

Motivation, Constraints & Task

<u>Motivation</u>: \sim \$3 billion annual vehicle repair bills related to pothole damage in the U.S.

<u>Constraints</u>: limited computational power (e.g., no GPU) on edge devices installed on the road inspection vehicles; need fast inference speed.

Task: Given a fixed number of training epochs and a lightweight model to be trained, what can practitioners do to improve the pothole classification accuracy?

Contributions

- We propose Iterative Self Knowledge Distillation (ISKD), which outperforms the SOTA self **KD** methods from pothole classification (RDD, simplex, complex) to generic (CIFAR-10, CIFAR-100), fine-grained (Oxford 102 Flower, Oxford IIIT Pet, Caltech-UCSD Birds 200), and medical imaging classification (COVID-19 Radiography).
- We provide more evidence showing that a teacher model with accuracy lower than the baseline can still result in a student model outperforming the baseline.
- ISKD is flexible with respect to parameter selection.

Iterative Self Knowledge Distillation



 S_k 's total loss function:

 $L_{KD} = (1 - \alpha) L_c + \alpha KLD (\mathbf{z}, \mathbf{z}^t)$

KLD: KL divergence; $\mathbf{z}^t / \mathbf{z}$: output probability distribution of T_k/S_k ; α : weight of KLD.

Iterative Self Knowledge Distillation – From Pothole Classification to Fine-Grained and COVID Recognition Kuan-Chuan Peng Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA, USA

