Supplementary Material for ORC: Network Group-based Knowledge Distillation using Online Role Change

Junyong Choi^{1,2}, Hyeon Cho¹, Seokhwa Cheung¹, and Wonjun Hwang^{1,3} ¹Ajou University, Korea, ²Hyundai Motor Company, ³Naver AI Lab

chldusxkr@hyundai.com, {ch0104, shjeong008, wjhwang}@ajou.ac.kr

1. Implementation Details on Multiple Networks

In this supplementary material, we describe the details of the networks used in our experiments in Tables 2 and 3 of the main paper such as ResNet[1], WRN[5], VGG[4].

We explain the details of used networks according to experimental settings. As introduced in the main paper, we experimented with the teacher network and the student network in two cases. In the first case, the teacher network and the student network have the same architecture, and in the second case, networks with a different architecture are used.

1.1. Similar Architecture Networks for CIFAR-100

We conducted the experiment by dividing the case where the structure of the teacher network and the student network are the same into a total of 7 types. The networks used in the experiment of Table 2 in main paper are described in Table 1, and a general ResNet, a ResNet with increased channels, a general wide residual network, and VGG were used.

Teacher Student WRN40.2 WRN34.2 WRN28.2 WRN16.2 WRN40.2 WRN40.1 WRN40.1 WRN40.1 ResNet56 ResNet44 ResNet32 ResNet20 Resnet110 ResNet56 ResNet44 ResNet32 ResNet20 ResNet110 ResNet56 ResNet44 ResNet32 ResNet32 ResNet32x4 ResNet26x4 ResNet14x4 ResNet8x4 VGG13 VGG11 VGG9 VGG8					
WRN40.2 WRN34.2 WRN28.2 WRN16.2 WRN40.2 WRN40.1 WRN40.1 WRN40.1 ResNet56 ResNet44 ResNet32 ResNet20 Resnet110 ResNet56 ResNet44 ResNet32 ResNet20 ResNet110 ResNet56 ResNet44 ResNet32 ResNet32 ResNet310 ResNet56 ResNet44 ResNet32 ResNet32 ResNet32x4 ResNet26x4 ResNet14x4 ResNet8x4 VGG13 VGG11 VGG9 VGG8	Teacher				Student
WRN40_2 WRN40_1 WRN40_1 WRN40_1 ResNet56 ResNet44 ResNet32 ResNet20 Resnet110 ResNet56 ResNet44 ResNet32 ResNet20 ResNet110 ResNet56 ResNet44 ResNet32 ResNet20 ResNet110 ResNet56 ResNet44 ResNet32 ResNet32 ResNet32x4 ResNet26x4 ResNet14x4 ResNet8x4 VGG13 VGG11 VGG9 VGG8	WRN40_2	WRN34_2	WRN28_2		WRN16_2
ResNet56ResNet44ResNet32ResNet20Resnet110ResNet56ResNet44ResNet32ResNet20ResNet110ResNet56ResNet44ResNet32ResNet32x4ResNet26x4ResNet14x4ResNet8x4VGG13VGG11VGG9VGG8	WRN40_2	WRN40_1	WRN40_1		WRN40_1
Resnet110ResNet56ResNet44ResNet32ResNet20ResNet110ResNet56ResNet44ResNet32ResNet32x4ResNet26x4ResNet14x4ResNet8x4VGG13VGG11VGG9VGG8	ResNet56	ResNet44	ResNet32		ResNet20
ResNet110ResNet56ResNet44ResNet32ResNet32x4ResNet26x4ResNet14x4ResNet8x4VGG13VGG11VGG9VGG8	Resnet110	ResNet56	ResNet44	ResNet32	ResNet20
ResNet32x4ResNet26x4ResNet14x4ResNet8x4VGG13VGG11VGG9VGG8	ResNet110	ResNet56	ResNet44		ResNet32
VGG13 VGG11 VGG9 VGG8	ResNet32x4	ResNet26x4	ResNet14x4		ResNet8x4
	VGG13	VGG11	VGG9		VGG8

Table 1. Networks of Similar Architecture

1.2. Different Architecture Networks for CIFAR-100

We conducted the experiment by dividing the case where the structure of the teacher network and the student network are different into 6 types. The networks used for the experiment are described in Table 2, and the network used for the teacher used the same network as the networks tested in the same case, but ShuffleNetV1[6], ShuffleNetV2[2], and MobileNetV2[3] were additionally used as the student networks.

