Hierarchical Adaptive Routing and Selection: A Novel Computational Model for Visual Attention Processing

Khodami M. Ahsan[1] · Hosseini S.M[2]
[1]Department of General Psychology, University of Padua, Italy · [2]Bayes Business School, City St George's, University of London, UK

Motivation

Current visual attention models inadequately [1] handle the routing problem in hierarchical processing, fail to maintain spatial precision across levels, struggle to separate overlapping neural activations [2], and fail to adequately model contextual influences [3].

Key Limitations

Classical saliency and deep attention models often (i) lack principled routing in hierarchies, (ii) lose spatial precision as receptive fields grow, (iii) suffer cross–talk under clutter, and (iv) treat context heuristically without guarantees. extcolorred.

Proposed Solution

HARS offers dynamic, activity-dependent routing across levels with context-aware normalization and edge-preserving refinement, maintaining spatial precision, reducing cross-talk, and enabling multiple foci in a tractable, testable pipeline.

Multi-Scale Feature Extraction

For an input image $I(x,y)$, we compute a set of feature maps

$$F_{i,s}(x,y) = \Phi_i(I(x,y), s)$$
$$F_{i,s}(x,y) = \sum_{u,v} I(x-u, y-v) \cdot K_i(u,v,s)$$
$$K_i(u,v,s,t) = K_i^0(u,v,s) \cdot A_i(u,v,s,t)$$
$$A_i(u,v,s,t) = \frac{1}{1 + \alpha_i \cdot \int_{t-\tau}^{t} |F_{i,s}(u,v,t')|\, dt'}$$

where Φ_i represents the operator for feature type i (such as intensity, color opponency, orientation, motion, etc.) at scale s, and $K_i(u,v,s)$ is the kernel for feature type i at scale s.

Contextual Integration and Normalization

To address the context problem identified by Tsotsos, we introduce an explicit contextual normalization mechanism

$$\hat{F}_{i,s}(x,y) = \frac{F_{i,s}(x,y) - \mu_{i,s}(x,y,\omega)}{\sigma_{i,s}(x,y,\omega) + \epsilon}$$

where $\mu_{i,s}(x,y,\omega)$ and $\sigma_{i,s}(x,y,\omega)$ represent the mean and standard deviation in a spatial neighborhood ω around position (x,y), and contextual is refined by statistics of the image

$$\widetilde{F_{i,s}}(x,y) = \widehat{F_{i,s}}(x,y) \cdot \left(1 + \beta_i \cdot \frac{|F_{i,s}(x,y) - \mu_{i,s}^{global}|}{\sigma_{i,s}^{global}}\right),$$

where $\mu_{i,s}^{global}$ and $\sigma_{i,s}^{global}$ are the global mean and standard deviation for feature type i at scale s, and β_i is a parameter controlling the influence of global distinctiveness.

Temporal Integration and Prediction

To account for the temporal dynamics of visual attention, we incorporate a predictive component that anticipates the future state of attended objects:

$$\hat{S}(x,y,t+\Delta t) = S(x,y,t) + \kappa \cdot \frac{\partial S(x,y,t)}{\partial t} \cdot \Delta t + \xi \cdot M(x,y,t) \cdot \Delta t,$$

where $\hat{S}$ is the predicted saliency, $\frac{\partial S(x,y,t)}{\partial t}$ represents the temporal derivative of saliency, $M(x,y,t)$ is a motion field extracted from the input, and κ and ξ are parameters controlling the influence of temporal change and motion respectively

Parameter Optimization and Model Adaptation

The HARS model contains numerous parameters that require proper tuning. Rather than setting these parameters arbitrarily, we propose an optimization framework that adjusts parameters based on human behavioral data

$$\Theta^* = \arg\min_{\Theta} \sum_{i=1}^{N_{data}} ||B_{human}(i) - B_{model}(i,\Theta)||^2,$$

where Θ represents the full parameter set of the model, $B_{human}(i)$ are human behavioral measurements (such as reaction times, detection rates, or eye movement patterns), and $B_{model}(i,\Theta)$ are the corresponding model predictions with parameter set Θ.

Hierarchical Integration with Recurrent Connectivity

To address the routing problem, At each level l of the hierarchy, feature maps are progressively combined into more complex representations

$$H_l(x,y,t) = \sum_{i,s} w_{i,s,l}(t) \cdot P_{l-1,l}(\widetilde{F_{i,s}}(x,y,t)),$$

where $P_{l-1,l}$ is a projection operator that maps features, and $w_{i,s,l}(t)$ are dynamic weights that control the contribution of each feature type and scale to the higher-level representation.

$$w_{i,s,l}(t) = w_{i,s,l}^0 + \Delta w_{i,s,l}^{bu}(t) + \Delta w_{i,s,l}^{td}(t),$$

