Human Instance Matting via Mutual Guidance and Multi-Instance Refinement

Yanan Sun[1,2] Chi-Keung Tang[1] Yu-Wing Tai[2]

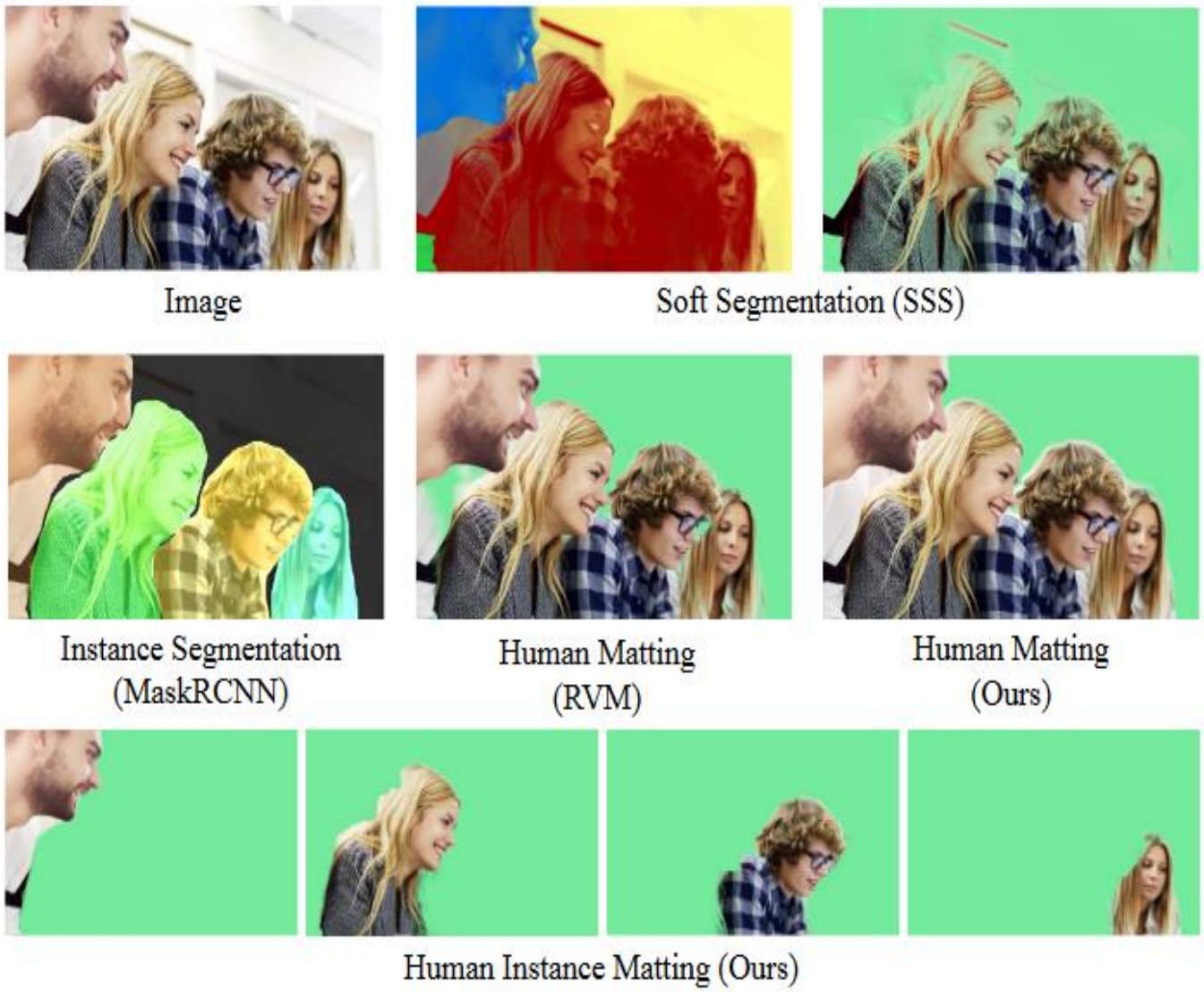
[1]HKUST [2]Kuaishou Technology

{now.syn, yuwing}@gmail.com cktang@cs.ust.hk

CVPR 2022（Oral）

New Matting Settings

- Instance human matting
  - Automatically extract precise alpha matte for **each human instance** in a given image;
  - No need for extra guidance;
  - Propose a new multi-human image matting dataset and benchmark.
  - Using a new metric to measure both instance recognition and matting quality (IMQ);



New Matting Dataset

- HIM2K Dataset
  - Synthetic Subset (1680 foregrounds with labels)
    - Each image includes 2-5 human foreground;
    - Foregrounds will be iteratively composited on non-human background;

$$I_i = I_0 \prod_{j=1}^{i} (1 - \alpha_j) + \sum_{j=1}^{i} \alpha_j F_j \prod_{k=j}^{i} (1 - \alpha_k)$$

  - Natural Subset (320 natural images with labels)
    - Cope with the domain gap;
    - Ground truth is obtained using PS.

New Matting Dataset

Figure 4. HIM2K examples: top is synthetic and bottom is natural.

New Matting Metric

- IMQ (Instance Matting Quality)
  - An extension of traditional matting metrics(MSE, MAD, Conn, etc)
  - Instance Matching
    - Variations of IoU between predictions and groundtruths using Hungarian matching;
    - TP(IoU > 0.5), FP and FN(false instance);
  - Similarity Measurement
    - Measurement of matting quality;

$$S(\alpha, \hat{\alpha}) = 1 - \min(w\mathcal{E}(\alpha, \hat{\alpha}), 1)$$

    - Adapted to MSE, MAD, Grad and Conn, where **ε** is original metric

New Matting Metric

Similarity measurement **Instance Matching**

$$\mathrm{IMQ} = \underbrace{\frac{\sum_{\alpha,\hat{\alpha}\in TP} \mathcal{S}(\alpha,\hat{\alpha})}{|TP|}}_{\text{Matting Quality (MQ)}} \underbrace{\frac{|TP|}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|}}_{\text{Recognition Quality (RQ)}}$$

