



Deep Image Composition

牛力

<http://www.ustcnewly.com>



上海交通大學

SHANGHAI JIAO TONG UNIVERSITY



Image Composition



- **Overall**
- Image harmonization
- Shadow generation
- Object placement
- Foreground object search
- Generative composition



Image Composition



Image Composition: address the inconsistency between foreground and background.

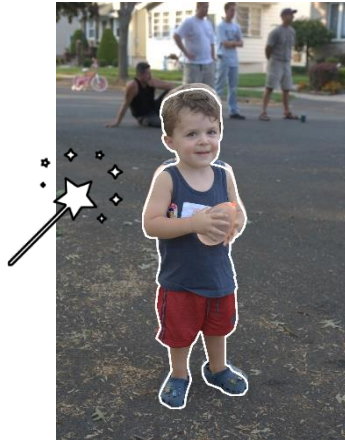
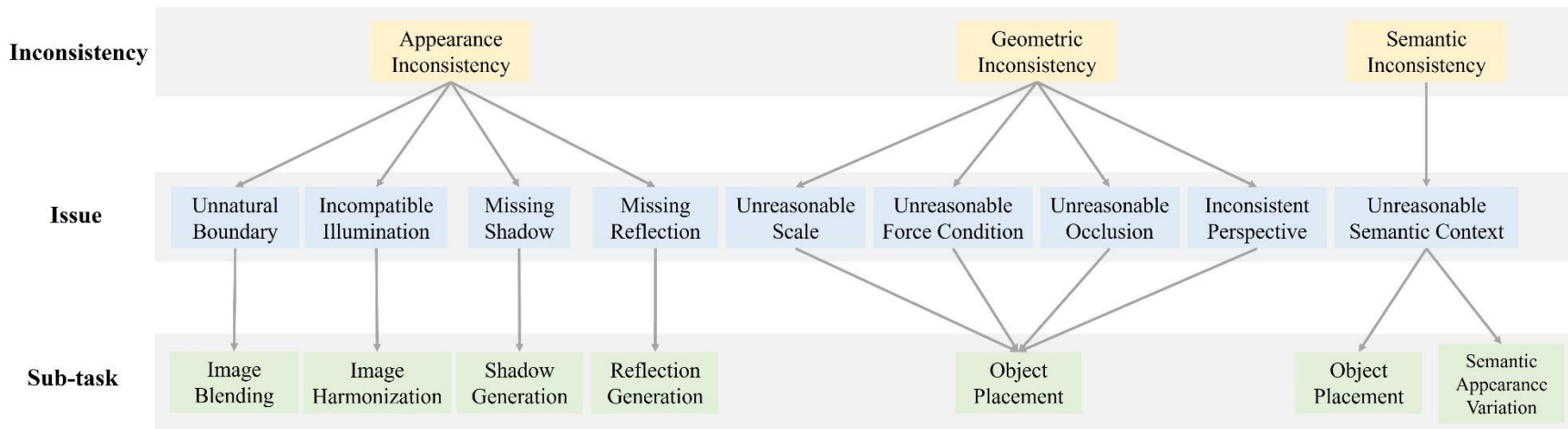




Image Composition



	2019	2020	2021	2022	2023	2024	2025
图像和谐化							
艺术图像和谐化							
物体放置							
阴影生成							
倒影生成							
生成式图像合成							



Image Composition



Potential Applications



virtual social
events

post design

advertisement
image/video generation



Image Composition



- **Overall**
- **Image harmonization**
- Shadow generation
- Object placement
- Foreground object search
- Generative composition

Image Harmonization



Given a composite image, image harmonization aims to making the pasted foreground compatible with the background.



Unrealistic composite image

Image
Harmonization



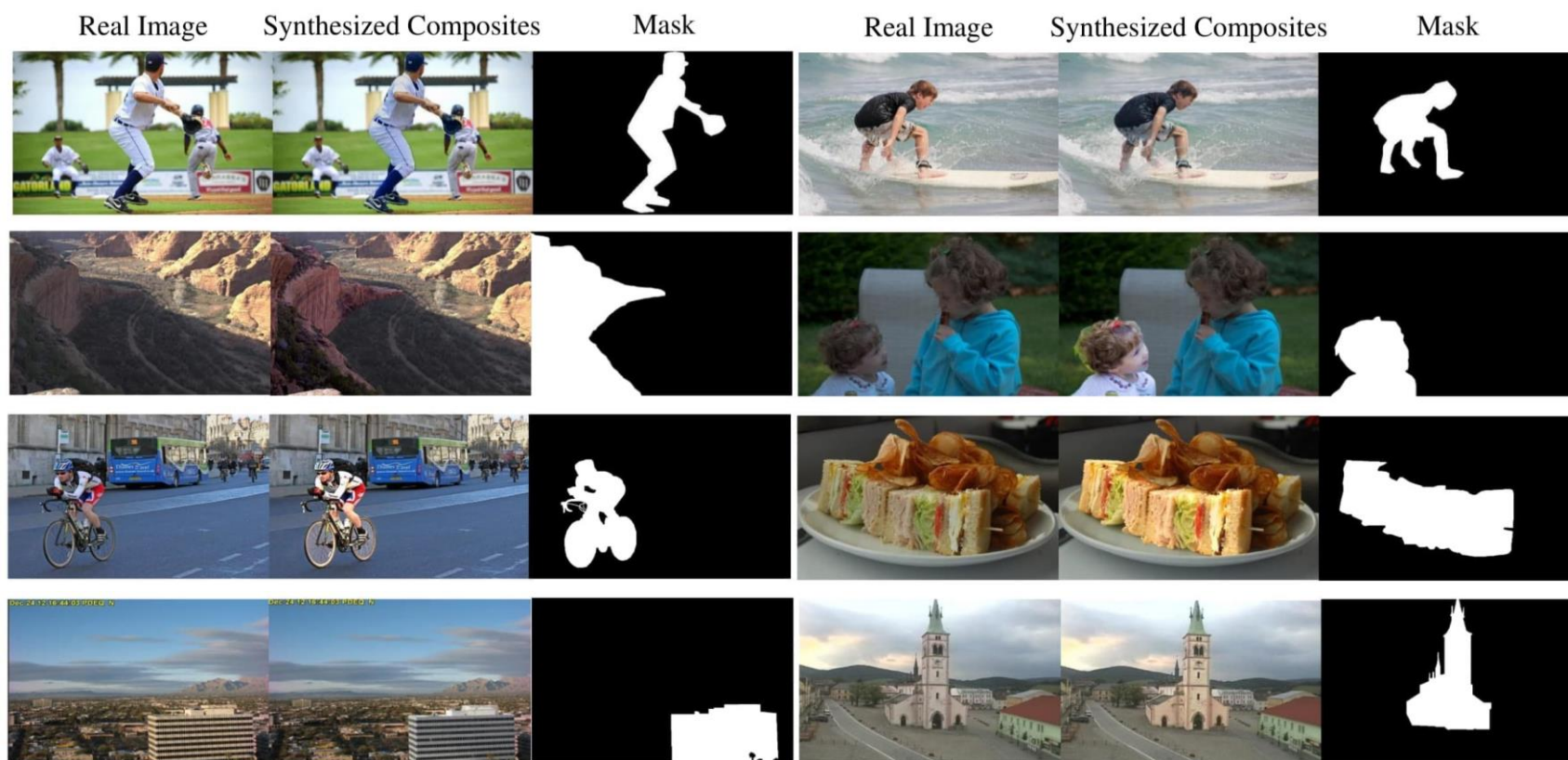
Realistic harmonized result



Image Harmonization



We release **the first large-scale image harmonization dataset** iHarmony 4.



Wenyan Cong, Jianfu Zhang, **Li Niu**, Liu Liu, Zhixin Ling, Weiyuan Li, Liqing Zhang, "DoveNet: Deep Image Harmonization via Domain Verification", **CVPR, 2020**.

Image Harmonization

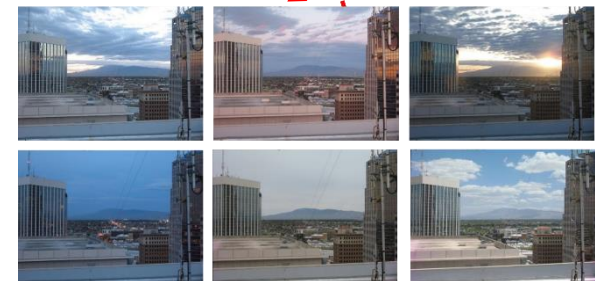


iHarmony4 has four subdatasets: HCOCO, HAdobe5k, HFlickr, Hday2night

	HCOCO	HAdobe5k	HFlickr	Hday2night	iHarmony4
Training set	38545	19437	7449	311	65742
Test set	4283	2160	828	133	7404



Long-term capture of the same scene



Wenyan Cong, Jianfu Zhang, **Li Niu**, Liu Liu, Zhixin Ling, Weiyuan Li, Liqing Zhang, "DoveNet: Deep Image Harmonization via Domain Verification", **CVPR, 2020**.

Image Harmonization



DoVeNet: Domain Verification Network



- We leverage an attention enhanced generator and two complementary discriminators.
- Domain verification discriminator** extracts the domain representations of foreground and background, and matches the foreground domain to background domain.

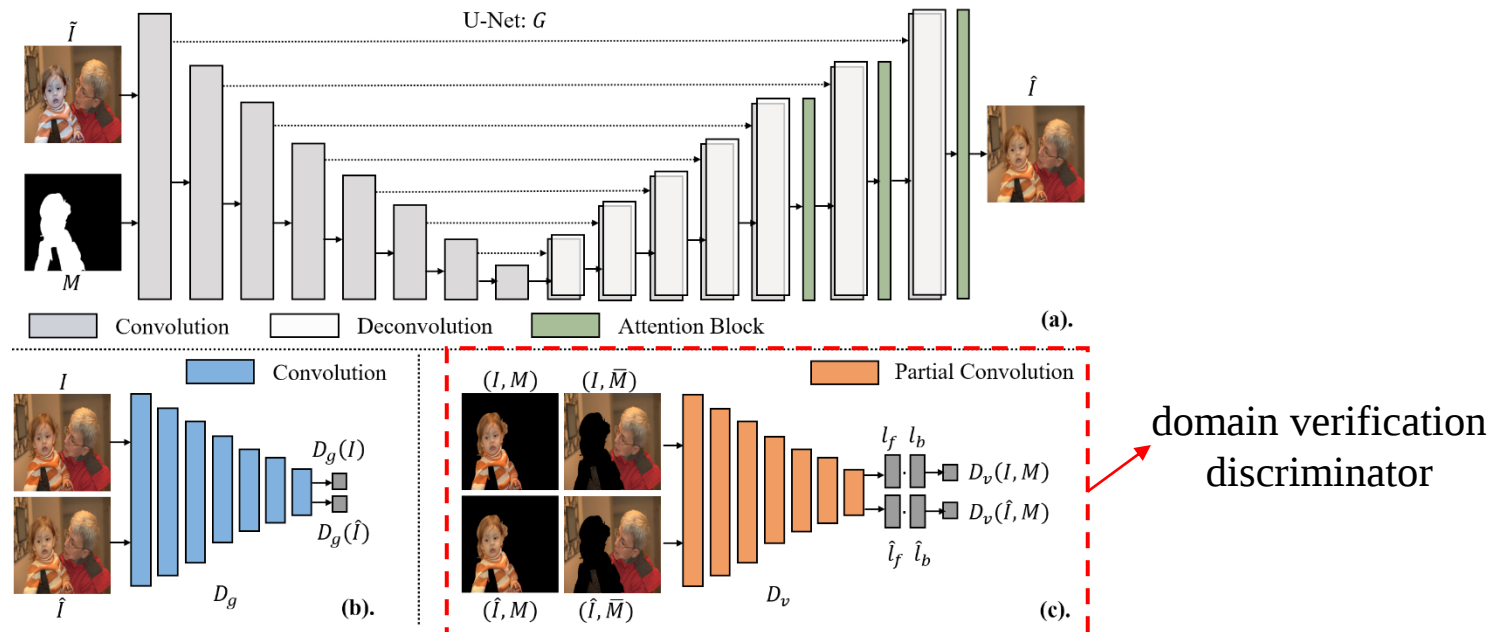
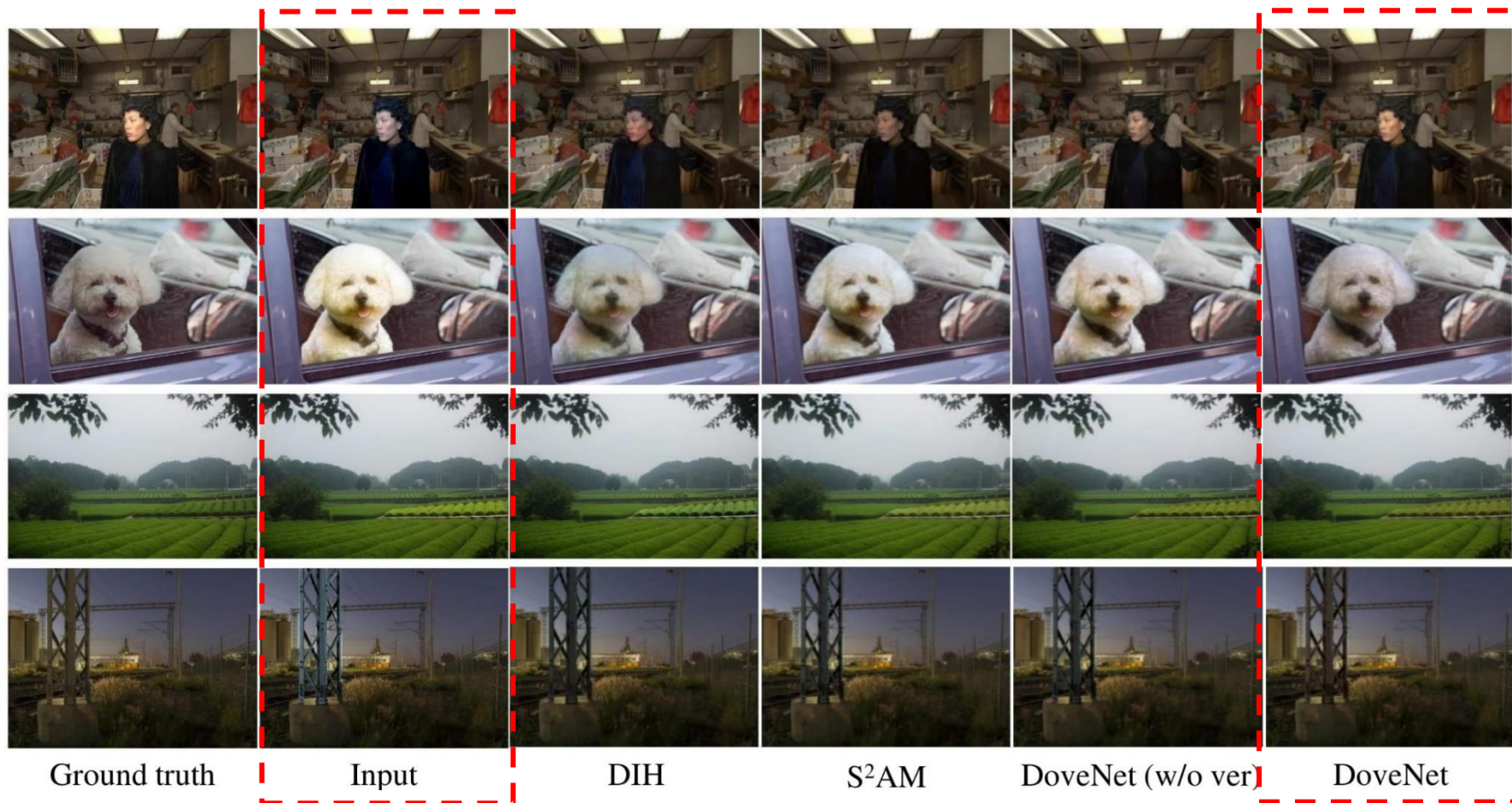




Image Harmonization

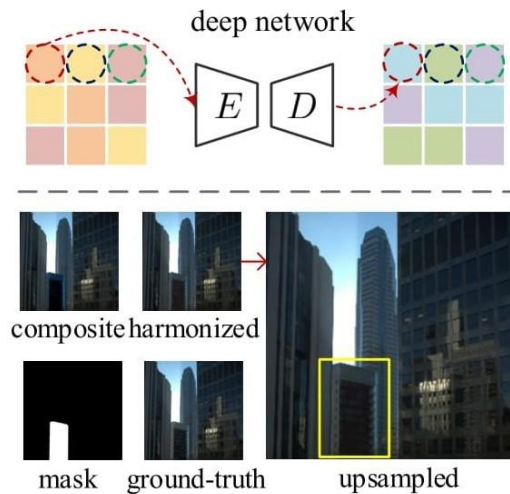


Wenyan Cong, Jianfu Zhang, **Li Niu**, Liu Liu, Zhixin Ling, Weiyuan Li, Liqing Zhang, "DoveNet: Deep Image Harmonization via Domain Verification", **CVPR, 2020**.

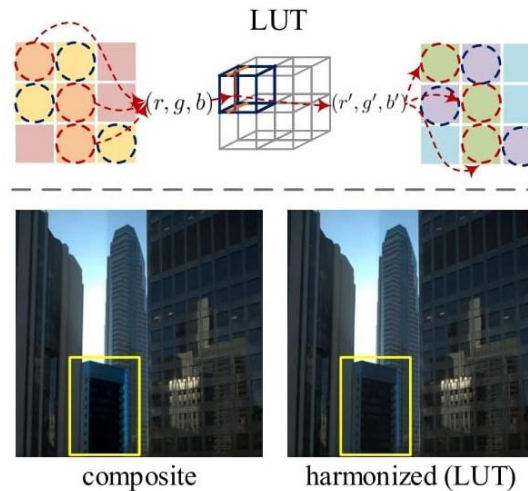
Image Harmonization



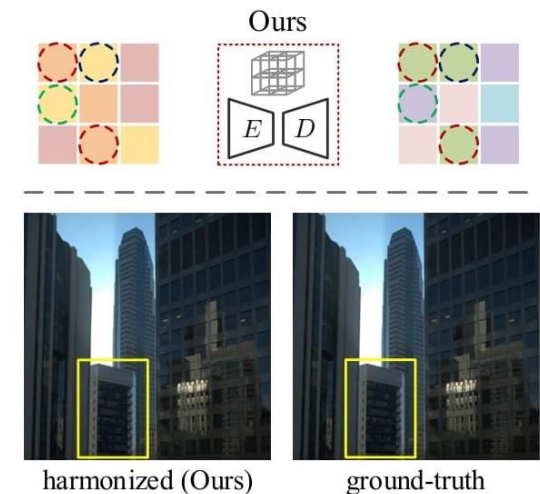
Pixel-to-pixel transformation



Color-to-color transformation



Dual transformation (Ours)



- **The first work on high-resolution image harmonization.**
- Color-to-color transformation is more efficient than pixel-to-pixel transformation.

