QUALITY EVALUATION FOR IMAGE RETARGETING WITH INSTANCE SEMANTICS

L. Li, Y. Li, J. Wu, L. Ma and Y. Fang

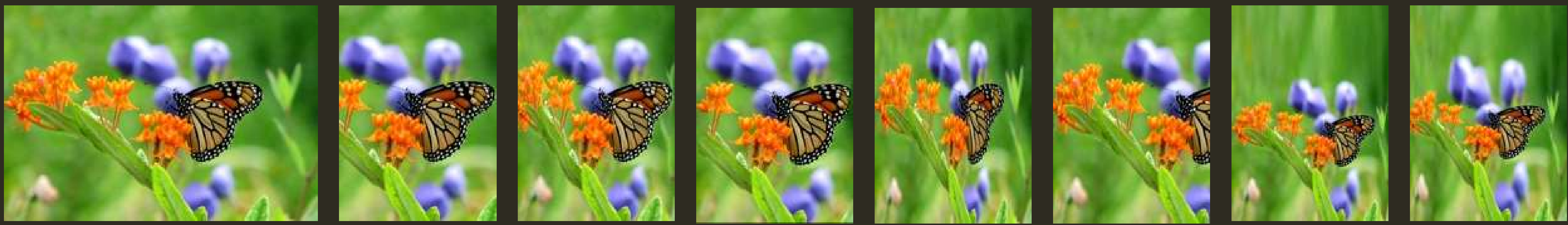
IEEE Transactions on Multimedia,
vol. 23, pp. 2757-2769, 2021

CONTENT

01 INTRODUCTION

INTRODUCTION

- image retargeting approaches
- Image Retargeting Quality Assessment (IRQA)
- INstance SEMantics (INSEM) -> semantic content



(a) Original image (b) CR (c) SC (d) SM (e) SCL (f) WARP (g) SV (h) SNS

INTRODUCTION

The contributions of this work are summarized as follows :

- A new IRQA metric based on instance semantics.
- A top-down pipeline to extract retargeting-aware semantic features to portray the distortions.
- Semantic-based self-adaptive pooling
- We conduct extensive experiments and ablation studies to demonstrate the superiority of the proposed metric over the state-of-the-art methods.

02 PROPOSED IMAGE RETARGETING QUALITY METRIC

PROPOSED IMAGE RETARGETING QUALITY METRIC

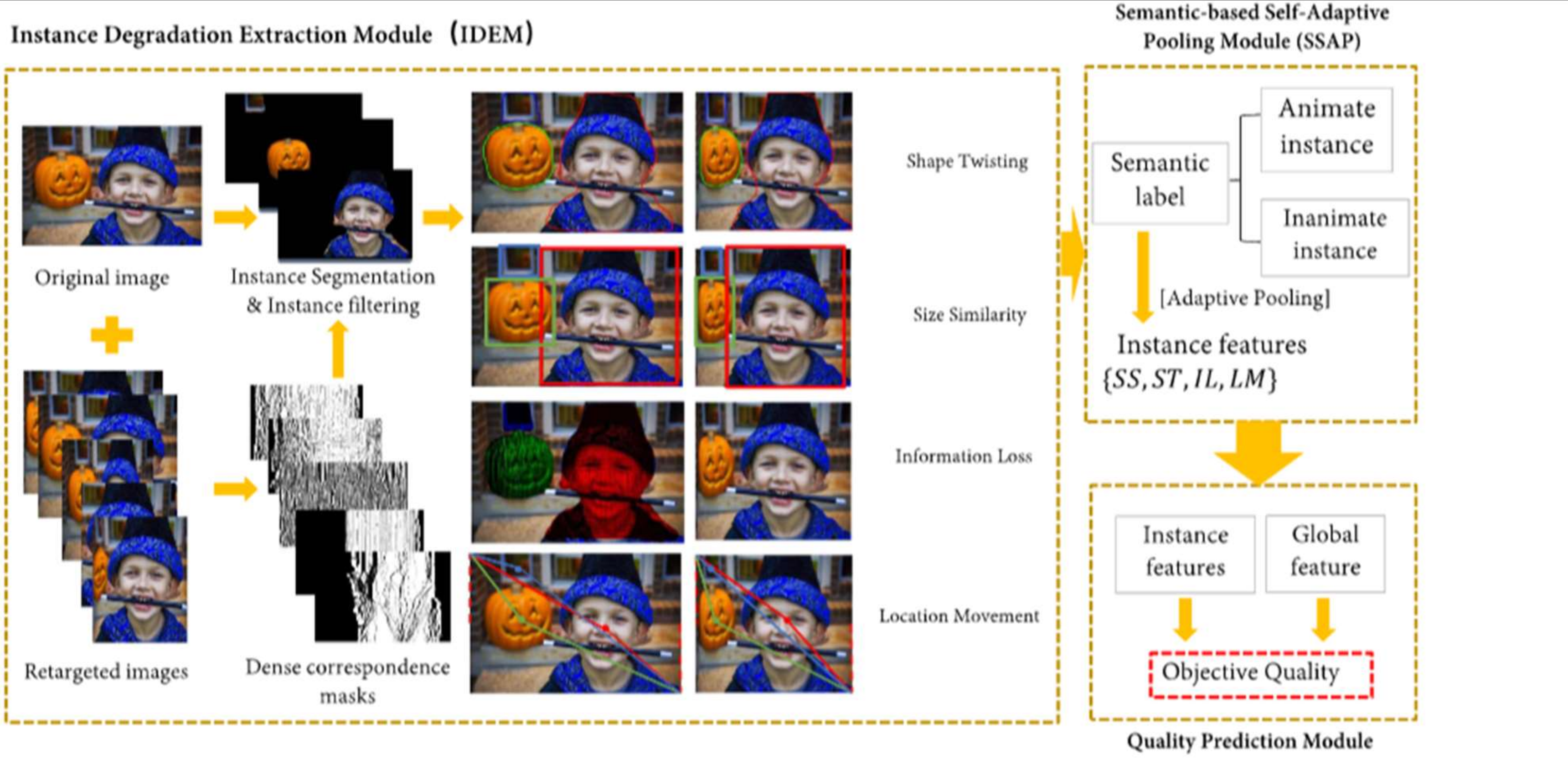


Fig. 1. Diagram of the proposed INSEM metric. INSEM consists of three modules: 1) Instance Degradation Extraction Module (IDEM); 2) Semantics-based Self-Adaptive Pooling (SSAP) module; and 3) quality prediction module.

INSEM

1) The instance quality degradation extraction module (IDEM)
2) The semantic-based self-adaptive pooling (SSAP)
3) The quality prediction module

INSEM

1) The instance quality degradation extraction module (IDEM)
2) The semantic-based self-adaptive pooling (SSAP)
3) The quality prediction module

