Deep predictive coding networks partly capture neural signatures of short-term temporal adaptation in human visual cortex

Amber M. Brands
Informatics Institute, University of Amsterdam, Amsterdam, The Netherlands

Paulo Ortiz
Informatics Institute, University of Amsterdam, Amsterdam, The Netherlands

Iris I. A. Groen (i.i.a.groen@uva.nl)
Informatics Institute, University of Amsterdam, Amsterdam, The Netherlands
Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands

Abstract

Predictive coding is a prominent theory of cortical function that proposes that the brain continuously generates predictions about sensory inputs through a hierarchical network of top-down and bottom-up connections. Prior studies demonstrate that PredNet, a deep neural network built on predictive coding principles, indeed captures key characteristics of neural responses observed in primate visual cortex. However, one widespread neural phenomenon that remains unexplored in this context is short-term visual adaptation: the modulation of neural activity over time in response to static visual inputs that are prolonged or repeated. Here, we investigate whether PredNet exhibits two hallmark signatures of temporal adaptation previously identified in human intracranial recordings. We find that, similar to human visual cortex, activations of error units in the first layer of PredNet exhibit subadditive temporal summation to prolonged stimuli, reflecting nonlinear accumulation of response magnitude with increased stimulus duration. However, unlike the neural data, PredNet shows systematic responses to stimulus offsets. For repeated stimuli, PredNet exhibits slight response suppression for consecutively presented images, but no repetition suppression, a stronger response reduction to identical than non-identical images that is robustly observed in visual cortex. These discrepancies are consistent across different training diets, optimization strategies and model unit types. Overall, our results show that PredNet's activation dynamics only partly capture short-term temporal adaptation signatures in human visual cortex, suggesting that this particular instantiation of predictive coding does not fully account for neural adaptation phenomena.

Keywords: predictive coding; temporal adaptation; subadditive temporal summation; response suppression; repetition suppression; deep neural network; human iEEG

Introduction

Predictive coding is a prominent theory of brain function and sensory information processing (Rao & Ballard, 1999; Friston, 2005), which posits that neural circuits learn and predict representations that reflect the statistical regularities of the natural world, signaling deviations from such regularities to higher processing centers. Rao & Ballard (1999) proposed and implemented a hierarchical architecture for predictive coding, referred to as the *predictive coding scheme*, that explains certain important properties of the visual cortex; this in turn inspired several subsequent works explaining various perceptual and neurophysiological phenomena (Hohwy et al. 2008; Spratling 2008; Summerfield & Egner 2009; Auksztulewicz & Friston 2016), while also providing biologically plausible neural dynamics and synaptic update rules (Friston, 2003; Lillicrap et al., 2020; Millidge et al., 2020).

Previous studies have also casted the predictive coding scheme by Rao & Ballard (1999) into a modern deep learning framework. An implementation that has been intrinsically designed according to predictive coding theory is known as PredNet (Lotter et al., 2016, 2020). This deep neural network (DNN) predicts future visual inputs, whereby each layer in the network makes local predictions in a top-down fashion and computes prediction errors by comparing those predictions to the input from the layer below. These prediction errors are then in turn fed to subsequent upper network layers, whereby the network learns in a recursive way to construct and update an internal model of its environment. In addition to implementing a predictive coding scheme, this deep learning architecture learns representations from video, allowing exploitation of the temporal structure in naturalistic visual environments of biological organisms (including humans). Together, these factors have given rise to the hypothesis that response dynamics in PredNet should align with those observed in visual cortex.

In support of this hypothesis, previous work has demonstrated PredNet's ability to capture several phenomena related to temporal dynamics observed in visual cortex, including on/off responses, sequence learning effects and perceptual motion illusions (Watanabe et al., 2018; Fonseca, 2019; Lotter et al., 2020; Kirubeswaran & Storrs, 2023). However, one prominent and ubiquitous property of neural responses that has not yet been studied in detail is short-term visual adaptation, which exhibits interesting and complex temporal dynamics, as reported in a series of recent studies on intracranial EEG (iEEG) recordings in human visual cortex (Zhou et al., 2019; Groen et al., 2022; Brands et al., 2024). First, neu-

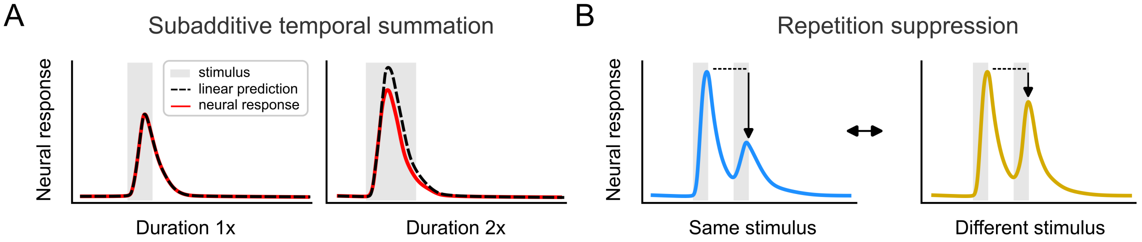

Figure 1: Neural population responses show adaptation over time. A: Subadditive temporal summation for prolonged stimuli with varying duration. B: Response suppression for repeated stimuli with a stronger reduction when the two images are identical, compared to when they are different (repetition suppression).

ral responses show subadditive temporal summation, a nonlinear accumulation of response magnitudes when a static visual stimulus is prolonged in time (**Fig. 1A**). Second, neural responses show a reduction in response magnitude when two visual stimuli are shown in quick succession (response suppression), with stronger reductions when the stimuli are identical than when they are different (repetition suppression, RS; **Fig. 1B**). While Lotter et al. (2020) showed that PredNet exhibits response suppression as a result of statistical learning, mirroring monkey visual cortex (Meyer & Olson, 2011), it is not clear whether it captures these two basic neural signatures of short-term temporal adaptation observed in humans.

To test this, we ran an analogous experiment in PredNet as in our recent human intracranial EEG study (Brands et al., 2024), using similar temporal stimulus manipulations and the same naturalistic scene images. To assess the degree to which PredNet shows brain-like temporal dynamics, we extracted error unit activations across all network layers and time steps for each stimulus condition and compared the activation dynamics with human visual cortex responses. Overall, our results show that while the PredNet captures several aspects of the neural adaptation signatures, it also deviates from the neural data and therefore not fully accounts for temporal adaptation phenomena in visual cortex.

Methods

Predictive coding networks

Architecture The original description of the PredNet can be found in Lotter et al. (2016). Briefly, the PredNet consists of a hierarchical stack of layers with each layer l containing four different unit types, namely representational (R_l), target ($\hat{A}_l$), prediction (A_l) and error (E_l) units (**Fig. 2A**). At each timestep t updating of unit activations occurs through two passes. First, the states of the representational units R_l are updated in a top-down pass via a convolutional LSTM (Hochreiter & Schmidhuber, 1997; Shi et al., 2015), which receives inputs from the error E_l^{t-1} and representational R_l^{t-1} units from the previous timestep and representational units R_{l+1}^t from the layer above. Following this top-down pass, a bottom-up pass is made where the prediction for the next frame is generated and the difference is computed between the actual A_0^t and predicted $\hat{A}_0^t$ target. These errors E_0^t are then used as input to the next layer from which a new prediction is generated (via a convolution). Based on a previously performed hyperparameter search (Lotter et al., 2016), a four-layer model with 3 × 3 filter sizes for all convolutions and stack sizes per layer of 3, 48, 96 and 192 for the E and R modules was adopted.

Pretrained PredNet We examined activations derived from a pretrained PredNet[1] which was optimized on 10-frame sequences of 128×160 pixel RGB videos from the KITTI dataset (Geiger et al., 2013), which consists of a collection of videos obtained from a car-mounted camera while driving in Germany. Image sequences were sampled from the "City", "Residential" and "Road" categories. Note that the model is trained in a self-supervised manner to perform next-frame prediction, without external labels or other forms of supervision.

Datasets To test the robustness of the results of the pretrained network, we additionally trained a number of PredNet instances from scratch[2], on four different videos. First, we retrained one PredNet on videos belonging to the KITTI dataset mentioned above, to serve as a controlled comparison with the results obtained from the pretrained network. The other three videos belonged to the "Walking Tours" dataset (Venkataramanan et al., 2023). This dataset contains a set of first-person hours-long videos, captured in a single uninterrupted take, depicting a large number of objects and actions with natural scene transitions. We selected two videos recorded in urban areas, specifically *Amsterdam* and *Venice*, and one video from a *wildlife* safari.

Training procedures All four videos were preprocessed, which consisted of downsampling to 15 frames-per-second and resizing to 128×160 pixels. Each network was trained for 150 epochs on one of the videos, with each epoch consisting of 500 samples. Samples consisted of a series (batch size of 4) of 10-frame sequences that were randomly selected. Unless stated otherwise, mean squared error loss was com-

[1] Pretrained model can be found at https://github.com/coxlab/prednet.git

[2] Code is available on https://github.com/ABra1993/tAdaptation_PredNet.git and neural data is available on https://openneuro.org/datasets/ds004194.

