Towards the quantitative evaluation of visual attention models


1. Introduction

Several decades of experimental research have uncovered a variety of neural and behavioral phenomena associated with visual attention. Physiological and brain imaging studies have been useful for exploring neural underpinnings of attention (Kastner & Ungerleider, 2000; Miller & Buschman, 2013), and psychophysical studies have examined various behavioral manifestations of human visual attention (Petersen & Posner, 2012; Simons & Chabris, 1999; Wolfe, 1998, 2007) (see also the ‘Course Readings’ section of the references). A synthesis of all this data is warranted; however, while it is unclear what it means to truly understand visual attention, these independent data points are likely insufficient. Instead, scientific progress is made by a meaningful compression of data, for example by constructing models that can explain and predict a diverse range of phenomena. In this domain, computational, rather than conceptual (or descriptive) models, have the advantage of providing quantitative explanations of the collected observations as well as making new predictions that are testable and verifiable. The use of computational models has led to progress in our understanding of various phenomena. For instance, developments in bottom-up attention modeling have led to an increased understanding of where people look in different images under varying conditions (Borji et al., 2013; Itti & Baldi, 2009; Judd, 2011; Tatler, 2007), computational models have been able to predict the effects of crowding on visual tasks (Balas, Nakano, & Rosenholtz, 2009; Rosenholtz et al., 2012), and to model top-down scene guidance for visual search tasks (Ehinger et al., 2009; Torralba, Oliva, Castelhano, & Henderson, 2006; Tsotsos, 2011). Taken together, this suggests that constructing computational models to solve specific visual attention tasks could lead to progress in understanding visual attention as a whole.

Nevertheless, we begin in Section 2 by highlighting the difficulties in model evaluation and comparison brought about by the simultaneous abundance of computational models of visual attention and the lack of model overlap across taxonomies. In Section 3 we advocate for quantitative evaluation via (i) operationalizing definitions of individual visual attention tasks and (ii) specifying rigorous protocols for measuring model performance under those tasks, and we provide some implementable examples. Operationalized task definitions are those that include sufficient detail and specificity so that the tasks may be put into practice,
implemented on a computer and quantitatively evaluated on meaningful input stimuli. We advocate against any abstract and ambiguous constructs that do not lend themselves easily to quantitative evaluation.

Next, in Section 4 we emphasize the need for large, multi-faceted, standardized benchmark datasets, and offer a discussion of the design considerations that surface. Finally, we outline the benefits of competition-style online benchmarks in Section 5 for measuring modeling progress. Altogether, this paper offers a number of suggestions and considerations that have proven successful at bringing structure and standardization to other computational areas (e.g. evaluation methodologies and benchmark datasets in saliency modeling (Borji et al., 2013; Bylinskii et al., 2014; Judd, Durand, & Torralba, 2012)), computer vision (Deng et al., 2009; Everingham et al., 2012; Lin et al., 2014; Torralba, Fergus, & Freeman, 2008; Xiao et al., 2010), and natural language processing (NIST, 2013; Voorhees, 2004; Voorhees & Harman, 2005)).

2. Moving beyond taxonomies

Many computational models of visual attention have been built during the past three decades. However, the sheer diversity of models makes comparison and evaluation of progress in the field of visual attention particularly difficult. In an attempt to understand the relationships between different models, various taxonomies and other categorizations have been introduced, some of which attempt to cover multiple types of computational models, and others that focus on specific subareas of visual attention or specific model structures. For instance, Frintrop, Rome, and Christensen (2010) classify models according to their structure, labeling them either as filter models, those that parse image features via image mapping, or connectionist models, those that employ neural network computations to process images. Tsotsos and Rothenstein (2011) divide computational models (themselves branching off from both computer and biological vision categories) into four types: selective routing models, saliency map models, temporal tagging models, and emergent attention models. Kimura, Yonetani, and HIRayama (2013) classify models as either bottom-up or top-down, each composed of several subcategories determined by the models’ algorithmic approach. Borji and Itti (2013) present a categorization of bottom-up and top-down models, qualitatively comparing 13 criteria.

