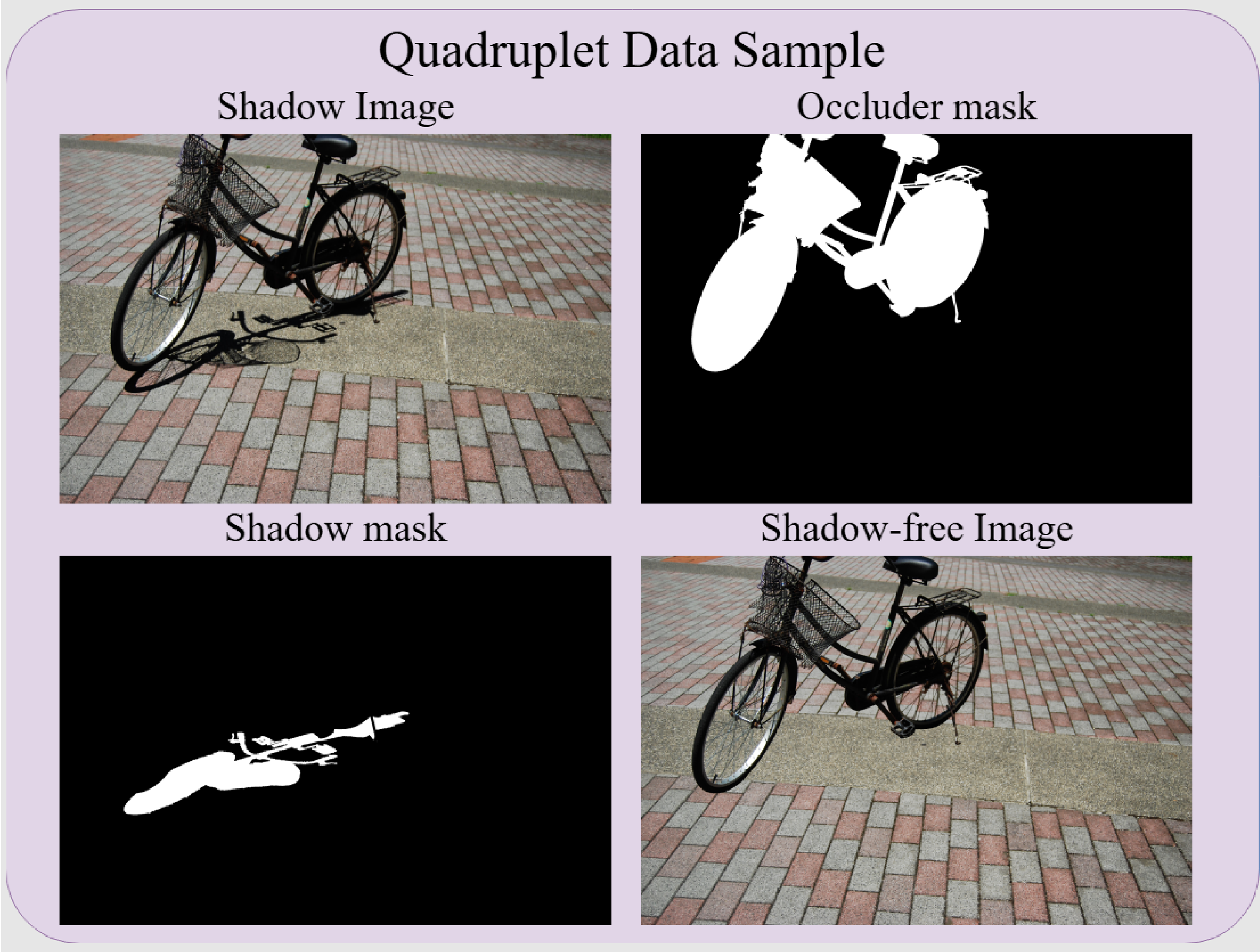
ADSP: Advanced Dataset for Shadow Processing, Enabling Visible Occluders Via Synthesizing Strategy.