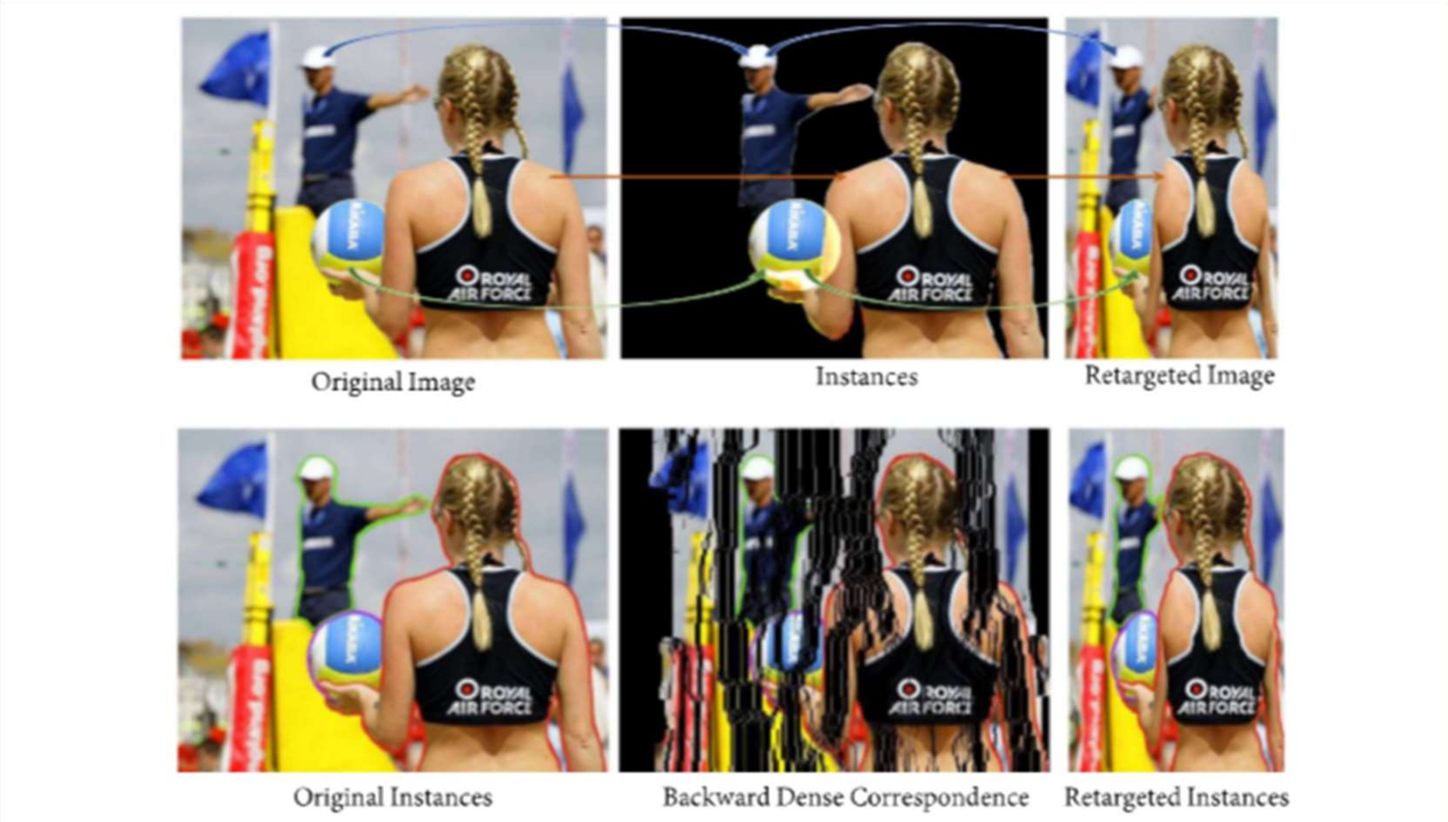
IDEM

1) Instance Detection and Filtering
2) Shape Twisting
3) Size Similarity
4) Information Loss
5) Location Movement

IDEM

1) Instance Detection and Filtering
2) Shape Twisting
3) Size Similarity
4) Information Loss
5) Location Movement

DISTORTION ANALYSIS IN IMAGE RETARGETING



IDEM | SSAP | quality prediction

INSTANCE DETECTION AND FILTERING

- instance segmentation
  - mask R-CNN model
  - binary instance mask, bounding box, and semantic label.
- multiple instances : remove the less important instances by defining the following instance saliency measure:

$$\mathcal{IS} = \frac{\sum_{(i,j)\in \mathrm{INS}} \mathbf{SM}(i,j)}{\sum_{(i,j)\in \mathbf{SM}} \mathbf{SM}(i,j)}, \qquad (1)$$

- instance filtering operation : IS ≥ τ ? (τ=0.25)

IDEM

1) Instance Detection and Filtering
2) Shape Twisting
3) Size Similarity
4) Information Loss
5) Location Movement

Fig. 3. Local and global shape twisting in image retargeting. Top row shows the original image, retargeted image and several shape twisting areas. Bottom row shows the original and retargeted shape contours of one instance 'person,' together with the Chamfer distance map ($\mathbf{DM}$).

