Improving filling level classification with adversarial training.

Apostolos Modas, Alessio Xompero, Ricardo Sanchez-Matilla, Pascal Frossard, Andrea Cavallaro

Deep Learning in the real world

Deep learning has achieved impressive results in huge benchmarks like ImageNet. Part of its success: access to millions of training data!

Reality is different: in many real world tasks the access to **training data might be very limited!**

A real world example

Human-robot collaboration in daily tasks. In such scenario the systems should be able to infer the "world" from just a few observations.

A use-case: manipulation and handovers of containers such as cups and drinking glasses.

Infer the weight of the container:

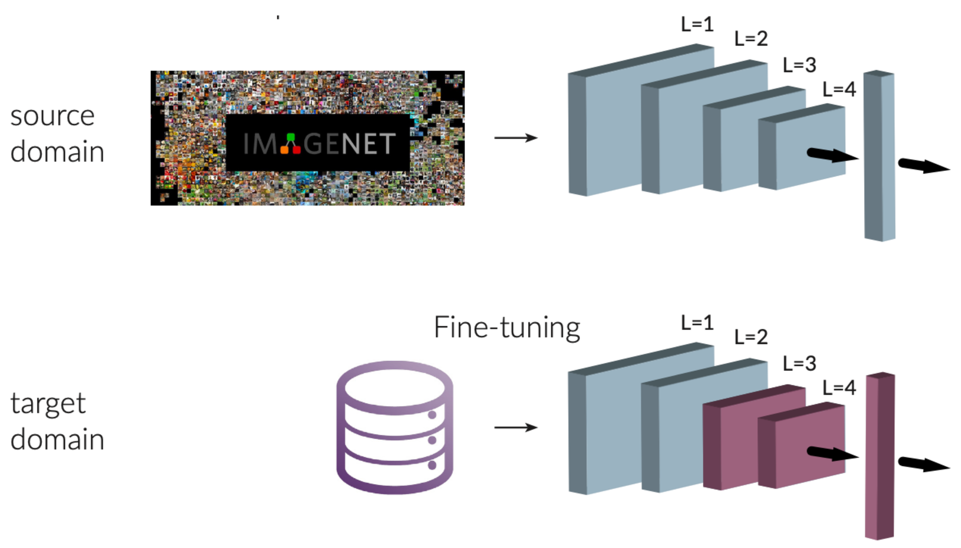
- Volume of container
- Filling level (amount of content)

Filling level estimation

This ostensibly simple scenario can be quite challenging!

- Training data are scarce
- Can be constrained: RGB still images
- Large variability: shape, material, content, transaprency, occlusions

Transfer learning



Prior work has used **transfer learning** to mitigate the limited training data issue.

Approach the problem as a classification task.

However: the results are a bit better than **random chance**!

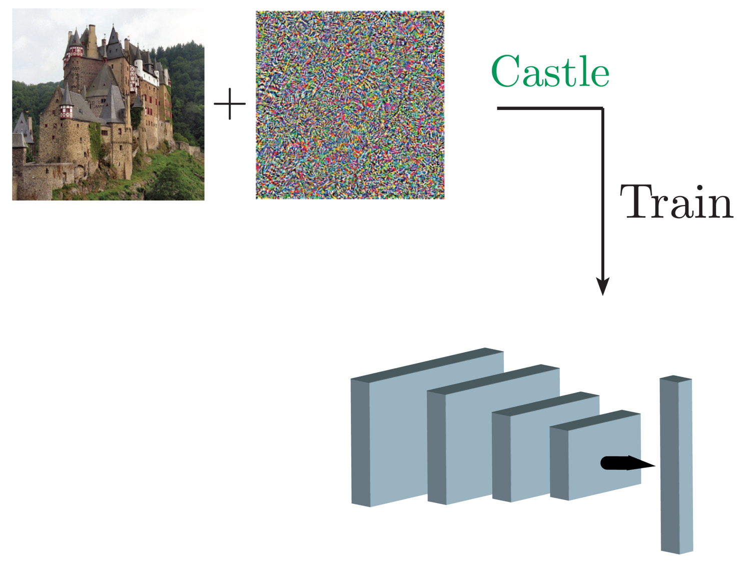
How to improve filling level classification using transfer learning?

With adversarial training!

