



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

공학석사학위논문

필기체 인식을 위한

컨벌루션 신경망 모델

Convolutional Neural Network Models for
Handwriting Recognition

2015년 8월

서울대학교 대학원

전기.컴퓨터공학부

김태준

논문 초록

요약

모바일 기기가 대중화 되면서 필기체 인식은 더욱 중요해졌고, 이와 관련해 필기체 인식의 성능을 높이기 위한 다양한 방법이 연구되고 있다. 필기체 인식의 성능을 높이기 위한 방법으로는 주로 인공 신경망을 이용한 연구가 활발히 이루어졌는데, 그 중에서도 컨벌루션 신경망을 이용한 연구가 좋은 결과를 보여주었다. 컨벌루션 신경망은 심층 학습의 일종으로 커널이라는 필터를 이용해 학습 데이터에 다양한 변화를 주는데, 본 논문에서는 이러한 특징을 앙상블 기법과 연관시켜 보았다. 또한 컨벌루션 신경망에 대해 많은 연구가 이루어지지 않은 모델 복잡도에 대해 실험해보며 기존의 신경망과 어떤 차이점이 있는지를 알아보았다. 컨벌루션 신경망의 모델 복잡도에 대해 알아보기 위해 계층의 수, 학습 횟수, 파라미터의 수 등을 변화시키며 모델 복잡도와 컨벌루션 신경망의 성능의 관계를 알아보았고, 기존 신경망에 비해 견고한 특징이 있는 것을 확인하였다.

주요어: 컨벌루션 신경망, 필기체 인식, 앙상블, 모델 복잡도

학번: 2013-20780

목차

1. 서론	1
1.1 연구배경	1
1.2 관련연구	2
2. 컨벌루션 신경망	3
2.1 배경	3
2.2 컨벌루션 계층	8
2.2.1 커널	8
2.2.2 가중치 공유	9
2.3 서브 샘플링 계층	10
3. 앙상블 기법	10
4. 실험 방법 및 결과	14
4.1 실험 데이터	14
4.2 실험 내용	14
4.3 결과 분석	19
5. 결론 및 향후 연구	20

그림 목차

그림 1 소문자 a와 대문자 A의 입력 예	1
그림 2 다층 퍼셉트론 구조	4
그림 3 CNN과 시각피질의 구조	7
그림 4 Kernel 구조	9
그림 5 Weight sharing 구조	9
그림 6 Pooling 구조	10
그림 7 배경 알고리즘	12
그림 8 부스팅 알고리즘	13
그림 9 MNIST 필기체 데이터 예시	14
그림 10 4-layer CNN 구조	15
그림 11 6-layer CNN 구조	15
그림 12 CNN 계층 수에 따른 성능	16
그림 13 학습 횟수에 따른 성능 비교	17
그림 14 CNN 파라미터 수에 따른 성능	17
그림 15 NN 파라미터 수에 따른 성능	18

1. 서론

1.1 연구배경

필기체 인식은 손으로 쓴 글씨를 종이나 사진, 터치 인터페이스 등을 통해서 입력 받아 인식하는 기술이다. 스마트폰과 태블릿 PC의 대중화로 터치 인터페이스의 중요성이 커지면서 스마트 디바이스 상에서의 필기체 인식의 중요성은 더욱 커지고 있다. 그러나 그림 1에서 볼 수 있는 것처럼 필기체 데이터는 작성자의 필기 습관과 방법에 따라서 같은 문자라도 그 모양이 크게 변화하게 되며, 동일한 작성자의 경우에도 글씨를 입력하는 환경 및 방법에 따라서 모양 및 크기 등에서 큰 차이가 발생하게 된다. 이러한 특성으로 인해 필기 데이터는 데이터에 존재하는 분산이 매우 크고, 이로 인해서 일반적인 기계학습 알고리즘은 학습에 많은 어려움이 발생하게 된다. 이러한 경우에 발생하는 가장 일반적인 상황은 데이터의 크기에 따른 학습 시간의 지수적 증가하거나 학습을 위해 사용한 모델이 수렴하지 않고 발산하게 되는 것이다. 필기체 인식의 또 다른 어려움은 많은 클래스 수이다. 영어 알파벳에는 총 52개의 클래스(대문자 26개, 소문자 26개)가 존재하고, 여기서 숫자나 특수문자 등이 추가되어 클래스 수가 증가하게 되어 필기체 인식을 더욱 어렵게 만든다[1].

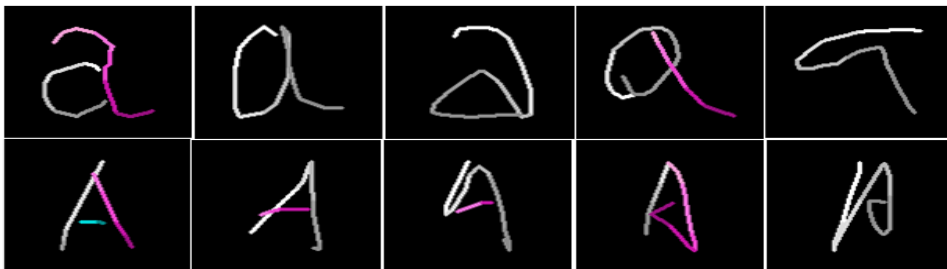


그림 1 소문자 a와 대문자 A의 입력 예

온라인 필기체 인식에서는 광학 문자 인식(Optical character recognition, OCR) 등에서 사용하는 이미지로부터 얻어낸 오프라인 특징(Offline feature)와 함께 획순과 획의 방향 등의 온라인 특징(Online feature)를 동시에 사용하게 된다. 텍스트 이미지로부터 오프라인 특징을 추출하기 위해서는 히스토그램(Histogram), 에지맵(Edge maps), 이미지 사영(Image projections) 등의 방법이 주로 사용되며, 온라인 특징을 추출하기 위해서는 문자의 획을 긋는 순서, 방향, 속도 등의 추가적인 정보를 이용하게 된다[1-3].

