BLIND PROPOSAL QUALITY ASSESSMENT VIA DEEP OBJECTNESS REPRESENTATION AND LOCAL LINEAR REGRESSION

Qingbo Wu[1], *Hongliang Li*[1], *Fanman Meng*[1], *King N. Ngan*[1,2], *and Linfeng Xu*[1]

[1]University of Electronic Science and Technology of China
[2]The Chinese University of Hong Kong
{qbwu,hlli,fmmeng,knngan,lfxu}@uestc.edu.cn

ABSTRACT

The quality of object proposal plays an important role in boosting the performance of many computer vision tasks, such as, object detection and recognition. Due to the absence of manually annotated bounding-box in practice, the quality metric towards blind assessment of object proposal is highly desirable for singling out the optimal proposals. In this paper, we propose a blind proposal quality assessment algorithm based on the Deep Objectness Representation and Local Linear Regression (DORLLR). Inspired by the hierarchy model of the human vision system, a deep convolutional neural network is developed to extract the objectness-aware image feature. Then, the local linear regression method is utilized to map the image feature to a quality score, which tries to evaluate each individual test window based on its *k*-nearest-neighbors. Experimental results on a large-scale IoU labeled dataset verify that the proposed method significantly outperforms the state-of-the-art blind proposal evaluation metrics.

***Index Terms*—** Blind proposal quality assessment, deep objectness representation, local linear regression

1. INTRODUCTION

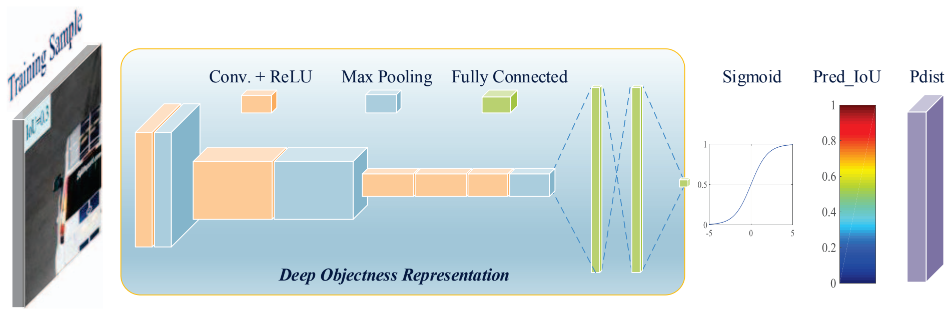
Following the increasing demands for fast object detection system [1, 2, 3], the proposal algorithm has become an active research field in recent years, which aims at generating small amounts of candidate windows to avoid exhaustively searching for massive sliding windows [4, 5]. The generated proposals are preferred to cover the objects as tightly as possible, where the popular Intersection-over-Union (IoU) index [6, 7, 8] is typically used to quantitatively measure the quality of each window. As a full-reference metric, the IoU performs well in evaluating the proposal quality when the manual annotations are given. However, in many automatic recognition systems, the interactive information are unavailable, which brings urgent demand for blind proposal quality assessment (BPQA) to approach the IoU index.

Existing proposal algorithms have made a great effort in blindly estimating the proposal quality, which can be roughly classified into two categories. The first class of methods model the BPQA as a foreground/background segmentation problem, where the foreground regions are considered to possess higher proposal quality than the backgrounds. In [9], Sande *et al.* utilized the iterative superpixel merging to obtain foreground regions, where the size and texture similarities are two crucial clues to activate the mergence operation. The similar idea is also employed in [10], and Manén *et al.* proposed a random sampling based maximum spanning tree algorithm to accelerate the merging process. To improve the segmentation accuracy of previous low-level feature based merging scheme, Chang *et al.* [11] integrated two visual attention clues, i.e., saliency and objectness into a graph model. Then, the segmentation was formulated as a energy function minimization problem, which was solved by the alternative optimization.

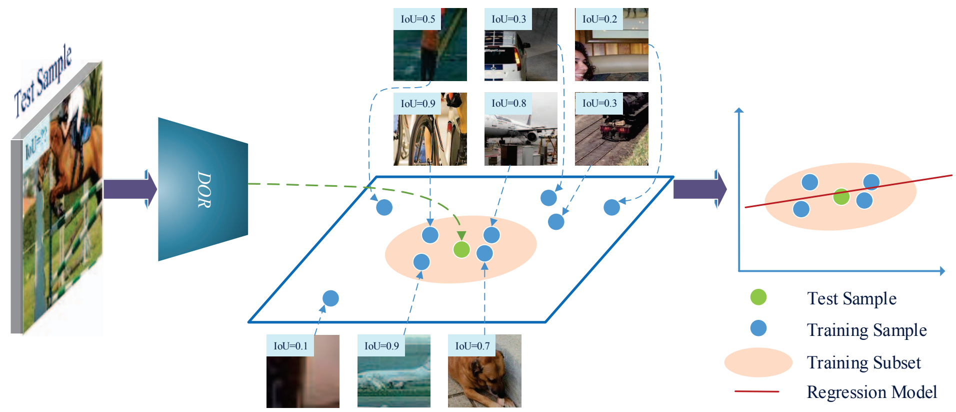
It is worth nothing that these segmentation based methods could only roughly identify the proposal quality in a binary mode (i.e., foreground/background), which fairly limits their capability in quantitatively evaluating the proposal. To cope with this issue, the second class of BPQA methods pay more attentions on developing kinds of window ranking or scoring functions in terms of specific image cues. In [12], Rahtu *et al.* proposed three objectness related features, i.e., the superpixel boundary, boundary edge and window symmetry, to feed a cascaded ranking model [13]. The similar idea could be found in [14], and a non-maximal suppression strategy was applied to remove the candidate windows significantly overlapped with the others. Endres *et al.* explored rich appearance features in [15], and a diversity rewarding function was developed to rank the proposals. In addition to the ranking scheme, many scoring models are also discussed in recent literatures. In [16], Zitnick *et al.* scored a bounding box by using the sum of edge strength within this box to subtract the edge strength of the contours straddling the box's boundary. In [4], Alexe *et al.* computed the proposal quality by combining four objectness measures, i.e., multi-scale saliency, color contrast, edge density and superpixels straddling, with a Bayesian model. Cheng *et al.* proposed a computationally efficient feature,

This work was supported in part by National Natural Science Foundation of China, under grant numbers 61601102, 61525102 and 61502084.

978-1-5090-6067-2/17/$31.00 ©2017 IEEE



(a) Phase I: Deep Objectness Representation



(b) Phase II: Local Linear Regression

Fig. 1. The framework of the proposed method consisting of two phases. (a) shows the process of phase I, which is used to train a CNN based objectness attribute representation. (b) shows the process of phase II, which is used to learn the local linear regressor from the Bayesian neighbors of the test sample.

i.e., binarized normed gradients, in [17] and mapped this image feature to the proposal quality with a linear model. In [18], Carreira *et al.* extracted a large set of mid-level features to capture the object regularities, and the random forest regression was used to learn the quality prediction function. Recently, the deep convolutional neural network (CNN) is utilized to integrate the proposal algorithm into a detection system [19], where the end-to-end method is employed to learn the proposal generator and evaluator simultaneously.

Although developed from different clues of generic objects, these BPQA algorithms share one prominent common ground. That is, each quality metric is specifically designed for a proposal algorithm. When this metric is applied to the other proposal algorithms, its evaluation results are very likely inaccurate due to the usage of different image clues and training data. This is applicable to a typical content-unaware detection or recognition system, which supposes there is unique "optimal" proposal algorithm for all of given images. However, since only limited object-relevant information are explored, the performances of different proposal algorithms would change across the images with diverse visual contents.

