J. Internet Comput. Serv.
ISSN 1598-0170 (Print) / ISSN 2287-1136 (Online)
http://www.jics.or.kr

오토인코더 기반의 잡음에 강인한 계층적 이미지 분류 시스템☆

A Noise-Tolerant Hierarchical Image Classification System based on Autoencoder Models

이 종 관[1*]
Jong-kwan Lee

요 약

본 논문은 다수의 오토인코더 모델들을 이용한 잡음에 강인한 이미지 분류 시스템을 제안한다. 딥러닝 기술의 발달로 이미지 분류의 정확도는 점점 높아지고 있다. 하지만 입력 이미지가 잡음에 의해서 오염된 경우에는 이미지 분류 성능이 급격히 저하된다. 이미지에 첨가되는 잡음은 이미지의 생성 및 전송 과정에서 필연적으로 발생할 수밖에 없다. 따라서 실제 환경에서 이미지 분류기가 사용되기 위해서는 잡음에 대한 처리 및 대응이 반드시 필요하다. 한편 오토인코더는 입력값과 출력값이 유사하도록 학습되어지는 인공신경망 모델이다. 입력데이터가 학습데이터와 유사하다면 오토인코더의 출력데이터와 입력데이터 사이의 오차는 작을 것이다. 하지만 입력 데이터가 학습데이터와 유사성이 없다면 오토인코더의 출력데이터와 입력데이터 사이의 오차는 클 것이다. 제안하는 시스템은 오토인코더의 입력데이터와 출력데이터 사이의 관계를 이용한다. 제안하는 시스템의 이미지 분류 절차는 2단계로 구성된다. 1단계에서 분류 가능성이 가장 높은 클래스 2개를 선정하고 이들 클래스의 분류 가능성이 서로 유사하면 2단계에서 추가적인 분류 절차를 거친다. 제안하는 시스템의 성능 분석을 위해 가우시안 잡음으로 오염된 MNIST 데이터셋을 대상으로 분류 정확도를 실험하였다. 실험 결과 잡음 환경에서 제안하는 시스템이 CNN(Convolutional Neural Network) 기반의 분류 기법에 비해 높은 정확도를 나타냄을 확인하였다.

☞ 주제어 : 이미지 분류, 딥러닝, 머신러닝, 오토인코더, 잡음

ABSTRACT

This paper proposes a noise-tolerant image classification system using multiple autoencoders. The development of deep learning technology has dramatically improved the performance of image classifiers. However, if the images are contaminated by noise, the performance degrades rapidly. Noise added to the image is inevitably generated in the process of obtaining and transmitting the image. Therefore, in order to use the classifier in a real environment, we have to deal with the noise. On the other hand, the autoencoder is an artificial neural network model that is trained to have similar input and output values. If the input data is similar to the training data, the error between the input data and output data of the autoencoder will be small. However, if the input data is not similar to the training data, the error will be large. The proposed system uses the relationship between the input data and the output data of the autoencoder, and it has two phases to classify the images. In the first phase, the classes with the highest likelihood of classification are selected and subject to the procedure again in the second phase. For the performance analysis of the proposed system, classification accuracy was tested on a Gaussian noise-contaminated MNIST dataset. As a result of the experiment, it was confirmed that the proposed system in the noisy environment has higher accuracy than the CNN-based classification technique.

☞ keyword : Image classification, Deep learning, Machine learning, Autoencoder, Noise

1. 서 론

최근 딥러닝을 이용한 이미지 처리 기술이 급격히 발달하여 영상인식, 물체인식, 얼굴인식 등의 분야에서 활발히 연구가 이루어지고 있다. 영상 처리를 위한 대표적인 딥러닝 기술인 CNN(Convolutional Neural Network)은 인공신경망의 한 부류로 기존의 이미지 처리 알고리즘에 비해 상대적으로 데이터 전처리 과정이 적다는 큰 장점

1 Department of Computer Science, Korea Military Academy, Seoul, 01805, Korea.
* Corresponding author (jklee64@kma.ac.kr)
[Received 25 June 2020, Reviewed 7 August 2020(R2 18 Septebmer 2020, R3 4 November 2020), Accepted 2 December 2020]
☆ 본 논문은 20년 한국정보통신학회 춘계학술대회 발표논문의 후속 연구 결과임.
☆ 본 논문은 화랑대연구소와 한국연구재단의 지원으로 수행되었음. (No. 2019R1G1A100303012)

이 있다. 다시 말해, 데이터의 특징(feature)을 자동적으로 추출한다. 기존 알고리즘들이 데이터의 특징을 사용자가 직접 추출해야 하고, 추출된 특징에 따라 이미지 분류 성능이 크게 좌우된다는 한계를 CNN은 극복한 것이다[1, 2]. 일반적으로 CNN은 컨볼루션(convolution) 계층과 폴링(pooling) 계층을 반복적으로 수행하여 데이터의 계층적 특징들을 추출하고, 완전 연결 계층(fully connected layer)을 통해 이미지를 분류한다.

