

深層学習による画像処理を用いた人物の行動認識システムに関する研究

担当者：清水 健太郎

指導教員：長田 茂美 教授

1. まえがき

近年、人物の行動を認識するシステムの実現が期待されており、特に、深層学習を用いた画像処理による行動認識の手法が注目されている。この深層学習の一つに、映像の現時刻の画像から次時刻画像を予測する Deep Pred Net(Deep Predictive Coding Network)^[1]というニューラルネットワークがある。Deep Pred Net は予測画像と実際の画像との予測誤差を出力するため、予測誤差の値を用いれば学習させた映像と学習させていない映像を区別することができると考えられる。そこで、本研究では、Deep Pred Net を用いた人物の行動認識システムの実現を目指す。

2. システム概要

本研究では、Deep Pred Net(Deep Predictive Coding Network)というニューラルネットワークの人物の行動認識への適用可能性を検討する。図1に、Deep Pred Net の構造を示す。Deep Pred Net は、映像の現時刻の画像から次時刻画像の予測、生成が可能であり、予測画像と実際の画像との予測誤差を算出できる。具体的には、Convolutional LSTM で予測画像を生成し、予測画像と実際の画像との差分を予測誤差としている。

今回、検討した人物の行動認識システムは、Deep Pred Net が出力する予測誤差の二乗和をとり、映像内での予測誤差の二乗和の最大値、最小値、および平均値を行動の認識の閾値と比較することで、入力された映像が学習をさせた行動かそれ以外の行動かを判断し、人物の行動認識を行っている。図2に、今回開発した人物の行動認識システムの処理の流れを示す。

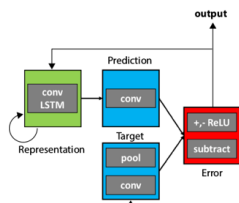


図1 Deep Pred Net の構造^[1]

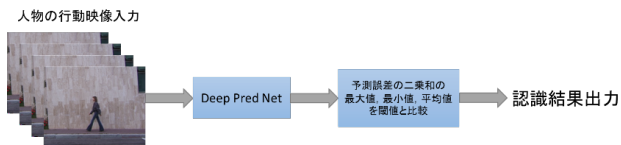


図2 開発した人物の行動認識システム

3. システム評価

人物の行動映像には、10種類の行動を9人の被験者が同一の場所で行っている映像データセットの1つであるワイツマンデータセットを用いた。Deep Pred Net は映像の次時刻画像を予測するため、映像のフレーム間の違いが大きければ人物の行動認識を行う際に有効なのではないかという仮説のもと、データセットの元々のフレームレートである25fpsの映像の場合と5fpsに縮めた映像の場合とで人物の行動認識を行い認識精度の比較評価を行った。

表1、表2に、25fpsの映像の認識結果と5fpsの映像の認識結果を示す。1つの行動ごとにネットワークの学習を行い、認識結果を出しているため、表は左が認識を行う行動、上が入力した行動映像、値がその映像を学習した行動と認識した数である。

表1 25fpsの映像の認識結果

	入力行動									
	bend	jack	jump	pjump	run	side	skip	wak	wave1	wave2
bend	2/4	2/4	4/4	4/4	3/4	3/4	4/4	4/4	4/4	3/4
jack	2/4	4/4	2/4	3/4	4/4	2/4	4/4	3/4	4/4	4/4
jump	2/4	4/4	4/4	4/4	4/4	4/4	4/4	4/4	4/4	4/4
pjump	2/4	4/4	4/4	4/4	3/4	4/4	4/4	4/4	4/4	4/4
run	2/4	4/4	2/4	3/4	4/4	2/4	4/4	4/4	3/4	3/4
side	3/4	4/4	4/4	4/4	4/4	4/4	4/4	4/4	4/4	4/4
skip	2/4	4/4	2/4	3/4	4/4	3/4	4/4	4/4	4/4	4/4
wak	1/4	3/4	3/4	3/4	3/4	3/4	4/4	4/4	3/4	3/4
wave1	4/4	1/4	4/4	4/4	2/4	3/4	4/4	4/4	4/4	4/4
wave2	4/4	2/4	4/4	4/4	3/4	4/4	4/4	4/4	4/4	4/4

表2 5fpsの映像の認識結果

	入力行動									
	bend	jack	jump	pjump	run	side	skip	wak	wave1	wave2
bend	3/4	1/4	1/4	3/4	1/4	1/4	1/4	1/4	4/4	3/4
jack	2/4	4/4	1/4	3/4	3/4	1/4	3/4	3/4	3/4	4/4
jump	3/4	1/4	3/4	3/4	4/4	3/4	3/4	4/4	2/4	2/4
pjump	3/4	1/4	1/4	4/4	0/4	0/4	0/4	0/4	4/4	3/4
run	2/4	4/4	3/4	4/4	3/4	3/4	4/4	4/4	4/4	3/4
side	3/4	4/4	4/4	4/4	4/4	4/4	4/4	4/4	4/4	4/4
skip	2/4	4/4	2/4	3/4	4/4	2/4	4/4	4/4	4/4	4/4
wak	2/4	4/4	2/4	3/4	3/4	2/4	4/4	3/4	4/4	4/4
wave1	4/4	1/4	1/4	3/4	4/4	1/4	4/4	1/4	4/4	2/4
wave2	4/4	2/4	2/4	4/4	1/4	0/4	1/4	1/4	4/4	4/4

25fpsの映像は、誤認識が多く行動の認識がうまくできていない。その理由は、映像のフレーム間の動き情報が小さいことで、行動の特徴を学習できなかったことが原因である。5fpsの映像は、その場で行う行動のネットワークで移動する行動との誤認識が少なくなっており、25fpsの映像と比較して行動の認識ができていない。移動する行動のネットワークで誤認識が多い理由は、撮影場所は同じであるが撮影角度が違う映像があり、背景の違いによって出力される予測誤差の値が異なり、行動の認識の閾値を決める際に閾値が高くなってしまったことが原因であった。このことから、Deep Pred Net は映像の背景に大きく影響を受けてしまい汎化性能が高くないことが分かる。そのため、撮影条件が同一の映像ごとに学習を行えば行動の認識を行うことができるであろう。また、その場で行う行動の予測誤差の値を見てみるとそれぞれのネットワークごとに似た値をとっており、現状のシステムでは違いを認識することが困難であることが分かった。

4. システムの適用法についての考察

Deep Pred Net を用いた人物の行動認識システムは撮影条件が同一の映像を用いれば、その場で行う行動と移動する行動の認識が可能である。そのため、エレベータなどの撮影条件が変化せず、人物が静止していることが一般的な状況下で人物が激しく動くような異常行動を行った際に異常行動を検知するシステムに使用することができる。

また、システムの改善策としてDeep Pred Net を階層化し、高次の層に低次の層で算出した予測誤差の入力を行い、単層では予測できなかった人物の行動の特徴を学習させ、より詳細に人物の行動の特徴を捉えることができるようにすることで、現状のシステムでは違いを認識できていない詳細な行動の認識を可能にすることが挙げられる。

5. むすび

本研究では、Deep Pred Net を用いた人物の行動認識システムの試作を行い、評価実験に基づき、システムの適用法について考察した。今後の課題として、Deep Pred Net を階層化し、現状のシステムでは違いを認識できていない詳細な行動の認識を可能にすることが挙げられる。

参考文献

- [1] William Lotter, Gabriel Kreiman, David Cox: “Deep Predictive Coding Networks for Video Prediction and Unsupervised Learning,” arXiv:1605.08104, 2016.