



Introduction

Knowledge distillation^[1] has achieved remarkable results in supervised learning.

However, there are two major problems with existing knowledge distillation methods.

Teacher's supervision is sometimes **misleading**.

Student's prediction is **not accurate** enough.

To address the first issues, we apply **knowledge** adjustment to correct teachers' supervision using ground truth.

For the second problem, we use the **selective** classification framework^[3] to train the student model. In particular, the deep gambler loss^[2] is adopted to predict with reservation by introducing the extra class.

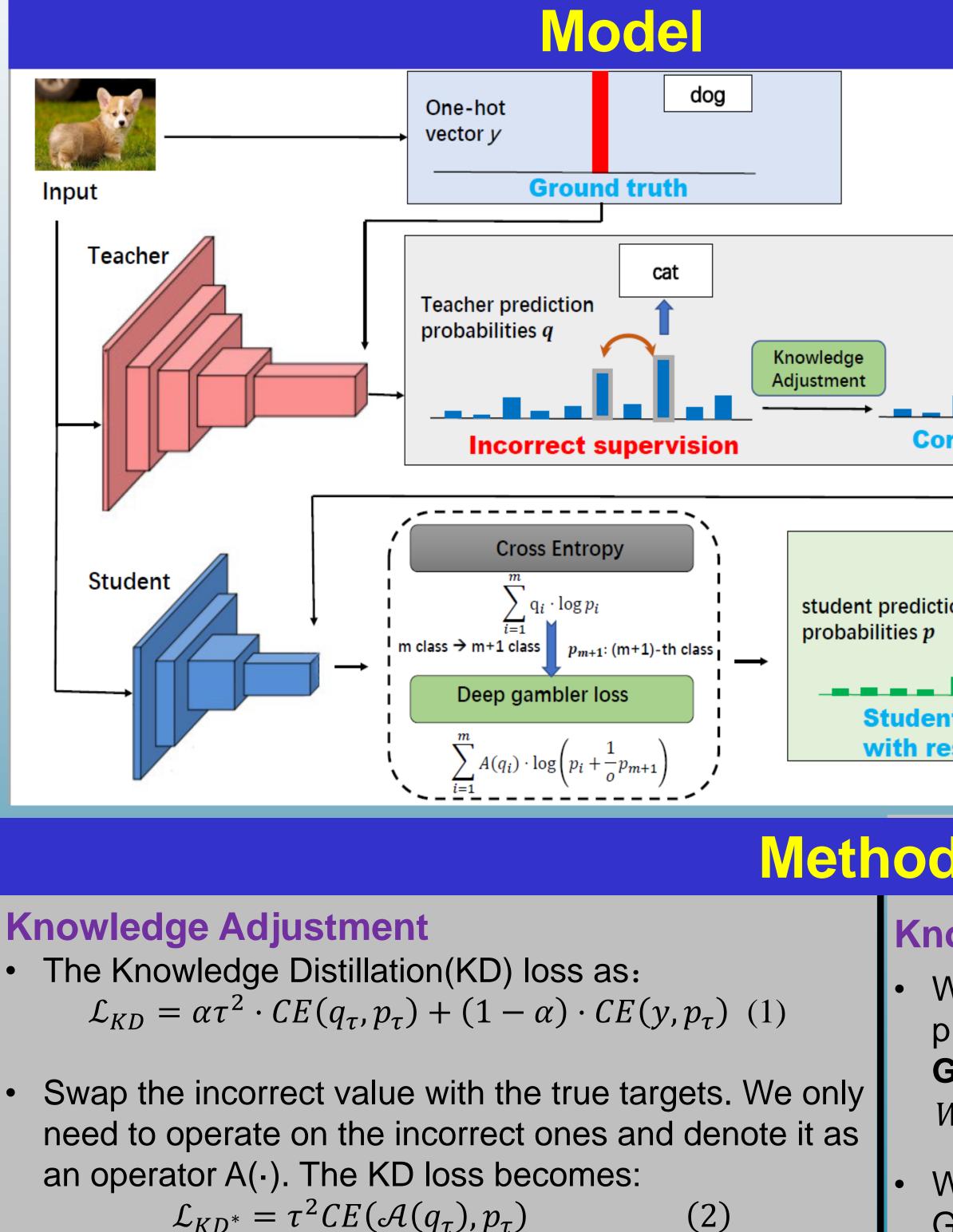
Motivation

- We propose to use knowledge adjustment to revise teacher's incorrect supervision using the ground truth label.
- 2. We propose to use the deep gambler loss to train the student network in an end-to-end way.
- 3. We evaluate the proposed method under two knowledge distillation settings. i.e., knowledge distillation across different network structures and distillation across networks with different depths.

