# Labeling tools are great, but what about quality checks?

## A UX and AI approach to quality

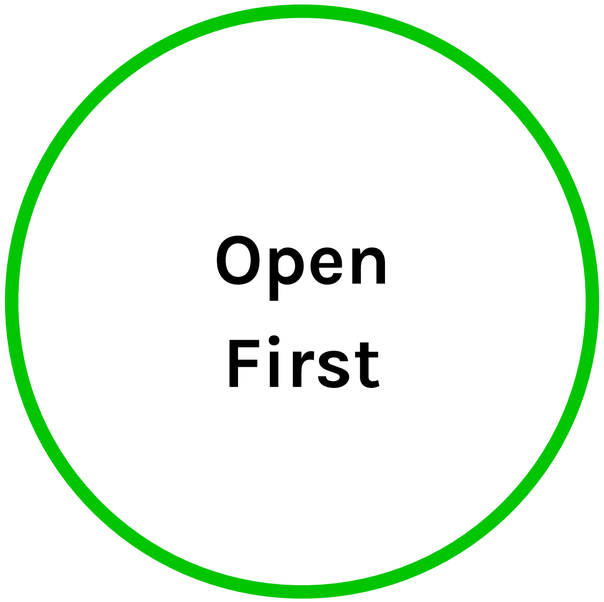
**Jakub Piotr Cłapa**

Senior AI Researcher @ Collabora Ltd.

# Open Source, tailored for you

We provide Guidance, Training, Development,
Integration, Optimization, Maintenance for:

- Artificial Intelligence: PyTorch, ONNX, TVM, mlpack, MLflow
- Linux kernel (core kernel, DRM/KMS, drivers/modules, ...)
- Bootloaders, secure and verified boot
- Multimedia: Streaming (webRTC, SRT, RTSP), GStreamer, PipeWire
- Graphics: GPU, OpenGL, Vulkan, Wayland, Weston
- Augmented Reality: OpenXR, Monado, XRdesktop
- Linux distributions: PetaLinux, Yocto, Debian, APERTIS



# The QA problem

- Labeling is a difficult cognitive task that requires a Quality Assurance (QA) process.
- Not all errors average out and can be ignored – systematic biases transfer to the model.

> In many industries where giant data sets simply don't exist, I think the focus has to shift from big data to good data.
>
> – Andrew Ng "UNBIGGEN AI"

# The QA tooling problem

- Most tools have only minimal support for review.
- Frequently the QA process is more difficult (and expensive!) then labeling.

> Annotations are reviewed four times in order to confirm accuracy. Two annotators label a given object, a supervisor then checks the quality of their work.
>
> – keymakr, a leading annotation provider


Image from the Mapillary Traffic Sign Dataset
by Mapillary Research
https://www.mapillary.com/dataset/trafficsign

