NADS: Neural Architecture Distribution Search for Uncertainty Awareness

Randy Ardywibowo [1] **Shahin Boluki** [1] **Xinyu Gong** [2] **Zhangyang Wang** [2] **Xiaoning Qian** [1]

1. Fixed Model Ablation Study

Table 1. OoD detection results on various training and testing experiments comparing our method with a baseline ensembling method that uses a fixed architecture trained multiple times with different random initializations.

D_{in}	D_{out}	FPR% at TPR 95%		AUROC%		AUPR%	
		Base Ensemble	Ours	Base Ensemble	Ours	Base Ensemble	Ours
CIFAR10	SVHN	50.07	**17.05**	93.48	**97.65**	95.98	**99.07**
	Texture	6.22	**0.25**	97.68	**99.81**	97.44	**99.86**
	Places365	1.03	**0.00**	99.59	**100**	99.97	**100**
	LSUN	34.35	**0.44**	91.55	**99.83**	92.15	**99.89**
	CIFAR100	65.13	**36.36**	78.44	**91.23**	79.44	**91.60**
	Gaussian	**0.00**	**0.00**	**100**	**100**	**100**	**100**
	Rademacher	**0.00**	**0.00**	**100**	**100**	**100**	**100**

[1]Department of Electrical and Computer Engineering, Texas A&M University, College Station, Texas, USA [2]Department of Computer Science and Engineering, Texas A&M University, College Station, Texas, USA. Correspondence to: Randy Ardywibowo <randyardywibowo@tamu.edu>.

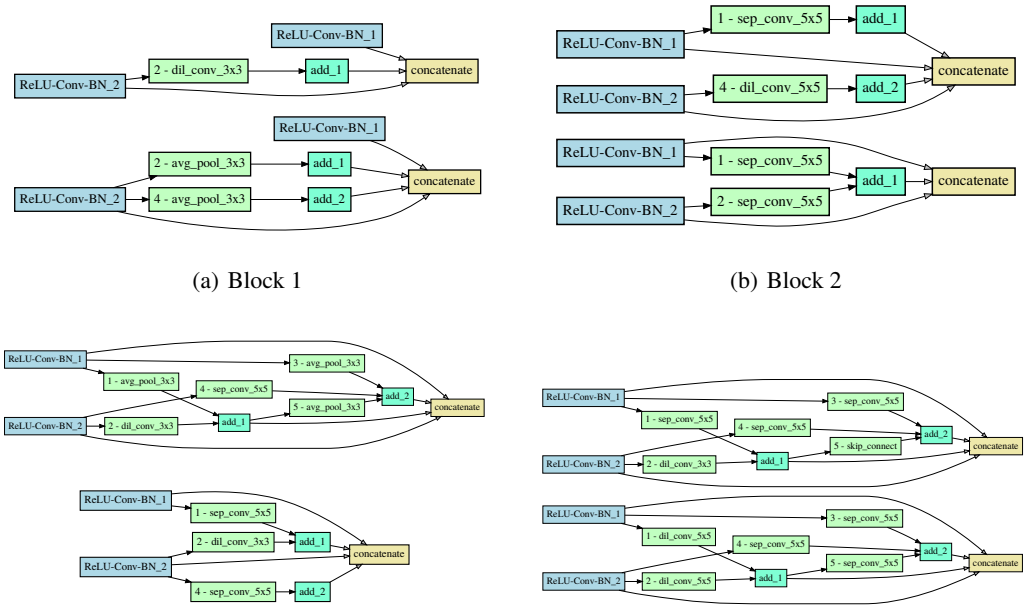
Proceedings of the 37^{th} International Conference on Machine Learning, Online, PMLR 119, 2020.

2. OoD Detection Performance Comparison with ODIN

Table 2. OoD detection results on various training and testing experiments comparing our method with ODIN (Liang et al., 2017).

D_{in}	D_{out}	FPR% at TPR 95%		AUROC%		AUPR%	
		ODIN	Ours	ODIN	Ours	ODIN	Ours
MNIST	not-MNIST	8.7	**0.00**	98.2	**100**	98.0	**100**
	F-MNIST	65	**0.00**	88.6	**100**	90.5	**100**
	K-MNIST	36.5	**0.76**	94.0	**99.80**	94.6	**99.84**
SVHN	Texture	33.9	**0.07**	92.4	**99.26**	88.2	**97.75**
	Places365	22.2	**0.00**	94.9	**99.99**	99.8	**99.99**
	LSUN	26.8	**0.02**	93.5	**99.99**	93.1	**99.99**
	CIFAR10	21.6	**0.37**	94.8	**99.92**	94.4	**99.83**
CIFAR10	SVHN	36.5	**17.05**	89.7	**97.65**	95.6	**99.07**
	Texture	76.2	**0.25**	81.4	**99.81**	76.7	**99.86**
	Places365	44.0	**0.00**	89.0	**100**	99.6	**100**
	LSUN	3.9	**0.44**	99.2	**99.83**	99.2	**99.89**
	CIFAR100	45.4	**36.36**	88.3	**91.23**	88.5	**91.60**
	Gaussian	0.1	**0.00**	**100**	**100**	99.9	**100**
	Rademacher	0.3	**0.00**	99.9	**100**	99.8	**100**
CIFAR100	SVHN	**32.8**	45.92	90.3	**94.35**	95.3	**96.01**
	Texture	78.9	**0.42**	75.7	**99.76**	64.5	**99.81**
	Places365	63.3	**0.012**	79.0	**99.99**	99.1	**99.99**
	LSUN	**17.6**	38.85	**96.8**	90.65	**96.5**	90.61
	CIFAR10	78.2	**45.62**	70.6	**83.27**	69.7	**81.48**
	Gaussian	1.3	**0.00**	99.5	**100**	97.8	**100**
	Rademacher	13.8	**0.00**	92.7	**100**	75.0	**100**

