Title
Identifying Body Parts in the Spatial Context of Pairwise Relations: Human Psychophysics and Model Simulations

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 45(45)

Authors
Liu, Ziwei
Kersten, Daniel

Publication Date
2023

Peer reviewed

Identifying Body Parts in the Spatial Context of Pairwise Relations: Human Psychophysics and Model Simulations

Ziwei Liu (liu00964@umn.edu)
Department of Psychology, University of Minnesota

Daniel Kersten (kersten@umn.edu)
Department of Psychology, University of Minnesota

Abstract

The ability to detect and analyze a human figure is key to our survival and social interactions. Efficient and robust identification of body parts can help to interpret images when bodies are partially occluded. While previous studies emphasized the configural processing of whole bodies using simplified stimuli, it still remains unclear how spatial contexts about body parts are integrated to resolve ambiguities (e.g., from occlusion) regarding the identities of or spatial relations among body parts. In a series of online experiments, we asked human observers to identify an ambiguous target body part in the presence of another "context" part. Our results showed that humans can use various amounts of spatial context to discount local ambiguities in natural images of pairs of parts, and are sensitive to low- and mid-level cues such as alignment and connectedness. Further simulations using deep convolutional neural networks (DCNNs) exhibited comparable similar sensitivity to spatial context variations, despite being trained solely on local part appearances without explicit prior knowledge of body structure. However, discrepancies between human and model performance were also observed, with humans showing greater sensitivity to spatial relations compared to the models. Our findings suggest that while both humans and models utilize low- and mid-level features for body part recognition, humans possess a stronger prior knowledge of body structure that influences their perception. These results contribute to our understanding of how humans integrate spatial context to resolve ambiguities and provide insights into the computational mechanisms underlying body perception.

Keywords: human body parts; spatial relations; natural images; psychophysics; model simulation

Introduction

Recognizing and interpreting the rich visual information conveyed by human bodies is a computationally demanding task that is crucial for human survival and daily social interactions. This task involves dealing with substantial appearance variations caused by factors such as clothing, viewpoint, lighting, poses, and occlusion. To accomplish this, the human brain must engage in complex computations to extract meaningful features and integrate them into coherent representations of the body. Previous research has primarily focused on configural processing of bodies, which proposes that the human visual system integrates information from multiple body parts to form a coherent whole-body representation (Reed et al., 2003, 2006; Stekelenburg & de Gelder, 2004). The human visual system relies on specific configural information, such as relative positions and distances among body parts, to form a complete global representation of a body, especially in simplified stimuli such as stick figures and point-light walkers (Johansson, 1973). However, an open question remains as to how local information extracted across different body parts can be integrated and utilized to resolve local ambiguities in body figures, which could facilitate establishing a configural representation of whole bodies, particularly in scenarios where only limited visible parts are available due to partial or heavy occlusion.

Despite limited spatial context, previous research has demonstrated that the human visual system is capable of effectively utilizing local information in natural images to consistently recognize partially visible objects and object parts, even when they are severely reduced in size or resolution (McDermott, 2004; Ullman et al., 2016). This capability extends to body perception, with recent research showing that humans can identify body parts with above-chance-level accuracy, even when only 40% of the part is visible in an image patch (Liu & Kersten, 2022). In fact, in natural images, views of only a limited number of parts can be sufficient to constrain the localization and interpret the identities of other hidden parts as well as the configuration of a full body posture. In contrast to relying on whole body structures or motions for part identification in context-free body figures (e.g., point-light walkers and stick figures), each sub-region of a static natural image offer a wealth of spatial contextual cues necessary for interpreting complex everyday scenes. However, it is currently not well understood how the spatial context of body parts in natural images contribute to resolving local ambiguities (e.g., from occlusion) and facilitating the recognition of identities or spatial relationships among body parts. Computational studies have demonstrated that by recognizing individual parts and their spatial relationships with only one nearby part, it is possible to reconstruct whole 2D poses from natural images without the need for a whole-body prior model (Chen & Yuille, 2014). In order to investigate this question in humans, we examined human ability to identify natural image patches of pairs of body parts with varying spatial contexts. The utilization of natural images allowed us to encompass the real-world complexity and variations in body part appearance and context. By incorporating rich low- (e.g., edge, color, brightness) and mid-level (e.g., contour, shape, texture) features inherent in these natural images, we could explore the role and importance of these features in body perception in a more ecologically valid manner. Moreover, by focusing on pairs of body parts, we are able to investigate

