1483
휴먼옛체	1.9231	2.3589	1.4818	2.7129

5.2. Experiments for the New Ground-Truth Validation

5.2.1. Analysis of the Influence of Stroke Elements by Combination

In this section, the influence of each stroke element is evaluated based on the ground-truth newly derived in Section 4.1. The influence of the stroke element by combination means the influence of the individual stroke element on the combinations generated when the combination operation is performed from one stroke element to eight stroke elements. The accuracy value of the recommended result of each combination is used for the stroke factor influence calculation. For example, if the font recommendation accuracy is 0.88 when recommending a font with a combination of four stroke elements: "bbichim", "kkokjijum", "dadhim", "sangtu", "dadhim" are calculated by adding 0.88 to the influence values of "buri", "bbichim", "kkokjijum", "dadhim", "sangtu", "dadhim". The formula for calculating the influence of the stroke element is shown in Equation (3).

s refers to a combination of one or more stroke elements, and si $(1 \leq i \leq n)$ refers to each stroke element included in the combination. It is assumed that Sx is an arbitrary target stroke element to calculate the influence, C is the accuracy value of the combination, and sx is a variable to store the cumulative sum of the accuracy values C of the combination to which the target stroke element Sx belongs. The cumulative sum of the stroke factor influence values was ranked from first to eighth according to size.

$$s = \{s_1, s_2\, s_3\, s_4 \ldots s_n\} = \mathrm{C} \tag{3}$$

$$\begin{cases} \mathrm{s_x} \in \mathrm{s},\ \mathrm{S_x} = \mathrm{S_x} + \mathrm{C} \\ \quad \mathrm{s_x} \notin \mathrm{s},\ \mathrm{S_x} \end{cases}$$

The experimental results can be confirmed in Table 10, and the following results were derived. For example, in the combination of the two stroke elements, the highest similarity was derived with the combination of "bbichim" and "sangtu", and in the three combinations, the "bbichim", "sangtu" and "kkokjijum" was the highest. In the four combinations, the two or three combinations excluded the "bbichim" and added "gaji" and "kkeokim", indicating a combination of "kkeokim", "gaji", "kkokjijum", and "sangtu". Based on the experimental results above, it can be concluded that in order to build a similar font recommendation system based on the stroke element, changes in the stroke element that vary by combination should be considered.

Table 10. Ranking of influence of stroke elements by combination.

	Number of Stroke Elements			
Rank	**1**	**2**	**3**	**4**
1	bbichim	bbichim sangtu	bbichim kkokjijum sangtu	gaji kkeokim kkokjijum sangtu
2	kkokjijum	kkokjijum sangtu	kkeokim kkokjijum sangtu	bbichim kkeokim kkokjijum sangtu
3	dolchul	bbichim kkokjijum	bbichim kkeokim sangtu	bbichim dolchul kkokjijum sangtu
4	kkeokim	kkeokim kkokjijum	gaji kkeokim sangtu	gaji dolchul kkokjijum sangtu
5	dadhim	dolchul kkokjijum	gaji kkokjijum sangtu	dolchul kkeokim kkokjijum sangtu
6	sangtu	kkeokim sangtu	bbichim dolchul kkokjijum	bbichim gaji kkokjijum sangtu
7	buri	bbichim dolchul	dolchul kkeokim kkokjijum	dolchul gaji kkeokim sangtu
8	gaji	dolchul kkeokim	bbichim kkeokim kkokjijum	bbichim dadhim kkokjijum sangtu
9		bbichim kkeokim	dolchul kkokjijum sangtu	dolchul gaji kkeokim kkokjijum
10		dolchul gaji	dadhim kkokjijum sangtu	dadhim gaji kkokjijum sangtu

Table 11 shows the existing ground-truth-based stroke element influence ranking. Looking at the results of the calculated stroke element influence, “sangtu” showed the greatest influence, and “dadhim” had the smallest influence.

Table 11. Existing stroke element influence values and rankings in previous study.

