On the Interplay of Convolutional Padding and Adversarial Robustness

Paul Gavrikov [1] Janis Keuper [1,2]

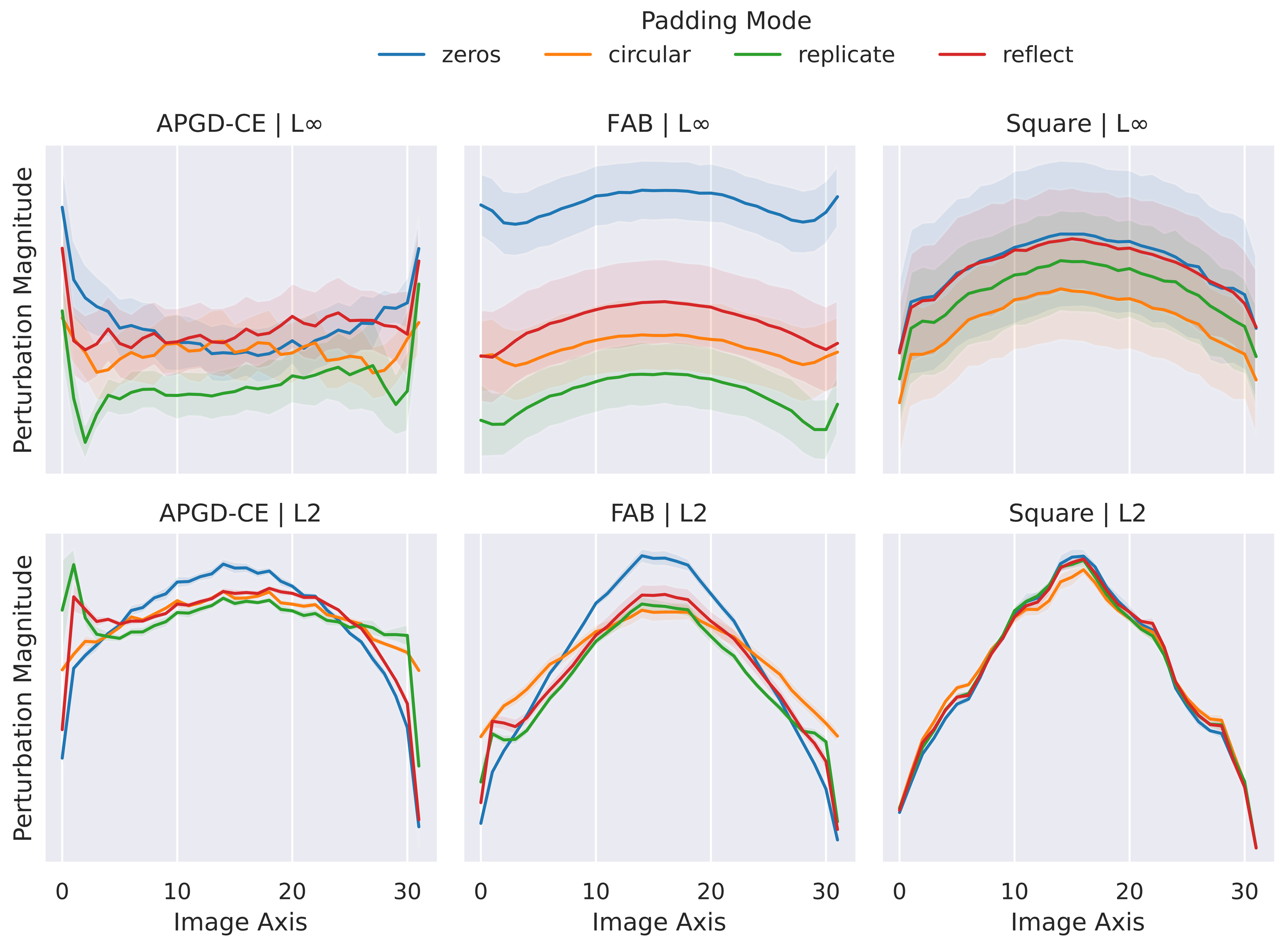
[1]IMLA, Offenburg University [2]Fraunhofer ITWM

