SHANGHAI JIAO TONG UNIVERSITY

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Image Cropping with Spatial-aware Feature and Rank Consistency

Chao Wang   Li Niu*   Bo Zhang   Liqing Zhang*

Shanghai Jiao Tong University

WED-AM-174

*Corresponding author

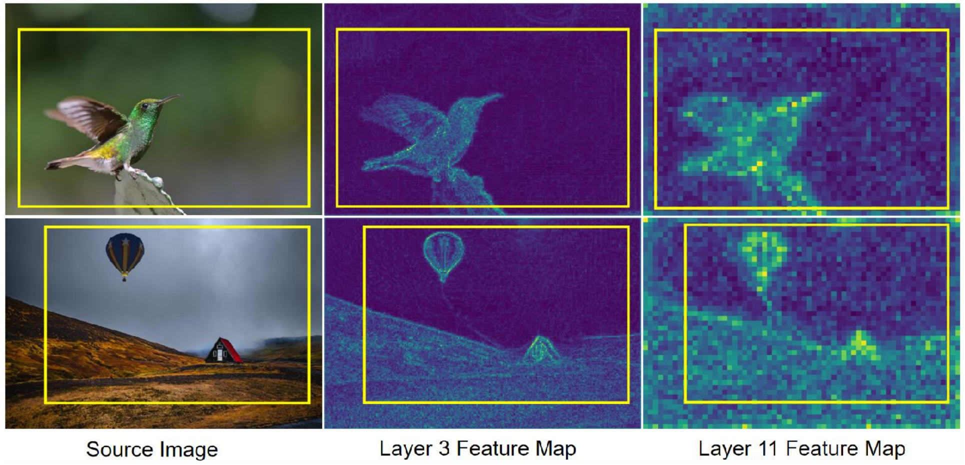
Quick Preview

- Image cropping aims to find visually appealing crops in an image.

- Drawbacks of previous methods.
  - They are weak in capturing the spatial relationship between crops and aesthetic elements.
  - The potential of unlabeled data awaits to be excavated.

Quick Preview

- We propose spatial-aware feature to encode the spatial relationship between candidate crops and aesthetic elements.

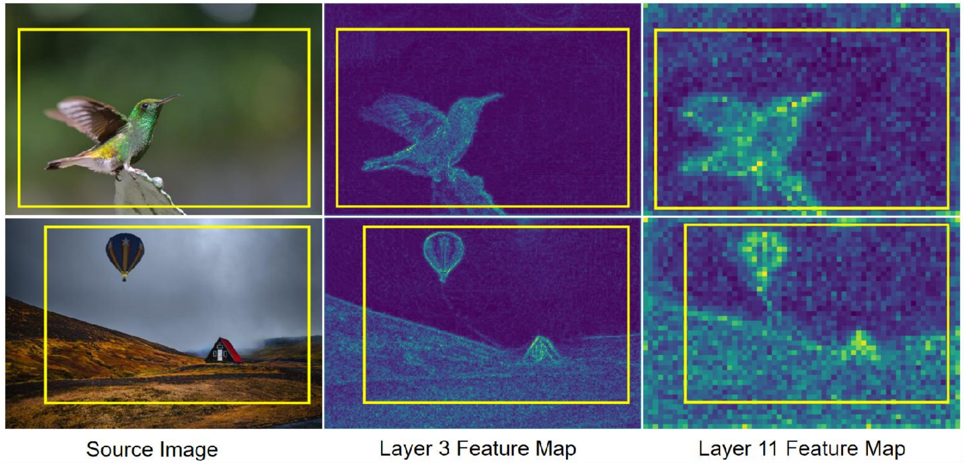


- We train a pair-wise ranking classifier on labeled images and transfer the ranking knowledge to unlabeled images to enforce rank consistency.

- Experimental results on the benchmark datasets show that our proposed method performs favorably against state-of-the-art methods.

Motivation of Spatial-aware Feature

- The spatial relationship between crops and aesthetic elements (e.g., salient objects, semantic edges) is very critical for image cropping.
- The crop should enclose the salient object, or should not cut through the semantic edges.