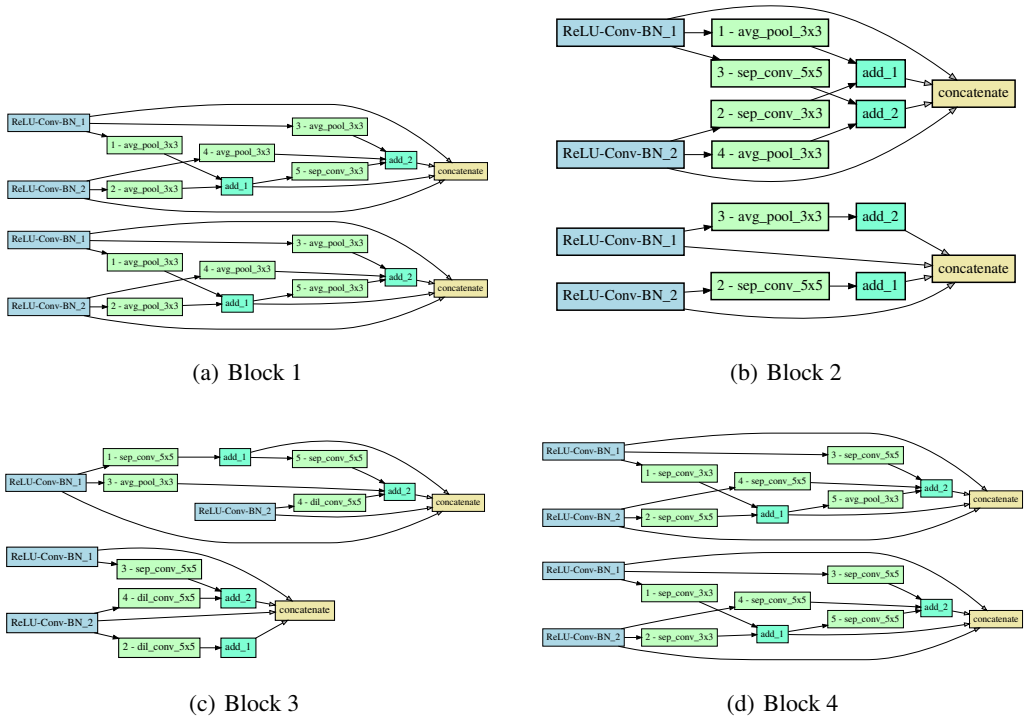
3. Additional Sample Architectures



(a) Block 1 (b) Block 2

(c) Block 3 (d) Block 4

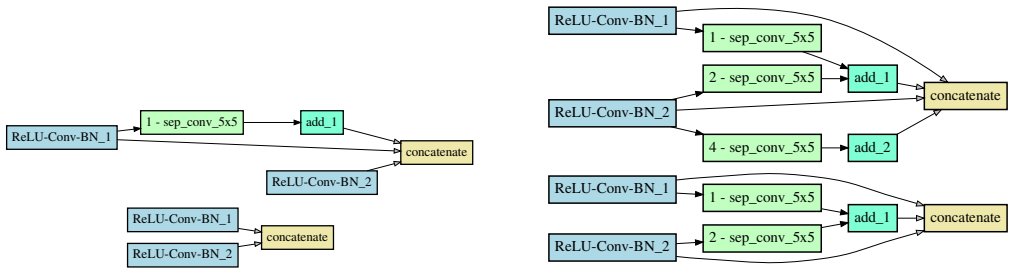
Figure 1. Maximum likelihood architectures inferred by our search algorithm on CelebA. Shown are two samples taken from each block.



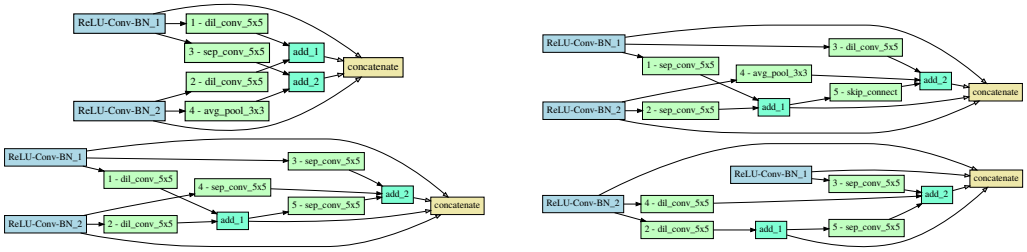
(a) Block 1 (b) Block 2

(c) Block 3 (d) Block 4

Figure 2. Maximum likelihood architectures inferred by our search algorithm on MNIST. Shown are two samples taken from each block.

