

Eyeriss Project

Gabriel Almeida Gonçalves	9292761
Linkon Louvison	7143698
Vitor Carneiro	8066183

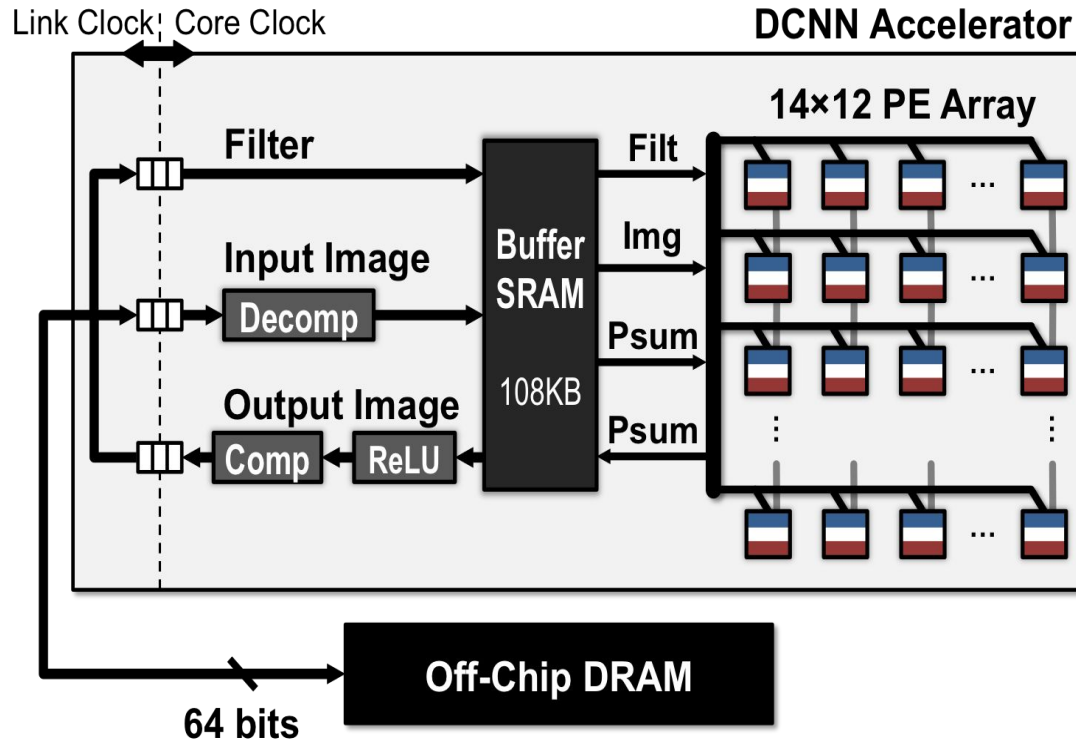
Eyeriss

- Um otimizador configurável para Redes Neurais Convolucionais Profundas
- Foco em economia de energia e classificação e processamento de imagens
- Primeira vez exibida no ISCA 2016, por Yu-Hsin (principal autor do projeto)
- Desenvolvida a partir de uma parceria do MIT e Nvidia

Rede Neural Convolutacional

- Convolução: uma operação matemática aplicada em duas funções a fim de obter-se uma terceira
- Mais conhecido como CNN (Convolutional Neural Network)
- É uma classe de rede neural artificial, com foco em processamento de imagens e vídeo
- Desenvolvida com base no Córtex Visual dos animais
- A rede aprende os filtros necessários para o funcionamento sem necessidade de implementação manual
- Uma CNN tende a demandar um nível mínimo de pré-processamento quando comparada a outros algoritmos de classificação de imagens.

