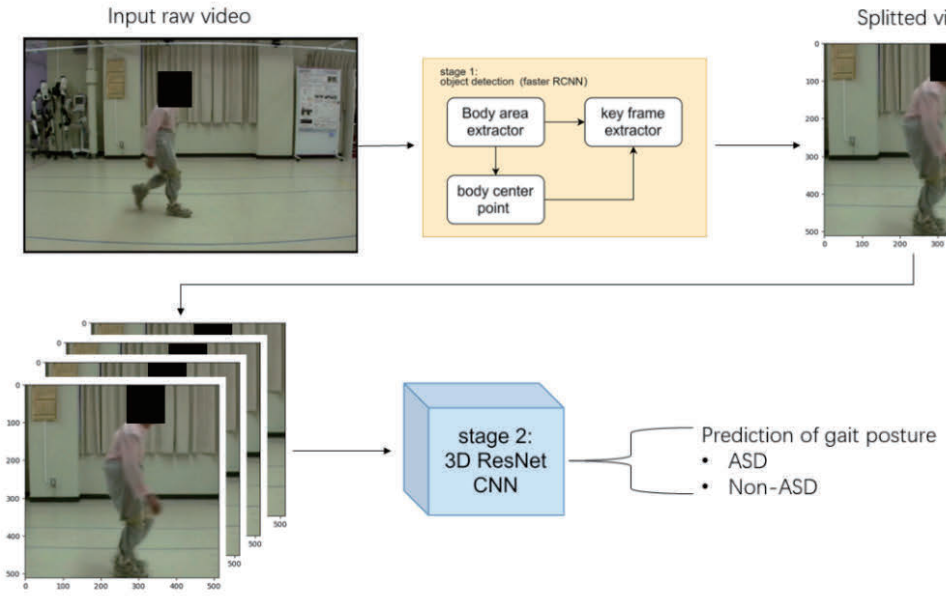
176　三次元動作解析と深層学習を用いた腰痛症の歩行解析	三浦　紘世

【目的】過去の研究より慢性腰痛を有する成人脊柱変形患者の歩行動作を評価することは、臨床的に有用な情報が得られる。しかし、三次元動作解析は機器が高価であることから、一般的な検査とするのは困難であり、歩行動作評価をより一般的に行えるように、本研究では、深層学習を用いた歩行動作評価により慢性腰痛症の病態診断を試みることを目的とした。

【方法】整形外科専門医が、成人脊柱変形（ASD）、腰部脊柱管狭窄症（LCS）、首下がり症候群（DHS）、変形性股関節症（HipOA）の 4 種類の疾患の診断のラベリングを行い、本研究では入院中に得られた患者の歩行動画より、短時間での入力動画から正確に疾患を分類できるシステムの構築を目指した。まずは、ASD と non-ASD の 2 つのクラスに分類するために、2 段階の解析手法を用いた。第 1 段階では、物体検出と姿勢推定法を用いて、患者の位置を特定し、患者がいる領域を抽出した。患者の歩行姿勢の動きは、ビデオフレーム全体の小さな領域を占めるため、また、患者の中には、転倒防止のために医師の付き添い歩きの必要があり、分類精度の悪化　に繋がり得るため、これらの手法を行った。この手法では、患者のおおよその位置を特定し、患者の体幹に関する重要な特徴を含むフレームを確保可能であった。続いて、第 2 段階では、3 次元 CNN ネットワークを用いて、抽出したフレームから空間的・時間的情報（動的運動）を取得し、患者の歩行姿勢から ASD と non-ASD を識別した。

【結果】我々が収集した 81 症例のビデオデータセットについて、CNN が RGB 動画において ASD と non-ASD を区別できるかを検証した結果、1 秒 8 フレームの動画から整形外科専門医のラベリングと比べて 75.53%の診断精度という比較的信頼できる結果を示した。