Knowledge Distillation with a Precise Teacher and **Prediction with Abstention**

Yi Xu, Jian Pu, Hui Zhao

School of Software Engineering, East China Normal University, Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University Email: 51184501068@stu.ecnu.edu.cn, jianpu@fudan.edu.cn, hzhao@sei.ecnu.edu.cn



		Method	Fashion-	MNIST	SVHN	CIFAR10	CIFA
		Method	Accura	cy(%)	Accuracy(%)	Accuracy(%)	Accura
XX	St	udent(AlexNe	t) 92.0	50	94.86	85.57	60.
		acher(ResNet5	*		97.47	95.42	76.
		Deep Gambler			95.02	87.17	61.
	Pro	oposed Metho	od 92.9	94	95.05	87.11	61.
	BLE II: The comparison work structures.	of Sum Co	C	r in 0%-1	100% and 7	70%-100% by	knowle SVHN
	Method	Sum C	Coverage Error		erage Error	Sum Coverage	
			(0,100)		,100)	(0,100)	
	Softmax Respon		130.21		0.70	218.48	
	Deep Gamble		119.03		5.63	195.96	
	Proposed Metho	od	114.58		5.88	181.96	01010
	Method			AR10	ana a Errar	Sum Courses	CIFAR1
		11	Coverage Error (0,100)		erage Error ,100)	Sum Coverage 1 (0,100)	Error S
	Softmax Respor		287.52		2.33	1826.63	
d supervision	Deep Gamble		265.85				
	200 Sunole		400000	2	7.62	1612.89	1
	Proposed Metho	od	276.37	22	7.62 0.54	1612.89 1598.24	for
	Proposed Metho Stillation a TABLE III: The comparis Stude Teach Dec Prop	od CCCOS son of accura Method ent(ResNet18) her(ResNet50) ep Gambler posed Method	276.37 55 Ne acy on four da Fashion-MNIST Accuracy(%) 93.64 93.95 93.79 93.79 93.92	220 two tasets by ki <u>SVHN</u> <u>Accuracy</u> 97.25 97.47 97.20 97.41	0.54 rk wi nowledge dist N CIFAF y(%) Accurac 5 95.1 7 95.4 0 95.1 1 95.4	1598.24 th Dif tillation across d rillation across d CIFAR10 cy(%) Accuracy 4 76.42 12 76.86 2 76.86	lifferent n
	Proposed Metho Stillation a TABLE III: The comparis	od CCCOS son of accura Method ent(ResNet18) her(ResNet50) ep Gambler posed Method	276.37 55 Ne acy on four da Fashion-MNIST Accuracy(%) 93.64 93.95 93.79 93.79 93.92	220 two tasets by ki <u>SVHN</u> <u>Accuracy</u> 97.25 97.47 97.20 97.41	0.54 rk wi nowledge dist N CIFAF y(%) Accurac 5 95.1 7 95.4 0 95.1 1 95.4	1598.24 th Dif tillation across d rillation across d CIFAR10 cy(%) Accuracy 4 76.42 12 76.86 2 76.86	lifferent n
	Proposed Metho Stillation a TABLE III: The comparis Stude Teach Dec Prop	od CCCOS son of accura Method ent(ResNet18) her(ResNet50) ep Gambler posed Method of Sum Cove	276.37 SSNe acy on four da Fashion-MNIST Accuracy(%) 93.64 93.95 93.79 93.79 93.92 erage Error in Fashion-MNIST	220 two tasets by ki Accuracy 97.25 97.47 97.20 97.41 0%-100%	0.54 rk wi nowledge dist N CIFAF (%) Accurac 5 95.1 7 95.4 0 95.1 1 95.4 and 70%-100	1598.24 th Dif tillation across d R10 CIFAR10 cy(%) Accuracy 4 76.42 12 76.86 12 76.86 16 Y 17 76.86 16 Y N	lifferent n
