A Framework for Safe Execution of User-Uploaded Algorithms

Benchmark as an Online Service (BOS)

Toni Tan, René Weller, Gabriel Zachmann
Computer Graphics and Virtual Reality Research Lab
University of Bremen, Germany

Web3D, 2-4 November 2022

Benchmarks are Important for Scientific Progress

- Replication Crisis
  - Lack of instructions, missing data/codes, incompatible hardware/software, etc
  - In computer graphics [Bonneel et al, 2020] :
    - 374 papers from SIGGRAPH 2014, 2016, and 2018
    - 151 software packages available (133 source codes, 18 pre-compiled softwares),
    - 68 source codes need modification to work, 19 technical issues, and 5 hardware issues
- Open Benchmark for reproducible and comparable results

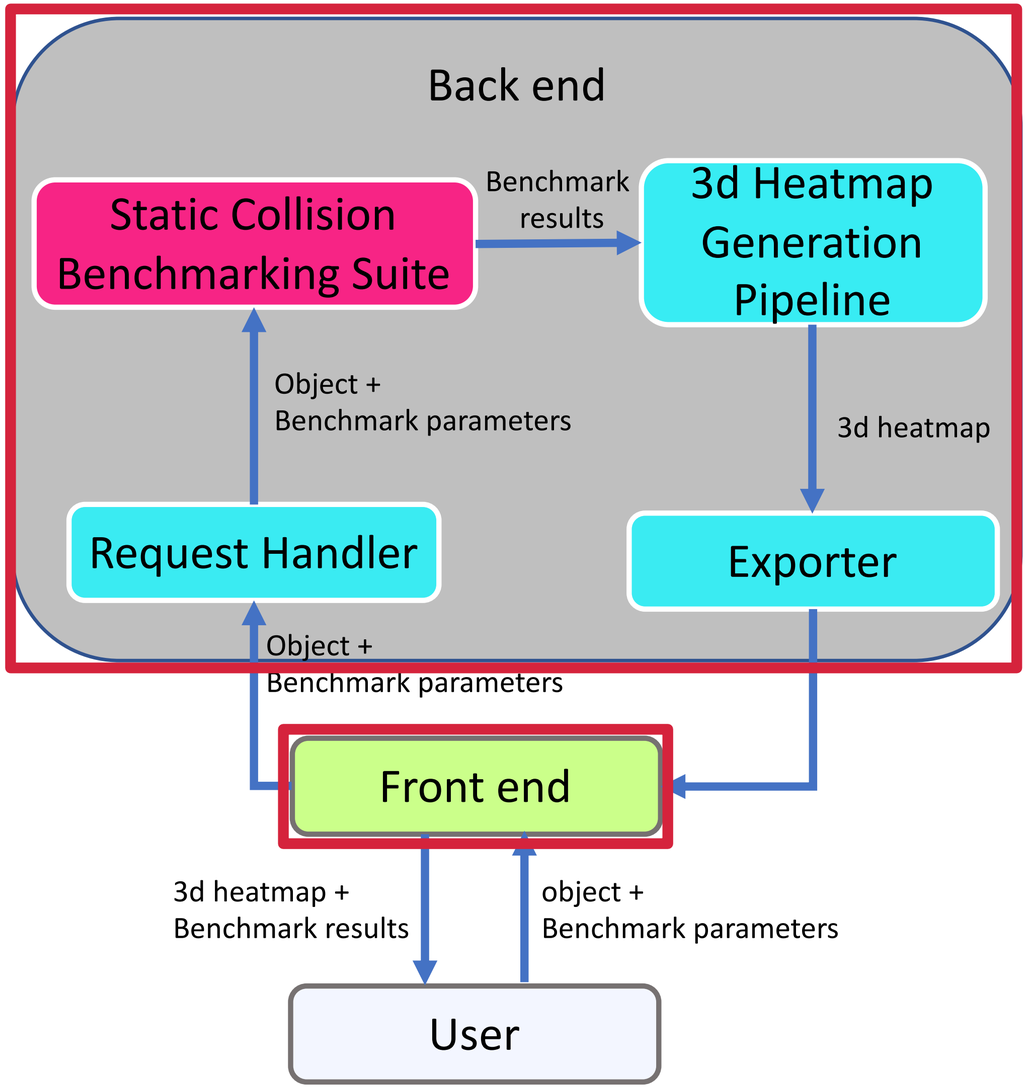
Open Benchmark - Requirements

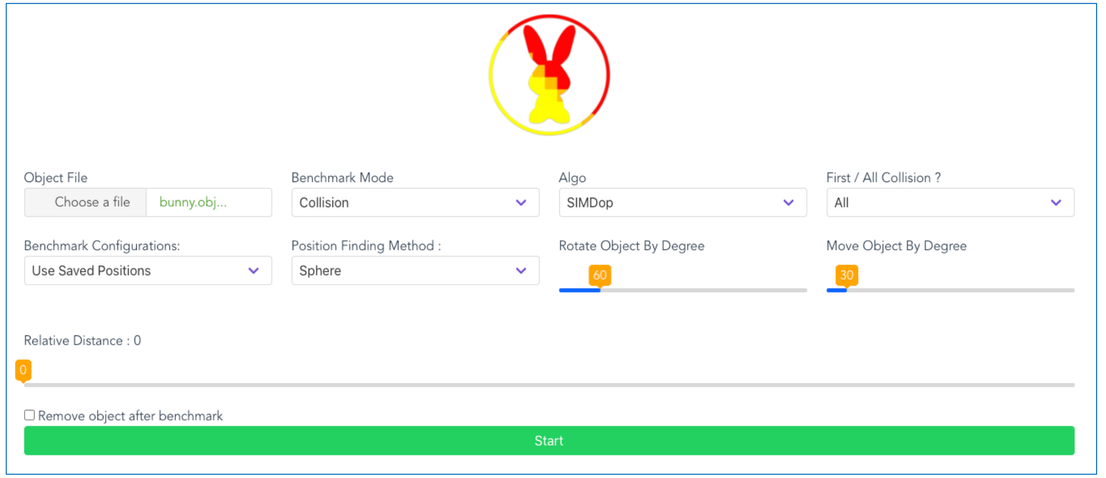
- Reproducible and comparable results
  - Same hardware and software environment
- Easy-to-use
- Sustainability
  - New hardware or software

