

## A IMPLEMENTATION DETAILS FOR NPSS

Algorithm 2 shows the NPSS optimization step (*OptimizeNPSS*) in our Spatial-Channel Optimization (SCO) as detailed in the main text.

---

**Algorithm 2:** *OptimizeNPSS* - Optimize over  $r$  rows of  $P_k \in [0, 1]^{r \times c}$  to find the subset of rows  $S^*$  with the highest NPSS score  $F^*$ .

---

**Input :**  $P_k \in [0, 1]^{r \times c}$

**Output:**  $F^*, S^*$

```

1  $F^* \leftarrow -1$ ;
2  $S^* \leftarrow \emptyset$ ;
3 for  $\alpha$  in  $T = \text{LinearSpace}(0, 1)$  do
4    $\text{sorted\_priority} \leftarrow \text{SortByPriority}(r, \alpha)$ ; /* Sort  $r$  rows by  $\frac{N_\alpha}{c}$  or the
      proportion of p-values  $< \alpha$  across  $c$  columns. */
5    $\text{score}, \text{subset} = \phi(\text{sorted\_priority})$ ; /* Score  $r$  subsets of  $\text{sorted\_priority}$ 
      by iteratively including elements one at a time in the
      sorted order using NPSS. */
6   if  $\text{score} > \text{max\_score}$  then
7      $F^* \leftarrow \text{score}$ ;
8      $S^* \leftarrow \text{subset}$ ;
9 return  $F^*, S^*$ 

```

---

## B PATCH-BASED ATTACKS GENERATION

In Figure 7 we can observe the effect of the adversarial attack patch sizes on FlowNetC (Dosovitskiy et al. 2015) trained on raw KITTI (Geiger et al. 2013) dataset. In Figure 8 we apply Principal Component Analysis (PCA) on each RGB channel of the clean and attacked samples in KITTI 2015 and visualize the distribution. Even though the clean and attacked samples are from a disjoint set, it is hard to distinguish the two distributions. In Table 5 we can see the effect of various patch attack sizes on EPE of the flow estimations from four flow networks we consider, i.e., FlowNetC, FlowNet2, PWCNet, and RAFT, on KITTI 2015 and MPI-Sintel dataset.



Figure 7: Visualization of the effect of adversarial attack patches on FlowNetC (Dosovitskiy et al. 2015) trained on raw KITTI dataset. In each panel, from left to right: attacked input image 1, attacked input image 2, original estimation of flow, and attacked estimation of flow.

## C EXTENDED RESULTS ON PATCH-BASED ATTACK DETECTION

Figure 9 shows the distribution of  $F_{max}$  of the clean and attacked set of KITTI 2015 (top) and MPI-Sintel (bottom) dataset across two patch attack sizes ( $p = 153, 51$ ) and flow networks (FlowNetC, FlowNet2, PWCNet, and RAFT) corresponding to the results listed in Table 2. Table 6 shows



Figure 8: Example image from the clean and attacked set from KITTI 2015 dataset, and distribution of the two sets after applying Principal Component Analysis (PCA) to reduce their dimension to 2. Even though the clean and attacked images come from disjoint sets, it is hard to distinguish them through PCA.

Network	Dataset	Non-attacked EPE	Attacked EPE			
			$p = 153$	$p = 102$	$p = 51$	$p = 25$
FlowNetC	KITTI 2015	11.50	38.85	38.43	31.60	13.29
	MPI-Sintel	3.18	42.45	15.16	29.64	3.48
FlowNet2	KITTI 2015	10.07	12.24	12.48	12.09	11.45
	MPI-Sintel	2.22	3.10	2.93	2.71	2.56
PWCNet	KITTI 2015	12.55	17.01	17.04	16.27	15.21
	MPI-Sintel	3.98	5.37	5.10	4.77	4.58
RAFT	KITTI 2015	5.86	8.85	8.52	7.37	6.93
	MPI-Sintel	1.63	2.85	2.57	2.23	2.00