It inspires us to develop a generic BPQA method, which could measure the objectness of a bonding box without any limitation about its generation methods. When the test images change from one scene to another one, more robust object-covered windows could be selected across multiple proposal algorithms by means of our BPQA metric. To this end, a CNN based deep feature extractor is utilized to describe the objectness attribute. Then, a local linear regression method is developed to predict the quality for each given proposal. Unlike the poplar end-to-end method that typically trains a CNN with all training data, we focus on capturing specific local property of the training set that are mostly relevant to each test image. More specifically, we carefully build the training subset for each test window, where only its k nearest neighborhoods (kNN) with high probability of sharing similar qualities are collected from the whole train set. Then, the parameters of the linear regressor are learned from this subset. Due to the flexibility of capturing content related training data, the proposed method could adaptively update the regression parameters for each individual test sample, which is crucial for delivering better regression accuracy. As shown in our experiments, the proposed method could significantly outperform many state-of-the-art metrics in estimating the proposal quality.

Table 1. The structure of the deep feature extractor

Layer	Type	Kernel size	Pooling size	Stride	Pad
conv1	Conv.+ReLU	11	-	4	0
pool1	Max pooling	-	3	2	0
conv2	Conv.+ReLU	5	-	1	2
pool2	Max pooling	-	3	2	0
conv3	Conv.+ReLU	3	-	1	1
conv4	Conv.+ReLU	3	-	1	1
conv5	Conv.+ReLU	3	-	1	1
pool5	Max pooling	-	3	2	0
fc6	Conv.+ReLU	6	-	1	0
fc7	Conv.+ReLU	1	-	1	0
fc8	Conv.	1	-	1	0
sg8	Sigmoid	-	-	-	-

2. PROPOSED METHOD

As shown in Fig. 1, given the cropped windows and their associated IoU values, we propose to train the BPQA in two phases, which include 1) deep objectness representation, and 2) local linear regression. Motivated by the success of CNN in many image recognition tasks [20, 21], a deep neural network based feature extractor is utilized to describe the objectness attribute. In the following, the quality regressor is learned from the *k*NN of each individual test window.

2.1. Deep Objectness Representation

Inspired by the neurophysiological insights towards visual recognition, the multi-layered convolutional neural network was developed to simulate the hierarchy model of the mammals' visual nervous system [22], which are composed of resemble simple cells and complex cells in the visual cortex. More specifically, the underlying convolutional layers perform in the role of feature extractor, which works like the simple cells and tries to capture the local image features located in the receptive fields. Let $\omega_i^k \in R^{m\times n}$ denote the kernel of kth filter in the ith convolution layer, and $b_i^k \in R$ denote its corresponding bias. Then, the kth feature map produced by the ith convolution layer is represented by

$$h_i^k = R(\omega_i^k * \mathbf{h}_{i-1} + b_i^k) \tag{1}$$

where $\mathbf{h}_{i-1} = \{h_{i-1}^1, \cdots, h_{i-1}^{K_{i-1}}\}$ denotes the feature maps outputted from K_{i-1} channels of the previous convolution layer, and $\mathbf{h}_0$ corresponds to the input image. $R(x) = \max\{0, x\}$ is the rectifier activation function to enforce the sparsity-inducing regularization [23].

Followed by the local feature description, the spatial pooling operation is applied to the feature map h_i^k to obtain the translation invariant mid-level image feature, which imitates the function of complex cells in the mammalian visual cortex [24]. In our method, we apply the max pooling operation [25] to each feature map, which computes the the max value of the neighboring feature values in each non-overlapped patch.

After five loops of the convolution and pooling operations, the feature maps are fully connected to 4096 hidden neurons in two times, which represents the input image with a 4096-dimension feature vector. Finally, the output layer encodes the proposal quality with one neuron, where the sigmoid function [26] is followed to map the quality score into the range [0, 1].

