Revista sítio novo
Instituto Federal do Tocantins

Modelo de Rede Neural para avaliação desportiva

Jeremias Fontinele da Silva [(1)]
Carlos Alberto de Sousa Parente Rodrigues [(2)]
Carlos Henrique Corrêa Tolentino [(3)] e
Wandro Bequiman Maciel [(4)]

Data de submissão: 7/1/2021. Data de aprovação: 29/4/2021.

Resumo – Trata-se de uma pesquisa que objetiva comprovar a possibilidade de utilizar um Modelo em Rede Neural capaz de avaliar o movimento desportivo. O diferencial deste estudo encontra-se no fato de a máquina-servidor ser totalmente em *cloud*, o que torna viável sua futura utilização por dispositivos *mobile* devido ao não comprometimento da capacidade de processamento destes. Outro fato relevante é o emprego de duas Redes Neurais (Convolucional e Recorrente) na análise do movimento desportivo. Quanto à metodologia investigativa, este trabalho tem por alicerce uma revisão bibliográfica sobre Rede Neurais e estimativa de pose humana. Isso significa que a fundamentação teórica foi desenvolvida tendo por suporte estudos já realizados e publicados sobre a temática. Como resultado, conclui-se que a utilização de redes convolucionais para a análise de estimativa de pose possui uma acurácia satisfatória, mas que carece de tratamento de ruídos para que a análise da execução do movimento desportivo possa ser feita de fato.
Palavras-chave: Rede Neurais. Estimativa de Pose. Movimento Desportivo.

Neural Network model for sporty evaluation

Abstract – This is a research that aims to prove the possibility of using a Neural Network Model capable of evaluating sporty movement. The differential of this study is in the fact that the server machine is completely in the cloud, which makes its future use via mobile devices viable, due to the non-compromising of their processing capacity. Another relevant fact is the use of two Neural Networks (Convolutional and Recurrent) in the analysis of the sport movement. As for the investigative methodology, this work is based on a bibliographic review on Neural Network and human pose estimation. This means that the theoretical foundation was developed based on studies already carried out and published on the subject. As a result, it is concluded that the use of convolutional networks for the analysis of pose estimation has a satisfactory accuracy, but that it needs noise treatment so that the analysis of the sporty movement execution can be done in fact.
Keywords: Neural Network. Pose Estimation. Sporty Movement.

Introdução

Um estudo da Organização Mundial da Saúde (OMS), publicado na revista *The Lancet Global Health*, revela que o sedentarismo no mundo cresceu nos últimos 15 anos (GUTHOLD *et al.*, 2018). A pesquisa levou em conta dados de 2001 até 2018 de vários países, incluindo o Brasil. De acordo com os padrões estabelecidos pela OMS na pesquisa (150 minutos de

[1] Mestre em Ensino em Ciências e Saúde (UFT); Pós-Graduando em Telemática do *Campus* Palmas (IFTO). *jeremias.fontinele@mail.uft.edu.br. ORCID: https://orcid.org/0000-0002-2412-3473.
[2] Mestre em Engenharia Elétrica e da Computação (UFG); Pós-Graduando em Telemática do *Campus* Palmas (IFTO). *carlos.ccomp@gmail.com. ORCID: https://orcid.org/0000-0002-3498-0313.
[3] Mestre em Ciência da Computação (UFSC) e docente da Universidade Estadual do Tocantins e do *Campus* Palmas (IFTO). *chtolentino@ifto.edu.br. ORCID: https://orcid.org/0000-0002-7222-6880.
[4] Bacharel em Ciência da Computação (UFT). Pós-Graduando em Telemática do *Campus* Palmas (IFTO). *wandrobeckman2@gmail.com. ORCID: https://orcid.org/0000-0003-4907-891X.

atividade de intensidade moderada ou 75 minutos de exercícios em alta intensidade por dia), aproximadamente 47% da população brasileira não se exercita o suficiente. Os números do Brasil surpreenderam e ficaram acima de países como Estados Unidos e Reino Unido, com 40% e 36%, respectivamente.

Por outro lado, é sabido que a prática de exercícios sem o devido cuidado com as articulações e tendões, bem como o mal posicionamento postural durante a execução do movimento, é um hábito que pode acarretar lesões a médio e longo prazo.

Quanto à evolução tecnológica, conforme Schwab (2019), vivemos a Quarta Revolução Industrial (Indústria 4.0), que é caracterizada pelo domínio de um conjunto de tecnologias, como: robótica; inteligência artificial (IA); realidade aumentada, virtual e mista; big data (análise de volumes massivos de dados); nanotecnologia; impressão 3D (manufatura aditiva); biologia sintética (SynBio); Sistemas Ciber-Físicos (CPS); computadores quânticos; teletransporte quântico e a chamada internet das coisas (IoT). Esse conjunto é denominado de tecnologias disruptivas, haja vista que provocaram uma ruptura com os padrões, modelos ou tecnologias já estabelecidas no mercado (CHRISTENSEN, BOWER, 1995).

