Unsupervised Difficulty Estimation with Action Scores

Octavio Arriaga[1] (arriagac@uni-bremen.de) and Matias Valdenegro-Toro[2] (matias.valdenegro@dfki.de)

[1]University of Bremen. [2]German Research Center for Artificial Intelligence. Bremen, Germany

Deutsches Forschungszentrum für Künstliche Intelligenz GmbH

Universität Bremen

Summary

- **Motivation:** It is well known that for ML models, classifying certain examples might be more difficult, due to variations in object pose, noise, object size, etc. This is important in determining biases in models and datasets, and a method that does not require changes in a model is preferable. We hypothesize that the per-sample loss during training contains information about difficulty.
- **Approach:** We define an action score as the accumulation of per-sample loss values across epochs during training. Our results show that visually difficult examples obtain a higher action scores, while easier examples obtain low scores. This confirms our hypothesis.
- **Contributions:** We propose the action score for difficulty estimation without any kind of supervision. Our approach does not require changes to the model, and allows to extract more information from the training process. We evaluate our method in CIFAR10 classification and SSD object detection on PASCAL VOC, which show that our method can be applied to simple and complex multi-task settings.

Action Scores

Given a loss function $\mathscr{L}$ and a model m with free parameters θ_n, we define the action $\mathscr{A}$ of a sample $x \in X$ with labels $y \in Y$ as

$$\mathscr{A}(x) = \sum_{n=0}^{N} \mathscr{L}(y, m(x; \theta_n))$$

where n represents epochs. Consequently, the action of a sample is the accumulated loss over all epochs. Our method characterizes the action of each sample as a measurement of its difficulty. Therefore, samples with a higher accumulated loss represent samples that are more difficult to learn.
Specifically, we argue that the action is directly proportional to its difficulty i.e.

$$D(x) \propto \mathscr{A}(x)$$

Note that action can be computed only during training, and it is available for training and validation sets.

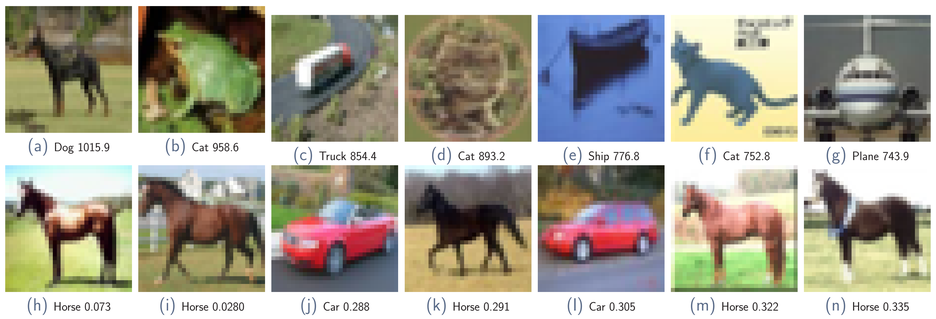
Experimental Evaluation and Analysis

- We evaluate on CIFAR10 classification, and SSD object detection on the PASCAL VOC dataset. For SSD, the loss can be decomposed into localization loss (bounding box regression), positive classification loss, and negative classification loss (for background windows). We compute an action score for each loss component separately. We group hard and easy examples by the top and bottom action scores.
- On CIFAR10, the easiest examples all contain objects in a canonical pose that is easy to classify, while the hardest examples have easy to confuse classes, difficult backgrounds, or incorrect labels.
- On PASCAL VOC object detection, we see similar issues, with the localization loss indicating easy examples to ones in easy poses, while the hardest examples contain small objects. Classification losses also show struggles with very small objects or cluttered scenes.

Conclusion & Future Work

- Our results show that the action score correlates with visual difficulty of training samples, which is useful to debug models and datasets.
- Future work will consider evaluating on more tasks (segmentation, pose estimation, etc), and evaluating the robustness of the action score to variations in the training process.

CIFAR10 Classification

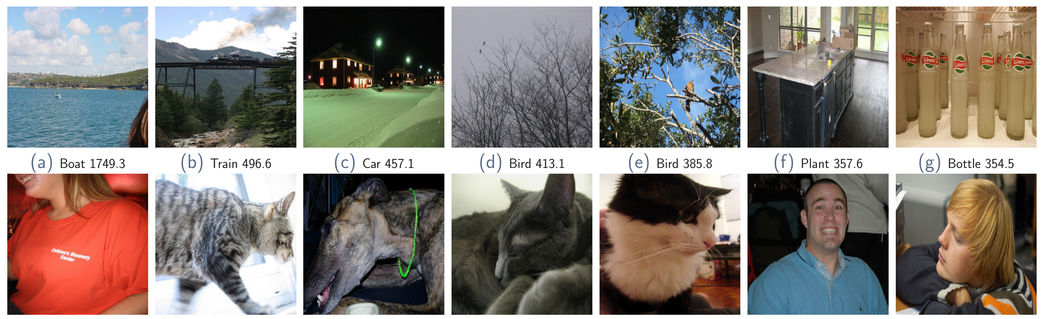


(a) Dog 1015.9 (b) Cat 958.6 (c) Truck 854.4 (d) Cat 893.2 (e) Ship 776.8 (f) Cat 752.8 (g) Plane 743.9

(h) Horse 0.073 (i) Horse 0.0280 (j) Car 0.288 (k) Horse 0.291 (l) Car 0.305 (m) Horse 0.322 (n) Horse 0.335

Figure: Most difficult (top-row) and easiest examples (bottom-row) in CIFAR10. Our proposed *action score* is displayed below each image as well as the true label.

SSD Object Detection on PASCAL VOC - Localization Loss



(a) Boat 1749.3 (b) Train 496.6 (c) Car 457.1 (d) Bird 413.1 (e) Bird 385.8 (f) Plant 357.6 (g) Bottle 354.5

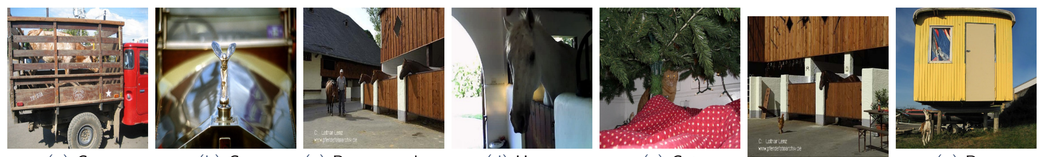
(h) Person 3.9 (i) Cat 4.1 (j) Dog 4.2 (k) Cat 4.7 (l) Cat 5.0 (m) Person 5.2 (n) Person 5.3

Figure: Most difficult (top-row) and easiest examples (bottom-row) in the VOC 2007-VAL with the SSD localization loss. The *action scores* are displayed below each image as well as the true label.

SSD Object Detection on PASCAL VOC - Classification Losses

(a) Person 1086.9 (b) Boat 626.9 (c) Table 548.1 (d) Person 476.6 (e) Table 452.9 (f) Chair 416.2 (g) Bicycle 401.5

Figure: Hardest Examples on PASCAL VOC 2007 with SSD validation negative loss. Action score is included in each caption.



(a) Cow 596.6 (b) Car 568.5 (c) Person and Horse 565.0 (d) Horse 544.6 (e) Cat 544.3 (f) Plant, Bottle and Horse 527.8 (g) Dog 524.7

Figure: Hardest Examples on PASCAL VOC 2007 with SSD validation positive loss. Action score is included in each caption.