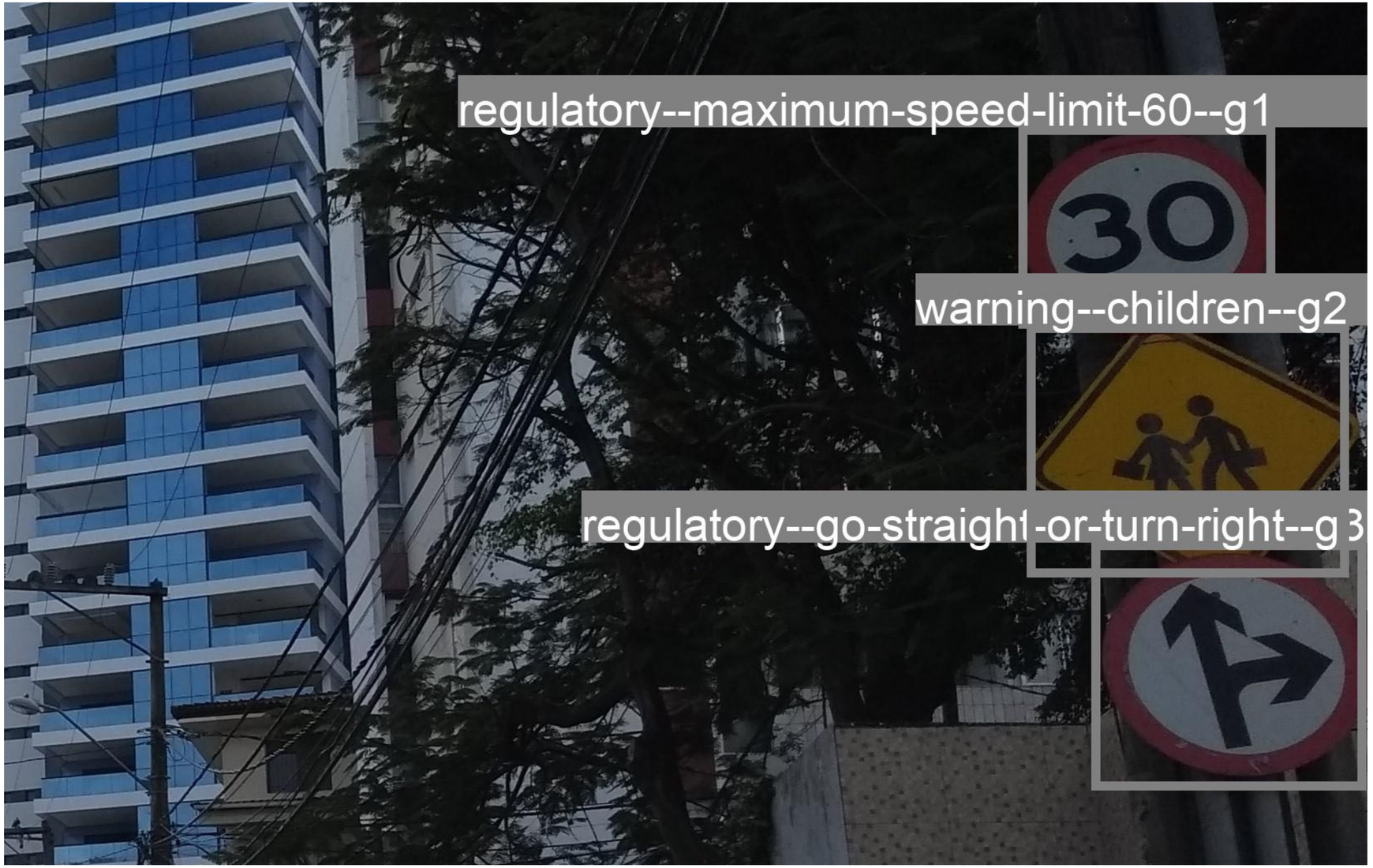
regulatory--maximum-speed-limit-60--g1


warning--children--g2

regulatory--go-straight-or-turn-right--g3

Image from the Mapillary Traffic Sign Dataset
by Mapillary Research
https://www.mapillary.com/dataset/trafficsign

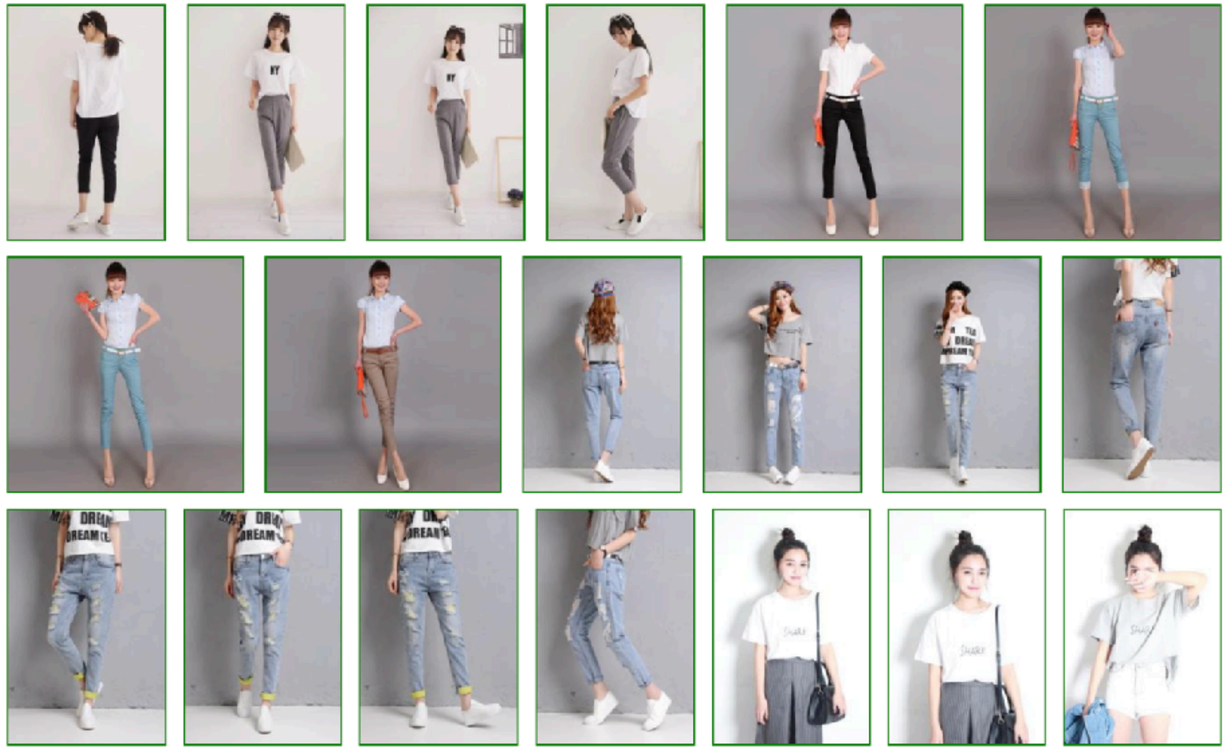
# Can we do better?