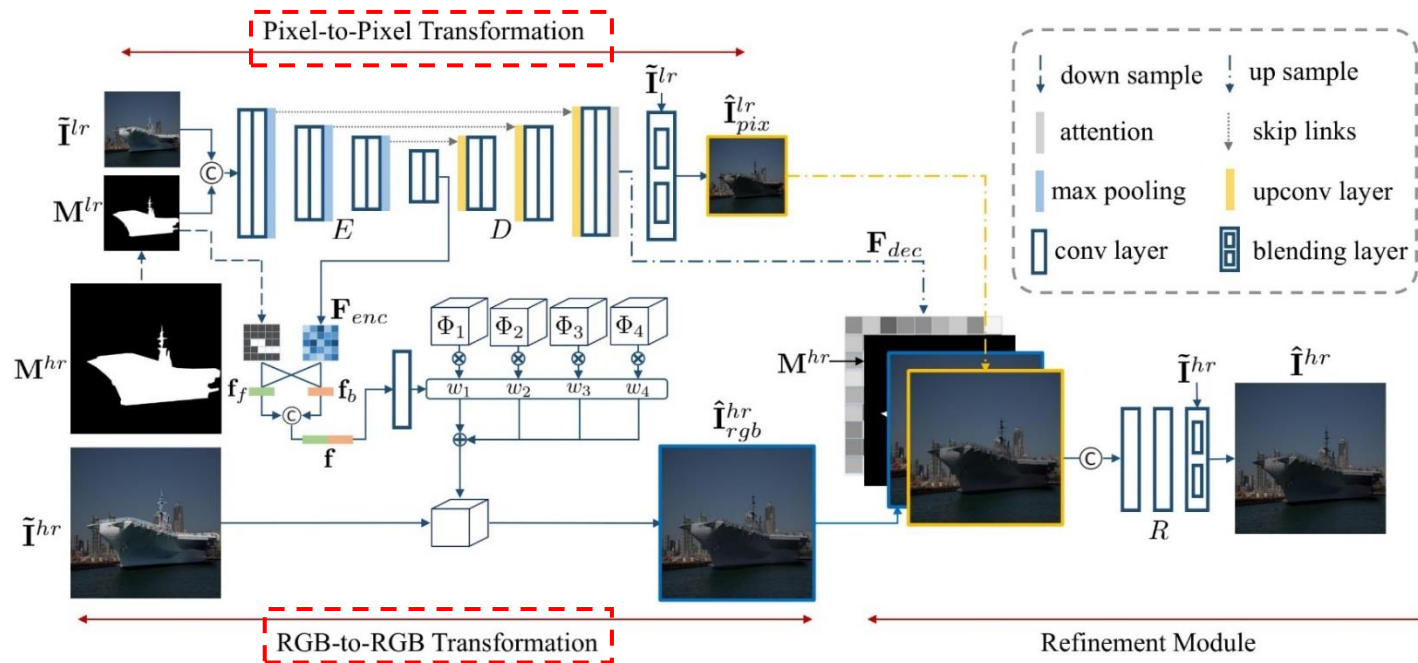
Wenyan Cong, Xinhao Tao, **Li Niu**, Jing Liang, Xuesong Gao, Qihao Sun, Liqing Zhang, "High-Resolution Image Harmonization via Collaborative Dual Transformations", **CVPR, 2022.**

Image Harmonization

Unify color-to-color transformation and pixel-to-pixel transformation under one framework.

Pixel-to-pixel: UNet

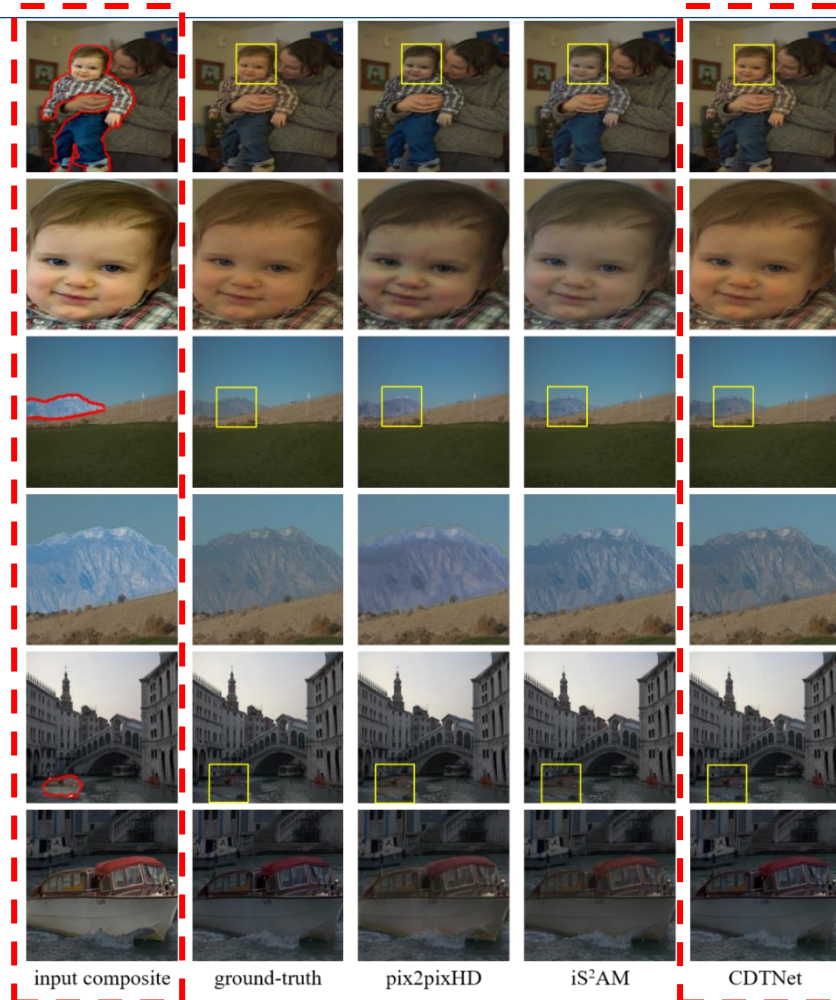
Color-to-color: predict the combination coefficients of basis LUTs



Wenyan Cong, Xinhao Tao, **Li Niu**, Jing Liang, Xuesong Gao, Qihao Sun, Liqing Zhang, "High-Resolution Image Harmonization via Collaborative Dual Transformations", **CVPR, 2022**.



Image Harmonization



Wenyan Cong, Xinhao Tao, **Li Niu**, Jing Liang, Xuesong Gao, Qihao Sun, Liqing Zhang, "High-Resolution Image Harmonization via Collaborative Dual Transformations", **CVPR, 2022**.

Image Harmonization



Use **color checker** to transit across different illumination conditions.

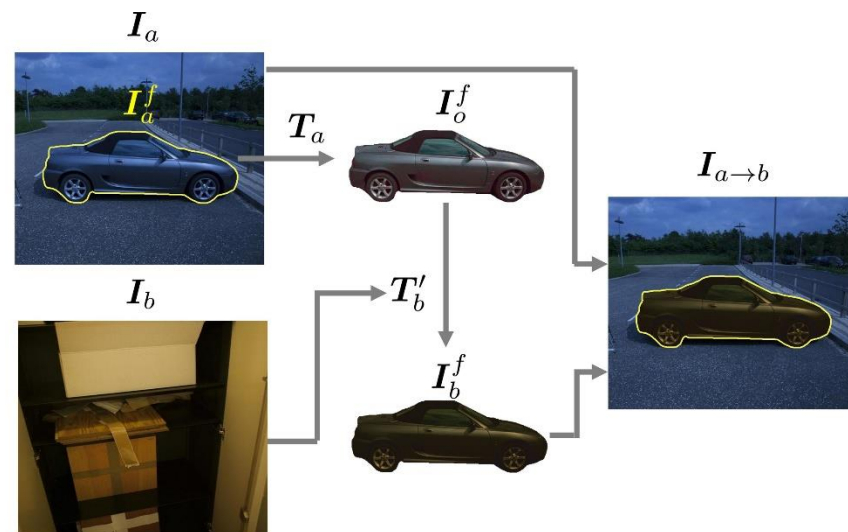
illumination A $\longrightarrow$ standard illumination $\longrightarrow$ illumination B



(a)



(b)



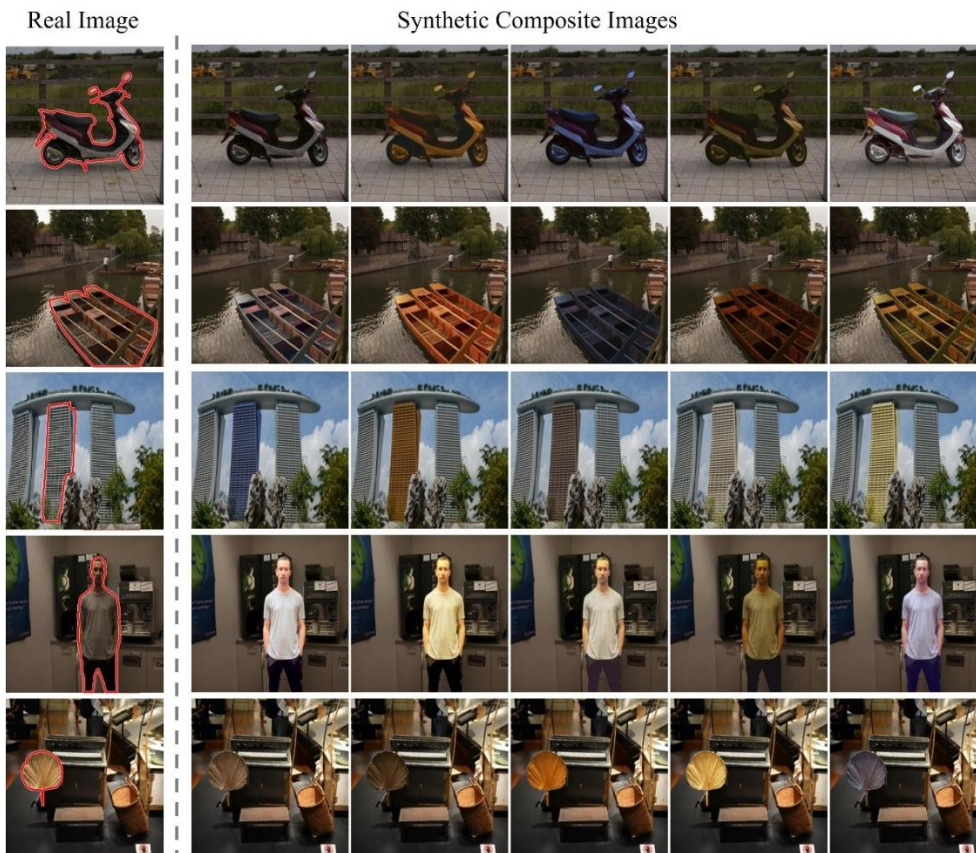
(c)



Image Harmonization



our ccHarmony dataset



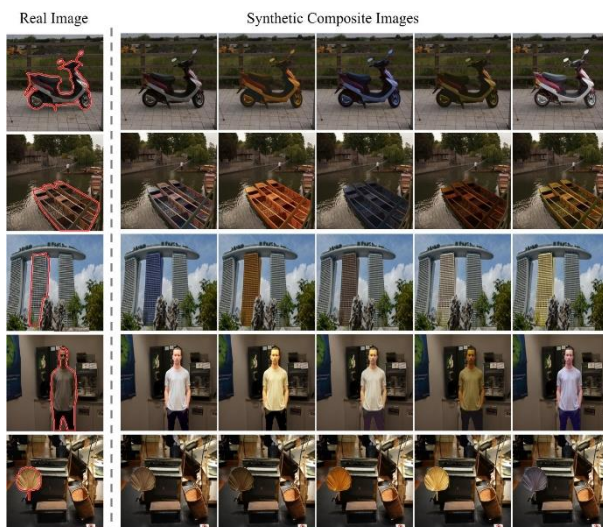
Li Niu, Linfeng Tan, Xinhao Tao, Junyan Cao, Fengjun Guo, Teng Long, Liqing Zhang, “Deep Image Harmonization with Globally Guided Feature Transformation and Relation Distillation”, ICCV, 2023.

Image Harmonization



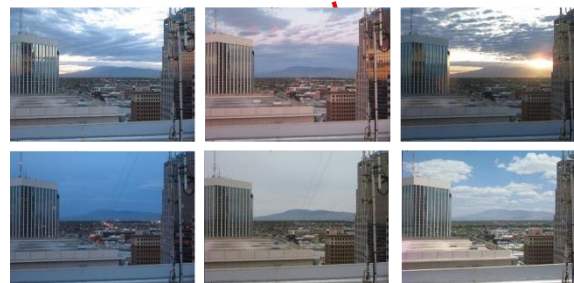
We observe that the model trained on ccHarmony performs well on Hday2night, which verifies that ccHarmony can faithfully reflect the illumination variation.

ccHarmony



Hday2night

Long-term capture of the same scene



Painterly Image Harmonization

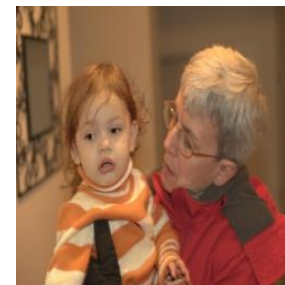


Background: photorealistic image; **Foreground:** photorealistic image



composite image

Image
Harmonization



harmonized image

Background: artistic image; **Foreground:** photorealistic image



composite image

Painterly Image
Harmonization



harmonized image

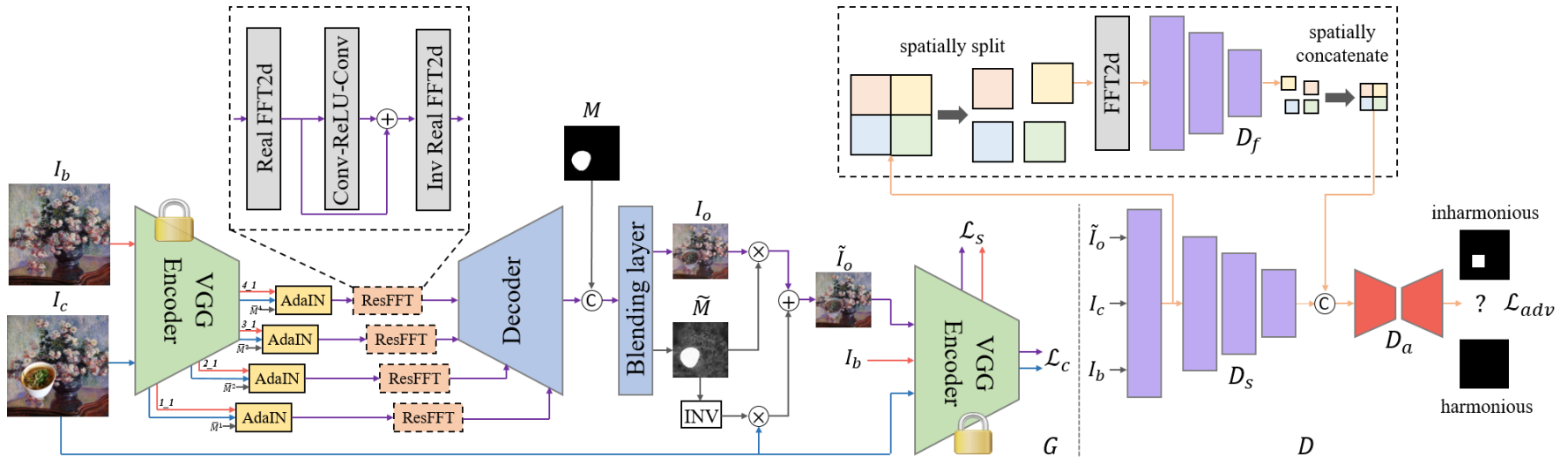
Painterly Image Harmonization



Perform harmonization in dual domains.

