Predictive Networks, Illusions and Schizophrenia

by

Umar Shaik

A Report submitted to the
School of Graduate Studies
in partial fulfillment of the
requirements for the degree of
Master of *Science*

Department of *Computer Science*
Memorial University of Newfoundland

August 2019

St. John's Newfoundland

Abstract

An essential task in Artificial Intelligence (AI) is the prediction of future inputs from a given sequence, such as predicting the next frame of a video, with prominent applications such as self-driving cars. The human brain is very good at this task, without this ability we would be too slow to catch a ball or jump out of the way. A theory called "predictive coding" in neuroscience introduced by Ballard and Rao in 1999, explains the phenomena.

A recently developed predictive network from Massachusetts Institute of Technology (MIT) called 'PredNet' leverages the ideas of predictive coding for next-frame video prediction. Interestingly, PredNet has shown a very brain-like ability to be fooled by illusions of motion (when a static image appears to be moving). This leads us to pose the question of whether PredNet can be used to study perceptual disruptions in the brain associated with mental disorders, particularly schizophrenia.

In this study, we show how several types of modifications to PredNet designed to simulate different models of perception disruption in schizophrenia from neuroscience literature, affects its ability to be fooled by illusions of motion, and the overall effect these disruptions have on the predictive ability.

Acknowledgements

I would like to express my gratitude to Dr. Antonina Kolokolova, my research supervisor, for her patient guidance, enthusiastic encouragement and useful critique on this research work.

I also wish to thank various people for their contribution to this project, Daniel Power and Hilary Sinclair for their valuable time and support in this project.

Contents

List of Figures

1 Introduction

Recent years saw an explosion of research in neural networks, in particular deep neural networks inspired by the brain structure. In particular, convolutional neural networks have been very successful at tasks such as classifying visual information and labeling objects in an image.

While being able to correctly process a given image is an important basic task, the next step is to process a sequence of images such as frames of a video. There, in addition to processing a given image the system needs to be able to estimate the future state of the object several frames ahead, given its current state and history ([Villegas et al., 2017], [Mathieu et al., 2015a]). Two prominent applications of this are self-driving cars (waymo) and Action-Conditional Video Prediction in Atari Games [Oh et al., 2015]. In particular, self-driving cars can predict locations of pedestrians or other cars based on their observed behaviour and steer away from trouble, making commuting safer. Similar is the case of Action-Conditional Video Prediction in Atari Games, where given a current state of the game like pacman trying to navigate through the maze safely, the network can not only predict the next frame but also take conditional actions depending on the situation in various games.

One way this is made possible is using theories that explain the information processing mechanism of the brain, in particular predictive coding. Predictive coding postulates that the information processing mechanism in the brain uses feedback connections to carry prediction from the higher abstract layers of the cortex down to lower-layer sensory layers, with feedforward connections carrying the difference between prediction and the sensory input (the error) upwards [Rao and Ballard, 1999].

A recently developed deep predictive neural network from MIT called 'PredNet' leverages the ideas of predictive coding for next-frame video prediction [Lotter et al., 2016]. PredNet learns to predict future frames, with each layer in the network making local predictions and forwarding only the difference in error generated from the prediction and input to the upper layers.

While human visual perception is exceptionally accurate, mistakes are occasionally made, such as in the case of visual illusions — for example, the rotating snakes illusion (link). Despite being a static image, the rotating snakes illusion induces a strong perception of motion illusion in most humans. However, some conditions, most notably schizophrenia, make people

susceptible to illusions, including illusions of motion.

Interestingly, PredNet has shown an ability similar to the brain of being fooled by illusions of motion. When presented with images such as the rotating snake illusion, PredNet detects motion as most humans perceive it (ie, clockwise rotation in the rotating snake illusion). This leads us to question whether PredNet can be used to study perceptual disruption in the brain associated with mental disorders, particularly schizophrenia.

In this project, we discuss the different models of perception disruption in schizophrenia (2.4). We then investigate the effects of implementing these models on the PredNet (3) and further discuss the effects of these models on different types of datasets, including illusions and real.

Each model of perception disruption inspired various ideas of disruption in the PredNet. The error modification disruption (3.2.1) was inspired by precision weighted error model(2.4.1). The forget top-down prediction (3.2.3) and forget current prediction (3.2.2) were inspired by the second model which is disrupted efference copies(2.4.2). Finally, the mix the current prediction with sensory (3.2.4), mix the top-down prediction with sensory (3.2.5) and effects of extra convolution (3.2.6) were inspired by the circular inference model (2.4.3).

While the error modification made PredNet less susceptible to illusions (Figure 9). on the other hand, the forget current prediction made PredNet produce less accurate predictions (Figure 15), and the forget top-down prediction disruption made PredNet repeat the same mistake on every frame (Figure 18). Addition to that, the last three disruptions had different results as well. The mix current prediction with sensory made PredNet create strong opposite illusory motion (Figure 22), the mix top-down prediction with sensory made PredNet create ghostly afterimage (Figure 28) and color inversions (Figure 26) and finally, the effects of extra convolution made PredNet produce strong outward flowing vectors (Figure 31).

2 Background

2.1 Predictive coding

Predictive coding proposes a model of visual processing in which the feedback connections carry predictions to lower level activities from a higher level, whereas feedforward connections carry errors between the predictions and

the actual lower level activity.[Rao and Ballard, 1999]

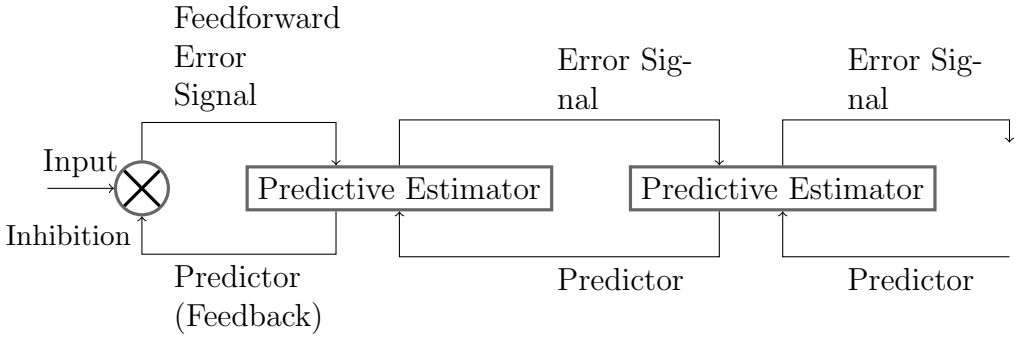


Figure 1: Hierarchical network for predictive coding [Rao and Ballard, 1999]

The lowest level in the hierarchical model network represents the image (Input). Each level in the model attempts to predict the responses at the next lower level via the feedback connections (Figure 1). The residual error between this prediction and the actual response from visual (Input) is then sent back to the higher level via the feedforward connections. This error signal is then used by the predictive estimator to correct the estimation of the input signal at each level and generate the next prediction.

Top-down information influences the lower-level estimates, and bottom-up information influences higher level estimates of the input signal because the prediction and the error correction cycles occur concurrently throughout the hierarchy.

2.2 PredNet

While many strides have been made in using deep learning algorithms to solve supervised learning tasks, the problem with unsupervised learning is that training on unlabeled examples to learn about the structure of a domain remains a difficult problem. This problem arises mainly on computer vision models that are typically trained on static images whereas, in the real world, visual objects are alive with movement, driven both by the self-motion of the viewer and the objects within a scene.

To address this, Bill Lotter, Gabriel Kreiman, and David Cox have developed a deep convolutional recurrent neural network they called PredNet [Lotter et al., 2016]. While building on the previous work in next-frame video

prediction [Softky, 1996, Palm, 2012, Goroshin et al., 2015, Mathieu et al., 2015b, Wang and Gupta, 2015], the architecture of PredNet is heavily inspired by the concept of predictive coding from neuroscience literature [Rao and Ballard, 1999]. PredNet attempts to continually predict the appearance of future video frames, using a deep, recurrent convolutional network with bottom-up and top-down connections. Top-down connections convey these predictions, which are compared against the actual inputs/observations in order to generate an error signal. This signal is then propagated back up the hierarchy, eventually leading to updated predictions.

Consistent with this idea, PredNet was able to demonstrate that prediction requires knowledge of object structures, i.e., the networks were able to learn the internal representation of an object. This prediction contributes to the further recognition and decoding of object parameters such as identity, view, and rotation speed.

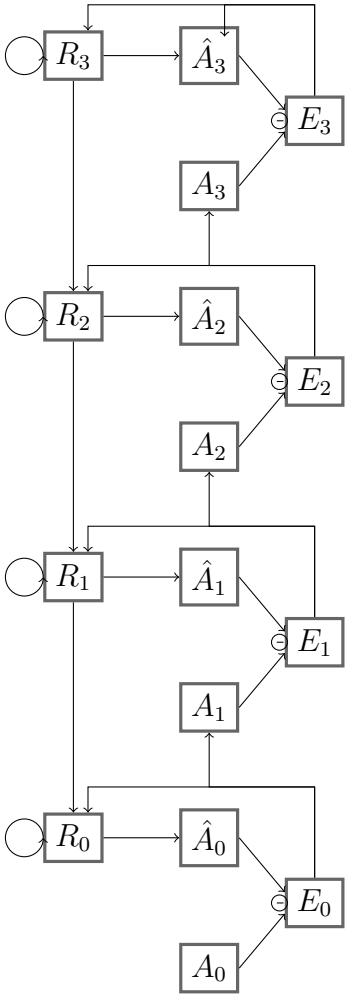


Figure 2: Information flow within layers

The PredNet architecture (Figure 2) consists of a series of repeating stacked modules. Each layer of PredNet stores a representation of its current state R_l, which is used to generate the prediction $\hat{A}_l$ for this layer. When an input A_l comes into a layer from the layer below, the difference E_l of A_l with the prediction $\hat{A}_l$, the error signal, is computed and passed on to the layer above. After the input has been processed by all layers, there is a top-down sequence of updates of the states of the layers, providing new predictions. Additionally, the number of channels increases and resolution decreases as the error is propagated upwards.

More specifically, the following equations are used to update the variables listed above at each iteration of PredNet (corresponding to reading one frame of the input video).

$$A_l^t = \begin{cases} x_t, & \text{if} l = 0 \\ MAXPOOL(ReLU(CONV(E_{l-1}^t))), & l > 0 \end{cases} \tag{1}$$

$$\hat{A}_l^t = ReLU(CONV(R_l^t)) \tag{2}$$

$$E_l^t = [ReLU(A_l^t - \hat{A}_l^t); ReLU(\hat{A}_l^t - A_l^t)] \tag{3}$$

$$R_l^t = CONVLSTM(E_l^{t-1}, R_l^{t-1}, UPSAMPLE(R_{l+1}^t)) \tag{4}$$

PredNet follows the above rules to generate predictions, accept inputs and calculate errors. Consider a sequence of images, x_t. The target for the lowest layer of input is set to the actual sequence itself. The targets for the higher levels A_l^t when $l > 0$ are computed by a convolution over the error units from the layer below, E_{l-1}^t followed by ReLU (Rectified Layer Unit) activation and max pooling. For representation, PredNet specifically uses convolutional LSTM units. R_l^t is updated according to R_l^{t-1}, E_l^t, R_{l+1}^t which is upsampled due to the pooling present in the feedforward path. The predictions $\hat{A}_l^t$, are made through the convolution of the R_l^t followed by ReLU activation. For the lowest layer, $\hat{A}_l^t$, predictions are set to the max pixel value. Finally, the error response, E_l^t, is calculated from the difference between $\hat{A}_l^t$ and A_l^t and is split into ReLU activated positive and negative prediction errors, which are concatenated. [Lotter et al., 2016].

2.3 PredNet and illusions of motion

A team from Japan studied motion detection of illusions on PredNet. Following the previous results of a random hyperparameter search for learning of a natural image sequence [Lotter et al., 2016], a four-layer model with 3X3 filter sizes for all convolutions and 3, 48, 96 and 192 stack sizes per layer was adopted. The models were retrained using videos from the First Person Social Interactions Dataset [Fathi et al., 2012]. The cameras were mounted on the caps and contained day-long videos of eight subjects at the Disney World Resort in Orlando, Florida. Model weights were optimized using the Adam Algorithm [Kingma and Ba, 2014] with default parameters. Optical flow vectors were calculated by the Lucas-Kanade method [Lucas and Kanade, 1981] using a customized python script.

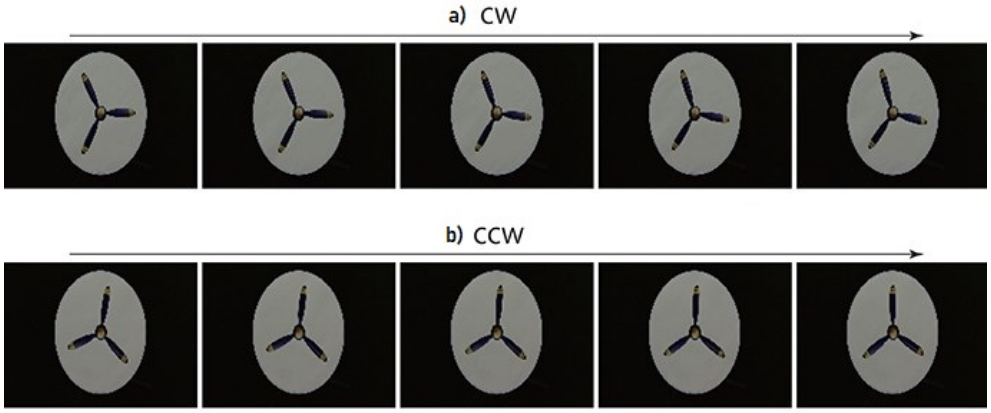


Figure 3: a) mirrored propeller CW direction, b) propeller in CCW direction [Watanabe et al., 2018]

Initially, to determine whether the trained networks of PredNet were capable of predicting the real motion of an object, videos of propellers rotating at 15 rpm were given as input(Figure 3). The motion vectors were generated using optical flow analysis, and it was observed that the predicted rotating motion was stronger in the Counter Clockwise (CCW) direction (Figure 4). From the results above, the trained DNNs appeared to respond sensitively to the subtle differences [Watanabe et al., 2018].

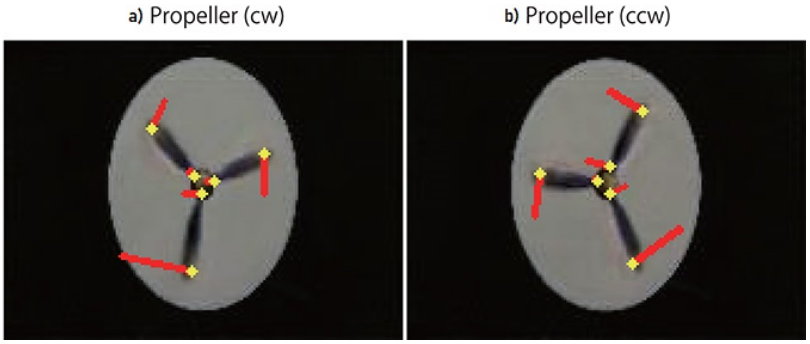


Figure 4: a) mirrored propeller with strong motion detection Clockwise (CW) direction, b) propeller with strong motion detection in CCW direction [Watanabe et al., 2018]

For the next prediction, 20 serial pictures of the rotating snake illusion

were input into the trained networks. The optical flow analysis revealed that the detected motion was in agreement with the rotation direction of the illusory motion perceived by humans.When negative control static images were input, rotational motions other than small optical flows were not predicted, which was similar to humans not perceiving illusionary motion at all (Figure 5).

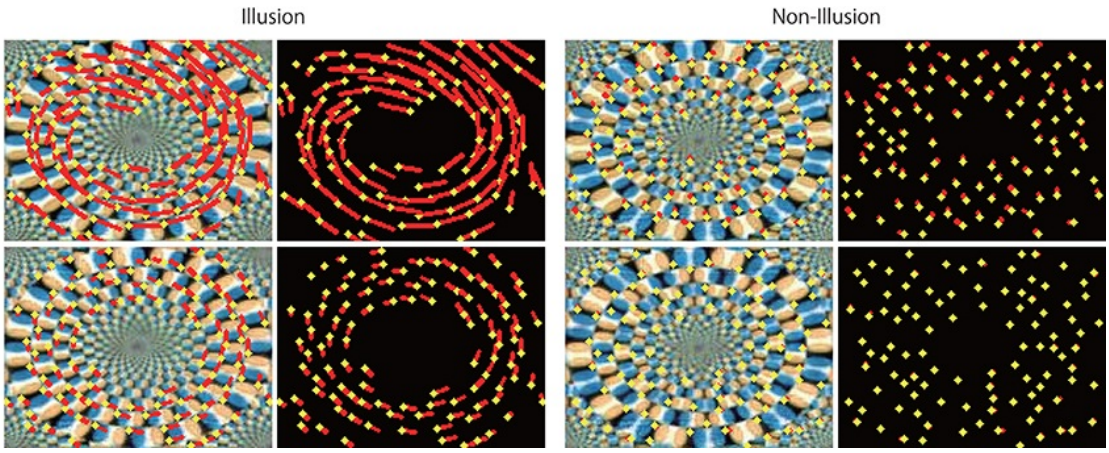


Figure 5: Optical flow vectors detected in the illusion. The left is a single ring of the rotating snake illusion, and the right is the negative control image.[Watanabe et al., 2018]

From the experiments above, it can be inferred that PredNet is capable of predicting motion in illusion and no motion in control images similar to motion perception of humans.

