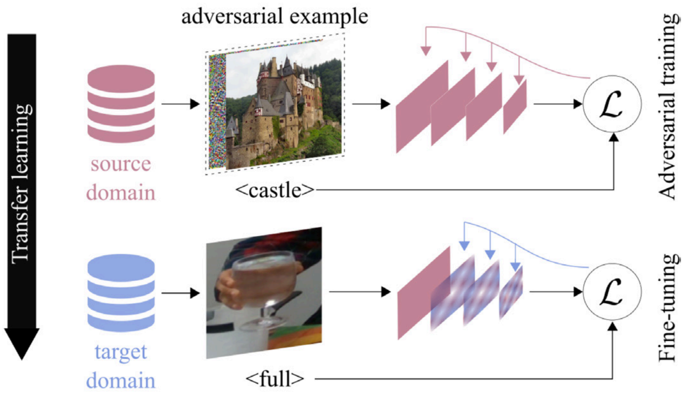
Improving filling level classification with adversarial training.

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



Deep Learning

SOTA in many different tasks

Requirement: "tons" of training data

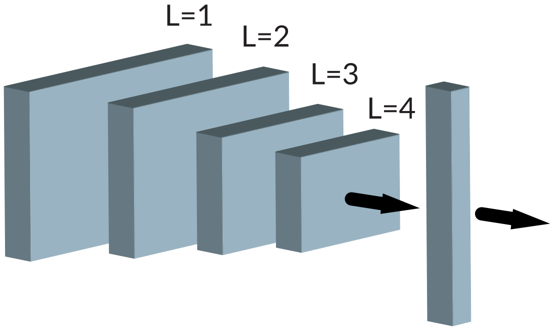
Reality: not always the case!

