Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Imperial College London

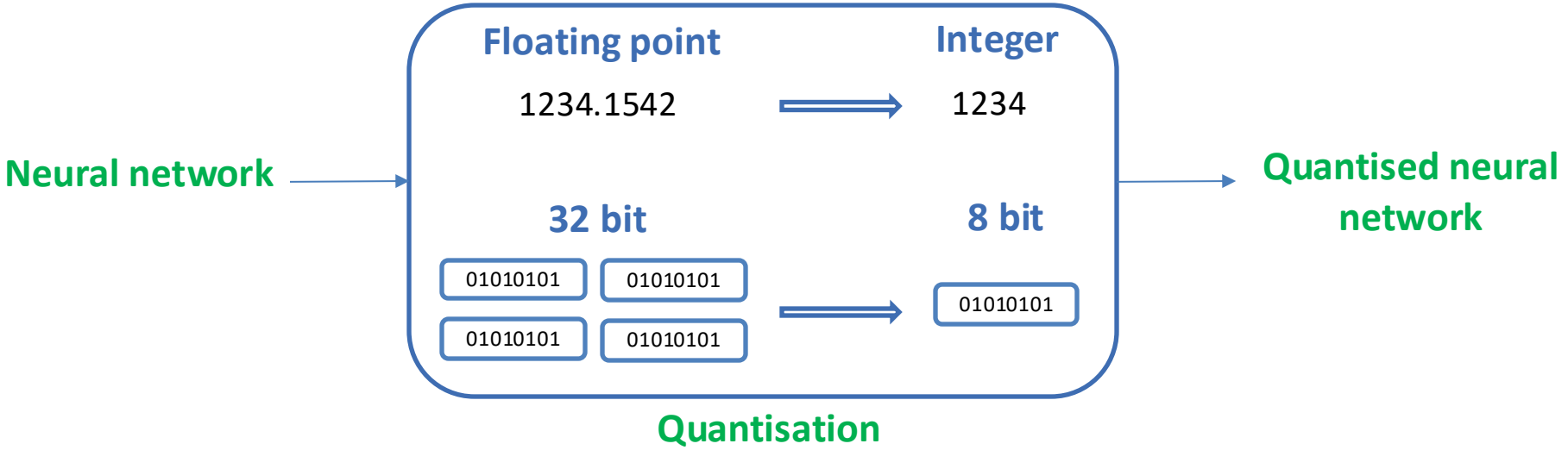
Certifiably Quantisation-Robust Training and Inference of Neural Networks

Hue Dang, Matthew Robert Wicker, Goetz Botterweck, Andrea Patane

Overview

1. **Introduction**
2. **Methods**
3. **Experiments**
4. **Discussion**

Introduction



Formal guarantees on the behavior of quantised models?

Introduction

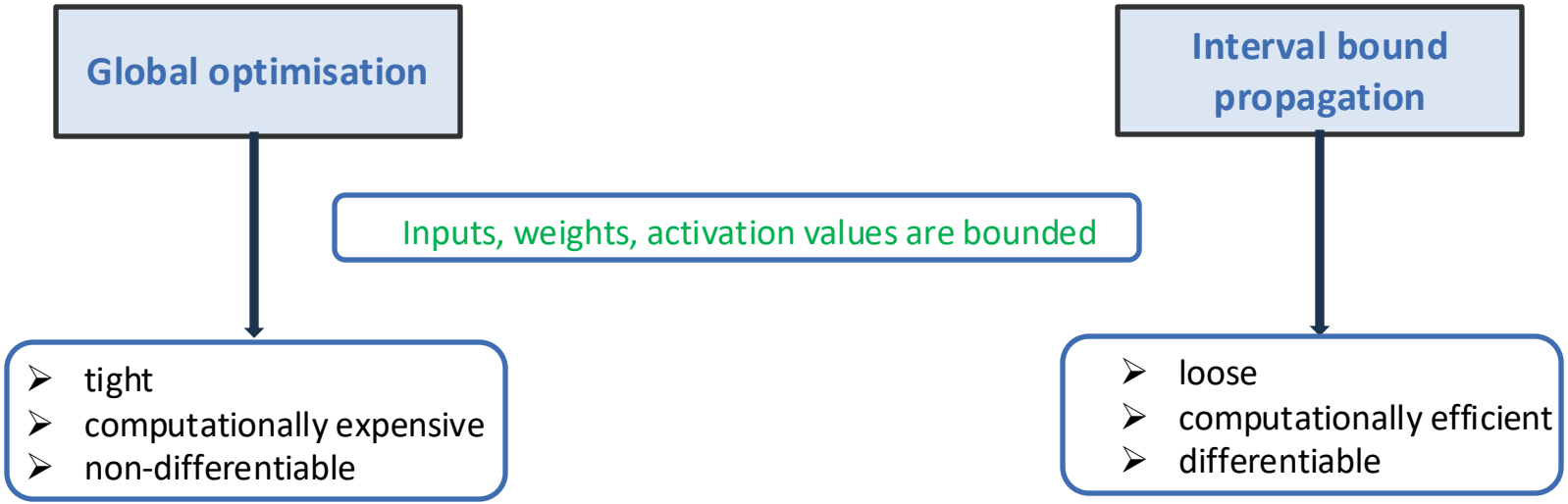
What are our contributions?

