COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 49 (2025) e70080

ISSN: 1551-6709 online
DOI: 10.1111/cogs.70080

Evaluating Dogs' Real-World Visual Environment and Attention

Madeline H. Pelgrim, Shreyas Sundara Raman, Thomas Serre, Daphna Buchsbaum

Department of Cognitive & Psychological Sciences, Brown University

Received 6 August 2024; received in revised form 29 May 2025; accepted 4 June 2025

Abstract

Dogs have a unique evolutionary relationship with humans, yet little is known about the visual information available to them or how they direct their visual attention within their environment. The present study, inspired by comparable work in infants, classified the items available to be gazed at by dogs during a common daily event, a walk. We then explored the statistics over the availability of those categories and over the dogs' visual attention. Using a head-mounted eye-tracking apparatus that was custom-designed for dogs, 11 dogs walked on a predetermined route outdoors under naturalistic conditions generating a total of 11,698 gazes for analysis. Image stills from these fixations were analyzed using computer vision techniques to explore the items present, the space within the visual field those items occupied, and which of the items the dog was gazing at. On average, dogs looked proportionally most at buses, plants, people, the pavement, and construction equipment; however, there were significant individual differences. The results of this project provide a foundational step toward understanding how dogs look at and interact with their physical world, opening up avenues for future research into how they learn and make decisions, both independently and with a human social partner.

Keywords: Domestic dog; Eye-tracking; Visual attention; Preferential gaze; Visual statistics

1. Introduction

What does your dog see on a walk? What catches their visual attention? Studying how individuals visually interact with the world from their own perspective gives insight into their cognition in contexts ranging from a pet dog scanning for squirrels to an urban search and rescue dog navigating rubble to find missing people. Egocentric vision research captures the

Correspondence should be sent to Madeline H. Pelgrim, Department of Cognitive & Psychological Sciences, Brown University, 190 Thayer St., Providence, RI 02912, USA. E-mail: mpelgrim@gmail.com

visual information available to an individual, such as the items present in their environment, from their first-person perspective, as well as how they allocate their attention to those items. This approach has become widely used to explore how human infants' visual environments and interactions change over the course of development and how these changes impact, among other things, their developing ability to recognize faces and their acquisition of language. At present, however, real-world egocentric vision research and a better understanding of how they allocate attention to and learn from their real-world environment have been, with some exceptions, mostly limited to humans (Rodin, Furnari, Mavroeidis, & Farinella, 2021; Smith, Yu, Yoshida, & Fausey, 2015).

The present research uses head-mounted eye-tracking to explore the visual statistics of the environment and gaze behaviors of the domestic dog, a species closely related to humans by emotional bond and relied upon by humans in a variety of working roles and as companions. There is a growing interest in how dogs visually interact with their world. Dogs are relied upon to navigate the human-built visual world in a variety of work settings (i.e., as guide dogs for the blind) and as pet dogs, yet little is known about how they allocate their visual attention to complete these tasks. Dogs have worse visual acuity and less sensitive color perception than humans. In contrast, dogs are more sensitive to flicker rates, and they surpass human visual performance in dim lighting conditions (Byosiere, Chouinard, Howell, & Bennett, 2018). Researchers have also made advances in understanding how dogs respond to visual stimuli using stationary, screen-based eye-tracking (Karl, Boch, Virányi, Lamm, & Huber, 2020; Somppi, Törnqvist, Hänninen, Krause, & Vainio, 2012). By presenting precise stimuli on screens and recording dogs' eye movements, we know that dogs can recognize photos of familiar human and conspecific individuals (Somppi, Törnqvist, Hänninen, Krause, & Vainio, 2014) and that they respond differentially to human faces expressing different emotions (Karl et al., 2020; Kis, Hernádi, Miklósi, Kanizsár, & Topál, 2017; Somppi et al., 2016). Dogs tend to direct their visual attention to living creatures in the foreground (vs. the background), a pattern also observed in chimpanzees and humans (Kano & Tomonaga, 2013; Törnqvist, Somppi, Kujala, & Vainio, 2020).

In all, screen-based eye-tracking studies and other screen-based research have made important advances, and they capture detailed and highly precise measurements of looking behaviors in response to images and videos. However, these systems are not appropriate for all research questions, and participants' eye movements in a naturalistic context (i.e., while walking) cannot be explored. As such, it is not currently known how dogs (or canids in general) visually interact with their daily real-world environment, including the kinds of items that are present in their field of view and how they allocate their visual attention. Dogs, like many species, might perceive their world as structured into high-level visual categories. In particular, dogs may use a primarily top-down approach, actively directing their visual attention to certain categories or classes of items. In this case, dogs, like human infants by approximately 6 months of age (e.g., see Hunnius & Bekkering, 2010; Reynolds, 2015), could be visually attending to objects in a sophisticated manner, responding to the type or category of object. Dogs may also display signs of anticipatory looking, directing their attention to locations where items of interest could appear. On the other hand, it is also possible that dogs may only have our shared bottom-up processes but lack a categorical- or identity-based understanding of their visual world. In this case, dogs may be using vision, but their looking behavior is not primarily based on item category or identity. Dogs' vision could be best explained by a bottom-up process, where they are primarily responding to low-level visual cues and saliency. If this is the case, we would expect that a model of image saliency, a bottom-up approach that identifies the most salient region of an image on the basis of low-level cues like brightness and luminance, would capture dogs gazing behaviors. Finally, while unlikely, as a possibly predominantly olfactory species, dogs' visual attention may be random or driven by their noses and not by visual features of the environment. If true, we would expect gaze to be randomly distributed.

We aimed to capture how dogs direct their visual attention using head-mounted eye-tracking to record what items were in the dogs' field of view (their relative frequencies and sizes) and which of the items they looked at. We also incorporated computer vision techniques to identify these items, something that was previously extremely labor-intensive. Finally, we compared dogs' visual attention behaviors to existing image-saliency models to explore if dogs' gazing behaviors were accurately captured by low-level visual-perceptual cues. Identifying how dogs direct their visual attention outdoors is particularly relevant given the numerous real-world working roles dogs serve in, where they must traverse complex outdoor environments.

1.1. Head-mounted eye-tracking

To study gaze in more naturalistic environments, researchers have begun using head-mounted eye-trackers. These systems capture the wearer's first-person view of the world, as well as record their eye movements. This captures both the items present in the environment and which of those items the participant looked at. They can be worn in a variety of ways (i.e., caps, glasses, goggles), allowing for their use in exploring natural behaviors with head shapes ranging from peacocks to lemurs (Shepherd & Platt, 2006; Yorzinski, Patricelli, Babcock, Pearson, & Platt, 2013), and more recently, dogs (Pelgrim, Espinosa, & Buchsbaum, 2022). Head-mounted eye-tracking has also been used to facilitate cross-species comparisons of visual behavior, such as by comparing how cats and humans coordinate eye and head movements (Einhäuser et al., 2009). The largest take-up of this method, though, has come from research on young infants.

Describing what participants have visual access to is inherently useful and necessary to explain behavior and available cognitive mechanisms. Head-mounted eye trackers have given us insight into how infants respond to their mother's voices (Franchak, Kretch, Soska, & Adolph, 2011) and how infants and parents coordinate joint attention to items (Yu & Smith, 2013). They have also been used to capture how visual attention changes over development. As children transition from crawling to walking, they move from looking at the floor in front of them while in motion to looking at walls, items, and caregivers. This changes their frequency of looks to caregivers because in order to look at their caregiver, crawlers have to stop and either sit or crane their heads, whereas walkers can stay in motion and look ahead (Kretch, Franchak, & Adolph, 2014). An improved understanding of infants' visual experiences has informed theories on language acquisition, namely, that a statistical learning framework can reasonably be applied to word learning because of the distribution of item frequencies in naturalistic scenes. For example, the first nouns that children learn tend to be those items most frequently present in their environments (e.g., spoon, bowl) (Clerkin, Hart, Rehg, Yu, & Smith, 2017; Smith, Jayaraman, Clerkin, & Yu, 2018). In sum, the items present in a child's environment support their learning and development. Capturing the types and distribution of items in children's visual environments has informed our understanding of their development across cognitive and social domains (Jayaraman & Smith, 2020).