Rank	Stroke Element	Value of Stroke Element Influence
1	sangtu	26.4869
2	buri	15.0577
3	gaji	14.0024
4	dolchul	12.2639
5	kkeokim	7.4040
6	kkokjijum	7.3690
7	bbichim	5.5579
8	dadhim	3.6782

Existing ground-truth is built on fonts of the same size, all without considering the font size. However, the reliability of the data used as ground-truth is poor because the character size is not considered. Therefore, the influence of the stroke element by character size was calculated using the new ground-truth. At this time, the average value of the influence of small letters, middle letters, and large letters was used as the influence of the integrated stroke element. Table 12 shows the influence of the integrated size and the influence of the size-specific size.

Table 12. Results of converting stroke element influence values into weights.

	Integrated Size		Small Size		Medium Size		Large Size	
Rank	**Stroke Element**	**Value of Stroke Element Influence**	**Stroke Element**	**Value of Stroke Element Influence**	**Stroke Element**	**Value of Stroke Element Influence**	**Stroke Element**	**Value of Stroke Element Influence**
1	sangtu	18.7006	sangtu	24.4613	sangtu	18.5556	kkokjij um	18.6540
2	buri	16.8693	buri	15.9557	kkokjiju m	17.6909	buri	15.8788
3	dolchul	14.9908	kkeokim	14.0555	buri	15.8192	dolchul	15.8570
4	kkokjij um	14.9595	dolchul	10.3460	dolchul	11.1747	gaji	13.8310
5	gaji	8.2870	kkokjiju m	10.2571	gaji	10.0992	bbichi m	9.2744
6	bbichim	6.4716	gaji	10.2183	bbichim	8.3007	sangtu	9.2268
7	kkeoki m	5.5395	bbichim	3.6657	dadhim	5.4818	dadhim	5.5270
8	dadhim	5.5250	dadhim	2.7196	kkeokim	3.6213	kkeoki m	2.6981

The most influential stroke elements in small and medium letter sizes were "sangtu" and "kkokjijum" in large letter sizes. On the other hand, the least influential elements were "dadhim" in small letter size and "kkeokim" in medium letter size and large letter size. The change in the influence ranking of the stroke element according to size is shown in Figure 8.

