1 Supplementary response

Q3: First, why is core output obtained through Eq.4? Since the mean of the noise distribution is 0, then at smaller noise scales Eq.4 is theoretically approximately equal to by Taylor expansion.

R3: Thanks for your query and we further elucidate the concept of 'core output'. As introduced in Section 4.1, we define robustness certification as the guarantee that the output consistently resides within a subspace centered around the core output. To enhance certification, we endeavor to locate the gathering place of smoothing results of the basic solver f which allows us to confine the robustness certification to a smaller subspace. Regarding your mention of the approximately equal result, it is imperative to consider smoothing noise within the context of robustness certification; to disregard it would imply that the solver is nearly impervious to attack, thereby rendering our robustness certification moot.

Q4: More importantly, it is less understandable why the challenge was solved(line 382- line 384) since both z_1 and z_2 have their own independent covariance Σ_1 and Σ_2.

R4: Thanks for your question and acknowledge that the aspect was not thoroughly delineated within our paper, potentially leading to misunderstandings. Hence, we provide a comprehensive clarification herein. C_1 introduces that perturbations for visual GM are paired and mutually constrained within the certified space. The discourse between lines 382-384 offers an initial design aimed at addressing C_1, whereas the solution is presented in Eq.9. It is within Eq.9 that the impact of paired perturbations on the robustness guarantee is demonstrated, signifying an advancement over the existing robustness guarantees for individual perturbations.

Q5: Second, the design of Σ appears strange to me. For two different nodes u and v, the Σ can be either 0, or σb depending on their order among all nodes. Is there any reason to design in this way? Because Σ constructed by the method in Sec 4.2 imposes certain constraints on noise compared to independent Gaussian noise. Then the improvement in certification results may be due to the decrease in noise capability.

R5: Thanks for your question and will now provide a more detailed introduction to the joint smoothing distribution applied in this paper.

Firstly, we note in C2 that employing independent Gaussian distributions can easily disrupt the graph structure, which is derived from the Delaunay triangulation of keypoint positions. Therefore, we aim to utilize a joint smoothing distribution to maintain the graph structure as much as possible during the smoothing process. One might wonder why we do not directly explore the relationships between keypoints but instead fix the design of Σ.

This is because we need to apply the same smoothing distribution to all data; otherwise, we risk data-specific certification, leading to non-soundness in the certification. Designing Σ in such a way provides a means to preserve the graph structure for the whole dataset. This initial design does indeed possess the randomness you mentioned—that is, it may not necessarily maintain the graph structure. However, Σ is realized through an optimization process, and its design inherently includes the special case of $b = 0$. Thus, when Σ cannot adequately reveal the structure, the value of b will decrease.

Secondly, upon your insightful reminder, we have discerned that preprocessing the input data engenders enhanced certification outcomes. The preprocessing procedure is delineated as follows: commencing with an arbitrary keypoint, one sequentially selects the nearest keypoint among the remaining ones as the subsequent point, and this process is iterated until the arrangement of the keypoint matrix is consummated. After this, a smoothing operation is applied utilizing $N(0, \Sigma)$. Such a maneuver ensures that Σ is efficacious not solely through optimization but is inherently revelatory of the interconnections between keypoints. That is, when the proximity between keypoints is closer, they are more susceptible to analogous smoothing noise perturbations, thereby facilitating the preservation of their intrinsic graph structure and mitigating the performance degradation induced by smoothing.

To elucidate the efficacy of this preprocessing step, we have conducted a comparative analysis of the ACR of CR-OSRS for NGMv2 on Pascal VOC with initial $\sigma = 0.5$, examining scenarios with

Table 1: ACR of CR-OSRS for NGMv2 on Pascal VOC with initial $\sigma = 0.5$ when checking the effect of preprocessing.

Method	$\|\delta\|_{upper}$	$\|\delta\|_{volume}$	$\|\delta\|_{lower}$
preprocessing	1.448	1.574	1.261
no preprocessing	1.439	1.567	1.253

Table 2: ACR of CR-OSRS for NGMv2 on Pascal VOC with initial $\sigma = 2$ when checking value for g_0.

