Robustness to Distribution Shifts of Compressed Networks for Edge Devices

Lulan Shen, Ali Edalati, Brett Meyer, Warren Gross, James J. Clark

Department of Electrical and Computer Engineering, McGill University, Canada

INTRODUCTION

It is necessary to develop efficient deep neural networks (DNNs) deployed on edge devices with limited computation resources. However, the compressed networks often execute new tasks in the target domain, which is different from the source domain where the original network is trained. We investigate the robustness of compressed networks in two types of data distribution shifts: domain shifts and adversarial perturbations.

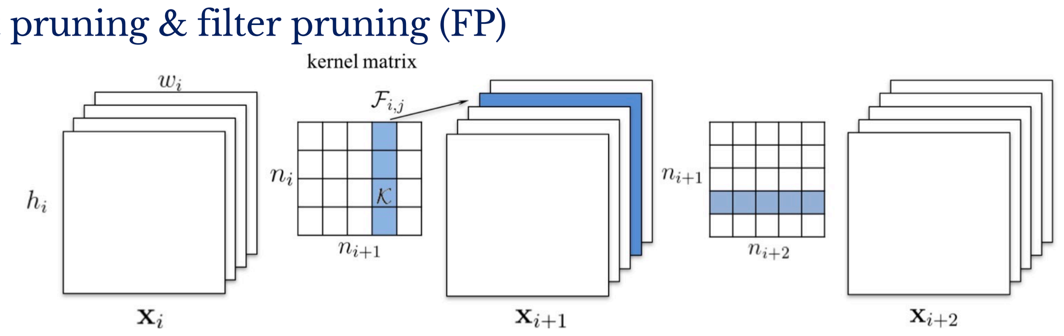
Edge Devices:

BACKGROUND

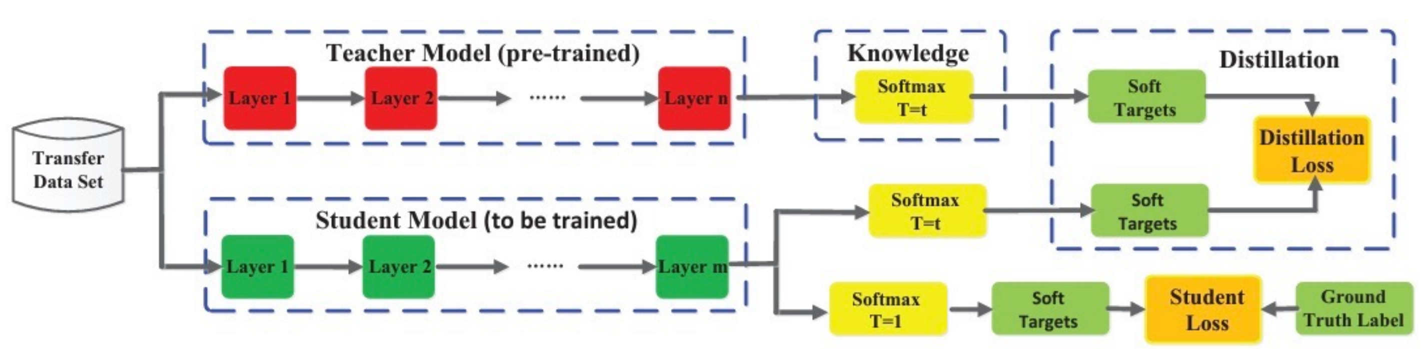
1) *Model Compression:*

- Pruning: weight pruning & filter pruning (FP)

L1-Norm FP (Li et al. 2017)



- Knowledge Distillation (KD) (Hinton et al. 2014)

