Real-time Semantic Computer Vison for Co-robotics

——Towards Realistic Predictors

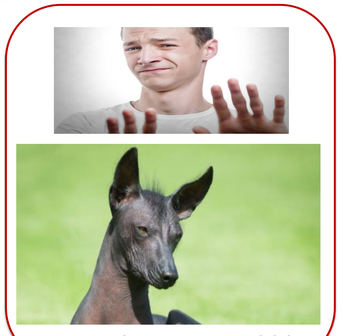
Nuno Vasconcelos

Statistical Visual Computing Laboratory, Dept. of Electrical and Computer Engineering, University of California, San Diego

Introduction

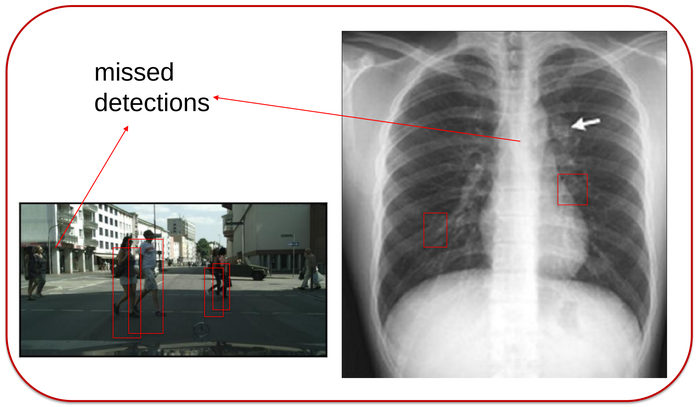
- Most machine learning and computer vision models are over-optimistic. They are trained to perform as well as possible on every example, without regard for how hard the task is.
- This is not what humans do! They analyze the difficulty of each task first, accept tasks that are doable and refuse tasks that are too hard. We refer this as **realism**.

Dalmatian

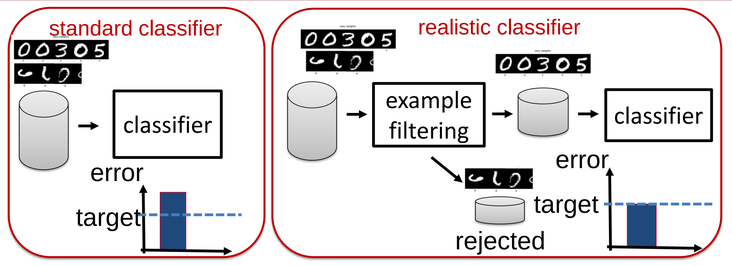


Peruvian Inca Orchid

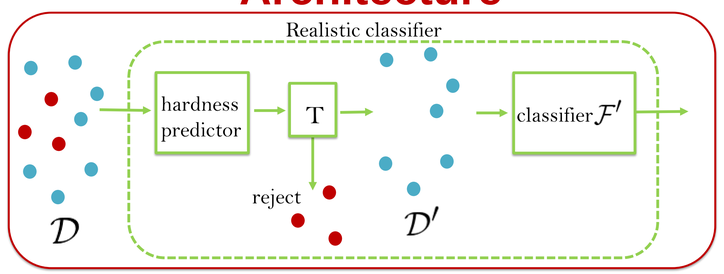
This trait is critical for applications where it is OK not to classify all examples so as to guarantee performance rates on the ones that are classified, like circumventing risk in autonomous driving and tumor detection and classification;

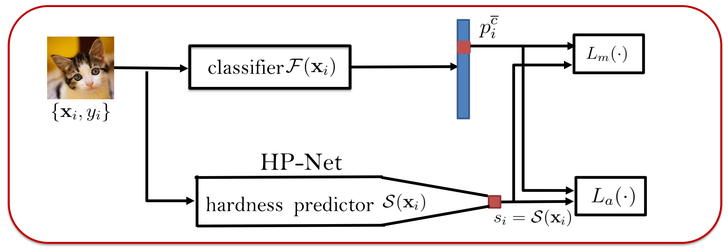


- Vision system should either
  - 1. Refuse to perform the hard task;
  - 2. Request additional information from other sensors.
- We define and propose a new type of classifiers, **realistic classifiers** that reject examples deemed too hard so as to guarantee a target performance on the accepted examples.



Architecture



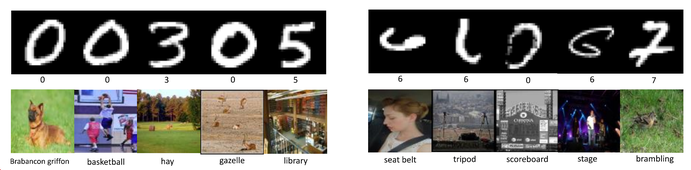


- **Training**
  - Train classifier $\mathcal{F}$ and HP-Net S jointly on training set D;
  - Run S on D and eliminate hard examples, to create realistic training set D′;
  - Learn realistic classifier $\mathcal{F}'$ on D′, with S fixed;
  - Output pair S and $\mathcal{F}'$.
- **Testing**
  - Run test examples x by S, reject hard examples, classify remaining with $\mathcal{F}'$. In all cases, x rejected if $S(x) > T$, for some threshold T.
- **Loss function**

$$L_m(\{\boldsymbol{x}_i, y_i, s_i\}) = \sum_i s_i \, l(\boldsymbol{x}_i, y_i)$$

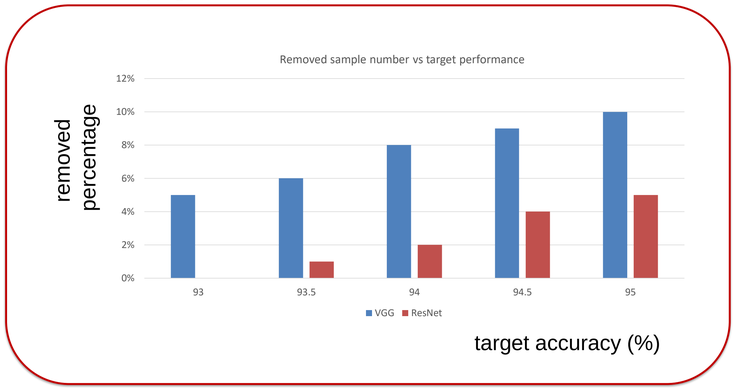
$$L_a(\{\boldsymbol{x}_i, y_i, s_i\}) = -\sum_i \{p_i^c \log(1 - s_i) + (1 - p_i^c) \log s_i\}$$

Experiments



- Easy samples are clearly written, close to prototypical digits, while hard samples are "in between" digits.
- **Evaluations of realistic predictors**

- Realistic predictor guarantees a target performance while accepting and classifying more examples than non-realistic one.



- VGG is a weaker model, but equals the ResNet when rejecting 5% of the examples, and beats it by 2% when rejecting 10%.

SVCL UCSD

Award ID#: 1637941