CNN 기반의 딥러닝을 활용한 기술의 발전으로 이미지 분류의 오류율이 크게 감소되었다. CNN기반의 이미지 분류 기술인 AlexNet[3]이 2012년에 ImageNet에 대한 분류 오류율을 15.4%까지 낮추어 CNN의 우수성을 입증하였다. 이후 VGGNet[4], GoogLeNet[5], ResNet[6], DenseNet[7], MobileNet[8], SENet[9], NASNet[10] 등 CNN 기반의 수많은 딥러닝 모델들이 잇달아 발표되었으며, 현재는 이미지 분류 문제에 있어서 딥러닝 모델이 인간의 이미지 분류 능력에 버금가는 성능을 나타내고 있다.

이미지 분류기의 성능은 데이터에 포함된 각종 노이즈에 의해서 저하될 수 있다. 또한 입력 데이터에 잡음이 없을 수는 없다. 이는 이미지를 획득하고 전송하는 과정에서 필연적으로 잡음이 첨가될 수밖에 없기 때문이다. 인간은 이미지에 포함된 노이즈를 인식하여 이를 무시하고 이미지를 분류한다. 그런데 분류기는 이미지에 포함된 잡음도 분류를 위한 계산과정에 포함시킨다. 따라서 분류 대상 이미지가 각종 노이즈에 의해서 오염되었다면 분류기의 성능은 저하될 수밖에 없다. 또한 잡음에 대한 모델의 성능 저하는 보안 취약점으로 악용될 수도 있다[11-13]. 따라서 분류기의 평가시 원본 이미지 뿐 아니라 잡음에 의해서 오염된 이미지에 대한 성능이 함께 고려되어야 한다.

잡음에 의한 분류기 성능 저하 문제를 해결하기 위한 방법은 크게 2가지로 나눠볼 수 있다. 첫 번째 방법은 분류기에 이미지를 입력하기 전에 잡음을 제거하는 것이다. 대표적인 방법은 중앙값(median) 필터, 가우시안(gaussian) 필터 등 적절한 필터를 사용하는 것인데, 잡음이 첨가되지 않은 픽셀에 대해서도 동일하게 적용되기 때문에 중요한 이미지 특징들도 함께 열화(blurring)시키는 단점이 있다[14]. 또한 잡음의 발생 시기, 크기, 형태는 많은 경우 예측하기 어렵다. 따라서 잡음을 확률 모델로 추정하여 완벽하게 제거하는 것은 현실적으로 쉬운 과업이 아니다.

Sudipta 등은 CNN 모델과 오토인코더 모델을 결합한 형태의 잡음에 강인한 이미지 분류 모델인 RSDAECNN을 제안하였다[15]. 낮은 수준의 잡음이 첨가된 이미지로 학습된 오토인코더와 원본 이미지로 학습된 오토인코더의 조합으로 3가지 분류기를 구성하고 각 분류기의 분류 결과를 종합하여 이미지를 분류하는 방식이다. 이를 통해 CNN 기반의 분류기에 비해 높은 분류 정확도를 달성하였다. 하지만 분류기의 성능은 학습시 사용한 잡음 데이터에 크게 좌우되는 문제가 있다.

잡음에 의한 분류기 성능 저하를 극복하기 위한 두 번째 방법은 잡음에 강인한 분류기 모델을 구성하는 것이다. 본 논문은 두 번째 방법에 해당하는 것으로 다수의 오토인코더 모델을 이용하여 잡음에 강인한 이미지 분류 시스템을 제안한다. 오토인코더 모델이 클래스 A의 이미지들로만 학습되었다면, 클래스 A에 속하는 이미지에 대한 모델 출력값은 입력값과 매우 유사할 것이다. 하지만 클래스 A에 속하지 않는 이미지에 대한 모델 출력값은 입력값과 크게 다를 것이다. 제안하는 시스템은 이러한 관계를 이용하여 잡음 이미지에 대한 분류 정확도를 향상시킨다. 또한 오토인코더 모델 학습시 원본 데이터만을 사용하기 때문에 RSDAECNN과 달리 학습시 사용되는 이미지의 잡음 정도에 따른 모델 성능의 의존성이 없다.

