

멀티모달 스파이킹 뉴럴 네트워크의 성능 분석

한혜림^{1,2}, 조정희², 박종길², 박성식²

서울과학기술대학교 전자 IT 미디어공학과¹, 한국과학기술연구원 인공지능융합연구단²
e-mail: *hanhyerim@seoultech.ac.kr*, *{jh.jo, jongkil, seong.sik.park}@kist.re.kr*

Performance Analysis of Multi-Modal Spiking Neural Networks

Hyerim Han^{1,2}, Jeonghee Jo², Jongkil Park², Seongsik Park²

Department of Electronic and IT Media Engineering, Seoul National University of
Science and Technology¹

Center for Neuromorphic Engineering, Korea Institute of Science and Technology²

Abstract

Multimodal models have been studied as promising model that can overcome the disadvantages of the unimodal models. However, not much research has been conducted on the effectiveness of multimodality in spiking neural networks (SNNs), which have been considered a next-generation artificial neural network for their energy efficiency. Thus, in this paper, we analyzed the effectiveness through experiments on modality, model size, and noise. According to our analysis, we validated that SNNs showed greater effectiveness in multimodality than DNNs.

I. 서론

최근 ChatGPT [1]의 등장으로 인공지능은 1956년 세상에 등장한 이래로 가장 큰 관심을 받고 있다. 이는 이미지(Image) 처리에 특화된 Convolutional Neural Networks(CNNs), 연속적인 데이터 처리에 주로 사용되는 Recurrent Neural Networks(RNNs) 등 다양한 딥러닝(Deep Learning) 구조의 등장과 함께 발전해왔다.

그러나 고성능의 딥러닝은 필연적으로 많은 컴퓨팅

자원과 데이터를 요구했다. 자연스레 엄청난 전력의 소모를 가져왔고, 이를 극복하기 위한 대안으로 대두되는 것이 바로 뇌의 동작을 모방한 spiking neural network (SNN)이다.

3세대 신경망이라고도 불리는 SNN은 모든 뉴런의 계산이 동시에 진행되는 기존 deep neural network (DNN)과 달리 스파이크가 발생하는 뉴런만 계산하므로 에너지 효율이 높다. 이처럼 SNN은 DNN에 비해 저전력으로 정보를 처리하는 데 적합하다는 점이 알려지면서 주목받게 되었다 [2]. 하지만 비연속적으로 동작하는 특성 때문에 스파이크 함수를 미분할 수 없다는 단점이 있다. 따라서 기존 DNN에서 높은 성능을 보여주었던 경사 하강법(gradient descent) 기반 알고리즘을 적용하기 어려워 DNN에 필적하는 성능을 보여주지 못하고 있다. 이를 해결하기 위해 spike-timing dependent-plasticity (STDP)와 같은 SNN 학습 알고리즘이 지속적으로 연구되고 있다 [3]. 따라서 본 논문에서는 멀티모달을 활용하여 기존 SNN의 한계를 극복하고 DNN에 필적하는 성능을 보여줄 수 있는지에 대해 연구하려 한다.

II. 본론

2.1 Spiking Neural Networks

SNN은 사람 뇌의 정보 처리 메커니즘을 모방한 차세대 인공지능망이다. 사람의 뇌는 뉴런(neuron)과 시

냅스(synapse)로 구성되어 있으며, 뉴런들은 스파이크(spike)를 생성한다. 생성된 스파이크는 시냅스 전 뉴런(pre-synaptic neuron)에서 시냅스 후 뉴런(post-synaptic neuron)으로 시냅스를 통해 전달된다.

뉴런은 다른 뉴런에서 생성된 스파이크들을 전달받았을 때, 스파이크는 각 시냅스의 가중치(synaptic weight)를 곱한 뒤 뉴런 내부의 막전위(membrane potential)에서 통합된다. 만약 막전위가 임계 전압(threshold voltage)을 넘었을 경우, 스파이크를 생성한다. 생성된 스파이크는 다른 뉴런으로 전송되고, 막전위는 휴지 전위로 초기화된다 [2]. 이를 리셋 메커니즘(reset mechanism)이라 하고, 하드 리셋(hard reset)과 소프트 리셋(soft reset)이 가장 일반적으로 사용된다.

