Can Peripheral Representations Improve Clutter Metrics on Complex Scenes?

Arturo Deza
Dynamical Neuroscience
Institute for Collaborative Biotechnologies
UC Santa Barbara, CA, USA
`deza@dyns.ucsb.edu`

Miguel P. Eckstein
Psychological and Brain Sciences
Institute for Collaborative Biotechnologies
UC Santa Barbara, CA, USA
`eckstein@psych.ucsb.edu`

Abstract

Previous studies have proposed image-based clutter measures that correlate with human search times and/or eye movements. However, most models do not take into account the fact that the effects of clutter interact with the foveated nature of the human visual system: visual clutter further from the fovea has an increasing detrimental influence on perception. Here, we introduce a new foveated clutter model to predict the detrimental effects in target search utilizing a forced fixation search task. We use Feature Congestion (Rosenholtz et al.) as our non foveated clutter model, and we stack a peripheral architecture on top of Feature Congestion for our foveated model. We introduce the Peripheral Integration Feature Congestion (PIFC) coefficient, as a fundamental ingredient of our model that modulates clutter as a non-linear gain contingent on eccentricity. We show that Foveated Feature Congestion (FFC) clutter scores ($r(44) = -0.82 \pm 0.04, p < 0.0001$) correlate better with target detection (hit rate) than regular Feature Congestion ($r(44) = -0.19 \pm 0.13, p = 0.0774$) in forced fixation search; and we extend foveation to other clutter models showing stronger correlations in all cases. Thus, our model allows us to enrich clutter perception research by computing fixation specific clutter maps. Code for building peripheral representations is available[1].

1 Introduction

What is clutter? While it seems easy to make sense of a cluttered desk *vs* an uncluttered desk at a glance, it is hard to quantify clutter with a number. Is a cluttered desk, one stacked with papers? Or is an uncluttered desk, one that is more organized irrelevant of number of items? An important goal in clutter research has been to develop an image based computational model that outputs a quantitative measure that correlates with human perceptual behavior [19, 12, 24, 21]. Previous studies have created models that output global or regional metrics to measure clutter perception. Such measures are aimed to predict the influence of clutter on perception. However, one important aspect of human visual perception is that it is not space invariant: the fovea processes visual information with high spatial detail while regions away from the central fovea have access to lower spatial detail. Thus, the influence of clutter on perception can depend on the retinal location of the stimulus and such influences will likely interact with the information content in the stimulus.

The goal of the current paper is to develop a foveated clutter model that can successfully predict the interaction between retinal eccentricity and image content in modulating the influence of clutter on perceptual behavior. We introduce a foveated mechanism based on the peripheral architecture proposed by Freeman and Simoncelli [9] and stack it into a current clutter model (Feature Congestion [23, 24]) to generate a clutter map that arises from a calculation of information loss with retinal eccentricity but is multiplicatively modulated by the original unfoveated clutter score. The new

---

[1]*Piranhas* Toolkit: `https://github.com/ArturoDeza/Piranhas`

30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain.

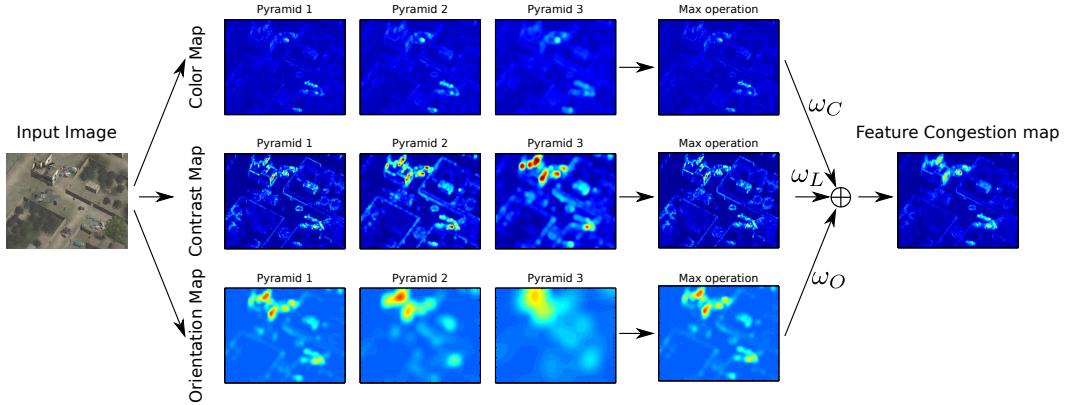



Figure 1: The Feature Congestion pipeline as explained in Rosenholtz *et al.* [24]. A color, contrast and orientation feature map for each spatial pyramid is extracted, and the max value of each is computed as the final feature map. The Feature Congestion map is then computed by a weighted sum over each feature map. The Feature Congestion score is the mean value of the map.

measure is evaluated in a gaze-contingent psychophysical experiment measuring target detection in complex scenes as a function of target retinal eccentricity. We show that the foveated clutter models that account for loss of information in the periphery correlates better with human target detection (hit rate) across retinal eccentricities than non-foveated models. Although the model is presented in the context of Feature Congestion, the framework can be extended to any previous or future clutter metrics that produce clutter scores that are computed from a global pixel-wise clutter map.

