Spring 5-2023

Handwritten Hangul Recognition

Joshua Cho
University at Albany, State University of New York, jcho4@albany.edu

Handwritten Hangul Recognition

Thesis

presented to the Department of Computer Science
University at Albany, State University of New York
in partial fulfillment of the requirements
for graduation with Honors in Computer Science & Applied Mathematics
and graduation from the Honors College

Bachelor of Science

by

Joshua Cho

Reviewed and approved by the following:

Ming-Ching Chang, PhD
Associate Professor, Department of Computer Science

March 27, 2023

Abstract

Despite significant advancement in Optical Character Recognition (OCR), Handwritten Hangul (Korean) Recognition, abbreviated as HHR, remains largely unsolved due to the similarity found in Hangul handwriting and a much larger number of syllables ($\sim 11,000$) to classify compared to English or Latin script languages. The state-of-the-art approach toward HHR on the SERI95a handwritten Korean dataset was based on the AlexNet trained with data augmentation and hybrid learning. While this method has achieved a higher classification performance than the traditional approaches, it is limited in learning feature representations at various levels. In this study, we adopt a variation of Xception which exploits depthwise separable convolutions to enhance the recognition rate of individual syllables. In addition, we propose a progressive training method from Progressive Growing GAN to provide a stabilized training process with reduced training time. Through this approach, the empirical results demonstrate that the proposed approach outperforms the state-of-the-art method and demonstrates its effectiveness on HHR.

Keywords: OCR, Handwritten Hangul Recognition, HHR, Korean, Xception, progressive training, depthwise separable convolution, SERI95a.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my God for His unconditional love and mercy. He is the source of my happiness, comfort, peace, strength, and fear (reverence). Thank You for allowing me to grow both spiritually and academically and for always being with me during the path of my life journey.

I would like to render my warmest thanks to my advisor, Professor Ming-Ching Chang, whose kind understanding, guidance, and teaching have made this work possible. I am deeply indebted to him for his continuous support in every step of my academic journey and I attribute the wonderful four years of my undergraduate experience to Professor Chang.

I would also like to express my earnest gratitude to Yuwei Chen, Abhineet Kumar Pandey, Zhenfei Zhang, Rui Wang, An Yu, Nidhi Vadnere, Srujan Vepuri, and Ragini Rani, my colleagues in the Computer Vision and Machine Learning (CVML) lab.

The greatest indirect contribution to this work is my family. I would like to express a sincere thankfulness to my family, whose importance is indescribable, for their love and wisdom and for always being by my side. I cannot express how much you mean to me and thank you for being my father, my mother, and my sister.

I would like to extend my deepest gratitude to my American grandparents and their family who have shaped into the person I am today. I can confidently state that I have the best childhood memory because of my American grandparents, and I cherish the memories of the time I spent with them. Although it will never be enough, thank you for your guidance, wisdom, prayer, and love.

I am also greatly indebted to my Korean grandparents. Thank you for your prayer, love, unwavering support, warm hugs, and for always bringing a smile to my face. Your influence during my childhood and adulthood is truly beyond measure, and I treasure the happy time that I have spent with you.

I would like to extend my sincere thanks to my uncle Minsoo Koh, my aunt Jenny Noh, and their children −Yuna and Yuchan. They have celebrated by my side in the best of times and supported me during the times of difficulty. I am deeply indebted to them for their sacrifice, prayer, guidance, warm hospitality, and for the enjoyable times that we have spent together.

Lastly, I would like to extend my sincere thanks to my Senior Academic Advisor Todd

J. Schnitzer, Lecturer Michael Phipps, Professor Penghang Yin, and Professor Vladimir Kuperman for the friendship and their kind support in my graduate school preparation, Administrative Manager Mark E. Seabury II for his warm kindness throughout my undergraduate years, Professor Dola Saha and her graduate students at MESA lab for their teaching and guidance during my freshman year, Professor Hui-Ching Chang and Dr. Kim Boyer for their teaching, friendly smiles, and for allowing memorable undergraduate experiences, and my church members at Oakwood Bible Church, Troy, NY.

