

<https://doi.org/10.7236/JIIBC.2024.24.1.181>

JIIBC 2024-1-26

회전된 객체 분류를 위한 CNN 기법들의 성능 비교 분석

Comparative Analysis of CNN Techniques designed for Rotated Object Classification

한희일*

Hee-Il Hahn*

요약 이미지 공간에서 무작위로 회전된 객체에 대한 분류 성능이 우수한 기법으로는 군 등변 CNN과 steerable 필터를 이용한 CNN 등이 있다. 본 논문에서는 이들의 수학적 구조를 설명하고 구현 방법을 소개한다. 기존의 CNN을 포함한 세 개의 모델에 대하여 동일한 필터 수를 갖도록 구현한 다음, 무작위로 회전된 *MNIST*를 이용하여 실험하고 이들의 성능을 비교분석한다. 실험 결과에 의하면 steerable CNN은 CNN보다 6.5% 이상의 인식을 향상을 보여준다. 특히, steerable CNN은 학습할 파라미터의 수가 상대적으로 적어서 훈련 데이터셋의 크기를 줄여도 성능 열화가 비교적 크지 않음을 실험 결과로 확인한다.

Abstract There are two kinds of well-known CNN methods, the group equivariant CNN and the CNN using steerable filters, which have excellent classification performances for randomly rotated objects in image space. This paper describes their mathematical structures and introduces implementation methods. We implement them, including the existing CNN, which have the same number of filters, then compare and analyze their performances by simulating them with the randomly rotated *MNIST*. According to the experimental results, the steerable CNN, which shows a classification improvement over the others, has a relatively small number of parameters to learn, so performance degradation is relatively small even when the size of the training dataset is reduced.

Key Words : Deep Learning, Filter Design, Steerable Filter, Group Equivariant Convolution, Group Equivariant CNN

1. 서 론

딥러닝의 대표적 모델인 CNN은 이미지, 오디오 신호 등에 널리 적용되는 강력한 모델로서 가중치 공유를 통하여 비교적 적은 양의 매개 변수를 이용한다. 이러한 이유로 FCN(fully-connected neural network)에 비하

여 계산량을 크게 줄이면서도 우수한 성능을 보여주는 특징이 있다[1,2,3]. 이미지나 오디오 같은 신호는 이동에 거의 불변이기 때문에 이들 신호의 각 영역을 분석하거나 처리할 때 동일한 가중치를 활용할 수 있어서 FCN보다 훨씬 적은 매개변수를 이용하고도 오히려 보다 우수한 성능을 보여 줄 수 있는 장점이 있다. 다시 말해서,

*정회원, 한국외국어대학교 공과대학 정보통신공학과
접수일자 2024년 1월 9일, 수정완료 2024년 1월 30일
게재확정일자 2024년 2월 9일

Received: 9 January, 2024 / Revised: 30 January, 2024 /
Accepted: 9 February, 2024

*Corresponding Author: hihahn@hufs.ac.kr

Dept. Information and Communications Eng., College of Engineering, Hankuk University of Foreign Studies, Korea

이미지 $f(x,y)$ 와 필터 커널 $\psi(x,y)$ 의 합성 곱(convolution)인 $o(x,y)=\psi(x,y) * f(x,y)$ 에서 $f(x,y)$ 를 공간 이동시키면 그 결과는 $o(x,y)$ 를 이동시킨 결과와 동일하다. 이러한 특성을 공간 이동에 등변(equivariant)이라고 부른다.

CNN에서 합성 곱 계층은 이동에 등변이기 때문에 심층 망에서 효과적으로 이용될 수 있다. 즉, 이미지를 이동한 다음 합성 곱을 적용한 결과는 합성 곱을 먼저 수행한 다음 이동한 것과 동일하므로 각 계층마다 이동 대칭성이 보존되는 특징이 있다. 이러한 이동 대칭성으로 말미암아 검출하고자 하는 객체의 위치에 상관없이 동일한 필터로 객체 검출을 가능하게 해준다. 하지만, 객체가 무작위적으로 회전되어 있으면 검출 성능이 급격히 저하되는 문제가 발생한다. 이러한 문제를 해결하기 위하여 Cohen, et al.[4]은 이동 대칭성 외에 회전, 거울 대칭(reflection) 등의 다른 대칭성을 추가하기 위하여 군-등변 CNN(Group Equivariant CNN)을 제안한다. 기존의 CNN은 이동 대칭성만을 이용하지만 군-등변 CNN은 이동 대칭성 외에 회전, 거울 대칭(reflection) 등의 다른 대칭성을 더 활용함으로써 가중치 공유를 보다 더 증가시키면서도 향상된 성능을 보여준다. 그런데, 디지털 이미지는 직교 격자로 양자화된 데이터 구조이어서 직교 격자 위의 커널을 임의의 각으로 회전시키면 커널의 크기가 변하기 때문에 현실적으로 군 등변 CNN 구현을 위하여 90° 단위인 4개의 각으로 표현된 커널밖에 이용할 수 없는 한계가 있다. 본 논문에서 구현한 군-등변 CNN이나 게이지-등변(gauge equivariant) CNN 등은 최근 들어 수학, 물리 전공자들을 중심으로 연구가 진행되고 있다. 대표적인 예로, Aronson[5]과 Gerken, et al.[6]은 동차 공간(homogeneous space)에서의 동차 벡터 다발(homogeneous vector bundle) 이론을 이용하여 군-등변 CNN을 기하학적으로 해석하는 기반을 제공한다.

