Context-Aware Object Detection for Enhanced Neural Network Training in Limited-Data Scenarios

Sepehr Akbari Sugata Banerji

Department of Mathematics & Computer Science, Lake Forest College

Abstract

We present Objectness, a modular framework for generating context-aware and semantically meaningful image crops, designed to enhance neural network training in data-limited scenarios where the context of the data is critical for learning. In applications such as fine-grained classification or rare object detection, traditional augmentation techniques often fail to capture the contextual relationships that define each class, especially when training examples are scarce. Our approach integrates a Faster-RCNN model with Region Proposal Network (RPN) for deep learning-based saliency detection and Binarized Normed Gradients (BING), a fast and efficient method for objectness estimation. By focusing on the most informative and contextually relevant regions, Objectness produces crops that preserve the semantic integrity of the original images, enriching limited training sets with meaningful, representative samples. Domain-agnostic and adaptable, Objectness enables more effective training and improved classification performance in environments where both data and context are constrained.

Deep Context Extraction & Refinement

We extract semantically meaningful regions using a Faster R-CNN architecture (ResNet-50 backbone + FPN) [1]. This builds a multi-scale "Context Pyramid" to detect objects of varying sizes without losing resolution [2].

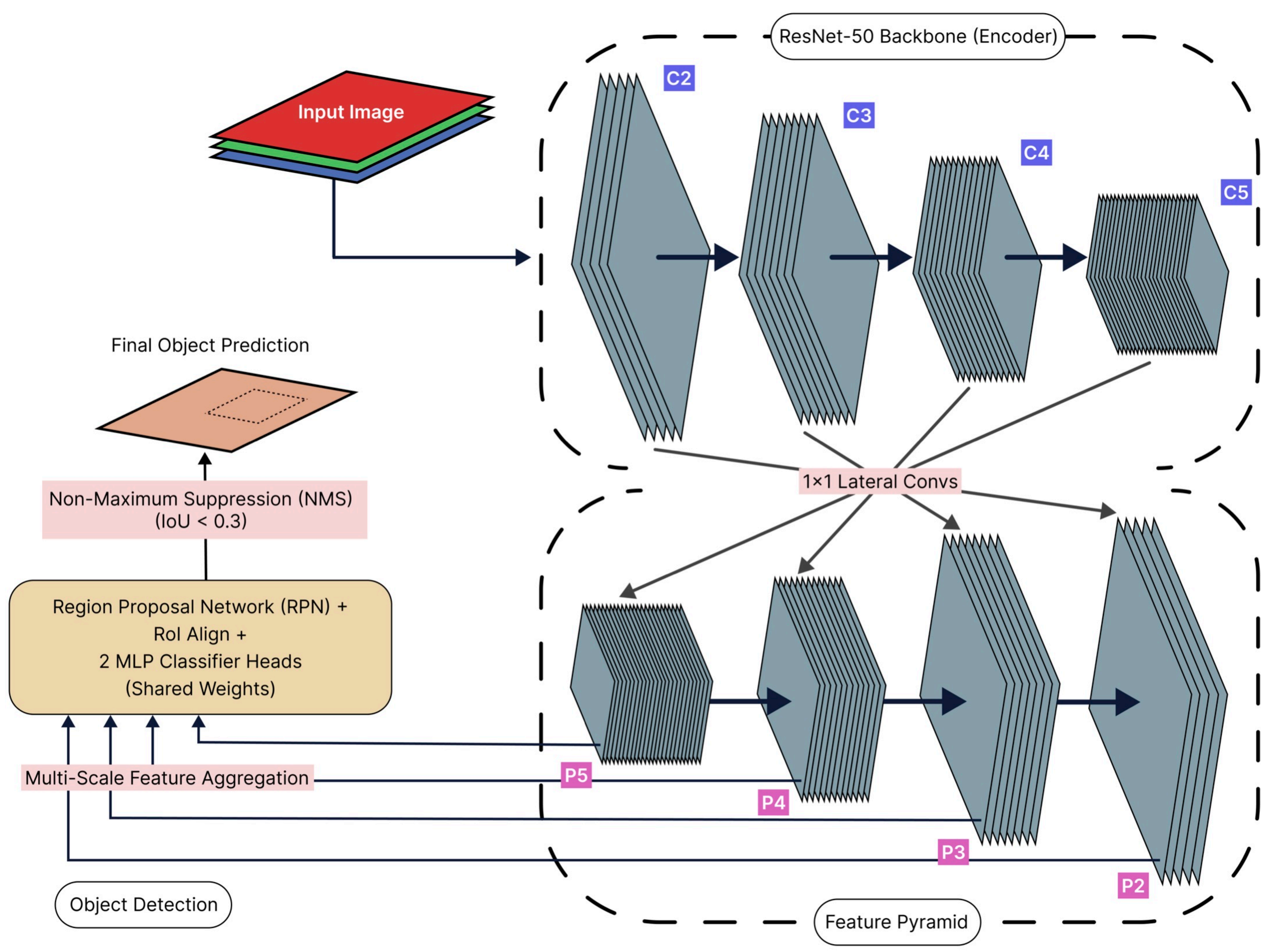

Figure 1. Faster R-CNN architecture with ResNet-50 backbone and Feature Pyramid Network (FPN) for multi-scale feature extraction.

1. Multi-Scale Context (FPN): The top-down pathway ($P_5 \to P_2$) restores spatial detail from deep semantic maps. By merging these pathways, the network captures both broad structures and fine artistic brushstrokes.

2. Region Proposal & Bounding Box Regression: The Region Proposal Network (RPN) slides over the feature pyramid, predicting scale-invariant coordinate transformations (t_x, t_y, t_w, t_h) relative to anchor boxes (x_a, y_a, w_a, h_a):

$$t_x = \frac{x - x_a}{w_a}, \quad t_y = \frac{y - y_a}{h_a}, \quad t_w = \log\left(\frac{w}{w_a}\right), \quad t_h = \log\left(\frac{h}{h_a}\right)$$

3. Strict Filtering & RoI Align: To prevent redundancy in our limited training set, the RPN enforces a strict confidence threshold. We apply Non-Maximum Suppression (NMS) to eliminate overlapping candidate regions, discarding bounding box B if its Intersection over Union (IoU) with the highest-scoring box A exceeds our threshold of 0.3:

$$\mathrm{IoU}(A, B) = \frac{|A \cap B|}{|A \cup B|} < 0.3$$

Finally, RoI Align uses bilinear interpolation to extract exact feature maps for these continuous coordinates, bypassing the quantization errors of standard max-pooling.