Table 2. Networks of Different Architecture

Table	2. 1100 WOLKS OF	Different me	millecture
Teacher			Student
ResNet50	ResNet34	ResNet18	MobileNetV2
ResNet50	ResNet34	ResNet32	VGG8
ResNet32x4	ResNet26x4	ResNet14x4	ShuffleNetV1
ResNet32x4	ResNet26x4	ResNet14x4	ShuffleNetV2
WRN40_2	WRN34_2	WRN28_2	ShuffleNetV1
VGG13	VGG11	VGG9	MobileNetV2

1.3. Network Architectures for ImageNet

This section explains the networks' architecture used in ImageNet. This experiment was conducted using ResNet and designed using 4 residual blocks. Referring to Table 3, unlike the frequently used ResNet34 and ResNet18, ResNet28 and ResNet22 are newly designed to have the largest number of the 3rd block, which is a feature of ResNet.

Table 3. ResNet for ImageNet Dataset

	ResNet28		ResNet22		ResNet18			
Conv7x7, 64								
		BN, I	ReLU					
	Conv3x3, 64		Conv3x3, 64		Conv3x3, 64			
2	BN, ReLU	2	BN, ReLU		BN, ReLU			
лэ	Conv3x3, 64	1.0	Conv3x3, 64	72	Conv3x3, 64	X2		
	BN, ReLU		BN, ReLU		BN, ReLU			
	Conv3x3, 128		Conv3x3, 128		Conv3x3, 128			
4	BN, ReLU	x3	BN, ReLU	x3	BN, ReLU	x2		
X4	Conv3x3, 128		Conv3x3, 128		Conv3x3, 128			
	BN, ReLU		BN, ReLU		BN, ReLU			
	Conv3x3, 256		Conv3x3, 256		Conv3x3, 256			
	BN, ReLU		BN, ReLU	2	BN, ReLU			
xo	Conv3x3, 256	X4	Conv3x3, 256	хэ	Conv3x3, 256			
	BN, ReLU		BN, ReLU		BN, ReLU			
	Conv3x3, 512		Conv3x3, 512		Conv3x3, 512			
9	BN, ReLU	9	BN, ReLU	0	BN, ReLU	9		
хэ	Conv3x3, 512	xə	Conv3x3, 512	X2	Conv3x3, 512	x2		
	BN, ReLU		BN, ReLU		BN, ReLU			
AveragePool								
FC-100								
Softmax								
	x3 x4 x6 x3	ResNet28 X3 Conv3x3, 64 BN, ReLU Conv3x3, 64 BN, ReLU Conv3x3, 128 BN, ReLU Conv3x3, 128 K4 EN, ReLU Conv3x3, 128 EN, ReLU Conv3x3, 256 EN, ReLU X6 EN, ReLU Conv3x3, 256 EN, ReLU X6 EN, ReLU Conv3x3, 512 EN, ReLU X3 EN, ReLU Conv3x3, 512 EN, ReLU MN, ReLU Conv3x3, 512 BN, ReLU Conv3x3, 512 B	ResNet28 Conv7 BN, I Conv3x3, 64 BN, ReLU x3 BN, ReLU x3 Conv3x3, 64 x3 BN, ReLU x3 BN, ReLU x3 BN, ReLU x3 BN, ReLU x3 Conv3x3, 128 x3 BN, ReLU x3 Conv3x3, 526 x4 BN, ReLU x4 BN, ReLU x4 BN, ReLU x4 BN, ReLU x3 BN, ReLU