To Reviewer # 4:

We appreciate the positive feedback from the reviewer. We found the comments extremely insightful.

To Comment # 1:

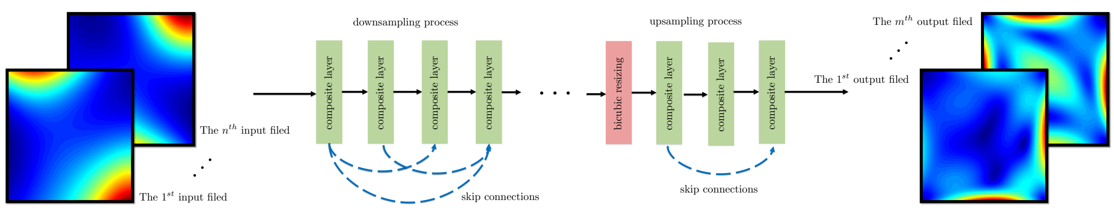
Concern: ***Please shorten Section 3.1 as the content has been widely known.***

Response: Yes, thank you for bringing this to our attention. We agree with the reviewer that some contents in Section 3.1 can be removed. To address this, we rewrote the whole section. The state-of-the-art techniques, i.e. the composite function, DenseNet design and resizing upsampling adopted in this paper are explicitly stated.

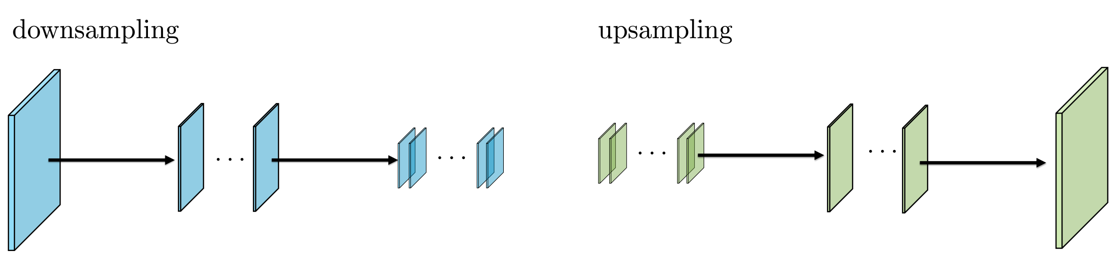
To Comment # 2:

Concern: ***Please provide more details about the hyperparameter identification and architecture design. How sensitively does the hyperparameters of CNNs, e.g. the depth, impact the experimental results?***

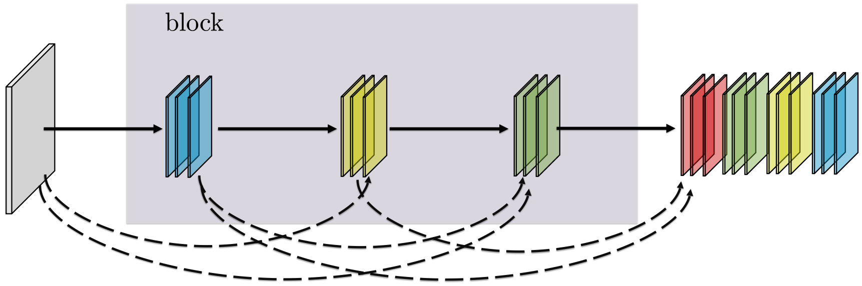
Response: Yes, thank you for raising this point. To address this concern, we have added a subsection to the manuscript, discussing details about the hyperparameter identification and architecture design. In addition, following pages are our step-by-step model construction experience that we would like to share with the reviewer. Motivated by the benchmark SegNet [1], the overall architecture of the initial network design looks like:



To reduce the number of model parameters, data propagates in a downsampling-upsampling manner within the proposed architecture:



This is achieved by the adoption of skip connections [2,3]. Figuratively, each text arrow shown above indicates a block that looks like:



Again, each arrow indicates a composite function that is defined in Section 3.1 of the revised manuscript. On the other hand, the main design considerations for our field-to-field regression problem broadly fall into the following two categories: (1) model hyperparameters and (2) learning hyperparameters.

Model hyperparameters	**Design considerations**
network	number of layers
block	number of blocks, skip connections or not
activation function	ReLU, TanH, SoftPlus
downsampling process	pooling, no pooling
upsampling process	deconvolution, image resizing
convolution kernel	kernel size, stride number, zero padding
Learning hyperparameters	**Design considerations**
optimizer	SGD, Adam, Adagrad
regularizations parameters	L1 or L2 penalty, weight decay, dropout ratio, early stopping, batch size
learning	epoch number, initial learning rate
annealing strategy	annealing frequency, annealing rate

Table 1: Summary of the hyperparameters tested in this paper

Grid, as well as random searching, have been adopted for obtaining optimized hyperparameters.

For the last part of this comment, we assume the type of hyperparameters mentioned refers to the model hyperparameters. Following the Occam's Razor rule, we gradually increase the complexity of our model. Once a prescribed accuracy has been achieved, adding more layers has trivial impacts on the experimental results. It should be noted that the spatial dimension of the coarsest level has bigger impacts than the depth in terms of training efficiency and prediction accuracy. This is due to the data explanation capacity of the latent feature space. More theoretical and experimental details are available at [4].

Reference

1. Badrinarayanan, V., Kendall, A. and Cipolla, R., 2017. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence, 39(12), pp.2481-2495.

2. He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

3. 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).

4. He, K., Zhang, X., Ren, S. and Sun, J., 2016, October. Identity mappings in deep residual networks. In European conference on computer vision (pp. 630-645). Springer, Cham.

To Comment # 3:

Concern: ***Please have a discussion on how to fit the fixed input/output size of CNNs in various input/output. This could be the major limitation of the method.***

Response: Yes, you have raised an important point here. Using CNNs to obtain solutions to PDEs on general/arbitrary domains is challenging. We would like to address this issue by combining data argumentation and hybrid learning. For instance, the domain of interest is irregular. First, we would like to integrate this irregular domain to a larger regular domain. Then, we would like to train the surrogate model using regression loss first and subsequently use classification loss to reinforce the learned results. The first phase can use the proposed technique and the second phase can use image segmentation techniques. A graphic illustration is given below: