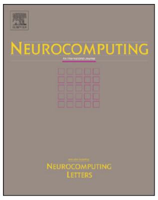
Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom



# Predicting task from eye movements: On the importance of spatial distribution, dynamics, and image features

Jonathan F.G. Boisvert, Neil D.B. Bruce [*]

*Department of Computer Science, University of Manitoba, 66 Chancellors Circle, Winnipeg, Manitoba, Canada*

ARTICLE INFO

*Article history:*
Received 10 November 2015
Received in revised form
13 May 2016
Accepted 15 May 2016
Communicated by Hantao Liu

*Keywords:*
Eye movements
Fixations
Task prediction
Features
Visual search
Yarbus

ABSTRACT

Yarbus' pioneering work in eye tracking has been influential to methodology and in demonstrating the apparent importance of task in eliciting different fixation patterns. There has been renewed interest in Yarbus' assertions on the importance of task in recent years, driven in part by a greater capability to apply quantitative methods to fixation data analysis. A number of recent research efforts have examined the extent to which an observer's task may be predicted from recorded fixation data. This body of recent work has raised a number of interesting questions, with some investigations calling for closer examination of the validity of Yarbus' claims, and subsequent efforts revealing some of the nuances involved in carrying out this type of analysis including both methodological, and data related considerations. In this paper, we present an overview of prior efforts in task prediction, and assess different types of statistics drawn from fixation data, or images in their ability to predict task from gaze. We also examine the extent to which relatively general task definitions (free-viewing, object-search, saliency-viewing, explicit saliency) may be predicted by spatial positioning of fixations, features co-located with fixation points, fixation dynamics and scene structure. This is accomplished in considering the data of Koehler et al. (2014) [30] affording a larger scale, and qualitatively different corpus of data for task prediction relative to existing efforts. Based on this analysis, we demonstrate that both spatial position, as well as local features are of value in distinguishing general task categories. The methods proposed provide a general framework for highlighting features that distinguish behavioural differences observed across visual tasks, and we relate new task prediction results in this paper to the body of prior work in this domain. Finally, we also comment on the value of task prediction and classification models in general in understanding facets of gaze behaviour.

## 1. Introduction

Early seminal work in analyzing eye movement patterns by Buswell [9] and Yarbus [58] remains influential in shaping scientific discourse addressing the role of task in gaze behaviour. This includes the notion that an observer's task may be inferred from examining their eye movements.

There is a rich literature on research efforts targeted at predicting human gaze patterns, in most instances in the absence of a specific task Bruce et al. [7], Shen and Zhao [41], Han et al. [24], Loyola et al. [35], Borji et al. [3], Kümmerer et al. [31], Wilming et al. [55].

Contemporary research efforts have further examined the interaction between task and fixations, in some cases directly considering Yarbus' claims about the predictability of task from fixations [10,48]. Some of this analysis has leveraged modern techniques in pattern classification to directly predict task from recorded fixation data. One prior effort modelled heavily on the methodology of Yarbus' experiments, [19], considers three different classifiers applied to aggregate eye-movement statistics for task prediction. None of these classifiers yielded performance above chance in considering the aggregate fixation features. More recent work considering the same data, but instead using low-resolution fixation density patterns [1], or other statistics achieved above chance performance by revising the feature set and inference methods used. Further improvements for the same data set have also been achieved in assuming knowledge of the participant viewing the image, or the specific image under consideration [29]. Accounting for covert attention in models of this variety has also lead to higher prediction accuracies [22]. The data provided by

[*] Correspondence to: E2-408 EITC, Department of Computer Science, University of Manitoba, Winnipeg, MB, Canada R3T 2N2.

*E-mail addresses:* umboisve@myumanitoba.ca (J.F.G. Boisvert), bruce@cs.umanitoba.ca (N.D.B. Bruce).

http://dx.doi.org/10.1016/j.neucom.2016.05.047
0925-2312/© 2016 Published by Elsevier B.V.

Greene et al. [19] has also inspired analysis of the possible impact of tasks requiring both visual processing and language processing [13]. There are evidently important individual differences that factor into viewing patterns [44], and such results also point to the notion that features such as fixation position and duration carry significant information about a viewers task.

In the context of attention modeling, there are many studies that address the relative contribution of *bottom-up* vs. *top-down* influences on the deployment of overt attention and fixation patterns. In the context of this paper, bottom-up refers to exogenous attention that is driven by properties of the visual stimulus, and independent of task or semantics. In referring to top-down processes, we refer to endogenous aspects of attention and guidance of gaze that are under executive control and may involve the influence of task directives, and working memory.

The spirit of Yarbus' assertions are closely related to an active overt attention process, drawing heavy influence from top-down bias. The critical importance of task in attention, and overt attention specifically is well supported as discussed in a number of recent studies and review papers addressing the relative importance of task and top-down cues [25,12,40,27,47,57,46,2].

While recent efforts leave little doubt that observed fixations provide a window into cognitive state or task directives, there is remaining benefit in examining task predictability from gaze statistics. One evident benefit from an applied perspective is the capability to infer task or intentions from fixations for applications in human machine interaction and human centric computing. Additional benefits of a more general nature arise from examining the ease with which different tasks may be distinguished based on gaze patterns, and in determining which features successfully discriminate between tasks. The degree of task separability has value in understanding similarity among visual and attentive mechanisms recruited for different tasks. Determining specific factors that distinguish tasks also points to targets for more careful examination in targeted experimental studies. In the domain of task prediction, prior work has focused heavily on confirming or denying the hypothesis that task may be predicted from gaze. For this reason, there has been a strong emphasis on prediction accuracy with less consideration of the role of different image or gaze related statistics in determining prediction performance. The work presented in this paper expands on the body of research involving task prediction in addressing the following important questions:

1. *Do relatively coarse grained tasks present distinct gaze statistics*? While several different sets of tasks have been examined in the literature, we consider task definitions that reside at a relatively general or coarse-grained level of abstraction. This serves to contribute to the growing body of efforts examining task prediction, and also to add diversity in the types of task sets considered.
2. *What methodological considerations are most critical to drawing value from efforts in task prediction*? If the goal of task prediction is to achieve something beyond confirming or denying Yarbus' assertions, there is value in ensuring that methodology allows for analysis beyond comparing prediction accuracies. We therefore employ methods for task prediction that are amenable to considering the relative importance of associated gaze and image related statistics, and discuss additional considerations of importance at the level of methodological details.
3. *Which gaze statistics are most important*? In considering a set of relatively coarse grained task directives and choosing suitable methods, we aim to establish which gaze statistics or image derived features seem to diverge most across different task definitions. This provides insight into information represented within different types of gaze or image related statistics.
4. *What does task prediction tell us (and not tell us) about vision*? Task prediction establishes that different tasks may be distinguished on the basis of gaze statistics. Results presented in this paper further reveal the relative importance of different types of features for the coarse-grained tasks examined in this paper, and also within existing studies. However, it is important to address the limitations on what studies in task prediction are able to convey about human vision. An additional goal of this paper is therefore to establish what benefits and limitations exist in examining task specific behaviour within a task prediction paradigm.

These four questions form the core of motivation for the work presented in this paper, and its novelty. First, we consider a very different set of task defined gaze data than has been considered previously, providing some new observations about challenges in this problem domain. More importantly though, we also explore in detail methodological considerations for approaching work in predicting $X$ from gaze, where $X$ might correspond to an assigned task, affective response to images or environment, or any other measurable factors. This is achieved in highlighting the importance of considering different types of features, and associated subsets while also applying predictive methods that offer feedback on the relative value of said features.

The balance of the paper is structured as follows: in Section 2 we present a survey of studies that emphasize task prediction highlighting differences in the set of tasks considered, methods and accuracies achieved across these studies. Section 3 presents the experimental methods that are exercised in this paper. This includes further details on the dataset, and types of features considered for task categorization. Following this, in Section 4, we present the relative classification performance that is achieved in considering the spatial distribution of fixations, local features at fixated locations, fixation dynamics and global scene structure. This analysis considers different conditions, including pooled fixation data across all observers, as well as fixations for single observers subject to different methods of partitioning the image set. Various combinations of 4-way, 3-way and binary classifications are considered where appropriate to shed further light on factors that separate tasks. We discuss the broader implications of this analysis in Section 5, including limitations and possible fruitful directions forward. Finally, Section 6 summarizes important results from this paper in addressing the role of task in observed fixation behaviour.

## 2. Prior work in task prediction

There are a variety of recent studies that consider this problem with direct reference to Yarbus' work, or specifically involving a classification paradigm for assessing the predictability of task from fixation data. DeAngelus and Pelz [15] re-examined Yarbus' work, including tools, methods, and implications of Yarbus' findings. They also replicated Yarbus' original experiment using updated methods and a larger pool of participants and paintings for fixation recording. Their results demonstrated patterns consistent with Yarbus' data for Repin's painting using modern eye tracking devices, while restricting observations to a shorter time course. Castelhano et al. [10] showed that an observer's task (object search and memorization) influences eye movement behaviour at the level of fixation durations and saccade amplitudes, specifically at the level of aggregate eye movement measures rather than individual fixation or saccade statistics. In the case of the memorization task, a larger area of the image was fixated and while the average fixation duration did not vary significantly between tasks, certain areas were re-fixated (approximately) increasing their total fixation duration. These results were observed in viewing photographs of natural scenes for an object search task and a scene memorization task. Mills et al. [36] also examined the impact of task on spatio-temporal fixation statistics with similar findings. A different task set was used in this case with free-viewing and *pleasantness* tasks added. The memorization task focused on the scene rather than objects, and the search task involved finding an “N” or “Z” that had been added to the scene, rather than a contextually appropriate object. While many of the results of the Mills et al. [36] study mirror those of Castelhano et al. [10], the average fixation duration varied more between tasks, specifically in the first few seconds of observation. The authors attributed this discrepancy to differences in task design or details of how results were analyzed. Tatler et al. [48] provide a detailed examination of Yarbus' body of work and the historical context surrounding it. Following this biographical overview an experiment was performed using a photograph of Yarbus. Results demonstrated that task influences the features fixated in viewing faces, and also that the importance of task in viewing extends to simpler types of visual stimuli.

Greene et al. [19] present a study suggesting that the relation between task and fixation behaviour may be overstated based on an inability to predict task from the fixation data captured for Yarbus style experiments. The feature vector used in this paper was composed of summary statistics of the observers fixations: (1) number of fixations, (2) the mean fixation duration, (3) mean saccade amplitude, and (4) percent of image covered by fixations assuming a one-degree fovea. These statistics had been used in previous work on scanpath analysis [36,10]. Also considered, was the proportion of fixation duration on various regions of interest: (5) faces, (6) human bodies, and (7) objects. The results from the Greene et al. [19] study demonstrated the ability to identify both the image and observer identity but an inability to predict the corresponding task. Subsequent efforts using this same data have demonstrated that while this presents a challenging classification task, above chance scores are possible with careful feature selection [2], division of data [29], or using Probability Density Functions and Hidden Markov Models to model dynamics of scan-paths and latent contributions of covert attention [22].