The present study seeks to describe and categorize both the visual information available to dogs in their daily environments, as well as characterize how they direct their attention within that space. We took an ecologically valid approach to understanding dogs' attention in their daily environment by having dogs walk with their guardians in a normal fashion along a predetermined route.

Our study had three major aims. First, we explored the relative frequency that items were present in the dogs' field of view, providing us with an understanding of the items available for dogs to look at in an ecologically relevant context. Second, we evaluated how dogs looked at the items in their field of view, exploring for consistency in looks across exposures to particular item classes (e.g., people) when those items were present in the environment (i.e., if they looked at a person each time there was a person in their field of view or if they rarely looked at people while people were in their field of view). We evaluated this using the proportion of time dogs fixated on the item relative to the amount of time that item was in their field of view. We particularly wanted to explore the social domain, namely, if people were looked at consistently across exposure and if they were looked at more than other nonsocial items. We also examined if dogs' visual attention to items in their field of view was predicted by the item categories present in their visual environment or could instead be explained by low-level properties of items rather than the identity of the item class, using image saliency analysis. Our third aim was to look for any individual differences in visual attention to item classes between dogs, considering both the items in their view and which of those items they looked at. It is possible that dogs may differ in what items they find visually interesting, which could result in differences both in the items in their field of view and in which of those items they looked at. Historically, this work would have been logistically challenging due to the time required to annotate the items present in the dogs' field of view. We developed a computer vision pipeline to identify each of the items in the image on a pixel level, segmenting out the individual items from the broader image and then applying class labels to each pixel-identified item. Potential classes were determined in advance based on common items present across multiple dogs' walks. We then integrated the eye-tracking data to provide a label for the item the dog was gazing at for each fixation.

2. Behavioral methods

2.1. Ethical note

This study was approved by the Brown University Institutional Animal Care and Use Committee (IACUC), protocol number 20-05-0002. Procedures were in accordance with the

ASAB/ABS Guidelines for the Use of Animals in Research and complied with the United States Department of Agriculture Animal Welfare Act & Regulations. All participants were pet dogs who were walked by their guardians throughout the session.

2.2. *Participants*

Participants were 11 dogs (female = 7, mean age = 6.5 years) recruited for participation in a broader eye-tracking training program. Dog breeds represented were one Miniature American Shepherd, one Australian Labradoodle, two Golden Retrievers, two Labrador Retrievers, and five Mixed Breeds. Each dog was recorded walking along the route once. An additional two dogs were excluded from data analysis due to camera displacement ($n = 1$) and a refusal to walk while in the goggles ($n = 1$). Dogs were chosen for suitability with the eye-tracking training program and the guardian's willingness to complete the training. Prior to participation in this experiment, dogs were trained at home by their guardians to wear the eye-tracking goggles using commercially available dog goggles following the methods described by Pelgrim et al. (2022), see Supplementary Materials for more details. The purpose of this training was to acclimate dogs to the goggles so that they were comfortable wearing them. Dogs were approved to begin participation if guardians reported they were comfortable walking and behaving normally at home and outdoors, wearing training goggles for at least 10 min.

2.3. *Procedure and materials*

Throughout sessions, dogs wore a custom-developed head-mounted eye-tracker consisting of two cameras affixed to dog goggles (Positive Science, Inc.). One camera records the dog's right eye via an infrared eye camera with an adjacent infrared emitting diode (hereafter the eye camera). The other camera (hereafter the scene camera) recorded the dogs' first-person perspective, recording a field of view of 101.55° horizontal and 73.60° vertical. Videos from both cameras were digitized at 29.96 frames per second. This apparatus, acclimation, and calibration procedures are adapted from comparable models in other species and have been validated in dogs using alternative methods (Fig. 1; Pelgrim et al., 2022). Dogs also wore a harness throughout their session to hold the video recording pack and its battery.

Prior to starting their walk, dogs first completed a calibration procedure. This procedure allowed for the dogs' eye movements as recorded from the eye camera to be mapped onto the field of view recording from the scene camera, the result being the dogs' gaze (or where in their environment they were looking), which could be extrapolated offline for the entire recording, after the session (Pelgrim et al., 2022). Consistent with previous work, we calibrated the eye-tracker when the experimenter drew dogs' attention to specific points using a treat. The dog's guardian held their dog's head stationary while the dog followed treats held by an experimenter, via eye movements alone, through five unique points in space. Each point in space where the dog looked at the experimenter provided a known point, meaning that the positioning of the dogs' pupil and corneal reflection was linked to where in their first-person view they were looking. The five points were chosen to be spread widely across the dogs' first-person view, thus requiring a wide range of eye movements. These eye movements made

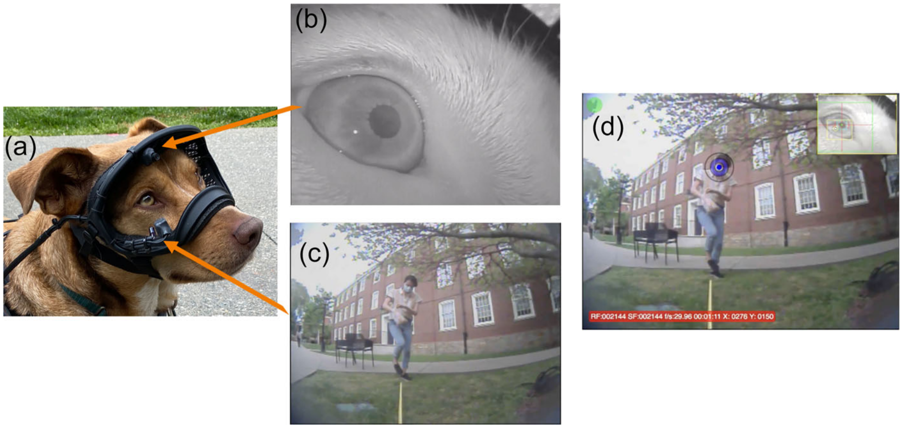


Fig. 1. (a) A dog wearing the eye-tracking goggles which had two cameras. (b) The dog's eye recorded from the eye camera. (c) The dog's view recorded by the scene camera. (d) The dog's view with their gaze indicated by the blue circle.

the offline extrapolation of the point of regard for the entirety of the walk using eye-tracking software more accurate.

During the calibration procedure, a removable handheld screen was plugged into the recording pack to allow the experimenter to verify the eye camera was recording a clear and centered image. After the dog looked at the experimenter in all five points, as judged by the experimenter, the LCD screen was removed. The calibration procedure was completed both before and after the walk to provide enough known points for (1) a successful calibration and (2) verification of that calibration accuracy (more details in Data Coding and Preparation).

Following the first calibration, dogs walked with their guardians, following the experimenter's directions, along a preset route. The route was 0.5 miles and was chosen for its variety of scenery, including both city streets and quiet campus greenspaces. Guardians were instructed to walk their dogs as normal, and the experimenter followed behind or adjacent to the dog−guardian pair. If at any point during the session, the eye-tracker was disturbed or shifted (i.e., the dog shook their body or brushed against a wall), the fit was adjusted, and the eye image was verified. In the event of tracking disruption, the calibration procedure was also repeated.

2.4. Data coding and preparation

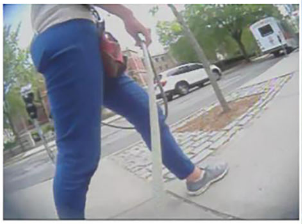
After the session, video data recorded from the eye and scene cameras were combined as described above, using between 4 and 9 calibration points (Yarbus eye-tracking software, Positive Science Inc.). This process identified both the timing and the direction of dogs' gaze, referred to from here on as fixations. Fixations were defined as a stable eye positioning lasting for 100 ms or more as determined by the eye-tracking software in keeping with past work (Pelgrim et al., 2022). More specifically, the timing (start and stop in ms) and the dog's point of regard in the visual scene (defined by x-y coordinates from the scene camera) was identified for each fixation.

This eye-tracking system has previously been established to have a spatial accuracy of approximately 2−4° in humans (Franchak et al., 2011, Watalingam, Richetelli, Pelz, & Speir, 2017), around 4° in peacocks (Yorzinski et al., 2013) and around 3.6° in dogs indoors (Pelgrim et al., 2022). The spatial accuracy of the eye camera mapping onto the scene for our outdoor walks was calculated, as in past work, using the unused points from the calibration procedure. The distance between the extrapolated point of regard (where the eye-tracking software calculated the dog to be looking) and the known point of regard (where the dog was known to actually be looking, namely, at the treat bag in the experimenter's hand) was calculated for approximately 20 frames across the calibration points not used for the mapping, providing the spatial accuracy of the mapping, in degrees, for each dog. The spatial accuracy for the present sample was 5.32°. This is less precise than past implementations mentioned above; however, it was to be expected given that this is the first time this system has been deployed outdoors under natural variable lighting conditions. Further, the spatial accuracy of the tracking for each dog was accounted for when determining dogs' fixated items, creating a region of fixation for each look.

Each fixation was segmented into the items present (and, therefore, available to be looked at), and coded for the classes of those items (see Fig. 2), the identity of the item (or items) the dog was looking at, and the relative proportion of the dog's field of view taken up by each item (a function of both the item's proximity to the dog and the item's absolute size, see below), using computer vision techniques (see Computer Vision section). One limit of our relative size calculations is that our eye-tracking cameras do not record depth information, so object size and proximity cannot be separated (objects that are very close will be recorded as larger, and large objects in the distance will be recorded as smaller). The classes of items were chosen ahead of data collection as they were consistently present on the predetermined route. Fifteen items were chosen in total based on the items available in the real world for multiple dogs (Table 1).[1]

The primary focus of this study was how dogs interact with their world visually, so instances where dogs were relying on another sense (such as while sniffing) were removed. Sniffing bouts were defined as looks where two or fewer items were present in the environment. As an example, when a dog was sniffing a pole, the only items visible in the environment were the pole and the plant and both were so close to the camera that it is unlikely that the dog was visually considering them (see Supplementary Materials).

In some cases, the dog could be fixated on at a location occupied by more than one item (see Fig. 3). To determine which item(s) the dog was gazing at, we examined the items identified by our computer vision output that were present in the fixated region (the target of fixation plus the calibration margin of error for each dog). For each item present in the fixated region, we weighted the item according to how much of the fixated region it occupied (Fig. 3).[2] For example, as in Fig. 3, if approximately 50% of the pixels in the fixated region were occupied by an item in the class horizontal plants, and approximately 25% each by items in the classes vertical plants and buildings, the fixation weight would be 0.25 each for the building and vertical plants, and 0.5 for the horizontal plants. For each fixation, we also weighted the duration of time spent looking at each class contained within the fixation region based on



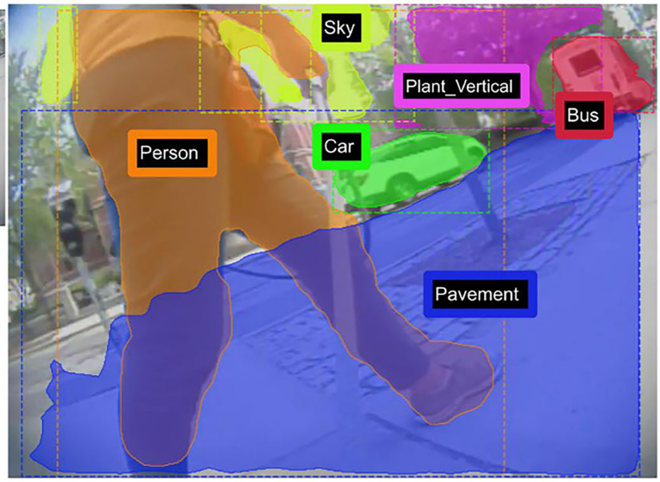


Image of a dog looking to its guardian. The computer vision algorithm identified a person, the pavement, the sky, a car, a vertical plant, and a bus.

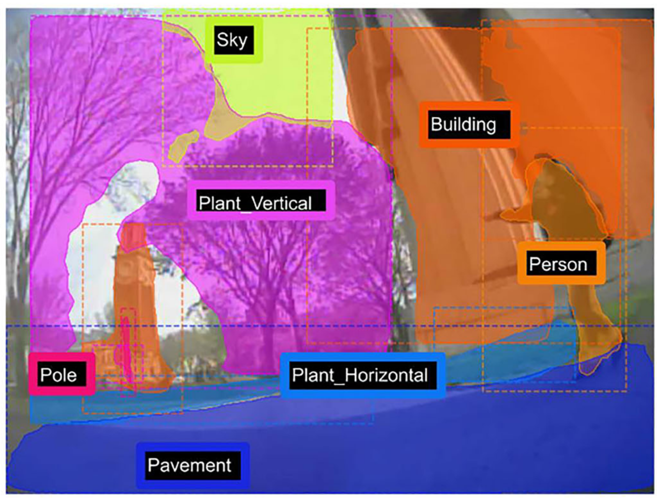


Image of a dog at the path ahead and the person. The algorithm identified a person, the pavement, the sky, buildings, horizontal and vertical plants, and a pole

Fig. 2. Two examples of images segmented into items from the test set, using our MaskRCNN segmentation approach. Left—original image, captured by head-mounted camera; Right—fine-tuned MaskRCNN's predicted segmentation masks. The shaded field indicates the identified item and the dashed rectangle shows the outer boundaries of each item. Unique class labels are bolded for easier visualization.

the proportion of the fixated region that item occupied (i.e., a 200-ms fixation to a region containing 50% sky and 50% building would count as 100 ms to each of the two classes).

2.5. Image saliency analysis

Dogs could be directing their attention based on interest in different item classes, or it could be a more bottom-up process driven by low-level image features contributing to increased saliency of regions of the image. To examine what a bottom-up visual approach would look

Table 1
Looking behaviors to classes across dogs

Object	Avg. proportion of time in view (SE)	Avg. proportion of time fixated while in view (SE)	Avg. size when in view (SE)
bench/chair	0.033 (0.005)	0.012 (0.004)	0.01 (0.001)
bicycle	0.024 (0.007)	0.099 (0.032)	0.031 (0.003)
building	0.878 (0.022)	0.144 (0.012)	0.145 (0.015)
bus	0.008 (0.005)	0.348 (0.094)	0.187 (0.024)
car	0.299 (0.05)	0.064 (0.015)	0.027 (0.003)
construction	0.011 (0.003)	0.145 (0.04)	0.044 (0.007)
pavement	0.885 (0.047)	0.381 (0.069)	0.336 (0.027)
person	0.389 (0.067)	0.157 (0.034)	0.131 (0.017)
plant_horizontal	0.616 (0.041)	0.174 (0.032)	0.18 (0.019)
plant_vertical	0.934 (0.018)	0.269 (0.058)	0.212 (0.02)
pole	0.168 (0.019)	0.027 (0.008)	0.022 (0.003)
scooter	0.008 (0.002)	0.077 (NA)	0.012 (0.006)
sculpture	0.037 (0.005)	0.036 (0.01)	0.015 (0.004)
sign	0.013 (0.003)	0.049 (0.011)	0.028 (0.009)
sky	0.838 (0.022)	0.07 (0.026)	0.075 (0.01)

Note. Only one fixation to a scooter occurred in our sample, so there is no standard error for fixations to that class.

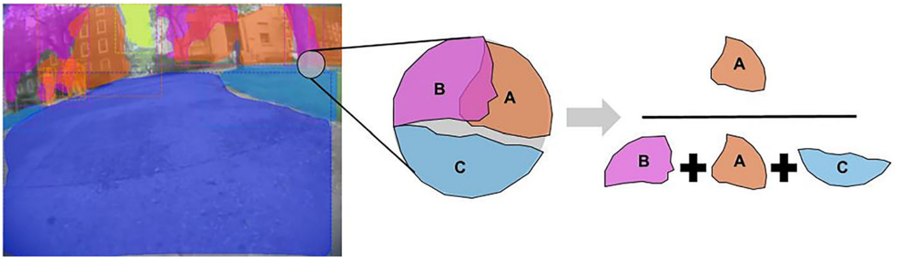


Fig. 3. The procedure used to estimate the probability distribution over fixated objects. Left-to-right: Given the fixation region in the image identified in gray (Left), we consider portions of predicted object masks that intersect the fixation region (Center) in our estimate. We estimate the probability of the fixation toward a given class (e.g., class "A") by normalizing the total number of pixels in the intersection belonging to the given class (i.e., the pixels in the intersection belonging to "A") over the total number of pixels belonging to all classes in the intersection (i.e., the pixels in the intersection belonging to "A," "B," and "C").

like, we conducted image saliency analyses using Saliency Toolbox Version 2.3, a well-established and validated model (Walther & Koch, 2006). For a given image, the Saliency Toolbox generates feature maps for standard low-level image features (i.e., luminance, color intensity and opponency, orientations) which are then combined eventually resulting in a saliency map. Dogs have different color vision than humans do, so to account for this, we conducted our image saliency analysis first in grayscale, and then in full color. This allowed us to compare dogs' observed real-world fixations to their expected fixations within the same images if their gaze was primarily driven by low-level image features. If fixations predicted by the saliency model generally align with dogs' real fixations, it would suggest that dogs' visual attention is primarily driven by low-level image features. In contrast, if the saliency model is a poor predictor of dogs' real-world fixations, while item class is a better predictor, it would suggest that dogs' visual attention is being driven by more complex factors, including item category or identity.

3. Computer vision

3.1. Methods

Our head-mounted eye-tracking approach generates an average of 1063 fixations in an 11-min walk. It is, therefore, not practical to manually annotate the items present in each fixation. Manually annotating the items present from the look image at each fixation would have taken around 1200 hours based on the time taken to annotate the data set used for training and testing our computer vision approach. Using computer vision significantly reduced the time required to identify the items in dogs' look images. After the computer vision approach was trained, each dog's complete walk could be annotated in a matter of minutes. Our aim was to automate the segmentation of images into items, the classification of those items, and the identification of the items the dog was gazing at. To do this, we utilized an automated computer vision method for *panoptic* segmentation (simultaneous identification of items within the image and their item classes). This approach served to identify and label the item classes present at each fixation (Fig. 2). We used an off-the-shelf object identification algorithm (Mask R-CNN with ResNet-101-based Feature Pyramid Network backbone, He, Gkioxari, Dollár, & Girshick, 2018) that was pretrained on a publicly available image data set (Microsoft Common items in Context, Lin et al., 2015). We then adapted and trained it using a training set of annotated scene images taken from our own data. We used 610 manually annotated images from our data set for training and an additional 621 images for testing. Training and test sets contain comparable distributions of item classes. See Fig. 2 for an illustration of how the model functions. We find that our fine-tuned model performs very well in matching the ground-truth annotations from our test set. After confirming our model's performance on the test set, we ran it on our remaining data. For complete details on data set curation and annotation and model training, see Supplementary Materials.

After the images were segmented into items and labeled, we integrated the eye-tracking fixation data (in conjunction with the spatial accuracy of the eye-tracking system) to create a probability distribution of the potential fixated item class, in order to create a fixation weight as described above (Fig. 3). We used the calibration error of the eye-mounted camera to estimate the radius of error around the fixation pixel; we assumed that the true fixation is expected to lie within this fixation region. To estimate a distribution over fixated items, we normalized the pixel counts for each identified object within the intersection by the total number of pixels in the fixated region assigned to an item class; this generated a probability distribution over possible fixation items for the image. Fig. 3 highlights a visual example of this probability distribution computation, as discussed earlier (see Supplementary Materials for additional details).

Across the 11 dogs, a total of 11,698 fixations were recorded spanning 102,868 frames. After eliminating sniffing bouts, or look images that contained two or fewer items in the view as described in data coding and analysis, as well as those that did not result in any predictions from our computer vision analysis, 10,296 look images remained. This set was used in subsequent calculations and analyses.

3.2. *Computer vision results*

We first verified that our computer vision model was generalized. To be successful, our computer vision algorithm needed to accurately identify item classes across the whole of the image, matching in overall coverage and class label with our human annotated images (our ground truth). To evaluate our model performance, we evaluated performance across metrics, specifically overall image coverage by identified items, class-specific coverage, and fixated region class identification. We also examined both the true positive rate, false positive rate, and the confusion on a class-by-class level (Fig. 4). All model evaluations were done by comparing our model annotated test set to the ground truth data.

First, we compared the overall coverage of the image from our computer vision model to our ground truth. The average percent of coverage over the frame (consisting of the total area of all model-identified item classes), across the four dogs in the test set, was smaller in area than ground-truth masks. Our fine-tuned model covers on average 84% of the image, relative to the 93% average of our human annotators. Given this reduction in coverage, we next wanted to evaluate if this resulted in changes in the classes identified in our fixated regions, a critical part of our analysis. We found that they did not significantly impact class distributions for fixation predictions.

Next, we examined the overlap of the generated classes to the ground-truth classes. We evaluated the Intersection Over Union (IoU) or the number of pixels from each of the model-identified classes overlapping with the ground-truth classes, normalized by the number of pixels in the union of the predicted and ground-truth identified classes. We found that the fine-tuned model achieves IoU substantially greater than random chance—indicating that the generated masks identifying the item classes are semantically relevant. The median IoU across all classes is 60% (10 $\times$ higher than chance) with a maximum IoU of 85% for the "bus" class and a minimum IoU of 42% for the "pole" class. In comparison, the maximum IoU by random chance is 8% for the "sky" class. For complete model performance metrics, see Supplementary Materials.

4. Statistical analyses

As a first step to understanding how dogs' view natural scenes, we quantify the items present in dogs' field of view. We report on the proportion of time each item class was present. We then conducted a 15 (item class) by 11 (dog identity) ANOVA to explore the impact of class and dog identity on the average proportion of time items were in view for each dog.[3]

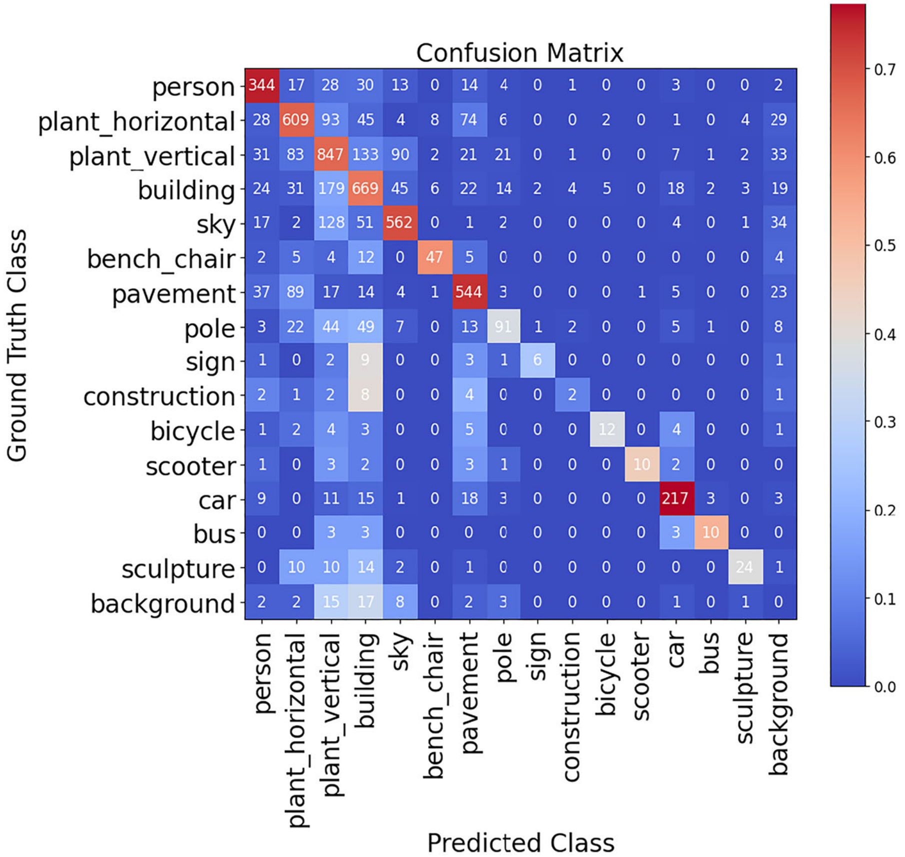


Fig. 4. Confusion matrix comparing, for our test set, the frequency of identification of an instance of a class relative to its ground-truth occurrence. Items classed as “background” mean that our algorithm failed to identify the item as one of the itemized classes.

