Multilingual Image Corpus

Svetla Koeva

Bulgarian Academy of Science

Main contributions of the project

The **Multilingual Image Corpus** is a large dataset containing thousands of images and object annotations in four thematic domains (Sport, Transport, Arts and Security) represented by 130 subdomains.

Main contributions of the project

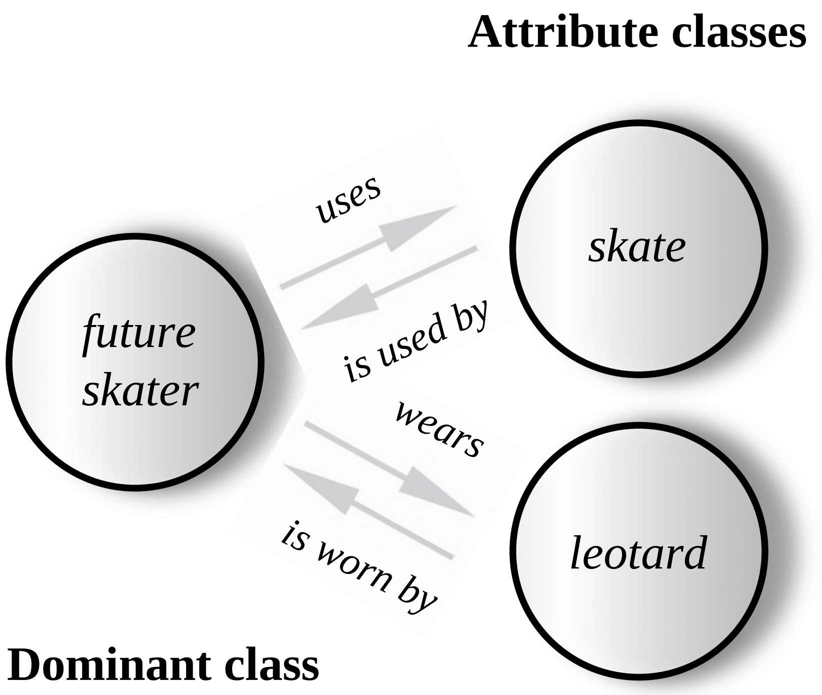
cc Search

WIKIMEDIA PEXELS flickr pixabay

The images are collected from a range of repositories offering APIs.

Main contributions of the project

The object segmentation and multi-label classification is provided for 706 object classes organised in a specially designed **Ontology of visual objects.**



