



Frequency Information Matters for Image Matting

Rongsheng Luo¹, Rukai Wei², Changxin Gao¹, and Nong Sang¹(✉)

¹ Key Laboratory of Image Processing and Intelligent Control, School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, China

{rongshengluo, cgao, nsang}@hust.edu.cn

² Wuhan National Laboratory For Optoelectronics, Huazhong University of Science and Technology, Wuhan, China
weirukai@hust.edu.cn

Abstract. Image matting aims to estimate the opacity of foreground objects in order to accurately extract them from the background. Existing methods are only concerned with RGB features to obtain alpha mattes, limiting the perception of local tiny details. To address this issue, we introduce frequency information as an auxiliary clue to accurately distinguish foreground boundaries and propose the **F**requency **M**atting **N**etwork (FMN). Specifically, we deploy a Frequency Boosting Module (FBM) in addition to the Discrete Cosine Transform (DCT) to extract frequency information from input images. The proposed FBM is a learnable component that empowers the model to adapt to complex scenarios. Furthermore, we design a Domain Aggregation Module (DAM) to effectively fuse frequency features with RGB features. With the assistance of frequency clues, our proposed FMN achieves significant improvements in matting accuracy and visual quality compared with state-of-the-art methods. Extensive experiments on Composition-1k and Distinctions-646 datasets demonstrate the superiority of introducing frequency information for image matting.

Keywords: Image matting · Frequency matting network · Frequency boosting module · Domain aggregation module

1 Introduction

Natural image matting is a crucial task in computer vision which involves extracting a high-quality alpha matte (i.e., the opacity of foreground object at each pixel) from an image, as shown in Fig. 1. This task has numerous applications in modern life, particularly in fields such as virtual reality, film production, and digital image processing. Generally, the input image is represented as a linear combination of foreground and background colors with alpha mattes [22]. This representation can be expressed mathematically as:

$$I_i = \alpha_i F_i + (1 - \alpha_i) B_i, \alpha_i \in [0, 1], \quad (1)$$

where α_i refers to the opacity of foreground objects at pixel i , I_i refers to the RGB color at pixel i , and F_i and B_i refer to the RGB colors of the foreground and background at pixel i , respectively. Obviously, the image matting problem is highly ill-posed since there are seven values to be determined, but only three values are known for each pixel of a given image.

Most matting approaches typically require a well-annotated trimap as an auxiliary input [2, 9, 10, 16, 18, 27, 28]. Since they depend excessively on the quality of trimaps to predict alpha mattes, automatic matting algorithms [13, 14, 20, 28, 29] have recently garnered significant attention from the academic community due to their ability to eliminate auxiliary guidance.

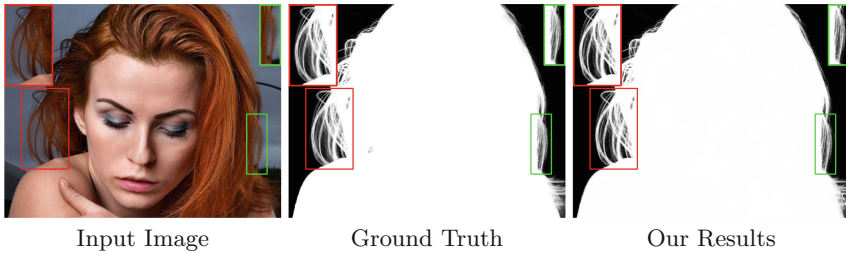


Fig. 1. The alpha matte generated by our FMN and the corresponding ground truth on Composition-1k test set.

All these matting methods [13, 14, 18–20, 28] share one common characteristic, i.e., they rely solely on RGB information for feature extraction, which limits the model to perceive local tiny details, e.g., changes of boundaries and illumination in hair. According to the studies of biology and psychology [32], some animals may beat humans in visual recognition since they have more wavebands (e.g., frequency information) than humans. Therefore, in this study, we claim that matting problems should not be limited to replicating the visual perception abilities of humans in the single RGB domain, but rather should incorporate the additional frequency information for superior recognition ability.

