논문 21-46-08-10 The Journal of Korean Institute of Communications and Information Sciences '21-08 Vol.46 No.08
https://doi.org/10.7840/kics.2021.46.8.1284

이종 문자가 혼합된 필기체 이미지 대상 글자 인식기의 구현

김 홍 숙◆°, 김 정 시*

Implementation of Character Recognizer for Heterogeneous Character Set

Hongsoog Kim◆°, Jeong-Si Kim*

요 약

본 논문에서는 한글, 영문, 숫자 및 특수 기호 4종의 글자들이 혼합된 필기체 이미지를 입력으로 하는 글자 인식 엔진의 구현 과정에 대하여 설명한다. 심층 학습 기반 신경망 모델 학습을 위한 대용량 데이타셋은 폰트 기반의 글자 이미지를 기반으로 데이타 증강 기법을 통하여 확보하였다. 글자 인식 엔진의 핵심부인 CNN 기반 심층 신경망 모델은 고성능 GPU가 장착된 데스크톱에서 학습하였다. 학습이 완료된 신경망 모델의 인식 정확도는 Top-1 accuracy 0.98의 성능을 보였다. 데스크톱에서 학습된 모델은 상대적으로 컴퓨팅 리소스가 한정된 임베디드 시스템상에서 실행할 수 있도록 추론 기능만을 포함하는 심층 신경망 모델로 이식하였다. 임베디드 시스템에 이식된 심층 신경망 모델은 낱글자 인식 및 간단한 양식 내의 단어 인식을 위한 응용 프로그램에서 글자 인식 엔진으로 활용되었다.

Key Words : Deep Learning, Hand-Written Character Recognition, Embedded System

ABSTRACT

In this paper, we present the implementation process of the character recognition engine that inputs hand-written images where four types of characters, Hangul, English, numbers, and special symbols, were mixed. Big data sets for training deep neural network models were prepared through data augmentation techniques based on font-based character images. The CNN-based deep neural network model, the core part of the character recognition engine, was trained on a desktop with a high-performance GPU. The recognition accuracy of the trained neural network model showed a performance of Top-1 accuracy of 0.98 . The model trained on the desktop was ported to a lightweight neural network model that includes only inference capabilities, so that it could be executed on an embedded system of relatively limited computing resources. In the embedded system, the trained model was utilized as a character recognition engine in application programs for individual character recognition and word recognition in a simple form of table.

※ 이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(No.2017-0-00142, 스마트기기를 위한 온디바이스 지능형 정보처리 가속화 SW플랫폼 기술 개발)을 받아 수행된 연구임.
※ A preliminary version of the paper was presented at 1st Korea Artificial Intelligence Conference, pp. 68-69, Jesu Island, Korea, 16-18, Dec. 2020[30].
◆° First and Corresponding Author : Electronics and Telecommunications Research Institute, kimkk@etri.re.kr, 정회원
* Electronics and Telecommunications Research Institute, sikim00@etri.re.kr
논문번호 : 202103-048-C-RE, Received March 1, 2021; Revised April 9, 2021; Accepted May 22, 2021

Ⅰ. 서 론

빅 데이터에 따른 학습에 사용 가능한 데이터양의 증가 및 GPGPU(General Purpose Graphic Processor Unit)를 통한 고속 병렬처리 등의 기술적/환경적 변화에 따라, 이론적인 신경망 모델 제시 수준에 머물렀던 심층 신경망(DNN: Deep Neural Network)의 실제적 학습이 가능해졌다. 이에 따라, 심층 신경망을 활용하는 심층 학습 (deep learning)기술이 급부상하고 있다[1]. 심층 학습 기술은 컴퓨터 비전, 음성 인식, 자율주행차, 로보틱스 등의 다양한 분야에서 활용되고 있다.

기존의 기계 학습 (mahcine learning) 기술에서는, 먼저 응용 분야의 전문 지식을 가진 전문가가 해당 도메인 지식을 사용하여 입력 데이터의 특징을 나타내는 특성들(features)을 찾아서 벡터화하고, 벡터화된 특성들을 기계 학습 알고리즘의 입력으로 사용하여 데이터 특성과 출력 간의 관계를 찾아내는 2단계 방식의 작업을 수행하였다. 반면 심층 학습 기술에서는 데이터 특성들을 도메인 전문가가 찾는 것이 아니라, 빅데이터 기반의 대량 데이터 학습을 통하여 데이터 특성 자체를 심층 신경망 모델에서 자동으로 찾아낸다. 심층 신경망 모델은 이렇게 찾아낸 데이터 특성과 출력 간의 관계도 자동으로 매핑해준다. 즉, 심층 학습 기술은 대량의 데이터로부터 데이터 특성 자체를 학습하는 표현 학습(representational learning) 능력을 특징으로 한다[2]. 심층 학습 기술을 통하여 가능해진 표현 학습 능력 덕분에 많은 응용 분야에서 심층 신경망 모델은 해당 분야의 도메인 전문가조차도 찾을 수 없었던 숨어 있는 중요한 데이터의 특성들을 추출하고, 찾아낸 특성을 활용하여 인간의 성능을 능가하고 있다.

심층 학습 기술은 인간이 시각을 통해 영상 또는 이미지를 인식하는 능력을 심층 신경망 모델로 구현하는 컴퓨터 비전 분야의 다양한 응용에서 활용되어 상용화된 서비스들도 많이 출시되었다. 컴퓨터 비전의 이미지 분류 기술에 해당하는 필기체 인식은 사람이 작성한 문서나 종이에 쓴 글자, 사진에 보이는 글자 등을 인식하는 기술이다.

