Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2019-128-AC2, 2019

Geoscientific Model Development Discussions
Open Access
EGU

Interactive comment on "DINCAE 1.0: a convolutional neural network with error estimates to reconstruct sea surface temperature satellite observations" *by* Alexander Barth et al.

Alexander Barth et al.

a.barth@uliege.be

Received and published: 17 December 2019

We thank the review again for this quick response and very insightful comments.

1. "This is indeed interesting to see"

Thank you for your encouragement.

2. The channels of input data increased from 8 to 10. But the filter size, the number

Printer-friendly version

Discussion paper

of output feature maps, and layer size, number of layers stay the same. Thus,the parameters of the network should stay the same. Right?

In addition to the SST (divided by the expected error variance), we are also providing the inverse of the expected error variance. The number of input channels changed from 4+3*2 = 10 to 4+5*2 = 14. The filter size of the first convolution network stayed at 16 filters. While previously the 3x3 convolution was realized with 3 x 3 x 10 x 16 (width x height x input channels x output channels), in the version with more time instances the convolution matrix had the dimensions 3 x 3 x 14 x 16 (so a 40In the first submitted version of the manuscript we indeed wrote that the total size of the array is 8 x 112 x 112 x 5266. This should be 10 x 112 x 112 x 5266 and this is corrected in the revised version. We apologize for this confusion.

3. This is helpful. Previously, the author introduced two variables with no explaination of what and why. Previously, from the formula only, it seems like the value of these two variables will affect strongly the computation. e.g. delta = 100 vs.delta = 0.01

We agree that this part was unclear in the first submission and thank the reviewer for highlighting that the parameters were not properly discussed.

4. I am not very sure I fully understand it. But I will leave it to other reviewers!

Maybe it is clearer with an (admittedly) extreme example: if there is some part of the domain where there are no training data and this domain is dynamically completely disconnected from the rest, then its value is (per construction) completely unconstrained, except for an a priori information of reasonable values. So e.g. 10°C is probably as good a guess as 14°C. The neural network tends to oscillate between these two values because there are no constraints from the

	last reconstruction	average reconstruction
missing data change every epoch	0.4968	0.3834
missing data change every 10 epochs	0.4423	0.4146
missing data change every 20 epochs	0.4387	0.3984

Table 1. CV error for different experiments keeping the mask of the missing data constant for several epochs

data. Assume further that we have a validation data point of 13°C in this area, then the average RMS would be (abs(10-13) + abs(14-13))/2 = 2 °C but the RMS of the average is abs(12-13) = 1 °C. (for a single value the RMS is directly related to the absolute value which we use here to simplify the notation, but the same results are true for a series of numbers).

5. I would guess the fundamental reason why the RMSE and Loss fluctuate so much is that the random mark missing data in every mini batch. Because in every epoch, the spatial correlation of missing and available data is disrupted due to random marking, hence what the network has learned in previous epoch is disrupted as well, which eventually is reflected in RMSE and Loss. The fluctuations may not have so much to do with mini-batch optimization. Perhaps one way to check is to use same random mark missing data for every 20 epochs, and average at every 20 epochs. Just my opinion

We conducted this experiment and the reviewer is right that it had indeed a quite significant/dramatic effect on the convergence of the cost function (see the attached figure). Unfortunately, the average reconstructed SST or the reconstruction from the last epoch are not better than the best experiments that we had already in the manuscript.

The experiment "missing data change every epoch" is the experiment "DINCAE

(all skip connections and average pooling)" from the manuscript. Despite that the results are not better in this case, the reviewer's idea is promising and we include it as an option in our code. Application to other cases will tell if this proposed option (keeping the same data marked as missing for every 20 mini-batch) should rather be preferred.

Another way to interpret the marking of some data as missing would be to view it as a drop-out layer as the value of zero does indeed represent an "infinitely" large error. Change the mask of missing data at every epoch seems to help the generalization.

We also verified that the current version of our code reproduces exactly the same results as the code when the article was submitted if the same random seeds are used to exclude the possibility that any other change to the code has an impact here.

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-128, 2019.

Fig. 1. Cost function when changing the data marked as missing change every 10 or 20 epochs