본 논문은 다음과 같이 구성된다. 2장에서 제안하는 시스템의 주요 구성요소인 오토인코더에 대해서 살펴보고, 3장에서는 제안하는 시스템의 이미지 분류 과정에 대해서 구체적으로 설명한다. 4장에서 제안하는 시스템의 성능을 실험을 통해 살펴보고 CNN 기반의 기법들과 비교한다. 마지막으로 5장에서 결론을 맺는다.

2. 오토인코더 소개

본 장에서는 제안하는 시스템의 주요 구성요소인 오토인코더에 대해서 간략히 살펴본다.

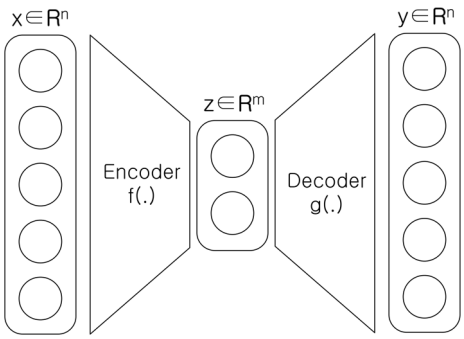
오토인코더는 모델의 입력값이 출력값과 유사해 지도록 학습하는 비지도학습(unsupervised learning) 형태의 인공신경망이다. 그림 1은 오토인코도의 기본 구조를 나타낸다.

기본적으로 오토인코더는 인코더(encoder) $f(\bullet)$와 디코더(decoder) $g(\bullet)$ 부분으로 구성된다. 모델의 입력값을 $x \in R^n$라 할 때 인코더의 출력과 디코더의 출력은 식(1), 식(2)와 같다.

$$z = f(x) \in R^m,\ n > m \tag{1}$$

$$y = g(f(x)) \in R^n \tag{2}$$

오토인코더는 식(3)과 같이 x와 y의 오차를 손실함수로 정의하고 손실함수가 최소가 되도록(즉, $x \approx y$) 모델의



(그림 1) 기본 오토인코더의 구조
(Figure 1) The structure of basic autoencoder

파라미터(가중치와 바이어스) θ를 경사하강법(gradient descent) 등을 이용하여 최적화한다.

$$\theta^{*} = \arg\min_{\theta} L(x, g(f(x))) \quad (3)$$

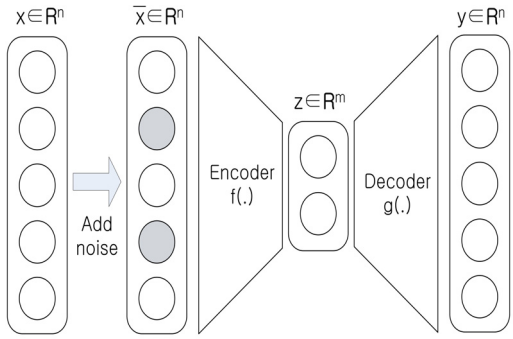
그림 1에서 보는 바와 같이 $n > m$ 이기 때문에 인코더는 입력 데이터의 특징을 함축적으로 표현할 수 있다. 이러한 특성은 데이터 압축, 차원 축소(dimensionality reduction), 사전학습(pre-training), 잡음제거(denoising) 등에 활용된다[16, 17].

오토인코더에서 데이터가 처리되는 과정을 살펴보면 입력데이터는 인코더에 의해서 데이터가 함축적으로 표현되고 디코더에 의해서 다시 복원된다. 즉, 데이터가 인코더에 의해서 표현될 때 불필요한 정보들은 생략된다. 다시 말해, 입력 데이터를 표현하는 핵심적인 요소들을 제외한 요소들은(예, 잡음) 제거되는 것이다. 이러한 특성을 이용하여 잡음제거 목적으로 오토인코더가 사용될 수 있다. 그림 2는 잡음제거 오토인코더의 구조를 나타낸다.

잡음제거 오토인코더는 원본 데이터(x)에 인위적으로 잡음을 첨가한 잡음 데이터($\bar{x}$)를 인코더의 입력으로 하고 출력 데이터(y)가 원본 데이터에 가깝도록 학습한다.

$$\theta^{*} = \arg\min_{\theta} L(x, g(f(\bar{x}))) \quad (4)$$

식 (4)에서 보는 바와 같이 잡음제거 오토인코더는 입력데이터에 잡음을 첨가하는 절차를 제외하고는 기본 오토인코더와 동일하다.



(그림 2) 기본 잡음제거 오토인코더의 구조
(Figure 2) The structure of basic denoising autoencoder

3. 제안하는 이미지 분류 시스템

본 장에서는 제안하는 이미지 분류 시스템에 대해서 구체적으로 살펴본다. 표 1은 제안하는 시스템을 설명하기 위한 기호들이다.