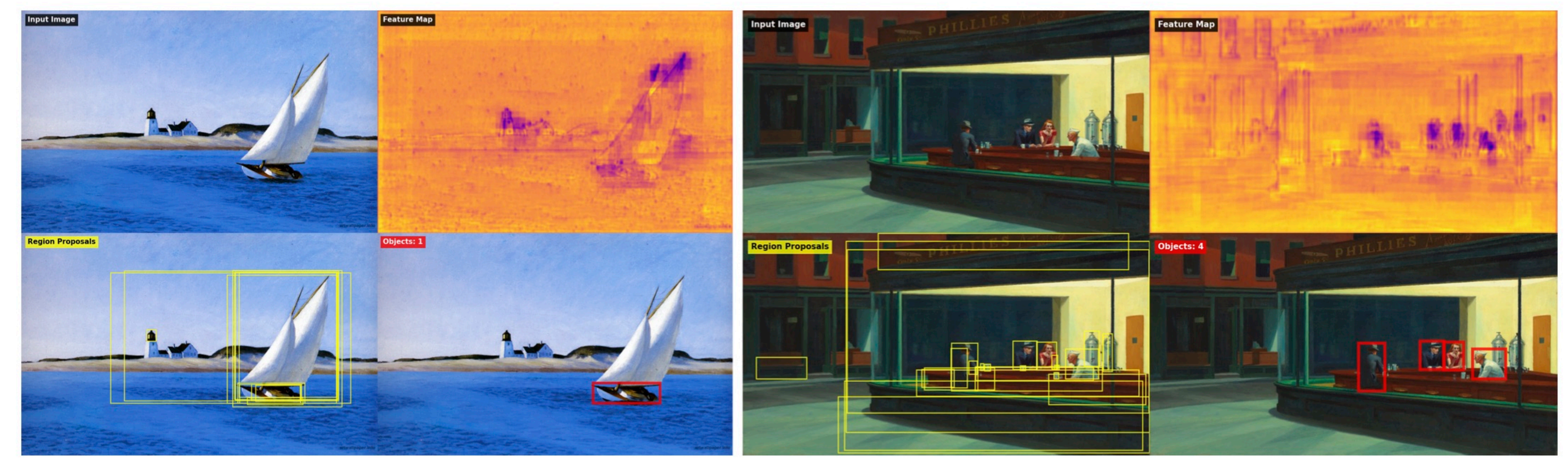

Figure 2. (i,ii) ResNet-50 feature activations, (iii) unfiltered RPN proposals, (iv) top proposals after NMS thresholding.

High-Speed Structural Recall

To ensure robust dataset coverage when our deep learning model is overly selective, we integrate Binarized Normed Gradients (BING) for ultra-fast, class-agnostic objectness estimation at 300fps [3].

Figure 3. Normed gradient maps.

1. Binarized Feature Space & Bitwise Scoring: Candidate regions are resized to a compact 8×8 grid. We extract the Normed Gradient magnitude,

$$g = \sqrt{\nabla_x^2 + \nabla_y^2}$$

which is highly sensitive to dense, closed boundaries. To achieve high-speed performance, these gradient maps are thresholded into a 64-bit binary feature vector, $b \in \{0, 1\}^{64}$.

The objectness score s is evaluated using a learned linear SVM weight vector w. By utilizing binarized features, the standard dot product is approximated via hardware-level bitwise operations (e.g., POPCNT), entirely bypassing expensive floating-point arithmetic:

$$s = \langle w, b \rangle \approx \sum_{i=1}^{64} w_i \cdot b_i$$

2. Structural Saliency (Object vs. Background): While Faster R-CNN relies on semantic deep features, BING acts as a structural safety net. It assigns high objectness scores to generic "blob-like" structures with strong edge alignment, capturing distinct artistic elements that may bypass class-specific filters.