Attribute classes

future skater

uses

is used by

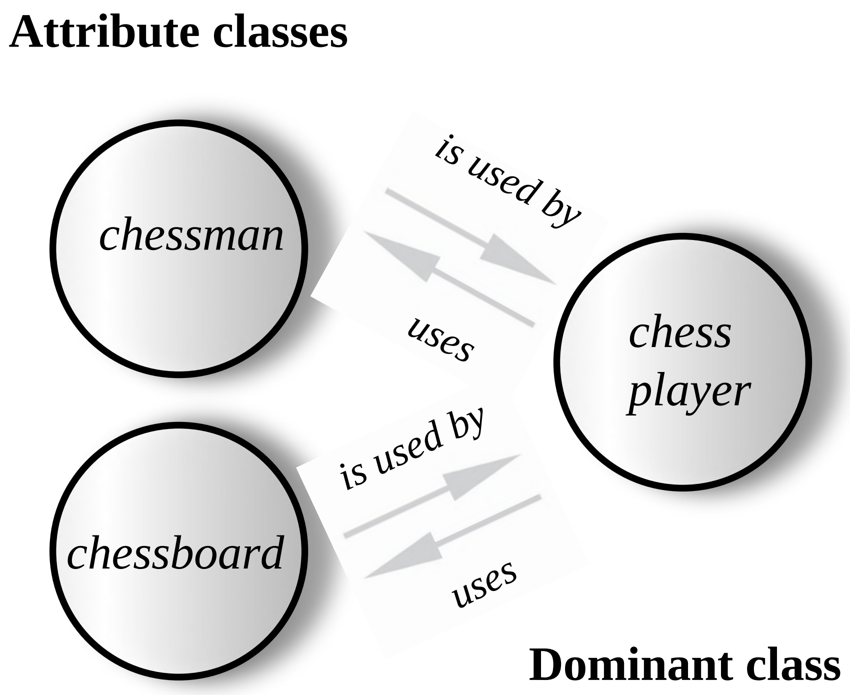
skate

wears

is worn by

leotard

Dominant class



Attribute classes

chessman

is used by

uses

chess player

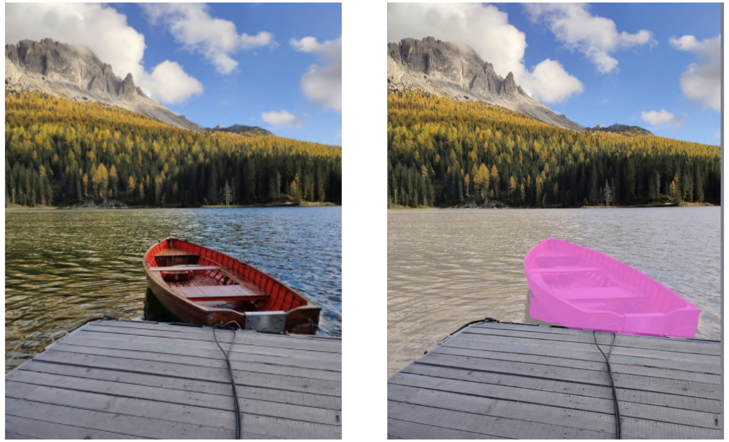
is used by

uses

chessboard

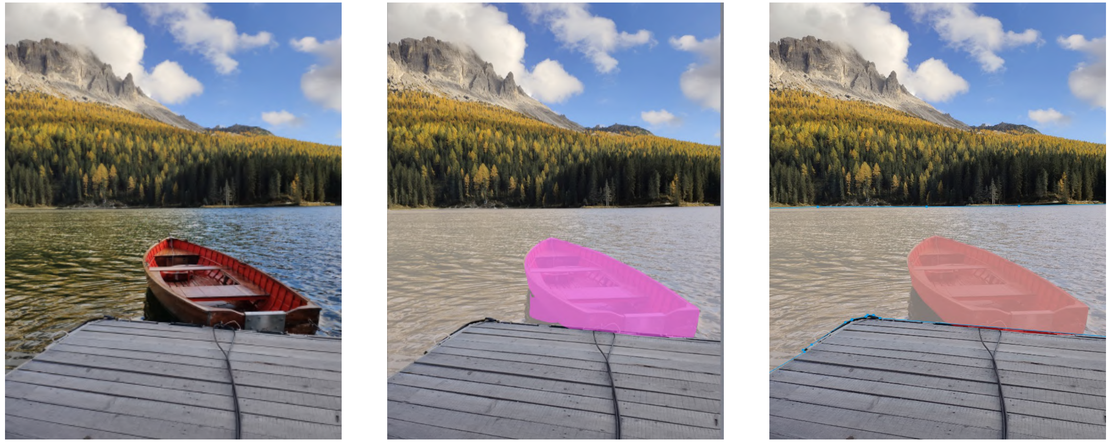
Dominant class

Object annotation



The task for the annotators was to outline polygons for individual objects in the image and to classify the objects against the classes from the predefined ontology.

Object annotation

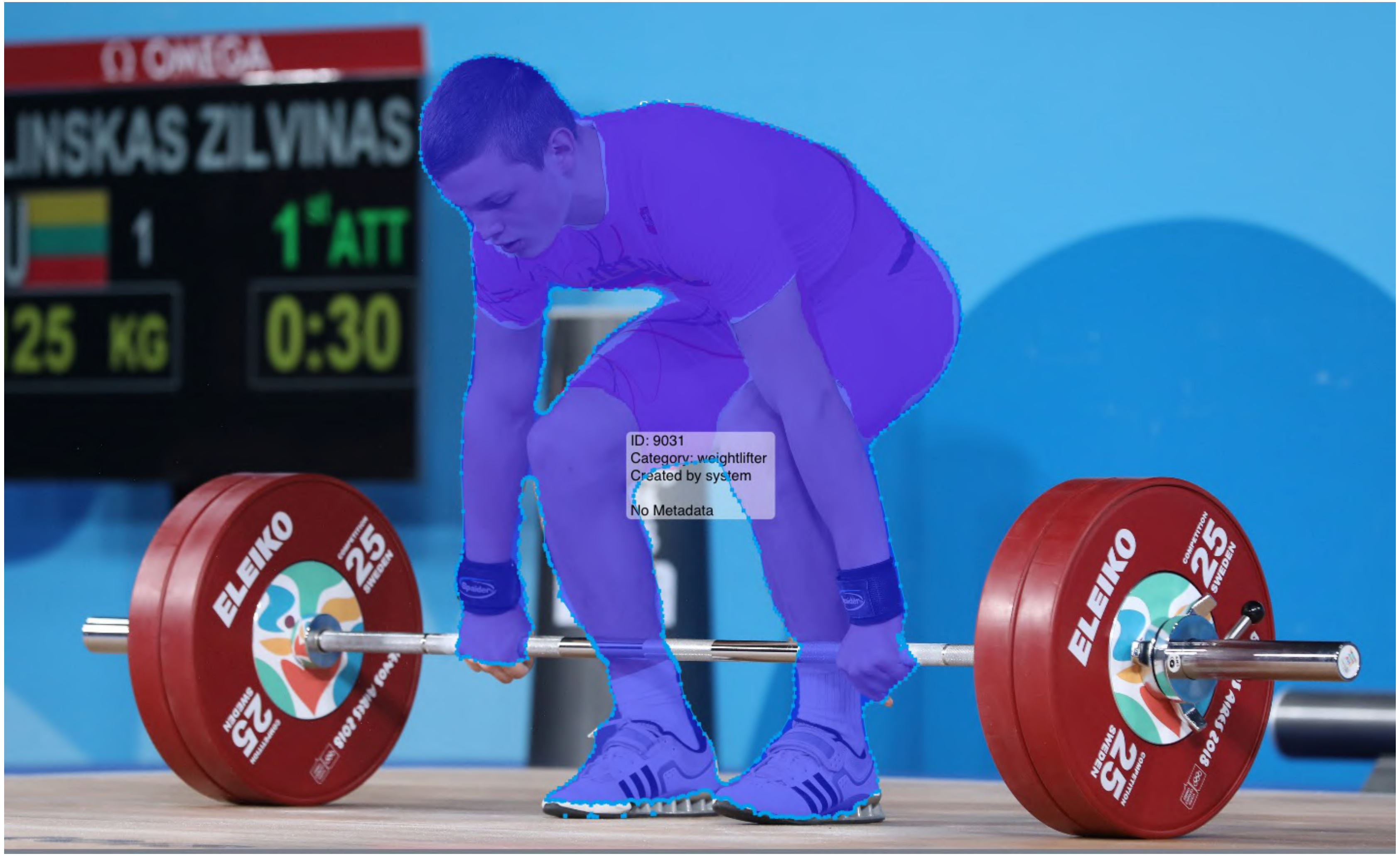


The task for the annotators was to outline polygons for individual objects in the image and to classify the objects against the classes from the predefined ontology.

We have used the COCO Annotator, which allows for collaborative work within a project, and offers tracking object instances and labelling objects with disconnected visible parts.

To accelerate the manual annotation, an image processing pipeline for object detection and segmentation was developed.

