2 뉴로모픽 반도체를 활용한 동작 인식 어플리케이션

글_남기령 석사과정 · 박종길 선임연구원 | 한국과학기술연구원

1. 서론

최근 인공지능은 COVID-19 팬데믹으로 인해 본격화된 언택트 시대로 인해 그 발전이 더욱 가속화되고 있다. 우리의 생활 전반적인 영역은 비대면, 비접촉이 일상이 되었고, 이는 인공지능 기술의 발전을 더욱 요구하고 있다. 하지만, 무어의 법칙이 한계에 도달하며 기존 인공지능 하드웨어 설계 기술의 확장으로 얻을 수 있는 성능 향상과 전력 효율 개선은 한계에 이르고 있다. 또한 현재 사용하는 인공지능 컴퓨팅 방식은 실제 인간의 뇌가 학습하는 방법과 근본적으로 다르기 때문에 학습 과정에서 대규모의 데이터를 순차적으로 연산 처리하기 위해서는 많은 시간과 막대한 전력이 소모되는 등 비효율적인 요소들이 존재한다.

이를 해결하기 위해 최근 인간의 뇌 동작 방식과 구조를 모방하여 학습하는 뉴로모픽 하드웨어와 이를 적용한 컴퓨팅 방식인 스파이킹 신경망 (spiking neural networks, SNN)이 향후 차세대 컴퓨팅 개발을 위한 촉매제로 떠오르고 있다. 최근에는 뉴로모픽 기술을 활용한 객체 인식 [1,2], 음성 처리 [3-5], 패턴 인식 [6-9] 등 스파이킹 신경망을 기반으로 한 다양한 어플리케이션 관련 연구도 진행 중이다. 본 논문에서는 그 중 스파이킹 신경망의 시공간적 특성을 적합하게 반영한 이벤트 기반 동작 인식에 관해서 더 자세히 살펴볼 것이다. 그에 앞서 먼저 2장에서는 스파이킹 신경망과 스파이킹 뉴런 모델의 종류에 대해 간단히 살펴본 후 3장에서 스파이킹 신경망 학습에 적합하게 설계된 이벤트 기반 센서와 이벤트 기반 센서로 만든 벤치마크 데이터셋에 대해서 간단히 알아보고자 한다 [10]. 4장에서는 이를 활용하

여 구현한 시공간적 정보를 가진 동작을 인식하는 어플리케이션의 다양한 구현 예에 대해 살펴볼 것이다 [10-12]. 마지막으로 5장에서 뉴로모픽 기술 활용의 향후 전망에 대해 이야기하며 마무리한다.

2. 뉴로모픽 관련 기술 설명

2.1 스파이킹 신경망

스파이킹 신경망은 실제 생물학적으로 인간의 뇌가 동작하는 방법에서 영감을 받아 인공 신경망을 구현한 알고리즘을 말한다. 기존의 심층 신경망과 가장 큰 차이점 중 하나는 뉴런 간의 정보가 전달되는 방법이다. 기존의 심층 신경망은 뉴런으로 들어온 입력 정보가 활성화 함수를 거쳐 나온 실수(real number) 출력 값이 다음 뉴런으로 전달되며 정보가 처리된다. 반면에 스파이킹 신경망은 입력된 정보가 시공간적 정보를 가진 이산적인 스파이크로 인코딩 되어 처리된다. 이러한 특징으로 인해 스파이킹 신경망은 생물학적 시스템을 보다 유사하게 모사하여 동작한다. 또한 스파이킹 신경망 기반 뉴로모픽 컴퓨팅은 비동기식 스파이크 동작 메커니즘에 의하여 데이터 처리의 병렬성을 높이고 처리해야 하는 데이터의 양을 줄임으로써 적은 전력 소모, 높은 에너지 효율, 낮은 지연시간과 같은 장점을 가지고 있다.

2.2 스파이킹 뉴런 모델

스파이킹 신경망에서 일반적으로 사용되는 스파이킹 뉴런 모델은 Integrate-and-Fire (IF) 모델 [13,14], Leaky Integrate-and-Fire (LIF) 모델 [15], Hodgkin-Huxley 모델 [16] 그리고 Spike Response Model (SRM) [17] 등이 있다. 이 중 대표적으로 생물학적 타당성을 유지하면서 대규모 하드웨어에서 구현이 용이한 LIF 모델이 가장 보편적으로 사용되고 있다.

