Bayesian Convolutional Neural Networks for Image Classification with Uncertainty Estimation

Fatma Z. Bessai-Mechmache
Research Centre on Scientific and Technical Information, CERIST
Algiers, Algeria
fatmazohrabm@gmail.com

Maya N. Ghaffar
Department of Mathematics and Computer Science, University of Algiers 1
Algiers, Algeria
mayanghaffar@gmail.com

Rayan Y. Laouti
Department of Mathematics and Computer Science, University of Algiers 1
Algiers, Algeria
l.rayaneyasmine@gmail.com

***Abstract*—Over the past decade, deep learning has led to a cutting-edge performance in a variety of fields. However, it faces a fundamental constraint which is the treatment of uncertainty. The representation of the model's uncertainty is of significant importance in areas subject to strict safety or reliability requirements. Bayesian deep learning offers a new approach that showcases the degree of reliability of predictions made by neural networks. The present work tests deep learning with Bayesian thinking through a case study of image classification. It puts into practice Bayesian inference to tackle the problem of uncertainty in deep learning and shows its correlation with data quality and model accuracy. To reach this goal, we have implemented a Bayesian convolutional neural network using the variational inference algorithm, Bayes by Backprop. The proposed model was evaluated on an image classification task, with two benchmark datasets. The results' review allowed for validation of the Bayesian approach and showed that it obtains comparable results to those of a non-Bayesian convolutional neural network. Furthermore, the uncertainty of the model was estimated in terms of aleatory and epistemic uncertainty.**

***Index Terms*—bayesian deep learning, convolutional neural networks, uncertainty estimation, image classification, bayes by backprop.**

I. Introduction

In recent years, deep learning has led to remarkable success in several fields of artificial intelligence, including computer vision [1], speech recognition [2], and natural language processing [3]. Despite the success of standard deep learning methods in solving various real-world problems, deep neural networks are potentially prone to overfitting and thus require weight regularization measures and similar techniques to reduce it. On the other hand, these networks cannot provide information about the reliability of their predictions. In fact, deep neural networks make point estimates, without taking uncertainty into account [4]. In safety-critical systems, such as self-driving cars and medical diagnoses, being able to tell whether a model is certain of its output can be as important as the accuracy of the prediction. The use of such models often leads to overconfident predictions, and therefore less reliable decisions which can have disastrous consequences.

On May 7, 2016, an assisted driving vehicle collided with a tractor-trailer crossing an uncontrolled intersection on a highway west of Williston, Florida, resulting in the death of the driver. Data obtained from the vehicle indicated that the perception system mistook the white side of the trailer for a bright sky [5]. In a second example, a Google image classification system misidentified two African-Americans as gorillas [6], raising concerns about racial discrimination. Such malfunctions can be avoided if we correctly estimate the uncertainty of the machine learning system.

Over the last few years, Bayesian deep learning (BDL) has emerged as a probabilistic framework that closely integrates deep learning and probability theory to address these problems. BDL offers a solid approach based on the principle of uncertainty that provides an accurate measure of confidence in deep neural networks' predictions.

The present work proposes a Bayesian deep learning approach for image classification using convolutional neural networks (CNNs) based on the Bayes by Backprop algorithm and uncertainty estimation. It also offers a comparison with the standard approach through experiments and results.

This paper is organized as follows: Section II presents related works and the paper's contribution. Section III is devoted to the description of the proposed Bayesian deep learning approach. Section IV gives experimental results and shows the effectiveness of the Bayesian convolutional neural networks. Section V concludes the paper.

II. Related Works

While the number of theoretical contributions to Bayesian deep learning is still growing, the practical applications of Bayesian methods are quite limited. This is mainly due to the lack of efficient algorithms to overcome the computational difficulties imposed by Bayesian inference methods, as well as the lack of understanding of recent contributions to tackle these challenges.

Deodato et al. [7] implemented Bayesian neural networks with variational inference where they estimated the uncertainty of the predicted class labels by calculating a confidence score based on the predictive variance. Their Bayesian approach was tested on MNIST and EMNIST and then applied to a biomedical image dataset. Through their uncertainty analysis, they showed that one can achieve better accuracy by only relying on predictions with high confidence, and identified images as out-of-distribution samples although they shared the same ground-truth labels.

2022 3rd International Conference on Embedded & Distributed Systems (EDiS) | 979-8-3503-9754-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/EDIS57230.2022.9996478