Learning multiple visual domains with residual adapters

Sylvestre-Alvise Rebuffi*, Hakan Bilen† and Andrea Vedaldi*

*University of Oxford †University of Edinburgh

Abstract

We study how to learn diverse visual domains with a single deep network, proposing the Visual Decathlon Challenge. We introduce the residual adapters, generic plug-and-play adaptation modules for deep networks that can learn diverse domains while achieving a high degree of parameter sharing. Our adapters also allow to learn new domains sequentially without forgetting.

Motivation

Goal: multiple domain learning

We want to:
- Learn diverse visual domains
- Use only one model
- Use few domain specific parameters

We do not want to:
- Forget the previous domains when adding new ones

Existing Approaches

Finetuning:
- Adapts a pretrained network to each new domain
- Requires a new network per domain as it forgets the original domain
- Good performance, but requires lots of parameters due to forgetting

Shared network:
- Use a pretrained network as a feature extractor
- Worse performances, but few parameters as there is no forgetting

Residual Adapters and architectures

Residual adapters: \( y = (l d + a) \ast w \ast x = ((l d + a) \ast w) \ast x \)
- Are plugged all along the network (here a ResNet [He, 2016])
- Produce a compact parametric family of networks
- Require only 10% of domain specific parameters
- Do not forget (by construction)
- Boils down to identity with enough shrinkage
- Allows per domain overfitting control

Residual Adapters used with a LearNet

LearNet [Bertinetto, 2016]:
- Uses a light ResNet
- Achieves 99.8% accuracy in domain prediction
- Achieves automatic and dynamic domain selection

Visual Decathlon Challenge

10 datasets:
- Different difficulty levels (small, medium and big datasets)
- Different domains (digits, textures, dynamic and natural images)
- Resized to 72 pixels

Decathlon score: \( S \sim \max(0, E_{\text{baseline}} - E)^2 \)
- Improvement relative to a baseline
- Every dataset has same importance
- Power law to favour big improvements wrt the baseline

Results

Accuracy (Decathlon deltas)

Model Size

Comments:
- Results on par or better than finetuning with 5x less parameters
- Using specific Batch Normalization like in [Bilen, 2016] or FC layers is not enough
- These adapters could be used for other transfer learning problems
- Submitted extension with better Decathlon scores and more compressed adapters