Non-Local Deep Features for Salient Object Detection

Zhiming Luo[1,2], Akshaya Mishra[3], Andrew Achkar[3], Justin Eichel[3], Shaozi Li[1], Pierre-Marc Jodoin[2]
[1]Xiamen University, China
[2]University of Sherbrooke, Canada
[3]Miovision Technologies Inc., Canada

IEEE 2017 Conference on Computer Vision and Pattern Recognition
CVPR July 21-26 HONOLULU 2017

Xiamen University · miovision rethink traffic · UNIVERSITÉ DE SHERBROOKE

Introduction

Goal

Highlight the most relevant objects in an image (**saliency**).

Proposed method

- Novel 4 × 5 multiresolution CNN grid structure
- Contrast features
- Loss function inspired by the Mumford-Shah functional
- No CRF, no superpixels

Outcome

- Top performing method on 6 datasets
- Real-time, high performance saliency detection.

Mumford-Shah Function [5]

$$F^{MS} = \underbrace{\sum_j \lambda_j \int_{\mathbf{v}\in\Omega_j} \left|I(\mathbf{v}) - u_j\right|^2 d\mathbf{v}}_{\text{data fidelity}} + \underbrace{\sum_j \gamma_j \oint_{\mathbf{v}\in C_j} d\mathbf{v}}_{\text{boundary length}}$$

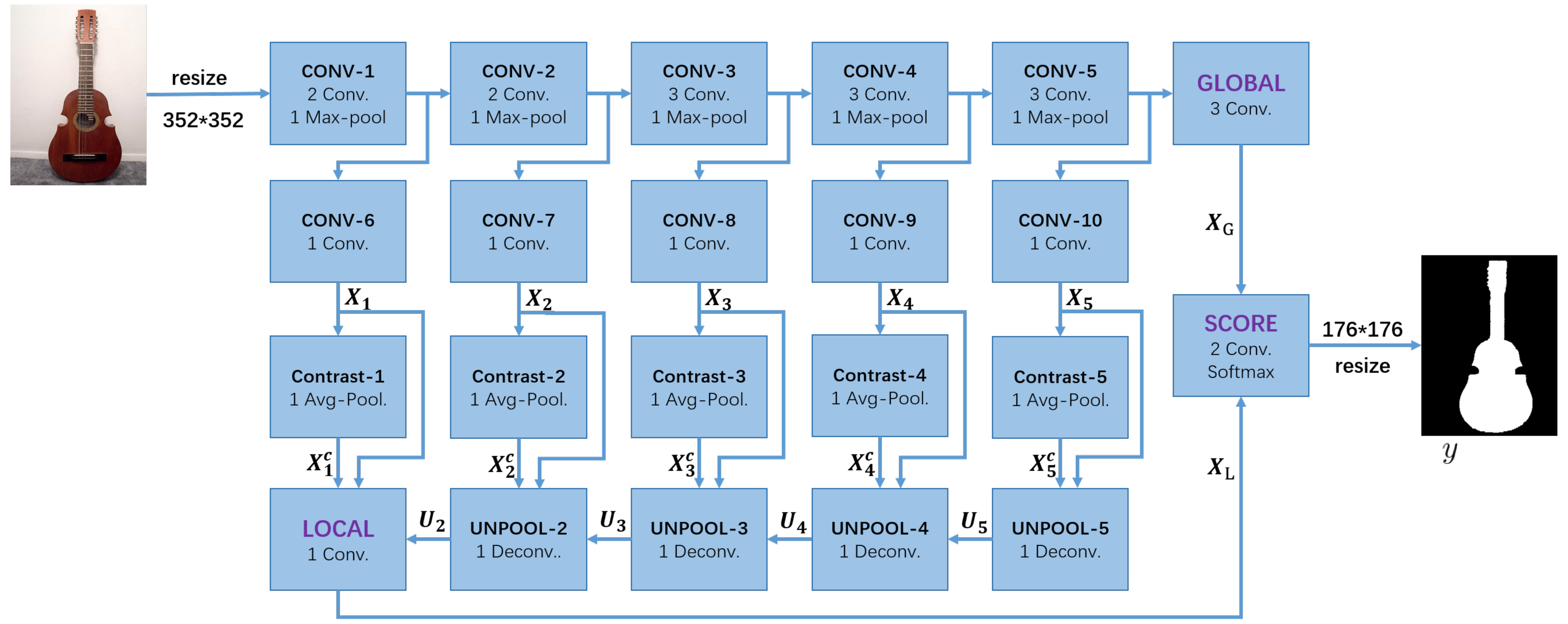
Bayesian statistical approximation [6]

$$F^{MS} \approx \underbrace{\sum_j \lambda_j \int_{\mathbf{v}\in\Omega_j} \log p_j(I(\mathbf{v}),\mathbf{v}) d\mathbf{v}}_{\text{data fidelity}} + \underbrace{\sum_j \gamma_j \oint_{\mathbf{v}\in C_j} d\mathbf{v}}_{\text{boundary length}}$$