1.2 관련연구

스마트 폰이 대중화 되면서 필기체 인식의 성능을 높이기 위한 연구는 더욱 활발히 이루어 졌다. 전통적으로는 데이터들의 분포를 파악해 분산이 가장 큰 방향 벡터를 분석하는 주성분 분석(Principal Component Analysis)과 조건부 확률을 이용해 현재의 상태에서 가장 적절한 다음 상태로 변화할 확률을 구하는 은닉 마코브 모델(Hidden Markov Model)을 이용한 방법[4], 그리고 다양한 분류기를 조합해 성능을 높이는 방법[5] 등이 연구되었다. 최근에는 필기체 인식 문제를 해결하기 위해 회귀형 신경망(Recurrent neural network)[6]이나 심층학습(Deep learning)[7], KNN과 SVM을 결합한 하이브리드 모델[8]을 이용한 다양한 시도들이 좋은 결과를 보이고 있다. 특히 심층학습 중에서도 컨벌루션 신경망(Convolutional Neural Network)은 음성이나 사진 같은 복잡한 실세계 데이터에 매우 좋은 성능을 보여준다.

본 논문에서는 컨벌루션 신경망의 계층별 특징과 기존의 인공 신경망과 어떤 차이가 있는지를 모델 복잡도 관점에서 알아 보았다. 컨벌루션 신경망은 기본적으로 컨벌루션 계층과 서브 샘플링 계층으로 이루어 지는데, 각 계층에는

커널과 통합이라는 기법을 사용한다. 컨벌루션 계층에서 사용하는 커널은 데이터를 리샘플링 하는 것과 비슷하다. 이는 한정된 학습 데이터에서 다양한 특징들을 추출할 수 있게 해주기 때문에 앙상블 기법과도 연관시켜볼 수 있다.

신경망의 모델 복잡도에 영향을 주는 파라미터들은 계층의 수와 뉴런의 수, 그리고 학습 횟수가 있다. 본 논문에서는 이 3가지 요소들을 변화시키며 어떤 변화가 발생하는지를 알아보았다. 컨벌루션 신경망의 모델 복잡도와 관련된 실험은 아직 활발하게 연구되지 않은 분야이다. 베이지안 모델의 복잡도와 관련된 연구[9]는 많은 연구가 이루어 졌지만 컨벌루션 신경망의 모델 복잡도에 관한 연구는 그렇지 않기 때문에 그 점에서 의의가 있다.

2. 컨벌루션 신경망

2.1. 배경

신경망은 인간 두뇌의 신경생물학적 학습과 계산방법을 모방한 모델이다. 단일 뉴런 j 는 두 단계로 정보를 처리한다. 우선 뉴런의 시냅스로 들어오는 활성화 x_i 로부터 신경망의 입력 net_j 을 계산한다.

$$net_j = \sum_{i \in \text{Pre}}^I w_{ji} x_i \quad (1)$$

그리고 신경망의 입력을 sigmoid함수와 같은 비선형 함수 $\sigma(\cdot)$ 를 이용해 변환한다.

$$\sigma(net) = \frac{1}{1 + \exp(-net)} \quad (2)$$

뉴런은 일반적으로 이전 뉴런 층으로 연결되는 층상 구조로 구성되고, 같은 층의 뉴런간에는 연결이 없다. 가장 잘 알려진 신경망 구조 중 하나는 입력 층(Input layer)과 출력 층(Output layer)사이에 하나 이상의 은닉 층(Hidden layer)을 추가하는 다층 퍼셉트론(Multilayer Perceptron, 이하 MLP)이다. 그림 2를 보면 은닉 층의 수가 많아질수록 분류(Classification) 성능이 좋아지는 것을 볼 수 있다. 하지만 이것은 은닉 층의 수가 증가할수록 weight의 수도 증가해 학습이 어렵다는 단점이 있다. Rumelhart등은 오류역전파알고리즘(Error Backpropagation Algorithm)을 이용해 이를 해결하였다[10].

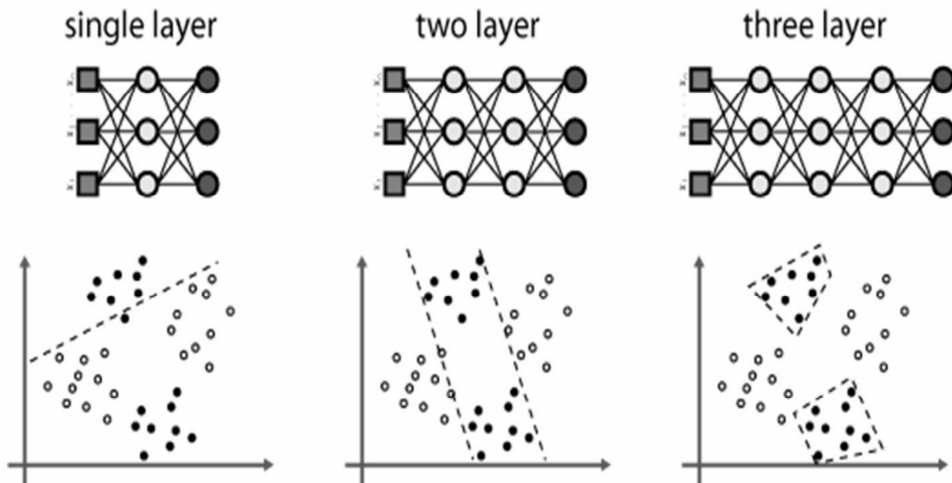


그림 2 다층 퍼셉트론 구조

만약 입력 층의 뉴런을 I , 은닉 층의 뉴런을 H 라 하고, k 번째 MLP의 출력 값을 표현하면 다음과 같다.

$$f_k(\mathbf{x}, \mathbf{w}) = \sigma_k \left(\sum_{h=1}^H w_{kh}^{(2)} \sigma_h \left(\sum_{i=1}^I w_{hi}^{(1)} x_i \right) \right) \quad (3)$$