x4 BN, ReLU x3 BN, ReLU x3	ResNet28 ResNet22 Conv3x3, 64 BN, ReLU X3 BN, ReLU Conv3x3, 64 BN, ReLU X3 BN, ReLU Conv3x3, 64 BN, ReLU Conv3x3, 64 BN, ReLU X3 BN, ReLU Conv3x3, 128 Conv3x3, 128 Conv3x3, 128 MAR Conv3x3, 256 Conv3x3, 256 N, ReLU Conv3x3, 256 EN, ReLU Conv3x3, 256 Conv3x3, 256 BN, ReLU X4 EN, ReLU Conv3x3, 256 Conv3x3, 512 X4 EN, ReLU Conv3x3, 256 BN, ReLU X4 EN, ReLU Conv3x3, 256 BN, ReLU X4 EN, ReLU Conv3x3, 256 BN, ReLU X4 EN, ReLU Conv3x3, 512 X3 EN, ReLU X3 BN, ReLU Conv3x3, 512 SN, ReLU Conv3x3, 512 BN, ReLU X3 BN, ReLU Conv3x3, 512 BN, ReLU X4 BN, ReLU Conv3x3, 512 BN, ReLU SN, ReLU SN	ResNet28 ResNet22 Conv7x7, 64 BN, ReLU BN, ReLU Conv3x3, 64 BN, ReLU x3 BN, ReLU Conv3x3, 64 BN, ReLU x3 BN, ReLU BN, ReLU Conv3x3, 64 BN, ReLU Conv3x3, 128 BN, ReLU Conv3x3, 128 BN, ReLU Conv3x3, 128 BN, ReLU Conv3x3, 256 BN, ReLU Conv3x3, 256 BN, ReLU Conv3x3, 256 BN, ReLU Conv3x3, 256 BN, ReLU Conv3x3, 512 BN, ReLU NR RELU X3 BN, ReLU Conv3x3, 512 BN, ReLU X3 Conv3x3, 512 BN, ReLU Conv3x3, 512 BN, ReLU Conv3x3, 512 BN, ReLU S0ftmax X4	ResNet28 ResNet22 ResNet18 Conv7x7, 64 BN, ReLU BN, ReLU x3 BN, ReLU Conv3x3, 64 BN, ReLU Conv3x3, 64 x3 Conv3x3, 64 BN, ReLU BN, ReLU x4 BN, ReLU Conv3x3, 64 BN, ReLU Conv3x3, 64 BN, ReLU x4 Conv3x3, 128 BN, ReLU Conv3x3, 128 Conv3x3, 128 BN, ReLU x4 BN, ReLU X3 BN, ReLU Conv3x3, 128 BN, ReLU Conv3x3, 128 BN, ReLU X3 BN, ReLU Conv3x3, 128 BN, ReLU Conv3x3, 128 BN, ReLU Conv3x3, 128 BN, ReLU Conv3x3, 128 BN, ReLU Conv3x3, 128 Conv3x3, 256 BN, ReLU Conv3x3, 256 Conv3x3, 256 BN, ReLU BN, ReLU Conv3x3, 512 X4 BN, ReLU BN, ReLU Conv3x3, 512 BN, ReLU X3 BN, ReLU Conv3x3, 512 BN, ReLU BN, ReLU Conv3x3, 512 X4 BN, ReLU Conv3x3, 512 BN, ReLU <		

1.4. Network Architectures for CIFAR-100

In this section, the details of the networks' architecture used in the CIFAR-100 experiments are described. The networks used in these experiments include commonly used networks such as ResNet56 and ResNet32, but there are additionally designed networks such as ResNet44 and ResNet14x4 to use multi-teacher networks.