LIF 모델 뉴런에서 프리 시냅틱 뉴런의 스파이크가 입력되면 해당 시냅스의 연결 강도를 나타내는 가중치만큼 포스트 시냅틱 뉴런의 막전위 값에 더해지며 막전위 값이 문턱 전압을 넘게 되면 출력 스파이크를 발생시킨다. 스파이크가 발현되면 뉴런 내부의 막전위 값은 초기화되며 일정 기간의 휴지기(refractory period)를 가지게 된다.

3. 뉴로모픽 반도체에 최적화된 비전 센서와 데이터셋

3.1 동적 비전 센서

동적 비전 센서(dynamic vision sensor, DVS)는 생물학적으로 인간이 홍채를 통해 시각 정보를 받아들이고 두뇌로 정보를 전달하는 방식을 모델링 하여 개발된 이벤트 기반 비전 센서이다 [18].

그림 1은 프레임 기반 비전 센서와 동적 비전 센서의 차이를 나타낸 그림이다. 기존의 프레임

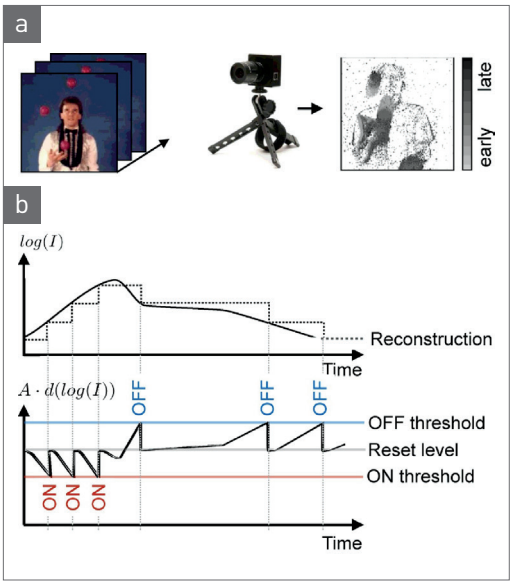

그림 1 ▸ (a) Difference between frame-based sensor and event-based sensor (b) DVS principle [19].

기반 비전 센서는 이미지를 매 프레임 촬영을 하여 데이터를 획득한다. 반면에 동적 비전 센서는 픽셀 단위의 빛의 세기의 변화를 감지하여 그 픽셀 위치 정보를 On/Off 형태의 이벤트로 전송하게 되며 (On: 빛의 세기 증가, Off: 빛의 세기 감소) 마이크로 초 단위의 시간 해상도(temporal resolution)를 가진다. 이는 1초당 수천 프레임을 촬영하는 초고속 카메라 보다 더욱 빠르게 움직임을 포착한다는 것을 의미한다. 이러한 특성 덕분에 동적 비전 센서를 사용하면 전력 소모 및 저장해야 하는 데이터의 규모를 굉장히 줄일 수 있다. 또한 동적 비전 센서에서는 움직이는 객체의 윤곽선 정도만 표현되기 때문에 모니터링되는 객체의 사생활 보호에도 유용하다.

동적 비전 센서는 기존 카메라의 연속적인 프레임에 포함되는 불필요한 메모리 낭비를 줄이며 더욱 정확한 정보를 포함하도록 한다. 따라서 동적 비전 센서는 빠른 동작이 필요한 연구 사례에 다양하게 활용되고 있다. 움직이는 물체를 즉시 파악하여 반응하고 회피해야 하는 자율 주행차, 로봇, 드론과 같은 자율 시스템에서 더욱 유용하게 활용될 것으로 기대된다.

