OPTIMIZATION OF SEMANTIC IMAGE RETARGETING BY USING GUIDED FUSION NETWORK

R.Pradeep[1], P.Sangamithra[2], S.Shankar[3], R.Venkatesh[4],K.Santhakumar[5]
[1234]UG Students,[5]Associate Professor,Department Of ECE,
Nandha Engineering College(Autonomous),Erode-52,Tamil Nadu

Abstract- Image retargeting has been applied to display images of any size via devices with various resolutions (e.g., cell phone and TV monitors). To fit an image with the target unimportant regions need to be deleted or distorted, and the key problem is to determine the importance of each pixel. Existing methods in a bottom-up manner via eye fixation estimation or saliency detection. In contrast, the predict pixel wise importance proposed the pixel-wise importance based on a top-down criterion where the target image maintains the semantic meaning of the original image. To this end, several semantic components corresponding to foreground objects, action contexts, and background regions are extracted.

KEY WORDS: Image retargeting, semantic component, semantic collage, classification guided fusion network.

I.INTRODUCTION

Image retargeting is a widely studied problem that aims to display an original image of arbitrary size on a target device with different resolution by cropping and resizing. Considering a source image is essentially a carrier of visual information, we define the image retargeting problem as a task to generate the target image that preserves the semantic information of the original image. For example, the image in Figure 1 shows a boy kicks a ball on a pitch (sports field), which contains four semantic components including boy, kicking, ball and pitch. Based on the source image, four target images can be generated as shown in Figure 1. The first three target images are less informative as certain semantic components are missing. The last target image is the only one that preserves all four semantic components .Existing retargeting methods operate based on an importance map which indicates pixel-wise importance. To generate a target image in Figure 1 that preserves semantics well, the pixels corresponding to semantic components, e.g., boy and ball, should have higher weights in the importance map such that these are preserved in the target image. In other words, an importance map needs to preserve semantics of the original image well.

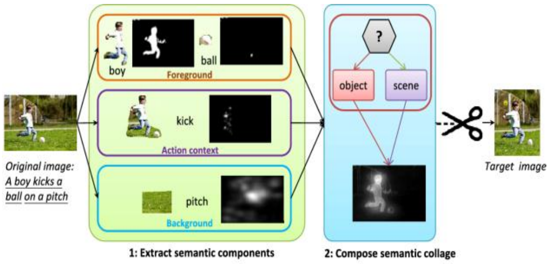


Fig1: Image Retargeting

II. Conventional Image Retargeting

Early image retargeting methods are developed based on saliency detection that models the human eye fixation process. As these bottom-up methods are driven by low-level visual cues, edges and corners in images are detected rather than semantic regions. Although the thumb-nailing method uses similar images in an annotated dataset to construct a saliency map for cropping this task-driven approach does not exploit or preserve high-level visual semantics. In contrast, the proposed SP- DIR algorithm can better preserve semantic meanings for image retargeting. Other retargeting methods crop images to improve visual quality of photographs However, these schemes do not explicitly preserve visual semantics, which may discard important contents for the sake of visual quality and aesthetics.

A. Semantic-Based Image Retargeting

In recent years, more efforts have been made to analyze image contents for retargeting. Luo detects a number of classes, e.g., skin, face, sky and grass, to crop photos. In Yan et al. extend the foreground detection method of with a human face detector to crop images. The semantic components introduced in Section III-A have several limitations. First, although the state-of-the-art deep modules are used, the semantic component maps may not be accurate. For example, the detection module are likely to generate false positives or negatives. Second, the context information between different semantic components is missing. For example, in Figure 2, the spatial relationship between boy and ball is missing in the individual semantic component maps. To address these issues, we propose a classification guided fusion network to integrate all component maps. While the importance maps have been used in the existing image retargeting methods, we emphasize the semantic collage in this work effectively preserves semantics and integrates multiple semantic component maps based on different cues.