In Fig. 1 we visualize the number of models that are considered by each of 4 categorizations (Borji & Itti, 2013; Frintrop et al., 2010; Kimura et al., 2013; Tsotsos & Rothenstein, 2011). We can see that relatively few models occur in more than one taxonomy/categorization, making comparisons very difficult. Each categorization covers only a subset of models and proceeds by carve up ing these models according to some author-defined set of characteristics. Another observation is that the sheer number of visual attention models that have been developed over the past few decades is staggering, and continues to grow.

Let us consider a single model categorization in greater detail. According to Borji and Itti (2013), there are a total of 13 criteria by which many of these models may be compared: bottom-up, top-down, spatial/spatiotemporal, task-type, space-based/object-based, features, model type, static, dynamic, synthetic, natural, measures, and dataset used. The first 7 criteria correspond to the models themselves, and the latter 6 are specific to task completion and evaluation. As Borji and Itti note, these criteria help establish the scope of applicability of these different models. In Fig. 2a, we visually represent this taxonomy by projecting down the model characteristics onto 3 dimensions. Gaussian noise was added to the projections to visualize models with identical 3-dimensional projections. The resulting representation accurately captures the factor similarity of models, i.e. models that are spatially clustered together share many taxonomical attributes. The dimensions of this representation are principal components that represent a linear combination of factors, although they do align fairly well with the factors: bottom-up/top-down, dynamic/static, and synthetic/natural. In 2b, we hold this spatial layout of models fixed, and overlay on top of it multiple model characteristics (represented by the coloring of models). From such a visualization we can see that models are clustered together in model space, with many overlapping and correlated characteristics. For example, bottom-up and top-down models are segregated along the first dimension of this representation, while models with synthetic versus natural stimuli are segregated along the third dimension. Thus, although the quantity of models is large, many reuse the same principles and computational approaches, and thus have similar application areas (use cases).

Taxonomies thus provide a way to describe models, but not with a method of sorting through them to discover the most accurate representation of human visual attention. We can use taxonomies to describe the characteristics of different models, or to identify models which may be sensibly compared, because they solve similar tasks or use comparable computational approaches. However, if a quantitative evaluation is sought, these descriptions need to be supplemented with a methodology of comparison. Quantitative evaluation can help us isolate the model characteristics that are essential to performance on different visual attention tasks.

Even though some attempts have been made to quantitatively evaluate a wide varieties of models according to some predefined criteria (Borji, Sihite, & Itti, 2012; Filipe & Alexandre, 2013; Heinke & Humphreys, 2005; Judd et al., 2012; Koehler et al., 2014), these endeavors only provide a comparison of a relatively
small subset of computational models, which highlight further the need for more coherence within the field. Thus it is evident that there is much work left to be done in making comparison and evaluation more quantitative. We suggest that this can be accomplished via developments in: (1) operationalizing the definitions of visual attention tasks (Section 3), and (2) defining success on these tasks via benchmarking datasets (Section 4). In this paper, we focus on these two particular issues with respect to the computational modeling of visual attention.

3. Operationalizing definitions of visual attention tasks (for model evaluation)

William James famously said “Everyone knows what attention is.” Nonetheless, attention has been variously described by those attempting to study it as an emergent property (Desimone & Duncan, 1995), a controlling factor (Rensink, 2000), a mental ability that allows for the selection of behaviorally relevant stimuli (Corbetta, 1998) and a reallocation of visual processing resources while preserving reactivity to rapid changes in the environment (Foley, Grossberg, & Mingolla, 2012). Here, we propose that instead of attempting to reach agreement on a particular semantic definition of attention, more progress can be achieved by moving towards operational definitions of different visual attention tasks, including free viewing and visual search (Table 1). These operational definitions should provide specifications of the phenomena (behavioral, physiological, etc.) that should be reproducible by image-computable models of attention. Image-computable models take as input an image, a task, and a system state, and algorithmically compute a function of the image. Image-computability makes models of visual attention testable and comparable to other models.