Arquitetura do Eyeriss



Fonte: <http://eyeriss.mit.edu>

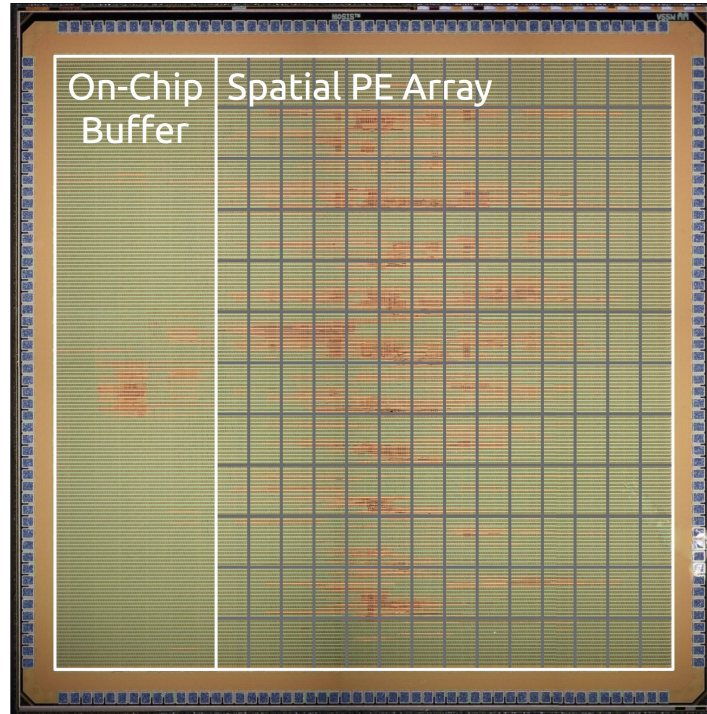
Especificação do Chip e comparação com GPU

Technology	TSMC 65nm LP 1P9M
Core Area	3.5mm×3.5mm
Gate Count	1852 kGates (NAND2)
On-Chip Buffer	108 KB
# of PEs	168
Scratch Pad / PE	0.5 KB
Supply Voltage	0.82 – 1.17 V
Core Frequency	100 – 250 MHz
Peak Performance	33.6 – 84.0 GOPS (2 OP = 1 MAC)
Word Bit-width	16-bit Fixed-Point
Filter Size*	1 – 32 [width] 1 – 12 [height]
# of Filters*	1 – 1024
# of Channels*	1 – 1024
Stride Range	1–12 [horizontal] 1, 2, 4 [vertical]

	<i>This Work</i>	NVIDIA TK1 (Jetson Kit)
Technology	65nm	28nm
Clock Rate	200MHz	852MHz
# Multipliers	168	192
On-Chip Storage	Buffer: 108KB Spad: 75.3KB	Shared Mem: 64KB Reg File: 256KB
Word Bit-Width	16b Fixed	32b Float
Throughput¹	34.7 fps	68 fps
Measured Power	278 mW	Idle/Active ² : 3.7W/10.2W
DRAM Bandwidth	127 MB/s	1120 MB/s ³