Table 5: Effect of adversarial patch attacks on four top optical flow estimators, FlowNetC, FlowNet2, PWC-Net, and RAFT on KITTI2015 and MPI-Sintel dataset. We use  $p \times p$  adversarial patches ( $p = 153, 102, 51, 25$ ) to attack our test images. These patches are trained on KITTI raw and MPI-Sintel raw dataset, and are evaluated on KITTI 2015 and MPI-Sintel datasets, respectively, using end-point error (EPE).

the detection performances (AUC) of all four flow networks across various sizes of patch attacks, i.e.,  $p = 153, 102, 51, 25$ , on KITTI 2015 and MPI-Sintel. Overall, we see increasing detection performances as we increase the patch attack sizes.

## D EXTENDED RESULTS ON PATCH LOCALIZATION

Table 7 shows the localization performances (AP/AR) of our proposed work across various patch attack sizes ( $p = 153, 102, 51, 25$ ) for KITTI 2015 and MPI-Sintel on four flow networks we consider, FlowNetC, FlowNet2, PWCNet, and RAFT. See Figure 10 for some examples of our localization predictions across these patch sizes on all four networks. Generally, we see higher localization performances for larger patches. An exception is the results for  $153 \times 153$  patch attacks for FlowNetC on MPI-Sintel, which show higher performances when  $k$  is optimized (see Figure 6)

## E EXTENDED RESULTS ON ABLATION STUDY

### E.1 PERFORMANCE ACROSS LAYERS

Table 8 shows the detection performances (AUC) across various intermediate layers of each flow network with Layer 1 being the earliest layer and Layer 5 being the deepest layer we consider. See Figure 11 for example visualizations of localized attacks on various layers of FlowNetC for across various patch attack sizes. Similar to the observation in the main paper, we see higher detection power in the deeper layers for FlowNetC and in the earlier layers for the other three networks.

### E.2 PERFORMANCE WITHOUT THE PROPOSED COMPONENTS

Figure 12 shows the overview of our proposed work without the proposed components, i.e., (a) without PC and (b) without PC and SCO. Figure 13 shows the distribution of  $F_{max}$  for the clean and attacked samples, Table 9 lists the corresponding detection performances (AUC), and Figure 14