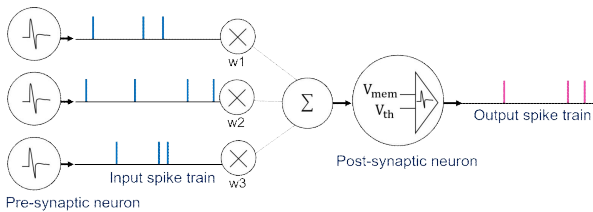


그림 1. SNN computational models [4]

막전위의 행동 양상을 결정하는 뉴런 모델 중 가장 많이 사용되는 모델은 바로 IF(Integrate and Fire) Model과 LIF(Leaky Integrate and Fire) Model이다 [5]. 본 논문에서는 surrogate gradient를 이용해 SNN을 학습했다.

2.2 Multimodal Model

모달리티는 다양한 감각기관을 통해 어떤 일을 경험하는 방식을 의미한다. 이러한 모달리티가 여러 개를 포함할 때 멀티모달이라 한다 [6]. 멀티모달은 학습할 때 각 모달리티에서 상호 보완하는 정보를 추출해 훈련하면 유니모달에 비해 훨씬 향상된 성능을 얻을 수 있다. 따라서 멀티모달 모델은 유니모달 모델의 한계를 극복하고, 더 우수한 성능을 내기 위한 방법 중 하나이다 [6].

이러한 멀티모달 모델은 보다 좋은 성능을 위해 고려해야 하는 5가지 요소가 있다. Representation은 멀티모달 데이터를 표현하고 요약하는 방식을 뜻한다. 즉, 여러 모달리티의 정보가 합쳐져 만들어진 종합적인 표현이다. Translation은 하나의 모달리티를 다른 모달리티로 변환하는 방법을 말한다. Alignment는 둘 이상의 모달리티 사이 요소 간의 관계를 식별하는 것이다. Fusion은 둘 이상의 모달리티의 정보를 결합하는 것이다. 마지막으로 Co-learning은 서로 다른 모달

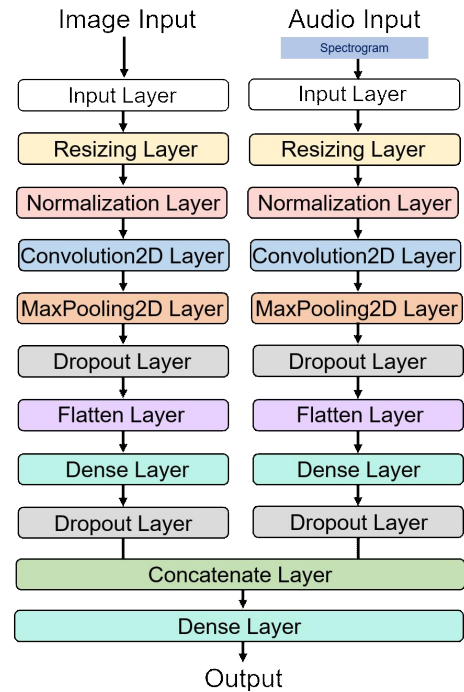


그림 2. 멀티모달 DNN과 SNN 구조

리티가 상호작용하여 함께 학습하는 방법이다 [6].

본 논문에서는 Fusion 중에서도 Early Fusion을 선택하여 각 모달리티가 분류되기 전 통합했다. 이는 단일 모델의 학습만 하면 돼 편리하다 [6]. 본 논문의 기본적인 멀티모달 SNN의 구조는 그림 2와 같다.

III. 구현

3.1 실험환경

환경이 변함에 따라 멀티모달 SNN가 어떤 성능을 내는지 알아보기 위해 총 6번의 실험을 진행했다. 모든 실험은 batch size 100, epoch 200을 기준으로 진행했다. 또한 정확도는 epoch 200 중 가장 높은 검증 정확도를 기준으로 표기했다. S는 기준이 되는 모델 크기로, 모델의 파라미터 수를 뜻한다. S는 약 1.6M이다. S/2는 S의 2분의 1인 약 0.8M이고, S/3은 S의 3분의 1인 약 0.5M이다.

연구에서 사용된 데이터는 MNIST 이미지 데이터셋과 Spoken_Digit 오디오 데이터셋[7]이다. 둘 다 숫자 0부터 9까지 총 10개의 class로 나뉘어 있다. 이미지는 훈련 데이터가 60,000개, 테스트 데이터가 10,000개이고, 오디오는 훈련 데이터가 2,700개, 테스트 데이터가 300개이다. 단, 멀티모달로 훈련할 때는 오디오 데이터의 반복을 통해 60,000개, 10,000개로 이미지에 맞춰 진행했다. 또한 오디오의 특징 추출은 short-time Fourier transform (STFT)을 통해 spectrogram으로

변환해 사용했다. DNN과 SNN 학습 시 이미지와 멀티모달은 validation dataset을 train dataset의 20%로, 오디오는 약 18.5%로 진행하였다. 이외의 설정값은 표 1과 같다.

