NASA Goddard Workshop on Artificial Intelligence, 28NOV18

Hybrid Semantic Image Segmentation using Deep Learning for On-board Space Processing

Investigators and Students

Sebastian Sabogal, Alan D. George – SHREC University of Pittsburgh

Gary Crum – NASA Goddard Space Flight Center

Research highlighted in this presentation was supported by SHREC members and by the I/UCRC Centers Program of the National Science Foundation under Grant No. CNS-1738783.

Deep Learning for Space

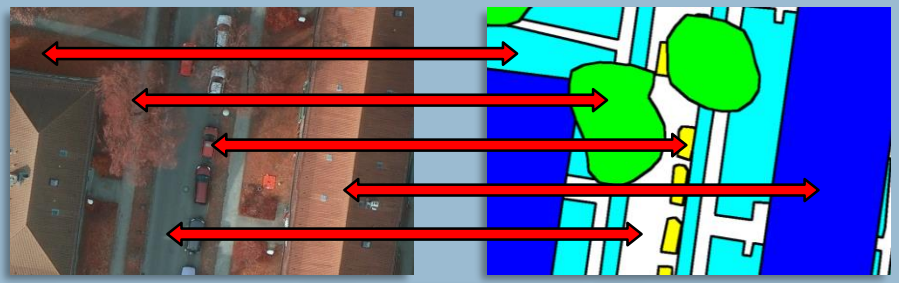
- Space Computing Challenges
  - Escalating demands for high-performance on-board processing
    - Convert large volumes of raw-sensor data into actionable data or scientific knowledge to overcome downlink bandwidth
    - Enable real-time systems for autonomous spacecraft operations
  - Unique constraints in size, weight, power, and cost (SWaP-C)
  - Unique hazards in radiation, vibration, thermal, and vacuum

- Deep Learning in Space
  - Numerous opportunities for enhanced scientific methods, autonomous operations, and intelligent space applications
  - Deep learning is computationally prohibitive on traditional radiation-hardened (rad-hard) processors

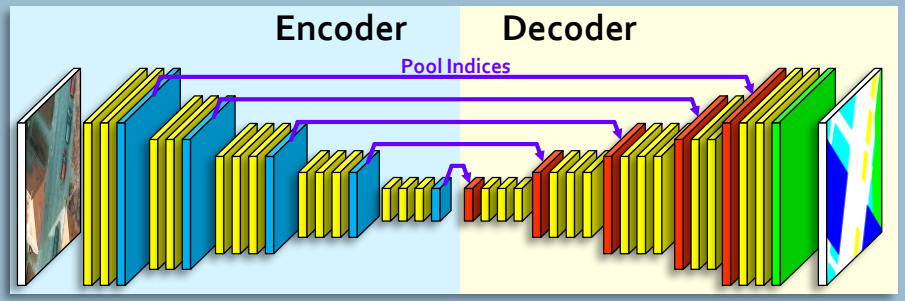
Mission-Critical Computing
NSF CENTER FOR SPACE, HIGH-PERFORMANCE, AND RESILIENT COMPUTING (SHREC)

Semantic Image Segmentation

- Computer Vision / Machine Learning Process
  - Assign labels to all pixels
  - Pixels with same label share semantic characteristics
  - Output roughly resembles input
- Space Applications
  - **Science**: Earth observations and remote sensing
  - **Defense**: reconnaissance and intelligence gathering
- SegNet Model
  - Convolutional neural network (CNN)
  - Deep learning (86 layers)
  - Encoder-decoder architecture
    - **Encoder:** classify objects
    - **Decoder:** upsample feature maps using pooling indices



Roads | Low Vegetation | Automobiles
Buildings | Trees



Convolution + Batch Normalization + ReLU | Pool | Unpool | Softmax

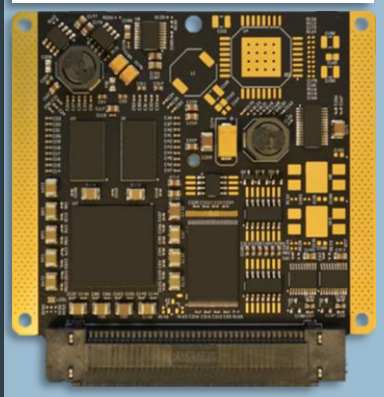
NASA
Mission-Critical Computing
NSF CENTER FOR SPACE, HIGH-PERFORMANCE, AND RESILIENT COMPUTING (SHREC)

