We thank the reviewers for their positive and helpful feedback. We are encouraged that they find our method scalable (R1, R3), widely applicable (R3) and our handling of interpolation errors (R1, R2) and evaluation extensive (R1). In a thorough revision we will include the reviewers' suggestions. Specifically we will (i) improve the outline of the paper (R1), (ii) clarify in Sec. 4.1 the need of a certifiably robust classifier h_E (R1) and (iii) add to Sec. 5.3 the explanation to calculate the inverse of the method in general (R1, R3). Below we answer the individual questions.

R1; Can you handle high dimensional transformations (stAdv/Wasserstein AdvEx)? BASESPT and DISTSPT can be applied directly to highly parametrized transformations like ℓ^2 bound vector field transformations ($\sum_p \|v_p\|_2^2 \leq \tau$, v_p denotes the displacement of pixel p, similarly to stAdv). However, INDIVSPT can only be applied for small τ as calculating a tight inverse is challenging, because interval splitting (L145) is not feasible. A not-tight inverse can be obtained by the relaxation that $\|v_p\|_2^2 \leq \tau$ for all pixels p individually. The Wasserstein transformation does not use interpolations, thus BASESPT is sound (cf. [arXiv:1910.10783]) and DISTSPT/INDIVSPT are not needed.

R1; Can you show results without vignetting and Gaussian blur? Yes. The results for BASESPT (Sec. 6.1) were obtained without these techniques (we will clarify this).

For DISTSPT (Sec. 6.2) the error estimates without vignetting or Gaussian blur are shown in Table 1. The setup was the same as in Sec. 6.2, but for ImageNet we used 10000 instead of 700000 samples. Both, vignetting and Gaussian blur improve the error bound significantly. On CIFAR10 and ImageNet vignetting is very impactful because the corners of images are rarely black. [13] uses vignetting for the same reason. Gaussian blur helps to shrink the errors for images particularly sensitive to interpolation, i.e. a chess board.

Dataset	Paper	-V	-G	-V-G
MNIST	**0.39**	**0.39**	2.38	2.49
CIFAR10	**0.77**	4.64	2.48	21.04
ImageNet	**0.95**	70.66	9.25	75.69

Table 1: Maximum observed errors and without gaussian blur (G) and without vignetting (V).

For INDIVSPT vignetting is crucial, even for MNIST, as we can make no assumptions for parts that are rotated into the image. Thus we need to set these pixels to the full $[0, 1]$ interval (see Fig. 2 in the paper and Fig. 1 here). Without Gaussian blur, the verification rate drops from 99.6% to 11.6% on MNIST. We will include details in an appendix.

R1, R3; Can you compare to prior work more extensively?
R1; How would vignetting and Gaussian blur benefit them?
Yes, we extended [11] (Table 1 in their paper) to include vignetting. The results are shown in Table 2. We also trained a CIFAR10 model with vignetting (CIFAR10+V) for completeness. While vignetting on MNIST slightly helps (+1 image verified) on CIFAR10 it leads to a significant drop. Including Gaussian blur into [11] would require non-trivial adaption of the method. However, we implemented this for interval analysis (on which their method is build) and found no impact on results. We will extend the our discussion (L274ff., L312ff.) similar to this discussion and more directly compare with our CIFAR10 results (App. E). Other related work is either in a fundamentally different setting or subsumed by the discussed works.

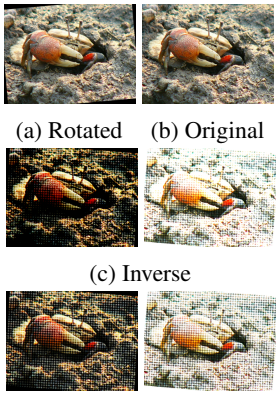
Model	Correct	[11]	[11]+V
MNIST	98	86	87
CIFAR10	74	65	32
CIFAR10+V	78	63	23

Table 2: Correct classifications and by the model and verifications by DeepG [11], with and without vignetting (V), out of 100 images.

R3; Does your scalability originate from randomized smoothing? Yes. Methods relying on convex relaxation (e.g., [11]) for neural network verification suffer from accumulation of overapproximation and the slow runtime. While we still use interval analysis for DISTSPT and INDIVSPT to bound the interpolation error, we circumvent both problems by verifying with randomized smoothing.

R3; Can your method be applied to combinations of transformations? Yes, Theorem 3.2 can be applied to *composable transformations*, that is $\psi_\beta \circ \psi_\gamma = \psi_{\beta+\gamma}$ (L92-93). In Sec. 4.1 we consider the case where this holds approximately. Unfortunately, as rotations R and translations T do not commute, $\psi_\beta := R_{\beta_1} \circ T_{\beta_2,\beta_3}$ is not composable, i.e. $\psi_\beta \circ \psi_\gamma \neq \psi_{\beta+\gamma}$ (L309-311).

R4; Can you show Figure 2 for a context rich RGB image? Yes, see Fig. 1 for an example from ImageNet. As outlined in L284-285 there are images with very large error (> 50). This error stems from the regions where the inverse algorithm can't determine strong constraints on the pixel value, visible as the circular pattern. Since the submission we have investigated improvements for such images and found partial success by replacing the circular vignette with an adaptive mask based on the local quality of the inverse. We will supplement the paper with example images and further discussion.



(a) Rotated (b) Original (c) Inverse (d) $10\times$ refined inverse

Figure 1: Image with high error. In (c) & (d): Lower and upper bound.