Detection Defenses: An Empty Promise against Adversarial Patch Attacks on Optical Flow

Erik Scheurer*[1] Jenny Schmalfuss*[2] Alexander Lis Andrés Bruhn[2]
Institute for Visualization and Interactive Systems, University of Stuttgart
first.last@{[1]simtech,[2]vis}.uni-stuttgart.de

Abstract

Adversarial patches undermine the reliability of optical flow predictions when placed in arbitrary scene locations. Therefore, they pose a realistic threat to real-world motion detection and its downstream applications. Potential remedies are defense strategies that detect and remove adversarial patches, but their influence on the underlying motion prediction has not been investigated. In this paper, we thoroughly examine the currently available detect-and-remove defenses ILP and LGS for a wide selection of state-of-the-art optical flow methods, and illuminate their side effects on the quality and robustness of the final flow predictions. In particular, we implement defense-aware attacks to investigate whether current defenses are able to withstand attacks that take the defense mechanism into account. Our experiments yield two surprising results: Detect-and-remove defenses do not only lower the optical flow quality on benign scenes, in doing so, they also harm the robustness under patch attacks for all tested optical flow methods except FlowNetC. As currently employed detect-and-remove defenses fail to deliver the promised adversarial robustness for optical flow, they evoke a false sense of security. The code is available at https://github.com/cv-stuttgart/DetectionDefenses.

1. Introduction

Adversarial attacks have an enormous potential to mislead optical flow methods into predicting the wrong apparent 2D motion from image sequences. Among them, adversarial patches [35,46,49] are the most safety-critical as they distort the optical flow predictions largely independent of their location and orientation, *cf*. Fig. 1, and are printable for effective physical-world attacks [35,48]. On top of that, embedding the distorted optical flow into high-level recognition methods, *e.g*. for flow-based action recognition [9, 16], often corrupts the downstream application [17, 50].

*Equal contribution.

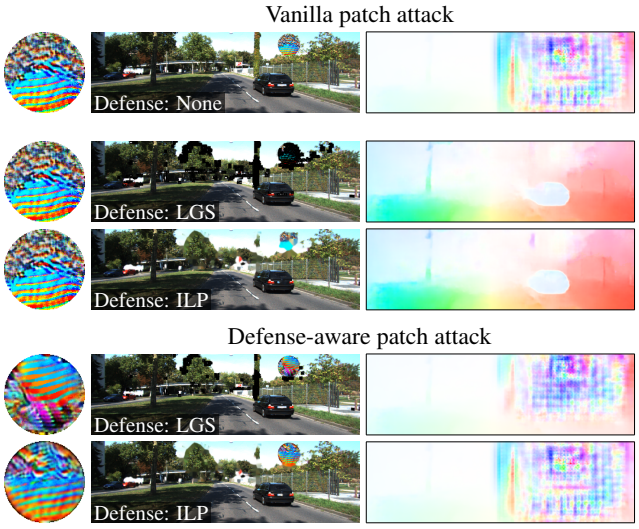

Figure 1. Standard patch attack [35] (vanilla) and defense-aware attacks on FlowNetC's [11] optical flow prediction. Left: Adversarial patch. Middle: Attacked image with applied defense (LGS or ILP, if any). Right: Optical flow. While both LGS and ILP defenses can defend against the vanilla patch attack [2, 35] (top) neither defense withstands defense-aware patch attacks (bottom).

To protect methods against the negative effects of adversarial patches, a straightforward defense concept is to first detect the patch and then render it harmless, *e.g*. by masking the former patch area [14, 30]. However, for classification, it was soon discovered that early defenses do not withstand attacks that take the defense mechanism into account [3, 10]. Such broken defenses are useless for practical applications because attackers aware of the method design (including potential defense mechanisms) can easily overcome them [3, 7, 45]. Among the broken defenses [10] for patch attacks on classification is Local Gradient Smoothing (LGS) [30], which also has been considered to defend optical-flow based action recognition pipelines [2]. Because LGS simply blackens the detected adversarial patch, Inpainting with Laplacian Prior (ILP) [2] was proposed. ILP

inpaints the patch region using neighborhood information to improve the classification accuracy for patch-attacked action recognition pipelines. However, the evaluation of LGS and ILP on the optical-flow component for action recognition in [2] has two major problems: First, it is unclear whether LGS or ILP withstand defense-aware attacks in the context of optical flow – given the results for LGS in classification [10], this is unlikely. And second, an evaluation of how these defenses affect the quality and robustness of optical flow methods is missing, which significantly impacts their practical applicability for all optical-flow-based problems. This work addresses both aspects by providing the first comprehensive analysis of detection defenses against patch attacks proposed in the context of optical flow.

Contributions. We make four contributions. (i) We develop *defense-aware patch attacks* on the ILP and LGS defense for optical flow estimation, by making the defenses differentiable (replacing gradient-free operations) and by avoiding patch detection by the defense (with tailored loss terms). (ii) Moreover, we *investigate the effectiveness* of ILP and LGS on a large set of optical flow methods. Surprisingly, the defenses not only lower the quality of benign (unattacked) predictions but also decrease the robustness for standard (vanilla) and defense-aware attacks – leaving no advantage of defended methods over undefended ones. (iii) Then, we find these *significant defense shortcomings to be caused* by the delocalized destruction of image information for benign scenes, which currently prevents viable detection defenses for optical flow estimation. (iv) Finally, we formulate evaluation advice for defenses on pixel-wise prediction tasks like optical flow, to help avoid common evaluation mistakes in future defense proposals.

2. Related work

Adversarial (patch) attacks on optical flow. Optical flow methods take a pair of input frames $I_1, I_2 \in \mathbb{R}^{M\times N\times 3}$ to predict the 2D vectors that describe the apparent motion, or optical flow $f \in \mathbb{R}^{M\times N\times 2}$ from I_1 to I_2. Adversarial attacks then modify the input frames to *corrupt the optical flow* prediction. The first adversarial attacks on optical flow go back to Ranjan *et al*. [35] who considered patches [5]. Since then, patch attacks were extended to include transparencies [46], simultaneously harm depth estimators [49] or were used to attack flow-based action recognition [17]. Meanwhile, image-wide attacks on optical flow range from global [1, 39, 40] over semantically constrained attacks [20] to adversarial weather [37, 38]. Here, we investigate adversarial patches for being a threat in the physical world.