	Proposed Metho Stillation a TABLE III: The comparis Stude Teach Dec Prop	od CCCOS son of accura Method ent(ResNet18) her(ResNet50) ep Gambler bosed Method of Sum Cove	276.37 SSNe acy on four da Fashion-MNIST Accuracy(%) 93.64 93.95 93.79 93.92 erage Error in Fashion-MNIS rage Error Sum	220 two tasets by ki Accuracy 97.25 97.47 97.20 97.41 0%-100% a ST Coverage Er	0.54 rk wi nowledge dist N CIFAH y(%) Accurac 5 95.1 7 95.4 0 95.1 1 95.4 and 70%-100	1598.24 th Dif tillation across d R10 CIFAR10 cy(%) Accuracy .4 76.42 .2 76.86 .2 76.86 .2 76.86 .2 76.86 .2 76.86 .2 76.86 .2 76.86 .2 76.86 .2 76.86 .2 76.86	e distillat
Dis	Proposed Metho Stillation a TABLE III: The comparis Stude Teach Dec Prop TABLE IV: The comparison on the twork scales.	od CCCOS son of accura Method ent(ResNet18) her(ResNet50) ep Gambler oosed Method of Sum Cove (0,1	276.37 SSNe acy on four da Fashion-MNIST Accuracy(%) 93.64 93.95 93.79 93.92 erage Error in Fashion-MNIS rage Error Sum	220 two tasets by ki Accuracy 97.25 97.47 97.20 97.41 0%-100%	0.54 rk wi nowledge dist N CIFAH y(%) Accurac 5 95.1 7 95.4 0 95.1 1 95.4 and 70%-100 Tor Sum Cov (0,	1598.24 th Dif tillation across d R10 CIFAR10 cy(%) Accuracy 4 76.42 12 76.86 12 76.86 16 Y 17 76.86 16 Y N	lifferent 1
	Proposed Metho Stillation a TABLE III: The comparis Stude Teach Dea Prop TABLE IV: The comparison on tetwork scales. Method Softmax Response Deep Gambler	od Image: Constraint of the section of the sec	276.37 SSNe acy on four da Fashion-MNIST Accuracy(%) 93.64 93.95 93.79 93.92 erage Error in Fashion-MNIST Fashion-MNIST 93.95 93.79 93.92 erage Error in Sao 30	220 two tasets by kn Accuracy 97.25 97.47 97.20 97.41 0%-100% ST Coverage Er (70,100) 98.05 82.55	0.54 rk wi nowledge dist N CIFAI y(%) Accurac 5 95.1 7 95.4 0 95.1 1 95.4 and 70%-100 ror Sum Cov (0, 111 11 12	1598.24 th Dif tillation across d R10 CIFARIO cy(%) Accuracy 4 76.42 12 76.86 12 76.86 12 76.86 12 76.86 12 76.86 12 76.86 12 76.86 12 76.86 12 76.86 12 76.86 12 76.86 12 76.86 13 80 14 76.86 15 80 100 5.18 100 5.18 100 5.18 100 5.18	00 (%)
	Proposed Metho Stillation a TABLE III: The comparis Stude Teach Dec Prop TABLE IV: The comparison on network scales. Method Softmax Response	od Image: Constraint of the section of the sec	276.37 SSNE acy on four da Fashion-MNIST Accuracy(%) 93.64 93.95 93.79 93.92 erage Error in Fashion-MNIS rage Error Sum 00) 3.30 .18	220 two tasets by kn Accuracy 97.25 97.47 97.20 97.41 0%-100% ST Coverage Er (70,100) 98.05	0.54 rk wi nowledge dist N CIFAI y(%) Accurac 5 95.1 7 95.4 0 95.1 1 95.4 and 70%-100 ror Sum Cov (0, 111 11 12	1598.24 th Dif tillation across d R10 CIFARIO cy(%) Accuracy 4 76.42 12 76.86 12 76.86 13 76.86 14 76.86 15 8 100 SVHN Verage Error Sum 1000 Sum 100 Sum 100 Sum	e distilla Coverage (70,100) 61.19