regulatory--maximum-speed-limit-60--g1
30
warning--children--g2
regulatory--go-straight-or-turn-right--g3


60
60
60
60
30
60

# Key insight: Show similar images together!

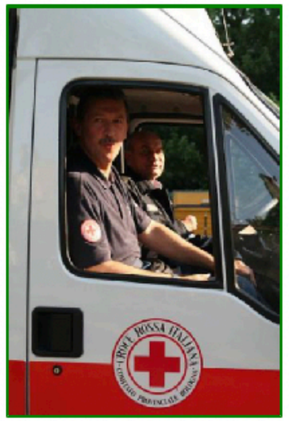
- Reduces the cognitive load in QA.
- Highlights the real-world variety of the samples.
- Even with more heterogeneous classes or unlabeled data we can sort the images automatically using AI!



Images from the DeepFashion2 dataset
by Yuying Ge and Ruimao Zhang and Lingyun Wu and Xiaogang Wang and Xiaoou Tang and Ping Luo
https://github.com/switchablenorms/DeepFashion2

# Key insight: Show similar images together!







Images from the ImageNet1k validation set

# Key insight: Show similar images together!








Images from the ImageNet1k validation set

# Case Study: Traffic signs

- We looked at the Mapillary Traffic Sign dataset and found quite a few errors.
- We estimated the error rate in a few easy classes to be at least 1%.
- This is in a high-quality dataset from a subsidiary of Meta.

# Case Study: Traffic signs

# Case Study: Traffic signs

# Case Study: Traffic signs

- We took all the ground truth crops from the Mapillary dataset.
- We trained a simple ResNet50 classifier.
- We selected ≈700 “other-sign” labels that our model thought were actually “regulatory--no-entry--g1“ signs.
- We used MLfix and found 170 no-entry signs mislabeled as “other-sign”!

# Case Study: Image search engines

- We created a new public dataset of men and women in t-shirts of various colors.
- 20 classes, 20k images total
- ≈80% error rate

# Why image search?

- Image search engines are frequently used to gather vast amounts of images to train language/vision models (the DALL-E dataset for example).

- The quality of image search results is not amazing.

