

Yandex

Efficient Data Annotation for Self-Driving Cars via Crowdsourcing on a Large-Scale

Alexey Drutsa, Denis Rogachevsky, Olga Megorskaya, Daria Baidakova, Ivan Semchuk

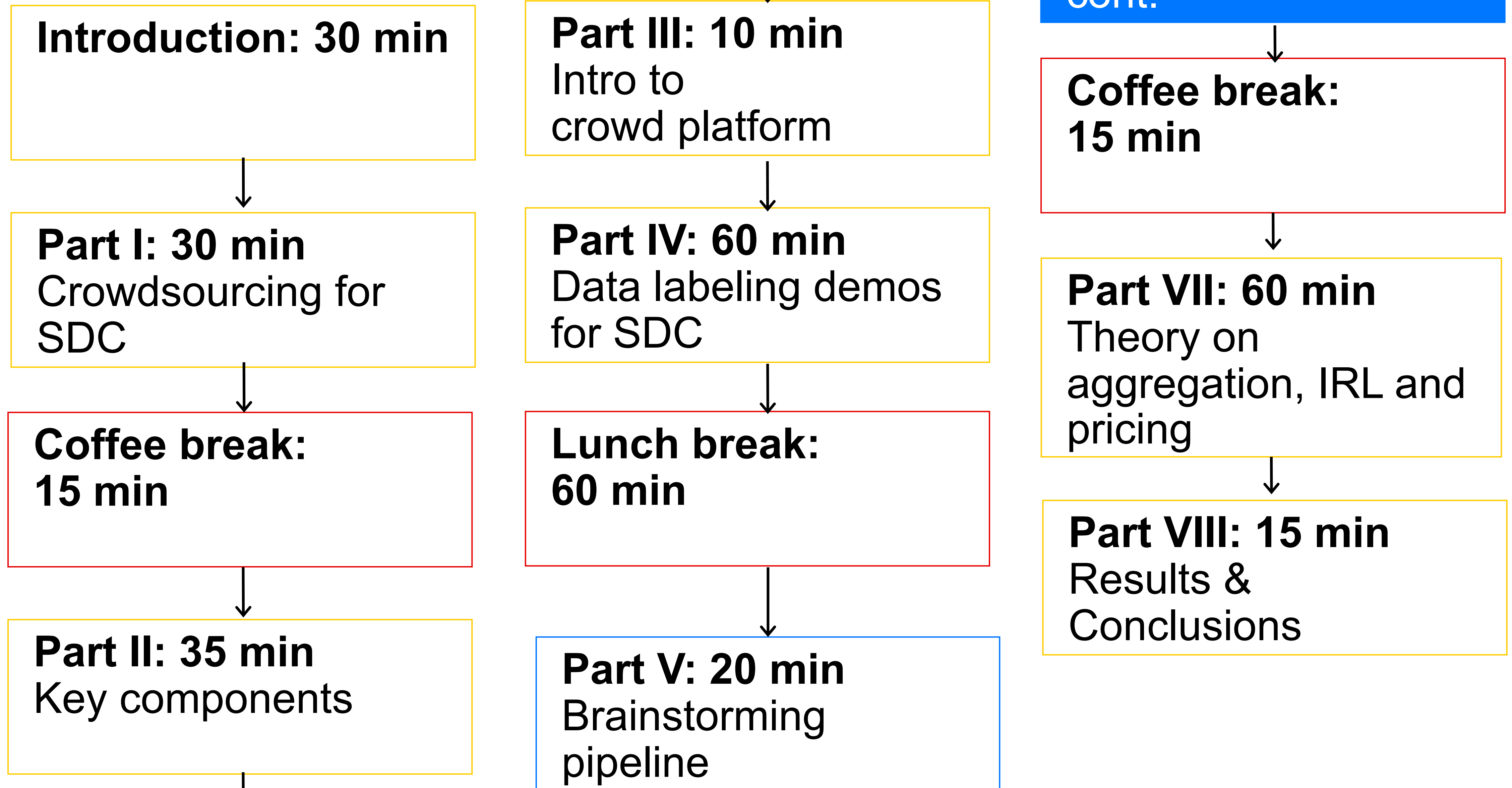
Part III:

Brainstorming the pipeline

Daria Baidakova,
Project Manager

Yandex.Toloka is a service of Swiss company Yandex Services AG

Tutorial schedule



Practice session

Our practice session will consist of two parts:

Part I (now)

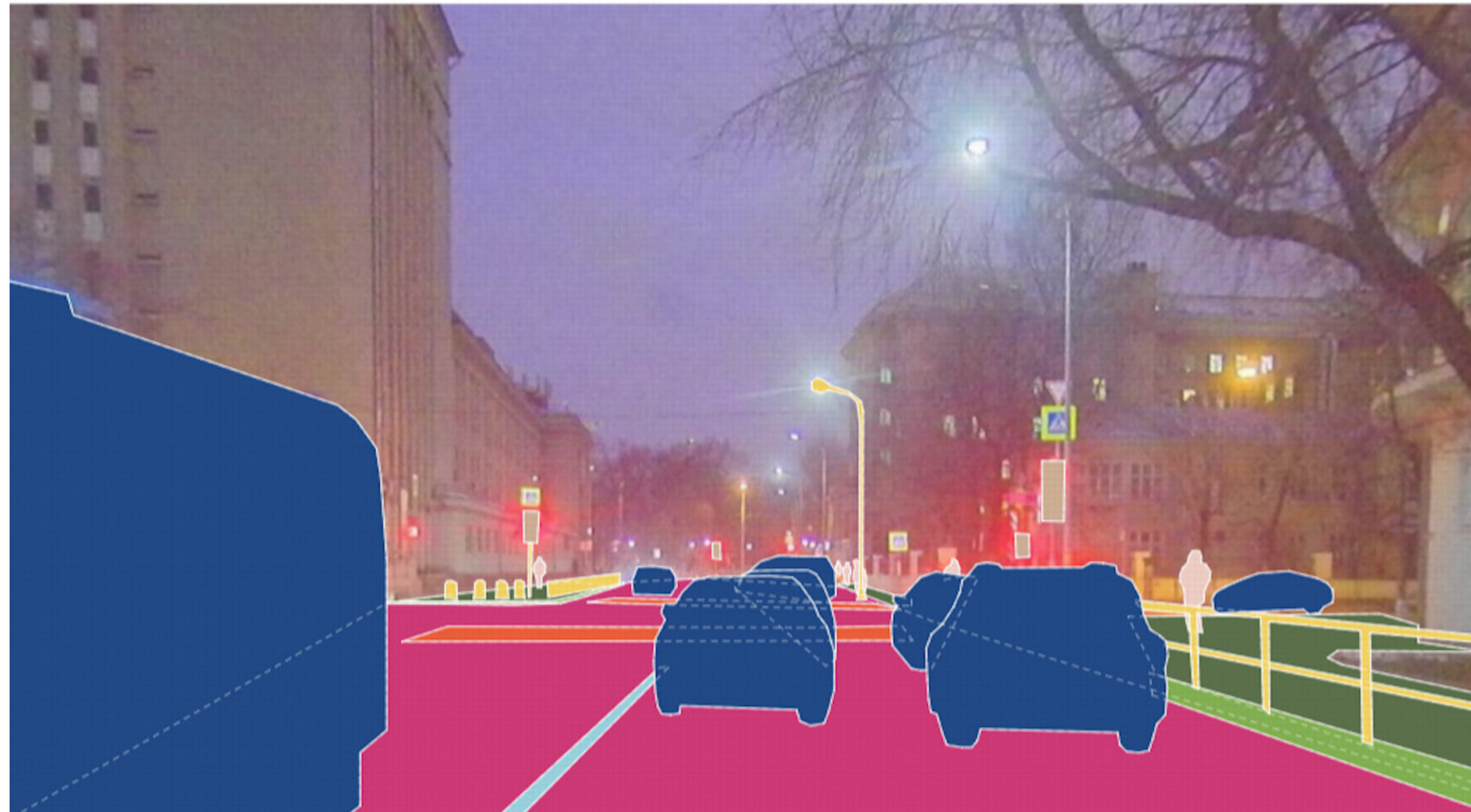
Think and discuss
how you would design a
crowdsourcing pipeline