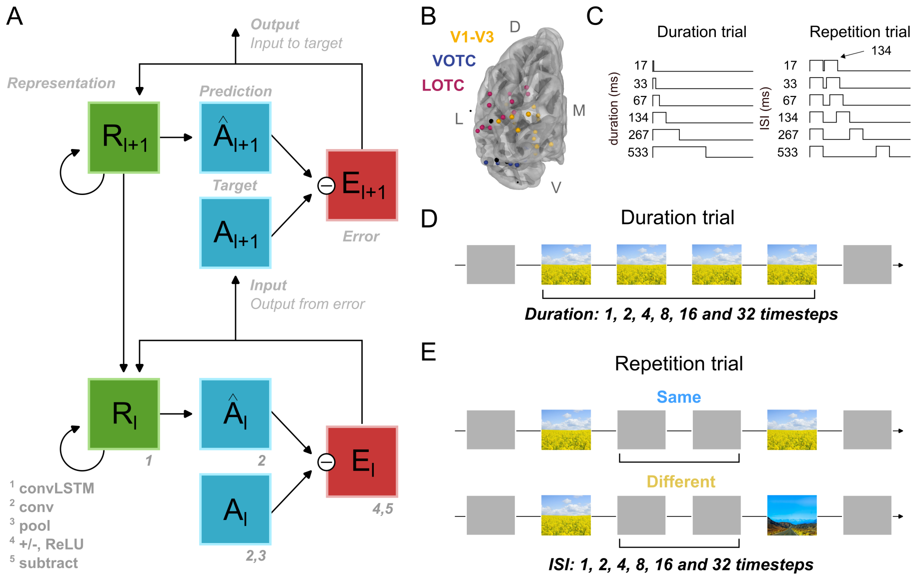

Figure 2: Experimental design. A: Information flow within the deep predictive coding network PredNet. Figure is adapted from Lotter et al. (2016). B: Location of the electrodes in the iEEG dataset (Brands et al., 2024). Electrodes were located in V1-V3 ($n = 17$), VOTC ($n = 11$) and LOTC ($n = 13$). Electrodes denoted in black were not assigned to any of the visual groups. C: Temporal structure of the stimulus presentations in the human experiment consisting of two trial types. *Left*, Single stimuli with varying duration. *Right*, Repeated stimuli with varying inter-stimulus interval. D-E: Analogous experimental settings to obtain temporal adaptation dynamics in the PredNet, including single (D) and repeated (E) stimulus trials with varying duration or inter-stimulus-interval, respectively.

puted for the first (pixel) layer aggregated over all time steps. Additionally, we also trained a set of network instances that used a layered predictive loss, where the error was computed over all, instead of only the first layer. Model weights were updated after each batch and an Adam optimizer was used with a learning rate of 0.001, reduced by a factor of 0.9 each 100 batches. To assess reliability, multiple instances ($n = 3$) with different random initial weights were trained for each video.

Human brain recordings

To determine to what degree emergent temporal dynamics in the PredNet exhibit the same neural adaptation signatures as human visual cortex responses, we reanalyzed an open dataset from Brands et al. (2024). In this study, iEEG data were collected from four participants implanted with subdural electrodes for clinical purposes, who were presented with naturalistic visual images for various durations and repetition intervals (see below). Raw voltage time courses from clinical strip, grid and depth electrodes were referenced to the common average for each electrode strip, and then filtered into separate 10 Hz wide frequency bands ranging between 50-200 Hz. This was followed by calculating the power envelope of each band-pass filtered time course, which were then averaged across bands to yield a time-varying broadband time course. By aggregating responses across patients, we obtained 41 visually response electrodes which were separated in one lower-level group (V1-V3, $n = 17$) and two higher-level groups, covering ventral-occipital cortex (VOTC, $n = 11$) and lateral-occipital cortex (LOTC, $n = 13$) (**Fig. 2B**).

Stimuli

For the experimental setup, PredNet instances were presented with the same stimuli that were used during collection of the neural dataset described above (Brands et al., 2024). This stimulus dataset consisted of 288 images (569×568 pixels) from six categories: buildings, bodies, faces, objects, scenes and scrambled. All categories except scenes consisted of the visual category depicted on a gray background, while PredNets are optimized on image sequences of video frames covering all pixels. To minimize potential effects of input distribution shifts on PredNet performance and activations, we used only the 48 images from the scene category, consisting of indoor, outdoor man-made and outdoor natural scenes.

Experimental design

To compare the emergent temporal dynamics in the PredNet with those in the iEEG dataset, we emulated the stimulus conditions from the human experiment and presented the network with two different trial types. In the human experiment, these two trial types were referred to as duration and repetition trials (**Fig. 2C**). Duration trials showed a single stimulus for one of six durations (**Fig. 2C**, *left*), namely 17, 33, 67, 134, 267 and 533 ms. Repetition trials contained a repeated presentation of either the same or two different images with fixed duration

(134 ms) but variable inter-stimulus interval (ISI) (**Fig. 2C**, *right*), ranging between 17-533 ms with the same temporal step sizes as the duration trials. The PredNets were also presented with duration (**Fig. 2D**) and repetition (**Fig. 2E**) trials, whereby the stimulus duration and ISI were similarly varied across six different temporal conditions defined in powers of two, i.e. 1, 2, 4, 8, 16 and 32 model time steps. Each trial consisted of 45 model timesteps; pixel values of inputs at model timesteps with no image were set to 0.5 (mid-gray).

Summary metrics

To characterize the temporal dynamics of the PredNets' error unit activations and the iEEG broadband responses, we computed several summary metrics. We here focus on the error units because these were previously found to exhibit the highest similarity with neural data (Lotter et al., 2020). However, results for the other three unit types (presentation, target and prediction) are provided in the Supplementary Section (**Fig. S1**-**Fig. S2**.)

Degree of subadditive temporal summation To determine the degree of subadditive temporal summation of either the neural or PredNet responses with increasing stimulus duration, we fitted the response magnitude (computed as the sum over the stimulus-on period, thereby excluding the offset) across durations with a logarithmic, *y = a + b · log(t)*, or linear, *y = c· x + d* function, where *y* is the response magnitude, *x* the stimulus duration and *[a, b]* and *[c, d]* are free parameters for the logarithmic and linear function, respectively. To then quantify the degree of subadditivity, we computed the ratio between the coefficient of determination (R^2) for the logarithmic and linear fit, whereby $R^2 < 1$ suggests that temporal summation is linear, while $R^2 > 1$ indicates that the summation of responses is subadditive.

Recovery from adaptation for repeated stimuli To quantify the degree of adaptation to repeated stimuli, we computed a recovery value defined as the response magnitude of the second stimulus as a proportion of the first. For the neural data, we averaged the response time courses for the 134 ms duration stimulus for single trials and all the repeated stimulus trials up to the onset of the second stimulus, thereby obtaining an average of the response to the first stimulus. This average was then subtracted from the repeated stimulus trials to isolate the second response. Recovery from adaptation was then defined as the Area Under the Curve (AUC) of the second response relative to the AUC of the first. For the PredNet activations, we used a similar approach, whereby we subtracted the first response, obtained from a duration trial with one model timestep, from the repeated sequence trial, after which we computed the ratio between the first and second response (AUC of the second response/AUC of the first response). To then describe the degree of recovery from response suppression both for the neural responses and PredNet responses, we fitted the recovery values across temporal conditions with the logarithmic function introduced above, *y = a + b · log(t)*.

Results

Error units in the first layer of PredNet capture subadditive temporal summation We first describe temporal dynamics in responses to duration-varying stimuli in the iEEG broadband data. Neural timecourses exhibit subadditive temporal summation as illustrated in **Fig. 1A**, which refers to the phenomenon that additional visual exposure resulting from longer presentation durations does not accumulate into a linearly increasing neural response. This subadditivity results from transient-sustained dynamics in the neural response time course, which can be clearly seen in **Fig. 3A** (*top*): Responses to visual stimulus onsets show an initial transient, which for short stimulus durations is the only part of the response. As stimulus duration increases, this transient response saturates, and a lower-amplitude sustained response emerges. Because prolonging stimulus duration no longer increases the peak transient, and only adds more of the relatively lower-amplitude sustained component, the overall response to the stimulus grows progressively less with longer stimulus durations (**Fig. 3B**). This compressive effect is evidenced by a qualitatively better fit for a logarithmic than a linear function between stimulus duration and response magnitude in V1-V3 (dependent-samples T-test computed over the electrodes, $t_{(16)}$ = -5.41, $p < 0.001$) and LOTC ($t_{(12)}$ = -6.23 , $p < 0.001$), with a similar pattern in VOTC (**Fig. 3C**).

To determine whether PredNet also exhibits subadditive temporal summation, we emulated the iEEG experiment by presenting the network with the same duration-varying stimuli. Similar to the iEEG broadband responses, the PredNet exhibits transient-sustained dynamics (**Fig. 3A**, *bottom*): error units across all layers show a transient response at stimulus onset, with a sustained response emerging as the stimulus duration increases. The qualitative response shape of the error units differ somewhat from the neural data, evident by the quicker decay after the transient and a lower sustained response. Here, units in the first network layer (*E1*) exhibit subadditive temporal summation, evidenced by the best fit with a logarithmic function (**Fig. 3DE**, dependent-samples T-test computed over the images, $t_{(47)}$ = -21.92, $p < 0.001$). Higher-level layers, including E2, E3 and E4, show comparatively less subadditive summation, demonstrated by the best fit for a linear function (E2, $t_{(47)}$ = 56.45, $p < 0.001$; E3, $t_{(47)}$ = 3.51, $p < 0.001$; E4, $t_{(47)}$ = 13.72, $p < 0.001$).

