NTIRE 2023 Image Super-Resolution (x4) Challenge Factsheet -Global Swin Transformer for Image Super-Resolution-

- Zhao Zhang, Baiang Li, Huan Zheng, Suiyi Zhao, Yangcheng Gao, Yanyan Wei and Jiahuan Ren -

March 22, 2023

This factsheet template is meant to structure the description of the contributions made by each participating team in the NTIRE2023 challenge on Image Suoer.

Ideally, all the aspects enumerated below should be addressed. The provided information, the codes/executables and the achieved performance on the testing data are used to decide the awardees of the NTIRE2023 challenge.

Reproducibility is a must and needs to be checked for the final test results in order to qualify for the NTIRE awards.

The main winners will be decided based on overall performance and a number of awards will go to novel, interesting solutions and to solutions that stand up as the best in a particular subcategory the judging committee will decided. Please check the competition webpage and forums for more details.

The winners, the awardees and the top ranking teams will be invited to co-author the NTIRE2020 challenge report and to submit papers with their solutions to the NTIRE2020 workshop. Detailed descriptions are much appreciated.

The factsheet, source codes/executables, and final **FULL** manipulation results of "all the images" should be sent to **all of the NTIRE2020 challenge organizers (Cosmin Ancuti, Codruta O. Ancuti and Radu Timofte)** by email. The image results should be reproducible and match the **last** submission on Codalab. You still need to upload your results on Codalab! As the file size is expected to be large, we accept download links from your website, google drive, dropbox, etc. When using cloud services, please remember to allow sharing and consider the traffic limit of your service provider.

Organizer name	Email
Cosmin Ancuti	cosmin.ancuti@gmail.com
Codruta O. Ancuti	codruta.ancuti@gmail.com
Radu Timofte	radu.timofte@vision.ee.ethz.ch

Email final submission guide

To: cosmin.ancuti@gmail.com; codruta.ancuti@gmail.com; radu.timofte@vision.ee.ethz.ch
cc: your_team_members
Title: NTIRE 2020 NonHomogeneous Dehazing Challenge - TEAM_NAME

Body contents should include:
a) the challenge name
b) team name
c) team leader's name and email address
d) rest of the team members
e) team members with NTIRE2020 sponsors
f) team name and user names on NTIRE2020 CodaLab competitions
g) executable/source code attached or download links.
h) factsheet attached
i) download link to the results of all of the test frames

The executable/source code should include trained models or necessary parameters so that we could run it and reproduce results. There should be a README or descriptions that explains how to execute the executable/code. Factsheet must be a compiled pdf file together with a zip with .tex factsheet source files. Please provide a detailed explanation.

1 Team details

- Team name

  LVGroup_HFUT

- Team leader name

  Zhao Zhang

- Team leader address, phone number, and email:

  9th Floor, Block A, Science And Education Building, Hefei University of Technology (Emerald Lake Campus), 485 Danxia Road, Hefei 230601,China.

  cszzhang@gmail.com,

  (+86)18356138079

- Rest of the team members:

  Baiang Li, Huan Zheng, Suiyi Zhao, Yangcheng Gao, Yanyan Wei and Jiahuan Ren.

- Team website URL (if any)
- Affiliation

  Hefei University of Technology
- Affiliation of the team and/or team members with NTIRE2023 sponsors (check the workshop website)
- User names and entries on the NTIRE2020 Codalab competitions (development/validation and testing phases)

  MotaLee, 5(validation) and 2(test)(1 success and 1 fail)
- Best scoring entries of the team during development/validation phase development:

  best scoring entry 5, score: 30.941821 test: best scoring entry 2, score: 30.676472
- Link to the codes/executables of the solution(s): This link
- Link to the restoration results of all frames: This link

2 Contribution details

- Title of the contribution:

  Global Swin Transformer for Image Super-Resolution
- General method description

  There have been a large number of previous Transformer-based works[1,2,3] that have achieved remarkable results in the field of image restoration. However, existing representative methods pay less attention to the global features of images, and they mainly focus on the local features of images. Therefore, we tried to make a GlobalSwinIR based on the Swin IR model. Specifically, we follow the basic architecture of SwinIR[1]. While increasing the depth of SwinIR to obtain better detail recovery quality, we add another global feature branch[2] to better capture the global features of the image, thereby guiding image recovery.
- Representative image / diagram of the method(s):See Fig 1
- References

  [1] Jingyun Liang Jiezhang Cao Guolei Sun Kai Zhang 0008 Luc Van Gool Radu Timofte SwinIR: Image Restoration Using Swin Transformer. 1833-1844 2021 ICCVW.