kpeng@merl.com

Datas

SLICS C		antita	itive	Res	SUITS					
RDD	simplex	comp	lex CII	FAR-10	CIFAR-1	100 Oz	xford-102	Oxford-	37 CUB-200	COVID
es 8511	4736	748	9 5	0000	50000)	6552	3680	3000	14815
es 2140	650	604	- 1	0000	10000)	818	3669	3033	6350
2	2	2		10	100		102	37	200	4
pothe	ole / norma	l road ima	ges o	bjects	object	S	flowers	cats & do	ogs birds	chest x-ray
infr	astructure	maintenan	ce	generic o	classificati	on	fine-g	rained class	sification	types of pneumoni
				datas	et stat	istics				
experin	nent ID	E_1	E_2	E_3	E_4	E_5	E_6	E_7	E_8	E_9
model $\setminus K$	D iteration	$i_1 \pmod{\mathrm{KD}}$	i_2	i_3	i_4	i_5	i_6	$i_1 \sim i_6$	i_1 (large-epoch)	$i_1 + \text{Tf-KD}_{self}$ [2]
ResNet	-18 [23]	91.54 ₅₀	92.7150	92.99 ₅₀	93.04 ₅₀	93.08 50	n/a	93.08 ₂₅₀	92.24250	92.34 ₂₅₀
SqueezeNe	et v1.1 [24]	89.67 ₅₀	89.9150	90.28_{50}	90.51_{50}	90.70 50	90.70 50	90.70 300	90.47300	90.28300
ShuffleNet	v2 x0.5 [25]	90.05_{50}	90.14_{50}	90.98_{50}	91.40_{50}	91.40_{50}	n/a	91.40_{250}	92.66 ₂₅₀	91.22_{250}
ShuffleNet	v2 x1.0 [25]	92.01_{50}	92.15 ₅₀	92.6650	93.22 ₅₀	93.22 ₅₀	n/a	93.22 ₂₅₀	93.13_{250}	93.13 ₂₅₀
ResNet	-18 [23]	81.85_{50}	90.4650	93.69 ₅₀	96.92 ₅₀	98.62 ₅₀	99.08 50	99.08 300	83.38300	92.00300
SqueezeNe	et v1.1 [24]	81.85_{50}	86.7750	87.6950	88.15 50	88.15 50	n/a	88.15_{250}	84.62_{250}	87.69_{250}
ShuffleNet	v2 x0.5 [25]	90.00_{50}	95.69 ₅₀	99.38_{50}	100.00_{50}	n/a	n/a	100.00_{200}	92.46200	97.54_{200}
ShuffleNet	v2 x1.0 [25]	90.31_{50}	96.3150	98.77_{50}	100.00_{50}	n/a	n/a	$\boldsymbol{100.00}_{200}$	93.38_{200}	97.54_{200}
ResNet	-18 [23]	61.92 ₅₀	74.14 ₅₀	77.65 ₅₀	79.80 ₅₀	83.7750	84.93 ₅₀	84.93 300	62.58300	82.95300
SqueezeNe	et v1.1 [24]	59.27_{50}	70.7050	76.16 ₅₀	76.16 50	n/a	n/a	76.16 ₂₀₀	62.25_{200}	71.03_{200}
ShuffleNet	v2 x0.5 [25]	56.79_{50}	78.9750	88.58_{50}	88.91 50	n/a	n/a	88.91 ₂₀₀	65.89_{200}	79.80_{200}
ShuffleNet	v2 x1.0 [25]	60.43_{50}	78.3150	86.59 50	86.42_{50}	n/a	n/a	86.42_{200}	71.85_{200}	82.45_{200}
ResNet	-18 [23]	90.75_{50}	91.8150	92.0550	92.07 ₅₀	n/a	n/a	92.07 200	91.53200	92.01200
ShuffleNet	v2 x1.0 [25]	82.63 ₅₀	85.1650	88.45 ₅₀	90.30_{50}	91.03 ₅₀	91.58 50	91.58 ₃₀₀	90.68300	85.55300
ResNet	-18 [23]	80.15_{50}	81.0550	81.6450	82.1750	82.3050	82.67 50	82.67 300	81.34300	81.62300
CIFAR-100 [26] ShuffleNet v2	v2 x1.0 [25]	58.95_{50}	65.60 ₅₀	72.1650	75.59_{50}	77.27_{50}	77.85 ₅₀	77.85 ₃₀₀	77.61 ₃₀₀	65.78300
ResNet	-18 [23]	96.5850	97.3150	97.6850	97.80 50	n/a	n/a	97.80 200	96.94200	97.43200
ShuffleNet	v2 x1.0 [25]	94.74_{50}	97.19 ₅₀	98.17 ₅₀	98.41 ₅₀	n/a	n/a	98.41 ₂₀₀	98.41 ₂₀₀	97.19 ₂₀₀
ResNet	-18 [23]	90.57 ₅₀	90.9850	91.3350	91.80 50	91.5850	n/a	91.80 200	91.20200	91.31200
ShuffleNet	v2 x1.0 [25]	79.69_{50}	84.3050	85.96 ₅₀	86.59 ₅₀	86.59 ₅₀	n/a	86.59 ₂₅₀	86.10_{250}	84.46 ₂₅₀
ResNet	-18 [23]	42.0750	46.4950	48.6650	49.82 50	49.49 ₅₀	n/a	49.82 200	46.03200	47.58200
ShuffleNet	v2 x1.0 [25]	41.54_{50}	45.4050	48.5350	49.69 ₅₀	50.28 50	49.95 ₅₀	50.28 ₂₅₀	48.99_{250}	46.95_{250}
RecNet	-18 [23]	04 11-0	94 68-0	94 80	95 01-0		n/a	95 01.000	<u>94</u> 70 ₂₀₀	94 88200
ShuffleNet	$v^2 \times 10[23]$	90.76ro	92 43ro	93 07-0	93 81 ₅₀	93 98 ×0	n/a	93 98 050	9272_{050}	93 10oro