위의 식에서 \mathbf{x} 는 학습 데이터가 되고 $\mathbf{w} = (\mathbf{w}^{(1)}, \mathbf{w}^{(2)})$ 는 weight 벡터이다. 신경망에서 가장 흥미로운 특징 중 하나는 모델의 학습 능력이다. 신경망은 프로그래밍 없이 학습 데이터로부터 문제를 해결하기 위해 자동적으로 학습이 가능하다. 예를 들어,

$$D_N = \{(\mathbf{x}_n, \mathbf{y}_n) \mid n = 1, 2, \dots, N\} \quad (4)$$

위와 같이 \mathbf{x}_n 은 n번째 입력 패턴이고, \mathbf{y}_n 은 목표로 하는 출력 패턴이다. 오류역전파알고리즘의 에러를 최소화 시키는데 사용되는 것이 Gradient descent 방법이다. 이는 미분계수에 비례해 weight 벡터 \mathbf{w} 를 수정하는 것이 특징인데 그 방법은 다음과 같다. 입력 데이터 \mathbf{x}_n 을 이용해 신경망의 출력 벡터 $\mathbf{f}(\mathbf{x}_n, \mathbf{w})$ 를 계산한다. 그리고 나서 실제 출력 값 $\mathbf{f}(\mathbf{x}_n, \mathbf{w})$ 과 예측 값 \mathbf{y}_n 사이의 에러 $E_n(\mathbf{w})$ 를 구한다. 식은 아래와 같다.

$$E_n(\mathbf{w}) = \frac{1}{2} \|\mathbf{y}_n - \mathbf{f}(\mathbf{x}_n, \mathbf{w})\|^2 \quad (5)$$

그 후, weight값 w_{ji} 를 learning rate η 를 이용해 에러를 최소화 시킬 수 있는 방향으로 바꾸어 준다. 식은 아래와 같다.

$$w_{ji} \leftarrow w_{ji} - \eta \frac{\partial E_n}{\partial w_{ji}} \quad (6)$$

이러한 신경망의 학습 능력 때문에 패턴인식(Pattern recognition), 컴퓨터 비전(Computer vision), 자연어처리(Natural language processing; NLP), 그리고 로봇틱스(Robotics) 등 인공지능의 다양한 분야에서 사용된다. 그리고 신경망은 인지과학(Cognitive science)에서 인간이 어떻게 정보를 표현하고, 처리하고, 학습하는지를 조사하는 도구로서도 사용이 가능하다.

전통적으로, 2-layer MLP가 가장 많이 사용되어왔다. 하지만 최근 연구에서 2개 보다 더 많은 층을 갖는 깊은 신경망(Deep networks)이 기존의 얇은 신경망(Shallow network)보다 뛰어나다는 것을 보였다. 이론적으로 다층 퍼셉트론은 하나의 은닉층만으로도 복잡한 비선형 함수(Nonlinear function)를 학습할 수 있다는 것이 증명되었지만 이미지나 음성과 같은 복잡한 데이터를 학습하는데 깊은 신경망이 더 유용하다는 것이 증명되었다. 깊은 신경망에는 두 종류가 있다. 하나는 다층의 Restricted Boltzmann machines(RBMs)을 이용한 Deep belief network(DBN)이다[11-14]. RBM은 뉴런들이 모두 연결된 2-layer로 이루어져 있다. DBN은 이 RBM을 Gibbs sampling에 기반한 Contrastive divergence (CD) 방법으로 학습을 시킨다[9].

Deep networks의 다른 하나는 Convolutional neural network 이다. 컨벌루션 신경망은 DBN과는 반대로 부분적으로 연결된 뉴런 층을 갖는다. 기존의

신경망은 입력된 이미지의 특성을 고려하지 않아 이미지의 작은 변화에 대응하지 못하였다. 반면 컨벌루션 신경망은 다양한 계층을 통해 데이터의 특징들을 추출하고 이미지를 단순화 시키기 때문에 이미지가 변화해도 인식이 가능하다. 컨벌루션 신경망은 뇌의 시각피질(Visual cortex)이 시각정보를 인지하는 방법을 본 따 만든 모델로 그림 3과 같은 구조를 가진다.

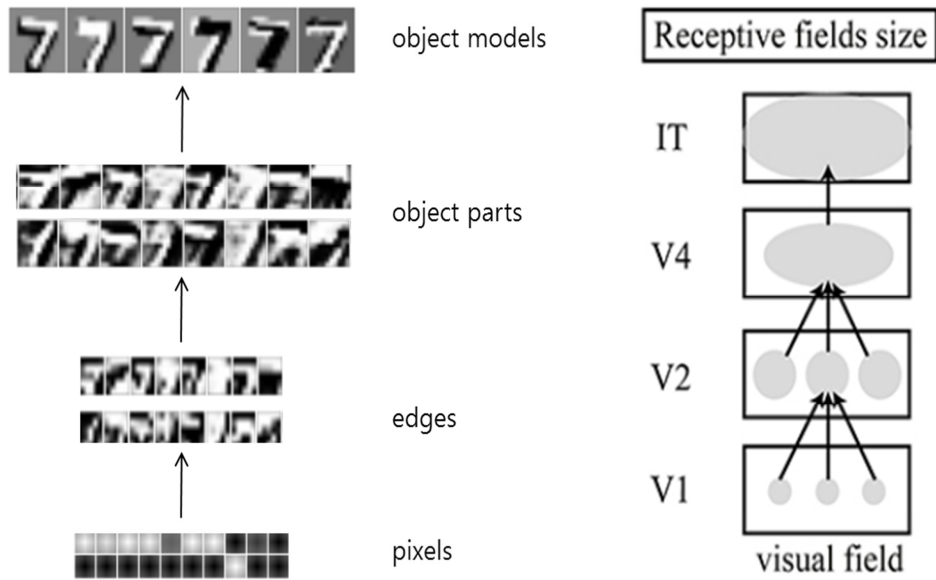


그림 3 CNN과 시각피질의 구조

컨벌루션 신경망은 여러 개의 컨벌루션 계층과 서브 샘플링 계층으로 이루어져 있다. 컨벌루션 계층은 많은 수의 feature map들을 가지고 있는데, 각 feature map은 입력된 이미지의 부분적인 수용장(Receptive field)을 갖는다. 즉, 컨벌루션 신경망에서 입력된 이미지의 몇몇 local patch들이 서브 샘플링 계층으로 연결된다. 서브 샘플링 계층도 여러 개의 feature map들로 구성되어 있는데, 이

feature map은 이전 계층에서 전달된 feature map의 최대 값으로 구해진다. 서브 샘플링 계층의 각 유닛들은 학습 가능한 계수(Coefficient)와 학습 가능한 bias를 갖고, 컨벌루션 신경망의 학습을 위해 오류역전파 알고리즘이 사용된다.[15-19]

