High-Resolution Natural Image Matting by Refining Low-Resolution Alpha Mattes

Xianmin Ye, Yihui Liang, Mian Tan, Fujian Feng, Lin Wang, and Han Huang, *Senior Member, IEEE*

***Abstract*—High-resolution natural image matting plays an important role in image editing, film-making and remote sensing due to its ability of accurately extract the foreground from a natural background. However, due to the complexity brought about by the proliferation of resolution, the existing image matting methods cannot obtain high-quality alpha mattes on high-resolution images in reasonable time. To overcome this challenge, we introduce a high-resolution image matting framework based on alpha matte refinement from low-resolution to high-resolution (HRIMF-AMR). The proposed framework transforms the complex high-resolution image matting problem into low-resolution image matting problem and high-resolution alpha matte refinement problem. While the first problem is solved by adopting an existing image matting method, the latter is addressed by applying the Detail Difference Feature Extractor (DDFE) designed as a part of our work. The DDFE extracts detail difference features from high-resolution images by measuring the image feature difference between high-resolution images and low-resolution images. The low-resolution alpha matte is refined according to the extracted detail difference feature, providing the high-resolution alpha matte. In addition, the Matte Detail Resolution Difference (MDRD) loss is introduced to train the DDFE, which imposes an additional constraint on the extraction of detail difference features with mattes. Experimental results show that integrating HRIMF-AMR significantly enhances the performance of existing matting methods on high-resolution images of Transparent-460 and Alphamatting. Project page: https://github.com/yexianmin/HRAMR-Matting**

***Index Terms*—High-resolution image matting, natural image matting, alpha matte detail, detail difference feature.**

Received 29 July 2024; revised 26 February 2025 and 3 May 2025; accepted 18 May 2025. Date of publication 2 June 2025; date of current version 9 June 2025. This work was supported in part by the National Natural Science Foundation of China under Grant 62276103 and Grant 62271130; in part by the Science and Technology Projects of Guizhou Province under Grant QKHJCZK2023YB143 and Grant QKHJCZK2022YB197; in part by Guangdong Basic and Applied Basic Research Foundation under Grant 2023A1515010066; in part by Zhongshan Science and Technology Research Project of Social Welfare under Grant 2021B2006; in part by the Innovation Team Project of General Colleges and Universities in Guangdong Province under Grant 2023KCXTD002; in part by the Analysis of Urban Events Based on Low-Altitude Drones under Grant 2024BQ010011; in part by the Fundamental Research Funds for the Central Universities under Grant 93K172024K24; in part by the Youth Science and Technology Talent Growth Project of Guizhou Province under Grant QJJ2024273; in part by the Natural Science Research Project of Education Department of Guizhou Province under Grant QJJ2023061 and Grant QJJ2022047; and in part by the Science and Technology Innovation Talent Team Project of Data Science and Computing Intelligence of Guizhou Province under Grant QKHRC-CXTD2025038. The associate editor coordinating the review of this article and approving it for publication was Dr. Nikos Deligiannis. *(Xianmin Ye and Yihui Liang are co-first authors.) (Corresponding authors: Yihui Liang; Mian Tan.)*

Xianmin Ye, Mian Tan, Fujian Feng, and Lin Wang are with the College of Data Science and Information Engineering, Guizhou Minzu University, Guiyang 550025, China, and also with Guizhou Key Laboratory of Pattern Recognition and Intelligent System, Guizhou Minzu University, Guiyang 550025, China (e-mail: yexianmin_gz@outlook.com; tanmian@gzmu.edu.cn; fujian_feng@gzmu.edu.cn; wanglin@gzmu.edu.cn).

Yihui Liang is with the School of Computer Science, Zhongshan Institute, University of Electronic Science and Technology of China, Zhongshan 528400, China (e-mail: yihuiliang@outlook.com).

Han Huang is with the School of Software Engineering, South China University of Technology, Guangzhou 510006, China, also with the Key Laboratory of Big Data and Intelligent Robot (SCUT), MOE of China, Guangzhou 510006, China, also with Guangdong Engineering Center for Large Model and GenAI Technology, Guangzhou 510006, China, and also with the Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun 130012, China (e-mail: hhan@scut.edu.cn).

Digital Object Identifier 10.1109/TIP.2025.3573620

I. Introduction

HIGH-resolution natural image matting is the process of precisely extracting the foreground from the background in a high-resolution image by accurately determining the opacity of the foreground. It is essential in several key applications, such as image editing [1], film-making [2], remote sensing [3] and autonomous driving [4], [5]. High-resolution image matting methods provide highly accurate foregrounds extracted from natural images, which can be used to synthesize realistic images and videos. In addition, it plays a key role in remote sensing, contributing to the accurate analysis of remote sensing images with clouds. In image matting problem, the color I_i of the pixel i is modeled as a convex combination of foreground color F_i and background color B_i:

$$I_i = \alpha_i F_i + (1 - \alpha_i) B_i, i \in \{1, 2, \cdots, N\} \tag{1}$$

where $\alpha_i \in [0, 1]$, represents the opacity of the foreground object at pixel i. N is the number of pixels in the image. The value of N can reach 10^6 for high-resolution (such as 2K resolution) images. As shown in Fig. 1, the high-resolution image depicted in Fig. 1(b) provides more details than the low-resolution image presented in Fig. 1(c). At higher resolutions, even small errors become very noticeable, yet this increased detail poses a challenge for the matting algorithm. Therefore, as the number of pixels increases, accurately determining the alpha value of each pixel within reasonable time becomes crucial for maintaining the image matting efficiency and quality. As there are three unknowns and only one available value in Eq. (1), the image matting problem is inherently ill-posed. Accordingly, image matting methods rely on a trimap to distinguish known foreground, known background, and unknown regions, allowing only the alpha values in the unknown region to be solved.

(a) Original image (b) High-res image (c) Low-res image (d) MGM-trimap[6] (e) Matteformer[7] (f) Ours

Fig. 1. The visual comparison of high-resolution image matting results on the Alphamatting [8] high-resolution dataset. The original image is shown in figures (a), with low-resolution and high-resolution enlargements of the original image provided in (b) and (c), respectively. Figures (d) and (e) show close-ups of the visual results of MGM-trimap [6] and Matteformer [7] on high-resolution images, respectively. Figure (f) shows a close-up of the visual results of our method on high-resolution images.

Image matting can be divided into traditional methods [9], [10] and deep learning-based methods [11], [12], [13], [14]. Traditional methods include propagation-based methods [9], [15], [16], [17] and optimization-based [10], [18], [19], [20], [21] methods. Propagation-based methods typically involve a pixel-by-pixel analysis to ascertain the degree of similarity. They analyze pixel similarity using graph models with Laplace matrices and optimization techniques [22]. They are impractical for use on high-resolution images, as the sheer volume of pixels results in a prohibitively large number of comparisons that need to be made. Although optimization-based methods treat matting as a pixel-pair optimization problem, these methods also encounter significant challenges when dealing with high-resolution images because the complexity and size of the search space increases exponentially with the resolution. To address these challenges and reduce the processing time and resources, researchers have employed swarm optimization techniques [20] and micro-scale searching algorithms [21]. While capable of navigating complex solution spaces, these methods may struggle with the high-dimensional nonlinear nature of the matting problem and may converge to a local optimal solution rather than a global solution. In addition, parameter tuning of these algorithms can be complex, and the computational costs associated with pixel affinity for high-resolution images can still be prohibitive.

Deep learning-based methods have advanced image matting significantly. Deep Image Matting (DIM) [11] introduced CNNs for alpha matte estimation, using an encoder-decoder architecture with refinement modules, and created a large-scale dataset for training. Researchers have proposed techniques like attention mechanisms [23], adaptive upsampling [12], multi-branch information mining [14], and progressive refinement [6] to address the limitations of CNNs in capturing long-range dependencies and to enhance the details of alpha mattes. Other researchers have explored computational efficiency in high-resolution image matting by introducing techniques such as image patch [2], [13] and sparse maps [24] to reduce computational costs. However, limited by the receptive field of the CNN, they are still not ideal in preserving fine details of images. Recent studies have used self-attention mechanisms [25], such as Matteformer [7], ELGT-Matting [26], ViTMatte [27] and DiffMatte [28], to model global context and preserve complex structures, providing a promising direction to overcome these limitations. However, such methods still have difficulty in preserving fine details in complex semi-transparent regions when applied to high-resolution images.

None of the aforementioned methods can accurately extract foreground details from high-resolution images in a timeframe acceptable for most practical applications. Therefore, even with advanced methods like Matteformer [7] or MGM [6], detail on high-resolution images is not adequately captured, as evident from Fig. 1(d) and 1(e).

To overcome these shortcomings, we introduce a high-resolution image matting framework based on alpha matte refinement from low-resolution to high-resolution (HRIMF-AMR), allowing us to transform the complex high-resolution matting problem into a low-resolution matting problem and a high-resolution alpha matte refinement problem. To achieve this objective, the HRIMF-AMR employs a Detail Difference Feature Extractor (DDFE) module to extract fine details unique to high-resolution images, which are then used to refine the corresponding low-resolution alpha matte. Additionally, a Matte Detail Resolution Difference (MDRD) loss function is incorporated to improve the extraction of these resolution-specific details, ensuring that the detailed differences between high-resolution and low-resolution images are effectively captured within the alpha matte.