Borji and Itti [2] re-considered the Greene et al. [19] data, but chose to include low-resolution fixation density patterns as part of the evaluated feature set. This resulted in above chance performance of 34.14% accuracy across 4 tasks (25% chance-level). They also conducted a second experiment, mirroring Yarbus' original 7 task experiment, with prediction results of 24.21% when considering 7 tasks (14.29% chance-level). Their improvement of accuracy in the inference problem by nearly 10% above chance was achieved principally by revising the feature set and inference methods used. Kanan et al. [29] approached the Greene et al. [19] data by first reproducing the original experiment with similar results. To preserve the temporal information for eye movements a Fisher Kernel Learning (FKL) algorithm was used, that allowed the variable number of time-series statistics to be compressed into a single feature vector. They also performed task prediction using two within-subject prediction experiments. For the first this involved leave-one-out cross-validation using 19 of the 20 trials to train the SVM, and testing with the remaining single trial, this was repeated for each possible leave-one-out combination (20 total). In the second condition they trained the SVM using 4 of the 5 trials for each task, and used the remaining 4 trials (1 per task) as a test set, repeating this for each combination (625 total). Due to the increased number of training and testing trials to consider in the second approach, the number of states in the FKL was decreased from 10 to 5. The first approach resulted in a prediction accuracy of 52.9% (25% chance-level) using the FKL algorithm while the second approach returned a lower 34.1% (25% chance-level) accuracy.

In more recent work Haji-Abolhassani and Clark [22] focused their attention on the Greene et al. [19] data, having previously demonstrated good prediction accuracies on synthetic images [20,21]. An important contribution of their work is the use of Hidden Markov Models to estimate hidden state information from observed fixations to simulate the role of covert attention. The authors noted that in many cases the Centre of Gaze (COG) did not match the Focus of Attention (FOA). This was motivated by the observation that fixations did not always land on task specific objects, but observers where shown to be aware of the “overlooked” objects. To account for covert attention the authors used a Gaussian Mixture Model to capture task relevant spatial positions and define the observation likelihoods via probability density functions. Probability densities correspond to different states derived from task-relevant objects/regions. To apply this to the Greene et al. [19] data they used k-means clustering on the aggregated fixations for each image-task pair, to generate a set of regions likely to be task-relevant. Using this approach they were able to achieve a prediction accuracy of 59.64% (chance: 25%)

Coco and Keller [13] also revisited the Greene et al. [19] data, hypothesizing that the tasks considered may have required only visual processing rather than including other cognitive modalities such as language processing. By their hypothesis, similar processing requirements should result in similar strategies for allocation of attention, leading to harder inferences. To verify that tasks requiring different cognitive processes can be classified with greater ease, Coco and Keller [13] carried out an alternative experiment. Tasks chosen for the experiment were: visual search, object naming and scene description. Each of these tasks required a different mixture of visual and language processing. Each task was shown to a different group of participants, 25 for search and scene description each and 24 for object naming. The authors used the same 7 features as Greene et al. [19] as well as another set of 15 corresponding to temporal fixation measures. Along with in-depth analysis of the recorded eye movement features across the various tasks, the authors also trained 3 different types of regression models to predict the observers tasks: multinomial regression, least-square angle regression, and support vector machines. Their accuracy averaged over all 3 tasks was 81% (33.33% chance-level). They also produced classification accuracy of 76% using only the features from Greene et al. [19].

In contrast, the data presented by Henderson et al. [26], includes 196 natural images and 140 images of text, and 4 tasks consisting of search, memorization, reading and *pseudo-reading*. Task performance in this case was approximately 80% likely owing in part to the distinct nature of the chosen tasks. In addition, the mix of natural scenes and text involves very different stimuli, implying the possibility of stimulus driven differences as opposed to principally task driven differences [37].

Bulling et al. [8] have presented a system to infer high-level contextual cues called EyeContext. Four participants were fitted with mobile eye tracking equipment consisting of 5 electrodes centred around the right eye, and were tasked with self-annotating various cues encountered during the day. The 4 categories participants were asked to track were: “*social* (*interacting with somebody vs. no interaction*), *cognitive* (*concentrated work vs. zleisure*), *physical* (*physically active vs. not active*), *and spatial* (*inside vs. outside a building*)”. The eye movements recorded by the electrodes were encoded into fixed-length words composed of symbols, with a saccade to the left represented by the character ‘L’ and diagonal right represented by a ‘B’. A sum total of 42.5 h of data was collected across the 4 participants. They trained a string kernel SVM with 70% of the data and tested classification using the remaining 30%, using a 5-fold cross-validation. The mean precision and recall obtained were 76.8% and 85.5% respectively. This research illustrates a novel way that task inference and eye

Please cite this article as: J.F.G. Boisvert, N.D.B. Bruce, Predicting task from eye movements: On the importance of spatial distribution, dynamics, and image features, Neurocomputing (2016), http://dx.doi.org/10.1016/j.neucom.2016.05.047

tracking can be approached as this set of experiments more closely resembles real-life tasks and conditions.

A related effort considers the problem of recognizing the type of document an observer is currently reading. Kunze et al. [32] performed an experiment involving 8 participants situated in 5 different environments and reading 5 different document types. This is implicitly a task inference prediction as the different documents types are likely to elicit behaviours typical of highly learned document dependent strategies for reading. To track eye movements in various environments, observers were fitted with eye tracking glasses. In each case, 10 min of document reading was recorded and subsequently divided into 1 min windows, as input to a decision tree (C4.5/J48). On a per-observer basis the decision tree achieved a accuracy of 99%, whereas average accuracy was 74% when observer identity was unknown and all observer data was pooled. Prediction accuracy increased to 90% for this latter condition when using majority voting over the entire 10 min of captured data. As noted by the authors this research could be used in reading assistance and logging of reading activities, and may also be expanded to other non-reading tasks.

Cerf et al. [11] investigated the power of saliency maps to predict which image an observer fixated using only their scanpath data. While they achieved good results by combining the scanpaths from all observers, superior results were achieved when individual observers were factored into the prediction. They also proposed a metric to quantify the “Decodability” of datasets from certain feature sets (e.g. scanpaths), which may allow for the clustering of observers based on the features appealing to each individual. One of the uses of such features involves separating samples from special patient populations from those of control observers. This was investigated by Tseng et al. [51] who distinguished children with attention deficit hyperactivity disorder (ADHD) or fetal alcohol spectrum disorder (FASD) from a control group using a variety of features with an emphasis on saliency-based features. They also found that elderly patients with Parkinson's disease (PD) could be identified using primarily oculomotor related measurements. In related work, Jones and Klin [28] studied infants from the age of 2 to 24 months and noted that infants that were later diagnosed with autism spectrum disorders (ASDs) exhibited a mean decline in fixations of caretakers eyes between 2 and 6 months of age.

In this paper we employ Random Forests [4] for classification, and the details associated with this process are further explained in Section 3. The reason for choosing Random Forests for classification in our work, is the capacity to identify the value of different features in successfully predicting the observer's task. This presents the possibility to identify important, and sometimes subtle differences in behaviour corresponding to different tasks, from a vantage point defined by the features that are considered. Random Forests are also employed by Sugano et al. [42] to predict which of two images presented side-by-side on a computer monitor is preferred by an observer, based on eye movements. Their experiment was split into two phases: In the first phase, 11 individual observers were shown 80 pairs of image with no instruction. This was followed by 400 pairs of images with instructions to explicitly choose a preference between the two via a manual key press. Finally, the original 80 pairs were again shown but with the preference instructions. A total of 25 features were computed from fixations and saccades and used to train the Random Forests. The mean accuracy was 73% (50% chance) when tested on the 80 image pairs with explicit preference instructions. When testing on the 80 image pairs with no instructions (free viewing) accuracy was only 61% (50% chance).

Important characteristics of efforts involving task prediction from gaze are summarized in Table 1.

## 3. Experimental methods

In the following, we outline the details of the data considered in our analysis, classification methods and feature types considered for task prediction. In short, the data considered focuses on relatively general task directives related to free viewing, object driven viewing, or saliency driven viewing. Features include the spatial distribution of fixations, local oriented edge structure/contrast at fixated regions, holistic scene structure, and fixation specific statistics (e.g. saccade amplitudes). While a number of different classifiers have been examined, we have primarily considered Random Forest based bootstrap aggregation [4] in our analysis. The reasons for this, details of features considered, and the specifics of the dataset are further discussed in the remainder of this section.

### 3.1. Fixation data

In this paper we examine the data from Koehler et al. [30] which includes fixation data for a set of 800 images associated with a number of different tasks. For each of the 800 images, each participant was asked to perform 1 of 3 viewing tasks (free viewing; object search; saliency viewing) or an explicit judgement task that required identifying the most salient location in the image through manual selection. A minimum of 19 (19–22) participants performed each viewing task and 100 participants performed the explicit judgment task given that each sample from participants yields only one observation per image for this task. In the free viewing task, observers were instructed to freely view the image with no further direction given. For the Object Search task the participants were given the name of an object to find in the scene, with the object present in 50% of cases shown. In the Saliency Search task observers were asked to judge whether the left or right half of a scene contained the most salient region. Finally in the explicit judgement task, observers were asked to use the computer mouse to click on what they believed was the most salient point in each image. Fixation data was collected using an EyeLink 1000 System at 250 Hz. For more detail on the precise methods, the reader is referred to the original experimental description [30]. Representative samples of images from this dataset, and accompanying heatmaps showing fixation (or click) distribution across observers is shown in Fig. 1. Each exemplar image has 4 associated heatmaps corresponding to explicit judgement (top left), free viewing (top right), object search (bottom left) and saliency viewing (bottom right) respectively. In practice, an ideal visualization of task differences might involve decomposition of the data according to temporal windows (e.g. first second, first 3 seconds etc.). Given that this dataset is based on a relatively small number of observers, and short time course for viewing, such visualizations result in a distribution of gaze points that is sufficiently sparse that it does not lend itself well to visualization in the form of a heatmap for shorter time windows. With that said, as results in the later parts of this paper reveal, dynamics of fixations may be one factor that is especially variable as a function of task, and therefore there is great potential to garner understanding of task relevant differences based on temporal slicing of data, provided the volume of data permits this.

Fixation data for all three of the viewing tasks, and the explicit judgement data is used in our analysis. The explicit judgment data was only used in one of the prediction experiments presented in this paper, since the distinct nature of this data and absence of spatial fixation bias makes discriminating this task from the gaze based tasks relatively easy. In addition, given only one data-point per image for the explicit judgement case, the only natural comparison using this data is in aggregating data across all participants. Therefore, for the aggregate classification case, the fixation