Adversarial defenses and their evaluation. Adversarial defense mechanisms are designed to protect methods against the perturbing effects of adversarial attacks. Typical defense strategies are adversarial training as a form of data augmentation [13, 21, 36, 42], upstream strategies that filter perturbations from the inputs [14, 23, 30] and certified defenses that come with robustness guarantees [4, 10, 22, 33, 47]. However, many early defenses based on filtering operations were found to be ineffective if the attacker takes the defense mechanism into account [3, 7, 45]. Therefore, a defense's effectiveness has to be shown under defense-aware adversarial attacks to justify its merit. In the process, one has to adequately (*i.e*. effectively) include the defense into the defense-aware attack: Prior work [3, 45] demonstrated in the context of *classification* that by neglecting this fact, many defenses appear unjustifiedly strong despite being evaluated with defense-aware attacks. Hence in this work, following the evaluation guidelines from [3, 45], we design the first defense-aware attacks on optical flow.

Defenses against adversarial patch attacks. Very few defenses are specialized to optical flow [2, 50]. Hence, we first discuss general adversarial patch defenses related to *classification*. Certifiable defenses against patch attacks on classification are provably robust [10, 22, 47], but often lead to smaller robustness improvements. Adversarial training [36] and architectural modification [29] have been also shown to improve the robustness against patch attacks. A last class of patch defenses aims to detect the patch in order to remove it [14, 23, 26, 30]. Among them, digital watermarking uses saliency maps [14], while Local Gradient Smoothing (LGS) detects anomalies in the input gradients [30]. Both use non-differentiable operations that hinder the backpropagation to train adversarial patches, but if these operations are replaced by differentiable ones [3], both defenses are ineffective against defense-aware attacks [10].

In the context of *optical flow*, LGS has been applied to defend optical-flow-based action recognition [2]. An optical-flow-specific improvement is Inpainting with Laplacian Prior (ILP) [2], which yields visually pleasing defended images. Also for action recognition, [50] proposed an optical-flow defense based on self-supervised counter-perturbations against noise-like perturbations. Since we focus on defenses against *patch attacks* for optical flow, this leaves LGS and ILP as potential methods. However, for such defenses, no analysis with defense-aware attacks has been performed, and neither have optical flow methods been considered independent of action recognition.

3. Defending optical flow with LGS and ILP

We begin by providing technical details for the LGS [30] and ILP [2] defenses. Both defenses detect the adversarial patch based on large image gradients and then replace these regions to remove the adversarial patch.

Patch detection. To detect the patch, the image is split into overlapping blocks $B = K \times K$ of size K and overlap O. Then, a subset of blocks containing potential adversar-

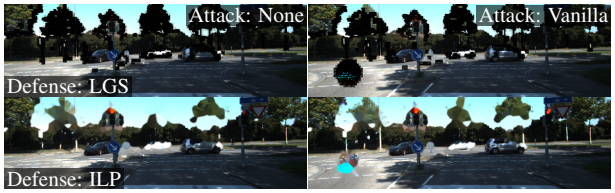

Figure 2. ILP (top) and LGS (bottom) defenses on unattacked (left) and attacked (right) images of the KITTI 2015 dataset [28]. Defenses degrade the visual quality, but LGS more than ILP.

ial modifications is selected. As adversarial patches often have large color changes (*cf*. Fig. 1), the gradient magnitude is accumulated for each block to identify blocks with the largest gradients. The gradient magnitude computation differs for ILP and LGS: While LGS considers first derivatives of the input image $I \in \mathbb{R}^{M\times N\times 3}$, ILP uses second derivatives, resulting in the gradient fields $G \in \mathbb{R}^{M\times N}$:

$$G_{\text{LGS}} = ||\nabla I||, \tag{1}$$

$$G_{\text{ILP}} = ||\Delta I||. \tag{2}$$

Normalizing gradients per image yields scale invariance:

$$\bar{G}_{i,j} = \frac{G_{i,j} - \min_{i,j\in M\times N} G_{i,j}}{\max_{i,j\in M\times N} G_{i,j} - \min_{i,j\in M\times N} G_{i,j}}. \tag{3}$$

Based on $\bar{G}$, adversarially modified pixels are marked: Per pixel (i,j), we denote all enclosing blocks by $B_{(i,j)}$, and let these blocks vote whether the sum of their gradients exceeds a threshold $t \in [0,1]$ (t is relative to the distribution of $\bar{G}$). If at least one block has large gradients, the respective pixel is marked as adversarial in a binary mask $M \in \mathbb{R}^{M\times N}$:

$$M_{i,j} = \begin{cases} 1 & \text{if } \exists\, B \in B_{(i,j)} : \sum_{k,l\in B} \bar{G}_{k,l} > t, \\ 0 & \text{else.} \end{cases} \tag{4}$$

Through this procedure, a block with large gradients causes all contained pixels to be marked as adversarial. To remove incorrectly marked (non-adversarial) pixels, ILP performs a reevaluation of candidates in M. After scaling with s_{ILP}, their gradients must exceed a threshold t_{ILP} to yield ILP's final mask M_{ILP}, with $\odot$ as pixel-wise multiplication:

$$M_{\text{ILP}} = M \odot \text{tr}(s_{\text{ILP}} \cdot \bar{G}_{\text{ILP}} > t_{\text{ILP}}). \tag{5}$$

Patch removal. Next, the defenses replace these potentially adversarial pixels from M for LGS and from M_{ILP} for ILP. LGS reduces the gradients in the detected area, which results in the modified image

$$I_{\text{LGS}} = (1 - \text{clip}_{[0,1]}(b_{\text{LGS}} \cdot \bar{G} \odot M)) \odot I, \tag{6}$$

where b_{LGS} is a smoothing parameter. If b_{LGS} is large, this darkens the selected adversarial pixels. ILP instead inpaints the selected pixels with Telea's algorithm [44] with radius r_{Telea} for more pleasing visual results. This smoothes colors from the edges of the selected areas into their center:

$$I_{\text{ILP}} = \text{Telea}(M_{\text{ILP}}, I, r_{\text{Talea}}), \tag{7}$$

Fig. 1 and Fig. 2 show LGS- and ILP-defended images. Our final hyperparameters for ILP and LGS are $K = 16$, $O = 8$, $t = 0.15$, $t_{\text{ILP}} = 0.5$, $s_{\text{ILP}} = 15$, $b_{\text{LGS}} = 15$ and $r_{\text{Telea}} = 5$; The supplement provides details on their selection.

4. Defense-aware patch attacks for optical flow

LGS and ILP defenses were only used to attack optical flow predictions for action recognition in a black-box way so far [2], meaning adversarial patches were trained without knowledge about the defense. According to best-practice for defense evaluation [3, 7, 45], the *defended* model must be evaluated under *defense-aware* attacks to show that it indeed offers protection. In the following, we develop white box patch attacks on the ILP and LGS defenses for optical flow. Our defense-aware attacks expand on Chiang *et al*. [10] who successfully attacked LGS for classification, but neither considered ILP nor the optical flow problem.