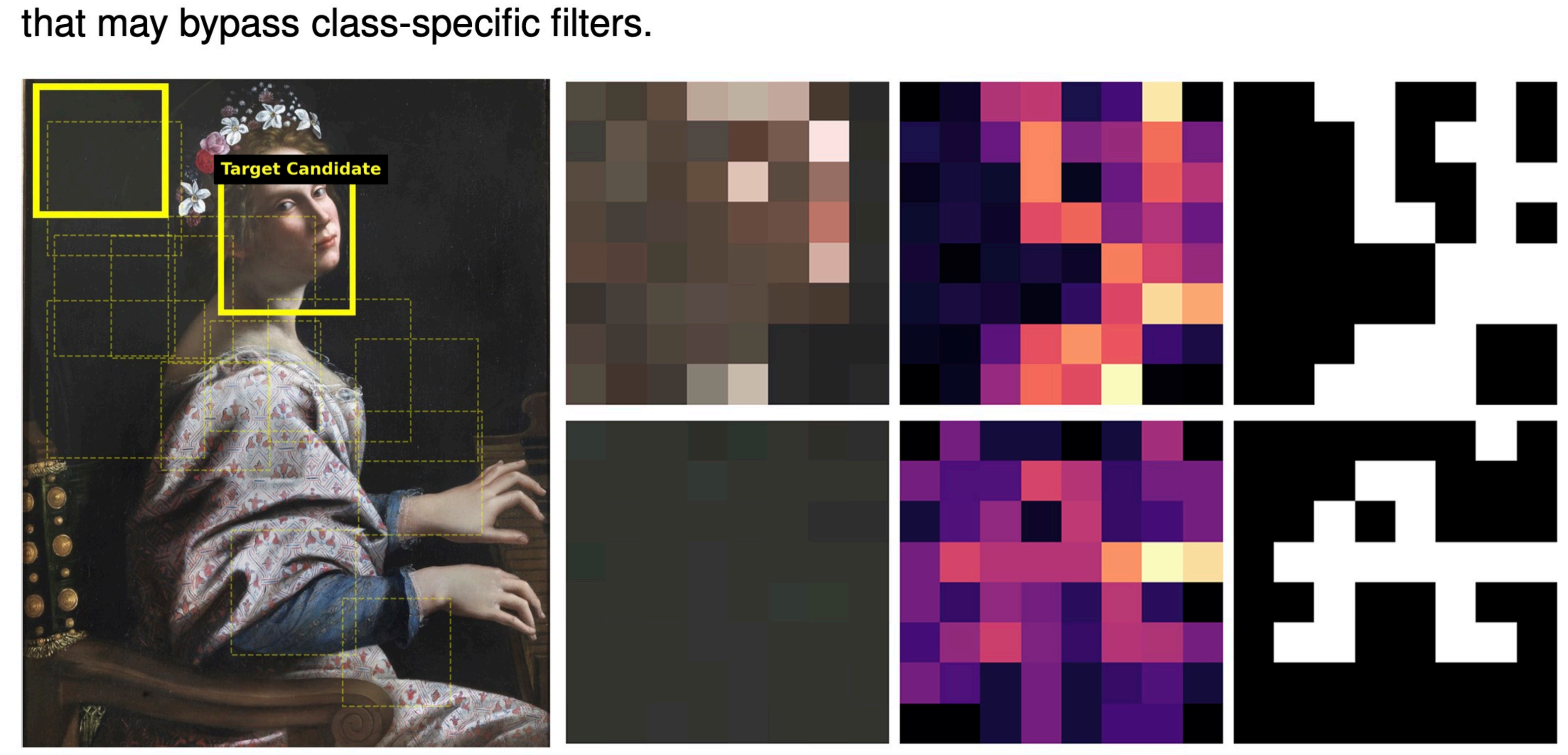

Figure 4. 8×8 binarized feature maps. Upper: High objectness. Lower: Low objectness.

Hybrid Orchestration Pipeline

The system is managed by an Orchestrator that routes images through our detection models to ensure consistency.

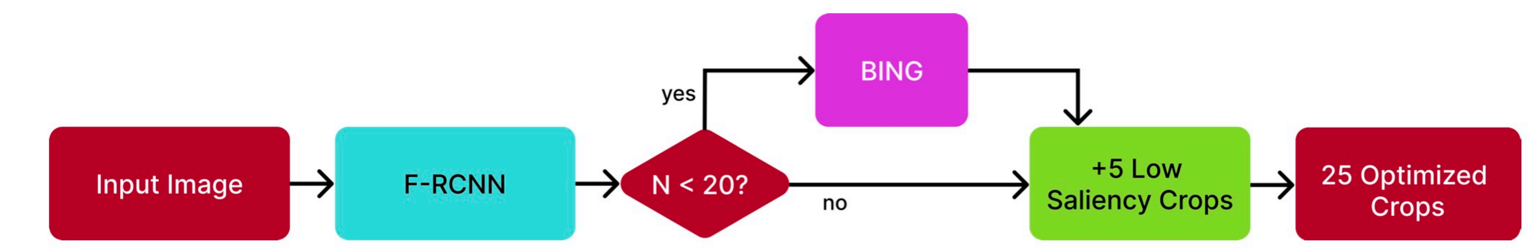

Figure 5. Orchestrator pipeline.