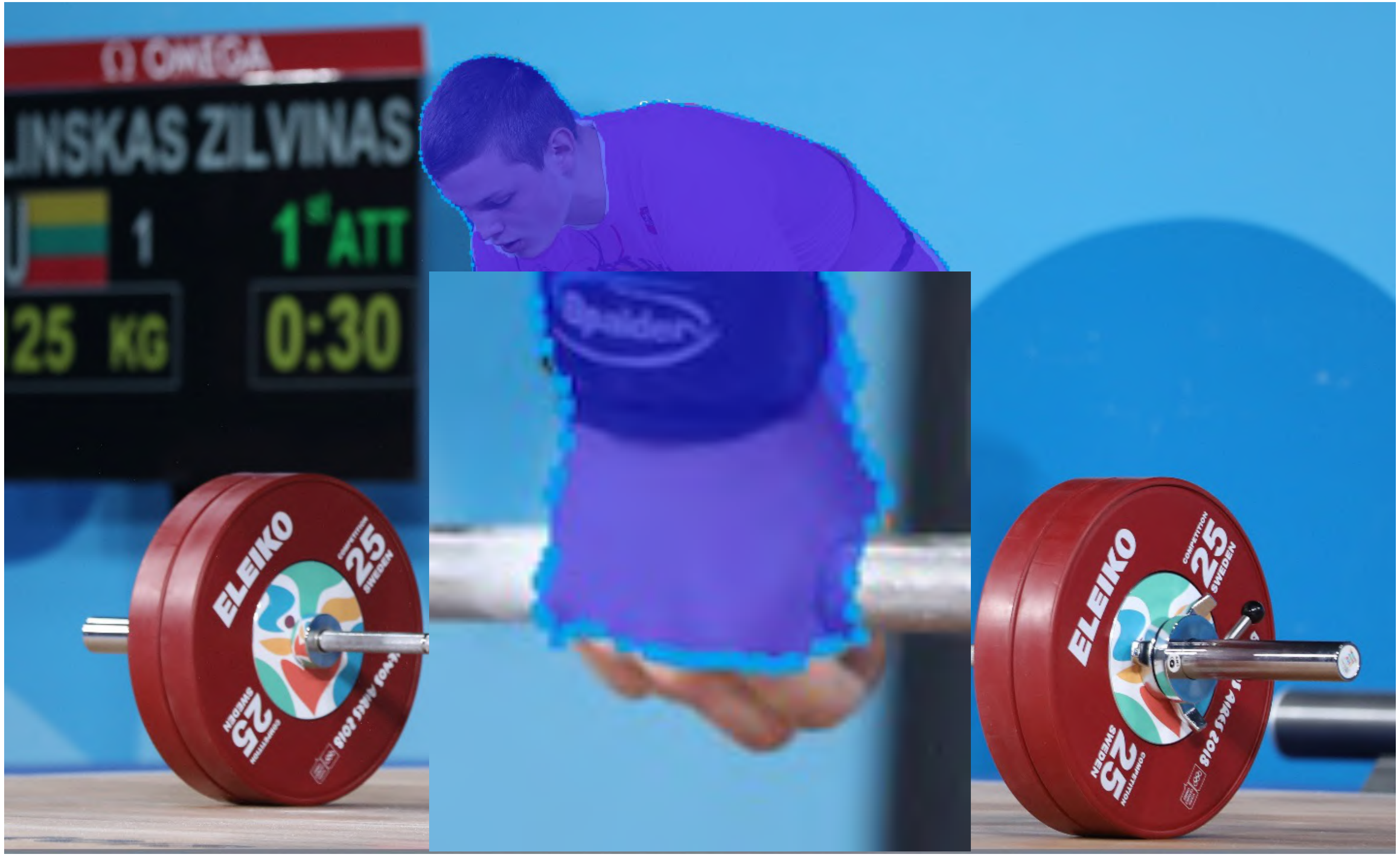
Two software packages – YOLACT and Detectron2, and Fast R-CNN models trained on the COCO dataset were used for the generation of annotation suggestions.



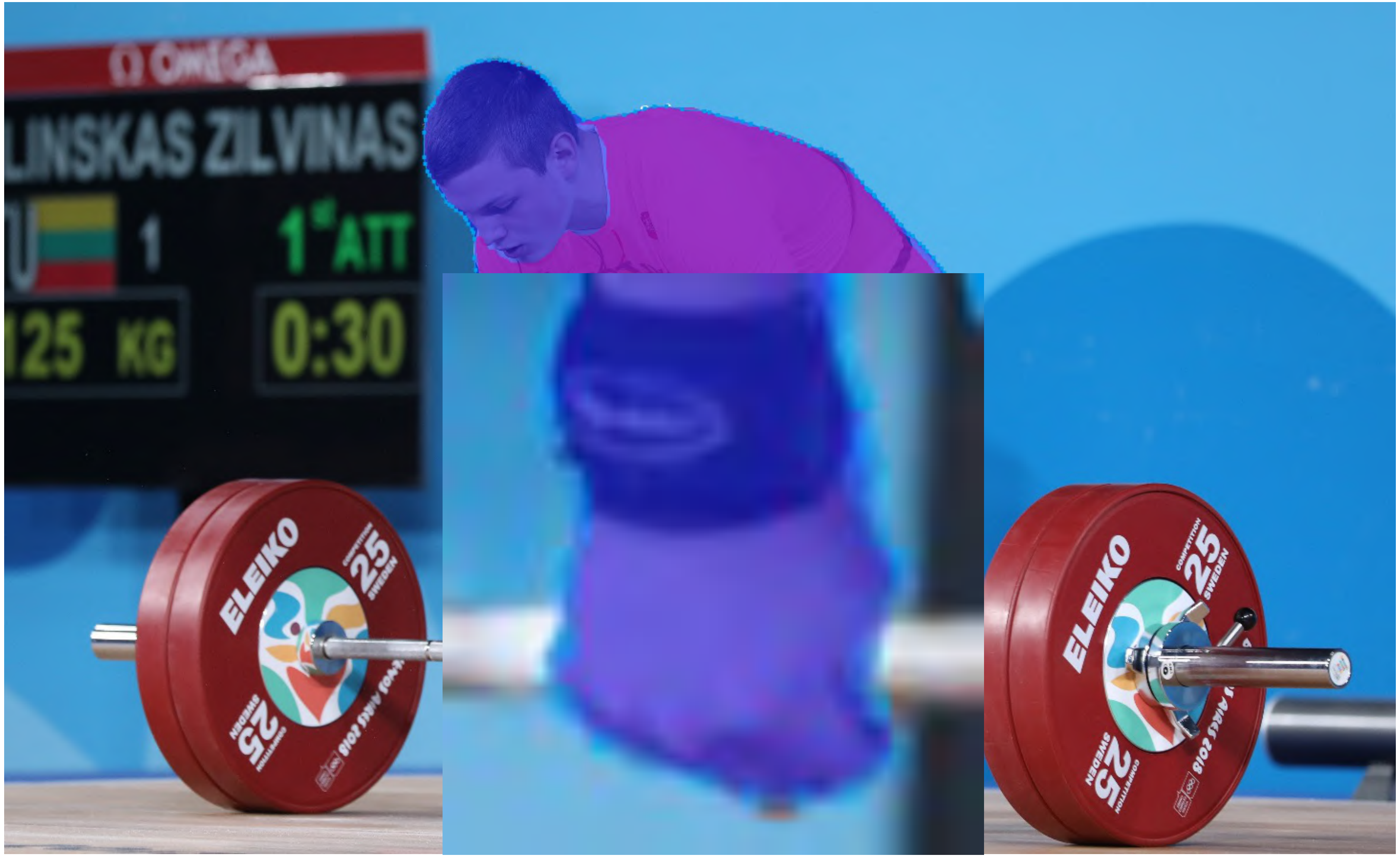
ID: 9031
Category: weightlifter
Created by system