Nesse contexto de busca por uma higidez física mais saudável e eficiente, bem como considerando as tecnologias digitais e a computação ubíqua vigentes na Quarta Revolução Industrial, surge a seguinte indagação, que problematiza este artigo: *é possível um Modelo de Rede Neural capaz de avaliar o movimento desportivo?*

A relevância deste artigo reside na proposição de utilização das tecnologias disruptivas, mais especificamente Inteligência Artificial (IA) e armazenamento em nuvem (*cloud*), no auxílio aos profissionais desportivos e/ou seus treinadores quando em atividade física, pois, ao executar-se determinado exercício utilizando-se a postura e a técnica corretas, é possível equilibrar músculos e ossos de forma a proteger as estruturas de suporte, diminuindo a sobrecarga nas articulações e permitindo a eficiência máxima no movimento em qualquer atividade desportiva.

A aplicabilidade reside na possibilidade de utilização da IA em aplicativos (apps), softwares ou, diretamente, na World Wide Web (WWW) para ajudar a avaliar algum movimento executado pelo usuário em momentos nos quais não se possa contar com supervisão humana. Além disso, no caso de um treinador, a máquina-servidor é capaz de informar dados que podem ser transpostos, utilizando software específico, para o formato de relatórios detalhados sobre o movimento avaliado.

O diferencial deste estudo encontra-se no fato de a máquina-servidor ser totalmente em *cloud* (nuvem), o que torna viável sua futura utilização por dispositivos *mobile* devido ao não comprometimento da capacidade de processamento destes. Outro fato relevante é o emprego de duas Redes Neurais (Convolucional e Recorrente) na análise do movimento desportivo.

Portanto, o objetivo geral deste artigo é comprovar a possibilidade de um Modelo em Rede Neural capaz de avaliar o movimento desportivo por meio de uma máquina-servidor totalmente em *cloud*.

Materiais e métodos

A metodologia investigativa (bibliográfica e documental) que alicerça a revisão da literatura deste artigo foi realizada sob a temática “Rede Neurais e estimativa de pose humana”. Isso significa que a fundamentação teórica foi desenvolvida tendo por suporte estudos já realizados e publicados sobre a temática.

A busca foi realizada por meio da Comunidade Acadêmica Federada (CAFe), a qual permite acesso a bases de dados, como Scielo, ERIC (Education Resources Information Center), Periódicos CAPES e o buscador Google Acadêmico, tendo sido utilizados os seguintes operadores lógicos booleanos na busca: “Rede Neural” AND “Movimento”; “Rede Neural” AND “Estimativa de pose”; “Rede Neural” AND “Atividade Física”.

Para alcançar o objetivo geral desta pesquisa fez-se uma revisão da literatura e o delineamento experimental de quatros objetivos específicos: 1 – Definir as técnicas a serem empregadas; 2 – Selecionar o conjunto de dados inerente ao treinamento de aprendizagem videogramétrica; 3 – Implementar a Rede Neural; e 4 – Validar os resultados obtidos.

Revisão da Literatura

Este trabalho tem por foco a implementação de uma rede neural artificial que possa analisar o movimento corpóreo humano quando em atividade desportiva, e os principais fatores que impactam na qualidade e acurácia nessa análise são: as técnicas utilizadas; as tecnologias de rastreamento 3D utilizadas; e os métodos de interpretação dos dados usados na inferência dos movimentos.

Nesse sentido, faremos uma breve explanação sobre os sistemas de captura de movimentos, estimativa de pose, modelos de estimativa de pose e rede neural convolucional e recorrente (LSTM). Por se tratar de um artigo científico, não nos aprofundaremos acerca de todos os conceitos, de modo que serão ressaltados apenas os aspectos necessários para o entendimento deste artigo.

Sistemas de capturas de movimentos

Um sistema de captura de movimento é o processo de gravar um evento em movimento ao vivo e traduzi-lo, em termos matemáticos, por meio do rastreamento de uma série de pontos-chave no espaço ao longo do tempo e combinando-os para obter uma representação tridimensional (3D) única do desempenho (MENACHE, 2000, p. 2). Ou seja, é a tecnologia que permite o processo de tradução de uma performance ao vivo em uma performance digital.

Entendem-se por pontos-chave as áreas que melhor representam o movimento das diferentes partes móveis do evento observado. Para um ser humano, por exemplo, alguns dos pontos-chave são as articulações, que atuam como pontos de articulação e conexões para os ossos.

A localização de cada um desses pontos é identificada por um ou mais sensores, marcadores ou potenciômetros que são colocados no sujeito e que servem, de uma forma ou de outra, como condutores de informações para o dispositivo principal de coleta (MENACHE, 2000). Esse sistema de captura de movimento é complexo e pode ser fragmentado em: inicialização; rastreamento; estimação de pose e reconhecimento do movimento.

Quadro 1 - Sistema de captura de movimento

ETAPA	DESCRIÇÃO
Inicialização	Abrange as ações necessárias para assegurar que o sistema inicie a sua operação com uma correta interpretação da cena atual.
Rastreamento	É a detecção e a localização recursiva de objetos ou, mais geralmente, de padrões em sequências de imagens (vídeos). Em sua forma mais simples o rastreamento compõe-se de um modelo de observação do espaço sensorizado, um modelo de representação do objeto rastreado e um algoritmo de rastreamento.
Estimação Pose e Reconhecimento	São, respectivamente, a identificação de como um corpo humano (ou outro objeto) está configurado no espaço 3D (ângulos e orientações de juntas); e a classificação do tipo de movimento capturado.