The main contributions of this work are summarized below:

- We present a high-resolution image matting framework based on alpha matte refinement from low-resolution to high-resolution, HRIMF-AMR, which transforms the complex problem of high-resolution image matting into two simpler problems: the low-resolution image matting problem and the high-resolution alpha matte refinement problem.
- We present Detail Difference Feature Extractor (DDFE) and the Matte Detail Resolution Difference (MDRD) loss function for solving the high-resolution matte refinement problem. DDFE provides foreground details for high-resolution matte refinement by capturing the subtle detail differences between high-resolution and low-resolution images. MDRD introduces additional constraints to the feature extraction to ensure that the extracted features can represent the detail difference along with the matte.
- Extensive experimental results demonstrate that integrating the presented HRIMF-AMR significantly improves the performance of existing image matting methods on

high-resolution images. This integration enables existing image matting methods to achieve state-of-the-art performance and makes them applicable to high-resolution image matting tasks.

II. Related Work

This section provides a brief literature review, focusing on image matting methods, which are categorized according to the underlying principle, i.e., (1) traditional matting methods that rely on statistical algorithms and (2) deep learning-based matting methods that rely on deep neural networks.

A. Traditional Methods

As one of the traditional methods, the propagation-based method [9], [15], [16], [17] propagates alpha values from known regions to unknown regions by measuring similarity between unknown pixels and known foreground and background pixels. On the other hand, the optimization-based method [10], [18], [19] models the image matting problem as a pixel-pair optimization problem and estimates the alpha value by solving the optimization problem for each unknown pixel. Closed-form matting [9] is based on the assumption of local smoothing of foreground and background colors, allowing an alpha-based Color-Line model to be established. Spectral matting [15] introduces spectral clustering dependent on a properly defined Laplacian matrix, which is based on Closed-form matting [9]. KNN matting [16] uses color similarity and spatial proximity to solve the image problem globally for K nearest neighbors, which speeds up interference while maintaining accuracy. Information flow matting [17] sets multiple information channels and adjusts their respective propagation modes to accurately generate alpha mattes. Huang et al. [29] proposed pixel-level discrete multi-objective sampling (PDMS) method, which effectively solves the problems of incomplete sample space and multi-sampling standard conflict by formalizing the color sampling process as a multi-objective optimization problem (MOP). The main advantage of PDMS stems from its ability to minimize color differences and spatial distances between unknown pixels and known pixels, along with its capacity to make adaptive tradeoffs between conflicting sampling criteria. However, extensive computations are required to evaluate and select the best foreground and background color sample pairs, resulting in high computational complexity.

In order to reduce the computational burden of the matting algorithm, some interesting schemes [20], [21], [30], [31] have been proposed. For example, Feng et al. [20] developed an innovative group competition optimization algorithm that capitalizes on the color similarity of pixels in unknown regions to cluster them effectively. This algorithm fosters group cooperation to streamline the optimization process, thereby significantly diminishing the computational complexity for high-resolution image matting. Still, as the algorithm necessitates a substantial number of iterations to attain optimal matting outcomes, its practical utility is relatively low. MS-AM [21] incorporates micro-search and solves the problem of high-resolution image matting by effective decision subset exploration. While this method reduces the search space and improves the processing efficiency in theory, to ensure the best performance in practical applications, parameters must be carefully adjusted according to the specific features of the image. As a proxy model-based matting method, IMBSM [30] estimates high-quality alpha mattes in reasonable time and outperforms traditional pixel-pair optimization techniques. MCSS [31] relies on a multi-criteria sampling strategy that, when combined with the Gaussian process proxy model, effectively improves the matting quality under resource constraints.

As demonstrated above, even when strategies to alleviate the pressure on computational resources are incorporated, traditional image matting methods are still incapable of dealing with high-resolution images. These methods often rely on complex similarity measures and precise operations at the pixel level, which are insufficiently flexible and efficient and thus cannot be applied to the extensive data volumes associated with high-resolution images. In addition, they may struggle to accurately distinguish between foreground and background in detail-rich regions, resulting in unsatisfactory matting results. These disadvantages are particularly prominent in the field of modern image processing, especially in scenes where a large number of high-resolution images need to be processed quickly and accurately.

B. Deep Learning-Based Methods

In recent years, deep learning has produced extremely promising results for the image matting task. For instance, Deep Image Matting (DIM) [11] introduced large-scale Adobe Composite-1K dataset and an end-to-end encoder-decoder network, marking a milestone in the field. However, DIM [11] is limited by its receptive field in natural image matting, as its convolutional neural network (CNN) architecture struggles to capture long-range context, affecting the accurate estimation of alpha values, particularly in regions with fine details like hair or transparency.

To address the limitations of CNNs in capturing long-range dependencies and fine details, several methods have been proposed. GCA-Matting [23] incorporates a guided contextual attention module that learns low-level affinities and propagates high-level opacity information globally, enhancing detail preservation. HDMatt [13] leverages the Cross-Patch Contextual Module (CPC) to capture long-distance context dependencies between image patches. However, this method may unintentionally lose details due to inconsistencies between local and global contexts caused by chunking. A^2U [12] improves the handling of fine structures by learning affinity during upsampling, exploiting pairwise interactions in images. MGM [6] introduces a Progressive Refinement Network (PRN) that refines details progressively through a self-guided decoding process. A two-stage framework addresses trimap dependency and model complexity: the Segmentation Network (SN) captures semantics and classifies pixels into unknown, foreground, and background regions, while the Matting Refine Network (MRN) captures detailed texture information and regresses accurate alpha values [32]. However, without trimap, it cannot achieve the same level of detail preservation as trimap-based methods. In addition, MODNet [33] obtains a real-time portrait matting model

by decomposing the trimap-free portrait matting task into three explicit sub-goals, namely semantic estimation, detail prediction and semantic detail fusion. However, the detail prediction part is limited by using the tensor concatenated from the downsampled image and the intermediate layer features of semantic estimation as input, which may struggle with complex fine details on high-resolution images. BGMV2 [2] replaces the trimap with a background image as input, providing prior information for matting. While effective for static backgrounds, it struggles with dynamic backgrounds, often losing fine details. SparseMat [24] reduces spatial complexity by using sparse maps, skipping regions where the foreground is already determined. However, this approach compromises the ability to handle abrupt changes in discontinuous areas, leading to detail loss in complex regions.

Recent advancements have shifted toward self-attention mechanisms to better capture long-range dependencies and fine details. Matteformer [7] introduces Swin-Transformer [34] into natural image matting, presenting trimap-based prior tokens for improved detail preservation and context modeling. ELGT-Matting [26] addresses the limited receptive fields of CNNs with a local-global transformer block, combining global context learning and local feature integration through window-level global MSA and local-global window MSA modules. ViTMatte [27] leverages Vision Transformers [35] to capture intricate details and long-range dependencies, achieving outstanding performance due to the pre-trained, semantically rich representations of Vision Transformers [35]. A domain alignment module with a dynamic attention pruning mechanism based on transformers is proposed in references [25], designed to locate domain-sensitive regions and enable robust performance on both synthetic and natural images. However, it often struggles with insufficient user control and tends to lose fine details, particularly in complex areas like transparency and hair objects [6]. DiffMatte [28] introduces a diffusion model [36] that iteratively refines the alpha matte through a pixel-level denoising process, addressing the limitations of one-step prediction methods in complex cases. By adopting a self-attention-based backbone network, DiffMatte [28] enhances its ability to capture long-range dependencies and refine details. However, the large number of pixels in high-resolution images reduces sampling efficiency, impacting computational performance and fine detail preservation.

Current image matting methods often fail to preserve fine details like hair and transparency, especially in high-resolution scenarios, resulting in blurred outputs or loss of intricate structures. While self-attention mechanisms have improved detail retention, they still struggle with discontinuous edges and transparent regions. A more effective approach is needed to enhance matting quality for high-resolution images.

III. Methodology

We begin this section with a brief description of the approach adopted in this work and overall network structure, followed by a detailed description of the Detail Difference Feature Extractor (DDFE) and the Matte Detail Resolution Difference (MDRD) loss function incorporated into our high-resolution matting design.

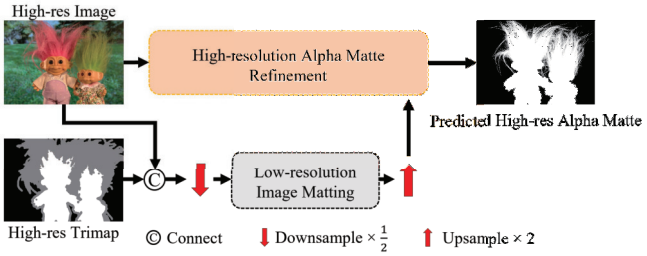


Fig. 2. Pipeline of the HRIMF-AMR and the relationship between the two problems transformed by the proposed approach.

A. Streamlined Approach to High-Resolution Matting

As shown in Fig. 2, our HRIMF-AMR employs an innovative two-step process to deliver high-quality high-resolution alpha mattes. Initially, we reduce the computational load by scaling down the image resolution, which simultaneously generates a preliminary alpha matte. This step streamlines the image matting process without compromising the final alpha matte quality. Subsequently, we refine the low-resolution matte with the intricate details of the high-resolution image. By integrating these details through fusion, we significantly enhance the clarity and precision of the high-resolution matte. By adopting this strategy, the HRIMF-AMR described here not only ensures the high quality of the high-resolution alpha matte but also circumvents the high computational costs associated with high-resolution image matting, achieving a dual optimization of efficiency and quality.