	Proposed Metho Stillation a TABLE III: The comparis Stude Teach Dea Prop TABLE IV: The comparison on tetwork scales. Method Softmax Response Deep Gambler	od Image: Constraint of the section of the sec	276.37 SSNe acy on four da Fashion-MNIST Accuracy(%) 93.64 93.95 93.79 93.92 erage Error in Fashion-MNIS rage Error Sum 00) 3.30 .18 CIFAR10	220 two tasets by kn Accuracy 97.25 97.47 97.20 97.41 0%-100% ST Coverage Er (70,100) 98.05 82.55 68.82	0.54 rk wi nowledge dist N CIFAF y(%) Accurac 5 95.1 7 95.4 0 95.1 1 95.4 and 70%-100 ror Sum Cov (0, 11 12 10	1598.24 th Dif tillation across d R10 CIFAR10 cy(%) Accuracy 4 76.42 12 76.86 2% by knowledg SVHN SVHN rerage Error Sum 1000 Sum 5.18 Sum 26.93 CIFAR100	Coverage (70,100) 61.19 64.80 55.89
Dis	Proposed Metho Stillation a TABLE III: The comparis Stude Teach Dec Prop TABLE IV: The comparison on the twork scales. Method Softmax Respons Deep Gambler Proposed Method	od Image: Constraint of the section of the sec	276.37 SSNe acy on four da Fashion-MNIST Accuracy(%) 93.64 93.95 93.79 93.92 erage Error in Fashion-MNIST Fashion-MNIST Accuracy(%) 93.64 93.95 93.79 93.92 erage Error Sum 30 .30 .18 CIFAR10 rage Error Sum	220 two tasets by kn Accuracy 97.25 97.47 97.20 97.41 0%-100% ST Coverage Er (70,100) 98.05 82.55	0.54 rk wi nowledge dist N CIFAF y(%) Accurac 5 95.1 7 95.4 0 95.1 7 95.4 0 95.1 1 95.4 and 70%-100 ror Sum Cov (0, 11 12 10 ror Sum Cov 0 Sum Cov	1598.24 th Dif tillation across d R10 CIFAR10 cy(%) Accuracy 4 76.42 12 76.86 2% by knowledg SVHN SVHN rerage Error Sum 100) Sum 5.18 CIFAR100	lifferent 00 (%) e distilla Coverage (70,100) 61.19 64.80 55.89 Coverage
Di:	Proposed Metho Stillation a TABLE III: The comparis Stude Teach Dec Prop TABLE IV: The comparison on the twork scales. Method Softmax Respons Deep Gambler Proposed Method	od Image: Construction of accuration of accuratio	276.37 SSNe acy on four da Fashion-MNIST Accuracy(%) 93.64 93.95 93.79 93.92 erage Error in Fashion-MNIST Fashion-MNIST Accuracy(%) 93.92 erage Error Sum .00) 3.30 .18 CIFAR10 rage Error Sum .00)	220 two tasets by ki Accuracy 97.25 97.47 97.20 97.41 0%-100% ST 1 Coverage Er (70,100) 98.05 82.55 68.82 1 Coverage Er	0.54 rk wi nowledge dist N CIFAH y(%) Accurac 5 95.1 7 95.4 0 95.1 1 95.4 and 70%-100 ror Sum Cov (0, 111 12 100 ror ror Sum Cov (0, 0 86	1598.24 th Dif tillation across d R10 CIFAR10 cy(%) Accuracy 4 76.42 12 76.86 12 76.86 12 76.86 16 SVHN Verage Error Sum 1000 SURN 15.18 CIFAR100 CIFAR100 Sum	00 (%)