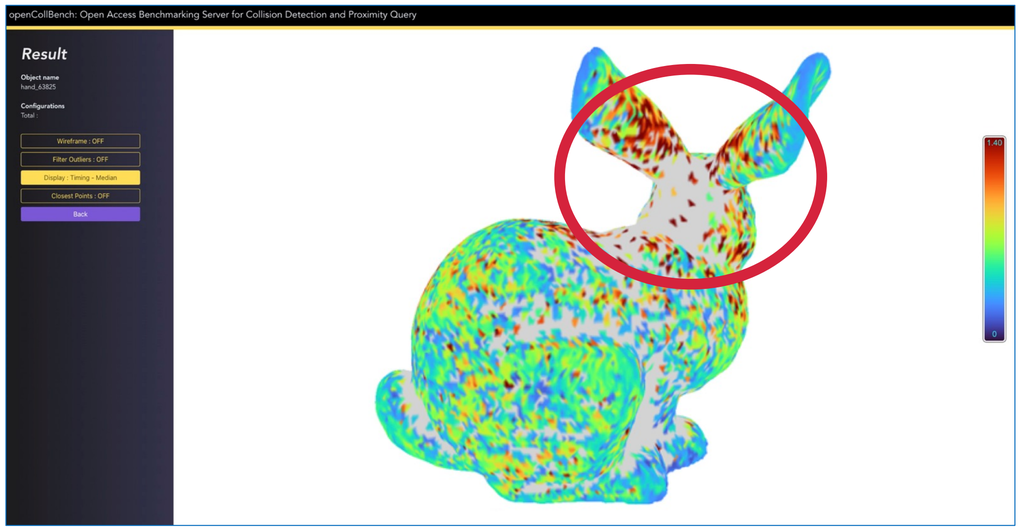
Benchmark as an Online Service (BOS)

For collision detection algorithms, i.e., OpenCollBench [Tan et al. 2019]

BOS: OpenCollBench







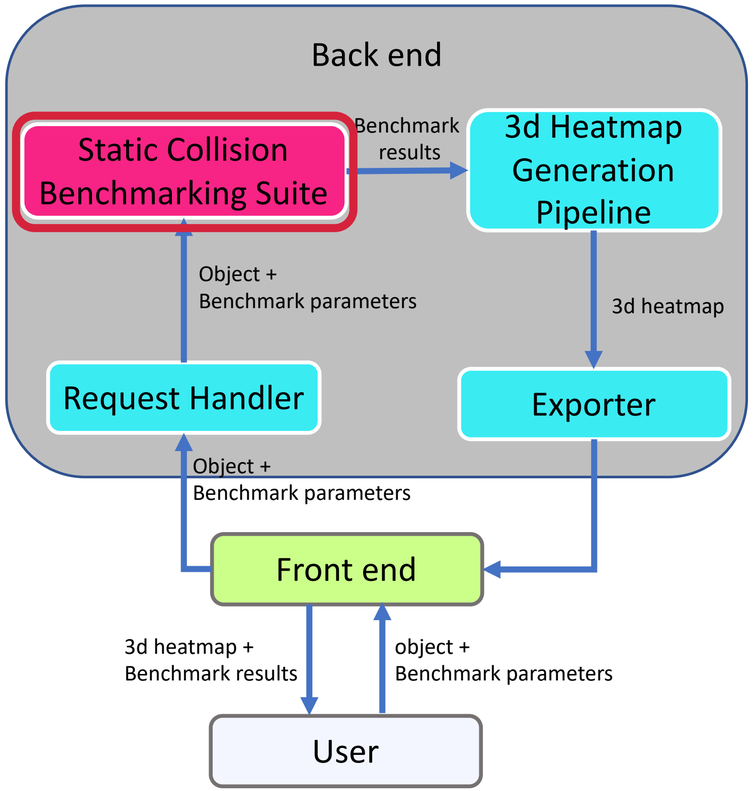
BOS Challenge: Integration of New Algorithms

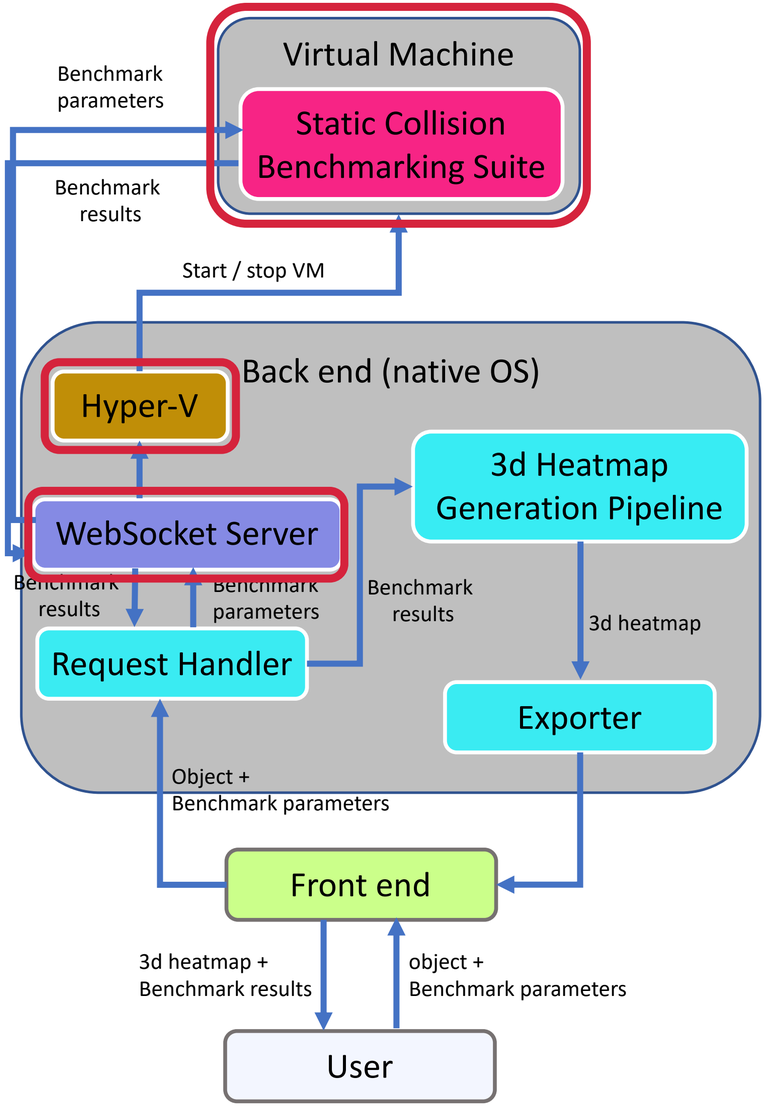
- New algorithms as wrapper *Dynamic Link Library* (DLL)
- Risk of running unkown user-uploaded code
  - Analyzing and validating code is not trivial
- Manual integration is time-consuming & problematic
  - e.g., in work-in-progress development or non-disclosure agreements.

