

ゴール領域に着目したバスケットボール番組のシュートシーン検出に関する研究 A Method to Detect Shoot Scene of Basketball Programs Focusing on Goal Area

迫田 峻* 大野 将樹** 獅々堀 正幹**
Shun Sakoda Masaki Oono Masami Shishibori

1. はじめに

近年、映像配信技術の発達によって膨大な量の映像の映像が配信され、それら全ての映像を視聴することは困難である。特にスポーツ映像を例に挙げると、サッカー、野球、バスケットボールなどのスポーツは 1 試合全て視聴する必要があるため、スポーツ番組を複数見てしまうと多大な時間を費やしてしまう。そこで、動画の内容を理解でき、かつ、スポーツ動画を興味のあるシーンのみを抽出し、動画を短縮する技術が必要不可欠である。

本研究では、長時間の使用が必要になるスポーツ動画の重要シーンの検出を行う。特にスポーツ動画の中でもバスケットボールについて注目する。バスケットボールにおいてシュートシーンはプレイの始点であるため重要シーンと言える[1]。そこで本稿では、深層学習に基づく物体検出器とクラス分類器を用いてバスケットボールの動画からシュートシーンを検出する手法を提案する。

本手法では、まずバスケットボールの動画中ではゴール領域が出現する位置の変化が小さいことに着目し、Single Shot Multibox Detector(SSD)[2]と呼ばれる物体検出器を用いてゴール領域の画像を取得する。次にゴール領域の画像をVGG-16と呼ばれるConvolutional Neural Network(CNN)[3][4]に基づくクラス分類器を用いて、シュートシーン画像と非シュートシーン画像の 2 クラスに分類する。更に、前後のフレームの変化からオプティカルフロー[5]を生成し、そのオプティカルフローを画像に書き込んだものをクラス分類することで、3 ポイントシュートのように動きの速いシュートシーンを検出する。

2. 関連研究

Vinay Bettadapure ら[1]は、放送された映像から観客や解説の音声、シュートを決めた選手、シュートが決まった際の得点差とシュートの種類、プレイ中の選手の動きの量をそれぞれスコア化し、重要シーンを抽出する研究がある。この研究は、菅火薬や解説者の音声はシュートが決まった際に大きくなることから、音声の大きさをスコア化し、スター選手の得点は他の選手の得点と比べて観客達の興奮度が高まることから、選手一人の平均得点をスコア化し加算する。これらのスコアを合計し、ハイライトを生成している。しかし、得点の判断を表示されているテロップやスコアシートを後付けしてから行っている。そこで本研究では、シュートシーンの検出をテロップから行わず、ゴール領域から検出する。

*徳島大学大学院先端技術科学教育部 Graduate School of Advanced Technology and Science Faculty of Engineering, Tokushima University

**徳島大学大学院社会産業理工学研究部 Graduate School of Technology, Industrial and Social Science, Tokushima University

3. 提案手法

3.1 システムの概要

本研究の提案手法の流れを図 3.1 に示す。まず動画データを入力すると、SSD でゴール領域の抽出を行う。次に抽出されたゴール領域を画像化する。その画像をシュートシーンと非シュートシーンの 2 クラスで学習を行ったモデル、そしてゴール領域のオプティカルフローを表示させた画像をシュートシーンと非シュートシーンの 2 クラスで学習を行ったモデルを用いて、シュートシーンと非シュートシーンを判別する。

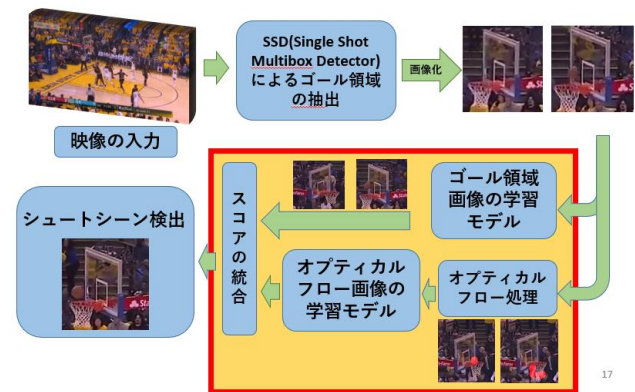


図 3.1: 提案手法の流れ

3.2 SSD によるゴール領域抽出

まず SSD を用いてゴール領域を検出し、画像化し取得する。SSD は物体検出を行うのに非常に優秀であるが、入力した画像サイズに対して小さい物体を検出しづらいという欠点がある。これは小さい物体を検出する層ほど浅めの層であり、検出する物体の特徴を抽出しきれないことによるものである。そこで SSD でシュートシーンの判別をするのではなく、ゴール領域だけを抽出するために用いる。

SSD のみでシュートシーンと非シュートシーンの判別を行うとシュートシーンの識別率が高いものの、非シュートシーンでない部分もシュートシーンと判別しているものが多く、正しく判別することができない。そこでシュートシーンの識別を SSD の物体検出としてのアプローチで行うのではなく、VGG-16 の Fine-tuning の画像分類のアプローチで行う。

