Learning Visual Importance for Graphic Designs and Data Visualizations

Learning Visual Importance for Graphic Designs and Data Visualizations

Zoya Bylinskii[1] **Nam Wook Kim**[2] **Peter O'Donovan**[3] **Sami Alsheikh**[1] **Spandan Madan**[2]
Hanspeter Pfister[2] **Fredo Durand**[1] **Bryan Russell**[4] **Aaron Hertzmann**[4]
[1] MIT CSAIL, Cambridge, MA USA {zoya,alsheikh,fredo}@mit.edu
[2] Harvard SEAS, Cambridge, MA USA {namwkim,spandan_madan,pfister}@seas.harvard.edu
[3] Adobe Systems, Seattle, WA USA {podonova}@adobe.com
[4] Adobe Research, San Francisco, CA USA {hertzman,brussell}@adobe.com

ABSTRACT

Knowing where people look and click on visual designs can provide clues about how the designs are perceived, and where the most important or relevant content lies. The most important content of a visual design can be used for effective summarization or to facilitate retrieval from a database. We present automated models that predict the relative importance of different elements in data visualizations and graphic designs. Our models are neural networks trained on human clicks and importance annotations on hundreds of designs. We collected a new dataset of crowdsourced importance, and analyzed the predictions of our models with respect to ground truth importance and human eye movements. We demonstrate how such predictions of importance can be used for automatic design retargeting and thumbnailing. User studies with hundreds of MTurk participants validate that, with limited post-processing, our importance-driven applications are on par with, or outperform, current state-of-the-art methods, including natural image saliency. We also provide a demonstration of how our importance predictions can be built into interactive design tools to offer immediate feedback during the design process.

ACM Classification Keywords

H.5.1 Information Interfaces and Presentation: Multimedia Information Systems

Author Keywords

Saliency; Computer Vision; Machine Learning; Eye Tracking; Visualization; Graphic Design; Deep Learning; Retargeting.

INTRODUCTION

A crucial goal of any graphic design or data visualization is to communicate the relative *importance* of different design elements, so that the viewer knows where to focus attention and how to interpret the design. In other words, the design should provide an effective management of attention [39]. Understanding how viewers perceive a design could be useful for many stages of the design process; for instance, to provide feedback [40]. Automatic understanding can help build tools to search, retarget, and summarize information in designs and visualizations. Though saliency prediction in natural images has recently become quite effective, there is little work in importance prediction for either graphic designs or data visualizations.

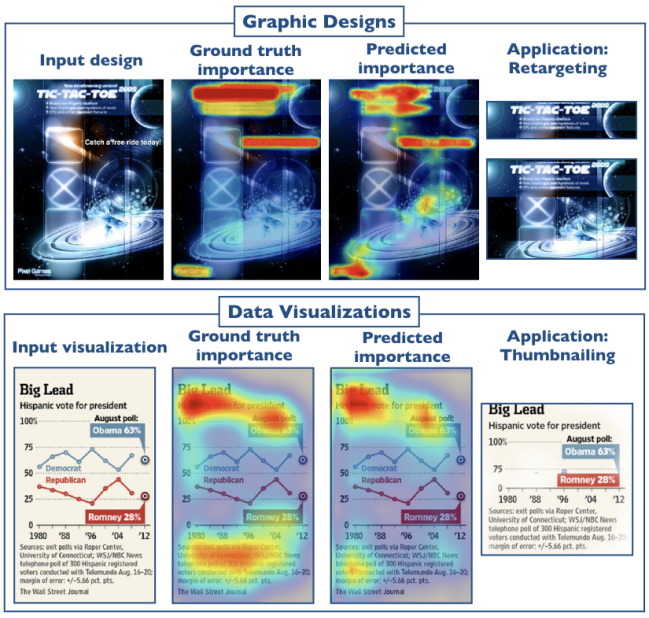

Figure 1. We present two neural network models trained on crowdsourced importance. We trained the graphic design model using a dataset of 1K graphic designs with GDI annotations [33]. For training the data visualization model, we collected mouse clicks using the BubbleView methodology [22] on 1.4K MASSVIS data visualizations [3]. Both networks successfully predict ground truth importance and can be used for applications such as retargeting, thumbnailing, and interactive design tools. Warmer colors in our heatmaps indicate higher importance.

UIST 2017, October 22–25, 2017, Quebec City, QC, Canada
 ISBN 978-1-4503-4981-9/17/10…$15.00
DOI: https://doi.org/10.1145/3126594.3126653

Our online demo, video, code, data, trained models, and supplemental material are available at visimportance.csail.mit.edu.

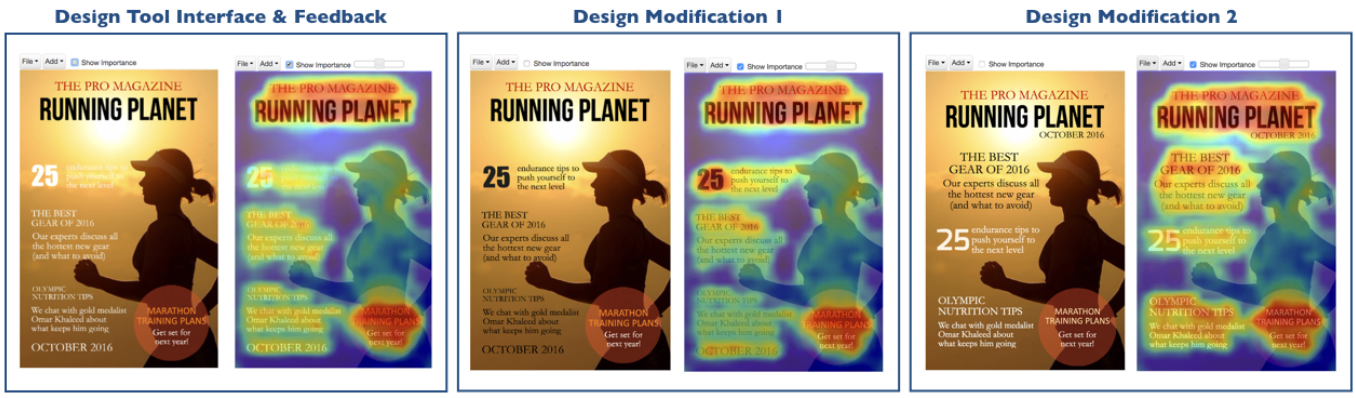

Figure 2. We show an interactive graphic design application using our model that lets users change and visualize the importance values of elements. Users can move and resize elements, as well as change color, font, and opacity, and see the updated realtime importance predictions. For instance, a user changes the color of the text to the left of the runner to increase its importance (middle panel). The rightmost panel includes a few additional changes to the size, font, and placement of the text elements to modify their relative importance scores. A demo is available at visimportance.csail.mit.edu.

We use "importance" as a generic term to describe the perceived relative weighting of design elements. Image saliency, which has been studied extensively, is a form of importance. However, whereas traditional notions of saliency refer to bottom-up, pop-out effects, our notion of importance can also depend on higher-level factors such as the semantic categories of design elements (e.g., title text, axis text, data points).

This paper presents a new importance prediction method for graphic designs and data visualizations. We use a state-of-the-art deep learning architecture, and train models on two types of crowdsourced importance data: graphic design importance (GDI) annotations [33] and a dataset of BubbleView clicks [22] we collected on data visualizations.

Our importance models take input designs in bitmap form. The original vector data is not required. As a result, the models are agnostic to the encoding format of the image and can be applied to existing libraries of bitmap designs. Our models pick up on some of the higher-level trends in ground truth human annotations. For instance, across a diverse collection of visualizations and designs, our models learn to localize the titles and correctly weight the relative importance of different design elements (Fig. 1).

We show how the predicted importance maps can be used as a common building block for a number of different applications, including retargeting and thumbnailing. Our predictions become inputs to cropping and seam carving with almost no additional post-processing. Despite the simplicity of the approach, our retargeting and thumbnailing results are on par with, or outperform, related methods, as validated by a set of user studies launched on Amazon's Mechanical Turk (MTurk). Moreover, an advantage of the fast test-time performance of neural networks makes it feasible for our predictions to be integrated into interactive design tools (Fig. 2). With another set of user studies, we validate that our model generalizes to fine-grained design variations and correctly predicts how importance is affected by changes in element size and location on a design.

Contributions: We present two neural network models for predicting importance: in graphic designs and data visualizations. This is the first time importance prediction is introduced for data visualizations. For this purpose, we collected a dataset of BubbleView clicks on 1,411 data visualizations. We also show that BubbleView clicks are related to explicit importance annotations [33] on graphic designs. We collected importance annotations for 264 graphic designs with fine-grained variations in the spatial arrangement and sizes of design elements. We demonstrate how our importance predictions can be used for retargeting and thumbnailing, and include user studies to validate result quality. Finally, we provide a working interactive demo.

RELATED WORK

Designers and researchers have long studied eye movements as a clue to understanding the perception of interfaces [9, 16]. There have also been several recent studies of eye movements and the perception of designs [2, 12]. However, measuring eye movements is an expensive and time-consuming process, and is rarely feasible for practical applications.

Few researchers have attempted to automatically predict importance in graphic designs. The DesignEye system [40] uses hand-crafted saliency methods, demonstrating that saliency methods can provide valuable feedback in the context of a design application. O'Donovan et al. [33] gather crowdsourced importance annotations, where participants are asked to mask out the most important design regions. They train a predictor from these annotations. However, their method requires knowledge of the location of design elements to run on a new design. Haass et al. [11] test three natural image saliency models on the MASSVIS data visualizations [3], concluding that, across most saliency metrics, these models perform significantly worse on visualizations than on natural images. Several models also exist for web page saliency. However, most methods use programmatic elements (e.g., the DOM) as input to saliency estimation rather than allowing bitmap images as input [4, 47]. Pang et al. predict the order in which people will look at components on a webpage [36] by making

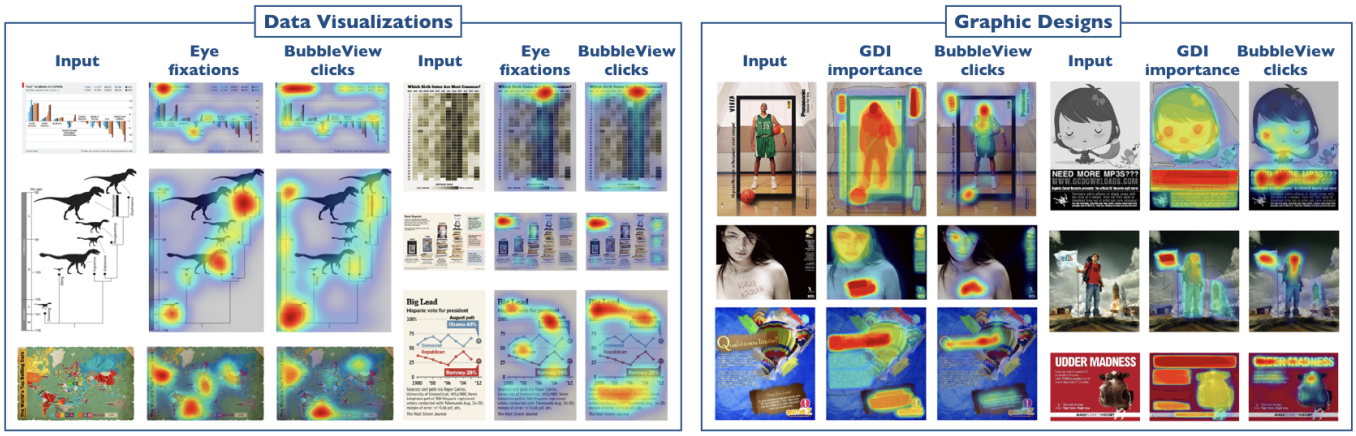

Figure 3. Left: Comparison of eye movements collected in a controlled lab setting [2], and clicks that we crowdsourced using the BubbleView interface [22, 23]. Right: Comparison of importance annotations from the GDI dataset, and clicks that we crowdsourced using the BubbleView interface. These examples were chosen to demonstrate some of the similarities and differences between the modalities. For instance, compared to eye fixations, clicks are sometimes more concentrated around text. Compared to the GDI annotations, clicks do not assign uniform importance to whole design elements. Despite these differences, BubbleView data leads to similar importance rankings of visual elements (*Evaluation*).

use of the DOM and manual segmentations. Other works use the web page image itself as input to predict saliency [43, 45]. Most of these methods use older saliency architectures based on hand-crafted features that are inferior to the state-of-the-art neural networks we use in our approach. Our work also relates to the general program of applying computer vision and machine learning in the service of graphic design tools [27, 28, 42].

Predicting eye movements for natural images is a classic topic in human and computer vision. The earliest natural image saliency methods relied on hand-coded features (e.g., [15]). Recently, deep learning methods, trained on large datasets, have produced a substantial jump in performance on standard saliency benchmarks [7, 8, 14, 29, 35, 49]. However, these methods have been developed exclusively for analyzing natural images, and are not trained or tested on graphic designs. Our work is the first to apply neural network importance predictors to both graphic designs and data visualizations.

DATA COLLECTION

To train our models we collected BubbleView data [22, 23] for data visualizations, and used the Graphic Design Importance (GDI) dataset by O'Donovan et al. [33] for graphic designs. We compared different measurements of importance: BubbleView clicks to eye movements on data visualizations, and BubbleView clicks to GDI annotations on graphic designs.

Ground truth importance for data visualizations

Large datasets are one of the prerequisites to train neural network models. Unfortunately, collecting human eye movements for even hundreds of images is extremely expensive and time-consuming. Instead, we use the BubbleView interface by Kim et al. [22, 23] to record human "attention" that is correlated with eye fixations. Unlike eye tracking, which requires expensive equipment and a controlled lab study, BubbleView can be used to to collect large datasets with online crowdsourcing.

In BubbleView, a participant is shown a blurry image and can click on different parts of the image to reveal small regions, or *bubbles*, of the image at full resolution. Initial experiments by Kim et al. [23] showed a high correlation between eye fixations collected in the lab and crowdsourced BubbleView click data. In this paper, we confirm this relationship.

Concurrent work in the computer vision community has applied a similar methodology to natural images. SALICON [18] is a crowdsourced dataset of mouse movements on natural images that has been shown to approximate free-viewing eye fixations. Current state-of-the-art models on saliency benchmarks have all been trained on the SALICON data [8, 14, 29, 35, 49]. BubbleView was concurrently developed [23] to approximate eye fixations on data visualizations with a description task. Some advantages of BubbleView over SALICON are discussed in [22].

Using Amazon's Mechanical Turk (MTurk), we collected BubbleView data on a set of 1,411 data visualizations from the MASSVIS dataset [3], spanning a diverse collection of sources (news media, government publications, etc.) and encoding types (bar graphs, treemaps, node-link diagrams, etc.). We manually filtered out visualizations containing illegible and non-English text, as well as scientific and technical visualizations containing too little context. Images were scaled to have a maximum dimension of 600 pixels to a side while maintaining their aspect-ratios to fit inside the MTurk task window. We blurred the visualizations using a Gaussian filter with a radius of 40 pixels and used a bubble size with a radius of 32 pixels as in [22]. MTurk participants were additionally required to provide descriptions for the visualizations to ensure that they meaningfully explored each image. Each visualization was shown to an average of 15 participants. We aggregated the clicks of all participants on each visualization and blurred the click locations with a Gaussian filter with a radius of 32 pixels, to match the format of the eye movement data.

We used the MASSVIS eye movement data for testing our importance predictions. Fixation maps were created by aggregating eye fixation locations of an average of 16 participants viewing each visualization for 10 seconds. Fixation locations were Gaussian filtered with a blur radius of 32 pixels. Fig. 3a includes a comparison of the BubbleView click maps to eye fixation maps from the MASSVIS dataset.

Ground truth importance for graphic designs

We used the Graphic Design Importance (GDI) dataset [33] which comes with importance annotations for 1,078 graphic designs from Flickr. Thirty-five MTurk participants were asked to label important regions in a design using binary masks, and their annotations were averaged. Participants were not given any instruction as to the meaning of "importance." To determine how BubbleView clicks relate to explicit importance annotations, we ran the BubbleView study on these graphic designs and collected data from an average of 15 participants per design. Fig. 3b shows comparisons between the GDI annotations and BubbleView click maps. In both data similar elements and regions of designs emerge as important.