SHAPE TWISTING (ST)

$$\mathcal{ST} = \sqrt{\frac{\sum_{i=1}^{N}\left(\left(1 + e^{-\alpha\cdot(LSDC_i - 1)}\right)\cdot \mathbf{DM}(p_i')\right)^2}{N}}, \quad (8)$$

- where α is a coefficient applied to control the relative contribution of local shape twisting. A large ST value indicates that the shape twisting is severe.

IDEM

1) Instance Detection and Filtering
2) Shape Twisting
3) Size Similarity
4) Information Loss
5) Location Movement

SIZE SIMILARITY (SS)

- The size change of an instance is decided by two respects in this paper:
  1) aspect ratio : relative size change
  2) Scale : absolute size change
- The size similarity (SS) of instances is defined as

$$SS = \left(e^{-\beta \cdot \left(\frac{r_w + r_h}{2} \right)^2} \right) \cdot \frac{2 \cdot r_w \cdot r_h + c_0}{r_w^2 + r_w^2 + c_0},$$

IDEM

1) Instance Detection and Filtering
2) Shape Twisting
3) Size Similarity
4) Information Loss
5) Location Movement

INFORMATION LOSS (IL)

- superpixel segmentation
- simple linear iterative clustering (SLIC)
  - segment size = 256
  - compactness index = 20

$$\mu = \frac{1}{n} \sum_{(i,j)\in \mathbf{S}} \mathbf{S}(i,j),$$

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{(i,j)\in \mathbf{S}} (\mathbf{S}(i,j) - \mu)^2}.$$

$$\mathbf{INF} = \{\mu_1, \sigma_1, n_1, \mu_2, \sigma_2, n_2, \ldots, \mu_{N_S}, \sigma_{N_S}, n_{N_S}\}.$$

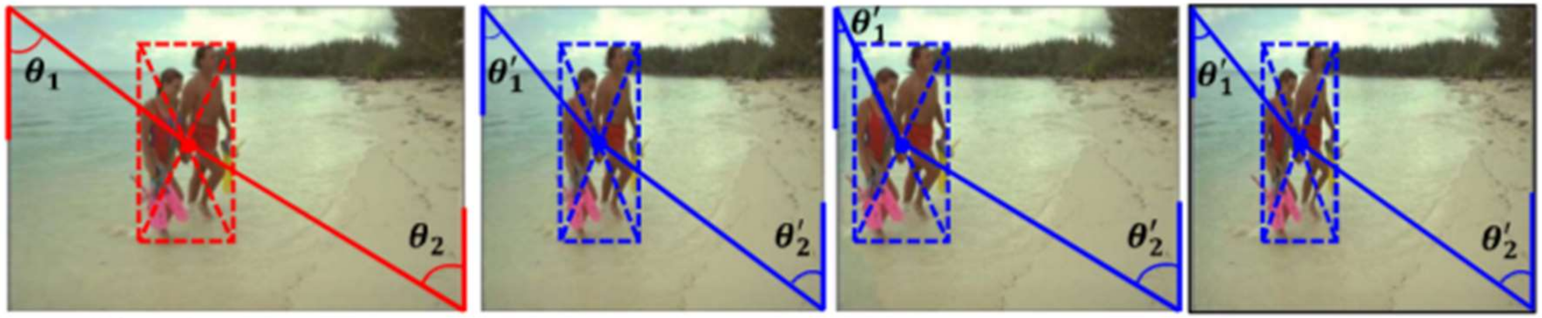
$$\mathcal{IL} = \frac{2 \cdot \mathbf{INF} \cdot \mathbf{INF}' + c_0}{\mathbf{INF}^2 + \mathbf{INF}'^2 + c_0}.$$

IDEM

1) Instance Detection and Filtering
2) Shape Twisting
3) Size Similarity
4) Information Loss
5) Location Movement

LOCATION MOVEMENT (LM)

- superpixel segmentation
- simple linear iterative clustering (SLIC)
  - segment size = 256
  - compactness index = 20



(a) Reference image (b) WARP (c) SM (d) SC

	(b) WARP		(c) SM		(d) SC
MOS:	60	*vs.*	**18**	*vs.*	48
ST feature:	6.605	*vs.*	**0**	*vs.*	8.186
LM feature:	0.942	*vs.*	**0.478**	*vs.*	0.886

Fig. 4. An exemplar illustration of location movement in the perception of retargeting quality. The MOS, ST, LM feature values are given for comparison. Even image (c) has no shape twisting ($\mathcal{ST} = 0$), the significant location change leads to the worst quality (MOS $= 18$).

LOCATION MOVEMENT (LM)

- the relative angle changes are computed as

$$r_{LM_1} = \frac{\tan(\theta'_1)}{\tan(\theta_1)}, \quad r_{LM_2} = \frac{\tan(\theta'_2)}{\tan(\theta_2)}.$$

- Then, the location movement feature is defined as

$$\mathcal{LM} = \frac{2 \cdot r_{LM_1} \cdot r_{LM_2} + c_0}{r_{LM_1}^2 + r_{LM_2}^2 + c_0}.$$

- When the LM score is close to 1, the location movement of an instance is small, and the retargeted image is expected to have relatively high quality.

INSEM

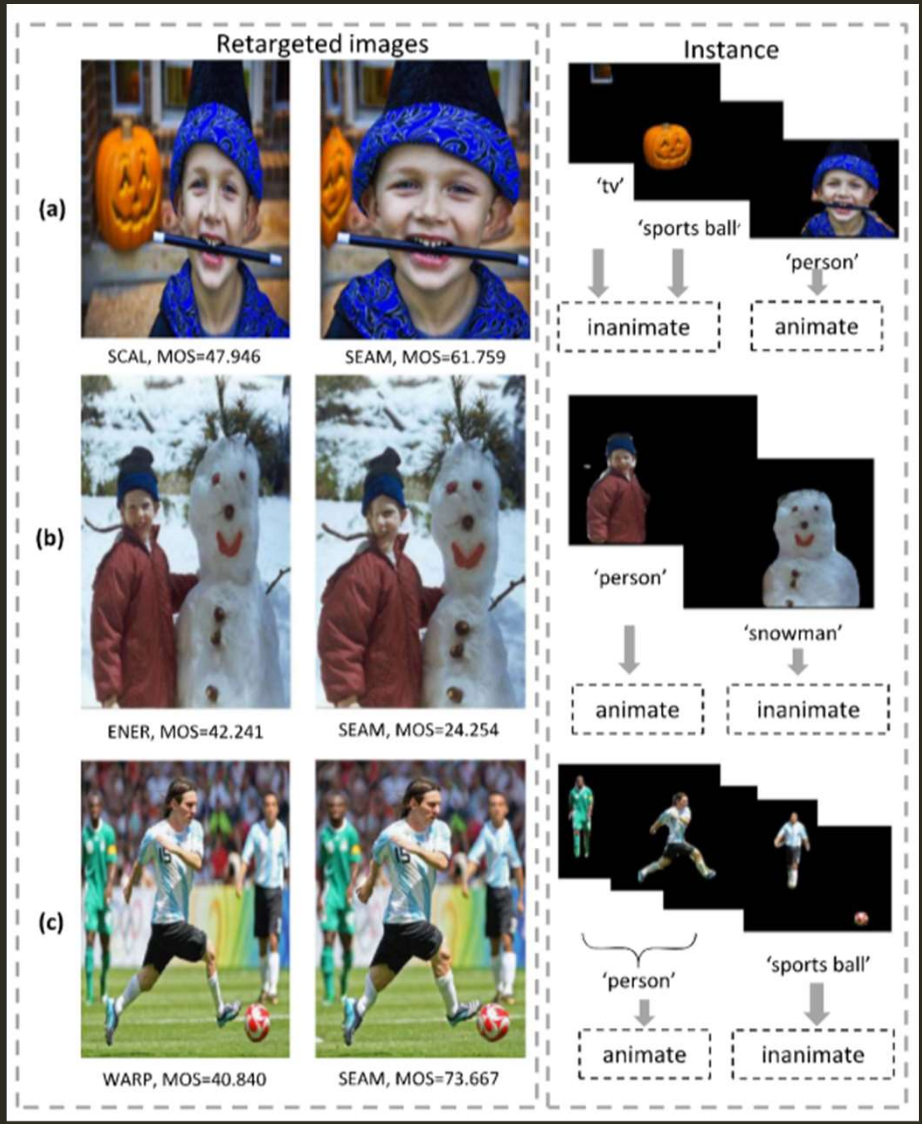
1) The instance quality degradation extraction module (IDEM)

