Inharmonious Region Localization by Magnifying Domain Discrepancy

Jing Liang[1], Li Niu[1]*, Penghao Wu[1], Fengjun Guo[2], Teng Long[2]

[1]MoE Key Lab of Artificial Intelligence, Shanghai Jiao Tong University

[2]INTSIG

AAAI 2022

Motivation

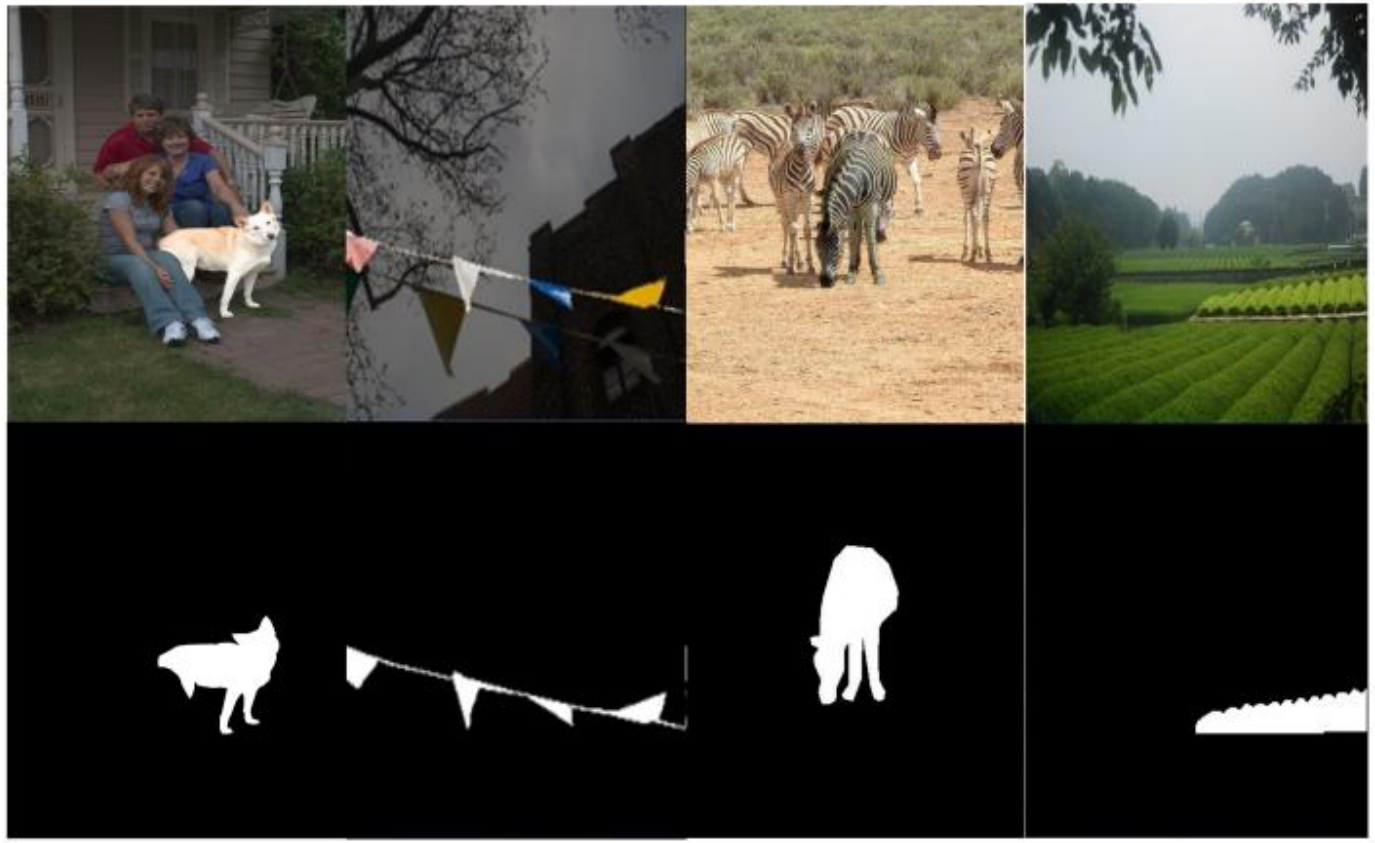


Figure 1: We show the examples of inharmonious synthetic images in the top row and their inharmonious region masks in the bottom row.