2.4 Schizophrenia and Models of perception

Schizophrenia is a mental disorder which commonly manifests by the state of psychosis, including symptoms such as hallucinations, delusions and aberrant thought processes and perception. There have been numerous attempts to understand what causes this debilitating disorder. In particular, the following models have been proposed to explain the distortions in perception in people with schizophrenia.

2.4.1 Precision Weighted Errors

One of the key problems with schizophrenia is a diminished ability to integrate new experiences with stored knowledge of previous experiences [Hemsley, 2005].

Fletcher and Frith (2009) attribute this to false prediction errors propagated upwards in the mind of someone with schizophrenia. Since the errors are false, any adjustments cannot fully resolve the problem, and as a result, prediction errors will be propagated even further up the system when calculated [Fletcher and Frith, 2009].

So, when a new experience (sensory input) is encountered, the prediction is weighted more than the sensory input due to false prediction errors, which might lead to positive symptoms of schizophrenia.

2.4.2 Disrupted efference copies

An efference copy is an internal copy of a movement producing signal generated by the motor system in humans. It is created by our movement and not those of other people, and, in healthy people, is automatically taken into account when interpreting sensory input. This is the reason why other people can tickle us (no efference copies are generated, so we are surprised by the sensation), but we cannot tickle ourselves (efference copies let us expect the sensation, avoiding surprise).

According to Pynn and DeSouza (2013), in people with schizophrenia, the efference copies fail to get created, and as a result, they cannot accurately differentiate between the cause of new sensory input as produced by actions of themselves versus the actions of others or their environment, which results in such common positive symptoms of schizophrenia as a feeling of being controlled by an outside force [Pynn and DeSouza, 2013].

2.4.3 Circular inference

Circular inference theory was proposed by Deneve and Jardri [Deneve and Jardri, 2016] to explain both the perceptual abnormalities and overconfidence observed in patients with schizophrenia. Circular inference is a phenomenon where bottom-up sensory evidence and top-down prior expectation reverberate through the network. It is based on the theory of belief propagation. Circular inference talks primarily about two loops - the climbing loop and descending loop are where the sensory evidence and prior expectations are overcounted, re-

spectively. The combination of both loops might lead to reduced sensitivity to perceptual illusions, which is a positive symptom of schizophrenia.

3 Our results

3.1 Methods

The version of PredNet which resulted from learning of a natural image sequence [Lotter et al., 2016] consisting of a four-layer model with 3x3 filter sizes for convolutions and stack sizes per layer of 3, 48, 96, and 192 was adopted for this study. Models which were pre-trained on the KITTI dataset [Geiger et al., 2013] were used for all the experiments (keras_models.zip). Since the models were trained on KITTI dataset, it should be noted that the model has an initial bias about everything moving towards the viewer.

The testing datasets that were used to conduct the experiments consisted of two types: real and illusion.

The real datasets were:

1. KITTI dataset.

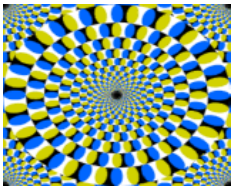
   10 frames of a driving car with a mounted camera on top.

2. Bike dataset.

   10 frames of a bicycle moving in a direction.

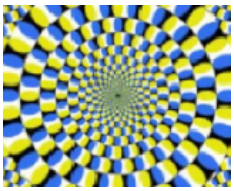
The illusion datasets were:

1. Snake dataset.



   10 frames of a cropped rotating snakes illusions (link).

2. Snake2 dataset.



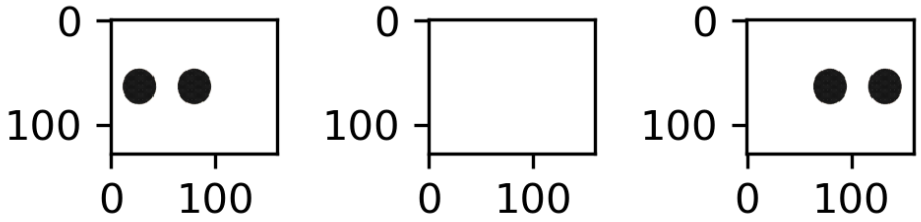
10 frames of a cropped rotating snakes illusions (link).

Other datasets not included in the paper but researched upon were:

1. Occlusion (Real)

10 frames of a still bicyclist and people moving in the background.

2. Ternus



3 frames of ternus (alternating three frames are input) illusion (link).

3. Lilac

11 frames of lilac illusion (link).

4. Hollow mask

10 frames of hollow mask illusion (link).

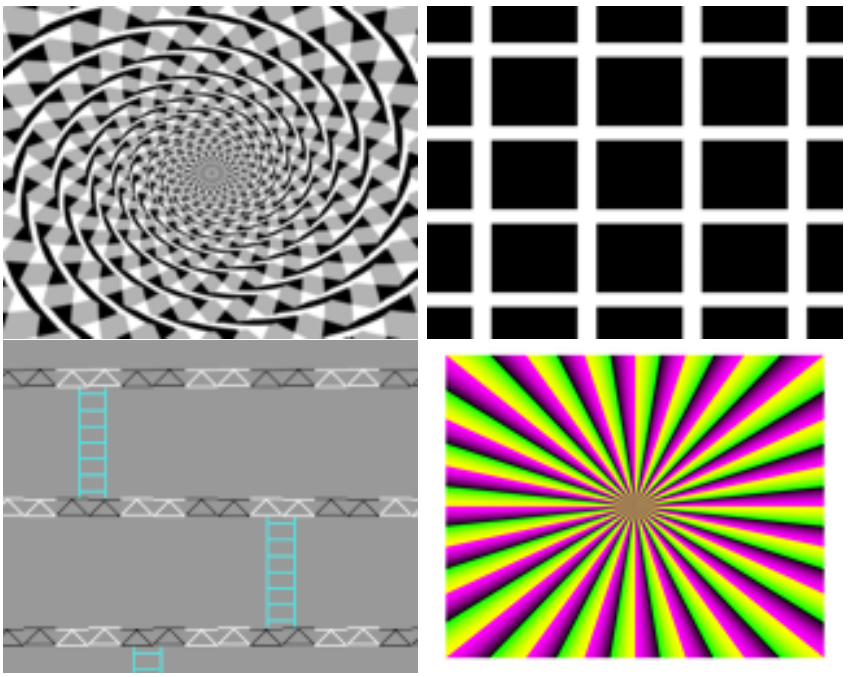


Figure 6: top-left: Fraser , top-right: Hermann, bottom-left: donkey kong, bottom-right: Spiral

5. Fraser

   10 frames of fraser illusion (link).

6. Hermann

   10 frames of hermann illusion (link).

7. Donkey kong

   10 frames of donkey kong illusion(link).

8. Spiral 10 frames of spiral illusion (link).

3.2 Experiments

In this section we talk about the different types of disruption that were experienced by the PredNet and the respective results. For all the datasets, the input images and the output predictions are 10 sequential frames.

3.2.1 Error Modification

The error modification disruption is inspired by the Precision Weighted Errors model of schizophrenia by Fletcher and Frith[Fletcher and Frith, 2009]; see section2.4.1 for an overview of this model. As explained in the architecture of the PredNet, the error response, E_l^t, is calculated from the difference between $\hat{A}_l^t$ and A_l^t and is split into ReLU activated positive and negative prediction errors, which are concatenated.

Since the model talks about false error propagation, we make the errors false by adding parameter changes to the calculation of $\hat{A}_l^t$ (prediction) and A_l^t (sensory) for both the positive and negative prediction errors.

So modifying 3rd rule of the PredNet results in,

$$E_l^t = [ReLU(\alpha A_l^t - \beta \hat{A}_l^t); ReLU(\beta \hat{A}_l^t - \alpha A_l^t)] \tag{5}$$

Equation 5 (above) has parameter changes to the prediction (A_l^t) and sensory (A_l^t) which are multiplied by α,β respectively.

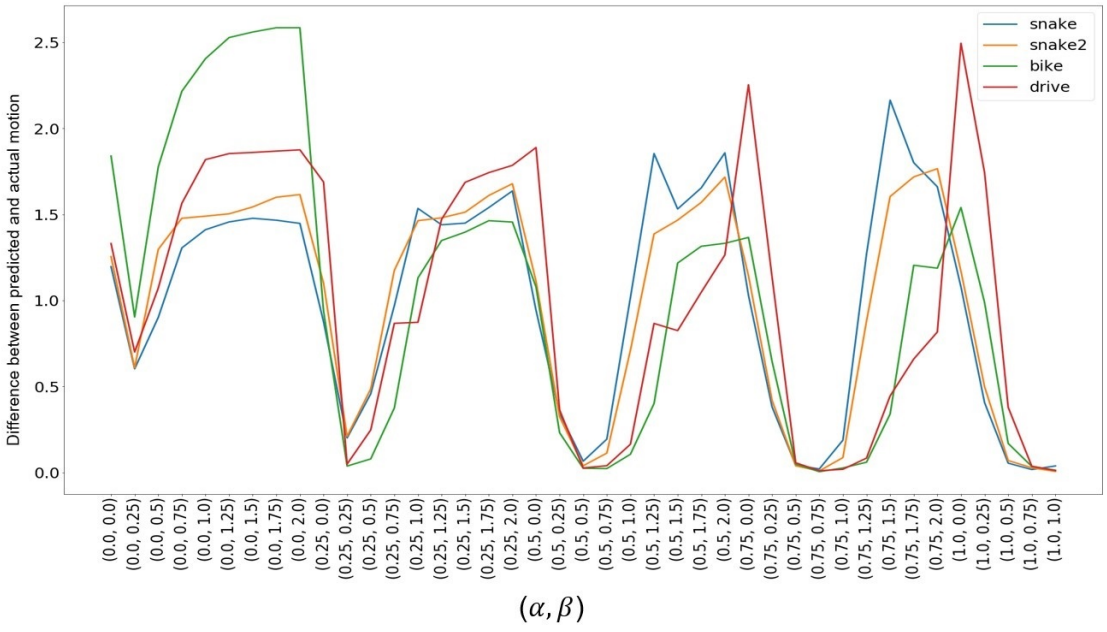


Figure 7: Effects of modifying the relative weight of positive and negative errors on the accuracy of motion prediction. Here, α=1=β corresponds to unmodified PredNet; decreasing α downweights the prediction, whereas decreasing β downweights the sensory input.

The model was run with parameter changes starting from 0 and ending at 2 with a step difference of 0.25 for both α and β. Higher disruption of β and α would mean the predictions and sensory are weighted more respectively.

The points (1,1) on the graph is the default prediction and the points which are closer to 0 on the y-axis are more accurate to predicted results. we can see from the graphs that when α is greater than β, the points are more closer to 0 on y-axis and when β is greater than α the points are closer to 1 on y-axis. While the model generated a lot of resulting images, the most interesting results were obtained when α is greater than β and when β is greater than α as is evident from the graphs.

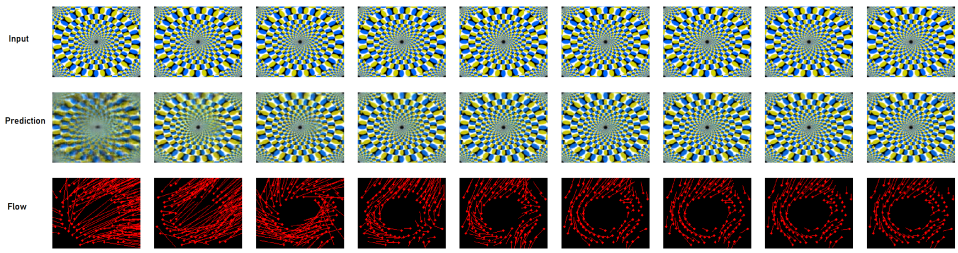


Figure 8: Snake, with unmodified parameters

Figure 8 shows the snake illusion being predicted on unmodified parameters by the PredNet. As the predictions are blurry in the beginning, the flow vectors don't have a proper sense of direction and as more frames are input the prediction gets better and a spiral CCW flow of directions can be seen in the final flow vectors.

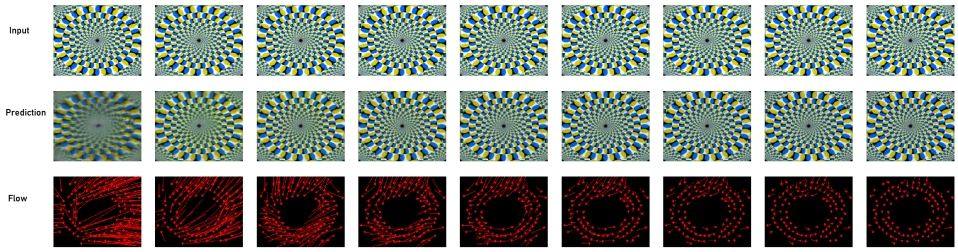


Figure 9: Snake, with $\alpha = 0.95$, $\beta = 0.85$

Figure 9 shows an example of snake when α is greater than β. As seen from the unmodified PredNet, the initial predictions frames don't have a sense of direction but as more frames are input into the PredNet, we can see that the motion vectors have less motion and by the last frame the motion vectors have less motion in comparison to unmodified predictions.

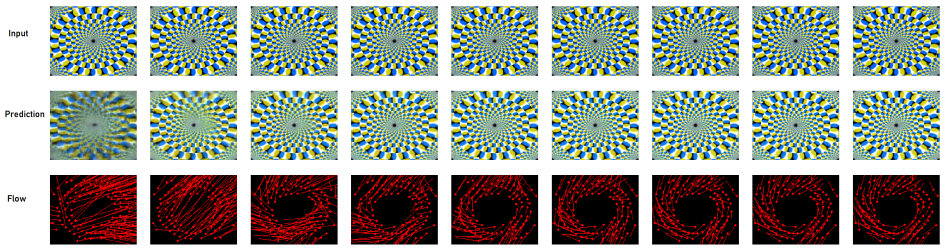


Figure 10: Snake, with $\alpha = 0.85$, $\beta = 0.95$

Figure 10 shows an example of snake when β is greater than α. As more frames are input into the PredNet, we can see that the predicted motion vectors have greater intensity of motion and by the last frame the motion vectors have amplified CCW flow vectors in comparison to unmodified predictions.

Let us see the results in the bike dataset,

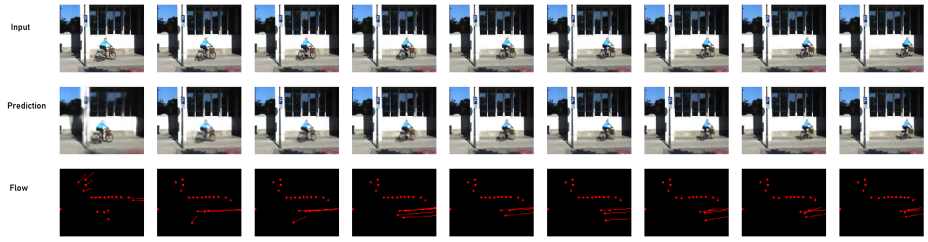


Figure 11: Bike, with unmodified parameters

Figure 11 shows the bike being predicted on unmodified parameters by the PredNet. As the predictions are blurry in the beginning, the flow vectors don't have a proper sense of direction and as more frames are input the prediction gets better and a directional flow of bike towards the right can be seen in the final flow vectors.

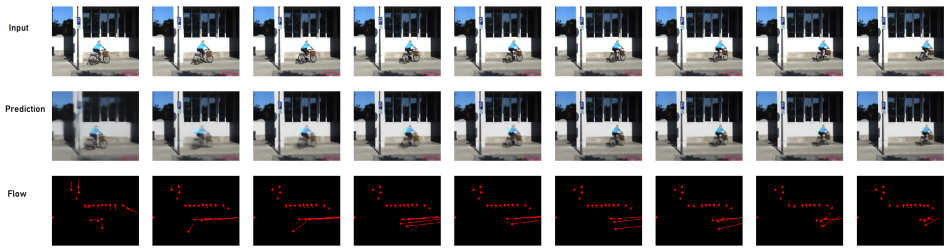


Figure 12: Bike, with $\alpha = 0.95$, $\beta = 0.85$

Figure 12 shows an example of bike when α is greater than β. The behaviour of bike is pretty similar to the behaviour of snake, the motion vectors have lesser flow vectors as they reach the final prediction.

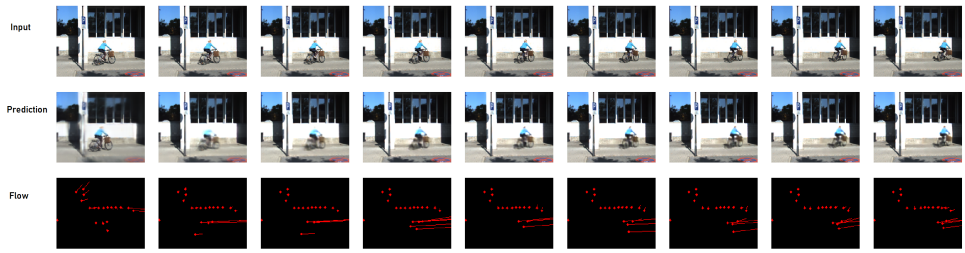


Figure 13: Bike, with $\alpha = 0.85$, $\beta = 0.95$

Figure 13 shows an example of bike when β is greater than α. Even in this case, the bike behaves pretty similar to the snake dataset, the motion vectors have amplified motions towards the direction of the bike, but there is also a peculiar object direction, if observed closely PredNet actually detects the windows moving towards upside direction. Nonetheless, the motion vectors are still amplified in the direction.

