卒 業 論 文 概 要 書

Summary of Bachelor's Thesis

Date of submission: 02/01/2022

学科名 Department	情報通信	氏名 Name	足立 翔平	指 導 教 員 Advisor	渡辺 裕 ㊞
研究指導名 Research guidance	オーディオビジュアル情報処理研究	学籍番号 Student ID number	CD 1w183004-3		
研究題目 Title	スポーツにおける動作品質評価の性能改善に向けた特性解析 Characteristic Analysis for Performance Improvement of Action Quality Assessment in Sports				

1 まえがき

スポーツ競技には，採点評価の基準から三つの分類がある．一つ目は対人競技，二つ目は競走，三つ目は採点競技である[1]．このうち採点競技は，人間が主観を用いて採点を行う特性上，採点基準の曖昧さや審査の公平性という点において課題が挙げられている．これらの課題に対して，熟練の審査員の導入や複数国の審査員の動員といった改善手法が取られているが，採点の判定に対する問題は依然として解決には至っていない[2]．そこで，近年においては上記課題の解決のために技の採点評価に機械学習を取り込む動きが注目されている．

人間の動作の評価を行う技術として Action Quality Assessment (以降 AQA と記述する)がある．また，この分野の最先端手法の一つとしてマルチタスク化を施した Multi Task Learning-AQA (以降 MTL-AQA とする)が存在する．C3D-AVG ネットワークとして提案されている MTL-AQA はスポーツの採点評価に用いることが可能であるが，公式な大会で用いるには現状の精度では不十分である．

そこで本研究では, C3D-AVG ネットワークのマルチタスク部分が個々にどれほどの性能への影響力を持っているかを解析し，MTL-AQA の性能向上への足掛かりとする．

2 AQA

AQA は，人間の動作を表す映像を入力として，その動作のスコアを判定する技術である．入力された映像を 3 次元畳み込みニューラルネットワークに通して特徴量を抽出し，得られた特徴量およびアノテーションされたスコアの正解ラベルに基づいて動作の品質評価を学習する．

3 C3D-AVG ネットワーク

C3D-AVG ネットワークは Paritosh Parmar らが MTL-AQA として提案した AQA のネットワークモデルである[3]．従来の AQA タスクに新たに動作分類のタスクと動作の説明の文章であるキャプション生成のタスクを追加し，精度を向上させたマルチモーダルなネットワークモデルとなっている．モデルの概要を図 1 に示す．

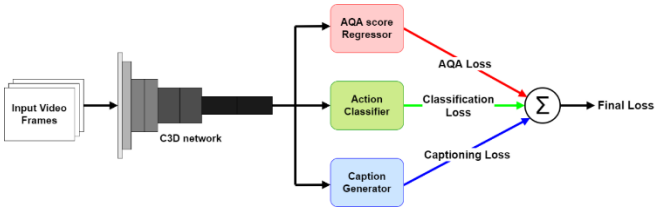

図 1 C3D-AVG ネットワークの構造

AQA タスクの Loss を L_{AQA}，動作分類タスクの Loss をL_{Cls}，キャプション生成タスクの Loss をL_{Cap} と表した際の最終損失L_{Final}を式(1)に示す．

$$L_{Final} = \alpha L_{AQA} + \beta L_{Cls} + \gamma L_{Cap} \quad (1)$$

ここで，α，β，γは各損失の重み係数である．

4 提案手法

本研究では, C3D-AVG ネットワークが持つ三つのタスクが，品質評価の精度にそれぞれどの程度の影響力を持っているかを調査する．その方法として，まず式(1)に示した最終 Loss の損失関数における各タスクの重み係数を変化させ, C3D-AVG ネットワークの精度との関係をグラフ化し，精度が最大となる場合の各タスクの重みを読み取る．その後，各損失の範囲を揃えるために 3 タスクの損失が取りうる値

の範囲を測定する．最後に，各損失の範囲を考慮した重み係数の比率をもとに，最終的な精度の向上に対する各タスクの影響力を推定する．本提案では C3D-AVG ネットワークの精度を評価する指標としてスピアマンの順位相関係数を用いる．

5 実験

5.1 C3D-AVG の精度最大となる重み測定

データセットとして，1412 個のダイビング動画に動作認識のラベル及びキャプションのラベルが対応付けられた MTL-AQA データセット[3]を用いた．

各タスクの重み係数α，β，γを$\frac{\alpha}{\beta}, 1, \frac{\gamma}{\beta}$として正規化し，$\frac{\alpha}{\beta}, \frac{\gamma}{\beta}$を変化させたことによる C3D-AVG ネットワークの精度の値を図 2 のグラフに示す．

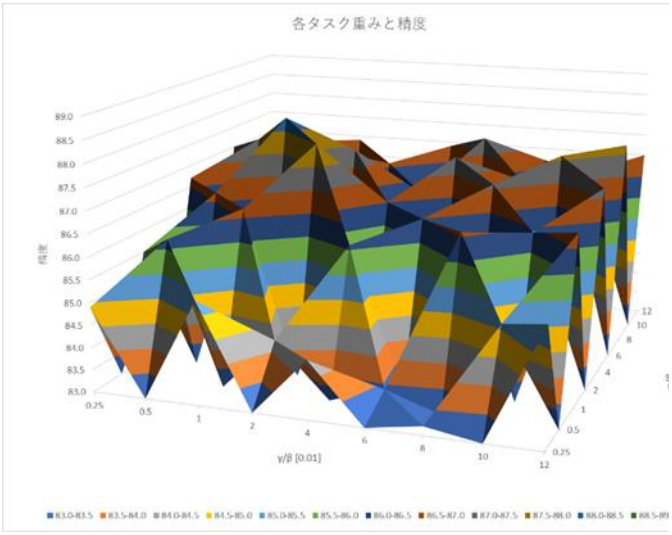

図 2 重み変化と精度のグラフ

図 2 のグラフの上に凸となるピークの部分から，C3D-AVG ネットワークの精度最大となる際の重みは$\frac{\alpha}{\beta} = 4, \frac{\gamma}{\beta} = 0.01$となる．よって各重みの値は式(2)に示す通りとなる．