Fonte: (SIMAS *et al*., 2007, p. 60).

Encontramos na literatura computacional diferentes formas de classificação dos métodos de Captura de Movimentos, sendo as principais: síncronos ou assíncronos, ativos ou passivos, marcadores existentes ou ascendentes, ou conforme os princípios físicos utilizados.

Estimativa de pose (movimento)

A estimativa de pose humana é um problema difícil, porque o corpo humano tem muitos graus de liberdade e articulações, o que torna a captura complexa. Também é de suma importância conseguir superar algumas das dificuldades relacionadas à variação da pose devido a roupas, formato do corpo, tamanho, iluminação, entre outros (BRITO, 2019, p. 15).

Mesmo diante de tamanha complexidade inerente ao movimento humano, é exigido que os resultados obtidos sejam eficazes mesmo que partes do corpo se sobreponham, como, por exemplo, a mão de uma pessoa cobrindo parcialmente uma articulação ou algum outro membro do corpo humano.

Contudo, os problemas da estimativa de pose podem ser classificados quanto a sua representação em dimensões, que podem ser 2D (duas dimensões - x, y) ou 3D (três dimensões - x, y, z), conforme nos ensina Brito (2019):

> Os problemas de *pose estimation* podem ser divididos em dois tipos: 2D *pose estimation* e 3D *pose estimation*. O primeiro estima uma posição 2D (x, y) de coordenadas para cada junta no espaço na qual a pessoa se situa, a partir de uma imagem. Já o segundo estima uma posição 3D (x, y, z) de coordenadas neste mesmo espaço métrico a partir de uma imagem (BRITO, 2019, p. 15).

Existe uma variedade de técnicas para trabalhar com a estimativa de movimento, contudo há dois artífices matemáticos que são cotidianamente utilizados nesse processo: matrizes e tensores. Nesse aspecto, os cálculos matemáticos são fundamentais para as estimativas de pose e para solucionar o problema de visão computacional.

Modelo para estimativa de pose (OpenPose)

A partir dos avanços tecnológicos na área da inteligência artificial e aprendizagem de máquina, diversas técnicas de processamento de vídeo, sem a necessidade de utilizar marcadores, com o viés de extrair parâmetros destes, foram propostas, como a OpenPose, escolhida para este trabalho.

O OpenPose é um modelo *open source* para estimativa de pose desenvolvido por Cao *et al*. (2017) que utiliza a sistemática *bottom-up*, ou seja, tem por ponto de partida o conjunto de todos os pontos detectados para efetuar a montagem das poses individuais.

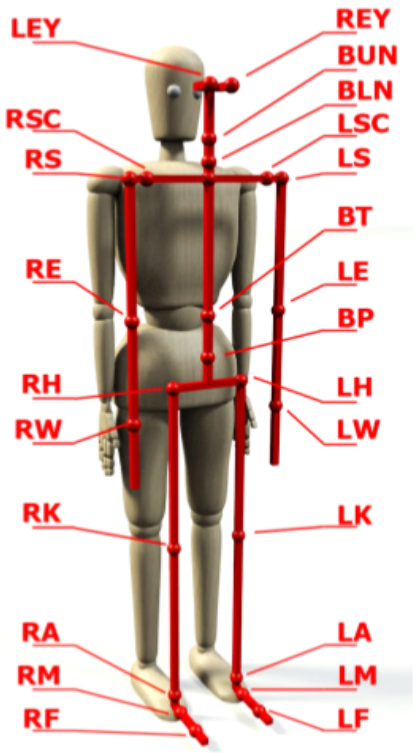
Segundo Cao *et al*. (2017), essa abordagem é justificada pelos problemas com a opção *top-down*, como por exemplo a necessidade inicial de segmentar a imagem entre os diferentes indivíduos — o que torna o processo crítico, haja vista que a qualidade geral da pose dependerá da qualidade do recorte da imagem —, o fator tempo utilizado nessa segmentação e o fato de ela não resolver totalmente o problema de haver oclusões ou sobreposição de membros.

Neste estudo, considerando o movimento escolhido — polichinelo —, o OpenPose opera com uma rede de dois estágios que culminam na identificação de 15 articulações corporais da imagem. Conforme Leite (2020, p. 21), "no primeiro estágio, o método cria mapas de confiança das posições das juntas e o segundo estágio prediz campos de afinidade entre as partes encontradas". Ainda segundo o autor, a afinidade é representada por um vetor 2D, que codifica a posição e a orientação de cada membro do corpo.

Para Gong *et al* (2016), em visão computacional, o corpo humano pode ser considerado como um objeto articulado que consiste em parte móveis rígidas, conectadas através das articulações.

Quanto aos graus de liberdade e limites de rotação para cada articulação do corpo humano, utilizamos o modelo em corpo rígido do corpo humano desenvolvido por Terlemez *et al*. (2014) para o projeto Master Motor Map (MMM).

Figura 1 - Graus de liberdade e limites de rotação para cada articulação do corpo humano.



