



일반논문 (Regular Paper)

방송공학회논문지 제29권 제5호, 2024년 9월 (JBE Vol.29, No.5, September 2024)

<https://doi.org/10.5909/JBE.2024.29.5.739>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

# 스파이킹 뉴럴 네트워크의 다양한 인코딩 기법에 따른 영상 분류 성능 분석

오범석<sup>a)</sup>, 김선옥<sup>a)†</sup>

## Analysis of Image Classification Performance on Various Encoding Methods In Spiking Neural Networks

Beomseok Oh<sup>a)</sup> and Sunok Kim<sup>a)†</sup>

### 요약

인간의 신경 시스템을 모방한 스파이킹 뉴럴 네트워크(Spiking Neural Networks, SNNs)는 3세대 인공 신경망으로, 인공 신경망(ANNs)에 비해 낮은 전력 소모량, 높은 에너지 효율성을 띠며 많은 연구자들의 관심을 받고 있다<sup>[1]</sup>. 그러나 SNN은 이산적인 스파이킹 신호를 처리하므로 부동소수점 연산인 ANN과 다르게 오차 역전파를 이용한 학습에 어려움이 있어 성능이 낮고, 입력 이미지에 대해 이산 스파이킹 신호로 변환해 주는 과정인 인코딩이 필요하다<sup>[2]</sup>. 본 연구에서는 입력 이미지를 이산 스파이킹 신호로 변환하는 각종 인코딩 기법에 대해 탐구하며 인코딩이 적용된 SNN과 ANN에 데이터 증강 기법을 적용해 실험을 진행한다. 일반 컬러 이미지와 합성 개구 레이더(Synthetic Aperture Radar, SAR) 영상을 이용해 동일한 조건에서의 성능을 비교 및 분석하고, 저전력의 환경에서 SNN이 ANN을 대체하여 분류 문제를 수행할 수 있음을 시사하고자 한다.

### Abstract

Spiking Neural Networks (SNNs), which mimic the human nervous system, are the third generation of Artificial Neural Networks (ANNs) and have attracted the attention of many researchers due to their low power consumption and high energy efficiency compared to Artificial Neural Networks (ANNs)<sup>[1]</sup>. However, SNNs process discrete spiking signals, so unlike ANNs, which are floating-point operations, they have difficulty learning using error backpropagation, resulting in low performance, and require encoding, a process that converts input images into discrete spiking signals<sup>[2]</sup>. In this study, we explore various encoding techniques to convert input images into discrete spiking signals and apply data augmentation techniques to SNNs and ANNs with encoding techniques. We use color images and Synthetic Aperture Radar (SAR) images to compare and analyze their performance under the same conditions, and suggest that SNNs can replace ANNs for classification problems in low-power environments.

Keyword : Spiking Neural Networks, Encoding Methods, SAR Image, Classification

## I. 서론

인공신경망(Artificial Neural Networks, ANN)이 발전하며 다양한 분야에 인공지능이 많은 이목을 끌고 있다. 특히 심층신경망(Deep Artificial Neural Networks, DNN)은 현재 강력한 성능을 바탕으로 여러 분야에 사용되고 있다. 그러나 DNN의 경우, 많은 양의 컴퓨팅 자원을 필요로 하며 이에 따라 자연스럽게 많은 양의 전력 소모가 발생한다. 특히 위성을 통해 얻을 수 있는 SAR 이미지의 경우, 위성 환경의 특성상 높은 전력을 사용하기 어려우므로 높은 전력 소모량의 고성능 DNN 모델을 쉽게 적용하기 어려운 점이 있다. 이를 보완하기 위한 전략으로 3세대 인공신경망이라 불리는 스파이킹 뉴럴 네트워크(Spiking Neural Networks, SNN)가 주목받고 있다.

SNN은 부동소수점 연산으로 진행되는 DNN과 달리 이산적인 스파이크 신호를 이용해 학습하므로 연산량이 대폭 감소하고 이에 따라 전력 소모율도 감소한다. 특히 많은 양의 전력을 사용할 수 없는 엣지 디바이스(Edge Device)에 SNN을 도입할 수 있어 많은 관심을 받고 있다.

위 같은 장점에도 불구하고 SNN이 DNN만큼 널리 쓰이지 못하고 있는데, 역설적으로 앞서 언급된 이산적인 스파이크 신호 때문이다. DNN에선 미분을 기반으로 하는 오차 역전파 학습 기법이 등장했고, 이를 통해 엄청난 발전을 이뤄왔다. 하지만 SNN의 0과 1의 이산적인 스파이크 신호를 이용하는 특성으로 인해 기존 DNN의 오차 역전파 알고리즘은 사용할 수 없게 된다. 이를 해결하기 위한 STDP (Spike-Timing-Dependent Plasticity)<sup>[3]</sup>, Surrogate gradient descent<sup>[4]</sup> 등 여러 연구가 진행되고 있다. 그러나 아직은 DNN만큼 계층을 깊게 쌓는 등 성능을 끌어올리기 위한 연구가 부족하여 DNN을 비롯한 다른 인공신경망에 준하는

성능을 발휘하지 못하고 있다. 또한 3채널의 컬러 이미지 또는 합성 개구 레이더(Synthetic Aperture Radar, SAR) 이미지와 같은 단일 채널의 흑백 이미지를 이산 스파이크 신호로 변환하는 인코딩 기법 역시 필요로 한다.