3.2 DVS 동작 데이터셋

현재까지 프레임 기반 카메라를 이용하여 많은 인공지능 학습용 동작 데이터셋들이 만들어져 왔지만 이벤트 기반의 컴퓨터 비전 연구를 진전시키기에 해당 데이터셋들은 적절하지 않았다. 따라서 이벤트 기반 컴퓨터 비전의 발전을 위해서는 동적 비전 센서로 만든 데이터셋이 필요했다. IBM에서는 DVS로 완전히 이벤트 기반으로 구성된 새로운 동작 데이터셋인 DVS 동작 데이터셋을 만들었다 [10]. DVS 동작 데이터셋은 총 1,342개의 데이터들로 구성되어 있는데 이는 10가지 특정 동작과 1가지의 자율 동작으로 이루어져 있으며 29명의 참가자들이 3가지의 조명 환경에서 행하여 만들어진 데이터이다. 분류 성능을 평가하는데 사용하기 위해 23명의 참가자들은 훈련 데이터셋으로 나머지 6명의 참가자들은 테스트 데이터셋으로 분류되어 사용된다. 각각의 동작은 평균 6초 정도의 길이를 가지고 있으며 동작에 따라 픽셀의 빛의 변화가 감지된 곳에 해당하는 On 이벤트, Off 이벤트로 기록되었다. DVS 동작 데이터셋은 이벤트 기반 컴퓨터 비전 연구에 활발히 사용되고 있다.

4. 뉴로모픽 동작 인식 연구 배경과 동향

4.1 뉴로모픽 동작 인식 연구 배경

일반적으로 인공지능을 이용하여 동작을 학습시키는 여러 연구들은 프레임 기반 비전 센서와 동기식 프로세서를 사용해 왔으며 학습에 사용되는 전력을 감소시키는 방향으로 발전해 왔다. 하지만 동작 인식 기술의 반응 속도는 프레임 기반 비전 센서의 프레임 캡처 속도에 제한되며 이는 실시간으로 동작을 인식하는 것엔 한계가 있다. 반응 속도를 높이기 위해서는 빠른 속도로 프레임을 획득하는 센서를 사용해야 하며 이를 처리하기 위해 필요한 전력 소모는 함께 증가한다. 이러한 점을 극복하고자 이벤트 기반 센서와 비동기식 뉴로모픽 프로세서를 활용한 스파이킹 신경망을 통해 동작을 학습하고 인식하는 다양한 알고리즘들이 연구되고 있다.

스파이킹 신경망은 기존 심층 신경망을 학습 시키는 것과 다르게 다음과 같이 고려해야 하는 문제점이 있다. 스파이킹 뉴런의 스파이크를 표현하는 함수는 시간적으로 연속적이지 않아 미분이 불가능하다. 이러한 점은 오류역전파법에서 필요한 오류 보정을 위한 미분값을 구하기 어렵게 만든다. 또한, 스파이킹 뉴런은 연속적인 스파이크 입력의 영향을 포함하고 있기 때문에 입력 스파이크의 시간적 기여도를 할당하기 어려운 문제가 있으며, 뉴런의 상태변수가 시간축에서 지역적인 정보만 가지고 있다는 점에서 Back-Propagation-Through-Time의 사용이 어렵다는 점이 있다. 이러한 문제점으로 인해 기존 심층 신경망을 학습시키는 알고리즘을 직접 스파이킹 신경망에 적용할 수 없으며 스파이킹 신경망 학습에 필요한 학습 방법을 새롭게 제안할 필요가 존재한다.

4.2절에서는 이벤트 기반 센서와 비동기식 뉴로모픽 프로세서를 활용하여 스파이킹 신경망을 통한 동작 인식을 하는 방법에 대해 소개한다. 4.3절과 4.4절에서는 앞서 언급한 스파이킹 신경망을 통해 학습하는데 있어 고려해야 할 스파이킹 뉴런 특성의 문제점들을 해결하는 학습 알고리즘을 제시한다.

4.2 뉴로모픽 반도체와 동적 비전 센서를 활용한 효율적인 동작 인식

비동기식 뉴로모픽 프로세서를 활용한 동작 인식 어플리케이션은 IBM에서 개발한 TrueNorth [20]를 이용하여 구현한 것이 먼저 소개되었다 [10,21]. 스파이킹 신경망의 학습은 심층 컨벌루션 신경망 Energy efficient deep networks (Eedn) 알고리즘을 이용하여 기존 심층 신경망 학습방법을 이용하여 오프라인에서 학습을 하고 학습된 신경망 구조를 TrueNorth를 활용하여 실시간으로 동작인식을 구현한다. 신경망을 학습하기 위해 최초로 DVS 동작 데이터셋을 벤치마크 데이터셋으로 제작하여 논문에서 활용하였다. 비동기식 이벤트 기반의 뉴로모픽 컴퓨팅을 이용하여 저전력으로 동작을 인식할 수 있는 것을 보여준 연구로써 의미가 있다.

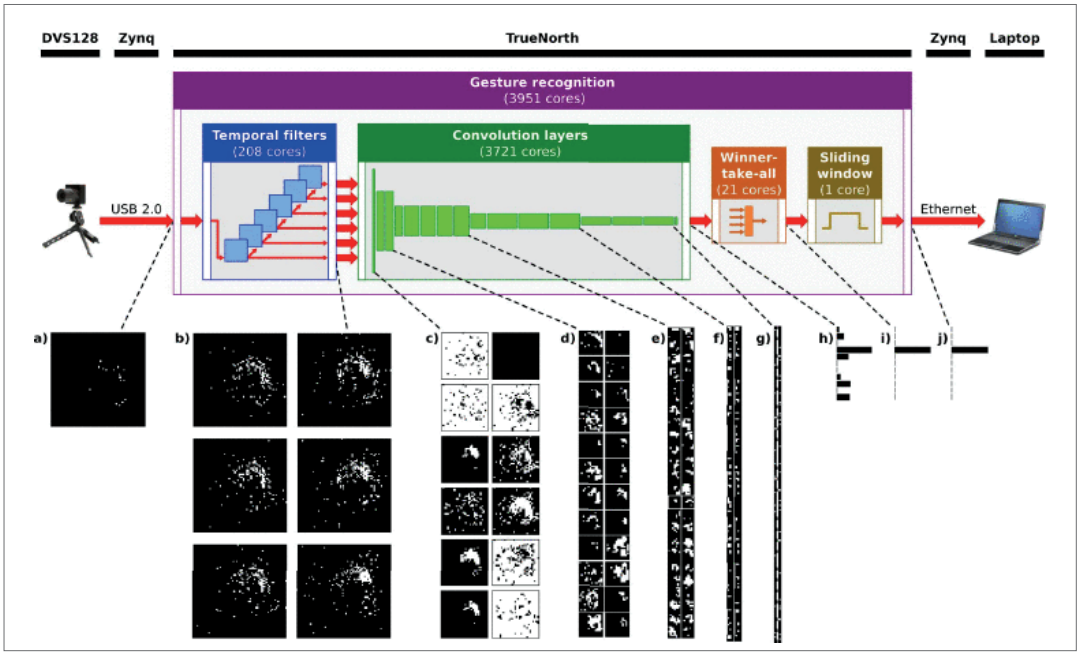

그림 2 ▸ Neural network configuration for the Energy efficient deep networks (Eedn)-based DVS gesture classification algorithm [10,21].

그림 2는 동작 인식 어플리케이션 구현을 위해 Eedn 알고리즘으로 학습된 신경망을 뉴로모픽 프로세서를 이용하여 구현한 구조를 보여준다. 먼저, 연속된 시간축 필터(temporal filter cascade)로 DVS 이벤트 스트림을 연속된 동작 프레임으로 캡쳐한다. 캡쳐된 동작들은 graphics processing unit (GPU)를 사용하여 오프라인으로 훈련된 컨벌루션 신경망에 입력으로 제공된다. 컨벌루션 신경망의 출력은 winner-take-all (WTA) 디코더로 입력되어 해당하는 동작의 클래스가 무엇인지 판별되는 과정을 거친다. WTA 디코더는 출력 뉴런 중 가장 많은 응답을 하는 뉴런이 다른 뉴런의 출력을 억제하고 승자로 출력되는 구조로 구성된 신경망이다. 따라서, WTA 디코더의 출력 뉴런의 스파이크 출력 개수로 컨벌루션 신경망의 출력이 어떤 동작의 입력으로 기인한지 알 수 있다. 마지막으로, 슬라이딩 윈도우 필터를 사용하여 매 밀리초마다 WTA 디코더에 의해 즉각적으로 생성되는 동작 인식 결과를 실시간으로 필터링하는 과정을 거친다.