Method

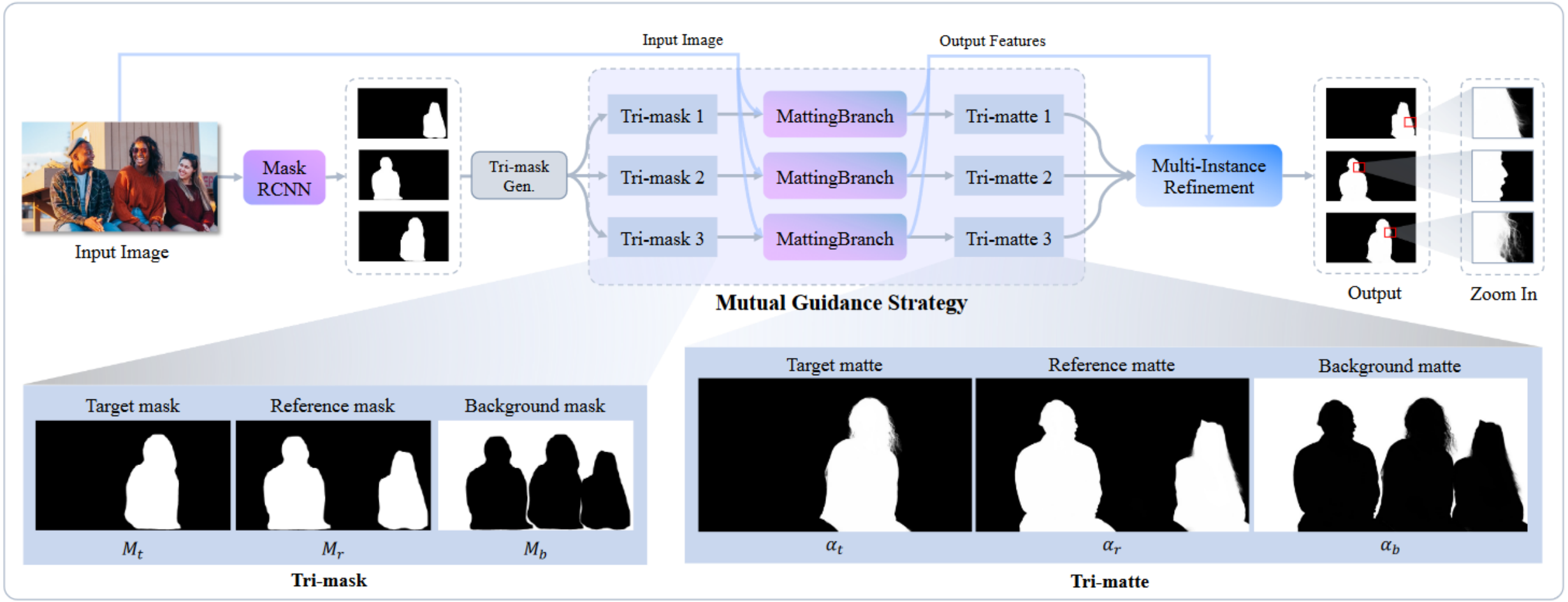


Figure 2. Overall **InstMatt** framework consisting of mutual guidance and multi-instance refinement. We first apply MaskRCNN to obtain instance masks, and then generate **tri-mask** for each instance to provide **mutual guidance** for the matting branch. Through mutual guidance strategy, we upgrade coarse tri-masks into fine **tri-mattes** for all instances. Finally, a **multi-instance refinement** module (illustrated in Figure 3) is designed to make use of the information difference of underlining tri-mattes to further promote the instance matte quality.

Method

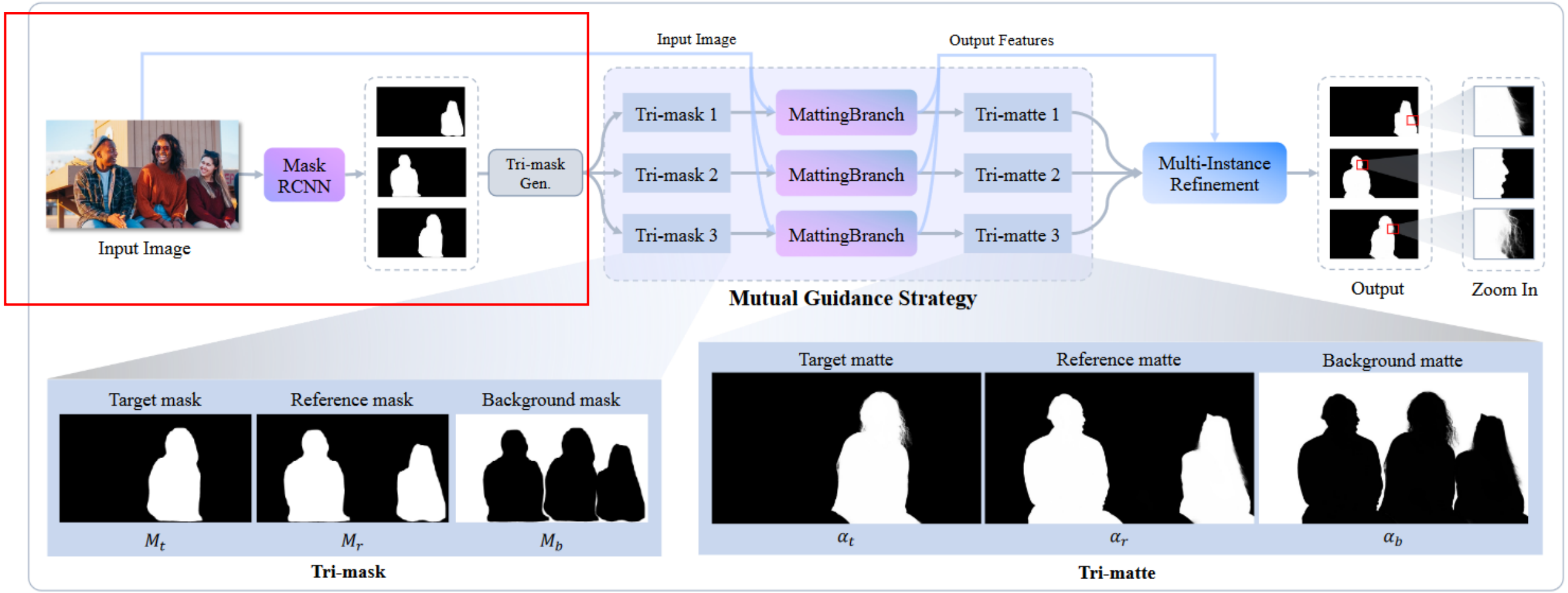


Figure 2. Overall **InstMatt** framework consisting of mutual guidance and multi-instance refinement. We first apply MaskRCNN to obtain instance masks, and then generate **tri-mask** for each instance to provide **mutual guidance** for the matting branch. Through mutual guidance strategy, we upgrade coarse tri-masks into fine **tri-mattes** for all instances. Finally, a **multi-instance refinement** module (illustrated in Figure 3) is designed to make use of the information difference of underlining tri-mattes to further promote the instance matte quality.