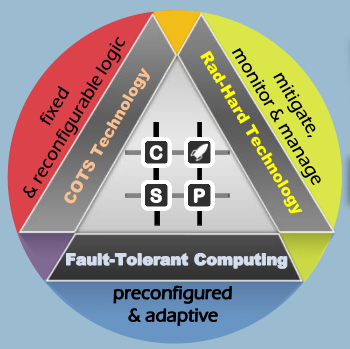
Hybrid Semantic Image Segmentation

- CHREC Space Processor (CSP)
  - Multifaceted **Hybrid** Space Computer
    - **Hybrid system-on-chip (SoC):** CPU + FPGA
    - **Hybrid architecture:** COTS + rad-hard
    - **Robust design:** Novel mix of COTS, rad-hard, and fault-tolerant computing
  - Operational on International Space Station since Mar'17 on **STP-H5/CSP**
  - Planned for numerous missions: NASA CeREs, Lockheed-Martin LunIR, and **STP-H6/SSIVP**

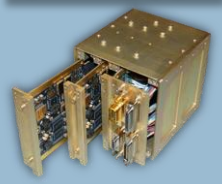
- Hybrid Semantic Image Segmentation
  - Control-flow parts → CPU
  - Data-flow parts → FPGA
  - Scatter-gather DMA (SGDMA)
    - High-throughput streams of multi-dimensional feature maps
    - Interleaving & deinterleaving architecture
  - ReCoN accelerator
    - High-performance CNN inferencing
    - Scalable, parameterizable, and optimized