Part II (in 20 min)

Run the best-practice pipeline
on a real crowd at Toloka

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

It is your goal during 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 objects
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

During your practice:

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

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

Example: before & after



Practice: you have 10 mins now

1. Assume we need to highlight every traffic sign on every photo with a bounding box
2. Design a crowdsourcing pipeline for it



a set of photos

How to get desired bounding boxes from a crowd?



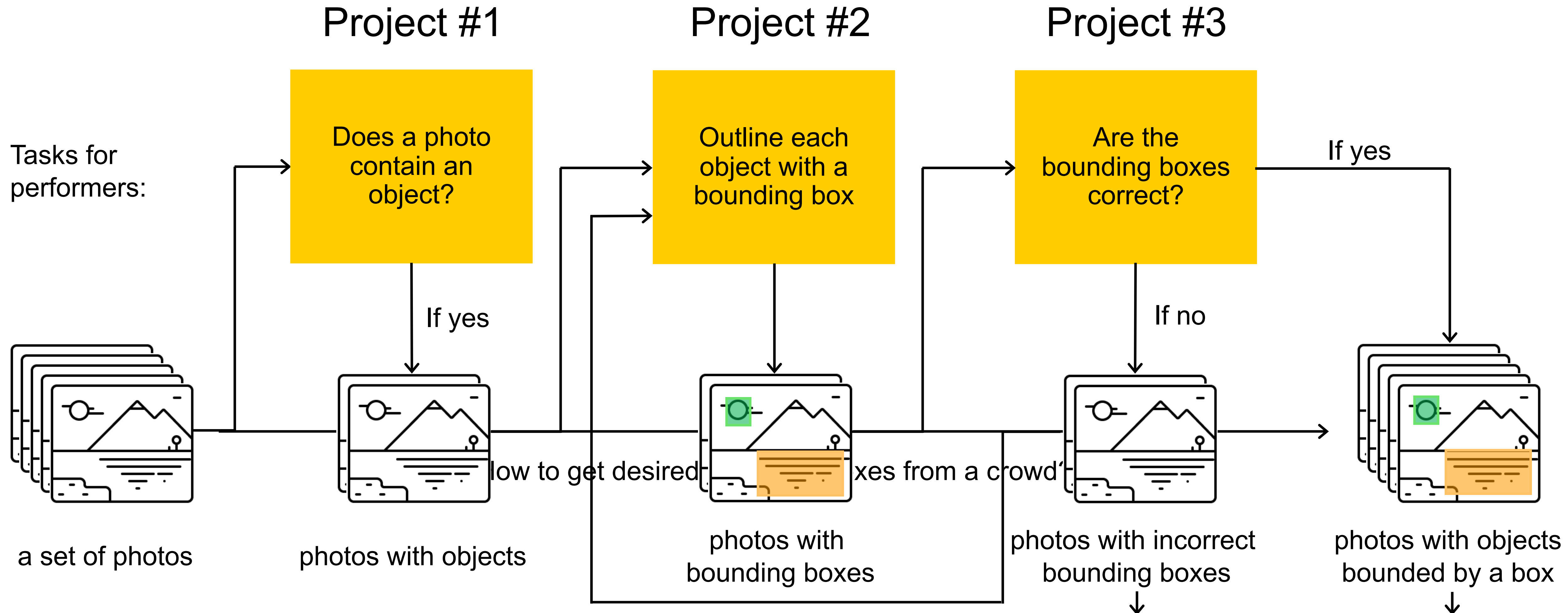
photos with objects
bounded by a box

Hint: discuss in groups how to

- › decompose the problem
- › make tasks simple
- › control quality

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

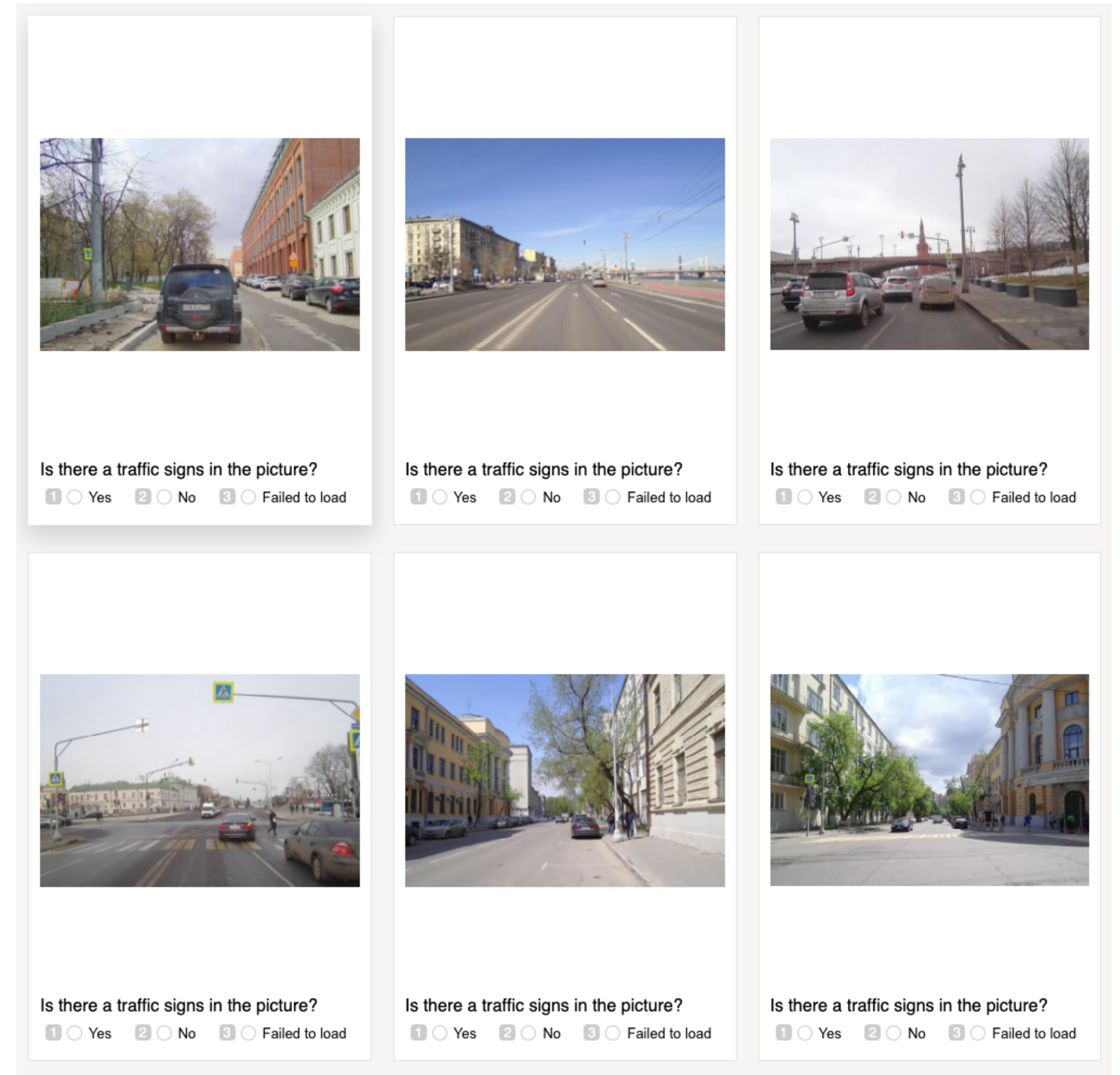
- › Does a photo contain objects of desired type?

