

Yandex Self-Driving Group



Kirill Danilyuk, Team Lead @ 2D Computer Vision Team



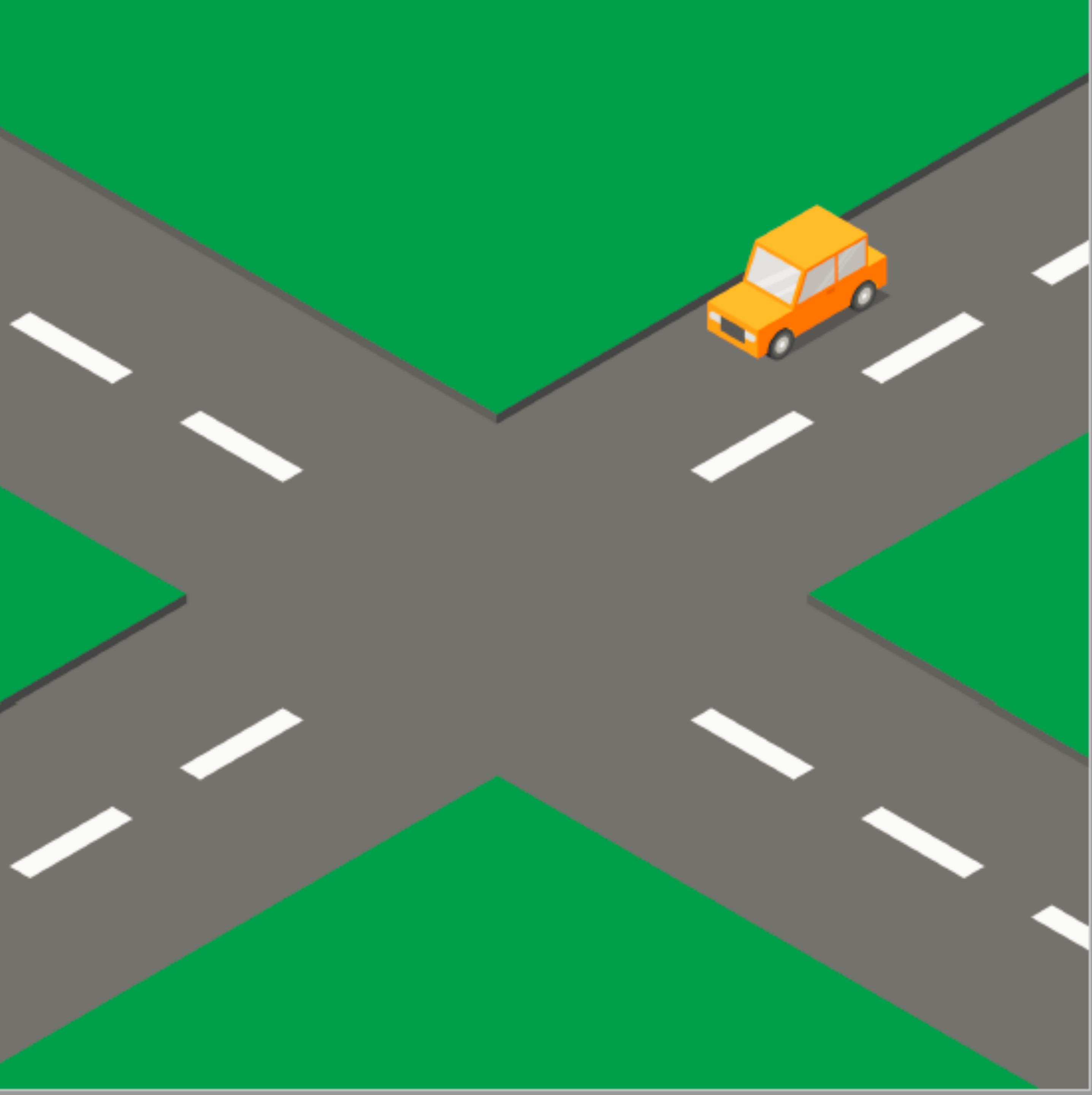
Visual Search at Yandex SDG

Kirill Danilyuk, Team Lead @ 2D Computer Vision Team
Yandex Self-Driving Group

PIPELINE 101

Self-Driving 101

Localization



Self-Driving 101

Localization

Perception



Self-Driving 101

Localization

Perception



Self-Driving 101

Localization

Perception



Self-Driving 101

Localization

Perception

Prediction



Self-Driving 101

Localization
Perception
Prediction
Planning



Self-Driving 101

Localization

Perception

Prediction

Planning



Self-Driving 101

