ISS-SP-037

双方向 LSTM を用いた 画像生成によるフレーム補間

長崎　大†　　宮崎　智††　　菅谷　至寛††　　大町　真一郎††
† 東北大学工学部電気情報物理工学科　　†† 東北大学工学研究科

1. はじめに

深層学習を用いた画像生成は盛んに研究されていて，これを活用することにより動画像における情報量の削減を図ることができると考えられる．本研究では，画像生成技術を用いて動画像における時間軸方向の補間をすることにより，補間するフレームの分だけ動画像の容量を削減することを目指す．

2. 関連研究

PredNet[1]は入力された直前までの動画像からその次のフレームの画像を予測するネットワークである．予測されたフレームとそのフレームの GroundTruth との Loss を計算し，そこから次のフレーム予測をするという構造になっている．またフレームを予測する際に LSTM[2]を用いることで直前までに入力されたフレームの情報を考慮した出力をなすようになっている．本研究ではこのネットワークを動画像における時間軸方向の補間に応用する．

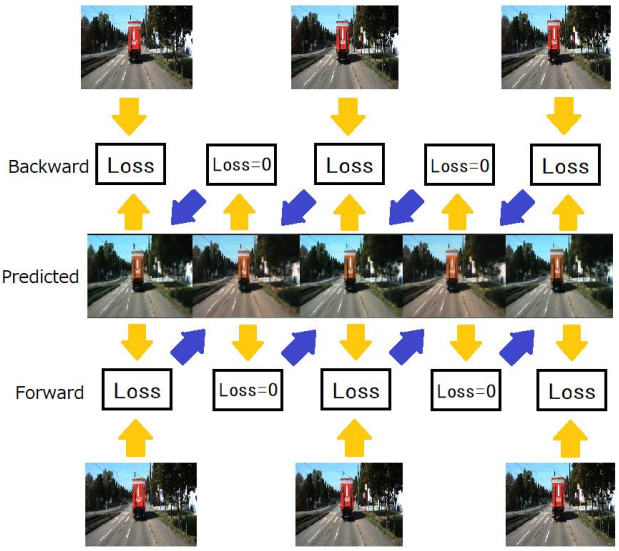

図 1. 提案手法のネットワークの動作

3. 提案手法

提案手法では，PredNet[1]を図 1 のように動作させている．双方向からネットワークに入力を行い，得られた二枚の画像を畳み込むことによって出力をしている．PredNet[1]をフレーム補間に用いる場合，予測されたフレームと実際のフレームの Loss から次のフレームを計算する部分において問題が生じる（補間するフレームは入力画像がないため）．そのため，提案手法では補間するフレームの部分に関しては Loss を「0」としてネットワークを動作させることにより，動画の補間を実現している．さらに図 1 のように今回の問題では未来の情報も既知であるため，LSTM を Bidirectional に用いることで未来方向と過去方向の両方の情報から補間フレームの導出をしている．

4. 実験結果

実験には KITTI データセットの画像を使用し，学習には 3200×40 枚，テストには 1100×40 枚を用いた．

実験結果は図 2 のようになった．補間はできたものの，補間されたフレームの PSNR は約 15 と非常に低い結果となった．出力された画像を見ると，全体的にフレームを補間しようとする動きが見られたが，実際の画像と見比べてみると非常に異なるものとなってしまった．　さらに，全体的に補間されたフレームは実際のフレームに対して非常にぼやけたものとなっていて，実際に動画にするとその画質の差で非常に違和感がある動画になってしまった．

原因としては，十分な学習が行えていないということがあげられる．さらに現状のネットワークでは最終的には単純な畳み込みによって出力されたものなので，画質には限界があると考えられる．

図 2. 動画像補間の結果

5. まとめ

本稿では，双方向 LSTM を用いた画像生成によるフレーム補間を提案した．しかし，結果からわかるように大きな動きの予測や画質に問題を抱えている．画質の問題を解決するためには，現状のネットワークでは非常に厳しいと考えられるため，GAN などの画像生成に向いているネットワークの導入なども考えていきたい．

参考文献

[1] William Lotter, Gabriel Kreiman and David Cox, Deep Predictive Coding Networks for Video Prediction and Unsupervised Learning, ICLR 2017

[2] Sepp Hochreiter and Jurgen Schmidhuber. Long short-term memory. Neural Computation, 1997