2 Previous Work

Previous studies have developed general measures of clutter computed for an entire image and do not consider the space-variant properties of the human visual system. Because our work seeks to model and assess the interaction between clutter and retinal location, experiments manipulating the eccentricity of a target while observers hold fixation (gaze contingent forced fixation) are most appropriate to evaluate the model. To our knowledge there has been no systematic evaluation of fixation dependent clutter models with forced fixation target detection in scenes. In this section, we will give an overview of state-of-the-art clutter models, metrics and evaluations.

2.1 Clutter Models

Feature Congestion: Feature Congestion, initially proposed by [23, 24] produces both a pixel-wise clutter score map as a well as a global clutter score for any input image or Region of Interest (ROI). Each clutter map is computed by combining a Color map in CIELab space, an orientation map [14], and a local contrast map at multiple scales through Gaussian Pyramids [5]. One of the main advantages Feature Congestion has is that each pixel-wise clutter score (Fig. 1) and global score can be computed in less than a second. Furthermore, this is one of the few models that can output a specific clutter score for any pixel or ROI in an image. This will be crucial for developing a foveated model as explained in Section 4.

Edge Density: Edge Density computes a ratio after applying an Edge Detector on the input image [19]. The final clutter score is the ratio of edges to total number of pixels present in the image. The intuition for this metric is straightforward: "the more edges, the more clutter" (due to objects for example).

Subband Entropy: The Subband Entropy model begins by computing N steerable pyramids [25] at K orientations across each channel from the input image in CIELab color space. Once each $N \times K$ subband is collected for each channel, the entropy for each oriented pyramid is computed pixelwise and they are averaged separately. Thus, Subband Entropy wishes to measure the entropy of each spatial frequency and oriented filter response of an image.

Scale Invariance: The Scale Invariant Clutter Model proposed by Farid and Bravo [4] uses graph-based segmentation [8] at multiple k scales. A scale invariant clutter representation is given by the power law coefficient that matches the decay of number of regions with the adjusted scale parameter.

2

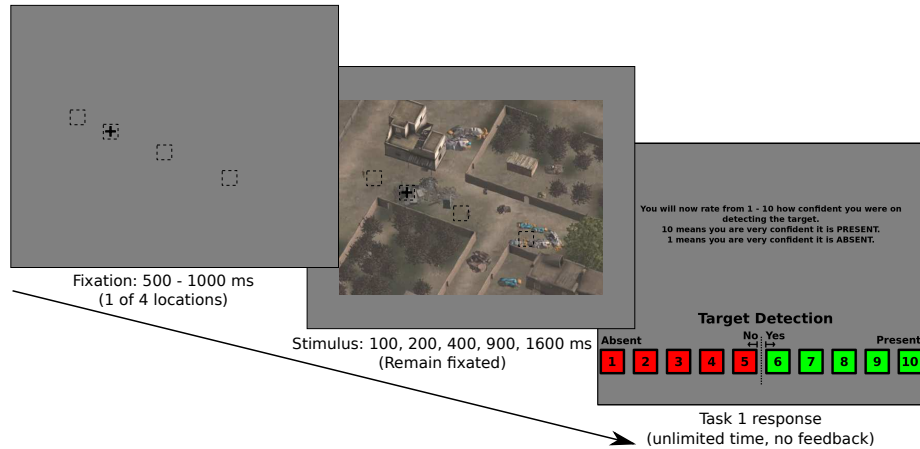



Figure 2: Experiment 1: Forced Fixation Search flow diagram. A naive observer begins by fixating the image at a location that is either 1, 4, 9 or 15 $\deg$ away from the target (the observer is not aware of the possible eccentricities). After fixating on the image for a variable amount of time (100, 200, 400, 900 or 1600 ms), the observer must make a decision on target detection.

ProtoObject Segmentation: ProtoObject Segmentation proposes an unsupervised metric for clutter scoring [27, 28]. The model begins by converting the image into HSV color space, and then proceeds to segment the image through superpixel segmentation [17, 16, 1]. After segmentation, mean-shift [11] is applied on all cluster (superpixel) medians to calculate the final amount of representative colors present in the image. Next, superpixels are merged with one another contingent on them being adjacent, and being assigned to the same mean-shift HSV cluster. The final score is a ratio between initial number of superpixels and final number of superpixels.

Crowding Model: The Crowding Model developed by van der Berg *et al.* [26] is the only model to have used losses in the periphery due to crowding as a clutter metric. It decomposes the image into 3 different scales in CIELab color space. It then produces 6 different orientation maps for each scale given the luminance channel; a contrast map is also obtained by difference of Gaussians on the previously mentioned channel. All feature maps are then pooled with Gaussian kernels that grow linearly with eccentricity, KL-divergence is then computed between the pre and post pooling feature maps to get information loss coefficients, all coefficients are averaged together to produce a final clutter score. We will discuss the differences of this model to ours in the Discussion (Section 5).