- Access to limited amount of training data

Transfer Learning

Alleviate the “few data” problem

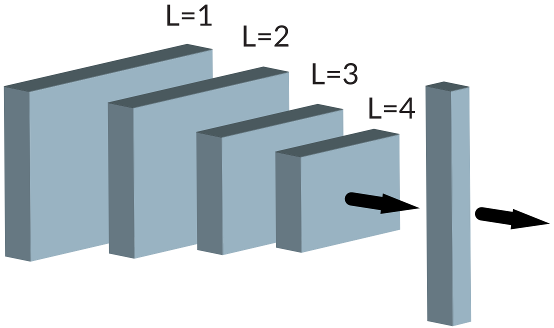
source
domain

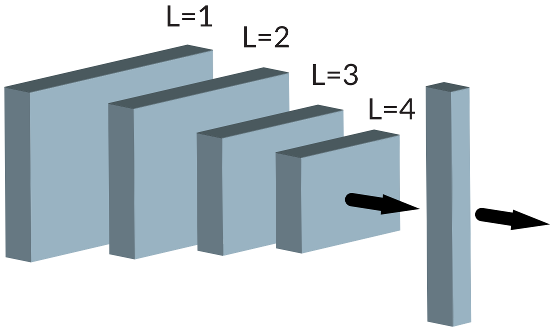


Transfer Learning

Alleviate the “few data” problem

source
domain

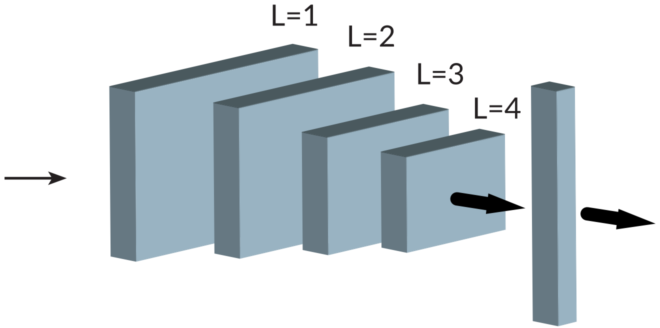




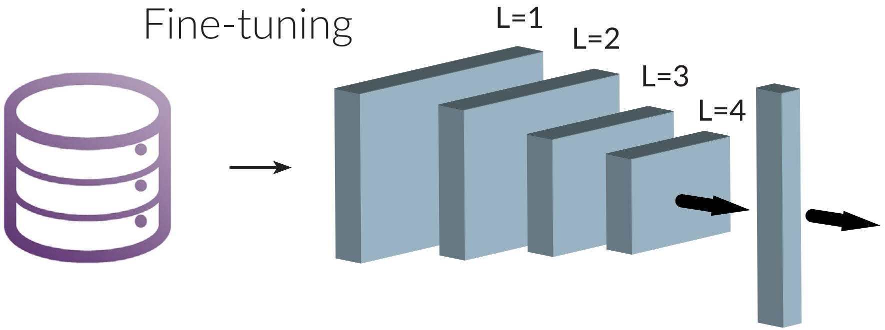
Transfer Learning

Alleviate the “few data” problem

source
domain



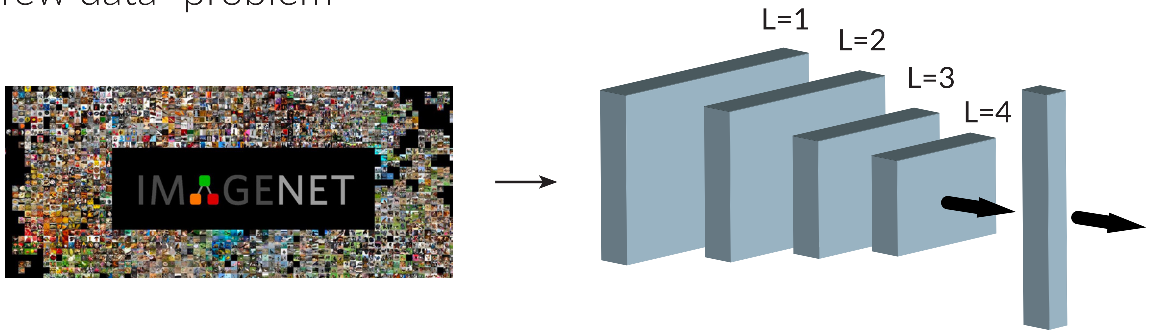
target
domain



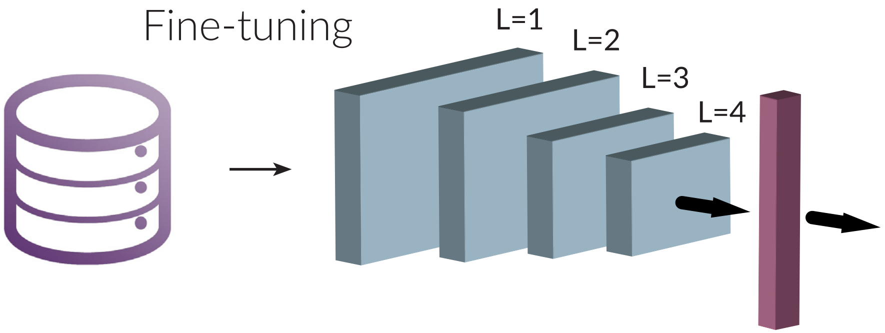
Transfer Learning

Alleviate the “few data” problem

source
domain



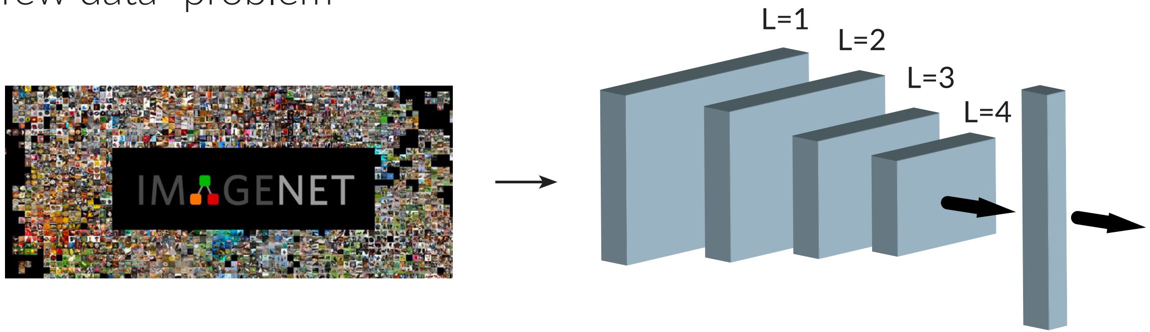
target
domain



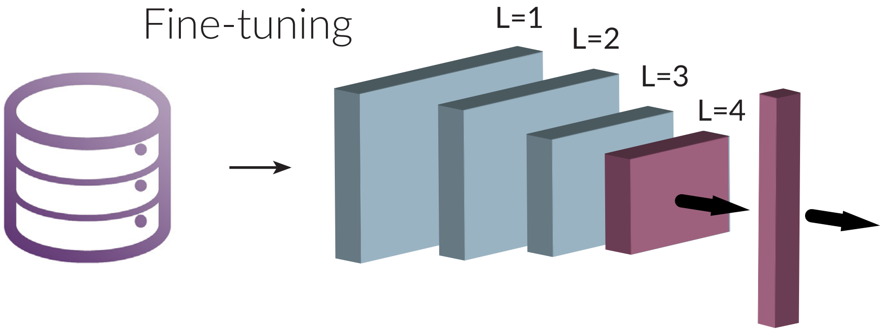
Transfer Learning

Alleviate the “few data” problem

source
domain



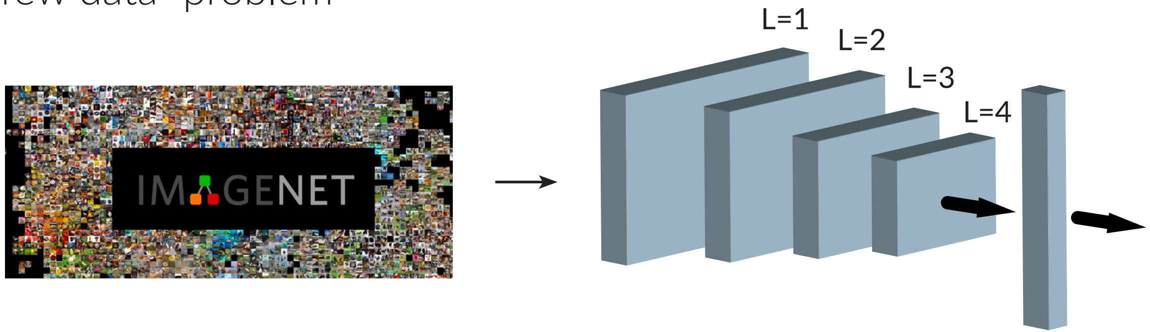
target
domain



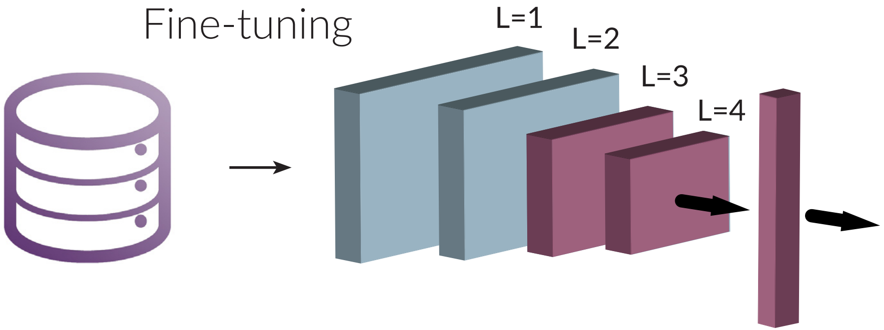
Transfer Learning

Alleviate the “few data” problem

source
domain



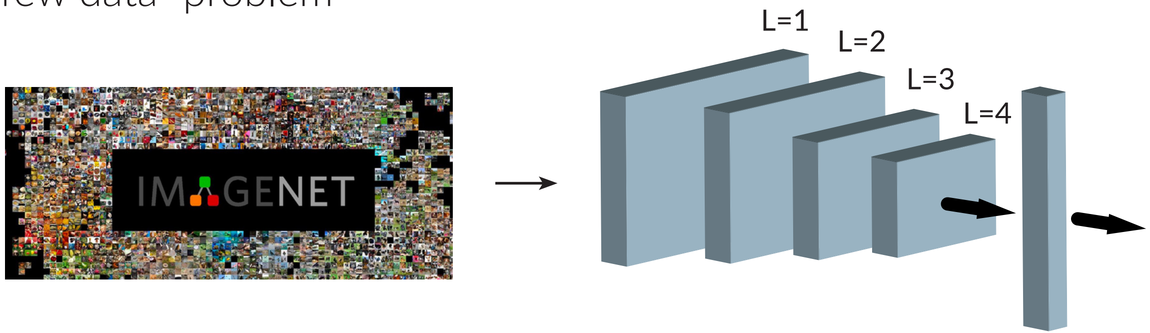
target
domain



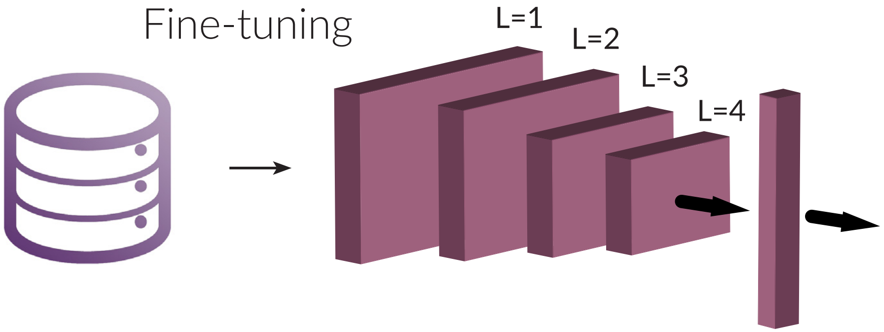
Transfer Learning

Alleviate the “few data” problem

source
domain



target
domain



A real world example

Human-robot collaboration on daily tasks

- Infer the “world” from a **few** observations

A real world example

Human-robot collaboration on daily tasks

- Infer the "world" from a **few** observations

Use-case: Manipulation and handovers of objects

- E.g., containers, drinking cups/glasses

CORSMAL: Collaborative Object Recognition, Shared Manipulation and Learning

A real world example

Human-robot collaboration on daily tasks

- Infer the “world” from a **few** observations

Use-case: Manipulation and handovers of objects

- E.g., containers, drinking cups/glasses

Important: estimate the container weight

- Infer dimensions/volume
- **Infer the amount of content within the container (filling level)**

CORSMAL: Collaborative Object Recognition, Shared Manipulation and Learning

Filling level estimation: challenges

This ostensibly simple scenario: **very challenging in fact!**

Filling level estimation: challenges

This ostensibly simple scenario: **very challenging in fact!**

- Constrained to vision madality: RGB data (no depth)

Filling level estimation: challenges

This ostensibly simple scenario: **very challenging in fact!**

- Constrained to vision madality: RGB data (no depth)

- Differences in shape


Filling level estimation: challenges

This ostensibly simple scenario: **very challenging in fact!**

- Constrained to vision madality: RGB data (no depth)