Notably, we also observe a strong discrepancy between PredNet unit activation timecourses and broadband iEEG responses: PredNet shows a second activation peak to the offset of the stimulus which is absent in the average neural response time courses. Moreover, we find that the other three unit types similarly exhibit transient-sustained dynamics and systematic offset responses, but deviate even more from the neural data, evident by the fact that there is no consistent subadditive temporal summation in any layer; only $\hat{A}4$ shows a slightly higher fit with a logarithmic function (**Fig. S1A-C**). Overall, these results show that error units in the first layer of PredNet capture subadditive temporal summation in neural re-

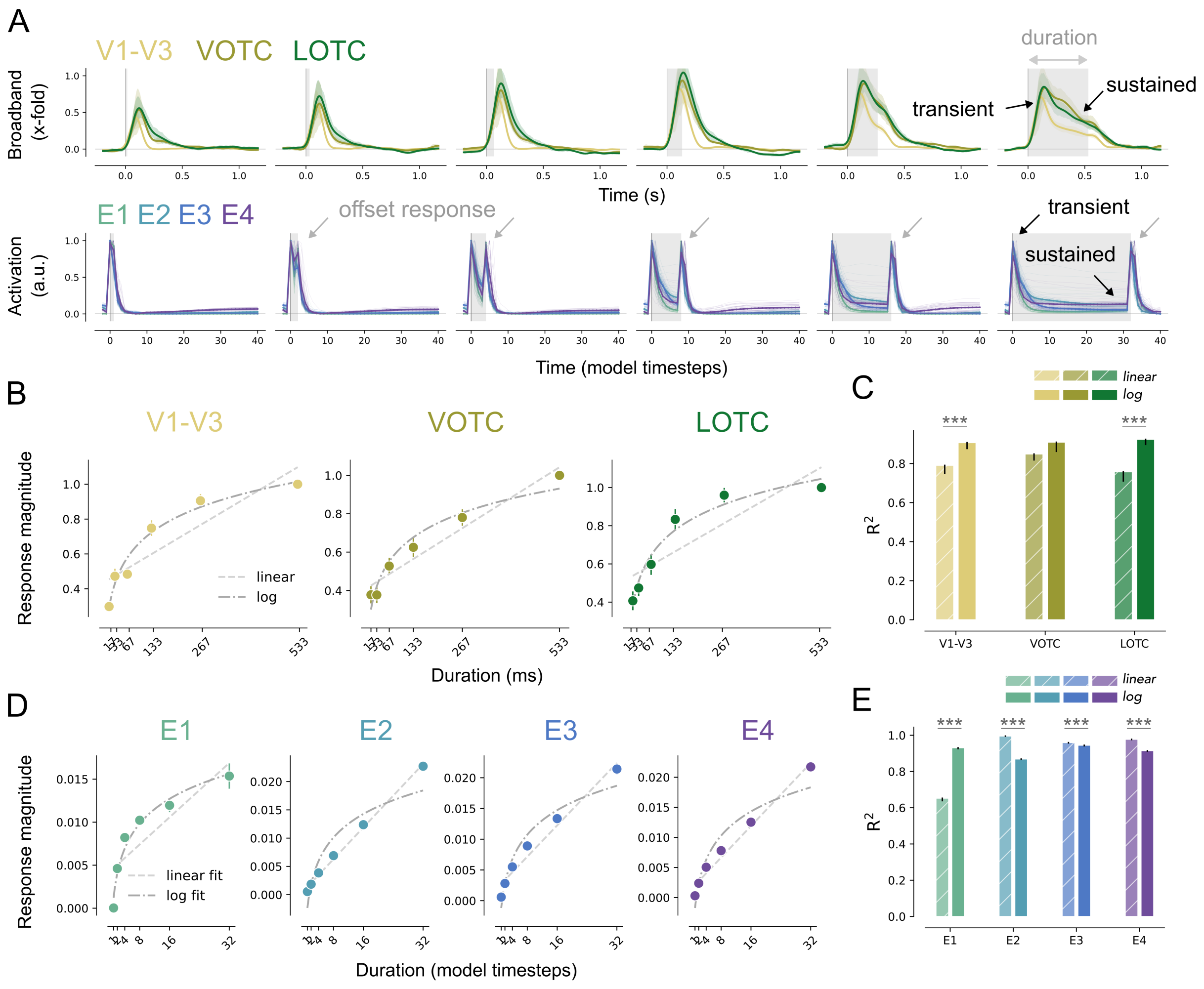

Figure 3: Neural responses and PredNet both exhibit subadditive temporal summation for single image presentations. A: *Top*, iEEG responses for electrodes assigned to V1-V3, VOTC or LOTC to single stimuli (gray) from short (left, 17 ms) to longest (right, 533 ms) stimulus durations. *Bottom*, PredNet activations of the error units for similar temporal conditions as in the human experiment across all model layers (from layer 1 to 4, E1, E2, E3 and E4). B: Sum of iEEG responses separately for each stimulus duration. The lines are fitted using either a linear or logarithmic function. C: Explained variance (coefficient of determination) of summed response magnitude per stimulus duration by a linear or logarithmic curve for each visual area. D-E: Same as B-C but for the error units of the PredNet model, separately for each model layer. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

sponses, but also show systematic offset responses throughout the network, thereby deviating from neural data.

Unlike visual cortex, PredNet does not exhibit short-term repetition suppression Another neural signature of temporal adaptation commonly observed in human visual cortex is a response reduction when stimuli are repeated, as illustrated in **Fig. 1B**. In our neural iEEG dataset, broadband responses to a second stimulus shown shortly after a first are indeed reduced (**Fig. 4A**, *top*). This suppression is strongest for shorter ISIs and gradually recovers as the gap between two stimuli increases. Moreover, we also observe repetition suppression: a stronger response reduction when the second stimulus is the same as the first compared to when it is different. We confirm these observations by quantifying the response magnitudes, revealing response suppression across all ISIs (**Fig. 4B**) and a substantially stronger reduction for same compared to different stimuli in all visual areas (dependent-samples T-test over electrodes, V1-V3, $t_{(16)}$ = -8.17, p < 0.001; VOTC, $t_{(10)}$ = -5.09, p < 0.001; LOTC, $t_{(12)}$ = -3.21, p = 0.008; **Fig. 4C**).

For PredNet, error units show some response suppression for short ISIs (one-sample t-test, degree of recovery averaged across all layers for shortest ISI vs. 1, $t_{(47)}$ = -7.06, $p < 0.001$), but unit activity quickly recovers as the number of model steps between stimuli increases (**Fig. 4A**, *bottom*). This recovery occurs comparatively much faster than for neural responses, which still exhibit substantial suppression for the longest ISI;