Texture Tiling Model: The Texture Tiling Model (TTM) is a recent perceptual model that accounts for losses in the periphery [22, 13] through psyhophysical experiments modelling visual search [7]: feature search, conjunction search, configuration search and asymmetric search. In essence, the Mongrels proposed by Rosenholtz *et al.* that simulate peripheral losses are very similar to the Metamers proposed by Freeman & Simoncelli [9]. We do not include comparisons to the TTM model since it requires additional psychophysics on the Mongrel versions of the images.

2.2 Clutter Metrics

Global Clutter Score: The most basic clutter metric used in clutter research is the original clutter score that every model computes over the entire image. Edge Density & Proto-Object Segmentation output a ratio, while Subband Entropy and Feature Congestion output a score. However, Feature Congestion is the only model that outputs a dense pixelwise clutter map before computing a global score (Fig. 1). Thus, we use Feature Congestion clutter maps for our foveated clutter model.

Clutter ROI: The second most used clutter metric is ROI (Region of Interest)-based, as shown in the work of Asher *et al.* [3]. This metric is of interest when an observer is engaging in target search, *vs* making a human judgement (Ex: "rate the clutter of the following scenes").

2.3 Clutter Evaluations

Human Clutter Judgements: Multiple studies of clutter, correlate their metrics with rankings/ratings of clutter provided by human participants. Ideally, if clutter model A is better than clutter model B, then the correlation of model scores and human rankings/ratings should be higher for model A than for model B. [28, 19, 26]

3

Response Time: Highly cluttered images will require more time for target search, hence more time to arrive to a decision of target present/absent. Under the previous assumption, a high correlation value between response time and clutter score are a good sign for a clutter model. [24, 4, 26, 3, 12]
Target Detection (Hit Rate, False Alarms, Performance): In general, when engaging in target search for a fixed amount of time across all trial conditions, an observer will have a lower hit rate and higher false alarm rate for a highly cluttered image than an uncluttered image. [24, 3, 12]

3 Methods & Experiments

3.1 Experiment 1: Forced Fixation Search

A total of 13 subjects participated in a Forced Fixation Search experiment where the goal was to detect a target in the subject's periphery and identify if there was a target (person) present or absent. Participants had variable amounts of time (100, 200, 400, 900, 1600 ms) to view each clip that was presented in a random order at a variable degree of eccentricities that the subjects were not aware of ($1 \deg$, $4 \deg$, $9 \deg$, $15 \deg$). They were then prompted with a Target Detection rating scale where they had to rate from a scale from 1-10 by clicking on a number reporting how confident they were on detecting the target. Participants have unlimited time for making their judgements, and they did not take more than 10 seconds per judgment. There was no response feedback after each trial. Trials were aborted when subjects broke fixation outside of a $1 \deg$ radius around the fixation cross.

Each subject did 12 sessions that consisted of 360 unique images. Every session also presented the images with aerial viewpoints from different vantage points (Example: session 1 had the target at 12 o'clock - North, while session 2 had the target at 3 o'clock - East). To control for any fixational biases, all subjects had a unique fixation point for every trial for the same eccentricity values. All images were rendered with variable levels of clutter. Each session took about an hour to complete. The target was of size $0.5 \deg \times 0.5 \deg$, $1 \deg \times 1 \deg$, $1.5 \deg \times 1.5 \deg$, depending on zoom level.

For our analysis, we only used the low zoom and 100 ms time condition since there was less ceiling effects across all eccentricities.

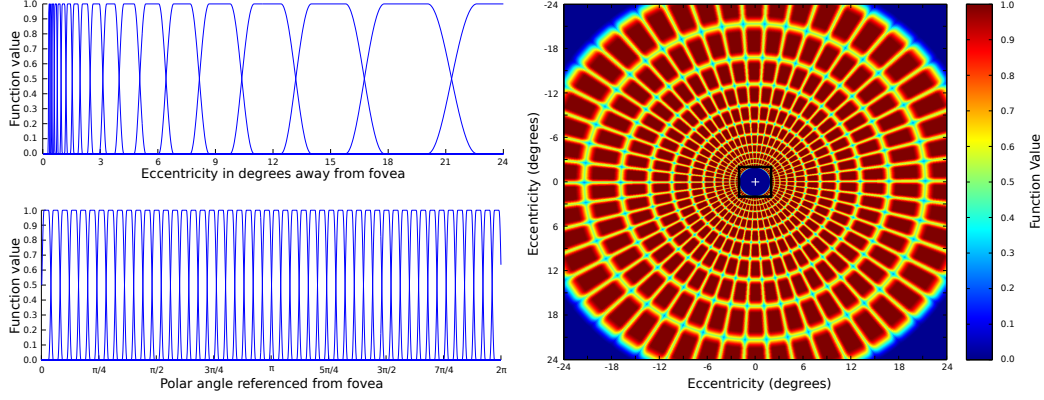
Stimuli Creation: A total of 273 videos were created each with a total duration of 120 seconds, where a 'birds eye' point-of-view camera rotated slowly around the center. While the video was in rotating motion, there was no relative motion between any parts of the video. From the original videos, a total of 360×4 different clips were created. Half of the clips were target present, while the other half were target absent. These short and slowly rotating clips were used instead of still images in our experiment, to simulate slow real movement from a pilot point of view. All clips were shown to participants in random order.