Our Contribution

Guarantee security as well as reproducibility and comparability of BOS by executing user-uploaded algorithms in virtualization environment

OpenCollBench: Extended System Overview

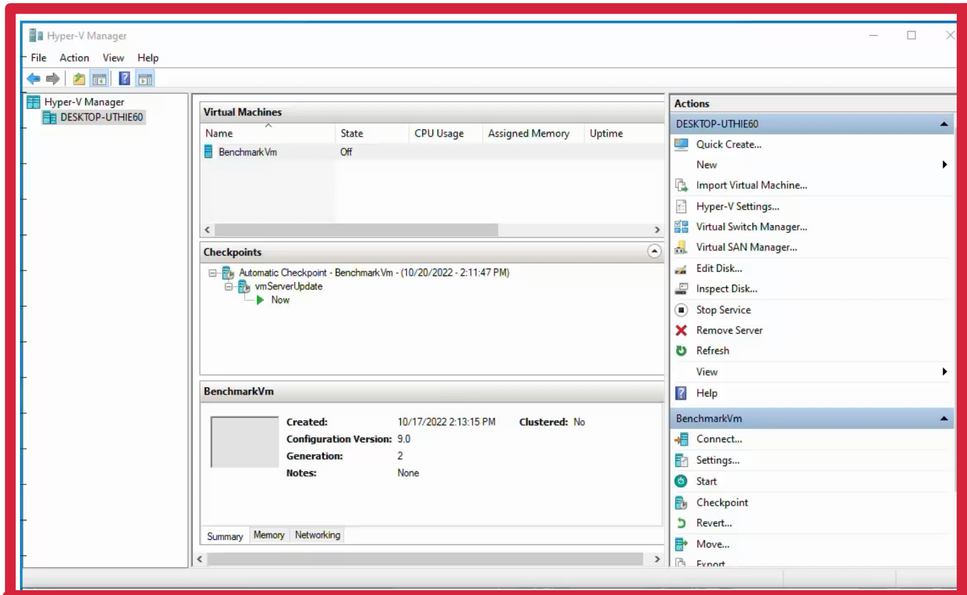


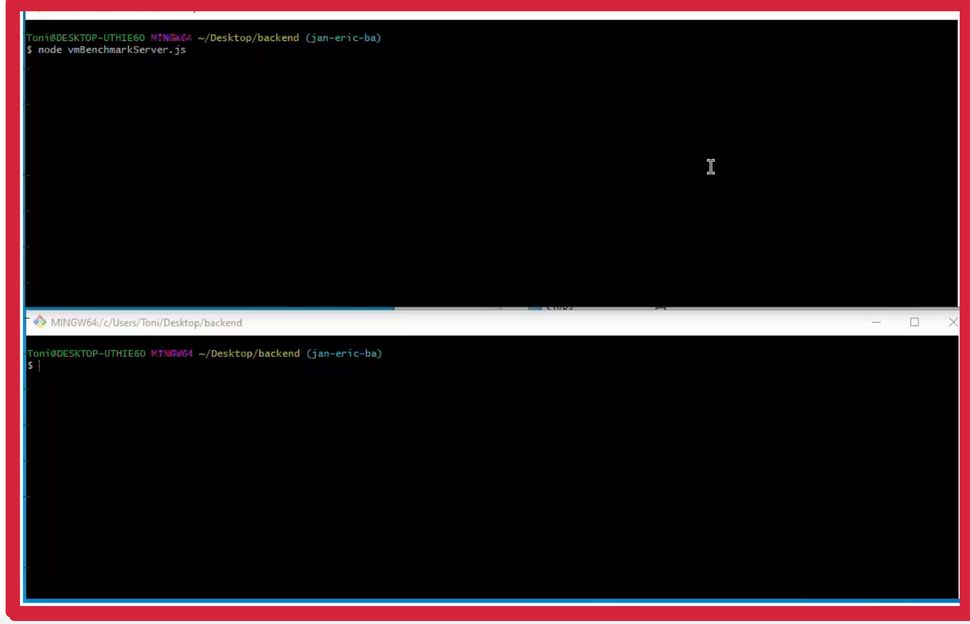


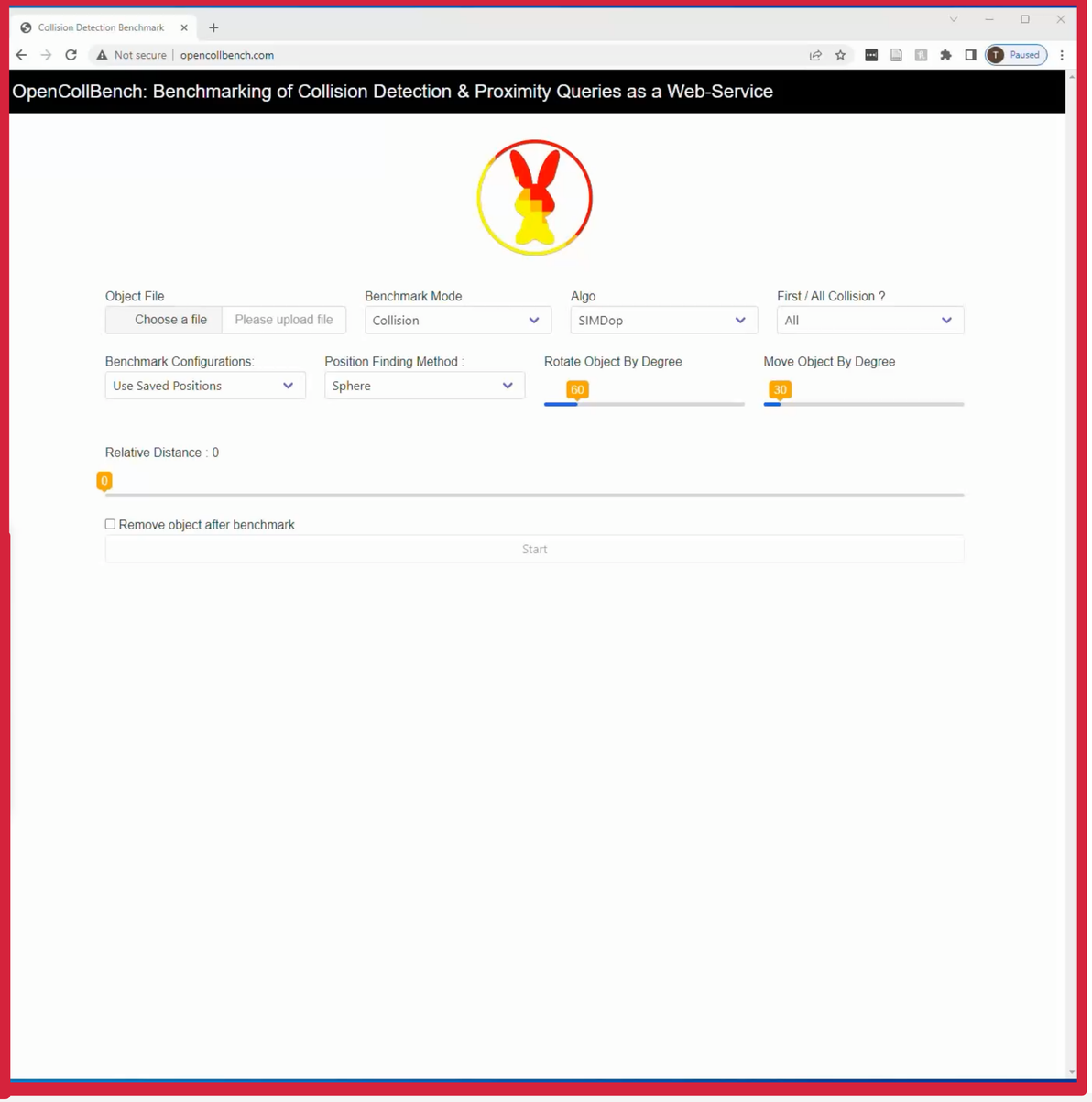
OpenCollBench: Extended System Overview

- Hyper-V supports SIMD (AVX-512) and GPU passthrough