2.2 컨벌루션 계층

2.2.1 커널

이 계층에서는 커널이라 불리는 필터를 이용해 이전 계층에서 전달된 이미지로부터 여러 개의 feature map을 만든다. 이렇게 만들어진 feature map은 기존의 데이터에는 없던 특성들이 생겨 모델을 더욱 강력하게 만들어 준다. 커널은 그림 4와 같은 구조를 가진다. 커널을 이용하면 한정된 데이터에서 다양한 특징들을 추출할 수 있는데 이는 한정된 데이터를 복원 추출을 통해 다양하게 만드는 앙상블 기법에서 사용하는 리샘플링과 비슷하다. 기존 이미지 데이터의 회전된 이미지 데이터를 얻기 위해 컨벌루션 매트릭스를 이용해 다양한 형태의 이미지 데이터를 만드는 연구가 시도된 바 있는데[20], 커널을 이용하면 이미지가 회전되거나 변화가 생겨도 인식이 가능하게 해준다. 본 논문에서는 커널과 앙상블 기법의 유사점에 대해 알아보았는데, 이러한 내용은 3장에 기술되어 있다.

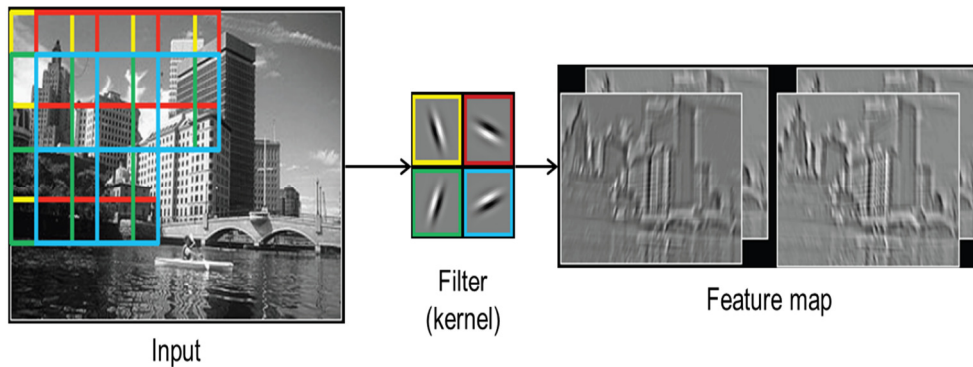


그림 4 Kernel 구조

2.2.2 가중치 공유

그리고 이렇게 만들어진 feature map의 각 픽셀 유닛들은 가중치 벡터를 공유하게 된다. 가중치 공유는 그림 5와 같은 구조를 가지는데, 같은 색으로 연결된 유닛들이 공유되는 가중치이다. 모든 유닛들이 연결되는 일반적인 인공 신경망에 반해 컨벌루션 신경망은 가중치를 공유 함으로써 더 적은 파라미터 수를 갖게 되어 학습 시간을 단축시키고, 학습 데이터를 과도하게 학습하는 것을 방지해 과적합(Overfitting)현상을 방지한다.

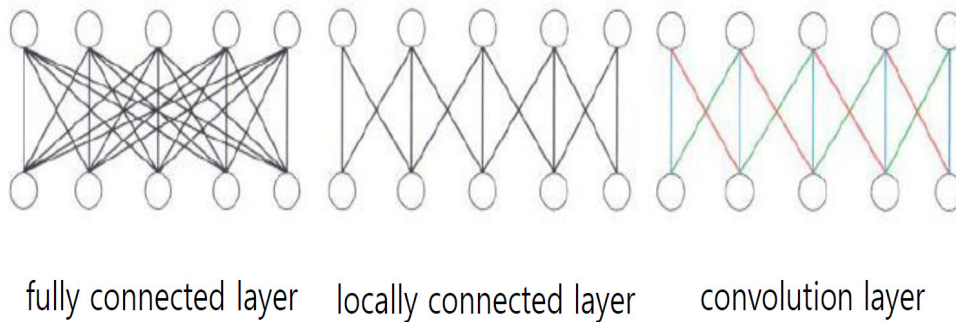


그림 5 Weight sharing 구조

2.3 서브 샘플링 계층

컨벌루션 계층 후에 서브 샘플링 계층이 오게 된다. 서브 샘플링 계층의 주요 기능은 통합(pooling)을 통해 이미지를 압축시키고 데이터의 차원을 축소해 이미지를 단순화 시켜 복잡한 이미지도 인식할 수 있게 해준다. 통합에는 여러 가지 방법이 있는데 벡터의 평균값을 취하는 방법과 최대값을 취하는 방법, 그리고 선형결합을 이용하는 방법이 있다. 일반적으로 최대값을 이용한 결과가 가장 좋다는 연구결과[21]가 있으며, 본 논문에서도 최대값을 이용해 실험하였다. 서브 샘플링 계층의 구조는 그림 7과 같다.

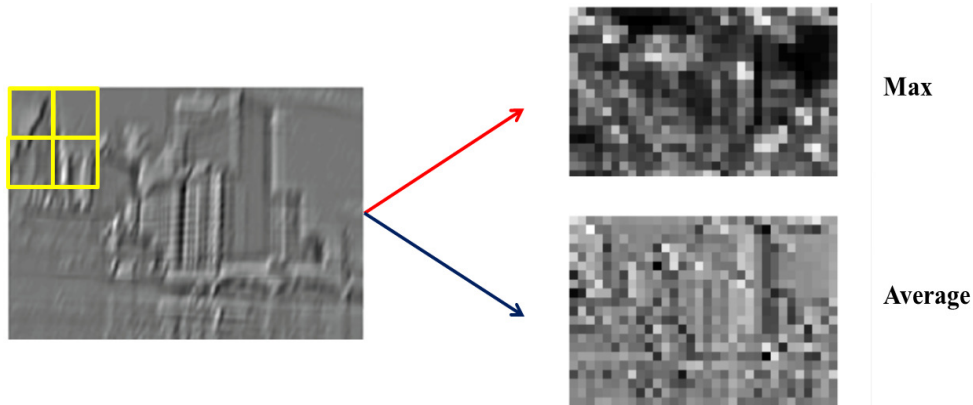


그림 6 Pooling 구조

3. 앙상블 기법

앙상블 기법은 학습 모델의 집합을 구성하고 그것들의 예측 결과를 투표를 통해 결정함으로써 알고리즘의 성능을 높이는 기법이다[22]. 전통적인 앙상블 기법으로는 베이지안 평균을 이용하는 방법이 있지만, 최근 연구로는 배깅과 부스팅이 대표적이다.