Gradient computations through the defense mechanism. The defensive properties of LGS and ILP are based on *shattered gradients*, *i.e*. the use of mathematical operations with nonexistent gradients that prevent adversarial optimization [3, 31]. To still optimize adversarial patches in a defense-aware manner, the Backward Pass Differential Approximation (BPDA) [3] replaces these operations with differentiable approximations during backpropagation. The forward pass is executed normally. Within LGS and ILP, the problematic operations are the block-wise filtering steps (LGS and ILP), thresholding (ILP) and clipping operations (LGS), and the inpainting step (ILP). Below, we describe how they are approximated to enable backpropagation.

In the block-wise filtering step for LGS and ILP, *cf*. Eq. (4), the gradients do not exist for the conditional selection. To bypass them with BPDA [3], the filtering is replaced with the differentiable identity function, resulting in $\nabla M = 1$. The thresholding in ILP's filtering has a similar problem, *cf*. Eq. (5), hence we also replace it with an identity function in the backward pass. For the clipping in LGS's smoothing, *cf*. Eq. (6), the true gradient is one when the argument is in $[0,1]$ and otherwise undefined. In practice, we find this operation responsible for most gradient shattering: Whenever the smoothing darkens values below zero, the clipping then sets them to zero, losing the gradient. Therefore, in the backpropagation, we approximate gradients with the identity if the value to clip is in $[0,1]$ and with zero otherwise. As the ILP inpainting is very time-consuming, *cf*. Eq. (7), we treat it as being gradient-free. Similar to the clipping approximation, we bypass it

with an identity operation for non-inpainted pixels and a zero-gradient for inpainted ones. To overcome optimization problems for zero-gradients in the clipping- and inpainting approximations, we introduce additional loss terms to improve the patch in areas with no gradient information.

Defense-aware loss functions. Optimizing defense-aware adversarial patches requires a loss function that encourages patches with a perturbing effect on the optical flow output. As baseline loss that defines the overall goal for the patch attack, we use the Average Cosine Similarity (ACS) which was used to train adversarial patches on optical flow in [35]. It encourages adversarial patches to invert the original optical flow prediction f to yield the adversarial flow $\check{f}$:

$$\mathcal{L}^{\text{ACS}}(f, \check{f}) = \frac{1}{NM} \sum_{i,j \in M \times N} \frac{\langle f_{i,j}, \check{f}_{i,j} \rangle}{\|f_{i,j}\|_2 \|\check{f}_{i,j}\|_2}. \quad (8)$$

Besides the ACS, another loss term is required to overcome the zero-gradients of BPDA. While the differentiable LGS and ILP approximations allow optimizing adversarial patches, these patches may still be detected by the defenses and hence be stopped from perturbing the flow. Therefore, we use loss terms to penalize large gradient magnitudes in the patches. To optimize defense-aware patches P, we therefore penalize first-order gradients $\|\nabla P\|$ for LGS and second-order gradients $\|\Delta P\|$ for ILP-awareness:

$$\mathcal{L}^{\text{LGS}}(f, \check{f}, P) = \mathcal{L}^{\text{ACS}}(f, \check{f}) + \alpha \|\nabla P\|, \quad (9)$$

$$\mathcal{L}^{\text{ILP}}(f, \check{f}, P) = \mathcal{L}^{\text{ACS}}(f, \check{f}) + \alpha \|\Delta P\|. \quad (10)$$

The parameter α balances the loss terms and is set to $\alpha = 1\text{e}{-8}$. With small gradient magnitudes, the patches are likely below the filtering threshold as it is relative to the remaining image gradients. This way, they evade the defenses and affect the optical flow output as in Fig. 1.

In the ACS implementation, we exclude the patch area from the computation. This measures to which extent the patch modifies the optical flow outside its direct area, *i.e.* it assesses the de-localized impact per patch. This is because one may take two points of view on the role of the patch: In the first view, the patch is an image part, with a zero ground-truth flow at the patch area. In the second view, the patch is an attack part, and defenses should mitigate its effect and restore the ground truth flow in its area. To refrain from assuming a "correct" optical flow for the patch, we exclude the patch region from our loss.

Defense-aware patch optimization. We test two different methods for optimization: The Iterative Fast Gradient Sign Method (I-FGSM) [21] and Stochastic Gradient Descent (SGD). To ensure a valid color range of the patch P after optimization, we consider clipping the values to their valid range in $[0, 1]$ after each update [35] and a change of variables (CoV) via tanh to optimize the values in $[-\infty, \infty]$ before transforming them back into the valid range [8, 39].

5. Metrics for defended quality and robustness

Including a defense to protect an existing method against attacks effectively creates a new method that consists of the original method plus defense D. Hence, we have to evaluate the quality and robustness of this new method instead of the original defense-free approach's metrics [7, 45].

Quality. To evaluate the quality of defended optical flow methods, one typically measures the average endpoint error (EPE) between the ground-truth flow f^* and the predicted flow f, where low errors indicate high quality:

$$\text{EPE}(f^*, f) = \frac{1}{MN} \sum_{i \in M \times N} \|f_i^* - f_i\|_2 \,. \quad (11)$$

Robustness. To evaluate robustness, we use the methodology from [39] and measure the distance between the benign and the attacked flow prediction. This quantifies how much an attack changes a method's output and is motivated by the Lipschitz continuity of functions. Due to the previously discussed two views on the adversarial patch, we evaluate the EPE for all pixels *outside* P, quantifying the patch's negative effect outside its immediate area. We denote the benign flow prediction of a method defended with D as f_{D} and its prediction under attack by A as f_{D}^{A}. Then, the robustness is

$$\text{EPE}_P(f_{\text{D}}, f_{\text{D}}^{\text{A}}) = \frac{1}{MN - P} \sum_{i \in M \times N \setminus P} \|(f_{\text{D}})_i - (f_{\text{D}}^{\text{A}})_i\|_2 \quad (12)$$

with low values for robust methods, as attacked predictions outside the patch should coincide with the unattacked ones.

Quality and robustness for pipelines. When a method is defended with a defense D and attacked with an attack that is defense-aware towards D, we call the setup a *full pipeline* with defense D. Its quality is $\text{Q}_{\text{D}} = \text{EPE}(f^*, f_{\text{D}})$ and the resulting pipeline robustness is $R_{\text{D}}^{\text{D}} = \text{EPE}_P(f_{\text{D}}, f_{\text{D}}^{\text{D}})$.

6. Experiments

We now assess how defenses against patch attacks impact the quality and robustness of optical flow methods. We begin by evaluating the quality of defended methods and then separately assess their robustness against patch attacks. Afterward, we jointly analyze both, quality and robustness, to find them being negatively impacted by defenses against patch attacks. Finally, we explore the reasons for their poor performance. All attacks and defenses are implemented with PyTorch [32], and available at https://github.com/cv-stuttgart/DetectionDefenses.