임의의 각도로 회전된 객체를 검출하기 위한 다른 방법으로는 steerable 필터를 활용하는 것이다. 즉, CNN의 합성 곱 필터들을 steerable 필터로 대체한다. steerable 필터 이론은 모든 필터를 기저 필터의 선형 결합으로 표현 가능하게 한다. 특히, 회전된 필터조차도 이러한 방식으로 구현 가능하게 하는 장점이 있다[7]. 여기서, 기저 필터를 가우시안 필터의 미분으로 정의하면 효율적으로 표현할 수 있는 장점이 있다. steerable 필터를 CNN에 적용할 때 가장 큰 장점 중의 하나는 학습시켜야 할 매개 변수의 수를 크게 줄일 수 있다는 점이다.

즉, CNN은 그 내부의 수 많은 필터들의 계수들을 모두 학습시켜야 하므로 $n \times n$ 크기의 이차원 필터가 k 개 할당된 경우에 학습시켜야 할 파라미터 수는 kn^2 에 비례하여 증가한다. 이로 인하여 계산량과 메모리 사용량이 급격히 증가하는 문제와 더불어, 학습 데이터셋의 크기가 충분하지 않으면 과대 적합될 위험을 내포한다. 이에 반해서 steerable 필터를 적용하는 기법은 모든 필터에 대하여 선형 결합 계수만 요구되므로 기저 필터의 수가 크지 않으면 학습시켜야 할 파라미터의 수가 크게 줄어드는 장점이 있다. 이러한 특성은 과대 적합을 피하면서도 학습 데이터셋의 크기를 줄일 수 있는 효과가 있다.

본 논문에서는 기존의 CNN과 군 등변 CNN, steerable 필터를 이용한 CNN 등을 각각 구현하여 MNIST 데이터셋과 이를 무작위적으로 회전시킨 rMNIST를 이용한 실험을 통하여 각 방식의 특징과 장단점 등을 비교분석한다. 본 논문의 구성은 다음과 같다. II절에서는 배경 이론을 리뷰하고, 군-등변 CNN과 steerable 필터를 이용한 CNN의 구현 내용을 III절에서 서술한다. IV절에서는 실험을 통하여 기존의 CNN과 군-등변 CNN 및 steerable 필터를 이용한 CNN 등의 성능을 비교 분석한 후, 마지막으로 V 절에서는 결론을 맺고 향후 연구 진행 방향에 대하여 논의한다.

II. 배경 이론

1. 군 등변 CNN

이미지에서 공간 이동뿐만 아니라 회전을 포함하는 일반적인 합성곱은 다음과 같이 나타낼 수 있다[4, 8].

$$[\kappa \star f](g) = \sum_{(u,v) \in \mathbb{Z}^2} \kappa(g^{-1}(u,v))f(u,v) \quad (1)$$

여기서 κ 는 커널이고 f 는 특징 맵(feature map)을 나타낸다. \star 는 상관 관계(correlation)를 나타내지만 본 논문에서는 합성 곱으로 부른다. $g \in \mathcal{G}$ 는 다음과 같은 이차원 유클리드 행렬 군(group)의 원소이다.

$$g = \begin{bmatrix} \cos\theta & -\sin\theta & x \\ \sin\theta & \cos\theta & y \\ 0 & 0 & 1 \end{bmatrix} \quad (2)$$

식(1)에서 $g^{-1}(u,v)$ 는 군의 연산으로 다음과 같이 표현할 수 있으므로 좌표의 회전과 공간 이동의 결합을 나타낸다.

