

水中における VGG による人物抽出手法に関する研究

21T308 岡本 和波 (最所研究室)

1. はじめに

競泳においてタイムを短縮するためにフォームの改善が重要である。フォームの改善方法の一つに、理想の選手のフォームを模倣することが挙げられる。しかし、理想とする選手と自分とでは身長や体格が異なるため、理想とする選手の映像を見て模倣するのは難しい。その為、自分と身長や体格が異なる選手でも模倣しやすい映像が必要である。

2. 本研究の最終目標

この映像を作るために理想のフォームを反映したユーザの競泳映像を生成するシステムの開発を目指す。このシステムはユーザの競泳映像と理想的なフォームで泳ぐ選手の競泳映像を入力することで、ユーザの競泳映像に理想とする選手のフォームが反映された映像を生成する。このシステムは二つの画像を入力し、一方の画像から輪郭を、他方の画像から色や模様を抽出し、それぞれの特徴を反映した新たな画像を生成するニューラルスタイル変換という技術を応用する。

3. 課題

ニューラルスタイル変換は、入力した画像から物体を認識し、認識する際に使用した特徴を基に変換を行う。しかし、想定する映像は水中映像であり、地上での物や人とは色彩が異なるだけでなく、撮影環境や時間によっても異なる。そのため、水中における人物を認識するのは難しく、ニューラルスタイル変換で必要となる特徴を抽出することはできない。

本研究では、水中映像から人物を認識できるようにし、ニューラルスタイル変換で扱える特徴を抽出できるようにすることを課題とする。

4. 水中における人物抽出の提案手法

課題解決の為に、ニューラルスタイル変換で使用されている VGG を使用し、その中間層から人物抽出ができるようにする。本研究では VGG の性能調査を行い、それに伴う提案手法を述べ、

4.1. 予備実験

競泳画像から泳者が映っている画像を人物画像、映っていない画像を背景画像としてデータセットを構築し、このデータセットを VGG に学習をさせモデル構築を行った。このモデルから認識画像を入力し、中間層の画像を確認したところ、人物特徴だけでなく、泳ぐ際に生じる泡や影といった他の要素も特徴として

反映されていた (図 1)。

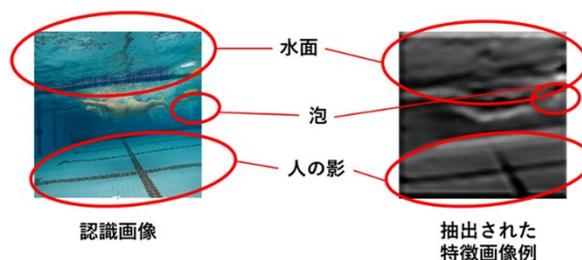


図 1 予備実験の特徴画像例

そのため、他の要素も特徴として残っているため人物抽出は行えていない。

4.2. 予備実験結果を踏まえた提案手法

予備実験の結果から学習データによる人物画像において、泳ぐ際に発生する他の要素でも学習が行われていると考えられる。そのため、違いを人物画像と背景画像の違いを人物だけにするために、学習画像における人物画像から人物抽出を行い、人物抽出を行った画像に背景画像を合成する (図 2)。これにより泳ぐ際に発生する要素が少なくなり、人物だけがさとして残るため、特徴として人物だけが抽出できると考えた。

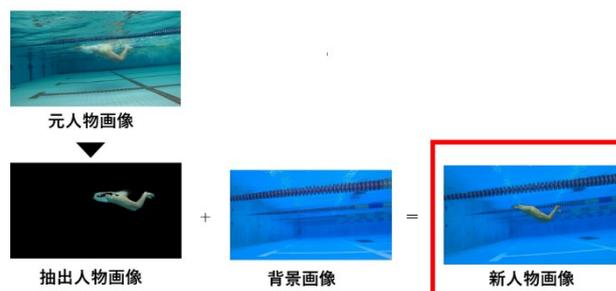


図 2 人物画像作成方法

5. 実装

4.2 節で提案したデータセットを作成するにあたり、画像における人物抽出手法として背景除去と泡の除去をおこなった。これらを踏まえ、データセットの作成および学習について述べる。

5.1. 背景除去

背景除去とは画像内で前景と背景と分類を行い、背景領域を消去する技術である。本研究ではこの技術を利用した rembg を使用した。rembg は画像内に映る顕著な領域を前景とみなし、背景を除去する技術である。本研究で使用する競泳画像は泳者に焦点をあててい