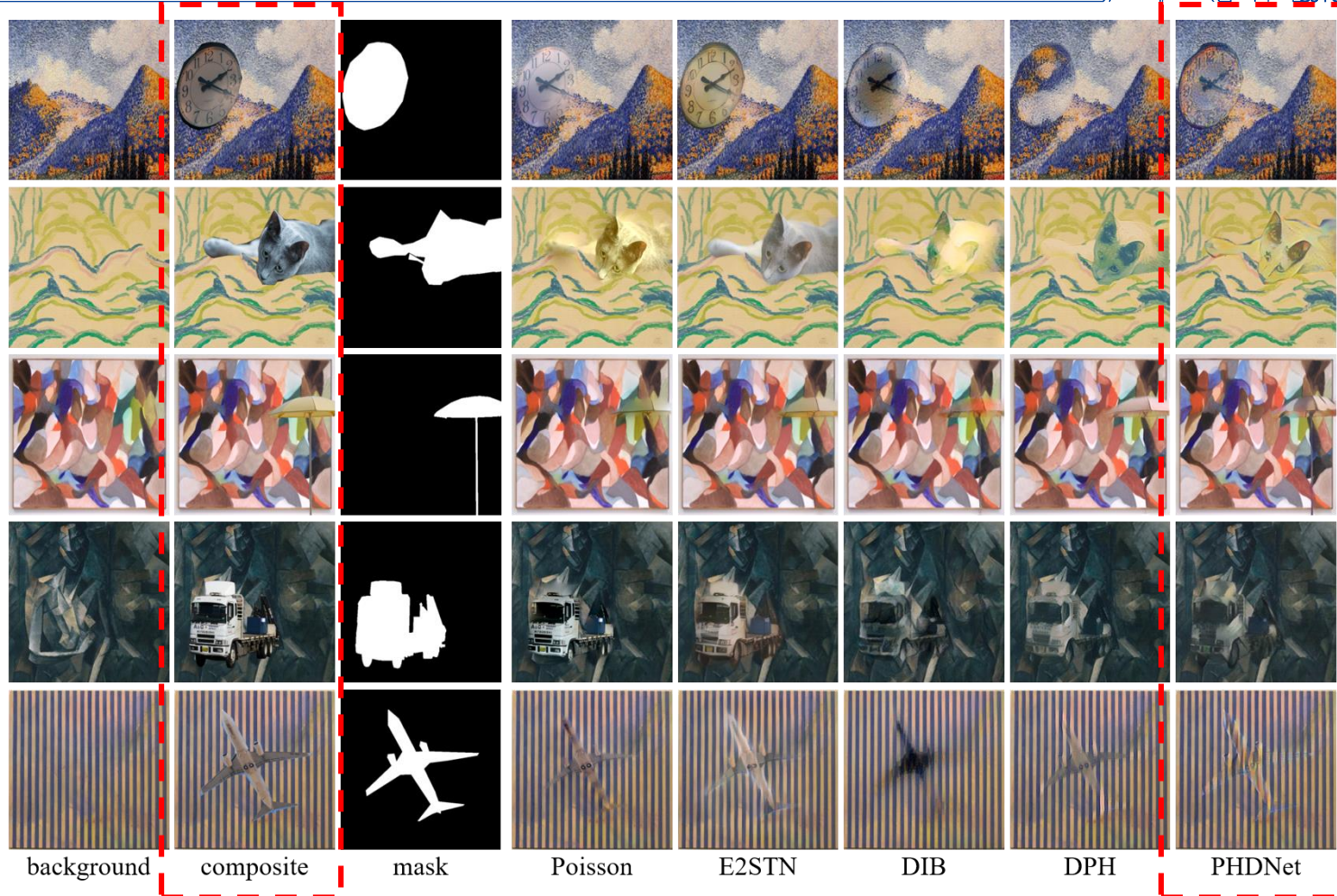
Spatial domain: AdaIN

Frequency domain: ResFFT



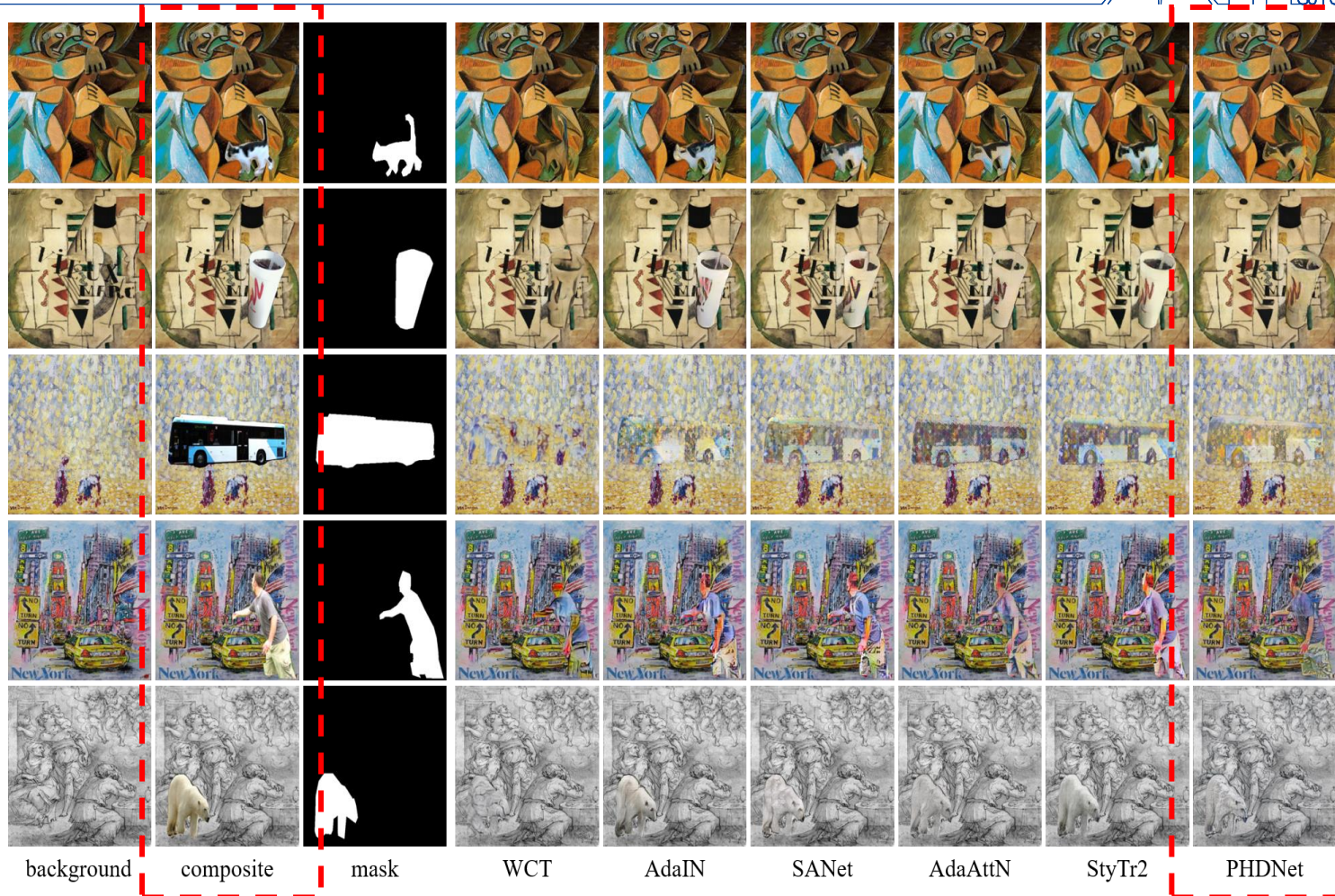


Painterly Image Harmonization





Painterly Image Harmonization

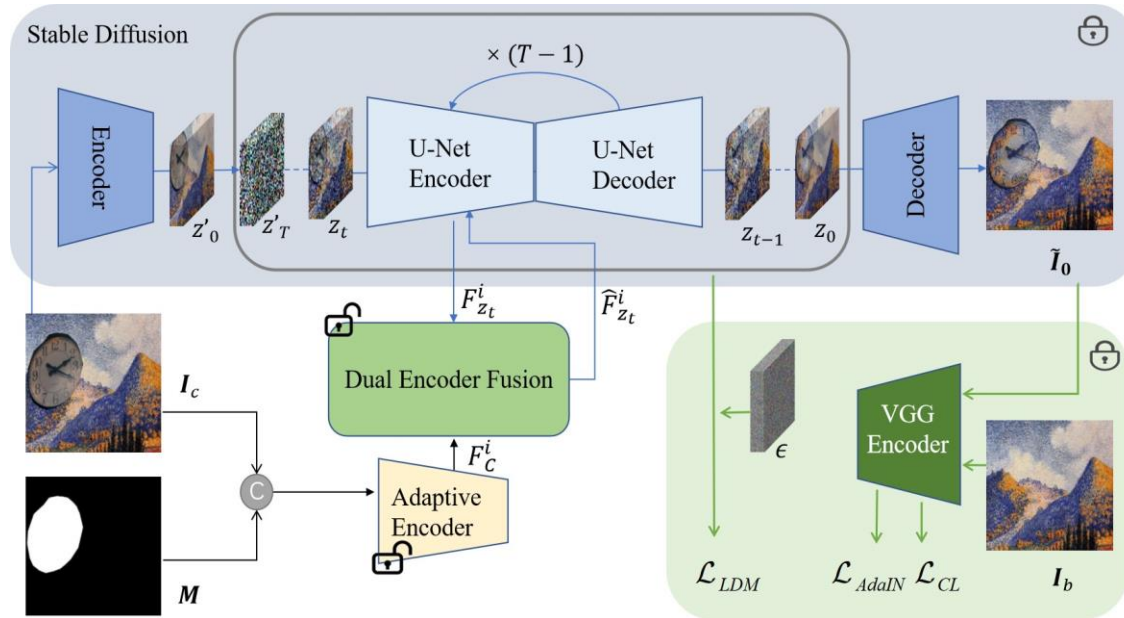


Painterly Image Harmonization



The first work using diffusion model for painterly image harmonization.

Outperform PHDNet significantly when the background has dense textures or abstract style.





Painterly Image Harmonization



Composite

Mask

InST

AdaIN

AdaAttn

SANet

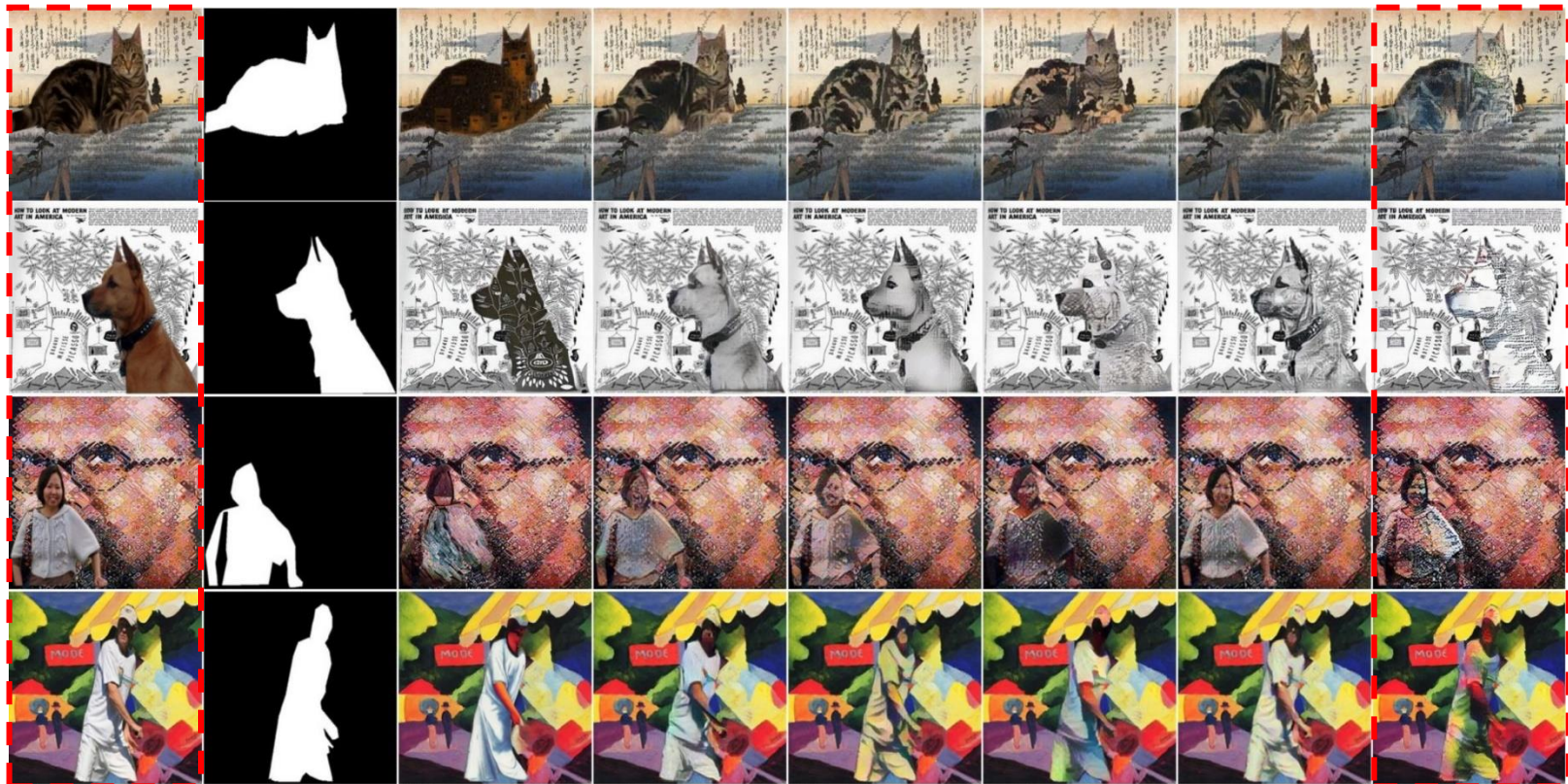
StyTr2

PHDiffusion

Lingxiao Lu, Jiangtong Li, Junyan Cao, **Li Niu**, Liqing Zhang, "Painterly Image Harmonization using Diffusion Model", **ACM MM, 2023.**



Painterly Image Harmonization



Composite

Mask

InST

AdaIN

AdaAttn

SANet

StyTr2

PHDiffusion

Lingxiao Lu, Jiangtong Li, Junyan Cao, **Li Niu**, Liqing Zhang, "Painterly Image Harmonization using Diffusion Model", **ACM MM, 2023.**



Painterly Image Harmonization



Composite

Mask

InST

AdaIN

AdaAttn

SANet

StyTr2

PHDiffusion

Lingxiao Lu, Jiangtong Li, Junyan Cao, **Li Niu**, Liqing Zhang, "Painterly Image Harmonization using Diffusion Model", **ACM MM, 2023.**



Painterly Image Harmonization



Composite

Mask

SDEdit

CDC

DIB

DPH

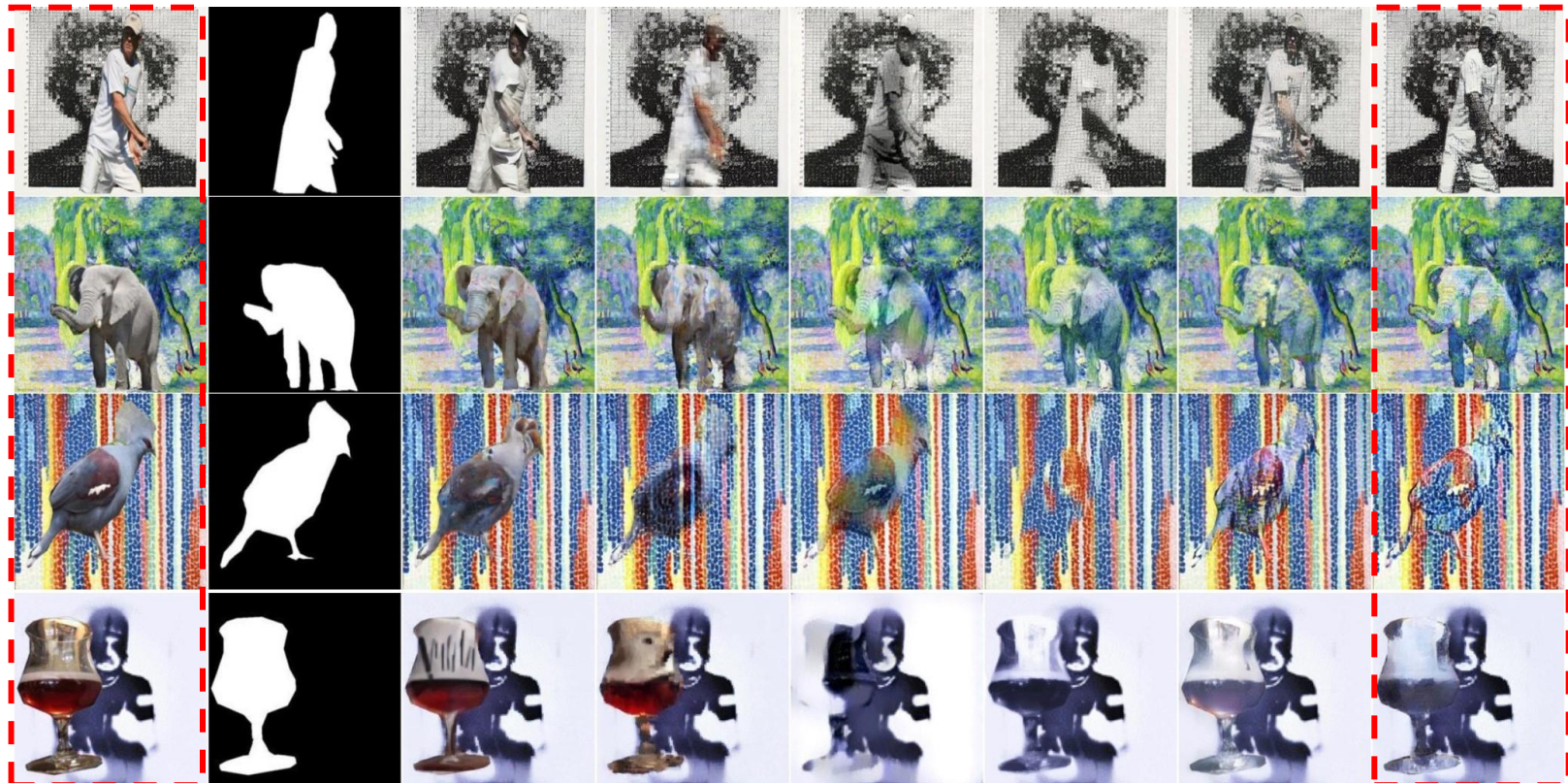
PHDNet

PHDiffusion

Lingxiao Lu, Jiangtong Li, Junyan Cao, **Li Niu**, Liqing Zhang, "Painterly Image Harmonization using Diffusion Model", **ACM MM, 2023.**