Our next goal was to examine how dogs directed their visual attention at the items present in their environment. We considered a number of factors that could impact how dogs look at item classes, specifically class identity, dog identity, and item size. To explore the impact of dog identity and item type, we conducted a logistic regression with a Firth’s correction using the logistf package in R exploring the coded binary fixation data as a function of the item class, dog identity, and an interaction between dog and item class. To conduct this analysis, we used a subset of the classes, exploring only those classes (11/15) that were present in all dogs’ views (removing buses, signs, construction equipment, and scooters). We also conducted post-hoc pairwise comparisons using a Tukey correction. To examine the impact of object size, we first conducted a Spearman correlation relating the average size of the item class to the time fixated on that class (relative to the time it was in view). Next, we used a linear mixed effects model to integrate the possible contributors of visual attention, examining the fixation duration as a function of the fixated item class(es), size of fixated item(s), and a possible interaction between the size and item class. We also included a random intercept for each dog.

As a final step, we evaluated the saliency models' predictions for where dogs would look, if their attention was driven primarily by low-level image features, against our fixation data by evaluating the Area Under the Curve (AUC) of the Receiver Operating Characteristics (ROC) (Bylinskii, Judd, Oliva, Torralba, & Durand, 2019). This compares the accuracy, using true and false positive rates, of the saliency model, relative to the ground truth from the fixation data. We used AUC-Judd, a variant of AUC with a repeated threshold approach. For each potential threshold on a pixel-by-pixel level, all points above the threshold are considered "fixations." Instances where the actual fixation overlaps with the predicted "fixations" based on the threshold are considered true positives, and the points that are above the threshold but do not overlap with the true fixated region are considered false positives. The ROC curve is plotted from these two values, and AUC-Judd scores can range from .5 (random classification) to 1 (perfect performance) (Bylinskii et al., 2012.; Judd, Durand, & Torralba, 2012).

5. Behavioral results and discussion

5.1. *Items in view*

Our first aim was to explore the items present in dogs' field of view during their walks. This is a necessary first step to understand how dogs direct their attention within their field of view. We found a significant effect of item identity, $F(14, 140) = 172.07, p < .001, \eta_p{}^2 = 0.95$, confirming that some item classes were more commonly present in dogs' view, and available to be looked at, than others.

On average, dogs tended to have common scene components—plants (horizontal and vertical), pavement, buildings, and the sky—present in their field of view on the majority (> 50%) of their fixations. In addition to these ubiquitous items, dogs had certain categories of items available for gaze moderate amounts of time (< 50% and > 10% of fixations). These included people, cars, and poles. Finally, some categories were rarely in dogs' view, including sculptures, benches/chairs, bicycles, signs, construction equipment, buses, and scooters. See Table 1 for a summary of the proportion of time all coded item categories were in view. We also found no effect of dog identity, $F(10, 140) = .879, p = .55, \eta_p{}^2 = 0.06$, meaning that the dogs in our study encountered similar item classes along their walk, and did not differ in the proportion of time that different item categories were in their view. Given that dogs walked the same route, they had the opportunity to orient their fields of view toward many of the same static items (e.g., buildings) for comparable amounts of time; however, it is still notable that dogs' different sizes, training experiences, and personalities did not result in them directing their heads differently on their walks.

5.2. *Relative time looking to items*

Our second aim was to explore how dogs allocated their visual attention to the items in their view (Fig. 5). As mentioned previously, it is possible that dogs are not attending to items

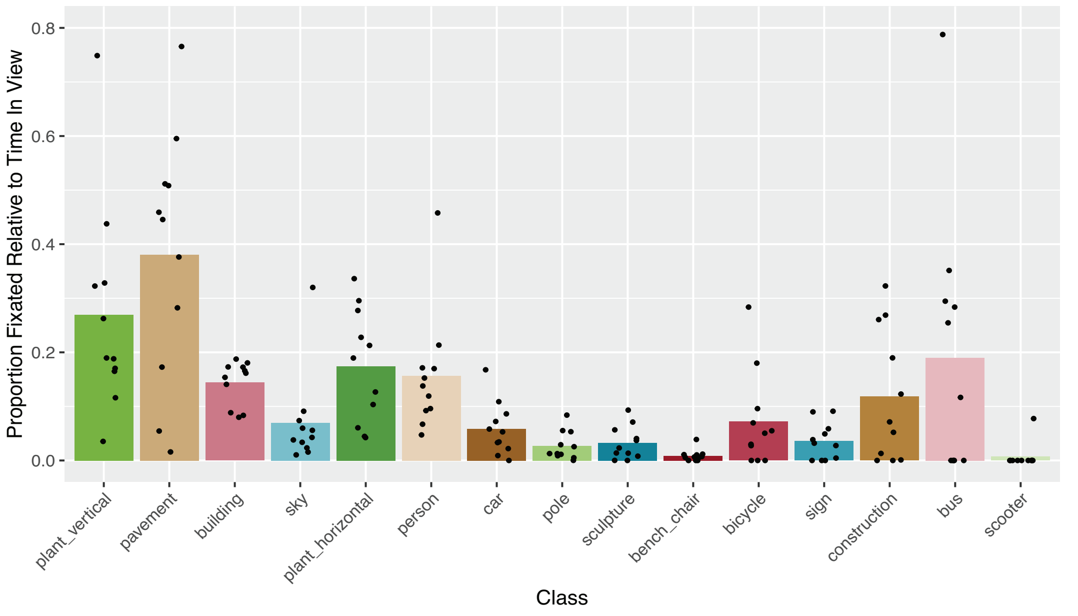


Fig. 5. Average proportion of fixated to each class, relative to the time the class was in view. Dots indicate individual dogs' averages.

on the basis of class, but are looking randomly at the items in their view. If this is the case, we would expect dogs' fixations to items in their view to be uniform across classes. In contrast, dogs may be actively directing their visual attention, selectively looking to certain classes more than others when they are available in the environment. Dogs may also differ in the way they direct their visual attention, and this individual difference may interact with the class identity. For example, some dogs may fixate on a class the majority of the time it is in their view, while others never look to that class.

We first explored in a binary fashion how often dogs fixated to an item class when it was in view, and if this differed across dogs or potential dog−class interaction. We found a significant main effect of item class, $X^2(10) = 1348.97$, $p < .001$. This suggests that dogs are not uniformly or passively observing the items in their field of view but are actively directing their attention to certain classes of items. For example, dogs looked frequently to people, 15.7% of the time they appeared. In contrast, dogs only looked to benches and chairs around 1.2% of the time they were in view, and post-hoc pairwise comparisons exploring class differences with a Tukey correction revealed that looks to these two classes were significantly different, $t(51{,}368) = -5.805$, $p < .001$. Dogs also looked significantly more to people than other asocial classes like sculptures, $t(51{,}368) = 5.137$, $p < .001$, and the sky, $t(51{,}368) = 16.346$, $p < .001$.

In past screen-based eye-tracking research in dogs, plants and sky have been used as the background material to explore how dogs look at the primary subject of the image, typically a person or animal present in the foreground (Törnqvist et al., 2020). Interestingly, in our sample, dogs were interested in plants and looked at them relatively frequently. Dogs looked at vertical plants (like trees and bushes that could not be easily moved through by the dog) 26.9% of the time they were in view. This was significantly more than they looked to people, $t(51{,}368) = -17.154$, $p < .001$. In contrast, another frequent background item from screen-based eye-tracking, sky, was almost never actually looked at by dogs. The sky was nearly ubiquitous in dogs' view, present and available for view on average 83.8% of fixations, yet dogs looked infrequently at it, only looking at the sky 7% of the time it was in their view (Table 1). This suggests that, unlike plants, in a real-world context, the sky is treated by dogs as a background, or at least not something that is worth attending to. Dogs also attend frequently to the ground in front of them, looking to the pavement 38% of the time it is in view and looking to horizontal plants (like grasses that could be walked on) around 17.4% of the time in view. These looks may be for navigational or wayfinding purposes, as past research in adult humans and in children suggests that when navigating in a real-world context, people often look to the ground, helping us to map a path forward and to avoid obstacles (20% to the ground in de Winter et al., 2021, 55.9% to street edges and the ground in Simpson, Thwaites, & Freeth, 2019, and 31.8% looking to ground obstacles in Franchak & Adolph, 2010).