본 연구에서는 입력 이미지를 스파이크 신호로 변환하는 여러 인코딩 기법의 특성을 파악하고, 데이터 증강 기법을 적용하여 이미지 분류 실험을 진행한 후 DNN을 포함해 SNN에서의 각 인코딩 모델의 성능을 비교 분석하고자 한다.

## II. 본론

### 1. 스파이킹 뉴럴 네트워크

SNN은 인간의 뇌가 정보를 어떻게 처리하는지를 모방하여 인공신경망에 적용한 3세대 인공신경망이다. 인간의 뇌를 구성하는 가장 작은 단위는 뉴런이고, 이 뉴런들은 시냅스를 통해 연결되어 있다. 뇌의 뉴런들은 스파이크를 생성하며, 시냅스를 기준으로 시냅스 전 뉴런의 스파이크를 시냅스 후 뉴런에게 전달한다. 이때 시냅스를 통해 전달받은 스파이크는 시냅스 후 뉴런 내부의 막전위(Membrane potential)를 변화시킨다. 이때 막전위가 어떠한 임계 전압을 넘게 되면 스파이크가 생성되고, 이를 시냅스를 통해 시냅스 후 뉴런에게 전송한다. 스파이크를 생성한 후 전송까지 마친 뉴런은 휴지 전위(Resting potential) 상태가 된다. 이 상태가 유지되는 동안은 스파이크가 입력으로 들어와도 막전위가 변하지 않는다. 위와 같은 과정이 연쇄적으로 작동하며 스파이크가 생성되고 전달된다. 이를 회로를 이용해 표현할 수 있는데 그것을 LIF(Leaky Integrate and Fire) 모델이라고 한다. 그림 1에서 전체적인 과정을 확인할 수 있다.

LIF 모델을 이용해 SNN을 학습시키기 위한 여러 방법이 있지만, 본 논문에서는 LIF 모델과 함께 Surrogate gradient descent<sup>[4]</sup> 방식을 이용해 SNN을 학습하였다.

Surrogate gradient descent<sup>[4]</sup> 방식이란 그림 2와 같이 이산적인 스파이크의 발생을 연속적인 함수로 근사하여 미분 가능 하도록 하는 방법이다. 기존에는 스파이크의 발생이 계단 함수와 같은 꼴을 가져 미분 시 기울기가 0이 되어버리는, 죽은 뉴런 문제(Dead Neuron problem)가 발생한다.

a) 한국항공대학교 대학원 인공지능학과(Dept. of Artificial Intelligence, Graduate School, Korea Aerospace University)

‡ Corresponding Author : 김선옥(Sunok Kim)

E-mail: sunok.kim@kau.ac.kr

Tel: +82-2-300-0262

ORCID: <https://orcid.org/0000-0002-9665-4214>

※ This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIP) (NRF-2021R1C1C2005202). (Corresponding author: Sunok Kim)

· Manuscript September 4, 2024; Revised September 9, 2024; Accepted September 9, 2024.

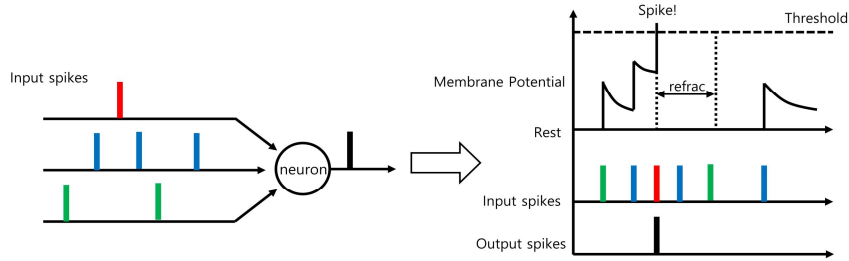


그림 1. LIF 모델 : 모델에 입력이 들어오고 SNN의 작동 방식에 의해 기준점을 넘게 된다면 Spike라는 이산적인 출력물이 생성 되는 구조

Fig. 1. LIF model : When input comes into the model and if crosses a threshold, it produces a discontinuous output called a spike, which is how SNNs work

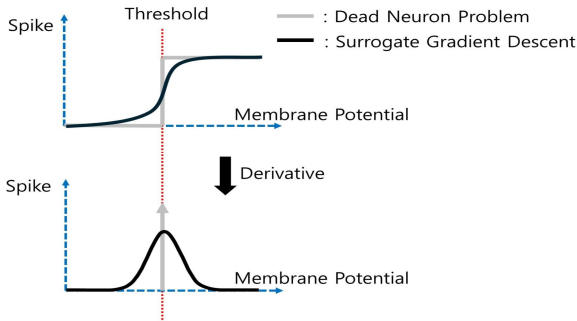


그림 2. Surrogate gradient descent : 기존 SNN의 Spike 생성 함수는 미분 시 기울기가 0이 되어 역전파 시 정보가 사라지는 Dead Neuron Problem이 발생하고 이를 기존 함수와 매우 비슷한 비선형적인 함수로 근사하여 미분 시 정보가 소실되는 문제를 방지하는 방법

Fig. 2. Surrogate gradient descent : The spike generation function of a traditional SNN suffers from the dead neuron problem, where the slope goes to zero on differentiation and information is lost on back-propagation, so we approximate it with a nonlinear function that is very similar to the original function to avoid the problem of losing information on differentiation

원활한 학습을 위해 계단 함수를 아주 유사한 연속적인 미분 가능한 함수 꼴로 근사하여 학습을 가능케 하는 방법이다.