CSPv1 Rev. B

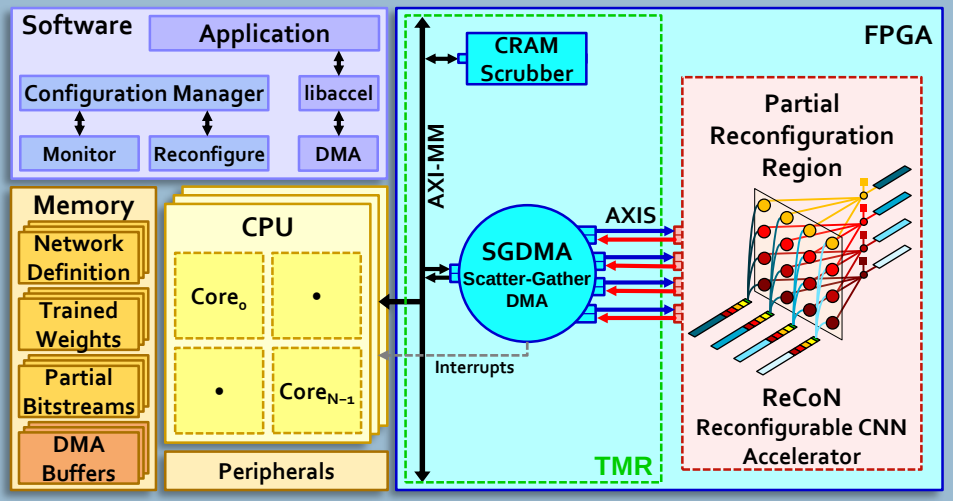




STP-H5/CSP



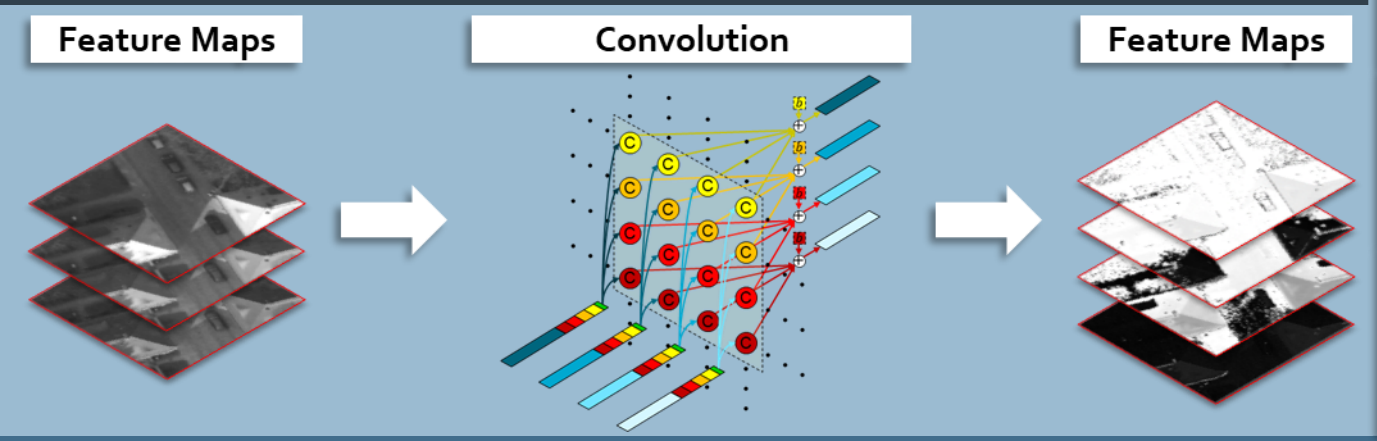
STP-H6/SSIVP

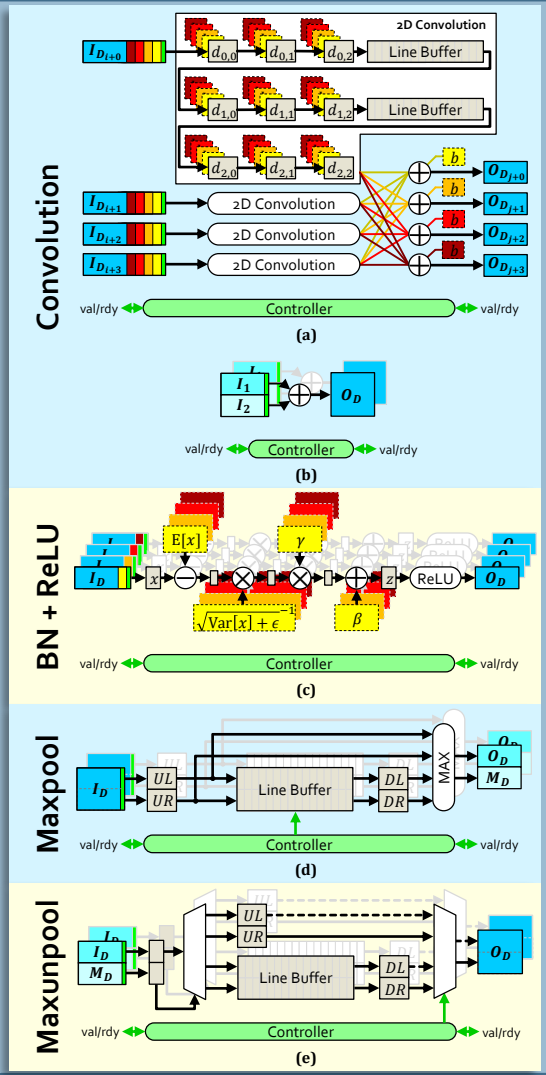


NASA
Mission-Critical Computing
NSF CENTER FOR SPACE, HIGH-PERFORMANCE, AND RESILIENT COMPUTING (SHREC)

ReCoN Accelerator

- Designed for accelerating parts of SegNet Model
  - Convolution, BatchNorm (BN) + ReLU, Maxpool, and Maxunpool
- Scalable and Parameterizable
  - Scale accelerator size to accommodate various platforms
  - Parameterize accelerator to support various mission applications
- Data-flow Optimizations
  - Process pipelines in parallel
  - Maximize input/output stream bandwidths
- Quantization Optimizations
  - Reduce area overhead at small cost of reduced accuracy