To this end, we propose a novel matting method named **F**requency **M**atting **N**etwork (FMN). We first follow existing methods [13, 28, 29] to deploy the CNN encoder for RGB feature extraction. Meanwhile, we use Discrete Cosine Transform (DCT) to transform the image into the frequency domain for frequency feature extraction. However, the fixed offline DCT algorithm fails to address real-world images with various foreground objects and complicated backgrounds. This motivates us to design a Frequency Boosting Module (FBM), which consists of band-wise boosting and space-wise boosting modules (See Sect. 3.3). In this way, the model can adapt to complex scenarios for accurate frequency information collection. Furthermore, we propose a Domain Aggregation Module (DAM), achieving effective feature fusion from the two domains (i.e., frequency domain and RGB domain) to generate informative feature maps (See Sect. 3.4). Finally, we adopt the Progressive Refinement Module (PRM) following [29] and

define a frequency perception loss to provide additional supervision for further frequency modeling.

Our main contributions can be summarized as:

- Different from previous matting approaches which only extract features in the RGB domain, we claim that matting should go beyond the RGB domain. Therefore, we present a new perspective that matting can be improved by incorporating frequency information with RGB information.
- To leverage information in the frequency domain, we present a novel matting network, i.e., **Frequency Matting Network (FMN)**, which comprises a **Frequency Boosting Module (FBM)** and a **Domain Aggregation Module (DAM)**. The former assists in enhancing frequency signals and the latter effectively fuses features from RGB domain and frequency domain. Furthermore, we design a frequency perception loss to provide supervision in the frequency domain.
- Comprehensive experiments on two widely-used matting datasets (i.e., Adobe Composition-1k and Distinctions-646) show that the proposed method outperforms state-of-the-art methods by a large margin.

2 Related Work

In this section, we provide a brief overview of the image matting methods, including traditional methods and deep-learning methods as well as knowledge about learning in the frequency domain.

2.1 Image Matting

Traditional Methods. Traditional matting methods typically rely on color model established from the input image. According to the manner additional inputs are utilized, traditional matting approaches are further divided into sampling-based approaches and affinity-based approaches. Sampling-based methods [10, 23] mainly calculate alpha mattes by representing each pixel inside transition regions with a pair of known foreground and background color. Affinity-based methods [1, 4, 11, 12] propagate the alpha values from known regions to unknown ones based on the affinities among adjacent pixels, resulting in high computational complexity.

Deep-learning Methods. Deep-learning methods typically provide superior performance than traditional methods and compensate for their shortcomings. Trimap-based learning methods require annotated trimaps as additional inputs. In the pioneering work, [27] proposes an encoder-decoder network that takes an RGB image and its corresponding trimap as inputs to estimate alpha matte. [9] presents a context-aware natural image matting method for simultaneous foreground and alpha mattes estimation. [2] proposes AdaMatting, which disentangles the matting task into trimap adaptation and alpha estimation. To address cross-patch dependency and consistency issues between patches, [28] proposes

a patch-based method for high-resolution inputs. [18] proposes a transformer-based architecture with prior-tokens which imply global information of each trimap region as global priors.

In contrast to trimap-based methods, trimap-free methods predict alpha mattes without trimaps. [30] proposes a structure with two decoders to classify foreground and background and fuses them in an extra network. [20] employs spatial and channel attention to integrate appearance cues and pyramidal features. [29] uses a binary mask as additional input and proposes a method to progressively refine the uncertain regions through the decoding process. [13, 14] predict trimap parallel to alpha matte of the transition region and then fuse them to obtain the final alpha matte.