- Differences in shape

- Differencies in transparency

Filling level estimation: challenges

This ostensibly simple scenario: **very challenging in fact!**

- Constrained to vision madality: RGB data (no depth)

- Differences in shape
- Differencies in transparency
- Occlusions by the hand

Filling level estimation: challenges

This ostensibly simple scenario: **very challenging in fact!**

- Constrained to vision madality: RGB data (no depth)

- Differences in shape

- Differencies in transparency

- Occlusions by the hand

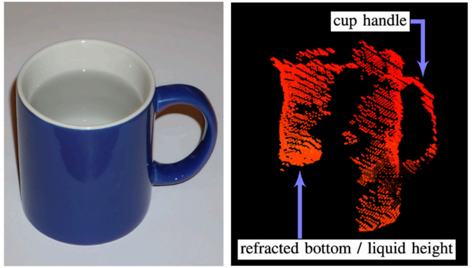
- More: material, type of content, illumination, background …

Filling level estimation: prior work

Observe the action of pouring content in the container

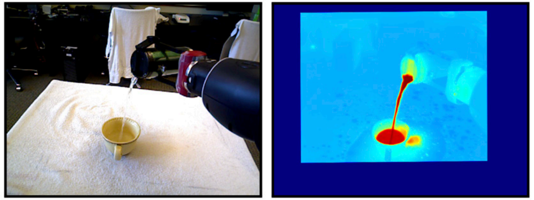
RGB-D

- Track the level during pouring [1],[2]



RGB-D + Thermal

- Identify pixels of “heated” liquid [3]



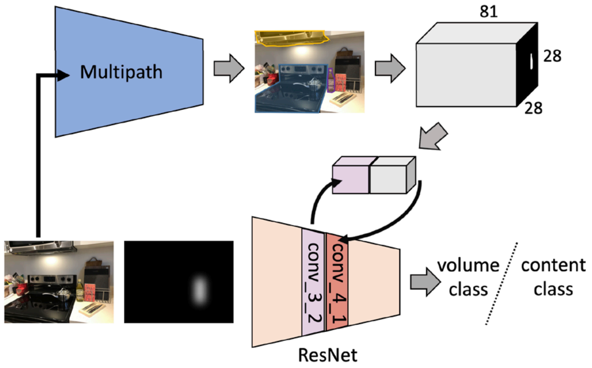
(a) RGB (b) Thermal

[1] C. Do et al. “A probabilistic approach to liquid level detection in cups using RGB-D camera”, IEEE IROS 2016
[2] C. Do et al. “Accurate pouring with an autonomous robot using an RGB-D camera”, AISC 2018
[3] C. Schenck et al. “Visual closed-loop control for pouring liquids”, IEEE ICRA 2017

Filling level estimation: prior work

Single RGB (still) images [1]

- Most challenging case (for vision)
- No depth - temporal - or material information
- Plus: the “few” data problem



Best solution (classification): **Transfer learning**

- ImageNet + fine-tuning

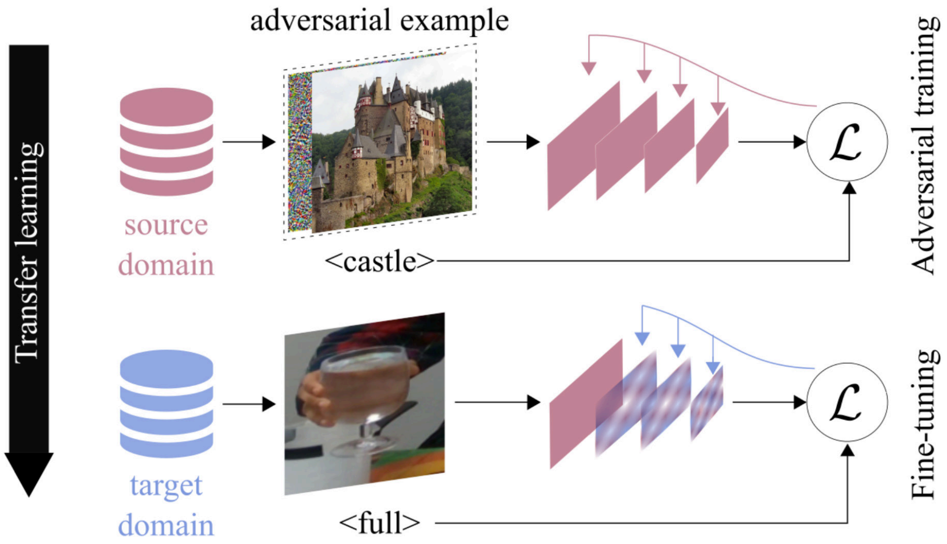
Yet… the performance is **marginally better than random chance!**

What if Transfer Learning could be improved?

[1] R. Mottaghi et al. “See the glass half-full: Reasoning about liquid containers, their volume and content”, IEEE ICCV 2017

Our work

Adversarial Training + Transfer Learning

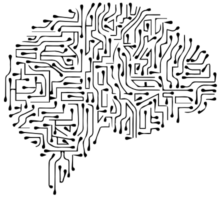


Preliminaries

Adversarial Examples

$\boldsymbol{x}$, y : Castle

$f_{\boldsymbol{\theta}}$

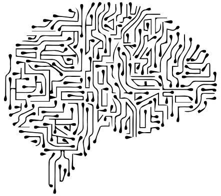


Adversarial Examples

$\boldsymbol{x},\ y$: Castle

$\xrightarrow{\boldsymbol{x}}$

$f_{\boldsymbol{\theta}}$



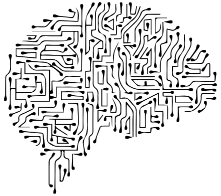
$\longrightarrow$ $f_{\boldsymbol{\theta}}(\boldsymbol{x}) = y$ Castle

Adversarial Examples

$\boldsymbol{x},\ y$: Castle

$f_{\boldsymbol{\theta}}$

$\xrightarrow{\boldsymbol{x}}$



$\longrightarrow \quad f_{\boldsymbol{\theta}}(\boldsymbol{x}) = y$ Castle

$\boldsymbol{x}$ $\boldsymbol{\delta}$

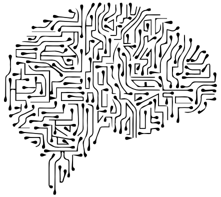
$+$

Adversarial Examples

$\boldsymbol{x},\ y$: Castle

$f_{\boldsymbol{\theta}}$

$\xrightarrow{\boldsymbol{x}}$



$\longrightarrow$ $f_{\boldsymbol{\theta}}(\boldsymbol{x}) = y$ Castle

$\boldsymbol{x}$ $\boldsymbol{\delta}$

$+$

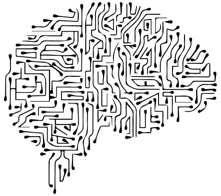
$\|\boldsymbol{\delta}\|_2 \leq \epsilon$

