Image Retargeting

2022 연구실 동계 세미나

민 성 준

Vision & Display Systems Lab.

Dept. of Electronic Engineering, Sogang University

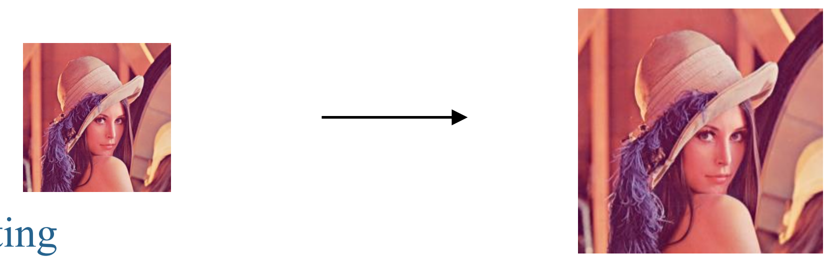
Introduction

Outline

- Introduction
- Cycle-IR
- Two-Directional Image Retargeting

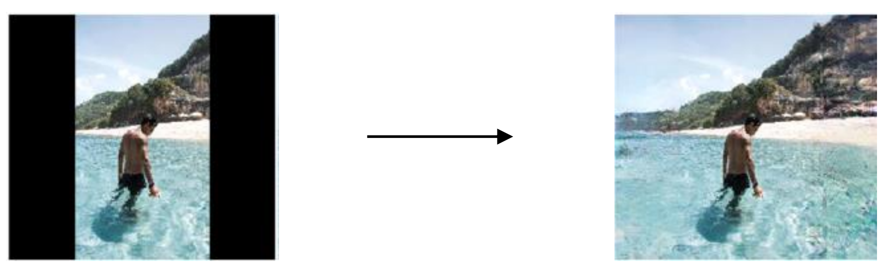
Introduction

- Image Resolution
  - Super Resolution



  - Image Retargeting

  - Image Outpainting

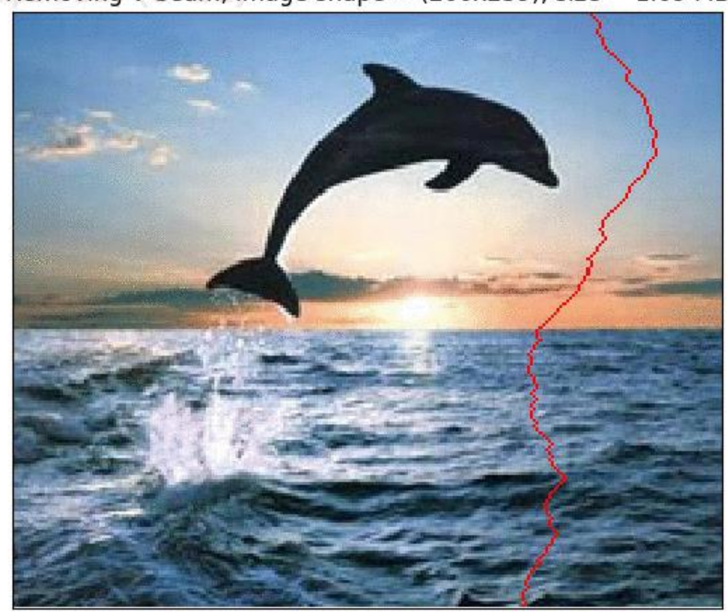


Introduction

- Image Retargeting
  - Seam carving
    - – Seam: 화소간 에너지 차이를 비교하여 가장 작은 에너지 패턴

Removing V-Seam, image shape = (225x225), size = 1.16 MB

Removing V-Seam, image shape = (200x239), size = 1.09 MB

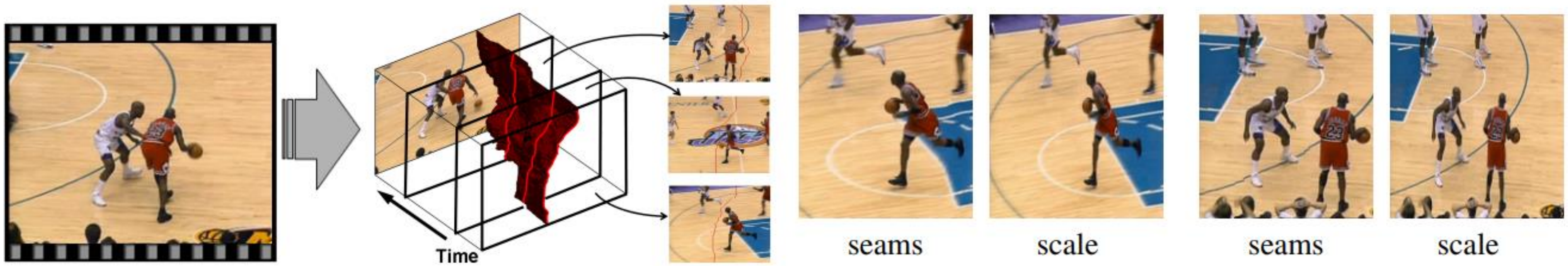


Introduction

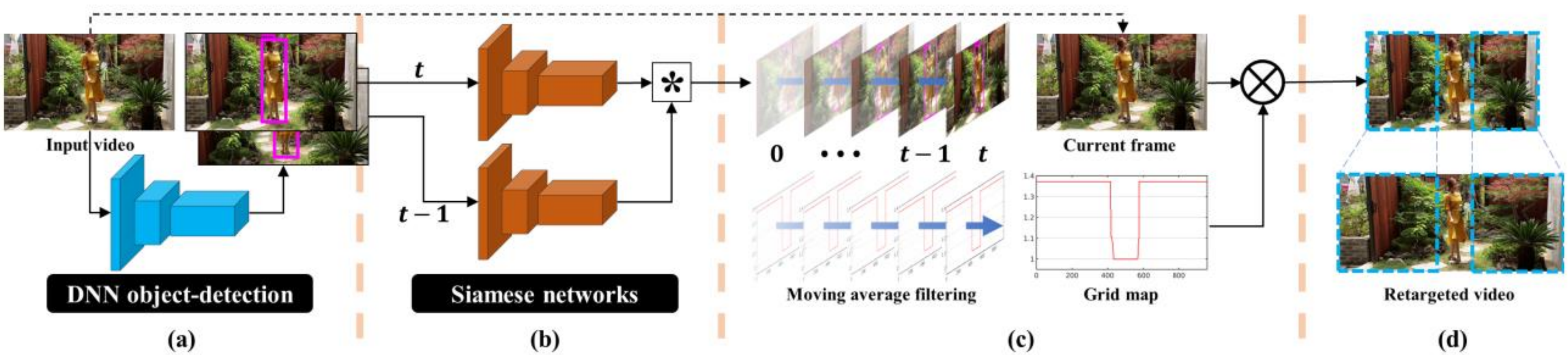
- Video Retargeting
  - Image Retargeting + Temporal coherence

Introduction

- Video Retargeting
  - Image Retargeting + Temporal coherence



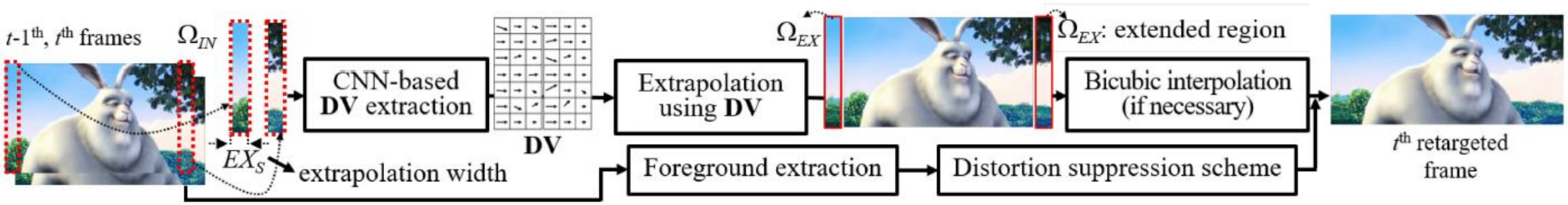
< Video cube-based method >



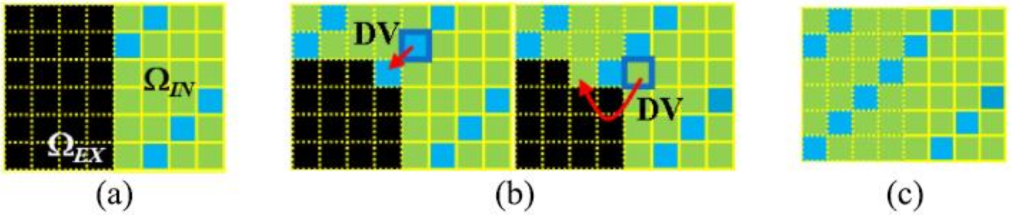
< Object-tracking & nonlinear warping method >

Introduction

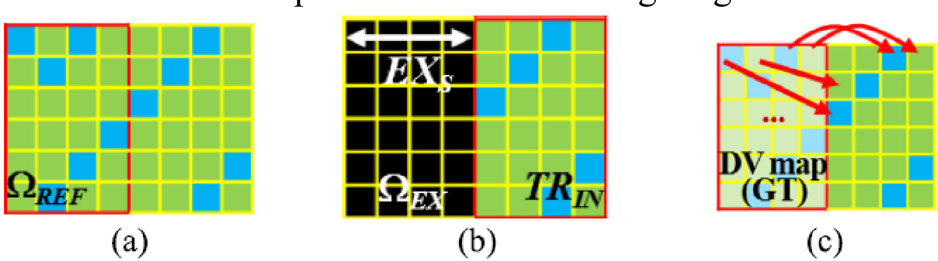
- Video Retargeting
  - Image Retargeting + Temporal coherence



< The proposed image-to-warping vector CNN-based extrapolation for video retargeting >



< Extrapolation for video retargeting >



< Training data generation for CNN-based extrapolation >

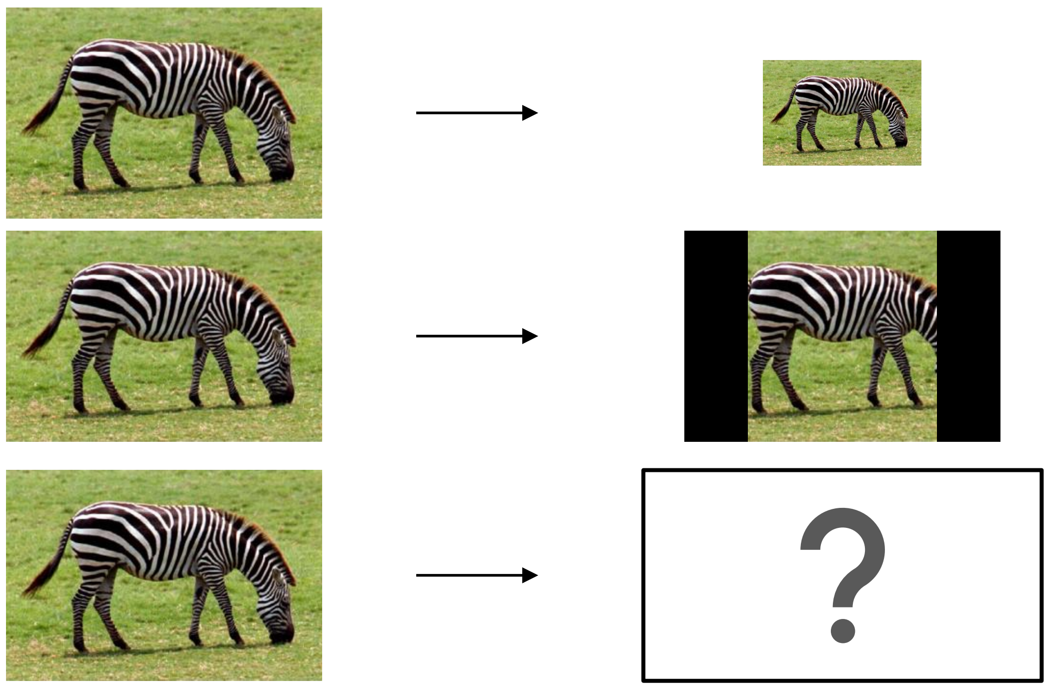
Cycle-IR[1]

- Motivation
  - GT – input image pair…?
    - The way to produce the ideal retargeting results is uncertain
      - Unsupervised! → doesn't require any labeled data!

SR

Outpainting

Retargeting



Cycle-IR

- Method
  - Downscale → upscale = input
  - Upscale → downscale = input
    - How..?
      - Dual output & cyclic perception loss

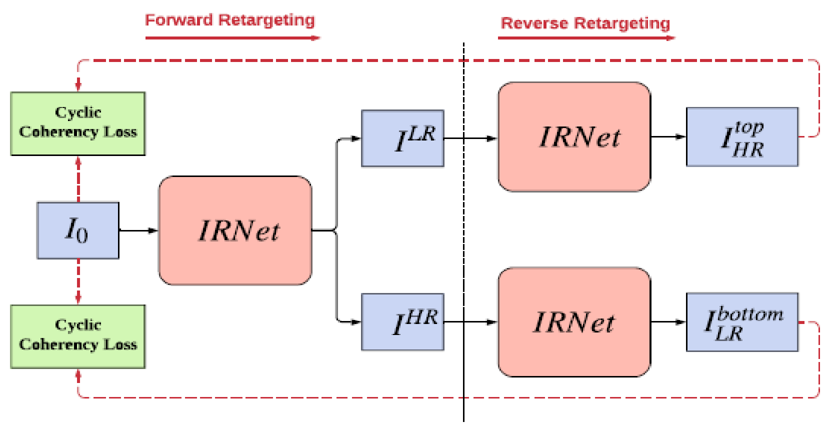


Fig. 1. Overview of the proposed Cycle-IR framework. By cyclically utilizing the retargeting results of the forward inference of deep image retargeting network (IRNet), our IRNet can be trained in an unsupervised way, without the requirement of any manual label.

Cycle-IR

- Method
  - Forward Retargeting

$$I^{LR}, I^{HR} = IRNet_{FWD}(I_0)$$

  - Reverse Retargeting

$$I^{top}_{LR}, I^{top}_{HR} = IRNet_{REV}(I^{LR})$$

$$I^{bottom}_{LR}, I^{bottom}_{HR} = IRNet_{REV}(I^{HR})$$

  - Cyclic Perception Coherence Loss
    - f_l: pretrained VGG16 l^{th} layer

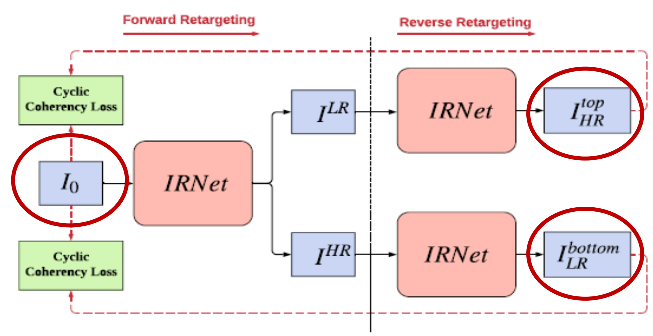


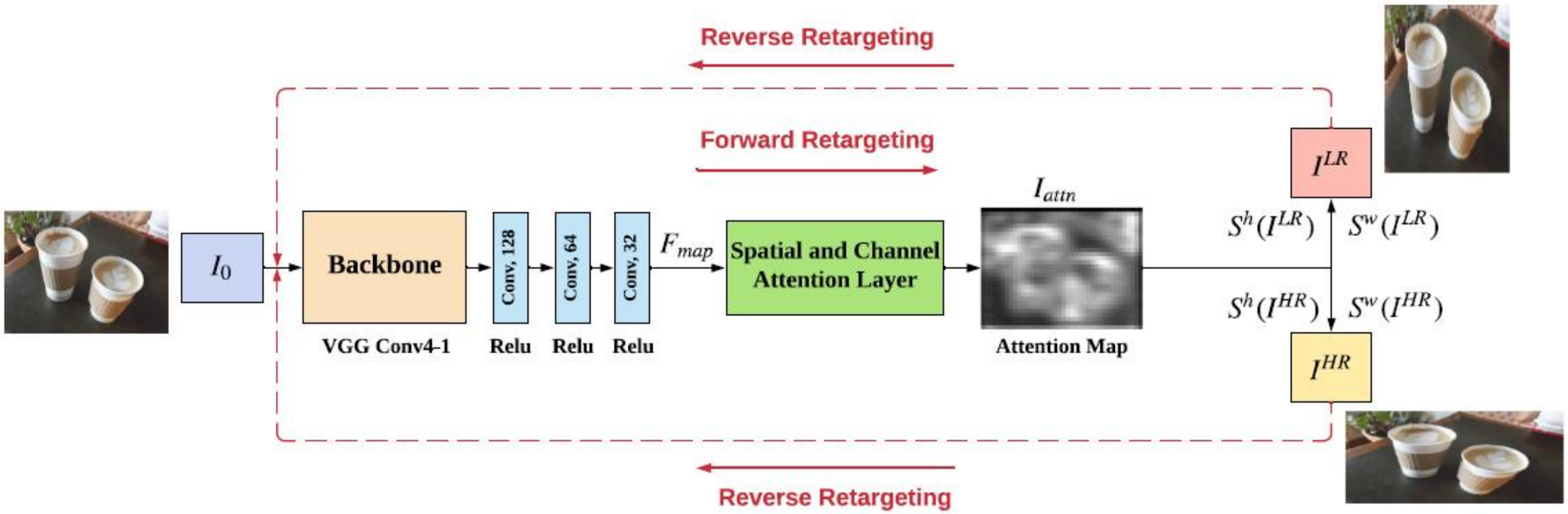
Fig. 1. Overview of the proposed Cycle-IR framework. By cyclically utilizing the retargeting results of the forward inference of deep image retargeting network (IRNet), our IRNet can be trained in an unsupervised way, without the requirement of any manual label.