Localization

Perception

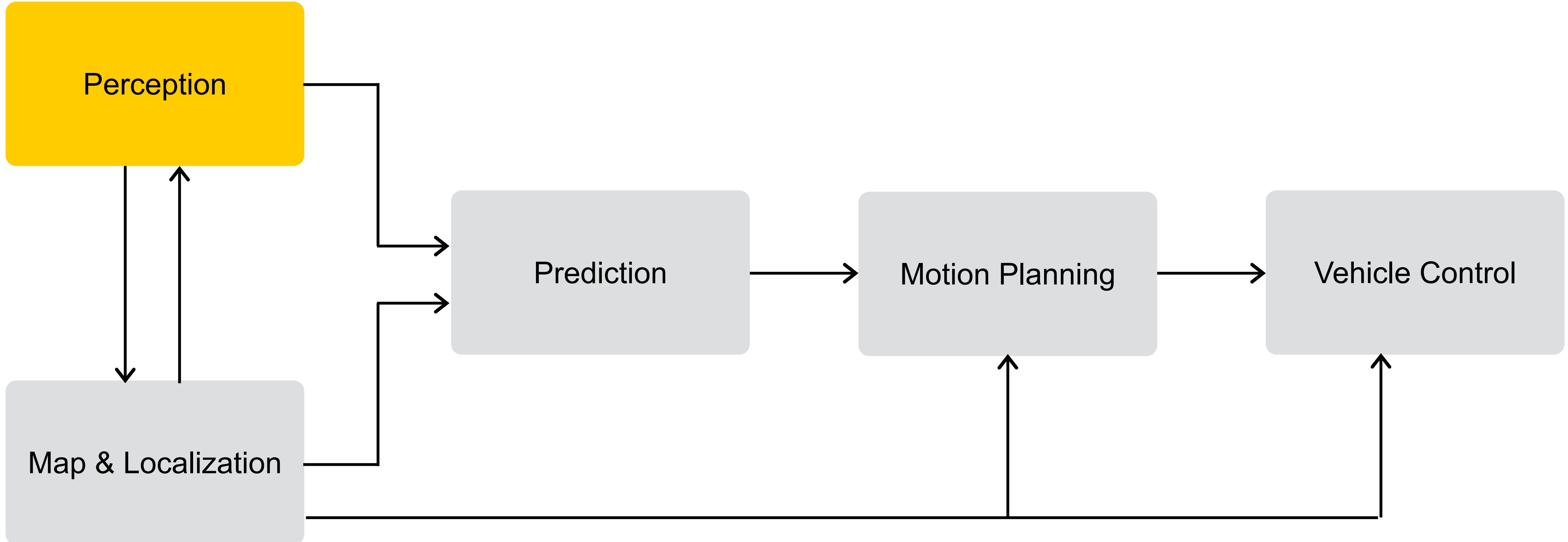
Prediction

Planning

Control



Self-Driving Pipeline

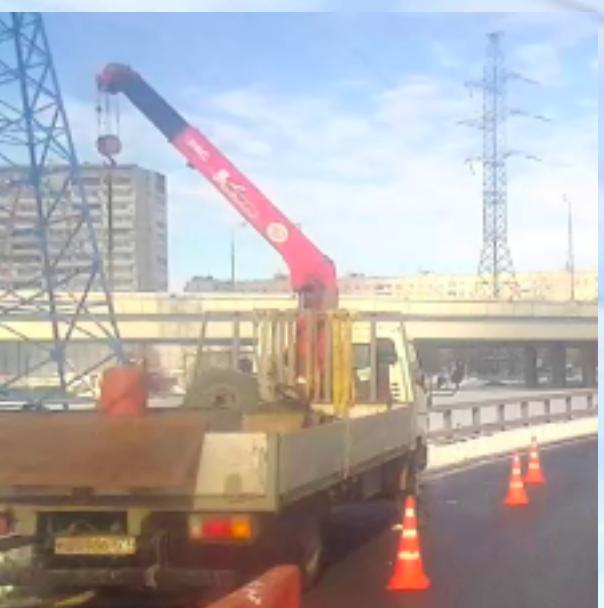


VISUAL SEARCH: MOTIVATION

Real World Is Complex



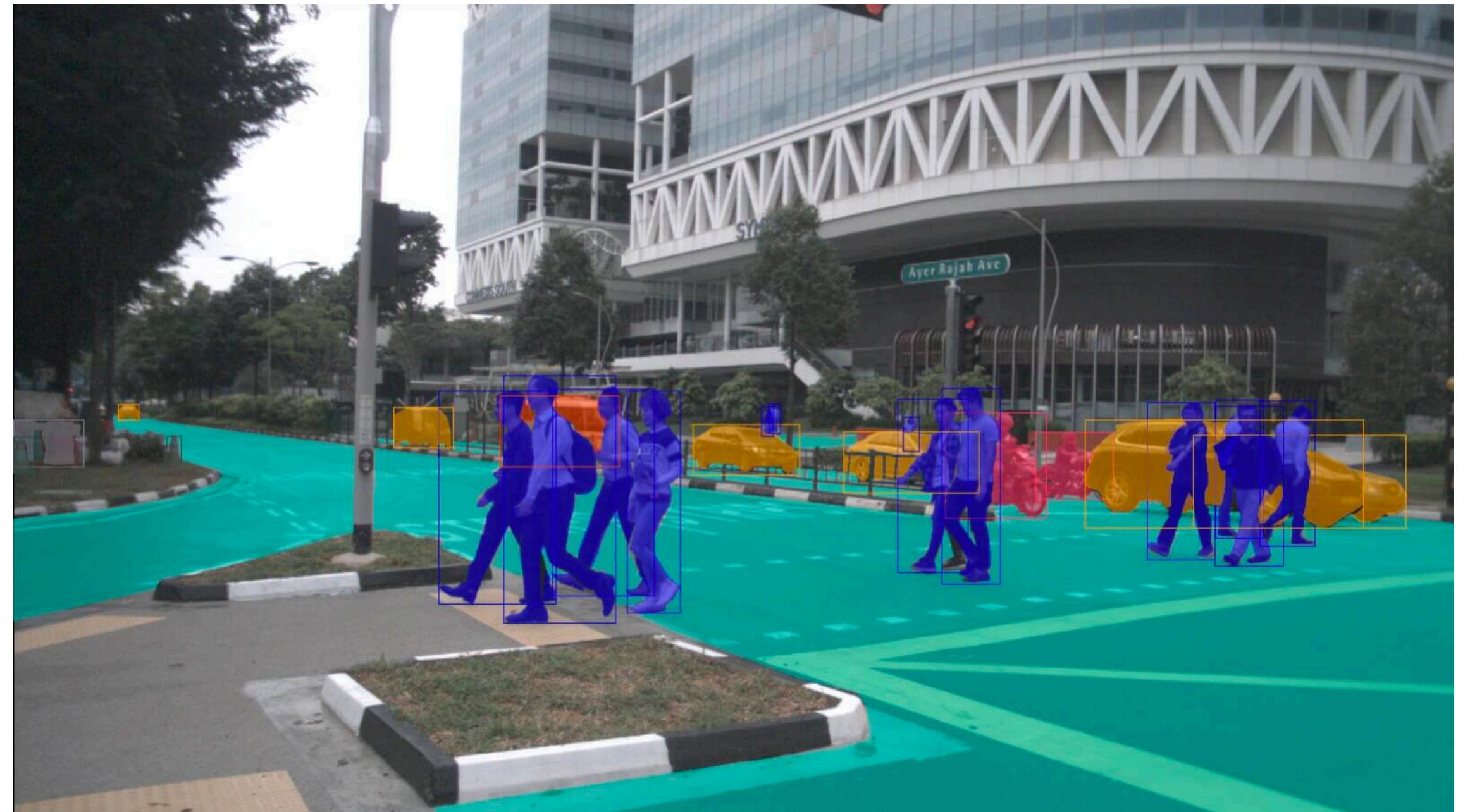
Real World Is Complex



NuScenes

One of the largest self-driving datasets available.

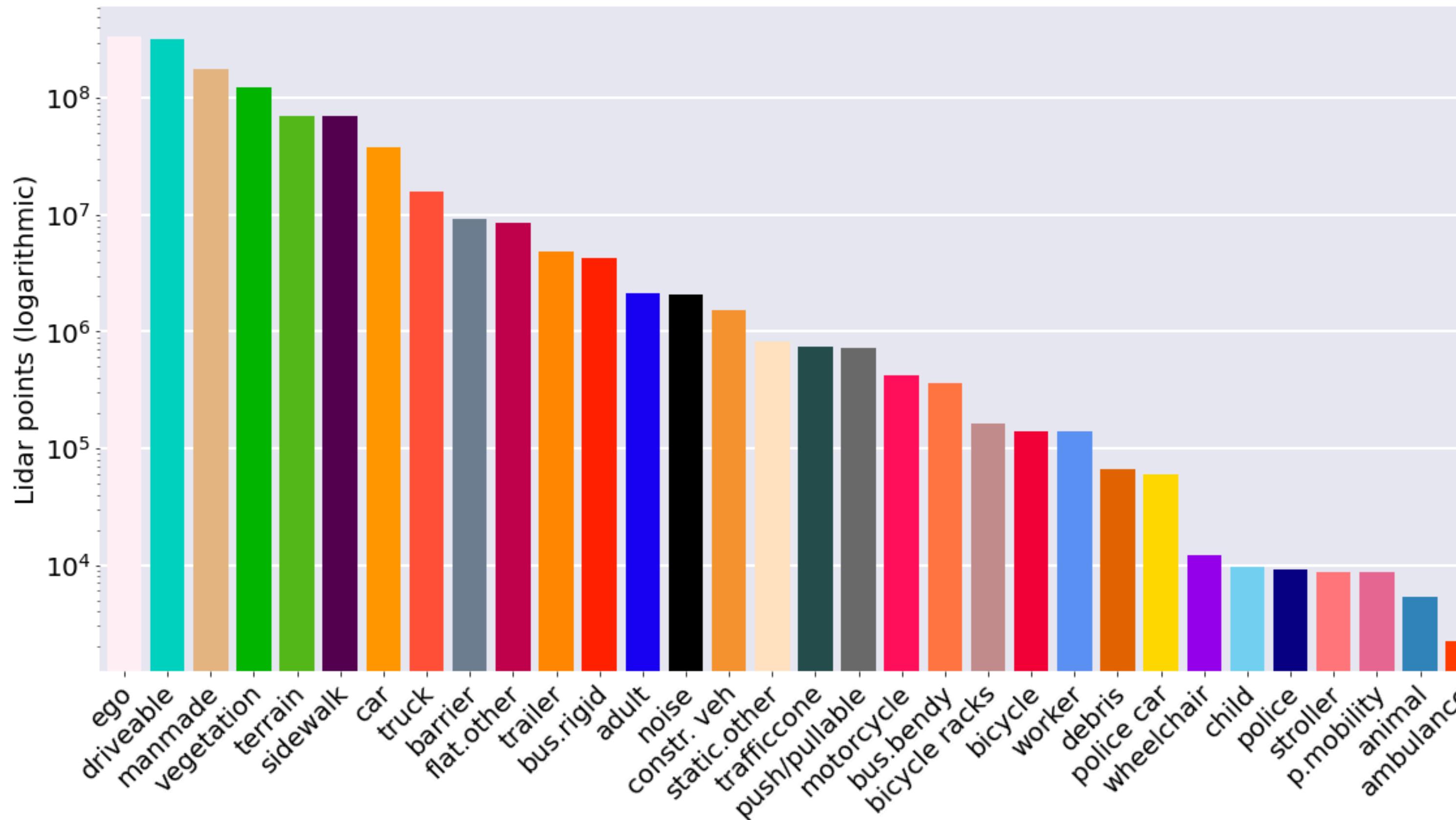
- › 1,000 scenes
- › 1.4M camera images
- › 23 object classes



! What is the problem here?

NuScenes

One of the largest self-driving datasets available.

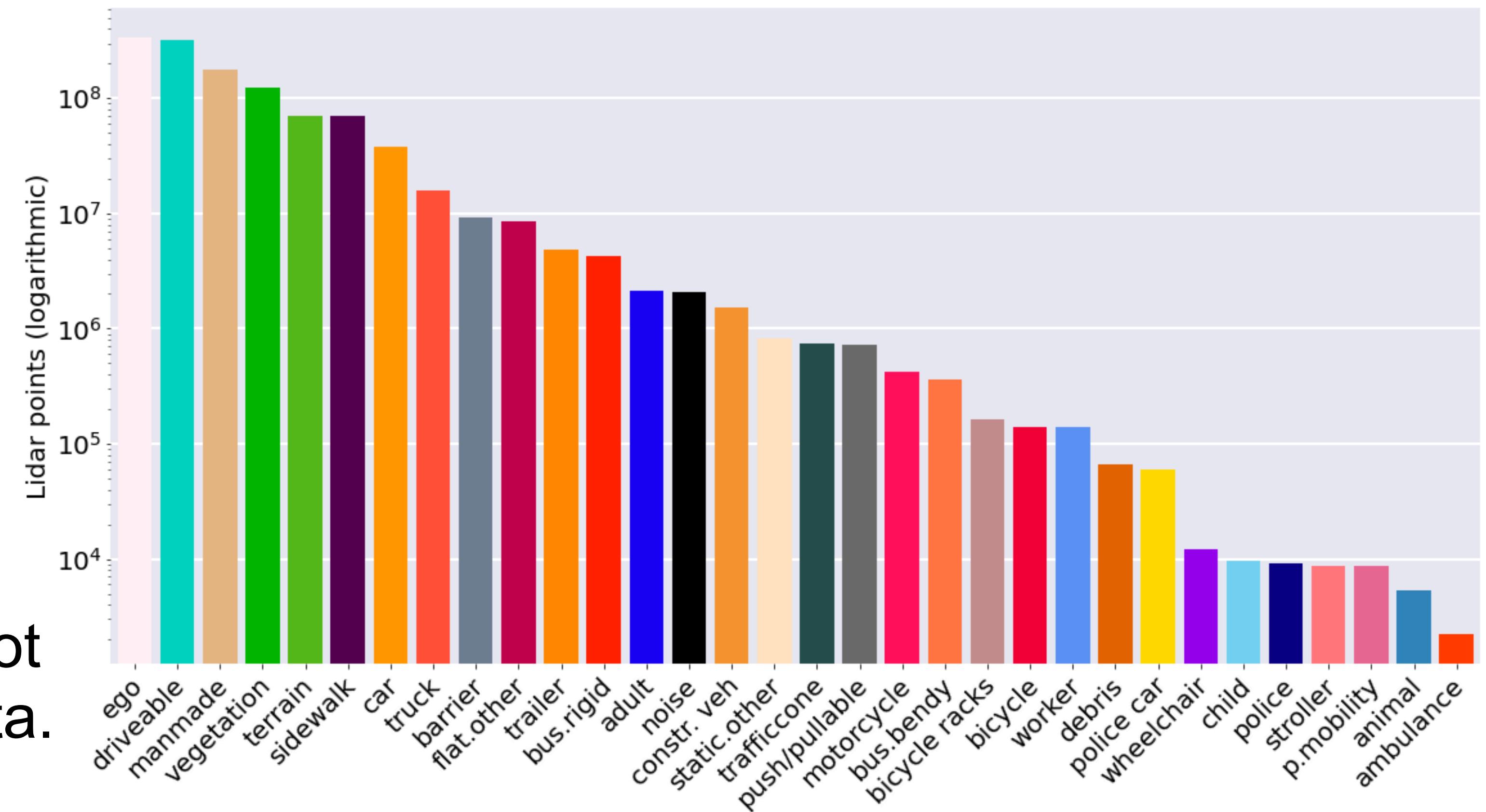


! Can you spot the problem?