Optical flow methods and attack setups. As optical flow methods, we select FlowNetC (FNC) [11], SpyNet [34] and PWCNet (PWC) [41] as milestone architectures in flow estimation, RAFT [43], GMA [18] and FlowFormer

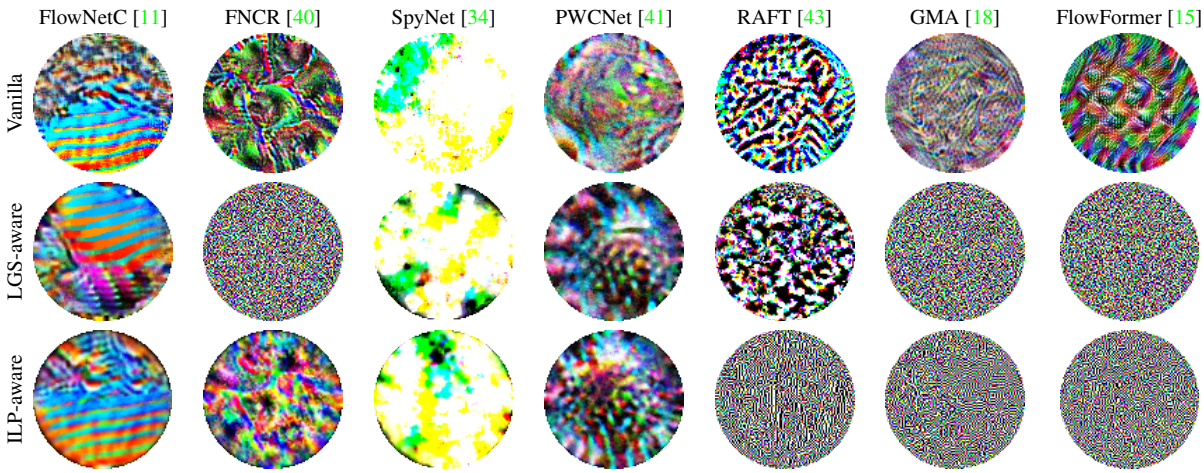

Figure 3. Overview of adversarial patches with size 100 for vanilla, ILP- and LGS-aware patch attacks against all tested networks.

(FF) [15] as current state-of-the-art, and FlowNetCRobust (FNCR) [40] as it improves FlowNetC's patch robustness.

We generate effective adversarial patches by optimizing learning rates, box constraints and optimizer choice for each optical flow network. As optimizers, we consider I-FGSM [21] and SGD, learning rates from 0.1-100 (optimizer-dependent), box constraints via change of variables (CoV) or clipping. Following the protocol for adversarial patches from [35], we optimize patches of size 100 using KITTI Raw [12] and evaluate on KITTI train [28]. Evaluations using Sintel [6], Driving [25], HD1K [19] and Spring [27] are shown in the supplement. For defense-aware patches, we choose $\alpha = 1\text{e}{-8}$ for the loss function in Eq. (9) and Eq. (10). Per flow method, we generate vanilla adversarial patches (without defense awareness, as in [35]) and defense-aware patches for LGS [30] or ILP [2]. We train 4 patches per parameter combination, and average the evaluated metrics. Among the parameters, we select the strongest adversarial configuration for the worst robustness.

The full evaluation of the best parameters per flow method and defense strategy is in the supplement. Fig. 3 visualizes the most effective patches. Both defenses detect high gradient magnitudes, causing smooth defense-aware patches with small derivatives. Interestingly, defense-aware patches for methods like RAFT, GMA or FlowFormer contain high-frequent noise. This calls for detection rather than evasion by ILP or LGS, which we explore in Sec. 6.4.

6.1. Quality of defended optical flow methods

To begin our investigation of defenses D, we assess the quality of defended and undefended optical flow methods on unattacked input frames. Tab. 1 lists the endpoint errors $\text{EPE}(f^*, f_\text{D})$ on KITTI train, where the ground truth flow is available. Across all optical flow methods, we find the lowest errors when no defense is applied; Both ILP and LGS lead to larger errors. But for accurate methods, the errors rise more than for less accurate ones, *i.e.* by 156% for GMA and 4% for SpyNet, using LGS vs. no defense. On average, ILP increases the error less than LGS, *i.e.* by 129% instead of 156% for GMA, compared to no defense. ILP performs better due to its more sophisticated image restoration, which adds pixel-wise filtering with inpainting rather than smoothing, *cf*. Fig. 2. In the figure, applying ILP and LGS to unattacked images visually degrades them, leading to worse predicted flows. Overall, detect-and-remove defenses lower the accuracy on unattacked frames, as they strongly affect the image quality, which harms the flow quality.

Table 1. Quality $\text{Q}_\text{D} = \text{EPE}(f^*, f_\text{D})$ for optical flow pipelines with defense D on the KITTI train dataset [28]; Best quality is **bold**. All defenses lead to a worse quality on unattacked frames.

Defense		FNC	FNCR	SpyNet	PWC	RAFT	GMA	FF
None	Q	**15.42**	**11.10**	**24.03**	**13.26**	**0.63**	**0.61**	**0.62**
LGS	Q_LGS	16.70	13.13	25.15	14.61	1.42	1.55	1.42
ILP	Q_ILP	16.46	12.77	24.74	14.52	1.36	1.39	1.30

6.2. Robustness under defense-aware patch attacks

To study the robustness of defended flow methods under defense-aware attacks, we measure $R^\text{A}_\text{D} = \text{EPE}_P(f_\text{D}, f^\text{A}_\text{D})$ for all combinations of defenses D (None, LGS and ILP) and attacks A (vanilla, LGS- and ILP-aware). Tab. 2 gives the full results. In the analysis process, we (i) evaluate whether defense-aware attacks bypass the defenses, and then (ii) identify the most effective defense for each attack.

Most effective attack per defense. First, we evaluate if our defense-aware attacks evade the defenses. In practice,

Table 2. Robustness scores for all combinations of defended methods and defense-aware attacks on optical flow methods on KITTI train [28]. For a given defense D and attack A, the robustness is defined as $R_{\mathrm{D}}^{\mathrm{A}} = \mathrm{EPE}_P(f_{\mathrm{D}}, f_{\mathrm{D}}^{\mathrm{A}})$. Per attack, the robustness values of the best defense are **bold**. Per defense, the robustness values for the attack it is most vulnerable to are underlined. Full pipelines are highlighted in gray, and provide the corresponding robustness values to the quality scores from Tab. 1.