  - Required by CD algorithms, i.e., simdop
- Run 1 VM at one time
  - Comparability
  - Avoid users from overloading host system
- VM always start from initial state
  - Prevent system changes by OS or previous algorithms

OpenCollBench: Demo





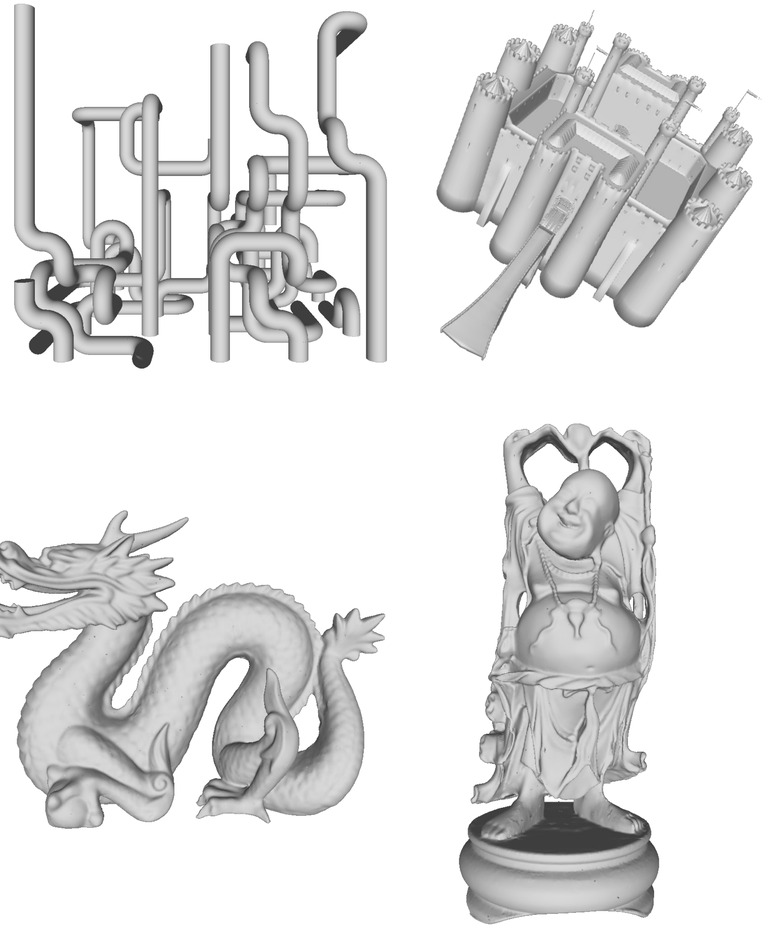


Benchmark Results: Reliability

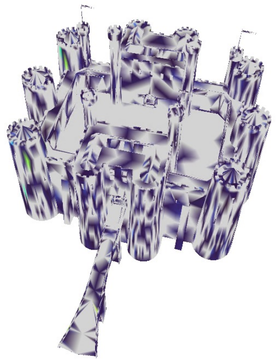
Is the benchmark results reliable?