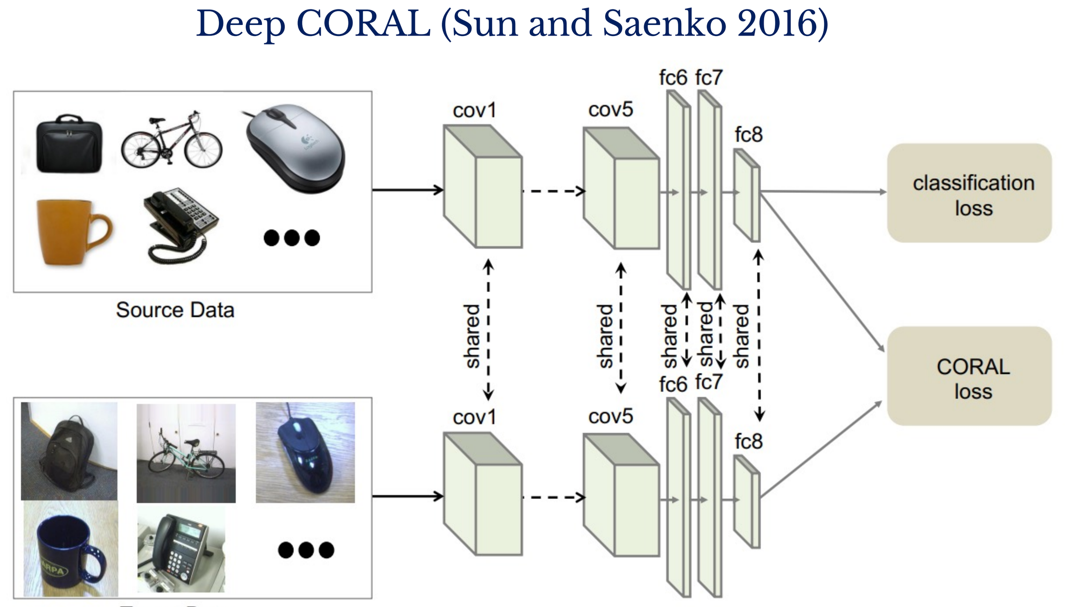


- Quantization: post-training static quantization (PTSQ)
- Low-rank factorization
- Neural Architecture Search (NAS)

The compressed networks often execute new tasks in the target domain, which is different from the source domain where the original network is trained.

2) *Domain Adaptation:*

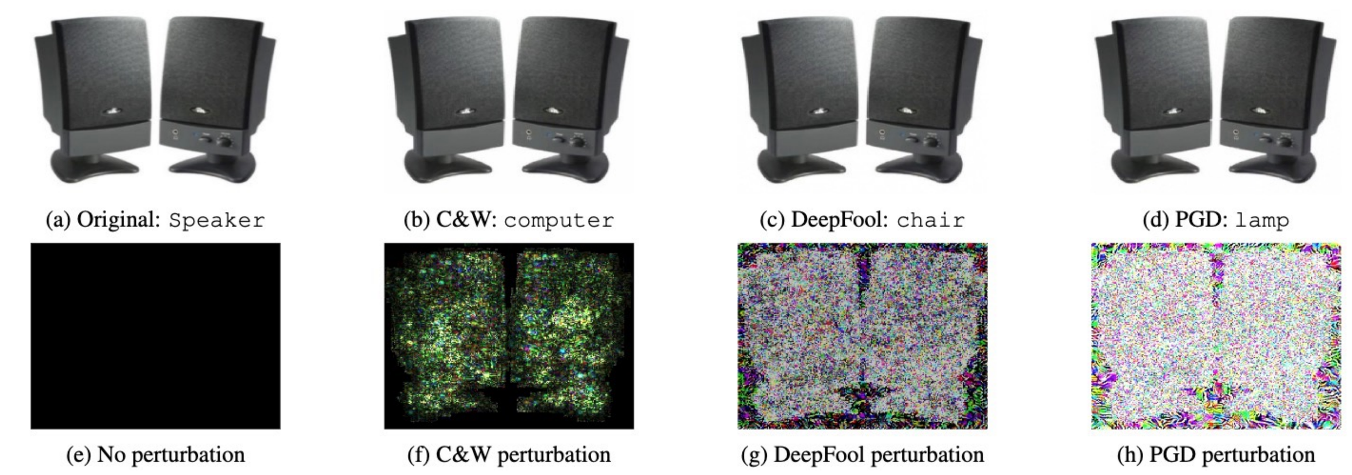
Deep CORAL (Sun and Saenko 2016)



$$L_{coral} = \frac{1}{4d^2}\left\|C_{source} - C_{target}\right\|_F^2, \quad L_{loss} = \sum_{i=1}^{k} \lambda_i L_{coral} = L_{class} + \lambda L_{coral}$$

The current deep domain adaptation methods for computer vision that minimize the distribution difference between the two domains do not consider network compression.

