

Lipschitz Optimization for Formal Verification of Homographies

Jean-Guillaume Durand
 Joby Aviation

jpg.durand@jobyaviation.com

Maxime Gariel
 Joby Aviation

maxime.gariel@jobyaviation.com

Panagiotis Kouvaros
 Safe Intelligence

panagiotis@safeintelligence.ai

Alessio Lomuscio
 Safe Intelligence

alessio@safeintelligence.ai

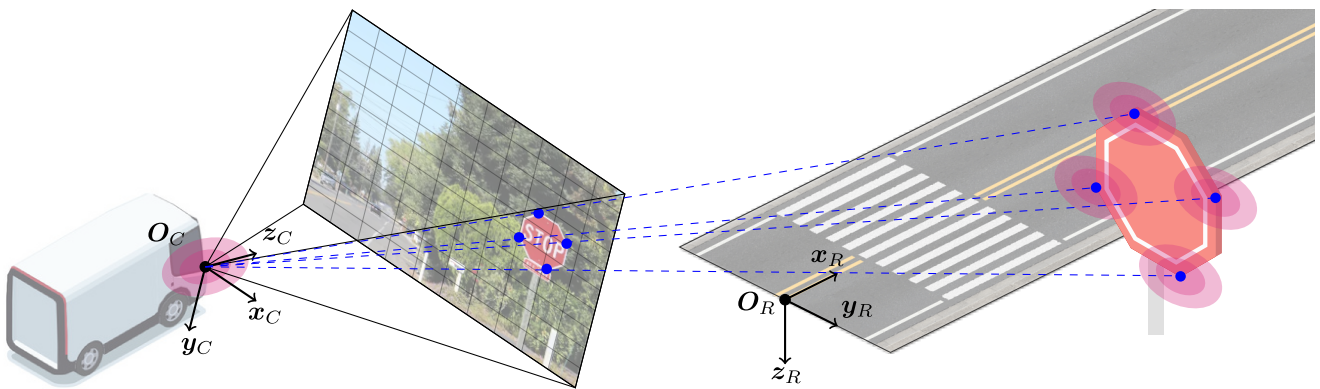


Figure 1. **Example application: autonomous driving.** The scene geometry defines the coordinate transform $T_C^R \in \text{SE}(3)$ from road frame (R) to camera frame (C). Assuming a planar traffic sign and a pinhole camera model, the stop sign features (blue points) are projected from the 3D world to the 2D image plane through projective geometry. Such a plane-to-plane mapping is a homography. We then study the robustness of traffic sign recognition to uncertainty or perturbations on the camera pose and stop sign localization (magenta ellipsoids). The formal verification of such homographies can support the certification of safety-critical applications in robotics.

Abstract

The adoption of vision neural networks in regulated industries requires formal robustness guarantees, especially in safety-critical domains such as healthcare, autonomous vehicles, and aerospace. However, current approaches are confined to incomplete statistical verification or robustness to ℓ_p -norm and affine transforms, which cover only a narrow subset of perturbations to the image formation process. In particular, robustness to camera motion remains an open problem despite being key to deploy many vision applications. We present a formal verification approach that targets robustness against 3D motion perturbations of the capturing camera. We first establish a closed-form mapping from camera pose to pixel values. By analyzing the continuity properties of the resulting homographies, we show that recent work on Lipschitz optimization and piecewise continuity can be extended to derive tight linear bounds on perturbed pixel values. Our approach applies to scenes with

predominantly planar structure, such as ground planes in augmented reality, road markings and traffic signs in autonomous driving, or planar workspaces in robotic manipulation. This enables the first formal verification of projective geometry transforms, without complex simulation, surrogate networks, or explicit image-formation models. We validate our implementation and show up to 89% speedup and 7% tighter bounds over prior work. We then evaluate our method on the VNN-COMP benchmark and reveal systematic weaknesses to projective perturbations. Finally, we demonstrate a real-world case study on a safety-critical runway classifier, highlighting practical vulnerabilities to camera motion, and addressing a key challenge in the certification of learned models. Data and code are publicly available at <https://github.com/jeangud/homography-verification>.

1. Introduction

The deployment of deep neural networks in safety-critical computer vision systems is rapidly accelerating across diverse applications, from augmented reality and medical imaging, to autonomous vehicles and aerospace (Fig. 1). Deep learning has proven a key enabler, significantly improving system performance and generalization [5, 14, 32]. However, these models exhibit instabilities and non-linear behaviors [19, 45], complicating robustness analysis and ultimately their regulatory compliance. Notwithstanding, for these systems to be trusted and certified in regulated domains, they require formal guarantees demonstrated under foreseeable operational scenarios [50]. Such robustness evidence is often produced by formal methods, which provide strong mathematical guarantees on system outputs.

To date, formal verification for neural networks (VNN) has largely focused on robustness against ℓ_p -bounded perturbations [9, 33], which act directly on the pixel space (e.g. contrast, color). This approach may not model some scenarios where perturbations originate in the physical space of the scene (e.g. viewpoint, shadows), before impacting image formation. While a sufficiently large ℓ_p -ball might contain the resulting images, it fails to capture the essential constraints defining the transformation. This approach also over-approximates the true manifold of perturbed images (Fig. 2), leading to uninformative loose bounds that are insufficient to prove network robustness [16, 37].

Consequently, verifying robotics applications requires handling a comprehensive range of perturbations within the Operational Design Domain (ODD), including 3D scene transformations. In this context, homographies emerge as a crucial class of transforms that capture perspective changes: they describe the inherent projective mapping between two different views of a planar surface [21]. This plane-to-plane mapping is ubiquitous in computer vision and is central to a vast array of use cases. For instance, augmented reality applications track ground planes for placing virtual objects. Autonomous cars identify planar objects like traffic signs, billboards, and road markings from various distances and angles. Similarly, tasks like image stitching, robotic manipulation, and visual servoing often depend on estimating homographies between views.

Existing research on 3D robustness is confined to either: statistical methods, which provide weaker guarantees than formal verification [39]; formal methods for 2D affine transforms [7, 43], which cannot model perspective effects; or complex simulations and explicit 3D scene models [25]. To bridge this gap, we position homographies as a principled and analytically tractable middle ground for predominantly planar scenes (i.e. limited parallax and occlusion).

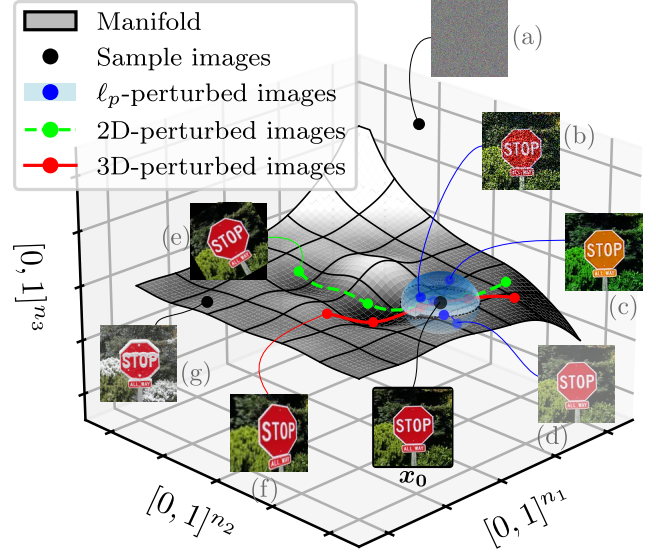


Figure 2. **The space of image perturbations (notional).** Most vectors in $[0, 1]^{h \times w}$ are noise (a). The subset of natural images forms a complex manifold of lower dimension (grey surface), on which lies our original image x_0 . The ball $B_p(x_0, \epsilon)$ of ℓ_p -norm perturbations (blue) can capture noise (b), hue (c) and brightness (d) variations for example. Beyond $B_p(x_0, \epsilon)$ and along constrained walks on the manifold, we may follow affine 2D perturbations (green) such as translation or rotation (e). We also encounter 3D perturbations (red): the same scene captured from a different perspective (f). Further on the manifold, lie more complex semantic perturbations like lens flare or snow (g). More natural images (e.g. different signs, cats) form other subsets of this manifold.

