

Asynchronous Neural Network for Neuromorphic Event-Based Image Sensor Driven by Change of the Environment

February 2020

Yusuke Sekikawa

報告番号	Ⓐ 乙 第	号	氏 名	関川 雄介
主 論 文 題 名 : Asynchronous Neural Network for Neuromorphic Event-Based Image Sensor Driven by Change of the Environment (生物のように環境の変化に反応するセンサーのための 非同期型のニューラルネットワーク)				
(内容の要旨) 本論文では、イベントカメラと呼ばれる生物の目に着想を得た非同期型のカメラを用いた非同期型のニューラルネットワークについて検討を行った。 既存のフレームベースのパラダイム—同期型のビデオストリームの取得と、それを使った同期型のニューラルネットワークによる処理—は、演算量の問題で高い時間解像度のデータや、高出力レートが必用な用途への適用が困難である。この計算量の増大は、主に、変化のないピクセルの輝度値を取得／送信する冗長な観測メカニズムと、CNN といった信号処理メカニズムの冗長性に起因する。例えば、convolutional neural network (CNN) では変化のない入力に対しても積和演算を行うため、これが計算量増大の原因となっている。さらには、時空間データのニューラルネットワークによるモデリングによく用いられる 3D CNN では、同じフレームが処理ウィンドウ内に存在する間、何度も処理されるため、さらに計算量が増加する。 近年、これらの課題を解決するため、生物の視覚システムを模し輝度に“変化”があった画素の情報のみを取得するイベントカメラと呼ばれるカメラが注目を集めている。イベントカメラは、動きの情報のみを取得する仕組みのカメラで、その非冗長かつスパースな“変化”情報を上手く処理することで、生物のように非常に効率的なデータ取得・処理を可能にするポテンシャルがある。 しかしながら、スパースなイベントデータは、従来のデンスなフレームデータとは大きく異なるため、CNN 等のフレームベースのニューラルネットワークが適用できない。スパースなイベント信号をデンスなフレーム画像のような形式に変換し、CNN を適用する例も存在するが、このアプローチではイベントデータの特徴であるスパース性が失われるため、高時間レートなイベントデータをリアルタイム処理することが困難となる。 本研究では、このスパース性を活かした非同期型のニューラルネットワークの検討を行った。まず、入力信号に局所的な線形性を仮定することで、スパースなイベント信号系列から低レベルの特徴（エッジの時空間運動）を逐次的に抽出するための 3 次元畳み込み演算アルゴリズム (cv3dconv) を提案した。等速直線運動に制約された畳み込みカーネルを考えた場合、計算量の大きな 3 次元畳み込み演算が、計算量の小さな 2 次元畳み込み演算と和の合成に分解できることに着目し効率的な逐次演算を実現している。さらに、このようにして得られた低次特徴や生のイベントを End-to-End でモデリングするための逐次処理型のニューラルネットワークとして、生物の脳を模した非同期型のネットワークである Spiking Neural Network に着想を得た逐次処理型のニューラルネットワーク (EventNet) を提案した。EventNet は、DVS から得られる非同期でスパースな時系列データを少数の積和演算で逐次的に処理することできるため、変化のないピクセルについても演算を行う従来型のフレームワークと比べて、大幅な演算量の低下を実現した。実データを用いた実験で数万個のイベントデータをリアルタイム処理ができることを確認した。				

Thesis Abstract

No.1

Registration Number	☒ “KOU” ☐ “OTSU” No. *Office use only	Name	Yusuke Sekikawa
Thesis Title: Asynchronous Neural Network for Neuromorphic Event-Based Image Sensor Driven by Change of the Environment			
Thesis Summary In this dissertation, we study the asynchronous neural network architecture for a bio-inspired change detection based asynchronous image sensor called event-based camera to realize fast and computationally efficient image processing. The paradigm of existing frame-based synchronous neural network--dense synchronous video stream acquisition at a fixed followed by dense synchronous processing by neural networks--is difficult to scale to higher frame rates or finer temporal resolutions because the computational complexity grows linearly with the processing rate or temporal resolution. The growth of the computational complexity comes from the redundant, synchronous measurement/transmission of dense intensity frames for unchanged pixels and the following redundant neural networks, such as convolutional neural networks (CNNs), which compute the sum of products, even for the unchanged pixels. Furthermore, in 3D CNN, which is a typical architecture for processing spatiotemporal data, the same frames are computed multiple times (temporal sliding window operation) for modeling temporal dependencies. Event-based cameras are bio-inspired vision sensors that mimic mammalian retinas to asynchronously report per-pixel intensity changes rather than outputting an actual intensity image at regular intervals. This new image sensor offers significant potential advantages, namely, fast and efficient signal processing by utilizing the sparse and non-redundant natures of event data. Unfortunately, however, most of the existing artificial neural network architecture, such as CNNs, requires dense synchronous input data, and therefore, cannot make use of the sparseness of the data. An efficient neural network architecture that can make use of the sparse nature of the event data is required to unlock the outstanding characteristics of the asynchronous camera. To this end, we investigate the possibilities of asynchronous neural network architecture for the asynchronous data from the event-based camera, especially suitable for processing on CPU or GPU. First, we propose a novel recursive 3D convolutional algorithm for extracting the low-level spatiotemporal feature (such as moving edge), which can be computed recursively by considering the constant movement of edges. Using the restricted class of 3D-kernel (constant velocity), we have derived a novel formulation, where computationally-intensive 3D convolution is decomposed into computationally-efficient 2D convolution followed by simple summation, thus achieves significant speedup over the existing frame-based			

counterpart.

Next, we propose an end-to-end trainable recursive and asynchronous neural network architecture that runs quite efficiently by utilizing novel recursive formulation and look-up-table.

As a result, at inference time, networks using the architecture operate in an event-wise manner that is realized with very few sum-of-products operations, which enables processing of few mega or more events per second on standard CPU.