$$\alpha = 4, \quad \beta = 1, \quad \gamma = 0.01 \tag{2}$$

5.2 損失範囲の測定

3 タスクの損失の範囲を揃えるために，各損失が取りうる値の範囲を C3D-AVG ネットワークの学習過程における損失の結果から求める．得られた損失範囲を式(3),(4),(5)に示す．

$$0 \leq L_{AQA} \leq 25.5 \tag{3}$$

$$0 \leq L_{Cls} \leq 8.81 \tag{4}$$

$$0 \leq L_{Cap} \leq 776 \tag{5}$$

この結果をもとに，損失の範囲を[0,1]に揃えた損失をそれぞれ$\frac{L_{AQA}}{25.5} = L'_{AQA}, \frac{L_{Cls}}{8.81} = L'_{Cls}, \frac{L_{Cap}}{776} = L'_{Cap}$とする．

5.3 特性解析の結果

4.1 節の結果から，C3D-AVG ネットワークの精度最大となる際の最終損失は式(6)に示す通りとなる．

$$L_{final} = 4L_{AQA} + L_{Cls} + 0.01L_{Cap} \tag{6}$$

しかし，この時点では損失の範囲が考慮されていないため，4.2 節で求めた損失の範囲を利用し式(7)に示す通りに変形する．

$$L_{final} = 4 * 25.5 * \frac{L_{AQA}}{25.5} + 8.81 * \frac{L_{Cls}}{8.81} + 0.01 * 776 * \frac{L_{Cap}}{776}$$

$$= 102L'_{AQA} + 8.81L'_{Cls} + 7.76L'_{Cap} \tag{7}$$

式(7)における各損失の重み係数が，損失の範囲を考慮した重みとなる．よって，この三つの重みの比率が C3D-AVG ネットワークの精度に与える各タスクの影響力となる．それぞれの比率の数値を表 1 に示す．

表 1 各タスクの影響力

AQA	動作分類	キャプション生成
86.03 [%]	7.43 [%]	6.54 [%]

6 まとめ

本研究では提案手法に基づく C3D-AVG ネットワークの特性解析を行った．結果として，AQA タスクの影響力は約 86%，動作分類タスクの影響力は約 7.4%，キャプション生成タスクの影響力は約 6.5%であった．したがって，メインタスクの影響力が最も大きいながらも，二つのサブタスクによる精度向上の影響力も十分に存在することが解析できた．

参考文献

[1] 佐藤国正，“スポーツ審判に関する研究”，桐蔭論叢，第 43 号，pp.57-63，Dec. 2020.

[2] 田野有一，“体操競技の特性と問題”，帯広大谷短期大学紀要，10 巻，pp.55-64, Mar. 1973.

[3] P. Parmar, and B. T. Morris, "What and How Well You Performed? A Multitask Learning Approach to Action Quality Assessment", IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp.304-313，June 2019.

2021 年度　卒業論文

スポーツにおける動作品質評価の性能改善に向けた特性解析

Characteristic Analysis for Performance Improvement of Action Quality Assessment in Sports

提出日　2022 年 2 月 1 日

指導教員　渡辺　裕　教授

早稲田大学　基幹理工学部　情報通信学科

1W183004-3

足立　翔平

目次

第1章　序論

1.1 研究の背景

スポーツとは，私たちの日常に非常になじみの深い活動である．このスポーツを特にスポーツ競技として考えた際，審判の採点評価の基準から三つの分類に分けることができる[1]．一つ目は柔道やサッカーに代表される対人競技．二つ目は競泳や競輪に代表される競走．三つ目はフィギュアスケートや体操競技に代表される採点競技である．これらのスポーツ競技において，特に採点競技は人間が主観を用いて選手の技に点数をつけるという採点評価の特性上，採点基準の曖昧さや審査の公平性という点について課題が挙げられている．これらの課題に対する改善策として，熟練の技術を持つ審査員を多数動員することによる曖昧さの排除や，複数国家の審査員を動員することによる公平性の確保などが行われてきたが，採点の判定に対する問題は依然として解決には至っていない[2]．そこで，近年においては上記課題を解決するために，技の採点評価に機械学習を取り込む動きが活発化している．この例として，富士通は 3D センシング技術および技認識技術を用いた体操競技の採点支援システムを提案し，国際体操連盟によって承認を得ている[3]．機械による客観的な採点がもたらす正確さや公平性が注目されている．

1.2 本研究の目的

機械学習によって人間の動作の評価を行う技術として Action Quality Assessment（以降 AQA と記述する．詳細は第 2 章 2.2 節で述べる）がある．これはスポーツの採点評価にも利用可能であるが，スポーツ競技の大会で人間の審査員に代わり採点評価を行うには現状の精度では不十分である．そこで本論文では，公式な競技シーンにおける AQA 利用のための性能改善を研究の目的とする．

AQA の最新の手法の一つとして Multi Task Learning-AQA（以降 MTL-AQA と記述する．詳細は第 2 章 2.3 節で述べる）が存在する．C3D-AVG ネットワークとして提案されている MTL-AQA は，従来の AQA の主タスクである品質評価に複数の補助タスクを加え性能向上を図った手法である．しかし，C3D-AVG ネットワークにおいて，各補助タスクがどれほどの比重で最終的な性能向上に貢献しているかという研究結果は報告されていない．そこで本論文は C3D-AVG ネットワークの特性解析を行うことで，MTL-AQA の機能向上のために注目すべきタスクを特定し，性能向上への足掛かりを得る手法を提案する．また，MTL-AQA は AQA の拡張を行う構想のシステムであるため，この研究の結果は将来的に AQA 全般の性能向上につながる．

1.3 本論文の構成

本論文の構成を以下に示す.

