

# OBIFormer: A Fast Attentive Denoising Framework for Oracle Bone Inscriptions

Jinhao Li<sup>a</sup>, Zijian Chen<sup>b</sup>, Tingzhu Chen<sup>c,\*</sup>, Zhiji Liu<sup>d</sup> and Changbo Wang<sup>a</sup>

<sup>a</sup>School of Computer Science and Technology, East China Normal University, Shanghai, 200062, China

<sup>b</sup>Institute of Image Communication and Information Processing, Shanghai Jiao Tong University, Shanghai, 200240, China

<sup>c</sup>School of Humanities, Shanghai Jiao Tong University, Shanghai, 200030, China

<sup>d</sup>Center for the Study and Application of Chinese Characters, East China Normal University, Shanghai, 200241, China

## ARTICLE INFO

### Keywords:

Oracle bone inscriptions  
Channel-wise self-attention  
Glyph information  
Image denoising  
Deep learning

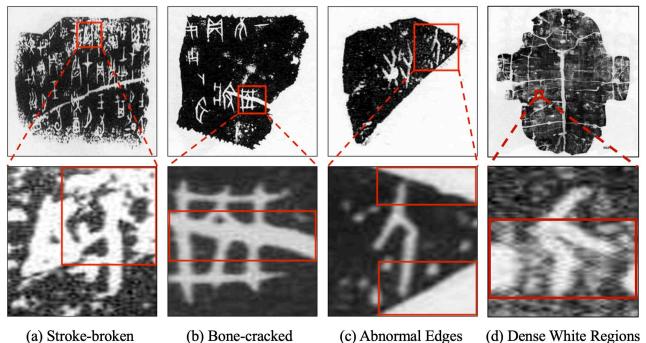
## ABSTRACT

Oracle bone inscriptions (OBIs) are the earliest known form of Chinese characters and serve as a valuable resource for research in anthropology and archaeology. However, most excavated fragments are severely degraded due to thousands of years of natural weathering, corrosion, and man-made destruction, making automatic OBI recognition extremely challenging. Previous methods either focus on pixel-level information or utilize vanilla transformers for glyph-based OBI denoising, which leads to tremendous computational overhead. Therefore, this paper proposes a fast attentive denoising framework for oracle bone inscriptions, i.e., OBIFormer. It leverages channel-wise self-attention, glyph extraction, and selective kernel feature fusion to reconstruct denoised images precisely while being computationally efficient. Our OBIFormer achieves state-of-the-art denoising performance for PSNR and SSIM metrics on synthetic and original OBI datasets. Furthermore, comprehensive experiments on a real oracle dataset demonstrate the great potential of our OBIFormer in assisting automatic OBI recognition. The code will be made available at <https://github.com/LJHolyGround/OBIFormer>.

## 1. Introduction

Oracle bone inscriptions (OBIs) represent China's earliest known mature and systematic writing system. Carved on materials such as oxen or turtle bones for divination and record-keeping during the late Shang and Zhou Dynasties, they provide profound insights into Chinese social culture for anthropologists and archaeologists. Therefore, the recognition of OBIs is an indispensable step in exploring the history of ancient China. However, despite the numerous fragments excavated, only a small percentage of these OBIs has been successfully recognized. Since the annotation task requires a high demand of time and labor with domain knowledge from OBI experts, there is an urgent need to develop an automatic OBI recognition algorithm. Unfortunately, many oracle bones have suffered considerable degradation over millennia due to natural weathering, corrosion, and man-made destruction, making automatic OBI recognition extremely challenging.

As shown in Fig. 1, noise in oracle bone inscriptions can be categorized into four types: stroke-broken, bone-cracked, abnormal edges, and dense white regions [1]. Stroke-broken refers to the clusters of white regions near the strokes, complicating OBI recognition. Bone-cracked is caused by occlusion and typically runs through the center of the image, disrupting the glyph information. Abnormal edges also arise



**Figure 1:** Four types of noise in real rubbings. The red rectangle indicates the corresponding noise. (a) Stroke-broken, (b) Bone-cracked, (c) Abnormal edges, (d) Dense white regions.

from occlusion and usually appear along the image boundaries. Besides, dense white regions represent fog-like noise, which obscures the structure and exacerbates ambiguity.

Traditional methods were first applied for OBI denoising. For example, Huang *et al.* [2] conducted a comprehensive comparative study of image denoising techniques relying on methods such as the anisotropic diffusion filter, Wiener filter, total variation, non-local means, and bilateral filtering. However, these methods do not perform well due to the complex degradation in OBIs. Therefore, some tailored denoising methods have been designed for the different noise types. Gu *et al.* [3] proposed an in-painting algorithm based on Poisson distribution and fractal geometry to remove erosion noise in OBIs. Khankasikam [4] proposed a binarization method based on adaptive multilayer information to restore

\*Corresponding author.

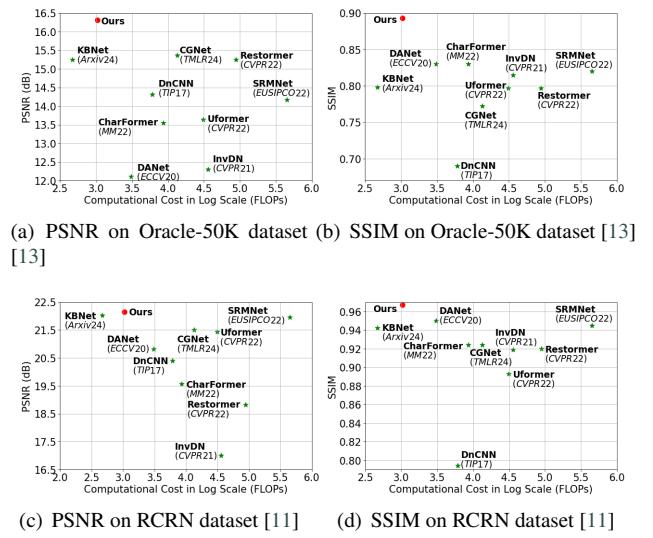
✉ 1omljhoax@stu.ecnu.edu.cn (J. Li); zijian.chen@sjtu.edu.cn (Z. Chen); tingzuchen@sjtu.edu.cn (T. Chen); 251755227@qq.com (Z. Liu); cbwang@cs.ecnu.edu.cn (C. Wang)  
ORCID(s): 0000-0002-8502-4110 (Z. Chen)

uneven backgrounds in degraded historical document images. Robust Kronecker component analysis (RKCA) [5, 6] has recently been proposed for handling structured data with inherent tensorial properties. Subsequently, Zhang *et al.* [7] designed the Kronecker component with the low-rank dictionary (KCLD) and the Kronecker component with the robust low-rank dictionary (KCRD), which takes the Nuclear norm into RKCA to better capture the low-rank property of the two dictionaries in the basic sparse representation model. Considering the internal structural, spatial, and spectral information of the image block, they proposed the robust low-rank analysis with adaptive weighted tensor (AWTD) [8], a method that applies the adaptive weight tensor to the low-rank tensor model for image denoising.

With the prevalence of deep learning, numerous methods have been proposed for OBI denoising. For example, DnCNN [9] was first introduced to handle blind noise with unknown noise levels. Zamir *et al.* [10] designed Restormer, an efficient encoder-decoder transformer for high-resolution image restoration. However, these generic denoising models mainly focus on pixel-level information while neglecting glyph information, which leads to poor performance. RCRN [11] first utilizes character skeleton information to achieve real-world character image restoration. Nevertheless, it extracts skeleton information from the input images, which results in unsatisfactory results. CharFormer [12] attempts to leverage glyph information from the target images to address the problem, but it adopts a simple feature fusion strategy that limits its performance. Moreover, it leverages three vanilla transformer blocks in the residual self-attention block, which leads to tremendous computational overhead.

Additionally, automatic OBI recognition also suffers from data scarcity. In the early days, data augmentation methods primarily relied on techniques such as random rotation and flipping. Later, Han *et al.* [13] introduced an Orc-Bert Augmentor pre-trained by self-supervised learning to recover masked input and generate pixel-format images as augmented data. Recently, Wang *et al.* [14] proposed a structure-texture separation network (STSN), which disentangles features into structure and texture components. It swaps texture information between any pairs of images so that transformed images are realistic and diverse.

To address these problems, we propose a fast attentive denoising framework for OBIs, i.e., OBIFformer, which utilizes channel-wise self-attention, glyph extraction, and selective kernel feature fusion. Specifically, our OBIFformer comprises an input projector, an output projector, an additional feature corrector, and several OBIFformer blocks (OFBs). The input projector extracts shallow features from the input images. The OBIFformer block is composed of channel-wise self-attention blocks (CSAB), glyph structural network blocks (GSNB), and a selective kernel feature fusion (SKFF) module. The SKFF module aggregates the reconstruction and glyph features captured by CSAB and GSNB. Finally, the output projector reconstructs the denoised image, and the additional feature corrector obtains the skeleton image. We conduct comprehensive experiments



**Figure 2:** Our model achieves state-of-the-art performance on the OBI denoising task while being computationally efficient.

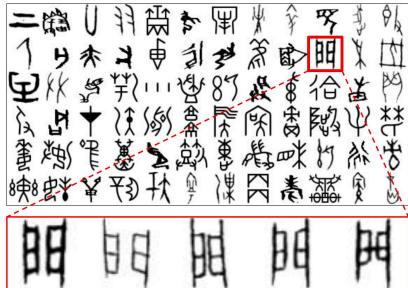
and demonstrate the effectiveness of our OBIFformer for OBI denoising tasks on Oracle-50K [13] and RCRN [11] datasets (See Fig. 2). Furthermore, extensive experiments on a real oracle dataset (i.e., the OBC306 dataset [1]) demonstrate its great potential in assisting automatic OBI recognition. Finally, we provide ablation studies to show the effectiveness of architectural designs and experimental choices. The main contributions of this work can be summarized as follows:

- We propose a fast attentive denoising framework for OBIs, i.e., OBIFformer, based on channel-wise self-attention, glyph extraction, and selective kernel feature fusion. Additionally, we apply domain adaptation for the Oracle-50K dataset to synthesize noisy images for experiments.
- Comprehensive experiments on synthetic and original OBI datasets demonstrate the superiority of our OBIFformer for OBI denoising tasks. Furthermore, our OBIFformer is computationally efficient compared to other baseline methods.
- Extensive experiments on the OBC306 dataset show the strong generalization ability of the proposed OBIFformer. It demonstrates the great potential of OBIFformer in assisting automatic OBI recognition.

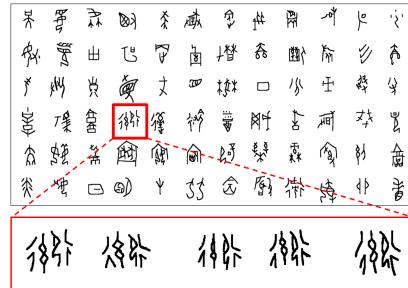
## 2. Related Works

### 2.1. Oracle Bone Inscriptions Datasets

In recent years, various OBI datasets have been established for OBI-related computer vision tasks, such as oracle bone recognition, rejoining, classification, retrieval, and deciphering [15]. From the perspective of the content sources, OBI datasets can be categorized into handprints and rubbings, as shown in Fig. 3. Furthermore, we summarize their statistical information in Table 1.



(a) Oracle-50K [13]



(b) HWOBC [16]



(c) EVOBI [17]



(d) OBC306 [1]



(e) OBI125 [18]



(f) EVOBC [19]

**Figure 3:** Examples of oracle character images in different OBI datasets: (a) Oracle-50K [13], (b) HWOBC [16], (c) EVOBI [17], (d) OBC306 [1], (e) OBI125 [18], and (f) EVOBC dataset [19]. The zoomed-in images are different structural variations of the same character.

**Table 1**

Summary of the existing oracle bone inscription datasets.

Type	Dataset	Year	#Classes	#Total	Avg. Res.	Application	Availability
Handprints	Oracle-20K [20]	2016	261	20,039	50×50	Recognition	✓
	Yang <i>et al.</i> [21]	2018	39	21,373	-	Recognition	✗
	Liu <i>et al.</i> [22]	2018	5,491	44,868	-	Recognition	✗
	Oracle-50K [13]	2020	2,668	59,081	50×50	Recognition	✓
	HWOBC [16]	2020	3,881	83,245	400×400	Recognition	✓
	EVOBI [17]	2022	972	4,860	105×105	Recognition	✓
Rubbings	OBC306 [1]	2019	306	309,551	<382×478	Recognition	✓
	Wang <i>et al.</i> [23]	2019	-	7,824	-	Denoising & Recognition	✗
	Liu <i>et al.</i> [24]	2020	682	-	-	Recognition	✗
	OracleBone8000 [25]	2020	-	129,770	-	Rejoining & Recognition	✗
	Yoshiyuki <i>et al.</i> [26]	2022	27	649	-	Detection & Recognition	✗
	OBI125 [18]	2022	125	4,257	<278×473	Recognition	✓
	OB-Rejoin [27]	2022	-	998	<1408×1049	Rejoining	✗
	Oracle-MINST [28]	2023	10	30,222	28×28	Recognition	✓
	EVOBC [19]	2024	13,714	229,170	<465×857	Generation & Recognition	✓
	HUST-OBC [29]	2024	10,999	140,053	<400×520	Recognition	✓
Hybrid	O2BR [15]	2025	-	800	<2664×2167	Recognition	✓
	OBI-rejoin [15]	2025	-	200	<2913×1268	Rejoining	✓
	Oracle-241 [14]	2022	241	78,565	<588×700	Generation & Recognition	✓
	Oracle-P15K [30]	2025	239	14,542	128×128	Generation & Denoising	✓

Handprinted OBI datasets contain distortion-free rubbing images rewritten by OBI experts. For example, the Oracle-20K dataset [20] was collected and labeled by Guo

*et al.* from the Chinese etymology website<sup>1</sup>. It encompasses 20,039 oracle character instances across 261 categories.

<sup>1</sup><http://www.chineseetymology.org/>

Later, Han *et al.* [13] further expanded the number of instances and proposed the Oracle-50K dataset, which consists of 59,081 instances belonging to 2,668 categories. Similarly, Li *et al.* built the HWOBC dataset [16] for training automatic OBI recognition models. It became the largest handprinted OBI dataset with 83,245 character-level samples grouped into 3,881 categories. Moreover, the EVOBI dataset [17] was proposed for OBI interpretation, which was crawled from the Chinese master website<sup>2</sup>. Nevertheless, these datasets only provide handprinted images, contributing minimally to automatic OBI recognition. To synthesize more realistic OBI data, Wang *et al.* [14] constructed the Oracle-241 dataset to transfer knowledge across handprinted characters and scanned data. It comprises 78,565 handprinted and scanned images of 241 categories. Most recently, Li *et al.* [30] proposed Oracle-P15K, a structure-aligned OBI dataset consisting of 14,542 images infused with domain knowledge from OBI experts, to achieve realistic and controllable OBI generation. Comprehensive experiments demonstrate that the augmented images can mitigate the long-tail distribution problem in existing OBI datasets.

Rubbing OBI datasets consist of scanned images from oracle bone publications, most of which suffer from severe and distinctive noise caused by thousands of years of natural weathering, corrosion, and man-made destruction. One of the representative datasets, OBC306 [1], encompasses 309,551 samples classified into 306 classes from eight authoritative oracle bone publications. The OracleBone8000 dataset [25] contains 129,770 images with character-level annotations. However, it is highly imbalanced and sparse, limiting itself to serving as a comprehensive benchmark for automatic OBI recognition tasks. Yue *et al.* [18] built the OBI125 dataset, which is composed of 4,257 images scanned from the collection of the Shanghai Museum (Volume I) [31]. For the oracle bone rejoining task, Zhang *et al.* [27] proposed the OB-Rejoin dataset, which covers different writing styles and fonts, featuring 249 pairs of known rejoining manually found by OBI experts over the past few decades. Moreover, the Oracle-MINST dataset [28] was released to benchmark the oracle bone character classification task, following the same data format as the well-known MNIST dataset [32]. Similar to EVOBI, the EVOBC dataset [19] was constructed to trace the evolution of Chinese characters from oracle bone inscriptions to their contemporary forms, including written representations of the same character across different historical periods. Wang *et al.* [29] introduced the HUST-OBC dataset, which comprises 141,053 images sourced from five different origins. Among them, 77,064 images spanning 1,781 categories have been deciphered, while 62,989 images across 9,411 categories remain undeciphered. Most recently, Chen *et al.* [15] proposed O2BR and OBI-rejoin datasets to evaluate the recent large multi-modal models (LMMs) in OBI recognition and rejoining tasks. Each image in these datasets is equipped with several questions alongside correct answers.

<sup>2</sup><http://www.guoxuedashi.net/>

## 2.2. Oracle Bone Inscription Recognition

The traditional OBI recognition adopts a three-stage pipeline paradigm: data pre-processing, feature extraction, and recognition. In the early days, Guo *et al.* [20] proposed a hierarchical representation that integrates a Gabor-related low-level representation and a sparse-encoder-related mid-level representation with CNN-based models, which achieved better performance over both approaches. Similarly, Yang *et al.* [21] also applied a feature extraction technique in CNN, but CNN completes both feature extraction and OBI recognition tasks. However, the simple feature representation limits the performance of traditional OBI recognition methods. Later, Du *et al.* [22] designed a two-branch deep learning framework consisting of two pretext tasks for rotation and deformation. Experimental results demonstrated that their method could effectively learn features of OBIs and provide good feature representation for the downstream tasks.