Painterly Image Harmonization



Composite

Mask

SDEdit

CDC

DIB

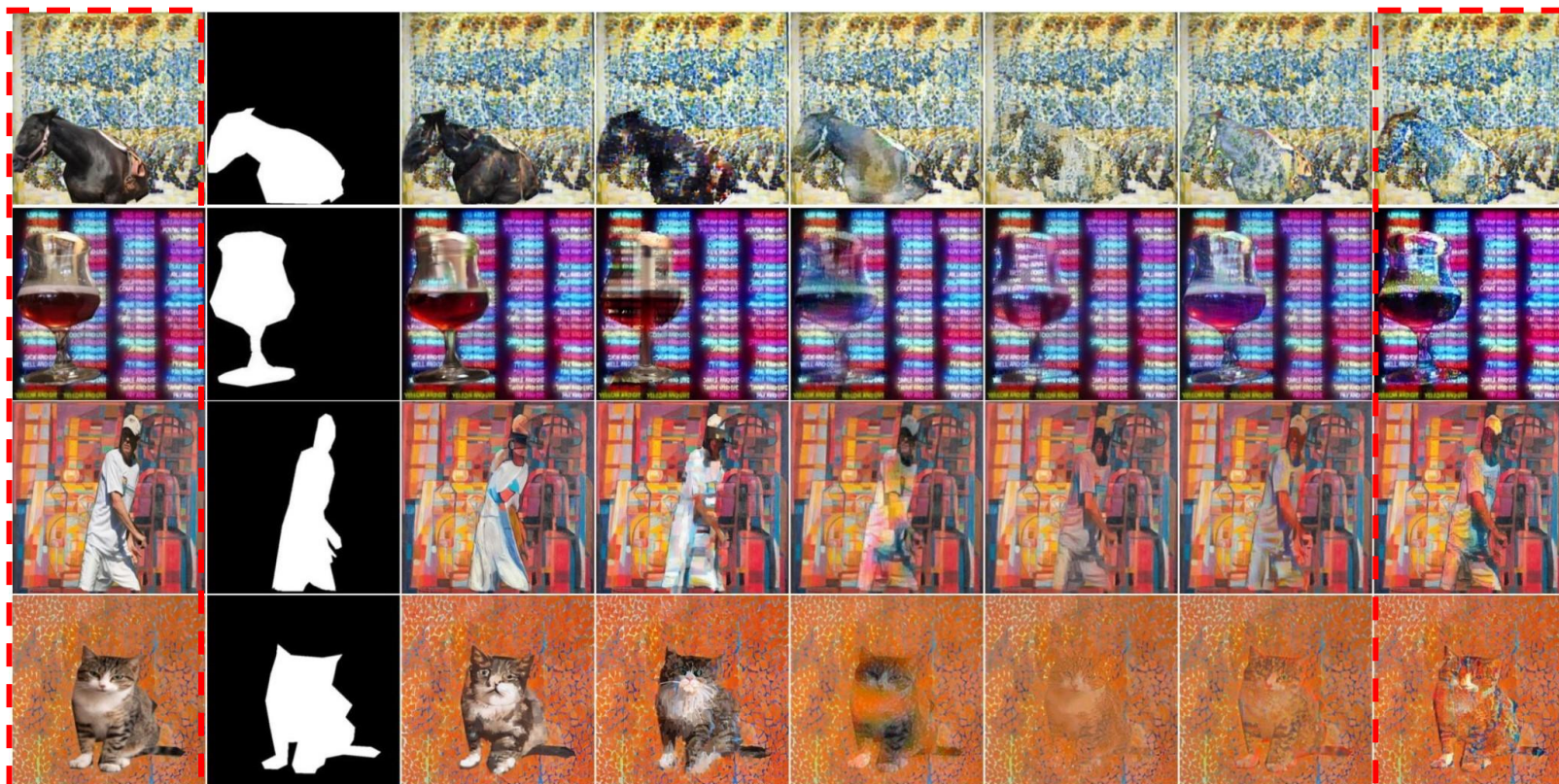
DPH

PHDNet

PHDiffusion

Lingxiao Lu, Jiangtong Li, Junyan Cao, **Li Niu**, Liqing Zhang, "Painterly Image Harmonization using Diffusion Model", **ACM MM, 2023.**

Painterly Image Harmonization



Composite

Mask

SDEdit

CDC

DIB

DPH

PHDNet

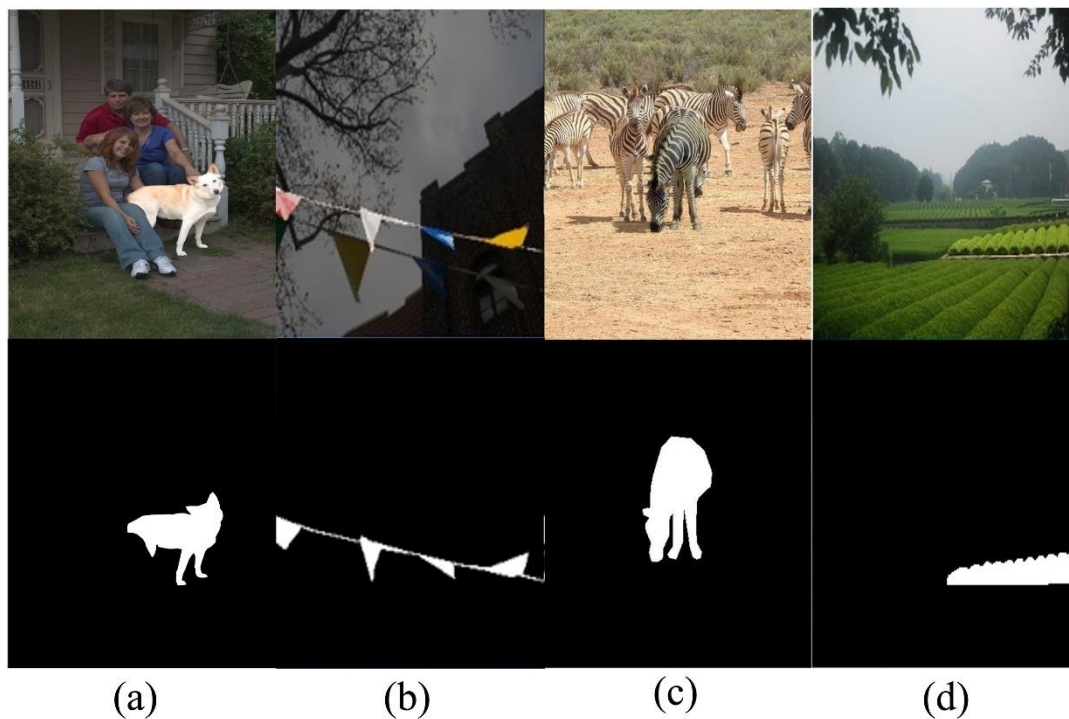
PHDiffusion

Lingxiao Lu, Jiangtong Li, Junyan Cao, **Li Niu**, Liqing Zhang, "Painterly Image Harmonization using Diffusion Model", **ACM MM, 2023.**

Inharmonious Region Localization

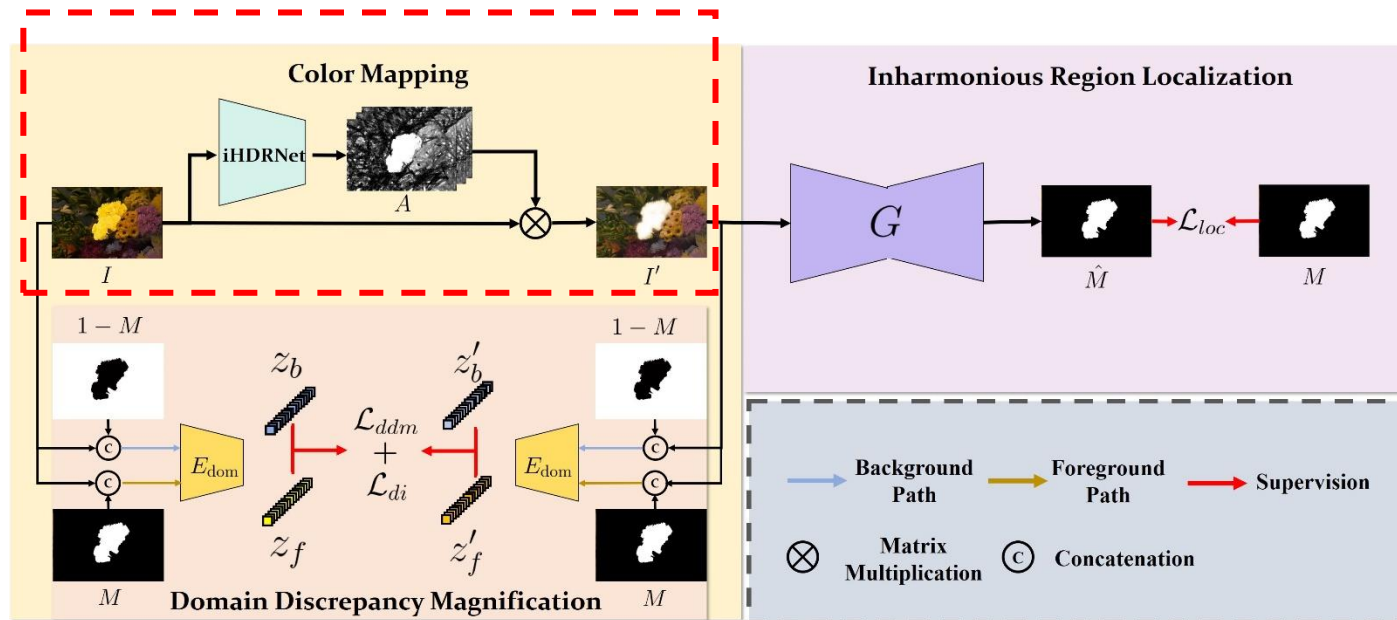


Inharmonious region localization aims to localize the inharmonious region in an image, which is an adversarial task of image harmonization.



Inharmonious Region Localization

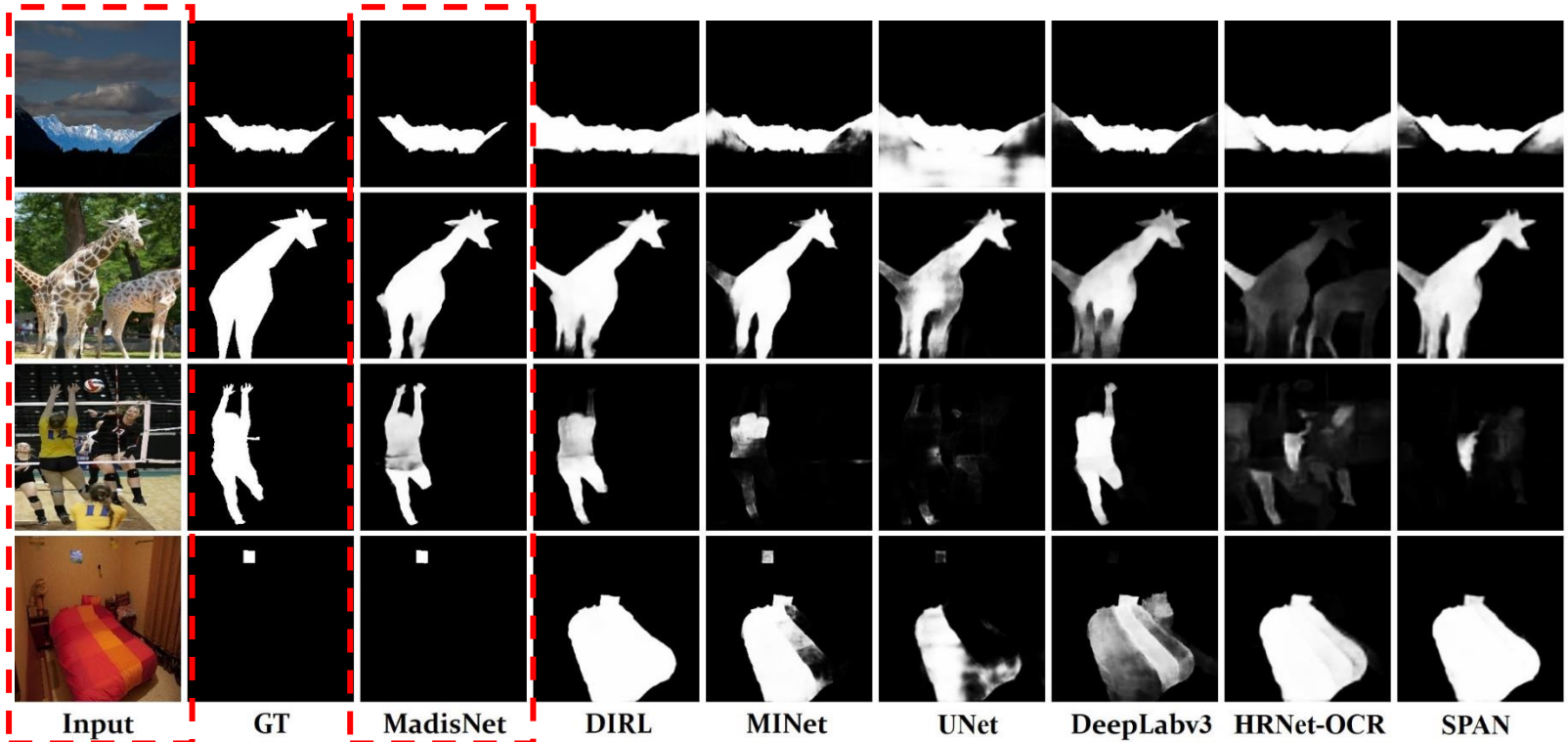
Learn a color mapping to magnify the domain discrepancy between foreground and background.



Jing Liang, **Li Niu**, Penghao Wu, Fengjun Guo, Teng Long, "Inharmonious Region Localization by Magnifying Domain Discrepancy", **AAAI**, 2022.



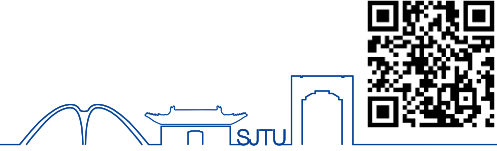
Inharmonious Region Localization



Jing Liang, **Li Niu**, Penghao Wu, Fengjun Guo, Teng Long, "Inharmonious Region Localization by Magnifying Domain Discrepancy", **AAAI**, 2022.



Image Composition



- **Overall**
- **Image harmonization**
- **Shadow generation**
- Object placement
- Foreground object search
- Generative composition

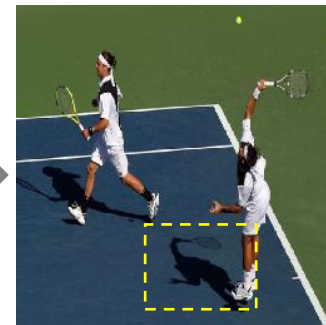
Shadow Generation



The inserted object may cast a shadow on the background, which is ignored by image harmonization.



Shadow
Generation



Unrealistic composite image

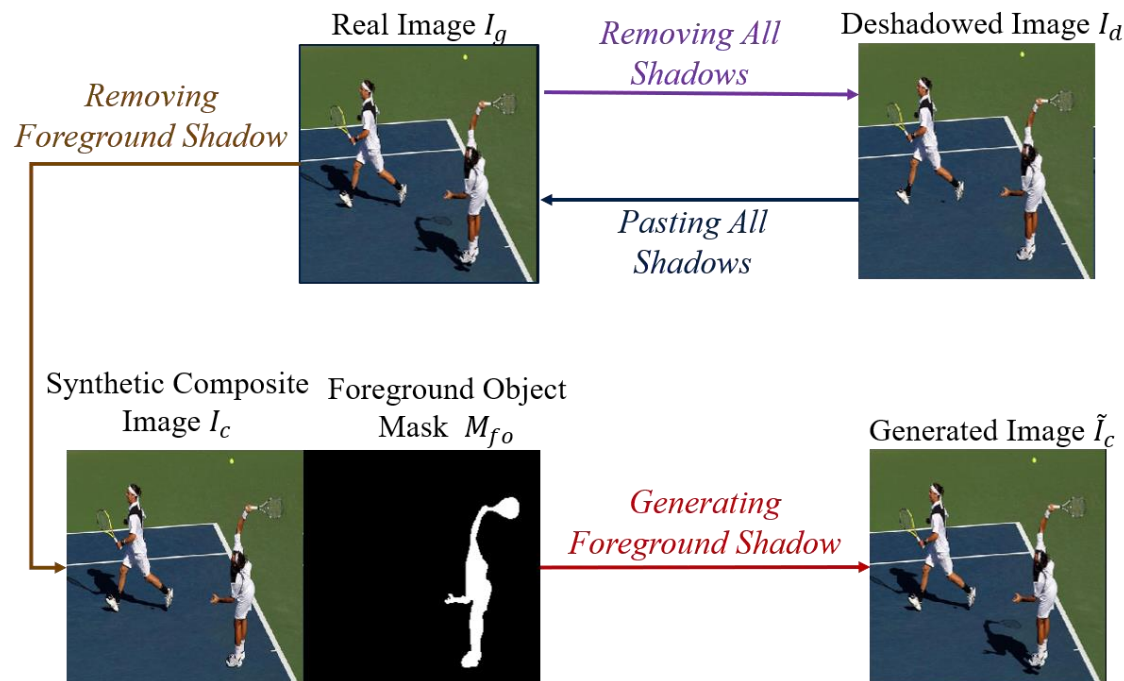
Realistic result with shadow



Shadow Generation



We release **the first real-world shadow generation dataset** DESOBA.

