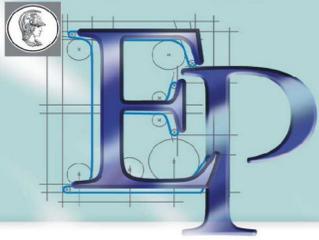


Projeto de Formatura – Turmas 2018



PCS - Departamento de Engenharia de Computação e Sistemas Digitais

Engenharia Elétrica – Ênfase Computação

Tema: **Implementação de rede neural convolucional em FPGA para classificação de imagem**

Objetivo

O objetivo do projeto foi desenvolver um hardware para fazer reconhecimento de padrões e classificação de imagens através do algoritmo da rede neural. O hardware alvo é um FPGA, que consiste em um chip com grande poder de reconfiguração no nível de portas lógicas. É uma plataforma muito eficiente para o processamento paralelizável e que exige grande velocidade para cálculos simples. O kit de desenvolvimento utilizado foi a placa De1-SoC com Cyclone V SoC da Terasic.

Redes neurais artificiais

A rede neural artificial é uma técnica computacional inspirados no sistema nervoso humano e possui capacidade de aprendizado de máquina para reconhecimento de padrões. A unidade básica de processamento dessa rede é chamada neurônio e ele é capaz de fazer cálculos simples. Contudo, o agrupamento de diversos desses neurônios em camadas conectadas permite criar estruturas complexas com alto poder de adaptação e aprendizado em um processo chamado treinamento.

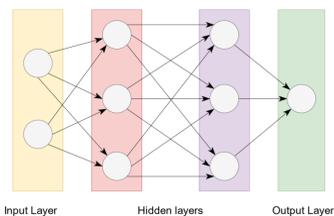


Figura 1: Rede neural totalmente conectada e Rede neural convolucional



A rede neural convolucional é um caso mais avançado de rede neural. A entrada desse tipo de rede são imagens e o componente básico da rede é um filtro ou kernel de convolução. Logo, diferente dos neurônios que realizam operações simples, a rede neural convolucional conta com camadas que fazem convolução, uma operação matemática que consegue identificar melhor padrões localizados. Assim, é o algoritmo mais recomendado para procurar padrões em imagens.

Além das camadas totalmente conectada e convolucionais, é comum a presença de camadas do tipo max-pooling. Esse tipo de camada tem a função de reduzir a dimensão da imagem de entrada para diminuir a complexidade e a variância à translação.

Integrantes:

Christian Kazuyoshi Miyoshi
Jean René Joseph Canteneur

Professor Orientador: Edson Satoshi Gomi
Co-orientador: Bruno de Carvalho Albertini

Arquitetura do sistema

O sistema foi projetado para ter entidades para cada tipo de camada de uma rede neural de forma que elas possam ser reutilizadas. Isso permite ter uma grande flexibilidade para que o hardware possa executar o algoritmo para diversas redes neurais.

A camada totalmente conectada contém uma entidade neurônio responsável pela adições, multiplicações e aplicação da função e ativação. Além disso, ela é comandada pelos sinais do controlador do neurônio, unidade responsável pelos sinais de controle e pelo endereçamento das entradas e pesos da memória. O início e término do processamento são determinados pelo controlador da camada que recebe os sinais de um controlador principal externo.

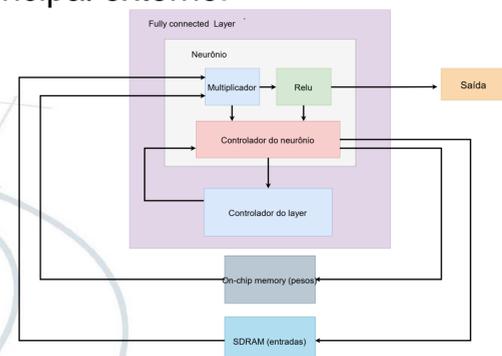


Figura 2: Arquitetura da camada totalmente conectada

A entidade da camada convolucional é baseada em um pipeline dos cálculos intermediários. Ao fazer a leitura do valor do pixel da entrada, todos os resultados das multiplicações são obtidos e acumulados em registradores que são deslocados para coincidir com as operações da convolução. Isso permite obter o resultado com menos ciclos de clock e de forma mais eficiente para fazer a escrita do resultado na memória da placa.

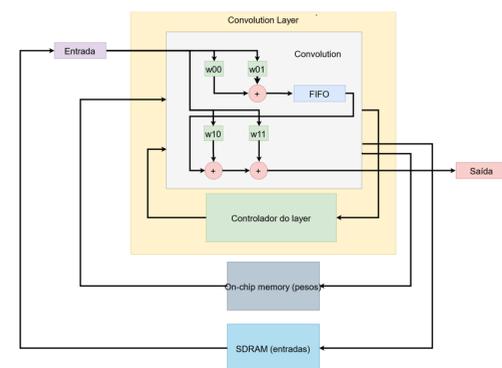


Figura 3: Arquitetura da camada convolucional.

Referências Bibliográficas

- A. Muthuramalingam, S. Himavathi, E. Srinivasan [2008]. Neural Network Implementation Using FPGA: Issues and Application.
- GSCHWEND, David [2006]. ZynqNet: An FPGA-Accelerated Embedded Convolutional Neural Network.
- KAIYUAN Guo, SHULIN Zeng, JINCHENG Yu, YU Wang, HUAZHONG Yang. A Survey of FPGA Based Neural Network Accelerator
- FARABET Clément, POULET Cyril, HAN Jefferson LECUN Yann. CNP: Na FPGA-based processor for convolutional networks.