	ResNet ShuffleNet	RDDsimplexRDDsimplexes 8511 4736 s 2140 650 22pothole / normainfrastructuremodel \ KDKesNet-18[23]SqueezeNet v1.1ShuffleNet v2 x0.5ShuffleNet v2 x1.0[25]ShuffleNet v2 x1.0ShuffleNet v2 x0.5SqueezeNet v1.1[24]ShuffleNet v2 x0.5SqueezeNet v1.1[24]ShuffleNet v2 x0.5ShuffleNet v2 x0.5SqueezeNet v1.1[24]ShuffleNet v2 x1.0[25]ResNet-18[23]SqueezeNet v1.1[24]ShuffleNet v2 x1.0[25]ResNet-18[23]ShuffleNet v2 x1.0[25]ResNet-18[23]	RDD simplex complex RDD simplex complex es 8511 4736 748% s 2140 650 604 2 2 2 pothole / normal road ima infrastructure maintenam model \ KD iteration i_1 (no KD) ResNet-18 [23] 91.5450 SqueezeNet v1.1 [24] 89.6750 ShuffleNet v2 x0.5 [25] 90.0550 ShuffleNet v2 x1.0 [25] 92.0150 ResNet-18 [23] 81.8550 SqueezeNet v1.1 [24] 81.8550 SqueezeNet v1.1 [24] 81.8550 SqueezeNet v1.1 [24] 81.8550 SqueezeNet v1.1 [24] 59.2750 ShuffleNet v2 x0.5 [25] 56.7950 ShuffleNet v2 x1.0 [25] 90.3150 ResNet-18 [23] 90.7550 ShuffleNet v2 x1.0 [25] 56.7950 ShuffleNet v2 x1.0 [25] 56.7950 ShuffleNet v2 x1.0 [25] 58.9550 ShuffleNet v2 x1.0 [25] 90.7550 ShuffleNet v2 x1.0 [25] 90.5750	RDD simplex complex CII es 8511 4736 7489 5 s 2140 650 604 1 2 2 2 2 0 pothole / normal road images infrastructure maintenance 0 0 0 experiment ID E_1 E_2 0 0 geodel \ KD iteration i_1 (no KD) i_2 0 0 SqueezeNet v1.1 [24] 89.6750 89.9150 90.1450 ShuffleNet v2 x0.5 [25] 90.0550 90.1450 92.1550 ResNet-18 [23] 81.8550 86.7750 91.9450 SqueezeNet v1.1 [24] 81.8550 96.3150 96.3150 SqueezeNet v1.1 [24] 59.2750 70.7050 91.8150 ShuffleNet v2 x0.5 [25] 90.7550 91.8150 91.950 ShuffleNet v2 x1.0 [25] 56.7950 78.9750 91.8150 ShuffleNet v2 x1.0 [25] 56.7950 78.9750 91.8150 ShuffleNet v2 x1.0 [25] 58.9550	RDDsimplexcomplexCIFAR-10es85114736748950000s21406506041000022210pothole / normal road imagesobjectsinfrastructure maintenancegeneric ofdatasexperiment ID E_1 E_2 model \ KD iteration i_1 (no KD) i_2 infrastructure maintenance92.9715092.9950SqueezeNet v1.1 [24]89.675089.915090.2850ShuffleNet v2 x0.5 [25]90.055090.145090.8850ShuffleNet v2 x1.0 [25]92.015092.155092.6650ResNet-18 [23]81.855090.465093.6950SqueezeNet v1.1 [24]81.855086.775087.6950ShuffleNet v2 x0.5 [25]90.015096.315098.7750ResNet-18 [23]61.925074.145077.6550SqueezeNet v1.1 [24]59.275070.705076.1650SqueezeNet v1.1 [24]59.275078.975088.5850ShuffleNet v2 x1.0 [25]56.795078.975088.5850ShuffleNet v2 x1.0 [25]80.155081.055081.6450ShuffleNet v2 x1.0 [25]90.755091.815092.0550ShuffleNet v2 x1.0 [25]94.745097.195098.1750ResNet-18 [23]90.575090.985091.3350ShuffleNet v2 x1.0 [25]79.695084.305085.9650ShuffleNet v2 x1.0 [25]79.695084.305085.9650ResNet-18 [23]	RDD simplex complex CIFAR-10 CIFAR-10 es 8511 4736 7489 50000 50000 s 2140 650 604 10000 10000 pothole / normal road images objects object object infrastructure maintenance generic classificati experiment ID E_1 E_2 E_3 E_4 model \ KD iteration i_1 (no KD) i_2 i_3 i_4 ResNet-18 [23] 91.5450 92.7150 92.9950 93.0450 SqueezeNet v1.1 [24] 89.6750 89.9150 90.2850 91.4050 ShuffleNet v2 x0.5 [25] 90.0050 95.6950 93.2250 93.2250 ResNet-18 [23] 81.8550 90.4650 93.6950 96.9250 SqueezeNet v1.1 [24] 81.8550 90.4650 93.6950 96.9250 SqueezeNet v1.1 [24] 81.8550 90.4650 93.6950 96.9250 SqueezeNet v1.1 [24] 59.2750 70.7050 76.1650 <td< td=""><td>RDD simplex complex CIFAR-10CIFAR-10CIFAR-10CIFAR-100O28851147367489500005000022210100pothole / normal road imagesobjectsobjectsobjectsgeneric classificationdataset statisticsdataset statisticscase experiment ID$E_1$$E_2$$E_3$$E_4$$E_5$model \ KD iteration$i_1$ (no KD)$i_2$$i_3$$i_4$$i_5$SqueezeNet v1.1 [24]$89.6750$$99.9150$$90.2850$$90.5150$$90.7050$ShuffleNet v2 x1.0 [25]$92.0150$$92.1550$$92.650$$93.2250$$93.2250$ResNet-18 [23]$91.5450$$92.4650$$93.6950$$93.2250$$93.2250$ResNet-18 [23]$91.5450$$92.1550$$92.6950$$93.2250$$93.2250$ResNet-18 [23]$91.5450$$92.1550$$92.6950$$93.850$$100.0050$$n'a$SqueezeNet v1.1 [24]$81.8550$$90.4650$$93.6950$$88.1550$$88.1550$ShuffleNet v2 x1.0 [25]$90.0750$$74.1450$$77.6550$$79.8050$$83.7750$SqueezeNet