ResNet. We use standard ResNet and ResNetx4 which increases the number of channels by 4 times in the experiment. For the structure of the networks, various models with different depths are designed by changing the total number of layers by changing the number of blocks. The number of layers of the network are designed while maintaining the structure that takes a large number of third blocks. Referring to Table 4, the contents of five standard ResNet and four ResNetx4 are shown. Unlike the architecture of ResNet described above, the networks used in the CIFAR dataset consist of three blocks.

ResNet11	ResNet110 ResNet56 ResNet44 ResNet32 ResNet20											
	Conv3x3, 16											
Conv3x3, 16 BN, ReLU Conv3x3, 16 BN, ReLU	x18	Co: Bl Co: Bl	nv3x3, 16 N, ReLU nv3x3, 16 N, ReLU	x9	Conv3x BN, Re Conv3x BN, Re	ReLU 3, 16 2LU 3, 16 2LU	x7	Conv3x BN, Re Conv3x BN, Re	3, 16 2LU 3, 16 2LU	x5	Conv3x3, 16 BN, ReLU Conv3x3, 16 BN, ReLU	x3
Conv3x3, 32 BN, ReLU Conv3x3, 32 BN, ReLU	x18	Co: Bl Co: Bl	nv3x3, 32 N, ReLU nv3x3, 32 N, ReLU	x9	Conv3x BN, Re Conv3x BN, Re	3, 32 ±LU 3, 32 ±LU	x7	Conv3x BN, Re Conv3x BN, Re	3, 32 2LU 3, 32 2LU	x5	Conv3x3, 32 BN, ReLU Conv3x3, 32 BN, ReLU	x3
Conv3x3, 64 BN, ReLU Conv3x3, 64 BN, ReLU	x18	Co: Bl Co: Bl	nv3x3, 64 N, ReLU nv3x3, 64 N, ReLU	x9	Conv3x BN, Re Conv3x BN, Re	3, 64 LU 3, 64 LU	x7	Conv3x BN, Re Conv3x BN, Re	3, 64 2LU 3, 64 2LU	x5	Conv3x3, 64 BN, ReLU Conv3x3, 64 BN, ReLU	x3
	AveragePool EC 100											
	Softmax											
ResNet32x4 ResNet26x4 ResNet14x4 ResNet8x4												
					BN,	sx3, . ReLU	32 J					
Conv3x3,	Conv3x3, 64 Conv3x3, 64 Conv3x3, 64 Conv3x3, 64											
BN, ReLU Conv3x3, (BN, ReLU	ReLU x5 BN, ReLU 3x3, 64 x5 Conv3x3, 64 ReLU BN, ReLU					B Co B	N, I mv3 N, I	ReLU x3, 64 ReLU	x2	E Co E	BN, ReLU onv3x3, 64 BN, ReLU	x1
Conv3x3, 128 Conv3x3, 128 Conv3x3, 128 Conv3x3, 128 BN, ReLU BN, ReLU BN, ReLU BN, ReLU Conv3x3, 128 BN, ReLU BN, ReLU Conv3x3, 128 BN, ReLU Conv3x3, 128 BN, ReLU BN, ReLU BN, ReLU Conv3x3, 128 BN, ReLU BN, ReLU BN, ReLU BN, ReLU Sonv3x3, 128 BN, ReLU								x1				
Conv3x3, 2 BN, ReLU Conv3x3, 2 BN, ReLU	256 J 256 J	x5	Conv3x BN, F Conv3x BN, F	3, 2 teLU 3, 2 teLU	56 56 x4	Co B Co B	nv3> N, I nv3> N, I	:3, 256 ReLU :3, 256 ReLU	x2	Co E Co E	nv3x3, 256 BN, ReLU nv3x3, 256 BN, ReLU	x1
AveragePool												
	FC-100											
	Softmax											

Wide ResNet. We use a standard Wide Residual Network (WRN) for the experiment. Like ResNet, various networks are designed to have different depths by changing the total layer by changing the number of blocks. Referring to Table 5, the architectures of WRN40_2 to WRN16_2 used as the teacher network

and WRN40_1 used as the student network are shown. Also, like ResNet, each cell means one Residual block, and it consists of consecutively as many as the number on the right.