배깅은 Breiman이 처음 제안한 방법으로 여러 개의 약분류기의 예측 결과를 결합함으로써 단일 분류기보다 신뢰도가 높은 예측 값을 얻는 앙상블 방법 중 하나이다[23][24]. 전체 데이터 T에서 복원추출을 통하여 n개의 데이터 집합 T1, T2, ..., Tn을 생성한 후 이를 학습 데이터로 사용하여 약분류기h1, h2, ..., hn을 만든다. 이렇게 만들어진 약분류기들의 예측 값을 취합한 후 최종 결과를 도출하게 된다. 일반적으로 다수결(majority voting)이나 가중치 투표(weighted voting)에 의해 예측 값을 결합한다. 배깅은 구현이 쉽고 학습을 전체 데이터의 일부만을 이용하기 때문에 학습 속도가 비교적 빠르다는 장점이 있다. 이러한 특징은 모델의 학습속도가 데이터의 크기에 선형적으로 증가할 경우에는 각각의 약분류기를 병렬적으로 학습하지 않을 경우에는 큰 도움이 되지 못하지만, 필기체 인식의 경우와 같이 데이터의 크기 증가에 따라서 학습속도가 지수적으로 증가하는 경우에 있어서는 전체 데이터를 사용한 단일 분류기에 비하여 월등히 빠른 학습 속도를 보여준다[25]. 기본적인 알고리즘은 그림 7과 같다.


```

Bagging ( $T, L_b, M$ )
  For each  $m = 1, 2, \dots, M$ 
     $T_m = \text{Sample\_With\_Replacement}(T, N)$ 
     $h_m = L_b(T_m)$ 
  Return  $h_{\text{fin}}(x) = \text{argmax}_{y \in Y} \sum_m^M I(h_m(x) = y)$ 

Sample\_With\_Replacement( $T, N$ )
   $S = \varnothing$ 
  For  $i = 1, 2, \dots, N$ 
     $r = \text{random\_integer}(1, N)$ 
    Add  $T[r]$  to  $S$ 
  Return  $S$ 

 $T$  : original training set of  $N$  examples
 $M$  : # of base models to be learned
 $L_b$  : base model learning algorithm
 $I(A)$  : indicator function that returns 1 if event  $A$  is true
and 0 otherwise

```

그림 7 배깅 알고리즘

부스팅 역시 예측력이 약한 약분류기들을 결합하여 강한 예측모형을 만드는 것이다. 부스팅의 대표적인 알고리즘은 Freund와 Schapire가 개발한 AdaBoost[26]가 있다. 배깅과의 차이점은 처음 관측값의 가중치는 동일한 상태에서 시작하며 잘못 예측된 관측값은 다음번에 높은 가중치를 주고, 제대로 예측된 관측값은 반대로 가중치를 낮게 부여한다. 관측값들의 가중치가 재조정된 새로운 학습 데이터를 생성하여 오분류율이 일정 수준에 도달도달할 때 위의 과정을 반복하게 된다. 기본적인 알고리즘은 그림 8과 같다.

Boosting (T, M)

Initialize weights $D(i) = 1 / m$

For $t = 1, \dots, T$:

$$\text{Find } h_t = \arg \min_{h_j \in H} \varepsilon_j = \sum D_t(i) [y_i \neq h_j(x_i)]$$

If $\varepsilon_t \geq 1/2$ then stop

$$\text{Set } \alpha_t = \frac{1}{2} \log\left(\frac{1 - \varepsilon_t}{\varepsilon_t}\right)$$

Update

$$D_{t+1}(i) = \frac{D_t(i) \exp(-\alpha_t y_i h_t(x_i))}{Z_t}$$

Output the final classifier :

$$H(x) = \text{sign}\left(\sum_{t=1}^T \alpha_t h_t(x)\right)$$

T : original training set of N examples

M : # of base models to be learned

Z : normalisation factor

그림 8 부스팅 알고리즘

양상블 기법은 여러 개의 약분류기들을 학습시키기 위해 한정된 데이터에 부트스트랩, 잭나이프 등의 리샘플링 기법을 이용한다. 이러한 리샘플링 기법을 학습 데이터에 이용할 경우 한정된 데이터의 수를 늘리고 다양성을 증가시킬 수 있다. 또한 리샘플링은 모집단의 분포를 추정하기 위해서도 사용되는데, 이미지 데이터에서는 컨벌루션 매트릭스를 리샘플링의 용도로 사용하기도 한다[27]. 컨벌루션 신경망에서는 학습 데이터의 다양한 특징들을 추출하기 위해 커널을

이용하는데, 앙상블 기법과 컨벌루션 신경망 모두 데이터의 다양성을 이용해 최종 성능을 높인다는 공통점을 발견할 수 있었다.

4. 실험 방법 및 결과

4.1 실험 데이터

본 연구를 위해 실험 데이터로 MNIST 데이터를 이용했다. MNIST 데이터는 필기체로 된 숫자 이미지들로 구성되어 있다[28]. 이 데이터는 National Institute of Standards and Technology(NIST)에서 배포했다. 숫자들은 고정된 크기(28x28 픽셀)로 정규화되어 있고 중앙에 집중해 있다. 실험을 위해 학습 데이터와 테스트 데이터로 0부터 9까지의 숫자들만 이용했다. MNIST 데이터의 예시는 아래 그림 9와 같다.



그림 9 MNIST 필기체 데이터 예시

4.2 실험 내용

신경망의 모델 복잡도에 영향을 주는 요소는 학습 횟수, 파라미터의 수, 계층의 수 등이 있다. 컨벌루션 신경망의 모델 복잡도에 따른 성능(Training and test error)을 알아보기 위해 총 4가지 실험을 진행하였다. 우선 컨벌루션 신경망이 4-layer일 때(그림 10)와 6-layer(그림 11)일때의 성능을 비교해 보았다. 두 번째로 학습 횟수에 따른 성능을 비교해 보았다. 세 번째로 모델 복잡도를 가장 잘

반영하는 파라미터의 변화에 따른 성능을 알아 보았다. 마지막으로 파라미터의 변화에 따른 성능을 컨벌루션 신경망이 아닌 인공 신경망에 대해 진행해 보았다.