Let $\mathbf{w}_i = \{\omega_i^1, \cdots, \omega_i^{K_i}, b_i^1, \cdots, b_i^{K_i}\}$ denote the parameters of the ith convolutional layer. Then, the trainable parameter set for the L layers deep neural network could be represented by $\mathbf{W} = \{\mathbf{w}_i, \cdots, \mathbf{w}_L\}$. Let $f(x_j; \mathbf{W})$ denote the output of the deep neural network including all convolution, pooling and sigmoid operations for the jth training window x_j, and y_j denote its IoU value. The inference for $\mathbf{W}$ could be solved by minimizing the prediction error between y_j and $f(x_j; \mathbf{W})$, i.e.,

$$\hat{\mathbf{W}} = \arg\min \frac{1}{N} \sum_{j=1}^{N} \|y_j - f(x_j; \mathbf{W})\|_2 \tag{2}$$

The l_2-norm based loss function is defined in the Pdist layer and the back-propagation algorithm [27] is utilized to solve it. For clarity, the detailed structure of the deep feature extractor is shown in Table 1.

2.2. Local Linear Regression

It is noted that the deep neural network trained from above end-to-end strategy tries to apply an universal prediction function $f(x_j; \mathbf{W})$ to all test windows. Since the off-line trained model focuses on minimizing the mean prediction error across all training data, the prediction accuracy towards each individual test sample will reduce when its local distribution deviates from the global one in the feature space.

To address this issue, we develop a simple yet efficient local linear regression method, which adaptively updates the parametric prediction function for each test sample. As shown in Fig. 1 (b), for each given test window, we first search for its *k*NN from all IoU labeled training data in terms of their Euclidean distances of the deep features. Let z_k denote the feature vector of the kth nearest training sample for the test window x_j, and $\mathbf{Z} = \{z_1, z_2, \cdots, z_K\}$ is the observation matrix. Meanwhile, let $\mathbf{Y} = \{y_1, y_2, \cdots, y_K\}$ denote the ground-truth IoU vectors for the K local training samples. Then, the proposal quality prediction function is represented by

$$\mathbf{Y} = \mathbf{Z}^T \beta + \varepsilon \tag{3}$$

where β is the regression coefficient vector and ε is prediction error vector for the local training samples.

The optimization objective is to estimate β to minimize the mean prediction error on the *k*NN based local training data, i.e.,

$$\begin{aligned}\hat{\beta} &= \arg\min_{\beta \subset R^d} E[\|\mathbf{Z}^T\beta - \mathbf{Y}\|_2^2] \\ &= \arg\min_{\beta \subset R^d} E\Big[(\mathbf{Z}^T\beta - \mathbf{Y})^T(\mathbf{Z}^T\beta - \mathbf{Y})\Big] \\ &= \arg\min_{\beta \subset R^d} E\Big[\beta^T A\beta + c\beta + p\Big]\end{aligned} \tag{4}$$

where $\|\cdot\|_2$ denotes the l_2-norm operator, $A = \mathbf{Z}^T\mathbf{Z}$, $c = -2\mathbf{Y}^T\mathbf{Z}^T$, and $p = \mathbf{Y}^T\mathbf{Y}$ is a constant value.

The loss function in (4) formulates a standard quadratic programming problem, and the sequential minimal optimization algorithm [28] is employed to solve it in this paper.

3. EXPERIMENTAL RESULTS

3.1. Data Collection

Existing detection or recognition databases only provide the locations of the object-covered windows. There is no annotated samples including both the proposals and their associated IoU values. In this section, we firstly introduce a homogeneous sampling method to collect the IoU labeled training and testing windows. Given the manually labeled object locations, the random sampling may be the most straightforward method to generate the candidate windows and their associated IoU values. Because the foreground objects only cover limited regions in the whole image, most sampled windows will locate on the backgrounds, which induces significant IoU bias for the value of 0.

However, as discussed in Section I, the main task of the generic BPQA metric focuses on sorting out optimal candidate windows across various top-ranked proposals, which usually possess nonzero IoU values and deviate far from the randomly sampled windows. Another alternative method is to generate the candidate windows via existing proposal algorithms. While, in this way, the training data will be highly dependent on specific proposal generation process, which reduces the generality of the BPQA trained from them.