We propose the following contributions:

1. **Formal verification of projective transforms:** we derive closed-form homographies induced by 6 DOF pose perturbations. We show that these parameterized homographies are compatible with recent verification work using piecewise-linear constraints and Lipschitz optimization. By extending these techniques, we derive provably tight bounds on pixel values under non-affine transforms.
2. **Implementation and integration with existing verifiers:** we propose an efficient implementation of our method, introducing several accuracy and run time improvements to the PWL algorithm [7]. We also adapt the VENUS verifier to handle piecewise linear bounds. We profile and validate our implementation against PWL, demonstrating faster run times and tighter bounds.
3. **Benchmarking and case study:** on VNN-COMP benchmarks, we reveal systematic weaknesses of some networks to 3D camera-motion perturbations. We also instantiate our method on a runway visibility classifier, illustrating how homography-based verification can reveal practical vulnerabilities in safety-critical perception, and address key regulatory certification challenges.

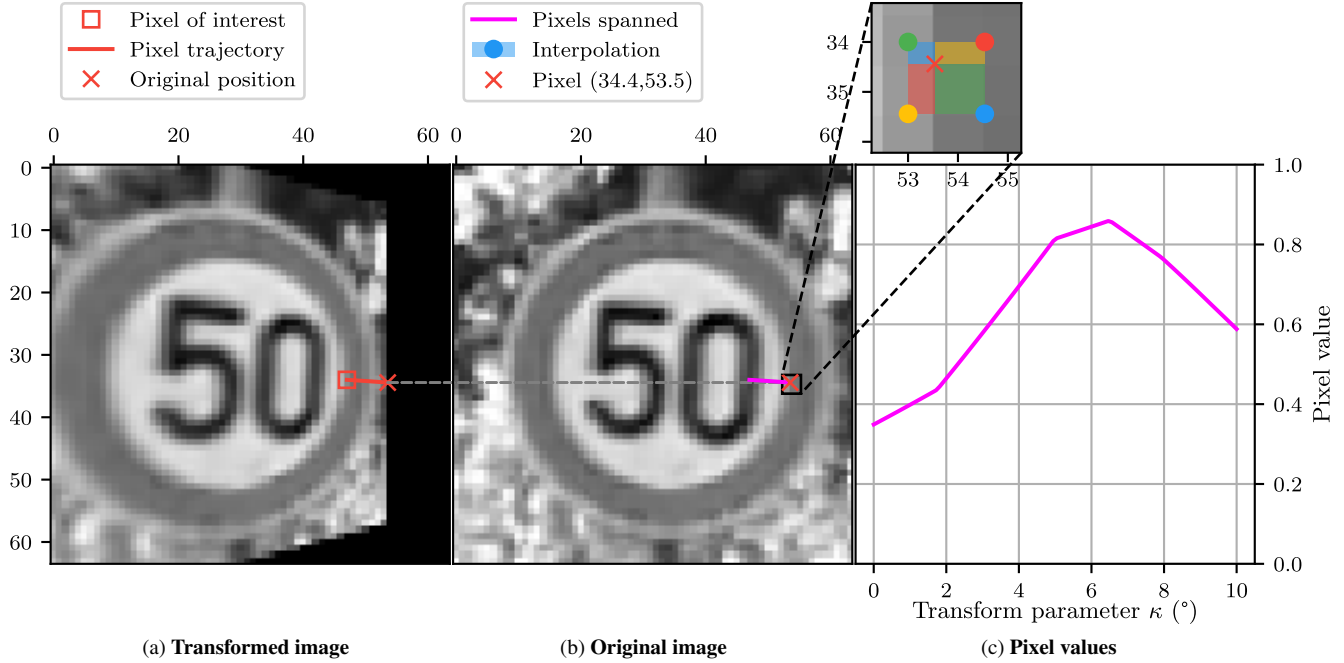


Figure 3. **Pixel values under perspective transform.** As original image x_0 (b) undergoes a yaw angle perturbation $\kappa \in \mathcal{B} = [0^\circ, 10^\circ]$, we look at the values $G_{34,47}^{x_0}(\kappa)$ of pixel $p = (34, 47)$, the red square in transformed image (a). Transform T is a homography and, as non-affine, is able to represent a change of perspective of the traffic sign. Applying the inverse transform, we find the corresponding position $T^{-1}(\kappa) = (34.4, 53.5)$ in the original image, the red cross in (a) and (b). We then compute the transformed pixel value with bilinear interpolation (zoomed region in (b)). As the yaw angle κ varies, pixel p follows a non-linear behavior in position (magenta path in (b)), and in values (c): transitioning from black (low values), to white (high values), to gray. Black padding is used to highlight the transform, but other padding techniques are possible for a more realistic rendering (see supplementary material Sec. 9).

2. Preliminaries

This section provides a brief theoretical exposition to key components of our proposed method.

Neural network verification. Let $\mathcal{I} \triangleq [0, 1]^{h \times w}$ be the set of all single-channel images of height h and width w . Multi-channel images are handled by applying this setup to each channel independently. We consider an original image $x_0 \in \mathcal{I}$ which undergoes a geometric transform T (Fig. 3). The parameters κ of the transform are confined to intervals $\mathcal{B} \subseteq \mathbb{R}^d$. As κ spans parameter space $\mathcal{B}$, the set of possible images obtainable from x_0 defines the *attack space*:

$$\Omega_{\mathcal{B}}(x_0) \triangleq \left\{ x \in \mathcal{I} \mid x_{i,j} = G_{i,j}^{x_0}(\kappa); \right. \\ \left. \forall \kappa \in \mathcal{B}, \quad \forall (i, j) \in \llbracket 0, h \rrbracket \times \llbracket 0, w \rrbracket \right\} \quad (1)$$

where $G_{i,j}^{x_0}(\kappa)$ is the value of pixel (i, j) in the image transformed from x_0 , by parameter κ . Neural network verification then over-approximates $\Omega_{\mathcal{B}}(x_0)$ using linear bounds for each pixel. To simplify notation in subsequent equations, we assume a fixed pixel (i, j) in the transformed image x from x_0 , and drop the indexing.

To obtain a pixel value in the transformed image, the corresponding coordinates in the original image are first computed through the preimage as $(u_0, v_0) = T^{-1}(\kappa)$. Since (u_0, v_0) may not correspond to exact pixel coordinates, bilinear interpolation is often applied to neighboring pixels for a better-looking result [7, Eq.2]. The pixel value in the transformed image is thus obtained as $G(\kappa) = I(u_0, v_0)$:

$$G_{i,j}^{x_0}(\kappa) \triangleq G(\kappa) = I \circ T^{-1}(\kappa) \quad (2)$$

Following [7], $G(\kappa)$ is then approximated by q piecewise-linear bounds $w, b \in \mathbb{R}^d$ as:

$$\begin{cases} \text{LB}(\kappa) = \max_{j \in \llbracket 1, q \rrbracket} \{ \underline{w}_j^\top \kappa + \underline{b}_j \} \leq G(\kappa) \\ \text{UB}(\kappa) = \min_{j \in \llbracket 1, q \rrbracket} \{ \overline{w}_j^\top \kappa + \overline{b}_j \} \geq G(\kappa) \end{cases} \quad (3)$$

These linear bounds can then be propagated through neural network verifiers to assure that the network output is unchanged under the perturbation (e.g. [7, eq.1]).