From the same binary fixation analysis, we found no significant main effect of dog identity, $X^2(10) = 4.15$, $p = .94$. However, we did observe a significant interaction between dog identity and class on fixation behaviors, $X^2(100) = 3997.19$, $p < .001$. This suggests that there were individual differences in the way that dogs fixated to classes. For some classes, there were significant differences between dogs, and dogs were not all alike in which classes they found visually interesting. As an example, when considering the relative amount of time dogs spent looking to people in their environment, one dog looked at people 46% of the time they were in her view. Post-hoc pairwise comparisons using a Tukey correction found that she looked to people significantly more than most other classes, including items like buildings, $t(51{,}368) = -5.96$, $p < .001$, horizontal plants, $t(51{,}368) = 10.843$, $p < .001$, and sculptures, $t(51{,}368) = 4.60$, $p < .001$. In contrast, another dog only looked at people around 5% of the time they were in her view and pairwise comparisons showed she did not look significantly more at people than at any other class ($p < .05$). Post-hoc pairwise comparisons using a Tukey correction also found that these two dogs looked to people significantly different from each other, $t(51{,}368) = 4.21$, $p = .001$. In contrast, for other classes like benches and chairs, dogs were quite consistent in their looking behaviors. On average, dogs looked rarely to benches and chairs when in view and there were no significant ($p < .05$) individual differences in fixations to benches and chairs. See Supplementary Materials for visualizations of individual dogs. Our results suggest that dogs differ in how visually interesting they find some, but not all, classes of items. There is little doubt that dogs do experience the world differently, and their size and life experiences may play a role in this.

5.3. *Impact of item size*

In addition to the identity of the fixated items present in dogs' view, we explored how dogs looked at items as a function of class size. It is possible that dogs are not gazing selectively, but are simply gazing at the things in their environment that take up the most space. If they are selective, we could expect that dogs would occasionally look to items that are comparatively smaller, either because the object itself is smaller or because it is far away.

Across walks, items differed in size, both within and between items. Some items, like sculptures and benches/chairs, were consistently very small in dogs' views ($M_{\text{sculpture}} = .01$, $SD = .04$, $M_{\text{bench/chair}} = .01$, $SD = .01$). In contrast, other items tended to take up large portions of dogs' view and varied widely in size (i.e., $M_{\text{pavement}} = .35$, $SD = .17$, $M_{\text{plant\_horizontal}} = .19$, $SD = .19$). On 5843 trials, or 56.75% of trials, the largest class in view was one of the fixated items, as determined from the probabilistic fixation data across the region of fixation. On these looks, the object took up an average of $M = .46$, $SD = 0.14$ of the dogs' field of view. This means that on just over half of their total fixations, dogs are looking to the largest item class in their view at that moment, and that item class on average occupied 46% of dogs' view.

The average size of the item was correlated with dogs' fixation to that item. Using a Spearman correlation, we found a significant positive correlation between the average size of item class and time fixated on that class (relative to the time it was in view), $r_s = .751$, $p < .001$. While it is interesting that dogs are tending to look at larger items, conclusions from these results should be limited, as the directionality is unknown. More specifically, it is likely that dogs will turn their heads and/or move toward items of interest, thus making them bigger. However, it is also likely that larger items may be more attention-grabbing, and dogs may be more likely to see them (vs. smaller items which could be equally or more interesting but more easily missed).

A major advantage of head-mounted eye-tracking is the use of real-world stimuli and mobile participants, and in this case, our results suggest that dogs may consider at least some plants to be items of interest, exploring them visually and through other sensory modalities (the majority of sniffing bouts that were removed included plants and poles); however, they do not appear to consider the sky as an item class worth investigating. Further research is needed to make more nuanced conclusions about dogs' response to plants and other potential background items.

We explored the duration of each of the dogs' fixations ($n = 10{,}296$, $M = 301.8$ ms, $SD = 481.09$ ms) as a function of the class of item they were looking at, the size of that item they were looking at, and a possible interaction between the size and item class using a linear mixed effects model with a random intercept for each dog. We had a total of 16,166 individual targets of fixation, and after distributing the fixation duration across the region, the average fixation time was reduced ($M = 190.11$ ms, $SD = 368.27$ ms). Dogs fixated longer on some item classes than others, $X^2(14) = 111.49$, $p < .001$. As an example, dogs' fixations to benches/chairs tended to last for a very short time ($M = 166.67$ ms, $SD = 47.14$) but dogs' fixations to cars (another infrequently appearing item) lasted much longer, and were much more variable ($M = 380.87$ ms, $SD = 582.55$). From the same model, we found no main effect of the fixated item size relative to the current view, $X^2(1) = .19$, $p = .66$, suggesting that dogs' fixations were not significantly longer based on the size of the item they were fixating on. Finally, in the same linear model, we found a significant interaction between item class and fixated item size, $X^2(13) = 37.91$, $p < .001$. This suggests that for some item classes, dogs had longer individual fixations in response to larger instances of the item, but this was not true for all classes. We also found some variance across dogs in the duration of their fixations ($SD = 84.66$).

5.4. Image saliency analysis

For our full-color analyses, AUC-Judd was 0.67. Similarly, the AUC-Judd for grayscale analyses was 0.66. This suggests that the saliency model is a poor predictor of dogs' looking behaviors, or that it is unlikely dogs' fixations are driven solely by low-level image saliency features. The ROC curve as well as further details on AUC calculations and saliency analysis can be found in Supplementary Materials.

6. Conclusion

The present study aimed to explore both the items present in dogs' visual environments as well as which of those items they chose to look at. We integrated computer vision techniques to identify and classify items from the dogs' perspectives. We found that we were able to identify and classify the classes of items available in the dogs' view and how they directed their attention to them. Within our sample, dogs were very similar in the item classes they had in their view, mostly encountering plants, buildings, the sky, and pavement. Dogs' visual attention is active and is not just driven by the low-level image saliency features (e.g., luminance, color, orientation). Instead, dogs attend selectively to certain item classes.

In our sample, relative to the frequency of the items in their view, dogs fixated proportionally the most to buses, plants, pavement, people, and construction equipment. Of these, buses, people, and construction equipment were all relatively uncommon in the dogs' field of view. Dogs looking to plants relatively often was unexpected, as plants are often considered to be background material, similar to the sky which dogs rarely looked to. Further research is needed in an ecologically valid context like the one presented here to further explore what components of plant material dogs are interested in.

We also found suggestions of visual neophilia in dogs, with dogs looking more to unusual and highly variable items, specifically construction equipment. However, dogs were not simply visually attracted to items that are uncommon along our walking route. Infrequently appearing items along our route that are ubiquitous in dogs' daily lives (i.e., benches and chairs) were rarely looked to. It is unlikely that any of the dogs in our sample were encountering entirely novel item classes and future research could consider incorporating novel item classes based on dog's past experiences to evaluate if this interest in construction equipment is related to its rarity in dogs' lives or something more tied to the class itself (large brightly colored equipment placed in unusual situations). In our sample, dogs did fixate differently to different classes of items. Some classes like benches and chairs or signs were consistently uninteresting to dogs, but other classes, like people, had significant variation between dogs. The relative consistency in item class presence suggests there may be possible true individual differences in attention to stimuli, including attention to social versus nonsocial item classes. This is an interesting interaction, and future work should aim to detect more nuanced individual differences, potentially by recruiting more dogs from a variety of breeds or with more diverse life experiences. Despite this, our findings of individual differences must be treated as preliminary, due to variations in factors such as the time of day, and the presence of mobile item classes (e.g., people, buses) between walks, and variations in the behavior of different dog−guardian dyads (we return to this below).

Like human adults (de Winter et al., 2021; Simpson et al., 2019), dogs frequently attend to classes (i.e., pavement, horizontal plants) that are directly related to ground navigation, helping to find a smooth path and monitor for potential obstacles. Prior work in children (i.e., Franchak & Adolph, 2010) has suggested that children also fixate on obstacles in their environment while wayfinding to a similar extent. Future work could explore whether dogs' fixation patterns to obstacles not related to wayfinding in their environment are similar to children, both in their frequency and timing. This would provide interesting comparisons to what is attention-grabbing between the two species. Young children and dogs' similar height means that the obstacles and items they encounter in daily life are approached from a similar visual angle, and future work can consider direct comparisons to how they direct their visual attention in complex real-world environments.