From the results above, it has lead us to believe that when α is greater than β i.e. when prediction is weighted more than sensory, then the illusions have lesser illusory motions and when β is greater than α the opposite holds true.

3.2.2 Forgetting the current prediction

Forgetting the current prediction disruption is inspired from the absence of the efference copies model (2.4.2). Since the model talks about forgetting

efference copies, we roughly translated it to forgetting predictions and so, we disrupt the R_l^{t-1} which is responsible for producing predictions at every layer.

The 4th rule of the PredNet was modified and the result rule is,

$$R_l^t = CONVLSTM(E_l^{t-1}, \alpha R_l^{t-1}, UPSAMPLE(R_{l+1}^t)) \tag{6}$$

The modified model was run with disruption to α starting from 0 and ending at 1. As the disruption increases from 0 to 1, the more the prediction remembers. So when $\alpha = 0$ it mostly forgets predictions, and when $\alpha = 1$ it mostly remembers predictions.

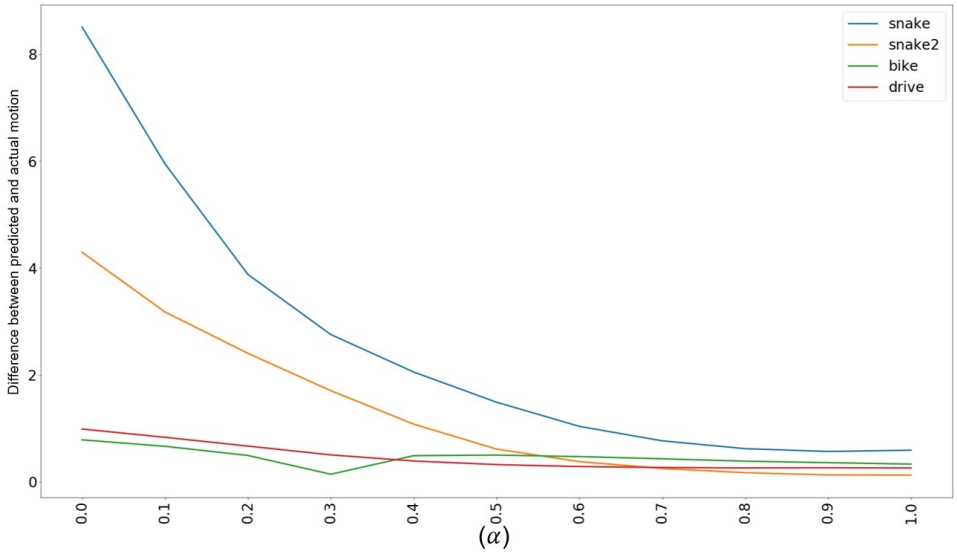


Figure 14: Effects of modifying the current prediction on the accuracy of the motion prediction. $\alpha = 1$ corresponds to unmodified PredNet. Decreasing α downweights current prediction

From the graph, it can be seen that as disruption moves from 1 to 0, the model is slowly forgetting the predictions resulting in greater difference between the predicted and actual motion.

Let us see the results when $\alpha = 0.25$ and $\alpha = 0.75$ which is the start of incremental and decremental spikes in the graphs respectively.

Refer to figure 8 for unmodified prediction of snake

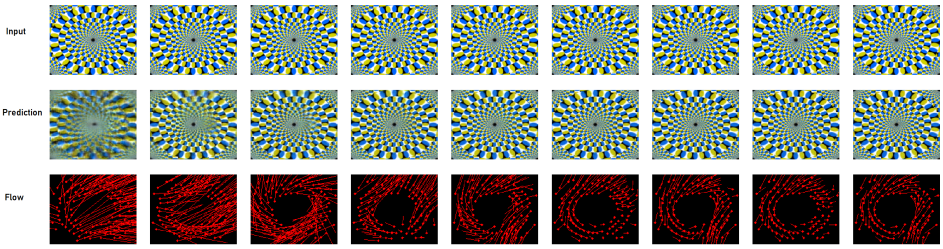


Figure 15: Snake, with $\alpha = 0.75$

An alternating behaviour can be seen from figure 15, PredNet generates a good prediction and forgets some of that prediction in the consecutive images. The final prediction doesn't have a proper sense of direction.

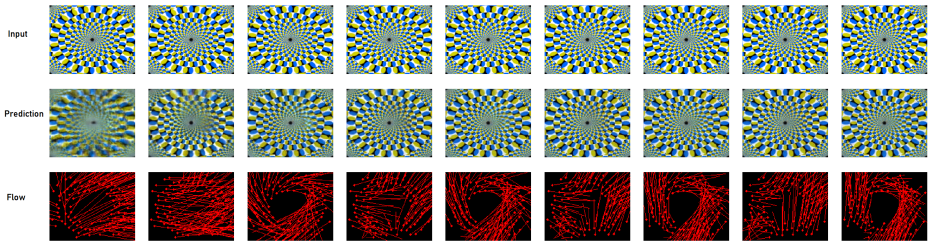


Figure 16: Snake, with $\alpha = 0.25$

An alternating behaviour similar to figure 15 can be seen in figure 16, PredNet generates a good prediction and forgets some of that prediction in the consecutive images. However, the final prediction has a better sense of direction comparatively.

From the results above, it has lead us to believe that as disruption (α) decreases from 1 to 0, it takes more time to generate better predictions.

3.2.3 Forgetting top-down predictions

Forgetting the top-down prediction disruption is also inspired from the absence of the efference copies model (2.4.2). Instead of disrupting the R_l^{t-1} which is responsible for producing predictions at every layer, we disrupt the top-down R_{l+1}^t.

The modified version of the 4th rule is,

$$R_l^t = CONVLSTM(E_l^{t-1}, R_l^{t-1}, UPSAMPLE(\alpha R_{l+1}^t)) \tag{7}$$

The modified model was run with disruption to α starting from 0 and ending at 1. From the graph(below), it can be seen that as disruption moves from 0 to 1, the model has peculiar trend on the snake at points where there's is a sudden flat (0.4 to 0.5 on x-axis) followed by decrease(0.5 to 1 on x-axis).

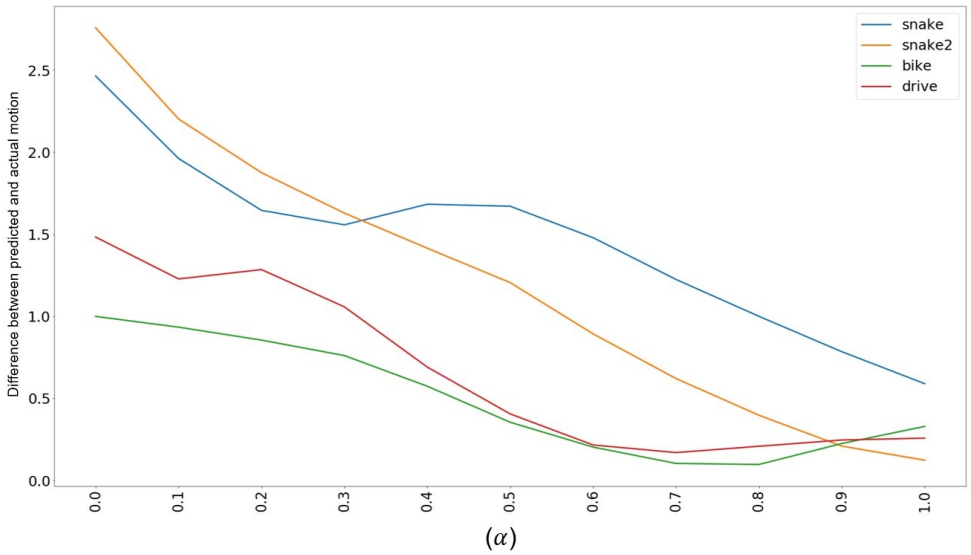


Figure 17: Effects of modifying the top-down prediction on the accuracy of the motion prediction. $\alpha = 1$ corresponds to unmodified PredNet. Decreasing α downweights top-down prediction

Let us see the results when $\alpha = 0$ which has a higher value of difference between predicted and actual motion and $\alpha = 0.5$ which is a peak point before decrease. Please refer to figure 8 for unmodified prediction of snake.

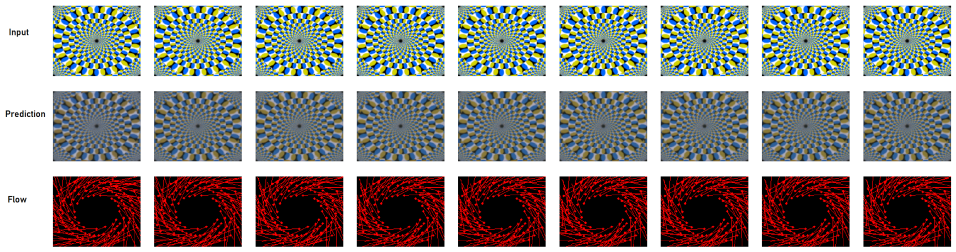


Figure 18: Snake, with $\alpha = 0$

In Figure 18, the prediction images are darker on every frame and the predictions have a very strong perceived flow from the beginning.