NuScenes

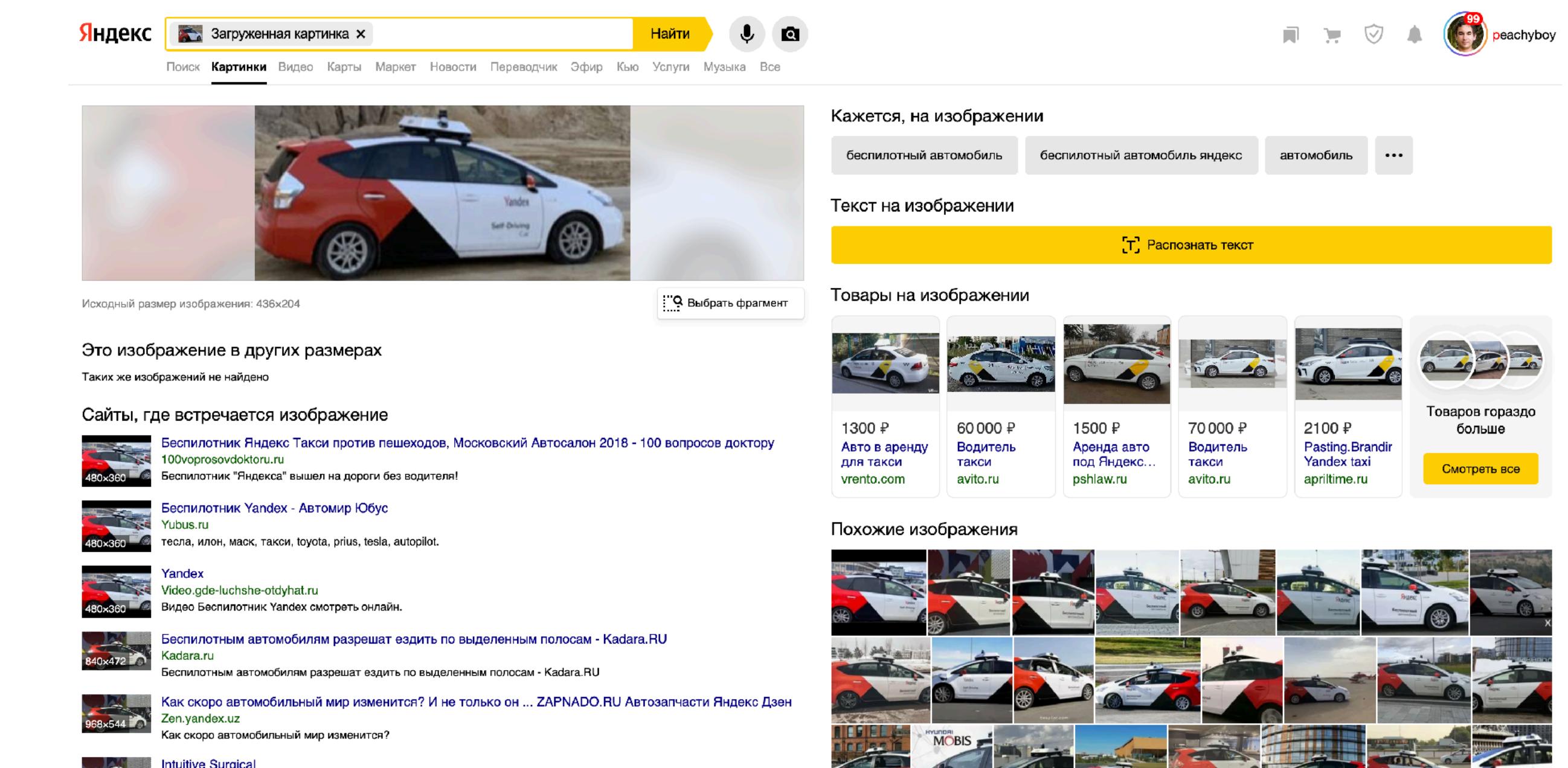
One of the largest self-driving datasets available.

- › Rare classes are exceptionally hard to mine with brute force.
- › Neural networks need to be trained on balanced data.
- › Oversampling, stratification does not solve the lack of data.



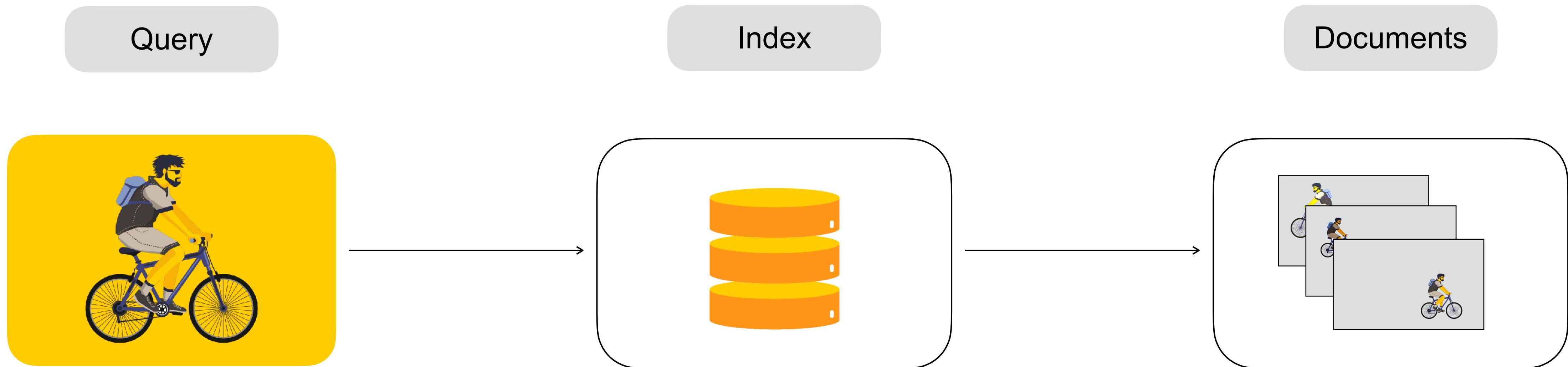
Yandex Image Search for SDG

- › Let's make an image search service **for our own data**, powered by Yandex Visual Search.
- › Idea: automate data mining, add more new, rare training data easily.
- › Instead of random sampling or heuristics sampling, let's query our index and retrieve scenes with images we need.