Homographies. A homography is a projective transformation that maps lines to lines [21, Sec.2.3]. In the context of computer vision, homographies arise in various scenarios such as central projection from 3D world to a 2D

camera plane (i.e. pinhole camera model), or when taking several images of the same planar surface. Using homogeneous coordinates, the homography between image points $\mathbf{p}_0 = (u_0, v_0, 1)^\top$ and $\mathbf{p} = (u, v, 1)^\top$ can be represented as a 3×3 matrix so that $\mathbf{p} = \mathbf{H}\mathbf{p}_0$. In Cartesian coordinates:

$$\begin{cases} u &= \frac{H_{1,1}u_0 + H_{1,2}v_0 + H_{1,3}}{H_{3,1}u_0 + H_{3,2}v_0 + H_{3,3}} \\ v &= \frac{H_{2,1}u_0 + H_{2,2}v_0 + H_{2,3}}{H_{3,1}u_0 + H_{3,2}v_0 + H_{3,3}} \end{cases} \quad (4)$$

Affine transforms are a special case where $H_{3,1} = H_{3,2} = 0$ and $H_{3,3} = 1$. For general homographies, this denominator is not 1. It represents a normalization by the depth along the optical ray, which models perspective distortions introduced by the imaging process.

Lipschitz optimization. A real-valued function $f : \mathcal{B} \rightarrow \mathbb{R}$ is said *Lipschitz-continuous* if its gradient is bounded, intuitively it is “limited in how fast it can change” (see Fig. 4b): $\exists L \in \mathbb{R}_+$ s.t. $\forall (\boldsymbol{\kappa}_0, \boldsymbol{\kappa}) \in \mathcal{B}^2$,

$$|f(\boldsymbol{\kappa}) - f(\boldsymbol{\kappa}_0)| \leq L \cdot \|\boldsymbol{\kappa} - \boldsymbol{\kappa}_0\|_2 \quad (5)$$

This effectively bounds the values taken by the function when stepping away from a given point $\boldsymbol{\kappa}_0$:

$$f(\boldsymbol{\kappa}_0) - L \cdot \|\boldsymbol{\kappa} - \boldsymbol{\kappa}_0\|_2 \leq f(\boldsymbol{\kappa}) \leq f(\boldsymbol{\kappa}_0) + L \cdot \|\boldsymbol{\kappa} - \boldsymbol{\kappa}_0\|_2 \quad (6)$$

Lipschitz maximization [3, 36] leverages this property: it evaluates f at n_{splits} sample points in the search space, and uses the Lipschitz constant to bound potential values of the global maximum f^* with the upper bound $\text{UB}(x) = \min_{i \in \llbracket 1, n_{\text{splits}} \rrbracket} (f(x_i) + L \cdot \|x - x_i\|_2)$ (see Fig. 4b). This process is iterated to refine the solution until a convergence criterion is met. In Batten et al. [7], the procedure is adapted to find an ε -optimal value $\widehat{f}^*$ such that $|f^* - \widehat{f}^*| \leq \varepsilon$.

3. Method

This section outlines our approach: we first model pose perturbations and derive closed-form solutions for homographies which arise from the geometry of many robotics problems. We then adapt Batten et al. [7] to compute piecewise-linear bounds for these non-affine transforms.

3.1. Parameterized homographies

We consider how ODD perturbations affect the projection of a planar scene feature (e.g. ground plane, traffic sign, road markings) onto the camera image plane (see Fig. 1). In the scope of geometric transformations, such perturbations can be a change in camera position, orientation, uncertainty in feature location, etc. We assume the camera lens distortion known, so that we can work with undistorted images and camera rays (e.g. pinhole camera model). We also neglect parallax effects so that there is no obstruction of the scene feature by objects as the viewpoint changes.

Under these assumptions, the feature plane induces a homography between different viewpoints as the camera is moving. From [21, Eq.13.2], this homography is given by:

$$\mathbf{H} = \mathbf{K} \left(\mathbf{R} - \frac{1}{d} \mathbf{t} \mathbf{n}^\top \right) \mathbf{K}_0^{-1} \quad (7)$$

with $\mathbf{K}_0 [\mathbf{I} | \mathbf{0}]$ and $\mathbf{K} [\mathbf{R} | \mathbf{t}]$ the projection matrices of the first and second viewpoint, and $\boldsymbol{\pi} = (\mathbf{n}^\top, d)^\top \in \mathbb{R}^4$ the homogeneous vector of the feature plane such that $\boldsymbol{\pi}^\top \mathbf{x} = 0$. All coordinates are expressed in the coordinate frame of the first camera. We assume the camera not aligned with the plane, ignoring the degenerate case $d = 0$.

We now derive the form of this homography for a plane $z = 0$ in the world frame (W), thus given by $\boldsymbol{\pi}^W = (0, 0, 1, 0)^\top$. We use notations from Sec. 7 in the supplementary material. We assume that both viewpoints are acquired by the same camera so that $\mathbf{K} = \mathbf{K}_0$. The plane equation becomes $0 = (\boldsymbol{\pi}^W)^\top \mathbf{x}^W = (\boldsymbol{\pi}^W)^\top \mathbf{T}_{C_0}^W \mathbf{x}^{C_0}$, with $\mathbf{T}_{C_0}^W = [\mathbf{R}_{C_0}^W | \mathbf{t}_{W \rightarrow C_0}^W]$ the passive homogeneous transform from the first camera (C_0) coordinates, to the world (W) coordinates. This gives the plane equation in the first camera frame as $(\boldsymbol{\pi}^{C_0})^\top \mathbf{x}^{C_0} = 0$ with $\boldsymbol{\pi}^{C_0} = (\mathbf{T}_{C_0}^W)^\top \boldsymbol{\pi}^W$ [21, Eq.3.6]. The transform $\mathbf{T}_{C_0}^W$ corresponds to the first camera pose, and is parameterized by the camera Euler angles roll ϕ , θ , ψ for roll, pitch, yaw, and the translation by $\mathbf{t}_{W \rightarrow C_0}^W = (x, y, z)^\top$. Note that the rotation angles are defined with respect to x -forward, y -right, as opposed to the typical camera frame convention with z -forward and x -right. In the same fashion, the transform from first camera (C_0) to second camera frame (C) coordinates is $\mathbf{T}_{C_0}^C = [\mathbf{R} | \mathbf{t}] = [\mathbf{R}_{C_0}^C | \mathbf{t}_{C \rightarrow C_0}^C]$, parameterized by a delta in pose $(\Delta\phi, \Delta\theta, \Delta\psi, \Delta x, \Delta y, \Delta z)^\top$. Overall, the homography $\mathbf{H}(\boldsymbol{\kappa})$ between two camera viewpoints is parameterized by vector $\boldsymbol{\kappa} = (\Delta\phi, \Delta\theta, \Delta\psi, \Delta x, \Delta y, \Delta z, \phi, \theta, \psi, x, y, z)^\top$, and given by Eq. (7) with:

$$\mathbf{K} = \mathbf{K}_0 = \begin{bmatrix} f & 0 & x_c \\ 0 & f & y_c \\ 0 & 0 & 1 \end{bmatrix} \quad (7a)$$

$$\mathbf{R} = \mathbf{R}_{C_0}^C(\Delta\phi, \Delta\theta, \Delta\psi) \quad (7b)$$

$$\mathbf{t} = \mathbf{t}_{C \rightarrow C_0}^C(\Delta x, \Delta y, \Delta z) \quad (7c)$$

$$\mathbf{n} = (\pi_1, \pi_2, \pi_3)^\top \text{ and } d = \pi_4 \quad (7d)$$

$$\boldsymbol{\pi} = \boldsymbol{\pi}^{C_0} = [\mathbf{T}_{C_0}^W(\phi, \theta, \psi, x, y, z)]^\top \boldsymbol{\pi}^W \quad (7e)$$

with f the camera focal length and (x_c, y_c) the principal point. The full expressions of $\mathbf{R}_{C_0}^C$ and $\mathbf{t}_{C \rightarrow C_0}^C$ are given in supplementary material Sec. 11.