**Table 1**
Past task-prediction contributions.

| Authors | Obs. | Size | Tasks | Features | Methods | Performance |
|---|---|---|---|---|---|---|
| Henderson et al. [26] | 12 | ⋄ 196 images ⋄ 140 texts | ⋄ Search ⋄ Memory ⋄ Reading ⋄ Pseudo-reading | ⋄ Eye movement measures | ⋄ Multivariate pattern analysis | 68–80% chance:25% |
| Haji-Abolhassani and Clark [23] (Ex1) | 6 | 180 images | ⋄ Counting 6 object types | ⋄ Gaussian mixture model capturing task relevant spatial positions | ⋄ Hidden Markov models | 71.5–88.5% Indv task predictions |
| Haji-Abolhassani and Clark [23] (Ex2) | 6 | 26 images | ⋄ Spelling 3 letter words | ⋄ Gaussian mixture model capturing task relevant spatial positions | ⋄ Hidden Markov models | 75.8–87.7% |
| Lethaus et al. [33] | 10 | 70 km simulated driving | ⋄ Overtaking ⋄ Following ⋄ Overtaking mult. | ⋄ Fixation time distribution across 4/5 zones during 5/10 s window | ⋄ Artificial neural net ⋄ Bayesian net ⋄ Naive Bayes classifier | Real-time:85% 1 s delay:90% |
| Bulling et al. [8] | 4 | Recording of workday (∼10 h) | ⋄ Social ⋄ Cognitive ⋄ Physical ⋄ Spatial | ⋄ Eye movement encodings | ⋄ SVM | Precision:76.8% Recall:85.5% chance:25% |
| Kunze et al. [32] | 8 | 5 books | ⋄ 5 document types | ⋄ Saccade direction counts, mean & variance ⋄ 95% quartile distance ⋄ Slope over fixations | ⋄ Decision tree | Independent:74% Dependent:99% chance:20% |
| Greene et al. [19] | 16 | 64 images | ⋄ Memory ⋄ Decade ⋄ People ⋄ Wealth | ⋄ Eye movement statistics | ⋄ Linear discriminant ⋄ Correlation methods ⋄ SVM | 25.9% chance:25% |
| Kanan et al. [29] | 16 | 64 images | ⋄ Memory ⋄ Decade ⋄ People ⋄ Wealth | ⋄ Eye movement statistics ⋄ Cartesian coordinates | ⋄ SVM | Within subject: 37.9% chance:25% |
| Haji-Abolhassani and Clark [23] | 16 | 64 images | ⋄ Memory ⋄ Decade ⋄ People ⋄ Wealth | ⋄ GMM for task relevant positions determined by fixation clusters (k-means) | ⋄ Hidden Markov models | 59.64% chance:25% |
| Borji and Itti [1] (Ex1) | 16 | 64 images | ⋄ Memory ⋄ Decade ⋄ People ⋄ Wealth | ⋄ Eye movement Statistics ⋄ Spatial density | ⋄ KNN ⋄ RUSBoost | 34.14% chance:25% |
| Borji and Itti [1] (Ex2) | 21 | 15 images | ⋄ Yarbus' original 7 tasks | ⋄ Eye movement statistics ⋄ Spatial density | ⋄ K-NN ⋄ RUSBoost | 24.21% chance:14.29% |
| Coco and Keller [13] (Ex2) | 24 | 24 images | ⋄ Visual search ⋄ Object-naming ⋄ Scene description | ⋄ Eye movement statistics | ⋄ Multinomial regression ⋄ Least-square angle regression ⋄ SVM | 81% chance:33.33% |
| Sugano et al. [42] | 11 | 480 img. pairs | ⋄ Free view ⋄ Preference | ⋄ Fixation & Saccade measure statistics | ⋄ Random Forests | 61% and 73% chance:50% |
| Boisvert and Bruce (this paper) (Agg) | 19 | 800 images | ⋄ Free view ⋄ Object search ⋄ Saliency ⋄ Explicit sal. | ⋄ Spatial density | ⋄ Random Forests | 69.63% chance:25% |
| Boisvert and Bruce (this paper) (Indv) | 19 | 800 images | ⋄ Free view ⋄ Object search ⋄ Saliency | ⋄ Spatial density ⋄ HOGs ⋄ LM filters ⋄ Gist | ⋄ Random Forests | 56.37% chance:33.33% 76.93% (binary task prediction) |

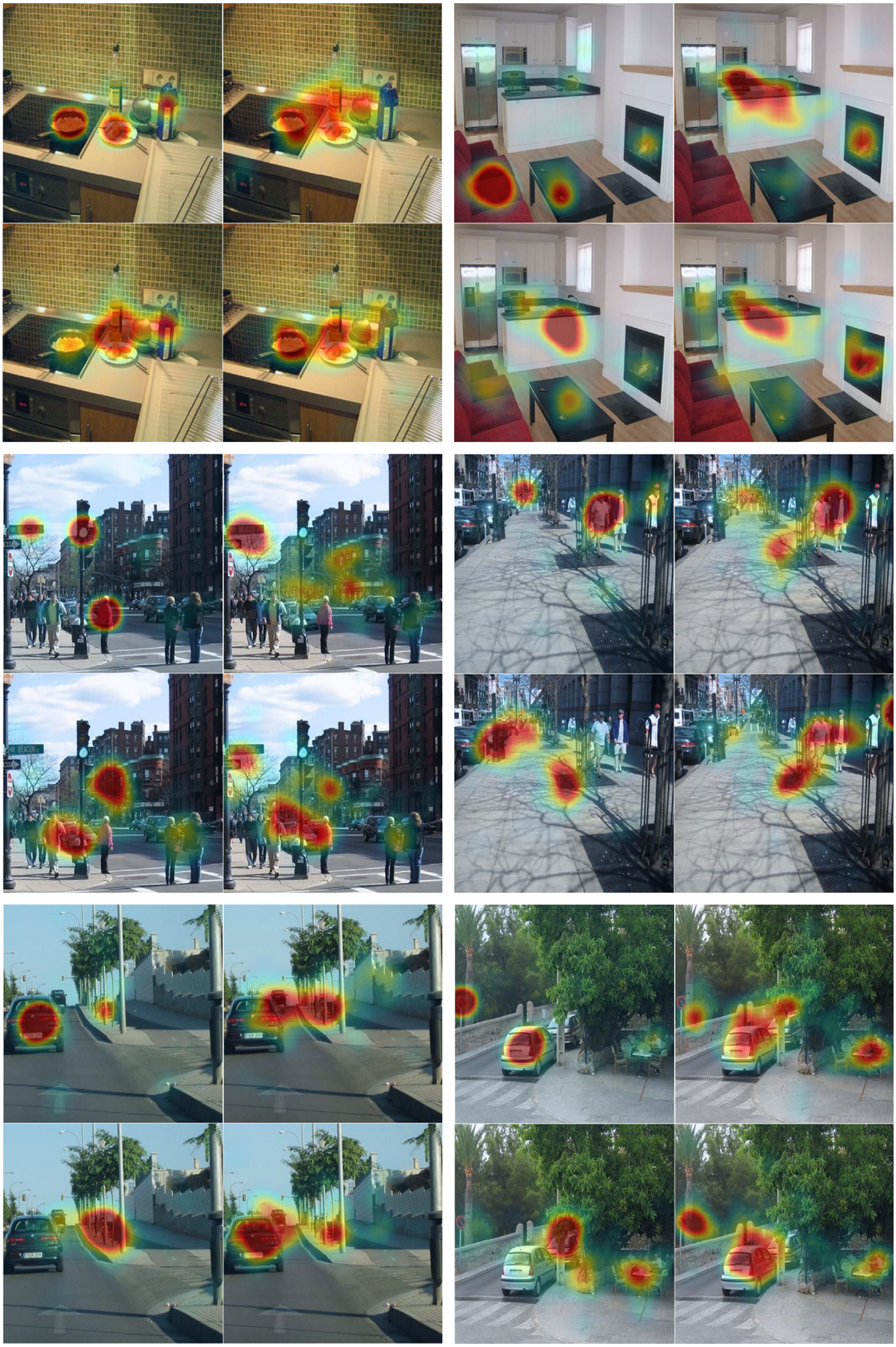


**Fig. 1.** A representative sample of images from the Koehler et al. dataset [30]. A heatmap is superimposed on each image showing the distribution of fixations (or clicks for explicit judgement) across each task. Images and their corresponding heatmaps for each set of images correspond to: explicit judgement (top left), free viewing (top right), object search (bottom left), saliency viewing (bottom right).

Please cite this article as: J.F.G. Boisvert, N.D.B. Bruce, Predicting task from eye movements: On the importance of spatial distribution, dynamics, and image features, Neurocomputing (2016), http://dx.doi.org/10.1016/j.neucom.2016.05.047

data for each image/task combination is pooled across all observers. In contrast, the *individual classification* cases that make up most of the analysis in this paper consider each image-participant-task sample as a separate case. It is important to note that in tests of classification performance, our results consider explicitly the role of prior exposure to the test image set during training providing additional observations relevant to task prediction in a broad sense.

### 3.2. *Features*

A central goal of the current work lies in evaluating both the role of spatial position (or density) of fixations in task prediction, but also the diagnostic value of specific features or image structure at fixated locations. We have also examined whether scene structure combined with the spatial distribution of fixation patterns might provide a useful diagnostic. To this end, we have selected a number of features for testing task classification performance. Each type of feature is considered in isolation and also in various combinations with other types of features. With respect to choice of features, we have attempted to minimize the complexity of features to facilitate the basic understanding of factors driving gaze that may be drawn out of the analysis. This includes features that characterize local, and global image structure (e.g. edge content and configurations), overall density of fixations, and gaze related statistics. With that said, it is important to note that one contribution of this paper is a methodological framework for analyzing modes of viewing that are elicited by different tasks. For this reason, there is no reason that more complex features might be considered. One natural possibility in this respect is computational measures of visual saliency. We have intentionally excluded saliency from consideration given that interaction between saliency and this data has been examined by Koehler et al. [30] (albeit with a different objective), and some tasks have strong ties to saliency. In the later part of this paper, we also touch on the apparent performance of semantically relevant patterns, and features motivated by modern methods in deep learning will no doubt bear fruit in understanding task relevant factors that contribute to observed gaze patterns.

#### 3.2.1. *Fixation density map (density)*

An alternative to considering raw fixation positions given their relative sparsity (in pixel terms) is considering a more continuous density map derived from the raw fixations. This may be produced by way of convolution of the fixation map with a Gaussian profile [5], and/or sub-sampling to produce a coarse-grained continuous spatial density map for task prediction [2]. The first set of features we have considered are spatial densities of fixations. We represent this quantity by generating a density map of the fixations on the $405 \times 405$ pixel images. A continuous density map is produced by convolving the image with a 2D Gaussian envelope with standard deviation corresponding to 1 degree of visual angle (27 pixels). These density maps are then down-sampled by a factor of 15, which results in a $27 \times 27$ map or a $1 \times 729$ feature vector.

*Aggregate density map* (*aggregate density*): We also tested an aggregate density map approach, where the fixations from all observers (as opposed to single observers) for a single image were merged to form a density map. That is, for image number $i \in \{1, \ldots, 800\}$ and task $j \in \{1, \ldots, 3\}$ fixations across the 19 observers were aggregated into a single density map. For the explicit judgment task, the aggregated data is based on pooling of the explicit judgements across 100 participants. This yields a number of observations per image similar to the fixation data. This pooling of the fixation data serves primarily to facilitate a comparison between data from viewing tasks, and from the explicit judgment task.

#### 3.2.2. *The Leung–Malik (LM) filter bank*

The LM filter bank [34] consists of 4 Gaussians filters corresponding to different spatial scales, 8 Laplacian of Gaussian (LoG) filters at different spatial scales, and first and second derivatives of Gaussians for each combination of 3 spatial scales and 6 different orientations (36 additional filters) for a total of 48 filters. To create an LM based feature vector, each image was convolved with the LM-Filter bank at its original scale, and the response of the 48 filters was sampled at each fixated location. Given that the total number of fixations is variable for any given image (generally from 7–15 fixations), filter responses across all fixations for an image were converted to summary statistics given by the mean response, and the standard-deviation of the response for each filter type across all observed fixations. This produces a 96 dimensional feature vector that captures the mean response, and variability in response of each of the filter channels for fixated regions of an image. This results in a $(19 \times 800 \times 3 \times 96)$ set of features for all combinations of image, participant and task. It is worth noting that the LM filter bank is chosen in part for its similarity to model simple cells represented within V1 and characterized by Gabor-like and center-surround receptive field profiles. This choice is relevant to making stronger assertions about human vision through task analysis by classification, and is an important consideration for future efforts in task prediction going forward.

#### 3.2.3. *Histogram of oriented gradients*

Histograms of Oriented Gradients are widely used in the computer vision literature [14], and have shown success in a range of tasks including object detection [17] or scene classification [56]. The HOG descriptor consists of histograms corresponding to oriented edge structure at different spatial scales within a local window of the image. Such features therefore capture coarse-grained summary statistics on the distribution of angular and radial frequencies represented within a local region of an image. While it's evident how such a representation may be used to determine whether an object is present at a given location, it's also natural to consider whether there is some inherent bias in edge content expressed in viewing behaviour across different tasks. Fixation based HOG features were generated in a fashion similar to the LM filter bank: At each fixated location in the original image, HOG features are extracted corresponding to a $65 \times 65$ image patch centred at the fixated location [16]. This results in a 31-dimensional feature vector for each fixation. Again, given variable numbers of fixations, the 31 dimensional HOG feature vector was converted to a summary representation, in considering the mean and standard deviation of HOG features across all fixations. Means and standard deviations are again coupled with a count of the total number of fixations for the reason stated above, yielding a 63 dimensional feature vector ($45{,}600 \times 63$ for all data).