Each representation has potential advantages. The GDI annotations assign a more uniform importance score to whole elements. This can serve as a soft segmentation to facilitate design applications like retargeting. BubbleView maps may be more appropriate for directly modeling human attention.

MODELS FOR PREDICTING IMPORTANCE

Given a graphic design or data visualization, our task is to predict the importance of the content at each pixel location. We assume the input design/visualization is a bitmap image. The output importance prediction at each pixel i is $P_i \in [0,1]$, where larger values indicate higher importance. We approach this problem using deep learning, which has lead to many recent breakthroughs on a variety of image processing tasks in the computer vision community [25, 38], including the closely related task of saliency modeling.

Similar to some top-performing saliency models for natural images [14, 26], our architecture is based on fully convolutional networks (FCNs) [32]. FCNs are specified by a directed acyclic graph of linear (e.g., convolution) and nonlinear (e.g., max pool, ReLU) operations over the pixel grid, and a set of parameters for the operations. The network parameters are optimized over a loss function given a labeled training dataset. We refer the reader to Long et al. [32] for more details.

We predict real-valued importance using a different training loss function from the original FCN work, which predicted discrete object classes. Given ground truth importances at each pixel i, $Q_i \in [0,1]$, we optimize the sigmoid cross entropy loss for FCN model parameters Θ over all pixels $i = 1,\ldots,N$:

$$L(\Theta) = -\frac{1}{N}\sum_{i=1}^{N}\left(Q_i \log P_i + (1-Q_i)\log(1-P_i)\right) \qquad (1)$$

where $P_i = \sigma\left(f_i(\Theta)\right)$ is the output prediction of the FCN $f_i(\Theta)$ composed with the sigmoid function $\sigma(x) = (1+\exp(-x))^{-1}$. Note that the same loss is used for binary classification, where $Q_i \in \{0,1\}$. Here, we extend it to real-valued $Q_i \in [0,1]$.

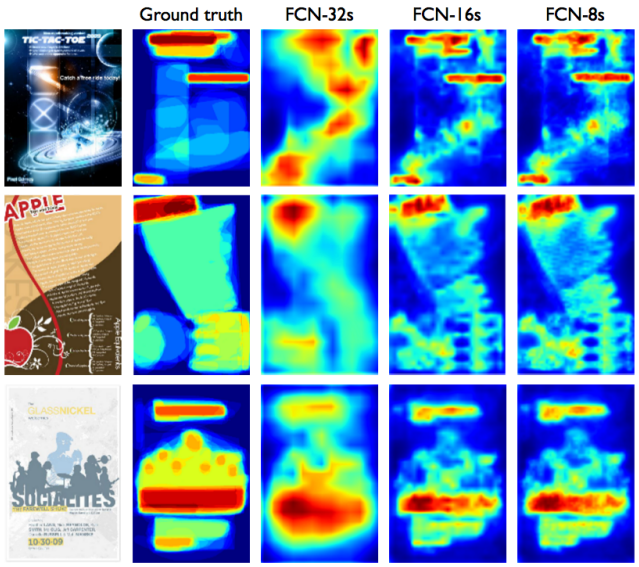

Figure 4. We increase the precision of our FCN-32s predictions by combining output from the final layer of the network with outputs from lower levels. The resulting predictions, FCN-16s and FCN-8s, capture finer details. We found FCN-16s sufficient for our model for graphic designs, as FCN-8s did not add a performance boost. For our model for data visualizations, we found no performance gains beyond FCN-32s.

We use a different loss than other saliency models based on neural networks that optimize Euclidean [26, 34], weighted Euclidean [8], or binary classification losses [29, 49]. Our loss is better suited to $[0,1]$ values, and is equivalent to optimizing the KL loss commonly used for saliency evaluation.

We trained separate networks for data visualizations and for graphic designs. For the data visualizations, we split the 1.4K MASSVIS images for which we collected BubbleView click data into 1,209 training images and 202 test images. For the test set we chose MASSVIS images for which eye movements are available [2]. For the graphic designs, we split the 1,078 GDI images into 862 training images and 216 test images (80-20% split). We used the GDI annotations [33] for training. We found that training on the GDI annotations rather than the BubbleView clicks on graphic designs facilitated the design applications better, since the GDI annotations were better aligned to element boundaries.

Model details: We converted an Oxford VGG-16 convolutional neural network [44] to an FCN-32s model via network surgery using the implementation in Caffe [17]. The model's predictions are 1/32 of the input image resolution, due to successive pooling layers. To increase the resolution of the predictions and capture fine details, we followed the procedure in Long et al. [32] to add skip connections from earlier layers to form FCN-16s and FCN-8s models, that are respectively, 1/16 and 1/8 of the input image resolution. We found that the FCN-16s (with a single skip connection from *pool4*) improved the graphic design importance maps relative to the FCN-32s model (Fig. 4), but that adding an additional skip connection from *pool3* (FCN-8s) performed similarly. We found that skip

connections lead to no gains for the data visualization importance. For our experiments we used the trained FCN-16s for graphic designs and the FCN-32s for data visualizations.

Since we have limited training data we initialized the network parameters with the pre-trained FCN32s model for semantic segmentation in natural images [32], and fine-tuned it for our task. The convolutional layers at the end of the network and the skip connections were randomly initialized. Training details are provided in the Supplemental Material.

We opted for a smaller architecture with fewer parameters than some other neural network saliency models for natural images. This makes our model more effective for our datasets, which are currently an order-of-magnitude smaller than the natural image saliency datasets.

EVALUATION OF MODEL PREDICTIONS

We compare the performance of our two importance models to ground truth importance on each dataset. For data visualizations, we compare predicted importance maps to bubble clicks gathered using BubbleView, and to eye fixations from the MASSVIS dataset. For graphic designs, we compare predicted importance maps to GDI annotations.

Evaluation criteria

We evaluate the similarity of our predicted and ground truth importance maps using two metrics commonly used for saliency evaluation [6]: Kullback-Leibler divergence (KL) and cross correlation (CC). CC measures how correlated the pixel-wise values are in the two maps, and treats both false positives and false negatives equally. KL, however, measures how well one distribution predicts another. Our importance maps can be interpreted as providing, for each pixel, the probability that the pixel would be considered important by ground truth observers. KL highly penalizes missed predictions, so a sparse map that fails to predict a ground truth important location will receive a large KL value (poor score). Given the ground truth importance map Q and the predicted importance map P, KL is computed as:

$$KL(P,Q) = \sum_{i=1}^{N} \left(Q_i \log Q_i - Q_i \log P_i\right) = L(P,Q) - H(Q), \quad (2)$$

where $H(Q) = -\sum_{i=1}^{N} \left(Q_i \log Q_i\right)$ is the entropy of the ground truth importance map and $L(P,Q)$ is the cross entropy of the prediction and ground truth. Note the similarity to the loss in Equation (1), which is over a Bernoulli random variable; here the random variable is instantiated. A large KL divergence indicates a high dissimilarity between maps, whereas $KL(P,Q) = 0$ indicates two maps are identical. KL is in principle unbounded, so to provide a feasible range, we include chance baselines in our experiments. CC is computed as:

$$CC(P,Q) = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(P_i - \bar{P}\right)\left(Q_i - \bar{Q}\right)}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(P_i - \bar{P}\right)^2}\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(Q_i - \bar{Q}\right)^2}}, \quad (3)$$

where $\bar{P} = \frac{1}{N}\sum_{i=1}^{N} P_i$, and respectively for Q. CC ranges from -1 to 1, where 1 indicates maximal correlation between two maps P and Q. For further intuition about how KL and CC metrics score similarity, we provide scores above each image in Fig. 5, and additional examples in the Supplemental Material, showing high- and low-scoring predictions.

Model	CC score ↑	KL score ↓
Chance	0.00	0.75
Judd [21]	0.11	0.49
DeepGaze [29]	0.57	3.48
Our model	**0.69**	**0.33**

Table 1. How well can our importance model predict the BubbleView click maps? We add comparisons to two other top-performing saliency models and a chance baseline. Scores are averaged over 202 test data visualizations. A higher CC score and lower KL score are better.

