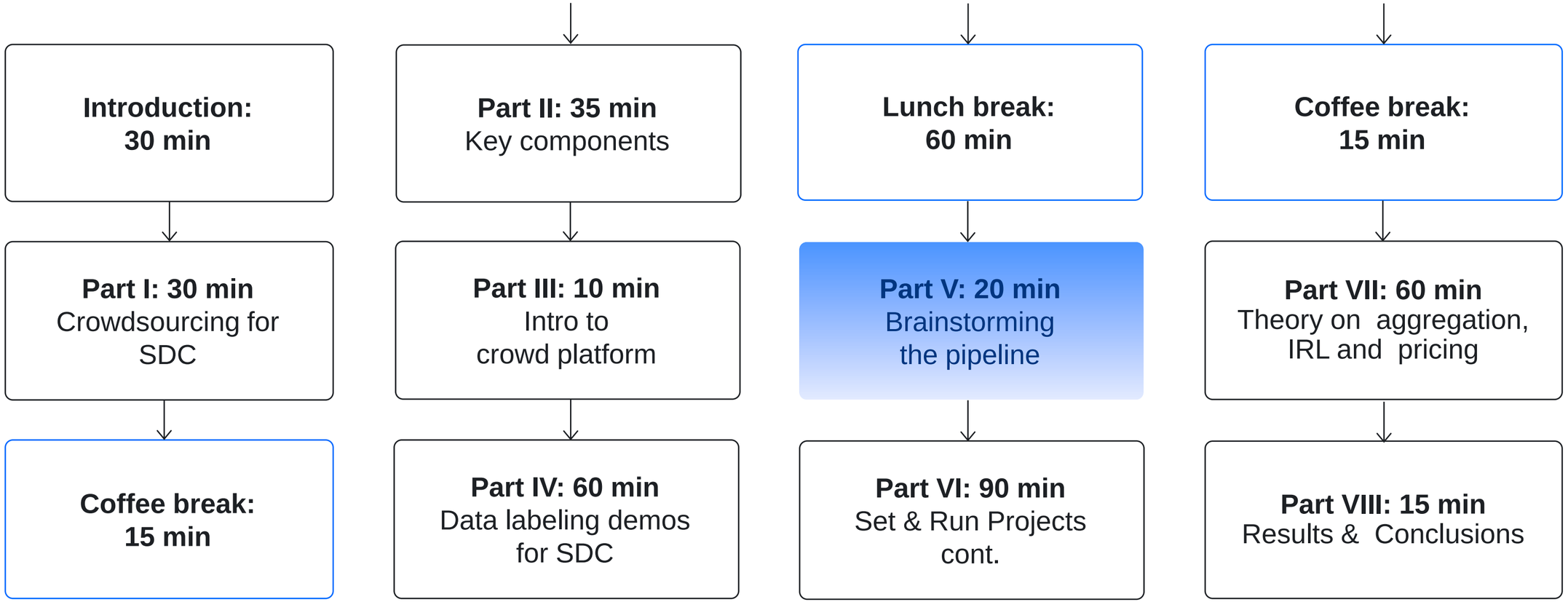
Part V

Brainstorming the pipeline

Daria Baidakova, Project Manager, Toloka

Tutorial schedule



Practice session

Our practice session will consist of two parts:

Part I (now)