The bottom-up weight adjustment $\Delta w_{i,s,l}^{bu}(t)$ is computed based on the global informativeness of each feature map

$$\Delta w_{i,s,l}^{bu}(t) = \gamma_{bu} \cdot \frac{\sum_{x,y} |\widetilde{F}_{i,s}(x,y,t)|}{\sum_{i',s'} \sum_{x,y} |\widetilde{F_{i',s'}}(x,y,t)|},$$

$$\text{top-down weight} = \Delta w_{i,s,l}^{td}(t) = \gamma_{td} \cdot \frac{\sum_{x,y} T_i(x,y,t) \cdot H_{l+1}(x',y',t-1)}{\sum_{i'} \sum_{x,y} T_{i'}(x,y,t) \cdot H_{l+1}(x',y',t-1)},$$

Selective Routing Mechanism

a selective routing mechanism that dynamically gates information flow through the hierarchy

$$G_l(x,y,t) = \sigma\left(\frac{S(P_{0,l}(x,y),t) - \theta_l(t)}{\tau_l}\right),$$

where G_l is a gating signal at level l, σ is the sigmoid function, $P_{0,l}$ maps input coordinates to level l coordinates, $\theta_l(t)$ is an adaptive threshold, and τ_l is a temperature parameter controlling the sharpness of the gating.

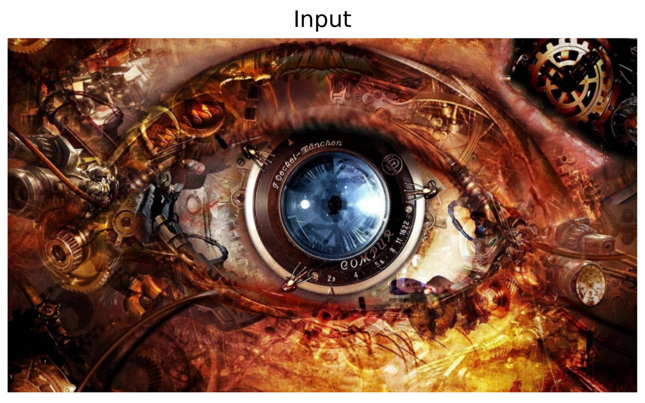
IOR with Object-Based Properties

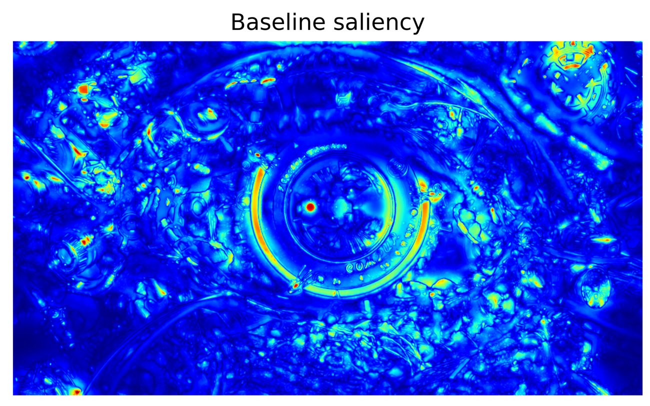
IOR mechanism that operates in both spatial and object-based coordinates

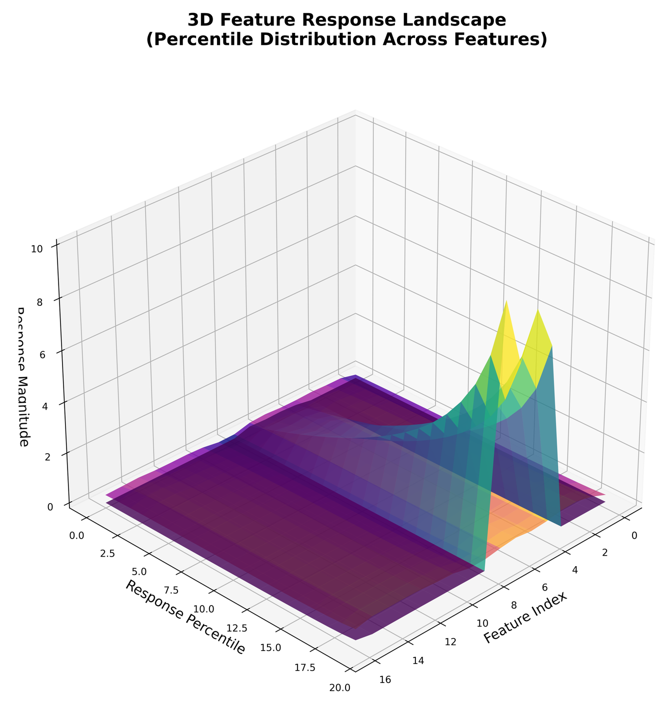
$$IOR(x,y,t) = \sum_{i=1}^{N_{attended}} \mathbf{exp}\left(-\frac{||[x,y]-[x_i,y_i]||^2}{2\sigma_{spatial}^2}\right) \cdot \mathbf{exp}\left(-\frac{t-t_i}{\tau_{IOR}}\right),$$

