Machine Learning Research Project

Anonymous CVPR submission

Paper ID ****

Abstract

The ABSTRACT is to be in fully-justified italicized text, at the top of the left-hand column, below the author and affiliation information. Use the word "Abstract" as the title, in 12-point Times, boldface type, centered relative to the column, initially capitalized. The abstract is to be in 10-point, single-spaced type. Leave two blank lines after the Abstract, then begin the main text. Look at previous CVPR abstracts to get a feel for style and length.

1. Introduction

In machine learning we see two types of research. The first is application-oriented research, where the paper will attempt to create a solution to a specific problem using machine learning. The second is generalized research, where the paper will push the state-of-the-art models in order to improve them, using standardized benchmarks to demonstrate the improvements. Application-oriented research often uses specialized datasets relevant to the examined domain. TrashNet [3] is a dataset used in waste classification. It contains 2527 images, taken from mobile devices, and categorized as: glass, paper, cardboard, plastic, metal, and trash. As most domain-specific datasets, TrashNet is extremely small; however, a number of techniques, that will be discussed later, are used to increase the sample number. In this paper we will be focusing on generalized research datasets. These datasets usually contain a large number of samples and features, they are extremely well documented, and often have many variants, curated versions of the original dataset that caters to a specific learning requirement.

CIFAR [2][7][8] is perhaps the most well-known families of datasets for image classification machine learning research. The most popular variants are the original CIFAR 100 and its predecessor CIFAR 10. CIFAR 10 [12] consists of 60000 32x32 color images divided equally in 10 classes. CIFAR 100 [13] while having the same number of samples and the same resolution as its predecessor, greatly increases the number of classes to 100. The 100 classes are equally divided to 20 super classes providing more general categorization. For example, the classes beaver, dolphin, otter, seal, and whale are part of the superclass aquatic mammals. Caltech 256, much like CIFAR 100, is a direct increase on Caltech 101's size and complexity. It contains 30,607 color images of different sizes, categorized unequally in 256 object classes and one additional clutter class, with each class containing at least 80 samples [10].

MNIST is a very popular education and benchmarking dataset comprised of 70.000 images of handwritten digits. As a modified subset of NIST special databases 3 and 1, MNIST samples have been subjected to several normalization techniques in order to achieve a unified 29x29 grayscale resolution [15]. The most interesting aspect of this dataset is the amount of variants that share its name. The largest variant of MNIST is Fashion MNIST [14], a dataset of the same size and sample resolution as the original, meant to be a direct replacement of the original by its creator. The images in Fashion MNIST are of articles of clothing and the classes (remaining 9 as the digit classes of the original MNIST) are general clothing categories such as bag or trouser. MRDBI, or MNIST with Rotated Digits plus Background Images, dramatically increases the complexity of the original dataset in order to challenge the training models. Apart from the changes in the images indicated by the name, another factor of increased difficulty is the unusual split of the data to 12.000 training samples and 50.000 testing samples [2]. Other less known variants of MNIST include Chinese MNIST [11], which is comprised of 15.000 64x64 images of hand-drawn Chinese numbers, and sign language MNIST [16], a collection of 34627 24x24 grayscale samples representing hand signs categorized in 25 classes, one for each letter with no cases for 9=J and 25=Z since they require motion gesturing. The variants of MNIST are exceptionally different from the original; however, most generalized datasets like CIFAR contain a very large number of variations created by slightly altering the images (i.e., flipping, mirroring, adding noise or distortion and even altering the colors). These techniques allow for exponentially increasing the number of samples in a dataset, or training models in specific ways [1][9].

ImageNet [4] was the leading competition in image classification with machine learning. Despite the discontinuation of the competition, the benchmark datasets used to facilitate it remain and are still widely used. Being perhaps the largest collection of labeled images, ImageNet contains over 15 million labeled high resolution pictures belonging to approximately 22,000 categories. Due to how vast the complete dataset is, there exist many subsets in order to make usage more accessible. One of those datasets is the one used in the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC). It contains approximately 1000 images in each category for a 1.2 million total training images, 50,000 validation images, and 150,000 testing images. Due to the increasing rate of production for state-of-the-art models, there is a high demand for more streamlined and robust benchmarking solutions. Researchers [5][6][17][18] create methodologies and datasets that are meant to be used for the comparison of the performance for machine learning models for image classification.

