Curriculum Learning Effectively Improves Low Data VQA

Narjes Askarian[1], Ehsan Abbasnejad[2], Ingrid Zukerman[1], Wray Buntine[1], Gholamreza Haffari[1]

[1]Dept. of Data Science and AI, Monash University [2]Australian Institute for Machine Learning, The Univ. of Adelaide

1. Motivations

Current VQA models are commonly trained on large-scale datasets to achieve state of the art performance. However, such datasets are not available for many domains. Further, these models on small datasets significantly reduces their high performance.

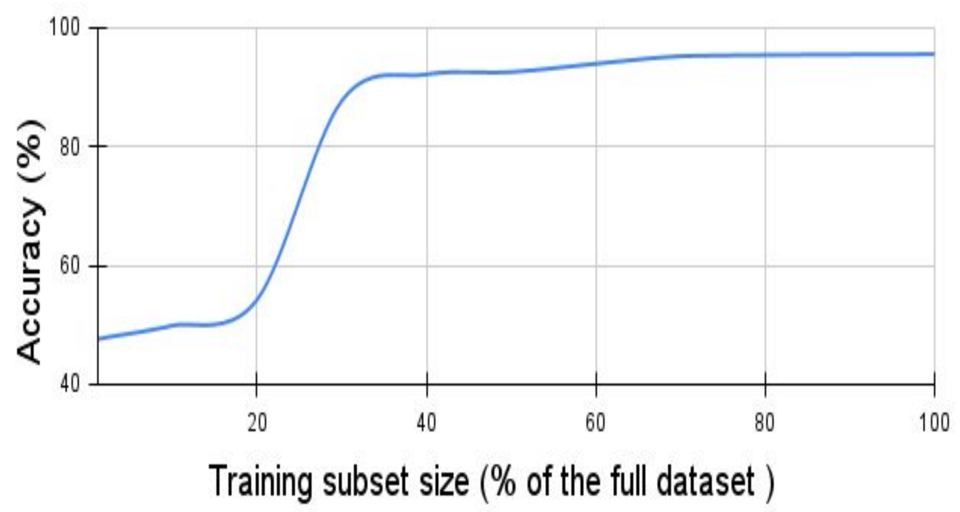


Figure 1: Accuracy of a VQA model when training on different-sized training set

Our goal is to train a modular VQA model from scratch by using only a small amount of labeled data without using any other resources. Specifically, we take the Curriculum learning approach to tackle the problem of VQA models' low performance under low data conditions.

2. Curriculum Heuristics

Curriculum by program length

We consider the length of the program corresponding to a question as an indicator of question length. Under the program length curriculum, the network is fed with easy-to-hard ranked examples starting from shorter programs and gradually increasing programs' length.

Curriculum by answer hierarchy

we define another measure based on a hand-crafted answer hierarchy in order to shift the focus from questions to answers. The higher level in the hierarchy includes a coarser categorization of each answer type, and the answer types are vertically extended downward to finer classes of types, *e.g., digit* at a lower level is divided into three groups, such as *0*, *1* and *many*.

Curriculum by hard examples

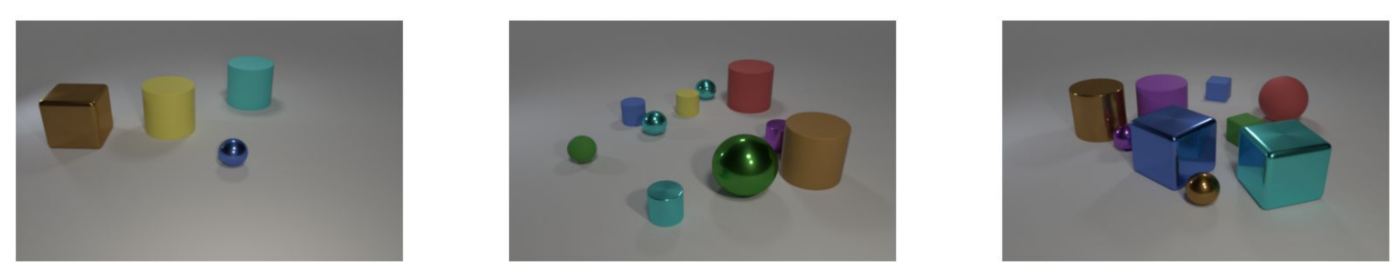
The intuition of this heuristic is to focus training on the hard examples where the learner does not perform well and consequently the loss is high [1]. We employ a dynamic hardness criterion H_t based on the running average of *instantaneous hardness* r_t , which is defined as the loss difference between two consecutive training iterations. γ is the discount factor.

$$r_t(i) = |\ell_t(a_i - \mathcal{E}(\mathbf{x}_i, p_i; w_t)) - \ell_{t-1}(a_i - \mathcal{E}(\mathbf{x}_i, p_i; w_{t-1}))|$$

$$H_{t+1}(i) = \begin{cases} \gamma \times r_t(i) + (1-\gamma) \times H_t(i) & \text{if } i \in S_t \\ H_t(i) & \text{else} \end{cases}$$

Hardness	Epoch					
	1	10	25	50	75	98
Easy	0.90	0.81	1.16	0.93	1.16	1.12
Medium	5.49	1.87	2.31	1.40	1.33	1.27
Hard	11.78	3.57	1.74	1.10	0.94	1.40

Table 1: Hardness scores of three examples at different epochs with various levels of difficulty. The hardness scores decrease as training progresses.