Adversarial Examples

$\boldsymbol{x}$, y : Castle

$f_{\boldsymbol{\theta}}$

$\xrightarrow{\boldsymbol{x}}$



$\longrightarrow$ $f_{\boldsymbol{\theta}}(\boldsymbol{x}) = y$ Castle

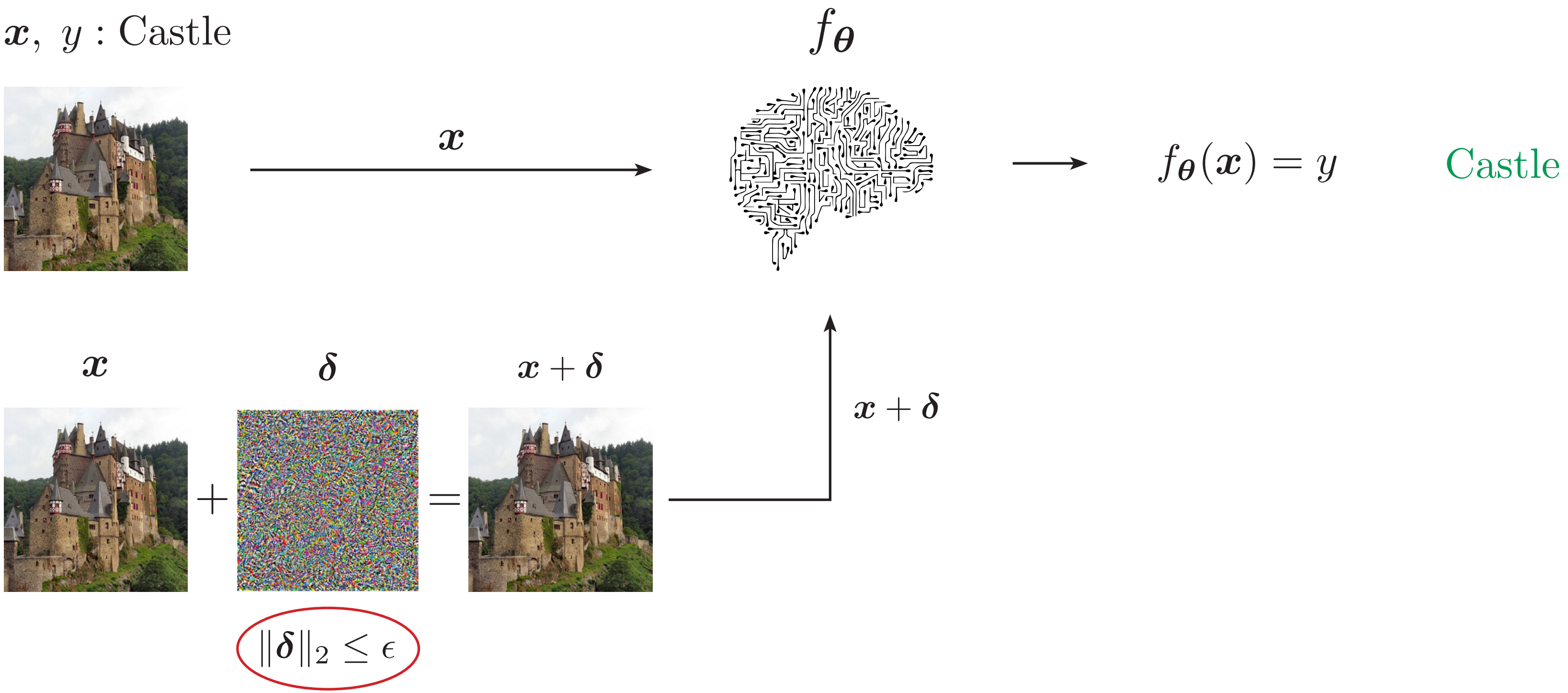
$\boldsymbol{x}$ $\boldsymbol{\delta}$ $\boldsymbol{x} + \boldsymbol{\delta}$