심층 학습 기술을 활용한 글자 인식 기술은 서구권의 라틴어 계열의 알파벳이나 숫자를 대상으로 한 글자 인식에 있어서 높은 인식률을 이미 달성하였지만, 한글을 대상으로 한 글자 인식률은 그리 높지 못하였다. 이는 한글의 초성, 중성 종성을 조합하여 하나의 글자를 만드는 한글의 구조적 특이성도 있지만, 학습에 필요한 한글 이미지 데이터 세트의 부족이 가장 큰 원인이다.

영문이나 숫자의 경우, 26개 또는 10개의 서로 다른 글자를 인식하면 되지만, 한글의 경우, 조합 가능한 글자들이 실생활에서 전부 사용되지는 않지만, 현대 국어 문법에 맞게 조합 가능한 글자의 개수는 무려 11,172자나 된다[3]. 유니코드에서 동아시아 문자 세트 중, 중국 한자 다음으로 한글이 많은 부분을 차지하고 있다.

심층 학습 기반의 글자 인식에 필요한 학습 데이터 세트의 경우, 영문이나 숫자의 경우, NIST-SD19[4], ICDAR COCO-text[5], SVT[6]등과 같이 인쇄체, 필기체, 장면 텍스트(scene text)에 대한 데이터 세트가 오래전부터 구축되어 심층 학습 기반 글자 인식 기술에 활용됐지만, 한글의 경우 최근에 Ai Hub Korea를 통해 체계적으로 구축된 한글 이미지 데이터 세트[7]를 제외하면, 심층 신경망 모델의 학습 (이하, 심층 학습)에 사용할 만한 규모의 데이터 세트는 부족한 현실이다. 특히 필기체 이미지 데이터 세트의 경우, 수집 및 처리에 많은 시간과 노력이 필요하므로, 단기간에 구축하기도 쉽지가 않은 측면도 있다.

Ⅱ. 문자 인식 관련 연구 동향

CNN(Convolutional Neural Network) 기반의 심층 신경망을 사용한 숫자 및 영어 알파벳 필기체 인식 연구는 1998년 LeCun 등의 연구에서 발표한 LeNet에서 시작되었다고 볼 수 있다[8]. 2011년에 발표된 Dan Clausdiu Ciresan 등의 연구 결과에 따르면, NIST-SD19 데이터 세트를 대상으로 CNN(Convolutional Neural Network) 기반 심층 학습을 통하여 99.73%의 인식률을 달성하였다[9].

아랍 문자의 경우, 총 29자로 구성되는데, Chaouki Boufenar 등의 연구에서는 CNN 기반 심층 신경망 모델을 사용하여 필기체 아랍 문자 인식에서 97.32%의 정확도를 보고하였다[10].

일본 문자의 경우, 카타카나 51종, 히라가나 75종, 한자를 일본어에 맞게 변형한 칸지의 경우 878종을 합쳐 총 1,004개의 문자로 이루어진다. AIST(National Institute of Advanced Industrial Science and Technology)에서 구축한 ETL Character Database[11]을 대상으로 Charlie Tsai의 기술 보고서에 따르면 CNN 기반 심층 학습 후 테스트한 결과 99.53%의 정확도를 보고하였다[12]. 일본어 필기체 인식의 높은 성능은 ETL 데이터베이스에 포함된 120만여 장의 문자 이미지도 큰 역할을 했겠지만, 일본 문자의 경우, 문

자 간 유사성 매우 적어서 글자 인식 면에서 유리한 것으로 분석된다.

한글보다 글자의 종류가 많은 중국의 한자 인식의 경우, 전체 한자의 개수가 3만 6천여 개고 중국 내에 통용되는 한자가 7천여 개, 상용한자가 3,500개 정도이다. Weixin Yangu 등의 연구에서는 CASIA (Chinese Academy of Sciences Institute of Automation)에서 구축한 CASIA-OLHWDB 1.0의 필기체 한자 3,740종과 CASIA-OLHWDB 1.1의 필기체 한자 3,755종으로 구성된 데이터 세트[13]을 대상으로 도메인 지식이 반영된 심층 학습 후 테스트한 결과 96.72% 및 96.35%의 인식률을 보고 하였다[14]. 중국 한자의 경우, 3,800여 종으로 상당히 많지만, 데이터 세트 내의 이미지들이 110만여 장이나 되어, 필기체 인식에서 좋은 결과를 낸 것으로 분석된다.

필기체 한글 인식 연구의 경우, 심층 학습에 필요한 데이터 세트의 부족함에도 불구하고, 관련 연구들이 꾸준히 진행되어 왔다.

김연규 등은 PHD08 (PHD: Printed Hangul Database) 한글 데이터 세트[15]를 대상으로 간소화된 GoogLeNet에 학습시킨 후, PHD08 대상의 검증 세트에 대해 99% 이상의 Top-1 테스트 정확도와 분류 테스트 결과 평균 89.14%의 분류 성공률을 보고하였다[16]. 모델 학습에 사용된 PHD08 데이터 세트는 9종의 한글 폰트를 사용하여 다양한 조건으로 생성된 한글 문자를 인쇄한 후 출력물을 스캔한 이미지를 저장한 데이터베이스로 2008년에 구축되었다.

이규철 등은 음식 메뉴 83개를 대상으로 무료 OCR 엔진인 Tesseract[17]을 이용하여 학습을 수행하고, 테스트 영상을 사용한 추론 시, 추론 결과를 사전에 준비된 단어 사전과 비교하여 유사한 단어로 매칭하는 후처리를 통하여 92.8% 인식률을 보고하였다[18]. 이 연구는 학습 대상 글자가 제한적이라는 한계를 가진다.

이승훈 등은 Tesseract 기반으로 한글 인식 정확도를 향상하기 위해 카메라로 입력받은 영상의 밝기를 조절하고 명암대비를 최대화한 후 이진화 처리와 LPF (Low Pass Filter)를 통해 노이즈를 제거한 후 문자를 인식하여 83.21%의 인식 정확도를 보고하였다[19]. 심층 학습을 하지 않고도 입력 영상의 처리만을 통하여 기존 Tesseract 대비 34.71%의 성능향상을 끌어냈다는 점에서 의미가 있다.

