

深層予測符号化ネットワークにおける 映像フレーム予測特性と時間過大視との関係に関する考察

1235063 汐見 興明 【コンピュータ構成学研究室】

A study on time overestimation illusion implied by video frame prediction characteristics of deep predictive coding network

1235063 Kohmei Shiomi 【Advanced Computer Engineering Lab.】

1 はじめに

近年、脳科学における大脳新皮質の理論仮説を実装し、ヒトが空間パターン認識において錯覚した際と同様の予測を行うDNN(PredNet)が提案された[1]。このPredNetの振る舞いを分析することによって、錯覚に影響する脳機能を推察できる。

時間錯覚の一種である時間過大視は、動的刺激と静的刺激を同時呈示した際、動的刺激の方が長く呈示されたと錯覚する現象である。この錯覚現象を模倣できるDNNの振る舞いを分析すれば、ヒトの時間知覚と視覚情報処理の関係性の解明に寄与できる可能性がある。

本研究では、PredNetの映像フレーム予測特性と時間過大視との関係性に着目した。学習方法・アーキテクチャの異なるDNNを対象に残像生成状況を分析することによって、時間知覚との関係性を考察した。

2 予測符号化が可能なDNN

予測符号化とは、脳科学における大脳新皮質の理論仮説である。ヒトが視覚で得た情報から未来の網膜像を予測し、実際に得られた網膜像との誤差情報を伝搬し学習する。予測符号化を実装したDNNには様々なものがある。

2.1 PredNet[1]

PredNetの構成を図1に示す。conv, Pool層にて予測画像と実際の画像の特徴を計算し、ReLU層にてそれらの誤差を計算する。計算した値をConvLSTM層にて受け取り、誤差が0になるよう学習する。ヒトが空間パターン認識において錯覚した際と同様の予測を行うPredNetにおいて、刺激の残像生成と時間過大視との関係性が生じるのではないかと？

2.2 CrevNet[2]

CrevNetは、2方向のオートエンコーダを搭載する可逆ネットワークであり、時系列データの相互間の時間・空間的特徴を抽出し学習する。PredNetに比べ、情報伝搬の相互性を実装しているため、PredNetとCrevNetの残像生成状況を比較することで、時間知覚との関係性を解明できる可能性がある。

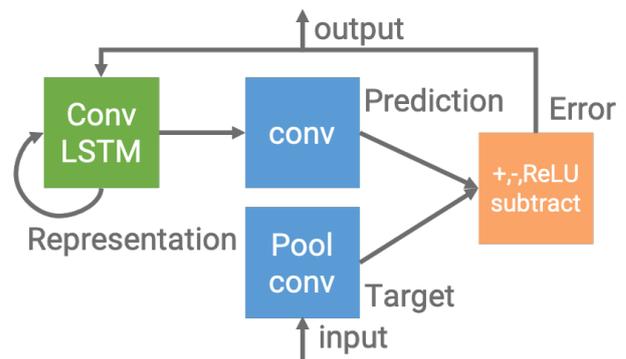


図1 PredNetの構成

3 時間過大視

時間過大視とは、時間錯覚の一種で、動的・静的刺激を順番に同じ時間呈示した際、動的刺激の方が長く呈示されたと錯覚する現象である。

この現象を用いて、視覚情報処理と時間知覚の関係性を調査した研究がある。ヒトは空間・時間周波数、速度の順に知覚する。そこで空間・時間周波数、速度を調整可能なガボールパッチを使用し、時間過大視の錯視量において空間・時間周波数、速度のどれが決定要因となるか実験した報告がある[3]。結果は、速度であった。つまり、視覚情報処理の後期にて速度を知覚するため、時間過大視、つまり時間知覚は視覚情報処理の後期以降で生じると言える。

本研究では、ヒトが空間パターン認識において錯覚した際と同様の予測を起こすPredNetと、予測符号化を実装し情報伝搬の相互性を実装するCrevNetを対象とし、先行研究と同様の実験を行い、DNNによる刺激の残像生成と空間・時間周波数、速度の変化による影響、残像生成状況と時間知覚の関係性について考察する。

4 実験

時間過大視を用いた視覚情報処理と時間知覚の関係性についての研究と同様の指標を用いて実験した[3]。テストデータセットの入力例は図2となる。テスト画像と推論画像の比較は、MSE・PSNR・SSIMを用いた。

3つの呈示時間条件についての結果をまとめるため、

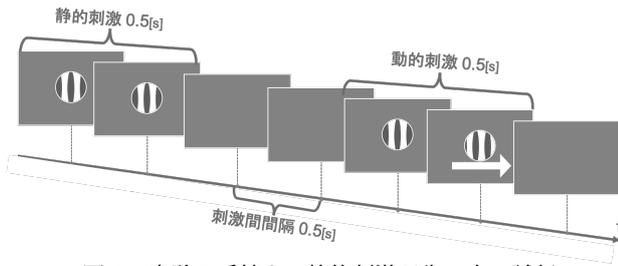


図2 実験の手続き。静的刺激が先に出る試行と動的刺激が先に出る試行を行った。

表1 動的刺激の呈示時間における誤差、一様刺激の呈示時間における誤差、それぞれを決定する要因と回帰曲線の傾き (決定係数)

