LEARNING TO SEGMENT ON TINY DATASETS: A NEW SHAPE MODEL

Maxime Tremblay and André Zaccarin

Department of Electrical and Computer Engineering, Université Laval, Québec (QC) Canada

Motivation

The main goal of this work is to **detect** and **segment** objects using only **tiny datasets**. To this extent, we propose a new automatic part-based object segmentation algorithm for non-deformable and semi-deformable objects in natural backgrounds.

Shape Descriptor

- Need shape descriptors that model strong boundaries
- Must be robust to small shape variations
- Can model straight and curved lines

Our shape descriptor is a quantized SIFT descriptor on the ground truth binary masks of the objects. They are used to generate part-based shape prior for our detection and segmentation framework.

Quantization:

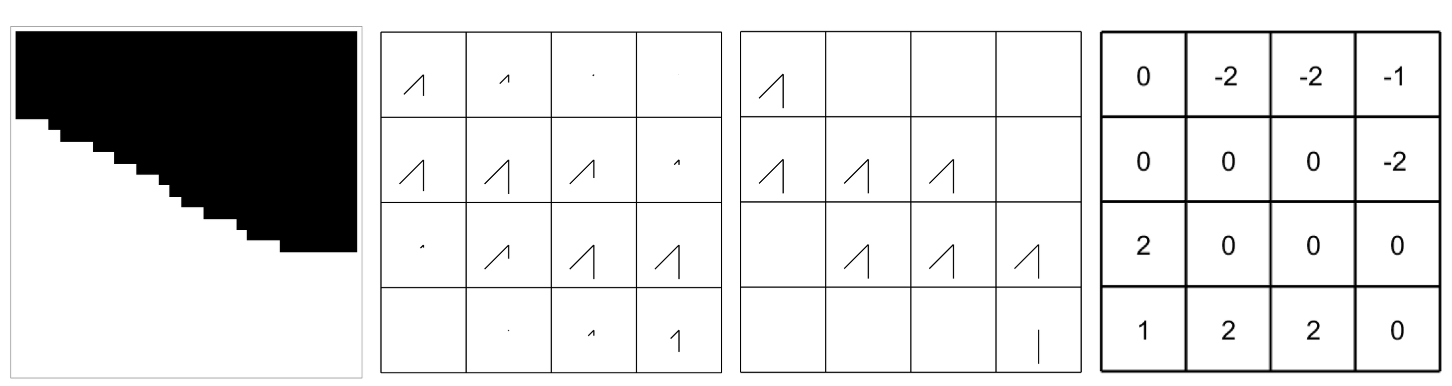
$$\mathcal{D}_k(i) = \frac{\text{sgn}\left(\frac{d_k(i)}{m} - 1\right) + 1}{2}, \quad m = \beta \max_i d_k(i)$$

Foreground/background prior:

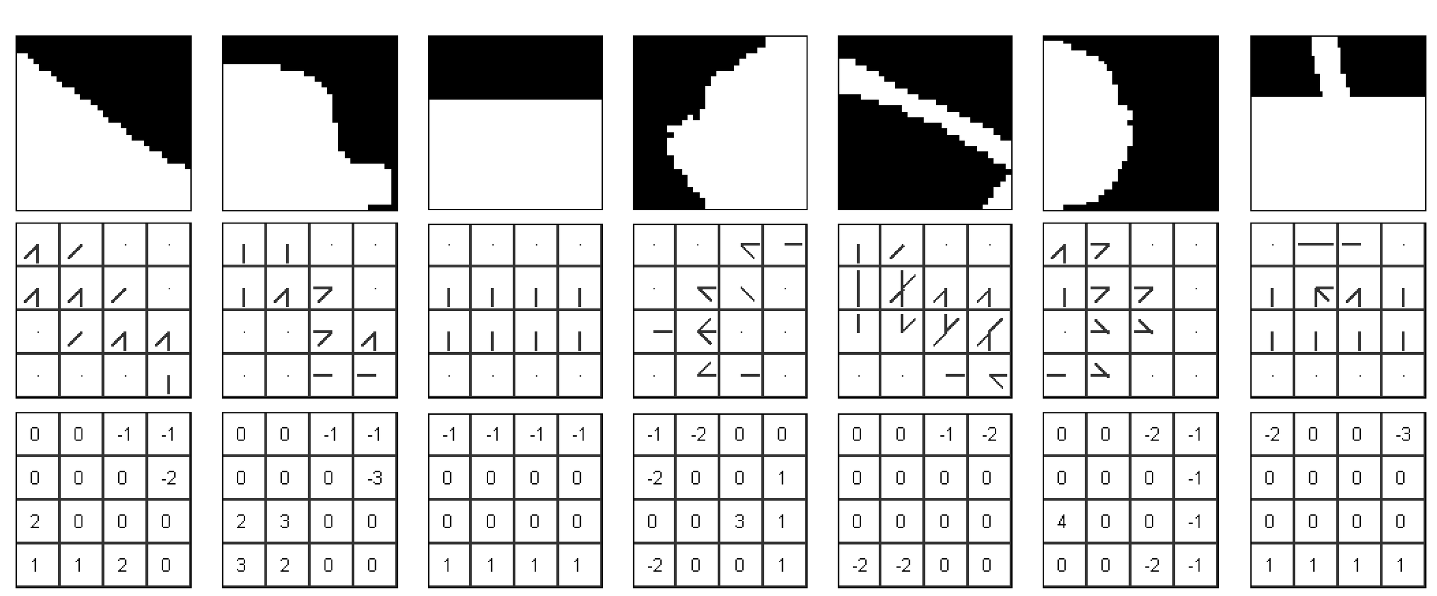
$$v_k(i) = \sum_{j=0}^{7} h_j((j+4) \bmod 8) - h_j(j), \quad \text{if} \sum_{k=0}^{7} h_i(k) = 0$$

Propagation to isolated cell:

$$v_i = \max_j v_j + \min_j v_j, \quad \text{iff} \sum_{j=0}^{7}\sum_{k=0}^{7} h_j(k) = 0$$



Ground truth patch | Ground truth SIFT | Shape descriptor | Fg/bg prior



More shape descriptor examples

References

- P. Krähenbühl and V. Koltun. Parameter Learning and Convergent Inference for Dense Random Fields. *ICML*. 2013.
- B. Leibe, *et al.* Robust Object Detection with Interleaved Categorization and Segmentation. *IJCV*, 2008.
- D. R. Magee and R. D. Boyle. Detecting Lameness Using 'Re-sampling Condensation' and 'Multi-Stream Cyclic Hidden Markov Models'. *IVC*, 2002.
- P. O. Pinheiro, *et al.* Learning to Refine Object Segments. *ECCV*. 2016.
- J. Shotton, *et al.* TextonBoost: Joint Appearance, Shape and Conext Modeling for Muli-class object Recognition and Segmentation. *CVPR*. 2006.
- S. Zagoruyko, *et al.* A MultiPath Network for Object Detection. *BMVC*. 2016.

Detection and Segmentation Framework

Part-based object detection

We use a **bag-of-words** approach based on Leibe *et al.* [2] work. Contrarily to standard bag-of-words approach, codewords extracted solely from the foreground and are not used for any general representation of an image.

Training

1. Extract foreground features (Harris-Lagrange + SIFT)
2. Extract shape descriptors
3. Hierarchical clustering (codewords)
4. Keep occurrences (l_x, l_y, s)

Test

1. Extract features (Harris-Lagrange + SIFT)
2. Compare with codewords
3. Every match votes for an object hypothesis
4. Mean-shift mode estimation to identify acceptable hypothesis
5. Non-maximum suppression on detections

An object is detected if enough occurrences vote at the position in the voting domain (x, y, s)

Segmentation

We frame the segmentation problem as a dense CRF which we solve using the mean field approximation of Krähenbühl *et al.* [1].

Energy function:

$$E(A) = \eta \sum_{x_i} \psi_u(x_i|A) + (1-\eta) \sum_{x_i,x_j} \psi_p(x_i, x_j|A)$$

Unary term:

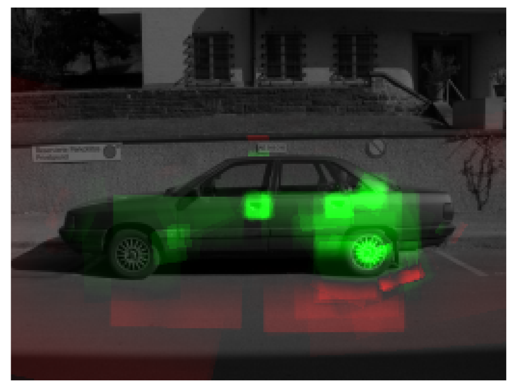
$$\psi_u(x_i|A) = \lambda_1 \psi_{shape}(x_i|A) + \lambda_2 \psi_{color}(x_i|A) + \lambda_3 \psi_{roi}(x_i|A)$$

$\psi_{shape}(x_i|A)$ is created by projecting coherent occurrences v_i onto the image domain.