Motivation

- The only existing inharmonious region localization method is DIRL, which attempted to fuse multi-scale features and avoid redundant information. However, DIRL is a rather general model without exploiting the uniqueness of this task, that is, the discrepancy between inharmonious region and background.
- We refer to each suite of color and illumination statistics as one domain following. Thus, the inharmonious region and the background belong to two different domains.

To achieve this goal, we propose a framework composed of two components: one **color mapping module** and one inharmonious region localization network.

Contributions

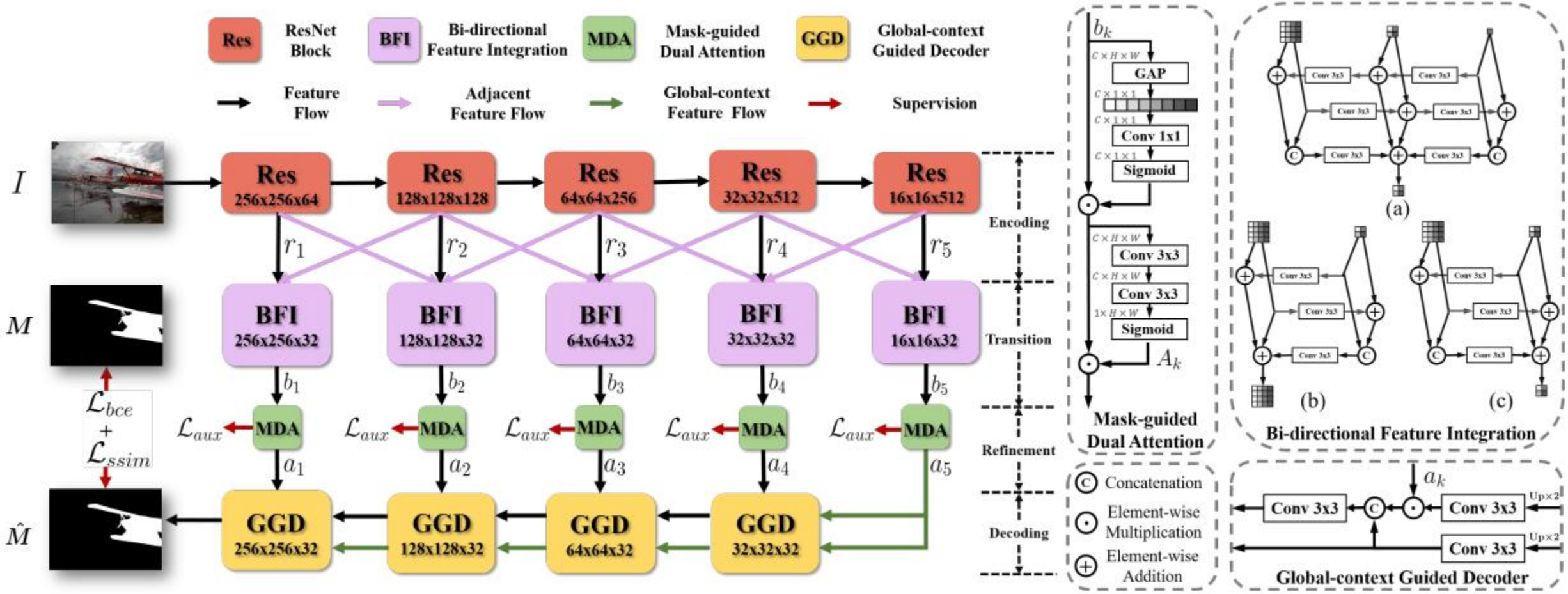
(a) We devise a simple yet effective inharmonious region localization framework which can accommodate any region localization method.

(b) We are the first to introduce adaptive color transformation to inharmonious region localization, in which improve HDRNet is used as the color mapping module.