No Metadata

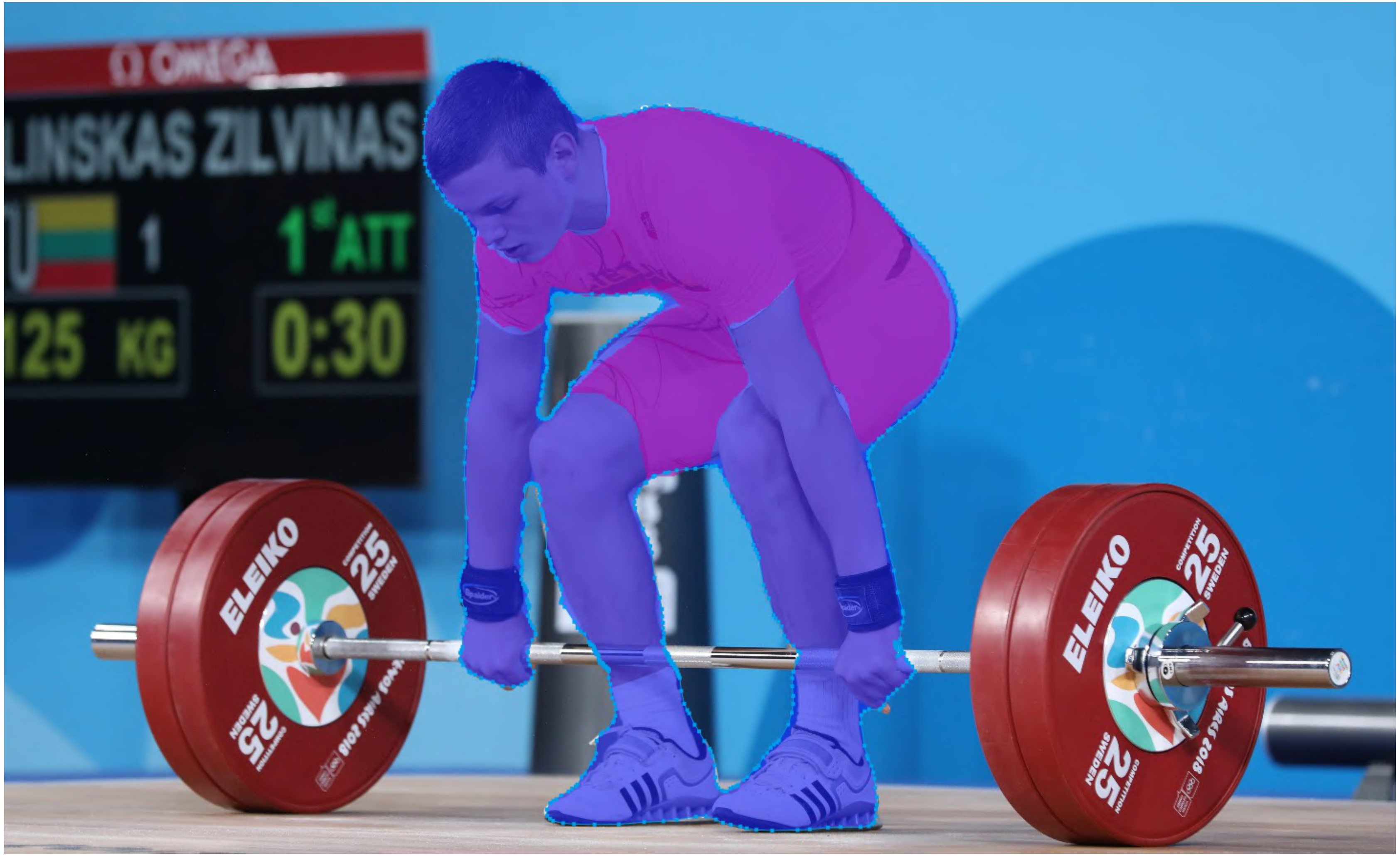
The manual annotation required adjusting or deleting object polygons in case they are not precise.



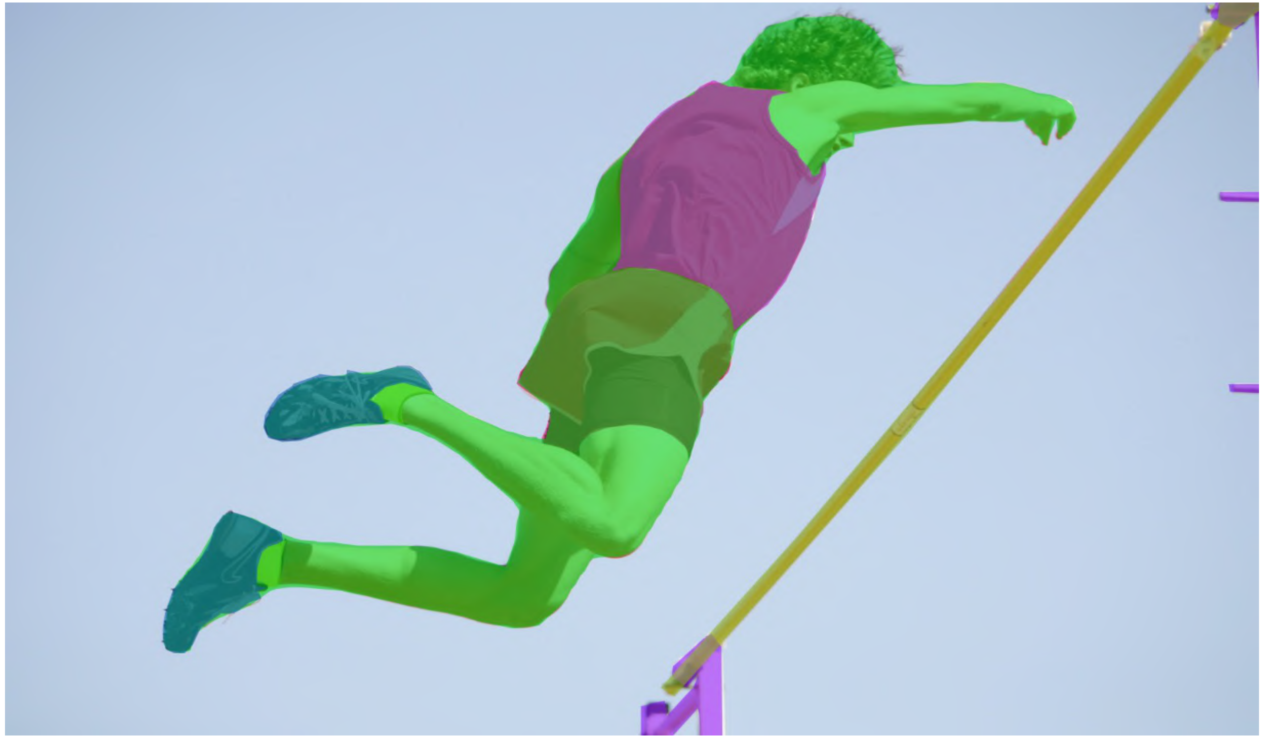
The manual annotation required adjusting or deleting object polygons in case they are not precise.



For the objects that are not within the COCO categories, the annotators created new polygons and classified the objects.



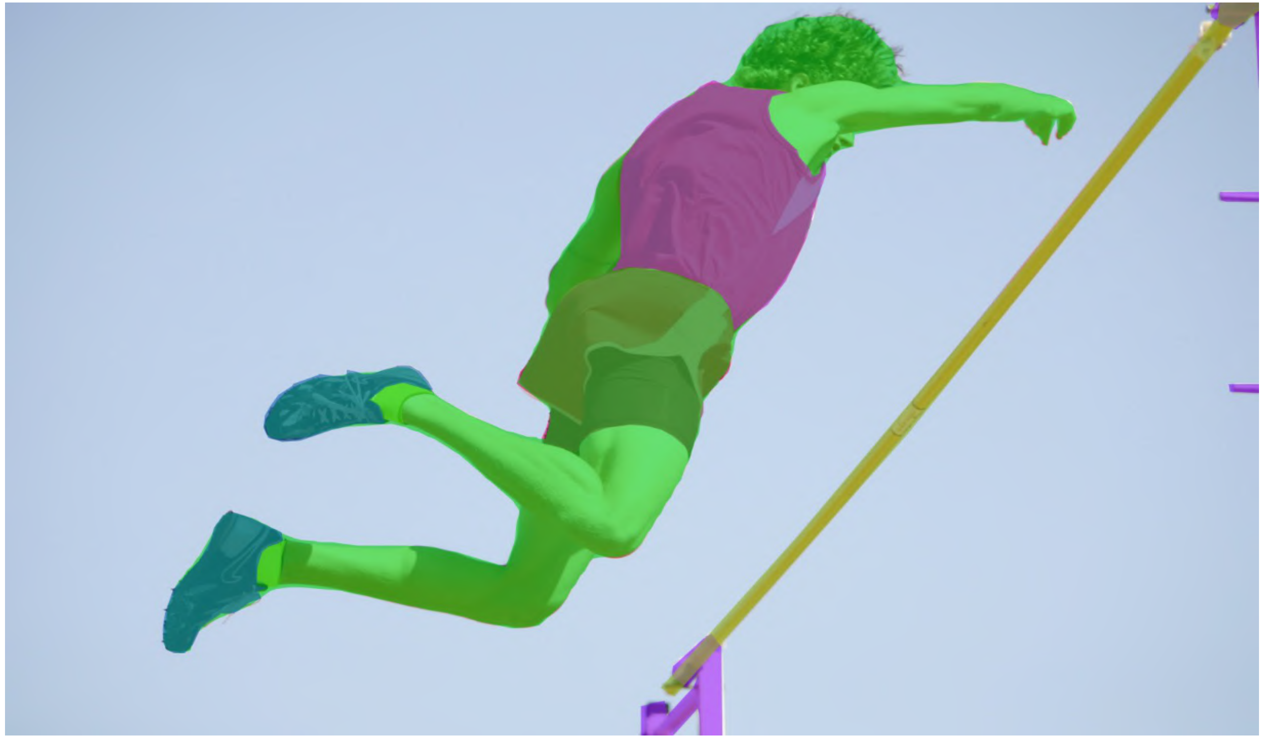
Multilingual Description of Ontology Classes



All Ontology classes (used as annotation labels) have been presented in 25 languages: *English* (Princeton WordNet), *Bulgarian, Albanian, Basque, Catalan, Croatian, Danish, Dutch, German, Greek, Finnish, French, Galician, Icelandic, Italian, Lithuanian, Polish, Portuguese, Romanian, Russian, Serbian, Slovak, Slovene, Spanish, Swedish.*

(1) Extracting translations from **WordNet**.
(2) From **BabelNet**.
(3) MT.

