

Appendix

| | |
|---|---|
| A Note On Terminology | 1 |
| B Representative Samples & Aggregations | 1 |
| C T-NID Algorithm | 2 |
| D Test Suite Of Synthetic Functions | 2 |
| E Additional Architectures For N-Way Interactions | 2 |
| F. COCO Multi-Object Detection | 2 |
| G Grad-CAM On Relational Reasoning | 2 |
| H Interactional Relation Network (IRN) | 2 |
| I. Taylor-CAM Pipeline | 3 |
| J. Biomedical Analysis | 3 |
| K Architecture Configurations | 9 |

A. Note On Terminology

In colloquial terms, two things are said to interact when they depend on each other in some way. Similar to [4], this can be formalized as follows:

Definition A.1. Entity Interaction Given an entity e_1 with attributes (a_1, \dots, a_n) , an interaction exists with another entity e_2 with attributes (b_1, \dots, b_n) if some a_i depends on some b_j or some b_j depends on some a_i .

Now we will define mathematical relation.

Definition A.2. Relation Given sets A and B , the binary relation from A to B is a subset of the Cartesian product $A \times B$.

We would like to unify our colloquial understanding of interaction in Definition A.1, our mathematical definition of relation in Definition A.2, and our definition for statistical interaction effects in Definition 1 of the main paper.

To connect this to Definition 1, we will reframe features as entities with the following theorem:

Theorem 1. *Given a function $F(\mathbf{x})$ and feature x_i , let entity e_i consist of attributes $(x_i, F/\partial x_i)$. An interaction exists between e_1 and e_2 if there is a nonzero interaction effect between x_1 and x_2 .*

Proof. If there is a nonzero interaction effect between x_1 and x_2 , then $\partial^2 F(\mathbf{x})/(\partial x_1 \partial x_2) \neq 0$ for some input \mathbf{x} . Then $F/\partial x_1$ depends on x_2 and consequently, there exists an interaction between entities e_1 and e_2 . \square

We have shown that our statistical interaction implies an interaction according to our colloquial understanding. An interaction exists between e_1 and e_2 if (but not only if, since the change need not be local) $F(\mathbf{x})/(\partial x_1 \partial x_2) \neq 0$, meaning $F/\partial x_1$ depends on x_2 . This is considered a binary relation between the two attributes, as all functions are relations, though not all relations are functions. Formally: given a function $F(\mathbf{x})$, a feature x_i , and entity e_i consisting of attributes $(x_i, F/\partial x_i)$, if there is a nonzero interaction effect between x_1 and x_2 , then a relation exists between the attributes of the two entities.

We have shown that, under this framing, an interaction effect is a relation, and if the interaction effect is nonzero, there must be a dependency/interaction between those entities. Since feature vectors in CNNs could be treated as entities [15, 19, 14], and if one interprets their gradients on the output to be implicit attributes, computing interaction effects between CNN feature vectors is equivalent to identifying the colloquial interactions and relations described in this formulation.

This is trivially generalized to interactions/relations of higher orders.

To summarize, a mathematical relation is implied by a colloquial interaction is implied by a statistical interaction, and this hierarchy can be formalized by regarding a feature x_i as an entity whose attributes include its gradients with respect to the function of interest. Thus, we offer a simple, formal connection between our statistical interaction effects definition and mathematical relations, as well as an integration of both into the colloquial understanding of “interaction” as merely a dependency between two “things.”

B. Representative Samples & Aggregations

| Aggregation Of Representative Samples | AUC Score |
|---------------------------------------|-----------|
| Mean Of Mean-Min-Mode-Rand | 0.61825 |
| Mean Of Med-Min-Mode-Rand | 0.61825 |
| Mean Of Mean-Med-Min-Mode-Rand | 0.61775 |
| Mean Of Mean-Min-Max-Mode-Rand | 0.6155 |
| Mean Of Med-Min-Max-Mode-Rand | 0.6155 |
| Med Of Mean-Min-Mode-Rand | 0.61525 |
| Med Of Med-Min-Mode-Rand | 0.61525 |
| Mean Of Mean-Med-Min-Max-Mode-Rand | 0.61525 |
| Mean Of Mean-Min-Rand | 0.614 |
| Mean Of Med-Min-Rand | 0.614 |

Table 1: Top average (across all orders) AUC scores for different aggregations of representative samples

Table 1 displays the top 10 aggregations and representative samples discovered via our power sweep.

C. T-NID Algorithm

Our complete T-NID is described in Algorithm 1. Note that each derivation of interaction effect using Definition 1 of the main paper for an interaction $I = \hat{I} \cup j$ of size $\hat{\ell}$ where $|\hat{I}| = \hat{\ell} - 1$ for sample x can be derived as a single-order partial derivative $\partial IE_{\hat{I}} / \partial x_j$ and does not need to be recomputed from the ground up.