최종적으로는 80ms의 슬라이딩 윈도우 필터를 사용하여 11가지 동작 (10가지 특정 동작 + 1가지 자율 동작) 인식의 경우 94.59% 정확도를 달성했으며, 10가지 특정 동작 인식은 96.49%의 정확도를 보였다. 또한 동작 인식 시 낮은 지연시간 (105ms) 과 낮은 전력 (178.8 mW) 소비를 보였다. 이로 인해 일반적으로 기존 동기식 프로세서에서 프레임 기반으로 데이터를 처리하는 과정에서 손실되는

에너지 효율을 이벤트 기반의 비동기식 뉴로모픽 프로세서 상에 구현하여 기존 시스템보다 전력소모, 지연시간, 정확도 등에서 더욱 효과적으로 동작 인식 어플리케이션이 구현 가능하다는 것을 보여주었다.

4.3 뉴런의 스파이킹 상태 변화 가능성을 활용한 스파이킹 신경망 학습

Spike layer error reassignment in time (SLAYER) 방법은 스파이크 함수의 미분 불가능한 특성을 스파이킹 함수의 막전위 값 변동성에 따른 뉴런의 스파이킹 상태 변화 가능성을 도입하여 해결한 알고리즘이다 [11]. 이 방법은 기존 신경망의 오류 역전파 알고리즘을 통한 학습과 비슷한 논리를 스파이킹 신경망에 적용시킬 수 있는 학습방법이다. SLAYER 방법을 이용하여 스파이킹 신경망에서도 미분을 통한 오류 역전파 학습을 할 수 있음을 보여주었으며, DVS 동작 데이터셋 학습에 정확도 93.64%를 달성하였다. SLAYER는 앞 절에서 설명한 Eedn 네트워크보다 훨씬 적은 수의 뉴런과 작은 네트워크를 사용한다. Eedn 네트워크를 통한 DVS 동작 학습에서는 데이터가 입력되기 전에 추가 뉴런들을 사용하여 데이터 전처리 과정을 거치지만 SLAYER는 DVS로부터 나온 이벤트 기반 스파이크 데이터들이 바로 네트워크로 입력된다는 점에서 차이가 있다. 또한 SLAYER는 DVS 동작 학습 이외에 MNIST, NMNIST, TIDIGITS 등의 데이터셋에 적용되어 기존의 인식 문제도 스파이킹 신경망을 통해 효율적인 학습이 가능함을 보여주었다.

4.4 온라인 지역학습법을 활용한 스파이킹 신경망 학습

스파이킹 신경망 학습에 있어 연산의 지역성과 뉴런의 미분 불가능한 특성은 핵심 문제이다. Deep continuous local searning (DECOLLE)은 이러한 문제를 스파이킹 뉴런의 surrogate gradient descent와 오직 지역적인 정보로만 학습하는 로컬 학습을 통해 해결하였다 [12,22]. DECOLLE은 앞 절에서 설명했던 오프라인으로 학습하는 Eedn과 SLAYER와는 다르게 온라인으로 DVS 동작 데이터셋을 학습, 인식하며 95.54%의 정확도를 보였다.

그림 3은 DECOLLE을 통해서 DVS 동작 데이

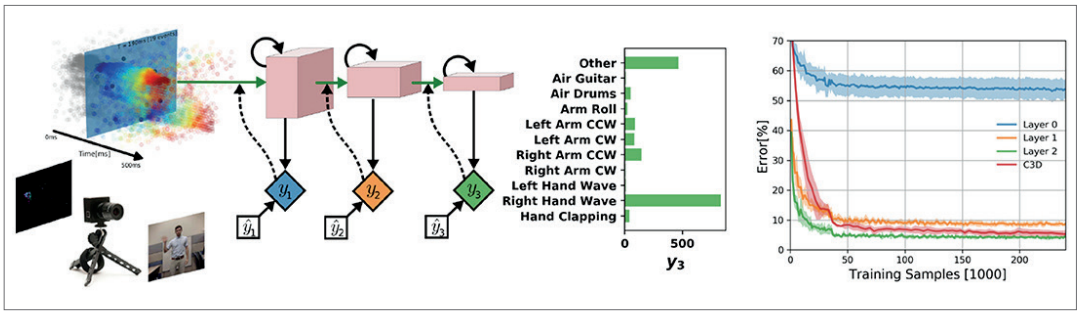

그림 3 ▸ DVS DECOLLE experimental setup for DVS gesture recognition (left) and classification error for the task (right) [12,22].