$$\mathcal{L}_{pair} = \frac{1}{L}\sum_{l=4}^{L}[(f_l(I_0) - f_l(I^{top}_{HR})) \times \beta_l]^2$$

$$+ \frac{1}{L}\sum_{l=4}^{L}[(f_l(I_0) - f_l(I^{bottom}_{LR})) \times \beta_l]^2$$

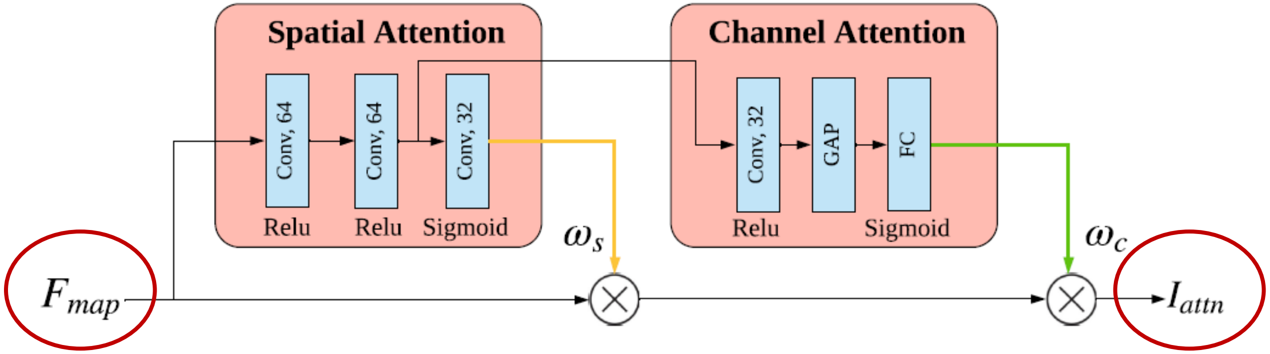
Cycle-IR

- Obtaining Visual Attention Map
  - F_{map}: Backbone (Conv4-1 of VGG16) + 3 convolution layer
  - Spatial and Channel Attention Layer
    - $I_{attn} = \Gamma_{attn}(F_{map})$
      - Attention map



Cycle-IR

- Obtaining Visual Attention Map
  - Spatial and Channel Attention Layer
    - F_{map}: extracted deep representation
    - I_{attn} : attention map (visualization)
      - $I_{attn} = \Gamma_{attn}(F_{map})$
      - Input of channel attention component
        - ✓Output of spatial attention, not F_{map}



Cycle-IR

- Generating desired target image
  - Calculate scaling factor for each grid cell for reconstructing I^{LR}
    - Using guidance of visual attention map I_{attn}

$$S_i^h(I^{LR}) = \frac{1}{N}\sum_{j=1}^{N}\frac{1}{1+e^{-I_{attn}(i,j)}}$$

$$S_j^w(I^{LR}) = \frac{1}{M}\sum_{i=1}^{M}\frac{1}{1+e^{-I_{attn}(i,j)}}$$

  - Calculate scaling factor for each grid cell for reconstructing I^{HR}

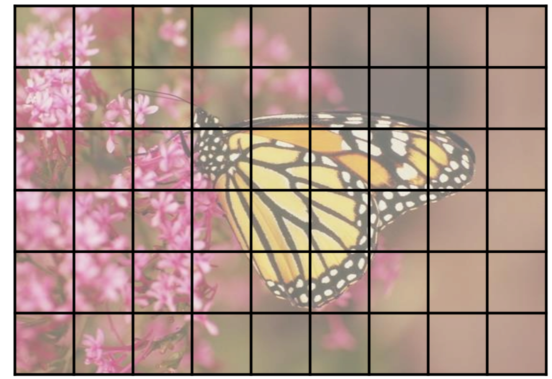
$$S_i^h(I^{HR}) = 1 - S_i^h(I^{LR}) + \psi_h$$

$$S_j^w(I^{HR}) = 1 - S_j^w(I^{LR}) + \psi_w$$

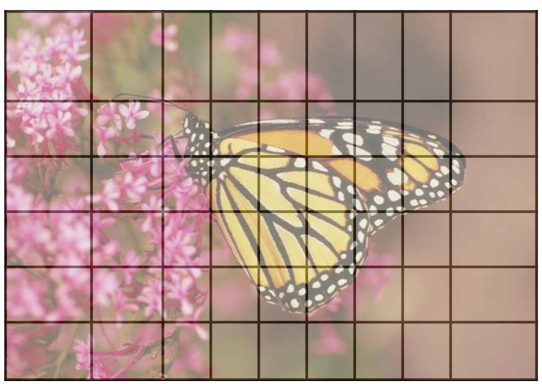
Cycle-IR

- Generating desired target image

Grid cell



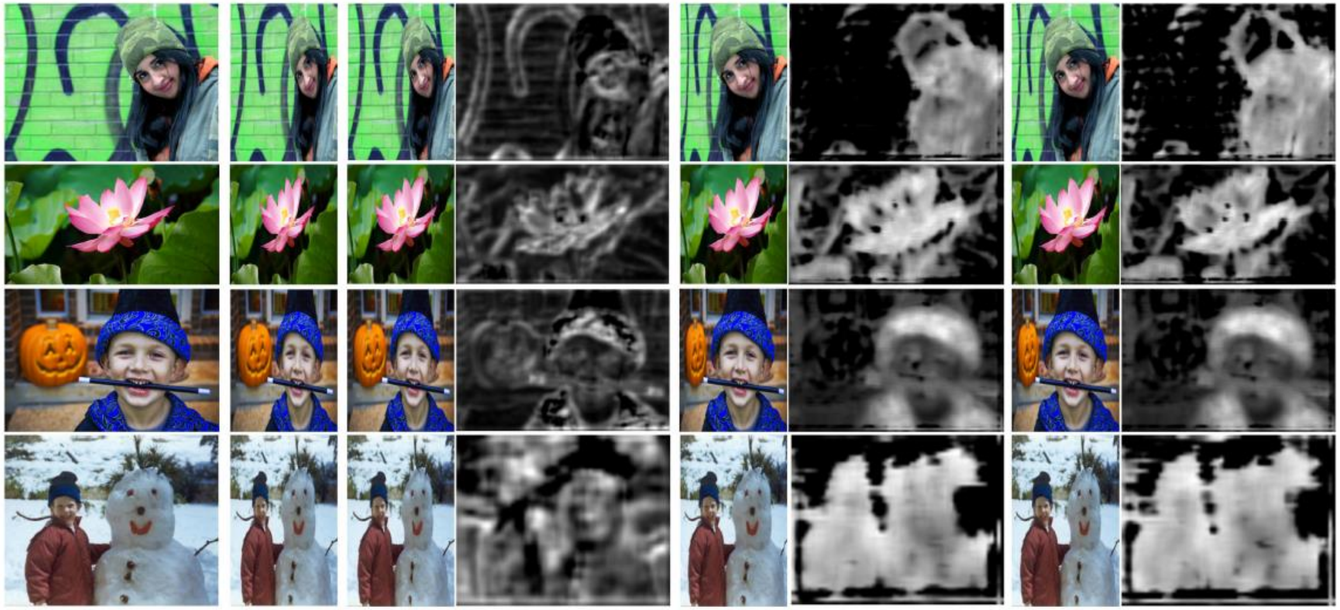
Scaling factor
${S_i}^h(I^{LR}), {S_j}^w(I^{LR})$



Generate result

Cycle-IR

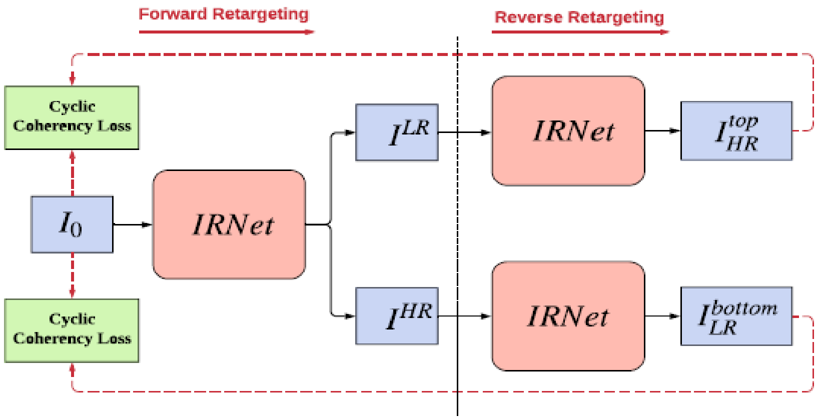
- Pixel-wise Loss vs Perceptual Loss
  - Pixel-wise loss
  - **Perceptual loss**
  - Pixel-wise + perceptual loss



(a) Input image (b) SCL (c) Pixel-Wise Loss (d) Perceptual Loss (e) Pixel-Wise + Perceptual

Cycle-IR

- Single Cycle Loss vs Pair Cycle Loss
    - Single Cycle Loss (top)
    - Single Cycle Loss (bottom)
    - **Pair Cycle Loss**



(a) Input image (b) Single Cycle Loss (bottom) (c) Single Cycle Loss (top) (d) Pair Cycle Loss (employed)

Cycle-IR

- Saliency Guidance
  - Saliency 정보를 활용하자
    - HKU-IS dataset: priori information (GT saliency) embedded
    - Pixel-wise loss + saliency map
      - Fail..
    - Perceptual loss + saliency map
      - Fail…

(a) Input image (b) Saliency map (c) Learned importance map

(a) Input image (b) Pixel-Wise Loss + Saliency map (c) Perceptual Loss + Saliency map

Fig. 7. Demonstration of integrating saliency maps as guidance with pixel-wise loss or perceptual loss.

