Analysis of scores, datasets, and models in visual saliency prediction

Ali Borji†  Hamed R. Tavakoli†  Dicky N. Sihite†  Laurent Itti†
† Department of Computer Science, University of Southern California, Los Angeles
† Center for Machine Vision Research, University of Oulu, Finland

Abstract

Significant recent progress has been made in developing high-quality saliency models. However, less effort has been undertaken on fair assessment of these models, over large standardized datasets and correctly addressing confounding factors. In this study, we pursue a critical and quantitative look at challenges (e.g., center-bias, map smoothing) in saliency modeling and the way they affect model accuracy. We quantitatively compare 32 state-of-the-art models (using the shuffled AUC score to discount center-bias) on 4 benchmark eye movement datasets, for prediction of human fixation locations and scanpath sequence. We also account for the role of map smoothing. We find that, although model rankings vary, some (e.g., AWS, LG, AIM, and HouNIPS) consistently outperform other models over all datasets. Some models work well for prediction of both fixation locations and scanpath sequence (e.g., Judd, GBVS). Our results show low prediction accuracy for models over emotional stimuli from the NUSEF dataset. Our last benchmark, for the first time, gauges the ability of models to decode the stimulus category from statistics of fixations, saccades, and model saliency values at fixated locations. In this test, ITTI and AIM models win over other models. Our benchmark provides a comprehensive high-level picture of the strengths and weaknesses of many popular models, and suggests future research directions in saliency modeling.

1. Introduction

A large number of models has been proposed for predicting where people look in scenes [1]. But due to a lack of an exhaustive coherent benchmarking system, to address several issues such as evaluation measures (e.g., at least 4 types of AUC measures have been used; supplement), center-bias, map characteristics (e.g., smoothing), and dataset bias, a lot of inconsistencies still exist. The discrepancy of results in previous works calls for a unified approach for gauging the progress in this field and for fair comparison of models.

Importance. Modeling visual saliency broadens our understanding of a highly complex cognitive behavior, which may lead to subsequent findings in other areas (object and scene recognition, visual search, etc.) [27][5]. It also benefits many engineering applications (e.g., object detection and segmentation, content-aware image re-targeting, image in-painting, visual tracking, image and video compression, crowds analysis and social gaming [2][6][24][37][30], determining the importance of objects in a scene [48, 44], memorability of image regions [49], and object recall [50].

Our contributions. We offer 3 main contributions: (1) discussing current challenges and directions in saliency modeling (evaluation metrics, dataset bias, model parameters, etc.) and proposing solutions, (2) comparing 32 models and their pros and cons in a unified quantitative framework over 4 widely-used datasets for fixation prediction (on classic and emotional stimuli) as well as scanpath prediction, and (3) stimuli/task decoding using saliency and fixation statistics. Hopefully, our study will open new directions and conversations and help better organize the saliency literature.

Previous benchmarks. Few attempts has been made for saliency model comparison, but their shortcomings have driven us to conduct a new benchmark considering the latest progress. Borji et al. [30] compared 36 models over 3 datasets. Judd et al. [41] compared 9 models over only 300 images. Some works have compared salient object detection and region-of-interest algorithms [42]. Some other benchmarks have compared models over applications such as image quality assessment [52]. While being very effective, previous comparisons have not correctly addressed all challenging parameters in model accuracy. For example, map smoothing which influences fixation prediction accuracy of models [43], or center-bias (tendency of humans to look towards the center of an image, such that a trivial model that predicts salience near the center may surpass other saliency models) [28, 25], have not been addressed in previous benchmarks. Here, we thoroughly investigate these shortcomings with additional comparison of models over scanpath sequences. We provide the latest update on saliency modeling, with the most comprehensive set of models, challenges/parameters, datasets, and measures.

2. Basic concepts and definitions

Here, we lay out the ground for the rest of the paper and explain some basic concepts of visual attention.

There is often confusion between saliency and atten-
3. Analysis of challenges and open problems

We discuss challenges that have emerged as more models have been proposed. These are open issues not only for research but also for performing fair model comparison.
Every model adds a central Gaussian. This adds Gaussian size and its weight as two additional parameters, (2) Collecting datasets with no CB. This is difficult since, even if we have an approach to uniformly distribute image content, viewing strategy still exists, and (3) Designing suitable metrics which we consider as the most reasonable approach.

**Evaluation metrics.** Traditionally, saliency models have been evaluated against eye movement datasets. In some cases, accuracy is whether one can predict what changes people will notice, or what they will remember or annotate [1]. We use three popular metrics for saliency evaluation: (1) Correlation Coefficient (CC) between a model \((s)\) and human \((h)\) saliency maps: \(\text{CC}(s,h) = \frac{\text{cov}(s,h)}{\sigma_s \sigma_h}\),

(2) Normalized Scanpath Saliency (NSS): the average of saliency values at n fixations in a normalized map \(\text{NSS} = \frac{1}{n} \sum_{i=1}^{n} \frac{s(x_i, y_i) - \mu_s}{\sigma_s}\) [51], and (3) Area Under the ROC Curve (AUC) where human fixations are considered as the positive set and some points from the image are uniformly chosen as the negative set. The saliency map is then treated as a binary classifier to separate the positive samples from negatives. By thresholding over this map and plotting true positive rate vs. false positive rate an ROC curve is achieved and its underneath area is calculated. Please see supplement for a subtle discussion of variations of AUC metrics. KL [9] and earth mover distance (EMD) [36] measures have also been used for model evaluation. Some studies have evaluated the sequence of fixations in scanpath [32, 31].

Fig. 2 shows analysis of how the above scores are affected by smoothness of the saliency map and possible center bias in the reference data. We generated some random eye fixations (sampled from a Gaussian distribution) and made a saliency map by convolving it with a Gaussian filter with variable sigma \(\sigma_1\). Shown in the 1st column, increasing \(\sigma_1\), reduces all 3 scores. Over AUC, however, the drop is moderate and the range is very small meaning as long as the hit rates are high, the AUC is high regardless of the false alarm rate [36].

Shown in the 2nd column, we placed a Gaussian at the center of the image and calculated the score again by varying the \(\sigma_2\) of the central Gaussian as well as \(\sigma_1\) of the Gaussian convolved with fixations (only for CC since for NSS and AUC, fixation positions are used). Increasing \(\sigma_2\), raises 3 scores up to the maximum match between Gaussian and MEP map and then drops or saturates. CC scores are raised by increasing \(\sigma_1\). Third column in Fig. 2 shows that by increasing the border size, scores reach a maximum and then drop, a similar effect to center-bias. These analyses show that smoothing the saliency maps and the size of the central Gaussian affect scores and should be accounted for fair model comparison. NSS score is more sensitive to smoothing. All of these scores suffer from center-bias.

Two other issues regarding scores are sensitivity to map normalization (a.k.a re-parameterization) and having well-defined bounds (and chance level). Some scores are invariant to continuous monotonic nonlinearity (e.g., KL) while some others are not (CC, NSS, and AUC). All scores are invariant to saliency map shifting and scaling. Some scores have well defined bounds (CC and AUC have lower and upper bounds) while some do not (KL and NSS; KL has lower bound and NSS has upper bound and chance level of 0).

A proper score for tackling CB is shuffled AUC (sAUC) [25] with the only difference to AUC being that instead of selecting negative points randomly, all fixations over other images are used as the negative set. This score is not affected by \(\sigma_2\) and border size in Fig. 2. sAUC value for a central Gaussian and a white map is near 0.5 (i.e., fixations from other images as the negative set [28]). When using the method in [25] (i.e., saliency from other images but at fixations of the current image), this type of AUC leads to the exact value of 0.5 for the central Gaussian (See Supp.).

**Features for saliency detection.** Traditionally, intensity, orientation, and color (in LAB and RGB spaces) have been used to derive saliency from static images. For dynamic scenes, flicker and motion features have been added. Furthermore, several other low-level features have been used to estimate saliency (size, depth, optical flow, etc.). High-level features (prior knowledge) such as faces [2], people [2], cars [2], symmetry [8], signs, and text [35] have been also incorporated. One challenge is detecting affective (emotional) features and semantic (high-level knowledge) scene properties (e.g., causality, action-influence) which have been suggested to be important in guiding attention (location and fixation duration) [4]. Models usually use all channels for all sorts of stimuli which makes them highly dependent on the false positive rates of employed feature

---

1 Note that in [36] and [2], AUC is calculated by thresholding the saliency map and then measuring hit rate which is different from what we (and also [25, 28, 43]) do by spreading random points on the image.
detectors (e.g., face or car detector). Since existing models use linear features, they render highly textured regions more salient. Non-linear features (e.g., famous egg in the nest or birthday candle images [25]) has been proposed but has not been fully implemented.