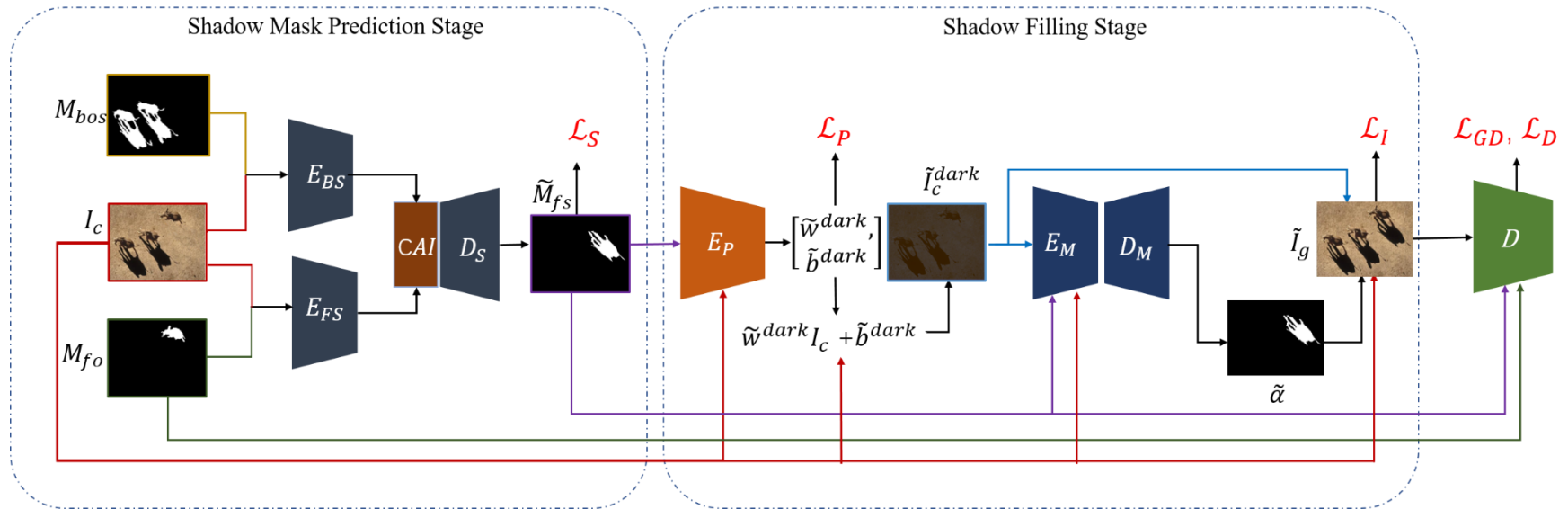


Yan Hong, **Li Niu**, Jianfu Zhang, "Shadow Generation for Composite Image in Real-world Scenes", **AAAI**, 2022.

Shadow Generation



Two-stage network: first predict shadow mask and then fill in the shadow region



Yan Hong, **Li Niu**, Jianfu Zhang, "Shadow Generation for Composite Image in Real-world Scenes", **AAAI**, 2022.



Shadow Generation

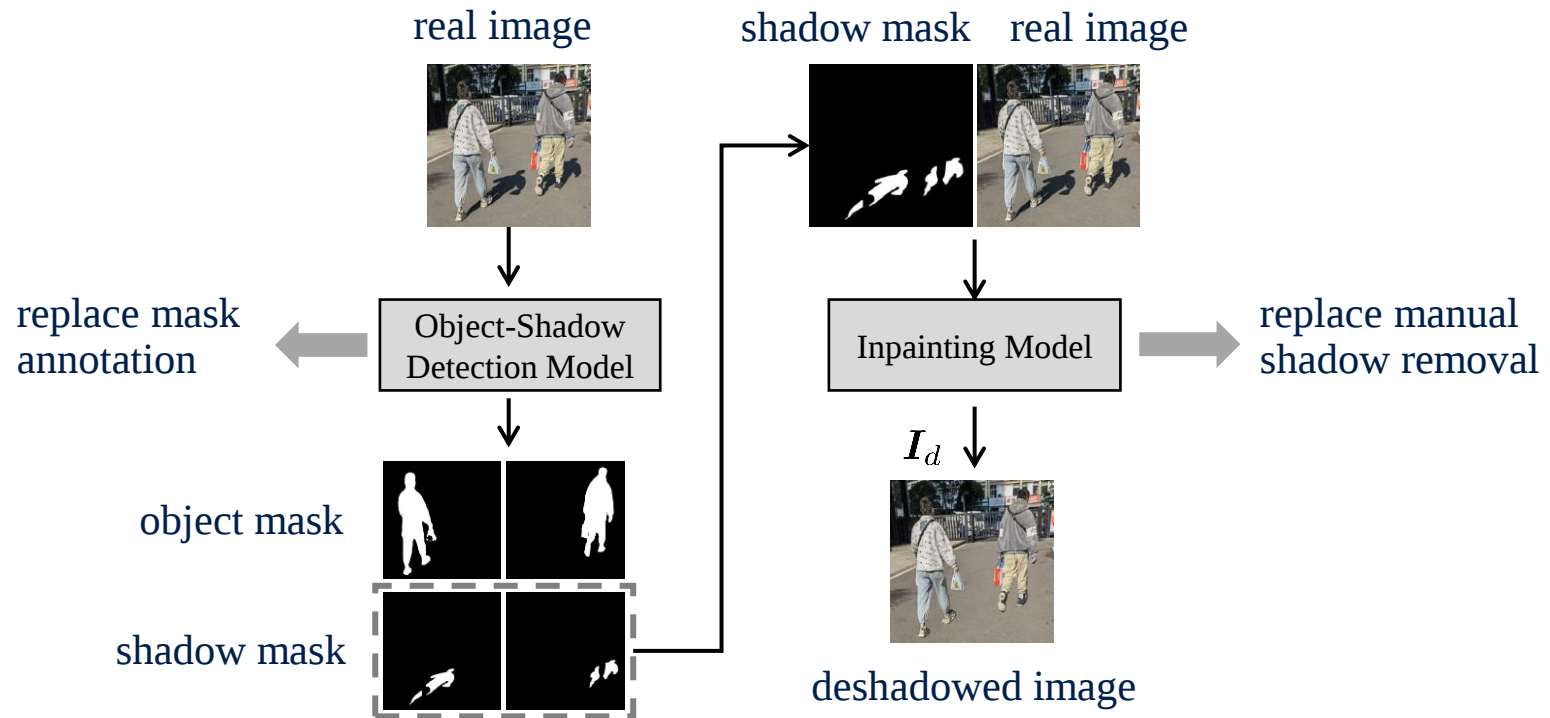


Yan Hong, **Li Niu**, Jianfu Zhang, "Shadow Generation for Composite Image in Real-world Scenes", **AAAI**, 2022.

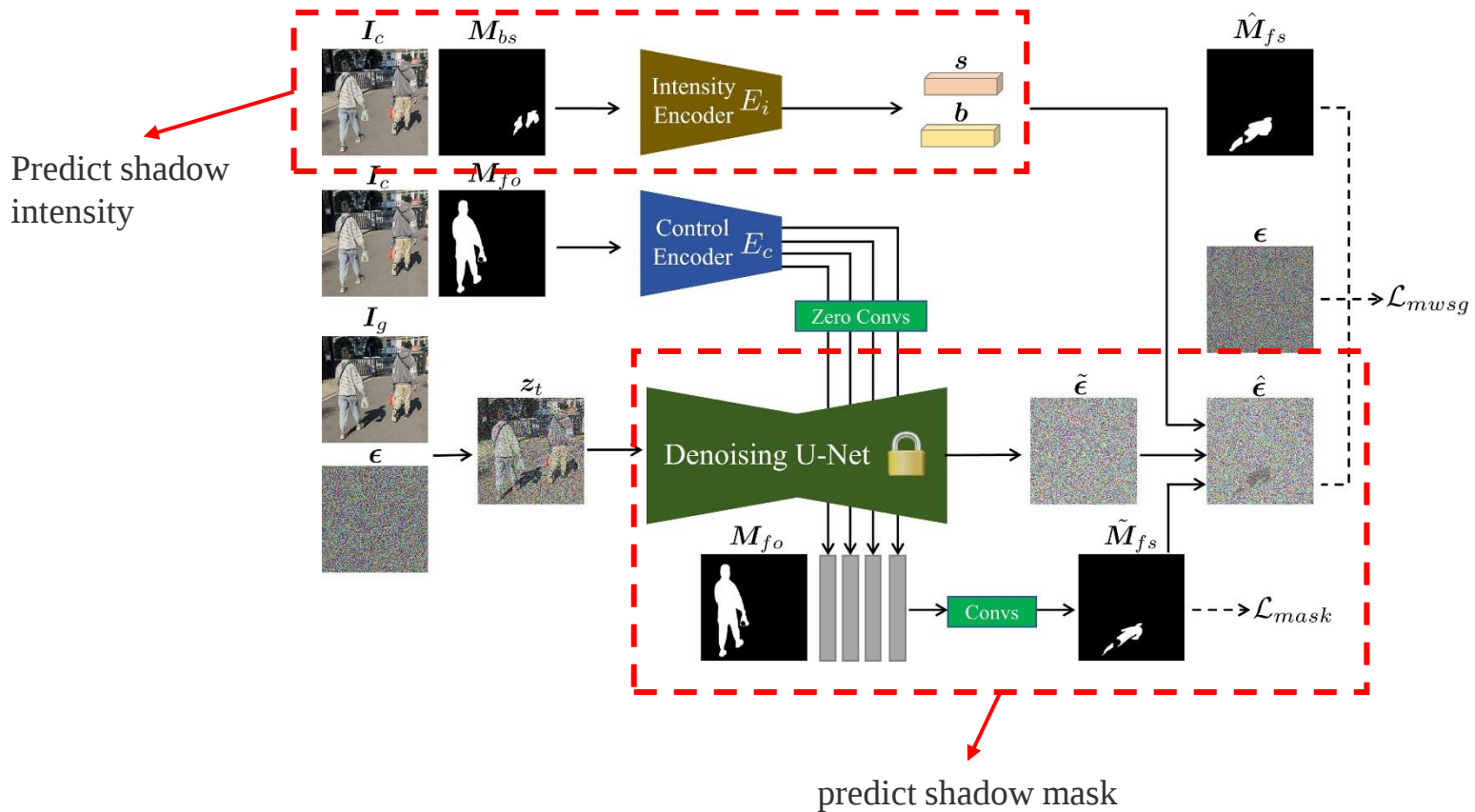
Shadow Generation



We extend DESOBA dataset (1k images) to DESOBAv2 (20k images).

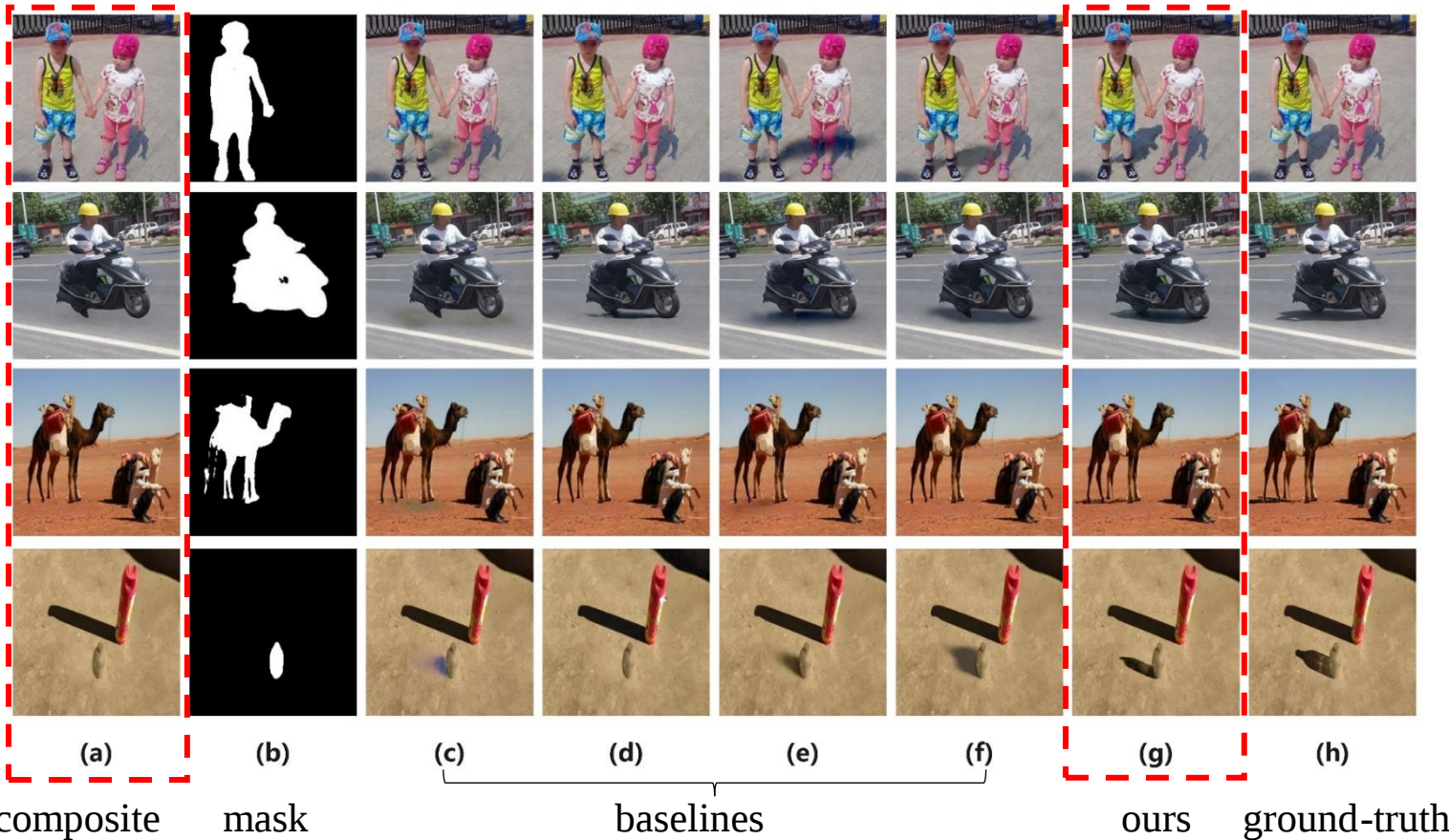


Shadow Generation





Shadow Generation





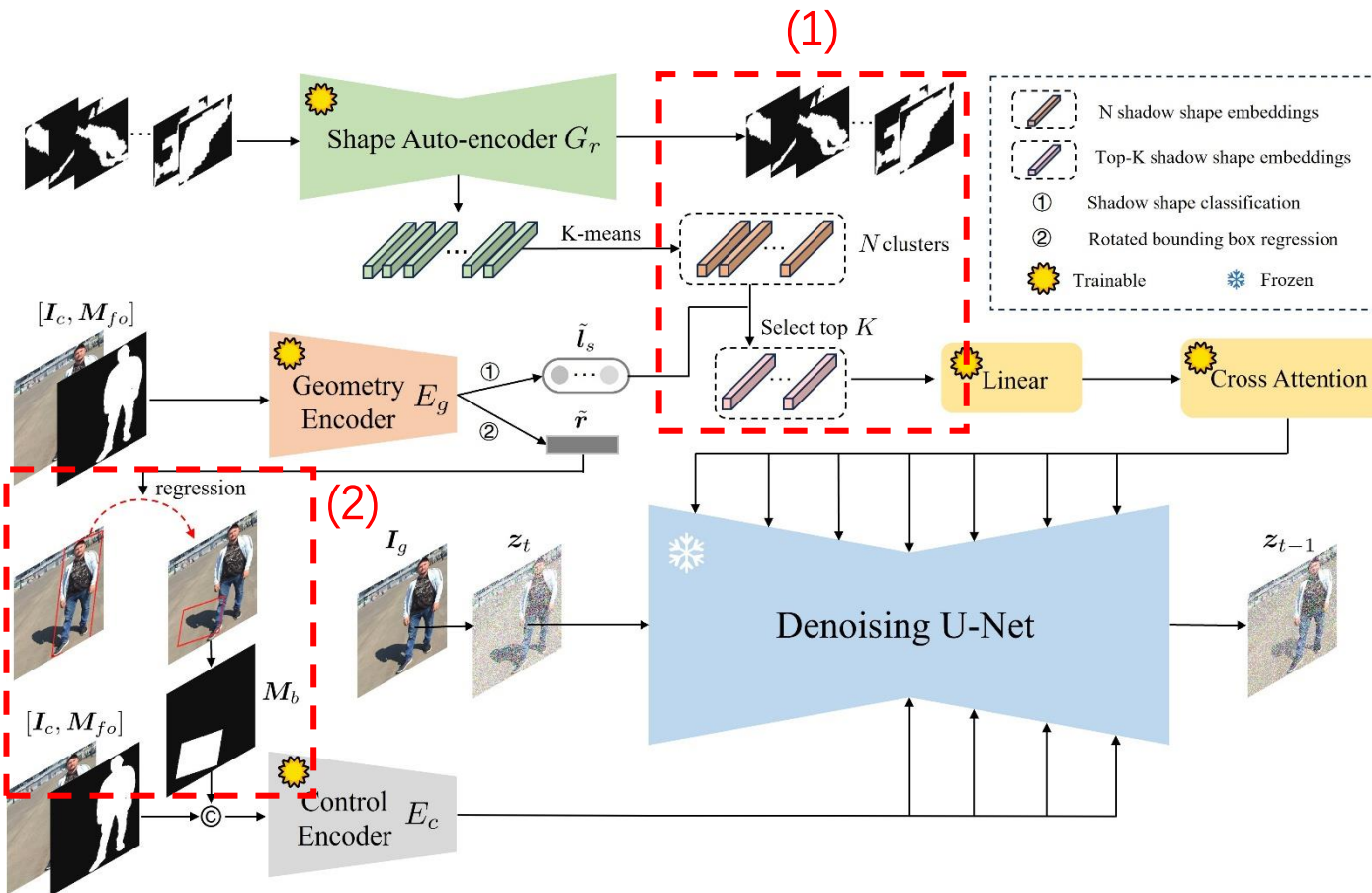
Shadow Generation



Existing shadow generation methods are struggling to generate shadows with plausible geometry in complex cases.

Haonan Zhao, Qingyang Liu, Xinhao Tao, **Li Niu**, Guangtao Zhai, "Shadow Generation Using Diffusion Model with Geometry Prior", **CVPR, 2025**.