Method

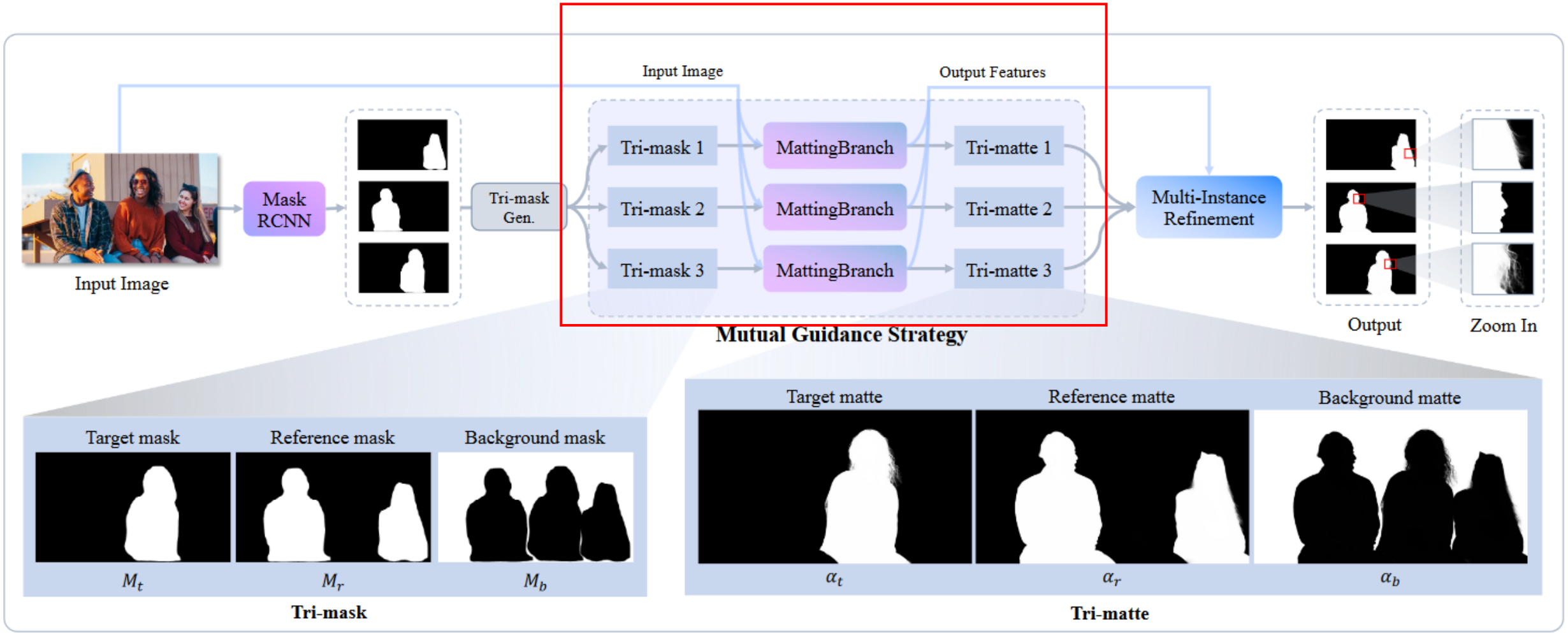


Figure 2. Overall **InstMatt** framework consisting of mutual guidance and multi-instance refinement. We first apply MaskRCNN to obtain instance masks, and then generate **tri-mask** for each instance to provide **mutual guidance** for the matting branch. Through mutual guidance strategy, we upgrade coarse tri-masks into fine **tri-mattes** for all instances. Finally, a **multi-instance refinement** module (illustrated in Figure 3) is designed to make use of the information difference of underlining tri-mattes to further promote the instance matte quality.

Method

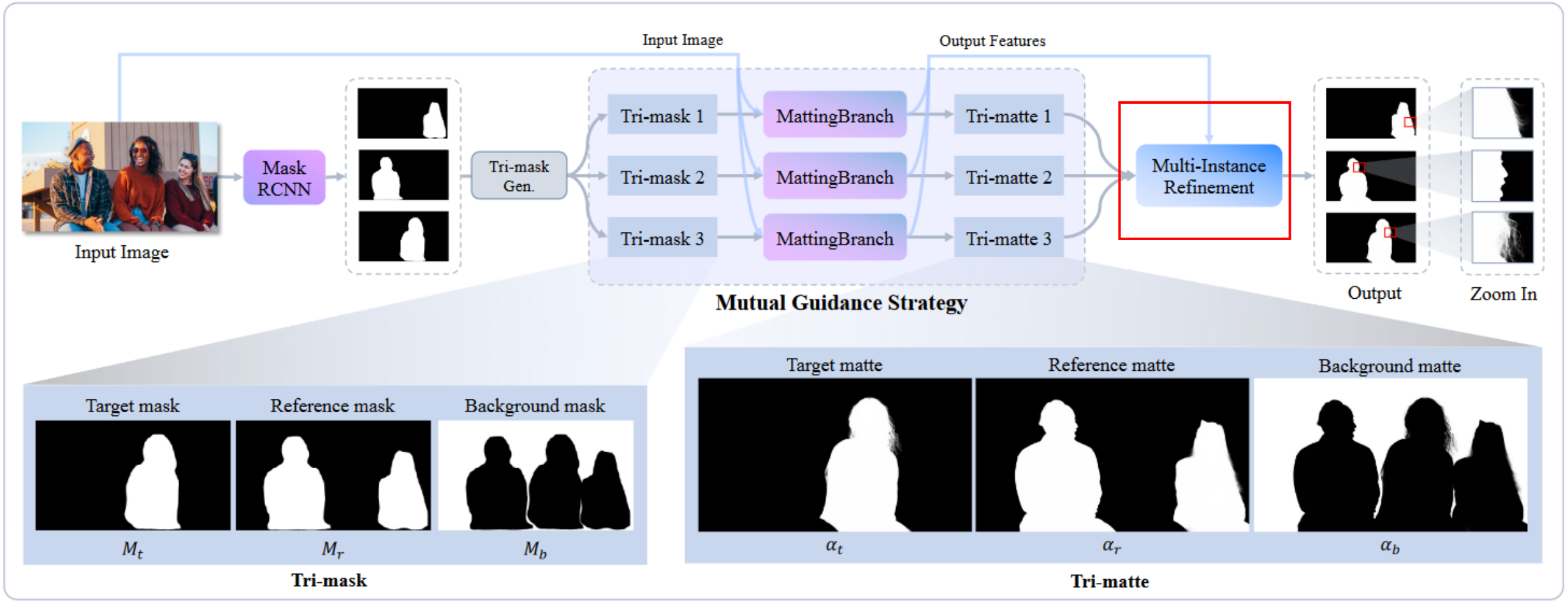


Figure 2. Overall **InstMatt** framework consisting of mutual guidance and multi-instance refinement. We first apply MaskRCNN to obtain instance masks, and then generate **tri-mask** for each instance to provide **mutual guidance** for the matting branch. Through mutual guidance strategy, we upgrade coarse tri-masks into fine **tri-mattes** for all instances. Finally, a **multi-instance refinement** module (illustrated in Figure 3) is designed to make use of the information difference of underlining tri-mattes to further promote the instance matte quality.