Cycle-IR

- Results

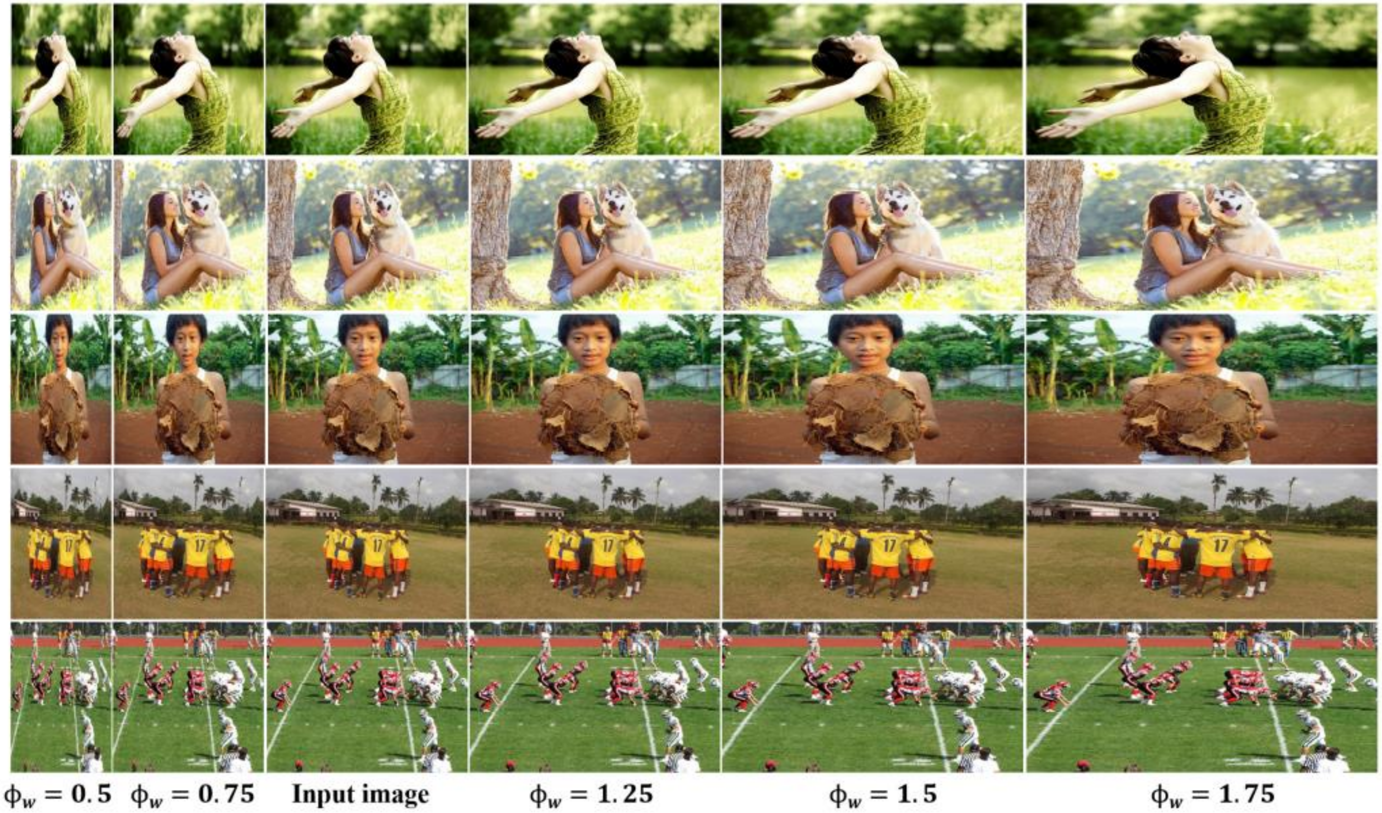


Fig. 8. Demonstration of the ability of our Cycle-IR to generate target images with arbitrary sizes. Despite the large scale span (from 0.5 to 1.75), our Cycle-IR is able to generate satisfactory target images.

Cycle-IR

- Results

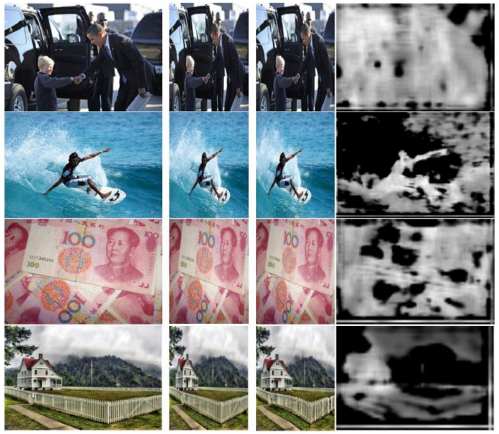
Fig. 9. Visual comparison of our Cycle-IR with representative retargeting approaches. The images are retargeted to half width.

Cycle-IR

- Results

↓ Outperforms →	ASAP	SC	SM	SNS	Warp	Multiop	Cycle-IR	Total	Prefer
ASAP	-	134	165	134	159	80	80	752	55.7%
SC	91	-	139	94	133	71	74	602	44.6%
SM	60	86	-	71	121	50	54	442	32.7%
SNS	91	131	154	-	157	72	71	676	50.1%
Warp	66	92	104	68	-	24	27	381	28.2%
Multiop	145	154	175	153	201	-	111	939	69.5%
Cycle-IR	145	151	171	154	198	114	-	933	69.1%

< User preference study >



(a) Input image (b) SCL (c) Our Cycle-IR

< Some failure cases >

Cycle-IR

- Conclusion
  - Unsupervised deep image retargeting approach
    - Doesn’t need any labeled data, additional parameter settings, or human assistance
  - IRNet model outputs a pair of retargeted images
  - Cyclic perception coherence loss
    - Evaluates cycle coherence between the forward-backward mapping results
    - Can be applied to other image retargeting methods
  - Spatial and channel attention layer
    - Able to discover visually interesting areas of input images
      - Visualization
      - Learnable network to help assist the IRNet in learning accurate attention map

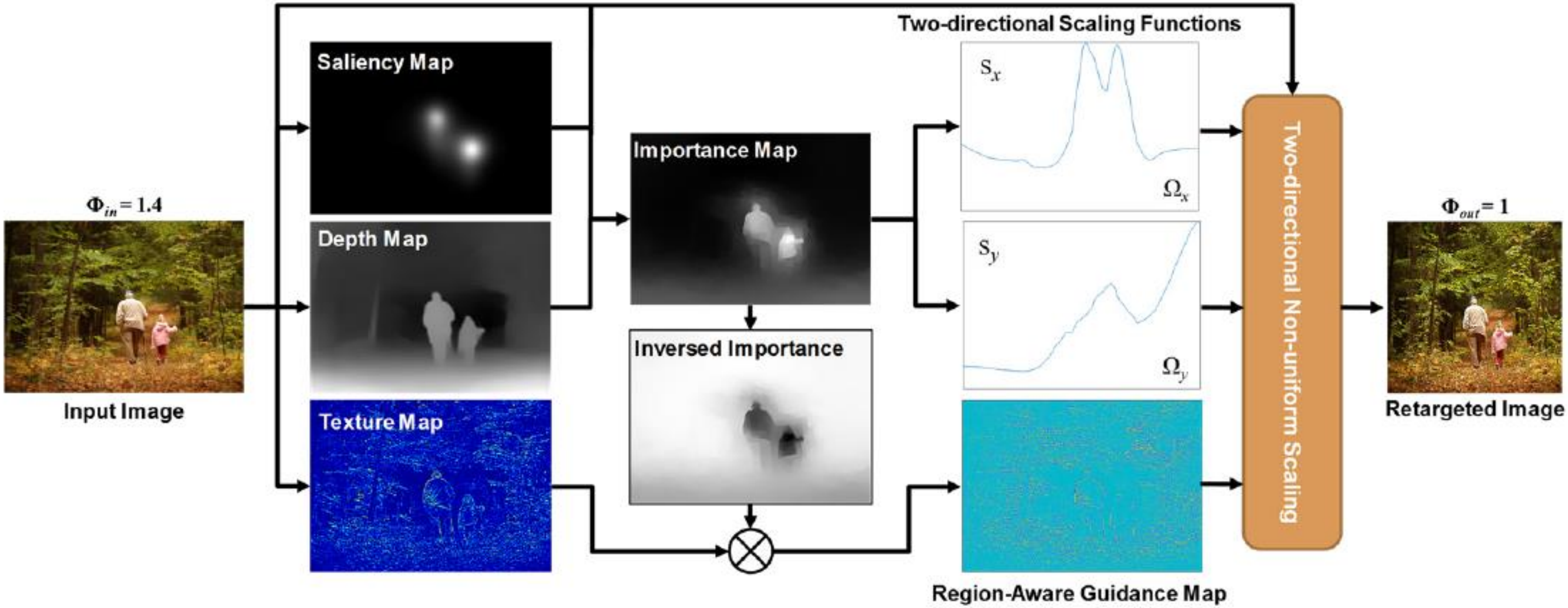
Two-Directional Image Retargeting[2]