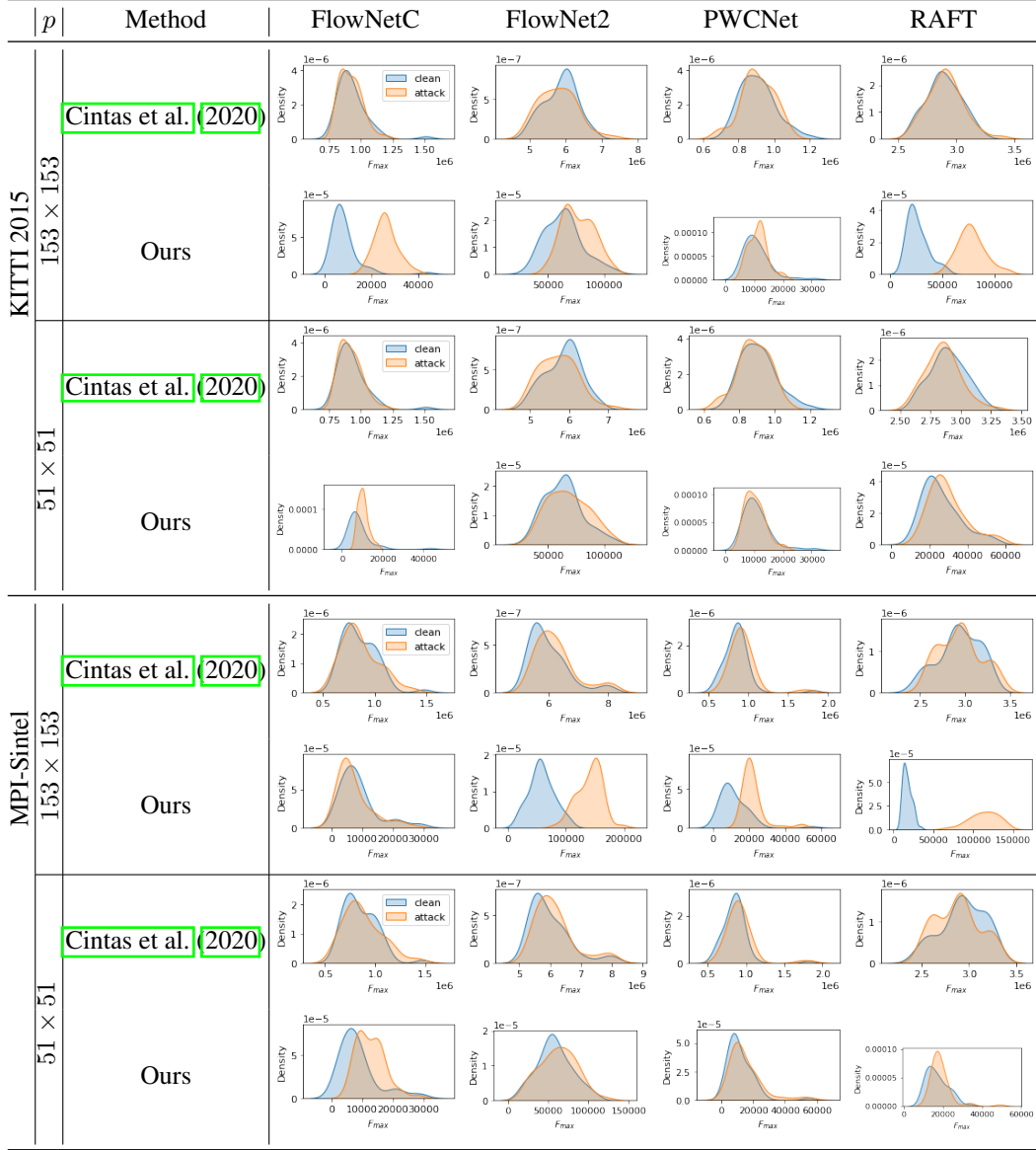


Figure 9: Distribution of anomalous scores obtained from the clean and attacked test set on FlowNetC, FlowNet2, PWCNet, and RAFT corresponding to the results listed in Table 2. Attacked test set contains samples with  $p \times p$  patch attacks ( $p = 153, 51$ ).

Table 6: Performance (AUC) of patch-based attack detection across various sizes of patch attacks ( $p \times p$ ) for FlowNetC, FlowNet2, PWCNet, and RAFT on KITTI 2015 and MPI-Sintel dataset.

	Method	$p$	FlowNetC	FlowNet2	PWCNet	RAFT
KITTI 2015	Cintas et al. (2020)	153	0.50	0.55	0.53	0.52
		102	0.51	0.57	0.51	0.54
		51	0.51	0.59	0.54	0.58
		25	0.52	0.59	0.57	0.59
	ours	153	0.98	0.72	0.58	1.00
		102	0.88	0.70	0.68	0.91
		51	0.78	0.57	0.52	0.61
		25	0.52	0.56	0.53	0.61
MPI-Sintel	Cintas et al. (2020)	153	0.50	0.64	0.66	0.54
		102	0.53	0.62	0.65	0.51
		51	0.54	0.59	0.63	0.53
		25	0.53	0.59	0.62	0.57
	ours	153	0.58	1.00	0.88	1.00
		102	0.59	0.84	0.83	0.98
		51	0.77	0.56	0.62	0.61
		25	0.56	0.56	0.54	0.53

Table 7: Localization performances (AP/AR) of our proposed method across various patch attack sizes ( $p \times p$ ) on four flow networks we consider, i.e., FlowNetC, FlowNet2, PWCNet, and RAFT.