第 1 章は本章であり，本研究の背景，目的について述べる.

第 2 章では先行研究を始めとする本研究に関連する技術について述べる.

第 3 章では本研究で提案する特性解析手法について述べる.

第 4 章では実験結果及び考察について述べる.

第 5 章では本論文の結論と今後の課題について述べる.

第2章　　関連技術

2.1 まえがき

本章では本研究で用いる技術の基本となる AQA，および AQA の発展形であり本研究の特性解析対象である C3D-AVG ネットワークについて述べる．また，これらの技術を構成する C3D,S2VT キャプション生成，動作分類の技術について述べる．

2.2 AQA

AQA は, 人間の動作を表す映像を入力として, その動作のスコアを判定する技術である．入力された映像を 3 次元畳み込みニューラルネットワークに通して特徴量を抽出し，得られた特徴量およびアノテーションされたスコアの正解ラベルに基づいて動作の品質評価を学習する．この際用いられる 3 次元畳み込みニューラルネットワークは後述する C3D ネットワークが主流となっている．一般的な AQA のネットワークモデルを図 2.1 に示す．

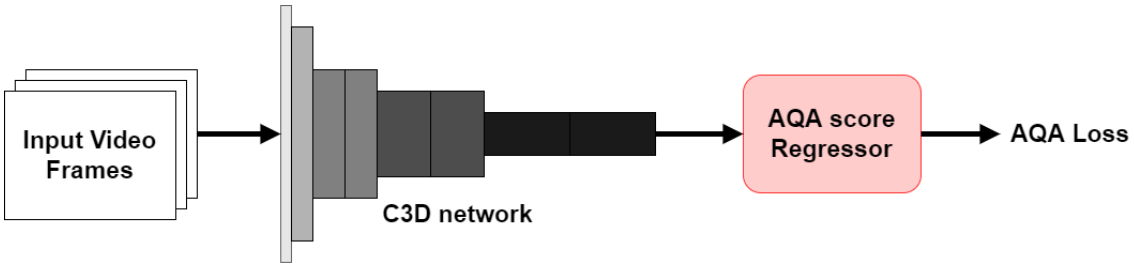

図 2.1 AQA ネットワークの構造例

2.3 C3D-AVG ネットワーク

C3D-AVG ネットワークは Parmar らが MTL-AQA として提案した AQA のネットワークモデルである[4]．従来の AQA タスクに新たに動作分類のタスクと動作の説明の文章であるキャプション生成のタスクを追加し，精度を向上させたマルチモーダルなネットワークモデルとなっている．モデルの概要を図 2.2 に示す．

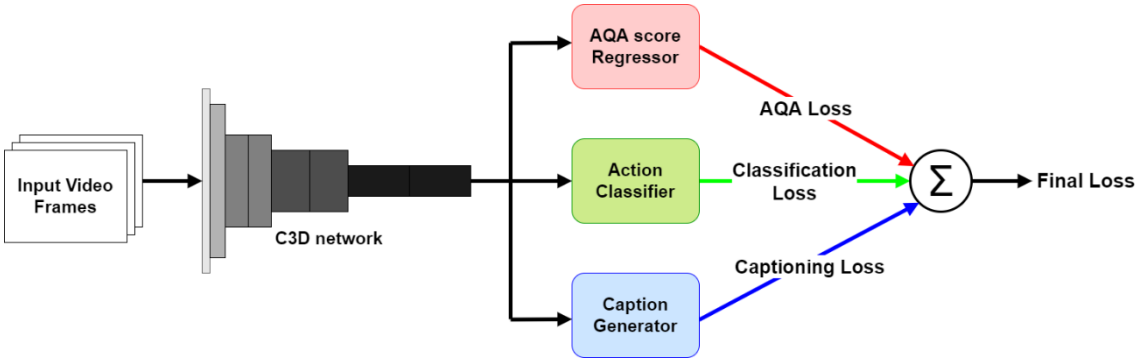

図 2.2 C3D-AVG ネットワークの構造

動作の品質は通常，何をどの程度行ったかを基準に評価される．そのため，動作品質評価のサブタスクは，何を行ったかを表す動作分類タスクと，どの程度行ったかを表すキャプション生成タスクによって構成されている．