B. Semantic Component

The semantic components including foreground, action context and background are extracted to describe an image for retargeting. Semantic Foreground Components: The salient objects in an image are considered as the semantic foreground components. For example, the image contains two main foreground components, i.e., boy and ball. We use the state of-the- art image parsing and classification modules to locate foreground components. Image parsing. Apply the pre-trained fully convolutional network to parse each input image into 59 common categories defined in the Pascal- Context dataset. The 59 categories, though still limited, include common objects that frequently occur in general images. Use all 59 parsing confidence maps where each semantic component map is denoted by Mp. As shown in, the semantic component maps highlight the objects, i.e., person and building, well. First, for concreteness use 59 categories defined in the Pascal-Context dataset to demonstrate the effectiveness of the proposed algorithm. While limited, they include common objects that frequently occur in general images. Second, several larger semantic segmentation datasets are released recently. For example, the ADE20K dataset contains 150 object and stuff classes with diverse annotations of scenes, objects, parts of objects, and in some cases even parts of parts. Third, it requires extensive manual labeling work to extend to a large number of categories, i.e., 3000 categories. One feasible approach is to resort to the weakly supervised semantic segmentation methods where bounding box] or image level annotations are available. Image classification use the VGG-16network pre-trained under image the ILSVRC 2012 dataset to predict a label distribution over 1, 000 object categories in an image. As each classification is carried out on the image level, an importance map is obtained via a back propagation pass from the VGG network output]. The semantic component map induced by the classification output using 1- channel image is denoted by The semantic collage Mg is obtained by Mg = c(o|M) · ro(M) + c(s|M) ·rs(M) (1) where M = {Mp, Mc, Ms, Ma} is the concatenation of all semantic component maps to be fused and contains 62 channels. In the above equation, ro(·) and rs(·) are regression functions for object-oriented and scene- oriented, respectively. In addition, c(o) and c(s) are the confidences that the image belongs to object or scene- oriented one. The semantic collage can be generated by a soft or hard fusion based on whether c is the classification confidence or binary output.

C. Network Training

The training process involves 3 stages by increasingly optimizing more components of network

Stage1. The classification sub-network is trained first as its results guide the regression sub-network. Here only he parameters related to the classification sub-network are updated. The loss function L1 at this stage is a weighted multinomial logistic loss: L1 = 1 N X N i=1 ωi log(ˆωi) (2) where ωi ∈ {0, 1} is ground truth classification label, ωˆi is the probability predicted by the classification sub-network, and N is the training set size Stage 2 . We train both classification and regression sub networks without CRF-RNN layers in this stage. The loss function L2 is: L2 = L1 + 1 [N X N] i=1 X W x=1 X H y=1 kIi,x,y − ˆIi,x,yk 2 (3) where I and ˆI are the ground truth and estimated semantic collages. In addition, W and H are width and height of input image, respectively. Stage 3. The CRF- RNN layers are activated. The loss function of this stage is the same as L2

III.PROBLEMIDENTIFICATION

Semantic components may not be extracted well in an image. Numerous image retargeting methods have been developed. visual quality can be reduced. It can generate only one input of array of pixels. Fixed resolution. Three semantic components including foreground action context and background.

IV.PROPOSED METHOD

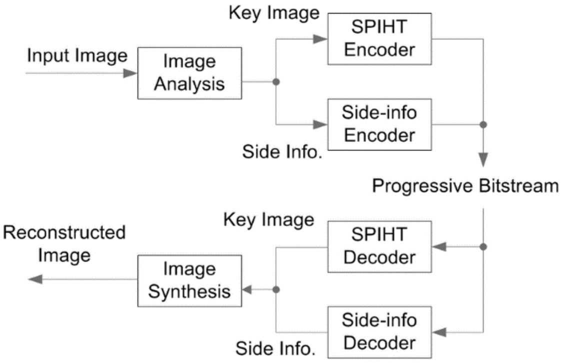


Fig2: Bitstream Image

Are extracted from an image. For example, in the image of Figure the boy and ball are foreground components, kick and pitch belong to the action context, and the rest is background. Semantic components are extracted by using the stage-of the-art modules based on deep learning. Wiener filtering is used to avoid the damage in pixel clarity. Destored rectification algorithm is used in neural network.34.5 classifiers is the format of the image.

V.IMAGE COLLAGE:

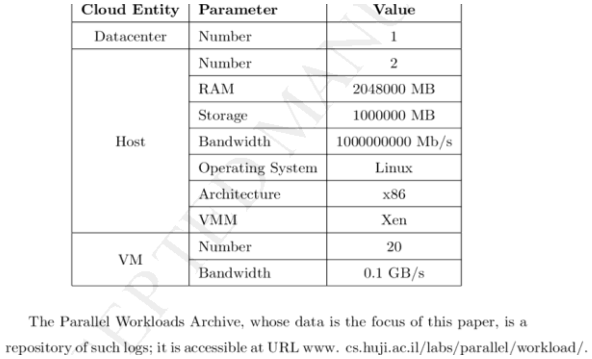
Fig-3: Collaged Image

VI. IMAGE RETARGETTING

We select images from the Pascal VOC datasets. In addition, we collect images from Google and Bing search engines .Based on the contents, all images are divided into 6 categories including single person, multiple people, single as well as multiple objects, and indoor as well as outdoor scenes. The images in single person, multiple people, single object and multiple objects classes are object- oriented while other images are scene- oriented. Table I shows the properties of the S-Retarget dataset. Some representative images are shown in Figure. The dataset is split into train/val/test subsets, containing images respectively. The distribution of the 6 categories are almost the same in the three sets. Semantic collage. We ask 5 subjects to annotate the pixel relevance based on the semantics of an image. The labeling process consists of two stages. In the first stage, each subject annotates the caption of an image. Several image captions are show. In the second stage, the annotators rate all pixels by referring to the image caption provided in the first stage. To facilitate labeling, each image is over segmented 5 times using multiple over-segmentations methods including 3 times and Quick Shift twice with different segmentation parameters, e.g., number of super pixels and compactness factors. Each annotator then assigns a value to each image segment where a higher score corresponds to high relevance.