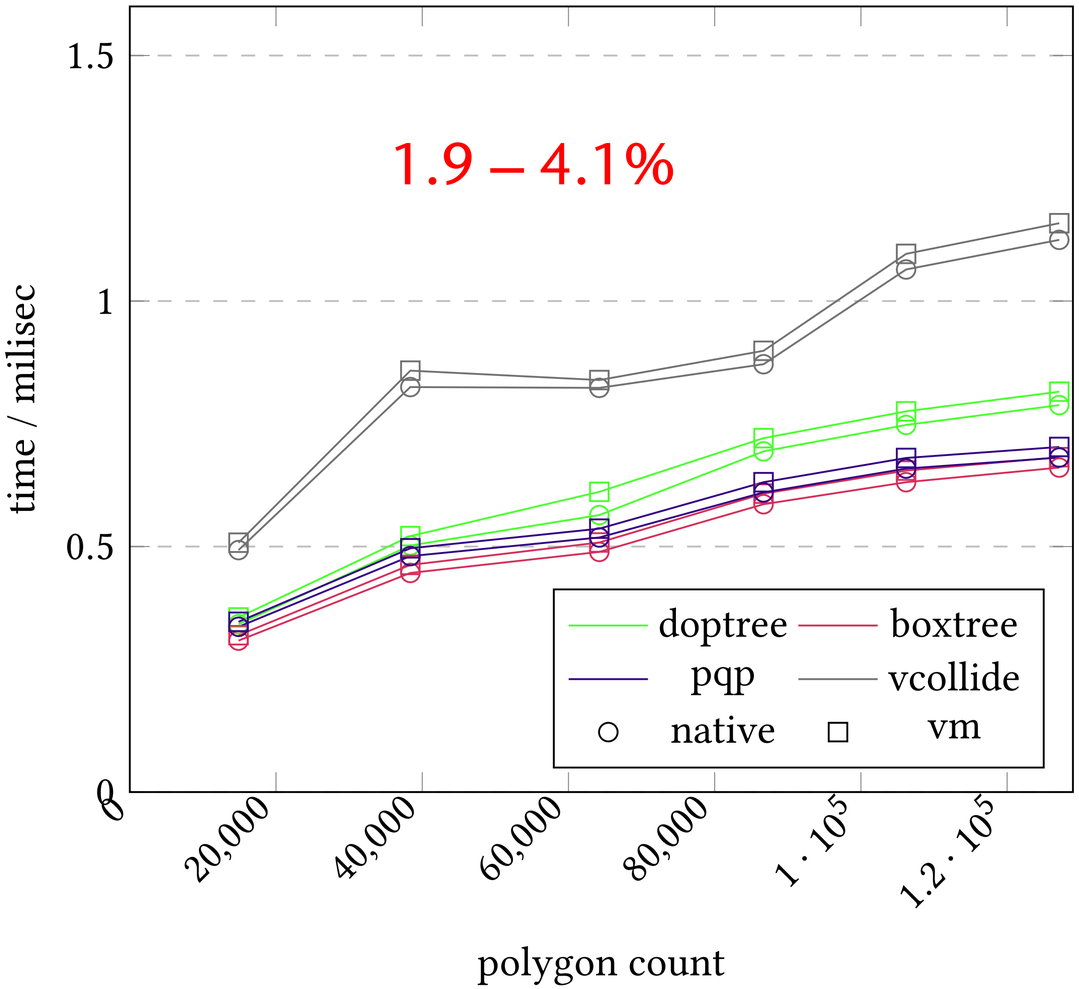
Benchmark Setup

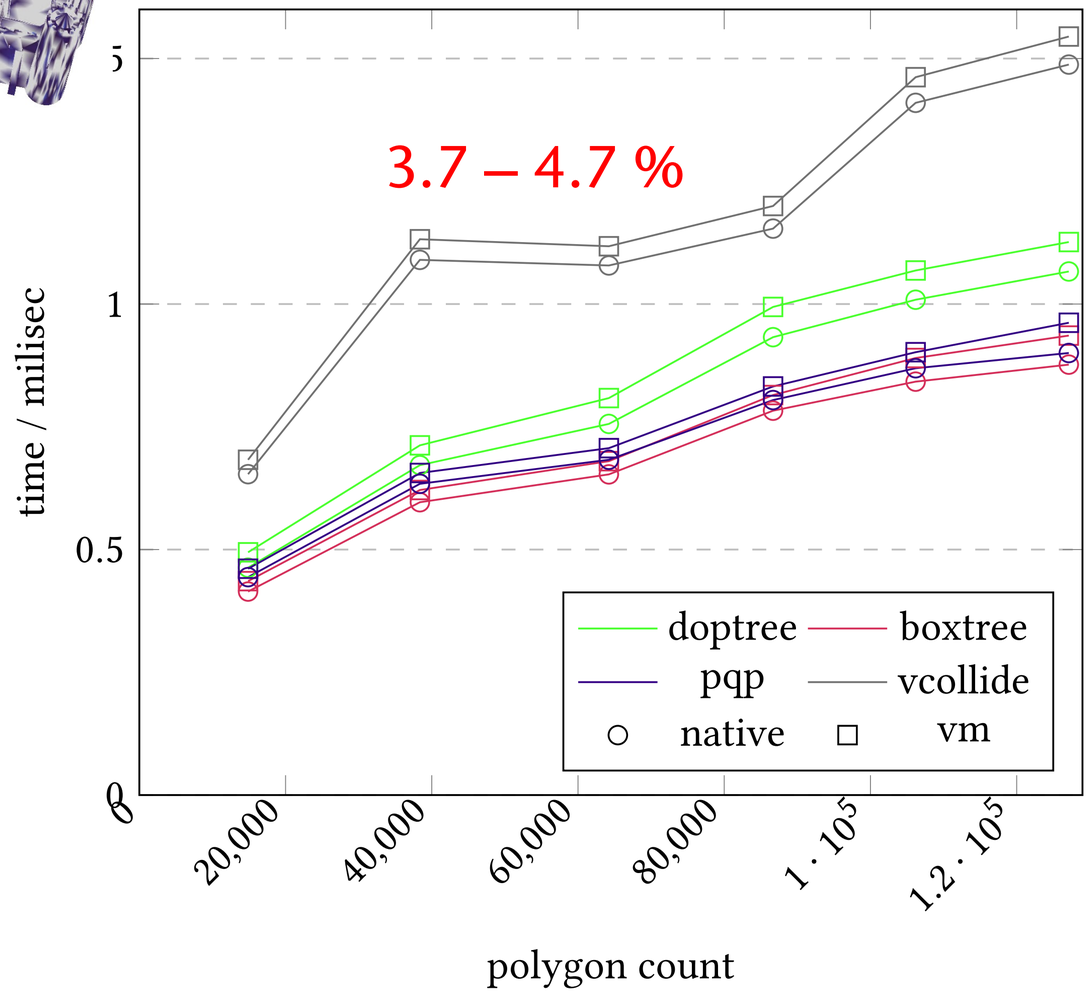
- Benchmark in native and vm
- Objects with various polygon count from 10k to 120k
- Ca 200k different configurations at distance 0.0
  - Up to 20 minutes to finish
- Common CD algorithms, e.g., doptree, boxtree, pqp, vcollide, and simdop



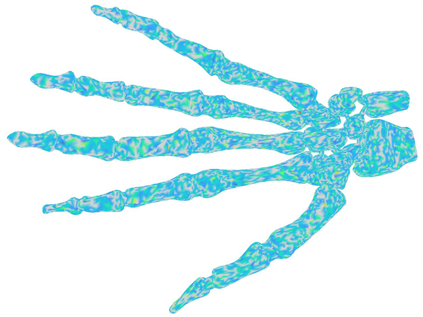
Results: Native vs VM (Different CPUs)