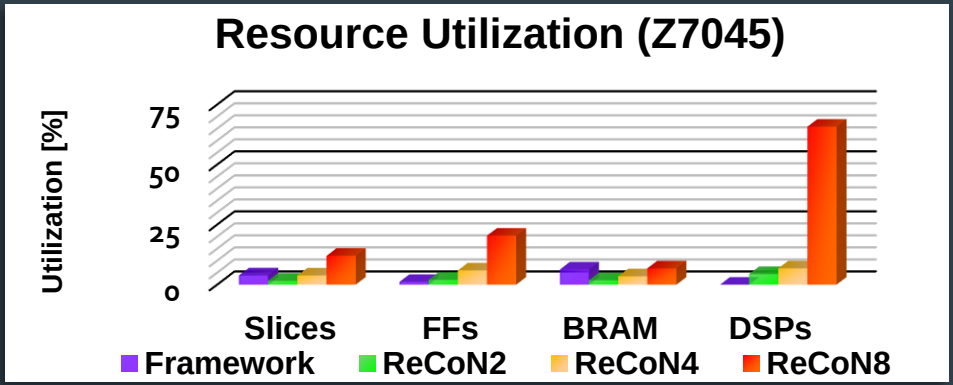


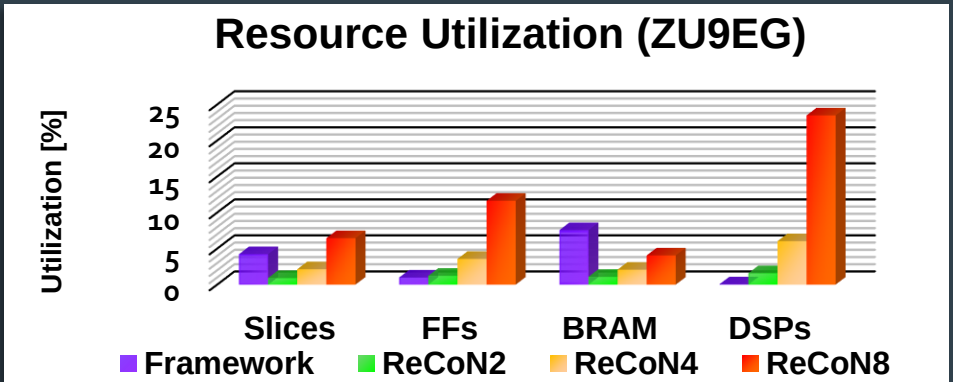


NASA
Mission-Critical Computing
NSF CENTER FOR SPACE, HIGH-PERFORMANCE, AND RESILIENT COMPUTING (SHREC)

Accuracy and Resource Utilization

Prediction Accuracy/Error	Net (7.38M weights)	Net½ (1.85M weights)	Net¼ (465K weights)
Prediction Accuracy (RGB)	90.17%	89.63%	88.30%
Prediction Accuracy (IRRG)	90.00%	89.95%	88.92%
Accelerator Error (floating-point)	0.00%	0.00%	0.00%
Accelerator Error (Q9.16)	0.73%	0.40%	0.30%
Accelerator Error (Q9.18)	0.72%	0.39%	0.29%





Subsystem	Slices (218600)	FFs (437200)	BRAM (545)	DSPs (900)
Framework	3.87%	1.14%	6.33%	0.00%
ReCoN2	1.62%	2.03%	1.84%	4.44%
ReCoN4	3.81%	5.94%	3.49%	16.89%
ReCoN8	12.02%	20.39%	6.79%	65.78%

Subsystem	Slices (274080)	FFs (548160)	BRAM (912)	DSPs (2520)
Framework	4.23%	1.00%	7.57%	0.04%
ReCoN2	0.93%	1.24%	1.09%	1.59%
ReCoN4	2.14%	3.57%	2.08%	6.03%
ReCoN8	6.45%	11.64%	4.05%	23.49%