$$g^{-1}(u,v) = \begin{bmatrix} \cos\theta & -\sin\theta & x \\ \sin\theta & \cos\theta & y \\ 0 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}$$

$$\simeq \begin{bmatrix} \cos\theta & \sin\theta & -x \\ -\sin\theta & \cos\theta & -y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \simeq r^{-1} \begin{bmatrix} u \\ v \end{bmatrix} \oplus \begin{bmatrix} u-x \\ v-y \end{bmatrix} \quad (3)$$

따라서, $[\kappa \star f](g)$ 는 사실상 $[\kappa \star f](k, x, y)$ 로 나타낼 수 있다. 여기서, k 는 다음과 같이 회전을 나타내는 인수이다.

$$r(k) = \begin{bmatrix} \cos\theta(k) & -\sin\theta(k) \\ \sin\theta(k) & \cos\theta(k) \end{bmatrix} \quad (4)$$

즉, 기존의 합성 곱은 이차원 합성 곱인데 반하여, 군-등변 합성 곱은 위치와 회전 정보를 인수로 갖는 삼차원 합성 곱이라고 볼 수 있다. 그런데, 식(1)을 보면, 입력 계층에서 f 는 이미지가어서 위치 정보만의 함수이므로 CNN의 첫 번째 계층에서만 적용된다. 두 번째 계층부터는 f 가 특징 맵이어서 κ 과 f 에 회전 인수가 추가되므로 다음과 같이 확장된다[4].

$$[\kappa \star f](g) = \sum_{h \in G} \kappa(g^{-1}h) f(h) \quad (5)$$

즉, 입력 이미지 f 는 격자 Z^2 의 함수이지만 특징 맵 $\kappa \star f$ 는 g 의 함수(삼차원 (k, x, y) 의 함수)이기 때문이다. 이를 군-등변 합성 곱이라고 부른다. 군-등변 CNN은 군-합성 곱을 이용하는데, 이는 일반적인 합성 곱보다 가중치 공유를 더욱 증가시킬 수 있고, 계산량 면에서 큰 부담 없이 구현 가능하다[4, 8]. 필터를 임의의 각도로 회전하면 그 크기가 변하기 때문에 등변 조건을 만족시키기가 사실상 불가능하다. 이러한 이유로 구현을 위해서는 식(2)에 주어진 g 를 $\theta = \frac{k\pi}{2}$, $0 \leq k < 4$ 로 이산화시킨 부분 군으로 제한할 수 밖에 없다. 따라서, $g \in G_2$ 에 대한 변환은 정수 이동과 90° 단위의 회전에 분리될 수 있다[4].

$$G_2 = \begin{bmatrix} \cos \frac{k\pi}{2} & -\sin \frac{k\pi}{2} & u \\ \sin \frac{k\pi}{2} & \cos \frac{k\pi}{2} & v \\ 0 & 0 & 1 \end{bmatrix} \quad (6)$$

신경망을 이용하여 이미지 내의 객체를 분류하거나 인식하기 위해서는 기본적으로 다양한 위치에 있거나 회전된 객체를 인식하기 위하여 그에 해당하는 이미지들을 충분히 획득하여 학습 데이터셋에 포함시켜야 신경망 학습이 제대로 이루어진다. 이 과정에서 엄청난 양의 학습 데이터(data augmentation)가 요구된다. 하지만, 군-등변 CNN은 객체의 위치 변화를 커널의 회전과 이동을

통한 대칭성으로 해결할 수 있기 때문에 학습 데이터셋의 양을 크게 줄일 수 있는 장점이 있다[4].

2. steerable 필터를 이용한 CNN

1991년에 Freeman과 Adelson[7]은 임의의 각도로 회전된 필터는 기저 필터의 선형 결합으로 구현할 수 있음을 증명한다. 이러한 필터를 steerable 필터라고 부른다. 예를 들어, 원형 대칭인 가우시안 함수 $G(x, y) = e^{-(x^2+y^2)}$ 에 대하여 x 방향과 y 방향으로의 편미분은 다음과 같이 구할 수 있다.

$$G_x = \frac{\partial}{\partial x} e^{-(x^2+y^2)} = -2xe^{-(x^2+y^2)} \quad (7)$$

$$G_y = \frac{\partial}{\partial y} e^{-(x^2+y^2)} = -2ye^{-(x^2+y^2)} \quad (8)$$

G_x 와 G_y 는 각각 x 축과 y 축 방향을 향하는 필터로 작용한다. 즉, G_y 는 G_x 를 90° 만큼 회전시킨 필터로 볼 수 있다. 이를 이용하여 임의의 각 θ 로 회전시킨 필터 G^θ 는 다음과 같이 $G_x (= G^{0^\circ})$ 와 $G_y (= G^{90^\circ})$ 의 선형 결합으로 구할 수 있다[7].

