
Convolutional Neural Network를 통한 대규모 한글 데이터 학습

김연규* · 차의영*

*부산대학교

Learning of Large-Scale Korean Character Data through the Convolutional Neural Network

Yeon-gyu Kim* · Eui-young Cha*

*Pusan National University

E-mail : dusrb2003@naver.com, eycha@pusan.ac.kr

요 약

CNN(Convolutinal Neural Network)을 사용하여 다양한 분야에 대한 심화 학습이 진행되고 있으며 이 이미지 인식 분야에서 특히 높은 성능을 보이고 있다. 본 논문에서는 5,000,000개 이상의 대규모 한글 문자 데이터베이스를 사용하여 한글을 Convolutional Neural Network에 학습 시킨 후 테스트 정확도를 확인한다. 실험에 사용된 CNN 구조는 AlexNet에 기반하여 새로 만들어진 KCR(Korean Character Recognition)-AlexNet 이며 학습 결과 98% 이상의 테스트 정확도를 보였다. 실험에 사용된 데이터베이스는 대규모 한글 데이터 데이터베이스인 PHD08로 총 2,350개의 한글 문자에 대해 각 문자마다 2,187개의 샘플을 가져 총 5,139,450 개의 데이터가 존재한다. 본 연구를 통해 KCR-AlexNet이 한글 데이터베이스인 PHD08을 학습하는데 우수한 구조임을 보인다.

ABSTRACT

Using the CNN(Convolutinal Neural Network), Deep Learning for variety of fields are being developed and these are showing significantly high level of performance at image recognition field. In this paper, we show the test accuracy which is learned by large-scale training data, over 5,000,000 of Korean characters. The architecture of CNN used in this paper is KCR(Korean Character Recognition)-AlexNet newly created based on AlexNet. KCR-AlexNet finally showed over 98% of test accuracy. The experimental data used in this paper is large-scale Korean character database PHD08 which has 2,187 samples for each Korean character and there are 2,350 Korean characters that makes total 5,139,450 sample data. Through this study, we show the excellence of architecture of KCR-AlexNet for learning PHD08.

키워드

심화 학습, Convolutional Neural Network, 한글 학습, PHD08

1. 서 론

영상 인식 분야에 CNN(Convolutional Neural Network)이 본격적으로 사용되기 시작하면서 여러 상황에서의 성능이 향상되고 있다[1-3]. 문자 인식 분야에서도 CNN을 적용하여 인식 성능을 개선시키는 연구가 진행되고 있으며 중국어 필

기체 인식 연구에서 뚜렷한 성과를 보여주고 있다. Zhong[4] 등은 학습 데이터마다 방향에 따른 특징 맵을 생성한 후 기존의 학습 데이터와 함께 CNN에 학습시켜 향상된 학습 정확도를 보였다. Yang[5] 등은 한문의 도메인 정보를 이용하여 CNN에 적용하는 실험을 진행하였다. 하지만 한글학습에 대해서는 CNN을 이용한 심화학

습이 연구되고 있지 않다. 따라서 본 논문에서는 한글 인식을 위한 CNN 구조를 설계하고 합대성 [6] 등이 제작한 대규모 한글 데이터베이스인 PHD08을 사용하여 한글 학습을 진행하였다.

먼저 2장에서 Convolutional Neural Network에 대한 간단한 소개를 한 후 본 논문에서 제안하는 KCR-AlexNet의 구조를 살펴본다. 그 후 3장에서 KCR-AlexNet에 PHD08을 학습시킨 결과를 분석해보고 4장에서 결론을 요약한다.

II. 한글 인식을 위한 CNN

2.1 Convolutional Neural Network의 소개

입력 데이터의 지역적 특징 정보를 추출하여 학습에 사용하는 CNN은 숫자를 인식하기 위해 LeCun[7]이 설계한 LeNet부터 본격적으로 사용되기 시작했다. 컨볼루션, 풀링, 완전 연결과 같은 계층을 반복적으로 사용하는 신경회로망의 일종인 CNN은 이미지 인식을 위한 심화학습에서 가장 많이 사용되는 구조로 자리매김했다. 최근 비선형 활성화 함수인 ReLU(Rectified Linear Unit)계층과 Drop-out 계층이 등장하면서 오버피팅을 줄이고 성능이 더욱 향상되는 모습을 보이고 있다. CNN의 컨볼루션 계층에 데이터가 입력되면 출력 값은 다음 식 (1)과 같이 계산되어 다음 계층의 노드로 전달된다.