Table 5. Wide ResNet for CIFAR-100

$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$		-			0 10001.00	10.		-0.	•		
$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$	WRN40_2	WRN40.2 WRN34.2 WRN28.2 WRN16.2 WRN40.1									
$\begin{array}{ c c c c c c c c c c c c c c c c c c c$					BN, ReLU						
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $		Conv3x3, 16									
Conv3x3, 32	BN, ReLU		BN, ReLU		BN, ReLU		BN, ReLU		BN, ReLU		
BN, ReLU X0 BN, ReLU X1 BN, ReLU X2 BN, ReLU X0 Conv3x3, 32 Conv3x3, 32 Conv3x3, 32 Conv3x3, 32 Conv3x3, 16 Conv3x3, 16 Conv3x3, 16 Conv3x3, 16 Conv3x3, 16 Conv3x3, 16 BN, ReLU BN, ReLU BN, ReLU SO Conv3x3, 16 Conv3x3, 16 Conv3x3, 16 Conv3x3, 16 Conv3x3, 12 Conv3x3, 12 Conv3x3, 64 Conv3x3, 64 Conv3x3, 12 Conv3x3, 64 Conv3x3, 12 Conv3x3, 64 Conv3x3, 12 Conv3x3, 12 Conv3x3, 128 Conv3x	Conv3x3, 32		Conv3x3, 32		Conv3x3, 32	4	Conv3x3, 32		Conv3x3, 16		
Conv3x3, 32 Conv3x3, 32 Conv3x3, 32 Conv3x3, 32 Conv3x3, 32 BN, ReLU Conv3x3, 32 Conv3x3, 32 Conv3x3, 32 Conv3x3, 16 COnv3x3, 64 x6 Conv3x3, 64 x6 Conv3x3, 64 BN, ReLU BN, ReLU BN, ReLU Conv3x3, 64 Conv3x3, 64<	BN, ReLU	xo	BN, ReLU	xə	BN, ReLU	X4	BN, ReLU	X2	BN, ReLU	xo	
BN, ReLU BN, ReLU BN, ReLU BN, ReLU BN, ReLU BN, ReLU Conv3x3, 64 Conv3x3, 32 N ReLU Conv3x3, 32 R R R R R R R R R R R R R R R R R R R R R R R R R R R R R R	Conv3x3, 32		Conv3x3, 32		Conv3x3, 32		Conv3x3, 32		Conv3x3, 16		
Conv3x3, 64 x6 Conv3x3, 64 x5 Conv3x3, 64 x4 Conv3x3, 64 x2 Conv3x3, 32 x5 BN, ReLU Conv3x3, 64 x6 BN, ReLU Conv3x3, 64 x8 BN, ReLU BN, ReLU Conv3x3, 64 x8 BN, ReLU Conv3x3, 64 x4 BN, ReLU Conv3x3, 32 BN, ReLU Conv3x3, 64 X8 Conv3x3, 64 X2 BN, ReLU Conv3x3, 32 BN, ReLU Conv3x3, 64 X4 BN, ReLU Conv3x3, 64 X2 Conv3x3, 32 BN, ReLU Conv3x3, 128 Conv3x3, 128 Conv3x3, 64 X4 BN, ReLU Conv3x3, 128 Conv3x3, 64 X6 S0 Conv3x3, 64 X6 S0 Conv3x3, 64 X6 S0 Conv3x3, 64 X6 S0 Conv3x3, 64 S0 S0 S0 S0 S0 S0 Conv3x3, 64 S0 S0 S0 S0	BN, ReLU		BN, ReLU		BN, ReLU		BN, ReLU		BN, ReLU		