Easy Q: There is an object that is both right of the yellow rubber object and behind the large brown thing; what is its color? **A:** cyan

Medium Q: What number of large objects are cyan metallic spheres or yellow spheres? **A:** 0

Hard Q: What size is the metal block right of the brown metal thing right of the blue thing in front of the small blue rubber thing? **A:** large

(A) Easy Question (B) Medium Question (C) Hard Question

Figure 2: Examples of easy, medium and hard VQA tasks according to their hardness score.

3. Curriculum Training

- **Training by length-based curriculum:** CL training with a batching method as the selection function and a linear paced scheduler
- **Training by answer hierarchy curriculum:** CL training with a self-paced scheduler. he scheduler updates the curriculum where the normalized difference of accuracy between two consecutive iterations goes higher than a predefined threshold.
- **Training by hard examples curriculum:** CL training with a warm-up phase where the model sweeps all training examples and then a curriculum learning where the examples are ranked according to their hardness score.

4. Results

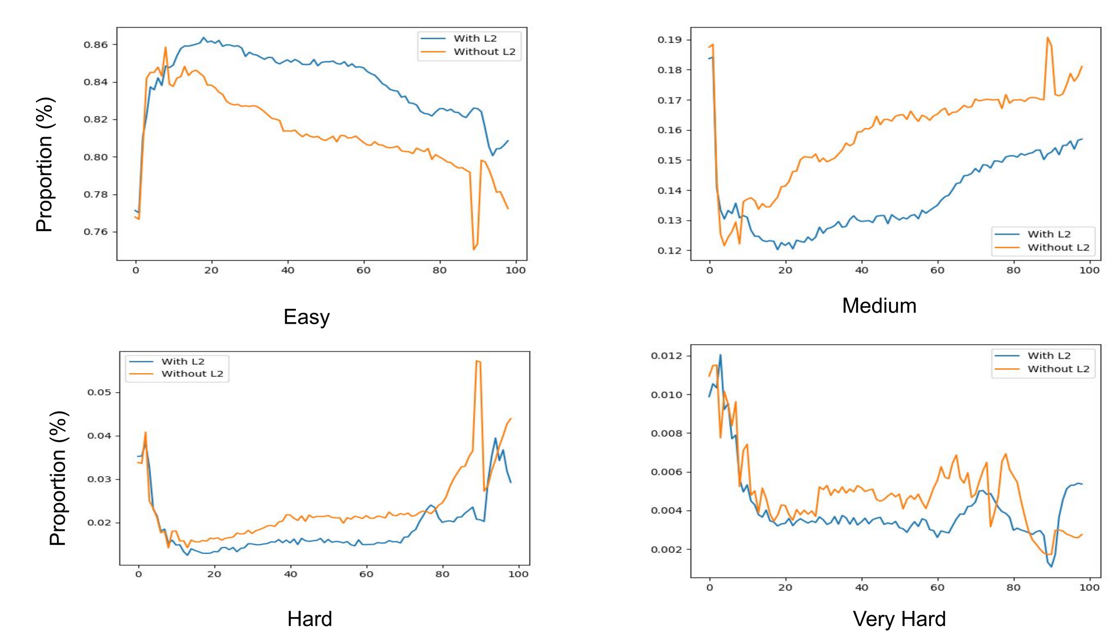
To simulate a low-data scenario, we select four small subsets from the CLEVR training set [2] with different sizes denoted as a percentage of the full dataset. We evaluate our approach using the modular VQA model proposed in [3]. To investigate the impact of regularization, we asses our experiments under three conditions: no regularizer (**No-Reg),** with dropout applied on the last layer of the model **(Drop-out),** and with L2 regularization **(L2-norm)**.

Method	No-Reg				Drop-out				L2-norm			
	5%	10%	15%	20%	5%	10%	15%	20%	5%	10%	15%	20%
No-CL	46.91	48.77	49.68	51.25	46.94	48.36	49.67	49.92	46.71	50.25	52.20	54.34
Length-CL	46.55	46.67	47.83	48.12	46.68	47.33	47.61	47.71	47.89	49.65	50.98	51.50
AnswerH-CL	47.42	48.59	49.73	51.65	47.43	47.73	48.60	50.24	48.62	49.03	48.70	48.95
HardEx-CL	47.93	50.04	51.97	53.14	48.80	49.94	51.69	56.29	48.95	51.49	53.27	**87.62**±1.3

Table 2: The model accuracy (%) on CLEVR val when training on training subsets of size 5%, 10%, 15% and 20% with three different choices of curriculum. The length-based (Length-CL) and answer hierarchy (AnswerH-CL) curriculum does not improve the performance while hard example (HardEx-CL) outperforms the vanilla baseline (No-CL) in all experiments.

Regularization impact

Our ablation studies shows that in contrast to dropout and L1-norm, using L2 regularization results in improved performance in almost all the experiments. The following plots shows that L2-norm prevents forgetting the patterns learned from easy examples by forcing the sampling function to incorporate more samples from easy category.



References:

[1] Tianyi Zhou, Shengjie Wang, and Jeffrey Bilmes. 2020, *Curriculum Learning by Dynamic Instance Hardness.* In Advances in Neural Information Processing Systems (NeurIPS)

[2] Justin Johnson, Bharath Hariharan, Laurens van der Maaten, Li Fei-Fei, C Lawrence Zitnick, and Ross Girshick. 2017. *Clevr: A diagnostic dataset for compositional language and elementary visual reasoning.* In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 2901–2910.

[3] Justin Johnson, Bharath Hariharan, Laurens van der Maaten, Judy Hoffman, Li Fei-Fei, C Lawrence Zitnick, and Ross Girshick. 2017. *Inferring and executing programs for visual reasoning.* In IEEE International Conference on Computer Vision (ICCV), pages 2989–2998.