## 2. 인코딩 기법

입력 이미지를 스파이크로 변환하는 인코딩 기법에는 많은 종류가 있지만, 본 논문에서는 대표적으로 사용되는 Direct<sup>[5]</sup>, Rate, Time-to-First-Spike(TTFS) 인코딩 방식을 비교한다.

### 2-1. Direct 코딩

Direct 코딩 방식은 그림 3과 같이 입력 이미지가 어떠한 전처리 없이 첫 번째 계층에 입력되어 부동소수점으로 출력물이 나오는 방식이다. 이는 시간 단계 T 동안 반복적으로 수행되며, 그 후 부동소수점 형태의 출력물을 토대로 스파이킹 계층으로 전해져 이산적인 스파이크 신호를 생성한다. 이전 연구들에 따르면<sup>[6]</sup>, Direct 코딩 방식<sup>[5]</sup>은 다른 인코딩 방식에 비해 높은 성능을 얻을 수 있다. 이는 부동소수점 연산의 정밀도가 유지되기 때문에 더 복잡한 패턴을 학습하고 인식할 수 있기 때문이다. 그러나 약의적인 기울기 공격에 취약하다는 단점이 있으며, 연산이 약간 더 복잡하다는 점도 고려해야 한다. 위 방법은 완벽히 이산적인 연산만 이뤄지는 것은 아니지만, 높은 성능을 얻고자 할 때 유의미하게 고려될 수 있다.

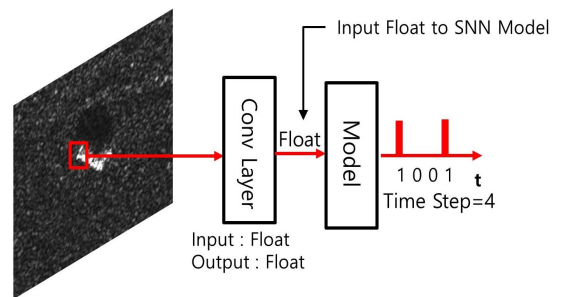


그림 3. Direct 코딩 : 부동 소수점 입력이 컨볼루션 계층을 거쳐 부동 소수점으로 출력된 후 SNN 모델에 입력되는 방식

Fig. 3. Direct coding : Floating-point input is passed through a convolution layer, output as floating point, and then fed into the SNN model

### 2-2. Rate 코딩

Rate 코딩 방식은 입력 이미지의 픽셀값에 비례하여 스파이크 수가 결정되는 방식이다. 이는 시간 단계 T 동안 반복되며, 스파이크가 푸아송 분포에서 샘플링되는 것과 동일한 원리이다. 이러한 방식은 몇 가지 장점을 가지고 있는데, 첫 번째로 잡음에 더 견고한 특성을 가지고 있다. 이는 스파이크 발생이 확률적으로 이루어지기 때문에 잡음이 일정 수준 이하일 경우 성능에 크게 영향을 미치지 않기 때문이다. 또한 얇은 네트워크뿐 아니라 깊은 네트워크나 많은 양의 데이터셋이 있을 때 유용하다. 또한 ReLU 활성화 함수와의 유사성으로 인해 정보 표현 능력이 우수하다<sup>[7][8]</sup>.

그러나 Rate 코딩 방식은 성능 면에서 Direct 코딩에 비해 부족하다. 특히, 지연시간(Latency)이 길고 에너지 효율성이 낮다. 그럼에도 불구하고, 실제 환경에서 무수히 많은 잡음이 발생하는 상황에서는 Rate coding 방식이 유리할 수 있다. Rate 코딩의 동작 방식은 그림 4를 통해 확인할 수 있다.

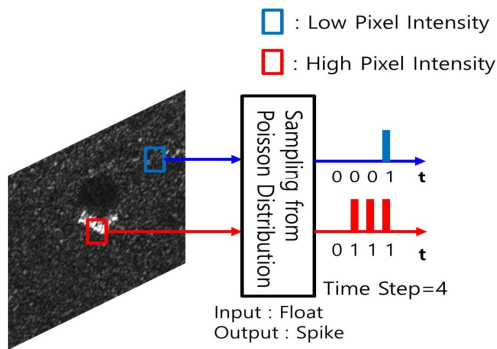


그림 4. Rate 코딩 : 픽셀의 값이 클수록 스파이크가 더 많이 생성되는 방식  
 Fig. 4. Rate coding : how the larger the value of a pixel, the more spikes it generates

### 2-3. Time-to-First-Spiking 코딩

Time-to-First-Spiking(TTFS) 코딩 방식은 Rate 코딩이 특정 시간 단계 T 동안 발생하는 스파이크의 빈도를 사용하여 정보를 전달하는 것과 달리, 그림 5처럼 첫 번째 스파이크가 발생하는 시간 정보를 이용해 인코딩하는 방식이다. 이는 입력 자극이 클수록 뉴런이 스파이크를 빠르게 발생시키는 원리에 기반한다. 첫 번째 스파이크가 발생하는 것에 초점을 두기 때문에 위 두 방식에 비해 가장 빠르고 높은 에너지 효율성을 보인다. 그러나 첫 번째 스파이크에 초점을 두므로 잡음에 취약할 수 있고, 구현이 복잡하다는 단점이 있다. 실제 환경에서는 안정적인 성능을 보장하기 어려울 수 있지만, 최근 TTFS 방식에 관한 연구에 따르면, TTFS 방식은 실시간 처리 또는 엣지 디바이스 등 빠른 처리 속도 또는 에너지 효율성이 중요한 상황에서 다른 인코딩 방식에 비해 적극적으로 고려된다고 한다<sup>[9]</sup>.