	$p$	FlowNetC		FlowNet2		PWCNet		RAFT	
		AP	AR	AP	AR	AP	AR	AP	AR
KITTI 2015	153	0.95	0.35	0.38	0.27	0.33	0.19	0.95	0.66
	102	0.91	0.41	0.36	0.40	0.35	0.32	0.74	0.65
	51	0.63	0.73	0.01	0.05	0.02	0.07	0.01	0.04
	25	0.00	0.02	0.00	0.04	0.00	0.02	0.00	0.02
MPI Sintel	153	0.02	0.00	0.92	0.66	0.66	0.29	0.95	0.70
	102	0.59	0.24	0.59	0.62	0.52	0.48	0.82	0.75
	51	0.50	0.73	0.04	0.09	0.19	0.38	0.18	0.34
	25	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00

show example visualization of the predicted localization. Using both our proposed components yields the best performances.

### E.3 DETECTED ANOMALOUS CHANNELS

Figure 15 shows the histograms of indices of detected anomalous channels or filters across the four networks we consider on KITTI 2015 (top) and MPI-Sintel (bottom). Generally, almost all  $p$ -values  $P$  across the channel dimension are detected to show anomalous behaviors except for FlowNetC and RAFT. Our future work will involve understanding this behavior in the channel dimension.















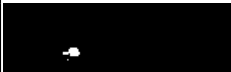
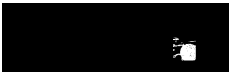
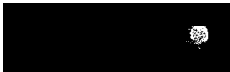
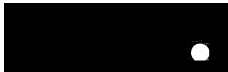
















Attack Size	FlowNetC	FlowNet2	PWCNet	RAFT
$153 \times 153$				
				
$102 \times 102$				
				
$51 \times 51$				
				
$25 \times 25$				
				

Figure 10: Example of the detected subset of anomalous locations (white) across four different sizes of patch attacks for FlowNetC, FlowNet2, PWCNet, and RAFT using the same examples as the ones in Figure 5 for KITTI 2015 dataset. Each panel shows examples with different sizes of patch attacks ranging from  $153 \times 153$  (top) to  $25 \times 25$ . In each panel, top row shows the true mask of where the patch attack occurs, and the bottom row shows the predicted location of the patch attacks.





















$p$	Attacked $I_1$	True Attack Mask	Encoder Layer	Cost Volume	Decoder Layer
$153 \times 153$					
$102 \times 102$					
$51 \times 51$					
$25 \times 25$					

Figure 11: Example of the detected subset of anomalous locations (white) across three layers for FlowNetC, each from its encoder, correlation, and decoder module using  $p \times p$  patch attacks ( $p = 153, 102, 51, 25$ ).

Table 8: Performance (AUC) of patch-based attack detection on various intermediate layers of FlowNetC, FlowNet2, PWCNet, and RAFT on KITTI 2015 and MPI-Sintel dataset.

	Network	Method	Layer 1	Layer 2	Layer 3	Layer 4	Layer 5
KITTI 2015	FlowNetC	Cintas et al. (2020) ours	0.56 <b>0.57</b>	0.60 <b>0.67</b>	0.50 <b>0.98</b>	0.72 <b>1.00</b>	0.77 <b>0.99</b>
	FlowNet2	Cintas et al. (2020) ours	0.55 <b>0.72</b>	0.50 <b>0.54</b>	0.51 <b>0.53</b>	0.51 <b>0.55</b>	0.52 0.52
	PWCNet	Cintas et al. (2020) ours	0.51 <b>0.58</b>	0.58 <b>0.60</b>	0.53 <b>0.58</b>	0.53 <b>0.54</b>	<b>0.60</b> 0.58
	RAFT	Cintas et al. (2020) ours	0.85 <b>0.98</b>	0.52 <b>1.00</b>	<b>0.59</b> 0.58	<b>0.54</b> 0.51	0.52 <b>0.55</b>
MPI-Sintel	FlowNetC	Cintas et al. (2020) ours	0.63 <b>0.88</b>	<b>0.62</b> 0.54	0.50 <b>0.58</b>	0.60 <b>1.00</b>	0.82 <b>0.88</b>
	FlowNet2	Cintas et al. (2020) ours	0.64 <b>1.00</b>	<b>0.57</b> 0.52	<b>0.55</b> 0.50	0.53 <b>0.85</b>	0.52 <b>0.59</b>
	PWCNet	Cintas et al. (2020) ours	<b>0.65</b> 0.53	<b>0.61</b> 0.58	0.66 <b>0.88</b>	0.63 <b>0.71</b>	<b>0.64</b> <b>0.55</b>
	RAFT	Cintas et al. (2020) ours	0.93 <b>1.00</b>	0.54 <b>1.00</b>	<b>0.65</b> 0.51	<b>0.60</b> 0.58	0.52 <b>0.55</b>

