CS239: Video Analytics

Lecture 1: Introduction

Ravi Netravali

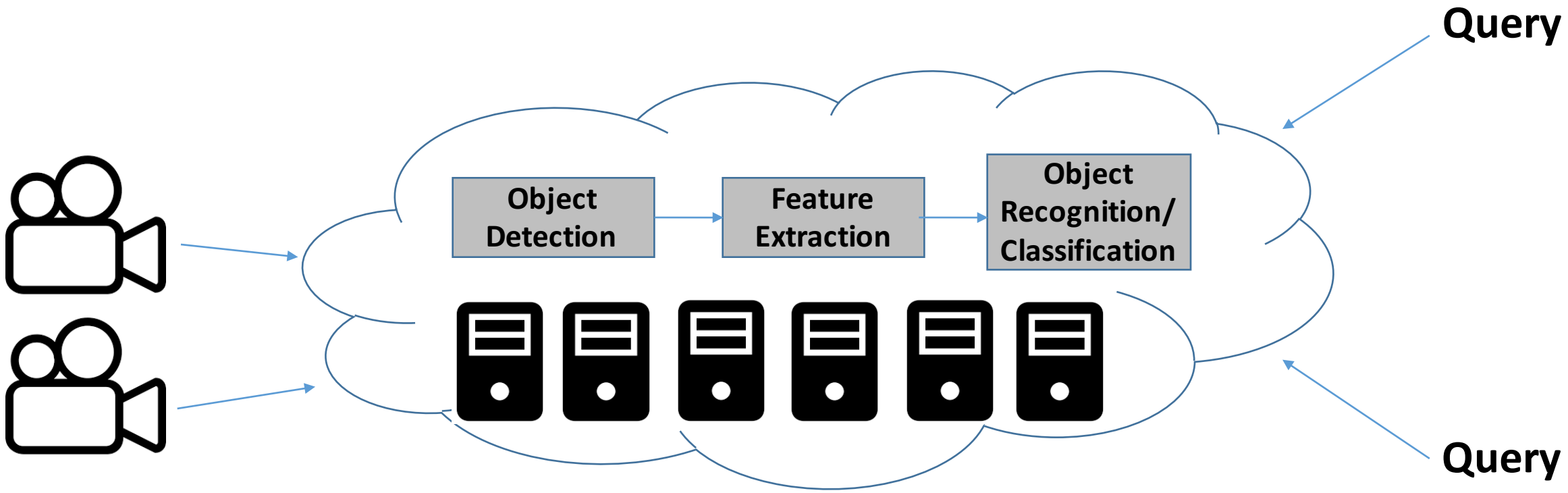
https://web.cs.ucla.edu/~ravi/

Today's Agenda

- Overview of topics

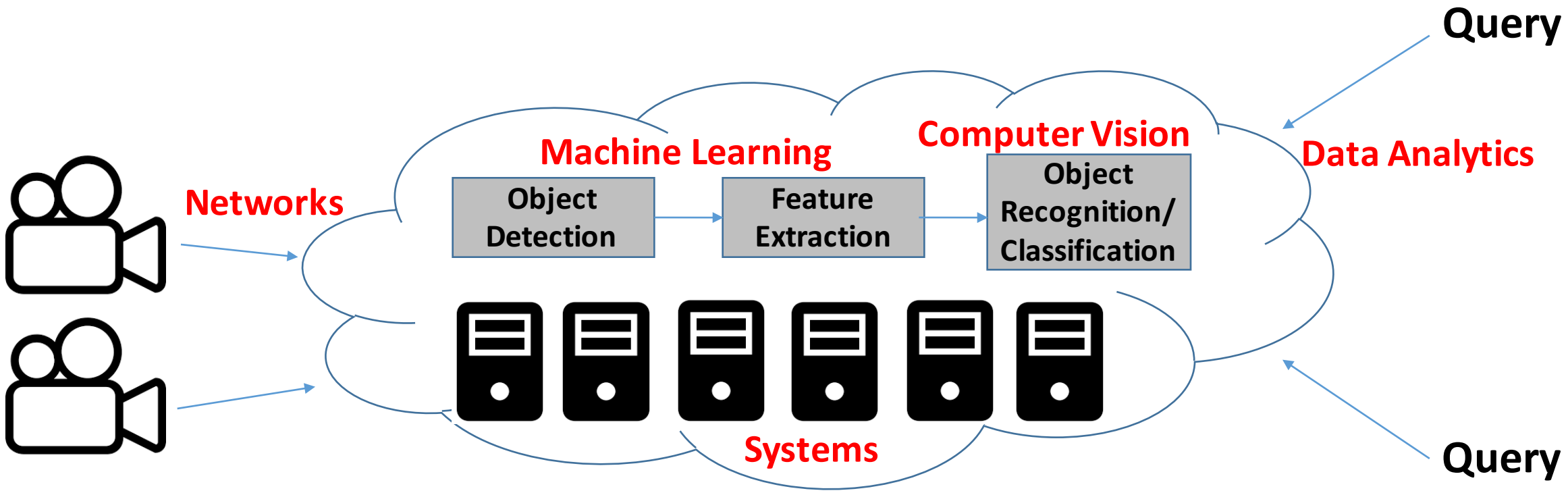
- Logistics
  - Class structure
  - Grading
  - Research project
  - Expectations and goals