(표 1) 기호와 의미
(Table 1) Symbol and meaning

기 호	의 미
n	Number of class
m	Number of selected classes in the first classification phase
T_i	Training data for class i
x / $\hat{x}$	Input data / Ouput data
x_i	i^{th} pixel value of data x
$\bar{x}$	Noisy data
$\check{x}$	Data with inverted contrast
AE_i	1-AE trained with data of class i
AE_{ij}	2-AE trained with data of class i and j
L_i	Loss value of AE_i
L_{ij}	Loss value of AE_{ij}
γ	Noise factor
$PC(x)$	Predicted class for $\bar{x}$

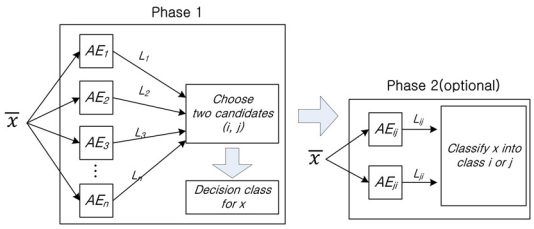
3.1 시스템 구성

제안하는 시스템은 2단계로 구성된다. 첫 번째 단계에서는 입력 이미지를 n개의 오토인코더에 각각 입력하여 출력 손실값을 계산한다. 손실값이 가장 작은 모델에 해

당하는 클래스로 입력 이미지를 분류한다. 만약 가장 작은 두 개의 손실값의 차이가 크지 않으면 2단계로 이동한다. 다시 말해, 분류의 확신도가 높지 않은 경우, 2개의 후보에 대해 추가적인 분류 절차를 거치는 것이다.

2단계에서는 1단계에서 선택된 클래스들의 데이터로 학습된 오토인코더를 이용하여 입력 이미지의 클래스를 최종적으로 판단한다. 1단계에서와 마찬가지로 입력 이미지를 오토인코더에 각각 입력하여 출력 손실값을 계산하고, 손실값이 가장 작은 모델에 해당되는 클래스로 입력 이미지를 최종적으로 분류한다.

그림 3은 제안하는 시스템의 구성을 개념적으로 나타낸 것이다.



(그림 3) 제안하는 시스템의 개념도
(Figure 3) The conceptual diagram of the proposed system

3.2 제안하는 시스템에서 오토인코더의 종류

제안하는 시스템에 사용되는 오토인코더의 종류는 두 가지이며, 각각을 1차 오토인코더(1-AE), 2차 오토인코더(2-AE)로 명명한다. 1-AE는 제안하는 시스템의 1단계에서 사용되며, 2-AE는 2단계에서 사용된다. 한편, 1-AE와 2-AE의 구조는 동일하지만 학습에 사용되는 데이터가 서로 다르다.

1-AE는 동일한 클래스의 데이터들로 학습된다. 즉, AE_i는 T_i 데이터로 학습된다. 반면, 2-AE는 2개 클래스의 데이터들이 학습에 사용된다. AE_{ij}는 T_i와 T_j 데이터로 학습된다. 첫 번째 클래스의 데이터(즉, T_i)에 대한 출력은 입력 데이터와 동일하지만, 두 번째 클래스의 데이터(T_j)에 대한 출력은 식(5)와 같다. $\breve{x}$는 x를 반전시킨 데이터를 의미하며, $max(x)$는 입력되는 이미지 픽셀의 최대값을 나타낸다. 즉, 출력데이터는 입력데이터의 반전된 형태이다. 이는 각 픽셀값의 차이를 인위적으로 크게 하여 두 번째 클래스의 데이터에 대한 출력손실이 첫 번째 클래스의 데이터에 대한 출력손실 보다 크도록 하기 위함이다.

$$\breve{x}_i = -x_i + \max(x) \tag{5}$$

그림 4는 숫자 6 이미지를 AE_{60}과 AE_{06}에 입력했을 때의 출력 이미지를 나타낸다. AE_{60}은 T_6과 반전된 T_0로 학습되었기 때문에 숫자 6의 이미지를 입력했을 때 입력 이미지와 유사하게 출력된다. 하지만, AE_{06}은 T_0과 반전된 T_6로 학습되었기 때문에 숫자 6의 이미지를 입력했을 때 반전된 이미지가 출력된다. 이때 L_{60}은 L_{06}에 비해 매우 작은 값을 갖게 된다. 따라서 손실값을 토대로 입력 이미지는 숫자 0이 아니라 숫자 6으로 쉽게 분류할 수 있다.