- Contribution
  - Two-directional non-uniform scaling according to the statistical distribution of scaling functions
    - Salient regions of input image can be preserved
      - Two-directional non-uniform scaling allows more input marginal regions
  - Use of region-aware guidance map from inversed importance map
    - Compensates quality degradation from non-uniform scaling

< Difference of proposed approach compared with the unidirectional non-uniform scaling >

Two-Directional Image Retargeting

- Overview of Proposed Approach



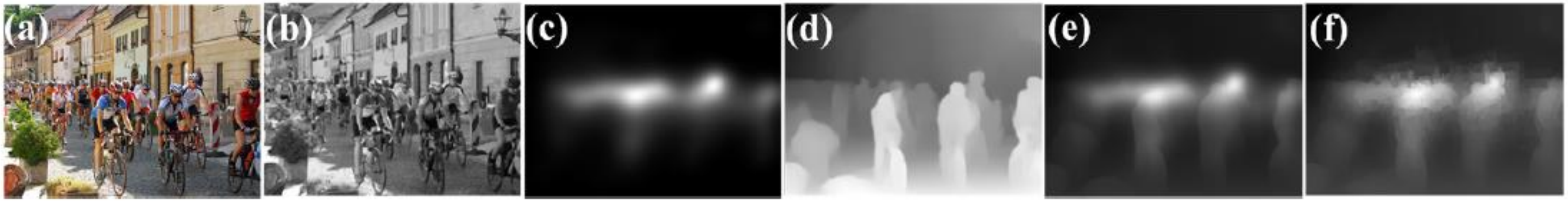
< Overview of proposed approach >

Two-Directional Image Retargeting

- Importance Map
  - Saliency detection
    - $B(i,j)$: UNISAL[3]
  - Monocular depth estimation
    - $Q(i,j)$: MiDas[4]
  - Importance Map = saliency & depth
    - $A_{init}(i,j) = \lambda \cdot B(i,j) + (1-\lambda) \cdot Q(i,j)$
    - Refinement with WLS solution

$$(\mathbf{A}-\mathbf{A}_{init})^T(\mathbf{A}-\mathbf{A}_{init})+\alpha(\mathbf{A}^T\mathbf{D}_x^T\mathbf{P}_x\mathbf{D}_x\mathbf{A}+\mathbf{A}^T\mathbf{D}_y^T\mathbf{P}_y\mathbf{D}_y\mathbf{A}),$$

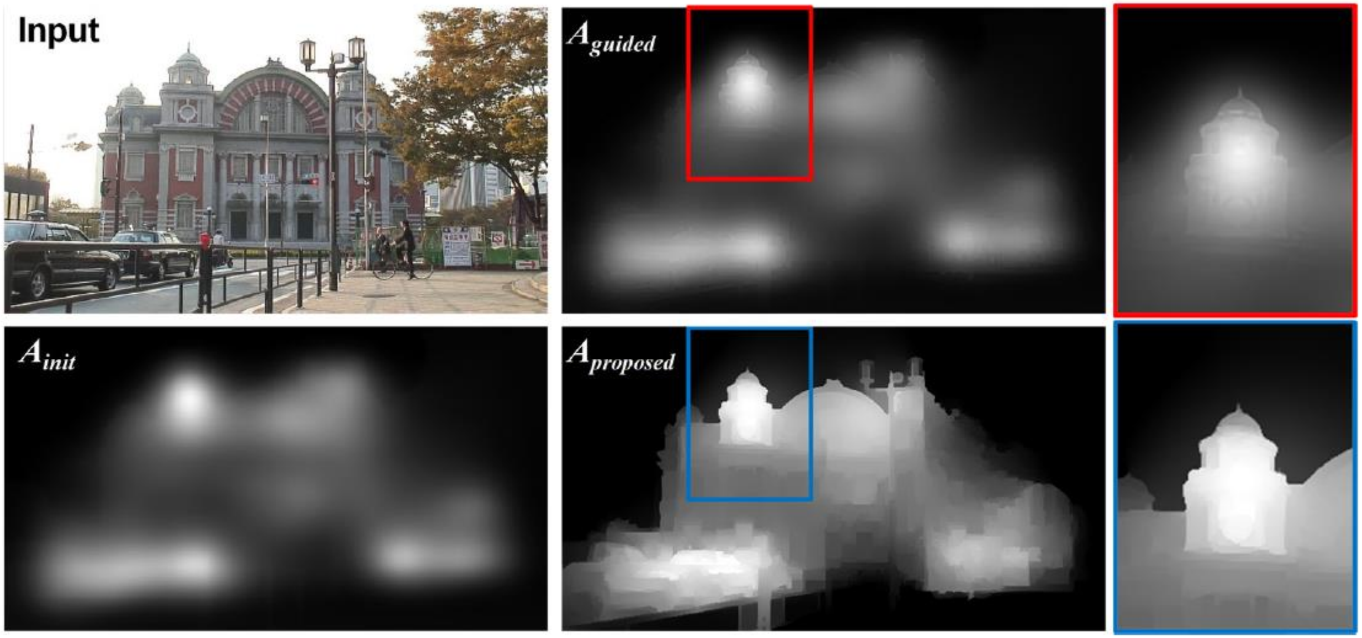
$$\mathbf{A}=\left(\mathbf{U}+\alpha(\mathbf{D}_x^T\mathbf{P}_x\mathbf{D}_x+\mathbf{D}_y^T\mathbf{P}_y\mathbf{D}_y\right)^{-1}\mathbf{A}_{init},$$



< Importance map estimation >

Two-Directional Image Retargeting

- Importance Map
  - Refinement
    - Bilateral filetr / guided filter / WLS(Weighted Least Squares filter
    - Removing halo effect



< Importance map estimation – guided filter / proposed >

Two-Directional Image Retargeting

- Two-Directional Scaling Functions
  - Proposed method defines two-directional scaling ratios in each direction
    - Using entropy-based dispersion measure
      - How “evenly” is the probability distributed
        - ✓Scaling function is used
          - Importance map is used

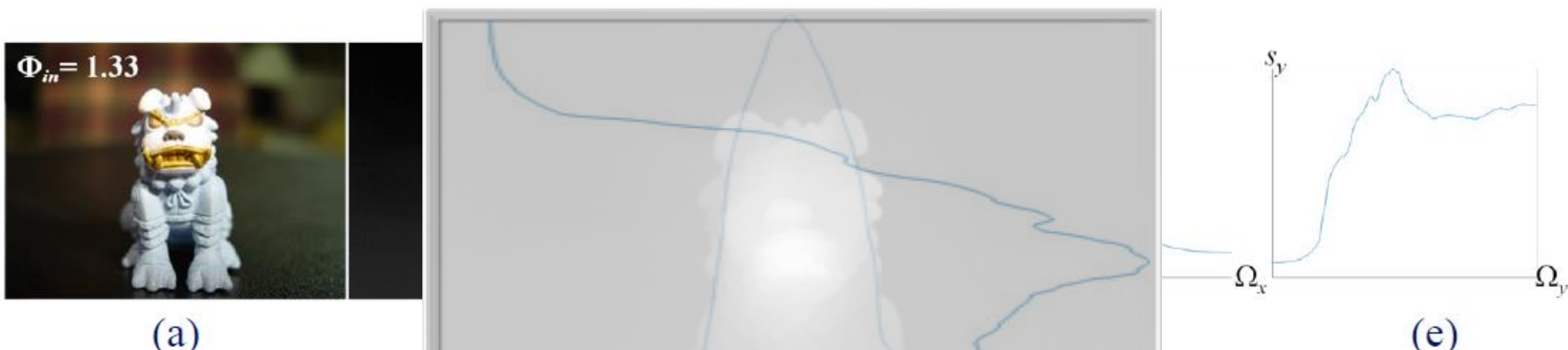


Figure 5: **Bidirectional** … image (400×300), (b) Importance map, (c) Re… scaling function s_x (φ_x = 0.6139, Ω_x = 339), (e) V… = 339).