B. High-Resolution Natural Image Matting Framework

Fig. 3 shows our proposed high-resolution image matting framework based on alpha matte refinement from low-resolution to high-resolution (HRIMF-AMR), which is an innovative framework for high-resolution image matting. The main advantage of the HRIMF-AMR stems from the application of transformation and conquest, which transforms the complex high-resolution image matting problem into two relatively simple problems: the low-resolution image matting problem and the high-resolution alpha matte refinement problem. First, we concatenate a high-resolution RGB image $I_h \in R^{H\times W\times 3}$ and its corresponding trimap $T_h \in R^{H\times W\times 1}$ by channel and input them to the low-resolution image matting branch of HRIMF-AMR. The high-resolution images and trimap are downsampled by a ratio of $\frac{1}{2}$ and then input into the low-resolution image matting network ($I_l \in R^{\frac{1}{2}H\times\frac{1}{2}W\times 3}$, $T_l \in R^{\frac{1}{2}H\times\frac{1}{2}W\times 1}$). Matteformer [7] is a competitive matting method that accurately extracts alpha mattes from images and is used in the low-resolution image matting branch of HRIMF-AMR. The resolution of the image input to the minutiae Difference Feature Extractor (DDFE) is the original resolution. The DDFE consists of a series of convolutional layers. This module contrasts details derived from high-resolution images with those obtained from low-resolution images with the aim of capturing only the details present in high-resolution images. Rather than relying on a special design, we simply use fusion of the matte detail difference feature with the upsampled low-resolution alpha matte to predict the final

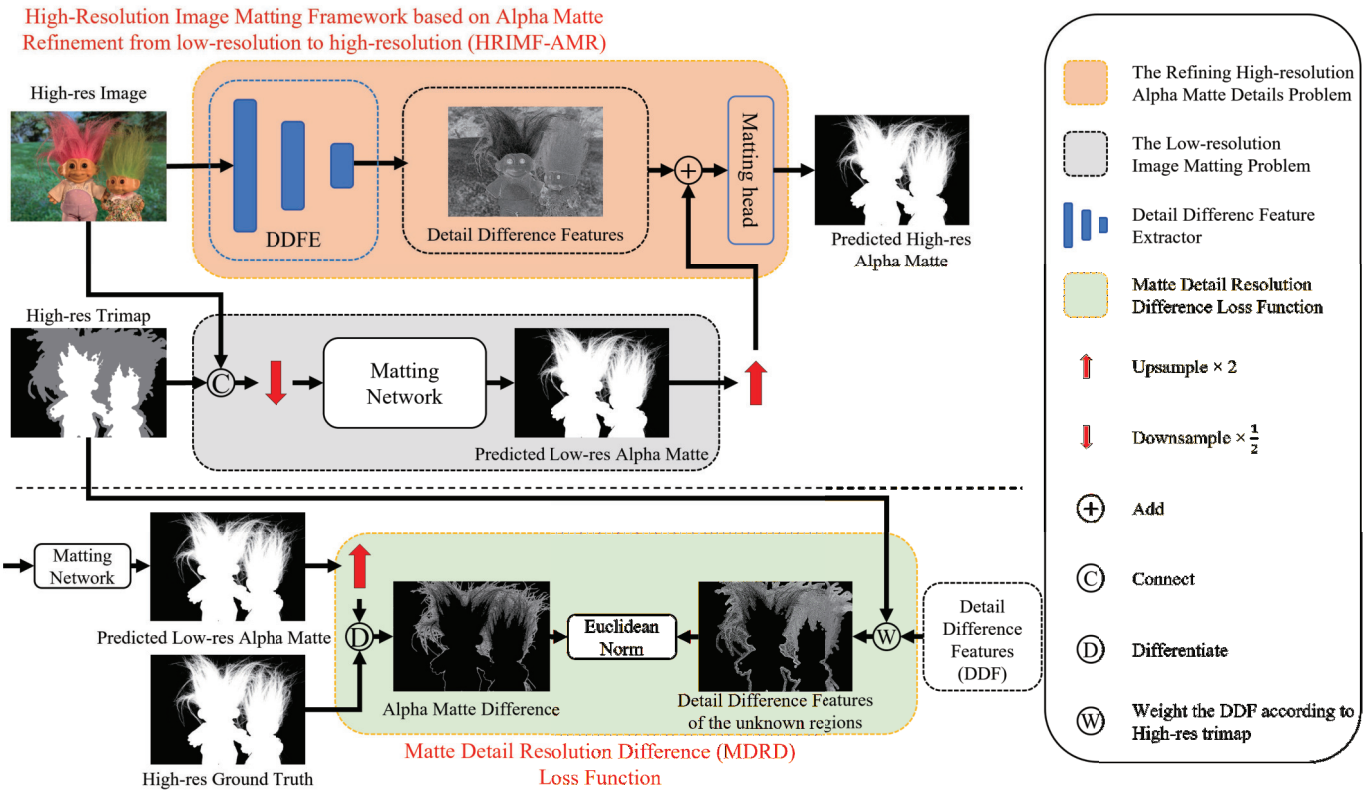


Fig. 3. An overview of the proposed high-resolution image matting framework based on alpha matte refinement from low-resolution to high-resolution (HRIMF-AMR). The framework overcomes the challenge of high-resolution matting by transforming it into two simple problems: high-resolution detail refinement and low-resolution matting. Detail Difference Feature Extractor (DDFE) is proposed to solve the high-resolution detail refinement problem. By incorporating the Matte Detail Resolution Difference (MDRD) loss function into the DDFE training, additional guidance for the extraction of features that reflect the subtleties in matte detail resolution is provided.

alpha matte $\alpha_h \in R^{H\times W\times 1}$. At the fusion stage, the detail difference features corresponding to the unknown regions of the trimap are incorporated into the upsampled low-resolution alpha mattes through an addition operation.

It should be noted that our HRIMF-AMR allows us to easily integrate and replace various natural image matting networks. Furthermore, the modularity and plug-and-play nature of HRIMF-AMR allows each component to be explored and optimized independently.

C. Detail Difference Feature Extractor

As high-resolution images comprise a significant number of pixels, they capture a greater level of detail compared to low-resolution images, allowing for finer local variations and texture information. This enhanced detail is crucial for high-resolution image matting, as it provides a wealth of visual cues that can significantly improve the matting process accuracy. Thus, by leveraging the inherent differences between the detail features present in high-resolution and low-resolution images our objective is to identify and isolate the features that are distinct in high resolution but may be less pronounced or even lost in low resolution. This preliminary idea is theoretical and involves the use of subtraction to highlight these differences. By focusing on these differences, the aim is to develop a difference feature representation that encapsulates only the high-resolution image details. This representation forms the basis for our image matting algorithms, ensuring that they concentrate on the most relevant aspects of the high-resolution alpha matte. The ultimate goal is to improve the precision and effectiveness of the matting process by leveraging the full potential of high-resolution imagery for superior matting results.

As depicted in Fig. 4, the DDFE is at the heart of our HRIMF-AMR, allowing the differences in detail between high-resolution and low-resolution images to be extracted, while the downsampling module quickly reduces the dimensionality of high-resolution images for subsequent matting network stages. Given the need to minimize the computational complexity in high-resolution natural image matting, we have selected the nearest neighbor method [37] for downsampling. However, from an information theory standpoint, the significant reduction of pixels in high-resolution images can lead to the loss of fine details that, while seemingly insignificant, are crucial for accurate matting. To address this limitation, the DDFE is specifically engineered to capture these subtle details and transform them into detail difference features. These features are then integrated with low-resolution mattes from the matting network to reconstruct more precise high-resolution mattes.

Calculation of the detail difference features involves three steps. First, we apply a simple convolutional neural network layer to the high-resolution image, aiming to capture

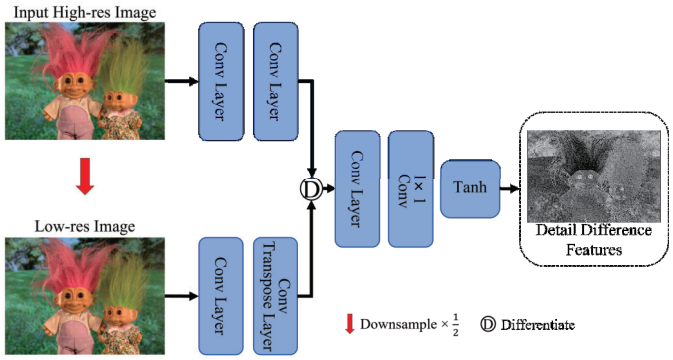


Fig. 4. The detail difference feature extractor structure. A simple convolutional neural network layer is applied to the high-resolution image as well as the low-resolution image, aiming to capture detail features in the image using simple structures. DDFE aligns the detail features of the low-resolution image to those of the high-resolution image, and extracts the detail features unique to the high-resolution image by capturing the differences between the features extracted from the high-resolution and the low-resolution images.

its fine detail features using simple structures. Next, we downsample the high-resolution image to generate a low-resolution image and extract its features. Finally, by comparing the features extracted from the high-resolution image and the low-resolution image, we extract the unique details that appear only in the high-resolution image. These steps are accomplished using the following formulas:

$$\begin{aligned} HDF &= Conv(I_h), \quad I_h \in R^{H\times W\times 3} \\ LDF &= Conv_T(I_l), \quad I_l \in R^{\frac{1}{2}H\times\frac{1}{2}W\times 3} \\ DDF &= Conv(HDF - LDF) \\ DDF_{offset} &= Tanh(Conv_{1\times1}(DDF)) \end{aligned} \tag{2}$$

where HDF and LDF respectively represent the detail features of the high-resolution and the low-resolution image, and $Conv$, $Conv_{1\times1}$, and $Conv_T$ are simple convolutional layers, each playing a specific role. Namely, $Conv$ is utilized for feature extraction, $Conv_{1\times1}$ is required for dimensionality reduction, and $Conv_T$ is adopted in the transposition convolutional layer. In order to restrict the eigenvalues to a normalized range, the calculated result is normalized by applying the $Tanh$ function. This step ensures that the DDF ranges from -1 to 1, resulting in an offset for alpha matte refinement, which is necessary to correct the upsampled results of the low-resolution alpha matte.