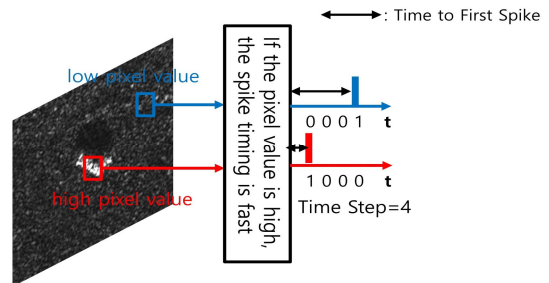


그림 5. TTFS 코딩 : 픽셀의 값이 클수록 스파이크가 더 빨리 생성되는 방식  
 Fig. 5. TTFS coding : how the larger the value of a pixel, the faster a spike is generated

이처럼 각 인코딩 방법은 고유의 장단점이 존재하고, 이는 본문과 표 1을 통해 확인할 수 있다.

표 1. 각 인코딩 기법의 장단점 및 특성

Table 1. Pros, Cons, and Characteristics of Each Encoding Method

	Direct	Rate	TTFS
장점 (Advantages)	높은 정확도(High accuracy), 구현하기 쉬움(Easy to code)	더 깊은 네트워크와 더 많은 데이터가 있을 때 좀 더 실용적(More useful when have large amounts of data and a deeper network), 노이즈에 견고(Robust to noise)	빠른 속도, 높은 에너지 효율성(Fastest and high energy efficiency)
단점 (Disadvantages)	완벽히 이산적인 연산만으로 구성되지 않음 (Not a completely discrete operation)	긴 지연시간(High latency), 낮은 에너지 효율(Low energy efficiency)	구현하기 어려움(Hard to code)
특성 (Characteristics)	부동 소수점 연산을 이용함 (Using floating-point operations)	푸아송 분포에서 샘플링하는 것과 유사 (Similar to sampling from Poisson distribution)	입력 자극이 클수록 Spike가 빠르게 생성됨(The larger the input stimulus, the faster the neuron fires spikes)

본 논문에서는 인코딩 방식에 따른 이미지 분류 성능을 알아보기 위해 동일한 조건 아래에서 앞서 살펴본 인코딩 방식을 적용하여 실험을 진행하였다. 각 인코딩 방식의 동작 과정은 그림 2, 3, 4를 통해 확인할 수 있다.

### 3. 데이터 증강 기법

본 논문에서는 단일 채널의 흑백 이미지로 합성 개구 레이다 영상 데이터인 MSTAR 데이터셋과 RGB 3채널의 컬러 이미지로 CIFAR-10 데이터셋을 사용하였다. 각 데이터

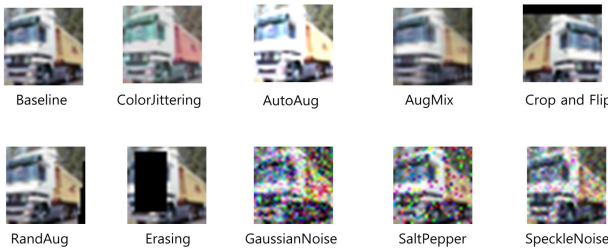


그림 6. Cifar 10 데이터셋에 적용된 데이터 증강 기법들에 대한 예시  
 Fig. 6. Data augmentations applied for CIFAR-10 dataset

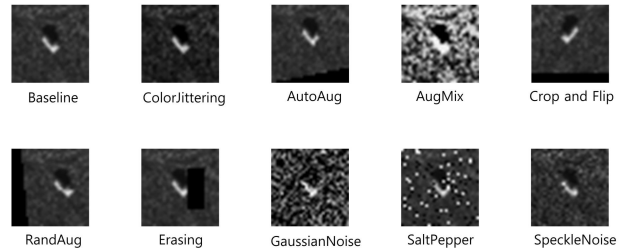


그림 7. MSTAR 데이터셋에 적용된 데이터 증강 기법들에 대한 예시  
 Fig. 7. Data augmentations applied for MSTAR dataset

셋에 대해 데이터 증강이 적용된 이미지는 그림 6, 7을 통해 확인할 수 있다.