Method

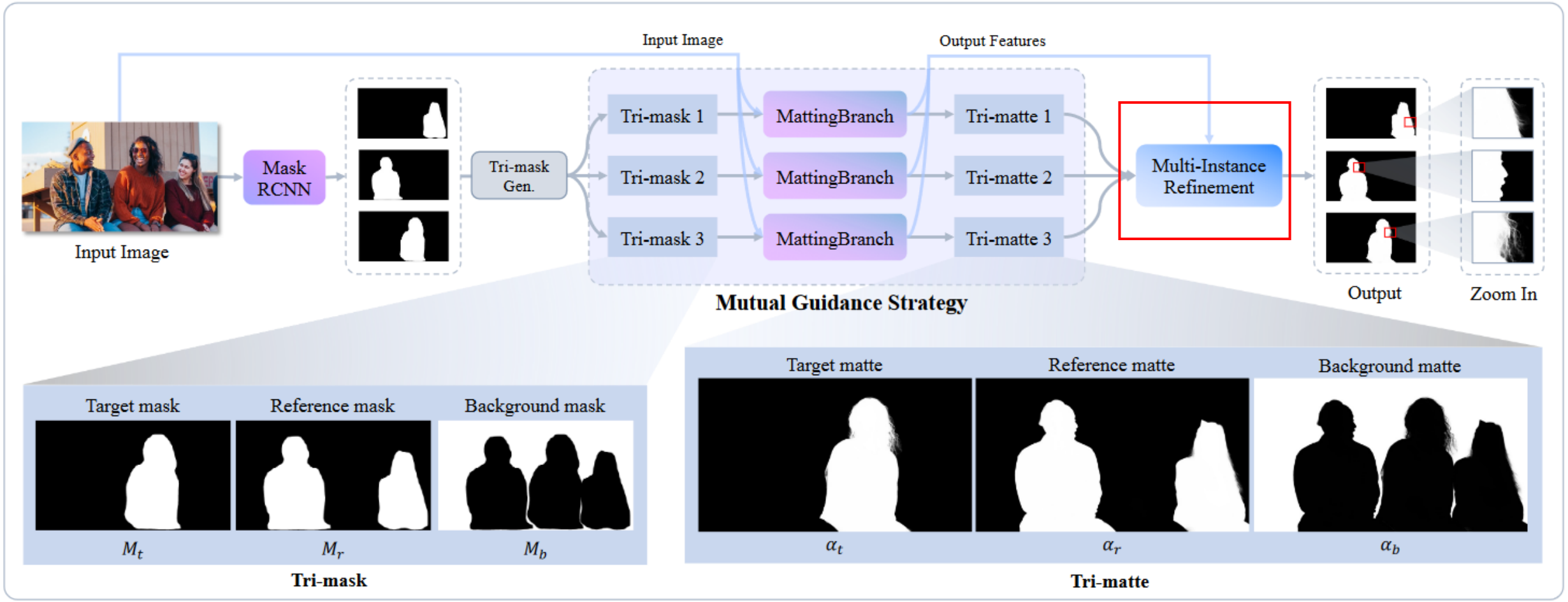
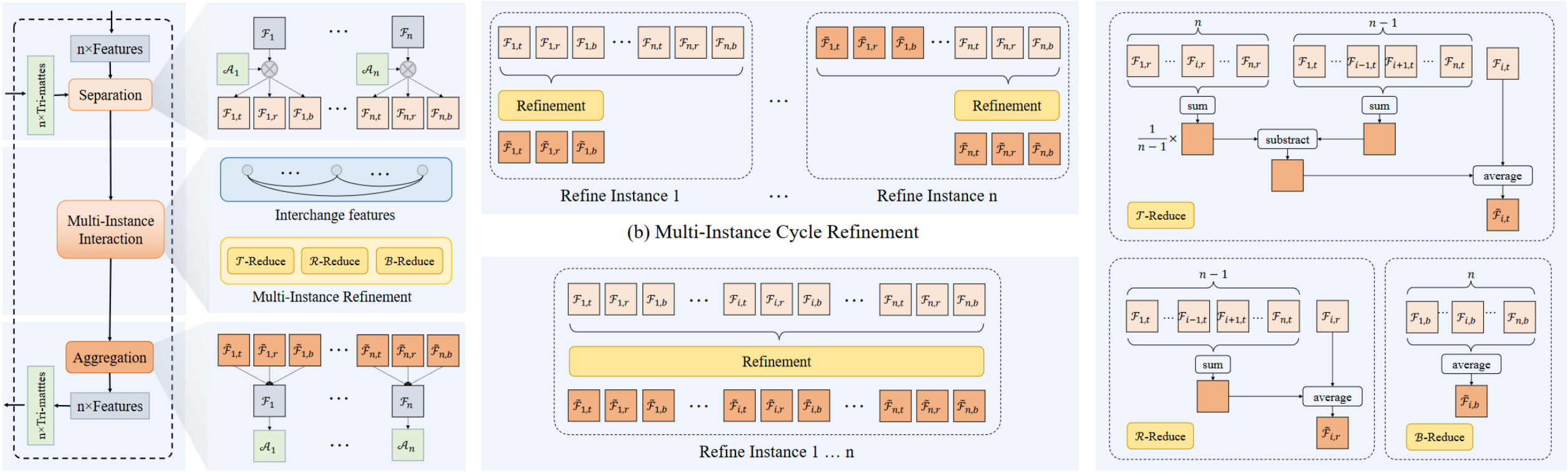


Figure 2. Overall **InstMatt** framework consisting of mutual guidance and multi-instance refinement. We first apply MaskRCNN to obtain instance masks, and then generate **tri-mask** for each instance to provide **mutual guidance** for the matting branch. Through mutual guidance strategy, we upgrade coarse tri-masks into fine **tri-mattes** for all instances. Finally, a **multi-instance refinement** module (illustrated in Figure 3) is designed to make use of the information difference of underlining tri-mattes to further promote the instance matte quality.

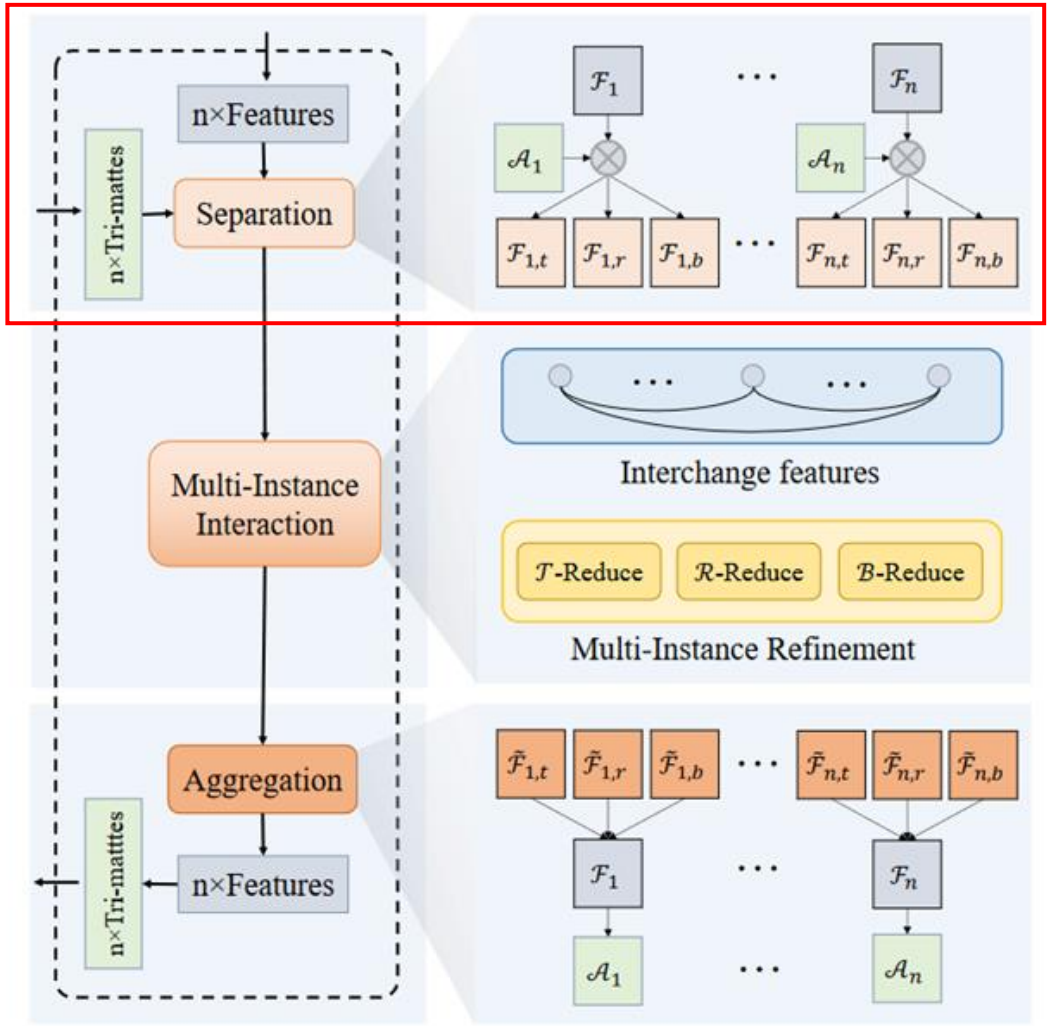
Multi Instance Refinement Module



(a) Overall (b) Multi-Instance Cycle Refinement (c) Multi-Instance Parallel Refinement (d) Feature Reduction

Figure 3. (a) Structure of our multi-instance refinement module, where instances exchange information among each other to refine their features through a multi-instance interaction layer. Two representative multi-instance refinement strategies, i.e. (b) cycle refinement and (c) parallel refinement are proposed and discussed. Figure (d) illustrates three feature reduction operations used in the two refinement ways.