Shadow Generation

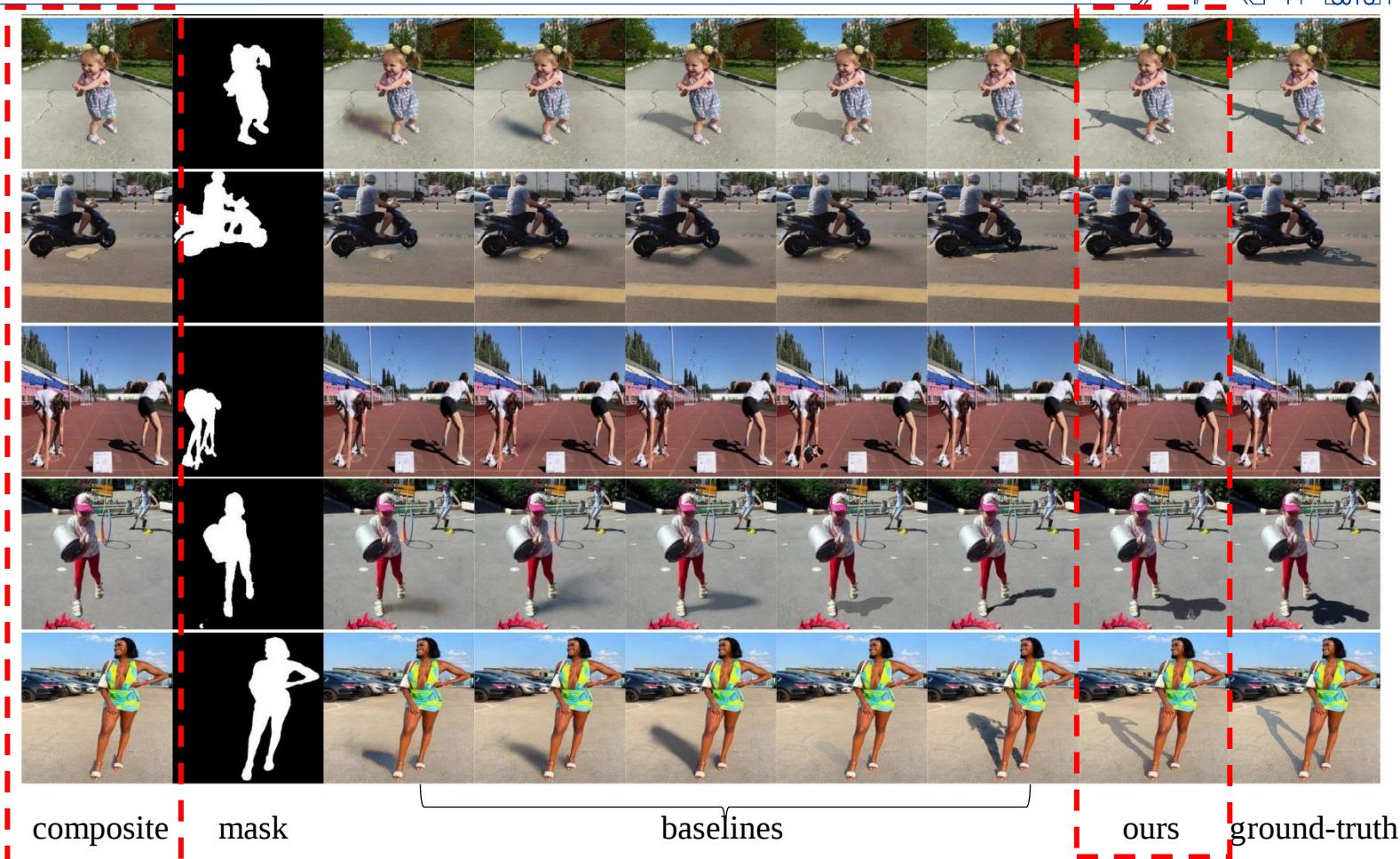


(1) Retrieve compatible shadow shapes

(2) Predict rotated shadow bounding box



Shadow Generation



composite

mask

baselines

ours

ground-truth

Haonan Zhao, Qingyang Liu, Xinhao Tao, **Li Niu**, Guangtao Zhai, "Shadow Generation Using Diffusion Model with Geometry Prior", **CVPR**, 2025.



Image Composition



- **Overall**
- **Image harmonization**
- **Shadow generation**
- **Object placement**
- Foreground object search
- Generative composition

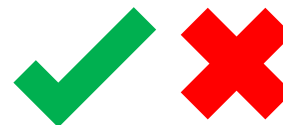


Object Placement



Three levels of tasks

Verify reasonability



Generate one
reasonable solution



Generate all
reasonable solutions



...





Object Placement



1. Foreground object too large or too small
2. Hanging in the air
3. Appearing in the wrong place
4. Unreasonable occlusion
5. Unreasonable perspective

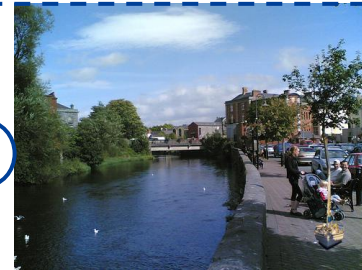
1



2



3



4



5

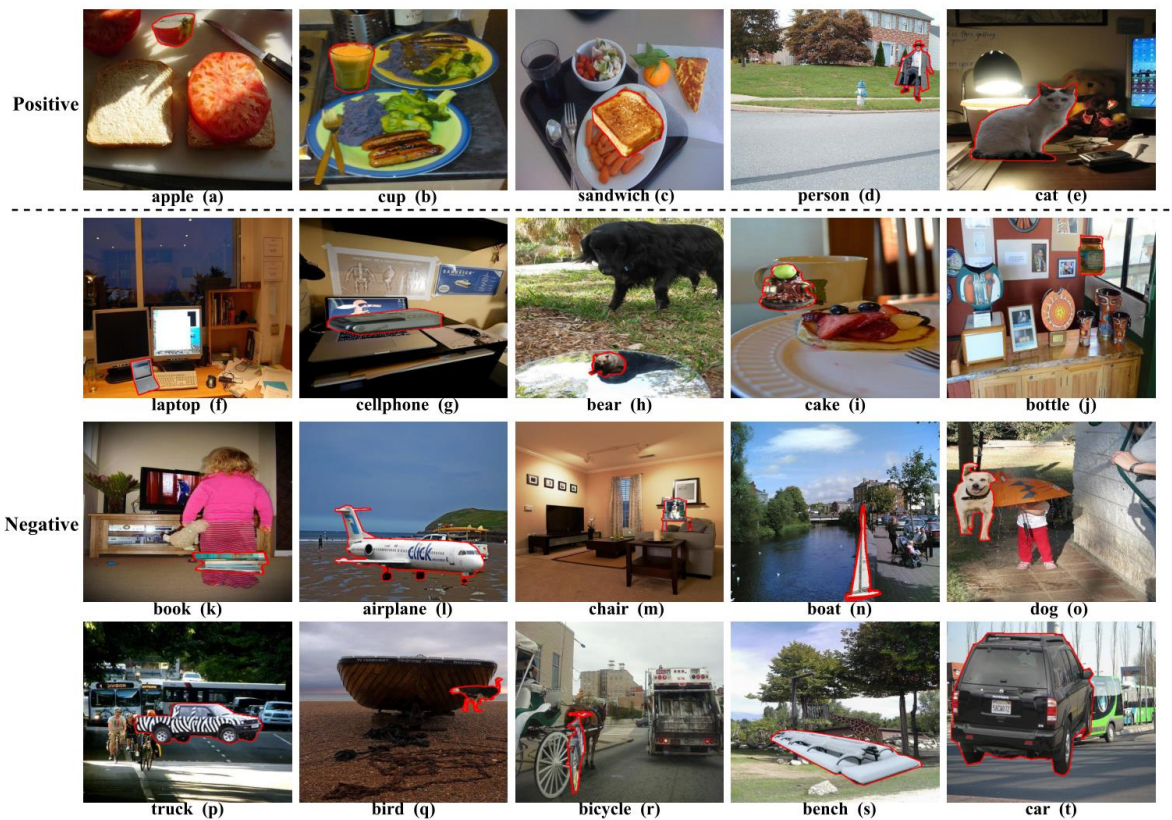




Object Placement



We release **the first object placement dataset OPA**.

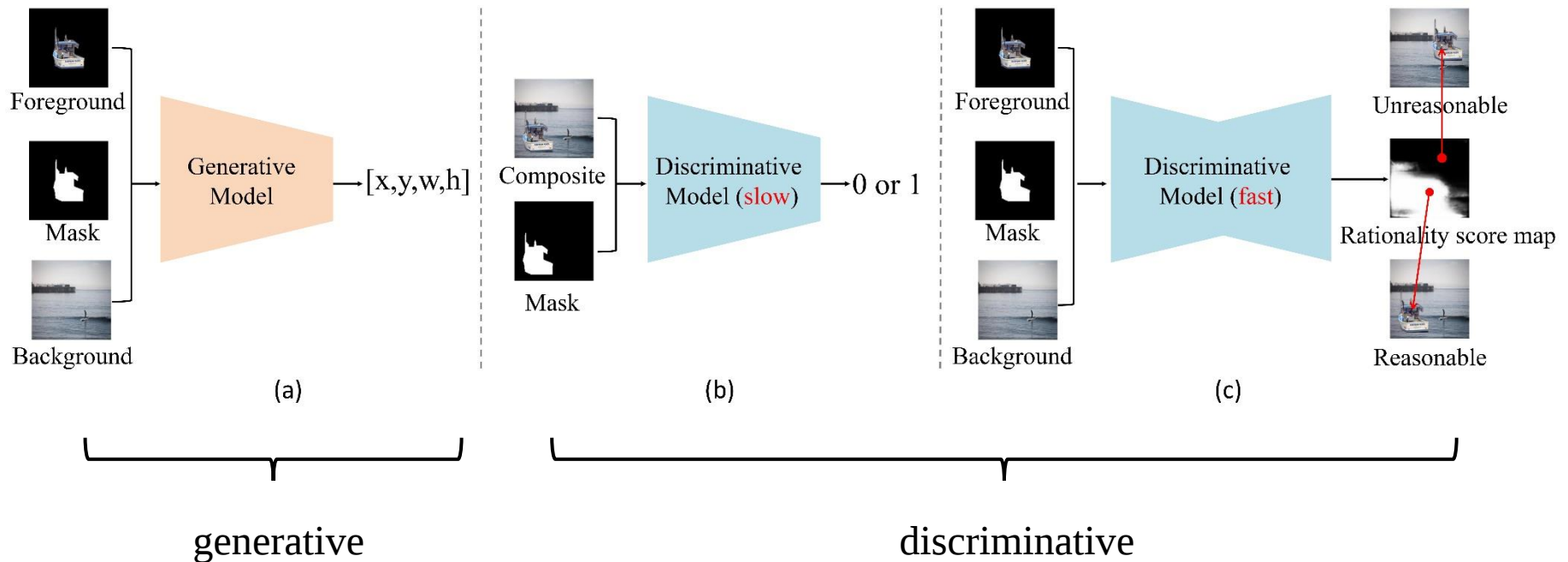


Generative v.s. Discriminative



Slow Object Placement Assessment (SOPA)

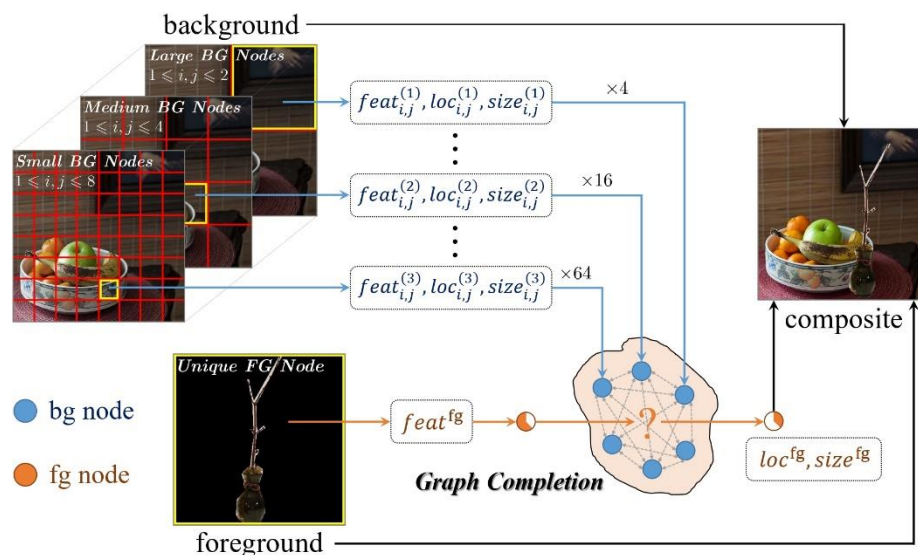
Fast Object Placement Assessment (FOPA)



Generative Method



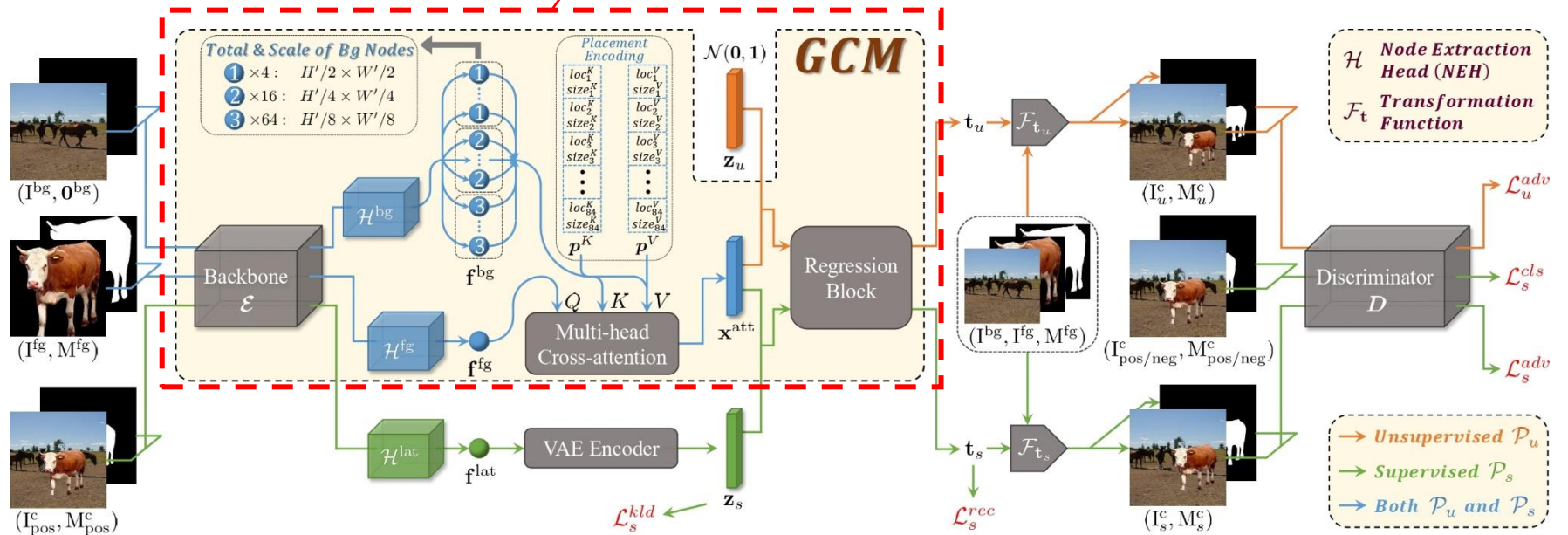
Formulate object placement as a **graph completion task**: background nodes have both content and placement information, while foreground node only has content information.



Generative Method



Graph Completion Module (GCM)





Generative Method



TERSE



PlaceNet



GracoNet

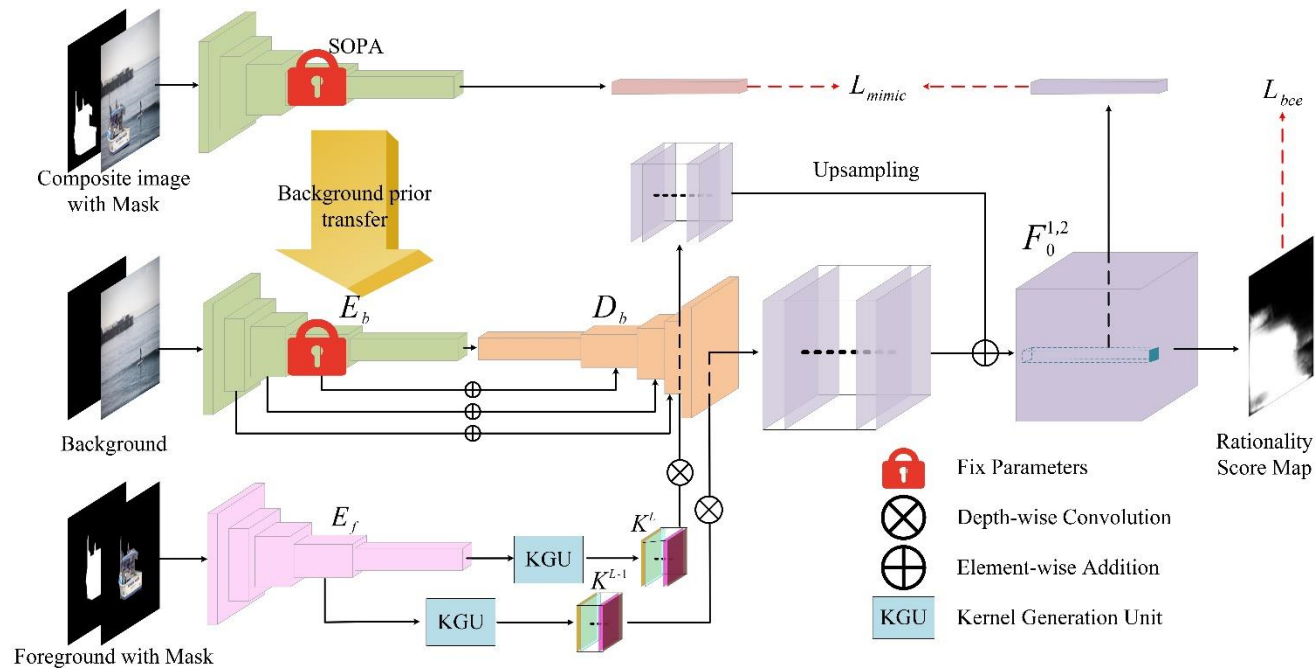


Siyuan Zhou, Liu Liu, **Li Niu**, Liqing Zhang, "Learning Object Placement via Dual-path Graph Completion", **ECCV, 2022**.