Attack type	Defense		FNC	FNCR	SpyNet	PWC	RAFT	GMA	FF
Vanilla	None	R^{Van}	73.74	**1.78**	**1.48**	**2.17**	**0.33**	**0.56**	**0.57**
	LGS	$R_{\mathrm{LGS}}^{\mathrm{Van}}$	**3.75**	2.97	3.97	3.34	1.45	1.31	1.30
	ILP	$R_{\mathrm{ILP}}^{\mathrm{Van}}$	4.66	3.11	3.34	3.29	1.43	0.99	1.41
+LGS (LGS-aware)	None	R^{LGS}	50.46	**0.46**	**1.40**	**2.10**	**0.25**	**0.27**	**0.44**
	LGS	$R_{\mathrm{LGS}}^{\mathrm{LGS}}$	23.36	3.27	4.05	4.13	1.46	1.60	1.67
	ILP	$R_{\mathrm{ILP}}^{\mathrm{LGS}}$	**23.04**	4.21	3.32	3.35	1.48	1.54	1.80
+ILP (ILP-aware)	None	R^{ILP}	56.56	**1.02**	**1.45**	**2.16**	**0.20**	**0.26**	**0.45**
	LGS	$R_{\mathrm{LGS}}^{\mathrm{ILP}}$	**10.99**	3.68	3.03	4.06	1.47	1.57	1.68
	ILP	$R_{\mathrm{ILP}}^{\mathrm{ILP}}$	55.26	3.25	3.36	4.25	1.49	1.51	1.81

this corresponds to choosing a defense to observe how the defended model fares against different attacks. For a fixed defense D in Tab. 2 we underline the worst robustness, *i.e.* the most effective attack. Hence we compare $R_{\mathrm{D}}^{\mathrm{Van}}$, $R_{\mathrm{D}}^{\mathrm{LGS}}$ and $R_{\mathrm{D}}^{\mathrm{ILP}}$ (*e.g.* the 2nd line in each block for D=LGS).

Without defenses every method is most vulnerable towards the vanilla attack: $R^{\mathrm{Van}} \geq R^{\mathrm{LGS}}, R^{\mathrm{ILP}}$. This is plausible, as LGS- and ILP-aware attacks impose additional constraints on the patches, which impairs their effectiveness for an undefended model. For defended models, the corresponding defense-aware attacks are often most effective, *e.g.* $R_{\mathrm{LGS}}^{\mathrm{LGS}}$ is largest for the LGS-defended FlowNetC, FlowNetCRobust, SpyNet and GMA. This confirms that our adaptive attacks are truly defense-aware, as they are most effective on the defended models, *i.e.* $R_{\mathrm{D}}^{\mathrm{D}} \geq R_{\mathrm{A}}^{\mathrm{D}}$ for the majority of models. Still, in some cases, an ILP-aware attack performs better on an LGS-defended model and vice versa. This indicates transferable patches for LGS and ILP, as the differences are small in these cases, *e.g.* LGS-defended RAFT scores $R_{\mathrm{LGS}}^{\mathrm{LGS}} = 1.46$ and $R_{\mathrm{ILP}}^{\mathrm{LGS}} = 1.47$.

Overall, our defense-aware attacks are most effective w.r.t. the respective defended models, which validates their design and implementation. Likewise, the defense-aware attacks are less effective on other defenses, as inappropriate constraints hinder the patch's effectiveness. Nonetheless, we find that LGS- and ILP-aware patches are transferable.

Most effective defense per attack. Next, we analyze the most effective defense for a given attack; or in other words, which defense withstands most attacks. Per fixed attack A in Tab. 2, we **boldface** the best robustness per network, comparing R^{A}, $R_{\mathrm{LGS}}^{\mathrm{A}}$ and $R_{\mathrm{ILP}}^{\mathrm{A}}$ with differing defenses.

Focusing first on the vanilla attack (Tab. 2 block 1), the undefended robustness R^{Van} strongly differs. FlowNetC is particularly vulnerable [2, 35, 39, 40] due to a limited field of vision that was expanded in FlowNetCRobust [40] and improved its robustness by 97%, making it comparable to SpyNet or PWCNet. Most robust against vanilla patch attacks are the state-of-the-art methods RAFT, GMA and FlowFormer. Their robustness appears linked to their quality, as they detect the static patches in Fig. 4, correctly estimating the zero motion. This retains correct flow predictions around the patch and results in low robustness scores.

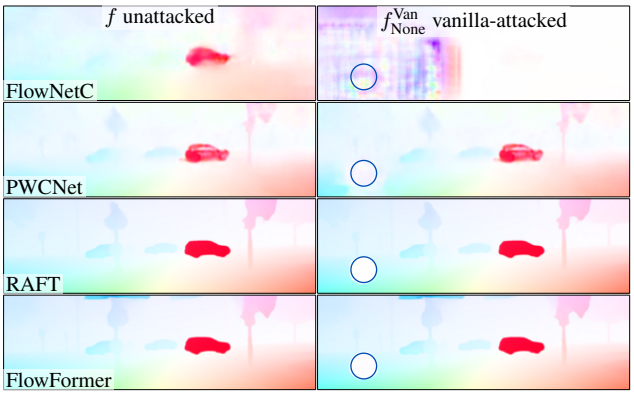

Figure 4. Optical flow estimations for selected methods on a KITTI frame that is unattacked (left) and attacked with the vanilla patch attack (right). Blue circles indicate the patch location. An overview of all optical flow methods is in the supplement.

For methods that are robust against vanilla attacks without defense, *i.e.* all except FlowNetC, defending harms their robustness scores: $R^{\mathrm{A}} < R_{\mathrm{LGS}}^{\mathrm{A}}, R_{\mathrm{ILP}}^{\mathrm{A}}$, independent of the attack A (vanilla, LGS- or ILP-aware). This renders the defenses ineffective, as improving robustness against attacks is their sole purpose. In contrast, defending FlowNetC improves its robustness against vanilla attacks from 73.74 to 3.75 with LGS and 4.66 with ILP. For action recognition, a similar improvement was seen with FlowNetC [2], but compared to other methods, FlowNetC is not robust even when defended and therefore should not be used.

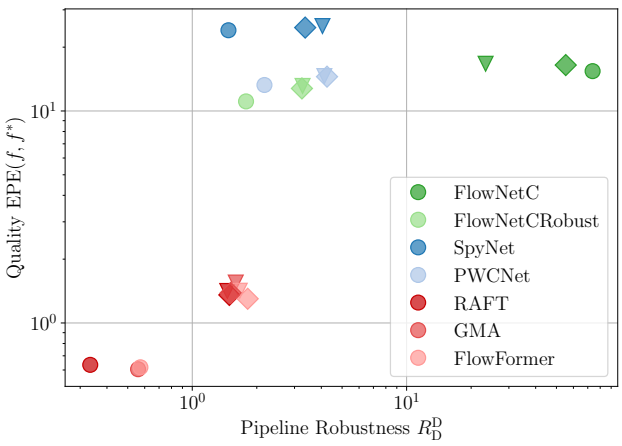

Figure 5. Quality vs. robustness of flow networks on KITTI train in a double logarithmic plot. An ideal method would be in the origin. Undefended networks are circles ◯, networks defended with LGS are triangles ▽ and networks defended with ILP are diamonds ♢. ILP and LGS deteriorate quality and robustness.