Apparatus: An EyeLink 1000 system (SR Research) was used to collect Eye Tracking data at a frequency of 1000Hz. Each participant was at a distance of 76 cm from a LCD screen on gamma display, so that each pixel subtended a visual angle of $0.022 \deg /\text{px}$. All video clips were rendered at 1024×760 pixels ($22.5 \deg \times 16.7 \deg$) and a frame rate of 24fps. Eye movements with velocity over $22 \deg /s$ and acceleration over $4000 \deg /s^2$ were qualified as saccades. Every trial began with a fixation cross, where each subject had to fixate the cross with a tolerance of $1 \deg$.

4 Foveated Feature Congestion

A regular Feature Congestion clutter score is computed by taking the mean of the Feature Congestion map of the image or of a target ROI [12]. We propose a Foveated Feature Congestion (FFC) model that outputs a score which takes into account two main terms: 1) a regular Feature Congestion (FC) score and 2) a Peripheral Integration Feature Congestion (PIFC) coefficient that accounts the lower spatial resolution of the visual periphery that are detrimental for target detection. The first term is independent of fixation, while the second term will act as a non-linear gain that will either reduce or amplify the clutter score depending on fixation distance from the target.

In this Section we will explain how to compute a PIFC, which will require creating a human-like peripheral architecture as explained in Section 4.1. We then present our Foveated Feature Congestion (FFC) clutter model in Section 4.2. Finally, we conclude by making a quantiative evaluation of the FFC (Section 4.3) in its ability to predict variations of target detectability across images and retinal eccentricity of the target.

4

(a) Top: $g_n(e)$ function. Bottom: $h_n(\theta)$ function.

(b) Peripheral Architecture.

Figure 3: Construction of a Peripheral Architecture *a la* Freeman & Simoncelli [9] using the functions described in Section 4.1 are shown in Fig. 3(a). The blue region in the center of Fig. 3(b), represents the fovea where all information is preserved. Outer regions (in red), represent different parts of the periphery at multiple eccentricities.

4.1 Creating a Peripheral Architecture

We used the *Piranhas* Toolkit to create a Freeman and Simoncelli [9] peripheral architecture. This biologically inspired model has been tested and used to model V1 and V2 responses in human and non-human primates with high precision for a variety of tasks [20, 10, 18, 2]. It is described by a set of pooling (linear) regions that increase in size with retinal eccentricity. Each pooling region is separable with respect to polar angle $h_n(\theta)$ and log eccentricity $g_n(e)$, as described in Eq. 2 and Eq. 3 respectively. These functions are multiplied for every angle and eccentricity (θ, e) and are plotted in log polar coordinates to create the peripheral architecture as seen in Fig. 3.

$$f(x) = \begin{cases} cos^2(\frac{\pi}{2}(\frac{x-(t_0-1)/2}{t_0})); & -(1+t_0)/2 < x \le (t_0-1)/2 \\ 1; & (t_0-1)/2 < x \le (1-t_0)/2 \\ -cos^2(\frac{\pi}{2}(\frac{x-(1+t_0)/2}{t_0})) + 1; & (1-t_0)/2 < x \le (1+t_0)/2 \end{cases} \quad (1)$$

$$h_n(\theta) = f\left(\frac{\theta - (w_\theta n + \frac{w_\theta(1-t_0)}{2})}{w_\theta}\right); w_\theta = \frac{2\pi}{N_\theta}; n = 0, ..., N_\theta - 1 \quad (2)$$

$$g_n(e) = f\left(\frac{\log(e) - [\log(e_0) + w_e(n+1)]}{w_e}\right); w_e = \frac{\log(e_r) - \log(e_0)}{N_e}; n = 0, ..., N_e - 1 \quad (3)$$

The parameters we used match a V1 architecture with a scale of $s = 0.25$, a visual radius of $e_r = 24 \deg$, a fovea of $2 \deg$, with $e_0 = 0.25 \deg$ [2], and $t_0 = 1/2$. The scale defines the number of eccentricities N_e, as well as the number of polar pooling regions N_θ from $\langle 0, 2\pi]$.

Although observers saw the original stimuli at $0.022 \deg$/pixel, with image size 1024×760; for modelling purposes: we rescaled all images to half their size so the peripheral architecture could fit all images under any fixation point. To preserve stimuli size in degrees after rescaling our images, our foveal model used an input value of $0.044 \deg$/pixel (twice the value of experimental settings). Resizing the image to half its size also allows the peripheral architecture to consume less CPU computation time and memory.