Joint	DoF	X-Limits	Z-Limits	Y-Limits
LF/RF	1+1	[-30°,45°]	-	-
LM/RM	1+1	-	[-30°,45°]	-
LA/RA	3+3	[-40°,30°]	[-30°, 30°]	[-20°, 20°]
LK/RK	1+1	[-130°,0°]	-	-
LH	3	[-50°,95°]	[-45°,45°]	[-20°,65°]
RH	3	[-50°,95°]	[-45°,45°]	[-65°,20°]
LW	2	[-30°,20°]	[-70°,50°]	-
RW	2	[-30°,20°]	[-50°,70°]	-
LE/RE	2+2	[0°,160°]	[-90°,90°]	-
LS	3	[-70°,190°]	[-70°,60°]	[0°,160°]
RS	3	[-70°,190°]	[-60°,70°]	[-160°,0]
LSC/RSC	2+2	-	[-20°,20°]	[-20°,20°]
LEY/REY	2+2	[-60°,60°]	-	[-60°,60°]
BUN	3	[-20°,30°]	[-20°,20°]	[-15°,15°]
BLN	3	[-45°,15°]	[-15°,15°]	[-20°,20°]
BT	3	[-35°,27°]	[-36°,36°]	[-20°,20°]
BP	3	[-50°,35°]	[-45°,45°]	[-20°,20°]

Fonte: MACEDO; SANTOS. 2019, p. 15.

No modelo cinemático do corpo humano acima encontramos os limites angulares de acordo com cada eixo (X-Limits, Z-Limits e Y-Limits) por articulação, outrossim o número de graus de liberdade (DoF), conforme desenvolvido no projeto Master Motor Map (MMM). As siglas seguem os termos originais, por exemplo: RK = Knee (joelho) Right (direito) e LK = Knee (joelho) Left (esquerdo).

Segundo Macedo e Santos (2019, p. 15), o processo de estimação de pose, no contexto corpo humano, pode ser definido como "a identificação da posição bidimensional ou tridimensional desses pontos de articulação no corpo, de forma a realizar suas interconexões e formar um sticker, ou boneco, com a pose estimada".

Rede Neural Convolucional (CNN)

As origens das Redes Neurais Convolucionais (CNN) remontam à década de 1970. Contudo o conceito moderno acerca das redes convolucionais surgiu em um artigo datado de 1998, intitulado *Gradient-based learning applied to document recognition*[5], escrito pelos cientistas da computação Yann LeCun, Léon Bottou, Yoshua Bengio e Patrick Haffner. Para LeCun, "a inspiração neural [biológica] em modelos como redes convolucionais é muito tênue. É por isso que eu os chamo de redes convolucionais e não redes neurais convolucionais, da mesma forma os nós eu chamo de unidades e não neurônios".

As CNN utilizam, além dos parâmetros da combinação linear, parâmetros de filtros convolucionais que são implementados nas primeiras camadas da rede. Segundo Leite (2020, p. 16) as operações de convolução, além de serem um subtipo de redes profundas, permitem que a rede aprenda características de baixo nível nas primeiras camadas e combine-as nas camadas seguintes para aprender características de alto nível.

[5] http://yann.lecun.com/exdb/publis/pdf/lecun-98.pdf

As redes neurais convolucionais baseiam-se sobre três pilares: campos receptivos locais; pesos compartilhados e *pooling*. Quanto ao processamento, temos três etapas: entrada de dados (*input*), aprendizado de características (*feature learning*) e classificação (*classification*).

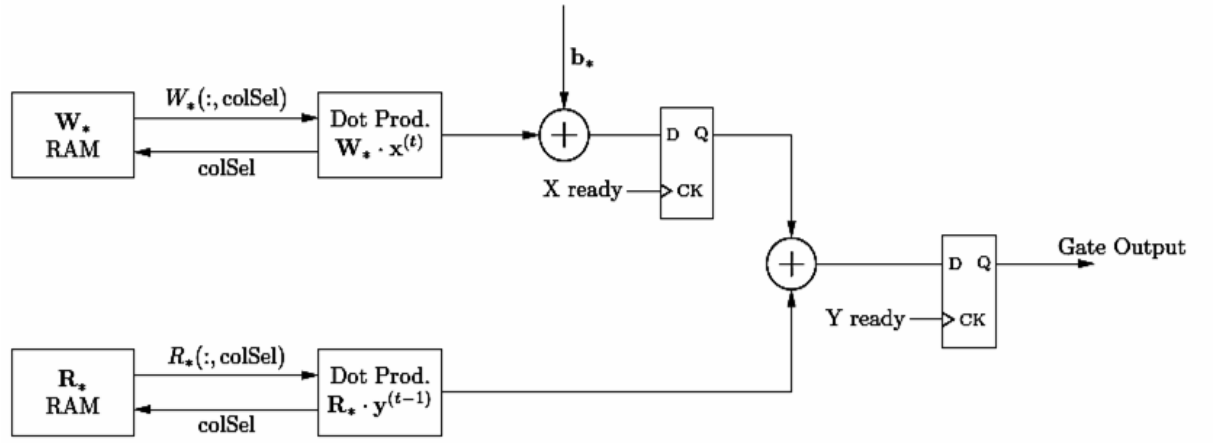
Redes Neurais Long Short Term Memory (LSTM)

As Redes Neurais Recorrentes (Recurrent Neural Network – RNN) superam o problema de reconhecimento de sequência, contudo elas falham em reter dependências de longo prazo. Evidente que o processo de treinamento com pesos é, em si, uma forma de memória, mas o problema é que a atualização do peso é muito mais lenta do que as ativações, e, portanto, essa memória retém apenas dependências de curto prazo, devido ao Problema de Vanishing Gradients.