Two-Directional Image Retargeting

- Two-Directional Scaling Functions
  - Ω_x, Ω_y : Target height, width
  - φ_x, φ_y : Horizontal, vertical scaling ratio
  - $s_x(i), s_y(j)$: Horizontal, vertical scaling function
  - $A(i,j)$: Importance map

$$s_x(i) = \frac{\sum_j A(i,j)}{\sum_i \sum_j A(i,j)} \Omega_x, \qquad s_y(j) = \frac{\sum_i A(i,j)}{\sum_j \sum_i A(i,j)} \Omega_y,$$

$$\Omega_x = w \pm \varphi_x \Delta, \qquad \Omega_y = h \mp \varphi_y \Delta,$$

$$\varphi_x = \frac{\rho_x}{\rho_x + \rho_y}, \qquad \varphi_y = \frac{\rho_y}{\rho_x + \rho_y},$$

$$\rho_x = \exp\left[-C_h(s_x(i))\right], \qquad \rho_y = \exp\left[-C_h(s_y(j))\right],$$

$$\Delta = \begin{cases} \dfrac{h \cdot \Phi_{out} - w}{\varphi_x + \varphi_y \cdot \Phi_{out}} & , \Phi_{out} > \Phi_{in} \\ -\dfrac{h \cdot \Phi_{out} - w}{\varphi_x + \varphi_y \cdot \Phi_{out}} & , \Phi_{out} < \Phi_{in} \end{cases}$$

Two-Directional Image Retargeting

- Two-Directional Scaling Functions

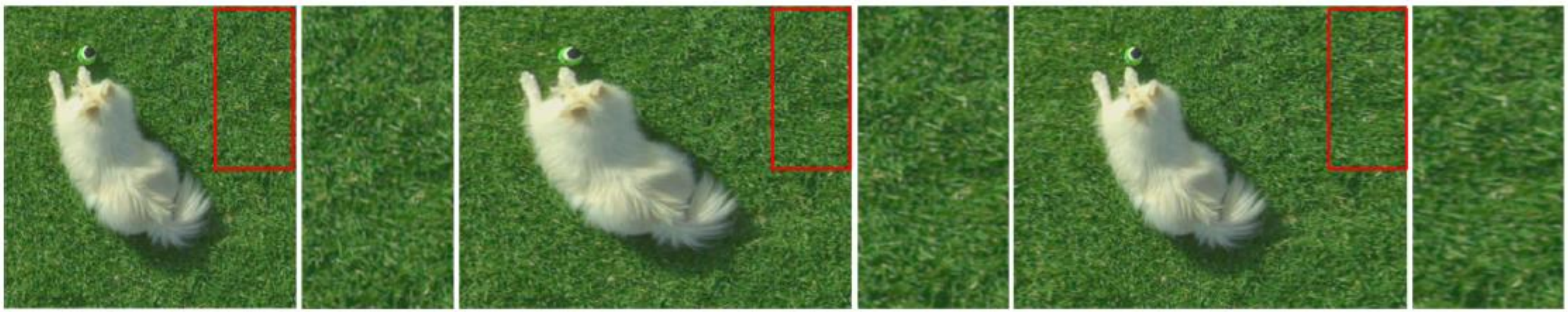
(a) (b) (c) (d) (e)

Figure 5: **Bidirectional scaling functions and ratio.** (a) Input image (400×300), (b) Importance map, (c) Retargeted image (339×339), (d) Horizontal scaling function s_x (φ_x = 0.6139, Ω_x = 339), (e) Vertical scaling function s_y (φ_y = 0.3861, Ω_y = 339).

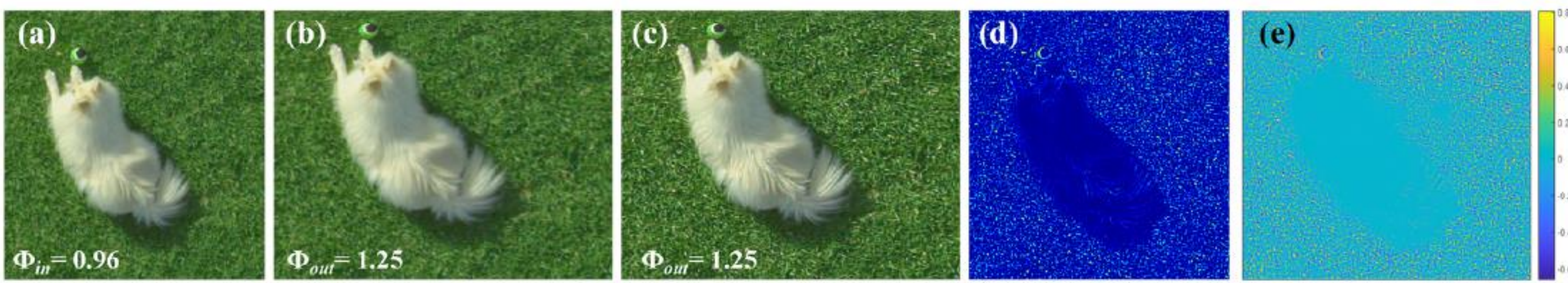
- $400 \times 300 \rightarrow 339 \times 339$
- Compute importance map $A(i, j)$
- Get scaling function $s_x(i), s_y(j)$
- Get scaling ratio φ_x, φ_y
- Get target height, width Ω_x, Ω_y

Two-Directional Image Retargeting

- Region-aware guidance map
  - Problem with uniform scaling
  - Problem with non-uniform scaling