2) The semantic-based self-adaptive pooling (SSAP)

3) The quality prediction module

SSAP

- different saliency preferences
- different biological natures
  - animate instance
  - inanimate instances



SSAP

- the semantic-aware weight of the k-th instance is defined as

$$w_k = \begin{cases} \Omega_{\mathrm{A}} \cdot \frac{\mathcal{IS}_k}{\sum_{i=1}^{N_{\mathrm{INS}}} \mathcal{IS}_i}, & \text{for animate instance,} \\ \Omega_{\mathrm{INA}} \cdot \frac{\mathcal{IS}_k}{\sum_{i=1}^{N_{\mathrm{INS}}} \mathcal{IS}_i}, & \text{for inanimate instance,} \end{cases}$$

- instance saliency

$$\mathcal{IS} = \frac{\sum_{(i,j)\in \mathrm{INS}} \mathbf{SM}(i,j)}{\sum_{(i,j)\in \mathbf{SM}} \mathbf{SM}(i,j)}, \quad (1)$$

SSAP

- Only one instance category detected, regardless of animate or inanimate, the semantic category weight is not calculated, and in this case

$$w_k = \frac{\mathcal{IS}_k}{\sum_{i=1}^{N_{\text{INS}}} \mathcal{IS}_i}.$$

- The overall instance-level feature vector for the whole image as

$$\mathbf{F} = \sum_{k=1}^{N_{\text{INS}}} w_k \cdot \mathbf{F}_{\text{INS}}^k,$$

$$\mathbf{F}_{\text{INS}} = \{\mathcal{ST}, \mathcal{SS}, \mathcal{IL}, \mathcal{LM}\}.$$

INSEM

1) The instance quality degradation extraction module (IDEM)

2) The semantic-based self-adaptive pooling (SSAP)

3) The quality prediction module

03 EXPERIMENTS

DATABASES AND PROTOCOLS

1) MIT RetargetMe database
2) train the quality model
3) NRID database

PARAMETER SELECTION

1) local shape twisting coefficient α

$$\mathcal{ST} = \sqrt{\frac{\sum_{i=1}^{N} \left(\left(1 + e^{-\alpha \cdot (LSDC_i - 1)}\right) \cdot \mathbf{DM}(p_i')\right)^2}{N}},$$

2) absolute size coefficient β

$$\mathcal{SS} = \left(e^{-\beta \cdot \left(\frac{r_w + r_h}{2}\right)^2}\right) \cdot \frac{2 \cdot r_w \cdot r_h + c_0}{r_w^2 + r_w^2 + c_0},$$

3) Ω_A in SSAP

4) Animate instances have a greater impact on perceived quality than inanimate instances.