Contents

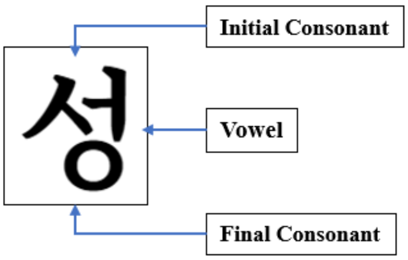
List of Figures

List of Tables

Chapter 1

Introduction

In 1443, King Sejong the Great created Hangul, a Korean alphabet, to address the illiteracy among the lower classes and establish a national identity. Hangul is a scientific alphabet consisting of 14 consonants modeled after the shapes of vocal organs when forming their sounds and 10 vowels that symbolize the sky (·), earth (ㅡ), and human (ㅣ). When these consonants and vowels are combined, they produce 19 initial consonants, 21 vowels, and 27 final consonants, resulting in 11,172 Korean syllables (see Fig. 1.1). Its simplicity has led Korea to have one of the highest literacy rates globally [7] and its scientific and aesthetic beauty is lauded among renowned linguistics [16, 22]. In the supervised image classification tasks, however, this large number of Korean syllables poses great difficulty in Handwritten Hangul Recognition, as the model must handle a much larger number of classes.



Consonants (14)	ㄱ ㄴ ㄷ ㄹ ㅁ ㅂ ㅅ ㅇ ㅈ ㅊ ㅋ ㅌ ㅍ ㅎ
Vowels (10)	ㅏ ㅑ ㅓ ㅕ ㅗ ㅛ ㅜ ㅠ ㅡ ㅣ

Initial Consonants (19)	ㄱ ㄴ ㄷ ㄹ ㅁ ㅂ ㅅ ㅇ ㅈ ㅊ ㅋ ㅌ ㅍ ㅎ ㄲ ㄸ ㅃ ㅆ ㅉ
Vowels (21)	ㅏ ㅑ ㅓ ㅕ ㅗ ㅛ ㅜ ㅠ ㅡ ㅣ ㅘ ㅚ ㅙ ㅝ ㅞ ㅢ ㅒ ㅔ ㅟ ㅖ ㅐ
Final Consonants (27)	ㄱ ㄴ ㄷ ㄹ ㅁ ㅂ ㅅ ㅇ ㅈ ㅊ ㅋ ㅌ ㅍ ㅎ ㄲ ㄳ ㄵ ㄶ ㄺ ㄻ ㄼ ㄽ ㄾ ㄿ ㅀ ㅄ ㅆ

Figure 1.1: **Hangul Formation:** (left) Korean syllable format, (right) Korean syllables.

Existing HHR on the SERI95a dataset has centered on deploying statistical methods and deep convolutional neural networks. A statistical method, which utilizes a statistical classifier on feature vectors, is a technique used in character recognition that has shown high performance in Handwritten Chinese Character Recognition (HCRR) [18]. By using

Modified Quadratic Discriminant Function (MQDF) as its statistical classifier [14], Park *et al.* have achieved an accuracy of 93.71% [20]. In [21], Ryu and Kim continued their work on a statical method by applying a discriminative nonlinear normalization algorithm to improve the accuracy that expands the critical region that distinguishes one character from another as it could be lost during preprocessing step and was able to achieve 90.11% accuracy. In [13], Kim and Xie further continued their work, and applied 4 convolution layered DCNN network architecture based on AlexNet rather than a statistical method. By applying the data augmentation method to increase the number of image samples and reduce overfitting and the edge operator method, which initializes the training convolution masks with good initial weights, they have achieved a higher accuracy of 95.96%. In [12], Kim, *et al.* further improved their work by applying a hybrid learning method on the DCNN network, which utilizes a loss function that learns recognition and discrimination among similar syllables, resulting in an accuracy of 97.67%.