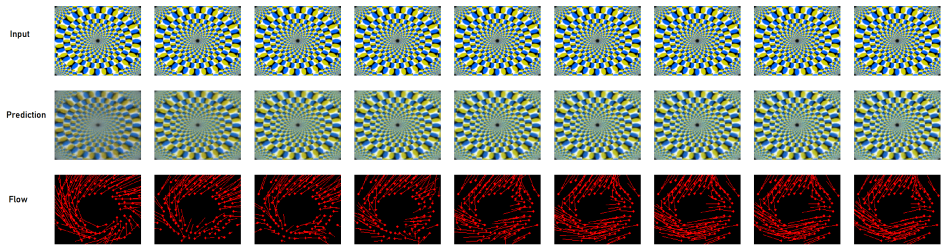


Figure 19: Snake, with $\alpha = 0.5$

In Figure 19, the prediction images are pretty similar to the unmodified snake, but as the predictions get better, the flow changes from very strong CCW movement to a CW movement and CCW movement on the left and right sides respectively.

From the results above, it has lead us to believe that as disruption (α) decreases from 1 to 0, the predictions tend to think they are better predictions from the beginning.

3.2.4 Mixing the current state with the sensory input

This disruption is inspired from the circular inference model by Deneve and Jardri[Deneve and Jardri, 2016], see 2.4.3 for an overview of this model. In this model, the sensory input "reverberates through the network", getting mixed with the prediction. To emulate that, we made changes to mix the predictions and sensory at every layer, by adding $\hat{A}_l$ to R_l.

The modified 4th rule is[1]

$$R_l^t = CONVLSTM(E_l^{t-1}, \alpha R_l^{t-1} + (1-\alpha)A_l^{t-1}, UPSAMPLE(R_{l+1}^t)) \quad (8)$$

The modified model was run with disruption to α starting from 0 and ending at 1. As disruption increases from 0 to 1, the snake2 has a gradual decrease of motion in the graph below.

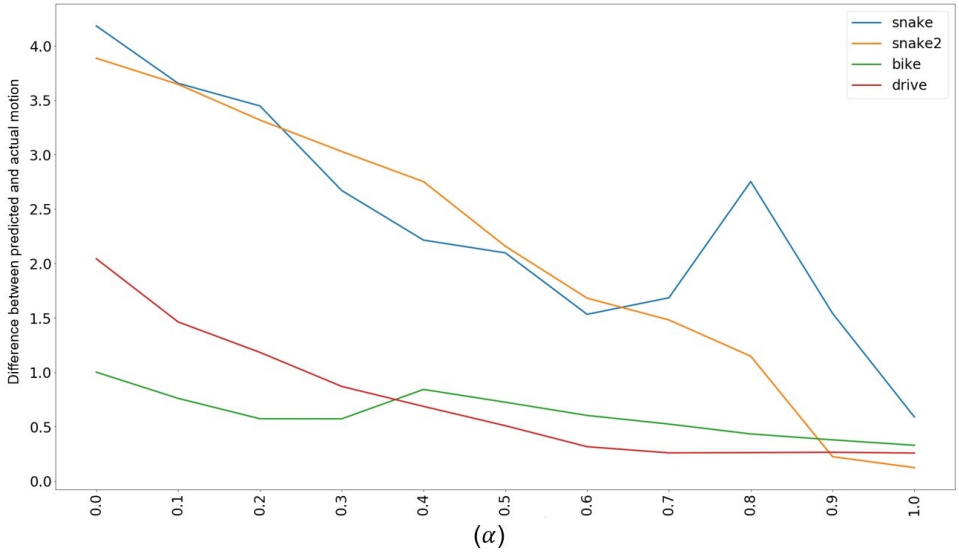


Figure 20: Effects of mixing the current prediction with sensory input on the accuracy of the motion prediction. $\alpha = 1$ corresponds to unmodified PredNet. Decreasing α downweights current prediction and increasing α downweights sensory input

Let us examine snake2 when $\alpha = 0.75$, $\alpha = 0.25$ and $\alpha = 1$ (unmodified). It should be noted that when disruption is high, the prediction is weighted more and when disruption is low, the input is weighted more.

[1]Note that since the actual prediction is $\hat{A}_l = ReLu(CONV(R_l))$, it would be more correct to add a deconvolved A_l rather than A_l as is. However, there seemed to be no mechanism in Keras to do this deconvolution at that point. Instead, we ran a separate experiment mixing R_l with $\hat{A}_l$, to estimate the effects of mixing in that extra convolution (and a ReLu).

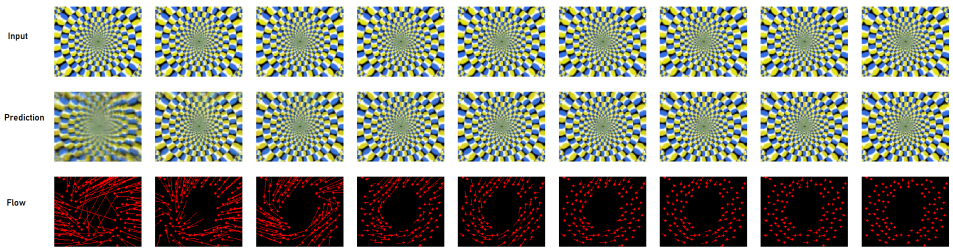


Figure 21: Snake2, with $\alpha = 1$, unmodified predictions

Figure 21, shows the unmodified snake2. As the predictions get better, the snake2 dataset has CCW flow vectors and as it nears the final predictions, the flow vectors have a more refined CCW directional flow.

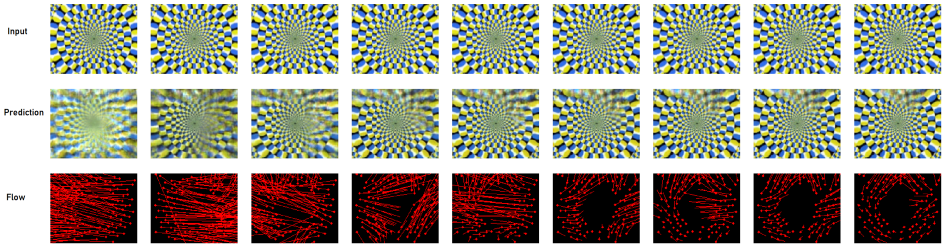


Figure 22: Snake2, with $\alpha = 0.75$

In Figure 22, when the prediction is weighted more, the flow vectors have a initial changed CW direction, but as the predictions progresses the directional flow is split between CW and CCW directions.

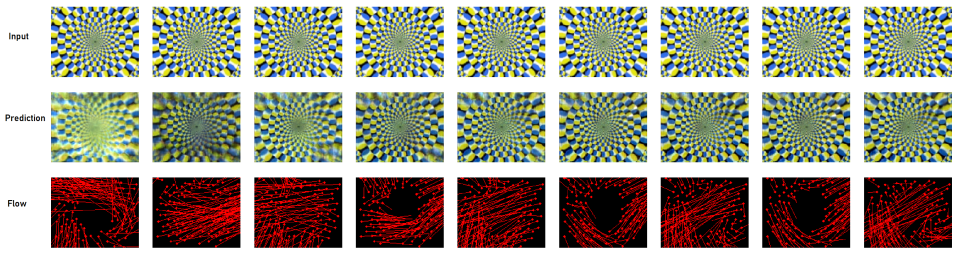


Figure 23: Snake2, with $\alpha = 0.25$

In Figure 23, when the sensory is weighted more, we can clearly see that

the flow vectors are in CW direction and as prediction progresses they get better and return to the original CCW direction with unstable flows.

From the results above, it can be said that as disruption is more on sensory i.e., as α decreases, it creates a strong opposite directional shift in the predictions.

3.2.5 Mixing top-down predictions with the sensory input

This disruption is also inspired from the circular inference model (2.4.3). In this disruption we made changes to mix the top-down predictions and sensory at every layer. We disrupted R_{l+1}^t which is the top-down prediction at every layer and also we disrupt A_l^{t-1} which is responsible for sensory input.