Para superar o problema de "não lembrar as dependências de longo prazo", Sepp Hochreiter e Jürgen Schmidhuber, em 1997, propuseram uma nova abordagem para os RNNs, denominada *Long Short Term Memory*. A LTSM está diretamente relacionada à forma de pensar da máquina, ou seja, lembrar informações por longos períodos de tempo. São redes com *loops*, permitindo que as informações persistam.

Fonseca (2016, p. 15) nos diz que "as redes LSTM são, hoje em dia, um dos algoritmos de última geração em aprendizado profundo, e seu desempenho é superior ao de outros tipos de RNNs e modelos de Markov ocultos". O LSTM tem a capacidade de remover ou adicionar informações, cuidadosamente, reguladas por estruturas chamadas portas. Existem três portas LSTM: porta de entrada, porta de saída e porta esquecer. Os módulos porta são responsáveis por produzir os vetores de sinais internos para $\mathbf{z}^{(t)}$, $\mathbf{i}^{(t)}$, $\mathbf{f}^{(t)}$, $\mathbf{o}^{(t)}$. Uma LTSM em propagação direta, quanto às portas, pode ser representada conforme a figura abaixo:

Figura 2 - Exemplo de vetores em uma LTSM.



Fonte: FONSECA. 2016, p. 29.

A leitura correta da figura 2 é a seguinte: I - Multiplica-se uma matriz (**W**) pelo vetor de input $\mathbf{x}^{(t)}$; II - Multiplica-se uma matriz (**R**) pelo vetor de input $\mathbf{y}^{(t-1)}$; e III - Some o vetor de polarização (**b**) aos resultados de produto escalar do vetor-matriz restante.

Seleção da técnica e do movimento desportivo

O polichinelo foi escolhido como movimento desportivo alvo de análise neste trabalho, pois se trata de um movimento que não necessita de equipamentos especializados e é bastante praticado pelo público em geral. Além disso, a parte mais significativa do movimento ocorre em apenas um plano, sendo possível sua avaliação sem depender de técnicas de reconstrução 3D a partir de 2D. Esse movimento também pode ser filmado, sem haver problemas com oclusão de partes relevantes do corpo do atleta, para a análise do movimento.

A respeito das técnicas computacionais, este trabalho propõe a utilização de duas redes de aprendizado de máquina. A primeira, convolucional, recebe via API a sequência de imagens

do cliente e a processa, encontrando pontos relevantes para a estimativa de pose (punhos, cotovelos, ombros, extremos do quadril, entre outros). A segunda, recorrente, analisa o histórico de posição e ângulos entre esses pontos durante o movimento para avaliar o exercício executado pelo usuário.

A aplicação de duas técnicas se dá pelo fato de o trabalho ter duas tarefas principais, sendo a primeira a extração de elementos que caracterizam a pose do atleta em cada quadro de uma captura em vídeo e a segunda a análise da série temporal de poses durante toda a execução da atividade. Ao dividir o problema, foi possível encontrar na literatura métodos que lidam com cada etapa da análise do movimento de forma mais especializada.

A primeira técnica escolhida foi a rede convolucional, e sua tarefa é extrair as posições de pontos específicos de controle que compõem a pose de um ser humano sendo filmado, sem considerar informações temporais de sequências de imagens ou analisar a qualidade do movimento.

Para este trabalho, foi utilizado o modelo OpenPose, presente na literatura e disponível para uso. Esse modelo retorna 15 pontos cartesianos, nas coordenadas da imagem, a partir de uma imagem ou quadro de um vídeo, conforme exemplifica a figura 4. A utilização de um modelo já treinado e consolidado traz como principal vantagem a segurança de empregar um método confiável para a extração das informações mais relevantes para a análise do movimento.

Figura 3 - Pontos de controle representando a pose de um atleta retornados pelo OpenPose

Fonte: Autores.

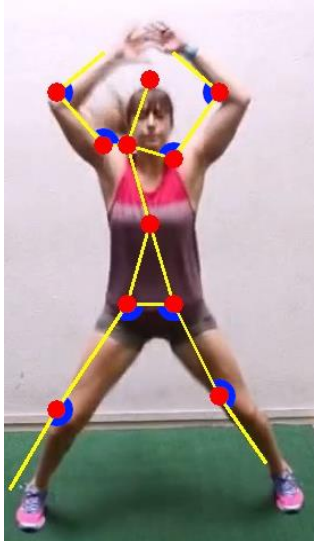
Após a extração dos pontos de controle da imagem, foi necessário representar a informação de pose em um formato mais adequado para seu posterior processamento. Dessa forma, para o movimento polichinelo, foram escolhidos oito ângulos relevantes para sua posterior análise, sendo eles o ângulo entre ambos os cotovelos, ombros e pernas em relação ao quadril e aos joelhos, conforme ilustrados em azul na figura 3.