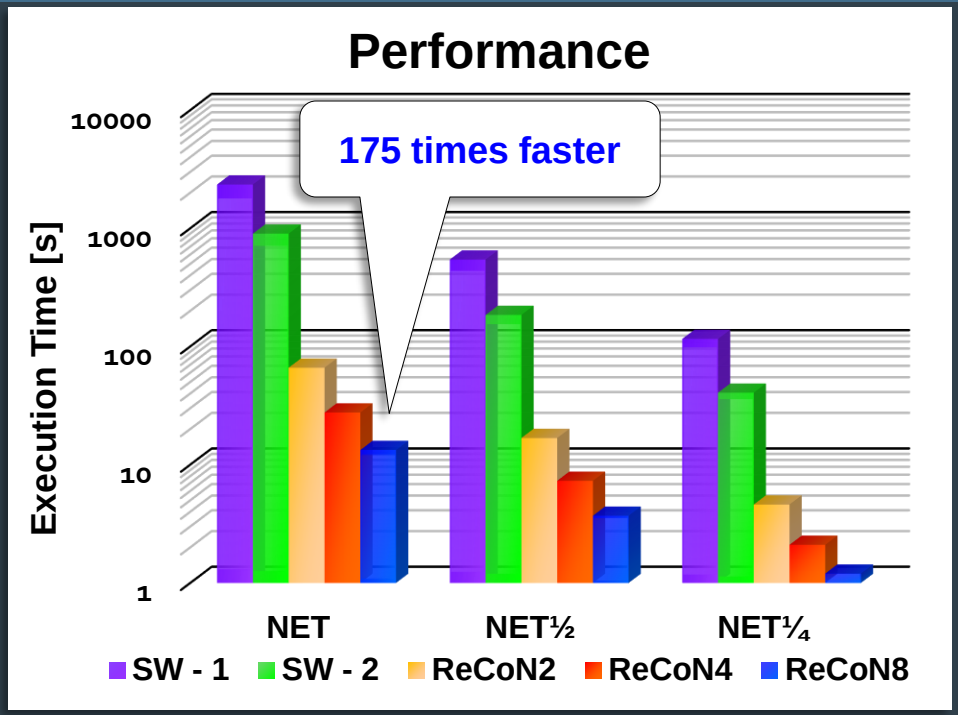
Platform: Xilinx ZC706 (Z7045)
Platform: Xilinx ZCU102 (ZU9EG)

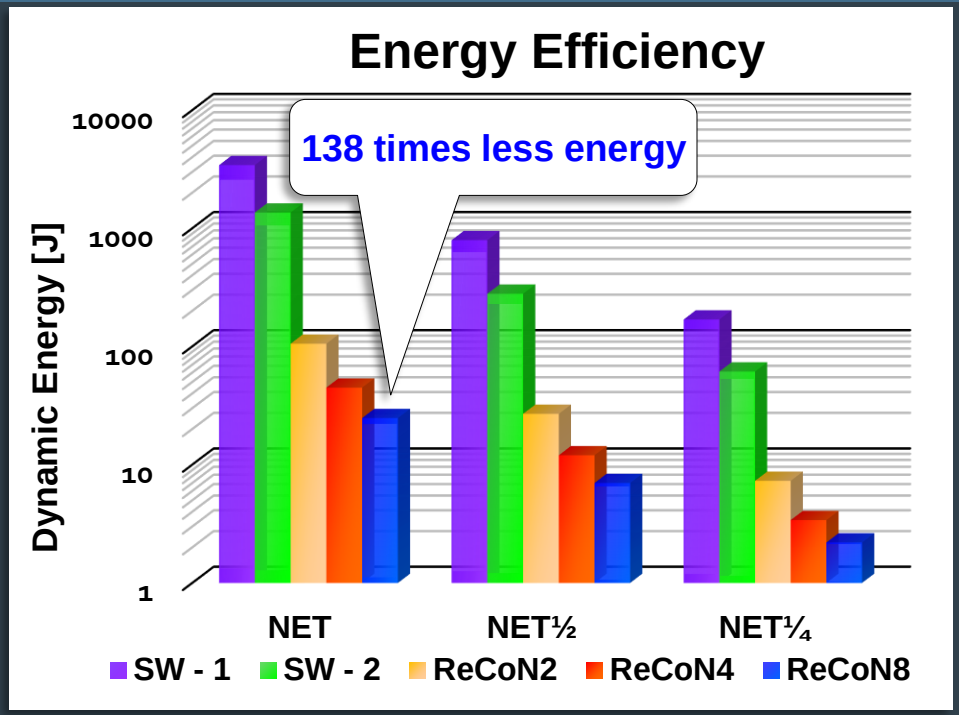
Best on CSPv1

Data: 512×512 RGB – ISPRS Potsdam

Mission-Critical Computing
NSF CENTER FOR SPACE, HIGH-PERFORMANCE, AND RESILIENT COMPUTING (SHREC)