We provide examples of a few visual tasks in Table 1 below. Operationalization allows us to turn our attention away from arguing over whether a model is a good model of visual attention, to measuring how the model accomplishes a particular visual attention task, or set of tasks.

For example, one behavioral manifestation of stimulus-based bottom-up attention is the pattern of eye movements made by humans on images under some fixed task constraints (e.g. free-viewing). We can operationally define free-viewing as human eye movements when given an image to look at and no specific task instructions. Based on this definition, we can then choose a measurement of eye movements: location of fixations, ordered sequence of fixations (scanpath), first fixation, dwell time per fixation, saccade extent, etc. Once we choose a measurement, this directly provides us with a methodology for evaluating models: for instance, if we are interested in fixation locations, we can use similarity metrics to compare model-predicted likelihoods of fixating different image regions with human fixation maps. Saliency models are image-computable, taking images as input, and returning topographic maps indicating the likelihood of fixation (or “saliency”) at each location in the image. Viewed as a distribution over an image, a saliency map can be compared to a human fixation map using various similarity metrics. Another possible model of bottom-up attention might be one that takes as input an image and the previous fixation location (as a system state) and returns the next location of fixation as its output.

---

3 For instance, although imperfect, Intelligence Quotient (IQ) tests provide a way of operationalizing intelligence, by reducing it to a measurable quantity that can be studied empirically (for instance, to investigate the impact of a particular condition on intelligence, we can quantify its effect on IQ score). Note that multiple operationalizations can exist (as is the case for intelligence), and thus to facilitate comparison across studies, research findings should make explicit reference to the particular operational definitions employed.

4 However, different experiments claiming to record free viewing fixations often provide participants with implicit tasks (e.g. to remember an image) and a very constrained experimental setting (e.g. fixed distance to screen, image of finite size displayed for short period of time). This further highlights the difficulty, but necessity, of having operationalized definitions for all experimental components: the phenomena studied, the tasks implemented, the constraints and assumptions used, etc.
In Fig. 3 we include the performances of a subset of the bottom-up models from Fig. 2 that are also available on the MIT Saliency Benchmark (Bylinskii et al., 2014). Seven metrics are provided, allowing for multiple comparisons across, and ranking of, computational models. Whereas taxonomic classifications can point out the differences between models (with respect to their

Table 1
Sample operational definitions for a few visual attention tasks. Under each definition and experimental measurement, the dataset requirements and metrics naturally fall out. Note that many other possibilities exist for each task, for which we only provide a few examples. For instance, we provide behavioral definitions, whereas physiological definitions can also be applied (e.g, a description of the neuronal firing patterns and rates expected under different conditions).

<table>
<thead>
<tr>
<th>Task</th>
<th>Operational definition</th>
<th>Experimental measurement</th>
<th>Dataset</th>
<th>Evaluation metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Free-viewing</td>
<td>Human eye movements when given an image to look at and no task instructions</td>
<td>Fixation locations</td>
<td>Set of images with ground-truth human fixations (see Bylinskii et al., 2014 for a listing of popular ones)</td>
<td>Similarity between human fixations and model saliency map (see Riche et al., 2013 for some examples)</td>
</tr>
<tr>
<td>Visual search with fixed gaze</td>
<td>Human response when given an image with or without a target and distractors and the instructions to respond to target absence/presence without changing fixation (speeded or unspeeded)</td>
<td>Response time; RT × set size; Target prediction</td>
<td>Set of images with varying number of distractors (with and without a target) and accompanying human response times; Set of images (as before) and ground-truth target locations</td>
<td>Similarity between human and predicted response times; Similarity between human and predicted RT slopes; Detection accuracy (see Macmillan and Creelman, 1991)</td>
</tr>
<tr>
<td>Visual search with free gaze</td>
<td>Human response and eye movements when given an image with target and distractors and the instructions to respond to target absence/presence by fixating target</td>
<td>Same as for visual search with fixed gaze, potentially also fixation locations and scanpaths</td>
<td>Same as for visual search with fixed gaze, but with the additional possibility of peripheral displays in the images, and ground-truth human fixations</td>
<td>Same as for visual search with fixed gaze, potentially also similarity between human fixations and model saliency map</td>
</tr>
</tbody>
</table>

characteristics), metrics allow us to step further and make quantitative judgements about which model performs better on a given task, and under which metric.