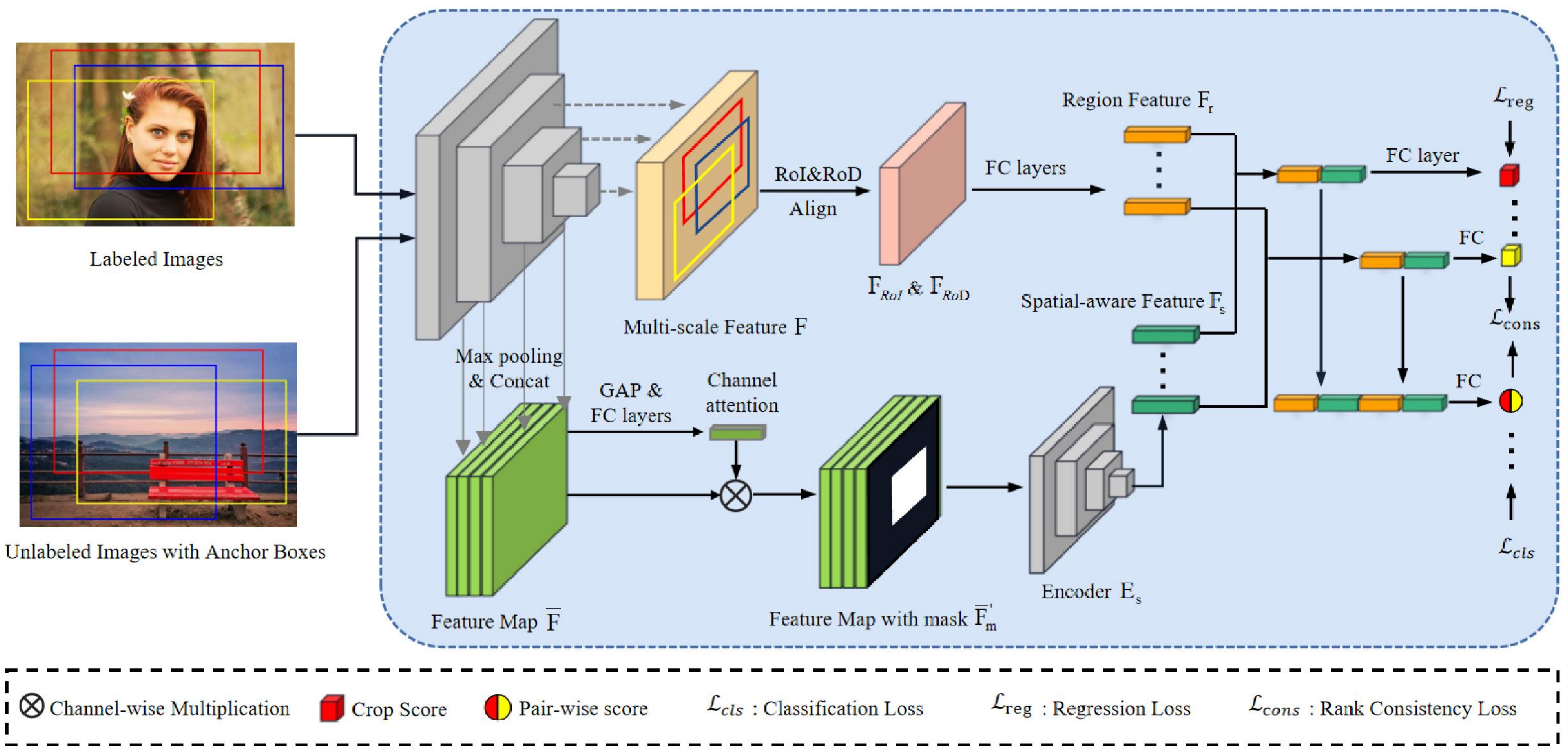
The feature map obtained using channel-wise max pooling can emphasize some aesthetic elements. The low-level feature maps emphasize semantic edges and the high-level feature maps emphasize salient objects.

Motivation of Rank Consistency

- The cost of crop annotation is very high.
- The rank of candidate crops should be consistent between labeled data and unlabeled data.
- We expect that the knowledge of comparing the aesthetic quality of two crops with similar content could be transferred to unlabeled data.
- Semi-supervised/Transductive Learning paradigms.

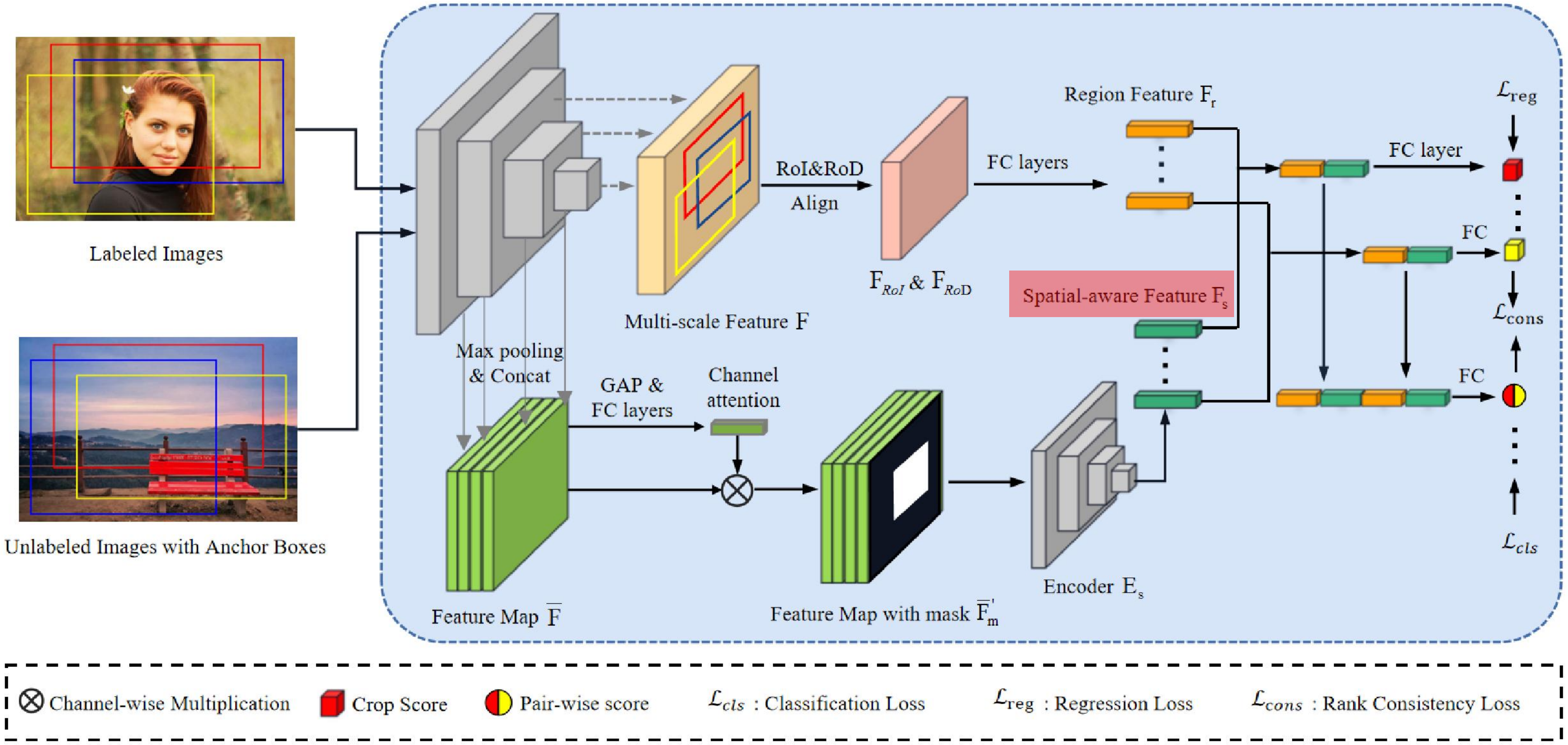
Proposed Method

- Image Cropping with Spatial-aware Feature and Rank Consistency



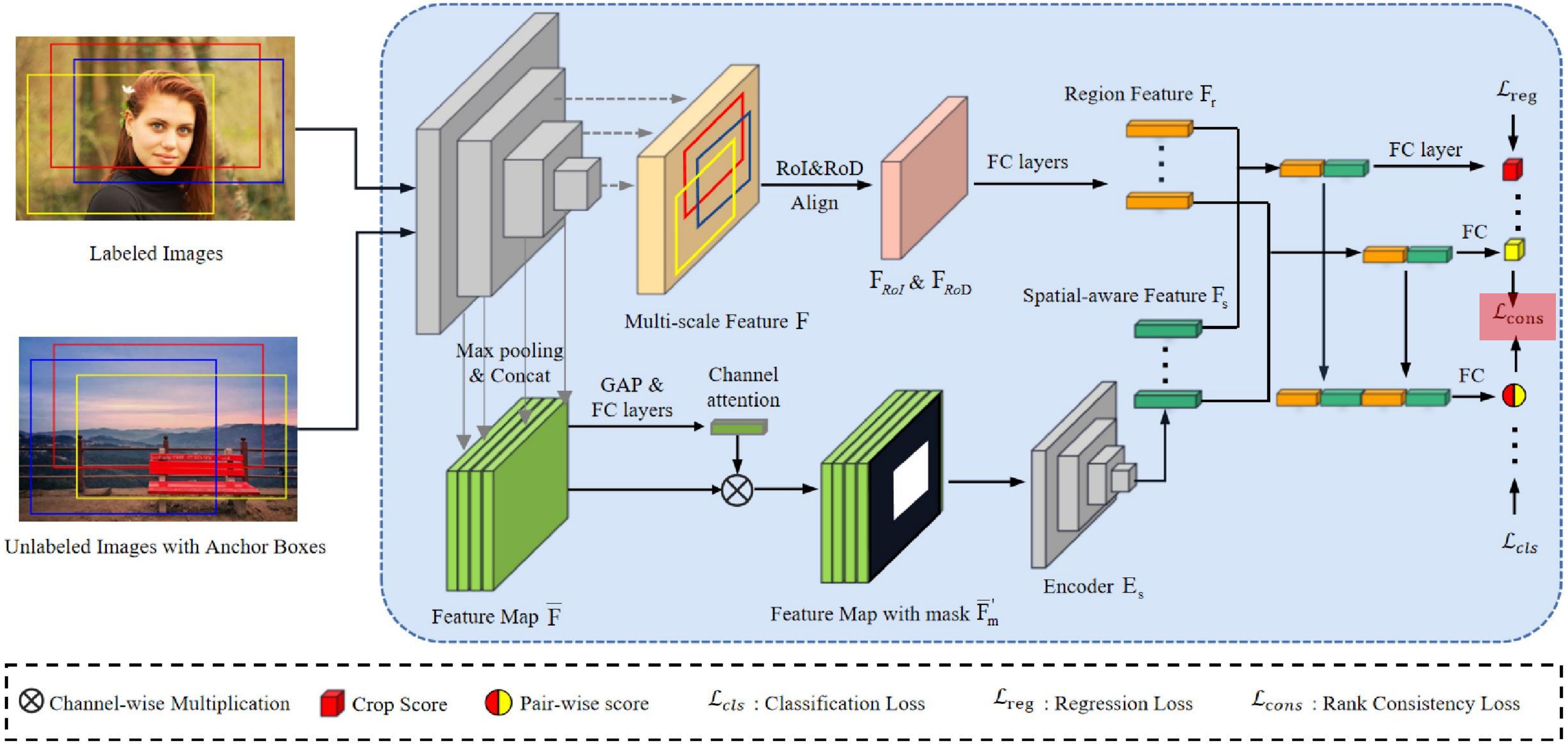
Proposed Method

- Spatial-aware Feature
  - Feature Maps Activation.
  - Channel Attention Block.
  - Spatial Relationship Modeling



Proposed Method

- Rank Consistency
  - Pair-wise Ranking Classifier.



Proposed Method

- Training with a multi-task loss function in an end-to-end manner

$$\mathcal{L}_{total} = \mathcal{L}_{labeled} + \mathcal{L}_{cons}$$

$$\mathcal{L}_{labeled} = \mathcal{L}_{reg} + \lambda_{cls}\mathcal{L}_{cls}$$

Regression Loss

$$\mathcal{L}_{reg} = \frac{1}{N}\sum_{i}^{N}\mathcal{L}_{s1}(y_i - \hat{y}_i)$$

Classification Loss

$$\mathcal{L}_{cls} = \frac{1}{P}\sum_{n=1}^{P} -q_n \cdot logp_n - (1-q_n)\cdot log(1-p_n)$$

Consistency Loss

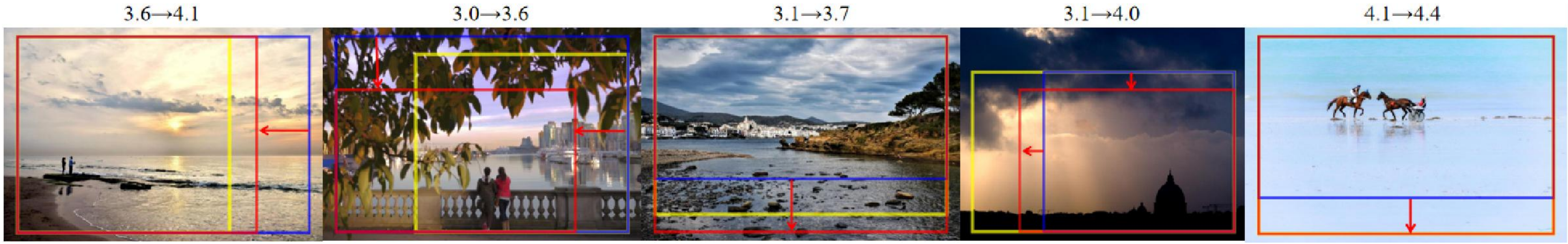
$$\mathcal{L}_{cons} = \frac{2}{(N^2 - N)}\sum_{i=1}^{N}\sum_{j=i+1}^{N} l(C_i, C_j)$$

$$l\left(C_i, C_j\right) = \max\{0, \delta + sign(p_n - 0.5)(\hat{y}_j - \hat{y}_i)\}$$

Proposed Method

➢ Spatial-aware Feature

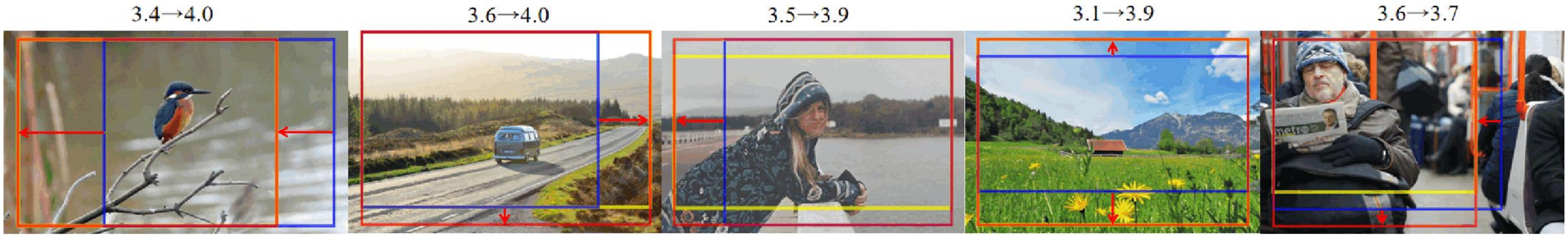
- Spatial-aware feature helps locate the crop better.



