Towards Understanding Adversarial Robustness of Optical Flow Networks

Simon Schrodi Tonmoy Saikia Thomas Brox

University of Freiburg, Germany

schrodi,saikiat,brox@cs.uni-freiburg.de

[o] Vision
COMPUTER VISION University of Freiburg

Paper

Code

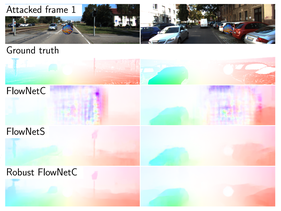
CVPR JUNE 19-24 2022 NEW ORLEANS • LOUISIANA

Motivation

Ranjan et al. found FlowNetC to be vulnerable to a physical, patch-based adversarial attack

Is encoder-decoder architecture the actual cause as suggested by Ranjan et al.?

Knowing the actual cause, we can fix the vulnerability

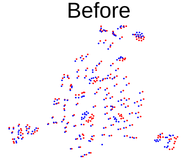


Network	Unattacked EPE	Attacked EPE 102x102	Attacked EPE 153x153
FlowNetC	11.50	52.66	51.99
FlowNetS	14.33	17.35	17.92
Robust FlowNetC	9.95	12.33	12.20

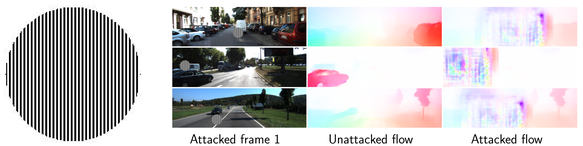
Analysis

Attacked features are separated from the unattacked ones by the correlation layer

Replace features of	Without replacement	With replacement
conv3(a,b)	25.95	11.31
corr	25.95	12.67



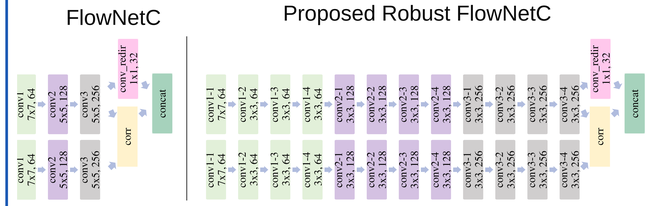
Since correlations are the actual cause, we can attack without optimization



Rectifying the vulnerability

Self-similar patterns causing ambiguities are related to the classical *aperture problem*

Fix: Increase the aperture by increasing the receptive field before the correlation layer



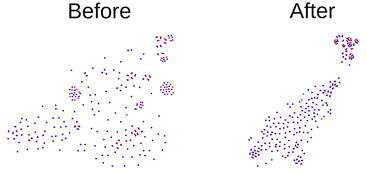
Aperture problem

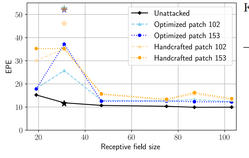
Motion of homogeneous contour is locally ambiguous

Caused by a finite receptive field (aperture)

Results

Increasing the receptive field alleviates the vulnerability and keeps the features well-aligned after the correlation layer





Kernel size	Convs per resolution level	Receptive field
3	1	19
5	1	31
3	2	47
3	3	75
5	2	87
3	4	103

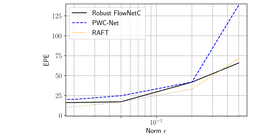
Making a robust network vulnerable

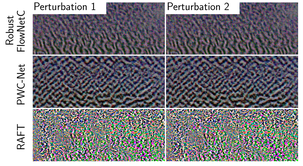
RAFT becomes vulnerable when its receptive field before the correlation layer is decreased

FlowNetC Encoder	Without Context Encoder	Unattacked EPE	102x102 (2.1%) Worst	153x153 (4.8%) Worst
-	-	5.86	8.69	8.96
-	✓	6.88	10.09	11.31
✓	-	5.84	10.50	11.60
✓	✓	6.33	19.12	20.99

Perturbation attacks

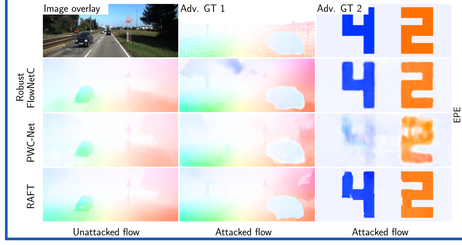
Flow networks are surprisingly robust against universal perturbation attacks

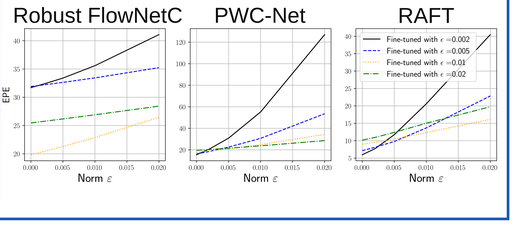




Flow networks are vulnerable to (targeted) perturbation attacks

This can be rectified through adversarial data augmentation





Conclusion

Self-similar patterns in conjunction with the correlation layer explain the vulnerability

This can be fixed by increasing the aperture

An attacker can create virtually any desired flow with (targeted) perturbation attacks

However, the attacker requires to access both the image stream and flow network

Acknowledgements

Funded by the Deutsche Forschungsgemeinschaft (DFG) – BR 3815/10-1, INST 39/1108-1, and the German Federal Ministry for Economic Affairs and Climate Action" within the project KI Delta Learning – 19A19013N.