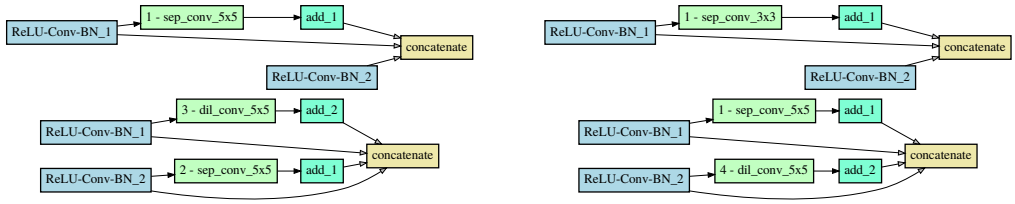


(a) Block 1 (b) Block 2

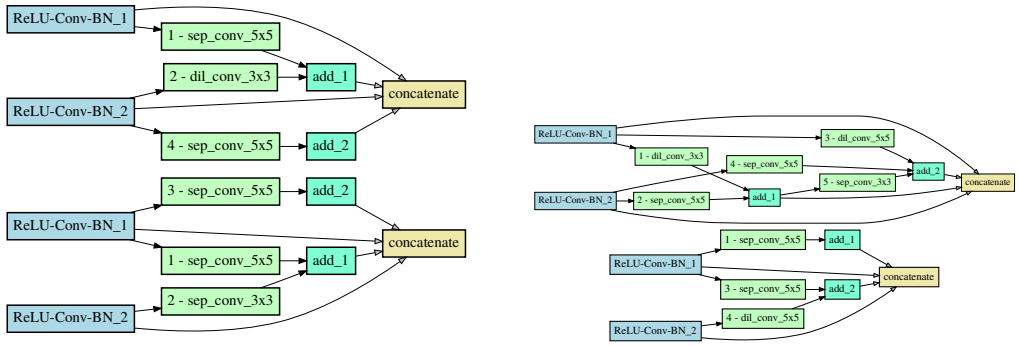


(c) Block 3 (d) Block 4

Figure 3. Maximum likelihood architectures inferred by our search algorithm on SVHN. Shown are two samples taken from each block.



(a) Block 1 (b) Block 2



(c) Block 3 (d) Block 4

Figure 4. Maximum likelihood architectures inferred by our search algorithm on CIFAR-10. Shown are two samples taken from each block.

4. Image Generation Samples

Figure 5. Samples taken from randomly sampled NADS architectures searched on CelebA. Images were not cherry-picked and the architectures were sampled without further retraining.

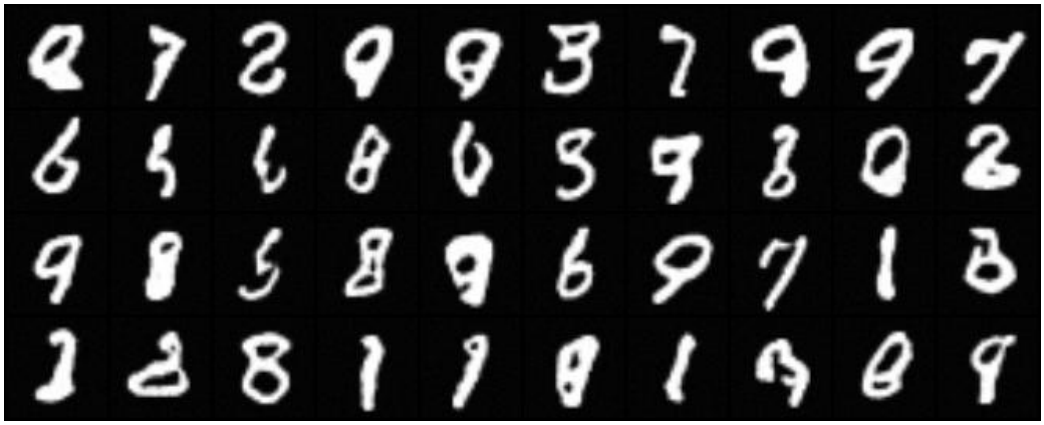


Figure 6. Samples taken from randomly sampled NADS architectures searched on MNIST. Images were not cherry-picked and the architectures were sampled without further retraining.

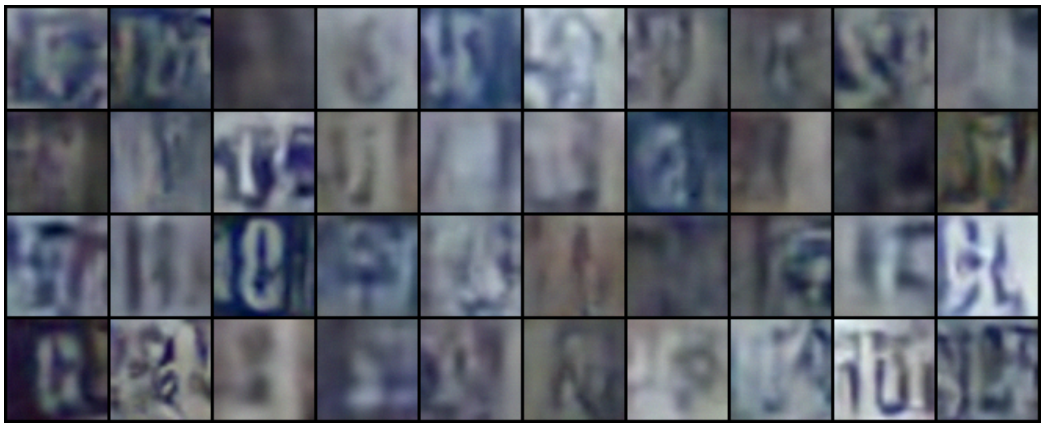


Figure 7. Samples taken from randomly sampled NADS architectures searched on SVHN. Images were not cherry-picked and the architectures were sampled without further retraining.