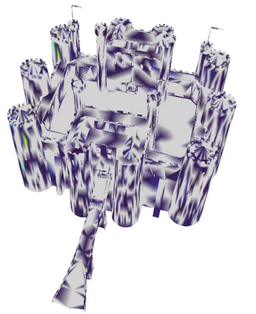


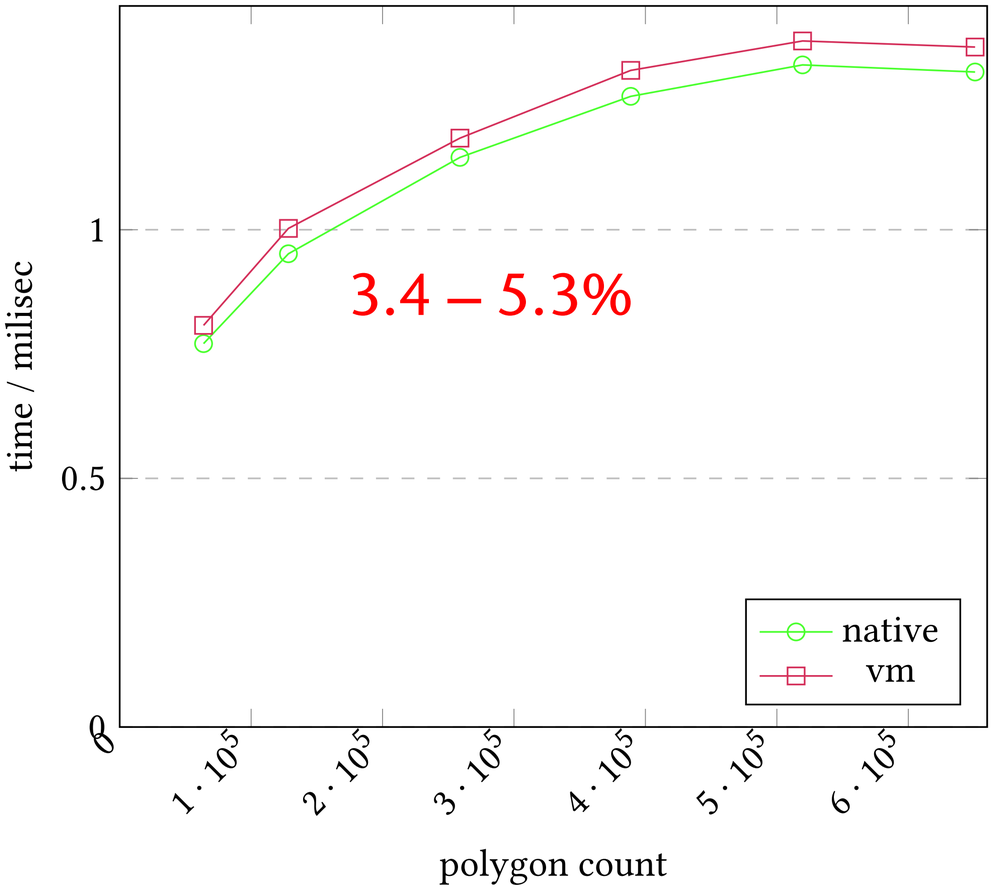


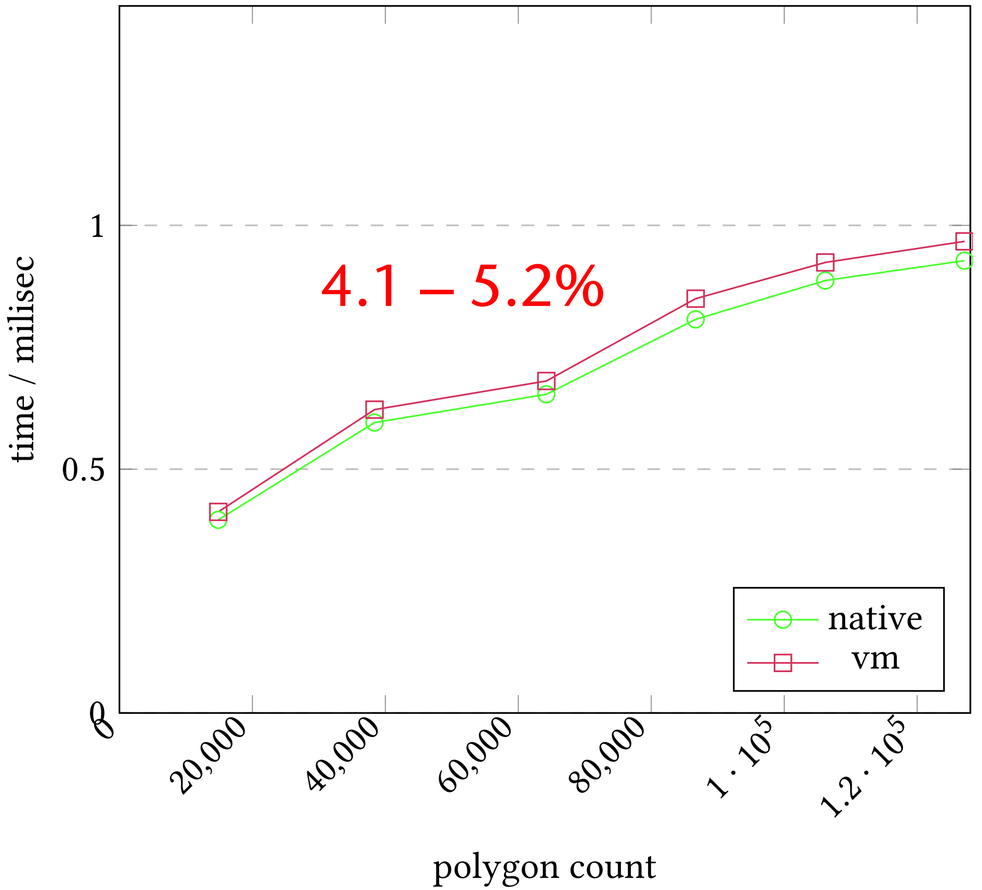


Results: SIMD (AVX-512) Performance In SIMDop

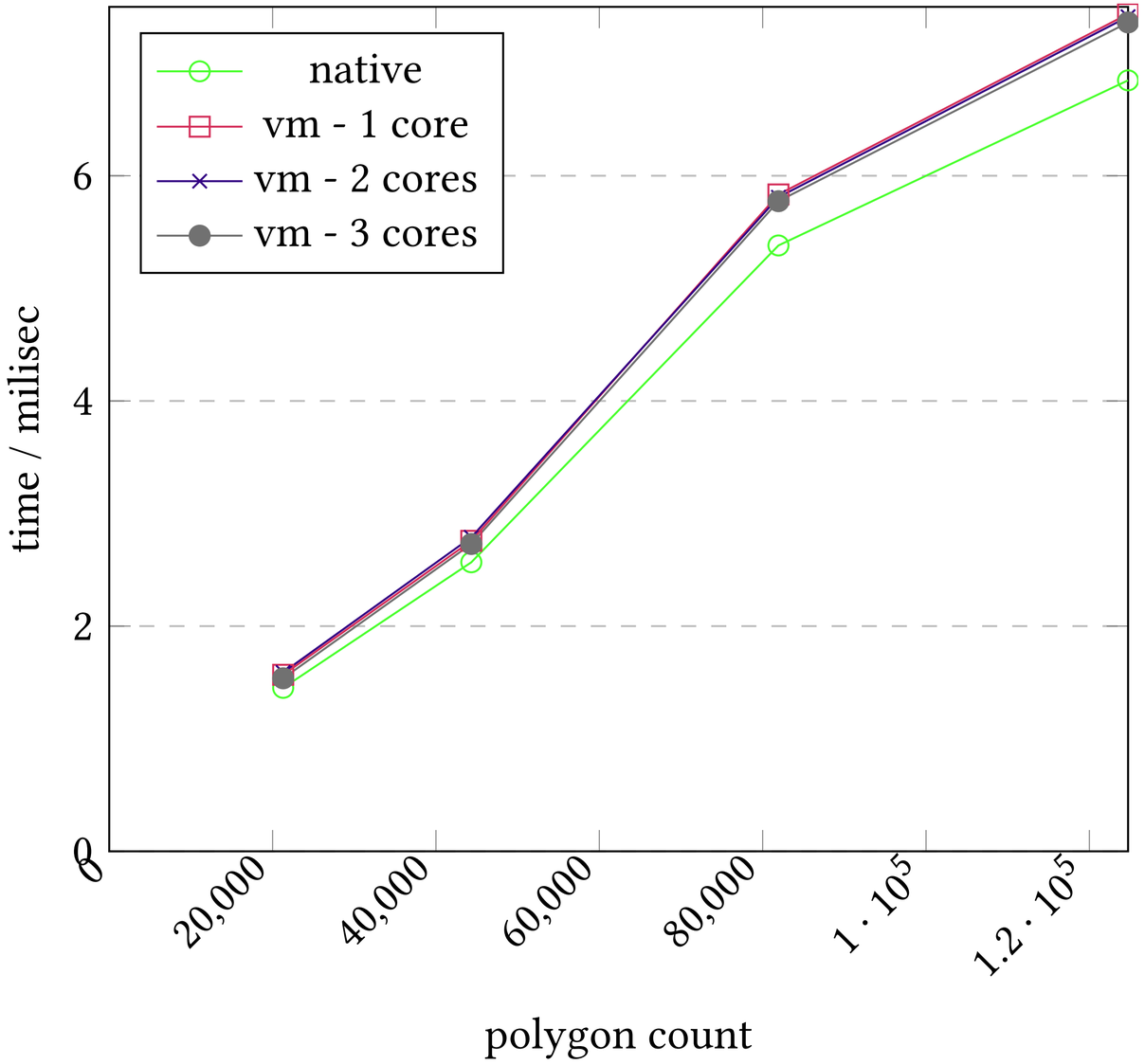


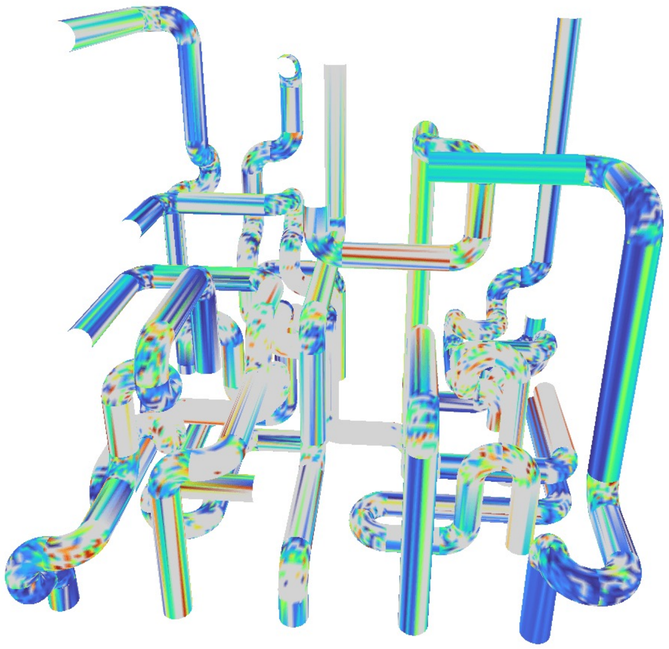






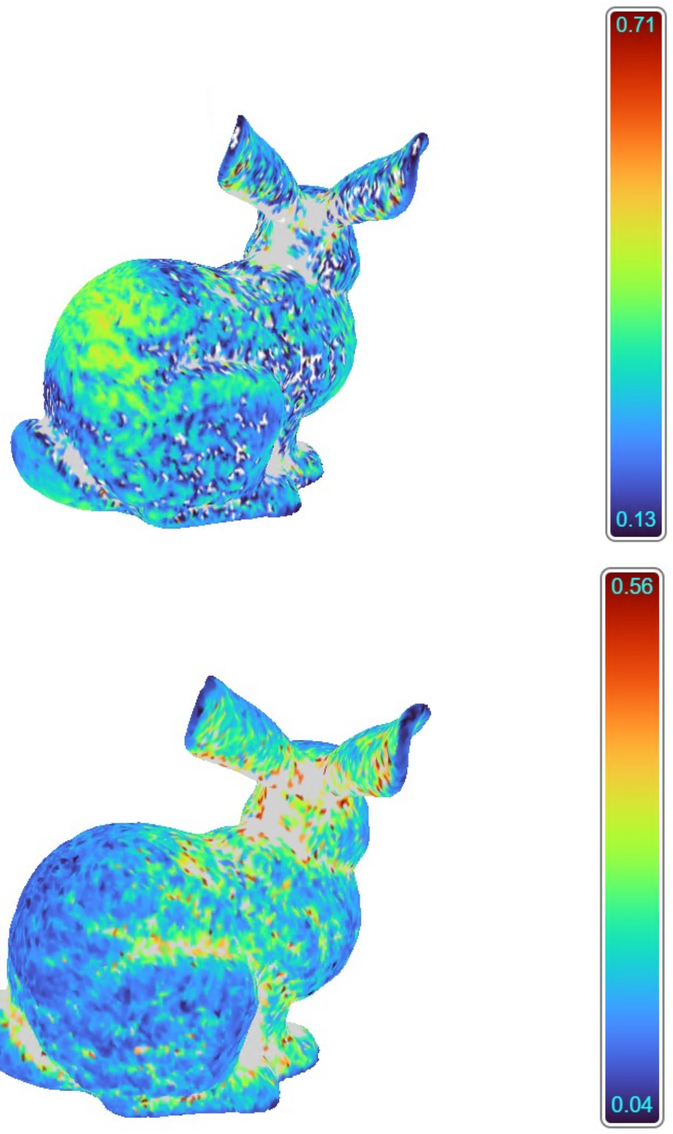
Result: VM with Different CPU Cores



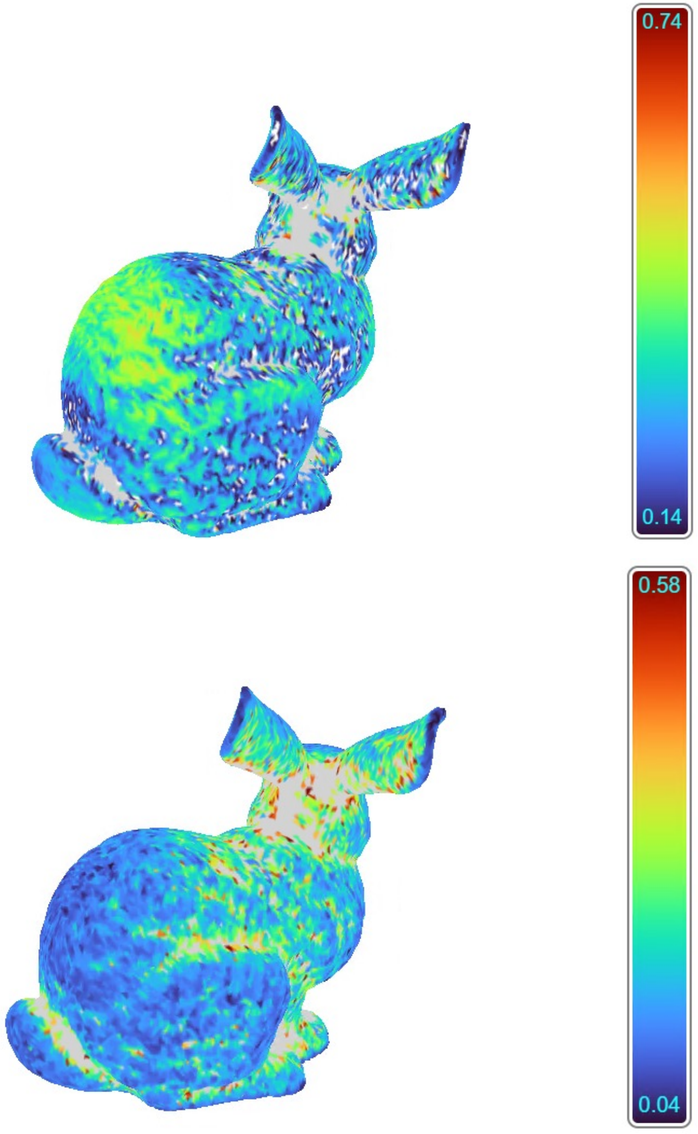


Result: Heatmap Comparison (Boxtree & Vcollide)

Boxtree (native)



Boxtree (vm)