入力刺激	決定要因	回帰直線の傾き
1. 動的, 一様	速度, 空間	1(0.72), -9(0.89)
2. 静的, 一様, 動的, 一様	速度, 空間	1(0.60), -8(0.90)
3. 動的, 一様, 静的, 一様	速度, 速度	0.7(0.77), -0.08(0.81)
4. 動的, 一様, 動的, 一様	速度, 空間	1(0.60), -9(0.94)

時系列の最後に入力される一様刺激の呈示時間におけるテスト画像と推論画像の誤差の平均値を求める。横軸を空間・時間周波数, 速度, 縦軸を誤差としてそれぞれプロットし, 回帰直線を計算し決定係数を求めた。決定係数が大きいほど, 誤差を決定する要因となる。結果を表1に示す。表1から, 動的刺激における誤差の決定要因は速度であり, 入力刺激3以外は動的刺激の速度の速さと誤差の大きさに正の相関がある。また, 一様刺激の呈示時間位における誤差を決定する要因は, 入力刺激3のみ速度, それ以外は全て空間周波数である。更に回帰曲の傾きから, 入力刺激3以外は動的刺激の空間周波数の大きさと一様刺激の呈示時間における誤差は大きさに負の相関がある。また, 横軸を空間・時間周波数, 速度, 縦軸を誤差としてそれぞれプロットしたグラフから, 静的・一様刺激の順序と入力刺激3におけるの一様刺激の呈示時間における誤差は, 後者が大きい。つまり入力刺激3の前半の刺激の入力による影響が見られる。更に, 一様刺激の呈示時間における誤差の大きさは入力刺激2 > 入力刺激4 > 入力刺激1となっており, 前半の刺激の入力による影響が見られる。静的・一様刺激と動的・一様刺激の一様刺激の呈示時間におけるの誤差は前者の方が大きい。

CrevNetにて, 入力刺激3の順序で入力した所, どの変数においても決定係数が0.08以下となった。

5 考察

動的刺激の空間周波数の大きさと一様刺激の呈示時間における誤差の大きさにおける負の相関

表1から決定要因が空間周波数における回帰曲線の傾きは全て負である。空間周波数が小さいほど, CNNの空間特徴の学習が速く, LSTMの学習が収束する。学習が進むほど, その後種類の異なる刺激が入力された

際, 刺激を推論できず, 誤差が大きくなる。

動的刺激の空間周波数が一様刺激の呈示時間における誤差の決定要因となる理由

この結果は, CNN・LSTMの学習速度による影響だと考えられる。動的・静的刺激または一様刺激の呈示時間における誤差は, 前者の方が大きい。つまり, CNNよりLSTMの学習速度が速い。そのため, 最後の一様刺激の呈示時間における誤差に最も影響を与えるのはCNNが処理する空間周波数となる。

刺激の入力順序と一様刺激の呈示時間における誤差の関係性

表1の入力刺激1・2・4の一様刺激の呈示時間における誤差の決定要因は空間周波数であり, 回帰直線の傾きの差異は1以下である, これは, 最後に入力される一様刺激の呈示時間における誤差に最も影響を与えるのが動的刺激であり, 後半に動的刺激が入力されるほど誤差の変化が同一となるためであると考えられる。

時間過大視との関係性

静的・一様刺激の順序, 動的・一様刺激の順序が後半の入力に見られる場合において, 前者の方が一様刺激の呈示時間における誤差が大きい。誤差の大きさと時間知覚の長さには負の相関がある場合, 動的・静的刺激を順番に同じ時間呈示した際, 動的刺激の方が長く呈示されたと錯覚する時間過大視が成り立つ。つまり, ヒトは刺激の誤差が小さいほど時間を長く知覚すると考えられる。

6 まとめ

本研究では, 錯視を起こす深層予測符号化ネットワークにおける映像フレーム予測特性と, 時間過大視の関係について考察した。PredNetでは, 一様刺激の呈示時間におけるテスト画像と推論画像間の誤差を説明する変数は, 空間周波数が最も適切であり, その結果と先行研究の結果の関連性から, 誤差が小さいほど時間過大視を知覚すると考えられる。また, 時系列データの相互間の時間・空間的特徴を学習することで, 時間過大視をより大きく知覚すると考えられる。

参考文献

- [1] Watanabe E, et al., "Illusory Motion Reproduced by Deep Neural Networks Trained for Prediction," *Frontiers in Psychology*, 2018;9:345. doi:10.3389/fpsyg.2018.00345, 2018.
- [2] W. Yu, et al., "Efficient and information-preserving future frame prediction and beyond," in *International Conference on Learning Representations, ICLR 2020*, 2020.
- [3] Kaneko S, et al., "Perceived duration of visual motion increases with speed," *Journal of Vision*, Vol.9, 14, 1-12, 2009.