システム構成図



176. 三次元動作解析と深層学習を用いた腰痛症の歩行解析

三浦　紘世

筑波大学 医学医療系 整形外科

Key words：歩行姿勢解析，人間動作認識，コンピュータビジョン，3D CNN

緒　言

現在本邦では、超高齢社会により高齢者の腰痛症を診療する機会が急増している。厚生労働省が打ち出した健康日本 21 という施策の中で、高齢化社会において国民の健康の増進のためにロコモティブシンドローム（運動器症候群）の予防を挙げており、そのロコモティブシンドロームの原因として腰痛が注目されている［1，2］。また、腰痛は慢性的な運動器の痛みの中で最も頻度が高く、移動能力の低下をきたし寝たきりのリスクとなることから社会的問題となっている［3］。青壮年で多い急性腰痛と異なり、高齢者の場合には多くは慢性腰痛であり、その要因が多因子（椎間板変性、脊柱変形、腰部脊柱管狭窄症、サルコペニア、骨粗鬆症など）に渡り従来の画像検査（レントゲン、CT、MRI など「静的」な検査）では病態の診断が青壮年に比べて困難なことが多い。その病態の違いは、手術の方法や、保存的加療の選択（薬物や運動療法）など治療戦略にも大きく関わる。腰痛症では、安静時は症状が楽な状態であり、歩行など動作によって症状が顕在化もしくは悪化が見られるのに対して、従来の画像検査では、動作により悪化する際の機序の評価が十分には行えないことが病態の診断が困難な理由である。従って、高齢者の腰痛症に対して「動的」な評価を加えることが病態の診断に重要な役割を果たすと推測される。

筆者らは、これまで歩行時の「動的」な評価を行うため、三次元動作解析と筋電図解析を同期させる新たな歩行解析システムを開発し、歩行中の体幹・下肢関節の三次元における運動角度の「動的」な評価、さらに歩行周期内や連続歩行動作中の「動的」な筋活動変化を同時に捉えることを可能とした［4］。一方で、本手法で用いた三次元動作解析や筋電図解析はモーションキャプチャーのためのカメラやセンサーを多く必要とするため、システムの設置や費用の面から、一般的検査として臨床に適用するのは非常に困難である。そこで、深層学習を用いた歩行動作解析を組み合わせることを着想し、歩行動画データから、深層学習により腰痛の病態診断に補助的に有用なプログラム作成できるのではないかと仮説した。精度の向上により病態診断に十分な診断精度が担保されれば、患者自身や診療所から歩行動画データを送付するだけで、遠隔による診断で行うことも可能となれば、その意義は大きい。

方　法

1. 関連研究

医療用歩行動作の動画を分類する課題（action recognition）については、ある時刻の断層画像や静止画を対象とした研究が主となっており、動画像を対象とした融合研究はまだ少ない状況である。良く使用されている手法としては事前に動画をスライスして、普通の画像としてモジュールに流し込んで学習させるという研究が多い。現在では、YOLO をベースにした研究や、Faster RCNN と組み合わせた位置付けの研究が多く行われている。しかし、学習前に動画をスライスに分割すると、動画のフレーム間の情報が失われ、結果としてモデルの精度が低下してしまう。医療での動画、特に脊椎疾患における動画を用いた研究は、いまだごく少数である。主流の動画タスクのデータセットは、日常生活の動画（UCF101、Kinetics-700、YouTube8m など）を題材にしている

ことから、診療時の歩行動画のデータセットに関して、本研究はユニークと言える。動画の分類タスクに関しては、2 つの方法が一般的である。一つ目は 3D CNN［5］をベースにしたもの、もう一つは光ストリーミング情報（flow）を利用した 2 ストリーム（two-stream）［6］という方法である。本研究では、3D CNN に焦点を当て、我々のオリジナルデータセットに基づいて一連の改良を行い、提案モジュールが診療時の歩行動画データセットで適用した。

2. 提案手法（図 1）

まず、動画を画像に分割して学習させるのではなく、動画の情報をメタデータとして学習させることにした。メリットとしては、オプティカルフロー（flow）の情報を失うことなく、モデルの学習が可能であることが挙げられる。次に、動画データを使用することで、（動画を画像に分割する場合と比較して）データセットの準備に必要な作業量が減り、データセットをインポートする際の柔軟性が高まる（学習用に異なるフレームレートを選択することが可能である）。次に、患者の歩行動画に対して行うタスクであるため、動画の編集時にどの部分が必要で、どの部分が不要かを考えることが重要であった。患者の歩行速度の違いや、歩行姿勢もモデルに影響を与えるため、データセットの準備に注力した。