표 1 ▸ Comparisons.

Model	Error	Training	Iterations
DECOLLE [12]	4.46 %	Online	0.16 M
SLAYER [11]	6.37 %	BPTT	0.27 M
C3D [23]	5.46 %	BPTT	0.32 M
IBM Eedn [10]	5.51 %	BPTT	64 M

터셋이 학습되는 전반적인 네트워크 구조를 보여준다. DECOLLE의 구체적인 네트워크 구조는 영상의 시공간 분류에 일반적으로 사용되는 3D 컨볼루션 네트워크인 Convolutional 3D (C3D) 네트워크 [23]와 비슷하지만 더욱 간단한 네트워크 구조를 가지고 있다. DECOLLE은 뉴런들의 시간 역학을 처리함에 있어 스파이킹 신경망과 순환 신경망(recurrent neural network, RNN)의 유사성을 반영하여 미분값을 찾는 접근법 덕분에 기존에 존재하는 머신러닝 프레임워크의 자동 미분 도구들의 사용이 가능하도록 하였다. 또한 계층별로 오류 함수를 계산하게 하여 공간적인 지역성(spatial locality) 문제를 해결하고 BPTT가 아닌 real-time recurrent learning (RTRL)을 하여 학습에 필요한 정보의 시간적 지역성(temporal locality) 문제를 해결하였다.

표 1은 여러 네트워크 모델들을 이용한 DVS 동작 데이터셋 학습 성능을 비교한 표이다. DECOLLE은 Eedn보다 학습하는 데에 소요되는 반복 횟수가 크게 줄어들고 네트워크가 작아졌는데도 불구하고 우수한 정확도를 보였고, 더욱 큰 네트워크 구조를 가진 C3D와는 비슷한 성능을 보였다.

5. 결론

본 논문에서는 뉴로모픽 반도체를 활용한 스파이킹 신경망이 가지는 강점인 저전력, 저지연 특성들에 대해 논의하고 스파이킹 신경망 학습의 어려움을 알아보며 학습을 구현하기 위해 제안된 몇 가지 알고리즘과 이를 검증하기 위해 구현된 동작 인식 어플리케이션들에 대해 알아보았다. 인간의 뇌의 신경망 구조 및 동작 원리를 모방한 뉴로모픽 기술을 사용하여 실제 뉴런의 동작 학습 원리를 적용한 스파이킹 신경망을 통해 학습하는 알고리즘은 향후 인공지능을 발전시키는데 전력 소모 및 데이터 관리 차원에서 큰 도움이 될 것으로 보인다. 또한 본 논문에서 다룬 동작 인식뿐만 아니라 이미지 인식, 목소리 인식 등 인간이 수행할 수 있는 다양한 분야에서도 활용 가능할 것으로 기대된다.

참 / 고 / 문 / 헌

[1] P. U. Diehl and M. Cook, *Front Comput Neurosci*, 9, 99 (2015). [doi:10.3389/fncom.2015.00099]

[2] S. R. Kheradpisheh, M. Ganjtabesh, S. J. Thorpe, and T. Masquelier, *Neural Network.*, 99, 56 (2018). [doi:10.1016/j.neunet.2017.12.005]

[3] S. Loiselle, J. Rouat, D. Pressnitzer, and S. Thorpe, *2005 IEEE International Joint Conference on Neural Networks*, 4, 2076 (2005). [doi: 10.1109/IJCNN.2005.1556220]

[4] S. G. Wysoski, L. Benuskova, and N. Kasabov, *Neural Networks*, 23, 819 (2010). [doi:10.1016/j.neunet.2010.04.009]