Results

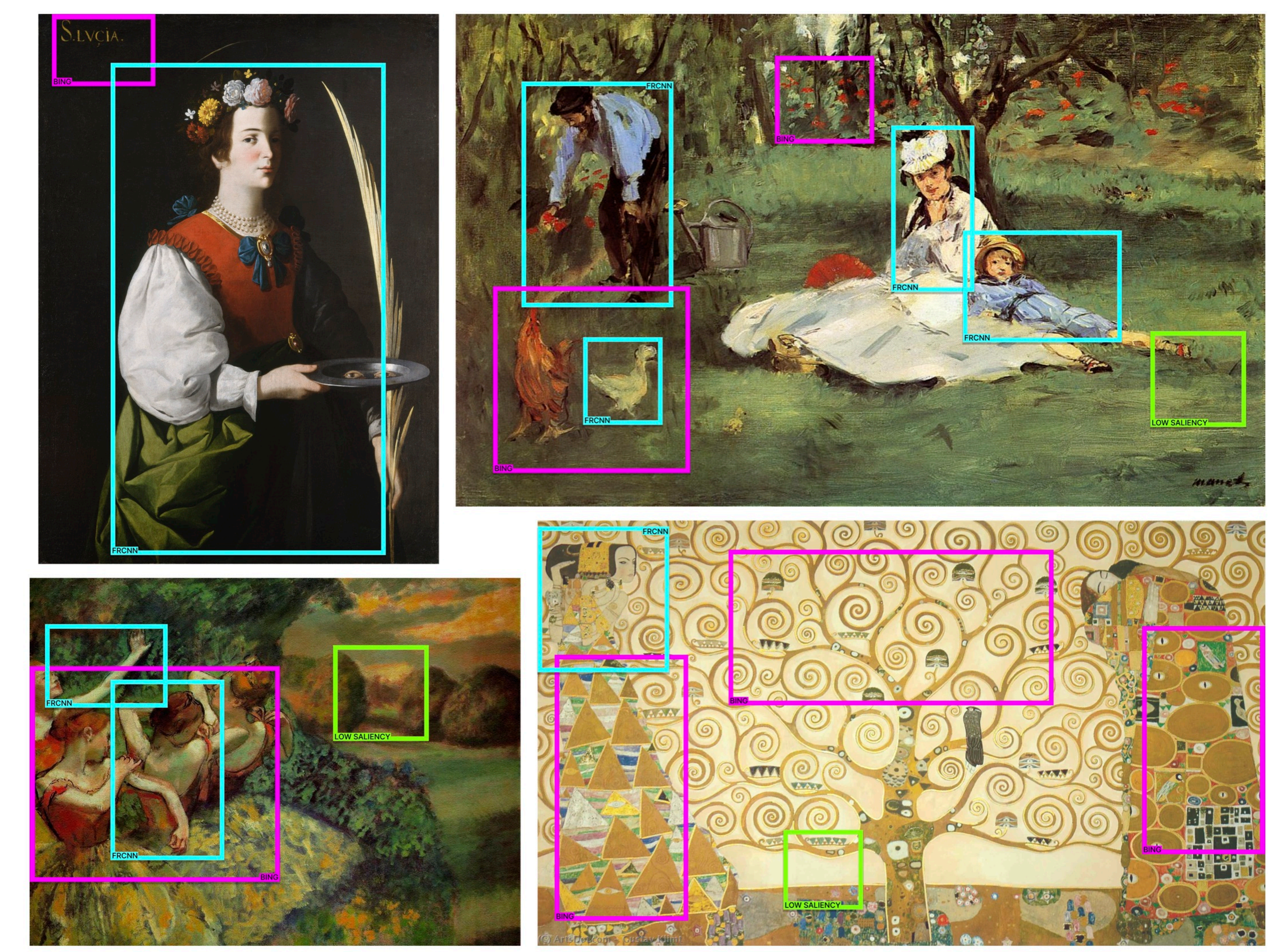

Figure 6. Cyan: Faster R-CNN proposals, Magenta: BING proposals, Green: Low-saliency crops.

Summary

Challenge: Training classifiers on limited datasets (e.g., 30 per class in 100 classes) leads to overfitting. Standard augmentations (rotation, jitter, etc.) destroy artistic style.

Solution:

- Faster R-CNN with FPN captures multi-scale semantic context, generating high-quality proposals focused on salient objects.
- BING with a fast, class-agnostic safety net to recall structurally salient regions that may be missed by the RPN.
- Low saliency crops to preserve contextual information and prevent overfitting to prominent objects.
- An orchestrator to dynamically balance precision and dataset volume, ensuring a consistent number of meaningful crops per image.

Result: "context-aware" crops that focus on standout objects to increase training data volume without losing semantic meaning.

References

[1] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems (NeurIPS)*, 2015.

[2] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[3] Zhi Cheng, Yujia Liu, Liang Wang, Yuhang Zhang, and Xiang Li. Bing: Binarized normed gradients for objectness estimation at 300fps. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (1), 2014.