표 1. 실험조건

실험	실험조건
모델의 크기(파라미터 수)	S(약 1,600,000), S/2, S/3
가우시안 노이즈의 표준 편차 크기	0.1, 5
가우시안 노이즈의 위치	Activation Function 앞, 뒤
리셋 메커니즘 (Reset Mechanism)	Hard Reset, Soft Reset
뉴런 모델(Neuron Model)	LIF, IF
타임 스텝(Time step)	4, 8, 16

표 2. SNN 학습 시 설정값

설정	값
Model Size	약 1,600,000
Time Step	4
Reset Mechanism	Soft Reset
Neuron Model	LIF model

3.2 실험결과

(1) 모델 크기에 따른 성능 비교

멀티모달의 효율성을 알아보기 위해 모델 크기를 바꿔가며 DNN과 SNN 검증 결과를 확인했다. 그림 3에서 볼 수 있듯이, DNN에서 멀티모달은 모델의 크기가 줄어들어도 99%의 정확도를 유지했다. 반면 유니모달은 모델의 크기가 작아짐에 따라 최대 99%에서 최소 95%대까지 변동하는 결과를 보였다.

그림 4를 확인하면, SNN에서 오디오는 모델 크기가 작아져도 오차 범위 2% 내에서 성능을 유지하고 있다. 멀티모달 또한 일관적으로 99.9%의 정확도를 유지한다. 반면 이미지는 모델 크기가 작아짐에 따라 정확도가 크게 낮아지고 있다. 또한 멀티모달이 유니모달보다 높은 정확도를 내는 것은 물론, 멀티모달은 모델 크기에 따른 차이가 0.1% 이내로 작았다. 따라서 작은 모델로 높은 성능을 낼 수 있는 멀티모달의 효율성을 확인했다.

유니모달과 멀티모달의 성능을 직관적으로 확인하기 위해 두 유니모달 모델의 평균 검증 정확도와 멀티모달의 검증 정확도를 비교했다. 그림 5를 통해 모델 크기가 S일 때 DNN과 SNN에서 유니모달보다 멀티모달이 월등히 높은 성능을 내고 있음을 확인할 수 있다.

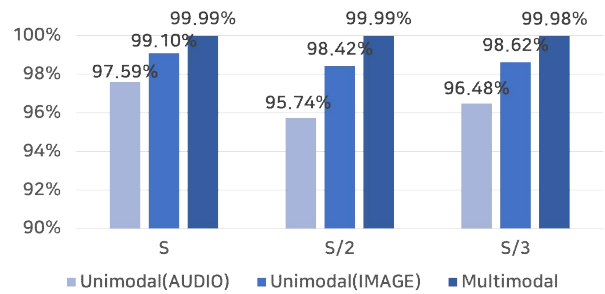


그림 3. 모델 크기에 따른 DNN 성능

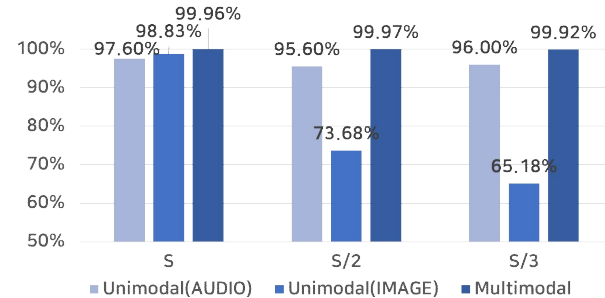


그림 4. 모델 크기에 따른 SNN 성능 비교

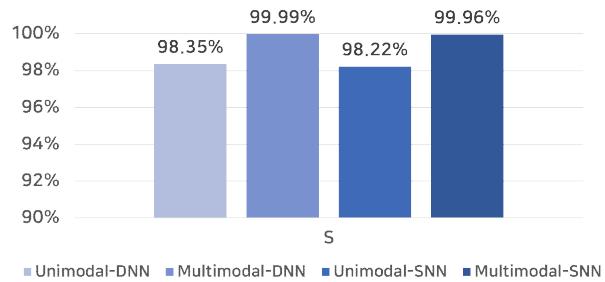


그림 5. 모델 크기 S일 때 DNN과 SNN 성능 비교

(2) 모델의 노이즈 크기에 따른 성능 비교

본 논문에서는 평균이 0인 가우시안 분포 $N(0, \sigma^2)$ 를 갖는 가우시안 노이즈(Gaussian Noise)를 사용했다. 학습에는 노이즈를 배제하고, 검증 시에만 노이즈를 적용했다. 실험에서 노이즈 크기, 즉 표준 편차의 값은 큰 값인 5와 작은 값인 0.1로 진행했다.