$$G^\theta = \cos\theta G_x + \sin\theta G_y \quad (9)$$

다시 말하면, G_x 와 G_y 는 G^θ 의 기저 필터의 역할을 한다. 이 개념을 일반화시키면 원형 대칭(radially symmetric)인 원도우 함수와 x, y 에 대한 다항식의 곱은 steerable한 성질이 있다. 따라서, G_x 와 G_y 가 steerable하듯이 $G(x, y)$ 를 여러 번 미분하여도 $G(x, y)$ 와 다항식의 곱으로 표현되므로 steerable하다. $G(x, y)$ 를 보다 일반적으로 다음과 같이 표현하면,

$$g(x, y) = e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (10)$$

이에 대한 4차까지의 편미분은 다음과 같다.

$$g_x = -\frac{x}{\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad (11)$$

$$g_{xx} = \left(\frac{x^2}{\sigma^4} - \frac{1}{\sigma^2} \right) e^{-(x^2+y^2)/2\sigma^2}$$

$$g_{xxx} = -\left(\frac{x^3}{\sigma^6} - \frac{3x}{\sigma^4} \right) e^{-(x^2+y^2)/2\sigma^2}$$

$$g_{xxxx} = \left(\frac{x^4}{\sigma^8} - \frac{6x^2}{\sigma^6} + \frac{3}{\sigma^4} \right) e^{-(x^2+y^2)/2\sigma^2}$$

여기서, 2차 편미분은 g_{xx}, g_{xy}, g_{yy} 등, 세 개가 있으며, 일반적으로 n 차 편미분의 갯수는 $n+1$ 개다. 식(10)과 식(11)에서 $\sigma^2 = 1/2$ 일 때, $g(x, y)$ 와 곱해진 다항식 $H_0(x) = 1$, $H_1(x) = 2x$, $H_2(x) = 4x^2 - 2$, $H_3(x) = 8x^3 - 12x$ 등을

에르미트 다항식(Hermite polynomial)이라고 부르는데, 다음과 같은 직교 관계를 만족시킨다[9].

$$\int_{-\infty}^{\infty} e^{-x^2} H_m(x) H_n(x) dx = 2^n n! \sqrt{\pi} \delta_{mn} \quad (12)$$

여기서, δ_{mn} 은 크로네커 델타 함수를 나타낸다. 따라서 식(10)과 식(11)의 함수들을 선택하여 적절히 정규화시키면 직교성(orthonormal) 있는 기저 필터를 구할 수 있다. 예를 들어, $x \rightarrow x \cos \theta - y \sin \theta$ 로 치환하면, g_{xx} , g_{xy} , g_{yy} 를 이용하여 θ 만큼 회전된 필터 $\phi^\theta(x, y)$ 를 다음과 같이 구할 수 있다[7, 10].

$$\begin{aligned} \phi(x, y)^\theta &= g_{xx} \Big|_{x \rightarrow x \cos \theta - y \sin \theta} \\ &= \{4(x \cos \theta - y \sin \theta)^2 - 2\} G(x, y) \\ &= \{4(x^2 \cos^2 \theta - 2xy \sin \theta \cos \theta + y^2 \sin^2 \theta) - 2\} G(x, y) \\ &= \{\cos^2 \theta (4x^2 - 2) - 2 \sin \theta \cos \theta (2xy) + \sin^2 \theta (4y^2 - 2)\} G(x, y) \\ &= \cos^2 \theta g_{xx} - 2 \sin \theta \cos \theta g_{xy} + \sin^2 \theta g_{yy} \end{aligned} \quad (13)$$

여기서, 계산 상의 편의를 위해 $\sigma^2 = 1/2$ 로 정한다. 이 과정은 편미분 차수에 관계 없이 항상 성립한다. 따라서, 적절한 steerable 기저 필터를 구하면 임의의 각도로 회전된 필터도 기저 필터의 선형 결합으로 구할 수 있다.