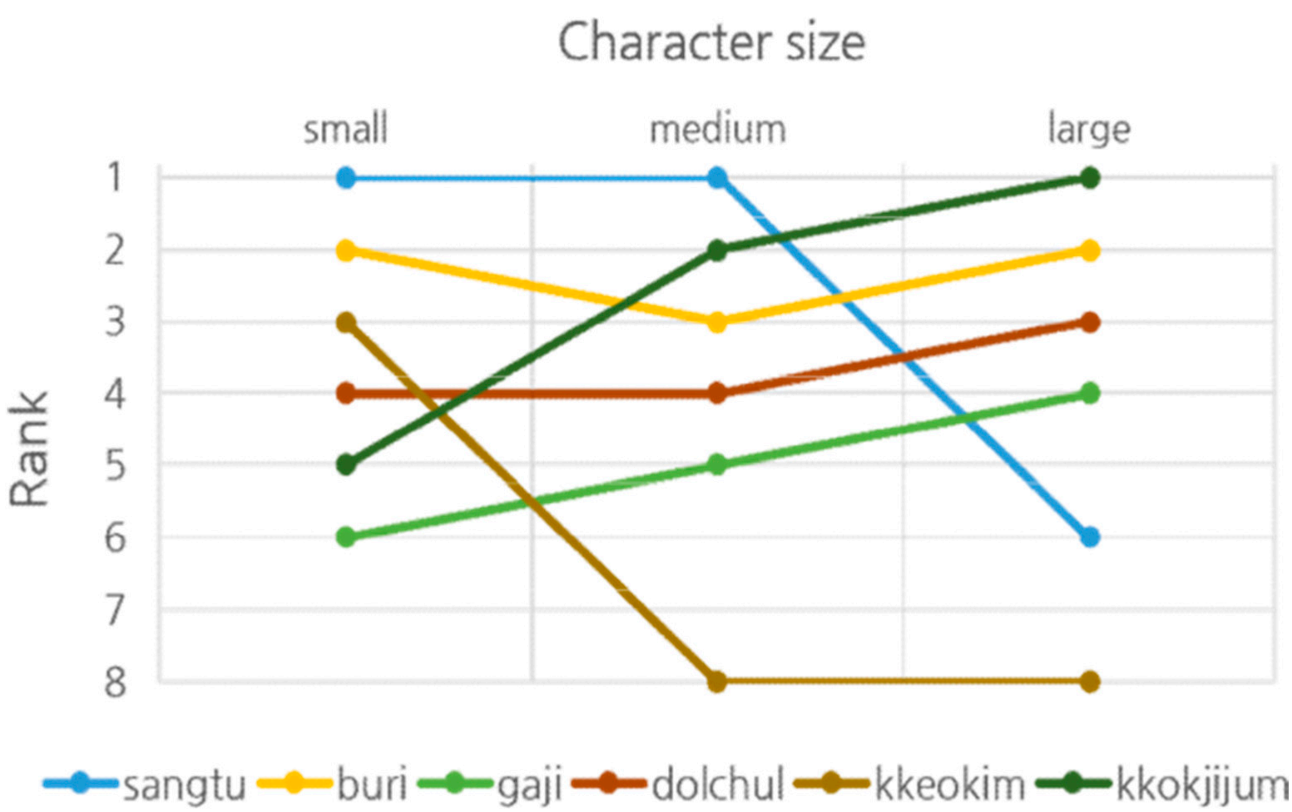


Figure 8. Change the rank of stroke element influences by character size.

5.2.2. Reliability Verification Results of Existing Ground-Truth

In order to verify the reliability of the existing user evaluation ground-truth, we conducted a re-validation by comparing the existing stroke element influence ranking with the size-integrated stroke element influence ranking. To check the correlation between the two rankings, the ranking was compared with the Kendall Tau correlation coefficient [27] to calculate the correlation, and if there is a correlation, the existing user evaluation ground-truth was considered reliable. τ is the Kendall Tau coefficient and p is the significant probability. By comparing the two rankings, we assume that the concordant pair is C and the discordant pair is D. The hypothesis for the test of the influence ranking of the stroke element is as follows and is Equation (4).

$$\tau = \frac{\mathbf{C}-\mathbf{D}}{C+D} \tag{4}$$

As a result of calculating the existing stroke element influence ranking and the size integrated stroke element influence ranking with the Kendall Tau correlation coefficient, the significance probability p-value is 0.014, which is less than the significance level of 0.05, so the null hypothesis is rejected. Therefore, the existing stroke element influence ranking is associated with the integrated stroke element influence ranking, and the Kendall Tau coefficient value (τ)) is 0.714, confirming that the existing stroke element influence ranking and the size integrated stroke element influence ranking are strongly correlated.

5.3. Necessity of a Stroke Element Weight Model by Character Size

Looking at the experimental results of Section 5.2, the ranking of the influence of the stroke element differs for each character size. Therefore, in font recommendation, it is possible to recommend a similar font more efficiently when used for recommendation by weighting the stroke element for each character size.

To support the above argument, the correlation between the existing ground-truth-derived stroke element influence ranking (Table 11) and the stroke element influence ranking (Table 12) derived by letter size was analyzed, and the results are shown in Table 13. Verification of the influence ranking of character elements by character size uses Kendall's Tau correlation coefficient.

Table 13. Verification of the influence ranking of character elements by character size using Kendall's Tau correlation coefficient.

Character Size	Small Size	Medium Size	Large Size
p-value	0.01	0.10	0.54
Similarity between ranks	similar	different	different
Kendall's τ coefficient	0.714	0.499	0.214
Correlation	strong	-	-

As a result of comparing the influence ranking of the small character size stroke element, the significance probability, p-value, was less than the significance level of 0.05, showing a correlation and a strong correlation.

Compared with the existing stroke factor influence ranking, the significance probability p-value was higher than the significance level of 0.05, confirming that the stroke factor influence ranking between the middle and large letters was not correlated with the existing stroke factor influence ranking.

Therefore, in the case of the small character size, there is no difference from the existing stroke element influence ranking, so it is okay to recommend it according to the existing stroke element influence ranking when recommending fonts. However, since there is a difference between the middle letter and the large letter size from the existing stroke element influence ranking, the accuracy of the font recommendation will be reduced if the existing stroke element influence ranking is applied.

Therefore, when recommending fonts for medium and large letter sizes, it can be concluded that it is efficient to apply a weight model that reflects high weights to high-impact stroke elements, rather than to recommend all stroke elements equally.