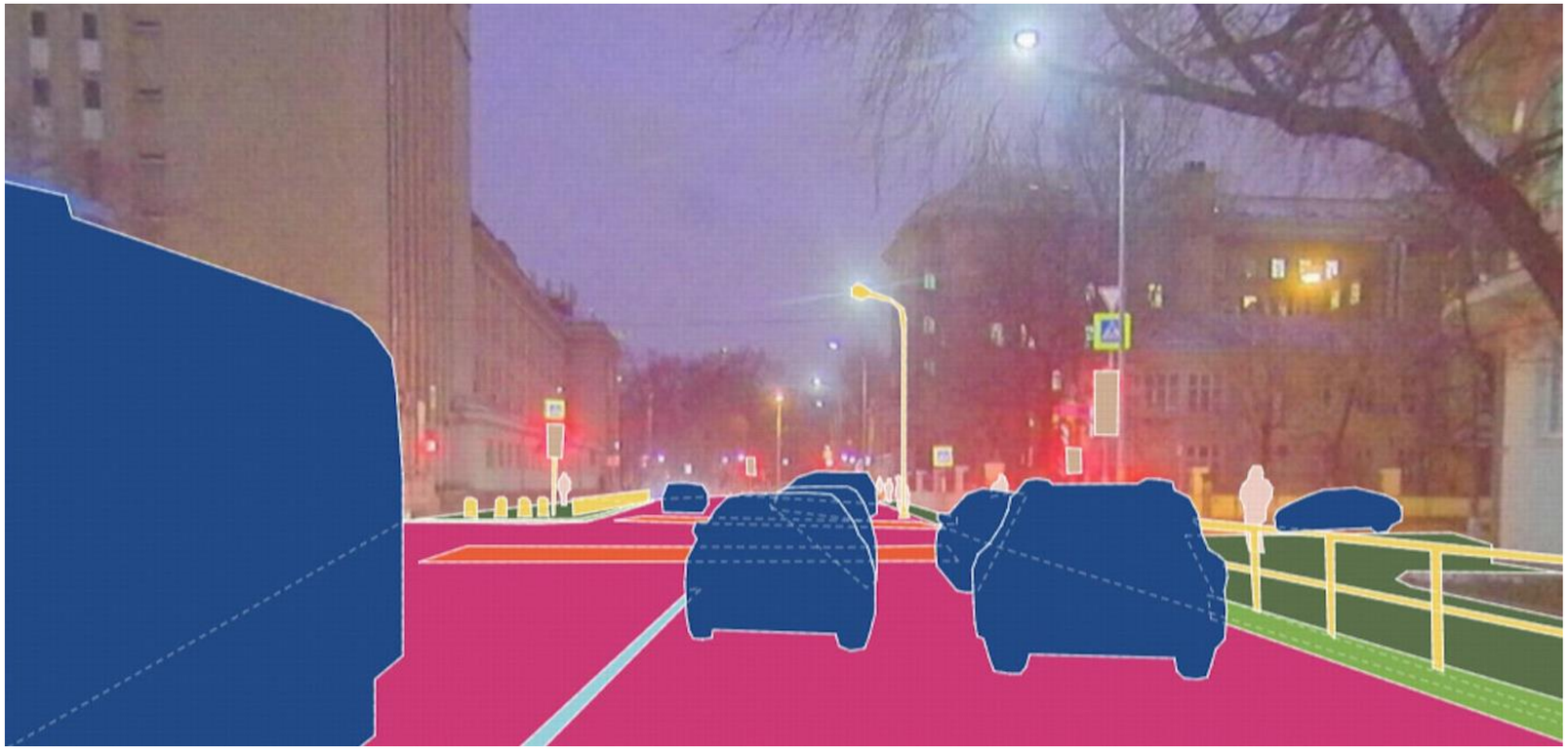
Think and discuss
how you would design
a crowdsourcing pipeline

Part II (in 20 min)

Run a best-practice
pipeline on the real
Toloka crowd

Practice session: scope

Imagine that you develop a machine learning pipeline
to help self-driving cars behave better on real roads



Practice session: scope

Imagine that you develop a machine learning pipeline
to help self-driving cars behave better on real roads

- You have already collected a large set of photos of roads
- You need to outline different objects in each of these photos
- These collected labels will further be used to train a CV system

This is your goal
for the practice
session of our tutorial

Dataset under study: real photos of city roads

Objects to be outlined in the photos

Each photo may contain objects of different types, for example:

- People
- Transport
- Road
- Curb
- Traffic lights
- Traffic signs
- Sidewalk
- Pedestrian crossing
- Other objects

During your practice:

Choose one type of object that you want to outline in the photos

For example: Traffic signs

Formal setup: get objects bounded by rectangles

- Each object of a selected type in each photo from the dataset needs to be outlined with a rectangle (bounding box)
- **Let us do it via crowdsourcing**

Example: I decided to outline all traffic signs, so our pipeline would be like..

