World Academy of Science, Engineering and Technology
International Journal of Computer and Systems Engineering
Vol:14, No:06, 2020

Embedded Semantic Segmentation Network Optimized for Matrix Multiplication Accelerator

Authors : Jaeyoung Lee

Abstract : Autonomous driving systems require high reliability to provide people with a safe and comfortable driving experience. However, despite the development of a number of vehicle sensors, it is difficult to always provide high perceived performance in driving environments that vary from time to season. The image segmentation method using deep learning, which has recently evolved rapidly, provides high recognition performance in various road environments stably. However, since the system controls a vehicle in real time, a highly complex deep learning network cannot be used due to time and memory constraints. Moreover, efficient networks are optimized for GPU environments, which degrade performance in embedded processor environments equipped simple hardware accelerators. In this paper, a semantic segmentation network, matrix multiplication accelerator network (MMANet), optimized for matrix multiplication accelerator (MMA) on Texas instrument digital signal processors (TI DSP) is proposed to improve the recognition performance of autonomous driving system. The proposed method is designed to maximize the number of layers that can be performed in a limited time to provide reliable driving environment information in real time. First, the number of channels in the activation map is fixed to fit the structure of MMA. By increasing the number of parallel branches, the lack of information caused by fixing the number of channels is resolved. Second, an efficient convolution is selected depending on the size of the activation. Since MMA is a fixed, it may be more efficient for normal convolution than depthwise separable convolution depending on memory access overhead. Thus, a convolution type is decided according to output stride to increase network depth. In addition, memory access time is minimized by processing operations only in L3 cache. Lastly, reliable contexts are extracted using the extended atrous spatial pyramid pooling (ASPP). The suggested method gets stable features from an extended path by increasing the kernel size and accessing consecutive data. In addition, it consists of two ASPPs to obtain high quality contexts using the restored shape without global average pooling paths since the layer uses MMA as a simple adder. To verify the proposed method, an experiment is conducted using perfsim, a timing simulator, and the Cityscapes validation sets. The proposed network can process an image with 640 x 480 resolution for 6.67 ms, so six cameras can be used to identify the surroundings of the vehicle as 20 frame per second (FPS). In addition, it achieves 73.1% mean intersection over union (mIoU) which is the highest recognition rate among embedded networks on the Cityscapes validation set.

Keywords : edge network, embedded network, MMA, matrix multiplication accelerator, semantic segmentation network
Conference Title : ICCV 2020 : International Conference on Computational Vision
Conference Location : Venice, Italy
Conference Dates : June 22-23, 2020