The matting head fuses the Detail Difference Features (DDF) obtained by the Detail Difference Features Extractor (DDFE) and the low-resolution alpha matte obtained by the low-resolution matting branch. The structure of the matting head includes three convolutional layers with a kernel size of 3×3, and one convolutional layer with a kernel size of 1×1.

D. Loss Function

In the domain of image processing, particularly in the context of high-resolution image matting, the transition from low to high resolution is not merely a matter of scaling up. It involves capturing the intricate details that are often lost in the process. A variety of loss functions were employed during the training phase, including Regression, Compositional, Laplacian, and the Matte Detail Resolution Difference (MDRD) loss. The MDRD loss function introduces a novel perspective to high-resolution image matting, particularly in the context of enhancing details that are often neglected by traditional loss functions. Below, we provide a concise explanation of the Regression loss, Compositional loss, and Laplacian loss. We also highlight the distinctive contributions of the MDRD loss and contrast it with the aforementioned loss functions.

1) Regression Loss: The definition of regression loss, also known as alpha 1-norm loss, involves calculating the mean absolute error between the actual and estimated alpha mattes within the domain of the unknown region:

$$\mathcal{L}_{rec} = \frac{1}{|\mathcal{U}|}\sum_{i\in\mathcal{U}}|\hat{\alpha}_i - \alpha_i| \tag{3}$$

where $\mathcal{U}$ is the unknown region marked in the trimap, and $\hat{\alpha}_i$ and α_i denote the predicted and ground-truth alpha matte at position i in the trimap, respectively.

2) Compositional Loss: In deep image matting [11], the compositional loss is defined as the absolute difference between the RGB colors of two composited images. One image is composited using the predicted alpha matte on the ground - truth foreground and background, and the other is composited using the ground - truth alpha matte, as defined below:

$$\mathcal{L}_{comp} = \sqrt{\left(c_{pre} - c_{gt}\right)^2 + \epsilon^2} \tag{4}$$

where c_{pre} denotes an image composited by predicted alpha matte, c_{gt} denotes an image composited by ground-truth alpha matte, and ϵ denotes a small number for avoiding zero compositional loss.

3) Laplacian Loss: Laplacian loss is defined as the difference between the predicted alpha matte $\hat{\alpha}$ and the ground-truth alpha matte α, as described below:

$$\mathcal{L}_{lap} = \sum_{i=1}^{5} 2^{i-1}\left\|L^i(\hat{\alpha}) - L^i(\alpha)\right\|_1 \tag{5}$$

where L^i denotes the i^{th} layer of the Laplacian pyramid of the alpha map, $\hat{\alpha}$ and α respectively represent the predicted and ground-truth alpha mattes.

4) Matte Detail Resolution Difference Loss: The Matte Detail Resolution Difference (MDRD) loss function introduces a novel strategy in the domain of high-resolution image matting by focusing on the supervision of the Detail Difference Feature Extraction (DDFE) process. This loss is distinct from other loss functions that typically target the final prediction outcomes.

Unlike the Regression Loss, which emphasizes overall accuracy, the MDRD loss is specifically tailored to capture and accentuate the subtleties of high-resolution mattes. It provides a more nuanced approach to detail preservation, ensuring that the transition from low to high resolution is not just a matter of scale but also of quality and clarity. In contrast to the Laplacian Loss, which measures structural content across different layers of the image, MDRD is more sensitive to the resolution-specific details that are intrinsic to high-resolution images. This enhanced resolution sensitivity ensures that the details

are not only upscaled but are also accurately and distinctly represented, reflecting the true essence of high-resolution imagery. Moreover, the MDRD loss function complements the Compositional Loss by focusing on the level of detail. While the Compositional Loss ensures color accuracy in image compositing, MDRD elevates this by ensuring that the color accuracy is accompanied by a high level of detail fidelity. This dual focus on both composition and detail makes MDRD a comprehensive loss function that addresses the multifaceted requirements of high-quality image matting.

MDRD is uniquely positioned to enhance the fine details characteristic of high-resolution mattes. It does not directly supervise the predicted alpha mattes but instead supervises the intermediate feature extraction step, ensuring that the learned features are rich in detail and resolution-specific information. This proactive supervision allows for a more refined and accurate representation of high-resolution details during the feature learning phase. The formulation of MDRD is as follows:

$$\mathcal{L}_{mdrd} = \left\| DDF^{\mathcal{U}}_{offset} - Diff_{\alpha} \right\|_2 \tag{6}$$

$DDF^{\mathcal{U}}_{offset}$ is calculated from the DDF_{offset} using a high-resolution trimap. Specifically, the unknown region value of the trimap are set to 1, while the other regions value are set to 0, and these values are then weighted and applied to the DDF_{offset}. $Diff_{\alpha}$ is calculated by measuring the difference between the ground-truth alpha mattes and the low-resolution alpha mattes predicted by the existing image matting model. $\|*\|_2$ is defined as the Euclidean norm.

MDRD is a complement to the total loss, where MDRD supervises DDFE to focus on extracting matte details that only exist in high-resolution images, regression loss supervises the overall quality of high-resolution matte, compositional loss supervises the color accuracy in image synthesis, and Laplacian loss supervises the structural content of different layers of alpha mattes. Combining multiple losses can achieve the effect of accurately extracting high-resolution alpha matte details. The integration of MDRD into the total loss $\mathcal{L}$ calculation is represented as:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{rec} + \lambda_2 \mathcal{L}_{comp} + \lambda_3 \mathcal{L}_{lap} + \lambda_4 \mathcal{L}_{mdrd} \tag{7}$$

IV. Experiment

In this section, the organization of the section was described. The dataset and five evaluation metrics used in the experiments were introduced, followed by a description of the implementation details and training setup. Five experiments were conducted: (1) validation of the HRIMF-AMR framework's effectiveness; (2) assessment of its adaptability; (3) ablation studies on DDFE and MDRD; (4) complexity analysis of HRIMF-AMR; and (5) discussion of its limitations.

A. Datasets

To assess the practical efficacy of the proposed approach, we carry out test experiments with high-resolution datasets. We employ the portion of Alphamatting [8] featuring high-resolution images, along with the Transparent-460 [38] dataset, given that their image resolutions are typically 2K and above. Fig. 5 illustrates the image resolution distribution of the datasets used in the experiments.

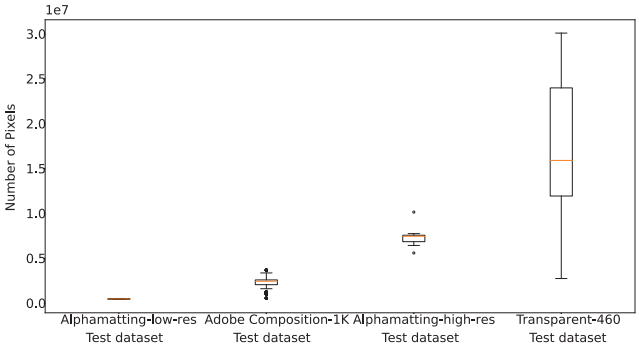


Fig. 5. Boxplot of image resolution size (replaced by the product of height and width of the image) for the dataset used in this article. Adobe Composition-1K-Test [11] and Alphamatting-low-Test [8] may be considered representative of low-resolution image datasets, and Transparent-460-Test [38] and Alphamatting-high-Test [8] may be considered representative of high-resolution image datasets. Alphamatting-low-Test and Alphamatting-high-Test are obtained from the same dataset but differ in resolution size.

1) Transparent-460: The Transparent-460 [38] dataset contains 460 well-annotated high-fidelity alpha mattes, with 410 images in the training subset and 50 images in the test subset. The background is based on the Adobe Composition-1K from the Microsoft COCO [39] and PASCAL VOC 2012 [40] datasets, respectively. It contains 41,000 training samples and 1,000 test samples, following the same composition rules as in [11]. In addition, images comprising the Transparent-460 test dataset have greater resolution, with an average size of 3915×4059 (ranging from 1661×1661 to 4480×6720).

2) Alphamatting: The Alphamatting [8] dataset consists of 8 test images and 27 training images, but provides both high- and low-resolution version for each image. Since the Adobe Composition-1K dataset integrates images from Alphamatting-train, only eight images from Alphamatting-test are used in the experiments. The average resolution of these eight images is 3127×2364, ranging from 2689×2085 to 3908×2600. However, the corresponding low-resolution image has an average resolution of 800×607 (with a minimum of 800×532 and a maximum of 800×671). All images in this dataset are natural (i.e., none are synthetic images).

3) Adobe Composition-1K: The Adobe Composition-1K [11] dataset comprises 43,100 images, each accompanied by an alpha matte resulting from the fusion of 431 distinct foreground elements with a corresponding number of background images. These backgrounds were randomly chosen from the Microsoft COCO [39] collection. Additionally, the test dataset encompasses 1,000 images, which are a blend of 50 unique foreground images with backgrounds sourced from the Pascal VOC 2012 [40] dataset. The average resolution of images in this test dataset varies from 1120×502 to 1920×1920, with an average of 1655×1380.

B. Metrics

We rate the performance of HRIMF-AMR via five metrics, four of which measure the quality of the predicted alpha mattes, while the remaining one reflects the computational complexity.

The Sum of Absolute Difference (SAD) metric measures the overall error by calculating the sum of the absolute differences of all pixels between the predicted alpha mattes and the true alpha mattes. It is the most intuitive error metric because it directly accumulates the prediction error for each pixel. SAD is calculated using the following expression:

$$SAD = \sum_{i=1}^{n} |\alpha_i - \hat{\alpha}_i| \tag{8}$$

where α_i is the ith pixel value of the predicted alpha matte, $\hat{\alpha}_i$ is the ith pixel value of the true alpha matte, and n is the total number of pixels.

The Mean Squared Error (MSE) metric emphasizes the effects of larger errors by squaring the prediction errors and then averaging them. This measure makes the contribution of a single large error to the overall error more significant, and helps to identify and improve significant defects in the algorithm. The formula for calculating MSE is:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (\alpha_i - \hat{\alpha}_i)^2 \tag{9}$$

for clarity, MSE is defined in this article as 10^{-3}.

Gradient error (Grad) metrics focus on assessing the edge quality of predicted alpha mattes by calculating the difference between the gradients of predicted and true alpha mattes. This metric is especially important for matting tasks that require high edge quality because it reflects the smoothness and accuracy of edges. The gradient error is calculated as:

$$Grad = \sum (\|\nabla\alpha_i - \nabla\hat{\alpha}_i\|^q) \tag{10}$$

where $\nabla\alpha_i$ and $\nabla\hat{\alpha}_i$ are the gradients of the predicted and true alpha mattes, respectively, and q is a constant, usually 2.

The Connectivity error (Conn) metric measures the difference between the connectivity of foreground pixels in the predicted alpha matte and the connectivity in the true alpha matte. This index is critical to ensure the integrity and connectivity of foreground objects in image matting. The connectivity error is calculated as:

$$Conn = \sum_i (\varphi(\alpha_i, \Omega) - \varphi(\hat{\alpha}_i, \Omega)) \tag{11}$$

where $\varphi(\alpha_i, \Omega)$ and $\varphi(\hat{\alpha}_i, \Omega)$ denote the connectivity of pixel i in the predicted and true alpha mattes, respectively, and Ω denotes the foreground region.

Latency is a key performance metric that measures the system response time, reflecting the total time that elapses from the request issuance to the response receipt.

C. Implementation Details

To provide alpha mattes for the comparisons of image matting quality, we integrate our HRIMF-AMR into Matteformer [7]. All methods were implemented under the environment of Python 3.9, PyTorch 1.8, CUDA 11.1, and cuDNN 8.0.5. We trained and tested our model on workstation clusters—CPU: Intel Xeon Gold 6226R CPU, GPU: NVIDIA A100 Tensor Core GPU with 40GB of graphics memory. It should be noted that the experiments discussed in this article (unless otherwise noted) were performed with the same resource configuration of the workstation cluster, i.e., with six core CPUs and one GPU. For fair comparison with other state-of-the-art methods, the method integrated with HRIMF-AMR was also trained on the Adobe Composition-1K [11] dataset.

In the training phase, input images were cropped randomly to 1024×1024. The random seed was set to 8282. These images are then subjected to random affine transformation, cropping, and real-world enhancement according to the strategy described in [6]. Affine transformations include random rotation, scaling, cropping, and vertical and horizontal flipping. To speed up the training process and prevent overfitting problems, the pre-trained weights of Matteformer [7] was loaded and the whole model was trained in an end-to-end manner. For loss optimization, the Adam [41] optimizer with $\beta_1 = 0.5$ and $\beta_2 = 0.999$ was set. Further, we set the initial optimizer learning rate to 1e-3, and use the warm-up strategy for 2,500 iterations, after which we adjust the learning rate according to the decay law of the learning rate. In the fine-tuning phase, the optimal HRIMF-AMR model is trained in 16 batches on an NVIDIA A100 for 50,000 iterations. In the inference phase, we input the high-resolution images and the trimap into the network to predict high-resolution alpha mattes.

D. Comparison Results on High-Resolution Images

As described in Subsection IV-B, sum of absolute difference (SAD), mean square error (MSE), gradient error (Grad) and connectivity (Conn) are used to assess the method integrated with HRIMF-AMR performance when applied to high-resolution images dataset. As outlined in Subsection IV-A, Adobe Composition-1K [11] and the portion of low-resolution images in the Alphamatting [8] test datasets are excluded from the comparison due to low resolution. For fair comparison, the proposed and the state-of-the-art methods used for the comparison are trained on the same dataset.

In the Table I, HRIMF-AMR achieves 32.65% to 80.41% improvement in SAD and 58.46% to 94.74% improvement in MSE compared to CNN-based image matting methods. HRIMF-AMR achieves 9.84% to 33.79% improvements in SAD and 37.55% to 61.44% improvements in MSE compared to Transformer-based image matting methods. These significant advancements in SAD and MSE metrics are primarily attributed to the contributions of Detail Difference Feature Extractor (DDFE) and Matte Detail Resolution Difference (MDRD) loss function. In addition, HRIMF-AMR achieves 2.19% to 88.13% improvement in Grad and 29.22% to 63.01% improvement in Conn compared to CNN-based image matting methods. HRIMF-AMR achieves 1.70% to 26.41% improvements in Conn compared to Transformer-based image matting methods. Although our method is inferior to DiffMatte-Swin [28] in terms of the Grad metric, the latter consumes a large amount of computational resources. DiffMatte-Swin [28] requires approximately 100GB of memory to run on the Transparent-460 [38] dataset. While our method only uses 23GB of graphic memory, enabling our method to run on consumer GPUs. Although HRIMF-AMR underperforms than the self-attention-based method in the Grad metrics, it outperforms in the MSE, SAD and Conn metrics. The main

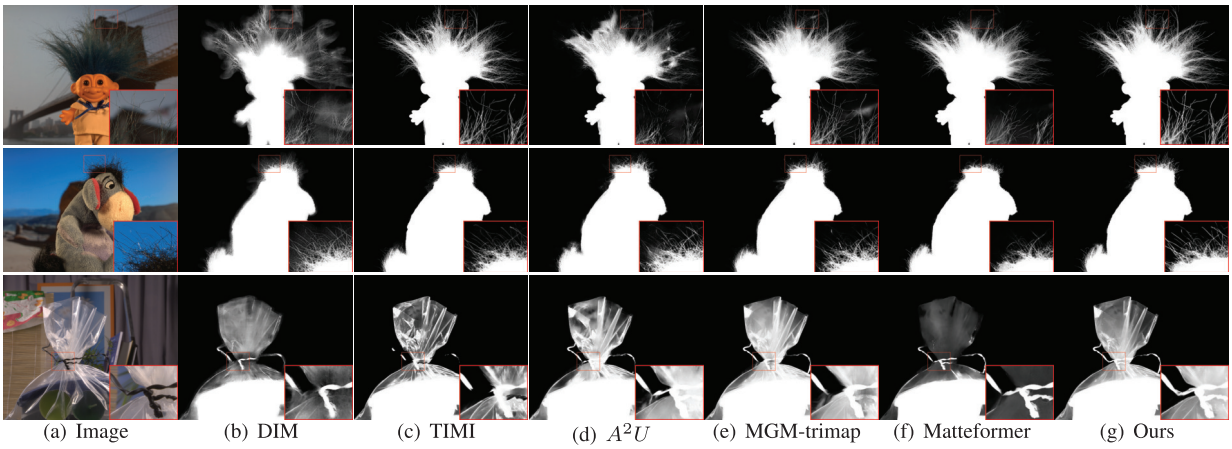


Fig. 6. Visual comparison of our HRIMF-AMR against other methods worked on the portion of high-resolution images in the Alphamatting [8] dataset. HRIMF-AMR demonstrates state-of-the-art performance in capturing fine details.