We then restrict the perturbation space $\mathcal{B}$ to specific scenarios, which simplify the form of $\mathbf{H}(\boldsymbol{\kappa})$. We focus here

on an example perturbation $\Delta\psi$ in yaw angle, with all other Δ parameters set to zero. The result is given in Eq. (8) and the complete derivation in supplementary material Sec. 12. Note that since the camera center is fixed, Eq. (8) does not depend on the camera pose $(\phi, \theta, \psi, x, y, z)^\top$. This is a special case where $\mathbf{H} = \mathbf{K}\mathbf{R}\mathbf{K}^{-1}$ remains valid even if the projected feature is non-planar [21, Sec.8.4.5]. Additional scenarios are detailed in supplementary material Sec. 10.

$$\mathbf{H}(\kappa = \Delta\psi) = \begin{bmatrix} \cos(\Delta\psi) + \frac{x_c \sin(\Delta\psi)}{f} & 0 & \frac{(-f^2 - x_c^2) \sin(\Delta\psi)}{f} \\ \frac{y_c \sin(\Delta\psi)}{f} & 1 & \frac{y_c(f \cos(\Delta\psi) - f - x_c \sin(\Delta\psi))}{f} \\ \frac{\sin(\Delta\psi)}{f} & 0 & \cos(\Delta\psi) - \frac{x_c \sin(\Delta\psi)}{f} \end{bmatrix} \quad (8)$$

3.2. Homography verification

With $\mathbf{H}(\kappa)$ known, the value of transformed pixel $\mathbf{p} = (u, v, 1)^\top$ is given by the preimage and interpolation from Eq. (2). T^{-1} is now the inverse of the homography Eq. (7), as opposed to an affine transform. The pixel position $T^{-1}(\kappa) = \mathbf{p}_0 = (u_0, v_0, 1)^\top$ in the original image is then:

$$(u_0(\kappa), v_0(\kappa)) = \left(\frac{h_1(\kappa)}{h_3(\kappa)}, \frac{h_2(\kappa)}{h_3(\kappa)} \right) \quad (9)$$

with $\mathbf{h} = \mathbf{H}^{-1}(\kappa) \mathbf{p}$. In our example of yaw deviation Eq. (8) with $\kappa = \Delta\psi$, this corresponds to:

$$\begin{cases} u_0(\Delta\psi) = x_c + f \cdot \frac{f \sin(\Delta\psi) + (u - x_c) \cos(\Delta\psi)}{f \cos(\Delta\psi) - (u - x_c) \sin(\Delta\psi)} \\ v_0(\Delta\psi) = y_c + f \cdot \frac{v - y_c}{f \cos(\Delta\psi) - (u - x_c) \sin(\Delta\psi)} \end{cases} \quad (10)$$

For verification, we then over-approximate $G(\Delta\psi) = I[u_0(\Delta\psi), v_0(\Delta\psi)]$ with linear bounds that can be efficiently propagated through a neural network verifier. Given the high non-linearity of geometric transforms, especially with perspective distortions, piecewise-linear bounds enable a tighter approximation of $G(\kappa)$.

Following Batten et al. [7], we first partition the parameter space $\mathcal{B} = \bigcup_{j \in [1, q]} \mathcal{B}_j$. For each subdomain, we evaluate the pixel value function at n_s sample points, and derive an unsound linear bound by minimizing the error to the bound. For a lower bound, the relaxed linear program is:

$$(\mathbf{w}_j^*, \mathbf{b}_j^*) = \arg \min_{\mathbf{w}_j, \mathbf{b}_j} \left(\frac{1}{n_s} \cdot \sum_{\kappa_i \in \mathcal{B}_j} [G(\kappa_i) - (\mathbf{w}_j^\top \kappa_i + \mathbf{b}_j)] \right) \quad (11)$$

An overall unsound bound LB is then obtained by combining the piecewise segments from all subdomains (Eq. (3)). Following [7], we then apply Lipschitz maximization (see Fig. 4) to cost function $\underline{J}(\kappa) = \text{LB}(\kappa) - G(\kappa)$,

Require: $J, \mathcal{B}, \epsilon > 0$

Ensure: estimated max. $\widehat{J}^*$ with certificate ϵ

```

1:  $J_{\max} \leftarrow 0$ 
2:  $(\mathcal{B}_i)_{i \in [1, n_{\text{splits}}]} \leftarrow \text{split}(\mathcal{B})$ 
3: while  $\text{Card}(\mathcal{B}) > 0$  do
4:    $\mathcal{B}_j \leftarrow \text{pop}[(\mathcal{B}_i)_i]$ 
5:    $\mathcal{J}_{\text{samples}} \leftarrow \text{sample}(J, \mathcal{B}_j)$ 
6:    $J_{\max} \leftarrow \max(\{J_{\max}\} \cup \mathcal{J}_{\text{samples}})$ 
7:    $L \leftarrow \sup_{\kappa \in \text{Diff}(\mathcal{B})} |\nabla_{\kappa}^\top J(\kappa) \cdot e|$ 
8:    $[\kappa_0, \kappa_n] \leftarrow \mathcal{B}_j$ 
9:    $J_{\text{bound}} \leftarrow \frac{J(\kappa_0)}{2} + \frac{J(\kappa_n)}{2} + \frac{L}{2} \cdot \|\kappa_n - \kappa_0\|$ 
10:  if  $J_{\text{bound}} > J_{\max} + \epsilon$  then
11:     $\mathcal{B} \leftarrow \mathcal{B} \cup \text{split}(\mathcal{B}_j)$ 
12:  end if
13: end while
14: return  $\widehat{J}^* = J_{\max} + \epsilon$ 

```

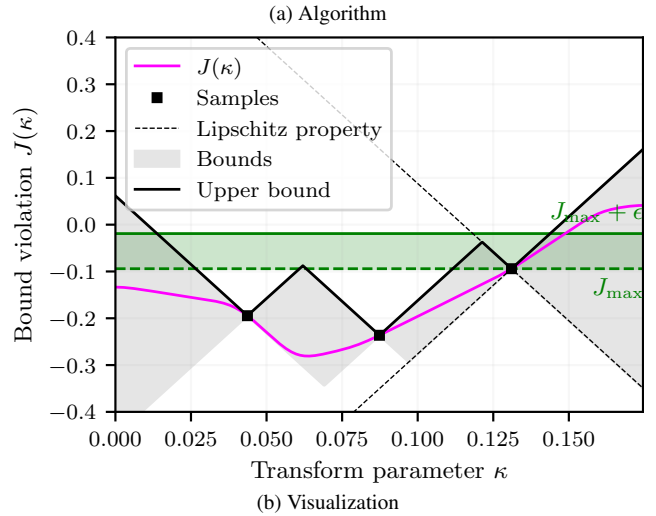


Figure 4. **Lipschitz optimization.** Adapted from [7], procedure 4a finds an ϵ -maximum of function J . On Fig. 4b, a few samples (black) approximate the true function (magenta). The Lipschitz property of J constrains, with slope $\pm L$, possible values for the function (gray bands). In this example, since the maximum of the Lipschitz bound (black) is above $J_{\max} + \epsilon$, the interval is split a few more times to better estimate $J_{\max}$, and capture a ϵ -optimum.

which represents the bound violation. By finding an analytical ϵ -majorant $\widehat{J}^*$, we can guarantee $\forall \kappa, \underline{J}(\kappa) \leq \widehat{J}^* + \epsilon \Leftrightarrow \forall \kappa, \text{LB}(\kappa) - (\widehat{J}^* + \epsilon) \leq G(\kappa)$.