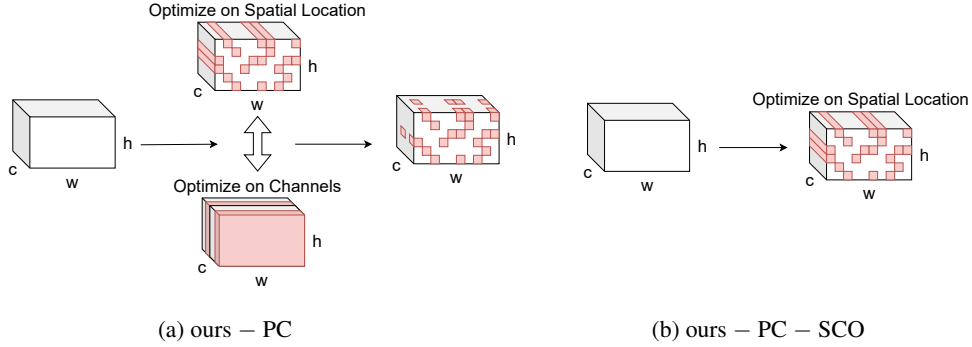


Figure 12: Simplified versions of our proposed method in Figure 2 where for a), we get rid of the proximity constraint (PC) detailed in Section 13 and apply our spatial-channel optimization step on the entire map instead of  $k \times k$  region, for b) we remove the spatial-channel optimization (SCO) step detailed in 13 from a) and optimize once across the spatial location. We show their corresponding detection performances in Table 9,  $F_{max}$  distributions in Figure 13 and localization visualizations in Figure 14.

Table 9: Performance (AUC) changes of patch-based attack detection without some of the proposed components, i.e., proximity constraint (PC) and spatial-channel optimization (SCO). In each column, we bold the method with the best detection performance. Our proposed method with all the components performs the best overall.

	SCO	PC	FlowNetC		FlowNet2		PWCNet		RAFT	
			$p = 153$	$p = 51$	$p = 153$	$p = 51$	$p = 153$	$p = 51$	$p = 153$	$p = 51$
KITTI 2015	✓	✓	<b>0.98</b>	<b>0.78</b>	<b>0.72</b>	0.57	<b>0.58</b>	<b>0.52</b>	<b>1.00</b>	<b>0.61</b>
	✓		0.79	0.52	0.52	<b>0.58</b>	<b>0.58</b>	<b>0.52</b>	0.79	0.51
			0.89	0.60	0.55	0.56	<b>0.58</b>	<b>0.52</b>	0.90	0.51
MPI-Sintel	✓	✓	0.58	<b>0.77</b>	<b>1.00</b>	0.56	<b>0.88</b>	<b>0.62</b>	<b>1.00</b>	0.61
	✓		<b>0.63</b>	0.54	0.65	0.59	0.65	0.59	0.84	0.55
			0.52	0.67	0.59	<b>0.65</b>	0.66	0.59	0.96	<b>0.65</b>

