Lossy compression for lossless prediction

EECS Seminar: Advanced Topics in Machine Learning

Romain Graux

March 15, 2022

Motivation

10^{21} - 10^{23} bytes data collected per year

Motivation

10^{21} - 10^{23} bytes data collected per year

$\rightarrow$ But most data is processed by algorithms performing **downstream tasks**.

Motivation

10^{21} - 10^{23} bytes data collected per year

$\rightarrow$ But most data is processed by algorithms performing **downstream tasks**.

Motivation

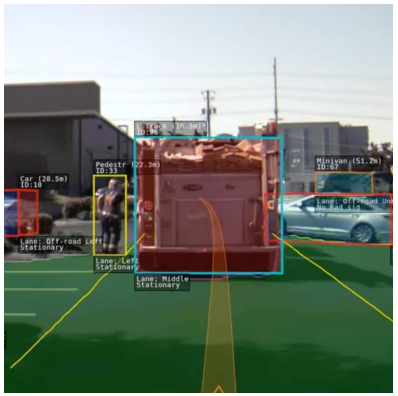
10^{21} - 10^{23} bytes data collected per year

$\rightarrow$ But most data is processed by algorithms performing **downstream tasks**.

Motivation

10^{21} - 10^{23} bytes data collected per year

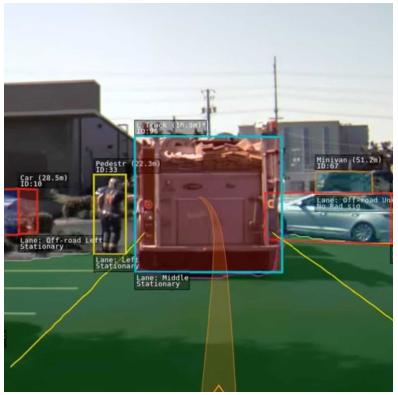
$\rightarrow$ But most data is processed by algorithms performing **downstream tasks**.



Motivation

10^{21} - 10^{23} bytes data collected per year

$\rightarrow$ But most data is processed by algorithms performing **downstream tasks**.



Motivation

10^{21} - 10^{23} bytes data collected per year

$\rightarrow$ But most data is processed by algorithms performing **downstream tasks**.

Yet current compressors optimize high **perceptual** fidelity

Motivation

10^{21} - 10^{23} bytes data collected per year

$\rightarrow$ But most data is processed by algorithms performing **downstream tasks**.

Yet current compressors optimize high **perceptual** fidelity

- Stores too much not needed information
- Does not ensure good task performance

Motivation

10^{21} - 10^{23} bytes data collected per year

$\rightarrow$ But most data is processed by algorithms performing **downstream tasks**.

Yet current compressors optimize high **perceptual** fidelity

- Stores too much not needed information
- Does not ensure good task performance

Source

Motivation

10^{21} - 10^{23} bytes data collected per year

$\rightarrow$ But most data is processed by algorithms performing **downstream tasks**.

Yet current compressors optimize high **perceptual** fidelity

- Stores too much not needed information
- Does not ensure good task performance

Source

High bitrate

Motivation

10^{21} - 10^{23} bytes data collected per year

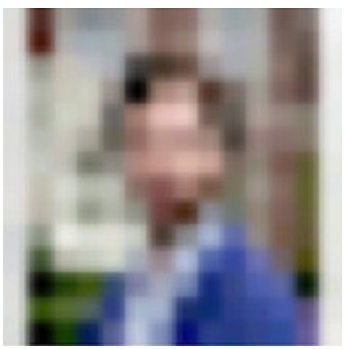
$\rightarrow$ But most data is processed by algorithms performing **downstream tasks**.

Yet current compressors optimize high **perceptual** fidelity

- Stores too much not needed information
- Does not ensure good task performance

Source

High bitrate



Low bitrate

Motivation

10^{21} - 10^{23} bytes data collected per year

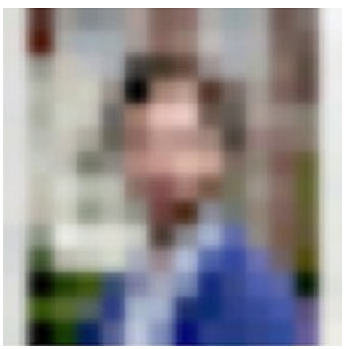
$\rightarrow$ But most data is processed by algorithms performing **downstream tasks**.

Yet current compressors optimize high **perceptual** fidelity

- Stores too much not needed information
- Does not ensure good task performance

Source

High bitrate



Low bitrate

Desired