SOLUTION BRIEF

MOFFETT AI

XILINX

Xilinx Accelerated, Cloud Based Sparse Neural Network Accelerated Visual Search

INTRODUCTION

Moffett AI offers a breakthrough Visual Search as a Service Solution architecture based on Sparse Processing on FPGAs— for Face Identification and Visual Search which is widely used in surveillance, smart retail, social media and autonomous vehicles.

Through software-hardware co-optimization, Moffett AI and Xilinx can offer low-power, low-cost alternative to high-end GPUs for low-latency visual search on huge databases.

KEY BENEFITS

- Sparse computation reduce ops and data storage by 10x – 100x, deliver 50x Query per second (QPS) throughput versus NVIDIA Tesla T4 GPU
- 20x Throughput per $
- 20x Lower Cost

Product Highlights:

- Industry First FPGA based Sparse Neural Network Accelerator Solution
- Toolchain for popular ML Frameworks including Tensor Flow, PyTorch, ONNX frameworks
- Moffett Visual Search as a Service Offered on Global Cloud and On-Premises:
  - AWS F1 Xilinx Instance
  - Alibaba Cloud F3 Instance
- Advanced Toolchain & SDK works with existing workflows, allows for advanced optimization, performance modeling, and estimator

XILINX

Visual Search

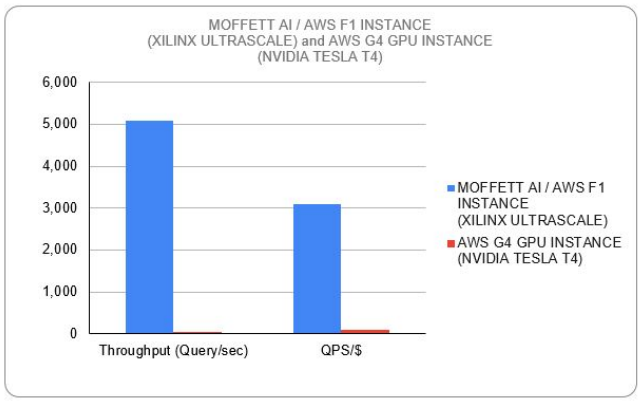
uses real-world images (screenshots, Internet images, or photographs) as the stimuli for online searches.

Modern visual search technology uses AI (artificial intelligence) to understand the content and context of these images and return a list of related results.

VIsual Search is a Killer Application for Moffett AI's unique Sparse Processing Engine coupled with Xilinx Datacenter FPGAs.

Moffett AI Visual Search Engine

The World's
Most Powerful (Throughput)
Fastest (Latency)
Most Energy Efficient
(Throughput/Watt)
AI Inference Platform

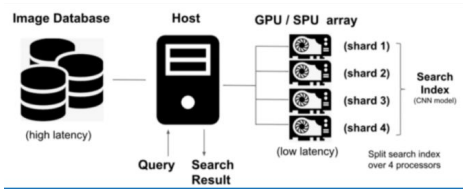


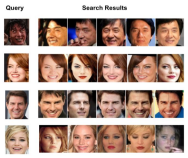
MOFFETT AI

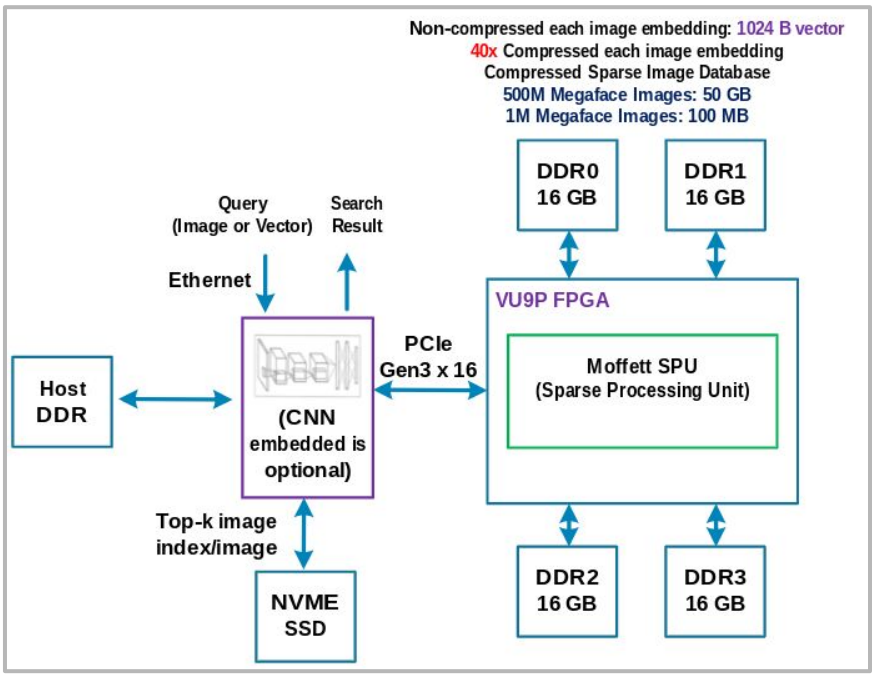
SOLUTION OVERVIEW

Moffett Visual Search as a Service (VSaaS)

Moffet AI's FPGA Accelerated Visual Search Engine and Service are ready made to accelerate Visual Search inference across a wide variety of industries and applications, such as surveillance, smart retail, content search, social media and autonomous vehicles, delivering an order of magnitude higher efficiency via reduced need for OPS & smaller network size (storage), delivering up to 15x lower cost compared to GPU based solutions.

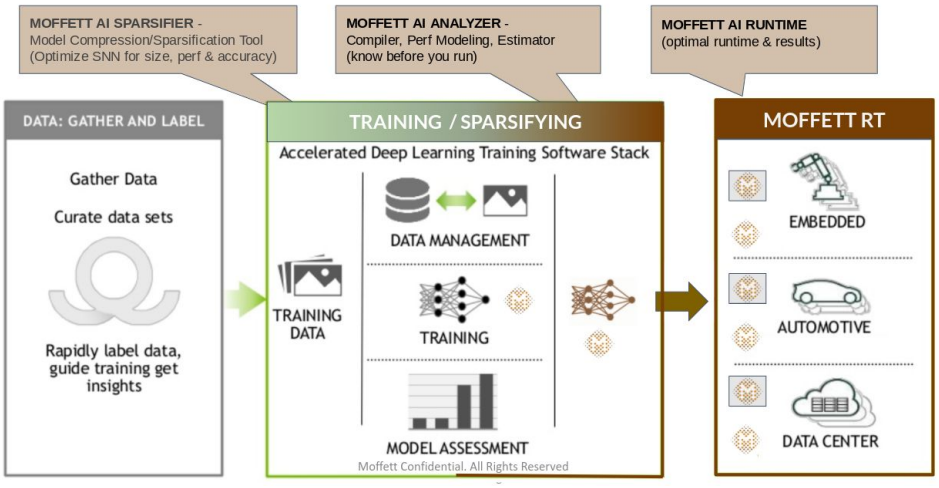






DEPLOYMENT

- VSaaS Offered on Global Cloud (AWS F1, Alibaba F3 Instance)
- Advanced, Compatible Toolchain & SDK integrates with existing workflows, allows for advanced optimization
- Compatible with ML Frameworks including: Tensor Flow, PyTorch, ONNX frameworks



TAKE THE NEXT STEP

- Contact Moffett AI about access to the Visual Search Engine Machine Instance and flexible SDK/APIs for Xilinx FPGA Instances on AWS and Alibaba Cloud for demo and Proof of Concept (contact: demo@moffett.ai)