(R1, R3) Why it makes sense to deblur the extracted features? The derivations of the feature-based Wiener deconvolution strictly hold in a linear feature space. As stated in L136–142, we are interested to examine how to leverage powerful learned feature extractors $\{\mathbf{F}_i\}$ beyond hand-crafted ones. To this end, we develop a feature extraction network (FEN) to estimate $\{\mathbf{F}_i\mathbf{y}\}$. As the FEN with ReLUs is piece-wise linear [17, 25], the linearity assumption of the Wiener filter holds locally. Violations can be successfully compensated by the feature refinement. To evaluate the feasibility and effectiveness of this piece-wise linear FEN, we have analyzed the effect of using features extracted by the FEN and hand-crafted features on the classical Wiener deconvolution in Tab. 8 of the supplementary. The results show that the method with the features extracted by the FEN performs much better than those with hand-crafted features, increasing the PSNR by at least 0.72dB on the dataset of [19] and 1.11dB on the dataset of [22]. We will clarify this.

(R1, R3) What if the blur is non-uniform, or the linear model assumption fails? Although our method is based on a uniform blur model, it can be extended to handle non-uniform blur by applying the local uniform approximation method [53].[1] As pointed out in [52], the gamma correction and non-linear CRF can be corrected before the deconvolution steps. We will discuss these in detail and add corresponding results in the revised paper.

(R1, R3) Parameters of the feature refinement. Our feature refinement network shares the parameters across all scales except for the first encoder block, as the input channel number for scale l=1 and l>1 differs (Eq. 9). For the other blocks (except the first), the number of features are the same and thus the parameters can be shared.

(R1, R4) Marginal improvement of learned features in Tab. 3? The reason why the improvement in Tab. 3 is not so prominent is because *Ours* and $Ours_{I+G}$ are incorporated with our multi-scale feature refinement, which is good at recovering clear images from deconvolved features and can compensate some errors caused by hand-crafted features. However, we have shown that using learned features is much more effective for applying the Wiener deconvolution than using hand-crafted features, cf. 34.78dB for $Wiener_D$ *vs.* 33.37dB for $Wiener_{I+G}$ on [19] in Tab. 8 of the suppl.

(R1) Will the method work at higher noise levels? Our PSNR results are 25.57 (15%), 24.01 (20%), 21.79 (30%) on the dataset of [19], which are 0.93dB, 1.33dB, and 0.82dB higher than the best results among the competing methods.

(R1) Why PSNRs in Tabs. 3 and 5 are so different? Tabs. 3 and 5 are evaluated on [19] and [16], respectively, with 1% Gaussian noise, where the kernel size ranges from 13×13 to 27×27 and 51×51 to 101×101. We will clarify.

(R2) Contribution and motivation. We develop a simple and effective deep learning-based non-blind image deblurring method by integrating the domain knowledge of image deconvolution. The contributions are summarized in L50–61. The motivation to use Wiener deconvolution is that we can use it to derive an effective feature-based deconvolution operator with SNR estimation, which can be effectively embedded into an end-to-end network for better deblurring.

(R3, R4) Some recent approaches are not retrained. As we use the same training dataset as [49], we do not retrain this model. The training codes of [8, 14, 38, 48] are not available, hence we do not retrain their models. However, we have finetuned their adjustable parameters for different datasets and noise levels for best results. We will clarify this.

(R4) Role of feature extractor. The feature extractor is not a denoiser but is used to provide more useful information for the Wiener deconvolution (L131–142). We further compare the learned features for blurry images w/o and w/ noise $(1\%, 5\%, 10\%)$ and find that the higher the noise level of the blurry image is, the more noise the learned feature contains. We will discuss in detail and visualize the learned features for various noise levels in the revised paper.

(R4) Why feature-based Wiener deconvolution is better than image-based method. First, the feature-based Wiener deconvolution can utilize more useful information beyond the image intensity to better constrain the deconvolution process (L110–114). Second, finer-scale detail is better modeled in the feature space. Especially, the deep feature extractor in our end-to-end network can adaptively learn useful features for the final image restoration (Figs. 3 and 6).

(R4) Robustness to noise & inaccurate kernels vs. competing methods. First, benefitting from the proposed feature-based Wiener deconvolution, our method can adaptively estimate the SNR from blurry features. Second, our end-to-end network facilitates the feature extractor learning useful features for deconvolution with fewer artifacts and benefits the feature refinement in handling deconvolved features to reconstruct clearer images. The robustness is also evaluated on real-world images in Figs. 1, 5, 17–19 (suppl.), where the noise is unknown and estimated kernels are inaccurate.

(R4) Effect of SNR estimation. The robustness and effectiveness of the SNR estimation have been demonstrated in Tabs. 1–2, 9 (suppl.) and Figs. 4–5, 11–19 (suppl.). We further carry out a sensitivity analysis w.r.t. the SNR estimation using [19] by adding 0–20% perturbation to our estimated SNR (no retraining). The PSNRs differ no more than 0.06dB, suggesting the robustness of our method to the SNR estimate. In addition, as we do not model spatially-variant noise, our method may not be so robust to non-stationary noise. This is worth further study.

(R4) "Ours w/o Wiener" is not clear. This is the method that removes the Wiener deconvolution in our model, thus is not guided by the blur kernel. We use this baseline to illustrate the importance of the Wiener deconvolution on the end-to-end network for non-blind image deblurring, which can effectively incorporate the kernel information.

(R4) Improvement of the proposed method in Fig. 3. As both the methods in Fig. 3 use the same multi-scale feature refinement, the improvement is due to the feature-based Wiener deconvolution module.

We thank the reviewers for constructive and detailed comments. We will revise the paper according to the above responses and add all other suggested references & experimental results and correct the unclear statements as suggested.

[1] [53] O. Whyte, J. Sivic, and A. Zisserman. Deblurring shaken and partially saturated images. IJCV, 110(2):185-201, 2014.