Algorithm 1 T-NID algorithm in pseudocode

Inputs ℓ -times differentiable trained neural network F , dataset X with i th sample features X_{i1}, \dots, X_{in} , order ℓ , orders without subsampling o , subsampling size k .

Outputs Interaction effects IE_I for top estimated interactions $I \subseteq \{1, \dots, n\}$, where $|I| \leq \ell$.

Get representative samples:

For j th aggregation \in mean, minimum, mode, random
 $c = \operatorname{argmin}_i \|X_i - \text{aggregation}(X, \text{axis} = 0)\|$
 $r_j = X_c$

For each representative sample:

For $r_j \in r$

Compute all non-redundant partial derivatives up to order o :

For $I \subseteq \{1, \dots, n\}$, where $|I| \leq o$

$I = \operatorname{sort}(I)$

If $IE_I^{(j)}$ uninitiated

Initiate $IE_I^{(j)}$ according to Definition 1 of the main paper

Compute remaining partial derivatives up to order ℓ by subsampling top k from previous orders:

For $\hat{\ell} \in o + 1, \dots, \ell$

For $\hat{I} \in$ top k argmax of $IE_I^{(j)}$, where $|\hat{I}| = \hat{\ell} - 1$

For $I \subseteq \{1, \dots, n\}$, where $|I| = \ell$ and $\hat{I} \subset I$

If $IE_I^{(j)}$ uninitiated

Initiate $IE_I^{(j)}$ according to Definition 1 of the main paper

Take the mean interaction effects across representative samples:

For $I \subseteq \{1, \dots, n\}$ if $IE_I^{(j)}$ initiated for some j

$IE_I = \operatorname{mean}(IE_I^{(j)})$ for all j where $IE_I^{(j)}$ initiated

Return IE

D. Test Suite Of Synthetic Functions

The test-suite of synthetic functions used to evaluate T-NID may be found in Table 2, courtesy of [17].

E. Additional Architectures For N -Way Interactions

Table 3 shows results for T-NID + MLP-M (T-NID using a neural network equipped with a main effects network as well as trained with sparsity regularization) and NID + MLP-M, the architecture used in [17].

F. COCO Multi-Object Detection

The task is to identify whether a pair of objects are each present in tandem. If only one is present, then the class label is negative. We tested this on the objects ‘‘car’’ and ‘‘person’’ in the COCO annotated-image dataset. We configured the frequency of the labels such that an even amount of positive and negative samples were in the training set. We found the COCO task to be somewhat inconclusive, because of model overfitting and rather low test accuracy, but still observed reasonable explanations, as seen in Figure 2. Taylor-CAM often prioritizes the car-person interaction correctly.

G. Grad-CAM On Relational Reasoning

Grad-CAM is a first-order explanatory tool that ranks different image regions and produces a heatmap of saliences. As shown qualitatively in Figure 3, Grad-CAM’s heatmaps are much harder to interpret and to reverse engineer questions and objects from compared to the results obtained from Taylor-CAM, shown in Figure 3 of the main paper, as corroborated quantitatively with our human study.

H. Interactional Relation Network (IRN)

A standard RN pools a set of feature vectors $O = \{o_1, \dots, o_n\}$, their corresponding positional encodings $C = \{c_1, \dots, c_n\}$, and a question q as follows:

$$\operatorname{RN}(O, C, q) = f_\phi \left(\sum_{i,j} g_\theta(o_i, o_j, c_i, c_j, q) \right), \quad (1)$$

where f and g are modeled by neural networks parameterized by ϕ and θ respectively.

We observed through Taylor-CAM that many of the top interactions in the RN’s reasoning were between individual regions and themselves, even when we zeroed out diagonals, such as in Figure ?? of the main paper. We found that we could mitigate this by making a simple modification to the RN architecture which we found to yield better test accuracy:

$$\operatorname{IRN}(O, C, q) =$$

$$f_\phi \left(\sum_{i,j} g_\theta(h_\psi(o_i, c_i, q), h_\psi(o_j, c_j, q), c_i, c_j, q) \right), \quad (2)$$

