革新的コンピューティング技術の開拓
2019 年度採択研究者

2021 年度
年次報告書

孫鶴鳴

早稲田大学理工総合研究所
次席研究員（研究院講師）

リアルタイム低電力深層学習適用による革新的な動画像圧縮システム

§1. 研究成果の概要

This year, we mainly aim at the FPGA prototype of the learned video codec. We have confirmed that our developed system can reach real-time coding for 720p@30fps.

FPGA based neural network (NN) accelerator can be classified into two categories (generic and pipeline) according to the paradigms of hardware resources utilization. Fluctuation of computation-to-communication (CTC) ratio among layers will degrade the performance of the generic architecture, while it can be addressed in the pipeline architecture. However, the previous pipeline architecture cannot exploit the overall computational potential of FPGA. Therefore, we propose a fine-grained DSP allocation algorithm based on a layer-wise pipeline architecture. Besides, an adaptive activation buffer between two consecutive layers is designed to support a fine-grained granularity for DSP allocation. Furthermore, a CTC adjustment algorithm is developed to satisfy the off-chip bandwidth constraints. By using the proposed FPGA neural engine, we can build the system to satisfy various requirements of video resolution and speed.

In addition to the hardware design, we also worked on some algorithmic improvements. The related results have been published in several IEEE journals and conferences.