Model	CC score ↑	KL score ↓
Chance	0.00	1.08
Judd [21]	0.19	0.74
DeepGaze [29]	0.53	3.10
Our model	**0.54**	**0.63**
Bubble clicks	0.79	0.28

Table 2. How well can human eye fixations be predicted? We measured the similarity between human fixation maps and various predictors. Scores are averaged over 202 test data visualizations. A good model achieves a high CC score and low KL score. Our neural network model was trained on BubbleView click data, so that is the modality it can predict best. Nevertheless, its predictions are also representative of eye fixation data. As an upper bound on this prediction performance, we consider how well the BubbleView click data predicts eye fixations, and as a lower bound, how well chance predicts eye fixations.

Prediction performance on data visualizations

We include predictions from our importance model in Fig. 5. Notice how we correctly predict important regions in the ground truth corresponding to titles, captions, and legends. We quantitatively evaluate our approach on our collected dataset of BubbleView clicks. We report CC and KL scores averaged over our dataset of 202 test images in Table 1.

We compare against the following baselines: chance, Judd saliency [21], and DeepGaze [29], a top neural network saliency model trained on the SALICON dataset [18] of mouse movements on natural images. The chance baseline, used in saliency benchmarks [6, 20], is computed by uniformly sampling a real value between 0 and 1 at each image pixel. Our approach out-performs all baselines. KL is highly sensitive to false negatives and drastically penalizes sparser models [6][1], explaining the high KL values for DeepGaze in Table 1. Post-processing or directly optimizing models for specific metrics can yield more favorable performances [30].

How well does our neural network model, trained on clicks, predict eye fixations? We find that the predicted importance is representative of eye fixation patterns as well (Table 2), although the difference in scores indicates that our model might be learning from patterns in the click data that are different from fixations.

[1]Because of the sensitivity of KL to output regularization, we advise against using it (solely) to compare models [6].

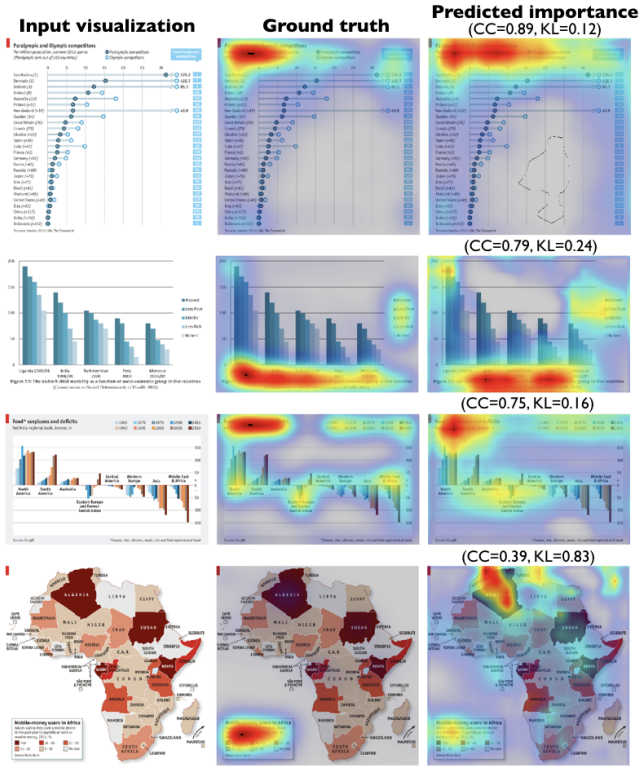

Figure 5. Importance predictions for data visualizations, compared to ground truth BubbleView clicks and sorted by performance. Our model is good at localizing visualization titles (the element clicked on, and gazed at, most by human participants) as well as picking up the extreme points on graphs (e.g., top and bottom entries). We include a failure case where our model overestimates the importance of the visual map regions. More examples in the Supplemental Material.

Which elements are most important? For our analysis, we used the element segmentations available for the visualizations in the MASSVIS dataset [2]. We overlapped these segmentations with normalized maps of eye fixations, clicks, and predicted importance. We computed the max score of the map within each element to get an importance ranking across elements[2]. Text elements, such as titles and captions, were the most looked at[3], and clicked on, elements, and were also predicted most important by our model (Fig. 6). Even though our model was trained on BubbleView clicks, the predicted importance remains representative of eye fixation patterns. With regards to differences, our model overpredicts the importance of titles. Our model learns to localize visualization titles very well (Fig. 5).

Prediction performance on graphic designs

The closest approach to ours is the work of O'Donovan et al. [33] who computed an importance model for the GDI dataset. We re-ran their baseline models on the train-test split used for our model (Table 3). To replicate their evaluation, we

[2]A similar analysis was used to rank the relative importance of objects in natural images [7, 18].

[3]Among the text and other content in a visualization, titles tend to be best remembered by human observers [2].

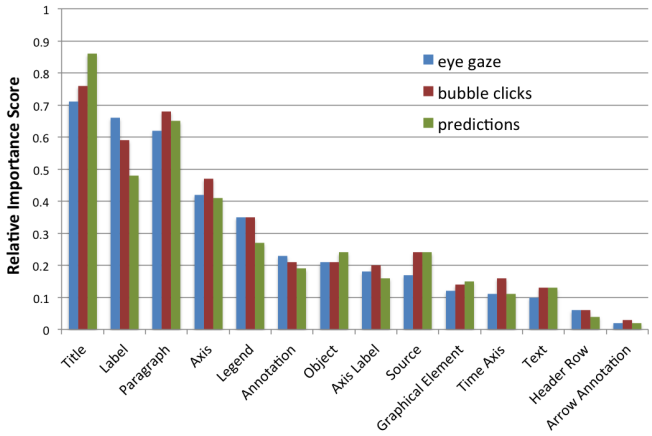

Figure 6. Relative importance scores of different elements in a data visualization assigned by eye fixation maps, BubbleView click maps, and model predictions. Scores were computed by overlapping element segmentations with normalized importance maps, and taking the max of the map within each element as its score. The elements that received the most clicks also tended to be highly fixated during viewing (Spearman's $r_s = .96, p < .001$). Text (titles, labels, paragraphs) received a lot of attention. Our neural network model correctly predicted the relative importance of these regions relative to eye movements ($r_s = .96, p < .001$).

report root-mean-square error (RMSE) and the R^2 coefficient, where $R^2 = 1$ indicates a perfect predictor, and $R^2 = 0$ is the baseline of predicting the mean importance value (details in Supplemental Material). The full O'Donovan model (*OD-Full*) requires manual annotations of *text*, *face*, and *person* regions, and would not be practical in an automatic setting. For a fair comparison, we evaluate our automatic predicted importance model (*Ours*) against the automatic portion of the O'Donovan model, which does not rely on human annotations (*OD-Automatic*). We find that our model outperforms *OD-Automatic*. Our model is also 100X faster, since it requires a single feed-forward pass through the network (~0.1 s/image on a GPU). O'Donovan's method requires separate computations of multiple CPU-based saliency models and image features (~10 s/image at the most efficient setting).

In Table 3, we include the performance of *Ours+OD*, where we added our importance predictions as an additional feature during training of the O'Donovan model, and re-estimated the optimal weights for combining all the features. *Ours+OD* improves upon *OD-Full* indicating that our importance predictions are not fully explainable by the existing features (e.g., text or natural image saliency). This full model is included for demonstration purposes only, and is not practical for interactive applications.

We also annotated elements in each of the test graphic designs using bounding boxes, and computed the maximum importance value in each bounding box as the element's score (Fig. 7). We obtain an average Spearman rank correlation of 0.56 between the predicted and ground truth scores assigned to the graphic design elements.

Some examples of predictions are included in Fig. 8. Our predictions capture important general trends, such as larger

Model	RMSE ↓	R^2 ↑
Saliency	.229	.462
OD-Automatic	.212	.539
Ours	**.203**	**.576**
OD-Full	.155	.754
Ours+OD	**.150**	**.769**

Table 3. A comparison of our predicted importance model (*Ours*) with the model of O'Donovan et al. [33]. Lower *RMSE* and higher R^2 are better. Our model outperforms the fully automatic O'Donovan variant (*OD-Automatic*). Another fully automatic model from [33] is *Saliency*, a learned combination of 4 saliency models: Itti&Koch [15], Hou&Zhang [13], Judd et al. [21], and Goferman et al. [10]. We also report the results of the semi-automatic OD-Full model, which includes manual annotations of *text*, *face*, and *person* regions. When we combine our approach with OD-Full (*Ours+OD*), we can approve upon the OD model. More comparisons are included in the Supplemental Material.