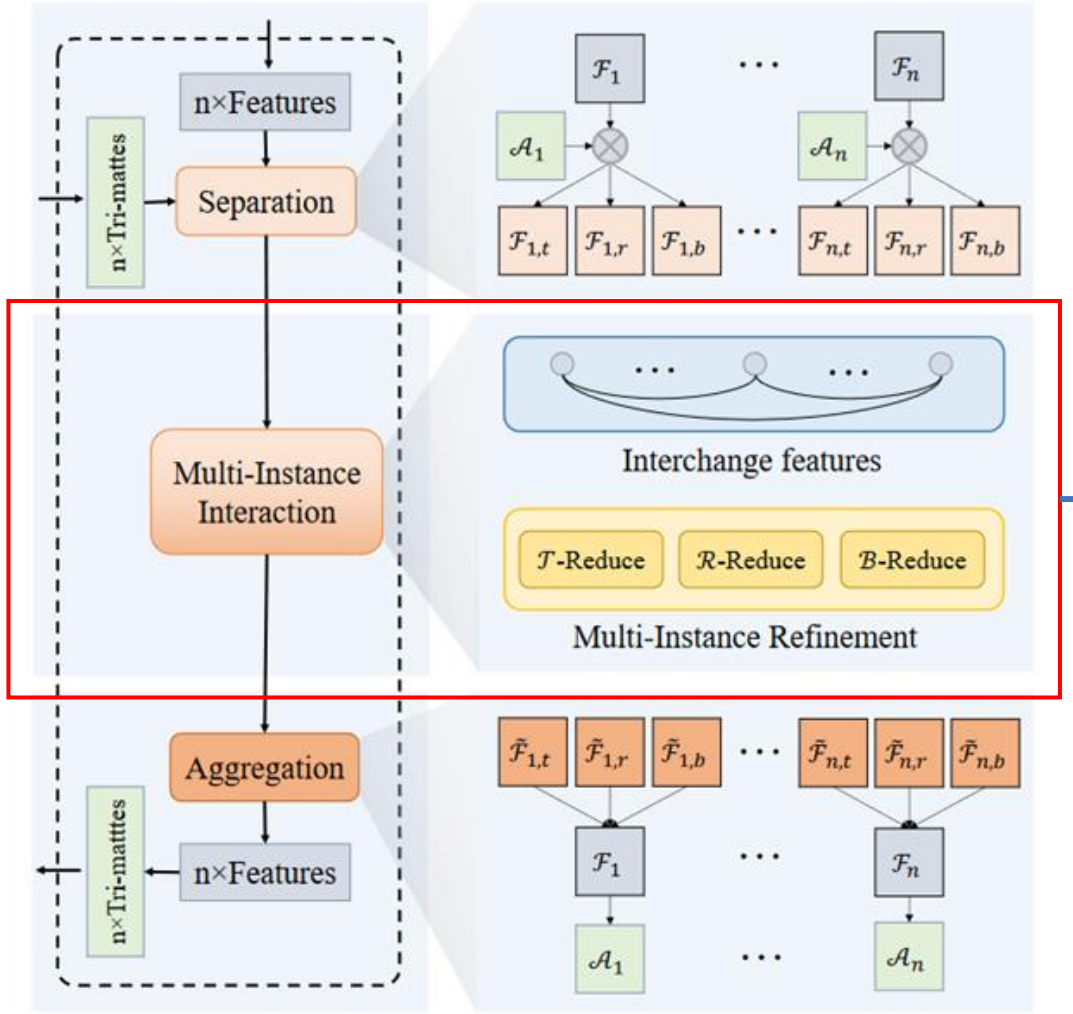
Multi Instance Refinement Module



(a) Overall

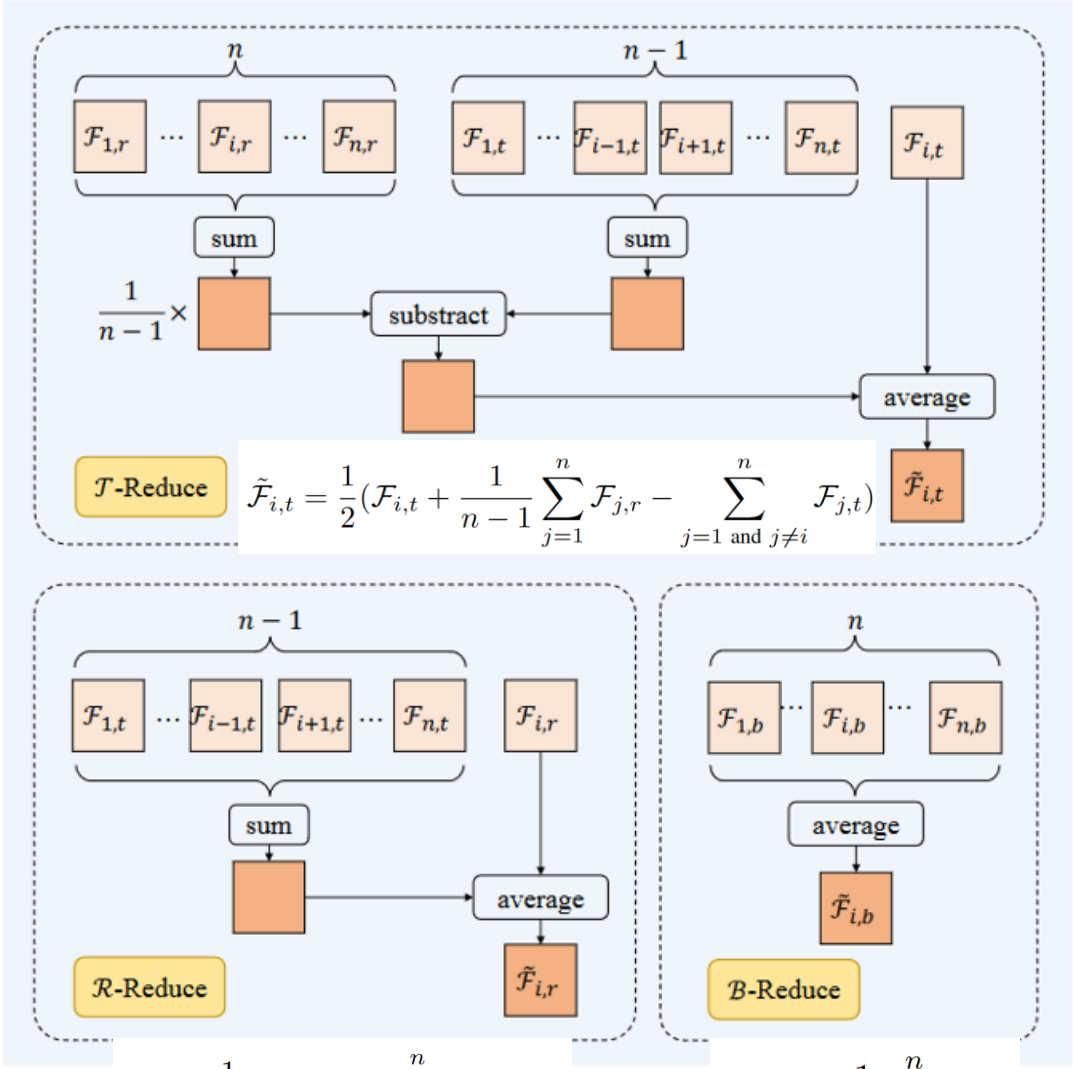
Feature Separation

Multi Instance Refinement Module



(a) Overall

Interaction

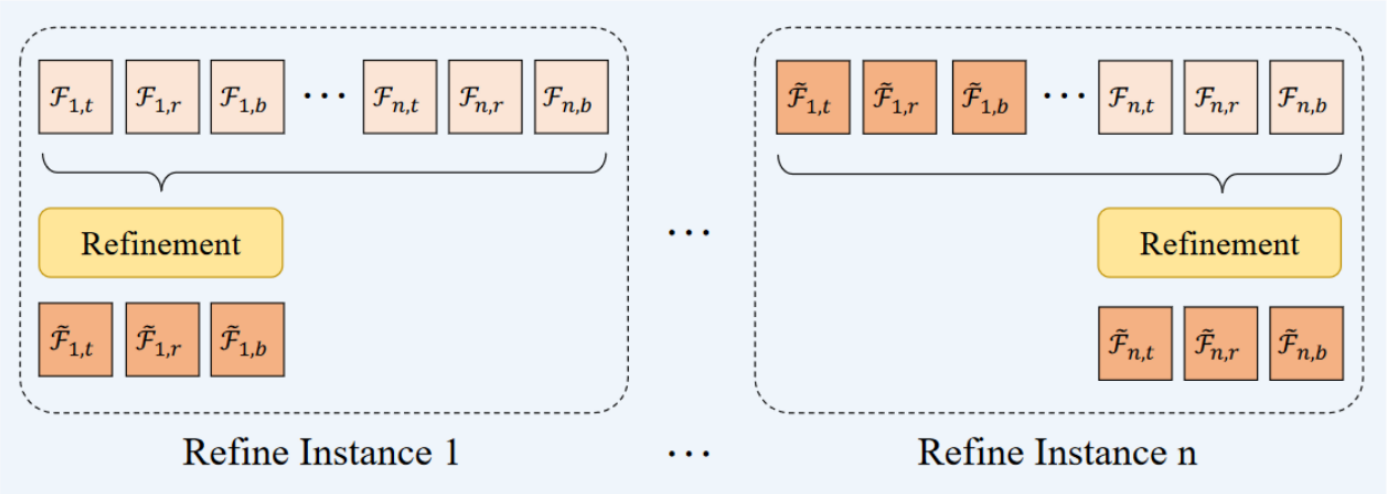