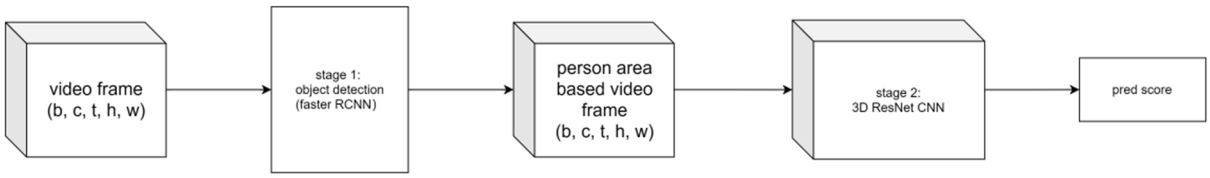


図 1. 提案手法

提案手法は、前処理ネットワークと動き分類ネットワークの 2 つのステージで構成される。

ステージ 1：前処理ネットワークでは、患者の位置を抽出して、無関係の影響を除去するために、画像サイズを所定のサイズに調整し、画像内の位置を調整する。

ステージ 2：モーション分類ネットワークは、3D CNN ネットワークを使用し、病型を予測する。

a）Preprocessing networks

本研究の患者の歩行動画撮影は、片側に固定された 1 台の RGB カメラを用いて、患者が固定カメラの前を往復して歩く条件であった。しかし、映像には、テーブル、扇風機、ホワイトボード、カーテンなど、患者の身体以外の無関係な情報が含まれてしまうことと、患者の歩行能力によっては、転倒防止に医師が付き添って歩く必要があるため、フレーム内に医師がいるシーンが発生することが問題となった。前処理ネットワークでは，前者の問題を身体領域抽出器によって解決し，後者の問題を患者が医師と一緒に歩くバウンディングボックスを抽出する方法によって解決した。

b）Motion classification network

この段階で、ResNet 形式［7］の 3D CNN をベースに患者の動き分類ネットワークを構築した。動き分類ネットワークは、第 1 段階から抽出された一連のフレーム画像（つまり、患者領域のみが使用されている）を受け取った。

結　果

1. データ構成（図 2）

データセットとしては、81 人の被験者が 4 種類の歩行病室内を行う様子を収録した。動画全体は 1 分～5 分長、フルフレームレート（30FPS または 60FPS）で記録された。各映像の解像度は 1020×1080 の HD に設定された。フォーマットは MP4、各映像のコーデックは H.264 に設定された。

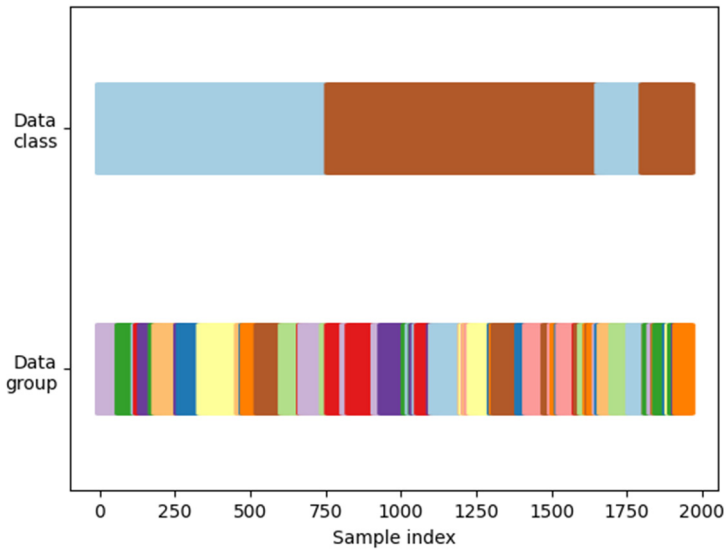


図 2. データセット全体の構成

上の data class は今回分類する種類、2 種類ラベルがあって、ASD と non-ASD である。下の data group は異なる患者数である。異なる色は異なる患者を示して、今回は 81 症例、動画数全体は 1,957 本あった。

2. 実験準備

提案手法の有効性を評価するために、分類実験を行った。整形外科専門医の診断により、各患者に病名ラベルを貼付し、評価ライセンスで訓練された専門家によるアノテーションに基づいて、各映像にラベルを付与した。成人脊柱変形（ASD）、首下がり症候群（DHS）、腰部脊柱管狭窄症（LCS）、変形性股関節症（HipOA）のラベルを用いた。さらに、ASD の診断能を評価するため、ASD は ASD、DHS、LCS、HipOA は non-ASD で 2 つの分類にした。分類性能を評価するために、平均精度スコアの平均値を使用して精度スコアを算出した。

