NTIRE 2024 Efficient SR Challenge Factsheet -title of the contribution-

Yubin Wei
XIAMEN UNIVERSITY
YubinWei@stu.xmu.edu.cn

Haonan Chen
XIAMEN UNIVERSITY
23220231151779@stu.xmu.edu.cn

Xiaole Yan
XIAMEN UNIVERSITY
23220231151812@stu.xmu.edu.cn

Xiaole Yan
XIAMEN UNIVERSITY
libinren@stu.xmu.edu.cn

Second Author
XIAMEN UNIVERSITY
secondauthor@i2.org

Second Author
XIAMEN UNIVERSITY
secondauthor@i2.org

1. Introduction

Transformer-based approach shows impressive performance in low-level visual tasks such as image super-resolution. However, we find that these networks can only utilize a limited spatial range of input information through attribution analysis. This means that the potential of Transformer is still underutilized in existing networks. In order to activate more input pixels for better reconstruction, we propose **a novel two-domain attention mixer** (EHAT) , which is based on a novel hybrid attention converter (HAT) developed by Xiangyu Chen et al. , based on the combination of channel-based attention and window-based self-attention scheme, **a new learning channel-based self-attention mechanism** (BSA) proposed by autonomous researchers is integrated, in addition, to better capture the global feature dependencies of spatial and channel dimensions, we introduce **a dual-attention network** (DAnet) , which can be used to identify the global feature dependencies of spatial and channel dimensions, by adaptively integrating local features and global dependencies, we further expand the scale of the model to demonstrate that the performance of this task can be greatly improved.

2. Email final submission guide

To: yawei.li@vision.ee.ethz.ch
bin.ren@unitn.com
nancy.mehta@uni-wuerzburg.de
timofte.radu@gmail.com
cc: your_team_members

Title: NTIRE 2024 Efficient SR Challenge - TEAM_NAME - TEAM_ID

To get your TEAM_ID, please register at Google Sheet. Please fill in your Team Name, Contact Person, and Contact Email in the first empty row from the top of sheet. Body contents should include:

a) team name

b) team leader's name and email address

c) rest of the team members

d) user names on NTIRE 2024 CodaLab competitions

e) Code, pretrained model, and factsheet download command, e.g. `git clone ...`, `wget ...`

f) Result download command, e.g. `wget ...`

- Please provide different urls in e) and f)

Factsheet must be a compiled pdf file together with a zip with .tex factsheet source files. Please provide a detailed explanation.

3. Code Submission

The code and trained models should be organized according to the GitHub repository. This code repository provides the basis to compare the various methods in the challenge. **Code scripts based on other repositories will not be accepted.** Specifically, you should follow the steps below.

1. Git clone the repository.

2. Put your model script under the `models` folder. Name your model script as `[Your_Team_ID]_[Your_Model_Name].py`.

3. Put your pretrained model under the `model_zoo` folder. Name your model checkpoint as `[Your_Team_ID]_[Your_Model_Name].[pth or pt or ckpt]`

4. Modify `model_path` in `test_demo.py`. Modify the imported models.

5. `python test_demo.py`

Please send us the command to download your code, e.g. `git clone [Your repository link]` When submitting the code, please remove the LR and SR images in `data` folder to save the bandwidth.

4. Factsheet Information

The EHAT network we used was based on the HAT network by fusing a learnable channel-based self-attention mechanism (BSA) to improve the model's global statistics and strong local fitting capabilities. The training data set was DF2K, and the model was modeled as A, to validate the data set as DIV2K, the model is first learned incrementally, applying the HAT's pre-training weights, and then fine-tuned in the learning rate and loss function. More details are described below.