$$\tilde{\mathcal{F}}_{i,t} = \frac{1}{2}(\mathcal{F}_{i,t} + \frac{1}{n-1}\sum_{j=1}^{n}\mathcal{F}_{j,r} - \sum_{j=1 \text{ and } j\neq i}^{n}\mathcal{F}_{j,t})$$

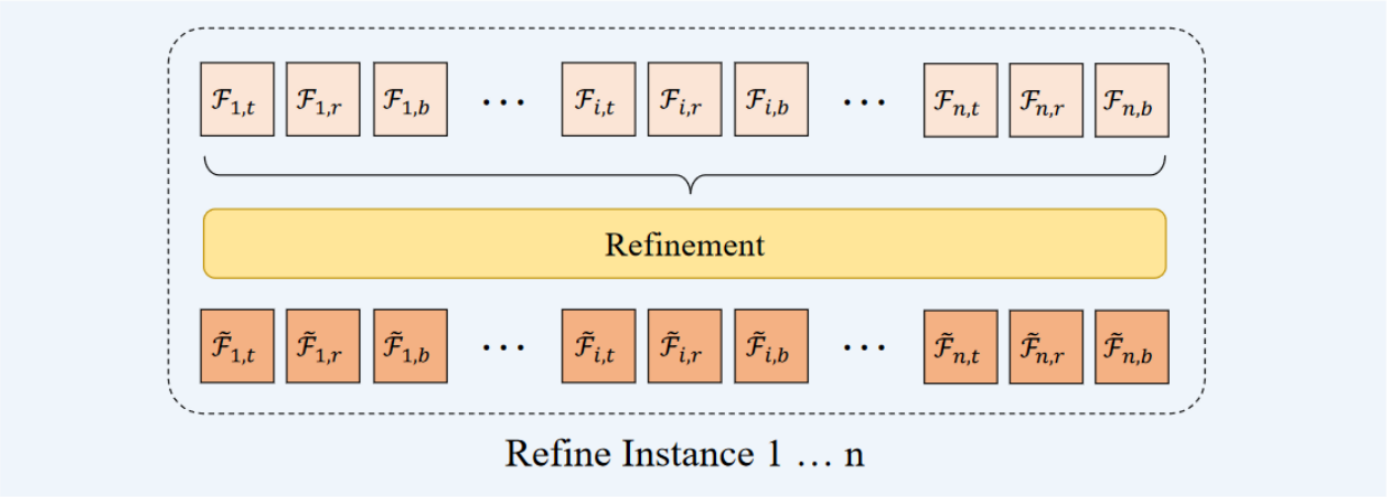
$$\tilde{\mathcal{F}}_{i,r} = \frac{1}{2}(\mathcal{F}_{i,r} + \sum_{j=1 \text{ and } j\neq i}^{n}\mathcal{F}_{j,t})$$