4.2 Creating a Foveated Feature Congestion Model

Intuitively, a foveated clutter model that takes into account target search should score very low when the target is in the fovea (near zero), and very high when the target is in the periphery. Thus, an observer should find a target without difficulty, achieving a near perfect hit rate in the fovea, yet the observer should have a lower hit rate in the periphery given crowding effects. Note that in the

---

[2]We remove regions with a radius smaller than the foveal radius, since there is no pooling in the fovea.

5

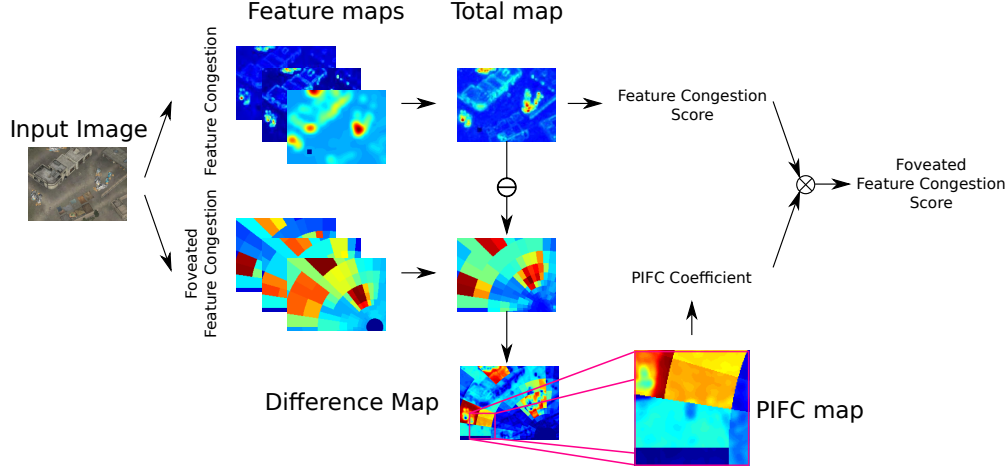



Figure 4: Foveated Feature Congestion flow diagram: In this example, the point of fixation is at $15\deg$ away from the target (bottom right corner of the input image). A Feature Congestion map of the image (top flow), and a Foveated Feature Congestion map (bottom flow) are created. The PIFC coefficient is computed around an ROI centered at the target (bottom flow; zoomed box). The Feature Congestion score is then multiplied by the PIFC coefficient, and the Foveated Feature Congestion score is returned. Sample PIFC's across eccentricities can be seen in the Supplementary Material.

periphery, not only should it be harder to detect a target, but it is also likely to confuse the target with another object or region affine in shape, size, texture and/or pixel value (false alarms). Under this assumption, we wish to modulate a clutter score (Feature Congestion) by a multiplicative factor, given the target and fixation location. We call this multiplicative term: the PIFC coefficient, which is defined over a $6\deg \times 6\deg$ ROI around the location of target t. The target itself was removed when processing the clutter maps since it indirectly contributes to the ROI clutter score [3]. The PIFC aims at quantifying the information loss around the target region due to peripheral processing.

To compute the PIFC, we use the before mentioned ROI, and calculate a mean difference from the foveated clutter map with respect to the original non-foveated clutter map. If the target is foveated, there should be little to no difference between a foveated map and the original map, thus setting the PIFC coefficient value to near zero. However, as the target is farther away from the fovea, the PIFC coefficient should be higher given pooling effects in the periphery. To create a foveated map, we use Feature Congestion and apply max pooling on each pooling region after the peripheral architecture has been stacked on top of the Feature Congestion map. Note that the FFC map values will depend on the fixation location as shown in Fig. 4. The PIFC map is the result of subtracting the foveated map from the unfoveated map in the ROI, and the score is a mean distance value between these two maps (we use L1-norm, L2-norm or KL-divergence). Computational details can be seen in Algorithm 1. Thus, we can resume our model in Eq. 4:

$$\text{FFC}_I^{f,t} = \text{FC}_I \times \text{PIFC}_{ROI(t)}^f \qquad (4)$$

where FC_I is the Feature Congestion score [24] of image I which is computed by the mean of the Feature Congestion map R_{FC}, and $\text{FFC}_I^{f,t}$ is the Foveated Feature Congestion score of the image I, depending on the point of fixation f and the location of the target t.

4.3 Foveated Feature Congestion Evaluation

A visualization of each image and its respective Hit Rate vs Clutter Score across both foveated and unfoveated models can be visualized in Fig 5. Qualitatively, it shows the importance of a PIFC weighting term to the total image clutter score when performing our forced fixation search experiment. Futhermore, a quantitative bootstrap correlation analysis comparing classic metrics (Image, Target, ROI) against foveal metrics (FFC_1, FFC_2 and FFC_3) shows that hit rate *vs* clutter scores are greater for those foveated models with a PIFC: Image: $(r(44) = -0.19 \pm 0.13, p = 0.0774)$, Target: $(r(44) = -0.03 \pm 0.14, p = 0.4204)$, ROI: $(r(44) = -0.25 \pm 0.14, p = 0.0392)$, FFC_1 (L1-norm):