The shifted bound $\text{LB}^*(\kappa) = \text{LB}(\kappa) - (\widehat{J}^* + \epsilon)$ thus constitutes a *sound* lower bound to G . A similar procedure is followed for the upper bound. Batten et al. [7] showed that for a given number of splits, the derived bounds are optimal, irrespective of the downstream robustness task.

The missing argument is then to find a valid upper bound

to the Lipschitz constant. In [7], it is estimated with the directional gradient as $L_m = \sup_{\kappa \in \text{Diff}(\mathcal{B})} |\nabla_{\kappa}^{\top} J(\kappa) \cdot e_m|$ for coordinate m . We extend this formula in the case of homographies where the denominator $h_3(\kappa)$ may render G non-continuous (Eq. (9)) and break its Lipschitz properties. With additional scenarios developed in supplementary material Sec. 10, our yaw perturbation gradient from Eq. (8) is:

$$\nabla_{\Delta\psi}^{\top} T^{-1}(\Delta\psi) = \begin{pmatrix} \frac{f[f^2 + (u-x_c)^2]}{[f \cos(\Delta\psi) - (u-x_c) \sin(\Delta\psi)]^2} \\ \frac{f(v-y_c)[f \sin(\Delta\psi) + (u-x_c) \cos(\Delta\psi)]}{[f \cos(\Delta\psi) - (u-x_c) \sin(\Delta\psi)]^2} \end{pmatrix} \quad (12)$$

From Eq. (10) we first notice that T^{-1} is not continuous at the critical angle $\Delta\psi_c = \arctan\left(\frac{f}{u-x_c}\right)$ if $u \neq x_c$, and $\frac{\pi}{2}$ otherwise. Thus $\text{Diff}(\mathcal{B}) = \mathcal{B} \setminus \{\Delta\psi_c \pmod{\pi}\}$. For the scope of small yaw perturbations around $\Delta\psi = 0^\circ$, the pixel-dependent critical angle does not belong to the parameter space $\mathcal{B}$. We also note that for diverging pixel coordinates at $\Delta\psi_c$, the pixel function G would remain bounded as the interpolation uses padding values. In practice, we found $|\Delta\psi_c| \in [72^\circ, 90^\circ] \not\subset \mathcal{B}$. With $\text{Diff}(\mathcal{B})$ now defined, we derive candidate locations for maximum gradient values (proof in supplementary material Sec. 14). From Eq. (12), and for $\mathcal{B} = [\Delta\psi_{\min}, \Delta\psi_{\max}]$:

$$\arg \max_{\Delta\psi \in \text{Diff}(\mathcal{B})} |\nabla_{\Delta\psi} u_0(\Delta\psi)| \in \left\{ \Delta\psi_{\min}, \Delta\psi_{\max}, \arctan\left(\frac{x_c - u}{f}\right) \pmod{\pi} \right\} \quad (13)$$

$$\arg \max_{\Delta\psi \in \text{Diff}(\mathcal{B})} |\nabla_{\Delta\psi} v_0(\Delta\psi)| \in \{\Delta\psi_{\min}, \Delta\psi_{\max}\} \quad (14)$$

For the gradient of the discontinuous interpolation function, we use the maximum gradient from the different interpolation regions [6, 7]. Finally, remembering our notation simplification $G = G_{i,j}^{x_0}$, the process is repeated for each pixel (i, j) to obtain bounds for all pixels of the transformed image. These bounds are then passed to a verifier to guarantee the network output is unchanged under the perturbation.

4. Evaluation

This section introduces our experimental setup and results. Experiments are performed on a laptop using an Intel i7-13800H processor (20 threads, 5GHz), 32GB RAM, and a NVIDIA RTX 2000 GPU (driver 550.120, CUDA 12.4).

Datasets. We evaluate our method on the VNN competition benchmarks [9]. We choose datasets of increasing complexity (model and image size) that are relevant to robotics applications: digit classification (MNIST), object recognition (CIFAR-10), and autonomous driving (GT-SRB). We use 100 images from each test set, and estimate the focal length to yield reasonable angular perturbations

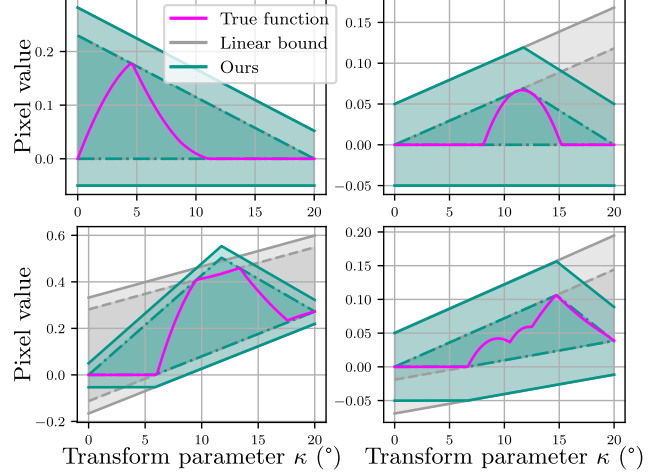


Figure 5. **Example bounds.** Pixel values (magenta) are over-approximated. Unsound bounds (dashed) are shifted to sound bounds (solid) by Lipschitz optimization. Depending on complexity, curves have zero (linear), one (PWL [7]), or two splits (ours).

that are similar between datasets (see supplementary material Sec. 10). Additionally, we use a runway visibility classifier developed on the LARD dataset [14].

Verification. Our bounds are solver-agnostic, decoupling transformation modeling (front end) from the network verification engine (back end). We choose VENUS as the only baseline supporting a MILP formulation for *complete* robustness of piecewise-linear bounds. We enhanced its bound-propagation framework to derive pre-activation bounds from linear input constraints rather than fixed intervals. This enables VENUS to construct tighter ReLU relaxations and better constrain the underlying MILP problem.

Metrics. First, we evaluate our method independently of any downstream verification task: polytope tightness, number of branch-and-bound steps (BaB), and run time per image. Then, we use our bounds to evaluate model robustness: we measure average run times, and number of robust cases (i.e. a correct model output is unchanged by the transform).

Validation. Analytical Lipschitz constants were validated with finite differences by oversampling the pixel curve. Unit tests validated VENUS [8] modifications for piecewise-linear bounds. For re-implementing PWL, we profiled the code and enforced bit-identical results at multiple stages of the algorithm, before adding our modifications.

4.1. Comparison to Batten et al. [7]

We re-implement PWL [7] to handle perspective transforms by adding closed-form homographies, associated gradients, and Lipschitz constants. Accuracy and exactness are improved through algorithmic changes: 1) half-pixel corrections for exact gradients; 2) inclusion of omitted κ values in $\text{Diff}(\mathcal{B})$ to yield extra Lipschitz constant candidates; 3)

dichotomy on $\mathcal{B}$ from the first iteration, reducing branch-and-bound steps; 4) tighter bounds by using $q = 2$ for both upper and lower bounds; and 5) parallelization at the pixel level. Since PWL supports only affine transforms, we compare implementations under 2D rotations (Tab. 1).

5 degrees	MNIST		CIFAR-10	
	PWL	Ours	PWL	Ours
Polytope area ($\times 10^{-3}$) $\downarrow$	9.48	9.42	10.94	10.83
BaB steps $\downarrow$	124.0	17.5	124.0	35.5
Time (s) $\downarrow$	87.6	7.0	910.3	98.7
Time @ fixed BaB (s) $\downarrow$	87.6	12.9	910.3	148.1
20 degrees				
Polytope area ($\times 10^{-2}$) $\downarrow$	9.33	8.81	15.1	14.0
BaB steps $\downarrow$	137.0	72.5	124.2	90.5
Time (s) $\downarrow$	90.3	36.2	891.0	530.9
Time @ fixed BaB (s) $\downarrow$	90.3	24.0	891.0	415.2

Table 1. Average performance comparison with PWL [7].

