000

007

011

014

1

⁰⁰¹ Supplementary Material : Knowledge ⁰⁰² Diversification in Ensembles of Identical ⁰⁰⁴ Neural Networks

BMVC 2022 Submission # 0798

1 Additional Results

Table 1: Results of experiments on a single layer ConvNet with M filters trained on MNIST.
The middle column indicates accuracy results obtained with a single network of M filters.
The right column indicates the accuracy results obtained with two networks with M/2 filters
each, trained using feature difference loss across the two networks. Results indicate that two
networks trained with FDL learns better representations, achieves higher accuracy and thus
makes better use of the model capacity.

021	Filters	1x Network (%)	2x Half-Networks (%)
022			(M/2 fitzer each) + EDI
023	(M)	(M mers)	(M/2 litters each) + FDL
024	1	87.87	-
025	2	91.96	92.7
026	4	94.99	94.55
027	8	97.06	97.52
028	16	98	98.04
029	32	98.23	98.37
030	64	98.27	98.49
031	128	98.31	98.55
032	256	98.32	98.54
033			
034			
035			
036			
037			
038			
039			
040			
041			
042			

© 2022. The copyright of this document resides with its authors.

¹⁴⁵ It may be distributed unchanged freely in print or electronic forms.

Model		CIFA	R-10		FDL
Widder	Network 1	Network 2	Network 3	Network 4	Ensemble
VGG-16 (x1)	93.66				93.66
VGG-16 (x2)	93.57	93.96			94.93
VGG-16 (x3)	93.53	93.93	93.60		95.14
VGG-16 (x4)	93.82	93.79	93.97	93.58	95.22
ResNet-20 (x1)	92.20				92.20
ResNet-20 (x2)	91.79	92.02			93.56
ResNet-20 (x3)	92.36	92.51	92.28		94.29
ResNet-20 (x4)	92.18	91.98	92.10	92.70	94.44
ResNet-32 (x1)	93.21				93.21
ResNet-32 (x2)	92.37	92.10			93.81
ResNet-32 (x3)	93.33	93.48	93.11		94.78
ResNet-32 (x4)	92.69	92.91	93.03	93.02	94.88

 Table 2: FDL Ensemble of multiple neural networks for the CIFAR-10 dataset. We report 046

 accuracy metrics of each base model along with the ensemble. Best individual and ensemble

 047

 accuracies are marked in bold.

Table 3: FDL Ensemble of multiple neural networks for the CIFAR-100 dataset. We report065accuracy metrics of each base model along with the ensemble. Best individual and ensemble066accuracies are marked in bold.067

Model		CIFA	R-100		FDL
	Network 1	Network 2	Network 3	Network 4	Ensemble
VGG-16 (x1)	74.61				74.61
VGG-16 (x2)	74.27	74.44			77.02
VGG-16 (x3)	74.03	74.42	73.72		77.66
VGG-16 (x4)	74.82	73.88	73.73	74.26	78.34
ResNet-20 (x1)	67.82				67.82
ResNet-20 (x2)	67.63	68.26			71.48
ResNet-20 (x3)	67.57	67.57	68.50		73.43
ResNet-20 (x4)	67.11	67.76	67.48	67.71	73.71
ResNet-32 (x1)	69.42				69.42
ResNet-32 (x2)	69.25	68.80			73.46
ResNet-32 (x3)	69.40	69.70	68.77		75.38
ResNet-32 (x4)	69.20	69.60	69.45	69.10	76.30