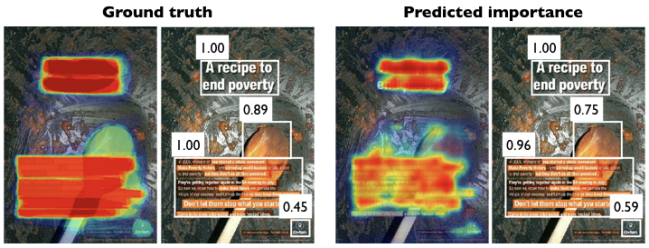

Figure 7. An example comparison between the predicted importance of design elements and the ground truth GDI annotations. The heatmaps are overlapped with element bounding boxes and the maximum score per box is used as the element's importance score (between 0 and 1).

and more central text and visual elements being more important. However, text regions are not always well segmented (predicted importance is not uniform over a text element), and text written in unusual fonts is not always detected. Such problems could be ameliorated through training on larger datasets. Harder cases are directly comparing the importance of a visual and text, which can depend on the semantics of the text itself (how informative it is) and the quality of the visual (how unexpected, aesthetic, etc.).

Prediction performance on fine-grained design variations

To check for feasibility of an interactive application we perform a more fine-grained test. We want the importance rankings of elements to be adjusted accurately when the user makes changes to their current design. For example, if the user makes a text box larger, then its importance should not go down in the ranking. Our predicted importance model has not been explicitly trained on systematic design variations, so we test if it can generalize to such a setting.

We used the Design Improvement Results dataset [33] containing 11 designs with an average of 35 variants. Across variants, the elements are preserved but the location and scale of the elements varies. We repeated the MTurk importance labeling task of O'Donovan et al. [33] on a subset of 264 design variants, recruiting an average of 19 participants to annotate the most important regions on each design. We averaged all participant annotations per design to obtain ground truth importance heatmaps. We segmented each design into elements

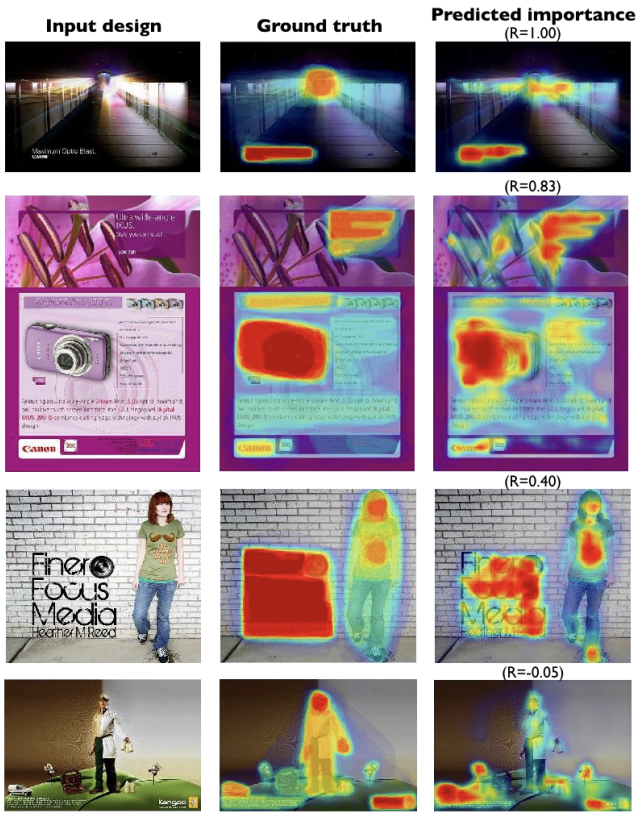

Figure 8. Importance predictions for graphic designs, sorted by performance. Performance is measured as the Spearman rank correlation (R) between the importance scores assigned to design elements by ground truth (GDI annotations) and predicted importance maps. A score of 1 indicates a perfect rank correlation; a negative score indicates the element rankings are reversed. The predicted importance maps distribute importance between text and visual features. We include a failure case where the importance of the man in the design is underestimated. More examples in the Supplemental Material.

and used the ground truth and predicted importance heatmaps to assign importance scores to all the elements, calculating the maximum heatmap value falling within each segment. The predicted and ground truth importance scores assigned to these elements achieved an average Spearman's correlation $r_s = .53$. As Fig. 9 shows, even though we make some absolute errors, we successfully account for the impact of design changes such as the location and size of various elements.

APPLICATIONS

We now demonstrate how automatic importance prediction can enable diverse applications. An importance map can provide a common building block for different summarization and retrieval tasks, including retargeting, thumbnailing, and interactive design tools. These prototypes are meant as proofs-of-concept, showing that our importance prediction alone can give good results with almost no additional post-processing.

Retargeting

The retargeting task is to take a graphic design as input, and to produce a new version of that design with specific dimensions.

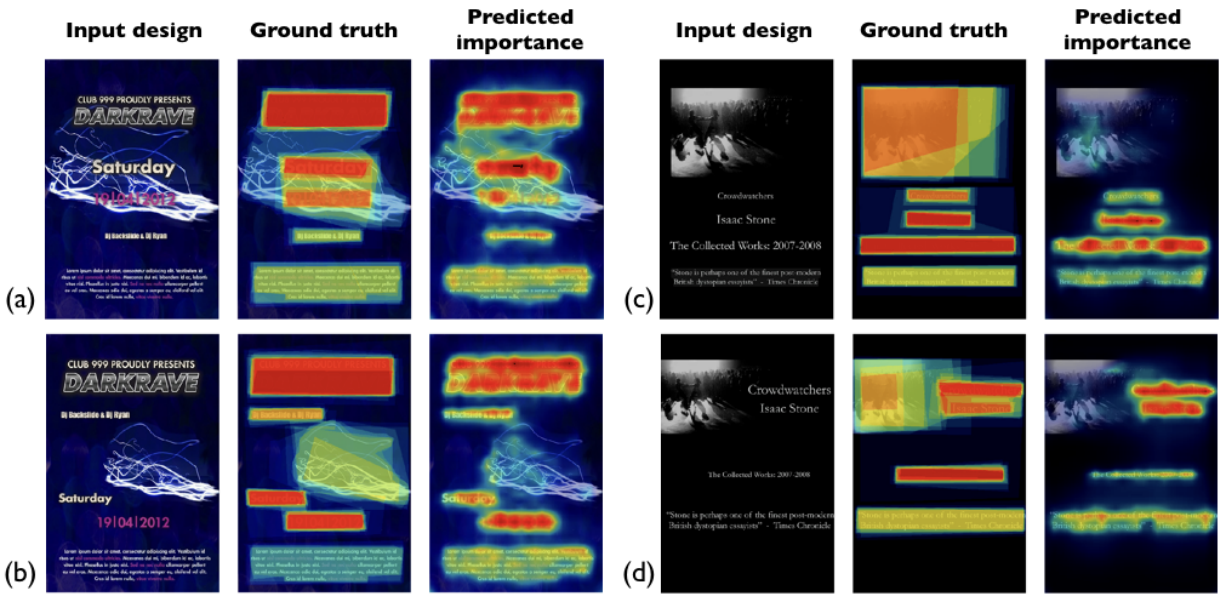

Figure 9. Sample input designs, and how the relative importance of the different design elements changes as they are moved, resized, and otherwise modified. For instance, compared to in (a), the event date stands out more and gains importance when it occurs at the bottom of the poster, in large font, on a contrasting background (b). Similarly, when the most important text of the design in (c) is moved to the upper righthand corner where it is not surrounded by other text, it gains prominence (d). Our automatic model makes similar predictions of the relative importance of design elements as ground truth human annotations.

Retargeting is a common task for modern designers, who must work with many different output dimensions. There is a substantial amount of work on automatic retargeting for natural images, e.g., [1, 41]. Several of these methods have shown that saliency or gaze provide good cues for retargeting, to avoid cropping out image content that people are likely to pay most attention to, such as faces in photographs.

The only previous work on retargeting graphic designs is by O'Donovan et al. [33]. They assumed knowledge of the underlying vector representation of the design and used an expensive optimization with many different energy terms. The method we propose uses bitmap data as input, and is much simpler, without requiring any manual annotations of the input image.