In this paper, we propose a variation of Xception [3] to improve the performance of Handwritten Hangul Recognition to classify 520 most frequently used Korean syllables on the SERI95a database (see Fig. 3.2 and Fig. 3.3). Akin to Xception architecture, the proposed model consists of an entry flow, a middle flow that repeats two times, and an exit flow introduced in [3]. To overcome the limitation of the state-of-the-art method based on DCNN in the learning of feature representations at various levels, a deeper and wider network based on depthwise separable convolution which is an extreme version of an inception module is designed. Through this model configuration, a higher classification performance than both the original Xception architecture and the state-of-the-art method was obtained by achieving an accuracy of 98.10%. Our second contribution is to adopt a progressive training method originally used in the Progressive Growing GAN (ProGAN) [10]. As the well-trained weights from the prior model before increasing the network capacity are utilized, our experiments show that this training approach can stabilize the training process and meanwhile reducing the training time. This training approach facilitates a smooth transition from a small network to a desirable and more complex network, and Fig. 2.1 illustrates the proposed framework with a progressive approach.

Chapter 2

Method

2.1 Xception

Proposed by François Chollet, Xception [3] which stands for "extreme inception"is an extension of the Inception V3 [23]. In this architecture, the Inception modules are replaced with an extreme version of the Inception module. The motivation behind this architecture lies in the fact that the inception module in Inception V3 performs cross-channel correlation first, followed by any n x n spatial correlation. Hence, Chollet interpreted the inception modules as an intermediate point between the regular convolution and the depthwise separable convolution. Thus, an extreme version of the Inception module performs 1x1 convolution for cross-channel correlation followed by 3x3 convolution for spatial correlations independently. In comparison to Inception V3, Xception slightly outperforms on the ImageNet dataset and significantly outperforms on the JFT dataset [8] with greater computational efficiency. Due to this robustness, Xception is also widely used as a backbone model in transfer learning [1, 2, 19, 25].

2.2 Progressive Training

Progressive Training is a widely-accepted training methodology that has revolutionized the field of high-resolution image generation. First introduced in Progressive Growing Generative Adversarial Networks (ProGAN) [10], the approach was later adapted as the base architecture for StyleGAN2 [11]. The concept of progressive training was inspired by several previous works in the field, including GMAN [5] which extends GANs to multiple discriminators, MAD-GAN [6] that consists of multiple generators and one discriminator, and multi-scale generator and discriminator architecture proposed in [24]. Although some adaptations have been made to the original methodology, progressive training is also applied in image classification tasks [4], [26], [17]. Due to the gradual increase in the network capacity and well-trained weights from the prior model before the addition of a new layer, progressive training offers stabilized training with reduced training time.

Algorithm 1 HHR with Progressive Training

With Initial V = $\{V_0, V_n\}$, label y, feature map F:
for $i \in [0, stages\_num)$ **do**
    for $j \in [0, epoch\_num)$ **do**
        for $k \in [0, steps\_num)$ **do**
            if j $<$ epochs_num - 1 **then**
                $\hat{y}_k = Softmax(O(V))$
                Backprop($\mathscr{L}_{CE}(\hat{y}_k, y_k)$)
            else if j $==$ epochs_num - 1 **then**
                $\alpha_k = k/(steps\_num - 1)$
                $F_k = (1 - \alpha_k) * O(V_{0\sim i}) + \alpha_k * O(V_{0\sim i+1})$
                $\hat{y}_k = Softmax(O(V_n(F_k)))$
                Backprop($\mathscr{L}_{CE}(\hat{y}_k, y_k)$)
                if k $==$ steps_num-1 **then**
                    $V \leftarrow Concat(V, V_{i+1})$
                end if
            end if
        end for
    end for
end for
Train Final V = $\{V_0, ..., V_n\}$ for N epochs

2.3 The Proposed Model Architecture

The overall architecture of the model is presented in Fig. 2.1, which comprises an entry flow with six blocks, a middle flow with two blocks, and an exit flow with 3 blocks that contains a convolution layer. Each block is followed by batch normalization [9] although it is not included in the diagram. Furthermore, a regularization was applied by utilizing dropout at every iteration with a probability of 50% before the fully connected layer in order to avoid overfitting. As the input is a small and simple (60,60,1) image rather than a large (299,299,3) image, a variation of Xception rather than its original architecture was used.

