

Contributions of Visual Pathway Segregation to Learning Multiple Behaviors with Divergent Visual Demands

Aida Mirebrahimi¹

David Plaut¹

Marlene Behrmann^{1,2}

¹ Neuroscience Institute, Carnegie Mellon University
² Department of Ophthalmology, University of Pittsburgh
amirebra@andrew.cmu.edu

Abstract

The primate visual system can support both object recognition and geometry-based grasping from the same input, while they impose different learning demands on the system. E.g. recognition benefits from invariance to changes in viewpoint and scale, whereas grasping requires sensitivity to those same features. In the brain, these behaviors are supported by two partially-segregated dorsal and ventral pathways. Although these pathways have been shown to differ in their representational properties, it remains unclear whether this anatomical segregation plays a role in resolving the multi-task learning tension between these behaviors. Here we test this using computational models, comparing a fully shared single-pathway (SP) architecture with an early-shared, late-segregated dual-pathway (DP) one, alongside single-task baselines. DP matched SP on recognition while learning grasping faster and generalizing better as distribution shifts grew. The analysis of task gradients showed that segregation reduced recognition-dominated interference in shared representations while preserving early feature sharing. Partial pathway segregation thus acts as an inductive bias for efficient simultaneous learning under competing visual demands.

Introduction

The primate visual system processes shared input through two partially segregated pathways: a ventral pathway supporting object recognition and a dorsal pathway supporting visually guided action such as grasping (Goodale & Milner, 1992; Mishkin et al., 1983; Ungerleider & Mishkin, 1982). Both behaviors rely on object shape but on different aspects of it (Freud et al., 2017): recognition benefits from invariance to viewpoint, scale, and position (DiCarlo et al., 2012; Grill-Spector & Malach, 2004), whereas grasping requires sensitivity to those same variables (Culham et al., 2003; Murata et al., 2000). Learning to serve both from a single input therefore poses a computational tension the visual system must resolve. Decades of work have characterized the functional differences between these pathways (Shmuelof & Zohary, 2005; Ungerleider & Haxby, 1994), yet whether this anatomical segregation itself aids learning under divergent demands remains unclear. The question cannot be studied in the brain, where anatomy is fixed. However,

computational models grounded in neuro-anatomy let us manipulate architecture directly during learning. Recent work shows that training a two-pathway neural network on separate dorsal- and ventral-aligned objectives makes them differentially predict dorsal and ventral neural responses (Choi et al., 2023), but does not address how pathway segregation contributes to learning behaviors simultaneously from shared input under representational conflict. We address this gap by jointly training single-pathway and dual-pathway models on recognition and grasping under matched conditions, testing whether partial segregation improves learning efficiency and generalization by reducing task interference.

Methods

Models and training: All models share a capacity-matched AlexNet-style convolutional backbone (Krizhevsky et al., 2012) and differ only after the convolutional trunk. The dual-pathway multitask model (*DP_MT*) splits into recognition and grasping branches, with each loss updating only its branch, while both losses update the backbone. The single-pathway multitask model (*SP_MT*) routes both tasks through one shared sparse pathway, where each layer matches the unit and connection count of *DP_MT*. Single-task baselines use the same *SP_MT* architecture but train on recognition or grasping, alone. Recognition used 15-way cross-entropy; grasping used summed MSE over population codes (20 units per output) for center-of-mass x/y , wrist rotation, and grip aperture. All models and runs used matched optimization (SGD; learning rate = 0.01; momentum = 0.9; batch size = 16).

Dataset: We rendered a shape-centric dataset without color or texture, forcing both tasks to rely on object geometry: 450 3D-meshes across 15 categories (30 instances each), generated from text prompts, and rendered in Blender from 250 viewpoints per instance (Blender Online Community, 2023; Meshy AI, 2023). Each image was paired with a category label and precision-grip targets produced by a normative grasping model, which accurately predicts human behavior (Klein et al., 2020).

