Neural Acceleration for General-Purpose Approximate Programs

Hadi Esmaeilzadeh
Adrian Sampson
Luis Ceze
Doug Burger

University of Washington

Microsoft Research

University of Washington

MICRO 2012

sampa

Many applications

do not need

perfect correctness.

augmented reality

voice recognition

physical simulation

signal processing

Many applications
do not need
perfect correctness.

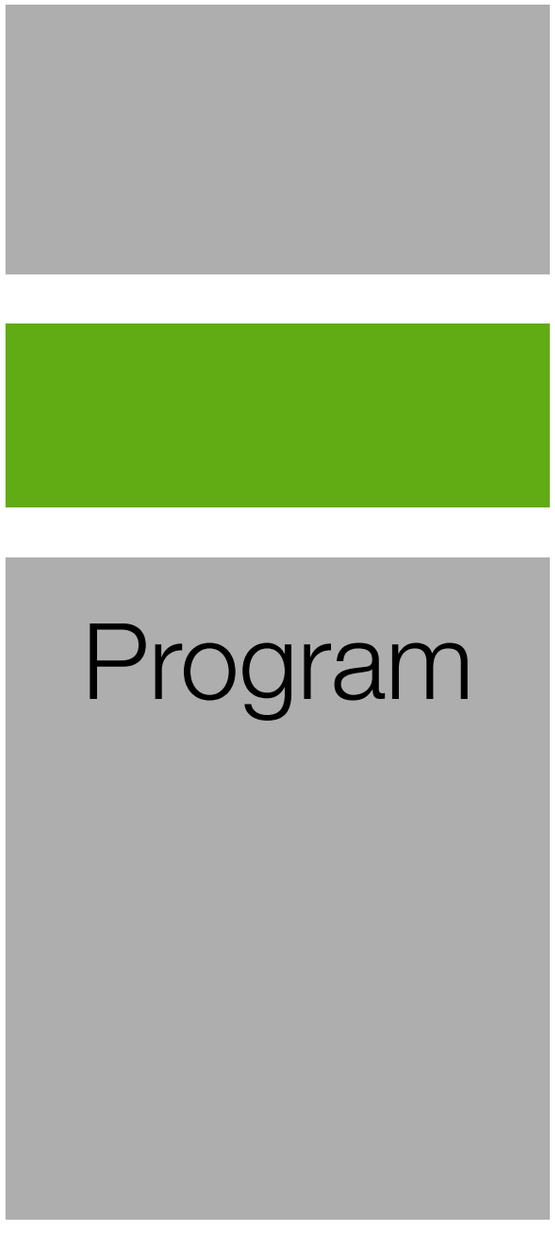
sensor data

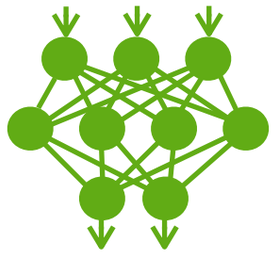
image rendering

machine learning

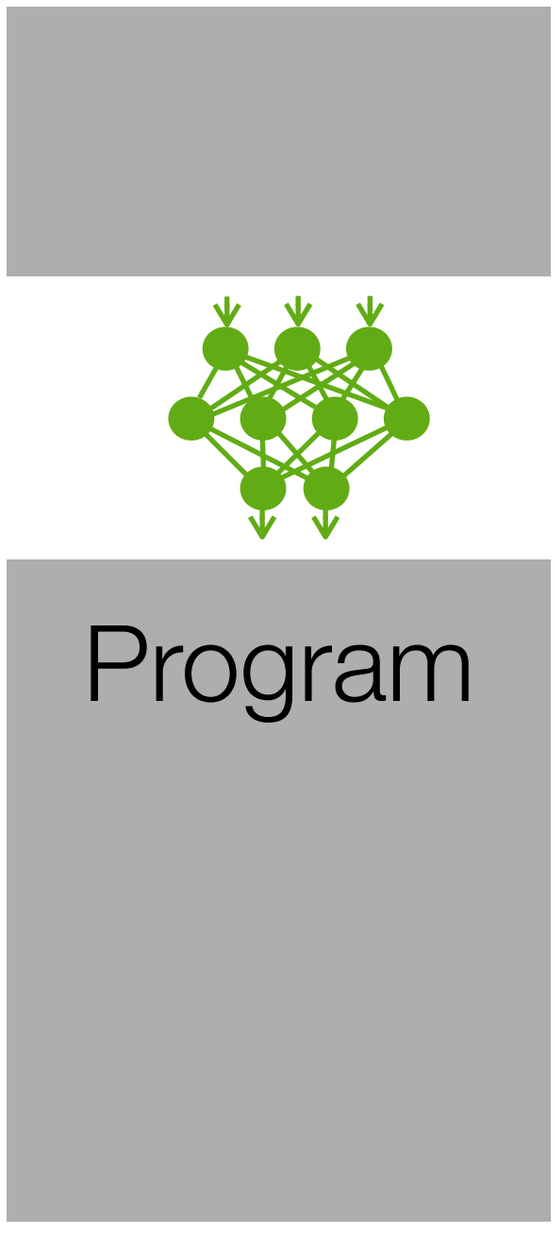
search

Neural networks can efficiently approximate functions from programs written in conventional languages.