#### 3.2.4. *Scene gist*

We have also considered a representation that captures the holistic structure of a scene based on the Gist descriptor [38]. The Gist descriptor is produced in sampling the responses of local filters sensitive to intensity gradients at different spatial scales, and over a grid of sub-windows on the image. These are subsequently converted to low dimensionality holistic receptive fields through PCA. This representation has been demonstrated as capable of classifying the type of scene (indoor, outdoor, forest, city, etc.) [38], and also having use in improving performance of models for predicting gaze locations [50]. The motivation for this set of features, is to examine whether general holistic scene structure is able to augment the ability to predict task when coupled with spatial densities of fixations.

### 3.2.5. Feature combinations

The relative value of individual base features types is important, but we are also interested in additional value that may be had in leveraging multiple distinct feature types for prediction. To this end, a number of composite feature sets have also been considered based on various combinations of spatial fixation density maps, fixated image features, scene structure and dynamics. One motivation for the incremental composition of features in our analysis is in establishing a more detailed understanding feature importance. The incremental gains in evaluating different subsets of features, provides additional information on redundancy in information captured by different feature sets with respect to the statistics that define task boundaries.

### 3.3. Classifiers

For results presented in this paper, Random Forests [4] are used for classification.[1] In all cases, 50% of the data was used for training, and the other 50% for testing. Classification is performed on different combinations of spatial and/or structural features to assess the relative efficacy of different cues and determine prediction performance.

Random Forest classification relies on the consensus predictions of a number of distinct decision trees. Each decision tree comprises a hierarchical arrangement of decision nodes. For example, the learning process might result in a root node that passes on an observation to one child node if the spatial fixation density for a particular location in the image is above some threshold, and to the other child node if the density is below this threshold, as depicted in Fig. 2. After a series of decisions of this type, which may also include branches on nodes that include feature based (HoG, LM), contextual statistics (Gist), or fixation statistics (number of fixations, saccade amplitudes), a leaf node is eventually reached that indicates the predicted task. An effective strategy for classification is to use a number of such decision trees in concert with an overall classification decision based on a majority weighted vote from individual decision trees. This can help to control against overfitting the data, but also brings additional benefits in diagnosing the value of features. In generating a collection of decision trees that make up a Random Forest, each decision tree is produced from independent data samples. Data samples from the training set are selected with replacement to produce a unique training set for each individual tree. Samples from the training set that are not included in a given sample are referred to as *out of bag* samples. Performance for a decision tree on *out of bag* samples provide one useful characteristic for understanding sample and feature importance. In the training process, when considering features to branch on in the tree, only a subset of the total variables/features is considered. This also introduces additional randomness that serves to produce diversity in the structure of trees that make up the Random Forest. In our evaluation, these bootstrap samples for decision nodes were based on $\sqrt{N}$ samples, where $N$ is the total number of statistics (features) used by the classifier. Factors that have a greater impact on prediction performance have more diagnostic value and are more important features in separating different task categories. A determination of feature importance may be made by permuting the values of a particular feature across different data samples. For example, the mean response of one of the LM filters across fixations provides a predictive statistic for each image. If these values are shuffled across the samples, it is possible to measure the impact on performance. This is performed for out of bag samples, providing a measure of feature importance for all of the individual features. This is an important property of this classification strategy, as it brings the additional value of discerning relative feature importance to understanding task-feature relationships. The detailed mathematical justification for this analysis is beyond the scope of this work, but the interested reader may refer to the statistical motivation given by Breiman [4].

A range of values was considered for the total number of voting tree-based classifiers including 50, 100, 200, 500, 1000, and 2000 trees respectively. This is in consideration of determining a ceiling on classification accuracy, but also provides confidence on the stability of the classification method used in much of the analysis. That is, a variable number of trees are considered to ensure that performance differences are due to feature differences rather than the complexity of the classifier.

## 4. Results

In the following, we examine the performance for the Random Forest classifier across the various feature sets, and in considering various combinations of features. This has been examined for the aggregate case (all observer data pooled), as well as the individual cases. In addition, we also examine multi-way classification and pair-wise binary classification to examine the separability of different task directives based on the fixation data. The following demonstrates that most of the features considered present significantly above chance performance, and also notably, that accuracy is dependent on the division of images among training and test sets.

### 4.1. Aggregate observers

Aggregating data across observers for each image/task combination results in 3200 ($4 \times 800$) image/task pairs. Half of these instances were used to train a Random Forest while the other half were used for testing only. Only the spatial densities were used for this test case as the resulting prediction rate is sufficient to show a strong degree of predictability based on spatial bias alone. This is especially true of explicit judgements since the associated data is not subject to the same gaze driven noise factors such as center bias or imprecise saccade targeting [52,43]. More detailed analysis is therefore reserved for the 3-way classification case, where the explicit judgement data across 100 observers is not included. Given that the spatial density features have a topographical organization, it is possible to visualize the importance profile in a topographical layout (Fig. 3). It is evident that the central region is of greatest importance, and that the degree of spatial (central)

[1] Alternative classification methods were also evaluated. The rationale for testing an array of classification methods was to confirm that results are representative of features chosen for prediction, and not the choice of classifier. All other classification methods performed no better than Random Forest based classification,and Random Forest based classification was more stable (similar performance using alternative methods sometimes required careful choice of parameters).

1. *Neural networks*: This evaluation employed multi-layer neural networks comprised of 2 or 3 layers of sparse autoencoders [53], with the weights of the sparse autoencoder training used to initialize a standard 2 or 3 layer back-propagation [54] network for classification. Performance was not any different than using Random Forests for the best cases, however sensitivity to parameters resulted in much poorer performance without careful tuning.
2. *Lasso regression* [49]: Regularized L1 Logistic regression was also evaluated for classification performance. With an appropriate choice of $\lambda$, results were on par with accuracy using Random Forests, however determination of this value was non-trivial, in part because this was dependent on features used for classification. In addition, mixed feature sets required normalization to achieve equivalent performance.
3. *AdaBoost* [18]: Several variants of AdaBoost were also tested. This method was relatively easy to obtain good prediction results, although in most cases accuracy values fell somewhat short of those produced by Random Forests. Among adaptive boosting methods, performance was best for the pseudo-loss variant of the standard algorithm [18] (as compared with LPBoost, TotalBoost and RUSBoost).

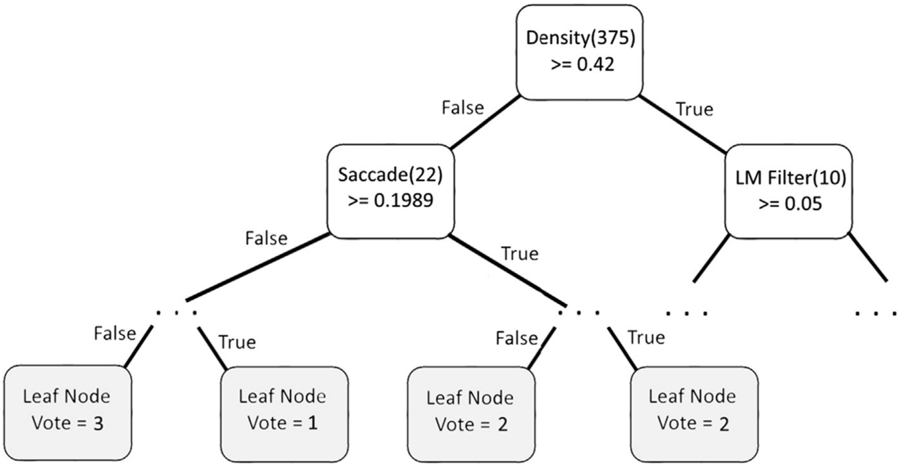


**Fig. 2.** A sample of decisions that may be taken by the nodes within a decision tree, the resulting leaf node corresponds to the tree's vote. Votes are collated across all trees comprising the Random Forest to determine decision of the ensemble.

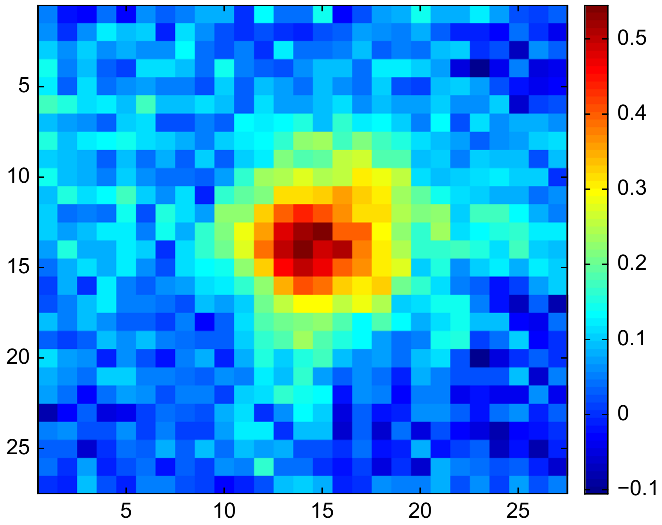


**Fig. 3.** Importance of spatial density statistics to classification performance in the aggregate density map. Importance measures are arranged topographically corresponding to their position in the fixation density map. This indicates that the relative importance of observations follows a concentric spatial profile.

**Table 2**
Aggregate density map results.

| Trees | All | Free/Obj | Free/Sal | Free/Exp | Obj/Sal | Obj/Exp | Sal/Exp |
|---|---|---|---|---|---|---|---|
| 50 | 70.87 | 84.38 | 66.13 | 88.25 | 89.88 | 97.50 | 90.00 |
| 100 | 69.31 | 83.25 | 65.75 | 89.75 | 89.75 | 97.75 | 89.75 |
| 200 | 69.87 | 83.63 | 65.38 | 89.38 | 90.00 | 97.62 | 89.88 |
| 500 | 68.87 | 84.13 | 65.75 | 89.25 | 89.75 | 97.50 | 89.88 |
| 1000 | 69.44 | 83.50 | 66.75 | 89.38 | 90.00 | 97.75 | 90.12 |
| 2000 | 69.19 | 83.87 | 66.37 | 89.50 | 90.25 | 97.62 | 89.62 |

bias, is one factor that has diagnostic value in determining the task being performed. This is consistent with many observations of the prominent degree of central bias present in gaze patterns, although this also hints that the degree and shape of central bias is task variant. Results for variable numbers of trees appear in Table 2 revealing the relative stability as a function of number of trees. Subsequent results for individual observers present only results corresponding to 2000 trees given this relative stability.

### 4.2. Individual observers

For task prediction based on data from individual observers, we consider the first 19 observers across the 3 task conditions to equate the number of observers per task. In total, there are 19 observers × 3 tasks × 800 images for a total of 45,600 cases to be distributed among training or testing. To provide deeper insight into the task prediction problem in general, we have considered two different partitions of this data as follows: *Partition* I: all of the observers, and all tasks are represented for half of the images. This implies that classifier predictions are not based on any patterns specific to individual images seen in training. *Partition* II: all observers and images are represented, but only half of the tasks carried out by each observer appear in the training set, and the other half in the test set (with equal number of samples of each task in training and test sets). This allows the relative importance of the specific images used in training, and importance of the size of the image set to be discerned. These two data partitions are referred to as P(I) and P(II) from hereon.

