Deep Learning-Based Blind Image Quality Assessment for Super-Resolution

Shawn LEE (Supervisor : Dr. Aaron Zhang)

University of BRISTOL

(1) Introduction

In the digital era, images inevitably undergo distortions during acquisition, compression, transmission, and super-resolution, which degrade visual quality and user experience. Image Quality Assessment (IQA) is essential and can be subjective, which rely on human ratings, or objective, where computational models predict quality. Since subjective evaluations are costly and time-consuming, objective IQA has become more practical. Early conventional IQA methods used hand-crafted metrics based on the Human Visual System and natural scene statistics, which are fast but often inaccurate. With advent of deep learning, IQA now leverages neural networks to extract features and predict quality scores through regression.

Super-resolution (SR) algorithms introduce new distortions at both local and global scales. Despite their significance, SR-induced distortions have received limited attention in IQA research. Current state-of-the-art IQA models typically use either a CNN to capture local features, a Transformer for global features, or a dual-branch approach to fuse both, aiming to improve prediction accuracy.

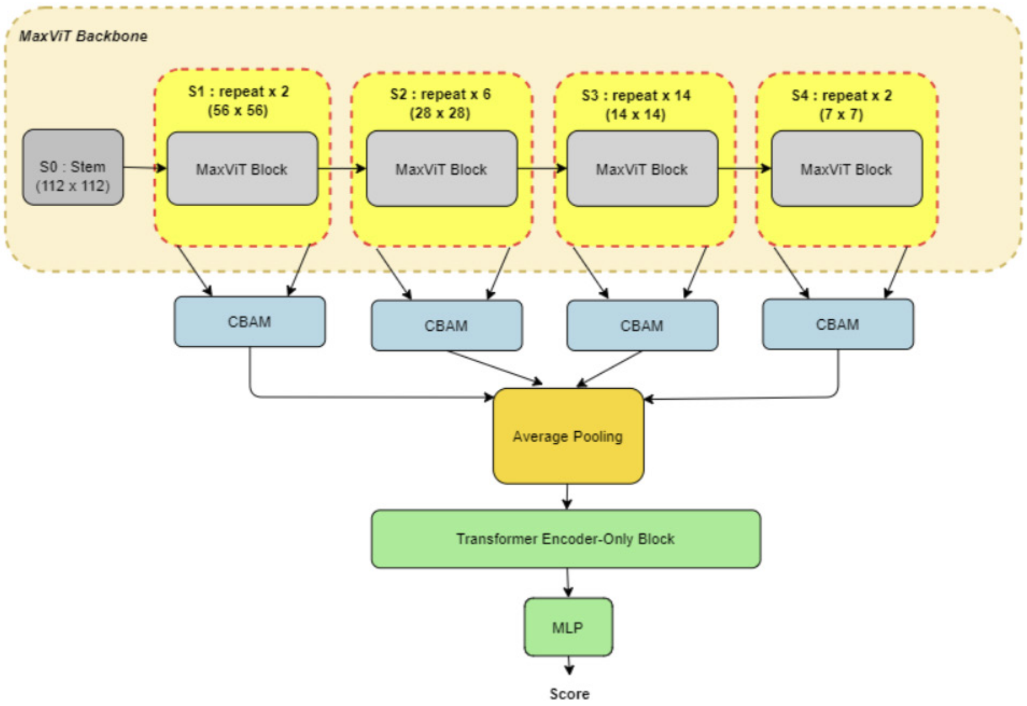
(2) Research Questions

- Performance of Existing Models on SR Datasets
  - Traditional FR-IQA performance vs. Traditional NR-IQA performance
  - CNN-based IQA vs. Transformer-based IQA
- Hybrid Backbone Efficacy
  - Can a hybrid CNN-transformer backbone provide superior performance when assessing the perceptual quality of super-resolved images?
  - Which feature maps extracted from the backbone enhance performance?
  - Does increasing crop size improve performance?
- Additional Head Modules Efficacy
  - Which modules effectively refine the features extracted from the backbone to yield better performance?

(3) Datasets

Dataset	Ref images	Image resolution	Distort images	Scaling factor	SR method types	Judgement type
PIPAL(Train)	250	288×288	23200	2,3,4	12	MOS(Elo rating system)
PIPAL(Test)	25	288×288	1000	2,3,4	12	MOS(Elo rating system)
QADS	20	500×380	980	2,3,4	21	MOS(0-5)
SR4KIQA	24	3840×2160	768	2,3,4,8	11	MOS(0-5)

(4) My Model Architecture



(5) Intra & Cross dataset Results

Table 1: Performance comparison on the PIPAL dataset

Method	PIPAL PLCC	PIPAL SROCC
TReS	0.2776	0.269
MANIQA	0.6187	0.6408
MAXIQA (Ours)	0.6305	0.6333

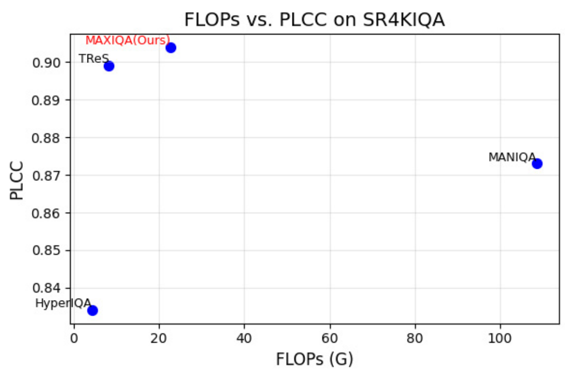
Table 2: Performance comparison on the SR4KIQA dataset

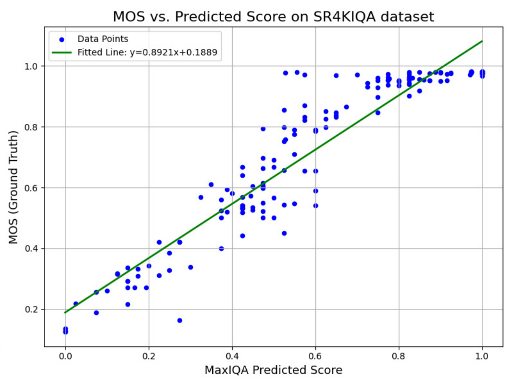
Method	SR4KIQA PLCC	SR4KIQA SROCC
PSNR	0.674	0.567
SSIM	0.641	0.665
MS-SSIM	0.584	0.802
BRISQUE	0.169	0.231
NIQE	0.512	0.527
PIQE	0.468	0.525
NRQM(Ma)	0.412	0.389
PI	0.564	0.601
HyperIQA	0.834	0.839
TReS	0.899	0.912
MANIQA	0.873	0.882
MAXIQA (Ours)	0.904	0.911

Table 3: Performance comparison when pre-trained on PIPAL and inference on SR4KIQA and QADS cross-dataset.

Pre-Trained on	Method	SR4KIQA PLCC	SR4KIQA SROCC	QADS PLCC	QADS SROCC
PIPAL	MANIQA	0.7507	0.7602	0.8456	0.8377
	MAXIQA (Ours)	0.7528	0.7474	0.8627	0.8682

(6) Analysis





(7) Challenges

- **Resource Constraints & Overfitting:** Limited GPU memory restricts batch sizes, and a small dataset leads to overfitting despite data augmentation.
- **HPC Policy Restrictions:** The Bluepebble HPC policy hinders fine-tuning, complicating model optimization.
- **Model Complexity vs. Optimization Trade-off:** Each model has different numbers of parameters and FLOPS, making it challenging to balance computational cost and optimization efficiency.
- **Novel Backbone:** No prior IQA studies with a MaxViT backbone required extra effort to determine optimal feature maps.

(8) Conclusion

Our proposed MAXIQA outperforms existing methods not only on intra-dataset evaluation but also generalizes well across different SR datasets.
By leveraging both CNN and transformer backbones and optimizing feature fusion strategies, our model achieves superior prediction accuracy and robustness to diverse SR distortions.