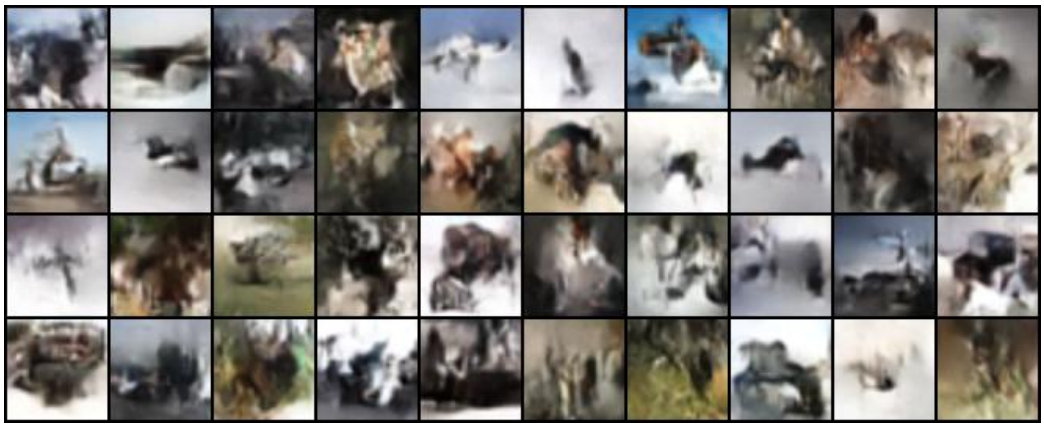


Figure 8. Samples taken from randomly sampled NADS architectures searched on CIFAR-10. Images were not cherry-picked and the architectures were sampled without further retraining.

5. Likelihood Estimation Models Assign Higher Likelihood to OoD Data

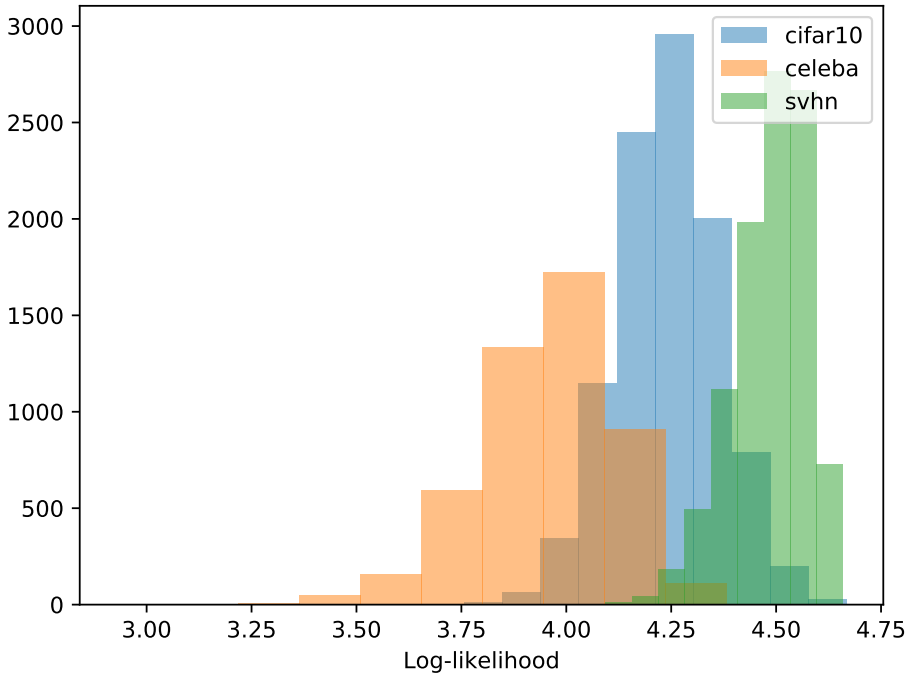


Figure 9. Likelihood distributions of different datasets evaluated on a Glow model trained on CelebA. The model assigns higher likelihood to OoD samples from CIFAR-10 and SVHN.

