Towards Low-bit Quantization of Deep Neural Networks with Limited Data

Yong Yuan, Chen Chen, Xiyuan Hu and Silong Peng
Institute of Automation, Chinese Academy of Sciences

CVPL
Associazione Italiana per la ricerca in Computer Vision, Pattern recognition e machine Learning (CVPL- ex-GIRPR)

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA
Centro Interdipartimentale di Ricerca Industriale SOFTECH: ICT per le Imprese

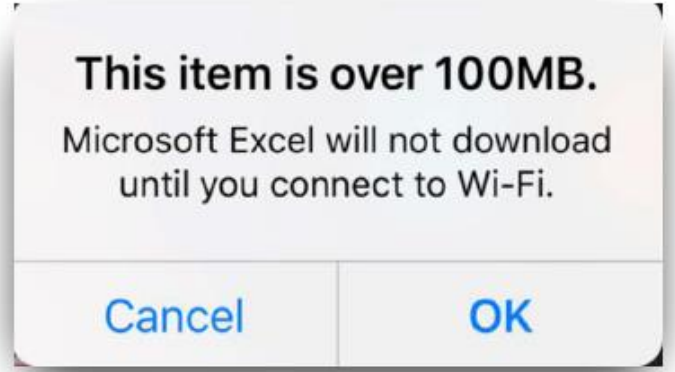
UNIVERSITÀ DEGLI STUDI FIRENZE
MICC
Centro per la Comunicazione e l'Integrazione dei Media

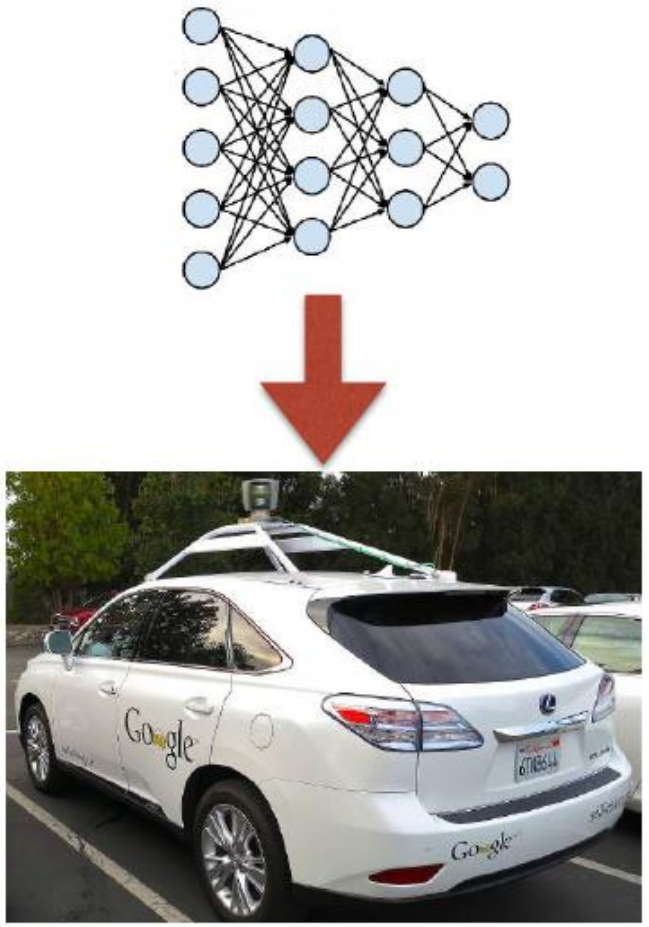
cini
consorzio interuniversitario nazionale per l'informatica