オーグメンテーションには、空間的ジッタリングと時間的ジッタリングの両方を使用した。各ビデオを 30FPS に統一し、安定したトレーニングを行った。また、トレーニング中に時間的なジッターを適用し、ランダムに開始時刻を選択し、フレーム、T フレームをデコードした。

3. 混同行列（図 3）

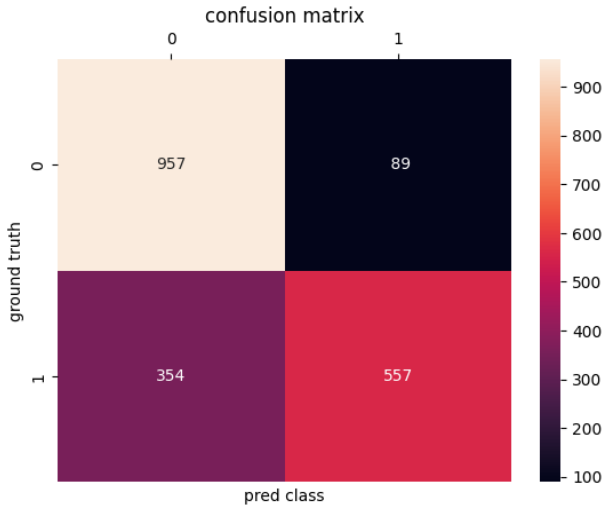


図 3. 実験結果の混同行列

左が実ラベルで、上のレジスタがモデル推論の結果。0 は ASD 症例、1 は non-ASD 症例を示す。右側はカラーバーを表し、薄い色は数字が大きく、濃い色は数字が小さいことを示す。図から、提案手法は病気を分類する能力をある程度持っていることが判明した。

4. AUCROC（Area Under the Receiver Operating Characteristic Curve）（図 4）

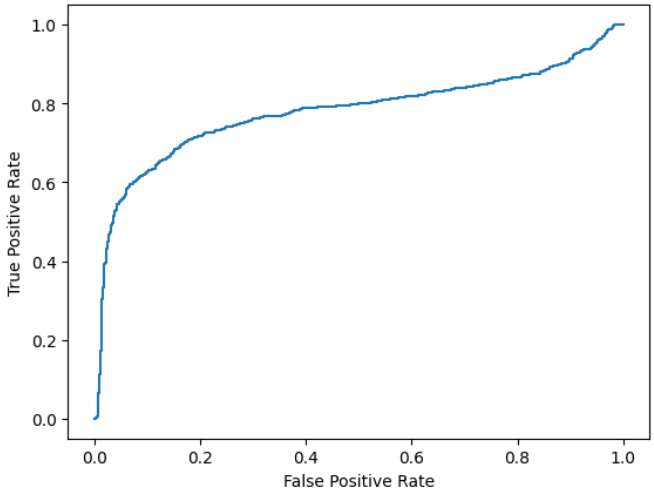


図 4. ROC 曲線

左が true positive rate で、下が false positive rate である。AUCROC（Area Under the Receiver Operating Characteristic Curve）は 0.7786 であった。AUROC スコアは、ROC 曲線を複数の閾値に対するモデルの性能を同時に表す一つの数値にまとめたものである。特に、AUROC スコア 1 が満点で、AUROC スコア 0.5 がランダム推測に対応する。

5. Attention map（図 5）

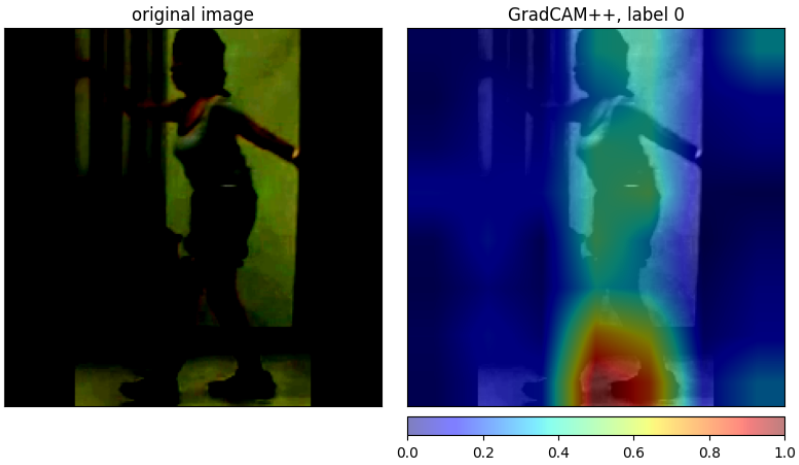


図 5. モデル可視化結果

左が original image で、右が GradCAM＋＋適用した結果である。赤い部分は、モデルがより関心を寄せている場所を示している。その結果、下肢をより重視したモデルであることが判明した。