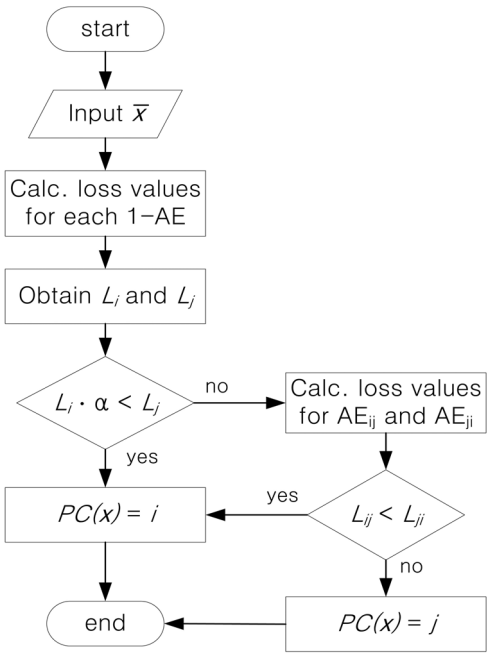
(a) Original img. (b) Ouput of AE_{60} (c) Output of AE_{06}

(그림 4) 2-AE 모델의 출력 이미지 예
(Figure 4) The example of outputs of 2-AEs

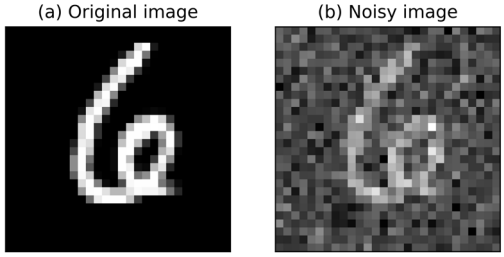
3.3 이미지 분류 절차

제안하는 시스템에서 잡음이 첨가된 이미지를 분류하는 세부 절차는 아래와 같고, 그림 5는 제안하는 시스템의 세부 이미지 분류 절차를 도식화한 것이다.

1) 모든 1-AE에 잡음 데이터 $\bar{x}$를 입력하여 각 모델별 손실값을 계산한다.
2) 손실값 n개 중 가장 작은 손실값 L_i와 L_j를 선택한다.
3) 만약 $L_i \times \alpha < L_j$ 이면 x를 클래스 i로 분류하고, 그렇지 않으면 분류 2단계로 이동한다. 여기서 α는 1이상의 실수값으로 2단계에 진입하는 정도를 조정하는 파라미터이다. 즉, α가 1에 가까울수록 2단계에 진입할 확률이 낮고, α가 1보다 클수록 2단계에 진입할 확률이 높다.
4) 2-AE인 AE_{ij}와 AE_{ji}에 $\bar{x}$를 입력하여 손실값 L_{ij}와 L_{ji}를 계산한다.
5) $L_{ij} < L_{ji}$인 경우 x를 클래스 i로 분류하고, 그렇지 않으면 x를 클래스 j로 분류한다.



(그림 5) 제안하는 시스템의 잡음 데이터 분류 절차
(Figure 5) Classification flowchart for noisy data in the proposed systems

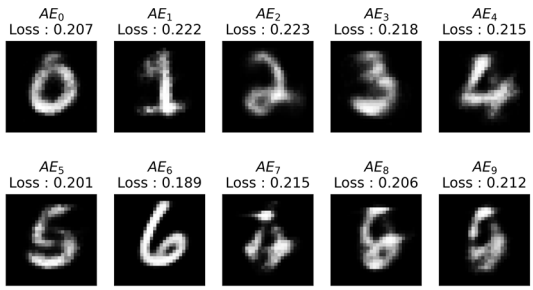


(그림 6) 원본 이미지와 잡음 이미지
(Figure 6) Original image and Noisy image

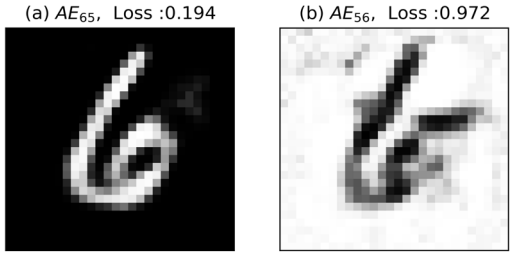
3.4 이미지 분류 예제

제안하는 시스템의 이미지 분류 예제를 간단히 살펴본다. 숫자 0부터 9까지의 이미지 중 그림 6과 같이 가우시안 잡음으로 오염된 숫자 6 이미지를 분류하는 과정이다.