Though achieving good accuracy, most traditional methods rely heavily on manually designed complex features at multiple levels. Therefore, they are not suitable for dealing with large-scale datasets. In contrast, deep learning methods are known for their representation capacity when processing large-scale datasets. Huang *et al.* [1] first introduced the standard deep CNN-based evaluation (i.e., AlexNet [33], Inception-v4 [34], VGG16 [35], ResNet-50, and ResNet-101 [36]) for the OBC306 dataset, which served as a benchmark. However, the long-tail distribution problem in OBI datasets hampers the performance of deep learning methods. Hence, Zhang *et al.* [37] uses a CNN to map the character images to a Euclidean space where the nearest neighbor rule classifies them. Most recently, Wang *et al.* [14] proposed a structure-texture separation network (STSN), which separates texture from the structure to avoid the negative influence caused by degradation. Nevertheless, the recognition accuracy of scanned images is still unsatisfactory. Afterward, they trained an unsupervised discriminative consistency network (UDCN) [38] with an unsupervised transition loss and leveraged pseudo-labeling to make the predictions of scanned samples consistent under different noise.

## 2.3. Image Denoising

Image denoising plays a vital role in different image processing applications, such as medical imaging [39, 40], remote sensing [41], and oracle bone research [11, 12]. Since the rise of deep learning, researchers have proposed various models for image denoising. For example, Zhang *et al.* [9] leveraged residual learning and batch normalization to introduce DnCNN, a model capable of handling blind noise with unknown noise levels. Later, Fan *et al.* [42] presented SRMNet, a blind real-image denoising network utilizing a hierarchical architecture improved from U-Net [43], which advanced the performance on two synthetic and two real-world noisy datasets. However, it requires tremendous computational overhead due to the complex hierarchical architecture. Therefore, Zhang *et al.* [44] proposed KBNet, which

consists of a kernel basis attention module and a multi-axis feature fusion block. It performs state-of-the-art on over ten image denoising benchmarks while maintaining low computational overhead. Besides, Liu *et al.* [45] designed a lightweight, information-lossless, and memory-saving invertible neural network (INN) based model, namely InvDN, which replaces the noisy latent representation with another one sampled from a prior distribution during reversion and achieves promising results.

Generative adversarial network (GAN) based models have recently been applied for image denoising. Yue *et al.* [46] introduced DANet, a unified framework synthesizing noisy-clean image pairs simultaneously with two new metrics. To preserve the structural consistency of characters, Shi *et al.* [11] proposed RCRN, which comprises a skeleton extractor (SENet) and a character image restorer (CiRNet). It is capable of handling complex degradation and specific noise types in OBIs.

More recently, researchers have introduced visual transformer (ViT) [47] to image denoising tasks. Shi *et al.* [12] designed a glyph-based attentive framework, i.e., CharFormer, which maintains critical features of a character during the denoising process. Later, Wang *et al.* [48] proposed Uformer based on a locally-refined window (LeWin) transformer and a learnable multi-scale restoration modulator. It excels at capturing local and global dependencies, making it highly effective for image restoration. However, though these transformer-based models obtain impressive performance, their computational complexity grows quadratically with the spatial resolution. Therefore, Zamir [10] applied multi-Dconv head transposed attention (MDTA) and gated-Dconv feed-forward network (GDFN) in Restormer to balance the performance and computational overhead. Most recently, Ghasemabadi *et al.* [49] presented CGNet to capture global information without self-attention, thus reducing the computational complexity significantly while maintaining outstanding performance.

### 3. Method

In this section, we elaborate on the overall pipeline and the hierarchical structure of our OBIFormer. Specifically, we provide the details of the OBIFormer block (OFB), which consists of channel-wise self-attention blocks (CSABs), glyph structural network blocks (GSNBs), and a selective kernel feature fusion (SKFF) module. The SKFF injects glyph features extracted by GSNBs into the denoising backbone, thereby guiding the model in removing the complex noise while preserving the inherent glyphs.

#### 3.1. Overall Pipeline

As shown in Fig. 4, the overall structure of our OBIFormer is a U-shaped encoder-decoder network with skip connections between the OFBs. Specifically, given a degraded image  $\mathbf{I} \in \mathbb{R}^{3 \times H \times W}$ , the input projector applies a  $3 \times 3$  convolution layer with LeakyReLU to extract the shallow features  $\mathbf{F}_0 \in \mathbb{R}^{C \times H \times W}$  from the input image. Then, the shallow features  $\mathbf{F}_0$  are passed through  $N + 1$  encoder

stages. Each stage contains an OFB and a downsampling layer. The OFB is designed for precise OBI image denoising based on the glyph feature extraction and aggregation. Given intermediate shallow features  $\mathbf{F}_i$ , the output of OFB<sub>*i*</sub> is:

$$\mathbf{F}_{i+1} = \text{OFB}_{i+1}(\mathbf{F}_i), 0 \leq i \leq N. \quad (1)$$

In the downsampling layer, we downsample the maps and double the channels using  $4 \times 4$  convolution with stride 2.

Next, the latent features  $\mathbf{F}_{N+1}$  are passed through  $N$  decoder stages. Each stage contains an OFB and an upsampling layer. Moreover, we apply skip connections between the OFBs in the encoder and decoder to transmit extraction results on different scales and restore spatial features. Therefore, the deep features  $\mathbf{F}_{i+1}$  can be formulated as:

$$\mathbf{F}_{i+1} = \text{OFB}_i(\mathbf{F}_i + \mathbf{F}_{2N-i+1}), N < i \leq 2N \quad (2)$$

where  $n$  is the stage number of the decoder, OFB <sub>$2N-i+1$</sub>  represents the  $(2N - i + 1)$ -th OFB. For feature upsampling, we apply a transposed convolution operation, which upsamples the maps and halves channels using  $2 \times 2$  convolution with stride 2.

The last OFB produces  $\mathbf{F}_{2N+1}$ , which consists of reconstruction features  $\mathbf{F}_{2N+1}^R$  and glyph features  $\mathbf{F}_{2N+1}^G$ . Finally, reconstruction features  $\mathbf{F}_{2N+1}^R$  are aggregated with the shallow features  $\mathbf{F}_0^R$ . The output projector reconstructs the restored image  $\mathbf{I}'$  with a  $3 \times 3$  convolution layer:

$$\mathbf{I}' = \text{Conv}(\mathbf{F}_0^R + \mathbf{F}_{2N+1}^R). \quad (3)$$

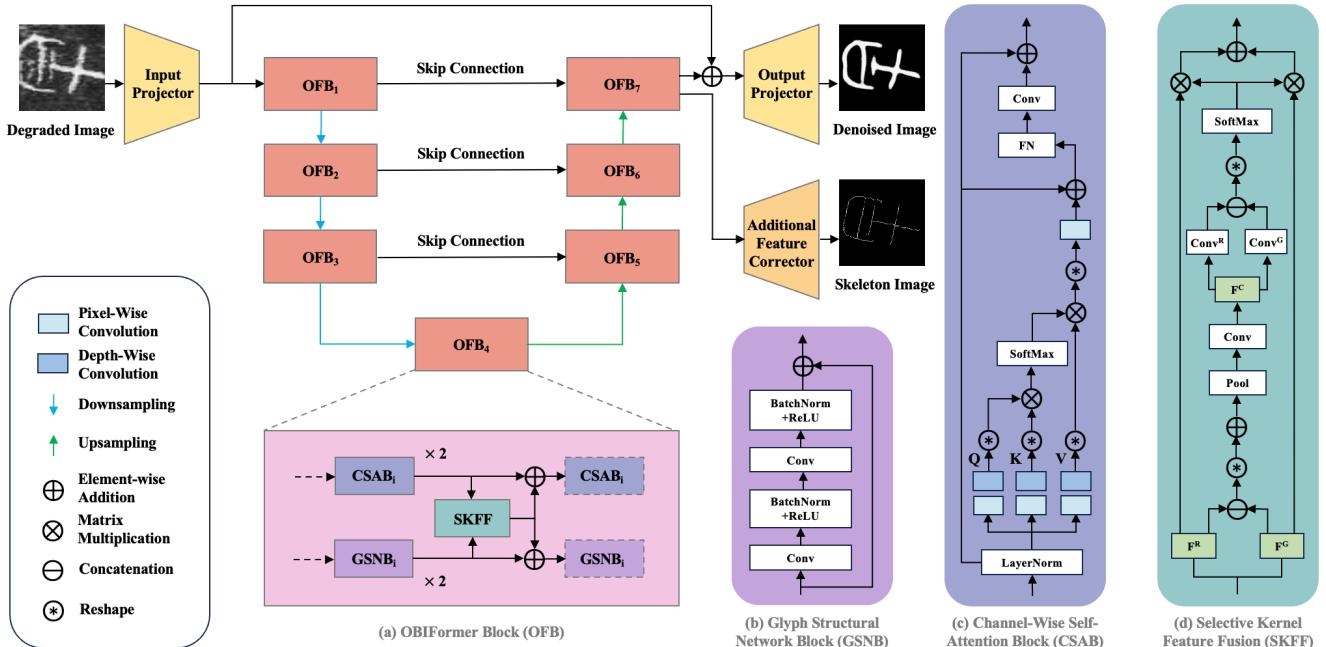
Similarly, the additional feature corrector adjusts the glyph features  $\mathbf{F}_{2N+1}^G$  to obtain the restored skeleton  $\mathbf{S}'$  with a  $3 \times 3$  convolution layer:

$$\mathbf{S}' = \text{Conv}(\mathbf{F}_{2N+1}^G). \quad (4)$$

#### 3.2. OBIFormer Block

As shown in Fig. 4(a), the OBIFormer block (OFB) consists of two residual channel-wise self-attention blocks (CSABs) as the denoising backbone, two glyph structural network blocks (GSNBs) for glyph extraction, and a selective kernel feature fusion (SKFF) module for feature aggregation. The primary purpose of the OFB is to alleviate the computational overhead and aggregate glyph features. We elaborate on the designs of these modules as follows:

**Residual Channel-Wise Self-Attention Block.** The residual channel-wise self-attention block (CSAB) is a residual block with a modified transformer layer, as shown in Fig. 4(c). Recently, transformer-based models have swept over various tasks with impressive results. However, the complexity of self-attention grows quadratically with the spatial resolution. Therefore, we apply channel-wise self-attention instead of spatial-wise self-attention, which remains effective while being computationally efficient. Specifically, each transformer layer performs channel-wise self-attention (CSA) and feed-forward (FN). For a normalized tensor  $\mathbf{F}_N$ , query ( $\mathbf{Q}$ ), key ( $\mathbf{K}$ ), and value ( $\mathbf{V}$ ) are generated by applying  $1 \times 1$  convolutions to aggregate pixel-wise cross-channel



**Figure 4:** The overall architecture of our OBIFformer. (a) OBIFformer block (OFB) that injects glyph information into the denoising backbone, (b) Glyph structural network block (GSNB) that extracts glyph features, (c) Channel-wise self-attention block (CSAB) that generates channel-wise self-attention effectively and efficiently, (d) Selective kernel feature fusion (SKFF) module that aggregates reconstruction features and glyph features.

context followed by  $3 \times 3$  depth-wise convolutions to encode channel-wise spatial context:

$$\mathbf{Q} = \mathbf{W}_d^Q \mathbf{W}_p^Q \mathbf{F}_N, \quad (5)$$

$$\mathbf{K} = \mathbf{W}_d^K \mathbf{W}_p^K \mathbf{F}_N, \quad (6)$$

$$\mathbf{V} = \mathbf{W}_d^V \mathbf{W}_p^V \mathbf{F}_N, \quad (7)$$

where  $\mathbf{W}_d^{(\cdot)}$  and  $\mathbf{W}_p^{(\cdot)}$  denote the projection matrices obtained by  $1 \times 1$  point-wise convolution and  $3 \times 3$  depth-wise convolution with no bias, respectively. Then, we reshape the query and key projection so that their dot product generates a transposed-attention map of size  $\mathbb{R}^{C \times C}$  instead of the regular attention map of size  $\mathbb{R}^{HW \times HW}$ :

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{SoftMax}\left(\frac{\mathbf{K}\mathbf{Q}}{\alpha}\right)\mathbf{V}, \quad (8)$$

where  $\mathbf{Q} \in \mathbb{R}^{HW \times C}$ ,  $\mathbf{K} \in \mathbb{R}^{C \times HW}$ , and  $\mathbf{V} \in \mathbb{R}^{HW \times C}$  matrices are obtained after reshaping.  $\alpha$  is a learnable scaling parameter. Therefore, the output features of the transformer layer in CSAB can be formulated as follows:

$$\mathbf{F}'_l = \mathbf{W}_p \text{CSA}(\text{LN}(\mathbf{F}_{l-1})) + \mathbf{F}_{l-1}, \quad (9)$$

$$\mathbf{F}_l = \text{FN}(\text{LN}(\mathbf{F}'_l)), \quad (10)$$

where LN represents layer normalization.  $\mathbf{F}'_l$  refers to the intermediate output features of the  $l$ -th transformer layer.  $\mathbf{F}_{l-1}$  and  $\mathbf{F}_l$  denote the final output features of the  $(l-1)$ -th and  $l$ -th transformer layer, respectively.

**Glyph Structural Network Block.** As shown in Fig. 4(b), the glyph structural network block (GSNB) consists of two

$3 \times 3$  convolution layers with batch normalization and ReLU activation function in a residual block. Due to the strong capacity to capture local dependencies of CNNs, GSNB can effectively extract glyph features that will be fused with reconstruction features. Similarly, GSNB incorporates downsampling and upsampling layers to maintain the same scale as the corresponding RSAB. Given input glyph features  $\mathbf{F}_i^G \in \mathbb{R}^{C \times H \times W}$ , the GSNB in  $n$ -th OFB will output  $\mathbf{F}_{i+1}^G$  which has the same size as  $\mathbf{F}_{i+1}^R$ .

**Selective Kernel Feature Fusion.** To better aggregate the glyph features, we perform selective kernel feature fusion (SKFF) [50] instead of simple addition or concatenation. As shown in Fig. 4(d), the SKFF module dynamically adjusts the receptive field via a triplet of operators: *Split*, *Fuse*, and *Select*. The *Split* operation generates reconstruction features  $\mathbf{F}^R$  and glyph features  $\mathbf{F}^G$  with different convolution layers. Then, the *Fuse* operation combines them to obtain a compact feature representation  $\mathbf{F}^C$  by applying element-wise summation, global average pooling, and pixel-wise convolution, which can be formulated as:

$$\mathbf{F}^C = \text{Conv}(\text{Pool}(\mathbf{F}^R + \mathbf{F}^G)). \quad (11)$$

Finally, the *Select* operation guides two other convolution layers followed by the softmax attention to enhance certain features:

$$\text{Attn}^R = \text{SoftMax}(\text{Conv}^R(\mathbf{F}^C), \text{Conv}^G(\mathbf{F}^C)), \quad (12)$$

$$\text{Attn}^G = \text{SoftMax}(\text{Conv}^R(\mathbf{F}^C), \text{Conv}^G(\mathbf{F}^C)), \quad (13)$$

where  $\text{Attn}^R$  and  $\text{Attn}^G$  represent the softmax attention of reconstruction and glyph features. The fused reconstruction

and glyph features are computed by multiplying the softmax attention with the  $\mathbf{F}^R$  and the  $\mathbf{F}^G$ , respectively:

$$\mathbf{F}^{FR} = \text{Attn}^R \mathbf{F}^R, \quad (14)$$

$$\mathbf{F}^{FG} = \text{Attn}^G \mathbf{F}^G, \quad (15)$$

where  $\mathbf{F}^{FR}$  and  $\mathbf{F}^{FG}$  are fused reconstruction and glyph features. The fused features  $\mathbf{F}^F$  are obtained by summing the  $\mathbf{F}^{FR}$  and the  $\mathbf{F}^{FG}$ :

$$\mathbf{F}^F = \mathbf{F}^{FR} + \mathbf{F}^{FG}. \quad (16)$$

### 3.3. Loss Function

We train the model with PSNR loss for the reconstructed OBI image  $\mathbf{I}'$  and perceptual loss for the reconstructed skeleton image  $\mathbf{S}'$ :

$$\mathcal{L} = \alpha_1 \mathcal{L}_1(\mathbf{I}') + \alpha_2 \mathcal{L}_2(\mathbf{I}') + \alpha_3 \mathcal{L}_1(\mathbf{S}') + \alpha_4 \mathcal{L}_2(\mathbf{S}'), \quad (17)$$

where  $\{\alpha_i\}_{i=1}^4$  are hyperparameters.  $\mathcal{L}_1$  and  $\mathcal{L}_2$  refer to PSNR and perceptual loss. Given an input image  $\mathbf{X}$ ,  $\mathcal{L}_1$  and  $\mathcal{L}_2$  are defined as:

$$\mathcal{L}_1(\mathbf{X}) = 10 \log \left( \frac{\max^2(\mathbf{X})}{\text{MSE}(\mathbf{X}_{GT}, \mathbf{X})} \right), \quad (18)$$

$$\mathcal{L}_2(\mathbf{X}) = \| \text{VGG}(\mathbf{X}_{GT}) - \text{VGG}(\mathbf{X}) \|_1, \quad (19)$$

where max and MSE denote the maximum value representing the color of the image pixel and the mean square error.  $\mathbf{X}_{GT}$  refers to the ground truth of the input image.  $\mathcal{L}_2$  is computed based on a VGG16 [35] model pre-trained on the ImageNet [51] dataset. The ground truth of the skeleton is obtained by an existing method [52] based on mathematical morphology.