VII. Experimental Settings Implementation details.

In the training process, we use 3 × 10−5 as learning rate in the first two stages and 3 × 10−6 in the last stage Datasets and baseline methods. We carry out experiments on the Retaret and S-Retarget datasets (see Section IV). Evaluation metric. We use the metrics of the MIT saliency benchmark dataset for evaluation including the Earth Mover's Distance (EMD), Pearson linear coefficient (CC), Kullback- Leibler divergence (KL), histogram intersection (SIM), and mean absolute error (MAE). For EMD, KL, MAE, the lower the better while for C Cand SIM, the higher the better. The other three metrics in the MIT saliency benchmark are not adopted as they require eye fixation as ground truth. We carry out user study to evaluate the retargeted results from different methods using the Amazon mechanical turk (AMT). Each AMT worker



Cloud Entity	Parameter	Value
Datacenter	Number	1
Host	Number	2
	RAM	2048000 MB
	Storage	1000000 MB
	Bandwidth	1000000000 Mb/s
	Operating System	Linux
	Architecture	x86
	VMM	Xen
VM	Number	20
	Bandwidth	0.1 GB/s

The Parallel Workloads Archive, whose data is the focus of this paper, is a repository of such logs; it is accessible at URL www. cs.huji.ac.il/labs/parallel/workload/.

Table1: Details Of Implemetation

VIII. Sensitivity analysis

Each generated semantic collage is fed into a carrier to generate the target image by removing or distorting less important pixels. In this experiment, we randomly select 60 images from each subsets in the S-Retarget to evaluate the proposed semantic collage with 6 baseline importance map generation methods using 3 carriers, i.e., AAD, multi-operator and importance filtering (IF) The baseline map generation methods and carriers are the same as discussed in Section V-A.. The results of all 6 subsets are presented in Table V where we use AMT scores for evaluation. For the Single person subset, the semantic collage + AAD method is preferred by 155 persons while the e DN + AAD scheme is favored for 50 times. Overall, the proposed semantic collage performs favorably against all the baselines in all subsets.

IX Comparison between S-Retarget and ECSSD

To demonstrate the merits of the proposed S-Retarget dataset, we compare the retarget results generated by the models trained on different datasets. In addition to the proposed dataset, we also consider the ECSSD data base . For fair comparisons, we use the following experimental settings. The ECSSD dataset is split into a training and a test set with 900 and 100 images respectively. We also select 100 images from the test set of the S- Retarget dataset. The selected 200 images from both datasets (100 from each one) form an unbiased test set. Our SP-DIR model is trained both on the S-Retarget and ECSSD datasets, and then evaluated on the new unbiased test set. We use training data salience method to denote different training dataset and salience method settings. In addition to our SP-DIR method, we also test with the state-of-the-art MC method for saliency detection. Therefore, there are 4 different experiment settings including: retargeting method.

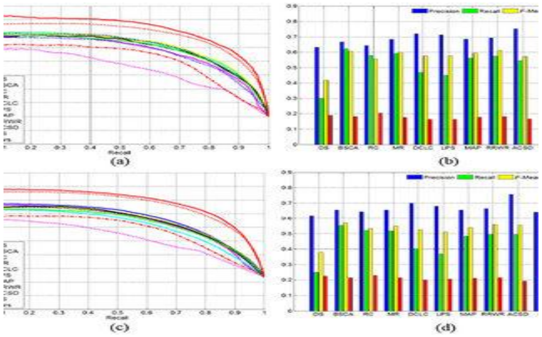


Fig4 : Graph Between S-target & ECSSD

X.CONCLUSION

In this paper, we propose a deep image retargeting algorithm that preserves the semantic meaning of the original image. A semantic collage that represents the semantic meaning carried by each pixel is generated in two steps. First, multiple individual semantic components, i.e., including foreground, contexts and background, are extracted by the state-of-the-art deep understanding modules. Second, all semantic component maps are combined via classification guided fusion network to generate the semantic collage. The network first classifies the image as object or scene- oriented one. Different classes of images have their respective fusion parameters. The semantic collage is fed into the carrier to generate the target image. Our future work include exploring image caption methods for calculating retargeting and related problems. In addition, we plan to integrate the Pixel CNN.