本論文において，AQA タスクの Loss を L_{AQA}，動作分類タスクの Loss をL_{Cls}，キャプション生成タスクの Loss をL_{Cap}と表し，それぞれの損失関数を式(2.1)，(2.2)，(2.3)に示す．各タスクにおいて Loss を計算したのち，式(2.4)に示すように最終的な Loss であるL_{Final}を求める．ここで，式(2.4)におけるα，β，γは各タスクの Loss に対する重み係数であり，Parmar らの論文[4]においては$\alpha = 1$，$\beta = 1$，$\gamma = 0.01$が設定されている．

$$L_{AQA} = -\frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2 + |x_i - y_i| \tag{2.1}$$

$$L_{Cls} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{sa}\sum_{j=1}^{k_{sa}} y_{i,j}^{sa}\log\left(x_{i,j}^{sa}\right) \tag{2.2}$$

$$L_{Cap} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{sl}\ln\left(x^{cap}{}_{y^{cap}}\right) \tag{2.3}$$

$$L_{Final} = \alpha L_{AQA} + \beta L_{Cls} + \gamma L_{Cap} \tag{2.4}$$

C3D-AVG ネットワークのサブタスクにおいて，動作分類の技術として本章 2.5 節で述べる動作分類戦略[5]を利用している．また，キャプション生成の技術として本章 2.6 節で述べる Sequence to Sequence Video to Text (以降 S2VT と記述する)[6]を利用している．

2.4 C3D ネットワーク

C3D ネットワーク[7]は，２D 畳み込みネットワークに時間軸の演算を加えた 3D 畳み込みネットワーク[8]のモデルの一つである．入力動画を 16 フレームごとに分割した動画クリップを入力とし，このフレーム間の時間軸の情報を保持した 3 次元の特徴量を出力する．従来の 2 次元画像を対象とする畳み込みネットワークと比較して入力のパラメータが多いため,畳み込み層は一般的な２D 畳み込みネットワークより浅い 8 層となっている．C3D ネットワークの概要を図 2.3 に示す．

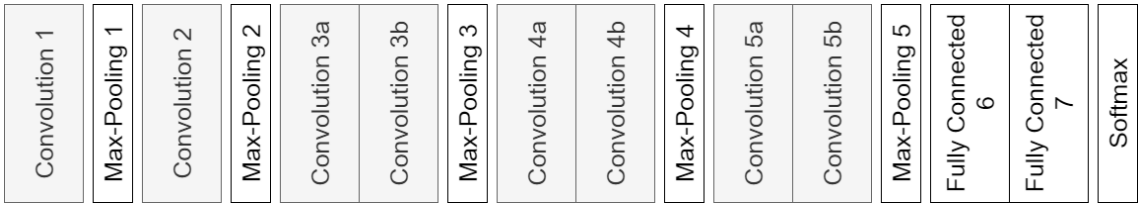

図 2.3 C3D ネットワークの構造

C3D-AVG ネットワークにおいては C3D ネットワークの 5 層目のプーリング層までを各タスク共通のバックボーンとして利用している．

2.5 動作分類

動作分類とは，映像に映っている対象がどのような動作を行っているかを認識，分類する技術である．OpenPose[9]に代表されるような対象の骨格を推定し，その動きを解析する手法や，C3D 特徴量をもとに動作を解析する手法が存在する．

本研究で扱う C3D-AVG ネットワークは採点評価の対象としてダイビングを設定している．ここで問題となるのが，ダイビングの分類ラベルの多様さである．ダイビングの演技種目は 3~4 桁の数字と 1 文字のアルファベットの組み合わせで表され，それぞれの数字およびアルファベットが演技のグループ，宙返りの回数，捻りの回数，演技の型といった内容を表す[10]．これによって表されるダイビングの分類数は数千種類にものぼり，すべての分類に対して十分なデータセットを用意することは現実的でない．そこで，C3D-AVG ネットワークにおいては Nibali らが提案したダイビングの分類手法[5]を用いている．数千ある分類ラベルをそのまま扱うのではなく，演技グループ，宙返り回数，捻り回数，演技の型といった要素を分けて扱い，それぞれの要素に対して動作分類を行っている．

2.6 S2VT

S2VT モデル[6]は Subhashini らが提案した，フレームを順次に読み込み，これに対応して単語を順次出力する形式の動画像キャプション生成モデルである．このモデルは可変長の動画像入力とそれに伴う可変長の単語列出力を可能にするために Recurrent Neural Network (RNN)の手法の一つである Long Short Term Memory (LSTM)ネットワークを用いている．S2VT モデルの概要図を図 2.4 に示す．

