JNDMIX: JND-BASED DATA AUGMENTATION FOR NO-REFERENCE IMAGE QUALITY ASSESSMENT

Jiamu Sheng¹, Jiayuan Fan^{*1}, Peng Ye², Jianjian Cao²

¹Academy for Engineering and Technology, Fudan University, Shanghai, China ²School of Information Science and Technology, Fudan University, Shanghai, China

ABSTRACT

Despite substantial progress in no-reference image quality assessment (NR-IOA), previous training models often suffer from over-fitting due to the limited scale of used datasets, resulting in model performance bottlenecks. To tackle this challenge, we explore the potential of leveraging data augmentation to improve data efficiency and enhance model robustness. However, most existing data augmentation methods incur a serious issue, namely that it alters the image quality and leads to training images mismatching with their original labels. Additionally, although only a few data augmentation methods are available for NR-IQA task, their ability to enrich dataset diversity is still insufficient. To address these issues, we propose a effective and general data augmentation based on just noticeable difference (JND) noise mixing for NR-IQA task, named JNDMix. In detail, we randomly inject the JND noise, imperceptible to the human visual system (HVS), into the training image without any adjustment to its label. Extensive experiments demonstrate that JNDMix significantly improves the performance and data efficiency of various state-of-the-art NR-IQA models and the commonly used baseline models, as well as the generalization ability. More importantly, JNDMix facilitates MANIQA to achieve the state-of-the-art performance on LIVEC and KonIQ-10k.

Index Terms— No-reference image quality assessment, Data augmentation, Just noticeable difference

1. INTRODUCTION

No-reference image quality assessment (NR-IQA) is a significant task to automatically predict image quality scores without using reference image information. With the development of deep learning techniques, a large number of learning-based NR-IQA methods achieve good performance. Using pre-trained models based on a large-scale dataset like ImageNet [1] can help NR-IQA models to extract relevant features for distorted images with the limited scale of used



Fig. 1: An original image is processed by five different data augmentation methods. Visually, the other four data augmentation methods cause highly noticeable image quality degradation, except JNDMix.

datasets. Therefore, many recent works [2, 3, 4, 5, 6] utilize ImageNet pre-trained backbone to achieve the state-of-the-art performance on the NR-IQA benchmarks.

A key challenge of learning-based NR-IQA methods is the lack of sufficiently large labeled datasets, as manually annotating image quality labels is expensive and timeconsuming. As a result, the learning-based NR-IQA methods usually suffer from over-fitting when training deep models [7]. To some extent, using pre-trained models can alleviate this issue, but this metric is not specifically designed for NR-IQA task, and the model performance bottleneck still exists. Therefore, it is necessary to design a novel approach to improve the data efficiency of the NR-IQA model.

Data augmentation is a practical and simple way to improve generalizability and data efficiency of models, which can effectively alleviate the above problems on NR-IQA task. So far, data augmentation methods are widely used in various vision tasks, that include geometric transformations (e.g., random cropping [8], horizontal flipping [9], cutout [10]), color space transformations (e.g., color jittering [11]), mixing images (e.g., mixup [12], CutMix [13]), resizing [14], noise injection [15], etc. Unfortunately, most of data augmentation methods mentioned above, are mainly developed for specific vision tasks and not available for the NR-IQA task, since the highly noticeable image quality degradation is caused [16].

^{*}Corresponding author.

This work was supported in part by the National Natural Science Foundation of China under Grant 62101137 and Grant 62071127, and in part by the Zhejiang Lab Project under Grant 2021KH0AB05.

As shown in Fig. 1, noise injection with Gaussian noise directly introduces distortions to the image; resizing, cutout and random cropping can impact image composition, thus altering the image quality. Although a few augmentation methods such as horizontal flipping and random cropping are widely used in NR-IQA task [17], they are not specifically designed for NR-IQA task and are still limited in their effectiveness in enriching dataset diversity. More seriously, random cropping leads to the loss of image information and introduces geometric deformation, resulting in generated patches mismatching with their labels, while existing models using random cropping ignore the above problem and still use the original labels to train. Hence, a general and effective data augmentation method is urgently needed for the NR-IQA task.