## 4. Experiments

In this section, we conduct comprehensive experiments on three representative OBI datasets, i.e., Oracle-50K [13], RCRN [11], and OBC306 [1], to evaluate the effectiveness of the proposed OBIFormer for the OBI denoising task. The details are as follows:

### 4.1. Datasets

**Oracle-50K:** The Oracle-50K dataset [13] is a large OBI dataset designed for OBI recognition and classification tasks. It contains various instances sourced from three different collections. The instances from Xiaoxuetang<sup>3</sup> and Chinese Etymology<sup>4</sup> are gathered by a custom web-crawling tool, while other instances are generated with a TrueType font file. Considering the long-tail distribution of oracle character instances in the Oracle-50K dataset, we solely select the top 100 characters with the highest frequency for our experiments.

**RCRN:** The RCRN dataset [11] is sampled from historical Chinese character and oracle document datasets. Due to its

complex real-world degradation, it is an essential benchmark for OBI denoising tasks [11, 12]. While the test set is not publicly available, we use the training set for training, validation, and testing, which consists of 900 noisy-clean image pairs. For each pair, we split it into a noisy image and a clean image for experiments.

**OBC306:** The OBC306 dataset [1] is a well-known rubbing dataset constructed using eight authoritative oracle bone publications worldwide. These publications are first scanned and encoded using a six-character/number code. The characters in scanned images are retrieved with the help of an oracle bone dictionary tool by comparing the text in the images with A List of Oracle Characters [53]. Finally, the best-matching character samples are added to the dataset.

### 4.2. Implementation Details

We use Oracle-50K, RCRN, and OBC306 datasets for our experiments. Among them, the RCRN and OBC306 datasets are rubbing datasets, and the Oracle-50K dataset is a handprint dataset. Specifically, we utilize STSN [14] to apply the domain adaptation for the Oracle-50K dataset. Then, these datasets are used to train denoising models. Our model is implemented using the PyTorch platform and trained on an NVIDIA RTX 4090 GPU for 300 epochs with the AdamW optimizer [54]. The learning rate is set to 2e-4, and the weight decay is set to 0.01. We use a batch size of 10, and all images are resized to  $256 \times 256$ . For the RCRN dataset, due to its limited data scale, we adopt a data augmentation technique consisting of random rotation and horizontal and vertical flipping. Considering various quality assessment metrics [55, 56, 57, 58], we apply two common metrics of low-level vision tasks, i.e., peak signal-to-noise ratio (PSNR) and the structural similarity index measure (SSIM) [59]. Higher PSNR and SSIM values indicate better denoising results. For Oracle-50K and RCRN datasets, their target images serve as the ground truth to compute the PSNR and SSIM metrics.

### 4.3. Baseline Algorithms

To evaluate the performance of the proposed OBIFormer, we select nine representative models, i.e., Denoising Convolutional Neural Networks (DnCNN) [9], Dual Adversarial Network (DANet) [46], Selective Residual M-Net (SRMNet) [42], Kernel Basis Network (KBNet) [44], Invertible Denoising Network (InvDN) [45], CharFormer [12], U-Shaped Transformer (Uformer) [48], Restoration Transformer (Restormer) [10], and CascadedGaze Network (CGNet) [49] for comparisons.

### 4.4. Comparison with the State-of-the-Art

Table 2 reports the quantitative comparisons of baseline methods and our OBIFormer on Oracle-50K and RCRN datasets. On the Oracle-50K dataset, OBIFormer achieves 16.31 dB on PSNR and 0.893 on SSIM, surpassing all the other methods by at least 1.06 dB and 0.063, respectively. Similarly, on the RCRN dataset, OBIFormer attains 22.19 dB on PSNR and 0.969 on SSIM, outperforming all the other methods by at least 0.16 dB in PSNR and 0.019 in SSIM.

<sup>3</sup><http://xiaoxtue.iis.sinica.edu.tw/jiaguwen>

<sup>4</sup><https://hanziyuan.net/>

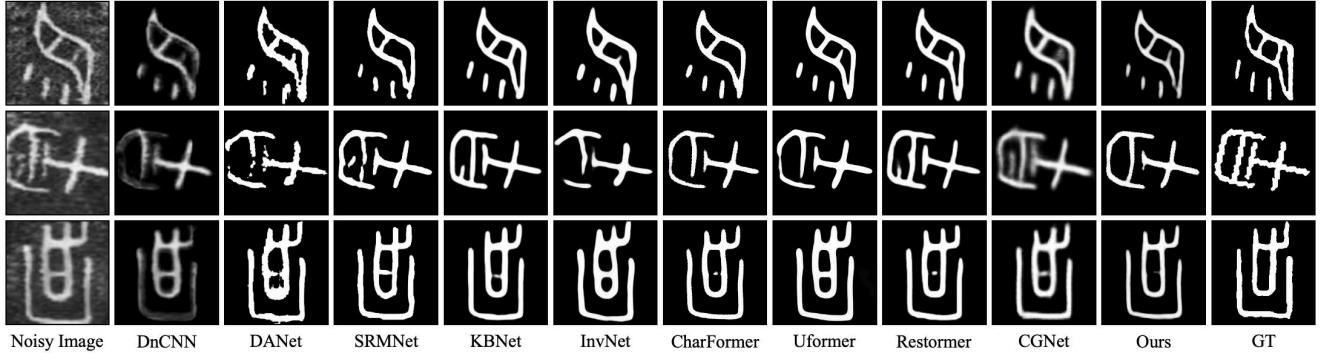


Figure 5: Qualitative comparisons of baseline methods and our OBIFformer on Oracle-50K dataset [13].

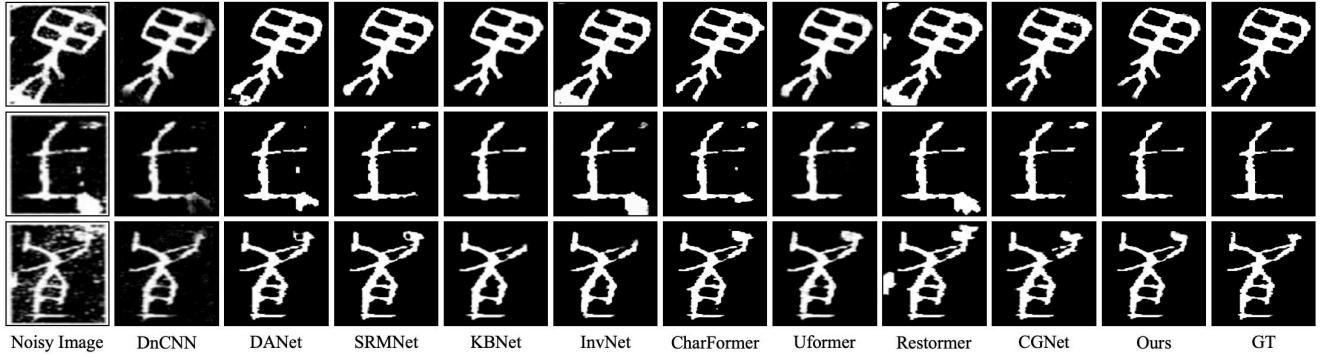


Figure 6: Qualitative comparisons of baseline methods and our OBIFformer on RCRN dataset [11].

**Table 2**

Quantitative comparisons of baseline methods and our OBIFformer on Oracle-50K [13] and RCRN [11] datasets. The compared methods include CNN, INN, GAN, and transformer-based models. The best and second-best results are in bold and underlined, respectively.

Methods	Oracle-50K [13]		RCRN [11]	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑
Raw Image	10.25	0.099	9.111	0.611
DnCNN [9]	14.31	0.690	20.40	0.794
DANet [46]	12.11	<u>0.830</u>	20.81	<u>0.950</u>
SRMNet [42]	14.17	0.820	21.96	0.945
KBNet [44]	<u>15.25</u>	0.798	<u>22.03</u>	0.942
InvDN [45]	12.30	0.815	18.96	0.941
CharFormer [12]	13.55	<u>0.830</u>	19.57	0.924
Uformer [48]	13.64	<u>0.797</u>	21.44	0.893
Restormer [10]	<u>15.25</u>	0.797	21.59	0.945
CGNet [49]	14.82	0.819	21.51	0.944
Ours	<b>16.31</b>	<b>0.893</b>	<b>22.19</b>	<b>0.969</b>

Besides, DANet achieves 0.830 and 0.950 in terms of SSIM on Oracle-50K and RCRN, respectively. KBNet obtains PSNR values of 15.25 on Oracle-50K and 22.03 on RCRN. However, these methods excel in only one metric while exhibiting poor performance in the other. In contrast, our OBIFformer achieves state-of-the-art performance on both

PSNR and SSIM, with a particularly notable improvement on SSIM, highlighting the effectiveness of GSNB and SKFF.