| | |
|----------------------|---|
| $F_1(\mathbf{x})$ | $\pi^{x_1 x_2} \sqrt{2x_3} - \sin^{-1}(x_4) + \log(x_3 + x_5) - \frac{x_9}{x_{10}} \sqrt{\frac{x_7}{x_8}} - x_2 x_7$ |
| $F_2(\mathbf{x})$ | $\pi^{x_1 x_2} \sqrt{2 x_3 } - \sin^{-1}(0.5x_4) + \log(x_3 + x_5 + 1) + \frac{x_9}{1+ x_{10} } \sqrt{\frac{ x_7 }{1+ x_8 }} - x_2 x_7$ |
| $F_3(\mathbf{x})$ | $\exp x_1 - x_2 + x_2 x_3 - x_3^{\frac{2 x_4 }{x_3}} + \log(x_4^2 + x_5^2 + x_7^2 + x_8^2) + x_9 + \frac{1}{1+x_{10}^2}$ |
| $F_4(\mathbf{x})$ | $\exp x_1 - x_2 + x_2 x_3 - x_3^{\frac{2 x_4 }{x_3}} + (x_1 x_4)^2 + \log(x_4^2 + x_5^2 + x_7^2 + x_8^2) + x_9 + \frac{1}{1+x_{10}^2}$ |
| $F_5(\mathbf{x})$ | $\frac{1}{1+x_1^2+x_2^2+x_3^2} + \sqrt{ x_4 + x_5 + x_6 + x_7 + x_8 x_9 x_{10}}$ |
| $F_6(\mathbf{x})$ | $\exp(x_1 x_2 + 1) - \exp(x_3 + x_4 + 1) + \cos(x_5 + x_6 - x_8) + \sqrt{x_8^2 + x_9^2 + x_{10}^2}$ |
| $F_7(\mathbf{x})$ | $(\arctan(x_1) + \arctan(x_2))^2 + \max(x_3 x_4 + x_6, 0) - \frac{1}{1+(x_4 x_5 x_6 x_7 x_8)^2} + \left(\frac{ x_7 }{1+ x_9 }\right)^5 + \sum_{i=1}^{10} x_i$ |
| $F_8(\mathbf{x})$ | $x_1 x_2 + 2^{x_3+x_5+x_6} + 2^{x_3+x_4+x_5+x_7} + \sin(x_7 \sin(x_8 + x_9)) + \arccos(0.9x_{10})$ |
| $F_9(\mathbf{x})$ | $\tanh(x_1 x_2 + x_3 x_4) \sqrt{ x_5 } + \exp(x_5 + x_6) + \log((x_6 x_7 x_8)^2 + 1) + x_9 x_{10} + \frac{1}{1+ x_{10} }$ |
| $F_{10}(\mathbf{x})$ | $\sinh(x_1 + x_2) + \arccos(\tanh(x_3 + x_5 + x_7)) + \cos(x_4 + x_5) + \sec(x_7 x_9)$ |

Table 2: Synthetic test-suite functions

| | T-NID 3-Way | NID 3-Way | T-NID 4-Way | NID 4-Way | T-NID 5-Way | NID 5-Way |
|----------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| $F_1(\mathbf{x})$ | 0.831 ± 0.064 | 0.122 ± 0.028 | 0.777 ± 0.389 | 0.555 ± 0.456 | N/A | N/A |
| $F_2(\mathbf{x})$ | 0.629 ± 0.165 | 0.07 ± 0.011 | 0.032 ± 0.065 | 0.185 ± 0.37 | N/A | N/A |
| $F_3(\mathbf{x})$ | 0.991 ± 0.013 | 0.095 ± 0.008 | 0.997 ± 0.006 | 1.0 ± 0.0 | N/A | N/A |
| $F_4(\mathbf{x})$ | 0.993 ± 0.007 | 0.09 ± 0.028 | 0.96 ± 0.08 | 0.996 ± 0.009 | N/A | N/A |
| $F_5(\mathbf{x})$ | 0.493 ± 0.009 | 0.035 ± 0.005 | N/A | N/A | N/A | N/A |
| $F_6(\mathbf{x})$ | 0.103 ± 0.025 | 0.034 ± 0.005 | N/A | N/A | N/A | N/A |
| $F_7(\mathbf{x})$ | 0.417 ± 0.264 | 0.156 ± 0.031 | 0.363 ± 0.322 | 0.711 ± 0.046 | 0.303 ± 0.367 | 0.536 ± 0.453 |
| $F_8(\mathbf{x})$ | 1.0 ± 0.0 | 0.141 ± 0.008 | 1.0 ± 0.0 | 0.994 ± 0.008 | N/A | N/A |
| $F_9(\mathbf{x})$ | 0.838 ± 0.146 | 0.113 ± 0.01 | 0.859 ± 0.068 | 0.618 ± 0.084 | 0.988 ± 0.024 | 0.549 ± 0.452 |
| $F_{10}(\mathbf{x})$ | 1.0 ± 0.0 | 0.03 ± 0.002 | N/A | N/A | N/A | N/A |
| average | 0.73 ± 0.069 | 0.089 ± 0.014 | 0.713 ± 0.133 | 0.723 ± 0.139 | 0.646 ± 0.200 | 0.543 ± 0.453 |

Table 3: N-Way AUC scores for T-NID + MLP-M and NID + MLP-M, both using a main effects network and sparsity regularization, as described in [17]

where h is an MLP parameterized by ψ .