본 논문에서는 식(10)을 포함한 2차까지의 편미분 함수 g , g_x , g_y , g_{xx} , g_{xy} , g_{yy} 등, 6 개의 함수를 기저 필터로 이용한다. 그 이유는 기저 필터의 수를 변경하면서 CNN을 구현하였을 때, 3차 이상의 편미분 함수는 추가하여도 성능 개선 효과가 무시할 정도로 적었기 때문이다. 기존의 CNN은 각 계층 당 다수의 필터를 지정하고 각 필터의 계수를 학습시킨다. 하지만, steerable 필터를 이용하면 각 계층 당 6 개의 기저 필터만을 이용하여 특징 공간과 합성 곱을 수행한 다음, 6 개의 기저 필터 출력의 선형 결합으로 각 할당된 필터에 대한 출력을 계산할 수 있다. 예를 들어, CNN의 임의의 한 계층에 특징 맵은 $f(x, y)$ 이고 필터가 N 개 할당되어 있다고 가정한다. 그 계층에서 2D 필터 커널은 다음과 같이 주어진다.

$$\psi^k(x, y) = \alpha_0^k g + \alpha_1^k g_x + \dots + \alpha_5^k g_{yy} \quad (14)$$

여기서, $0 \leq k < N$ 이다. 그리고 각 기저 필터의 출력을 다음과 같이 나타내면,

$$\begin{aligned} v_0(x, y) &= g \star f, \quad v_1(x, y) = g_x \star f \dots \\ v_5(x, y) &= g_{yy} \star f \end{aligned} \quad (15)$$

각 필터 커널의 출력은 다음과 같이 구할 수 있다[7, 11].

$$\begin{aligned} \psi^k \star f(x, y) &= (\alpha_0^k g + \alpha_1^k g_x + \dots + \alpha_5^k g_{yy}) \star f(x, y) \\ &= \sum_{i=0}^5 \alpha_i^k v_i(x, y) \end{aligned} \quad (16)$$

이 때, 기저 필터는 고정된 필터이고, 학습은 각 필터에 할당된 선형 결합 계수 α_j^k : ($0 \leq j < 6, 0 \leq k < N$)에 대해서만 이루어진다. 그런데, CNN은 3×3 또는 5×5 커널을 주로 사용하기 때문에 각 계층 당 학습시켜야 할 파라미터의 수가 $9N$ 또는 $25N$ 인 반면, steerable 필터를 이용하면 $6N$ 으로 줄어드는 특징이 있다.

이에 대한 가장 큰 장점으로서는 작은 크기의 훈련 데이터셋을 이용하면서도 과잉 적합을 피할 수 있을 것이라는 점이다. 본 논문에서는 이에 초점을 맞추어 실험을 통하여 steerable 필터를 이용한 CNN의 특성을 확인하고 CNN, 균 등변 CNN 등과 성능을 비교분석한다.

III. 제안 알고리즘 구현 방안

본 논문에서는 기존의 CNN과 균 등변 CNN, steerable 필터를 이용한 CNN 등의 특성과 장단점 등을 비교분석하기 위하여 최대한 동일한 구조를 갖도록 이들을 설계하고 구현한다. 기존 모델을 정하기 위하여 CNN을 세 개의 합성곱 층과 완전 연결 한 계층으로 구성한다. 합성곱 층은 모두 3×3 필터를 사용하는데 첫 번째 층은 32 개, 다른 층은 64개의 필터를 각각 할당한다. 각 층은 (2, 2) 풀링을 적용하고 렐루(relu) 활성화 함수를 채택한다. 완전 연결 층에서는 활성화 함수로 소프트맥스 함수를 이용한다. 교차 엔트로피 함수를 손실함수로 정하고 Adam 방식으로 파라미터를 학습한다. 이를 기반으로 균 등변 CNN과 steerable 필터를 이용한 CNN을 각각 구현한다. 균 등변 CNN은 위에서 설명한 CNN과 기본적으로 거의 동일한 방식으로 동작된다. 다만, 계층의 각 필터는 90° 씩 회전하면서 특징 맵과 합성 곱을 계산한 다음 네 개의 값에 대한 평균을 출력시킨다는 점이 다르다. 따라서, 균 등변 CNN이 기존의 CNN보다 4배 정도 합성 곱을 더 연산한다는 점 외에는 위 두 방식은 프로그램 구조 면에서 매우 유사하다.

steerable 필터를 이용한 CNN은 앞에서 설명한 바와 같이 2차까지의 편미분 함수 g , g_x , g_y , g_{xx} , g_{xy} , g_{yy} 등, 6 개의 함수를 기저 필터로 선정하여, 이를 기반으로 각 계층에 필요한 필터를 구성하고 합성 곱 연산을 수행한다. 여기서, $\sigma = 1$ 로 고정한다. 구현하기 위해서는 우선, 기저 필터들을 이산화시켜야 한다. 본 절에서는 g_{xx} , g_{xy} , g_{yy} 에 대한 이산화 과정을 중심으로 설명하기로 한다. 기본적으로 모든 기저 필터는 분리 가능(separable)