6. Results and Discussion

6.1. Results and Analysis of Semi-Structured/Unstructured Font Extension Experiments

When checking the mAP indicators for each stroke element in Group A, the mAP values were high with 'Kkokjijum' (100%), 'Buri' (100%), and 'Gaji' (92.31%), 'Kkeokim' (73.57%) and 'Bbichim' (53.85%) showed relatively low performance. Lastly, 'Sangtu' (30.77%), 'Dadhim' (14.62%), and 'Dolchul' (19.23%) had the lowest detection performance. Group A, 'Kkeokim' was detected in all the ground-truth target letters, but False Positive existed, and it seems to have been detected by recognizing the upper part of 'ㅁ' as 'Kkeokim'. In the case of a 'Sangtu', detection failed in a relatively long 'ㅇ' shape.

Overall, except for 'Kkeokim', the performance seems to have been lowered due to the failure of the detection. This shows a lower detection accuracy than previous studies, but it is due to the expansion of representative letters from simple letters to six form combinations. Considering this, the shape of consonants and vowels in Hangul changes slightly depending on the combined form, it is not judged to be a significant performance degradation. When the mAP indicators for each stroke element of Group B were checked, 'Gaji' (100%), 'Buri' (87.5%), 'Kkokjijum' (87.5%), 'Bbichim' (82.81%), and 'Kkeokim' (75%) showed relatively high mAP values, and on the contrary, 'Dolchul' (28.12%), 'Dadhim' (19.20%), and 'Sangtu' (18.75%) showed low mAP values. In addition, in the case of Group B, the 'Bbichim' was about 83%, and there were many cases of incorrect detection (FP) by confusing the vowel conclusion and pouting. An example image of the detection result of Group B is shown in Figure 9 below.

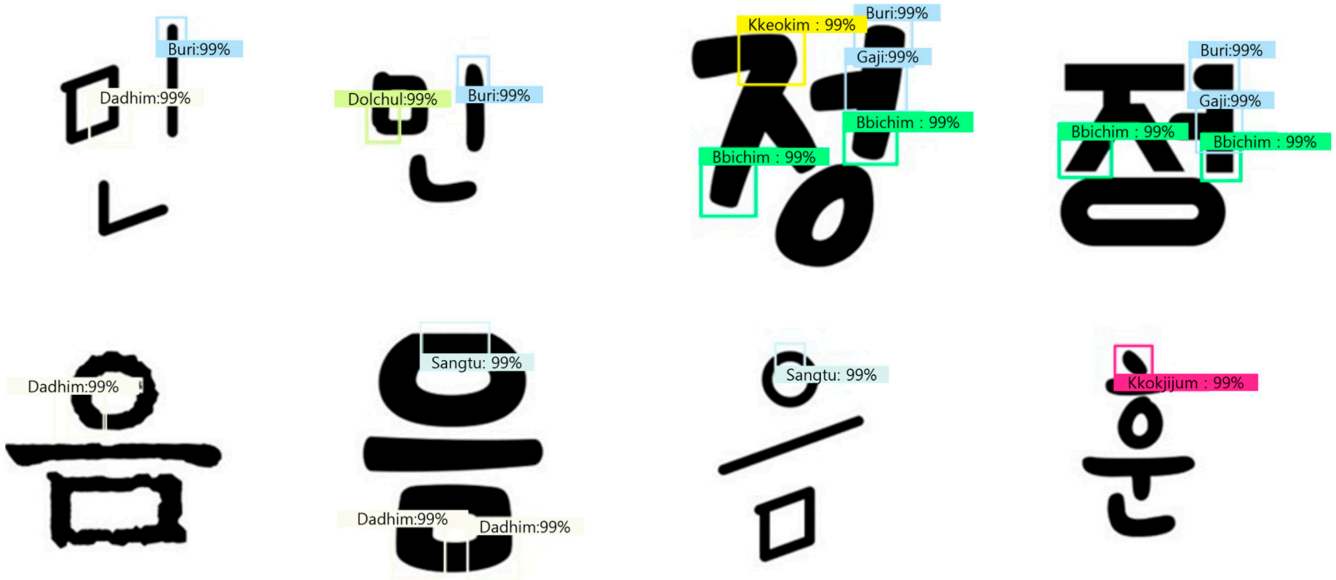


Figure 9. An example image of the detection result of Group B.