v1.1 [24]$59.2750$$78.9750$$86.5950$$86.4250$$n'a$ShuffleNet v2 x1.0 [25]$50.750$$78.9750$$86.5950$<t< td=""><td>$\begin{array}{ c c c c c c c c c c c c c c c c c c c$</td><td>$\begin{array}{ c c c c c c c c c c c c c c c c c c c$</td><td>RDD simplex complex CIFAR-10 CIFAR-100 Oxford-102 Oxford-37 CUB-200 cs 8511 4736 7489 50000 50000 6552 3680 3000 s 2140 650 604 10000 1000 818 3669 3033 2 2 10 100 102 37 200 pothole / normal road images objects objects flowers cats & dogs birds infrastructure maintenance generic classification flowers cats & dogs birds gueezeNet v1.1 121 91.5450 92.7150 92.9930 93.0450 93.0850 n/a 93.0850 n/a 93.0850 n/a 93.0850 90.7050 90.70300 90.47300 90.47300 90.47300 90.47300 90.47300 90.47300 91.40250 93.2250 93.2250 93.2250 93.2250 93.2250 93.2250 93.2250 93.2250 93.2250 93.2250 93.2250 93.2250</td></t<></td></td<>	RDD simplex complex CIFAR-10CIFAR-10CIFAR-10CIFAR-100O28851147367489500005000022210100pothole / normal road imagesobjectsobjectsobjectsgeneric classificationdataset statisticsdataset statisticscase experiment ID E_1 E_2 E_3 E_4 E_5 model \ KD iteration i_1 (no KD) i_2 i_3 i_4 i_5 SqueezeNet v1.1 [24] 89.6750 99.9150 90.2850 90.5150 90.7050 ShuffleNet v2 x1.0 [25] 92.0150 92.1550 92.650 93.2250 93.2250 ResNet-18 [23] 91.5450 92.4650 93.6950 93.2250 93.2250 ResNet-18 [23] 91.5450 92.1550 92.6950 93.2250 93.2250 ResNet-18 [23] 91.5450 92.1550 92.6950 93.850 100.0050 $n'a$ SqueezeNet v1.1 [24] 81.8550 90.4650 93.6950 88.1550 88.1550 ShuffleNet v2 x1.0 [25] 90.0750 74.1450 77.6550 79.8050 83.7750 SqueezeNet v1.1 [24] 59.2750 78.9750 86.5950 86.4250 $n'a$ ShuffleNet v2 x1.0 [25] 50.750 78.9750 86.5950 <t< td=""><td>$\begin{array}{ c c c c c c c c c c c c c c c c c c c$</td><td>$\begin{array}{ c c c c c c c c c c c c c c c c c c c$</td><td>RDD simplex complex CIFAR-10 CIFAR-100 Oxford-102 Oxford-37 CUB-200 cs 8511 4736 7489 50000 50000 6552 3680 3000 s 2140 650 604 10000 1000 818 3669 3033 2 2 10 100 102 37 200 pothole / normal road images objects objects flowers cats & dogs birds infrastructure maintenance generic classification flowers cats & dogs birds gueezeNet v1.1 121 91.5450 92.7150 92.9930 93.0450 93.0850 n/a 93.0850 n/a 93.0850 n/a 93.0850 90.7050 90.70300 90.47300 90.47300 90.47300 90.47300 90.47300 90.47300 91.40250 93.2250 93.2250 93.2250 93.2250 93.2250 93.2250 93.2250 93.2250 93.2250 93.2250 93.2250 93.2250</td></t<>	$\begin{array}{ c c c c c c c c c c c c c c c c c c c$	$\begin{array}{ c c c c c c c c c c c c c c c c c c c$	RDD simplex complex CIFAR-10 CIFAR-100 Oxford-102 Oxford-37 CUB-200 cs 8511 4736 7489 50000 50000 6552 3680 3000 s 2140 650 604 10000 1000 818 3669 3033 2 2 10 100 102 37 200 pothole / normal road images objects objects flowers cats & dogs birds infrastructure maintenance generic classification flowers cats & dogs birds gueezeNet v1.1 121 91.5450 92.7150 92.9930 93.0450 93.0850 n/a 93.0850 n/a 93.0850 n/a 93.0850 90.7050 90.70300 90.47300 90.47300 90.47300 90.47300 90.47300 90.47300 91.40250 93.2250 93.2250 93.2250 93.2250 93.2250 93.2250 93.2250 93.2250 93.2250 93.2250 93.2250 93.2250

Table 1: Comparing the classification accuracy (%) of the iterative self knowledge distillation (KD) method versus the baselines. The numbers are in the format of $[accuracy]_{[e_s]}$, where e_s is the number of epochs which the student model is trained for.

Teacher-Student Accuracy Relation









nethod	ISKD	prior work (backbone)
0 [26]	82.67	81.60 [27] (Wide-ResNet-28-10)
2 [17]	97.80	91.10 [28] (ResNet152-SAM)
7 [18]	91.80	91.60 [28] (ResNet50-SAM)