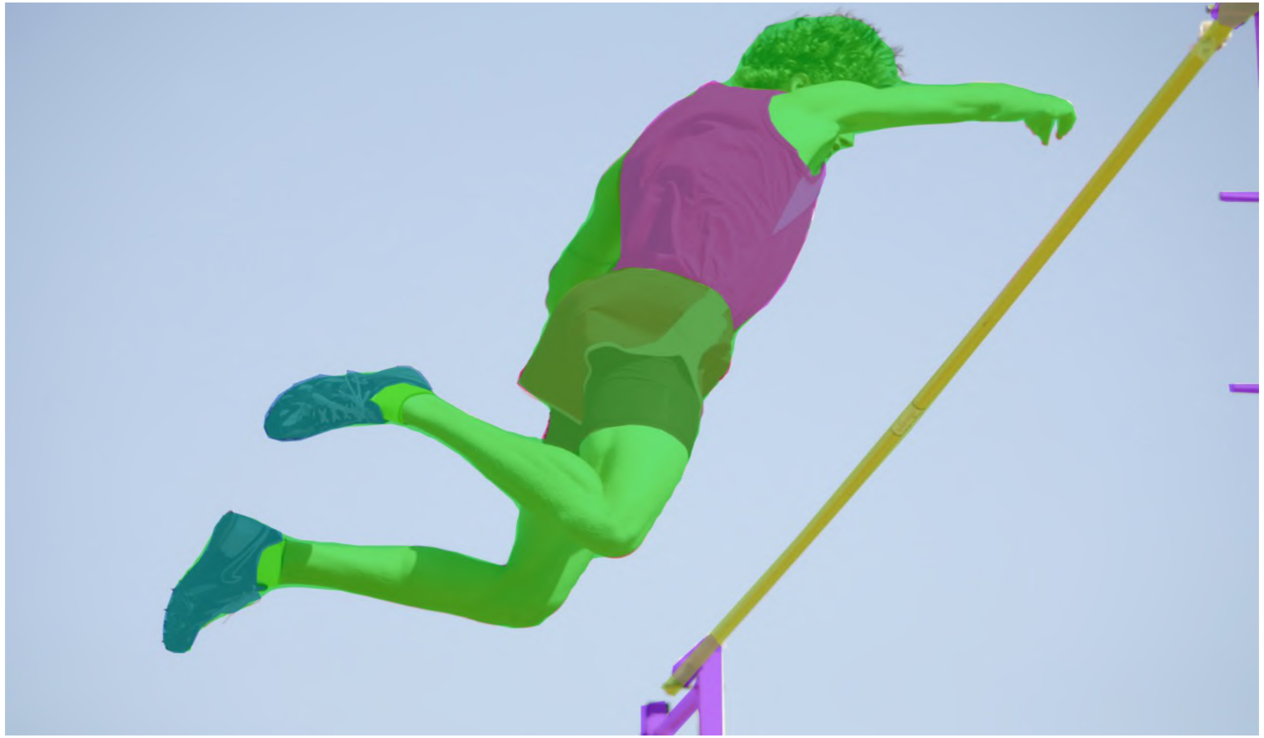
Multilingual Description of Ontology Classes



English (WN):

vaulter; pole vaulter; pole jumper

Multilingual Description of Ontology Classes

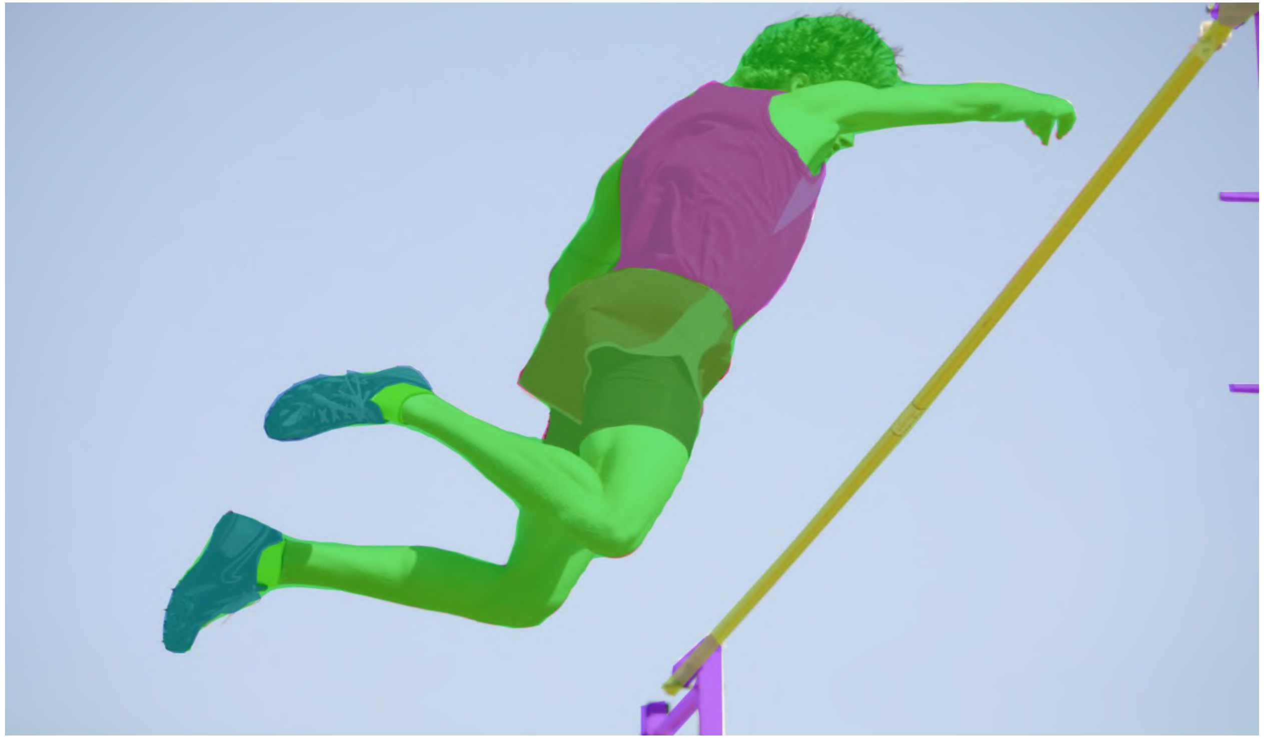


English (WN):
vaulter; pole vaulter; pole jumper

Bulgarian (WN):
скачач на овчарски скок;
състезател на овчарски скок

Polish: (WN)
tyczkarz

Multilingual Description of Ontology Classes



Definition:
‘an athlete who jumps over a high crossbar with the aid of a long pole’

English (WN):

vaulter; pole vaulter; pole jumper

Bulgarian (WN):