(a) Input image (b) Uniform scaling (c) Non-uniform scaling

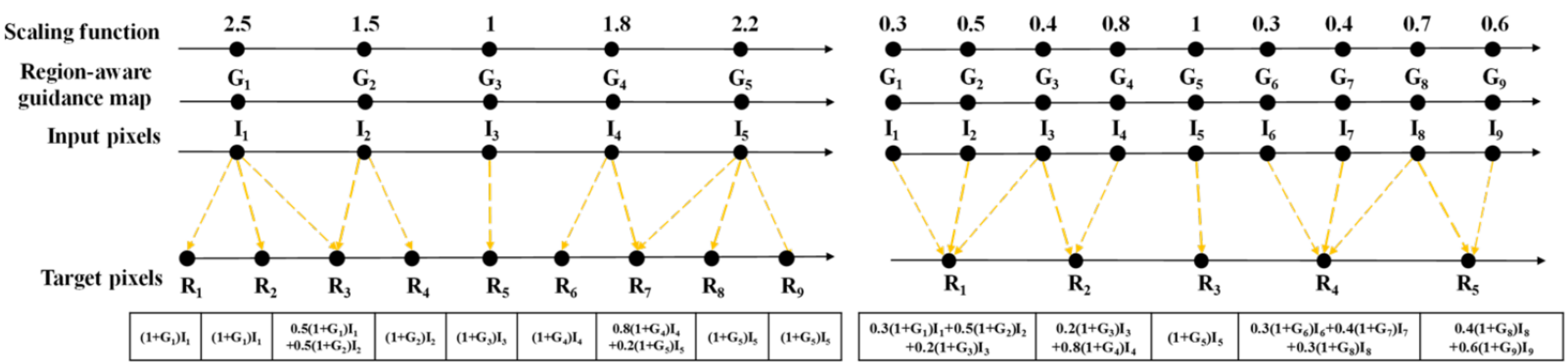


Two-Directional Image Retargeting

- Region-aware guidance map

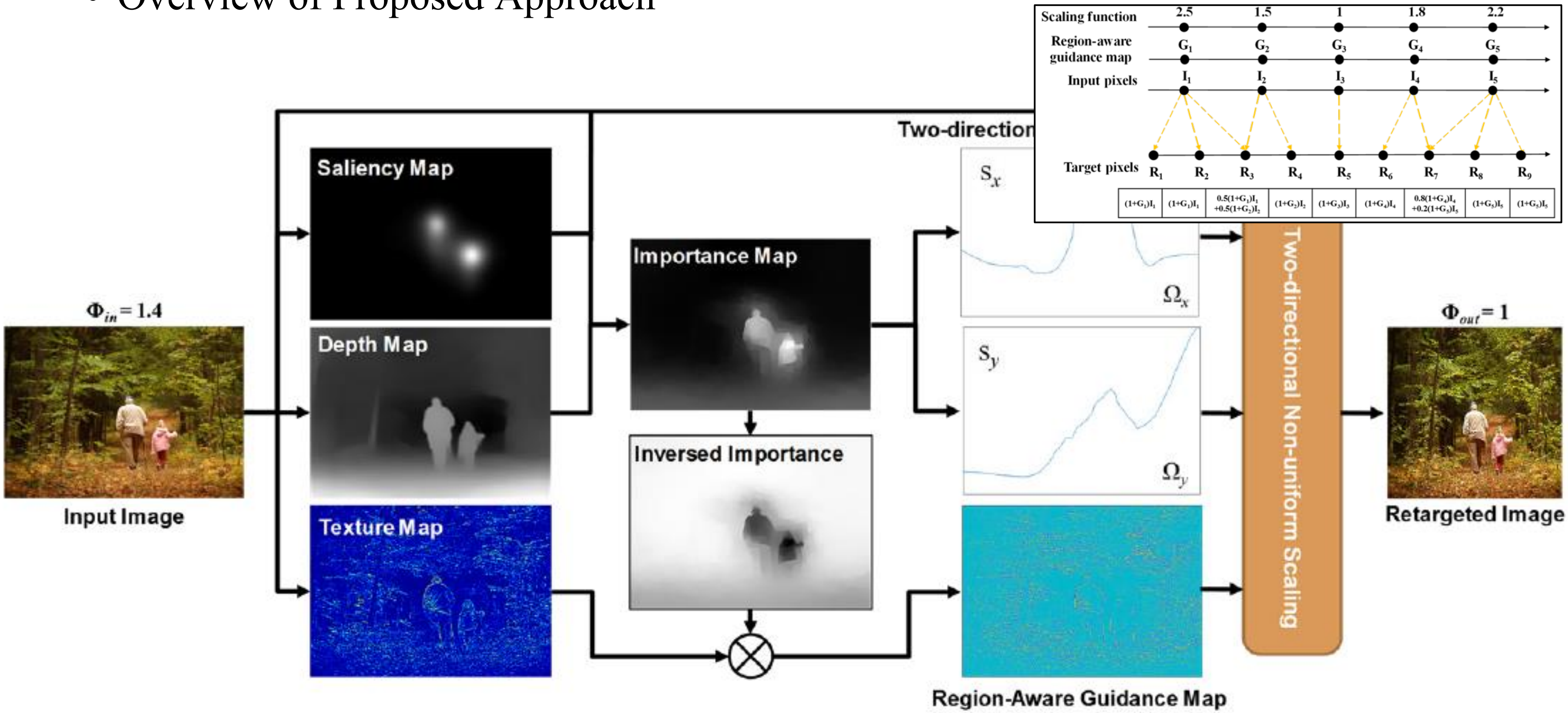
  - To enhance degradation of non-uniform scaling
  - Region-aware guidance map = inversed importance map * texture map

$$G(i,j) = (1 - A(i,j)) \cdot T(i,j),$$



Two-Directional Image Retargeting

- Overview of Proposed Approach



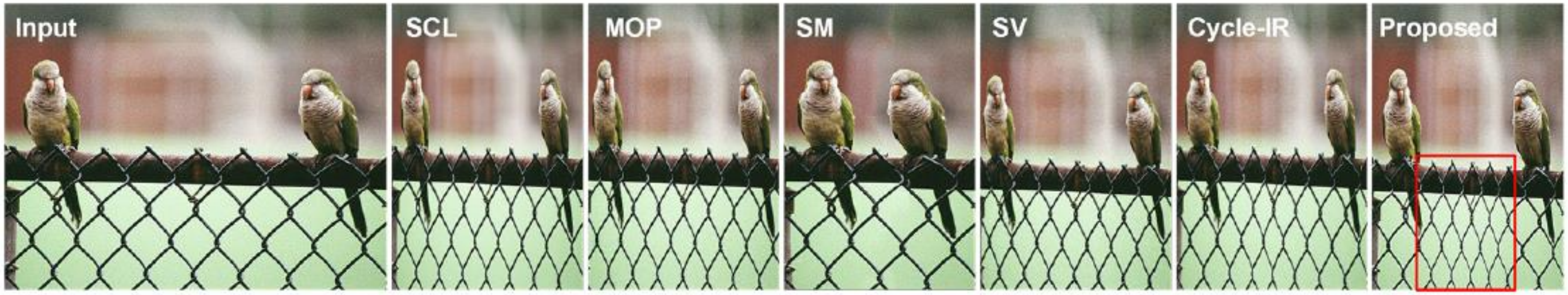
< Overview of proposed approach >

Two-Directional Image Retargeting

- Results

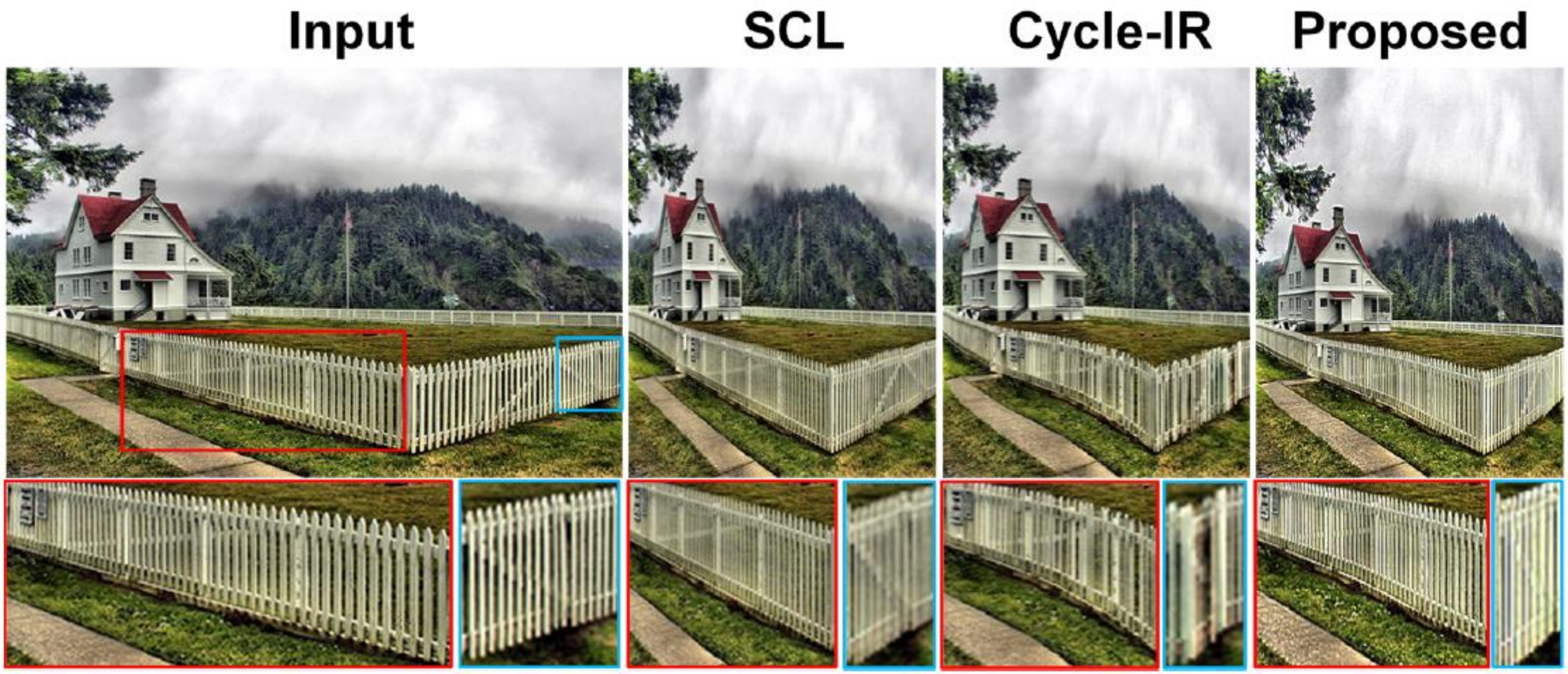
$0.5\Phi_{in}$ $0.75\Phi_{in}$ Φ_{in} $1.25\Phi_{in}$ $1.5\Phi_{in}$