Intuitively, it is reasonable that the noise injection augmentation without noticeable image quality degradation is suitable for NR-IQA task. As the characteristic of the human visual system (HVS), just noticeable difference (JND) is defined as the minimum visual content changes that human can perceive [18]. Namely, for each pixel, JND is a threshold, and the HVS will not perceive any changes under it. Hence, we believe the change under JND as a kind of noise, named JND noise, which can be used in noise injection augmentation. In NR-IQA datasets, the labels are the quality of images subjectively scored by the human. Due to the imperceptibility of JND noise, the image injecting JND noise will have the same subjective quality score as the original image.

In this paper, we are the first to propose a general data augmentation method based on JND noise mixing (JNDMix) specifically for NR-IQA task. In detail, we randomly inject JND noise to the training image, and it will not alter the image quality visually so that the generated image can remain the original label. The seminal proposal of JNDMix tackles the long-standing problem with existing data augmentation methods available for NR-IQA task, in particular, further enriching the diversity of training images. Hence, JNDMix can help NR-IQA models effectively to improve model performance while enhancing data efficiency over previous stateof-the-art methods. We conduct extensive experiments on two IQA benchmarks, LIVEC and KonIQ-10k. And the results demonstrate that JNDMix bring the significant improvements in performance and data efficiency of various state-of-the-art NR-IQA models and baseline models. More importantly, our proposed JNDMix helps MANIQA to achieve the state-ofthe-art performance on LIVEC and KonIQ-10k. The crossdataset evaluation results also verify the great generalization ability of our proposed method for NR-IQA models. Besides, our approach is general and can be flexibly added to any NR-IQA method.

2. METHOD

In this section, we will present the details of our proposed general data augmentation method, depicted in Fig. 2.



Fig. 2: Overview of JNDMix data augmentation workflow.

2.1. Random JND Noise Injection Procedure

Let $x \in \mathbb{R}^{W \times H \times C}$ and y denote an original training image and its label, respectively. The goal of JNDMix is to generate a new training image \tilde{x} and its associated label \tilde{y} . The original training image x and the associated generated JND image x_{ind} are formed into an image pair to jointly generate a new training image. Firstly, each pixel of the JND image x_{ind} is multiplied by the random ratio λ to generate a random JND noise x_{ind} where each pixel is less than the JND threshold. Then we inject our random JND noise x_{ind} into the associated training image. For each pixel, the injection noise can either add the noise to the original pixel value or subtract the noise, so we have to multiply a random matrix $r \in \mathbb{R}^{W \times H \times C}$ that is random on each pixel, with only 1 and -1. Finally, the novel training image, named JND distorted image, is generated after the above process. This random JND noise injection procedure can be formulated as follows:

$$x_{\text{noise}} = \lambda x_{\text{ind}} \tag{1}$$

$$\tilde{x} = x \oplus (r \cdot x_{\text{noise}}) \tag{2}$$

where \oplus is element-wise sum. (·) is element-wise product. Random ratio λ is sampled from the uniform distribution (0, 1).

2.2. Model Training

In the training phase, the generated JND distorted image \tilde{x} replaces the original training image x in the model, but we do not need to adjust the associated quality score label y. The image quality scores in the dataset are obtained by subjective human scoring, while JND noise can not be perceived by humans theoretically, so the subjective scores of JND distorted image will not change. Thus, the following equation is obtained:

$$\tilde{y} = y \tag{3}$$

Then, the new training sample (\tilde{x}, \tilde{y}) generated by JNDMix is used to train the model with its original training settings and loss function. In each training iteration, each JNDMix-ed sample (\tilde{x}, \tilde{y}) is generated by multiplying the JND image by a resampled random ratio to enrich the training sample's diversity, and diverse training samples reduce over-fitting. Our method is very general and concise, so it can be flexibly added to any NR-IQA method.