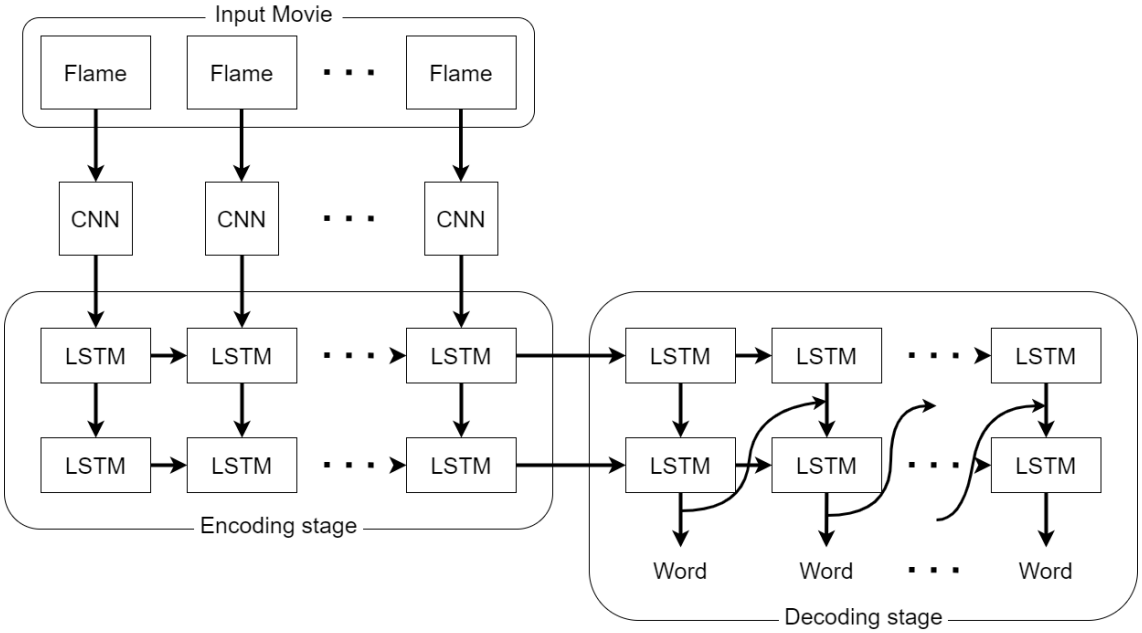

図 2.4 S2VT モデルの構造

最初の段階として動画像の各フレームを CNN に入力し，得られた CNN の出力を LSTM

に入力する．LSTM がフレーム毎のエンコードを完了させた段階でデコードを開始し，単語単位での文章生成を行う．

2.7 むすび

本章では本研究で用いる技術の基本となる AQA，および AQA の発展形であり本研究の特性解析対象である MTL-AQA について述べた．また，これらの技術を構成する 3D Convolutional Networks，動作分類，S2VT キャプション生成の技術について述べた．

第3章　C3D-AVG の特性解析手法

3.1 まえがき

本章では，MTL-AQA の代表例として C3D-AVG ネットワークの特性解析を行う手法を提案する．なお，本研究で扱う動作品質評価の対象はダイビングの動画とし，学習及びテストのデータとして MTL-AQA データセット[4]を用いる．

3.2 手法の概要

本研究では，2.3 節で述べた C3D-AVG ネットワークが持つ三つのタスクが，品質評価の精度にそれぞれどの程度の影響力を持っているかを調査する．その方法として，まず式(2.4)に示した最終 Loss の損失関数における各タスクの重み係数を変化させ，C3D-AVG ネットワークの精度との関係をグラフ化して読み取る．その後，C3D-AVG ネットワークの精度が最大となる場合の各タスクの重みの割合から，最終的な精度の向上に対する各タスクの影響力を推定する．本論文では Parmar らの手法[4]と同様に，C3D-AVG ネットワークの精度を評価する指標として 3.3 節で述べるスピアマンの順位相関係数[11]を用いる．

3.3 スピアマンの順位相関係数

スピアマンの順位相関係数は，2 グループの順位データから両者の相関関係を求める指標である．通常の相関係数は 2 変量の線形的な相関関係を扱うが，スピアマンの順位相関係数は 2 変量の順位の単調な関係を扱う．2 グループの変量をそれぞれx_i, $y_i (i = 1,2,3, \dots, n)$，スピアマンの相関係数を$r_s$として，$r_s$を求める数式を式(3.1)に示す．

$$r_s = 1 - \frac{6\sum_{i=1}^{n}(x_i - y_i)^2}{n(n^2-1)} \tag{3.1}$$

本研究において，x_iを予測スコア，y_iを正解スコアとして，両者の相関を C3D-AVG ネットワークの精度としている．

3.4 重み係数の正規化

3 タスクの重み変化と C3D-AVG ネットワークの精度の関係を読み取るため，三つそれぞれの重みを変化させた際の C3D-AVG ネットワークの精度の結果を表にまとめグラフ化を行う．しかし，AQA の重み，動作分類の重み，キャプション生成の重み，そして C3D-AVG ネットワークの精度の四つの要素があるためこのままではグラフとして扱うことが難しい．そこで，2 タスクの重みを残り 1 タスクの重みで正規化し，正規化した 2 タスクの重みと C3D-AVG ネットワークの精度の 3 軸のグラフを作成する．式(2.4)に示す通り，AQA の重みをα, 動作分類の重みをβ, キャプション生成の重みをγとする．正規化の様子を式(3.2)および式(3.3)に示す．

$$\alpha:\beta:\gamma \tag{3.2}$$

$$\frac{\alpha}{\beta}:1:\frac{\gamma}{\beta} \tag{3.3}$$

本手法ではαとγをβで正規化し，$\frac{\alpha}{\beta}$と$\frac{\gamma}{\beta}$および C3D-AVG ネットワークの精度の 3 要素の関係を 3 次元のグラフとする．

3.5 損失の範囲を考慮した重みの取り扱い

本節では最終的な各タスクの影響力を推定する際の各重みの取り扱いについて述べる．C3D-AVG ネットワークの精度最大となったときの AQA の重みαを A，動作分類の重みβを B，キャプション生成の重みγを C とする．すると，最終損失L_{final}は式(3.4)に示す通りとなる．

$$L_{final} = AL_{AQA} + BL_{Cls} + CL_{Cap} \tag{3.4}$$

最終損失における各タスクの損失の占める割合が各タスクの影響力と考えるため，この重み A, B, C の比率で各タスクの影響力を推定するのが理想であるが，実際には各損失L_{AQA}，L_{Cls}，L_{Cap}の取りうる範囲が異なるという点を考慮する必要がある．これらの損失の理論上の範囲はそれぞれの損失関数の式(2.1)，(2.2)，(2.3)より求められるが，ネットワークの学習段階の損失の推移を鑑みても，損失を求めるうえで各項全ての予測値が正解値と完全な乖離を起こすことは稀なため，現実に損失が最大値をとる可能性は低い．そこで，現実に各損失が取りうる値の上限を実験により求める．本節では仮に，L_{AQA}の上限値をL_{AQA_MAX}，L_{Cls}の上限値をL_{Cls_MAX}，L_{Cap}の上限値をL_{Cap_MAX}とする．各損失の範囲を式(3.5)，(3.6)，(3.7)に示す．