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society.*

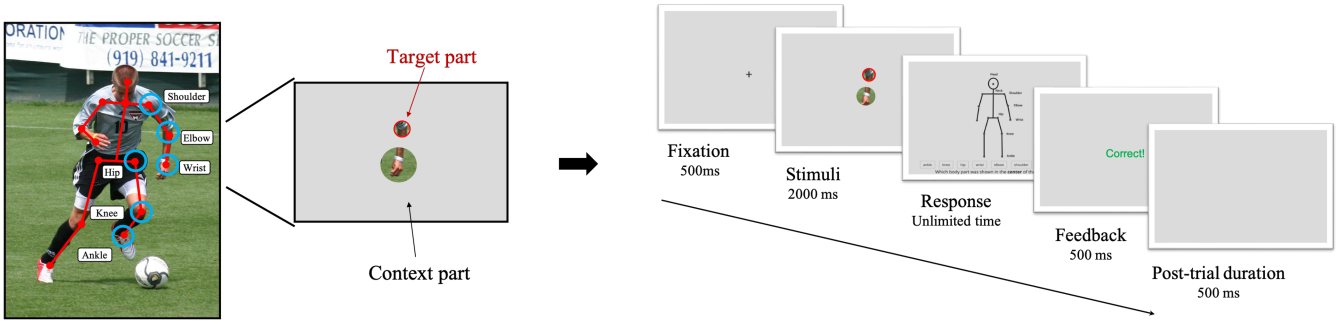

Figure 1: Illustration of stimuli generation (left) and experimental procedures (right).

local-to-local interactions between body parts with systematically varied contextual information, while minimizing potential confounds from global configural processing.

In a series of online experiments, we presented human observers with pairs of body parts and asked them to identify an ambiguous partially visible target body part in the presence of another "context" part (Figure 1). In Experiment 1, we varied the spatial relations (proper or improper) between the target parts and the context parts, and included image patches of different sizes to examine the effect of varying amounts of spatial context on the informativeness of the context parts (Figure 2a and 2b). In Experiment 2, we investigated the influence of local feature congruency between target and context parts. Specifically, we rotated the context parts at relative angles ranging from 0 to 180 degrees while maintaining their size constant (Figure 2c). In this way, we introduced "alignment" as another important type of spatial relation between body parts, where alignment was gradually disrupted from 0 to 90 degrees and recovered from 90 to 180 degrees in a reversed direction. We also considered spatial relations of adjacency, and specifically examined the factor of "connectedness" between two body parts. Different pairs of parts were divided into groups of skeletally connected or unconnected pairs (Figure 2a) for data analysis in both experiments. Research has shown that humans are sensitive to proper spatial context and congruency of features among local regions and parts in general scenes (Bar & Ullman, 1996; Mannion et al., 2015; Kaiser & Peelen, 2018). In this study, we hypothesized that improper spatial relations or incongruent contextual cues between body parts would disrupt the identification of an ambiguous body part, whereas proper spatial relations and congruent cues would facilitate the disambiguation of body parts.

To gain further insights, we contrasted human behavior with computational models on the same tasks to assess their sensitivity to spatial context in pairwise part relations. Specifically, we fine-tuned two widely-used feedforward DCNNs, VGG-19 and ResNet-50, by training both models using pairs of intact body part image patches with proper spatial relations. The VGG-Networks (Simonyan & Zisserman, 2014) are renowned for their simplest feedforward architecture that closely resemble the primate ventral visual stream (Tripp, 2017), while the residual neural networks (ResNets) are known for their faster speed and improved performance with deeper architecture and more efficient computational techniques (He et al., 2016). We evaluated the performance of both models on the same stimuli sets with partially visible body parts and systematically varied spatial context, as presented to human observers in Experiments 1 and 2. We expected to observe similarities and differences in the performance and sensitivity of humans and models to varying spatial contexts between body parts. Taken together, our results provided both behavioral and computational evidence in the human visual representations of various spatial relations between pairs of body parts. The results could pave the way for future exploration of finer-grained whole-body representations and underlying neural mechanisms.