The stroke elements with relatively high mAP values in Group C were 'Gaji' (52.08%), 'Buri'(64.82%), 'Kkokjijum' (37.50%), 'Bbichim' (41.59%), and 'Kkeokim' (48.61%), but showed lower mAP values on average than in Group B. In addition, 'Dolchul' (12.50%), 'Dadhim' (6.25%), and 'Sangtu' (6.25%) showed very low detection accuracy compared to other stroke elements.

Group C had the lowest detection performance of overall stroke elements, especially compared to other groups. In the case of the best-detected 'Buri', it was confirmed that the part other than the beak was recognized as the beak. In the case of 'Bbichim', like group B, there were cases of incorrect detection by confusing it with the vowel enclosure, and

'Sangtu' showed a failure to detect 'ㅇ' in a consonant like group B. In the case of 'Dadhim' it showed results that could not be detected in unstructured font files with a large difference from the general form of 'Dadhim'. In addition, when the consonant 'ㅇ' is in the located in the Jongsung (final consonant), as in the letter '정', there was a tendency to not detect the 'Sangtu'. An example image of the detection result of Group C is shown in Figure 10 below.

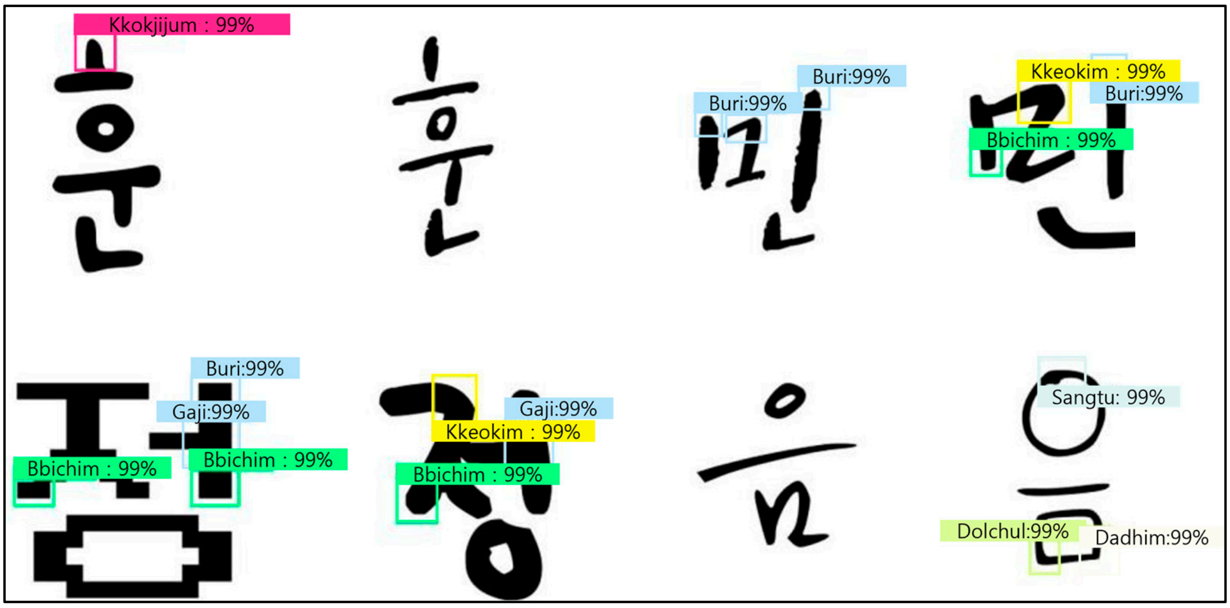


Figure 10. An example image of the detection result of Group C.

The experimental results above imply that Group C's performance needs to be improved. In response, this study discussed ways to improve performance for each group to be conducted in future studies. Group A and B showed relatively good and similar performances. As can be seen in Figure 11, there is a slight difference and it can be seen that the shape is similar. However, in the case of Group C, the shape is quite different and it is composed of freewheeling typefaces, so additional measures are needed. The proposed performance improvement plan is as follows, and we intend to observe performance changes by conducting this in future research. The proposed method is to define additional shape elements considering that the detection of the stroke element in unstructured fonts will fail. For example, characteristic elements other than stroke elements in letters include thickness, space, and skeleton. Using these additional elements, similarities can be calculated for characters that fail to detect the stroke element. For example, skeleton elements can be extracted from letters that fail to detect stroke elements, allowing inter-bone comparisons to be performed. In addition, it will be possible to perform tasks such as newly defining stroke elements specialized in unstructured fonts.