그림 7은 제안한 시스템의 1단계 분류 절차에서 각 오토인코더의 출력이미지를 나타낸다. 각 오토인코더의 손실값을 비교해 보면 AE_6의 손실값이 0.189로 가장 작다. 즉, 해당 이미지를 클래스 6으로 분류할 수 있다. 그런데 AE_5와의 손실값 차이가 크지 않다. 따라서 분류 결과의 신뢰도를 높이기 위해 2단계 분류 과정을 필요시 추가적으로 수행한다. 2단계 분류 과정 수행 여부는 α 값에 의해서 좌우된다.



(그림 7) 제안하는 시스템의 1단계 오토인코더의 출력 이미지
(Figure 7) Output images of autoencoders in the first phase of the proposed system



(그림 8) 제안하는 시스템의 2단계 오토인코더의 출력
(Figure 8) Output images of autoencoders in the second phase of the proposed system

그림 8은 2단계 분류 절차에서 AE_{65}과 AE_{56}의 출력이미지를 나타낸다. 그림에서 알 수 있는 바와 같이 해당 이미지는 숫자 6이기 때문에 해당 이미지의 반전된 이미지로 학습된 AE_{56}의 출력이미지는 입력이미지의 반전된 형태로 나타난다. 이것은 손실값을 매우 크게 만든다. 따라서 두 오토인코더의 손실값의 차이가 1단계에 비해서 매우 커졌다. 결과적으로 AE_{65}의 손실값이 AE_{56}에 비해 매우 작으므로 해당 이미지를 클래스 6으로 분류한다.

4. 성능 분석

본 장에서는 제안하는 시스템의 성능을 분석하기 위해

가우시안 잡음으로 오염된 MNIST 데이터셋을 대상으로 분류 정확도를 측정하고, CNN 시스템, 오토인코더와 CNN이 결합된 시스템, 그리고 제안하는 시스템의 성능을 비교, 분석한다.

4.1 실험 환경

MNIST 데이터셋는 0부터 9까지의 필기체 숫자 이미지로 28×28 픽셀의 그레이스케일 이미지 6만장의 학습데이터와 1만장의 테스트데이터로 구성되어 있다.

잡음 이미지는 식 (6)과 같이 MINIST의 원본 데이터에 가우시안 잡음을 첨가하여 생성하였다. 식(6)에서 확률변수 ϵ_i는 서로 독립이며 평균이 0이고 표준편차가 1인 가우시안 확률분포를 동일하게 따른다. 가우시안 잡음은 원본 데이터에 $\gamma \bullet \epsilon$ 만큼이 더해진다. γ는 원본 데이터에 더해지는 잡음의 정도를 조절하는 파라미터로 0~1까지의 값을 갖는다. γ가 클수록 첨가되는 잡음의 정도가 심해진다.

$$\overline{x_i} = x_i + \gamma \bullet \epsilon_i \tag{6}$$

그림 9은 γ의 크기에 따른 이미지의 모양을 나타낸다. γ=0인 경우는 잡음이 첨가되지 않은 원본 이미지를 의미한다. γ가 0.5 이상인 경우 잡음으로 인해 육안으로 이미지에 포함된 숫자를 인식하기 어렵다.

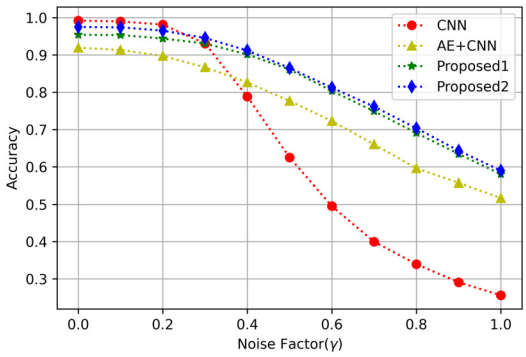
한편, 실험간 사용된 손실함수는 이미지의 픽셀 단위의 오차의 평균으로 식(7)과 같이 정의하였다.

$$L_i = \frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{x_i})^2 \tag{7}$$

4.2 실험 결과

그림 10은 가우시안 잡음으로 오염된 데이터에 대한 분류 정확도를 비교한 것으로 CNN을 적용한 결과, AE를 통해 잡음을 제거한 후 CNN을 적용한 결과(AE+CNN) 그리고 제안한 시스템의 1단계만을 적용한 결과(즉, α=1), 제안한 시스템의 1, 2단계를 모두 적용한 결과(즉, α를 매우 높은 값으로 설정)를 각각 나타낸다. AE+CNN 기법에서 오토인코더 모델은 모든 클래스의 데이터들(T_1, T_2, ... T_n)로 학습되었다.