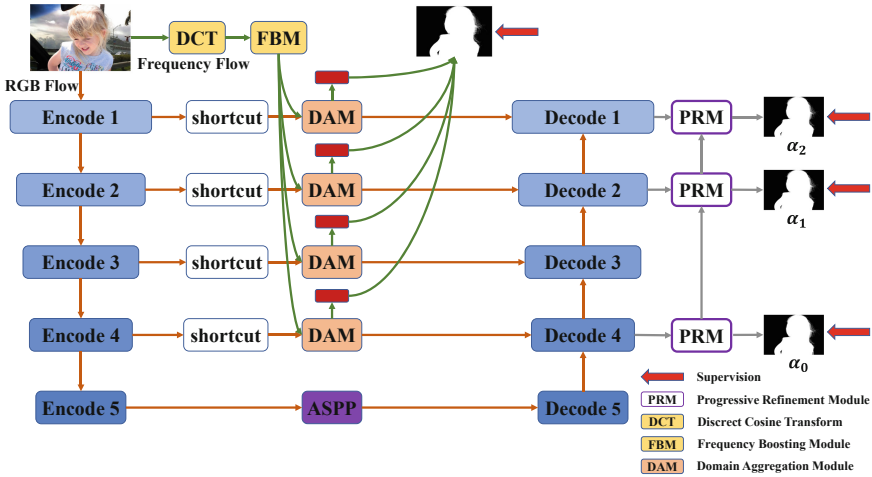


Fig. 2. The overall architecture of the proposed FMN. The frequency features from FBM and the RGB features from CNN encoder are integrated by DAM. Subsequently, the outputs from DAM, on the one hand, are fed to the decoder for alpha matte prediction at multiple resolutions. On the other hand, they are used for frequency perception to achieve more effective guidance for the network in the frequency domain.

2.2 Learning in the Frequency Domain

The frequency-domain compressed representations contain rich patterns for image understanding tasks. [8] conducts image classification task based on features extracted from frequency domain. [26] first converts information to frequency domain for better feature extraction and uses the SE-Block to select the beneficial frequency channels and simultaneously filter meaningless ones. [21] proposes that global average pooling(GAP) operation is dissatisfactory in capturing a range of features since it is equivalent to the lowest frequency elements

from perspective of the frequency domain. Our proposed FMN benefits from existing trimap-free methods in terms of the model design. We also innovatively introduce frequency information into the matting task to perceive more local details.

3 Frequency Matting Network

In this section, we present the overall network architecture of our **F**requency **M**atting **N**etwork (FMN) and provide details on its implementation. Additionally, we discuss the loss functions adopted in this paper.

3.1 Architecture Overview

We adopt ResNet34-UNet [15] with an Atrous Spatial Pyramid Pooling (ASPP) as the matting fundamental framework. As illustrated in Fig. 2, the input image is processed by two data flows, i.e., the RGB flow and the frequency flow. For the RGB flow, we use a CNN encoder to extract RGB features. While for the frequency flow, we utilize FBM after DCT to extract frequency features simultaneously. Then the features from two domains are fed into DAM for feature fusion. On the one hand, the output of DAM is processed by a convolution layer to reduce the dimension. The 1-d feature is used for frequency perception loss, which provides supervision in the frequency domain. On the other hand, the output is sent into decoder at the corresponding layer to reserve information lost in the decoding process. Finally, the outputs from Decoder 1, 2 and 4 are used by the Progressive Refinement Module (PRM) to selectively fuse features at different scales. We use weighted l_1 loss, composition loss and Laplacian loss to calculate loss in the RGB domain. Note that We provide supervision for the network in the both RGB domain and frequency domain. Therefore, we obtain high quality alpha mattes.

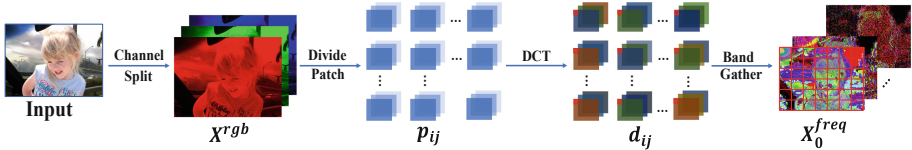


Fig. 3. The pipeline of Discrete Cosine Transform for an image.

3.2 DCT for Image Transformation

DCT utilizes an orthogonal cosine function as the basis function, which brings the energy of the image together and facilitates the extraction of features in the frequency domain. As shown in Fig. 3, the input RGB image x^{rgb} is firstly split

into three channels, then we can obtain $\{p_{i,j}^c | 1 \leq i, j \leq \frac{H}{8}\}$ by dividing x^{rgb} into a set of 8×8 patches. Specifically, we divide patches densely on slide windows of the image for further frequency processing. Finally, each patch of a certain color channel $p_{i,j}^c \in \mathbb{R}^{8 \times 8}$ is processed by DCT into frequency spectrum $d_{i,j}^c \in \mathbb{R}^{8 \times 8}$.

After operations discussed above, each value corresponds to the intensity of a certain frequency band. To group all components of the same frequency into one channel, we first obtain $d_{i,j} \in \mathbb{R}^{8 \times 8 \times 3}$ by concatenating each channel $d_{i,j}^c$, and then we flatten the frequency spectrum and reshape them to form $x_0^{freq} \in \mathbb{R}^{\frac{H}{8} \times \frac{W}{8} \times 192}$. In this way, we rearrange the signals in zigzag order within one patch and each channel of x_0^{freq} belongs to one band. Therefore, the original RGB input is transformed to the frequency domain.