Method

Human Psychophysics

Stimuli Image patches of body parts were generated from Leeds Sports Pose (LSP) dataset (Johnson & Everingham, 2010), a widely used benchmark human pose dataset composed of 2000 natural images of sportspersons. To ensure a good resolution, we only used images with a head length (i.e., the Euclidean distance between the head and neck coordinates) exceeding approximately 60 pixels, which were then downscaled to have a standard head length of 60 pixels. We selected a subset of 340 images for stimuli generation, which included diversified poses from a variety of activities with relatively balanced viewpoints (i.e., frontal, back, and side views), pose typicality, and occlusion levels to ensure the representativeness of the dataset.

We first generated image patches of individual body parts to acquire the baseline performance on single body part identification, which were then used to select pairs of target parts and context parts (see details in "Baseline data" and selection criteria below). In the experiments, six major joints of a headless body: shoulder, elbow, wrist, hip, knee, and ankle, were tested. The body part categories were not further subdivided into left and right sides. The image patches were centered on one of the six body parts (on either side) of the primary person in each image, and were cropped into circular shapes of three sizes: 36, 48, and 60 pixels in diameter, corresponding to visual angles of 3 to 5 degrees respectively. The sizes were

chosen to match the estimated head length (i.e., 60 pixels) of the primary person in each image after scaling, and to ensure that a single intact body part with minimal background interference was presented at the largest size (i.e., at 60 pixels diameter).

After the baseline testing, pairs of target and context parts were selected for the formal experiments based on the following criteria: (1) target image patches with an average accuracy below 50% at the smallest size (i.e., 36 pixels) were chosen, while context image patches were selected with an average accuracy above 80% at the largest size (i.e., 60 pixels); (2) both target and context image patches were cropped from the same person in an image; (3) neither target nor context image patch included any other body parts, i.e., there was no significant occlusion that could affect the validity of the ground truth label. This way, for each pair of body parts, the targets were "hard image patches" that were difficult to recognize by their own local features, while the context image patches were reliably recognizable and could provide informative cues at their full appearance. There were a total of 72 pairs of target and context image patches qualified for the formal experiments, and another 12 pairs were selected as example stimuli for practice trials.

In Experiment 1, we varied the spatial relations between body parts for each pair of body parts in two conditions: proper spatial relation, where the relative position was retained as in the original image, and improper spatial relation, where the body parts were presented side-by-side. We also varied the sizes of the context image patch, with circles of 36, 48, and 60 pixels in diameter, respectively. In Experiment 2, we manipulated the congruency between two image patches in each pair by rotating the context image patches, while maintaining their relative spatial positions. The context image patches were rotated at 0, 30, 60, 90, 135, and 180 degrees. The sizes of the target and context image patches were fixed at 36 and 60 pixels, respectively.

Data Collection The data was collected online from Prolific.co, an online platform for behavioral studies (Palan & Schitter, 2018). Prior to both baseline testing and formal experiments, a "virtual-chinrest" testing procedure (Li et al., 2020) was adopted. Stimuli were calibrated based on the estimated viewing distance and measured monitor size, such that image patches of the smallest size (i.e., 36 pixels) subtended approximately 3 degrees of visual angle. Abnormal operations, such as exiting full-screen mode or switching to other windows during the experiment, were recorded to exclude data from later analysis.

Baseline Data Sixty-four observers (28 females, 36 males) were tested on individual image patches of body parts that were later used as baseline measurements for selecting qualified pairs of image patches in the formal experiments. Each observer completed 1020 trials of a body part identification task, where the size of the image patch varied from trial to trial, with diameters of 36, 48, or 60 pixels. The number of trials was balanced across different levels of sizes and categories of body parts for each participant. This resulted in an average of 15 data points for each image patch in the baseline testing set.

