BiPedalNet

Binarized Neural Networks for Resource-Constrained On-Device Gait Identification

Daniel J. Wu[1] Avoy Datta[1] Vinay Prabhu[2]

[1]Stanford University

[2]UnifyID Inc.

Problem Statement

We are interested in **on-device user identification** using phone-collected gait acceleometry, which has applications in:

- Account security
- Two-factor authentication
- Theft detection
- Login convenience
- User privacy

Problem Setting

Practical limitations in developing countries include:

- Low processing power on mobile devices
- Network quality issues
- Limited device memory

Previous Work

Previous work[1] on user re-identification from gait data demonstrated high accuracy with deep CNNs, but also high computational overhead, making on-device, real-time, inference, intractable. We propose the usage of binarized neural networks to make on-device inference feasible, and in this work, we compare our method to IDNet on the 50-user Padova dataset.

[1] Matteo Gadaleta and Michele Rossi. "IDNet: Smartphone-based gait recognition with convolutional neural networks". In: *Pattern Recognition* 74 (2018), pp. 25–37. ISSN: 0031-3203. DOI: https://doi.org/10.1016/j.patcog.2017.09.005. URL: http://www.sciencedirect.com/science/article/pii/S0031320317303485.

Binarized Neural Networks

BNNs[2]

- have weights constrained to $\{-1, 1\}$.
- generally achieving lower accuracies than their full-precision equivalents.
- are an order of magnitude smaller, due to compact binary weight matrices.
- are faster, due to the usage of bitwise operations for matrix multiplication and activation functions.

Our binarized architecture is implemented with the LARQ[3] library.

[2] Matthieu Courbariaux et al. "Binarized neural networks: Training deep neural networks with weights and activations constrained to+ 1 or-1". In: *arXiv preprint arXiv:1602.02830* (2016).

[3] Lukas Geiger et al. *Larq: An Open-Source Deep Learning Library for Training Binarized Neural Networks*. Web page. 2019. URL: `https://larq.dev`.

Architecture

In addition to the binarized version of the original IDNet, we propose a lightweight, binarized neural network architecture, dubbed BiPedalNet:

Layer	Units	Filter Size	# 1-bit params	# 32-bit params
Conv2D	32	(3,3)	288	0
MaxPool2D	-	(2,2)	0	0
BatchNorm	-	-	0	96
Conv1D	64	(1,3)	6144	0
BatchNorm	-	-	0	192
Conv1D	64	(1,3)	12288	0
BatchNorm	-	-	0	192
Flatten	-	-	-	-
Dense	32	-	276480	0
BatchNorm	-	-	0	96
Dense	38	-	1216	0
Softmax	-	-	-	-

The BiPedalNet Architecture

Training

During training, we maintain latent real-valued weights that are updated by backpropagation. In the forward pass, these latent weights are set to 1 if positive; otherwise, clipped to -1. Note that these latent weights are not used at test time, and thus don't contribute to the memory usage of the deployed model.

We use the Adam optimizer[4] to update the latent weights of the architecture, with a default learning rate of 1e-3 and a scheduler that exponentially reduces the rate when validation accuracy plateaus. The multi-class cross-entropy metric is used as the optimization objective.

[4] Diederik Kingma and Jimmy Ba. "Adam: A Method for Stochastic Optimization". In: *International Conference on Learning Representations* (Dec. 2014).

Results

- Naive binarization of IDNet performed poorly.
- The hand-designed BiPedalNet achieved comparable performance to the full-precision IDNet.

Model	Number of Parameters	Size on disk (MB)	Top-1 Accuracy(%)
IDNet	335k	1.28	98.05
Binarized IDNet	335k	0.04	41.45
BiPedalNet	297k	0.04	**95.91**

Conclusions

- BNNs are well suited for low-resource machine learning.
- BNNs can achieve comparable accuracies to full precision architectures, at a fraction of the size, on gait identification tasks.
- Custom-designing binarized architectures is necessary for reasonable performance – naively binarizing a given full precision architecture is a recipe for disaster.

References

Courbariaux, Matthieu et al. "Binarized neural networks: Training deep neural networks with weights and activations constrained to+ 1 or-1". In: *arXiv preprint arXiv:1602.02830* (2016).

Gadaleta, Matteo and Michele Rossi. "IDNet: Smartphone-based gait recognition with convolutional neural networks". In: *Pattern Recognition* 74 (2018), pp. 25–37. ISSN: 0031-3203. DOI: https://doi.org/10.1016/j.patcog.2017.09.005. URL: http://www.sciencedirect.com/science/article/pii/S0031320317303485.

Geiger, Lukas et al. *Larq: An Open-Source Deep Learning Library for Training Binarized Neural Networks*. Web page. 2019. URL: https://larq.dev.

Kingma, Diederik and Jimmy Ba. "Adam: A Method for Stochastic Optimization". In: *International Conference on Learning Representations* (Dec. 2014).