To Reviewer # 2:

We appreciate the valuable insights from the reviewer. In the following pages are our point-by-point responses to each of the comments.

To Comment # 1:

Concern: ***For most physical or mechanical systems, the 2nd order stationarity does not hold. The statistics such as mean and correlation are usually highly dependent on the mechanism and external environments.***

Response: Yes, we understand the importance of nonlinearity, non-Gaussian, and non-Stationarity. In the revised manuscript, we have extended the current case study to nonlinear region. Second, we have tested different uncertainty models. Third, we have added some discussion on applying the current method to complex spatiotemporal systems in Section 5.

To Comment # 2:

Concern: ***The authors propose to use a plain CNN model to represent the physical dependency between different quantities with spatial variation. However, deep net models for similar work typically contain multiple different layers rather than convolutional layers only (e.g., Cha et al. 2017 as cited in this manuscript). There is no evidence that every layer of the model should be computed using convolution. I would recommend using a mixed deep net model, which have been validated in many similar applications.***

Response: Yes, thank you for bringing this to our attention. The proposed surrogate is not a plain CNN model. We admit that the concepts, terminologies, and definitions adopted in the manuscript are mainly from computer science. And we know this may result in misunderstanding. To address this issue, we have rewritten the methodology section to clarify the novelty of the architecture design. Moreover, we also prepared a figure illustrating the network design of the proposed model.

As it is mentioned, Cha's work contains different layers, i.e. the Convolution layer, Pooling layer, Activation layer, Auxiliary layers, and Softmax layer. In our work, we also have different model ingredients but with several differences. First, Convolution is kept. Second, we intentionally removed the max-pooling after extensive network architecture search. This is due to the fact that pixel-wise regression requires information conveying local variability compared to the standard image classification problem. Third, we use ReLU as the activation function (Same as Cha). Fourth, we also implemented dropout and Batch normalization (Auxiliary layers). Fifth, the softmax function is not considered in our work because it is specifically designed for classification problems. Instead, we use image resizing techniques to address the pixel-wise prediction problem. On the other hand, the proposed network architecture is designed in a downsampling-upsampling manner and Cha's work takes a downsampling form. This is due to the formulation of the problem. For crack image detection, the state-of-the-art techniques are centering on pixel-wise segmentation to identify and measure diverse cracks concurrently at the pixel level [1, 2].

1. Yang, X., Li, H., Yu, Y., Luo, X., Huang, T. and Yang, X., 2018. Automatic pixel-level crack detection and measurement using fully convolutional network. Computer-Aided Civil and

Infrastructure Engineering, 33(12), pp.1090-1109.

2. Li, S., Zhao, X. and Zhou, G., 2019. Automatic pixel-level multiple damage detection of concrete structure using fully convolutional network. Computer-Aided Civil and Infrastructure Engineering, 34(7), pp.616-634.

To Comment # 3:

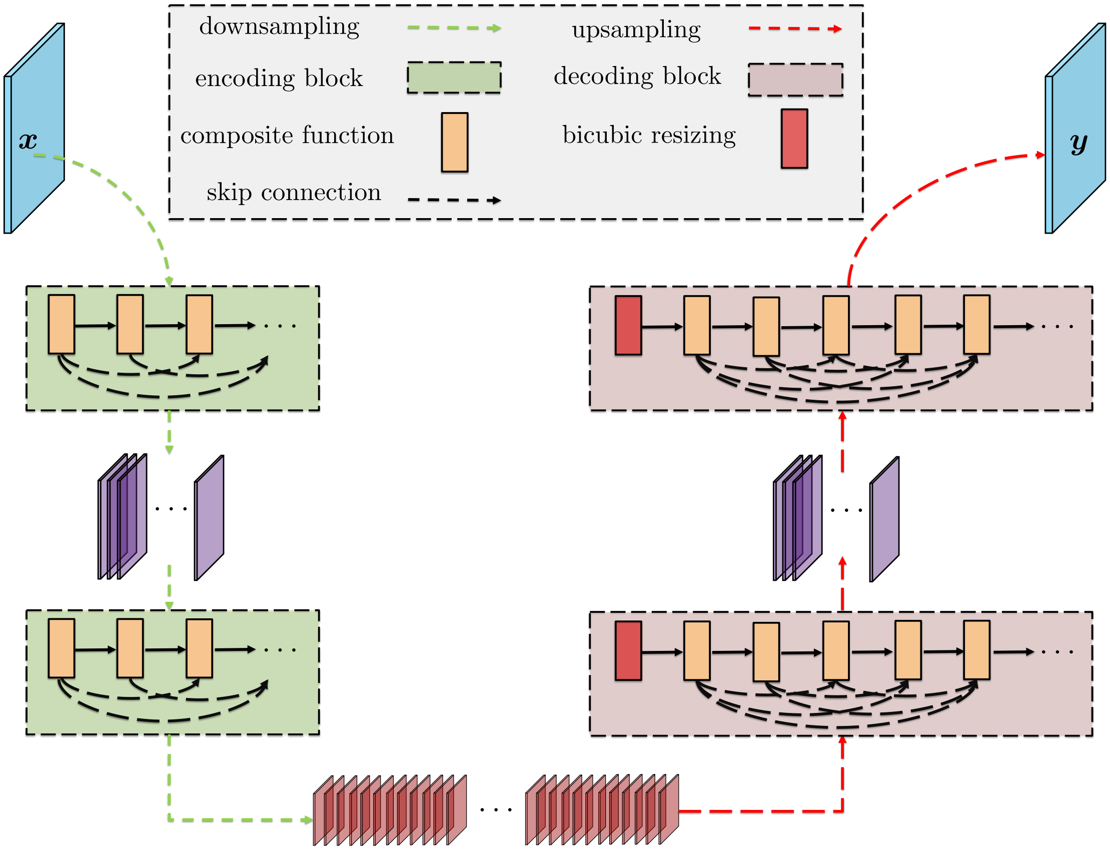
Concern: ***It makes sense that the intermediate layers should be standardized to avoid gradient explosion. But why the first layer is not standardized? If the scales of different input variables are significantly different (which is possible for a physical system), gradient explosion would be a problem.***

Response: Yes, thank you so much for catching this part. In fact, we have standardized the input and output data. This process is referred to as the data preprocessing in our manuscript.

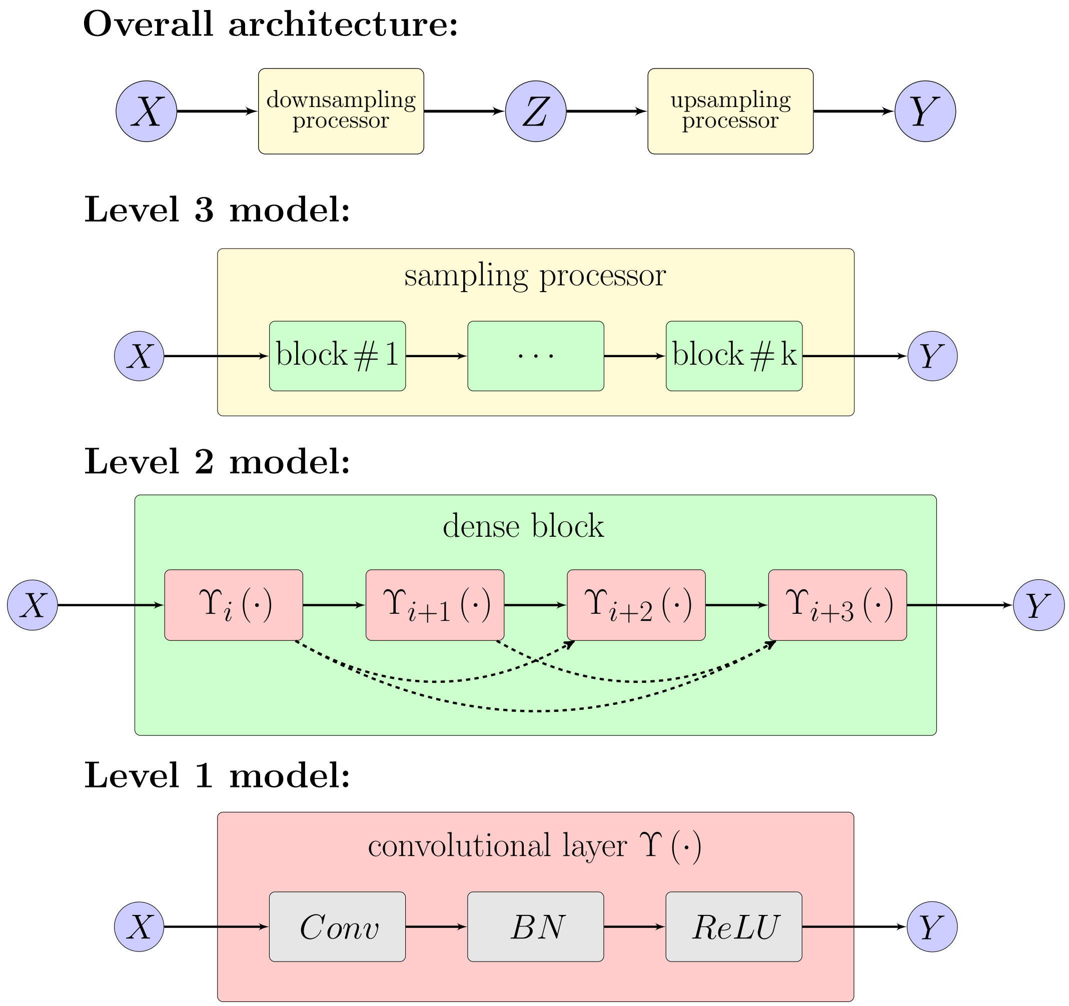
To Comment # 4:

Concern: ***Fig.2: It is hard to tell the relationship between different blocks in this figure. Also, why is Level 1 put behind all the other levels?***

Response: Yes, we have prepared a new figure illustrating the model architecture.



We used an ascend pipeline to describe the model. In particular, level 1 refers to the local functional operation and level 3 refers to the global information flow pattern.



To Comment # 5:

Concern: ***I understand that gradient vanishing could be a problem in general, and using connection layers can help solve this issue. However, there is no reason of connecting all the layers in a CNN model.***

Response: Yes, we completely agree with the comment made by the reviewer. In fact, we are not connecting all the layers, rather, we are only connecting layers coming from the same block. This block design is a state-of-the-art technique we adopted from [1].

1. Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K.Q., 2017. Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708).

To Comment # 6:

Concern: ***Similar issues appear in Section 3.1.3 and 3.2.2. There is no reason that we must use non-unit stride, L2 norm, and mini-batch gradient descent to solve a spatial***

problem. How to choose the model form, layer form, and optimization algorithm depends on the specific problems.

Response: Yes, we agree. In fact, we have carried out an extensive study on architecture design and hyperparameter identification. To address these issues, we have added a new subsection regrading the model design.

To Comment # 7:

Concern: ***Section 3.3: One important question the authors raised up is how to represent the uncertainties. However, neither RMSE nor R-square is a good indicator of the model uncertainty. The authors may consider some solid model performance evaluation criteria, such as DIC, Log-score, et al.***

Response: Yes, we agree. In fact, RMSE and R-square are used as performance evaluation criteria regarding model learning instead of the model uncertainty in this paper.

To Comment # 8:

Concern: ***Again, I am not convinced that using 21 plain convolutional layers is the best option. A combination of different types of layers (and possibly with fewer number of parameters) is very likely to provide a better representation of the system behavior.***

Response: Yes, we understand the concern. First, the model is not 21 plain convolutional layers. Second, we have carried out other studies where FR-21 is directly applied. The predictions results show the effectiveness of the proposed FR-21 in terms of dealing with different problems.

To Comment # 9:

Concern: ***In addition, Fig. 6 in the subsection 'uncertainty analysis results' should be Fig. 7.***

Response: Yes, thank you for this excellent observation. We have revised the manuscript accordingly.

To Comment # 10:

Concern: ***The authors mentioned multiple times of uncertainty quantification. The results only show the variability of the final outputs obtained from CNN. Only by looking at the variability of the inputs or outputs, I cannot see the value of using CNN but not a much simpler model.***

Response: In the uncertainty quantification results plot, we have summarized the mean and probability density functions of some points randomly selected from the domain in addition to the variance. The main advantage of the proposed surrogate over other technique such as Gaussian process or generalized polynomial chaos expansions is its excellent scalability in terms of high-dimensional problems.