High run times are intrinsic to *complete* branch-and-bound verifiers; no cheaper alternative provides complete certificates to piecewise-linear bounds. Our implementation reduces BaB iterations by more than 71% on small transforms, for an 89% total speedup. On larger transforms, this fades to 40% as the overhead of splitting non-converging candidates becomes dominant. At a fixed BaB budget, our per-pixel parallelization is still 53-85% faster. The choice of $q = 2$ yields marginally tighter bounds for small transforms, and up to 7% on more complex pixel curves (Fig. 5).

4.2. VNN competition benchmarks

We perform the first robustness evaluation of the VNN benchmark networks [9] against perspective transforms. We generate and propagate bounds through VENUS to verify network robustness against the benchmark specifications (Tab. 2). We use black padding, a Lipschitz-error of 0.01, and 5,000 maximum BaB iterations.

Perturbation	$\mathcal{B}$	MNIST	CIFAR-10	GTSRB
$\Delta\phi$ (roll)	$[0, 5]^\circ$	61	69	0
$\Delta\theta$ (pitch)	$[0, 5]^\circ$	5	10	0
$\Delta\psi$ (yaw)	$[0, 5]^\circ$	25	7	0
Δx	$[0, 1]_m$	50	21	0
Δy	$[0, 1]_m$	73	83	0
Δz	$[0, 1]_m$	53	24	0
Generation time (s) $\downarrow$		42.0	1858.6	292.6
Verification time (s) $\downarrow$		8.23	41.33	5323.8
Total timeouts $\downarrow$		0	2	178

Table 2. Ratio (%) of robust cases to non-affine perturbations.

Tab. 2 shows that even PGD-trained networks can be highly vulnerable to 3D geometric perturbations. MNIST and CIFAR models are less robust to pitch, yaw, Δx and Δz transforms than to the ℓ_p -norm perturbations reported in [4].

Roll and Δy perturbations appear more robust since, under our assumptions, they correspond exactly to affine transforms (rotation and shear, Sec. 10) better approximated by ℓ_p -norm attacks and commonly included in data augmentation. More surprisingly, the autonomous-driving models exhibit no certified robustness to 3D perturbations on GT-SRB, reinforcing prior evidence that these networks lack robustness even to simpler ℓ_p -norm attacks [9, 30].

4.3. Case study: runway visibility

Following current aviation safety recommendations [1, 2], we illustrate our method on a low-criticality application: a binary classifier that predicts runway visibility, informing the pilot’s decision to continue the approach or initiate a go-around. The network was trained on LARD [14] without robustness objectives. Applying our homography-based verification, we found limited robustness: only 16% of test images were certified under 10 cm translations and 1% under 1° rotations. This underscores the difficulty of achieving robustness to camera motion even in simple classifiers, and motivates future work on robust training in this setting.

4.4. Sensitivity and limitations

Amplitude. Large transforms exacerbate the non-linearities of the pixel curve and complicate its approximation by piecewise-linear bounds. Tab. 3 shows how certified robustness decreases with amplitude. This is a well-known challenge in formal verification, typically addressed through *input splitting* [52, 55] by partitioning the domain into smaller intervals: a near-linear regime precisely consistent with homography modeling and our planar assumptions.

Yaw angle	1°	2°	3°	4°	5°	10°	15°	20°
Verified (%)	76	68	57	41	25	0	0	0

Table 3. Robustness to yaw amplitude on MNIST.

Padding. Bilinear interpolation uses a padding value to fill in missing pixels transformed from outside the original image bounds (Fig. 3). Tab. 4 shows how padding that duplicates image content offers better robustness than padding with stark borders, which induce distribution shift. While black padding remains standard in benchmarks and aligns with CIFAR/MNIST training augmentations, it risks exposing GTSRB models to border artifacts (e.g., vignetting) rather than geometric shifts. However, substituting realistic reflection padding yields unchanged experiment results, reiterating the known brittleness of these networks [9, 30].

$\mathcal{B}$	Black	Gray	Replicate	Reflect	Wrap
$[0, 2.5]^\circ$	34	33	42	42	35
$[0, 5]^\circ$	7	7	8	9	7

Table 4. Robustness to yaw padding on CIFAR-10.

5. Related Work

We focus on the verification of pre-trained models, without considering *robust-training* techniques [56].

Robustness of neural networks. Neural networks, while capable of approximating complex functions, exhibit inherent instability stemming from repeated compositions of linear and non-linear operations [19, 45]. This vulnerability to input perturbations raises significant concerns about their reliability in safety-critical applications in the real-world, where inputs are subject to noise, distortions, or adversarial manipulations. Verification thus aims to provide robustness guarantees against input perturbations [26, 33, 34]. Without loss of generality, most methods focus on the feed-forward fully-connected architecture [26].

ℓ_p -norm perturbations. In this context, robustness to ℓ_p -norm perturbations is often modeled as an optimization problem [33]. *Incomplete* verifiers scale to large vision models, but may not always provide conclusive answers. In contrast, *complete* verifiers such as Reluplex [29], Marabou [52], or VENUS [8] focus on exact formulations through satisfiability modulo theory (SMT) or mixed-integer linear programming (MILP). They provide definitive satisfiability answers but are limited to small networks.

Interval bound propagation (IBP), used in DeepPoly [43] and Lipschitz-based methods [7], offers faster verification by relaxing the ReLU polytope, but often over-approximates the underlying optimisation constraints. Probabilistic methods [39] scale well and can handle arbitrary *black-box* architectures, but provide weaker, probabilistic guarantees than formal methods. The Branch-and-bound (BaB) paradigm combines complete and incomplete verification, thereby scaling to large models while still providing definitive robustness answers. BaB is the current state of the art, with α - β -CROWN [42, 49, 54, 55, 57, 58] dominating the past few VNN-COMP leaderboards [9].

Despite significant progress in ℓ_p -norm verification, recent work has highlighted its limitations in capturing the complexity of real-world perturbations [37]. As illustrated in Fig. 2, the space of possible perturbations extends far beyond the confines of ℓ_p -balls [13, 16]: pixel flow [53], data augmentation [18], multi-norm [27], universal adversarial perturbations (UAPs) [38], distribution [51], Wasserstein W_p -balls [44], composition [22], etc. In particular, ℓ_p -bounded attacks cannot directly model geometric perturbations such as changes of the camera viewpoint.

Geometric perturbations. Through non-linear mappings of pixel coordinates, geometric transformations exacerbate the inherent non-linearity of neural networks, and their vulnerability to input variations [16, 17, 28, 33]. While the set of rotations of an image does not form an ℓ_p -ball, a common approach consists in bounding the rotation parameters to retrieve a problem formulation that is compati-

ble with ℓ_p -norm verifiers [43, 48, 56]. Balunovic et al. [6] linked geometric parameter to pixel value bounds but for 2D transforms only. They also contributed a method for verifying bilinear interpolation, commonly used with geometric transforms. Batten et al. [7] achieved tighter bounds by employing piecewise-linear approximation and Lipschitz optimization, again for 2D transforms. Their approach generalizes to matrix-defined affine transforms [6, 15, 22, 31].