In this study, two architectures are explored: the model without progressive training

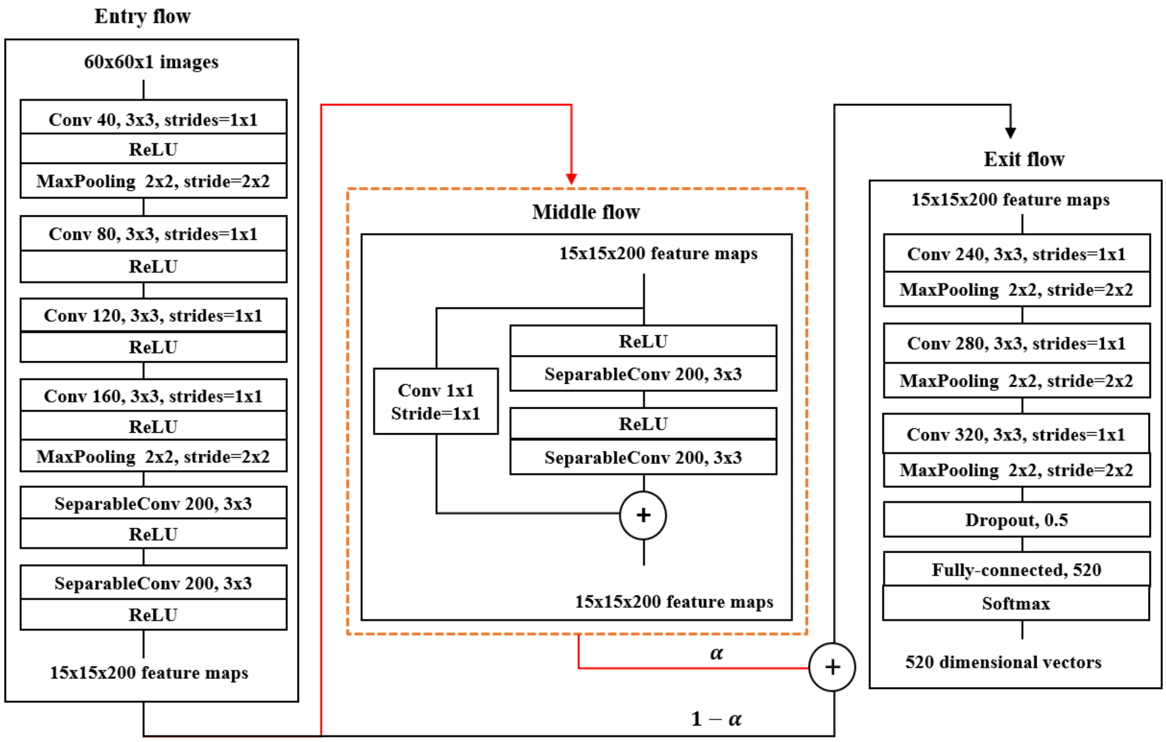


Figure 2.1: **The proposed Handwritten Hangual Recognition (HHR) network architecture.**

and the model with progressive training from ProGAN. The network is denoted by V = $\{V_0,...,V_n\}$, where V_0 is an entry flow, V_n is an exit flow, and V_i is an i^{th} middle flow with n = 3. The output tensor of an i^{th} flow is represented as $O(V_i)$ and the feature map at k^{th} step in an iteration is represented as $F_k \in \mathbb{R}^{H_k * W_k * C_K}$. In the model without progressive training, the training is performed on the model V = $\{V_0,...,V_n\}$. In the model with the progressive technique, however, the initial training is performed on the model V = $\{V_0,V_n\}$, i.e. without middle flows. Let $S \in [0,2)$ be the number of stages during which a new middle flow V_i is added to V, i.e. V = $\{V_0,V_1,V_3\}$ when S=0. During the addition of a new V_{i+1}, 1-α represents the weight of an old output from $V_{0\sim i}$ while α represents the weight of new output from $V_{0\sim i+1}$. For each S, all layers in the network remain trainable. Furthermore, the weight α increases linearly from 0 to 1 and the input to the exit flow is the sum of the old and the new output tensor. Algorithm 1 expresses the proposed model with a progressive training method in pseudocode.

