Review

The manuscript addressed an interesting topic, which was the use of a model already trained with data from a given region in another region (transfer learning) in the subject of precipitation nowcasting. This is especially interesting in the case of few radar data for a given region and low computational resources. **In the text, I suggest emphasizing this as a motivation and/or other factor that has motivated the present work and also pointing out the advantages and disadvantages of this method, with regard to the complex problem of precipitation nowcasting.**

However, the text lacked sufficient description of the data used (RADOLAN) compared to the data used by Shi et al. (2017): Does it need to be the same type of data (e.g. CAPPI 2km, reflectivity or precipitation)? What about time frequency and spatial resolution? HKO-7 used only cases with rain. It seemed to me that your work used the entire sequence, comprising rain and no rain events, without interruptions, right? **You should include a description of how you approached these issues and faced difficulties, as it adds value to the work and also helps other researchers to reproduce the results.**

Some analyses of the results were lacking to prove the gain or not of the method: use of more metrics; evaluation of the model's performance with the forecast time; influence of hyperparameters obtained from a model trained in another database; etc. I give some suggestions below, along the listed comments.

The text is well written, easy to read, even for a non-native reader. But there is a need for an English review due to minor errors in spelling, verb tenses, punctuation (e.g. lines 17, 27, 33, 49, 55, 56, 185, 194,202).

In the following, **I highlight some points that should be given more attention, and somehow should be included in the text**. They are separated into “content comments” and “text comments”. :

Content comments:

1) Sec. 2.1 (data). This section lacks some important information:
    a. Did you do any pre-processing on the images or is 480x480 pixels the original size of the image?
    b. Did you use reflectivity or precipitation? In case of precipitation, what was the Z-R relation used?
    c. Is it a CAPPI or PPI? Please, inform the height or sweep elevation;

d. The used period from 2017 to 2021 comprises only three years: from Apr/2017 to Mar/2021. Make it clear to the reader;
e. Did you use the complete sequence or only selected rain events, as Shi et al. (2017)?
f. Explain the selection of training, validation and testing sets: summer to test, winter to validate;
g. Instead of "*German RADOLAN*", use "*RADOLAN*".
h. Please, if possible, give more information: What is the weather radar type: band, polarization, Doppler? Where can the reader find this type of information?

2) Line 120: "*Other finetuning configurations were tested*" The authors should comment more on this;
3) Table 1:
    a. The first threshold includes rain rate = 0. What do you consider as no rain?
    b. The RADOLAN column sum more than 100%;
    c. How was this table calculated, with the complete sequence of the dataset or with selected rainy cases? Is it the distribution of pixel values in the image set?
    d. Shi et al. (2017) used selected rainfall events. Is the HKO-7 column considering only these events? (You do not need to repeat Shi et al.'s paper, but you should provide enough information for your reader to understand what you are talking about.)
4) Line 138: The authors introduced binary values (0, 1) based on thresholds. They must inform the meaning of the values above and below the thresholds;
5) Line 145: What "*measurements*" do you refer?
6) Line 145: Briefly describe Welch's t-test in sec. 3;
7) Sec. 4.1: The model predicts 20 images, from 5 to 100 min lead time. For which forecast times are the shown results?
8) Fig. 2 is equal as Fig. 3, the same pattern, but with different values. The authors should verify that it is correct. If correct, what is the gain of using such metrics, what does this prove? Why not use another metric to explore more information?
9) Section 4.1 (lines 157-163):
    a. The values are too small to draw conclusions. The authors forced a conclusion mainly with the expression "big increase" (lines 163, 225);
    b. What about the analysis of the result evolution with forecast time?

c. I suggest including other metrics, such as FAR and POD, and some metric to assess the image quality, since you are using a computer vision method;

10) Lines 165-167: "*We compare the model output for the real truth in the frame of two case studies. Using the ~~the~~ RADOLAN data set, we consider frontal systems at 4 May 2017, 14:40:00 UTC and at 12 May 2017, 07:40:00 UTC. These are two exemplary dates of clusters of mainly moderate precipitation crossing Germany, where the data is not part of the training data set of the model*." The authors should comment on this first of all in sec. 2.3, Experimental setup;

11) Sec. 4.2:
    a. You compare your results with Ravuri et al. (2021) and Ayzel et al. (2020). How many examples did these studies use to compute their statistics? Because yours considers just one case;
    b. Line 199: "*the positive effect of finetuning is clearly visible for higher precipitation intensities*." You are referring to a small gain of just one score;
    c. **What is the merit of the model used in your predictions?** The other models have different architectures; you should take this into account in your analysis;
    d. Why the case studies weren't done for both "*finetuned*" and "*scratch*"? How will you evaluate the gain of one against the other?
    e. Do not miss your goal as this is the scientific question you must answer. You need to structure your analyses so you do not mix up the results;

12) Figs. 2-5, 8: $R = 0.5$ or $R > 0.5$? As in Tab. 2. (The same for the other thresholds);

13) Lines 209-210: "*However, a statistical analysis would be wishful to confirm an improving CSI over the prediction time for higher thresholds as shown by the case study in Fig. 8 (a)*." Why haven't you done it yet? You already have the model outputs. **This must be included in the results**;

14) Lines 210-211: "*It can be recognized that the HSS, shown in Fig. 8 (b) and (d), provides higher scores than the CSI depicted in Fig. 8 (a) and (c)*." Fig. 8, as Figs. 2 and 3, shows the same pattern for CSI and HSS, with different values. **You should take a careful look at your results in case you missed something**;

15) Fig. 6:
    a. Why did you put reflectivity images on the 2nd row? What is the point you want to show?
    b. Is it prediction of rain or reflectivity? (See comment 11d.)
    c. In the color legend, what does it mean when the rain field is gray? The color legend is incomplete;

d. Which forecast times are included in Fig. 6? You should comment this in the caption and in the text;
e. Where are the "*scratch*" images?
f. What does negative reflectivity mean? Why didn't you filter the raw data?