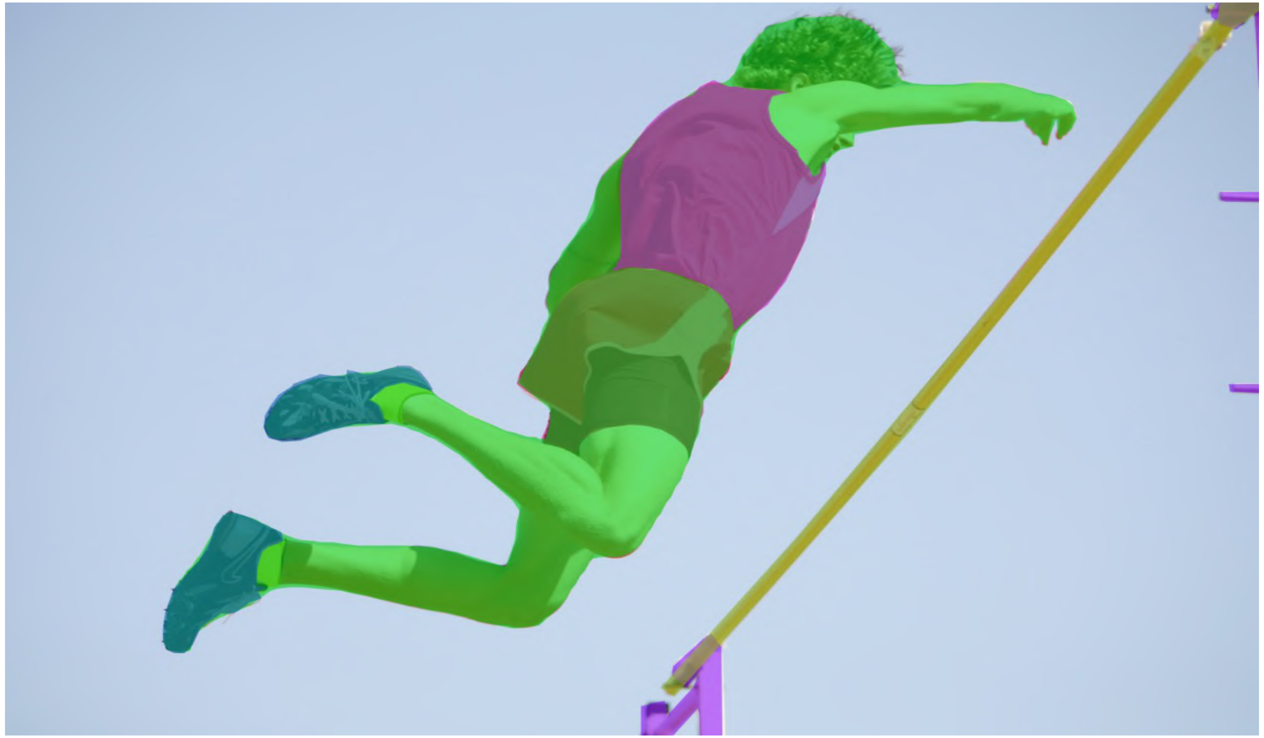
скачач на овчарски скок;

състезател на овчарски скок

Polish: (WN)

tyczkarz

Multilingual Description of Ontology Classes



English (WN):

vaulter; pole vaulter; pole jumper

Definition:
'an athlete who jumps over a high crossbar with the aid of a long pole'

Bulgarian (WN):

скачач на овчарски скок;

състезател на овчарски скок

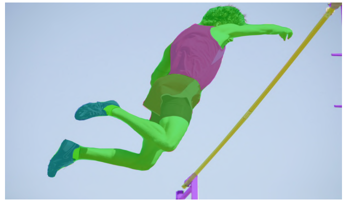
Definition:
'спортист, който се състезава в лекоатлетическата дисциплина овчарски скок'

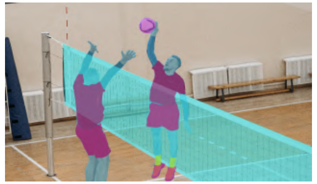
Polish: (WN)