Final loss function

$$F^{MS} \approx \underbrace{\sum_j \lambda_j \int_{\mathbf{v}\in\Omega_j} \mathrm{H}_j(y(\mathbf{v}),\hat{y}(\mathbf{v}))}_{\text{cross entropy}} + \underbrace{\sum_j \gamma_j (1 - \mathrm{IoU}(C_j,\hat{C}_j))}_{\text{boundary IoU loss}}$$

Network Architecture



Cross-Entropy Loss

$$\hat{y}(\mathbf{v}) = p\big(y(\mathbf{v}) = c\big) = \frac{e^{\overbrace{W_L^c X_L(\mathbf{v}) + b_L^c}^{\text{Local}} + \overbrace{W_G^c X_G + b_G^c}^{\text{global features}}}}{\sum_{c'\in\{0,1\}} e^{W_L^{c'} X_L(\mathbf{v}) + b_L^{c'} + W_G^{c'} X_G + b_G^{c'}}}$$

(Local & global features)

$$H_j(y(\mathbf{v}),\hat{y}(\mathbf{v})) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c\in\{0,1\}} (y(\mathbf{v}_i) = c)\big(\log(\hat{y}(\mathbf{v}_i) = c)\big)$$

IoU Boundary Loss (Dice Loss)

$$\text{IoU Loss} = 1 - \frac{2|C_j \cap \hat{C}_j|}{|C_j| + |\hat{C}_j|},$$

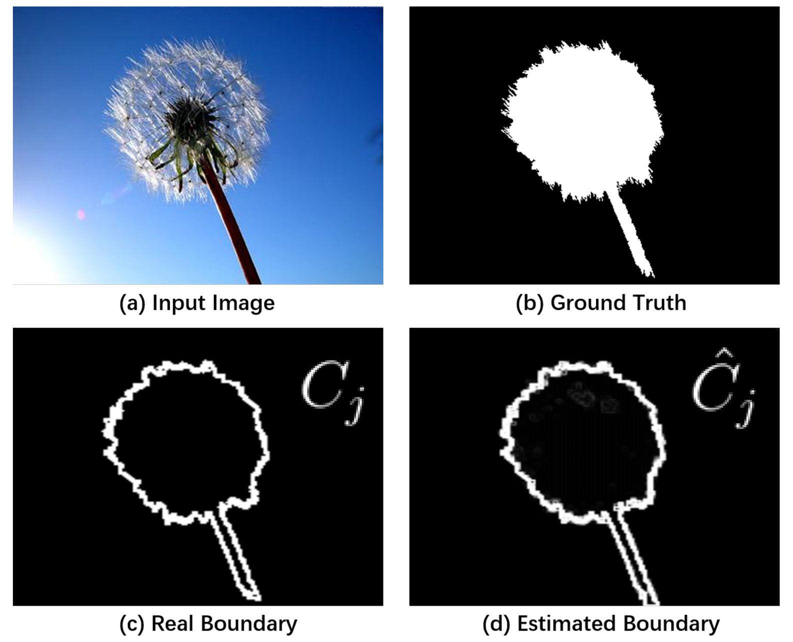


Figure 1. The IoU boundary loss, end-to-end trainable.

Figure 2. Visual comparison of saliency detection results with and without the boundary loss term.