Video Analytics



- Run ML pipelines (e.g., object recognition) to answer queries on live video
- Goals: accurate and low-latency query results

Video Analytics



Applications

- **Real Time (Live)**
  - Traffic coordination
  - Disaster Relief
  - Amber alert response
  - Factory monitoring
  - Surveillance

- **Retrospective (Delayed)**
  - City planning
  - After-the-fact security/investigation
  - Long-term data analysis (trends)

Query Types

- Query = DNN output + additional processing
  - In this class, we'll mainly focus on DNN output
- 3 main classes
  - Binary classification: is an object there or not
  - Counting: how many of an object type are there
  - Detection: bounding boxes for objects in the scene
- Many others
  - Segmentation
  - Additional processing can consider past results, etc.

Real -Time Challenges

- Server-side resource efficiency
  - ML models are expensive
  - Many cameras and many frames
  - Concurrent queries
- Network Bandwidth + Latency
  - Video is data intensive (worse with many cameras over same network)
  - Latency between camera and server → delayed responses
- Edge resource constraints
  - Solve latency issues by running on edge → exacerbates resource overheads!

Offline Challenges

- Main problem: **too much data**!
- Computation costs
  - Cannot run each query on all frames
- Storage + Network costs
  - Don’t know a priori which frames will be needed for future queries

Shared Challenges

- Query language
  - How can average users express rich queries?
- Privacy
  - Cameras in public settings are now commonplace
- ML vision models
  - Want high-accuracy models even with potentially low quality video

Course Logistics

Staff

- Instructor: Ravi Netravali
  - Assistant Professor (joined UCLA in January 2019)
  - Research interests: networks/distributed systems; performance and debugging of large-scale, distributed applications
  - Office hours: by appointment

Heads up

This is first offering of the course, and my first course overall so…

…expect a few hiccups

…don’t hesitate to provide suggestions!

Course Website

CS239 Home

Course Resources

Schedule

Review Submission

CS239: ML-driven Video Analytics Systems, Winter 2020

Instructor: ~~Ravi Netravali~~

Lectures: **Monday/Wednesday 12pm-1:50pm** in 5280 Boelter Hall

Office Hours

- Ravi: by appointment (ravi@cs.ucla.edu)

Course Overview

Video cameras are pervasive. As camera deployments expand, organizations increasingly rely on analyzing video feeds to guide numerous applications including traffic monitoring, surveillance, and amber alert response. Key to the success of such applications has been recent advances in computer vision, particularly neural network (NN)-based techniques for highly accurate object detection and recognition. Though effective at answering high-level queries about video content, these NN-based pipelines are resource intensive in terms of network and server-side compute overheads. This class will explore a wide range of systems and machine learning optimizations to improve the efficiency of modern video analytics pipelines, without violating latency and query accuracy expectations.

Grading

- 40% Participation in paper discussions
- 10% Paper summaries
- 20% Paper presentation
- 30% Final project (report and presentation)

Paper Summaries

This course will be **entirely** based on research papers. Prior to each class, students will be expected to read the listed research paper(s) and write up a brief summary for each. Paper summaries should be short and include the following components:

- Paper description (few sentences) including what problem the paper tackles, and how it does so
- Potential limitations of the solution (e.g., are any assumptions too general? are there scenarios where the proposed solution does not work?)
- Potential extensions (e.g., ways to support more scenarios, other domains/applications where the proposed solution could be useful)

http://web.cs.ucla.edu/~ravi/CS239_W20/

Who should take this course?

