

**Y**andex

# **Crowdsourcing Practice for Efficient Data Labeling: Aggregation, Incremental Relabeling, and Pricing**

Alexey Drutsa, Valentina Fedorova, Dmitry Ustalov, Olga Megorskaya, Evfrosiniya Zerminova, Daria Baidakova

# Part VIII

## **Discussion of the projects' results**

## **Conclusions**

Alexey Drutsa,  
Head of Efficiency and Growth Division, Yandex

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

# Tutorial schedule

**Introduction: 30 min**

**Part I: 30 min**  
Crowdsourcing for  
SDC

**Coffee break:**  
**15 min**

**Part II: 35 min**  
Key components

**Part III: 10 min**  
Intro to  
crowd platform

**Part IV: 60 min**  
Data labeling demos  
for SDC

**Lunch break:**  
**60 min**

**Part V: 20 min**  
Brainstorming  
pipeline

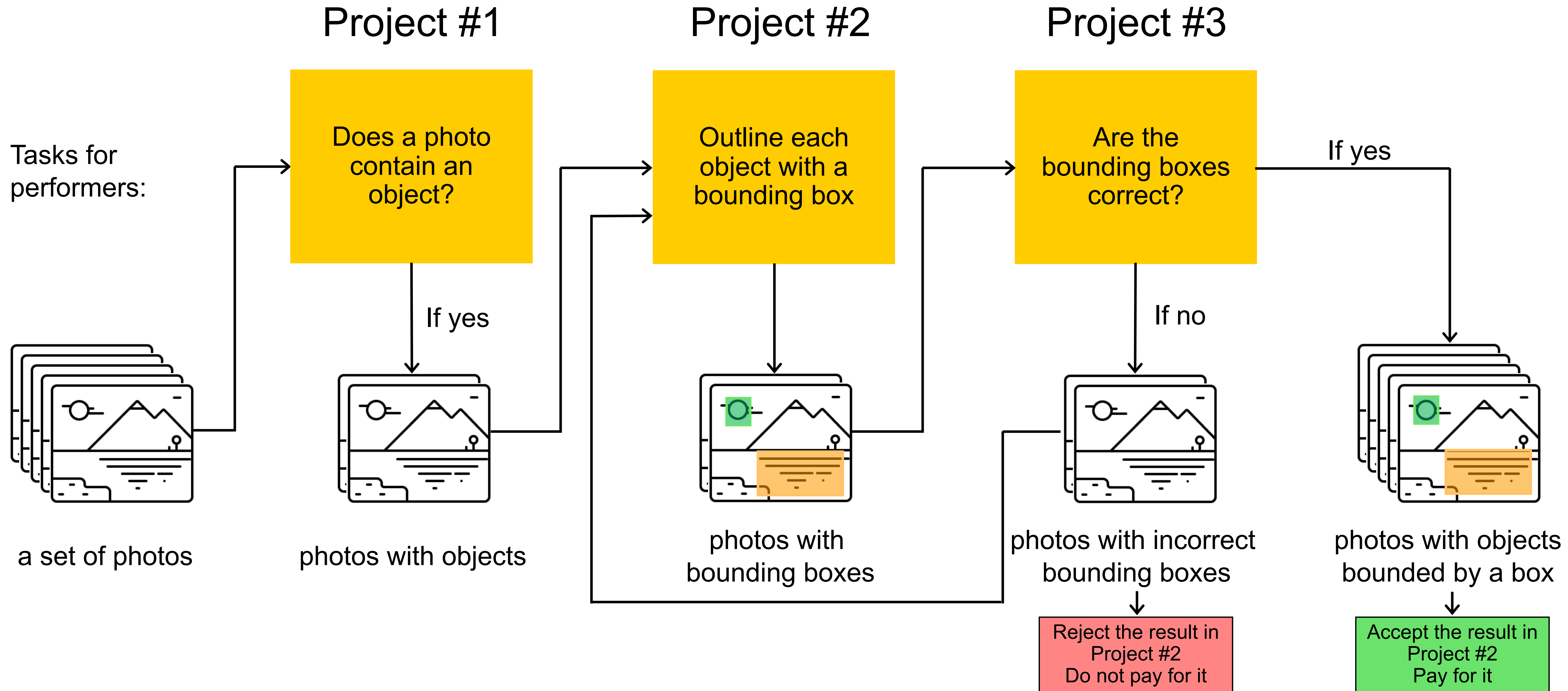
**Part VI: 90 min**  
Set & Run Projects  
cont.