**Parameters.** Models often have several design parameters such as the number and type of filters, choice of non-linearities, within-feature and across-scale normalization schemes, smoothing, and center-bias. Properly tuning these parameters is important in fair model comparison and is perhaps best left for a model developer to optimize himself.

## 4. Saliency benchmark

We chose four widely-used datasets for model comparison: Toronto [10], NUSEF [4], MIT [2], and Kootstra [8]. Table 1 shows 30 models compared here. Additionally, we implemented two simple yet powerful models, to serve as baselines: **Gaussian Blob** (Gauss) and **Human inter-observer** (IO). Gaussian blob is simply a 2D Gaussian shape drawn at the center of the image; it is expected to predict human gaze well if such gaze is strongly clustered around the center. For a given stimulus, the human model outputs a map built by integrating fixations from other subjects than the one under test while they watched that same stimulus. The map is usually smoothed by convolving with a Gaussian filter. This inter-observer model is expected to provide an upper bound on prediction accuracy of computational models, to the extent that different humans may be the best predictors of each other. We resized saliency maps to the size of the original images onto which eye movements have been recorded. Please note that, besides models compared here, some other models may exist that might perform well (e.g., [37]), but are not publicly available or easily accessible. We leave them for future investigations.

We first measure how well a model performs at predicting where people look over static image eye movement datasets. We report results using sAUC score as it has several advantages over others. Results over other scores are shown in supplement. Note however, sAUC score alone is not the only criterion for our conclusions as it gives more credit to the off-center information and favors true positives more. Next, we compare models for their ability of predicting the saccade sequence. Our conclusions are based on the premise that if a model is good it should perform well over all configurations (i.e., score, datasets, and parameters).

**Predicting fixation locations:** Model scores and average ranks using sAUC over four datasets are shown in Table 1. We smoothed saliency map of each model by convolving it with a Gaussian kernel (Fig. 3). We then plotted the sAUC of each model over the range of standard deviations of the Gaussian kernel in width (from 0.01 to 0.13 in steps of 0.01) and calculated the maximum value over this range for each model. Compared with our rankings with original maps (supplement), now some models get a better score. Although the ranking order is not the same over four datasets, some general patterns are noticeable. The Gaussian model is the worst (not significantly better than chance) over all datasets as we expected. There is a significant difference between models and IO model. This difference is more profound over NUSEF and MIT datasets as they contain many stimuli with complex high-level concepts.

AWS model is significantly better than all other models followed by LG, AIM, Global Rarity, Torralba, HouCVPR, HouNIPS, SDSR, and Judd. Over the largest dataset (i.e, MIT), AWS, LG, AIM, Torralba performed better than other models. Over the NUSEF dataset, AIM, LG, Torralba, HouCVPR, HouNIPS models did the best. Kootstra, STB, ITTI (due to different normalization and map sparseness than ITTI98), and Marat ranked at the bottom. Interestingly, AWS model on the Kootstra dataset performs as good as the human IO. Our analyses show that CC, NSS, and AUC produce very high scores for the Gaussian; almost better than all models thanks to its center-preference (see supplement). Therefore, we do not recommend using them for saliency model comparison. Considering rankings sAUC, CC, and NSS scores, we noticed that models that performed well over sAUC are also ranked on top using other scores.

Fig. 4 shows model performance over stimulus categories of the NUSEF dataset for each model and average over all models. There is no significant difference over different categories of stimuli averaged over all models (Inset; See also supplement) although it seems that models perform better over face stimuli and the worst over portrait and nude (this pattern is more clear considering only

![Figure 3. A sample saliency map smoothed by convolving with a variable-size Gaussian kernel (for the AWS model over an image of the Toronto dataset).](image)