Importance-based retargeting for graphic designs should preserve the most important regions of a design, such as the title and key visual elements. Given a graphic design bitmap as input and specific dimensions, we use the predicted importance map to automatically select a crop of the image with highest importance (Fig. 10). Alternative variants of retargeting (e.g., seam carving) are discussed in the Supplemental Material.

Evaluation: We ran MTurk experiments where 96 participants were presented with a design and 6 retargeted variants, and were asked to score each variant using a 5-point Likert scale with 1 = very poor and 5 = very good (Fig. 11). Each participant completed this task for 12 designs: 10 randomly selected from a collection of 216 designs, and another 2 designs used for quality control. We used this task to compare crops retargeted using predicted importance to crops retargeted using ground truth GDI annotations, Judd saliency [21], DeepGaze saliency [29], and an edge energy map. We extracted a crop with an aspect ratio of 1:4 from a design using the highest-valued region, as assigned by each of the saliency/importance maps. As a baseline, we selected a random crop location.

After an analysis of variance showed a significant effect of retargeting method on score, we performed Bonferonni paired t-tests on the scores of different methods. Across all 216 designs, crops obtained using ground truth GDI annotations had the highest score (Mean: 3.19), followed by DeepGaze (Mean: 2.95) and predicted importance (Mean: 2.92). However, the difference between the latter pair of models was not statistically significant. Edge energy maps (Mean: 2.66) were worse, but not significantly; while Judd saliency (Mean: 2.47) and the random crop baseline (Mean: 2.23) were significantly worse in pairwise comparisons with all the other methods ($p < .01$ for all pairs). Results of additional experimental variants are reported in the Supplemental Material.

Our predicted importance outperforms Judd saliency, a natural saliency model commonly used for comparison [31, 33]. Judd saliency has no notion of text. Predicted importance, trained on less than 1K graphic design images, performs on par with DeepGaze, the currently top-performing neural network-based saliency model [5] which has been trained on 10K natural images, including images with text. Both significantly outperform the edge energy map, which is a common baseline for retargeting. These results show the potential use case of predicted importance for a retargeting task, even without any post-processing steps.

Thumbnailing

Thumbnailing is similar to retargeting, but with a different goal. It aims to provide a visual summary for an image to make it easier to find relevant images in a large collection [19, 46]. Unlike previous methods, our approach operates

directly on a bitmap input, rather than requiring a specialized representation as input. For this example our domain is data visualizations rather than graphic designs.

Given a data visualization and an automatically-computed importance map as input, we generate a thumbnail by carving out the less important regions of the image. The importance map is used as an energy function, whereby we iteratively remove image regions with least energy first. Rows and columns of pixels are removed until the desired proportions are achieved, in this case a square thumbnail. This is similar to seam carving [1, 41], but using straight seams, found to work better in our setting. The boundaries of the remaining elements are blurred using the importance map as an alpha-mask with a fade to white. Qualitatively, the resulting thumbnails consist of titles and other main supporting text, as well as data extremes (from the top and bottom of a table, for instance, or from the left and right sides of a plot).

Evaluation: We designed a task intended to imitate a search through a database of visualizations. MTurk participants were given a description and a grid of 60 thumbnails, and were instructed to find the visualization that best matches the description. We ran two versions of the study: with the original visualizations resized to thumbnails (Fig. 12a), and another with our automatically-computed importance-based thumbnails (Fig. 12b). We measured how many clicks it took for participants to find the visualization corresponding to the description in each version.

A total of 400 participants were recruited for our study. After filtering, we compared the performance of 200 participants who performed the study with resized visualizations and 169 participants who saw the importance-based thumbnails.

Each MTurk assignment, containing a single search task assigned to a single participant, was treated as a repeated observation. We ran an unpaired two-sample t-test to compare the task performance of both groups. On average, participants found the visualization corresponding to the description in fewer clicks using the importance-based thumbnails (1.96 clicks) versus using the resized visualizations (3.25 clicks, $t(367) = 5.10$, $p < .001$). Our importance-based thumbnails facilitated speedier retrieval, indicating that the thumbnails captured visualization content relevant for retrieval.

Interactive applications

An attractive aspect of neural network models is their fast runtime performance (Table 4). As a prototype, we integrated our importance prediction with a simple design layout tool that allows users to move and resize elements, as well as change color, text font, and opacity (Fig. 2). With each change in the design, an importance map is recomputed automatically to provide immediate feedback to the user. The accompanying video and demo (`visimportance.csail.mit.edu`) demonstrate the interactive capabilities of our predictions. Our experiments in the *Evaluation* section provide initial evidence that our model can generalize to the kind of fine-grained design manipulations, like the resizing and relocation of design elements, that would be common in an interactive setting. Determining how best to use importance prediction to provide feedback to users is an interesting problem for future work. For example, importance prediction could help in formulating automatic suggestions for novice users to improve their designs.

Timing: On a Titan-X GPU, our model computes the importance map for a design in the GDI dataset (600×450 pixels) in 100 ms. Table 4 provides some timing information for our model on differently-sized images.

Image size (pixels/side)	300	600	900	1200	1500
Avg. compute time (ms)	46	118	219	367	562

Table 4. Time (in milliseconds) taken by our model to compute an importance map for differently-sized images, averaged over 100 trials.

LIMITATIONS

Our neural network model is only as good as the training data we provide it. In the case of data visualizations, there is a strong bias, both by the model and the ground truth human data, to focus on the text regions. This behavior might not generalize to other types of visualizations and tasks. Click data, gathered via the BubbleView interface, is not uniform over elements, unlike explicit bounding box annotations (i.e., as in the GDI dataset [33]). While this might be a better approximation to natural viewing, non-uniform importance across design elements might cause side-effects for downstream applications like thumbnailing, by cutting off parts of elements or text.

CONCLUSIONS

We curated hundreds of examples of graphic designs and data visualizations, annotated with importance, to train fully convolutional neural network models to predict importance maps for novel designs. We showed that our computational predictions approximate ground truth human data enough to be used for a number of automatic applications. Our importance maps act as a common underlying representation for retargeting of graphic designs, thumbnailing of data visualizations, and in a prototype interactive design application.

This paper presents the first neural network model for predicting saliency or importance in graphic designs and data visualizations, capable of generalizing to a wide range of design formats. Moreover, the fast test-time performance of our model makes it feasible for the predictions to be used in interactive design tools. Our approach is not limited to graphic designs and data visualizations. The methodology and models can easily be adapted to other visual domains, such as websites [22]. As better webcam-based eyetracking methods become available (e.g., [24, 37, 48]) possibilities also open up for directly training our model from eye movement data. Future work can also explore the use of importance predictions to offer more targeted design feedback and to provide automated suggestions to a user.

ACKNOWLEDGEMENTS

We would like to thank Joel Brandt and the anonymous reviewers for useful feedback. Thank you to Matthias Kümmerer and Adrià Recasens for help computing saliency models. ZB would like the acknowledge the support of the Natural Sciences and Engineering Research Council of Canada Postgraduate Doctoral Scholarship (NSERC PGS-D).

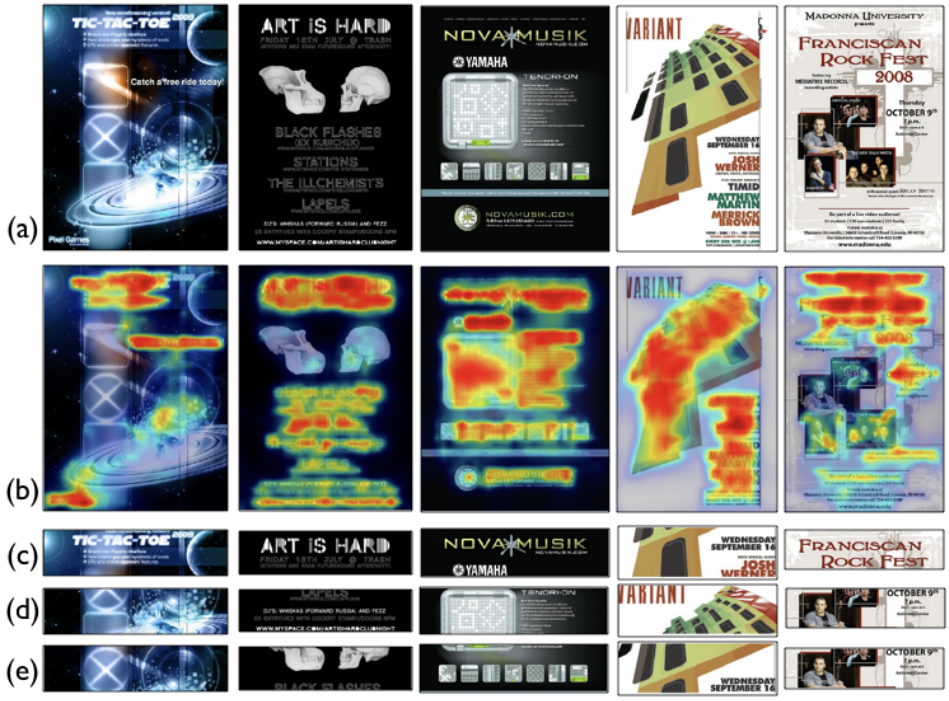

Figure 10. (a) Input designs, (b) our predicted importance maps, and (c) automatic retargeting results using the predicted importance maps to crop out design regions with highest overall importance. This is compared to: (d) edge-based retargeting, where gradient magnitudes are used as the energy map, and (e) Judd saliency, a commonly-used natural image saliency model. Additional comparisons are provided in the Supplemental Material.