4.1. Team details

- Team name: **SKDADDYS**
- Team leader name: **Yubin Wei**
- Team leader address, phone number, and email

  address: **XIAMEN UNIVERSITY**

  phone number: **+86 13459661560**

  email: **YubinWei@stu.xmu.edu.cn**
- Rest of the team members:

  Haonan Chen (email:23220231151779@stu.xmu.edu.cn)

  Xiaole Yan (email:23220231151812@stu.xmu.edu.cn)

  Binren Li (libinren@stu.xmu.edu.cn)

  Siqi Zhang (23220231151819@stu.xmu.edu.cn)

  Sihan Chen (23220231151780@stu.xmu.edu.cn)
- User names and entries on the NTIRE 2024 Codalab competitions (development/validation and testing phases)

  User names on NTIRE2024 CodaLab competitions:longguo

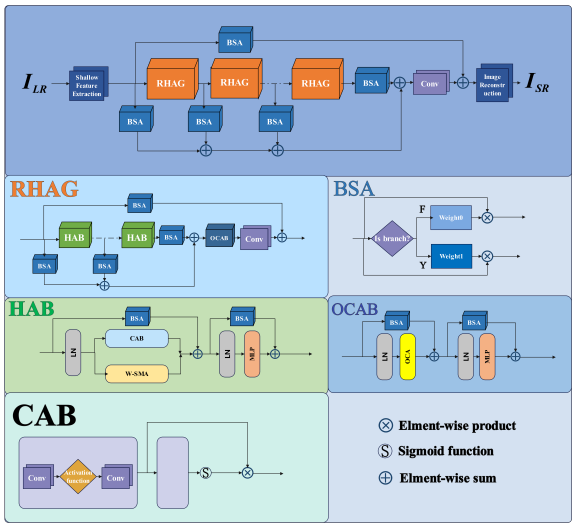

Figure 1. EHAT Neural Network Architecture Diagram

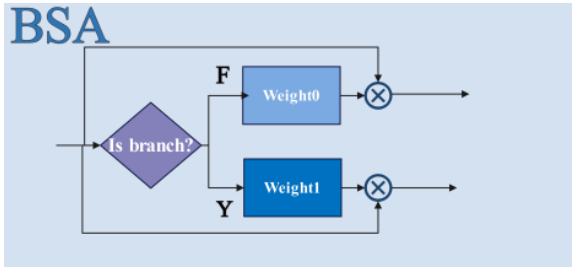

Figure 2. a new learning channel-based self-attention mechanism

- Best scoring entries of the team during development/validation phase

  31.109242
- Link to the codes/executables of the solution(s) **https://github.com/notiom/Ehat**

4.2. Method details

The EHAT(Figure 1.) network architecture integrates a new type of learnable channel-based self-attention mechanism (BSA) based on the HAT network, this module is suitable for networks with resnet and contact splicingto better capture the global feature dependencies of spatial and channel dimen- sions, we introduce a dual-attention network (DAnet) , which can be used to identify the global feature dependen- cies of spatial and channel dimensions, by adaptively inte- grating local features and global dependencies.

The BSA (a learning branch weight module) is shown in Figure.2

Our team did not perform tens of thousands of iterations of the model during the experiment due to hardware con-

Figure 3. EHAT Neural Network Architecture Diagram

straints, so we estimate that the model still has untapped potential. Here's how we did it:

• TRACK1 first, we use HAT-L on the hat website. PTH served as our pre-training model while acquiring the 10800 raw data set of DF2K, and then we performed 500 iterations of training on 1 Nvidia GeForce RTX 3090 GPU with a batch size of 2. After adding the BranchAttentionModule, the model is added several parameters, and the total parameters of the model are reached. In addition, the model is optimized by Adam, 1 = 0.9 and 2 = 0.99, and the weight is reduced to 0 by default. The initial learning rate was set to 1e-5 and initial training was performed using L1LOSS.

• TRACK2 at the end of TRACK1, we selected the weight file (referenced with PSNR values) that best performed in TRACK1 as the pre-training model for this round, and in addition, we set the learning rate to 1e-8, the remaining parameters were left unchanged for 500 iterations.

• Track3 finally, we used the pre-training model of TRACK2 to vary the learning rate sequentially from 1e-5 to 1e-8 and to fine-tune the loss function L1loss by replacing it with MSELOSS, which is then fed into the network for training; Get the final result.

The final experimental results are shown in figure.3. Compared with realesrgan, resize and hat, the results show that ehat has a significant improvement.

5. Other details

- Planned submission of a solution(s) description paper at NTIRE 2024 workshop.
- General comments and impressions of the NTIRE 2024 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.

References