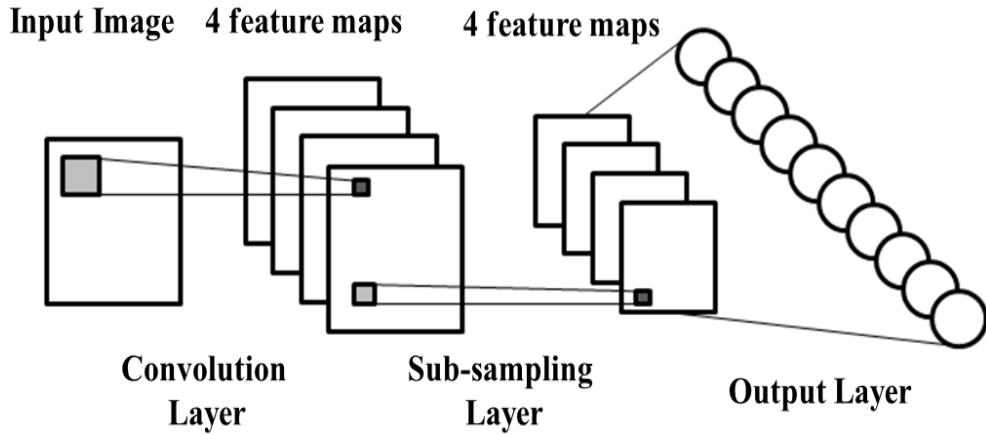


그림 10 4-layer CNN 구조

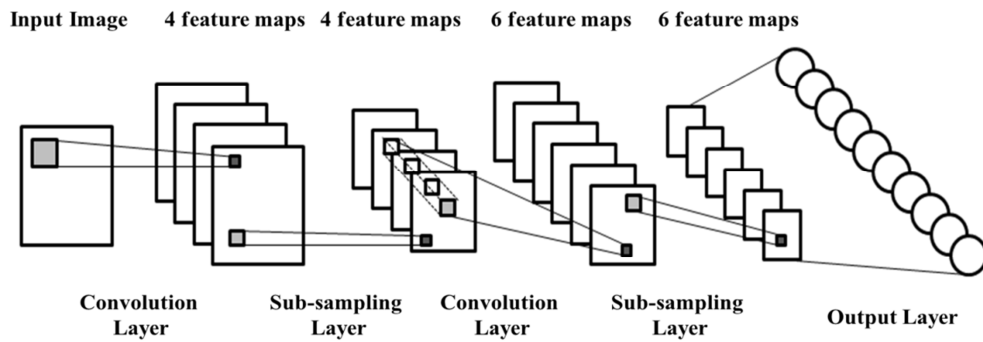


그림 11 6-layer CNN 구조

그림 12는 계층이 4개일 때와 6개일 때 컨벌루션 신경망의 성능을 보여준다. 학습 에러는 평균 제곱 편차(Mean Squared Error; MSE)로 측정되었고 테스트 에러는 오답변의 비율로 측정했다. 학습 데이터는 총 6,000개로 이루어져 있으며 테스트 데이터는 총 1,000개로 이루어져 있다. 이 실험에서는 유닛의 수를 8개로 학습 횟수는 1,000번으로 제한하였다. 이 실험의 결과 4-layer의 실험결과가 더

안 좋게 나왔는데, 이는 충분한 과정을 거치지 못해 학습 데이터의 추상적인 부분들을 학습하지 못하고 단편적인 부분만 학습했기 때문으로 보인다.



그림 12 CNN 계층 수에 따른 성능

그림 13는 컨벌루션 신경망의 학습 횟수에 따른 성능 변화를 보여준다. 학습 횟수가 많으면 학습횟수가 적을 때 보다 모델을 복잡하게 만든다. 따라서 학습 횟수를 증가시키는 것은 모델 복잡도를 증가시키는 것과 같은 효과가 있다[29]. 그림 13에서 학습 횟수가 증가하면 모델의 성능도 좋아지는 것을 볼 수 있다.