그림 10에서 알 수 있는 바와 같이 4가지 방법 모두 γ



(그림 10) 정확도 비교: CNN, AE+CNN, 제안한 시스템
(Figure 10) Accuracy Comparison: CNN, AE+CNN and Prop. systems

가 증가함에 따라 분류 정확도는 감소한다. 그런데 CNN은 γ가 0.3 이상으로 증가하면 정확도가 급격히 저하된다. 반면 AE+CNN 기법과 제안하는 기법은 γ가 증가함에 따라 감소되는 정확도의 폭이 CNN에 비해 상대적으로 크지 않다.

CNN은 γ가 낮은 경우에는 가장 우수한 정확도를 나타낸다. 하지만 γ가 증가할수록 정확도가 크게 감소되어 γ가 0.3 이상이 되면 제안하는 기법에 비해 정확도가 크게 낮아진다.

AE+CNN 기법은 γ가 작은 경우 4가지 방법 중 가장 낮은 정확도를 보인다. γ가 증가하게 되면 CNN 방법에 비해 높은 정확도를 나타내지만 제안하는 기법에는 미치지 못한다. AE+CNN 기법은 AE 모델에 의해서 잡음이 제거된 후 CNN 모델에 의해서 이미지가 분류된다. 따라서 잡음이 증가할수록 잡음 제거 효과에 의해 분류 정확도가 CNN 모델에 비해서 높게 나타난다. 하지만 AE 모델에 의한 잡음 제거가 완벽할 수 없기 때문에 γ가 증가할수록 정확도는 점점 저하된다.

제안한 기법들은 γ가 작을 때 CNN에 비해 정확도가 다소 낮지만 γ가 커질수록 다른 기법에 비해 높은 정확도를 나타낸다. 또한 제안한 기법 중 1, 2단계를 모두 적용한 결과가 1단계만을 적용한 결과에 비해 다소 높은 정확도를 나타낸다. 즉, 1단계에서 분류에 실패했더라도 2단계에서 분류가 성공한다는 것을 의미한다. 특히, γ가 0.7이상이면 심한 잡음으로 인간의 눈으로도 잡음 이미지를 분류하기가 쉽지 않다. 하지만 제안하는 기법은 약 70% 이상의 정확도로 이미지 분류가 가능하다. 표 2는 실

험결과의 수치를 나타낸다. 결론적으로 잡음이 많은 환경에서는 제안하는 기법이 효과적으로 활용될 수 있다.

(표 2) 실험 결과
(Table 2) Simulation results

γ	Accuracy			
	CNN	AE+CNN	Prop. (α=1)	Prop. (α=100)
0.0	0.991	0.919	0.954	0.975
0.1	0.990	0.913	0.953	0.974
0.2	0.981	0.897	0.944	0.965
0.3	0.930	0.866	0.930	0.945
0.4	0.788	0.826	0.901	0.911
0.5	0.624	0.776	0.860	0.864
0.6	0.495	0.722	0.803	0.812
0.7	0.399	0.660	0.748	0.761
0.8	0.339	0.595	0.690	0.703
0.9	0.291	0.557	0.634	0.643
1.0	0.256	0.516	0.580	0.589

한편 제안하는 기법은 AE+CNN 기법과 달리 절대적인 잡음제거 효과로 정확도가 결정되지 않고, 각 AE 모델별 상대적인 손실값에 의해 정확도가 결정된다. 즉, 완벽한 잡음제거 효과가 아니라 다른 AE 모델에 비해 상대적으로 우수한 잡음제거 효과만이 요구된다. 이러한 이유 때문에 제안하는 기법이 AE+CNN 기법에 비해 잡음 환경에서 보다 우수한 성능을 나타내는 것으로 판단된다.

5. 결 론

본 논문은 다수의 오토인코더 모델의 조합을 통해 잡음에 강인한 이미지 분류 시스템을 제안하였다. 제안하는 시스템은 2단계로 구성된다. 1단계에서 오토인코더들의 입력데이터와 출력데이터 간의 오차들을 비교하여 분류할 클래스를 선택한다. 만약 분류 확률이 가장 높은 2개의 클래스에 대한 분류 확률이 서로 유사하면 2단계에서 추가적인 분류 절차를 수행하여 이미지 분류의 정확도를 높인다. 가우시안 잡음으로 오염된 MNIST 데이터를 대상으로 실험했을 때 제안하는 시스템이 CNN 기반의 시스템, 오토인코더와 CNN이 결합된 시스템에 비해 잡음 환경에서 보다 높은 분류 정확도를 나타냈다.