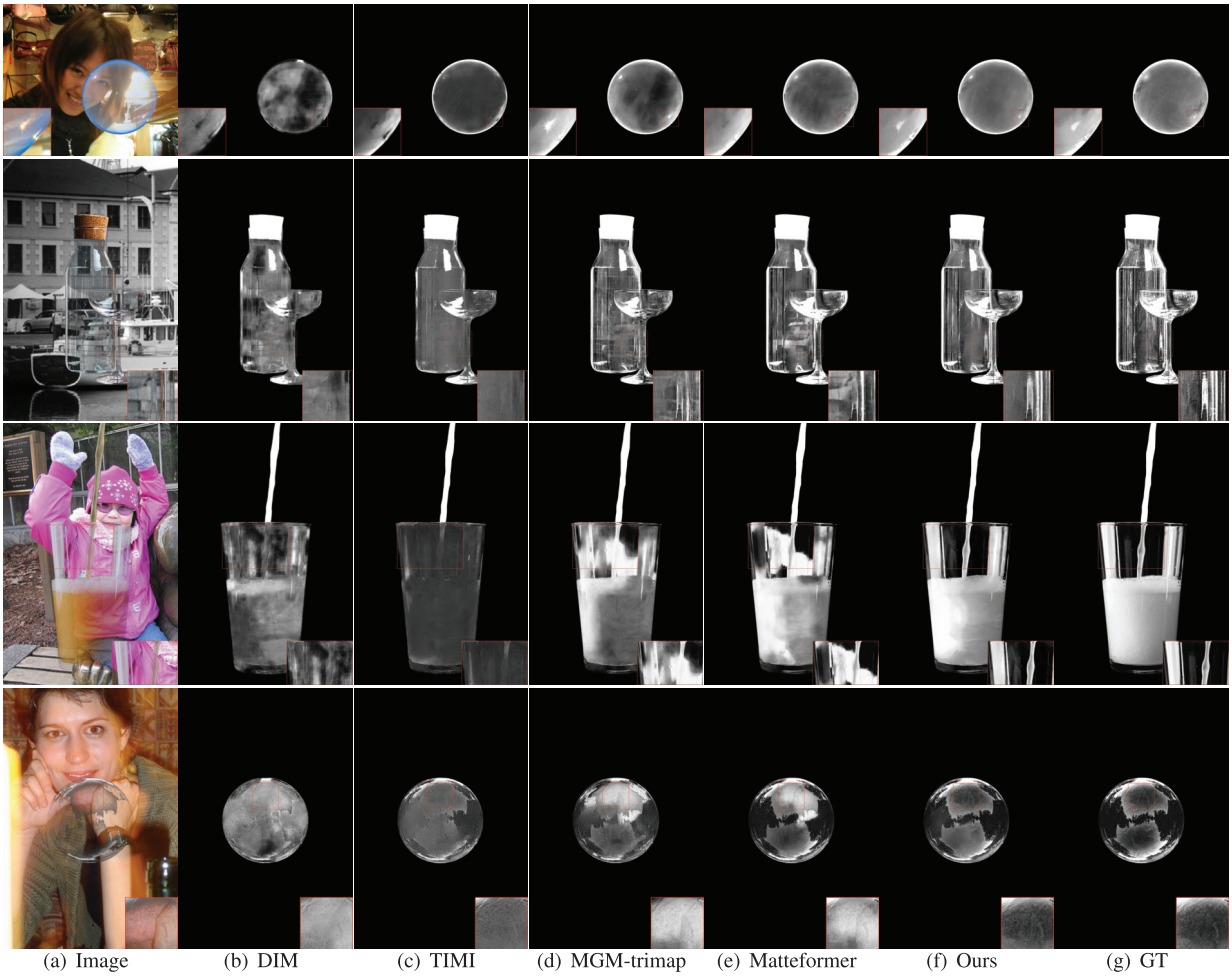


Fig. 7. Visual comparison of our HRIMF-AMR against other methods worked on the Transparent-460 [38] dataset. GT denotes the ground truth of the alpha matte. HRIMF-AMR excels in detail, outperforming other methods.

reasons are the difficulty of extracting edge details is caused by image downsampling in complex foreground structure scenes, and there is a quantitative deviation in index evaluation. This evidence demonstrates the effectiveness of HRIMF-AMR. By transforming the high-resolution image matting problem into a low-resolution image matting task and a high-resolution alpha matte refinement process, HRIMF-AMR significantly reduces computational complexity while maintaining high accuracy.

In Fig. 6 and Fig. 7, our method is visually compared with other state-of-the-art methods when applied to the portion of high-resolution images in the Alphamatting [8] and the Transparent-460 [38] test datasets, respectively. From Fig. 6, it

TABLE I

QUANTITATIVE RESULTS WERE ACHIEVED BY TESTING OUR METHOD AND OTHER STATE-OF-THE-ART METHODS ON THE TRANSPARENT-460 DATASET AFTER TRAINING ON THE ADOBE COMPOSITION-1K DATASET

Methods	SAD	MSE	Grad	Conn
DIM [11]	507.71	87.28	162.10	422.63
GCA-Matting [23]†	372.76	67.56	91.65	301.95
A^{2}U[12]*	350.37	56.11	106.35	292.21
MGM-trimap [6]	257.50	31.51	57.09	220.91
TIMI-Net [14]	328.08	44.20	142.11	289.79
SparseMat [24]	885.26	248.73	470.54	287.44
Matteformer [7]	194.73	21.59	36.97	161.75
TransMatting [38]	192.36	20.96	41.80	158.37
ELGT-Matting [26]*	261.93	33.95	49.87	212.46
VITMatte [27]†	-	-	-	-
DiffMatte-SwinT (S10) [28]†	203.37	25.81	**27.40**	159.05
Ours	**173.43**	**13.09**	55.84	**156.35**

The involved methods were run on a workstation equipped with an Inte1 Xeon Go1d 6226R CPU, 30 GB memory and an NVIDIA A100 GPU with 40GB of graphics memory.

∗ denotes the method was ran on a workstation equipped with an Intel Xeon Platinum 8352V CPU, 300 GB memory and an NVIDIA A800 GPU with 80GB of graphics memory.

† denotes the method was run on a personal computer equipped with an Intel Core i9-10900K CPU, 128 GB memory, and 1TB virtual memory (GPU acceleration was not available).

The experimental results of ViTMatte [27] cannot be provided because it ran out of 1024 GB memory.

TABLE II

ADAPTABILITY RESULTS OF THE PROPOSED FRAMEWORK

Methods	SAD	MSE	Grad	Conn
DIM [11]	507.71	87.28	162.10	422.63
DIM [11] + HRIMF-AMR	**399.37**	**55.81**	**132.69**	**336.92**
A^{2}U [12]	350.37	56.11	106.35	292.21
A^{2}U [12] + HRIMF-AMR	**286.52**	**39.64**	**90.35**	**250.19**
MGM-trimap [6]	257.50	31.51	**57.09**	220.91
MGM-trimap [6] + HRIMF-AMR	**212.41**	**20.03**	83.81	**186.46**
Matteformer [7]	194.73	21.59	**36.97**	161.75
Matteformer [7] + HRIMF-AMR	**173.43**	**13.09**	55.84	**156.35**

is evident that our method is still highly competitive in detail extraction from high-resolution natural images. It should also be noted that the images depicted in Fig. 6 have no ground-truth as it has not been published for this test set [8]. Likewise, because *www.alphamatting.com* does not provide an evaluation interface for high-resolution test images, no quantitative results are reported. Fig. 7 shows that, unlike other methods used in the comparison, our method can extract fine details from high-resolution images, such as the reflection of glass and the noise of transparent spheres.

E. Adaptability of HRIMF-AMR

To verify the adaptability of the proposed HRIMF-AMR framework, it has been integrated with classical methods such as DIM [11], A^{2}U [12], MGM-trimap [6], and Matteformer [7]. These methods have been carefully integrated, trained, and tested to ensure that our framework works optimally across different matting methods.

The adaptability results of HRIMF-AMR, shown in Table II, highlight the significant improvement in the overall performance of these classical methods after integration with HRIMF-AMR. Specifically, the accuracy of the DIM [11] method is improved by about by 21.4% (SAD), 36.3% (MSE), 18.1% (Grad), and 19.8% (Conn), respectively. Our HRIMF-AMR also addresses memory limitations encountered by certain algorithms, the A^{2}U [12] method being a prime example. Initially, it required more graphics memory than was available on a workstation equipped with an Inte1 Xeon Go1d 6226R CPU, 30 GB memory and an NVIDIA A100 GPU with 40GB of graphics memory. However, by integrating it with HRIMF-AMR, it was rendered capable of operating under these constraints. In addition, when incorporated into the HRIMF-AMR ensemble, the MGM-trimap [6] method exhibits significant accuracy improvements, with SAD reduced by 17.9%, MSE by 36.2%, and Conn by 15.7%. The Matteformer [7] method also demonstrates significant efficiency gains due to the HRIMF-AMR adoption. Its accuracy increased by 10.9% (SAD), 39.4% (MSE), and 0.03% (Conn). These results confirm that HRIMF-AMR can improve the accuracy of existing high-resolution image matting methods, while reducing the computational resources required for processing high-resolution images.

After integrating HRIMF-AMR into the system, a slight increase in Grad error is observed, which could be attributed to several underlying factors. It is conceivable that focusing on minimizing SAD and MSE may lead to an emphasis on overall image quality, possibly at the expense of refining edge details. In addition, although HRIMF-AMR is designed to optimize in a wider accuracy range, it may introduce small artifacts in the edge regions, which may affect the smoothness and accuracy of the edges, thus increasing the Grad error. However, our experimental results show that HRIMF-AMR has strong robustness and significantly improves the computational efficiency and accuracy of these image matting methods, which verifies the effectiveness of the HRIMF-AMR ensemble.