るため、泳者領域において顕著性が高くなることで前景とみなされると考えた。しかし、泳ぐ際に発生する泡は泳者付近で発生するため泡とも焦点が合っている。そのため、泳者だけでなく泡も残った画像になる。

5.2. 泡の低減

5.1 節では泳者とともに泡の残った画像ができています。ここで泡の性質について述べる。水中において青色成分が強く残り、青色が強調された色になる。泡もこの現象をうけ、青色成分が高くなる。また、泡は光の散乱特性と水の吸収特異性により周辺の水よりも青色成分が高くなるとされている。この性質を利用し、青色成分で一定の閾値を設けることで泡領域と人物領域を分割し、泡領域を低減した。

5.3. 背景との合成

5.1、5.2 の処理により人物抽出画像が完成された。この画像に競泳の泳者が映っていない画像を合成する。これにより背景画像と人物画像の差は人物だけとなる。

5.4. 学習

5.3 により作成した画像を新たな「人物画像」とし、人物画像に使用されていない背景画像を「背景画像」としてデータセットを構築する。このデータセットに VGG を学習させ、モデルを構築する。

6. 評価実験

本実験ではニューラルスタイル変換で用いることを前提としているため、5 章で構築したモデルを対象にどの層で人物抽出が行えるか、また、提案手法により精度向上はおきているのかを評価する。

6.1. 実験の目的

本実験では 5 章で構築したモデルが人物抽出をおこなっているのかを評価することを目的とする。また、泳ぐ際に発生する要素を減らしたことによる人物抽出性能がどれだけ変化したのかを評価することも目的とする。

6.2. 実験の方法

香川大学水泳部員 3 名から提供された水中映像をフレーム単位に分割し、人物が映っている画像を任意に 3 枚ずつ抽出する。そして、抽出した画像を提案手法のモデルと予備実験のモデルに入力し、認識時に得られる特徴画像を用いる。これらの特徴画像と筆者が手動で作成した正解画像を比較し、一致度を示す IoU (Intersection over Union) を算出する。また、各層ごとに求めた IoU の最大値を代表的な特徴画像として選定し、提案手法と予備実験モデルの最大 IoU を比較することで、それぞれの人物認識精度を評価する。

6.3. 実験の結果

提案手法の人物抽出の評価実験の結果を表 1、提案手法と予備実験の人物抽出精度の比較結果を表 2 にまとめる。

層	最大値	平均値	中央値	標準偏差
block1_conv1	0.207915	0.099022	0.081037	0.057874
block1_conv2	0.079554	0.055471	0.051152	0.016179
block2_conv1	0.132258	0.112029	0.121058	0.020765
block2_conv2	0.193562	0.148839	0.141984	0.027111
block3_conv1	0.372079	0.239044	0.227738	0.064646
block3_conv2	0.305728	0.196573	0.196758	0.052768
block3_conv3	0.312457	0.218340	0.206897	0.056927
block4_conv1	0.500000	0.375190	0.372872	0.075431
block4_conv2	0.662213	0.465984	0.425602	0.097096
block4_conv3	0.649002	0.520051	0.523214	0.081249
block5_conv1	0.598449	0.545647	0.557372	0.060324
block5_conv2	0.647377	0.544692	0.561609	0.079654
block5_conv3	0.579849	0.439248	0.456891	0.099363

表 1 VGG 各層ごとの IoU 統計値

予備実験(最大値)	提案手法(最大値)	変化(背景なし - 背景あり)
0.40	0.67	+0.27

表 2 提案手法と予備実験 IoU 最大値の差

7. 考察

人物抽出の評価より block4_conv3 から block5_conv2 で平均値 0.5 を超えていた。この 0.5 という指標は物体検知の最低基準であり、その点において人物抽出は行えているといえる。しかし、高度な物体検出には指標として 0.75 が使われておりこれには最大値であっても満たせていない。そのため、改善の余地がある。提案手法と予備実験の比較では提案手法の IoU が 0.27 と大きく増加しており、人物抽出性能は向上している。これは背景特徴が抽出されにくくなったことにより精度が向上したと考えられる。

8. おわりに

提案手法により最低限の人物抽出は行えるようになったが、精度としては十分ではない。そのため、学習画像における人物抽出の精度を向上させ、学習データを増やすことにより精度向上を目指す。

また、精度が上がり次第、ニューラルスタイル変換の応用も進め、理想のフォームを反映した自身の映像を生成するシステムの開発を目指す。