Figure 11. MTurk interface for evaluating retargeting results of predicted importance compared to other baselines. More experimental details are provided in the Supplemental Material.

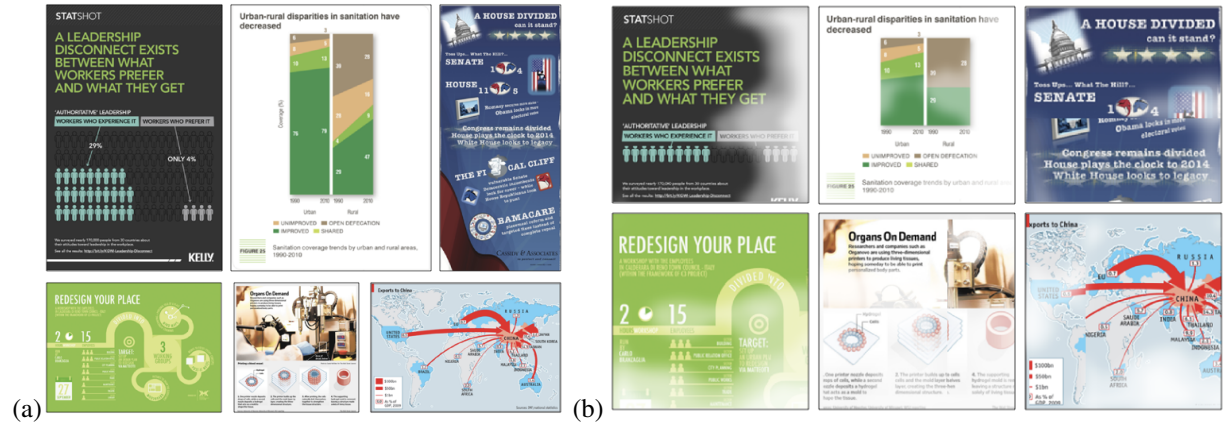

Figure 12. Given a set of data visualizations (a), we use our importance maps to automatically generate thumbnails (b). The thumbnails facilitate visual search through a database of visualizations by summarizing the most important content. More examples can be found in the Supplemental Material.

REFERENCES

1. Shai Avidan and Ariel Shamir. 2007. Seam Carving for Content-aware Image Resizing. *ACM Trans. Graph.* 26, 3, Article 10 (July 2007). DOI: http://dx.doi.org/10.1145/1276377.1276390

2. Michelle A. Borkin, Zoya Bylinskii, Nam Wook Kim, Constance May Bainbridge, Chelsea S. Yeh, Daniel Borkin, Hanspeter Pfister, and Aude Oliva. 2016. Beyond Memorability: Visualization Recognition and Recall. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (Jan 2016), 519–528. DOI: http://dx.doi.org/10.1109/TVCG.2015.2467732

3. Michelle A. Borkin, Azalea A Vo, Zoya Bylinskii, Phillip Isola, Shashank Sunkavalli, Aude Oliva, and Hanspeter Pfister. 2013. What Makes a Visualization Memorable? *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (Dec 2013), 2306–2315. DOI: http://dx.doi.org/10.1109/TVCG.2013.234

4. Georg Buscher, Edward Cutrell, and Meredith Ringel Morris. 2009. What Do You See when You'Re Surfing?: Using Eye Tracking to Predict Salient Regions of Web Pages. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 21–30. DOI: http://dx.doi.org/10.1145/1518701.1518705

5. Zoya Bylinskii, Tilke Judd, Ali Borji, Laurent Itti, Frédo Durand, Aude Oliva, and Antonio Torralba. 2012. MIT Saliency Benchmark. (2012).

6. Zoya Bylinskii, Tilke Judd, Aude Oliva, Antonio Torralba, and Frédo Durand. 2016a. What do different evaluation metrics tell us about saliency models? *CoRR* abs/1604.03605 (2016). http://arxiv.org/abs/1604.03605

7. Zoya Bylinskii, Adrià Recasens, Ali Borji, Aude Oliva, Antonio Torralba, and Frédo Durand. 2016b. *Where Should Saliency Models Look Next?* Springer International Publishing (ECCV), Cham, 809–824. DOI: http://dx.doi.org/10.1007/978-3-319-46454-1_49

8. Marcella Cornia, Lorenzo Baraldi, Giuseppe Serra, and Rita Cucchiara. 2016. A deep multi-level network for saliency prediction. In *2016 23rd International Conference on Pattern Recognition (ICPR)*. 3488–3493. DOI:http://dx.doi.org/10.1109/ICPR.2016.7900174

9. Andrew T Duchowski. 2007. Eye tracking methodology. *Theory and practice* 328 (2007).

10. Stas Goferman, Lihi Zelnik-Manor, and Ayellet Tal. 2012. Context-Aware Saliency Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34, 10 (Oct 2012), 1915–1926. DOI: http://dx.doi.org/10.1109/TPAMI.2011.272

11. Michael J. Haass, Andrew T. Wilson, Laura E. Matzen, and Kristin M. Divis. 2016. *Modeling Human Comprehension of Data Visualizations*. Springer International Publishing (VAMR), Cham, 125–134. DOI: http://dx.doi.org/10.1007/978-3-319-39907-2_12

12. Lane Harrison, Katharina Reinecke, and Remco Chang. 2015. Infographic Aesthetics: Designing for the First Impression. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 1187–1190. DOI: http://dx.doi.org/10.1145/2702123.2702545

13. Xiaodi Hou and Liqing Zhang. 2007. Saliency Detection: A Spectral Residual Approach. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*. 1–8. DOI: http://dx.doi.org/10.1109/CVPR.2007.383267

14. Xun Huang, Chengyao Shen, Xavier Boix, and Qi Zhao. 2015. SALICON: Reducing the Semantic Gap in Saliency Prediction by Adapting Deep Neural Networks. In *2015 IEEE International Conference on Computer Vision (ICCV)*. 262–270. DOI: http://dx.doi.org/10.1109/ICCV.2015.38

15. Laurent Itti, Christof Koch, and Ernst Niebur. 1998. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20, 11 (Nov 1998), 1254–1259. DOI:http://dx.doi.org/10.1109/34.730558

16. Robert Jacob and Keith S Karn. 2003. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. *Mind* 2, 3 (2003), 4.

17. Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross B. Girshick, Sergio Guadarrama, and Trevor Darrell. 2014. Caffe: Convolutional Architecture for Fast Feature Embedding. *CoRR* abs/1408.5093 (2014). http://arxiv.org/abs/1408.5093

18. Ming Jiang, Shengsheng Huang, Juanyong Duan, and Qi Zhao. 2015. SALICON: Saliency in Context. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 1072–1080. DOI: http://dx.doi.org/10.1109/CVPR.2015.7298710

19. Binxing Jiao, Linjun Yang, Jizheng Xu, and Feng Wu. 2010. Visual Summarization of Web Pages. In *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '10)*. ACM, New York, NY, USA, 499–506. DOI: http://dx.doi.org/10.1145/1835449.1835533

20. Tilke Judd, Frédo Durand, and Antonio Torralba. 2012. A Benchmark of Computational Models of Saliency to Predict Human Fixations. In *MIT Technical Report*.