Discriminative Method

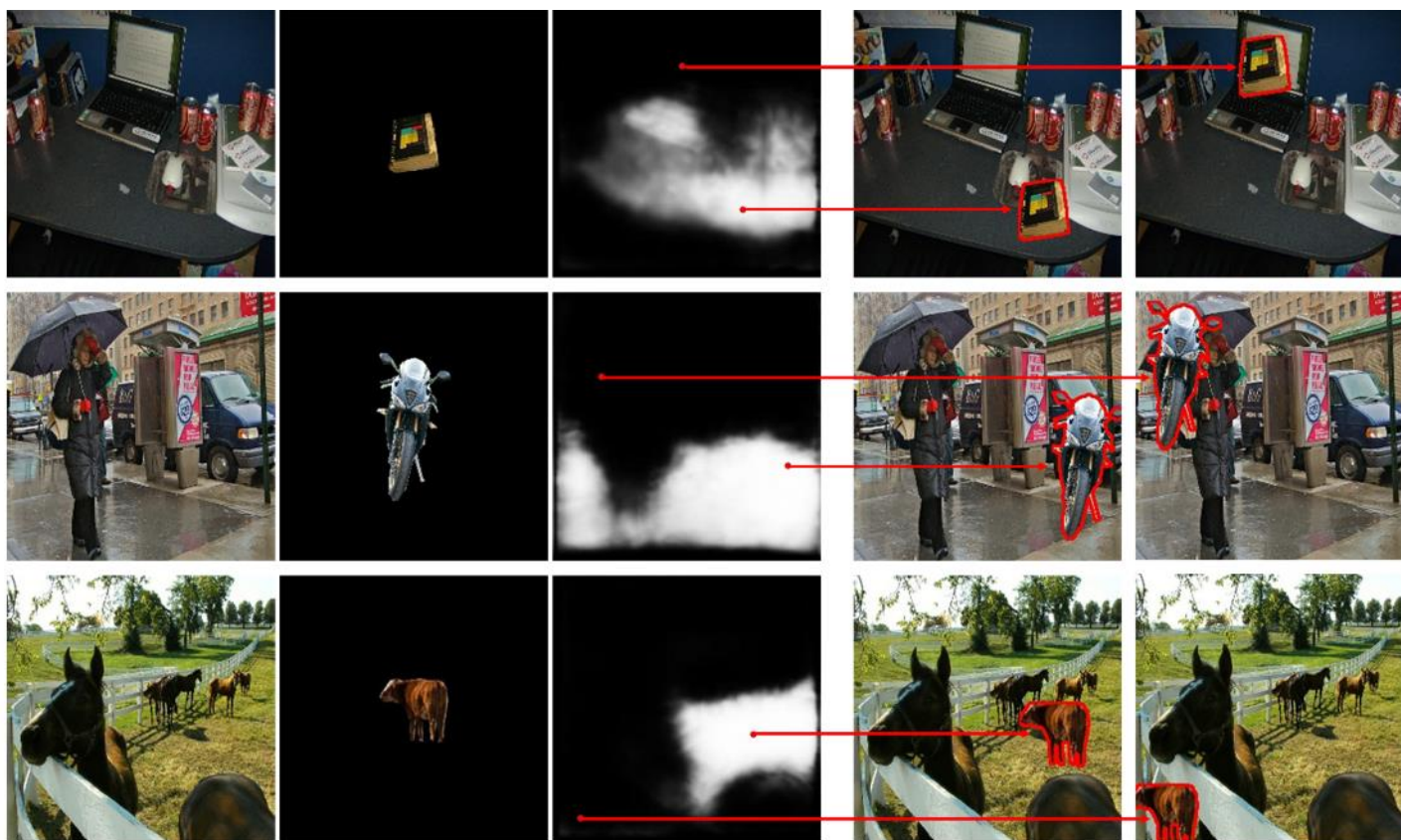


The first discriminative approach for object placement.



Li Niu, Qingyang Liu, Zhenchen Liu, Jiangtong Li, "Fast Object Placement Assessment." arXiv preprint arXiv:2205.14280 (2022).

Discriminative Method



background

foreground

FOPA

positive composite

negative composite

Li Niu, Qingyang Liu, Zhenchen Liu, Jiangtong Li, "Fast Object Placement Assessment." arXiv preprint arXiv:2205.14280 (2022).

Image Composition



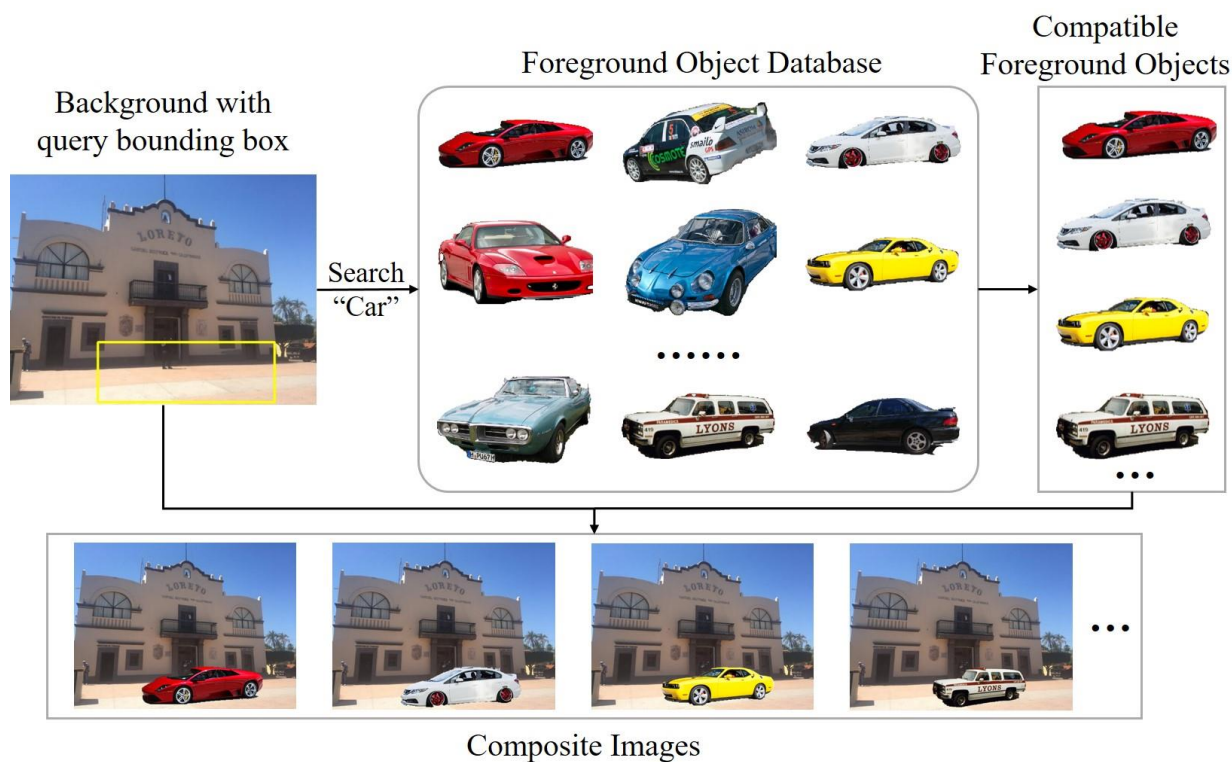
- **Overall**
- **Image harmonization**
- **Shadow generation**
- **Object placement**
- **Foreground object search**
- **Generative composition**

Foreground Object Search



Given a background and a bounding box, FOS aims to retrieve compatible (geometry, semantics) foreground objects from the foreground object database.

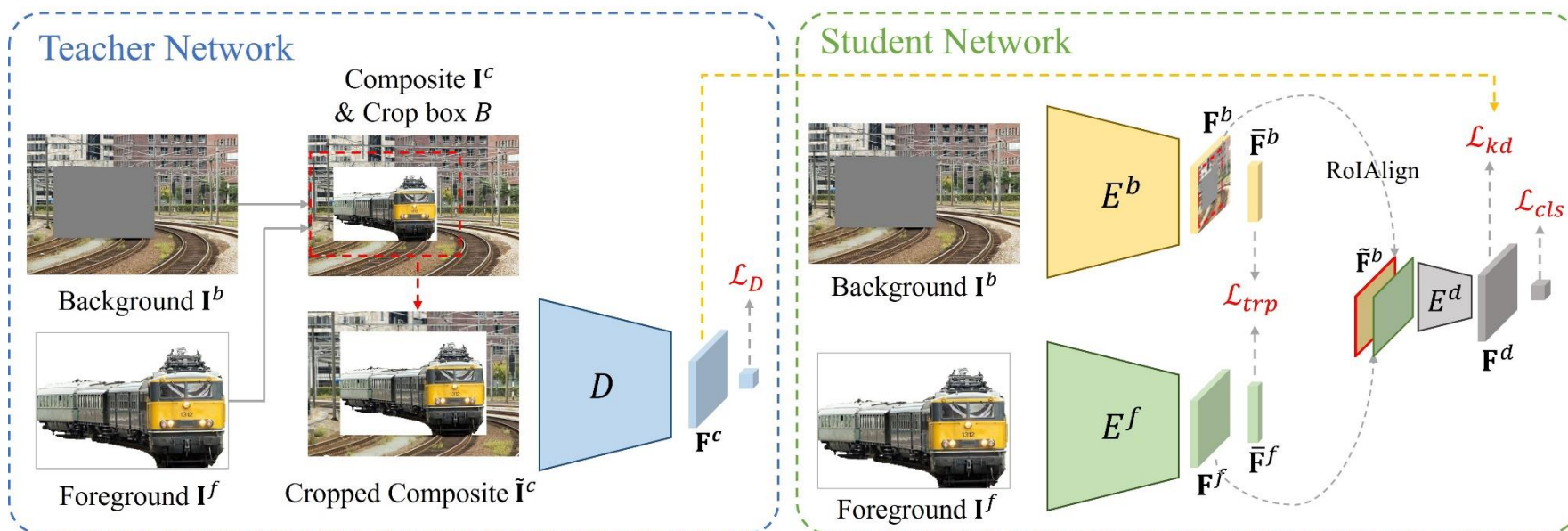
We release **the first two public FOS datasets**: S-FOSD and R-FOSD.



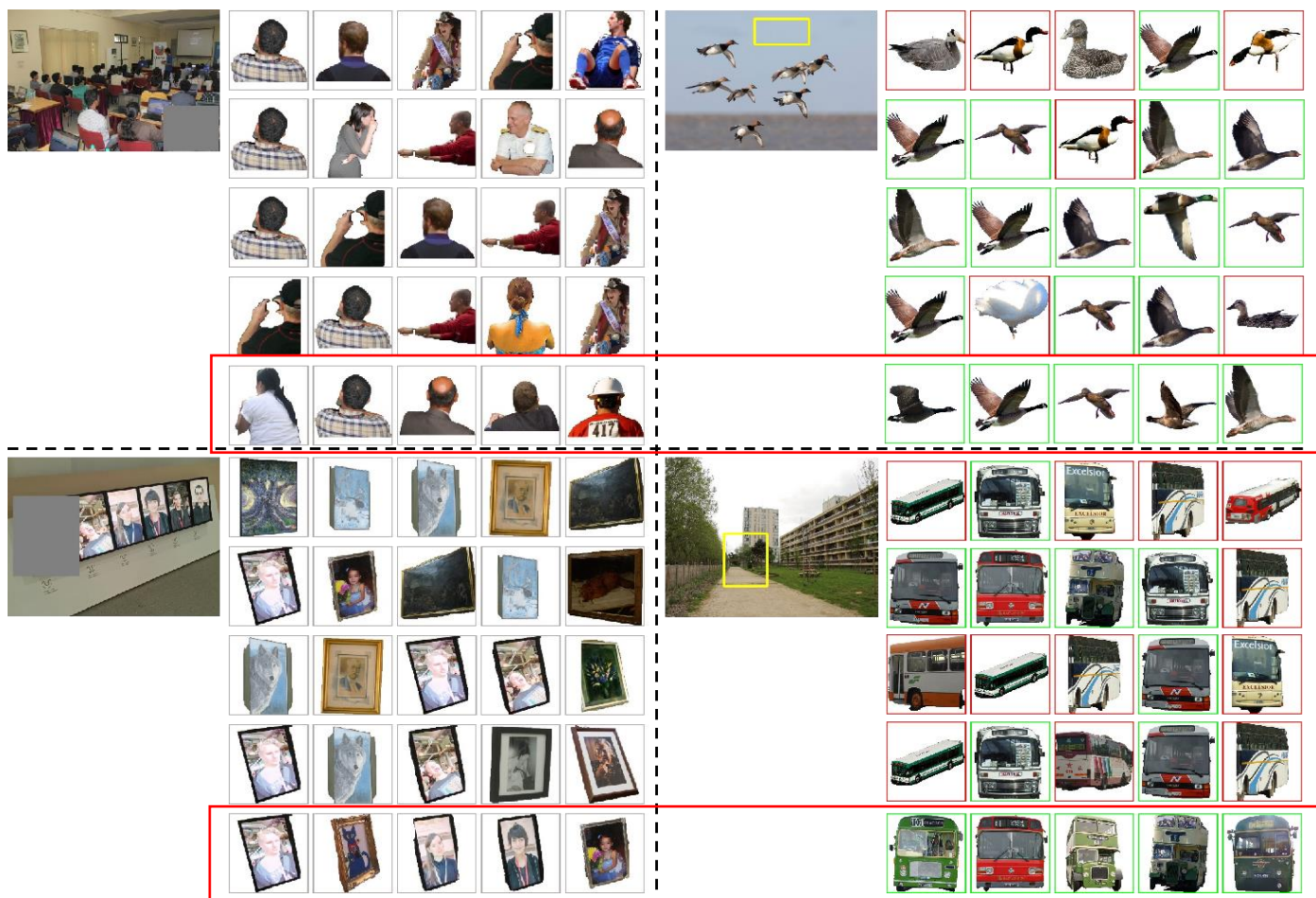
Foreground Object Search



Distill knowledge from composite image features in the teacher network.



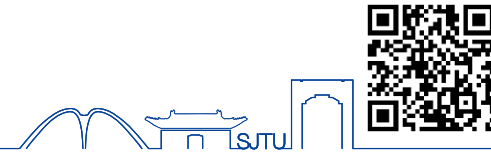
Foreground Object Search



our results

our results

Image Composition



- **Overall**
- **Image harmonization**
- **Shadow generation**
- **Object placement**
- **Foreground object search**
- **Generative composition**



Generative Composition



background with
bounding box



foreground



composite image

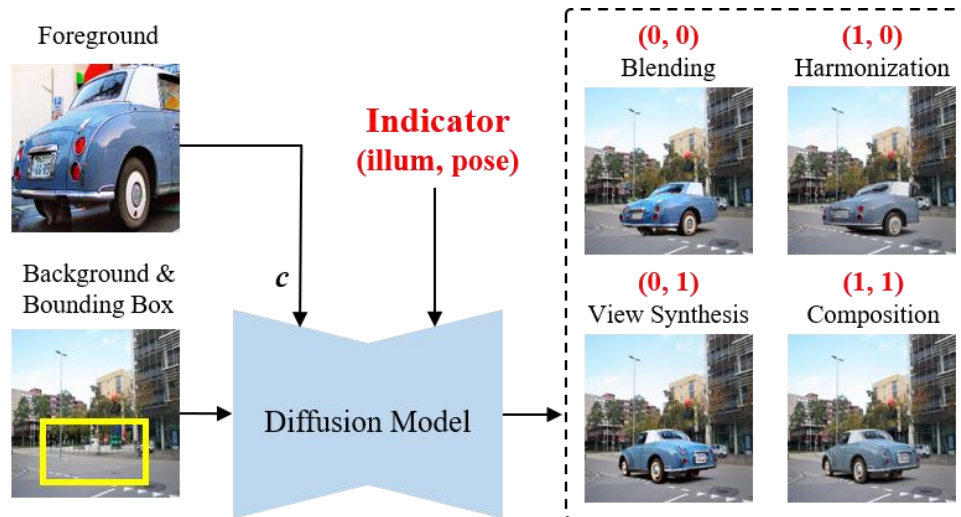


Generative Composition

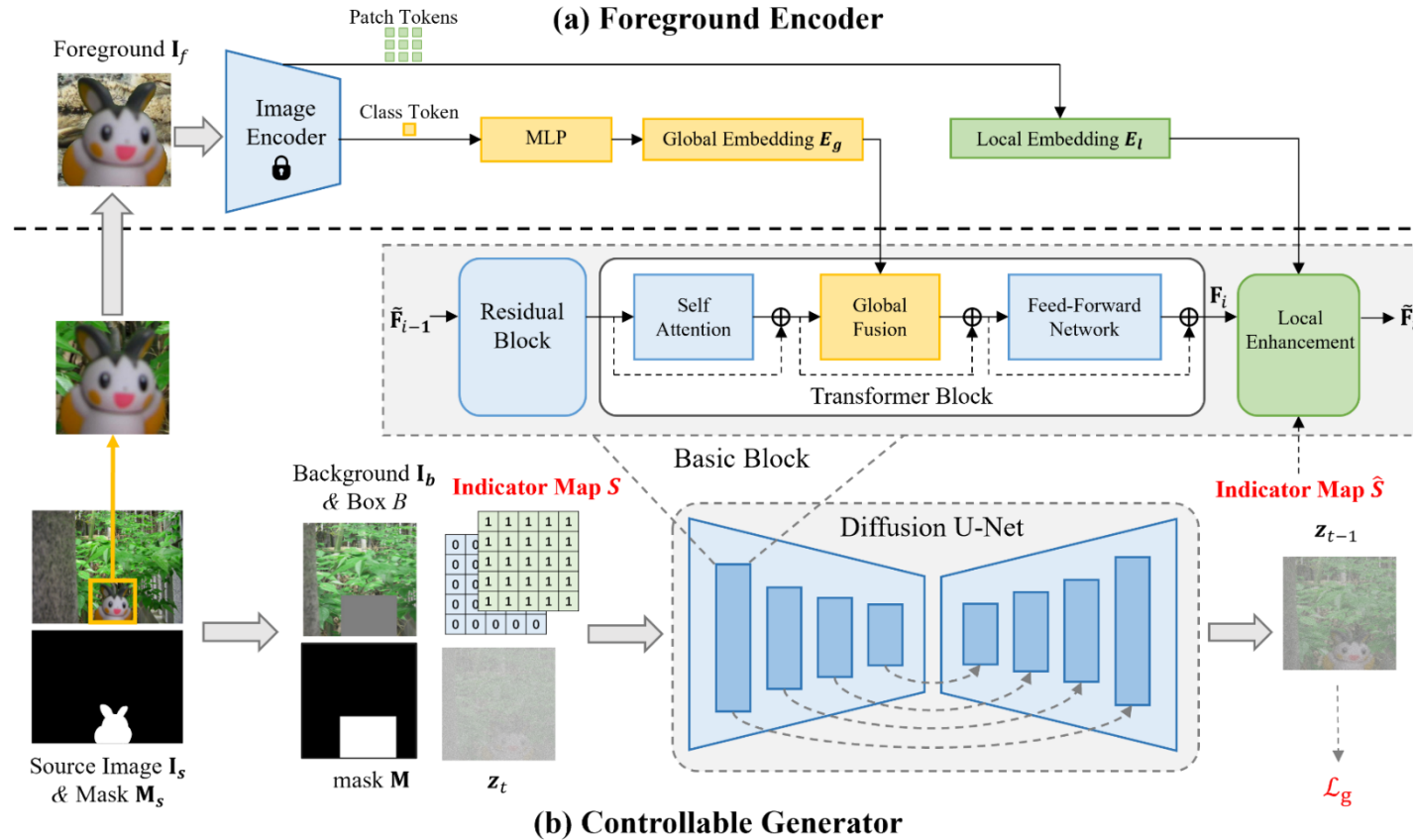


Controllable composition: selectively adjust illumination (A) and pose (B) of foreground to be compatible with background.