Paper

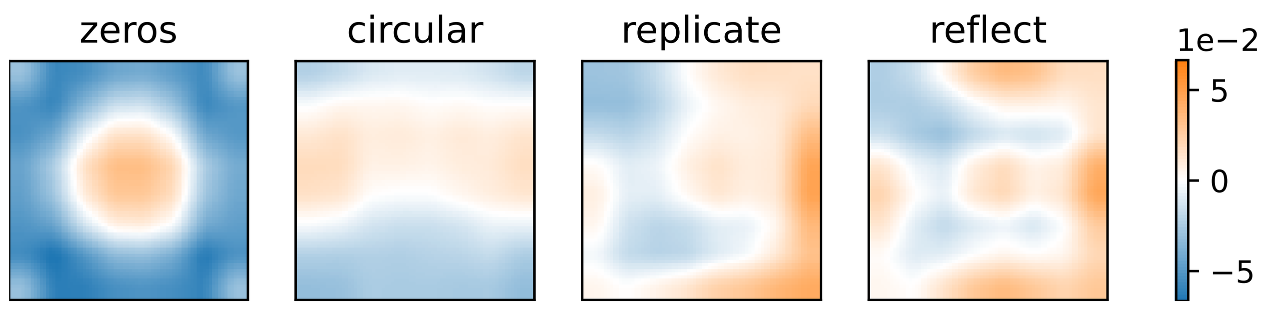
The padding mode is an essential yet rarely tuned CNN hyperparameter. How does its choice affect robustness?

Normal Training

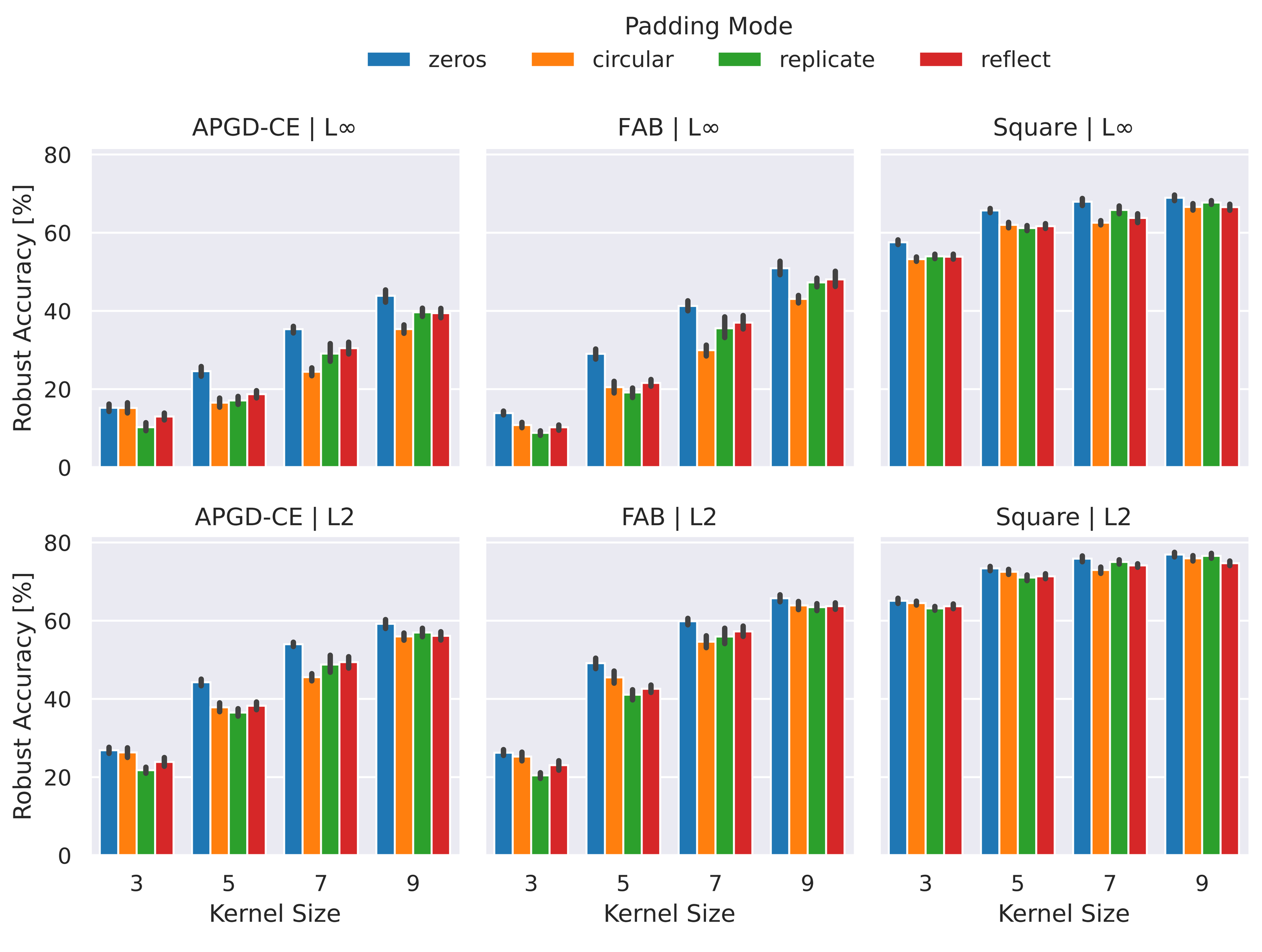
Anomalies in Perturbations in Padding Regions.