Performance and Energy-Efficiency on Zynq SoC





	Execution Time [s]			Improvement			Dynamic Power [W]	Dynamic Energy [J]			Improvement		
Version	**Net**	**Net½**	**Net¼**	**Net**	**Net½**	**Net¼**		**Net**	**Net½**	**Net¼**	**Net**	**Net½**	**Net¼**
SW (1 thread)	2345.1	543.2	115.7	1.0	1.0	1.0	1.46	3424.0	793.0	169.0	1.0	1.0	1.0
SW (2 threads)	900.9	184.6	40.6	2.6	2.9	2.9	1.51	1360.0	279.0	61.3	2.5	2.8	2.8
SW (4 threads)	–	–	–	–	–	–	–	–	–	–	–	–	–
ReCoN2	66.0	16.8	4.6	35.5	32.3	25.3	1.60	105.3	26.9	7.3	32.5	29.5	23.2
ReCoN4	**27.6**	**7.3**	**2.1**	**85.0**	**73.7**	**54.5**	**1.62**	**44.8**	**12.0**	**3.4**	**76.4**	**66.2**	**49.0**
ReCoN8	**13.3**	**3.7**	**1.2**	**175.8**	**145.0**	**96.2**	**1.86**	**24.8**	**7.0**	**2.2**	**138.0**	**114.0**	**75.5**

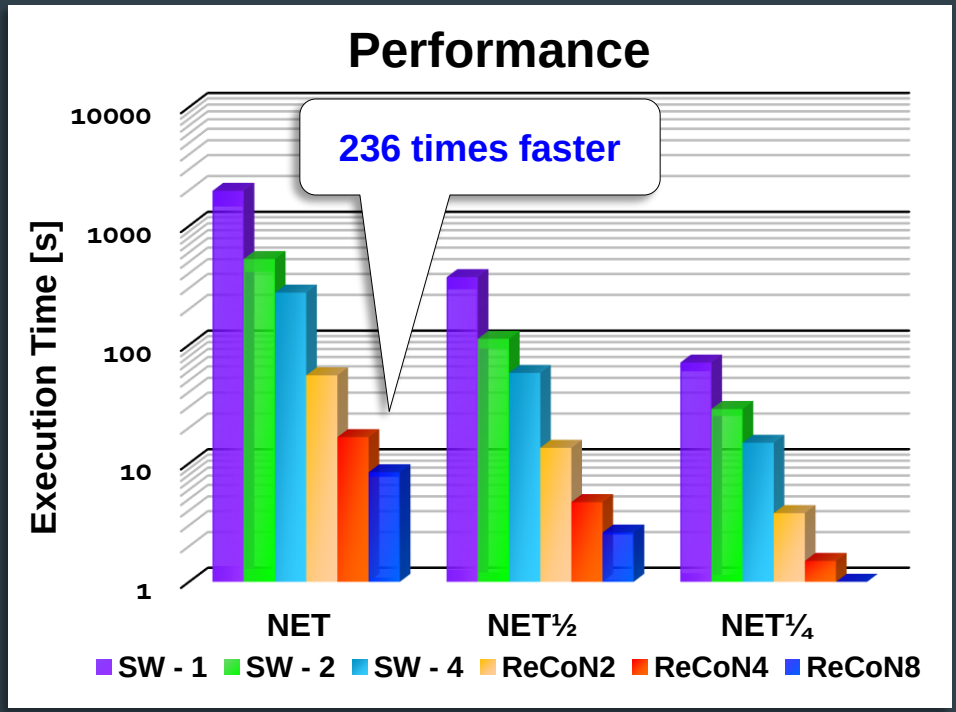
Platform: Xilinx ZC706 (Z7045)
SW/CPU: dual-core ARM Cortex-A9 with NEON, 667MHz, -O2