3) *Adversarial Attacks:*



(a) Original: `speaker` (b) C&W: `computer` (c) DeepFool: `chair` (d) PGD: `lamp`

(e) No perturbation (f) C&W perturbation (g) DeepFool perturbation (h) PGD perturbation

EXPERIMENTS

Model Initialization: Initialize ResNets with parameters pre-trained on the ImageNet

↓

Model baselines: Fine-tune on each source domain → Test on each target domain

↓

Model Compression: Obtain the compact model using FP/KD/PTSQ

↓

Domain Adpataion: Apply deep CORAL method (Sun and Saenko 2016) → Test on each target domain

Office-31 dataset (Saenko et al. 2010)

We take one domain as the source domain and one of the other as the target domain. In total, six domain shifts.

RESULTS – DOMAIN SHIFTS

Base Model	Compression Method	# Params (M)	A → W	A → D	W → A	W → D	D → A	D → W	Avg Acc
ResNet-18	-	11.19	64.15	60.84	50.05	97.39	47.43	92.70	68.76
	FP	4.51	57.36	59.64	47.39	96.99	44.73	91.32	66.24
	PTSQ	11.19(/4)	63.02	65.26	-	-	-	-	64.14
ResNet-34	-	21.30	67.80	68.27	54.67	98.59	52.01	92.70	72.34
	FP	11.19	48.93	51.61	46.22	95.78	43.66	89.69	62.65
	FP	4.51	16.73	14.06	16.68	63.86	10.47	57.36	29.86
	KD (ResNet18)	11.19	62.52	63.25	48.49	99.00	52.79	95.47	**70.25**
	PTSQ	21.3(/4)	64.28	62.47	-	-	-	-	63.38
ResNet-50	-	23.57	68.93	71.89	64.25	99.00	60.95	94.84	76.64
	FP	21.30	69.43	72.69	61.52	98.39	59.53	94.34	75.98
	FP	11.19	53.46	61.45	41.36	93.57	35.64	85.28	61.79
	FP	4.51	19.37	26.10	11.89	50.00	9.23	34.72	25.22
	KD (ResNet34)	21.30	57.74	59.04	56.59	98.80	58.18	97.11	**71.24**
	KD (ResNet18)	11.19	57.99	57.63	52.18	99.40	48.78	95.47	**68.58**
	PTSQ	23.57(/4)	74.45	72.69	-	-	-	-	73.57
ResNet-101	-	42.56	74.59	75.70	64.32	99.00	24.35	90.19	71.36
	FP	23.57	62.89	63.45	54.10	98.59	55.80	91.19	71.00
	FP	11.19	43.02	40.16	28.19	90.36	19.63	84.65	51.00
	FP	4.51	19.37	17.27	12.11	57.83	5.18	32.58	24.10
	PTSQ	42.56(/4)	74.21	74.50	-	-	-	-	74.36
ResNet-152	-	58.21	74.97	74.70	63.97	98.39	63.76	95.72	78.59
	FP	42.56	71.19	72.69	61.95	99.00	61.59	93.46	76.65
	FP	23.57	65.41	66.47	55.13	99.00	43.02	86.79	69.30
	FP	11.19	36.23	39.36	25.49	81.93	10.69	57.11	41.80
	FP	4.51	17.86	15.26	9.83	50.00	4.40	23.02	20.06
	KD (ResNet18)	11.19	58.49	62.25	48.49	98.39	50.80	95.85	**69.05**
	PTSQ	58.21(/4)	73.46	75.90	-	-	-	-	74.68

Table 1: The *test* accuracies (%) of ResNets on *target* domains of the Office-31 dataset other than what they were trained on. The baseline (uncompressed) ResNets are obtained after fine-tuning the pre-trained model on the ImageNet dataset. The pruned models are obtained using the $\mathcal{L}1$-FP method with different pruning ratios, and the distilled/student models are obtained using teacher networks of different sizes."A", "W", and "D" represent the domain of Amazon, Webcam, and DSLR.