Method	LIV	/EC	KonIQ-10k		
Wiethou	SRCC	PLCC	SRCC	PLCC	
DBCNN (100%)	0.851	0.869	0.906	0.923	
w/ JNDMix	0.854	0.876	0.915	0.930	
	+0.003	+0.007	+0.009	+0.007	
HyperIQA (100%)	0.859	0.882	0.906	0.917	
w/ JNDMix	0.870	0.883	0.911	0.924	
	+0.011	+0.001	+0.005	+0.007	
MANIQA (100%)	0.886	0.908	0.923	0.941	
w/ JNDMix	0.897	0.918	0.927	0.943	
	+0.011	+0.010	+0.004	+0.002	

 Table 1: Model performance improvements with JNDMix across different state-of-the-art models.

 Table 2: Model performance improvements with JNDMix across different baseline models.

Method	LIV	/EC	KonIQ-10k		
Wiethou	SRCC	PLCC	SRCC	PLCC	
Resnet34 (100%)	0.824	0.849	0.905	0.920	
w/ JNDMix	0.842	0.868	0.910	0.928	
	+0.018	+0.019	+0.005	+0.008	
Resnet50 (100%)	0.838	0.868	0.916	0.931	
w/ JNDMix	0.851	0.872	0.920	0.932	
	+0.013	+0.004	+0.004	+0.001	
ViT (100%)	0.776	0.816	0.901	0.915	
w/ JNDMix	0.815	0.858	0.906	0.923	
	+0.039	+0.042	+0.005	+0.008	

3. EXPERIMENTS

3.1. Experimental Settings

Datasets and Evaluation Metrics. We assess different NR-IQA models and JNDMix on two commonly used IQA datasets: LIVEC [19] and KonIQ-10k [20]. LIVEC contains 1162 images taken in the real world by various photos using various camera devices, resulting in complex and composite distortions. KonIQ-10k consists of 10073 images chosen from the large public multimedia dataset YFCC100m[21].

Two commonly used evaluation metrics Spearman's rank order correlation coefficient (SRCC) and Pearson's linear correlation coefficient (PLCC) are adopted to evaluate the performance of NR-IQA models.

These datasets are randomly split into 80% images for training, and 20% images for testing. In data-efficient experiments, we always use the 20% images for testing regardless of different fractions of the training dataset. We conduct 10 times of this random dataset splitting operation for each experiment and the average SRCC and PLCC values are reported to evaluate performance.

Models and Implementation Details. To evaluate our proposed data augmentation method, we utilize JNDMix on these three representative state-of-the-art NR-IQA models: DBCNN [2], HyperIQA [3], MANIQA [4]. Also, three baseline models are adopted to evaluate JNDMix effectiveness: Resnet34, Resnet50 [22] and Vision Transformer (ViT) [23], which are pretrained on ImageNet [1]. The JND images are generated by RGB-JND [24] model to generate JND noise.

Table 3	3:	Data-ef	ficiency	impr	oven	nents	with	JNDMix	in	the
state-of	f-t	he-art	models l	Туре	erIQA	A and	l DB	CNN.		

Method	LIV	/EC	KonIQ-10k		
Wiethou	SRCC	PLCC	SRCC	PLCC	
HyperIQA (100%)	0.859	0.882	0.906	0.917	
w/ JNDMix	0.870	0.883	0.911	0.924	
	+0.011	+0.001	+0.005	+0.007	
HyperIQA (50%)	0.820	0.843	0.892	0.908	
w/ JNDMix	0.821	0.855	0.902	0.913	
	+0.001	+0.012	+0.010	+0.005	
HyperIQA (25%)	0.787	0.809	0.880	0.896	
w/ JNDMix	0.790	0.818	0.885	0.900	
	+0.003	+0.009	+0.005	+0.004	
HyperIQA (10%)	0.700	0.736	0.851	0.866	
w/ JNDMix	0.740	0.773	0.861	0.874	
	+0.040	+0.037	+0.010	+0.008	
DBCNN (100%)	0.851	0.869	0.906	0.923	
w/ JNDMix	0.854	0.876	0.915	0.930	
	+0.003	+0.007	+0.009	+0.007	
DBCNN (50%)	0.814	0.843	0.902	0.918	
w/ JNDMix	0.815	0.849	0.907	0.924	
	+0.001	+0.006	+0.005	+0.006	
DBCNN (25%)	0.764	0.793	0.888	0.907	
w/ JNDMix	0.787	0.812	0.893	0.910	
	+0.023	+0.019	+0.005	+0.003	
DBCNN (10%)	0.699	0.740	0.869	0.889	
w/ JNDMix	0.718	0.756	0.873	0.893	
	+0.019	+0.016	+0.004	+0.004	