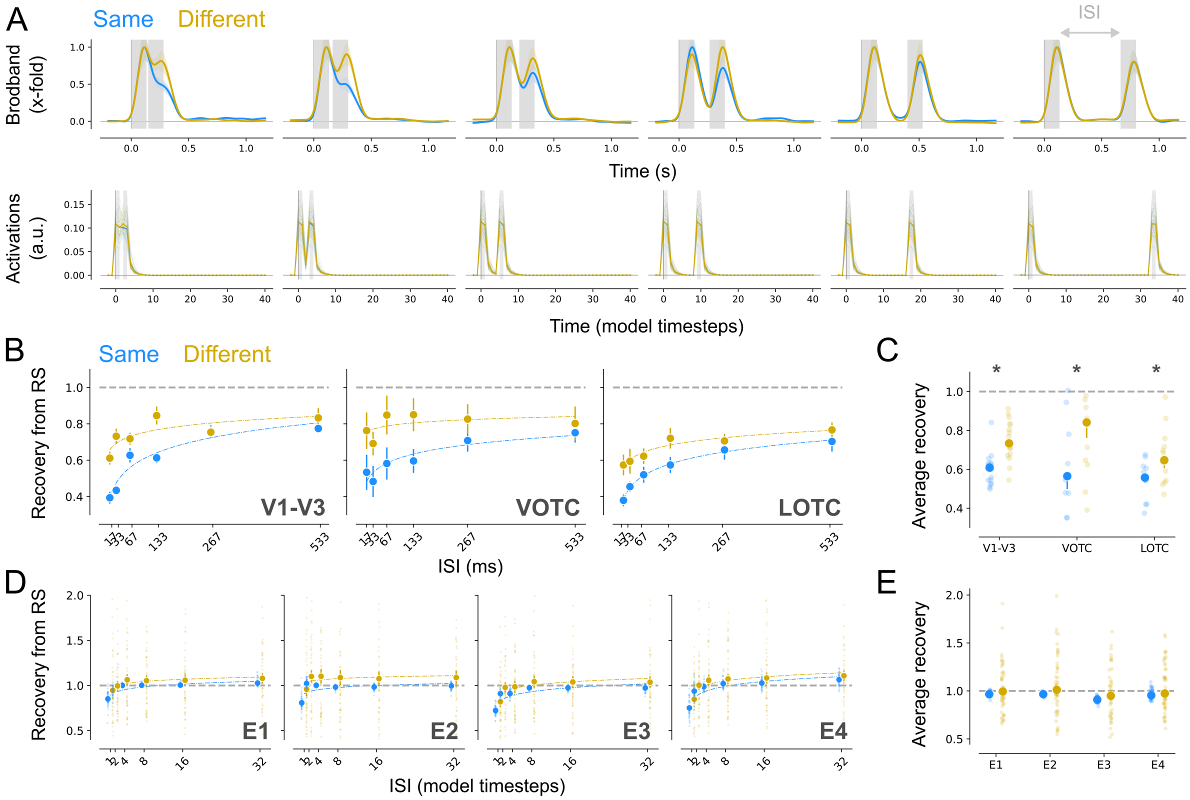

Figure 4: PredNet does not exhibit stronger suppression for same compared to different inputs shown in sequence. A: *Top*, iEEG responses for electrodes in V1-V3, VOTC or LOTC to repeated stimuli (gray) which are either the same or different, from short (left, 17 ms) to longest (right, 533) ISI. *Bottom*, PredNet activations of the error units in the first network layer for similar temporal conditions as in the human experiment. B: Recovery from adaptation computed as the ratio of the Area Under the Curve of the response to the first and second stimulus. C: Average recovery from adaptation averaged over al ISIs. D-E: Same as B-C but for the error units of PredNet, separately for each model layer (from layer 1 to 4, E1, E2, E3 and E4). * $p < 0.05$.

in contrast, we already observe near-full recovery across all PredNet layers for the second-to-shortest ISI (**Fig. 4D**).

Moreover, PredNet does not exhibit the hallmark signature of repetition suppression: For the shortest ISIs, the unit activations are slightly lower for same than different image pairs, but the response reduction for longer ISIs is remarkably similar, across all network layers (**Fig. 4E**). We observe similar patterns for the other three unit types (**Fig. S2A-C**), including the response reduction for short ISIs and the overall lack of repetition suppression.

Interestingly, the absence of response suppression in PredNet is very stable across same image pairs, while different image pairs exhibit a large spread around zero, showing both positive and negative recovery values which indicates both suppression and enhancement of unit activations. This lack of suppression is surprising given the a priori expectation that same image pairs should be more predictable to PredNet (resulting in lower error unit activations). Importantly, this pattern is again different than was found in the neural data, where responses showed similar variability in response suppression for both same and different image pairs (**Fig. 4C**).

One potential explanation for the modest suppression and the lack of RS in PredNet is that we presented stimuli for just one model timestep, whereby units did not have the opportunity to update their predictions during a longer exposure to the first image. To determine if PredNet exhibits stronger response suppression and repetition suppression for more prolonged exposure conditions, we also ran the repetition experiment using durations of 8 time steps per stimulus (**Fig. S3**). However, this analysis yielded similar results: PredNet units already exhibited near-full recovery from adaptation for relatively short ISIs, although interestingly, layers E3-4 showed slightly more suppression for longer ISIs. Notably, the overall degree of recovery was again similar for same and different image pairs, indicating a lack of RS.

All together, these findings demonstrate that PredNet shows less temporal adaptation for repeating stimuli compared to our neural recordings, and notably does not capture the strong repetition suppression effects observed in human visual cortex.

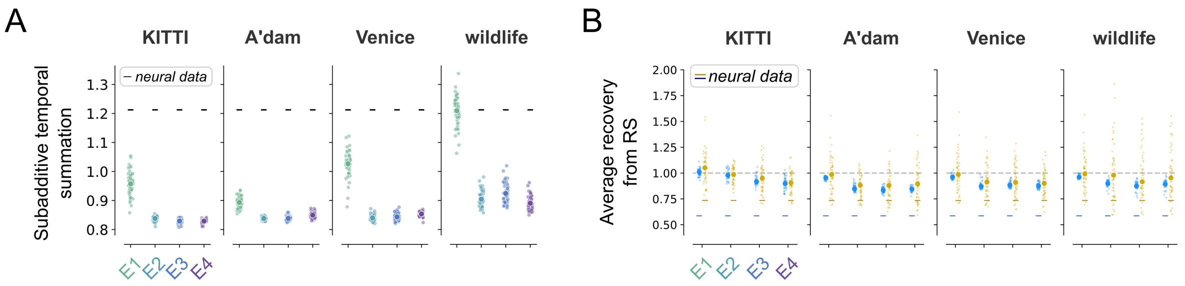

Figure 5: Temporal adaptation signatures in PredNet are robust across datasets. A: Subadditive temporal summation in PredNet instances trained on the different videos, defined as the ratio between a log and linear fit of the summed responses of varying stimulus durations for the error units in layer 1 to 4 (E1, E2, E3 and E4, resp.). B: Average degree of recovery by computing the response suppression for repeated stimulus pairs which are either the same (blue) or different (yellow) for the error units in all layers. Same network instances and stimuli as depicted in panel A. The dotted black line depicts no temporal adaptation (value of 1). The horizontal solid lines in panel A (black) and B (blue and yellow) depict the values derived from the neural data across all visual areas.

Consistent emergent temporal dynamics in PredNets trained with various datasets and loss functions To assess the robustness and generalizability of the temporal dynamics of the error units for prolonged and repeated stimuli observed in PredNet, we trained several new network instances on four different videos (see Methods), including three videos belonging to the "Walking Tours" dataset (Venkataramanan et al. 2023). We chose videos derived from walking tour footage because they contain more diverse and also more biologically plausible inputs than the *KITTI* dataset, since videos were collected from the perspective of a walking human, as opposed to a car-mounted camera; possibly, PredNet could show more human-like dynamics when trained on more-human-like visual inputs. Training was successful for each video and sample predictions demonstrate that our in-house trained networks were able to make accurate predictions on the visual content they were trained on, such as trajectory of passing cars and shadows on the road (*KITTI*, **Fig. S4A**), sliding motion of the camera (*Venice*, **Fig. S4B**), approaching persons on a pedestrian road (*Amsterdam*, **Fig. S4C**) and animal movements (*wildlife*, **Fig. S4D**).

In line with our findings in pretrained PredNets, unit responses of model instances trained on the different videos all exhibit most pronounced subadditive temporal summation in the first layer (**Fig. 5A**, **Fig. S5A**). All the network instances also again show much faster recovery from response suppression compared to the neural data and no repetition suppression (**Fig. 5B**, **Fig. S5B**). Notably, while results are thus generally similar across videos, PredNets trained on the wildlife footage seem to have somewhat distinct dynamics compared to the other three videos, showing relatively stronger subadditive temporal summation in all layers. We hypothesize that this stronger subadditivity results from the presence of a higher degree of motion continuity in this video, referring to the smooth and natural progression of movement between frames without abrupt transitions, putatively resulting in more accurate predictions (and lower errors) of future frames. Consistent with this hypothesis, the *wildlife* footage contains the highest degree of temporal autocorrelation, and PredNets trained on this video show relatively minimal performance degradation when trained on static images rather than video (see **Fig. S6ABC**), suggesting the visual inputs are overall easier to predict, resulting in less sustained error activity.

In a separate set of analyses, we additionally investigated the effect of the loss computation during training on the emergent network dynamics. We find that networks that are optimized using a layered predictive process by minimizing error across all, instead of only the first layer, exhibit highly similar dynamics for both prolonged (**Fig. S7A**) and repeated (**Fig. S7B**) stimuli. Here, we do observe that minimizing the error across all layers seems to strengthen the overall subadditive temporal summation for duration-varying stimuli. Nonetheless, these results indicate that the observed temporal dynamics in PredNet are robust across multiple training datasets and optimization functions.

Discussion

Our aim was to evaluate whether PredNet, a deep predictive coding network (Lotter et al., 2016), exhibits key signatures of short-term neural adaptation as observed in human visual cortex. To this end, we compared activation dynamics exhibited by the network, with temporal dynamics observed in a neural dataset of iEEG broadband responses of humans presented with single or repeated stimulus presentations. For single stimulus presentations, we demonstrate that error unit activations in the first layer of PredNet exhibit subadditive temporal summation, but, as opposed to the neural data, also show systematic offset responses. For repeated stimulus presentations, error unit responses show a slight reduction in response magnitude when an image is preceded by another, but fail to show repetition suppression as observed in the neural responses, which is considered a robust and ubiquitous phenomenon across the brain. These findings were consistent across datasets, optimization functions, and model unit types.

In human visual cortex, neural responses show adaptation over time, thereby exhibiting subadditive temporal summation and repetition suppression to single and repeated stimuli, respectively (Zhou et al., 2019; Groen et al., 2022). Here, we confirm these two signatures of neural adaptation in a recently collected iEEG dataset (Brands et al., 2024), evident by the fact that neural responses in both early and late visual areas exhibit subadditivity of the response magnitudes with prolonged stimulus durations and stronger response suppression for same as opposed to different inputs when shown in sequence. We show that PredNets capture some of the dynamics present in the neural data, namely subadditive temporal summation, but fails to reproduce others, namely repetition suppression. To determine the fidelity of the PredNet to biological systems, it is helpful to ask which features and components of the model are responsible or necessary to reproduce these neural dynamics. For example, one key feature of the PredNet is recurrent connectivity, which is required to observe temporal dynamics. Another important PredNet feature, motivated by predictive coding principles (Rao & Ballard, 1999), is that the network explicitly computes an error representation in a population of neurons, which is propagated from layer to layer in a feedforward manner and is used to update the network representations and consequently, network predictions. These two PredNet features - recurrence and explicit error representation - presumably allow the network to exhibit several of the response properties observed in the neural data.

