CENTER FOR BIOSYSTEMS AND BIOTECH DATA SCIENCE

Esla Timothy Anzaku, Arnout Van Messem, Wesley De Neve

Leveraging Human-Machine Interactions for Computer Vision Dataset Quality Enhancement

The importance of dataset quality and the ImageNet dataset

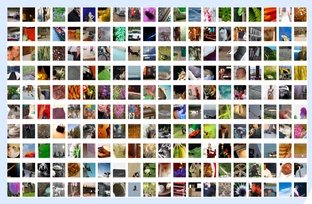
- Deep learning (DL) models learn from data and can learn data biases too
- The ImageNet dataset [1] is pivotal for progress in DL research
  - The popular ImageNet dataset comprises a million-plus images in 1,000 categories
  - Each image is assigned to only a single defined category
- Example applications for the ImageNet dataset
  - Benchmarking DL progress in supervised computer vision and self-supervised learning
  - Transfer learning and fine-tuning
  - Feature extraction for downstream tasks, such as object detection and segmentation

Research goal

- Design and implement a framework to effectively utilize 15 annotators and good pre-trained DL models to enhance the label quality for ImageNet-V2 [2]
- ImageNet-V2 is a more recent test dataset that is created using similar dataset creation protocols as ImageNet
  - Consists of 10,000 images for 1,000 categories
  - Useful for assessing DL progress on image recognition tasks

ImageNet validation set (50,000 images)

ImageNet V2 (10,000 images)



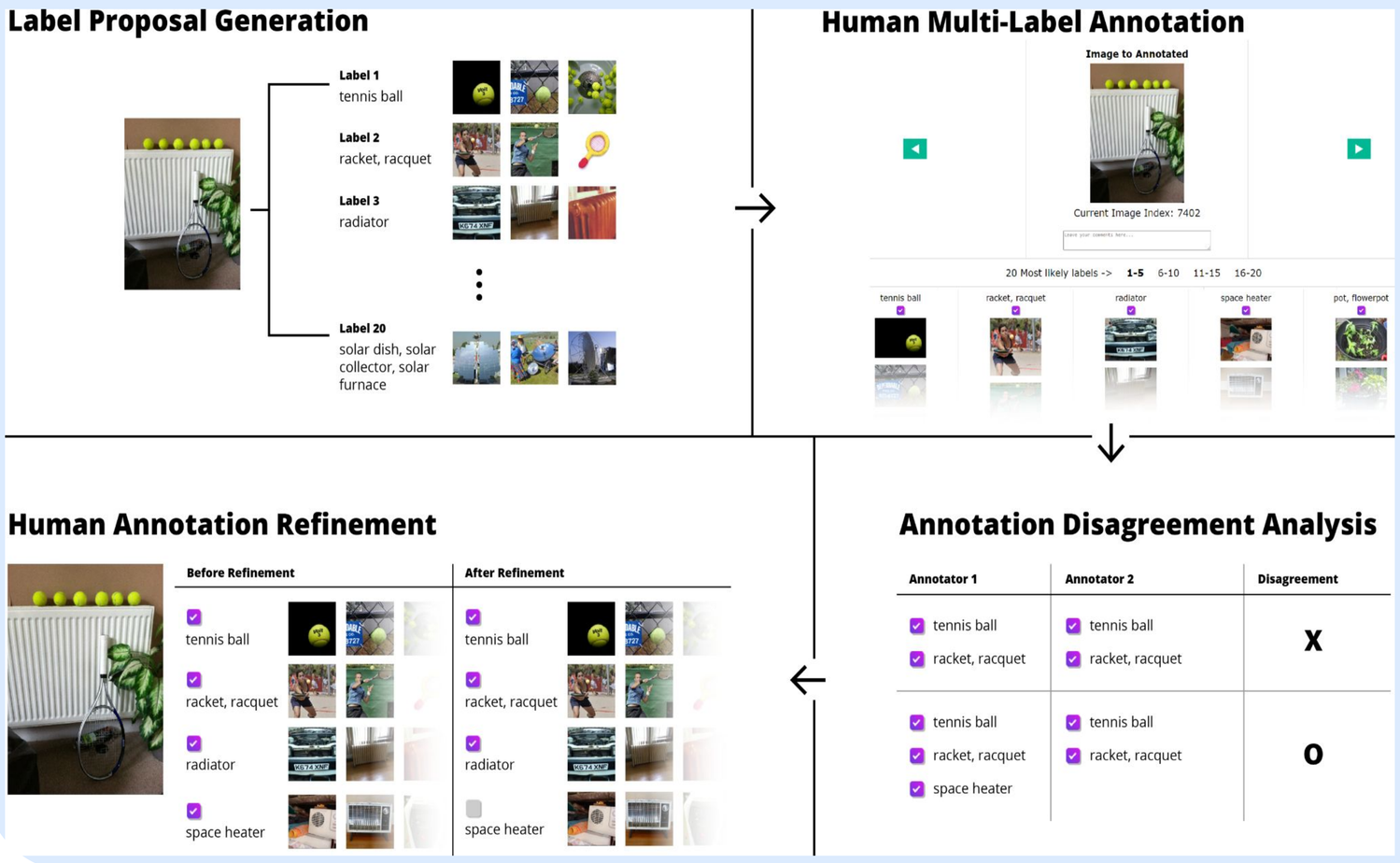
Why enhance dataset labels and leverage machine interaction (pretrained models)?