We refer to this as Interactional Relation Network (IRN) since it explicitly separates within its architecture the concerns of reasoning about interactions from reasoning about individual objects. While IRN does not relate to Taylor-CAM directly, it does highlight how visualizing a network’s relational reasoning can inspire potential ideas for improvement to a network’s architecture.

I. Taylor-CAM Pipeline

In Figure 1, we illustrate the full pipeline of Taylor-CAM. A model consisting of a CNN and Relation Network predicts answers based on images and questions. Taylor-CAM intercepts the model’s gradients, reverse engineers the question asked, and visualizes for a human observer.

Given three possible question categories (*closest*, *furthest*, and *same shape*), the user is able to interpret which the model is reasoning about by looking at Taylor-CAM’s proposed interactions, as shown quantitatively in Table ?? of the main paper.

J. Biomedical Analysis

We applied these techniques to the Parkinson’s Progression Marker Initiative (PPMI) study (<http://www.ppmi-info.org/>) dataset, which follows persons living with early-stage Parkinson’s disease for up to approximately eight years collecting clinical and biological data from participants. Parkinson’s disease (PD) is a neurodegenerative progressive disease, characterized clinically by motor (*e.g.*, tremor, rigidity) and non-motor (*e.g.*, cognition and autonomic dysfunction) symptoms that vary over time within and between patients. Progression of motor and non-motor symptoms are likely not independent of each other. Instead, collateral damage may be inflicted multilaterally with non-motor and motor pathological features progressing interdependently. As an example, depressive symptoms in Parkinson’s disease are common and may perpetuate motor and cognitive deficits, which could impact function, and ultimately diminish quality of life. Therefore, it is necessary to take as comprehensive of an approach as possible in unraveling the clinical progression of Parkinson’s disease.

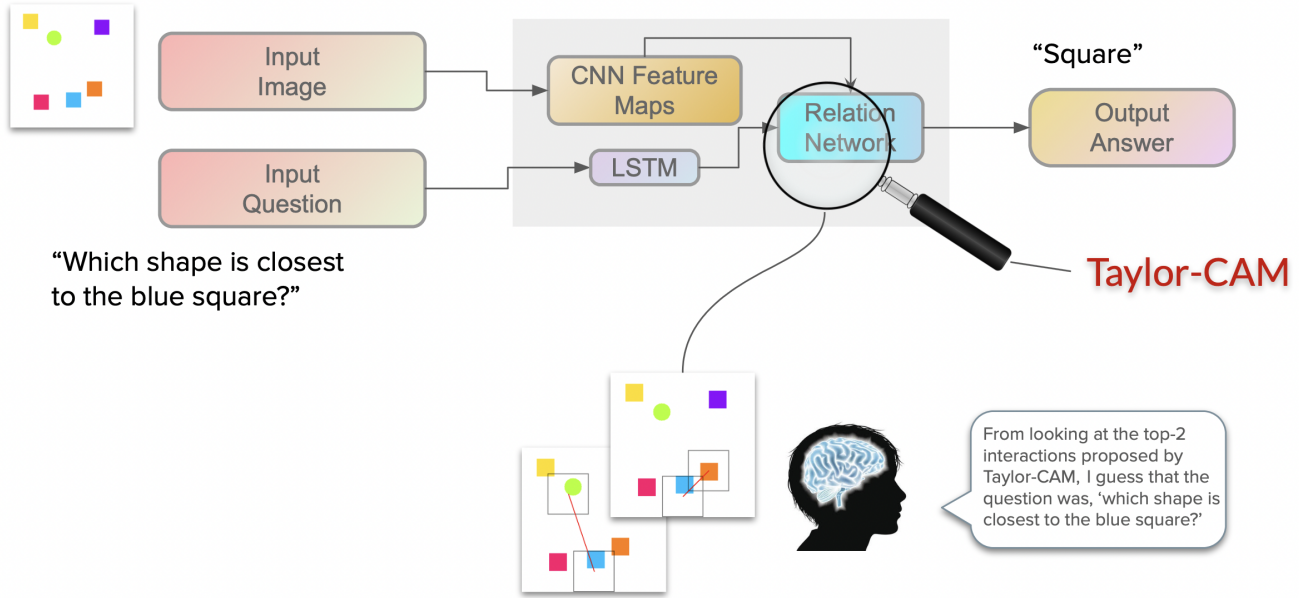


Figure 1: Full Taylor-CAM pipeline on the problem of relational visual question answering. A CNN + RN take in an image and question, and output an answer. Taylor-CAM is able to visualize the model’s reasoning from just its gradients such that a human can interpret what and how the model reasoned.