### 4. 모델 구조

SNN에 각종 인코딩 기법을 적용해 실험 하기 위해 그림 8과 같이 모델을 구성하였다. 기존 DNN 모델인 VGG를 모방하여 5개의 계층만을 이용해 구성하였으며 입력을 스파이킹 신호로 바꿔주는 인코딩 계층을 먼저 통과한 뒤, 컨볼루션 함수와 LIF 모델이 결합된 3개의 계층과 입력의 크

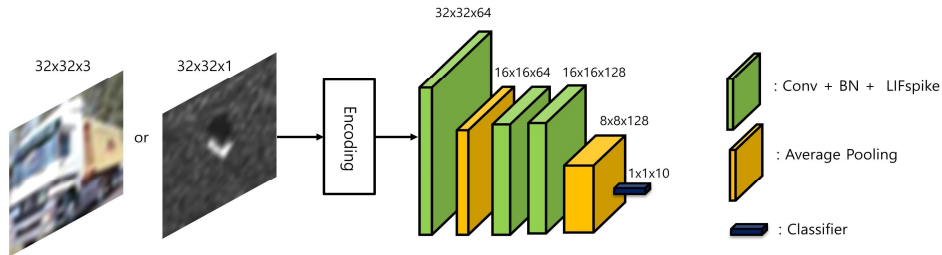


그림 8. 본 논문에서 제안한 SNN 모델 구조  
 Fig. 8. Proposed SNN Model Architecture

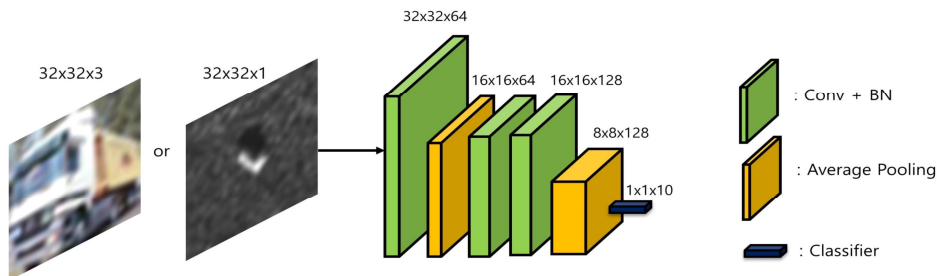


그림 9. 본 논문에서 제안한 ANN 모델 구조  
 Fig. 9. Proposed ANN Model Architecture

기를 감소시키는 2개의 Average 풀링 계층, 그리고 분류기를 통과하는 구조로 이루어져 있다.

DNN 모델은 SNN 모델과의 비교를 위해 그림 9와 같이 SNN 모델 고유의 인코딩 계층과 LIF 계층을 제외한 나머지 부분들은 전부 동일하게 구성하여 보다 공정한 비교가 가능하도록 하였다.

### III. 실험 및 결과

#### 1. 실험 환경

인코딩 방법과 데이터 증강에 따른 성능을 분석하기 위해 실험을 진행하였다. 실험은 Python 3.8, Pytorch 1.12.1, NVIDIA A5000 기반의 컴퓨터에서 진행되었다. 사용한 데이터셋은 MSTAR 데이터셋과 CIFAR-10, SVHN 데이터셋이다, 실험 시 모든 조건을 동일하게 하기 위해 데이터의 수가 더 적은 MSTAR 데이터셋을 기준으로 훈련용 데이터는 2,747개, 검증용 데이터는 2,425개로 설정하였고, 이미지 크기는 CIFAR-10 데이터셋의 이미지 크기인 32로 맞춰주었다. 시간 단계 T는 5로 고정하였으며 배치 크기는 32, 옵티마이저는 SGD를 사용하였다. 모델의 정확도는 전체 300 epoch 중 가장 높은 검증 정확도를 기준으로 하였다. 또한 실험을 진행할 때, Torch 라이브러리의 RandomApply 함수를 이용하여 원본 이미지와 증강 또는 잡음이 적용된 이미지를 확률적으로 입력되도록 해주었다. 실험에 사용한

표 2. 실험 환경

Table 2. Experiment Setting

	Cifar 10	MSTAR 10	SVHN
Model	VGG5	VGG5	VGG5
# of Train Data	2747	2747	2747
# of Test Data	2452	2452	2452
Image size	32	32	32
Batch Size	32	32	32
Num of Epoch	300	300	300
Learning Rate	0.01	0.01	0.01
Optimizer	SGD	SGD	SGD
Time Step	5	5	5
Augmentation Probability	50%	50%	50%

자세한 설정값은 표 2와 같다.

#### 2. 실험 결과 및 분석

##### 2.1 MSTAR 데이터셋 성능 분석

단일 채널의 흑백 이미지인 MSTAR 데이터셋에 대한 성능 비교를 진행하였다. 실험 결과는 표 3과 같다. 표 3을 보면, 대체로 높은 정확도를 보인다. 특히 Direct 코딩의 경우, 입력 이미지가 컨볼루션 계층을 통과하여 입력되기 때문에 다른 인코딩 기법들보다 상대적으로 높은 수치를 보이며, ANN에 근접한 성능을 보인다. Rate 코딩의 경우 앞서 서술된 단점인 긴 지연시간과 낮은 에너지 효율성으로 인해 다른 기법들보다 수렴하는 데 오랜 시간이 걸렸으며, 많은 데이터셋에서 강점을 보이는 기법이지만 반대로 작은 데이터셋에서는 부각을 드러내지 못함을 확인하였다. 실제로 Rate 코딩 방식은 SNN의 시간 단계 값을 크게 하고, 학습 시간을 늘릴수록 성능이 꾸준히 상승함을 알 수 있었다. TTFS 인코딩 기법의 경우 Direct 기법에 준하는 학습 속도와 성능을 나타냄을 알 수 있었다. 특히 각종 데이터 증강 기법과 잡음이 포함된 이미지에서 높은 성능 향상 폭을 보였다. 또한 매 실험 진행 시  $\pm 1\%$  정도의 오차가 존재하였