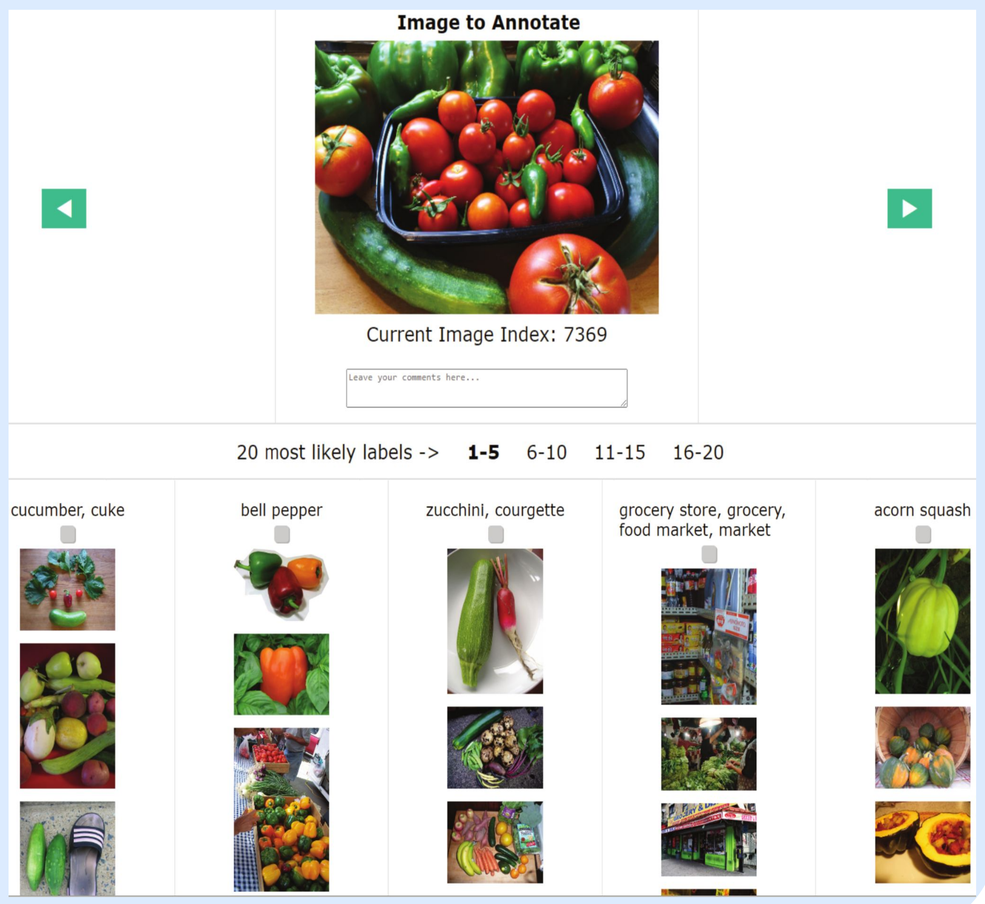
- Data labeling process is tedious and error-prone
  - Automation using pre-trained models can substantially reduce workload and minimize labelling errors due to human oversight or bias
- An image often contains multiple objects of interest
  - Assuming only one label per image oversimplifies the complexity, especially when DL models have substantial learning capacities
- DL can tolerate some noise in the dataset and still create useful models
  - DL models can generate useful insights, allowing us to use their output to propose enhancements to dataset labels
- Create a feedback loop for continuous improvements
  - A system that enables humans to refine machine suggestions can create a productive feedback loop that continually enhances the model's performance and the dataset's quality

Methodology (framework and web interface)

1. Pretrained models generate labels
2. Humans label images
3. Labels auto-analyzed by annotation disagreement methods to select images for the final stage
4. Humans refine conflicting labels

Lightweight, user-friendly, and intuitive web interface to reduce the labelers' fatigue and labeling errors