As PD progresses, cognitive impairment leading to dementia may affect up to 80% of patients, ultimately impairing one’s functional independence. Within the PPMI study, we tested 2-, 3-, 4- and 5-way interactions to understand multivariable features at baseline that distinguish patients with a more severe progression in decline of cognitive function (“fast progressors”) compared to those with a more benign course of cognitive changes, as measured by the Montreal Cognitive Assessment (MoCA) scale. Top 2-way interaction effects identified (Figure 4) among “fast progressors” included feature interactions between handedness (handed) and severity of rigidity in the neck (np3rign); presence of resting tremor at disease diagnosis (dxtremor) and severity of rigidity in the lower extremities (np3rigll); and, severity of tremor (np2trmr) and alternating trail making test from the MoCA scale (mcaalttm) – which ultimately is a measure of processing speed, mental flexibility, ability to sequence, and visuo-motor skills. Each of these features individually have some established associations with cognitive dysfunction or neuropsychological disorders; however, their interactions together have not been previously considered. For example, handedness, has been significantly associated with functional connectivity between language networks, as well as specific genetic loci implicated in the pathogenesis of neurologic disorders including Parkinson’s disease [18]. More severe rigidity symptoms in Parkinson’s disease are also associated with faster cognitive decline [13]. Our analysis, for the first time, suggests that measures of both handedness and rigidity severity together

are important to consider when predicting faster cognitive progression in Parkinson’s disease. As shown in Figure 4, we provide 3-, 4-, and 5-way interactions between features.

When broadening the interactions to 3-, 4-, and 5-way interactions between features that predict fast cognitive decline, we observe additional features with some consistency (Figure 5 provides 3-, 4-, and 5-way interactions between features). Broadly speaking, some of the most important interactions occurred between symptoms of autonomic dysfunction: urinary (np1urin, scau11, scau13) and constipation issues (np1cnst, scau5), problems tolerating cold/heat [scau20]); mood and sleep disturbances: depression (np1dprs), apathy (np1apat), and restless sleep (np1slpn, slplmbmv); postural instability and balance issues (np2walk, np3pstbl); overall severity of Parkinson’s disease (nhy); and, memory impairment: delayed recall (mcarec2, mcarec4). Each of these symptoms, singularly, have been thought to be associated with cognitive impairment [11, 8, 16, 9]. It is novel, yet biologically plausible to consider these symptoms interacting, as the neuropathology underlying Parkinson’s disease involves multiple areas of the brain and nervous system beyond the nigrostriatal dopamine pathway. For instance, Lewy Body pathology affects the limbic cortex and frontal neocortical areas, sympathetic ganglia and even the peripheral autonomic nervous system including the myenteric plexus [3].

We also performed our analysis to predict fast progression of ambulatory impairment which stems from worsening progression of motor symptoms and is a major source