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

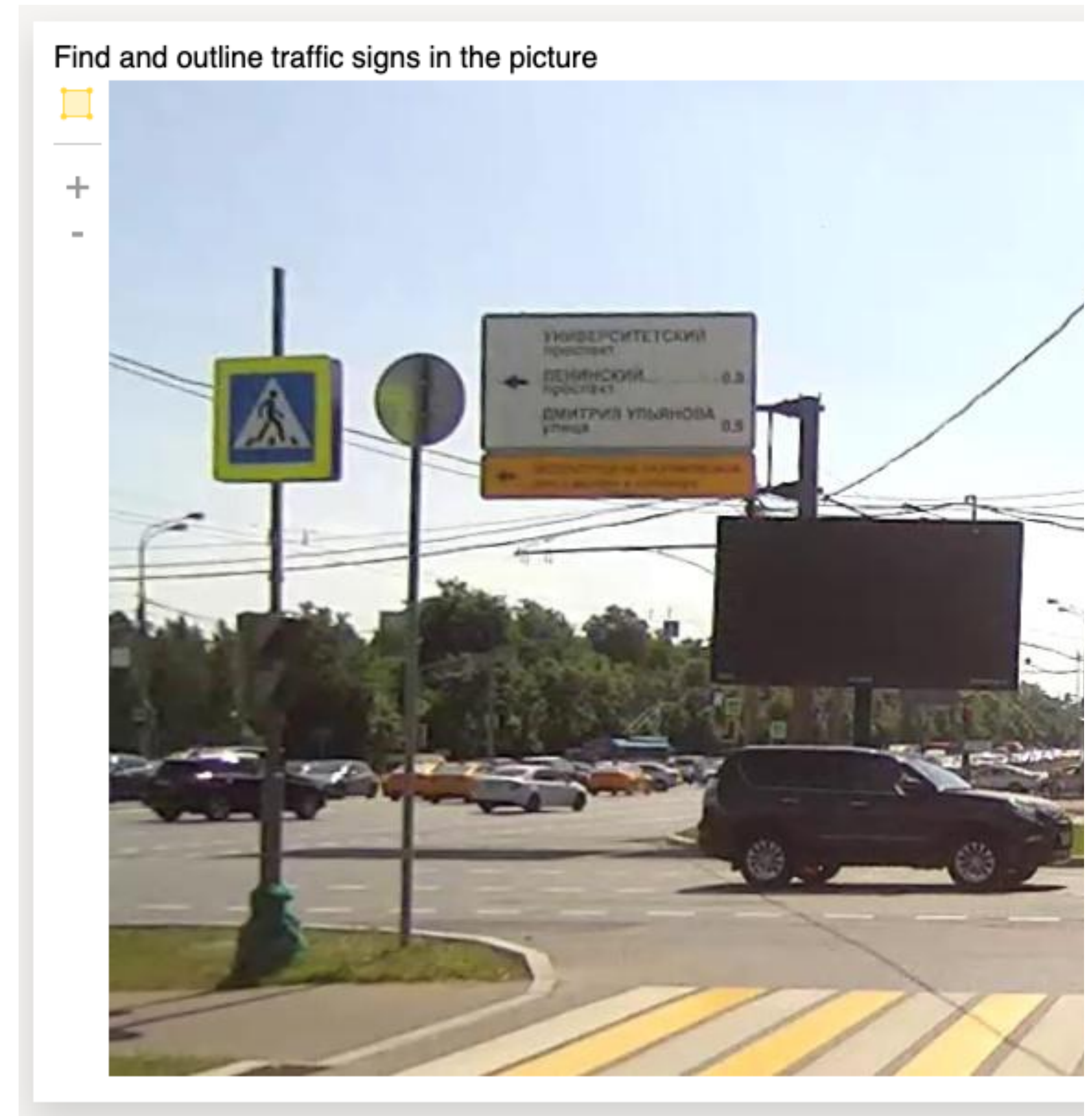
- › Outline each object of desired type with a bounding box