**Coffee break:**  
**15 min**

**Part VII: 60 min**  
Theory on  
aggregation, IRL and  
pricing

**Part VIII: 15 min**  
Results &  
Conclusions

# Reminder: we implemented the pipeline



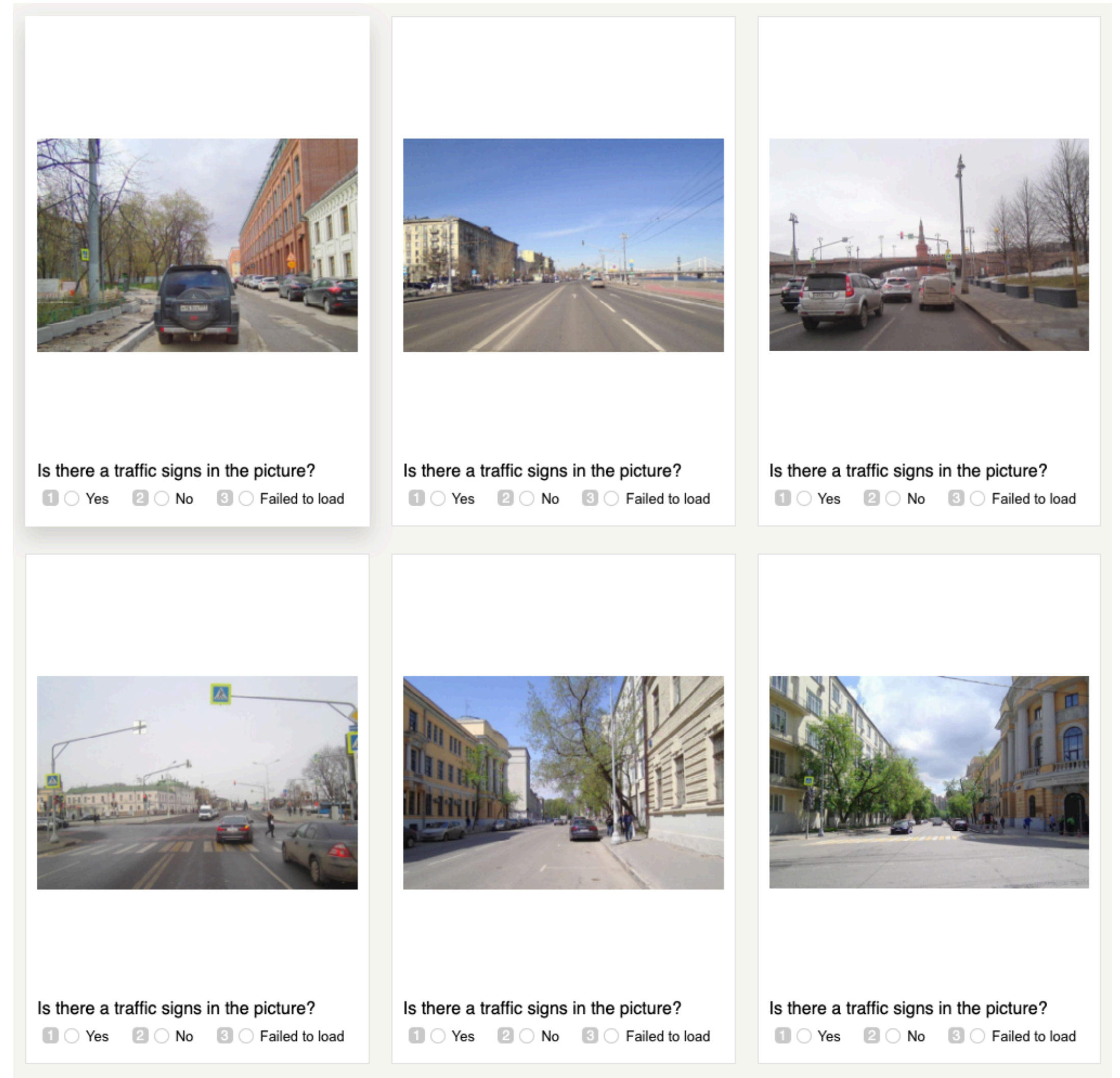
# Project #1: Filter out photos without objects