For rigorous comparison, the training settings for all models trained with JNDMix are the same as without it, including learning rate schedule, weight decay, data pre-processing and augmentations, etc. For three state-of-the-art models, we utilize their original codebases and training settings to train. For Resnet34 and Resnet50, we utilize AdamW [25] optimization with a mini-batch of 24, and set the base learning rate to 5e-5 with a cosine decay schedule. For ViT, we utilize AdamW optimization with a mini-batch of 8, and set base learning rate to 1e-5 with a cosine decay schedule.

3.2. JNDMix improves IQA benchmarks state-of-the-art

We train three representative state-of-the-art models in NR-IQA task, DBCNN, HyperIQA, and MANIQA, to study if JNDMix can improve state-of-the-art NR-IQA models on two IQA benchmarks, LIVEC and KonIQ-10k. Table 1 shows the results of applying JNDMix on the state-of-the-art models. All models using JNDMix have significant improvements based on their original state-of-the-art performance. Specifically, it provides a gain of +0.011 SRCC and +0.01 PLCC on LIVEC, resulting in MANIQA with the state-of-the-art performance of 0.897 SRCC and 0.918 PLCC on LIVEC, as well as achieve the state-of-the-art performance of 0.927 SRCC and 0.943 PLCC on KonIQ-10k.

Additionally, we reveal that JNDMix improves the performance of baseline models such as Resnet34, Resnet50 and Vision Transformer (ViT). Table 2 illustrates that we get excellent improvements over these baseline models. JNDMix improves ViT from 0.816 to 0.858 with a gain of +0.042 in

Method	LIV	/EC	KonIQ-10k		
Wiethou	SRCC	PLCC	SRCC	PLCC	
Resnet34 (100%)	0.824	0.849	0.905	0.920	
w/ JNDMix	0.834	0.855	0.910	0.928	
	+0.010	+0.006	+0.005	+0.008	
Resnet34 (50%)	0.788	0.816	0.890	0.908	
w/ JNDMix	0.811	0.834	0.898	0.914	
	+0.023	+0.018	+0.008	+0.006	
Resnet34 (25%)	0.738	0.746	0.869	0.888	
w/ JNDMix	0.758	0.766	0.881	0.900	
	+0.020	+0.020	+0.012	+0.012	
Resnet34 (10%)	0.635	0.609	0.850	0.871	
w/ JNDMix	0.674	0.653	0.859	0.878	
	+0.039	+0.044	+0.009	+0.007	
Resnet50 (100%)	0.835	0.868	0.916	0.931	
w/ JNDMix	0.842	0.872	0.920	0.932	
	+0.007	+0.004	+0.004	+0.001	
Resnet50 (50%)	0.801	0.827	0.902	0.917	
w/ JNDMix	0.812	0.840	0.910	0.923	
	+0.011	+0.013	+0.008	+0.006	
Resnet50 (25%)	0.751	0.768	0.886	0.902	
w/ JNDMix	0.767	0.794	0.892	0.908	
	+0.016	+0.026	+0.006	+0.006	
Resnet50 (10%)	0.657	0.682	0.863	0.880	
w/ JNDMix	0.681	0.695	0.868	0.884	
	+0.024	+0.013	+0.005	+0.004	

Table 4: Data-efficiency improvements with JNDMix in thebaseline models Resnet34 and Resnet50.

PLCC on LIVEC, and also achieves a great improvement on Resnet34 with a gain of +0.019 PLCC.