For prolonged stimuli, transient-sustained dynamics arise from an initial prediction error at stimulus onset, followed by error decay as the model updates its predictions. However, the appearance of the gray image at stimulus offset induces a strong prediction error in PredNet, while it is absent in the neural data. Previous studies show inconsistencies regarding the presence of offsets in neural responses, with some iEEG dataset showing a similar lack (Zhou et al., 2019; Groen et al., 2022), while other human iEEG (Zhou et al., 2019) and animal (Bair et al., 2002; Benucci et al., 2009) datasets did show an offset response in at least a subset of neural responses.

From a methodological perspective, the presence of an offset response in the neural timecourses can depend on several factors, such as the data type used for data collection (e.g., fMRI vs. iEEG), brain areas sampled, or experimental design. From an empirical perspective, it is still debated what causes offset responses or the lack thereof. A previous study noted that offset responses in an iEEG data were more pronounced for electrodes with peripherally tuned spatial receptive fields, suggesting a link between the offset response and spatial coverage of the stimulus (Zhou et al., 2019). Other work has hypothesized that offset dynamics are related to neural representation which reflects differences in information processing. More specifically, transient responses both on the on- and offset of stimuli indicate involvement in detecting temporal change, whereas the absence of an offset is related to other types of information processing, including object recognition and appearance (Zhou et al., 2018). Moreover, earlier work has proposed segregated neural pathways for onset and offset responses as a feature of many sensory computations, for example in motion detection (Westheimer, 2007), perceptual grouping of auditory stimuli (Bregman, 1994) and olfaction-related behavior (Chalasani et al., 2007). Thus, accurately capturing the heterogeneous neural offset response profiles for single, duration-varying stimuli may require implementing additional features in PredNet, such as spatial topography (e.g. Lu et al. 2023) or separate pathways for predicting motion and object identity (e.g. Choi et al. 2023).

For repeated stimuli, recurrence and explicit error representations do result in a slightly lower error during the second compared to first stimulus presentation, especially for when two stimuli are in close temporal proximity of each other and the two stimuli are the same, likely reflecting the lingering prediction of the first stimulus presentation during the presentation of the second stimulus. However, due to the fast updating of the network representations and possibly the intervening gray image in the ISI, the network "forgets" previous inputs as the time in between stimuli increases, resulting in a full recovery of the response suppression for short ISIs and no difference between same and different stimuli, thereby significantly deviating from the observations in the neural data. These results demonstrate that while features as recurrence and explicit error representation may effectively capture some of the neural signatures of temporal adaptation, there is a misalignment between "biological" time and the notion of "time" in the PredNet. Additional features, either on the side of the inputs (e.g. adjusting the sample rate of the frames, Butts et al. 2007) or in PredNet itself (e.g. controlling the rate with which representations are updated; Chien et al. 2021) might be necessary to flexibility adjust the temporal resolution of the PredNet such that it better matches that of the neural data.

Our findings are subject to several limitations. First, it is important to note that our results do not refute the predictive coding scheme at large, but its particular instantiation in PredNet. This specific model has been successful at capturing several neurophysiological phenomena observed in visual cortex (Lotter et al., 2020). Our comparison with short-term neural adaptation, specifically the failure of PredNet to reproduce repetition suppression effects, provides a notable exception to these earlier positive findings. To determine whether alternative implementations of the predictive coding scheme do account for short-term temporal adaptation in human visual cortex, future research could explore solutions provided by other studies (e.g. Rane et al. 2020; Heilbron & de Lange 2023; Gutlin & Auksztulewicz 2025). Second, in the current setup, the PredNet was tested on stimuli that differed from the training data, which may have contributed to the mismatch with the neural responses. A valuable future direction would therefore be to train the model on stimuli that better match the experimental paradigm, including abrupt onsets and offsets of stimuli, so that the model is familiarized with conditions more comparable to those in the experimental setup. An alternative direction could be to include explicit temporal inductive biases in the network, such as periodic-like movements (e.g. Perrinet et al. 2014), which may improve biological plausibility.

Conclusion

In this study, we highlight the potential of PredNet in modeling certain aspects of temporal adaptation, while also showing misalignments with the neural data, suggesting that predictive, top-down processes - as implemented in PredNet - are not sufficient to fully capture signatures of short-term temporal adaptation in human visual cortex.

Acknowledgments

This work was supported by a MacGillavry Fellowship from the University of Amsterdam to IIAG.

References

Auksztulewicz, R., & Friston, K. (2016). Repetition suppression and its contextual determinants in predictive coding. *cortex*, *80*, 125–140.

Bair, W., Cavanaugh, J. R., Smith, M. A., & Movshon, J. A. (2002). The timing of response onset and offset in macaque visual neurons. *Journal of Neuroscience*, *22*(8), 3189–3205.

Benucci, A., Ringach, D. L., & Carandini, M. (2009). Coding of stimulus sequences by population responses in visual cortex. *Nature neuroscience*, *12*(10), 1317–1324.

Brands, A. M., Devore, S., Devinsky, O., Doyle, W., Flinker, A., Friedman, D., ... Groen, I. I. A. (2024). Temporal dynamics of short-term neural adaptation across human visual cortex. *PLOS Computational Biology*, *20*(5), e1012161.

Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound*. MIT press.

Butts, D. A., Weng, C., Jin, J., Yeh, C.-I., Lesica, N. A., Alonso, J.-M., & Stanley, G. B. (2007). Temporal precision in the neural code and the timescales of natural vision. *Nature*, *449*(7158), 92–95.

Chalasani, S. H., Chronis, N., Tsunozaki, M., Gray, J. M., Ramot, D., Goodman, M. B., & Bargmann, C. I. (2007). Dissecting a circuit for olfactory behaviour in caenorhabditis elegans. *Nature*, *450*(7166), 63–70.

Chien, H.-Y. S., Turek, J. S., Beckage, N., Vo, V. A., Honey, C. J., & Willke, T. L. (2021). Slower is better: revisiting the forgetting mechanism in lstm for slower information decay. *arXiv preprint arXiv:2105.05944*.

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*, *36*, 50408–50428.

Fonseca, M. (2019). Learning to predict visual brain activity by predicting future sensory states. In *Real neurons* $\{\backslash\&\}$ *hidden units: Future directions at the intersection of neuroscience and artificial intelligence@ neurips 2019.*

Friston, K. (2003). Learning and inference in the brain. *Neural Networks*, *16*(9), 1325–1352.

Friston, K. (2005). A theory of cortical responses. *Philosophical transactions of the Royal Society B: Biological sciences*, *360*(1456), 815–836.

Geiger, A., Lenz, P., Stiller, C., & Urtasun, R. (2013). Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research*, *32*(11), 1231–1237.

Groen, I. I., Piantoni, G., Montenegro, S., Flinker, A., Devore, S., Devinsky, O., ... others (2022). Temporal dynamics of neural responses in human visual cortex. *Journal of Neuroscience*, *42*(40), 7562–7580.

Gutlin, D. C., & Auksztulewicz, R. (2025). Predictive coding algorithms induce brain-like responses in artificial neural networks. *bioRxiv*, 2025–01.

Heilbron, M., & de Lange, F. P. (2023). Higher-level spatial prediction during natural scene perception in mouse primary visual cortex.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735–1780.

Hohwy, J., Roepstorff, A., & Friston, K. (2008). Predictive coding explains binocular rivalry: An epistemological review. *Cognition*, *108*(3), 687–701.

Kirubeswaran, O., & Storrs, K. R. (2023). Inconsistent illusory motion in predictive coding deep neural networks. *Vision Research*, *206*, 108195.

Lillicrap, T. P., Santoro, A., Marris, L., Akerman, C. J., & Hinton, G. (2020). Backpropagation and the brain. *Nature Reviews Neuroscience*, *21*(6), 335–346.

Lotter, W., Kreiman, G., & Cox, D. (2016). Deep predictive coding networks for video prediction and unsupervised learning. *arXiv preprint arXiv:1605.08104*.

Lotter, W., Kreiman, G., & Cox, D. (2020). A neural network trained for prediction mimics diverse features of biological neurons and perception. *Nature machine intelligence*, *2*(4), 210–219.

Lu, Z., Doerig, A., Bosch, V., Krahmer, B., Kaiser, D., Cichy, R. M., & Kietzmann, T. C. (2023). End-to-end topographic networks as models of cortical map formation and human visual behaviour: moving beyond convolutions. *arXiv preprint arXiv:2308.09431*.

Meyer, T., & Olson, C. R. (2011). Statistical learning of visual transitions in monkey inferotemporal cortex. *Proceedings of the National Academy of Sciences*, *108*(48), 19401–19406.