Although the same progressive methodology was used, the primary distinction between training the proposed model with the progressive training and the ProGAN lies in its architecture. As ProGAN is used for image generation rather than classification, it introduces a higher-resolution convolutional layer incrementally, enabling the network to first identify large-scale features and then finer-scale details. In contrast, the proposed method incrementally introduces an equal dimension middle flow of the proposed model rather than introducing a new convolutional layer with an increased dimension. Through this approach, a smooth transition from a small network to a desirable and finer network could be constructed with stabilized training and reduced training time.

Chapter 3

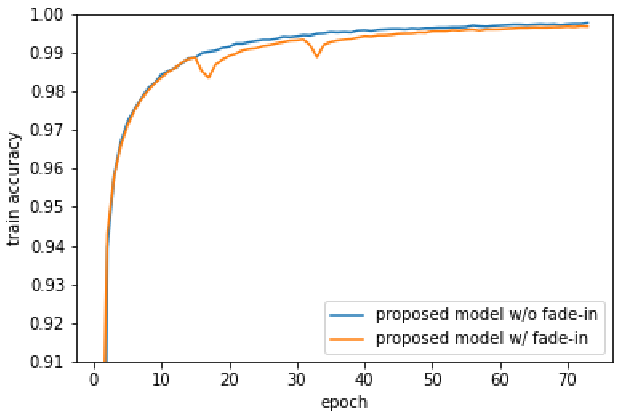
Experiments

3.1 The SERI95a HHR Dataset

A series of experiments were performed on the SERI95a handwritten Hangul database, which was collected by the System Engineering Research Institute (SERI) of Korea University, South Korea. The database comprises approximately $1,000$ instances per 520 most frequently used Hangul syllables. In accordance with prior studies [12, 13, 20, 21], a 90:10 ratio of training to test dataset with input grey-scale images with size 60x60 was used to ensure consistency in the experimental environment.

3.2 Implementation Details

During the training phase, the cross-entropy loss function $\mathscr{L}_{CE}$ = -$\sum_{k=1}^{C} y_k \log(\hat{y}_k)$ with C denoting the number of classes, and Adam optimizer [15] with a learning rate of 0.001 was utilized. In addition, the model for both with and without progressive method was trained for a total of 73 epochs with a batch size of 32. For the model with progressive training, *epochs_num*=16 and *N*=41 in Algorithm 1. To expedite training, a parallel-computing graphics card GeForce RTX 2080 was used, along with Intel Xeon Silver 4214 CPU and Tensorflow 2.8.0.



Method	Test Accuracy	Time per epoch (sec)
Proposed model w/o fade-in	**98.10%**	**317**
Proposed model w/ fade-in	**97.92%**	**Base: 235** **1^{st} insertion: 274** **2^{nd} insertion: 312**
Proposed model w/o middle flows	**97.86%**	**239**
Xception (original)	**97.49%**	**526**
Xception w/ 2 middle flows	**97.32%**	**336**

Figure 3.1: **Methods Comparison and Results.**

3.3 Experiment Results

Fig. 3.1 compares the model with and without progressive training as well as the original Xception. The results indicate that the progressive training approach, which smoothly increases the complexity of the model by utilizing well-trained weights from prior less complex models, enables the model to adapt to small and simple Hangul images (60,60,1) with nearly identical accuracy with reduced training time. Moreover, the proposed model offers benefits in both classification performance and training time compared to the original Xception architecture, demonstrating its effectiveness in HHR.

Method	Accuracy
MQDF [20]	93.71
Discriminative Normalization [21]	90.11
DCNN [13]	95.96
DCNN and Hybrid Method [12]	97.67
Ours	**98.10**

Table 3.1: **Test accuracy comparison.**

Table 3.1 presents the classification success rate of the proposed model on the SERI95a database in addition to the recognition rate from the prior works. The state-of-the-art method, which is based on DCNN and data augmentation, has limitations in learning feature representations at various levels. To overcome this limitation, the proposed model employs a depthwise separable convolution, which is an extreme version of an inception module. In addition, the state-of-the-art method has utilized data augmentation to increase the number of images which has led to an accuracy increase by 0.48% [12]. However, the accuracy of the proposed model is solely derived from the configuration of the model. As less amount of data has been used as we have not changed the number of images, this demonstrates the superior performance of the proposed model in HHR.