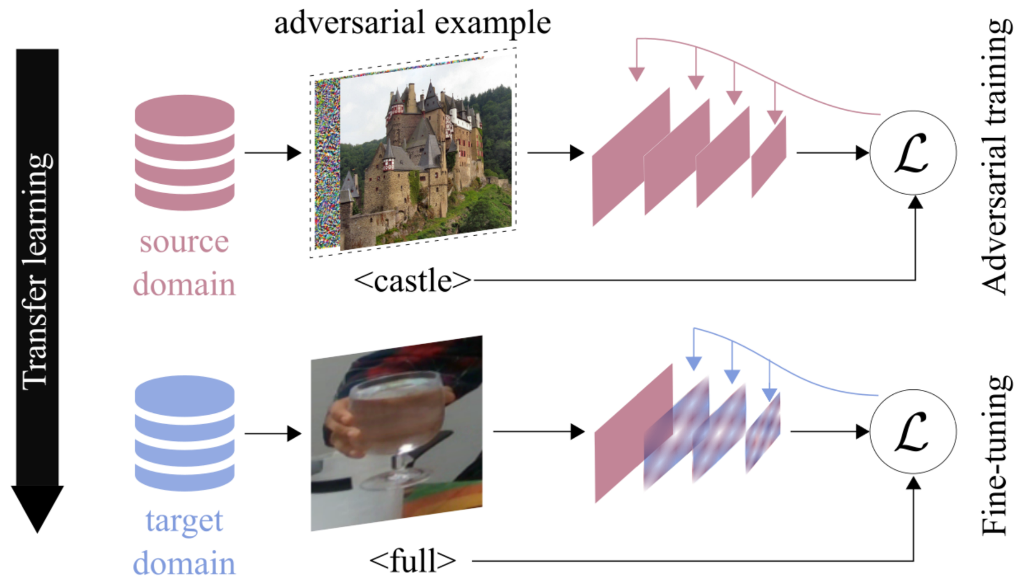
Adversarial training (AT)

During training replace the data with their **adversarial examples**.

Adversarially trained networks transfer better: **improve fine-tuning performance!**



Improving filling level classification with AT



AT on the source domain (ImageNet) and fine-tune for filling level classification.

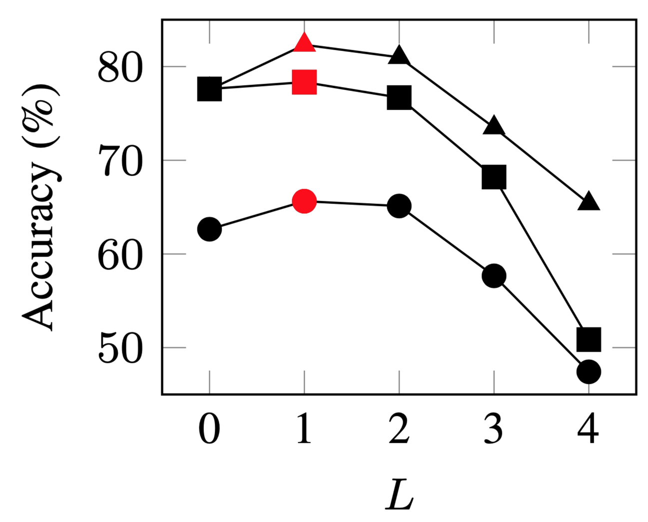
We further explore AT on the target domain.

The dataset: C-CCM

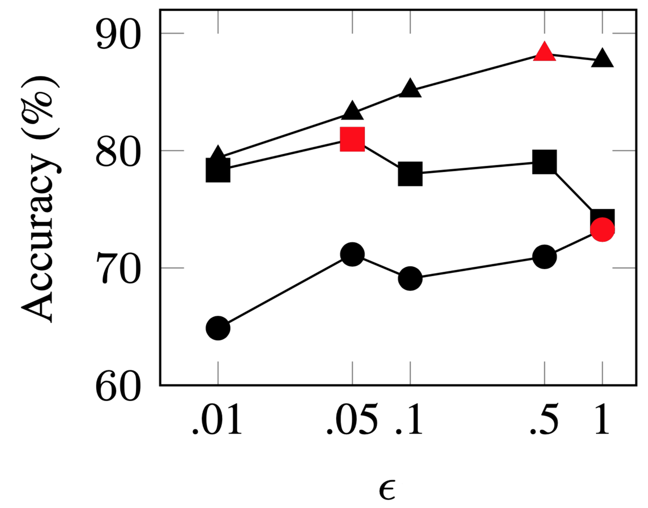
Image **C**rops from the **C**ORSMAL **C**ontainers **M**anipulation Dataset

- **8 objects:** 4 cups + 4 drinking glasses
- 10,216 **RGB images**
- Filling level: **0%, 50%, 90%, "unknown"**
- Filling type: **water, pasta, rice**