표 3. MSTAR 데이터셋 실험 결과

Table 3. Experiment Result of MSTAR dataset

	Direct	Rate	TTFS	ANN
No Aug	98.14	81.81	92.37	98.35
ColorJitter	99.26 (+1.12%)	83.27 (+1.46%)	96.04 (+3.67%)	98.72 (+0.37%)
AutoAug	99.46 (+1.32%)	82.68 (+0.87%)	94.27 (+2.1%)	99.38 (+1.03%)
AugMix	98.93 (+0.79%)	82.09 (+0.28%)	95.79 (+3.42%)	99.26 (+0.91%)
Crop and Flip	99.05 (+0.91%)	82.09 (+0.28%)	92.49 (+0.12%)	99.13 (0.78%)
RandAug	99.50 (+1.36%)	83.48 (+1.67%)	95.22 (+2.85%)	99.51 (+1.16%)
Erasing	98.56 (+0.42%)	82.31 (+0.5%)	92.66 (+0.29%)	99.01 (+0.61%)
Gaussian Noise	99.26 (+1.12%)	83.42 (+1.61%)	94.97 (+2.6%)	98.92 (+0.57%)
Speckle Noise	99.42 (+1.28%)	83.21 (+1.4%)	95.05 (+2.68%)	98.76 (+0.41%)
Salt Pepper Noise	99.21 (+1.07%)	82.92 (+1.11%)	92.87 (+0.5%)	99.34 (+0.99%)

고 이를 고려한다면 Direct 또는 TTFS 코딩 기법에서는 ANN과 근접한 성능이 보임을 알 수 있다. MSTAR 데이터셋과 같은 SAR 이미지는 인공위성을 통해 획득된다. 이미지가 획득되는 장소의 특수성과 더불어 ANN과 근소한 성능 차이를 가지는 SNN을 고려해 보았을 때, SAR 이미지가 생성되고 처리되는 특별한 환경에서는 SNN이 ANN을 대체할 여지가 충분하다고 보인다.

### 2.2 CIFAR-10 데이터셋 성능 분석

두 번째로, RGB 3채널의 컬러 이미지인 CIFAR-10 데이터셋에 대한 성능 비교를 진행하였다. 다만 학습 데이터 셋의 양을 MSTAR 데이터셋과 동일하게 맞춰주기 위해 기존 50,000개에서 2,747개로 대폭 감소하여 학습을 진행하였다. 그 결과 대부분의 시도에서 매우 낮은 정확도를 보였다. 그러나 ANN에서도 낮은 성능을 보이기 때문에 이는 온전히 SNN만의 문제라고 볼 수 없다. SNN과 ANN 모두 앞선 MSTAR 데이터셋에 대한 실험 결과에 비해 3채널의 컬러 이미지에 대해서는 낮은 성능을 보인다. 비록 낮은 정확도를 가지지만, 적용된 대부분의 데이터 증강에서도 SNN과 ANN 모두 성능이 증가하는 모습을 확인할 수 있다.

그러나 표 4에서 주목할 만한 점은, 컬러 이미지에서

표 4. CIFAR-10 데이터셋 실험 결과  
 Table 4. Experiment Result of CIFAR-10 dataset

	Direct	Rate	TTFS	ANN
No Aug	54.76	48.84	52.56	54.79
ColorJitter	63.55 (+8.79%)	54.41 (+5.57%)	53.88 (+1.32%)	56.90 (+2.11%)
AutoAug	59.04 (+4.28%)	52.04 (+3.2%)	55.70 (+2.14%)	58.68 (+3.89%)
AugMix	56.32 (+1.56%)	50.74 (+1.9%)	54.55 (+1.99%)	55.41 (+0.62%)
Crop and Flip	63.06 (+8.3%)	54.42 (+5.88%)	56.28 (+3.72%)	58.18 (3.39%)
RandAug	63.39 (+8.63%)	54.34 (+5.5%)	59.46 (+6.9%)	63.80 (+8.01%)
Erasing	58.67 (+3.89%)	50.79 (+1.95%)	53.80 (+1.24%)	57.48 (+2.69%)
Gaussian Noise	54.01 (-0.75%)	49.83 (+0.99%)	55.12 (+2.56%)	48.97 (-5.82%)
Speckle Noise	55.70 (+0.94%)	49.75 (+0.91%)	53.51 (+0.95%)	53.22 (-1.57%)
Salt Pepper Noise	55.25 (+0.49%)	50.33 (+1.59%)	55.37 (+2.81%)	50.37 (-4.42%)

잡음이 추가된 경우 ANN은 성능이 감소한 것에 비해, SNN은 성능이 증가하는 점이다. SNN의 경우 입력 이미지를 스파이크 신호로 변환하는 과정에서 잡음이 스무딩되고 일반화 성능이 증가하여 정확도가 향상했다고 볼 수 있다.