$$\tilde{\mathcal{F}}_{i,b} = \frac{1}{n}\sum_{j=1}^{n}\mathcal{F}_{j,b}$$

Multi Instance Refinement Module

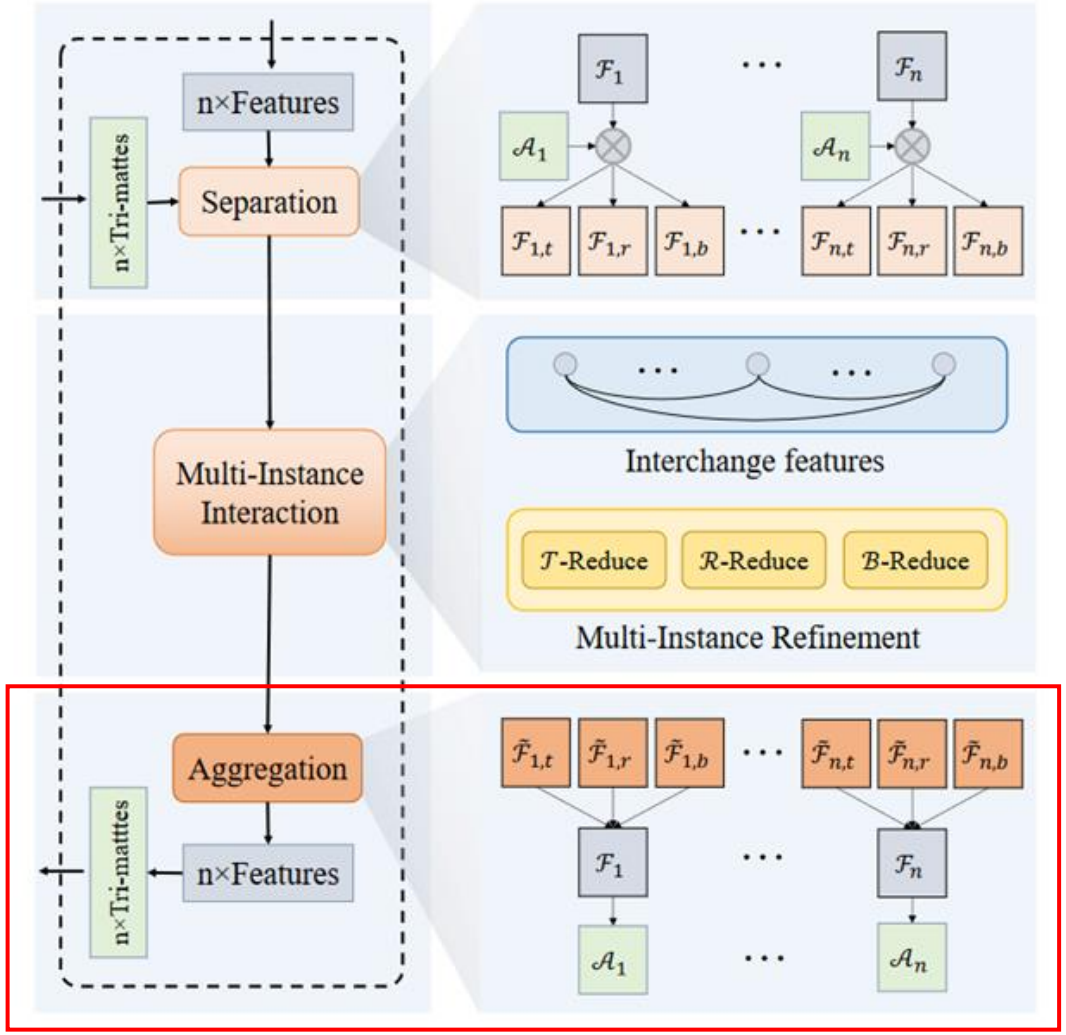


(b) Multi-Instance Cycle Refinement



(c) Multi-Instance Parallel Refinement

Multi Instance Refinement Module



Feature Aggregation

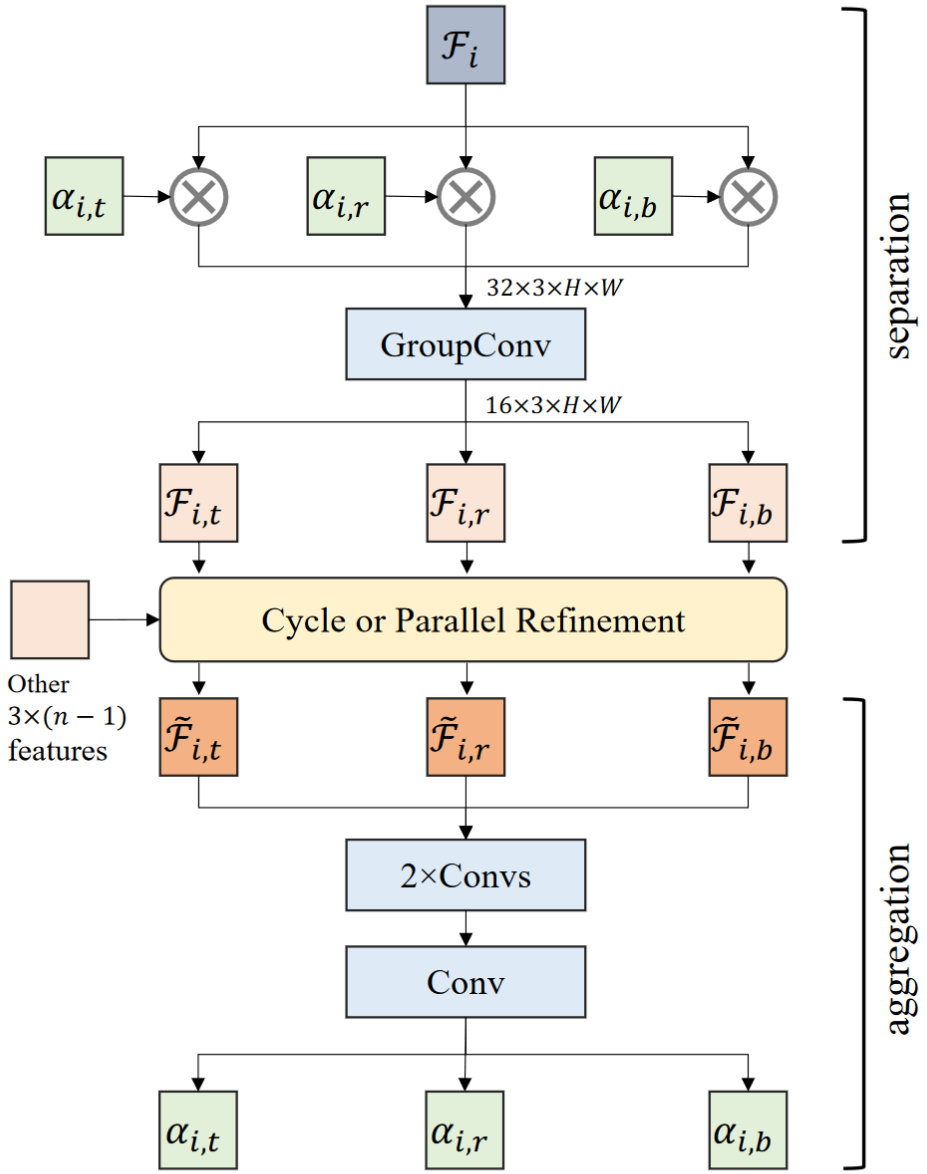


Figure 10. The structure of multi-instance refinement.

Loss Functions

- Traditional Alpha Loss
  - L1 loss
  - Pyramid Laplacian Loss
- Instance Matting Composition Loss

$$L_{mc} = ||\alpha_t F_t + \alpha_r F_r + \alpha_b F_b - I||_1$$

- Resolution Space Restriction Loss

$$L_{m\alpha} = ||\alpha_t + \alpha_r + \alpha_b - 1||_1$$

- Total Loss for Each Instance:

$$L = L_\alpha + L_{lap} + L_{mc} + L_{m\alpha}$$

Experiments

Method	HIM2K (Synthetic Subset)				HIM2K (Natural Subset)				RWP636	
	IMQ_{mad}	IMQ_{mse}	IMQ_{grad}	IMQ_{conn}	IMQ_{mad}	IMQ_{mse}	IMQ_{grad}	IMQ_{conn}	IMQ_{mad}	IMQ_{mse}
MaskRCNN [23]	18.37	25.65	0.45	19.07	24.22	33.74	2.27	26.65	20.26	25.36
MaskRCNN + CascadePSP [12]	40.85	51.64	29.59	43.37	64.58	74.66	60.02	67.20	42.20	52.91
MaskRCNN + GCA [36]	37.76	51.56	38.33	39.90	45.72	61.40	44.77	48.81	33.87	46.47
MaskRCNN + SIM [52]	43.02	52.90	40.63	44.29	54.43	66.67	49.56	58.12	34.66	46.60
MaskRCNN + FBA [18]	36.01	51.44	37.86	38.81	34.81	48.32	36.29	37.23	35.00	47.54
MaskRCNN + MaskGuided [61]	51.67	67.08	53.03	55.38	57.98	71.12	66.53	60.86	30.64	53.16
InstMatt (Ours)	**63.59**	**78.14**	**64.50**	**67.71**	**70.26**	**81.34**	**74.90**	**72.60**	**51.10**	**73.09**

Table 1. Quantitative comparisons on HIM2K and RWP636 [61]. The balance factor w in Equation 17 is set to 10. For IMQ_{mad}, IMQ_{mse}, IMQ_{grad} and IMQ_{conn}, the higher, the better. Bold numbers indicate the best performance.

Method	IMQ_{mad}	IMQ_{mse}
MaskRCNN [23]	18.44	18.48
MaskRCNN + CascadePSP [12]	30.54	33.37
InstMatt (Ours)	**30.67**	**39.56**

Table 2. Quantitative results on SPD [16].

M_t	M_r	M_b	MIR	IMQ_{mad}	IMQ_{mse}
✓	✗	✗	✗	57.98	71.12
✓	✓	✗	✗	62.25	74.35
✓	✓	✓	✗	69.40	79.74
✓	✓	✓	✓	**70.26**	**81.34**

Table 3. Results on tri-mask and multi-instance refinement.