  [2] Chenhongyi Yang Jiarui Xu Shalini De Mello Elliot J. Crowley Xiaolong Wang 0004 GPViT: A High Resolution Non-Hierarchical Vision Transformer with Group Propagation. 2022 abs/2212.06795

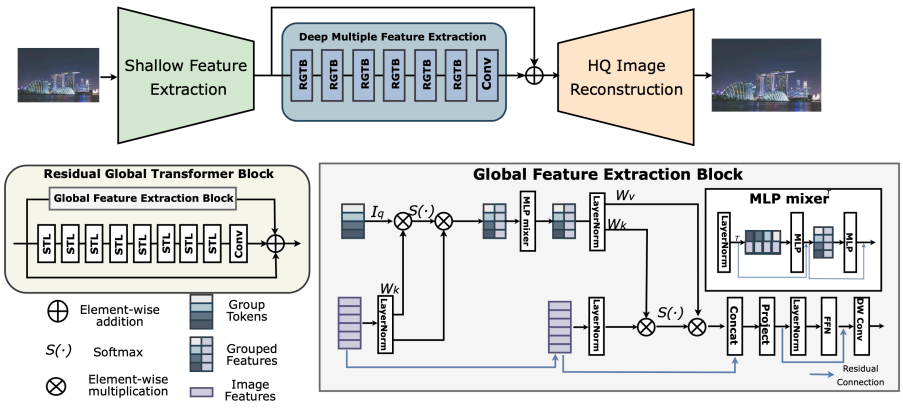

Figure 1: The structure of our proposed GlobalSwinIR. The Swin Transformer Layer(STL) is the same as SwinIR[1].

[3] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, Neil Houlsby: An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. ICLR 2021

3 Global Method Description

- Total method complexity: all stages

  Train stage: (resolution 48x48, channel:3) parameters: 80810742

- Which pre-trained or external methods / models have been used (for any stage, if any)

- Which additional data has been used in addition to the provided NTIRE training and validation data (at any stage, if any)

- Training description

  We trained totally 500,000 iterations to take the model to convergence. Our patch size(random crop) is 192 for high resolution image. We choose adam as our optimizer, and our initial learning rate is set to $2*10^{-4}$, and the learning rate is halved every 100,000 iterations. Meanwhile, in the first 400,000 iterations, we use L1 loss as our loss function, and in the last 100,000 iterations we use L2 loss. Our other settings are exactly the same as SwinIR[1]. There are no other tricks.

- Testing description The proposed solution is implemented based on PyTorch vision 1.11.0 and on python3.8, Cuda11.3. we use A100-PCIE-40GB with 40G memory.

- Quantitative and qualitative advantages of the proposed solution

  Our solution is 0.30dB PSNR higher than SwinIR baseline.

- Results of the comparison to other approaches (if any)
- Results on other benchmarks (if any)
- Novelty degree of the solution and if it has been previously published:

  The proposed architecture solution mainly refers to SwinIR[1] and GPViT[2].

- It is OK if the proposed solution is based on other works (papers, reports, Internet sources (links), etc). It is ethically wrong and a misconduct if you are not properly giving credits and hide this information.

4 Competition particularities

Any particularities of the deployed solution for the competition (if applicable)

5 Ensembles and fusion strategies

- Describe in detail the use of ensembles and/or fusion strategies (if any).
- What was the benefit over the single method?

  Our methods can achieve better results than baseline.

- What were the baseline and the fused methods?

  baseline:SwinIR[1], GPViT[2]

6 Technical details

- Language and implementation details (including platform, memory, parallelization requirements)

  We train our model using A100-PCIE-40GB with 40GB memory. We don't use parallelization.

- Human effort required for implementation, training and validation?

  Just one person is enough to finish it.

- Training/testing time? Runtime at test per image.

  For training, we train this model about 5 days to take it to convergence. For testing, one image takes about 1.71s(including calculating PSNR and SSIM and other preparations for the test.)

- Comment the robustness and generality of the proposed solution(s)? Is it easy to deploy it for other sets of downscaling operators?

  This model use SwinIR[1] as our benchmark, we add one block on it and add its depth, so there's no difficulties to deploy it for other sets of downscaling operators.

- Comment the efficiency of the proposed solution(s)?

  It's larger than SwinIR[1], but it achieves **higher quality recovery**.

7 Other details

- Planned submission of a solution(s) description paper at NTIRE2020 workshop.
- General comments and impressions of the NTIRE2020 challenge.
- What do you expect from a new challenge in image restoration, enhancement and manipulation?
- Other comments: encountered difficulties, fairness of the challenge, proposed subcategories, proposed evaluation method(s), etc.