![Figure 4. Model performance over categories of the NUSEF dataset.](image)
myGauss = fspecial('gaussian', 50, 10) which was then normalized to $[0, 1]$. In principle, smoothing Gaussian should be about
although there are many images with symmetric objects in this dataset. Numbers are rounded to their closest value (See supplement for more accurate
and models. Next hardest in the NUSEF dataset. Note that symmetry model of Kootstra cannot compete with the other models over the Kootstra dataset
followed by half-wave rectification. This gives rise to much sparser saliency maps, which are more useful than the ITTI98 maps when trying to decide on
feature map [3]. This gives rise to smooth saliency maps, which tend to correlate better with noisy human eye movement data. In the ITTI model [33], the
smoothness (Fig. 3). See supplement for optimal sigma values of Gaussian kernel
VOCUS [6] are two implementations of the Itti et al. [24] model. Numbers are maximum shuffled AUC scores of models by optimizing the saliency map
smoothness (Fig. 3). See supplement for optimal sigma values of Gaussian kernel $\sigma$ where models take their maximums ($\sigma$ from 0 : 0.1 : 0.1 in image width). Model category belongs to one of these categories [30]: Cognitive (C), Bayesian (B), Decision-theoretic (D), Information-theoretic (I), Graphical (G), Spectral-analysis (S), Pattern-classification (P), Others (O). We observe lower performance for the STB model compared with either ITTI or ITT98 models over free-viewing datasets. Thus, it’s use instead on the ITTI model (e.g., for model-based behavioral studies) is not encouraged. We employ two different versions of the Itti et al. model: ITT98 and ITTI which correspond to different normalization schemes. In ITT98, each feature map’s contribution to the saliency map is weighted by the squared difference between the globally most active location and the average activity of all other local maxima in the feature map [3]. This gives rise to smooth saliency maps, which tend to correlate better with noisy human eye movement data. In the ITTI model [33], the spatial competition for saliency is much stronger, and implemented in each feature map as 10 rounds of convolution by a large difference-of-Gaussians followed by half-wave rectification. This gives rise to much sparser saliency maps, which are more useful than the ITTI98 maps when trying to decide on the single next location to look at (e.g., in machine vision and robotics applications). Kootstra dataset is the hardest one for humans (low IO agreement) and models. Next hardest in the NUSEF dataset. Note that symmetry model of Kootstra cannot compete with the other models over the Kootstra dataset although there are many images with symmetric objects in this dataset. Numbers are rounded to their closest value (See supplement for more accurate values). Borji et al. [30], results are on original images while here we report optimized results over smoothed images. In our experiments here we used the
Figure 5. sAUC scores over emotional images of the NUSEF dataset. Results on the right-hand table are the maxima over smoothing range. The best performing models). The AWS model did the best over all categories. HouNIPS, Judd, SDSR, Yan, and AIM also ranked at the top. Faces are often located at the center while nude and event stimuli are mostly off-center. Humans are more correlated for portrait, event, and nude stimuli. A separate analysis over the Kootstra dataset showed that models have difficulty in saliency detection over nature stimuli where there are less distinctive and salient objects (See supplement). This means that much progress remains to be done in saliency detection over stimuli containing conceptual stimuli (e.g., images containing interacting objects, actions such as grasping, living vs. non-living, object regions inside a bigger object i.e., faces, body parts, etc).

Behavioral studies have shown that affective stimuli influence the way we look at images. Humphrey et al. [45] showed that initial fixations were more likely to be on emotional objects than more visually salient neutral ones. Here we take a close look at model differences over emotional (affective) stimuli, using 287 images from the NUSEF when deciding on the single next location to look at (e.g., in machine vision and robotics applications). Kootstra dataset is the hardest one for humans (low IO agreement) and models. Next hardest in the NUSEF dataset. Note that symmetry model of Kootstra cannot compete with the other models over the Kootstra dataset although there are many images with symmetric objects in this dataset. Numbers are rounded to their closest value (See supplement for more accurate values). Borji et al. [30], results are on original images while here we report optimized results over smoothed images. In our experiments here we used the

Figure 6. Sample emotional images with positive, negative, and neutral emotional valence from NUSEF dataset along with saliency maps of the AWS model. Note that in some cases saliency misses fixations.
and quantizing scanpaths. Here, we first define a measure to quantify scanpath and then compare models in terms of their ability to generate saccades similar to human scanpath. In some applications it is desired to predict gaze one or multiple steps ahead (e.g., in advertising).