Beyond affine transforms, Hu et al. [23, 24, 25] verify models against camera motion perturbations with pixel-wise smoothing, but require a point cloud of the scene. The Lipschitz procedures from DeepGO [40] and GeoRobust [47] are used to analyze the robustness *of the network*, rather than for linearly approximating the input bounds as we propose. ManiFool [28] models projective transforms by traversing the perturbation manifold to find counter-examples. However, the resulting metric is difficult to compare with standard robustness specifications (e.g., class score margin). Hanspal and Lomuscio [20] abstract complex 3D transforms through an encoding head, and verify specifications in the resulting latent space. This approach is very efficient but modifies the network, and does not leverage potential closed-form solutions.

Homographies in safety-critical applications. Homographies are central to many robotics and vision systems that exploit planar geometry. In autonomous driving, inverse perspective mapping (IPM) uses a ground-plane homography to generate bird’s-eye-view representations for lane detection and scene understanding [10, 35]. In robotic surgery, homographies support calibration to align robot-mounted cameras with the workspace [46], and augmented reality to overlay anatomical models from endoscopic or X-ray views [12]. Finally, for autonomous drones, vision systems often detect a known planar pattern (e.g., an AprilTag marker or helipad) and use homographies to estimate the vehicle’s relative pose for landing [11, 41].

6. Conclusion

Moving beyond the limitations of ℓ_p -norm and affine robustness, we propose the first general framework to formally verify 3D projective transforms, with applications across core computer-vision and robotics settings (e.g. augmented reality, autonomous vehicles, and manipulation). Our work derives closed-form homographies from 6-DOF camera pose and extends the PWL algorithm to compute provably tight pixel bounds, achieving up to 89% faster run times and 7% tighter bounds. Our experiments reveal previously unreported vulnerabilities of benchmark networks to 3D camera motions. Finally, a case study on a runway classifier illustrates how homography-based verification can be applied to perception modules in aerospace, marking progress towards the certification of AI-powered systems in safety-critical applications.

Acknowledgments

The authors would like to thank Ben Batten for helpful discussions on the PWL algorithm and the VENUS verifier.

References

- [1] Federal Aviation Administration. Roadmap for artificial intelligence safety assurance. Technical Report Version I, FAA, 2024. 31 pages. 7
- [2] European Union Aviation Safety Agency. EASA concept paper: Guidance for level 1 & 2 machine-learning applications. Technical Report Issue 02, EASA, 2024. 7
- [3] François Bachoc, Tommaso R Cesari, and Sébastien Gerchinovitz. The sample complexities of global Lipschitz optimization. *hal-03129721v2*, 2021. 4
- [4] Stanley Bak, Changliu Liu, and Taylor Johnson. The second international verification of neural networks competition (VNN-COMP 2021): Summary and results. *arXiv preprint arXiv:2109.00498*, 2021. 7
- [5] Giovanni Balduzzi, Martino Ferrari Bravo, Anna Chernova, Calin Cruceru, Luuk van Dijk, Peter de Lange, Juan Jerez, Nathanaël Koehler, Mathias Koerner, Corentin Perret-Gentil, Zoltan Pillio, Ruben Polak, Hugo Silva, Romeo Valentin, Ian Whittington, and Grigory Yakushev. Neural network based runway landing guidance for general aviation autoland. Technical Report DOT/FAA/TC-21/48, Federal Aviation Administration, 2021. 2
- [6] Mislav Balunovic, Maximilian Baader, Gagandeep Singh, Timon Gehr, and Martin Vechev. Certifying geometric robustness of neural networks. In *NeurIPS*, 2019. 6, 8
- [7] Ben Batten, Yang Zheng, Alessandro De Palma, Panagiotis Kouvaros, and Alessio Lomuscio. Verification of geometric robustness of neural networks via piecewise linear approximation and Lipschitz optimisation. In *ECAI*, pages 2362–2369, 2024. 2, 3, 4, 5, 6, 7, 8
- [8] Elena Botoeva, Panagiotis Kouvaros, Jan Kronqvist, Alessio Lomuscio, and Ruth Misener. Efficient verification of ReLU-based neural networks via dependency analysis. In *AAAI Conference on Artificial Intelligence*, 2020. 6, 8
- [9] Christopher Brix, Stanley Bak, Changliu Liu, and Taylor T Johnson. The fourth international verification of neural networks competition (VNN-COMP 2023): Summary and results. In *CAV*, 2023. 2, 6, 7, 8
- [10] Tom Bruls, Horia Porav, Lars Kunze, and Paul Newman. The right (angled) perspective: Improving the understanding of road scenes using boosted inverse perspective mapping. *2019 IEEE Intelligent Vehicles Symposium (IV)*, pages 302–309, 2018. 8
- [11] Andres Chavez, Dalton L’Heureux, Nirmal Prabhakar, Matthew Clark, Wai-Leuk Law, and Richard J Prazenica. Homography-based state estimation for autonomous UAV landing. In *AIAA Information Systems-AIAA Infotech@Aerospace*, 2017. 8
- [12] Long Chen, Wen Tang, and Nigel W John. Real-time geometry-aware augmented reality in minimally invasive surgery. *Healthcare technology letters*, 4(5):163–167, 2017. 8
- [13] Sihui Dai, Chong Xiang, Tong Wu, and Prateek Mittal. Position: Beyond robustness against single attack types. *arXiv preprint arXiv:2405.01349*, 2024. 8
- [14] Mélanie Ducoffe, Maxime Carrere, Léo Féliers, Adrien Gauffriau, Vincent Mussot, Claire Pagetti, and Thierry Sammour. LARD - Landing Approach Runway Detection - Dataset for Vision Based Landing. 2023. 2, 6, 7
- [15] Beranger Dumont, Simona Maggio, and Pablo Montalvo. Robustness of rotation-equivariant networks to adversarial perturbations. *arXiv preprint arXiv:1802.06627*, 2018. 8
- [16] Logan Engstrom, Brandon Tran, Dimitris Tsipras, Ludwig Schmidt, and Aleksander Madry. Exploring the landscape of spatial robustness. In *ICML*, 2019. 2, 8
- [17] Alhussein Fawzi, Seyed-Mohsen Moosavi-Dezfooli, and Pascal Frossard. The robustness of deep networks: A geometrical perspective. *IEEE Signal Processing Magazine*, 34(6):50–62, 2017. 8
- [18] Xiang Gao, Ripon K Saha, Mukul R Prasad, and Abhik Roychoudhury. Fuzz testing based data augmentation to improve robustness of deep neural networks. In *ICSE*, pages 1147–1158. IEEE, 2020. 8
- [19] Ian Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples. In *ICLR*, 2015. 2, 8
- [20] Harleen Hanspal and Alessio Lomuscio. Efficient verification of neural networks against LVM-based specifications. In *CVPR*, pages 3894–3903. IEEE, 2023. 8
- [21] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge University Press, 2003. 2, 3, 4, 5
- [22] Lei Hsiung, Yun-Yun Tsai, Pin-Yu Chen, and Tsung-Yi Ho. Towards compositional adversarial robustness: Generalizing adversarial training to composite semantic perturbations. In *CVPR*, pages 24658–24667, 2023. 8
- [23] Hanjiang Hu, Zuxin Liu, Linyi Li, Jiacheng Zhu, and Ding Zhao. Robustness certification of visual perception models via camera motion smoothing. In *CoRL*, pages 1309–1320. PMLR, 2022. 8
- [24] Hanjiang Hu, Changliu Liu, and Ding Zhao. Robustness verification for perception models against camera motion perturbations. In *ICML Workshop on Formal Verification of Machine Learning (WFVML)*, 2023. 8
- [25] Hanjiang Hu, Zuxin Liu, Linyi Li, Jiacheng Zhu, and Ding Zhao. Pixel-wise smoothing for certified robustness against camera motion perturbations. In *AISTATS*, pages 217–225. PMLR, 2024. 2, 8
- [26] Joey Huchette, Gonzalo Muñoz, Thiago Serra, and Calvin Tsay. When deep learning meets polyhedral theory: A survey. *INFORMS Journal on Computing*, 2026. 8
- [27] Enyi Jiang and Gagandeep Singh. Towards Universal Certified Robustness with Multi-Norm Training. *arXiv preprint arXiv:2410.03000*, 2024. 8
- [28] Can Kanbak, Seyed-Mohsen Moosavi-Dezfooli, and Pascal Frossard. Geometric robustness of deep networks: analysis and improvement. In *CVPR*, pages 4441–4449. IEEE, 2018. 8