Esse passo de extração de características traz como principais benefícios a representação da pose do atleta em um formato que não dependa diretamente do sistema de coordenadas da imagem, o que permite que essa técnica seja utilizada em diferentes configurações de captura de vídeo (resolução da câmera, distância do atleta para a câmera etc.).

Outra vantagem dessa representação é a diminuição da quantidade de entradas a serem processadas posteriormente por outros métodos, sem perda significativa de informações para a análise do movimento.

Dessa forma, essa etapa de extração de características tem como tarefa analisar as coordenadas de imagem enviadas pela rede convolucional de todos os quadros da captura de vídeo e transmitir para a próxima técnica um conjunto de 8 valores, representando os ângulos relevantes calculados, para cada quadro do movimento, conforme a figura 4 (arcos na cor azul).

Figura 4 - Ângulos relevantes para a extração de características para a aplicação da rede LSTM.



Fonte: Autores.

A segunda técnica de aprendizado de máquina escolhida por este trabalho é responsável pela análise de uma série temporal das características previamente extraídas e tem como finalidade a análise do movimento como satisfatório ou insatisfatório. Para esse passo, foi escolhida uma rede com duas camadas LSTM de 100 unidades. A entrada dessa rede são 100 amostras de descrições de pose regularmente espaçadas entre todo o conjunto do movimento. A saída dessa rede é um valor entre 0 e 1, sendo que valores que tendem a 1 indicam um movimento satisfatório.

Para a etapa de treinamento da rede LSTM, foram coletados 100 vídeos de polichinelos disponíveis na plataforma YouTube, sendo 90 de movimentos satisfatórios e 10 de movimentos insatisfatórios.

Matematicamente, temos: uma matriz de medidas no formato (2i x 4j); o ângulo (α) entre as articulações adota a distância em *pixels* e considera três pontos A, B e O (central) sendo, $a = \tan\left[\frac{Oy-By}{Ox-Bx}\right] - \tan\left[\frac{Oy-Ay}{Ox-Ax}\right]$ com limite $\int_0^{\pi} a$, conforme nos ensinam Macedo e Santos (2019, p. 33). Assim, o cálculo dos ângulos é realizado para cada *frame* $\mathcal{N}$ de entrada pela matriz $\mathcal{BJ}\times 2\mathcal{N}$, resultando em uma matriz $\mathcal{A}p\times n$, com *p* tendo dimensão 15, de acordo com o número definido de articulações avaliadas.

Quanto às métricas cinemáticas, utilizamos o trabalho desenvolvido por Victor Oliveira Corrieri de Macedo e Joyce da Costa Santos, por haver semelhança na análise realizada neste trabalho com o desenvolvido por eles. Entre as métricas, segundo Macedo e Santos (2019, p. 41), destacamos:

✓ Cadência (V): a velocidade de execução do movimento medida em abertura e fechamento por minuto, entendido neste estudo por SPM (em inglês, *Strokes Per Minute*) que se assemelha ao RPM (em inglês, *revolutions per minute*). Para calcular essa métrica, determinamos o período ou tempo de duração $\mathcal{T}_D$ (em segundos) do ciclo, de acordo com a seguinte relação: $V = \frac{60}{Td} = \frac{60}{fps \cdot Tc}$, onde *fps* é a taxa de amostragem do sinal, *Td* é o tempo de duração e $\mathcal{T}c$ é a quantidade de *frames* no ciclo;

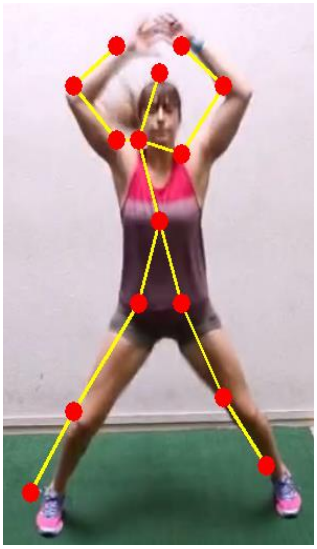
✓ Consistência da Cadência (V): avalia-se o gráfico da cadência estimada em função dos ciclos segmentados do vídeo;

✓ Consistência do Ângulo/Fase (V): avalia-se o gráfico de uma determinada angulação articular para uma determinada fase em função dos ciclos segmentados do vídeo.

O primeiro passo da validação dos resultados foi a análise dos pontos gerados pela rede convolucional OpenPose. A figura 5 exemplifica a saída da rede convolucional plotada com a imagem de entrada correspondente.

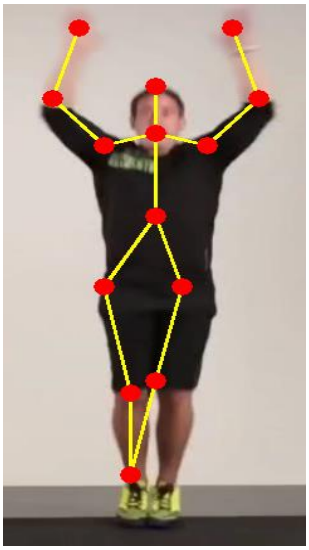
Durante a análise, foi observada a presença de ruídos que podem ou não influenciar na análise do movimento. Alguns casos, como por exemplo o demonstrado na figura 5, possuem um nível de ruído quase irrelevante para a análise. Porém, há situações em que o ruído influenciou a análise do movimento, como ilustrado pelas figuras 6 e 7, cujas saídas da rede OpenPose possuem ruído nas partes superior e inferior da pose do atleta, respectivamente.