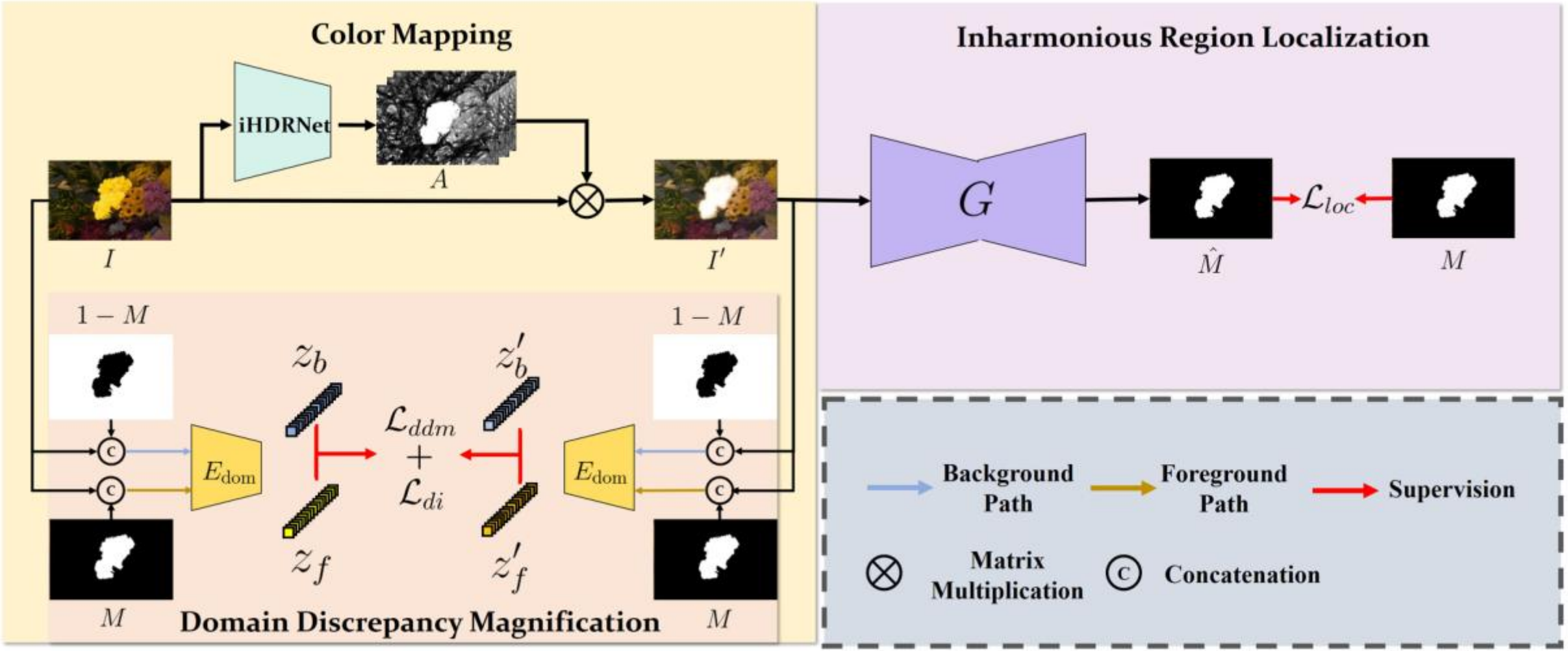
(c) We propose a novel domain discrepancy magnification loss to magnify the domain discrepancy between inharmonious region and background.

(d) Extensive experiments demonstrate that our framework outperforms existing methods by a large margin (e.g., IoU is improved from 67.85% to 74.44%).

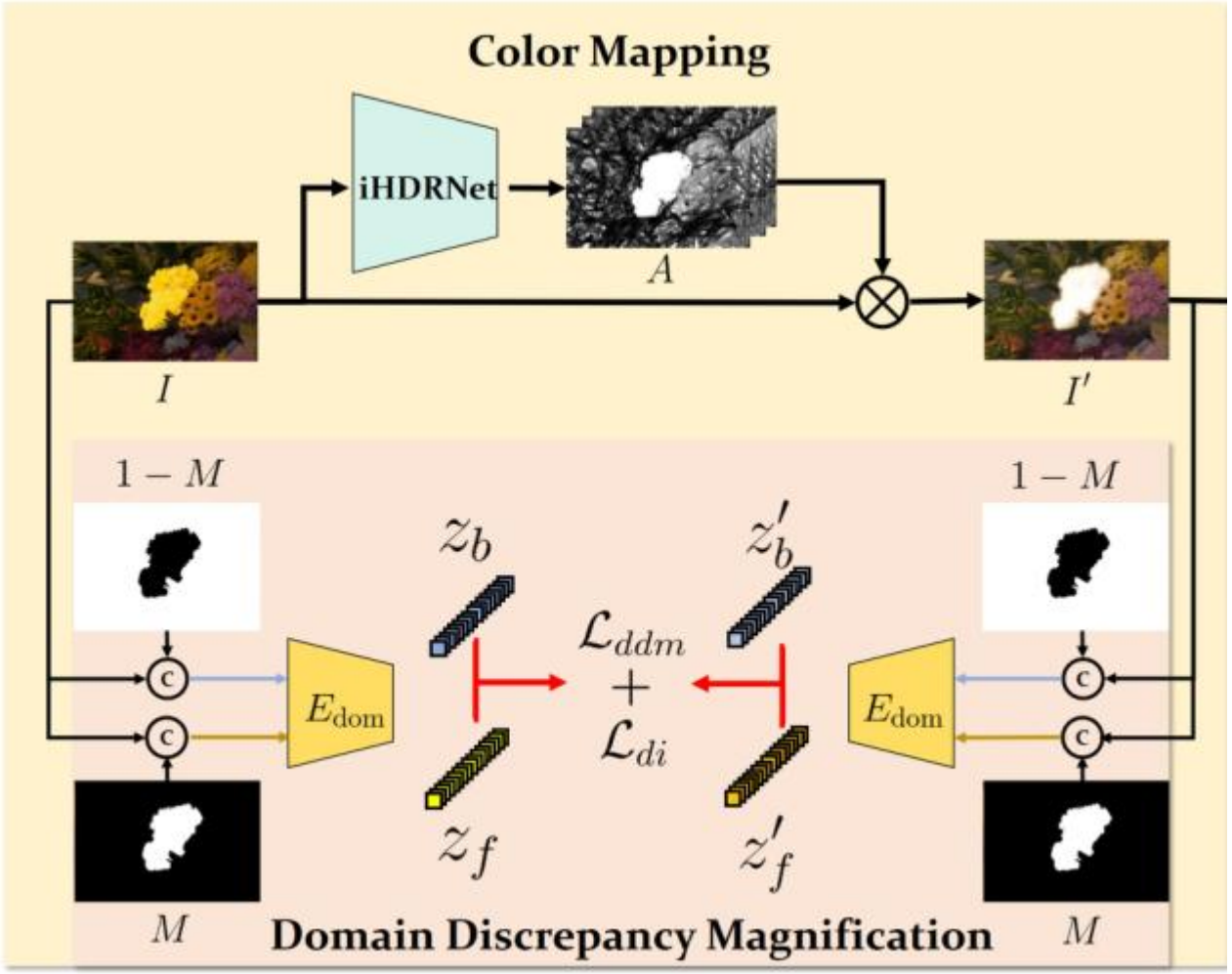
DIRL


Res
ResNet Block
BFI
Bi-directional Feature Integration
MDA
Mask-guided Dual Attention
GGD
Global-context Guided Decoder
Feature Flow
Adjacent Feature Flow
Global-context Feature Flow
Supervision
I
M
$\hat{M}$
Res 256x256x64
Res 128x128x128
Res 64x64x256
Res 32x32x512
Res 16x16x512
BFI 256x256x32
BFI 128x128x32
BFI 64x64x32
BFI 32x32x32
BFI 16x16x32
MDA
GGD 256x256x32
GGD 128x128x32
GGD 64x64x32
GGD 32x32x32
r_1 r_2 r_3 r_4 r_5
b_1 b_2 b_3 b_4 b_5
a_1 a_2 a_3 a_4 a_5
$\mathcal{L}_{aux}$
$\mathcal{L}_{bce} + \mathcal{L}_{ssim}$
Encoding
Transition
Refinement
Decoding
GAP
Conv 1x1
Sigmoid
Conv 3x3
Sigmoid
A_k
Mask-guided Dual Attention
(a)
(b)
(c)
Bi-directional Feature Integration
Concatenation
Element-wise Multiplication
Element-wise Addition
Up×2
Global-context Guided Decoder

MadisNet


Color Mapping
iHDRNet
A
I
I'
Inharmonious Region Localization
G
$\hat{M}$
$\mathcal{L}_{loc}$
M
1 − M
1 − M
z_b
z'_b
$\mathcal{L}_{ddm}$
+
$\mathcal{L}_{di}$
E_{dom}
E_{dom}
z_f
z'_f
M
Domain Discrepancy Magnification
M
Background Path
Foreground Path
Supervision
Matrix Multiplication
Concatenation

Methodology



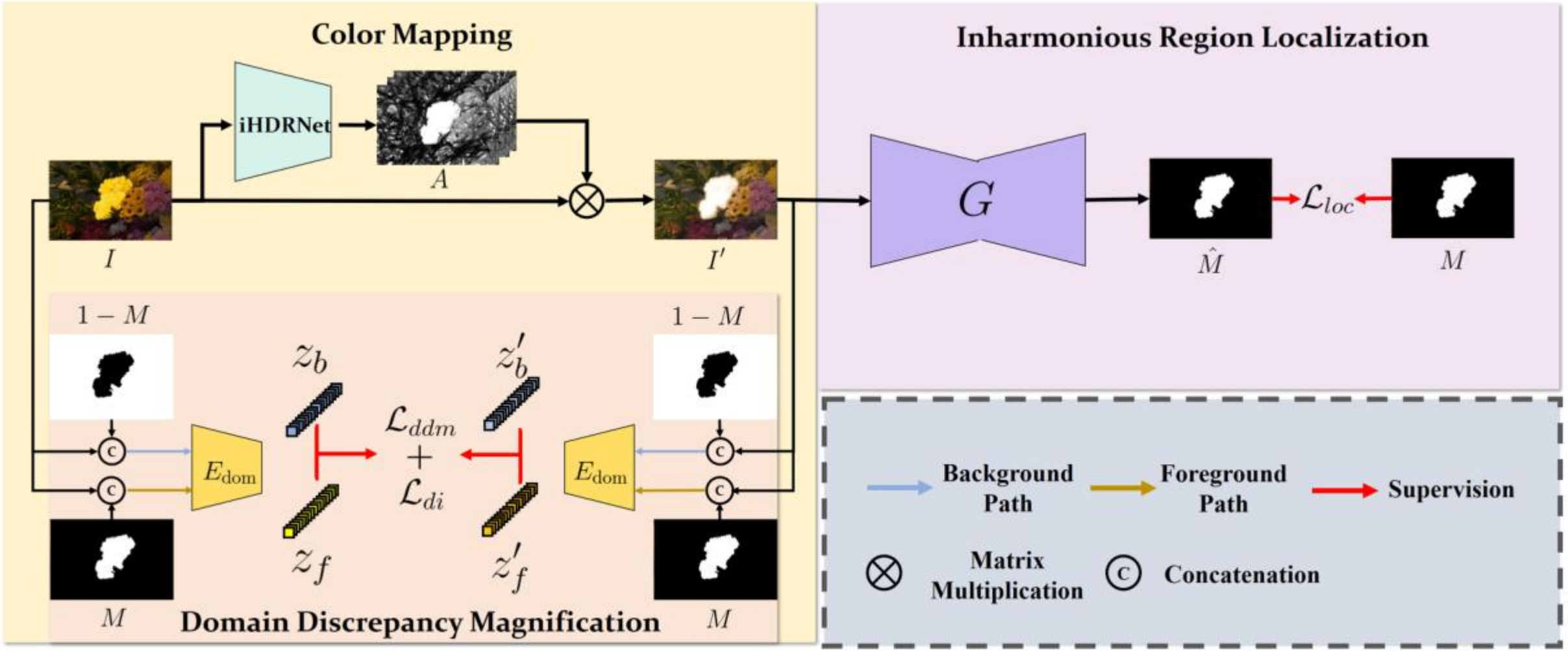
Domain Discrepancy Magnification (DDM) loss:

$$\mathcal{L}_{ddm} = \max\left(d(\boldsymbol{z}_f, \boldsymbol{z}_b) - d(\boldsymbol{z}'_f, \boldsymbol{z}'_b) + m, 0\right), \quad (1)$$

Direction Invariance Loss:

$$\mathcal{L}_{di} = 1 - \langle \Delta z, \Delta z' \rangle, \quad (2)$$

MadisNet



$$\mathcal{L}_{total} = \lambda_{ddm}\mathcal{L}_{ddm} + \lambda_{di}\mathcal{L}_{di} + \mathcal{L}_{loc}, \qquad (3)$$

Experiments

Methods	Evaluation Metrics		
	AP(%) ↑	F_1 ↑	IoU(%) ↑
UNet	74.90	0.6717	64.74
DeepLabv3	75.69	0.6902	66.01
HRNet-OCR	75.33	0.6765	65.49
MFCN	45.63	0.3794	28.54
ManTraNet	64.22	0.5691	50.31
MAGritte	71.16	0.6907	60.14
H-LSTM	60.21	0.5239	47.07
SPAN	65.94	0.5850	54.27
F3Net	61.46	0.5506	47.48
GATENet	62.43	0.5296	46.33
MINet	77.51	0.6822	63.04
S2AM	43.77	0.3029	22.36
DIRL	80.02	0.7317	67.85
MadisNet(UNet)	81.15	0.7372	67.28
MadisNet(DIRL)	**85.86**	**0.8022**	**74.44**

Table 1: Quantitative comparison with baseline methods on iHarmony4 dataset. The best results are denoted in boldface.