Classification results corresponding to the various feature/task combinations are summarized in Table 3 for P(I), and in Table 4 for P(II). There are some notable difference in the efficacy of different features, and also a significant impact on accuracies as a function of how data is partitioned. These points are discussed in detail in what follows, along with careful analysis of diagnostic measures of feature importance.

### 4.3. Spatial density

Similar to the case of classification based on aggregated observers, we assess the relative importance of different locations in the spatial density map (Fig. 4). The individual spatial densities shown in Fig. 4 also demonstrate significant weight in the importance of centrally located positions in the density map, albeit the individual case is characterized by a more pronounced peak at the very centre, accompanied by a more diffuse spread of feature importance over the scene outside of the centre. This again points to the importance of eccentricity of fixations as an importance distinction between tasks. It is also interesting to note that even relatively central fixations provide evidence diagnostic of task when considered as an ensemble. The importance of spatial (center) bias within studies of fixation behaviour has been explored in detail, and the data presented by Tatler [43] reveals a similar trend to these observations for two very similar task definitions but corresponding to a different set of image data.

**Table 3**
Individual density map results P(I) – all observers, all tasks, 50% of images.

| Spatial density | LM filters | HoG features | Gist | Sacc. ampl. | Num. fixations | All tasks | Free. vs. obj. | Free. vs. sal. | Obj. vs. sal. |
|---|---|---|---|---|---|---|---|---|---|
| Chance | | | | | | 33.33 | 50.00 | 50.00 | 50.00 |
| ✓ | | | | | | 48.34 | 66.55 | 58.89 | 66.13 |
| | ✓ | | | | | 42.98 | 65.05 | 52.32 | 62.18 |
| | | ✓ | | | | 41.06 | 61.98 | 51.36 | 60.20 |
| | | | | ✓ | | 54.54 | 76.89 | 56.69 | 75.65 |
| | | | | | ✓ | 50.65 | 73.11 | 54.81 | 71.38 |
| | ✓ | | | | ✓ | 49.45 | 73.23 | 51.84 | 71.26 |
| | | ✓ | | | ✓ | 49.73 | 73.35 | 53.52 | 71.33 |
| ✓ | | | ✓ | | | 48.47 | 66.54 | 58.80 | 66.12 |
| ✓ | ✓ | | | | ✓ | 50.09 | 69.94 | 59.27 | 68.09 |
| ✓ | | ✓ | | | ✓ | 50.26 | 69.38 | 59.17 | 68.19 |
| ✓ | ✓ | ✓ | ✓ | | | 51.59 | 71.34 | 59.53 | 69.32 |
| ✓ | ✓ | | ✓ | | ✓ | 50.34 | 69.35 | 59.00 | 68.15 |
| ✓ | ✓ | ✓ | ✓ | | ✓ | 51.62 | 71.00 | 59.48 | 69.12 |
| ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 53.42 | 73.41 | 59.59 | 71.01 |

**Table 4**
Individual density map results P(II) – all images and observers represented, 50% of tasks per observer.

| Spatial density | LM filters | HoG features | Gist | Sacc. ampl. | Num. fixations | All tasks | Free. vs. obj. | Free. vs. sal. | Obj. vs. sal. |
|---|---|---|---|---|---|---|---|---|---|
| Chance | | | | | | 33.33 | 50.00 | 50.00 | 50.00 |
| ✓ | | | | | | 54.6 | 76.24 | 58.22 | 75.18 |
| | ✓ | | | | | 42.89 | 65.07 | 51.96 | 62.83 |
| | | ✓ | | | | 43.79 | 64.60 | 52.19 | 62.95 |
| | | | | ✓ | | 54.45 | 76.86 | 56.69 | 75.37 |
| | | | | | ✓ | 51.04 | 73.42 | 54.65 | 72.13 |
| | ✓ | | | | ✓ | 49.79 | 73.56 | 53.25 | 71.58 |
| | | ✓ | | | ✓ | 50.39 | 74.19 | 53.70 | 71.89 |
| ✓ | | | ✓ | | | 54.74 | 76.27 | 58.17 | 75.27 |
| ✓ | ✓ | | | | ✓ | 56.05 | 77.82 | 59.58 | 76.28 |
| ✓ | | ✓ | | | ✓ | 56.16 | 77.70 | 59.58 | 76.35 |
| ✓ | ✓ | ✓ | | | ✓ | 56.91 | 78.70 | 59.81 | 76.88 |
| ✓ | ✓ | | ✓ | | ✓ | 56.11 | 77.86 | 59.59 | 76.33 |
| ✓ | ✓ | ✓ | ✓ | | ✓ | 56.74 | 78.50 | 60.20 | 76.82 |
| ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 58.14 | 79.98 | 60.16 | 77.85 |

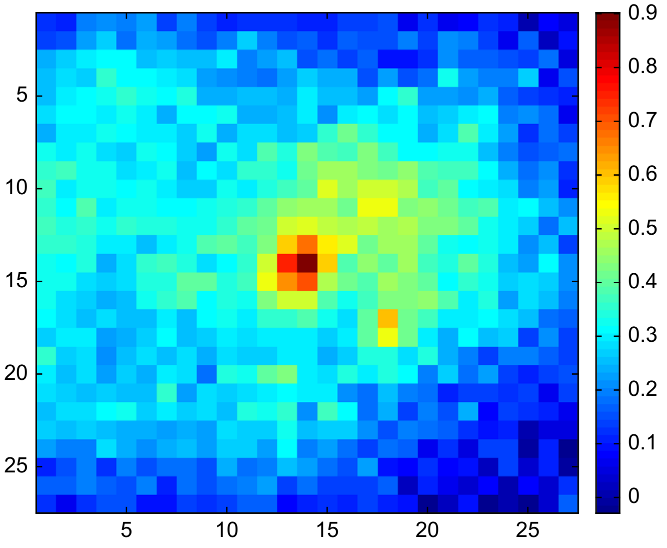


**Fig. 4.** Importance of spatial density statistics to classification performance in the individual density map. Importance measures are arranged topographically corresponding to their position in the fixation density map.

Differences in prediction accuracy between P(I) and P(II) are striking. While spatial density alone is among the more effective features for task prediction when all images appear in training (P (II)), its value is diminished significantly when training and test image sets are disjoint. This has important implications for how studies on task prediction are interpreted, especially in light of the relatively large image set associated with the data under consideration relative to prior efforts in task prediction. A more detailed discussion of the implications of this observation appear in Section 5.

### 4.4. Local image features

The LM Filters and HoG features alone present similar efficacy for both P(I) and P(II). This corresponds to approximately 42% accuracy for the 3-way classification tasks, and approximately 52–65% accuracy depending on task pairing, with free viewing and saliency viewing again most difficult to distinguish. In agreement with prior observations concerning the relative lack of importance of features at fixated locations, fixated image features alone are relatively poor at distinguishing between tasks. However, accuracy as high as 65% for some of the binary classifications suggests that statistical differences between these cases are not entirely spurious.

Important to understanding this observation, is the relative importance of different fixated features to distinguishing tasks. To support this analysis we present the relative importance of different features from the LM filter set in Fig. 5 based on out-of-bag analysis corresponding to the Random Forest based prediction. For illustrative purposes, first derivative filters corresponding to horizonal (red) and vertical (yellow) edge content are highlighted. There appears to be a consistent advantage to statistics associated with vertically oriented image structure at fixations in delineating task. A more detailed illustration of this difference is shown in

Fig. 6, which demonstrates the probability density associated with horizontal (left) and vertical (right) first derivative LM features for the 3 tasks. These correspond to features 7 and 10 appearing in Fig. 5. Free viewing, object search and saliency viewing correspond to the blue, red and yellow curves respectively. The bottom row depicts the *boxed* part of these curves at a higher level of zoom to better show the separation among feature distributions. For vertically oriented structure, overall separation is greater and also some degree of separation between the free viewing and object search conditions emerges. While this analysis does not reveal much about the reason for these differences, it demonstrates a subtle difference in structure of content at fixations that may be an important target for future analysis as it relates to task. It is also important to note that the value of this information is due chiefly to an accumulation of very weak evidence across a number of fixations in yielding task discrimination that is well above chance, and not any single fixation.

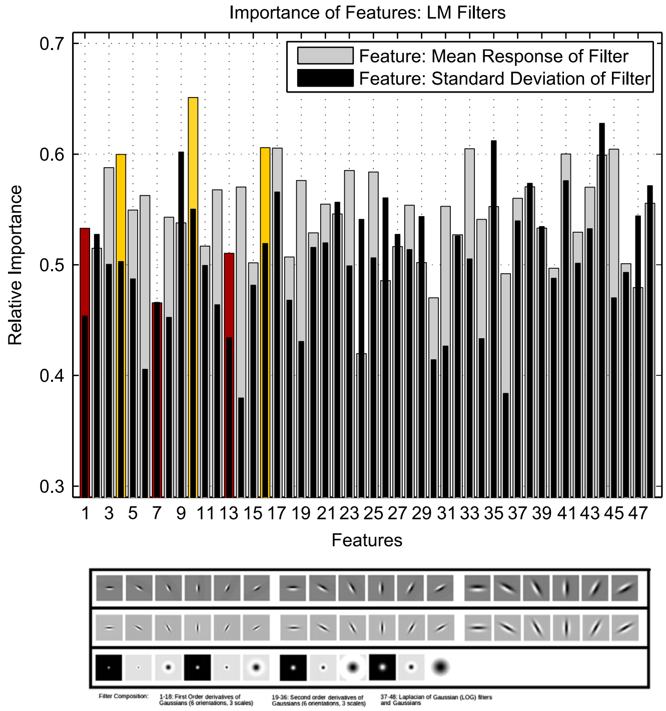


**Fig. 5.** Variable importance for statistics corresponding to mean and standard deviation of LM feature outputs across fixations. The order of bars in the plot mirrors the order of filters in the legend below: first row, left to right. Second row, left to right. Third row, left to right. First derivative filters show systematic variation as a function of orientation, with vertically oriented filters (yellow) carrying consistently higher predictive value than horizontal filters (red). (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this paper.)

An interesting asymmetry appears in the relative importance of spatial density, and local image structure for binary task classification. There is an advantage for spatial density over local features in delineating tasks independent of the data partitioning scheme, but this difference is much larger for P(II). However, the value of spatial density is invariant to the data partition scheme for discriminating saliency viewing from free viewing. An implication of this, is that the spatial density profile for saliency viewing and free viewing differ in a manner that is relatively independent of content specific to individual images, while object search appears to carry a spatial density profile that is more highly image specific. This is also revealing with respect to the level of specificity of image patterns driving statistical differences across different tasks.