Sensitivity analysis

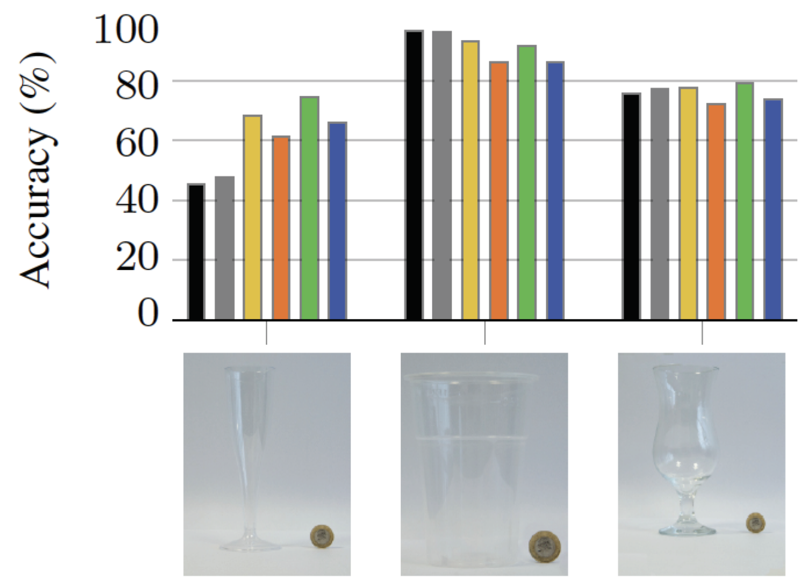


Fixing the 1st layer results in the highest test accuracy.

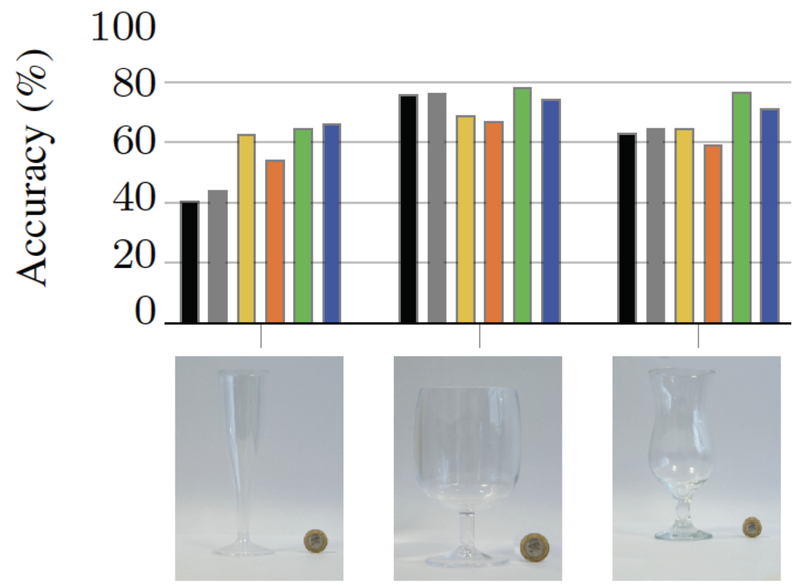


The optimal value of ε depends on the dataset.

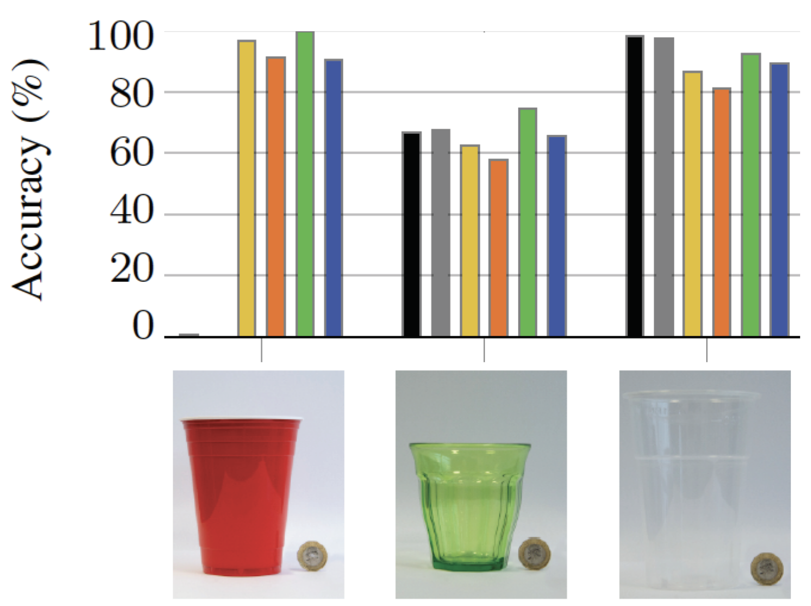
Performance evaluation



- **Champagne flute**: hard to cope with the narrow shape above stem
- **Beer & Cocktail**: shape (also above stem) that appears in the train set
- Robust fine-tuning: improves the generalization performance



- **Wine glass**: quite regular shape above stem
- **Cocktail**: the absence of stem in the train set causes performance drop
- Robust fine-tuning: improves the generalization performance



- **Red cup**: always predicts 99% with some opaque content (pasta, rice)
- **Green glass**: color + reflection that can harden the problem
- Robust fine-tuning: improves the generalization performance