그림 6을 확인하면, 노이즈의 표준 편차가 0.1일 때 DNN에서 멀티모달은 모델 크기가 줄어들어도 99.9%를 유지했다. 또한 오디오-이미지-멀티모달 순으로 정확도가 높다. 그림 7의 멀티모달은 SNN에서도 여전히 99.9%를 유지한다. 하지만 이미지는 셋 중 가장 정확도가 떨어질 뿐만 아니라, 그림 4와 비교했을 때 노이즈가 추가되면 S에서는 약 4%, S/3에서는 25% 가량 정확도가 떨어진다. 이미지는 노이즈 유무에 따른 차이가 가장 크며, 노이즈에 취약한 경향을 보인다.

그림 8에서는 노이즈의 표준 편차가 0.1과 5로 차이 날 때 DNN과 SNN의 성능을 비교했다. 크기가 S

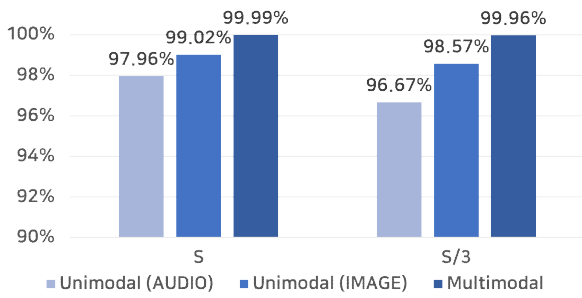


그림 6. 모델 크기에 따른 Noise-DNN 성능 비교 (Gaussian Noise $N(0, 0.1^2)$)

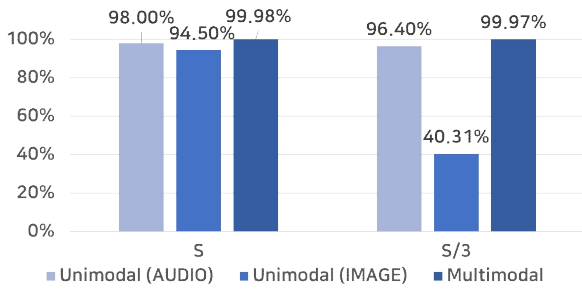


그림 7. 모델 크기에 따른 Noise-SNN 성능 비교 (Gaussian Noise $N(0, 0.1^2)$)

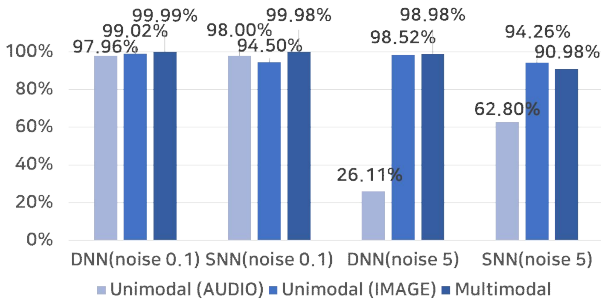


그림 8. Noise 크기에 따른 DNN과 SNN 성능 비교

일 때, 유니모달의 경우 SNN이 DNN보다 더 노이즈에 강건함을 확인했다. 오디오는 노이즈가 커졌을 때 DNN에서 약 70%의 성능이 하락하지만, SNN에서는 하락의 폭이 약 35%로 훨씬 작다. 또한 이미지도 DNN에서 정확도가 0.5% 하락했지만, SNN에서는 0.24%라는 적은 폭으로 하락했다. 멀티모달의 경우, DNN에서 노이즈가 증가해도 약 99%를 유지했다. 하지만 SNN에서는 9% 하락했다. 즉 노이즈가 있는 상황에서 멀티모달은 DNN, 유니모달은 SNN에서 더 좋은 성능을 보였다.