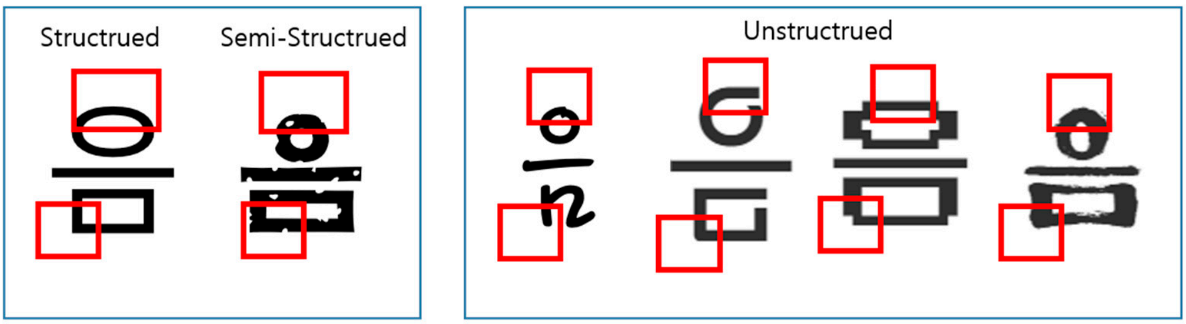


Figure 11. Difference in appearance of the same stroke element by group.

6.2. Weight Calculation Model for Each Stroke Element

Based on the experiments and verifications performed in Section 4, this section describes a planning element-specific weight calculation model that will finally be used in the Hangul font recommendation system. When recommending fonts, more accurate font recommendations are possible if the proportion of stroke elements that greatly affect similarity judgment increases and the proportion of stroke elements with small impact decreases. Therefore, we propose a stroke element weight calculation model that applies the stroke element reflection ratio differently according to the influence value of each stroke element.

To this end, the individual stroke factor influence values calculated in Section 3 are normalized to a value between 1 and 10 and used as a weight for font recommendation [28]. Assuming that a, b is a normalization range (a < b) and x is a normalization target variable *xnormalized* is a normalized variable, the equation is the same as (5).

$$xnormalized = (b - a) * (x - \min(x)/(\max(x) - \min(x)) + a \tag{5}$$

Table 14 shows the results of adjusting the range of influence values of existing stroke elements, size integrated stroke elements and font size with the above weight calculation model, and you can adjust the reflection ratio of stroke elements by font size in the font recommendation system later.

Table 14. Results of converting stroke element influence values into weights.

	Existing		Integrated Size		Small Size		Medium Size		Large Size	
Rank	**Stroke Element**	**Weight**	**Stroke Element**	**Weight**	**Stroke Element**	**Weight**	**Stroke Element**	**Weight**	**Stroke Element**	**Weight**
1	sangtu	10.0	sangtu	10.0	sangtu	10.0	sangtu	10.0	kkokjiju m	10.0
2	buri	5.49	buri	8.95	buri	6.48	kkokjiju m	9.48	buri	8.44
3	gaji	5.07	dolchul	7.47	kkeoki m	5.69	buri	8.35	dolchul	8.42
4	dolchul	4.39	kkokjiju m	7.45	dolchul	4.16	dolchul	5.55	gaji	7.28
5	kkeoki m	2.47	gaji	2.89	kkokjiju m	4.12	gaji	4.9	bbichim	4.71
6	kkokjiju m	2.46	bbichim	1.65	gaji	4.10	bbichim	3.82	sangtu	4.68
7	bbichim	1.74	kkeoki m	1.01	bbichim	1.39	dadhim	2.12	dadhim	2.6
8	dadhim	1.0	dadhim	1.0	dadhim	1.0	kkeoki m	1.0	kkeokim	1.0

7. Conclusions

Since the stroke element is an important discriminating element in the font, when calculating the similarity between fonts based on this, it is possible to calculate the similarity more accurately than calculating the similarity with the entire letter.

However, in previous studies, a deep learning model was developed for only structured fonts when extracting stroke elements. This causes limitations in the development of a font recommendation system with completeness in the future. Therefore, there was a need for supplementation and verification of existing studies and expansion for versatile font recommendations.