Pairwise term:

$$\psi_p(x_i, x_j|A) = \sum_{m=1}^{C} \mu^{(m)}(x_i, x_j|A) k^{(m)}(f_i - f_j)$$

Parameters λ_1, λ_2, λ_3, and η are found in validation.

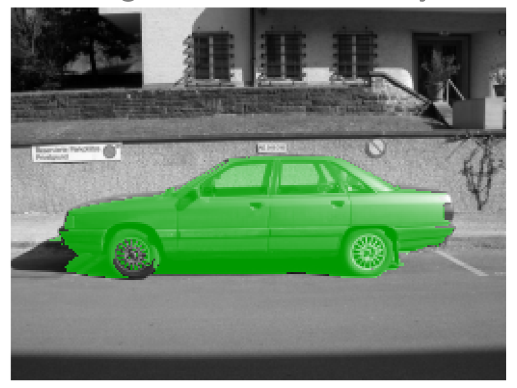


Shape unary

Color unary

Region-of-interest unary

Combined



Segmentation

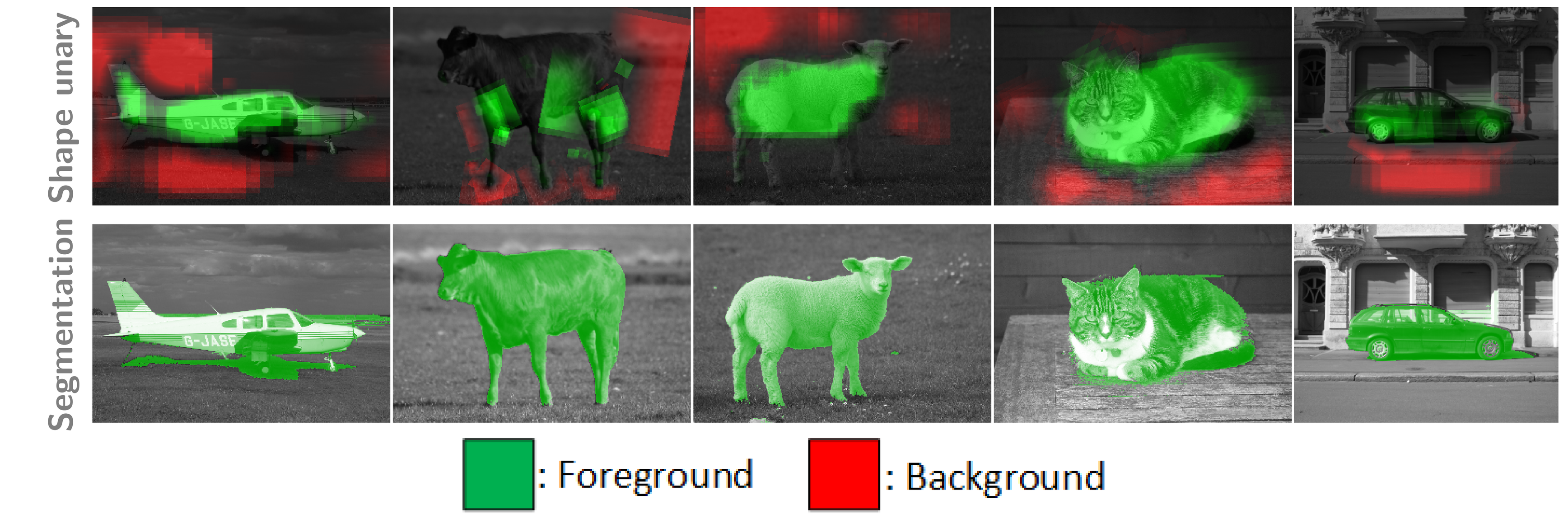
Experiments

Performance were evaluated on two small image sets with detection and segmentation ground truth: TUDarmstadt Object Dataset (TUD) [3] and MSRC21 [5].