XI.REFERENCES:

[1]Y.-S. Wang, C.-L. Tai, O. Sorkine, and T.-Y. Lee, "Optimized scale- and stretch for image resizing," ACM TOG, 2008. 1, 2, 7

[2]D. Panozzo, O. Weber, and O. Sorkine, "Robust image retargeting via axis-aligned deformation," in EUROGRAPHICS, 2012. 1, 2, 3, 7, 8

[3]M. Rubinstein, A. Shamir, and S.Avidan,"Multi-operator media retargeting," ACM TOG, 2009. 1, 2, 3, 7, 8

[4]J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," CoRR, vol. abs/1411.4038, 2014. 1, 3

[5]A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutionalneural networks," in NIPS, 2012. 1

[6]M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Learning and transferring mid-level image representations using convolutional neural networks," in CVPR, 2014. 1, 3

[7]B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva, "Learning deep features for scene recognition using places database," in NIPS, 2014. 1, 3, 4

[8]S. Bhattacharya, R. Sukthankar, and M. Shah, "A framework for photoquality assessment and enhancement based on visual aesthetics," in MM, 2010. 1, 4

[9]Y. Ding, J. Xiao, and J. Yu, "Importance filtering for image retargeting," in CVPR, 2011. 1, 2, 3, 7, 8

[10]J. Sun and H. Ling, "Scale and object aware image thumb nailing," IJCV, vol. 104, no. 2, pp. 135–153,2013. 2, 7

[11]M. Rubinstein, A. Shamir, and S. Avidan, "Improved seam carving for video retargeting," in ACM TOG,
2008. 2, 7

[12]G.-X. Zhang, M.-M. Cheng, S.-M. Hu, and R. R. Martin, "A shapepreserving approach to image resizing," Computer Graphics Forum, 2009.

[13.]Han, B.; Zhu, H.; Ding, Y. Bottom-up saliency based on weighted sparse coding residual.
In Proceedings of the ACM International Conference on Multimedia, Scottsdale, AZ, USA, 28 November–1 December 2011; pp. 1117–1120.]

[14.] Yang, J.; Yang, M.-H. Top-down visual saliency via joint CRF and dictionary learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 2296–2303.]

[15.]Mehmood, I.; Sajjad, M.; Ejaz, W.; Baik, S.W. Saliency-directed prioritization of visual data in wireless surveillance networks. *Inf. Fusion* **2015**, *24*, 16–30.]]

[16.]Sajjad, M.; Ullah, A.; Ahmad, J.; Abbas, N.; Rho, S.; WookBaik, S. Integrating salient colors with rotational invariant texture features for image representation in retrieval system. *Multimed. Tools Appl.* **2018**, *77*, 4769–4789.

[17.]Sajjad, M.; Ullah, A.; Ahmad, J.; Abbas, N.; Rho, S.; WookBaik,S.Saliency-weighted graphs for efficient visual content description and their applications in real-time image retrieval systems. *J. Real-Time Image Process.* **2017**, *13*, 431–447

[18.]Borji, A.; Itti, L. Exploiting local and global patch rarities for saliency detection. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Providence, RI, USA, 16–21 June 2012; pp. 478–485.]

[19.]Duan, L.; Wu, C.; Miao, J.; Qing, L.; Fu, Y. Visual saliency detection by spatially weighted dissimilarity. In Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Colorado Springs, CO, USA, 20–25 June 2011; pp. 473–480.

[20.]Lu, H.; Li, X.; Zhang, L.; Ruan, X.; Yang, M.H. Dense and Sparse Reconstruction Error Based Saliency Descriptor. *IEEE Trans. Image Process.***2016**, *25*, 1592–1603.

[21.]Huo, L.; Yang, S.; Jiao, L.; Wang, S.; Shi, J. Local graph regularized coding for salient object detection. *Infrared Phys. Technol.* **2016**, *77*, 124–131.

[22.]Huo, L.; Yang, S.; Jiao, L.; Wang, S.; Wang, S. Local graph regularized reconstruction for salientobject
detection. *Neurocomputing* **2016**, *194*, 348–359

[23.]Yang, C.; Zhang, L.; Lu, H. Graph Regularized Saliency Detection With Convex-Hull-Based Center Prior. *IEEE Signal Process. Lett.* **2013**, *20*, 637–640.

[24.]Hou, X.; Zhang, L. Dynamic visual attention: Searching for coding length increments. Advances in Neural Information Processing Systems 21. In Proceedings of the 22nd Annual Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 8–11 December 2008; pp. 681–688

[25.]Shen, X.; Wu, Y. A unified approach to salient object detection via low rank matrix recovery. In Proceedings of the IEEE Conference on Computer Vision Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 853–860.