Value	$\|\delta\|_{upper}$	$\|\delta\|_{volume}$	$\|\delta\|_{lower}$
0.5	13.742	17.033	12.631
0.6	13.742	17.033	12.631
0.7	13.737	17.026	12.626
0.8	13.713	17.008	12.605
0.9	13.580	16.882	12.484

and without the application of preprocessing as follows. Tab. 1 shows that preprocessing can slightly improve our original certification.

Thirdly, regarding your query about whether the improvement in certification results is due to noise constraints, this is not the case. Our design only alters the structure of the smoothing distribution and thus the shape of the certified space. This results in a variation of the smoothing noise in comparison to RS. However, this distinction is separate from altering the adversarial perturbations, denoted by δ, that the model encounters. Consequently, our method does not diminish the model's resistance to disturbances; rather, it attenuates the influence of such disturbances on the visual GM process.

Q6: In equation 7, where did the 1/2 come from? I assume it means the confidence of g is over 1/2. What happens if the value is adjusted?

R6: Thanks for your inquiry, and we will now expound upon the probabilistic design employed herein.

The fraction $\frac{1}{2}$ signifies that when the majority of smoothing outcomes—exceeding $\frac{1}{2}$—are within a subspace, we designate $g = 1$ (indicating that the matching result is within the subspace). Despite the feasibility of incrementing the numerical value as follows to impose more rigorous standards for the determination of g, such an operation would decline provable robustness.

We change the value of [0.5,0.6,0.7,0.8,0.9], and calculate the ACR of CR-OSRS for NGMv2 on Pascal VOC with initial $\sigma = 2$ as in Tab. 2. With the increase of the value, the robustness guarantee will be reduced to some extent.

Q7: The σ and b can be optimized by Eq.10. So why is the result based on the fixed σ? Could you show the optimization result of σ and b?

R7: Thanks for your question.

The rationale for uniformly designating σ as the heading is due to its role as the original parameter from which optimized results are derived in our method. For a given initial parameter σ, Randomized Smoothing (RS) directly employs it to construct an independent Gaussian smoothing distribution. In contrast, our approach utilizes the optimized Σ to establish a joint Gaussian smoothing distribution. Hence, a comparison between the two is indeed valid.

In our experiment conducted on the ngmv2 solver with an initial setting of $\sigma = 0.5$, we obtain the optimized parameters of $\sigma = 0.702$ and $b = 0.016$. We now compute the ACR for RS-GM using $\sigma = 0.702$ as the parameter and compare it with CR-OSRS with an initial setting of $\sigma = 0.5$, which is presented in Tab. 3.

Tab. 3 demonstrates that our method retains advantages over the RS approach, even when compared with the optimized parameter σ. However, we do not consider this a fair comparison. Given

Table 3: ACR of RS-GM and CR-OSRS for NGMv2 on Pascal VOC with initial $\sigma = 0.5$.

Method	$\|\|\delta\|\|_{upper}$	$\|\|\delta\|\|_{volume}$	$\|\|\delta\|\|_{lower}$
RS-GM+AUG+REG($\sigma = 0.702$)	1.363	1.363	1.699
CR-OSRS+AUG+REG($\sigma = 0.5$)	1.425	1.586	1.934

an initial σ, the RS method employs fixed parameters and does not engage in the optimization of parameters in proximity to σ that may be more conducive to certification.

2 Refenrence

[1] Ren Q, Bao Q, Wang R, et al. Appearance and structure aware robust deep visual graph matching: Attack, defense and beyond[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 15263-15272.

[2] Wang R, Yan J, Yang X. Learning combinatorial embedding networks for deep graph matching[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 3056-3065.

[3] Fey M, Lenssen J E, Morris C, et al. Deep graph matching consensus[J]. arXiv preprint arXiv:2001.09621, 2020.