R1Q1: Relationship with HRNET: Our work is inspired by the first work which has been cited in the paper, but we missed the second work which will be reviewed and credited in the revision. Noticed that, HRNET is a handcrafted network, which does not abstract the three modules (i.e., building a search space) for NAS like this paper does.

R1Q2: A universal architecture for multiple tasks: As suggested, we conduct the experiments of using the architecture searched for denoising to train for deraining, the PSNR is 26.51 and SSIM is 0.8519 on Rain800.

R1Q3, R4Q1, R2Q1: Specific characteristics for restoration: The multi-resolution (MR) has shown effectiveness in image restoration, and all of them are based on the handcrafted design. As it is unclear (also crucial) to know when/where fuse or parallelly extract the resolutions, CLEARER is proposed. It has a MR searching space and MR feature fusion that benefits to image restoration. See L39–44 for more details. Considering the effectiveness of multi-resolution in high-level tasks, we have extended this idea and achieved some encouraging results. In fact, it is exciting to handle various tasks using a unified framework like Wang's work with the development of deep learning.

R1Q4, R1Q5, R4Q2: Additional experiments on new data: Due to time limitation, we only obtained the result on RAIN1400 (12600 training images + 1400 testing images). The PSNR and SSIM of our CLEARER are 30.42 and 0.9056. The architecture searched on it has 4 of 12 modules different from the architecture searched on RAIN800.

R2Q2: Exchange the parallel module and the transition module: The sequence of the parallel module and the transition module could be exchanged, but correspondingly, the number of parallel subnetworks will automatically change to deal with the input.

R2Q3: Iteration optimization vs. performance: Ideally, the architecture parameters are optimized for the converged network weights. Due to the expensive inner optimization, an iterative optimization was proposed as an approximation whose performance and efficiency has been proved by a series of works like DARTS. Noticed that, the major contribution of this work is designing a specific search space (multi-resolution) for image restoration first time.

R2Q4: Scales interactions: Yes. On one hand, the proposed fusion module could process cross-scale/-resolution fusion. On the other hand, many residual connections are existing in the parallel modules. Both the two folds provide the interactions between scales.

R3Q1, R4: Insufficient contributions: Most of the existing NAS are designed for the classification task, to the best of our knowledge, there is NO specific NAS for image restoration so far. Here, the specific, we refer to multi-resolution interaction which has shown effectiveness in a variety of handcrafted models, e.g., image denoising [29], deraining [37], dehazing [17]. Noticed that, although HiNAS (CVPR'20) could be one of the first NAS for denoising, it does not consider the specific characters of image denoising like our idea. Experiments show the image restoration tasks could remarkably benefit from such our specific design. Is it not valuable and novel to specifically design a NAS for the tasks besides the classification task?

R3Q2: Do the authors refer to "multi-scale" as the Parallel module? No, it refers to the multi-resolution features flowing through the network which consists of the parallel, fusion, and transition module. Only all of these together could refer to multi-scale.

R3Q3+Correctness: The loss $\mathcal{L}_{Arch}$ in Eqn.(4) is contrary/error to the description. Thanks! We missed a minus in the right-hand side. It should be $\mathcal{L}_{Arch} = -\frac{1}{N}(\sum_{a\in\{\alpha^p,\alpha^f\}} \alpha \log \alpha + (1-\alpha)\log(1-\alpha))$.

R4Q1: The reason for not searching the inner structure and the global connections: The reason are two folds. First, prohibitively temporal and spatial costs for searching the whole architecture. In fact, we have attempted to simultaneously search the interior structure, global connection, and modules, but the model is too large even for the 32GB v100 GPU. This problem cannot be solved by adding more GPUs since the model cannot be loaded into different GPUs. Second, this work is not devoted to search for a good interior structure or the global connections, but automatize the integration and design of multi-resolution neural architectures, especially when to fuse or parallelly extract the multi-resolution features. The reason is that multi-resolution has shown effectiveness in a lot of image restoration tasks, but it is crucial and unknown when and where the multi-resolution are fused and higher-resolution is expected.

R4Q2: Model complexity vs. inference accuracy of other methods: On one hand, our networks are more complex than other methods due to the multi-resolution architecture which consists of multiple subnetworks. On the other hand, other methods do not explicitly consider this character. In fact, we believe that NAS is promising in real applications of image restoration because the trade-off between model complexity and performance could be precisely controlled.

R4Q3+Correctness: As a bilevel optimization method, the relationship between three losses and objective functions is unclear. We would clarify that our contribution is NOT developing a bilevel optimizer, which instead designing a specific searching space for image restoration. In our loss, $\mathcal{L}_{Res}$ is the restoration loss which is used both in the optimization of architecture parameters and network weights. $\mathcal{L}_{Arch}$ and $\mathcal{L}_{Comp}$ are the architecture regularizations, which only used to optimize the architecture parameters.