PARAMETER SELECTION

1) local shape twisting coefficient α

$$\mathcal{ST} = \sqrt{\frac{\sum_{i=1}^{N}\left(\left(1+e^{-\alpha\cdot(LSDC_i-1)}\right)\cdot \mathbf{DM}(p_i')\right)^2}{N}},$$

2) absolute size coefficient β

$$\mathcal{SS} = \left(e^{-\beta\cdot\left(\frac{r_w+r_h}{2}\right)^2}\right)\cdot \frac{2\cdot r_w\cdot r_h + c_0}{r_w^2 + r_w^2 + c_0},$$

3) Ω_A in SSAP

4) Animate instances have a greater impact on perceived quality than inanimate instances.

PERFORMANCE EVALUATION

1) Performance Comparison on the CUHK Database

Metric	PLCC	SRCC	RMSE	OR
SIFT-flow [37]	0.314	0.290	12.817	0.146
BDS [38]	0.290	0.289	12.922	0.216
EMD [40]	0.277	0.290	12.977	0.170
IRQA [14]	0.437	0.466	12.141	0.152
IRSSIM [16]	0.230	0.240	13.140	0.176
Liang [21]	0.443	0.467	12.105	0.181
Ma [26]	0.537	0.493	-	0.193
Jiang [23]	0.644	0.616	10.763	-
ARS [20]	0.684	0.669	9.855	0.070
DEEP [25]	0.701	0.673	8.364	0.057
Proposed INSEM	**0.798**	**0.748**	**7.905**	**0.023**

PERFORMANCE EVALUATION

2) Performance Comparison on the MIT RetargetMe Database With Labeled Attributes

Metric	Mean KRCC for each subset						Total	
	Lines Edges	*Faces People*	*Foreground Objects*	*Texture*	*Geometric Structure*	*Symmetry*	mean KRCC	std KRCC
SIFT-flow [37]	0.097	0.252	0.218	0.161	0.085	0.071	0.145	0.262
BDS [38]	0.040	0.190	0.167	0.060	-0.004	-0.012	0.083	0.268
EMD [40]	0.220	0.262	0.226	0.205	0.237	0.500	0.251	0.272
IRQA [14]	0.097	0.290	0.293	0.161	0.053	0.150	0.164	0.263
IRSSIM [16]	0.309	0.452	0.377	0.321	0.313	0.333	0.363	0.271
Liang [21]	0.351	0.271	0.304	0.381	0.415	**0.548**	0.399	-
Jiang [23]	-	-	-	-	-	-	0.413	0.282
ARS [20]	0.463	0.519	0.444	0.330	0.505	0.464	0.452	0.283
DEEP [25]	0.466	0.512	0.452	0.434	0.515	0.443	0.476	0.243
Ma [26]	0.229	0.273	0.182	0.218	0.252	0.484	0.477	-
Proposed INSEM	**0.586**	**0.562**	**0.552**	**0.607**	**0.594**	0.469	**0.537**	**0.188**

PERFORMANCE EVALUATION

3) Performance Comparison on the NRID Database With Labeled Attributes

Metric	Mean KRCC for each subset						Total	
	Line Edges	*Faces People*	*Foreground Objects*	*Texture*	*Geometric Structure*	*Symmetry*	mean KRCC	std KRCC
SIFT-flow [37]	-0.013	-0.040	0.090	-0.017	-0.025	0.267	-0.010	0.500
BDS [38]	-	-	-	-	-	-	0.131	0.527
EMD [40]	0.213	0.480	0.375	0.266	0.400	0.133	0.361	0.362
IRQA [14]	0.093	0.240	0.013	0.050	0.025	0.233	0.154	0.512
IRSSIM [16]	0.347	0.440	0.313	0.267	0.200	0.333	0.383	0.554
Jiang [23]	-	-	-	-	-	-	0.577	**0.334**
ARS [20]	0.373	0.667	0.467	0.330	0.475	**0.600**	0.514	0.398
DEEP [25]	-	-	-	-	-	-	0.598	0.412
Proposed INSEM	**0.667**	**0.673**	**0.788**	**0.467**	**0.475**	0.533	**0.640**	0.433

ABLATION STUDY

- Instance Features

  vs.