BN, ReLU X0 BN, ReLU X3 BN, ReLU X4 BN, ReLU X2 BN, ReLU Conv3x3, 64 BN, ReLU BN, ReLU BN, ReLU BN, ReLU BN, ReLU Conv3x3, 64 Conv3x3, 64 Conv3x3, 32 Conv3x3, 32 BN, ReLU BN, ReLU BN, ReLU BN, ReLU BN, ReLU BN, ReLU Conv3x3, 32 BN, ReLU Conv3x3, 128 Conv3x3, 128 BN, ReLU Conv3x3, 128 BN, ReLU Conv3x3, 64 SN, ReLU Conv3x3, 64 SN, ReLU Conv3x3, 64 Conv3x3, 64 SN, ReLU <td< td=""><td>Conv3x3, 64</td><td></td><td>Conv3x3, 64</td><td></td><td>Conv3x3, 64</td><td> 4</td><td>Conv3x3, 64</td><td></td><td>Conv3x3, 32</td><td></td></td<>	Conv3x3, 64		Conv3x3, 64		Conv3x3, 64	4	Conv3x3, 64		Conv3x3, 32		
Conv3x3, 64 Conv3x3, 64 Conv3x3, 64 Conv3x3, 64 Conv3x3, 32 BN, ReLU BN, ReLU BN, ReLU BN, ReLU BN, ReLU Conv3x3, 128 Conv3x3, 128 BN, ReLU Conv3x3, 128	BN, ReLU	xo	BN, ReLU	xə	BN, ReLU	X4	BN, ReLU	X2	BN, ReLU	xo	
BN, ReLU BN, ReLU BN, ReLU BN, ReLU BN, ReLU BN, ReLU Conv3x3, 128 BN, ReLU BN, ReLU Conv3x3, 128 Conv3x3, 128 Conv3x3, 128 Conv3x3, 128 X4 BN, ReLU Conv3x3, 128 Conv3x3, 64 NC Conv3x3, 128 Conv3x3, 128 Conv3x3, 128 Conv3x3, 128 Conv3x3, 128 Conv3x3, 64 NC	Conv3x3, 64		Conv3x3, 64		Conv3x3, 64		Conv3x3, 64		Conv3x3, 32		
Conv3x3, 128 BN, ReLU x6 BN, ReLU Conv3x3, 128 BN, ReLU x5 BN, ReLU Conv3x3, 128 BN, ReLU x4 BN, ReLU Conv3x3, 128 BN, ReLU x2 Conv3x3, 128 Conv3x3, 64 BN, ReLU x6 Conv3x3, 128 Conv3x3, 128 x4 Conv3x3, 128 x2 Conv3x3, 64 BN, ReLU x6 Conv3x3, 128 x6 Conv3x3, 128 x7 Conv3x3, 64 x6 Conv3x3, 128 Conv3x3, 128 Conv3x3, 128 x6 Conv3x3, 64 x6 Conv3x3, 128 Conv3x3, 128 Conv3x3, 128 Conv3x3, 128 Conv3x3, 64 x6 Conv3x3, 128 Conv3x3, 128 Conv3x3, 128 Conv3x3, 128 Conv3x3, 64 X6 Conv3x3, 128 Conv3x3, 128 Conv3x3, 128 Conv3x3, 64 X6 FC-100	BN, ReLU		BN, ReLU		BN, ReLU		BN, ReLU		BN, ReLU		
BN, ReLU X0 BN, ReLU X3 BN, ReLU X4 BN, ReLU X2 BN, ReLU X0 Conv3x3, 128 Conv	Conv3x3, 128		Conv3x3, 128		Conv3x3, 128	4	Conv3x3, 128		Conv3x3, 64		
Conv3x3, 128 Conv3x3, 128 Conv3x3, 128 Conv3x3, 64 BN, ReLU BN, ReLU AveragePool FC-100 Softmax Softmax	BN, ReLU	ло	BN, ReLU	ло	BN, ReLU	A-4	BN, ReLU	12	BN, ReLU	10	
BN, ReLU AveragePool FC-100 Softmax	Conv3x3, 128		Conv3x3, 128		Conv3x3, 128		Conv3x3, 128		Conv3x3, 64		
AveragePool FC-100 Softmax		BN, ReLU									
FC-100 Softmax		AveragePool									
Softmax		FC-100									
					Softmax						