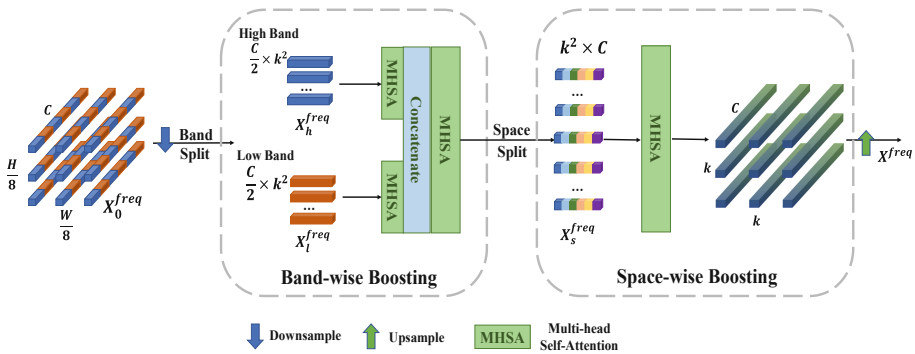


Fig. 4. The illustration of Frequency Boosting Module. It comprises two parts, i.e., band-wise boosting and space-wise boosting for interactions within individual patches and between patches, respectively.

3.3 Frequency Boosting Module

Although DCT is capable of transforming the image from RGB domain into frequency domain, its characteristic of having no learnable parameters makes it difficult to be adapted to complex scenarios. To solve this problem, we design a Frequency Boosting Module, and the framework is shown in Fig. 4. Specifically, we boost the signals from two aspects, including within individual patches and between patches. On the one hand, we enhance the coefficients in local frequency bands, i.e., band-wise boosting, and on the other hand, we establish interactions between patches, i.e., space-wise boosting. Firstly, we downsample and partition the signals into two parts, the low x_l^{freq} and high signals $x_h^{freq} \in \mathbb{R}^{96 \times k^2}$, where k means the size. To boost the signals in the corresponding frequency bands, we feed them into two multi-head self-attention (MHSA) separately and concatenate their outputs to recover the original shape. Secondly, we utilize another

MHSA to reconcile all the different frequency bands, whereas the rich correlation information between each channel in the input features is captured. We denote the output of band-wise boosting as x_f^{freq} . However, the above procedures only enable interactions between different frequency spectrums within a single patch. Therefore, we need to establish connections between patches. To this end, we first reshape x_f^{freq} to $x_s^{freq} \in \mathbb{R}^{k^2 \times C}$ and use MHSA to model the relationships among all the patches. Finally, we upsample these features and get the enhanced frequency signals x^{freq} .

3.4 Domain Aggregation Module

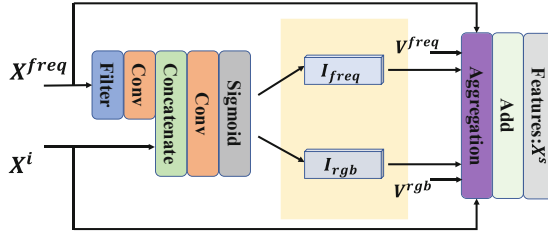


Fig. 5. The illustration of Domain Aggregation Module. It is designed to fuse features from the RGB domain and the frequency domain.

We have already obtained RGB features and frequency features by CNN and FBM, respectively. However, it remains a challenge to aggregate features from two different domains. To this end, we design the Domain Aggregation Module (DAM) to fuse these features, as shown in Fig. 5. The feature aggregation is a mutually reinforcing process, where frequency features are discriminative for local details while RGB features have a larger receptive field to perceive global semantics.

As CNNs are more sensitive to low-frequency channels, we first apply a filter to extract high-frequency channels manually. For an input frequency domain feature x^{freq} , the network can focus on the most important spectrum automatically. Specifically, we use a binary base filter f^{base} that covers the high-frequency bands and a Conv block to adjust the channels of frequency features for concatenation. Then we feed the aggregated features into another Conv block with two output channels and a sigmoid. In this way, we obtain the matrix I_{freq} for the frequency domain and I_{rgb} for the RGB domain, separately. Secondly, we aggregate features from two domains. Multiplied with the matrix and a learned vector $v \in \mathbb{R}^{1 \times C}$ to adjust the intensity of each channel, the aggregated features of each domain can be defined as:

$$X_s^{rgb} = I_{rgb} X^{rgb} \otimes V^{rgb}, \quad X_s^{freq} = I_{freq} X^{freq} \otimes V^{freq}, \quad (2)$$

Finally, we can obtain the fused features by adding two domain features: $X^s = X_s^{rgb} + X_s^{freq}$. In this way, we can make full use of discriminative frequency information while maintaining semantic information to ensure that both integrity and details of the foreground can be preserved.

3.5 Network Supervision

To further capture the frequency information that differs from human perception, we introduce a novel loss, i.e., frequency perception loss. Besides calculating loss directly in the RGB domain, we also intend to provide supervision for the network in the frequency domain. And we assume that the predictions should be correct not only at each pixel location but also in the coefficients after DCT when they act on the original images. As a result, we design the frequency perception loss to make the network mine more information in the frequency domain. We can define frequency perception loss as:

$$L_f(\alpha_p, \alpha_g) = \|DCT(\alpha_g) - DCT(\alpha_p)\|_2^2/q, \quad (3)$$

where q is the quantization table, α_p refers to predicted alpha matte and α_g refers to ground truth. α_p should be upsampled to the same size as α_g before loss calculation.

As can be seen in Fig. 2, the four predicted alpha mattes under different resolutions are rescaled to the input image size and then supervised by the frequency perception loss L_f in the frequency domain. The overall loss functions in the frequency domain are as follows:

$$L_{freq} = \sum_l w_l L_f(\alpha_p, \alpha_g) \quad (4)$$

where w_l is the loss weight of different scales. We set $w_{\frac{1}{8}} : w_{\frac{1}{4}} : w_{\frac{1}{2}} : w_1 = 1 : 2 : 2 : 3$ in our experiments.

We also provide supervision in the RGB domain. Previous scale alpha mattes preserve relatively complete profiles while they may suffer from ambiguous details, and current scale alpha mattes retain detail information while they may be subjected to background noises. Therefore, following [29], we adopt Progressive Refinement Module (PRM) to selectively fuse the alpha mattes from different scales with a self-guidance mask. Meanwhile, we employ their loss functions as overall loss functions in the RGB domain:

$$L_{rgb} = \sum_l w_l L(\alpha_p, \alpha_g) \quad (5)$$

$$L(\alpha_p, \alpha_g) = L_{l1}(\alpha_p, \alpha_g) + L_{comp}(\alpha_p, \alpha_g) + L_{lap}(\alpha_p, \alpha_g)$$

where w_l is the loss weight of different scales. We set $w_{\frac{1}{8}} : w_{\frac{1}{4}} : w_1 = 1 : 2 : 3$ in our experiments.

The final loss function for the FMN can be expressed as:

$$L_{final} = L_{rgb} + L_{freq} \quad (6)$$

4 Experiments

In this section, we evaluate the proposed Frequency Matting Network (FMN) on two datasets: Adobe Composition-1k [27] and Distinctions-646 [20]. We first compare FMN with SOTA methods both quantitatively and qualitatively. Then we perform ablation studies for FMN on Composition-1k and Distinctions-646 to demonstrate the importance of several crucial components.



Fig. 6. The visual comparisons on Composition-1k test set.

4.1 Datasets and Evaluation Metrics

Datasets. The first dataset is the public Adobe Composition-1k [27]. It consists of 431 foreground objects for training and 50 foreground objects which are composed with 20 background images chosen from PASCAL VOC [6] for testing. The second one is the Distinctions-646 dataset which improves the diversity of Composition-1k. It comprises 596 foreground objects for training and 50 foreground objects for testing, and then we produce 59, 600 training images and 1000 test images according to the composition rules in [27].

Table 1. The quantitative comparisons on Composition-1k test set. Upper part: trimap-based approaches. Lower part: trimap-free approaches.