To balance the generality and diversity, we design a homogeneous sampling method to collect our training data. As shown in Fig. 2, we lay out five sampling windows around each manually annotated bounding box, where all sampling windows are required to share the *same* IoU value. The windows 1$\sim$4 just cover part of the object, which are located in the northeast, northwest, southeast and southwest directions. The window 5 could box the whole object but also involves part of the background. All windows follow the same aspect ratio with respect to the ground-truth bounding box. In this example, the IoU values of the five cropped windows are all 0.5. To cover diverse overlap conditions, we sample the candidate windows in nine IoU values, which range from 0.1 to 0.9 at the interval of 0.1. For each object-covered bounding box, there will be at most 45 sample windows. It is noted that the sample windows will not be collected into the training data when they go beyond the image boundary.

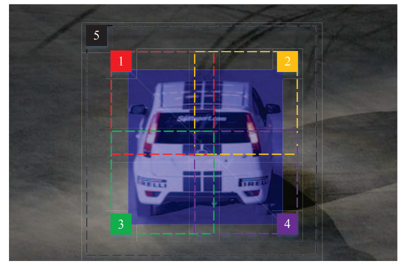


Fig. 2. The homogeneous sampling process for the windows with the same IoU values. The ground-truth bounding box is labeled by the light blue region, and each sampling window is represented by a dotted box with different colors. All windows follow the same aspect ratio with respect to the ground-truth bounding box.

All training images from PASCAL VOC2007 [8] are used for building the IoU labeled proposal evaluation dataset. Particularly, we carefully select 32798 cropped windows, which cover 20 categories of visual contents. To facilitate the evaluation and comparison of different BPQA metrics, we separate these IoU annotated windows into non-overlapped training and testing subsets just like [8]. Each subset takes over 50% of annotated samples.

3.2. Consistency Evaluation

We evaluate the performance of the proposed DORLLR method on the large-scale proposal evaluation dataset introduced in Section 3.1. It is noted that the proposed DORLLR metric is trained based the *k*NN of each test sample. Too few neighbors would easily cause the overfitting problem. On the contrary, too many neighbors would significantly increase the complexity of the training process. In our experiment, the number of neighbors is empirically set to 2500 to balance the regression accuracy and complexity. For comparison, four state-of-the-art BPQA algorithms are involved in this investigation, i.e., objectness (OB) [4], BINGS [17], Edge Boxes (EB) [16], and Region Proposal Network (RPN) [29]. To verify the superiority of the LLR with respect to the end-to-end strategy, the AlexNet [20] fine tuned on our proposal evaluation dataset is also tested here.

The rank-order based indice is used for measuring the consistency between the BPQA predictions and the ground-truth IoU values across all test samples, i.e., the Spearman's rank correlation coefficient (SRCC) [30, 31]. Particularly, a value that is more close to 1 indicates a better prediction performance.

The proposal quality prediction accuracy on 20 categories of the test set are reported in Table 2, respectively. For clarity, the best result across all BPQA metrics are highlighted by boldface for each category. It is seen that the proposed