잡음 관련하여 한 가지 더 이목을 끄는 점은, Direct나 Rate 인코딩 방식에 비해 TTFS 인코딩 방식이 높은 성능 증가 폭을 보이는 점이다. 특히 Direct 코딩 방식은 가우시안 잡음에서 성능이 감소하는 모습을 보였는데, TTFS 코딩은 2.56%P 상승하였다. 소금 후추 잡음에서도 2.81%P라는 높은 향상 폭을 보인다.

### 2.3 SVHN 데이터셋 성능 분석

세 번째로, RGB 3채널의 컬러 이미지면서 복잡한 개체가 아닌 숫자를 분류하는 SVHN 데이터셋에 대한 성능 비교를 진행하였다. 위 2개의 다른 데이터셋과 마찬가지로 모든 실험 환경을 동일하게 설정하였다. CIFAR-10 데이터셋과 동일한 컬러 이미지지만, 분류 작업의 난이도가 더 쉬워 전반적으로 높은 정확도를 보인다. 전체 실험 결과는 표 5를 통해 확인할 수 있다.

표 5. SVHN 데이터셋 실험 결과  
 Table 5. Experiment Result of SVHN dataset

	Direct	Rate	TTFS	ANN
No Aug	78.88	66.94	69.21	76.32
ColorJitter	81.24 (+2.36%)	72.52 (+5.58%)	74.91 (+5.70%)	77.35 (+1.03%)
AutoAug	87.31 (+8.43%)	73.64 (+6.70%)	76.49 (+7.82%)	86.36 (+10.04%)
AugMix	80.99 (+1.11%)	69.50 (+2.56%)	72.36 (+3.15%)	79.50 (+3.18%)
Crop and Flip	81.74 (+2.88%)	69.59 (+2.65%)	69.26 (+0.05%)	76.77 (+0.45%)
RandAug	86.40 (+7.52%)	74.26 (+7.32%)	76.36 (+7.15%)	84.42 (+8.10%)
Erasing	76.32 (-2.56%)	66.90 (-0.04%)	66.94 (+0.73%)	75.24 (-1.08%)
Gaussian Noise	75.95 (-2.93%)	67.98 (+1.04%)	72.36 (+3.15%)	74.38 (-1.94%)
Speckle Noise	75.09 (-3.79%)	68.60 (+1.66%)	71.57 (+2.36%)	76.11 (-0.21%)
Salt Pepper Noise	77.39 (-0.49%)	68.35 (+1.41%)	70.29 (+1.08%)	74.75 (-1.57%)

CIFAR-10과 같은 3채널의 컬러 이미지지만, 객체를 예측하는 것보다 더 쉬운 숫자를 예측하는 작업이어서 전반

적으로 높은 정확도를 보여준다. 특히 주목할 만한 점은 잡음을 이용한 증강 이외에 다른 데이터 증강 방식에서 ANN 과 여타 다른 인코딩 방법을 적용한 모델들보다 Direct 방법을 적용한 모델의 정확도가 제일 높다는 점, 그리고 Erasing과 잡음을 적용했을 때 입력을 이산적인 스파이크 신호로 변환하여 사용하는 Rate 코딩 방식과 TTFS 코딩 방식의 성능이 다른 방식에 비해 더욱 강인함을 확인할 수 있다.

위 실험은 데이터의 성질만 다르고 이외의 것은 모두 동일한 환경에서 진행되었다. 위 실험들로 미루어 볼 때, 실생활에 가까운 3채널의 컬러 이미지를 분류하는 상황에서 잡음이 많은 환경이라면 잡음과 Occlusion에 강인한 SNN이 충분히 ANN을 대체할 수 있는 부분이라 생각한다.

또한 표 6을 통해 모델 별 파라미터 수와 크기, 메모리 사용량, 초당 부동소수점 연산횟수(FLOPs, Floating point operations per second)를 살펴보았을 때, ANN에 비해 SNN 모델들의 파라미터 수가 매우 적고 모델의 크기 역시 매우 작은 것을 알 수 있다. 메모리 사용량 역시 소폭이지만 더 적은 것을 알 수 있다. 다만 ANN에 비해 인코딩이라는 추가적인 연산과 Time Step이라는 추가적인 차원이 들어가기 때문에, 모델의 FLOPs와 추론 시간에서는 뒤지는 모습을 보인다. 그러나 모델의 파라미터 수와 크기에서 알 수 있듯, ANN을 사용할 수 없는 저전력의 하드웨어 환경 또는 전력 공급이 원활하지 않은 상황에서 인공신경망에 준하는 기술에 대한 요구가 수반될 때, 그 해결책으로 SNN이 기용될 수 있는 신호로 보인다.

