

**2017-06-29**

Tom Rochette <tom.rochette@coreteks.org>

July 24, 2025 — [daae079c](#)

## **0.1 Context**

## **0.2 Learned in this study**

## **0.3 Things to explore**

# **1 Problems faced**

- How to deal with loading and batching huge amount of data, more particularly in the form of images?
  - Loading thousands of images directly from the filesystem is efficient due to a lot of system calls
  - It seems straightforward to pack these images into more concise structures, such as numpy arrays and using compressed files such as npz
  - However, how does one deal with loading all this data at training time, such that 10 GB of compressed data does not equal 20 GB of RAM used all throughout training?

# **2 Overview**

# **3 See also**

# **4 References**