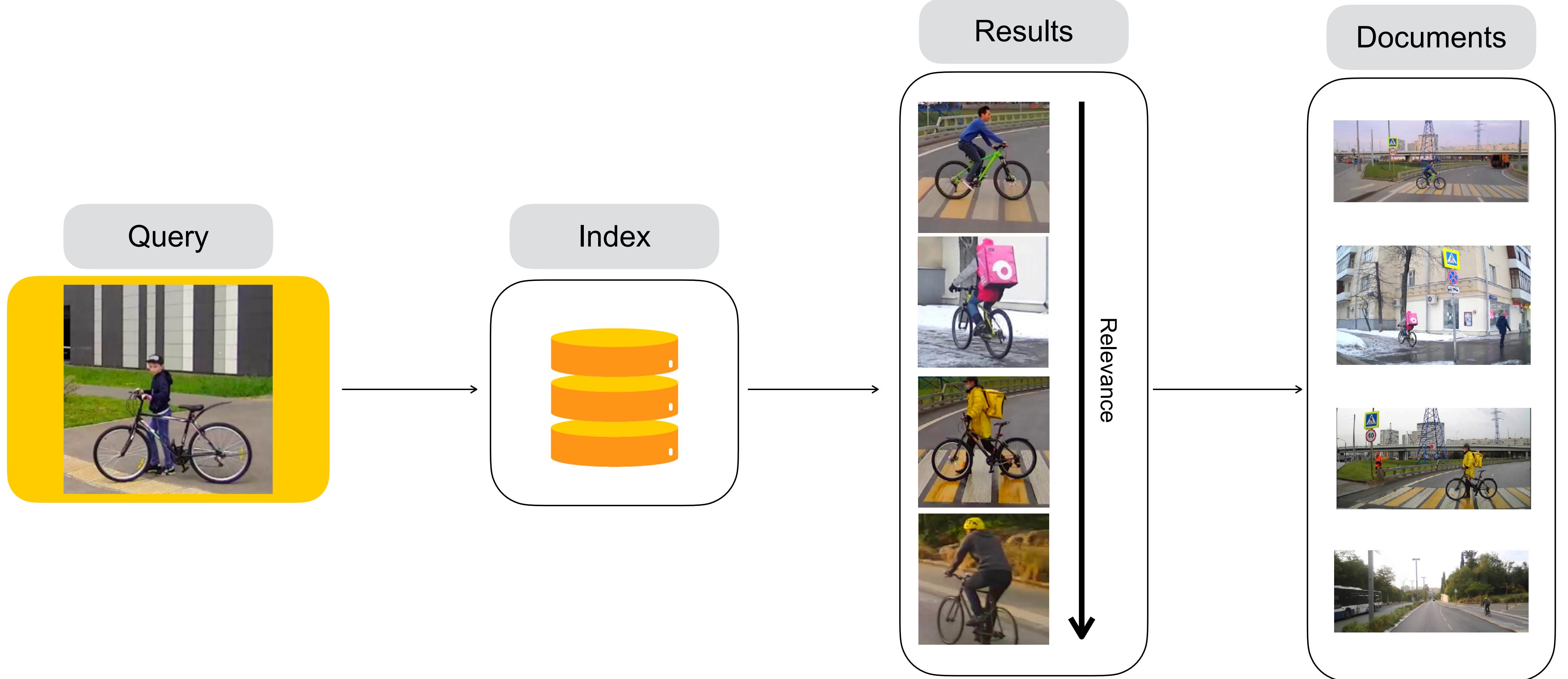


VISUAL SEARCH

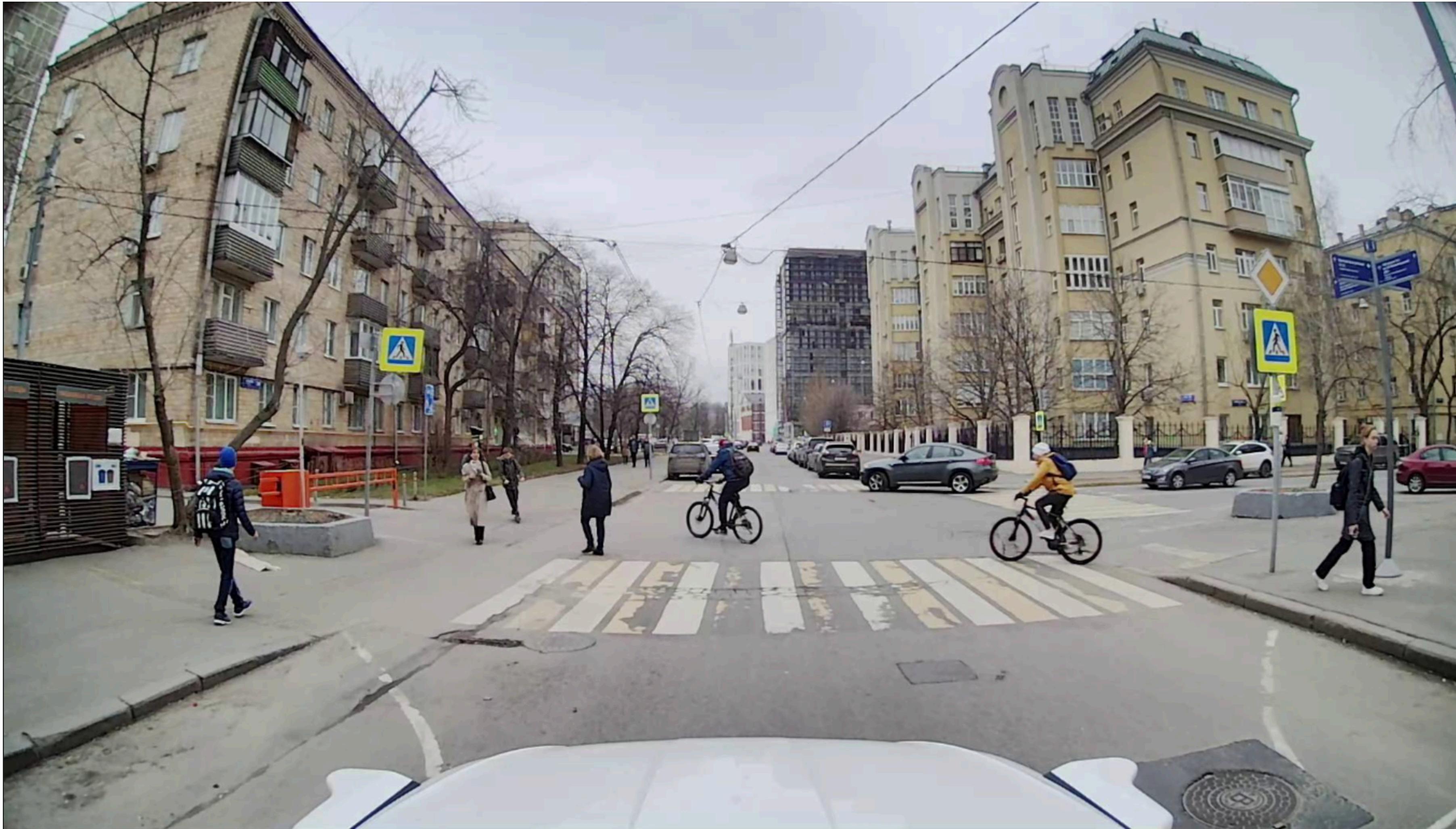
Visual Search User Workflow



Documents Retrieval

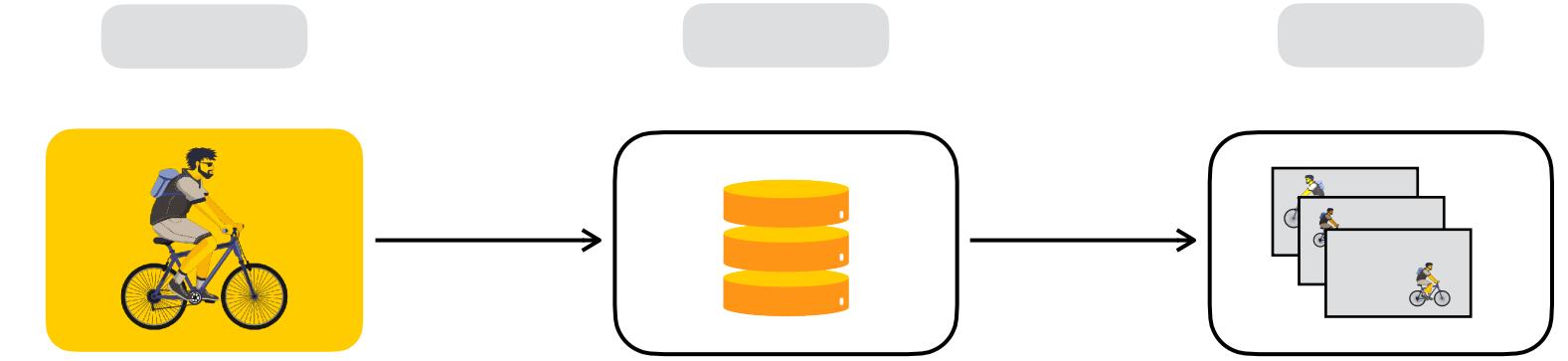


Documents: We Collect Petabytes of These



Basic Indexing Questions

- › What to store in the index?
- › How to compare the query with objects in the index?
- › How to retrieve most relevant documents?



Reviewing the Index

Index

- › Index contains crops of images, descriptors and ids of documents they are extracted from.
- › What to consider an image crop to be put into the index?
- › How to select most interesting crops to be put into the index?



image_crop

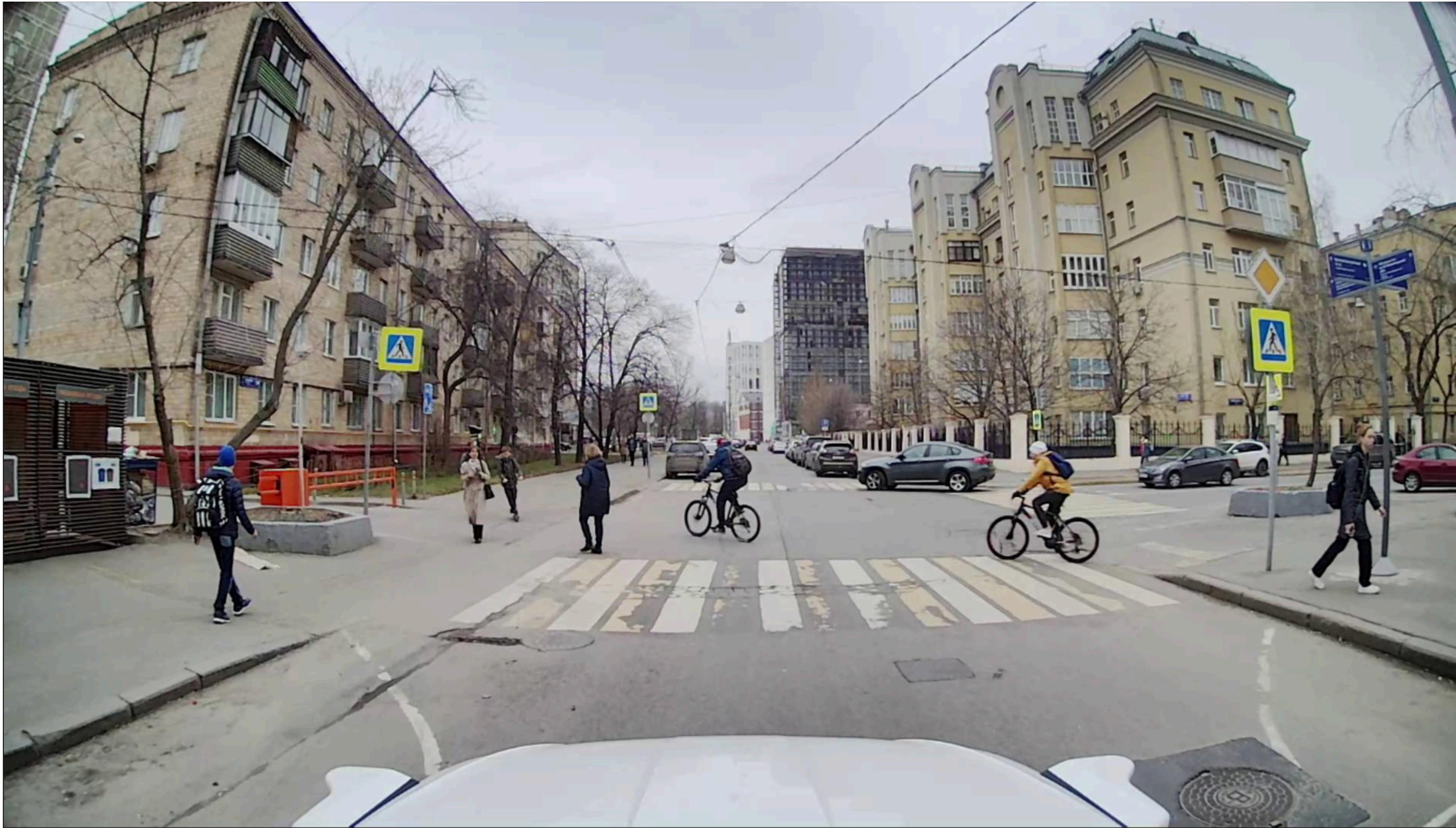
[-0.013,
-0.024,
0.045,...]