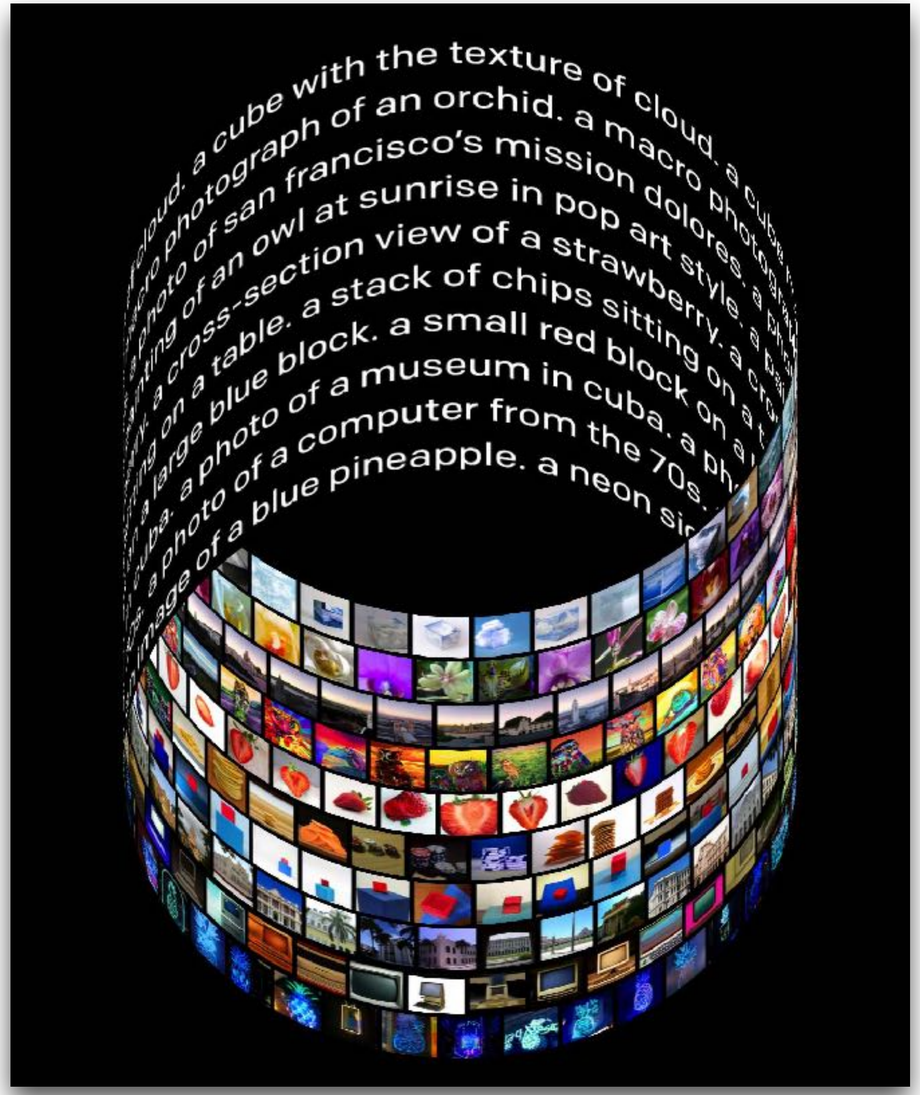



Image from the DALL·E: Creating Images from Text blog post
by OpenAI
https://openai.com/blog/dall-e/