하므로 다음과 같이 이산화시킨다.

$$\begin{aligned} g_{xx} &\rightarrow f_1(n)f_2(n) \\ g_{xy} &\rightarrow f_3(n)f_3(n) \\ g_{yy} &\rightarrow f_2(n)f_1(n) \end{aligned} \quad (17)$$

여기서, $f_1(n)$, $f_2(n)$, $f_3(n)$ 은 다음과 같은 샘플링 과정을 통하여 구한다[7].

$$\begin{aligned} f_1(n) &= (4x^2 - 2)e^{-x^2} \Big|_{x=0.67n} \\ f_2(n) &= e^{-(0.67n)^2} \quad -4 \leq n \leq 4 \\ f_3(n) &= 0.67e^{-(0.67n)^2} \end{aligned} \quad (18)$$

steerable 필터의 선형 결합 계수의 학습 기법

Input : $l-1$ 번째 계층의 i 번째 특징맵 o_i^{l-1} , desired output $\{d_0, \dots, d_9\}$, 기저필터 $\{\phi_0, \dots, \phi_9\}$

1. o_i^{l-1} 와 $\{\phi_0, \dots, \phi_9\}$ 의 합성곱 $\{o_j^l, \dots, o_j^l\}$ 계산
2. l 번째 계층의 j 번째 특징맵 $o_j^l = \alpha_{i,j,0}^l \nu_0 + \dots + \alpha_{i,j,5}^l \nu_5$ 계산
3. 최종계층의 o_j^l 을 이용하여 n 번째 뉴런 $d_{j,n}$ 계산
4. 교차 엔트로피 E 계산
5. 파라미터 α_{ijk}^l 에 대하여 gradient $\frac{\partial E}{\partial \alpha_{ijk}^l}$ 계산
6. 파라미터 학습 : $\alpha_{ijk}^l = \alpha_{ijk}^l - \eta \frac{1}{N} \sum_{m=1}^N \left(\frac{\partial E}{\partial \alpha_{ijk}^l} \right)_m$

Output : 학습된 파라미터 α_{ijk}^l

그림 1. 경사 하강법으로 선형 결합 계수 α_j^k 를 학습하는 기법
 Fig. 1. Algorithm of learning the parameters α_j^k of linear combination in the gradient descent method.

예를 들어, g_{xx} 에 해당되는 이차원 필터는 모두 0으로 채워진 9×9 행렬의 정 중앙 위치에만 1을 저장한 다음, $f_1(n)$ 과 $f_2(n)$ 를 x 축과 y 축 방향으로 각각 합성곱 연산을 함으로써 구한다. 동일한 방법으로 다른 기저 필터의 계수들을 구할 수 있다. 각 합성곱 계층에 6개의 기저 필터가 할당되고 식(16)에 제시한 방법으로 각 필터의 출력이 계산된다. 본 실험에서 구현한 CNN과 군 등변 CNN은 모두 3×3 필터를 사용하기 때문에 합성곱 계층에서 요구되는 필터 파라미터 수는 $(32+64+64) \times 9 = 1,440$ 개이다. 하지만, steerable 필터를 이용한 CNN에서는 이에 대한 파라미터 수가 $(32+64+64) \times 6 = 960$ 개로 줄어드는 장점이 있다. 그림 1은 steerable 필터의 선형 결합 계수에 대한 학습 방법을 설명하고 있다^[11, 12].

IV. 실험 결과 및 분석

본 논문에서는 기존의 CNN과 군 등변 CNN, steerable 필터를 이용한 CNN 등을 각각 구현하여 *MNIST* 숫자 이미지 데이터셋과 이를 무작위적으로 회전시킨 *rMNIST* [13]를 이용한 실험으로 각 방식의 특성과 장단점 등을 비교분석한다. *MNIST*가 60,000 + 10,000개의 이미지로 구성된 반면, *rMNIST*는 그림 2에 제시한 바와 같이 무작위적으로 회전된 50,000 + 12,000개의 데이터를 가진다.