Many popular image datasets with human gaze data exist for evaluating saliency models (reviews of some of these can be found in Borji et al. (2013), Judd et al. (2012), and a regularly-updated online listing can be found in Bylinskii et al. (2014)). A discussion of commonly-used evaluation metrics is provided in Riche et al. (2013). Additionally, other behavioral (e.g. human ratings) and physiological (e.g. neural recordings) manifestations of bottom-up attention are amenable to evaluation on separate datasets, and other sets of metrics may be appropriate. We have yet to see the development of benchmark datasets and large-scale evaluation methodologies for bottom-up attention beyond the prediction of eye movements and saliency.

3.1. Additional considerations

In a more general review of computational models of selective attention, Heinke and Humphreys (2005) put forth several possible evaluation metrics, though indicating that comparison between all models is not feasible given the wide variety of model designs and tasks. They note that a chief difficulty lies (a) in the variability of parameter settings, and (b) the lack of a consensus on how to weight biological plausibility when evaluating models. On these two issues, we make the following observations:

(a) Reporting the influence of different parameter choices on model performance (e.g. via sensitivity analysis5) is crucial for standardizing evaluation, simply because a single setting of parameters does little to disentangle the power of the underlying model formulation from fine parameter tuning. Not only does the lack of such reporting not help progress, it actually hampers it by potentially offering misleading information about what works well and what does not.

(b) To quantify biological plausibility, models should be evaluated both in their ability to explain the most comprehensive behavioral data (e.g. reaction times, gaze patterns, success rate, etc.) as well as the available underlying biological data at various levels of abstraction (e.g. neuronal activity, local field potentials, EEG/MEG, neuro-imaging, etc.). An integrated, complete computational description should unite behavioral findings with the underlying physiological mechanisms.

To summarize, in order to quantitatively measure progress within a given subarea of visual attention, we advocate for the use of computational models evaluated under a rigorous quantitative protocol. This involves first selecting and operationalizing particular visual attention tasks, and specifying evaluation protocols under those task definitions. The model evaluation protocols should additionally detail (1) what data and task constraints are to be used; (2) which metrics models will be compared on; (3) what additional aspects of models should be reported (e.g. parameter choices, model complexity, underlying biological assumptions, etc.).

4. Need for benchmarking datasets (for model comparison)

Given a set of operationally-defined tasks comprising visual attention, the ultimate goal would be to have a model capable of performing well on a maximum number of these tasks. This would hopefully lead to a unified computational understanding of the mechanisms underlying visual attention. Such a model would become the reference model of visual attention.6

In contrast, in the case of saliency, there has been significant noise in the reporting of successes because many papers introducing novel saliency models report the performance of their model on some choice of dataset(s), under some choice of metric(s), compared against some choice of other models. All these choices afford too much flexibility to the model authors as to how the success of their model’s performance is quantified. For instance, in the classification of models provided in Borji and Itti (2013), most of the 63 models considered are included along with the metrics and datasets the models were originally evaluated on. Over 27 distinct datasets are listed and an additional 8 mentions of authors gathering their own data. From this information alone, it is impossible to infer which models are objectively doing well at predicting human fixations. Thus, while some coarse but useful comparisons can be made across models, there is currently no clear conclusion of a winning reference model.