6

Algorithm 1 Computation of Peripheral Integration Feature Congestion (PIFC) Coefficient

1: **procedure** COMPUTE PIFC OF ROI OF IMAGE I ON FIXATION f
2: Create a Peripheral Architecture $\mathbf{A} : (N_\theta, N_e)$
3: Offset image I in Peripheral Architecture by fixation $f : (f_x, f_y)$.
4: Compute Regular Feature Congestion (R_{FC}) map of image I
5: Set Peripheral Feature Congestion $(P_{FC}^f) \subset I\!R_+^2$ map to zero.
6: Copy Feature Congestion values in fovea r_0: $P_{FC}^f(r_0) = (R_{FC}(r_0))$
7:     **for** each pooling region r_i overlapping I, s.t. $1 \leq i \leq N_\theta \times N_e$ **do**
8:         Get Regular Feature Congestion (FC) values in r_i
9:         Set Peripheral FC value to max Regular FC value: $P_{FC}^f(r_i) = \max(R_{FC}(r_i))$
10:     **end for**
11: Crop PIFC map to ROI: $p_{FC}^f = P_{FC}^f(ROI)$
12: Crop FC map to ROI: $r_{FC} = R_{FC}(ROI)$
13: Choose Distance metric D between r_{FC} and p_{FC}^f map
14: Compute Coefficient = mean$(D(r_{FC}, p_{FC}^f))$
15: **return** Coefficient
16: **end procedure**

(a) Feature Congestion with image ID's (b) Foveated Feature Congestion with image ID's

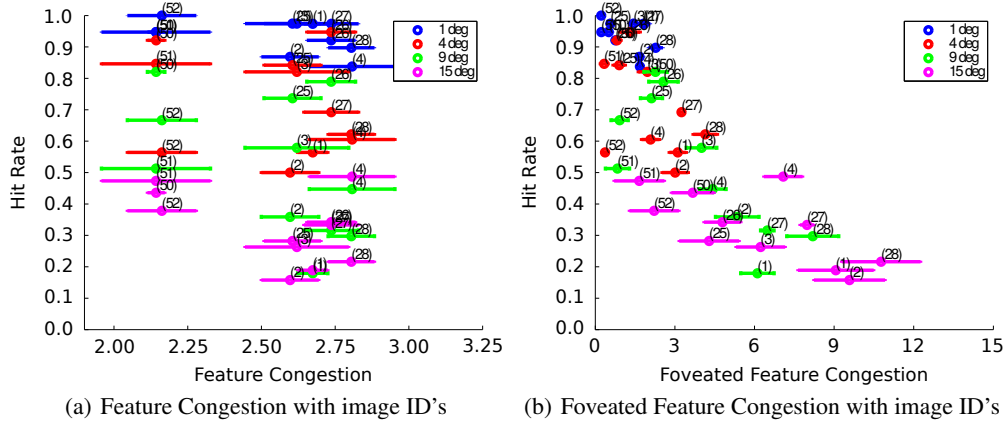



Figure 5: Fig. 5(a) shows the current limitations of global clutter metrics when engaging in Forced Fixation Search. The same image under different eccentricities has the same clutter score yet possess a different hit rate. Our proposed foveated model (Fig. 5(b)), compensates this difference through the PIFC coefficient, and modulates each clutter score depending on fixation distance from target.

$(r(44) = -0.82 \pm 0.04, p < 0.0001)$, FFC$_2$ (L2-norm): $(r(44) = -0.79 \pm 0.06, p < 0.0001)$, FFC$_3$ (KL-divergence): $(r(44) = -0.82 \pm 0.04, p < 0.0001)$.

Notice that there is no difference in correlations between using the L1-norm, L2-norm or KL-divergence distance for each model in terms of the correlation with hit rate. Table **??**(Supp. Mat.) also shows the highest correlation with a $6 \times 6 \deg$ ROI window across all metrics. Note that the same analysis can not be applied to false alarms, since it is indistinguishable to separate a false alarm at $1 \deg$ from $15 \deg$ (the target is not present, so there is no real eccentricity away from fixation). However as mentioned in the Methods section, fixation location for target absent trials in the experiment were placed assuming a location from its matching target present image. It is important that target present and absent fixations have the same distributions for each eccentricity.

5 Discussion

In general, images that have low Feature Congestion have less gain in PIFC coefficients as eccentricity increases. While images with high clutter have higher gain in PIFC coefficients. Consequently, the difference of FFC between different images increases nonlinearly with eccentricity, as observed in Fig. 6. This is our main contribution, as these differences in clutter score as a function of eccentricity do not exist for regular Feature Congestion, and these differences in scores should be able to correlate with human performance in target detection.