All in all, the reported robustness enhancement through ILP and LGS in [2] can not be confirmed for our large test body of optical flow methods. Instead, we find that LGS and ILP defenses harm the robustness of competitive optical flow methods for all tested patch attacks.

6.3. Quality and robustness for defended methods

After separately considering quality and robustness of defended methods under adversarial patch attacks, we now jointly analyze both aspects. Perfect defenses decrease the vulnerability to adversarial attacks without negatively impacting the quality. As models, we consider all tested flow methods with no defense, LGS or ILP. Their quality is taken from Tab. 1. Their corresponding pipeline robustness, *i.e.* defended model's robustness under the respective defense-aware attacks, which is highlighted in gray in Tab. 2.

For all optical flow methods, Fig. 5 visualizes the quality-robustness pairs per defense, *e.g.* Q_{LGS} with R_{LGS}^{LGS}. An ideal method with low scores for quality and robustness would be positioned at the origin. An improvement in robustness moves the defended point to the left, ideally without decreasing quality. For all methods except FlowNetC, the undefended standard model (◯) is closest to the origin and therefore offers the best robustness *and* the best quality, without any trade-off. Using LGS (▽) or ILP (♢) defenses worsen both metrics to a similar extent. The only outlier is FlowNetC, where both defenses improve the robustness while keeping the quality nearly constant, with larger improvements for LGS than for ILP. Overall, our investigation shows that almost all optical flow methods are harmed by the detect-and-remove defenses ILP and LGS, as they worsen method quality and robustness alike.

Figure 6. Effect of the LGS defenses on KITTI [28] frames (left) and the resulting optical flow prediction with RAFT [43] (right). Black areas in the input frame are filtered by the LGS defense. Blue circles mark the area of the adversarial patch, the red boxes highlight an area with prominent differences in the flow predictions. Note that the robustness calculation omits the blue circle.

6.4. Flaws explained: Manual patch attack

From Tab. 2 we saw significant robustness reductions for high-quality methods like RAFT, GMA or FlowFormer when defended with ILP or LGS. Yet, the reductions are caused by high-frequent defense-aware patches, *cf*. Fig. 3, which seems to contradict the optimization for smoothness to evade detection by the ILP and LGS gradient filtering.

To understand this behavior, we compare the flows entering into the robustness calculation – the unattacked flow f_D of the defended method and the flow f_D^A after applying a defense-aware attack to the defended model. Fig. 6 shows RAFT's original prediction (unattacked, no defense, Row 1) together with flows for the LGS-defended version. Comparing defended and undefended flows, *e.g.* the car in the red box, the flow from the *unattacked* LGS-defended RAFT (f_{LGS}, Row 2) is very erroneous compared to the *attacked* LGS-defended flow (f_{LGS}^{LGS}, Row 3). In other words, the gradient filtering of the defense destroys important visual information throughout the image, which yields low-quality optical flow predictions in the absence of adversarial attacks. If the alterations in an unattacked image are scattered throughout its domain, a patch attack can maximize flow changes by aggregating alterations in a single location, *i.e.* the patch itself. Incidentally, this *improves* the optical flow prediction in large areas (Rows 1, 3; Red box).

Therefore, we hypothesize that the bad robustness scores of high-quality methods are driven by large distortions in unattacked frames caused by the defense. To test this, we design a manual patch (see Fig. 7) consisting of a checkerboard pattern to maximize first- and second derivatives. With the manual patch, we then attack defended and undefended optical flow methods. Their robustness in Tab. 3 confirms our hypothesis. For undefended methods, the

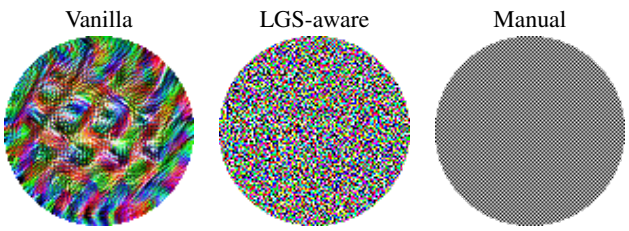

Figure 7. Visual comparison of patches obtained for vanilla, LGS-aware and manual patch attack on FlowFormer [15]. The manual patch imitates high derivatives in the LGS-aware adversarial patch.

Table 3. Robustness $R_{\mathrm{D}}^{\mathrm{Man}} = \mathrm{EPE}_P(f_{\mathrm{D}}, f_{\mathrm{D}}^{\mathrm{Man}})$ against a manual patch attack (Man) of optical flow methods with different defenses D. The best robustness is **bold**.

Defense		FNC	FNCR	SpyNet	PWC	RAFT	GMA	FF
None	R^{Man}	**1.19**	**0.51**	**1.15**	**0.90**	**0.19**	**0.25**	**0.44**
LGS	$R_{\mathrm{LGS}}^{\mathrm{Man}}$	3.69	3.26	3.86	3.40	1.45	1.56	1.61
ILP	$R_{\mathrm{ILP}}^{\mathrm{Man}}$	3.84	3.35	3.18	3.49	1.53	1.57	1.86

patch hardly affects the optical flow prediction. For defended methods, however, the robustness significantly degrades, even though the patch impacts the defense, not the network. Also, we obtain similar results to those of defense-aware attacks in Tab. 2, where robustness degrades most significantly for accurate methods, *e.g.* RAFT, GMA and FlowFormer. However, because FlowNetC's robustness $\mathrm{EPE}_P(f_{\mathrm{D}}, f_{\mathrm{D}}^{\mathrm{A}})$ is driven by the attacked flow $f_{\mathrm{D}}^{\mathrm{A}}$ rather than the unattacked f_{D}, *cf.* Fig. 4, defenses can succeed as every attacked-flow improvement directly serves the robustness.

In summary, the sub-par quality and robustness of defended high-quality methods are a direct consequence of the defense itself, which causes visual distortions in unattacked frames. These distortions not only reduce the quality of benign frames but also are the empirically-confirmed cause for the low robustness scores of high-quality methods.

7. Discussion

We take a moment to condense the findings from analyzing defended optical flow methods into actionable evaluation advice, to encourage meaningful defense evaluations in the future. While evaluation advice has been formulated before and should be adhered [3, 7, 24, 45], we refresh some points, reinforce their importance and add discussions specific to pixel-wise prediction tasks like optical flow.

Quality changes with defense. Defending a method creates a modified method and thus modifies its quality characteristics. Therefore, the quality of the defended method Q_{D} should be explicitly reported. Particularly for pixel-wise prediction tasks, subtle changes in the inputs can cause significant output changes over large areas, making it indispensable to *reevaluate the defended quality*.