Algorithm 1 Scanpath evaluation

- **PHASE 1:** generating human scanpath and clusters
  
  **Input:** subjects’ fixations
  
  **Output:** fixation clusters & human scanpath strings.
  
  1. \(BW \leftarrow \text{List of all possible band-width values}\)
  2. **for each band-width \(BW\) do**
  3. Compute a clustering using the mean-shift algorithm
  4. Compute interaction between clusters
  5. **end for**
  6. Chose the band-width with the highest interaction rate
  7. Repeat the clustering for the chosen band-width and store clusters
  8. Assign a unique character to each cluster
  9. **for each subject do**
  10. Generate a string for scanpath by finding the closest cluster centers
  11. Store the string sequence
  12. **end for**
  13. Repeat the above process for all the images in the database

- **PHASE 2:** Model evaluation on a given test image
  
  **Input:** model generated saliency maps.
  
  **Output:** model’s overall score.
  
  1. **for each image do**
  2. Take a model saliency map
  3. Load the sequence strings and clusters corresponding to that image
  4. **for each subject do**
  5. Generate an equal length sequence running IOR
  6. Generate a string sequence by finding the closest cluster
  7. Assign a matching score to the two strings
  8. **end for**
  9. Save the average score over all subjects
  10. **end for**
  11. Report the average matching score over all images

Algorithm 1 shows our method for scanpath evaluation. Shortly, first for an image some clusters are derived from its human fixation map and then scanpath of each subject is coded into a string using these clusters. Then for a saliency map, an inhibition of return (IOR) mechanism is used to generate a sequence. Using Needleman-Wunsch [53] string matching algorithm, a model’s scanpath is compared against a subject’s scanpath and then the average score over all subjects is calculated (Supplement).

Results of scanpath prediction over Toronto and MIT datasets are shown in Fig. 7. Over the Toronto dataset, Tavakoli, GBVS, and Judd performed significantly better then the other models. Over the MIT dataset, GBVS, Judd, and AWS ranked on top. To investigate the relationship between fixation location and sequence prediction, we plotted these two scores versus each other (Fig. 8). There is a moderate linear correlation between these two scores. While there is not a single winning model using both scores, some models score well using both (e.g., AWS, HouNIPS, AIM, and SDDR). A random white noise map achieves the lowest sequence score (about 0.1 on Toronto and 0.05 on MIT) while a Gaussian model scores very well as it has some overlap with all subject sequences. Fig. 9.b shows sample images and their corresponding sequence measures.

Figure 7. Sequence prediction results over Toronto and MIT datasets.

5. Model comparison over applications

Since models are already good at predicting fixations, some new measures are necessary to draw distinctions among them. An application of saliency modeling is to decode task (illustrated by the classic study of Yarbus [63]), stimulus category, or different populations of human subjects (e.g., control vs. ADHD patients vs. normal subjects, elderly vs. young). Here, as an example we intend to decode the category of the stimulus from features augmented from statistics of fixations, saccades, and saliency at fixations. Note that here we intentionally do not use image features to keep our method general for different applications.

The task is to decode the stimulus category of a subset of images from the NUSEF dataset containing 5 categories over a total of 409 images (Event: 36, Face: 52, Nude: 20, Portrait: 123, Others: 178). For each image we extract a 872 dimensional vector. Features from statistics of fixations include: fixation histogram (256D) and fixation duration histogram (60 bins), from saccades: saccade length (50 bins), saccade orientation (36 bins), saccade duration (60 bins), saccade velocity (50 bins), and saccade slope (30 bins), from saliency: saliency map (300D), saliency histogram at fixated locations (10 bins), and Top 10 salient locations. To compute the histograms for a given image, we initially compute corresponding features (e.g., saccade velocity, etc.) for each observer and quantize the values into several bins. Later, the histogram of saccade statistic for an image is computed from all the observers and is L1 normalized. Fixation histogram is made by dividing the image into a grid pattern (16 × 16) and counting the number of fixations in each grid. Saliency-based features include histogram of saliency values at fixation points (10 bins), vectorized saliency map of size 20×15 (a vector of size 1×300) and coordinates of ten most salient points obtained by applying IOR to the saliency map (a vector of size 1×20).
Classification was performed using a multi-class SVM classifier using binary C-SVC support vector classification with RBF kernel. Initially, we picked 20 images from each scene category and optimized libsvm parameters (but not weights) using a 5-fold cross validation scheme (over 100 runs). Later, we categorized images by a 5-fold cross validation. This process was repeated 10 times. To cope with the imbalanced data over categories, we performed oversampling to reach the same number of samples in each category (i.e., number of samples in the largest class). For measuring Naive Bayes chance, we shuffled class labels and trained the SVM classifier again.