$$0 \le L_{AQA} \le L_{AQA\_MAX} \tag{3.5}$$

$$0 \le L_{Cls} \le L_{Cls\_MAX} \tag{3.6}$$

$$0 \le L_{Cap} \le L_{Cap\_MAX} \tag{3.7}$$

これらの損失の範囲を[0,1]に統一するため，式(3.5)，(3.6)，(3.7)についてそれぞれの上限値を除算する．このとき，上限値は 0 でないとする．結果を式(3.8)，(3.9)，(3.10)に示す．

$$0 \le \frac{L_{AQA}}{L_{AQA\_MAX}} \le 1 \tag{3.8}$$

$$0 \le \frac{L_{Cls}}{L_{Cls\_MAX}} \le 1 \tag{3.9}$$

$$0 \le \frac{L_{Cap}}{L_{Cap\_MAX}} \le 1 \tag{3.10}$$

損失の範囲を考慮した重みを取り扱うために式(3.4)を式(3.11)のように変形する．

$$L_{final} = AL_{AQA\_MAX}\frac{L_{AQA}}{L_{AQA\_MAX}} + BL_{Cls\_MAX}\frac{L_{Cls}}{L_{Cls\_MAX}} + CL_{Cap\_MAX}\frac{L_{Cap}}{L_{Cap\_MAX}} \tag{3.11}$$

式(3.11)において，$\frac{L_{AQA}}{L_{AQA\_MAX}}$，$\frac{L_{Cls}}{L_{Cls\_MAX}}$，$\frac{L_{Cap}}{L_{Cap_MAX}}$は式(3.8)，(3.9)，(3.10)より，範囲[0,1]で統一された損失であり，AL_{AQA_MAX}，BL_{Cls_MAX}，CL_{Cap_MAX}は損失の範囲の違いを考慮した各タスクの重みとなる．よって，本研究ではAL_{AQA_MAX}，BL_{Cls_MAX}，CL_{Cap_MAX}の比率をもとに各タスクの影響力を解析することとする．

3.6 むすび

本章では，C3D-AVG ネットワークの特性解析を行う手法および利用する評価指標について述べた．

第4章　　実験結果と考察

4.1 まえがき

本章では，実験で利用するデータセットおよび提案手法に基づく損失範囲の測定，特性解析の結果について述べる．また，これらの実験結果から得られる考察について述べる．

4.2 データセット

本実験では，大規模なダイビングのデータセットである MTL-AQA データセット[4]を利用する．このデータセットには 16 回の国際大会での競技シーンが含まれており，サンプル数は 1412 個である．それぞれの飛び込み動画サンプルには得点，捻り数や回転数といった動作認識の要素，実況解説によるキャプションのラベルが対応付けられている．動画のサイズは 171×128 画素となっている．

4.3 C3D-AVG ネットワークの精度最大となる重み測定

正規化した重み$\frac{\alpha}{\beta},\frac{\gamma}{\beta}$を変化させたことによる C3D-AVG ネットワークの精度の値を表 4.1 に示す．

表 4.1 重み変化と精度の表

γ/β [0.01] / α/β	0.25	0.5	1	2	4	6	8	10	12
0.25	84.9		85.3		84.3		83.2		85.6
0.5		86.2		84.2		83.4		85.1	
1	85.5		86.7		86.1		86.6		86.8
2		86.7		88.0		86.5		86.4	
4	86.7		88.3		86.5		87.1		87.2
6		87.5		87.4		87.4		87.6	
8	87.0		87.1		86.2		86.7		87.7
10		87.0		86.6		87.4		86.7	
12	86.5		86.9		86.9		86.7		87.0

実験に費やすことのできた時間の都合上，表 4.1 において一つ飛ばしの要素についての実験を行うこととした．この表 4.1 の結果を三次元グラフに示したものを図 4.1 に示す．