## Task

- › Does a photo contain objects of desired type?

## Our results:

- › 100 photos evaluated
- › within 4 min on real performers
- › cost: \$0.3 + Toloka fee





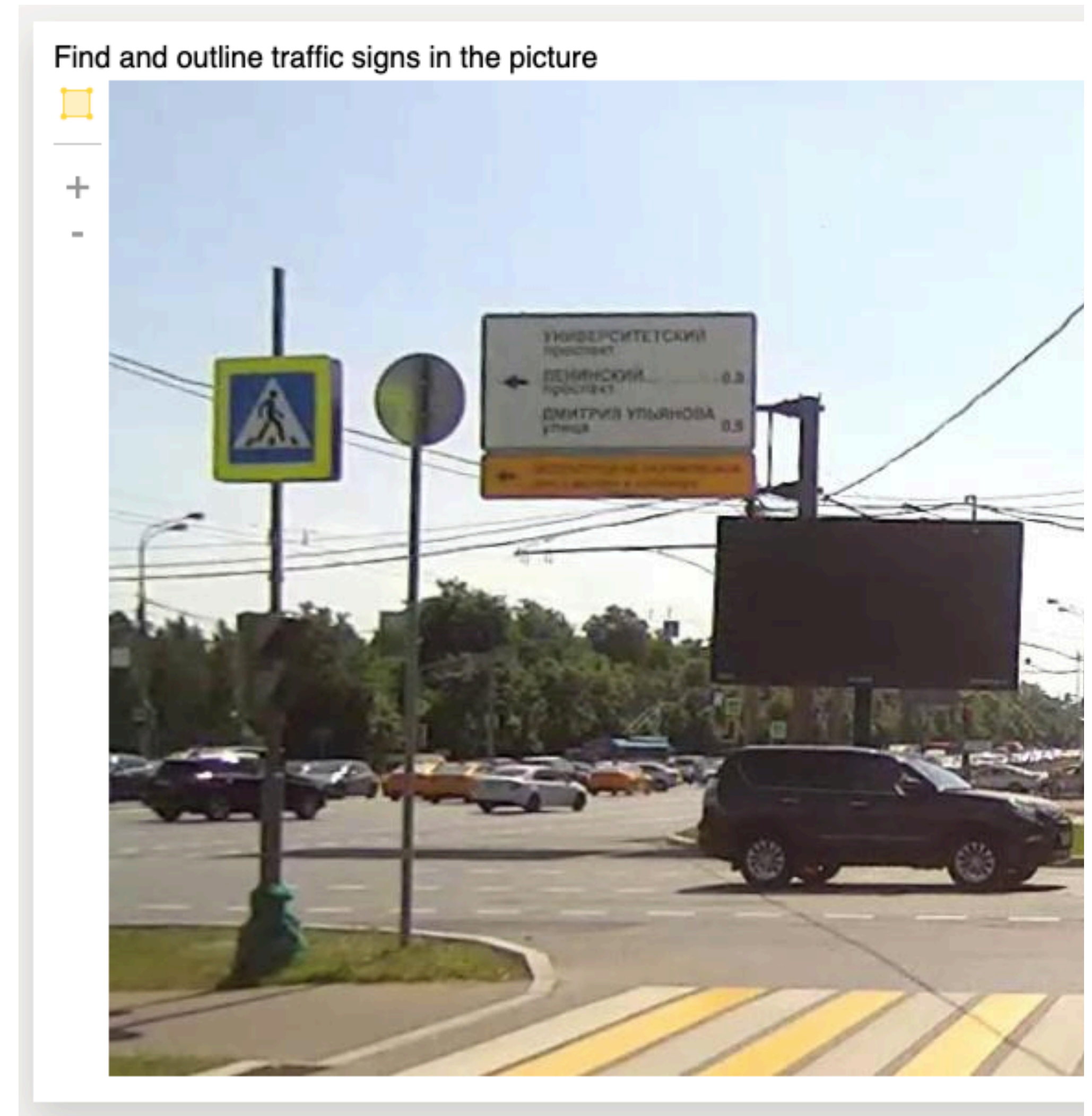
# Project #2: Outlining objects with rectangles