3.3. JNDMix helps data-efficiency

In this section, we demonstrate that JNDMix is effective to improve model performance across different dataset sizes and helps data efficiency of models. Table 3 illustrates that JNDMix always improves the perfomance of HyperIQA and DBCNN across all fractions of LIVEC and KonIQ-10k, which implies the data efficiency improvement of the state-ofthe-art models. JNDMix provides a surprising improvement of +0.04 SRCC in the low data regime (10% of data) while still being effective with a improvement of +0.007 PLCC in the high data regime on LIVEC. We detect that the performance growth brought by JNDMix is increasing with the decrease of the dataset sizes, which is also reflected in the difference in performance improvements between the two datasets, where the images of LIVEC is 10 times less than those of KonIQ-10k. Table 4 reveals that JNDMix also greatly helps data efficiency of baseline models with the maximum gain of +0.044 PLCC in the low data regime (10% of data).

3.4. JNDMix improves generalization ability

To show the generaliazation ability of NR-IQA models, the cross dataset evaluations are performed. We train MANIQA on different fractions of KonIQ-10k and test it on LIVEC. Table 5 proves that JNDMix is effective to improve generalization ability of NR-IQA models across different dataset sizes.

Table 5: Results of cross dataset evaluations trained on different fractions of datasets adopting the state-of-the-art model MANIQA.

Training	Testing	Method	SRCC	PLCC
		MANIQA (100%)	0.844	0.868
		w/ JNDMix	0.856	0.880
			+0.012	+0.012
		MANIQA (50%)	0.837	0.858
		w/ JNDMix	0.843	0.863
KanIO 10h	LIVEC		+0.006	+0.005
KoniQ-10k	LIVEC	MANIQA (25%)	0.828	0.854
		w/ JNDMix	0.837	0.862
			+0.009	+0.008
		MANIQA (10%)	0.797	0.828
		w/ JNDMix	0.808	0.836
			+0.011	+0.008
		MANIQA (100%)	0.790	0.849
LIVEC	KonIQ-10k	w/ JNDMix	0.803	0.861
			+0.013	+0.012

 Table 6: Performance of MANIQA with different noise injection on 100% LIVEC.

Method	SRCC	PLCC
MANIQA	0.886	0.908
w/ JNDMix	0.897	0.918
w/ JND injection	0.891	0.911
w/ Gaussian noise injection	0.796	0.825

3.5. Ablation Study

To study the contribution of JND noise injection, we conduct ablation study by injecting different noises. Table 6 shows the performance of MANIQA with different noise injection trained on 100% LIVEC. 'JND injection' means we do not use JND noise generated by multiplying the JND image with a resampled random ratio, but directly add the JND image to the training image. 'Gaussian noise injection' means we inject the Gaussian noise to generate the new training image. The results reveal that injecting Gaussian noise leads to performance degradation and adding the JND image results in smaller improvement than JNDMix.

4. CONCLUSION

Data augmentation is vital to many vision tasks, and however, no effective data augmentation method specifically designed for NR-IQA task is proposed. In this paper, we are the first to propose a general data augmentation method for NR-IQA task, and rigorously study the JNDMix, finding it highly effective and general. JNDMix works well across a variety of different model architectures and helps their data efficiency, both on LIVEC and KonIQ-10k. The JNDMix augmentation method is simple and can be flexibly plugged into any NR-IQA model codebase. We wish that our method will be widely used in NR-IQA task and the lack of data augmentation methods in NR-IQA task will receive attention.