crop_descriptor

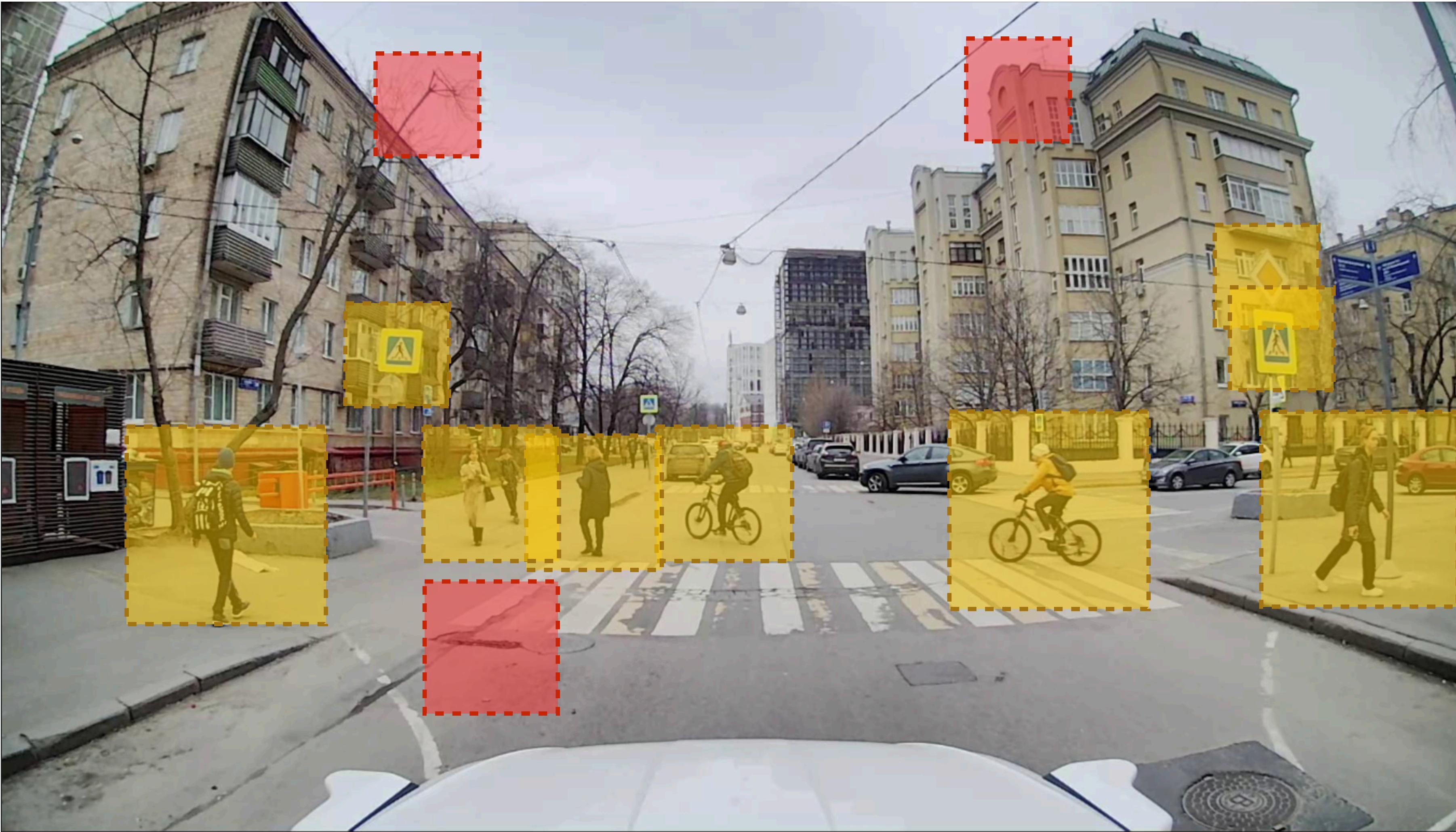


← document_id

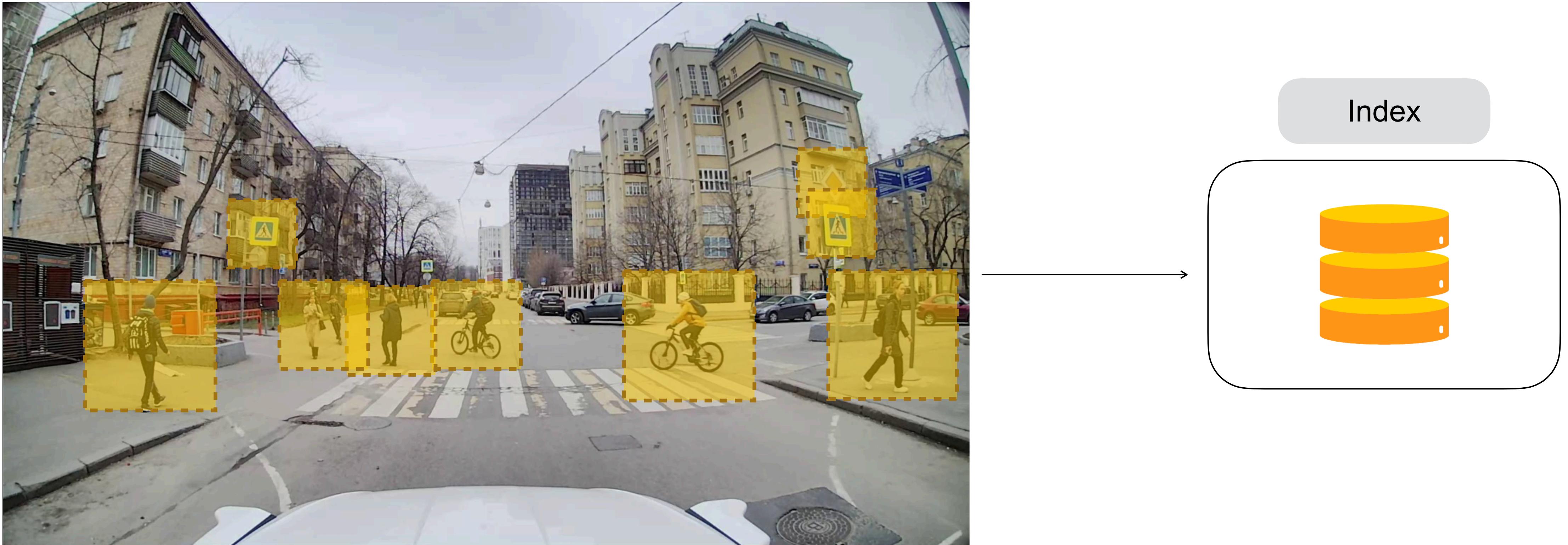
How to Select Proposals?



How to Select Proposals?

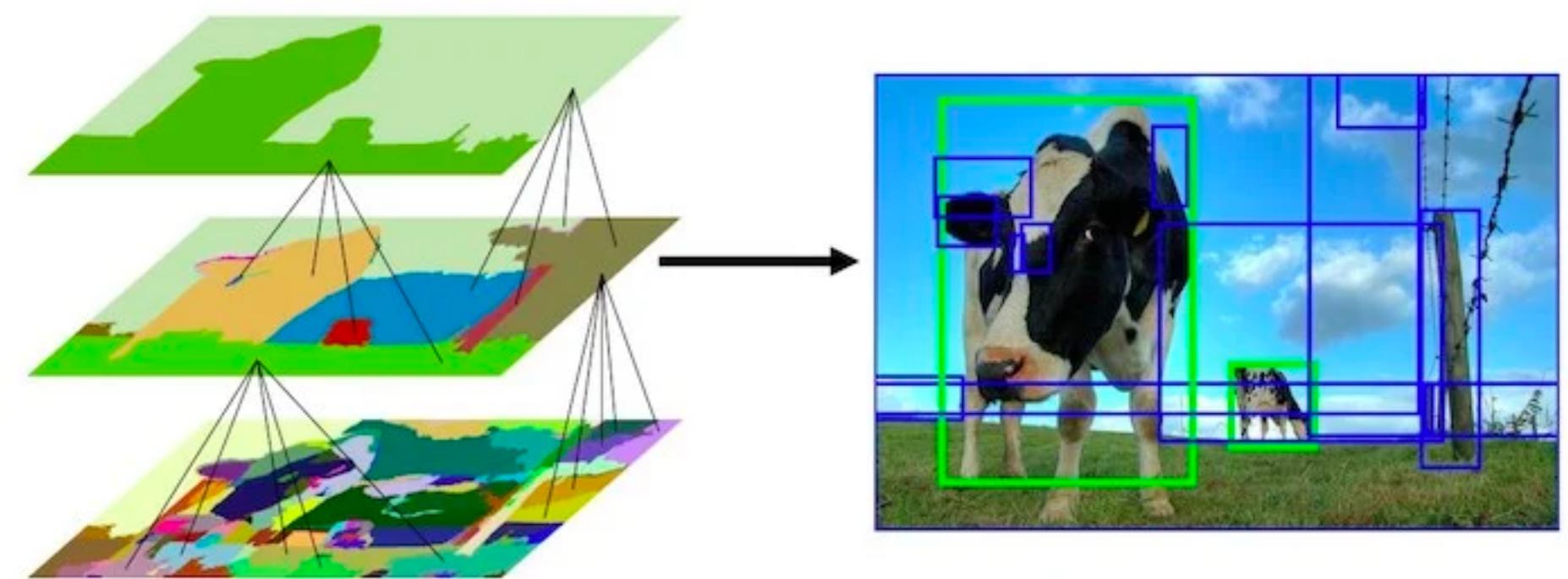
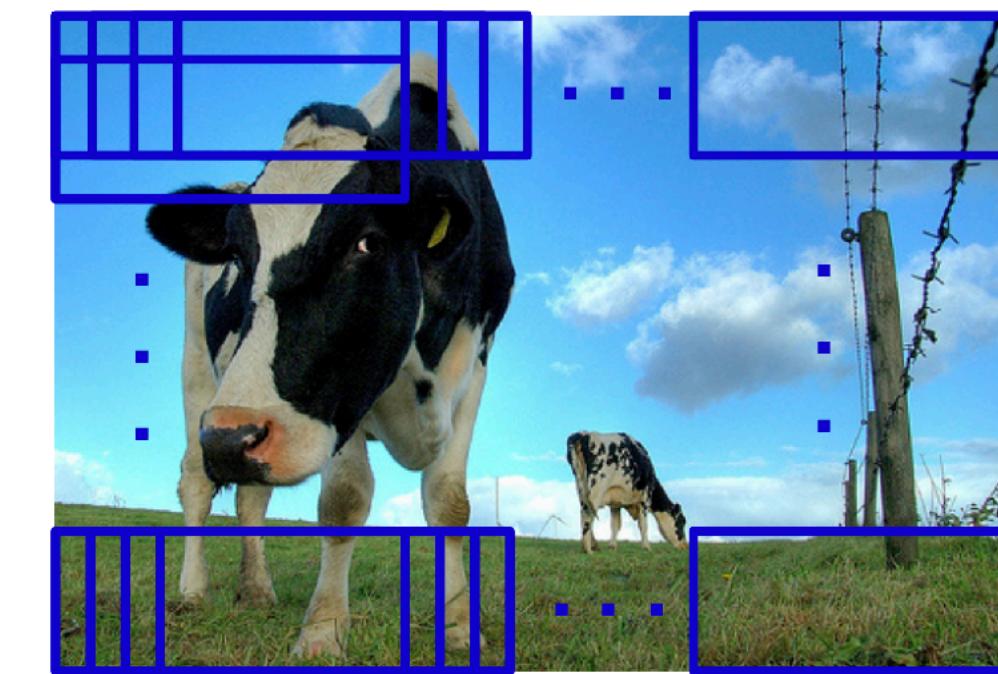


Interesting Proposals



Proposals Selection (1)

- › **Sliding windows:** slide a bounding box over an image, cover several different scales. Exhaustive search. Potentially high recall at the cost of precision.
- › **Selective search:** super-pixel grouping (oversegments) using similarity measurements (color, texture, shape). High recall, low precision, slow. Used in RCNN.