Furthermore, we provide qualitative comparisons of baseline methods and our OBIFformer on Oracle-50K and RCRN datasets. As illustrated in Fig. 5 and Fig. 6, OBIFformer effectively removes complex noise while preserving glyph details, whereas other methods struggle to maintain glyph consistency. Notably, OBIFformer generates cleaner and visually closer images to the ground truths than other algorithms in challenging scenarios where noise and strokes are difficult to distinguish. For example, most other methods fail to remove the spindle-shaped noise in the second case in Fig. 6. In the first case in Fig. 5 and the third case in Fig. 6, only our OBIFformer succeeds in restoring the broken strokes precisely. Overall, OBIFformer achieves state-of-the-art performance on Oracle-50K and RCRN quantitatively and qualitatively.

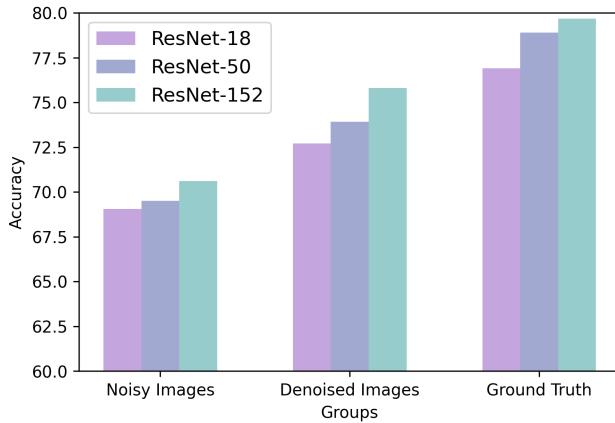
#### 4.5. OBI Recognition

To further validate the effectiveness of OBI denoising in enhancing recognition accuracy, we employ ResNet-18, ResNet-50, and ResNet-152 [36] for the OBI recognition task on the test set of the Oracle-50K dataset. The test set is divided into training and testing subsets in a 7:3 ratio. The model is pre-trained on the ImageNet dataset [51] and fine-tuned on the Oracle-50K dataset for 100 epochs using the Adam optimizer. We adopt a data augmentation technique of random rotation. The learning rate is set to 1e-3 (5e-4 for

**Table 3**

The number of parameters (#Param.), FLOPs, and inference time of baseline methods and our OBIFomer.

	DnCNN	DANet	SRMNet	KBNet	InvDN	CharFormer	Uformer	Restormer	CGNet	Ours
#Param. (M)	0.67	63.01	37.59	104.93	2.64	13.10	50.88	26.10	119.22	8.35
FLOPs (G)	44.02	32.73	285.17	14.47	95.08	51.04	89.46	140.99	62.10	20.45
Inference Time (ms)	1.94	1.19	15.73	36.60	3.04	19.92	19.23	46.34	10.88	10.35



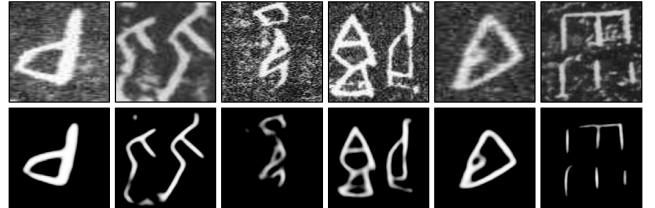
**Figure 7:** Recognition results of ResNet-18, ResNet-50, and ResNet-152 [36] on Oracle-50K dataset [13].

ResNet-152) with a batch size of 256 (128 for ResNet-50, 64 for ResNet-152). We compare the original Oracle-50K dataset with the denoising results generated by OBIFomer trained on the same dataset.

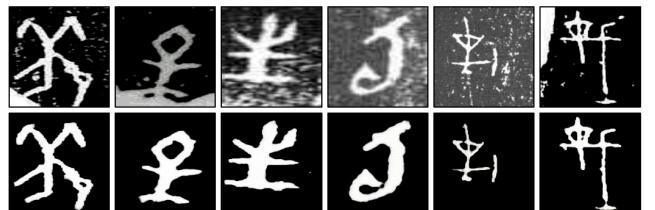
As illustrated in Fig. 7, ResNet-18 obtains a significant gain of 3.65% in the recognition accuracy on denoised images compared to the noisy images, demonstrating the effectiveness of OBI denoising in improving recognition accuracy. In addition, we observe a more remarkable improvement (4.42% and 5.19%) when deeper networks were applied, i.e., ResNet-50 and ResNet-152. This is attributed to the fact that denoised images contain more discriminative features, which are better captured by deeper networks. However, the recognition accuracy of the denoised image is still far from the ground truth.

#### 4.6. Computational efficiency

Table 3 presents comprehensive comparisons of the computational efficiency of baseline methods and our OBIFomer. We evaluate the number of parameters (#Param.), FLOPs, and inference time. To ensure a consistent comparison, #Param., FLOPs, and inference time are calculated based on a random input patch with a  $256 \times 256$  resolution. We perform 50 iterations of GPU warm-up for the inference time before the measurement. Consequently, OBIFomer has fewer FLOPs than most baseline methods. Compared to CharFormer, OBIFomer has  $3.02\times$  fewer parameters and runs  $4.76\times$  faster. When deployed on the 13th Gen Intel(R) Core(TM) i9-13900K CPU @ 3.00GHz, 32GB RAM, and an NVIDIA GeForce RTX 4090 GPU for acceleration,



**Figure 8:** Denoising results of our OBIFomer (trained on Oracle-50K dataset [13]) on OBC306 dataset [1].



**Figure 9:** Denoising results of our OBIFomer (trained on RCRN dataset [11]) on OBC306 dataset [1].

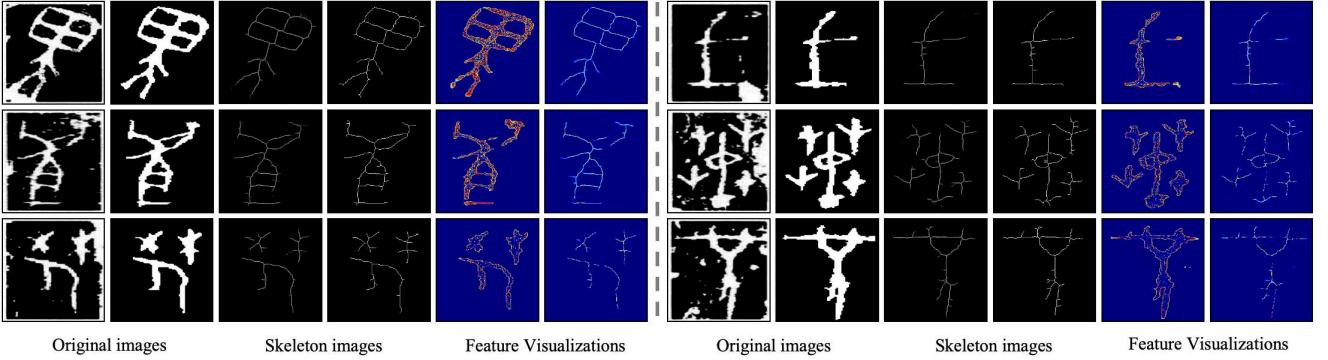
OBIFomer processes an image in just 10.35 ms, yielding competitive inference efficiency.

#### 4.7. Exploration of the Generalization Ability

To explore the generalization ability of our OBIFomer, we test it on a real oracle dataset (i.e., the OBC306 dataset) after training it on Oracle-50K and RCRN datasets. As shown in Fig. 8 and Fig. 9, OBIFomer shows a strong generalization ability on the OBC306 dataset. On the one hand, the denoising results of OBIFomer trained on the Oracle-50K dataset demonstrate the effectiveness of the adapted texture, which can be easily obtained with STSN or other OBI generation methods. On the other hand, even when we train OBIFomer on the RCRN dataset, which contains only 900 noisy-clean image pairs, it still performs well on the OBC306 dataset. Therefore, with a large amount of synthetic noisy-clean image pairs of OBIs, the generalization ability of OBIFomer can be further improved. This study demonstrates the great potential of OBIFomer in assisting automatic OBI recognition.

#### 4.8. Ablation Studies

To validate the effectiveness of our OBIFomer, we conduct ablation studies on the RCRN dataset. In our experiments, we analyze the contribution of each core component in OBIFomer and the impact of specific hyperparameters.



**Figure 10:** Visualization results of our OBIFformer on RCRN dataset [11]. For each case, the first two images are the noisy character image and its ground truth, the second two refer to the reconstructed skeleton image and its ground truth, and the last two are the visualization of reconstruction and glyph features.

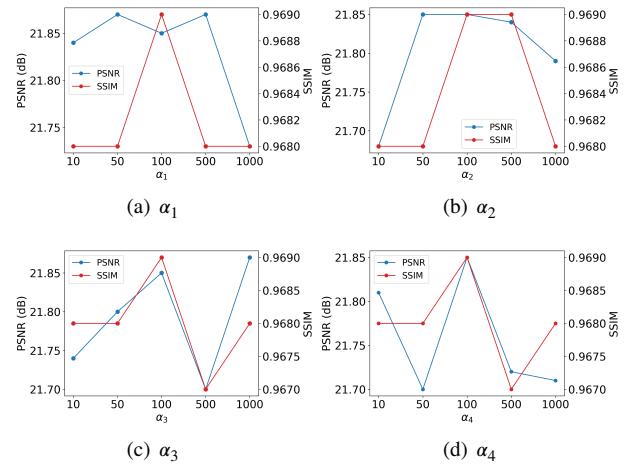
**Table 4**  
Effects of different feature fusion strategies.