# Demo

# How does it work?

1. We start by pretraining a ResNet18 model with Barlow Twins.

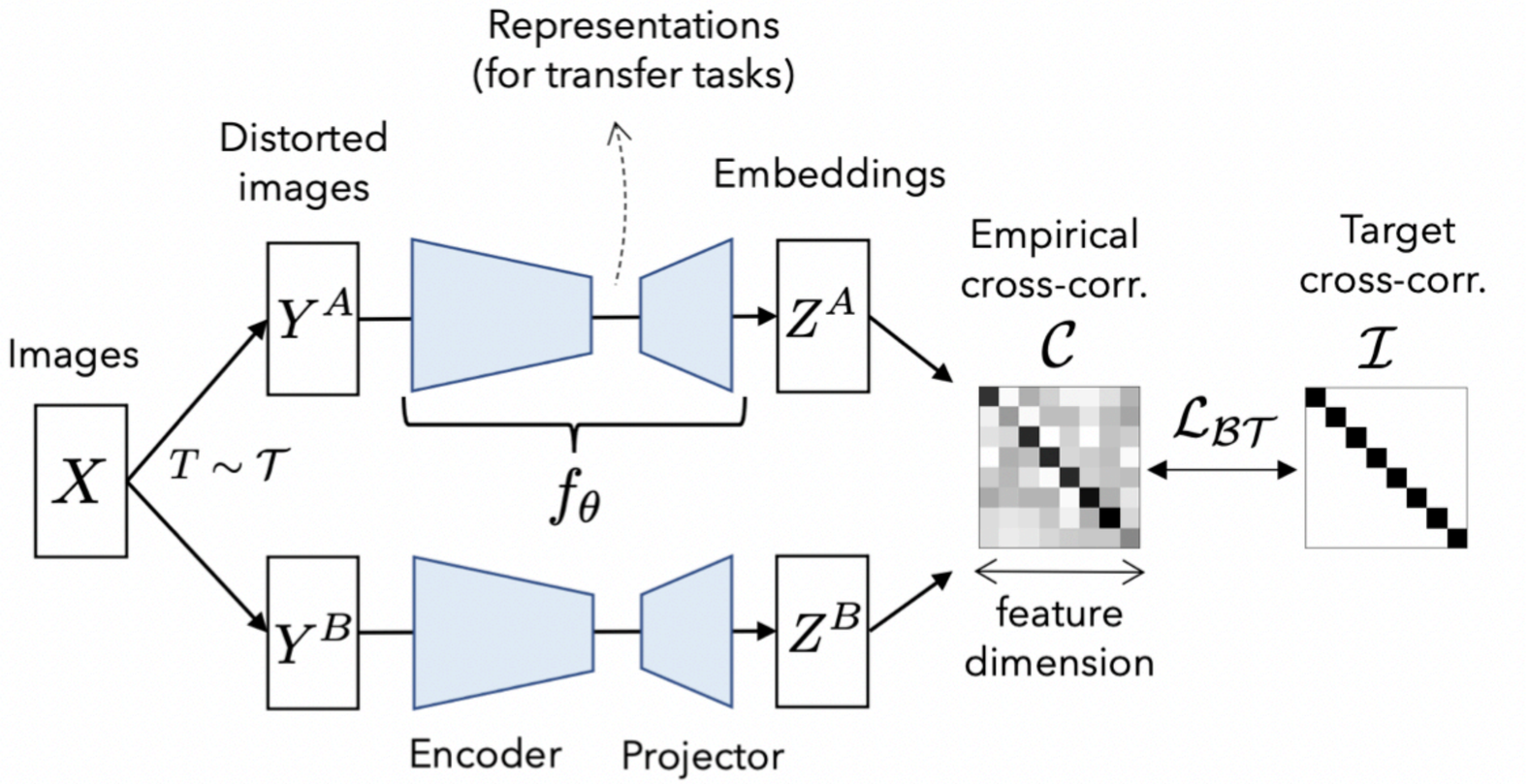



Image from Barlow Twins: Self-Supervised Learning via Redundancy Reduction by Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, Stéphane Deny
https://arxiv.org/abs/2103.03230

# How does it work?

2. We take all the feature vectors from the second to last layer (a 14x14 grid with 256 floating point numbers in each cell).

3. We cluster them into 1024 categories (“visual words”) using KMeans.


Image from Recognizing and Learning Object Categories
by Li Fei-Fei (Stanford), Rob Fergus (NYU), Antonio Torralba (MIT)
https://people.csail.mit.edu/torralba/shortCourseRLOC/

# How does it work?

4. We create a 1024-bit vector for each image with ones for each visual words that occurred in this image (inspired by the Bag of Visual Words method).