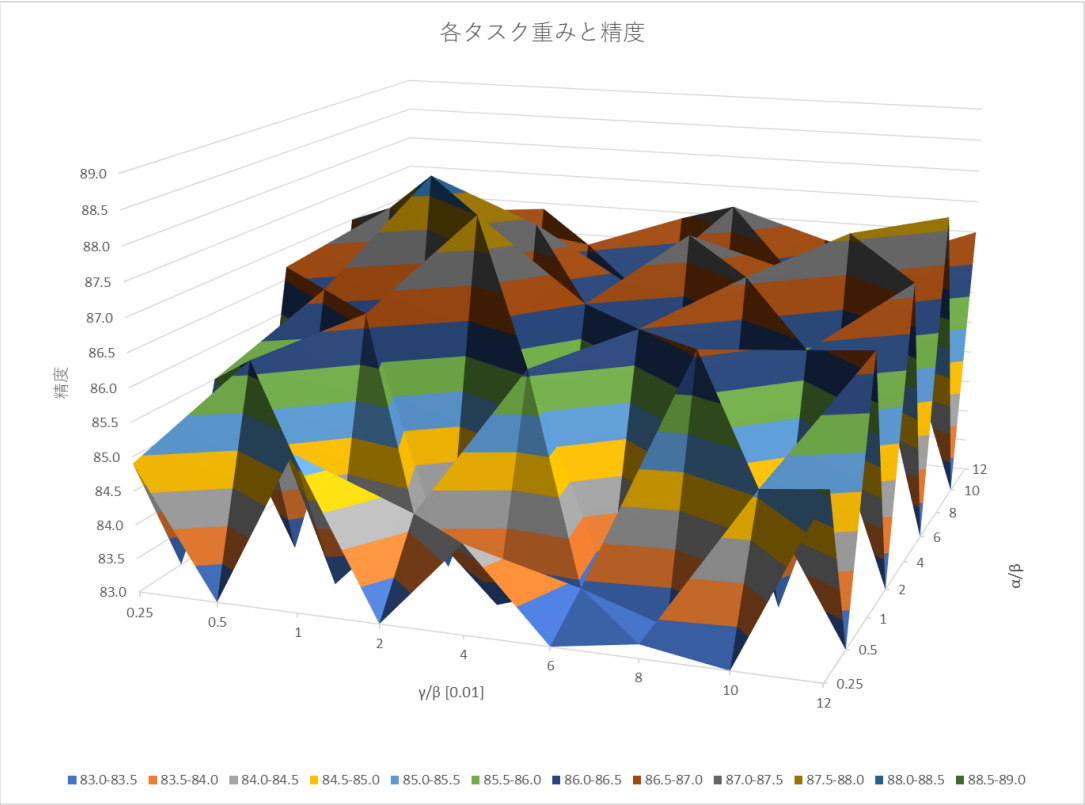

図 4.1 重み変化と精度のグラフ

図 4.1 のグラフの上に凸となるピークの部分から，C3D-AVG ネットワークの精度最大となる際の重みは$\frac{\alpha}{\beta}=4,\frac{\gamma}{\beta}=0.01$となる．よって各重みの値は式(4.1)に示す通りとなる．

$$\alpha=4,\qquad \beta=1,\qquad \gamma=0.01 \tag{4.1}$$

4.4 損失範囲の測定

3.5 節の提案手法に基づき，実験を行った中で各損失が取り得た最大値に基づいた各損失の範囲を式(4.2),(4.3),(4.4)に示す．

$$0\leq L_{AQA}\leq 25.5 \tag{4.2}$$

$$0\leq L_{Cls}\leq 8.81 \tag{4.3}$$

$$0\leq L_{Cap}\leq 776 \tag{4.4}$$

また，式(3.8)，(3.9)，(3.10)に示すように各損失の範囲を[0,1]に揃えることで式(4.5)(4.6)(4.7)に示す通りとなる．

$$0\leq \frac{L_{AQA}}{25.5}\leq 1 \tag{4.5}$$

$$0\leq \frac{L_{Cls}}{8.81}\leq 1 \tag{4.6}$$

$$0 \leq \frac{L_{Cap}}{776} \leq 1 \tag{4.7}$$

ここで，範囲[0,1]に揃えた損失をそれぞれ$\frac{L_{AQA}}{25.5} = L'_{AQA}, \frac{L_{Cls}}{8.81} = L'_{Cls}, \frac{L_{Cap}}{776} = L'_{Cap}$とする．

4.5 特性解析の結果

4.3 節の結果から，C3D-AVG ネットワークの精度最大となる際の最終損失は式(4.8)に示す通りとなる．

$$L_{final} = 4L_{AQA} + L_{Cls} + 0.01L_{Cap} \tag{4.8}$$

しかし，この時点では損失の範囲が考慮されていないため，4.4 節で求めた損失の範囲を利用し，式(3.11)に示す通りに変形する．結果を式(4.9)に示す．

$$\begin{aligned} L_{final} &= 4 * 25.5 * \frac{L_{AQA}}{25.5} + 8.81 * \frac{L_{Cls}}{8.81} + 0.01 * 776 * \frac{L_{Cap}}{776} \\ &= 102L'_{AQA} + 8.81L'_{Cls} + 7.76L'_{Cap} \end{aligned} \tag{4.9}$$

式(4.9)における各損失の重み係数が，損失の範囲を考慮した重みとなる．よって，この三つの重みの比率が C3D-AVG ネットワークの精度に与える各タスクの影響力となる．それぞれの比率の数値を表 4.2 に示す．

表 4.2 AQA，動作分類，キャプション生成タスクの影響力

AQA	動作分類	キャプション生成
86.03 [%]	7.43 [%]	6.54 [%]