RESULTS – ADVERSARIAL ATTACKS

Base/Teacher Model	Compr. Method	# Params (M)	In-Domain $\mathcal{D}_A$ Acc (%)	DeepFool$_{L_\infty}$ (ϵ0.004)	PGD$_{L_\infty}$ (ϵ0.004)	FGSM (ϵ0.004)	C&W$_{L_2}$ (ϵ0.4)	DDN (ϵ0.24)	EAD (ϵ10)	SP (ϵ15)
ResNet-18	-	11.19	89.32	2.13	1.78	34.52	14.23	20.64	50.89	31.67
	FP	4.51	89.68	20.28	11.73	49.47	38.43	48.75	62.28	16.73
	PTSQ	11.19(/4)	88.26	-	-	-	87.9	-	87.9	44.84
ResNet-34	-	21.30	90.75	3.91	0.36	47.69	14.95	18.51	49.47	50.89
	FP	11.19	85.41	40.57	32.38	58.01	48.75	58.36	64.41	40.93
	FP	4.51	75.44	30.60	21.71	46.95	40.21	49.82	48.04	21.71
	KD	11.19	90.04	66.90	61.92	75.44	70.82	76.16	75.80	61.57
	PTSQ	23.57(/4)	89.32	-	-	-	88.97	-	88.79	46.98
ResNet-50	-	23.57	90.04	0.36	0.00	45.51	13.17	6.05	48.04	42.35
	FP	21.30	92.17	6.76	1.78	53.74	22.06	26.33	57.30	55.16
	FP	11.19	87.90	0.71	0.35	32.38	4.63	1.78	37.37	40.57
	FP	4.51	75.44	0.00	0.00	9.25	0.71	0.00	17.08	12.46
	KD	21.30	90.04	69.03	64.06	73.31	70.82	77.58	76.87	66.55
	KD	11.19	90.39	62.63	58.72	71.89	67.97	75.80	75.44	49.11
	PTSQ	23.57(/4)	87.19	-	-	-	87.9	-	87.9	58.01
ResNet-101	-	42.56	91.81	6.05	1.78	59.07	12.10	11.39	57.30	64.41
	FP	23.57	90.04	41.28	27.76	66.19	53.38	62.28	70.46	56.23
	FP	11.19	86.12	28.47	17.44	49.82	43.42	53.38	56.94	27.40
	FP	4.51	72.95	2.13	0.71	13.52	10.68	12.81	19.93	6.05
	PTSQ	42.56(/4)	89.68	-	-	-	88.26	-	87.54	61.92
ResNet-152	-	58.21	90.75	7.12	4.98	63.34	15.30	12.46	61.92	65.12
	FP	42.56	91.46	5.69	1.07	61.56	9.96	7.12	56.23	52.67
	FP	23.57	89.68	1.42	0.71	55.87	7.83	5.34	50.53	35.59
	FP	11.19	82.56	0.35	0.00	29.54	4.27	1.07	22.42	7.12
	KD	11.19	90.04	69.75	65.48	75.80	72.95	78.65	77.94	53.74
	PTSQ	58.21(/4)	88.26	-	-	-	89.32	-	88.97	58.72

Table 2: Accuracies of baseline-A, quantized and compressed ResNets under *heavy* adversarial perturbations.

CONCLUSION

- As the compression ratio increases, the compressed models perform more poorly in the unseen domain due to distribution shifts.
- Compressed networks originating from smaller models demonstrate better generalization abilities in the target domain, indicating that they are more robust to distribution shifts compared to networks that were originally as large.
- The pruning technique is known for generating highly sensitive compressed networks that are vulnerable to domain shifts and adversarial perturbations. On the other hand, compact networks produced through KD are less affected by these issues.
- The quantized networks (compressed to ~25% of their original size) offer significantly more robustness to distribution shifts, particularly in the case of domain shifts, than other compressed networks.

REFERENCE

- Li, H.; Kadav, A.; Durdanovic, I.; Samet, H.; and Graf, H. P. 2017. Pruning Filters for Efficient ConvNets. In *International Conference on Learning Representations*.
- G.Hinton; O.Vinyals; and J.Dean. Distilling the knowledge in a neural network, in *Advances in Neural Information Processing Systems Workshop*, 2014.
- Sun, B.; and Saenko, K. 2016. Deep CORAL: Correlation Alignment for Deep Domain Adaptation. In *European Conference on Computer Vision Workshop.β*
- Saenko, K.; Kulis, B.; Fritz, M.; and Darrell, T. 2010. *Adapting Visual Category Models to New Domains.* In *European Conference on Computer Vision.*

ACKNOWLEDGEMENT

Thank both MEIL (McGill Edge Intelligence Lab) and CIM (Center for Intelligent Machines) for powerful local computing resources and for dedicating significant advanced computing resources for research from Compute Canada. I am grateful to MEDA (McGill Engineering Doctoral Award) and FRQNT (Fonds de Recherche du Québec Nature et Technologies) for providing grants that funded my research.