## Task

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

## Our results:

- › 67 photos processed
- › within 5.5 min on real performers
- › cost: \$0.67 + Toloka fee





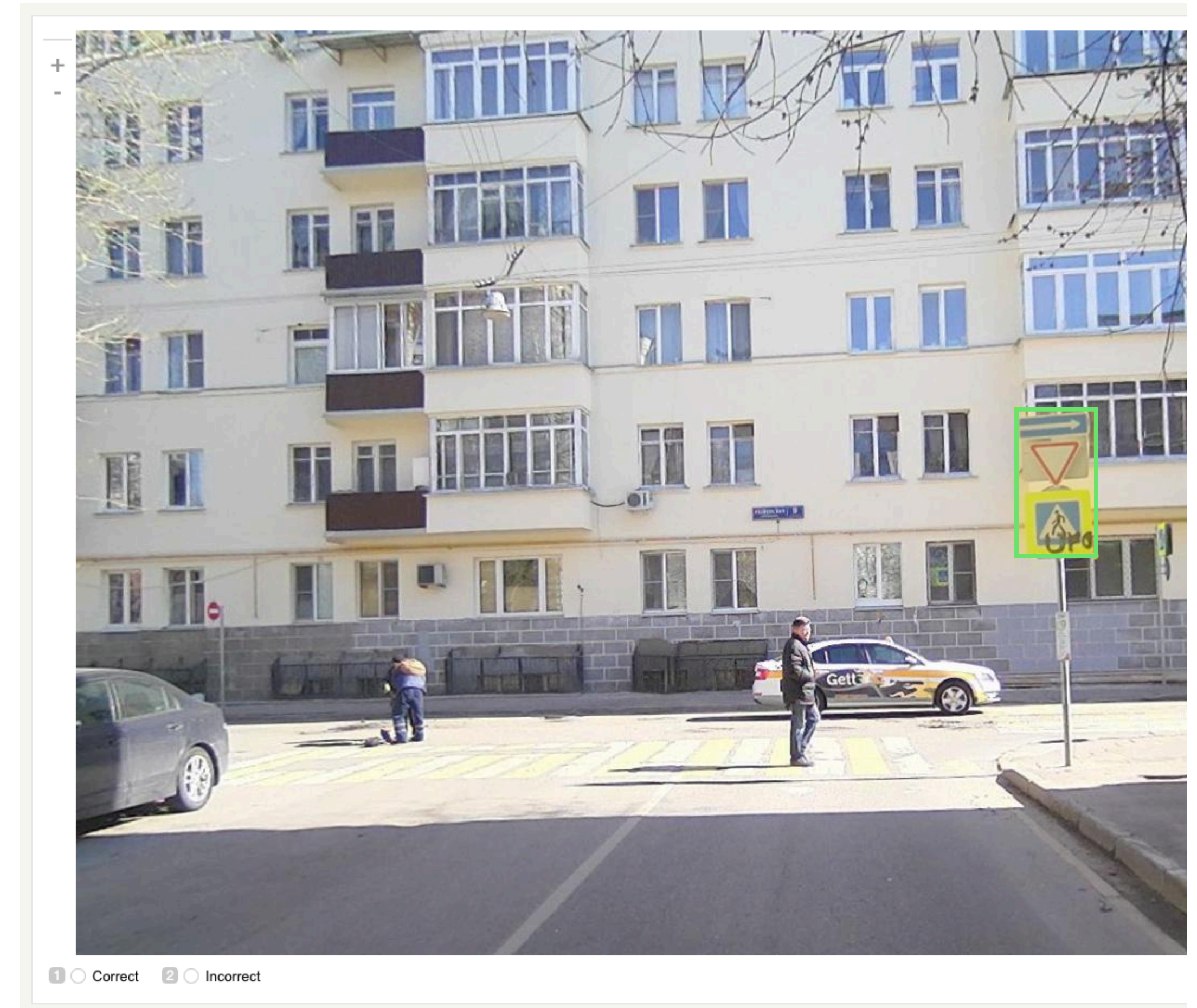
# Project #3: Accept correct bounding boxes