Use defense-aware attacks. Every defense proposal must be evaluated with a sufficiently *strong adaptive attack* [3, 7, 45] and report the pipeline robustness $R_{\mathrm{D}}^{\mathrm{D}}$. Showing that it withstands adversarial samples for the original method is *not enough*. While many defense-circumvention strategies for classification [45] may apply, individual tuning to pixel-wise prediction is needed for strong adaptive attacks.

Components matter. Defenses for specific components of complex methods should be evaluated on the *specific part*, not only on the full method. *E.g.* when defending optical flow for action recognition [2, 50], defended quality Q_{D} and defense-aware robustness $R_{\mathrm{D}}^{\mathrm{D}}$ should be reported for the flow component. Otherwise, the defense effectiveness and method sensitivity towards the component are entangled.

8. Limitations

This work solely focuses on detect-and-remove defenses for optical flow estimation. Hence, it covers neither defenses for problems unrelated to optical flow, nor optical flow defenses against non-patch attacks (of which none were published so far). Our findings that detection defenses do not protect against patch attacks could transfer to future defenses based on the gradient magnitude, as we found ILP-aware patches to transfer to LGS-defended methods and vice versa. Nonetheless, defending optical flow may be possible with more specialized techniques.

9. Conclusion

We investigated detect-and-remove defenses against adversarial patch attacks on optical flow methods. To this end, we designed defense-aware patches that avoid detection by LGS and ILP defenses, allowing us to break both defenses on a large variety of optical flow methods. On top of that, we found that both defenses reduce the optical flow quality and even failed to increase the robustness against standard (*i.e.*, not defense-aware) attacks. We could attribute this discouraging performance to the severe image quality degradation resulting from pixel replacements in the defenses. As image quality is crucial for pixel-wise motion estimation, this illustrates that defenses for classification methods, like LGS, do not automatically protect optical flow. Consequently, flow pipelines' robustness and quality must be thoroughly investigated for every defense, to promote trust instead of making empty promises.

Acknowledgments. We thank Filip Ilic for helpful discussions. The International Max Planck Research School for Intelligent Systems supports JS. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 251654672 – TRR 161 (B04).

References

[1] Shashank Agnihotri, Steffen Jung, and Margret Keuper. CosPGD: A unified white-box adversarial attack for pixel-wise prediction tasks. *arXiv:2302.02213*, 2023. 2

[2] Adithya Prem Anand, H. Gokul, Harish Srinivasan, Pranav Vijay, and Vineeth Vijayaraghavan. Adversarial patch defense for optical flow networks in video action recognition. In *Proc. IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 1289–1296, 2020. 1, 2, 3, 5, 6, 7, 8

[3] Anish Athalye, Nicholas Carlini, and David Wagner. Obfuscated gradients give a false sense of security: Circumventing defenses to adversarial examples. In *Proc. International Conference on Learning Representations (ICML)*, pages 274–283, 2018. 1, 2, 3, 8

[4] Wieland Brendel and Matthias Bethge. Approximating CNNs with bag-of-local-features models works surprisingly well on ImageNet. In *Proc. International Conference on Learning Representations (ICML)*, 2019. 2

[5] Tom B. Brown, Dandelion Mané, Aurko Roy, Martín Abadi, and Justin Gilmer. Adversarial patch. In *NeurIPS Workshop on Machine Learning and Computer Security (NeurIPS-MLCS)*, 2017. 2

[6] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black. A naturalistic open source movie for optical flow evaluation. In *Proc. European Conference on Computer Vision (ECCV)*, pages 611–625. Springer, 2012. 5

[7] Nicholas Carlini and David Wagner. Adversarial examples are not easily detected: Bypassing ten detection methods. In *ACM Workshop on Artificial Intelligence and Security (AiSec)*, pages 3–14, 2017. 1, 2, 3, 4, 8

[8] Nicholas Carlini and David Wagner. Towards evaluating the robustness of neural networks. In *IEEE Symposium on Security and Privacy (SP)*, pages 39–57, 2017. 4

[9] Joao Carreira and Andrew Zisserman. Quo vadis, action recognition? a new model and the Kinetics dataset. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 1

[10] Ping-Yeh Chiang, Renkun Ni, Ahmed Abdelkader, Chen Zhu, Christoph Studor, and Tom Goldstein. Certified defenses for adversarial patches. In *Proc. International Conference on Learning Representations (ICLR)*, 2020. 1, 2, 3

[11] Alexey Dosovitskiy, Philipp Fischer, Eddy Ilg, Philip Hausser, Caner Hazirbas, Vladimir Golkov, Patrick van der Smagt, Daniel Cremers, and Thomas Brox. FlowNet: Learning optical flow with convolutional networks. In *Proc. IEEE/CVF International Conference on Computer Vision (ICCV)*, 2015. 1, 4, 5

[12] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The KITTI dataset. *International Journal in Robotics Research (IJRR)*, 32(11):1231–1237, 2013. 5

[13] Thomas Gittings, Steve Schneider, and John Collomosse. Vax-a-Net: Training-time defence against adversarial patch attacks. In *Proc. Asian Conference on Computer Vision (ACCV)*, 2020. 2

[14] Jamie Hayes. On visible adversarial perturbations & digital watermarking. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1597–1604, 2018. 1, 2

[15] Zhaoyang Huang, Xiaoyu Shi, Chao Zhang, Qiang Wang, Ka Chun Cheung, Hongwei Qin, Jifeng Dai, and Hongsheng Li. FlowFormer: A transformer architecture for optical flow. In *Proc. European Conference on Computer Vision (ECCV)*, pages 668–685, 2022. 5, 8

[16] Filip Ilic, Thomas Pock, and Richard P. Wildes. Is appearance free action recognition possible? In *Proc. European Conference on Computer Vision (ECCV)*, pages 156–173, 2022. 1

[17] Nathan Inkawhich, Matthew Inkawhich, Yiran Chen, and Hai Li. Adversarial attacks for optical flow-based action recognition classifiers. *arXiv:1811.11875*, 2018. 1, 2

[18] Shihao Jiang, Dylan Campbell, Yao Lu, Hongdong Li, and Richard Hartley. Learning to estimate hidden motions with global motion aggregation. In *Proc. IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9772–9781, 2021. 4, 5

[19] Daniel Kondermann, Rahul Nair, Katrin Honauer, Karsten Krispin, Jonas Andrulis, Alexander Brock, Burkhard Gussefeld, Mohsen Rahimimoghaddam, Sabine Hofmann, Claus Brenner, and Bernd Jähne. The HCI benchmark suite: Stereo and flow ground truth with uncertainties for urban autonomous driving. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 19–28, 2016. 5