#### IV. 결 론

본 논문에서는 SNN에서의 다양한 인코딩 방법과 데이터 증강을 이용해 ANN과 SNN에서의 성능을 분석하였다. 각 인코딩 방법이 가지는 장단점을 파악하고, 실험을 통해 SNN이 차세대 인공신경망으로써 DNN에 준하는 성능을 가질 수 있는지 확인해 보았다. 이에 SAR 이미지와 컬러 이미지에 대해 동일 조건에서 비교 실험을 진행하였다. SNN은 ANN에 비해 훨씬 적은 파라미터와 작은 모델 크기를 가지며, SAR 이미지에 대해서는 ANN에 준하는 정확도를 보이고 컬러 이미지에서는 잡음에 강인한 모습을 보인다. 계층 5개뿐인 얇은 모델과 3,000개가 미처 되지 못하는 소량의 데이터셋에서도 ANN과 유사한 성능을 보이는 것으로 미루어 보아, SNN 모델 구조에 대한 연구가 ANN만큼 왕성해지고 위 실험에 사용한 데이터보다 더 많은 양의 데이터를 사용했을 때 과연 SNN과 ANN 간 어떠한 결과가 나올지 매우 기대되는 바이다. 대개 더 큰 모델과 더 많은 데이터셋의 조합은 모델 성능 향상으로 이어지는 것은 앞서 진행된 많은 연구를 통해 알려진 사실이지만, SNN은 아직 초기 단계이고 SNN에 맞는 완벽한 학습 방법이 부재하므로 ANN을 따라가기 위해 노력하면서도 SNN만의 학습 방법을 찾기 위한 연구가 수반되어야 한다고 생각한다. 본 논문에서 진행된 실험 결과는 하드웨어 또는 전력, 잡음 등에 관련된 여러 가지 문제로 ANN을 활용하지 못하는 특수한 상황이 있다면, SNN이 ANN을 대체할 여러 근거를 긍정적으로 제시한다.

표 6. 모델 별 파라미터 수, FLOPs 비교  
Table 6. Comparison of parameters, FLOPs by model

	Direct	Rate	TTFS	ANN
Num of parameters	305,892	305,892	305,892	8,697,482
Model size (MB)	1.17	1.17	1.17	33.18
Memory usage (MB)	35.79	35.77	35.77	39.35
M FLOPs	590.40	590.23	590.23	118
Inference time (second)	0.0031	0.0034	0.0036	0.0008

#### 참 고 문 헌 (References)

- [1] Gopalakrishnan Srinibasan et al, "Training Deep Spiking Neural Networks for Energy-Efficient Neuromorphic Computing," International Conference on Acoustics, Speech, and Signal Processing (ICASSP), May 2020.  
doi: <https://doi.org/10.1109/ICASSP40776.2020.9053914>
- [2] Wenzhe Guo et al, "Neural Coding in Spiking Neural Networks: A Comparative Study for Robust Neuromorphic Systems," Frontiers, Vol.15, March 2021.  
doi: <https://doi.org/10.3389/fnins.2021.638474>
- [3] H. Markram et al, "Spike-timing-dependent plasticity: a comprehensive overview," Frontiers, July 2012.



- doi: <https://doi.org/10.3389/fnsyn.2012.00002>
- [4] Emre O. Neftci et al, "Surrogate Gradient Learning in Spiking Neural Networks," IEEE Signal Processing Magazine, Vol.36, pp.51-63, November 2019.  
doi: <https://doi.org/10.1109/MSP.2019.2931595>
- [5] Yujie Wu et al, "Direct Training for Spiking Neural Networks: Faster, Larger, Better," AAAI Conference on Artificial Intelligence, Vol.33, pp.1311 - 1318, 2019.  
doi: <https://doi.org/10.1609/aaai.v33i01.33011311>
- [6] Youngeun Kim et al, "Rate Coding or Direct Coding: Which One is Better for Accurate, Robust, and Energy-efficient Spiking Neural Networks?," International Conference on Acoustics, Speech, and Signal Processing (ICASSP), April 2022.  
doi: <https://doi.org/10.1109/ICASSP43922.2022.9747906>
- [7] Xinyi Chen, Qu Yang, Jibin Wu, Haizhou Li, Kay Chen Tan, "A Hybrid Neural Coding Approach for Pattern Recognition with Spiking Neural Networks," IEEE Transactions on Pattern Analysis and Machine Intelligence(TPAMI) Vol. 46, pp.3064 - 3078, December 2023.  
doi: <https://doi.org/10.1109/TPAMI.2023.3339211>
- [8] Tong Bu, Wei Fang, Jianhao Ding, PENGLIN DAI, Zhaofei Yu, Tiejun Huang, "Optimal ANN-SNN Conversion for High-accuracy and Ultra-low-latency Spiking Neural Networks," The International Conference on Learning Representations(ICLR), January 2022.  
doi: <https://doi.org/10.48550/arXiv.2303.04347>
- [9] Fanfan Li et al, "An artificial visual neuron with multiplexed rate and time-to-first-spike coding," Nature Communications, Vol.15, May 2024.  
doi: <https://doi.org/10.1038/s41467-024-48103-9>

---

## 저 자 소 개



### 오 범 석

- 2018년 ~ 2024년 : 한국항공대학교 소프트웨어학과 학사
- 2024년 ~ 현재 : 한국항공대학교 대학원 인공지능학과 석사과정
- ORCID : <https://orcid.org/0009-0003-6125-0214>
- 주관심분야 : 스파이킹 뉴럴 네트워크, 이미지 분류, 준지도학습, 비지도학습



### 김 선 옥

- 2009년 ~ 2014년 : 연세대학교 전기전자공학과 학사
- 2014년 ~ 2019년 : 연세대학교 전기전자공학과 박사
- 2019년 ~ 2021년 : 연세대학교 박사후연구원
- 2021년 ~ 현재 : 한국항공대학교 인공지능학과 조교수
- ORCID : <https://orcid.org/0000-0002-9665-4214>
- 주관심분야 : 컴퓨터비전, 인공지능, 3차원 영상처리