### 4.5. Global image features

While fixation density for most locations (especially those proximal to the center) is relevant to task discrimination, feature importance for Gist features approaches 0 for all Gist feature dimensions. One possible explanation for this is that the holistic spatial envelope has a relatively small influence on gaze targets insofar as it interacts with task. That is, the influence of holistic scene structure may be relatively strong overall, but task invariant. An alternative possibility is that the task dependent influence of holistic/structural differences are already reflected implicitly in the

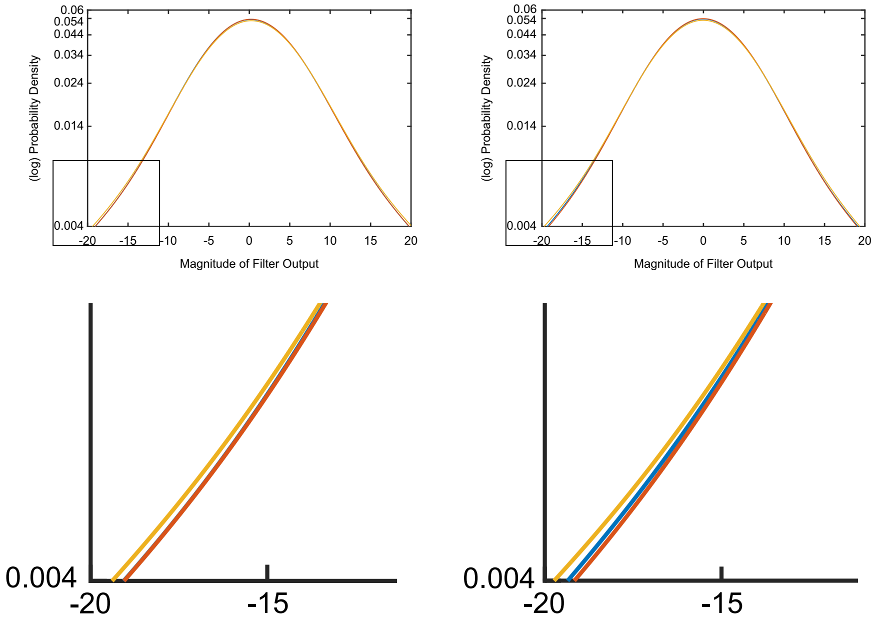


**Fig. 6.** Probability density associated with the response of two different LM filters (see text for details). Colours correspond to response densities for free viewing (blue), object search (red), and saliency viewing (yellow). Density profiles are shown for the horizontally oriented first derivative of Gaussian LM filter (left), and for the vertically oriented first derivative of Gaussian LM filter (right). The lower frames depict a magnified view of the lower left section of each curve. (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this paper.)

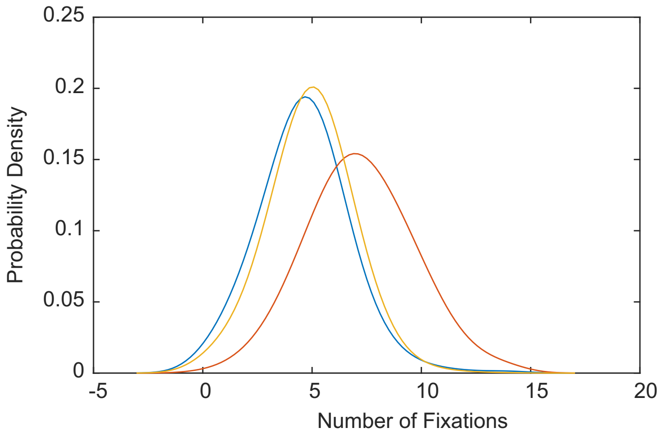


**Fig. 7.** Probability densities reflecting the number of fixations made for each image presentation for free viewing (blue), object search (red) and saliency viewing (yellow) tasks. (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this paper.)

spatial density profiles. That is, the Gist based structural representation may be a weaker cue when coupled with spatial density given redundancy in the information they capture. It is also the case that the number of image samples is small relative to what is typical for scene classification efforts [50].

### 4.6. Fixation dynamics

Fixation dynamics associated with the different tasks are characterized by the total number of fixations, and the amplitude of saccades observed within each task. These measurements are surprisingly effective in distinguishing between the tasks considered. Given that there are significant differences in total number of fixations (and latency) for the object search task compared with free viewing, and saliency viewing it is evident that this is a valuable statistic in distinguishing among these tasks. The probability density associated with the number of fixations for each class is shown in Fig. 7.

Perhaps more surprising is the strength of saccade amplitudes alone in distinguishing between tasks. In particular, for the challenging case of free viewing vs. saliency viewing, these are among the most important features alongside spatial fixation density. To examine this observation in more detail, we plot the relative importance of first, second and additional saccades made at the start of each trial in Fig. 8.

Fig. 8 reveals that the amplitude of the very first saccade is highest in importance, but there is also a high value to the several subsequent saccades in discriminating between tasks. The probability density associated with saccade amplitudes is shown in Fig. 9. This reveals that initially saccade amplitudes for object search are quite disparate from the other tasks, however, with an increasing number of saccades object search and saliency viewing converge, and distinguish themselves from free viewing. This observation is important in revealing the apparent value of fine grained temporal dynamics in providing defining traits associated with different tasks. Given that there exist differences between features at fixation for long vs. short saccades that are task dependent, interaction between relative spatial position of saccades and content at fixation is also likely to be relevant to inferring task [45] even for relatively general tasks definitions such as those examined in this paper. This also has implications for the role of task prediction for applications in human centric applications that make use of eye movements, with the assumption that there may be a significant advantage to classification models that employ a rich characterization of temporal dynamics.

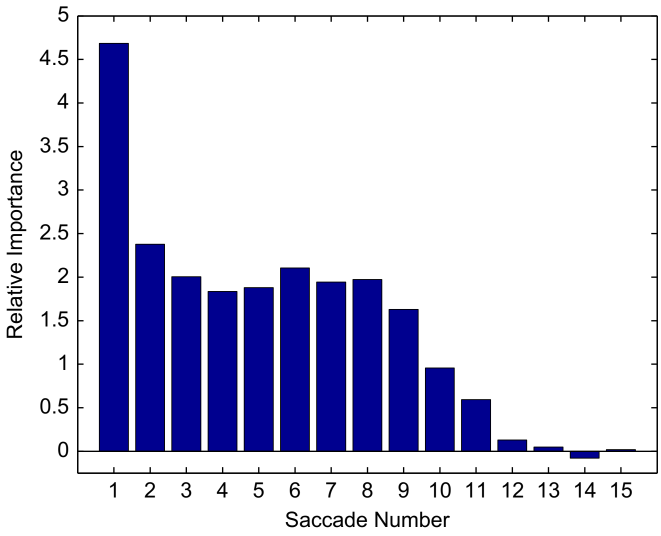


**Fig. 8.** Relative importance values for the first several saccade amplitudes based on out-of-bag analysis corresponding to the Random Forest Classifier.

### 4.7. Feature combinations

Performance for binary task classification is important to the overall interpretation of results as this provides a sense of similarity among behavioural observations among each pair of tasks. It is also evident in some of the preceding discussion within this section, that some important observations concerning relative importance of specific features or partitioning of data are possible only with a granularity of task prediction results that includes different subsets of features. This includes determining the degree of redundancy in information captured by different types of features. For example, relatively small gains are observed in combining LM and HoG features compared to their independent prediction accuracies. Slightly larger gains result from combining spatial densities and fixated features, and even larger gains in combining saccade amplitudes with spatial density and fixated features. These observations largely fit with *a priori* intuition concerning the overlap in information represented among such features. However, such analysis also helps to support or rule out other suspicions concerning the nature of task differences. For example, one might posit that feature differences at fixation are due primarily to bias in the spatial density profile of fixations that varies with task in combination with bias in how images are composed (framed and targeted by the photographer). However, the improvement seen in combining these features seems to deny the possibility that feature level differences are entirely spatial in their impetus. It is important therefore to note the methodological value of decomposition of both features, and task pairing in prediction to derive a deeper understanding of task relatedness.

## 5. Discussion

We have considered the extent to which relatively general task definitions, such as free viewing vs. search for objects may be distinguished on the basis of either the spatial density profile of fixations, features at fixation, scene structure or fixation dynamics. There are a variety of interesting observations that emerge from the classification experiments that include establishing the relative importance of different features in discriminating tasks, and highlighting important methodological considerations in how analysis by classification may be conducted.

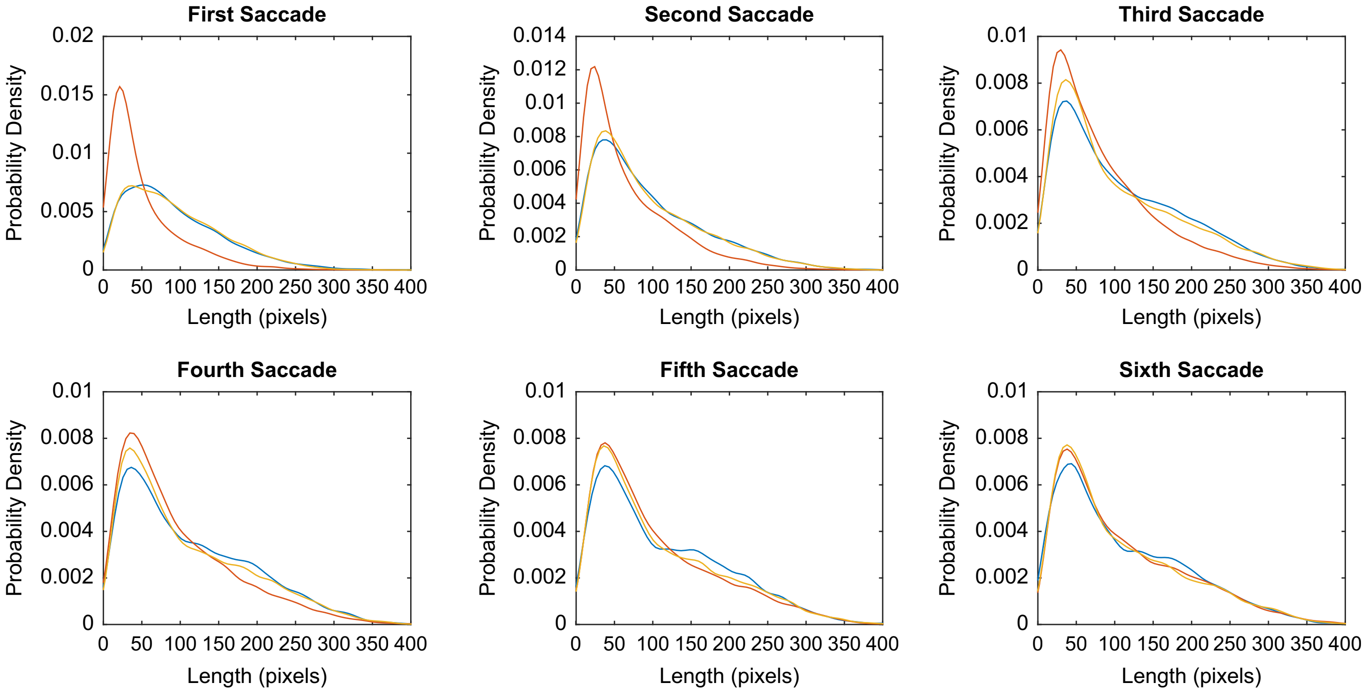


**Fig. 9.** Probability densities associated with saccade amplitudes (in pixels) for the first 6 saccades. Colours indicate free viewing (blue), object search (red) and saliency viewing (yellow). (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this paper.)

While free viewing and saliency viewing produce gaze derived statistics that are quite similar, these tasks are distinguished from object search with relative ease. Employing methods that place an emphasis on feature importance allows for subtle differences between tasks to be identified for more careful focused examination. For example, the relative importance of saccade amplitudes for the first several fixations in free viewing and saliency viewing may not be as readily observable outside of analysis of feature importance. In this view, some of the value of task prediction may reside in the capacity to test a large number of features in their value for task discrimination to identify targets for subsequent analysis using alternative methods. In this view, there is a role for task discrimination as a means for *high-throughput* screening for important feature dimensions.

While a task description may carry a relatively clear intent or definition, the associated neural and behavioural mechanisms that any task definition elicits may be relatively obscured in comparison. With that said, there is reason to be optimistic that a further proliferation of studies focused on eye movements including a larger variety of task definitions will help to clarify this relationship. This will allow for a stronger functional characterization of the nature of different *tasks* at various levels of abstraction (high-level or specific), while also distilling out the distinct associated neural or behavioural mechanisms that are recruited for particular tasks. (For a different set of methods towards this goal, see also [6,39]). Task prediction accuracies may be of value in determining task similarity, however, the specific features that are most discriminative in separating tasks may also provide important clues concerning how tasks are related. Finally, it is important to note that there are certain limitations to this type analysis and these are discussed in greater detail in Section 5.2.