$$x_i^l = \sum x^{l-1} * k_i^l + b_i^l \quad (1)$$

여기서 x_i^l 는 l^{th} 계층의 i^{th} 노드가 가지는 연산 결과 값이며 이는 l^{th} 계층에 연결되어 있는 i^{th} 커널 맵과 $l-1^{th}$ 계층의 연결된 노드 값을 곱해 누적한 결과에 바이어스 b_j^l 을 더한 값으로 계산된다. CNN의 최종 계층에서 결과가 출력되면 출력 값이 실제 값과 얼마나 가까운지 나타내는 값으로 흔히 Loss 또는 Error 라는 용어를 사용하며 전체 데이터 셋 D의 모든 데이터 객체 $|D|$ 에 대한 평균 Loss값은 다음 식 (2)와 같이 계산된다.

$$L(W) = \frac{1}{|D|} \sum_i f_w(X^{(i)}) + \lambda r(W) \quad (2)$$

W 는 현재 네트워크의 가중치 매개변수 맵이며 $f_w(X^{(i)})$ 는 데이터 객체 $X^{(i)}$ 의 Loss값이다. $r(W)$ 는 가중치 상수 λ 와 함께 사용되는 제약 조건이다. 식 (2)에서 Loss 값의 계산이 완료되면 학습마다 Loss 값의 변화에 대해 가중치 매개변수 맵을 새로 업데이트해야 하며 이는 식 (3)과 같이 SGD(Stochastic Gradient Descent) 방식을 사용하여 수행된다. 최종적으로 Loss 값을 최소화하는 것이 학습의 목표이다.

2.2 KCR-AlexNet

본 논문에서 설계한 CNN 구조인 KCR-AlexNet은 Krizhevsky의 AlexNet[8]에 약간의 변형을 준 구조이다. AlexNet은 이미지 인식 대회인 ILSVRC(ImageNet Large Scale Visual Recognition Challenge)-2012[9]에서 우승한 CNN 구조이다. KCR-AlexNet은 기존의 AlexNet이 가지고 있는 계층들의 구조와 같은 구조로서 그림 1과 같이 5개의 컨볼루션 계층, 3개의 최대 풀링 계층, 3개의 완전 연결 계층이 사용되었다. 또한, KCR-AlexNet의 모든 컨볼루션 계층과 완전 연결 계층에는 ReLU 활성화 함수가 적용된다. 기존 AlexNet과의 차이점은 출력 계층에서 ILSVRC 클래스를 위한 1,000개의 노드를 가지는 대신에 PHD08이 가지는 클래스의 개수인 2,350개의 노드를 가지며 입력 데이터의 크기도 글자 크기를 위해 기존보다 작은 56x56의 데이터가 사용된다. ILSVRC에서 제공하는 자연 영상 데이터 크기인 256x256보다 작은 크기이다.

III. 실험 결과

3.1 실험 데이터

학습에 사용된 한글 데이터베이스는 PHD08이다. PHD08은 다양한 조건으로 생성된 한글 문자를 인쇄한 후 스캔한 이미지를 바이너리 형태로 저장한 데이터베이스이다. KS 완성형 문자인 2,350개의 한글 문자에 대해 각 문자별로 2,187개의 샘플을 가지고 있으며 샘플들은 각각 서로 다른 폰트, 글자 크기, 스캔 해상도, 노이즈 레벨, 회전, 각도 등을 가지고 있다. 실험은 PHD08의 바이너리 데이터를 이진 이미지로 바꾸어 CNN의 입력 데이터로 사용한다. 단, CNN의 입력 데이터는 모두 같은 크기의 데이터를 사용해