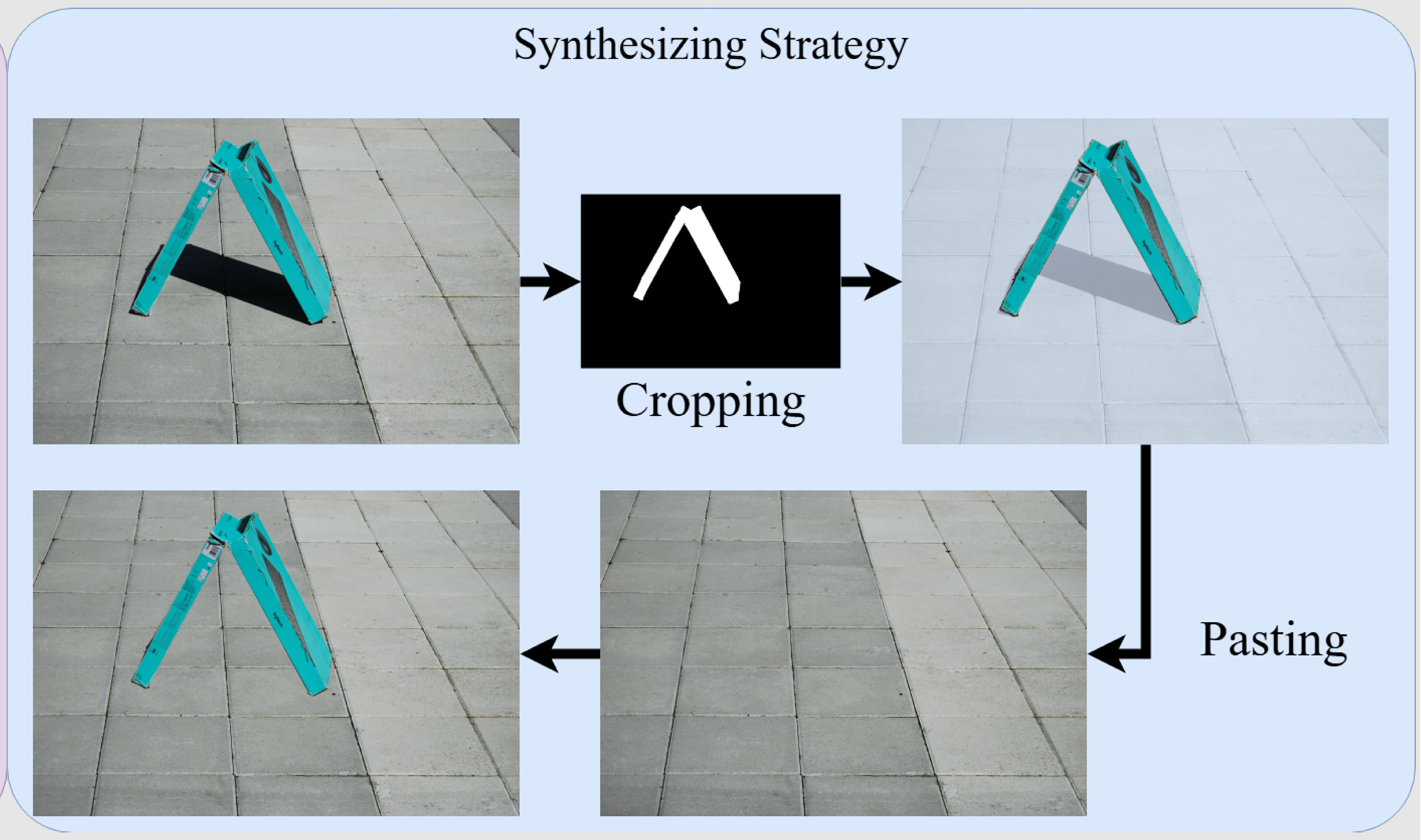
Hsieh Chang-Yu [1] Ding Jian-Jiun [1]

[1]Graduate Institute of Communication Engineering, National Taiwan University, Taipei, Taiwan

Introduction

Shadow is a challenging degradation to high-level computer vision tasks, weakening their performance and robustness. Thus, shadow removal, which recovers underlying information from polluted images, is a crucial problem. However, most existing works deal with simple, sometimes impractical conditions. In this work, we proposed a new **Advanced Dataset for Shadow Processing (ADSP)**, handling these problems at the dataset level. It adopts a novel **synthesizing strategy** to generate shadow-free images with **visible occluders**, which makes its contents closer to real-world applications. Shadow and occluder masks that have pixel-level accuracy also make it the first shadow dataset with **quadruplet data**. Furthermore, benefiting from the rigorous acquiring process, the inconsistency problems were well-suppressed. Overall, the ADSP has high quality and excellent generalization capability, which is helpful in training models with higher robustness to in-the-wild images. We also proposed a three-staged baseline model named Segmented Refinement Removal Network (SRRN) to set a reference for the removal task, achieving state-of-the-art performance on our ADSP.