Predicted scores are lift by Spatial-aware feature.The annotated best crops are in yellow, the predicted best crops by the basic model and our proposed method are in blue and red respectively. The numbers above the images are their predicted scores.

Proposed Method

- Rank Consistency
  - Rank consistency helps rank candidate crops more accurately.



Predicted scores are lift by Rank consistency.The annotated best crops are in yellow, the predicted best crops by the basic model and our proposed method are in blue and red respectively. The numbers above the images are their predicted scores.

Experiments

- Quantitative comparison on GIACD dataset.
  - PCC evaluates the linear correlation.
  - SRCC measures the ranking order correlation.
  - Acc5/Acc10 measures the ability to return the best crops.

Model	$Acc_{1/5}$	$Acc_{2/5}$	$Acc_{3/5}$	$Acc_{4/5}$	$\overline{Acc_5}$	$Acc_{1/10}$	$Acc_{2/10}$	$Acc_{3/10}$	$Acc_{4/10}$	$\overline{Acc_{10}}$	$\overline{SRCC}$	$\overline{PCC}$
A2RL [21]	23.2	-	-	-	-	39.5	-	-	-	-	-	-
VPN [52]	36.0	-	-	-	-	48.5	-	-	-	-	-	-
VFN [5]	26.6	26.5	26.7	25.7	26.4	40.6	40.2	40.3	39.3	40.1	0.485	0.503
VEN [52]	37.5	35.0	35.3	34.2	35.5	50.5	49.2	48.4	46.4	48.6	0.616	0.662
GAIC [57]	68.2	64.3	61.3	58.5	63.1	84.4	82.7	80.7	78.7	81.6	0.849	0.874
CGS [23]	63.0	62.3	58.8	54.9	59.7	81.5	79.5	77.0	73.3	77.8	0.795	-
CGS* [23]	66.2	63.0	59.6	56.5	61.3	84.4	81.4	78.9	76.9	80.4	0.850	0.874
TransView [36]	69.0	**66.9**	61.9	57.8	63.9	85.4	84.1	81.3	78.6	82.4	0.857	0.880
Ours (w/o te)	68.4	65.1	62.1	59.2	63.7	86.2	83.1	81.4	79.5	82.6	0.865	0.889
Ours	**70.0**	**66.9**	**62.5**	**59.8**	**64.8**	**86.8**	**84.5**	**82.9**	**79.8**	**83.3**	**0.872**	**0.893**

Experiments

➢ Quantitative comparison on FCDB dataset.

- IoU and Disp measure the overlap and offset degree.

Method	Training Set	IoU ↑	Disp ↓
A2RL [21]	AVA	0.663	0.089
A3RL [22]	AVA	0.696	0.077
VPN [52]	CPC	0.711	0.073
VEN [52]	CPC	0.735	0.072
ASM [46]	CPC	0.749	0.068
GAIC [57]	GAICD	0.672	0.084
CGS [23]	GAICD	0.685	0.079
TransView [36]	GAICD	0.682	0.080
Ours (w/o te)	GAICD	0.686	0.078
Ours	GAICD	0.695	0.075