Fig. 10.a shows decoding results using saliency maps of four models and the MEP map. Using all features (and ITTI98 saliency maps), the best decoding accuracy is 0.512 followed by AIM (+ all features = 0.435). Note that the chance level here is 0.2 (Naive Bayes chance: SVM trained with shuffled labels is 0.36). MEP combined with other features scores 0.432 (performance alone is: 0.38) indicating different average fixation locations over stimulus categories. Fig. 10 shows performance of individual features in classification. Fixation histogram, saliency (here MEP), and saccade length histogram were more useful than the others. Confusion matrices of models show better accuracy over other, portrait, and face stimuli (See supplement).

6. Conclusions and future directions

Learned lessons: Our comparisons show that in general AWS, LG, HouNIPS, Judd, Rarity-G (smoothed version), AIM, and Torralba models performed higher than other models. Note that some recent high performing models (e.g., ∆QDCT [59], CovSal [60]) and benchmarking efforts were not considered here (e.g., [29, 58]). Analysis of scores shows that CC and NSS suffer from center-bias and conclusions should not be based just on them. The shuffled AUC score tackles the center-bias issue with Gaussian model nearly at the chance level (sAUC ≃ 0.5). This comes with the drawback of giving less credit to central fixations. Smoothing saliency maps affects performance and should be taken into account for fair model comparison. We found that some stimulus categories are harder for models (e.g., nature, nude, and portrait) which warrant more attention in future works. Finally, we showed that it is feasible to decode the stimulus category from a feature vector combined from saliency, saccade, and fixation statistics. This could be because of two reasons: similar saliency patterns across scenes of a category and/or systematic/semantic biases of fixations in each category. In this regard, it will also be interesting to test the feasibility of predicting whether a scene is natural or man-made from saliency and fixations.

We believe it is important to constantly measure the gap between the IO model and models to find out in which directions models lag behind human performance. Our analyses over IO model indicate: (1) Observers show poor agreement for complex cluttered scenes such as landscapes, buildings, and street views and (2) Observers show good agreement on scenes with people, animals, faces, and text.

Future directions: Our results show a small gap (but statistically significant) between the best models and human performance (e.g., AWS = 0.72, Human IO = 0.73 on Toronto dataset) indicating a significant progress in this field over the last few years. However, there is a danger that we are fitting ourselves to existing datasets working in a closed world. Thus we believe it is important to gather larger datasets, especially over new stimulus categories. Investigating cases where the current best models fail can help discover additional cues that guide human saccades and should be added to models (e.g., motion, vanishing point, actor and action, focus of expansion, signs, social cues including gaze and hand direction, image clutter, object importance, and text). For example, it has been shown that humans look at the center objects [61, 62]. Some saliency models implicitly emphasize the central parts of objects (e.g., [17]). Explicit central object-bias may lead to even higher fixation predic-

---

Our online challenge: https://sites.google.com/site/saliencyevaluation.
tion accuracies. Another direction is finding better ways to combine local and global saliences (i.e., regions with the same global saliency could have different local saliences).

We showed that, from statistics of fixations, saccades, and saliency at fixations, it is possible to decode the stimulus category. Future extensions would be designing standard and more challenging decoding scenarios and considering other features such as saccade sequence (scapath) information for better discriminating tasks, populations of patients, or stimulus category. Our results are particularly important, if we notice that a recent study [57] reported a failure in decoding subject’s task from fixations. Another promising research direction is designing better saliency evaluation scores which: (1) are able to better distinguish fixated vs. non-fixated locations, and (2) are able to discount confounding parameters such as center-bias.

This work was supported by NSF (CCF-1317433 and CMMI-1235539) and ARO (W911NF-11-1-0046 and W911NF-12-1-0433).

References

[14] T. Avraham, M. Lindenbaum. Esaliency (Extended Saliency): Meaningful saliency at fixations, it is possible to decode the stimulus category. Future extensions would be designing standard and more challenging decoding scenarios and consid- ering other features such as saccade sequence (scapath) information for better discriminating tasks, populations of patients, or stimulus category. Our results are particularly important, if we notice that a recent study [57] reported a failure in decoding subject’s task from fixations. Another promising research direction is designing better saliency evaluation scores which: (1) are able to better distinguish fixated vs. non-fixated locations, and (2) are able to discount confounding parameters such as center-bias.

This work was supported by NSF (CCF-1317433 and CMMI-1235539) and ARO (W911NF-11-1-0046 and W911NF-12-1-0433).

References