Investigation of DNN Model Robustness Using Heterogeneous Datasets

Wen-Hung Liao, Yen-Ting Huang

Dept. of Computer Science, National Chengchi University, Taipei, TAIWAN
Pervasive Artificial Intelligence Research (PAIR) Labs, TAIWAN
Email : {whliao, ythuang}@nccu.edu.tw

Abstract

Deep learning frameworks have been successfully applied to tackle many challenging tasks in pattern recognition and computer vision thanks to its ability to automatically extract representative features from the training data. Such type of data-driven approach, however, is subject to the criticism of too much dependency on the training set. In this research, we attempt to investigate the validity of this statement: 'deep learning is only as good as its data' by evaluating the performance of deep learning models using heterogeneous data sets, in which distinct representations of the same source data are employed for training/testing. We have examined three cases: low-resolution image, severely compressed input and halftone image in this work. Our preliminary results indicate that such dependency indeed exists. Classifier performance drops considerably when the model is tested with modified or transformed input. The best outcomes are obtained when the model is trained with hybrid input.

HETEROGENEOUS DATA SETS

Definition: Multiple distinct representations of the "same" source

- **Heterogenerous Representation**
  - **Original vs. Low-Resolution Images**
  - **Original vs. Compressed Images**

Figure 1:(a) Original (b) 60% compressed (c) 80% compressed images.

  - **Original vs. Halftone Images**

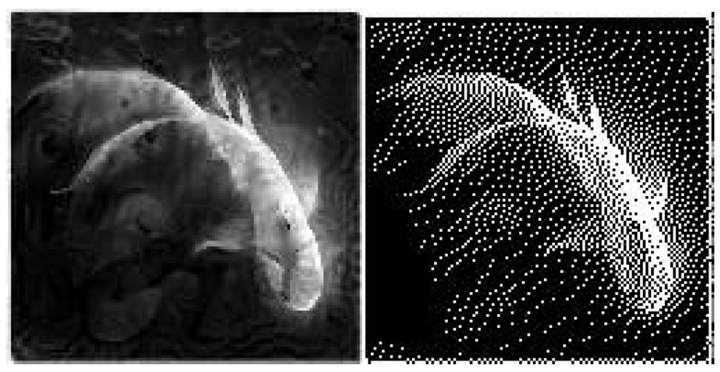

Figure 2:(a) Original vs. (b) halftone images using Floyd-Steinberg dithering.

- **Training with Heterogeneous Data**
  - **Hybrid Training:** Train the model with both data sets simultaneously (data augmentation perspective)
  - **Feature Concatenation:** Train the network with two branches and merge the extracted features (feature fusion perspective)
  - **The Order of Training:** Investigate the effect of continual learning by training the model with a pre-arranged order (incremental learning perspective)

Experimental Results

- **Resolution Change**

Training		Original	4x ↓	9x ↓	Hybrid - Original + 4x ↓	Hybrid - Original + 9x ↓
Evaluation (Acc.)	Original	0.61	0.02	0.01	0.62	0.57
	4x ↓	0.17	0.57	0.28	0.60	0.47
	9x ↓	0.08	0.34	0.51	0.35	0.53

Table 1:Accuracy using original vs. low-resolution inputs

- **Compression Ratio Change**

Training		Original	60% ↓	80% ↓	Hybrid - Original + 60% ↓	Hybrid - Original + 80% ↓
Evaluation (Acc.)	Original	0.61	0.47	0.17	0.58	0.53
	60% ↓	0.36	0.52	0.24	0.53	0.48
	80% ↓	0.03	0.05	0.32	0.09	0.39

Table 2:Accuracy using original vs. compressed input

- **Original vs. Halftone Images**

Training		Grayscale	FS Halftone	Feature Concatenation	Hybrid Training
Evaluation (Acc.)	Grayscale	0.61 (0.77)	0.42 (0.64)	0.45 (0.69)	0.66 (0.85)
	FS Halftone	0.01 (0.03)	0.61 (0.80)	0.48 (0.71)	0.59 (0.82)

Table 3:Accuracy using grayscale vs. halftone input

- **The Effect of the Order of Training**

Training Order		4x then Original	Original then 4x	9x then Original	Original then 9x
Evaluation (Acc.)	Original	0.60	0.02*	0.56	0.01*
	4x ↓	0.22	0.55	0.29	0.27
	9x ↓	0.07	0.36	0.11	0.51

Table 4:How the order of learning affects accuracy for resolution change experiment (*:catastrophic forgetting)

- **Hybrid Training**

Training		Hybrid
Evaluation (Acc.)	Original	0.59
	4x ↓	0.57
	9x ↓	0.54
	60% ↓	0.55
	80% ↓	0.40

Table 5:Accuracy using hybrid training

Acknowledgements

This work was partially supported by The Ministry of Science and Technology, Taiwan, under GRANT No. MOST108-2221-E-004-008 and MOST109-2634-F-004-001 through Pervasive Artificial Intelligence Research (PAIR) Labs.