Knowledge Distillation with Deep Gambler

We can add a class to stand for abandoning predictions and reservations according to **Deep Gambler Loss^[2]** in selective classification: $W(b(f),p) = \sum_{i} \log[o \cdot f_{w}(x_{i})_{j} + f_{w}(x_{i})_{m+1}]$ (3)

We proposed the loss function that utilizes Deep Gambler (DG) loss to the KA method.

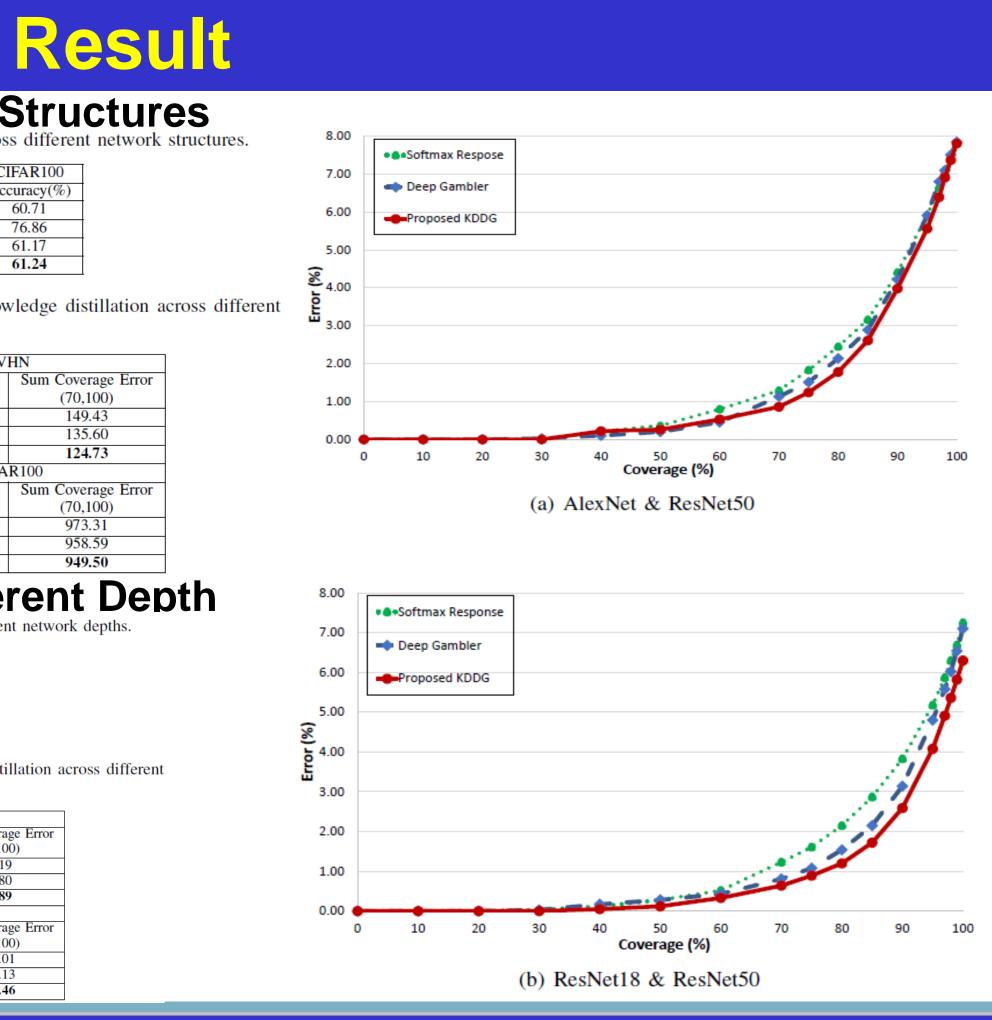
 $\mathcal{L} = \sum_{i} \mathcal{A}(q_{\tau}^{i}) \log \left(p_{\tau}^{i} + \frac{1}{o} p_{\tau}^{m+1} \right) (4)$

[1] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," arXiv preprint arXiv:1503.02531, 2015.

[2] Z. Liu, Z. Wang, P. P. Liang, R. R. Salakhutdinov, L.-P. Morency, an dM. Ueda, "Deep gamblers: Learning to abstain with portfolio theory," in Advances in Neural Information Processing Systems, pp. 10623–10633,2019.

[3] Y. Geifman and R. El-Yaniv, "Selective classification for deep neural networks," in Advances in Neural Information Processing Systems, pp. 4878–4887, 2017.





Reference