- Containing **visible occluder** helps to provide semantic information closer to the real world's practical situation.
- Quadruplet data are feasible for training most shadow-related tasks, like detection, removal, and generation. It is also feasible for **multi-task training strategy**.
- The proposed synthesizing strategy has **enormous potential** to provide high-quality, diverse data. In addition, it is also feasible for indoor data collection.

Data-acquiring process

The collection follows the steps below, designed for **mitigating inconsistency problems**, **avoiding remaining shadows**, and **enabling visible occluders**.

1. Collecting multiple highly manipulable pure scene images, denoted as I_{scene} (w/o visible shadows and occluders), as the candidate synthesizing baseboard.
2. Adding occluders into the environment and collecting the shadow-affected images I_s (w/ shadow and occluder).
3. Labeling I_s to get shadow mask M_s and occluder mask M_o by LabelMe.
4. Performing the proposed synthesizing strategy to generate I_{sf} with I_s, M_o, and the optimal I_{scene}.

During collection, **tripod** and **wireless controller** were adopted to suppress inconsistency caused by camera movement, and camera mode was set to **manual** to fix each parameter. During labeling, we manually annotated masks to ensure they reached pixel-level accuracy.

Baseline model

The baseline model has three subnets and two-staged training. For stage 1, we adopted the Charbonnier Loss to train a Removal Network to perform preliminary removal. Then, in stage 2, two refinement networks were optimized by different loss functions to mitigate two artifacts that frequently appear in removal. Shadow region loss was used in the Color Adjustment Network to deal with the color bias of the shadow area. On the other hand, the penumbra loss drives the Boundary Smoothing Network to eliminate the boundary effect among the penumbra region. Experimental results show that the SRRN can the reduce visually unpleasantness caused by artifacts effectively.

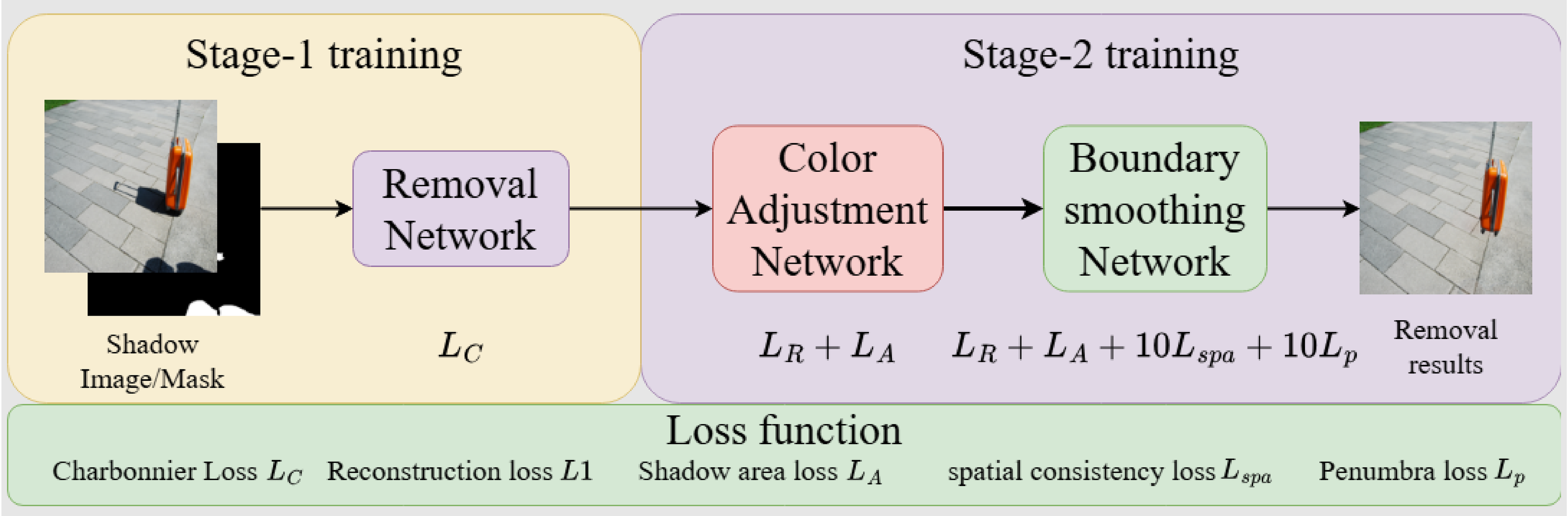


Figure 1. The simplified procedure of the proposed baseline model.