$+$

$=$

$\|\boldsymbol{\delta}\|_2 \leq \epsilon$

Adversarial Examples

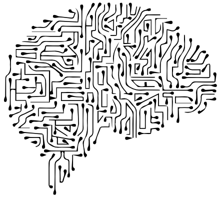


Adversarial Examples

$\boldsymbol{x},\ y$: Castle

$\xrightarrow{\boldsymbol{x}}$

$f_{\boldsymbol{\theta}}$



$\longrightarrow$

$f_{\boldsymbol{\theta}}(\boldsymbol{x}) = y$ Castle

$f_{\boldsymbol{\theta}}(\boldsymbol{x}+\boldsymbol{\delta}) \neq y$ Boat

$\boldsymbol{x}$ $\boldsymbol{\delta}$ $\boldsymbol{x}+\boldsymbol{\delta}$

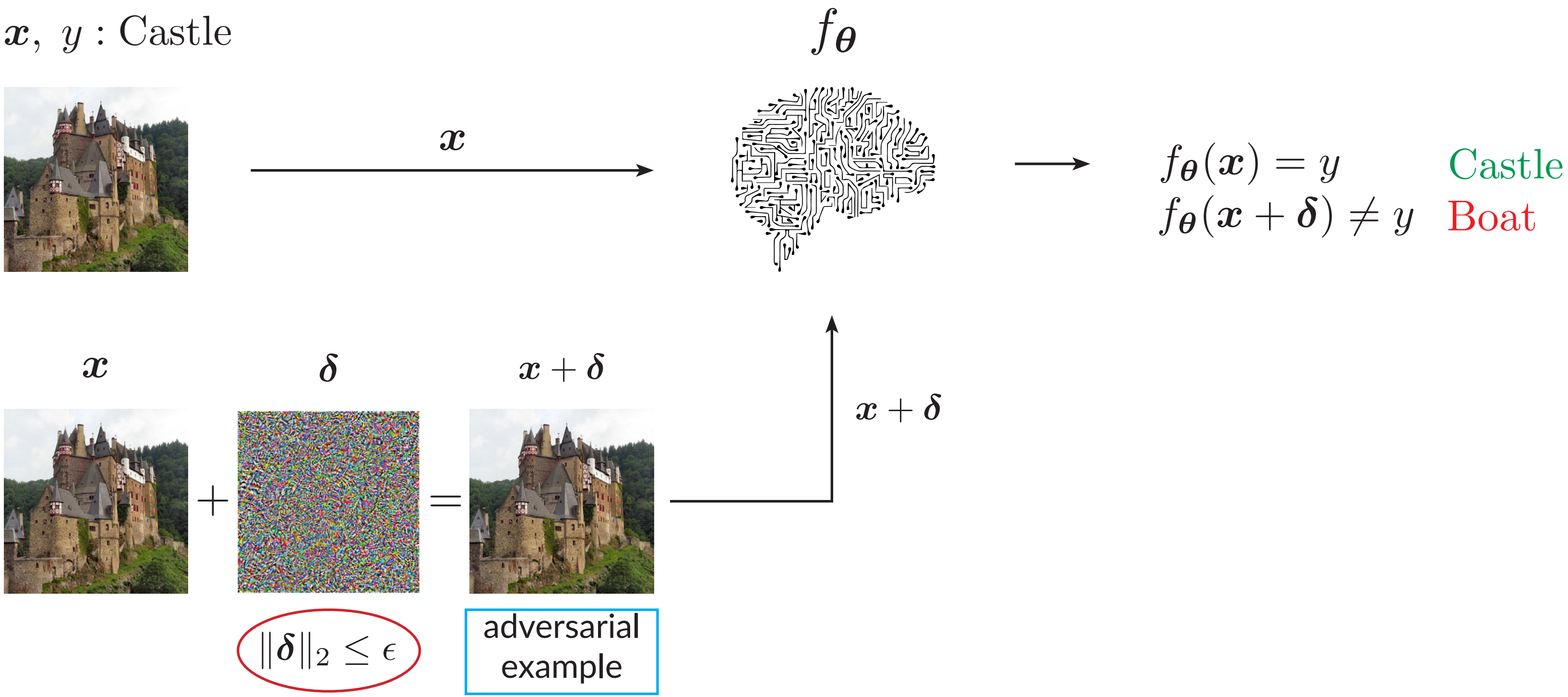
$+$

$=$

$\boldsymbol{x}+\boldsymbol{\delta}$

$\|\boldsymbol{\delta}\|_2 \leq \epsilon$

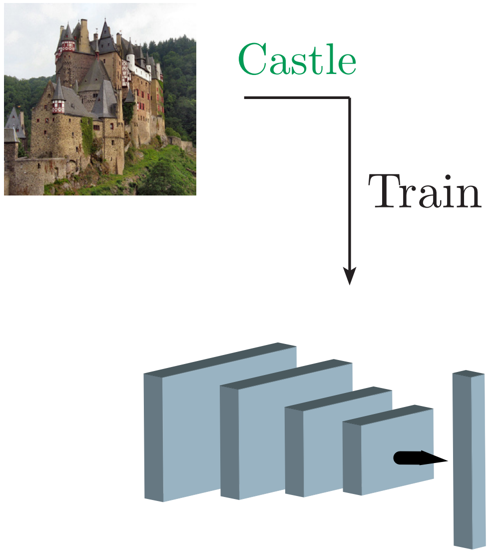
Adversarial Examples

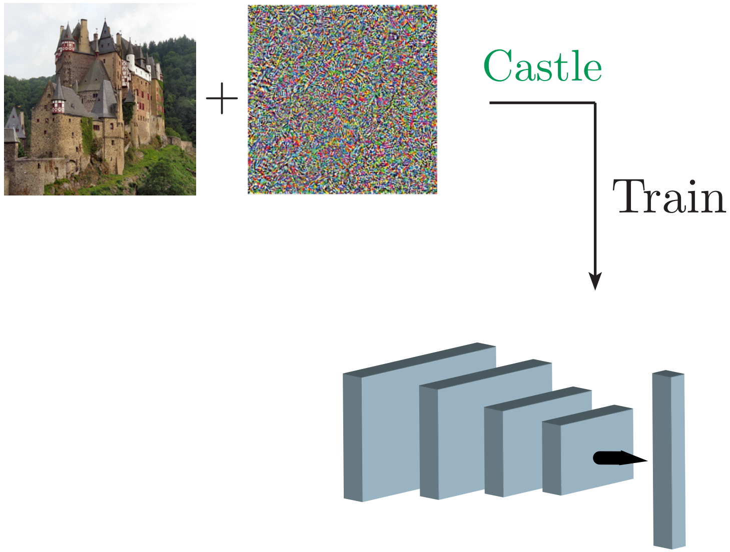


Adversarial Training

How to make the network **robust** = **Adversarial Training (AT)**

Instead of training with natural examples $\longrightarrow$ Train with adversarial examples





Why adversarial training?

AT improves transfer learning! [1]

- AT <u>on the source</u> domain, then fine-tune on the target
- Better results than standard transfer learning
- Evaluated and holds for many computer vision tasks!

[1] H. Salman et al. "Do adversarially robust ImageNet models transfer better?", NeurIPS 2020

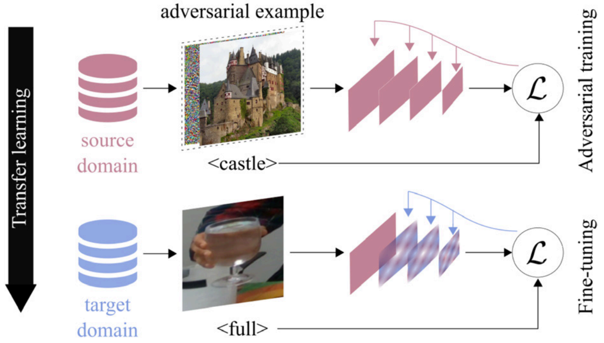
Why adversarial training?

AT improves transfer learning! [1]

- AT <u>on the source</u> domain, then fine-tune on the target
- Better results than standard transfer learning
- Evaluated and holds for many computer vision tasks!

Question: would it hold for filling level estimation?

- Quite novel task
- What paramteres should be used?
- What if we also perform AT on the target domain?

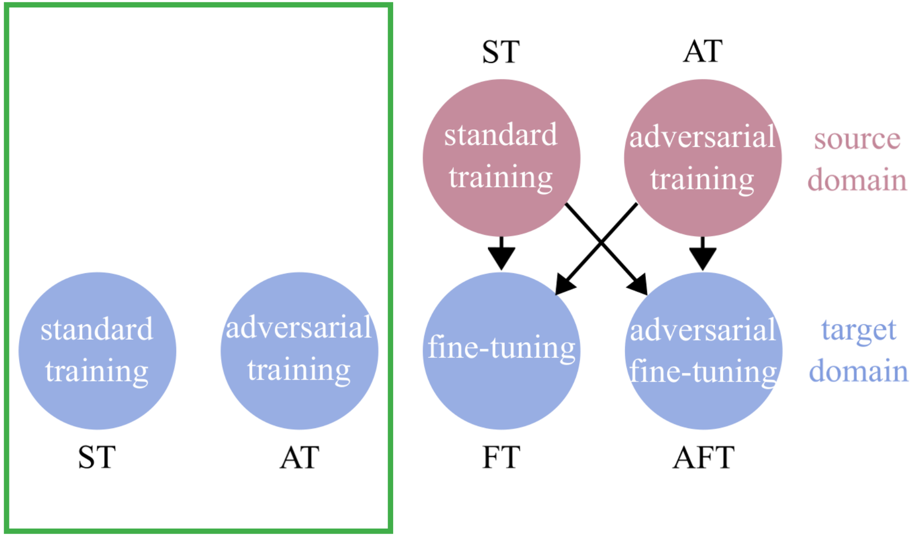


[1] H. Salman et al. "Do adversarially robust ImageNet models transfer better?", NeurIPS 2020

Setup

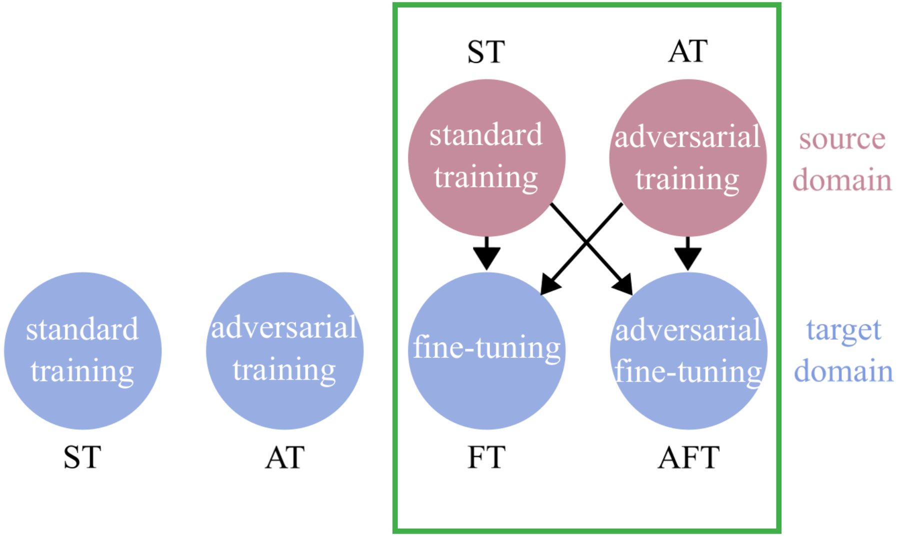
The training strategies

ResNet-18



The training strategies

ResNet-18



The dataset

C-CCM: Image crops from the CORSMAL Containers Manipulation Dataset [1]

