Shu Kong
shuk@andrew.cmu.edu
Robotics Institute
Carnegie Mellon University

# Visual Perception and Learning in an Open World

**Notice**. The content is for demonstration purpose. The book is being written and under contract with the Morgan & Claypool Publishers. The final content might be different from the current one.

## Content

1. **Computer Vision in the Open World**
   a. **Introduction**
      i. **Status quo of computer vision systems**
         learning-based methods really work, visually demonstrating the AI power.
      ii. **The life cycle** of a vision system in the real world:
         *visual sensing/perception → data collection & labeling → visual learning → vision system deployment → visual sensing/perception*
      iii. **Academic computer vision research** focuses on isolated stages/components of the life cycle to explore algorithmic study, oftentimes ending up with closed-world visual modules that lack sufficient applicability in the open world.
      iv. **The world is open**, containing situations that are dynamic, vast, and unpredictable
      v. **Visual perception and learning in an open world**
   b. **Examples of open-world vision systems**
      i. **autonomous vehicles**
      ii. **vision recognition for biological data analysis**
      iii. **fashion recommendation** (Ravi)
      iv. **video surveillance** (DIVA)
      v. **examples of failure cases**
         - Tesla incident1 and incident2, Nio Inc., Uber's incident
         - Google's and Facebook's offensive predictions for underrepresented minorities
         - Visual recognition wrongfully predicts criminality
   c. **The outline of the book**
         - introduction
         - open-world challenges for visual perception and learning
         - open-world data collection and labeling
         - unpredictability of open worlds
         - visual learning in open worlds
         - testing and benchmarking of open-world vision tasks
         - outlook and conclusion
   d. **The goal of the book and the target reader**

- a collection of insightful discussions on topics related to visual perception and learning in an open-world at the Open World Vision workshop
- the target reader: researchers, students, and practitioners
- a reference for the reader w.r.t computer vision research in open-world setups
- the book is *not* a technical survey nor an introductory textbook because open-world computer vision is relatively new, and even its understanding has not reach a consensus in the community.

2. **Open-world Challenges for Vision Perception and Learning**
   Open world contains situations that are vast, dynamic, and unpredictable
   a. **Mismatch between data distribution and vision tasks**
      i. **long-tailed distributions, and rare-classes**
      ii. **data scarcity and label scarcity**
         - corner cases are rarely occurrence, hence data scarcity
         - annotations are expensive that demand expert effort, hence label scarcity
      iii. **data collection: annotation, cost, domain expertise**
         - How to collect and label data that follows a long-tailed distribution?
         - How to obtain more examples for rare cases?
         - How to better spend human efforts (ref. domain experts)?
         - How to actively and continually label data? e.g., detect novel/unkown examples
   b. **The world is dynamic**
      i. **data distribution shift and concept drift** -- continual learning and domain adaptation
      ii. **streaming data** -- early prediction (e.g., early event recognition and forecasting)
   c. **The open-world is vast and not always predictable**
      a closed-world dataset will not span the whole open world
      i. **domain gaps in both space and time**
      ii. **rare / corner cases** concerning robustness, safety, and ethics
      iii. **handling the unknowns**: open-set, out-of-distribution/anomaly, counterfactual queries
   d. **The world is not always visually/reliably sensed/measured**
      i. **Measuring the hardly-seen**, e.g., under low-light environment = -- infrared
      ii. **Measuring the geometry**, decision making and planning -- LiDAR
      iii. **Measuring the invisible**, e.g., occluded and distant objects -- audio
3. **Open-World Data Collection and Labeling**
   A naive approach: labeling random examples
   a. How to automatically **discover interesting/novel objects** in the wild?
   b. How to **collect rare examples**?
   c. **Efficient and active labeling**
   d. **Interface for labeling and analysis**
4. **Unpredictability of Open Worlds**
   a. **The unknowns: out-of-distribution, anomaly, and the open-set**

   b. **Compositionality and novelty in attributes** (e.g., overturned truck, laughing in public) counterfactual interaction
   c. **Domain gaps: in-domain training and out-of-domain experience**
   d. **Interface for human-machine interaction**
5. **Visual Learning in Open Worlds**
   a. **Fundation models and pretraining with out-of-domain data**
   b. **Self-supervised feature learning with domain data**
   c. **Class-imbalanced learning and long-tailed recognition**
   d. **Few-shot learning in the tail**
   e. **Continual learning**
6. **Testing and Benchmarking for Open-World Vision Algorithms**
   a. **Recap typical isolated tasks included in an open-world vision system**
   b. **Open-world learning** for recognition, and detection
   c. **End-to-end testing,** e.g., driving safety, planning, etc.
   d. **Realistic setups and evaluations**
      i. Realistic semi-supervised learning
      ii. Realistic continual learning
      iii. Realistic object discovery
7. **Conclusion and Outlook**
      i. Open World Vision witnesses a wide range of applications.
      ii. Learning with less-vs-more data / time / computation.
      iii. Data of multiple modalities.
      iv. Robustness of open world systems, robotics
      v. evaluation by downstream-task metrics
      vi. system view, not cognitive view, not a real AI.