1. AlexNet Convolutional Layers Only
2. Board Power
3. Modeled from [Tan, SC11]

Arquitetura do Eyeriss



Fonte: <http://eyeriss.mit.edu>

Características do Eyeriss

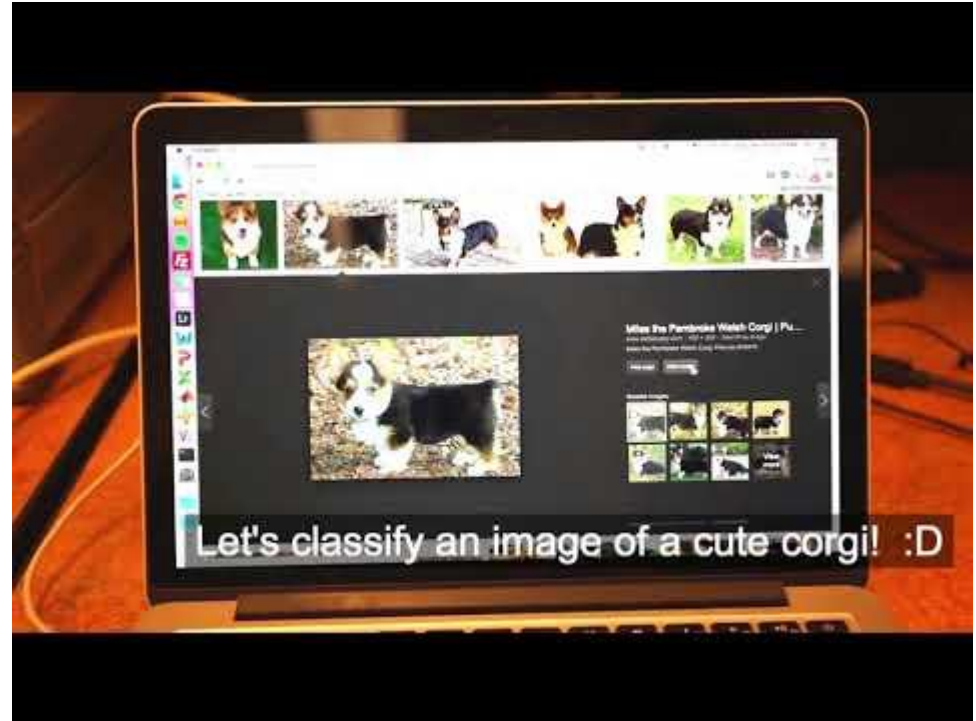
- Acelerador de rede neural convolucional profunda com eficiência energética.
- 168 elementos de processamento(PE).
- Integrado com o framework de deep learning Caffe.
- Alimentado por uma rede multicast reconfigurável on-chip que lida com muitas formas e minimiza a movimentação de dados utilizando da reutilização dos mesmos.

Vantagens no uso

- Até dez vezes mais eficiente que um GPU móvel tradicional.
- Implementar redes neurais artificiais para dispositivos móveis com um consumo de energia reduzido.
- Cada PE possui memória independente para armazenar e analisar os dados.
- Design de circuito único permitindo cada núcleo maximizar a carga que consegue processar sem acesso à memória principal.

Demonstração

O vídeo ao lado demonstra uma tarefa de classificação de imagem de 1.000 classes em tempo real usando o AlexNet pré-treinado que é executado em nosso sistema Eyeriss Caffe.



Referências

- Eyeriss: An Energy-Efficient Reconfigurable Accelerator for Deep Convolutional Neural Networks - Yu-Hsin Chen, 2016 IEEE - http://www.rle.mit.edu/eems/wp-content/uploads/2016/02/eyeriss_isscc_2016_slides.pdf?fbclid=IwAR30qUCuxMuiDgzvzhz0IyYv8HQymb5UBNShGKUNP2kmxurUEQV2RQiNnIQU
- Eyeriss: An Energy-Efficient Reconfigurable Accelerator for Deep Convolutional Neural Network - Yu-Hsin Chen , Tushar Krishna , Joel Emer1, Vivienne Sze, 2016 IEEE - http://www.rle.mit.edu/eems/wp-content/uploads/2016/02/eyeriss_isscc_2016.pdf
- Understanding Convolutions - Christopher Olah, Jul 2014 - <http://colah.github.io/posts/2014-07-Understanding-Convolutions/>
- O que é uma Rede Neural Convolutacional? - Alex Fernandes Mansano, Jul 2017 - <https://www.linkedin.com/pulse/o-que-%C3%A9-um-rede-neural-convolutacional-alex-fernandes-mansano>

Referências

- Eyeriss Project Website - Yu-Hsin Chen, Tushar Krishna, Joel Emer, Vivienne Sze - IEEE ISSCC 2016 - <http://eyeriss.mit.edu/>
- Eyeriss Related Papers - Yu-Hsin Chen, Tushar Krishna, Joel Emer, Vivienne Sze - IEEE ISSCC 2016 - <http://eyeriss.mit.edu/#papers>