- Large variability (transparency, shape, etc)
- 8 objects: **4 cups and 4 drinking glasses**
- In total: **10,216 RGB images**
- Filling level: **0%, 50%, 90%, "*unknown*"**
- Filling type: **water, pasta, rice**

[1] A. Xompero et al. "CORSMAL Containers Manipulation Dataset (1.0)", https://doi.org/10.17636/101CORSMAL1

Dataset configurations

Config. 1 (S_1)

Train | Test

- Champagne flute in test set

Config. 2 (S_2)

Train | Test

- All stems in test set

Config. 3 (S_3)

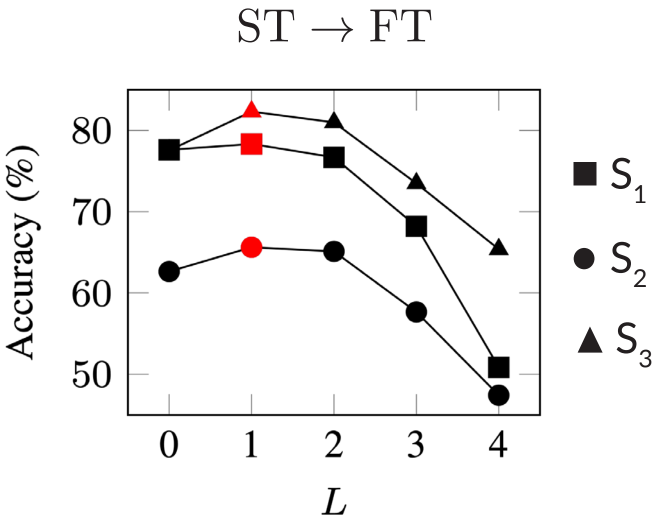
Train | Test

- All stems in train set
- Color & opaque in test set

Experimental Results

Sensitivity analysis

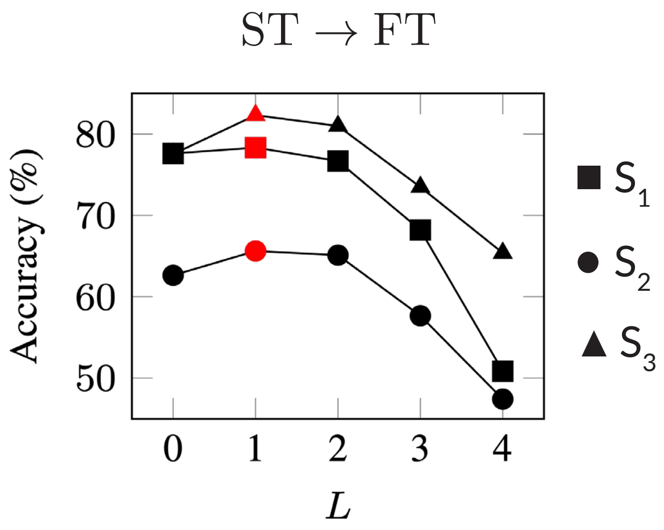
<u>Freezing layers during standard fine-tuning</u>



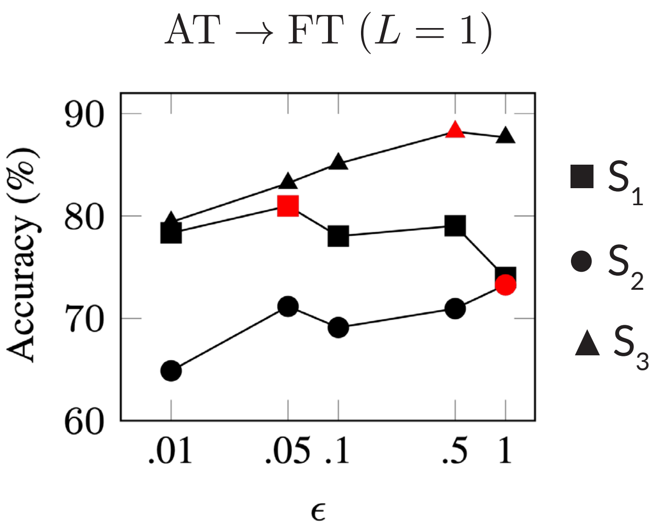
- Fixing the 1st layer results in the highest test accuracy

Sensitivity analysis

Freezing layers during standard fine-tuning

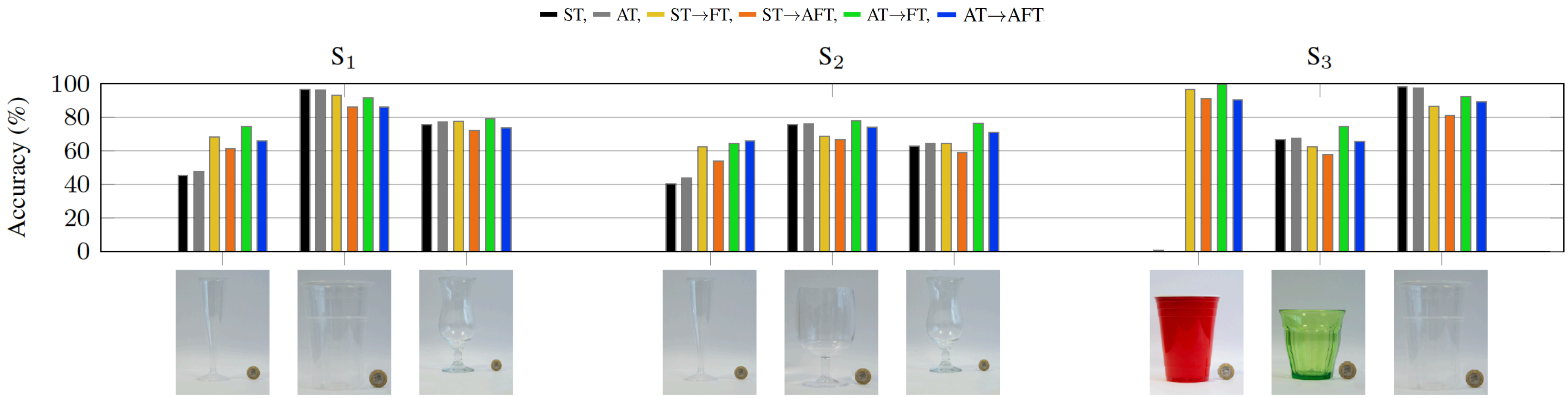


Perturbation size ε (source domain)



- Fixing the 1st layer results in the highest test accuracy
- Robust models (source) adversarially trained with different ε lead to highest test accuracy