A systematic approach to determining a possible reference model would be to compare all existing models on standard benchmark evaluation datasets under the same set of metrics. Standardized datasets and evaluation metrics have proved successful in the fields of computer vision (Deng et al., 2009; Everingham et al., 2012; Lin et al., 2014; Torralba et al., 2008; Xiao et al., 2010) and natural language processing (NIST, 2013; Voorhees, 2004; Voorhees & Harman, 2005). Unfortunately, benchmark datasets are difficult to construct, and are consequently rare in visual attention. Benchmarks are, however, becoming increasingly popular for the evaluation of saliency models, and this is helping to drive progress and eliminate some of the aforementioned noise in model evaluation. The problem is that saliency benchmark datasets are now where computer vision datasets were a decade ago (Antonio Torralba, personal communication, 4/28/14). By image count alone, computer vision has now moved on to datasets of tens of thousands to many millions of images in size (Deng et al., 2009; Everingham et al., 2012; Lin et al., 2014; Torralba et al., 2008; Xiao et al., 2010), while saliency datasets, in particular, have not grown beyond a few hundred to a few thousand images (Borji et al., 2013; Judd et al., 2012).

4.1. Dataset design considerations

Large datasets alone are insufficient for capturing the space of complex behaviors that are attributable to visual attention. A good benchmark dataset for testing computational models of visual attention should offer multiple tasks on which performance can be reported. As an analog, a popular image benchmark in computer vision (the PASCAL challenge Everingham et al. (2012)) consists of separate competitions that computational models can be evaluated on (e.g. classification, detection, segmentation, etc.). Computing model performance on different tasks may help illuminate what aspects of visual attention the model is most predictive of Koehler, Guo, Zhang, and Eckstein (2014), as well as which models may have complementary functionalities. Testing models on multiple tasks also provides a way of differentiating a solid underlying model framework (capable of generalizing) from one that owes its performance to the implementation details and parameter tuning specific to a task.

5 Sensitivity analysis can be performed in many different ways, but comes down to testing the robustness of a model under changing inputs and parameter settings, to carefully measure effects on performance and quantify uncertainty (Saltelli et al., 2008) provides a thorough treatment of the subject.

6 By reference model we mean a standardized model that would be the accepted reference point for other models to compare against, at a fixed point in time. To obtain this status, a model would need to be capable of explaining the most complete set of observations on the broadest set of tasks.
Furthermore, summary performance numbers may be insufficient to tease models apart: is a model performing reasonably across all conditions, or is its performance confined to a small set of successes? Do different models make similar mistakes, or are they complementary in the data they can explain? Reporting multiple performance numbers on each dataset may reveal common and complementary aspects of different models.

Additionally, reporting results on multiple datasets can help alleviate dataset bias, which can have severe implications on model evaluation (Andreadopoulos & Tsotsos, 2012; Torralba & Efros, 2011). For instance, center-bias has been a prevalent problem in saliency model evaluation (Borji et al., 2013; Judd et al., 2012), with a simple center prior model, completely ignorant of image content, often outperforming many other models. One potential cause is photographer bias (placing the main content in the center), which could be avoided, for instance, by a more careful selection of images. An alternative to modifying the data is designing evaluation metrics to compensate for dataset biases. For instance, the shuffled AUC (sAUC) metric (Borji et al., 2013; Zhang et al., 2008) assigns chance performance to a center prior model. However, Tatier (2007) has found that observer bias is a major contributing factor to center bias (a kind of optimal viewing strategy). Wloka and Tsotsos have shown that center-bias seems to be a feature of fixed-size image viewing, not present in natural free-head viewing (Wloka & Tsotsos, 2013). In this case, center bias may be natural under the task constraints of no head movements and fixed-size input. In any case, center bias remains an issue to resolve.

These are the types of questions that must be carefully considered when putting together benchmark datasets and evaluation methodologies: what are the possible biases? Are they a property of the data or the task? Should dataset bias be compensated for by testing on different datasets or using appropriate metrics? Should task bias be a property of the models?