Future research can expand on the item classes identified here. Anecdotally, we noticed that on many fixations to buildings, dogs were specifically looking to the doors and windows of buildings. This pattern was also observed with buses, with dogs often looking at the door of the bus. It is possible that dogs are looking to these portions of buildings and buses because of anticipated functionality (i.e., dogs look to doors of the buildings because they could enter through the door) or for anticipatory social reasons (i.e., people often appear in doorways so doorways are more interesting because of this potential). While this was an interesting observation, due to the limited spatial accuracy of the eye-tracking system given the variable outdoor lighting conditions, we were not able to determine with confidence what component of buildings dogs were looking to on all instances. With increased camera resolution or image enhancement, future work can explicitly identify whether, for instance, dogs look more to the door of buildings and cars than they do to the walls of buildings and bumpers or tires of cars (strictly nonsocial components). This would provide additional insight into dogs' potential preference for visually attending to social parts of their environment, which could further our understanding of how dogs' complete complex social tasks.

Developing a more nuanced understanding of dogs' attention to social and potentially social areas of their environment would also provide an interesting point of comparison to other nondomesticated canids. While there are no eye-tracking studies in other canid species, prior work comparing dogs' and wolves' gaze behavior has suggested that domestication led to an increased attention to humans and their faces, evidenced in dogs increased looking back behaviors and general visual-social attention (Gácsi et al., 2005; Miklósi et al., 2003). Further, comparative studies can also explore how children and dogs visually engage with social partners on comparable tasks, such as when following a point or working with a partner toward a shared goal. We know that children look to social partners and engage in joint attention tasks (i.e., Yu & Smith, 2013) and we could use a similar approach in dogs to explore if their visual engagement patterns are similar.

Reactive dogs (those that have a strong arousal response to a given stimuli such as a dog or a person on a bicycle) would also be promising candidates for use in this paradigm to understand how they scan their environments and how they, temporally, respond to triggers. Future work can consider applying this method, using what dogs' find visually interesting to create a predictive model of reactive dogs' behavior. This could be particularly important for working dogs, predicting suitability for service roles based on how dogs distribute their visual attention. We also found, descriptively, that other features of item classes such as movement were also not large drivers of dog attention. However, many of our item classes moved in a subtle way where the video quality made potential movement challenging to identify. Future work can consider staging instances of identical classes in motion versus stationary to formally examine the impact of movement on dogs' visual attention.

Another potential source of social input is the dogs' guardian. In our study, dog guardians were instructed to walk their dogs as they would normally. No dog guardians actively directed their dogs' visual attention; however, there were other differences in the behaviors of dog guardians while walking their dogs. In particular, guardians used different walking styles, varying in how far their dog typically is from them, and in whether they provided their dog with verbal navigational directions. It is possible that these different walking styles may have contributed to differences in dog gaze behavior between different dog–guardian dyads. Limited conclusions can be drawn about these different walking styles at this time given our limited sample. While in this study we opted to aim for ecological validity and encourage normal walking styles, future work can examine different interaction styles between dogs and their guardians.

While we excluded intensive sniffing bouts where it was unlikely dogs were engaging in intentional visual processing of the items in their view, there is still the broader question of the relationship between dogs' visual and olfactory attention. Recent work has suggested that dogs' olfaction may be integrated into their visual processing, perhaps into a joint representation of their visual and olfactory environment (Andrews, Pascalau, Horowitz, Lawrence, & Johnson, 2022). Therefore, the odors in dogs' environment may be encouraging them to orient their visual attention to a particular visual target or region (and visual attention may similarly be guiding olfaction). In these instances, unlike in our sniffing bouts, attention could still be visual as items are at a distance and reasonable to process visually. Future work can consider examining how dogs integrate olfactory and visual features in a naturalistic context, such as examining how search and rescue or other detection dogs direct their gaze while searching for a person or item.

This study was the first to record how dogs observe their physical environment in a naturalistic setting. Building upon this understanding, we can now expand into how dogs complete more complex social tasks. Very little is currently known about how, visually, dogs build a bond through play with a human companion or how they navigate a complex physical environment with social guidance, such as in dog agility. Understanding the visual behaviors that dogs are utilizing to complete daily tasks, and how those differ from key visual features seen in human–human interactions, will provide insight into social learning and cooperation in a unique cross-species context. Additionally, being able to measure how dogs visually interact with the world while completing tasks is a first step to building an eventual model of dogs' visual behaviors, something with significant implications for both working dog and Artificial Intelligence training.

Author contributions

Madeline H. Pelgrim: Conceptualization, methodology, formal analysis, investigation, data curation, writing—original draft, visualization. **Shreyas Sundara Raman**: Methodology, software, validation, formal analysis, data curation, writing—original draft. **Thomas Serre**: Methodology, resources, writing—review and editing. **Daphna Buchsbaum**: Supervision, conceptualization, methodology, resources, formal analysis, writing—review and editing, funding acquisition.

Acknowledgments

Thank you to our participants, their people, the research assistants of the Brown Dog Lab, and the team at PositiveScience. A preliminary version of this data was presented at the 44th Annual Meeting of the Cognitive Science Society and published in the proceedings (Pelgrim & Buchsbaum, 2022). This paper contains significantly expanded data analysis, results, and discussion, as well as a larger sample size.

Funding

Thank you to our funding sources: MHP: NIH Grant Number T32MH115895. TS: ONR (N00014-19-1-2029), NSF (IIS-1912280 and EAR-1925481), DARPA (D19AC00015), NIH/NINDS (R21 NS 112743), and the ANR-3IA Artificial and Natural Intelligence Toulouse Institute (ANR-19-PI3A-0004). Computing hardware supported by NIH Office of the Director grant S10OD025181 and the Center for Computation and Visualization (CCV).

Conflict of interest

The authors declare no conflict of interest.

Data availability statement

The code used for computer vision is available here (https://github.com/Shreyas-S-Raman/dogvision-eyefixations) and all data used is available here (https://osf.io/resj5/?view_only=cb2e89a7abda4d779eec2770fafe5).

Ethics approval statement

This study was approved by Brown University's Institutional Animal Care and Use Committee (IACUC) Protocol 20-05-0002. Procedures were in accordance with the ASAB/ABS Guidelines for the Use of Animals in Research and complied with the United States Department of Agriculture Animal Welfare Act & Regulations.

Notes

1 We were initially interested in recording conspecifics and/or prey animals (i.e., birds and squirrels) that were in dogs' field of view or gazed at by dogs. Unfortunately, neither of these categories were observed from the dogs' perspectives on the walks so these were not coded for, resulting in a total of 15 item classes present and observed on walks.

2 For each analysis of fixations described, we also conducted an alternative analysis using a winner takes all approach, meaning that for each fixation, we assigned a single target of fixation based on the majority item in the fixated region. This did not significantly change the results.

3 In addition to the results presented here, we also evaluated the proportion that classes were in view from a binary perspective, evaluating the proportion of total fixations that they were present rather than the proportion of time. These results and visualizations are consistent with the results presented here and are presented in Supplementary Materials.

References

Andrews, E. F., Pascalau, R., Horowitz, A., Lawrence, G. M., & Johnson, P. J. (2022). Extensive connections of the canine olfactory pathway revealed by tractography and dissection. *Journal of Neuroscience*, *42*(33), 6392–6407. https://doi.org/10.1523/JNEUROSCI.2355-21.2022

Bylinskii, Z., Judd, T., Durand, F., Oliva, A., & Torralba, A. (2012). *MIT Saliency Benchmark*. Retrieved from http://saliency.mit.edu/

Bylinskii, Z., Judd, T., Oliva, A., Torralba, A., & Durand, F. (2019). What do different evaluation metrics tell us about saliency models? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *41*(3), 740–757. https://doi.org/10.1109/TPAMI.2018.2815601

Byosiere, S.-E., Chouinard, P. A., Howell, T. J., & Bennett, P. C. (2018). What do dogs (*Canis familiaris*) see? A review of vision in dogs and implications for cognition research. *Psychonomic Bulletin & Review*, *25*(5), 1798–1813. https://doi.org/10.3758/s13423-017-1404-7