참고문헌(Reference)

[1] Wu, Jianxin. "Introduction to convolutional neural networks," National Key Lab for Novel Software Technology, Nanjing University, 2020. https://cs.nju.edu.cn/wujx/teaching/15_CNN.pdf

[2] Kuo, C-C. Jay. "Understanding convolutional neural networks with a mathematical model," Journal of Visual Communication and Image Representation, Vol. 41, pp. 406-413, 2016. https://doi.org/10.1016/j.jvcir.2016.11.003

[3] Krizhevsky, A., Sutskever, I., & Hinton, G. E. "Imagenet classification with deep convolutional neural networks," Commun. ACM, Vol. 60, No. 6, pp. 84-90, 2017. https://doi.org/10.1145/3065386

[4] Simonyan, K., & Zisserman, A., "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014. https://arxiv.org/abs/1409.1556

[5] C. Szegedy, et al. "Going deeper with convolutions," Proceedings of the IEEE conference on computer vision and pattern recognition, 2015. https://doi.org/10.1109/cvpr.2015.7298594

[6] K. He, et al. "Deep residual learning for image recognition," Proceedings of the IEEE conference on computer vision and pattern recognition, 2016. https://doi.org/10.1109/cvpr.2016.90

[7] G. Huang, et al. "Densely connected convolutional networks," Proceedings of the IEEE conference on computer vision and pattern recognition, 2017. https://doi.org/10.1109/cvpr.2017.243

[8] A. G. HOWARD, et al. "Mobilenets: Efficient convolutional neural networks for mobile vision applications," arXiv preprint arXiv:1704.04861, 2017. https://arxiv.org/abs/1704.04861

[9] J. Hu, Li Shen, and Gang Sun, "Squeeze-and-excitation networks," Proceedings of the IEEE conference on computer vision and pattern recognition, 2018. https://doi.org/10.1109/cvpr.2018.00745

[10] B. Zoph, "Learning transferable architectures for scalable image recognition," Proceedings of the IEEE

conference on computer vision and pattern recognition. 2018.
https://doi.org/10.1109/cvpr.2018.00907

[11] Sun, Lu, Mingtian Tan, and Zhe Zhou. "A survey of practical adversarial example attacks," Cybersecurity, 2018.
https://doi.org/10.1186/s42400-018-0012-9

[12] Cisse, Moustapha M., et al. "Houdini: Fooling deep structured visual and speech recognition models with adversarial examples," Advances in neural information processing systems, 2017.
https://papers.nips.cc/paper/7273-houdini-fooling-deep-structured-visual-and-speech-recognition-models-with-adversarial-examples.pdf

[13] H. Kwon, Y. Kim, H. Yoon and D. Choi, "Selective Audio Adversarial Example in Evasion Attack on Speech Recognition System," IEEE Transactions on Information Forensics and Security, vol. 15, pp. 526-538, 2020.
https://doi.org/10.1109/tifs.2019.2925452

[14] R. H. Chan, Chung-Wa Ho and M. Nikolova, "Salt-and-pepper noise removal by median-type noise detectors and detail-preserving regularization," IEEE Transactions on Image Processing, Vol. 14, No. 10, pp. 1479-1485, 2005.
https://doi.org/10.1109/tip.2005.852196

[15] Roy, S. S., Hossain, S. I., Akhand, M. A. H., & Murase, K. "A robust system for noisy image classification combining denoising autoencoder and convolutional neural network," International Journal of Advanced Computer Science and Applications, Vol. 9, No. 4, pp. 224-235, 2018.
https://doi.org/10.14569/ijacsa.2018.090131

[16] Charte, David, et al. "A practical tutorial on autoencoders for nonlinear feature fusion: Taxonomy, models, software and guidelines," Information Fusion Vol 44, pp. 78-96, 2018.
https://doi.org/10.1016/j.inffus.2017.12.007

[17] Goodfellow, Ian, Yoshua Bengio, and Aaron Courville, "Deep learning," MIT press, 2016.
https://www.deeplearningbook.org

◐ 저 자 소 개 ◑

이 종 관(Jong-kwan Lee)
2000년 육군사관학교 전자공학과(공학사)
2004년 한국과학기술원 전기 및 전자공학과(공학석사)
2014년 아주대학교 일반대학원 NCW공학과(공학박사)
2018년~현재 육군 사이버전 연구센터 연구실장
2017년~현재 육군사관학교 컴퓨터과학과 조교수
관심분야 : 머신러닝, 딥러닝, 전술 네트워크, 사이버전
E-mail : jklee64@kma.ac.kr