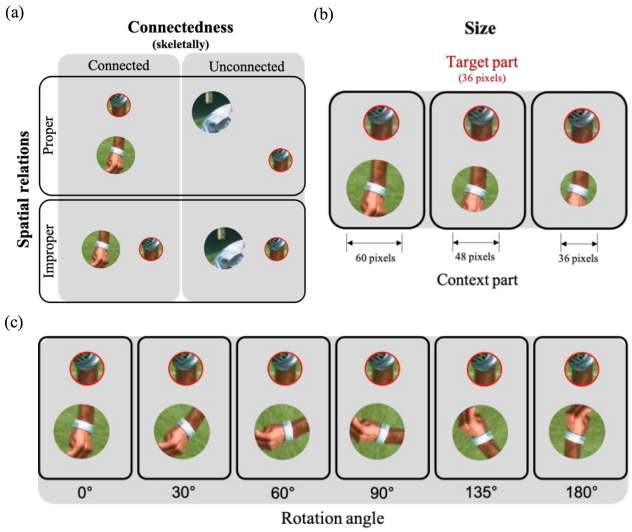

Figure 2: Illustration of stimuli generation (left) and experimental procedures (right).

Experiment 1: Spatial Relations

Two groups of 30 participants each were recruited for the proper and improper spatial relation conditions, respectively. The task procedure was the same for both conditions, as illustrated in Figure 1. In each trial, participants first fixated on a cross for 500 ms, after which a pair of image patches was presented at the center of the screen for 2000 ms. The target image patch was circled in red. Participants were then asked to select a label name from the six categories to indicate which body part they saw in the target aperture. Participants were encouraged to provide their best guess if they were unsure of the answer, and at least one response was required for each trial. There were no time limits for responses, and feedback was given after each response. The next trial was automatically initiated 500 ms after the end of the previous trial. Each participant completed 72 trials, with each pair of image patches presented once at a randomly selected size of the context image patch (Figure 2b). The number of trials for each level of the size of context parts and categories of target parts was balanced across all trials for each participant. In the proper spatial relation condition, pairs of body parts were presented with proper relative positions as in the original image, while in the improper spatial relation condition, two body parts were presented side by side, as shown in Figure 2a.

Experiment 2: Congruency of Spatial Context

A total of 121 observers participated in Experiment 2. The experimental procedure was the same as in Experiment 1, ex-

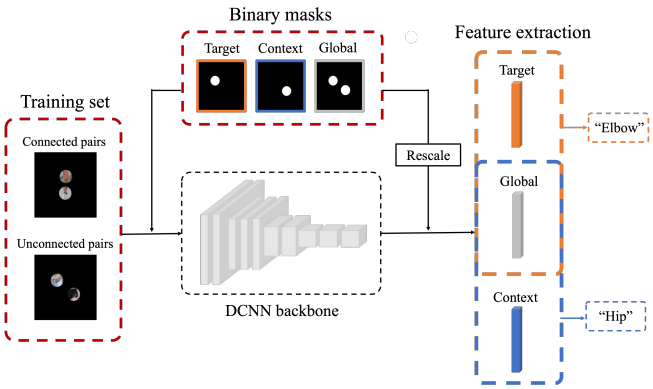

Figure 3: Illustration of model simulations. Binary masks were shown for the example image of unconnected pair. See text for detailed training procedure.

cept that each participant completed 36 formal trials, with a balanced number of trials at each level of the rotation angle of the context part (Figure 2c) and the category of the target body part in a random order.

Model Simulations

Model Training We build our models on top of two state-of-the-art feedforward deep networks, i.e., VGG-19 and ResNet-50, both pretrained on the ImageNet classification (Deng et al., 2009). Using the 1660 held-out images from the LSP dataset, we created a training set consisting of pairs of both connected and unconnected parts. To ensure consistency with human psychophysics, all images were scaled to maintain a head length of approximately 60 pixels in the primary human figure. The models were trained solely on intact body parts with proper spatial relations, with all target and context parts cropped at Size 60 without rotation, and presented with retained relative positions. We included all possible combinations of connected pairs among the six body parts with side separation (i.e., only connecting the right wrist to the right elbow), and to maintain a balance between the number of connected and unconnected pairs, we randomly sampled a subset from each possible combination of unconnected pairs.

Each pair was centered on a black background image, adjusted to a standard size of 300 x 300 pixels based on the furthest distance between all pairs of body parts. During training, three binary masks were utilized per image to selectively include different regions of interest: two local masks to localize the target and context parts respectively, and one global mask that is the union of local masks, which retained information of relative positions and global features (Figure 3). Features extracted from masked local and global regions were concatenated and used for target and context part classification. We applied full-size local masks before the first layer of the network to filter out information outside the target and context body parts, and rescaled local and global masks after the last convolutional layers for feature extraction. The local features for the target and context parts were directly extracted from the last convolutional layer with masking and global average pooling, while the global features were computed by passing the masked features with two additional layers. The two types of features were concatenated together to predict the labels for the target and context parts, respectively.

Model Evaluation The models were evaluated on the same stimulus sets used in human psychophysics experiments. Specifically, in Experiment 1, the models were tested with the target parts fixed at Size 36 and evaluated on all three sizes (36-, 48- or 60-pixel diameter) of context parts. The context parts were presented in either proper spatial relations with retained relative positions or improper spatial relations presented side-by-side, relative to the target part. In Experiment 2, the models were tested on all rotation levels from 0 to 180 degrees. Both connected and unconnected pairs were included. Binary masks were applied to each image, as during the training phase, to filter the local or global regions of interest.

Results

Experiment 1: Effects of Spatial Relations

In Experiment 1 (Figure 4A(a)), we found significant main effects of both spatial relations between target and context parts ($F(1,58) = 69.49, p < .01$) and the size of the context parts ($F(2,116) = 16.17, p < .01$), as well as a significant interaction effect between the two variables ($F(2,116) = 7.68, p < .01$). In both proper and improper spatial relation conditions, observers performed significantly better when the two body parts are skeletally connected than with unconnected pairs ($F(1,58) = 4.90, p < .05$). Specifically, in the proper spatial relation condition, accuracy was consistently enhanced in identifying ambiguous target parts, compared to the baseline performance of isolated target parts without the presence of context parts, for both connected and unconnected pairs. The accuracy increased linearly as a function of the size of the context part. Whereas in the improper spatial relation condition, accuracy increased monotonically as the size of the context part increased in connected pairs, but dropped in unconnected pairs at the largest context size, suggesting that improper spatial relations caused greater interference.

Experiment 2: Effects of Spatial Context Congruency

In Experiment 2 (Figure 4B(a)), we observed a significant effect of the congruency of spatial context ($F(5,600) = 30.81, p < .01$) and connectedness between pairs of parts ($F(1,120) = 64.17, p < .01$), as well as a significant interaction effect between the two variables ($F(5,600) = 7.08, p < .01$). Similarly, observers performed better in connected pairs than in unconnected pairs across all rotation angles. Specifically, in unconnected pairs, identification accuracy decreased monotonically as the rotation angle increased from 0 to 180 degrees. Whereas for skeletally connected pairs, the identification accuracy stopped dropping after 90 degrees, where