Evaluation We evaluated grasping across three increasingly out-of-distribution splits: novel viewpoints (T1),



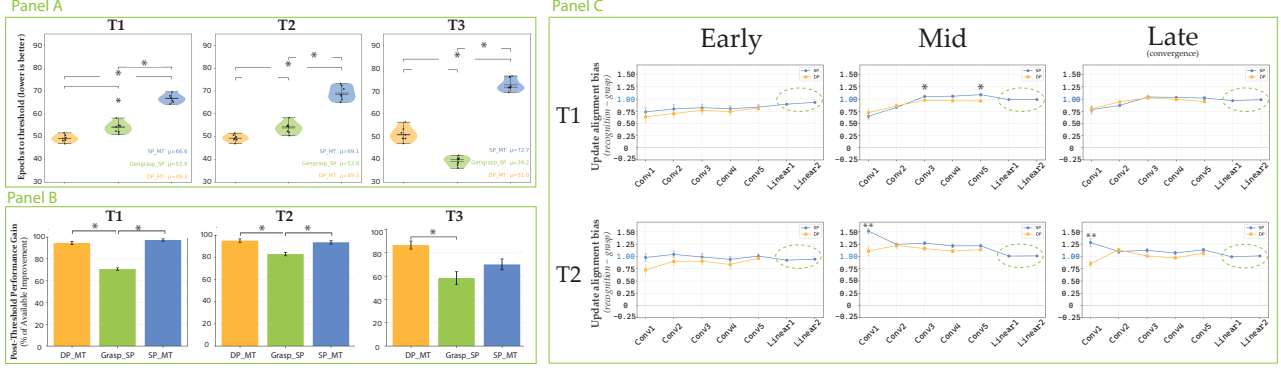


Figure 1: (A) Grasp training efficiency and (B) post-threshold gain for *DP_MT*, *Grasp_SP*, and *SP_MT* across test splits T1–T3. (C) Layer-wise update-alignment bias (recognition – grasp) for *SP_MT* vs. *DP_MT* across training stages on test splits T1–T2. Stars denote BH–FDR-significant differences (* $p < 0.05$; ** $p < 0.01$).

novel instances within trained categories (T2), and held-out categories (T3), using $n = 7$ seeds per model type. Training efficiency (Fig. 1A) was measured as epochs to reach a split-specific test threshold, with lower values indicating faster learning. Converged generalization (Fig. 1B) was measured as post-threshold gain: the fraction of the headroom between threshold performance and the best converged performance achieved by any run in that test split (0% = threshold; 100% = best converged loss across all models and runs). Models were compared using run-label permutation tests with BH–FDR correction. To probe mechanism, we computed recognition and grasp gradients separately on test images and measured how the combined multitask update aligned with each task across layers and training stages on T1 and T2, where recognition labels and gradients were meaningful.

Results

Training Efficiency and Generalization Recognition learning was largely unaffected by architecture: both *DP_MT* and *SP_MT* reached ceiling recognition accuracy within 25–30 epochs, matching the recognition-only baseline, so architectural effects emerged only in grasping. *DP_MT* reached the grasp threshold significantly earlier than *SP_MT* on every split (T1: $\Delta = -17.4$; T2: $\Delta = -19.8$; T3: $\Delta = -21.6$ epochs; all BH–FDR significant), the gap widening with the distribution shift (Fig. 1A). At convergence, both multitask models surpassed *Gen-grasp_SP* on T1 and T2 generalization, with *SP_MT* matching *DP_MT*—joint training thus led to better grasp generalization for novel viewpoints and instances of trained categories (Fig. 1B). Under the strongest shift (T3, unseen categories), only *DP_MT* retained an advantage over the baseline. Thus the *SP_MT* outperformed the baseline only when tasks were more aligned (T1, T2).

Mechanism: Task Update Alignment To probe the mechanism for the grasp learning advantages seen

in *DP_MT*, we measured how the combined multitask update aligned with each task gradient across layers and training stages (Fig. 1C). Update-alignment bias is near zero when updates support both tasks, near one when the update is recognition-dominated with little grasp contribution, and above one when it conflicts with grasp. Across models, layers, and training, updates were strongly recognition-biased, reflecting a large gradient-magnitude imbalance (mean $\log_{10} |g_{\text{rec}}|/|g_{\text{grasp}}| = 2.86 \pm 0.84$). In *SP_MT*, shared late linear layers were almost entirely recognition-aligned (recognition alignment ≈ 1.0) and nearly orthogonal to grasp (grasp alignment ≈ 0.0), whereas *DP_MT* avoided this competition by routing grasp updates through a dedicated branch (green ovals in Fig. 1C). Segregation also affected the earliest shared layer: under the divergent T2 condition, *SP_MT* conv1 updates conflicted with grasp mid-training and at convergence (bias > 1 ; $\Delta \approx 0.41$ mid, 0.43 late, $p_{\text{fdr}} < 0.005$), while *DP_MT* maintained more grasp aligned update < 1 ; no comparable difference appeared on T1. Thus, segregation did not globally rebalance learning, but reduced recognition dominance where task conflict was greatest.