Experiments

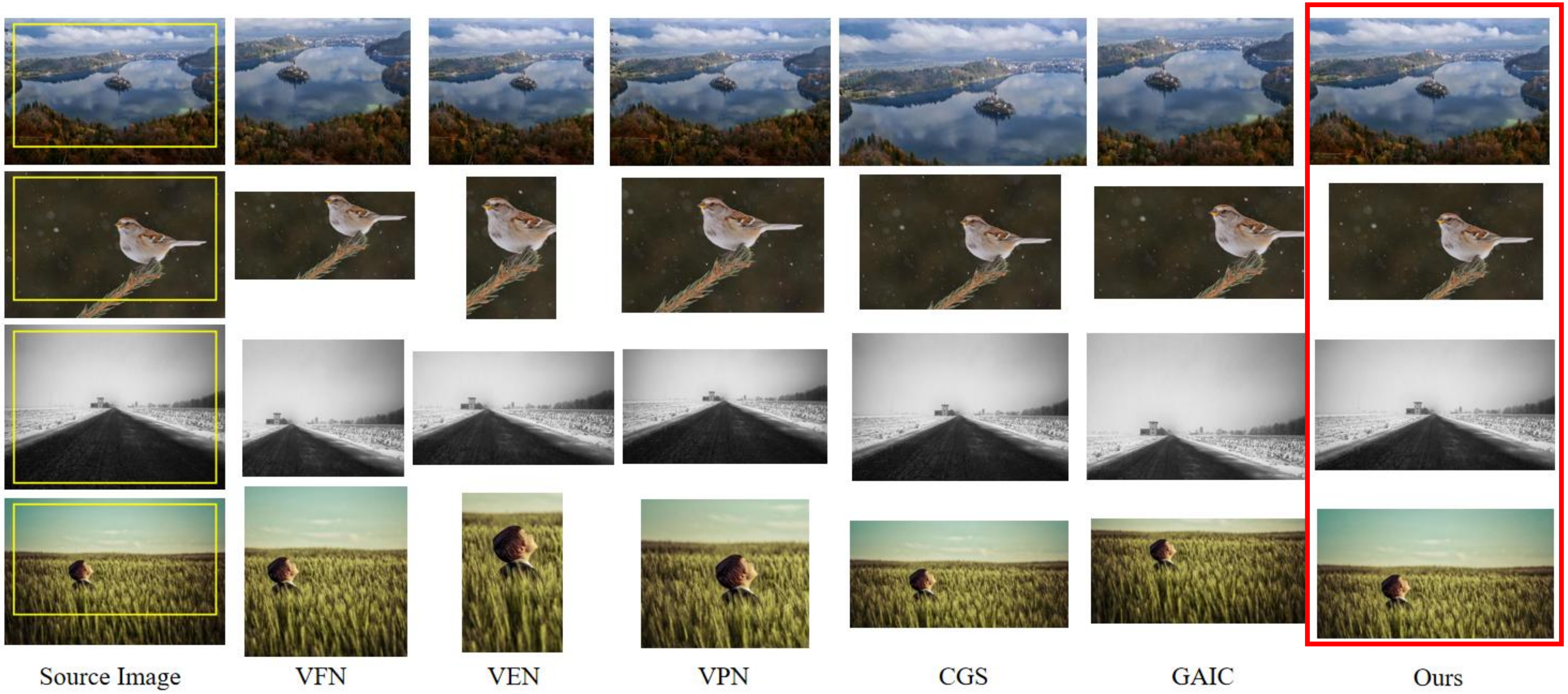
- Model complexity and runtime.
  - Our model is lighted-weighted and efficient for mobile device application.

Method	Backbone	#Parameters	Runtime
VFN	Alexnet	14.88M	2491ms
VEN	VGG16	40.93M	5331ms
VPN	VGG16	65.31M	149ms
CGS	VGG16	21.25M	31ms
GAIC	MobileNetv2	5.91M	24ms
Ours(basic)	MobileNetv2	5.91M	25ms
Ours	MobileNetv2	7.10M	32ms

All models are run on the PC with Intel(R) Core(TM) i7-9700K CPU and one single NVIDIA GTX 1080Ti GPU.