16) Fig. 7:
    a. Again, what is the predicted variable, rain or reflectivity?
    b. What is the range of your data so I can understand the differences in the images?
17) Line 220: Here you say "*similar*", but before you said "*slightly better*" (line 155);
18) Lines 226-228: "*Comparing the here obtained results with recent publications on deep learning algorithms to precipitation nowcasting based on radar data (Ayzel et al., 2020; Ravuri et al., 2021) the finetuned TrajGRU shows slightly higher scores with less decrease with prediction time.*" See comments 10 and 11;
19) Lines 228-230: "*While Chen et al. (2020); Ayzel et al. (2020) show that their models are suitable to predict precipitation up to 60 minutes (Chen et al., 2020; Ayzel et al., 2020), we achieve comparable scores for a 100 minute prediction time*." I couldn't find this analysis throughout the text;
20) Lines 230-231: "*A statistical analysis of the prediction time for more than the here presented two case studies would be wishful for future research*." You have not presented any analysis regarding forecast time (line 209). **What you showed here is still not enough for a publication.** You already have the data and the outputs of the models, you need to explore further analysis;
21) Lines 232-233: "*To further optimize the results in future studies, different finetuning methods and finetuning hyperparameters could be taken into account.*" You should add some examples in the text;
22) Line 236: "*finally*" This is an ambitious comment regarding the precipitation nowcasting problem itself. When completed described, your presented solution can help in some ways, but in my opinion, based on my research and experience, I think one product is not enough to solve the complex problem of precipitation nowcasting. I suggest changing to "*contribute positively*" to be more realistic.

Text comments:

23) In the abstract (line 6): "*The positive effectiveness of finetuning a network pretrained at a different location and for different precipitation intensity thresholds is*

demonstrated." I suggest changing it to "*many*", since you used the same thresholds as the TrajGRU.

24) Line 12: please properly cite the reference WMO in the text, it must contain a name and a year;
25) Line 30: Why "*loosely*"? From Ayzel et al. (2020), that is what they said about RainNet: "*Its design was inspired by the U-Net and SegNet families of deep learning models*".
26) In the Introduction, 3rd paragraph (lines 32-43), the text is not clear. I suggest informing at the beginning of the paragraph that the model used was TrajGRU (Shi et al., 2017) and then developing the other concepts;
27) Still in the Introduction, 4th paragraph (lines 44-54), the text is also not clear. There you present the objective of the present work. I suggest highlighting the objective, reviewing this paragraph and making any necessary changes to make it clear to the reader;
28) Lines 42,44: instead of "*standardized benchmark dataset*" and "*standardized dataset*", I suggest using "*benchmark dataset*";
29) Line 45: revise "*ideally one wants to use*";
30) Line 52: Replace "In this work" with "In the present work", to make it clear that it refers to the present work and not to another;
31) In the Introduction, 5th paragraph (lines 55-63), the text is confusing as you mixed sec. 2 and sec. 3. I suggest revising;
    a. Also, you cite specific results in this paragraph "*The results show that finetuning not only reaches the same performance of the network trained from scratch, and this significantly faster, but it is also able to improve it, especially for heavy rainfall events.*" (lines 60-61), but that is not the place to comment on results, you should delete it;
32) Page 4: The description of the equations is insufficient in relation to Shi et al. (2017), some terms are missing, such as "*H*", "*σ*" and "*warp*". You should summarize and cite appropriately;
33) Line 101: "*keeping the weights*", but the weights were used only in the initialization;
34) Line 146: Instead of "*different skill scores thresholds*", write "*different thresholds*";
35) Table 3: Fig.2 already shows this point. The values are the same in both CSI and HSS columns. I suggest removing this table;
36) Lines 149 and 154: use the acronyms CSI and HSS;
37) Figs. 2 e 3: Instead of "*from scratch*", write "*scratch*", as in Tab. 2;
38) Line 165: What is "*real truth*"? Write "*ground truth*" instead;

39) Lines 171-172: "*While the ~~the~~ precipitation around the outer boundaries of precipitating cells is slightly underestimated by the TrajGRU, the rainfall in regions of intense precipitation is overestimated*." From where you got these analyses, Fig 6 or Fig. 7? You should reorder Figs and the comments about Figs in the text;

40) Lines 188, 192, 194, 204: Instead of "*finetuned TrajGRU*", use only "*finetuned*".

41) Sec. 5: I suggest removing the sentence "*The TrajGRU is developed by Shi et al. (2017), who improved their own Convolutional Long Short-Term Memory (LSTM) network for radar data (Shi et al., 2015) by the inclusion of error functions. The authors demonstrate its applicability to a standardized data set of the ChineseWeather Service*." (lines 216-218) and rewriting the beginning of the paragraph from "*The deep learning algorithm TrajGRU (Trajectory Gated Recurrent Unit) for precipitation nowcasting (…)*" (line 213) as "*The deep learning algorithm Trajectory Gated Recurrent Unit (TrajGRU; Shi et al., 2017) for precipitation nowcasting (…)*";

42) Line 220: "*fully newly trained network*" means "*scratch*". It is important to keep the nomenclature, or at least put it in parentheses so that the reader who has come this far can connect what he/she read before with the conclusions.