강가현 등은 한글의 초성, 중성, 종성의 모든 경우의 수를 조합하여 글자의 이미지를 생성하여, CNN을 이용한 심층 학습을 수행하고, Tesseract와 성능 비교 결과, Tesseract 대비 58.8%의 성능향상을 보고하였다[20]. 글자 이미지 생성에 있어서 인쇄체만을 사용한 한계와 Tesseract가 라틴 언어 계열에서는 높은 인식 성능을 보이지만, 한글과 같은 동아시아 언어 문자에 대해서는 학습이 부족하여 인식 성능이 떨어지므로, 성능 비교 결과의 공정성 면에서 아쉬움을 남긴다.

박선우는 STR(Scene Text Recognition) 구조를 사용해 변환, 추출, 시퀀스, 예측 모듈에 가능한 24가지 모델 조합에 대하여 성능 평가를 통하여 한글 문장에 적합한 모델 조합을 찾아서 한글 인식 연구를 글자 단위에서 문장 단위로 확장하였다. 데이터 세트는 99종의 폰트를 사용하여 5종의 한글 문장 데이터 세트를 만들어 사용하였다. 테스트 데이터 세트에 AI Hub Korea의 한국어 글자체 이미지 데이터 세트에서 인쇄체 한글과 단어만을 사용했다[21].

한글 인식기 관련 연구의 동향을 정리해보면, CNN 기반의 심층 신경망 모델을 사용한 필기체 또는 인쇄체 한글 인식 연구들이 시도되었으나, 학습에 사용된 데이터 세트가 심층 학습에 사용하기에는 소규모라는 점에서 아쉬움이 있다. 또한, 데이터 세트를 어느 정도 갖춘 경우라도 인쇄체 기반의 데이터 세트를 사용한 점과 라틴어 기반의 문자에 최적화된 Tesseract를 추론 엔진으로 사용하거나, 성능 비교 대상으로 했다는 점에서 한계가 있었다. Tesseract의 경우, 공개 소스로 제공되기에 손쉬운 접근을 할 수 있다는 점에서 이해가 가기는 하지만, 한글 데이터 세트로 Tesseract를 학습을 시도한 논문은 찾아볼 수 없었다. 블로그 등에서 Tesseract OCR4 를 사용하여 한글 학습을 시도한 문서들은 근래에 다수 찾을 수 있었지만, 성능 평가 결과를 찾을 수는 없었다.

본 논문의 구성은 다음과 같다. 3장에서는 한글, 영문, 숫자 및 특수 기호를 포함한 2,448개의 서로 다른 글자들의 필기체 인식을 위한 CNN 기반 심층 학습을 위한 대량의 인쇄체 및 필기체 글자 이미지 데이터 세트의 준비, 데스크톱 서버에서 심층 학습에 관련된 내용을 설명한다. 4장에서는 학습 완료된 심층 신경망 모델의 파라미터들을 임베디드 시스템상에서 학습 기능을 제외한 추론 기능만을 수행하는 신경망 모델에 이식 및 이식된 신경망 모델을 글자 인식 엔진으로 사용하는 글자 인식 응용 프로그램의 개발 과정을 기술한다. 마지막 5장에서는 현재 연구의 한계점에 대하여 정리하고, 추후 연구 방향을 제시한다.

Ⅲ. 심층 학습용 데이터 세트 및 심층 학습

본 연구의 목표인 한글, 영문, 숫자 및 특수 기호가

혼합된 필기체 인식을 위한 CNN 기반 심층 신경망 모델의 성공적 심층 학습을 위해서는 2장의 관련 연구 동향에서 살펴본 바와 같이 대규모의 인쇄체/필기체 글자 이미지 세트의 구축이 선행되어야 한다.

3.1 심층 학습용 데이터 세트의 구축

본 연구에서 목표로 하는 인식 대상 글자의 범위는 KS-X-1001규격에서 정의하는 완성형 한글 2,350자, 알파벳 대소문자 52자, 숫자 10자, 주요 화폐 기호를 포함한 특수 기호 36자를 포함한 총 2,448개이다.

한글의 경우 현대 한글 맞춤법에 따라 조합 가능한 글자의 수는 이론상 11,172종이 가능하지만, 실생활에서 거의 사용되지 않는 글자들이 대부분이고, 데스크톱과 비교해 상대적으로 계산 능력이 떨어지는 임베디드 시스템상에서의 추론을 위하여 CNN 기반 심층 신경망 모델의 구조를 최소화할 필요가 있었기에, 한글 인식 대상 범위를 완성형 한글에서 지원하는 2,350글자로 한정하였다.

심층 학습 및 검증을 위한 필기체 이미지 데이터 세트들은 AI Hub Korea에서 배포하는 한국어 글자체 이미지 데이터 세트, 자체 기보유 중인 한, 영, 숫자 이미지 데이터 세트, NIST-SD19 영문 숫자 필기체 이미지 데이터 세트를 기본으로 하였다.

영문/숫자 필기체 이미지 데이터 세트인 NIST SD19에서 추출한 영문/숫자 이미지의 개수가 총 731,668개이고, 영문 숫자를 합쳐 62개의 글자가 있으므로, 글자당 평균 11,801장의 이미지를 사용할 수 있다.

AI Hub Korea에서 배포하는 한국어 글자체 이미지 데이터 세트에는 학습에 적절하지 않은 글자 이미지가 다수 포함되어 있어, 완성형 한글 필기체 학습에 사용 가능한 그레이 스케일 이미지들만을 필터링한 결과 271,858장의 이미지들만이 사용 가능한 것으로 확인되었다. 이는 전체 글자 이미지 개수 624,359장의 약 35% 정도에 해당하며, 완성형 글자 수 2,350 클래스를 고려하면 글자당 평균 115장 이미지가 된다. 자체 보유 중인 한, 영, 숫자 이미지 세트의 경우, 완성형 한글 글자만을 추출한 결과, 2,071,811개의 이미지를 확보하여, 글자당 이미지 882장을 확보하여, AI Hub Korea의 한국어 글자체 이미지 데이터 세트에서 확보한 글자당 평균 이미지 115장을 더하여도 완성형 한글의 경우, 글자당 이미지의 개수는 평균 997장에 불과하였다. 한글용 데이터 세트의 경우, CNN 기반 심층 학습에 사용하기에는 규모가 작아서 무리하게 학습을 진행한다면, 언더피팅(underfitting) 문제에 봉착할 가능성이 컸다.

성공적인 심층 학습을 위해서는 양질의 대규모 데이터셋이 필수적인 선결사항이다. NIST-SD19 영문 숫자 필기체 이미지 데이터 세트 수준의 한글 글자 이미지 세트 및 특수 기호 36자에 대응하는 이미지 데이터 세트 확보를 위하여 폰트 기반 필기체 글자 이미지 프로그램(ALICE: Automatic Letter Image Creation with Effects)을 사용하였다. ALICE는 주어진 기본 글자 이미지에 회전, 임의 탄성 왜곡(RED: Random Elastic Distortion), X/Y축 Shearing, Erosion/Dilation 등의 이미지 변환 기능을 가진 데이터 증강(data augmentation)용 프로그램이다.

데이터셋 증강을 위하여 사용된 폰트의 수는 한글 폰트 486종, 영문 폰트 171종 모두 총 657종이다. 원본 폰트 글자 이미지에 X축 Shear (4단계) 효과 및 RED 수행 시, alpha=36, sigma=6 및 step=6 설정에 따른 6가지 변환 효과를 사용하였다. 따라서 한 개의 폰트당 한 개의 글자에 대하여 1 + 4 + (4*6) = 29개의 효과로 글자 이미지를 생성하도록 하였다.

한글의 경우, 사용한 총 폰트의 개수가 486개이므로 이론적으로 글자당 29개 효과*486폰트 = 14,094장의 이미지가 가능하고, 완성형 한글의 클래스 개수가 2,350이므로 완성형 한글의 한글 전체에 대하여 계산할 경우, 이론적으로 14,094*2,350= 33,120,900개의 이미지를 생성할 수 있다. 그러나, ALICE에서 로딩하지 못하는 폰트들과 로딩이 된 폰트라도 글자에 따라서 해당 폰트에서 지원되지 않는 경우도 있어서 실제 생성된 이미지의 총 개수는 29,309,807장이었다.

AI Hub Korea 한국어 글자 이미지 데이터 세트를 통하여 확보한 271,858장에 자체 기확보 한글 이미지 2,071,811까지 포함하면 총 31,653,476장의 이미지를 확보하였다. 클래스당 평균 13,469 (=31,653,476/2,350)장으로, NIST-SD19의 114%에 해당하는 양이다.

특수 문자의 경우, 한글 폰트와 영문 폰트를 합친 657종의 폰트를 모두 사용하였으며, 실제 생성된 이미지의 총 개수는 454,227이었다. 특수 문자당 평균 이미지의 개수는 12,617(=454,227/36)장으로 NIST-SD19의 106% 에 해당하는 양이다.

Table 1에서 심층 학습용 데이터 세트 준비 과정을 통해 확보한 데이터 세트에 대하여 통계 정보를 정리하였다. 데이타 셋트의 크기가 부족했던 한글과 특수 문자의 경우, NIST-SD19 데이터 셋트과 비교했을 때, 각각 114% 및 106%로 심층 학습에 충분한 양질의 데이터 세트를 확보하였다.

확보된 32,839,371장의 글자 이미지 파일에 대하

Table 1. Image Dataset Statistics

Type	Source	Total Num. of Images	Num. of Classes	Avg. # of Images per Class
Alphabet, Digits	NIST-SD19	731,668	62	11,801
Korean Syllables	Proprietary, AI Hub, ALICE	31,653,476	2,350	13,469
Special Symbols	ALICE	454,227	36	12,617
Total	-	32,839,371	2,448	13,415

여, 클래스별로 9:1의 비율로 나누어 학습용 세트(training set)로 29,557,881장, 검증용 세트(validation set)로 3,281,490장을 사용하였다. 즉, 글자당 평균 12,074장을 학습에 사용하고, 글자당 평균 1,340장을 검증에 사용하였다.

3.2 데스크톱에서의 CNN 기반 심층 학습

2,448개의 클래스로 이루어진 필기체 인식용 심층 신경망 모델은 ResNet 모델을 기반으로 한글 인식에 맞게 수정하였다. CNN 기반의 심층 신경망 모델의 경우, 레이어가 깊어질수록 학습에 필요한 파라미터 계산에 필요한 그라디언트 계산 과정 중에 그라디언트 소멸 또는 폭발 (gradient vanishing or exploding) 발생 가능성이 커지고, 이에 따라 학습이 잘 이루어지지 않아 오히려 성능이 떨어지는 현상이 발생하는 것으로 알려졌다. ResNet은 어떤 레이어의 출력을 한 단계를 건너뛴 그 다음 레이어의 입력에 바로 연결하게 하는 스킵 커넥션 (skip connection)을 사용하여 최소 그라디언트가 1이 되도록 만들어 줌으로써 그라디언트 소멸의 문제점을 해결하였으며, 이미지 분류 문제에 있어서 우수한 성능을 달성하였다[22]. ResNet은 심층 신경망 모델을 구성하는 레이어 중 학습을 통하여 최적화되는 파라메터가 포함된 레이어, 즉, 컨볼루션 레이어(convolution layer) 및 완전 연결 레이어(fully connected layer)의 개수에 따라, ResNet-18, ResNet-34, ResNet-50, ResNet-101, ResNet-152 등으로 구분한다. 데스크톱과 비교했을 때컴퓨팅 리소스가 상대적으로 제한되는 임베디드 시스템에서 추론을 위하여 작은 규모의 모델인 ResNet-18을 선택하였다.