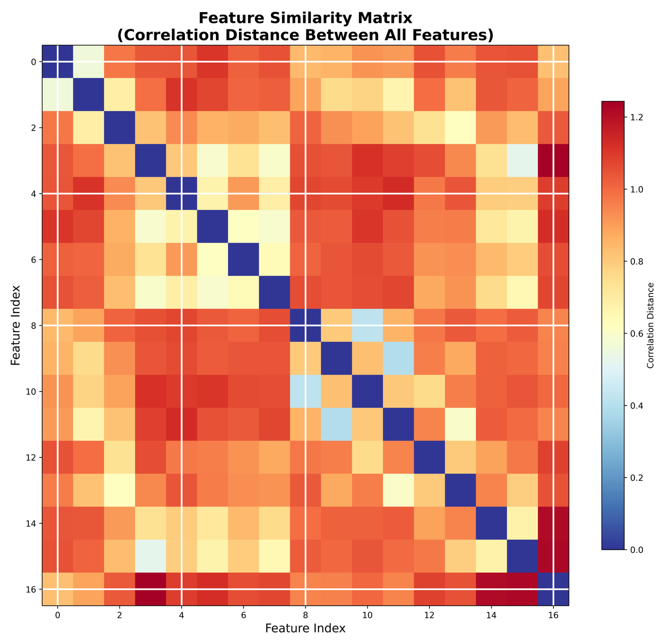
For object-based IOR, we extend this mechanism to include feature similarity:

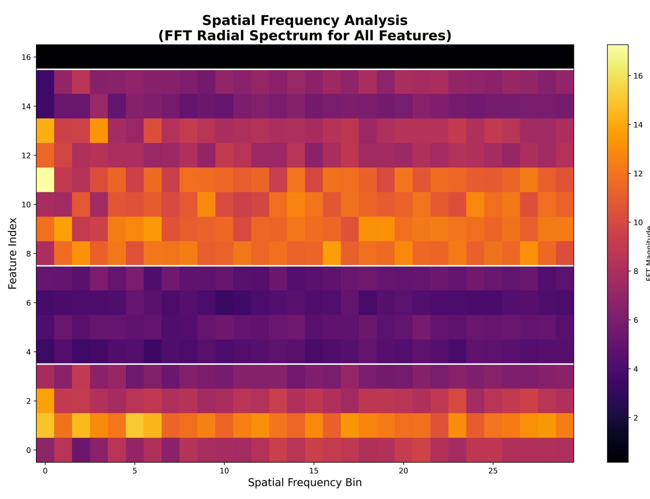
$$IOR_{obj}(x,y,t) = IOR(x,y,t) \cdot \sum_{i=1}^{N_{attended}} \mathbf{exp}\left(-\frac{||F(x,y,t) - F(x_i,y_i,t_i)||^2}{2\sigma_{feature}^2}\right),$$

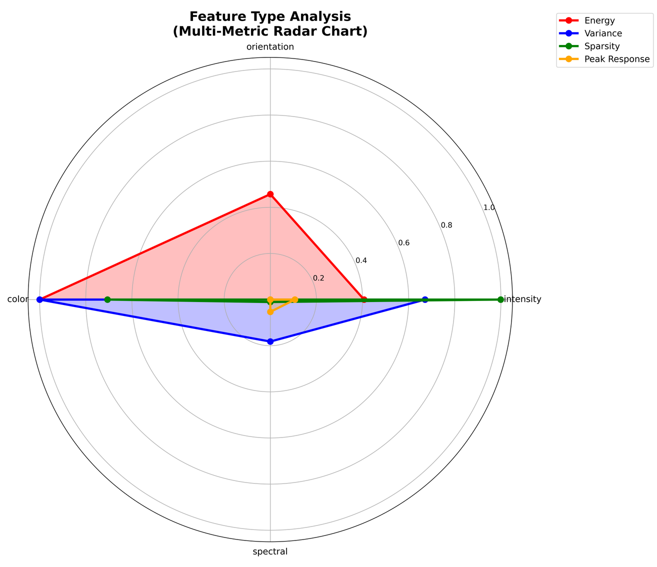












Acknowledgements

MAK has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 101034319 and from the European Union – NextGenerationEU

Bibliography

[1] Itti, L. (2005). Models of bottom-up attention and saliency. In Neurobiology of attention (pp. 576-582). Academic Press.

[2] Krasovskaya, S., & MacInnes, W. J. (2019). Salience models: A computational cognitive neuroscience review. Vision, 3(4), 56.

[3] Menon, V., & Uddin, L. Q. (2010). Saliency, switching, attention and control: a network model of insula function. Brain structure and function, 214(5), 655-667.

SCAN ME

Be in touch OR download my python packages

ETFORMAT:
EyeTracking conversion and Analysis Package

SCIA:
A Single-Case Design Analysis Package