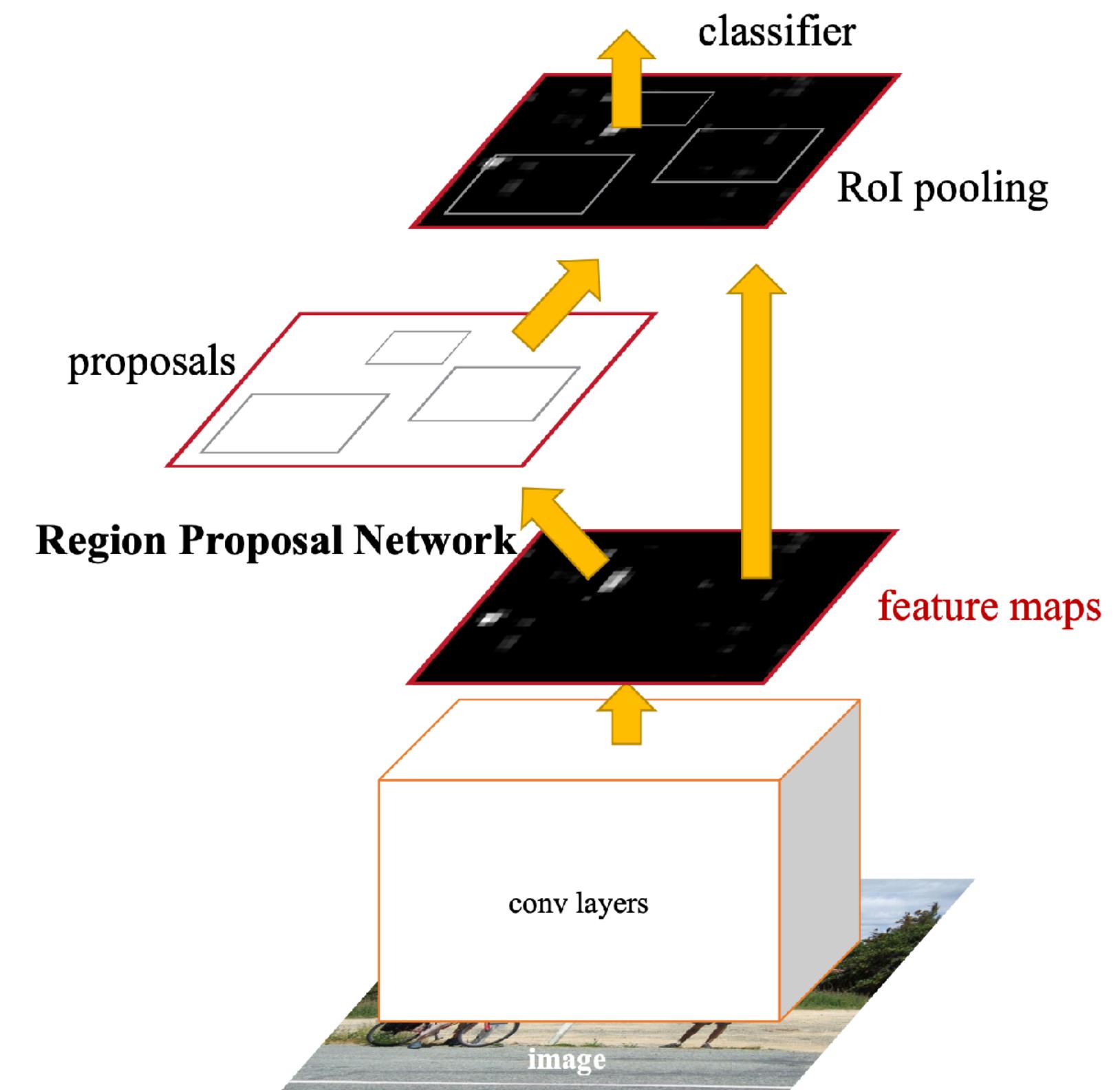


<https://www.koen.me/research/pub/vandesande-iccv2011-poster.pdf>

Proposals Selection (2)

- › **Trainable Proposals:** use a region proposal network (RPN) to generate proposals. Better precision, requires training and fine-tuning.

Used in Faster-RCNN.



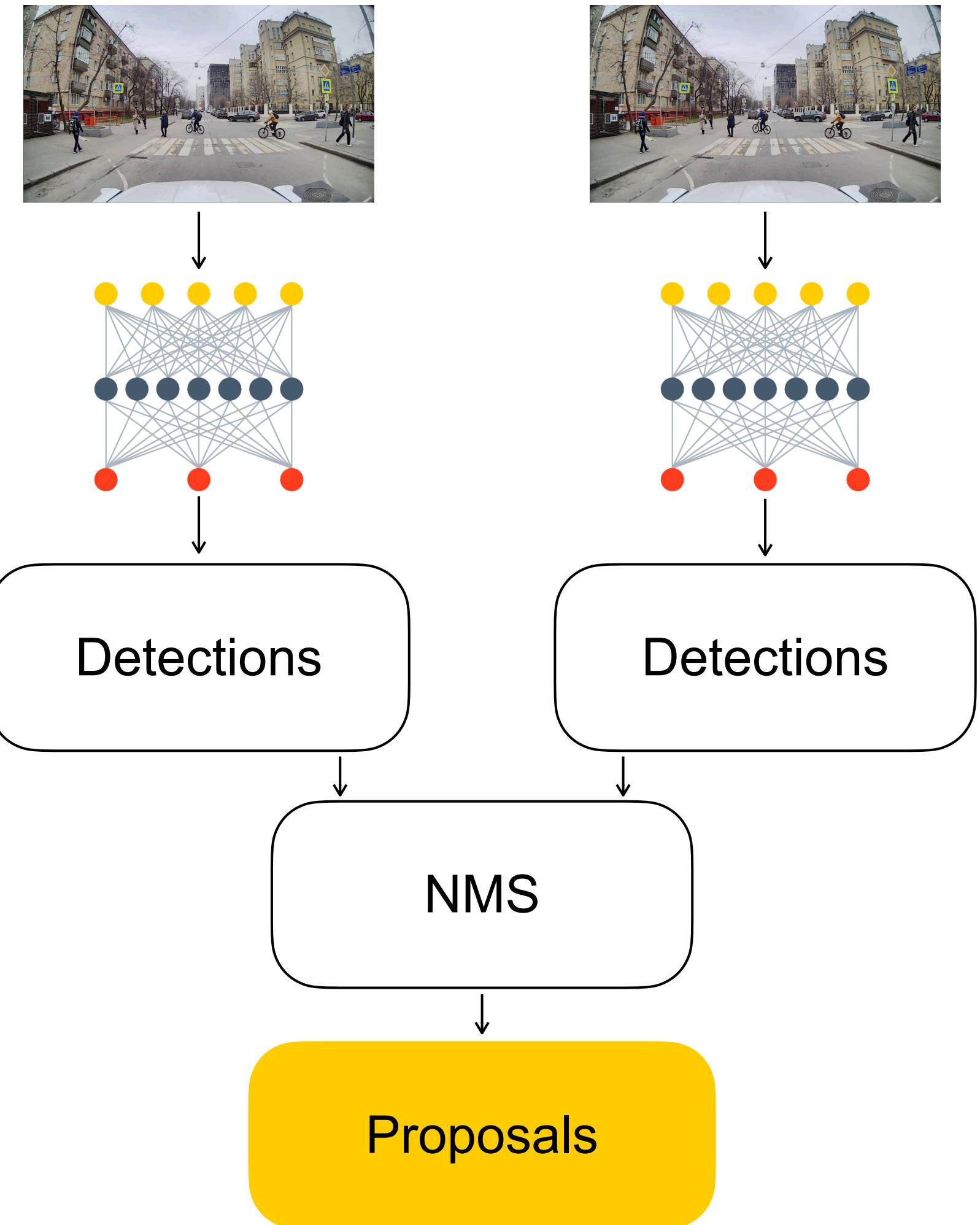
<https://arxiv.org/pdf/1506.01497.pdf>

Proposals Selection (3)

- › **Object detections from a trained object detection network:** combine detections from one or many pretrained object detection networks trained on relevant datasets.

Apply NMS to combined detections to obtain final proposals.

Even higher precision, lower recall.



Reviewing the Index

Index

- › Index contains crops of images, descriptors and ids of documents they are extracted from.
- › What to consider an image crop to be put into the index?
- › How to select most interesting crops to be put into the index?



image_crop

[-0.013,
-0.024,
0.045,...]

crop_descriptor



← document_id

Reviewing the Index

- › Index contains crops of images, descriptors and ids of documents they are extracted from.
- ›  What to consider an image crop to be put into the index?
- ›  How to select most interesting crops to be put into the index?

Index



image_crop

[-0.013,
-0.024,
0.045,...]

crop_descriptor



← document_id

Reviewing the Index

Index

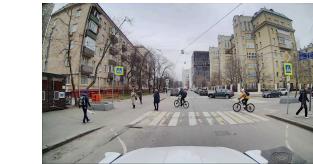
- › Index contains crops of images, descriptors and ids of documents they are extracted from.
- ›  What to consider an image crop to be put into the index?
- ›  How to select most interesting crops to be put into the index?
- ›  How to obtain an image crop descriptor?
- ›  How to make descriptors capture relevant semantic information?



image_crop

[-0.013,
-0.024,
0.045,...]

crop_descriptor



← document_id

Problem with Descriptors

- › Need to capture semantics and variance of any objects.
- › Color, shape descriptors are not useful here.
- › HOG / SIFT features are not robust enough.



Problem with Descriptors

- › Need to capture semantics and variance of any objects.
- › Color, shape descriptors are not useful here.
- › HOG / SIFT features are not robust enough.