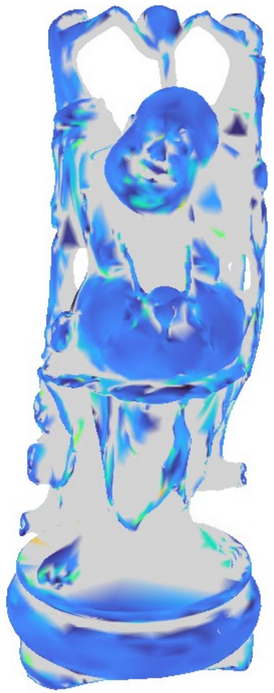
Vcollide (native)

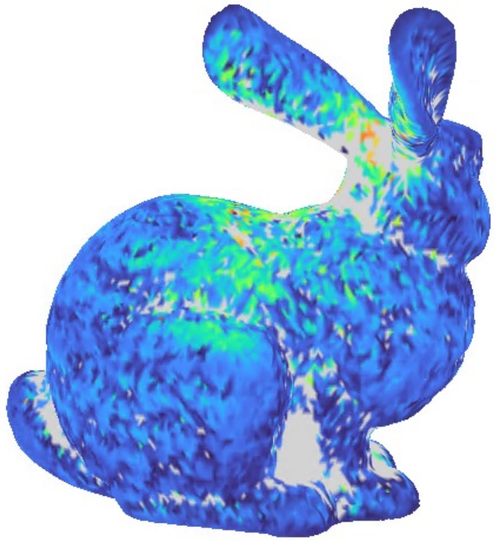
Vcollide (vm)

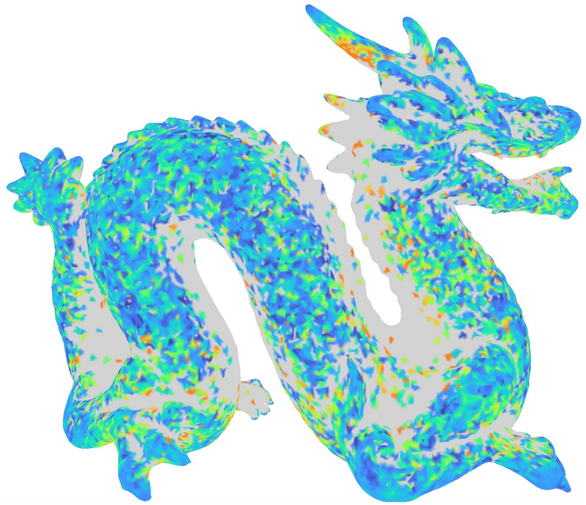
Conclusions

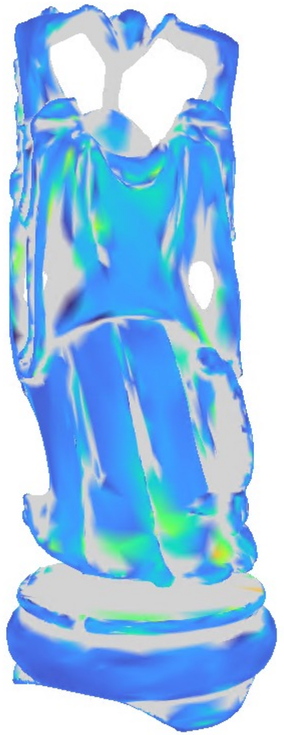
- A framework for the secure execution of user-uploaded algorithms in virtualization environment for BOS
- The entire process is automated and implemented on top of exisiting BOS Framework, i.e., OpenCollBench
- Benchmark results are reliable
- Future work:
  - Implementation of REST endpoint
    - Automatic benchmark within continuous integration pipeline, or
    - As a plugin for IDE to directly assess effects of change in algorithms during development

Thank You!



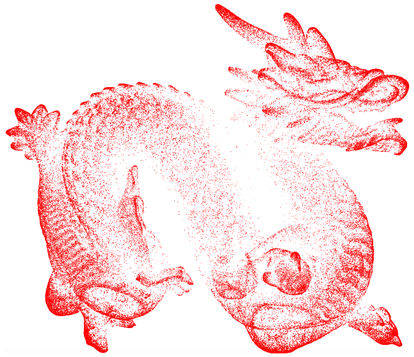






Toni Tan, René Weller, Gabriel Zachmann

{toni, weller, zach}@cs.uni-bremen.de



References

- Bonneel, Nicolas, David Coeurjolly, Julie Digne, and Nicolas Mellado. "Code replicability in computer graphics." *ACM Transactions on Graphics (TOG)* 39, no. 4 (2020): 93-1.