21. Tilke Judd, Krista Ehinger, Frédo Durand, and Antonio Torralba. 2009. Learning to predict where humans look. In *2009 IEEE 12th International Conference on Computer Vision*. 2106–2113. DOI: http://dx.doi.org/10.1109/ICCV.2009.5459462

22. Nam Wook Kim, Zoya Bylinskii, Michelle A. Borkin, Krzysztof Z. Gajos, Aude Oliva, Frédo Durand, and Hanspeter Pfister. 2017. BubbleView: an interface for crowdsourcing image importance maps and tracking

visual attention. *TOCHI* (2017). DOI: http://dx.doi.org/10.1145/3131275

23. Nam Wook Kim, Zoya Bylinskii, Michelle A. Borkin, Aude Oliva, Krzysztof Z. Gajos, and Hanspeter Pfister. 2015. A Crowdsourced Alternative to Eye-tracking for Visualization Understanding. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '15)*. ACM, New York, NY, USA, 1349–1354. DOI: http://dx.doi.org/10.1145/2702613.2732934

24. Kyle Krafka, Aditya Khosla, Petr Kellnhofer, Harini Kannan, Suchendra Bhandarkar, Wojciech Matusik, and Antonio Torralba. 2016. Eye Tracking for Everyone. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2176–2184. DOI: http://dx.doi.org/10.1109/CVPR.2016.239

25. Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2012. ImageNet Classification with Deep Convolutional Neural Networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems (NIPS'12)*. Curran Associates Inc., USA, 1097–1105. http://dl.acm.org/citation.cfm?id=2999134.2999257

26. Srinivas S. Kruthiventi, Kumar Ayush, and R. Venkatesh Babu. 2015. DeepFix: A Fully Convolutional Neural Network for predicting Human Eye Fixations. *CoRR* abs/1510.02927 (2015). http://arxiv.org/abs/1510.02927

27. Ranjitha Kumar, Arvind Satyanarayan, Cesar Torres, Maxine Lim, Salman Ahmad, Scott R. Klemmer, and Jerry O. Talton. 2013. Webzeitgeist: Design Mining the Web. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 3083–3092. DOI: http://dx.doi.org/10.1145/2470654.2466420

28. Ranjitha Kumar, Jerry O. Talton, Salman Ahmad, and Scott R. Klemmer. 2011. Bricolage: Example-based Retargeting for Web Design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 2197–2206. DOI: http://dx.doi.org/10.1145/1978942.1979262

29. Matthias Kümmerer, Lucas Theis, and Matthias Bethge. 2014. Deep Gaze I: Boosting Saliency Prediction with Feature Maps Trained on ImageNet. *CoRR* abs/1411.1045 (2014). http://arxiv.org/abs/1411.1045

30. Matthias Kümmerer, Thomas S. A. Wallis, and Matthias Bethge. 2017. Saliency Benchmarking: Separating Models, Maps and Metrics. *CoRR* abs/1704.08615 (2017). http://arxiv.org/abs/1704.08615

31. Sharon Lin and Pat Hanrahan. 2013. Modeling How People Extract Color Themes from Images. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 3101–3110. DOI: http://dx.doi.org/10.1145/2470654.2466424

32. Jonathan Long, Evan Shelhamer, and Trevor Darrell. 2017. Fully Convolutional Networks for Semantic Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39, 4 (April 2017), 640–651. DOI:http://dx.doi.org/10.1109/TPAMI.2016.2572683

33. Peter O'Donovan, Aseem Agarwala, and Aaron Hertzmann. 2014. Learning Layouts for Single-PageGraphic Designs. *IEEE Transactions on Visualization and Computer Graphics* 20, 8 (Aug 2014), 1200–1213. DOI: http://dx.doi.org/10.1109/TVCG.2014.48

34. Junting Pan, Kevin McGuinness, Elisa Sayrol, Noel E. O'Connor, and Xavier Giró i Nieto. 2016a. Shallow and Deep Convolutional Networks for Saliency Prediction. *CoRR* abs/1603.00845 (2016). http://arxiv.org/abs/1603.00845

35. Junting Pan, Elisa Sayrol, Xavier Giro-i Nieto, Kevin McGuinness, and Noel E O'Connor. 2016b. Shallow and deep convolutional networks for saliency prediction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 598–606.

36. Xufang Pang, Ying Cao, Rynson W. H. Lau, and Antoni B. Chan. 2016. Directing User Attention via Visual Flow on Web Designs. *ACM Trans. Graph.* 35, 6, Article 240 (Nov. 2016), 11 pages. DOI: http://dx.doi.org/10.1145/2980179.2982422

37. Alexandra Papoutsaki, Patsorn Sangkloy, James Laskey, Nediyana Daskalova, Jeff Huang, and James Hays. 2016. Webgazer: Scalable Webcam Eye Tracking Using User Interactions. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI'16)*. AAAI Press, 3839–3845. http://dl.acm.org/citation.cfm?id=3061053.3061156

38. Ali Sharif Razavian, Hossein Azizpour, Josephine Sullivan, and Stefan Carlsson. 2014. CNN Features Off-the-Shelf: An Astounding Baseline for Recognition. In *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW '14)*. IEEE Computer Society, Washington, DC, USA, 512–519. DOI: http://dx.doi.org/10.1109/CVPRW.2014.131

39. Ronald A Rensink. 2011. *The management of visual attention in graphic displays*. Cambridge University Press, Cambridge, England.

40. Ruth Rosenholtz, Amal Dorai, and Rosalind Freeman. 2011. Do Predictions of Visual Perception Aid Design? *ACM Trans. Appl. Percept.* 8, 2, Article 12 (Feb. 2011), 20 pages. DOI: http://dx.doi.org/10.1145/1870076.1870080

41. Michael Rubinstein, Diego Gutierrez, Olga Sorkine, and Ariel Shamir. 2010. A Comparative Study of Image Retargeting. *ACM Trans. Graph.* 29, 6, Article 160 (Dec. 2010), 10 pages. DOI: http://dx.doi.org/10.1145/1882261.1866186

42. Manolis Savva, Nicholas Kong, Arti Chhajta, Li Fei-Fei, Maneesh Agrawala, and Jeffrey Heer. 2011. ReVision: Automated Classification, Analysis and Redesign of Chart Images. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology (UIST '11)*. ACM, New York, NY, USA, 393–402. DOI: http://dx.doi.org/10.1145/2047196.2047247

43. Chengyao Shen and Qi Zhao. 2014. *Webpage Saliency*. Springer International Publishing (ECCV), Cham, 33–46. DOI:http://dx.doi.org/10.1007/978-3-319-10584-0_3

44. Karen Simonyan and Andrew Zisserman. 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. *CoRR* abs/1409.1556 (2014). http://arxiv.org/abs/1409.1556

45. Jeremiah D Still and Christopher M Masciocchi. 2010. A saliency model predicts fixations in web interfaces. In *5 th International Workshop on Model Driven Development of Advanced User Interfaces (MDDAUI 2010)*. Citeseer, 25.

46. Jaime Teevan, Edward Cutrell, Danyel Fisher, Steven M. Drucker, Gonzalo Ramos, Paul André, and Chang Hu. 2009. Visual Snippets: Summarizing Web Pages for Search and Revisitation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 2023–2032. DOI: http://dx.doi.org/10.1145/1518701.1519008

47. Allison Woodruff, Andrew Faulring, Ruth Rosenholtz, Julie Morrsion, and Peter Pirolli. 2001. Using Thumbnails to Search the Web. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '01)*. ACM, New York, NY, USA, 198–205. DOI:http://dx.doi.org/10.1145/365024.365098

48. Pingmei Xu, Krista A. Ehinger, Yinda Zhang, Adam Finkelstein, Sanjeev R. Kulkarni, and Jianxiong Xiao. 2015. TurkerGaze: Crowdsourcing Saliency with Webcam based Eye Tracking. *CoRR* abs/1504.06755 (2015). http://arxiv.org/abs/1504.06755

49. Rui Zhao, Wanli Ouyang, Hongsheng Li, and Xiaogang Wang. 2015. Saliency detection by multi-context deep learning. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 1265–1274. DOI: http://dx.doi.org/10.1109/CVPR.2015.7298731