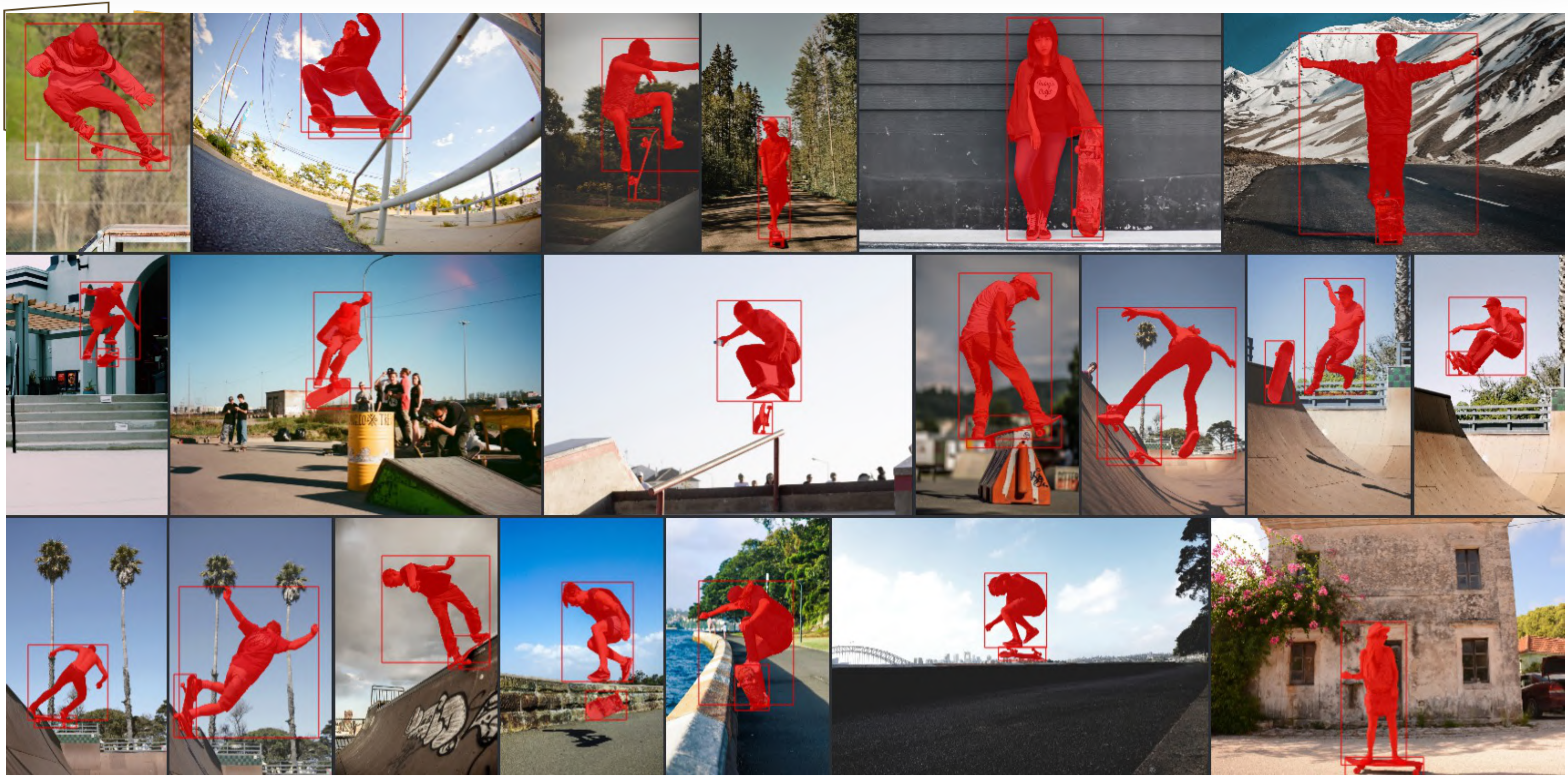
tyczkarz

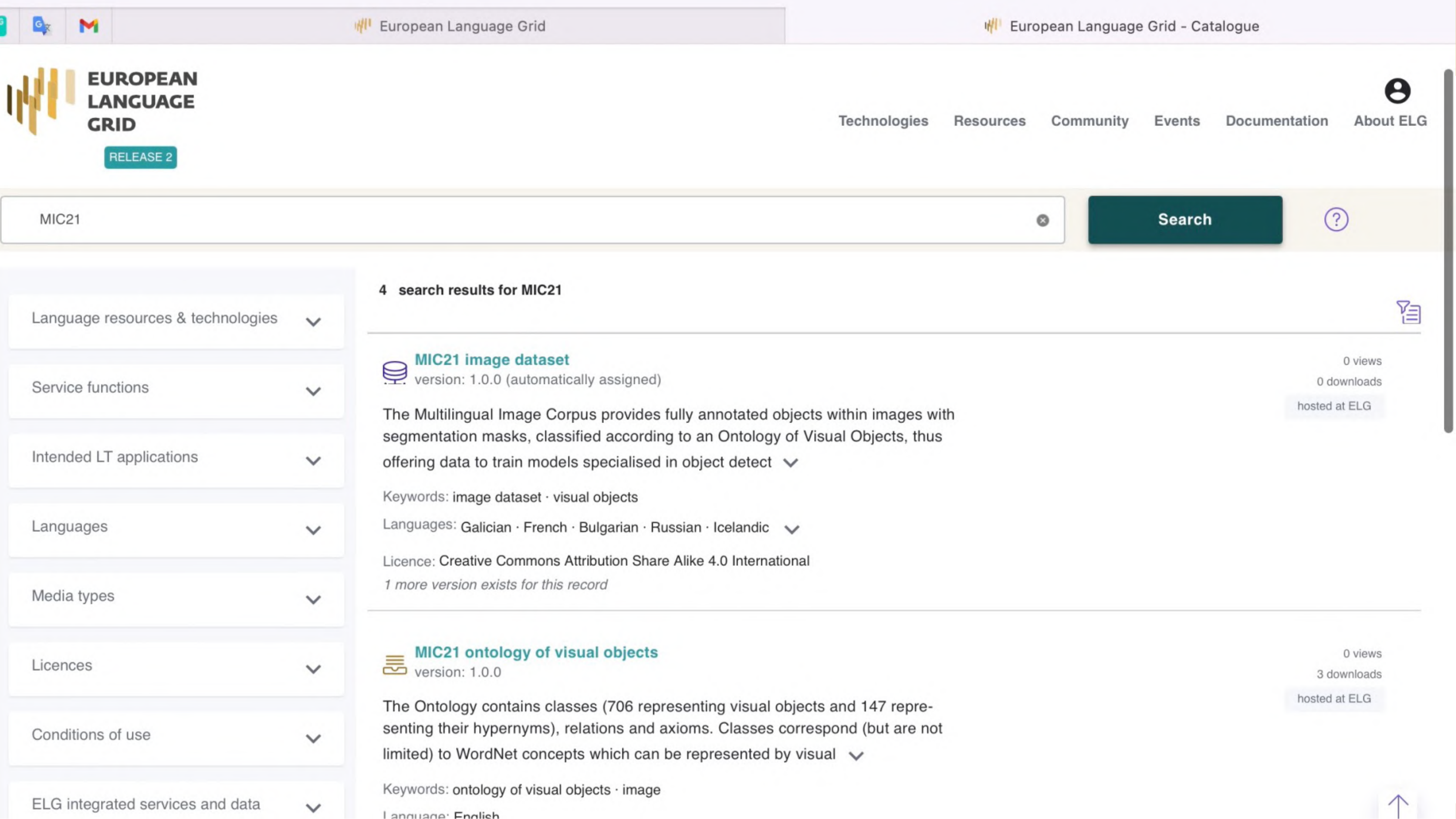
The Multilingual Image Corpus in numbers

Domain	Images	Annotations
Sport	6 915	65 482
Transport	7 710	78 172
Arts	3 854	24 217
Security	2 837	35 916
MIC21 in total	21 316	203 797









4 search results for MIC21

MIC21 image dataset
version: 1.0.0 (automatically assigned)

0 views
0 downloads
hosted at ELG

The Multilingual Image Corpus provides fully annotated objects within images with segmentation masks, classified according to an Ontology of Visual Objects, thus offering data to train models specialised in object detect

Keywords: image dataset · visual objects

Languages: Galician · French · Bulgarian · Russian · Icelandic

Licence: Creative Commons Attribution Share Alike 4.0 International

1 more version exists for this record

MIC21 ontology of visual objects
version: 1.0.0

0 views
3 downloads
hosted at ELG

The Ontology contains classes (706 representing visual objects and 147 representing their hypernyms), relations and axioms. Classes correspond (but are not limited) to WordNet concepts which can be represented by visual

Keywords: ontology of visual objects · image

Language: English