[5] A. Tavanaei and A. Maida, *Neural Information Processing. ICONIP 2017* (Liu D., Xie S., Li Y., Zhao D., El-Alfy ES.) (Springer, Cham, 2017), p. 899. [doi:10.1007/978-3-319-70136-3_95]

[6] B. Han and T. M. Taha, *Applied Optics*, 49, b83 (2010). [doi:10.1364/AO.49.000B83]

[7] K. Dhoble, N. Nuntalid, G. Indiveri, and N. Kasabov, *The 2012 International Joint Conference on Neural Networks(IJCNN)*, (IEEE, Brisbane, QLD, Australia, 2012). p. 1. [doi: 10.1109/IJCNN.2012.6252439]

[8] A. Mohemmed, S. Schliebs, S. Matsuda, and N. Kasabov, International Journal of Neural Systems, 22, 1250012 (2012). [doi: 10.1142/S0129065712500128]

[9] N. Kasabov, K. Dhoble, N. Nuntalid, and G. Indiveri, *Neural Networks*, 41, 188 (2013). [doi:10.1016/j.neunet.2012.11.014]

[10] A. Amir et al., *2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR)* (IEEE, Honolulu, HI, USA, 2017). p. 7388. [doi:10.1109/CVPR.2017.781]

[11] S. B. Shrestha and G. Orchard, *arXiv preprint arXiv*, 1810, 08646 (2018).

[12] J Kaiser, H. Mostafa, and E. Neftci, *Front. Neurosci*, 14, 424 (2020). [doi:10.3389/fnins.2020.00424]

[13] J. Feng, *Neural Network.*, 14, 955 (2001). [doi:10.1016/S0893-6080(01)00074-0]

[14] J. Feng and D. Brown, *Bulletin of Mathematical Biology.*, 62, 467 (2000). [doi:10.1006/bulm.1999.0162]

[15] Y. H. Liu and X. J. Wang, *Journal of computational Neuroscience.*, 10, 25 (2001). [doi:https://doi.org/10.1023/A:1008916026143]

[16] M. Nelson and J. Rinzel, *The Book of GENESIS* (James M. Bower., David Beeman. New York, 1998). p. 29. [doi:10.1007/978-1-4612-1634-6_4]

[17] W. Gerstner, W. M. Kistler, R. Naud, and L. Paninski, *Neuronal Dynamics: From Single Neurons to Networks and Models of Cognition* (Cambridge University Press, Cambridge, 2014).

[18] P. Lichtsteiner, C. Posch, and T. Delbruck., *IEEE Journal of Solid-State Circuits*, 43, 566 (2008)., [doi:10.1109/JSSC.2007.914337]

[19] N. K. Medathati, H. Neumann, G.S. Masson, and P. Kornprobst, *Computer Vision and Image Understanding*, 150, 1 (2016). [doi:10.1016/j.cviu.2016.04.009]

[20] F. Akopyan et al., *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 34, 1537 (2015). [doi: 10.1109/TCAD.2015.2474396]

[21] S. K. Esser et al., *Proceedings of the National Academy of Sciences*, 113, 11441 (2016). [doi:10.1073/pnas.1604850113]

[22] E. O. Neftci, H. Mostafa, and F. Zenke, *IEEE Signal Processing Magazine*, 36, 51 (2019). [doi: 10.1109/MSP.2019.2931595]

[23] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, *2015 IEEE International Conference on Computer Vision (ICCV)* (IEEE, Santiago, Chile, 2015) p. 4489. [doi:10.1109/ICCV.2015.510]

저/자/약/력

성명 남기령

학력	2020년	한양대학교 (ERICA) 전자공학부 공학사
	2020년 ~ 현재	고려대학교 전기전자공학부 석사과정
경력	2020년 ~ 현재	한국과학기술연구원 학생연구원

성명 박종길

학력	2007년	고려대학교 전기전자전파공학부 공학사
	2010년	University of California, San Diego, Electrical and Computer Engineering, 공학석사
	2014년	University of California, San Diego, Electrical and Computer Engineering, 공학박사
경력	2014년 ~ 2018년	한국전자통신연구원 연구원
	2018년 ~ 현재	한국과학기술연구원 선임연구원