Experiments

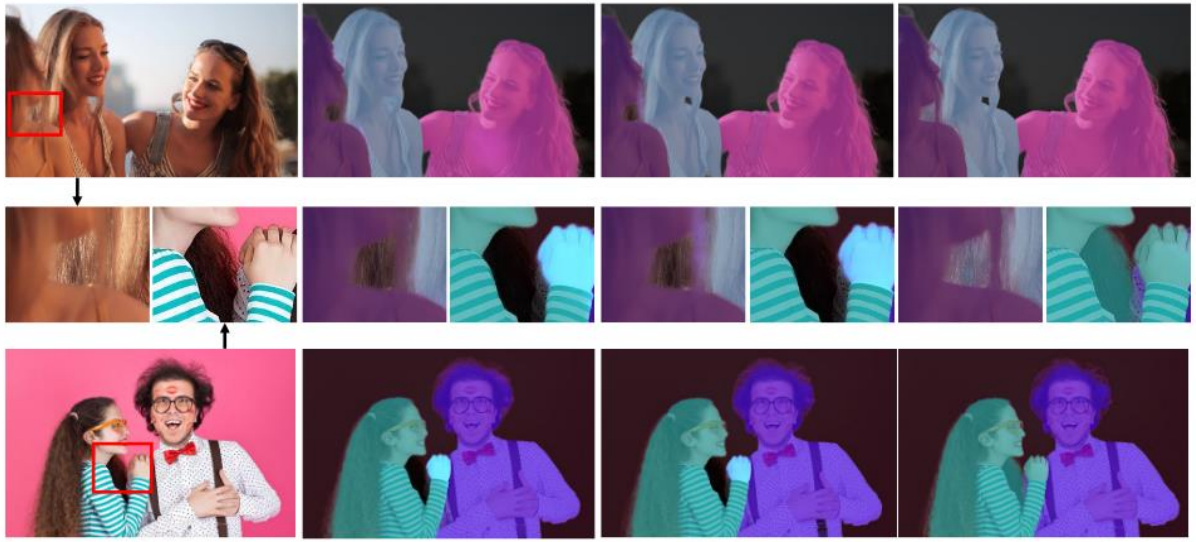


Figure 8. Comparisons among different mask guidance settings. See in particular the zoom-ins showing the best results when all three components are enabled, where the blonde's hairs and the man's shoulder are clearly delineated.

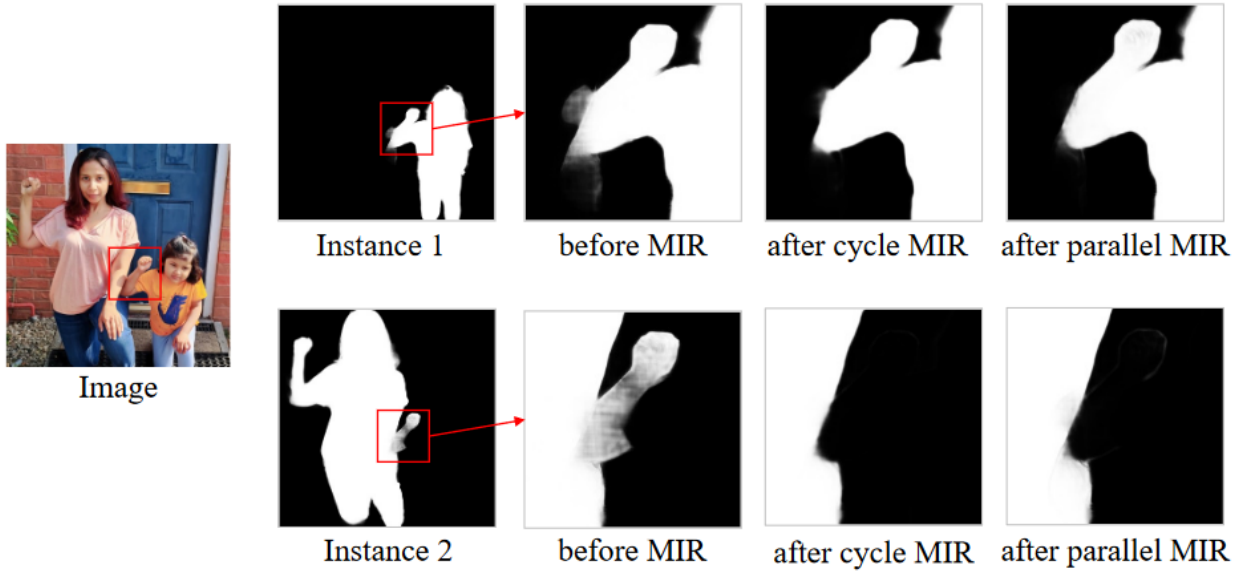


Figure 9. Alpha matte before and after multi-instance refinement.

Experiments

Image MaskRCNN CascadePSP SIM MaskGuided InstMatt (Ours)

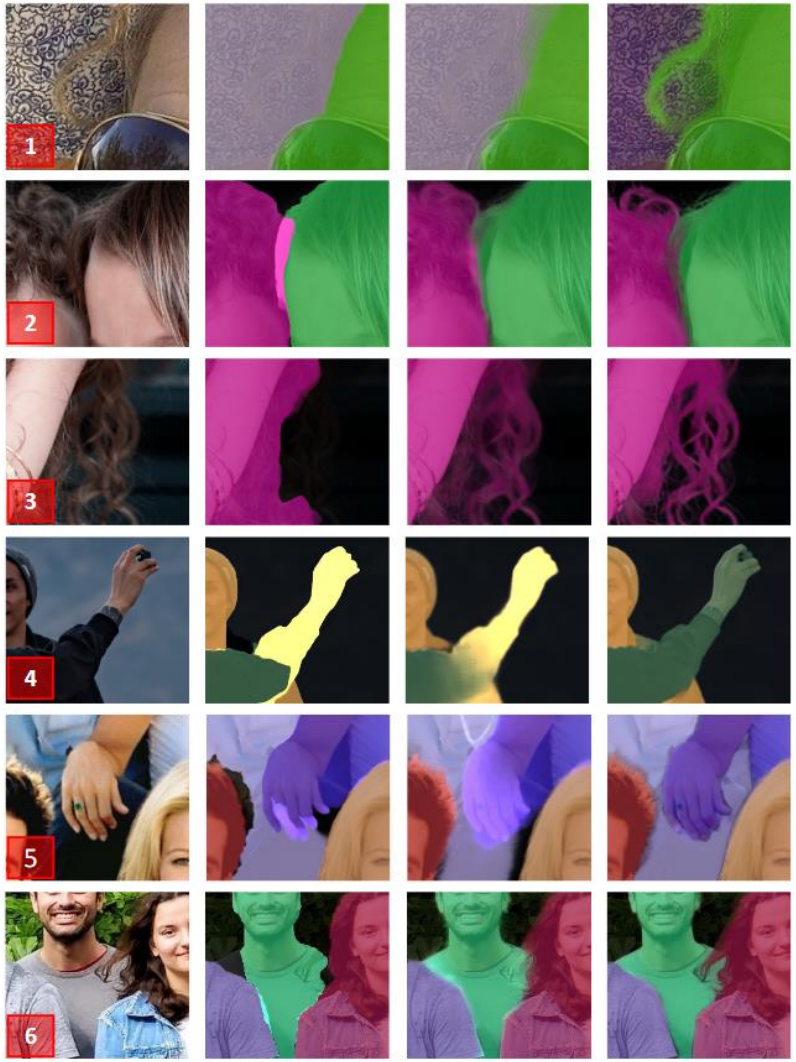


Image Patch CascadePSP MaskGuided InstMatt