3.4 Failure Cases

The images on which the network trained in the proposed model configuration as illustrated in Fig. 2.1 has failed to make a correct prediction can be primarily classified as the following: incorrect labels from the dataset (Type I error) and alike syllables (Type II error), as shown in Fig. 3.3. A fair amount of incorrect label attribution was found in the dataset, and we believe a much higher accuracy could be obtained if the dataset with proper labels were used. In Type II, the embedded syllables are very difficult to recognize as the heavy cursive writing creates (i) extra strokes which could cause a syllable to appear as another syllable and (ii) ’ ㅇ’ and ’ ㅁ’ to be very similar in shape.

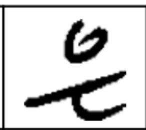
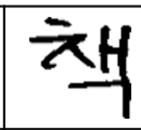
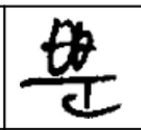
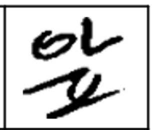
Image					
Prediction	은	빛	책	뿐	앞
Label	은	빛	책	뿐	앞

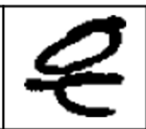
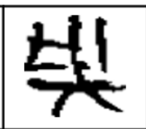
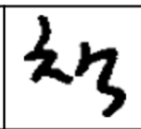
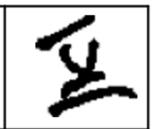
Image					
Prediction	은	빛	책	본	프
Label	은	빛	책	본	프

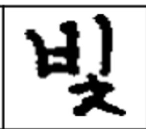
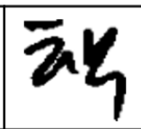
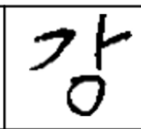
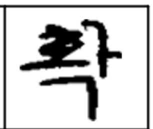
Image					
Prediction	은	빛	책	강	확
Label	은	빛	책	강	확

Figure 3.2: **Classification Results on SERI95a dataset.**

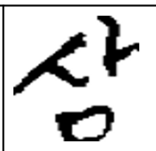
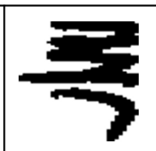
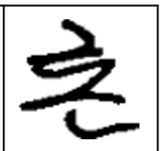
Image (Type I)					
Prediction	은	의	상	록	론
Label	으	외	니	룩	른

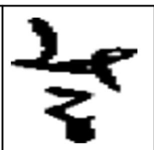
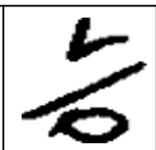
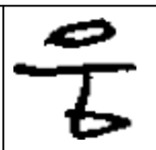
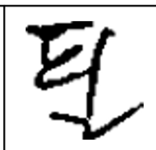
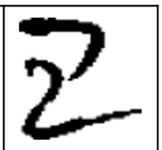
Image (Type II)					
Prediction	놓	능	웅	될	고
Label	렇	승	움	필	코

Figure 3.3: **Misclassified Image Examples.**

Chapter 4

Conclusions

In this paper, a deeper and wider network based on Xception architecture on the SERI95a database was utilized for classification tasks in HHR. The recognition accuracy of 98.10% is obtained, exceeding that of the state-of-the-art method, and has demonstrated the effectiveness of the network on Handwritten Hangul Recognition. Furthermore, this study demonstrates that progressive training can facilitate a smooth transition from a small network to a desirable and more complex network while maintaining nearly identical accuracy and reduced training time.

Future work. Further study should be conducted to evaluate whether the proposed model will still be effective in other languages with a large number of syllables like Hangul, including but not limited to Chinese and Japanese, in the context of handwritten syllable recognition.

Bibliography

[1] Ahmad, S. and Choudhury, P. K. (2022). On the performance of deep transfer learning networks for brain tumor detection using mr images. *IEEE*.

[2] Arbane, M., Benlamri, R., Brik, Y., and Djerioui, M. (2021). Transfer learning for automatic brain tumor classification using mri images. *2020 2nd International Workshop on Human-Centric Smart Environments for Health and Well-being (IHSH)*, pages 210–214.