092	Algorithm 1: Training routine for two identical base networks in	an FDL ensemble.
093	Given: Identical networks N_1 and N_2 , Ensemble Head network	k N _E
094	Data: I=Image, y=Label.	
095		
096	Pretrain N_1 for one epoch.	/*Phase 0*/
097	Pretrain N_2 for one epoch.	
098	10	
100	while $iter \leq iter_{max}$ do	
101	$\Delta (M(I))$	(Dhara 1 (
101	$\begin{array}{c} \mathbf{y}_1 \leftarrow I\mathbf{v}_1(I) \\ I \leftarrow I = (\hat{\mathbf{x}} - \mathbf{v}) \end{array}$	/*Phase I*/
102	$L_1 \leftarrow L_X(\mathbf{y}_1, \mathbf{y})$ Back propagate Λ nd U nd at $a_{Y}(\mathbf{I}_{+})$	
103	$\hat{\mathbf{v}}_{2} \leftarrow N_{2}(I)$	
104	$\begin{array}{c} \mathbf{y}_2 \land \mathbf{i}_{\mathbf{y}_2}(\mathbf{i}) \\ \mathbf{I}_2 \leftarrow \mathbf{I}_{\mathbf{y}_1}(\hat{\mathbf{y}}_2 \mathbf{y}) \end{array}$	
105	$B_2 (J_2, J_3)$ BackpropagateAndUpdate _N (L ₂)	
100	$\sum u o n p \circ p o g ou o n n u o p u o n o n o n o n o n o n o n o n o n o$	
107	$L_1 \leftarrow L_X(N_1(I), \mathbf{v})$	/*Phase 2*/
100	$L_2 \leftarrow L_X(N_2(I), \mathbf{y})$	
110	$S^{N_1,N_2} = (L_1 - L_2)^2$	
110	BackpropagateAndUpdate _{N1 N2} (S)	
112		
112	$\mathbf{\hat{y}}_1 \leftarrow N_1(I)$	/*Phase 3*/
11/	$\mathbf{\hat{y}}_2 \leftarrow N_2(I)$	
115	Collect all feature tensors into buckets F_{N_1} and F_{N_2} .	
116	Compute feature difference loss, L_{FDL} from Eq. ??.	
117	$BackpropagateAndUpdate_{N_1,N_2}(-L_{FDL})$	
118		
119	$\hat{\mathbf{y}}_c \leftarrow [N_1(I)^T, N_2(I)^T]^T$	/*Phase 4*/
120	$\hat{\mathbf{y}}_E \leftarrow N_E(\hat{\mathbf{y}}_c)$	
121	$L_E \leftarrow L_X(\hat{\mathbf{y}}_E, \mathbf{y})$	
122	$BackpropagateAndUpdate_{N_E}(L_E)$	
123		
124	<i>iter</i> + +	
125 -	end	
126		
127		
128		
129		
130		
131 ′	2 Hypernarameters	
132		
133		
134		
135		
136		