In Table 1, we provide some examples of how benchmarking can be applied to different visual attention tasks under the operationalized definitions proposed. A benchmark provides models with a standard dataset of inputs composed of images, task constraints, and system state (e.g. current context). Model output is then evaluated according to standardized evaluation metrics on ground-truth measurements (human fixations, neuronal recordings, etc.). In some cases, the ground-truth measurements are held out and known only by the benchmark curators. This prevents models from being specifically tailored to fit the data. In other cases, part of the ground-truth measurements may be released for training models, but a held-out test set is ultimately what models are evaluated on. This is the case for the MIT Saliency Benchmark (Bylinskii et al., 2014), an online benchmarking website which publicly releases only images, and not the ground-truth human fixations on those images. Model designers submit saliency maps computed on this set of images, and the benchmark curators evaluate the saliency maps against human fixation maps according to 7 standardized metrics (as of this manuscript’s publication date). The website makes available the code for the metrics and for optimization, to allow model designers to re-adjust saliency map parameters to compensate for dataset biases. The website also includes an up-to-date list of other fixation datasets that can be used for model evaluation.

Similarly, for standardizing model comparison in other subareas of visual attention, we propose that benchmark datasets should be constructed: (1) to cover multiple visual attention tasks; (2) to include multiple measures of performance and comparison per task; and (3) to permit model testing on multiple datasets to minimize the effects of dataset bias. The additional benefits of putting benchmarks online are discussed in the following section.

5. Future directions

Established evaluation protocols and benchmark datasets can help organize and categorize the wide range of phenomena and corresponding computational models in visual attention research. While comprehensive reviews are immensely useful for performing this synthesis, we believe that an important next step is to establish a systematic, high-throughput and rapidly-updated evaluation protocol. We propose having online, up-to-date competitions, whereby all model entries are evaluated in the same manner, via the same metrics, on the same data, compared against the same set of models. The latest progress in the field can thus be documented. Such a benchmark already exists for saliency modeling (Bylinskii et al., 2014), but most other subareas of visual attention have no such benchmark available. Benchmarks have been crucial for pushing progress in the computer vision (Deng et al., 2009; Everingham et al., 2012; Lin et al., 2014; Torralba et al., 2008; Xiao et al., 2010) and natural language processing (NIST, 2013; Voorhees, 2004; Voorhees & Harman, 2005) communities by having regularly-updated lists of the best-ranking models. Note that we are not advocating for the reduction of all model evaluation to one summary number, as this is uninformative and a poor measure of success. Instead, as discussed in the previous section, evaluation should be multi-faceted – covering multiple tasks, datasets, and metrics. In such a way, evaluation is better able to capture a broader set of capabilities of a model. Making evaluation quantitative is precisely what will allow us to make comparisons rigorous, serving as a better indicator of progress.

Given the established groundwork and success of computer vision competitions, this evaluation paradigm is most easily implemented by studying attentional effects as human responses to sequences of image stimuli. This suggests that the first-generation visual attention benchmarks should focus on evaluating image-computable models. The ground truth for this evaluation paradigm can be multi-layered, spanning data from single-unit neuronal responses all the way to behavioral outputs. For instance, models of visual search may attempt to predict behavioral patterns of human subjects performing visual search experiments, as well as neurons in the non-human primate attention network that modulate their activity based on search behavior. Already, there are many published results on human and non-human primate visual search that can be used to bootstrap the benchmarking process. Good models of visual search should be capable of explaining as comprehensive a set of the published results as possible.

When consensus about the specific behavioral measures and similarity metrics cannot be attained, or when new measures are introduced, multiple measures can be considered simultaneously. Similarly, the establishment of a competition-style benchmarking system allows for relatively rapid updating with new data and new models. Finally, by augmenting the test data on a yearly basis, established models will be constantly challenged and must prove their value to remain in contention for the position of a reference model.

The computational modeling of visual attention is a rapidly-developing area. In addition to making significant contributions to our understanding of visual attention, these models have made significant impacts on other scientific domains, via applications including efficient image search for object detection and recognition, surveillance, image segmentation, image...
compression, photo retargeting, infographic design, user-interface design, brain–machine interfaces, medical diagnosis, robot navigation, as well as educational and promotional content design (discussed in greater detail by Judd (2011), Borji & Itti (2013)).

The growing interest in applications of computational models of attention is likely, as a byproduct, to stimulate progress in this field. As the number of computational models available continue to increase, it may become increasingly difficult to find order and structure among them, making scientific progress difficult to evaluate. Thus, now is a crucial time to establish standardized evaluation methodologies. In this paper, we have offered some approaches for standardization, borrowing ideas from other computational fields that have proven successful.