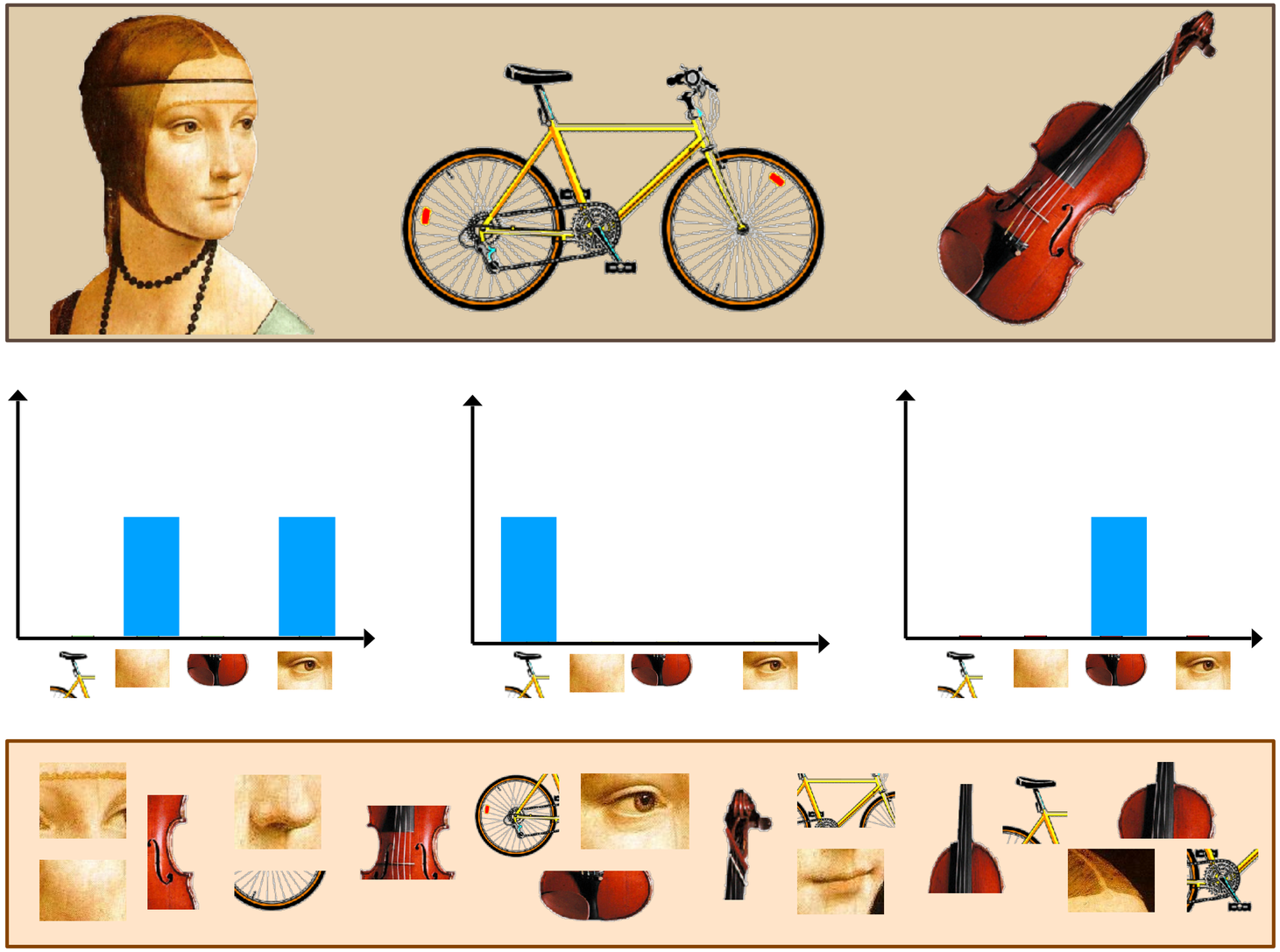


Image from Recognizing and Learning Object Categories
by Li Fei-Fei (Stanford), Rob Fergus (NYU), Antonio Torralba (MIT)
https://people.csail.mit.edu/torralba/shortCourseRLOC/

