Open problem: Structure vs data

© 2019 Philipp Krähenbühl and Chao-Yuan Wu

Vision and action

- In humans and animals
  - vision developed as a side product for action
  - no explicit supervision for vision
  - emerges from model structure (connections)

Image source: https://en.wikipedia.org/wiki/Cambrian_explosion#/media/File:Opabinia_BW2.jpg

Vision and action

- In computer vision
  - lots of data and labels
    - explicit supervision
    - emerges from data
- Classical robotics
  - Planing after computer vision

example 0: ( , Zebra)

example 1: ( , Zebra)

example 2: ( , Zebra)

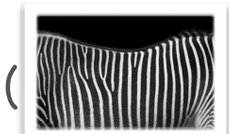
⋮

example 999: ( , Zebra)

Deep Network

Open problem

- Why this disconnect?

example 0: ( , Zebra)

example 1: ( , Zebra)

example 2: ( , Zebra)

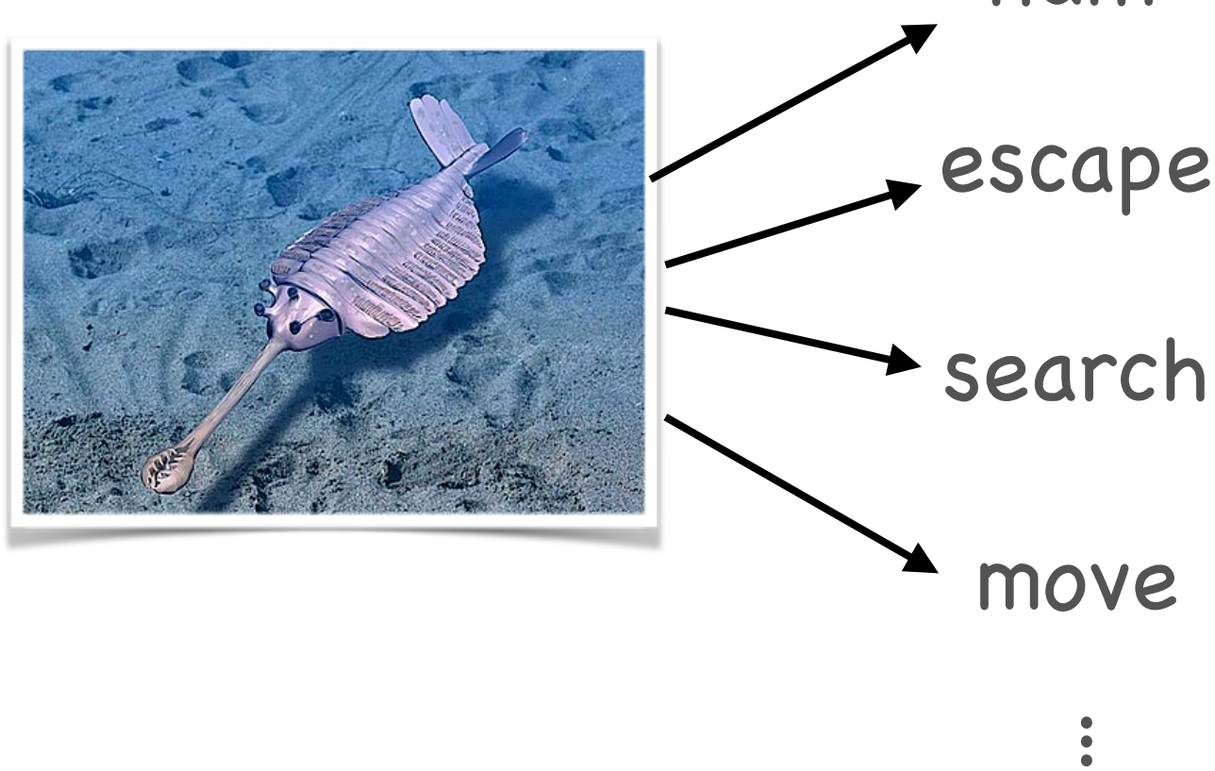
⋮

example 999: ( , Zebra)

Deep Network

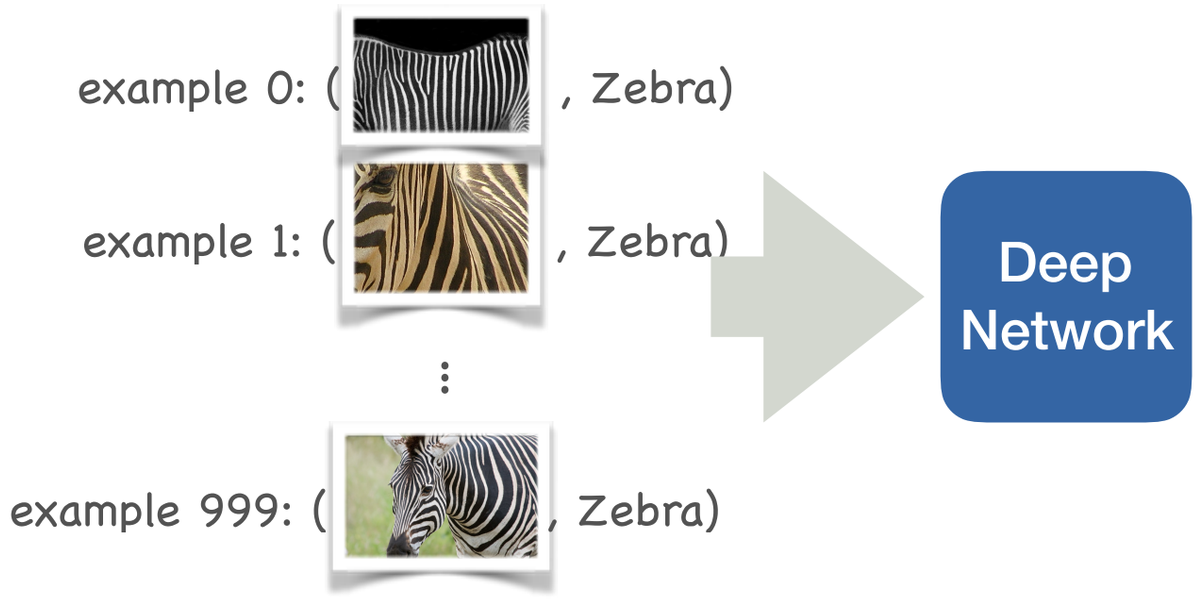
Hypothesis 1 – Too narrow tasks

- On single (narrow) task
  - Data and labels always win
- On multiple tasks
  - Generalization between tasks creates visual representation



Hypothesis 2 - Wrong models and algorithms

- Backprop + SGD biased
- Doesn't work on all tasks and architectures equally well

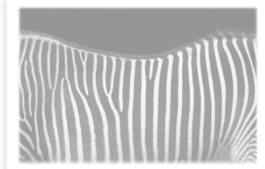


Hypothesis 3 - No evolution

- Insufficient optimization of models in outer loop
  - Meta-learning can find visual representations without much supervision
    - supervision: acting well → survival

Implications

- If any of above hypothesis are right
  - We are wasting time with labeled data

example 0: ( , Zebra)

example 1: ( , Zebra)

example 2: ( , Zebra)

example 3: ( , Zebra)

example 4: ( , Zebra)

⋮

example 999: ( , Zebra)

Hypothesis 4 - Value of labels

- Our current approach is fine
- labeled data provides abstract representation without need for evolution and massive optimization