Experiments

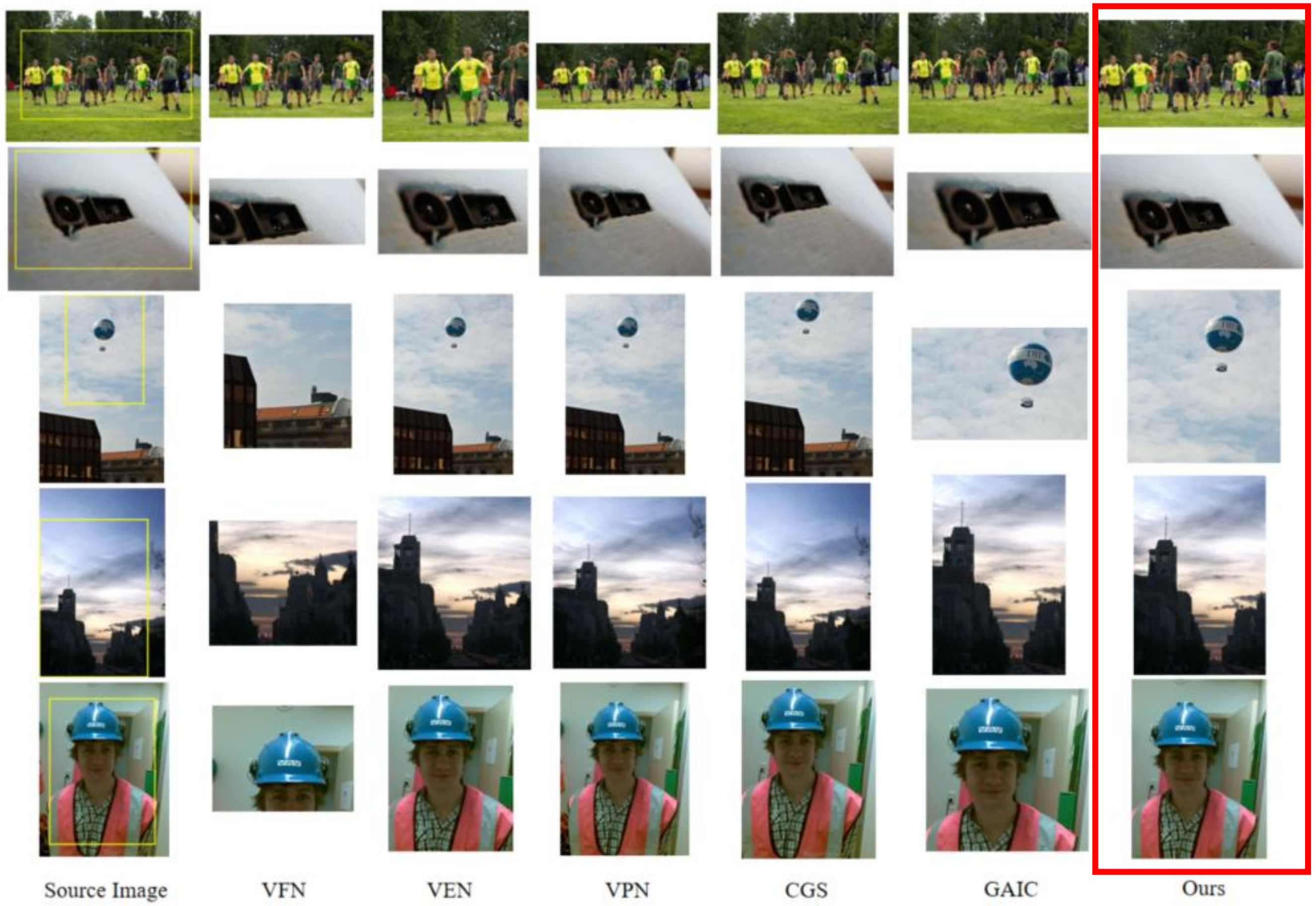
➢ Qualitative comparison on GAICD dataset.



The annotated best crop (yellow bounding box) in the source image is in the left column and top-1 crops obtained by different methods are in the rest of the columns.

Experiments

- Qualitative comparison on FCDB dataset.



The annotated best crop (yellow bounding box) in the source image is in the left column and top-1 crops obtained by different methods are in the rest of the columns.

Contributions

- Spatial-aware Feature: capture the spatial relationship between candidate crops and aesthetic elements.
- Rank Consistency: transfer ranking knowledge from labeled images to unlabeled images.
- Quantitative and qualitative comparisons have shown that our method obtains the state-of-the-art performance on benchmark datasets.

SHANGHAI JIAO TONG UNIVERSITY

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Thanks for watching !

Image Cropping with Spatial-aware Feature and Rank Consistency

Chao Wang   Li Niu*   Bo Zhang   Liqing Zhang*

Shanghai Jiao Tong University

WED-AM-174

*Corresponding author