VGG. We use a standard VGG for the experiment. Referring to Table 6, the network architectures from VGG13 to VGG8 are distinguished by the difference in layer depth. VGG13, VGG11, and VGG9 are used as the teacher network, and VGG8 is used as the student network. The deeper the model, the more convolution operations with high channels are used.

Table 6. VGG for CIFAR-100

VGG13	VGG13 VGG11 VGG9 VGG8									
Conv3x3, 64	x2	Conv3x3, 64	x1	Conv3x3, 64	x1	Conv3x3, 64	x1			
BN, ReLU		BN, ReLU		BN, ReLU		BN, ReLU				
Conv3x3, 128	v 2	Conv3x3, 128	v1	Conv3x3, 128	v 1	Conv3x3, 128	v1			
BN, ReLU	77	BN, ReLU	×1	BN, ReLU	71	BN, ReLU				
Conv3x3, 256		Conv3x3, 256		Conv3x3, 256	1	Conv3x3, 256	1			
BN, ReLU	X2	BN, ReLU	X2	BN, ReLU	XI	BN, ReLU	XI			
Conv3x3, 512		Conv3x3, 512		Conv3x3, 512	1	Conv3x3, 512	1			
BN, ReLU	X2	BN, ReLU	X2	BN, ReLU	XI	BN, ReLU	XI			
Conv3x3, 512		Conv3x3, 512		Conv3x3, 512		Conv3x3, 512	1			
BN, ReLU	X2	BN, ReLU	X2	BN, ReLU	X2	BN, ReLU	XI			
MaxPool										
FC-512										
FC-512										
FC-512										
	Softmax									

2. Effectiveness of Online Role Change

2.1. Reducing False Knowledge After ORC

After applying the online role change, we measured the change in performance of the networks included in the student group. As a result, in Fig. 1, it can be confirmed that the overall performance of the student network has improved for each class that was difficult.

2.2. Online Role-Change at every interation

We perform online role change at every iteration to utilize the advantages of student networks for each mini-batch. In other words, for each mini-batch, false



Figure 1. Accuracy improvement of student groups on difficult samples after ORC.

knowledge contamination by temporary teachers can be prevented by taking advantage of student networks. Therefore, we utilize the network which shows most confidence at specific mini-batch as a temporary teacher. Further, we compare the performance of different iterations per epoch to analyze the advantage of the number of switching. As shown in Table 7, the student network shows better accuracy as the number of role changes increases. It can be said that the strengths of students are highlighted and used as much as possible.

Table 7. ImageNet Top-1 ERROR on the num. of iterations.

Model	4,688 iters.	9,375 iters.	18,750 iters.
ResNet18	29.48	28.72	28.00

Limitation. We note that the multiple networkbased KD results in more computational complexity in training, but it does not affect the complexity of the student network.

References

- K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 770–778, Jun. 2016. 1
- N. Ma, X. Zhang, H. Zheng, and J. Sun. Shufflenet v2: Practical guidelines for efficient cnn architecture design. In *Proceedings of the European conference on computer* vision (ECCV), pages 116–131, 2018.
- [3] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4510– 4520, 2018.
- [4] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014. 1
- [5] S. Zagoruyko and N. Komodakis. Wide residual networks. British Machine Vision Conference, pages 87.1– 87.12, Sept. 2016. 1
- [6] X. Zhang, X. Zhou, M. Lin, and J. Sun. Shufflenet: An extremely efficient convolutional neural network for mobile devices. In *Proceedings of the IEEE conference* on computer vision and pattern recognition, pages 6848– 6856, 2018. 1