This paper conducted an experiment with two main themes, and the results are as follows. First, as a result of confirming the performance of the stroke element detection model through font classification and target character expansion, the performance of the font recommendation system was derived in Group A, a structured font, and Group B, a semi-structured font. On the other hand, Group C, an unstructured font with too much diversity, had the lowest performance in extracting stroke elements.

In conclusion, it was concluded that if the results of this experiment are reflected in the font recommendation system in which only Group A was used, it is possible to add the font of Group B, which does not have a significant performance difference from Group A. On the other hand, Group C showed low detection performance because of the severe change in fonts and the model has a difficult form of stroke elements to train.

Group B is also a semi-structured font with consonant and vowel variations, but it can be included within the scope of use of the recommendation system due to similar characteristic elements to the formal font, so it will be applied to the font recommendation system later. In the case of Group C, the performance of detecting stroke elements is very low, so research to increase performance is needed.

Second, the reliability of the existing user evaluation ground-truth used in previous studies [5–7] was verified in the process of determining the similarity between fonts when recommending Korean fonts, and whether there is a difference in the influence of Korean stroke elements according to the font size. As a result of verification, the existing ground-truth was validated using Kendall's Tau correlation coefficient using the existing stroke factor influence ranking, and the existing user evaluation was found to be reliable. In addition, it was concluded that the influence of individual planning elements was different, and the difference in influence was large when using the combination of planning elements, so it was necessary to reflect it when applied to the recommendation system. Accordingly, when using the stroke element influence value in the font recommendation system, a stroke element weight setting model was developed to reflect the weight according to the influence difference.

Lastly, in the case of small character size, there is no significant difference from the existing stroke element influence, so it is not necessary to calculate the stroke element influence value separately since the middle character size and the large character size were different from the existing stroke element influence. In general, small letter sizes are used for the body, medium letter sizes are used for the title of the presentation material or for the body, and large letter sizes are used for poster design or the title of the presentation material, so letter sizes can be associated with the purpose. Through this result, font recommendations for future font recommendations can also be performed, and it can be expected to be expanded to a sophisticated font recommendation system. As a future study, we intend to develop a multi-lingual font recommendation model after developing the definition of English and Chinese character design elements and a stroke element extraction model.

Author Contributions: Conceptualization, S.-B.L. and Y.S.; methodology, S.-B.L.; software, S.-B.L.; validation, S.-B.L., Y.S. and J.L.; formal analysis, Y.S.; writing—original draft preparation, X.Z.; writing—review and editing, Y.S. and X.Z.; project administration, J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2021R1I1A4A01059550). This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (grant number: RS-2022-00165818).

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ju, I.; Lee, H.; Sherrick, B. Consumer Responses to Covert Marketing Communications: A Case of Native Advertising Disclosure in News Contexts. *J. Promot. Manag.* **2022**, *28*, 1107–1128. [CrossRef]
2. Ilonga, E.; Mapunda, G. Complementarity of communicative modes on meaning making in Tanzania's digital telecom marketing: A social semiotic multimodal perspective. *South. Afr. Linguist. Appl. Lang. Stud.* **2022**, *40*, 87–99. [CrossRef]