Clerkin, E. M., Hart, E., Rehg, J. M., Yu, C., & Smith, L. B. (2017). Real-world visual statistics and infants' first-learned object names. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711), 20160055. https://doi.org/10.1098/rstb.2016.0055

de Winter, J., Bazilinskyy, P., Wesdorp, D., de Vlam, V., Hopmans, B., Visscher, J., & Dodou, D. (2021). How do pedestrians distribute their visual attention when walking through a parking garage? An eye-tracking study. *Ergonomics*, *64*(6), 793–805. https://doi.org/10.1080/00140139.2020.1862310

Einhäuser, W., Moeller, G. U., Schumann, F., Conradt, J., Vockeroth, J., Bartl, K., Schneider, E., König, P. (2009). Eye-head coordination during free exploration in human and cat. *Annals of the New York Academy of Sciences*, *1164*, 353–366. https://doi.org/10.1111/j.1749-6632.2008.03709.x

Franchak, J. M., & Adolph, K. E. (2010). Visually guided navigation: Head-mounted eye-tracking of natural locomotion in children and adults. *Vision Research*, *50*(24), 2766–2774. https://doi.org/10.1016/j.visres.2010.09.024

Franchak, J. M., Kretch, K. S., Soska, K. C., & Adolph, K. E. (2011). Head-mounted eye tracking: A new method to describe infant looking. *Child Development*, *82*(6), 1738–1750. https://doi.org/10.1111/j.1467-8624.2011.01670.x

Gácsi, M., Győri, B., Miklósi, Á., Virányi, Z., Kubinyi, E., Topál, J., & Csányi, V. (2005). Species-specific differences and similarities in the behavior of hand-raised dog and wolf pups in social situations with humans. *Developmental Psychobiology*, *47*(2), 111–122. https://doi.org/10.1002/dev.20082

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2018). *Mask R-CNN*. Retrieved from http://arxiv.org/abs/1703.06870

Hunnius, S., & Bekkering, H. (2010). The early development of object knowledge: A study of infants' visual anticipations during action observation. *Developmental Psychology*, *46*(2), 446–454. https://doi.org/10.1037/a0016543

Jayaraman, S., & Smith, L. B. (2020). The infant's visual world: The everyday statistics for visual learning. In J. J. Lockman & C. S. Tamis-LeMonda (Eds.), *The Cambridge handbook of infant development* (1st ed., pp. 549–576). Cambridge University Press. https://doi.org/10.1017/9781108351959.020

Judd, T., Durand, F., & Torralba, A. (2012). A benchmark of computational models of saliency to predict human fixations. MIT Technical Report.

Kano, F., & Tomonaga, M. (2013). Head-mounted eye tracking of a chimpanzee under naturalistic conditions. *PLoS ONE*, *8*(3), e59785. https://doi.org/10.1371/journal.pone.0059785

Karl, S., Boch, M., Virányi, Z., Lamm, C., & Huber, L. (2020). Training pet dogs for eye-tracking and awake fMRI. *Behavior Research Methods*, *52*(2), 838–856. https://doi.org/10.3758/s13428-019-01281-7

Karl, S., Boch, M., Zamansky, A., van der Linden, D., Wagner, I. C., Völter, C. J., Lamm, C., Huber, L. (2020). Exploring the dog−human relationship by combining fMRI, eye-tracking and behavioural measures. *Scientific Reports*, *10*(1), 22273. https://doi.org/10.1038/s41598-020-79247-5

Kis, A., Hernádi, A., Miklósi, B., Kanizsár, O., & Topál, J. (2017). The way dogs (*Canis familiaris*) look at human emotional faces is modulated by oxytocin. An eye-tracking study. *Frontiers in Behavioral Neuroscience*, *11*, 210. https://doi.org/10.3389/fnbeh.2017.00210

Kretch, K. S., Franchak, J. M., & Adolph, K. E. (2014). Crawling and walking infants see the world differently. *Child Development*, *85*(4), 1503–1518. https://doi.org/10.1111/cdev.12206

Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., & Dollár, P. (2015). *Microsoft COCO: Common Objects in Context*. Retrieved from http://arxiv.org/abs/1405.0312

Miklósi, Á., Kubinyi, E., Topál, J., Gácsi, M., Virányi, Z., & Csányi, V. (2003). A simple reason for a big difference: Wolves do not look back at humans, but dogs do. *Current Biology*, *13*(9), 763–766. https://doi.org/10.1016/S0960-9822(03)00263-X

Pelgrim, M. H., & Buchsbaum, D. (2022) Categorizing Dogs' Real World Visual Statistics. *Proceedings of the 44th Annual Conference of the Cognitive Science Society*, *44*. https://escholarship.org/uc/item/769022r9

Pelgrim, M. H., Espinosa, J., & Buchsbaum, D. (2022). Head-mounted mobile eye-tracking in the domestic dog: A new method. *Behavior Research Methods*, *55*, 1924–1941. https://doi.org/10.3758/s13428-022-01907-3

Reynolds, G. D. (2015). Infant visual attention and object recognition. *Behavioural Brain Research*, *285*, 34–43. https://doi.org/10.1016/j.bbr.2015.01.015

Rodin, I., Furnari, A., Mavroeidis, D., & Farinella, G. M. (2021). Predicting the future from first person (egocentric) vision: A survey. *Computer Vision and Image Understanding*, *211*, 103252. https://doi.org/10.1016/j.cviu.2021.103252

Shepherd, S. V., & Platt, M. L. (2006). Noninvasive telemetric gaze tracking in freely moving socially housed prosimian primates. *Methods*, *38*(3), 185–194. https://doi.org/10.1016/j.ymeth.2005.12.003

Simpson, J., Thwaites, K., & Freeth, M. (2019). Understanding visual engagement with urban street edges along non-pedestrianised and pedestrianised streets using mobile eye-tracking. *Sustainability*, *11*(15), 15. https://doi.org/10.3390/su11154251

Smith, L. B., Jayaraman, S., Clerkin, E., & Yu, C. (2018). The developing infant creates a curriculum for statistical learning. *Trends in Cognitive Sciences*, *22*(4), 325–336. https://doi.org/10.1016/j.tics.2018.02.004

Smith, L. B., Yu, C., Yoshida, H., & Fausey, C. M. (2015). Contributions of head-mounted cameras to studying the visual environments of infants and young children. *Journal of Cognition and Development*, *16*(3), 407–419. https://doi.org/10.1080/15248372.2014.933430

Somppi, S., Törnqvist, H., Hänninen, L., Krause, C. M., & Vainio, O. (2014). How dogs scan familiar and inverted faces: An eye movement study. *Animal Cognition*, *17*(3), 793–803. https://doi.org/10.1007/s10071-013-0713-0

Somppi, S., Törnqvist, H., Hänninen, L., Krause, C., & Vainio, O. (2012). Dogs do look at images: Eye tracking in canine cognition research. *Animal Cognition*, *15*(2), 163–174. https://doi.org/10.1007/s10071-011-0442-1

Somppi, S., Törnqvist, H., Kujala, M. V., Hänninen, L., Krause, C. M., & Vainio, O. (2016). Dogs evaluate threatening facial expressions by their biological validity – Evidence from gazing patterns. *PLOS ONE*, *11*(1), e0143047. https://doi.org/10.1371/journal.pone.0143047

Törnqvist, H., Somppi, S., Kujala, M. V., & Vainio, O. (2020). Observing animals and humans: Dogs target their gaze to the biological information in natural scenes. *PeerJ*, *8*, e10341. https://doi.org/10.7717/peerj.10341

Walther, D., & Koch, C. (2006). Modeling attention to salient proto-objects. *Neural Networks*, *19*(9), 1395–1407. https://doi.org/10.1016/j.neunet.2006.10.001

Watalingam, R. D., Richetelli, N., Pelz, J. B., & Speir, J. A. (2017). Eye tracking to evaluate evidence recognition in crime scene investigations. *Forensic Science International*, *280*, 64–80. https://doi.org/10.1016/j.forsciint.2017.08.012

Yorzinski, J. L., Patricelli, G. L., Babcock, J. S., Pearson, J. M., & Platt, M. L. (2013). Through their eyes: Selective attention in peahens during courtship. *Journal of Experimental Biology*, *216*(16), 3035–3046. https://doi.org/10.1242/jeb.087338

Yu, C., & Smith, L. B. (2013). Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand coordination. *PLOS ONE*, *8*(11), e79659. https://doi.org/10.1371/journal.pone.0079659

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Supporting Information