1.	Record Nr.	UNINA9910337657903321
	Autore	Moons Bert
	Titolo	Embedded Deep Learning : Algorithms, Architectures and Circuits for Always-on Neural Network Processing / / by Bert Moons, Daniel Bankman, Marian Verhelst
	Pubbl/distr/stampa	Cham : , , : Springer International Publishing : , , : Imprint : Springer, , 2019
	ISBN	3-319-99223-6
	Edizione	[1st ed. 2019.]
	Descrizione fisica	1 online resource (216 pages)
	Disciplina	370.285
	Soggetti	Electronic circuits Signal processing Image processing Speech processing systems Electronics Microelectronics Circuits and Systems Signal, Image and Speech Processing Electronics and Microelectronics, Instrumentation
	Lingua di pubblicazione	Inglese
	Formato	Materiale a stampa
	Livello bibliografico	Monografia
	Nota di contenuto	Chapter 1 Embedded Deep Neural Networks -- Chapter 2 Optimized Hierarchical Cascaded Processing -- Chapter 3 Hardware-Algorithm Co-optimizations -- Chapter 4 Circuit Techniques for Approximate Computing -- Chapter 5 ENVISION: Energy-Scalable Sparse Convolutional Neural Network Processing -- Chapter 6 BINAREYE: Digital and Mixed-signal Always-on Binary Neural Network Processing -- Chapter 7 Conclusions, contributions and future work.
	Sommario/riassunto	This book covers algorithmic and hardware implementation techniques to enable embedded deep learning. The authors describe synergetic design approaches on the application-, algorithmic-, computer architecture-, and circuit-level that will help in achieving the goal of reducing the computational cost of deep learning algorithms. The impact of these techniques is displayed in four silicon prototypes for

embedded deep learning. Gives a wide overview of a series of effective solutions for energy-efficient neural networks on battery constrained wearable devices; Discusses the optimization of neural networks for embedded deployment on all levels of the design hierarchy – applications, algorithms, hardware architectures, and circuits – supported by real silicon prototypes; Elaborates on how to design efficient Convolutional Neural Network processors, exploiting parallelism and data-reuse, sparse operations, and low-precision computations; Supports the introduced theory and design concepts by four real silicon prototypes. The physical realization's implementation and achieved performances are discussed elaborately to illustrated and highlight the introduced cross-layer design concepts.