Figura 5 - Saída da rede convolucional OpenPose.



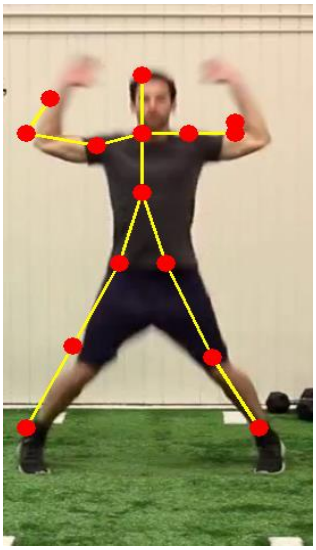
Fonte: Autores.

Figura 6 - Ruído na parte inferior.



Fonte: Autores.

Figura 7 - Ruído na parte superior.



Fonte: Autores.

Resultados e discussões

Um dos desafios encontrados nos primeiros vídeos analisados foi separar o atleta avaliado quando havia outras pessoas no campo de visão da imagem videogramétrica ou quando a técnica produzia falsos positivos indicando pessoas holograma (persona Ghost).

Esse problema ocorre devido ao *frame* e não em virtude do algoritmo, pois este fez a detecção e separação dos pontos de cada atleta imagem por imagem. Já quanto ao *frame*, verificamos que, em alguns casos, o atleta identificado ocupava a mesma posição no vetor resultante de outro, não sendo possível uma forma direta de distingui-las.

Entretanto, esse problema pode ser resolvido aplicando um método de extração da pessoa principal, citado por Macedo e Santos (2019, p. 29), que "consiste em calcular a mínima área retangular ocupada pelas articulações estimadas de cada pessoa. A partir dos pontos detectados de cada pessoa, sua área [Ap] pode ser calculada por meio da seguinte equação": $AP = (xmax - xmin) \times (ymax - ymin)$.

> Onde $xmax$ representa o máximo valor no eixo x, dentre as coordenadas dos pontos de articulação encontrados, e $xmin$ representa o valor mínimo, seguindo a mesma lógica para o y. Uma vez calculada a área de ocupação de cada pessoa, o vetor de pontos é reorganizado e a pessoa com maior área é separada para análise, produzindo uma matriz no formato $BJ \times 2 \times N$, a qual contém informação cinemática de apenas uma pessoa, em geral, da pessoa no plano central do vídeo. (MACEDO; SANTOS. 2019, p. 29).

Considerando o movimento do polichinelo, foram definidos quinzes pontos de articulação para serem considerados na análise: ombro; cotovelo; pulso; quadril; joelho e tornozelo, além do esterno como ponto de flexão do tronco em curvatura para frente. O movimento foi reduzido ao plano sagital e, devido à execução frontal do movimento, foi considerada apenas a parte frontal do corpo. Esses pontos estão demonstrados nas figuras acima, que representam um movimento de polichinelo no plano sagital frontal.

No tocante aos falsos negativos da rede, ou seja, os pontos de articulação não identificados, verificamos, a partir dos testes realizados, entre outras possíveis, duas causas que consideramos como principais para a ocorrência desse problema: 1- a geração, pelo mapa de probabilidade, de valores abaixo do *threshold*, predeterminado, na região articular, o que, concluímos, pode ocorrer ou por oclusão ou por limitações da rede em condições específicas;

2 – por erro de caracterização, ou seja, quando a rede detecta o ponto, mas o caracteriza como parte de outro atleta.

Para resolvermos essas ocorrências, utilizamos um filtro de Kalman para estimar os pontos perdidos; e, considerando um processamento *offline* com todos os *frames*, utilizamos a interpolação dos pontos perdidos.

Foram observadas oscilações de alta frequência na linha trajetória das articulações, as quais sanamos utilizando a etapa de filtragem. Segundo Macedo e Santos (2019), uma forma de evitar as oscilações seria a utilização câmeras mais robustas quando da captação dos movimentos em formato vídeo:

> Borramento da imagem devido à perda de foco da câmera com o movimento, levando o sistema a considerar uma maior região de probabilidade para a articulação e, consequentemente, produzindo alterações ou oscilações não desejadas na coordenada estimada de um *frame* para o outro. Esse problema poderia ser minimizado usando câmeras mais robustas a movimentos no vídeo. (MACEDO; SANTOS. 2019, p. 31).

Considerando a biomecânica do movimento polichinelo, este, ao ser executado corretamente, descreve um movimento periódico. Para a trajetória vertical, em específico, esse movimento é aproximadamente um duplo arco de parábola cuja frequência define a cadência, ou seja, o abrir e fechar dos braços ocorre em função do tempo (sincronismo com as pernas).

Portanto, neste estudo, aplicamos um filtro de passa-baixas ideal (Kalman) de forma genérica estimando o movimento da partícula como um ponto descrevendo uma trajetória em função do tempo bidimensional. Apesar de termos obtido resultados satisfatórios na redução dos ruídos, faz-se necessário um delineamento experimental mais apurado para validar essa técnica.