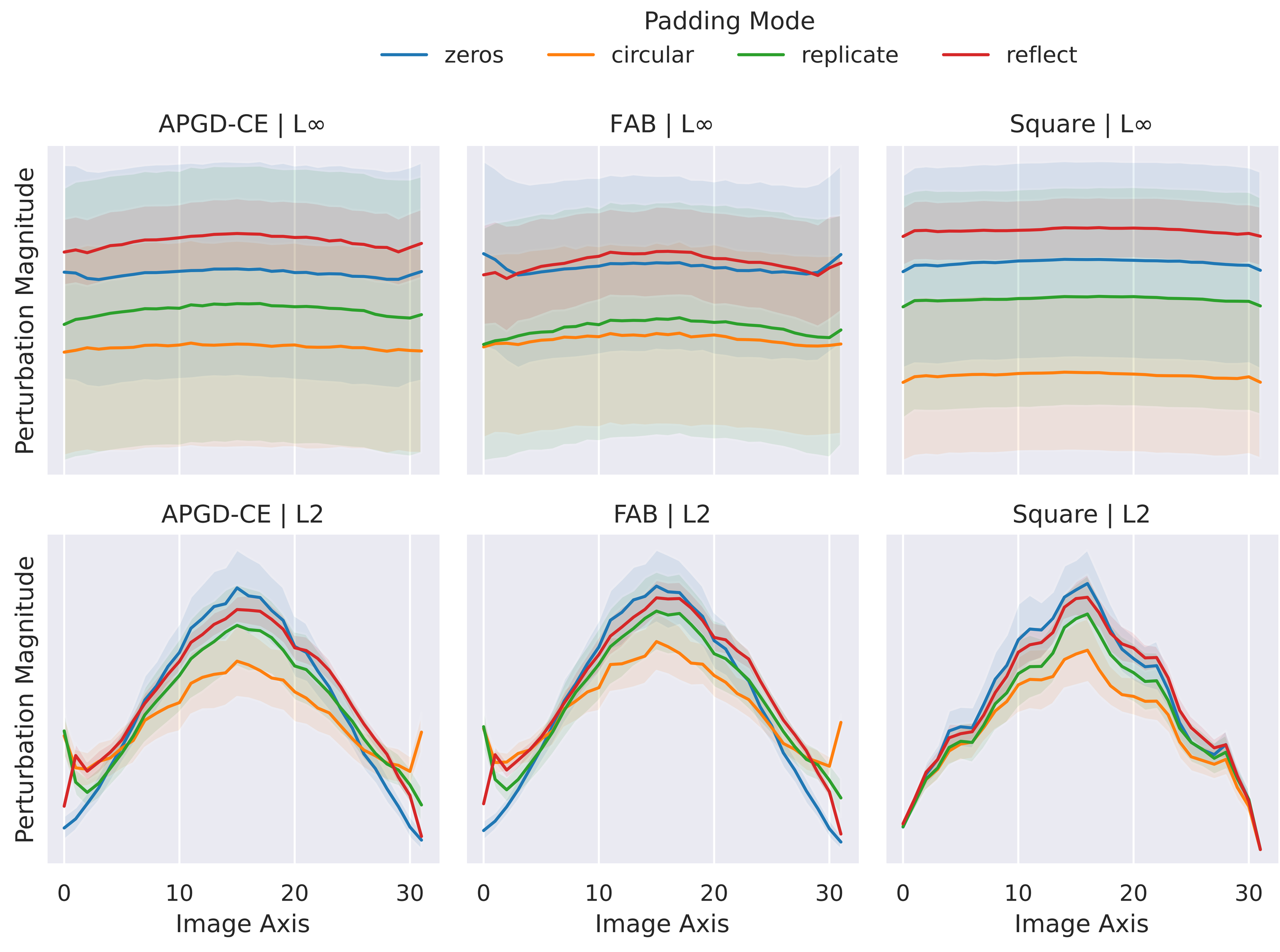
Effect on Model Decisions.