Experiments

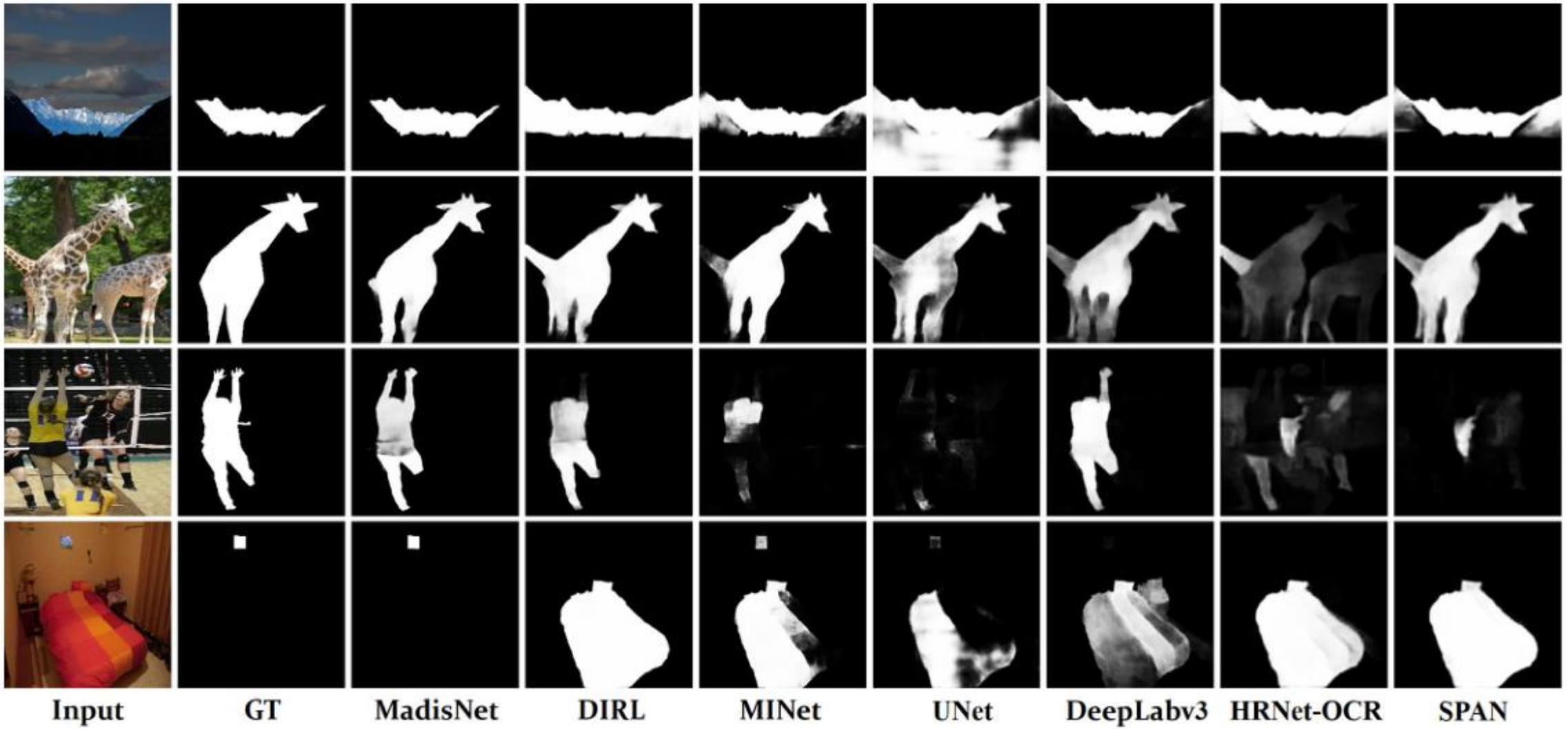


Figure 3: Qualitative comparison with baseline methods. GT is the ground-truth inharmonious region mask.

Q&A

Thank you!