CAFFE 심층 학습 프레임워크[23]기반의 Github에 공개된 ResNet-18 모델[24]을 수정하여 사용하였다. ResNet 계열의 심층 신경망 모델들은 1,000개의 클래스를 가진 총 백만 개가 넘는 이미지로 구성된 ILSVRC2012 (ImageNet Large Scale Visual Recognition Challenge 2012) 데이터 세트[25]를 학습하는 용도로 설계되었기에 본 연구의 목표인 2,448종의 글자 인식에 맞게 모델 구조를 수정하고, 레이어 파라미터들도 수정하였다. 학습에 사용할 GPU의 성능에 맞게 하이퍼 파라미터도 조정하였다. Fig. 1에서 최종 수정 완료된 심층 신경망 모델(이하 DASH2448: Digit, Alphabet, Symbol, Hangul 2448)의 구조를 도시하였다. 하단 중앙에 있는 블록 (b)를 보면 ResNet 모델의 특징인 스킵 커넥션을 통하여 출력을 다음 레이어를 스킵하고, 그다음 레이어의 입력으로 연결하는 구조의 Residual 블럭을 사용하고 있음을 확인할 수 있다. ResNet-18에서는 Residual 블록을 총 4개 사용하였으나, DASH2448에서는 사물이 아닌 글자를 분류하기 위한 용도이므로 지나치게 추상화된 저해상도

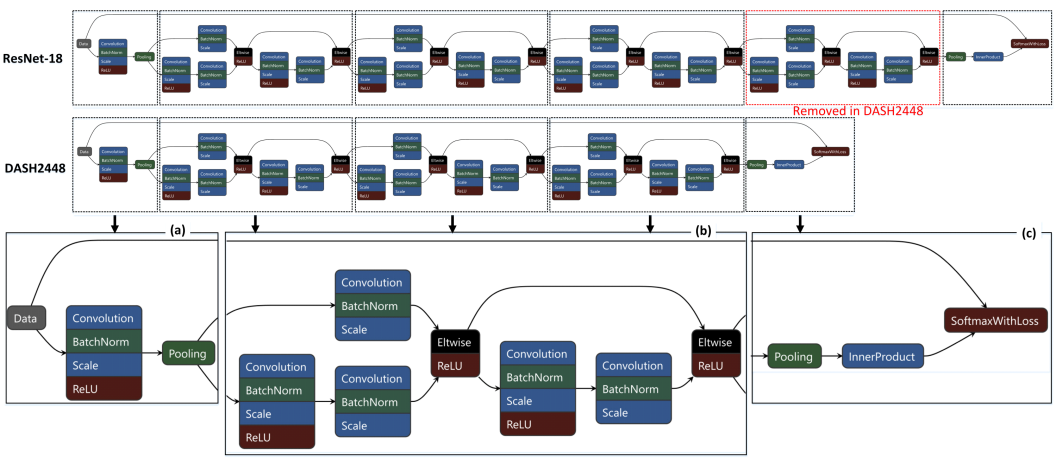

Fig. 1. DASH2448 Deep Neural Network Model Architecture

의 feature map을 피하고, 네트워크 전체 크기를 줄이기 위하여 3개만을 사용하였다.

DASH2448의 학습 및 검증을 위하여 연구용으로 널리 사용되는 Berkeley AI Research (BAIR)/The Berkeley Vision and Learning Center (BVLC) CAFFE 심층 학습 프레임워크를 사용하였다. CAFFE는 C++, Python, Matlab 프로그래밍 인터페이스를 제공한다. 임베디드 시스템상에서의 구현 및 성능을 고려하여 C++ 인터페이스를 지원하는 CAFFE를 선택하였다. CAFFE의 경우, 명령어 기반 인터페이스만을 제공하므로, 실험을 편리성과 학습 과정의 학습률 및 정확도의 변화를 모니터링을 하기 위하여, 독일 뮌스터 대학에서 개발하여 배포하는 파이썬 기반 CAFFE GUI 프로그램인 Barista[26]를 사용하였다. 학습 중 모델 최적화를 위하여 Adam 옵티마이저를 사용하였다. Adam 옵티마이저는 동적으로 learning rate를 변경하므로 base learining rate는 고정값으로 하였다. Table 2는 CNN 기반 필기체 인식용 심층 신경망 모델인 DASH2448의 학습에 사용된 데스크톱 컴퓨팅 하드웨어 사양, 학습 모델, 학습 및 검증용 데이터 세트, 하이퍼 파라메터 설정값들에 대하여 정리한 것이다.