3.3 ゴール領域画像のシュートシーン判別

次に SSD で抽出したゴール領域を画像化し、VGG-16 モデルを Fine-tuning した CNN を用いてシュートシーンと非シュートシーンの 2 クラス分類を行う。学習に用いたデータ数は表 3.1 に示す。学習の際に偏りが出ないように、シュートシーンと非シュートシーンの画像数を同じにした。

表 3.1:ゴール領域画像の学習数

学習データ	シュート	500枚
	非シュート	500枚
バリレーションデータ	シュート	200枚
	非シュート	200枚

3.4 オプティカルフロー画像のシュートシーン判別

SSD で取得したゴール領域の画像にオプティカルフローを書き込む。そうすることで 1 枚の画像にボールの動きやゴールリングのネットの動きを表示することが可能になり、時間の流れを確認できるようになる。またオプティカルフローを書き込む際に、画像的に判別しやすくなるように、フローの太さと色を変更した。実際にオプティカルフローを書き込んだ画像を図 3.2 に示す。



[a]例 1

[b]例 2

図 3.2:オプティカルフローを書き込んだ画像

オプティカルフロー画像の学習は、ゴール領域画像の学習と同様に行う。表 3.2 に学習に用いたデータ数を示す。

表 3.2:オプティカルフロー画像の学習数

学習データ	シュート	100枚
	非シュート	100枚
バリレーションデータ	シュート	40枚
	非シュート	40枚

3.5 クラススコアの統合

ゴール領域の学習での分類は、シュートシーン時にボールが確認できて、非シュートシーンとの違いが分かりやすい時に精度が良い。またオプティカルフローを表示させた画像では、ボールが移動したときのオプティカルフローが大きく表示されている時に精度が良い。つまり非シュートシーンやボールが確認できている時は、ゴール領域画像でのクラス分類を行い、ボールの動きが速く、オプティカルフローが大きく表示される時は、オプティカルフロー画像でのクラス分類を行うことで、精度向上を試みる。ゴール領域画像のクラス分類の尤度を A 、割合を α 、オプティカルフロー画像のクラス分類の尤度を B 、割合を β として計算する。計算式は以下の通りである。

$$\text{統合スコア} = \frac{(A * \alpha + B * \beta)}{10} \quad (1)$$

4. 評価

評価は再現率・適合率・F値で評価する。

$$\text{再現率} = \frac{\text{正しく検出できたシュートシーン数}}{\text{実際のシュートシーン数}} \quad (2)$$

$$\text{適合率} = \frac{\text{正しく検出できたシュートシーン数}}{\text{手法によって検出したシュートシーン数}} \quad (3)$$

$$F \text{ 値} = \frac{2 * \text{再現率} * \text{適合率}}{\text{再現率} + \text{適合率}} \quad (4)$$

スコア統合の結果を表 4.1 に示す。ゴール領域画像の 2 クラス分類のみは再現率が高く、適合率が低い。これは、ボールの動きが速いシュートシーンの時にボールがほとんど見えず、画像上で認識できなかったためである。そこでオプティカルフロー画像の 2 クラス分類のスコアを組み合わせることで、ボールの認識が難しいシュートの分類の精度をあげた。またスコアを統合する際に、ゴール領域画像の 2 クラス分類の割合をオプティカルフロー画像の 2 クラス分類の割合よりも高くすると、全体のシュートシーンの検出率を上げることが可能になる。またオプティカルフロー画像の 2 クラス分類の割合をゴール領域画像の 2 クラス分類の割合よりも高くすると、検出率の精度を上げることが可能になる。

表 4.1:再現率, 適合率, F 値

	再現率	適合率	F 値
ゴール領域画像のみ	0.845	0.637	0.726
オプティカルフローのみ	0.758	0.698	0.727
$\alpha : \beta = 4 : 6$	0.849	0.714	0.776
$\alpha : \beta = 5 : 5$	0.855	0.698	0.769
$\alpha : \beta = 6 : 4$	0.859	0.693	0.767

5. まとめ

本論文では、深層学習に基づく物体検出器とクラス分類器を用いて、バスケットボールの動画におけるシュートシーン検出を提案した。ゴール領域画像の 2 クラス分類だけでは検出することが難しかった、ボールの動きが速いシュートシーンの検出を可能にするために、オプティカルフローを用いてボールの動きの認識を可能にした。結果として、ゴール領域画像の 2 クラス分類とオプティカルフロー画像の 2 クラス分類で得られたスコアを組み合わせることで精度が向上した。またスコアの割合を調整することで、ユーザーが求めるシーンを抽出することも可能である。

6. 参考文献

- [1] Vinary Bettadapure, Caroline Pantofaru and Irfan Essa "Leveraging Contextual Cues for Generating Basketball Highlights", Proceeding of the 2016 ACM on Multimedia Conference pp.908-917, 2016
- [2] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu and Alexander C. Berg, "SSD: Single Shot Multibox Detector", ECCV 2016, pp.21-37, 2016
- [3] Karen Simonyan, Andrew Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition", 2014
- [4] 中山英樹, "深層畳み込みニューラルネットワークによる画像特徴抽出と転移学習", 電子情報通信学会音声研究会 7 月研究会, pp.55-59, 2015
- [5] B.K.P.Horn and B.G.Schunck, "Determining Optical Flow", Artificial Intelligence, vol.17, pp.185-203, 1981