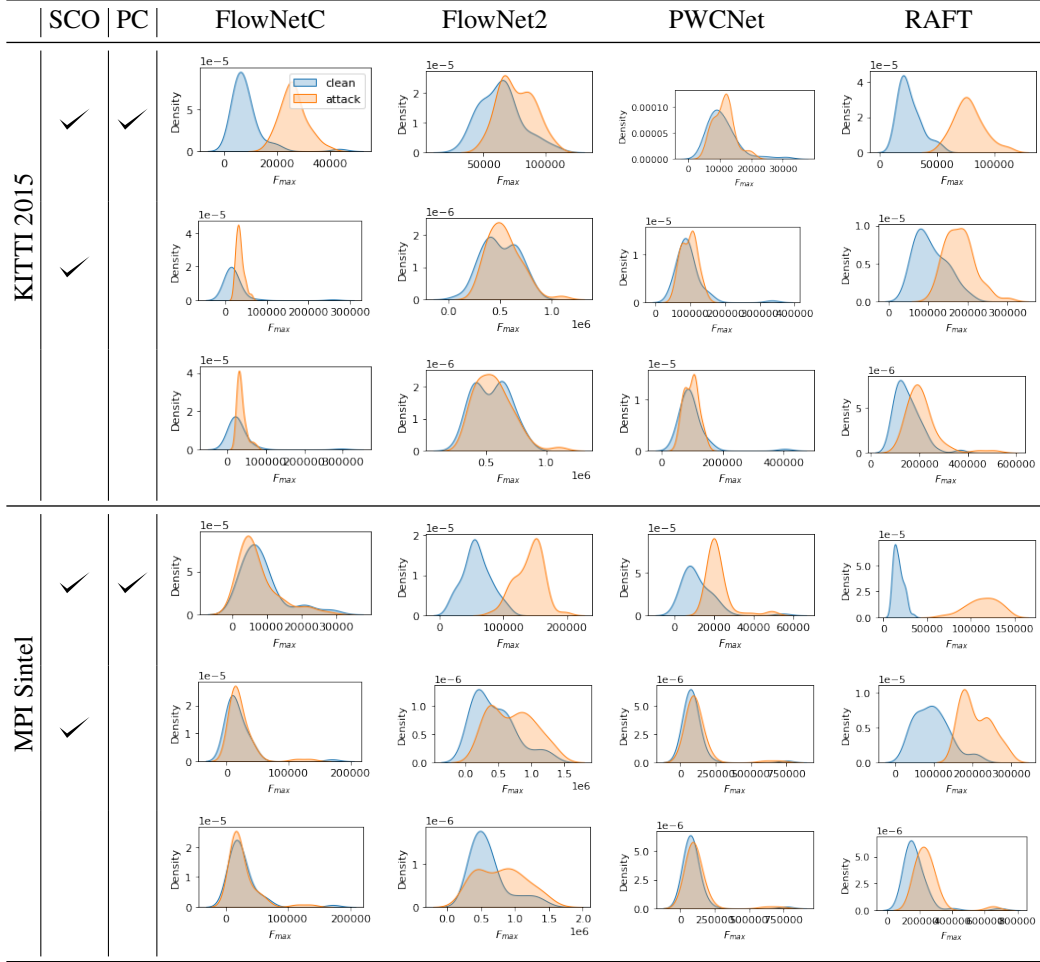


Figure 13: Distribution of  $F_{max}$  obtained from the clean and attacked test set on FlowNetC with different components of our proposed network corresponding to the results in Table 4