Table 2. SRCC performance on large-scale proposal quality evaluation dataset

Meric	aeroplane	bicycle	bird	boat	bottle	bus	car	cat	chair	cow
BINGS	0.362	0.091	0.220	0.056	0.197	0.206	0.153	0.193	-0.037	0.115
EB	0.310	0.331	0.344	0.274	0.190	0.338	0.290	0.362	0.274	0.384
OB	0.149	0.127	0.122	0.123	0.109	0.150	0.106	0.159	0.030	0.094
RPN	0.286	0.213	0.290	-0.013	0.125	0.266	0.207	0.322	0.223	0.274
AlexNet	0.763	0.724	0.753	0.737	0.690	0.740	0.729	0.734	0.704	0.811
DORLLR	**0.795**	**0.762**	**0.771**	**0.781**	**0.729**	**0.761**	**0.767**	**0.778**	**0.746**	**0.835**
Meric	diningtable	dog	horse	motorbike	person	pottedplant	sheep	sofa	train	tvmonitor
BINGS	0.084	0.156	0.237	0.151	0.107	0.091	0.070	0.116	0.247	-0.003
EB	0.340	0.305	0.327	0.482	0.304	0.385	0.370	0.254	0.372	0.421
OB	0.141	0.120	0.100	0.144	0.101	0.177	0.280	0.072	0.036	0.022
RPN	0.220	0.280	0.365	0.353	0.289	0.210	0.319	0.233	0.261	0.341
AlexNet	0.692	0.671	0.766	0.746	0.715	0.773	0.790	0.580	0.734	0.817
DORLLR	**0.741**	**0.723**	**0.794**	**0.775**	**0.741**	**0.802**	**0.816**	**0.600**	**0.758**	**0.839**

Table 3. Overall SRCC performance on the whole test set

Metric	BINGS	EB	OB	RPN	AlexNet	DORLLR
SRCC	0.123	0.314	0.107	0.258	0.730	**0.760**

method delivers the best prediction accuracy and significantly outperforms the state-of-the-art four BPQA metrics including OB, EB, BINGS, and RPN, whose SRCC performances are smaller than 0.5 for all categories of visual contents. By contrast, the fine-tuned AlexNet and our DORLLR method achieve much higher SRCC performances, which are close to or larger than 0.7 for each category of test samples. In comparison with the AlexNet trained by end-to-end strategy, our local regression based method still shows superior accuracy across all categories of test samples. The overall prediction performance on the whole test set is also tabulated in Table 3 for different BPQA metrics, where the proposed DORLLR method achieves the highest SRCC value as well.

4. CONCLUSION

In this paper, we propose a highly efficient blind proposal quality assessment algorithm based on the deep objectness representation and local linear regression. Instead of exploring the objectness-aware clues from specific proposal generation algorithm, the proposed method learns the deep feature extractor from massive object-covered windows labeled by the IoU values. In the following, the query-specific quality evaluator is trained from the *k*NN of the test image by local linear regression. In comparison with the conventional proposal quality metrics, the proposed method does not depend on any proposal generation process, and works well on any given object-covered windows. In addition, due to the capability of updating the quality evaluator for different visual contents, the proposed method shows superior prediction accuracy over many state-of-the-art blind proposal quality metrics.

5. REFERENCES

[1] Kaiqi Huang, Liangsheng Wang, Tieniu Tan, and Steve Maybank, "A real-time object detecting and tracking system for outdoor night surveillance," *Pattern Recognition*, vol. 41, no. 1, pp. 432 – 444, 2008.

[2] Joseph Redmon, Santosh Kumar Divvala, Ross B. Girshick, and Ali Farhadi, "You only look once: Unified, real-time object detection," *CoRR*, vol. abs/1506.02640, 2015.

[3] Paul Viola and Michael J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, May 2004.

[4] B. Alexe, T. Deselaers, and V. Ferrari, "What is an object?," in *IEEE Conference on Computer Vision and Pattern Recognition*, June 2010, pp. 73–80.

[5] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2001, vol. 1, pp. I–511–I–518.

[6] Faruk Ahmed, Dany Tarlow, and Dhruv Batra, "Optimizing expected intersection-over-union with candidate-constrained crfs," in *IEEE International Conference on Computer Vision*, Washington, DC, USA, 2015, pp. 1850–1858.

[7] Sebastian Nowozin, "Optimal decisions from probabilistic models: The intersection-over-union case," in *IEEE Conference on Computer Vision and Pattern Recognition*, Washington, DC, USA, 2014, pp. 548–555.

[8] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL Visual Object Classes Challenge 2007 (VOC 2007) Results," http://www.pascal-network.org/challenges/VOC/voc2007/workshop/index.html.

[9] K. E. A. van de Sande, J. R. R. Uijlings, T. Gevers, and A. W. M. Smeulders, "Segmentation as selective search for object recognition," in *International Conference on Computer Vision*, Nov 2011, pp. 1879–1886.