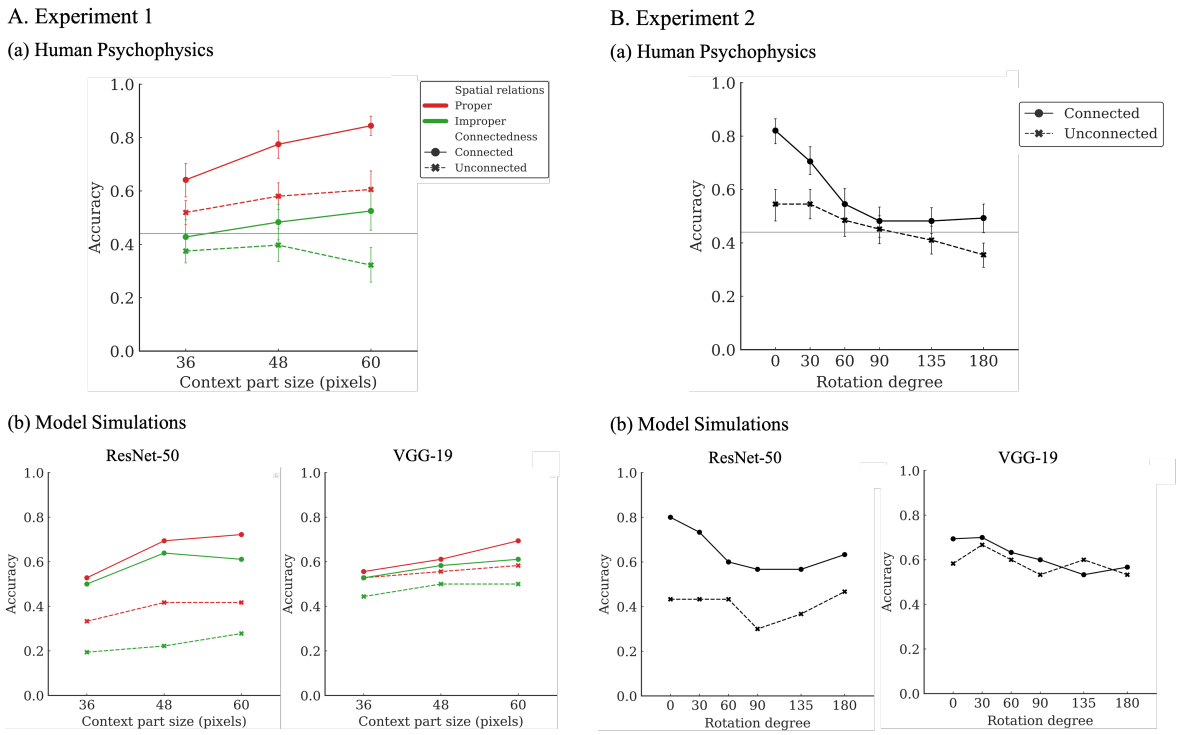

Figure 4: Results in Experiment 1 (A) and Experiment 2 (B). The black horizontal lines represent averaged baseline performance across all target parts at Size 36 (without rotation) in human psychophysics. Error bars represent 95% confidence interval.

there was maximum disruption of alignment between two connected parts. There was an insignificant but slight increase from 90 to 180 degrees, where the alignment gradually recovered in the reversed direction, suggesting human sensitivity to alignment cues regardless of direction between two connected parts.

Model Simulations

In Experiment 1 (Figure 4A(b)), we observed that the accuracy of both models increased as the diameter of the context parts increased from 36 to 48 pixels, but started to saturate or slightly drop as it extended to 60 pixels. Moreover, our results revealed that, unlike human observers, the models were more influenced by the connectedness of pairs of body parts than their spatial relations, particularly in the case of ResNet-50. Specifically, both models showed better performance on connected pairs than unconnected pairs, irrespective of the spatial relations between the target and context parts. However, it is noteworthy that both models still exhibited better performance when body parts were presented in proper spatial relations compared to improper ones, which is consistent with human observers.

In Experiment 2 (Figure 4B(b)), both models exhibited a general decrease in accuracy as the rotation angle increased from 0 to 90 degrees for both connected and unconnected pairs. Mixed patterns were observed after 90 degrees. Specifically, ResNet-50 showed increased accuracy from 90 to 180 degrees, while VGG-19 did not show a clear pattern of accuracy change in this range. Similarly, ResNet-50 was more influenced by the connectedness of pairs than VGG-19, especially after 90 degrees of rotation.

Discussion

Our study aimed to explore the extent to which humans are sensitive to and able to utilize local spatial context in order to resolve local ambiguities in the identity and spatial relations of body parts. Using natural image patches of pairs of body parts, we were able to systematically vary the spatial context between parts, allowing us to quantitatively measure human behavioral performance as functions of different spatial relations. We specifically examined several types of spatial relations relevant to human body and body part perception, including relative positions, connectedness, and congruency between local features such as alignment and direction. In Experiment 1, we found that the proper relative positions between body parts significantly facilitated identification performance, whereas disrupted relative positions caused interference in identification. Additionally, participants performed better in both conditions when two parts were skeletally connected. In Experiment 2, we observed that increasing incongruence between the local features of body parts led to decreased identification performance. However, cues of alignment between the parts mitigated the effect, even when the parts were aligned in a reversed direction. Our results provide