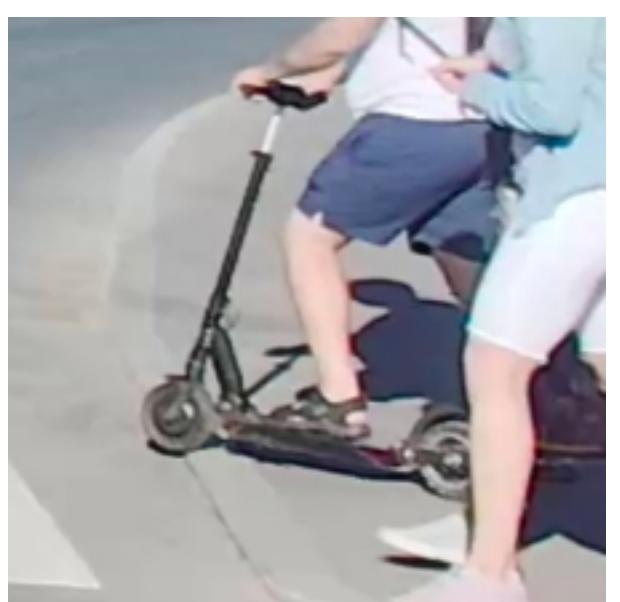
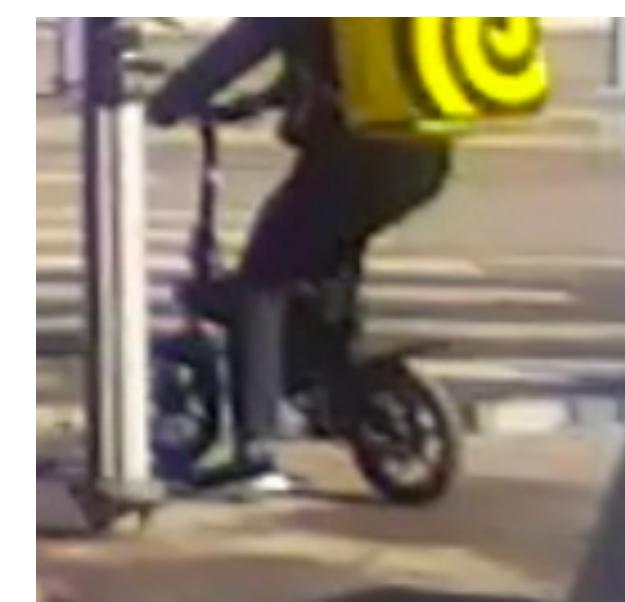
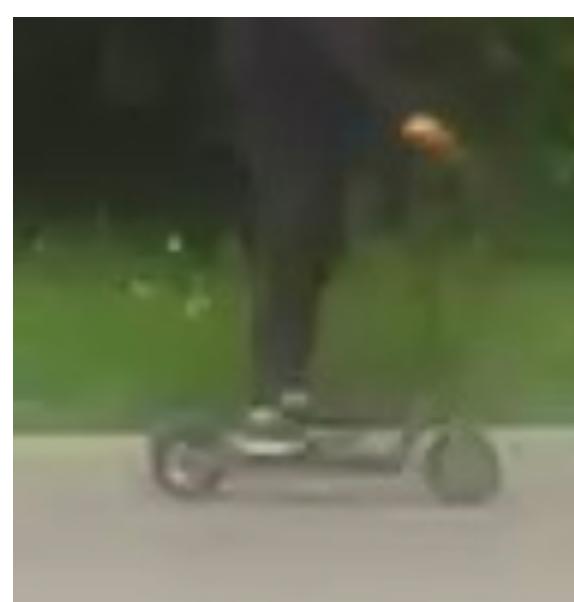
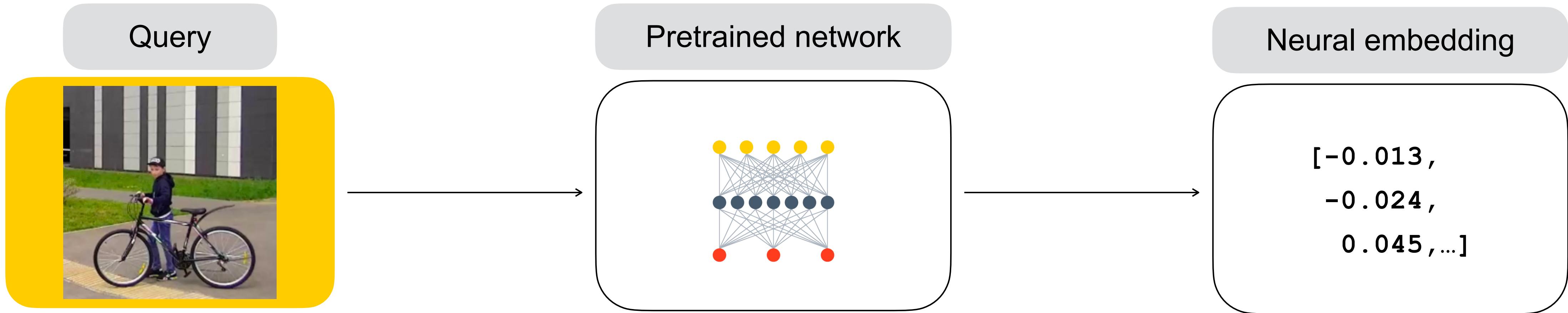


Image Embeddings



Neural Network for Embeddings



- › Applies the principle of knowledge transfer: a network used for image classification captures generic image features and semantics.
- › **Requirement:** should be very generic, should have a large feature extractor / backbone.
- › **Idea:** the CNN used for Yandex Image Search suits us well, let's use it.

The screenshot shows the Yandex Image Search results page. At the top, there is a search bar with the text "Загруженная картинка" and a "Найти" button. Below the search bar, there are navigation links for Поиск, Картинки, Видео, Карты, Маркет, Новости, Переводчик, Эфир, Кью, Услуги, Музыка, and Все. On the right side of the header, there are user profile icons and a notification badge with the number 98. The main content area displays a thumbnail of a person riding a bicycle. Below the thumbnail, the text "Исходный размер изображения: 224x224" is shown, along with a "Выбрать фрагмент" button. A section titled "Кажется, на изображении" lists tags: "на велосипеде", "человек", "ездить на велосипеде", and "катается на велосипеде". Below this, a section titled "Похожие изображения" shows a grid of smaller thumbnail images of people riding bicycles. At the bottom of the page, there is a link "Больше похожих".

Reviewing the Index

Index

- › Index contains crops of images, descriptors and ids of documents they are extracted from.
- ›  What to consider an image crop to be put into the index?
- ›  How to select most interesting crops to be put into the index?
- ›  How to obtain an image crop descriptor?
- ›  How to make descriptors capture relevant semantic information?



image_crop

[-0.013,
-0.024,
0.045,...]

crop_descriptor

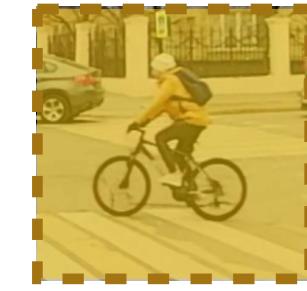


← document_id

Reviewing the Index

- › Index contains crops of images, descriptors and ids of documents they are extracted from.
- ›  What to consider an image crop to be put into the index?
- ›  How to select most interesting crops to be put into the index?
- ›  How to obtain an image crop descriptor?
- ›  How to make descriptors capture relevant semantic information?
- ›  How to compare embeddings?

Index



image_crop

[-0.013,
-0.024,
0.045,...]