그림 2. 회전된 MNIST 데이터셋의 예
 Fig. 2. Samples of the rotated MNIST dataset

CNN은 세 개의 합성곱 계층과 완전 연결 한 계층으로 구성한다. 합성곱 층은 모두 3×3 필터를 사용하는데 첫 번째 층은 32개, 다른 층은 64개의 필터를 할당한다. 군 등변 CNN은 위에서 설명한 CNN과 기본적으로 동일한 구조로 구성된다. 다만, 각 필터는 90° 씩 회전하면서 합성곱을 계산한 다음 평균을 내서 출력을 계산한다는 점이 다르다. steerable 필터를 이용한 CNN은 앞에서 설명한 바와 같이 2차까지의 편미분 함수 g , g_x , g_y , g_{xx} , g_{xy} , g_{yy} 등, 6개 함수를 기저 필터로 선정하여, 이를 기반으로 합성곱 연산을 수행한다. 우선, CNN의 세가지 모델에 대하여, *MNIST* 데이터셋으로 학습 데이터셋의 크기를 변경하면서 50 이포크 동안 학습한 후에 10,000개의 테스트 데이터로 구한 인식 성능을 그림 3에 제시한다.

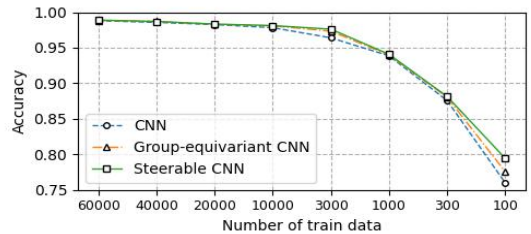


그림 3. *MNIST*로 학습 데이터셋의 크기를 가변시키면서 학습 시, CNN, 군 등변 CNN, steerable CNN의 인식 성능의 변화

Fig. 3. Classification performance of CNN, group-equivariant CNN and steerable CNN when the various subsets of the *MNIST* dataset are used as training data.

이 그림에서 알 수 있듯이, 거의 정 방향으로 위치한 *MNIST* 데이터에 대해서는, 학습 데이터의 수가 감소함에 따라 steerable 필터를 이용한 CNN이 매우 근소한 차이지만 보다 나은 성능을 보이고 있으나 기본적으로 세 방식 모두 우열을 가리기 어려운 정도의 우수한 성능을 보여주고 있다.

그림 4는 회전된 *rMNIST* 데이터셋에 대하여 위와 동일한 실험을 수행하여 구한 인식률을 보여준다. 학습 데이터의 수가 충분히 클 때에는 세 방식 모두 성능이 우수하지만 학습 데이터의 수가 감소함에 따라 CNN은 인식률이 급격히 떨어지는 반면, 다른 두 방식은 비교적 완만한 감소 속도를 보이고 있음을 확인할 수 있다. steerable 필터를 이용한 알고리즘이 군 등변 CNN 보다 학습 데이터의 수가 작을 때 특히 인식률이 더 높은 이유는 전자가 후자에 비해 학습 파라미터의 수가 2/3 밖에 안되어 과잉 적합이 발생할 가능성이 그만큼 줄어들기 때문인 것으로 해석된다.

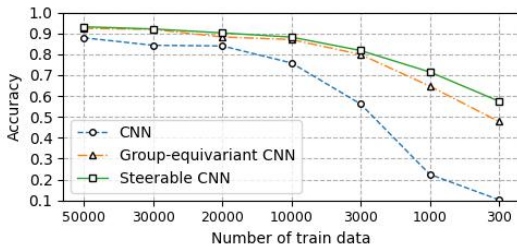


그림 4. *rMNIST*로 학습 데이터셋의 크기를 가변시키면서 학습 시, CNN, 군 등변 CNN, steerable CNN의 인식 성능의 변화

Fig. 4. Classification performance of CNN, group-equivariant CNN and steerable CNN when the various subsets of the *rMNIST* dataset are used as training data.

V. 결 론

본 논문에서는 이미지 공간에서 무작위로 회전된 객체에 대한 검출 성능을 확인하기 위하여, 동일한 구조로 CNN과 군 등변 CNN, steerable 필터를 이용한 CNN 등을 각각 구현한 다음, 실험을 통하여 이들의 성능을 비교분석하였다. 실험에 의하면, 객체가 임의의 방향으로 회전되어 있으면 학습 데이터가 충분히 많아도 학습이 잘 이루어지지 않음을 확인하였다. 반면에 군 등변 CNN과 steerable 필터를 이용한 CNN은 기존의 CNN에 비

하여 월등한 성능을 보여주고 있다. 군-등변 CNN은 학습된 각 계층의 필터를 4 개의 고정된 방향으로 합성 곱한 결과를 일률적으로 평균하여 특징 맵을 구하기 때문에 객체의 회전 정도에 관계 없이 동일한 방법으로 출력을 구한다. 이에 반하여 steerable 필터를 이용한 CNN은 객체의 회전량을 기저 필터의 선형 결합 계수에 반영되도록 학습한다는 점에서 구조적 우위에 있다고 평가될 수 있을 것이다. 향후에는 군 등변 CNN과 steerable 필터 이론을 결합하여 회전에 등변인 CNN으로 확장시키는 연구를 계속 진행할 계획이다.