behavioral evidence for human visual representations of various spatial relations between pairs of body parts. It is worth noting that such representation is beyond merely semantic associations between pairs of parts based on their identities regardless of image-dependent local part appearances. We revealed the crucial role of low-level and mid-level representations in the integration of local spatial context, including connectedness and alignment between body parts. However, despite these observations, it remains unclear which specific low- or mid-level image features, such as local evidence of contour alignment, boundary ownership, or medial axis (i.e., the central line connecting joints) of body parts, contribute to this phenomenon. Further research utilizing quantitative measurements of various image features is necessary to fully understand the underlying mechanisms of the visual representation of spatial relations between body parts.

In addition to investigating human psychophysics, we conducted model comparisons that exposed both similarities and differences in the way humans and models represent pairwise part relations. By fine-tuning two DCNNs, we showed that both models gained comparable benefits to humans from proper spatial relations, connectedness, and local congruency between pairs of parts, as well as an increase in the size of contextual parts. Remarkably, despite being trained solely on local part appearances from natural image patches of body parts, without any explicit prior structural knowledge of bodies or excessive architecture engineering, both models exhibited sensitivity to various relational properties between body parts, such as the level of connectedness and alignment. This provided computational evidence that the utilization of low- and mid-level features is crucial in both body part recognition and the representation of pairwise part relations, and that mid-level representational features, such as contour alignment and connectivity, could be efficiently learned from pixel-based low-level features. However, the comparison with human performance also revealed several disparities. Firstly, models showed less sensitivity to spatial relations compared to connectedness between parts, while humans demonstrated a greater susceptibility to disruptions in spatial relations. We speculate the discrepancy stems from the strong prior knowledge of human body structure that humans possess, which imposes stricter constraints on human perception. Secondly, mixed patterns of performance were observed between the two models, with VGG-19 exhibiting higher levels of translational and rotational invariance in terms of the relative positions and degree of rotation angles between parts. We speculate that the discrepancies may be attributed to different downsampling strategies employed by the models, as the max-pooling layers utilized in VGG-19 focus primarily on the maximum value within a local region and thus is less sensitive to spatial variations.

Lastly, there are some limitations to consider and potential avenues for future research in addressing these findings. Firstly, we did not explicitly differentiate between the factors of "connectedness" and "proximity," which are closely related in the spatial relations of body parts. Future investigations on the influence of relational knowledge about proximity on body part perception would be valuable. This exploration could involve exploring whether proximity is based on the inherent body structure, following proximity along body skeletons, or characterized by the frequency of co-occurrence and proximity in relation to the relative positions between parts in natural images. For instance, certain body parts like the "hip" and "wrist" may not be skeletally adjacent but could exhibit frequent spatial proximity in natural occurrences. Previous research has suggested the existence of a distinct representation system in human perception that captures the statistical distributions of typical spatial relations among objects and object parts in the natural world (Bonner & Epstein, 2021; Kaiser et al., 2014, 2019). However, it remains unclear whether such a system extends to body perception and how the brain forms and represents such relational knowledge about body structures. Secondly, while our study primarily investigated pairwise relations between body parts, further research on the integration of multiple body parts would offer a more comprehensive understanding of how local contextual information can be effectively integrated to resolve global ambiguities related to whole-body structure and pose estimation. Moreover, our results have provided empirical support for brain imaging studies on body part representation. Previous studies have identified two sub-areas in the body selective region, the fusiform and extrastriate body areas (EBA and FBA), that exhibit distinct responses to images of whole bodies or body parts (Urgesi et al., 2004, 2007; Taylor et al., 2007). Spatial relation representations of body parts may serve as an important intermediary link between the two sub-areas of the body selective region. Future brain imaging studies that investigate the spatial representation of body parts in these regions are necessary, as they can provide a finer-grained characterization of the hierarchical structure of the body representation.

Conclusion

In summary, the current study provides novel insights into the mechanisms by which humans visually represent various spatial relations between body parts during natural image processing, both behaviorally and computationally. The results highlight the significance of low- and mid-level contextual features employed by human observers, which aid in resolving local ambiguities, in conjunction with higher-level structural knowledge and semantic-level associations between body parts.