On the balance of evidence from studies that aim to predict task from eye movements, there is evidently support for Yarbus' assertions concerning the importance of task in determining gaze behaviour, and recent efforts have demonstrated that fixation data may successfully predict task. However, analysis that emphasizes feature importance is necessary to understand specific factors that distinguish tasks, and also to understand which features are of universal value vs. those that are discriminative only for specific task pairings. This consideration has relevance to prior work in task prediction, and this consideration is discussed further in the remainder of this section.

### 5.1. *Fixation dynamics, covert attention and data partitioning*

There is a clear benefit for the tasks examined in this paper to considering fixation dynamics in conjunction with fixation densities or local image features. While viewing a scene is marked by a sequential scanpath that results from interaction between overt and covert attention mechanisms, analysis based on density alone fails to capture such temporal dynamics. This is evident in the work of Haji-Abolhassani and Clark [22], in which accuracy in task prediction exceeds prior efforts that consider primarily static spatial characteristics [1,29]. Moreover, modeling the problem in the framework of a Hidden Markov Model (HMM) allows the problem to be cast in terms of gaze points as observations, with the focus of attention as a hidden variable. This has the benefit of providing an explicit mechanism to relate overt attention to latent covert attention, and this framing of the problem also has apparent additional value in characterizing factors that distinguish different tasks.

One important property of the HMM characterization used by Haji-Abolhassani and Clark [22] is in modeling the relationship between observed gaze locations and attended locations. That a face is not fixated explicitly does not imply that it was not attended. This non-locality within the model allows for a richer construct for capturing the role of content not directly at the center of gaze in task prediction. In the characterization presented in this paper, there is some non-locality to prediction in that fixation density extends outside of the centers of gaze, and image characteristics rely on features or regions that correspond to a region surrounding the center of gaze. This is more restrictive than the model of Haji-Abolhassani and Clark [22], but does allow for some non-locality of spatial and structural information. It is also the case, that the HMM formalism considered by Haji-Abolhassani and Clark [22] requires stronger explicit definition of stimulus characteristics such as target position, or foci of interest as gleaned from fixation clusters.

The role over overlap among images in training/testing is highly relevant to the interpretation of results from prior studies that consider the Greene et al. [19] data. In the case of spatial densities, the small image set and overlap among training and testing data implies that information on spatial position of content that is image specific may be leveraged by a classifier that considers only spatial density. For the HMM based analysis of Haji-Abolhassani and Clark [22], this is also true as clusters of fixations, or labeled focal points within images are critical to the statistical representation in the affinity between observed fixation positions and these key locations. In contrast to general task dependent differences in center bias, or low level structure (e.g. edges) at fixation, both of these schemes present the possibility that image specific positions of high level features (e.g. objects) are a strong factor in boundaries drawn by a classifier. That is, in contrast to differences in dynamics, or general differences in spatial profile, success in delineating tasks may be relatively specific to known object positions for a known set of images.

To examine this point further, we consider differences in normalized fixation densities across two different tasks (Memory and People). The choice of this pair is based on the relatively small degree of confusion that exists between this pairing in prior classification studies. Spatial fixation densities are produced as discussed in the methods section, and subsequently the density maps for each image are averaged within task. This implies a density that is based on an average of 4 observer's viewing patterns. Fig. 10 shows the difference between densities associated with the Memory and People tasks, such that red regions indicate regions for which observed density is higher in the People task, and blue regions those for which density is higher in the Memory task. The visualization in Fig. 10 seems to leave little doubt that there are task dependent differences that are specific to certain types of image content. One might surmise that the failure to classify tasks above chance in the study of Greene et al. [19] is due primarily to a lack of features that capture positions of specific relevant objects within the images considered. In contrast, alternative classification efforts [1,29,22] carry a relatively strong encoding of this type of information, even if not explicit. Although the analysis of Greene et al. [19] includes a measure of dwell time corresponding to specific object categories, the correspondence to discrete localized regions, and non-spatial nature of this measure may be limiting. Moreover, it is possible that measuring dwell time rather than instances of fixations might limit the discriminative value of this type of feature. As a whole, the observations concerning the importance of specific object positions suggests that models capable of identifying and localization a large array of object types, or patterns of semantic relevance might achieve a much higher degree of success for task prediction given novel images or scenarios.

### 5.2. What can task prediction reveal about vision?

In recent years, studies involving task prediction have focused principally on prediction accuracies, and on confirming Yarbus' assertions concerning task predictability. In taking a more comprehensive account of task predictability, it is important to consider what studies involving task prediction are able to convey about human vision. To this end, we discuss a few types of analysis for which task prediction may be a valuable tool while also highlighting some of the associated limitations:

1. *Task similarity*: It is evident that binary classification accuracies might be used as a measuring stick for task similarity, and that the specific types of features that are discriminative for different task pairings may aid in this determination. One can imagine establishing an embedding (or topology) of task relatedness based on a large array of tasks, and suitably chosen features. However, failure to observe differences among tasks may be due

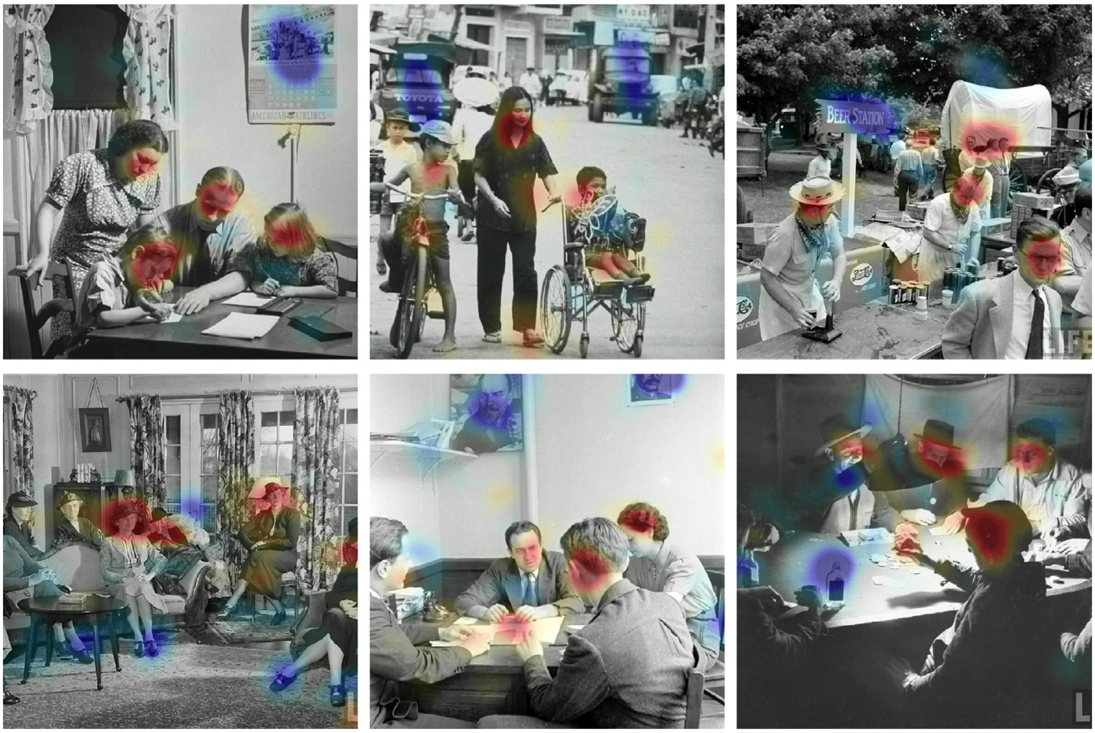


**Fig. 10.** A visualization of differential fixation density between the *People* and *Memory* tasks. Red regions correspond to those for which a higher density is observed with the *People* task, and blue the *Memory* task. (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this paper.)

to the set of features chosen. Moreover, the distances among task categories or ease of discriminating between tasks is also dependent on choosing the *right* set of features. This does not imply that the goal of establishing a representation of *task space* should be abandoned, but does call for caution in how results are interpreted. Pushing the ceiling on accuracies achieved in task prediction will help to establish the most relevant set of features. There are also alternative sources of data, such as brain imaging data, that may provide an adjunct source of statistics for considering measures of similarity and as a useful basis for comparison to task prediction results.

2. *Discriminative features*: One advantage of analysis by task prediction that is exposed in the results presented in this paper, is the capacity to identify subtle factors that delineate different tasks. As a tool for identifying relevant features, this provides the capacity to identify relevant experimental factors that may otherwise be ignored. This also affords the potential to probe a potentially large set of features in their capacity to discriminate between tasks, and to draw out factors of greatest significance.

## 6. Conclusions

In this paper, we have conducted a number of tests to determine the extent to which an observer's task may be determined based on eye tracking data. This has been carried out with a focus on the value of different factors such as spatial distribution of fixations, the nature of content at fixations, or fixation dynamics. This analysis is accompanied by discussion of methodological considerations important to interpreting results derived from task prediction. This provides a number of important observations relevant to the specific tasks considered as well as prior studies involving task prediction, and to methods associated with task prediction:

1. Unlike alternative studies involving task classification, we include a heavy emphasis on feature importance. This provides an indication of the relative importance of spatial fixation density, local image structure, holistic scene structure and fixation dynamics in distinguishing between different tasks. We observe that spatial density, and timing and length of saccades are important factors for classification. We also present evidence that for finer task distinctions, specific information about spatial positions of important objects may be relatively important to defining task differences.
2. While previous work has included measures of fixation density, saccade statistics and image salience, the current study marks the first effort to examine the value of local image content at fixated locations in predicting task for a Yarbus style paradigm. Results indicate that fixated content is of significant value in delineating some tasks, even if less important than other factors. While one might expect some bias in fixated structure due to differences in spatial profile, the combined strength of spatial fixation densities and structure of content at fixation indicates that fixated features carry diagnostic information that is independent of spatial position.
3. We have demonstrated that the means of partitioning experimental data is a very important factor in interpreting outcomes from task prediction. In particular, the value of spatial densities may be relatively specific to the content of known images for some task pairings, and less so for others. This also presents the possibility that success in some prior task prediction studies may be due to implicit representation of relatively high-level contextual, or object specific factors within the spatial density associated with a smaller set of *known* images.
4. The analysis presented in this paper is distinguished from prior efforts in the size of the data set, the task definitions considered, and methods that are used. Differences in gaze behaviour depend on both task definitions, and image content. It is our hope that as further research efforts in this domain continue to cover a more diverse range of task/data combinations that stronger conclusions may be possible concerning task relatedness, important principles, and underlying mechanisms.
5. Strategies that are successful in task prediction are also of value in application domains that include human-machine interaction, perceptual user interfaces and assistive technology for physiological or neurological conditions. This paper contributes to the growing body of strategies for task prediction towards supporting this goal. In particular, models that include a strong representation of both dynamics, and recognition of patterns with semantic relevance (e.g. objects) may be expected to be especially capable for these types of applications.

As a whole, this work contributes to the growing body of efforts that support Yarbus' assertions concerning fixation patterns. On the basis of the results presented, we are also optimistic that future efforts that emphasize task relevance will be fruitful in understanding task-gaze interaction. Further proliferation of efforts of this variety may also provide a window into the *bigger picture* of generalized task relatedness, task-data interaction and individual and general differences in gaze-task relationships.

## Acknowledgements

We gratefully acknowledge the financial support of the Natural Sciences and Engineering Research Council Canada Discovery Grants program and the University of Manitoba Graduate Enhancement of Tri-Council Stipends program. The authors also wish to thank Kathryn Koehler, Fei Guo, Sheng Zhang, and Miguel P. Eckstein for making their data available.