- A0B0: adjust neither illumination nor pose (image blending)
- A1B0: only adjust illumination (image harmonization)
- A0B1: only adjust pose (view synthesis)
- A1B1: adjust both illumination and pose (image composition)

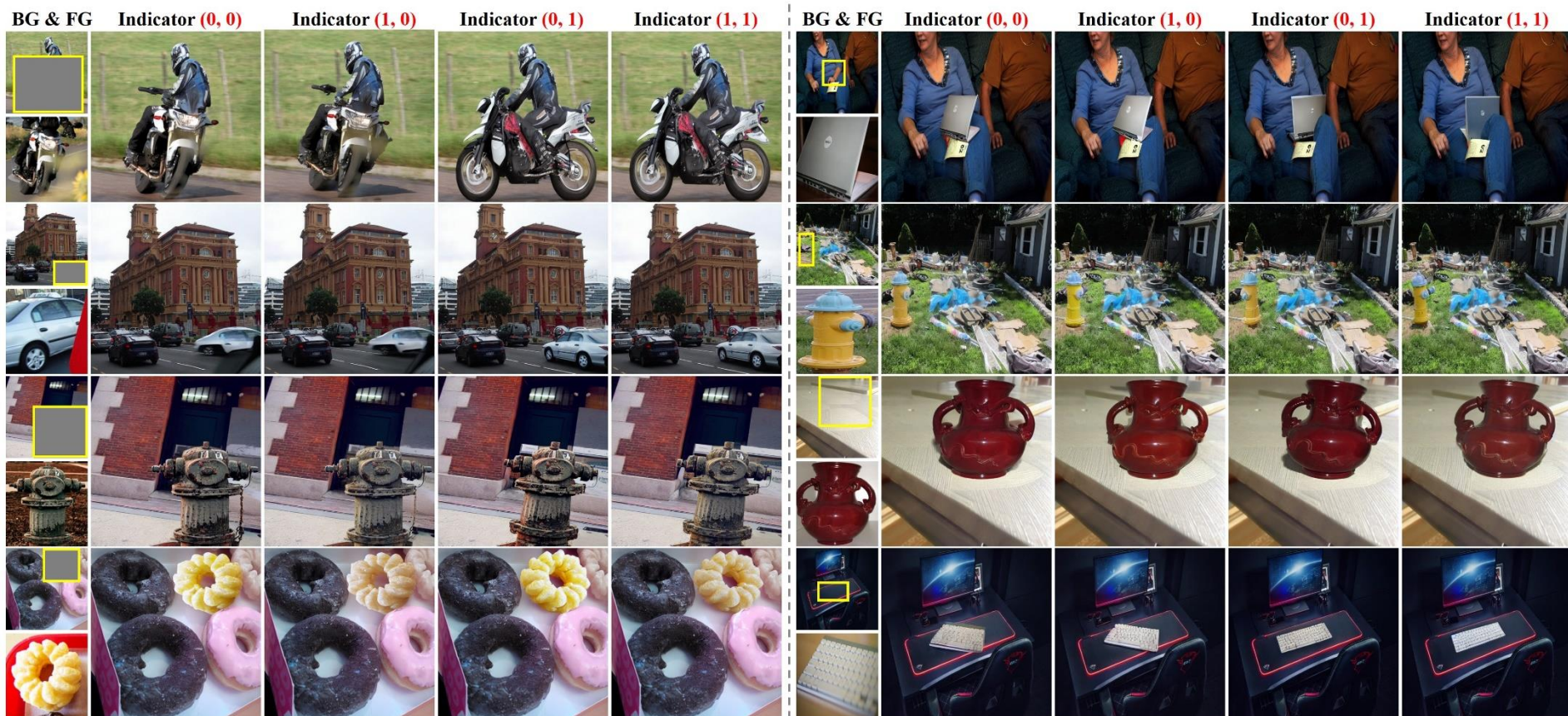


Generative Composition



Bo Zhang, Yuxuan Duan, Jun Lan, Yan Hong, Huijia Zhu, Weiqiang Wang, **Li Niu**, "Controlcom: Controllable image composition using diffusion model." arXiv preprint arXiv:2308.10040 (2023).

Generative Composition





Ours



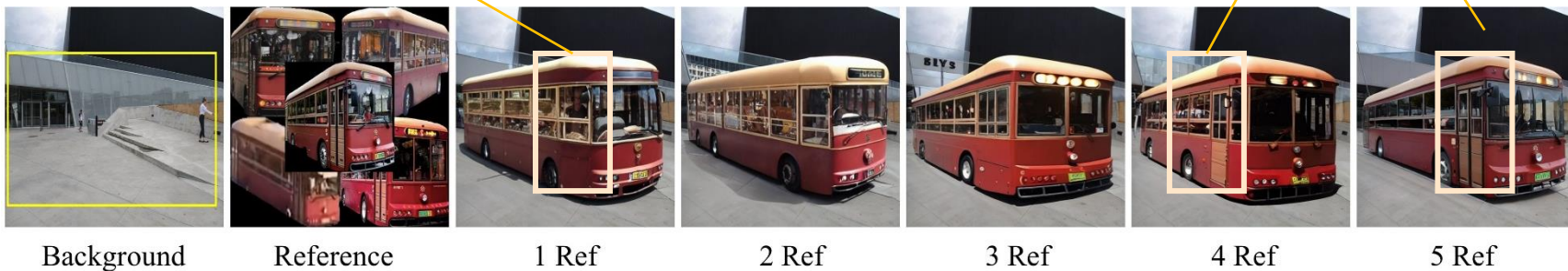
Generative Composition



In some cases, it is necessary to use multiple reference images, because they can provide supplementary information.

When using one reference image, there is **no door** on the right side.

When using more reference images, there is **door** on the right side.



Generative Composition

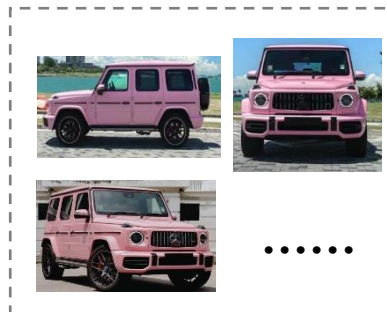


- If there are a large number of reference images, we can use **Foreground Object Search (FOS)** to retrieve the suitable reference image.

background with bounding box



foreground reference images



FOS

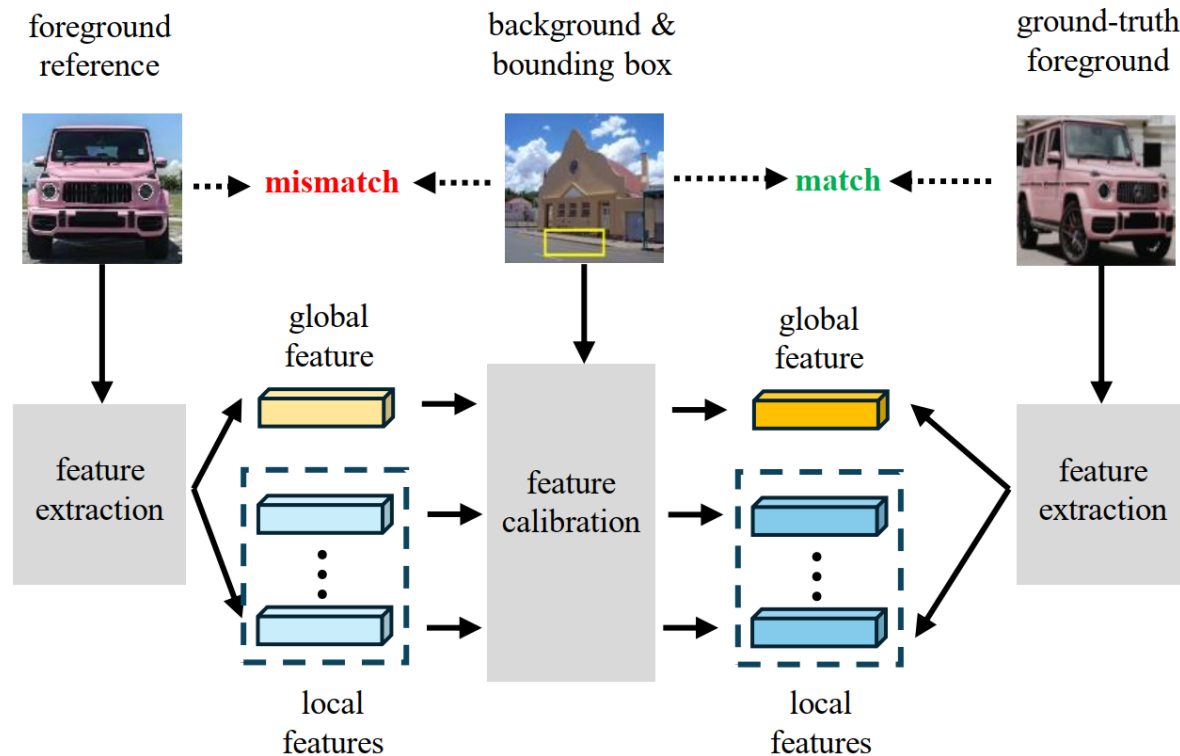


- In practice, there are only a few reference images, how about hallucinating more reference images?
 - We hallucinate reference features instead of reference images.
 - We attempt to hallucinate matched reference features.

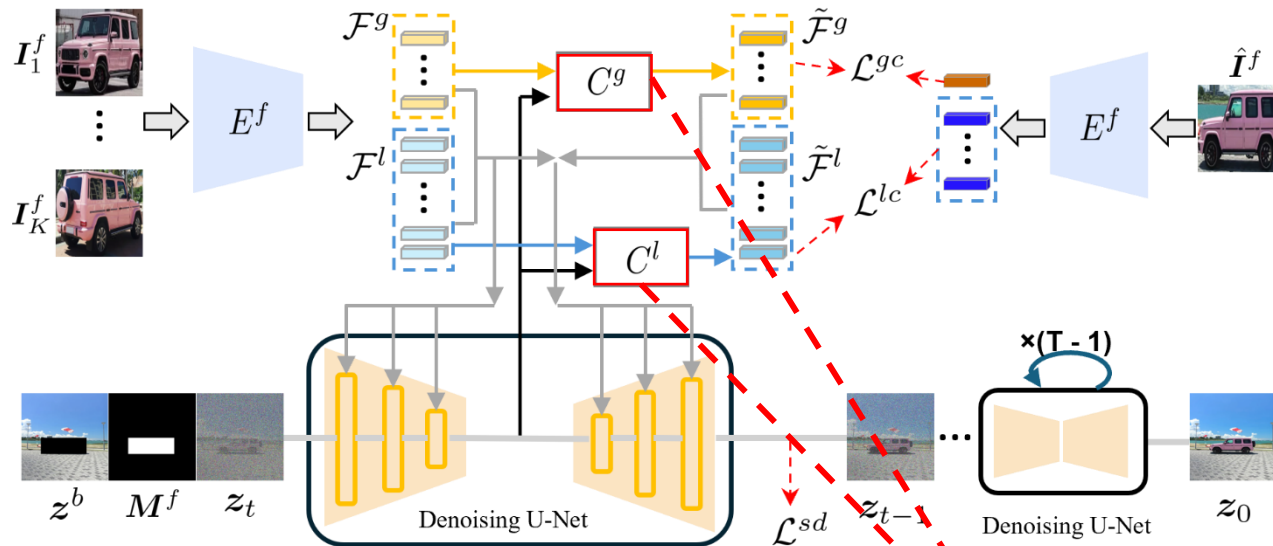
Generative Composition



Based on the background and foreground bounding box, we calibrate the features of foreground reference images to match those of ground-truth foreground.

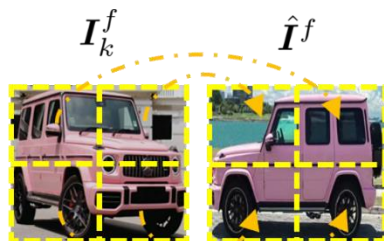


Generative Composition



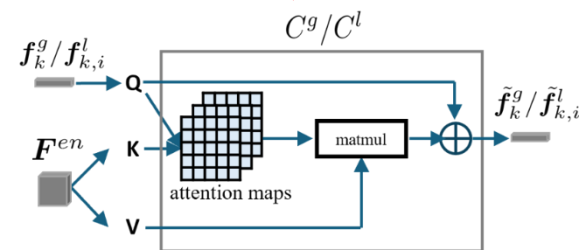
(a)

local correspondence



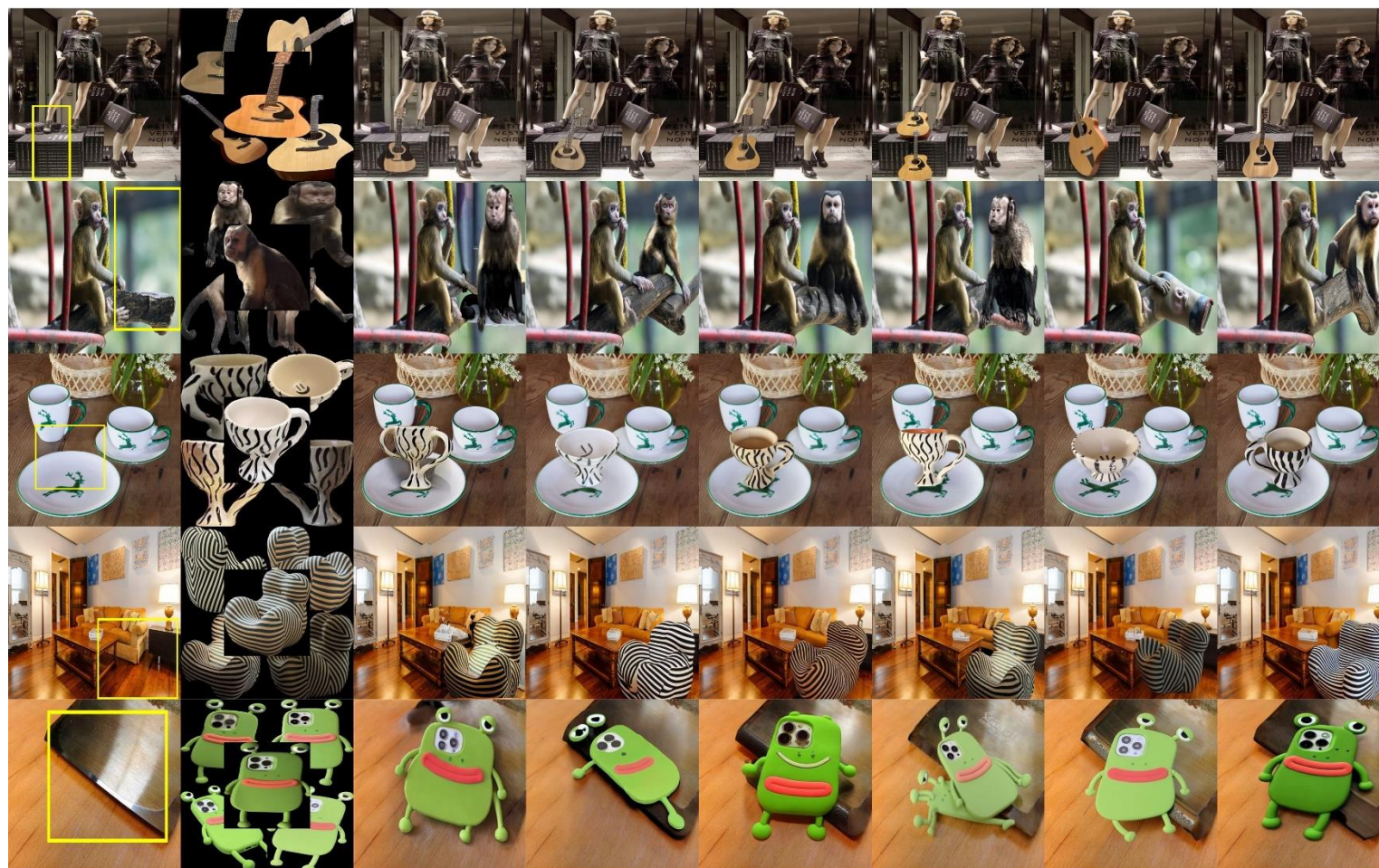
(b)

feature calibration



(c)

Generative Composition



Background

Reference

Anydoor

ControlCom

ObjectStitch

Insert Anything

UniCombine

Ours



Image Composition Resources



Survey: <https://arxiv.org/abs/2106.14490>

Toolbox (libcom): GitHub: <https://github.com/bcmi/libcom>

Document: <https://libcom.readthedocs.io/en/latest/>

Thanks!