6. Effect of Ensemble Size

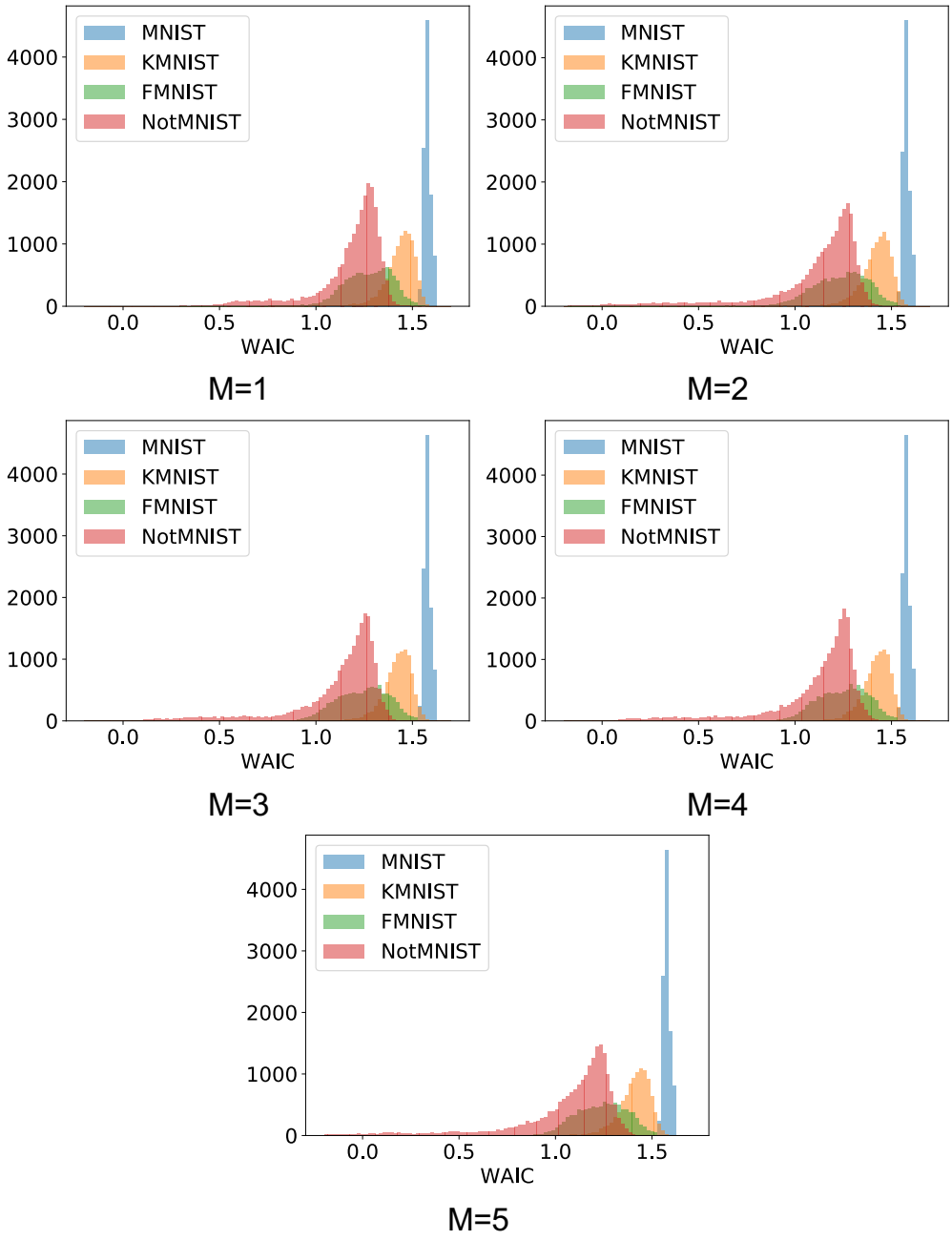


Figure 10. Effect of ensemble size to the distribution of WAIC scores estimated by model ensembles trained on MNIST.

