Large-Scale Counterfactual Training for Object Classification

By: Sebastian Palacio

Description

Object classification problems for machine learning can be subject to biases coming from regularities of the background. End-to-end training schemes or regularization techniques cannot prevent parametrized models to base predictions on said background. The goal of this project is to implement novel training schemes that maximize entropy for features found in the background, concurrently with the original classification objective. Effectiveness of this approach will be evaluated on large-scale datasets like ImageNet or MS-COCO.

Requirements

- Python
- Numpy
- Matplotlib
- PyTorch
- Scikit-learn
- Machine Learning
- Deep Learning

Test Task

Start by taking a Resnet18 (not pre-trained) and train for 10 epochs using SGD with learning rate of 1e-1 and momentum of 9e-1 on the CIFAR10 training set. Make sure the data is normalized to the unit cube (e.g., pixel values are mapped to the interval [0, 1]). Report **accuracy per-class** on the test set of CIFAR10.
Repeat the same process (training and evaluation) but this time, using only the red channel as input (you'll still need to make a 3D tensor by copying the data of the red channel in the blue and green channels).
Finally, run the training and evaluation two more times, using the green and blue channels as input respectively.
Present all results in a single table where the accuracy of each class is reported for the 4 experiments. What conclusions or insights can you gather from these results?
Make a PDF with the results and upload the code to a repository of your choosing (GitHub, Gitlab, BitBucket, etc).
Send the PDF and a link to the repository to sebastian.palacio@dfki.de