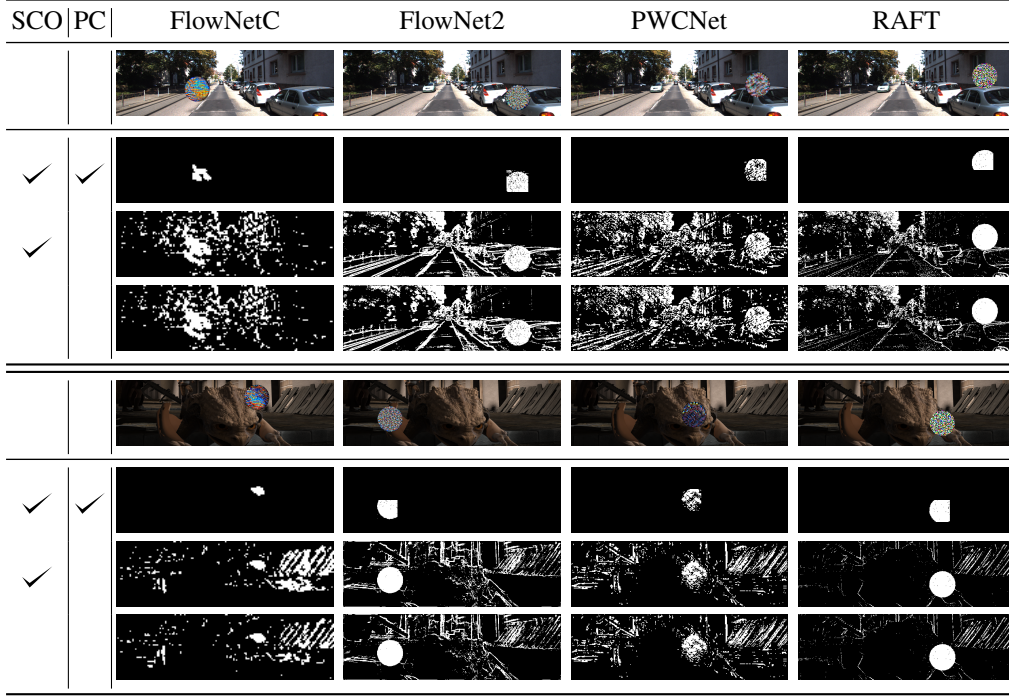


Figure 14: Example of the detected subset of anomalous locations (white) with different components of our proposed method, i.e., proximity constraint (PC) detailed in Section 13 and spatial-channel optimization (SCO) detailed in Section 13.

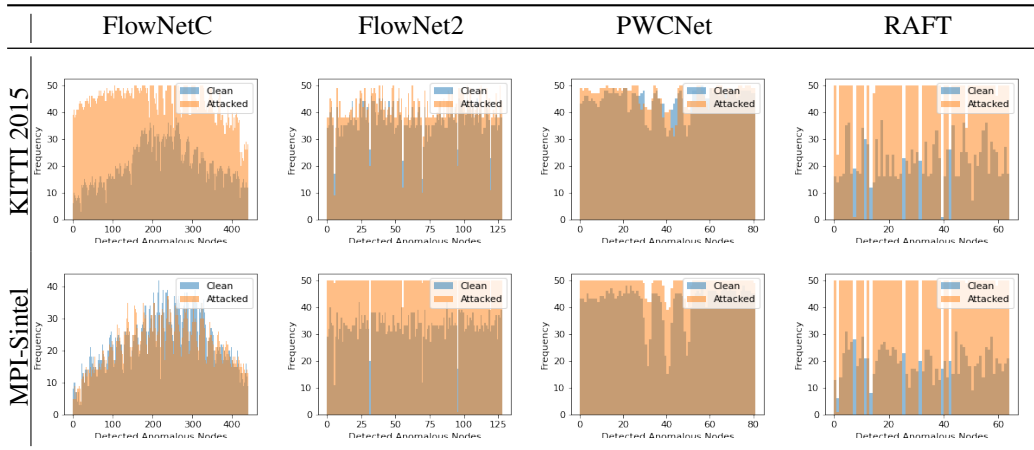


Figure 15: Histogram of the indices of detected anomalous channels or filters across four flow networks we consider, i.e., FlowNetC, FlowNet2, PWCNet, and RAFT, on KITTI 2015 (top) and MPI-Sintel (bottom) dataset.