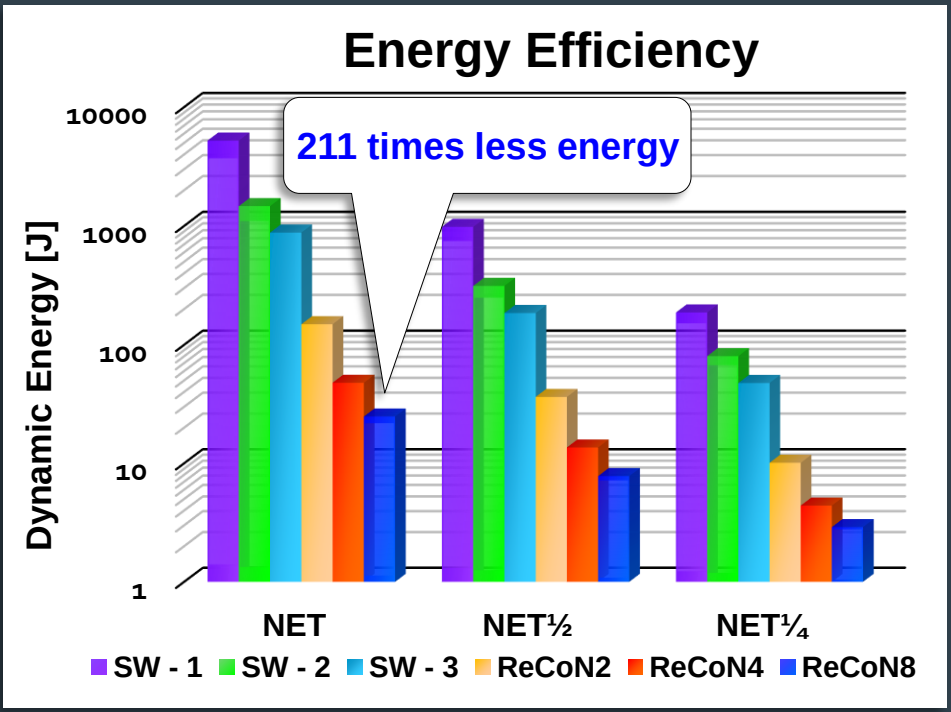
Best on CSPv1 | Best Outcome

Data: 512×512 RGB

Mission-Critical Computing
NSF CENTER FOR SPACE, HIGH-PERFORMANCE, AND RESILIENT COMPUTING (SHREC)

Performance and Energy-Efficiency on Zynq UltraScale+ MPSoC





	Execution Time [s]			Improvement			Dynamic Power [W]	Dynamic Energy [J]			Improvement		
Version	**Net**	**Net½**	**Net¼**	**Net**	**Net½**	**Net¼**		**Net**	**Net½**	**Net¼**	**Net**	**Net½**	**Net¼**
SW (1 thread)	1973.4	370.5	70.7	1.0	1.0	1.0	2.65	5230.0	982.0	187	1.0	1.0	1.0
SW (2 threads)	526.2	112.1	28.7	3.8	3.3	2.5	2.79	1468.0	313.0	80.2	3.6	3.1	2.3
SW (4 threads)	274.3	57.8	14.9	7.2	6.4	4.8	3.20	879.0	185.0	47.6	6.0	5.3	3.9
ReCoN2	55.2	13.5	3.8	35.7	27.5	18.8	2.70	148.8	36.3	10.1	35.1	27.1	18.5
ReCoN4	16.6	4.7	1.5	118.9	78.7	46.6	2.88	47.8	13.6	4.4	109.0	72.4	42.9
ReCoN8	**8.4**	**2.6**	**1.0**	**236.1**	**141**	**72.1**	**2.97**	**24.8**	**7.8**	**2.9**	**211.0**	**126.0**	**64.3**

Platform: Xilinx ZCU102 (ZU9EG)
SW/CPU: quad-core ARM Cortex-A53 with NEON, 1.2GHz, -O2

Best Outcome
Data: 512×512 RGB

Mission-Critical Computing
NSF CENTER FOR SPACE, HIGH-PERFORMANCE, AND RESILIENT COMPUTING (SHREC)

Conclusions

- **Major Challenges Lie Ahead**
  - Escalating app demands in harsh environments
  - Tightening constraints of platform, budget, process
  - ***Necessitates adeptly doing more with less***
- **Hybrid Computing for On-board Processing**
  - Enable intelligent spacecraft
    - Accelerate compute-intensive deep-learning applications
    - Improve application performance and energy-efficiency
  - Deep learning on-board
    - Enable and extend spacecraft capabilities for science and defense applications

Questions?

Dr. Alan D. George (PI), Department Chair, R&H Mickle Endowed Chair, Professor of ECE, and NSF SHREC Center Director
NSF Center for Space, High-Performance, and Resilient Computing (SHREC)
ECE Department, University of Pittsburgh
1238D Benedum Hall, 3700 O'Hara Street Pittsburgh, PA 15261 412-624-9664
Email: **alan.george@pitt.edu** or **alan.george@nsf-shrec.org**

Mission-Critical Computing
NSF CENTER FOR SPACE, HIGH-PERFORMANCE, AND RESILIENT COMPUTING (SHREC)

www.nsf-shrec.org