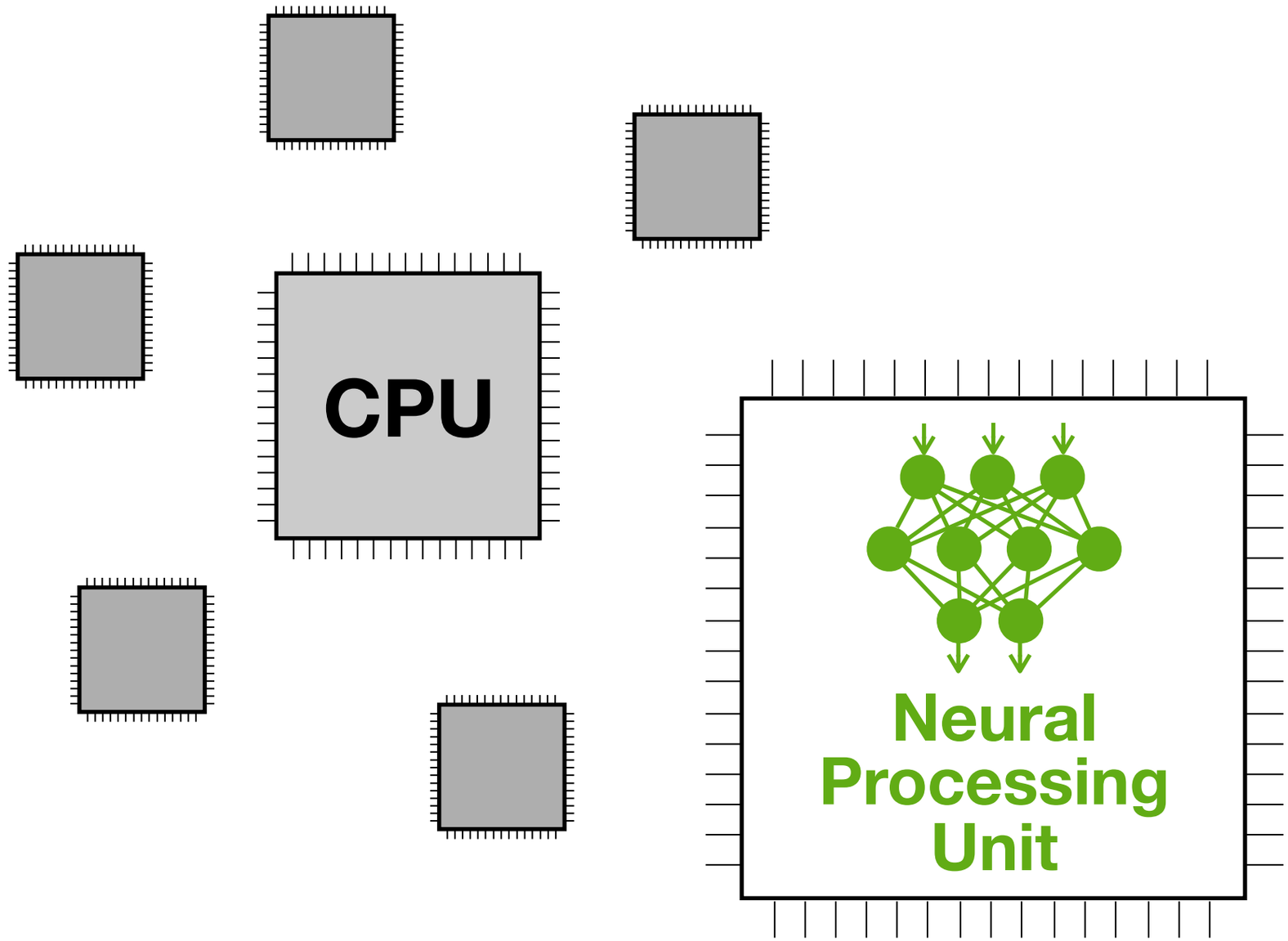




Neural networks can efficiently approximate functions from programs written in conventional languages.



Neural networks can efficiently approximate functions from programs written in conventional languages.


CPU
Neural
Processing
Unit

2.3x

speedup across 6 benchmarks

3.0x

average energy reduction

<10%

output quality loss

Last session

Wednesday at 11:30am

Neural Acceleration for General-Purpose Approximate Programs

Hadi Esmaeilzadeh
Adrian Sampson
Luis Ceze
Doug Burger

University of Washington

Microsoft Research

University of Washington

MICRO 2012

sampa