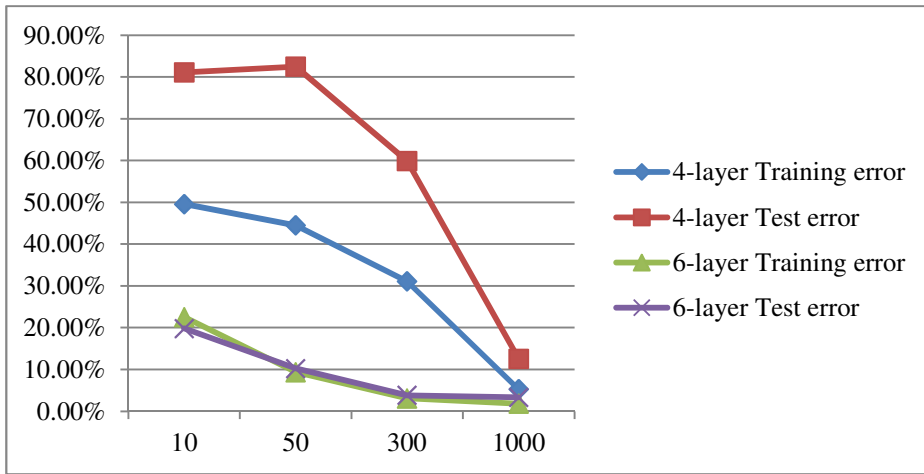


그림 13 학습 횟수에 따른 성능 비교

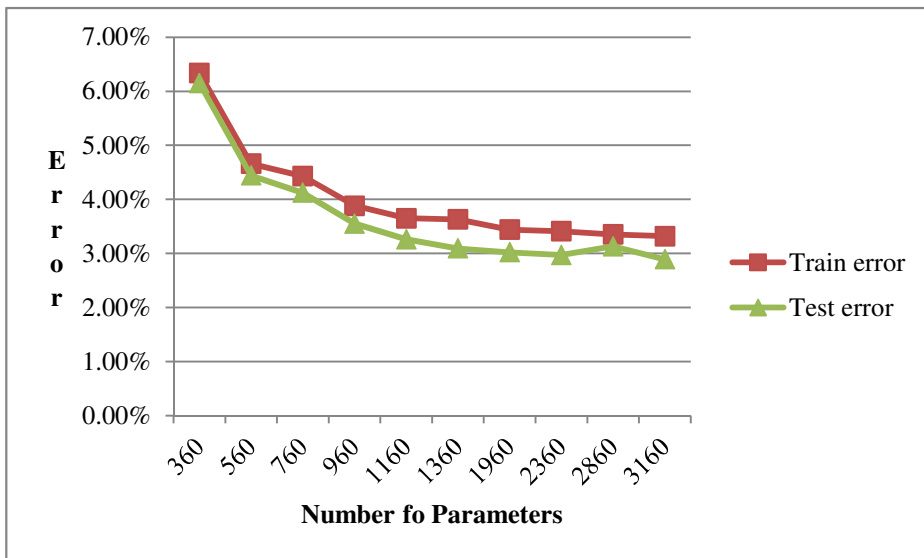


그림 14 CNN 파라미터 수에 따른 성능



그림 15 NN 파라미터 수에 따른 성능

그림 14는 컨벌루션 신경망이 파라미터의 변화(360~3160)에 따라 성능이 어떻게 변하는지를 보여준다. 이 실험에서 파라미터는 각 계층의 뉴런들만을 이용했다. 실험 결과 파라미터의 수가 증가할수록 성능도 대체적으로 좋아짐을 알 수 있었다. 또한 이 실험결과를 통해 컨벌루션 신경망이 굉장히 견고한 모델임을 알 수 있었다. 반면에, 인공 신경망의 파라미터 변화에 따른 실험에서는 과적합 현상을 발견할 수 있었다(그림 15). 이 실험에서 입력 뉴런(784)과 출력 뉴런(10)은 고정 시키고 은닉 뉴런을 100부터 280까지 변화시켰다. 이 실험을 통해 파라미터의 수가 증가할수록 Training error는 계속해서 내려가지만 Test error는 파라미터의 수가 119100 부근에서 증가하는 것을 알 수 있다. 이것은 인공 신경망이 컨벌루션 신경망보다 과적합 현상에 더 취약하다는 것을 의미한다. 컨벌루션 신경망이 과적합 현상이 나타나지 않은

이유에는 여러 가지 요소들이 복합적으로 작용한다. 과적합 현상은 기본적으로 모델이 학습 데이터를 지나치게 학습한 경우에 나타나게 되는데, 컨벌루션 신경망은 가중치 공유를 통해 모델의 파라미터 수를 낮추고, 통합 과정을 통해 이미지 데이터를 단순화 시킴으로써 과적합 현상을 방지하게 된다. 하지만 과적합을 방지했다는 것은 그만큼 모델이 학습 데이터에 덜 적합한 모델이라는 의미일 수도 있다. 과적합을 피하기 위해 학습시간이 짧은 단순한 모델을 만들면 데이터의 특징들을 제대로 파악하지 못하는 과소적합(Underfitting) 문제가 발생하게 된다. 이러한 과적합과 과소적합 사이의 문제를 편향-분산 트레이드오프(Bias-variance tradeoff) 문제라고 하는데, 오차를 최소화시키기에는 어느 정도 한계가 있다. 다만, 데이터 집합을 K 개의 집합으로 구성하고 $K-1$ 까지를 학습 데이터로, 나머지 1개를 테스트 데이터로 이용하는 방법을 K 번 반복하는 교차검증(Cross-validation) 방법을 이용한다면 파라미터 최적화와 과적합 방지를 일정 수준 해결할 수 있다.

4.3 결과 분석

앞의 4가지 실험 결과 컨벌루션 신경망은 모델의 복잡도가 증가할수록 성능 또한 계속해서 좋아지는 모습을 보여주었고, 기존의 인공 신경망은 그렇지 않음을 보여주었다. 이러한 결과가 나온 이유에는 여러 가지가 있을 수 있다. 앞에서 언급한 컨벌루션 계층에서 가중치를 공유하고, 다양한 데이터 특징들을 얻는 것이 컨벌루션 신경망을 인공 신경망보다 견고하게 만든 원인 중 하나라고 볼 수 있다. 또한 서브 샘플링 계층의 통합 과정을 통해 데이터를 단순화 시켜 데이터의 추상적인 부분을 학습하는 것이 가능해지고, 이미지의 왜곡이나 변화 등에 영향을 덜 받게 된다. 이러한 여러 계층을 거치면서 컨벌루션 신경망은 이미지의 변화에 상관없이 인식이 가능하게 된다.

5. 결론 및 향후 연구

본 논문에서는 컨벌루션 신경망의 구조와 모델 복잡도에 따른 성능에 대해 알아보았다. 컨벌루션 신경망의 컨벌루션 계층에서 사용하는 커널은 학습 데이터의 다양성을 증가시키기 위한 방법으로 앙상블 기법과 연관시켜 볼 수 있었다. 그리고 이러한 앙상블 기법의 리샘플링 과정이 학습 데이터에 다양성을 주어 모델의 최종 성능에 긍정적인 영향을 줄 수 있다는 것을 보여주었다. 또한, 다양한 실험을 통해 컨벌루션 신경망이 일반적인 신경망에 비해 견고하다는 것을 보여주었다. 커널을 통해 서로 다른 분포를 가지는 학습 데이터가 만들어지고, 이미지 데이터의 변화에 무관해지기 때문에 분산이 큰 필기체 데이터에 대해서도 좋은 성능을 보여준다. 그리고 심층학습 방법에 대한 새로운 알고리즘을 개발한 것은 아니지만, 컨벌루션 신경망의 모델 복잡도와 관련된 실험이 아직 활발히 연구가 되지 않은 분야이기 때문에 그 분야를 위한 초석이 되고자 한 것에 그 의의가 있다.

커널은 분명 최종 모델의 성능에 긍정적인 영향을 끼치지만 커널로 만들어진 모든 데이터들이 학습 성능을 높이는데 영향을 주는지에 대해서는 고려하지 못하였다. 최근 이와 관련해 심층 학습에서 **drop-out**을 이용한 연구가 활발히 이루어지고 있다. **Drop-out**은 랜덤하게 학습에 사용되는 뉴런을 고르고 나머지 뉴런은 학습에 관여를 시키지 않는 것이다. 이렇게 되면 학습 시간은 조금 더 필요하지만 더 소수의 뉴런들로 데이터를 학습하기 때문에 좀 더 추상적인 특징들을 학습하게 되어 최종 성능에 긍정적인 영향을 준다. 이와 비슷한 방법으로 노드들의 조합을 무작위로 선택하고 이 중 가장 좋은 조합을 이용하는 하이퍼네트워크가 있는데, 컨벌루션 신경망과 하이퍼네트워크를 결합하면 보다 강력한 신경망 모델이 될 것으로 생각된다[30].