5. REFERENCES

- Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.
- [2] Weixia Zhang, Kede Ma, Jia Yan, Dexiang Deng, and Zhou Wang, "Blind image quality assessment using a deep bilinear convolutional neural network," *IEEE Transactions on Circuits* and Systems for Video Technology, vol. 30, no. 1, pp. 36–47, 2018.
- [3] Shaolin Su, Qingsen Yan, Yu Zhu, Cheng Zhang, Xin Ge, Jinqiu Sun, and Yanning Zhang, "Blindly assess image quality in the wild guided by a self-adaptive hyper network," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 3667–3676.
- [4] Sidi Yang, Tianhe Wu, Shuwei Shi, Shanshan Lao, Yuan Gong, Mingdeng Cao, Jiahao Wang, and Yujiu Yang, "MANIQA: Multi-dimension Attention Network for No-Reference Image Quality Assessment," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 1191–1200.
- [5] Peng Ye, Baopu Li, Yikang Li, Tao Chen, Jiayuan Fan, and Wanli Ouyang, "b-DARTS: Beta-Decay Regularization for Differentiable Architecture Search," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, 2022, pp. 10874–10883.
- [6] Peng Ye, Baopu Li, Tao Chen, Jiayuan Fan, Zhen Mei, Chen Lin, Chongyan Zuo, Qinghua Chi, and Wanli Ouyang, "Efficient joint-dimensional search with solution space regularization for real-time semantic segmentation," *International Journal of Computer Vision*, vol. 130, no. 11, pp. 2674–2694, 2022.
- [7] Hancheng Zhu, Leida Li, Jinjian Wu, Weisheng Dong, and Guangming Shi, "MetaIQA: Deep meta-learning for noreference image quality assessment," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 14143–14152.
- [8] Chun-Hsiao Yeh, Cheng-Yao Hong, Yen-Chi Hsu, and Tyng-Luh Liu, "SAGA: Self-Augmentation with Guided Attention for Representation Learning," in *IEEE International Conference on Acoustics, Speech and Signal Processing.* IEEE, 2022, pp. 3463–3467.
- [9] Niall McLaughlin, Jesus Martinez Del Rincon, and Paul Miller, "Data-augmentation for reducing dataset bias in person reidentification," in *IEEE International conference on advanced video and signal based surveillance*. IEEE, 2015, pp. 1–6.
- [10] Terrance DeVries and Graham W Taylor, "Improved regularization of convolutional neural networks with cutout," arXiv preprint arXiv:1708.04552, 2017.
- [11] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer* vision and pattern recognition, 2015, pp. 1–9.
- [12] Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, and David Lopez-Paz, "mixup: Beyond Empirical Risk Minimization," in International Conference on Learning Representations, 2018.

- [13] Sangdoo Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, and Youngjoon Yoo, "Cutmix: Regularization strategy to train strong classifiers with localizable features," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 6023–6032.
- [14] Ross Girshick, Ilija Radosavovic, Georgia Gkioxari, Piotr Dollár, and Kaiming He, "Detectron," 2018.
- [15] Francisco J Moreno-Barea, Fiammetta Strazzera, José M Jerez, Daniel Urda, and Leonardo Franco, "Forward noise adjustment scheme for data augmentation," in *IEEE symposium series on computational intelligence*. IEEE, 2018, pp. 728–734.
- [16] Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang, "Musiq: Multi-scale image quality transformer," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 5148–5157.
- [17] Qingsen Yan and Dong Gong, "Two-stream convolutional networks for blind image quality assessment," *IEEE Transactions* on *Image Processing*, vol. 28, no. 5, pp. 2200–2211, 2018.
- [18] Charles F Hall and Ernest L Hall, "A nonlinear model for the spatial characteristics of the human visual system," *IEEE Transactions on systems, man, and cybernetics*, vol. 7, no. 3, pp. 161–170, 1977.
- [19] Deepti Ghadiyaram and Alan C Bovik, "Massive online crowdsourced study of subjective and objective picture quality," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 372–387, 2015.
- [20] Vlad Hosu, Hanhe Lin, Tamas Sziranyi, and Dietmar Saupe, "KonIQ-10k: An ecologically valid database for deep learning of blind image quality assessment," *IEEE Transactions on Image Processing*, vol. 29, pp. 4041–4056, 2020.
- [21] Bart Thomee, David A Shamma, Gerald Friedland, Benjamin Elizalde, Karl Ni, Douglas Poland, Damian Borth, and Li-Jia Li, "YFCC100M: The new data in multimedia research," *Communications of the ACM*, vol. 59, no. 2, pp. 64–73, 2016.
- [22] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceed-ings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [23] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," in *International Conference on Learning Representations*, 2020.
- [24] Jian Jin, Dong Yu, Weisi Lin, Lili Meng, Hao Wang, and Huaxiang Zhang, "Full RGB Just Noticeable Difference (JND) Modelling," arXiv preprint arXiv:2203.00629, 2022.
- [25] Ilya Loshchilov and Frank Hutter, "Decoupled weight decay regularization," in *International Conference on Learning Representations*, 2018.