The 4th rule was modified which resulted in,

$$R_l^t = CONVLSTM(E_l^{t-1}, R_l^{t-1}, UPSAMPLE(\alpha R_{l+1}^t + (1-\alpha)A_{l+1}^t)) \quad (9)$$

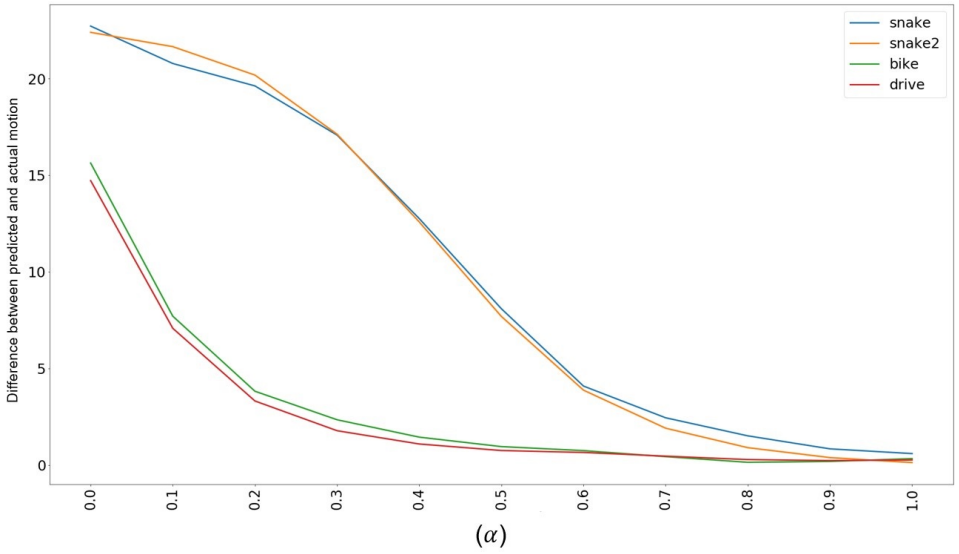


Figure 24: Effects of mixing the top-down prediction with sensory input on the accuracy of the motion prediction. $\alpha = 1$ corresponds to unmodified PredNet. Decreasing α downweights top-down prediction and increasing α downweights sensory input

The modified model was run with disruption to α starting from 0 and ending at 1 similar to the previous model. As disruption increases from 0 to 1, the pattern of two illusions dataset was very similar to each other, and curiously, it was a similar phenomena with the two real datasets.

So lets examine the snake and bike dataset with $\alpha = 0.25$ and $\alpha = 0.75$ respectively.

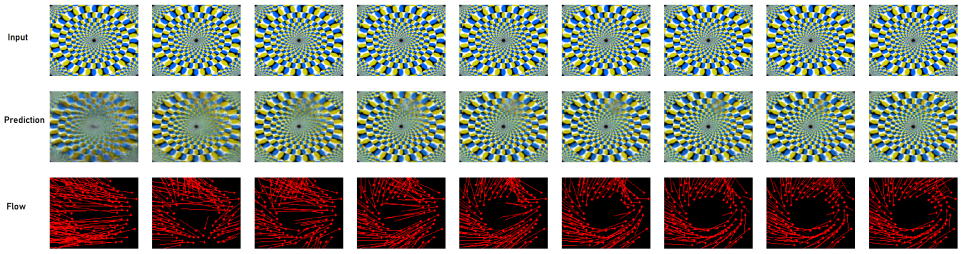


Figure 25: Snake, with $\alpha = 0.75$

In Figure 25, the prediction is weighted more than sensory. The predictions have initial directional changes but as more predictions are generated the flow is intensified. which can be expected due to the nature of the change which is similar to forgetting top-down prediction.

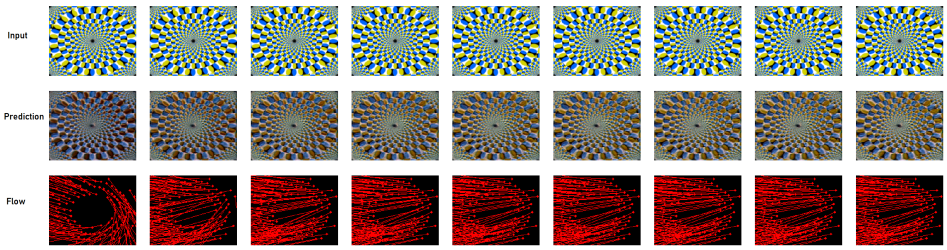


Figure 26: Snake, with $\alpha = 0.25$

In figure 26, the sensory is weighted more than prediction and it has some really peculiar results. The predictions have a color inversion from the initial frame and as more predictions are generated, PredNet is trying to match the color to the real image and hence the color keeps changing throughout. The flow vectors are always flying outwards in all the predictions.

Now lets take a look at the bike dataset,

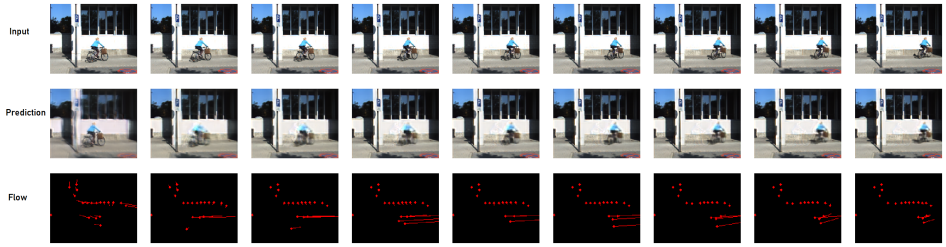


Figure 27: Snake, with $\alpha = 0.75$

In Figure 27, the prediction is weighted more than sensory. The predictions have initial directional changes but as more predictions are generated the flow is intensified and behaves in a similar way to forgetting top-down prediction.

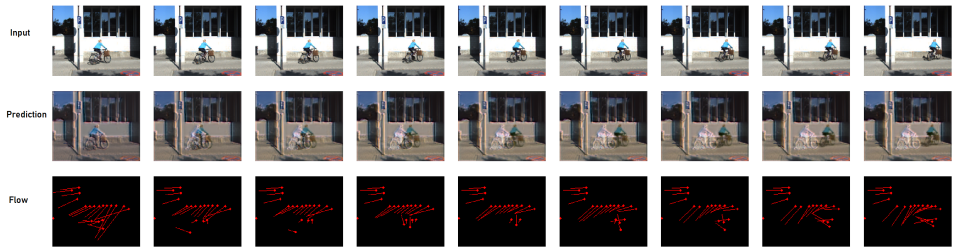


Figure 28: Snake, with $\alpha = 0.25$

In figure 27, the sensory is weighted more than prediction, the bike dataset showed some peculiar results as well. The predictions started to leave a ghost from the initial frame and as more predictions are generated, PredNet intensified the ghostly after image similar to the expectation of hallucination and delusion in schizophrenia. The flow vectors are trying to feature match two bicycles with the window movement in every frame.

To summarize, the results were different in real and illusion datasets. In real dataset when the sensory was weighted more than prediction we saw ghostly afterimages of the bike, whereas in the snake we saw color inversions.

3.2.6 Effects of extra convolution

This disruption is also inspired from the circular inference model (2.4.3). In this disruption we made changes to mix the predictions and sensory at

every layer. We disrupted R_l^{t-1} which is the prediction at every layer and also we disrupt $\hat{A}_l^{t-1}$ which is responsible for current prediction. The goal was to disrupt layer predictions with deconvolved current predictions. But as the implementation of keras restricts deconvolution without adding additional layer and any new layer would mean retraining the model, we made disruptions to the layer predictions and current predictions.

The 4th rule was modified which resulted in,

$$R_l^t = CONVLSTM(E_l^{t-1}, \alpha R_l^{t-1} + (1-\alpha)\hat{A}_l^{t-1}, UPSAMPLE(R_{l+1}^t)) \quad (10)$$

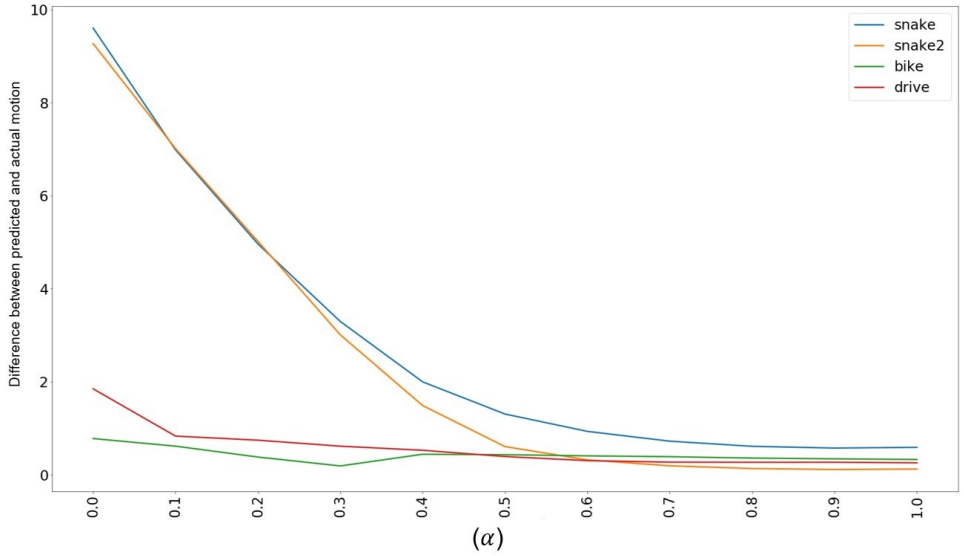


Figure 29: Effects of mixing the current prediction with layer prediction on the accuracy of the motion prediction. $\alpha = 1$ corresponds to unmodified PredNet. Decreasing α downweights current prediction and increasing α downweights layer prediction

The modified model was run with disruption to α starting from 0 and ending at 1 similar to the previous model. As disruption increases from 0 to 1, the pattern of two illusions dataset was very similar to each other, and it was a similar phenomena with the two real datasets.