Comparisons



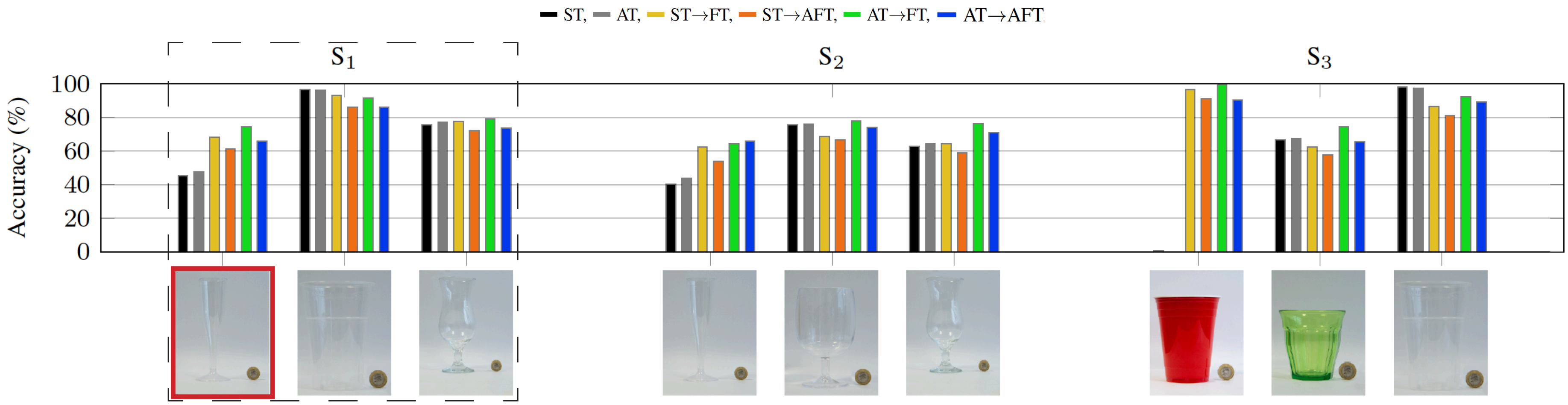
AT→FT : best results most of the times

ST→FT : ImageNet features reduce biases

AT→FT : ImageNet features are aslo filtered by AT
and improve generalization even further

: AT on the target domain is not really helpful

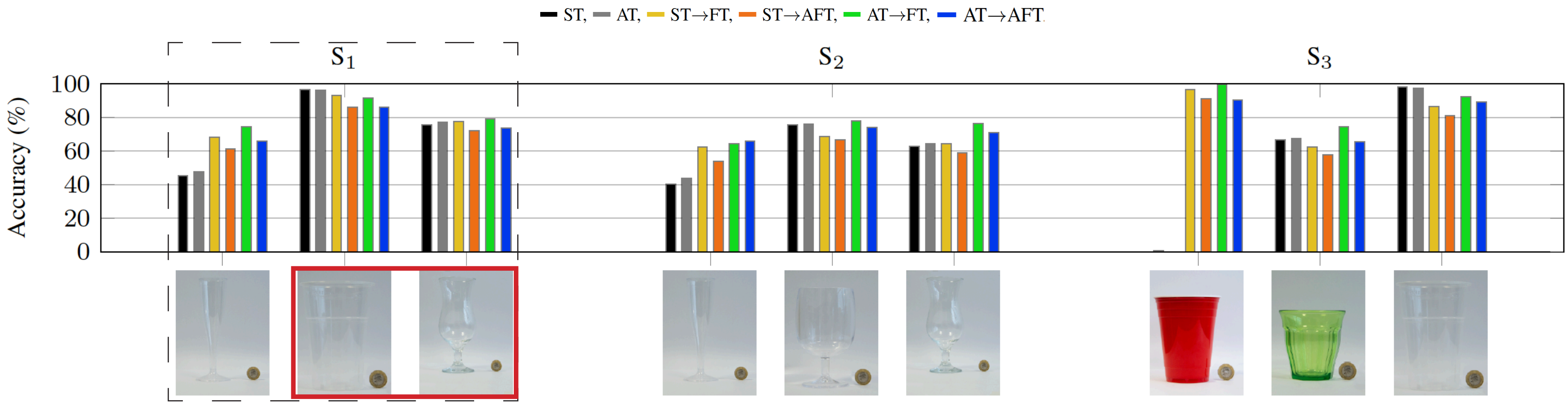
Comparisons



ST, AT : Cannot cope with shape above stem

AT→FT : Improves by 1.8x the performance

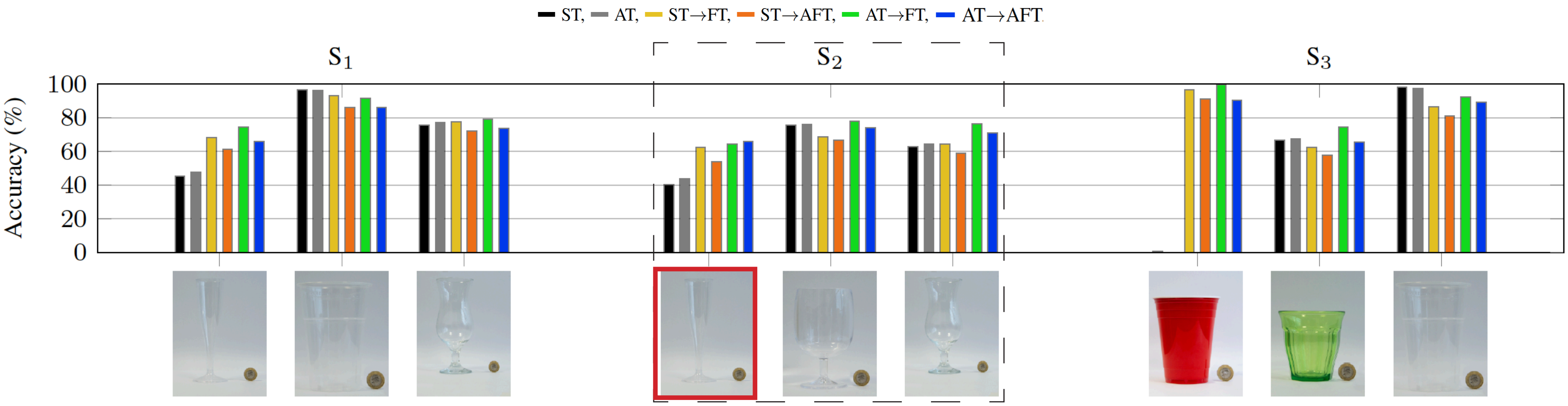
Comparisons



Beer cup: same shape as small transparent cup of the train set, but just bigger

Cocktail glass: many similarities with wine glass of the train set, but still not exact same shape