## References

[1] A. Borji, L. Itti, Defending Yarbus: eye movements reveal observers' task, J. Vis. 14 (2014) 29.

[2] A. Borji, A. Lennartz, M. Pomplun, What do eyes reveal about the mind?: algorithmic inference of search targets from fixations Neurocomputing 149 (2015) 788–799.

[3] A. Borji, D.N. Sihite, L. Itti, Quantitative analysis of human-model agreement in visual saliency modeling: a comparative study, IEEE Trans. Image Process. 22 (2013) 55–69.

[4] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32.

[5] N. Bruce, J. Tsotsos, Saliency based on information maximization, in: Advances in Neural Information Processing Systems, 2006, pp. 155–162.

[6] N.D. Bruce, Towards fine-grained fixation analysis: distilling out context dependence, in: Proceedings of the Symposium on Eye Tracking Research and Applications, ACM, Safety Harbor, FL, USA, 2014, pp. 99–102.

[7] N.D. Bruce, C. Wloka, N. Frosst, S. Rahman, J.K. Tsotsos, On computational modeling of visual saliency: examining what's right, and what's left, Vis. Res. 116 (2015) 95–112.

[8] A. Bulling, C. Weichel, H. Gellersen, Eyecontext: recognition of high-level contextual cues from human visual behaviour, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, Paris, France, 2013, pp. 305–308.

[9] G.T. Buswell, How People Look at Pictures, University of Chicago Press, Chicago, 1935.

[10] M.S. Castelhano, M.L. Mack, J.M. Henderson, Viewing task influences eye movement control during active scene perception, J. Vis. 9 (2009).

[11] M. Cerf, J. Harel, A. Huth, W. Einhäuser, C. Koch, Decoding what people see from where they look: predicting visual stimuli from scanpaths, in: Attention in Cognitive Systems, Springer, Berlin Heidelberg, 2009, pp. 15–26.

[12] X. Chen, G.J. Zelinsky, Real-world visual search is dominated by top-down guidance, Vis. Res. 46 (2006) 4118–4133.

[13] M.I. Coco, F. Keller, Classification of visual and linguistic tasks using eye-movement features, J. Vis. 14 (2014) 11.

[14] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, vol. 1, IEEE, San Diego, CA, USA, 2005, pp. 886–893.
[15] M. DeAngelus, J.B. Pelz, Top-down control of eye movements: Yarbus revisited, Vis. Cognit. 17 (2009) 790–811.
[16] P. Felzenszwalb, D. McAllester, D. Ramanan, A discriminatively trained, multiscale, deformable part model, in: 2008 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2008, IEEE, Anchorage, Alaska, USA, 2008, pp. 1–8.
[17] P.F. Felzenszwalb, R.B. Girshick, D. McAllester, D. Ramanan, Object detection with discriminatively trained part-based models, IEEE Trans. Pattern Anal. Mach. Intell. 32 (2010) 1627–1645.
[18] Y. Freund, R.E. Schapire, et al., Experiments with a new boosting algorithm, in: ICML, vol. 96, 1996, pp. 148–156.
[19] M.R. Greene, T. Liu, J.M. Wolfe, Reconsidering Yarbus: a failure to predict observers task from eye movement patterns, Vis. Res. 62 (2012) 1–8.
[20] A. Haji-Abolhassani, J.J. Clark, Realization of an inverse Yarbus process via hidden Markov models for visual-task inference, J. Vis. 11 (2011) 218–218.
[21] A. Haji-Abolhassani, J.J. Clark, Visual task inference using hidden Markov models, in: Twenty-Second International Joint Conference on Artificial Intelligence, 2011.
[22] A. Haji-Abolhassani, J.J. Clark, An inverse Yarbus process: predicting observers task from eye movement patterns, Vis. Res. 103 (2014) 127–142.
[23] A. Haji-Abolhassani, J.J. Clark, A computational model for task inference in visual search, J. Vis. 13 (3) (2013) 29.
[24] J. Han, L. Sun, X. Hu, J. Han, L. Shao, Spatial and temporal visual attention prediction in videos using eye movement data, Neurocomputing 145 (2014) 140–153.
[25] M. Hayhoe, D. Ballard, Eye movements in natural behavior, Trends Cognit. Sci. 9 (2005) 188–194.
[26] J.M. Henderson, S.V. Shinkareva, J. Wang, S.G. Luke, J. Olejarczyk, Predicting cognitive state from eye movements, PLoS One 8 (2013) e64937.
[27] Y. Hua, M. Yang, Z. Zhao, R. Zhou, A. Cai, On semantic-instructed attention: from video eye-tracking dataset to memory-guided probabilistic saliency model, Neurocomputing 168 (2015) 917–929.
[28] W. Jones, A. Klin, Attention to eyes is present but in decline in 2–6-month-old infants later diagnosed with autism, Nature 504 (2013) 427–431.
[29] C. Kanan, N.A. Ray, D.N. Bseiso, J.H. Hsiao, G.W. Cottrell, Predicting an observer's task using multi-fixation pattern analysis, in: Proceedings of the Symposium on Eye Tracking Research and Applications, ACM, Safety Harbor, FL, USA, 2014, pp. 287–290.
[30] K. Koehler, F. Guo, S. Zhang, M.P. Eckstein, What do saliency models predict? J. Vis. 14 (2014) 14.
[31] M. Kümmerer, T.S. Wallis, M. Bethge, Information-theoretic model comparison unifies saliency metrics, in: Proc. Natl. Acad. Sci. 112 (2015) 16054–16059.
[32] K. Kunze, Y. Utsumi, Y. Shiga, K. Kise, A. Bulling, I know what you are reading: recognition of document types using mobile eye tracking, in: Proceedings of the 17th Annual International Symposium on International Symposium on Wearable Computers, ACM, Zurich, Switzerland, 2013, pp. 113–116.
[33] F. Lethaus, M.R. Baumann, F. Köster, K. Lemmer, A comparison of selected simple supervised learning algorithms to predict driver intent based on gaze data, Neurocomputing 121 (2013) 108–130.
[34] T. Leung, J. Malik, Representing and recognizing the visual appearance of materials using three-dimensional textons, Int. J. Comput. Vis. 43 (2001) 29–44.
[35] P. Loyola, G. Martinez, K. Muñoz, J.D. Velásquez, P. Maldonado, A. Couve, Combining eye tracking and pupillary dilation analysis to identify website key objects, Neurocomputing 168 (2015) 179–189.
[36] M. Mills, A. Hollingworth, S. Van der Stigchel, L. Hoffman, M.D. Dodd, Examining the influence of task set on eye movements and fixations, J. Vis. 11 (2011) 17.
[37] T. O'Connell, D. Walther, Fixation patterns predict scene category, J. Vis. 12 (2012) 801, http://dx.doi.org/10.1167/12.9.801.
[38] A. Oliva, A. Torralba, Building the gist of a scene: the role of global image features in recognition, Prog. Brain Res. 155 (2006) 23–36.
[39] S. Rahman, N.D. Bruce, Factors underlying inter-observer agreement in gaze patterns: predictive modelling and analysis, in: Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications, ACM, Charleston, SC, USA, 2016, pp. 155–162.
[40] C.A. Rothkopf, D.H. Ballard, M.M. Hayhoe, Task and context determine where you look, J. Vis. 7 (2007) 16.
[41] C. Shen, Q. Zhao, Learning to predict eye fixations for semantic contents using multi-layer sparse network, Neurocomputing 138 (2014) 61–68.
[42] Y. Sugano, H. Kasai, K. Ogaki, Y. Sato, Image preference estimation from eye movements with a data-driven approach, J. Vis. 7 (2014) 1–9.
[43] B.W. Tatler, The central fixation bias in scene viewing: selecting an optimal viewing position independently of motor biases and image feature distributions, J. Vis. 7 (2007) 4.
[44] B.W. Tatler, R.J. Baddeley, I.D. Gilchrist, Visual correlates of fixation selection: effects of scale and time, Vis. Res. 45 (2005) 643–659.
[45] B.W. Tatler, R.J. Baddeley, B.T. Vincent, The long and the short of it: spatial statistics at fixation vary with saccade amplitude and task, Vis. Res. 46 (2006) 1857–1862.
[46] B.W. Tatler, M.M. Hayhoe, M.F. Land, D.H. Ballard, Eye guidance in natural vision: reinterpreting salience, J. Vis. 11 (2011) 5.
[47] B.W. Tatler, B.T. Vincent, The prominence of behavioural biases in eye guidance, Vis. Cognit. 17 (2009) 1029–1054.
[48] B.W. Tatler, N.J. Wade, H. Kwan, J.M. Findlay, B.M. Velichkovsky, Yarbus, eye movements, and vision, I-Percept. 1 (2010) 7.
[49] R. Tibshirani, Regression shrinkage and selection via the lasso, J. R. Stat. Soc. Ser. B (Methodol.) (1996) 267–288.
[50] A. Torralba, A. Oliva, M.S. Castelhano, J.M. Henderson, Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search, Psychol. Rev. 113 (2006) 766.
[51] P.-H. Tseng, I.G. Cameron, G. Pari, J.N. Reynolds, D.P. Munoz, L. Itti, High-throughput classification of clinical populations from natural viewing eye movements, J. Neurol. 260 (2013) 275–284.
[52] P.-H. Tseng, R. Carmi, I.G. Cameron, D.P. Munoz, L. Itti, Quantifying center bias of observers in free viewing of dynamic natural scenes, J. Vis. 9 (2009) 4.
[53] P. Vincent, H. Larochelle, Y. Bengio, P.-A. Manzagol, Extracting and composing robust features with denoising autoencoders, in: Proceedings of the 25th International Conference on Machine Learning, ACM, Helsinki, Finland, 2008, pp. 1096–1103.
[54] T.P. Vogl, J. Mangis, A. Rigler, W. Zink, D. Alkon, Accelerating the convergence of the back-propagation method, Biol. Cybern. 59 (1988) 257–263.
[55] N. Wilming, T. Betz, T.C. Kietzmann, P. König, Measures and limits of models of fixation selection, PLoS One 6 (2011) e24038.
[56] J. Xiao, J. Hays, K.A. Ehinger, A. Oliva, A. Torralba, Sun database: large-scale scene recognition from abbey to zoo, in: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Providence, RI, USA, 2010, pp. 3485–3492.
[57] H. Yang, G.J. Zelinsky, Visual search is guided to categorically-defined targets, Vis. Res. 49 (2009) 2095–2103.
[58] A.L. Yarbus, Eye Movements and Vision, Plenum, New York, 1967.

**Jonathan F.G. Boisvert** graduated with an Honours B.C.Sc. in Computer Science from the University of Manitoba in 2013, and his M.Sc. in Computer Science in 2016. His current research interests include computer vision and artificial intelligence as points of emphasis.

**Neil D.B. Bruce** is currently an Assistant Professor in the Department of Computer Science at the University of Manitoba, Canada. He has been a postdoctoral fellow at the Centre for Vision Research at York University, and at INRIA Sophia Antipolis in France. He holds a Ph.D. in Computer Science (York University, 2008), M.A.Sc. in System Design Engineering (University of Waterloo, 2003), and an Honours B.Sc. with Majors in Computer Science and Mathematics (University of Guelph, 2001). His research interests include human and computer vision, computational neuroscience, visual attention, HCI and visualization.

Please cite this article as: J.F.G. Boisvert, N.D.B. Bruce, Predicting task from eye movements: On the importance of spatial distribution, dynamics, and image features, Neurocomputing (2016), http://dx.doi.org/10.1016/j.neucom.2016.05.047