IAPR

Technically Co-Sponsored by
IEEE
IEEE COMPUTER SOCIETY

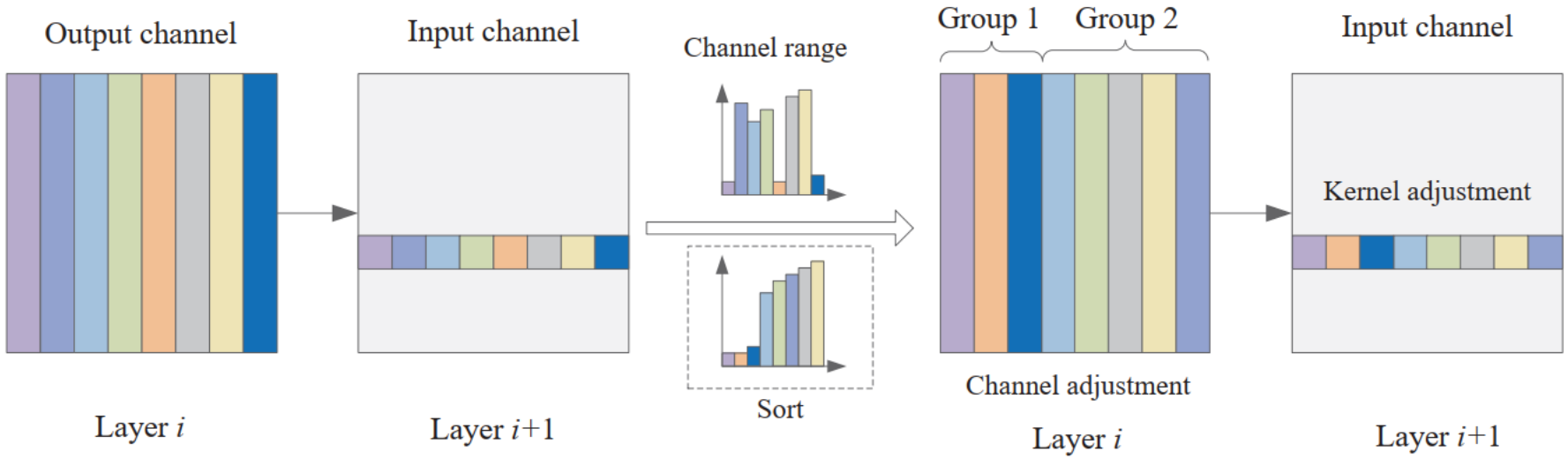
Motivation


This item is over 100MB.
Microsoft Excel will not download until you connect to Wi-Fi.
Cancel
OK

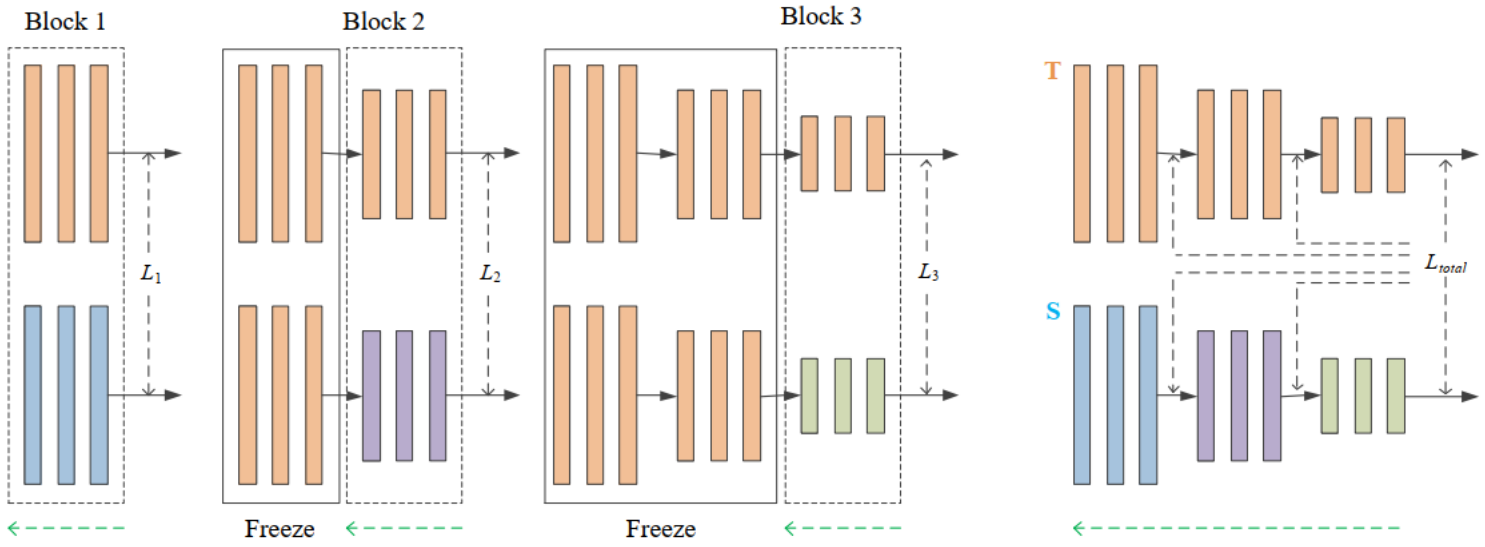

Google

Proposed Method

- **Group-wise quantization**



- # Two-stage knowledge distillation



(a) Independent optimization stage

(b) Joint optimization stage

Experiments

	W-bit	Channel	Layer	Equalization [25]	Group			
					2	4	8	16
ResNet18	4	55.50	0.20	0.25	12.95	31.59	46.29	52.28
	6	69.05	66.90	66.36	68.43	68.53	69.04	69.25
ResNet50	4	70.07	0.12	0.32	5.80	61.33	67.00	68.15
	6	75.86	72.51	73.26	75.27	75.61	75.71	75.70
ShuffleNetV2	4	24.75	0.10	0.44	0.42	5.00	9.92	15.14
	6	67.51	0.12	60.36	60.96	66.37	67.05	67.12
	8	68.89	62.84	68.65	68.89	68.97	69.84	68.80
MobileNetV2	4	13.38	0.10	0.19	0.43	1.23	4.37	8.24
	6	59.03	0.10	33.41	1.03	61.62	58.74	60.82
	8	69.03	0.10	68.90	65.96	68.14	69.57	69.17

	Original	W-bit	Channel				8-Group			
			Init	GS	BS	**Ours**	Init	GS	BS	**Ours**
ResNet18	69.76	4	55.50	67.80	67.93	**68.08**	46.29	67.41	67.57	**67.72**
ResNet50	76.15	4	70.07	74.48	74.75	**74.98**	67.00	74.11	74.36	**74.75**
ShuffleNetV2	69.36	4	24.75	62.79	62.79	**63.65**	10.91	60.84	60.25	**61.05**
MobileNetV2	71.88	4	13.38	64.45	64.89	**65.74**	4.37	64.26	64.50	**65.30**

Thanks for listening