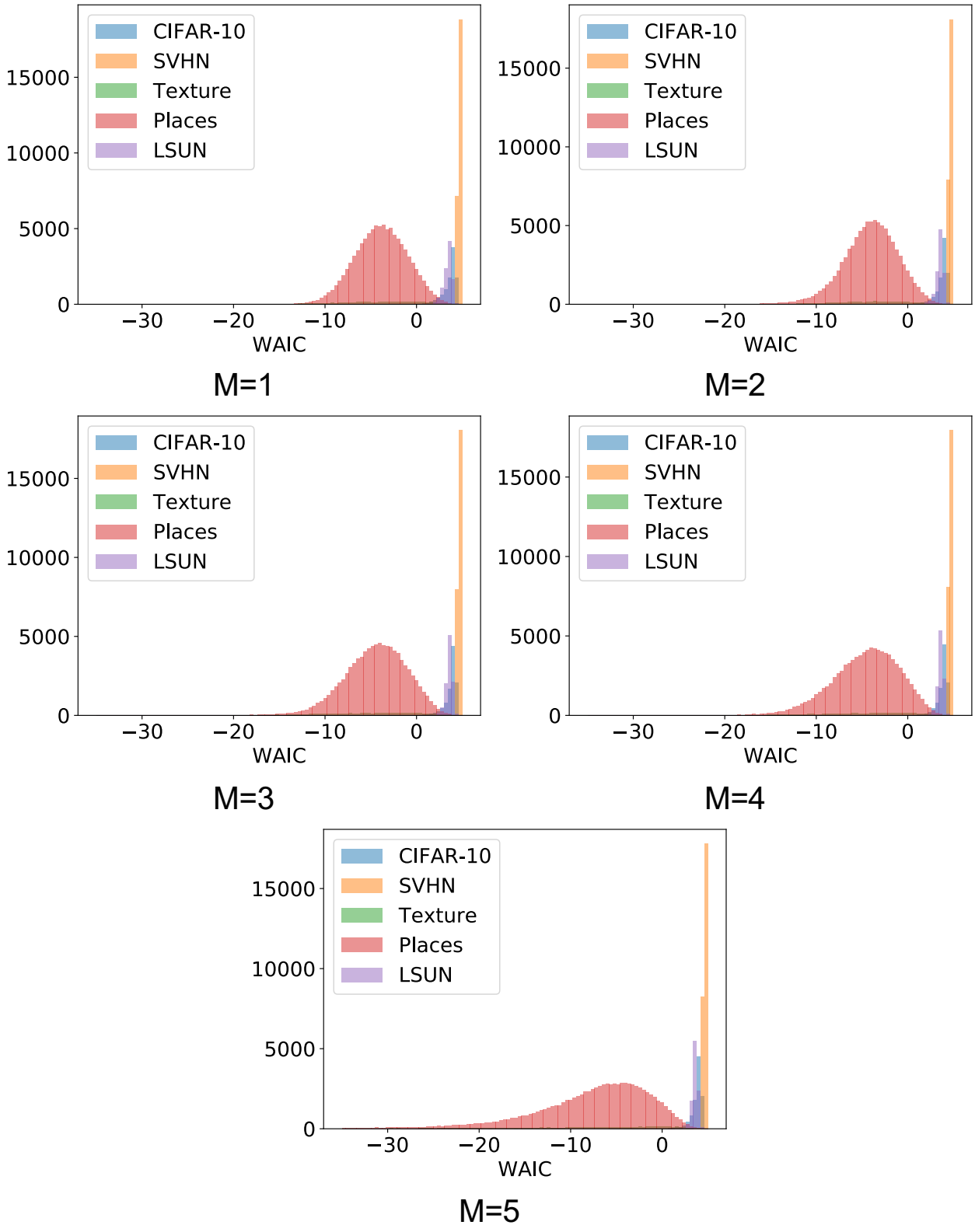


Figure 11. Effect of ensemble size to the distribution of WAIC scores estimated by model ensembles trained on SVHN.

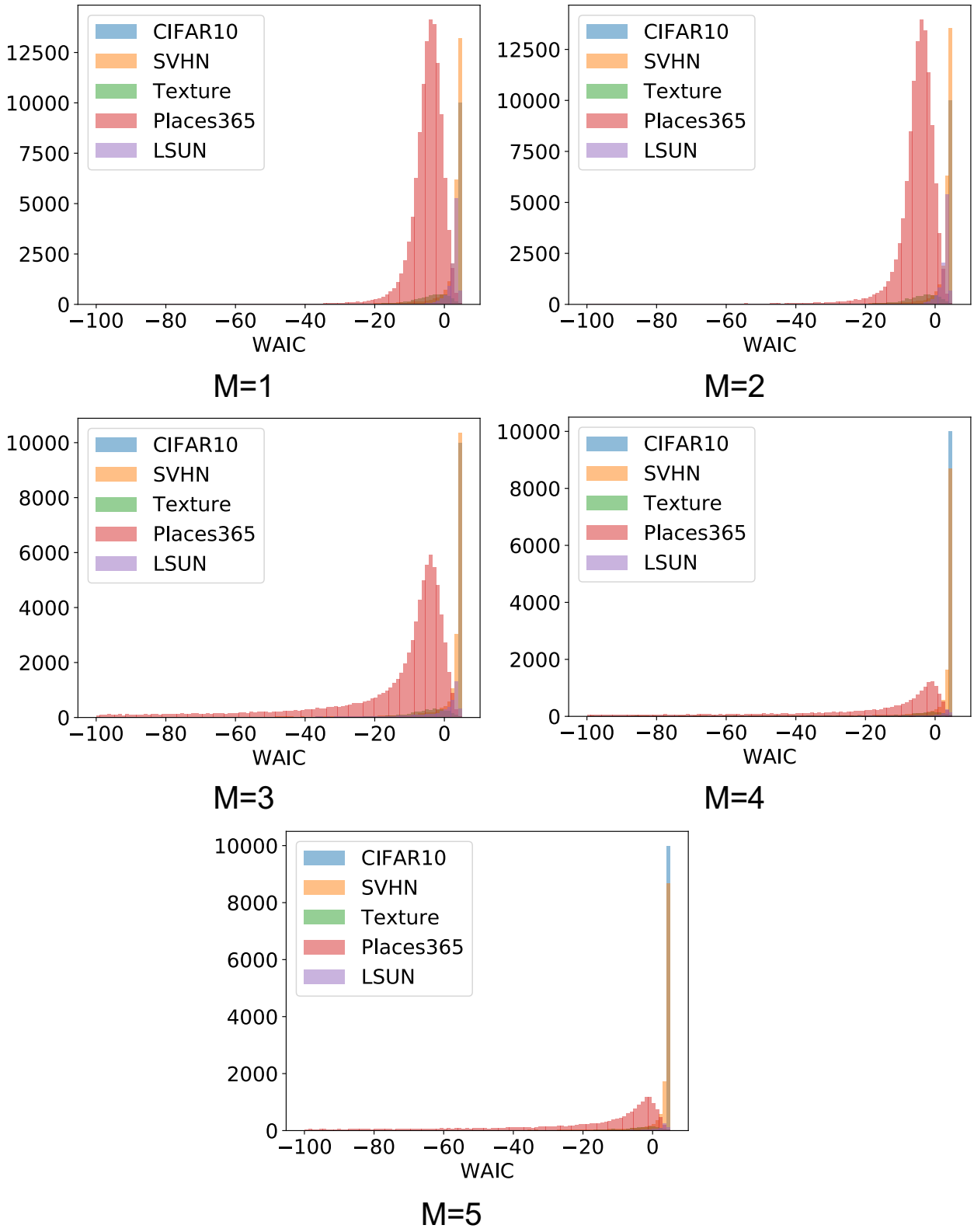
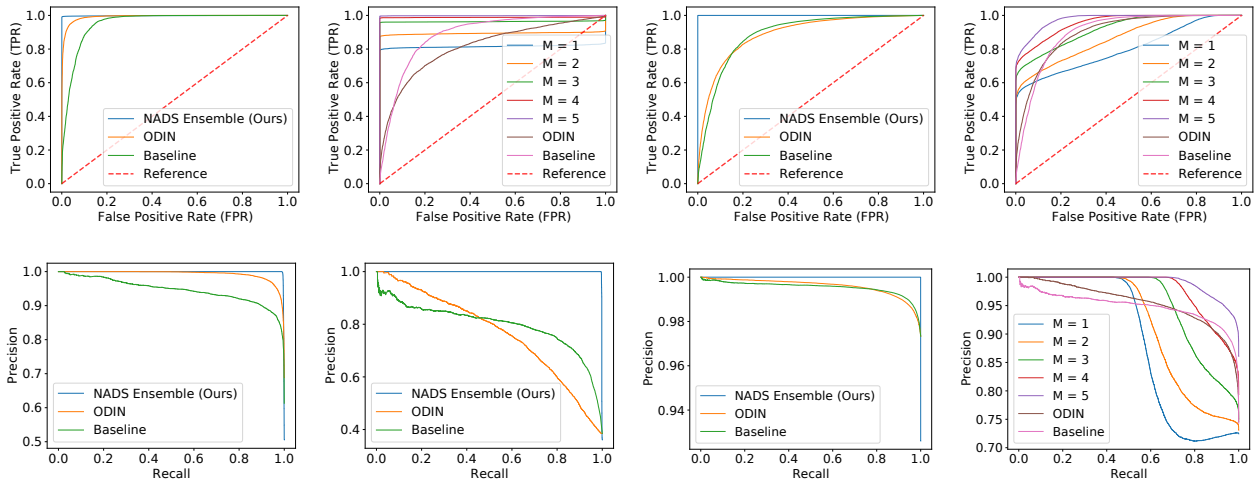


