Compositing-aware Image Search — Supplementary Material

Hengshuang Zhao[1⋆], Xiaohui Shen[2], Zhe Lin[3],
Kalyan Sunkavalli[3], Brian Price[3], Jiaya Jia[1,4]

[1]The Chinese University of Hong Kong, [2]ByteDance AI Lab,
[3]Adobe Research, [4] Tencent Youtu Lab
{hszhao,leojia}@cse.cuhk.edu.hk, shenxiaohui@bytedance.com,
{zlin,sunkaval,bprice}@adobe.com

1 Dataset Statistics

Training Set The training set statistics are shown in Table 1. Some datasets do not have annotations for certain categories, i.e., there are no 'dog' from ADE20K [1] and no 'painting' from MS-COCO [2] and PASCAL VOC 2012 [3]. Also, MS-COCO has already contained sufficient 'person' instance masks, and therefore we exclusively use the 'person' instances from MS-COCO for training.

Evaluation Set Each category has 10 background images with various scenes downloaded from Flickr[1]. And for candidate foreground objects, we utilize object instance masks from validation set of MS-COCO [2], PASCAL VOC 2012 [3] and ADE20K [1]. The statistics are shown in Table 2.

Table 1: Training set statistics. Number of training images of each category in different datasets. '-' stands for no images are chosen from the related dataset.

Category	MS-COCO	VOC 2012	ADE20K	All
boat	2527	475	229	3231
bottle	3385	511	193	4089
car	7002	1095	3382	11479
chair	10883	1530	6708	19121
dog	2910	1291	-	4202
painting	-	-	2962	2962
person	38418	-	-	38418
plant	2805	493	-	3298

⋆This work was partly done when H. Zhao was an intern at Adobe Research.
[1] https://flickr.com

Table 2: Evaluation set statistics. The second column stands for total number of foreground candidates of each category in the evaluation set. Column 3 to 12 stand for number of ground truths or positive foreground candidates of each background.

Category	all	1	2	3	4	5	6	7	8	9	10
boat	144	63	44	73	72	83	63	51	29	66	41
bottle	204	109	109	97	106	36	118	133	104	99	4
car	244	43	37	31	17	42	49	30	21	23	28
chair	249	10	20	20	33	19	16	17	16	7	10
dog	322	196	196	193	71	137	136	171	95	68	147
painting	114	28	34	27	24	23	11	34	35	25	20
person	364	51	94	12	79	61	19	25	97	69	38
plant	140	9	102	22	35	64	55	16	48	65	65

2 More Qualitative Results

Visual Search Results We show more visual search results in Fig. 1, Fig. 2 and Fig. 3. Compared to RealismCNN, shape information and classification features, the returned results of our approach contain more compatible foregrounds for image compositing.

Generalization to New Categories To further exhibit the representation ability of our learned shared feature across multiple classes, we test our method on new categories that have not been trained. The search results are illustrated in Fig. 4. Even without training on the new classes, our system still works reasonably well.

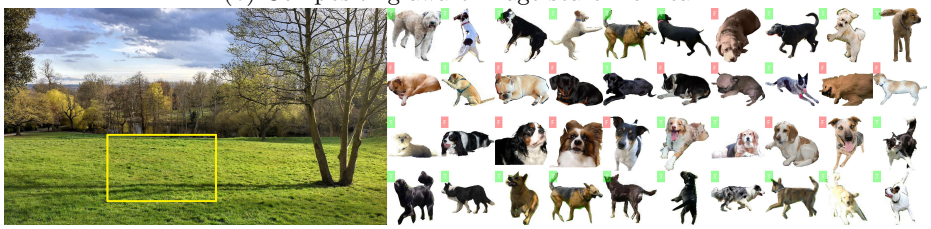
Foreground Similarity We essentially learned new feature representations to measure image similarity using image compositing as a proxy in a self-supervised manner. The learned features can not only be used for image compositing, but can also be applied in other search scenarios, *e.g.*, finding similar foregrounds from a query foreground using our learned foreground features. Some visual results of this task are shown in Fig. 5 and Fig. 6. We can see our features can capture more finegrained similarity in semantics, shape and pose, benefiting from the rich information learned through compositing.