比率をグラフとして表した結果を図 4.2 に示す．

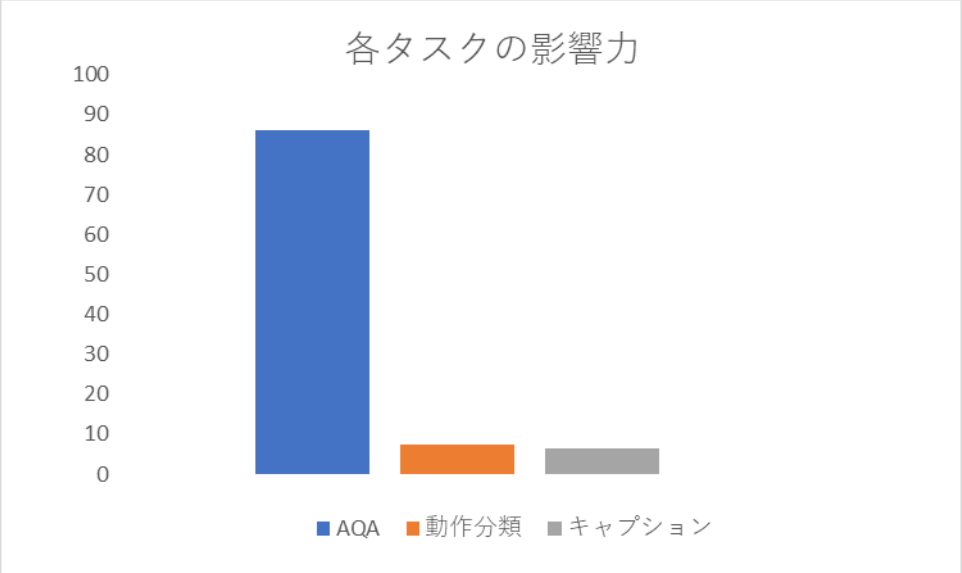

図 4.2 AQA，動作分類，キャプション生成タスクの影響力

4.6 考察

4.5 節の表 4.2 および図 4.2 に示した通り AQA，動作分類，キャプション生成の 3 タスクの中で最も C3D-AVG ネットワークの精度への影響力が最も大きいのは AQA タスクである．これは，MTL-AQA の手法が従来の AQA シングルタスクモデルをベースとして提案されている点から妥当な結果と言える．サブタスクとして採用されている動作分類タスクとキャプション生成タスクについてみると，二つ合わせて全体の 14%ほどを占めているため，サブタスクによる精度向上の効果が十分にあることがわかる．また，動作分類タスクとキャプション生成タスクの両者間の影響力の差は 1%程度であるため，各サブタスクのモデルの性能向上を目指すメリットは両者とも同様に存在する．

4.7 むすび

本章では提案手法に基づく C3D-AVG ネットワークの特性解析を行った．結果として，AQA タスクの影響力は約 86%，動作分類タスクの影響力は約 7.4%，キャプション生成タスクの影響力は約 6.5%であった．したがって，メインタスクの影響力が最も大きいながらも，二つのサブタスクによる精度向上の影響力も十分に存在することが解析できた．

第5章　結論と今後の課題

5.1 結論

本研究では MTL-AQA のモデルである C3D-AVG ネットワークについて，精度最大となる際の AQA タスク，動作分類タスク，キャプション生成タスクそれぞれの重みを求め，各損失の範囲を考慮したうえで 3 タスクの影響力を解析した．サブタスクである動作分類タスクとキャプション生成タスクによる影響力を確認でき，両者の影響力がおおよそ同等であることから，二つのサブタスクのモデルについて精度向上を図る試みは MTL-AQA の精度の向上に効果があることを示した．

5.2 今後の課題

本研究では C3D-AVG ネットワークの得点予測精度を最大化する各タスクの重みについて特性解析を実施した．そのため，より重み変化の刻み幅を細かくし，データ数を増やすことによってより正確な精度最大となる際の重みを測ることができると思われる．

また，本研究で示した結論より，動作分類とキャプション生成のサブタスクについて精度向上の可能性が十分にあることが示せたため，今後は動作分類やキャプション生成のモデル自体の性能向上や，より精度の高いモデルへの置き換え等を試みることで MTL-AQA の性能改善を目指すことが可能と考えられる．

謝辞

本研究に利用した PC 等の実験環境を整えてくださり，研究テーマや方針に熱心にご指導してくださった渡辺裕教授に深く感謝いたします．また，本論文の校閲を始めとし，多くのご指導を頂いた早稲田大学国際情報通信センターの石川孝明様に深く感謝申し上げます．

また，異なる研究に打ち込みながらも，自身のみでは得難い視点からのご指摘や助言をくださった同研究室の皆様に心より御礼申し上げます．

最後に，精神面や金銭面で多くの力添えをしていただいた家族に心より感謝致します．

参考文献

[1] 佐藤国正，“スポーツ審判に関する研究”，桐蔭論叢，第 43 号，pp.57-63，Dec. 2020.

[2] 田野有一，“体操競技の特性と問題”，帯広大谷短期大学紀要，10 巻，pp.55-64，Mar. 1973.

[3] 桝井，手塚，矢吹，佐々木，“3D センシング・技認識技術による体操採点支援システムの実用化”，情報処理，Vol61，No.11，デジタルプラクティスコーナー，Oct. 2020.

[4] P. Parmar, and B. T. Morris, “What and How Well You Performed? A Multitask Learning Approach to Action Quality Assessment”, IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp.304-313, June 2019.

[5] A. Nibali, Z. He, S. Morgan, D. Greenwood, “Extraction and Classification of Diving Clips from Continuous Video Footage”, 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp.94-104, Jul. 2017.

[6] S. Venugopalan, M. Rohrbach, J. Donahue, R. Mooney, T. Darrell, K. Saenko, “Sequence to Sequence -- Video to Text”, 2015 IEEE International Conference on Computer Vision (ICCV), pp.4534-4542, Dec. 2015.

[7] D. Tran, L. Bourdev, R. Fergus, L. Torresani, M. Paluri, “Learning Spatiotemporal Features with 3D Convolutional Networks”, 2015 IEEE International Conference on Computer Vision (ICCV), pp.4489-4497, Dec. 2015.

[8] S. Ji, W. Xu, M. Yang, K. Yu, “3D Convolutional Neural Networks for Human Action Recognition”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 35, Issue 1, pp.221-231, Jan. 2013.

[9] Z. Cao, G. Hidalgo, T. Simon, S. Wei, Y. Sheikh, “OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 43, Issue 1, pp.172-186, 17 Jul. 2019.

[10] C. Spearman, “The Proof and Measurement of Association between Two Things”, The American Journal of Psychology, Volume15, No.1, pp.72-101, Jan. 1904.

[11] 公益財団法人 日本水泳連盟，“飛込競技規則”，pp.1-3，1 Apr. 2018.

図一覧

表一覧