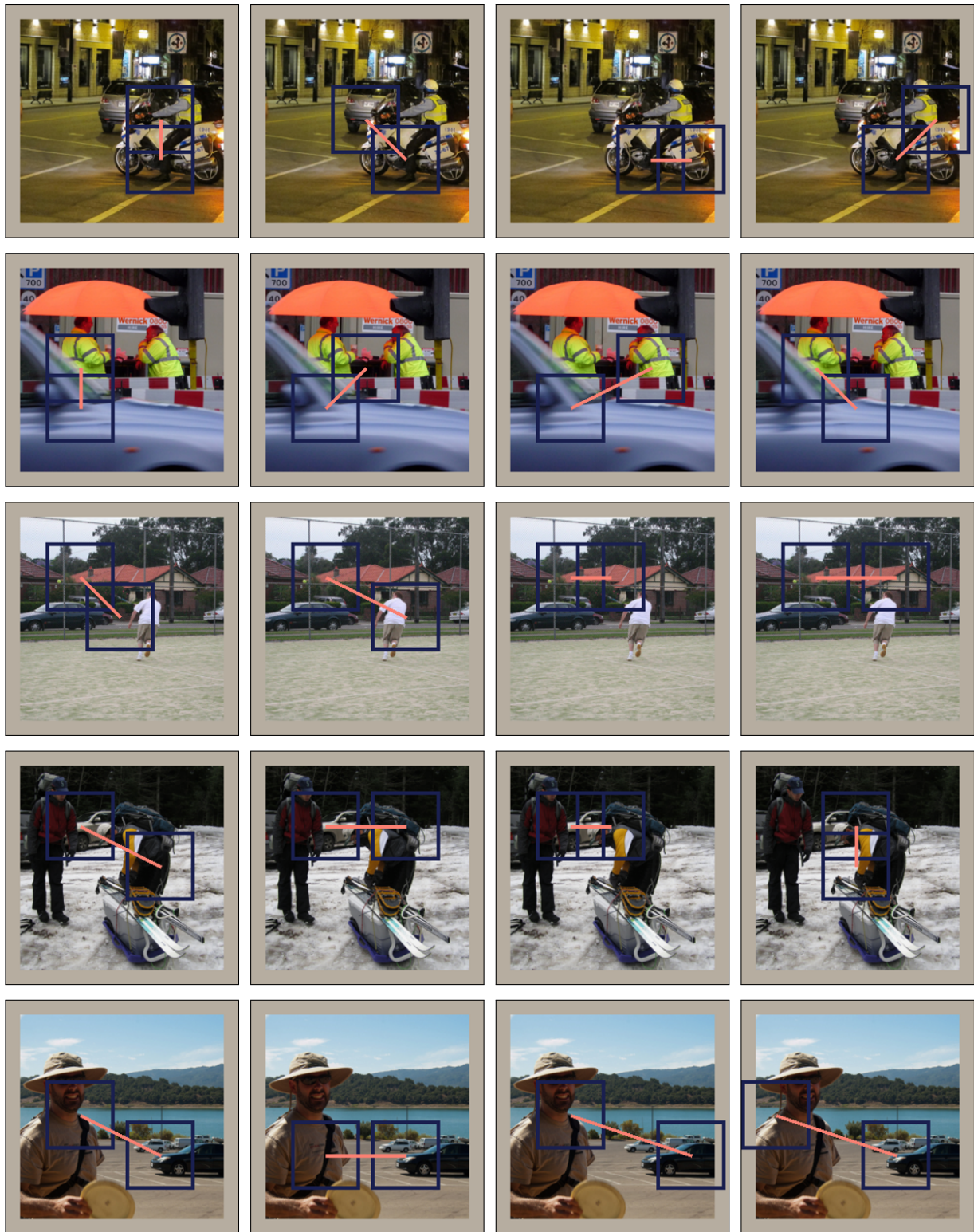
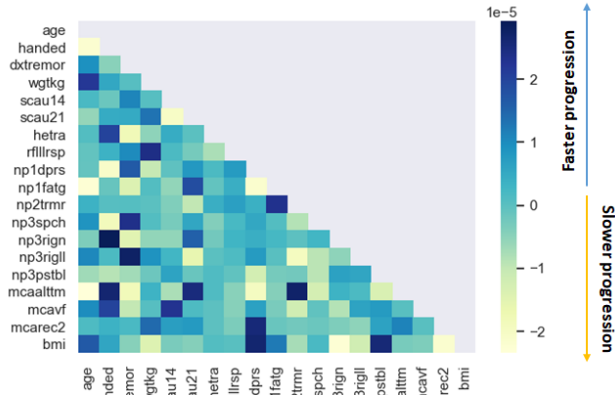


Figure 2: Taylor-CAM interacts “car” and “person” in the custom COCO multi-object detection task. Taylor-CAM’s top 4 discovered interactions per image are shown. As discussed in the main paper, in cases where the interaction is not present (as in 4th example), Taylor-CAM interacts immediately adjacent regions around whichever of the two objects is present.



Figure 3: Grad-CAM's saliency heatmaps per image/question. *Left*, the raw image. *Right*, Grad-CAM's explanatory heatmap.