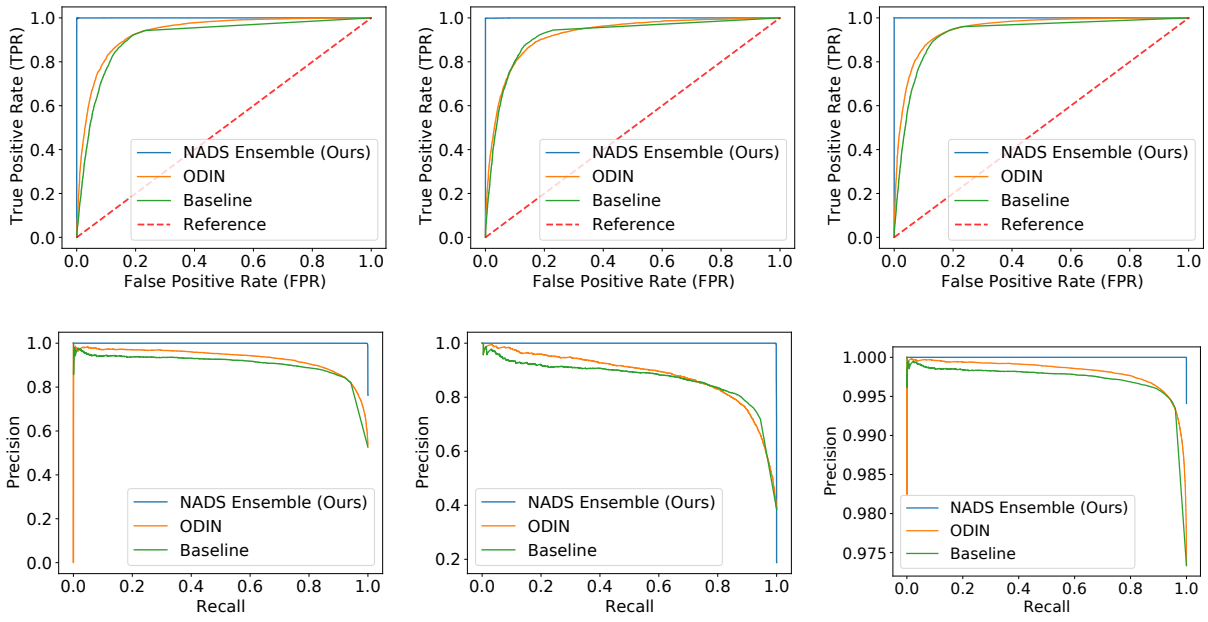
Figure 12. Effect of ensemble size to the distribution of WAIC scores estimated by model ensembles trained on CIFAR-10.