Conclusions

Partial pathway segregation improved grasp learning under simultaneous recognition and grasp training, acting as an inductive bias that reduced recognition-dominated interference while preserving the ability to exploit shared task structure where the tasks overlapped. The absence of a recognition advantage was due to task imbalance: recognition dominated grasping and reached ceiling quickly, so architectural effects surfaced only in grasping. We would expect recognition advantages in *DP_MT* if grasping dominated learning, which we will examine in future analyses. More broadly, our work offers a computational account of how dorsal–ventral organization can support both feature sharing and specialization when learning behaviors with divergent demands.

LLM Disclosure

Anthropic's Claude was used for condensing text and formatting references and text in LaTeX. All scientific content is the authors' own.

References

- Blender Online Community. (2023). *Blender—a 3D modelling and rendering package*. <https://www.blender.org>
- Choi, M., Han, K., Wang, X., Zhang, Y., & Liu, Z. (2023). A dual-stream neural network explains the functional segregation of dorsal and ventral visual pathways in human brains. *Advances in Neural Information Processing Systems (NeurIPS)*, 36, 50408–50428.
- Culham, J. C., Danckert, S. L., DeSouza, J. F. X., Gati, J. S., Menon, R. S., & Goodale, M. A. (2003). Visually guided grasping produces fMRI activation in dorsal but not ventral stream brain areas. *Experimental Brain Research*, 153(2), 180–189. <https://doi.org/10.1007/s00221-003-1591-5>.
- DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron*, 73(3), 415–434. <https://doi.org/10.1016/j.neuron.2012.01.010>.
- Freud, E., Culham, J. C., Plaut, D. C., & Behrmann, M. (2017). The large-scale organization of shape processing in the ventral and dorsal pathways. *eLife*, 6, e27576. <https://doi.org/10.7554/eLife.27576>.
- Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, 15(1), 20–25. [https://doi.org/10.1016/0166-2236\(92\)90344-8](https://doi.org/10.1016/0166-2236(92)90344-8).
- Grill-Spector, K., & Malach, R. (2004). The human visual cortex. *Annual Review of Neuroscience*, 27(1), 649–677. <https://doi.org/10.1146/annurev.neuro.27.070203.144220>.
- Klein, L. K., Maiello, G., Paulun, V. C., & Fleming, R. W. (2020). Predicting precision grip grasp locations on three-dimensional objects. *PLOS Computational Biology*, 16(8), e1008081. <https://doi.org/10.1371/journal.pcbi.1008081>.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems (NeurIPS)*, 25, 1097–1105.
- Meshy AI. (2023). *Meshy: Generative 3D AI platform*. <https://www.meshy.ai>
- Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: Two cortical pathways. *Trends in Neurosciences*, 6, 414–417. [https://doi.org/10.1016/0166-2236\(83\)90190-X](https://doi.org/10.1016/0166-2236(83)90190-X).
- Murata, A., Gallese, V., Luppino, G., Kaseda, M., & Sakata, H. (2000). Selectivity for the shape, size, and orientation of objects for grasping in neurons of monkey parietal area AIP. *Journal of Neurophysiology*, 83(5), 2580–2601. <https://doi.org/10.1152/jn.2000.83.5.2580>.
- Shmuelof, L., & Zohary, E. (2005). Dissociation between ventral and dorsal fMRI activation during object and action recognition. *Neuron*, 47(3), 457–470. <https://doi.org/10.1016/j.neuron.2005.06.034>.
- Ungerleider, L. G., & Haxby, J. V. (1994). 'what' and 'where' in the human brain. *Current Opinion in Neurobiology*, 4(2), 157–165. [https://doi.org/10.1016/0959-4388\(94\)90066-3](https://doi.org/10.1016/0959-4388(94)90066-3).
- Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of visual behavior*. MIT Press.