- Course is entirely research-focused (2-3 papers per week)
  - Reading each paper will take several hours
  - Understanding the paper (and related work) will take even more time
    - We are really trying to get into the paper details in this class
- No programming (other than potentially for research project)
- Course is mainly designed for PhD students
  - Masters students: welcome, but please note course focus
  - Undergrads: please discuss enrollment with me
- Prerequisites: knowledge of networking, OS, and distributed systems

Enrollment

- By PTE only

- Enrollment decisions will be made after today's lecture
  - If interested, see me after class to discuss

Course Goals

- Learn how to read network/systems research papers critically
  - Compare similar and seemingly different papers
- Articulate understanding and thoughts about paper
- Formulate and present research directions

Course Structure

- Before Class
  - Read papers
  - Submit paper critiques
- During Class
  - Paper presentations
  - **Lively** discussions
- Throughout the quarter
  - Research Direction

Paper Reading

- 1-2 papers per lecture (usually 1)

- "How to Read a Paper" by S. Keshav (https://blizzard.cs.uwaterloo.ca/keshav/home/Papers/data/07/paper-reading.pdf)
  - 1st pass: high-level (title, abstract, intro, section titles); categorize paper (by area/goals), is solution plausible, etc.
  - 2nd pass: more detail (graphs/illustrations); understand main contribution
  - 3rd pass: be able to re-implement paper solution from scratch and identify flaws

Paper Critiques

- Each paper review should include:
  - Paper summary (1 short paragraph): problem addressed, and how?
  - Potential limitations of the solution (e.g., cases where it won't work)
  - Potential extensions to make better or extend to other scenarios
  - Any questions about the paper or general topic
- **Looking for critique, not abstract only**
- You should submit a paper review for each paper (not per lecture)
- Graded on 1-5 scale (mostly on display of thought)

Paper Reviews

- Due by **10pm** the night **before** each lecture
  - Lets me identify questions that many people have
  - Important to give yourself time to think about the paper (helps discussion)

- Submit paper reviews using the form on the course website (https://web.cs.ucla.edu/~ravi/CS239_W20/review.html)

- You may skip 2 paper summaries without penalty

Paper Presentations

- Most likely individual presentations (subject to enrollment)

- “Conference style” presentations
  - Domain and relevant background for the paper
  - Problem statement and challenges
  - Solution
  - Results (along with setup details)
  - Potential limitations and improvements

- Presentations should be roughly 30 minutes

Presentation Sign-ups

- Spreadsheet sent out later today
  - Signups due by end of week (presentations start next week)
  - First come first serve

- Drop policy: please let me know ASAP via email

Paper Discussion

- Presenters lead the discussion (after talk is done)
  - **But everyone should participate**

- Presenters and audience should come prepared with:
  - Questions to discuss
  - Discussion of key takeaways from paper
  - Discussion points for limitations
  - Potential extensions (good time to get feedback on ideas!)

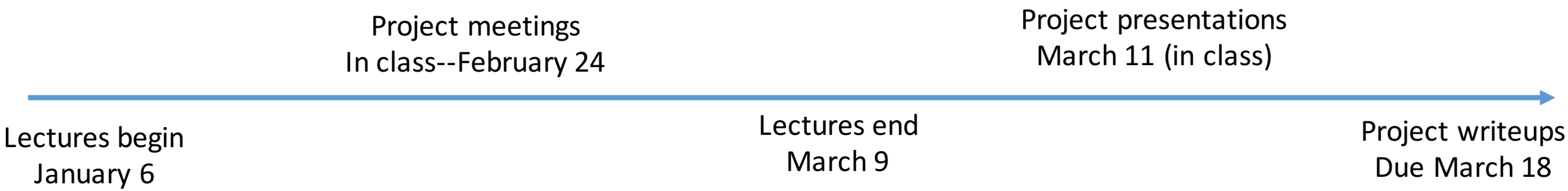
Research Project

- Topic: anything related to video analytics pipelines

- **Goal: come up with and motivate a *potential* research direction**
  - Motivational results will be helpful
  - No implementation of proposed idea required
  - Aim high!

- Done independently

Research Project Timeline/Deliverables



- **Project meetings:** explain high-level direction and motivation, related work, proposed solution
  - Okay to pivot!
- **Project presentations**: 7 minute in-class presentations
- **Project writeups** (3 pages): conference-style paper detailing motivation/problem, related work, challenges, high-level solution, and potential implementation details

Project Notes

- Fine to relate to ongoing research projects, but must be video analytics-related

- **Please start thinking early!**
  - I'm happy to discuss project ideas anytime

- **Example Areas**
  - Alternate ways to index video for retrospective queries
  - Approximate responses for live queries
  - System design for resource-constrained settings (e.g., limited storage, network, compute, etc.)

Grading

- **40% Participation in paper discussions**
- 10% Paper summaries
- 20% Paper presentation
- 30% Final project (presentation and writeup)

Other Notes

- Please drop early
  - Affects paper presentations
- Any issues (e.g., critique deadlines, project concerns, etc.) → please come see me early
- Not a lecture course!!
  - Please come prepared and participate so everyone can benefit

For Next Lecture

- Topic: Splitting video analytics across edge and cloud
- Presenter: Murali
- Paper:
  - **Glimpse** (SenSys 2015)
  - Paper Critique due tomorrow (Tuesday) by 10pm

Any Questions?