Millidge, B., Tschantz, A., Seth, A., & Buckley, C. L. (2020). Relaxing the constraints on predictive coding models. *arXiv preprint arXiv:2010.01047*.

Perrinet, L. U., Adams, R. A., & Friston, K. J. (2014). Active inference, eye movements and oculomotor delays. *Biological cybernetics*, *108*, 777–801.

Rane, R. P., Szügyi, E., Saxena, V., Ofner, A., & Stober, S. (2020). Prednet and predictive coding: A critical review. In *Proceedings of the 2020 international conference on multimedia retrieval* (pp. 233–241).

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience*, *2*(1), 79–87.

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., & Woo, W.-c. (2015). Convolutional lstm network: A machine learning approach for precipitation nowcasting. *Advances in neural information processing systems*, *28*.

Spratling, M. W. (2008). Reconciling predictive coding and biased competition models of cortical function. *Frontiers in computational neuroscience*, *2*, 300.

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in cognitive sciences*, *13*(9), 403–409.

Venkataramanan, S., Rizve, M. N., Carreira, J., Asano, Y. M., & Avrithis, Y. (2023). Is imagenet worth 1 video? learning strong image encoders from 1 long unlabelled video. *arXiv preprint arXiv:2310.08584*.

Watanabe, E., Kitaoka, A., Sakamoto, K., Yasugi, M., & Tanaka, K. (2018). Illusory motion reproduced by deep neural networks trained for prediction. *Frontiers in psychology*, *9*, 340023.

Westheimer, G. (2007). The on–off dichotomy in visual processing: from receptors to perception. *Progress in retinal and eye research*, *26*(6), 636–648.

Zhou, J., Benson, N. C., Kay, K., & Winawer, J. (2019). Predicting neuronal dynamics with a delayed gain control model. *PLoS computational biology*, *15*(11), e1007484.

Zhou, J., Benson, N. C., Kay, K. N., & Winawer, J. (2018). Compressive temporal summation in human visual cortex. *Journal of Neuroscience*, *38*(3), 691–709.

Supplementary Section

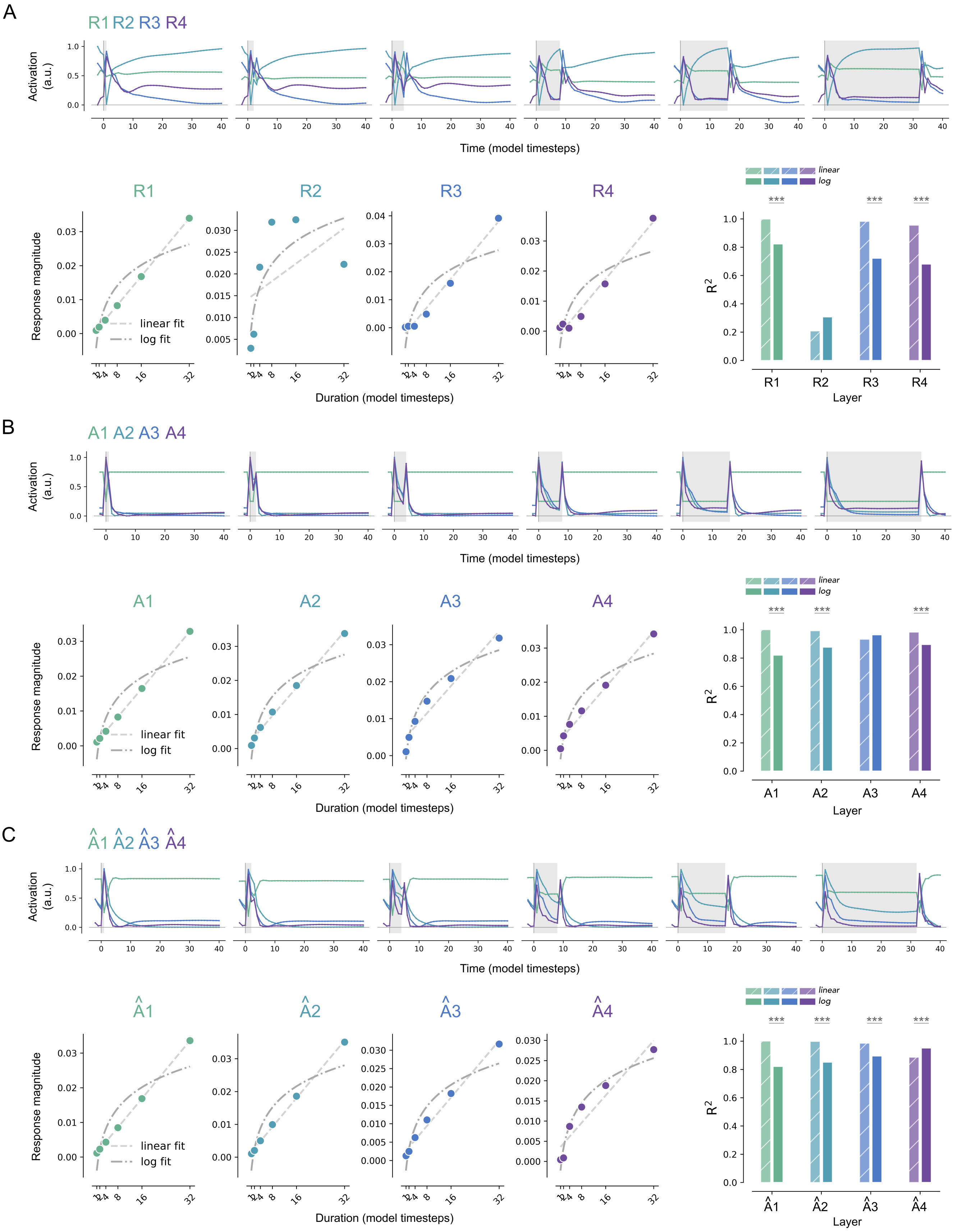
A
R1 R2 R3 R4
Activation (a.u.)
1.0
0.5
0.0
0 10 20 30 40
Time (model timesteps)
R1
R2
R3
R4
Response magnitude
linear fit
log fit
1 2 4 8 16 32
Duration (model timesteps)
linear
log
***
R²
Layer
R1 R2 R3 R4
B
A1 A2 A3 A4
Activation (a.u.)
Time (model timesteps)
A1
A2
A3
A4
Response magnitude
linear fit
log fit
Duration (model timesteps)
linear
log
***
R²
Layer
A1 A2 A3 A4
C
Â1 Â2 Â3 Â4
Activation (a.u.)
Time (model timesteps)
Â1
Â2
Â3
Â4
Response magnitude
linear fit
log fit
Duration (model timesteps)
linear
log
***
R²
Layer
Â1 Â2 Â3 Â4

Figure S1: *(previous page)* Representation, target and prediction units in the PredNet do not exhibit subadditive temporal summation for single image presentations. A: *Top*, PredNet activations of the representation units for similar temporal conditions as in the human experiment across all model layers. *Bottom left*, Sum of unit activations separately for each stimulus duration. The lines are fitted using either a linear or logarithmic function. *Bottom right*, Explained variance (coefficient of determination) of summed response magnitude per stimulus duration by a linear or logarithmic curve for each visual area. B-C: Same as panel A for target (B) and prediction (C) units. *** $p < 0.001$.