The most commonly used metrics for the calculation of the performance of a machine learning algorithm are the average accuracy, the top-k error, and the confusion matrix. The average accuracy is the number of correct predictions divided by the number of samples, it can be represented both as a fraction and as a percentile, though the latter is most common. The top-k error is a score that increases when the correct label is included in the top k predictions of the model based on probability. This means that for a top-1 error we will have an increase in score only if the top predicted class of the model is the correct one, while for a top-5 error it suffices for the correct label to be one of the top 5 predicted. Clearly, the larger the k, the easier it becomes to score high, and the more meaningless the score becomes. The confusion matrix is not only useful as a metric for performance but also as a visualization. It is a table where one dimension is all the actual labels, and the other is all the predictions of the model. For every sample, an overlap between the label on x and the label on y is a correct prediction. When constructing a confusion matrix, a diagonal is created that contains the overlaps, and thus the correct predictions. This allows us to easily see not only how many mistakes our algorithm made, but also allows us to identify clusters of incorrect predictions that might be caused by relations inside our data [2].

1.1. References

[1] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," arXiv:1512.03385 [cs], Dec. 2015, Accessed: May 09, 2022. [Online]. Available: http://arxiv.org/abs/1512.03385

[2] H. Iba and N. Noman, Deep Neural Evolution: Deep Learning with Evolutionary Computation. Springer Nature, 2020.

[3] N. Karthikeyan, "Review of Deep Transfer Learning Models for Image Classification," International Journal of Recent Contributions from Engineering, Science IT (iJES), vol. 10, no. 01, Art. no. 01, Mar. 2022, doi: 10.3991/ijes.v10i01.29783.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in Advances in Neural Information Processing Systems, 2012, vol. 25. Accessed: May 08, 2022. [Online].

[5] Z. Lin, J. Shi, D. Pathak, and D. Ramanan, "The CLEAR Benchmark: Continual LEArning on Real-World Imagery," p. 13, 2021.

[6] V. Lomonaco, CORe50. 2022. Accessed: May 08, 2022. [Online]. Available: https://github.com/vlomonaco/core50

[7] S. Sankar, A. Jain, R. Chellappa, and S.-N. Lim, "Regularizing deep networks using efficient layerwise adversarial training," May 2017.

[8] F. Wang et al., "Residual Attention Network for Image Classification," arXiv:1704.06904 [cs], Apr. 2017, Accessed: May 08, 2022. [Online]. Available: http://arxiv.org/abs/1704.06904

[9] M. D. Zeiler and R. Fergus, "Visualizing and Understanding Convolutional Networks," arXiv:1311.2901 [cs], Nov. 2013, Accessed: May 08, 2022. [Online]. Available: http://arxiv.org/abs/1311.2901

[10] "Caltech 256 Image Dataset." https://www.kaggle.com/jessicali9530/caltech256 (accessed May 09, 2022).

[11] "Chinese MNIST," Keggle. https://www.kaggle.com/gpreda/chinese-mnist (accessed May 08, 2022).

[12] "CIFAR-10 and CIFAR-100 datasets." https://www.cs.toronto.edu/ kriz/cifar.html (accessed May 10, 2022).

[13] "CIFAR-100 Python." https://www.kaggle.com/fedesoriano/cifar100 (accessed May 10, 2022).

[14] "Fashion MNIST." https://www.kaggle.com/zalando-research/fashionmnist (accessed May 10, 2022).

[15] "Papers with Code - MNIST Dataset." https://paperswithcode.com/dataset/mnist (accessed May 10, 2022).

[16] "Sign Language MNIST," Keggle. https://www.kaggle.com/datamunge/sign-language-mnist (accessed May 08, 2022).

[17] M. Caldeira, P. Martins, R. L. C. Costa, and P. Furtado, "Image Classification Benchmark (ICB)," Expert Systems with Applications, vol. 142, p. 112998, Mar. 2020, doi: 10.1016/j.eswa.2019.112998.

[18] T. Yellamraju, J. Hepp, and M. Boutin, "Benchmarks for Image Classification and Other High-dimensional Pattern Recognition Problems," arXiv:1806.05272 [cs,

stat], Jun. 2018, Accessed: May 10, 2022. [Online]. Available: http://arxiv.org/abs/1806.05272