참고문헌

- [1] Cruz, R. M., Cavalcanti, G. D., Ren, T. I., An Ensemble Classifier For Offline Cursive Character Recognition Using Multiple Feature Extraction Techniques, *In Neural Networks (IJCNN), The 2010 International Joint Conference on*, p. 1-8, 2010.
- [2] España-Boquera, S., Gorbe-Moya, J., Zamora-Martínez, F., Castro-Bleda, M. J., Hybrid HMM/ann models for bimodal online and offline cursive word recognition, *In Recognizing Patterns in Signals, Speech, Images and Videos*, p. 14-21, 2010.
- [3] AlKhateeb, J. H., Ren, J., Jiang, J., Al-Muhtaseb, H., Offline handwritten Arabic cursive text recognition using Hidden Markov Models and re-ranking, *Pattern Recognition Letters*, p.1081-1088, 2011
- [4] Plamondon, R., Srihari, S. N., Online and off-line handwriting recognition: a comprehensive survey, *Pattern Analysis and Machine Intelligence*, IEEE Transactions on, p. 63-84, 2000.
- [5] Xu, L., Krzyzak, A., Suen, C. Y., Methods of combining multiple classifiers and their applications to handwriting recognition, *Systems, man and cybernetics*, IEEE transactions on, p. 418-435, 1992
- [6] Graves, A., Schmidhuber, J., Offline handwriting recognition with multidimensional recurrent neural networks, *Advances in Neural Information Processing Systems*, p. 545-552, 2009.
- [7] Ciresan, D., Meier, U., Schmidhuber, J., Multi-column deep neural networks for image classification, *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, p. 3642-3649, 2012.

- [8] Zanchettin, C., Bezerra, B. L. D., Azevedo, W. W., A KNN-SVM hybrid model for cursive handwriting recognition, *In Neural Networks (IJCNN), The 2012 International Joint Conference on*, p. 1-8, 2012.
- [9] Spiegelhalter, D. J., Best, N. G., Carlin, B. P., Van Der Linde, A., Bayesian measures of model complexity and fit, *Journal of the Royal Statistical Society*, p. 583-639, 2002.
- [10] Rumelhart, D. E., Hinton, G. E., and Williams, R. J., Learning internal representations by errorpropagation, , *DTIC Document*, 1985.
- [11] Hinton, G.E., Salakhutdinov, R.R, Reducing the dimensionality of data with neural networks, *Science*, 313, 504–507, 2006.
- [12] Bengio, Y., Learning deep architectures for AI, *Foundations and Trends in Machine Learning*, 2, 1–127, 2009.
- [13] Hinton, G.E., Osindero, S., Teh, Y.-W., A fast learning algorithm for deep belief nets, *Neural Computation*, 18, 1527–1554, 2006.
- [14] Deng, L., Yu, D., Platt, J., Scalable stacking and learning for building deep architectures, *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, p. 2133–2136, 2012.
- [15] LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L.D., Backpropagation applied to handwritten Zip code recognition, *Neural Computation*, p. 541–551, 1989.
- [16] Palm, R.B., Prediction as a candidate for learning deep hierarchical models of data, *Technical Report, Technical University of Denmark*, 2012.
- [17] Ciresan, D.C., Meier, U., Gambardella, L.M., Schmidhuber, J., Deep big multilayer perceptron for digit recognition, *Neural Networks: Tricks of the Trade*, p.581–598, 2012.

- [18] Raiko, T., Valpola, H., LeCun, Y., Deep learning made easier by linear transformations in perceptrons, *In 15th International Conference on Artificial Intelligence and Statistics (AISTATS)*, p. 924–932, 2012.
- [19] Bouchain, D., Character recognition using convolutional neural networks, *Technical report, Institute for Neural Information Processing*, 2006.
- [20] Unser, M., Thevenaz, P., & Yaroslavsky, L., Convolution-based interpolation for fast, high-quality rotation of images, *Image Processing, IEEE Transactions on*, p. 1371-1381, 1995.
- [21] Boureau, Y. L., Jean P, and Yann L.C., A theoretical analysis of feature pooling in visual recognition, *International Conference on Machine Learning (ICML-10)*, 2010.
- [22] Dietterich, T. G., Ensemble methods in machine learning, *In Multiple classifier systems*, p. 1-15, 2000.
- [23] Quinlan, J.R., Bagging, boosting, and C4. 5, *Proceedings of the National Conference on Artificial Intelligence*, p.725-730, 1996.
- [24] Buhlmann, P., Yu, B., Analyzing bagging, *The Annals of Statistics*, p. 927-961, 2002.
- [25] 김태준, 장하영, 박정완, 황성택, & 장병택., 온라인 필기인식을 위한 증가하는 데이터를 이용한 앙상블 기법, *정보과학회논문지: 소프트웨어 및 응용*, p. 164-170, 2014
- [26] Freund, Y., Schapire, R., & Abe, N., A short introduction to boosting, *Journal-Japanese Society For Artificial Intelligence*, p.771-780, 1999.
- [27] Keys, R., Cubic convolution interpolation for digital image processing, *Acoustics, Speech and Signal Processing, IEEE Transactions on*, p. 1153-1160, 1981.
- [28] León, G.M., Moreno B. A., Magallanes Q. R., Valdez, R.D., Assessment in subsets of MNIST handwritten digits and their effect in the recognition rate, *Journal of Pattern*

Recognition Research, p. 244–252 , 2011.

[29] Kim, T. J., Zhang, D. S., Kim, J. S., Effects of Model Complexity on Generalization Performance of Convolutional Neural Networks, *ISIS2013*, 2013.

[30] Zhang, B.T., Hypernetworks: A molecular evolutionary architecture for cognitive learning and memory, *Computational Intelligence Magazine*, p. 49-69, 2008.