We thank the reviewers for their comments. The indices of references below are the same as references in our paper.

Response to Reviewer #1

"The paper proposed a J-invariance measure as the loss term in self-supervised denoising." **This summary of our work is neither accurate nor comprehensive. Our method should NOT be understood as just adding a $\mathcal{J}$-invariance measure term.** **(a)** The first term in Eqn.(6)/(8) is also different from the previous self-supervised (SS) loss in Eqn.(3). This brings benefits in efficiency, as shown in Fig.3; **(b)** We point out the practical results do not match the $\mathcal{J}$-invariance assumption in the theory behind using Eqn.(3). This is a flaw shared by previous SS denoising methods; **(c)** Our method is derived from a new and solid theory without the $\mathcal{J}$-invariance assumption, leading to improved performance.

"The J-invariance is useful for overfitting ... Table I that relate D(f) and PSNR has little meaning." **There is a misunderstanding about the results in Table 1. It is caused by statements in Line 115-117, which we will revise to make them accurate.** In fact, we are discussing the training process. $\mathcal{D}(f)$ is supposed to check whether f has the $\mathcal{J}$-invariance, which is an intrinsic property of f. Although $\mathcal{D}(f)$ in Table 1 is computed on testing data, the same results ($\mathcal{D}(f) >> 0$) can be observed for when computed on training data, for any f during training. **Therefore, results in Table 1 indicate that the model f does not have the $\mathcal{J}$-invariance, thus violating the assumption behind using the loss in Eqn.(3) in training. Besides, we fix all model configurations and training settings except for the masking strategy to make Table 1 reasonable.**

"Line 25 - Hard to agree N2N is supervised." We follow prior studies [1,12] to categorize N2N as a supervised method.

"Line 34,35: ... as they can still add AWGN to the noisy images to generate a noiser ones." **It is not true.** [16,26] only work on additive and known noise models, so that the noise from the same distribution can be simulated. Since adding AWGN does not work with unknown noise models or noise types other than Gaussian, [16,26] may not be applicable. Furthermore, N2N requires the noise in the pairs of images to be independent and identically distributed. Adding AWGN to the noisy images changes the original noise distribution and dissatisfies the independence required by N2N.

"Please explain how to determine the value of the weight of the J-invariance loss term." In most cases, we follow Eqn.(6) to set the weight to its default value 2. However, when observing extremely imbalanced L_{rec} and L_{inv} during training, we adjust the weight to balance them, as described in Appendix E.

Response to Reviewer #3

"Except for the psnr, I do not see a big visual difference relative to existing self-supervised trained methods." The visual difference does exist and is sharp especially for the ImageNet dataset. We recommend zooming-in for a better view. We'll consider adding some zoomed-in views to the visualization to make it more clear.

"I recommend the authors provide some comparisons about visually perceptual metrics e.g. NIQE, BRISQUE." We follow prior denoising studies to use PSNR, in order to make consistent comparisons. NIQE and BRISQUE may not be suitable since they are not for evaluating the denoising performance and half of our datasets are not natural images.

"In addition, I think the authors need to provide some comparisons in real noisy dataset ..." The Planaria dataset is a real noisy dataset, on which our method still outperforms the baselines.

"Why is the result of BM3D for BSD68 bolded in Table 3" It was bolded by mistake. We will unbold it. In Table 3, the bolded values correspond to the best results among self-supervised deep learning methods.

Response to Reviewer #4

"The idea of the paper is interesting but seems to be over-claimed ... This is somehow misleading." We agree that the "sub-optimal" statement is inappropriate and will revise accordingly. Nevertheless, we provide convincing analytical and experimental results to show why and how Noise2Same outperforms the baselines.

"In addition, the analysis in Sec. 4.2 did not show why the proposed method is better than the baselines ..." **Sec 4.2 theoretically analyzes the invariance term. The analytical result suggests that the invariance term has the similar effect as the post-processing in previous methods, as discussed in Line 219-224. This explains why our method achieves better performance, especially in the case where post-processing is not applicable.** In addition, as we point out in Sec 3, the $\mathcal{J}$-invariance assumption is violated in practice, making the theory behind using the loss in Eqn.(3) not applicable. This potentially limits the performance of baselines. On the contrary, our method is derived from a new and valid theory. In this case, better performance is expected.

"The results are on synthetic noise. It will be great to apply the proposed method on real noisy image Benchmarking." The Planaria dataset is a real noisy dataset, on which our method still outperforms the baselines.