Key setting

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

Peculiar properties

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



Project #3: Accept correct bounding boxes

Task

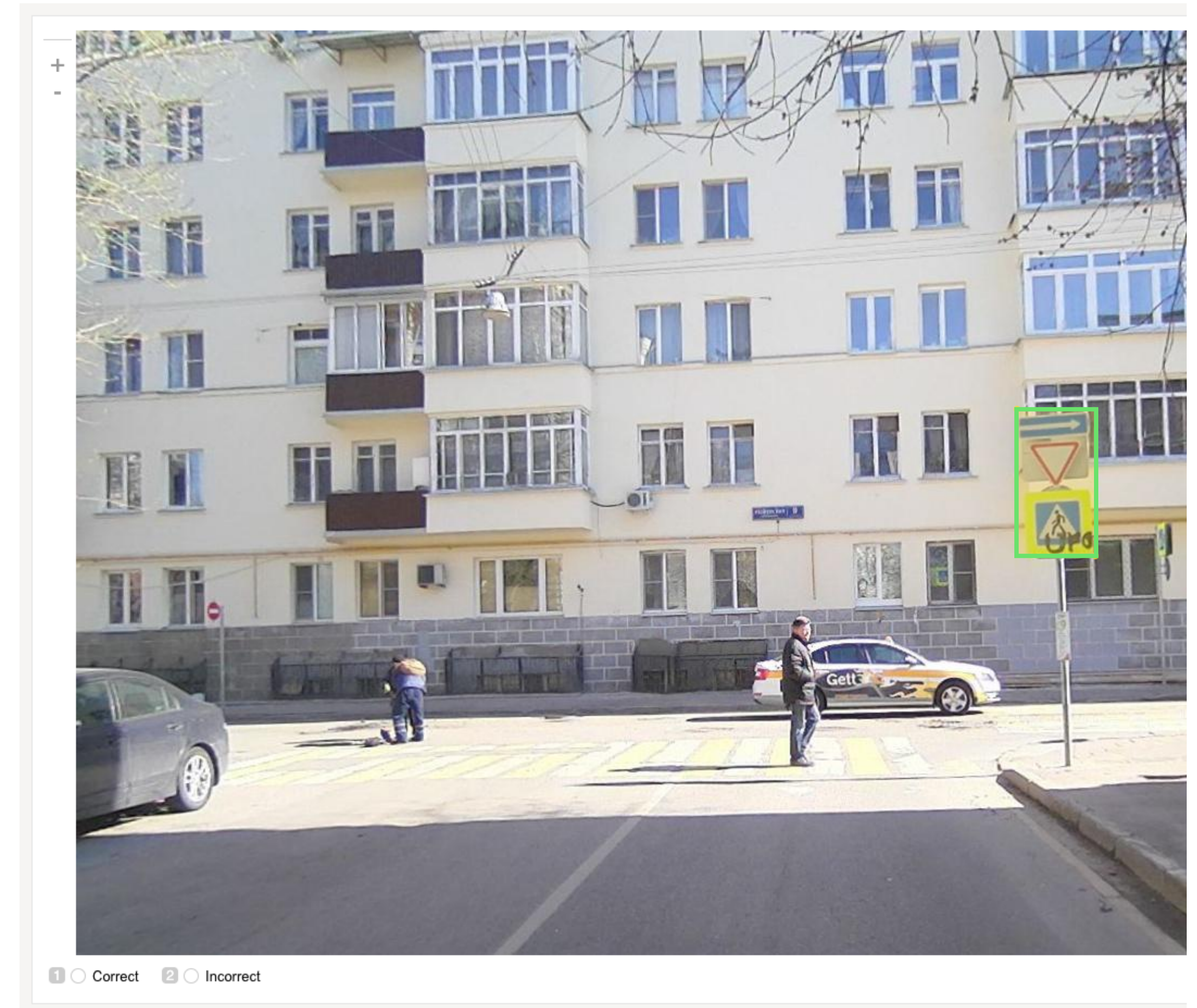
- › Are the objects of desired type outlined with the bounding boxes correctly?

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



Yandex

Thank you!
Questions?

Daria Baidakova

Project manager



dbaidakova@yandex-team.ru



<https://research.yandex.com/tutorials/crowd/cvpr-2020>