학습용 데이터 세트를 50 epochs동안 학습된 모델을 대상으로 검증용 데이터 세트를 사용하여 정확도를 검증한 결과 Top-1 accuracy 0.9777 및 Top-5 accuracy 0.9988을 확인하였다. Fig. 2는 학습 과정에

Table 2. IHW/SW Setting for DASH2448 Training

HW Spec.	CPU: Intel Core I7-6700K GPU: NVIDIA RTX2080 Ti Memory: 64GB (16GBx4) Deep Learning Framework: BVLC Caffe GUI-based Managing Tool: Barista
Training Model	DASH2448: Resnet-18 Modified version # of Learing Paramters: 3,404,864
Data Set	Training : Validation = 9 : 1 -# of Images in Training Set: 29,557,881 -# of Images in Validation Set: 3,281,490 Input image: 28x28 gray-scale
Hyper Parameter Setting	Total Epochs: 50 - Batch Size for Training: 1,280 - Batch Size for Validation: 256 - Overall Iterations: 1,154,650 Optimizer: ADAM base_lr: 0.001 momentum: 0.9 momentum2: 0.999 lr_policy: fixed

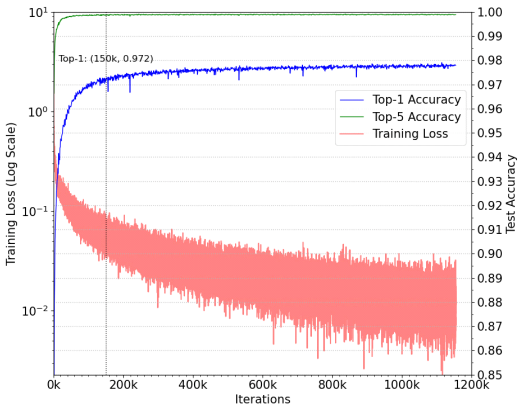

Fig. 2. Top-1/Top5 Accuracy, Loss, Learning Rate

서의 iteration에 따른 Training Loss, Top-1 Accuracy 및 Top-5 Accuracy를 값을 표시한 그래프이다. Fig. 2에서 파란색 커브와 초록색 커브는 각각 Top-1 Accuracy와 Top-5 Accuracy에 나타내며, 로그 스케일로 표시한 빨간색 커브는 Traing Loss를 나타낸다. 150,000 iteration 후에 Top-5 accuracy가 이미 0.97 이상을 달성하였지만, Training Loss의 감소가 지속적으로 관찰되어 조기 종료(Early Stopping) 없이 실험 초기에 설정한 50 epoch 에 해당하는 iteration을 계속하여 진행하였다. 50 epoch 학습에 소요된 시간은 3일 11시간 30분이었다.

Ⅳ. 임베디드 시스템으로의 심층 신경망 모델 이식 및 글자 인식기 기반 응용 프로그램 개발

고성능 GPU가 장착된 데스크톱에서 학습된 심층 신경망 모델을 실제 서비스용 타겟 시스템인 임베디드 시스템상에 이식하였다. 임베디드 시스템상에서 은행에서 사용하는 타행 무통장 입금증을 간략화한 양식에 사용자가 필기한 내용을 웹캠으로 촬영하고, 해당 이미지에 포함된 글자들을 인식하는 응용 프로그램을 구현하여 실용화 가능성을 평가하였다.

웹캠을 통하여 촬영된 필기체 양식 이미지의 경우, 스캐너와 같은 장비를 통하여 스캔한 이미지와 달리 촬영 각도, 주변 광원 등에 의한 노이즈 및 왜곡 효과가 필연적으로 포함될 수 밖에 없다. 따라서 촬영된 이미지에 포함된 노이즈및 왜곡 효과를 제거하기 위하여, 이미지 전처리 기능을 통과한 보정된 이미지를 사용하였다.

Fig. 3은 고정 양식 내의 개별 글자 영역 처리 과정을 도식화한 것이다. 먼저 (a) 단계에서 배경 있거나

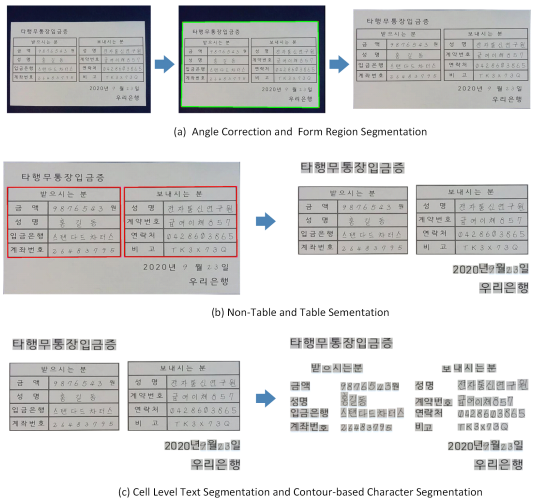

Fig. 3. Preprocessing Details on Sepration of Individual Character in Fixed Form

비스듬히 촬영된 경우, 직사각형화하는 촬영 각 보정(angle correction)을 통하여 양식 영역을 분리하고, 이진화과정의 전처리를 통하여 그레이스케일 이미지로 변환하고, (b) 단계에서 표 영역과 표가 아닌 영역의 텍스트 영역을 분리하고, (c) 단계에서 표 영역 내부의 개별 셀 영역을 구분한 후, 개별 셀 영역 내의 텍스트 영역들을 분리하고, 분리된 텍스트 영역에서 개별 글자 이미지들을 추출한다.

Fig. 4는 확보된 텍스트 영역에서 영역에 등고선 기반 추출(Contour-based Segmentation)기법과 캐니 에지 탐지 기법(Canny Edge Detection) 알고리즘을 과정을 도식화한 것이다.

최종적으로 추출된 개별 글자 이미지들은 DASH2448의 입력으로 주어진다. 응용 프로그램 내부에서는 고정 양식 폼에서의 표 영역의 좌표, 표가 아닌 영역에서는 텍스트의 좌표, 셀의 좌표 정보와 개별 글자들의 좌표 정보를 저장하여, DASH2448을 사용하여 추론한 결과를 입력 이미지에서 대응하는 글자의 좌표 위치에 매핑하는 자료구조를 유지하였다.

임베디드 시스템상에서 필기체 글자 추론 기능은 자체 개발한 추론 가속 라이브러리에 기반한 추론 기능만을 가진 심층 신경망 모델을 작성하고, 데스크톱에서 대규모 데이터 세트를 사용하여 학습된 DASH2448 심층 신경망 모델의 파라미터들을 적재하여 사용하는 전이학습(transfer learning)기술을 사용하였다.