[3] Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. *CVPR*.

[4] Du, R., Chang, D., Bhunia, A., Zie, J., Song, Y.-Z., Ma, Z., and Guo, J. (2020). Fine-grained visual classification via progressive multi-granularity training of jigsaw patches. *European Conference on Computer Vision (ECCV)*.

[5] Durugkar, I., Gemp, I., and Mahadevan, S. (2016). Generative multi-adversarial networks. *CoRR*.

[6] Ghosh, A., Kulharia, V., Namboodiri, V. P., Torr, P. H. S., and Dokania, P. K. (2017). Multi-agent diverse generative adversarial networks. *CoRR*.

[7] Gökmen, M. E. (1973). The aesthetic features of korean alphabetic system - hangul. *Dil Dergisi*.

[8] Hinton, G., Vinyals, O., and Dean, J. (2015). Distilling the knowledge in a neural network. *In NIPS Deep Learning and Representation Learning Workshop*.

[9] Ioffe, S. and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *ICML*.

[10] Karras, T., Aila, T., Laine, S., and Lehtinen, J. (2018). Progressive growing of gans for improved quality, stability, and variation. *ICLR*.

[11] Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., and Aila, T. (2020). Analyzing and improving the image quality of stylegan. *CVPR*, pages 8107–8116.

[12] Kim, I. J., Choi, C. B., and Lee, S.-H. (2016). Improving discrimination ability of convolutional neural networks by hybrid learning. *International Journal on Document Analysis and Recognition*.

[13] Kim, I. J. and Xie, X. H. (2015). Handwritten hangul recognition using deep convolutional neural networks. *Int. J. Doc. Anal. Recognit.*, 18(1):1–13.

[14] Kimura, F., Takashina, K., Tsuruoka, S., and Miyake, Y. (1987). Modified quadratic discriminant functions and the application to chinese character recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, pages 149–153.

[15] Kingma, D. P. and Ba, J. L. (2015). Adam: A method for stochastic optimization. *ICLR*.

[16] Ledyard, G. K. (1996). The korean language reform of 1446: The origin, background, and early history of the korean alphabet. *Michigan, Ann Arbor: University of Microfilms International: A Ph.D. thesis, University of California, Berkeley*.

[17] Li, C., Miao, F., and Gao, G. (2021). A novel progressive image classification method based on hierarchical convolutional neural networks. *Electronics*, (24).

[18] Liu, H. and Ding, X. (2005). Handwritten character recognition using gradient feature and quadratic classifier with multiple discrimination schemes. *IEEE*.

[19] Lo, W. W., Yang, X., and Wang, Y. (2019). An xception convolutional neural network for malware classification with transfer learning. *2019 10th IFIP international conference on new technologies, mobility and security (NTMS)*, pages 1–5.

[20] Park, G.-R., Kim, I.-J., and Liu, C.-L. (2013). An evaluation of statistical methods in handwritten hangul recognition. *Int. J. Document Analysis and Recognition*, 16(3):273–283.

[21] Ryu, S. J. and Kim, I.-J. (2014). Discrimination of similar characters using nonlinear normalization based on regional importance measure. *International Journal on Document Analysis and Recognition*.

[22] Sampson, G. (1985). Writing systems: A linguistic introduction. *Stanford University Press*.

[23] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). Rethinking the inception architecture for computer vision. *CVPR*.

[24] Wang, T.-C., Liu, M.-Y., Zhu, J.-Y., Tao, A., Kautz, J., and Catanzaro, B. (2017). High-resolution image synthesis and semantic manipulation with conditional gans. *CoRR*.

[25] Wu, X., Liu, R., Yang, H., and Chen, Z. (2020). An xception based convolutional neural network for scene image classification with transfer learning. *In 2020 2nd International Conference on Information Technology and Computer Application (ITCA)*, pages 262–267.

[26] Zhang, Z., Ning, G., Cen, Y., Li, Y., Zhao, Z., Sun, H., and He, Z. (2018). Progressive neural networks for image classification. *Computing Research Repository(CoRR)*.