Supplemental Document

Biomedical Optics EXPRESS

Student becomes teacher: training faster deep learning lightweight networks for automated identification of optical coherence tomography B-scans of interest using a student-teacher framework: supplement

JULIA P. OWEN,[1] MARIAN BLAZES,[1] NIRANCHANA MANIVANNAN,[2] GARY C. LEE,[2] SOPHIA YU,[2] MARY K. DURBIN,[2] ADITYA NAIR,[2] RISHI P. SINGH,[3] KATHERINE E. TALCOTT,[3] ALLINE G. MELO,[3] TYLER GREENLEE,[3] ERIC R. CHEN,[3] THAIS F. CONTI,[3] CECILIA S. LEE,[1] AND AARON Y. LEE[1,*]

[1] *Department of Ophthalmology, University of Washington, Seattle, WA 98195, USA*
[2] *Carl Zeiss Meditec, Inc., Dublin, CA 94568, USA*
[3] *Center for Ophthalmic Bioinformatics, Cole Eye Institute, Cleveland Clinic, Cleveland, OH 44195, USA*
[*] *leeay@uw.edu*

This supplement published with The Optical Society on 2 August 2021 by The Authors under the terms of the Creative Commons Attribution 4.0 License in the format provided by the authors and unedited. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

Supplement DOI: https://doi.org/10.6084/m9.figshare.15019737

Parent Article DOI: https://doi.org/10.1364/BOE.433432

Supplementary Table 1: ResNet and lightweight model details.

Model Family	Model Name	Variation on Family	# Parameters	FLOPS	Model Disk Usage (bytes)	Reference
ResNet	ResNet50	N/A	23,591,810	353,054,355	283,446,328	https://arxiv.org/abs/1512.03385
SqueezeNet	SN.1	Base model	727,042	1,448,162	3,178,024	https://arxiv.org/pdf/1602.07360
	SN.2	SN.1 with fire models redistributed	722,370	2,887,044	3,158,272	
	SRN.1	SN.1 with residual connections	727,042	4,335,206	3,181,488	
	SRN.2	SN.2 with fire models redistributed and residual connections	722,370	5,774,088	3,161,768	
SqueezeNext	SQN.1.1	Base model	600,594	6,946,731	4,439,008	https://arxiv.org/pdf/1803.10615
	SQN.1.2	SQN.1.1 with 1.5x wider channels	1,327,130	9,558,006	7,360,248	
	SQN.1.3	SQN.1.1 l with 2x wider channels	2,337,826	14,176,233	11,390,744	
	SQN.2.1	SQN.1.1 with redistribution of the number of modules at each stage	801,170	15,742,988	5,244,024	
	SQN.2.2	SQN.2.1 with 1.5x wider channels	1,773,146	19,235,735	9,150,640	
	SQN.2.3	SQN.2.1 with 2x wider channels	3,126,050	25,416,330	14,545,592	
MobileNet	M.1.1	Base model	3,230,338	31,844,257	7,983,000	https://arxiv.org/pdf/1704.04861
	M.1.2	M.1.1 with width multiplier of 0.5	830,274	33,488,472	3,799,592	
	FM.1.1	M.1.1 with fast downsampling	1,893,506	37,251,939	7,983,000	https://arxiv.org/pdf/1704.04861
	FM.1.2	FM.1.1 with width multiplier of 0.5	489,538	38,219,278	2,366,736	
	M.2.1	M.1.1 with inverted residual connection and linear bottleneck layers inserted into the convolutional blocks	2,261,120	42,690,419	9,943,928	https://arxiv.org/pdf/1801.04381

	M.2.2	M.2.1 with width multiplier of 0.5	705,376	44,073,784	3,723,104	
	M.3.1	M.2.1 with squeeze and excite in residual layer and layers with modified switch nonlinearities	4,227,426	52,482,036	17,856,104	https://arxiv.org/pdf/1905.02244
	M.3.2	M.3.1 with width multiplier of 0.5	2,069,626	56,596,344	9,223,240	
	M.3.3	M.3.1 with fewer layers	1,678,674	59,929,874	7,472,728	
	M.3.4	M.3.3 with width multiplier of 0.5	1,017,114	61,951,220	4,828,880	
ShuffleNet	SFN1.1	Base model	954,418	63,829,636	4,913,752	https://arxiv.org/pdf/1707.01083
	SFN1.2	SFN1.1 with 4 groups	908,834	65,605,930	5,220,152	
	SFN1.3	SFN1.1 with 8 groups	939,578	67,426,700	6,308,712	
	SFN1.4	SFN1.1 with 1 group and width multiplier 0f 0.5	253,190	67,922,213	1,740,560	
	SFN1.5	SFN1.1 with with 3 groups and width multiplier of 0.5	250,310	68,404,684	2,334,680	
	SFN2.1	SFN1.1 with an additional 1 × 1 convolution layer added before global averaged pooling and width multiplier of 0.41	351,362	69,095,650	2,245,824	https://arxiv.org/pdf/1807.11164
	SFN2.2	SFN1.1 with an additional 1 × 1 convolution layer added before global averaged pooling and width multiplier of 2.1	5,485,488	80,015,907	22,786,296	

Supplementary Table 2: Results of knowledge distillation with temperature adjusted soft-target learning, the first two rows did not use a temperature raised softmax.

Soft targets used	Unlabeled images used	Temperature	Alpha	MaximumValidation Accuracy
No	**No**	**NA**	**NA**	**89.6%**
No	**Yes**	**NA**	**NA**	**95.1%**
Yes	No	2	0.5	88.8%
Yes	No	2	1.0	89.5%
Yes	No	2	1.5	89.1%
Yes	No	3	0.5	89.5%
Yes	No	3	1.0	89.6%
Yes	**No**	**3**	**1.5**	**90.3%**
Yes	No	3	2	88.5%
Yes	No	3	5	88.7%
Yes	No	3	10	89.7%
Yes	No	4	0.5	89.6%
Yes	No	4	1.0	88.5%
Yes	No	4	1.5	90.0%
Yes	No	5	0.5	89.8%
Yes	No	5	1.0	89.4%
Yes	No	5	1.5	89.1%
Yes	No	10	1.5	89.0%
Yes	No	20	1.5	89.7%
Yes	**Yes**	**3**	**1.5**	**94.4%**