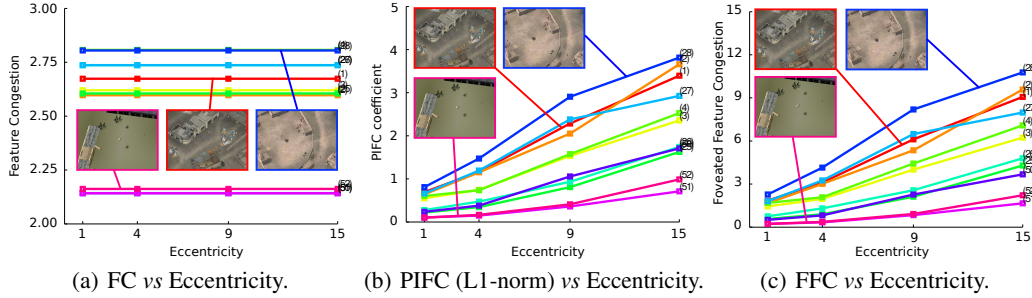
7

(a) FC *vs* Eccentricity.   (b) PIFC (L1-norm) *vs* Eccentricity.   (c) FFC *vs* Eccentricity.

Figure 6: Feature Congestion (FC) *vs* Foveated Feature Congestion (FFC). In Fig. 6(a) we see that clutter stays constant across different eccentricities for a forced fixation task. Our FFC model (Fig. 6(c)) enriches the FC model, by showing how clutter increases as a function of eccentricity through the PIFC in Fig. 6(b).

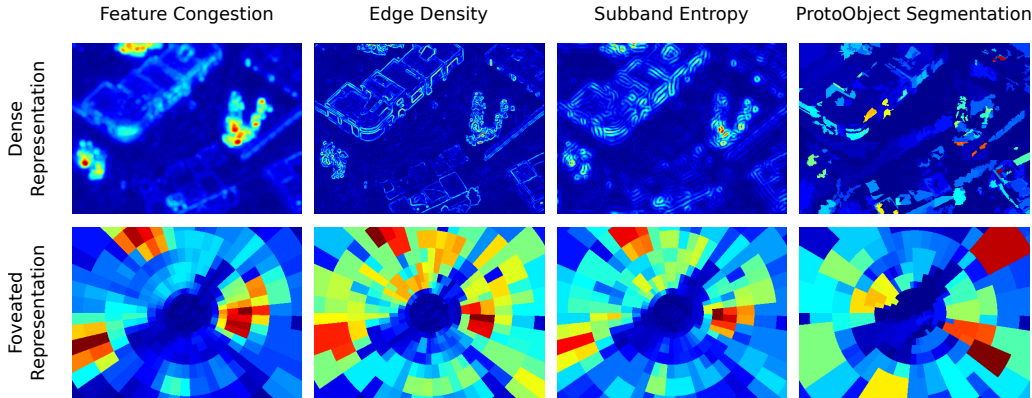



Figure 7: Dense and Foveated representations of multiple models assuming a center point of fixation.

Our model is also different from the van der Berg *et al.* [26] model since our peripheral architecture uses: a biologically inspired peripheral architecture with log polar regions that provide anisotropic pooling [15] rather than isotropic gaussian pooling as a linear function of eccentricity [26, 6]; we used region-based max pooling for each final feature map instead of pixel-based mean pooling (gaussians) per each scale (which allows for stronger differences); this final difference also makes our model computationally more efficient running at 700ms per image, *vs* 180s per image for the Crowding model ($\times 250$ speed up). A home-brewed Crowding Model applied to our forced fixation experiment resulted in a correlation of $(r(44) = -0.23 \pm 0.13, p = 0.0469)$, equivalent to using a non foveated metric such as regular Feature Congestion $(r(44) = -0.19 \pm 0.13, p = 0.0774)$.

We finally extended our model to create foveated(FoV) versions of Edge Density(ED) [19], Subband Entropy(SE) [25, 24] and ProtoObject Segmentation(PS) [28] showing that correlations for all foveated versions are stronger than non-foveated versions for the same task: $r_{ED} = -0.21$, $r_{ED+FoV} = -0.76$, $r_{SE} = -0.19$, $r_{SE+FoV} = -0.77$, $r_{PS} = -0.30$, but $r_{PS+FoV} = -0.74$. Note that the highest foveated correlation is FC: $r_{FC+FoV} = -0.82$, despite $r_{FC} = -0.19$ under a L1-norm loss of the PIFC. Feature Congestion has a dense representation, is more bio-inspired than the other models, and outperforms in the periphery. See Figure 7. An overview of creating dense and foveated versions for previously mentioned models can be seen in the Supp. Material.

6   Conclusion

In this paper we have introduced a peripheral architecture that shows detrimental effects of different eccentricities on target detection, that helps us model clutter for forced fixation experiments. We introduced a forced fixation experimental design for clutter research; we defined a biologically inspired peripheral architecture that pools features in V1; and we stacked the previously mentioned peripheral architecture on top of a Feature Congestion map to create a Foveated Feature Congestion (FFC) model – and we extended this pipeline to other clutter models. We showed that the FFC model better explains loss in target detection performance as a function of eccentricity through the introduction of the Peripheral Integration Feature Congestion coefficient which varies non linearly.