考　察

本研究では、歩行動画を元にした腰痛を有する患者の病態診断のサポートツール開発を目的に、深層学習を用いた歩行動作解析のプログラムとして 2 段階の分類法を提示した。まず、提案したメソッドに前処理ネットワークを導入し、映像から無関係な情報の除去を行い、続いて、歩行中の患者の時空間的な特徴表現を捉えるために、さらに構成されたビデオフレーム画像で ResNet スタイルの 3D CNN を導入した。実験結果では、無関係な要素へのオーバーフィッティングを抑制することが可能であった。また、Kinetics-400 を用いた転移学習実験では、学習した特徴が汎用的であることが示唆された。特に、動画分類の再学習を行うことにより、最も高い伝達学習性能を達成した。実験結果の混同行列より、最終的な分類精度は 1 秒 8 フレームの動画から 75.53%の精度という比較的信頼できる結果を得た。AUCROC（Area Under the Receiver Operating Characteristic Curve）は 0.7786 であった。モデルはある程度の分類能力を持つことが判明した。Attention map から、モデルの注目部分は患者の下肢を重視したモデルであることが判明した。

今後については、将来的に、ネットワーク構造を改良し、オプティカルフローを利用した 2 ストラムネットワークを使用することが期待される。また、タスクは歩行ベースであるため、特徴を抽出することでタスクのトレーニングができるようになると予想される。それにより、歩行動画を用いた深層学習を用いた歩行動作解析により、患者のさまざまな部位の情報から病態を分類することができる可能性がある。

共同研究者・謝辞

本研究の共同研究者は、筑波大学システム情報工学研究群の陳凱旭、家永直人、黒田嘉宏、筑波大学人間総合科学研究科疾患制御医学専攻・整形外科の朝田智之、坂下孝太郎である。

文　献

1) Iizuka Y, Iizuka H, Mieda T, Tajika T, Yamamoto A, Takagishi K. Population-based study of the association of osteoporosis and chronic musculoskeletal pain and locomotive syndrome: the Katashina study. J Orthop Sci. 2015 Nov;20(6):1085-9. doi: 10.1007/s00776-015-0774-9. Epub 2015 Sep 7. PMID: 26345242.

2) Hirano K, Imagama S, Hasegawa Y, Ito Z, Muramoto A, Ishiguro N. The influence of locomotive syndrome on health-related quality of life in a community-living population. Mod Rheumatol. 2013 Sep;23(5):939-44. doi: 10.1007/s10165-012-0770-2. Epub 2012 Sep 21. PMID: 22996232.

3) Nakamura M, Nishiwaki Y, Ushida T, Toyama Y. Prevalence and characteristics of chronic musculoskeletal pain in Japan. J Orthop Sci. 2011 Jul;16(4):424-32. doi: 10.1007/s00776-011-0102-y. Epub 2011 Jun 16. PMID: 21678085; PMCID: PMC3140943.

4) Miura K, Kadone H, Koda M, Abe T, Funayama T, Noguchi H, Mataki K, Nagashima K, Kumagai H, Shibao Y, Suzuki K, Yamazaki M. Thoracic kyphosis and pelvic anteversion in patients with adult spinal deformity increase while walking: analyses of dynamic alignment change using a three-dimensional gait motion analysis system. Eur Spine J. 2020 Apr;29(4):840-848. doi: 10.1007/s00586-020-06312-y. Epub 2020 Jan 31. PMID: 32002700.

5) Ji S, Yang M, Yu K. 3D convolutional neural networks for human action recognition. IEEE Trans Pattern Anal Mach Intell. 2013 Jan;35(1):221-31. doi: 10.1109/TPAMI.2012.59. PMID: 22392705.

6) Simonyan K, Zisserman A. Two-stream convolutional networks for action recognition in videos. Advances in neural information processing systems. 2014;27. doi.org/10.48550/arXiv.1406.2199

7) He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition 2016 (pp. 770-778). doi.org/10.48550/arXiv.1512.03385