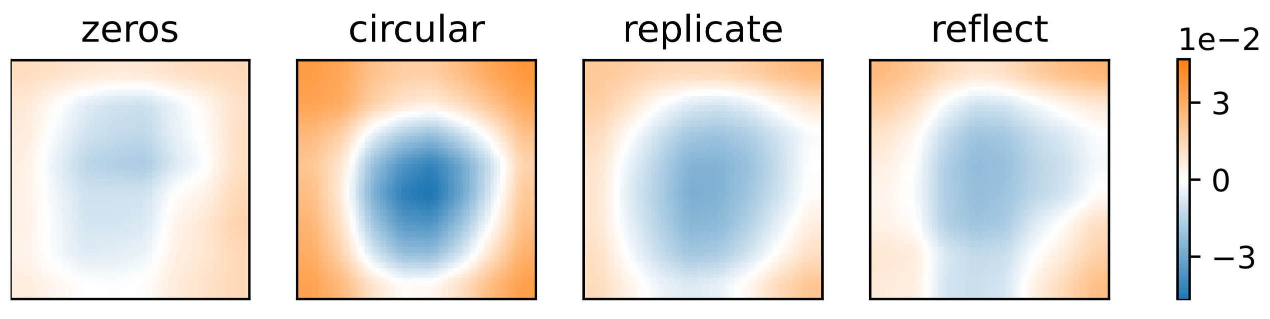


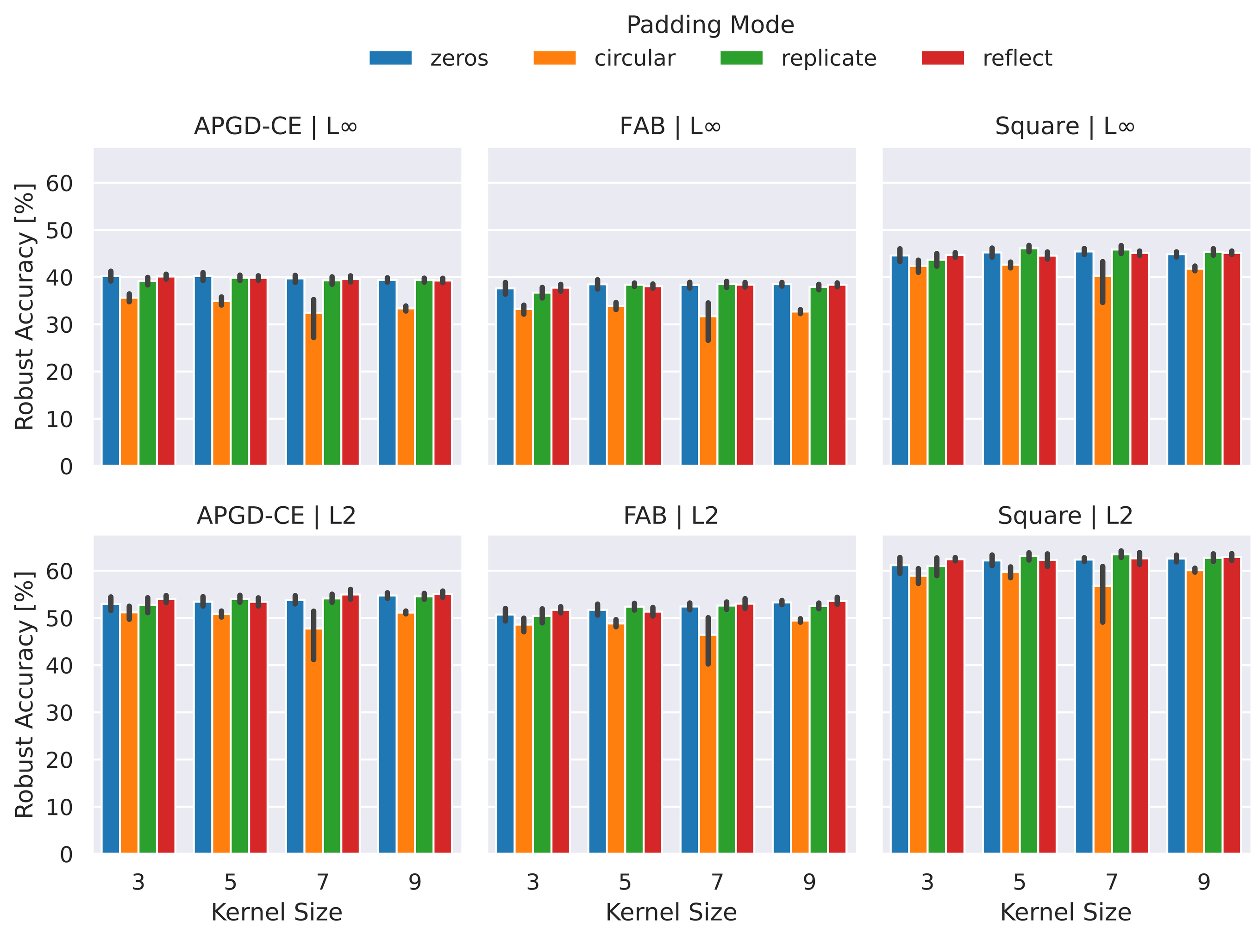
Robust Performance.



Adversarial Training







Comparison

		Clean Test [%] (↑)				AutoAttack [%] (↑)			
AT	k	zeros	circular	replicate	reflect	zeros	circular	replicate	reflect
✗	3	**90.26**	90.10	90.13	90.15	**8.52**	4.69	4.90	5.79
	5	**90.14**	89.66	89.82	89.67	**17.69**	10.44	11.12	12.33
	7	**89.36**	88.49	88.52	88.47	**29.06**	17.86	24.55	24.35
	9	**88.22**	87.50	87.03	87.25	**39.18**	30.52	36.39	34.81
✓	3	71.84	69.17	70.79	**73.11**	**36.88**	32.09	35.91	36.82
	5	73.72	71.34	**74.02**	73.08	**37.48**	32.34	37.30	37.12
	7	73.86	67.33	**73.89**	73.10	**37.42**	30.16	37.08	37.26
	9	73.51	71.53	72.24	**73.90**	**37.49**	31.09	36.89	37.25

Take-Home Messages

- Padding results in anomalies in the spatial distribution of adversarial attacks.
- Increasing the kernel size (and padding) natively improves robustness without adversarial training.
- *Zero* padding performs best in, both, clean and adversarial evaluation with normal training.
- Adversarial training balances the robust performance under different padding modes (except *circular*) and kernel sizes.
- When using adversarial training, *replicate*/*reflect* notably improves clean performance with marginal impairments in robust performance compared to *zero* padding.
- Padding is an essential operation. Removing padding results in deteriorated performance in clean and adversarial settings.
- Limitation: We only studied image classification on CIFAR-10 with ResNet-20. As with many "toy datasets", objects in question are usually perfectly centered in the images → not clear if the results transfer to real-world scenarios.