Strategies	Addition	Concatenation	SKFF
PSNR↑	20.69	20.96	22.19
SSIM↑	0.962	0.967	0.969

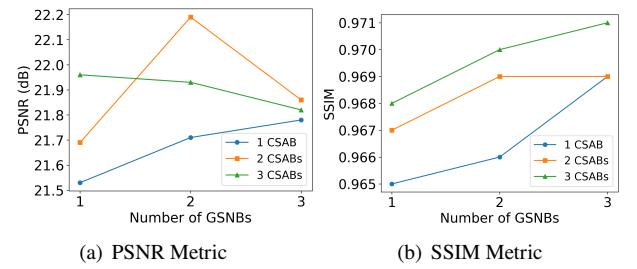
**Validation of Glyph Information Extraction.** To demonstrate that glyph structural network blocks (GSNBs) can effectively extract glyph information from input images, we visualize the output of the additional feature corrector. The results can be found in Fig. 10. For each case in Fig. 10, the first two images refer to the noisy character image and its ground truth, and the second two are the reconstructed skeleton image and its ground truth, where we can find that the glyphs are extracted properly. This visualization indicates that our OBIFformer can precisely extract glyph features from the input images.

**Evaluation of Feature Fusion Strategy.** To evaluate the effectiveness of the selective kernel feature fusion (SKFF) module, we apply simple addition and concatenation of the output of  $i$ -th CSAB and GSNB for comparison. As illustrated in Table 4, the SKFF module provides favorable gains of 1.50 dB and 1.23 dB compared to simple addition and concatenation in terms of PSNR. Similar performance gains can be observed in the SSIM metric. That's because the SKFF module can generate attention maps for the reconstruction and glyph features and aggregate them dynamically. Since the glyph features can guide the model in reconstructing the denoised image, a more significant improvement in PSNR is obtained compared to SSIM.

**Impact of the Hyperparameters.**  $\{\alpha_i\}_{i=1}^4$  are the hyperparameters for different losses in Eq. 17. To investigate the performance of OBIFformer when these parameters change, we conduct ablation experiments and show the results in Fig. 11. The models are trained for 200 epochs in all experiments for computational reasons. It can be observed that PSNR and SSIM metrics first increase and then decrease as  $\alpha_1$  and  $\alpha_2$  vary, demonstrating a desirable bell-shaped curve. In



**Figure 11:** The sensitivity of PSNR and SSIM to  $\{\alpha_i\}_{i=1}^4$ .



**Figure 12:** Effects of different CSABs and GSNBs settings.

principle, small values of  $\alpha_1$  and  $\alpha_2$  (e.g., 10) would limit the performance of the CSABs, while large values of  $\alpha_1$  and  $\alpha_2$  (e.g., 1000) would weaken the effect of GSNBs. We also observe that  $\alpha_3$  and  $\alpha_4$  lead to similar PSNR and SSIM metrics trends, indicating that glyph features can guide the model in reconstructing the denoised image.

Additionally, we adjust the number of CSABs and GSNBs to optimize the performance of our OBIFformer. Fig. 12 shows the effects of different CSABs and GSNBs settings. As the number of CSABs and GSNBs increases,

the PSNR metric initially rises but eventually declines due to overfitting as model complexity grows. In contrast, the SSIM metric continues to improve even when the model is overfitted. That can be attributed to the SKFF module, which effectively aggregates the overfitted reconstruction features and glyph features, mitigating the negative influence of excessive CSABs. Consequently, we use two CSABs and GSNBs in this paper.

#### 4.9. Visualization

To further demonstrate the effectiveness of our OBI-Former, we conduct visualization studies. As shown in Fig. 10, we visualize the deep features extracted by the final OFB, which consists of reconstruction and glyph features. For each case in Fig. 10, the last two images correspond to the visualizations of reconstruction and glyph features, respectively. For the reconstruction features, OBIFormer successfully captures character-related features while removing complex noise. For the glyph features, OBIFormer precisely learns the glyph information from the input images, aided by the GSNBs and SKFFs. These visualizations indicate that both reconstruction and glyph features are successfully learned between the OFBs.

### 5. Conclusion and Future Work

In this paper, we propose a fast attentive denoising framework for OBIs, i.e., OBIFormer. Specifically, our OBI-Former consists of an input projector, an output projector, an

additional feature corrector, and several OFBs. The OFB utilizes CSABs to extract reconstruction features and GSNBs to learn glyph information. Additionally, the SKFF module aggregates reconstruction features and glyph information dynamically. Extensive experiments demonstrate the superiority of OBIFormer on Oracle-50K and RCRN datasets quantitatively and qualitatively. Furthermore, OBIFormer shows a strong generalization ability on the OBC306 dataset, demonstrating its great potential in assisting automatic OBI recognition. Finally, we provide comprehensive ablations to validate the effectiveness of each module.

However, the performance of OBIFormer on OBIs with some types of noise, such as bone-cracked and dense white regions, remains unsatisfactory. That is because there are only a few images with such noise, which can be treated as a few-shot learning problem. Hence, one future trend is to investigate more generic methods, e.g., the conditional diffusion model, to utilize multi-modal conditions to generate a noise-balanced dataset for OBI denoising. Also, we can apply a conditional diffusion model for image denoising directly, where conditions guide the model in removing target noise.

### Acknowledgement

This work was supported by the National Social Science Foundation of China (24Z300404220) and the Shanghai Philosophy and Social Science Planning Project (2023BYY003).