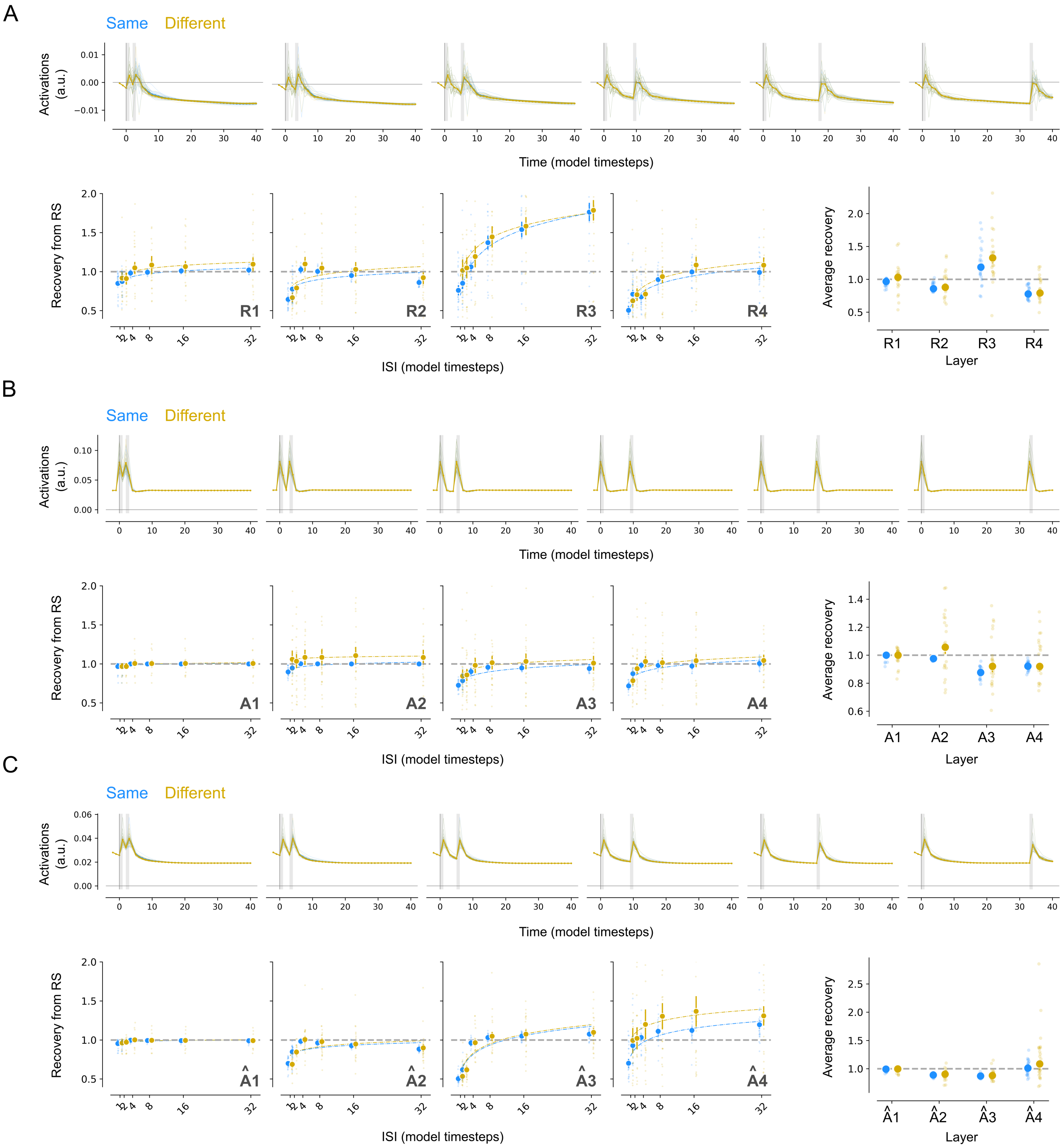
A
Same
Different
Activations (a.u.)
0.01
0.00
−0.01
0 10 20 30 40
Time (model timesteps)
Recovery from RS
2.0
1.5
1.0
0.5
R1
R2
R3
R4
1 2 4 8 16 32
ISI (model timesteps)
Average recovery
R1 R2 R3 R4
Layer
B
Same
Different
Activations (a.u.)
0.10
0.05
0.00
0 10 20 30 40
Time (model timesteps)
Recovery from RS
A1
A2
A3
A4
ISI (model timesteps)
Average recovery
1.4
1.2
1.0
0.8
0.6
A1 A2 A3 A4
Layer
C
Same
Different
Activations (a.u.)
0.06
0.04
0.02
0.00
Time (model timesteps)
Recovery from RS
Â1
Â2
Â3
Â4
ISI (model timesteps)
Average recovery
2.5
2.0
1.5
1.0
Â1 Â2 Â3 Â4
Layer

Figure S2: *(previous page)* Representation, target and prediction units in the PredNet do not exhibit stronger suppression for same compared to different inputs shown in sequence. A: *Top*, PredNet activations of the representation units in the third network layer for similar temporal conditions as in the human experiment. *Bottom left*, Recovery from adaptation computed as the ratio of the Area Under the Curve of the response to the first and second stimulus. *Bottom right*, Average recovery from adaptation averaged over al ISIs. B-C: Same as panel A for target (B) and prediction (C) units.

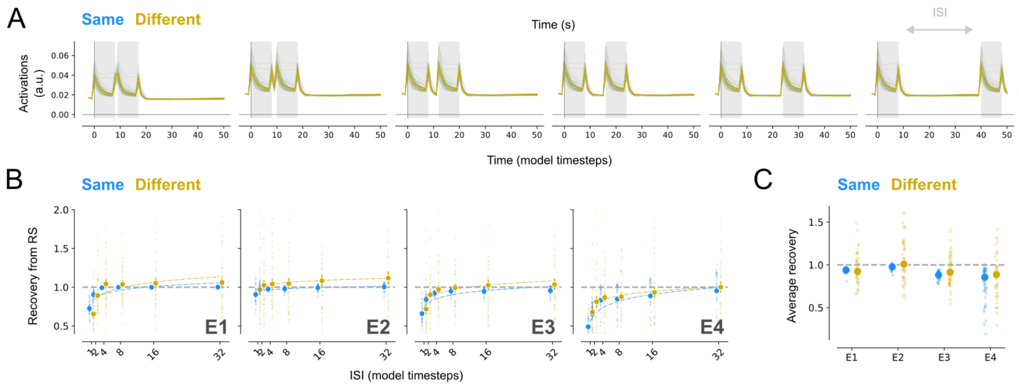

Figure S3: Recovery form repetition for trials with long stimulus durations. A: PredNet activations of the error units in the first network layer for the six different interstimulus intervals (ISI) between two stimuli with a duration of 8 model timesteps. B: Recovery from response suppression for same and different stimuli plotted separately per network layer (i.e. E1, E2, E3 and E4). The fitted curves express recovery as a function of the ISI. The dotted grey line depicts a recovery of 1 (i.e. when the magnitude of the first and second response is the same). C: Average degree of recovery computed over all ISIs plotted separately per network layer.

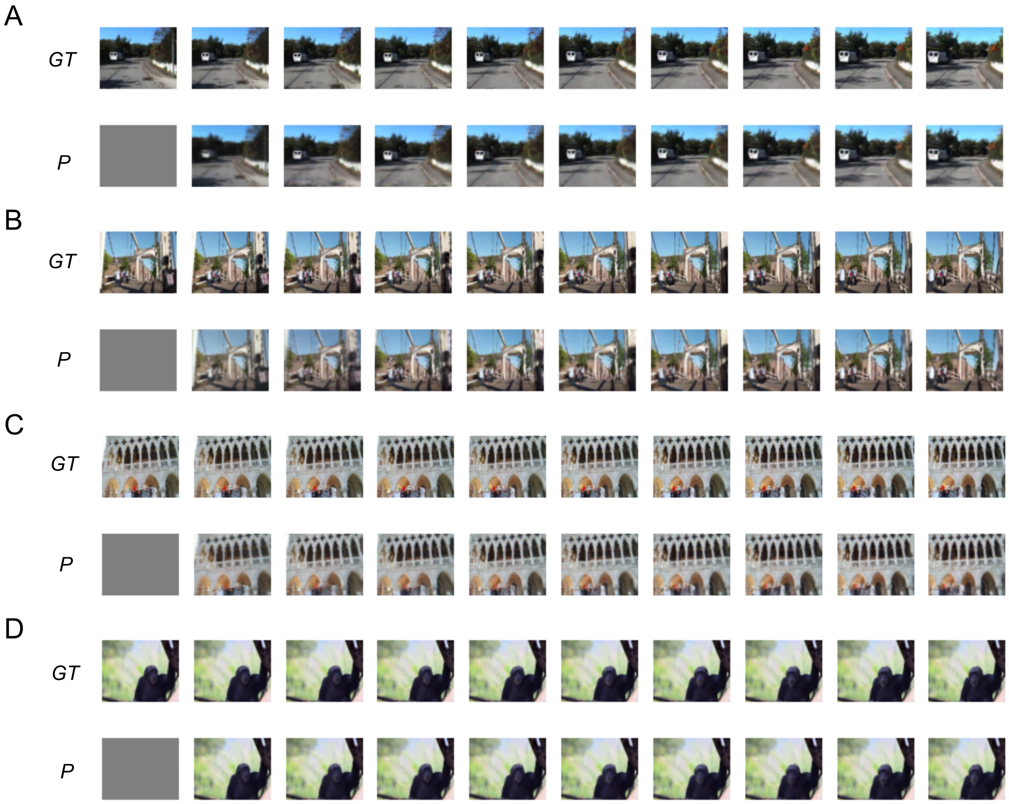

Figure S4: In-house trained PredNets accurately predict future frames. A-D: Next frame predictions of PredNet instances trained on four different videos: the *KITTI* dataset (A) and three videos from the "Walking Tours" dataset: *Amsterdam* (B), *Venice* (C) and a *wildlife* safari (D). GT = presented image (ground-truth); P = PredNet prediction.