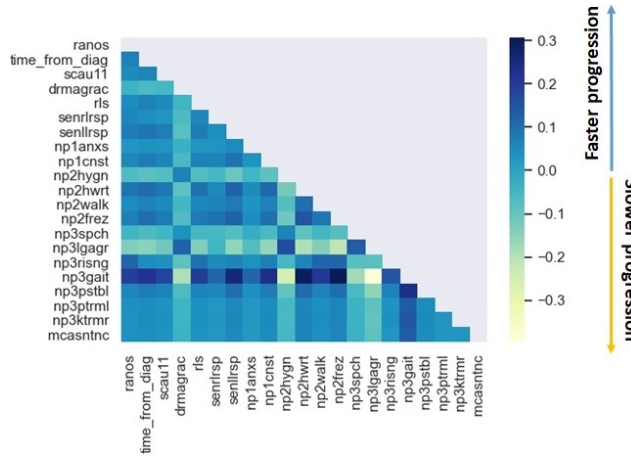


a) Top 2-way interactions for MoCA fast progression

| <i>N</i> -Way Interaction | Strength |
|--|----------|
| id_num, scau20, mcarec4 | 4.77E-06 |
| id_num, drmagrac, mcarec4 | 4.69E-06 |
| educyrs, np1apat, bmi | 4.56E-06 |
| np1dprs, np2walk, np3pstbl | 4.27E-06 |
| scau13, np1slpn, np1cnst, nhy | 6.00E-07 |
| scau11, scau13, scau20, bmi | 5.66E-07 |
| scau11, scau13, np1slpn, nhy | 5.64E-07 |
| scau13, scau20, np1urin, nhy | 5.43E-07 |
| slplmbmv, np1dprs, np2walk, np3rigru, np3pstbl | 1.23E-07 |
| slplmbmv, np1dprs, np2walk, np3rign, np3pstbl | 1.22E-07 |
| scau5, np1dprs, np2walk, np3rigru, np3pstbl | 1.19E-07 |
| slplmbmv, np1dprs, np2walk, np3pstbl, mcarec2 | 1.18E-07 |

b) Top 3-way, 4-way, and 5-way interactions for MoCA fast progression

Figure 4: Interaction effects for classifying fast clinical progression of MoCA scores from baseline



a) Top 2-way interactions for uMCA fast progression

| <i>N</i> -Way Interaction | Strength |
|--|----------|
| scau1, np3lgagr, np3risng | 1.97E+00 |
| scau1, scau9, np3lgagr | 1.41E+00 |
| scau1, np2hwrt, np3lgagr | 1.31E+00 |
| scau1, senllrsp, np3lgagr | 1.19E+00 |
| time_from_diag, scau1, np2frez, np3lgagr | 5.39E+00 |
| scau1, np2walk, np2frez, np3lgagr | 5.28E+00 |
| ranos, scau1, np2frez, np3lgagr | 5.25E+00 |
| scau1, rls, np2frez, np3lgagr | 5.21E+00 |
| scau1, np3lgagr, np3risng, np3gait, np3rtarl | 1.93E+01 |
| scau1, np2hygn, np3lgagr, np3risng, np3gait | 1.80E+01 |
| scau1, mslarsp, np3lgagr, np3risng, np3gait | 1.68E+01 |
| dxrigid, scau1, np3lgagr, np3risng, np3gait | 1.47E+01 |

b) Top 3-way, 4-way, and 5-way interactions for uMCA fast progression

Figure 5: Interaction effects for classifying fast clinical progression of uMCA scores from baseline

of disability for patients with Parkinson's disease. Severity of ambulatory impairment was measured by an ambulatory capacity score derived from sum of scores of the MDS-UPDRS items 2.13 (freezing), 2.12 (walking and balance), 3.10 (gait), 3.12 (postural stability), and 3.11 (freezing of gait). The top 2-way interaction among "fast progressors" of ambulatory capacity was between severity of freezing (np2frez) and gait (np3gait), which is unsurprising as both are components of the ambulatory capacity scale score. Interestingly, however, the next top 2-way interactions were between handwriting (np2hwrt) and gait (np3gait); and, sensory of legs (senllrsp) and gait (np3gait). Worsening of handwriting is often reported as an initial symptom of Parkinson's disease and is reported to be more problematic

in people with Parkinson's disease who experience freezing of gait [12, 6]. Periphery sensory defects in the lower limbs are also commonly noted in people with Parkinson's disease, and could be a main contributor to balance control issues and postural instability [7]. As interactions were expanded to 3-, 4-, and 5-ways, other items that were consistently identified were difficulty in swallowing and chewing (scau1), and leg agility (np3lgagr). While these items are not often considered as obvious predictors of ambulatory capacity by themselves, their interactions with some more apparent features (e.g., gait, freezing, arising from chair) provide new insights on how symptoms in Parkinson's disease patients contribute to disease progression.

K. Architecture Configurations

T-NID For T-NID, we trained a GELU-activated multi-layer perceptron with hidden layer sizes 140, 100, 60, and 20 for 200 epochs with a learning rate of 0.003 using early stopping [1] with a patience of 10. Results were averaged across 10 trials. Input data was normalized by standard deviation. T-NID hyperparameters were set as $\ell = 5, o = 2, k = 10$.

Taylor-CAM For our COCO [10] task, we used Pytorch’s ResNet-50 [5] pretrained on ImageNet [2], except we replaced the global average pooling layer with an additional convolutional layer composed of 1024 out-channels, size 2 kernel, 2 stride, and 2 padding, followed by 3 hidden linear layers of size 512, 256, 64, because global average pooling yields no higher-order derivatives. For our Relation Network, we used an open source reference implementation, which can be found here: <https://github.com/kimhc6028/relational-networks>, since [15] did not release code to the public. We trained for 50 epochs.