## References

- [1] Shuangping Huang, Haobin Wang, Yongge Liu, Xiaosong Shi, and Lianwen Jin. Obc306: A large-scale oracle bone character recognition dataset. In *2019 International Conference on Document Analysis and Recognition (ICDAR)*, pages 681–688. IEEE, 2019.
- [2] Zhi-Kai Huang, Zhi-Hong Li, Han Huang, Zhi-Biao Li, and Ling-Ying Hou. Comparison of different image denoising algorithms for chinese calligraphy images. *Neurocomputing*, 188:102–112, 2016.
- [3] ShaoTong Gu, GeFei Feng, XiaoHu Ma, and YiMing Yang. Restoration method of characters on jiagu rubbings based on poisson distribution and fractal geometry. *Science China Information Sciences*, 53:1296–1304, 2010.
- [4] Krisda Khankasikam. Restoration of degraded historical document image: An adaptive multilayer-information binarization technique. *J. Inf. Sci. Eng.*, 30(5):1321–1338, 2014.
- [5] Mehdi Bahri, Yannis Panagakis, and Stefanos Zafeiriou. Robust kronecker-decomposable component analysis for low-rank modeling. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3352–3361, 2017.
- [6] Mehdi Bahri, Yannis Panagakis, and Stefanos Zafeiriou. Robust kronecker component analysis. *IEEE transactions on pattern analysis and machine intelligence*, 41(10):2365–2379, 2018.
- [7] Lei Zhang and Cong Liu. Kronecker component with robust low-rank dictionary for image denoising. *Displays*, 74:102194, 2022.
- [8] Lei Zhang and Cong Liu. Robust low-rank analysis with adaptive weighted tensor for image denoising. *Displays*, 73:102200, 2022.
- [9] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE transactions on image processing*, 26(7):3142–3155, 2017.
- [10] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5728–5739, 2022.
- [11] Daqian Shi, Xiaolei Diao, Hao Tang, Xiaomin Li, Hao Xing, and Hao Xu. Rcrn: Real-world character image restoration network via skeleton extraction. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 1177–1185, 2022.
- [12] Daqian Shi, Xiaolei Diao, Lida Shi, Hao Tang, Yang Chi, Chuntao Li, and Hao Xu. Charformer: A glyph fusion based attentive framework for high-precision character image denoising. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 1147–1155, 2022.
- [13] Wenhui Han, Xinlin Ren, Hangyu Lin, Yanwei Fu, and Xiangyang Xue. Self-supervised learning of orc-bert augmentator for recognizing few-shot oracle characters. In *Proceedings of the Asian Conference on Computer Vision*, 2020.
- [14] Mei Wang, Weihong Deng, and Cheng-Lin Liu. Unsupervised structure-texture separation network for oracle character recognition. *IEEE Transactions on Image Processing*, 31:3137–3150, 2022.
- [15] Zijian Chen, Tingzhu Chen, Wenjun Zhang, and Guangtao Zhai. Obi-bench: Can lmms aid in study of ancient script on oracle bones?, 2025.
- [16] Bang Li, Qianwen Dai, Feng Gao, Weiye Zhu, Qiang Li, and Yongge Liu. Hwobc-a handwriting oracle bone character recognition database. In *Journal of Physics: Conference Series*, volume 1651, page 012050. IOP Publishing, 2020.
- [17] Mengru Wang, Yu Cai, Li Gao, Ruichen Feng, Qingju Jiao, Xiaolin Ma, and Yu Jia. Study on the evolution of chinese characters based on few-shot learning: From oracle bone inscriptions to regular script. *Plos one*, 17(8):e0272974, 2022.
- [18] Xuebin Yue, Hengyi Li, Yoshiyuki Fujikawa, and Lin Meng. Dynamic dataset augmentation for deep learning-based oracle bone inscriptions recognition. *ACM Journal on Computing and Cultural Heritage*, 15(4):1–20, 2022.
- [19] Haisu Guan, Jinpeng Wan, Yuliang Liu, Pengjie Wang, Kaile Zhang, Zhebin Kuang, Xinyu Wang, Xiang Bai, and Lianwen Jin. An open dataset for the evolution of oracle bone characters: Evobc. *arXiv preprint arXiv:2401.12467*, 2024.
- [20] Jun Guo, Changhu Wang, Edgar Roman-Rangel, Hongyang Chao, and Yong Rui. Building hierarchical representations for oracle character and sketch recognition. *IEEE Transactions on Image Processing*, 25(1):104–118, 2015.
- [21] Zhen Yang, Qiqi Wang, Xiuying He, Yang Liu, Fan Yang, Zhijian Yin, and Chen Yao. Accurate oracle classification based on deep convolutional neural network. In *2018 IEEE 18th International Conference on Communication Technology (ICCT)*, pages 1188–1191. IEEE, 2018.
- [22] Bingxin Du, Guoying Liu, and Wenying Ge. Deep self-supervised learning for oracle bone inscriptions features representation. In *2021 IEEE 4th International Conference on Information Systems and Computer Aided Education (ICISCAE)*, pages 7–11. IEEE, 2021.
- [23] Wang Huihui. Large-scale oracle bone inscriptions dataset construction and algorithm research. Master's thesis, Henan University, 2020.
- [24] Mengting Liu, Guoying Liu, Yongge Liu, and Qingju Jiao. Oracle bone inscriptions recognition based on deep convolutional neural network. *Journal of image and graphics*, 8(4):114–119, 2020.
- [25] Chongsheng Zhang, Ruixing Zong, Shuang Cao, Yi Men, and Bofeng Mo. Ai-powered oracle bone inscriptions recognition and fragments rejoining. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, pages 5309–5311, 2021.
- [26] Yoshiyuki Fujikawa, Hengyi Li, Xuebin Yue, CV Aravinda, G Amar Prabhu, and Lin Meng. Recognition of oracle bone inscriptions by using two deep learning models. *International Journal of Digital Humanities*, 5(2):65–79, 2023.
- [27] Chongsheng Zhang, Bin Wang, Ke Chen, Ruixing Zong, Bo-feng Mo, Yi Men, George Almanidis, Shanxiong Chen, and Xiangliang Zhang. Data-driven oracle bone rejoining: A dataset and practical self-supervised learning scheme. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 4482–4492, 2022.
- [28] Mei Wang and Weihong Deng. A dataset of oracle characters for benchmarking machine learning algorithms. *Scientific Data*, 11(1):87, 2024.
- [29] Pengjie Wang, Kaile Zhang, Yuliang Liu, Jinpeng Wan, Haisu Guan, Zhebin Kuang, Xinyu Wang, Lianwen Jin, and Xiang Bai. An open dataset for oracle bone script recognition and decipherment. *arXiv preprint arXiv:2401.15365*, 2024.
- [30] Jinhao Li, Zijian Chen, Runze Jiang, Tingzhu Chen, Changbo Wang, and Guangtao Zhai. Mitigating long-tail distribution in oracle bone inscriptions: Dataset, model, and benchmark, 2025.
- [31] Maozuo Pu. Oracle bone inscriptions in the collection of shanghai museum (volume i), 2009.
- [32] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
- [33] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.
- [34] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 31, 2017.
- [35] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [36] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [37] Yi-Kang Zhang, Heng Zhang, Yong-Ge Liu, Qing Yang, and Cheng-Lin Liu. Oracle character recognition by nearest neighbor classification with deep metric learning. In *2019 International Conference on Document Analysis and Recognition (ICDAR)*, pages 309–314. IEEE, 2019.

- [38] Mei Wang, Weihong Deng, and Sen Su. Oracle character recognition using unsupervised discriminative consistency network. *Pattern Recognition*, 148:110180, 2024.
- [39] Oscar Valbuena Prada, Miguel Ángel Vera, Guillermo Ramirez, Ricardo Barrientos Rojel, and David Mojica Maldonado. Statistical techniques for digital pre-processing of computed tomography medical images: A current review. *Displays*, 85:102835, 2024.
- [40] Kecheng Chen, Kun Long, Yazhou Ren, Jiayu Sun, and Xiaorong Pu. Lesion-inspired denoising network: Connecting medical image denoising and lesion detection. In *Proceedings of the 29th ACM International Conference on Multimedia*, pages 3283–3292, 2021.
- [41] Boyu Liu, Lingda Wu, Xiaorui Song, Hongxing Hao, Ling Zou, and Yu Lu. Indeandcoe: A framework based on multi-scale feature fusion and residual learning for interferometric sar remote sensing image denoising and coherence estimation. *Displays*, 79:102496, 2023.
- [42] Chi-Mao Fan, Tsung-Jung Liu, Kuan-Hsien Liu, and Ching-Hsiang Chiu. Selective residual m-net for real image denoising. In *2022 30th European Signal Processing Conference (EUSIPCO)*, pages 469–473. IEEE, 2022.
- [43] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5–9, 2015, proceedings, part III* 18, pages 234–241. Springer, 2015.
- [44] Yi Zhang, Dasong Li, Xiaoyu Shi, Dailan He, Kangning Song, Xiaogang Wang, Hongwei Qin, and Hongsheng Li. Kbnet: Kernel basis network for image restoration. *arXiv preprint arXiv:2303.02881*, 2023.
- [45] Yang Liu, Zhenyue Qin, Saeed Anwar, Pan Ji, Dongwoo Kim, Sabrina Caldwell, and Tom Gedeon. Invertible denoising network: A light solution for real noise removal. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 13365–13374, 2021.
- [46] Zongsheng Yue, Qian Zhao, Lei Zhang, and Deyu Meng. Dual adversarial network: Toward real-world noise removal and noise generation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part X 16*, pages 41–58. Springer, 2020.
- [47] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021.
- [48] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 17683–17693, 2022.
- [49] Amirhosein Ghaseabadi, Muhammad Kamran Janjua, Mohammad Salameh, Chunhua Zhou, Fengyu Sun, and Di Niu. Cascadedgaze: Efficiency in global context extraction for image restoration. *Transactions on Machine Learning Research*, 2024.
- [50] Xiang Li, Wenhui Wang, Xiaolin Hu, and Jian Yang. Selective kernel networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 510–519, 2019.
- [51] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.
- [52] W Jian-ping, Q Zi-tuo, W Jin-ling, and L Guo-jun. Chinese characters stroke thinning and extraction based on mathematical morphology [j]. *Journal of Hefei University of Technology (Natural Science)*, 11:017, 2005.
- [53] Jianhua Shen and Jinyan Cao. *Jiaguwen zixing biao (a list of oracle characters)*. Shanghai, China: Shanghai cishuchubanshe, 2008.
- [54] I Loshchilov. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.
- [55] Zicheng Zhang, Yingjie Zhou, Chunyi Li, Baixuan Zhao, Xiaohong Liu, and Guangtao Zhai. Quality assessment in the era of large models: A survey. *arXiv preprint arXiv:2409.00031*, 2024.
- [56] Zicheng Zhang, Haoning Wu, Chunyi Li, Yingjie Zhou, Wei Sun, Xiongkuo Min, Zijian Chen, Xiaohong Liu, Weisi Lin, and Guangtao Zhai. A-bench: Are lmms masters at evaluating ai-generated images? *arXiv preprint arXiv:2406.03070*, 2024.
- [57] Zicheng Zhang, Ziheng Jia, Haoning Wu, Chunyi Li, Zijian Chen, Yingjie Zhou, Wei Sun, Xiaohong Liu, Xiongkuo Min, Weisi Lin, et al. Q-bench-video: Benchmarking the video quality understanding of lmms. *arXiv preprint arXiv:2409.20063*, 2024.
- [58] Zijian Chen, Wei Sun, Haoning Wu, Zicheng Zhang, Jun Jia, Ru Huang, Xiongkuo Min, Guangtao Zhai, and Wenjun Zhang. Study of subjective and objective naturalness assessment of ai-generated images. *IEEE Transactions on Circuits and Systems for Video Technology*, 2024.
- [59] Alain Hore and Djemel Ziou. Image quality metrics: Psnr vs. ssim. In *2010 20th international conference on pattern recognition*, pages 2366–2369. IEEE, 2010.