## Task

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

## Our results:

- › 90 photos evaluated
- › within 5 min on real performers
- › cost: \$0.36 + Toloka fee





# Statistics over the whole pipeline

**100 photos processed to highlight desired objects**

**within 14.5 min on real performers**

**total cost: \$1.33 + Toloka fee**

**Quality of the final result (via manual assessment):**

- › Recall: 90% (measured on results of Project #1)
- › Precision: 86% (measured on results of Project #2)

# Afterparty: upgrade your pipeline

## **To obtain more comprehensive data**

- › Use Polygons instead of Bounding boxes
- › Highlight more object types

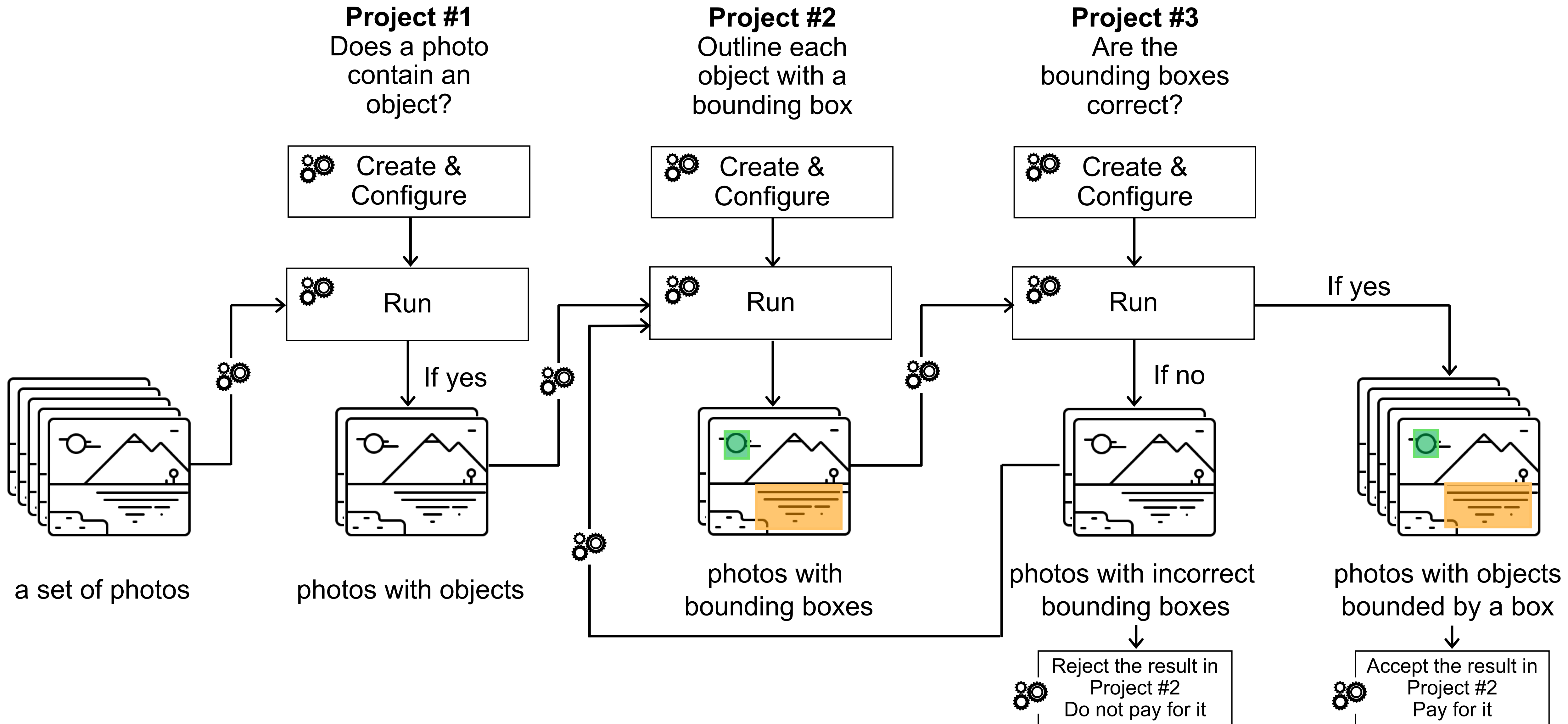
## **To reduce costs**

- › Use incremental relabeling aka Dynamic overlap

## **To improve quality**

- › Use dynamic pricing
- › Add more Golden Sets and hints
- › Experiment with aggregation methods
- › Add training for workers

# API of Yandex.Toloka





# API of Yandex.Toloka

**| Allows you to automate all steps of our pipeline**

Discover at: <https://yandex.com/dev/toloka/>

# Crowdsource all types of data

Search Relevance

Moderation

Generation of content

Computer vision

Speech Technologies

# References: Aggregation

- [1] Dawid, A. P. and Skene, A. M, Maximum likelihood estimation of observer error-rates using the EM algorithm, Applied statistics 1979
- [2] Whitehill, J., Wu, T., Bergsma, J., Movellan, J. R, Ruvolo, P. L, Whose vote should count more: Optimal integration of labels from labelers of unknown expertise}, NIPS 2009
- [3] Zhou, D., Liu, Q., Platt, J. C, Meek, C., Shah, N. B, Regularized minimax conditional entropy for crowdsourcing, arXiv preprint 2015
- [4] Raykar, V. C, Yu, S., Zhao, L. H, Valadez, G. H., Florin, C., Bogoni, L., Moy, L., Learning from crowds, JMLR 2010
