ST-ABN を用いた動画像認識における Temporal attention の修正による効果の調査

EP18048 坂場紀香　　指導教授：山下隆義

1. はじめに

動画像認識における視覚的説明の手法に Spatio-Temporal Attention Branch Network (ST-ABN) [1] がある．ST-ABN は，ネットワークの推論時における空間情報と時間情報に対する重要度を獲得し，認識処理に応用することで認識性能の向上を実現した．しかしながら，獲得した時間情報に対する重要度と注視するべき動作区間にズレが発生した場合，誤認識を誘発することがある．そこで本研究では，時間情報の重要度を修正することが精度向上に寄与するか調査を行う．

2. 関連研究

ST-ABN は，重要な空間情報と時間情報を同時に獲得可能な動画像認識の手法である．ST-ABN は，図 1 に示すように Feature extractor，Spatio-Temporal (ST) Attention branch，Perception branch の 3 つのモジュールで構築される．Feature extractor は，3 次元の複数の畳み込み層が含まれており，入力から特徴マップを獲得する．ST Attention branch は，空間情報に対する重要度を示す Spatial attention と時間情報に対する重要度を示す Temporal attention を獲得する．Perception branch は，Attention 機構により Spatial attention と Temporal attention を特徴マップに重み付けした特徴マップを入力し，各クラス確率を出力する．

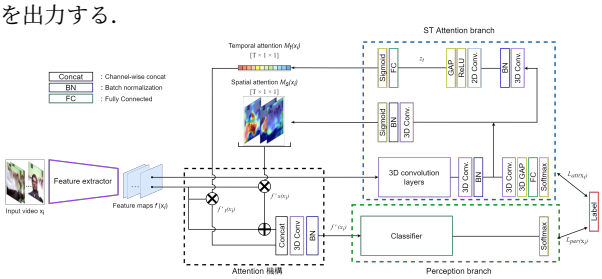

図 1：ST-ABN のネットワーク構造

3. 提案手法

3.1. Temporal Attenion 修正ツール

ST-ABN の Temporal Attention を人手により修正するために，図 3 に示す修正ツールを作成した．ツールの画面左上で修正する動画を選択し，選択した動画は上段中央で再生ができる．修正前と修正後の推論結果と順位を画面右上で確認ができる．また，Temporal attention の修正は下段の画像列に対して行う．

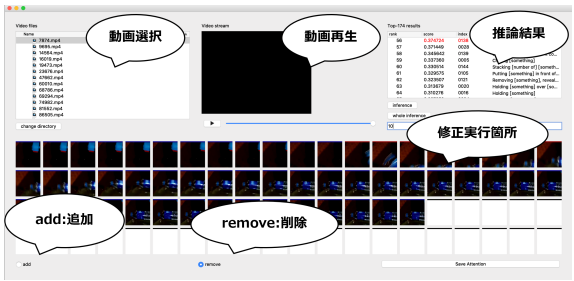

図 2：Temporal attention の修正ツール

3.2. 修正の流れ

修正の流れを以下に述べる．まず初めに，学習済みの ST-ABN を用いて動画を評価する．推論結果もとに修正する動画を選択する．下段に表示された各フレームの画像とその Tempotal attention を確認し，修正を行う．修正は注視すべき区間の Temporal attention が小さい場合 add を選択後クリックして Temporal attention を追加する．一方，注視すべきではない区間の Temporal attention が大きい場合 remove を選択後クリックして Temporal attention を削除する．実際の修正例を図 3 に示す．図 3 に示すように，クラスの識別に重要だと思われるフレームを 1（赤），その他のフレームを 0（青）とし，Temporal attention を修正する．そして，修正した Temporal attention を Attention 機構に用いて推論する．

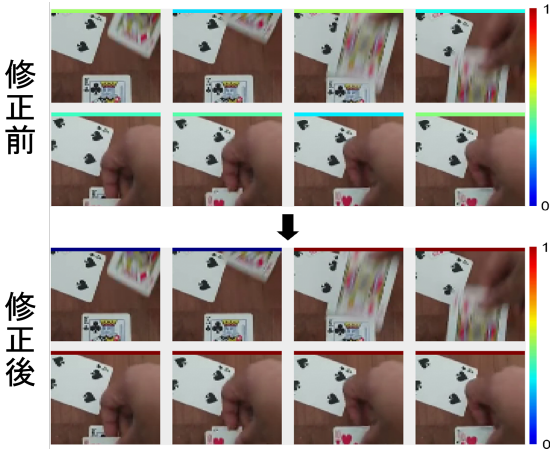

図 3：Temporal attention の修正ツール

4. 評価実験

4.1. 実験概要

Temporal attention 修正の有効性を評価するために，修正前と修正後の認識精度を比較する．本実験では，Something-Somrthing データセット V2 の中で，誤認識率が高い “Piling [something] up” クラスの動画を対象とし，Temporal Attention を修正する．790 本の内，学習済みモデルで誤認識しているかつ評価に必要なフレーム数が 32 以上の動画 394 本を修正する．評価に用いる ST-ABN のベースネットワークには，3D ResNet-50 を使用し，入力フレーム数は 32 フレームとする．

4.2. 評価結果

Temporal attention 修正前と修正後の結果を表 1 に示す．表 1 より Temporal attention を修正することによって，Top-1 になった動画が 2 本増えた．また，Top-5 になった動画は 16 本増え，認識精度が向上していることが確認できる．そのため，時間方向の重要度の修正は有効であることが確認できる．

表 1：評価結果 [%]

Top-1		Top-5	
修正前	修正後	修正前	修正後
48.10 (380/790)	**48.35 (382/790)**	52.78 (417/790)	**54.81 (433/790)**

5. おわりに

本研究では，ST-ABN における Temporal attention の修正による有効性を調査した．評価実験では，Temporal attention の修正により，認識精度の向上を確認した．今後は，修正した Temporal attention を用いた再学習法を検討する．

参考文献

[1] 三津原将弘 等，“Spatio-Temporal Attention Branch Network による時空間情報を考慮した視覚的説明”，MIRU, 2021.