TABLE III

ABLATION STUDY RESULTS OF THE PROPOSED DETAIL DIFFERENCE FEATURE MODULE AND MATTE DETAIL RESOLUTION DIFFERENCE LOSS FUNCTION

Methods	SAD	MSE	Grad	Conn
BL [7]	194.73	21.59	**36.97**	161.75
BL + bilinear [42]	264.58	48.57	299.17	172.96
BL + DDFE	174.93	13.45	56.71	157.09
BL + DDFE + MDRD	**173.43**	**13.09**	55.84	**156.35**

F. Ablation Study

Ablation studies are conducted by training on the Adobe Composition-1K [11] dataset and testing on the Transparent-460 [38] dataset. In this experiment, we investigate three schemes: training and testing with interpolation-based upsampling and downsampling [42], utilizing our DDFE, and incorporating MDRD loss as an additional constraint. The results presented in Table III show that using interpolation-based upsampling and downsampling for training and testing results in a considerable loss of detail information in high-resolution images, which is unacceptable for a dense predictive task such as high-resolution image matting. The reported

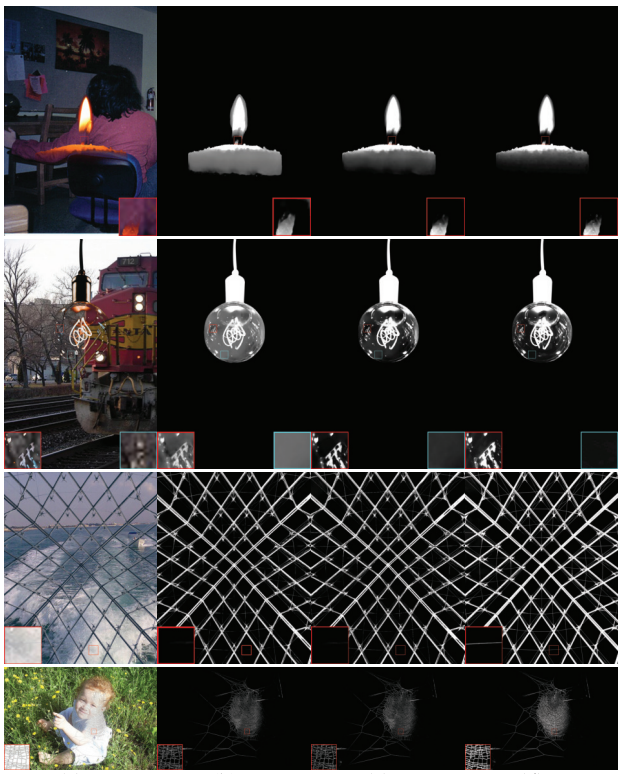


Fig. 8. Examples of paired low-resolution predicted alpha matte, high-resolution predicted alpha matte on the error correction. (a) high-resolution image, (b) predicted low-resolution alpha mattes obtained by the low-resolution image matting branch in HRIMF-AMR, (c) predicted high-resolution alpha mattes obtained by HRIMF-AMR and (d) high-resolution ground truth.

results thus reflect the effectiveness of DDFE and MDRD comprising our HRIMF-AMR. The DDFE significantly reduces the SAD and MSE values, reflecting the improved accuracy in capturing details. The addition of MDRD loss function further refines these metrics, indicating its subtle but positive effect on detail. As the Grad metric evaluates the edge quality of a predicted alpha matte by comparing the gradient of the predicted alpha matte with that of the true alpha matte, it initially increased with the addition of DDFE, suggesting that the edge smoothness is temporarily reduced. Although incorporating MDRD loss has a positive impact on improving the balance between detail preservation and edge smoothing, there is still scope for further optimization to achieve desirable edge quality. The Conn metric, which is critical to ensuring the integrity and connectivity of foreground objects, shows improvements over the DDFE module, highlighting the role of MDRD in preserving foreground connectivity. The MDRD loss further optimizes the Conn metric and strengthens the ability of the framework to maintain the alpha matte structure consistency.

G. Limitation

To assess error correction ability of HRIMF-AMR, we compared the alpha matte quality between the low-resolution branch and the final high-resolution output using mean absolute difference (MAD) and mean square error (MSE) metrics

TABLE IV

QUANTITATIVE RESULTS OF ERROR CORRECTION CAPABILITY OF THE PROPOSED FRAMEWORK

Alpha mattes	MAD	MSE
Predicted low-resolution alpha mattes	118.08	26.23
Predicted high-resolution alpha mattes	**67.31**	**13.09**

Predicted low-resolution alpha mattes were obtained by the low-resolution image matting branch in HRIMF-AMR.
Predicted high-resolution alpha mattes were obtained by HRIMF-AMR.

TABLE V

A COMPLEXITY ANALYSIS OF OUR METHOD ON THE TRANSPARENT-460 DATASET

Method	Params	Avg. FLOPs	Avg. Latency	MSE
DIM [11]	130.6M	7,914.14 G	2.39 s	87.28
A^2U[12]	8.1M	3,262.56 G	3.39 s	56.11
MGM-trimap [6]	29.7M	284.30 G	2.75 s	31.51
TIMI-Net [14]	34.89M	45.19 G	**1.30** s	44.20
Matteformer [7]	44.81M	1,744.26 G	3.94 s	21.59
ELGT-Matting [26]	53.0M	1,708.03 G	2.76 s	33.95
Ours	44.82M	498.56 G	2.29 s	**13.09**

on the Transparent-460 [38] dataset. MAD and MSE are two resolution-independent metrics, and the MAD is defined as 10^{-3}. For the low-resolution alpha matte, the metrics were computed against a downsampled ground truth using the same ratio as in the low-resolution branch. The results show that HRIMF-AMR improves MAD by 43% and MSE by 50.1% over the low-resolution branch, demonstrating its ability to correct inaccuracies (Table IV, Fig. 8). However, the method has limitations: (1) Severe errors in the low-resolution alpha matte, particularly in images with similar foreground and background characteristics (e.g., meshes, holes, or lines), cannot be fully corrected, as downsampling may cause critical detail loss. (2) Performance degradation occurs when low-resolution alpha mattes contain significant errors, as HRIMF-AMR's refinement depends on the quality of the low-resolution alpha matte, potentially increasing gradient error by nearly 50% compared to the baseline in challenging cases.

H. Complexity Analysis

This experiment studies the complexity of HRIMF-AMR. The FLOPs, number of parameters, and average latency of the involved methods were computed on the Transparent-460 [38] dataset. The average FLOPs were obtained by calculating the FLOPs for 1000 images in the Transparent-460 [38] Test dataset. The average latency of the involved methods was obtained by performing them on 1000 images in the Transparent-460 [38] Test dataset. In addition, MSE metric was used in the comparison to demonstrate their trade-off between computational complexity and matting accuracy. The involved methods were run on a workstation equipped with an Intel Xeon Platinum 8352V CPU, 300 GB memory, and an NVIDIA A800 GPU with 80 GB of graphics memory.

The experimental results shown in Table V demonstrate that HRIMF-AMR outperforms almost all involved image matting methods in terms of FLOPs and average latency. Specifically, HRIMF-AMR demonstrates significantly fewer FLOPs than

DIM [11], A^{2}U [12], Matteformer [7] and ELGT-Matting [26]. Although HRIMF-AMR does not surpass MGM-trimap [6] in terms of FLOPs, it achieves lower average latency and MSE. Similarly, while HRIMF-AMR is less efficient than TIMI-Net [14] in both FLOPs and average latency, it maintains a lower MSE. A particularly noteworthy observation is that while HRIMF-AMR and Matteformer possess a comparable number of parameters, HRIMF-AMR outperforms Matteformer by 71.42% in FLOPs on the Transparent-460 [38] dataset. These results collectively demonstrate that HRIMF-AMR achieves an appropriate balance between computational efficiency and performance in high-resolution image matting.

V. Conclusion

In this work, we introduced the high-resolution image matting framework based on alpha matte refinement from low-resolution to high-resolution (HRIMF-AMR), addressing the challenges associated with the low quality and efficiency in high-resolution image matting. By decomposing the problem into low-resolution matting and high-resolution refinement, our approach reduces its complexity. Appropriate existing methods are leveraged for the initial low-resolution matting. For the refinement, the Detail Difference Feature Extractor (DDFE) is introduced to capture fine details by comparing high-resolution and low-resolution image features. The matte is then enhanced using these extracted features. The novel Matte Detail Resolution Difference (MDRD) loss function is employed for training the DDFE, ensuring that the extracted features are aligned with the matte quality. The experimental results reported here demonstrate that the method integrated our HRIMF-AMR outperforms state-of-the-art methods when applied to high-resolution datasets in terms of Sum of Absolute Differences (SAD), Mean Square Error (MSE), and Connectivity error (Conn), validating its effectiveness and superiority in high-resolution image matting.

Our future work will involve designing novel framework components that can accurately extract information about discontinuous foreground changes when the background is similar, and can accurately capture detailed features when computational resources are limited.

References

[1] T. Wei et al., "Deep image matting with sparse user interactions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 2, pp. 881–895, Feb. 2024.

[2] S. Lin, A. Ryabtsev, S. Sengupta, B. L. Curless, S. M. Seitz, and I. Kemelmacher-Shlizerman, "Real-time high-resolution background matting," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2021, pp. 8762–8771.

[3] W. Li, Z. Zou, and Z. Shi, "Deep matting for cloud detection in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 12, pp. 8490–8502, Dec. 2020.

[4] Y. Liang, H. Huang, Z. Cai, Z. Hao, and K. C. Tan, "Deep infrared pedestrian classification based on automatic image matting," *Appl. Soft Comput.*, vol. 77, pp. 484–496, Apr. 2019.