Results and Evaluation

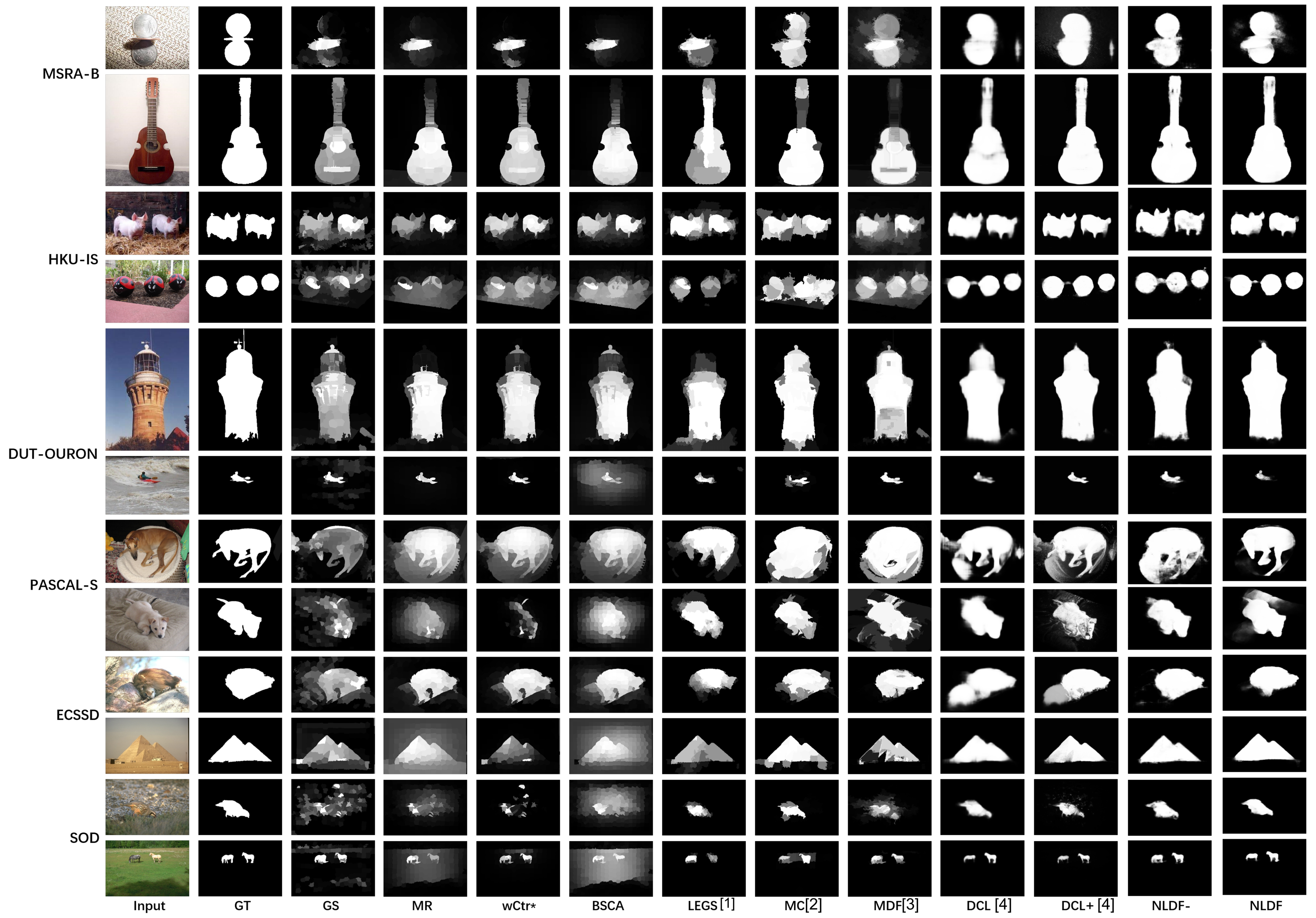


Figure 3. Visual comparison between previous approaches and our method (NLDF).

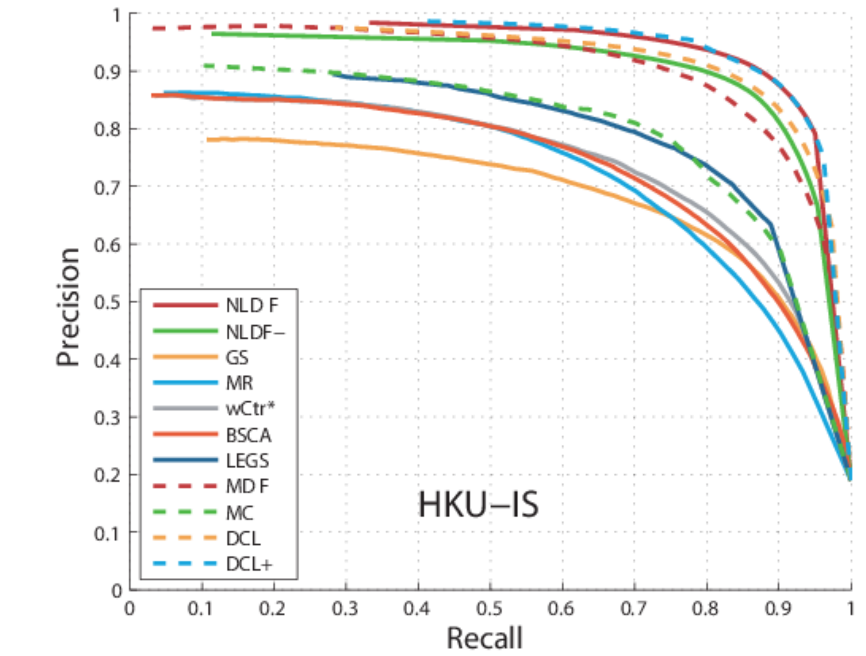


Figure 4. Precision and recall curve for the HKU-IS datasets

	LEGS[1]	MC[2]	MDF[3]	DCL[4]	DCL+[4]	**NLDF**
s/img	2	1.6	8	1.5	2.3	**0.08**

Table. Average execution time to process an image

Bibliography

[1] L. Wang, H. Lu, X. Ruan, and M. Yang. *Deep networks for saliency detection via local estimation and global search*. In Proc. CVPR, 2015.
[2] R. Zhao, W. Ouyang, H. Li, and X. Wang. *Saliency detection by multi-context deep learning.* In Proc. CVPR, 2015.
[3] G. Li and Y. Yu. *Visual saliency based on multiscale deep features*. In Proc. CVPR, 2015
[4] G. Li and Y. Yu. *Deep contrast learning for salient object detection.* In Proc. CVPR, 2016.
[5] D. Mumford and J. Shah. *Optimal approximations by piecewise smooth functions and associated variational problems*. Communications on pure and applied mathematics, 42(5):577–685, 1989.
[6] T. Brox and D. Cremers. *On local region models and a statistical interpretation of the piecewise smooth mumford-shah functional*. Int. J. Comput. Vision, 84(2):184–193, Jun. 2009.