[10] Santiago Manen, Matthieu Guillaumin, and Luc Van Gool, "Prime object proposals with randomized prim's algorithm," in *IEEE International Conference on Computer Vision*, 2013, pp. 2536–2543.

[11] Kai-Yueh Chang, Tyng-Luh Liu, Hwann-Tzong Chen, and Shang-Hong Lai, "Fusing generic objectness and visual saliency for salient object detection," in *International Conference on Computer Vision*, Nov 2011, pp. 914–921.

[12] E. Rahtu, J. Kannala, and M. Blaschko, "Learning a category independent object detection cascade," in *2011 International Conference on Computer Vision*, Nov 2011, pp. 1052–1059.

[13] Gökhan BakIr, Thomas Hofmann, Bernhard Schölkopf, Alexander J. Smola, Ben Taskar, and S. V. N Vishwanathan, *Support Vector Machine Learning for Interdependent and Structured Output Spaces*, pp. 85–103, MIT Press, 2007.

[14] Matthew B. Blaschko, Juho Kannala, and Esa Rahtu, *Non Maximal Suppression in Cascaded Ranking Models*, pp. 408–419, 2013.

[15] I. Endres and D. Hoiem, "Category-independent object proposals with diverse ranking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 2, pp. 222–234, Feb 2014.

[16] C. Lawrence Zitnick and Piotr Dollár, *Edge Boxes: Locating Object Proposals from Edges*, pp. 391–405, 2014.

[17] M. M. Cheng, Z. Zhang, W. Y. Lin, and P. Torr, "Bing: Binarized normed gradients for objectness estimation at 300fps," in *IEEE Conference on Computer Vision and Pattern Recognition*, June 2014, pp. 3286–3293.

[18] J. Carreira and C. Sminchisescu, "CPMC: Automatic object segmentation using constrained parametric min-cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 7, pp. 1312–1328, July 2012.

[19] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems*, 2015.

[20] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, pp. 1097–1105. 2012.

[21] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014.

[22] Kunihiko Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biological Cybernetics*, vol. 36, no. 4, pp. 193–202, 1980.

[23] Xavier Glorot, Antoine Bordes, and Yoshua Bengio, "Deep sparse rectifier neural networks.," in *Aistats*, 2011, vol. 15, p. 275.

[24] David H Hubel and Torsten N Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *The Journal of physiology*, vol. 160, no. 1, pp. 106–154, 1962.

[25] Dominik Scherer, Andreas Müller, and Sven Behnke, "Evaluation of pooling operations in convolutional architectures for object recognition," in *Proceedings of the 20th International Conference on Artificial Neural Networks: Part III*, 2010, ICANN'10, pp. 92–101.

[26] I. Arel, D. C. Rose, and T. P. Karnowski, "Deep machine learning - a new frontier in artificial intelligence research [research frontier]," *IEEE Computational Intelligence Magazine*, vol. 5, no. 4, pp. 13–18, Nov 2010.

[27] Raúl Rojas, *Neural Networks: A Systematic Introduction*, Springer-Verlag New York, Inc., New York, NY, USA, 1996.

[28] Francis R. Bach, Gert R. G. Lanckriet, and Michael I. Jordan, "Multiple kernel learning, conic duality, and the smo algorithm," in *Proceedings of the Twenty-first International Conference on Machine Learning*, 2004, ICML '04, p. 6.

[29] Ross Girshick, "Fast r-cnn," in *IEEE International Conference on Computer Vision*, 2015.

[30] Jean Dickinson Gibbons and Subhabrata Chakraborti, *Nonparametric Statistical Inference*, pp. 977–979, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[31] Q. Wu, H. Li, F. Meng, K. N. Ngan, B. Luo, C. Huang, and B. Zeng, "Blind image quality assessment based on multichannel feature fusion and label transfer," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 3, pp. 425–440, March 2016.