Fig. 5는 임베디드 시스템상에서 Qt 5.12.0 QML 2.0 개발 도구를 사용하여, 개발한 GUI 환경에서 고

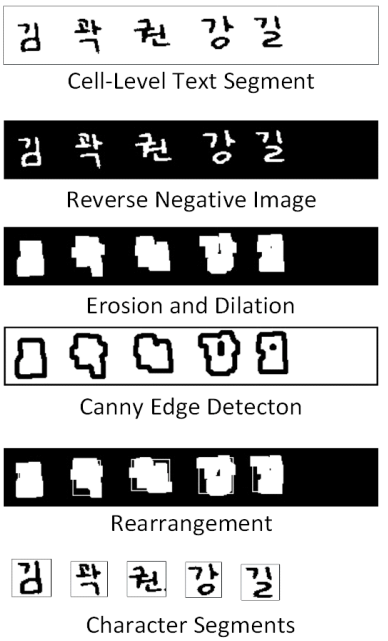

Fig. 4. Character Segmentation

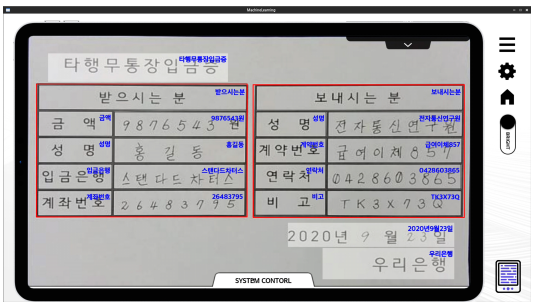

Fig. 5. Input Image with Inference Result Overy

정 양식에서 글자 이미지에 대하여 추론한 결과를 원래의 이미지상에 오버레이 한 최종 인식 결과 화면이다. 결과에서 보듯이 낱글자 영역만을 추출하기 때문에 공백 문자를 처리하지 못하여, 띄어쓰기는 무시되는 문제점이 있다. 간단한 양식 같은 경우 띄어쓰기가 중요하지 않을 수 있으나 문서의 경우 띄어쓰기에 의해 단어가 결정되므로, 중요한 문제가 될 수 있다.

V. 결 론

본 논문에서는 한글, 영문, 숫자 및 특수 기호를 포함한 2,448개의 서로 다른 글자 클래스의 필기체 인식을 위한 심층 학습을 위한 대용량 인쇄체 및 필기체 글자 이미지 데이터 세트의 준비 과정, 데스크톱 서버

에서 모델 학습 과정, 데스크 톱에서 학습된 심층 신경망 모델을 임베디드 시스템으로 추론 기능만을 가진 심층 신경망 모델로 이식하는 과정, 이식된 심층 신경망 모델을 사용하여 고정 양식 내에 포함된 글자들을 인식하는 응용 프로그램 개발 과정 및 이 과정에서 얻은 실무적 경험을 소개하였다.

본 연구는 심층 학습에 필요한 충분한 양의 한글, 특수문자에 대한 빅 데이터 세트를 확보하였다는 점에서 기존 연구와 차별성을 가진다. 학습된 심층 신경망 모델 학습에 대하여 학습용 데이터 세트와 독립된 별도의 검증용 데이터 세트를 통하여 테스트한 결과, Top-1 accuracy 0.9777 및 Top-5 accuracy 0.9988의 높은 정확도를 달성하였다는 점에서 의의가 있다.

임베디드 시스템상의 웹캠을 사용하는 실무 응용 프로그램 개발을 통하여, 카메라 촬영 환경에서는 발생하는 광원 및 촬영 각도등에 따른 글자 이미지 왜곡 현상을 보정하기 위한 전체 이미지 수준의 전처리가 매우 중요하였다는 점을 확인하였다.

본 연구에서는 4종의 다른 글자가 혼합된 필기체 글자 인식 기능의 확인이 주요 목표였기에 텍스트 영역 탐지 기능은 컴퓨터 비전 분야에서 널리 사용되는 기초적인 등고선 기반 글자 추출 기능만을 사용하였다. 따라서 띄어쓰기에 필요한 공백을 인식할 수 없어서 범용적으로 적용하기에는 한계가 있다. 범용적 실무 응용을 위하여, 띄어쓰기를 포함한 일정 크기 이상의 텍스트 영역 자체도 대량의 텍스트 이미지 데이터 학습을 통하여 자동으로 탐지하여야 한다. 이를 위하여 심층 신경망 모델 기반의 객체 탐지(Object Detection) 연구[27-29]의 결과를 활용하여 단어 단위의 텍스트 영역을 탐지하는 연구를 진행하고 있다.

References

[1] V. Sze, et al., “Efficient processing of deep neural networks: A tutorial and survey,” in *Proc. IEEE*, vol. 105, no. 12, pp. 2295-2329, Dec. 2017.

[2] I. Goodfellow, et al., *Deep Learning, MIT Press*, 2016.

[3] J. Byun, “An arrangement of hangul codes in unicode,” in *Proc. 30th Annu. Conf. Human and Cognitive Lang. Technol., Info. Scientist and Eng.*, pp. 234-236. Seoul, Korea, 2018.

[4] *NIST Special Database 19* (2016), Retrieved Feb., 26, 2021, from https://www.nist.gov/srd/nist-special-database-19.

[5] R. Gomez, et al., “ICDAR2017 robust reading challenge on COCO-Text,” in *IEEE 14th IAPR ICDAR*, Kyoto, Japan, Nov. 2017.

[6] *The Street View Text Dataset (SVT)* (2012), Retrieved Feb., 26, 2021, from http://www.iapr-tc11.org/mediawiki/index.php/The_Street_View_Text_Dataset.