Model	Dataset	z		Netw Optij	ork 1 nizer	Netw Opti	ork 2 mizer	Netv Opti	vork 3 imizer	O p	work 4 timizer	nis O	ilarity imizer	¯ õ	FDL timizer	Ensen Op	able Head timizer	в
			B	ㅂ	в	<u>н</u>	E	1	в	<u>۲</u>	в	=	в	<u>н</u>	E	ч	в	I
VGG-16	CIFAR-10	x1	128	1e-2	0.9					<u> </u>								
		x2	128	1e-2	0.9	1e-2	0.9	ı	,	ı	,	1e-2	0.9	le-5	0.85	1e-1	0.9	
		x3	128	1e-2	0.9	1e-2	0.9	1e-2	0.9	,		1e-2	0.9	1e-4	0.85	le-1	0.9	
		x4	128	1e-2	0.9	1e-2	0.9	1e-2	0.9	1e-2	0.9	1e-2	0.9	1e-4	0.85	le-2	0.0	Ta
		-	-					-		-				-		-		ble
	CIFAR-100	x1	128	1e-2	0.9	,	,	,	,	,	,	,	,	,	,	,	,	e 4
		X2	128	1e-2	0.9	1e-2	0.9					1e-2	0.9	1e-5	0.85	le-1	0.9	:
		х3	128	1e-2	0.9	1e-2	0.9	1e-2	0.9	,	,	1e-2	0.9	1e-5	0.85	le-1	0.9	H
		x4	128	1e-2	0.9	1e-2	0.9	1e-2	0.9	1e-2	0.0	1e-2	0.9	1e-4	0.85	le-2	0.0	ype
																		erp
ResNet-20	CIFAR-10	x1	128	1e-2	0.9	'	,	,		'	,	•	,	,		,	,	oar
		x2	128	1e-2	0.9	1e-2	0.9					1e-2	0.9	1e-4	0.85	1e-2	0.9	ar
		x3	128	1e-2	0.9	1e-2	0.9	1e-2	0.9			1e-2	0.9	1e-4	0.85	1e-2	0.9	ne
		x4	128	1e-2	0.9	1e-2	0.0	1e-2	0.9	1e-2	0.0	1e-2	0.0	1e-4	0.85	1e-2	0.0	tei
																		s f
	CIFAR-100	x1	128	1e-2	0.9										ı			or
		x2	128	1e-2	0.9	1e-2	0.9					1e-2	0.9	1e-5	0.85	1e-2	0.9	a
		x3	128	1e-2	0.9	1e-2	0.9	1e-2	0.9	,		1e-2	0.9	1e-5	0.85	1e-2	0.9	11 (
		x4	128	1e-2	0.9	1e-2	0.9	1e-2	0.9	1e-2	0.9	1e-2	0.9	1e-5	0.85	1e-2	0.0	exj
						-		-		-		-		-		-		per
ResNet-32	CIFAR-10	x1	128	le-2	0.0		,	,		•				,	ı			in
		x2	128	1e-2	0.9	1e-2	0.9	1	,	ı	,	1e-2	0.9	1e-4	0.85	1e-2	0.9	ne
		. x3	128	le-2	0.9 0.9	le-2	0.9	1e-2	0.9		, (1e-2	0.9	le-4	0.85	le-2	0.9 0.9	nts
		x4	128	le-2	0.9	1e-2	0.9	le-2	0.9	1e-2	0.9	1e-2	0.9	le-4	0.85	le-2	0.0	
		-	901	-	00	_		_		_		_		_		_		arri
		Υ γ	128	1e-2	0.0	- 101	- 00					- 1 1	- 00	- 19-5	- 0.85	- 10	- 00	ed
		ž ž	128	1e-2	0.0	1e-2	0.0	1e-2	0.0	,	,	1e-2	0.9	le-5	0.85	1e-2	6.0	0
		x4	128	1e-2	0.9	1e-2	0.9	1e-2	0.9	1e-2	0.9	1e-2	0.9	1e-5	0.85	le-2	0.9	ut.
ResNet-18	ImageNet	x1	256	1e-1	60	_				_		_				_		
0 1-100 10001	magence	ζ×	256	11	0.0	le-1	0.9					1e-5	0.9	2e-5	0.9	- 1e-4	0.9	
		2	- 2024	1-21	200	1-21	200					2		2	200		200	1
ResNet-50	ImageNet	x1	256	le-1	0.9		,		,		ı		,		,		,	
	0	x2	256	1e-1	0.9	le-1	0.9			'	·	1e-5	0.9	le-1	0.9	le-4	0.9	
																		I
17 17 18 18	17 17 17 17	17 17	17 17 17	16 16	16 16	16 16	16 16 16	18 16	15 15	15 15	15 15 15	15	14	14 14	14 14 14	14	13	13
79 30 31	75 76 77 77	73 74	(U 71 72	58 59	36 37	54 65	51 52 53	59 50	57 58	55 56	52 53 54	51	18 19	16 17	13 14 15	+1 12	39 10	38

AUTHOR(S): BMVC AUTHOR GUIDELINES

184 185	N = Number of networks in the ensemble. B = Batch Size, lr = Learning Rate, m = Momentum
186 187	Other hyperparameters that are common across all experiments are the following:
188 189	Weight Decay = $5e - 4$
190 191	Nesterov = $True$
192	
193	
194	
195	
196	
197	
198	
200	
200	
202	
203	
204	
205	
206	
207	
208	
209	
210	
211	
212	
213	
214	
215	
210	
218	
219	
220	
221	
222	
223	
224	
225	
226	
227	
228	
229	