So lets examine the snake2 with $\alpha = 0.25$ and $\alpha = 0.75$ respectively.

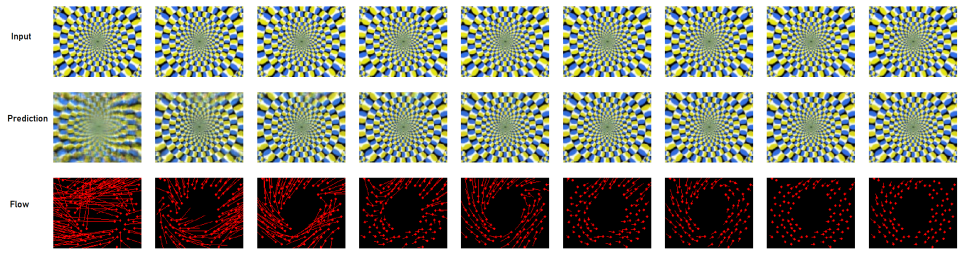


Figure 30: Snake2, with $\alpha = 0.75$

In Figure 30, the layer prediction is weighted more than current prediction. Initially, the predicted flows have lesser sense of direction but as more predictions are generated they get very similar to unmodified prediction. Refer to Figure 21 for unmodified snake2.

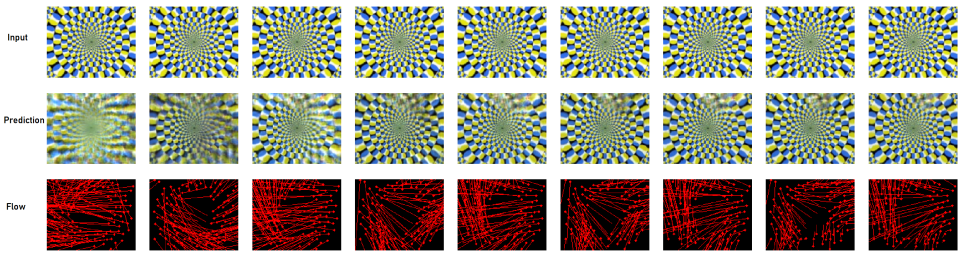


Figure 31: Snake2, with $\alpha = 0.25$

In Figure 31, the current prediction is weighted more than layer prediction which resulted in the outward directional flows till the final prediction. This might partially be due to the models being trained on the KITTI dataset.

To summarize, when layer prediction is more, the prediction frames are closer to the unmodified predictions but when current prediction is more, the directional vectors flow outwards corrupting the predictions.

3.3 Discussion

In this section we discuss the results produced from implementation of disruptions based on different models of schizophrenia.

PredNet perceived less illusory motion when sensory was down-weighted in the error modification disruption, which was inspired by the precision weighted error model.

Different results were acquired from forgetting current prediction, and top-down prediction disruptions, which were inspired by disrupted efference copies model. In the first disruption, we saw that when predictions were down-weighted, it took PredNet longer to make better predictions. However, in the next disruption, we saw that when predictions were down-weighted, PredNet started to fool itself by thinking the initial predictions were the best and repeated the same mistake over and over again.

The last group of disruptions was inspired by the circular inference model. Firstly, in mixing the current prediction with sensory, we saw that when sensory was weighted more than the prediction, PredNet created strong opposite directional motion in predicted frames. Secondly, In mixing the top-down prediction with sensory, we saw that when sensory was weighted more than the prediction, PredNet created a ghostly afterimage in bike and color inversion in the snake. As some of the effects could have been attributed to mixing the state before convolution into prediction with the sensory input, we also looked at the effects of mixing of the states with prediction; however, the opposite directional motion effect was absent in that scenario.

We find it remarkable that the PredNet reacted to some of the disruptions and produced results which could be reinterpreted in the terms of neuroscience models of schizophrenia, in particular that is PredNet showed lesser susceptibility to illusions and also showed evidence of hallucinations (ghostly afterimages). We hope that future work will help us better understand both the properties of predictive networks and the human brain.

4 Conclusions and future work

In this project, we have looked at how different disruptions, designed to mimic models of schizophrenia, affect PredNet's ability to predict real as well as illusory motion, and, generally, what are the effects of each type of disruption on PredNet's output. In particular, the only scenario in which we saw a reduction in illusory motion was disruption of the error calculation.

In the ongoing work, we are looking at additional datasets for both real and illusory motion, to better understand the effects of various disruptions and to try to glean what it is that allows PredNet to be fooled by illusions of motion. This has a potential to improve prediction ability of networks like PredNet in practical applications (for example, making it better in recognizing and predicting motion of bicycles, a notoriously tricky task for self-driving car systems). Going the other direction, from Computer Science to Neuroscience, what (if anything) do our experiments with PredNet tell us about models of schizophrenia, and, generally, theories of perception in mammalian brains? Can we conclude something about validity of theories of perception disruption which suggested modifications that did not affect illusory motion prediction? One challenge there is that in many of these models a crucial other parameter that is being disrupted is the estimation of the accuracy of the prediction; we are currently trying to understand how to get accuracy estimates from the PredNet.

Historically, Computer Science and Neuroscience have informed and enriched each other, and we hope that this project can contribute to the fruitful interaction between these two fields.

References

[Deneve and Jardri, 2016] Deneve, S. and Jardri, R. (2016). Circular inference: mistaken belief, misplaced trust. *Current Opinion in Behavioral Sciences*, 11:40–48.

[Fathi et al., 2012] Fathi, A., Hodgins, J. K., and Rehg, J. M. (2012). Social interactions: A first-person perspective. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1226–1233. IEEE.

[Fletcher and Frith, 2009] Fletcher, P. C. and Frith, C. D. (2009). Perceiving is believing: a bayesian approach to explaining the positive symptoms of schizophrenia. *Nature Reviews Neuroscience*, 10(1):48.

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

[Goroshin et al., 2015] Goroshin, R., Bruna, J., Tompson, J., Eigen, D., and LeCun, Y. (2015). Unsupervised learning of spatiotemporally coherent metrics. In *Proceedings of the IEEE international conference on computer vision*, pages 4086–4093.

[Hemsley, 2005] Hemsley, D. R. (2005). The development of a cognitive model of schizophrenia: placing it in context. *Neuroscience & Biobehavioral Reviews*, 29(6):977–988.

[Kingma and Ba, 2014] Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

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

[Lucas and Kanade, 1981] Lucas, B. D. and Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. In *IJCAI*.

[Mathieu et al., 2015a] Mathieu, M., Couprie, C., and LeCun, Y. (2015a). Deep multi-scale video prediction beyond mean square error. *arXiv preprint arXiv:1511.05440*.

[Mathieu et al., 2015b] Mathieu, M., Couprie, C., and LeCun, Y. (2015b). Deep multi-scale video prediction beyond mean square error. *arXiv preprint arXiv:1511.05440*.

[Oh et al., 2015] Oh, J., Guo, X., Lee, H., Lewis, R. L., and Singh, S. (2015). Action-conditional video prediction using deep networks in atari games. In *Advances in neural information processing systems*, pages 2863–2871.

[Palm, 2012] Palm, R. B. (2012). Prediction as a candidate for learning deep hierarchical models of data. *Technical University of Denmark*, 5.

[Pynn and DeSouza, 2013] Pynn, L. K. and DeSouza, J. F. (2013). The function of efference copy signals: implications for symptoms of schizophrenia. *Vision research*, 76:124–133.

[Rao and Ballard, 1999] Rao, R. P. and 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.

[Softky, 1996] Softky, W. R. (1996). Unsupervised pixel-prediction. In *Advances in neural information processing Systems*, pages 809–815.

[Villegas et al., 2017] Villegas, R., Yang, J., Hong, S., Lin, X., and Lee, H. (2017). Decomposing motion and content for natural video sequence prediction. *arXiv preprint arXiv:1706.08033*.

[Wang and Gupta, 2015] Wang, X. and Gupta, A. (2015). Unsupervised learning of visual representations using videos. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2794–2802.

[Watanabe et al., 2018] Watanabe, E., Kitaoka, A., Sakamoto, K., Yasugi, M., and Tanaka, K. (2018). Illusory motion reproduced by deep neural networks trained for prediction. *Frontiers in psychology*, 9:345.