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

Binren Li
XIAMEN UNIVERSITY
libinren@stu.xmu.edu.cn

Siqi Zhang
XIAMEN UNIVERSITY
23220231151819@stu.xmu.edu.cn

Sihan Chen
XIAMEN UNIVERSITY
23220231151780@stu.xmu.edu.cn)

1. Introduction

Transformer-based methods have demonstrated excellent performance in numerous vision tasks, including but not limited to image super-resolution reconstruction. However, we find that these networks can only utilize a limited spatial range of input information through attribution analysis. It means that the potential of Transformer is enormous. To overcome these limitations, we propose a novel network model **Enhanced Hybrid Attention Transformer (EHAT)** built upon the recent development of Hybrid Attention Transformer (HAT) by Xiangyu Chen et al., incorporating our independently designed innovative **Branch Self-Attention mechanism (BSA)**. The importation of BSA aims to address the issues of overlooking the importance of some pathways and having a single-weight setting for pathway in traditional approaches. Compared to simply setting pathway weights to fixed values, BSA dynamically adjusts their weights by learning mechanisms to accurately identify the importance of each pathway, so that to capture key features in the data more precisely. Ideally, BSA assigns larger weight coefficients to pathways with more significant features, thereby effectively enhancing the overall network performance. Through extensive empirical validation, we successfully demonstrate the effectiveness of BSA in improving network performance.**The source code and pre-trained models are available at https://github.com/notiom/Ehat**.

2. Method

The network architecture of EHAT **(Figure 1)** has been improved based on the HAT network to better adapt to com-

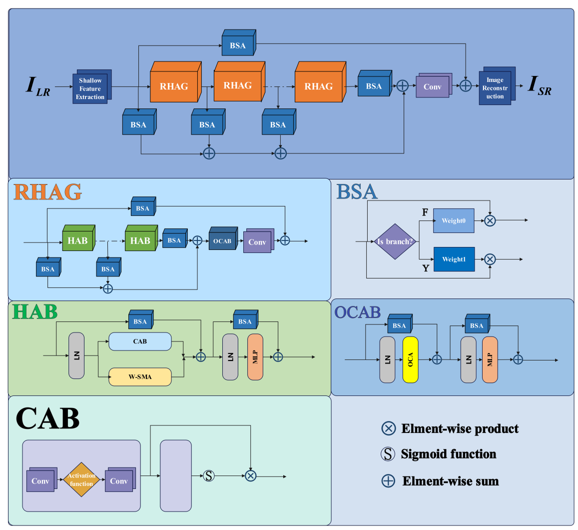

Figure 1. EHAT Neural Network Architecture Diagram

plex feature learning tasks. We introduce BSA **(Figure 2)** into the HAT network to address the issues of single-weight setting for pathway weights and overlooking the importance of some pathways in traditional methods. Through extensive empirical validation, we demonstrated the effectiveness of BSA in improving network performance. Given its significant advantages, we suggest applying this module to networks incorporating structures such as ResNet and concatenation to further enhance their performance.

Our team did not perform tens of thousands of iterations of the model during the experiment due to hardware constraints, so we estimate that the model still has untapped

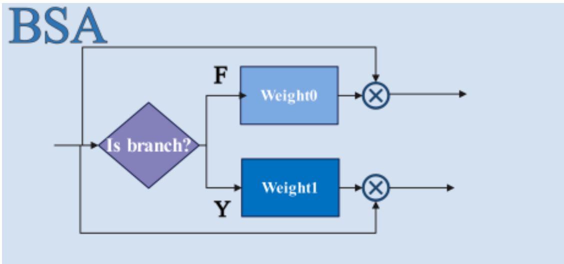

Figure 2. a new learning channel-based self-attention mechanism

potential. The details are as follows about how we did it:

- **Track1**: Firstly, we utilized the HAT-L_SRx4_ImageNet-pretrain.pth model from the official HAT website as our pre-trained model. We employed the original dataset of 800 images from DIV2K and 10,000 images from LSDIR as our training set. Subsequently, we improved the HAT model by introducing the BranchAttentionModule and fine-tuned the model with a batch size of 2. Training was conducted for 500 iterations on an NVIDIA GeForce RTX 3090 GPU. Additionally, the model was optimized using Adam with $\beta_1 = 0.9$ and $\beta_2 = 0.99$, and default weight decay of zero. The initial learning rate was set to 10^{-5} and preliminary training was performed using L1 loss.

- **Track2**: Following Track1, we selected the best weight (measured by PSNR) from Track1 as the pre-trained model for this round. Furthermore, we set the learning rate to 10^{-8} and continued training the model.

- **Track3**: Finally, utilizing the pre-trained model from Track2, we sequentially adjusted the learning rate from 10^{-5} to 10^{-8} and replaced the L1 loss function with MSELoss for fine-tuning. Subsequently, the data was sent into the network for training to get the final results.

  The final experimental results are shown in **figure 3**. The results show that EHAT has a significant improvement compared with RealESRGAN, HAT and HAT-GAN.

3. Team Information

- Team name: **SKDADDYS**

- Team leader

  name: **Yubin Wei**

  address: **XIAMEN UNIVERSITY**

  phone number: **+86 13459661560**

  email: **yubinwei@stu.xmu.edu.cn**

Figure 3. EHAT Neural Network Architecture Diagram

- Other Members:

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

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

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

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

  Sihan Chen (email:23220231151780@stu.xmu.edu.cn)

- User names on NTIRE2024 CodaLab competitions: **longguo**

- Best scoring entries of the team during development/validation phase: **31.109242**

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