Experiments

Table 1. Basic information of included datasets in domain shift experiments. Where $*$ means that such dataset contains majority contents with visible occluder.

Eval\Train	SRD(*)	ISTD	DESOBAv2*	ADSP* (ours)
No. of pair	2680/408	1330/540	20000/296	1100/120
Resolution	640x840	480x640	256x256	2592x3872

Table 2. Number of first/second place for each benchmarks of four metrics from two SOTA models, i.e. there are 8 ranking in each evaluation.

Eval\Train	SRD(*)	ISTD	DESOBAv2*	ADSP* (ours)
SRD(*)	-	3/2	2/2	3/4
ISTD	2/5	-	4/2	2/1
DESOBAv2*	1/6	0/1	-	7/1
ADSP* (ours)	1/5	0/2	7/1	-

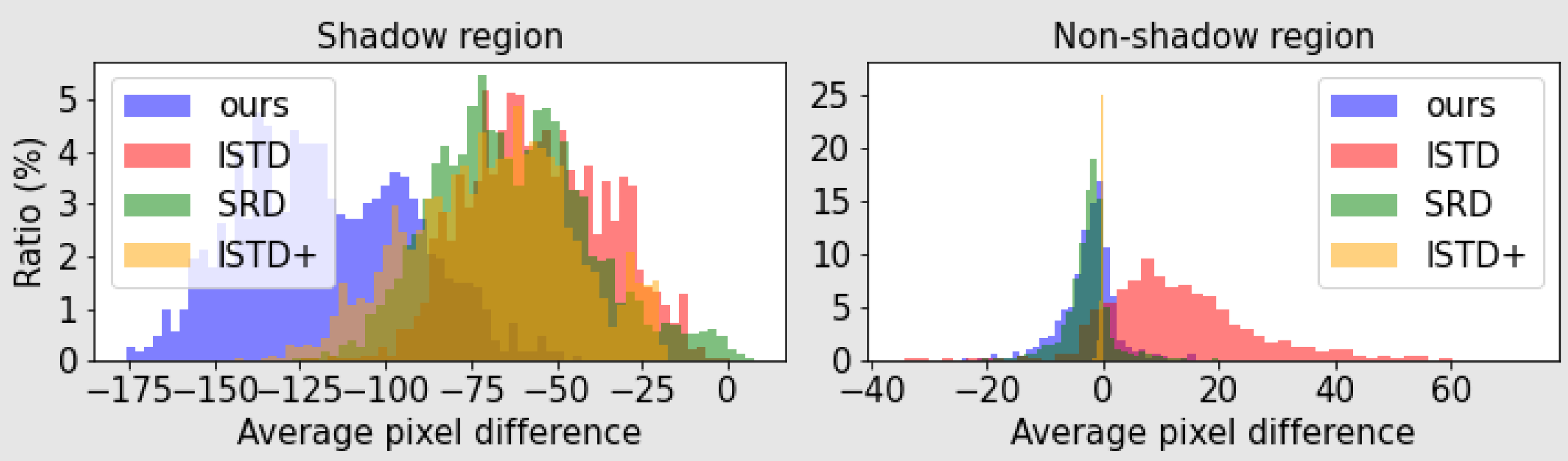


Figure 2. Pixel value difference among shadow/non-shadow regions.

Evaluations from statistical and learning perspectives verify the proposed ADSP's superiority. First, in statistical analysis, it shows **more significant challenges** because of the stronger degradation, resulting from darker shadows. Furthermore, a more rigorous setup enables it to suppress inconsistency problems effectively. On the other hand, in domain shift exps, our ADSP shows well **generalization capability** prevailing existing benchmarks even under the evident disadvantage of data amount. The ADSP received most of the top rankings when applying on in-the-wild images with visible occluder. While applied to popular data used currently, it also has enough strength.

github   project page   paper