[20] Tom Koren, Lior Talker, Michael Dinerstein, and Ran Vitek. Consistent semantic attacks on optical flow. In *Proc. Asian Conference on Computer Vision (ACCV)*, pages 1658–1674, 2022. 2

[21] Alexey Kurakin, Ian J. Goodfellow, and Samy Bengio. Adversarial machine learning at scale. In *Proc. International Conference on Learning Representations (ICLR)*, 2017. 2, 4, 5

[22] Alexander Levine and Soheil Feizi. (De)randomized smoothing for certifiable defense against patch attacks. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Proc. Conference on Neural Information Processing Systems (NeurIPS)*, pages 6465–6475, 2020. 2

[23] Jiang Liu, Alexander Levine, Chun Pong Lau, Rama Chellappa, and Soheil Feizi. Segment and complete: Defending object detectors against adversarial patch attacks with robust patch detection. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14973–14982, 2022. 2

[24] Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adria Vladu. Towards deep learning models resistant to adversarial attacks. In *Proc. International Conference on Learning Representations (ICML)*, pages 1–10, 2018. 8

[25] Nikolaus Mayer, Eddy Ilg, Philip Hausser, Philipp Fischer, Daniel Cremers, Alexey Dosovitskiy, and Thomas Brox. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation. In *Proc. IEEE/CVF*

Conference on Computer Vision and Pattern Recognition (CVPR), pages 4040–4048, 2016. 5

[26] Michael McCoyd, Won Park, Steven Chen, Neil Shah, Ryan Roggenkemper, Minjune Hwang, Jason Xinyu Liu, and David Wagner. Minority reports defense: Defending against adversarial patches. In *Proc. International Conference on Applied Cryptography and Network Security Workshops (ACNSW)*, pages 564–582, 2020. 2

[27] Lukas Mehl, Jenny Schmalfuss, Azin Jahedi, Yaroslava Nalivayko, and Andrés Bruhn. Spring: A high-resolution high-detail dataset and benchmark for scene flow, optical flow and stereo. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4981–4991, 2023. 5

[28] Moritz Menze and Andreas Geiger. Object scene flow for autonomous vehicles. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3061–3070, 2015. 3, 5, 6, 7

[29] Norman Mu and David Wagner. Defending against adversarial patches with robust self-attention. In *ICML Workshop on Uncertainty and Robustness in Deep Learning (ICML-UDL)*, 2021. 2

[30] Muzammal Naseer, Salman Khan, and Fatih Porikli. Local gradients smoothing: Defense against localized adversarial attacks. In *Proc. IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 1300–1307, 2019. 1, 2, 5

[31] Nicolas Papernot, Patrick McDaniel, Ian Goodfellow, Somesh Jha, Z. Berkay Celik, and Ananthram Swami. Practical black-box attacks against machine learning. In *ACM Asia Conference on Computer and Communications Security (ASIA-CCS)*, pages 506–519, 2017. 3

[32] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. PyTorch: An imperative style, high-performance deep learning library. In *Proc. Conference on Neural Information Processing Systems (NeurIPS)*, pages 8024–8035, 2019. 4

[33] Aditi Raghunathan, Jacob Steinhardt, and Percy Liang. Certified defenses against adversarial examples. In *Proc. International Conference on Learning Representations (ICLR)*, 2018. 2

[34] Anurag Ranjan and Michael J. Black. Optical flow estimation using a spatial pyramid network. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 4, 5

[35] Anurag Ranjan, Joel Janai, Andreas Geiger, and Michael J. Black. Attacking optical flow. In *Proc. IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019. 1, 2, 4, 5, 6

[36] Sukrut Rao, David Stutz, and Bernt Schiele. Adversarial training against location-optimized adversarial patches. In *Proc. IEEE/CVF International Conference on Computer Vision Workshops (ECCVW)*, pages 429–448, 2020. 2

[37] Jenny Schmalfuss, Lukas Mehl, and Andrés Bruhn. Attacking motion estimation with adversarial snow. *ECCV Workshop on Adversarial Robustness in the Real World (ECCV-AROW)*, 2022. 2

[38] Jenny Schmalfuss, Lukas Mehl, and Andrés Bruhn. Distracting downpour: Adversarial weather attacks for motion estimation. In *Proc. IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023. 2

[39] Jenny Schmalfuss, Philipp Scholze, and Andrés Bruhn. A perturbation-constrained adversarial attack for evaluating the robustness of optical flow. In *Proc. European Conference on Computer Vision (ECCV)*, pages 183–200, 2022. 2, 4, 6

[40] Simon Schrodi, Tonmoy Saikia, and Thomas Brox. Towards understanding adversarial robustness of optical flow networks. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8916–8924, 2022. 2, 5, 6

[41] Deqing Sun, Xiaodong Yang, Ming-Yu Liu, and Jan Kautz. PWC-Net: CNNs for optical flow using pyramid, warping, and cost volume. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 4, 5

[42] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus. Intriguing properties of neural networks. In *Proc. International Conference on Learning Representations (ICLR)*, 2014. 2

[43] Zachary Teed and Jia Deng. RAFT: Recurrent all-pairs field transforms for optical flow. In *Proc. European Conference on Computer Vision (ECCV)*, pages 402–419, 2020. 4, 5, 7

[44] Alexandru Telea. An image inpainting technique based on the fast marching method. *Journal of Graphics Tools (JGT)*, 9(1):23–34, 2004. 3

[45] Florian Tramer, Nicholas Carlini, Wieland Brendel, and Aleksander Madry. On adaptive attacks to adversarial example defenses. In *Proc. Conference on Neural Information Processing Systems (NeurIPS)*, pages 1633–1645, 2020. 1, 2, 3, 4, 8

[46] Benjamin Wortman. Hidden patch attacks for optical flow. In *ICML Workshop on Adversarial Machine Learning (ICML-AdvML)*, 2021. 1, 2

[47] Chong Xiang, Arjun Nitin Bhagoji, Vikash Sehwag, and Prateek Mittal. PatchGuard: A provably robust defense against adversarial patches via small receptive fields and masking. In *Proc. USENIX Security Symposium*, 2021. 2

[48] Koichiro Yamanaka, Ryutaroh Matsumoto, Keita Takahashi, and Toshiaki Fujii. Adversarial patch attacks on monocular depth estimation networks. *IEEE Access*, 8:179094–179104, 2020. 1

[49] Koichiro Yamanaka, Keita Takahashi, Toshiaki Fujii, and Ryuraroh Matsumoto. Simultaneous attack on CNN-based monocular depth estimation and optical flow estimation. *IEICE Transactions on Information and Systems*, (5):785–788, 2021. 1, 2

[50] Lingyu Zhang, Chengzhi Mao, Junfeng Yang, and Carl Vondrick. Adversarially robust video perception by seeing motion. *arXiv:2212.07815*, 2022. 1, 2, 8