

Impact of Data Quality on the Training Efficiency of a Convolutional Neural Network*

Marta BISTRONŃ, Michał PACZKOWSKI, Igor PAWLAK, Piotr ŚWIDZIŃSKI,
Monika CIERPICKA and Zbigniew PIOTROWSKI

Military University of Technology, gen. Sylwestra Kaliskiego 2, 00-902 Warsaw, Poland

Correspondence should be addressed to: Marta BISTRONŃ; marta.bistron@wat.edu.pl

* Presented at the 40th IBIMA International Conference, 23-24 November 2022, Seville, Spain

Copyright © 2022. Marta BISTRONŃ, Michał PACZKOWSKI, Igor PAWLAK, Piotr ŚWIDZIŃSKI,
Monika CIERPICKA and Zbigniew PIOTROWSKI

Abstract

This paper examines the impact of the size and quality of the learning dataset on the training efficiency of a convolutional network model. The transfer learning technique is characterized, and the results obtained for models trained from scratch for 40 and 100 epochs are compared with those obtained by transfer models built on 4 popular convolutional neural network architectures: AlexNet, VGG-16, ResNet-50, SqueezeNet. The impact of the quality of the dataset on the evaluation results of the selected model was also analyzed. To this end, neural network training using transfer learning was performed on 3 datasets: a baseline dataset, a dataset reduced by incorrect examples and a dataset reduced and then extended by additional correct examples.

Keywords: CNN, digital image classification, data selection, transfer learning