- [29] Guy Katz, Clark Barrett, David L Dill, Kyle Julian, and Mykel J Kochenderfer. Reluplex: An efficient SMT solver for verifying deep neural networks. In *CAV*, pages 97–117. Springer, 2017. 8
- [30] Konstantin Kaulen, Tobias Ladner, Stanley Bak, Christopher Brix, Hai Duong, Thomas Flinkow, Taylor T Johnson, Lukas Koller, Edoardo Manino, ThanhVu H Nguyen, et al. The 6th international verification of neural networks competition (VNN-COMP 2025): Summary and results. *arXiv preprint arXiv:2512.19007*, 2025. 7
- [31] Panagiotis Kouvaros and Alessio Lomuscio. Formal verification of CNN-based perception systems. *arXiv preprint arXiv:1811.11373*, 2018. 8
- [32] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015. 2
- [33] Linyi Li, Tao Xie, and Bo Li. SoK: Certified robustness for deep neural networks. In *IEEE Symposium on Security and Privacy*, 2023. 2, 8
- [34] Changliu Liu, Tomer Arnon, Christopher Lazarus, Christopher Strong, Clark Barrett, Mykel J Kochenderfer, et al. Algorithms for verifying deep neural networks. *Foundations and Trends in Optimization*, 4(3-4):244–404, 2021. 8
- [35] Yuexin Ma, Tai Wang, Xuyang Bai, Huitong Yang, Yuenan Hou, Yaming Wang, Yu Qiao, Ruigang Yang, and Xinge Zhu. Vision-centric BEV perception: A survey. *IEEE TPAMI*, 2024. 8
- [36] Cédric Malherbe and Nicolas Vayatis. Global optimization of Lipschitz functions. In *ICML*, pages 2314–2323. PMLR, 2017. 4
- [37] Ravi Mangal, Klas Leino, Zifan Wang, Kai Hu, Weicheng Yu, Corina Pasareanu, Anupam Datta, and Matt Fredrikson. Is certifying ℓ_p robustness still worthwhile? *arXiv preprint arXiv:2310.09361*, 2023. 2, 8
- [38] Seyed-Mohsen Moosavi-Dezfooli, Alhussein Fawzi, Omar Fawzi, and Pascal Frossard. Universal adversarial perturbations. In *CVPR*. IEEE, 2017. 8
- [39] Robert J. Moss, Mykel J. Kochenderfer, Maxime Gariel, and Arthur Dubois. Bayesian safety validation for failure probability estimation of black-box systems. *AIAA Journal of Aerospace Information Systems (JAIS)*, 2024. 2, 8
- [40] Wenjie Ruan, Xiaowei Huang, and Marta Kwiatkowska. Reachability analysis of deep neural networks with provable guarantees. In *IJCAI*, 2018. 8
- [41] Mohammad Sefidgar and Rene Landry Jr. Landing system development based on inverse homography range camera fusion (IHRCF). *Sensors*, 22(5):1870, 2022. 8
- [42] Zhouxing Shi, Qirui Jin, Zico Kolter, Suman Jana, Cho-Jui Hsieh, and Huan Zhang. Neural network verification with branch-and-bound for general nonlinearities. In *TACAS*, pages 315–335, 2025. 8
- [43] Gagandeep Singh, Timon Gehr, Markus Püschel, and Martin Vechev. An abstract domain for certifying neural networks. *PACMPL*, 3(POPL):1–30, 2019. 2, 8
- [44] Aman Sinha, Hongseok Namkoong, Riccardo Volpi, and John Duchi. Certifying some distributional robustness with principled adversarial training. *arXiv preprint arXiv:1710.10571*, 2017. 8
- [45] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus. Intriguing properties of neural networks. In *ICLR*, 2014. 2, 8
- [46] Roger Y Tsai, Reimar K Lenz, et al. A new technique for fully autonomous and efficient 3D robotics hand/eye calibration. *IEEE Transactions on Robotics and Automation*, 5(3): 345–358, 1989. 8
- [47] Fu Wang, Peipei Xu, Wenjie Ruan, and Xiaowei Huang. Towards verifying the geometric robustness of large-scale neural networks. *Proceedings of the AAAI conference on AI*, 2023. 8
- [48] Ruibin Wang, Yibo Yang, and Dacheng Tao. Art-point: Improving rotation robustness of point cloud classifiers via adversarial rotation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. 8
- [49] Shiqi Wang, Huan Zhang, Kaidi Xu, Xue Lin, Suman Jana, Cho-Jui Hsieh, and J Zico Kolter. Beta-CROWN: Efficient bound propagation with per-neuron split constraints for complete and incomplete neural network verification. *NeurIPS*, 34, 2021. 8
- [50] Kimberly Wasson and Robert Voros. Deobfuscating machine learning assurance and approval. In *43rd Digital Avionics Systems Conference (DASC)*, pages 1–10. IEEE, 2024. 2
- [51] Haoze Wu, Teruhiro Tagomori, Alexander Robey, Fengjun Yang, Nikolai Matni, George Pappas, Hamed Hassani, Corina Pasareanu, and Clark Barrett. Toward certified robustness against real-world distribution shifts. In *IEEE Conference on Secure and Trustworthy Machine Learning (SaTML)*, pages 537–553. IEEE, 2023. 8
- [52] Haoze Wu, Omri Isac, Aleksandar Zeljić, Teruhiro Tagomori, Matthew Daggitt, Wen Kokke, Idan Refaeli, Guy Amir, Kyle Julian, Shahaf Bassan, et al. Marabou 2.0: a versatile formal analyzer of neural networks. In *CAV*, pages 249–264. Springer, 2024. 7, 8
- [53] Chaowei Xiao, Jun-Yan Zhu, Bo Li, Warren He, Mingyan Liu, and Dawn Song. Spatially transformed adversarial examples. *arXiv preprint arXiv:1801.02612*, 2018. 8
- [54] Kaidi Xu, Zhouxing Shi, Huan Zhang, Yihan Wang, Kai-Wei Chang, Minlie Huang, Bhavya Kailkhura, Xue Lin, and Cho-Jui Hsieh. Automatic perturbation analysis for scalable certified robustness and beyond. In *NeurIPS*, pages 1129–1141, 2020. 8
- [55] Kaidi Xu, Huan Zhang, Shiqi Wang, Yihan Wang, Suman Jana, Xue Lin, and Cho-Jui Hsieh. Fast and Complete: Enabling complete neural network verification with rapid and massively parallel incomplete verifiers. In *ICLR*, 2021. 7, 8
- [56] Rem Yang, Jacob Laurel, Sasa Misailovic, and Gagandeep Singh. Provable defense against geometric transformations. In *ICLR*, 2023. 8
- [57] Huan Zhang, Tsui-Wei Weng, Pin-Yu Chen, Cho-Jui Hsieh, and Luca Daniel. Efficient neural network robustness certification with general activation functions. pages 4939–4948, 2018. 8
- [58] Huan Zhang, Shiqi Wang, Kaidi Xu, Linyi Li, Bo Li, Suman Jana, Cho-Jui Hsieh, and J Zico Kolter. General cutting planes for bound-propagation-based neural network verification. In *NeurIPS*, pages 1656–1670, 2022. 8