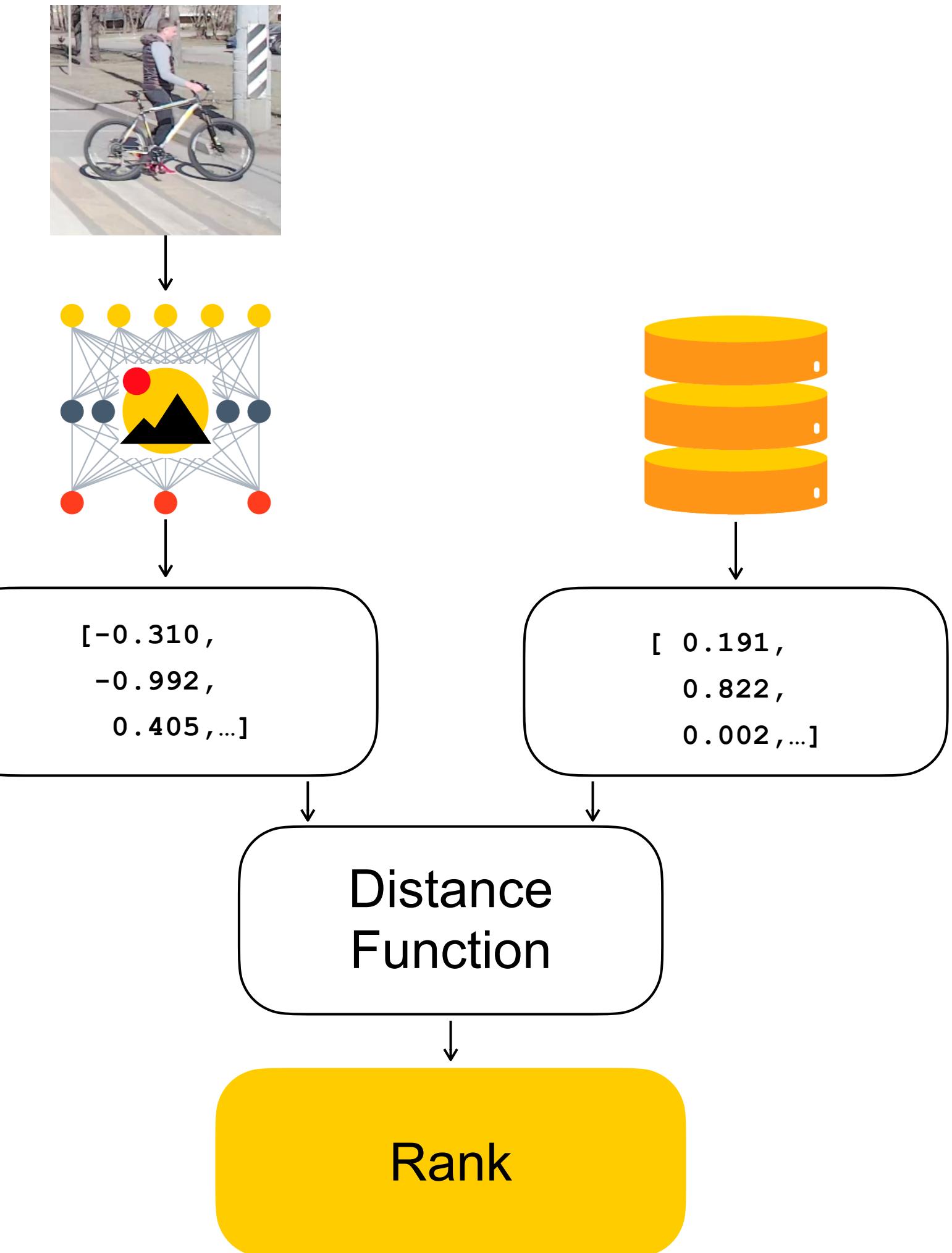
crop_descriptor



← document_id

Comparing Proposal Embeddings

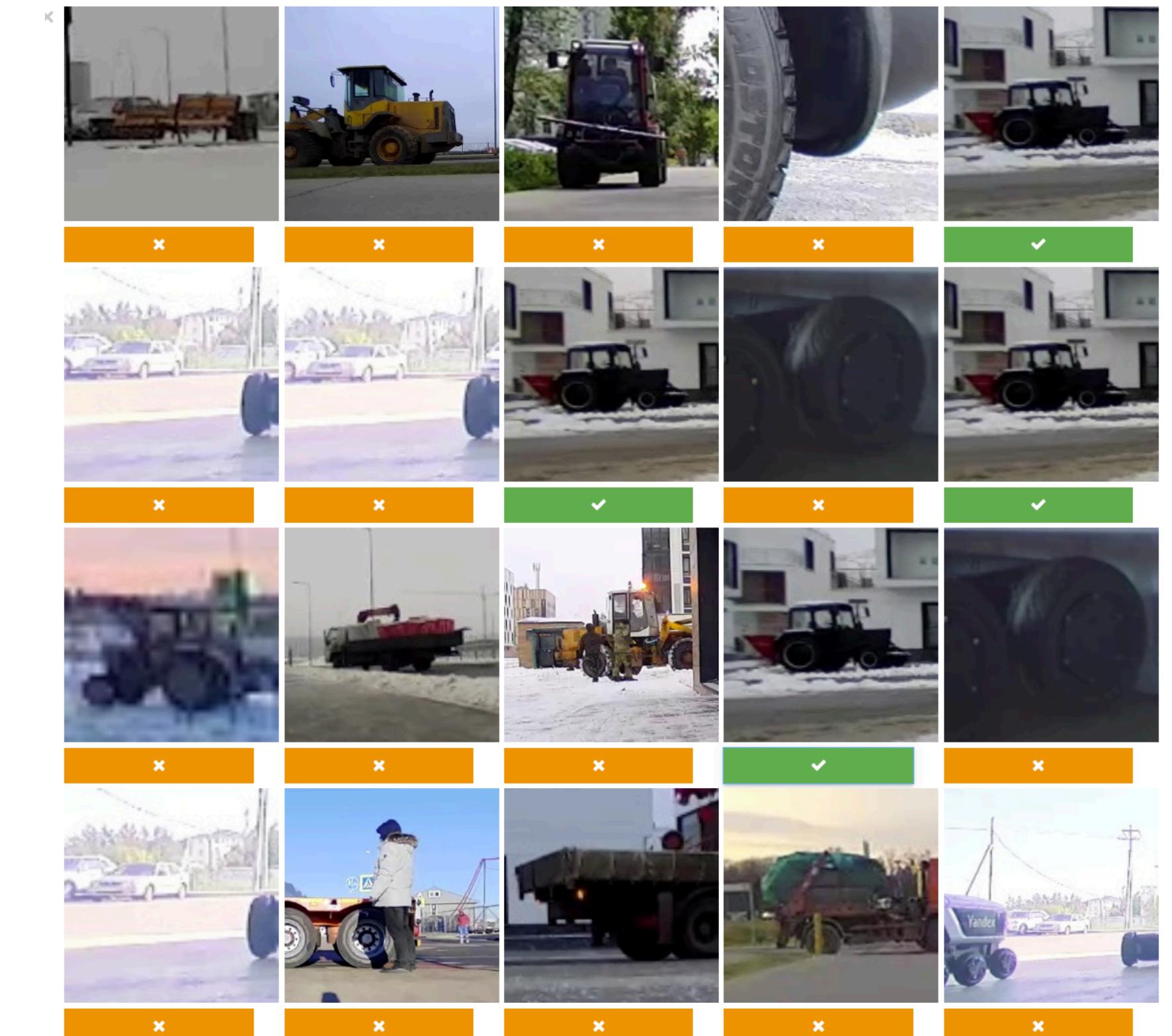
- › **Cosine similarity:** easiest to understand, requires no supervision, easy to implement.
Drawback: may not capture non-linear interactions in embeddings.
Good baseline distance function.



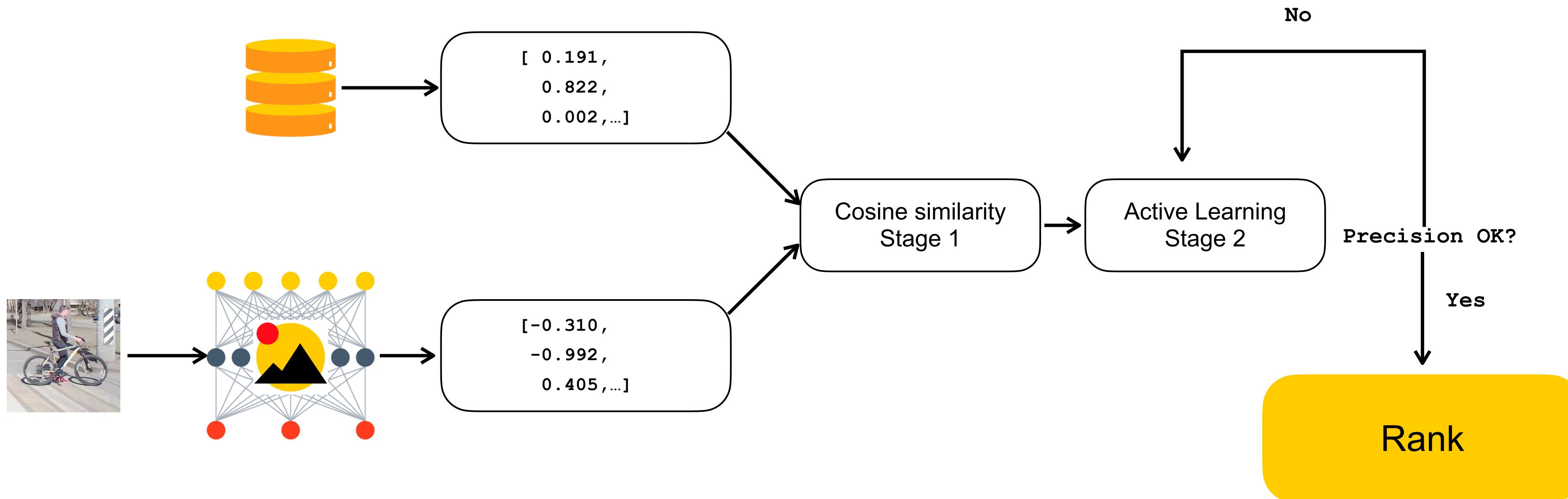
Comparing Proposal Embeddings

- › **Boosting as a distance function.** If we run a cosine similarity search, can we improve by asking a user to label (binary classification) a small chunk of results and by training gradient boosting (e.g. Catboost) on it?

The answer is **yes**.

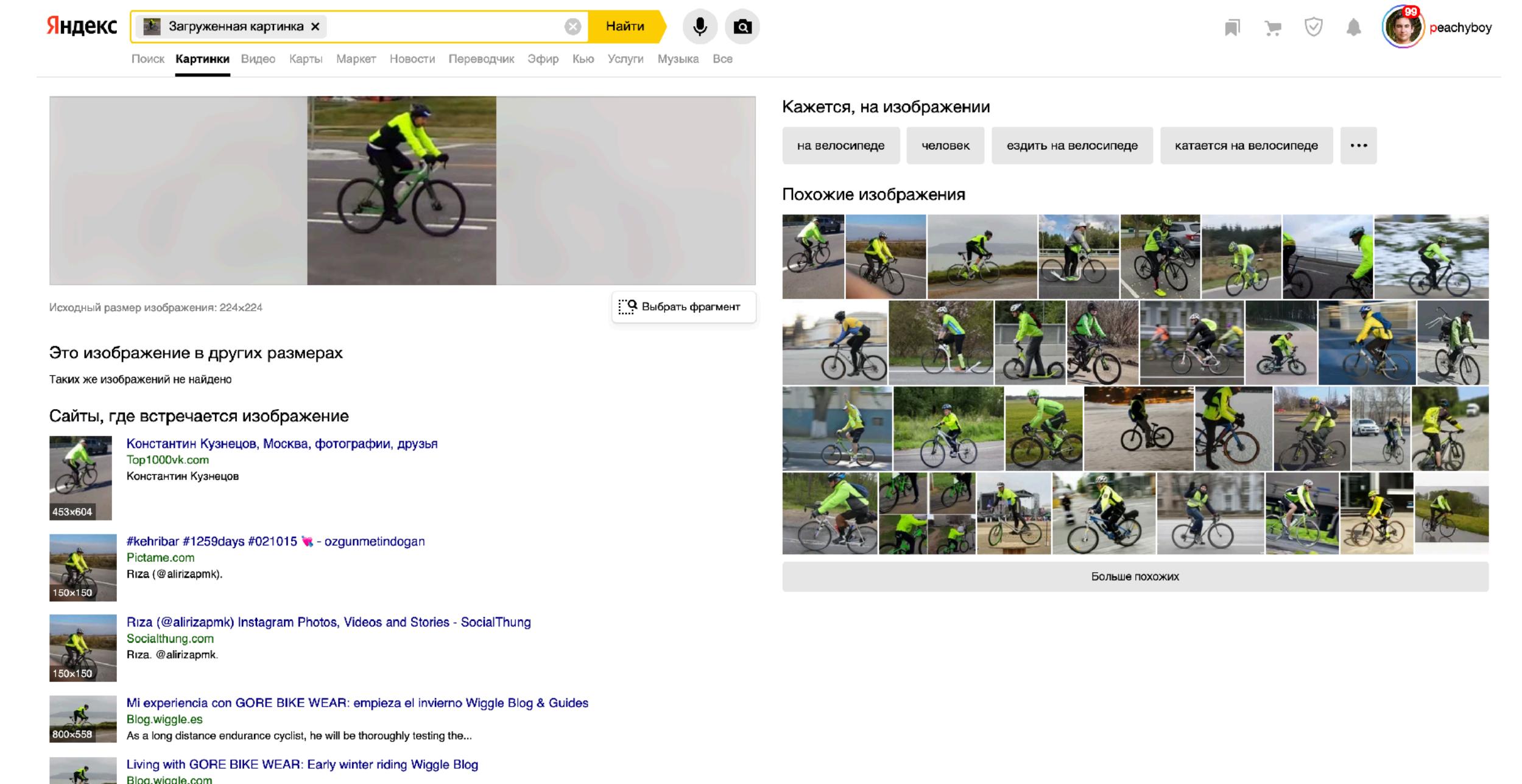


Two Stages of Proposals Comparison



Neural Network for Embeddings

- › Applies the principle of knowledge transfer: a network used for image classification captures generic image features and semantics.
- › **Requirement:** should be very generic, should have a large feature extractor / backbone.
- › **Idea:** the CNN used for Yandex Image Search suits us well, let's use it.





Thank you

Kirill Danilyuk

2D Computer Vision @ Yandex SDG

 dnkirill@yandex-team.ru

 [@kd_at_telegram](https://t.me/kd_at_telegram)