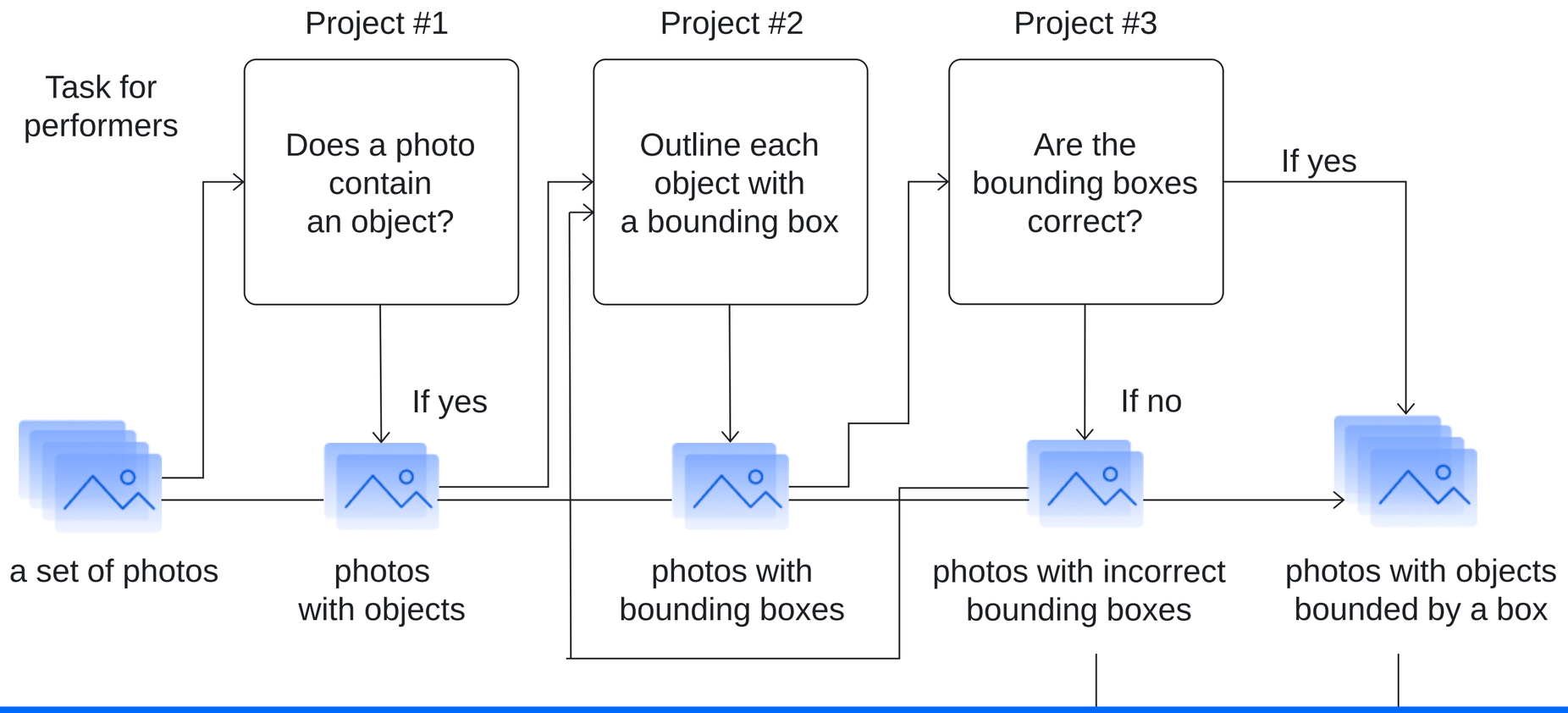
During your practice:

Think how you would design a crowd pipeline to collect these labels

Example: before & after

Suggested pipeline

We suggest the following pipeline



During the practical session we will help you implement and run this pipeline

Project #1: Filter out photos without objects

Task

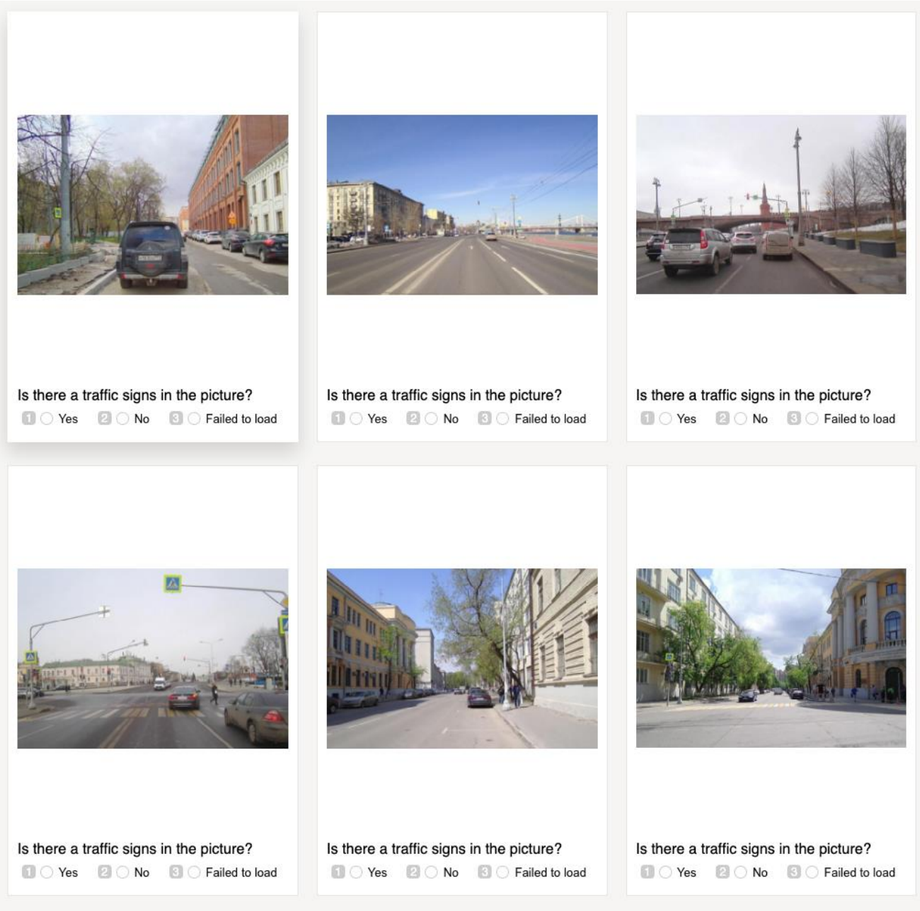
- Does a photo contain objects of desired type?

Key setting

- Type: classification
- Quality control: golden set
- Overlap: 3 answers per photo
- Pay: $0.01 per a suite of 10 photo

Why?

- Save money: no need to process further photos without desired objects



Project #2: Outlining objects with rectangles

Task

- Outline each object of desired type with a bounding box

Key setting

- Type: region selection in an image
- Quality control: post verification
- Overlap: 1 (with correct boxes)
- Pay: $0.01 per 1 photo

Why?

- Hard to use golden set and consensus
- Results will be verified in Project #3

Project #3: Accept correct bounding boxes

Task

- Are the objects of desired type outlined with the bounding boxes correctly?

Key setting

- Type: classification
- Quality control: consensus
- Overlap: 3 answers per photo
- Pay: $0.01 per a suite of 10 photo

Why?

- Need to verify the results obtained from Project #2