(3) 모델의 노이즈 위치에 따른 성능 비교

노이즈 크기가 아닌 위치는 결과에 어떤 영향을 주는 지 확인하기 위해 실험했다. activation 함수를 기준으로, 전에 노이즈를 배치하여 pre-synaptic neuron에

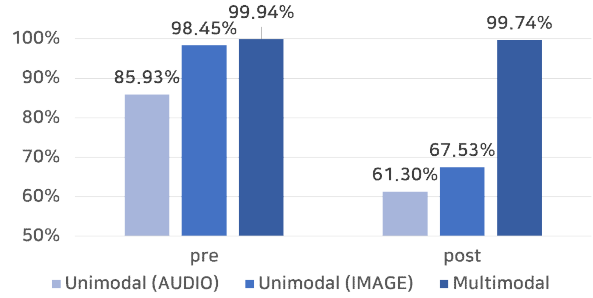


그림 9. Noise 위치에 따른 DNN 성능 비교

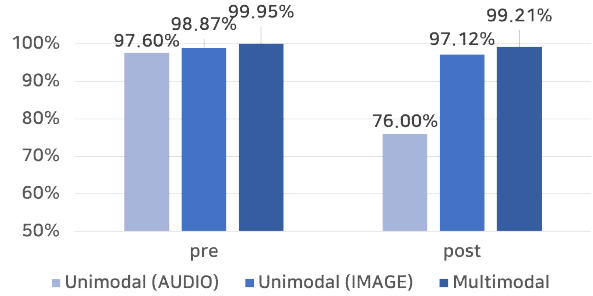


그림 10. Noise 위치에 따른 SNN 성능 비교

영향을 주었을 때와 후에 배치하여 post-synaptic neuron에 영향을 주었을 때의 결과를 비교했다. 모델 크기는 S, 노이즈 크기는 0.1로 통일했다.

그림 9에서 볼 수 있듯이, DNN에서는 activation 함수 이전에 노이즈를 위치했을 때 모든 모델이 좋은 성능을 냈다. 또한 그림 10을 확인하면, SNN에서도 모든 모델이 노이즈가 pre-synaptic neuron에 위치할 때 더 좋은 결과를 나타냈다. 이를 통해 pre-synaptic neuron이 노이즈에 강인함을 확인했다.

(4) 모델의 리셋 메커니즘에 따른 성능 비교

리셋 메커니즘이 SNN의 성능에 어떠한 영향을 미치는지 확인하기 위해 실험을 진행했다. 그림 11에서 두 유니모달 모델 모두 Hard Reset보다 Soft Reset에서 더 높은 정확도를 보였다. 특히 오디오는 10% 이상의 차이로 Soft Reset이 훨씬 높은 정확도를 나타냈다. 멀티모달은 약 99%로 리셋 메커니즘에 따른 큰 차이는 없지만, Soft Reset의 정확도가 더 높았다.

(5) 모델의 뉴런 모델에 따른 성능 비교

막전위의 행동 양상을 결정하는 뉴런 모델이 SNN의 성능에 미치는 영향을 알아보기 위해 뉴런 모델에 따른 검증 정확도를 비교했다. 그림 12에서 둘 다 오디오-이미지-멀티모달 순으로 좋은 성능을 냈다. 전체적인 정확도는 모두 오차 범위 0.2% 내로, LIF와 IF의 결과는 비슷한 양상을 나타냈다. 따라서 뉴런 모델이 유니모달과 멀티모달의 SNN 성능에 큰 영향을 미치지 않는 것을 확인했다.