Methods	SAD↓	MSE(10^3)↓	Grad↓	Conn↓
Share Matting [7]	125.37	0.029	144.28	123.53
Learning Based [31]	95.04	0.018	76.63	98.92
ClosedForm [12]	124.68	0.025	115.31	106.06
KNN Matting [4]	126.24	0.025	117.17	131.05
DCNN [24]	115.82	0.023	107.36	111.23
Info-Flow [1]	70.36	0.013	42.79	70.66
DIM [27]	48.87	0.008	31.04	50.36
AlphaGAN [17]	90.94	0.018	93.92	95.29
SampleNet [25]	48.03	0.008	35.19	56.55
CA Matting [9]	38.73	0.004	26.13	35.89
IndexNet [16]	44.52	0.005	29.88	42.37
GCA Matting [15]	35.27	0.004	19.72	31.93
A ² U [5]	33.78	0.004	18.04	31.00
Late Fusion [30]	58.34	0.011	41.63	59.74
HAttMatting [20]	44.01	0.007	29.26	46.41
HAttMatting++ [19]	43.27	0.006	27.91	44.09
PP-Matting [3]	46.22	0.005	22.69	45.40
FMN(Ours)	40.01	0.004	19.97	33.59

Evaluation metrics. We evaluate the alpha mattes following four common quantitative metrics: Sum of Absolute Differences (SAD), Mean Square Error (MSE), Gradient(Grad) and Connectivity (Conn) errors proposed by [27].

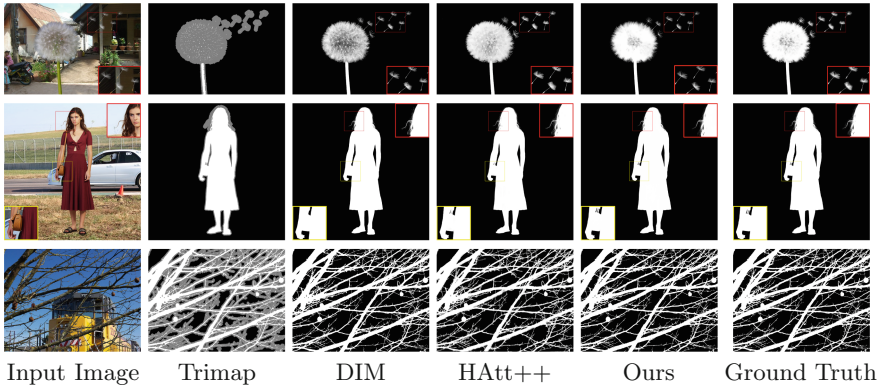
4.2 Evaluation Results

Evaluation on Composition-1k test set. We compare the FMN with 6 traditional hand-crafted algorithms as well as 11 deep learning-based methods. For the trimap-based methods, we can generate trimaps by dilating alpha mattes with random kernel size in the range of [1, 25]. As the qualitative and quantitative comparisons shown in Fig. 6 and Table 1, respectively, the proposed FMN exhibits significant superiority over traditional trimap-based approaches. With respect to trimap-based learning approaches, FMN still produces much better results than DIM [27], AlphaGAN [17], SampleNet [25], IndexNet [16] in terms of all the four metrics. For example, IndexNet achieves SAD 44.52 and MSE 0.005, while our method obtains a superior performance with SAD 40.01 and MSE 0.004. Moreover, our approach is slightly inferior to Context-aware [9] but a little worse than GCA Matting [15] and A²U [5]. However, our method can achieve equivalent performance without any auxiliary inputs, which is very convenient for novice users. The lower part of Table 1 illustrates that our FMN outperforms

Table 2. The quantitative comparisons on Distinctions-646 test set. Upper part: trimap-based approaches. Lower part: trimap-free approaches.

Methods	SAD↓	MSE(10^3)↓	Grad↓	Conn↓
Share Matting [7]	119.56	0.026	129.61	114.37
Learning Based [31]	105.04	0.021	94.16	110.41
ClosedForm [12]	105.73	0.023	91.76	114.55
KNN Matting [4]	116.68	0.025	103.15	121.45
DCNN [24]	103.81	0.020	82.45	99.96
Info-Flow [1]	78.89	0.016	58.72	80.47
DIM [27]	47.56	0.009	43.29	55.90
HAttMatting [20]	48.98	0.009	41.57	49.93
HAttMatting++ [19]	47.38	0.009	40.09	45.60
PP-Matting [3]	40.69	0.009	43.91	40.56
FMN(Ours)	34.28	0.006	19.93	27.23

the SOTA trimap-free approach to a great extent, which decreases SAD and Conn metrics heavily: from 46.22 and 45.40 to 40.01 and 33.59, respectively, indicating the effectiveness of our FMN.