3. Korea National Museum. Survey and Prospect of Font Industry. *Korea Natl. Mus. Res. Rep.* 2015. Available online: https://www.hangeul.go.kr/bbs/publicBbsView.do?curr_menu_cd=0105050000&pageIndex=1&search_type=title&search_text=&bbs_id=4&bbs_no=31&mode=/ (accessed on 10 November 2022).
4. Kim, N. A Study on The Typology for Fonts Identification—Focusing on the difference element in shape of font. *Korea Des. Forum* **2018**, *23*, 89–98.
5. Jeon, J.Y.; Lim, S.B. Analysis of Extraction Performance according to the Expanding of Applied Character in Stroke Element Extraction. *J. Korea Multimed. Soc.* **2020**, *23*, 1361–1371.
6. Jeon, J.Y.; Park, D.Y.; Lim, S.Y.; Ji, Y.S.; Lim, S.B. Automatic Extraction of Hangul Stroke Element Using Faster R-CNN for Font Similarity. *J. Korea Multimed. Soc.* **2020**, *23*, 953–964.
7. Park, D.Y.; Jeon, J.Y.; Lim, S.B. A Study on Influence of Stroke Element Properties to find Typeface Similarity. *J. Korea Multimed. Soc.* **2020**, *23*, 1552–1564.
8. Korean Intellectual Property Office. A Study on the Protection Scope and Similarity Judgment Criteria of Korean Typeface. 2003. Available online: http://dl.nanet.go.kr/law/SearchDetailView.do?cn=MONO1200720746#none (accessed on 10 January 2023).
9. Kim, J.P. *Formation Study*; Korea Institute of Publishing Culture: Seoul, Republic of Korea, 1990.
10. Lee, Y.J. Research on Representative Letterforms for Developing Effective Fonts. *J. Korean Soc. Typogr.* **2012**, *4*, 1110–1260.
11. Kim, H.Y.; Lim, S.B. Shape Property Study of Hangul Font for Font Classification. J. *Korea Multimed. Soc.* **2017**, *20*, 1584–1595.
12. Kim, H.Y.; Lim, S.B. Standardization Study of Font Shape Classification for Font Registration System. *J. Korea Multimed. Soc.* **2017**, *20*, 571–580. [CrossRef]
13. TTA. Font Production Guides for Font Registration System Part 2. Design Based Classification, e-Publishing Project Group 608, TTAK.KO-10.0906-part2, Telecommunication Technology Association. 2016. Available online: https://www.tta.or.kr/tta/ttaSearchView.do?key=77&rep=1&searchStandardNo=TTAK.KO-10.0906-Part2&searchCate=TTAS (accessed on 10 November 2022).
14. Jeon, J.; Lim, S. Machine Learning Based Automatic Extraction of Stroke Elements in Font. In Proceedings of the 8th Korea-Japan Joint Workshop on Complex Communication Sciences, Gyeong Ju, Republic of Korea, 4–6 January 2020; pp. 53–55.
15. WhatTheFont. Available online: https://www.myfonts.com/ (accessed on 13 December 2022).
16. Solli, M.; Lenz, R. A font search engine for large font databases. *ELCVIA Electron. Lett. Comput. Vis. Image Anal.* **2011**, *10*, 24–41. [CrossRef]
17. Sandollcloud. Available online: https://www.sandollcloud.com/ (accessed on 13 December 2022).
18. Wang, Z.; Yang, J.; Jin, H.; Brandt, J.; Shechtman, E.; Agarwala, A.; Wang, Z.; Song, Y.; Hsieh, J.; Kong, S.; et al. A system for font recognition and similarity. In Proceedings of the 23rd ACM International Conference on Multimedia, Brisbane, Australia, 26–30 October 2015; pp. 813–814.
19. Jiang, S.; Wang, Z.; Hertzmann, A.; Jin, H.; Fu, Y. Visual font pairing. *IEEE Trans. Multimed.* **2019**, *22*, 2086–2097. [CrossRef]
20. Jeon, J. CNN-Based Similar Font Recommendation Techniques Using Hangul Stroke Elements Information. Master's Thesis, Sookmyung Women's University, Seoul, Republic of Korea, 2021.
21. Hearst, M.A.; Dumais, S.T.; Osuna, E.; Platt, J.; Scholkopf, B. Support vector machines. *IEEE Intell. Syst. Appl.* **1998**, *13*, 18–28. [CrossRef]
22. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Tran. Patt. Anal. Mach. Intell.* **2015**, *39*, 1137–1149. [CrossRef] [PubMed]
23. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
24. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
25. Mustamo, P. Object Detection in Sports: TensorFlow Object Detection API Case Study. Bachelor's Thesis, University of Oulu, Oulu, Finland, 2018.
26. Janahiraman, T.V.; Mohamed, S.M.S. Traffic light detection using tensorflow object detection framework. In Proceedings of the 2019 IEEE 9th International Conference on System Engineering and Technology (ICSET), Shah Alam, Malaysia, 7 October 2019; pp. 108–113.
27. Schaeffer, M.S.; Eugene, E.L. Concerning Kendall's tau, a nonparametric correlation coefficient. *Psychol. Bull.* **1956**, *53*, 338. [CrossRef] [PubMed]
28. Patro, S.; Kishore, K.S. Normalization: A preprocessing stage. *arXiv* **2015**, arXiv:1503.06462. [CrossRef]