# CSCI-582 Project Report: "Inference Performance Study on Oak-D Camera"

Cameron Legg
Colorado School of Mines
Golden, CO, USA

Ben Hempelmann
Colorado School of Mines
Golden, CO, USA

## 1 INTRODUCTION

In this project, we did a study with the Oak-D camera, operating on the Nvidia Orin NX. We vary parameters such as the number of shaves (cores), inference threads, and video size. We then compare the results.

## 2 TARGETED DOMAIN [2 PTS]

Explain the application or the domain you have presented on and experimented with.

## 3 MOTIVATION: THE NEED FOR ACCELERATION [2 PTS]

Why do we need beyond CPU processing for this domain and/or application? Motivate it.

## 4 RELATED WORK [8 PTS]

Starting with the short summary of the paper you presented, explain how other work tackled the acceleration problem for the targeted domain. For the paper you presented, state the key idea(s), solution(s) and contribution(s) of the paper(s) and briefly talk about the technical details of their approach.

You don't need to read additional papers, however, in your summary above, please cite at least 3 other related works that you learned about while you are reading the paper you presented.

### 4.1 Current Limitations

Briefly talk about weaknesses of current approaches, if available.

## 5 PROJECT DESCRIPTION [2 PTS]

Explain/introduce the project you did. Including:

- Selected application/benchmark, the specific kernels that are being accelerated. The details of the parallelization, if available. (Don't repeat information you gave eairlier above)
- Targeted platform and accelerators
- What is/are the programming interface (s) and/or runtime framework (e.g., CUDA, tensorrt etc.) being used for acceleration? Details...

### 5.1 Challenges

What are the challenges you have faced while using this interface and how you tackled them?

## 6 EXPERIMENT SETUP [2 PTS]

### 6.1 Methodology

How did you set up your experiments? How many iterations of executions have you made? How did you eliminate the noise from your results? Etc.

### 6.2 Metrics measured and Comparisons made

What are the metrics (e.g. time, power, energy, etc.) being measured? What is being compared (e.g., applications and/or HW)? Through which methodology? What is your baseline?

## 7 RESULTS [9 PTS]

Present your results (with charts) here. For each experiment you did:

- Insert a chart/figure/table
- Explain the information that chart/figure/table gives
- Interprete results. Refer to more data, if available, to support your interpretation.

There is no min/max limit on the charts. Try to logically group the information if you could. For example, you may include one chart for time measurement, another for power/energy, and another for the breakdown of the performance (data movement overhead, etc.). I expect this to be the most detailed section of your report.

Also, please address the feedback I have given during your project presention.

## 8 LINK TO SOURCE AND OTHER RESOURCES [3 PTS]

Please insert the public repository link that includes the scripts you have used, the applications you developed (if you did), the presentations and the results. Also, if available, include the links to other tutorials you followed to set up the device and benchmarks.

## 9 PROJECT/CLASS TAKEAWAY [2 PTS]

Write a brief paragraph on what you have learned "new" regarding performance analysis and overall on "compute acceleration". Please also include your honest, brief opinion on how this new knowledge could help your future career. Focus only on what you have learned. I will use this information to evaluate and improve the learning objectives of the class.

## 10 FEEDBACK [+1 PTS (GOES TO PARTICIPATION)]

In addition to the anonymous and official class feedback (which I strongly encourage you to do), please include any other feedback and suggestions (format, content, evaluation, project, etc.) to improve the class.

## REFERENCES