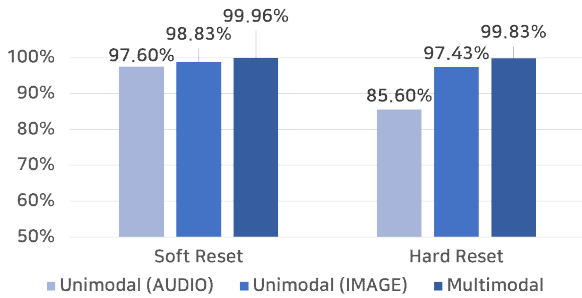


그림 11. Reset Mechanism에 따른 SNN 성능 비교

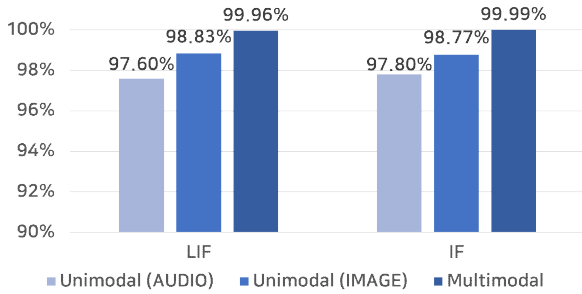


그림 12. Neuron Model에 따른 SNN 성능 비교

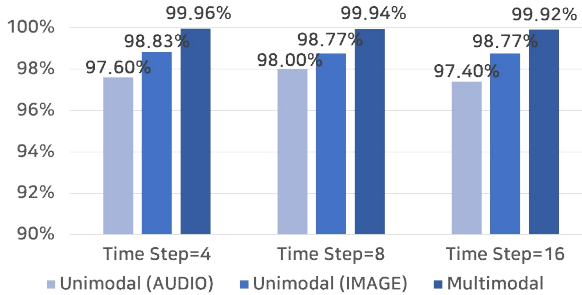


그림 13. Time Step에 따른 SNN 성능 비교

(6) 모델의 타임 스텝에 따른 성능 비교

타임 스텝이 SNN의 성능에 어떠한 영향을 미치는지 알아보기 위해 실험을 진행했다. 그림 13에서 타임 스텝에 따른 결과는 모델별 오차 범위 0.6% 내로 모두 비슷하게 나타났다. 이를 통해 모델이 충분한 타임 스텝으로 실험되었음을 확인했다.

IV. 결론 및 향후 연구 방향

본 논문에서는 다양한 실험환경 속에서 유니모달과 멀티모달의 SNN과 DNN 성능을 분석했다. 모델 크기가 작아져도, 멀티모달은 SNN과 DNN 모두 좋은 성능을 냈다. 또한 노이즈를 추가했을 때 유니모달은 SNN, 멀티모달은 DNN에서 높은 정확도를 유지했다. 모달리티 모두 노이즈의 위치가 activation 함수 이전 일 때 더 잘 버티는 것을 확인했다. 리셋 메커니즘은 전체적으로 하드 리셋보다 소프트 리셋에서 더 좋은

성능을 보였다. 멀티모달은 약 99% 정확도로 큰 차이는 없지만, 소프트 리셋일 때 더 높았다.

멀티모달 SNN은 노이즈가 5일 때를 제외하고 노이즈 위치나, 리셋 메커니즘, 뉴런 모델, 타임 스텝의 변화에도 높은 정확도를 유지했다. SNN과 멀티모달을 단독으로 각각 사용할 때보다 멀티모달 SNN으로 함께 사용할 때 더 좋은 결과를 얻을 수 있음을 확인했다.

다만 본 연구에서는 비교적 간단한 MNIST 데이터를 사용했고, 멀티모달 데이터를 fusion하는 데 있어서 early fusion 중 concatenation만을 실험했다는 한계가 있다. 또한 SNN에서 이미지 유니모달 모델은 모델 크기가 작아질 때마다 정확도가 크게 떨어지는 현상이 발생했다. 이는 모델 크기를 제한하여 생긴 구조적인 문제로 인한 현상으로 추정된다.

후속 연구에서는 이런 점을 고려해 다양한 데이터로 실험할 예정이다. 또한 fusion 방식에 있어서 late fusion이나 hybrid fusion도 적용해 비교해볼 예정이다.

Acknowledgements

이 연구는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원 (NRF-2021R1C1C2010454)과, 한국과학기술연구원(KIST) 주요 사업인 '인공뇌융합연구사업(2E32260)'의 지원을 받아 수행되었음

참고문헌

- [1] Andreas Holzinger et al. "AI for life: Trends in artificial intelligence for biotechnology", New Biotechnology, 5 February 2023.
- [2] Gopalakrishnan Srinivasan et al. "Training Deep Spiking Neural Networks for Energy-Efficient Neuromorphic Computing", IEEE, 09 April 2020.
- [3] 도우형 외, "깊은 스파이킹 신경망 학습 방법의 최신 연구 동향", 2020년 대한전자공학회 하계학술대회 논문집, pp 1,829-1,834, 2020.
- [4] Kaushik Roy et al. "Towards spike-based machine intelligence with neuromorphic computing", Nature volume 575, pages 607 - 617, 2019.
- [5] 홍윤표 외, "LIF/IF Model을 활용한 SNN Neuron의 H/W 구현", 2021년도 대한전자공학회 하계종합학술대회 논문집, pp. 2,293-2,295, 2021.
- [6] Tadas Baltrušaitis et al. "Multimodal Machine Learning: A Survey and Taxonomy", IEEE, 25 January 2018.
- [7] www.tensorflow.org/datasets/catalog/spoken_digit