Acknowledgments

We would like to thank Shi Chen for his invaluable input and technical support throughout the research process. This work was supported by the National Institutes of Health with grant NIH R01 EY029700.

References

Bar, M., & Ullman, S. (1996). Spatial context in recognition. *Perception*, *25*(3), 343–352.

Bonner, M. F., & Epstein, R. A. (2021). Object representations in the human brain reflect the co-occurrence statistics of vision and language. *Nature communications*, *12*(1), 4081.

Chen, X., & Yuille, A. L. (2014). Articulated pose estimation by a graphical model with image dependent pairwise relations. *Advances in neural information processing systems*, *27*.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 ieee conference on computer vision and pattern recognition* (pp. 248–255).

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 770–778).

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & psychophysics*, *14*, 201–211.

Johnson, S., & Everingham, M. (2010). Clustered pose and nonlinear appearance models for human pose estimation. In *bmvc* (Vol. 2, p. 5).

Kaiser, D., & Peelen, M. V. (2018). Transformation from independent to integrative coding of multi-object arrangements in human visual cortex. *Neuroimage*, *169*, 334–341.

Kaiser, D., Quek, G. L., Cichy, R. M., & Peelen, M. V. (2019). Object vision in a structured world. *Trends in cognitive sciences*, *23*(8), 672–685.

Kaiser, D., Stein, T., & Peelen, M. V. (2014). Object grouping based on real-world regularities facilitates perception by reducing competitive interactions in visual cortex. *Proceedings of the National Academy of Sciences*, *111*(30), 11217–11222.

Li, Q., Joo, S. J., Yeatman, J. D., & Reinecke, K. (2020). Controlling for participants' viewing distance in large-scale, psychophysical online experiments using a virtual chinrest. *Scientific reports*, *10*(1), 1–11.

Liu, Z., & Kersten, D. (2022). Visual recognition of single body parts in natural images. *Journal of Vision*, *22*(14), 4419–4419.

Mannion, D. J., Kersten, D. J., & Olman, C. A. (2015). Scene coherence can affect the local response to natural images in human v1. *European Journal of Neuroscience*, *42*(11), 2895–2903.

McDermott, J. (2004). Psychophysics with junctions in real images. *Perception*, *33*(9), 1101–1127.

Palan, S., & Schitter, C. (2018). Prolific. ac—a subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, *17*, 22–27.

Reed, C. L., Stone, V. E., Bozova, S., & Tanaka, J. (2003). The body-inversion effect. *Psychological science*, *14*(4), 302–308.

Reed, C. L., Stone, V. E., Grubb, J. D., & McGoldrick, J. E. (2006). Turning configural processing upside down: part and whole body postures. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(1), 73.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Stekelenburg, J. J., & de Gelder, B. (2004). The neural correlates of perceiving human bodies: an erp study on the body-inversion effect. *Neuroreport*, *15*(5), 777–780.

Taylor, J. C., Wiggett, A. J., & Downing, P. E. (2007). Functional mri analysis of body and body part representations in the extrastriate and fusiform body areas. *Journal of neurophysiology*, *98*(3), 1626–1633.

Tripp, B. P. (2017). Similarities and differences between stimulus tuning in the inferotemporal visual cortex and convolutional networks. In *2017 international joint conference on neural networks (ijcnn)* (pp. 3551–3560).

Ullman, S., Assif, L., Fetaya, E., & Harari, D. (2016). Atoms of recognition in human and computer vision. *Proceedings of the National Academy of Sciences*, *113*(10), 2744–2749.

Urgesi, C., Berlucchi, G., & Aglioti, S. M. (2004). Magnetic stimulation of extrastriate body area impairs visual processing of nonfacial body parts. *Current Biology*, *14*(23), 2130–2134.

Urgesi, C., Candidi, M., Ionta, S., & Aglioti, S. M. (2007). Representation of body identity and body actions in extrastriate body area and ventral premotor cortex. *Nature neuroscience*, *10*(1), 30–31.