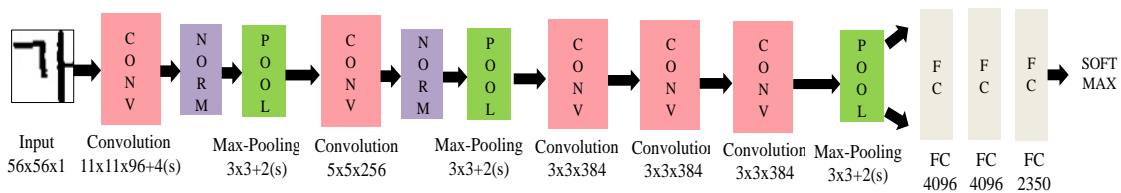


그림 1. KCR-AlexNet 의 구조

야 한다는 제약사항이 있으므로 기존 다양한 크기를 가진 PHD08 데이터들을 일괄적으로 56x56 크기로 변경한 후 사용한다. 크기를 변경하는 방법으로는 선형 보간법이 사용되었다. 최종적으로 사용되는 한글 입력 데이터의 예시는 그림 2에서 확인할 수 있다.

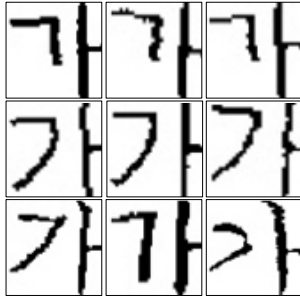


그림 2. PHD08로부터 변형된 입력데이터의 예시

본 논문에서는 학습에 사용된 데이터 수와 테스트에 사용된 데이터 수의 비율을 1:1로 하여 2,350개의 한글 문자에 대해 각각 1,093개와 1,094개의 샘플을 할당하여 실험을 진행하였다.

3.2 실험 환경

KCR-AlexNet에서 사용하는 여러 매개변수들의 초기 값을 설정한다. 학습에서 가장 중요한 역할을 하는 가중치 매개변수를 수정하는 방법은 2장에서 제시된 바와 같이 SGD방식을 사용하며 이에 사용되는 매개변수의 초기 값으로 학습율은 0.01, 모멘텀 상수로 0.9를 사용한다. 또한, 학습율은 매 학습 단계 10,000번 마다 0.96 배씩 작아진다. 매 학습 단계마다 입력 데이터로 사용되는 이미지의 개수를 뜻하는 배치 사이즈는 학습 단계에서 56, 테스트 단계에서 50이 사용되었다. 실험에는 빠른 연산을 위해 그래픽카드 GTX-970 (CUDA v7.0)을 통한 병렬 연산이 사용되었으며 공개 프레임워크인 CAFFE[10]를 통해 진행되었다.

3.3 실험 결과

PHD08이 가진 데이터를 학습과 테스트 데이터로 각각 1:1로 나눈 후 KCR-AlexNet에 학습시키면서 나타난 학습 단계에 따른 테스트 정확도를 그림 3에서, Loss 값의 변화를 그림 4에서 확인할 수 있다. 본 실험에서 100,000번의 학습 단계가 종료된 후 테스트 정확도와 Loss 값의 정확한 수치는 표 1에 명시하였다.