- Failure case
  - No regard for global consistency



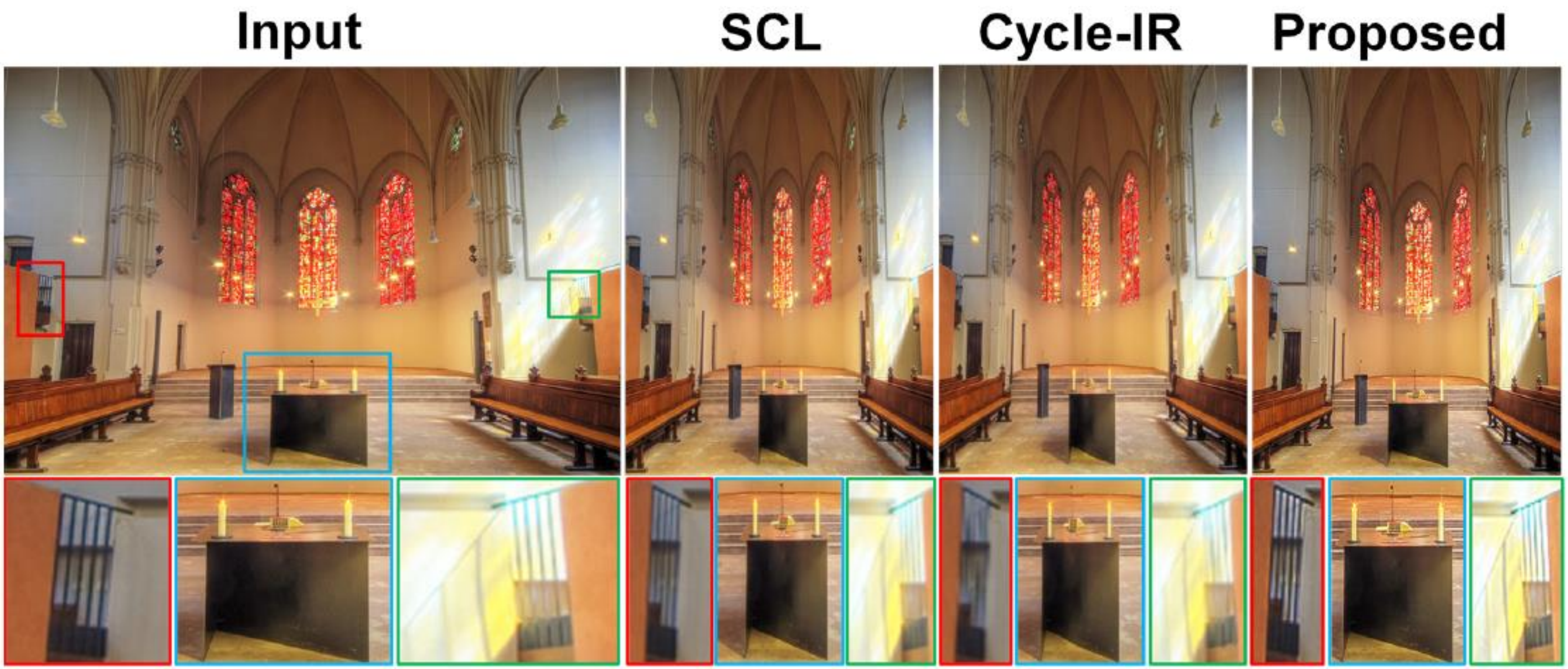
Two-Directional Image Retargeting

- Results
  - Cycle-IR : over-smoothing & twist



Two-Directional Image Retargeting

- Results
  - Cycle-IR : over-smoothing & twist

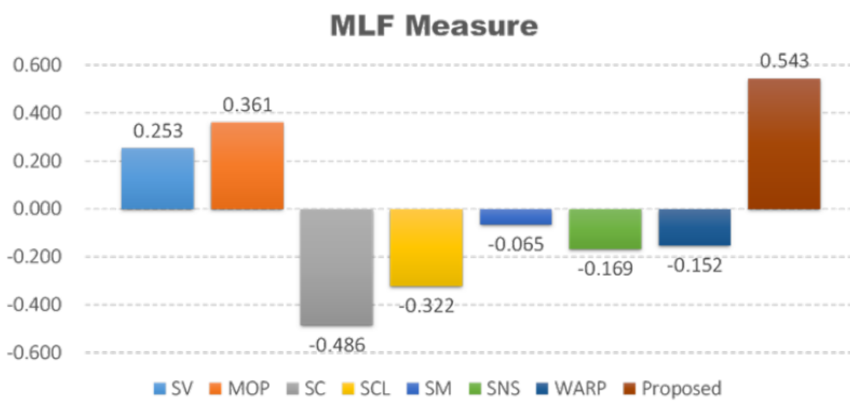


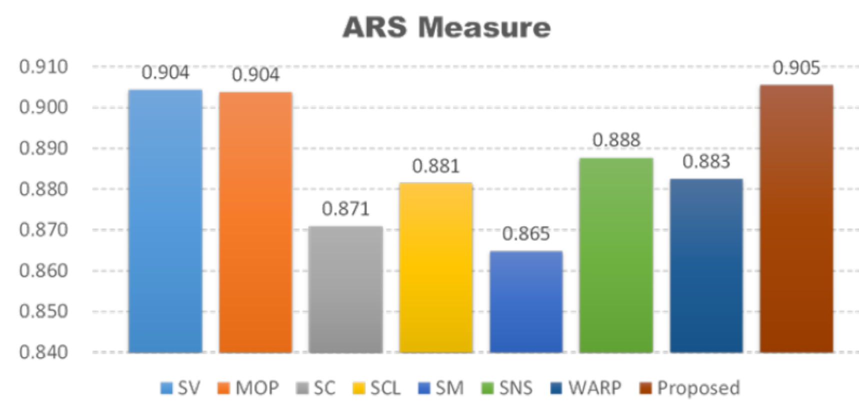
Two-Directional Image Retargeting

- Results

Two-Directional Image Retargeting

- Evaluation





  - MLF Measure
    - Multiple-Level Feature-Based Measure
  - ARS Measure
    - Backward Registration-Based Aspect Ratio Similarity Measure
  - User study / MLF / ARS / BDS

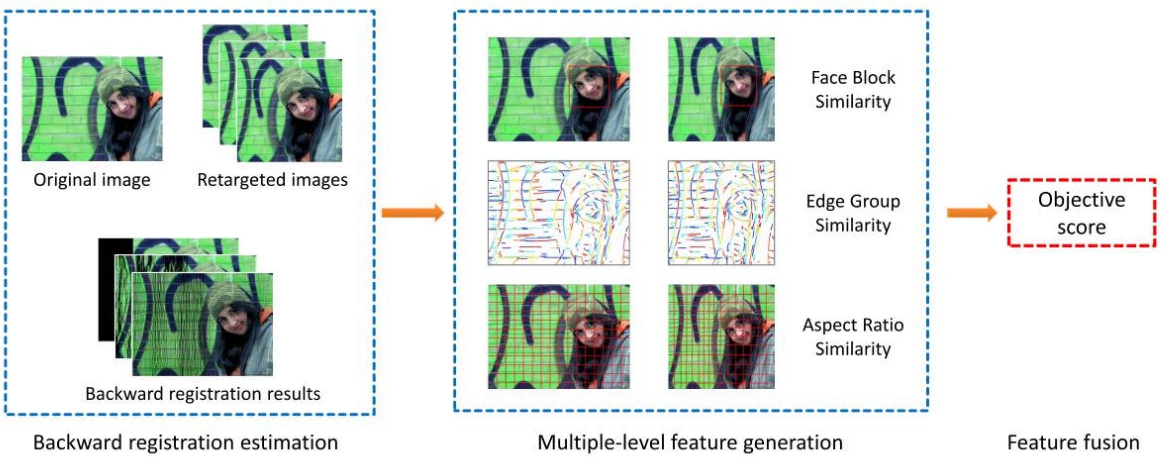
Two-Directional Image Retargeting

- Evaluation
  - MLF[5] Measure (Multiple-Level Feature-Based Measure)

Votes = 27 v.s. 40
ARS = 0.863 v.s. 0.838
Q_{EG} = 0.793 v.s. 0.818
(a)

Votes = 32 v.s. 12
ARS = 0.889 v.s. 0.947
Q_{EG} = 0.891 v.s. 0.836
(b)

Votes = 34 v.s. 11
ARS = 0.938 v.s. 0.956
Q_{EG} = 0.857 v.s. 0.824
(c)



Two-Directional Image Retargeting

- Conclusion
  - Steps
    - Estimate importance map by saliency, depth map, and median filtered input image
    - Two-directional scaling functions derived from importance map with measure of statistical dispersion
    - Generate target image using region-aware guidance map, by two-directional non-uniform scaling
  - Contribution
    - Two-directional non-uniform scaling
    - Region-aware guidance map

References

[1] Tan, Weimin, et al. "Cycle-IR: Deep cyclic image retargeting." *IEEE Transactions on Multimedia* 22.7 (2019): 1730-1743.

[2] Park, Dubok. "Two-directional Image Retargeting with Region-Aware Texture Enhancement." (BMVC, 2021)

[3] R. Droste, J. Jiao, and J. A. Noble. Unified Image and Video Saliency Modeling. In *Proc. ECCV, 2020.*

[4] R. Ranftl, K. Lasinger, D. Hafner, K. Schindler, and V. Koltun. Towards Robust Monocular Depth Estimation: Mixing Datasets for Zero-shot Cross-dataset Transfer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, doi: 10.1109/TPAMI.2020.3019967.

[5] Y. Zhang, W. Lin, Q. Li, W. Cheng, and X. Zhang. Multiple-Level Feature-Based Measure for Retargeted Image Quality. *IEEE Transactions on Image Processing*, vol. 27, issue 1, Jan. 2018.

Thank you