# How does it work?

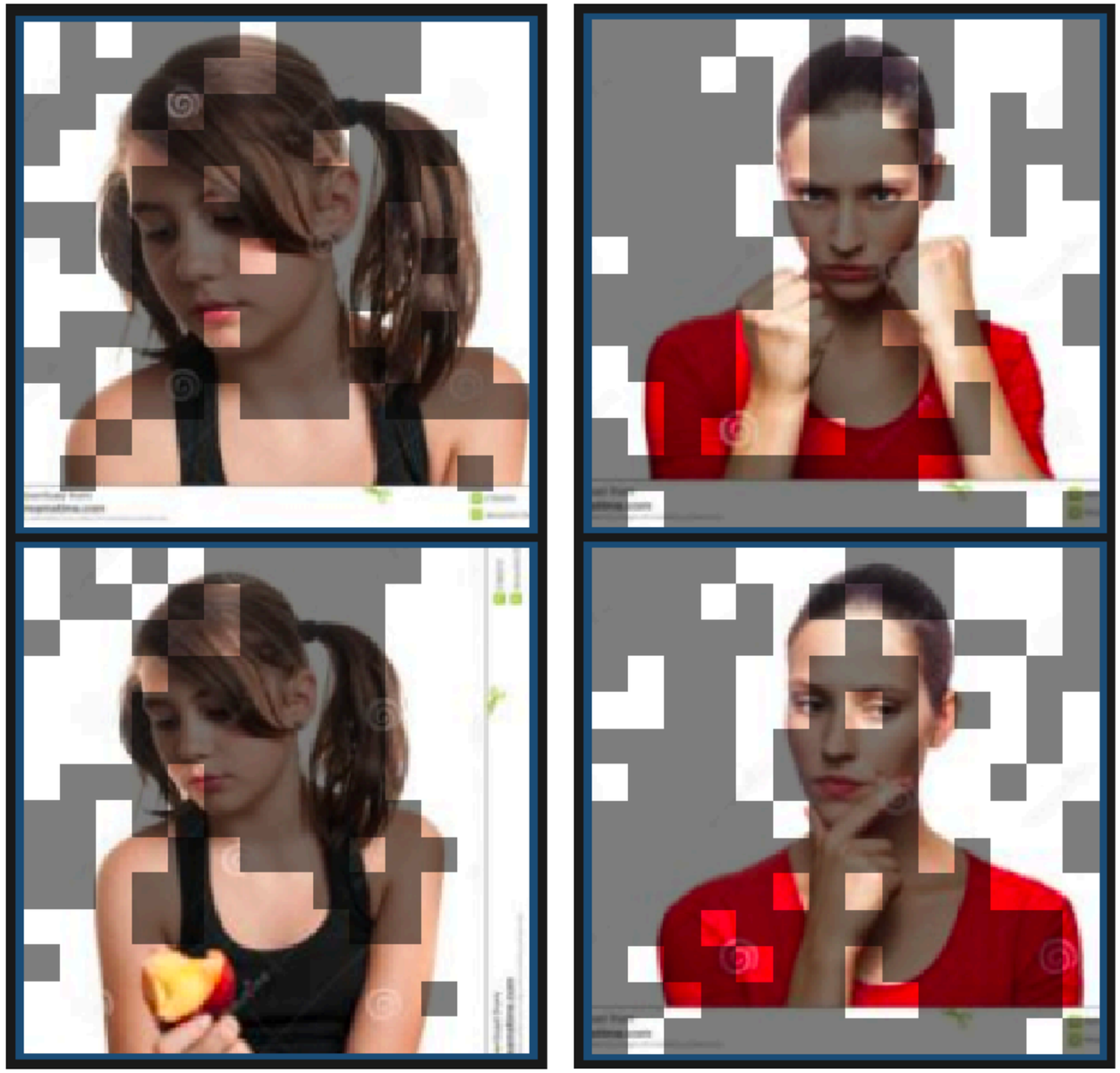
5. We start with a random image and then find the next one with highest count of common visual words:

   (x * y).sum()

6. Repeat until we run out of images.

# Limitations

- The model has no way to know what the thing we care about is.
- It does not even know what is the foreground object in the scene.



# What’s next?

- Make it easier to slice and dice the data by connecting the UI and your Python process.
- Should be very easy to run from a Jupyter Notebook.
- Add a quick way to fetch results back into Python for further analysis.

# What’s next?

- Add support for sorting techniques invented by other people:
  - CleanLab,
  - classification loss,
  - supervised model confusion matrices.

# What's next?

- Deep learning based clustering methods to replace KMeans (VQ-VAE looks interesting)