7. Additional ROC and Precision-Recall Curves



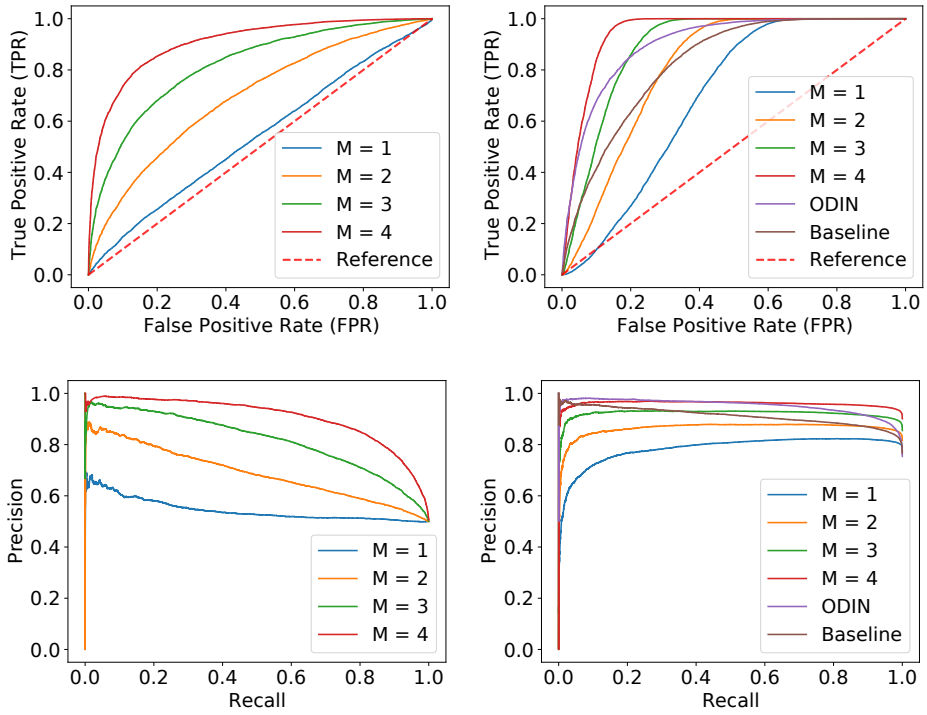
(a) LSUN (b) Texture (c) Places (d) SVHN

Figure 13. ROC and PR curve comparison of methods trained on CIFAR-10



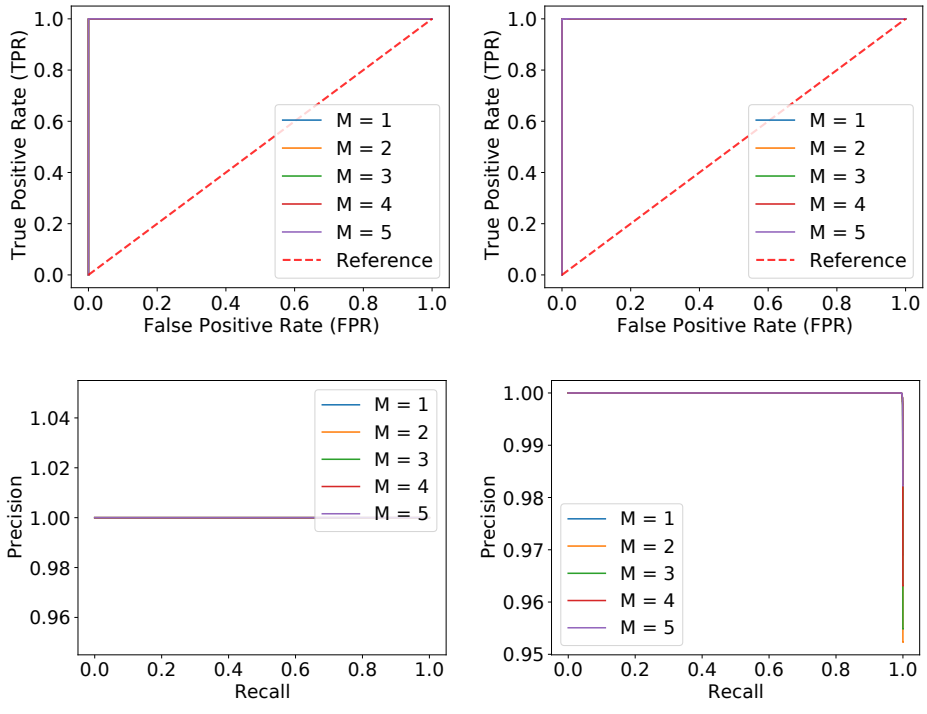
(a) LSUN (b) Texture (c) Places

Figure 14. ROC and PR curve comparison of methods trained on SVHN



(a) LSUN (b) SVHN

Figure 15. ROC and PR curve comparison of methods trained on CIFAR-100



(a) NotMNIST (b) FMNIST

Figure 16. ROC and PR curve comparison of methods trained on MNIST

References

Liang, S., Li, Y., and Srikant, R. Enhancing the reliability of out-of-distribution image detection in neural networks. *arXiv preprint arXiv:1706.02690*, 2017.