Considerações finais

O presente trabalho propôs uma arquitetura de duas técnicas de *machine learning* para a análise do movimento polichinelo a partir de uma captura de vídeo. O primeiro passo foi a escolha das técnicas utilizadas. Nesse caso, foram escolhidas as técnicas de rede convolucional para processar a imagem e extrair a pose e rede recorrente LSTM para processar a série temporal de poses e analisar o movimento.

Com as técnicas escolhidas, o conjunto de treinamento foi gerado a partir de capturas de vídeo públicas retiradas da ferramenta YouTube. Foram escolhidos 90 vídeos de movimentos considerados satisfatórios e 10 de movimentos considerados insatisfatórios. Para a implementação das redes foi utilizada a linguagem de programação Python, juntamente com as bibliotecas OpenCV, Scikit-Learn e NumPy.

Para a etapa da rede convolucional, a rede OpenPose foi escolhida da literatura. Mesmo utilizando um modelo previamente treinado e consolidado na literatura, foi observado um certo nível de ruído que pode alterar os resultados dos movimentos desportivos. Nesse contexto, deduzimos que seja necessária a introdução de uma técnica que reduza o ruído, visando alcançar os resultados esperados, e o método de filtro de Kalman é um candidato a ser testado.

Dessa forma, conclui-se que a possibilidade de um Modelo em Rede Neural capaz de avaliar o movimento desportivo, por meio de uma máquina-servidor totalmente em *cloud*, possui uma acurácia satisfatória, mas carece de tratamento de ruídos para a análise da execução do movimento. Feito isso, tem-se a implementação da rede LSTM para a análise do movimento de fato.

Referências

BRITO, Eduardo Stein. **Transcrição Musical Automática do Instrumento de Bateria a partir de Vídeos**. 2019. 76 f. Monografia (Especialização) - Curso de Engenharia de

Computação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2019. Disponível em: http://hdl.handle.net/10183/198581. Acesso em: 7 set. 2020.

CAO, Zhe. *et al*. Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. *In:* **IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**, pp. 7291-7299, 2017. Disponível em: https://openaccess.thecvf.com/content_cvpr_2017/papers/Cao_Realtime_Multi-Person_2D_CVPR_2017_paper.pdf. Acesso em: 15 set. 2020.

CHRISTENSEN, Clayton M.; BOWER, Joseph L. Disruptive Technologies: Catching the Wave. **Magazine Harvard Business Review** (HBR), 1995. Disponível em: https://hbr.org/1995/01/disruptive-technologies-catching-the-wave. Acesso em 02/09/2020.

FONSECA, José Pedro Castro. **FPGA implementation of a LSTM Neural Network**. 2016. 86 f. Dissertação (Mestrado) - Curso de Engenharia da Computação, Faculdade de Engenharia, Universidade do Porto, Porto-PT, 2020. Disponível em: https://repositorio-aberto.up.pt/bitstream/10216/90359/2/138867.pdf. Acesso em: 15 set. 2020.

GONG, Wenjuan. *et al*. **Human pose estimation from monocular images**: a comprehensive survey. Sensors (Switzerland), v. 16, n. 12, p. 1-39, 2016. Disponível em: https://doi.org/10.3390/s16121966. Acesso em: 7 jan. 2021.

LEITE, Guilherme Vieira. **Detecção de Quedas de Pessoas em Vídeos Utilizando Redes Neurais Convolucionais com Múltiplos Canais**. 2020. 52 f. Dissertação (Mestrado) - Curso de Ciência da Computação, Instituto de Computação, Universidade Estadual de Campinas, Campinas, 2020. Disponível em: http://repositorio.unicamp.br/jspui/handle/REPOSIP/341843. Acesso em: 15 set. 2020.

MACEDO, Victor Oliveira Corrieri de; SANTOS, Joyce da Costa. **Análise cinemática automática usando OpenPose e Dynamic Time Warping com aplicações no remo**. 2019. 56 f., il. Trabalho de Conclusão de Curso (Bacharelado em Engenharia Eletrônica) — Universidade de Brasília, Brasília, 2019. Disponível em: https://bdm.unb.br/handle/10483/24937. Acesso em: 15 set. 2020.

MENACHE, Alberto. **Understanding motion capture for computer animation and video games**. illustrated. 2. ed. [*S.l.*]: Morgan kaufmann, v. 1, 2000.

SCHWAB, Klaus. **A Quarta Revolução Industrial**. Tradução de Daniel Moreira Miranda. 1. ed. São Paulo: Editora Edipro, 2019.

SIMAS, Gisele Moraes. *et al*. Utilizando visão computacional para reconstrução probabilística 3d e rastreamento de movimento. **VETOR-Revista de Ciências Exatas e Engenharias**, v. 17, n. 2, p. 59-77, 2007. Disponível em: http://repositorio.furg.br/handle/1/6862. Acesso em: 15 set. 2020.

TERLEMEZ, Ömer *et al*. Master Motor Map (MMM)—Framework and toolkit for capturing, representing, and reproducing human motion on humanoid robots. *In:* **2014 IEEE-RAS International Conference on Humanoid Robots.** Madrid. IEEE, p. 894-901, 2014. Disponível em: https://doi.org/10.1109/HUMANOIDS.2014.7041470. Acesso em: 15 set. 2020.