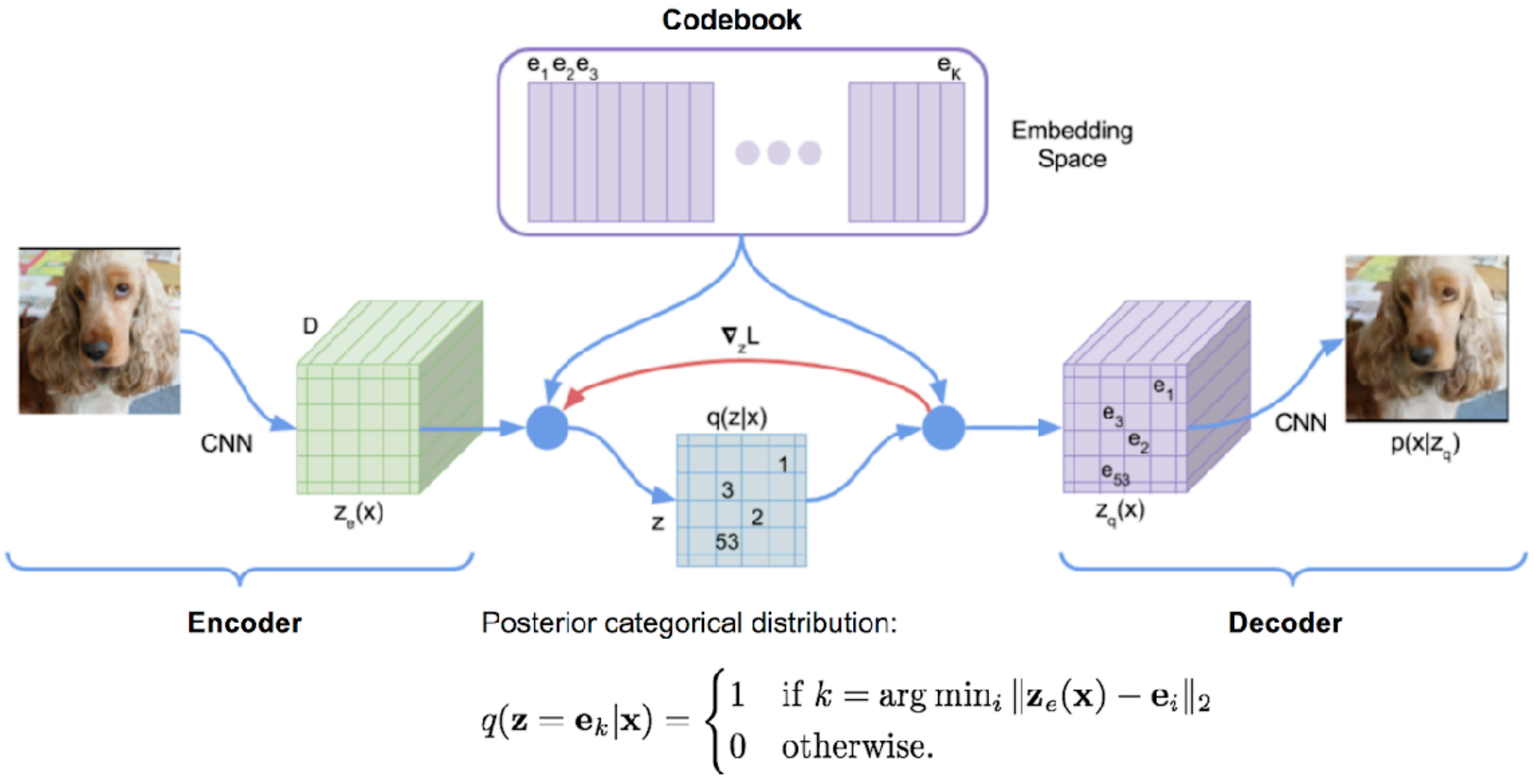


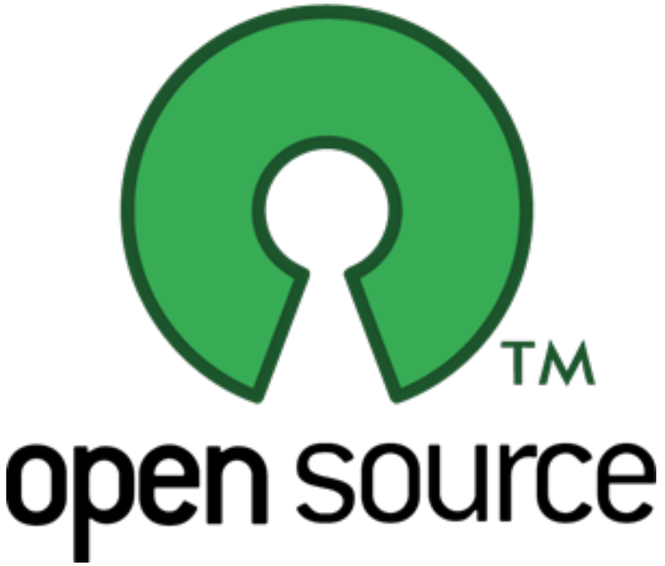

Image from Neural Discrete Representation Learning
by Aaron van den Oord, Oriol Vinyals, Koray Kavukcuoglu
https://arxiv.org/abs/1711.00937

# Summary

- The QA task is different and requires a new user-interface.
- We should improve both UIs and AIs.
- (Un)supervised deep learning can help augment the human intellect.

# Where to get it

- We released all our research under an Open Source license: https://github.com/collabora/MLfix
- We worked hard so you can use it on your datasets today and we are thrilled to help you do this: https://gitter.im/MLfix/community



# We are hiring

## col.la/careers