  Global Feature

Index	Feature combination				Database						
					CUHK			MIT RetargetMe		NRID	
	ST	SS	IL	LM	PLCC	SRCC	RMSE	mean KRCC	std KRCC	mean KRCC	std KRCC
1	√				0.473	0.504	11.239	0.308	0.248	0.354	0.524
2		√			0.729	0.688	9.313	0.253	0.297	0.354	0.490
3			√		0.692	0.643	9.799	0.268	0.283	0.314	0.532
4				√	0.128	0.028	13.466	0.106	0.299	0.286	0.347
5	√	√			0.749	0.713	8.669	0.324	0.213	0.491	0.292
6	√		√		0.752	0.722	8.685	0.276	0.141	0.514	0.330
7	√			√	0.755	0.722	8.608	0.251	0.191	0.257	0.225
8		√	√		0.755	0.719	8.604	0.270	0.179	0.497	0.320
9		√		√	0.753	0.718	8.710	0.249	0.192	0.343	0.245
10			√	√	0.753	0.719	8.831	0.247	0.191	0.342	0.250
11	√	√	√		0.789	0.735	7.878	0.330	0.202	0.503	0.289
12	√	√		√	0.795	0.745	7.783	0.342	0.221	0.469	0.310
13	√		√	√	0.799	0.745	7.792	0.380	0.186	0.423	0.306
14		√	√	√	0.791	0.740	7.913	0.328	0.199	0.503	0.293
15	√	√	√	√	0.784	0.730	7.946	0.510	0.221	0.491	0.288
16	Global feature				0.696	0.686	9.757	0.435	0.233	0.429	0.385
17	ST+SS+IL+LM+Global feature				0.788	0.726	7.962	0.521	0.194	0.584	0.288

ABLATION STUDY

- SSAP

  vs.

  Average Pooling

Database	Criterion	Average Pooling	Proposed SSAP
CUHK	PLCC	0.771	**0.798**
	SRCC	0.741	**0.748**
	RMSE	8.069	**7.905**
	OR	0.047	**0.023**
MIT RetargetMe	mean KRCC	0.421	**0.537**
	std KRCC	0.233	**0.188**
NRID	mean KRCC	0.560	**0.640**
	std KRCC	**0.302**	0.433

INSTANCE FEATURE AS UNIVERSAL IRQA MODULE

- Promoting Effect of the Proposed IDEM Module on Global-feature-based IRQA Metrics on CUHK/MIT/NRID Databases

Global Metric	CUHK				MIT RetargetMe		NRID	
	PLCC		SRCC		mean KRCC		mean KRCC	
SIFT-flow [37]	0.314	**0.399**(+27.1%)	0.291	**0.425**(+46.6%)	0.145	**0.255**(+75.9%)	-0.010	**0.173**(-)
EMD [40]	0.277	**0.402**(+45.1%)	0.290	**0.479**(+65.2%)	0.251	**0.397**(+58.2%)	0.361	**0.499**(+38.2%)
BDS [38]	0.290	**0.439**(+51.4%)	0.291	**0.422**(+45.5%)	0.083	**0.273**(+228.9%)	0.131	**0.336**(+156.5%)
IRQA [14]	0.437	**0.711**(+62.7%)	0.466	**0.754**(+61.8%)	0.164	**0.220**(+34.2%)	0.154	**0.189**(+22.7%)
IRSSIM [16]	0.230	**0.736**(+220.6%)	0.240	**0.689**(+187.1%)	0.363	**0.460**(+26.7%)	0.383	**0.488**(+27.4%)

04 CONCLUSION

CONCLUSION

- A novel image retargeting quality assessment metric based on instance semantics.
- Four kinds of instance-level semantic
- Animate and inanimate
- Semantics-based self-adaptive pooling strategy(SSAP)
- Performed extensive comparisons with state-of-the-art IRQA metrics.
- Both intradatabase and cross-database settings

Q & A