Evaluation

Datasets:

- TUD and MSRC21 have respectively 100 and 30 images per class.
- We split TUD and MSRC21 in respectively 3 and 5 random folds for each class.
- For TUD *sideviews-cars* images, we kept mirrored pairs in the same fold.



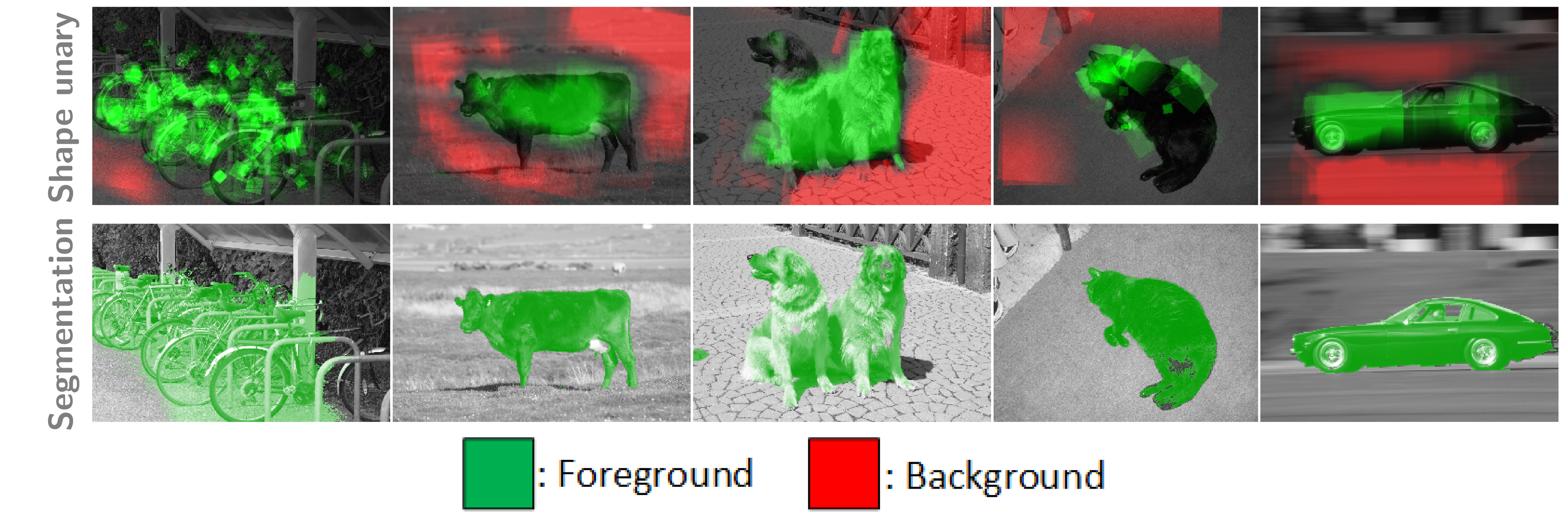
Upper bound - Trained on COCO's full training set

	TUD		MSRC21						
	sideviews cars	sideviews cows	plane	cow	car	bike	sheep	cat	dog
SharpMask	0.40	0.52	0.29	0.71	0.40	0.19	0.48	0.61	0.48
SharpMask + MPN	0.39	0.52	0.29	0.68	0.37	0.18	0.44	0.61	0.45

Our performance - Trained on 32-34 images on TUD and 18 on MSRC21

	TUD		MSRC21						
	sideviews cars	sideviews cows	plane	cow	car	bike	sheep	cat	dog
BSM	0.39	0.48	0.17	0.57	0.18	0.26	0.48	0.22	0.13

- Since SharpMask [4] does not produce any labeling; we funnel its segmentations to a MultiPath Network [6].
- SharpMask without MPN is evaluated on masks which overlap with the ground truth ($iou \geq 0.5$).
- mAP measurement uses the PASCAL recall step (0.1) instead of COCO's (0.01) considering the size of the sets.



Conclusion

- Perform well on really small sets of data (15-20 training images)
- Tight segmentation
- Good with occluded objects

This work was supported through funding from Auto21 (Canada) and the REPARTI strategic center (FRQ-NT, Québec).