8

Acknowledgements

We would like to thank Miguel Lago and Aditya Jonnalagadda for useful proof-reads and revisions, as well as Mordechai Juni, N.C. Puneeth, and Emre Akbas for useful suggestions. This work has been sponsored by the U.S. Army Research Office and the Regents of the University of California, through Contract Number W911NF-09-0001 for the Institute for Collaborative Biotechnologies, and that the content of the information does not necessarily reflect the position or the policy of the Government or the Regents of the University of California, and no official endorsement should be inferred.

References

[1] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk. Slic superpixels. Technical report, 2010.

[2] E. Akbas and M. P. Eckstein. Object detection through exploration with a foveated visual field. *arXiv preprint arXiv:1408.0814*, 2014.

[3] M. F. Asher, D. J. Tolhurst, T. Troscianko, and I. D. Gilchrist. Regional effects of clutter on human target detection performance. *Journal of vision*, 13(5):25–25, 2013.

[4] M. J. Bravo and H. Farid. A scale invariant measure of clutter. *Journal of Vision*, 8(1):23–23, 2008.

[5] P. J. Burt and E. H. Adelson. The laplacian pyramid as a compact image code. *Communications, IEEE Transactions on*, 31(4):532–540, 1983.

[6] R. Dubey, C. S. Soon, and P.-J. B. Hsieh. A blurring based model of peripheral vision predicts visual search performances. *Journal of Vision*, 14(10):935–935, 2014.

[7] M. P. Eckstein. Visual search: A retrospective. *Journal of Vision*, 11(5):14–14, 2011.

[8] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59(2):167–181, 2004.

[9] J. Freeman and E. P. Simoncelli. Metamers of the ventral stream. *Nature neuroscience*, 14(9):1195–1201, 2011.

[10] J. Freeman, C. M. Ziemba, D. J. Heeger, E. P. Simoncelli, and J. A. Movshon. A functional and perceptual signature of the second visual area in primates. *Nature neuroscience*, 16(7):974–981, 2013.

[11] K. Fukunaga and L. D. Hostetler. The estimation of the gradient of a density function, with applications in pattern recognition. *Information Theory, IEEE Transactions on*, 21(1):32–40, 1975.

[12] J. M. Henderson, M. Chanceaux, and T. J. Smith. The influence of clutter on real-world scene search: Evidence from search efficiency and eye movements. *Journal of Vision*, 9(1):32–32, 2009.

[13] S. Keshvari and R. Rosenholtz. Pooling of continuous features provides a unifying account of crowding. *Journal of Vision*, 16(39), 2016.

[14] M. S. Landy and J. R. Bergen. Texture segregation and orientation gradient. *Vision research*, 31(4):679–691, 1991.

[15] D. M. Levi. Visual crowding. *Current Biology*, 21(18):R678–R679, 2011.

[16] A. Levinshtein, A. Stere, K. N. Kutulakos, D. J. Fleet, S. J. Dickinson, and K. Siddiqi. Turbopixels: Fast superpixels using geometric flows. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(12):2290–2297, 2009.

[17] M.-Y. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa. Entropy rate superpixel segmentation. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 2097–2104. IEEE, 2011.

[18] J. A. Movshon and E. P. Simoncelli. Representation of naturalistic image structure in the primate visual cortex. In *Cold Spring Harbor symposia on quantitative biology*, volume 79, pages 115–122. Cold Spring Harbor Laboratory Press, 2014.

[19] A. Oliva, M. L. Mack, M. Shrestha, and A. Peeper. Identifying the perceptual dimensions of visual complexity of scenes.

[20] J. Portilla and E. P. Simoncelli. A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision*, 40(1):49–70, 2000.

[21] R. Pramod and S. Arun. Do computational models differ systematically from human object perception? In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1601–1609, 2016.

[22] R. Rosenholtz, J. Huang, A. Raj, B. J. Balas, and L. Ilie. A summary statistic representation in peripheral vision explains visual search. *Journal of vision*, 12(4):14–14, 2012.

[23] R. Rosenholtz, Y. Li, J. Mansfield, and Z. Jin. Feature congestion: a measure of display clutter. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 761–770. ACM, 2005.

[24] R. Rosenholtz, Y. Li, and L. Nakano. Measuring visual clutter. *Journal of vision*, 7(2):17–17, 2007.

[25] E. P. Simoncelli and W. T. Freeman. The steerable pyramid: A flexible architecture for multi-scale derivative computation. In *icip*, page 3444. IEEE, 1995.

[26] R. van den Berg, F. W. Cornelissen, and J. B. Roerdink. A crowding model of visual clutter. *Journal of Vision*, 9(4):24–24, 2009.

[27] C.-P. Yu, W.-Y. Hua, D. Samaras, and G. Zelinsky. Modeling clutter perception using parametric proto-object partitioning. In *Advances in Neural Information Processing Systems*, pages 118–126, 2013.

[28] C.-P. Yu, D. Samaras, and G. J. Zelinsky. Modeling visual clutter perception using proto-object segmentation. *Journal of vision*, 14(7):4–4, 2014.

9