References

- [1] H. I. Hahn, "Technique proposal to stabilize Lipschitz continuity of WGAN based on regularization terms," *Journal of The Institute of Internet, Broadcasting and Communication*, Vol. 20, No. 1, pp. 239-246, Feb. 2020.
DOI: <https://doi.org/10.7236/JIIBC.2020.20.1.239>
- [2] S. K. Kim and J. G. Ahn "Tomato Crop Diseases Classification Models Using Deep CNN-based Architectures," *Journal of the Korea Academia-Industrial cooperation Society*, Vol. 22, No. 5, pp. 7-14, 2021.
DOI: <https://doi.org/10.5762/KAIS.2021.22.5.7>
- [3] J. E. Lee, K. B. Jang, S. H. Lim, "Implementation and Performance Analysis of Mixed Precision-based CNN Inference," *Journal of Korean Institute of Information Technology*, Vol. 21, No. 12, pp. 77-85, 2023.
DOI: <http://dxdoi.org/10.14801/jkiit.2023.21.12.77>
- [4] T. S. Cohen and M. Welling, "Group Equivariant Convolutional Networks," *arXiv:1602.07576v3 [cs.LG]* 3, Jun. 2016.
DOI: <https://doi.org/10.48550/arXiv.1602.07576>
- [5] J. Aronson, "Homogeneous Vector Bundles and G-equivariant Convolutional Neural Networks", *arXiv:2105.05400v1 [cs.LG]*, May, 2021.
DOI: <https://doi.org/10.48550/arXiv.2105.05400>
- [6] J. E. Gerken, J. Aronsson, O. Carlsson, H. Linander, F. Ohlsson, and C. Petteron Aronson, "Geometric Deep learning and Equivariant Neural Networks", *arXiv:2105.13926v1 [cs.LG]*, 28 May, 2021.
DOI: <https://doi.org/10.48550/arXiv.2105.13926>
- [7] W. T. Freeman and E. H. Adelson, "The Design and Use of Steerable Filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 13, No. 9, pp.891-906, Sep. 1991.
DOI: <https://doi.org/10.1109/34.93808>
- [8] M. M. Bronstein, J. Bruna, T. Cohen, and P. Velickovic, "Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges," *arXiv:2104.13478*.

DOI: <https://doi.org/10.48550/arXiv.2104.13478>
DOI: <https://doi.org/107236/JIIBC.2020.20.1>

- [9] J. Indritz, "An Inequality for Hermite polynomials", Proceedings of the American Mathematical Society, 12 (6): 981-983, , 1961.
DOI: <https://doi.org/10.1090/S0002-9939-1961-0132852-2>
- [10] Y. Hel-Or and P.C. Teo, "Canonical decomposition of steerable functions," Proceedings CVPR IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 809-816, 1996.
DOI: <https://doi.org/10.1109/CVPR.1996.517165>.
- [11] J. Jacobsen, J. Gemert, Z. Lou, and A.W.M. Smeulders, "Structured Receptive Fields in CNNs," arXiv:1605.02971v2 [cs.CV] 13, May, 2016.
DOI: <https://doi.org/10.48550/arXiv.1605.02971>
- [12] T. Kobayashi, "Analyzing Filters Toward Efficient ConvNet," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, pp. 5619-5628.
DOI: <https://doi.org/10.1109/CVPR.2018.00589>.
- [13] H. Larochelle, D. Erhan, A. Courville, J. Bergstra, and Y. Bengio, "An Empirical Evaluation of Deep Architectures on Problems with Many Factors of Variation," Proceedings of the 24th International Conference on Machine Learning(ICML), 25, 2007.
DOI: <https://doi.org/10.1145/1273496.1273556>

저 자 소 개

한 희 일(정회원)



- 1984년 : 서울대학교 제어계측공학과 학사 졸업.
 - 1988년 : 서울대학교 제어계측공학과 석사 졸업.
 - 1995년 : University of Arizona 전 기및컴퓨터공학과 박사 졸업.
 - 2023년 ~ 현재 : 한국외국어대학교 공과대학 정보통신공학과 교수.
- 주관심분야 : 신호처리, 컴퓨터비전, 머신러닝, 게이지 이론

※ 본 논문은 2023년도 한국외국어대학교 교내 학술연구지원에 의하여 연구되었음