- [5] Snow, R., O'Connor, B., Jurafsky, D., Ng, A. Y, Cheap and fast---but is it good?: evaluating non-expert annotations for natural language tasks, EMNLP 2008
- [6] Ruvolo, P., Whitehill, J., Movellan, J. R, Exploiting Commonality and Interaction Effects in Crowdsourcing Tasks Using Latent Factor Models, NIPS '13 Workshop on Crowdsourcing: Theory, Algorithms and Applications
- [7] Faridani, S. and Buscher, G., LabelBoost: An Ensemble Model for Ground Truth Inference Using Boosted Trees, HCOMP 2013
- [8] Welinder, P., Branson, S., Perona, P., Belongie, S. J , The multidimensional wisdom of crowds, NIPS 2010
- [9] Jin, Y., Carman, M., Kim, D., Xie, L., Leveraging Side Information to Improve Label Quality Control in Crowd-Sourcing, HCOMP 2017
- [10] Imamura, H., Sato, I., Sugiyama, M., Analysis of Minimax Error Rate for Crowdsourcing and Its Application to Worker Clustering Model, arXiv preprint 2018



# References: Aggregation

- [11] Sheshadri, A. and Lease, M., Square: A benchmark for research on computing crowd consensus, HCOMP 2013
- [12] Kim, H. and Ghahramani, Z., Bayesian classifier combination, AISTATS 2012
- [13] Venanzi, M., Guiver, J., Kazai, G., Kohli, P., Shokouhi, M., Community-based bayesian aggregation models for crowdsourcing, WWW2014
- [14] Vuurens, J., de Vries, A. P, Eickhoff, C., How much spam can you take? an analysis of crowdsourcing results to increase accuracy, SIGIR Workshop CIR 2011
- [15] Chen, X. and Bennett, P. N and Collins-Thompson, K. and Horvitz, E., Pairwise ranking aggregation in a crowdsourced setting, WSDM 2013
- [16] Liu, C. and Wang, Y., Truelabel+ confusions: A spectrum of probabilistic models in analyzing multiple ratings, ICML 2012