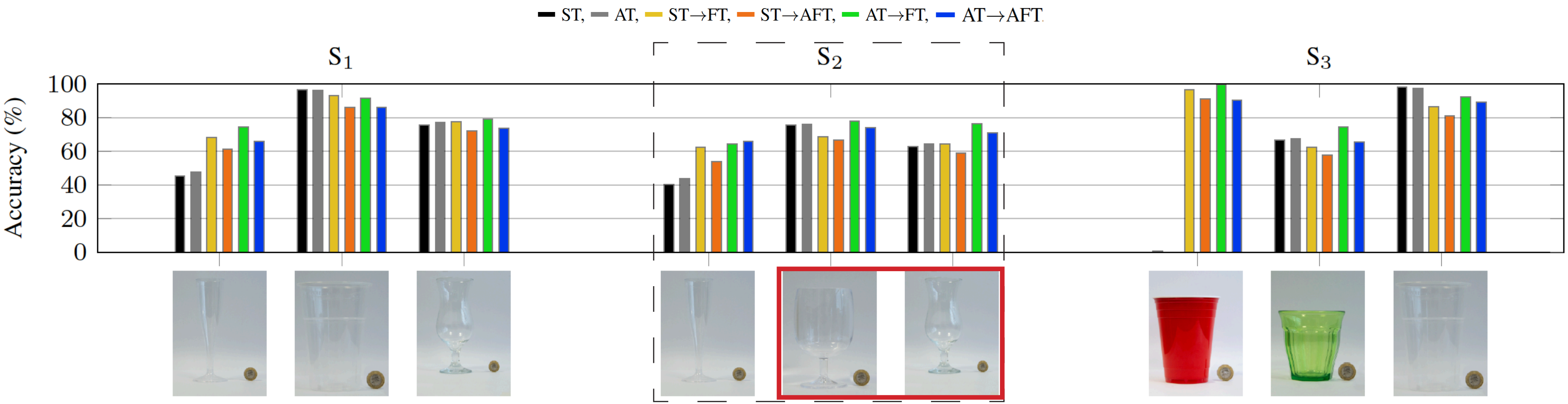
Comparisons



ST, AT : Cannot cope with shape above stem

AT→FT : Improves by 1.6x the performance

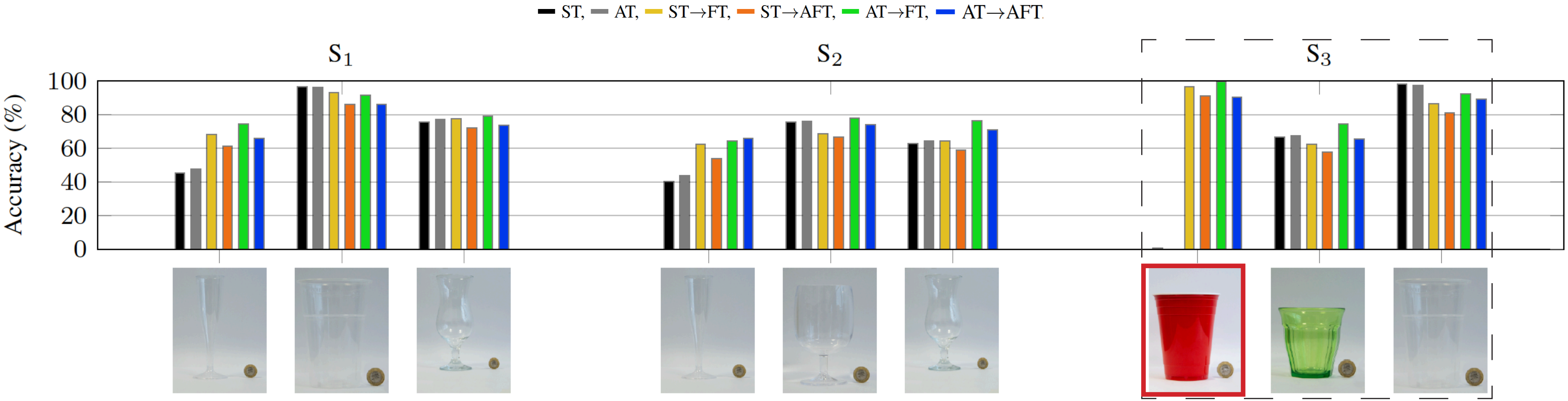
Comparisons



ST, AT : Good results - shape above stem is “sufficiently” regular

AT→FT : Much better than standard transfer learning

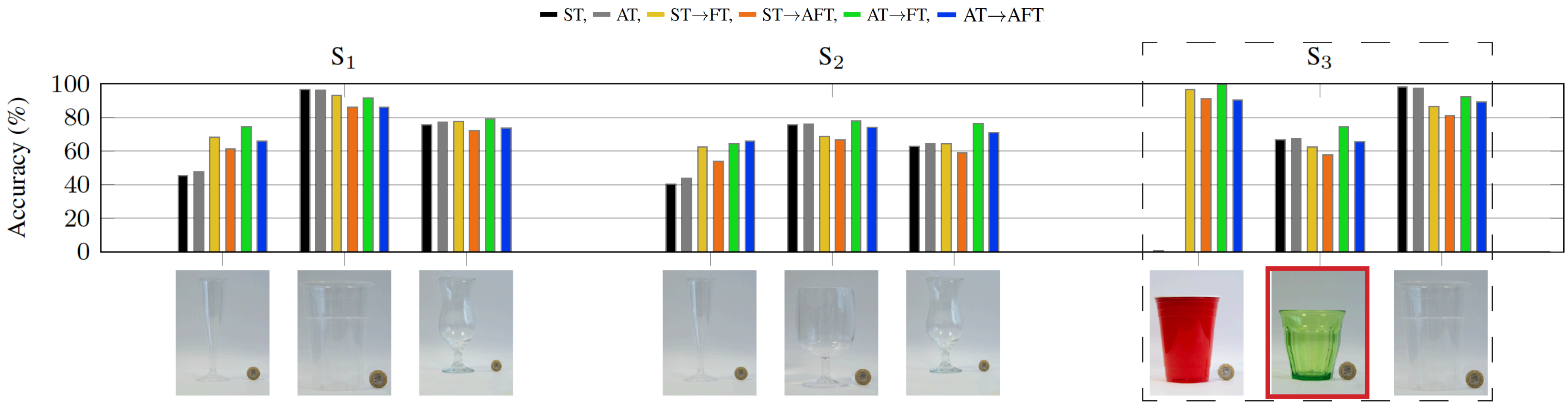
Comparisons



ST, AT : Almost **0%**! In fact, 99% of predictions are "90% full".
Possibly: the opaque red cup resembles a "transparent cup" + "90% with rice/pasta" of the train set

AT→FT : Superior performance - generally all transfer learning strategies improve

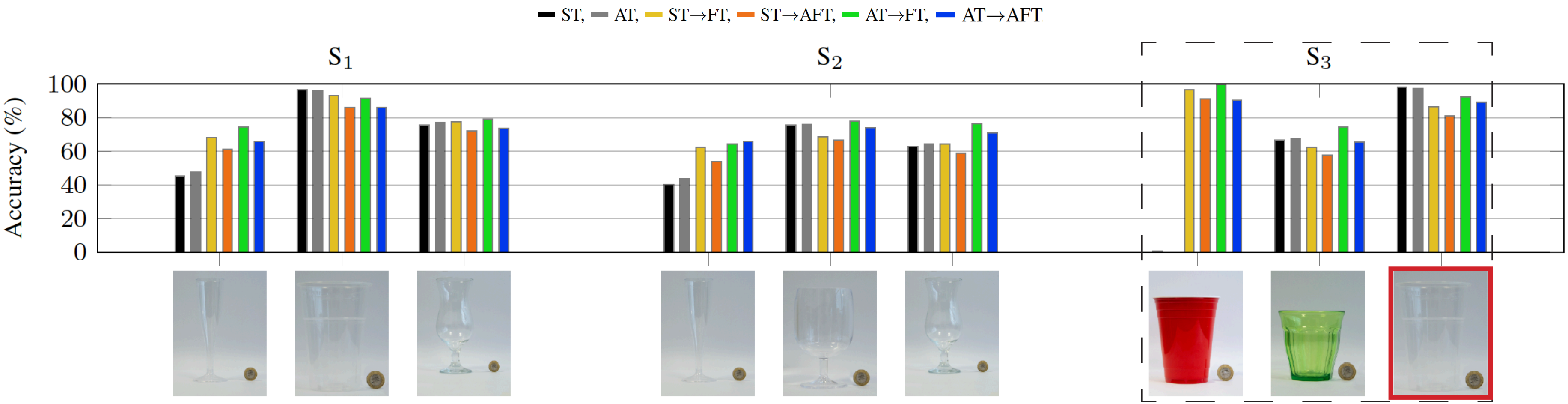
Comparisons



Almost every method performs similarly

AT→FT : Superior performance, almost +10% accuracy

Comparisons



All methods perform very well: same shape as the small transparent cup of the train set, but just bigger

Conclusions

Estimate the content level within a container

- Classification task

Release a new dataset: Cropped CORSMAL Containers Manipulation (C-CCM)

- Variability in shape, content, transparencies, occlusions

Training strategies

- Explored different training strategies
- With standard training: overfitting to specific features (ie, shape)

AT (source) + Fine-tuning

- Improves standard transfer learning
- Superior performance & eliminates biases

Improving filling level classification with adversarial training.

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

EPFL

Queen Mary
University of London