[5] J. Ruan, H. Cui, Y. Huang, T. Li, C. Wu, and K. Zhang, "A review of occluded objects detection in real complex scenarios for autonomous driving," *Green Energy Intell. Transp.*, vol. 2, no. 3, Jun. 2023, Art. no. 100092.

[6] Q. Yu et al., "Mask guided matting via progressive refinement network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 1154–1163.

[7] G. Park, S. Son, J. Yoo, S. Kim, and N. Kwak, "MatteFormer: Transformer-based image matting via prior-tokens," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 11686–11696.

[8] C. Rhemann, C. Rother, J. Wang, M. Gelautz, P. Kohli, and P. Rott, "A perceptually motivated online benchmark for image matting," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 1826–1833.

[9] A. Levin, D. Lischinski, and Y. Weiss, "A closed-form solution to natural image matting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 228–242, Feb. 2008.

[10] Y.-Y. Chuang, B. Curless, D. H. Salesin, and R. Szeliski, "A Bayesian approach to digital matting," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2, Jun. 2001, pp. 264–271.

[11] N. Xu, B. Price, S. Cohen, and T. Huang, "Deep image matting," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 311–320.

[12] Y. Dai, H. Lu, and C. Shen, "Learning affinity-aware upsampling for deep image matting," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 6837–6846.

[13] H. Yu, N. Xu, Z. Huang, Y. Zhou, and H. Shi, "High-resolution deep image matting," in *Proc. AAAI Conf. Artif. Intell.*, 2021, vol. 35, no. 4, pp. 3217–3224.

[14] Y. Liu et al., "Tripartite information mining and integration for image matting," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 7555–7564.

[15] A. Levin, A. Rav Acha, and D. Lischinski, "Spectral matting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 10, pp. 1699–1712, Oct. 2008.

[16] Q. Chen, D. Li, and C.-K. Tang, "KNN matting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 9, pp. 2175–2188, Sep. 2013.

[17] Y. Aksoy, T. O. Aydin, and M. Pollefeys, "Designing effective inter-pixel information flow for natural image matting," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 228–236.

[18] L. Grady, T. Schiwietz, S. Aharon, and R. Westermann, "Random walks for interactive alpha-matting," in *Proc. 5th IASTED Int. Conf. Vis., Imag., Image Process.*, 2005, pp. 423–429.

[19] E. S. L. Gastal and M. M. Oliveira, "Shared sampling for real-time alpha matting," *Comput. Graph. Forum*, vol. 29, no. 2, pp. 575–584, May 2010.

[20] Y. Liang, Z. Cai, H. Huang, Q. Wu, F. Feng, and X. Ling, "An alpha matting algorithm based on collaborative swarm optimization for high-resolution images," *SCIENTIA SINICA Informationis*, vol. 50, no. 3, pp. 424–437, Mar. 2020.

[21] F. Fujian, Y. Yuan, T. Mian, G. Hongshan, L. Yihui, and W. Lin, "An alpha matting algorithm based on micro-scale searching for high-resolution images," *Pattern Recognit. Artif. Intell.*, vol. 36, no. 6, p. 530, 2023.

[22] S. Liu, X. Wang, M. Weiszer, and J. Chen, "Extracting multi-objective multigraph features for the shortest path cost prediction: Statistics-based or learning-based?," *Green Energy Intell. Transp.*, vol. 3, no. 1, Feb. 2024, Art. no. 100129.

[23] Y. Li and H. Lu, "Natural image matting via guided contextual attention," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 7, pp. 11450–11457.

[24] Y. Sun, C.-K. Tang, and Y.-W. Tai, "Ultrahigh resolution image/video matting with spatio-temporal sparsity," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 14112–14121.

[25] Y. Wang, L. Tang, Y. Zhong, and B. Li, "From composited to real-world: Transformer-based natural image matting," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 4, pp. 2097–2111, Apr. 2024.

[26] L. Hu, Y. Kong, J. Li, and X. Li, "Effective local–global transformer for natural image matting," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 8, pp. 3888–3898, Aug. 2023.

[27] J. Yao, X. Wang, S. Yang, and B. Wang, "ViTMatte: Boosting image matting with pre-trained plain vision transformers," *Inf. Fusion*, vol. 103, Mar. 2024, Art. no. 102091.

[28] Y. Hu, Y. Lin, W. Wang, Y. Zhao, Y. Wei, and H. Shi, "Diffusion for natural image matting," in *Proc. 18th Eur. Conf. Comput. Vis.-ECCV*, Milan, Italy, Sep. 2024, pp. 181–199.

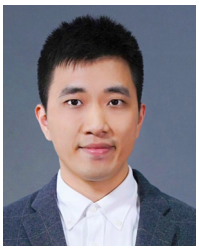
[29] H. Huang, Y. Liang, X. Yang, and Z. Hao, "Pixel-level discrete multiobjective sampling for image matting," *IEEE Trans. Image Process.*, vol. 28, no. 8, pp. 3739–3751, Aug. 2019.

[30] Y. Liang, H. Gou, F. Feng, G. Liu, and H. Huang, "Natural image matting based on surrogate model," *Appl. Soft Comput.*, vol. 143, Aug. 2023, Art. no. 110407.

[31] Y. Yang et al., "Multi-criterion sampling matting algorithm via Gaussian process," *Biomimetics*, vol. 8, no. 3, p. 301, Jul. 2023.

[32] Y. Zhong, B. Li, L. Tang, H. Tang, and S. Ding, "Highly efficient natural image matting," in *Proc. Brit. Mach. Vis. Conf.*, Jan. 2021, pp. 1–13.
[33] Z. Ke, J. Sun, K. Li, Q. Yan, and R. W. Lau, "MODNet: Real-time trimap-free portrait matting via objective decomposition," in *Proc. AAAI Conf. Artif. Intell.*, 2022, vol. 36, no. 1, pp. 1140–1147.
[34] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 10012–10022.
[35] A. Dosovitskiy et al., "An image is worth 16 x 16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.
[36] J. Ho, A. N. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Proc. 34th Int. Conf. Neural Inf. Process. Syst.*, Jan. 2020, pp. 6840–6851.
[37] Y. Xing, Q. Song, and G. Cheng, "Benefit of interpolation in nearest neighbor algorithms," *SIAM J. Math. Data Sci.*, vol. 4, no. 2, pp. 935–956, Jun. 2022.
[38] H. Cai, F. Xue, L. Xu, and L. Guo, "TransMatting: Enhancing transparent objects matting with transformers," in *Proc. Eur. Conf. Comput. Vis.-ECCV*, Jan. 2022, pp. 253–269.
[39] T. Lin et al., "Microsoft COCO: Common objects in context," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 740–755.
[40] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal visual object classes (VOC) challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, 2010.
[41] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[42] V. D. Earshia and M. Sumathi, "A comprehensive study of 1D and 2D image interpolation techniques," in *Proc. Int. Conf. Commun. Cyber Phys. Eng.*, Aug. 2018, pp. 383–391.

Xianmin Ye received the B.S. degree in statistics from Zunyi Normal University, China, in 2022, and the M.S. degree in statistics from Guizhou Minzu University, China, in 2025. His current research interests include alpha matting and image processing.



Yihui Liang received the B.S. degree in digital media technology from Xi'an University of Technology, China, in 2012, and the M.Eng. and Ph.D. degrees in software engineering from the South China University of Technology, China, in 2015 and 2019, respectively. He is currently an Associate Professor with the School of Computer Science, Zhongshan Institute, University of Electronic Science and Technology of China. His current research interests include alpha matting and image processing.

Mian Tan received the B.S. degree in mathematics and applied mathematics and the M.S. degree in probability theory and mathematical statistics from Guizhou Minzu University, Guiyang, Guizhou, China, in 2009 and 2012, respectively. She is currently an Associate Professor with Guizhou Key Laboratory of Pattern Recognition and Intelligent System, Guizhou Minzu University. Her current research interests include alpha matting and microcomputation.

Fujian Feng received the B.S. degree in computer science and technology from Shandong Technology and Business University, Yantai, China, in 2009, the M.S. degree in probability theory and mathematical statistics from Guizhou Minzu University, Guiyang, China, in 2013, and the Ph.D. degree in software engineering from the South China University of Technology, Guangzhou, China, in 2022. He is currently a Professor with Guizhou Key Laboratory of Pattern Recognition and Intelligent System, Guizhou Minzu University. His research interests include evolutionary computation and microcomputation.

Lin Wang received the B.S. degree in mathematics from Guizhou Minzu University, Guizhou, China, in 1987, and the Ph.D. degree in computer science from the University of Michel de Montaigne Bordeaux 3, France, in 2005. He is currently a Professor with Guizhou Key Laboratory of Pattern Recognition and Intelligent System, Guizhou Minzu University. His research interests include image processing, pattern recognition, and intelligent control.

Han Huang (Senior Member, IEEE) received the B.S. degree in information management and information system from the School of Mathematics, South China University of Technology (SCUT), Guangzhou, China, in 2003, and the Ph.D. degree in computer science from SCUT in 2008. He is currently a Full Professor with the School of Software Engineering, SCUT. His research interests include theoretical foundation and application of microcomputation, intelligent software engineering, data intelligence engineering, and biochemical computer. He is a Distinguished Member of CCF.