- ✓ Verification across *all possible quantised networks*, independent of specific schemes
- ✓ *Differentiable bounds* that enable both verification and robust training
- ✓ Formal guarantees for quantisation-robust neural networks

Methods

Verify all possible quantised networks, independent of specific

Find the worst−case scenario of each output dimension
Min $\hat{y}_i$ and Max $\hat{y}_i$



Methods

Differentiable bound Propagation for Quantisation-Robust Training

$$z^{(0),\mu}(x) = z^{(0)}(x)$$
$$z^{(0),r}(x) = \epsilon_{in}$$
$$h^{(k),\mu}(x) = W^{(k)} z^{(k-1),\mu}(x) + b^{(k)}$$
$$h^{(k),r}(x) = |W^{(k)}| z^{(k-1),r}(x) + W^{(k),\epsilon_w} |z^{(k-1),\mu}(x)| + W^{(k),\epsilon_w} z^{(k-1),r}(x) + b^{(k),\epsilon_w}$$

- Over-approximation of matrix multiplication between weights and layer inputs
- Differentiable bound
- In-training integration

Experiments

Training a neural network with differentiable bound propagation technique improves its robustness

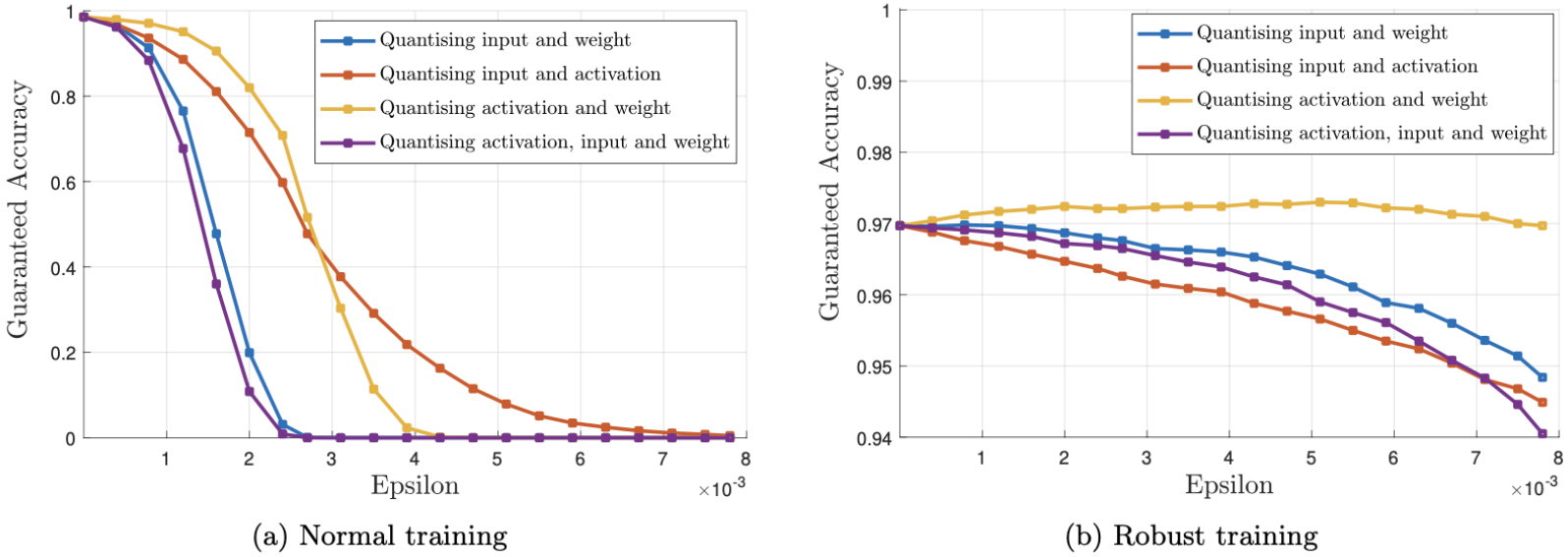


*Figure 1: Guaranteed accuracy computed by **IBP** method of the model (a) trained normally and (b) trained robustly*

Experiments

Verification of all possible quantised networks parameterised by 6/8/10 bit quantisation diameters

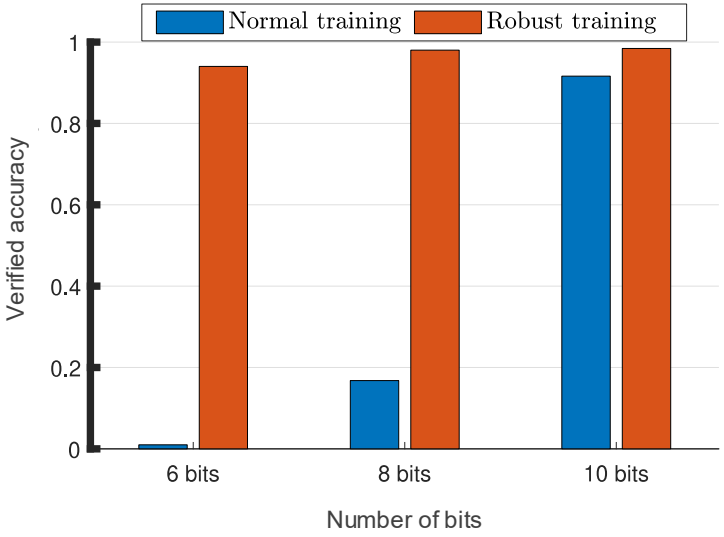


Figure 2: Comparison of normal training and robust training on verified accuracy computed by ***bilinear optimization method***

Experiments

Differentiable bound trained models are only robust but also easier to verify

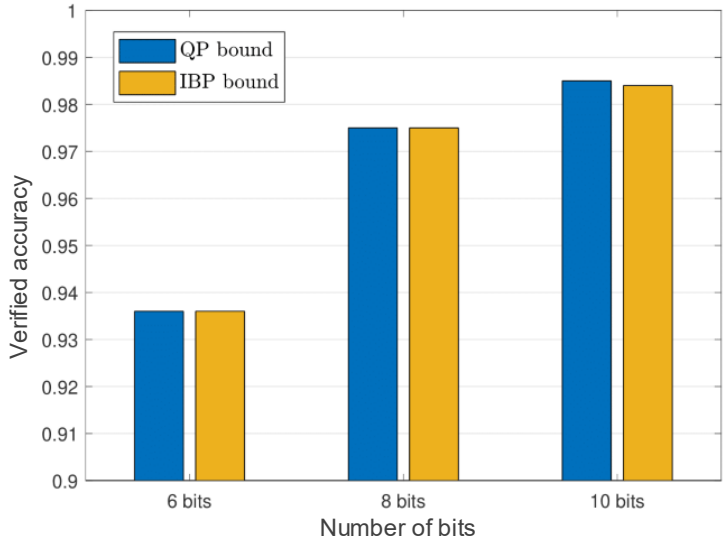


*Figure 3: Verified accuracy computed by IBP and BP bounds on **models trained via differentiable bound propagation technique***

Discussion

- Proposing formal verification and training methods for the robustness of neural networks against quantisation of their inputs, parameters and activation values

- The scalability of the verification and training methods