[7] *Korean Syllable Image AI Data* (2020), Retrieved Feb., 26, 2021, from https://aihub.or.kr/aidata/133.

[8] Y. LeCun, et al., “Gradient-based learning applied to document recognition,” in *Proc. IEEE*, vol. 86, no. 11, pp. 2278-2324, 1998.

[9] D. C. Ciresan, et al., “Convolutional neural network committees for handwritten character classification,” *2011 Int. Conf. Document Anal. and Recognition*, pp. 1135-1139, Beijing, China, 2011.

[10] C. Boufenar and M. Batouche, “Investigation on deep learning for off-line handwritten arabic character recognition using theano research platform,” *2017 ISCV*, pp. 1-6, Feb. 2017.

[11] *ETL Character Database* (2011), Retrieved Feb., 26, 2021, from http://etlcdb.db.aist.go.jp/

[12] C. Tsai, “*Recognizing handwritten Japanese characters using deep convolutional neural networks*,” Technical Report, Stanford University (2016). Retrieved Apr. 1, 2021, from http://cs231n.stanford.edu/reports/2016/pdfs/262_Report.pdf

[13] *CASIA Online and Offline Chinese Handwriting Databases* (2020) Retrieved Feb., 26, 2021, from http://www.nlpr.ia.ac.cn/databases/handwriting/Download.html.

[14] W. Yang, et al., “Improved deep convolutional neural network for online handwritten Chinese character recognition using domain-specific knowledge,” *2015 13th ICDAR*, pp. 551-555, Tunis, Tunisia, 2015.

[15] D. Ham, et al., “Construction of printed Hangul character database PHD08,” *J. Korea Contents Assoc.,* vol. 8 no. 11, pp. 33-40, 2008.

[16] Y. Kim and E. Cha, "Streamlined GoogLeNet algorithm based on CNN for Korean character recognition," *J. KIICE*, vol. 20, no. 9, pp. 1657-1665, Sep. 2016.

[17] R. Smith, "An overview of the tesseract OCR engine," in *Proc. ICDAR2007*, pp. 629-633, Parana, Argentina, Sep. 2007.

[18] G.-C. Lee and J. Yoo, "Development an android based OCR application for Hangul food menu," *J. KIICE*, vol. 21, no. 5, pp. 951-959, May 2017.

[19] S.-H. Lee, et al., "Korean prescription character recognition system using OCR technology," in *KSC 2017*, pp. 362-364, Busan, Korea, Dec. 2017.

[20] G.-H. Kang, et al., "A study on improvement of Korean OCR accuracy using deep learning," in *Proc. KIICE 2018*, vol. 23, no. 1, pp. 693-695, May 2018.

[21] S.-W. Park, "A study on the OCR of Korean sentence using deep learning," in *Proc. 31th Annu. Conf. Human and Cognitive Lang. Technol.*, pp. 470-474, Daejeon, Korea, Oct. 2019.

[22] K. He, et al., "Deep residual learning for image recognition," *29th IEEE Conf. CVPR2016*, pp. 770-778, Las Vegas, USA, Jun. 2016.

[23] Y. Jia, et al., "Caffe: Convolutional architecture for fast feature embedding," in *Proc. 22nd ACM Int. Conf. Multimedia*, pp. 675-678, Orlando Florida, USA, Nov. 2014.

[24] *ResNet-18 Caffemodel on ImageNet* (2019), Retrived Feb., 26. 2021, from https://github.com/HolmesShuan/ResNet-18-Caffemodel-on-ImageNet.

[25] *ImageNet Large Scale Visual Recognition Challenge* (2012). Retrived Apr., 2. 2021, from http://www.image-net.org/challenges/LSVRC/2012/results.html.

[26] S. Klemm, et al., "Barista - A graphical tool for designing and training deep neural networks," *CoRR*, arXiv preprint: 1802.04626, 2018.

[27] W. Liu, et al., "SSD: Single shot MultiBox detector," *ECCV LNCS*, vol. 9905, Springer, 2016.

[28] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," *IEEE Conf. CVPR 2017*, pp. 6517-6525, Honolulu, USA, 2017.

[29] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *CVPR*, 2018. Retrieved Feb., 26. 2021, from https://arxiv.org/abs/1804.02767

[30] H. Kim and J.-S. Kim, "Implementation of hand-written letter recognizer for text of Korean syllable, alphabet, digit and special symbol," *1st Korea Artificial Intell. Conf.*, pp. 68-69, Jeju Island, Korea, Dec. 2020.

김 홍 숙 (Hongsoog Kim)

1994년 2월 : 서강대학교 컴퓨터공학과 학사

1996년 2월 : 서강대학교 컴퓨터공학과 석사

2003년 3월 : 한국과학기술원 컴퓨터시스템 및 이론 트랙 박사

1996년 2월~1998년 2월 : 현대정보기술(주) 정보기술연구소 선임연구원

2001년 2월~2004년 9월 : 엔솔바이오사이언스(주) 바이오인포매틱스 연구소 개발팀장

2004년 10월~2004년 12월 : 한국정보통신대학원 대학교 부설 영재교육원 강사

2005년 4월~현재 : 한국전자통신연구원 인공지능연구소 책임연구원

<관심분야> Deep Learning for Vision Applications, Parallel Processing with Compiler Technologies, SON in 5G Networks.

[ORCID:0000-0003-3142-2700]

김 정 시 (Jeong-Si Kim)

1992년 2월 : 경상대학교 전산학과 학사
1994년 2월 : 경상대학교 전산학과 석사
1999년 2월 : 경상대학교 전산학과 박사
2000년1월~11월 : 한국과학기술원 Post-Doc.
2000년 12월~현재 : 한국전자통신연구원 책임연구원
<관심분야> 임베디드SW, 온디바이스 AI 컴퓨팅, 병렬 컴퓨팅
[ORCID:0000-0002-8460-2695]