6. Origin of this opinion paper

This presentation comes out of the culmination of a seminar course on “Understanding Visual Attention through Computation”. The course details are included in the Appendix. The authors come from five different institutions and represent a total of eight different research labs with diverse backgrounds and research interests, ranging from physiological to behavioral and computational. Their own research areas served as launching pads for initiating discussion about visual attention.

Acknowledgements

We would like to thank the guest lecturers: Salva Ardid, Neil Bruce, Robert Desimone, Mazyar Fallah, Ruth Rosenholtz, Thomas Serre, Antonio Torralba, Jeremy Wolfe, and Thilo Womelsdorf for their time and stimulating presentations.

JKT is grateful for the support of Robert Desimone and James DiCarlo during his sabbatical at MIT, during which this course was presented and this paper developed.

JKT, ZB, and RR acknowledge support from the Natural Sciences and Engineering Research Council of Canada. JKT additionally acknowledges support from the Canada Research Chairs Program. JZ is supported by the China Scholarship Council and National Natural Science Fund of China (Grant No. 61233011).

Appendix A. Models included in Figure 1

A.1. From Kimura et al. (2013)


A.2. From Frintrop et al. (2010)


A.3. Tsotsos and Rothenstein (2011)


Appendix B. Course information: understanding visual attention through computation

A full syllabus for the course taught by the senior author on “Understanding Visual Attention through Computation” can be found online. For more information, please visit the course website or contact the instructor directly.
found online. It is available as a resource and possible inspiration for future courses on visual attention. This highly interdisciplinary course explored many of the different approaches and perspectives in the current literature, within the historical context of research of the field. The intent was to develop a 'big picture' view of what this thing called visual attention might entail and how can we best further deepen our understanding. The course used different themes to explore different viewpoints and theories in order to develop an appreciation of both strengths and weaknesses, not only of the research but also the methodologies. Here we provide an outline of the lecture topics, guest lecturers, and course readings.

Appendix C. Lectures

1. Introduction to Attention, J.K. Tsotsos (Tsotsos, 2011)
5. Attention and Search, J. Wolfe and R. Rosenholtz (Rosenholtz, Kuzmova, & Sherman, 2011; Rosenholtz et al., 2012; Treisman & Gelade, 1980; Treisman, 2006; Wolfe, 1998; Wolfe, 2007; Wolfe, 1994; Wolfe et al., 2011)
6. Neurobiology of Attention, M. Fallback (Baluch & Itti, 2011; Corbetta, Patel, & Shulman, 2008; Fallah, Stoner, & Reynolds, 2007; Krauzlis, Lovejoy, & Zenon, 2013; Maunsell & Treue, 2006; Petersen & Posner, 2012; Salazar et al., 2012; Squire et al., 2013; Sundberg et al., 2006)
7. Saliency Map Models, M.D.B. Bruce (Borji & Itti, 2013; Bruce & Tsotsos, 2009; Koehler et al., 2014; Riche et al., 2013)
8. Selective Tuning Part II, J.K. Tsotsos (Tsotsos, 2011)
9. Dynamical systems models, S. Aridid (Ardid, Wang, & Compte, 2007; Ardid et al., 2010; Borgers, Epstein, & Kopell, 2008; Buia & Tiesinga, 2008; Mante et al., 2013)
10. The Roles of Gist, Context and Task, A. Torralba (Isola et al., 2011; Oliva et al., 2003; Oliva & Torralba, 2006; Torralba et al., 2006; Torralba, 2003)
12. Final Integrative Discussion, J.K. Tsotsos

References


References used in Figure 1


http://dx.doi.org/10.1016/j.visres.2015.04.007


Please cite this article in press as: Bylinskii, Z., et al. Towards the quantitative evaluation of visual attention models. *Vision Research* (2015), [http://dx.doi.org/10.1016/j.visres.2015.04.007](http://dx.doi.org/10.1016/j.visres.2015.04.007)