**Fig. 7.** The visual comparisons on Distinctions-646 test set.

Evaluation on Distinctions-646 test set. We compare the FMN with 10 recent matting methods. We also use random dilation to generate high-quality trimaps [27] and relevant metrics are computed on the whole image. As qualitative and quantitative comparisons on the Distinctions-646 dataset displayed in Fig. 7 and Table 2, respectively, our FMN shows a clear advantage compared to all the mentioned matting approaches. It is noted that FMN outperforms trimap-free matting approach by a large margin, especially in terms of Grad

and Conn metrics. We can see a sharp decrease on Grad and Conn metrics, i.e., from 43.92 and 40.56 to 19.93 and 27.23 for PP-Matting, which indicates that our model can achieve high-quality visual perception.

Table 3. Ablation Study on Composition-1k dataset.

FBM	DAM	L_{freq}	SAD↓	MSE(10^3)↓	Grad↓	Conn↓
			64.33	0.018	42.68	48.42
✓			42.61	0.007	22.16	34.57
✓	✓		40.24	0.004	20.26	34.05
✓	✓	✓	40.01	0.004	19.97	33.59

Table 4. Ablation Study on Distinctions-646 dataset.

FBM	DAM	L_{freq}	SAD↓	MSE(10^3)↓	Grad↓	Conn↓
			51.46	0.012	45.68	56.71
✓			35.88	0.009	23.16	36.57
✓	✓		34.38	0.009	22.64	36.77
✓	✓	✓	34.28	0.007	19.93	27.23

4.3 Ablation Study

We validate the effectiveness of different components on Composition-1k dataset and Distinctions-646 dataset, separately. The correlated evaluation values are summarized in Table 3 and Table 4. Compared with results in the first row, the utilization of FBM can bring considerable performance improvements on all the four metrics to a great extent. For example, SAD error decreases sharply from 64.33 to 42.61 on Composition-1k dataset and from 51.46 to 35.88 on Distinctions-646 dataset. The main reason is that FBM introduces frequency information into matting assignment and thus provides more precise local tiny details to compensate for RGB features. Moreover, the results also show that DAM plays a vital role in fusing RGB features and frequency features, which provides a slight rise in alpha matte quality compared to simply adding the features from two domains in the second row. In addition, the application of frequency perception loss to constrain frequency features has proven to be valuable, particularly in terms of Grad and Conn metrics. Specifically, we observe a significant improvement in these metrics, with values decreasing from 22.64 and 36.77 to 19.93 and 27.23, respectively.

5 Conclusion

In this paper, we utilize frequency information of an image to help predicting alpha values in the transition areas, i.e., foreground boundaries. To extract the

discriminative cues in the frequency domain for complex scenario perception, we design a Frequency Boosting Module (FBM), which comprises band-wise boosting and space-wise boosting, to boost the coefficients in all the frequency bands. Furthermore, we integrate features from the RGB domain and the frequency domain through the Domain Aggregation Module (DAM). Besides, by providing supervision in both RGB domain and frequency domain, we can compensate for RGB information, which may tend to provide a large receptive field with details from frequency features. Experiments demonstrate that our proposed FMN achieves better performance than state-of-the-art matting methods on two commonly-used benchmarks. This work will inspire researchers to explore the utilization of frequency information in computer vision.