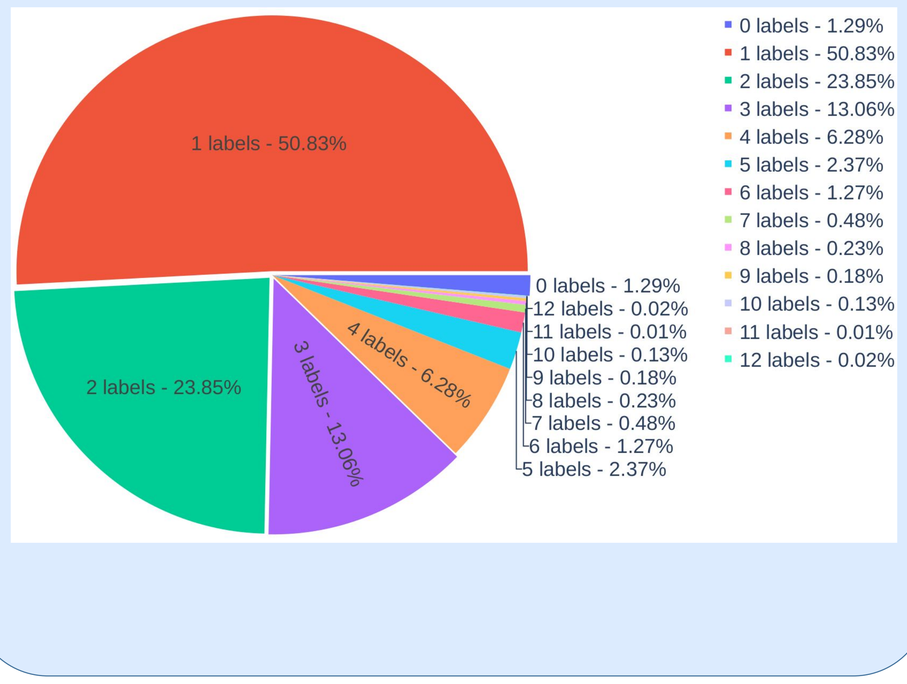




Results

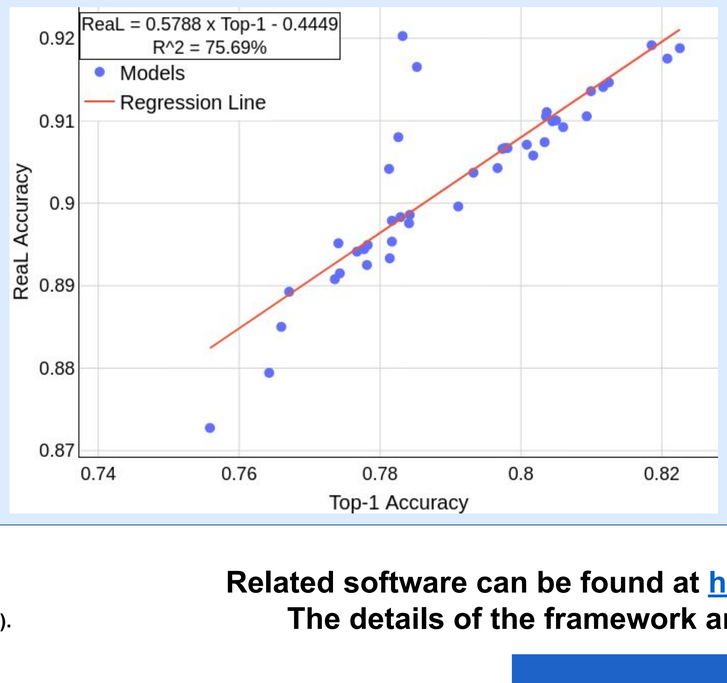
Proportion of multi-label images

- Close to 50% of the images have more than one label out of the 1,000 categories for the ImageNet-V2 dataset



Accuracy implications

- The performance of evaluated pre-trained models are underrated under Top-1 accuracy
  - 57 pre-trained models were evaluated under Top-1 and ReaL accuracy
    - Top-1: Correctness of topmost prediction
    - ReaL: Correctness when ground truth label belongs to the topmost 5 predictions
  - Higher ReaL accuracy indicates that models could perform better if we acknowledge alternative valid labels for the images



Conclusions

- DL models excel in performance but struggle with reliability due to sensitivity to even minor data variations
- As model-centric advancements progress, it is essential to also focus on data-centric improvements, particularly dataset quality enhancement, to ensure robust DL model creation and evaluation
- DL models trained on ImageNet exhibit substantial and unexpected reductions in effectiveness on ImageNet-V2. Our enhanced labels can facilitate further investigation into this issue
- Our lightweight, open-source framework reduces labeling effort and enables researchers to easily enhance dataset labels. This contributes toward data-centric approaches to improving DL robustness and reliability

[1] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and F. Li, ImageNet: A Large-Scale Hierarchical Image Dataset (2009).
[2] B. Recht, R. Roelofs, L. Schmidt, V. Shankar, Do ImageNet Classifiers Generalize to ImageNet? (2019).
[3] E.T. Anzaku et al., Leveraging Human-Machine Interactions for Computer Vision Dataset Quality Enhancement (2023).

Related software can be found at https://github.com/esla/Multilabelfy
The details of the framework and results can be found in [3].

GHENT UNIVERSITY GLOBAL CAMPUS

Center for Biosystems & Biotech Data Science

LIÈGE université Mathématique

FACULTY OF ENGINEERING AND ARCHITECTURE

For more inquiries and discussion, please contact me at
eslatimothy.anzaku@ghent.ac.kr