References

1. Zhou, B., Zhao, H., Puig, X., Fidler, S., Barriuso, A., Torralba, A.: Scene parsing through ade20k dataset. In: CVPR. (2017)
2. Lin, T., Maire, M., Belongie, S.J., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft coco: Common objects in context. In: ECCV. (2014)
3. Everingham, M., Gool, L.J.V., Williams, C.K.I., Winn, J.M., Zisserman, A.: The pascal visual object classes VOC challenge. IJCV (2010)

(a) Compositing-aware image search for 'boat'

(b) Compositing-aware image search for 'car'



(c) Compositing-aware image search for 'dog'

(d) Compositing-aware image search for 'person'

Fig. 1: Visual search results. In each example, the yellow box indicates the position of foreground object to be inserted. The 1st to the 4th rows show the retrieved results using RealismCNN, shape information, classification features and our approach, respectively. The text boxes with 'green' and 'red' color in the top left corner of the foregrounds represent 'positive' and 'negative' foregrounds respectively. Our returned results contain more compatible foregrounds for image compositing.

(a) Compositing-aware image search for 'boat'

(b) Compositing-aware image search for 'chair'

(c) Compositing-aware image search for 'painting'

(d) Compositing-aware image search for 'plant'

Fig. 2: Visual search results. In each example, the yellow box indicates the position of foreground object to be inserted. The 1st to the 4th rows show the retrieved results using RealismCNN, shape information, classification features and our approach, respectively. The text boxes with 'green' and 'red' color in the top left corner of the foregrounds represent 'positive' and 'negative' foregrounds respectively. Our returned results contain more compatible foregrounds for image compositing.

(a) Compositing-aware image search for ‘bottle’

(b) Compositing-aware image search for ‘dog’

(c) Compositing-aware image search for ‘person’

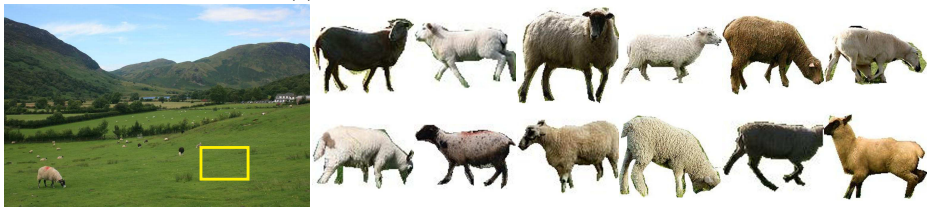
(d) Compositing-aware image search for ‘plant’

Fig. 3: Visual search results. In each example, the yellow box indicates the position of foreground object to be inserted. The 1st to the 4th rows show the retrieved results using RealismCNN, shape information, classification features and our approach, respectively. The text boxes with ‘green’ and ‘red’ color in the top left corner of the foregrounds represent ‘positive’ and ‘negative’ foregrounds respectively. Our returned results contain more compatible foregrounds for image compositing.

(a) Search for new category 'bus'

(b) Search for new category 'suitcase'

(c) Search for new category 'cow'



(d) Search for new category 'sheep'

(e) Search for new category 'cat'

Fig. 4: Generalization to new categories.

(a) Foreground search results on 'boat'

(b) Foreground search results on 'bottle'

(c) Foreground search results on 'dog'

(d) Foreground search results on 'chair'

(e) Foreground search results on 'person'

Fig. 5: Measuring foreground similarities. The leftmost one marked with blue box is the query foreground. In each case, the top row are the search results using ResNet50 classification features, and the bottom one are our results.

(a) Foreground search results on ‘dog’

(b) Foreground search results on ‘car’

(c) Foreground search results on ‘plant’

(d) Foreground search results on ‘painting’

(e) Foreground search results on ‘person’

Fig. 6: Measuring foreground similarities. The leftmost one marked with blue box is the query foreground. In each case, the top row are the search results using ResNet50 classification features, and the bottom one are our results.