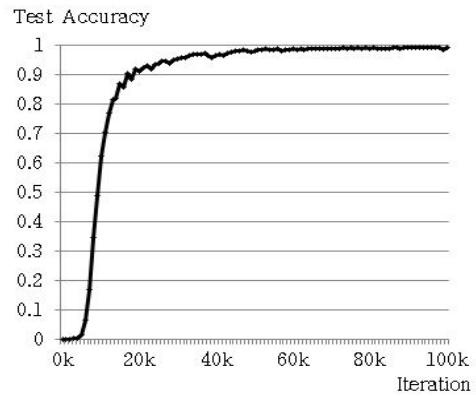


그림 3. 학습 단계에 따른 테스트 정확도

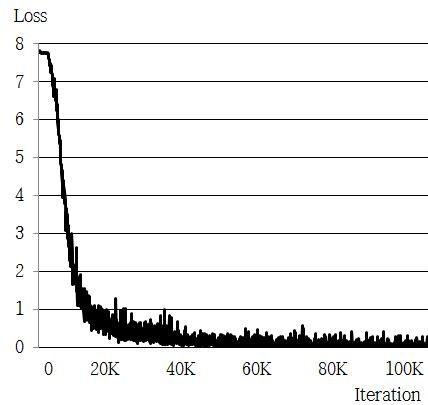


그림 4. 학습 단계에 따른 Loss

표 1. 학습 단계의 시작점과 종료시점에서의 테스트 정확도와 Loss 값

	테스트 정확도	Loss
학습 시작 단계	0.042 %	7.79
학습 종료 단계	98.49 %	0.017

표 1과 같이 학습 종료시점에서 98.49%라는 높은 테스트 정확도를 보인 것은 큰 의미가 있다. 먼저, 5백만개 이상의 데이터를 가진 PHD08 내부적으로 KCR-AlexNet이 98.49%의 분류 성공률을 보인다는 의미이며 이는 PHD08이 가진 폰트에 한해서는 상당히 높은 인식 정확도를 보여준다는 의미이다. 이는 기존 현 시점에서 존재하는 한글 데이터베이스 중 가장 큰 규모인 PHD08에서 얻어낸 수치라는 점에서 앞으로의 OCR 분야에 대한 활발한 응용이 기대되는 점이다.

IV. 결론

본 논문에서는 대규모 한글 데이터베이스인 PHD08을 이미지화 하여 KCR-AlexNet에 학습

시킨 후 테스트 정확도와 Loss 값의 변화 추이를 통해 KCR-AlexNet이 한글 이미지 데이터를 학습하는데 적절함을 보였다. 최종 적으로 학습이 종료된 100,000번의 학습 단계 시점에서 98.49%라는 높은 테스트 정확도를 보였다.

하지만 본 논문은 여전히 학습 데이터에 대해 한계가 존재한다. PHD08이 5백만개 이상의 대규모 데이터베이스지만 사용된 폰트는 여전히 유한하며 이를 통한 학습만으로는 사람이 작성하는 필기체 인식에는 한계점이 존재한다. 즉, 인쇄체 폰트가 아닌 필기체 한글의 인식 정확도는 여전히 해결되지 않은 과제이다. 이를 위해 KCR-AlexNet을 통한 추가적인 학습이 진행되어야 할 것이며 이는 필기체 인식을 위한 추가 학습 데이터베이스에 기반하여 이루어 져야한다. 또는 새로운 CNN 구조를 통해 나온 인식 성공률을 기대해야 할 것이다.

참고문헌

- [1] D. Cireşan, U Meier, J. Masci and J. Schmidhuber, "Multi-column deep neural network for traffic sign classification," *Neural Networks*, vol. 32, pp. 333-338, Aug. 2012.
- [2] N. Kalchbrenner, E. Grefenstette and P. Blunsom, "A Convolutional Neural Network for Modelling Sentences" arXiv preprint arXiv:1404.2188, 2014.
- [3] P. Callet, C. Viard-Gaudin and Dominique Barba, "A Convolutional Neural Network Approach for Objective Video Quality Assessment," *IEEE Transactions on Neural Networks*, vol. 17, no. 5, pp. 1316-1327, 2006.
- [4] Z.Zhong, L. Jin and Z. Xie, "High performance offline handwritten Chinese character recognition using GoogLeNet and directional feature map," in *Document Analysis and Recognition(ICDAR)*, 13th International Conference on. IEEE, pp. 846-850, 2015.
- [5] W,Yang, L. Jin, Z. Xie and Z.Feng, "Improved deep convolutional neural network for online handwritten Chinese character recognition using domain-specific knowledge," in *Document Analysis and Recognition(ICDAR)*, 13th International Conference on. IEEE, pp. 551-555, 2015.
- [6] D. S. Ham, D. Y. Lee, I. S. Jung and I. S. Oh, "Construction of Printed Hangul Character Database PHD08," *Journal of the Korea contents association*, vol. 8, no. 11, pp.33-40, Nov, 2008.
- [7] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-Based Learning Applied to Document Recognition," in *Proceeding of the IEEE 86.11*, pp. 2278-2324, 1998.
- [8] A. Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Network," in *Advances in neural information processing systems*, pp. 1097-1105, 2012.
- [9] ImageNet Large Scale Visual Recognition Challenge Available: <http://image-net.org/>
- [10] Caffe(Convolutional Architecture for Fast Feature Embedding) Available: <http://caffe.berkeleyvision.org/>