References

- [1] Rich Caruana, Steve Lawrence, and C Lee Giles. Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping. In *Advances in neural information processing systems*, pages 402–408, 2001. 9
- [2] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 9
- [3] Dennis W Dickson. Parkinson’s disease and parkinsonism: neuropathology. *Cold Spring Harbor perspectives in medicine*, 2(8):a009258, 2012. 4
- [4] Jerome H Friedman, Bogdan E Popescu, et al. Predictive learning via rule ensembles. *The Annals of Applied Statistics*, 2(3):916–954, 2008. 1
- [5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 9
- [6] Elke Heremans, Evelien Nackaerts, Sanne Broeder, Griet Vervoort, Stephan P Swinnen, and Alice Nieuwboer. Handwriting impairments in people with parkinson’s disease and freezing of gait. *Neurorehabilitation and neural repair*, 30(10):911–919, 2016. 8
- [7] Sungjae Hwang, Peter Agada, Stephen Grill, Tim Kiemel, and John J Jeka. A central processing sensory deficit with parkinson’s disease. *Experimental brain research*, 234(8):2369–2379, 2016. 8
- [8] Natalia Jozwiak, Ronald B Postuma, Jacques Montplaisir, Véronique Latreille, Michel Panisset, Sylvain Chouinard, Pierre-Alexandre Bourgouin, and Jean-François Gagnon. Rem sleep behavior disorder and cognitive impairment in parkinson’s disease. *Sleep*, 40(8), 2017. 4
- [9] VE Kelly, CO Johnson, EL McGough, A Shumway-Cook, FB Horak, KA Chung, AJ Espay, FJ Revilla, J Devoto, C Wood-Siverio, et al. Association of cognitive domains with postural instability/gait disturbance in parkinson’s disease. *Parkinsonism & related disorders*, 21(7):692–697, 2015. 4
- [10] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014. 9
- [11] Werner Poewe. Dysautonomia and cognitive dysfunction in parkinson’s disease. *Movement disorders: official journal of the Movement Disorder Society*, 22(S17):S374–S378, 2007. 4
- [12] Mirthe M Ponsen, Andreas Daffertshofer, Erik Ch Wolters, Peter J Beek, and Henk W Berendse. Impairment of complex upper limb motor function in de novo parkinson’s disease. *Parkinsonism & Related Disorders*, 14(3):199–204, 2008. 8
- [13] AH Rajput, A Voll, ML Rajput, CA Robinson, and A Rajput. Course in parkinson disease subtypes: a 39-year clinicopathologic study. *Neurology*, 73(3):206–212, 2009. 4
- [14] Adam Santoro, Ryan Faulkner, David Raposo, Jack Rae, Mike Chrzanowski, Theophane Weber, Daan Wierstra, Oriol Vinyals, Razvan Pascanu, and Timothy Lillicrap. Relational recurrent neural networks. In *Advances in neural information processing systems*, pages 7299–7310, 2018. 1
- [15] Adam Santoro, David Raposo, David G Barrett, Mateusz Malinowski, Razvan Pascanu, Peter Battaglia, and Timothy Lillicrap. A simple neural network module for relational reasoning. In *Advances in neural information processing systems*, pages 4967–4976, 2017. 1, 9
- [16] AI Troster, AM Paolo, KE Lyons, SL Glatt, JP Hubble, and WC Koller. The influence of depression on cognition in parkinson’s disease: a pattern of impairment distinguishable from alzheimer’s disease. *Neurology*, 45(4):672–676, 1995. 4
- [17] Michael Tsang, Dehua Cheng, and Yan Liu. Detecting statistical interactions from neural network weights. In *International Conference on Learning Representations*, 2018. 2, 3
- [18] Akira Wiberg, Michael Ng, Yasser Al Omran, Fidel Alfaró-Almagro, Paul McCarthy, Jonathan Marchini, David L Bennett, Stephen Smith, Gwenaelle Douaud, and Dominic Furniss. Handedness, language areas and neuropsychiatric diseases: insights from brain imaging and genetics. *Brain*, 142(10):2938–2947, 2019. 4
- [19] Vinicius Zambaldi, David Raposo, Adam Santoro, Victor Bapst, Yujia Li, Igor Babuschkin, Karl Tuyls, David Reichert, Timothy Lillicrap, Edward Lockhart, Murray Shanahan, Victoria Langston, Razvan Pascanu, Matthew Botvinick, Oriol Vinyals, and Peter Battaglia. Deep reinforcement learning with relational inductive biases. In *International Conference on Learning Representations*, 2019. 1