References

1. Aksoy, Y., Ozan Aydin, T., Pollefeys, M.: Designing effective inter-pixel information flow for natural image matting. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 29–37 (2017)
2. Cai, S., et al.: Disentangled image matting. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 8819–8828 (2019)
3. Chen, G., et al.: PP-matting: high-accuracy natural image matting. arXiv preprint [arXiv:2204.09433](https://arxiv.org/abs/2204.09433) (2022)
4. Chen, Q., Li, D., Tang, C.K.: KNN matting. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(9), 2175–2188 (2013)
5. Dai, Y., Lu, H., Shen, C.: Learning affinity-aware upsampling for deep image matting. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6841–6850 (2021)
6. Everingham, M., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A.: The pascal visual object classes (VOC) challenge. *Int. J. Comput. Vis.* **88**, 303–308 (2009)
7. Gastal, E.S., Oliveira, M.M.: Shared sampling for real-time alpha matting. In: *Computer Graphics Forum*, vol. 29, pp. 575–584. Wiley Online Library (2010)
8. Gueguen, L., Sergeev, A., Kadlec, B., Liu, R., Yosinski, J.: Faster neural networks straight from jpeg. In: *Advances in Neural Information Processing Systems 31* (2018)
9. Hou, Q., Liu, F.: Context-aware image matting for simultaneous foreground and alpha estimation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 4130–4139 (2019)
10. Karacan, L., Erdem, A., Erdem, E.: Image matting with KL-divergence based sparse sampling. In: Proceedings of the IEEE International Con Computer Vision, pp. 424–432 (2015)
11. Lee, P., Wu, Y.: Nonlocal matting. In: *CVPR 2011*, pp. 2193–2200. IEEE (2011)
12. Levin, A., Lischinski, D., Weiss, Y.: A closed-form solution to natural image matting. *IEEE Trans. Pattern Anal. Mach. Intell.* **30**(2), 228–242 (2007)
13. Li, J., Zhang, J., Maybank, S.J., Tao, D.: Bridging composite and real: towards end-to-end deep image matting. *Int. J. Comput. Vis.* **130**(2), 246–266 (2022)
14. Li, J., Zhang, J., Tao, D.: Deep automatic natural image matting. arXiv preprint [arXiv:2107.07235](https://arxiv.org/abs/2107.07235) (2021)

15. Li, Y., Lu, H.: Natural image matting via guided contextual attention. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 11450–11457 (2020)
16. Lu, H., Dai, Y., Shen, C., Xu, S.: Indices Matter: learning to index for deep image matting. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 3266–3275 (2019)
17. Lutz, S., Amliantitis, K., Smolic, A.: AlphaGAN: generative adversarial networks for natural image matting. arXiv preprint [arXiv:1807.10088](https://arxiv.org/abs/1807.10088) (2018)
18. Park, G., Son, S., Yoo, J., Kim, S., Kwak, N.: MatteFormer: transformer-based image matting via prior-tokens. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11696–11706 (2022)
19. Qiao, Y., et al.: Hierarchical and progressive image matting. ACM Trans. Multimed. Comput. Commun. Appl. **19**(2), 1–23 (2023)
20. Qiao, Y., et al.: Attention-guided hierarchical structure aggregation for image matting. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 13676–13685 (2020)
21. Qin, Z., Zhang, P., Wu, F., Li, X.: FcaNet: frequency channel attention networks. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 783–792 (2021)
22. Rhemann, C., Rother, C., Wang, J., Gelautz, M., Kohli, P., Rott, P.: A perceptually motivated online benchmark for image matting. In: 2009 IEEE Con Computer Vision and Pattern Recognition, pp. 1826–1833. IEEE (2009)
23. Shahrian, E., Rajan, D., Price, B., Cohen, S.: Improving image matting using comprehensive sampling sets. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 636–643 (2013)
24. Shen, X., Tao, X., Gao, H., Zhou, C., Jia, J.: Deep automatic portrait matting. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9905, pp. 92–107. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46448-0_6
25. Tang, J., Aksoy, Y., Oztireli, C., Gross, M., Aydin, T.O.: Learning-based sampling for natural image matting. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3055–3063 (2019)
26. Xu, K., Qin, M., Sun, F., Wang, Y., Chen, Y.K., Ren, F.: Learning in the frequency domain. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1740–1749 (2020)
27. Xu, N., Price, B., Cohen, S., Huang, T.: Deep image matting. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2970–2979 (2017)
28. Yu, H., Xu, N., Huang, Z., Zhou, Y., Shi, H.: High-resolution deep image matting. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, pp. 3217–3224 (2021)
29. Yu, Q., et al.: Mask guided matting via progressive refinement network. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1154–1163 (2021)
30. Zhang, Y., et al.: A late fusion CNN for digital matting. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 7469–7478 (2019)
31. Zheng, Y., Kambhampettu, C.: Learning based digital matting. In: 2009 IEEE 12th International Con Computer Vision, pp. 889–896. IEEE (2009)
32. Zhong, Y., Li, B., Tang, L., Kuang, S., Wu, S., Ding, S.: Detecting camouflaged object in frequency domain. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4504–4513 (2022)