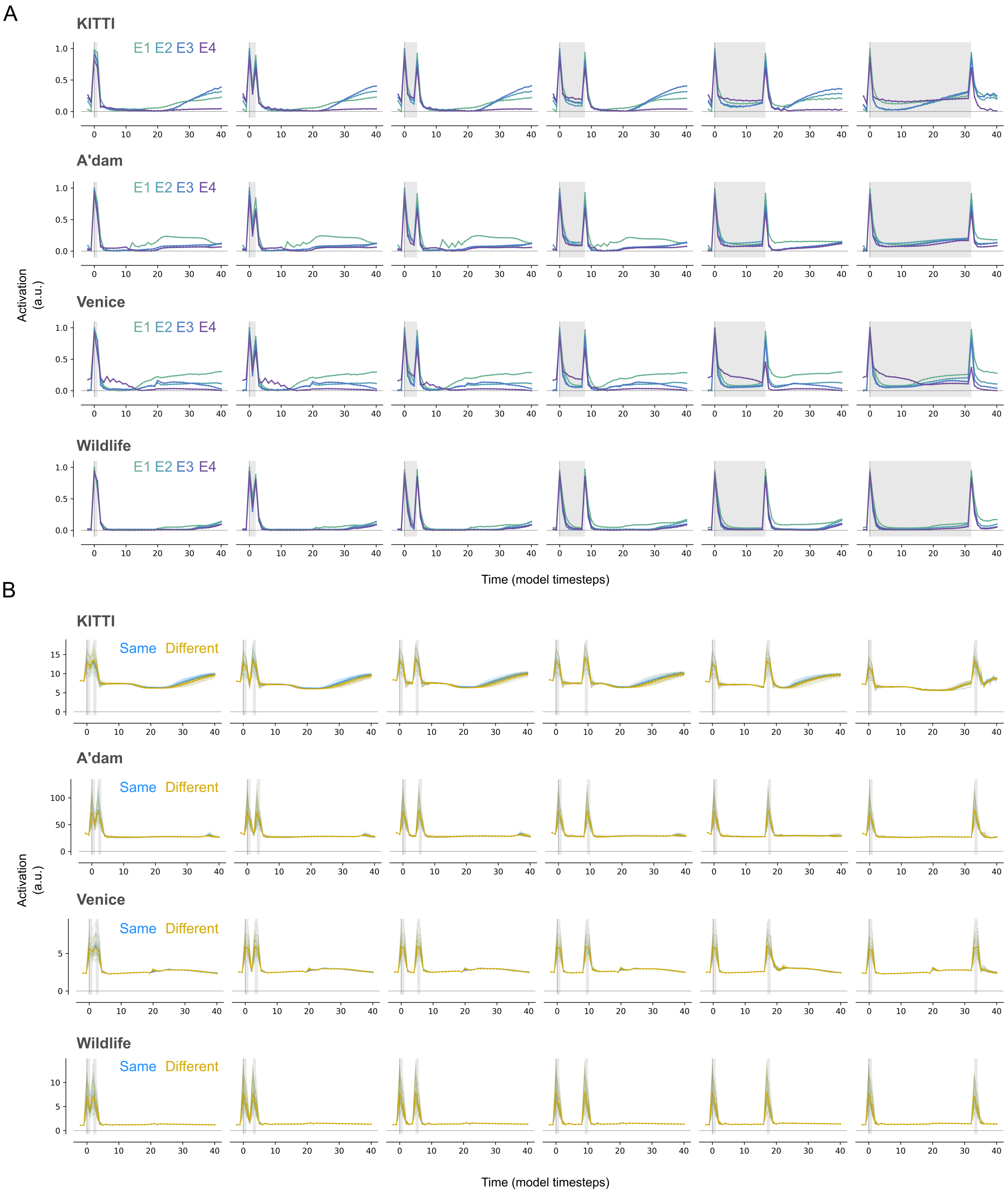
A
KITTI
E1 E2 E3 E4
A'dam
E1 E2 E3 E4
Venice
E1 E2 E3 E4
Wildlife
E1 E2 E3 E4
Activation
(a.u.)
1.0
0.5
0.0
0
10
20
30
40
Time (model timesteps)
B
KITTI
Same Different
15
10
5
0
A'dam
Same Different
100
50
0
Venice
Same Different
5
0
Wildlife
Same Different
10
5
0
Activation
(a.u.)
0
10
20
30
40
Time (model timesteps)

Figure S5: *(previous page)* Consistent dynamics of the error units across videos. A: PredNet activations of the error units for duration trials across all model layers (from layer 1 to 4, E1, E2, E3 and E4, resp.). Each row depicts results from training on one of four videos, namely a video belonging to the *KITTI* dataset, and three video's belonging to the “Walking tours” dataset, with footage recorded in *Amsterdam*, *Venice* or during a *wildlife* safari. B: PredNet activations of the error units averaged across network layers for repetition trials, including pairs of two stimuli that are the same or different, for networks trained on the same videos as shown in the rows of panel (A).

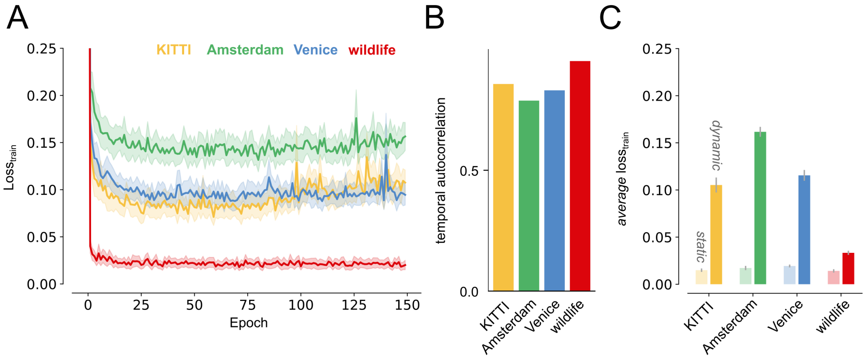

Figure S6: Training loss and image statistics for the different datasets. A: Average training loss across PredNet instances trained on the *KITTI* video dataset or one of three videos from the "Walking Tours" dataset, including *Amsterdam*, *Venice* and from a *wildlife* safari. Curves are smoothed with a Gaussian kernel with standard deviation of $\sigma = 10$. The shaded region depicts the SEM across network initializations ($n = 3$). B: Temporal autocorrelation, described by the Pearson product-moment correlation coefficients across video frames. Higher temporal autocorrelation occurs in videos with slow-moving objects or static scenes, where consecutive frames are very similar. Low temporal autocorrelation occurs in videos with fast motion or abrupt changes, where frames differ significantly over short intervals. C: Average loss across PredNet instances on the four videos introduced in panel (A), with either static (light) or dynamic (dark) frame sequences.

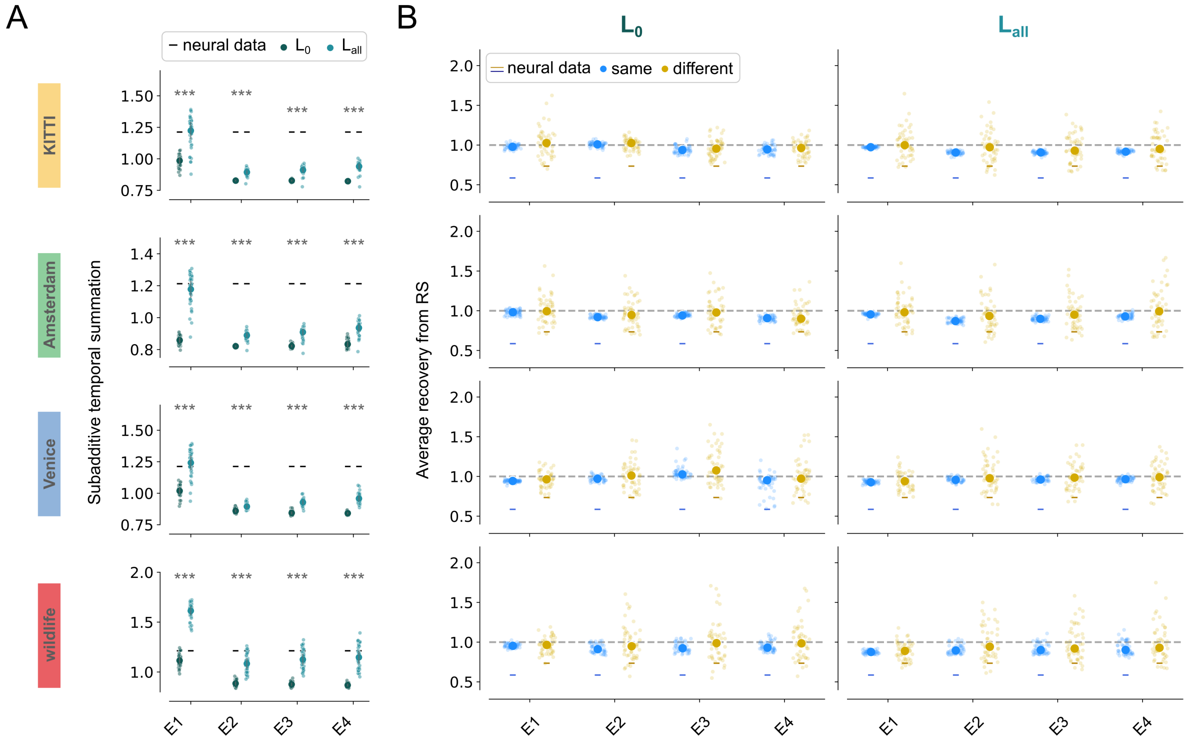

Figure S7: Error-minimization over solely the first or all network layers yield similar temporal dynamics. A: The degree of subadditive temporal summation for the error units in layer 1 to 4 (E1, E2, E3 and E4, respectively). PredNet instances were trained on one of the four videos (rows) with an L_0 (green) or L_{all} (blue) loss. Independent T-test, *** p $<$ 0.001. B: Average degree of recovery for same (blue) and different (yellow) repeated stimuli averaged over the inter-stimulus intervals, for the error units in layer 1 to 4. The dotted black line depicts no temporal adaptation (value of 1). The horizontal solid lines in panel A (black) and B (blue and yellow) depict the values derived from the neural data across all visual areas.