# References: Incremental relabeling & Pricing

- [17] Ipeirotis, P. G and Provost, F. and Sheng, V. S and Wang, J., Repeated labeling using multiple noisy labelers, KDD 2014
- [18] Abraham, I., Alonso, O., Kandylas, V., Patel, R., Shelford, S., Slivkins, A., How many workers to ask?: Adaptive exploration for collecting high quality labels, SIGIR 2016
- [19] Ertekin, S., Hirsh, H., Rudin, C., Learning to predict the wisdom of crowds, arXiv preprint 2012
- [20] Lin, C. H, Mausam, M., Weld, D. S, To Re(label), or Not To Re(label), HCOMP 2014
- [21] Zhao, L., Sukthankar, G., Sukthankar, R., Incremental relabeling for active learning with noisy crowdsourced annotations, PASSAT/SocialCom 2011
- [22] Wang, J., Ipeirotis, P. G, Provost, F., Quality-based pricing for crowdsourced workers, working paper, 2013
- [23] Cheng, J., Teevan, J., Bernstein, M. S, Measuring crowdsourcing effort with error-time curves, CHI 2015
- [24] Ho, C., Slivkins, A., Suri, S., Vaughan, J. W., Incentivizing high quality crowdwork, WWW 2015
- [25] Difallah, D. E., Catasta, M., Demartini, G., Cudr`e-Mauroux, P., Scaling-up the crowd: Micro-task pricing schemes for worker retention and latency improvement, HCOMP 2014
- [26] Yin, M., Chen, Y., Sun, Y., The effects of performance-contingent financial incentives in online labor markets, AI 2013
- [27] Shah, N., Zhou, D., Peres, Y., Approval voting and incentives in crowdsourcing, ICML 2015
- [26] Shah, N. and Zhou, D., No oops, you won't do it again: Mechanisms for self-correction in crowdsourcing, ICML 2016

# References: Tutorials

- [27] Crowdsourcing: Beyond Label Generation, Vaughan, J. W. KDD 2017
- [28] Crowd-Powered Data Mining, Li, G., Wang, J., Fan, J., Zheng, Y., Chai, C., KDD 2018
- [29] Social Spam Campaigns Social Spam, Campaigns, Misinformation and Crowdturfing, Lee, K., Caverlee, J., Pu, C., WWW2014
- [30] From Complex Object Exploration to Complex Crowdsourcing, Amer-Yahia, S., Roy, S.B., WWW 2015
- [31] Crowdsourced Data Management: Overview and Challenges, Li, G., Zheng, Y., Fan, J., Wang, J., Cheng, R., SIGMOD 2017
- [32] Spatial Crowdsourcing: Challenges, Techniques, and Applications, Tong, Y., Chen, L., Shahab, C., VLDB 2016
- [33] Truth Discovery and Crowdsourcing Aggregation: A Unified Perspective, Gao, J., Li, Q., Zhao, B., Fan, W., Han, J., VLDB 2015
- [34] Data-Driven Crowdsourcing: Management, Mining, and Applications, Chen, L., Lee, D., Milo, T., ICDE 15
- [35] Practice of Efficient Data Collection via Crowdsourcing at Large-Scale, Drutsa A., Fedorova V., Megorskaya O., Zerminova E., KDD 2019
- [36] Practice of Efficient Data Collection via Crowdsourcing: Aggregation, Incremental Relabelling, and Pricing, Drutsa A., Fedorova V., Ustalov D., Megorskaya O., Zerminova E., Baidakova D., WSDM 2020





**Thank you!**  
**Questions?**

**Alexey Drutsa**

Head of Efficiency and Growth Division



adrutsa@yandex-team.ru



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